



Hitachi Data Instance Director

User's Guide

FASTFIND LINKS

[Software Version](#)

[Getting Help](#)

[Contents](#)

© 2013 - 2015 Hitachi, Ltd. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or stored in a database or retrieval system for any purpose without the express written permission of Hitachi, Ltd.

Hitachi, Ltd. reserves the right to make changes to this document at any time without notice and assumes no responsibility for its use. This document contains the most current information available at the time of publication. When new and or revised information becomes available, this entire document will be updated and distributed to all registered users.

Some of the features described in this document might not be currently available. Refer to the most recent product announcement for information about feature and product availability, or contact Hitachi Data Systems Corporation at <https://portal.hds.com>.

Notice: Hitachi, Ltd. products and services can be ordered only under the terms and conditions of the applicable Hitachi Data Systems Corporation agreements. The use of Hitachi, Ltd. products is governed by the terms of your agreements with Hitachi Data Systems.

By using this software, you agree that you are responsible for: a) Acquiring the relevant consents as may be required under local privacy laws or otherwise from employees and other individuals to access relevant data; and b) Verifying that data continues to be held, retrieved, deleted, or otherwise processed in accordance with relevant laws.

Hitachi is a registered trademark of Hitachi, Ltd., in the United States and other countries. Hitachi Data Systems is a registered trademark and service mark of Hitachi, Ltd., in the United States and other countries.

Archivas, Essential NAS Platform, HiCommand, Hi-Track, ShadowImage, Tagmaserve, Tagmasoft, Tagmasolve, Tagmastore, TrueCopy, Universal Star Network, and Universal Storage Platform are registered trademarks of Hitachi Data Systems.

AIX, AS/400, DB2, Domino, DS6000, DS8000, Enterprise Storage Server, ESCON, FICON, FlashCopy, IBM, Lotus, MVS, OS/390, RS/6000, S/390, System z9, System z10, Tivoli, VM/ESA, z/OS, z9, z10, zSeries, z/VM, and z/VSE are registered trademarks or trademarks of International Business Machines Corporation.

All other trademarks, service marks, and company names in this document or website are properties of their respective owners.

Microsoft product screen shots are reprinted with permission from Microsoft Corporation.



Contents

Preface	xiii
Intended audience	xiv
Software version	xiv
Related documents.....	xiv
Document conventions	xv
Accessing product documentation	xv
Getting help	xvi
Comments	xvi
 Architecture overview.....	 1-1
Overview	1-2
Policy orientated data management	1-2
Groups	1-3
Live monitoring	1-4
Data Instance Director data movement	1-5
Granular file I/O data capture	1-5
Data transfer modes	1-5
Active Sync.....	1-6
Bandwidth throttling.....	1-6
Network compression	1-6
Physical to virtual replication.....	1-6
Data Instance Director data transfer topologies	1-6
Data Instance Director Repository	1-7
Multi-tier deduplication	1-8
Metadata Store	1-11
Virtual machine protection	1-12
Remote office protection.....	1-12
Agentless backups.....	1-12
Storage hardware orchestration framework.....	1-12

Installation 2-1

Prerequisites.....	2-2
Hitachi Content Platform indexing requirement	2-2
Firewall configuration	2-2
Hitachi Block and File requirements.....	2-2
Installing Data Instance Director.....	2-4
Installation: Windows, Linux	2-4
Installation: cluster environment.....	2-11
Installing Data Instance Director on the controlling node	2-12
Installing Data Instance Director on remaining nodes	2-13
Adding the Hub service	2-13
Licensing the software.....	2-14
Upgrading Data Instance Director	2-15
Upgrade order.....	2-15
Upgrade alert message	2-15
Upgrading clients from Master	2-17
Licensing Data Instance Director.....	2-18
Manual license activation	2-18
Cluster environment license activation	2-19
Uninstalling Data Instance Director	2-20
Completing optional, post-installation configuration	2-21
Setting up an Outlook form for viewing stubbed emails	2-21
Installing an Outlook form.....	2-21
Creating a public folder and setting permissions for it	2-22
Publishing the Outlook form	2-23
Uninstalling an Outlook form.....	2-25
Stubbed email information and security	2-25
Forwarding email information	2-25

Getting started 3-1

Overview.....	3-2
Backup, Live Backup, and Continuous Data Protection (CDP).....	3-2
CDP 3-2	
Backup or CDP configuration.....	3-2
Policy definition	3-3
Data flow definition	3-6
Distribution and activation.....	3-8
File archive, file versioning	3-8
Version configuration.....	3-9
Policy definition	3-9
Data flow definition	3-11
Distribution and activation.....	3-12
Replication	3-13
Replication configuration	3-13

Policy definition.....	3-13
Data flow definition	3-14
Distribution and activation	3-17
Mirror replication from the Data Flow screen.....	3-18
Distribution and activation	3-19
Information security: Block and Track.....	3-20
Block configuration	3-20
Policy definition.....	3-21
Data flow definition	3-22
Distribution and activation	3-24
Track configuration.....	3-24
Policy definition.....	3-24
Data flow definition	3-25
Distribution and activation	3-27
Application backups: MS SQL, MS Exchange, and Hyper-V	3-27
Implementing a backup strategy.....	3-27
Backing up MS SQL, MS Exchange Server, or MS Hyper-V	3-28
Application policy definition: SQL	3-29
Data flow definition	3-31
Distribution and activation	3-33
Backing up a Clustered Share Volume with Hyper-V.....	3-33
Application policy definition: Hyper-V	3-34
Data flow definition	3-35
Distribution and activation	3-37
Restoring a single Hyper-V virtual machine (cluster)	3-37
Agentless nodes.....	3-38
Backing up a Windows share.....	3-38
Node creation	3-38
Policy definition.....	3-40
Data flow definition	3-42
Distribution and activation	3-44
Backing up a VMware ESX virtual machine	3-44
Node creation	3-45
Policy definition.....	3-46
Data flow definition	3-48
Distribution and activation	3-50
Browsing and recovering V2I snapshots	3-53
Node creation	3-53
Browsing snapshots	3-54
Restoring snapshots.....	3-55
Mounting snapshots	3-56
Backing up to a tape library.....	3-56
Node creation	3-56
Policy definition.....	3-59
Data flow definition	3-61

Distribution and activation.....	3-62
Archiving Microsoft Exchange mailboxes to HCP	3-63
Node creation.....	3-63
Policy definition	3-68
Data flow definition	3-70
Distribution and activation.....	3-72
Viewing stubbed information in Microsoft Outlook.....	3-73
Restoring an Exchange server and database	3-73
Creation of a recovery storage group/database	3-74
Creation of a backup database	3-75
Restoring to the original server but different location (Standalone+DAG)	3-76
Archiving file system files to HCP or to Azure	3-77
Node creation.....	3-77
Policy definition	3-82
Data flow definition	3-84
Distribution and activation.....	3-86
Snapshotting a Hitachi block storage volume	3-88
Node creation.....	3-89
Policy definition	3-90
Data flow definition	3-92
Distribution and activation.....	3-94
Restore	3-94
Replicating a Hitachi block storage volume	3-98
Node creation.....	3-98
Policy definition	3-100
Data flow definition	3-101
Distribution and activation.....	3-103
Proxy backup/repurposing of a Hitachi block storage replication.....	3-103
Node creation.....	3-104
Policy definition	3-105
Dataflow Definition	3-106
Distribution and activation.....	3-108
Replicating and snapshotting a Hitachi block storage volume	3-110
Synchronization groups.....	3-110
Node creation.....	3-111
Policy definition	3-113
Data flow definition	3-115
Distribution and activation.....	3-117
Restore	3-118
Snapshotting a Hitachi File Storage (Hitachi NAS) path	3-119
Node creation.....	3-120
Policy definition	3-121
Data flow definition	3-123
Distribution and activation.....	3-124
Restore	3-124

Snapshotting an Oracle database on Hitachi block storage	3-127
Node creation	3-129
Policy definition.....	3-132
Data flow definition	3-134
Distribution and activation	3-135
Mount	3-136
Revert.....	3-137
Managing alert notifications	3-138

User interface 4-1

Node Manager screen	4-2
Screen layout	4-2
Node icons	4-3
Node attributes.....	4-6
Authorizing nodes.....	4-6
Deauthorizing nodes	4-7
Removing nodes	4-8
Creating Agentless/Hardware nodes	4-8
Policy definition screen	4-9
Screen layout	4-9
Defining a policy	4-11
Data classification items	4-11
Operation items	4-13
Data classification items and attributes	4-16
Standard	4-16
Applications.....	4-17
Storage	4-23
File Type	4-24
Advanced	4-28
Data operation items and attributes.....	4-33
Standard	4-33
Archive.....	4-38
Advanced	4-41
Macros.....	4-42
Application backup.....	4-43
Exchange, MS SQL Server, and MS Hyper-V.....	4-44
Other applications.....	4-44
Saving policies.....	4-44
Loading and modifying existing policies.....	4-46
Schedule Manager dialog	4-47
Defining a new schedule	4-48
Modifying and deleting schedules	4-51
Opening and modifying a schedule.....	4-51
Deleting a schedule.....	4-51

Creating custom calendars.....	4-52
Data flow screen.....	4-53
Screen layout.....	4-54
Element placements	4-56
Connecting elements.....	4-56
Method 1 – connect.....	4-56
Method 2 – drag and connect.....	4-57
Moving and removing nodes	4-57
Properties section.....	4-58
Source node – policy assignment.....	4-58
Mover – data movement option and routing	4-59
Destination – data processing options	4-61
Mirror destination options	4-62
Repository destination options	4-66
Tape destination options.....	4-68
Hitachi Block storage destination options	4-69
Hitachi File storage destination options.....	4-71
Reverting the data flow	4-71
Compiling and distributing rules.....	4-73
Rules for compilation	4-73
Compilation errors	4-75
Rules distribution	4-75
Monitor screen.....	4-77
Window layout.....	4-77
Stats	4-79
Badges	4-80
Disconnected nodes	4-81
Node actions.....	4-81
Common actions.....	4-81
Source-only actions	4-82
Destination-only actions.....	4-83
Progress	4-84
Repository statistics.....	4-84
Monitor tabs	4-86
Monitoring node groups.....	4-86
Log Manager	4-87
Screen layout.....	4-88
Log Results area	4-91
Session Viewer	4-92
Purging (deleting) logs	4-93
Clearing (acknowledging) all error logs	4-94
Exporting logs.....	4-94
Notification Management dialog	4-95
Email setup.....	4-96
Configuring notifications	4-96

Reports screen	4-98
Storage screen	4-106
Repositories tab.....	4-107
Repository Usage tab.....	4-111
Managed Storage tab.....	4-113
Library Management tab	4-117
Media tab.....	4-118
Media Requests tab	4-119
Restore screen	4-119
Searching for data	4-120
Browse and search results fields	4-122
Viewing search results	4-124
Browsing files	4-126
Result limit.....	4-127
Timelines	4-128
Restoring data.....	4-130
Restore options.....	4-131
Tape Restore	4-133
VMware restore options.....	4-134
Exchange mailbox restore options.....	4-136
Restore destination	4-137
Restore progress.....	4-137
Topology examples	5-1
Many-to-one	5-2
One-to-many	5-4
Forwarding	5-6
Smart repository sync.....	5-8
Constructing a complex data flow.....	5-10
Parallel solution	5-11
Serial solution.....	5-12
Command-based task management	6-1
Reviewing guidelines about running command line tools.....	6-2
Backing up, restoring node configuration files: mastersettings.exe	6-2
Backing up node configuration files.....	6-2
Options and usage.....	6-3
About the mastersettings.cfg file	6-3
Examples.....	6-4
Restoring node configuration files.....	6-5
Changing a client machine's Master node: setconfig.exe	6-7
Changing the behavior of a Data Instance Director node: diagdata.exe	6-8
Reducing the repository's size: repomaint.exe.....	6-10

Customizing alert notifications	6-11
About the log message file	6-13
Testing	6-14
Creating a blacklist for stubbed files.....	6-15
Usage examples	6-15

Troubleshooting 7-1

Troubleshooting the Exchange Mailbox.....	7-2
Node creation	7-2
Failed to authorize node <name> (Invalid username/password)	7-2
Rules compilation	7-2
Compiler error 10256.....	7-2
Logs.....	7-2
Failed to initialize rules: [...] (0x8007052e) [...]	7-2
Rules resolved to no user mailboxes	7-2
No mailboxes to synchronize [...]	7-3
Failed to initialize rules: [...], database may have moved [...]	7-3
User without mailbox.....	7-3
Failed to initialize rules: [...] (0x80072030) [...]	7-3
The account does not have permission to impersonate the requested user.....	7-3
Connection did not succeed. Try again later.	7-4
Sender had problems while performing a full synchronization of [...].....	7-4
Troubleshooting HCP.....	7-5
Node creation	7-5
Couldn't connect to server or HTTP status 503 Service Unavailable.....	7-5
HTTP status 403 Forbidden	7-5
Couldn't resolve host name	7-5
Logs.....	7-5
413 Request Entity Too Large	7-5
In operation.....	7-5
Files, email bodies, or attachments are not being stubbed/removed.	7-5
After a file is opened, the file's access time is not modified.	7-6
Troubleshooting a down machine	7-7
The machine from which I've backed up using HDID has a damaged OS .	7-7
Restoring a Hyper-V virtual machine	7-8
I need to restore a single Hyper-V virtual machine.	7-8
Troubleshooting Hitachi NAS	7-9
Node authorization	7-9
Authorization Failure dialog: Failed to authorize/deauthorize node [...]	7-9
Directory clone.....	7-9
Handler call failed: [...] cannot overlap with source directory	7-9
Replication.....	7-9
API call failed	7-9
Troubleshooting Hitachi VSP	7-10

Snapshotting	7-10
Handler call failed: [...] failed to create snapshot pair [...]	7-10
Handler call failed: [...] snapshots that are in a different pool....	7-10
Handler call failed: refreshLogicalDevices Failed:: Failed run process "Permission denied with the Resource Group [raidcom.exe get ldev -ldev_id 2503 -l2500]	7-10
Troubleshooting Oracle Database snapshotting on VSP.....	7-11
OracleHandler [...]: Oracle Database Files and Redo in the [...] .	7-11
OracleHandler [...]: normal or high redundancy used; revert.....	7-11
OracleHandler [...]: mount for 'Oracle DB: [...]' finished [...]	7-11
Handler call failed: [...] RAC Lock Database failed, check if [...] .	7-11
Handler call failed: [...] ASM Diskgroups [...] mounted.....	7-11
Checks to perform when an online backup or mount fails	7-11

Preface

This document describes how to install and use Hitachi Data Instance Director.

This preface includes the following information:

- ☐ [Intended Audience](#)
- ☐ [Software Version](#)
- ☐ [Related Documents](#)
- ☐ [Document Revision Level](#)
- ☐ [Document Conventions](#)
- ☐ [Accessing Product Documentation](#)
- ☐ [Getting Help](#)
- ☐ [Comments](#)

Intended audience

This document is intended for system administrators and other users who are responsible for installing, configuring, and operating Hitachi Data Instance Director.

Software version

This document revision applies to Data Instance Director version 5.1.

Related documents

- *Hitachi Data Instance Director Software Version 5.1 Release Notes*, RN-93HDID000-02
- *Hitachi Data Instance Director Quick Start Guide*, MK-93HDID002-01
- *Hitachi Data Instance Director Disaster Recovery Reference*, MK-93HDID004-00

Document conventions

This document uses the following typographic conventions:

Convention	Description
Bold	Indicates text on a window, including menus, menu options, buttons, user input controls and labels. Example: Click OK .
<i>Italic</i>	Indicates a variable, which is a placeholder for actual text provided by the user or system. Example: copy <i>source-file target-file</i> Note: Angled brackets (< >) are also used to indicate variables.
screen/code	Indicates text that is displayed on screen or entered by the user. Example: # pairdisplay -g oradb
< > angled brackets	Indicates a variable, which is a placeholder for actual text provided by the user or system. Example: # pairdisplay -g <group> Note: Italic font is also used to indicate variables.
[] square brackets	Indicates optional values. Example: [a b] indicates that you can choose a, b, or nothing.
{ } braces	Indicates required or expected values. Example: { a b } indicates that you must choose either a or b.
vertical bar	Indicates that you have a choice between two or more options or arguments. Examples: [a b] indicates that you can choose a, b, or nothing. { a b } indicates that you must choose either a or b.

This document uses the following icons to draw attention to information:

Icon	Label	Description
	Note	Calls attention to important and/or additional information.
	Caution	Warns the user of adverse conditions and/or consequences (e.g., disruptive operations).
	WARNING	Warns the user of severe conditions and/or consequences (e.g., destructive operations).

Accessing product documentation

The Data Instance Director user documentation is available on the Hitachi Data Systems Portal: <https://portal.hds.com>. Check this site for the most current documentation, including important updates that may have been made after the release of the product.

Getting help

The Hitachi Data Systems Support Center staff are available 24 hours a day, seven days a week. If you need technical support, log on to the Hitachi Data Systems Portal for contact information: <https://portal.hds.com>.

Before calling the Hitachi Data Systems Support Center, please gather as much information about the problem as possible, including:

- The circumstances surrounding the error or failure.
- The exact content of any error messages displayed on the host systems.

Comments

Please send us your comments on this document to: doc.comments@hds.com. Include the document title and number, including the revision level located on the front page), and refer to specific sections and paragraphs whenever possible. All comments become the property of Hitachi Data Systems Corporation.

Architecture overview

This section describes the data management capabilities of Data Instance Director.

- ☐ [Overview](#)
- ☐ [Multi-Tier Deduplication](#)
- ☐ [Metadata Store](#)
- ☐ [Virtual Machine Protection](#)
- ☐ [Remote Office Protection](#)
- ☐ [Agentless Backups](#)
- ☐ [Storage hardware orchestration framework](#)

Overview

At the heart of the data management process is the Master Node. This feature is the controlling force that provides instruction to all other nodes. Unless specifically instructed to do so, the Master Node does not take part in the data flow, and nodes pass data among themselves as instructed by the Master Node.

The different node types can be defined as:

- *Master Node*: The machine that controls the actions of all other nodes on the system. Only one active Master Node is deployed in a network (a second passive Master Node can be installed on a Windows failover cluster).
- *Source Node*: This node (server, workstation or virtual machine) stores data to be managed. The source node has an Active Data Change Agent, which is responsible for monitoring the host file system, and performing the relevant actions defined by the policies. Nodes need to be configured as a source node if they need to transfer locally stored data, or implement data tracking, blocking and auditing functions.
- *Destination Node*: This machine is capable of receiving data. This machine might be the Data Instance Director Repository, a general-purpose system designated as the recipient of data in a replication configuration.

A machine can be simultaneously designated as multiple node types. A Master Node can also be either a source or destination node, or both.

Policy orientated data management

Data Instance Director is designed to allow the administrator to implement a business data protection process by using policies. Policies are mapped to machines or groups, and to the data management agents that implement those policies. A large, complex environment can be configured in a matter of minutes rather than days through the use of only two screens (Figure 1-1).

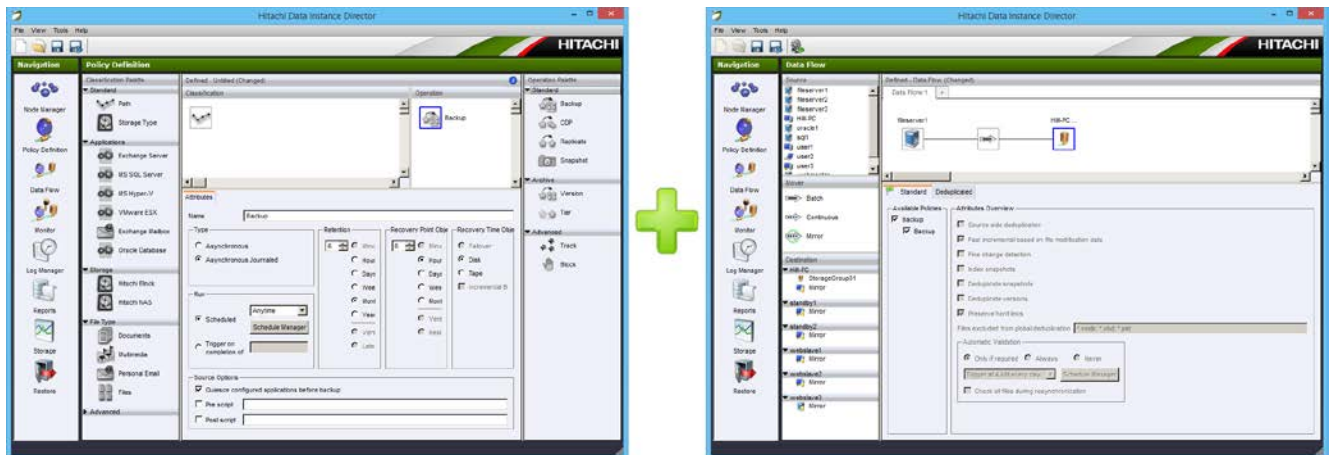


Figure 1-1 Building a policy

The **Policy Definition** screen allows you to design the business objective.

The **Data Flow** screen is used to define where those policies are applied, and the path of the data flow, reflecting the information life cycle.

Once this task is completed, you only need to apply that new topology. Data Instance Director compiles the policies into more granular instructions, and sends those instructions to each of the elements involved in data management activity.

Groups

Data Instance Director allows you to assign machines to groups through the **Node Manager** screen. A machine may belong to more than one group.

In the **Data Flow** screen, policies can be assigned to individual machines or to a group. This feature provides an easy method to manage large sets of machines, eliminating the need to create policies for specific machines.

Once the policies have been distributed, the whole environment can be monitored through the **Monitor** screen (Figure 1-2). You can expand a group in this screen to see activity within that group.

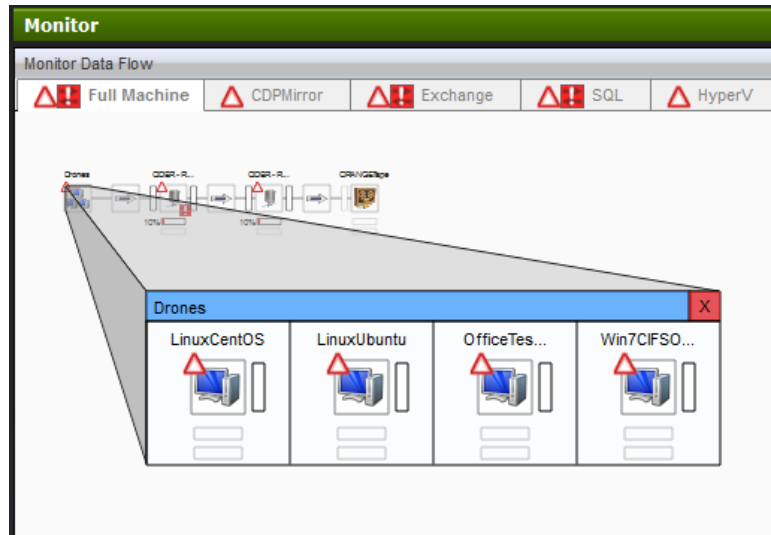


Figure 1-2 Node group expanded in monitor screen

Live monitoring

The replication activity can be monitored using the **Monitor** screen (Figure 1-3). Meters and icons provide the status.

- *Outgoing Network Traffic*: The amount of data being sent over the network.
- *Incoming Network Traffic*: The amount of data being received over the network.
- *Memory Cache Queue*: The amount of data being stored in memory awaiting transfer.
- *Disk Cache Queue*: The amount of data being stored on disk awaiting transfer.
- *Status Badges*: Indicating a node's current activity, status and error conditions.

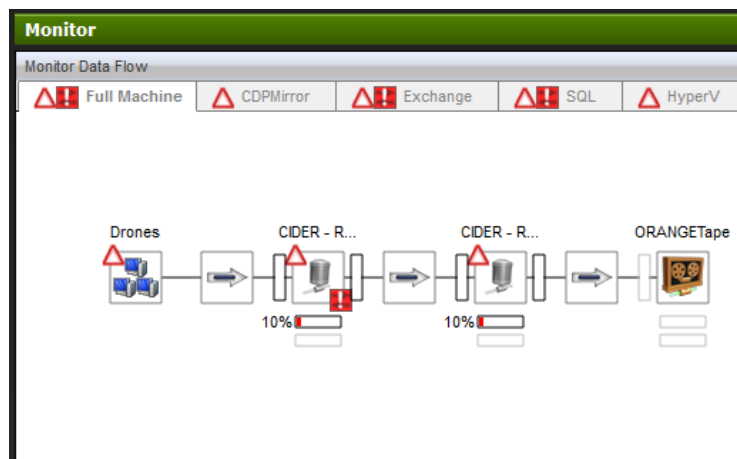


Figure 1-3 Monitor screen

Data Instance Director data movement

Data Instance Director creates and updates archive and backup files as changes occur, whenever possible. Traditional storage products repeatedly transfer the same data in bulk transactions, creating heavy loads on the source systems and putting undue stress on the networks. Data Instance Director sends the changed data as it happens. For example, if you receive an email message that is 10 KB, only 10 KB will be archived, not the entire 2 GB of an email archive (PST) file.

Granular file I/O data capture

The data capture system is highly granular, with the option to tightly define the type of data that is transferred within the policy classification. Unlike systems that replicate all contents on the volume, Data Instance Director replicates only specified data, thereby saving bandwidth and storage space.

In addition, unlike volume replication, the replication system only replicates changed bytes. So, if only 10 bytes are written to a file, then only 10 bytes will be replicated and not the entire 1K-4K volume block.

Data transfer modes

Asynchronous:

Change data is written to local memory on the source machine ready for transmission over the network to the destination machine. If the network connection is unavailable then this change data is cached locally to disk, ready for transmission once the network becomes available.

While this mode of operation is more efficient (because disk writes are not required on the source machine under normal conditions), a catastrophic failure of the source machine will cause the in-memory change data to be lost. A resynchronization between source and destination is required once the source machine is rebooted.

Asynchronous Journal:

Change data is written to local disk on the source machine ready for transmission over the network to the destination machine. If the network connection is unavailable then this change data is cached locally to disk, ready for transmission once the network becomes available.

This mode of operation is less efficient (because disk writes are required on the source machine under normal conditions), however a catastrophic failure of the source machine will only cause the very last change to be lost. Minimal resynchronization is required once the source machine is rebooted and therefore this mode is more robust.

Active Sync

Initially, a point of consistency must be established between the source and destination nodes. This step must be done during the initial setup, when both nodes are synchronized. If the nodes have been disconnected for a prolonged period, they must be resynchronized.

Traditionally, this synchronization required the file system on the source node to be static, meaning the applications had to be stopped while the synchronization/resynchronization process took place.

Data Instance Director offers an Active Sync system that allows synchronizing while the applications are running.

Active Sync also analyzes the data before it is sent. Data is not sent if it is identical to data that is already in the repository. The repository adds a reference to this data to ensure that only one copy is stored.

Bandwidth throttling

Data Instance Director allows you to control the amount of bandwidth that data transfers use so that a network connection can be used simultaneously with other applications in the environment. Throttling can be configured to be variable, allowing reduced bandwidth during the day and increased bandwidth during the evening.

If the amount of data that needs to be replicated continuously exceeds the current bandwidth, data is kept on the source until the bandwidth increases.

Network compression

Optionally, data sent across the network can be compressed. Compression decreases network utilization, but increases CPU usage.

Physical to virtual replication

Data Instance Director provides support for physical to virtual machine replication. Using VMware (or other virtual machine technologies), users can replicate numerous physical or virtual servers to other physical or virtual servers. This replication method increases efficiency and helps users work toward consolidated, virtualized infrastructure goals.

Data Instance Director data transfer topologies

Data Instance Director offers many different topology capabilities, ranging from One-to-One, Many-To-One, Many-To-Many, etc. Matching topology styles with Data Movers (Batch, Continuous, and Mirror) offers flexibility to meet your most challenging needs. See the section titled [‘Topology Examples’](#) for detailed information.

Data Instance Director Repository

The Data Instance Director Repository is a multi-functional object secondary storage system that can simultaneously perform multiple types of storage operations. Legacy storage systems are very different and inefficiently create duplicate copies of data. Some systems try to de-duplicate data (remove duplicates), but at best this is only effective within one subsystem. Figure 1-4 demonstrates the problem with legacy point solutions.

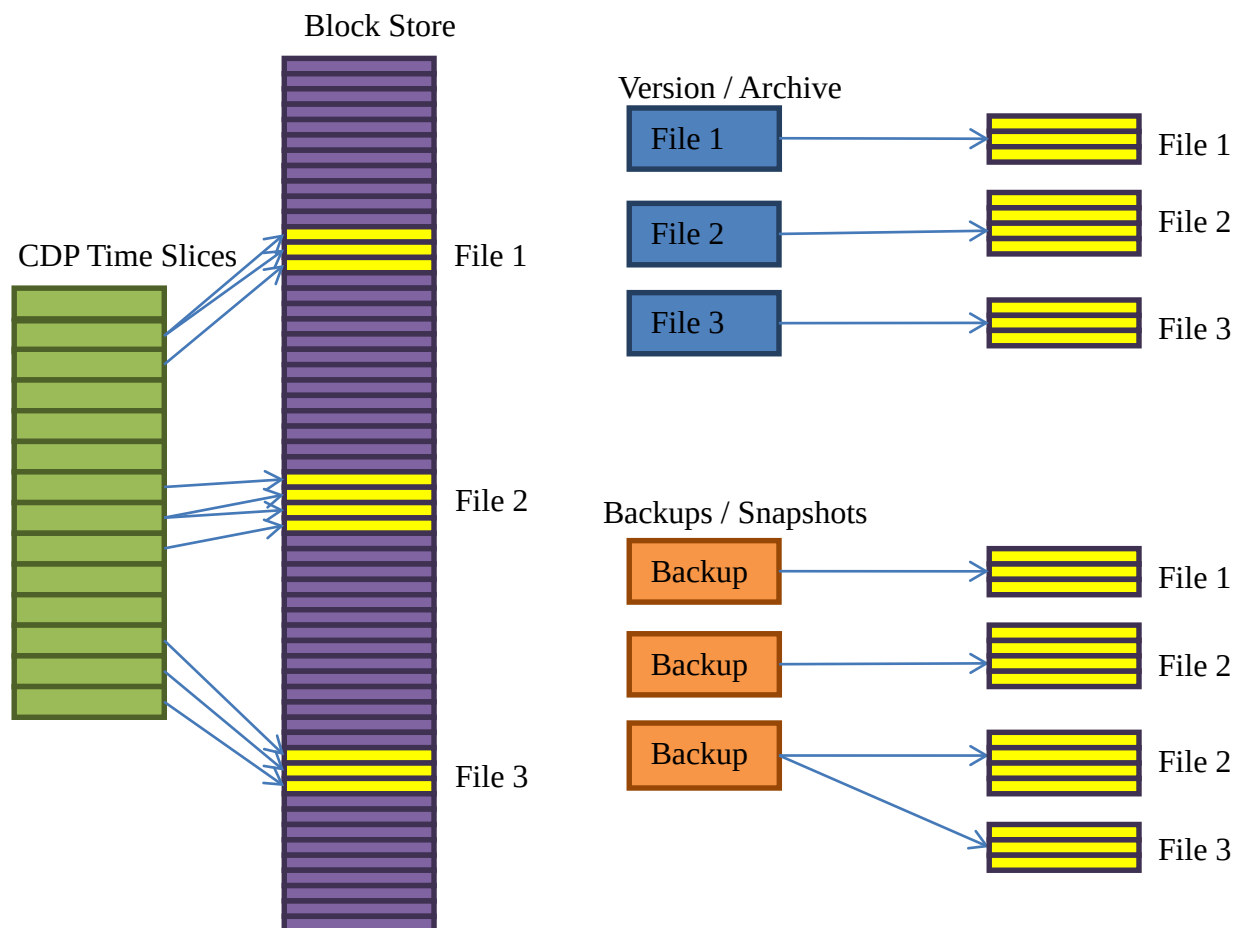


Figure 1-4 Legacy point solutions working independently

Multi-tier deduplication

Data Instance Director Repository's ability to provide backup, archiving and continuous data protection (CDP) storage enables it to employ several data reduction techniques delivering massive storage cost savings. Traditional storage systems require separate storage for backup, archiving and CDP. This model requires that data reside in several locations. With Data Instance Director, if a policy targets backup, archiving, and CDP to the same repository store, then the data is stored only once as seen in Figure 1-5. The data also is only sent once, reducing the amount of data that is moved around the system.

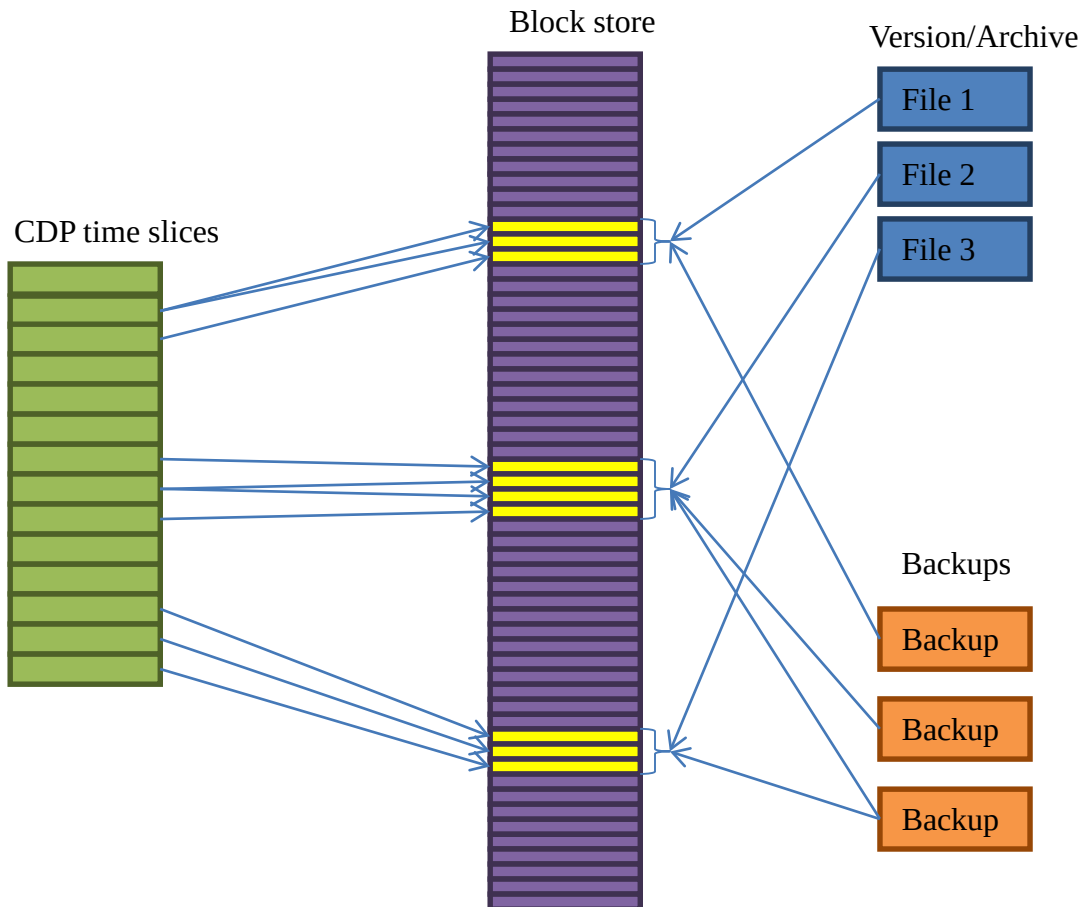


Figure 1-5 Data Instance Director Repository

Duplicate Transfer Avoidance: During initial synchronization Data Instance Director intelligently analyzes data and will not transmit any duplicates of data stored.

Byte Level Change Detection: Data Change Detection eliminates continuous scans, saving time, performance and capacity hits on all machines. Only changed bytes need to be transferred.

Multi-Level Single Instance Storage: Data Instance Director's Unified Repository stores only a single instance of a file, no matter if it is a CDP, archive, file or email message.

Post Process Deduplication: A final step runs post-processing deduplication algorithms within a repository across all data sets from all machines.

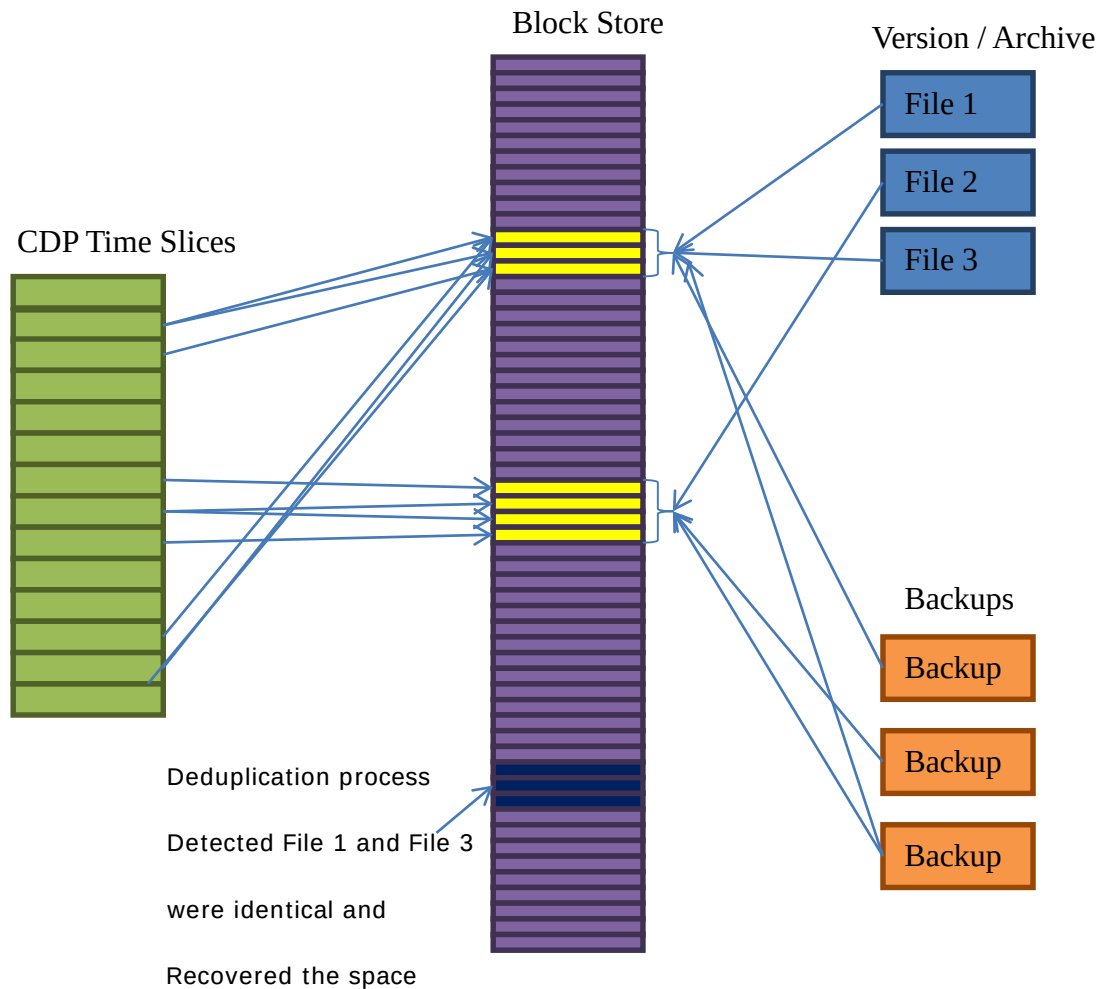


Figure 1-6 Data Instance Director Repository deduplicated

The repository detects duplicated data that might appear across different datasets and de-duplicates this data, further reducing the storage space required (Figure 1-6).

The Data Instance Director Repository receives real-time updates: It receives updates as the data from the source is changed. These changes are at the byte level. For example, if a row from a database is modified, only that row is sent, not the whole database. This capability is not limited to CDP, but is available also for backup and archiving.

The Repository enables the following:

- *Live Backup*: The repository is updated with live updates, which reduces the need for file system scans and bulk data transfers. During the initial, full resynchronization, the file system is scanned, a bulk data transfer occurs, and an initial fast incremental resynchronization is completed. If the destination store becomes out-of-sync (for example, during a Data Instance Director software upgrade, or if data stops being captured), then a fast incremental resynchronization should be completed. (By default, fast incremental resynchronization occurs at 4:00 AM).
- *Batch Backup*: The repository is updated periodically using scheduled resynchronizations. This method involves a scan of the source machine's file system: With live capture, the changed *bytes* are transferred. With batch resynchronization, the changed *blocks* are transferred. By default, the block size that is transferred is 2 MB; it is 16 KB if fine change detection is configured in the store template. An entire file is transferred if it is less than one block. Batch backup is useful for data that does not change often, such as data contained on the operating system disk. Fine change detection should be used with the batch mover if there are frequent, widely distributed changes to files, such as Microsoft SQL and Microsoft Exchange files.
- *Archiving / Versioning*: The repository has the ability to store versions of files. With the appropriate retention rules, this feature provides granular archiving. Unlike backups that are taken at set intervals, versions are taken when a file changes. For instance, when applied to office documents, every time a document is saved, the version is retained and indexed. This action allows a history of a file to be generated showing how the file changed during the generations. To fully facilitate audits and e-discovery, the extra information is retained, providing the file owner as well as which user and which program modified the file.
- *CDP*: Continuous data protection provides a method of capturing the state of a system on a very granular basis. CDP shares much of the functionality of Live Backup, except data is retained for a shorter period of time and is not indexed by the Metadata Store. Typically, CDP and Live Backup are used in conjunction.

Each repository has its own metadata store. A site might have more than one repository. The repositories work together to allow searches or restores that pull data from several repositories. On large sites where many repositories are required, the best practice is to assign repositories according to the type of data each is storing. Keeping similar data going to the same repository provides the greatest data reduction savings.

Metadata Store

The Metadata Store (MDS) is a metadata index for items held in the repository. Every repository store group has its own MDS and this is an integral component of a repository. Unlike traditional storage products, this feature is not a database, but is an indexing system similar to those used in search engines. Unlike a database, it does not have the limitation of requiring fixed field definitions. Fully indexed fields can be added to enrich the metadata at any time, without the need to restructure the database. The records do not need to reserve space for fields that are not used.

The MDS is closely coupled to the repository, but it also can be used to index data that is not in its local repository.

The MDS is fed events from the repository, which informs the MDS of new versions, new repository snapshots and block and file storage system snapshots, and so on. The MDS retrieves the metadata that repository has for a given object, and also appends its own metadata corresponding to the policy set. For objects that have retention, a retention record is attached to the object, keeping the object alive. You may have more than one retention record for a given object. Typically, this occurs if you have two or more policies that intersect. For example, you may have a policy for all users to keep files for a year and a policy for accounts that requests documents to be held for seven years. The accounts documents are only stored once, but have two retention records. The documents are not removed from the system until both retention records have expired.

For performance optimization, the MDS has events queued to it, allowing the repository to function without waiting for the MDS to complete its indexing workload. During particularly busy periods, a delay might occur between an event (such as a repository snapshot or a file version) and the time when the metadata store makes this available in the search results.

The MDS indexes both version and repository snapshot contents. When a snapshot is produced, the MDS indexes the items that are new or changed in that snapshot. This indexing allows a user to search for both files captured in version (archive) and snapshots. If a file appears in many snapshots but has not changed or moved, it will only appear once in the search results. The MDS also offers you the ability to browse a snapshot, providing a file system view of the data inside.

Versioning has more extensive functionality when it relates to how a file is indexed. Versioning has the notion that a file has generations and the MDS keeps a track of how a file changes and creates a relationship between its previous versions. A version search groups the search results so that files that are related are grouped together. If you create a file called New Microsoft Word Document.doc, make some changes, rename it to Fred.doc, and make more changes and then rename it to Fred Smith annual review 2009.doc, you are able to see how all these files are related.

Virtual machine protection

Data Instance Director automatically detects virtual machines (VM) by identifying them as a virtual node, and treats them the same way as physical machines when an agent is loaded onto the VM, providing real time data protection and monitoring to any virtual node. Data Instance Director can also backup VMDK (VMware vSphere and VMware vCenter) and VHD (Hyper-V) files backed up through the host, which are ideal for DR and flexible restore purposes. You will see mentions of virtual node usage throughout this user manual.

Remote office protection

Data Instance Director's distributed architecture provides easy setup with remote office environments and the ability to connect to machines easily across a WAN and to set up a Repository for remote office data protection and recovery, and provides easy setup of many different topologies for remote systems and offices.

Agentless backups

Data Instance Director is able to integrate data from agentless nodes into its backup policies. Files from Windows shares, Samba shares, Network-Attached Storage, and any other implementation of the Common Internet File System (CIFS) can be unified and archived alongside files from Data Instance Director-managed systems.

Storage hardware orchestration framework

Data Instance Director is able to control the snapshotting and replication of data stored on Hitachi block and file storage hardware (such as VSP and Hitachi NAS).

Data Instance Director's hardware orchestration framework is able to create snapshots and replications in an application consistent state by interacting with the production application (Oracle, Exchange and SQL Server) and the storage array hardware to ensure that data is in a consistent state during snapshotting or replication; it does this with minimal disruption to the production system. The orchestration framework discovers the underlining hardware paths for applications and provisions space for snapshots and replications when the policy is activated.

The orchestration framework is extensible, allowing other managed storage technologies and applications to be integrated in future releases. These capabilities are all easily accessible via Data Instance Director's intuitive, graphical policy and dataflow definition GUI.

Guidance on implementing policies using the hardware orchestration framework are described in the following sections:

[Hitachi Block and File requirements](#)

[Snapshotting a Hitachi block storage volume](#)

[Replicating a Hitachi block storage volume](#)

[Replicating and snapshotting a Hitachi block storage volume](#)

[Snapshotting a Hitachi File Storage \(Hitachi NAS\) path](#)

[Snapshotting an Oracle database on Hitachi block storage](#)

[Hitachi Block storage destination options](#)

[Hitachi File storage destination options](#)

Installation

This chapter describes the software installation, license activation, and initial configuration tasks.

- ❑ [Prerequisites](#)
- ❑ [Installing Data Instance Director](#)
- ❑ [Upgrading Data Instance Director](#)
- ❑ [Licensing Data Instance Director](#)
- ❑ [Uninstalling Data Instance Director](#)
- ❑ [Completing Optional, Post-installation Configuration](#)

Prerequisites

Before you install Data Instance Director, confirm that your site's hardware and software meet the requirements that are outlined on the Data Instance Director product website at <https://www.hds.com/products/storage-software/hitachi-data-instance-director.html>.

Hitachi Content Platform indexing requirement

If you plan to use the Hitachi Content Platform (HCP) with Data Instance Director, then the following **Search Facility Settings** must be selected within the **Settings** tab of the HCP, for the **Metadata Query Engine** (MQE):

- Enable indexing
- Enable indexing of custom metadata

Click **Update MQE Settings** to reflect the changes.

Firewall configuration

If a third party firewall is installed within your network, when processes are started as part of the Data Instance Director software installation, firewall warnings may be generated.

Configure the firewall to allow access for all Internet-connected nodes (or nodes that are separated by firewalls) on one open port: 30304 (TCP and UDP).

Additional ports that may need to be opened are:

- The HCP node (30306, by default)
- The Data Instance Director Microsoft Exchange notification client (30500)

Hitachi Block and File requirements

If you plan to perform snapshotting and replication of Hitachi Block and File volumes then the following prerequisites are required:

- General:
 - o A machine must be assigned as a proxy node that controls the storage device. This proxy node must be a Windows machine with:
 - The HDID repository capability installed
 - CCI (RAID Manager) installed
 - Access to a command device set up as follows:
 - Security disabled
 - User authentication enabled
 - Device group definition disabled

- o A dedicated user for HDID must be created in the Administration user group for the storage device
 - o If using virtual machines, they must use NPIV
 - o All paths or application data to be snapshotted or replicated must exist on the same block hardware device
- Hitachi Block:
 - o The following licensed features are required:
 - Dynamic Provisioning
 - Storage Navigator
 - ThinImage
 - ShadowImage
 - TrueCopy
 - Universal Replicator
 - Remote Replication Extended (for 3DC scenarios)
 - o Primary volumes must be set up
 - o For application consistent snapshots, the application must be installed and configured to use HDS block P-Vols
 - o The username specified in the VSP node's credentials must be a member of the Administration User Group on the VSP
 - o The password for authorizing a VSP node must contain only useable CCI command characters: A-Za-z0-9'-./:@__
 - o Must have adequate shared memory – see Provisioning and Technical Guides
 - o Pools must be created using Hitachi Storage Navigator prior to selecting the Target Storage in Data Instance Director:
 - For Thin Image (TI) the TI Pools must be set up
 - For ShadowImage (SI) the Dynamic Provisioning Pools must be set up
 - For TrueCopy (TC) and Universal Replicator (UR) the Dynamic Provisioning Pools must be set up
 - o Journals must be set up for UR – Refer to the UR Guide
- Hitachi File:
 - o The following licensed features are required:
 - FileClone
 - NFS
 - Replication
- Windows Nodes:
 - o None
- Linux Nodes:
 - o Needs 'rescan-scsi-bus.sh' CLI tool (see http://sg.danny.cz/sg/sg3_utils.html)
 - o Needs 'Multipath' CLI command
 - o Anything done directly on Hitachi Block (i.e. outside of HDID) is not supported by HDID

Installing Data Instance Director

Data Instance Director is distributed as an installation package for Windows and Linux.

The installation process allows you to set the installation destination and select which components are to be installed. The same installation program is used for the Master node and Client node installations. The following sections provide installation instructions for both nodes. Instructions for installing Data Instance Director in a Microsoft Windows cluster configuration are also provided.



Notes:

- Data Instance Director 32-bit version is supported for installation on a 32-bit operating system. If you attempt to install a 32-bit Data Instance Director on a 64-bit operating system, then you will be prompted to use the 64-bit installer instead.
- Data Instance Director 64-bit version is supported for installation on a 64-bit operating system. If you attempt to install a 64-bit Data Instance Director on a 32-bit operating system, then you will be prompted to use the 32-bit installer instead.
- For the master and repository installation, Data Instance Director is supported on 64-bit operating systems, only. Source node installation can be done on either a 32-bit or 64-bit operating system.
- Only one active Master node can be installed in an environment (a second standby Master can be installed on a Windows failover cluster – see [Installation: cluster environment](#)). Install the Master node before installing other types of nodes.
- Windows 7 and Windows Vista are considered to be workstation operating systems, and by default, have additional security measures known as User Account Control (UAC). Installation of the Data Instance Director master components on a Windows 7 or Vista machine requires that the UAC system be disabled. Ideally the Master node should be installed on server-based operating systems, such as Windows 2003 or Windows 2008.

If you must install Data Instance Director master components on a machine running Windows 7 or Vista, you can disable UAC in the **Control Panel**:

Go to **User Accounts** and select **Turn User Account Control on or off**.

Installation: Windows, Linux



Note: If this is a new installation on a 32-bit operating system, then the installation will be for a client node, only.



Note: When installing Data Instance Director on a Solaris node, the GCC runtime library (version 4.5 or later) must be installed first. This is distributed with Data Instance Director. Type the following command (here version 4.5 is specified but the version number may increase in future releases):

```
pkg install gcc-45-runtime
```

1. In Windows, double-click HDID-5.xxxx.exe in the directory where you downloaded it.
2. For a Linux Installation, ensure you have execute permissions and type the command:

```
chmod og+rx ./HDID-5.xxxx-Linux.run
```

The installation process is either presented via a set of dialogues, but if no windowing system is available on Linux the installations will display the same information via the text mode shell.

When Setup is started, it checks for previous installations of Data Instance Director. If a previous installation is found, Setup launches the upgrade process. For more information, see [Upgrading Data Instance Director](#).

3. Click **Next>** in the **Setup** welcome dialog (Figure 2-1).

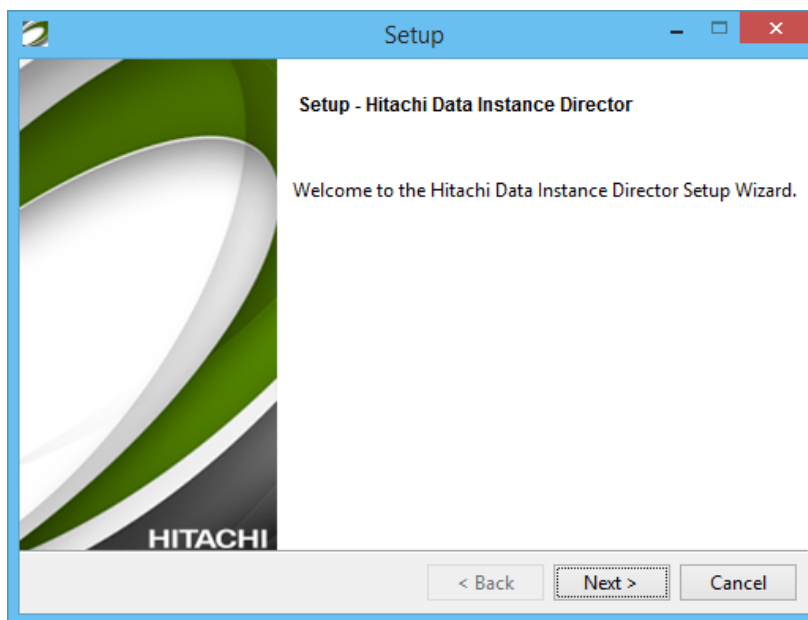


Figure 2-1 Data Instance Director installer

4. When prompted, accept the License Agreement, and click **Next>** (Figure 2-2).

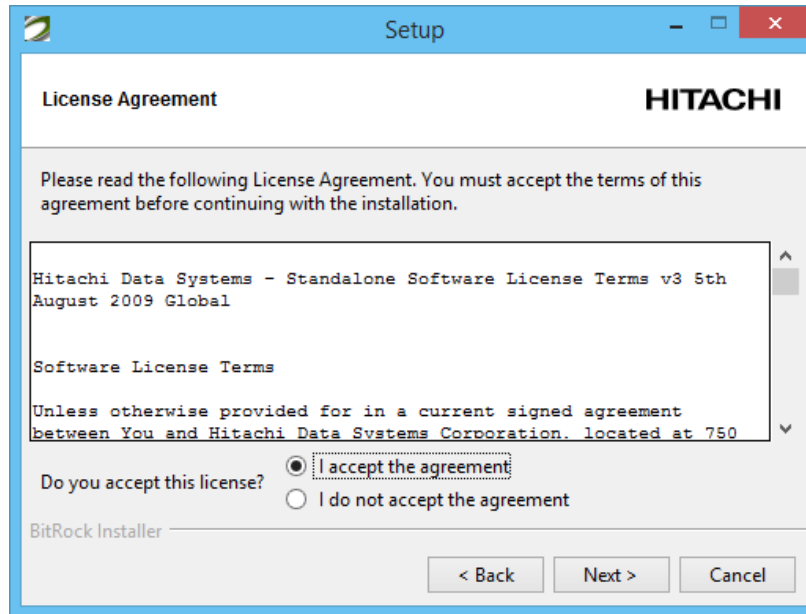


Figure 2-2 Accepting the License Agreement

5. The default installation directory for Data Instance Director is entered automatically:

- For Windows installations - C:\Program Files\Hitachi\HDID
- For Linux installations - /opt/Hitach/HDID

To install Data Instance Director in a different location, enter the path in the

Installation Directory edit box or use the folder browser  button. Click **Next>** (Figure 2-3).

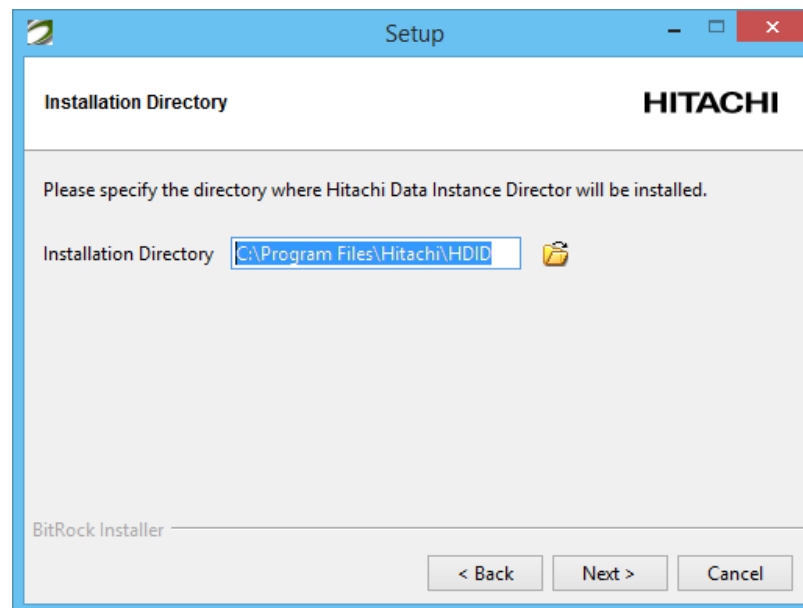


Figure 2-3 Installation Directory



Note: If this is a new installation on a 32-bit operating system, then the installation will be for a client node only, and steps 6, 7, and 8 are not applicable.

6. Select the Data Instance Director components to install, and click **Next>** (Figure 2-4). By default, **Client** is selected.

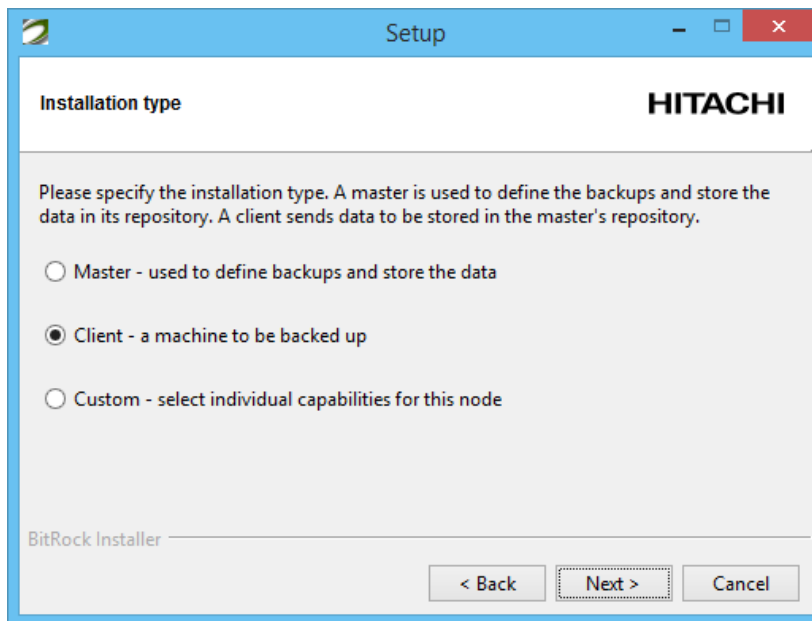


Figure 2-4 Installation type

- **Master:** (Only available on Windows) Installs the components required to make the current system the Master Node. Only one active Master Node can be installed in a network (a second standby Master can be installed on a Windows failover cluster – see [Installation: cluster environment](#)). The main Data Instance Director user interface is installed and only accessible on the Master Node. The Data Instance Director environment is set up and managed through the Master Node. A Master can also act as a Client and a Repository, enabling the node to have data captured from it and be able to store data from Source Nodes in a Repository.
- **Client:** This option enables the node to be a Source Node, and allows data to be captured on that node. This is required if the machine will send data to another Data Instance Director enabled system. A Client will also be able to have block or track operations performed on it.
- **Custom:** This option allows specific components to be installed, allowing the Data Instance Director installation to be tailored to a specific need.

If the **Master** option was selected, skip step 7.

If the **Client** option was selected, skip steps 7 and 8.

7. (**Custom** installations only) Select the Data Instance Director components to install, and click **Next>** (Figure 2-5). By default, **Data Source** is selected.

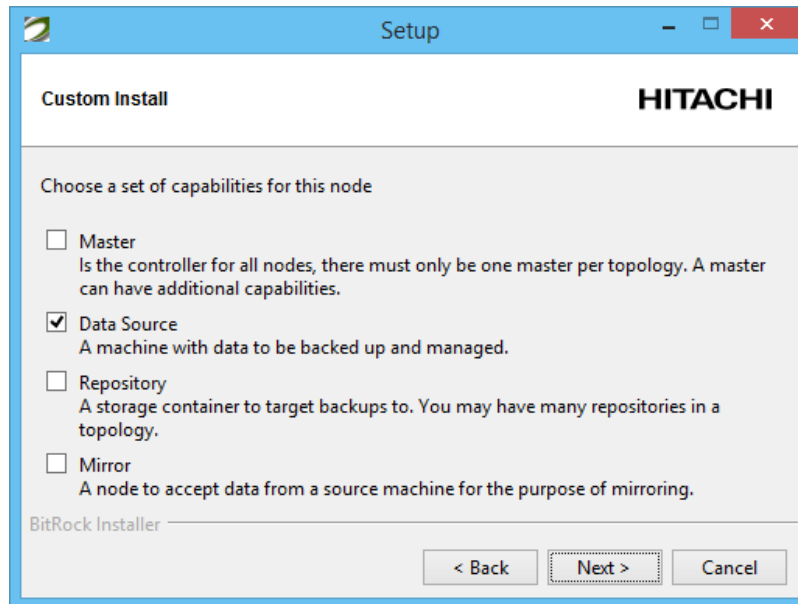


Figure 2-5 Custom Install

A Custom node can be any combination of the following components:

- **Master:** (Only available on Windows) Installs the components required to make the current system the Master Node. Only one active Master Node can be installed in a network (a second standby Master can be installed on a Windows failover cluster – see [Installation: cluster environment](#)). The main Data Instance Director user interface is installed and only accessible on the Master Node. The Data Instance Director environment is set up and managed through the Master Node.
- **Data Source:** This option enables the node to be a Source Node, and allows data to be captured on that node. This is required if the machine will send data to another Data Instance Director enabled system. The Data Source component is also required to perform block or track operations. Application aware hardware snapshotting and replication also requires this component.
- **Repository:** (Only available on Windows) This option enables the node to store data received from source nodes in a repository. Any machine used as a hardware proxy node for storage-based protection must also have this component installed.
- **Mirror:** This option allows the node to be designated as a Replication Node, making it suitable for replication operations (mirroring).

8. If a Repository is being installed, the installer prompts you for a **Repository Directory** (Figure 2-6). This directory allows the Repository and the data stored within it to be separated from the other Data Instance Director components and files which are installed.

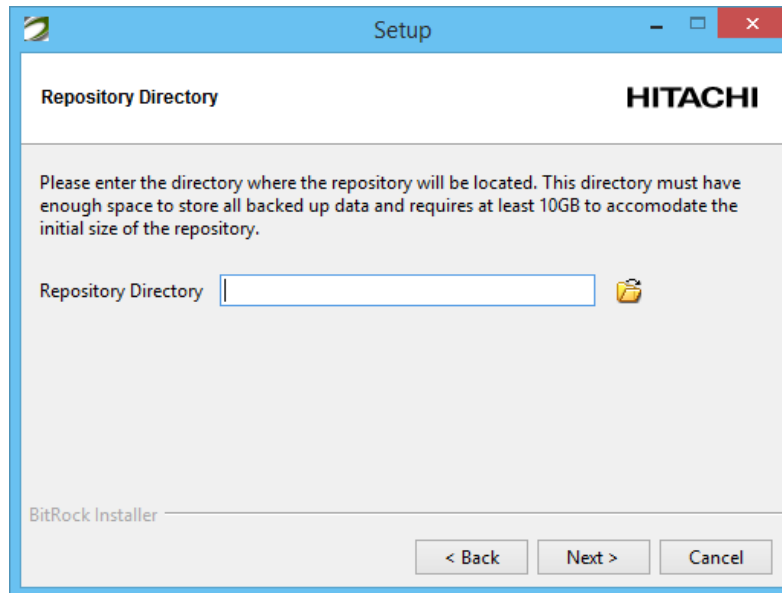


Figure 2-6 Repository install destination



Note: Because the Repository grows in large blocks, even for a small amount of stored data, ensure that the Repository Destination has a minimum of 200 GB of free space. For an ISM only (i.e. where no repository backups are required) 10GB of free space is required.

Specify a node name to be used within Data Instance Director (Figure 2-7). By default, the name is set to the machine's host name. This name is only for running Data Instance Director, and will not change the name set in the operating system. If it is known that the node will be operating over a non-secure network, such as the Internet, then enabling the **Internet connected node** option is recommended. This will encrypt transmitted data as an extra security precaution.

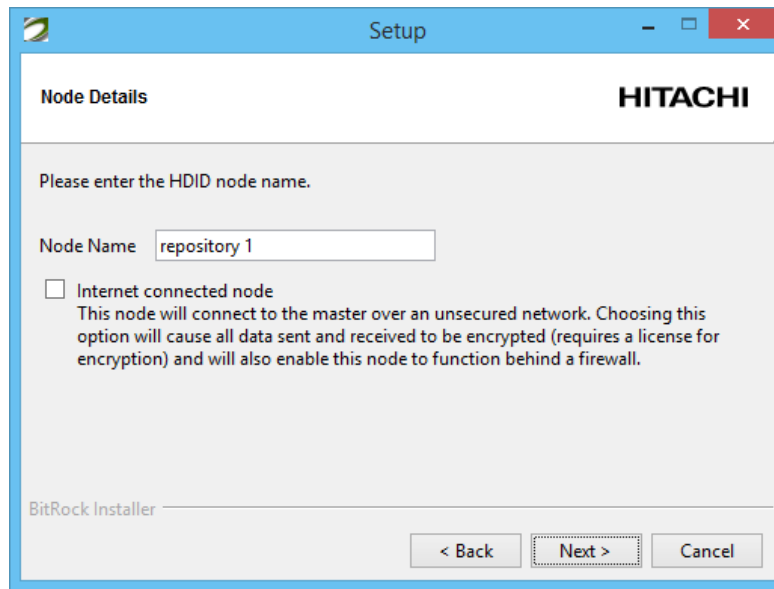


Figure 2-7 Node Name

9. If the Master component was not selected, the node is configured as a Client Node. Enter the **Master Hostname or IP** address or a DNS resolvable name of the Master node (Figure 2-8).

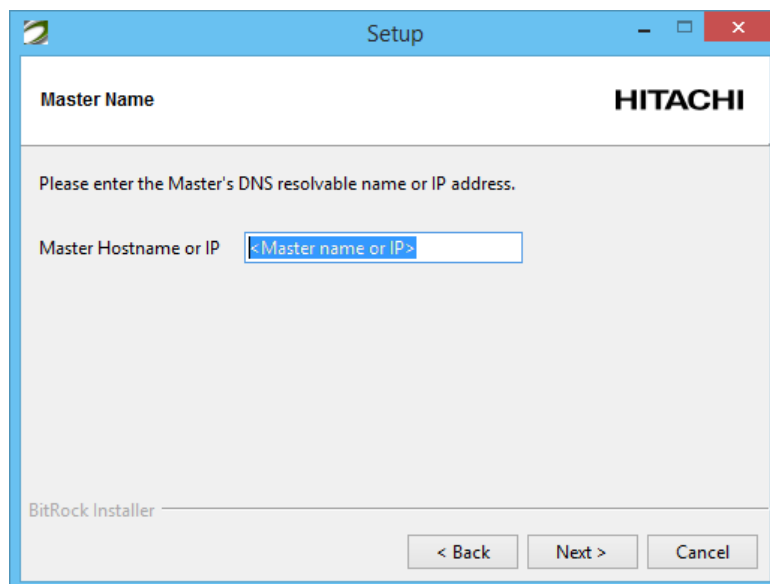


Figure 2-8 Enter Master IP address or DNS resolvable name

10. Indicate whether you want to place a Data Instance Director GUI shortcut on the desktop, then click **Next>** (Figure 2-9).

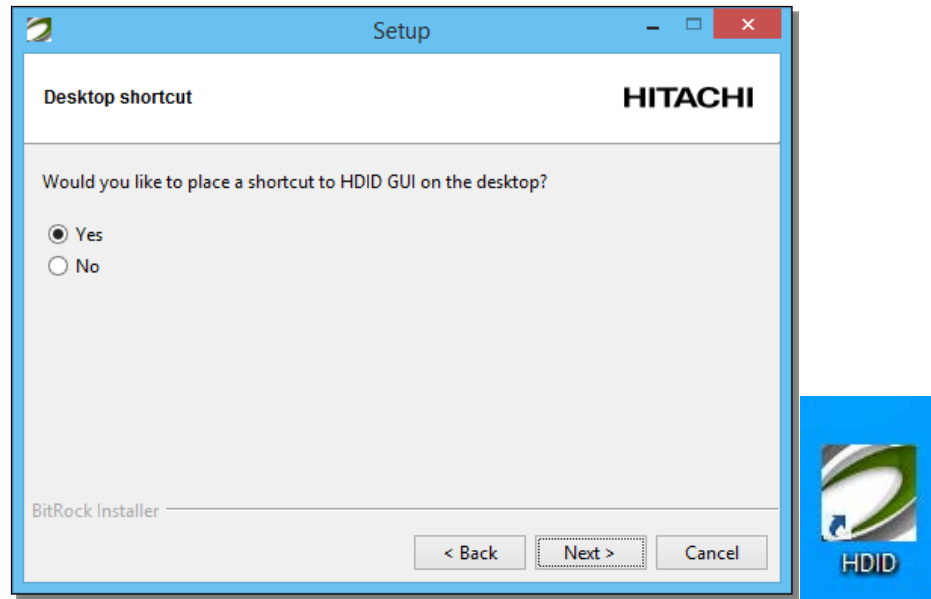


Figure 2-9 Create shortcut

- 11.** Click **Next>** when Setup indicates that it is ready to begin the installation. Data Instance Director files are copied to the designated directories and the necessary components installed.
- 12.** Click **Finish** when the message is displayed stating that the installation is complete.

You do not need to restart your machine after the installation is complete.

At the end of the file copying process, the installer starts all installed and required Data Instance Director components on the system.



Note: If a third party firewall is installed on the network, then starting these processes generates firewall warnings. Configure the firewall to allow access for all

Data Instance Director components on one open port: 30304 (TCP and UDP). Additional ports that might need to be opened are:

- The HCP node (30306, by default)
- The Data Instance Director Microsoft Exchange notification client (30500)

Installation: cluster environment

Data Instance Director can be installed on the servers which are in a Windows cluster. Before starting the installation, review the following:

- Ensure that all servers meet the software prerequisites that are outlined in the first section of this chapter, [Prerequisites](#).

- Confirm that a cluster environment is set up and working correctly.
- Ensure that you have all required license keys for Data Instance Director: A separate license key for each master node is required.

The installation for a cluster environment involves the following main tasks:

1. Installing Data Instance Director on the controlling node.
2. Installing Data Instance Director on all other nodes.
3. Adding the Hub service to the cluster.
4. Licensing the software.

Installing Data Instance Director on the controlling node

Start with the node that is in control (the node which has access to the shared disk). If needed, you can identify this node by launching the cluster manager.

To install Data Instance Director on the controlling node:

1. Double-click on HDID-5.xxxx.exe in the directory where you downloaded it on the server. Then, click **Next>** in the **Setup** welcome dialog.
2. Read and accept the license agreement, then click **Next>**.
3. Change the installation directory to the shared drive. For example, if your shared drive is Q:\ then the installation path might be:
Q:\Hitachi\HDID.



Note: Make a note of the specified path, as you will need to reference this later in the installation process.

4. Click **Next>**.
5. Select the **Custom** installation type, then click **Next>**.
6. Select **Master**, **Repository**, or both as these are the only supported clustered roles. Then, click **Next>**.
7. If the Repository role has been selected, select the location of the repository, then click **Next>**.



Note: The repository must be located on the shared drive; for example:
Q:\Hitachi\HDID

Also, make a note of the specified path, as you will need to reference this later in the installation process.

8. Change the default Node name as all cluster nodes will have the same name. Use a generic name such as Cluster, Master, or Repository. Click **Next>**.
9. Select an option for a desktop shortcut, then click **Next>**.
10. Click **Next>** to start the installation.
11. Click **Finish**, then shut down the machine to invoke failover.

Installing Data Instance Director on remaining nodes

To install Data Instance Director:

1. Double-click on HDID-5.xxxx.exe in the directory where you downloaded it on the server. Then, click **Next>** in the **Setup** welcome dialog.
2. Read and accept the license agreement, then click **Next>**.
3. Specify the same installation directory that you indicated in step 3 of the first task, 'Installing Data Instance Director on the Controlling Node'.
4. Select the **Custom** installation type, then click **Next>**.
5. Specify the same clustered role that you indicated at step 6 of the first task, 'Installing Data Instance Director on the Controlling Node'.
6. If the Repository role was selected at step 7 of the first task, 'Installing Data Instance Director on the Controlling Node', then select the same location for the repository. Click **Next**.
7. Specify the same node name that you indicated at step 8 of the first task, 'Installing Data Instance Director on the Controlling Node'.
8. Select the option for a desktop shortcut, then click **Next**.
9. Click **Next** to start the installation.
10. Click **Finish**, then shut down the machine to invoke failover.
11. Repeat steps 1-10 for each additional node. After Data Instance Director has been installed on all nodes, restart each node.

Adding the Hub service

Complete the following steps on one node of the cluster, only.

To add the Hub service:

1. Launch **Failover Clustering** on the server: Click **Start > Administrative Tools > Failover Cluster Manager**
2. If the console tree is collapsed, expand the tree for the cluster that you want to configure.
3. Right click **Services and Applications**, then click **Configure a Service or Application**.
4. At the **Before You Begin** wizard, click **Next**.
5. Select **Generic Service**, then click **Next**.
6. Select the Cofio Hub Service from the list.
7. Name the service (for example, CofioHubService), assign an IP address, then click **Next**.
8. At the **Select Storage** dialog, click **Next**.
9. At the **Replicate Registry Settings** dialog, Click **Next**.
10. At the confirmation prompt, check that all settings are correct.
 - If a setting is incorrect, click **Previous**.
 - If all settings are correct, click **Next**.

11. At the **Summary** dialog, click **Finish**.

Licensing the software

See the section, [Cluster Environment License Activation](#).

Upgrading Data Instance Director

The Data Instance Director installation package automatically detects any currently installed versions of Data Instance Director. If a previous version is found, an upgrade can be run with the same installation package.

An upgrade can be performed if a newer version of Data Instance Director is available or if the existing installation has become corrupt.

During an upgrade no options are available to change the installation directory, repository directory (if installed), machine name, installation type, capabilities or master name or IP address. The contents of any classifications will not be changed, so if the HDID install location appears in a classification it will not be changed.

When an upgrade starts, the Data Instance Director services are shutdown. These services are restarted when the upgrade process is exited or completed.



Note: We recommend upgrading Data Instance Director only after backups have been completed, and not while backups are currently in progress.

Upgrade order

It is recommended that nodes be upgraded in the following order:

- Master nodes
- Repository destination nodes
- Replication destination nodes
- Source nodes

After upgrading, the rules should be redistributed and all destination nodes should be manually resynchronized with their sources.

Upgrade alert message

An upgrade starts with the following alert message (Figure 2-10)

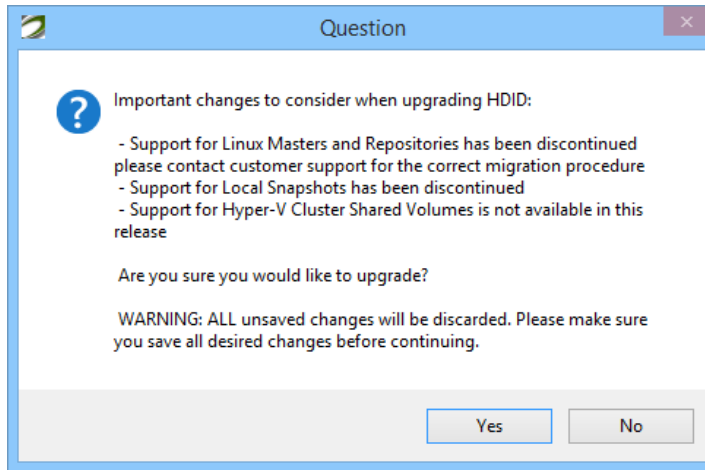


Figure 2-10 Upgrade alert



Note: If this is an upgrade installation on a 32-bit operating system, all previously defined roles are retained.

However, if you upgrade a 32-bit version Data Instance Director that is installed on a 64-bit operating system (in the installation default location), then in addition to retaining the previous roles, the installer converts Data Instance Director to a 64-bit version. The Hitachi file system structure (and the location of the filter log file) will be moved to the Program Files location, for example:

C:\Program Files\Hitachi

In this case, the following upgrade alert message is displayed (Figure 2-11):

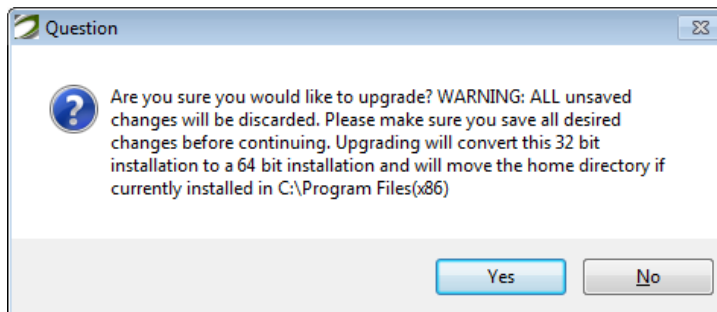


Figure 2-11 Upgrade warning message

The file system structure is not moved if Data Instance Director was installed in a non-default location.

Click **Yes** to proceed with the upgrade, or **No** to quit and exit Setup.

Upgrading clients from Master

While the Master can be upgraded by running the update executable, Data Instance Director clients, Source, Repository, and Replication Nodes, (virtual and physical) can be automatically updated from direct control from the Master.

To do this, place the updated Data Instance Director executable into the following directory on the Master:

```
<INSTALL>\Hitachi\HDID\runtime\updater
```

Then, select the **Upgrade Clients** option (Figure 2-12).

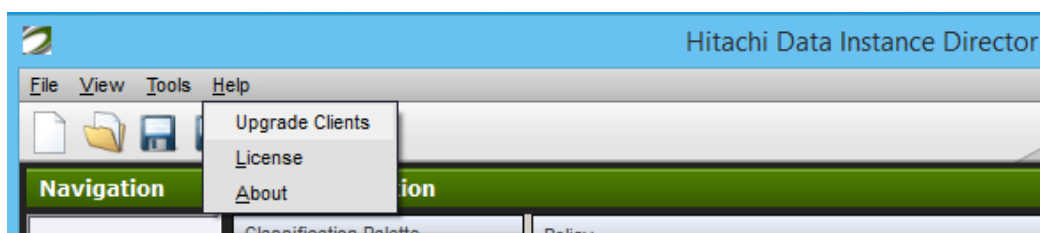


Figure 2-12 Upgrade Clients menu option

Afterward, the **Upgrade Clients** dialog launches and lists all Data Instance Director clients (Figure 2-13).

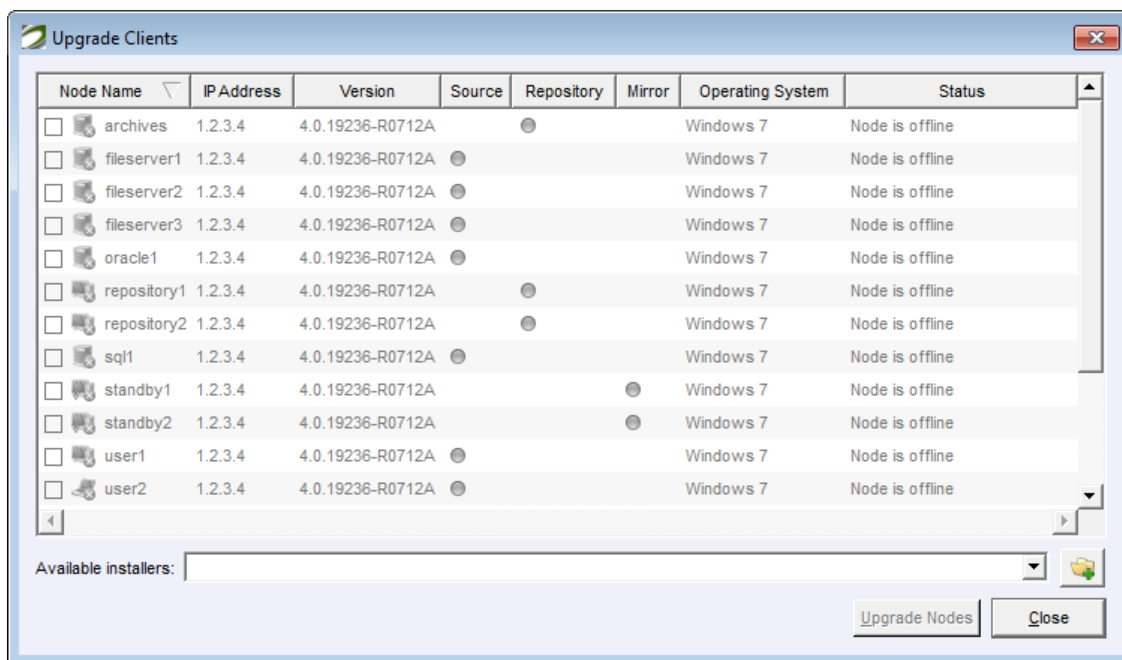


Figure 2-13 Data Instance Director Clients list

Select the check boxes on the nodes to upgrade and choose the installer version to use from the **Available installers** combo box at the bottom of the dialog. This will display all the executables that have been placed into the updater directory. The Master will push the install package and execute the upgrade automatically all together.



Note: The **Upgrade Clients** dialog only lists nodes that are running on a Windows OS.

Licensing Data Instance Director

To distribute rules, the master machine must have a valid Data Instance Director activation key.

As of Data Instance Director version 4.2.0, the software installation includes a built-in evaluation key for sites that want to use the product on a trial-to-permanent basis. After the software is installed, it can be used for 45 days, with up to 10 TB of maximum data usage, before a permanent license key is required. (Existing, purchased customers who upgrade to version 4.2.0 are unaffected by the feature: Their permanent license continues to convey during the upgrade.)

After a trial license has expired, the permanent license must be entered manually.

Manual license activation

If you have a valid activation key, then it can be entered directly within the Data Instance Director GUI.

1. Start Data Instance Director.
 - Double-click on the desktop shortcut.
 - or -
 - Access **Programs** → **Data Instance Director** → **Data Instance Director** from the **Start** menu.
2. Open the **Data Instance Director License** dialog (Figure 2-14). If Data Instance Director is unlicensed, then this will be opened automatically upon starting the user interface. Otherwise, it can be accessed through **Help** → **License**.

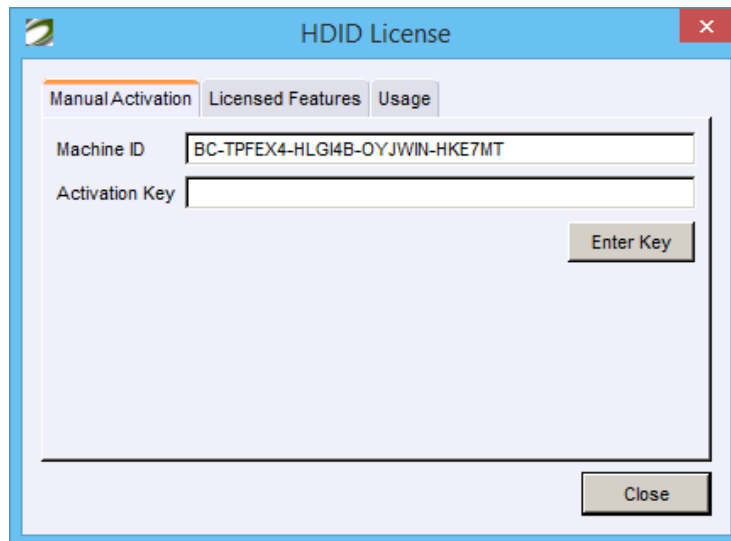


Figure 2-14 License dialog

3. Select the **Manual Activation** tab.
4. Enter the **Activation Key** then click **Enter Key**.

Cluster environment license activation

Because each machine in the cluster has a different machine ID, a separate license for each node is required.

1. Launch **Data Instance Director**.
The **Data Instance Director License** dialog displays and indicates the machine-specific ID.
2. Enter the license key for the machine, then click **Enter Key**.
This node is now licensed.
3. Close all **Data Instance Director** windows.
4. Browse to the installation directory that you specified in the first task, 'Installing Data Instance Director on the Controlling Node', then navigate to <Data Instance Director install>\db\config\
5. Right click on License.xml then open the file with WordPad.
6. Copy line 2 and insert as a new line after line 2.
7. Change the activation key to match your second node.
8. Repeat steps 6 and 7 for each additional node.
9. Save and close the file.

Uninstalling Data Instance Director

Data Instance Director can be uninstalled by navigating to its installation directory or the **Start** menu, and running **Uninstall Hitachi Data Instance Director**. If a repository was installed on the machine then the uninstaller will prompt the user to remove it manually.



Note: Uninstalling the Data Instance Director master does not deactivate active rules on client nodes. To do this, you must either clear the dataflow and redistribute before uninstalling, or uninstall each client individually.

Stubbed files and emails are not automatically un-stubbed; this must be done manually, prior to uninstalling Data Instance Director, by running the `sourcestubcli -r` command on each machine, restoring email then deleting the stubs.

Completing optional, post-installation configuration

If your site plans to stub email bodies and attachments, and if your site is currently running Microsoft Exchange Server 2007 (or later) with Microsoft Outlook 2007 (or later), then you can view that stubbed information directly within Microsoft Outlook. In order for Data Instance Director to unstub the viewed information, a Data Instance Director Outlook form must be configured, first.



Note: No stubbed email bodies or attachments can be viewed unless the Data Instance Director master is in the same domain as Microsoft Exchange.

We recommend setting up an Outlook form before completing the Getting Started and other tasks that are described in subsequent chapters.

Setting up an Outlook form for viewing stubbed emails

The general setup process involves the following tasks:

1. Installing the Outlook form.
2. Creating a public folder and setting permissions for it.
3. Publishing the Outlook form.

The following sections describe the implementation process. For information about file stubbing, refer to Chapters 3 and 4.

Installing an Outlook form

The Outlook Form must be installed within an Organizational level Exchange Public Folder within an Exchange Public Database, so it can be accessible by all Outlook users. You will use Exchange commands (cmdlets), which are available on a Microsoft Exchange Server system, to create and configure the Public Folder.

Public database

Only one Public Folder Database can be created on a specific Exchange Server.

To view available Public Folder Databases:

```
Get-PublicFolderDatabase
```

Applies to Exchange 2007:

[http://technet.microsoft.com/en-us/library/bb123687\(v=exchg.80\).aspx](http://technet.microsoft.com/en-us/library/bb123687(v=exchg.80).aspx)

To create a new Public Folder Database:

```
New-PublicFolderDatabase -Name <database_name> -StorageGroup  
<storage_group>
```

To view available Storage Groups:

```
Get-StorageGroup
```

To create the Public Folder Database on the first available Storage Group (typically this would be 'First Storage Group'):

```
New-PublicFolderDatabase -Name <Database_Name> -StorageGroup "First Storage  
Group"
```

Applies to Exchange 2010:

[http://technet.microsoft.com/en-us/library/bb397221\(v=exchg.141\).aspx](http://technet.microsoft.com/en-us/library/bb397221(v=exchg.141).aspx)

To create a new Public Folder Database:

```
New-PublicFolderDatabase -Name <database_name> -Server <server_name>
```

After the database is created it must be mounted:

```
Mount-Database -Identity <database_name>
```

Creating a public folder and setting permissions for it

The Public Folder used for the Organizational Form is created in a specific location. The following applies to Exchange 2010, or later.

To view available Public Folders:

```
Get-PublicFolder -Recurse
```

To create a new Public Folder:

```
New-PublicFolder -Path "\NON_IPM_SUBTREE\EFORMS REGISTRY" -Name  
"Organizational Forms Library"
```

Set the Language locale of the public folder:

```
Set-PublicFolder "\NON_IPM_SUBTREE\EFORMS REGISTRY\Organizational Forms  
Library" -EformsLocaleID EN-US
```

Public Folder Permissions for Exchange 2007:

[http://technet.microsoft.com/en-us/library/bb123509\(v=exchg.80\).aspx](http://technet.microsoft.com/en-us/library/bb123509(v=exchg.80).aspx)

Public Folder Permissions for Exchange 2010:

[http://technet.microsoft.com/en-us/library/bb310789\(v=exchg.141\).aspx](http://technet.microsoft.com/en-us/library/bb310789(v=exchg.141).aspx)

To be able to publish a form to the Public Folder, the publishing user must have the required permissions.

To view current permissions of the Public Folder:

```
Get-PublicFolderClientPermission -Identity "\\NON_IPM_SUBTREE\EFORMS  
REGISTRY\Organizational Forms Library"
```

To prevent possible issues with conflicting permissions, we recommend that a user which is not a member of any elevated groups (Administrators, etc...) publish the form. Then, grant temporary permissions to publish the form.

To set the permissions required to publish the form to the Organization Forms Library:

```
Add-PublicFolderClientPermission -AccessRights Owner -Identity  
"\\NON_IPM_SUBTREE\EFORMS REGISTRY\Organizational Forms Library" -User  
<domain\username>
```

Publishing the Outlook form

In the following task sequence, you will install and open a Data Instance Director Outlook form, 'HDID Exchange Email Destubber Outlook Form.fdm'. Figure 2-15 provides an example of the located form, after it has been installed.

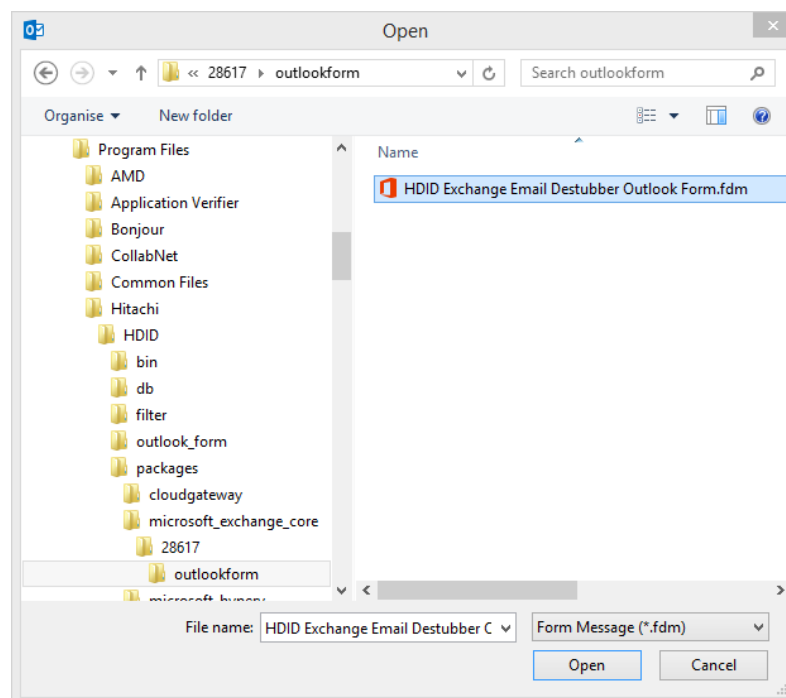


Figure 2-15 Locating the Data Instance Director Outlook Form

The instructions assume that Data Instance Director is installed on the same machine on which Microsoft Outlook is installed. However, you can copy the Data Instance Director Outlook form 'HDID Exchange Email Destubber Outlook Form.fdm', to any machine on which Microsoft Outlook is installed.

Once the form is published at the organization level, the following steps do not need to be repeated for every user.



Note: The following steps apply to Microsoft Outlook 2010.

1. Log on to Windows using the user account which has been given the **Owner Public Folder Client** permission.
2. Open **Microsoft Outlook** and if needed, configure it to connect to the user's Microsoft Exchange Mailbox.
3. From the **File** menu, select **Options**.
4. Select **Advanced** then scroll down to the **Developers** section. Click **Custom Forms...**
5. Select **Manage Forms...**
6. Confirm that the right pane displays **Organization Forms**. If it does not, complete the following:
 - a) Click **Set** to display the **Set Library To** dialog.
 - b) Select the **Forms Library** radio button.
 - c) Select **Organization Forms** in the drop-down list.



Note: If the **Organization Forms** option is not available, ensure that the **Organization Forms Library** has been created correctly.

- d) Click **OK** to return to the **Forms Manager** dialog.
7. Below the right pane, click **Install...**
8. Browse to the directory:
%HDID%\packages\microsoft_exchange_core\#####\outlookform\
(where ##### is the HDID build number, e.g. 28617).
9. On the bottom right (above the open and cancel buttons), change the file filter to **Form Message (*.fdm)**.
10. Select the file, 'HDID Exchange Email Destubber Outlook Form.fdm', then click **Open**.



Note: If Microsoft Outlook is not installed on the same machine as Data Instance Director, then you can copy HDID Exchange Email Destubber Outlook Form.fdm to any machine on which Microsoft Outlook is installed.

11. In the **Form Properties** dialog, click **OK**.
12. The Data Instance Director Outlook Form is now installed and should be available for all users.

13. Click **Close** and **OK** to return to the **Outlook Options** dialog.
14. Click **OK** to close the options dialog and return to Outlook.



Note: When initially viewing a Data Instance Director stubbed message, a dialog will display and will indicate that the form is being installed onto the machine. This indicates Outlook is copying the Organizational Form to the local computer for use when viewing messages.

Uninstalling an Outlook form

1. Log on to Windows using the user account which has been given the **Owner Public Folder Client** permission.
2. Open **Microsoft Outlook** and if needed, configure it to connect to the user's Microsoft Exchange Mailbox.
3. From the **File** menu, select **Options**.
4. Select **Advanced** then scroll down to the **Developers** section. Click **Custom Forms...**
5. Select **Manage Forms...**
6. Confirm that the right pane displays **Organization Forms**. If it does not, complete the following:
 - a) Click **Set**.
 - b) Select **Form Library**.
 - c) Select **Organization Forms**.
 - d) Click **OK** to return to the previous window.
7. From the right pane, select the **Data Instance Director_Outlook_Form**.
8. Click **Delete**.
9. Click **Yes** to confirm the deletion.
10. Click **Close** and **OK** to return to Outlook options.
11. Click **OK** to close the options dialog and return to Outlook.

Stubbed email information and security

Stubbed email authentication (whereby the recipient of the stubbed email is required to enter the NT logon information for their account, in order to open the stubbed email and any associated attachments and forward the item) is currently disabled in Data Instance Director version 5.0. It will be re-enabled in a later version.

Forwarding email information

As with standard email, stubbed emails and their attachments can be forwarded to other recipients. There are differences in handling, however, depending on whether the Web version of Microsoft Outlook is used:

- Within the Desktop version of Microsoft Outlook, after the email user enters their NT logon information, the stubbed email is unstubbed and can be forwarded to a designated recipient, who will receive the email as a standard email notification.
- Within the web version of Microsoft Outlook, after the email user forwards a stubbed email that contains a hyperlink, the recipient must access the cloud gateway and enter the password information of the originating email owner. In this case, we recommend that the originating owner of the email opens the stubbed email from the hyperlink, download the email from the cloud gateway, and then either forward that downloaded email or send it as an email attachment to the designated recipient.

Getting started

This chapter describes the steps that are required to set up each of the main policy types that are supported by the Data Instance Director repository and replication nodes. It also includes information about customizing alert notifications.

- ☐ [Overview](#)
- ☐ [Backup, Live Backup, and Continuous Data Protection \(CDP\)](#)
- ☐ [File Archive, File Versioning](#)
- ☐ [Replication](#)
- ☐ [Mirror Replication from the Data Flow panel dialog](#)
- ☐ [Information Security: Block and Track](#)
- ☐ [Application Backups: MS SQL, MS Exchange, and Hyper-V](#)
- ☐ [Agentless Nodes](#)
- ☐ [Managing Alert Notifications](#)

Overview

In order to describe the steps for setting up each of the main policy types, three nodes are used in this set of examples: A node that will be the source of the data called *filesaver1*, and two types of destination nodes. The mirror destination is labeled as *standby1* and the repository destination is labeled as *repository1*.

Backup, Live Backup, and Continuous Data Protection (CDP)

Using Data Instance Director, backups are achieved by continuously replicating data to a repository. When it is time to produce a backup, a repository snapshot is created. The resulting repository snapshot is given a retention period, and is available for full or partial restores from the **Restore** screen until it is time to retire it. Backups can be configured with a Batch Data Mover for standard backup or with a Continuous Data Mover for Live Backup. It is also possible to specify multi-level deduplication at source and target (repository).

CDP

CDP operations can be used alone, or together with backup operations to create more frequent recovery points of one minute to one hour. CDP does not provide post process deduplication due to its short retention cycle. Ideally, CDP should be used for application data such as databases and files for real-time. CDP is always used with the Continuous Data Mover and provides source level deduplication with Changed Byte Transfer.

Backup or CDP configuration

The following section described how to create a simple workflow to back up the data in C:\source on *filesaver1* to a Store Group, *StorageGroup01* on *repository1*.

The tasks to be performed when configuring a backup or CDP workflow are:



Policy definition – To define what data qualifies for back up.



Data Flow definition – To define which nodes should be backed up and where to.



Distribution and Activation – To put the policy into effect for each node in its data flow.

Policy definition

A policy defines two things: The data in question (*data classification*) and the process to be performed on it (*data operation*).

In this example, we need the contents of the C:\source directory to be backed up.

1. To define a policy, click the **Policy Definition** icon  in the Navigation column on the far left to navigate to the **Policy Definition** screen (Figure 3-1). See the [Policy Definition](#) section in Chapter 5 for more information.

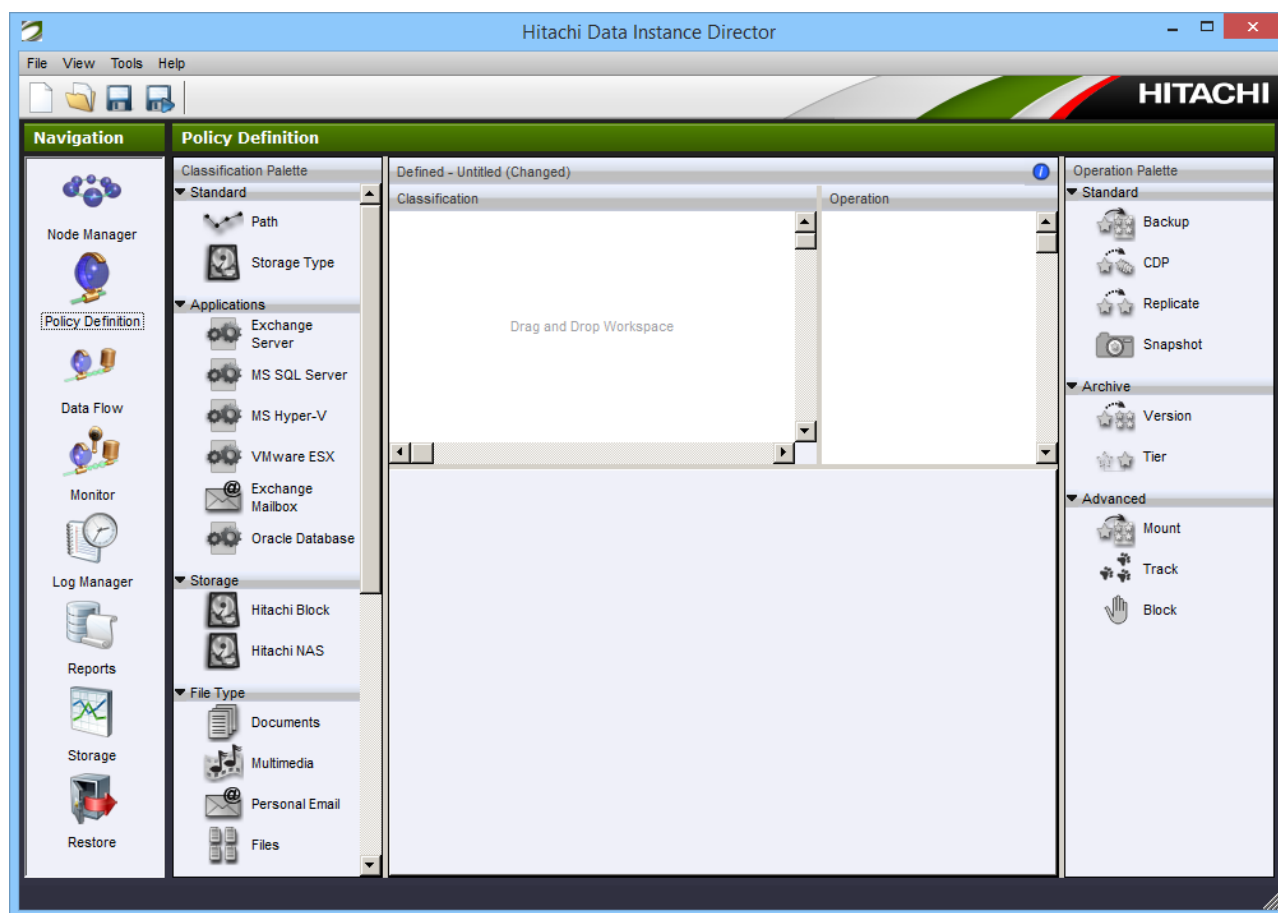


Figure 3-1 Policy Definition screen

2. The **Classification Palette** on the left is used to define the data captured by the policy. In this example the only requirement is that the data resides in a certain directory – c:\source. Click the **Standard** heading to expand it, then drag the **Path** icon from the **Classification Palette** into the **Classification** workspace in the middle of the screen (Figure 3-2).

Be sure that the **Path** icon that was just placed into the workspace is selected. Edit boxes to enter applicable paths are then displayed in the **Attributes** tab at the bottom of the screen.

3. Click the folder button  and a new dialog will appear. From the combo box, select the machine from which the path may be taken. Select the path and in the **Apply to** combo box in the same dialog choose to apply the path to all nodes, or to only this one node. The path will appear in the **Enter Paths** edit box, as shown in the following figure. C:\source is the only entry required, and only for this particular node (fileserver1), as this is the directory that contains the data we wish to back up.

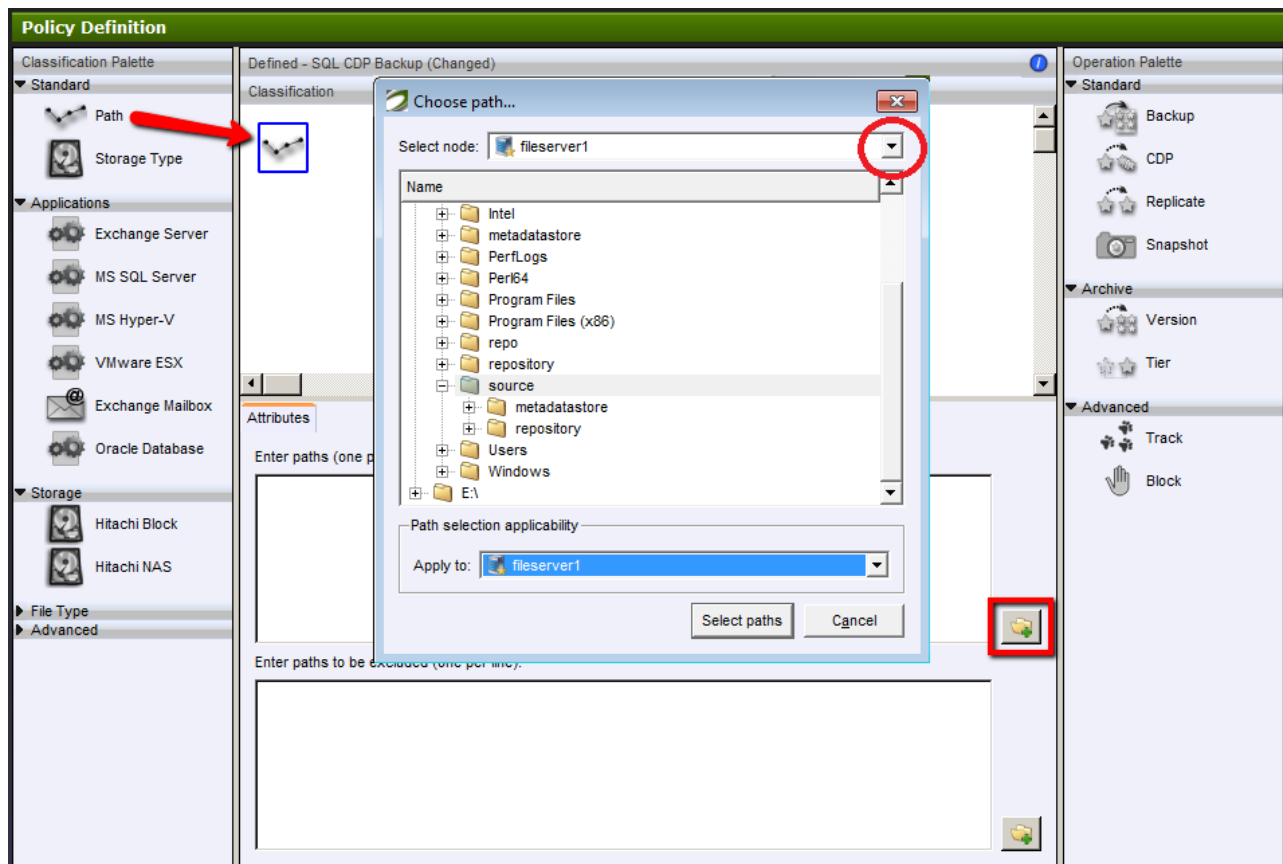


Figure 3-2 Defining a Policy Classification

4. The **Operation Palette** on the right is used to define the data operation. In this example, the data is to be backed up. Expand the Standard heading and drag the **Backup** icon into the **Operation** workspace just to the left of that palette (Figure 3-3). If CDP is used instead, the same principle applies, and the **CDP** icon can be dragged into the workspace instead.

- When the **Backup** icon is selected in the workspace area, the options related to the backup operation are displayed in the **Attributes** tab. In this example, a daily backup is required and the backup needs to be kept for two weeks, so set the Retention to **2 Weeks** and set the Recovery Point Objective to **1 Day**.

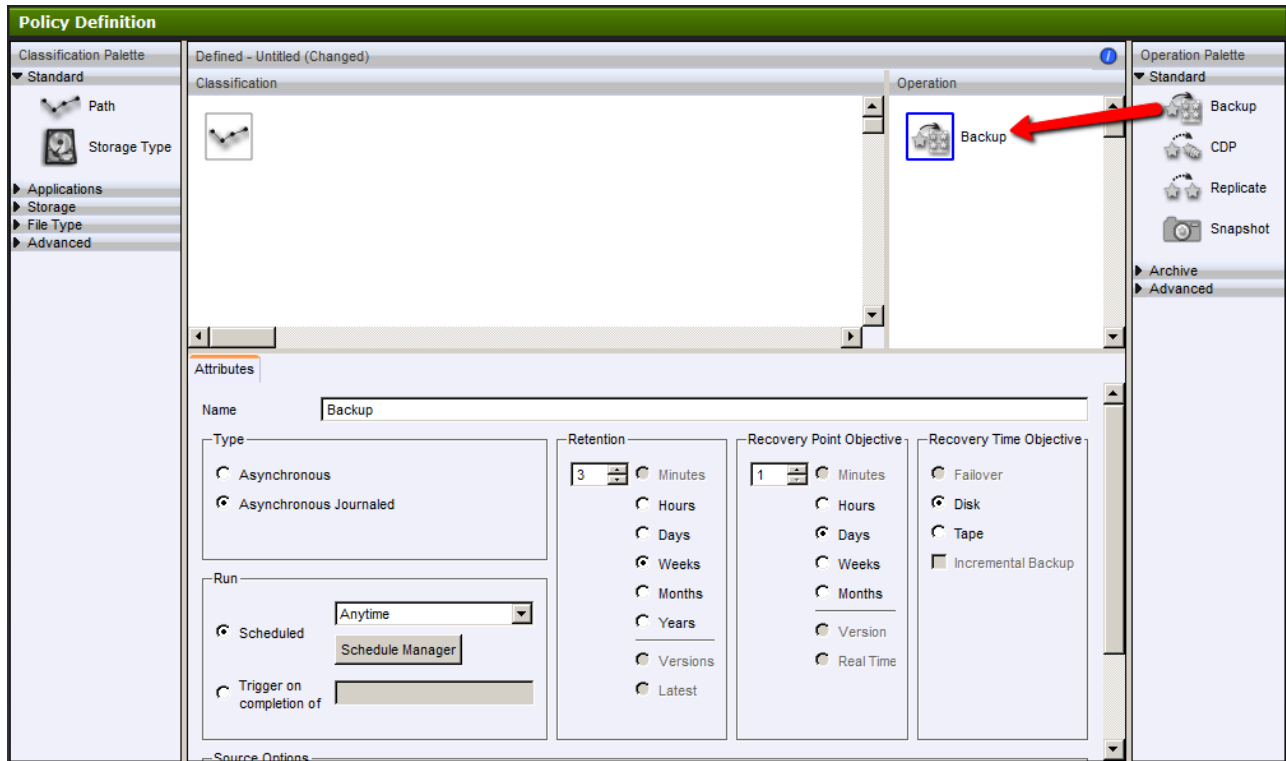


Figure 3-3 Defining a Policy operation

- Select **File** → **Save** from the main menu to save the policy.
- Enter a policy name and brief description, and click **Save** (Figure 3-4).

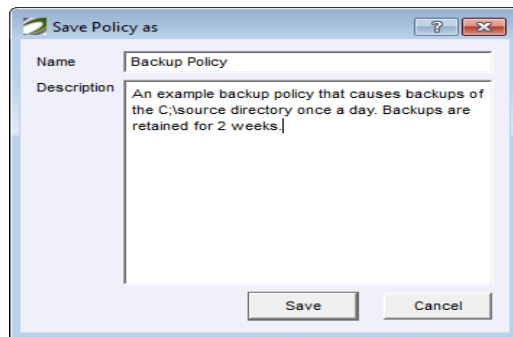


Figure 3-4 Saving a Policy

Data flow definition

The data flow defines how data moves between machines. In this example, data needs to move from *fileserver1* to *StorageGroup01* on *repository1*.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (Figure 3-5).

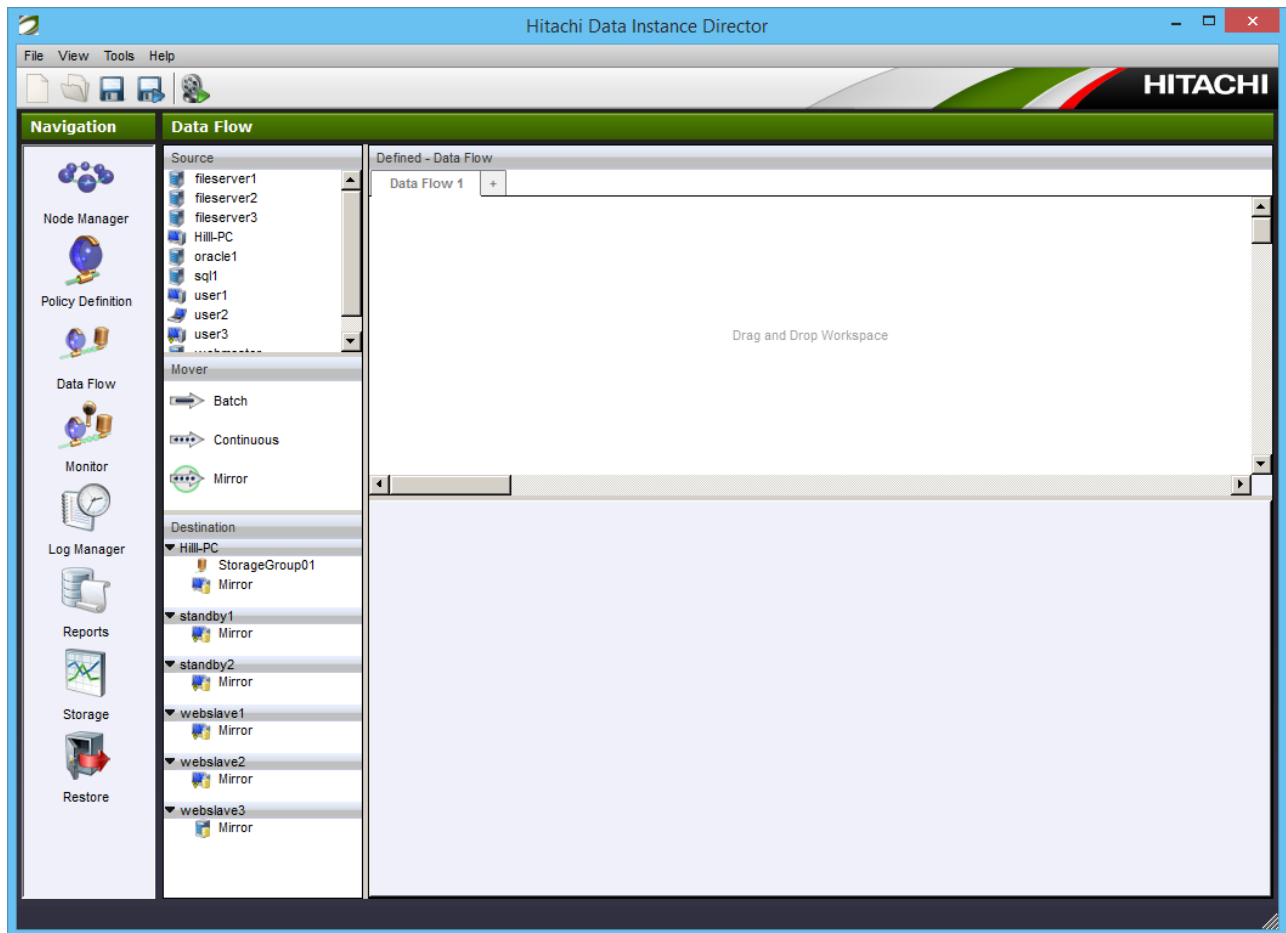


Figure 3-5 The Data Flow screen

The three palettes on the left side contain the source and destination nodes. It also provides Data Instance Director's *mover* operations.

Data Flow object type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed

Icons from the palettes are dragged and dropped into the workspace (Figure 3-6). The tabs below the workspace shows the specific options for each placed item.

In this example, data needs to move from *fileserver1* to the Store Group, *StorageGroup01* on *fileserver1*. This transfer is done with a *continuous* mover operation.

2. Drag the **fileserver1** icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Continuous** icon from the **Mover** palette to the **Data Flow** workspace. If the **Continuous** icon is dropped on top of the **fileserver1** icon, they are automatically connected. Note that if the source node is running Linux a Batch data mover must be used instead as Linux is unable to detect and send continuously-changing data at this time.
4. Drag the **StorageGroup01** icon from within *fileserver1* in the **Destination** palette and connect it by dropping it on to the **Continuous** icon.

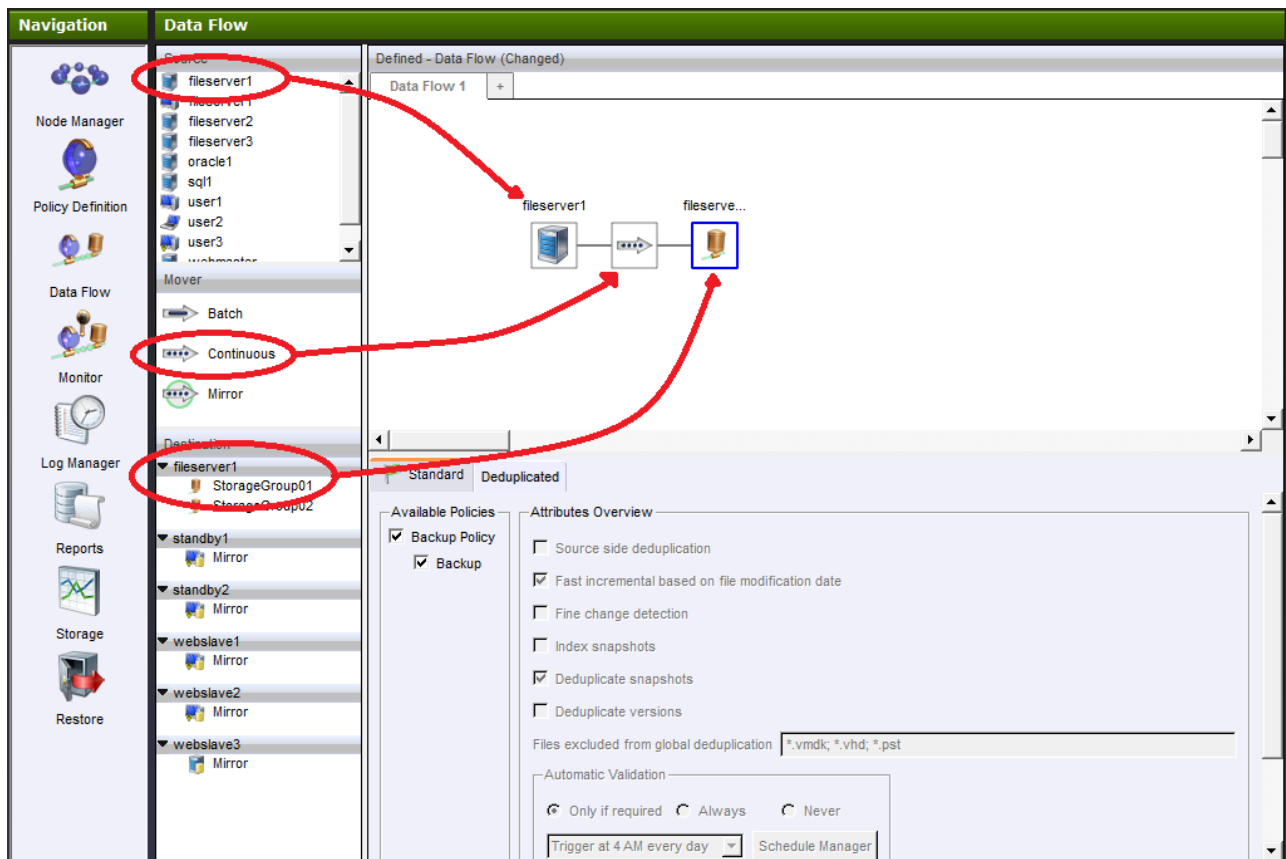


Figure 3-6 The Complete Data Flow

5. In the workspace, select the **fileserver1** icon to display the list of defined policies in the **Available Policies** pane in the **Standard** tab at the bottom

of the screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.

6. In this example, the Backup Policy needs to be assigned. Select the checkbox next to that policy.
7. To apply the policy to the destination, select the **StorageGroup01** icon in the workspace. Policies are displayed in the **Standard** tab's **Available Policies** pane.
8. Select the Backup Policy. This selection assigns all of the data operations belonging to this policy to this store.

A Store Group can have a number of named stores. Data Instance Director is preconfigured with two stores, Standard and Deduplicated. The Deduplicated store is selected for this example. Any store can be configured with various attributes, including type of deduplication.

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the Distribute Rules button after the policy is compiled and the rules will be distributed and the policy activated.

The system performs an initial synchronization and then begins real time replication to the repository, producing daily backups. Information regarding the process can be seen on the **Monitor** and **Log Manager** screens.

Once the initial synchronization is complete the backup can be seen by performing a search in the **Restore** screen. See the [Restore](#) section of Chapter 5 for details.

File archive, file versioning

Data Instance Director versioning is achieved by continuously replicating data to a repository. The repository stores changes to files whenever they are saved so that previous versions can be recovered if needed. The number of versions held in the repository is controlled indirectly by the retention period and change rate.

The data sent to the repository for backup and version operations are held as objects which are tagged and indexed. Data Instance Director source nodes (clients) can then utilize Data Instance Director as a central archive repository for files and search for them by name, extension or other search term. Versions can also be viewed easily by a timeline of creation and edit date, by checking the box on the Data Instance Director **Restore** screen. For more information on version search, retrieval, and restore, see the [Restore](#) section of Chapter 4 for details.

Version configuration

The tasks to be performed when configuring a version policy are:



Policy definition – To define what data qualifies for versioning



Data Flow definition – To define what node's data should be versioned and where to



Distribution and Activation – To put the policy into effect for each node in its data flow

Policy definition

A policy defines *data classification* and the *data operation*.

In this example, the contents of the C:\source directory will be versioned to a repository.



1. Click the **Policy Definition** icon to navigate to the **Policy Definition** screen (Figure 3-7) and define a policy, as shown earlier in this chapter.
2. In this example, a directory needs to be defined. Expand the **Classification Palette - Standard** heading and drag the **Path** icon from the **Classification Palette** into the **Classification** workspace.

Be sure that the **Path** icon remains selected in the workspace. The options specific to the Path classification are displayed in the **Attributes** tab at the bottom of the screen.



3. Enter this value in the **Enter Paths** edit box, or click the folder icon and select a path as shown in the following figure. In this example, enter C:\source to define the path for the one node, only.

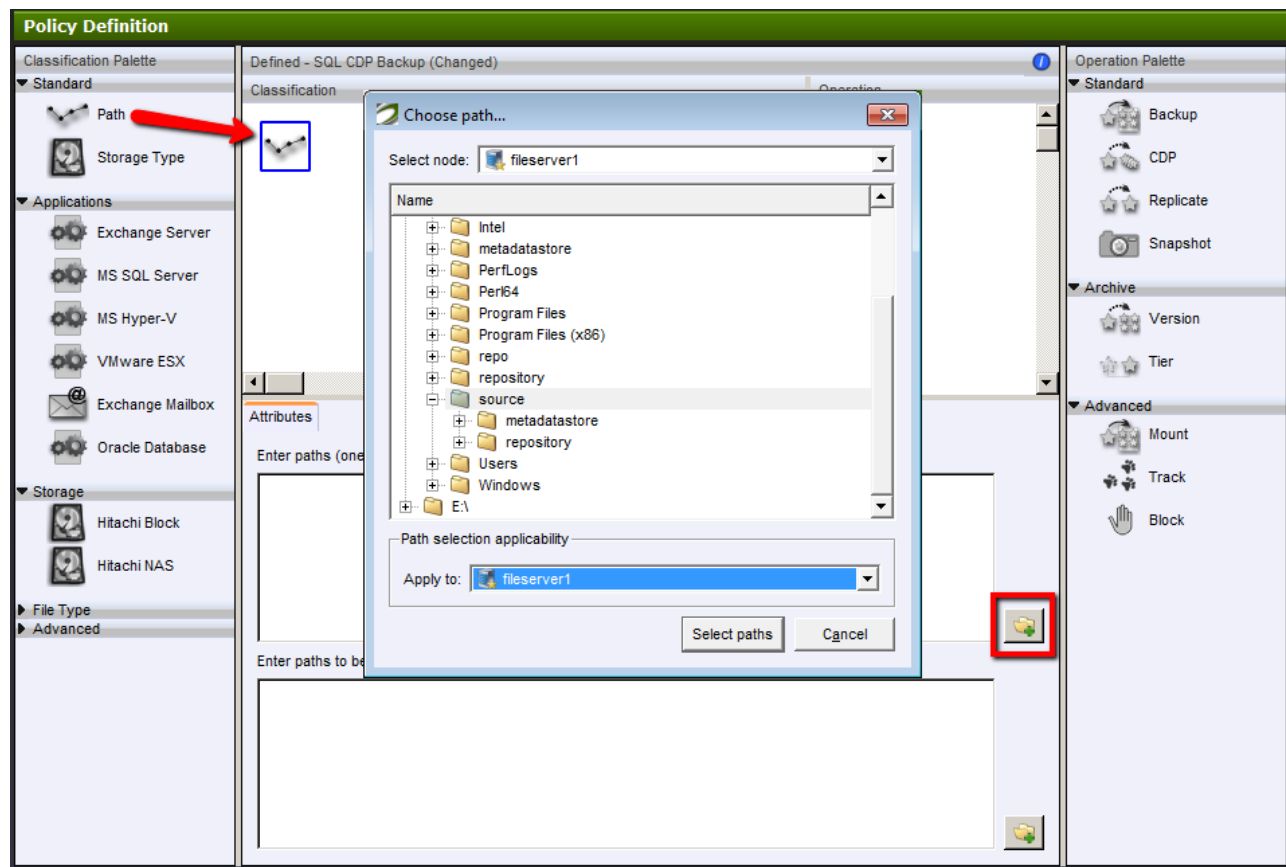


Figure 3-7 Version Policy path

4. Expand the **Operation Palette - Archive** heading and drag the **Version** icon from the **Operation Palette** into the **Operation** workspace.



Note: Linux supports versions but only in conjunction with backup. The policy must contain both a backup operation and a version operation. Each time the backup has completed, changed files are versioned.

5. When the **Version** icon is selected in the workspace area, the options related to the version operation are displayed in the **Attributes** tab (Figure 3-8). For this example, set Retention to 4 Weeks. Note that if only one version of a file exist in the repository then it will not be deleted when its retention period expires.

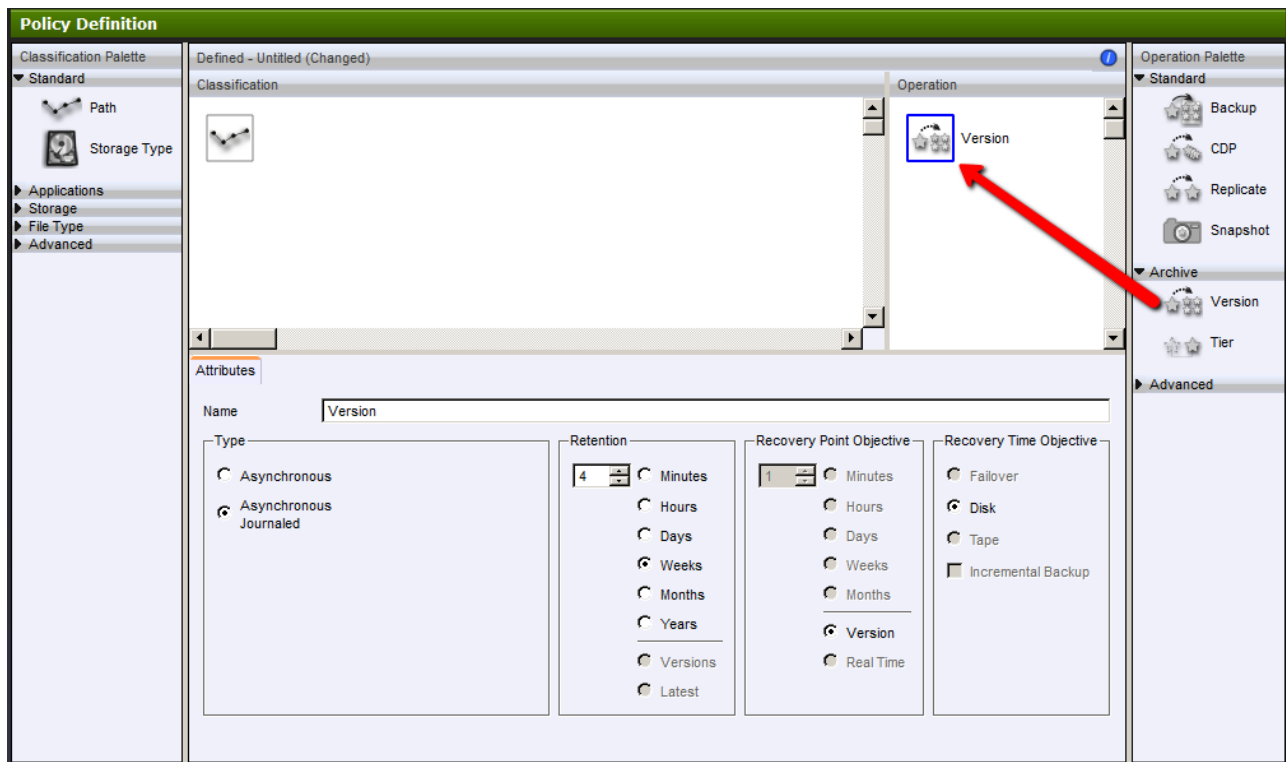


Figure 3-8 Defining a Version Policy

6. Select **File** → **Save** from the main menu to keep the policy.
7. Enter a policy name and brief description, then click **Save** (Figure 3-9).

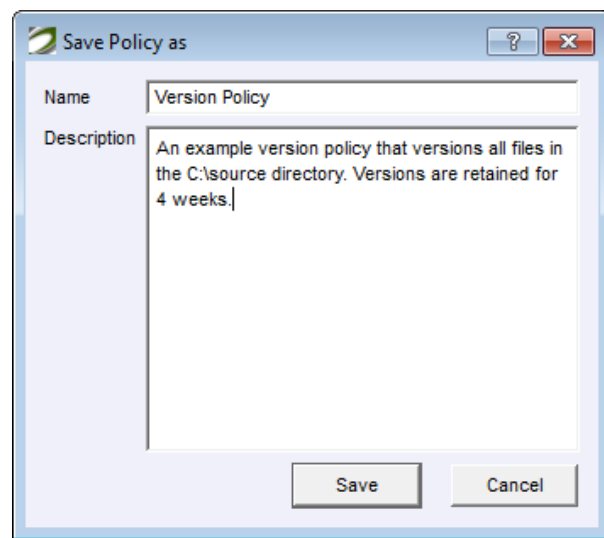


Figure 3-9 Saving a Version Policy

Data flow definition

In this example, data needs to be versioned from *filesaver1* into *repository1*.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen.

The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's **mover** operations.

Data Flow object type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed

In this example data, the **Continuous** mover operation will be used to move data from *fileserver1* to the Store Group, *StorageGroup01* on *repository1*.

2. Drag the **fileserver1** icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Continuous** icon from the **Mover** palette to the **Data Flow** workspace. If the **Continuous** icon is dropped on top of the **fileserver1** icon, they are automatically connected.
4. Drag the **StorageGroup01** icon from within *repository1* in the **Destination** palette and connect it by dropping it on to the **Continuous** mover icon.
5. In the workspace, select the **fileserver1** icon to display a list of defined policies in the **Policies to assign locally** tab at the bottom of the screen (Figure 3-10). To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.

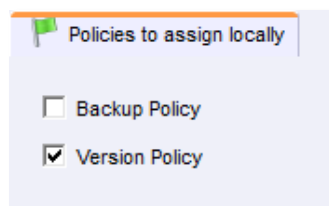


Figure 3-10 Selecting Version Policy

6. Check **Version Policy** to select that policy.
7. Select the **StorageGroup01** icon to apply the policy to the destination node.

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

The system performs an initial synchronization in order to create the initial versions. Information regarding the initial synchronization can be seen on the **Monitor** and **Log Manager** screens.

Once the initial synchronization is complete the versions can be seen by performing a search in the **Restore** screen. More information on searching can be found in the [Restore](#) section of Chapter 5. Modifying the data in `c:\source` on *fileserver1* will result in new versions, evident through performing another search within the **Restore** screen.

Replication

Data Instance Director replication is achieved by continuously replicating data from a source node to a destination node.



Note: Only Windows source nodes can be replicated using the continuous mover. Linux source nodes have to perform replication using the batch data mover.

Replication configuration

The tasks to be performed when configuring a replication policy are:



Policy definition – To define what data should be replicated



Data Flow definition – To define the source and destination of the replication



Distribution and Activation – To put the policy into effect for each node in its data flow

Policy definition

A policy defines *data classification* and the *data operation*.

In this example, we wish to replicate the contents of `C:\source\` directory on *fileserver1* to another server, called *standby1*.



1. Click the **Policy Definition** icon to navigate to the **Policy Definition** screen (Figure 3-11) and define a policy, as shown earlier in this chapter.
2. In this example, a directory needs to be defined. Expand the **Classification Palette - Standard** heading and drag the **Path** icon from the **Classification Palette** into the **Classification** workspace.

3. Be sure that the **Path** icon is selected in the workspace. The options specific to the **Path** classification are displayed in the **Attributes** tab at the bottom of the screen.
4. Enter a path in the **Enter Paths** edit box. In this example, enter C:\source to define the path.
5. Expand the **Operation Palette Standard** heading and drag the **Replicate** icon from the **Operation Palette** into the **Operation** workspace.

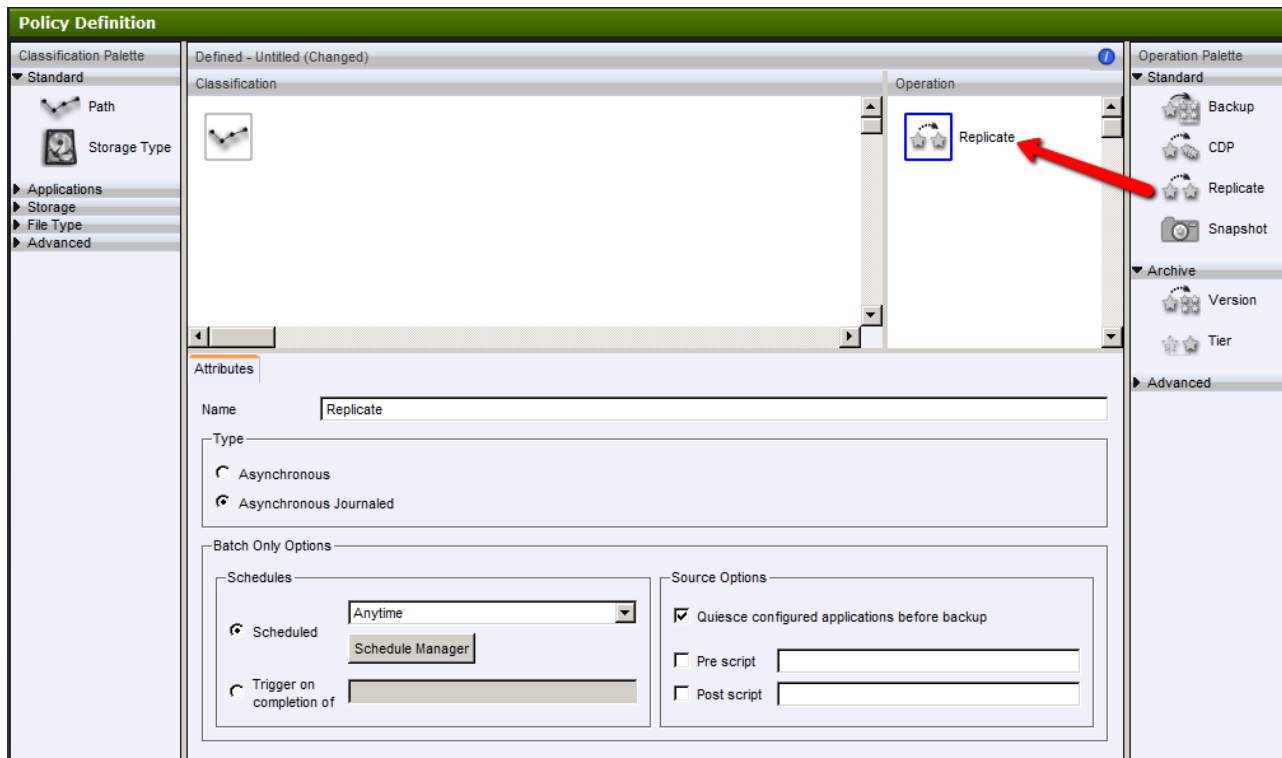


Figure 3-11 Replication Policy

6. Be sure that the **Replicate** icon is selected in the workspace area. The options related to the replication operation are displayed in the **Attributes** tab. In this example, the default settings are used. The **Schedule** and **Source Options** controls are under the heading **Batch Only Options**, this is because these options are only relevant when used with a batch mover.
7. Select **File** → **Save** from the main menu to keep the policy.
8. Enter a policy name and brief description, and click **Save**.

Data flow definition

In this example, data needs to move from *fileserver1* to *standby1*.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (Figure 3-12).

The three palettes on the left-hand side contain the source and destination nodes, plus the *mover* operations.

Data Flow object type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed

In this example, the *Continuous* mover operation will be used to move data from *fileserver1* to the standby server, *standby1*.

2. Drag the **fileserver1** icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Continuous** mover icon from the **Mover** palette to the **Data Flow** workspace. If the **Continuous** mover icon is dropped on top of the **fileserver1** icon, they are automatically connected.
4. Drag the **Mirror** icon from within *standby1* in the **Destination** palette and connect it by dropping it on to the **Continuous** mover icon.

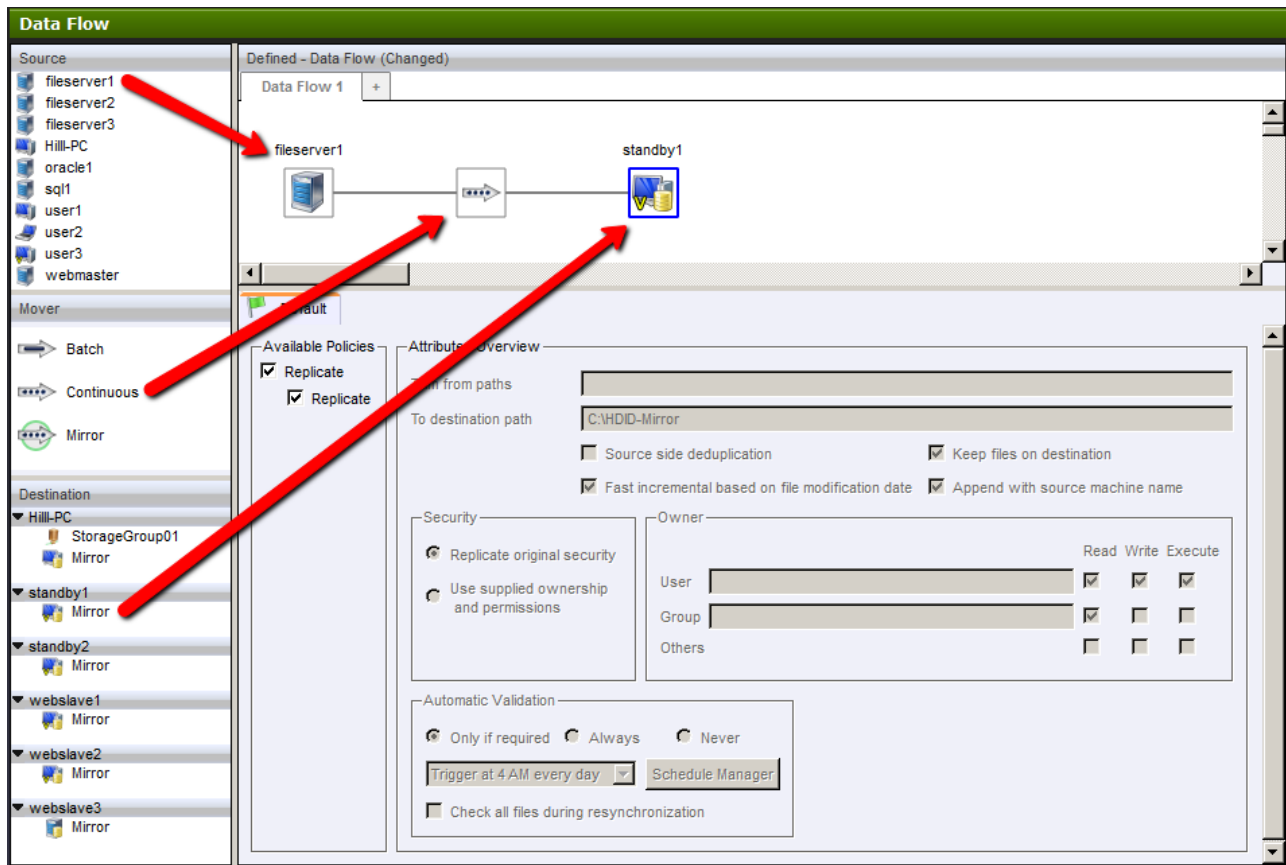


Figure 3-12 Replication Data Flow

5. In the workspace, select the **fileserver1** icon to display a list of defined policies in the **Available Policies** pane on the **Default** tab at the bottom of the screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.
6. Check the Replicate Policy to select that policy.
7. Select the **standby1** icon to apply the policy to the destination node.
8. Check the Replicate Policy on the **Default** tab for the Mirror Destination in at the bottom of the screen. This setting assigns all of the data operations belonging to the Replicate Policy to this store. In this screen, each of the data operations is displayed below the policy name. In this example, Replicate is the only data operation. The default mirror destination has a destination directory of C:\HDID-Mirror, which is used in this example. For a complete description of all destination options see the Mirror destination section later in the manual.



Note: The Replication node needs to be running the same or a newer version of the Windows operating system, for example, 2003 → 2008 is OK but 2008 → 2003 is not supported. Replication of same file systems is supported; for example, NTFS to NTFS, or ReFS to ReFS. Replication of dissimilar file system types is not supported.

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

The system performs an initial synchronization and then begins real time replication to the destination. Information regarding the process can be seen on the **Monitor** and **Log Manager** screens.

Once the initial synchronization is complete all the data that was in C:\source on *fileserver1* can be seen inside C:\HDID-Mirror on *standby1*. Modifying the data on *fileserver1* will result in real-time change to the corresponding data on *standby1*.

Mirror replication from the Data Flow screen

Data Instance Director also provides a simple way to mirror paths between two Source node machines in a more direct manner, through the **Data Flow** screen.

The tasks to be performed when configuring a Mirror Replication in the **Data Flow** screen are:



Data Flow definition – To define the folders and two source nodes for the mirror.



Distribution and Activation – To put the policy into effect for each node in its data flow.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** definition screen (Figure 3-13).

The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's **mover** operations.

Data flow object type	Description
Source	Nodes installed with the <i>Data Source</i> component.
Destination	Not required in this operation.
Mover	The type of data movement operation to be performed.

2. Drag two machines from the **Source** palette into the **Data Flow** workspace. One should be the 'primary' data node, where real data is being written to—make this the 'left side node'. The other should be a 'standby' node, where data will be copied to in real-time—make this the 'right side node'. Keep some space between them so you can move over the **Mirror**, drag and drop the Mirror mover onto the Source Node on the left (*FILE-SERV1*), then drag it on top of the node to the right (*FILE-SERV2*) so they are both connected.
3. Leave the Mirror data mover set to 'Left to Right'. Click on the folder for the Mirror Path (left side node) and elect the folder to mirror, C:\users in this example. Do the same for the right side node, also C:\users in this example (see **red** annotations). Activate rules and distribute for the new mirror partnership to begin initial sync. Data will now flow from Left side node to Right side node in real-time. In event of failure of the Left side node, you can re-assign the left side node as the primary and direct resources (file share permissions, etc.) within your Windows domain to that machine.

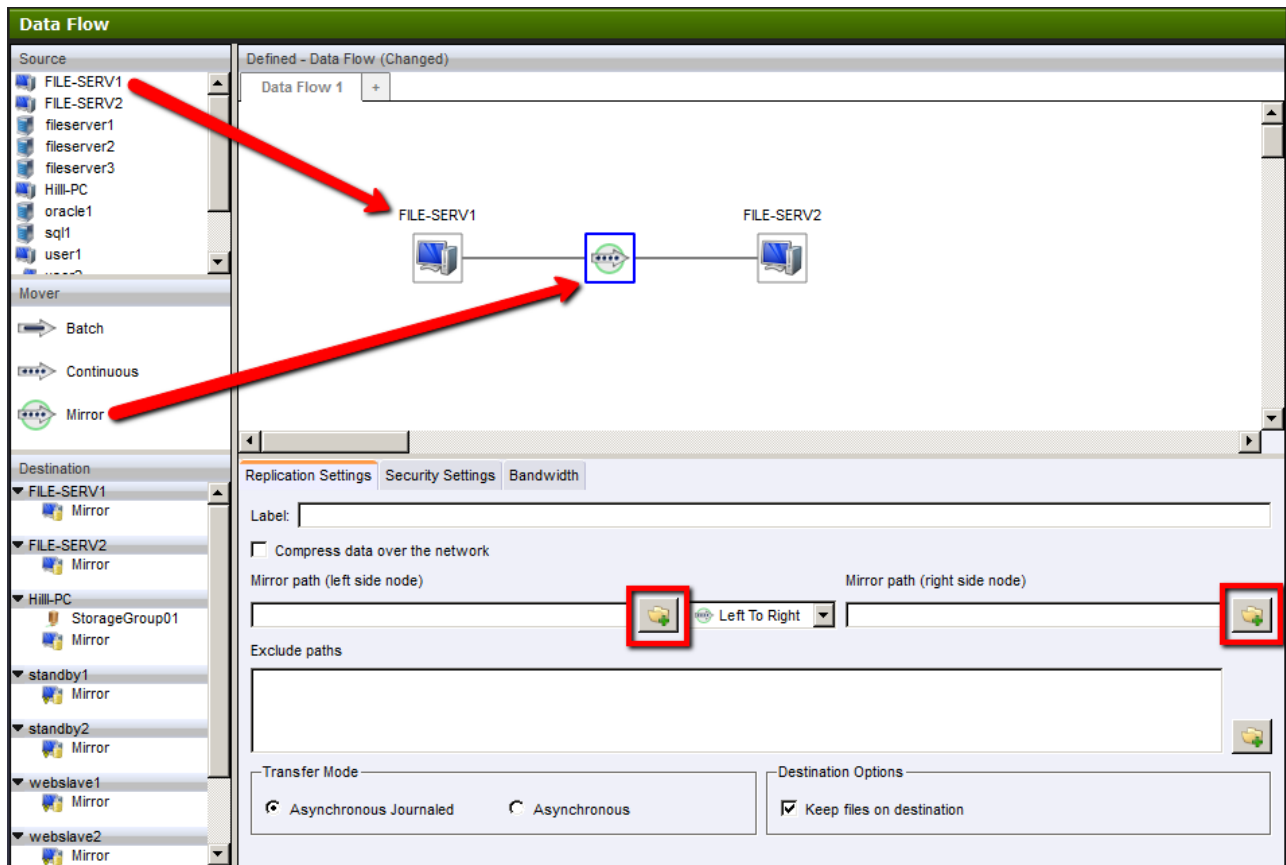


Figure 3-13 Mirror Replication Data Flow

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog displays.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

The system performs an initial synchronization and then begins real time replication to the destination. Information regarding the process can be seen on the **Monitor** and **Log Manager** screens.

Once the initial synchronization is complete all the data that was in C:\users on *FILE-SERV1* can be seen inside C:\users on *FILE-SERV2*. Modifying the data on *FILE-SERV1* will result in real-time change to the corresponding data on *FILE-SERV2*.

**Notes:**

- In the event that there is a failure with the left side node (primary), change the direction (shown in the center of the **Replication Settings** tab) from 'Left To Right' by clicking the combo box and selecting **Pause. Save and Activate**, then **Distribute**. Afterward, you can begin using the standby node FILE-SVR2 as the primary data source.
 - When you have rebuilt FILE-SERV1, you can redirect the flow back to FILE-SVR1 from FILE-SERV2: Change the direction of data flow again from 'Pause' to 'Right To Left', **Save and Activate**, then **Distribute**. When synchronization is completed, you can once again pause (save/activate/distribute) and then change back to 'Left To Right' to reset the mirror as it was before, with FILE-SERV1 as primary source and FILE-SERV2 as standby.
 - If your site uses a bidirectional mirror, files are not locked. In this case, do not simultaneously edit a file on both sides of the mirror: Doing so can have undesirable results.
-

Information security: Block and Track

Use Data Instance Director's Block and Track feature for security tasks such as data leak prevention, USB monitoring of file changes, copies, renames and movements.

Block configuration

Data Instance Director can be used to block access to files or directories on a given source machine.



Note: Only source nodes that are running on a Windows OS can be blocked.

The tasks to be performed when configuring a blocking policy are:



Policy definition – To define what data and actions should be blocked.



Data Flow definition – To define what node should implement blocking policies.



Distribution and Activation – To put the policy into effect for the node.



Caution: Do not combine Block with other Data Instance Director Operations (Backup, Version, etc.). *Block can only be combined with the Track operation.* When activated, Block places limitations on the file system for the classified

data addressed within the policy. Applying Block to Classifications such as 'OS Type' locks the system and requires the assistance of Support.

Policy definition

A policy defines *data classification* and the *data operation*.

In this example, we wish to block access to the contents of the directory, C:\blockdirectory\

1. Click the **Policy Definition** icon  to navigate to the **Policy Definition** screen (Figure 3-14) and define a policy, as shown earlier in this chapter.
2. In this example, a directory needs to be defined. Expand the **Classification Palette - Standard** heading and drag the **Path** icon from the **Classification Palette** into the **Classification** workspace.
Be sure that the **Path** icon is selected in the workspace. The options specific to the Path classification are displayed in the **Attributes** tab at the bottom of the screen.
3. Enter a path to block in the **Enter Paths** edit box, as shown in Figure 3-16. In this example, enter C:\blockdirectory to define the path.
4. Expand the **Operation Palette - Advanced** heading and drag the **Block** icon from the **Operation Palette** into the **Operation** workspace.
Be sure that the **Block** icon is selected in the workspace area. The options related to the Block operation are displayed in the **Attributes** tab. By default no options are selected, for this example select all options.

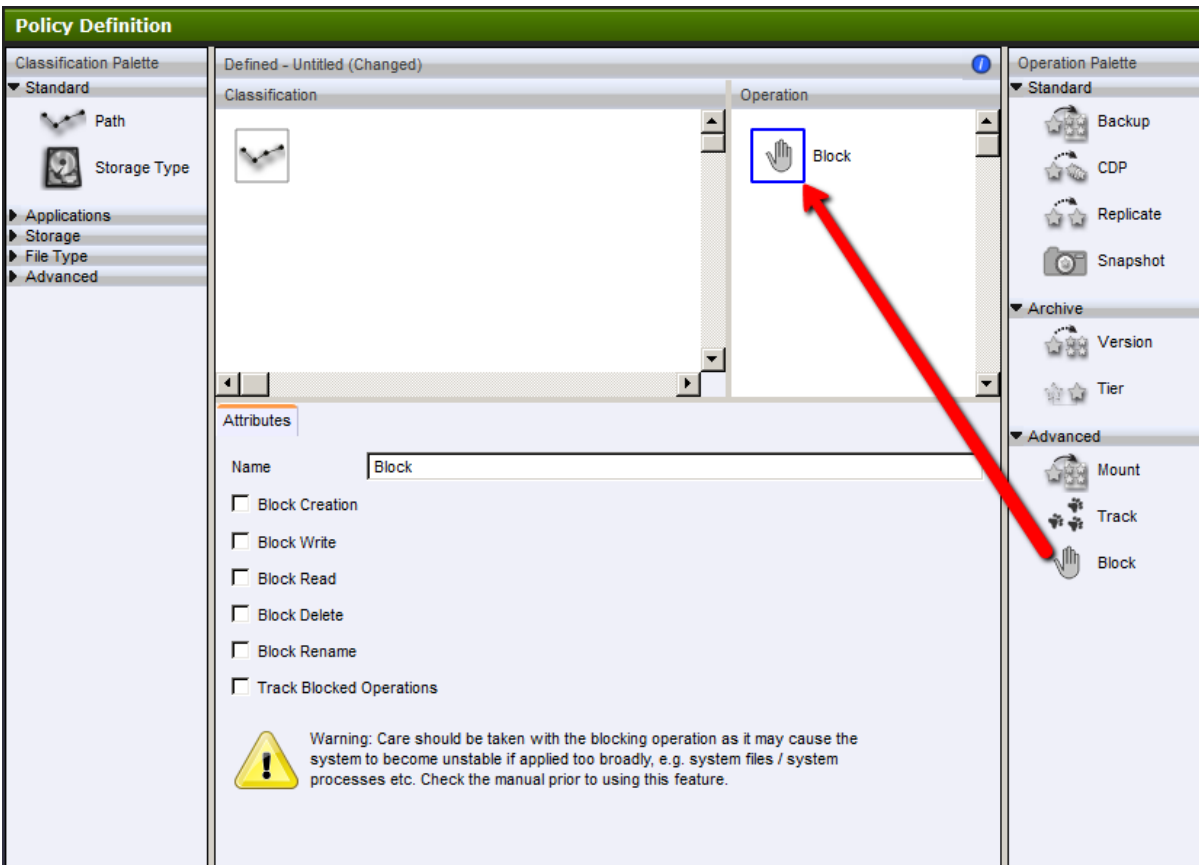


Figure 3-14 Block Policy Definition default settings

These options allow you to block selected operations on files in the directory listed for the path.

5. Select **File** → **Save** from the main menu to keep the policy.
6. Enter a policy name and brief description, and click **Save**.

Data flow definition

In the case of a blocking policy, the data flow defines on which machines the policy will be implemented.

In this example, the blocking policy will be assigned to the *fileserver1* node.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (Figure 3-15).

The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's **mover** operations.

Data Flow object type	Description
Source	Nodes installed with the <i>Data Source</i> component.
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option.
Mover	The type of data movement operation to be performed. Movers are not needed in blocking data flows.

In this example, place the source node, *fileserver1*, in the data flow.

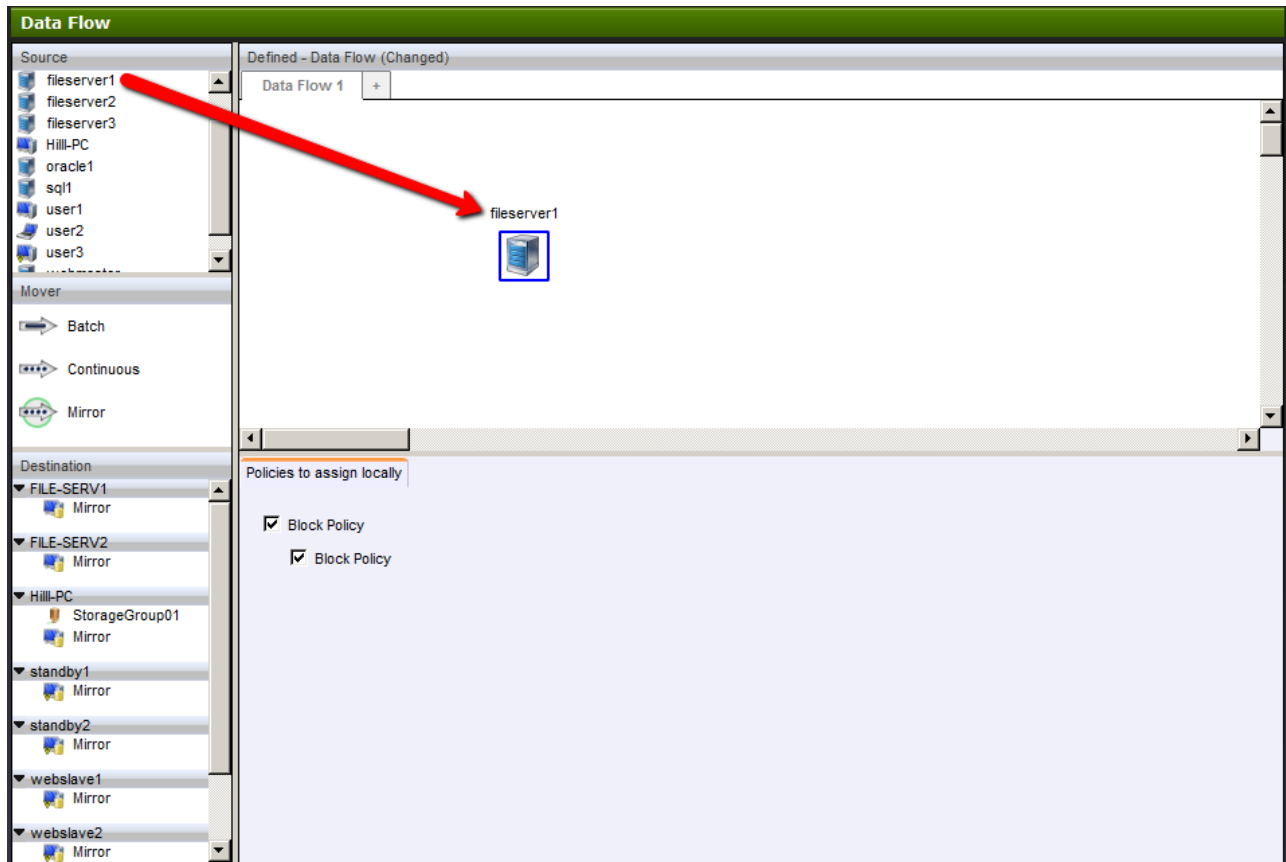


Figure 3-15 Block Policy Data Flow

2. Drag the **fileserver1** icon from the **Source** palette to the **Data Flow** workspace.

Be sure that the **fileserver1** icon is selected in the workspace. A list of defined policies is displayed in the **Policies to assign locally** tab at the bottom of the window screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.

3. Check Block Policy to select that policy.

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

Access to the data within c:\blockdirectory will now be restricted. If you selected 'Track block operations' during the policy definition stage then tracking messages regarding blocked operations can be seen in the **Reports** screen, under **Tracking**.

Track configuration

Data Instance Director can be used to track access to files or directories on any given source machine. Each time a file is accessed, a log message is generated that lists the type of access and the user who accessed the file.



Note: Only source nodes that are running on a Windows OS can be tracked.

The tasks to be performed when configuring a tracking policy are:



Policy definition – To define what data and actions should be tracked.



Data Flow definition – To define the node that should implement tracking policies.



Distribution and Activation – To put the policy into effect for the node.

Policy definition

A policy defines *data classification* and the *data operation*.

In this example, we wish to track access to the contents of C:\trackdirectory\ directory.



1. Click the **Policy Definition** icon  to navigate to the **Policy Definition** screen (Figure 3-16) and define a policy, as shown earlier in this chapter.
2. In this example, a directory needs to be defined. Expand the **Classification Palette - Standard** heading and drag the **Path** icon from the **Classification Palette** into the **Classification** workspace.
Be sure that the **Path** icon is selected in the workspace. The options specific to the path classification are displayed in the **Attributes** tab at the bottom of the screen.
3. Enter a path to block in the **Enter Path** edit box. In this example, enter C:\trackdirectory to define the path.

4. Expand the **Operation Palette - Advanced** heading and drag the **Track** icon from the **Operation Palette** to the **Operation** workspace.

Be sure that the **Track** icon is selected in the **Operation** workspace. The options related to the **Track** operation are displayed in the **Attributes** tab. In this example, all access is to be tracked, which is the default setting.

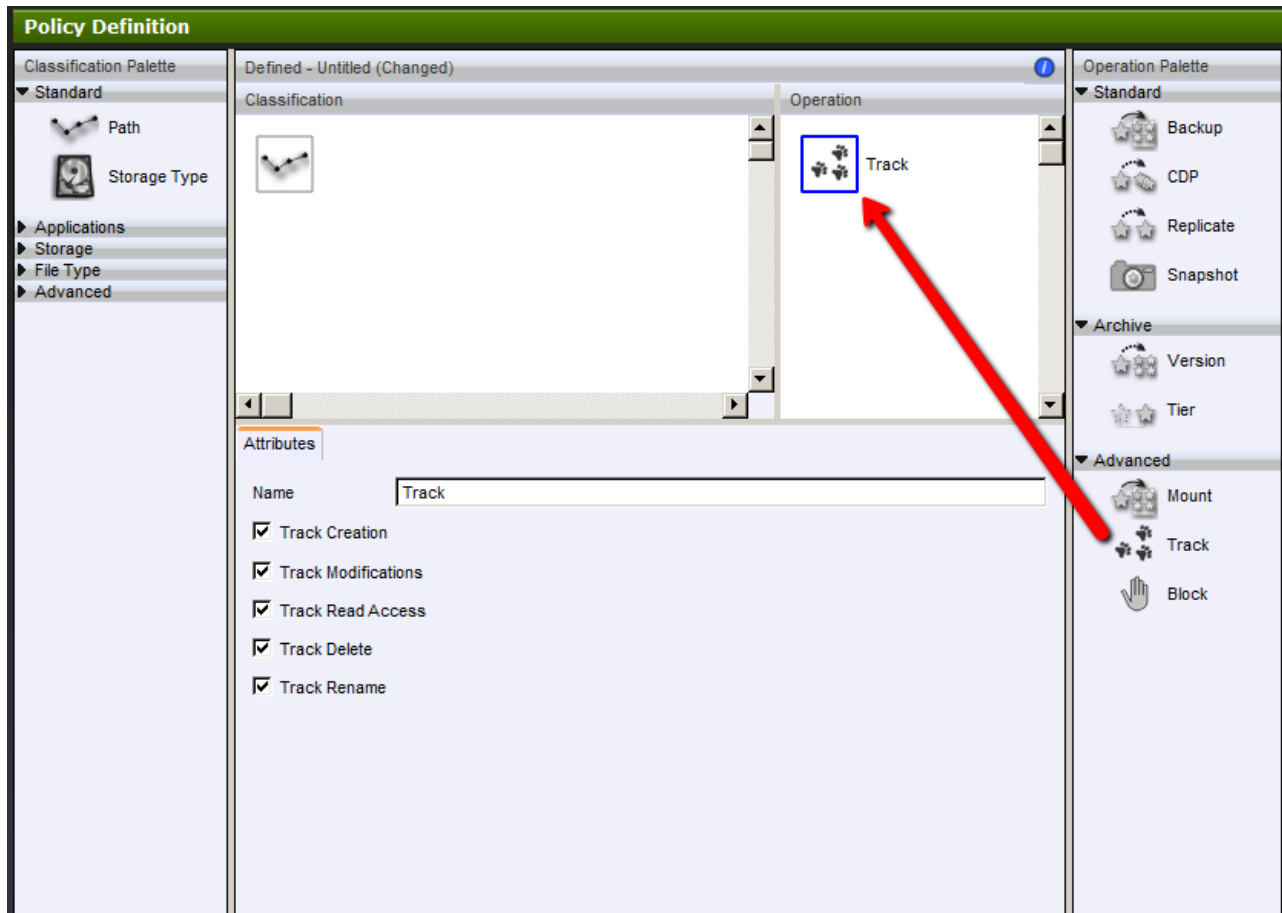


Figure 3-16 Track Policy Definition

5. Select **File** → **Save** from the main menu to keep the policy.
6. Enter a policy name and brief description, and click **Save**.

Data flow definition

In the case of a tracking policy, the data flow defines on which machines the policy will be implemented.

In this example, the tracking policy will be assigned to the *fileserver1* node.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (Figure 3-17).

The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's mover operations.

Data flow object type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed Movers are not needed in tracking data flows

2. Drag the **fileserver1** icon from the **Source** palette to the **Data Flow** workspace.

In the workspace, be sure that the **fileserver1** icon is selected. A list of defined policies is displayed in the **Policies to assign locally** tab at the bottom of the screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.

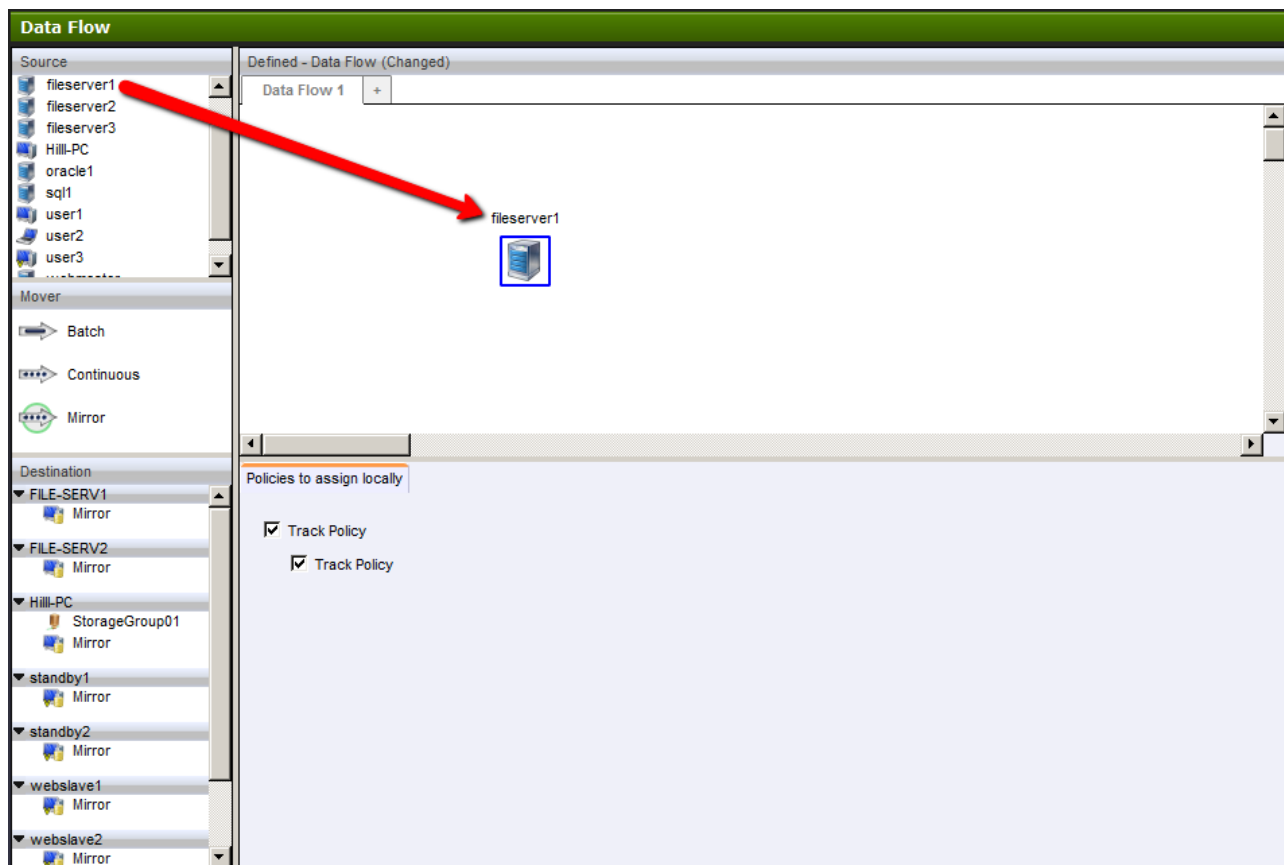


Figure 3-17 Track Policy data flow

3. Check **Track Policy** to select that policy.

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog displays.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

Modifying the data contained within C:\trackdirectory will now result in log messages. These can be viewed on the **Log Manager** and **Reports** screen.

Application backups: MS SQL, MS Exchange, and Hyper-V

Data Instance Director has the ability to back up certain applications such as MS SQL Server and Exchange Server, and it can back up virtual machines. Before backing up your applications, we recommend that the protection of your applications (such as Microsoft Exchange) is planned. Additionally, all intended restore options should be considered when planning the strategy to use to protect the application and application data.

Implementing a backup strategy

In order to ensure that all required data (such as configuration files and external index files) is available for a restore, and to ensure that a consistent state exists for all files (in the event that problems are introduced after the time of the last backup), we recommend that your site implements a suitable backup strategy. The backup strategy should include both the application classification along with a path classification, and should include all application files in the path classification, within a single policy, as described in this section.

If your site plans to restore a full machine, including a protected application, then in addition to the application backup, a full machine backup is required. Depending on the Recovery Point Objective (RPO, the frequency that the backup occurs) of the policies that are used, a Bare Metal Recovery (BMR) restore followed by an application restore may also be required. For information about completing a BMR, refer to the *Hitachi Data Instance Director Disaster Recovery Reference*.

Example:

Policy 1: Back up the entire machine (excluding the application and application files). Set the RPO to 24 hours.

Policy 2: Back up the application and application files. Set the RPO to 8 hours (or less, if required).

If a finer RPO is required to restore to the latest state available, adding CDP to the policy (with an RPO of 5 minutes) will cause a repository snapshot to be created every 5 minutes.

After the above policies have been implemented, all restore scenarios are available.



Notes: With live Exchange backup, it is necessary to define a policy with both an application classification along with a path classification that specifies all of the Exchange files. Configured applications should be quiesced (using the **quiesce configured applications before backup** option) before backing up the Exchange data on the source machine.

When restoring an Exchange server, you should complete the following tasks:

1. Stop Exchange services.
2. Move existing edb, chk, and log files to a temporary location. If space constraints prohibit the file move, you might need to delete the files.
3. Restore the required Exchange files.
4. Reboot the machine (or restart Exchange services).

If you are restoring an Exchange snapshot from a batch backup, then restoring only the data files which have been deleted is acceptable. However, if you are restoring an Exchange snapshot from a live backup, there is a greater risk of inconsistencies between Exchange data and other Exchange files. After the restore, there may be issues with accessing mailboxes. This problem will not occur if all Exchange files are restored and are in an application consistent state.

Backing up MS SQL, MS Exchange Server, or MS Hyper-V

This section relates specifically to backing up data from an application. Although the process is the same as a normal Backup, a unique Classification must be used. For both SQL and Exchange, the process is the same. Therefore, for this example, we will use MS SQL to show the required steps. Keep in mind, these steps also apply to Exchange Backups (including CDP). This example focuses on backing up the data from a MS SQL Server database to a Windows Store Group, *StorageGroup01* on *repository1*.



Note: Microsoft Windows Server 2012 supports Hyper-V hosted on Clustered Shared Volumes, with the following requirements:

- The Hyper-V for CSV configuration must use the **Batch** data mover operation.
 - **Quiesce configured applications** must be enabled for the Hyper-V for CSV configuration.
-

The tasks to be performed when configuring an application backup workflow are:



Policy definition – To define what data qualifies for backup.



Data Flow definition – To define what node's data should be backed up and where to.



Distribution and Activation – To put the policy into effect for each node in its data flow.

Application policy definition: SQL

A backup policy for a fully consistent application recovery point is created in a slightly different way from a typical backup policy. For this example, we will create a policy that utilizes CDP and Backup for the **MS SQL Server** application data. The backup operation will provide application consistent recovery points and the CDP operation will provide crash consistent recovery points.



1. Click the **Policy Definition** icon to navigate to the *Policy Definition* screen (Figure 3-18) and define a policy, as shown earlier in this chapter.
2. The **Classification Palette** on the left is used to define the data captured by the policy. In this example the only requirement is that we identify the SQL data. Expand the Applications heading and drag the **MS SQL Server** icon from the **Classification Palette** into the **Classification** workspace in the middle of the screen.

Be sure that the **MS SQL Server** icon that was placed into the workspace is selected. You can select Standard or Enterprise for type of SQL (this also applies to the Exchange Server classification).

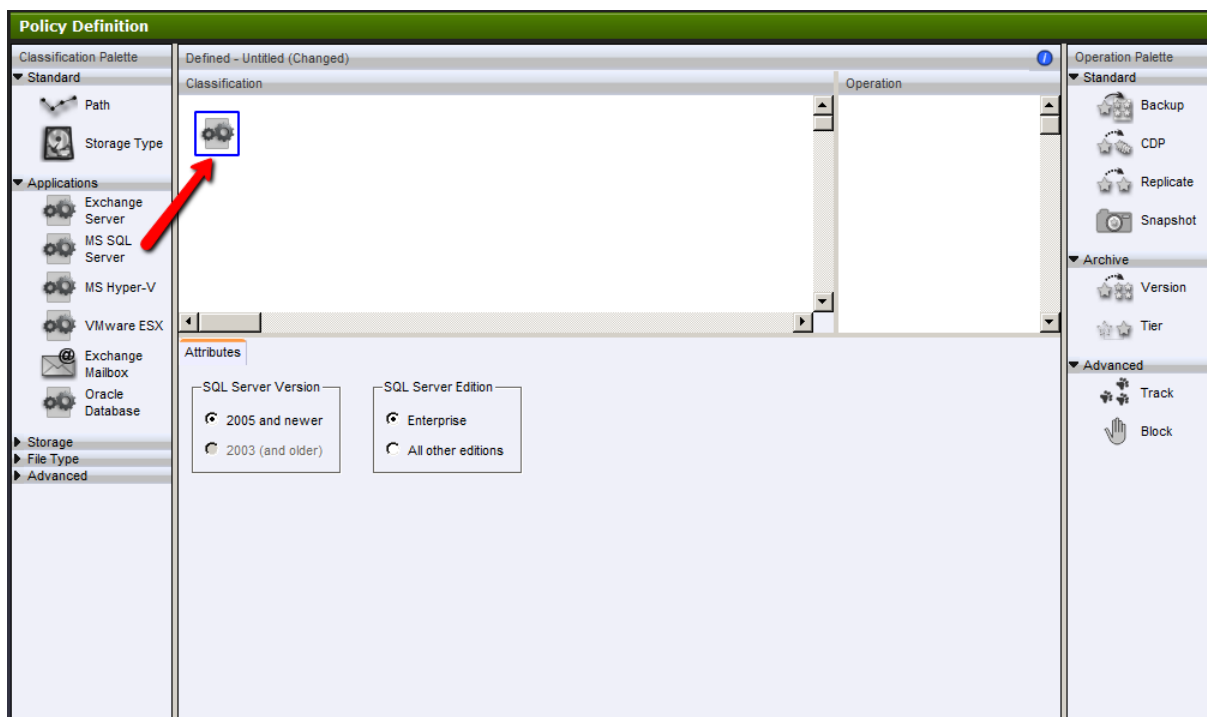


Figure 3-18 Defining a MS SQL Server Policy

3. The **Operation Palette** on the right is used to define the data operation. In this example, the data is to be protected by both CDP for granular RPO snapshots and Backup for broader RPO and longer term retention. Drag the **CDP** icon into the **Operation** workspace just to the left of that palette.
4. When the **CDP** icon is selected in the workspace area, the options related to the CDP operation are displayed in the **Attributes** tab (Figure 3-19 left). In this example, a 1 minute backup is required and the backup needs to be kept for 24 hours, so set the Retention to **24 Hours** and set the Recovery Point Objective to **1 Minute**.
5. Select the **Backup** icon and drop it below the **CDP** icon in the workspace area. You will see a small square appear under the **CDP** icon, drop the **Backup** icon on top of it. The options related to the Backup operation (Figure 3-19 right) need to be tuned a bit differently in the **Attributes** tab. In this example, a daily backup is required and the backup needs to be kept for four weeks, so set the Retention to **4 Weeks** and set the Recovery Point Objective to **1 Day**.
6. Under Backup attributes is a section called **Source Options** (Figure 3-20). Check **Quiesce configured applications before backup**. This only issues a quiesce during initial synchronization, but for all other snapshots Data Instance Director provides a quiesced and fully application consistent snapshot with zero impact upon the SQL services. This option is only available for the Backup operation. Select it for this example.

Retention	Recovery Point Objective
24 ☒ Minutes	1 ☒ Minutes
☒ Hours	☐ Hours
☐ Days	☐ Days
☐ Weeks	☐ Weeks

Retention	Recovery Point Objective
4 ☐ Minutes	1 ☐ Minutes
☐ Hours	☐ Hours
☐ Days	☒ Days
☒ Weeks	☐ Weeks

Figure 3-19 Retention and RPO settings for CDP and Backup respectively

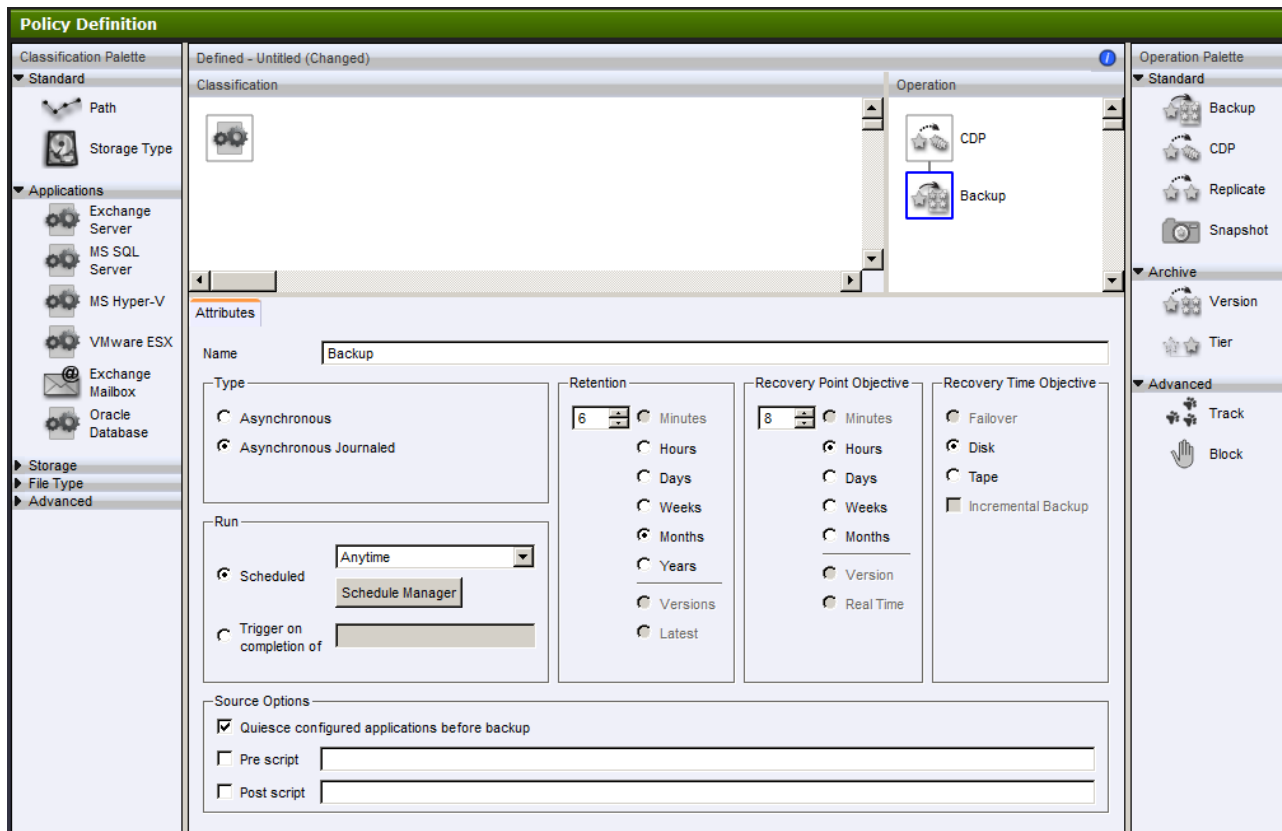


Figure 3-20 Backup operation with Quiesce selected for MS SQL Server

7. Select **File** → **Save** from the main menu to save the policy.
8. Enter a policy name and brief description, then click **Save**.

Data flow definition

A typical data flow for CDP and backup of MS SQL Server is similar to standard data flows for other policies already reviewed. Select the Source Node, select the Continuous Data Mover, and select the Store Group.

In this example, the source node will be *sql1* and the Store Group is called *StorageGroup01*.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (Figure 3-21). The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's **mover** operations.

Data flow object type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed

The **Continuous** mover operation must be used to activate real-time CDP and backup for the SQL data.

2. Drag the **sql1** icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Continuous** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *sql1*. If the **Continuous** icon is dropped on top of the **sql1** icon, they are automatically connected.
4. Drag the **StorageGroup01** icon from within *repository1* in the **Destination** palette and connect it by dropping it on to the **Continuous** mover icon.

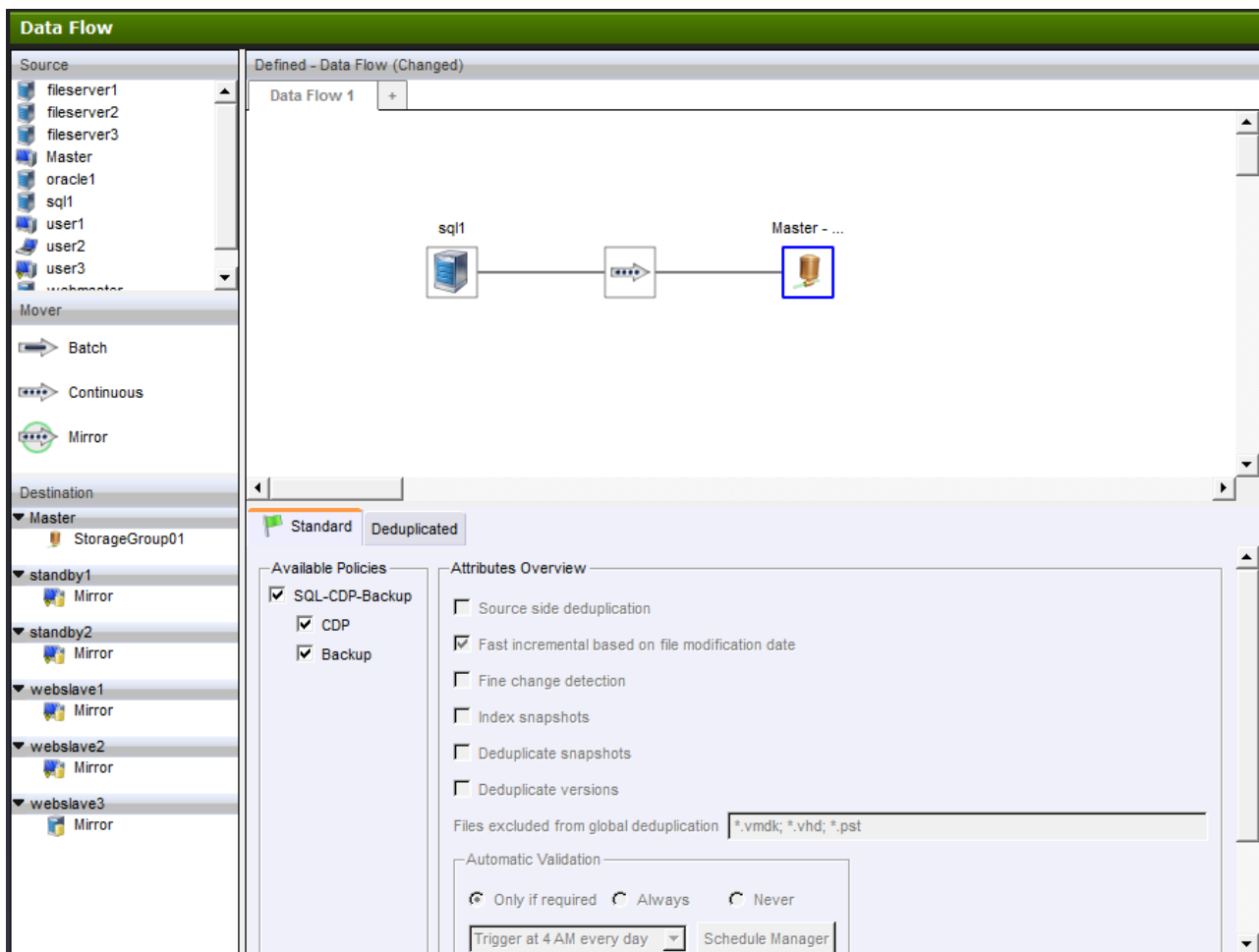


Figure 3-21 CDP and Backup Data Flow for MS SQL Server

5. In the workspace, select the **sql1** icon to display a list of defined policies in the **Available Policies** pane on the **Standard** tab at the bottom of the screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.
6. Check the SQL-CDP-Backup policy (the policy defined just now) to select that policy within the **Standard** tab.
7. Select the **StorageGroup01** icon and apply the policy items CDP and Backup to the destination node.

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the Distribute Rules button after the policy is compiled and the rules will be distributed and the policy activated.

The system performs an initial synchronization from the *sql1* node to the repository. Information regarding this process can be seen on the **Monitor** and **Log Manager** screens.

Once the initial synchronization is complete, the snapshots from the CDP and backup functions can be seen by performing a search in the **Restore** screen. More information on searching can be found in the [Restore](#) section.



Note: To backup an Exchange Server, follow the procedure outlined above for MS SQL Server; however, for Application Policy Definition step 2, substitute the **Exchange Server** icon. For Data Flow Definition step 2, use a Source Node which has Exchange Server installed.

To backup MS Hyper-V, follow the procedure outlined above for MS SQL Server; however, for Application Policy Definition step 2 substitute the **MS Hyper-V** icon. For Data Flow Definition step 2, use a Source Node which has Hyper-V installed.

Backing up a Clustered Share Volume with Hyper-V

The tasks to be performed when configuring an application backup workflow are:



Policy definition – To define what data qualifies for backup.



Data Flow definition – To define what node's data should be backed up and where to.



Distribution and Activation – To put the policy into effect for each node in its data flow.

Application policy definition: Hyper-V

For this example, we will create a policy that utilizes Backup for **MS Hyper-V** in a CSV environment.

1. Click the **Policy Definition** icon  to navigate to the *Policy Definition* screen (Figure 3-22) to define the policy.
2. The **Classification Palette** on the left is used to define the data captured by the policy. In this example the only requirement is that we identify the Hyper-V data. Expand the Applications heading and drag the **MS Hyper-V** icon from the **Classification Palette** into the **Classification** workspace in the middle of the screen.

With the **MS Hyper-V** icon selected, notice that the options on the **Attributes** tab are disabled and **All virtual machines** is selected.

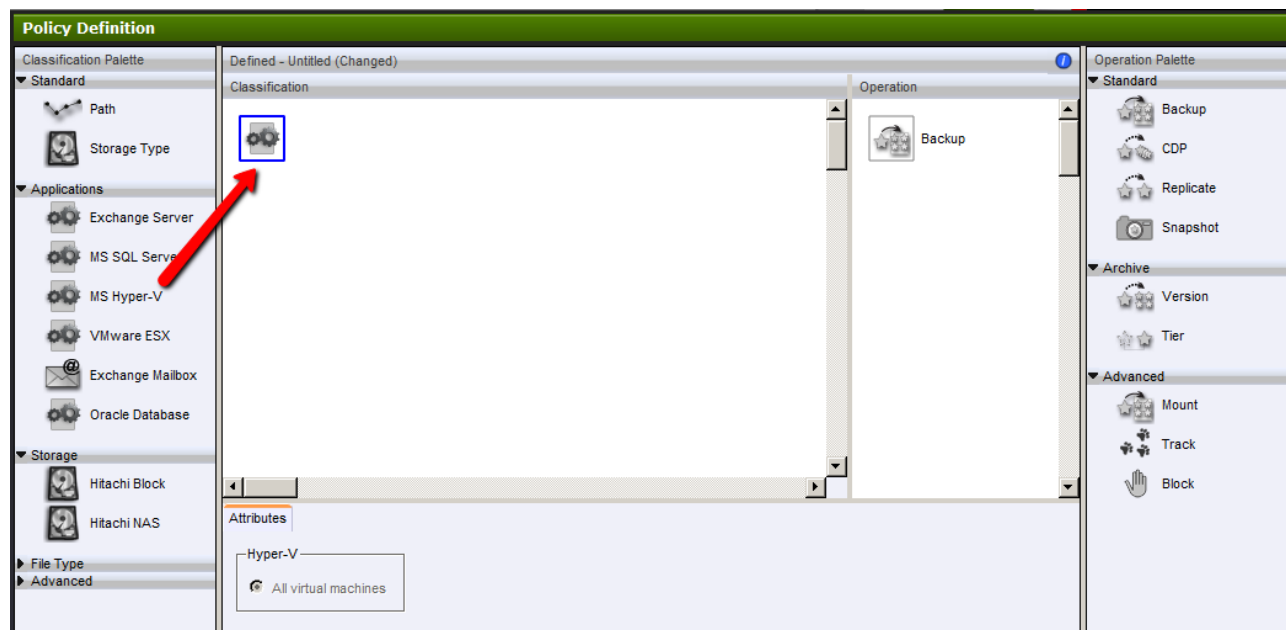


Figure 3-22 Defining a MS Hyper-V Policy

3. The **Operation Palette** on the right is used to define the data operation. In this example, the data is to be protected by Backup. Drag the **Backup** icon into the **Operation** workspace just to the left of that palette.
4. The options related to the Backup operation (Figure 3-23) need to be configured in the **Attributes** tab. In this example, a daily backup is required and the backup needs to be kept for four weeks, so set the Retention to **4 Weeks** and set the Recovery Point Objective to **1 Day**.
5. Under Backup attributes is a section called **Source Options** (Figure 3-24). Check **Quiesce configured applications before backup**. This option is only available for the Backup operation. Select it for this example.

Retention	Recovery Point Objective
4 ☐ Minutes ☐ Hours ☐ Days ☒ Weeks	1 ☐ Minutes ☐ Hours ☒ Days ☐ Weeks

Figure 3-23 Retention and RPO settings for Backup

Figure 3-24 Backup operation with quiesce selected for MS Hyper-V

6. Select **File** → **Save** from the main menu to save the policy as *HV-CSV-Backup*.
7. Enter a policy name and brief description, then click **Save**.

Data flow definition

A typical data flow for Backup of MS Hyper-V on CSV is similar to standard data flows for other policies already reviewed except that multiple source nodes are connected to the same Batch mover. Select the Source Nodes, select the Batch Data Mover, and select the Store Group.

In this example, the source nodes will be *CSVHV1* and *CSVHV2*. The *Store Group* is called *QAHO17 – StorageGroup01*.

8. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (Figure 3-25).

The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's **Mover** operations.

Data flow object type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed

The **Batch** mover operation must be used to activate backups for the Hyper-V data.

9. Drag the *CSVHV1* and *CSVHV2* icons from the **Source** palette to the **Data Flow** workspace.
10. Drag the **Batch** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *CSVHV1* and *CSVHV2*.
11. Drag the *StorageGroup01* icon from within *QAHO17* in the **Destination** palette and connect it by dropping it on to the **Batch** mover icon.

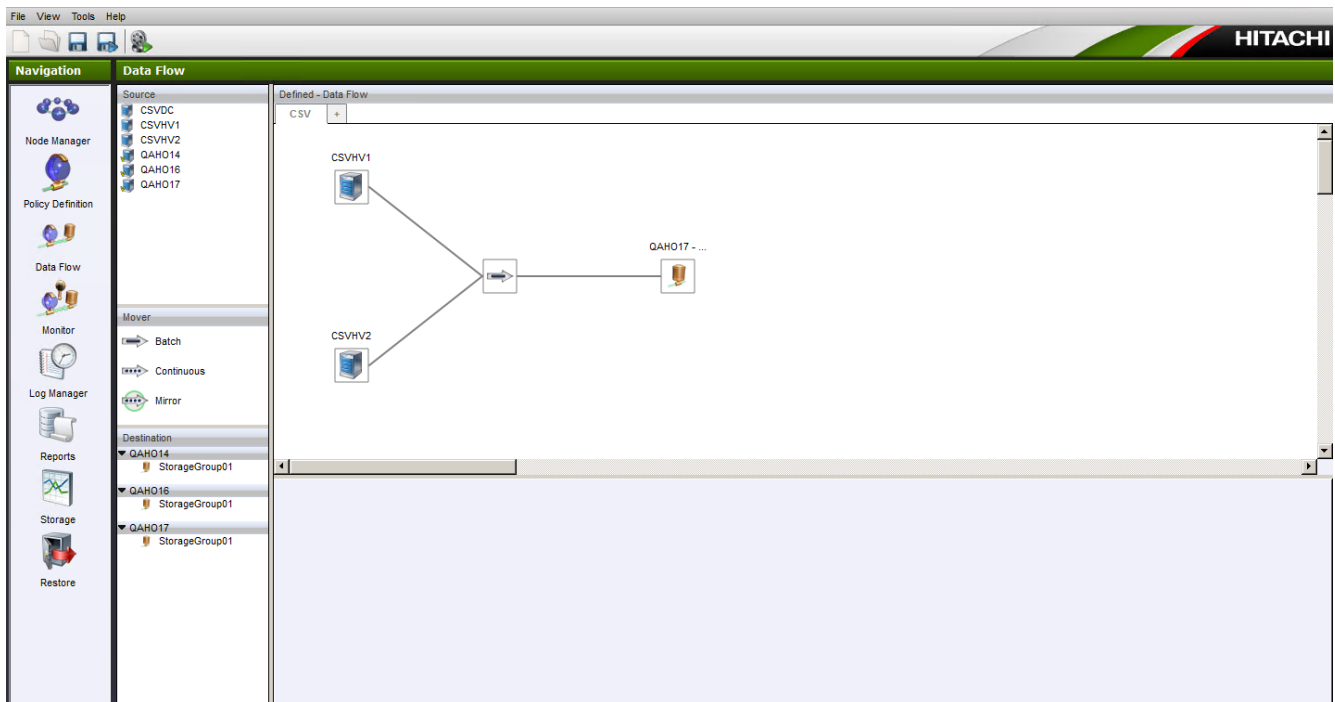


Figure 3-25 CDP and Backup Data Flow for MS SQL Server

12. In the workspace, select the CSVHV1 and CSVHV2 icons to display a list of defined policies in the **Available Policies** pane on the **Standard** tab at the bottom of the screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.
13. Check the *HV-CSV-Backup* policy (the policy defined above) to select that policy within the **Standard** tab.
14. Select the *QAHO7 - StorageGroup01* icon and apply the policy item Backup to the destination node.

Distribution and activation

15. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
16. Click the Distribute Rules button after the policy is compiled and the rules will be distributed and the policy activated.

The system performs an initial synchronization from the CSVHV1 and CSVHV2 nodes to the repository. Information regarding this process can be seen on the **Monitor** and **Log Manager** screens.

Once the initial synchronization is complete, the snapshots from the backup function can be seen by performing a search in the **Restore** screen.

Restoring a single Hyper-V virtual machine (cluster)

When restoring a single VM, either the entire VM can be restored or the virtual hard disk can be restored and a new VM created from it.

Before you begin, do the following:

- On the Hyper-V machine, shut down the originating VM.
- Some versions of Hyper-V may require stopping the **Virtual Machine Management Service**.

Note: The following procedure cannot be used to restore individual Hyper-V VM snapshots.

1. Identify the relevant virtual machine files within the Data Instance Director snapshot. (The .xml, .gui, and .bin files use a unique string to identify the virtual machine.)
2. Within Windows PowerShell, translate the string name to the virtual machine name, using the applicable command:
 - Server 2008 R2:


```
Get-WmiObject -Namespace root\virtualization -class msvm_computersystem | select elementname, name | ft -auto
```
 - Server 2012 R2:


```
Get-VM | select VMName, VMId
```
3. Delete the originating VM within the **Hyper-V Manager** (not in the **Failover Cluster Manager**).

4. Select the following items for restoration:
 - I. The directory and XML file which have the name identified in step 1.
 - II. The Virtual Hard Disk (.vhd, or .vhdx) file which needs to be restored, and which is located within the snapshot.
 - III. Any file that is prepended with the Virtual Machine name.
5. Restore to the original location with the **Differentiate file name** option deselected.

By default, the Virtual Machine should be re-added to the Hyper-V Manager. The restore process can take up to an hour to complete.

Agentless nodes

Data Instance Director allows the configuration of topologies involving agentless nodes, such as Windows shares, VMware ESX hosts and Tape libraries.

Backing up a Windows share

This section relates specifically to backing up data contained within a Windows share, although the general sequence of operation can be extended to any CIFS implementation. This example will focus on backing up the share data to a Windows Store Group, *StorageGroup01* on *repository1*.

The tasks to be performed when configuring a share backup policy are:



Node creation – To inform Data Instance Director of the share credentials and details.



Policy definition – To define what data should be captured from the share.



Data Flow definition – To define what share should have its data captured and where it should be backed up to.



Distribution and Activation – To put the policy into effect for each node in its data flow.

Node creation

First, a node representing the share needs to be created.



1. Select **Node Manager** in the **Navigation** pane to access the **node manager** screen.
2. In the **Agentless / Hardware Node** section, select CIFS from the combo box.
3. Click **Create**. A configuration dialog is displayed (Figure 3-26).

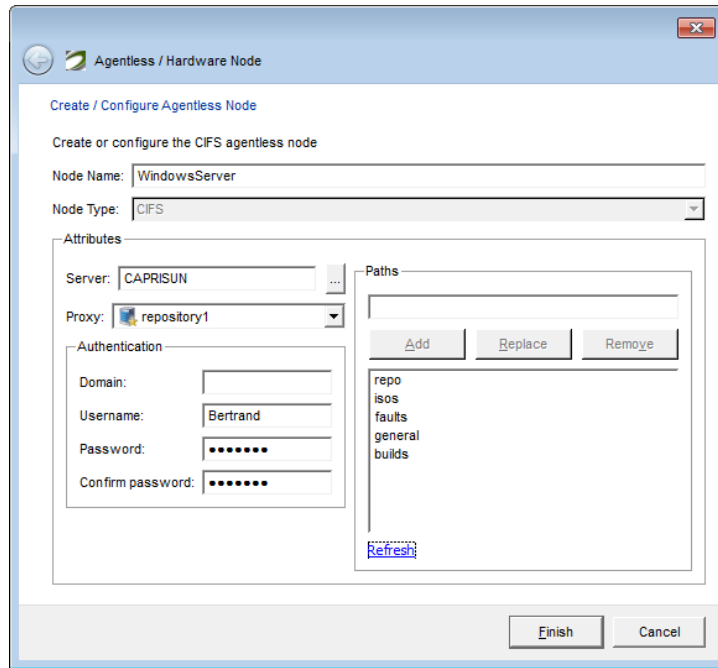


Figure 3-26 CIFS Configuration

4. Enter a name in the **Node Name** edit box at the top of the dialog by which to identify the node. We will use the name *WindowsServer*.
5. In the **Server** edit box enter the name or IP address of a machine with valid shares. Click the  button and browse the network for a suitable machine. This button is not available for Linux.
6. A proxy node that is responsible for interfacing with the server is required. This must be a Windows machine and, to avoid excessive network transfer, should be as close as possible to the eventual destination of the server's data. In our case, *repository1* is suitable.
7. Enter the credentials for the server chosen in the previous step. The user should have system level access to all the files and must either be in the Administrator or Backup Operator user groups. For security, we recommend using a **Backup Operator**, in which case additional security access privilege is required for the user on the remote machine:
 - a. Launch the **Local Security Policy** management console: Either type 'secpol.msc' in the **Run** dialog from the **Start** menu, or double-click the icon in **Administrative Tools** of the **Control Panel**.
 - b. Navigate to **Local Policies** → **User Rights Assignment**.
 - c. Double-click **Manage auditing and security log**.
 - d. Add the required user.

8. Add the shares that should be visible to Data Instance Director into the **Paths** list box on the right. A hyperlink style **Refresh** button is available at the bottom. This will auto-fill the **Paths** list box with all available shares detectable on the supplied server. If the server is not accessible or the credentials supplied are invalid, then an error message will be displayed.



Note: When entering shares manually, only the actual share name is required. The server name, and any auxiliary slashes (/, \) can be omitted.

9. Click **Finish** when ready.
10. The new node with the name *WindowsServer* will now appear in the **Unauthorized Nodes** list. Select the node in this list, then click the **Authorize** button  to authorize the node.

If the server cannot be accessed, its credentials are invalid, or a supplied share cannot be found, then the node will fail authorization. The configuration dialog can be reopened to fix errors by right-clicking on the node and selecting **Edit node**. Ensure UAC is turned off.



Caution: Common sources of errors during CIFS backups include the following:

- Credentials supplied when creating the node might be incorrect. Verify the supplied username and password by right-clicking on the node in **Node Manager** and selecting **Edit Node**.
 - The guest account might be enabled on the server, causing Windows to fall back to it when supplied with incorrect credentials. Check the system security logs to see which user logged on.
 - The user might have incorrect privileges. They should have system level access as described above. Privileges can be verified by checking the system security logs for 'Special Logon' records, and ensuring that SeSecurityPrivilege, SeBackupPrivilege and SeRestorePrivilege are present.
-

Policy definition

A backup policy for a CIFS share is created in a slightly different way from a typical backup policy. For this example, we will back up the contents of the general share on *CAPRISUN*.



1. Click the **Policy Definition** icon  to navigate to the **Policy Definition** screen (Figure 3-27) and define a policy, as shown earlier in this chapter.
2. The **Classification Palette** on the left is used to define the data captured by the policy. In this example, the only requirement is that the data resides within the general share. Expand the **Advanced** heading and drag the **Remote Path** icon from the **Classification Palette** into the **Classification** workspace in the middle of the screen.

Be sure that the **Remote Path** icon that was just placed into the workspace is selected. Then, **Include paths** and **Exclude paths list**

boxes to enter applicable remote paths are displayed in the **Attributes** tab at the bottom of the screen.

3. **Available Paths** (shares) are shown in the list box on the left. The general share is the only entry required. Move it to the **Included Paths** list box on the right by clicking **Include** or double clicking the entry.

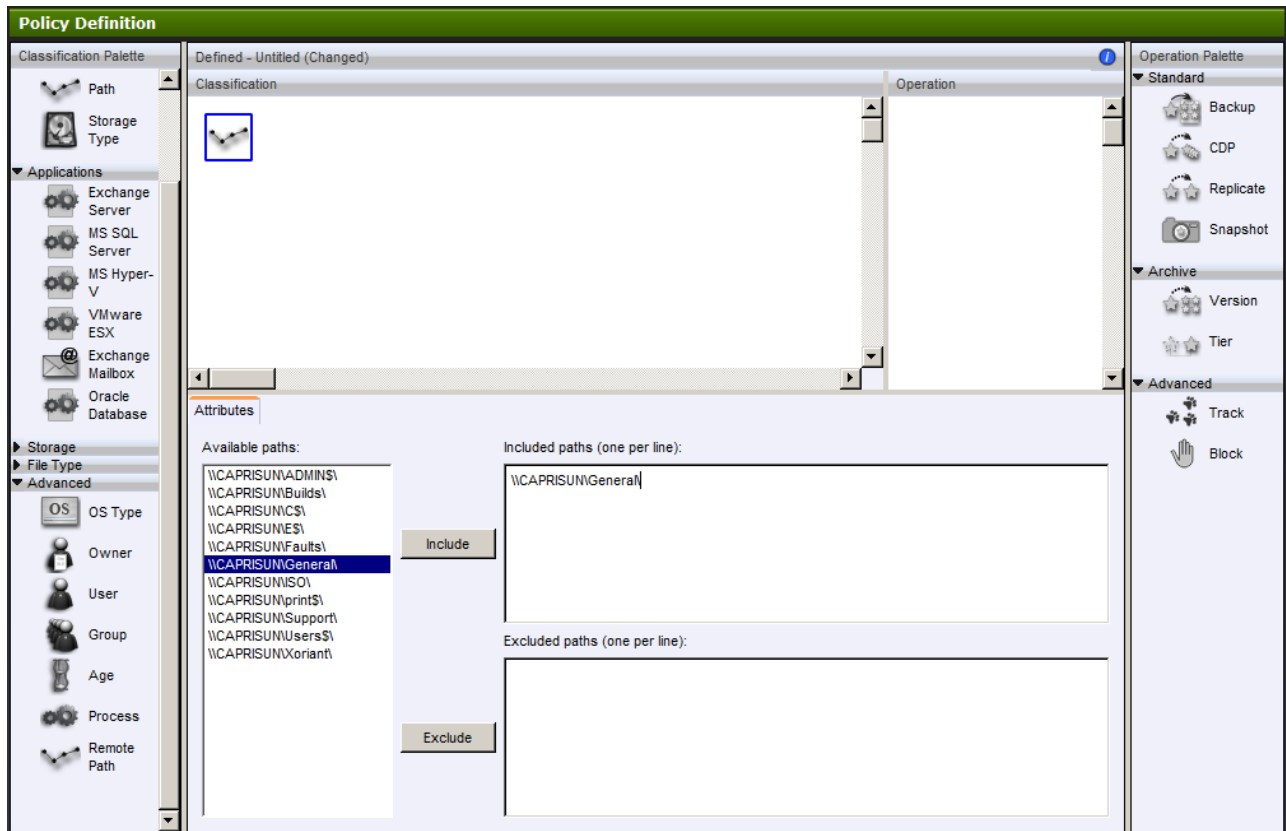


Figure 3-27 Defining a Remote Path Policy

4. If only data from a subdirectory of the share is required – as opposed to the whole share - the subdirectory name can be appended to the share name by typing into the **Included paths** list box.
5. The **Operation Palette** on the right is used to define the data operation (Figure 3-28). In this example, the data is to be backed up. Drag the **Backup** icon into the **Operation** workspace just to the left of that palette. Only Backup and Version operations are applicable to remote CIFS nodes.
6. When the **Backup** icon is selected in the workspace area, the options related to the backup operation are displayed in the **Attributes** tab. In this example, a daily backup is required and the backup needs to be kept for two weeks, so set the Retention to **2 Weeks** and set the Recovery Point Objective to **1 Day**.

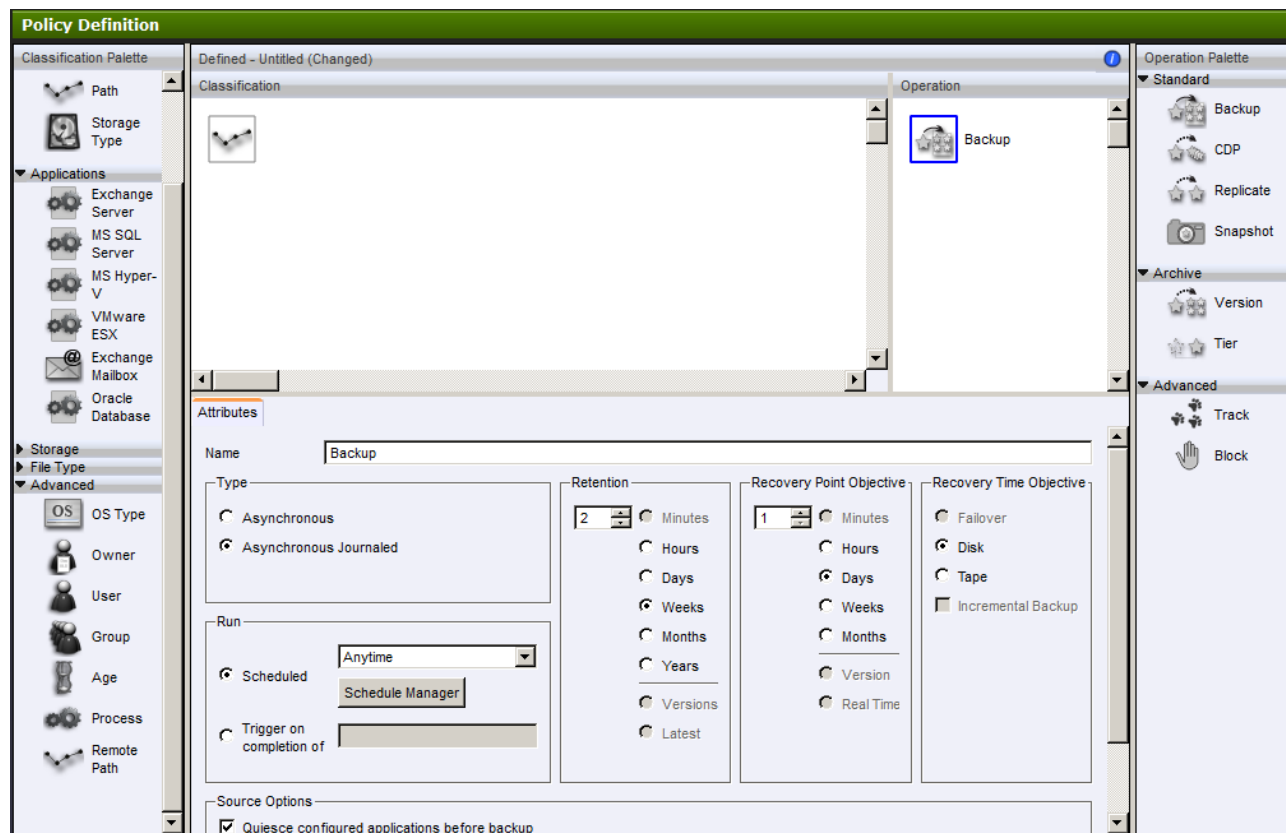


Figure 3-28 Defining a Backup Operation

7. Select **File** → **Save** from the main menu to save the policy.
8. Enter a policy name and brief description, and click **Save**.

Data flow definition

A typical data flow for backing up from a share consists of transferring the share data to the proxy, and transferring this data in turn to a repository. If the proxy and the repository are the same machine then the data can be transferred directly from the share to the repository.

In this example, the source node will be *WindowsServer* and the repository is called *StorageGroup01* on *repository1* (which in this case is also the proxy).



Note: It is only possible to specify a batch mover when defining a data flow from a CIFS node.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (Figure 3-29). The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's **mover** operations.

Data Flow Object Type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed

The *Batch* mover operation must be used to back data up from a share.

2. Drag the **WindowsServer** icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Batch** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *WindowsServer*. If the **Batch** icon is dropped on top of the **WindowsServer** icon, they are automatically connected.
4. Drag the **StorageGroup01** icon from within *repository1* in the **Destination** palette and connect it by dropping it on to the **Batch** mover icon.

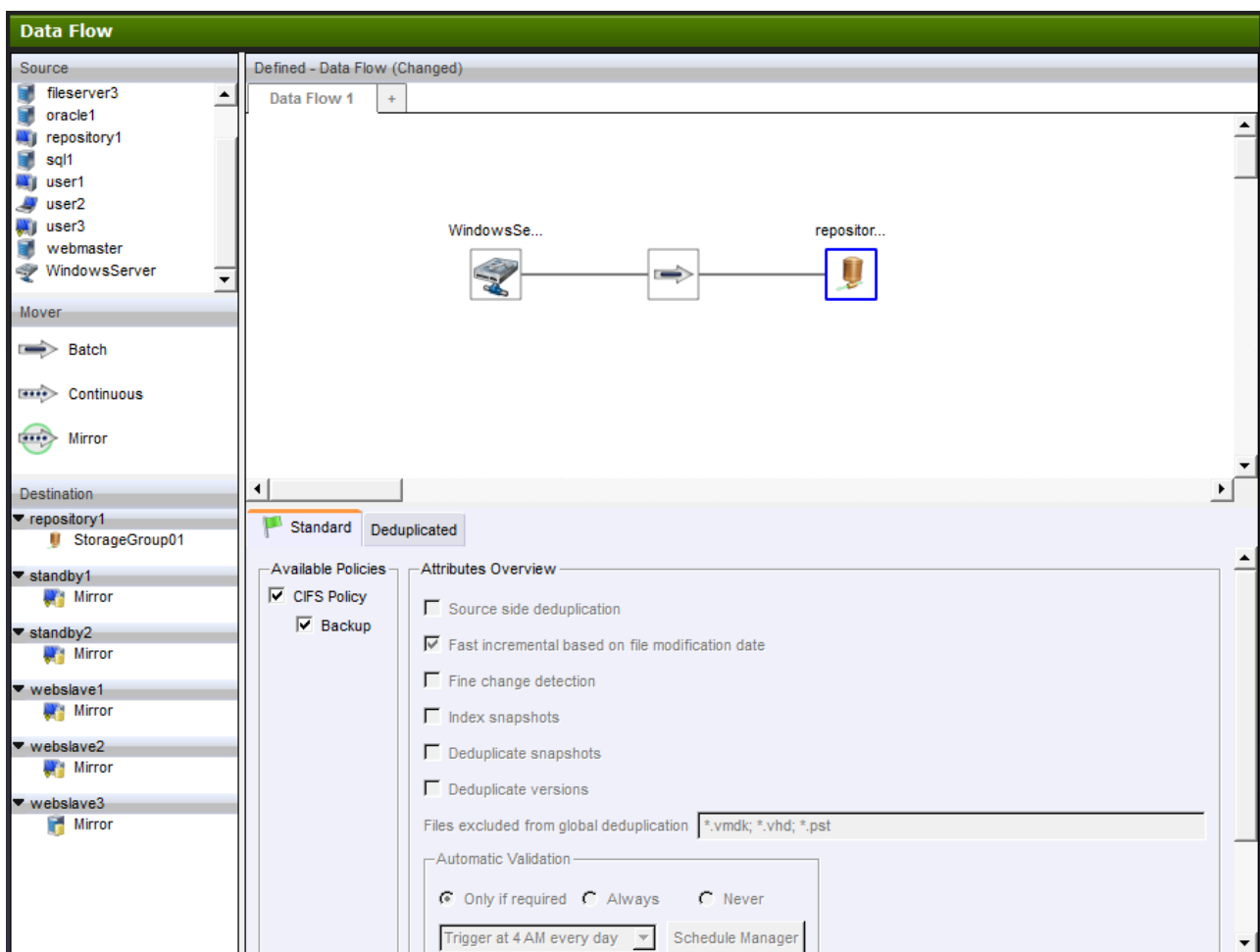


Figure 3-29 Backup Data Flow

5. In the workspace, select the **WindowsServer** icon to display a list of defined policies in the **Available Policies** pane on the **Standard** tab at the bottom of the screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.
6. Check the CIFS backup policy (the policy defined just now) to select that policy.
7. Select the **repository1** icon and apply the policy to the destination node.

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the Distribute Rules button after the policy is compiled and the rules will be distributed and the policy activated.

The system performs an initial synchronization from the share to the repository. Information regarding this process can be seen on the **Monitor** and **Log Manager** screens.

Once the initial synchronization is complete the backup can be seen by performing a search in the **Restore** screen. More information on searching can be found in the [Restore](#) section of Chapter 5.

Backing up a VMware ESX virtual machine

This section relates specifically to backing up virtual machines on an ESX host. This example will focus on backing up a virtual machine to a Windows Store Group, *StorageGroup01* on *repository1*.

The tasks to be performed when configuring an ESX backup policy are:

-  Node creation – To inform Data Instance Director of the ESX host credentials and details.
-  Policy definition – To define which virtual machines are to be backed up.
-  Data Flow definition – To define which host should have its virtual machines captured and where it should be backed up to.
-  Distribution and Activation – To put the policy into effect for each node in its data flow.



Note: This tutorial refers to a VMware ESX host, but the steps also apply to ESXi hosts and vCenter.

Node creation

First, a node representing the VMware ESX server needs to be created.

1. Select **Node Manager**  in the **Navigation** pane to access the **Node Manager** screen.
2. In the **Agentless / Hardware Node** section, select 'VMware ESX' from the combo box.
3. Click the **Create** button. A configuration dialog is displayed (Figure 3-30).

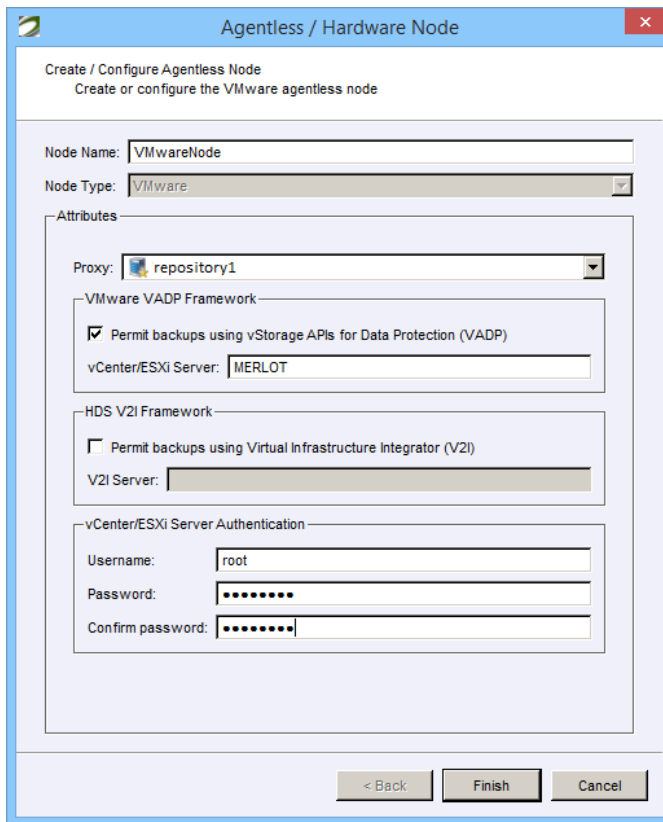


Figure 3-30 VMware ESX Agentless Node Configuration

4. Enter a **Node Name** in the text box at the top of the dialog by which to identify the node. We will use the name *VMwareNode*.
5. A **Proxy** node that is responsible for interfacing with the server is required. This must be a Windows machine and, to avoid excessive network transfer, should be as close as possible to the eventual destination of the server's data. In our case, the machine named *repository1* is suitable.
6. Check **Permit backups using vStorage APIs for Data Protection (VADP)**.
7. In the **vCenter / ESXi Server** text box enter the name or address of an ESX host, here we will use *MERLOT*.
8. Enter the credentials for the vCenter/ESXi server chosen in the previous step.

9. Click **Finish** when ready.
10. The new node with the name *VMwareNode* will now appear in the **Unauthorized Nodes** list. Select the node in this list and then click the **Authorize** button  to authorize the node.

If the server cannot be accessed or its credentials are invalid then the node will fail authorization. The configuration dialog can be reopened in order to fix any errors by right-clicking on the node and selecting **Edit node**.



WARNING: If you use vCenter to manage your ESX or ESXi host, then you must backup your VMs through vCenter. If you fail to do this, the VM cannot be restored due to security issues.



Note: If you change the proxy of an ESX node while the node is in an activate data flow, or if you modify the username or password of the ESX node, then you must redistribute the rules in order for the changes to take effect.

Policy definition

A backup policy for a VMware ESX host is created in a slightly different way from a typical backup policy. For this example, we will back up a single virtual machine on *VMwareNode*

1. Select Policy Definition  in the **Navigation** Pane to access the **Policy Definition** screen and define a policy (Figure 3-31).
2. The **Classification Palette** on the left is used to define the data captured by the policy. Expand the **Applications** heading and drag the **VMware ESX** icon from the **Classification Palette** into the **Classification** workspace in the middle of the screen.

Be sure that the **VMware ESX** icon that was just placed into the workspace is selected. Virtual Machines can be selected by clicking the browse button located to the right of the **Enter Virtual Machines** list box. In the **Choose virtual machine(s)** dialog, select the node from the combo box. This will display a list of Virtual Machines which can be selected to add to the policy.

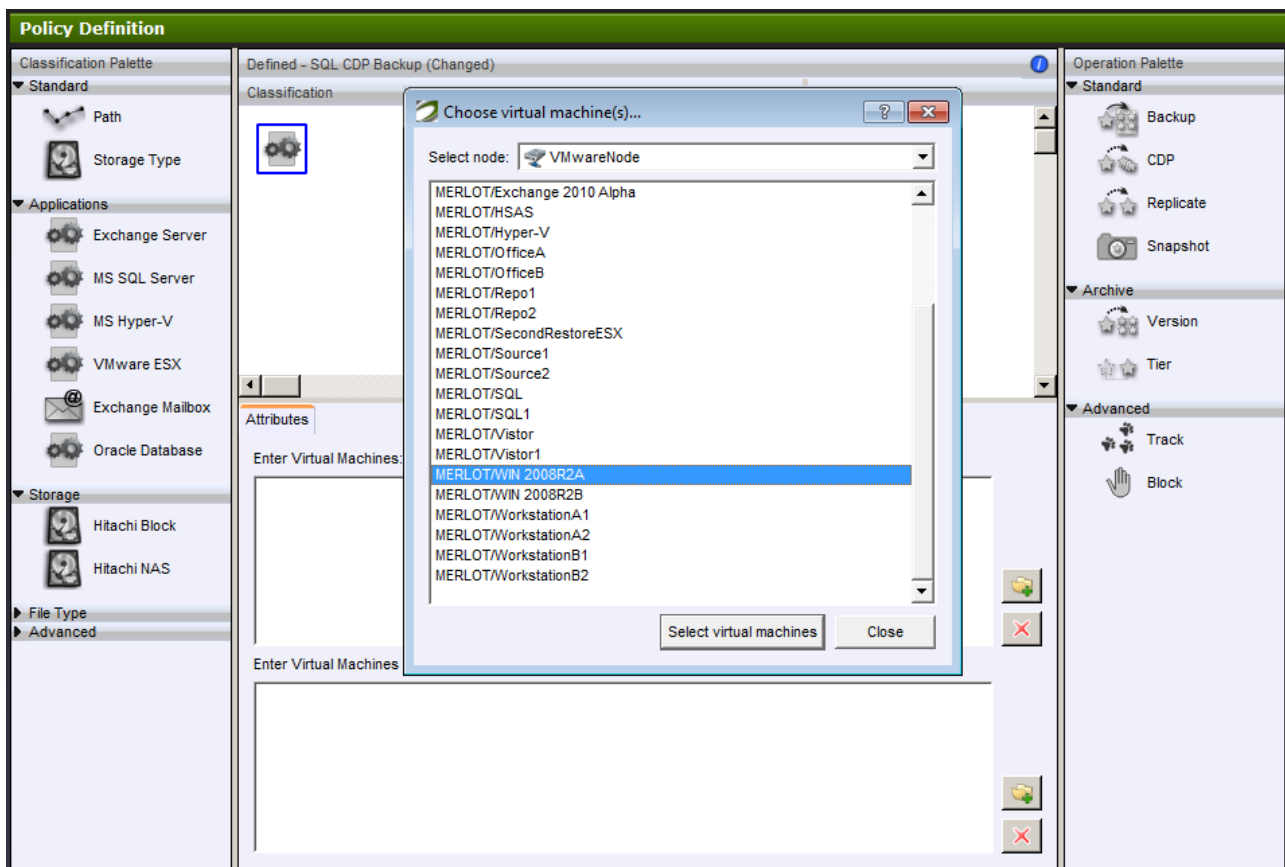


Figure 3-31 Defining a VMware ESX Policy

3. The **Operation Palette** on the right is used to define the data operation. In this example, the virtual machine(s) is to be backed up. Drag the **Backup** icon into the **Operation** workspace just to the left of that palette. Only the Backup operation is applicable to VMware ESX nodes.
4. When the **Backup** icon is selected in the workspace area, the options related to the backup operation are displayed in the **Attributes** tab (Figure 3-32). In this example, an hourly backup is required and the backup needs to be kept for two weeks, so set the Retention to **2 Weeks** and set the Recovery Point Objective to **1 Hour**.

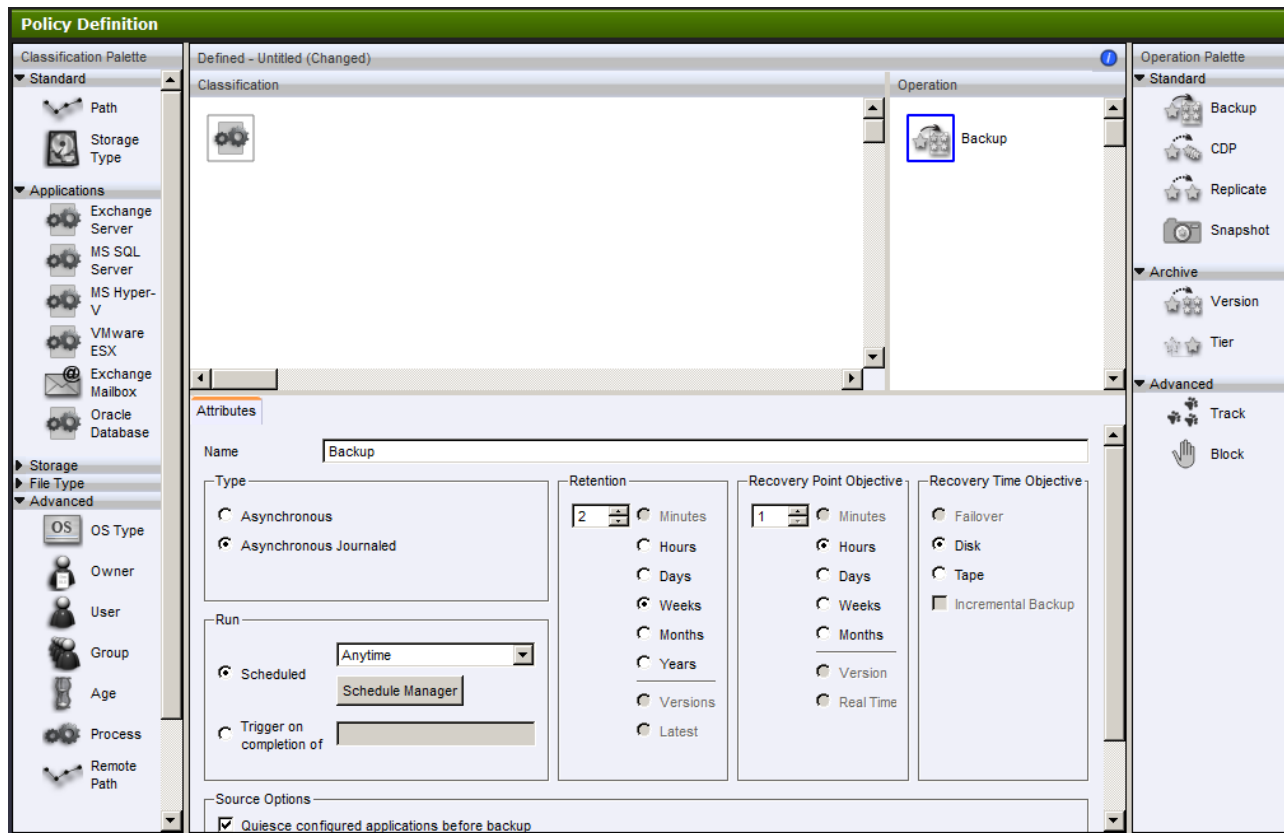


Figure 3-32 Defining a Backup Operation

5. Select **File** → **Save** from the main menu to save the policy.
6. Enter a policy name and brief description, and click **Save**.

Data flow definition

A typical data flow for backing up virtual machines from a VMware ESX host consists of transferring the data to the proxy, and transferring this data in turn to a repository. If the proxy and the repository are the same machine then the data can be transferred directly from the VMware ESX host to the repository.

In this example, the source node will be *VMwareNode* and the Store Group is called *StorageGroup01* on *repository1* (which in this case is also the proxy).



Note: It is only possible to specify a batch mover when defining a data flow from a VMware ESX node.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (Figure 3-33). The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's **mover** operations.

Data Flow Object Type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed

The *Batch* mover operation must be used to back data up from a VMware ESX host.

2. Drag the **VMwareNode** icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Batch** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *VMwareNode*. If the **Batch** icon is dropped on top of the **VMwareNode** icon, they are automatically connected.
4. Drag the the **StorageGroup01** icon from within *repository1* in the **Destination** palette and connect it by dropping it on to the **Batch** mover icon.

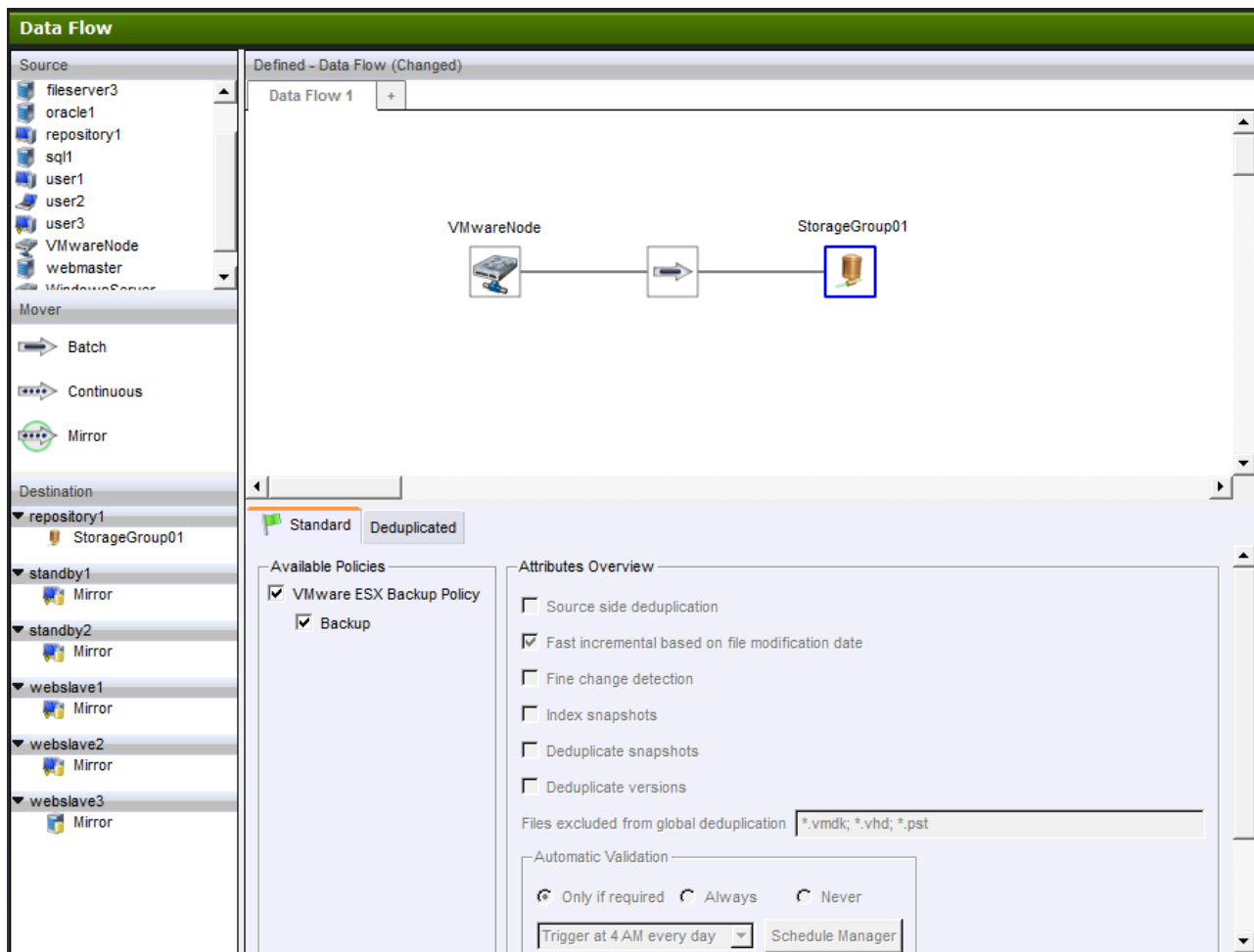


Figure 3-33 Backup Data Flow

5. In the workspace, select the **VMwareNode** icon to display a list of defined policies in the **Available Policies** pane on the **Standard** tab at the bottom of the screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.
6. Check the VMware ESX backup policy (the policy defined above) to select that policy.
7. Select the **StorageGroup01** icon and apply the policy to the destination node.

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.



Note: If the store group has an **Automatic Validation** schedule enabled (in the store group's **Attributes Overview** pane on the **Standard** tab), the compilation will return a warning. It is safe to distribute these rules despite the warning as the option is ignored for ESX backups.



Note: The VM's to be backed up must be placed in the powered off state prior to the initial synchronization. Data Instance Director will perform the initial synchronization of the VMware ESX host to the repository at a time dictated by the backup schedule. VM power down may thus need to be scheduled to occur just prior to initial synchronization in order to minimize disruption.

2. Click the Distribute Rules button after the policy is compiled and the rules will be distributed and the policy activated. Information regarding this process can be seen on the **Monitor** (Figure 3-34) and **Log Manager** screens. The initial synchronization will transfer only the used blocks on the virtual machines' disks.
3. After the initial synchronization is complete, the VMs can be returned to the powered on state. Subsequent backups will only transfer the changed blocks from the previous backup. Each backup however will be, in effect, a full backup so when performing a restore just the one restore operation is required.



Note: Data Instance Director will try to enable Changed Block Tracking (CBT) if it is not already enabled on any virtual machines on the ESX host that are to be backed up. For CBT to be enabled, the virtual machine must be shut down. If the virtual machine does not have CBT enabled or Data Instance Director fails to enable it, then the backup will fail.

It is likely that enabling CBT will need to be scheduled, given that the virtual machine needs to be in the powered off state for CBT to be turned on. Instructions to manually enable CBT can be found on the VMware knowledgebase:

<http://kb.vmware.com/selfservice/microsites/search.do?cmd=displayKC&externalId=1031873>.



Note: With CBT working correctly, the resynchronization progress will appear to show it sending the entirety of the virtual machines' disks, however, the synchronization will complete after it has transferred the changed parts of the disks.

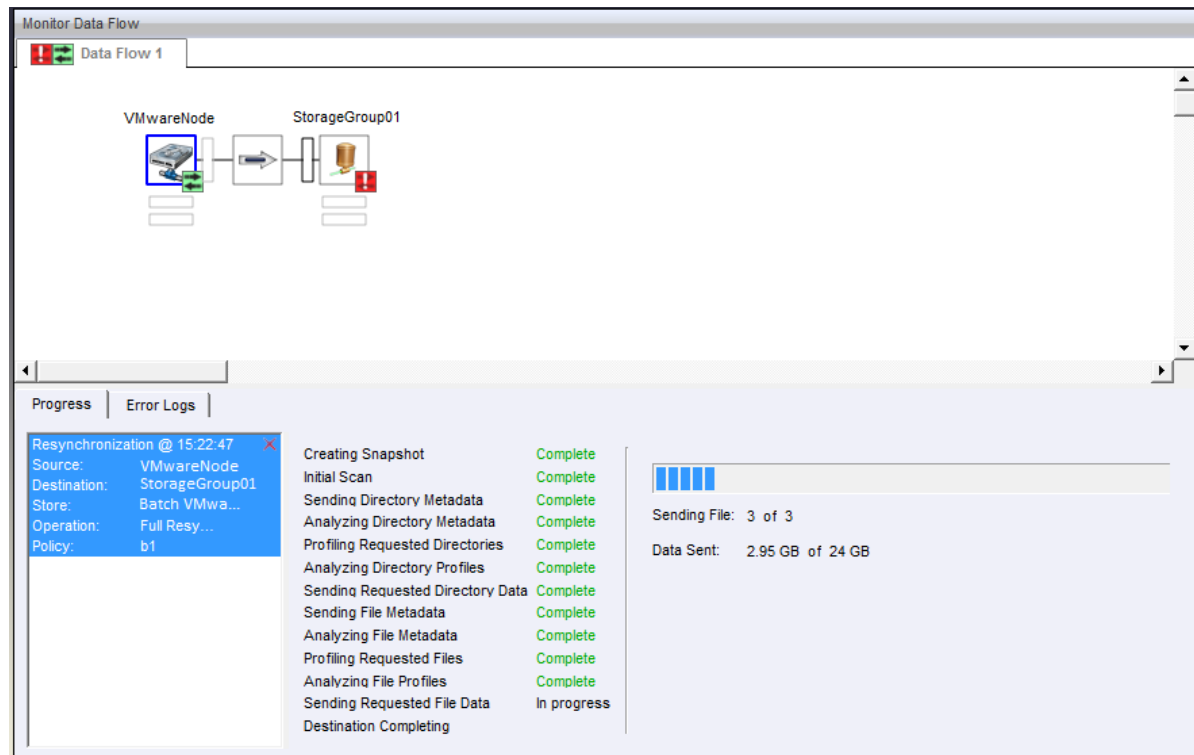


Figure 3-34 Monitoring the progress of an ESX backup

In this initial synchronization example, the ESX host appears to be sending the complete 24 GB virtual disk, however, the synchronization will complete after approximately 7 GB as that is the used space on the virtual disk.

There is a 'detail' level log that will give the correct amount transferred once the synchronization is complete, however if CBT is not working correctly it will transfer the entire disk.

Browsing and recovering V2I snapshots

This section describes how to configure Data Instance Director so that it can communicate with Hitachi V2I server to enable you to browse and recover V2I snapshots. Data Instance Director cannot currently invoke backups.

The tasks to be performed when configuring a V2I server node are:



Node creation – To inform Data Instance Director of the V2I Server details.



Policy definition – No policy is required since backups cannot be invoked.



Data Flow definition – No data flow is required since backups cannot be invoked.



Distribution and Activation – No rules are required since backups cannot be invoked.

Node creation

First, a node representing the vCenter server needs to be created.



1. Select **Node Manager** in the **Navigation** pane to access the **Node Manager** screen.
2. In the **Agentless / Hardware Node** section, select 'VMware' from the combo box.
3. Click the **Create** button. A configuration dialog is displayed (Figure 3-35).
4. Enter a **Node Name** in the text box at the top of the dialog by which to identify the node. We will use the name *V2I*.
5. A **Proxy** node that is responsible for interfacing with the server is required. This must be a Windows machine. In our case, the machine named *HyperV15* is suitable.
6. Check **Permit backups using Virtual Infrastructure Integrator (V2I)**.
7. In the **V2I Server** text box enter the name or address of an ESX host, here we will use *192.168.45.114*.
8. Enter the credentials for the vCenter/ESXi server chosen in the previous step.
9. Click **Finish** when ready.
10. The new node with the name *V2I* will now appear in the **Unauthorized Nodes** list. Select the node in this list and then click the **Authorize** button  to authorize the node.

If the server cannot be accessed or its credentials are invalid then the node will fail authorization. The configuration dialog can be reopened in order to fix any errors by right-clicking on the node and selecting **Edit node**.

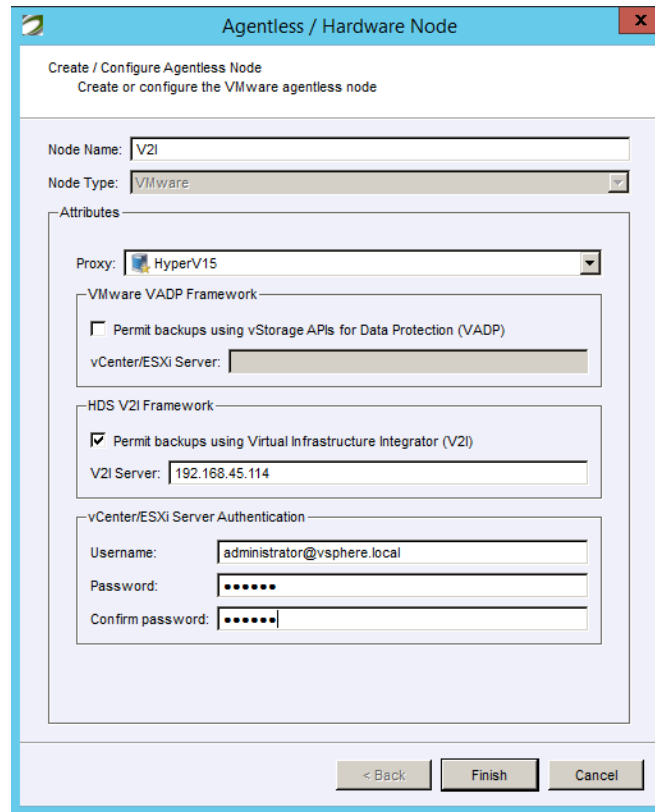


Figure 3-35 V2I Agentless Node Configuration

Browsing snapshots

V2I snapshots will now be listed on the **Restore** screen from where they can be restored or mounted. Notice that all V2I snapshots have their policy name prefixed with '--V2I External--' indicating that they were created externally and not by Data Instance Director.

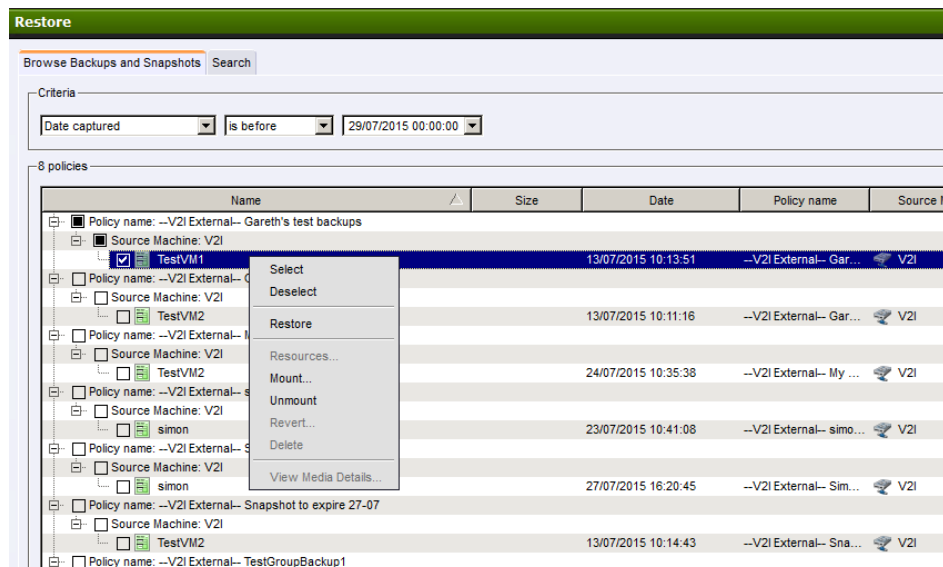


Figure 3-36 V2I Snapshots Listed on the Restore Screen

Restoring snapshots

To restore a V2I snapshot:

1. Right click on the required V2I snapshot on the **Restore** screen to display the context menu and select the **Restore** option (Figure 3-36). Note that restored VMs are not started automatically.
2. The **V2I Restore Options** dialog (Figure 3-37) is displayed allowing you to either:
 - a. **Restore virtual machine(s) overwriting the original(s)** or,
 - b. **Create virtual machine clone(s)** in which case you need to specify the **New Virtual Machine name**. If multiple VMs are selected then the name specified will be suffixed with '_1', '_2' etc.

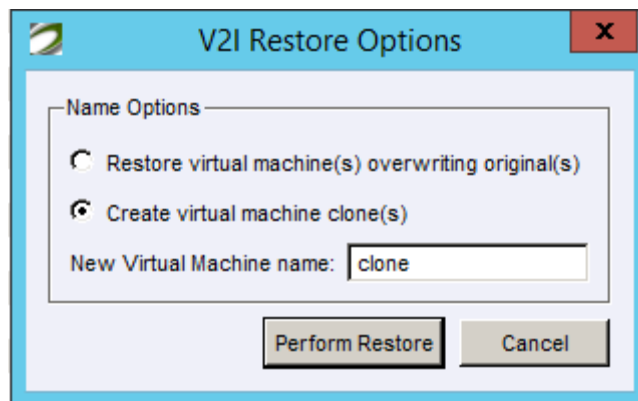


Figure 3-37 V2I Restore Options Dialog

Mounting snapshots

To mount a V2I snapshot:

1. Right click on the required V2I snapshot on the **Restore** screen to display the context menu and select the **Mount...** option (Figure 3-36).
2. The **V2I Mount Options** dialog (Figure 3-37) is displayed informing you that the snapshot will be mounted on the original virtual machine from which it was taken. Note that you may have to set these mounted volumes to 'online' in order to access them on Windows systems.

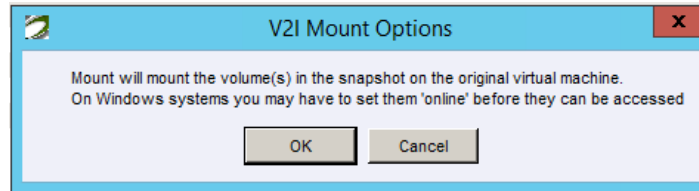


Figure 3-38 V2I Mount Options Dialog

Snapshots can be unmounted when no longer required by right clicking on the required snapshot in the **Restore** screen and selecting the **Unmount...** option.

Backing up to a tape library

This section relates specifically to backing up data to a tape library. This example will focus on backing up a Source node to a tape library, *TapeLib*.

The tasks to be performed when configuring a backup policy with a tape library destination are:

-  Node creation – To inform Data Instance Director of the Tape Library and details.
-  Policy definition – To define what data qualifies for back up to tape.
-  Data Flow definition – To define what node's data should be backed up and where to.
-  Distribution and Activation – To put the policy into effect for each node in its data flow.

Node creation

First, a node representing the Tape Library needs to be created.

1. Select **Node Manager**  in the **Navigation** pane to access the **Node Manager** screen.
2. In the **Agentless / Hardware Node** section, select 'Tape Library' from the combo box.
3. Click the **Create** button. A configuration wizard is displayed (Figure 3-34).

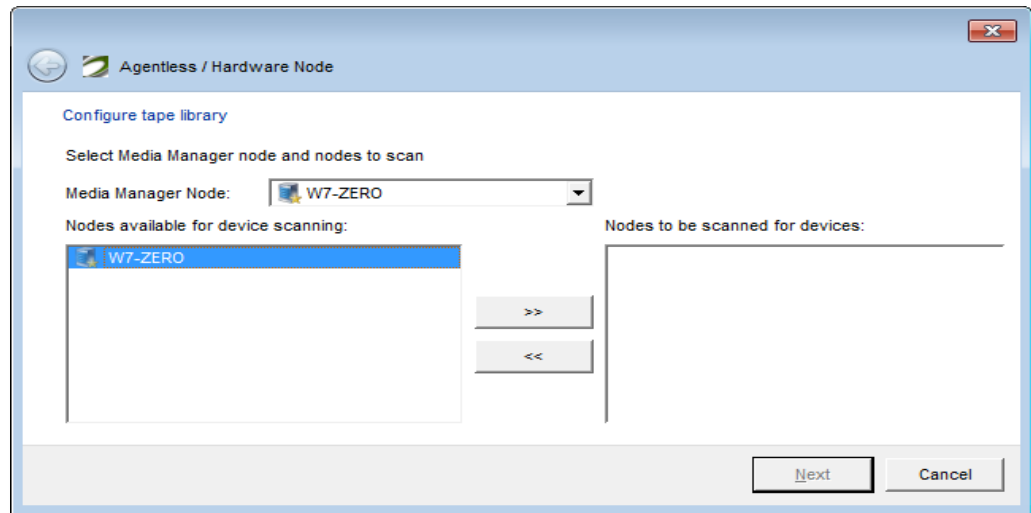


Figure 3-39 Tape Library Wizard

4. Select a node which will act as the Media Manager. A Tape Library Node can only have one Media Manager.
5. Then select which nodes are to be scanned for attached tape devices (tape libraries) then click **Next**.
6. The nodes will now be scanned and any tape device will show up in the list on the left-hand side of the scanning results dialog (Figure 3-40). When a library is selected details of it will be displayed. Select one of the listed libraries then click **Next**.

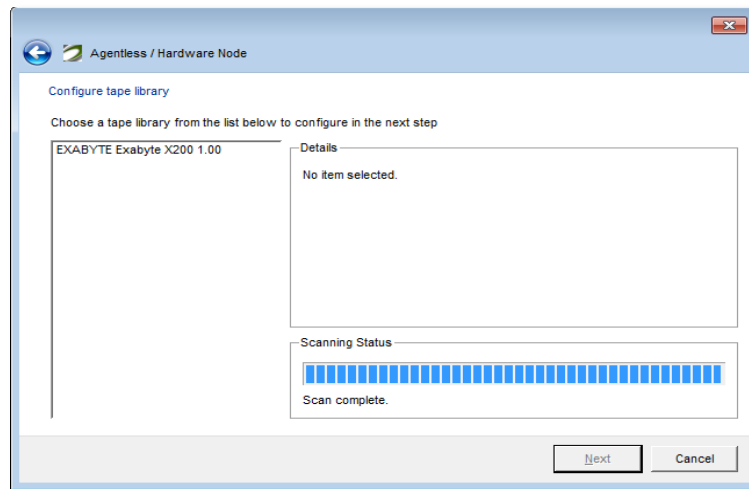


Figure 3-40 Scanning and Selecting a Library

7. Enter a name for the newly created tape library node, in this example it is named *TapeLib*. Selecting the drives or slots will present further options such as making them useable or read only (this option is only for drives). Click **Finish** when ready (Figure 3-41).

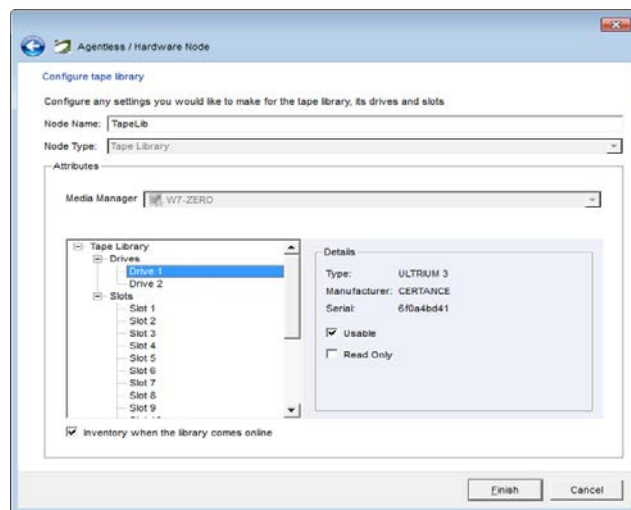


Figure 3-41 Naming and Drive Details

8. The new node with the name *TapeLib* will now appear in the **Unauthorized Nodes** list. Select the node in this list and then press the **Authorize** button  to authorize the node.

If the library cannot be accessed then the node will fail authorization. The configuration dialog can be reopened in order to change the settings for drives or slots by right-clicking on the node and selecting **Edit node**.



Note: If you intend to connect to a Sepaton Virtual Tape Library, its emulation mode must be set to LTO.

Policy definition

A backup policy for a Tape Library host is created in almost the same way as a typical backup policy. For this example, we will back up a directory on single Source node to a Tape Library

1. Select **Policy Definition**  in the **Navigation** pane to access the **Policy Definition** screen and define a policy (Figure 3-42).
2. The **Classification Palette** on the left is used to define the data captured by the policy. Expand the **Standard** heading and drag the **Path** icon from the **Classification Palette** into the **Classification** workspace in the middle of the screen.

Be sure that the **Path** icon that was just placed into the workspace is selected. List boxes to enter applicable paths are then displayed in the **Attributes** tab at the bottom of the screen.

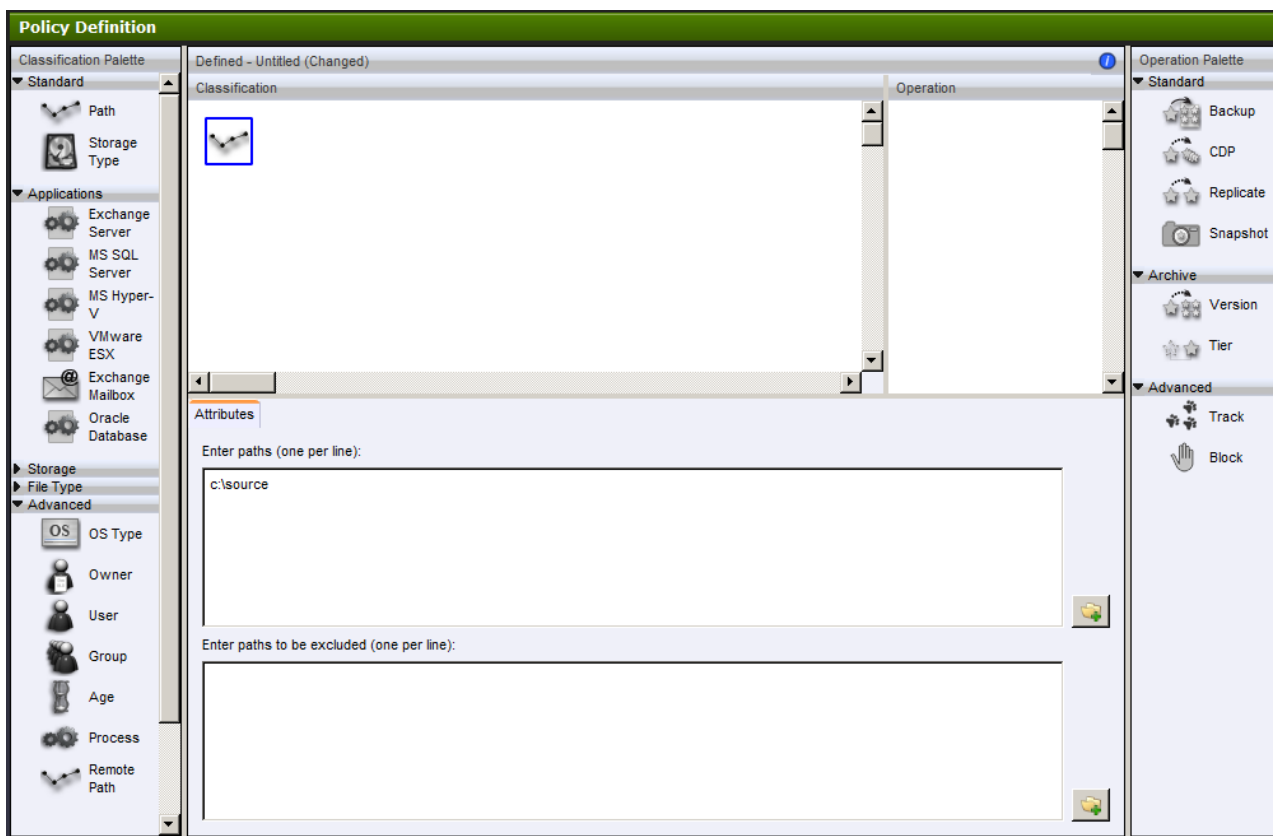


Figure 3-42 Defining a backup policy to tape

3. The **Operation Palette** on the right is used to define the data operation. In this example, the data is to be backed up. Drag the **Backup** icon into the **Operation** workspace just to the left of that palette. Only the Backup operation is applicable to Tape nodes.
4. When the **Backup** icon is selected in the workspace area, the options related to the backup operation are displayed in the **Attributes** tab (Figure 3-43). In this example, an hourly backup is required and the backup needs to be kept for two months, so set the Retention to **2 Months** and set the Recovery Point Objective to **1 Hour**. For backups to tape, Recovery Time Objective needs to be set to **Tape** or the rules will not compile. **Incremental Backup** will be checked for this example. This allows incremental backups for the data going to tape, as opposed to sending all the data for every Recovery Point Objective.

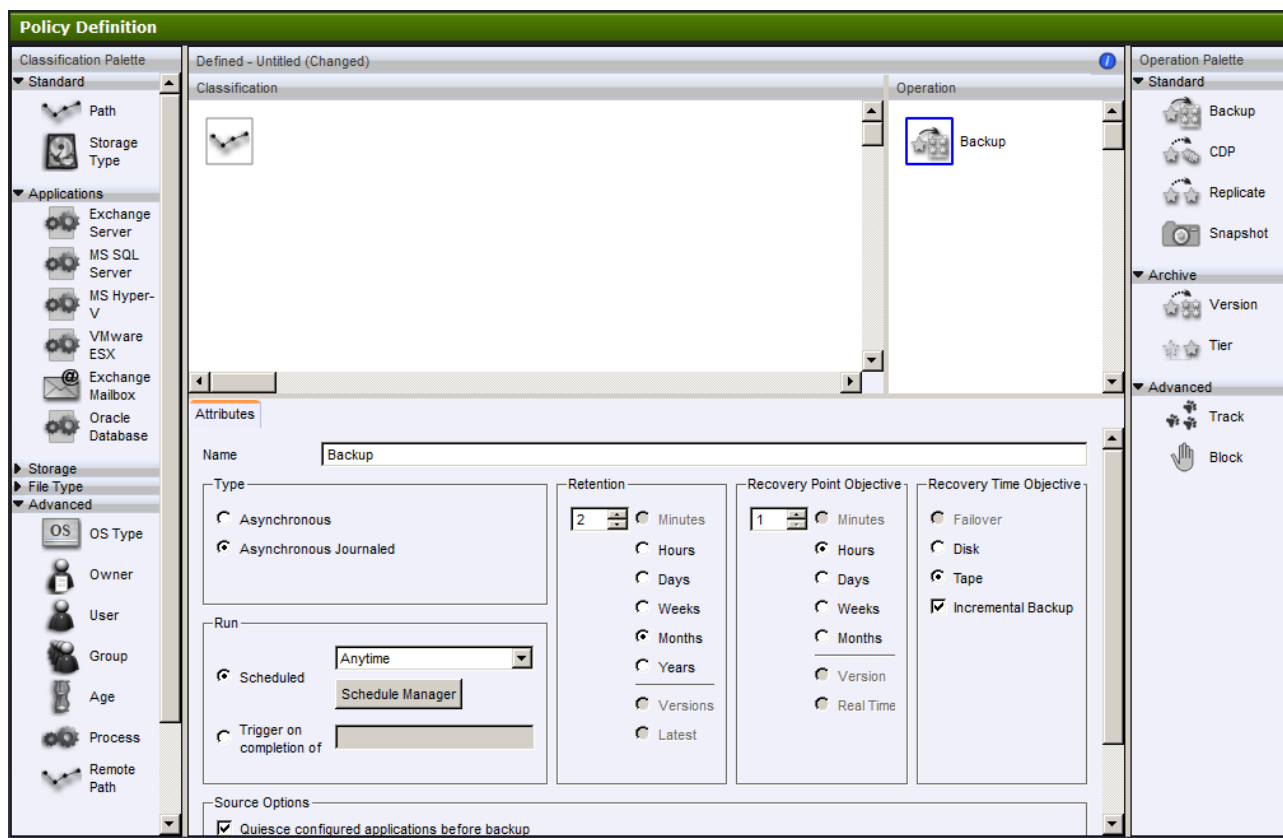


Figure 3-43 Defining a Backup Operation for Tape

5. Select **File** → **Save** from the main menu to save the policy.
6. Enter a policy name and brief description, and click **Save**.

Data flow definition

A data flow for Backup to a Tape library is similar to a backup to a repository.

In this example, the source node will be *Admin-PC* and the *Tape Library* is called *TapeLib*.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (Figure 3-44). The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's **mover** operations.

Data Flow Object Type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed.

The *Batch* mover operation must be used to back up data to a Tape Library.

2. Drag the **Admin-PC** icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Batch** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *Admin-PC*. If the **Batch** icon is dropped on top of the **Admin-PC** icon, they are automatically connected.
4. Drag the **TapeLib** icon from the **Destination** palette, and connect it by dropping it on the **Batch** mover icon.

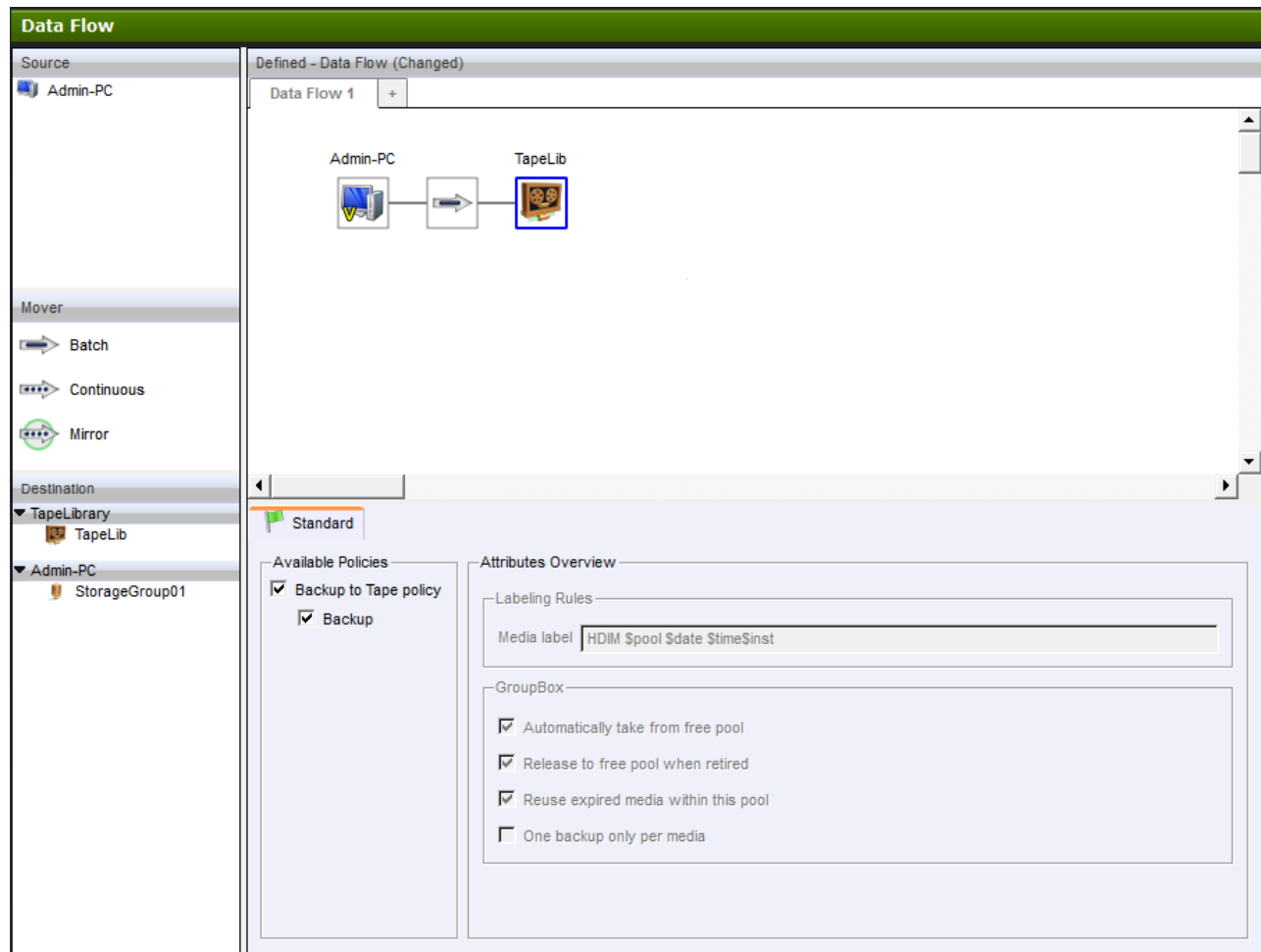


Figure 3-44 Backup Data Flow

5. In the workspace, select the **Admin-PC** icon to display a list of defined policies in the **Available Policies** pane on the **Standard** tab at the bottom of the window screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.
6. Check the Backup to Tape policy (the policy defined above) to select that policy.
7. Select the **TapeLib** icon and apply the policy to the destination node. Selecting *TapeLib* will present options in the Attributes Overview that can be edited as required by right-clicking the **Store** tab (in this example Standard) and selecting **Edit**.

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the Distribute Rules button after the policy is compiled and the rules will be distributed and the policy activated.

The system performs an initial synchronization from the Source node to the Tape Library. Information regarding this process can be seen on the **Monitor** and **Log Manager** screens.

If the **Incremental Backups** option was selected when defining the policy, then subsequent backups will only transfer data that has changed since the original backup was performed. However, each backup is essentially a full backup, so when performing a restore, only the one restore is required.

Archiving Microsoft Exchange mailboxes to HCP

This section relates specifically to archiving mailboxes from Microsoft Exchange to HCP. This example will focus on backing up Exchange mailboxes to a Windows Store Group, *StorageGroup01* on *repository1* and then archiving to HCP.



Note: The Exchange mailbox classification is only supported when you tier it to HCP.

The tasks to be performed when configuring an Exchange mailbox backup and archive policy are:



Node creation – To inform Data Instance Director of an Exchange server's Accounts and HCP credentials.



Policy definition – To define which Exchange Mailboxes should be captured from Exchange, and define the tiering to HCP.



Data Flow definition – To define where Exchange Mailboxes should be archived to.



Distribution and Activation – To put the policy into effect for each node in its data flow.

Node creation

First, a node representing Exchange needs to be created.



Note: Exchange mailbox archiving requires the default Autodiscover URLs to be resolvable from the designated HDID Exchange agentless proxy node. The following web page describes how to configure Autodiscover:

<http://www.msexchange.org/articles-tutorials/exchange-server-2010/management-administration/exchange-autodiscover.html>



1. Select **Node Manager** in the **Navigation** pane to access the **Node Manager** screen.
2. In the **Agentless / Hardware Node** section, select 'Exchange' from the combo box.

3. Click the **Create** button. A configuration dialog is displayed (Figure 3-45).

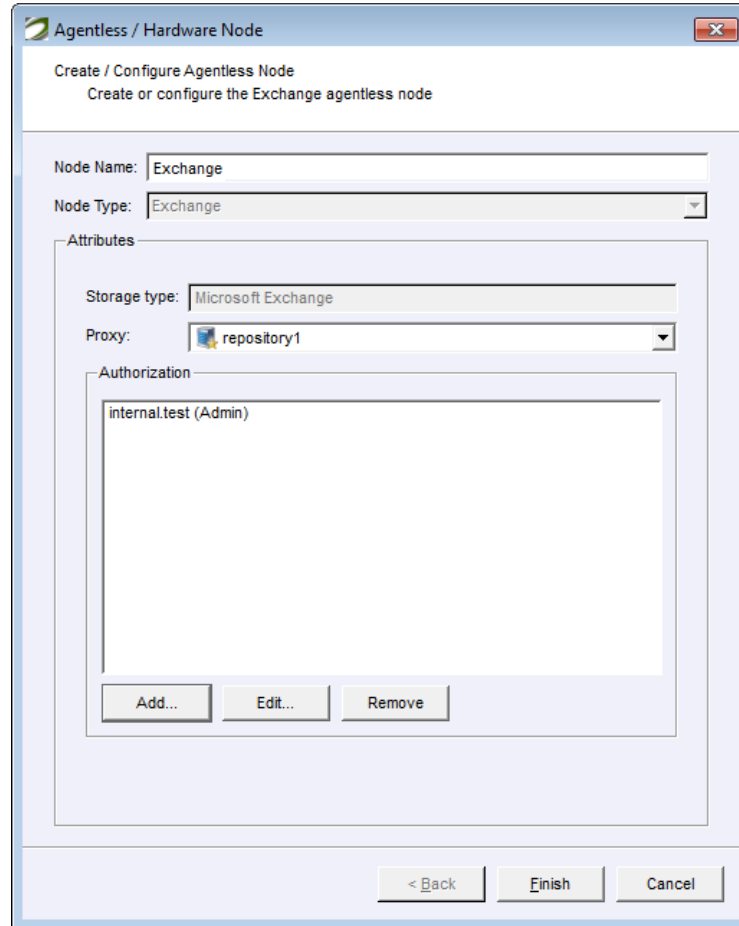


Figure 3-45 Exchange Configuration

4. Enter a **Node Name** in the text box at the top of the dialog by which to identify the node. We will use the name *Exchange*.
5. A proxy node is required, which is to be responsible for interfacing with Exchange Mailboxes. A list of potential proxy nodes is presented in the **Proxy** combo box. The Proxy node must be on the same domain as the Exchange Server. In our case, *repository1* is suitable.



Note: Proxy nodes must be Windows machines.

6. Click **Add...** to add credentials for a domain (Figure 3-46). The specified credentials must be for a domain account that can impersonate other accounts—this will be the 'backupmanager' account. This account will subscribe to the other mailboxes within the domain that are to be backed up—if needed, you can add multiple domains to an Exchange node.

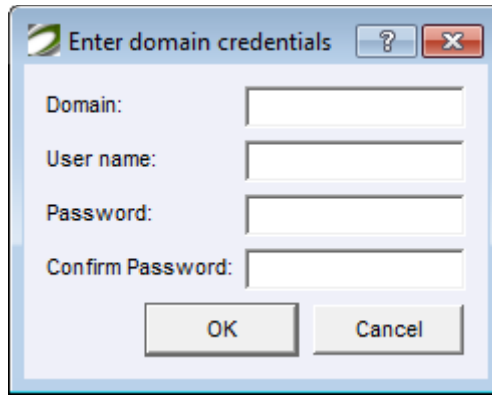


Figure 3-46 Domain Credentials Dialog



Note:

For Exchange 2010:

To configure Exchange Impersonation with organization wide scope for the '<DesiredAccount>' account, use the following command:

```
New-ManagementRoleAssignment -Name:impersonationAssignmentName  
-Role:ApplicationImpersonation -User:<DesiredAccount>
```

For Exchange 2007:

To configure Exchange Impersonation with server wide scope for the '<DesiredAccount>' account, use the following command:

```
Get-ExchangeServer | where {$_.IsClientAccessServer -eq $TRUE} | ForEach-Object  
{Add-ADPermission -Identity $_.distinguishedname -User (Get-User -Identity  
<DesiredAccount> | select-object).identity -extendedRight ms-Exch-EPI-  
Impersonation}
```

To configure Exchange Impersonation with database wide scope:

```
Get-MailboxDatabase | ForEach-Object {Add-ADPermission -Identity  
$_.DistinguishedName -User backupmanager -ExtendedRights ms-Exch-EPI-May-  
Impersonate}
```

For Exchange 2010:

The 'backupmanager' account provided for impersonation may need its maximum subscription value increased if it is to subscribe to more than 5000 mailboxes. This can be done using a custom throttling policy.

Refer to the following Microsoft article for more information about throttling policies:
[http://technet.microsoft.com/en-us/library/dd298094\(v=exchg.141\).aspx](http://technet.microsoft.com/en-us/library/dd298094(v=exchg.141).aspx)

7. Press **Finish** when ready.

8. The new node with the name Exchange will now appear in the **Unauthorized Nodes** list. Select the node in this list and then click the **Authorize** button  to authorize the node.
9. A node representing the HCP needs to be created:

1. Select **Node Manager**  in the **Navigation** pane to access the **Node Manager** screen.
2. In the **Agentless / Hardware Node** section, select **HCP** from the combo box.
3. Click the **Create** button. A configuration dialog is displayed (Figure 3-47).

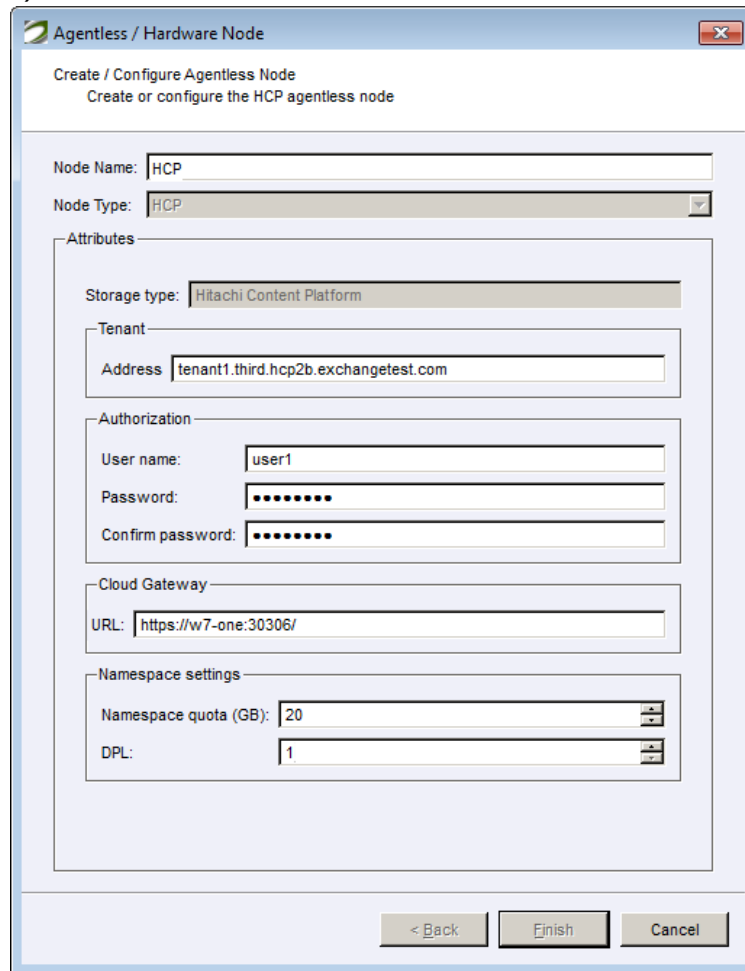


Figure 3-47 HCP Configuration

4. Enter a **Node Name** in the text box at the top of the dialog by which to identify the node. We will use the name HCP.
5. Enter a Tenant **Address**, this address must be the domain name of the tenant, a prefix of http or https is not required.



Note: The tenant must have 'Enable management through APIs' turned on via the Tenant Console on HCP.

6. Enter the credentials for the HCP chosen in the previous step. The HCP user should have Administrator and Compliance roles and have Search enabled.
7. The **Cloud Gateway** is the URL path that will be used to access the stubbed emails. The URL can use http or https, and the port number can be changed, but by default, it is required that the Data Instance Director Master node is specified in the gateway address, either by DNS name or IP address. Otherwise, viewing stubbed emails will fail.



Note: The default **Cloud Gateway** URL is specified as:

<scheme>://<host>:<port>/

Or:

<scheme>://<host>/

e.g.:

mw-win2008.test.cofio.com:9009/

Where:

<scheme> = http/https

<host> = the host name of the master or whatever resolves to the master using port forwarding

<port> = the port to use. If it is not specified, then it will use 80 for HTTP and 443 for HTTPS

You can change the default **Cloud Gateway** if the following conditions apply:

- You want users to access stubbed electronic mail outside the corporate firewall. In this case, use the DNS name to point to an external router, then forward the port specified in the URL to the master.
- The port is already in use by another process.
- You do not want to configure a certificate and do not want users to receive an error every time they attempt to view a stub (if users are accessing email within the corporate firewall). In this case, change the URL to use HTTP. This is not advised if users are accessing email outside the corporate firewall.

-
8. Change the Namespace quota and DPL (Data Protection Level) as required. In this example, the Default values are acceptable.
 9. Click **Finish** when ready.
 10. The new node with the name HCP will now appear in the **Unauthorized Nodes** list. Select the node in this list and then click the **Authorize** button  to authorize the node.

Policy definition

A backup and tier policy for Exchange Mailboxes is created in a slightly different way from a typical backup policy. For this example, we will back up the Mailboxes belonging to the distribution group '5Users'.

1. Click the **Policy Definition** icon  to navigate to the **Policy Definition** screen (Figure 3-48) and define a policy, as shown earlier in this chapter.
2. The **Classification Palette** on the left is used to define the data captured by the policy. Expand the Applications heading and drag the **Exchange Mailbox** icon from the **Classification Palette** into the **Classification** workspace in the middle of the screen.

Be sure that the **Exchange Mailbox** icon that was just placed into the workspace is selected.

3. A combination of any of the following can be used to specify which Mailboxes to backup: distribution group, security group, dynamic distribution group or individual email addresses in the format of aaa@bbb.com. In this example, we are backing up then archiving the distribution group '5Users'.

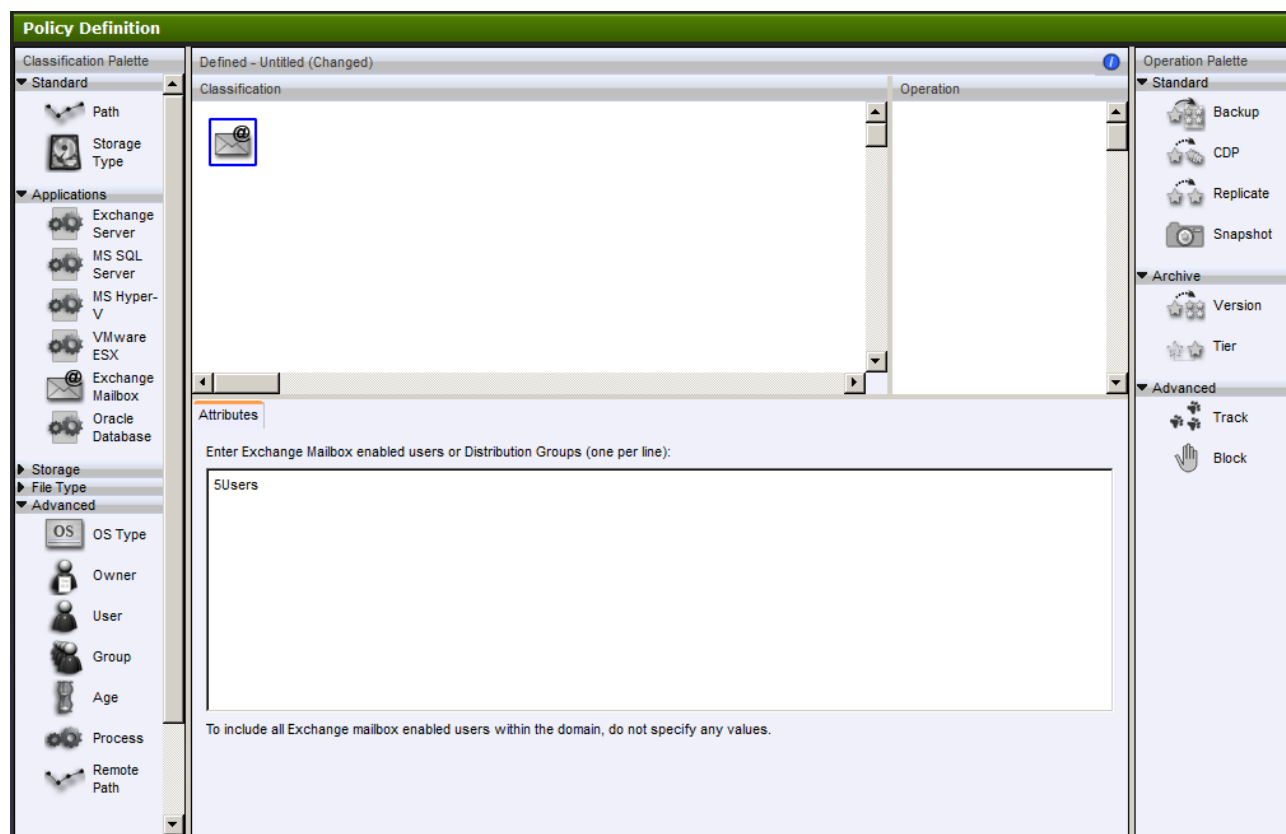


Figure 3-48 Defining an Exchange to HCP Archiving Policy

4. The **Operation Palette** on the right is used to define the data operation (Figure 3-49). In this example, the data is to be backed up and archived. Drag the **Backup** icon into the **Operation** workspace just to the left of that palette. Only Backup and Tier operations are applicable to Exchange Mailboxes.
5. When the **Backup** icon is selected in the workspace area, the options related to the backup operation are displayed in the **Attributes** tab. In this example, a daily backup is required and the backup needs to be kept for two weeks, so set the Retention to 2 weeks and set the Recovery Point Objective to **1 Day**.
6. In the **Operation Palette - Archive** heading, drag the **Tier** icon into the **Operation** workspace just to the left of that palette; placing the icon below the backup operation.

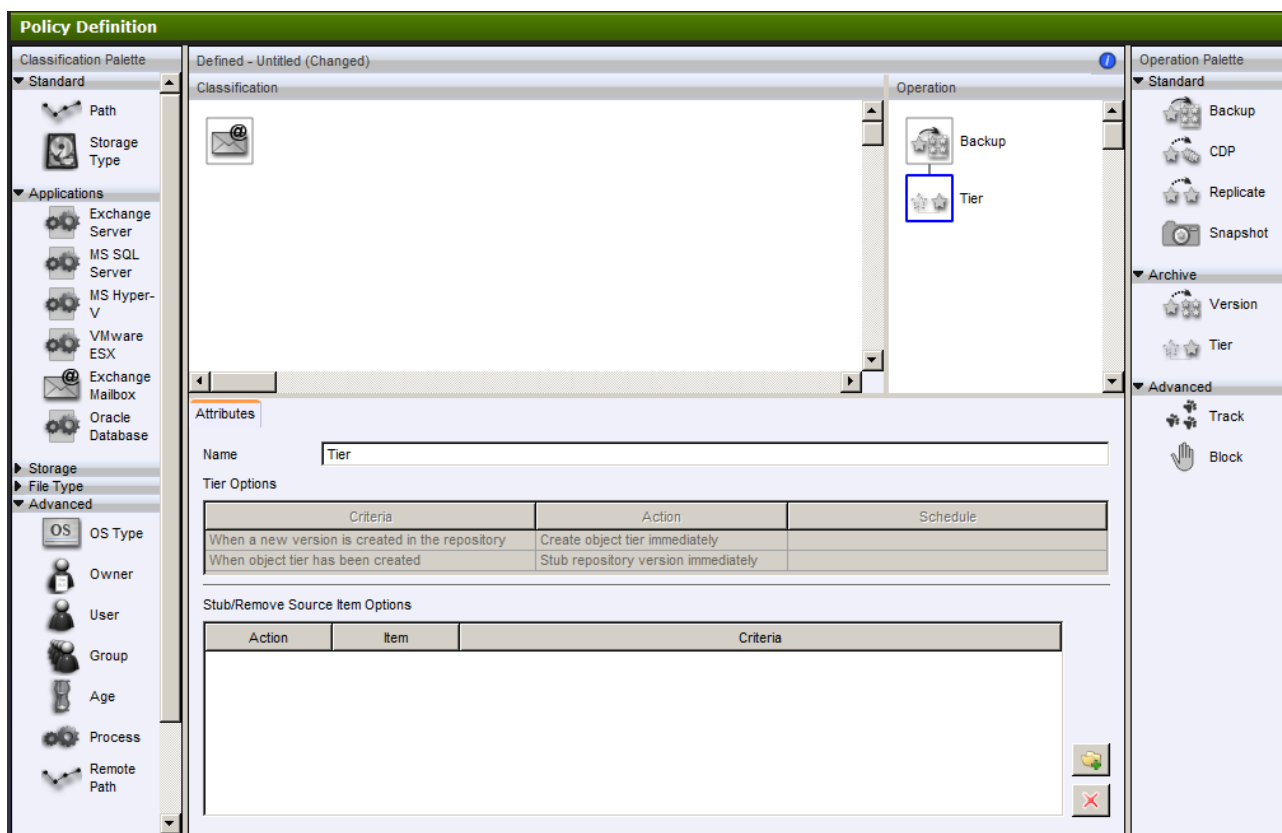


Figure 3-49 Defining a Backup and Tier Operation

7. When the **Tier** icon is selected in the workspace area, click the folder icon  to open the **Stub/Remove Source Item Options** dialog (Figure 3-50). This enables the stubbing or removing of email bodies or email attachments based on their age or size. In this example, stubbing of email bodies larger than 500KB is required; so, when an email body or file attachment exceeds 500KB in size, it will be sent to and stored on HCP.

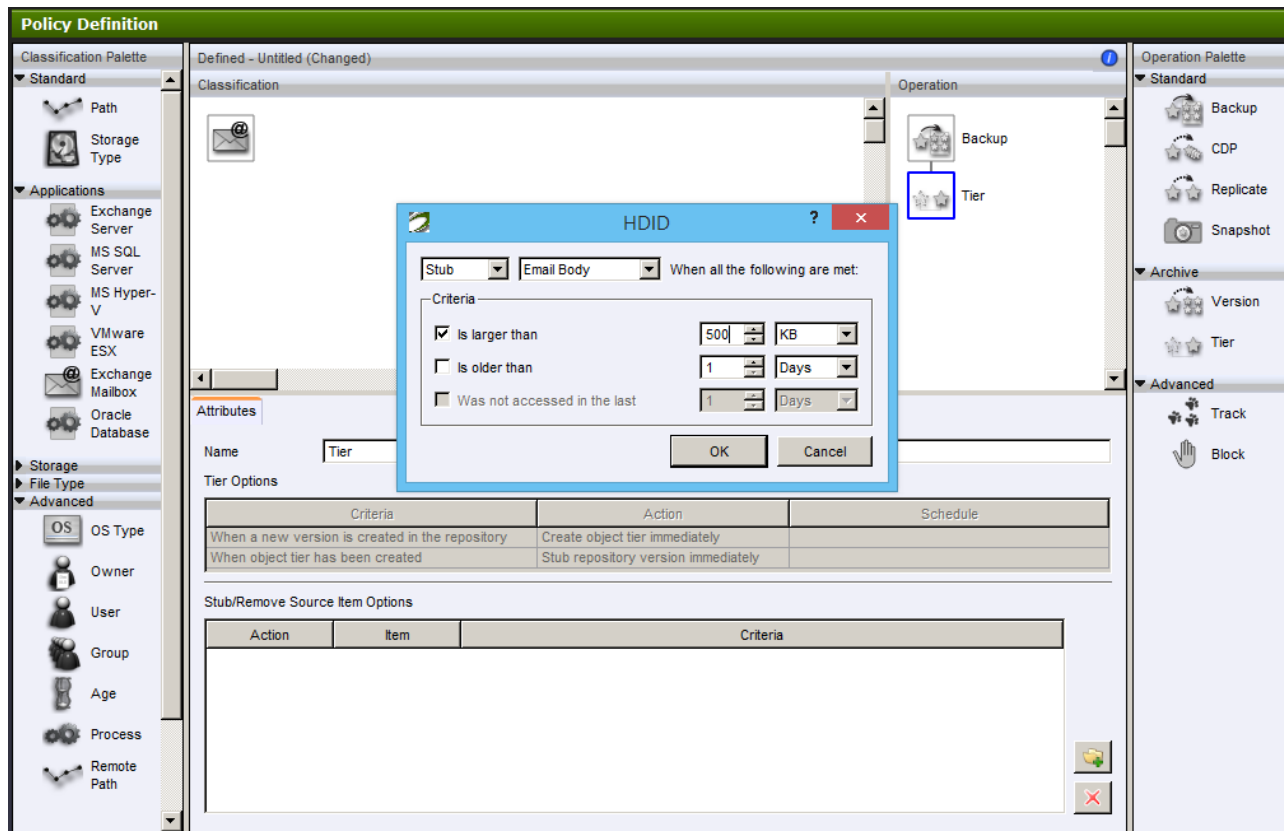


Figure 3-50 Defining Stubbing Options

8. Select **File** → **Save** from the main menu to save the policy.
9. Enter a policy name and brief description, and click **Save**.

Data flow definition

A typical data flow for backing up and archiving Exchange consists of transferring the Mailbox data to the proxy, and transferring this data to a repository. If the proxy and the repository are the same machine then the data can be transferred directly from Exchange to the repository.

In this example, the source node will be *Exchange* and the repository Store Group is called *StorageGroup01* on *repository1* (which in this case is also the proxy), and the HCP node will be *HCP*.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (Figure 3-51).

The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's **mover** operations.

Data Flow Object Type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed

The *Batch or Continuous* mover operation can be used to back up data from Exchange. For this example a continuous mover will be used. Only the batch mover can be used to connect the storage group to the HCP node.

2. Drag the **Exchange** icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Continuous** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *Exchange*. If the **Continuous** icon is dropped on top of the **Exchange** icon, they are automatically connected.
4. Drag the **StorageGroup01** icon from within *repository1* in the **Destination** palette and connect it by dropping it on to the **Continuous** mover icon.
5. Drag the **Batch** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *repository1*.
6. Drag the **HCP** icon in the **Destination** palette and connect it by dropping it on to the **Batch** mover icon.
7. In the workspace, select the **Exchange** icon to display a list of defined policies in the **Policies to assign locally** tab at the bottom of the screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.
8. Check the 5 User Mailboxes (the policy defined just now) to select that policy.
9. Select the **StorageGroup01** icon and apply the policy to the destination node.

10. Select the **HCP** icon and apply the policy to the destination node.

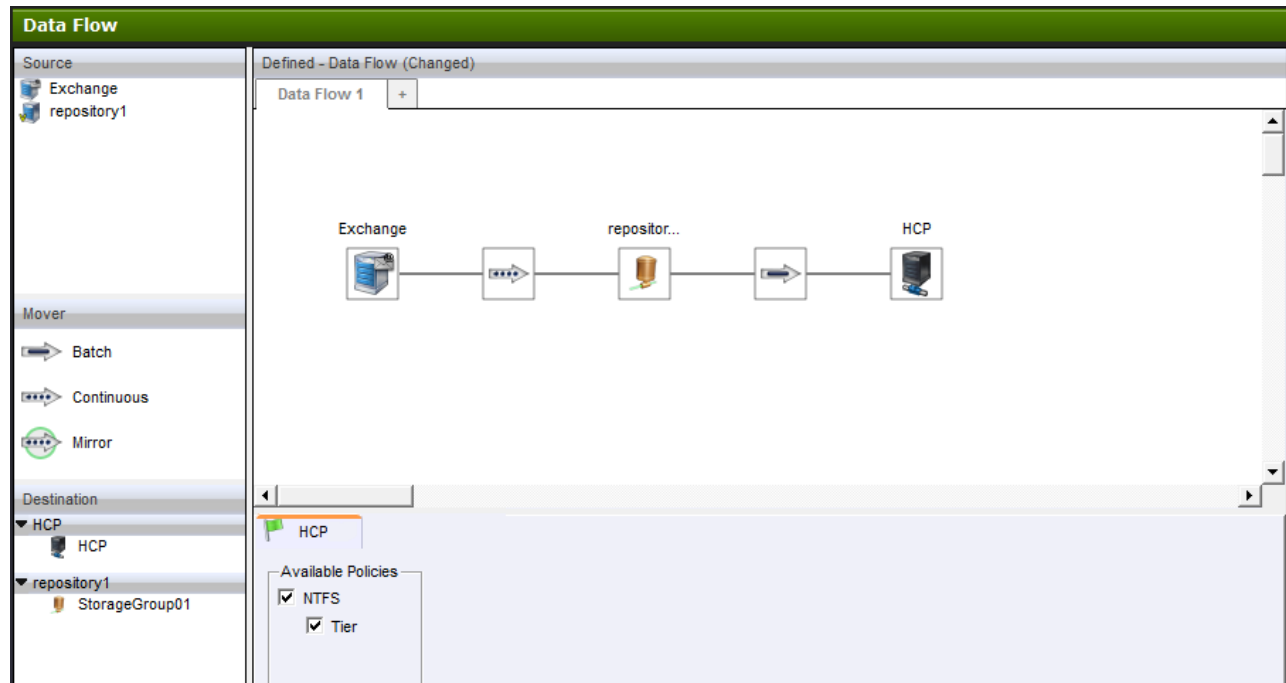


Figure 3-51 Exchange and HCP Data Flow

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the Distribute Rules button after the policy is compiled and the rules will be distributed and the policy activated.

The system performs an initial synchronization from Exchange to the repository. Information regarding this process can be seen on the **Monitor** and **Log Manager** screens.

Once the initial synchronization is complete the backup can be seen by performing a search in the **Restore** screen.

Viewing stubbed information in Microsoft Outlook

You can view stubbed email bodies and email attachments directly in Microsoft Outlook.



Note: No stubbed email bodies or attachments can be viewed unless the Data Instance Director master is in the same domain as Microsoft Exchange.

Webmail and other email clients will continue to display a link. When you view a stubbed email in a browser, there is a download button. This will download the email as a .eml file, which can be opened and forwarded.

In order to view and unstub the information, a Data Instance Director Outlook form must be configured. For details, refer to the Chapter 1 section, [Optional Post-installation Configuration](#).

You can optionally specify the files and directories to exclude from being stubbed and removed, and you can block specific programs from reading and unstubbing stubbed files. For information, refer to Section 6, [Creating a blacklist for stubbed files](#).

Restoring an Exchange server and database

When Data Instance Director restores Exchange database files, the EDB and log files are in a 'dirty shutdown' state. Sometimes the database files may need to be recovered and repaired before they can be used by the Exchange server. The PowerShell script RestoreExch.ps1 located in the Hitachi\HDID\bin folder is provided to automate this task. Before starting, ensure that:

- The Exchange server can be discovered from the HDID Master.
- The HDID Master can transfer the EDB database and log files to the Exchange server.
- The Exchange server is up and running.
- Exchange Management Shell can be launched successfully on the Exchange server
- PowerShell remoting is enabled on the Exchange server (using the Enable-PSRemoting cmdlet)

Restoring must be done on active databases since restoring of passive databases is not supported by Exchange.

Creation of a recovery storage group/database

In this scenario we want to recover one or several mailboxes, or even specific mailbox items from a point in time without having to overwrite the entire database. In Exchange Server 2010 and 2013 we can do this with a recovery database. A recovery database allows a copy of a database to be mounted on an Exchange server without having any impact on the live copy of that same database. In fact it can even be used to mount copies of databases from other servers within the same Exchange organization. The recovery procedure is as follows:

1. Restore the exchange database to a different location from original database. In this example we'll use C:\temp\ as the restore location.
2. Run the RestoreExch.ps1 script within PowerShell passing the following parameters to the script:

- DbName the database name
- Prefix the first 3 characters log file base name
- DestDir the location where the restored database files are
- Recovery:\$true the recovery flag
- Mount:\$true to mount the recovery database:

For example:

```
PS C:\temp>.\RestoreExch.ps1 -Server exch1-23 -DbName exch1 -
Prefix E01 -DestDir c:\Temp\exch1 -Recovery:$true -Mount:$true
```

3. The script should output the following:

```
___ Create the Recovery Database(exch1Recovery) for exch1 at: c:\temp\exch1. ___

Step 1) Recovering the edb file .....Completed (25% Done).
Step 2) Creating the Recovery Database .....Completed (50% Done).
Step 3) Restarting Microsoft Exchange service .....Completed (75% Done).
Step 4) Mounting the Database .....Completed (100% Done).
```

4. Verify the recovery database status by running the following command:

```
PS C:\temp>Get-Mailboxdatabase
```

The following output should be displayed:

Name	Server	Recovery	ReplicationType
Mailbox Database 1454312608	EXCH1-23	False	None
Mailbox Database 0036628388	EXCH1-24	False	None
exch1Recovery	EXCH1-23	True	None

Creation of a backup database

In this scenario we want to restore an Exchange server Database to a server. We can create a new database using the backup of the original database. The name of the new database can be the same or different from the original. The restore procedure is as follows:

1. Restore the exchange database to a different server at some location. In this example we'll use C:\temp\.
2. Run the RestoreExch.ps1 script within PowerShell passing the following parameters to the script:

- DbName the database name
- Prefix the first 3 characters log file base name
- DestDir the location where the restored database files are
- Backup:\$true the backup flag
- Mount:\$true to mount the backup database:

For example:

```
PS C:\temp> .\RestoreExch.ps1 -Server exch1-23 -DbName exch1 -  
Prefix E01 -DestDir c:\Temp\exch1 -Backup:$true -Mount:$true
```

3. The script should output the following:

```
___ Create the Backup Database(exch1Backup) for exch1 at: c:\temp\exch1. ___
```

```
Step 1) Recovering the edb file .....Completed (25% Done).
```

```
Step 2) Creating the Backup Database .....Completed (50% Done).
```

```
Step 3) Restarting Microsoft Exchange service .....Completed (75% Done).
```

```
Step 4) Mounting the Database .....Completed (100% Done).
```

4. Verify the recovery database status by running the following command:

```
PS C:\temp> Get-Mailboxdatabase
```

The following output should be displayed:

Name	Server	Recovery	ReplicationType
----	-----	-----	-----
Mailbox Database 1454312608	EXCH1-23	False	None
Mailbox Database 0036628388	EXCH1-24	False	None
exch1Backup	EXCH1-23	False	None

Restoring to the original server but different location (Standalone+DAG)

In this scenario we want to restore the Exchange database to the original server but at a different location; dismounting the original database during the operation. If the database is in a DAG, the passive database(s) will be removed first. The passive database(s) will be created and restored back to the remote Exchange server as the last step of the operation. The restore procedure is as follows:

1. Create a temp directory first, for example "C:\Temp", and restore the database files (edb and log files) to the temp directory as No. 2 of the prerequisites mentions.
2. Run the RestoreExch.ps1 script within PowerShell passing the following parameters to the script:

- Server the server name
- DbName the database name
- Prefix the first 3 characters log file base name
- TempEdbFile the path to the EDB file in temp the directory
- TempLogPath the path to the log files in the temp directory
- DestEdbFile the path to the EDB file final destination directory
- DestLogPath the path to the log files final destination directory

For example:

```
PS C:\temp>. \RestoreExch.ps1 -Server exch1-23 -DbName exch10 -
Prefix E01 -TempEdbFile c:\temp\exch10\exch10.edb -TempLogPath
c:\temp\exch10 -DestEdbFile c:\backup\exch10\exch10.edb -
DestLogPath c:\backup\exch10
```

3. The script should output the following:

```
_____ Restoring Database exch10 to the followed location _____
              (EDB): c:\backup\exch10\exch10.edb
              and (Log): c:\backup\exch10

Step 0) Preprocess Stage .....Completed.

Step 1) Recovering the edb file .....Completed (20% Done).

Step 2) Dis-mounting the mailbox database .....Completed (40% Done).

Step 3) Moving Database if the destination is different from the existing path ....Completed (60% Done).

Step 4) Copying the Database to the Destination .....Completed (80% Done).

Step 5) Mounting the Database .....Completed (90% Done).

Step 6) Postprocess Stage .....Completed (100% Done).

_____ End _____
```

4. Verify the database status on both local and remote Exchange servers by running the following commands:

[PS] C:\temp>Get-Mailboxdatabase

The following output should be displayed:

Name	Server	Recovery	ReplicationType
Mailbox Database 1454312608	EXCH1-23	False	None
Mailbox Database 0036628388	EXCH1-24	False	None
exch10	EXCH1-23	False	Remote

Then run the following command:

[PS] C:\temp>Get-Mailboxdatabase -server Exch1-24

Name	Server	Recovery	ReplicationType
Mailbox Database 0036628388	EXCH1-24	False	None
exch10	EXCH1-23	False	Remote

Archiving file system files to HCP or to Azure

This section describes the process of archiving files to HCP or to Azure. This example will focus on archiving file system files to a Windows Store Group, *StorageGroup01* on the *Master* node.

The tasks to be performed when configuring a backup workflow for either HCP or Azure are:



Policy definition – To define what data qualifies for back up.



Data Flow definition – To define what node's data should be archived.



Distribution and Activation – To put the policy into effect for each node in its data flow.

Node creation

A node representing HCP or Azure needs to be created.



1. Select **Node Manager** from the **Navigation** pane to access the **Node Manager** screen.
2. In the **Agentless / Hardware Node** section, select the applicable option from the combo box; either HCP or Azure.
3. Click the **Create** button. An **Agentless / Hardware Node** dialog is displayed (HCP Figure 3-52, Azure Figure 3-53).

Figure 3-52 HCP Configuration

Figure 3-53 Azure Configuration

4. Enter a **Node Name** in the edit box to identify the node. For this example, use the name *HCP* for the HCP node, and *Azure* for the Azure node.
5. Specify **Attributes** according to the node type as follows:

HCP Node	Azure Node
1. Enter a Tenant address. This address must be the domain name of the tenant. A prefix of http or https is not required.  Note: The tenant must have Enable management through APIs turned on using the HCP interface. 2. Enter the credentials for the HCP node. The user should have Administrator and Compliance roles and have Search enabled.	1. Enter the user Account and specify the (primary access) Key credentials for the Azure node. 2. Specify the URL for the Cloud Gateway. The Cloud Gateway is the URL path that will be used to access the stubbed files. The URL can use http or https, and the port number can be changed, but by default, it is required that the Master node is specified in the gateway address, either by DNS name or IP address. Otherwise, accessing stubbed files will fail.

HCP Node	Azure Node
3. The Cloud Gateway is the URL path that will be used to access the stubbed files. The URL can use http or https, and the port number can be changed, but by default, it is required that the Master node is specified in the gateway address, either by DNS name or IP address. Otherwise, accessing stubbed files will fail.  Note: Following is information about the default Cloud Gateway URL: scheme://host:port/ or scheme://host/ e.g. mw-win2008.test.cofio.com:9009/ Where: scheme = http/https host = the host name of the master or what resolves to the master using port forwarding port = the port to use. If it is not specified, then it will use 80 for HTTP and 443 for HTTPS	

HCP Node	Azure Node
You can change the default Cloud Gateway if the following conditions apply: • You want users to access stubbed electronic mail outside the corporate firewall. In this case, use the DNS name to point to an external router, then forward the port specified in the URL to the master. • The port is already in use by another process. • You do not want to configure a certificate and do not want users to receive an error every time they attempt to view a stub (if users are accessing email within the corporate firewall). In this case, change the URL to use HTTP. This is not advised if users are accessing email outside the corporate firewall.	
4. Change the Namespace quota and DPL as required. In this example, the default values are acceptable.	
 Note: Ensure the Namespace has a large enough quota for all current and future data, as there can only be one Namespace per Data Instance Director Store.	

6. Click **Finish** when ready.

7. The new node with the name **HCP** or **Azure** displays in the **Unauthorized Nodes** list. Select the node in this list, then click the **Authorize** button  to authorize the node.

Policy definition

A policy for archiving files to HCP is created in a slightly different way from a typical backup policy. In this example, we will archive the files in C:\Documents.

1. Click the **Policy Definition** icon  to navigate to the **Policy Definition** screen and define a policy (Figure 3-54).
2. The **Classification Palette** on the left is used to define the data captured by the policy. Expand the Standard heading, headings can be expanded by clicking on them, and drag the **Path** icon from the **Classification Palette** into the **Classification** workspace in the middle of the screen.

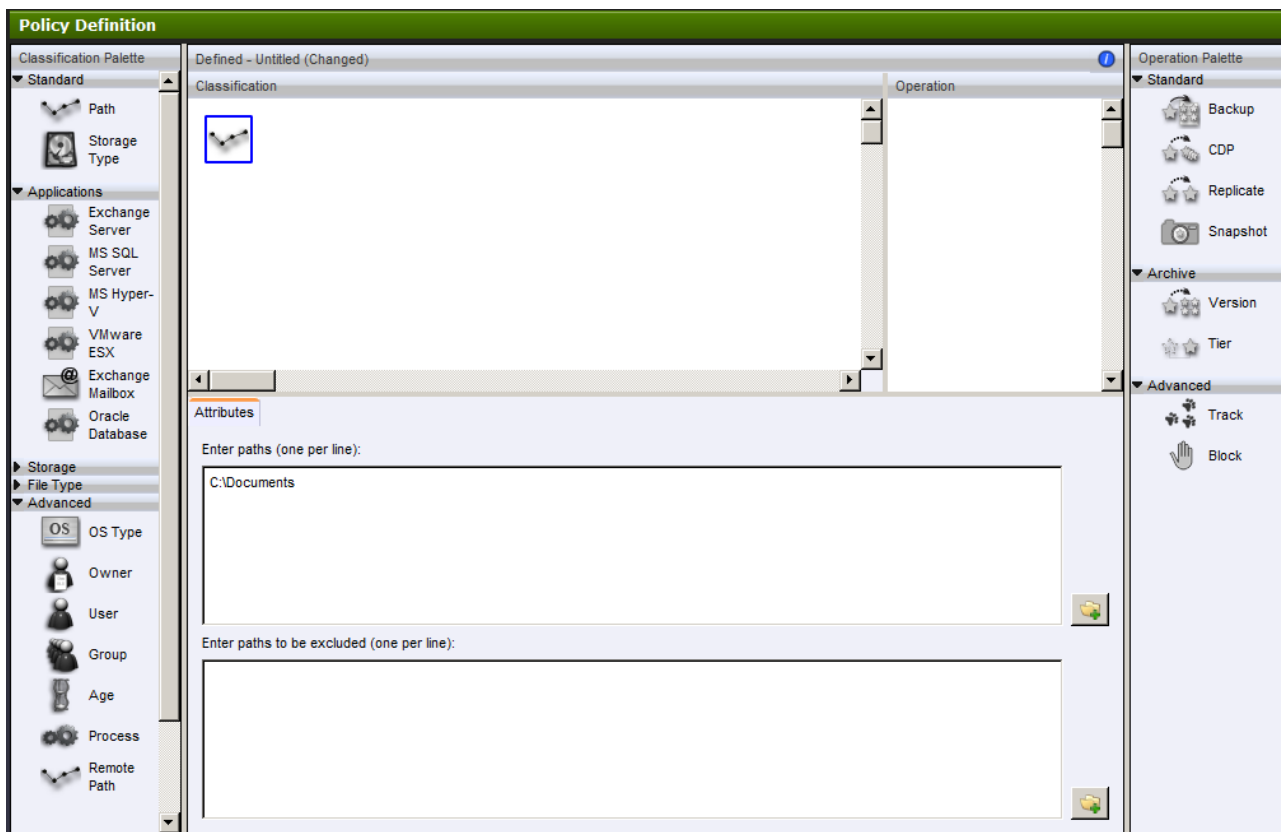


Figure 3-54 Defining a Remote Path Policy

3. The **Operation Palette** on the right is used to define the data operation. In this example, the data is to be backed up. Drag the **Backup** icon into the **Operation** workspace just to the left of that palette.
4. When the **Backup** icon is selected in the workspace area, the options related to the backup operation are displayed in the **Attributes** tab. In this example, a daily backup is required and the backup needs to be kept

for two weeks, so set the Retention to **2 weeks** and set the Recovery Point Objective to **1 Day**.

5. In the **Operation Palette** expand the **Archive** heading and drag the **Tier** icon into the **Operation** workspace just to the left of that palette, placing the icon below the backup operation (Figure 3-55).

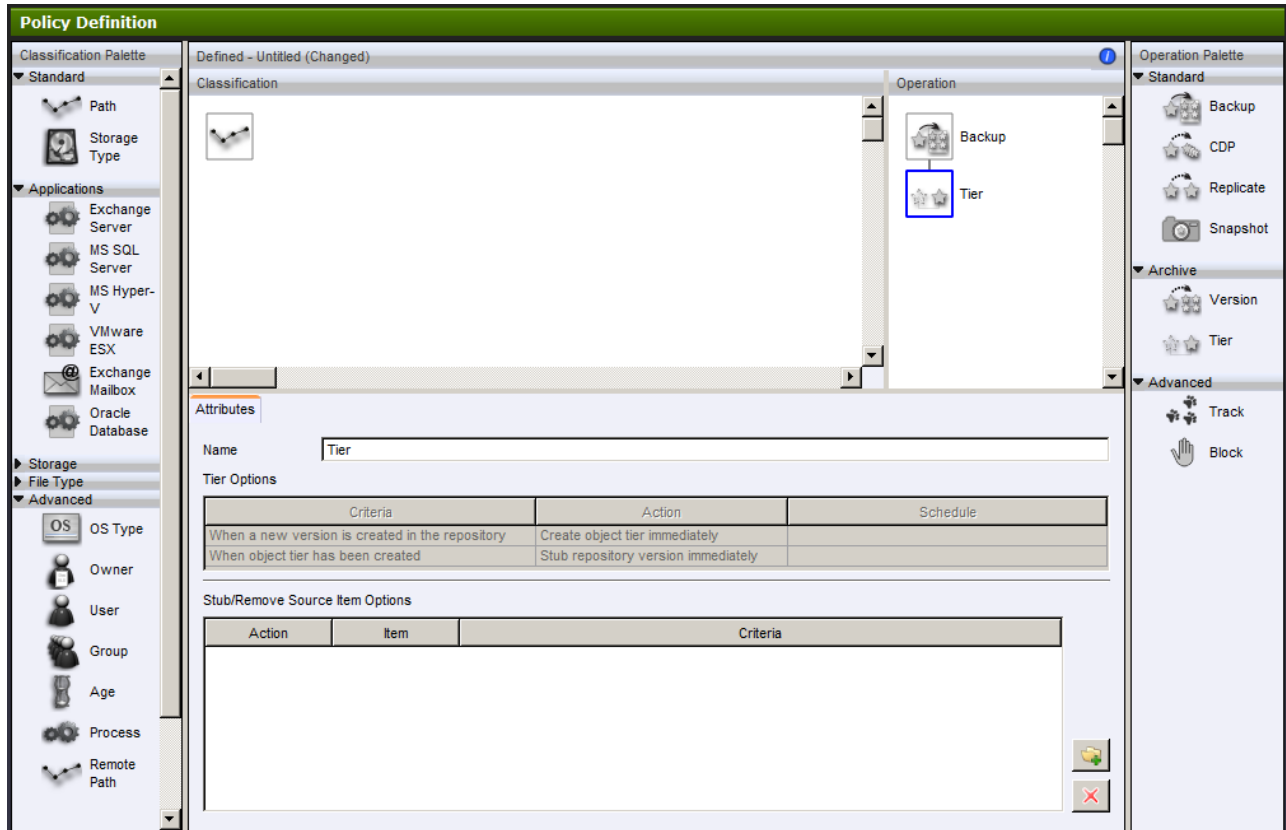


Figure 3-55 Defining a Backup Operation

6. When the **Tier** icon is selected in the workspace area, click the folder icon  to open the **Stub/Remove Source Item Options** dialog (Figure 3-56). This enables the stubbing or removing of files based on their age, size, or last accessed date. In this example, stubbing of all files larger than 500KB is required; so, when a file exceeds 500KB in size, it will be sent to and stored on HCP, and its attributes will change.

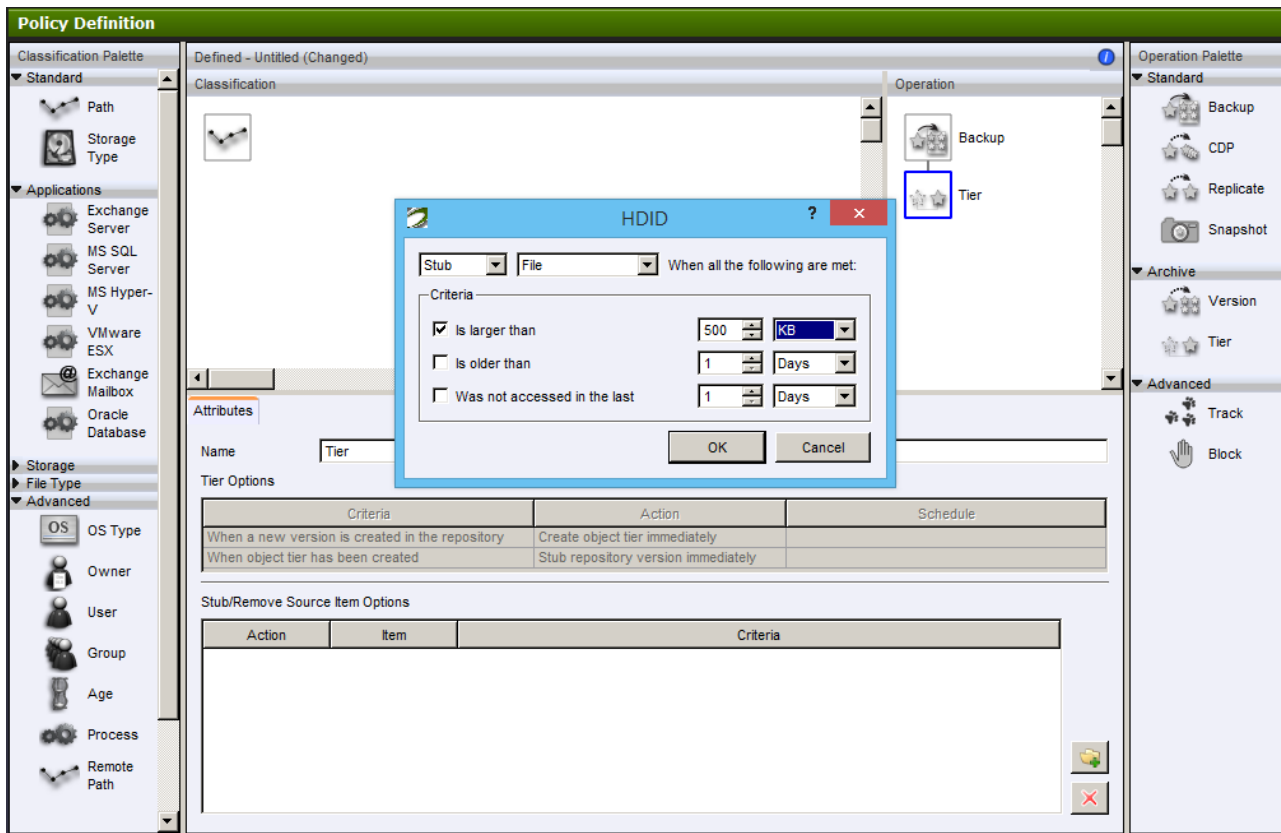


Figure 3-56 Defining Stubbing Options



Note: With applications such as Microsoft Notepad, the file's accessed time might not be modified when the file is opened. For additional information about file access time handling, see Chapter 7, [Troubleshooting HCP](#).

7. Select **File** → **Save** from the main menu to save the policy.
8. Enter a policy name and brief description, and click **Save**.

Data flow definition

The data flow defines how data moves between machines. In this example, data needs to move from *source* to *StorageGroup01* on *Master*, then to *HCP*.

In this example, the source node will be *Source* and the repository Store Group is called *StorageGroup01* on *Master* (which in this case is also the proxy), and the HCP node will be *HCP*.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow** screen (HCP Figure 3-57, Azure Figure 3-58).

The three palettes on the left-hand side contain the source and destination nodes. It also provides Data Instance Director's **mover** operations.

Data Flow Object Type	Description
Source	Nodes installed with the <i>Data Source</i> component
Destination	Nodes installed with the <i>Replication / Repository Destination</i> option
Mover	The type of data movement operation to be performed

The *Batch* or *Continuous* mover operation can be used to back up data. For this example, *Batch* will be used. Only the *Batch* mover can be used to connect the storage group to the HCP node.

2. Drag the **Source** icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Batch** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *Source*. If the **Batch** icon is dropped on top of the **Source** icon, they are automatically connected.
4. Drag the **StorageGroup01** icon from within *Master* in the **Destination** palette and connect it by dropping it on to the **Batch** mover icon.
5. Drag the **Batch** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *Master*.
6. Drag the **HCP** icon in the **Destination** palette and connect it by dropping it on to the **Batch** mover icon.
7. In the workspace, select the **Source** icon to display a list of defined policies in the **Policies to assign locally** tab at the bottom of the screen. To see more information about a policy, place the cursor over a policy. You can assign multiple policies to a source node.
8. Check the File Stub (the policy defined just now) to select that policy.
9. Select the **StorageGroup01** icon and apply the policy to the destination node.
10. Select the **HCP** icon and apply the policy to the destination node.

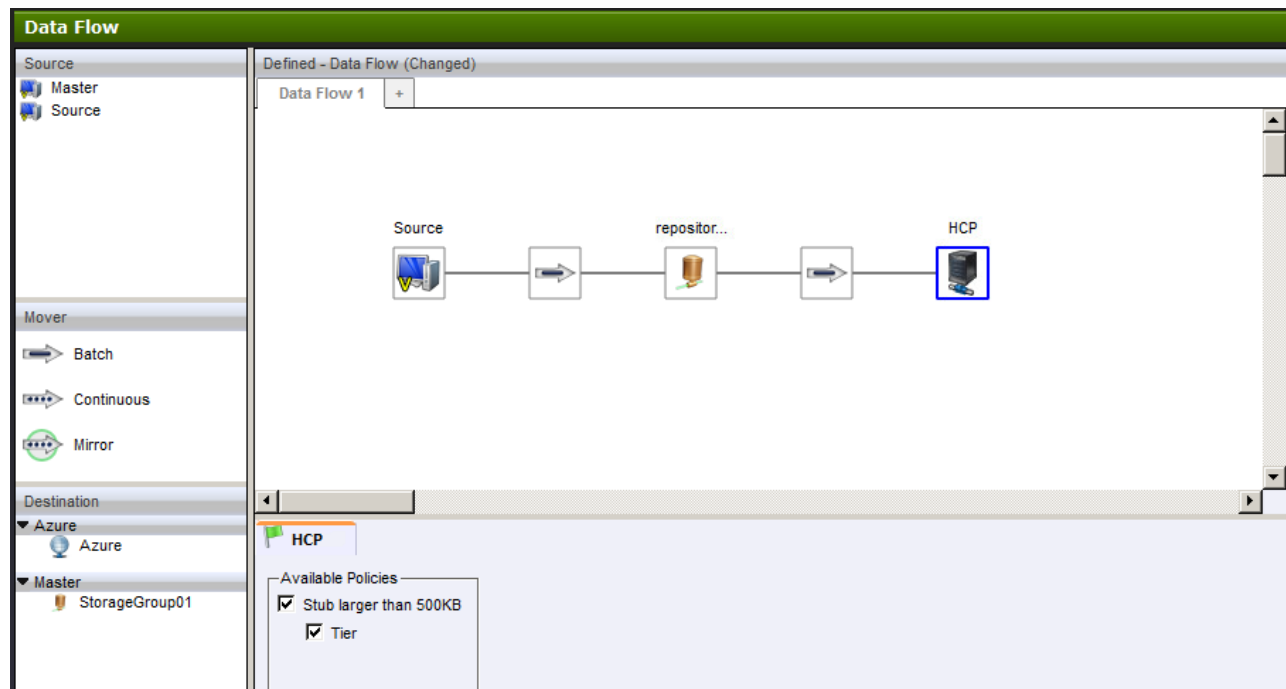


Figure 3-57 Exchange and HCP Data Flow

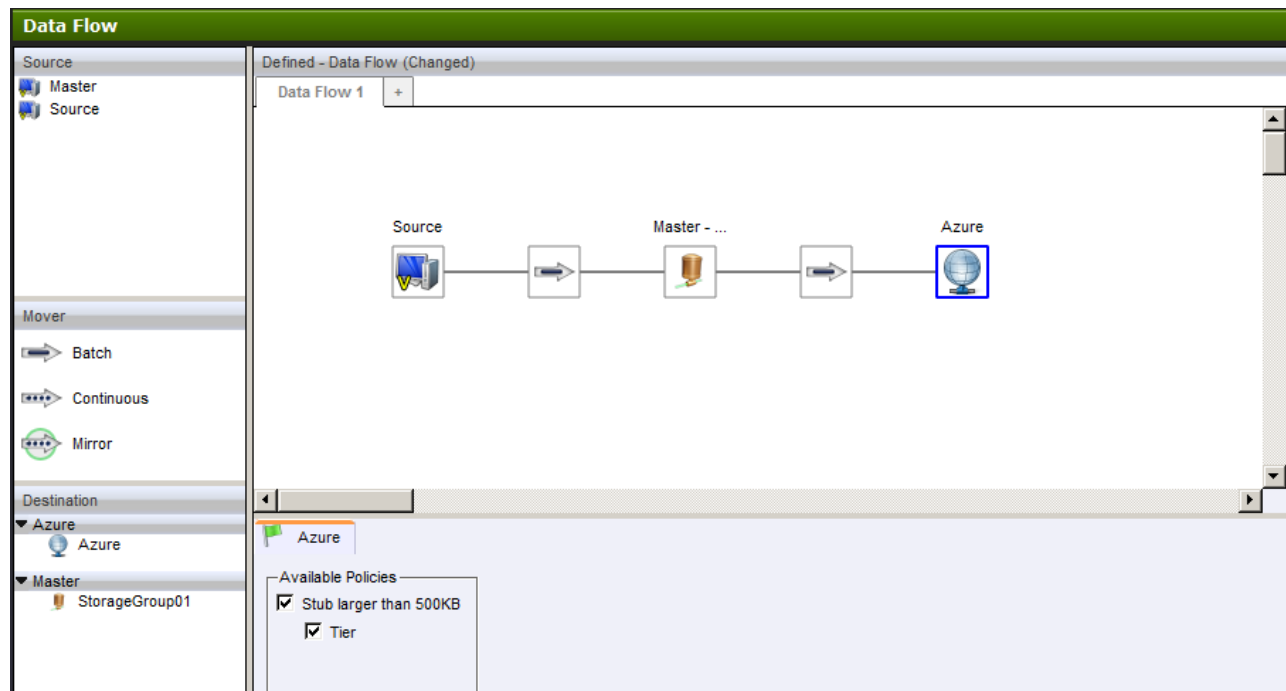


Figure 3-58 Azure Data Flow

Distribution and activation

1. To save and compile in one operation, click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog displays.
2. Click the **Distribute Rules** button after the policy is compiled to distribute rules and activate the policy.

The system performs an initial synchronization from the client node to the repository. Information regarding this process can be seen on the **Monitor** and **Log Manager** screens.

Once the initial synchronization is complete, the backup can be seen by performing a search in the **Restore** screen.

After the files have been stubbed, their associated icon and attributes will display differently within the Windows Explorer view, as shown in the following example.

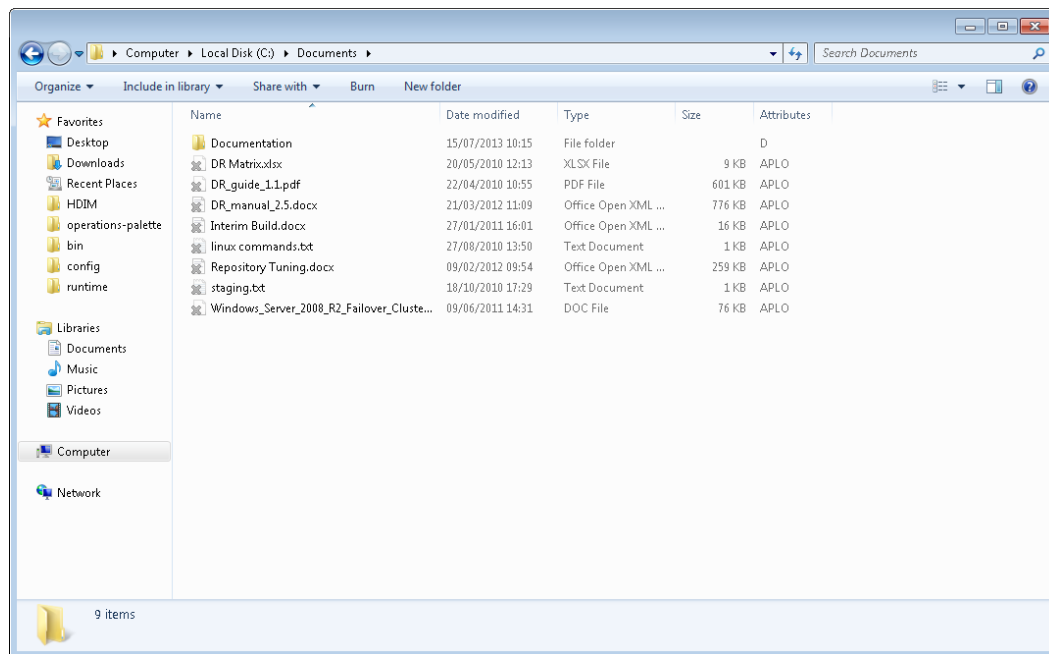


Figure 3-59 Stubbed Files

Stubbed files are indicated by a unique icon and the file information is dimmed, as shown in Figure 3-59. Stubbed files may have some or all of the following file attributes:

P (Sparse). Stubs are made sparse in order for the original file size to be reported and reserved by the local filesystem.

L (Reparse Point). Indicates that the file contains metadata that must be parsed by the Data Instance Director stubbing filter. If the stubbing filter is not present then file access will fail.

O (Offline). The file does not exist on the local filesystem and may take time to restore should a program try to access it. The accessing program may appear to hang until the file is restored. Programs that scan the local filesystem (e.g. antivirus software) must respect this attribute otherwise they will cause the file to be unstubbed. Refer to Chapter 6, [Creating a blacklist for stubbed files](#) (Stubbing Access Blacklist) to prevent unstubbing by such software.

You can un-stub a file by opening it (double-clicking the file). This restores the file from the HCP to the machine's disk drive.



Note: If a file is stubbed by Data Instance Director and then subsequently un-stubbed by a user before the next batch backup, then it will not be scheduled for stubbing again until the next time it is modified. The stubbing process is scheduled to run every 24 hours (configurable), users will therefore not see stubs appearing immediately. Any file that is un-stubbed by the user will be re-stubbed 2 weeks (configurable) later, subject to the stubbing policy's size rule or at a time determined by the stubbing policy's age rules currently in force.

Snapshotting a Hitachi block storage volume

This section relates specifically to using the Orchestration Framework for snapshotting a disk volume residing on a Hitachi Virtual Storage Platform (in this example the E: drive is specified using the Path classification, although the Hitachi Block, Exchange Server and MS SQL Server classifications could also be used). Note that a Data Instance Director repository is not involved in these data flows. Please refer to [Hitachi Block snapshotting and replication requirements](#) before starting.

The tasks to be performed when configuring a VSP Snapshot policy are:



Node creation – To inform Data Instance Director of the VSP credentials and details.



Policy definition – To define what volume to snapshot.



Data Flow definition – To define the machine on which the volume being snapshot resides.



Distribution and Activation – To put the policy into effect for each node in its data flow.



Restore – To mount a volume snapshot at a later date.



NOTE – Although this tutorial refers to a VSP volume, these steps are also applicable to a VSP-G or HUS VM volume. VSP, VSP-G and HUS VM are collectively referred to as Hitachi Block Storage on the GUI.



NOTE – A proxy node is required that is responsible for interfacing with the VSP. This must be a Windows machine with a repository installed and must have a connection to the VSP. The proxy

node must have CCI (RAID Manager) installed. Refer to the Hitachi Command Control Interface documentation available at <https://portal.hds.com>.



NOTE – If HDID is subsequently uninstalled, the existing snapshots, replication pairs and mounted volumes are left in place; they are not torn down.

Node creation

First, a node representing the VSP needs to be created, unless one already exists.



1. Select **Node Manager** in the navigation pane to access the **Node Management** screen.
2. In the **Agentless/Hardware Node** section, select 'Hitachi Block Storage' in the drop-down list.
3. Press the **Create** button. A configuration dialog (Figure 3-60) will appear.

Agentless / Hardware Node

Create / Configure Agentless Node
Create or configure the Hitachi Block Storage agentless node

Node Name: VSP

Node Type: Hitachi Block Storage

Attributes

Storage Proxy: HardwareDemo

Serial Number: 212418 Lookup

Authentication

Username: maintenance

Password:

Confirm password:

Settings

LDEV Range

Start: 0x0000 End: 0xFFFF

< Back Finish Cancel

Figure 3-60 VSP Agentless Node Configuration

4. Enter a **Node Name** in the text box at the top of the dialog by which to identify the node. We will use the name *VSP*.

5. Enter the **Storage Proxy** node name that is responsible for interfacing with the VSP. This must be a Windows machine with a repository, a connection to the VSP and have Hitachi CCI (RAID Manager) installed. The command device must **only** have user authentication enabled. In our case *HardwareDemo* is a suitable node.
6. Click **Lookup**, then in the **Serial Number** dropdown menu select the serial number of the VSP.
7. Enter the **Authentication** credentials for the VSP chosen in the previous steps. Note that the password must contain only those characters useable for CCI commands.
8. Specify an **LDEV Range** from which snapshots should be allocated.
9. Click **Finish** when ready.
10. The new node with the name *VSP* will now appear in the Unauthorized Nodes list. Select the node in this list and then press the **Authorize** button  to authorize the node.

If the VSP cannot be accessed or its credentials are invalid then the node will fail authorization. The configuration dialog can be reopened in order to fix any errors by right-clicking on the node and selecting **Edit node**.



NOTE – The Username specified in the VSP node's credentials must be a member of the Administration User Group on the VSP

Policy definition

A policy defines *data classification* and the *data operation*. In this example, we wish to snapshot E:



1. Select **Policy Definition**  in the Navigation pane to access the **Policy Definition** screen (Figure 3-61) and define a policy.
2. In this example, a directory needs to be defined. Expand the **Classification Palette - Standard** heading and drag the **Path** icon from the palette into the **Classification** workspace.

Be sure that the **Path** icon remains selected in the workspace. The options specific to the Path classification are displayed in the **Attributes** pane below the workspace.
3. Enter a path name in the **Enter Path** edit box. In this example type 'E:\' to define the volume to be snapshotted.
4. The **Operation Palette** on the right is used to define the data operation. In this example, the path 'E:\' is to be snapshotted so drag the **Snapshot** icon into the **Operation** workspace just to the left of the palette.

5. When the **Snapshot** icon is selected in the workspace area, the options related to the snapshot operation are displayed in the **Attributes** pane. In this example, a snapshot is required every 8 hours and needs to be kept for 2 weeks, so set the **Retention** to 2 **Weeks** and set the **Recovery Point Objective** to 8 **Hours**. Finally set **Mode** to **Hardware** and **Hardware Type** to **Hitachi Block**.



WARNING: VSP currently has a limit of 1024 snapshots per LDEV.

Ensure that the RPO and Retention periods are set such that this limit is not exceeded (i.e. $\text{Retention} / \text{RPO} \leq 1024$).

In the example shown here, the maximum number of snapshots stored will be 42, i.e.:

(2 weeks retention [x 7 days x 24 hours] / 8 hours RPO).

6. Select **File** → **Save** from the main menu to save the policy. Enter a policy name and brief description, and click **Save**.



NOTE – *The classifications that are applicable to VSP policies are Path, Exchange Server, MS SQL Server and Hitachi Block*

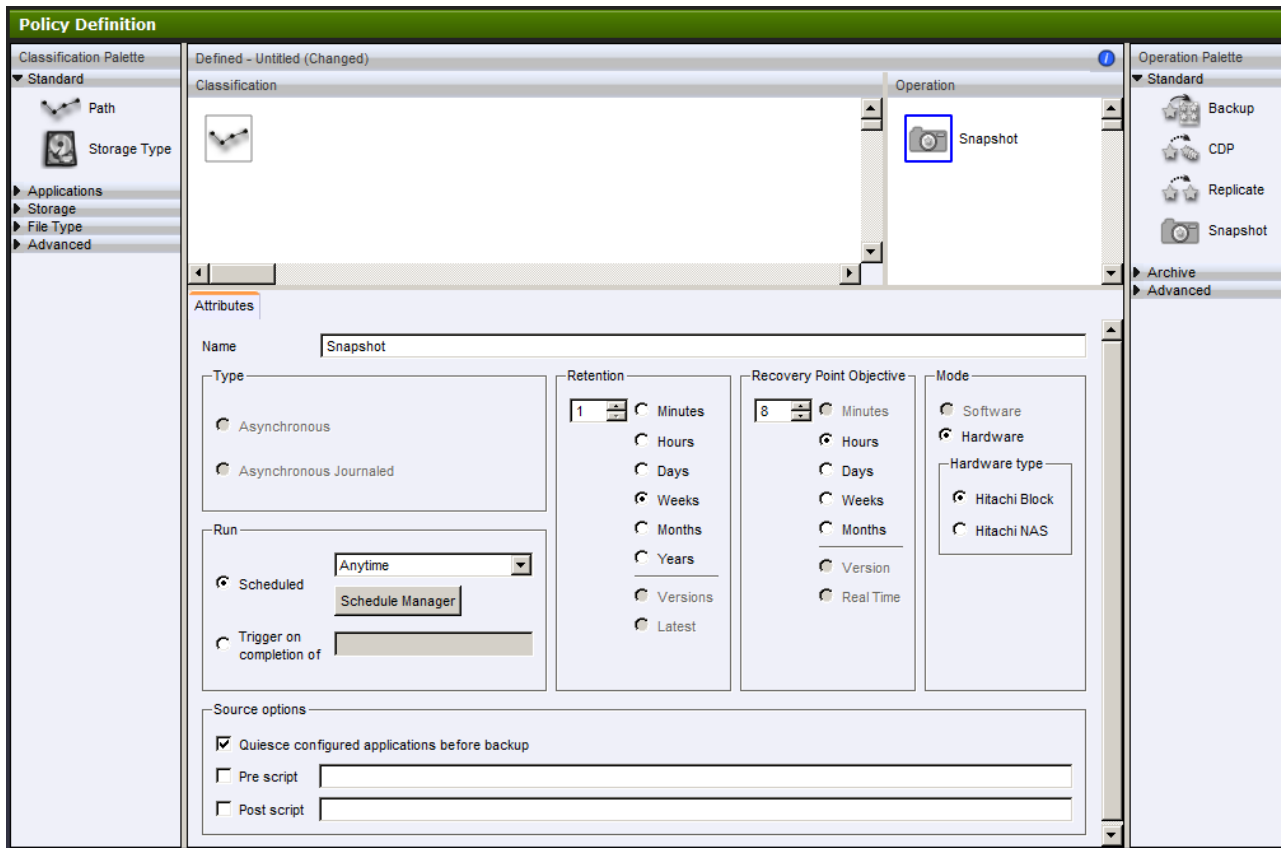


Figure 3-61 Defining a Snapshot Operation

Data flow definition

In the case of a snapshot policy, the data flow defines on which machines the policy will be implemented. The machine that this policy will apply to must be connected to the VSP and have the VSP volume (LDEV) mounted; in this example the *E:* drive.

The snapshot policy needs to be assigned to the *HardwareDemo* node connected to the VSP.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow Definition** screen (Figure 3-62).

The palette on the left-hand side contains the Source nodes, Movers and Destination nodes. Snapshotting a VSP volume only requires a source node to be specified.

2. Drag the *HardwareDemo* icon from the **Source** palette to the **Data Flow** workspace.

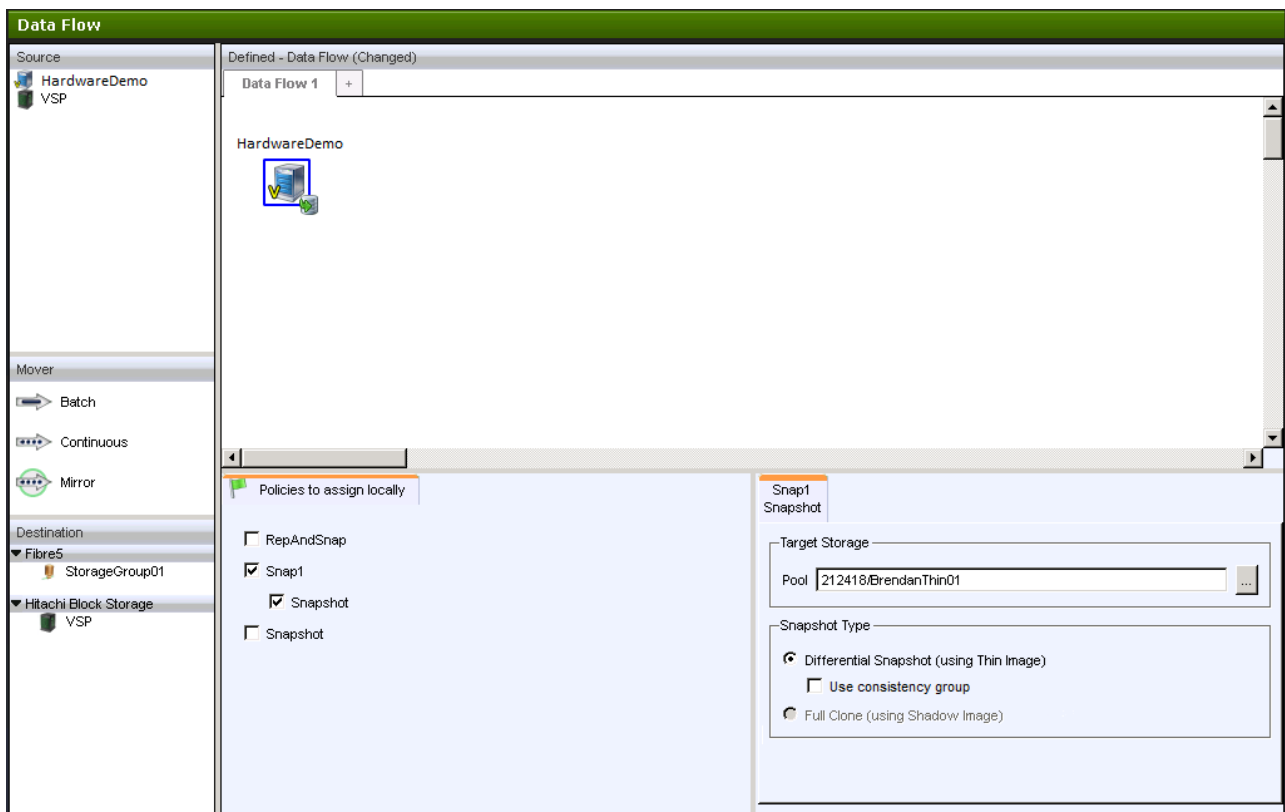


Figure 3-62 Hardware Snapshot Data Flow

3. In the workspace, select the *HardwareDemo* icon to display a list of defined policies in the **Policies to assign locally** tab below the workspace. To see more information about a policy, hover the cursor over a policy.
4. Check the *Snap1* policy (the policy defined above) to assign that policy.
5. Selecting this policy will give options in the **Snapshot** tab to the right.
6. Within **Target Storage**, select a **Pool** using the [...] button; the **Select VSP Storage Pool** dialog will appear.
7. Within the **Select VSP Storage Pool** dialog select the *VSP* node from the drop down, then select a suitable **Pool** from the list presented.
8. Select a **Snapshot Type** of **Differential Snapshot (using Thin Image)** for this example.
9. The **Use consistency group** checkbox can be selected to ensure that when a set of ThinImage pairs is split as part of a group, they will be split at the same time and are thus consistent. This is most useful when snapshotting non-application consistent volumes, e.g. a Hitachi Block classification.

✓ **NOTE** – *Thin Image pools must be created using Hitachi Storage Navigator prior to selecting the Target Storage in Data Instance Director.*

Distribution and activation

1. To save and compile in one operation, click the **Save and Compile** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

Hot Tip: The actions that are performed by the orchestration framework as it executes a backup policy are captured by the [Log Manager](#). Detailed logs and their attachments can be exported as a text file (refer to the section [Exporting logs](#)).

Hot Tip: Details of the hardware resources (LDEVs) used for a particular replication or snapshot can be found via the **Resources...** menu option on the [Managed Storage tab](#) of the Storage screen.

Restore

A snapshot of a VSP volume (LDEV) can be mounted via the **Restore** screen (Figure 3-63).

1. Enter **Criteria** into the **Browse Backup and Snapshots** tab, results of all snapshots meeting the search criteria will be displayed.
2. Expand the tree for the policy *Snap1*.
3. Select a snapshot within this policy, right-click it and select **Mount** from the context menu. The **Mount Snapshot Set** dialog is displayed.
4. From the **Mount host** drop down select the machine on which this snapshot is to be mounted and select the **Mount starting with the drive** radio button, in this example we will mount to 'J:' on '*HardwareDemo*'.
5. The host group to use to expose a snapshot to a machine can be set to **Automatically discover host group** or specified directly in the **Host Group** edit box or by selecting from those presented by pressing the [...] button to the right of the edit box.



WARNING: For AIX hosts, when performing unmount and revert operations on Hitachi File and Block storage, the user is required to first manually unmount the file systems from the host operating system.

When unmounting from an AIX host the following warning is displayed to the user:
“This file system must be unmounted manually before an 'unmount' operation is performed by HDID, failing to do so can cause a loss of data”.

When reverting on an AIX host the following warning is displayed to the user:
“This file system must be unmounted manually before a revert can take place, failing to do so may cause loss of data”.

When performing a mount operation on an AIX host:

The mount dialog options are forced to **Expose volumes to the host but do not mount them** and all other options disabled; thus the file system will not actually be mounted.

✓ ***NOTE** – Mount at original location is not currently supported if the original source volumes are present. The base OS may experience problems if there is already an instance of that file system mounted. This is because the snapshots share the same disk partition IDs. In extreme conditions it can cause corruption to both disks.*

✓ ***NOTE** – Unless the user selects a host group, the machine where the volume is to be mounted **MUST** have an existing volume on the same storage device. If there is no connection between the mount host and the block storage device then Data Instance Director will fail the mount operation timeout after 30 minutes.*

✓ ***NOTE** – When reverting a snapshot of an Exchange server, it is necessary to stop the Exchange services prior to reverting and then reboot the server after the revert operation is complete.*

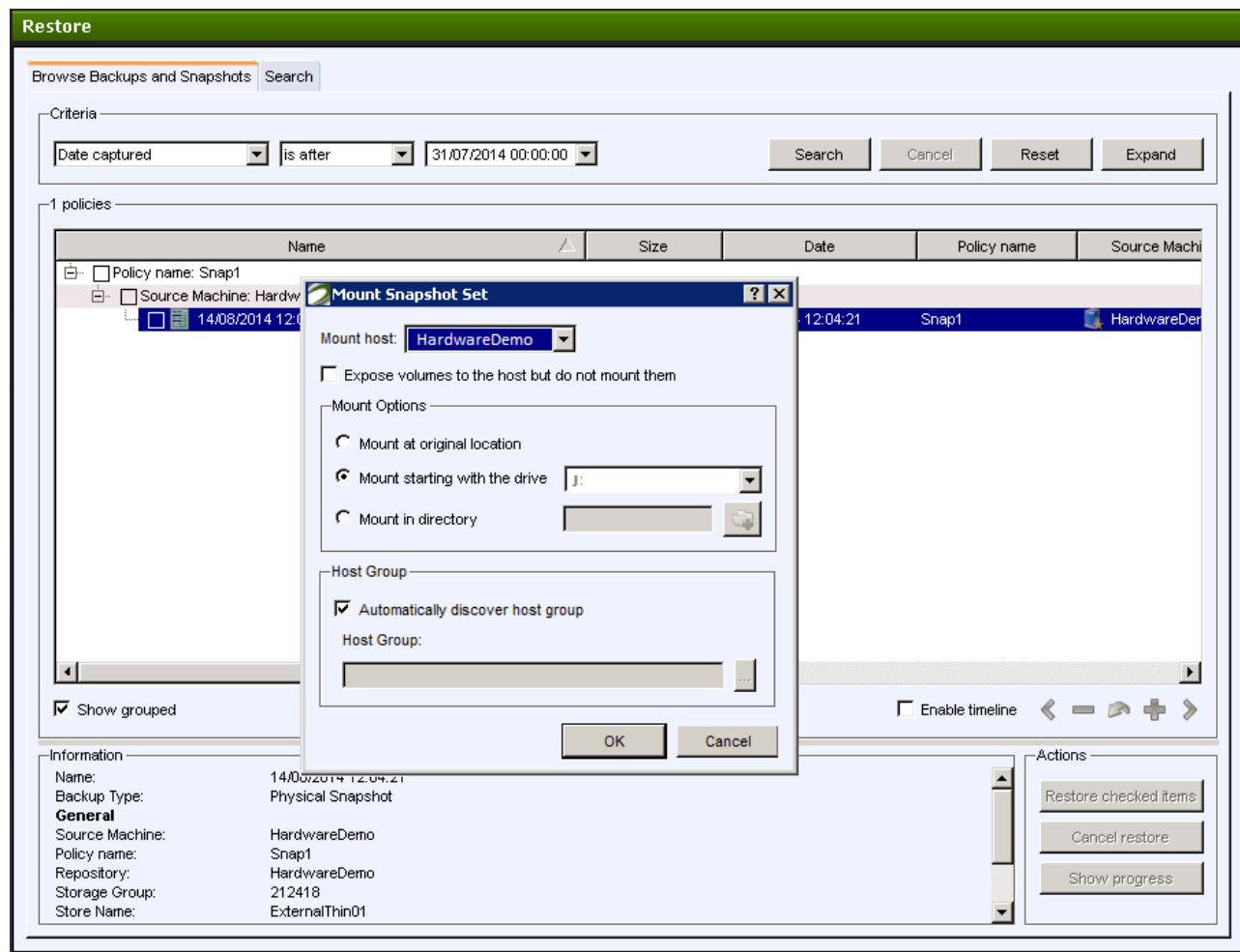


Figure 3-63 Hardware Snapshot mounting

A snapshot can be unmounted either in the **Restore** or **Storage** screens. However the **Storage** screen (Figure 3-64) shows which snapshots are mounted; either way the procedure to unmount is identical.

1. Select a snapshot and right-click it and select **Unmount**
2. The snapshot will be unmounted. It may take several minutes before the unmount is actually completed.

Storage

Repositories

Repository Usage

Managed Storage

Library Management

Media

Media Requests

Filter

Intelligent Storage Manager node

All

Select

Fibre5

Filter

Record range

All

Select

05/03/2015 00:00:00

to

20/03/2015 00:00:00

Intelligent Storage Records

Name	Information	Policy	Operation	Expiry	Application	Mounted	Av
Fibre5	ISM Node						
newVSP	Hitachi HUS-VM (212418)						
BrendanDynamic01	Pool						
Gaymer	Source Node						
19/03/2015 17:11:56	Refreshed Full Copy Set	GaymBrendanSITic...	Replicate	07/02/21 ...	Window...		
19/03/2015 13:04:57	Static Full Copy Set	GaymBrendanSibat...	Replicate	07/02/21 ...	Window...		

Figure 3-64 Snapshot shown as Mounted

Replicating a Hitachi block storage volume

This section relates specifically to using the Orchestration Framework for replicating a disk volume (in this example the E: drive); a volume residing on a Hitachi Virtual Storage Platform. A volume can either be replicated to its original VSP (Batch) or to a different VSP (Live). Please refer to [Hitachi Block snapshotting and replication requirements](#) before starting.

The tasks to be performed when configuring a VSP Snapshot policy are:



Node creation – To inform Data Instance Director of the VSP credentials and details.



Policy definition – To define what volume to replicate



Data Flow definition – To define the machine on which the volume being replicated resides



Distribution and Activation – To put the policy into effect for each node in its data flow.



NOTE – Although this tutorial refers to a VSP volume, these steps are also applicable to a VSP-G or HUS VM volume. VSP, VSP-G and HUS VM are collectively referred to as Hitachi Block Storage on the GUI.



NOTE – A proxy node is required that is responsible for interfacing with the VSP. This must be a Windows machine with a repository installed and must have a connection to the VSP. The proxy node must have CCI (RAID Manager) installed. Refer to the Hitachi Command Control Interface documentation available at <https://portal.hds.com>.



NOTE – If HDID is subsequently uninstalled, the existing snapshots, replication pairs and mounted volumes are left in place; they are not torn down.

Node creation

First, a node representing the VSP needs to be created, unless one already exists. In this example a TrueCopy replication will be performed, this requires both the source and destination VSP have nodes created to represent them.



1. Select **Node Manager** in the **Navigation** pane to access the **Node Management** screen.
2. In the **Agentless/Hardware Node** section, select 'Hitachi Block Storage' in the drop-down list.
3. Press the **Create** button. A configuration dialog (Figure 3-65) will appear.

Figure 3-65 Hitachi VSP Node Configuration

4. Enter a **Node Name** in the text box at the top of the dialog by which to identify the node. We will use the name *VSP*.
5. Enter the **Storage Proxy** node name that is responsible for interfacing with the VSP. This must be a Windows machine with a repository, it must have a connection to the VSP and have Hitachi CCI (RAID Manager) installed. The command device must ***only*** have user authentication enabled. In our case *HardwareDemo* is a suitable node.
6. Click **Lookup**, then in the **Serial Number** dropdown menu select the serial number of the VSP.
7. Enter the **Authentication** credentials for the VSP chosen in the previous step. Note that the password must contain only those characters useable for CCI commands.
8. Specify an **LDEV Range** from which the replica should be allocated.
9. Click **Finish** when ready.
10. The new node with the name *VSP1* will now appear in the Unauthorized Nodes list. Select the node in this list and then press the **Authorize** button  to authorize the node.

If the VSP cannot be accessed or its credentials are invalid then the node will fail authorization. The configuration dialog can be reopened in order to fix any errors by right-clicking on the node and selecting **Edit node**.



NOTE – The Username specified in the VSP node's credentials, must be a member of the Administration User Group on the VSP

Policy definition

A policy defines *data classification* and the *data operation*. In this example, we wish to replicate E:



1. Select **Policy Definition** in the Navigation Pane to access the **Policy Definition** screen (Figure 3-66) and define a policy.
2. In this example, a directory needs to be defined. Expand the **Classification Palette - Standard** heading and drag the **Path** icon from the palette into the **Classification** workspace. Be sure that the **Path** icon remains selected in the workspace. The options specific to the Path classification are displayed in the **Attributes** pane at the bottom of the screen.
3. Enter this value in the **Enter Path** edit box. In this example, enter 'E:\' to define the volume to be replicated.
4. The **Operation Palette** on the right is used to define the data operation. In this example, the path is to be replicated. Drag the **Replicate** icon into the **Operation** workspace just to the left of that palette.
5. When the **Replicate** icon is selected in the workspace area, the options related to the replicate operation are displayed in the **Attributes** pane. Since Replicate does not have any RPO retention options no further changes need to be made.

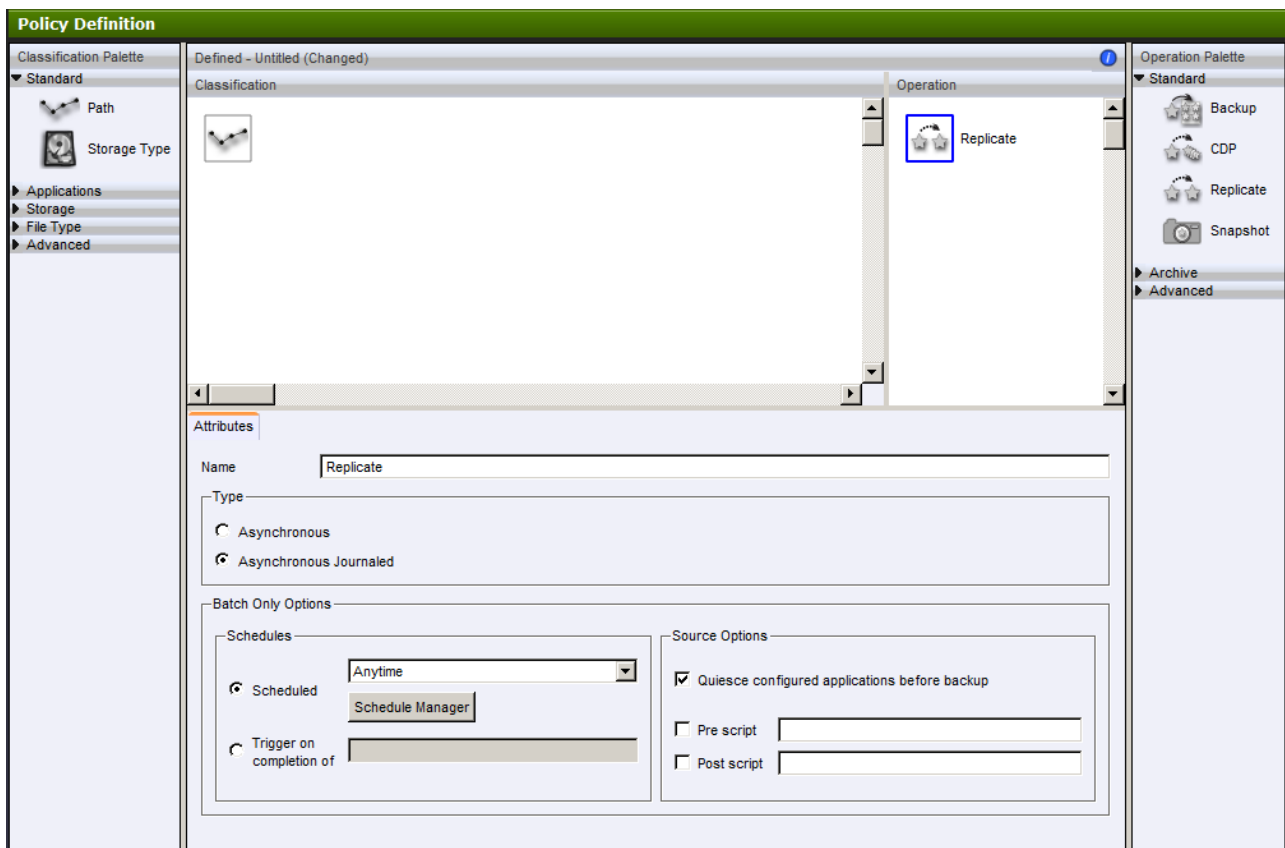


Figure 3-66 Defining a Replication Operation

6. Select **File** → **Save** from the main menu to save the policy. Enter a policy name and brief description, and click **Save**.

Data flow definition

In the case of a replication policy, the data flow defines on which machines the policy will be implemented. The machine that this policy will apply to must be connected to the VSP, and have a VSP volume (LDEV) mounted; in this example the *E:* drive.

The replication policy needs to be assigned to the *HardwareDemo* node connected to the VSP.

This example will be a TrueCopy replication. The VSP the data is replicated to is not the same one as where the source volumes originated from. *VSP1* is VSP where the volumes originate from, this node will not be used but must still be present for TrueCopy to work, *VSP2* is the destination VSP.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow Definition** screen (Figure 3-67).

The palette on the left-hand side contains the Source nodes, Movers and Destination nodes. It also provides Data Instance Director's operations. Replicating a VSP volume requires a source node, a mover and VSP destination node.

2. Drag the *HardwareDemo* icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Continuous** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *HardwareDemo*. If the continuous mover icon is dropped on top of the *HardwareDemo* icon, they are automatically connected.
4. Drag the *VSP2* icon from the **Destination** palette and connect it by dropping it on to the **Continuous** mover icon. *VSP2* is used as this node is not the node which has the source volumes, so this will be a TrueCopy replication.

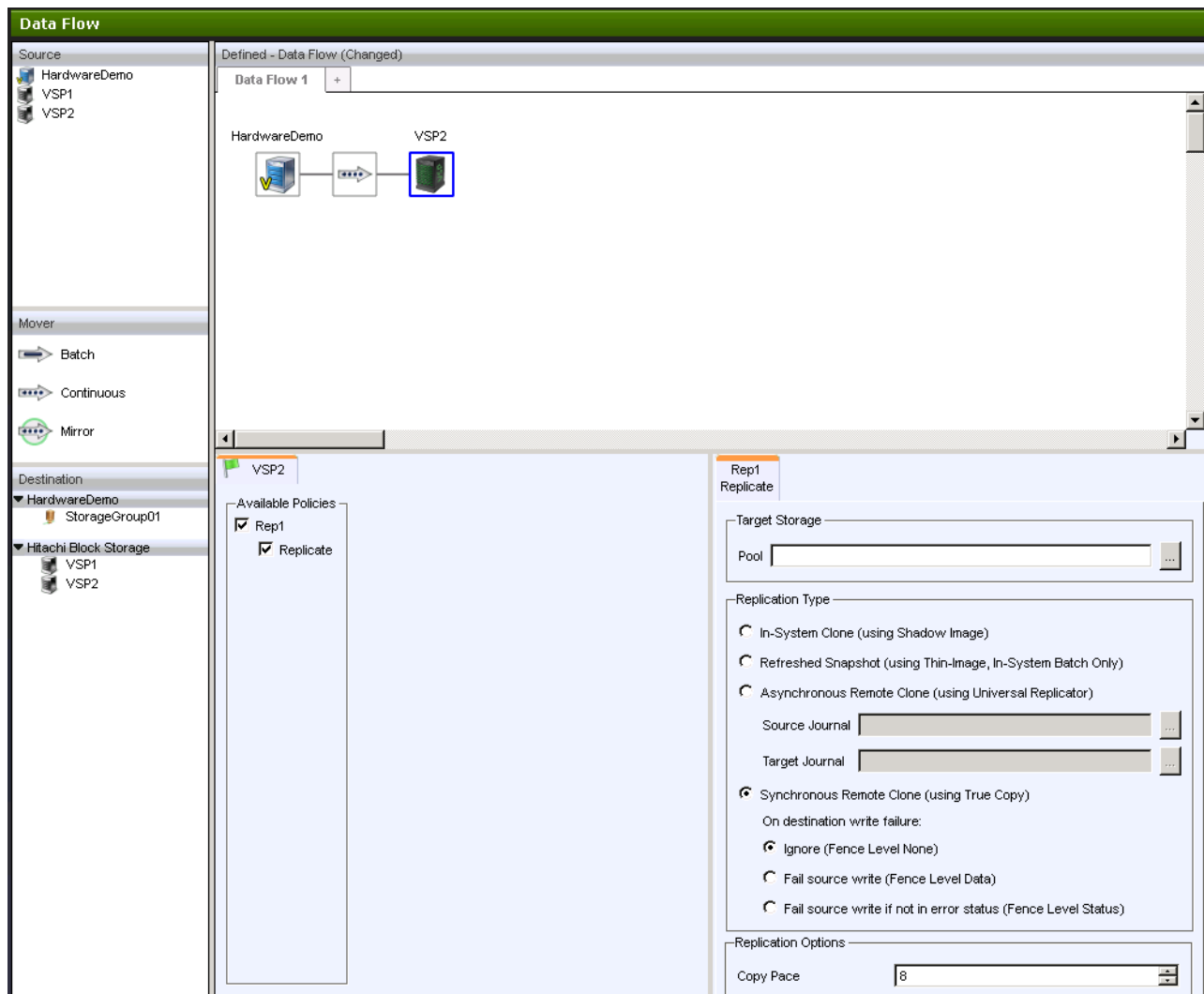


Figure 3-67 Replication Data Flow

5. In the workspace, select the *HardwareDemo* icon to display a list of defined policies in the **Policies to assign locally** tab below the workspace. To see more information about a policy, hover the cursor over a policy.
6. Check the *Rep1* policy (the policy defined above) to assign that policy.
7. Select the **VSP** icon and apply the policy to the destination node.
8. Selecting this policy will give options in the **Rep1 Replicate** tab to the right of the **VSP** tab.
9. Within **Target Storage**, select a **Pool** using the [...] button, the **Select VSP Storage Pool** dialog will appear
10. Within the **Select VSP Storage Pool** dialog select the VSP node from the drop down. This node can be the same VSP node that the disk being replicated is present on (for a ShadowImage replication) or a different VSP (for a TrueCopy replication). Then select a suitable **Pool** from the list presented, and close the dialog.
11. Select a **Replication Type** of **Synchronous Remote Clone (using TrueCopy)** for this example.
12. **Copy Pace** determines how quickly the storage array copies data. Leave this set to the HDID default value of 8.

✓ ***NOTE** – Dynamic Provisioning pools must be created using Hitachi Storage Navigator prior to selecting the Target Storage in HDID.*

Distribution and activation

1. To save and compile in one operation, click the **Save and Compile** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

Hot Tip: The actions that are performed by the orchestration framework as it executes a backup policy are captured by the [Log Manager](#). Detailed logs and their attachments can be exported as a text file (refer to the section [Exporting logs](#)).

Hot Tip: Details of the hardware resources (LDEVs) used for a particular replication or snapshot can be found via the **Resources...** menu option on the [Managed Storage tab](#) of the Storage screen.

Proxy backup/repurposing of a Hitachi block storage replication

This section relates specifically to using the Orchestration Framework for automatically mounting a replicated disk volume residing on a Hitachi Virtual Storage Platform. Please refer to [Hitachi Block snapshotting and replication requirements](#) before starting. This example builds upon the policy and data

flow described in the previous example [Replicating a Hitachi block storage volume](#).

The tasks to be performed when configuring a VSP Snapshot policy are:



Node creation – To inform Data Instance Director of the VSP credentials and details.



Policy definition – To define what volume to replicate and mount.



Data Flow definition – To define the machine on which the volume being replicated resides and how to mount the replicated volume.



Distribution and Activation – To put the policy into effect for each node in its data flow.



***NOTE** – Although this tutorial refers to a VSP volume, these steps are also applicable to a VSP-G or HUS VM volume. VSP, VSP-G, HUS VM (collectively referred to as Hitachi Block Storage) and also to HNAS (referred to as Hitachi File).*



***NOTE** – A proxy node is required that is responsible for interfacing with the VSP. This must be a Windows machine with a repository installed and must have a connection to the VSP. The proxy node must have CCI (RAID Manager) installed. Refer to the Hitachi Command Control Interface documentation available at <https://portal.hds.com>.*



***NOTE** – If HDID is subsequently uninstalled, the existing snapshots, replication pairs and mounted volumes are left in place; they are not torn down.*

Node creation

Refer to the [Node creation](#) steps described in the previous example [Replicating a Hitachi block storage volume](#). The Agentless/Hardware Node configuration should be similar to Figure 3-68.

Figure 3-68 Hitachi VSP Node Configuration

Policy definition

Refer to the [Policy definition](#) steps described in the previous example [Replicating a Hitachi block storage volume](#) in order to define the replication policy, then add the Mount operation as shown in Figure 3-69.

1. The **Operation Palette** on the right is used to define the data operation. In this example, the replicated volume is to be mounted. Drag the **Mount** icon into the **Operation** workspace just below the Replicate operation.
2. When the **Mount** icon is selected in the workspace area, the options related to the operation are displayed in the **Attributes** pane. In this example the replication is continuous (live), so specify a *trigger* **Schedule** that will be used to pause the replication and initiate the Mount operation. **Scripts** can be specified that are run on the mount host **Before unmount** and **After mount**.
3. Select **File** → **Save** from the main menu to save the policy. Enter a policy name and brief description, and click **Save**.

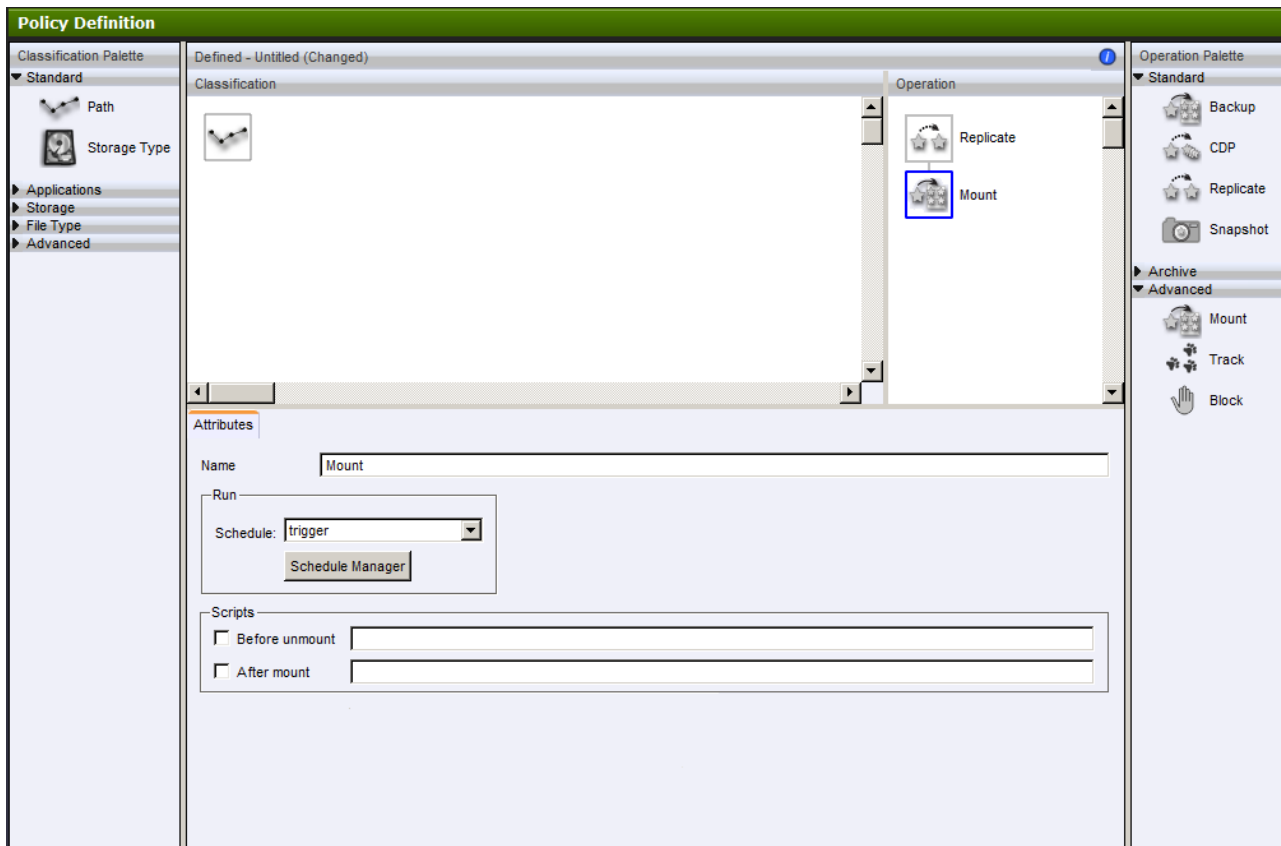


Figure 3-69 Defining a Mount Operation

Dataflow Definition

Refer to [Data flow definition](#) described in the previous example [Replicating a Hitachi block storage volume](#). The mount operation must now be applied to the VSP node.

1. Select the **VSP** icon and apply the **Mount** operation to the destination node.
2. Selecting this policy will give options in the **Rep1 Mount** tab to the right of the **VSP** tab.
3. Set the **Mount Type** to **Proxy Backup**. Refer to the sections [Proxy backup mount sequence](#) and [Repurposing mount sequence](#) below for a description of the differences in the mounting/unmounting process for these mount types.
4. The **Mount host** will be *merlot* in this example.
5. Set the **Mount Options** to **Mount starting with drive** and specify drive letter 'W:.'
6. Set the **Host Group** to **Automatically discover host group**.

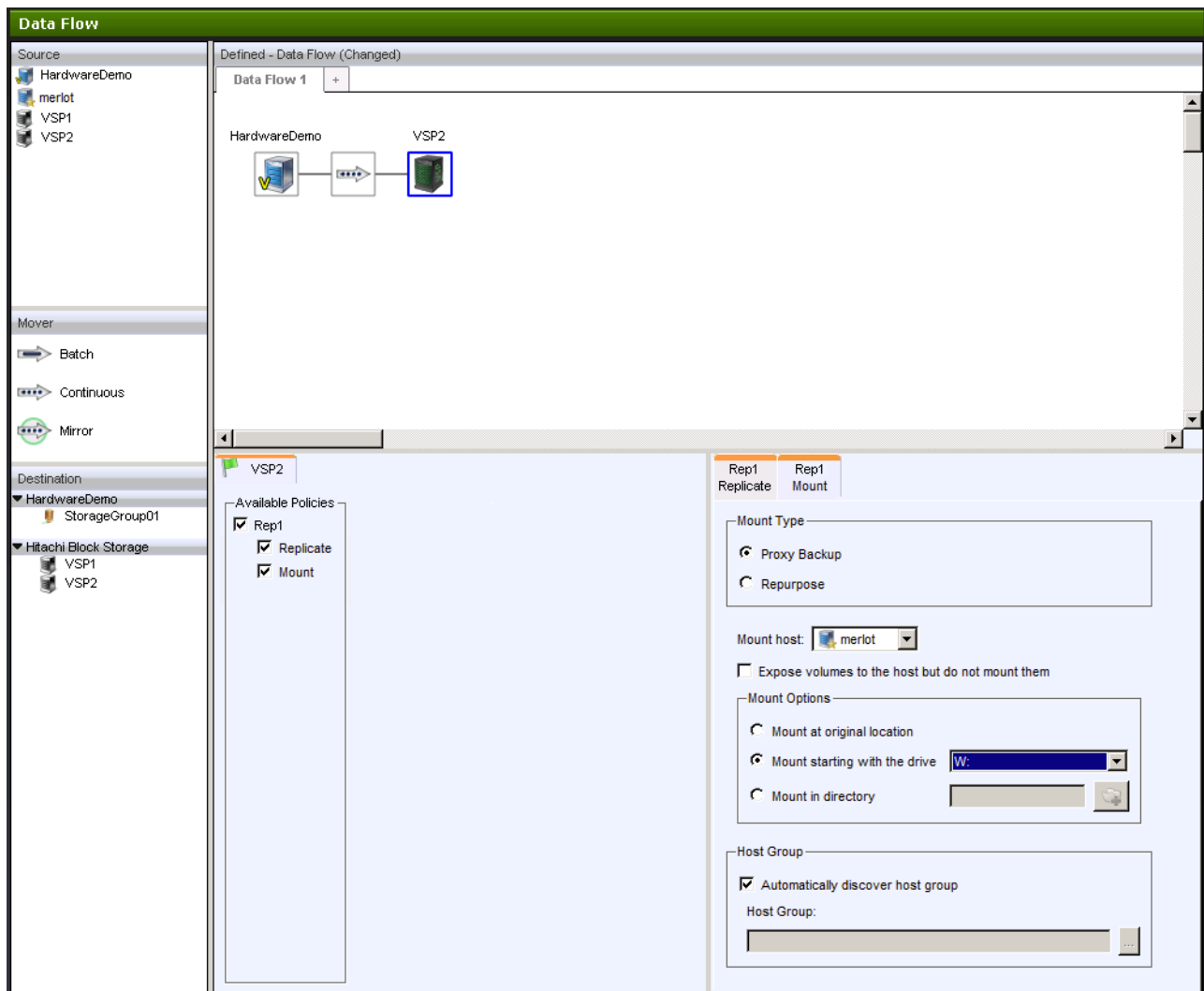


Figure 3-70 Proxy Backup Data Flow

Proxy backup mount sequence

Proxy Backup is valid for continuous (live) and batch backup. In this scenario the replication destination is mounted, work performed on it and then unmounted. The actual sequence is as follows:

- I. The replication is re-evaluated.
- II. If the replication is live then it is paused.
- III. The updated replicated volumes are mounted on the mount host.

- IV. The **'After mount'** script is invoked on the mount host. The script would typically cause the proxy backup to run, calling out to a third party backup product or invoke a Data Instance Director backup via a *trigger* schedule.
- V. The **'Before unmount'** script is invoked on the mount host.
- VI. The replicated volumes are unmounted from mount host.
- VII. If the replication is live then it is resumed.

✓ **NOTE** – A proxy backup of a live replication will generate a warning in the compiler to tell the user that their replication will be paused until the proxy backup is complete.

Repurposing mount sequence

Repurpose is not valid for continuous (live) replication; the rules compiler will issue an error if a continuous mover is used on the data flow in conjunction with this type of mount operation. In this scenario the replication destination remains mounted on the mount host and is available to the repurposing application. The replication is, depending on the schedule, periodically unmounted, re-evaluated then remounted. The actual sequence is as follows:

- I. The **'Before unmount'** script is invoked on the mount host. The script would typically stop any applications from accessing the repurposed volumes while they are being re-evaluated.
- II. The replicated volumes are unmounted from mount host.
- III. The replication is re-evaluated.
- IV. The updated volumes are remounted to mount host.
- V. The **'After mount'** script is invoked on the mount host. The script would typically restart any application accessing the repurposed volumes.

Distribution and activation

1. To save and compile in one operation, click the **Save and Compile** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

Hot Tip: The actions that are performed by the orchestration framework as it executes a backup policy are captured by the [Log Manager](#). Detailed logs and their attachments can be exported as a text file (refer to the section [Exporting logs](#)).

Hot Tip: Details of the hardware resources (LDEVs) used for a particular replication or snapshot can be found via the **Resources...** menu option on the [Managed Storage tab](#) of the Storage screen.

Replicating and snapshotting a Hitachi block storage volume

This section relates specifically to using the Orchestration Framework for snapshotting and replicating a Virtual Storage Platform (VSP) volume (in this example the E: drive); a volume residing on a Hitachi Virtual Storage Platform. Please refer to [Hitachi Block snapshotting and replication requirements](#) before starting. A volume can either be replicated to its original VSP (Batch) or to a different VSP (Live). This example demonstrates how operations can be synchronized with each other to ensure that both the snapshot and replication operations are performed at the same point in time and thus capture identical state of the source data. This mechanism is known as a sync group and is described in [Synchronization Groups](#).

The tasks to be performed when configuring a VSP Snapshot policy are:



Node creation – To inform Data Instance Director of the VSP credentials and details



Policy definition – To define what volume to snapshot



Data Flow definition – To define the machine on which the volume being snapshotted and replicated resides



Distribution and Activation – To put the policy into effect for each node in its data flow



Restore – To mount a volume snapshot at a later date



NOTE – Whilst this tutorial refers to a VSP volume, these steps are also applicable to a VSP-G or HUS VM volume. VSP, VSP-G and HUS VM are collectively referred to as Hitachi Block Storage on the GUI.



NOTE – A proxy node is required that is responsible for interfacing with the VSP. This must be a Windows machine with a repository installed and must have a connection to the VSP. The proxy node must have CCI (RAID Manager) installed. Refer to the Hitachi Command Control Interface documentation available at <https://portal.hds.com>.



NOTE – If HDID is subsequently uninstalled, the existing snapshots, replication pairs and mounted volumes are left in place; they are not torn down.

Synchronization groups

The purpose of synchronization groups is to ensure that, if required, multiple snapshots or batch replications of source data can be created on multiple nodes using the same schedule and yet remain consistent with each other and with the source data.

To this end any two operations in the same policy using the same schedule name will be placed in the same synchronization group. When a dataflow is created, any nodes implementing operations using that synchronization group will be scheduled all as one.



NOTE – The RPO of all snapshot operations in the same synchronization group must be identical. The rules compiler checks the RPO settings and if inconsistent generates error 10356: "Policy <policy> contains multiple hardware snapshot operations with different RPO settings but all use schedule '<schedule>'".

In a simple case of a batch dataflow A -> B -> C where 'B' and 'C' both implement a replicate operation within policy 'P' then if both those operations share the same schedule name 'S' they will be both be part of the same synchronization group and be scheduled as one. This ensures that the replica of 'B' that is created on 'C' is created only when the replica of 'A' created on 'B' has been completed and so the replicas on 'B' and 'C' remain identical.

When mixing live and batch data flows or implement different schedules (synchronization groups) at different points of the data flow, the basic rule for implementing these is that any operation is traced back towards the source, and if an operation sharing the same synchronization group is implemented on an immediately preceding node, then they will be scheduled together by the first preceding node that does not use that synchronization group. Note that live replications are considered to be in all synchronization groups since they should always match the source.

Where synchronization groups are not in use, replicate operations are scheduled on the sending node and snapshot operations are scheduled on the destination node.

Node creation

First, a node representing the VSP needs to be created, unless one already exists. In this example a TrueCopy replication will be performed, this requires both the source and destination VSPs have nodes created to represent them.



1. Select **Node Manager** in the **Navigation** pane to access the **Node Management** screen.
2. In the **Agentless/Hardware Node** section, select 'Hitachi Block Storage' in the drop-down box.
3. Press the **Create** button. A configuration dialog (Figure 3-71) will appear.

Figure 3-71 Hitachi VSP Node Configuration

4. Enter a **Node Name** in the text box at the top of the dialog by which to identify the node. We will use the name *VSP1*.
5. Enter the **Storage Proxy** node name that is responsible for interfacing with the VSP. This must be a Windows machine with a repository, it must have a connection to the VSP and have Hitachi CCI (RAID Manager) installed. The command device must ***only*** have user authentication enabled. In our case *HardwareDemo* is a suitable node.
6. Click **Lookup**, then in the **Serial Number** dropdown menu select the serial number of the VSP.
7. Enter the **Authentication** credentials for the VSP chosen in the previous step. Note that the password must contain only those characters useable for CCI commands.
8. Specify an **LDEV Range** from which the replica should be allocated.
9. Click **Finish** when ready.
10. The new node with the name *VSP1* will now appear in the Unauthorized Nodes list. Select the node in this list and then press the **Authorize** button  to authorize the node.

If the VSP cannot be accessed or its credentials are invalid then the node will fail authorization. The configuration dialog can be reopened in order to fix any errors by right-clicking on the node and selecting **Edit node**.



NOTE – The Username specified in the VSP node, must be a member of the Administration User Group on the VSP

Policy definition

A policy defines *data classification* and the *data operation*. In this example, we wish to snapshot and replicate E:



1. Select **Policy Definition** in the Navigation Pane to access the **Policy Definition** screen (Figure 3-72) and define a policy.
2. In this example, a directory needs to be defined. Expand the **Classification Palette - Standard** heading and drag the **Path** icon from the palette into the **Classification** workspace. Be sure that the **Path** icon remains selected in the workspace. The options specific to the **Path** classification are displayed in the **Attributes** pane at the bottom of the screen.
3. Enter this value in the **Enter Path** edit box. In this example, enter 'E:\' to define the volume to be snapshot and replicated.
4. The **Operation Palette** on the right is used to define the data operation. In this example, the path is to be snapshot and replicated. Drag the **Replicate** icon into the **Operation** workspace just to the left of that palette, then drag a **Snapshot** icon into the **Operation** workspace and then drag another **Snapshot** icon into the **Operation** workspace.
5. When the **Replicate** icon is selected in the workspace area, the options related to the replicate operation are displayed in the **Attributes** pane. Since replication does not have any RPO or retention options no further changes need to be made.
6. When the first **Snapshot** icon is selected in the workspace area, the options related to the snapshot operation are displayed in the **Attributes** pane. In this example, a snapshot every 8 hours is required and the snapshot needs to be kept for two weeks, so set the **Retention** to 2 **Weeks** and set the **Recovery Point Objective** to 8 **Hours**, also under **Mode** select **Hardware** and for **Hardware Type** select **Hitachi Block**. Change the **Name** of the operation, in this example change it to 'LocalSnapshot', to differentiate it from the second snapshot operation which will be synchronized with this operation.
7. When the second **Snapshot** icon is selected in the workspace area, enter a unique name for the second snapshot; in this example change it to 'VSPSnapshot'. Set the **Retention** to 1 **Weeks**. Notice that the **Recovery Point Objective** for the VSP snapshot is ignored since it will be synchronized

with the local snapshot. Under **Mode** select **Hardware** and for **Hardware Type** select **Hitachi Block**.

8. In order to synchronize the local and VSP snapshots, a schedule must be defined and assigned to both snapshot operations. Select the '*LocalSnapshot*' operation. Under **Run** select the **Scheduled** radio button then click **Schedule Manager** and define a schedule and name it '*Synch*'. Finally ensure '*Synch*' is selected in the Run dropdown menu for both the '*LocalSnapshot*' and '*VSPSnapshot*' operations.

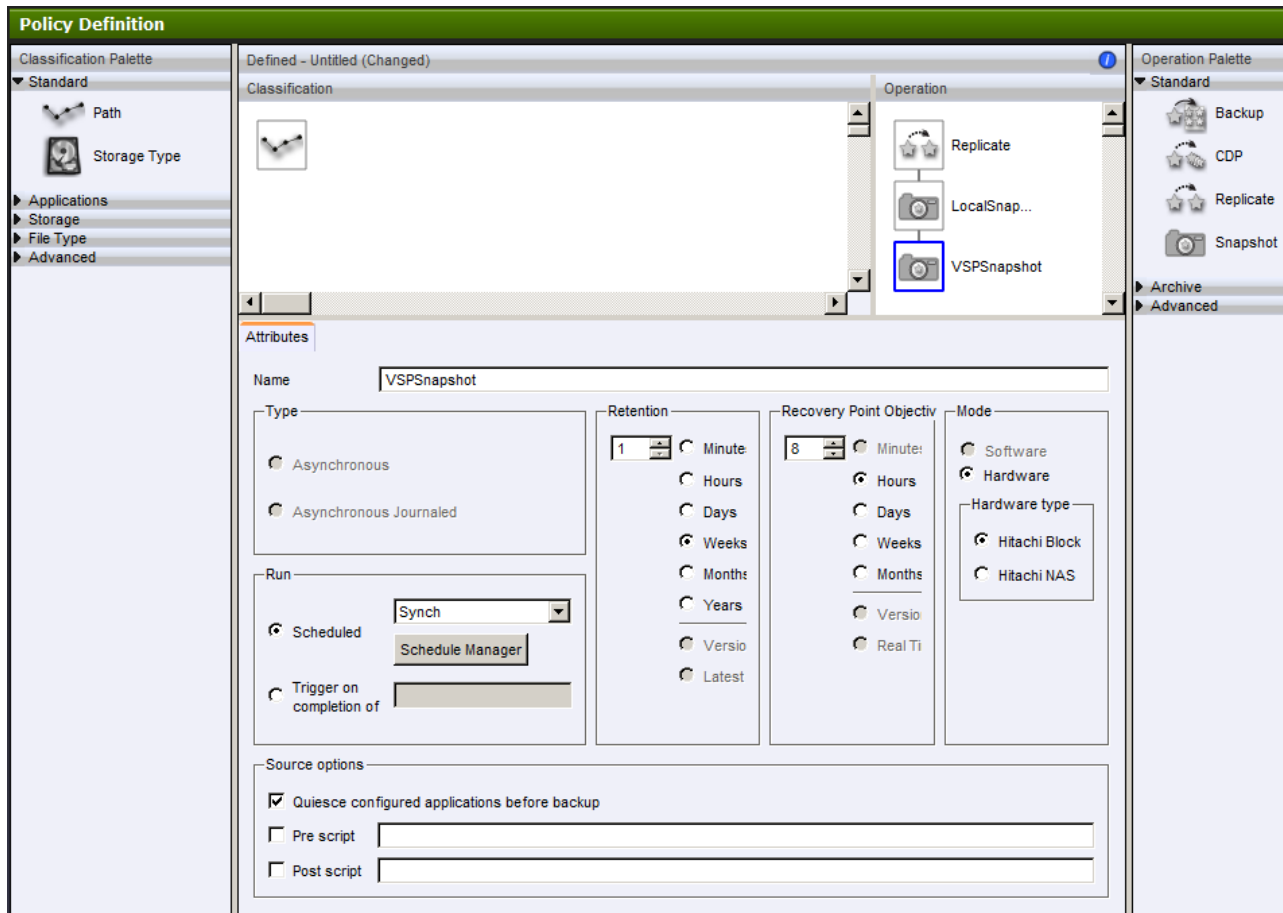


Figure 3-72 Defining a Replication and Snapshot Operation

9. Select **File** → **Save** from the main menu to save the policy. Enter a policy name and brief description, and click **Save**.

✓ **NOTE** – It is possible to just snapshot the replicated volume, instead of doing a synchronized snapshot of the source and replicated volumes. To do this, only drag one snapshot icon into the workspace and place it after the replicate icon (VSPSnapshot in the example above).

Data flow definition

In the case of a replicate and snapshot policy, the data flow defines on which machines the policy will be implemented. The machine that this policy will apply to must be connected to the VSP, and have a VSP volume (LDEV) mounted; in this example the E: drive.

The replicate and snapshot policy will be assigned to the *HardwareDemo* node connected to the VSP.

This example will be a TrueCopy replication. The VSP the data is replicated to is not the same one as where the source volumes originated from. *VSP1* is the VSP where the volumes originate from, this node will not be used but must still be present for TrueCopy to work, *VSP2* is the destination VSP

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow Definition** screen (Figure 3-73). The palette on the left-hand side contains the Source nodes, Movers and Destination nodes. It also provides Data Instance Director's **mover** operations. Replicating a VSP volume requires a source node, a mover and VSP destination node.
2. Drag the *HardwareDemo* icon from the **Source** palette to the **Data Flow** workspace.
3. Drag the **Continuous** icon from the **Mover** palette to the **Data Flow** workspace and connect it to *HardwareDemo*. If the continuous mover icon is dropped on top of the *HardwareDemo* icon, they are automatically connected.
4. Drag the *VSP* icon from the **Destination** palette and connect it by dropping it on to the **Continuous** mover icon.

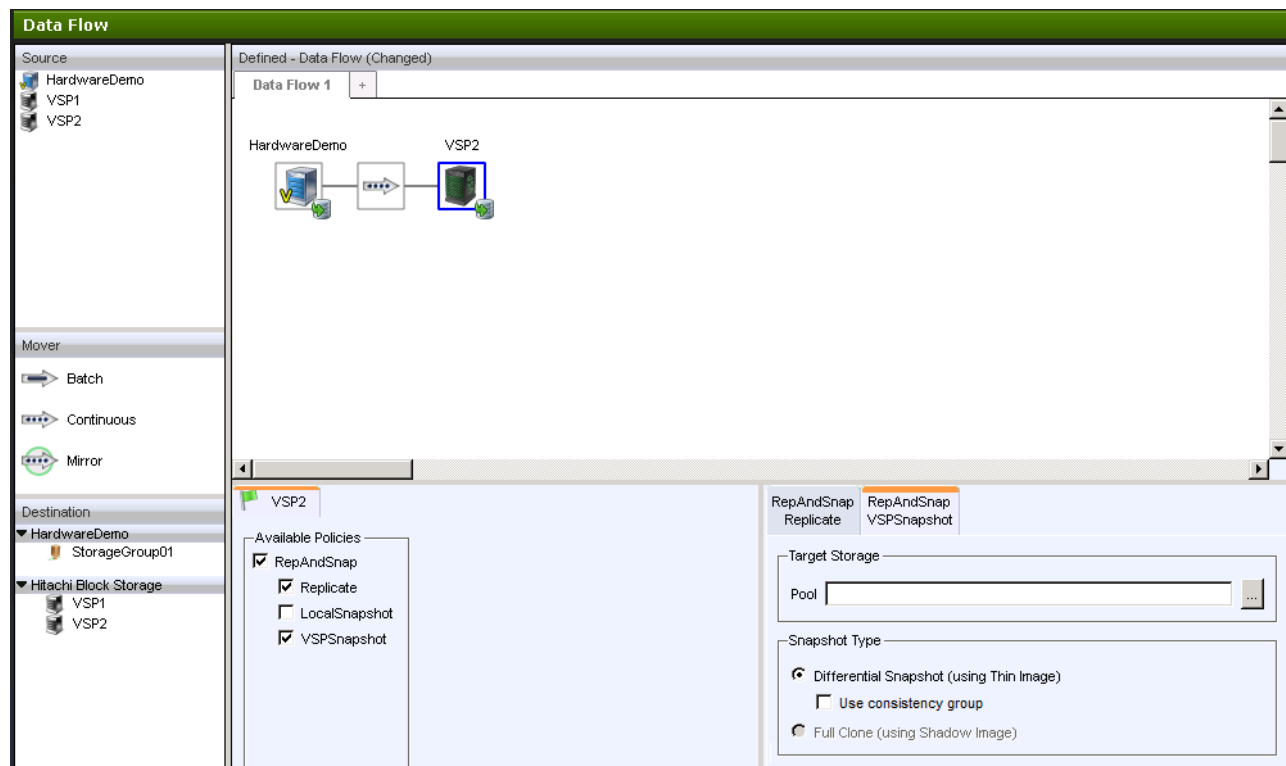


Figure 3-73 Replication and Snapshot data flow

5. In the workspace, select the *HardwareDemo* icon to display a list of defined policies in the **Policies to assign locally** tab below the workspace. To see more information about a policy, hover the cursor over a policy.
6. Check the *RepAndSnap* policy (the policy defined above) to assign that policy.
7. Deselect the *VSPSnapshot* operation on the source node, leaving just *LocalSnapshot* selected.
8. Select the **LocalSnapshot** tab.
9. Within **Target Storage**, select a **Pool** using the [...] button, the **Select VSP Storage Pool** dialogue will appear
10. Within the **Select VSP Storage Pool** dialog select the **VSP** node from the drop down (this node has to be the VSP node the disk being replicated and snapshotted is present on), then select a suitable **Pool** from the list presented.
11. Select a **Snapshot Type** of **Differential (using Thin Image)** for this example.
12. Select the **VSP** icon and apply the policy to this node.
13. Deselect the *LocalSnapshot* operation on the VSP node leaving *Replicate* and *VSPSnapshot* selected.
14. Selecting this policy will give options in the **RepAndSnap Replication** tab to the right of the policy tab.

15. Within **Target Storage**, select a **Pool** using the [...] button; the **Select VSP Storage Pool** dialog will appear.
16. Within the **Select VSP Storage Pool** dialog, select the VSP node from the drop down. This node can be the VSP node the disk being snapshotted is present on (for a ShadowImage replication) or a different VSP (for a TrueCopy replication). Then select a suitable **Pool** from the list presented.
17. Select a **Replication Type** of **Synchronous Remote Clone (using TrueCopy)** for this example then close the dialog.
18. Select the **RepAndSnap VSPSnapshot** tab on the **Data Flow** screen.
19. Within **Target Storage**, select a **Pool** using the [...] button; the **Select VSP Storage Pool** dialog will appear.
20. Within the **Select VSP Storage Pool** dialog select the same VSP node as used for replication from the drop down, then select a suitable **Pool** from the list presented. Close the dialog.
21. Select a **Snapshot Type** of **Differential (using Thin Image)** on the **Data Flow** screen for this example.

✓ ***NOTE** – Thin Image and Dynamic Provisioning pools must be created in Storage Navigator prior to selecting the Target Storage in HDID.*

✓ ***NOTE** – If you are only implementing the replication and destination snapshot scenario (described in the note in the previous section), do not perform the steps to assign a local snapshot.*

Distribution and activation

1. To save and compile in one operation, click the **Save and Compile** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

Hot Tip: The actions that are performed by the orchestration framework as it executes a backup policy are captured by the [Log Manager](#). Detailed logs and their attachments can be exported as a text file (refer to the section [Exporting logs](#)).

Hot Tip: Details of the hardware resources (LDEVs) used for a particular replication or snapshot can be found via the **Resources...** menu option on the [Managed Storage tab](#) of the Storage screen.

Restore

A snapshot of the VSP volume (LDEV) can be mounted via the **Restore** screen (Figure 3-63).

1. Search the **Browse Backup and Snapshots** tab, results of all snapshots meeting the search criteria will be displayed.
2. Expand the Tree for the policy VSP Volume Snapshot
3. Select a snapshot within this policy, right-click it and select **Mount**
4. Select from the **Mount host** drop down which machine this snapshot of the volume is to be mounted and select **Mount starting with the drive**, in this example we will mount to **HardwareDemo** and mount it to **J:**



WARNING: For AIX hosts, when performing unmount and revert operations on Hitachi File and Block storage, the user is required to first manually unmount the file systems from the host operating system.

When unmounting from an AIX host the following warning is displayed to the user: "This file system must be unmounted manually before an 'unmount' operation is performed by HDID, failing to do so can cause a loss of data".

When reverting on an AIX host the following warning is displayed to the user: "This file system must be unmounted manually before a revert can take place, failing to do so may cause loss of data".

When performing a mount operation on an AIX host:

The mount dialog options are forced to **Expose volumes to the host but do not mount them** and all other options disabled; thus the file system will not actually be mounted.

✓ ***NOTE** – Mount at original location is not currently supported if the original source volumes are present. The base OS may experience problems if there is already an instance of that file system mounted. This is because the snapshots share the same disk partition IDs. In extreme conditions it can cause corruption to both disks.*

✓ ***NOTE** – Unless the user selects a host group, the machine where the volume is to be mounted MUST have an existing volume on the same storage device. If there is no connection between the mount host and the block storage device then Data Instance Director will fail the mount operation timeout after 30 minutes.*

✓ ***NOTE** – When reverting a snapshot of an MS SQL server, the SQLWriter service must be running whilst performing the revert operation. It is also necessary to restart the SQL service (MSSQLSERVER) after the revert operation is complete.*

✓ **NOTE** – When reverting a snapshot of an Exchange server, it is necessary to stop the Exchange services prior to reverting and then reboot the server after the revert operation is complete.

Snapshotting a Hitachi File Storage (Hitachi NAS) path

This section relates specifically to using the Orchestration Framework for snapshotting a path ('/Documents') exported from a Hitachi Network Attached Storage device and mounted on a machine (in this example 'HardwareDemo').

The tasks to be performed when configuring an Hitachi NAS Snapshot policy are:



Node creation – To inform Data Instance Director of the Hitachi NAS credentials and details.



Policy definition – To define what path to snapshot.



Data Flow definition – To define the machine on which the path being snapshot resides.



Distribution and Activation – To put the policy into effect for each node in its data flow.



Restore – To mount a path snapshot at a later date.

✓ **NOTE** – A proxy node is required that is responsible for interfacing with the Hitachi NAS. This must be a Windows machine with a repository and must have a connection to the Hitachi NAS.

✓ **NOTE** – HDID requires an EVS (Enterprise Virtual Server) configured for file services and file system mounted on it. Snapshots take up space within the file system they are created on so sufficient space is needed to store the changes.

✓ **NOTE** – To identify which cluster it is talking to, HDID requires file system EVSes to be able to provide some information about the cluster they belong to. This is not enabled by default. To enable it log on to the administration EVS via SSH using as the supervisor user and run the commands:

```
mscfg http enable
mscfg http evsipaddr on
```

HDID requires the following Hitachi NAS management software and file system versions:

```
SMU: 12.2.3708.00
WFS-2
```

Node creation

First, a node representing the Hitachi NAS needs to be created, unless one already exists.

1. Select **Node Manager**  in the navigation pane to access the **Node Management** screen.
2. In the **Agentless / Hardware Node** section, select '*Hitachi Network Attached Storage*' in the drop-down box.
3. Press the **Create** button. A configuration dialog (Figure 3-74) will appear.

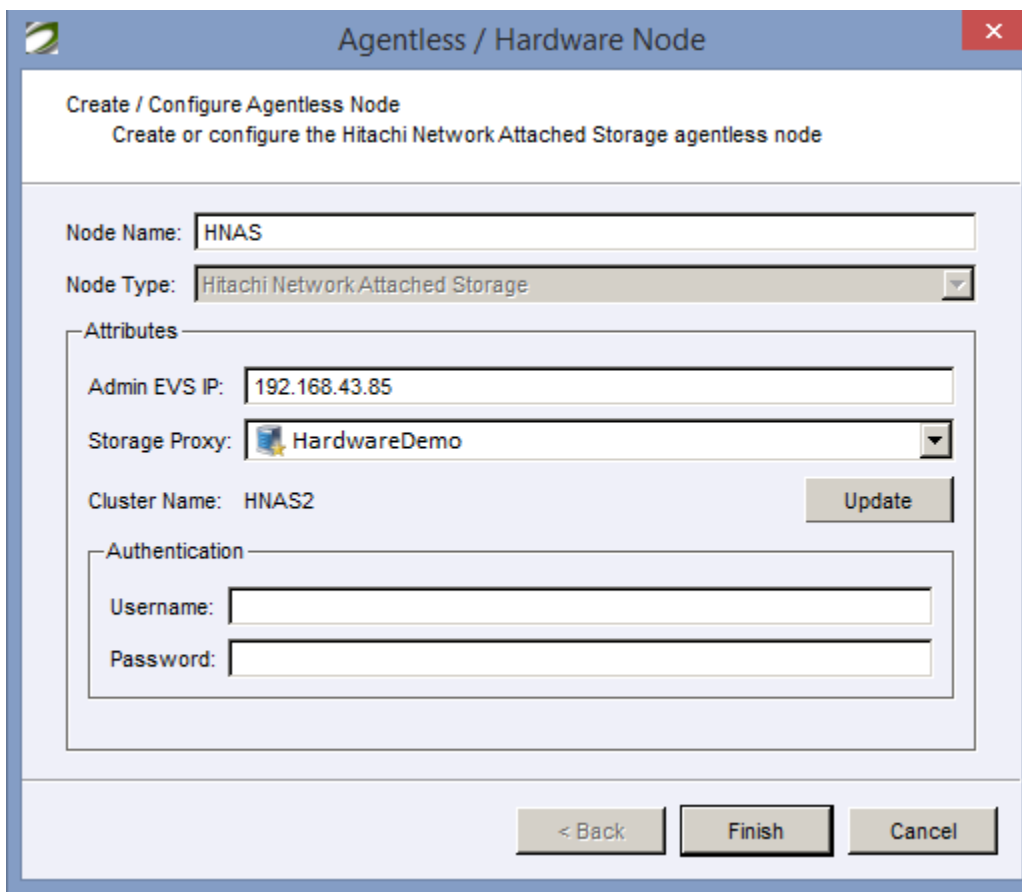


Figure 3-74 Hitachi NAS Agentless Node Configuration

4. Enter a **Node Name** in the text box at the top of the dialog by which to identify the node. We will use the name *Hitachi NAS*.
5. In the **Admin EVS IP** section enter either the DNS name or IP address of the Hitachi NAS EVS running the administrative service.

6. Select a node to use as the **Storage Proxy**. This node will be used to manage the Hitachi NAS cluster and store the records relating to snapshots and replication.
7. Click **Update**. This will retrieve the cluster name and ID and display it in the dialog using the previously entered EVS address and storage proxy node.
8. Under **Authentication**, enter the **Username** and **Password**. Note that this is not the password used to access the web interface.
9. Click **Finish** when ready.
10. The new node with the name *Hitachi NAS* will now appear in the Unauthorized Nodes list. Select the node in this list and then press the authorize button  to authorize the node. Note that unauthorized Hitachi NAS nodes are shown as inactive until they are authorized.

If the Hitachi NAS cannot be accessed or its credentials are invalid then the node will fail authorization. If the EVS address is changed to a different cluster then the cluster information will need to be updated again or the node will not authorize. The configuration dialog can be reopened in order to fix any errors by right-clicking on the node and selecting **Edit node**.



NOTE – The password specified in the Hitachi NAS node's credentials must be that of the supervisor user account.

Policy definition

A policy defines *data classification* and the *data operation*. In this example, we wish to snapshot the path *'/Documents'*



1. Select **Policy Definition** in the Navigation pane to access the **Policy Definition** screen (Figure 3-75) and define a policy.
2. In this example, a path needs to be defined. Expand the **Classification Palette - Standard** heading and drag the **Path** icon from the palette into the **Classification** workspace.

Be sure that the **Path** icon remains selected in the workspace. The options specific to the Path classification are displayed in the **Attributes** pane below the workspace.
3. Enter a path name in the **Enter Path** edit box. In this example type *'/Documents'* to define the mounted export to be snapshotted.



NOTE – The path specified must resolve to a directory or sub directory of a mounted Hitachi NAS NFS export.

4. The **Operation Palette** on the right is used to define the data operation. In this example, the path *'/Documents'* is to be snapshotted so drag the **Snapshot** icon into the **Operation** workspace just to the left of the palette.
5. When the **Snapshot** icon is selected in the workspace area, the options related to the snapshot operation are displayed in the **Attributes** pane. In this example, a snapshot is required every 8 hours and needs to be kept for 2 weeks, so set the **Retention** to **2 Weeks** and set the **Recovery Point Objective** to **8 Hours**. Finally set **Mode** to **Hardware** and **Hardware Type** to **Hitachi NAS**. **Quiesce configured applications before backup** must be unchecked for hardware snapshot policies because HDID does not know which applications are using the storage.

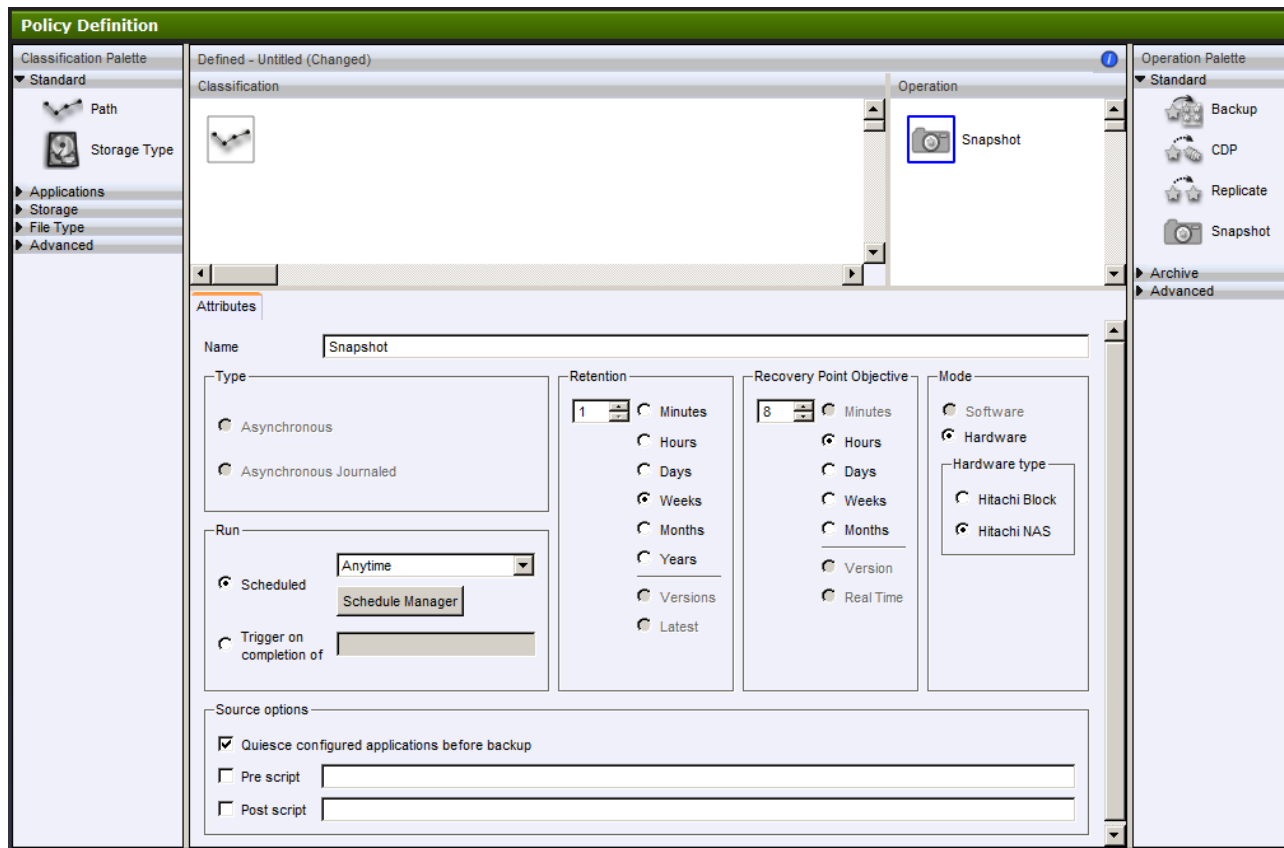


Figure 3-75 Defining an Hitachi NAS Snapshot Operation

6. Select **File** → **Save** from the main menu to save the policy. Enter a policy name and brief description, and click **Save**.

✓ **NOTE** – The classifications that are applicable to Hitachi NAS policies are *Path*, *Oracle* and *Hitachi NAS*

Data flow definition

In the case of a snapshot policy, the data flow defines on which machines the policy will be implemented. The machine that this policy will apply to must be connected to the Hitachi NAS and have the Hitachi NAS export mounted; in this example the *'/Documents'* folder.

The snapshot policy needs to be assigned to the *HardwareDemo* node connected to the Hitachi NAS.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow Definition** screen (Figure 3-76).

The palette on the left-hand side contains the Source nodes, Movers and Destination nodes. Snapshotting a mounted Hitachi NAS export only requires a source node to be specified.

2. Drag the *HardwareDemo* icon from the **Source** palette to the **Data Flow** workspace.

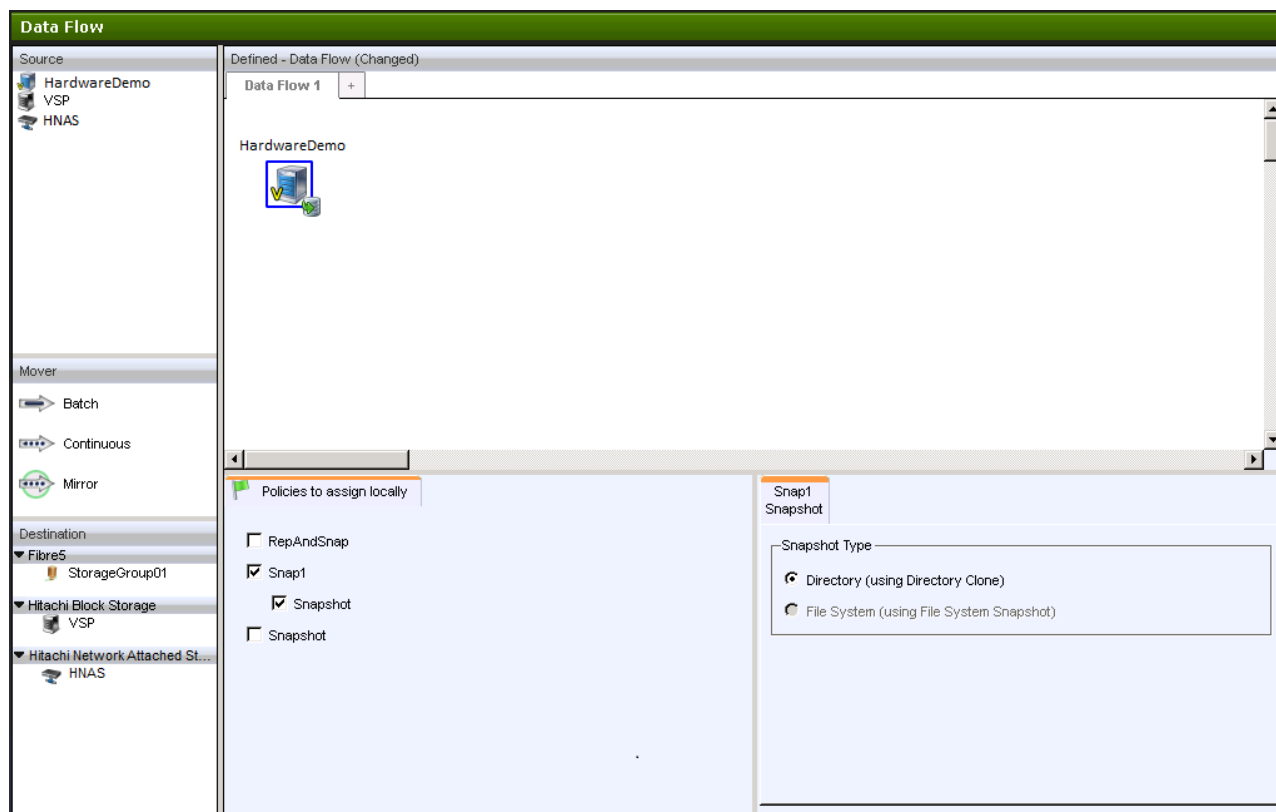


Figure 3-76 Hardware Snapshot Data Flow

3. In the workspace, select the *HardwareDemo* icon to display a list of defined policies in the **Policies to assign locally** tab below the workspace. To see more information about a policy, hover the cursor over a policy.

4. Check the *Snap1* policy (the policy defined above) to assign that policy.
5. Selecting this policy will give options in the **Snapshot** tab to the right.
6. Within **Snapshot Type**, select **Directory (using Directory Clone)**.



NOTE – HDID version 5.1 does not currently support the File System snapshot option.

Distribution and activation

1. To save and compile in one operation, click the **Save and Compile** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

Hot Tip: The actions that are performed by the orchestration framework as it executes a backup policy are captured by the [Log Manager](#). Detailed logs and their attachments can be exported as a text file (refer to the section [Exporting logs](#)).

Restore

A snapshot of an Hitachi NAS directory can be mounted via the **Restore** screen (Figure 3-77).

1. Enter **Criteria** into the **Browse Backup and Snapshots** tab, results of all snapshots meeting the search criteria will be displayed.
2. Expand the tree for the policy *Snap1*.
3. Select a snapshot within this policy, right-click it and select **Mount** from the context menu. The **Mount Snapshot Set** dialog is displayed.
4. From the **Mount host** drop down select the machine on which this snapshot is to be mounted and select the **Mount in directory** radio button, in this example we will mount to */mnt* on *HardwareDemo*.

A snapshot can be unmounted either in the **Restore** or **Storage** screens, however the **Storage** screen (Figure 3-78) shows which snapshots are mounted; either way the procedure to unmount is identical.

5. Select a snapshot and right-click it and select **Unmount**
6. The snapshot will be unmounted. It may take several minutes before the unmount is actually completed.



WARNING: For AIX hosts, when performing unmount and revert operations on Hitachi File and Block storage, the user is required to first manually unmount the file systems from the host operating system.

When unmounting from an AIX host the following warning is displayed to the user:
“This file system must be unmounted manually before an 'unmount' operation is performed by HDID, failing to do so can cause a loss of data”.

When reverting on an AIX host the following warning is displayed to the user:
“This file system must be unmounted manually before a revert can take place, failing to do so may cause loss of data”.

When performing a mount operation on an AIX host:

The mount dialog options are forced to **Expose volumes to the host but do not mount them** and all other options disabled; thus the file system will not actually be mounted.

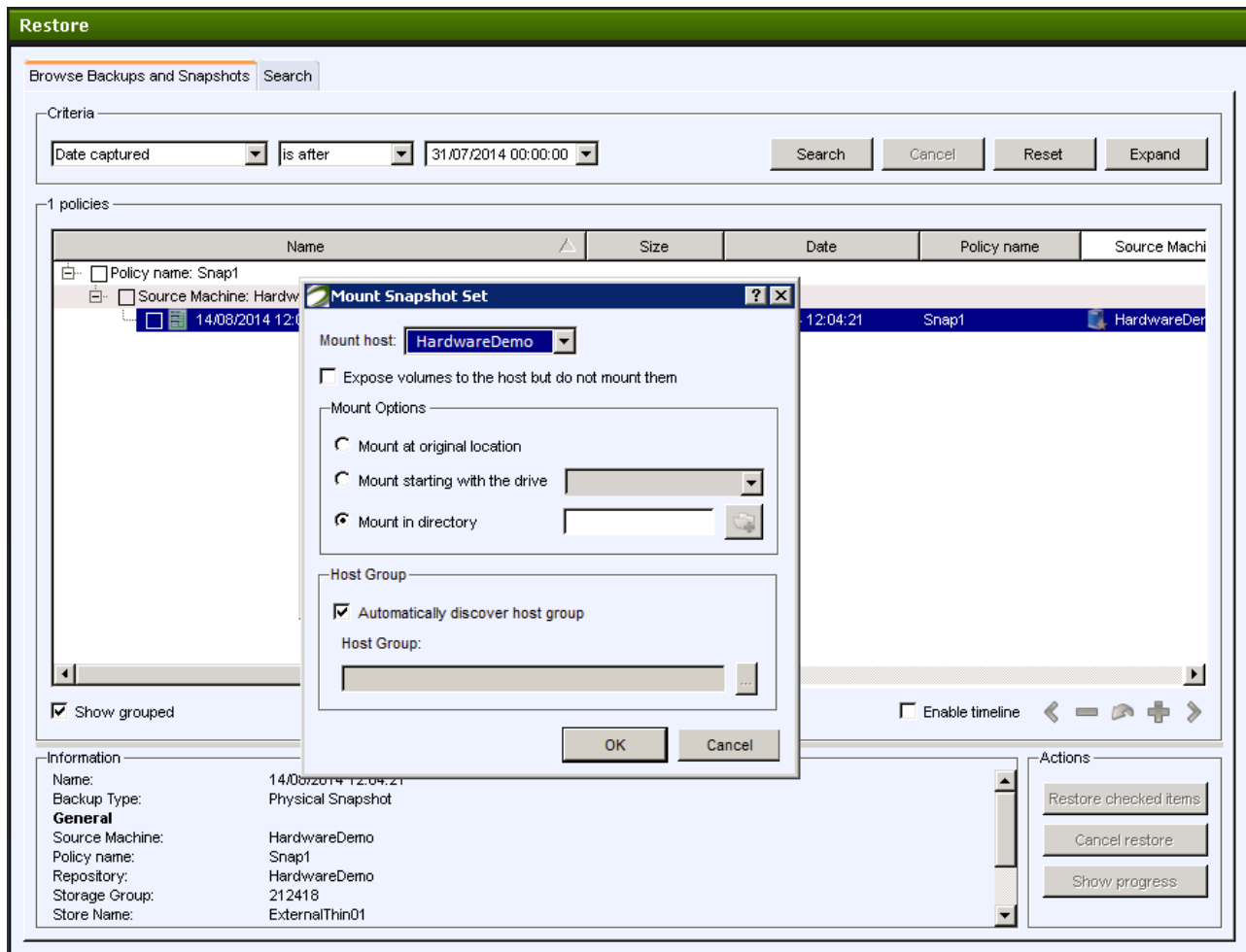


Figure 3-77 Hardware snapshot mounting

Storage

Repositories
Repository Usage
Managed Storage
Library Management
Media
Media Requests

Filter

Intelligent Storage Manager node
All
Select
Woodchuck

Snapshot range
All
Select
18/09/2014 00:00:00
to
01/10/2014 00:00:00

Intelligent Storage Snapshots

Name	Information	Policy	Used Size	Max Size	Available	Reusable	Expiry	Application	Mounted On
Woodchuck	ISM Node								
newHNAS	HNAS 212418								
PeteThin01									
Woodchuck	Application Host								
30/09/2014 09:17:06	Physical Snapshot	WoodPeteTlsq2909	546 MB	8 TB			07/10/2014 09:17:06	Volume Snapshot Service	
30/09/2014 01:17:06	Physical Snapshot	WoodPeteTlsq2909	462 MB	8 TB			07/10/2014 01:17:06	Volume Snapshot Service	
29/09/2014 17:17:07	Physical Snapshot	WoodPeteTlsq2909	420 MB	8 TB			06/10/2014 17:17:07	Volume Snapshot Service	
29/09/2014 09:17:06	Physical Snapshot	WoodPeteTlsq2909	420 MB	8 TB			06/10/2014 09:17:06	Volume Snapshot Service	THATCHERS
29/09/2014 09:06:36	Physical Snapshot	WoodPeteTlsq2909	378 MB	8 TB			06/10/2014 09:06:36	Volume Snapshot Service	

Figure 3-78 Snapshot shown as mounted

Snapshotting an Oracle database on Hitachi block storage

This section relates specifically to creating a snapshot of an Oracle database (running either as a single instance or as a RAC) located on Hitachi Block Storage (VSP). An Oracle database snapshot can either be mounted to its original location or to a different directory. Please refer to [Hitachi Block snapshotting and replication requirements](#) before starting.

The tasks to be performed when configuring an Oracle snapshot policy are:



Node creation – To inform Data Instance Director of the Oracle object credentials and information.



Policy definition – To define what databases to snapshot



Data Flow definition – To define the machine on which the database being snapshot resides



Distribution and Activation – To put the policy into effect.



Restore – To mount a database snapshot at a later date

Before configuring HDID to perform Oracle database snapshotting, the following points should be noted:

a. Oracle file location prior to backup:

It is necessary to separate Oracle data files from redo files, especially if you intend to perform in-system revert operations. If data and redo files are not separated, it is impossible to perform in system reverts (i.e. it is not possible to go beyond the point in time when the snapshot occurred), snapshots cannot be performed. Separation of Oracle files must be as follows:

- ASM disks have to be partitioned. The ASM Data has to be on the only partition of this disk. This is in accordance to the requirements of Oracle.
- Database files and redo logs have to be in separate directories/ASM Diskgroups
- Database files and archive logs have to be in separate directories/ASM Diskgroups
- Redo logs and archive logs can be in separate or one dedicated directory/ASM Diskgroup
- SPfiles are only transported if located within the datafile or archive log space
- Entries within the \$ORACLE_HOME/network/admin path (i.e. sqlnet.ora, tnsnames.ora, listener.ora) are not transported and have to be provided on the "Infrastructure to show the Point in Time Copy". This is the database administrator's responsibility and these files will need modification in any case.

- Control files are recommended to be placed in the archive log directory/ASM Diskgroup.
- For performance reasons it is recommended to have as few as possible files within the asm_diskstring. An asm_diskstring = '/dev/*' would require checking many hundreds of entries on an average Linux machine. Be precise when defining asm_diskstring, e.g. '/dev/oracleasm/disks/*' with only the ASM disk special nodes in this directory.

b. File movements performed by HDID during mount operations:

A copy of the following files are backed up by HDID in \$ORACLE_HOME/SnapShotBackup/<SID>/<BackupTimestamp> on the server that sees the mount operation:

- controlfile (binary and ascii format),
- spfile (binary and ascii format),
- files recording the startCN and endCN,
- oracle-Password file
- init.ora file

Control files and SPfile are only copied if they are either in the same ASM Diskgroup or in the same directory with the database files or archive logs/redo logs.

ACFS is not supported (although the underlying ASM Diskgroup is copied).

The following files are not copied. A separate operation must be put in place if they are required:

- Database binaries
- External tables
- External LOBs (BFiles)
- Encryption wallets

c. Infrastructures supported on the host:

HDID supports the following infrastructure types on the originating Oracle database (i.e. the database being backed up):

- Single instance with local or Hitachi NAS file systems
- Single instance with ASM (i.e. ASMLib or udev)
- RAC with NAS file system
- RAC with ASM (i.e. ASMLib or udev)
- Linux device mapper multipathing (but not HDLM)
- GPT partitioned disks (one primary partition)

d. Supported mount (destination) infrastructures

HDID supports the following infrastructure types on the destination Oracle database (i.e. the place where a database snapshot is to be mounted):

- Single instance with local file system
- Single instance with ASM (the ASM disk string is used, the standard location is /dev/oracleasm/disks,)
- The asm_diskstring for mount operations has explicit requirements:
 - o the first element of a chained asm_diskstring will be used
 - o the element must start with /dev
 - o the element must be a directory
 - o the element must not have sub-directories
 - o the element must end with '/*'
 - o all elements must not include /dev/disks/by-id
- No RAC support
- Linux device mapper multipathing (but not HDLM)

e. Privileges required to perform snapshotting

HDID requires the following privileges to perform snapshotting:

- Commands to Oracle need sysdba privileges
- Commands to ASM need sysasm privileges
- If no Oracle credentials are supplied, those of the OS user of \$ORACLE_HOME/bin/oracle are used for Oracle commands (for Linux this user must be a member of a group having sysdba and sysasm)
- If no ASM credentials are supplied, those of the OS user of \$ORACLE_HOME/bin/oracle are used for ASM commands (for Linux the 'grid' or 'oracle' user are used if possible)

f. ASM Diskgroup support for HDID operations

The following ASM protection modes are supported:

- External protection – all operations
- Normal (2 failure groups) or high protection (3 failure groups) – all operations except revert
- Normal or high protection with other configurations – not supported

Note - normal or high protection is degraded on a mount host since there is only one failure group available.

Node creation

First, a node representing the Oracle database needs to be created, unless one already exists..



1. Select **Node Manager** in the **Navigation** pane to access the **Node Management** screen.
2. In the **Agentless/Hardware Node** section, select 'Oracle' in the drop-down box.

3. Press the **Create** button. A configuration dialog (Figure 3-79) will appear.

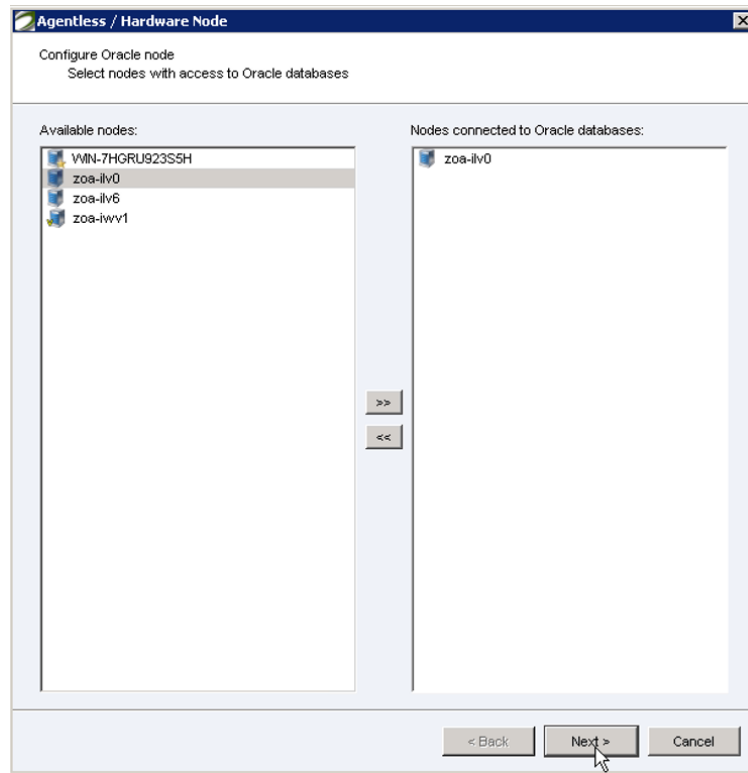


Figure 3-79 Oracle node configuration dialog (page 1)

4. To add a single instance, select the Oracle node in the **Available nodes:** list and click the [**>>**] button to add it to the **Nodes connected to Oracle databases:** list. To add a RAC cluster, select each node in the RAC and add them to the right hand list. Click **Next>** to display the second page of the **Agentless / Hardware Node** dialog (Figure 3-80).

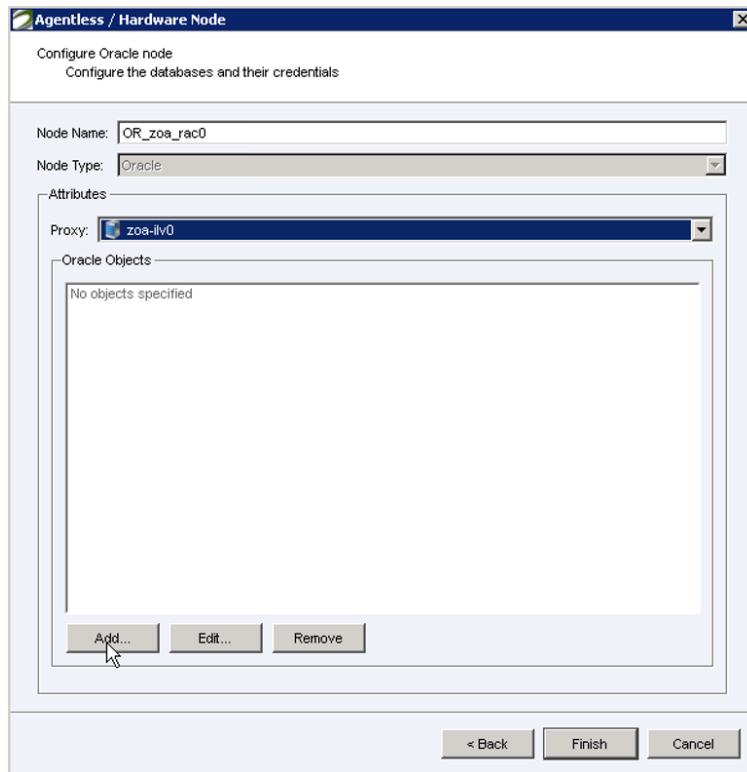


Figure 3-80 Oracle node configuration dialog (page 2)

5. Enter a **Node Name** in the text box at the top of the dialog by which to identify the agentless node. We will use the name *OR_zoa_rac0*.
6. Select the **Proxy** node name that is responsible for interfacing with the Oracle server. This must be either the Oracle server node (when dealing with a single instance setup) or one of the RAC nodes (when dealing with a RAC setup).
7. Click **Add...** to add objects to the **Oracle Objects** list. The **Enter object information** dialog is displayed (Figure 3-81).

The screenshot shows a Windows-style dialog box titled "Enter object information". It contains three main sections:

- Object Information:** Includes a dropdown menu for "Object Type" (set to "Oracle Database"), a text box for "Oracle SID" (containing "aoracle"), and a text box for "Host".
- Object Credentials:** Includes text boxes for "User name:", "Password:", and "Confirm Password:".
- OS Credentials:** Includes text boxes for "Domain:", "User name:", "Password:", and "Confirm Password:".

At the bottom of the dialog are "OK" and "Cancel" buttons. A mouse cursor is pointing at the "OK" button.

Figure 3-81 Oracle object information and credentials

8. Select 'Oracle Database' from the Object Type dropdown and enter the **Oracle SID** in the edit box provided. In this case HDID uses the credentials of owner of the Oracle binary to connect to the database.

✓ ***NOTE** – Object Credentials can be added if authentication is to be performed using an Oracle password file or via Oracle DB authentication (Oracle DB authentication will not work for offline backups). Specify OS Credentials if you want to use a special account for authentication. Optionally, ASM credentials can be specified by selecting 'Oracle ASM' in the Object Type dropdown.*

9. Click **OK** when ready.
10. The new node with the name `OR_zoa_rac0` will now appear in the Unauthorized Nodes list. Select the node in this list and then press the **Authorize** button to authorize the node. If the node cannot be accessed or its credentials are invalid then it will fail authorization. The configuration dialog can be reopened in order to fix any errors by right-clicking on the node and selecting **Edit node**.

Policy definition

A policy defines *data classifications* and the *data operations*. In this example, we wish to take snapshots of an Oracle database represented by the agentless node `OR_zoa_rac0`.

1. Select **Policy Definition**  in the Navigation Pane to access the **Policy Definition** screen (Figure 3-82) and define a policy.
2. In this example, an Oracle database needs to be defined. Expand the **Classification Palette - Applications** heading and drag the **Oracle Database** icon from the palette into the **Classification** workspace. Be sure that the icon remains selected in the workspace. The options specific to the **Oracle Database** classification are displayed in the **Attributes** pane at the bottom of the screen.
3. Select **Online** from the **Backup Mode** radio buttons (note that **Crash** is not currently supported). In the **Databases** area select 'aoracle' in the **Available databases** list then click **Include** to add it to the **Included databases** list – the ones that are to be snapshotted.

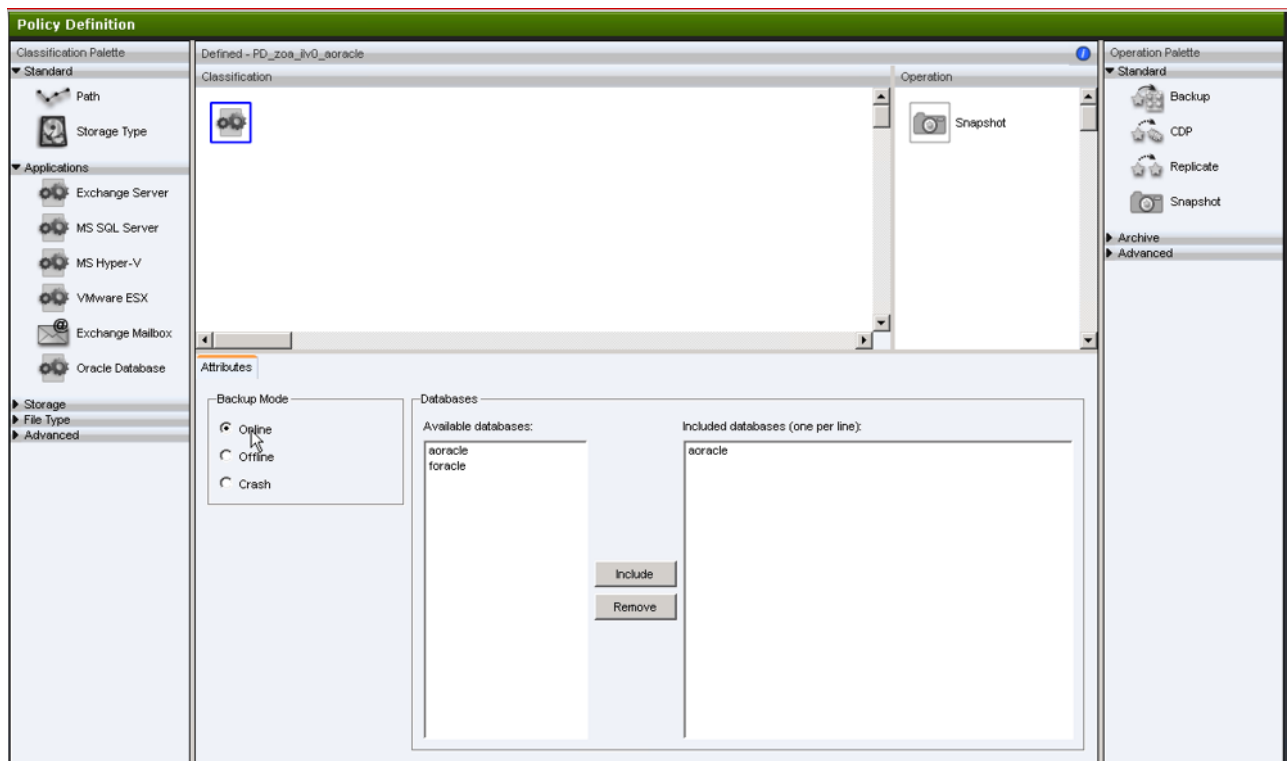


Figure 3-82 Defining an Oracle Classification

4. The **Operation Palette** on the right of the screen (Figure 3-83) is used to define the data operation. In this example, the database is to be snapshotted. Drag the **Snapshot** icon into the **Operation** workspace just to the left of that palette.
5. With the **Snapshot** icon selected in the workspace area, the options related to the snapshot operation are displayed in the **Attributes** pane. In this example, a snapshot every 8 hours is required and the snapshot needs to be kept for one week, so set the **Retention** to **1 Weeks** and set the **Recovery**

Point Objective to **8 Hours**, also under **Mode** select **Hardware** and for **Hardware Type** select **Hitachi Block**. Leave the **Name** of the operation as the default '*Snapshot*'.

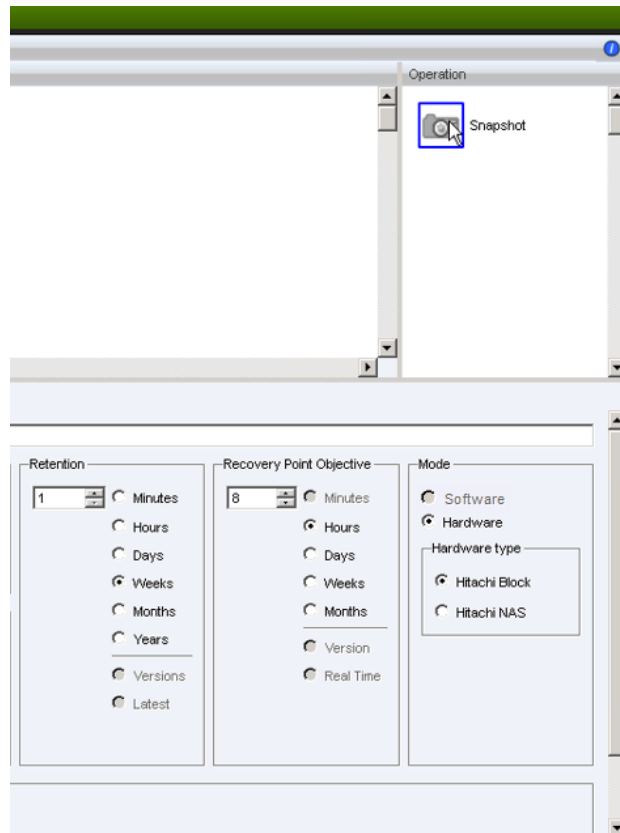


Figure 3-83 Defining a Snapshot Operation

6. Select **File** → **Save** from the main menu to save the policy. Enter the policy name '*PD_zoa_ilv0_aoracle*' and brief description, and click **Save**.

Data flow definition

In the case of an Oracle snapshot policy, the data flow defines on which machines the policy will be implemented. The machine that this policy will apply to must be connected to the Oracle server with the database residing on a VSP volume (LDEV).

The snapshot policy will be assigned to the *OR_zoa-ilv0* agentless node connected to the Oracle database server or RAC.

1. Select **Data Flow**  in the **Navigation** pane to access the **Data Flow Definition** screen (Figure 3-84). The palette on the left-hand side contains the **Source** nodes, **Movers** and **Destination** nodes. Snapshotting an Oracle database requires only the Oracle server agentless node to be specified.

2. Drag the *OR_zoa_ilv0* icon from the **Source** palette to the **Data Flow** workspace.

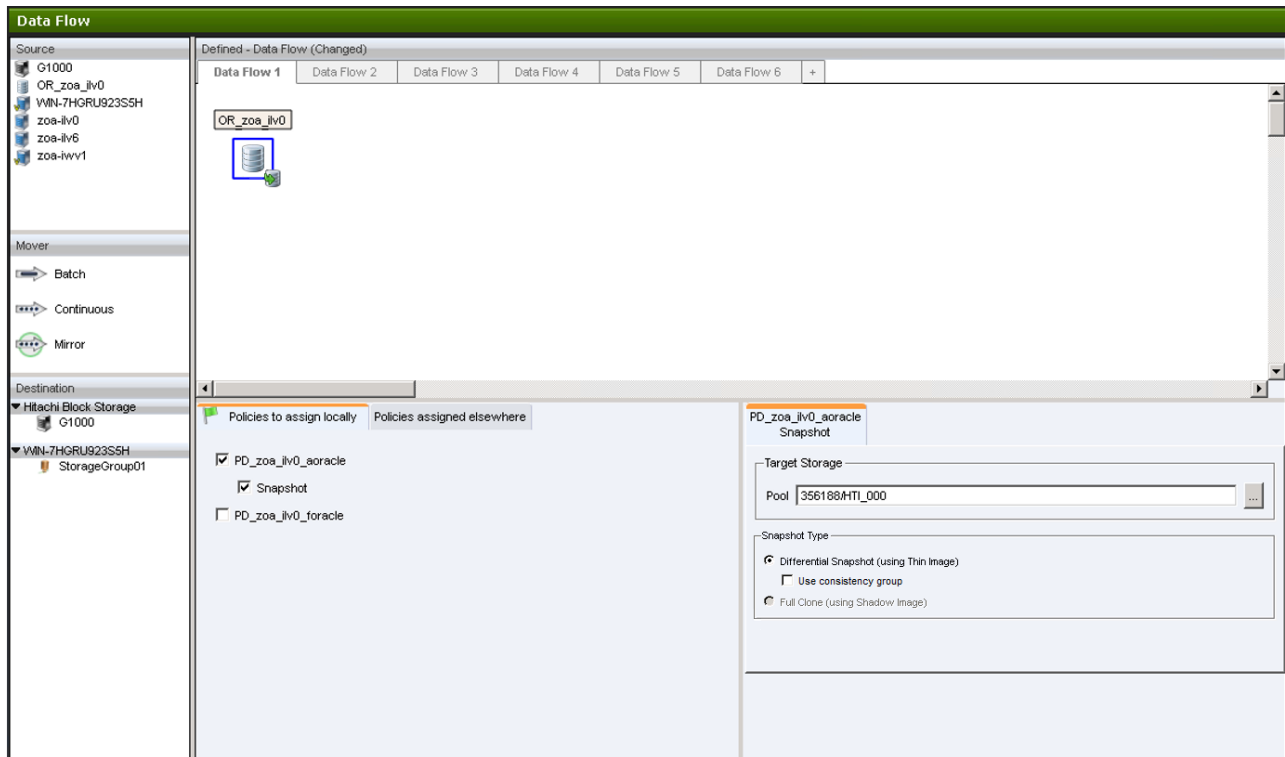


Figure 3-84 Oracle Data Flow

3. In the workspace, select the *OR_zoa_ilv0* icon to display a list of defined policies in the **Policies to assign locally** tab below the workspace. To see more information about a policy, hover the cursor over a policy.
4. Check the *Snapshot* policy (the policy defined above) to assign that policy.
5. Select the *PD_zoa_ilv0_aoracle Snapshot* tab.
6. Within **Target Storage**, select a **Pool** using the [...] button, the **Select VSP Storage Pool** dialogue will appear.
7. Within the **Select VSP Storage Pool** dialog select the **VSP** node from the drop down (this node has to be the VSP node the database being snapshotted is present on), then select a suitable **Pool** from the list presented where the snapshots will be stored.
8. Select a **Snapshot Type** of **Differential Snapshot (using Thin Image)** for this example.

Distribution and activation

1. To save and compile in one operation, click the **Save and Compile** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
2. Click the **Distribute Rules** button after the policy is compiled and the rules will be distributed and the policy activated.

Hot Tip: The actions that are performed by the orchestration framework as it executes a backup policy are captured by the [Log Manager](#). Detailed logs and their attachments can be exported as a text file (refer to the section [Exporting logs](#)).

Hot Tip: Details of the hardware resources (LDEVs) used for a particular replication or snapshot can be found via the **Resources...** menu option on the [Managed Storage tab](#) of the Storage screen.

Mount

Mounting can be performed from the **Storage – Managed Storage** or **Restore – Browse Backups and Snapshot** screens (Figure 3-85), by right-clicking on the required snapshot and selecting **Mount** from the menu.

1. Search the **Browse Backup and Snapshots** tab, results of all snapshots meeting the search criteria will be displayed.
2. Expand the Tree for the policy *PD_zoa_ilv0_aoracle*.
3. Select a snapshot within this policy, right-click it and select **Mount**

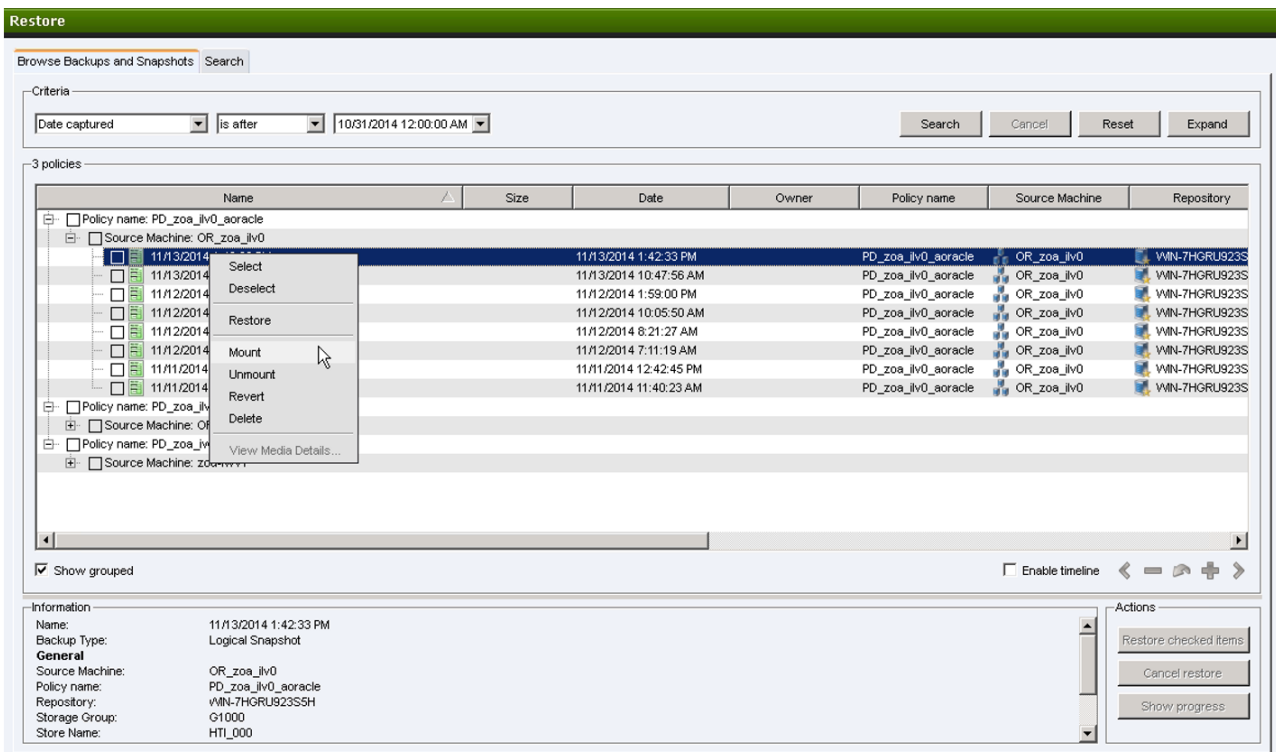


Figure 3-85 Mounting a Snapshot from the Restore Screen

4. The **Mount Snapshot Set** dialog (Figure 3-86) is displayed. Select from the **Mount host** drop down where this snapshot is to be mounted and click the

Mount at original location radio button, in this example we are mounting to 'zoa-illv6'. For ASM snapshots, selecting **Mount in directory** will mount the snapshot on the ASM instance, a path can't be specified.

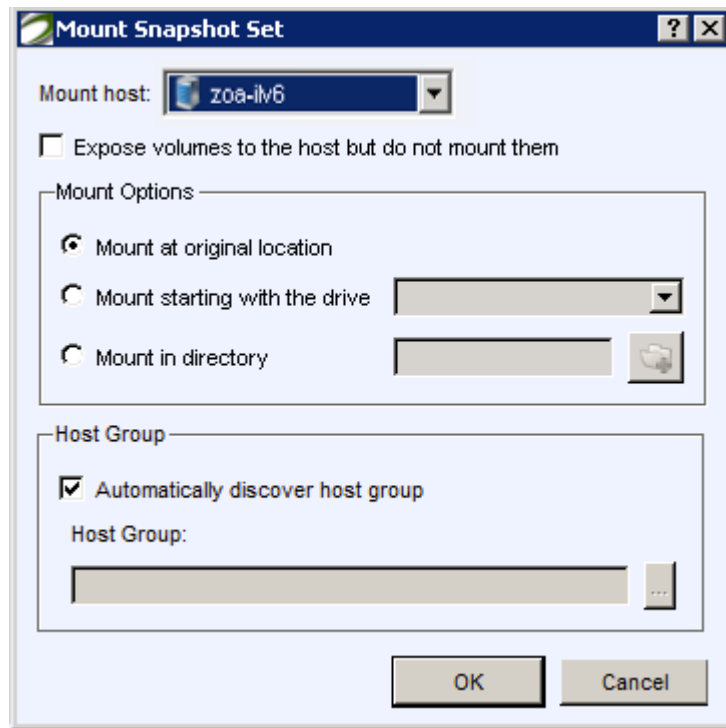


Figure 3-86 Specifying the Mount Location

Revert

Reverting can be performed from the **Storage – Managed Storage** or **Restore – Browse Backups and Snapshot** screens (Figure 3-85), by right-clicking on the required snapshot and selecting **Revert** from the menu.



WARNING: Revert will overwrite the original database and destroy all data in that database as a result.

1. The **Revert Warning and Confirmation** dialog (Figure 3-87) is displayed. To ensure the user does not accidentally perform a revert, the text '**REVERT**' must be typed in uppercase prior to clicking **OK**. The database will be shutdown prior to the reversion. Once reversion is complete the database must be restarted manually by the database administrator, who will need to choose how to recover the database (point-in-time, last known point etc.).



Figure 3-87 Restoring a Snapshot from the Restore Screen

Managing alert notifications

Although Data Instance Director supports emailed alerting, SNMP and MIB are not supported. If your site needs to receive alert notifications based on real-time or audit log events, then you can create a custom configuration file with instructions for which script/program Data Instance Director should run, and the arguments to pass to it. For instructions, see Chapter 6, [Customizing Alert Notifications](#).

User interface

This chapter introduces the Data Instance Director desktop.

- ☐ [Node Manager Screen](#)
- ☐ [Policy Definition Screen](#)
- ☐ [Schedule Manager Dialog](#)
- ☐ [Data Flow Screen](#)
- ☐ [Compiling and Distributing Rules Dialog](#)
- ☐ [Monitor Screen](#)
- ☐ [Log Manager Screen](#)
- ☐ [Notification Management Dialog](#)
- ☐ [Reports Screen](#)
- ☐ [Storage Screen](#)
- ☐ [Restore Screen](#)

Node Manager screen

Using the Node Manager, located on the Data Instance Director master node, you can control all Data Instance Director nodes.



To access the **Node Manager** screen, select the **Node Manager** icon in the Navigation bar.

Screen layout

The **Node Manager** screen is split into sections as shown in Figure 4-1.

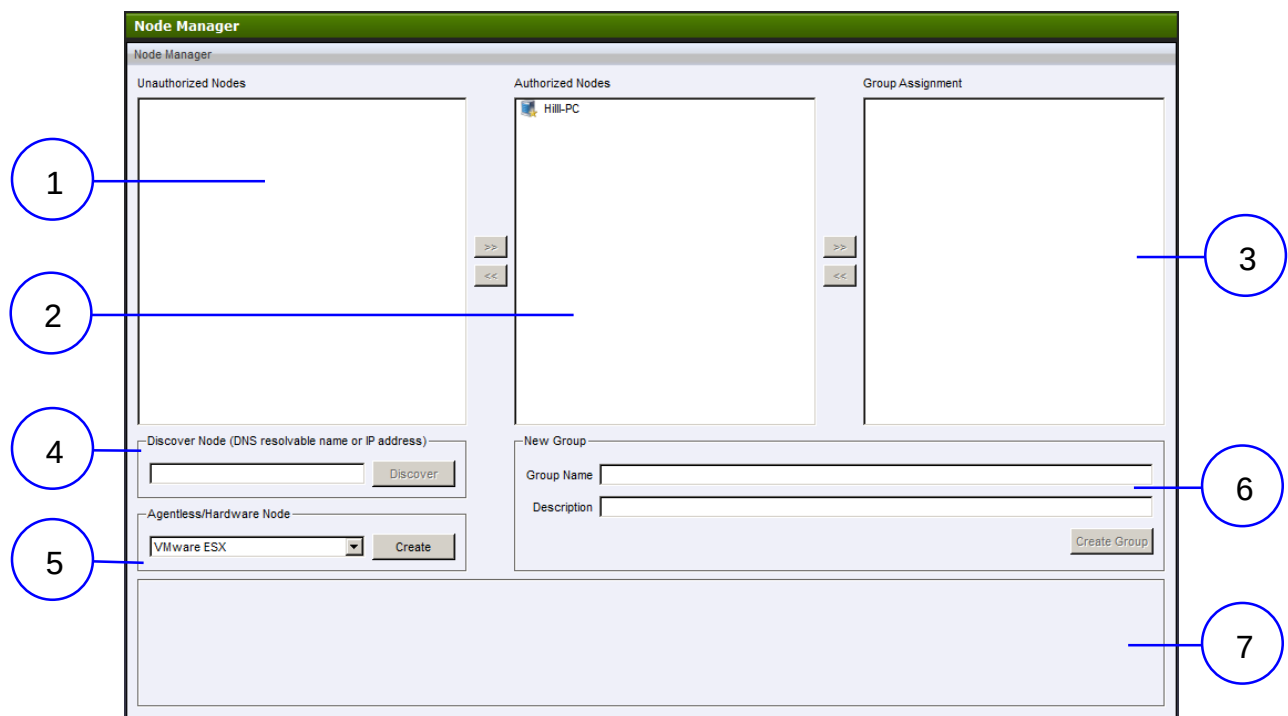


Figure 4-1 Node Manager

1

Unauthorized Nodes

The nodes on which Data Instance Director has been installed are automatically detected and displayed in the list of Unauthorized Nodes. Nodes in this list are unable to interact with other Data Instance Director nodes for security reasons and must be authorized in order to be used in policies.

The node names used in this list are the names assigned to the nodes when Data Instance Director was installed on them.

2

Authorized Nodes

The nodes that have been authorized by the Master node to perform Data Instance Director functions are displayed in the list of Authorized Nodes. The node names in this list are the names assigned to the nodes when Data Instance Director is installed on them.

3

Group Assignment

This section shows a tree, with all defined groups at the top level. Each group may be opened to show the nodes contained within that group.

To delete a group, right-click on the item and select Delete from the pop-up menu, or press Delete on the keyboard.

4

Discover Node

Nodes upon which Data Instance Director has been installed but not automatically discovered by the Master node can be added to the **Unauthorized Nodes** list by using the Discover Node tool.

Enter a DNS resolvable node name or the node's IP address in the **Discover Node** edit box and click **Discover** to contact that node. If the node is contacted successfully, then the node name appears in the **Unauthorized Nodes** list. The node name is defined when the Data Instance Director software is installed.

5

Agentless / Hardware Node

These controls allow the addition of agentless nodes, including CIFS and Tape nodes. See the section [Creating Agentless/Hardware Nodes](#).

6

New Group

You can create a new group in the New Groups section.

Enter a name for the new group, plus a short description, and then click **Create Group**. After the group is created successfully, then the group name is displayed in the Group Assignment section.

7

Node Attributes

This section provides information regarding the currently selected node. It displays the type of machine, its network addresses, and the Data Instance Director capabilities that were selected for it during installation.

Node icons

The following node icons and indicators are used in the lists of unauthorized and authorized nodes:

Node Types	
Icons that provide an at-a-glance indication of the function of a node.	
	Workstation Node - A node with the Data Instance Director services installed that is not a laptop and is not running a server operating system.
	Server Node - A node running a server operating system that also has the Data Instance Director services installed.
	Laptop Node - A laptop with the Data Instance Director services installed. This indicates the node is probably a client workstation and has unreliable uptime.
	CIFS Node - An agentless node containing network-shared data, and accessed through a separate proxy node. See the section Creating Agentless/Hardware Nodes .
	VMware ESX Node - An agentless node containing virtual machines, and accessed through a separate proxy node. See the section Creating Agentless/Hardware Nodes .
	Hitachi NAS Node - An agentless node containing Hitachi NAS storage, and accessed through a separate proxy node. See the section Creating Agentless/Hardware Nodes .
	Tape Library - An agentless node representing a Tape Library, and controlled using a media manager, with a device manager handling communications with the device. See the section Creating Agentless/Hardware Nodes .
	Exchange Node - An agentless node containing Exchange mailboxes, and accessed through a separate proxy node. See the section Creating Agentless/Hardware Nodes .
	Azure Node - An agentless node representing Microsoft Azure. See the section Creating Agentless/Hardware Nodes .
	HCP Node - An agentless node representing a Hitachi Content Platform (HCP). See the section Creating Agentless/Hardware Nodes .
	VSP Node - An agentless node representing a Hitachi Virtual Storage Platform (VSP). See the section Creating Agentless/Hardware Nodes .

	Disaster Recovery Node - This node is running inside the Data Instance Director Disaster Recovery environment. It cannot be involved in data flow topologies, but can have data restored to it.
	Oracle Node – Represents a cluster of servers running as an Oracle RAC that accesses databases on shared disks.
	Server Group - Represents a group of server nodes that can be treated identically when assigning policies or monitoring activities. See the sections Node Manager Screen and Monitoring Node Groups .
	Workstation Group – Represents a group of workstation nodes that can be treated identically when assigning policies or monitoring activities. See the sections Node Manager Screen and Monitoring Node Groups .
	Mixed Node Group - Represents a mixed group of nodes (laptops, workstations and servers) that can be treated identically when assigning policies or monitoring activities. See the sections Node Manager Screen and Monitoring Node Groups .
Status Indicators Small graphical indicators that provide visual cues as to a particular node's status	
	Master Node - A small star overlay indicates that this node is the master node.
	Virtual Node – An overlay of the letter 'V' indicates the node is being hosted on a virtual machine. Currently Data Instance Director can detect installations running in VMware, Hyper-V, VirtualPC and VirtualBox virtual environments.
	Snapshotted Node – An overlay of a storage device with a green arrow indicates that the node has a policy containing a snapshot operation applied to it.
	Inactive Node – An overlay of a red 'X' indicates that the node is not running the Data Instance Director hub process, or it is not currently attached to the network. The master node is unable to connect to the node. It is possible for it to be displayed alongside the <i>virtual</i> indicator above.
	Unknown Node - The master node is unable to determine any node details, and cannot connect to it.

Node attributes

When a node is selected either in the **Unauthorized Nodes** list or in the **Authorized Nodes** list, the details of that node are displayed in the **Node Attributes** pane at the bottom of the **Node Manager** screen.

Item	Description
Name	This name is assigned to the node during Data Instance Director installation.
Address(es)	A combo box containing all the IP addresses in use on the node. This is not displayed for agentless nodes.
Machine Type	The type of machine on which the node is running. This value is either <i>Laptop</i> , <i>Virtual</i> , <i>Server/Workstation</i> , or <i>Agentless/Hardware Node</i> .
Machine Role	Indicates whether the machine is a workstation or a server. This is not displayed for agentless nodes.
OS Type	The name of the operating system on the node. This is not displayed for agentless nodes.
Authorized	The checkbox is checked if the node has been authorized by the master as a member of the network.
Software Version	The version number of the Data Instance Director software running on the node. This is not displayed for agentless nodes, as they are not running the Data Instance Director software.
Capabilities Available	Shows the Data Instance Director functions that the node is permitted to perform: Data Source – May log file system operations for backup to a repository, perform blocking and/or tracking. Replication – May receive replication and replay data in order to function as a mirror. Repository – The node may backup and version data. Any Data Instance Director node may act as a proxy to an agentless node.
Capabilities Used	Shows the Data Instance Director functions that the node is currently performing in any distributed policies and data flows. The values are the same as used for Capabilities Available.

Authorizing nodes

Nodes must be authorized before they can be configured to perform any Data Instance Director functions or take an active part within a Data Instance Director network.

1. Select one or more active nodes from the **Unauthorized Nodes** list.
2. Click the node authorization button  located between the **Unauthorized Nodes** list and **Authorized Nodes** list.

If an attempt is made to authorize an inactive or unknown node, or if the master node fails to communicate with the node, an error message is displayed (Figure 4-2) and the node remains unauthorized.

When authorizing agentless/hardware nodes, Data Instance Director will not authorize if the remote machine is inaccessible, if the credentials provided are invalid, or – for CIFS nodes – if the shares specified cannot be accessed.

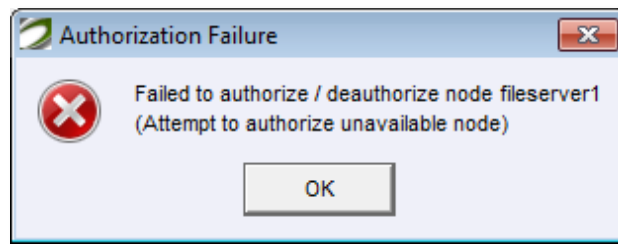


Figure 4-2 Authorization Failure

Deauthorizing nodes

When a node is no longer required to be authorized, it can be removed from the **Authorized Nodes** list. A node can only be removed if it is not part of an active data flow and is not running the most recently distributed rules.

1. Select one or more nodes in the **Authorized Nodes** list.
2. Click the node deauthorize button  located between the **Unauthorized Nodes** list and **Authorized Nodes** list.

If an attempt is made to deauthorize any node that is implementing rules in the currently active topology, an error message is displayed (Figure 4-3) for each such node and the node remains authorized.

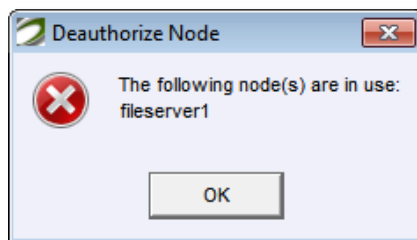


Figure 4-3 Deauthorization Failure

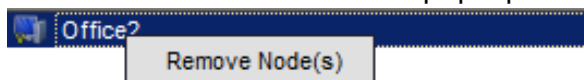


Note: Before deauthorizing nodes that are no longer required, they must be deleted from the current data flow definition first, then the new rules must be distributed.

Removing nodes

When a node is no longer required to be listed, it can be removed from the **Unauthorized Nodes** list.

1. In the **Unauthorized Nodes** list, right-click on the node that you would like to remove.
2. Select Remove Item from the pop-up menu:



Note: If the node is still running the Data Instance Director hub process and is configured to use the current master node, then the node will re-appear in the unauthorized nodes as it periodically reconnects to the master node. Data Instance Director should be uninstalled from the node in order to stop this periodic reconnection.

Creating Agentless/Hardware nodes

Nodes can be created to represent systems not explicitly managed by Data Instance Director. Data from these systems can then be integrated into the usual Data Instance Director workflow.

To do this, select the type of agentless node required from the combo box in the **Agentless / Hardware node** section, and then press **Create**. For details about creating an agentless node see the following sections:

[CIFS Agentless Node Creation](#)

[VMware ESX Agentless Node Creation](#)

[V2I Agentless Node Creation](#)

[Tape Library Agentless Node Creation](#)

[Exchange Agentless Node Creation](#)

[HCP/Azure Agentless Node Creation](#)

[Hitachi Block Storage Node Creation](#)

[Hitachi File Storage Node Creation](#)

After an agentless node has been created, it can be authorized in the usual way. Authorization will fail if the supplied agentless configuration is invalid. To alter the configuration, right-click on any agentless node and select **Edit Node**. Changes will come into effect from the next time rules are distributed.

Policy definition screen

The **Policy Definition** screen is used to define a dataset that is to have an operation performed on it by Data Instance Director nodes.

This screen is accessed by clicking the **Policy Definition** icon  on the **Navigation** bar.

Screen layout

The **Policy Definition** screen is split into five sections, as shown in (Figure 4-4).

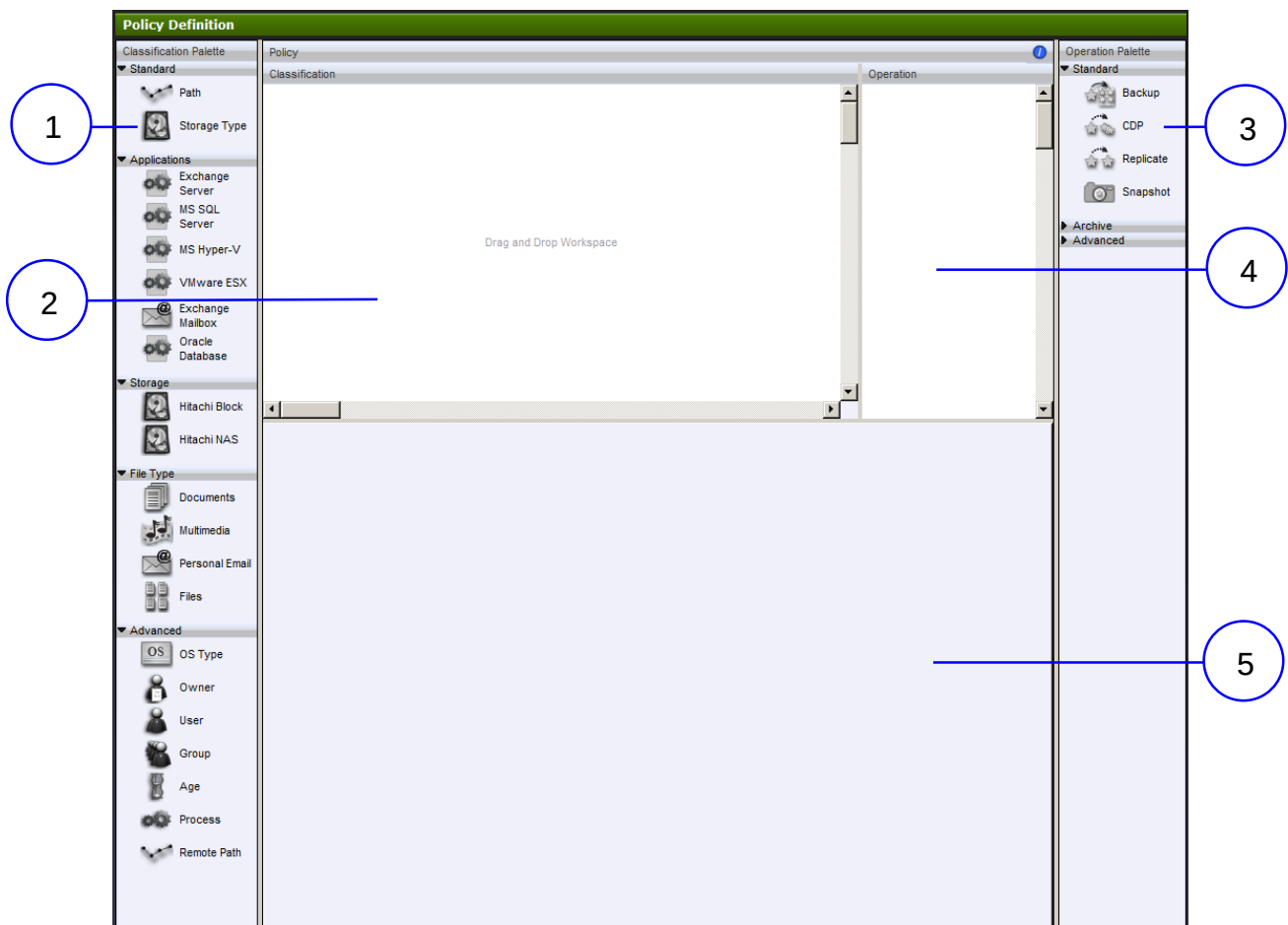


Figure 4-4 Policy Definition

1

Classification Palette

This palette lists the data classification types, all classifications are sorted into categories, these categories are: *Standard*, *Applications*, *File Type* and *Advanced*. Each type defines a set of criteria for selecting data on a node's file system. Click on the category heading to expand or collapse it (Figure 4-5).

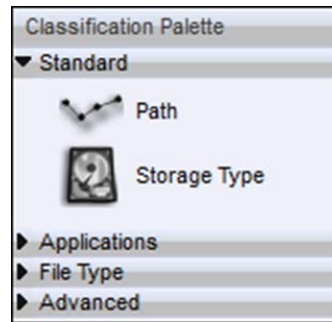


Figure 4-5 The Classification Palette

For example, the Path type may be used to select data by its location, or the User type may be used to select data accessed by one or more users. The data classification items and their attributes are defined later in this section.

2

Classification Workspace

This workspace is where the data classification for a policy is built. Classification items are dragged from the **Classification Palette** into this area. Items may be removed from the workspace by selecting the item and pressing the Delete key, or by accessing a pop-up menu with the right mouse button and selecting **Delete Item**.

3

Operation Palette

This palette lists data operation types, all operations are sorted into categories, these categories are: *Standard*, *Archive* and *Advanced*. Click on the category heading to expand or collapse the categories

Each type represents an action that may be performed on data that matches the classification. For example, Block may be used to prevent the file system from performing some or all operations on data in a defined directory.

4

Operation Workspace

This workspace is where the data operation for a policy is built. Data operation items are dragged from the **Operation Palette** into this area. Items may be removed from the workspace by selecting the item and

pressing the Delete key, or by accessing a pop-up menu with the right mouse button selecting Delete Item.

5

Attributes Tab

When the user selects a data classification item in the **Classification** workspace, or a data operation item in the Operation workspace, the item's attributes are displayed in the **Attributes** tab at the bottom of the screen. You can modify these attributes as needed. For example, if the Path item is dragged to the workspace, then the file system paths are displayed in the **Attributes** tab, and can be modified as needed.

If you select an item in the **Classification Palette**, or in the **Operation Palette**, the **Preview** tab displays a preview of the attributes which can be modified if the item is added to a workspace.

Defining a policy

A policy consists of two parts:

- One or more *Data Classification* items, which define what data is to be filtered.
- One or more *Operation* items, which define what action is to be performed on that data.

Data classification items

Select a *Data classification item* by clicking on an item in the **Classification Palette**, and then dragging it into the **Classification** workspace. The first item in the policy can be dropped anywhere in the workspace, although it will always be placed in the top left of the area. You can define the exact characteristics of an item by selecting that item in the **Classification** workspace, and then entering those characteristics in the **Attributes** tab (Figure 4-6).



Figure 4-6 Path Data Classification Attributes

Using AND and OR

Data classifications can be visualized as rows of sensors, placed over a conveyor belt, looking for the data of interest. If all the sensors on the same row detect data they recognize, as it passes beneath, then that data is subjected to all the operations defined in the adjacent **Operations** workspace.

By placing one row of detectors beneath another, it is possible to classify various types of data passing on the conveyor. The graphical representation of data classification in Data Instance Director mimics this idea in plan view; each group of sensors being represented by a horizontal row of data classification icons. The various arrays of detectors are placed one beneath the next, while the data flows downwards beneath them.

When adding multiple data classification items to a row, AND icons are automatically placed between each item (this indicates that each item in that row is logically conjunct to form the detection condition). The order in which classification items are placed in a row has no effect on their meaning.

When adding multiple sensor rows, the first item on each row is joined to the first item on the row above and below with an OR icon (this indicates that each row is logically inclusively disjunct to form the detection conditions).

By creating multiple rows a composite policy can be created. The vertical order in which rows occur has no effect on the overall meaning.

As each classification item is dragged into the **Classification** workspace (Figure 4-7 and Figure 4-8), the permitted drop locations for that item are denoted with yellow boxes.



Figure 4-7 Creating a Policy with AND and OR

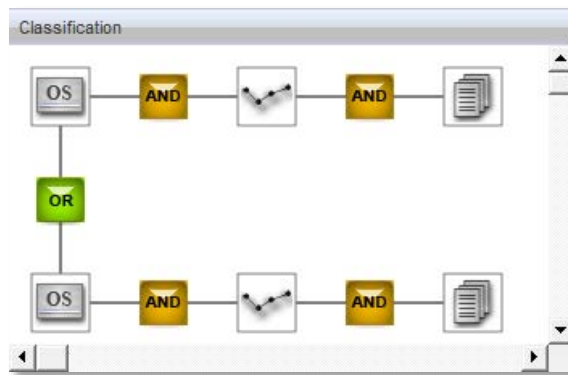


Figure 4-8 Joining Multiple Lines with OR



Note: Multiple classifications of the same type are not permitted on the same horizontal line.

Operation items

You can select data operation items by clicking on an item in the **Operation Palette** and dragging it into the **Operation** workspace (Figure 4-9). The first item in the policy may be dropped anywhere in the workspace, but it will always be placed at the top left of the area. You can define the exact characteristics of an item by selecting that item in the **Operation** workspace, and then entering those characteristics in the **Attributes** tab.

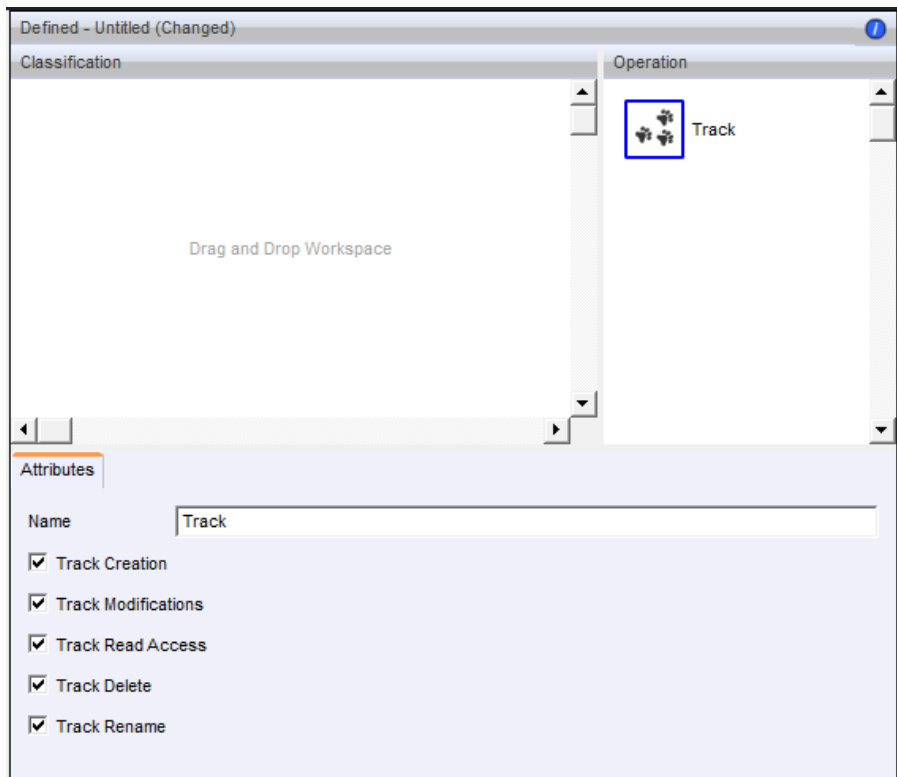


Figure 4-9 Adding an Operation Item

Multiple operations may be combined in a policy. When dragging an additional item into the **Operation** workspace, a yellow box shows where that item may be dropped. A typical example of multiple operations is a policy configured to both Track and Replicate the contents of a directory (Figure 4-10). The order in which operations are placed in the workspace has no effect on its meaning.

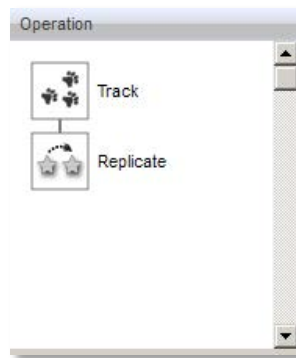


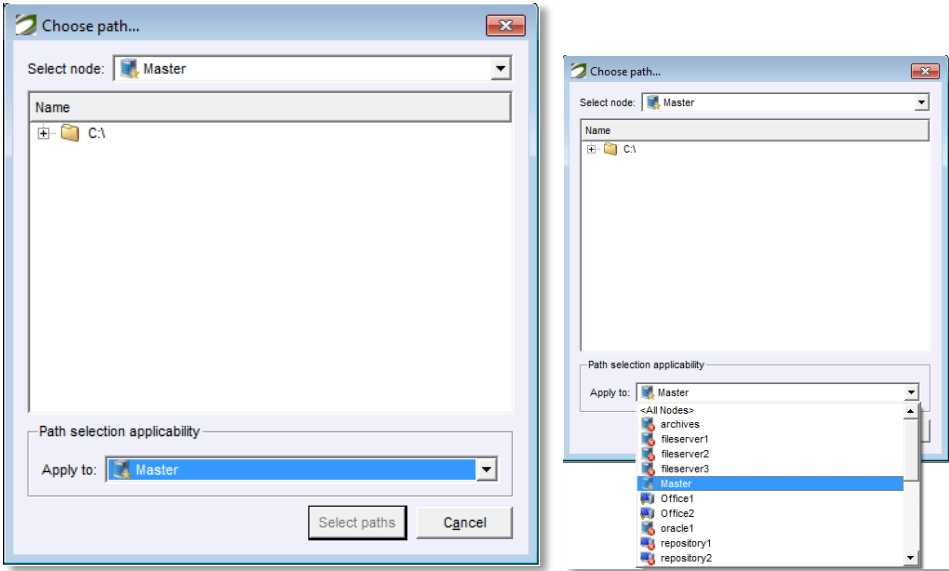
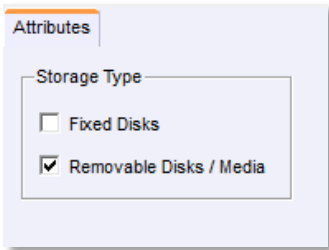
Figure 4-10 Multiple operations

A policy description may be defined by clicking on the information icon  above the **Operation** workspace. A brief description of the policy can be entered in the **Description** tab below the workspaces. This description can be viewed when saving and loading policies, and when selecting policies for use in the **Data Flow** screen.

Data classification items and attributes

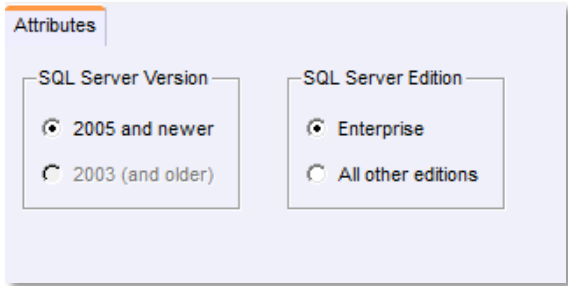
Standard

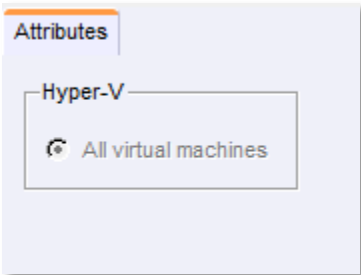
Classification	Attributes
 Path	Attributes Enter paths (one per line): Enter paths to be excluded (one per line): The Path classification allows you to define the path or paths for directories that contain files to be filtered. The targeted directory paths are entered in the Enter paths list box with one path per line. All files in each directory will be filtered. Files themselves cannot be entered – only directories. To specify files use the File Type classifiers described below in conjunction with this classifier. If a directory to be filtered contains one or more sub-directories that are not required, then enter the paths of these sub-directories in the Enter paths to be excluded list box with one path per line. It is possible to use wildcard characters (*) within each path entry. Only one wildcard character is permissible within each path element. That is, C:*\hello is valid, whereas C:*document*\hello is not valid. It is possible to use variables within a Path classification. See the section on Macros below this table for details. Existing paths can be chosen by clicking the folder icon  to the lower right side of the paths to be included or excluded list boxes. The Select Node and Apply To combo boxes in the Choose path dialog allow you to search a specific path within a node, and then to apply that path broadly to all nodes or apply it to one node only. When used in the context of the orchestration framework, Data Instance Director will discover the underlining hardware paths at runtime.

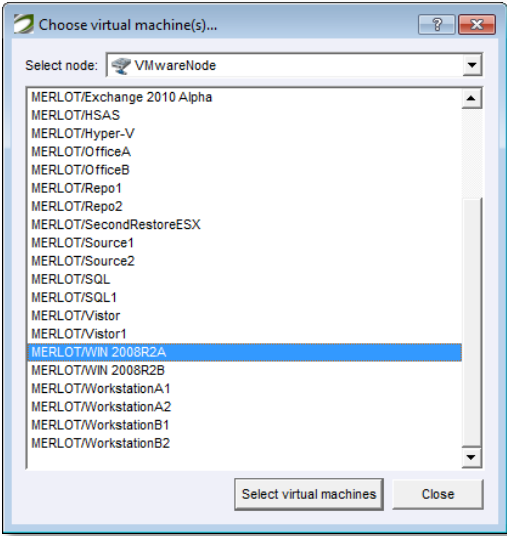
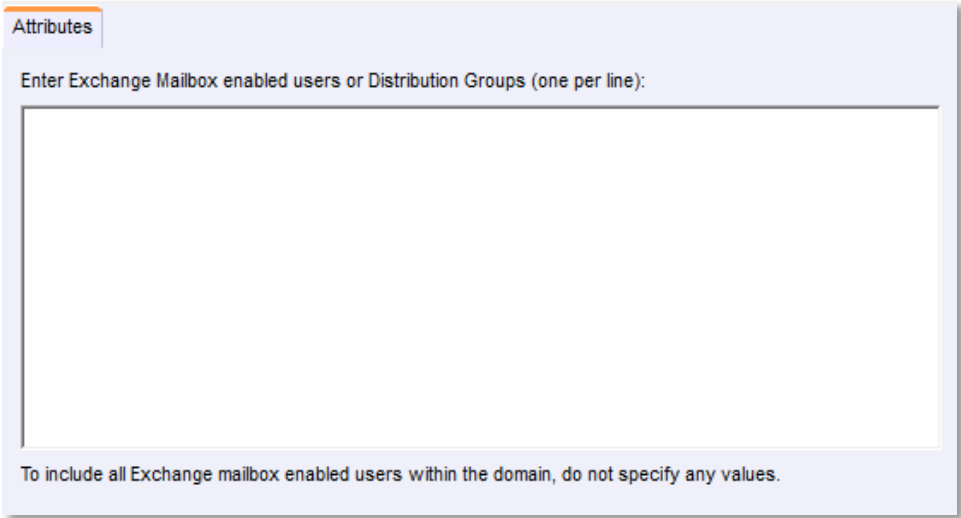
Classification	Attributes
	NOTE: For Hardware orchestration to work all the paths must exist on the same block hardware. 
 Storage Type	 The Storage Type classification is used to limit filtering to files that are located on fixed devices or to files that are located on removable disks. All floppy disks, CD/DVD drives, and drives that are attached to a USB or IEEE 1394 interface are considered removable. All drives connected to any other interface (SCSI, ATAPI, ATA, SAS, Fiber, RAID, etc.) are considered fixed.

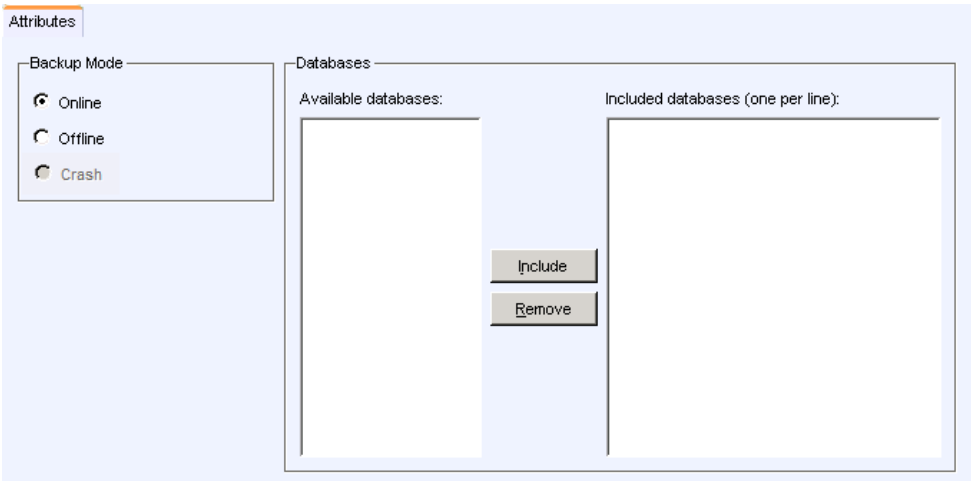
Applications

Classification	Attributes
 Exchange Server	Attributes Database Type ☐ All ☒ Active ☐ Passive The Exchange Server classification is used as a means of conveniently specifying the data associated with an Microsoft Exchange Server installation. This classifier can be used to ensure Exchange Server quiesces into a consistent state prior to a backup operation. For more information see the Applications section below. One of the following Database Types can be selected for Exchange 2007, 2010 and 2013 with .NET 4.0 installed (If earlier versions of Exchange are being used or .NET4.0 is not installed then All must be selected): • All: backs up/snapshots ALL DBs • Active: - o 2007/2010/2013 – backs up/snapshots ALL DBs - o 2010/2013 DAG – backs up/snapshots ACTIVE DBs. If there are no active DBs you get an empty backup and snapshots fail due to no volumes being selected • Passive: - o 2010/2013 – NO DBs are backup and snapshots fail due to no volumes being selected - o 2010/2013 DAG – backs up/snapshots PASSIVE DBs. If there are no passive DBs you get an empty backup and snapshots fail due to no volumes being selected

Classification	Attributes
	When used in the context of the orchestration framework, Data Instance Director will discover the underlining hardware paths at runtime. The following notes must also be observed when using this classification in conjunction with hardware orchestration backups: • All the paths must exist on the same block hardware. If Active or Passive is selected then the databases <u>must</u> exist on different LUNs.
 MS SQL Server	 The MS SQL Server classification is used as a means of conveniently specifying the data associated with a Microsoft SQL Server installation. The version and edition of the server must be provided (2003 and older editions are not currently supported). This classifier can be used to ensure SQL Server quiesces into a consistent state prior to a backup operation. For more information see the Application backup section below. When used in the context of the orchestration framework, Data Instance Director will discover the underlining hardware paths at runtime. NOTE: For Hardware orchestration to work all the paths must exist on the same block hardware.

Classification	Attributes
 MS Hyper-V	 The MS Hyper-V classification is used as a means of conveniently specifying the data associated with a Microsoft Hyper-V installation (selection of individual virtual machines is not currently supported). This classifier can be used to ensure Hyper-V quiesces into a consistent state prior to a backup operation. For more information see the Application section below.
 VMware ESX	 The VMware ESX classification is used as a means of conveniently specifying the virtual machines on a VMware ESX host. The virtual machines to be backed up must be selected by clicking the folder icon  to the lower right side of the virtual machines to be included or excluded list boxes. The Select Node combo box on the Choose virtual machine(s) dialog allows you to search a specific ESX host, and then choose which virtual machines to select for backup.

Classification	Attributes
	
 Exchange Mailbox	 The Exchange Mailbox classification is used to define which mailboxes are to be protected. Valid entries for this classification are: • Distribution Groups • Security Groups • Dynamic Distribution Groups • Individual email addresses (in the format name@domain.com)

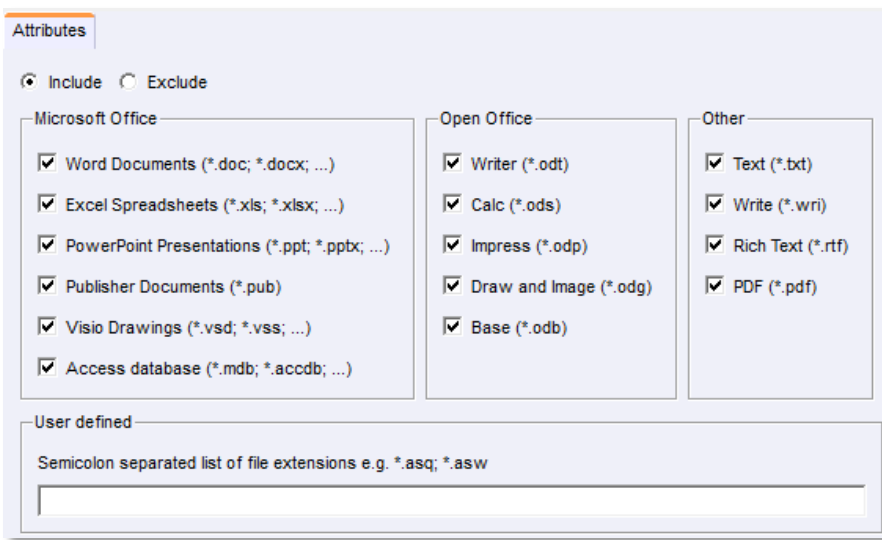
Classification	Attributes
	These must be entered one per line. If nothing is specified then all mailboxes will be backed up.
 Oracle Database	 The Oracle Database classification is used to define which databases are to be protected. A list of Available databases is presented from which those to be protected can be selected and moved to the Included databases list. The Backup Mode can be set to: • Online – Oracle is quiesced while an application consistent snapshot is made. • Offline – Oracle is taken offline while an application consistent snapshot is made. • Crash – Oracle remains running while a crash consistent snapshot is taken. This option is disabled in version 5.1. When used in the context of the orchestration framework, Data Instance Director will discover the underlining hardware paths at runtime. NOTE: For Hardware orchestration to work all the paths must exist on the same block hardware.

Storage

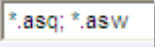
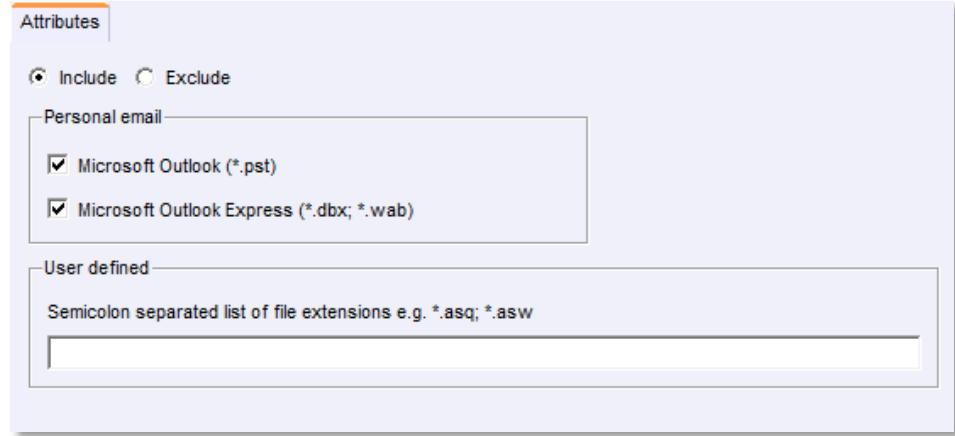
Classification	Attributes
 Hitachi Block	Attributes List LDEVs:   The Hitachi Block classification is used to define which LDEVs are to be protected. A list of LDEVs to be protected is presented. The LDEVs to be backed up must be selected by clicking the folder icon  to the lower right side of the list box. The Select Node combo box on the Select Block Storage dialog allows you to search a specific node, and then choose which block storage to select for backup.
 Hitachi NAS	Attributes Enter Paths:   The Hitachi NAS classification is used to define which paths on an Hitachi NAS are to be protected. A list of paths to be protected is presented. The paths to be backed up must be selected by clicking the folder icon  to the lower right side of the list box. The Select Node combo box on the Select Hitachi NAS path(s) dialog allows you to search a specific node, and then choose which file storage to select for backup.

Classification	Attributes
	If you want to specify paths to Hitachi NAS exports mounted on a client machine, use the Path classification in the Standard palette ✓ <i>NOTE – Data Instance Director version 5.1 has the following limitations for the Hitachi NAS classification:</i> • <i>Windows clients and CIFS the file system are not currently supported.</i> • <i>The root of the file system cannot be snapshotted. Root snapshots require using File System Snapshot (selected on the dataflow) and is not supported in this release.</i> ✓ <i>NOTE – Paths from different Hitachi NAS devices cannot be specified here since each icon relates to a single hardware device.</i>

File Type

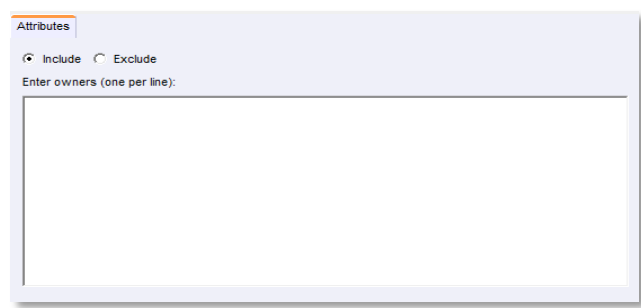
Classification	Attributes
 Documents	 The Documents classification is used to filter files created and modified by certain types of applications. Additional file types not covered by the checkboxes may be specified by entering them in the User defined text box, each preceded by an asterisk (*), and separated by a semicolon (;), as shown in this example: <code>*.asq; *.asw</code>

Classification	Attributes
	The specified file types can be included or excluded from processing by Data Instance Director by selecting the Include or Exclude radio buttons.

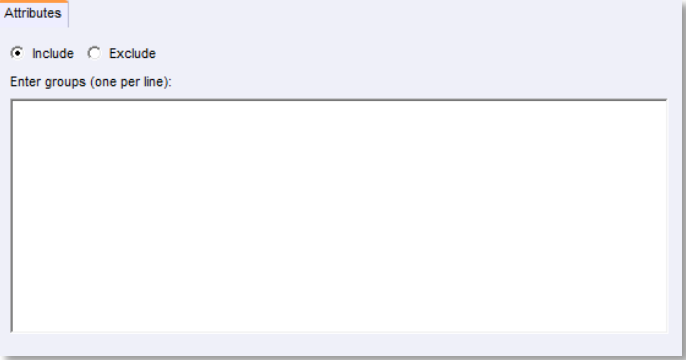
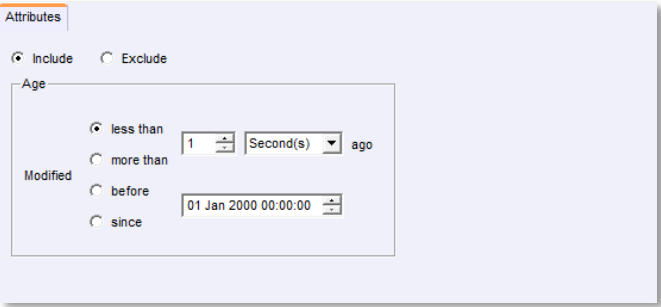
Classification	Attributes
 Multimedia	 The Multimedia classification is used to filter one or more types of multimedia files. The appropriate checkboxes should be checked for the file types to be filtered. Additional file types not covered by the checkboxes may be specified by entering them in the User defined edit box, each preceded by an asterisk (*), and separated by a semicolon (;), as shown in this example:  The specified file types can be included or excluded from processing by Data Instance Director by selecting the Include or Exclude radio buttons.
 Personal Email	 The Personal Email classification is used to filter the mail files for individual users. If the users are using Microsoft Outlook or Outlook Express, then the appropriate check box should be checked. For all other mail programs the file types created by that program should

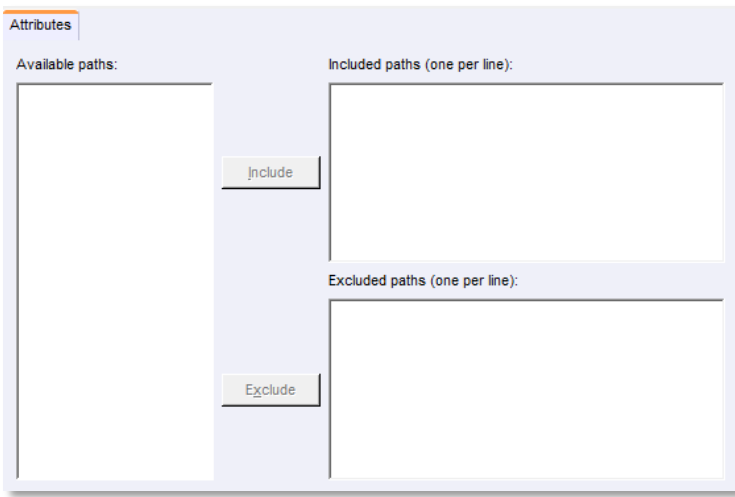
Classification	Attributes
	be entered in the User defined edit box, each preceded by an asterisk (*), and separated by a semicolon (;), as shown in this example: <pre>*.asq; *.asw</pre> The specified file types can be included or excluded from processing by Data Instance Director by selecting the Include or Exclude radio buttons.
Files	Attributes Enter file names (one per line): <pre>file.txt</pre> Enter file names to be excluded (one per line): <pre>not_this.bmp</pre> The Files classification allows you to define the name of files that should be filtered. All files, regardless of path, whose name matches the given name will be included. To specify paths use the Path classifier described above in conjunction with this classifier.

Advanced

Classification	Attributes
 OS Type	 The OS Type classification restricts the associated (conject) data classifications to those machines running one or more of the selected operating system types. This classifier is only intended to be used in conjunction with other classifiers. Used in isolation it has no effect.
 Owner	 The Owner classification is used to filter all file system activity on files owned by the listed user or users. Specify one user name per line. This classifier is only intended to be used in conjunction with other classifiers. Used in isolation it would cause all file activity for the specified owners to be filtered. The user names may be either local to the node or defined in a domain. A machine or domain name may be specified to limit matches to just that machine or name. Example: <pre>simon jonathon Administrator \\HeadOffice\auditor</pre> The specified owners can be included or excluded from processing by Data Instance Director by selecting the Include or Exclude radio buttons.

Classification	Attributes
 User	Attributes ☒ Include ☐ Exclude Enter users (one per line): The User classification is used to filter all file system activity performed by the listed user or users. Specify one user name per line. This classifier is only intended to be used in conjunction with other classifiers. Used in isolation it would cause all file activity for the specified users to be filtered. The user names may be either local to the node or defined in a domain. A machine or domain name may be specified to limit matches to just that machine or name. Example: <pre> simon jonathon Administrator \\HeadOffice\auditor </pre> The specified users can be included or excluded from processing by Data Instance Director by selecting the Include or Exclude radio buttons.

Classification	Attributes
 Group	 The Group classification is used to filter all file system activity performed by users that are a member of a listed group or groups. Specify one group name per line. This classifier is only intended to be used in conjunction with other classifiers. Used in isolation it would cause all file activity for the specified groups to be filtered. The group names may be either local to the node or defined in a domain. A machine or domain name may be specified to limit matches to just that machine or name. Example: <pre>Users Administrator \\HeadOffice\auditor</pre> The specified groups can be included or excluded from processing by Data Instance Director by selecting the Include or Exclude radio buttons.
 Age	 The Age classification is used to filter all activity on all files that were last modified more or less recently than a specified time duration in the past, or were modified before or after a specified date and time. This classifier is only intended to be used in conjunction with other classifiers. Used in isolation it would cause all file activity meeting the age specification to be filtered.

Classification	Attributes
	Files meeting the age specification can be included or excluded from processing by Data Instance Director by selecting the Include or Exclude radio buttons.
 Process	 The Process classification is used to filter all activity on files opened by a specified process or processes. Ensure the process extension is included when entering a process name. For instance, notepad.exe and notepad. Specify one process name per line. This classification is unsuitable for files opened by processes running on remote systems, since the process name will not be known by the local filesystem. The specified file types can be included or excluded from processing by Data Instance Director by selecting the Include or Exclude radio buttons.
 Remote Path	 The Remote Path classification is used instead of the Path classification to specify the shares and directories that should be accessed when retrieving data from CIFS nodes. Shares can be

Classification	Attributes
	include or excluded by selecting them in the Available paths list and clicking Include or Exclude. Subdirectories can be added to the end of the share names to further restrict what is copied.  Note: This classification is only applicable to CIFS nodes , and only the Backup and Version operations are effective when using it. Active paths do not have to be present on every CIFS node that the policy is applied to, and will be ignored when not present. This facilitates the creation of CIFS groups, if needed.

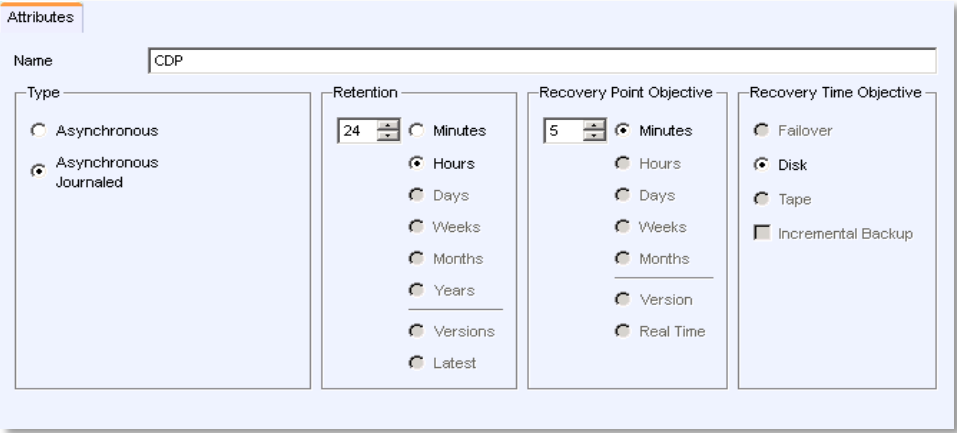
Data operation items and attributes

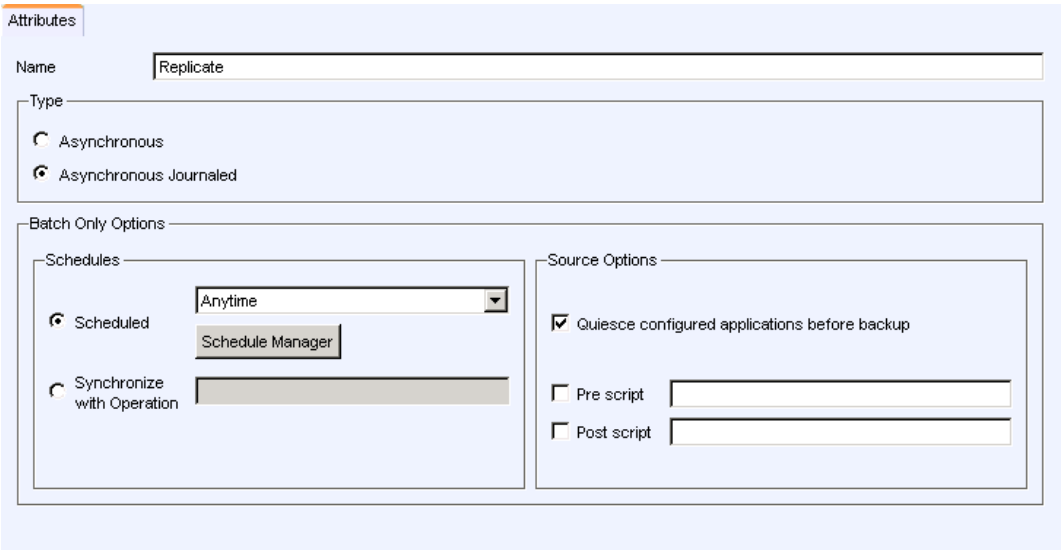


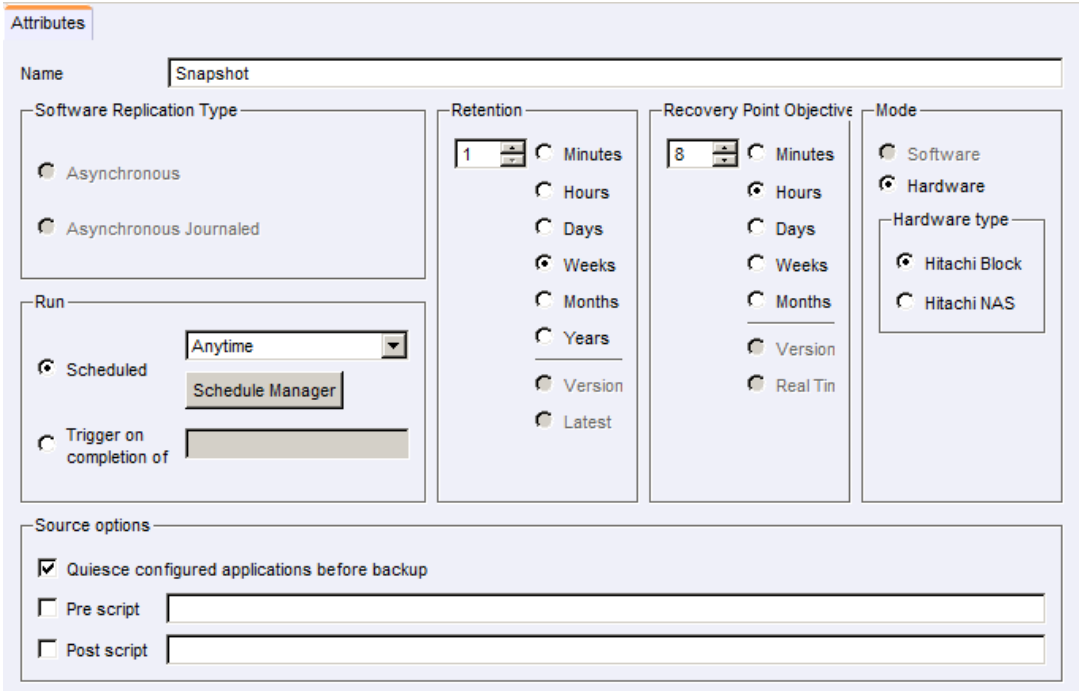
Note: When constructing data flows involving managed storage (e.g. VSP or HUS VM), only Replicate and Snapshot operations can be specified.

Standard

Operation	Attributes
 Backup	Attributes Name Backup Type ☐ Asynchronous ☒ Asynchronous Journalled Run ☒ Scheduled Anytime Schedule Manager ☐ Synchronize with Operation Source Options ☒ Quiesce configured applications before backup ☐ Pre script ☐ Post script Retention 6 ☐ Minutes ☐ Hours ☐ Days ☐ Weeks ☒ Months ☐ Years ☐ Versions ☐ Latest Recovery Point Objective 8 ☐ Minutes ☒ Hours ☐ Days ☐ Weeks ☐ Months ☐ Version ☐ Real Time Recovery Time Objective ☐ Failover ☒ Disk ☐ Tape ☐ Incremental Backup

Operation	Attributes
	created whenever the Recovery Point Objective time is reached, following the initial backup performed when the policy is activated. The Recovery Time Objective can be set to Disk or Tape. Selecting Tape allows the Incremental Backups option to be selected. If a backup policy having an RTO specified as Tape is assigned to a repository node (or vice-versa) then an error is displayed during rules compilation. The Schedules section contains the Schedule Manager button and the Start on Completion of operation edit box. <i>Start on completion of operation</i> is used when a policy contains more than one operation and it is necessary that the operations occur in a sequential order. See the Smart Repository Sync section for more detailed instructions on its use. The times, dates and/or windows within which backups are written to a repository or tape library can be controlled by clicking the Schedule Manager button to define a named schedule that is then selected from the Schedules combo box. The Source options controls provide options designed to ensure snapshot consistency. If the Quiesce configured applications before backup option is selected and any Exchange Server, MS SQL Server, VMware ESX, or MS Hyper-V classifiers are included in the policy, then these applications will be temporarily quiesced into a consistent state prior to the snapshot occurring. A Pre script and Post script can also be specified in order to ensure consistent state for other applications. The Application section below contains more information on all these options.
 CDP (Continuous Data Protection)	 The CDP operation makes an initial complete copy of the data (specified by the associated data classification) on a source node to a repository node. It then continuously records any changes made to that data and places markers in the repository at intervals determined by the Recovery Point

Operation	Attributes
	Objective. This mechanism enables very fine grain RPOs to be achieved, measured in minutes. The Name is used to identify the operation when the policy is used in a Data Flow diagram. The Type specifies how the file system operations on the matching data classifications are logged. The different options are discussed in the Data Transfer Modes section of Chapter 1. The Retention value can be set to a specified number of minutes or hours. The repository retains each snapshot of the policy for the specified time. Retention values of Days to Years, Versions and Latest are not applicable to continuous data protection operations. The Recovery Point Objective can be set to a specified number of minutes. The repository places markers in the stream of recorded changes that enables recovery to those points if required at a later time. The Recovery Time Objective can only be set to Disk, as the repository always maintains its database on disk.
 Replicate	 The Replicate operation is used to keep an identical copy of the latest data in the policy. Only the latest version of each file is retained. The replica can be stored either in a mirror or in a repository on the destination node. The Name is used to identify the operation when the policy is used in a Data Flow diagram. The Type specifies how the file system operations on the matching data classifications are logged. The different options are discussed in the Data Transfer Modes section of Chapter 1.

Operation	Attributes
	Replication can be performed in real time using a mirror mover or in batch mode using a batch mover. The Batch Only Options described below are only relevant when using a batch mover. The Schedules section contains the Schedule Manager button and a combo box that lists predefined and user defined schedules. The Source Options group contains options designed to ensure consistency of application files prior to mirroring. If the Quiesce configured applications before backup option is checked and any Exchange Server, MS SQL Server, or MS Hyper-V classifiers are included in the policy, then these applications will be temporarily quiesced into a consistent state prior to being mirrored. A pre script and post script can also be specified in order to ensure consistent state for other applications. The Application section provides more information about all of these options.
 Snapshot	 The Snapshot operation is used to trigger snapshots of entire volumes. Created snapshots can be accessed via the Restore screen where they can be mounted or used for reverting a volume to a previous state. The Name is used to identify the operation when the policy is used in the Data Flow diagram.

Operation	Attributes
	The Retention and Recovery Point Objective can be set to a specified period. These controls are used in conjunction with the Run controls. If no schedules are defined, then a snapshot is created on the source node whenever the Recovery Point Objective time is reached. The times and/or dates that these snapshots are or are not created can be controlled by using the Schedule Manager to define a schedule. This schedule can then be selected with the Schedules combo box. The Mode may only be set to: • Software – currently not available in this version. • Hardware – for snapshots of attached storage. If Hardware mode is selected then the Hardware type can be set to: • Hitachi Block – for VSP and HUS VM block storage • Hitachi NAS – for Hitachi NAS file storage The Source options group contains options designed to ensure snapshot consistency. A pre script and post script can be specified in order to ensure consistent state prior to the snapshot. The Application backup section contains more information about all of these options.  WARNING: Snapshot operations always apply to the whole selected volume. Any associated data classifications are ignored. HDID versions prior to 5.0 only supported Local Snapshot operations. These are not available in this release. If you have existing policies that contain Local Snapshot operations created using older HDID versions (pre version 5.0), then these will no longer work when upgrading to HDID version 5.0. Software snapshots will be re-instated in a later version. Snapshots will be automatically retired after the given retention period, indexed by the Metadata Store and can be mounted and reverted via the Restore screen. Local Snapshots created prior to upgrading to HDID version 5.0 will not be indexed by the Metadata Store or retired, but will continue to be accessible from Windows file explorer on the machine where they were created.

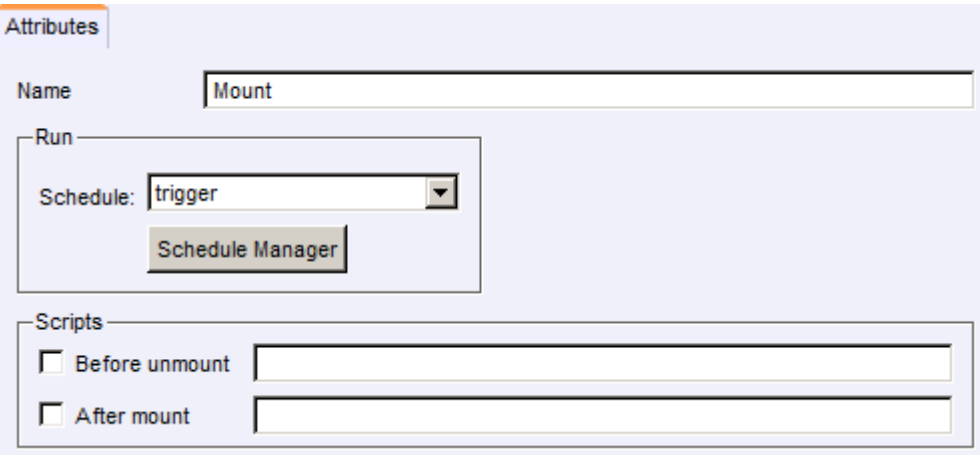
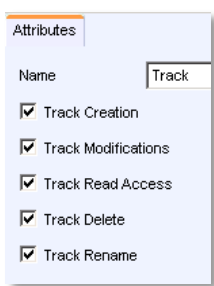
Archive

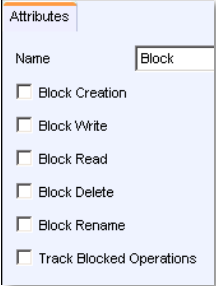
Operation	Attributes
 Version	Attributes Name Version Type ☐ Asynchronous ☒ Asynchronous Journalled Retention 3 Minutes Hours Days Weeks Months Years Versions Latest Recovery Point Objective 1 Minutes Hours Days Weeks Months Version Real Time Recovery Time Objective ☐ Failover ☒ Disk ☐ Tape ☐ Incremental Backup

Operation	Attributes
Tier	The Tier operation is used to send data to a Hitachi Content Platform (HCP) or Microsoft Azure. The Stub/Remove Source Item Options list box and associated <i>add</i> and <i>delete</i> buttons allow the conditions for tiering email bodies, email attachments, or files to be specified. These options can be specified by clicking the folder icon . The following dialog is displayed. Selecting Stub will cause items that meet the specified criteria to be moved to HCP or Azure; the moved item is replaced by a reference that points to the location on HCP or Azure where the original is located. Selecting Remove will remove any items from the source node that meet the specified criteria.

Operation	Attributes
	 Note: With applications such as Microsoft Notepad, the file's accessed time might not be modified when the file is opened. If you plan to stub files based on the Was not accessed in the last criterion, then you must ensure that the file access timestamp is correctly updated by modifying a registry key. For information, see Chapter 7, Troubleshooting HCP.

Advanced

Operation	Attributes
 Mount	 The Mount operation is used to automatically mount and unmount the destination of a Hitachi Block and File storage replication (snapshots cannot be used). The mount operation can have its own <i>trigger</i> Schedule that would be used when applied to a live backup. The operation can also be part of a synchronization group so that it can participate in a batch replication sequence. Scripts can be specified that run: • Before unmount - allows the script writer to put the application into a safe state before the volumes are unmounted from the mount host. • After mount - allows the script writer to invoke an action after the volumes have been mounted on the mount host, e.g. schedule a host based back-up. These scripts are run on the mount host.
 Track	 The Track operation is used to log various types of access to files specified in the associated classification.

Operation	Attributes
	The Name is used to identify the operation when the policy is used in a Data Flow diagram. The check boxes are used to specify the file system operations that are to be tracked for those files matching the data classifications. All matching activity is then logged and can be monitored in the Reports screen.
 Block	 The Block operation is used to prevent and optionally log various types of access to files specified in the associated classification. The Name may be used to identify the operation when the policy is used in a Data Flow diagram. The check boxes are used to specify the file system operations that are to be denied for those files matching the data classifications. If blocked operations are to be tracked, then check the Track blocked operations check box. All blocked activity is then logged and can be monitored in the Reports screen.  WARNING: Care should be taken not to block OS files and processes as this may make the system unstable or it may stop working completely. Use precise data classifications to minimize the risk of causing such problems.

Macros

There are occasions where it is desirable to define paths logically when constructing a policy, knowing that only the source machines that will receive the policy can actually determine where those paths should really be on the local system. For example, all the fixed drives or the Exchange directory, as these may differ from machine to machine.

Data Instance Director facilitates this requirement by allowing the use of macros when defining paths. These are given in the form `$(MACRO_NAME)` and are resolved during rules distribution upon each source machine. They are resolved by looking up the macro name in a `macro.cfg` file stored on each source machine in `C:\Program Files\Hitachi\HDID\db\config\`.

Entries within `macro.cfg` are given in the form:

```
<keyword>=<path>  
For instance:  
files=c:\files
```

To use this macro, `$(files)` would be entered in either the Include or Exclude list for a **Path** classification. Note that macros can be combined with normal text, so that `$(files)\subdirectory` would be a valid entry. If the macro is defined with a trailing slash then ensure it is not combined with *another* slash when combining it with plain text.



WARNING: Macro substitution is case sensitive. Ensure the case of macros in each `macro.cfg` file matches the case of macros when used to define a policy.

Multiple paths can be assigned to a single macro by delimiting each path with a semi-colon (;), as so: `<keyword>=<path_1>;<path_2>;<path_3>`

Application backup

Ordinarily, when a snapshot is performed, the state of the snapshot consists of the exact state of the file system at the point in time at which the snapshot was triggered. This can be problematic when the snapshot is eventually restored, as some applications that were running at the time may not, for example, have flushed all their in memory file buffers, and their data files on disk could therefore be in an inconsistent state.

To resolve this, it is possible to request certain applications to *quiesce* into a safe state prior to a snapshot being triggered, and then to restore their previous active state after the snapshot is triggered. This ensures their state during the snapshot is consistent and that they can run as expected if the snapshot is restored.

If an application cannot be quiesced then a snapshot of its files is said to be in a crash-consistent state, i.e. as if the application had crashed while running. Crash consistent snapshots may, depending on the application, be used to restore state to some point prior to the snapshot.

Exchange, MS SQL Server, and MS Hyper-V

Data Instance Director includes built-in classifications that enable **Exchange Server**, **MS SQL Server** and **MS Hyper-V** to quiesce into a consistent state prior to a Data Instance Director Backup being triggered. These are all described individually in the table above. To define a policy able to quiesce any of these applications, do the following:

1. Drag the application classification into the **Classification** workspace.
2. Select the relevant application edition within the classifications **Attributes** tab.
3. Drag a Backup operator into the **Operation** workspace.
4. Select the **Quiesce configured applications before backup** checkbox in the backup **Attributes** tab.

Now, when a backup is triggered in a workflow that makes use of this policy, the resulting snapshot will contain a quiesced version of the application's files.

Other applications

If other applications not mentioned in the section above need to put themselves into a certain state prior to a Backup or Local Snapshot operation, this can be done using the **Pre script** and **Post script** options.

Select the **Pre script** and **Post script** checkboxes within the **Attributes** tab for the Backup or Local Snapshot operator, and then enter a name of a script file. This script file must be present on the source node that implements the policy and must be placed inside the Data Instance Director scripts directory (C:\Program Files\Hitachi\HDID\scripts is the Windows default).

Saving policies

A newly created policy cannot be used until it is saved.

1. Click the Save button  on the toolbar,
or, select **File** → **Save As...** from the menu.
2. Enter a **Name** to be used to identify the policy in the **Save Policy as** dialog (Figure 4-11). Policy names can only use letters and /or numbers and the following symbols: () - _ + = ~ @ , and <space>.
3. Enter a short description of the policy in the **Description** edit box.
4. Click **Save**.

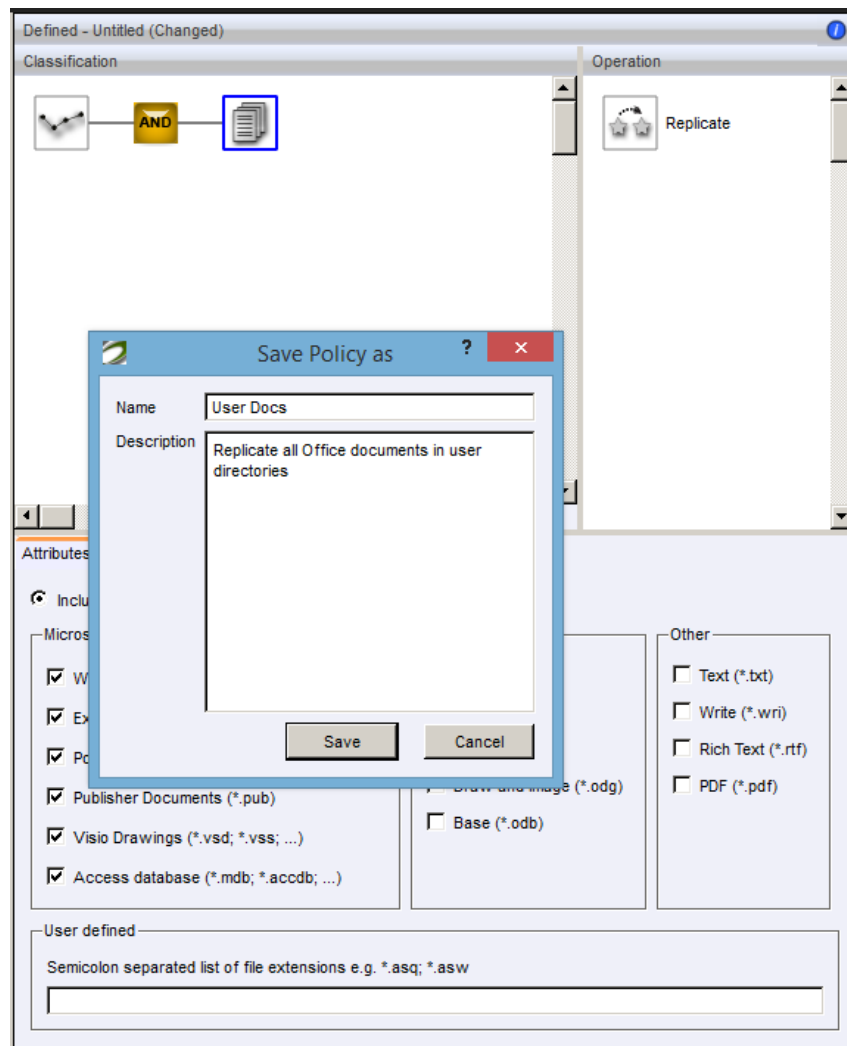


Figure 4-11 Naming and Saving a Policy

Loading and modifying existing policies

1. To load any existing policies, click the Load button  on the toolbar, or select **File** → **Open.....**. The **Open Policy** dialog is displayed (Figure 4-12).

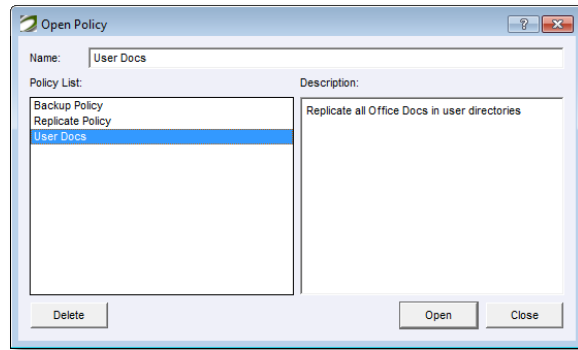


Figure 4-12 Opening an Existing Policy

All existing policies are shown in the **Policy List**. When a policy in the list is selected, the description of that policy is displayed in the **Description** text box.

2. Click **Open** to load the selected policy.
 - You can remove a policy by clicking on **Delete** as long as the policy is not in use in the current dataflow definition.
 - After a policy is loaded, you can modify it by changing the attributes for existing classification or operation items, or by adding or deleting classification or operation items.
 - When the required modifications have been made, keep the changes by clicking the Save button on the toolbar, or by selecting File → Save.

If the policy is in use in the current dataflow definition, the modifications made to the policy only take effect following recompilation and redistribution of the rules. If data operation items have been added or deleted, it may be necessary to first modify the dataflow definition to take account of these changes before the dataflow definition is recompiled and rules are distributed.



WARNING: If data operations are added to or removed from a policy that is in use in the current data flow definition, then that dataflow definition must be modified to add or remove information about how the data from that policy is passed about the network, for all routes originating at source nodes that implement the policy.

If a policy using the Replicate, Version, CDP or Backup data operations has any part of its data classification changed, then all destination nodes for that policy may need to be re-synchronized to include the changed scope of the policy.

Schedule Manager dialog

The Schedule Manager dialog (Figure 4-13) is used to specify when a Backup or Snapshot operation should run. Schedule Manager can be accessed from **Tools** → **Schedule Manager**, or by clicking the **Schedule Manager** button in the **Attributes** tab when configuring a Backup, Replicate or Snapshot operation.

Central to the operation of the Schedule Manager is an operation's Recovery Point Objective (RPO). This setting defines how regularly a backup should be triggered. An RPO of 12 hours tells the Schedule Manager to schedule a backup at that interval, while also obeying any other user-defined constraints imposed on the schedule. For example, a backup may only be allowed to occur between 12 a.m. and 6 a.m., and this restriction must be adhered to even if the RPO is only one hour. See the [Policy Definition](#) section earlier in this chapter for more details about RPOs.

The times a schedule allows a backup to occur depends on the constraints used to build the schedule. Three types of constraints are available:

- **Backup window** – A backup event can only occur during a backup window. By default, an implicit backup window is used that operates all day, every day. If a custom backup window element is created, then the default backup window is ignored.
- **Trigger** – This element is similar to a backup window, except that it mandates that a backup event must occur at the specified time or date, unless it occurs during an exclude window.
- **Exclude** – A backup event can never occur during an exclude window.

These elements can be used in any combination and in any amount within a single schedule. Essentially, a backup can only occur if one of the following is true:

- The backup has exceeded its RPO
- and a backup window is active
- and an exclude window is not active.

or

- A trigger is active
- and an exclude window is not active.

Defining a new schedule

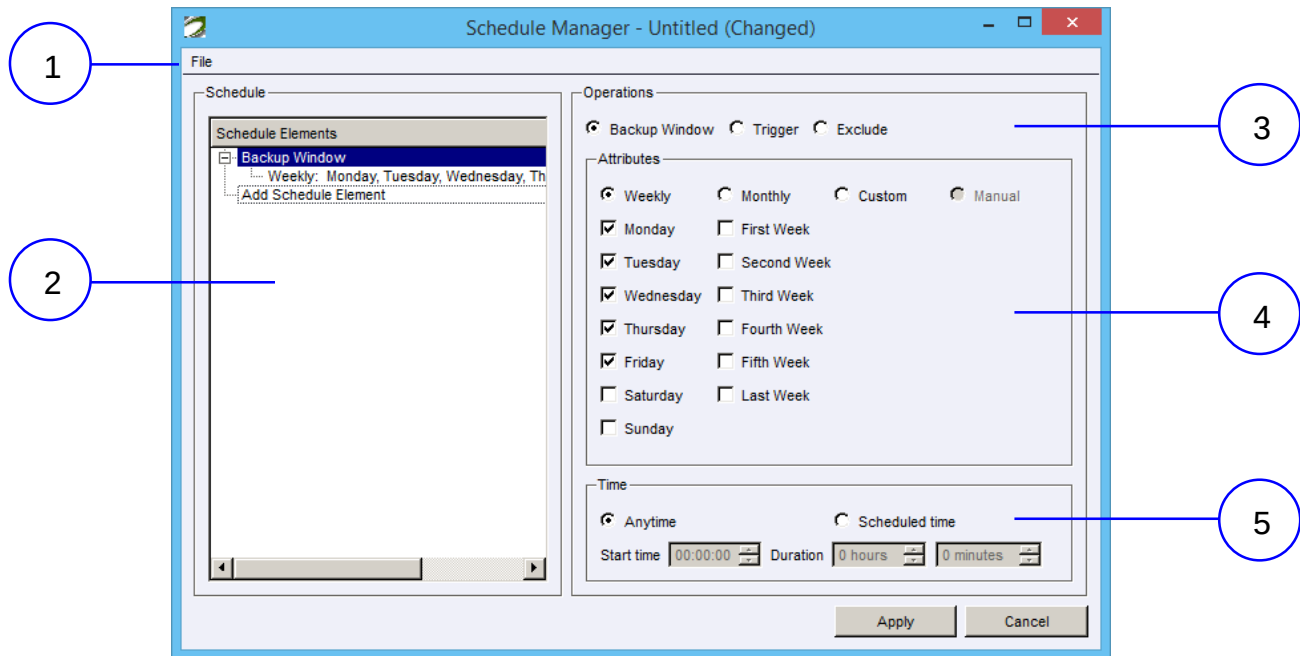


Figure 4-13 Schedule Manager

1

If the **Schedule Manager** is launched from the **Tools** → **Schedule Manager** main menu option, or by clicking the **Schedule Manager** button with the 'Anytime' schedule selected, then new (untitled) schedule is presented for creation. Otherwise, the currently selected schedule displayed in the **Schedules** combo box on the **Attributes** tab is presented for editing. You can also create a new schedule in **Schedule Manager** by selecting **File** → **New**.

2

A new schedule does not initially have any constraint elements and so all constraints to the right are disabled. Add a new element by double clicking the Add Schedule Element item in the **Schedule Elements** tree view on the left. You can select an element in the Schedule Elements tree view and press the Delete key to remove it.

3

By default, the new schedule element is opened with an **Operations** type of **Backup Window** and the **Attributes** set to **Weekly**, **Monday** through **Friday** at **Anytime**. You can select a different Operation, such as **Trigger** or **Exclude** by selecting the corresponding radio button.

4

The **Attributes** are set by first selecting the corresponding radio button (**Weekly**, **Monthly** or **Custom**), then by selecting the days, weeks and **Time**.

- **Weekly** – As displayed in the image above, the **Weekly** attribute is used to define schedules based on the day of the week. For example, you can create a schedule element that only occurs on weekends, or that never occurs on weekdays of the last week of the month. At least one day must be specified while weeks are optional (every week is assumed if none are selected).



Note: The weeks listed do not refer to calendar weeks. The First Week of a month always includes the days from the 1st until the 7th, the Second Week includes the days from the 8th to the 14th, and so on. The Last Week always refers to the 7 days prior to (and including) the last day of the month.

- **Monthly** – The Monthly attribute (Figure 4-14) is used to define schedules based on the day of the month. A monthly schedule element is applied on the same day or days each month. For instance, an element could be defined that triggers a backup on the last day of the month, or an element could be set to prevent backups from occurring on even numbered days. At least one day must be selected in order for a Monthly element to be valid. If the month in question does not contain the selected day (e.g. 29th Feb in a non-leap year) then the schedule element will not be applied.

Figure 4-14 Monthly Attribute

- **Custom** – The Custom attribute provides the greatest flexibility by allowing you to create a custom collection of dates on which a backup should or should not occur. A collection of dates in this context is called a *calendar*. For example, a calendar could be created that contains all public holidays for the next few years. More details on the Custom element are provided later in this section.
- **Manual** – If a **Trigger** operation is selected as a schedule element then the **Manual** option is also enabled. This attribute allows you to force a backup event to occur by executing a separate program using a schedule-specific keyword. To configure this attribute, enter a

`<trigger_keyword>` in the **Trigger Keyword** edit box. Then when you want to trigger the backup:

1. Open a command prompt.
2. Navigate to C:\Program Files\Hitachi\HDID\bin.
3. Run the following command:

```
triggerbackup.exe -t <trigger_keyword>
```

5

Schedule Manager also allows you to apply a **Time** setting (Figure 4-15) for a schedule element when configured using the **Weekly**, **Monthly**, or **Custom** attribute. By default, backups are set to occur at **Anytime** during the selected days. To set a window during which the backup may start, select the **Scheduled time** radio button, and specify the **Start time** and **Duration**.

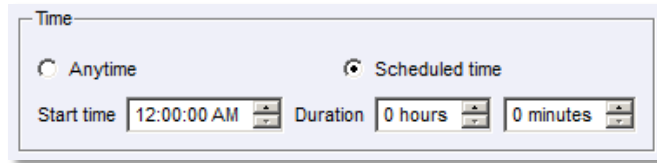


Figure 4-15 Time Setting



Note: If you provide a specific start time but with a duration of zero hours and zero minutes, the event can only occur at the start time given.

Modifying and deleting schedules

You can create, open, save, modify and delete schedules via the **File** menu displayed in the top left corner of the **Schedule Manager** dialog.

Opening and modifying a schedule

1. Open Schedule Manager.
2. Select **File** → **Open**. The **Open Schedule** dialog is displayed (Figure 4-16). The **Schedule List** includes all schedules that have been created.

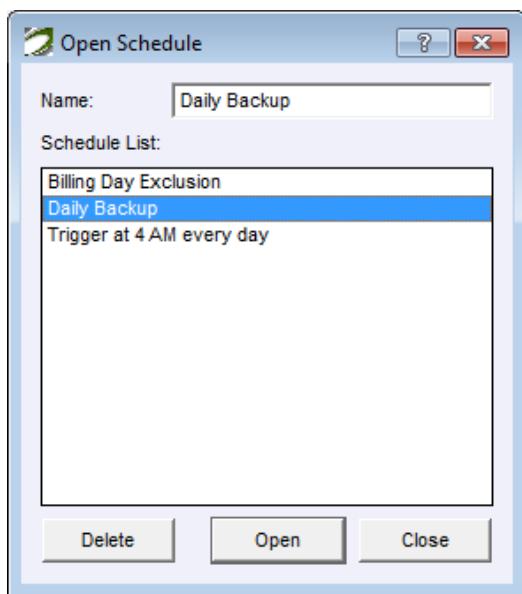


Figure 4-16 Open Schedule Dialog

3. Select a schedule from the list.
4. Click **Open** to open a schedule in **Schedule Manager**.
5. After making modifications, click Save.

Deleting a schedule

1. Open **Schedule Manager**.
2. Select a schedule from the list.
3. Click **Delete** to permanently remove a schedule.

Creating custom calendars

Calendars are lists of user specified dates that define custom schedule elements.

To create a new calendar:

1. In **Schedule Manager**, open a schedule.
2. Select the Custom radio button for an element.
3. Click **New**. The **Add Calendar** dialog appears (Figure 4-17). A list of dates can be constructed using this feature.

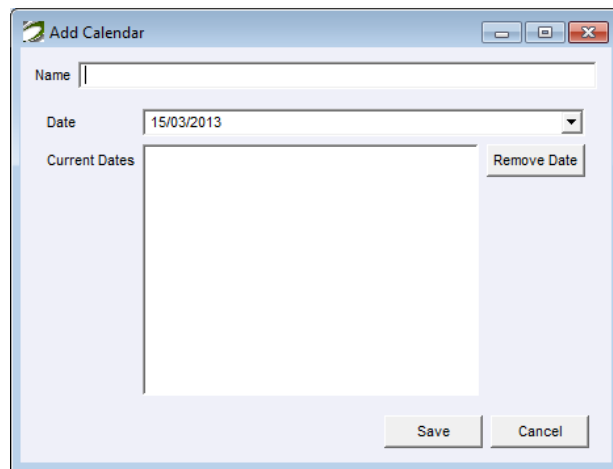


Figure 4-17 Add Calendar Dialog

4. Enter a name for the Calendar in the **Name** edit box.
5. Add new dates to the **Current Dates** list by clicking the **Date** combo box to display the calendar control (Figure 4-18). Select the required month by clicking the left/right arrows or the month dropdown menu. Select the required year by clicking the year spin control at the top of the calendar. Finally select the required day. The calendar control will then close and the selected date will be displayed in the **Date** calendar control and the **Current Dates** list. Dates can be entered manually by typing them in to the **Date** calendar control directly in DD/MM/YYYY format (note that each number is edited separately and the '/' delimiters cannot be edited). Press the enter key to add manually edited dates to the **Current Dates** list.

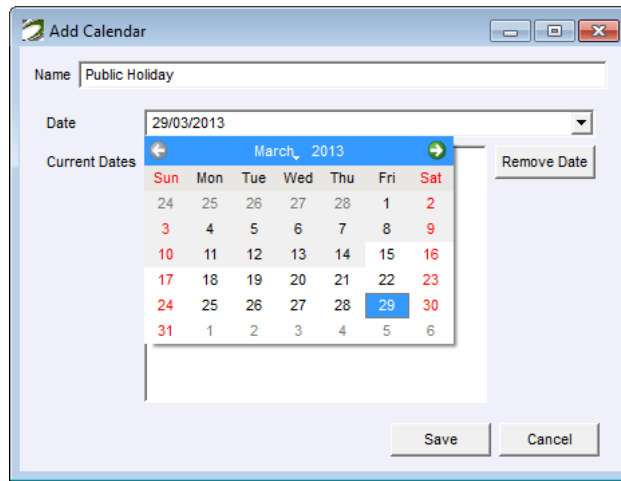


Figure 4-18 Adding Dates

6. Repeat step 5 to add more dates. If you select a date that you do not want to be included, select the date in the **Current Dates** list and click **Remove Date**.
7. Click **Save** to keep the calendar.

Data flow screen

The **Data Flow** screen (Figure 4-19) is designed to allow you to perform two principle functions:

- Assign policies to nodes on your network
- Define the data flow's routing path in cases where policies cause data movement from source nodes to destination nodes.

The **Data Flow** screen utilizes a drag and drop interface to define the topology.



To access the **Data Flow**, click the **Data Flow** icon in the Navigation bar.

Screen layout

The screen is split into five sections.

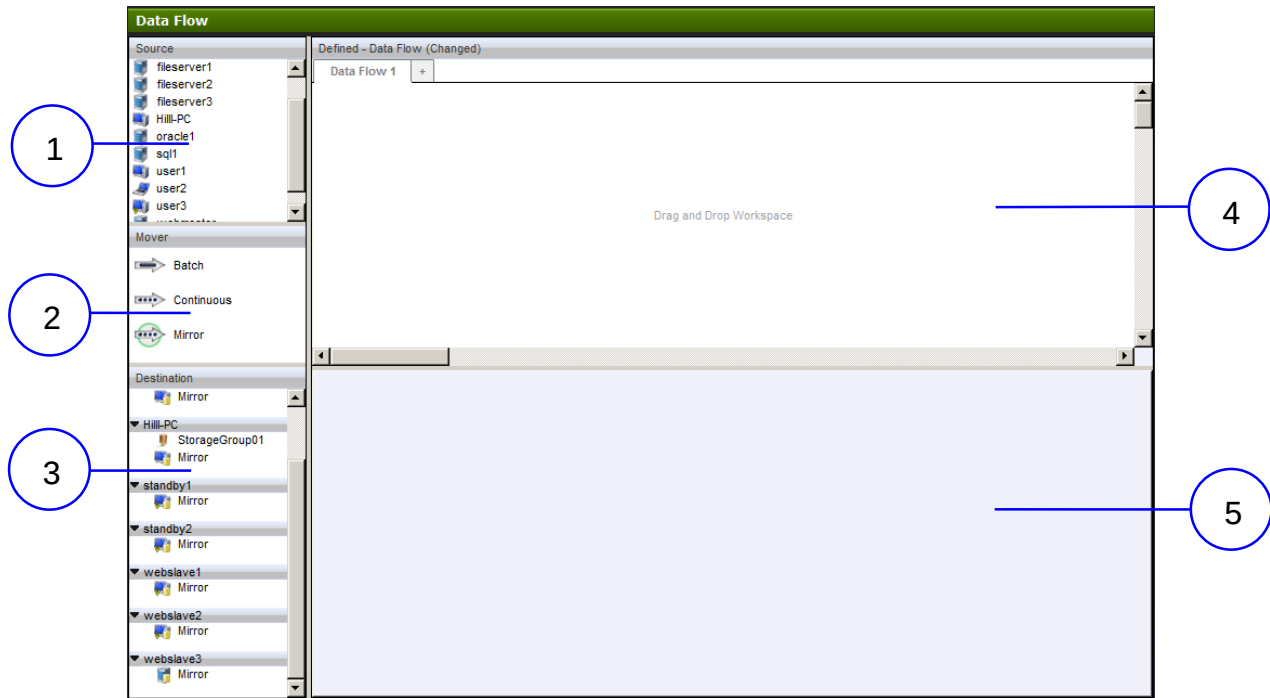


Figure 4-19 Data Flow Screen

1

Source

This is the list of source nodes and groups. Source nodes have the Data Source component installed or are CIFS nodes.

2

Mover

A mover is responsible for defining how data should be transferred from one node to another. Movers are not nodes; they are connectors that represent how data is transferred from source node to destination node over the network. The following types of mover are available:

- **Batch** – Transfers data from node A to node B only when explicitly triggered or scheduled by the user. Live replication does not occur. Node A must be a source node.
- **Continuous** – Transfers data from node A to node B in near real time, as and when it changes. Node A must be a source node.
- **Mirror** – Allows real-time replication of specific folders from one source node to another on the left and right of the Mirror data mover respectively. The Mirror mover provides for left to right movement, right to left movement, and bidirectional movement (which ensures that data from both sides of the mover are always present on each other). It is also possible to pause data mirroring. Source nodes must be placed on both sides of a mirror mover

irrespective of the direction selected. Note that file locking is not currently implemented for bidirectional mirrors, so care must be taken not to modify the same file on both sides of the mirror simultaneously.

3

Destination

Destination nodes have either the Replicate or Repository components installed, or are HCP, Azure or Tape Libraries. Each Repository node can have multiple Store Groups within it.

4

Drag and Drop Workspace

The workspace is where the data flow is defined. This area consists of multiple tabs that enable data flows to be logically grouped together, e.g. separate departmental networks. Initially, only one tab is displayed. Right clicking on the tab displays the following context menu options:

- **Select All** – selects all nodes and movers in the diagram
- **Paste** – pastes items in the paste buffer into the diagram
- **Add Tab** – adds a new data flow tab to the workspace
- **Delete Tab** – deletes the current data flow tab
- **Rename Tab** – enables the name on the tab to be edited
- **Disable/Enable Tab** – disables the data flow tab so that it is not included in the rules when next compiled and distributed.
- **Show Working Data Flow** – displays the data flow currently being edited
- **Show Distributed Data Flow** – displayed the data flow that is currently in operation (i.e. the current rules distributed to nodes in the system)

To add a new tab, click the plus sign (+) symbol at the far right of the tab headings.

To remove or rename a tab, right-click on the tab, and select **Delete Tab** or **Rename Tab**, respectively, from the context menu.

5

Properties tabs

These tabs show the properties for the currently selected item. Properties differ depending on the type of item that is selected.



Note: Some properties are not visible until other parts of the dataflow are configured.

Element placements

Data Instance Director allows you to drag and drop items from the **Source**, **Mover** and **Destination** palettes into the workspace.

There are only a few restrictions to the placement of items. Visual indicators show if a dragged object can successfully be placed in a location:



Item cannot be placed in the current location



Item can be placed in a location

Generally, an item cannot be placed in a location for the following reasons:

- The item is too close to the edge.
- The item is too close to another item.
- Moving the item would break the left-to-right flow. Source nodes must be to the left of a mover. Destination nodes must be to the right of the mover.

Connecting elements

A data flow is created by dragging items to the workflow area and creating connections between those items. Only compatible elements can be connected. For instance, agentless source nodes can only connect to batch movers.

There are three ways to connect elements:

Method 1 – connect

The Connect method is available from a pull down menu.

webmaster



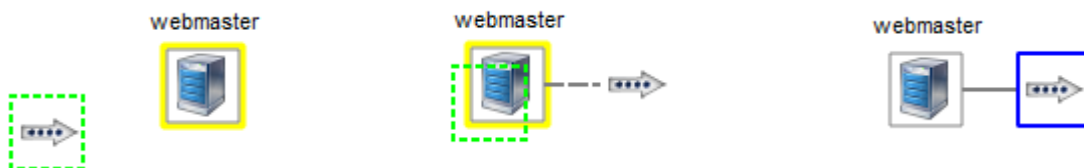
webmaster



1. When the connect method is selected, the cursor is replaced with a line. Any items that can be connected are highlighted with a yellow glow.
2. Click on the other item to complete the connection.

Method 2 – drag and connect

The best method for connecting items is to drag one item on top of another item.



1. Select an item from a palette. The items to which it can be connected are highlighted with a yellow box.

2. Drag the new item on top of the highlighted item. The dragged item is shown in its prospective position.

A connection line is drawn between the two items.

Moving and removing nodes

Nodes can be moved or removed by using options from the context menu (Figure 4-20) by right-clicking on a node.

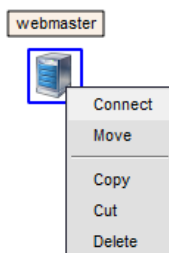


Figure 4-20 Pop-up menu

Multiple nodes can be moved or deleted at once by dragging a rectangle over the target nodes with the mouse. If a mistake is made, you can revert the dataflow to either the last saved version or the last distributed version. See the relevant section later in the manual for details.

When an item is selected, you can:

- Delete it using the Delete or Backspace key.
- Copy it using Ctrl-C.
- Cut it using Ctrl-X.
- Paste it using Ctrl-V.

Properties section

The property section displayed below the **Data Flow** workspace differs depending on the item selected. A selected item has a blue border. Here is an overview of the more common properties and how they relate to different items.

Source node – policy assignment

The properties of the source provide the method of assigning policies to a node or a group of nodes (Figure 4-21). Hovering over a policy's name displays a description as defined on the **Policy Definition** screen.

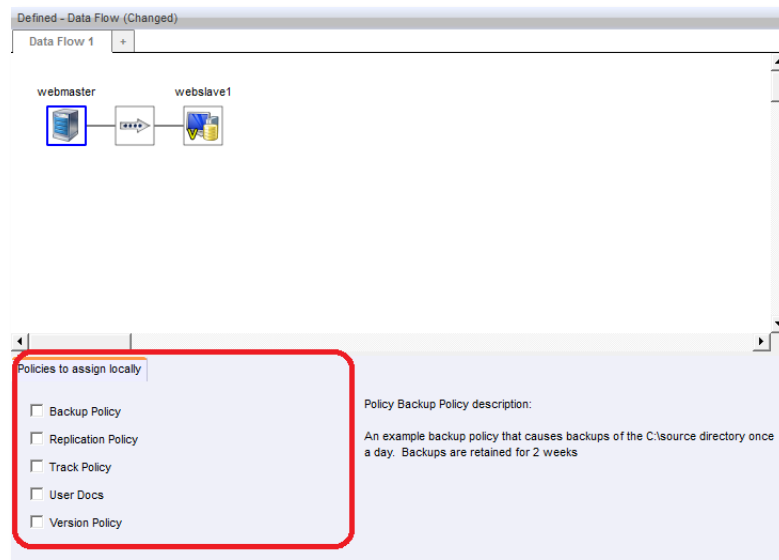


Figure 4-21 Using properties for policy assignment

If the selected policy requires another node to complete the policy, the source node displays a warning symbol (Figure 4-22) to indicate that the policy implementation is not complete.

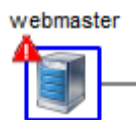


Figure 4-22 Policy Warning Indicator

Mover – data movement option and routing

The data mover property sheet (Figure 4-23) provides options on how to route data for policies that *move* data. By default, if a source node is configured to implement a policy, then any movers connected to it automatically route that policy. If the mover is not configured to implement a policy, then the nodes further downstream are not able to either configure or implement the policy.

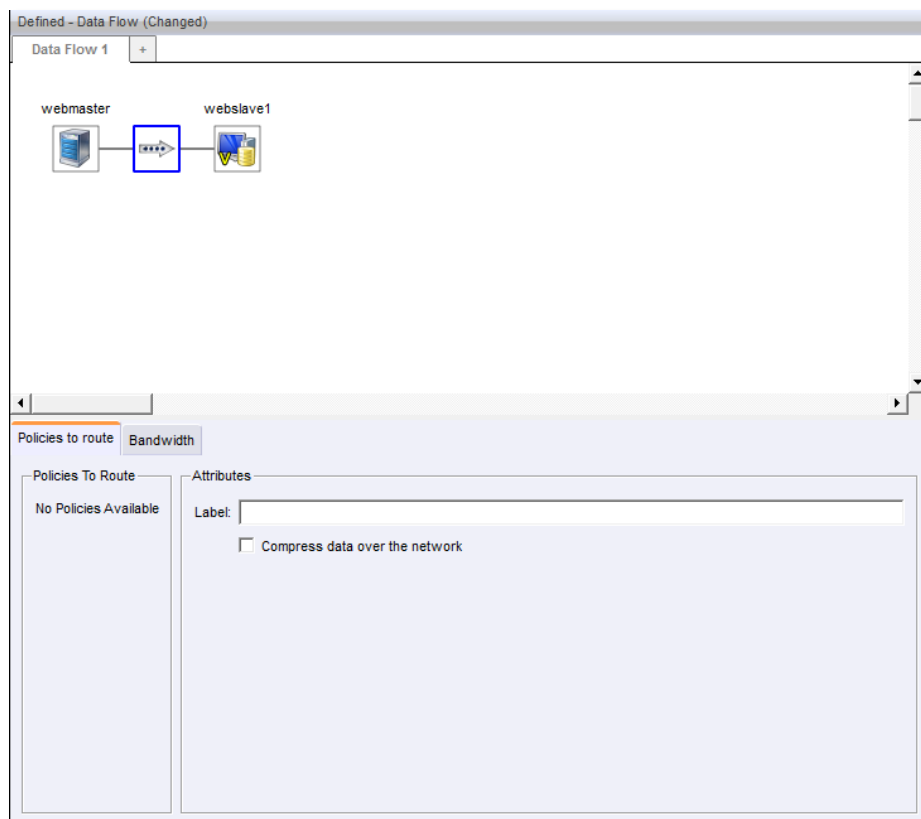


Figure 4-23 Mover policies

- **Label:** Movers do not have a name assigned to them by default. Assigning them a name in the edit box provides a name for that flow. This feature can be useful in situations where there is more than one mover connected to a node. Information about the different types of movers can be found in the [Screen](#) section above.
- **Compress data over the network:** You can compress data sent across the network by selecting this option. Compression decreases network utilization but increases CPU usage.

Bandwidth throttling

You can control the amount of bandwidth that data transfers use so that a network connection can be used simultaneously with other applications in the environment via the **Bandwidth** tab (Figure 4-24).

The screenshot shows the 'Bandwidth' tab in a 'Policies to route' window. It features a 'Default bandwidth' section with a value of 1000 Kb/s and a 'Throttle Default Bandwidth' checkbox. Below this is a section for custom bandwidth settings, including checkboxes for days of the week (Mon, Tue, Wed, Thu, Fri, Sat, Sun), a 'Custom bandwidth' field set to 15 Mb/s, a 'Start time' field set to 02:00:00, and a 'Duration' field set to 2 hours and 10 minutes. There are 'Add' and 'Remove' buttons. On the right, a list of configured policies is visible, showing start times and durations.

Start time	Duration
02:00:00	2 hours 10 minutes
13:38:02	2 hours 30 minutes
18:00:00	5 hours 30 minutes

Figure 4-24 Bandwidth settings tab

By checking the **Throttle Default Bandwidth** checkbox you can set a default parameter and keep a maximum level for bandwidth usage.

You can also assign specific bandwidth constraints for certain time windows by choosing day(s), Custom bandwidth, Start time and Duration, and then clicking Add.

For instance, you may decide to increase bandwidth during the workday when data is changing on a regular basis and decrease it at night when application batch processes are being run by other machines on the network.

If the amount of data that needs to be replicated continuously exceeds the current bandwidth, data is kept on the source until the bandwidth increases. Cache size on the Source Node can be adjusted to avoid any problems with non-transferred data reaching the cache limit. The configuration file `distributor.cfg` contains the values "MaxDiskCache" and "MaxMemoryCache" which can be increased if the cache limit becomes a problem.

Destination – data processing options

The **Available Policies** list on the left (Figure 4-25) defines what policies from source nodes can be implemented on this destination.

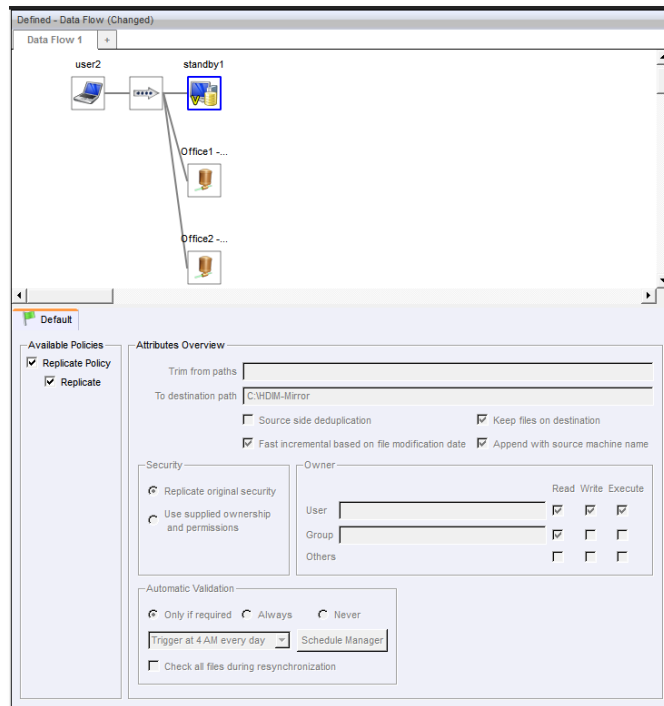


Figure 4-25 Destination attributes tab

If a policy on the source node dictates that data should be replicated, versioned or backed up (batch mode or CDP) and the Mover is configured to forward that policy, the destination node has the option of receiving that data, so long as it has the appropriate capabilities. For instance, a non-repository node cannot receive versioning policies. Other options exist that are specific to whether the destination is a mirror or a repository.



Note: A single policy may have several operations for a given data classification. The destination indicates which of those operations it can implement by displaying only the supported operation checkboxes.

Mirror destination options

When a mirror (or replication) destination node is configured to accept a mover policy, a destination path configuration must be assigned to it. Each mirror destination node can be associated with multiple path configurations. However, by default, there is only one configured, called *Default* (Figure 4-26).

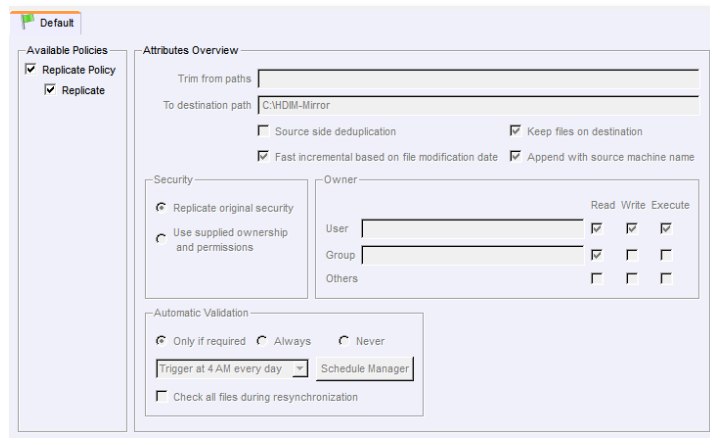


Figure 4-26 Default mirror destination tab

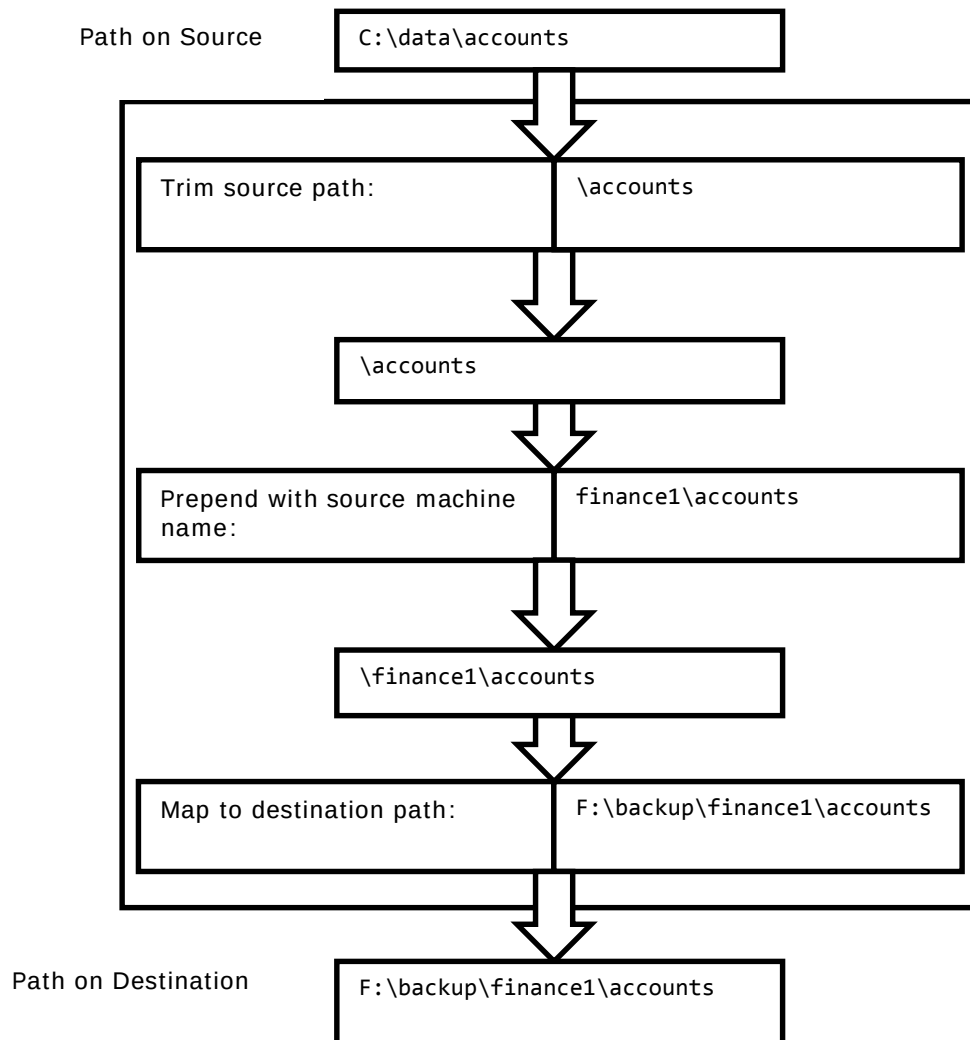
To add, rename, modify or delete a mirror destination, right-click in the tab bar and select the appropriate option from the context menu. Any newly created mirror destination configuration can be used in combination with any mirror destination node. Once a mirror destination is in use on a node, a green flag is displayed next to its name on the tab, as seen in the figure above. The green flag distinguishes destination configurations that are in use from those destination configurations that are inactive.

Every mirror destination configuration has the following settings associated with it that define where replicated files should be stored:

- **Trim from paths** – The part of each replicated item's source path that is removed before the remaining part is mapped to a destination path when stored. For Example: if the path of the replicated data from the Source node is C:\data\accounts, this path will, by default, be appended to the destination path on the Replication node F:\HDID-Mirror\. This will result in the long path F:\HDID-Mirror\C%\data\accounts (note the ':' is replaced by '%' to ensure a legal path name). By setting **Trim from paths** to C:\data\, this long path name can be abbreviated to F:\HDID-Mirror\accounts. See the diagram below for a visual overview of the path construction process.

- **To destination path** – Specifies where on the replication node to store incoming files. The source file's path is appended to this path. This path must be set and defaults to C:\HDID-Mirror. If a map path was also set, then this text replaces the mapped part of the path. The following table provides an example of how the path options work together, with the following assumptions:
 - Incoming Path = C:\data\accounts
 - Source node name = *finance1*.

Trim from Paths	To destination path	Append with source machine name	Resulting Path
<i>blank</i>	F:\backup	<i>not set</i>	F:\backup\C%\data\accounts
<i>blank</i>	F:\backup	<i>set</i>	F:\backup\finance1\C%\data\accounts
C:\data	F:\backup	<i>not set</i>	F:\backup\accounts
C:\data	F:\backup	<i>set</i>	F:\backup\finance1\accounts



Note: The mapping must correctly align to the source replication paths. For example, if a Trim from paths of `C:\data` is selected but the source being replicated is `C:\archives` then the files will not be replicated, as it is not possible to perform the path substitution.

Additional destination configuration options include:

- **Source side deduplication** – On occasions where many machines have identical roles within the dataflow and also contain very similar data (for instance, OS data and installed software on machines on a corporate network) then a network speedup can be obtained by avoiding sending the same data multiple times. Check this option to make use of this.

- **Keep files on destination** – This option instructs the synchronization process not to remove files on the destination if those files no longer exist on the source. Deselecting this option ensures that the destination directory is identical to the source directory, but results in data in the destination directory being permanently deleted when it no longer exist on the source. We recommend selecting this option to retain a copy of files deleted from the source.
- **Fast incremental based on file modification date** – If this is checked then Data Instance Director decides what files need resynchronizing based on whether the modification date has changed. This saves a lot of time when resynchronizing, but can be disabled if it is known that software is installed that will modify files without updating their size or modification date.
- **Append with source machine name** – Instructs the node to append the machine name of source nodes to the destination path defined above. This feature is useful if you are using the same policy for multiple nodes, as each will have its replication data in a different directory within the destination directory.
- **Replicate original security** – This option keeps the original security of the files on the source node. This option must be used with care, as it is possible that the destination node does not share the same users and privileges. In the instance that the destination does have the same users and privileges, the replication is errorless, but the files may be inaccessible on the destination.
- **Use supplied ownership and permissions** – Instead of using the original security credentials from the source node, this option allows the permissions to be mapped to a user on the destination node. This option enables the *Owner* controls so that the details can be provided.
- **Automatic Validation** – This is used to automatically resynchronize source to the destination based on a schedule.

Only if required – Will trigger resynchronization only if the destination is out of sync with the source(s).

Always – Will always perform a resynchronization with the source(s) based on the schedule.

Never – An automatic validation will never be performed. Checking this option will gray out the remaining options.

Schedule Manager – Schedules are created for automatic validation the same way as schedules for a policy, see the [Schedule Manager](#) section for more detailed instructions. A schedule for automatic validation must use trigger, by default a pre created schedule named *Trigger at 4 AM every day* is selected.

Check all files during resynchronization – Ignores file modification date and checks every file for changes.

Repository destination options

When using a repository as a destination for a dataflow a number of configuration options are available. These destination options for a repository node are contained within tab pages known as Store Templates and are shown in Figure 4-27. Each destination node can be associated with multiple stores templates. However, by default, there are only two displayed stores templates, *Default* and *Deduplicated*. Additional store templates can be created in the same way as is described for mirrors in the section above, via the context menu.

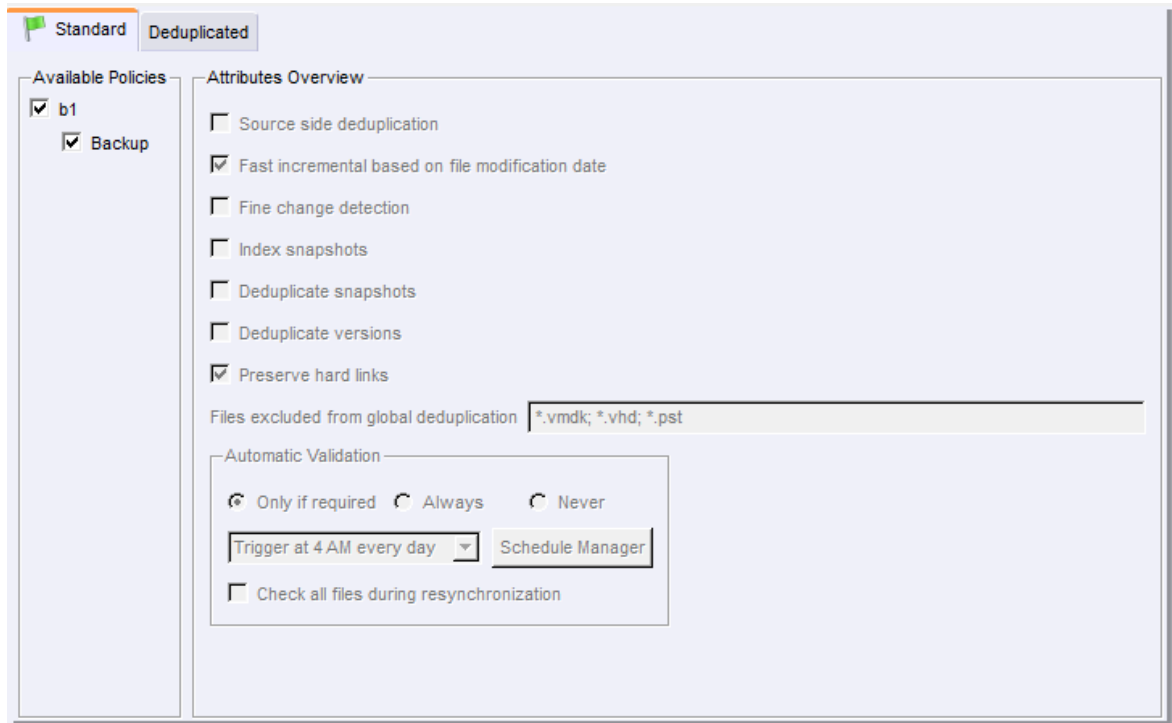


Figure 4-27 Default Stores

- **Source side deduplication** – This has the same effect as for mirror destinations. See the previous section for more information.
- **Fast incremental based on file modification date** – This has the same effect as for mirror destinations. See the previous section for more information.
- **Fine change detection** – Reduces the amount of data transferred and stored during a resynchronization. This is particularly useful for SQL and Exchange backups; for example, it will typically reduce Exchange data that is backed up from 90% of the total size to 30%. By default, the block size that is scanned for changes and transferred is 2 MB; it is 16 KB when **Fine change detection** is specified. An entire file is transferred if it has changed and is less than one block in size. This option should be used sparingly as there is a processing overhead.

- **Index snapshots** – If this is selected, then every item captured by the snapshot will be indexed so that it can be discovered quickly on an individual basis on the **Restore** screen's **Search** tab. If it is unselected then items within a snapshot will only be discoverable by browsing within a snapshot within the **Restore** screen. **Index snapshots** only applies to Backup operations, it has no affect for CDP since CDP snapshots cannot be indexed.
- **Deduplicate snapshots** – The store group is able to detect duplicated data across those snapshots transferred to it and can deduplicate it such that there is only one copy of the duplicated data. This can save disk space on the repository machine at the expense of extra CPU load.
- **Deduplicate versions** – The store group is able to detect duplicated data across those versions transferred to it and can deduplicate it such that there is only one copy of the duplicated data. This can save disk space on the repository machine at the expense of extra CPU load.
- **Preserve hard links** – If this is selected, then Data Instance Director checks that only one instance of the data is stored, regardless of the number of links pointing to it. This option increases the file system scan time during a resynchronization or batch backup.
- **Files excluded from global deduplication** – A semicolon-separated list of file extensions to be excluded from the store group's duplication detection can be entered here.

Tape destination options

Tape Libraries have a unique set of configuration options and these are contained within tab pages known as Media Pools. A tape library can be associated with multiple media pools. However, by default, there is only one media pool displayed, called *Standard* (Figure 4-28). Further media pools can be created in the same way as is described for mirrors in the section above, via the context menu.

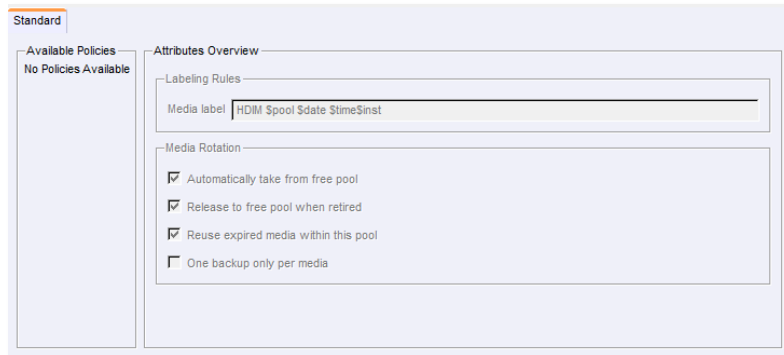


Figure 4-28 Default Media Pool

- **Media Label** – This is the label that will be attached to the individual media and is used to identify them. By default the media label is specified using substitution variables such as:

The following variables may be used:

\$pool – Appends which pool the media belongs to.

\$date – Appends the current date to the media.

\$time – Appends the current time to the media.

\$inst – Appends a unique instance number to the media.

\$node – Appends the node that is being backed up when this label is written.

- **Automatically take from free pool** – Media will be automatically taken from the free pool. The free pool is a pool of media that has yet to be allocated to a specific pool.
- **Release to free pool when retired** – When all the snapshots on the media have been expired it will be returned to the free pool where it can be reused.
- **Reuse expired media within this pool** – When media has been expired it will be reused without having been returned to the free pool.
- **One backup per media** – Only a single backup will be stored on each media.

Hitachi Block storage destination options

When using Hitachi Block storage as a destination in a dataflow a number of configuration options are available for configuring replication (Figure 4-29) and snapshotting (Figure 4-30) behavior.

The screenshot shows the 'Replicate' tab for a policy named 'SBP0003'. On the left, under 'Available Policies', 'SBP0003' and 'Replicate' are checked. The main area is divided into sections: 'Target Storage' with a 'Pool' dropdown set to '210613/ReplicationDPPool'; 'Replication Type' with radio buttons for 'In-System Clone (using ShadowImage)', 'Refreshed Snapshot (using Thin Image, In-System Batch Only)', 'Asynchronous Remote Clone (using Universal Replicator)', and 'Synchronous Remote Clone (using TrueCopy)'. The 'Synchronous Remote Clone' option is selected. Below it, 'On destination write failure:' has radio buttons for 'Ignore (Fence Level Never)', 'Fail source write (Fence Level Data)', and 'Fail source write if not in error status (Fence Level Status)'. The 'Ignore' option is selected. At the bottom, 'Replication Options' includes a 'Copy Pace' dropdown set to '8'. There are also input fields for 'Source Journal' and 'Target Journal' under the 'Asynchronous Remote Clone' option.

Figure 4-29 – Hitachi Block storage replicate tab

- **Pool** – specifies the target storage pool from which replication LDEVs are allocated. The browse button displays a dialog from which a pool can be selected.
- **In-System Cone (using ShadowImage)** – the replication is created as a cloned, in-system volume using Hitachi ShadowImage.
- **Refreshed Snapshot (using ThinImage, In-System Batch Only)** – the replication is created as a single, differential, in-system snapshot using Hitachi ThinImage. The snapshot is refreshed on each batch resync rather than creating a new snapshot for each resync.
- **Asynchronous Remote Clone (using Universal Replicator)** – the replication is created as a journalled, remote, asynchronous copy using Hitachi Universal Replicator.
 - o **Source Journal** – specifies the journal on the source side of the replication. The browse button displays a dialog from which a source journal can be selected.
 - o **Target Journal** – specifies the journal on the target side of the replication. The browse button displays a dialog from which a target journal can be selected.



NOTE – The selected journals must be unique to each operation and policy in the data flow.

- **Synchronous Remote Clone (using TrueCopy)** – the replication is created as a remote, synchronous copy using Hitachi TrueCopy.

On destination write failure:

- o **Ignore (Fence Level Never)** - When writing to the primary volumes, confirm the write regardless of whether the data has been copied successfully to secondary volumes.
 - o **Fail source write (Fence Level Data)** - Only confirm primary volume writes if the data is successfully copied to the secondary volume, generate a write error if not.
 - o **Fail source write if not in error status (Fence Level Status)** - Only generate a write error if the data is not successfully copied to the secondary volume *and* the replication hasn't been put into an error status (PSUE).
- **Copy Pace** – determines how quickly the storage array copies data (1 is slowest, 15 fastest). The array's default is 3, HDID defaults to 8.

Figure 4-30 – Hitachi Block storage snapshot tab

- **Pool** – specifies the target storage pool from which snapshot LDEVs are allocated. The browse button displays a dialog from which a pool can be selected.
- **Differential (using ThinImage)** – snapshots are created as a differential, in-system snapshots using Hitachi ThinImage.
 - o **Use consistency group** – when performing a crash consistent snapshot (e.g. via the hardware path classification), use consistency groups to make it truly crash consistent. The hardware has a limited number of consistency groups available so they should be used sparingly.
- **Full Clone (using ShadowImage)** – snapshots are created as a cloned, in-system volumes using Hitachi ShadowImage.

Hitachi File storage destination options

When using Hitachi File storage as a destination in a dataflow a number of configuration options are available for configuring replication (Figure 4-31).

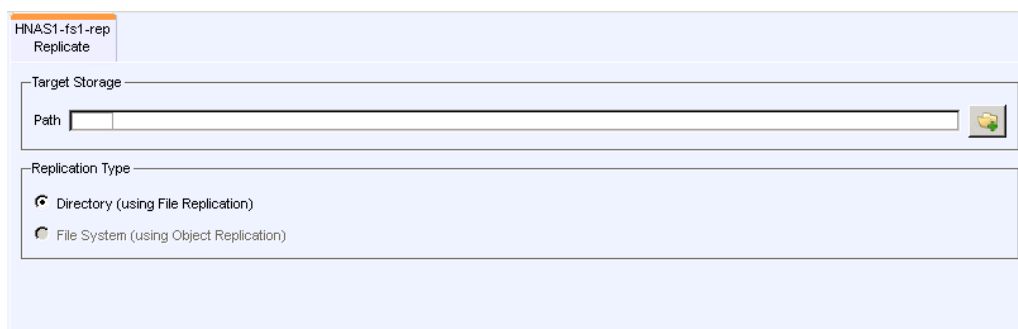
The screenshot shows a software interface for configuring replication. At the top, there is a tab labeled 'HNAS1-fs1-rep' with a sub-label 'Replicate'. Below this, there is a section titled 'Target Storage' which contains a 'Path' text input field and a folder icon button. Underneath the 'Target Storage' section is a 'Replication Type' section with two radio button options: 'Directory (using File Replication)' and 'File System (using Object Replication)'. The 'Directory' option is currently selected.

Figure 4-31 – Hitachi File storage replication tab

- **Path** – specifies the file system on the file storage hardware where the replication will be placed. This must be a normal file system not an Object Replication Target.
- **Directory (using File Replication)** – replication will be performed using Hitachi File Replication
- **File System (using Object Replication)** – replication will be performed using Hitachi Object Replication. This option is not available in version 5.1.

Reverting the data flow

You can revert the Data Flow to the last distributed state or last saved state. This feature allows you to undo modifications to the Data Flow currently being edited.



Note: The Data Flow reversion process is irreversible. Once the operation has been completed, you cannot restore to the state prior to reverting.

- Select **Tools** → **Revert to Distributed Data Flow** from the main menu to revert the Data Flow to the last distributed state.
- You can view the distributed Data Flow by either accessing the **Monitor** screen, or by selecting View → Show Distributed Data Flow from the main menu. Using the latter option displays the distributed Data Flow in *Read-Only* mode. The nodes can be selected and their attributes can be viewed, but the flow *cannot* be modified. The **Source**, **Mover** and **Destination** palettes are disabled and nodes cannot be added to the Data Flow. The title bar displays the title *Distributed dataflow – Read only*. Before choosing to view the distributed Data Flow, be sure to save any changes since unsaved changes will be lost.

- You can toggle back to viewing the working Data Flow (the one being edited) and the distributed Data Flow (Figure 4-32) by selecting **View** → **Show Working Data Flow** from the main menu. Note that any unsaved changes made before viewing the distributed Data Flow are not kept.

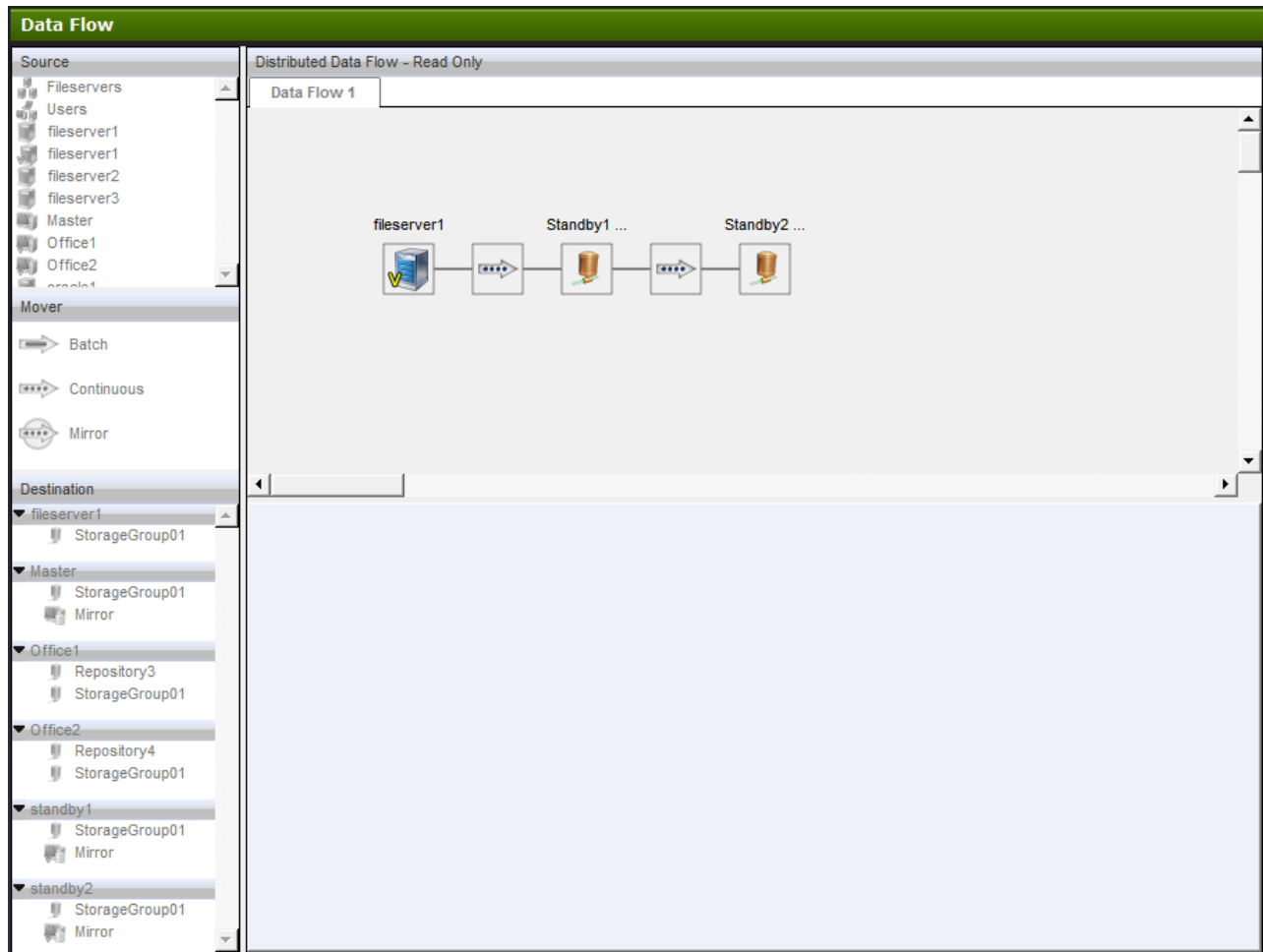


Figure 4-32 Read-only view of the distributed data flow

As well as reverting to the last distributed state, it is also possible to revert the Data Flow to the last saved state. This is the equivalent of restarting the GUI without saving changes to the Data Flow.

Select **Tools** → **Revert to Last Saved Data Flow**. The operation is not reversible, and current unsaved changes are discarded.

Compiling and distributing rules

Rules files control the operation of the Data Instance Director components on each node in the data flow definition. Rules files need to be generated by the *rules compiler* after all policies have been created, the data flow has been defined, and policies assigned to source and destination nodes.

The rules compiler is started by either:

- Clicking on the **Save and Activate** icon  on the tool bar, or
- Selecting **Tools** → **Save and Activate** from the main menu.

The **Compiler Results** dialog is displayed. It contains information, warning and error messages about the compilation process and if the compilation is successful, it includes a summary of what policies each node in the data flow will be implementing, forwarding or processing. Following a successful compilation, the rules files may either be distributed immediately or scheduled for distribution later.

Rules for compilation

If the data flow definition shown in Figure 4-33 is compiled:

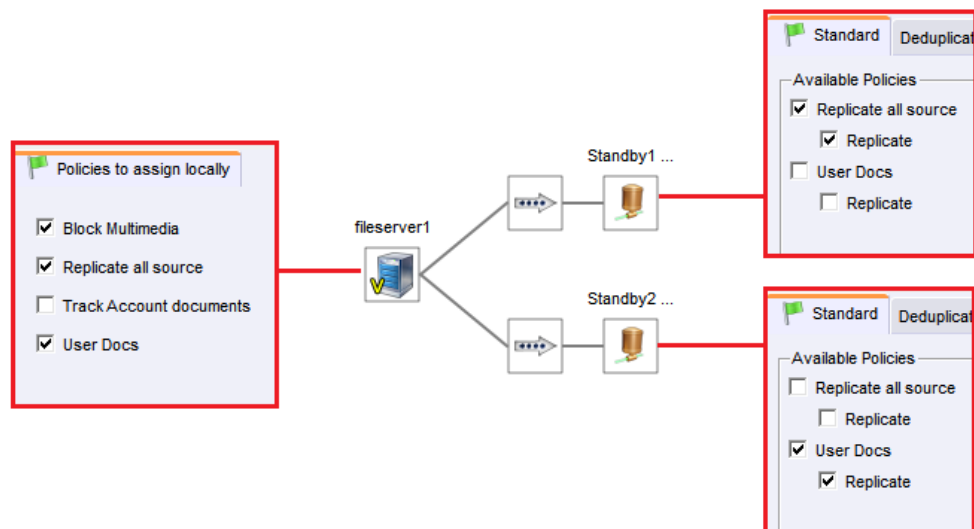


Figure 4-33 Compile and distribute, one-to-many

A **Compiler Results** dialog is displayed (Figure 4-34).

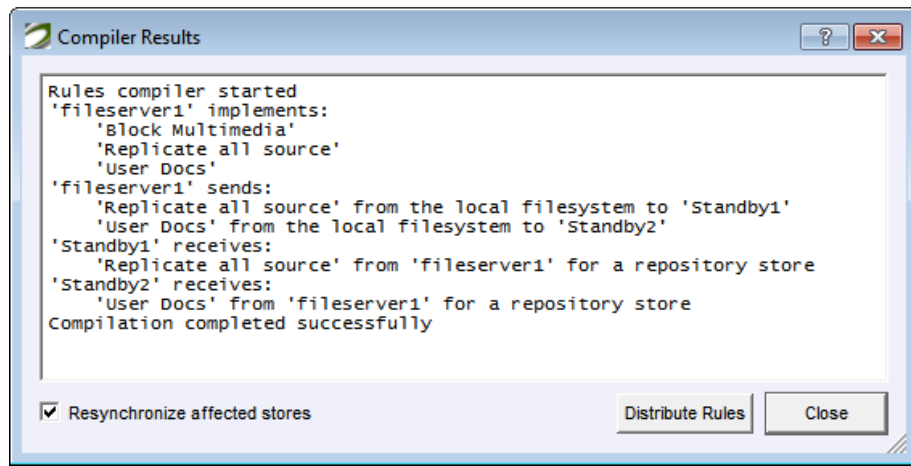


Figure 4-34 Compiler results, one-to-many

Following a successful compilation, a summary of the actions performed by each node in the data flow is displayed, indicating which nodes implement specific policies, and how the policy data is passed across the network to the nodes that are processing that data (Figure 4-35).

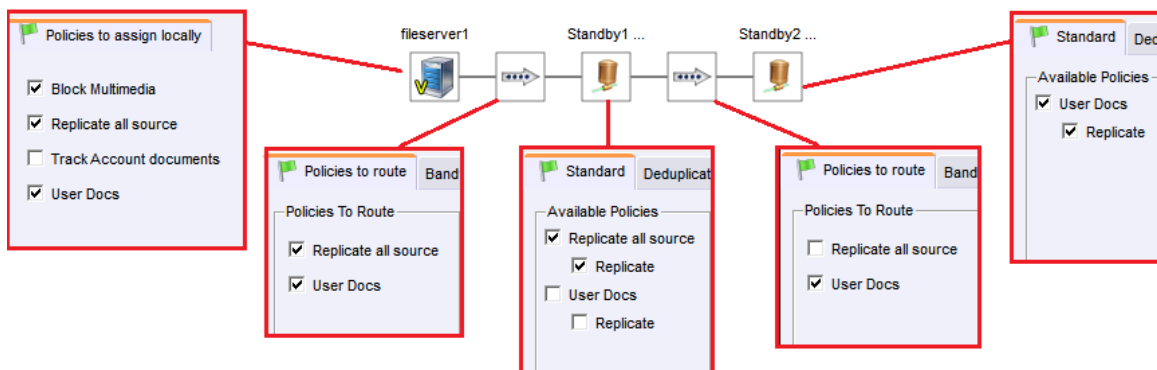


Figure 4-35 Compile and distribute, forwarding

Figure 4-36 shows the compiler output generated by a network using an intermediate node to forward data:

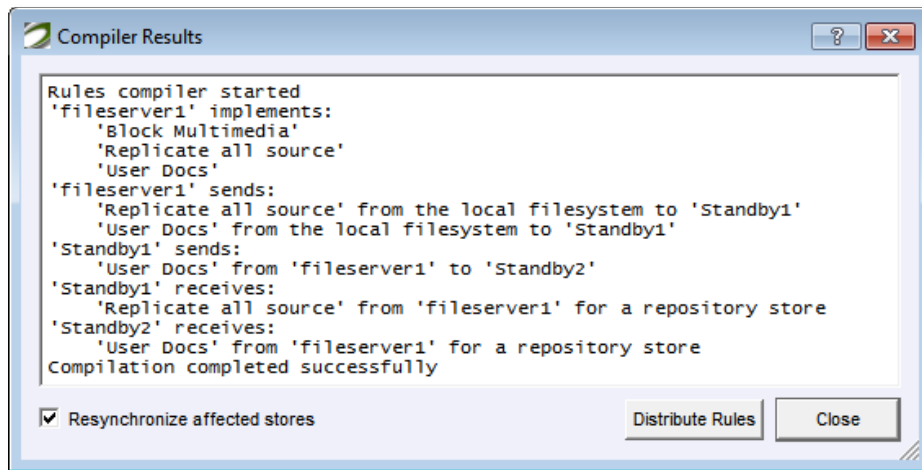


Figure 4-36 Compiler results - forwarding

Compilation errors

If any errors were made while creating the policies or the data flow definition, then a brief description of the fault is displayed in the **Compiler Results** dialog, and the compilation fails. The **Distribute Rules** button is disabled to prevent erroneous rules from being distributed to nodes.

Compiler errors may be the result of incorrect or incomplete policies, an incomplete dataflow definition, or incorrect or incomplete dataflow item attributes. Another cause can be incomplete or ambiguous routes for policies through the diagram, or missing parameters on destination nodes. Errors can occur when operations are changed, added or deleted in policies without updating the dataflow. When errors occur, carefully review the error messages reported in the Compiler Results dialog and make the necessary revisions.

Rules distribution

When the dataflow is compiled successfully, the new rules must be distributed to all the nodes used in the dataflow definition before they will take effect. This is done by clicking the Distribute Rules button on the **Compiler Results** dialog.

If all nodes in the dataflow are up and connected to the master, then the distribution should complete quickly, and all nodes will be activated with the new rules.

The **Compiler Results** dialog (Figure 4-37) shows the progress of the distribution.

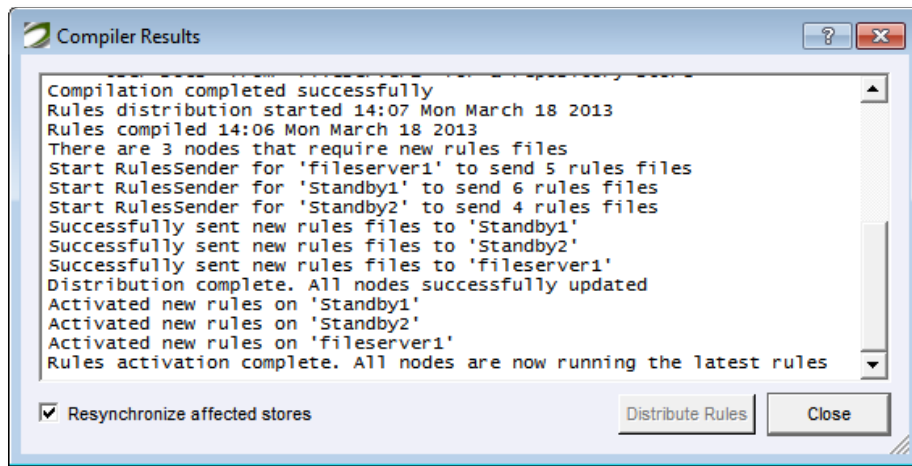


Figure 4-37 Successful compile and distribute

The **Resynchronize affected stores** option will, when checked, only resynchronize stores that are affected by these newly distributed rules. If this option is not checked then all stores are resynchronized regardless.

The **Log Manager** screen (Figure 4-38) can be used to watch or review the progress of a rules distribution:

Category	Level	Log Message
Node Manager		Rules distribution initiated
Node Manager		Rules distribution started at 14:07 Mon March 18 2013
Node Manager		Rules will be sent to 3 nodes
Node Manager		Start RulesSender for 'fileserver1' to send 5 rules files
Node Manager		Start RulesSender for 'Standby1' to send 6 rules files
Node Manager		Start RulesSender for 'Standby2' to send 4 rules files
Node Manager		Receiving new rules files
Node Manager		Receiving new rules files
Node Manager		Receiving new rules files
Node Manager		Completed distribution of rules files to 'Standby1'
Node Manager		Completed distribution of rules files to 'Standby2'
Node Manager		Successfully sent new rules files to 'Standby1'
Node Manager		Completed distribution of rules files to 'fileserver1'
Node Manager		Successfully sent new rules files to 'Standby2'
Node Manager		Successfully sent new rules files to 'fileserver1'
Node Manager		Distribution complete. All nodes successfully updated

Figure 4-38 Log Manager

If a node in the data flow definition is not connected to the master, then the rules files cannot be sent to that node. The distribution continues for all other nodes. When the distribution is complete, the new rules are activated on those nodes. The absent node is flagged as *out of date* in the **Monitor** screen with an appropriate icon . When the absent node becomes available and reconnects to the master, the new rules are sent automatically to that node, and activated immediately after they are received.

If there are any nodes that were present in the previous data flow definition, but are not in the current definition, then these nodes are deactivated when the new rules are activated on all other nodes or, if any such node is not currently available, as soon as it re-connects to the master.

Monitor screen

The **Monitor** screen displays information related to the current distributed

data flow. Access the **Monitor** screen by clicking the **Monitor** icon in the **Navigation** pane.



Window layout

The **Monitor** screen (Figure 4-39) is split into two resizable sections. The upper area shows the active topology within the available data flow tab pages, the nodes and their states. The lower section displays error logs and synchronization progress (source nodes only) for the node selected.

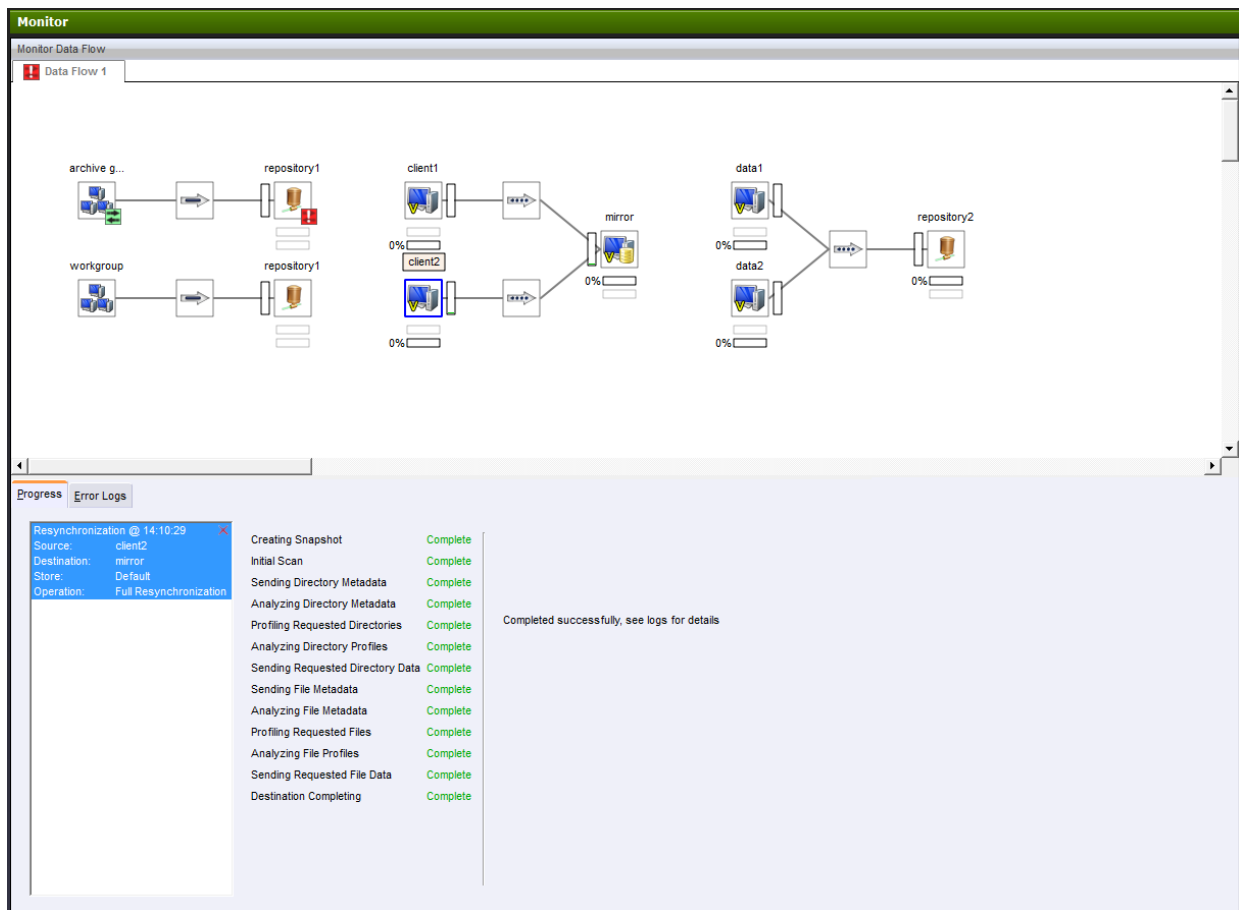


Figure 4-39 Monitor screen

Once the rules have been distributed, the associated data flow is shown and real-time information for the participating nodes is displayed.

Source nodes, movers and destination nodes are represented by the same icons as those on the **Data Flow** screen. Node names are displayed above the node. If a node has a long name, then the name is truncated. Hovering the cursor over the node icon will display the full node name.

Stats

Each node icon has up to four utilization bars. The bars on each side of a node represent the current cumulative network utilization of the node.



- The bar on the left represents incoming traffic (absent for source nodes).
- The bar on the right represents outgoing traffic (absent for destination nodes).

The bars below the icon represent the percent of memory or disk cache that is currently being used. The total memory or disk cache is the total used for all nodes connected to that node.

- The top bar represents the cache being used by incoming data, (the node's Local Cache).
- The bottom bar represents cache being used by outgoing data, the node's Remote Cache.

The network utilization bars are split into three sections. The lower section (light green) represents utilization of zero to 10 Mb/s. The mid-section (medium green) represents utilization of more than 10 Mb/s up to 100 Mb/s. The highest section (dark green) represents utilization of over 100 Mb/s to 1 Gb/s.

The network utilization hover hint (Figure 4-40) shows real time, precise information of the network utilization.

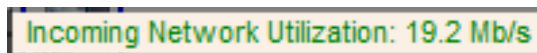


Figure 4-40 Network Utilization hover hint

The cache utilization bars represent memory cache in cyan, or disk cache in red. This can be seen in the hover hint for either cache utilization bar (Figure 4-41), along with which cache (local or remote) the bar represents.

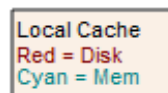


Figure 4-41 Cache bar hover hint

If a node has no incoming or outgoing connections (therefore never needing to cache data), the local or remote cache bar is faded. If the node has both incoming and outgoing data then both bars are active.

Badges

Various badges are superimposed on nodes to visually represent the state of a node and what it is currently doing. These badges and their meanings are listed below.

	Resynchronizing – This badge is displayed upon source nodes while they are synchronizing to a connected destination. This will only appear when an initial synchronization is occurring (because of rule distribution) or when a resynchronization has been explicitly triggered or scheduled.
	Validating – This badge is displayed upon a source node while its data is being validated against the corresponding data on the destination. If inconsistencies are found the destination node is flagged with the Resynchronization required badge below.
	Resynchronization required – This badge is displayed upon a destination node if the data stored upon it is out-of-sync with one of the source nodes it is connected to. This is essentially an error state, and should be rectified by right-clicking on the node in question and resynchronizing it with the problematic source node by using the Request Resynchronization option.
	Out of date rules – This badge is displayed if the master has updated and distributed new rules but the node in question was not connected so could not receive them. When the node is reconnected it will receive the updated rules.
	Error logged – Displayed when an error has occurred pertaining to the node. To clear this flag the errors must be acknowledged and cleared by the user. Convenient ways to acknowledge errors are detailed in the Node Actions section below, but they can also be cleared from the Log Manager screen.
	Stopped Data Capture – This badge is displayed when the Data Source component upon a source node has been stopped by the user. This is done by right-clicking upon a node and selecting Stop Data Capture . In this state the node will not be capturing live changes to its source data, and – when the Data Source component is re-enabled – the node will require a full resynchronization in order to receive any changes that may have occurred in the interim period.
	Stopped distribution – This badge is displayed when data distribution from a source node has been stopped by the user. This is done by right-clicking upon a node and selecting Stop Distribution . Any changes that occur to data on the source are cached in the source's memory and sent when the distributor is re-enabled. If distribution is stopped on a destination, the outcome is the same as if it were stopped on the source except that the transmitted data is cached on the destination side instead of the sender side.

	Restoring – This badge is displayed on any node that is restoring data from a repository.
	Pending Media Requests – This badge is displayed when there is a job running on the Tape Library.
	Media requests require user attention – This badge is displayed when there is a task that cannot be completed without user intervention, for example, tapes may need to be added, or a library may be offline.

Disconnected nodes

Nodes that are not contactable by the master, either because no network connection can be made or because the Data Instance Director services are not running on the node are displayed with a red background (Figure 4-42).



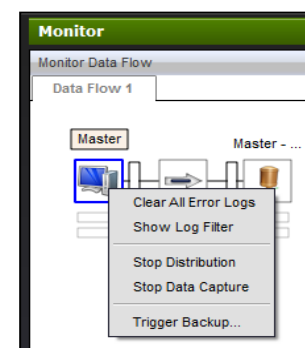
Figure 4-42 Disconnected node (with out-of-date rules)

Node actions

Source and destination nodes have a number of options available from the context menu. A number of these options are common to all types of nodes. However, some options are available only on specific types of node. Right click on any node to display the context menu.

Common actions

- Clear All Error Logs** - Attempts to acknowledge all unacknowledged error logs currently on the selected node. Once selected, an input pop-up is shown so that an acknowledgement comment can be entered; but is not mandatory. When all error logs have been acknowledged the nodes error flag () is removed.
- Show Log Filter** - Displays additional filter options for the selected node's log filtering (Figure 4-43). This option allows the displayed logs to be filtered by criteria other than their error level. It is even possible to display logs of *all* levels (the default is errors only). Refer to the Log Manager section below for instructions on building filter criteria.



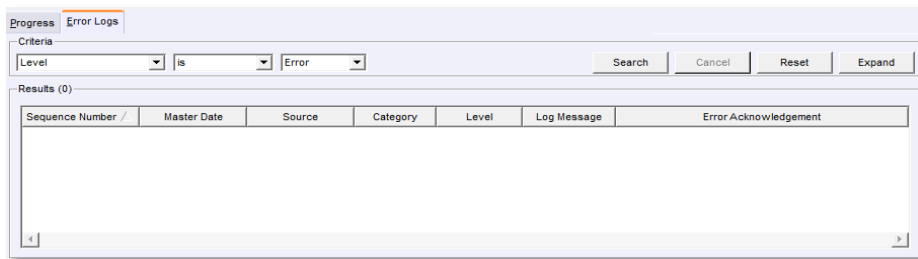


Figure 4-43 Error logs and filters

- **Stop / Start Distribution** - Stops or starts the distribution of data from the selected node. Distribution will be stopped for the node for all flows in which it acts as a source. More information can be found in the section regarding the Stopped Distribution badge  above.

Source-only actions

- **Stop / Start Data Capture** - Stops or starts the Data Source component which is responsible for monitoring changes to files on the system. More information can be found in the section regarding the Stopped Data Capture badge  above.
- **Trigger Backup...** - This option is enabled only for nodes that are acting as sources of data, this can include repository nodes in a SRS ([Smart Repository Sync](#)) setup. The **Trigger Backup** dialog (Figure 4-44) is displayed listing the policies available on the selected node. Selecting the required policy and clicking **Run now** will instantly start backup of the source to the destination.

This option will be grayed out for sources connected to a destination with a Continuous mover.

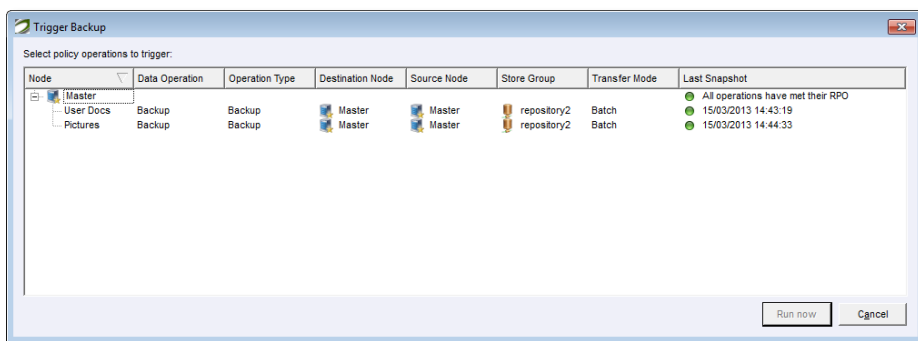


Figure 4-44 Trigger backup dialog

Destination-only actions

- **Request Resynchronization** - Allows the user to request a resynchronization of a destination. When this option is selected, a dialog is displayed (Figure 4-45) listing all available paths that can be resynchronized, along with their resynchronization state. Nodes that are already being resynchronized are not selectable and are greyed out.



Figure 4-45 Request resynchronization dialog

1. Select the source node that the destination should resynchronize with.
2. Click **Resynchronize**. If a resynchronization is not required for the selected path, a confirmation is requested. Click **Yes** to trigger Data Instance Director to force a resynchronization of the path.

The **Request Resynchronization** dialog also presents additional options that can be performed during a resynchronization:

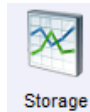
- **Check all files during resynchronization** – Ignores file modification date and checks every file for changes.
- **Create persistent snapshot** – Once the resynchronization is complete a persistent snapshot will be created, this snapshot can never be manually deleted.
- **Resynchronize twice, initially ignoring live changes** – This will perform two resynchronizations. The first will ignore any live changes that are made; the second will not ignore them. This option is only valid for a live policy and is ignored when using batch, it is to be used when there are problems with disk cache overflow, using this option will reduce the disk usage.
- **Request Validation** - Validation functions in a similar way to resynchronization. The **Validate Stores** dialog is displayed.
 1. Select the policy that requires validation.
 2. Click **Validate**.

Validated policies do not need to be revalidated, and requests are queued if validation is currently in progress.

- **Mount Repository** - Allows the user to mount a store group that has been unmounted, this option will be grayed out unless the store group is currently unmounted.
- **Unmount Repository** - Allows the user to unmount a store group that has been mounted, this option will be grayed out unless the store group is currently mounted.



Note: More options for Mounting and Unmounting are located within the **Storage** screen:



Progress

The **Progress** tab shows progress information for source nodes undergoing synchronization. The left-hand pane lists the activities being performed. An activity can be aborted by clicking the 'red X' in the top right corner of that activity's description. The list to the right of this shows the status of each step that is performed for the selected activity. The right hand area displays a summary of the selected activity once completed.

For all other node types this tab will be empty.

Repository statistics

Clicking on a store group within the **Monitor** screen and selecting the **Repository Statistics** tab (Figure 4-46) displays a number of statistics that can help determine if certain repository processes are queued or under way and what the current performance and resource usage is at any given time.

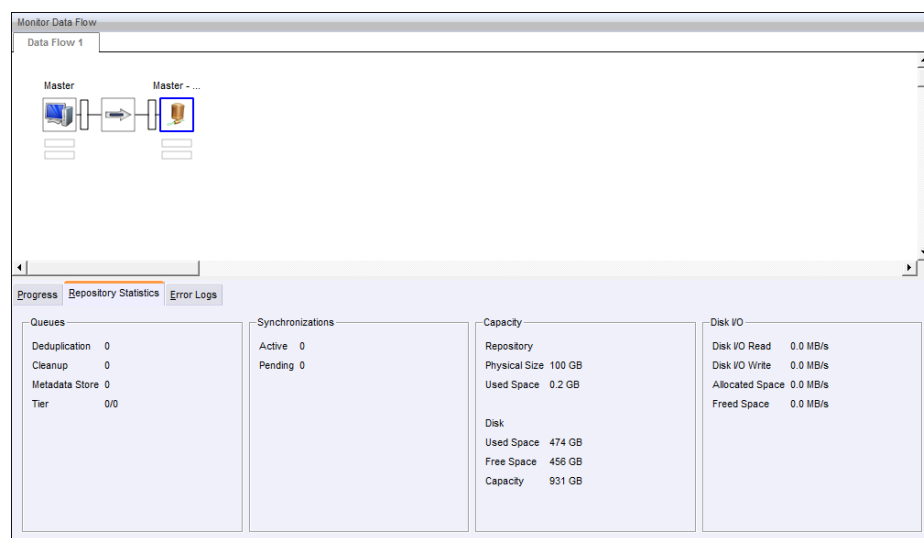


Figure 4-46 Repository statistics tab

The following parameters are reported:

- **Queues** statistics:
 - o **Deduplication** – the number of items that have been identified for deduplication and are awaiting processing.
 - o **Cleanup** – the number of blocks that have been marked for release from the repository and are awaiting processing.
 - o **Metadata Store** – the number of items that are awaiting indexing by the MDS.
 - o **Tier** – the number of files tiered to HCP or Azure. The first value indicates how many are currently being processed and the second is how many are waiting to be processed.
- **Synchronizations** statistics:
 - o **Active** – the number of batch backups or resynchronizations currently being processed.
 - o **Pending** – the number of batch backups and resynchronizations awaiting processing.
- **Capacity** statistics:
 - o **Repository:**
 - **Physical Size** – the total space reserved on disk for use by the repository. When this size is reached, the repository will grow by a pre-configured amount.
 - **Used Space** – the total space on disk currently in use by the repository.
 - o **Disk:**
 - **Used Space** – the total space in use on the volume where the repository is mounted.
 - **Free Space** – the total space remaining on the volume where the repository is mounted.
 - **Capacity** – the size of volume where the repository is mounted.
- **Disk I/O** statistics:
 - o **Disk I/O Read** – the read throughput to disk, attributed to repository activity.

- o **Disk I/O Write** – the write throughput from disk, attributed to repository activity.
- o **Allocated Space** – the rate at which disk space is being used up by the repository.
- o **Freed Space** – the rate at which disk space is being freed up by the repository. Note that this space is made available for reuse by the repository only.

Monitor tabs

Tabs created in the **Data Flow** (Figure 4-47) are displayed within the **Monitor** screen in an identical format, with the addition of node state icons.

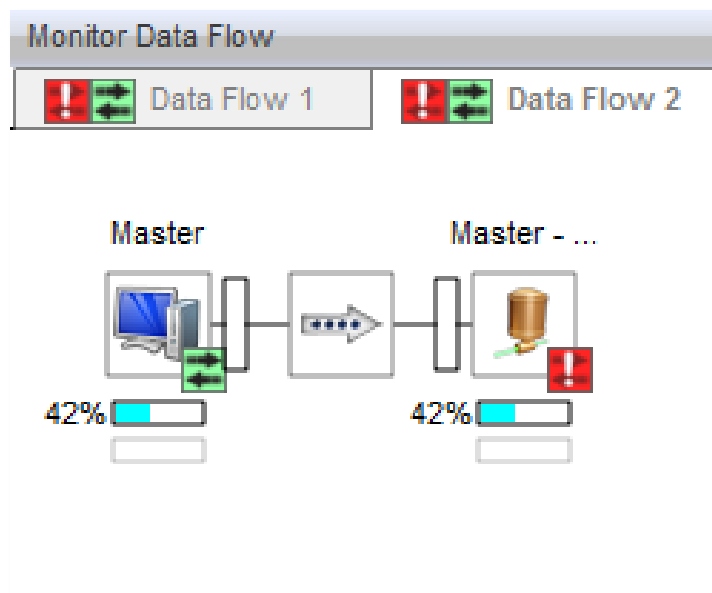


Figure 4-47 Monitor Data Flow

Monitoring node groups

The group node icon displays status in the same fashion as normal nodes. When a group is selected, error logs for all nodes within that group are displayed in the error logs section. A group can also be expanded by double-clicking to show the individual nodes within the group. In the expanded view nodes can be individually monitored and managed (Figure 4-48).

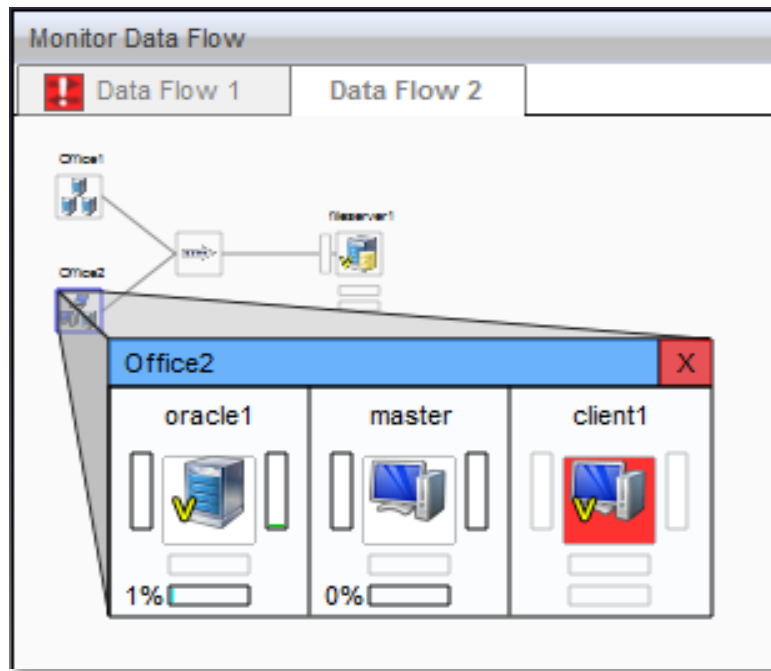


Figure 4-48 Monitor Data Flow Group Details

To monitor a group, either double-click on the node, or select the Monitor Members option from the node's context menu.

By default, node statistics (node cache and network utilization bars) are not displayed when a group is inspected. This feature prevents the system from simultaneously requesting these statistics from a large number of nodes.

To display the statistics for a node, click on the node's icon. Any error logs are also displayed for that node.

Node states, such as *unavailable*, *performing a resynchronization*, and other status icons, display on nodes within a group. However, only key icons; *error*, *resynchronization* and *out of sync*, are shown on the group icon.

Moving on and off the **Monitor** screen does not close the group inspection view. However, the statistics for the nodes do not remain displayed and need to be reselected. To close the group inspection view, press the red X button in the top right corner.

Log Manager

The **Log Manager** allows you to filter, view, and manage Data Instance Director application logs.



Access the **Log Manager** screen by clicking the **Log Manager** icon in the **Navigation** bar.

Screen layout

Log Manager is divided into two areas, **Criteria** and **Results** (Figure 4-49).

The screenshot shows the Log Manager interface. The top section is titled 'Log Manager' in a green bar. Below it, the 'Criteria' section contains search filters: 'Level' (dropdown), 'is at least' (dropdown), 'Information' (dropdown), 'Master Date' (dropdown), 'is after' (dropdown), and a date field '15/03/2013 00:00:00'. There are buttons for 'Search', 'Cancel', 'Reset', and 'Collapse'. The 'Results (578)' section displays a table with columns: Sequence Number /, Master Date, Local Date, Log Origin, Data Source, Category, Level, and a text field. The table contains 20 rows of log entries, including resynchronization and rules distribution logs.

Sequence Number /	Master Date	Local Date	Log Origin	Data Source	Category	Level	
000001459	15/03/2013 15:58:16	15/03/2013 15:58:16	Master	Master	Resynchroniza...	Information	No directory data requested for fast incremental res
000001461	15/03/2013 15:58:16	15/03/2013 15:58:16	Master	Master	Resynchroniza...	Information	No file data requested for fast incremental resynch
000001465	15/03/2013 15:58:16	15/03/2013 15:58:16	Master	Master	Repository	Information	Fast incremental resynchronization of 'Master' to rep
000001468	15/03/2013 15:59:51	15/03/2013 15:59:51	Master	Master	Replication	Information	User has requested a resynchronization of Master, t
000001469	15/03/2013 15:59:51	15/03/2013 15:59:51	Master	Master	Repository	Warning	A resynchronization of 'Master' to repository2 Live F
000001474	15/03/2013 15:59:52	15/03/2013 15:59:52	Master	Master	Resynchroniza...	Information	Performing a fast incremental resynchronization of '1
000001485	15/03/2013 15:59:56	15/03/2013 15:59:56	Master	Master	Resynchroniza...	Information	No directory data requested for fast incremental res
000001487	15/03/2013 15:59:56	15/03/2013 15:59:56	Master	Master	Resynchroniza...	Information	No file data requested for fast incremental resynch
000001491	15/03/2013 15:59:56	15/03/2013 15:59:56	Master	Master	Repository	Information	Fast incremental resynchronization of 'Master' to rep
000001513	15/03/2013 16:35:17	15/03/2013 16:35:17	Master	Master	Node Manager	Information	Rules distribution initiated
000001514	15/03/2013 16:35:17	15/03/2013 16:35:17	Master	Master	Node Manager	Information	Rules distribution started at: 16:35 Fri March 15 201
000001515	15/03/2013 16:35:17	15/03/2013 16:35:17	Master	Master	Node Manager	Information	Rules will be sent to 1 nodes
000001516	15/03/2013 16:35:17	15/03/2013 16:35:17	Master	Master	Node Manager	Information	Start RulesSender for 'Master' to send 7 rules files
000001517	15/03/2013 16:35:17	15/03/2013 16:35:17	Master	Master	Node Manager	Information	Receiving new rules files
000001518	15/03/2013 16:35:17	15/03/2013 16:35:17	Master	Master	Node Manager	Information	Completed distribution of rules files to 'Master'
000001519	15/03/2013 16:35:17	15/03/2013 16:35:17	Master	Master	Node Manager	Information	Activating rules set 1363365314
000001520	15/03/2013 16:35:17	15/03/2013 16:35:17	Master	Master	Node Manager	Information	Successfully sent new rules files to 'Master'
000001521	15/03/2013 16:35:17	15/03/2013 16:35:17	Master	Master	Node Manager	Information	Distribution complete. All nodes successfully update
000001522	15/03/2013 16:35:17	15/03/2013 16:35:17	Master	Master	Node Manager	Information	Rules distribution complete. (Attachment 1) *** A

Figure 4-49 Log Manager screen

Logs can be filtered and searched using the criteria controls at the top of the screen, allowing you to build up queries by combining and constraining the various fields that can be displayed. Each line within the **Criteria** group box represents a search condition and consists of:

- A combo box containing the field that should be searched. Each time a new field name is selected, a new search term is inserted below.
- A combo box containing the type of comparison to perform. For example, it may be necessary to specify either 'is less than' or 'is more than'.

- Another control allowing the user to enter the value that the field should be compared with when the search is conducted.

If a number of these conditions are entered it indicates that returned log messages should obey every one (i.e. the logical AND) of the entered conditions. A search condition can be deleted by clicking the red 'X' button to its right.

Once appropriate criteria have been chosen, the search is initiated by pressing the **Search** button. Results are displayed as they arrive within the **Results** list. If a search is taking too long to complete or was made in error, then clicking the **Cancel** button will abort it. The search criteria can be cleared by clicking the **Reset** button; the search is actually reset to show all log levels from midnight on the date the logging process started. The current search criteria can be hidden or shown by clicking the **Collapse / Expand** button.



Note: Search criteria that are hidden from view still affect what is displayed. Thus although a single search term is shown, other terms may be limiting what logs are displayed. Click the **Expand** button to ensure there are no other terms in effect.

A brief summary of each log field and its meaning follows:

Log Field Name	Description
Sequence Number	All logs have a sequence number associated with them representing the order in which they arrived at the master node.
Master Date	The date at which the log message was stored within the master node's log database.
Local Date	The date at which the message was actually logged by the machine that generated the message. Due to network speed, connectivity issues or different time zones this may differ from the Master Date.
Log Origin	The name of the machine that logged the message. Using the search criteria it is possible to limit search results to only return those messages that were logged by a particular machine, or by a particular group of machines.
Data Source	The machine from which the protected data is originating, regardless of whether the log is generated by a source or destination node.
Category	The Data Instance Director process or thread that generated the log. A special category named 'Session' is used to indicate the start and end of a

Log Field Name	Description
	sequence of log entries relating to a specific transaction or job. Refer to Session Viewer below for details.
Level	How important the log message is. This can be one of four values:  Detail – These log messages may describe intricacies of a process, or non-critical information and do not require any action by the user.  Information – These messages give a higher-level overview of a process, and usually indicate normal operation. These messages do not require any action by the user.  Warning – Messages logged at warning level may indicate that the user should investigate the subject of the message to see if it requires further action. Warning messages are not necessarily critical.  Error – Error messages are only logged when the attention of the user is required. They appear with a red background and cause an error badge to appear on relevant nodes within the Monitor screen. The error must be acknowledged in order to clear this badge (see the Error Acknowledgement field below). When searching using this field it is possible to specify that returned logs should have a level 'less than' or 'more than' a given level, in terms of their severity.
Log Message	The actual text content of the log message. When searching this field, all words that are entered in the search box must appear in the log message text in order for the log to be returned. A paperclip icon  displayed at the beginning of a log message indicates that an attachment is included with the log entry that provides further information that may aid understanding. Double click anywhere on the message text to open the attachment.

Log Field Name	Description
Error Acknowledgement	Double click on this field to acknowledge any log message and clear the error indicator badge in the Monitor screen. A note can then be entered and the date & time is appended automatically.

Log Results area

The **Results** area shows log messages in a table. By default, the log messages are sorted in ascending order by sequence number. Any log messages that are generated while the **Results** area is being viewed are automatically inserted in the table at a point dictated by the current sort order.

To change which columns are shown for each result, right-click anywhere on the table headings and check or uncheck the applicable menu item.

By right-clicking on a log message it is possible to create a notification filter based upon it. See the section on Additional Notification below for more information.

You can sort the log messages by clicking on the field header (Figure 4-50) on the column you want to sort by. The currently sorted field is denoted with an up or down arrow next to the field name, indicating ascending or descending order.

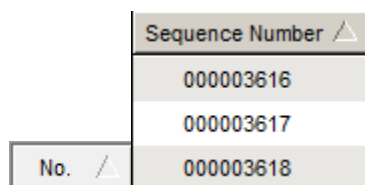


Figure 4-50 Sortable field headers

Log messages are displayed with alternating background shading to aid readability, as shown in the top two logs of Figure 4-51. Unacknowledged errors are displayed with a red background to make them more visible. Acknowledged errors are shown with a green background.

000000038	2008-09-03 02:33:07	fileserver1	132	Replication		User has requested a resync	
000000039	2008-09-03 02:33:10	fileserver1	132	Replication		User has requested a forced resync	
000000043	2008-09-03 02:33:12	fileserver1	132	Replication		Failed to log snapshot request	Error Acknowledged
000000051	2008-09-03 02:33:16	fileserver1	132	Replication		Failed to log snapshot request	

Figure 4-51 Log messages

Some logs are considered an auditable event. These logs cannot be deleted. An undeletable (immutable) log has a lock icon on the far left.

Session Viewer

A session is all the activity related to a specific job, the logs from the session can be viewed in the **Session Viewer**. Logs belonging to sessions are identified by having a Category type of 'Resynchronizations', 'Restores' and 'Validations'. Sessions are bracketed by log entries having the Category type 'Session' at the start ('Initial session started') and end ('Initial session completed...'). Several sessions can however be interleaved in the main log, making it difficult to trace a specific one.

Compared to the **Log Manager** screen the **Session Viewer** dialog (Figure 4-52) will only display a subset of logs that are related to the same session (i.e. have the same session ID). The Session log headings can be configured by right-clicking on the table headings and selecting the required fields from the menu. Entries in the session log can be sorted by clicking on the corresponding heading.

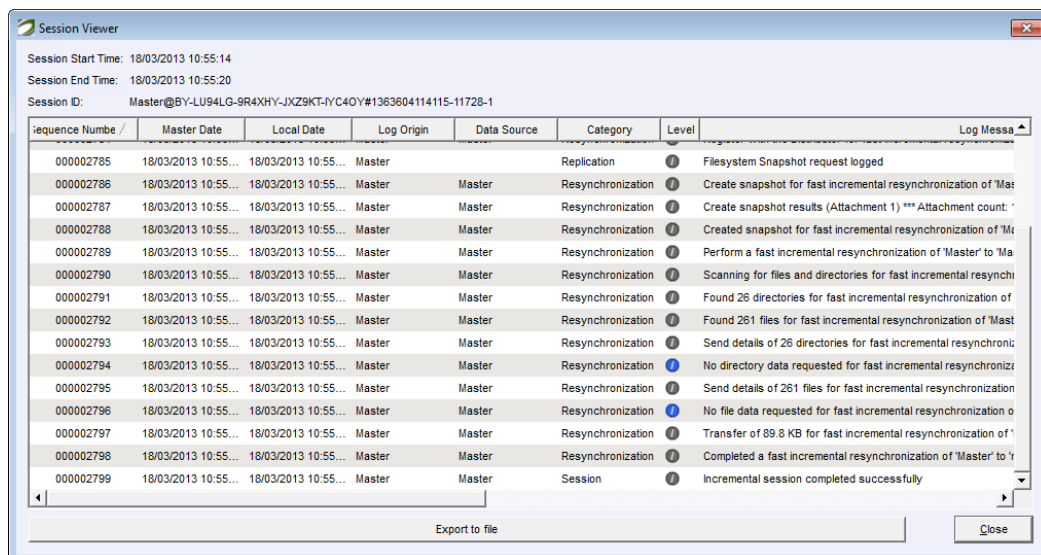


Figure 4-52 Log messages Session Viewer

To open the **Session Viewer** dialog, right-click on a log entry within the **Log Manager** screen and select **View Session** (Figure 4-53). If the selected log entry does not form part of a session then a dialog is displayed stating there is no session ID for the selected entry. The easiest way to find a session is to locate the corresponding Session start or Session end entry in the main log.



Figure 4-53 Log Manager Context Menu

Logs in the session viewer can be exported by clicking **Export to file**, and can be exported in HTML, PDF, or plain text formats.

Purging (deleting) logs

Entries can be deleted from the log database up to midnight on a specified date:

1. Select **Tools** → **Log Administration** from the main menu to display the **Log Administration** dialog (Figure 4-54).
2. Select Purge from the **Choose action** combo box. By default, the cut-off date is set to be two weeks before the current date. To change this cut-off date, select a date from the **From date and older** calendar control.

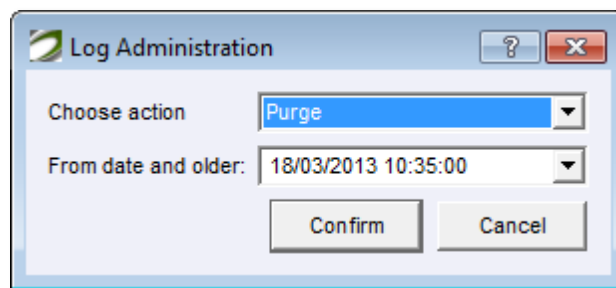


Figure 4-54 Log Admin function – purge action

3. After setting the required parameters, click **Confirm** to purge the log messages.



WARNING: Once purged, the log messages cannot be recovered, so be careful not to purge important log messages. Consider exporting logs before purging if in any doubt. Immutable logs  cannot be purged.

When the Master finishes purging log messages, the **Log Manager** screen will need to be refreshed manually by reviewing the search **Criteria** then clicking the **Search** button at the top of the **Log Manager** screen.

Clearing (acknowledging) all error logs

It is possible to acknowledge all outstanding errors in a single action rather than going through and acknowledging each one individually:

1. Select **Tools** → **Log Administration** from the main menu to display the **Log Administration** dialog (Figure 4-55).
2. Select **Clear All Errors** in the combo box. Enter an **Acknowledgement text**, the Acknowledgment will be appended with the date and time of acknowledgment upon confirmation.

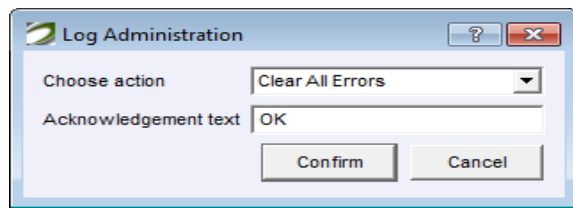


Figure 4-55 Log Admin Function – Clear All Errors Action

3. After entering an acknowledgement, click **Confirm** to acknowledge all the error messages simultaneously.

Exporting logs

To export the logs currently displayed in Log Manager to a text file, click on the **Save** icon  on the tool bar. This option displays the **Save file** dialog (Figure 4-56).

1. Within the **Save file** dialog, there is the **Save as type** option to save the logs as an HTML document, PDF document or a TXT plain text file.
2. Enter a **File name** and choose a location in the dialog.
3. Click **Save**.

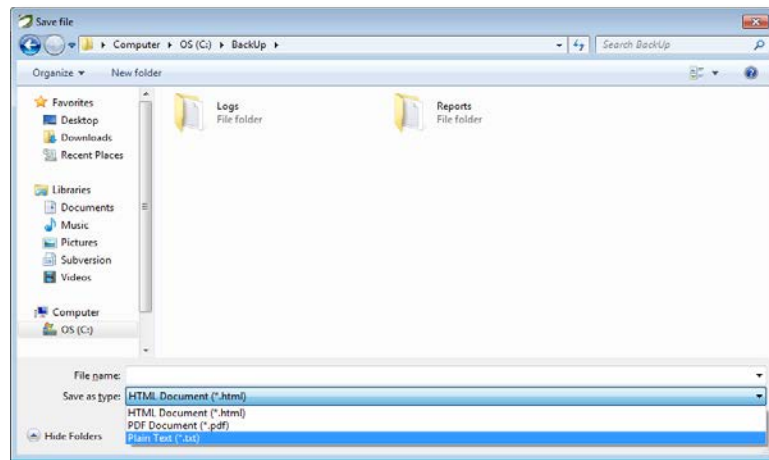


Figure 4-56 Export Logs (Save File) Dialog



Note: Only the logs currently displayed in the **Log Manager** are exported.

Notification Management dialog

Data Instance Director can send log messages by email and they can be entered into the operating system's event log. To access this feature, select **Tools** → **Notification Management** from the main menu to display the **Notification Management** dialog (Figure 4-57).

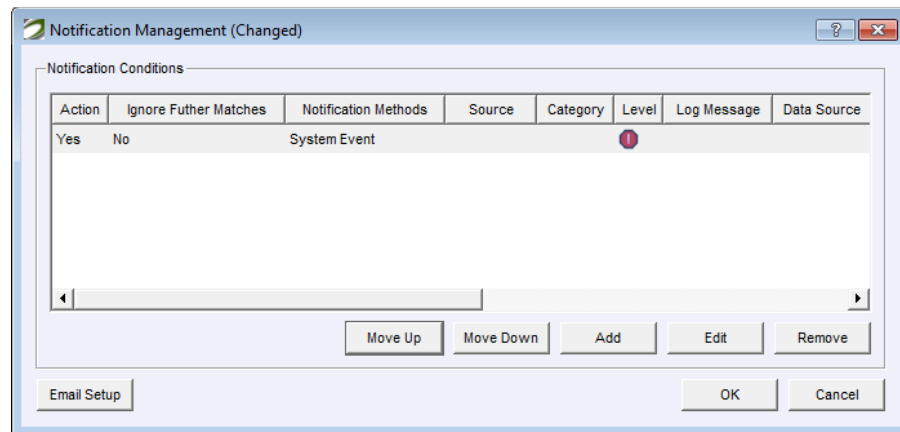


Figure 4-57 Notification Management Dialog

The **Notification Management** dialog shows a list of circumstances in which a notification may occur. Whenever a log message is received, it is checked against each item in the list, working downwards from the top. Each condition can issue a notification depending on whether the log message meets certain criteria and, if necessary, can stop any further notifications lower down the list being triggered. Expand the dialog horizontally or scroll left to view the extensive set of conditions that can be set.

Email setup

The email accounts from which notifications will be sent and by whom they will be received is configured by clicking the **Email Setup** button in the lower left of the **Notification Management** dialog (Figure 4-58).

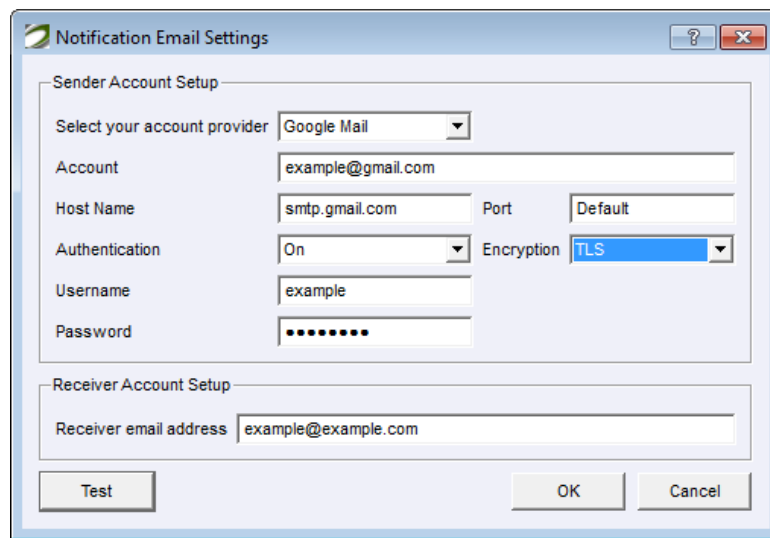
The image shows a Windows-style dialog box titled "Notification Email Settings". It is divided into two main sections: "Sender Account Setup" and "Receiver Account Setup". In the "Sender Account Setup" section, there is a dropdown menu for "Select your account provider" with "Google Mail" selected. Below this are text input fields for "Account" (containing "example@gmail.com"), "Host Name" (containing "smtp.gmail.com"), "Port" (containing "Default"), "Authentication" (a dropdown menu with "On" selected), "Encryption" (a dropdown menu with "TLS" selected), "Username" (containing "example"), and "Password" (containing masked characters). The "Receiver Account Setup" section has a single text input field for "Receiver email address" containing "example@example.com". At the bottom of the dialog are three buttons: "Test", "OK", and "Cancel".

Figure 4-58 Notification email settings dialog

Select an email account provider for the sender from the combo box, to automatically fill in the server settings. Alternatively, select 'Other' and enter these details manually. Finally, enter the Sender Account, Username, Password and Receiver email address. To specify multiple recipient addresses separate each one with a space.

Click the **Test** button to verify that an email can be sent from the sender to the recipient addresses.

Configuring notifications

To add a new rule to the list, click the **Add** button. To modify an existing rule, select it and then click the **Edit** button. Both of these methods will display the **Add Notification Condition** dialog (Figure 4-59). Right-clicking a log and selecting **Create Notification Condition** will also launch this dialog.

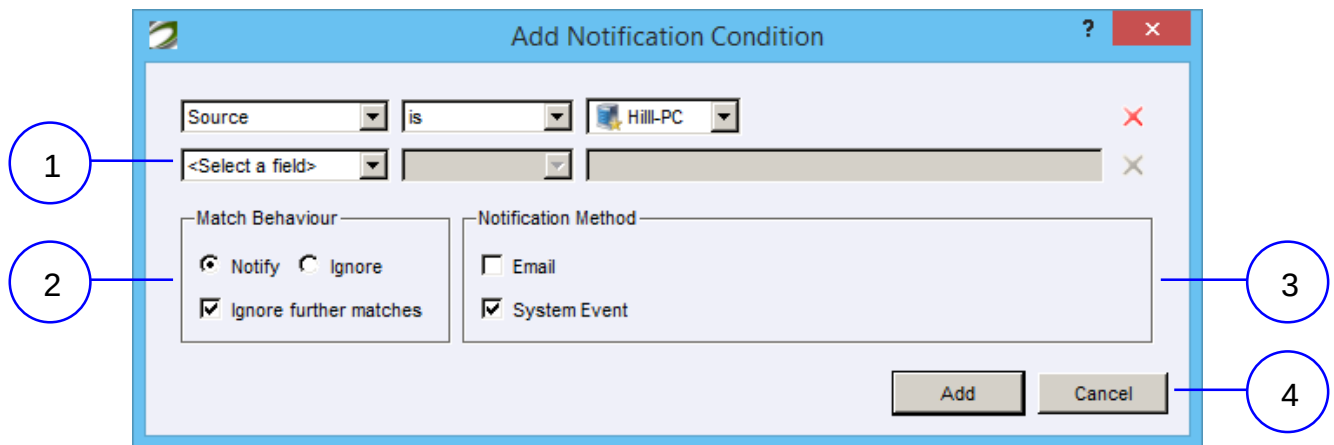


Figure 4-59 Edit Notification Condition dialog

1

Add Condition

Here you can select the condition that new log messages must match in order to trigger a notification. This works the same way as the search criteria controls on the **Log Manager** screen (i.e. a set of terms of the form 'field name – condition – value', all of which must be matched to generate a notification). Conditions can also be generated automatically from existing log messages by right-clicking on a log message in the **Log Manager** screen and selecting **Create Notification Condition**. This will generate a rule with conditions that will catch similar log messages.

2

Match behavior

This section defines what will occur when the inputted criteria above are matched.

- **Notify** – If selected then some kind of notification will be triggered when a condition is matched. If the checkbox–**Ignore further matches** is checked then no rules that appear below this one in the main notification list will be considered.
- **Ignore** – If selected then no notification will be triggered when a condition is matched.

3

Notification method

- **Email** – An email will be sent to a configured email account whenever a log meets the defined criteria. Currently there is no user interface for setting up the account details for emails to be sent to.
- **System event** – This will create a system event entry on the master when a matching log is found. The system event log can be found in the Application log section within the Event Viewer on a Windows system, and in /var/log/messages on a Linux system.

- **Custom handlers** – It is possible for advanced users to write their own custom handlers to add to this list. Refer to the section on [Customizing Alert Notifications](#) for details.

4

You can confirm the addition of the new rule by clicking Add, and cancel without saving by clicking **Cancel**.

Reports screen

The **Reports** screen allows the viewing of custom metadata that Data Instance Director processes can include when logging, in order to enrich the information they provide. This metadata is presented in a number of user selectable report views.



To access **Reports**, click the **Reports** icon in the **Navigation** bar.

When a report is selected from the **Select report** combo box at the top of the screen, the available **Filter** criteria adapt to the report type selected. The manner in which search terms are constructed is, in some cases, similar to those in the **Log Manager** screen.

Eight reports are available:

Destination RPO Report –For each destination node, this report (Figure 4-60) shows whether the RPO for the respective operations received from each source node have been met. Traffic lights are displayed to alert the user to RPOs that are out of specification by more than the amount shown in the **Warn if RPO is missed by** edit box with spin buttons. Checking **Hide operations which have met RPO** displays only the out of specification operations. This report applies to both Backup and CDP policies. It is useful for analyzing scenarios where many source nodes backup to one repository.

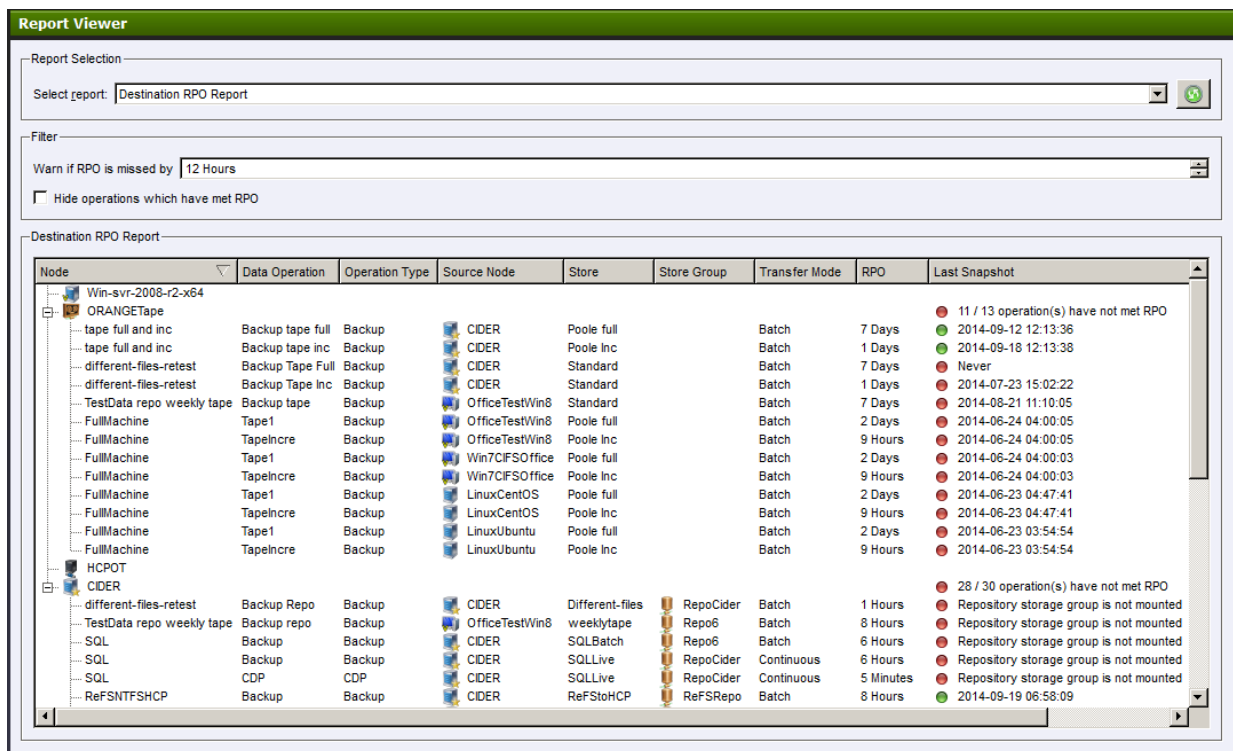


Figure 4-60 Reports Viewer: Destination RPO Report

- Source RPO Report** – For each source node, this report (Figure 4-61) shows whether the RPO for the respective operations sent to each destination node have been met. Traffic lights are displayed to alert the user to RPOs that are out of specification by more than the amount shown in the **Warn if RPO is missed by** edit box with spin buttons. Checking **Hide operations which have met RPO** displays only the out of specification operations. This report applies to both Backup and CDP policies. It is useful for analyzing scenarios where one source node backs up to many repositories.

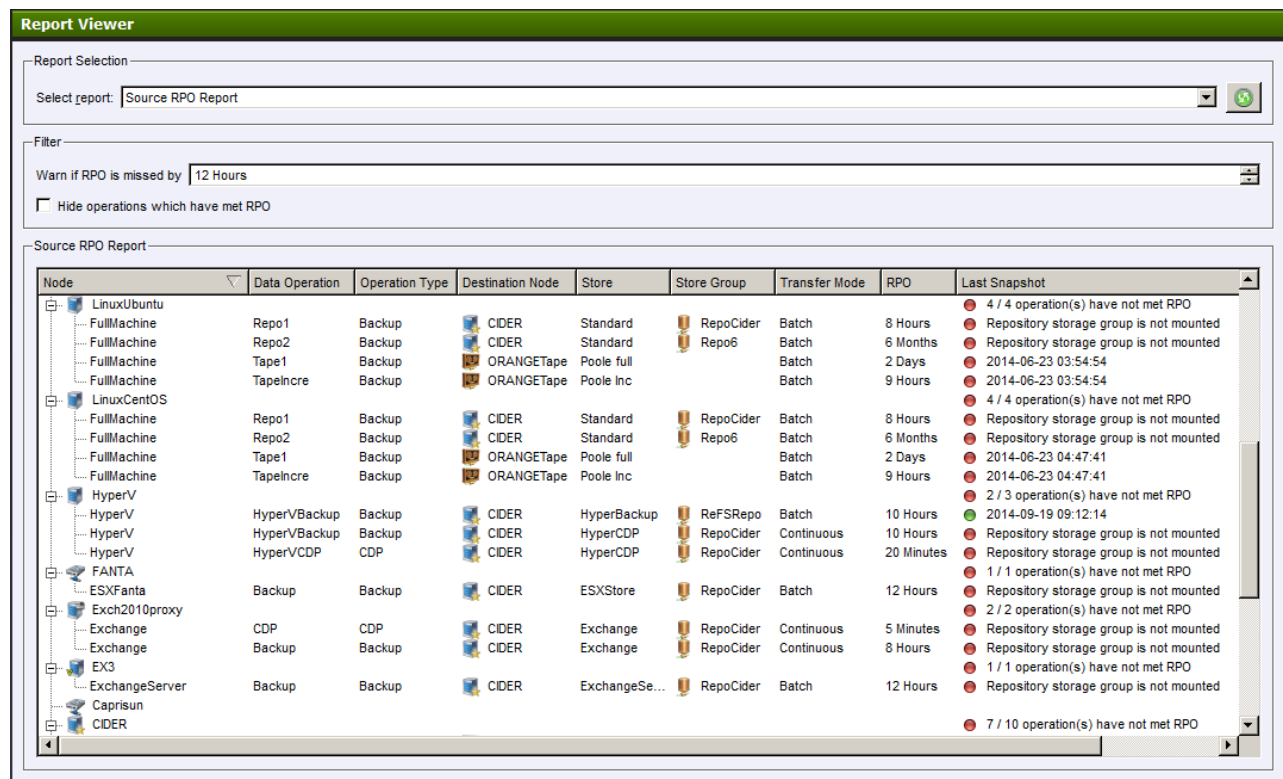


Figure 4-61 Reports Viewer: Source RPO Report

- Destination Policy Deployment** – This report (Figure 4-62) provides a graphic representation of Destination Nodes and which operations are being performed on a per source node and group basis. The Operation Type, Source Node, Store, and Transfer Mode, such as Continuous or Batch, are all organized for fast checking. This report has no associated filters.

Report Viewer						
Report Selection						
Select report: Destination Policy Deployment						
Destination Policy Deployment						
Node	Data Operation	Operation Type	Source Node	Store	Store Group	Transfer Mode
Win-svr-2008-r2-x64						
ORANGETape						
tape full and inc	Backup tape full	Backup	CIDER	Poole full		Batch
tape full and inc	Backup tape inc	Backup	CIDER	Poole Inc		Batch
different-files-retest	Backup Tape Full	Backup	CIDER	Standard		Batch
different-files-retest	Backup Tape Inc	Backup	CIDER	Standard		Batch
TestData repo weekly tape	Backup tape	Backup	OfficeTestW...	Standard		Batch
FullMachine	Tape1	Backup	OfficeTestW...	Poole full		Batch
FullMachine	Tape1	Backup	OfficeTestW...	Poole Inc		Batch
FullMachine	Tape1	Backup	Win7CIFSof...	Poole full		Batch
FullMachine	Tape1	Backup	Win7CIFSof...	Poole Inc		Batch
FullMachine	Tape1	Backup	LinuxCentOS	Poole full		Batch
FullMachine	Tape1	Backup	LinuxCentOS	Poole Inc		Batch
FullMachine	Tape1	Backup	LinuxUbuntu	Poole full		Batch
FullMachine	Tape1	Backup	LinuxUbuntu	Poole Inc		Batch
HCPOT						
ReFSNTFSHCP	Tier	Tier	CIDER	HCPOT		Batch
NTFSReFSHCP	Tier	Tier	CIDER	HCPOT		Batch
Exchange	Tier	Tier	Exch2010pr...	HCPOT		Batch
CIDER						
different-files-retest	Backup Repo	Backup	CIDER	Different-files	RepoCider	Batch
TestData repo weekly tape	Backup repo	Backup	OfficeTestW...	weeklytape	Repo6	Batch
SQL	Backup	Backup	CIDER	SQLBatch	Repo6	Batch
SQL	Backup	Backup	CIDER	SQLLive	RepoCider	Continuous
SQL	CDP	CDP	CIDER	SQLLive	RepoCider	Continuous
ReplicateHCP	Replicate	Replicate	CIDER	Default		Continuous
Replicate	Replicate	Replicate	Brendan-PC	Default		Continuous
Replicate	Replicate	Replicate	Alison-PC	Default		Continuous
Replicate	Replicate	Replicate	TaylorD-PC	Default		Continuous

Figure 4-62 Reports Viewer: Destination Policy Deployment

- Source Policy Deployment** – This report (Figure 4-63) provides a graphic representation of Source Nodes and which operations are being performed on a per destination node basis. The Operation Type, Destination Node, Store, and Transfer Mode, such as Continuous or Batch, are all organized for fast checking. This report has no associated search filters.

Report Viewer							
Report Selection							
Select report: Source Policy Deployment							
Source Policy Deployment							
Node	Data Operation	Operation Type	Destination Node	Store	Store Group	Transfer Mode	
vCenterDean							
Win7DroneShare							
CIFSBackup	BackupCIFS	Backup	CIDER	CIFS	EncryptedRepo	Batch	
Win7CIFSOffice							
FullMachine	Repo1	Backup	CIDER	Standard	RepoCider	Batch	
FullMachine	Repo2	Backup	CIDER	Standard	Repo6	Batch	
FullMachine	Tape1	Backup	ORANGETape	Poole full		Batch	
FullMachine	TapeIncr	Backup	ORANGETape	Poole Inc		Batch	
Win-svr-2008-r2-x64							
TaylorP-PC	Replicate	Replicate	CIDER	Default		Continuous	
CDPBackup	Backup	Backup	CIDER	Standard	EncryptedRepo	Continuous	
CDPBackup	CDP	CDP	CIDER	Standard	EncryptedRepo	Continuous	
Tanglefoot							
QAVCenter							
ESXQAVCenter	BackupQAVCenter	Backup	CIDER	ESXStore	RepoCider	Batch	
OfficeTestWin8							
TestData repo weekly tape	Backup repo	Backup	CIDER	weeklytape	Repo6	Batch	
TestData repo weekly tape	Backup tape	Backup	ORANGETape	Standard		Batch	
FullMachine	Repo1	Backup	CIDER	Standard	RepoCider	Batch	
FullMachine	Repo2	Backup	CIDER	Standard	Repo6	Batch	
FullMachine	Tape1	Backup	ORANGETape	Poole full		Batch	
FullMachine	TapeIncr	Backup	ORANGETape	Poole Inc		Batch	
LinuxUbuntu							
FullMachine	Repo1	Backup	CIDER	Standard	RepoCider	Batch	
FullMachine	Repo2	Backup	CIDER	Standard	Repo6	Batch	
FullMachine	Tape1	Backup	ORANGETape	Poole full		Batch	
FullMachine	TapeIncr	Backup	ORANGETape	Poole Inc		Batch	
LinuxCentOS							

Figure 4-63 Reports Viewer: Source Policy Deployment

- Store Status Changes** – This report (Figure 4-64) displays logs relating to the resynchronization status of destination machines. Whenever a destination goes out of sync with one of its sources for whatever reason, a log will be shown here. This makes it easy to see when and why a specific machine went in and out of sync. Search terms can be constructed, based on the field names in the report, using the **Filter** controls.

Report Viewer					
Report Selection					
Select report: Store Status Changes					
Filter					
<Select a field> 2014-09-05 00:00:00 Search Cancel Reset Expand					
Store Status Changes					
Date	Data Source	Destination	Store Name	Resync Req.	Event
2014-09-04 04:00:01	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-05 04:00:02	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-06 04:00:02	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-07 04:00:02	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-08 04:00:01	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-08 11:47:10	Alison-PC	CIDER	C:\FileSystem\	Yes	File operation failed
2014-09-09 04:00:01	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-09 04:07:05	Alison-PC	CIDER	C:\FileSystem\	Yes	Fast incremental resynchronization failed
2014-09-10 04:00:01	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-10 04:28:10	Alison-PC	CIDER	C:\FileSystem\	No	Fast incremental resynchronization succeeded
2014-09-11 04:00:02	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-12 04:00:01	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-13 04:00:02	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-14 04:00:01	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-15 04:00:02	TaylorP-PC	CIDER	C:\FileSystem\	Yes	AutoValidate
2014-09-15 07:55:27	TaylorP-PC	CIDER	C:\FileSystem\	No	Fast incremental resynchronization succeeded
2014-09-17 10:37:59	TaylorP-PC	CIDER	C:\FileSystem\	Yes	File operation failed
2014-09-18 04:08:56	TaylorP-PC	CIDER	C:\FileSystem\	No	Fast incremental resynchronization succeeded

Figure 4-64 Reports Viewer: Store Status Changes

- Tracking** – This report (Figure 4-65) provides a convenient display format for audit logs associated with tracking and blocking policies. In the **Log Manager** screen, the information provided by audit logs is presented in plain text in the log message field, making it harder to filter out unwanted information. However, in the **Reports** screen it is possible to search for audit logs based upon their unique metadata, such as the Process Name that caused the log, the User Action that the process was performing, and the location within the file system where the operation took place. Search terms can be constructed, based on the field names in the report, using the **Filter** controls.

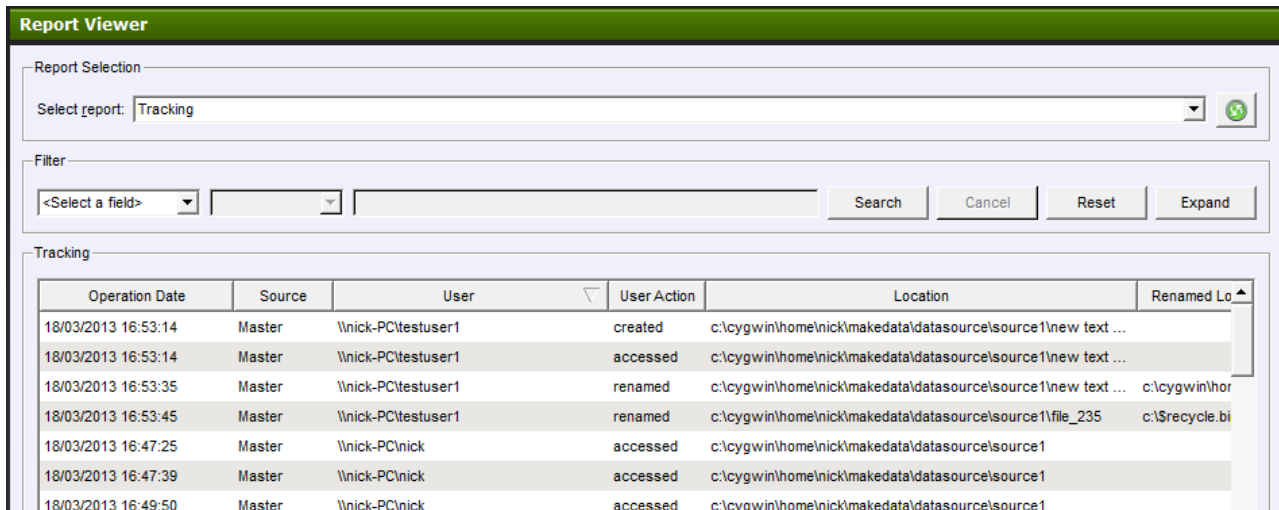


Figure 4-65 Reports Viewer: Tracking

- Session Summary** – This report (Figure 4-66) shows a history of the sessions that have taken place within the Data Instance Director system, including activity such as Resynchronization, Restore and Validation, and the details pertaining to that specific session, such as the type of session, the time the session started/finished and quantity of transferred data. This list of sessions can be filtered in a similar method to **Log Manager** to find a specific session or range of sessions. It is also possible to **View Session** details and **Create Notification Condition** (as per the **Log Manager** screen) from the context menu by right-clicking on an entry in the report.

Report Viewer

Report Selection

Select report: Session Summary

Filter

Master Date is after 2014-09-17 00:00:00 Search Cancel Reset Collapse

Activity equals Backup <Select a field>

Session Summary

Master Date	Local Date	Session Type	Source	Destination	Destination Store Type	Destination Store Group	Policy	Action
2014-09-18 23:01:03	2014-09-18 23:01:03	Incremental	CIDER	CIDER	Batch Filesystem	ReFSRepo	ReFSNTFSHCP	Backi
2014-09-18 23:12:57	2014-09-18 23:12:57	Incremental	HyperV	CIDER	Batch Filesystem	ReFSRepo	HyperV	Backi
2014-09-18 23:21:20	2014-09-18 23:21:20	Partial	TaylorP-PC	CIDER	Live Filesystem			Backi
2014-09-18 23:29:00	2014-09-18 23:29:00	Partial	TaylorP-PC	CIDER	Live Filesystem			Backi
2014-09-18 23:58:14	2014-09-18 23:58:14	Initial	LinuxCentOS	CIDER	Batch Filesystem	ORANGETape	FullMachine	Backi
2014-09-18 23:58:15	2014-09-18 23:58:15	Initial	LinuxUbuntu	CIDER	Batch Filesystem	ORANGETape	FullMachine	Backi
2014-09-18 23:58:16	2014-09-18 23:58:16	Initial	OfficeTestWin8	CIDER	Batch Filesystem	ORANGETape	FullMachine	Backi
2014-09-18 23:58:17	2014-09-18 23:58:17	Initial	Win7CIFSOffice	CIDER	Batch Filesystem	ORANGETape	FullMachine	Backi
2014-09-19 07:01:02	2014-09-19 07:01:02	Incremental	CIDER	CIDER	Batch Filesystem	ReFSRepo	ReFSNTFSHCP	Backi
2014-09-19 08:58:15	2014-09-19 08:58:15	Initial	LinuxCentOS	CIDER	Batch Filesystem	ORANGETape	FullMachine	Backi
2014-09-19 08:58:16	2014-09-19 08:58:16	Initial	LinuxUbuntu	CIDER	Batch Filesystem	ORANGETape	FullMachine	Backi
2014-09-19 08:58:17	2014-09-19 08:58:17	Initial	OfficeTestWin8	CIDER	Batch Filesystem	ORANGETape	FullMachine	Backi
2014-09-19 08:58:18	2014-09-19 08:58:18	Initial	Win7CIFSOffice	CIDER	Batch Filesystem	ORANGETape	FullMachine	Backi
2014-09-19 09:12:56	2014-09-19 09:12:56	Incremental	HyperV	CIDER	Batch Filesystem	ReFSRepo	HyperV	Backi
2014-09-19 09:52:53	2014-09-19 09:52:53	Partial	TaylorP-PC	CIDER	Live Filesystem			Backi

Figure 4-66 Reports Viewer: Session Summary

- Current Activity** – This report (Figure 4-67) displays a list of what resynchronizations are running or queued on the repository, as well providing details for these individual activities such as Source, Destination and when the resynchronization job started. This report has no associated search filters.

Report Viewer

Report Selection

Select report: Current Activity Report

Current Activity Report

Activity	Started	Status	Source	From	To	Policy	Operation
Fast Incremental Resynchronization	18/03/2013 17:13:16	Active	Master	Master	Master repository2		Standard Live Filesystem
Fast Incremental Resynchronization	18/03/2013 17:13:16	Pending	Master	Master	Master repository2		Deduplicated Live Filesystem

Figure 4-67 Reports Viewer: Current Activity Report

Storage screen

This screen (Figure 4-68) displays storage information relating to repository nodes, tape libraries and their snapshots displayed within the 6 available tabs:

- Repositories
- Repository Usage
- Managed Storage
- Library Management
- Media
- Media Requests



To access the **Storage** screen, select the **Storage** icon in the **Navigation** bar.

A screenshot of the Storage screen in the Data Instance Director. The top bar is green and labeled 'Storage'. Below it are six tabs: 'Repositories', 'Repository Usage', 'Managed Storage', 'Library Management', 'Media', and 'Media Requests'. The 'Repositories' tab is selected. A filter section shows 'Repository node' set to 'All' and 'Snapshot range' from '2014-09-05 00:00:00' to '2014-09-20 00:00:00'. The main area displays a table of repository results. The table has columns: Item, Information, File Count, Directory Count, Policy, Data Set Size, Used Size, Physical Size, and Free Disk Space. The data shows a hierarchy starting with 'CIDER' (Repository Machine), followed by 'EncryptedRepo (Unmounted)' (Repository), 'ReFSRepo (Mounted)' (Analyzed at 2014-01-06 13:42:25), and 'CIDER' (Data Source). Under 'CIDER' is 'ReFSStoHCP' (Batch Filesystem Store), which lists numerous 'Policy Snapshot' entries with their respective file counts, directory counts, and policies (all 'ReFSNTFSHCP'). At the bottom, 'HyperV' is listed as a 'Data Source' and 'Windows' as a 'Batch Filesystem Store'.

Figure 4-68 Storage screen with the Repositories tab selected

Repositories tab

The **Repositories** tab presents the available repositories as a forest. The trunk or each tree represents a machine that hosts a repository, the first branches of each tree are the storage groups within that repository, then each data source that backs up to that storage group, then each type of store (based on the repository templates in use), and finally the list of snapshots relating to the policies that write to that store.

The **Filter** controls enable the user to search for all repository nodes or just one specific repository node. It is also possible to set a date range to filter which snapshots are displayed in the **Repository Results** tree view. Click the **Filter** button to display a hierarchical list of repositories and snapshots that meet the criteria entered. By default, all snapshots in all repository nodes are displayed.

-  The trunk nodes in the Repository Results forest represent machines with the Repository component installed. Each Repository node can have multiple store groups. By right-clicking on the repository machine a context menu with one option is displayed:
 - **Create Store Group** –Displays the **Create Repository on Master** dialog (Figure 4-69) to create a new store group by specifying a **Repository Name** and an existing or new empty folder (**Repository Root Path**). Data Instance Director will automatically create repository configuration files within that storage area for a new store group. If **Mount Store Group** is not checked, the store will remain unmounted once created. A repository node can have multiple store groups mounted at the same time. A new store group can be created as an encrypted repository by checking the **Encrypt Repository** option, a **Repository Passphrase** will have to be entered twice.



WARNING: An encrypted repository CANNOT be recovered. If the passphrase is forgotten then the data in that repository will be inaccessible.

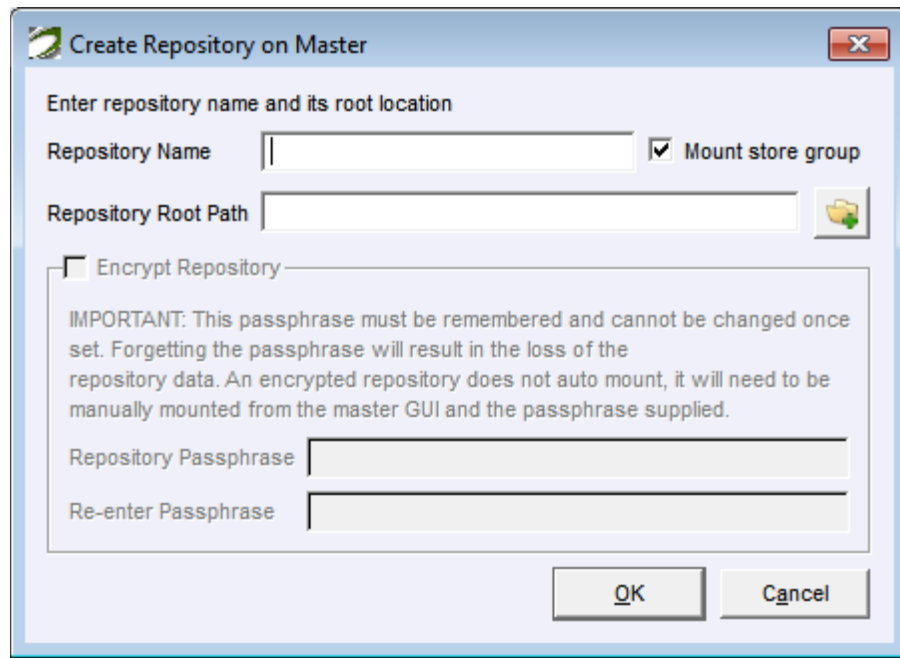


Figure 4-69 Creating a new store group

-  The next level down in the **Repository Results** tree represents the individual store groups. Each store group has a used size, a physical size and the free disk space associated with it. The physical size is the amount of space the store group has allocated to it on the disk it is contained on. The used size is the amount of space within the store group occupied by backed up data. The free disk space is the amount of free space left on the disk the store group is mounted on. By right-clicking on a store group, a context menu with further options is displayed:
 - **Analyze Used Store Size** – calculates the used size of each store on the store group.
 - **Details** –displays a dialog showing further details specific to the store group.
 - **Mount** –displays the **Mount Repository** dialog (Figure 4-70) which enables the mounting of a store group that is currently unmounted. The store group can be mounted to either its original location or to an alternative location. If the store group is being mounted to an alternate location a prompt (Figure 4-71) will ask for the creation of a new configuration file and, when mounted, a full resynchronization will be required.
 - **Unmount** –unmounts the store group and closes the repository files. An unmounted repository has the  icon.
 - **Remove Store Group** – This deletes the storage group from the forest. The associated files are not deleted from the disk.

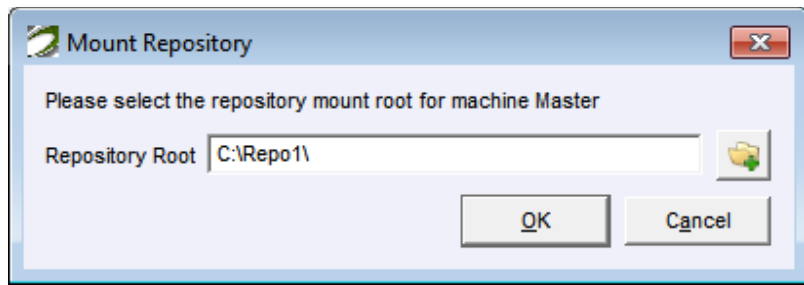


Figure 4-70 Select a path for mounting a repository

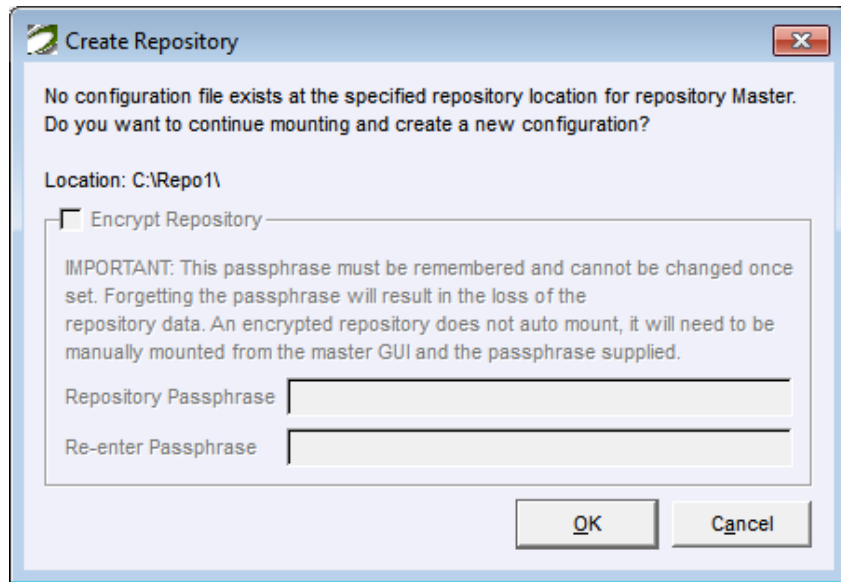


Figure 4-71 Warning displayed when selecting a new store group path

-  The next level down the tree has all the machines connected to the store group, all of which will have the Data Source component installed. There is no context menu for this item type.
-  The next level down represents the stores in the store group that contain data from the given data source. This will just be the default (Standard) store group unless extra store templates were manually configured in the data flow definition. (Note that the **ID** column indicates the name of the Namespace on HCP or the BLOB Container on Azure. By default this column is hidden.) By right-clicking on the store a context menu is displayed:

- **User Snapshot** – creates a single snapshot containing all the data in the store at that point in time. Snapshots of this type are shown with a  next to them. Note that user snapshots are retained forever or until they are manually deleted. The snapshot is of the whole store – not just an individual policy. The data from all policies whose destination is that store will be contained within the same user snapshot.
- **Delete** – This option is only available if the store is inactive – that is, no active policies currently transfer data to a store with this name. Selecting it will delete this store and all the snapshots contained within it.

The word 'Live' in the **Policy** column indicates that the store is receiving data as and when it changes. This means the store is connected to its source on the data flow by a **Continuous** mover, as opposed to a **Batch** mover.

- Finally, snapshots are displayed as leaf nodes of the tree. A green icon  indicates a successful policy snapshot, a yellow icon  indicates a policy snapshot that occurred while the store group was not synchronized with its sources, and a blue icon  indicates a user snapshot. User snapshots are created by right-clicking on the data store and selecting **User Snapshot** (see above). The policy each snapshot is associated with is also shown. Further options are available by *right-clicking* on the snapshot icon:
 - **Calculate Data Set Size** – This can be used in order to find out the logical size of all files contained within the snapshot. The Data Set Size is the amount of space required to restore the snapshot, not the amount of space used to store it in the store group. This must be done on an individual snapshot basis and may take a few seconds to several minutes to complete depending on repository load and snapshot size.
 - **Details** – (Unavailable unless **Calculate Data Set Size** is selected first). Once the snapshot data set size calculation has been completed, this brings up the **Snapshot Details** dialog (Figure 4-72) and provides a number of statistics such as directory count, file count, etc. and is ideal for testing against known file system structures. It can also display a list of files that have been modified or are new to that particular snapshot.
 - **Get Modifications/New File List**– This button is at the top-right of the **Snapshot Details** dialog and updates the list of files and statistics, as shown below.

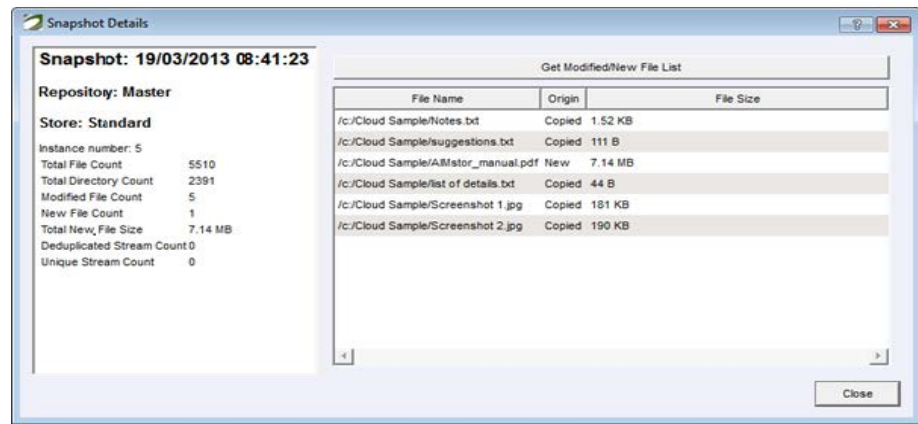


Figure 4-72 Snapshot Details Dialog

- **Delete** – This will schedule the snapshot for deletion from the store group. The snapshot may not be immediately deleted. The text '(Deleting)' is appended to the Information field for snapshots awaiting deletion. No further operations can be performed on snapshots pending deletion.

Repository Usage tab

The **Repository Usage** tab (Figure 4-73) displays a line graph of allocated and used repository space against time, and below that on a common time axis a set of event markers show when key events affecting repository size have occurred.

The mouse scroll wheel can be used to increase and decrease the scale of both axes simultaneously. Drag the graph area up and down or left and right with the mouse to change the Size and Time axis offsets respectively.

The **Criteria** controls allow you to filter results displayed on the graphs by date, repository node and storage group name. Click the **Refresh** to update the plotted data after changing any of the filter criteria.

The **Repository Size Trend** graph area displays the following data:

- **Total Repository Size** – Displays a line graph of the allocated capacity of a repository. As this space is used or freed up, this line steps up or down by preset increments (100Gb by default). The total size will always be greater than or equal to the used size.
- **Used Repository Size** – Displays a line graph of the actual logical size of the repository at any point in time. Typically this line will show rapid initial growth which rapidly reduces as the data change rate falls off.
- **Backup Retired** – Event markers are plotted along a common time axis indicating when backups have been retired, freeing up logical space within the repository.

- **Continuous Backup Created** – Event markers indicating the point in time when continuous backups were created.
- **Batch Backup Created** – Event markers indicating the point in time when batch backups were created.
- **Dedup Queue** – This graph shows the frequency and duration of I/O activity for file-level deduplication within the repository.

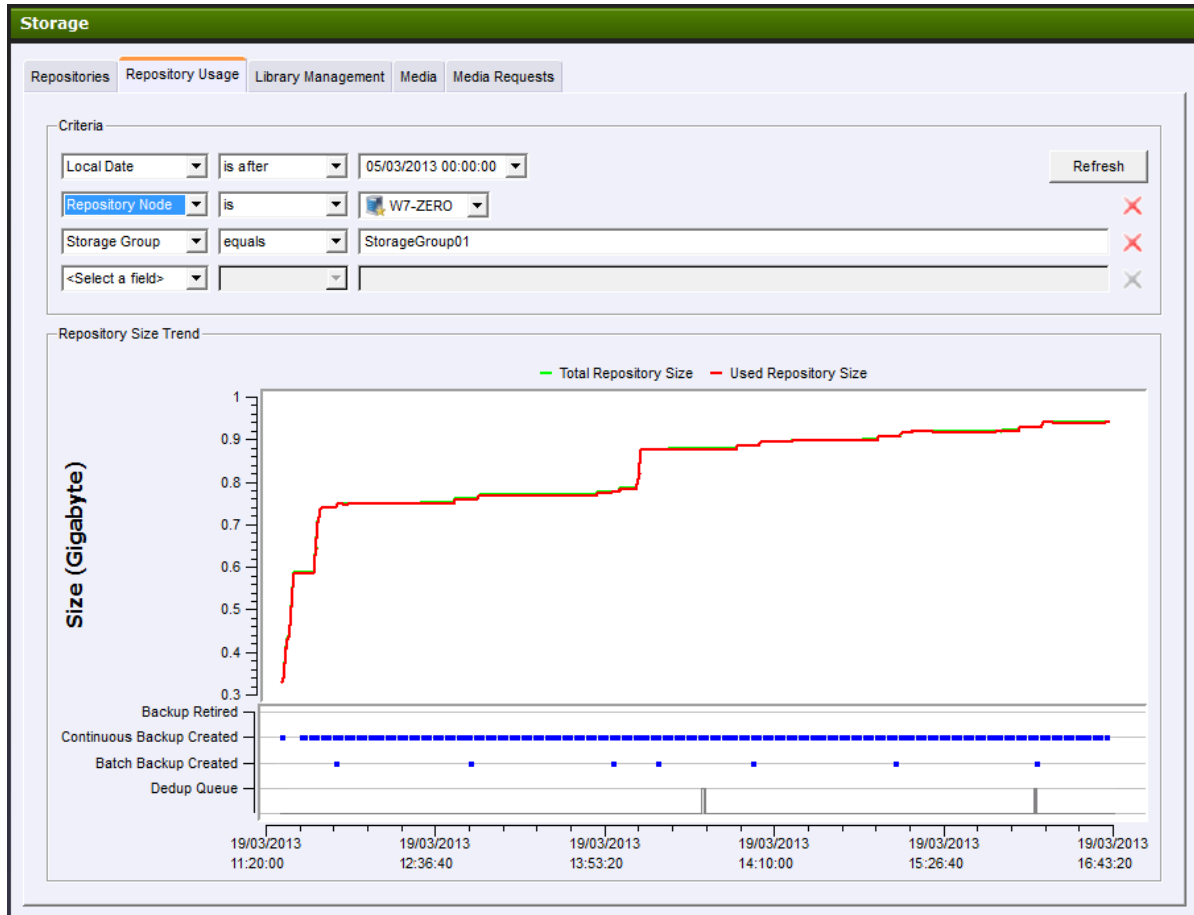


Figure 4-73 Repository Usage tab

Figure 4-74 provides another example of graphed Repository Usage. In this case, a system was completing backups of a set of data of approximately 250 GB. The data change rate was approximately 5% for each backup, and backups were retired after every three days.

After a time, the graph indicates that the repository growth has stopped, as it is accounting for the existing data change rate. (Data growth, however, will result in a gradual meandering of the plotted graph.)



Figure 4-74 Repository usage indicating change rates



Notes:

- The repository's size increases by the number of deltas that are retained. For example, if your site completes one backup each day, and that backup is retained for two days with a change rate of 5% per day, then the size of the repository would be 100% plus 5% increase per day.
- After a large set of data has been manually deleted, a large number of snapshots have expired, or a number of stores became inactive and were deleted from the repository, you will need to check the repository's status and remove the unused space. See Chapter 6, [Reducing the repository's size: repomaint.exe](#).

Managed Storage tab

The **Managed Storage** tab presents the available ISM (Intelligent Storage Manager) nodes as a forest. The trunk or each tree represents an ISM that manages block (VSP and HUS VM) or file storage (Hitachi NAS), the first branch of each tree are the storage devices managed by that ISM, then LDEV pools (for block storage devices only), then each data source that backs up to that storage device, and finally the list of snapshots and replications created by the operations within the associated policies.

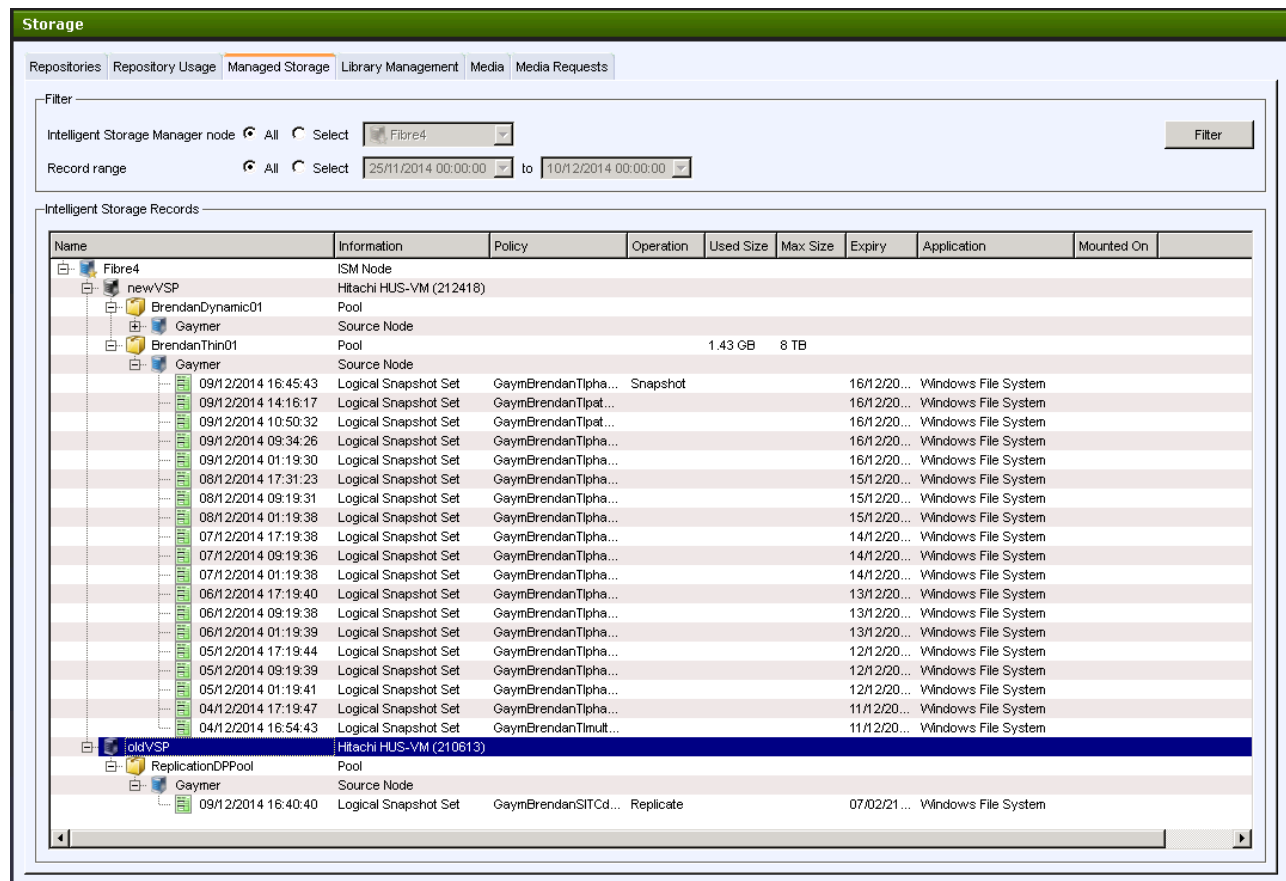


Figure 4-75 Managed Storage tab

The **Filter** controls enable the user to search for all ISM nodes or just one specific ISM node. It is also possible to set a date range to filter which snapshots and replications are displayed in the **Intelligent Storage Records** tree view. Click the **Filter** button to display a hierarchical list of ISMs and snapshots/replications that meet the criteria entered. By default, all ISM nodes are displayed.

-  The trunk nodes in the **Intelligent Storage Records** forest represent machines with the ISM component installed. Each ISM node can have multiple managed storage devices attached. There is no context menu for this item type.
-  The next level down represents block (VSP, VSP-G and HUS VM) and/or file storage (Hitachi NAS) devices managed by that ISM. There is no context menu for this item type.
-  If the storage is a block device then this level of the tree represents LDEV pools. Software snapshots (VSS and LVM) are placed in a separate branch. File storage devices do not have this sub-division. Right-clicking and selecting **Refresh** on this item enables the machines and snapshots listed under it to be refreshed.

-  The next level down has all the machines that have snapshots/replications stored in the LDEV pool or file storage device. There is no context menu for this item type.
-  Finally, snapshots/replications are displayed as leaf nodes of the tree. A **green** icon indicates a successful policy snapshot/replication, a **yellow** icon indicates a policy snapshot/replication that is inconsistent for some reason, and a **red** icon indicates an erroneous snapshot/replication. The policy each snapshot/replication is associated with is also shown. Further options are available by right-clicking on these items:
 - **Resources ...** - Displays details about the storage resources involved in the replication or snapshot.
 - **Pause/Resume** – Pauses and resumes an active replication.
 - **Mount** – This option allows you to expose volumes or mount the selected snapshot/replication. The mount operation may take several minutes to complete. The dialog shown in Figure 4-76 is displayed.

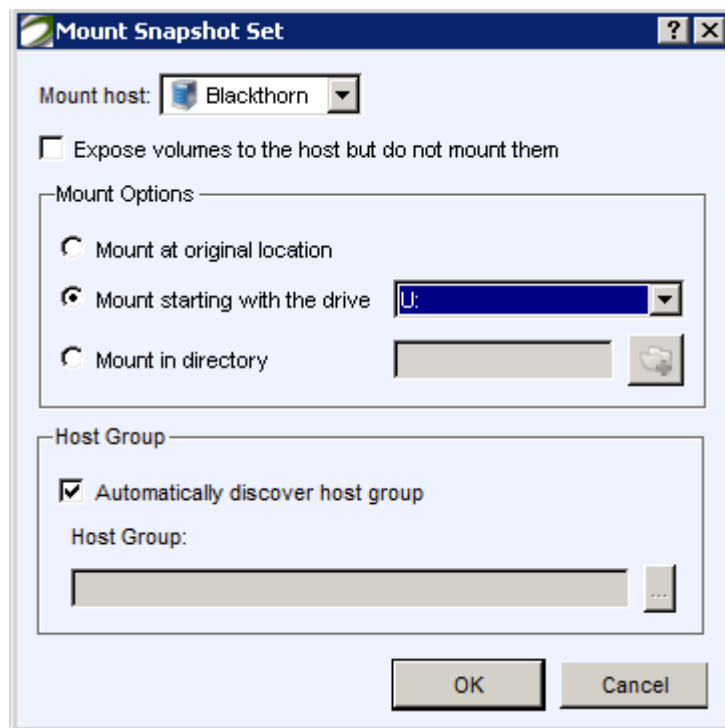


Figure 4-76 Mount dialog

- o **Mount host** – the machine to mount to or expose to.
- o **Expose volumes ...** – check this option if you want to expose the volumes and mount them later. For Hitachi NAS, Data Instance Director will create an export and log the command that needs to be run on the client to mount the volumes.

- o **Mount at original location** – snapshots are mounted at their original location.
 - o **Mount starting with the drive** – When mounting a snapshot that contains multiple volumes, the first volume will mount at on the specified drive and subsequent drives are used for each additional volume.
 - o **Mount in directory** – When mounting a snapshot that contains multiple volumes, each volume will be assigned a separate subdirectory.
 - o **Automatically discover host group** – automatically choose a host group to use to expose a snapshot to a machine.
 - o **Host Group** – manually specify a host group to use to expose a snapshot to a machine.
- **Unmount** – (Unavailable unless **Mount** is performed first). This option allows a previously mounted snapshot to be unmounted. The export is deleted. The unmount operation may take several minutes to complete.
 - **Revert** – This option allows the snapshot to be used to revert the volume(s) to a prior state, destroying any changes to data on those volumes that were made subsequent to the selected snapshot event. The revert operation may take several minutes to complete. The dialog shown in Figure 4-77 is displayed and the word REVERT must be explicitly typed in to confirm the action.

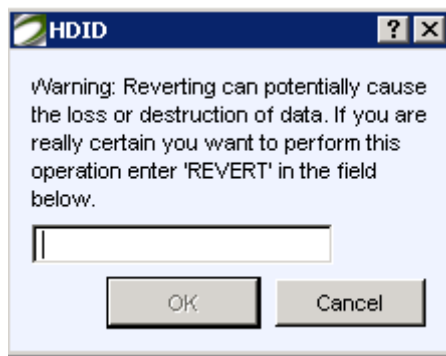
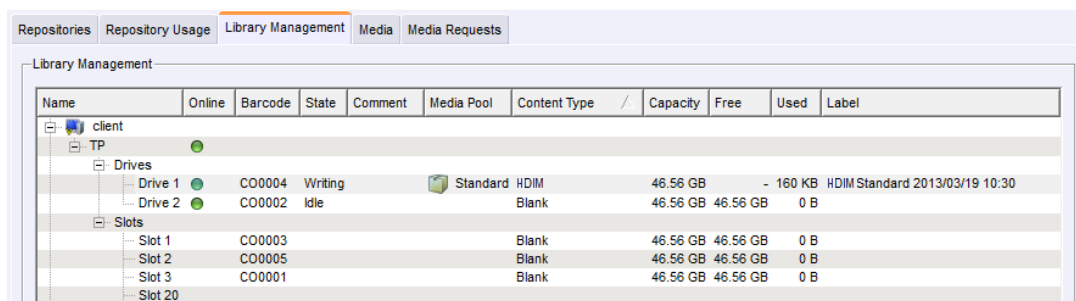


Figure 4-77 Revert dialog

- **Delete** – This will delete the snapshot from the storage device. A confirmation dialog is displayed to confirm the action. Replication records must be removed manually after they have been removed from the data flow. Records for replications currently in the data flow cannot be deleted.

Library Management tab

The **Library Management** tab (Figure 4-78) is used to manage any Tape Libraries and their associated tapes that are connected to Data Instance Director. The **Media Management** tree view presents the available Tape Libraries as a forest. The trunk or each tree represents a Machine that hosts a Tape Library, the first branches of each tree are the Tape Libraries attached to that Machine, then each Tape Library is organized into Drives that read and write to tapes and Slots that hold blank or used tapes. Some of the available parameters in the table heading are hidden by default, these can be displayed by right-clicking on the header bar and selecting the required items. The ordering of tapes listed in slots and drives can be changed by clicking the appropriate heading.



Name	Online	Barcode	State	Comment	Media Pool	Content Type	Capacity	Free	Used	Label
client										
TP										
Drives										
Drive 1		CO0004	Writing		Standard	HDIM	46.56 GB	- 160 KB	HDIM Standard 2013/03/19 10:30	
Drive 2		CO0002	Idle			Blank	46.56 GB	46.56 GB	0 B	
Slots										
Slot 1		CO0003				Blank	46.56 GB	46.56 GB	0 B	
Slot 2		CO0005				Blank	46.56 GB	46.56 GB	0 B	
Slot 3		CO0001				Blank	46.56 GB	46.56 GB	0 B	
Slot 20										

Figure 4-78 Library Management tab

Right-clicking on an occupied drive or slot gives a context menu of which several options are grayed out. This menu is also used within the **Media** tab, and all options detailed below are also available to be performed on tapes from within the **Media** tab. The available options are:

- **Unload** – (only available on occupied drives) Moves a tape from a drive to the first available empty slot.
- **Move...** – Displays a dialog (Figure 4-79) with options for moving the tape to a specific drive or a slot, if one is available.

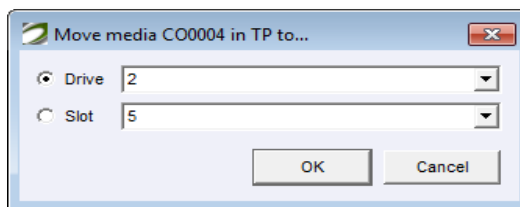


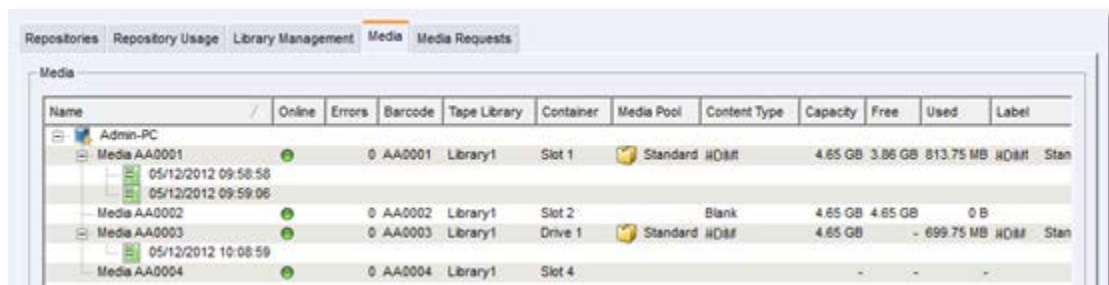
Figure 4-79 Moving media

- **Move to pool...** – Displays a dialog with options for allocating a tape to a specific media pool. This option will erase all existing data on the tape.

- **Import Backups** – Scans the tape to find backups and imports any Data Instance Director backups. If the tape is in a slot it is moved to a drive automatically.
- **Set Media Properties...** – Allows a comment and an offsite location to be added to the tape.
- **Reset Error Count** – Resets the error count for this tape, the error count (and other parameters) for a tape can be seen by right-clicking on the table header and selecting Errors.
- **Scan** – Forces the tape header to be scanned. This is done automatically if automatic inventorying is enabled.
- **Erase** – Displays a confirmation dialog to confirm the operation, then erases the tape.

Media tab

The **Media** tab (Figure 4-80) displays any tapes known to (i.e. in the media database of) Data Instance Director and any snapshots they contain. These snapshots are organized in a tree structure by machine name then tape name (media). Many of the available parameters in the table heading are hidden by default. These can be displayed by right-clicking on the header bar and selecting the required items. The ordering of tapes and snapshots can be changed by clicking the appropriate heading.



Name	Online	Errors	Barcode	Tape Library	Container	Media Pool	Content Type	Capacity	Free	Used	Label
Admin-PC											
Media AA0001	●	0	AA0001	Library1	Slot 1	Standard	HDiff	4.65 GB	3.86 GB	813.75 MB	HDiff Stan
05/12/2012 09:58:58											
05/12/2012 09:59:06											
Media AA0002	●	0	AA0002	Library1	Slot 2		Blank	4.65 GB	4.65 GB	0 B	
Media AA0003	●	0	AA0003	Library1	Drive 1	Standard	HDiff	4.65 GB	-	699.75 MB	HDiff Stan
05/12/2012 10:08:59											
Media AA0004	●	0	AA0004	Library1	Slot 4						

Figure 4-80 Media Tab

When right-clicking a Tape within this table, the same options are presented as when selecting a Tape in the **Library Management** tab, however, right-clicking a snapshot within the tape, presents two previously grayed out options:

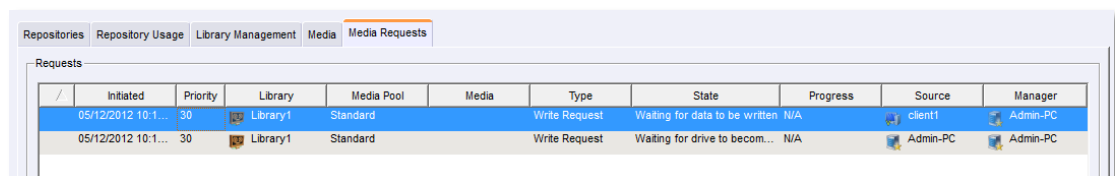
- **Import Data Index** – Imports the data index for a selected snapshot. The data index allows the snapshot to be browsed and restored in the **Restore** screen.
- **Remove** – deletes a tape from the media database. The tape cannot be removed from the database if it is currently loaded in a slot or drive.

- **Delete Data Index** – Deletes the data index for the selected snapshot. This will prevent the snapshot from being browsed and restored, it also removes the index from the media database held on disk. A snapshot with no data index has the  icon.

Media Requests tab

The **Media Requests** tab (Figure 4-81) displays a list of any jobs that the Media Manager is running and their status. Media request can also be viewed by selecting the Tape Library node within the **Monitor** screen and selecting the **Tape Request** tab.

Media requests can be canceled by right-clicking a request and selecting **Cancel**.



Initiated	Priority	Library	Media Pool	Media	Type	State	Progress	Source	Manager
05/12/2012 10:1...	30	Library1	Standard		Write Request	Waiting for data to be written	N/A	client1	Admin-PC
05/12/2012 10:1...	30	Library1	Standard		Write Request	Waiting for drive to become...	N/A	Admin-PC	Admin-PC

Figure 4-81 Media Request Tab

Restore screen

The **Restore** screen (Figure 4-82) is used for viewing and restoring the versions and snapshots currently stored in the various repositories and tape libraries connected to a Data Instance Director network.

To access the Restore screen:



- Click the **Restore** icon in the Navigation pane on the master Data Instance Director GUI.

– or –

- Run the Data Instance Director GUI from the node you want to restore data to. The GUI will automatically be restricted to just the **Restore** functionality unless the node is the master.

Various restrictions are placed on the capabilities of the **Restore** screen depending on whether it is run from a master or non-master node as follows:

Master Restore Capabilities	Non-master Restore Capabilities
Can search for data that originated from any machine within the Data Instance Director network.	Can only search for data that originated from itself.
Can restore data to any machine within the Data Instance Director network, regardless of its source.	Can only restore data to itself.

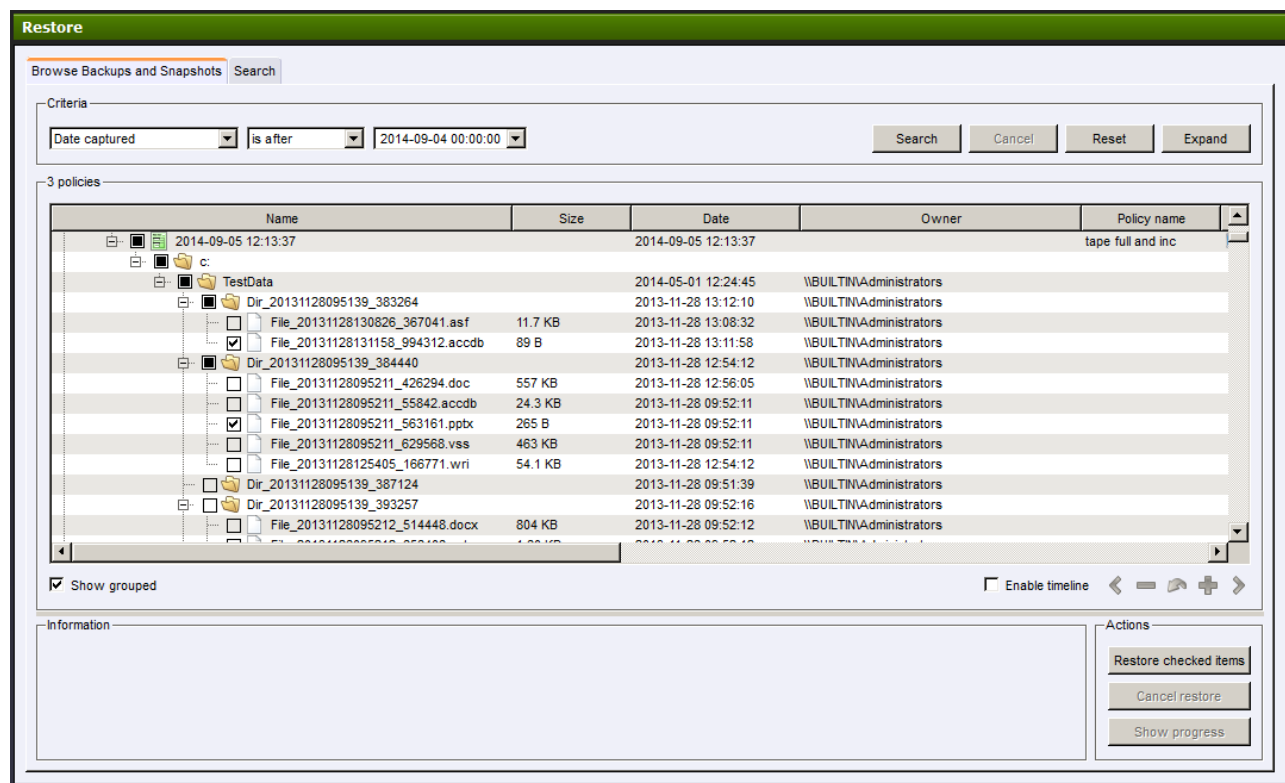


Figure 4-82 Restore screen

Searching for data

You can search for stored data on both the **Browse Backups and Snapshots** tab (Figure 4-83) and the **Search** tab. The **Browse Backups and Snapshots** tab is used to explore individual backups and snapshots in a hierarchical manner. The **Search** tab is able to display versions as well as individual files contained in any backups that have occurred.

Search queries are built up in the same way on both tabs, and in the same general manner as on the **Log Manager** screen. A query consists of zero or more conditions. The following are two example queries:

- Date captured is after 2014-09-04 00:00:00
- Status is Synchronized

The image shows a dialog box titled 'Browse Backups and Snapshots' with a 'Search' tab. Inside, there's a section labeled 'Criteria'. It contains three dropdown menus: the first is 'Date captured', the second is 'is after', and the third is '2014-09-04 00:00:00'. To the right of these are four buttons: 'Search', 'Cancel', 'Reset', and 'Expand'.

Figure 4-83 Search Criteria

The first term can be defined in this section at the top of the tab. The first combo box on the left is used to select the field on which the search will be done. This includes items such as size, name, directory, dates and extension.

The second combo box is used to select a comparison value (*comparator*). Typically, only a few comparators are provided for each item. The comparators change, depending on the type of field selected. For example, dates can be compared in terms of *before* and *after*. The last controls are where the value and units of the data to be searched for should be entered.

Click **Expand** to add further conditions to the search. Another line allowing the input of search fields, comparators and data is displayed (Figure 4-84). Clicking **Collapse** will hide the additional search terms. Click the X at the end of the line to delete a search term.

When you have constructed the query, click **Search** to send the search query to all repositories and tape libraries in the Data Instance Director network.



Note: If the search was conducted on the master, an immutable message will be logged for auditing purposes by any repository that receives the request.

The image shows the same dialog box as Figure 4-83, but with additional search criteria. The 'Criteria' section now has three rows. The first row is 'Date captured' is 'after' '2014-09-04 00:00:00'. The second row is 'Status' is 'is' 'Synchronized'. The third row starts with '<Select a field>'. To the right of the third row is a red 'X' icon. The buttons 'Search', 'Cancel', 'Reset', and 'Collapse' are still present.

Figure 4-84 Advanced search

Browse and search results fields

The fields used on the **Browse Backups and Snapshots** and **Search** tabs are described below. Items shown in **blue** only appear on the **Search** tab, those shown in **green** only appear on the **Browse** tab. The remaining items appear in both tabs.

Restore field name	Description
Name	The file's name, excluding its directory.
Location	Searches the name of any directory in an item's original directory structure on the source. For example, searching for <i>good</i> finds C:\good\hello.txt. (Search tab only)
Extension	The extension of the file to be searched for. The extension includes everything after the <i>first dot character</i> in the file name. (Search tab only)
Size	The size of the file. Searching for a file by size locates only the files that are exactly that size. For example, a 1 KB search only finds those files that are exactly that size in bytes (that is, 1024 bytes). To find files that are approximately 1 KB, you could use a ranged search with two terms, one of which compares in terms of smaller than, and one of which compares in terms of larger than.
Date	The date and time the snapshot corresponds to on the source. (Browse tab only)
Date Modified	The date and time the file was last modified on the source. (Search tab only)
Date Created	The date and time the file was created on the source. (Search tab only)
Owner	The owner of the file, for example \\BUILTIN\Administrators. Note however that only one component of the owner name should be entered in any individual search term. For example in the above name the individual components are 'BUILTIN' and 'Administrators'. To find all items owned by \\BUILTIN\Administrators add two distinct terms,

Restore field name	Description
	one of which searches for 'BUILTIN' and with the other searching for 'Administrators'.
Stream Count	The number of streams in the file.
Policy name	The name of the policy that generated the snapshot. (Browse tab only)
Source Machine	The machine from which the data originated. A name can be manually specified by changing the comparison value to 'is called'. Files originating from all machines within a particular group can be specified using the 'is in group' comparator. Group membership is decided at the time of the search. If a file was backed up from a machine in the group called <i>HR</i>, and then this machine is removed from that group, the file is not included in the results when searching for <i>HR</i> at later.
Repository	The name of the repository that the data is stored on.
Storage Group	The name of the Storage Group that the data is stored within.
Store Name	The name of the store within the repository in which the data is contained.
Date (repository)	The date and time the snapshot was completely stored into the repository. (Browse tab only)
Expiry Date	The date and time that the snapshot is due to expire. (Browse tab only)
Backup Type	On the Browse tab, two values are available: Snapshots and User snapshots. A value of Snapshots only returns those backups triggered automatically by Data Instance Director policy rules. User snapshots will only return those snapshots that were generated by the user on the Storage screen. See the Storage section for details on how this is done. By default, all snapshots are returned. On the Search tab, two values are available: Snapshot Files or Versions. Selecting Snapshot Files locates only data that was stored using the

Restore field name	Description
	backup or CDP operations in a policy. Selecting Versions locates only data that was stored using a version operation. By default, both types are found.
Date Modified (repository)	The date and time the file was last modified in the repository. (Search tab only)
Date Created (repository)	The date and time the file was first captured into the repository. (Search tab only)
User	The name of the user that triggered the version. This field only applies to versions and files within a snapshot are not associated with a user. (Search tab only)
Process	The name of the process that created the version. This field only applies to versions and files within a snapshot are not associated with a process. (Search tab only)
Status	Whether or not the snapshot contains unsynchronized data. This can occur if the source machine experienced downtime between snapshots and was not resynchronized. (Browse tab only)

Viewing search results

Once the query is sent, results are received from all the repositories and tape libraries. As the results arrive, they are displayed in the **Policies** or **Results** list depending on the tab selected. You can customize the data displayed in this list by right-clicking on the column header and selecting or deselecting the columns to be included from the context menu.

Selecting an item within the **Results** list displays all information associated with it in the **Information** group box at the bottom of the screen.

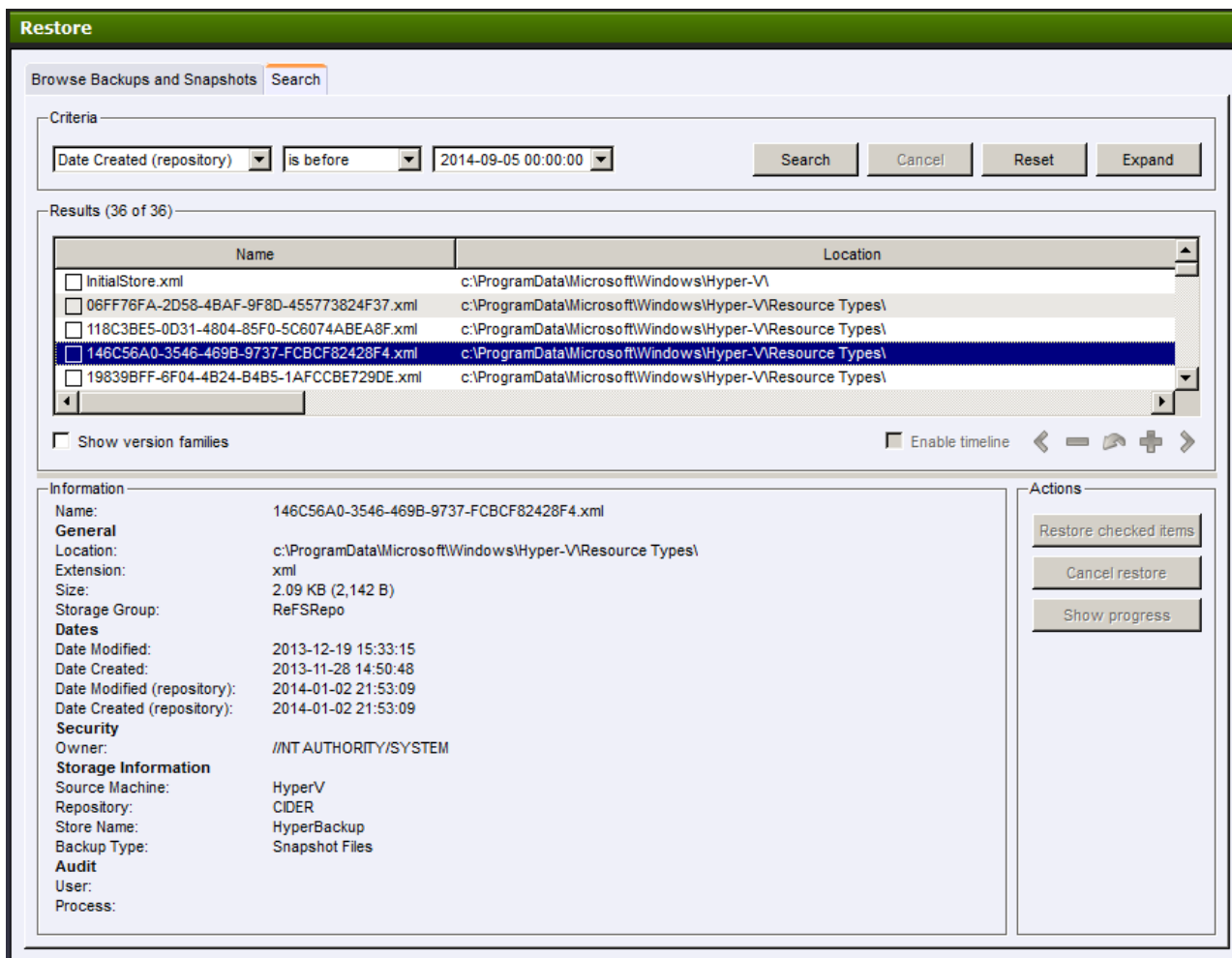


Figure 4-85 Restore screen Search tab results list

When searching from the **Search** tab (Figure 4-85), selecting the **Show version families** option below the **Results** list allows you to view the version history for a family of versions. Families of versions are grouped together under the same parent. Timelines can also be enabled in this view (see the [Timelines](#) section below).



Note: When viewing version history, only files backed up using a versioning policy are displayed. Normal backup files are ignored because they do not have a version history, and as such can only be viewed in the flat view.

Browsing files

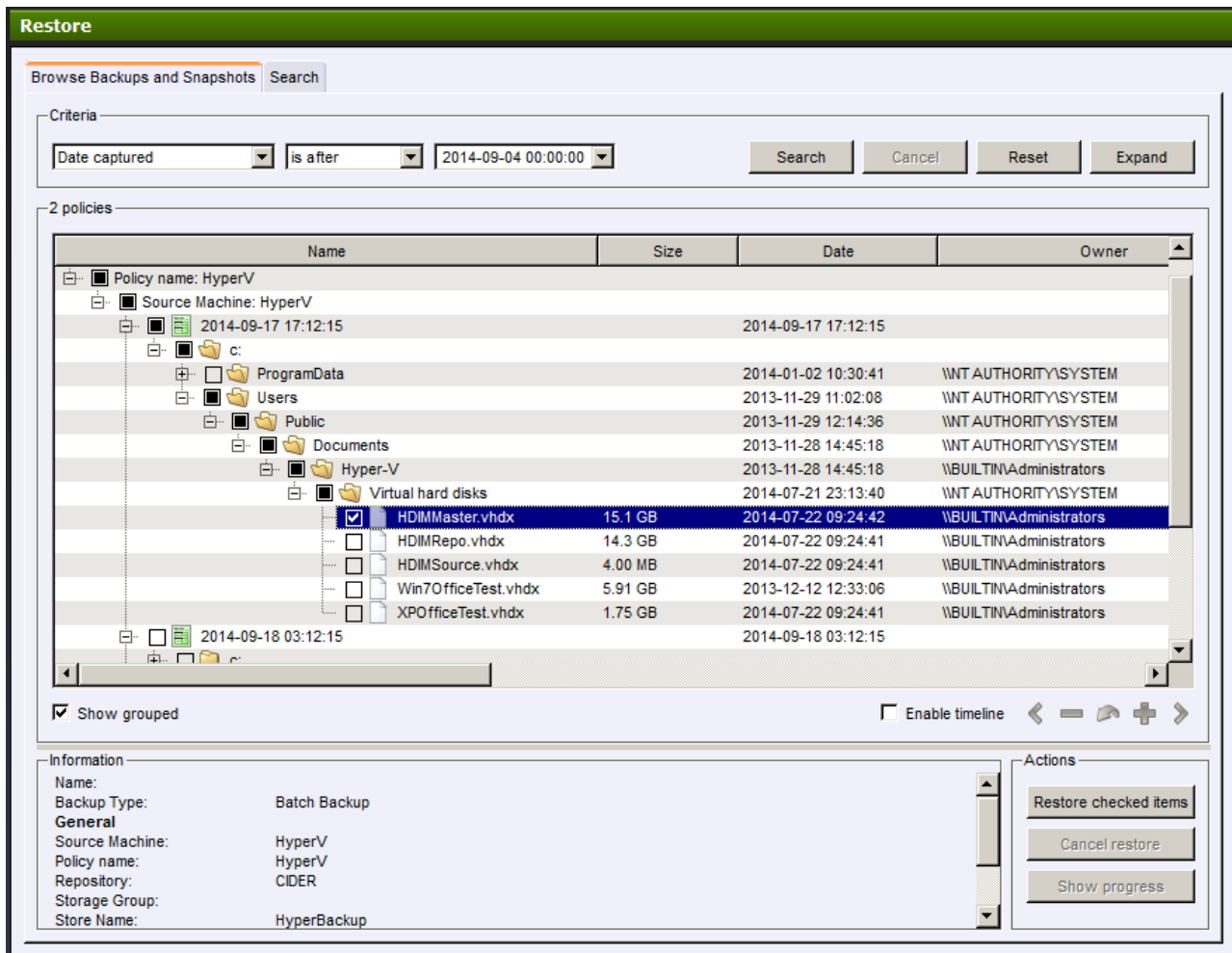


Figure 4-86 Browse backups and snapshots tab

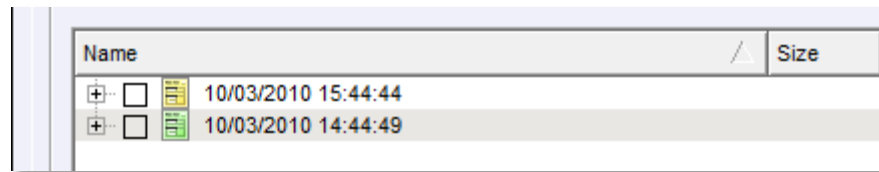
Search results in the **Browse Backups and Snapshots** tab (Figure 4-86) show backed up files in a directory structure. To browse an item, click the small plus symbol (+) to the left of its name. A request is sent to the repository or tape library where the file is located.

- If a directory is browsed, its yellow icon  briefly turns green  to show that its contents are being retrieved.
- If an error occurred for any reason while browsing, the icon of the node where the problem occurred turns red .

Once a directory has been browsed, it can be opened and closed freely, without making another request to the repository, except in the case where the [results limit](#) has almost been reached.

Click the **Show Grouped** checkbox underneath the **Results** list to view all snapshots grouped by the policy that created them and the machine they originated on. Timelines can also be enabled in this view (see the [Timelines](#) section below).

Next to each snapshot is an icon (Figure 4-87) indicating the status of the returned search results. **Green**  indicates that the snapshot is consistent, **yellow**  indicates the snapshot may have been created when the source and destination were out of sync and could result in data being stale or missing. Snapshots can be individually deleted within the **Storage** screen. A snapshot of an entire store (represented by the  icon) can be created by the user on the same screen. See the [Storage](#) section for further details.



Name			Size
+ ☐ 		10/03/2010 15:44:44	
+ ☐ 		10/03/2010 14:44:49	

Figure 4-87 Snapshot status indicators

Result limit

The number of results the **Restore** screen can display at once is limited. If the combined number of search results, snapshot results, and browse results exceeds approximately 32000, then no more results are displayed in order to conserve system resources. When this occurs a message box appears asking what course of action should be taken to rectify the issue. The following options are available:

- Clear all current search results from memory and repeat the search with tighter constraints so that fewer results are returned.
- Ignore the issue for now. All the results up to the result limit will still be displayed but no further results can arrive. By closing browsed tree nodes and clearing search results manually it is possible to free up memory enough for new browse requests to succeed.
- Increase the result limit. The search will have to be repeated.

Timelines

When viewing either the grouped snapshot view (**Show grouped** is checked on the **Browse Backups and Snapshot** tab (Figure 4-88)) or the version family view (**Show version families** is checked on the **Search** tab (Figure 4-89)) it is possible to enable timelines. This is done by checking the **Enable timeline** checkbox in the bottom right. When this view is enabled, it displays the distribution of versions or snapshots through time in a visually intuitive manner. Each event is represented at the appropriate point in the timeline by a green or yellow marker as follows:

- - Live Backup
- ▲ - Batch Backup
- - Continuous Data Protection or Version
- ★ - Tape Backup
- ◆ - Intelligent Storage Snapshot/Replication

If the source was not in sync with the destination at that point in time, the marker is displayed in yellow.

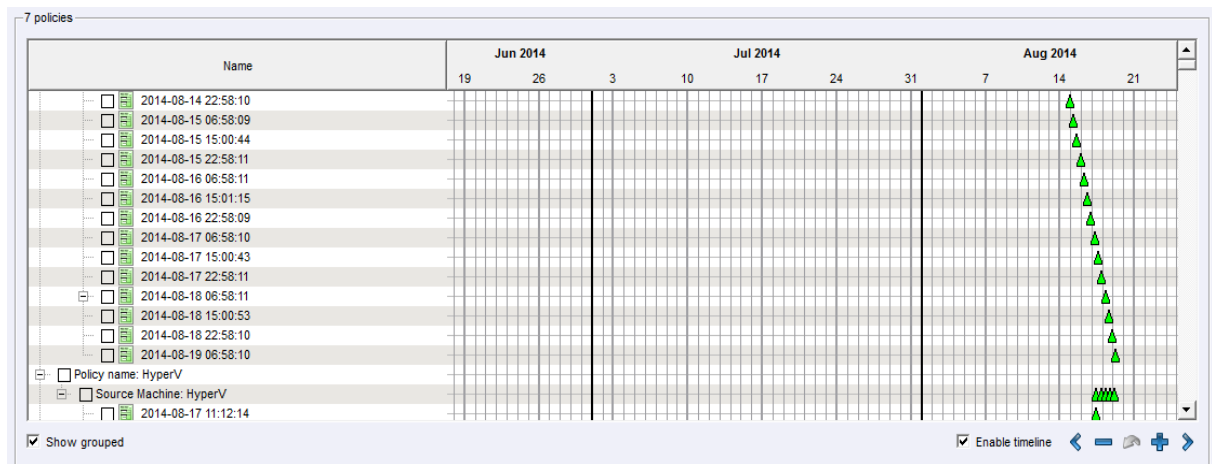


Figure 4-88 Timeline views for snapshots

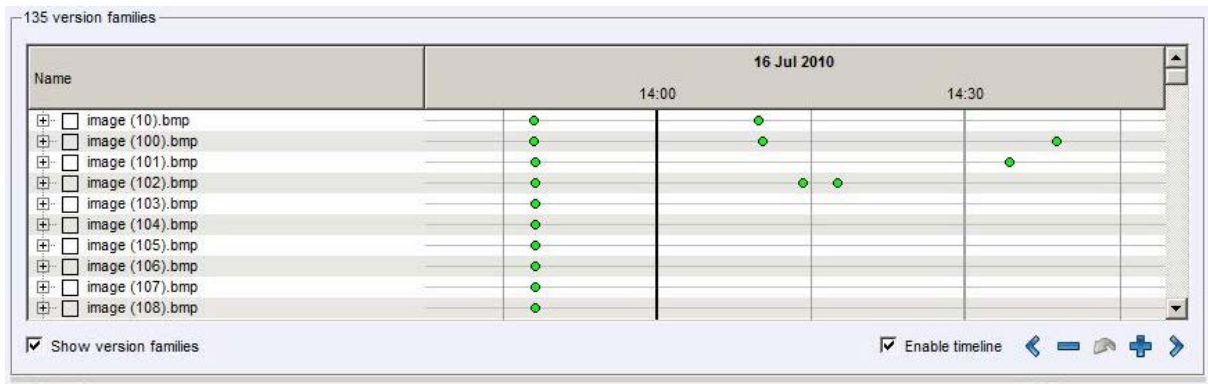


Figure 4-89 Timeline views for versions

At first, the timeline view is automatically sized so that all events can be seen within the timeline. By using the buttons in the bottom right , it is possible to move the timeline left and right (arrow buttons) and to zoom in and out (plus and minus buttons). The view can be reset to how it was originally by clicking the revert button.

By hovering over an item within the timeline, information about that specific item can be seen in the form of a hover hint. Clicking on an item will disable the timeline view and return the user to the previous view, but with the clicked item selected. Information about the item can then be seen in the **Information** pane at the bottom of the screen.

Restoring data

Data can be selected for restore by checking the checkbox to the left of the item's name. Individual versions of files, entire families of versions, entire snapshots, and individual components comprising a snapshot all can be restored to either the local machine and, if the user interface is being run on the master node, to any remote machines.

To restore data, first select the item to restore from the list on either tab of the **Restore** screen and then click the **Restore checked items** button in the bottom left corner.



Note: Following are the data restoration limitations:

- Because of issues with translating security streams, data originating from Linux machines cannot be restored to Windows machines and vice versa. For this reason, it is not possible to select Linux data alongside Windows data. They must be restored in separate transactions, and they can only be restored to compatible OS types.
 - Data can be restored to NAS using a CIFS agentless node, with the following conditions:
 - o The user account that was specified when the agentless node was created has write permissions on the share.
 - o If backups were done using a CIFS agentless node, then you can restore to the original location or a particular network location, by specifying a restore path in the following format: `\\machinename\share`
-

Restore options

Prior to a restore, the **Restore Options** dialog (Figure 4-90) is displayed, in which the exact behavior of the restore can be configured. See Figure 4-90 below.

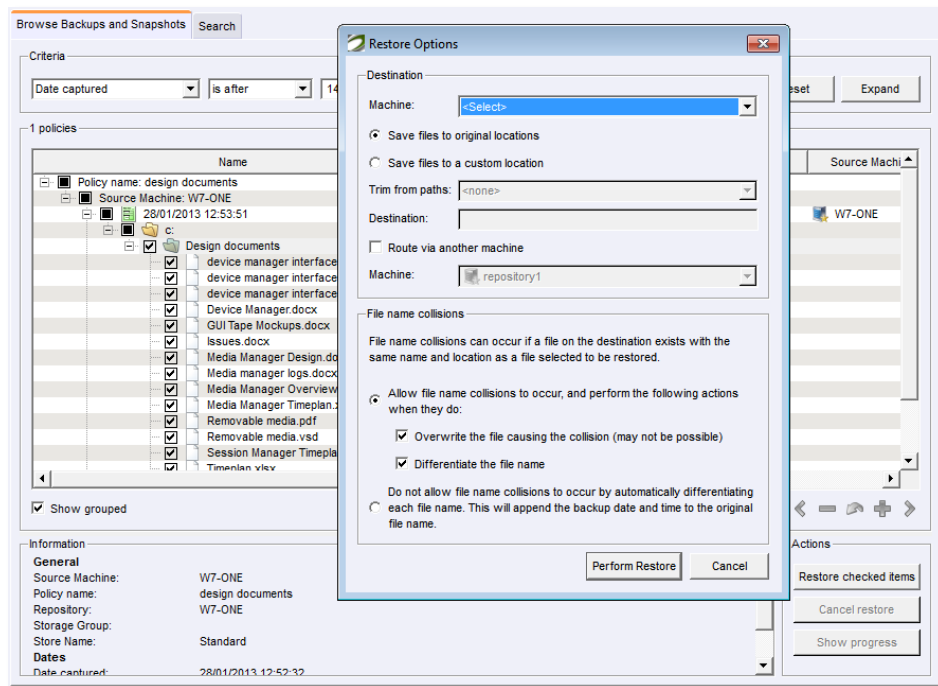


Figure 4-90 Restore Options dialog

This dialog contains a number of controls:

- **Destination** – these controls specify how and where to restore files:
 - o **Machine** – the target Data Instance Director node name for the restore. This is only selectable when restoring from the Master GUI.
 - o **Save file to original locations** – use the original path names when restoring files.
 - o **Save File to custom location** – use a new path based on the following specification (refer to [Mirror Destination](#) for guidance on path trimming):
 - **Trim from paths** – presents a list of path prefixes that can be removed from the files being restored, whilst ensuring files remain grouped together in their respective folders.
 - **Destination** – the path specification to be prepended once the paths are trimmed.
 - o **Route via another machine** - primarily used for Internet Connected Nodes where the destination is not directly connected to the repository node, so a proxy needs to be specified.
 - **Machine** – the Data Instance Director node name of the machine via which a connection to the remote node is to be routed.
- **File name collisions** – these controls determine the action taken if the files being restored already exist on the destination:
 - o **Allow** – if file names collide during a restore then perform the following actions:
 - **Overwrite** – just replace the old file with the new unless permissions or attributes prevent it. The existing file is permanently deleted.
 - **Differentiate** – If overwriting fails then the new file will be restored with a backup date and time prefix to differentiate it from the original file.
 - o **Do not allow** – differentiates all restored files by prepending the backup date and time (i.e.: <YYYYMMDDHHMM>_filename.ext) to the file name.



WARNING: Be careful when restoring data to a location that is in the scope of one or more Data Instance Director policies. The restored data will not be incorporated by Data Instance Director into its workflows until after a resynchronization has occurred and synchronization errors for the restored data may occasionally be logged until that point. Perform a manual resynchronization (see the [Node](#) section) to correct the inconsistent state.

When the settings are satisfactory, a restore can be initiated by clicking **Perform Restore**.



Note: If the restore was initiated by the master, an immutable message will be logged for auditing purposes by any repository systems that receive the request.

Tape Restore

Before restoring a tape backup, you can view details of which media are required for the restore. To do this, check a snapshot, right click and select **View Media Details...** the resulting dialog (Figure 4-91) will show the details of the snapshot. Click **OK** to close the dialog. Note that this dialog will appear automatically if you try to restore a snapshot, stored on a media that is offline.

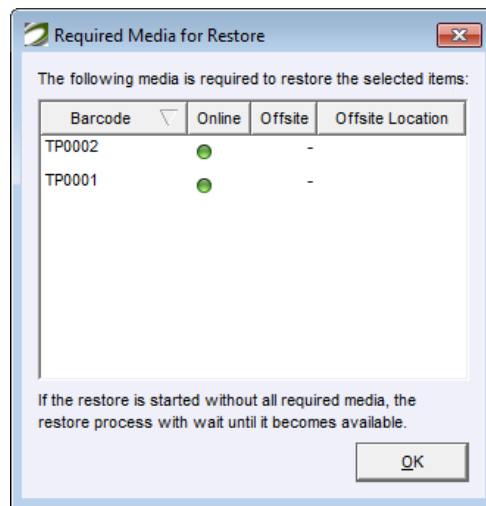


Figure 4-91 View media details dialog



Note: Options for viewing and managing media and libraries can be found in the **Storage** screen under the **Library Management**, **Media**, and **Media Request** tabs.

VMware restore options

A restore of a VMware virtual machine presents a different set of restore options to restoring regular file data (Figure 4-92).

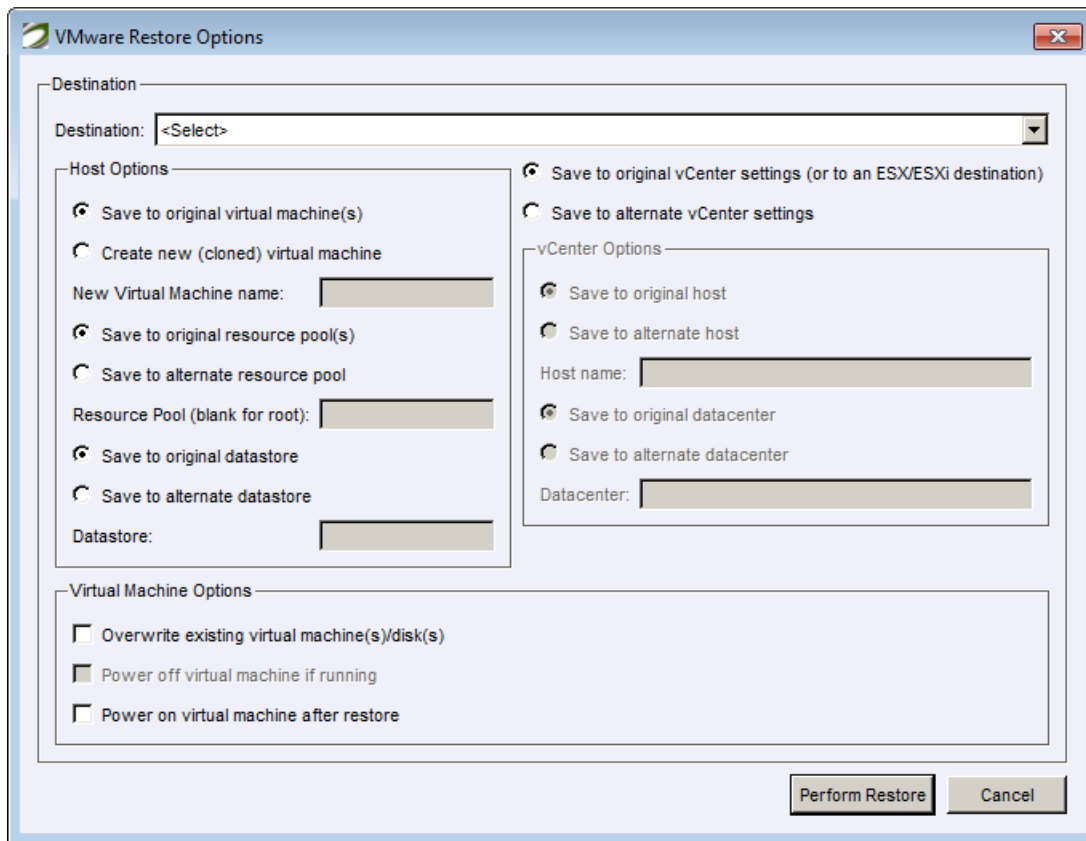


Figure 4-92 VMware restore dialog

- **Destination** – Specifies which ESX node the data is to be restored to.
- **Host Options:**
 - **Save to original virtual machine(s)** – restores to the original virtual machine.
 - **Create new (cloned) virtual machine** – This uses the backup of a virtual machine and allows it to be restored with a different name to create a clone of the original virtual machine.
 - **New Virtual machine name** – Specify the name for a cloned virtual machine.
 - **Save to original resource pool(s)** – If the virtual machine was in a resource pool it will be restored to the same resource pool.
 - **Save to alternate resource pool** – Will restore the virtual machine to a different resource pool than the one it was previously in.

- **Resource Pool (Blank for root)** – Specifies the name of the alternate resource pool.
- **Save to original datastore** – Will restore the virtual machine to the original datastore within the ESX host.
- **Save to alternate datastore** – Will restore the virtual machine to a different datastore than the one it was previously on within its ESX host.
- **Datastore** – Specifies the alternate datastore name.
- **Save to original vCenter settings (or to an ESX/ESXi destination)** – This will restore the virtual machine to its original host.
- **Save to alternate vCenter settings** – This option will enable some additional **vCenter Options** that are only applicable when using vCenter:
 - **Save to original host** – The virtual machine will be restored to its original host within vCenter.
 - **Save to alternate host** – This restores the virtual machine to a host that is not the one from which it was backed up.
 - **Host name** – Specifies the name of the alternate host.
 - **Save to original datacenter** – The virtual machine will be restored to its original datacenter within vCenter.
 - **Save to alternate datacenter** – Will restore the virtual machine to a different datacenter to the one it was backed up from.
 - **Datacenter** – Specifies the name of the alternate datacenter
- **Virtual Machine Options:**
 - **Overwrite existing virtual machine(s) / disk(s)** – If the virtual machine is being restored to a location where it already exists, for example, its original location, the conflicting (original) virtual machine will be overwritten by the restore process.
 - **Power off virtual machine if running** – If the virtual machine is being restored to a location where it already exists, for example, its original location, the virtual machine it is overwriting (the original) will be powered off when the restore process begins.
 - **Power on virtual machine after restore** – The virtual machine will be automatically powered on once the restore process has been completed.

Exchange mailbox restore options

A restore to an Exchange mailbox presents a different set of restore options than those for when regular file data is restored (Figure 4-93).

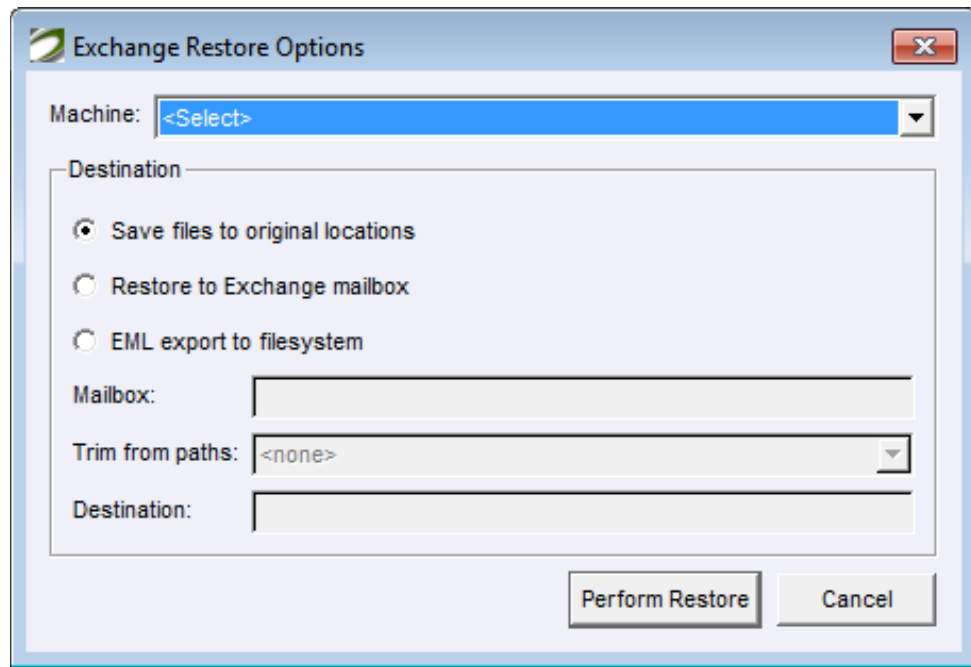


Figure 4-93 Exchange Restore dialog



Note: When specifying **Trim from paths** & **Destination**, ensure that the resultant path to which emails are restored is accessible to users.

- **Machine** – Specifies which exchange node the data is to be restored to.
- **Destination:**
 - **Save Files to original locations** – This option will restore any selected emails back to their original mailbox and to their original location.
 - **Restore to Exchange Mailbox** – This is used for to restore emails to a different location and can also be used to restore emails to a folder within a different mailbox, it will also enable the following options:
 - **Mailbox** – This is used to specify which mailbox the data is to be restored to. The mailbox is specified in the format of: aaa@bbb.com.
 - **Trim from paths** – This specifies what part of the email message's original path is removed prior to being mapped to a new path when restored.

- **Destination** – Specifies the folder within the mailbox which data is to be restored after any trimming.
- **EML export to filesystem** – Will export email messages to the file system of the Master node. Using this option enables the following options:
 - **Trim from paths** – This is where the email message's original path is removed and mapped to a new path when exported.
 - **Destination** – Specifies the directory on the Master node which data is to be restored to after any trimming.

Restore destination

Various caveats exist depending on the data being restored, the machine instigating the restore, and the destination machine selected:

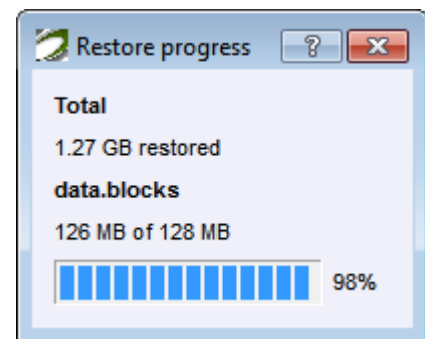
- Master nodes can restore data originating on a Linux system to any other Linux system in the Data Instance Director network, and can restore data originating on a Windows system to any other Windows system in the Data Instance Director network.
- Non-master nodes can only restore their own data, and can only restore it to themselves.
- CIFS nodes  can have data restored to any of their accessible shares. To do this enter a path in the form \\<machine name>\<share name>\<path> into the **Destination** edit box.

Restore progress

Once a restore is started, a progress dialog (shown to the right) is displayed.

This message box shows the name of the current file, the percentage of the file that has been restored, and the total amount of data that has been transferred during the entire restore process.

If the restore was initiated from the **Search** tab, then the overall restore percentage is also graphically displayed. However, snapshot restores do not show this.



Note: Only the main data stream of each file is used when measuring progress. It is possible for a file with a large alternate data stream to be restoring without any visual indication within the progress bars.

When the restore is completed, the progress dialog disappears and a message is displayed in the status bar indicating a successful restore.

A restore can complete successfully, complete with errors, fail prematurely, or get cancelled. In each case, more information can be found by checking the logs.

A restore can be cancelled at any time by clicking **Cancel Restore** on the **Restore** screen. If the **Restore Progress** dialog is closed at any time it can be reopened by clicking **Show Progress**.

Topology examples

This chapter reviews some common topologies that are used in the creation of data flows.

- ☐ [Many-to-One](#)
- ☐ [One-to-Many](#)
- ☐ [Forwarding](#)
- ☐ [Smart Repository Sync](#)
- ☐ [Constructing a Complex Data Flow](#)

Many-to-one

Many-to-one topologies are one of the standard building blocks used when creating complex data flows. These topologies allow data from multiple nodes to be stored on a single destination node, such as a backup server. For this example, the deployment requires one replication destination node and at least two source nodes. The process is very similar for a topology involving repository destinations.

First, the policy needs to be defined. The objective is for data to be replicated from the source nodes to the destination node. Place some files in the source folders before starting to see some data transfers occur when the setup is complete.

1. Open the **Policy Definition** screen (Figure 5-1).
2. Drag the **Replicate** icon to the **Operation** workspace.

Now, a classification for the policy to work upon needs to be created. This classification defines the files that should be transferred.

3. Drag a **Path** icon to the **Classification** workspace.
4. Enter the path of an existing directory into the upper edit box in the **Attributes** tab. This path is a simple but common classification that informs Data Instance Director that all the files and folders within the specified directory should be replicated.

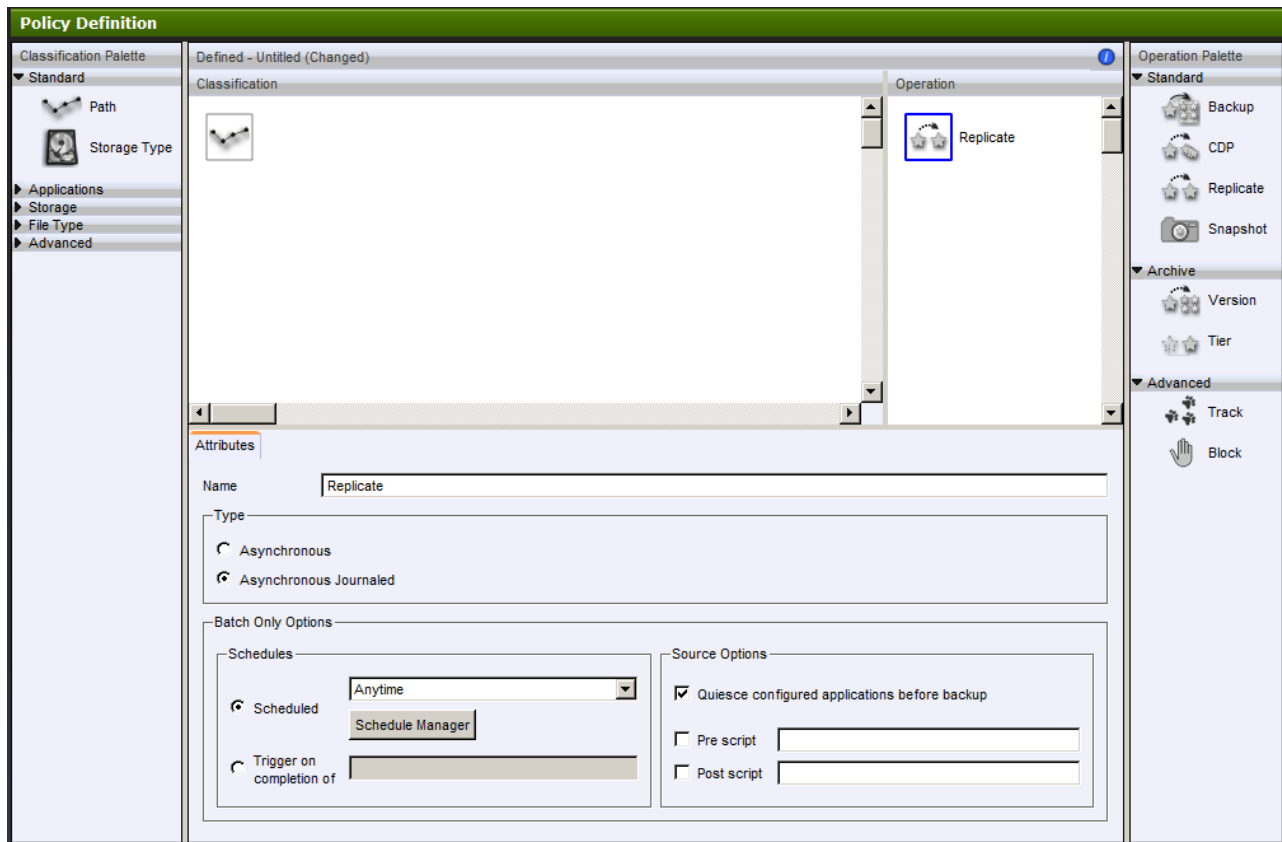


Figure 5-1 Defining the many-to-one policy

5. Select **File** → **Save**, or click the **Save** icon .
6. When prompted, enter a Name and Description (optional) for the policy, and click **Save**. The policy is now accessible from the **Data Flow** screen.
7. Select the **Data Flow** screen. See the [Data Flow](#) section in this manual for more information on using this screen.
8. In the **Data Flow** workspace, create a topology like the one shown in Figure 5-2.

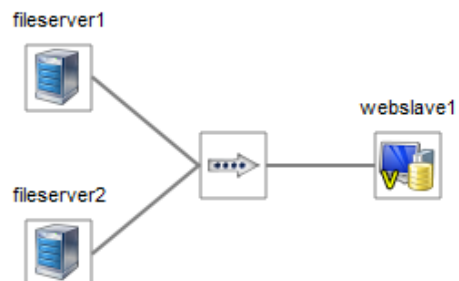


Figure 5-2 Many-to-one example

9. Now the policy needs to be applied to each node within this topology. Start by selecting a source node (*filesaver1*).
10. Tick the checkbox for the policy from the list in the **Policies to assign locally** tab. The policy is automatically applied to the mover.
11. The policy still needs to be applied to the other source node. Select the other source node (*filesaver2*), and apply the policy.
12. Lastly, the policy needs to be applied to the destination node. Select the destination node within the workflow.
13. Use the *Default* destination configuration for the destination node. This will store transferred data in the C:\HDID-Mirror directory by default, and – because the ‘Append with source machine’ checkbox is checked – will store replicated data in a directory named either *filesaver1* or *filesaver2* within the C:\HDID-Mirror directory depending upon which of the two source nodes it originated from.
14. Click the checkbox to select the policy in the **Available Policies** list.
15. Click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog displays.
16. Click **Distribute Rules** after the policy is compiled.

The **Monitor** and **Log Manager** screens allow you to see the results of the configuration. They are discussed later in this chapter.

Synchronization between source nodes and destination node takes place first. The interface indicates the synchronization is complete when all of the green Initial Sync badges  have disappeared from the source nodes shown on the Monitor screen. If you look at the destination directory – C:\HDID-Mirror – a separate folder can be seen corresponding to each source, with each source’s files contained within the folder.

One-to-many

Creating a *one-to-many* topology involves many of the same concepts and actions involved in creating a many-to-one topology. Again, a suitable number of authorized nodes must be available: one source node and at least two replication destination nodes. Like in the many-to-one example, the same principles apply when creating a topology involving repository destination nodes.

The policy is similar but slightly more complex than the previous example – it will instead only replicate text files.

Create some text files in the source folders before starting so some data transfer takes place when the setup is complete.

1. Open the **Policy Definition** screen (Figure 5-3).

2. Drag the **Replicate** icon to the **Operation** workspace.
3. Drag a **Path** icon to the **Classification** workspace and type a valid source directory path into the **Enter paths** list box.
4. Drag a **Documents** classification icon to the right of the **Path** icon. This action creates an **AND** relationship, meaning data must meet both conditions in order to be replicated.
5. In the **Attributes** tab, select Text (*.txt). Deselect the other options. This configuration tells Data Instance Director to target only text documents.

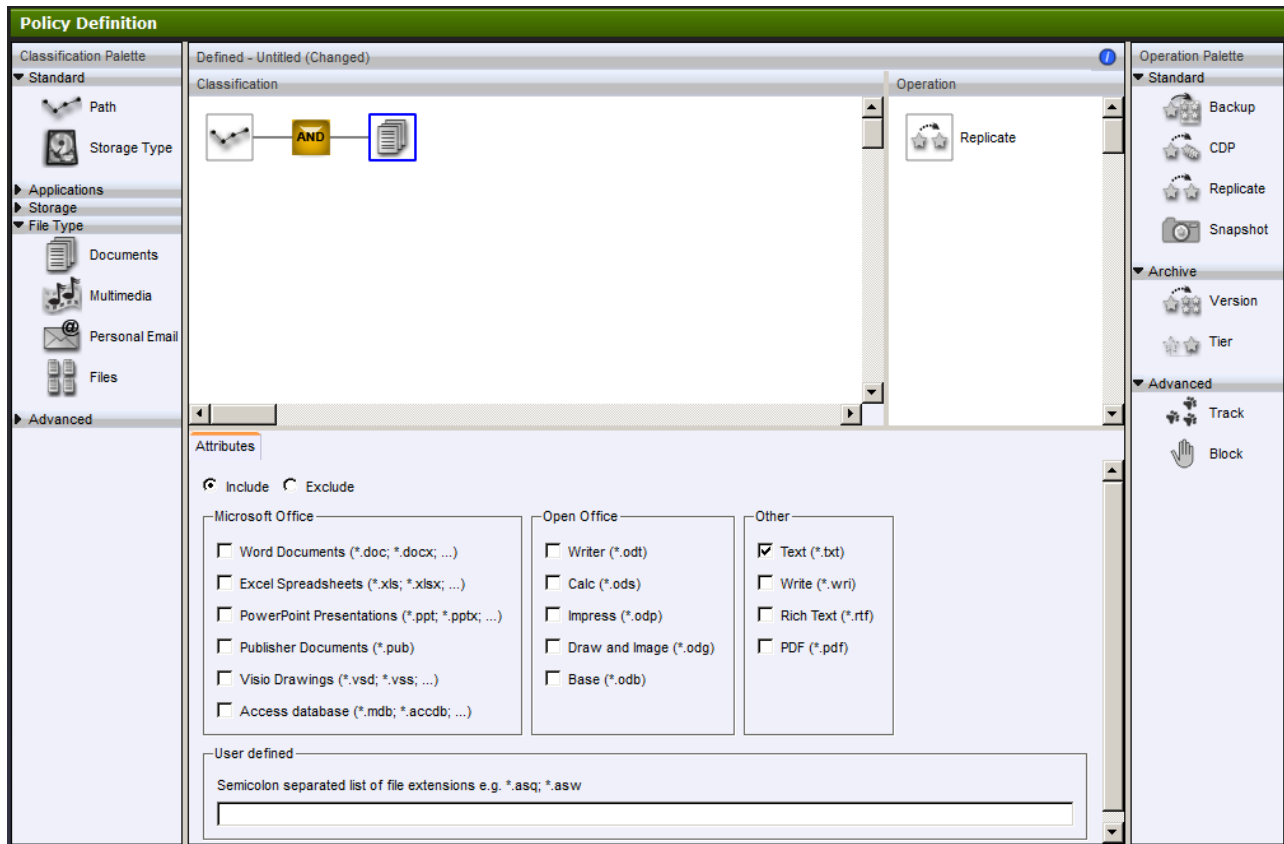


Figure 5-3 Defining the one-to-many policy

6. Select **File** → **Save** to keep this policy.
7. Open the **Data Flow** screen.
8. Create a data flow like the one shown in Figure 5-4.

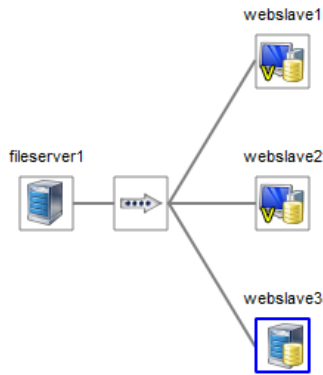


Figure 5-4 One-to-many topology

9. The policy needs to be applied to the source node. Select the source node (**filesaver1**).
10. Tick the checkbox for the policy from the list in the **Policies to assign locally** tab. The policy is automatically applied to the mover.
11. The Default destination configuration is fine for each of the destination nodes.
12. Select each of the destination nodes in turn, and tick the policy defined earlier from the list in **Available Policies** within the **Default** tab.
13. Click the **Save and Activate** icon  in the tool bar. The **Rules Compilation** dialog is displayed.
14. Click **Distribute Rules** after the policy is compiled.

The **Monitor** and **Log Manager** screens allow you to see the results of the configuration. They are discussed later in this chapter.

Synchronization between source nodes and destination node takes place first. The interface indicates the synchronization is complete when all of the green Initial Sync badges  have disappeared from the source nodes shown on the **Monitor** screen. Inspection of each destination node's C:\HDID-Mirror directory shows they have each successfully replicated the text files contained within the source directory.

Forwarding

In a forwarding topology, a node is used only to forward data from one set of nodes to another set of nodes. This node does not store the data. Typically, this topology is used to replicate data to an offsite location. By sending all data through the forwarding node, the number of systems that are visible to external observers can be limited.

Again, a suitable number of authorized nodes must be available: two source nodes and one replication destination node. Like in the last tutorial, the same principles apply when creating a topology involving repository destination nodes.

Create some text files in the source folder before starting so some data transfer takes place when the setup is complete.

1. Create the same policy as used for the one-to-many topology in the previous example.
2. Open the **Data Flow** screen.
3. Create a one-to-one topology (using the forwarding node) with three different systems, two of which are destination nodes. See Figure 5-5.



Figure 5-5 Forwarding topology

4. Assign the policy created in step 1 to the source node (*filesaver1*) and the final destination node (*client*). The movers automatically implement the policy. The policy is not assigned to the forwarding node (*standby1*) as that node only passes the data that it receives.
5. Select the destination node (*client*).
6. Create a new mirror destination tab that corresponds to the configuration shown in Figure 5-5:
 - a. Right-click on the destination configuration tab bar. This is the tab bar that contains the **Default** tab.
 - b. Select Add Replication Destination from the pop-up menu.
 - c. In the resulting dialog (Figure 5-6), enter a **Trim from paths**, such as C:\Source and set **To Destination Path** to be C:\Destination. Note that C:\Source should be replaced with the source directory that was specified in the policy definition. This rule maps every file and folder that originally existed within C:\Source to C:\Destination. For example, C:\Source\hello.txt is mapped to C:\Destination\hello.txt. Ordinarily, with a blank source map, this file would be replicated to C:\Destination\C%\Source\hello.txt, as Data Instance Director attempts to recreate the original data structure.
 - d. Deselect the checkbox for Append with source machine name.
 - e. Enter a name, and click **Save**.

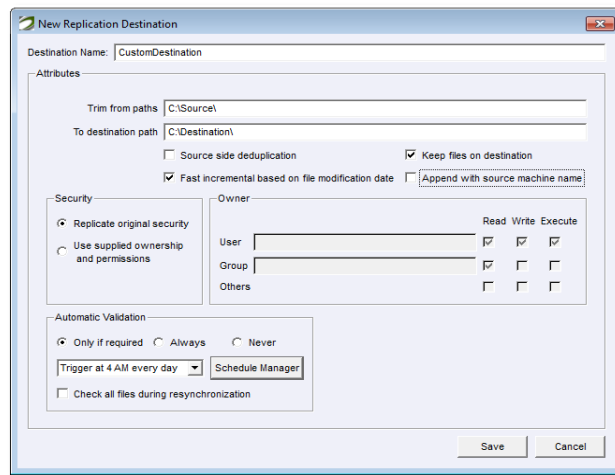


Figure 5-6 Updated destination path

7. View the tab for the newly-created configuration. Select the policy defined earlier from the **Available Policies** list.

8. Click the **Save and Activate** icon  in the tool bar.

The **Rules Compilation** dialog displays.

9. Click **Distribute Rules** after the policy is compiled.

On the **Monitor** screen, you can see the utilization bars moving for the forwarding node, indicating incoming and outgoing network traffic as it receives the data and then forwards it. Once this process has completed, the replicated files and folders can be seen on the destination system in their correctly mapped paths. None of this data is stored on the forwarding node at any point.

Smart repository sync

Smart repository sync enables a repository to create backups from another repository's backups. The secondary repository will take the incremental changes required to create a new snapshot from the primary repository, the source node is not involved in this process. It is also possible for the secondary repository to only request a subset of the data according the policy used.

To setup up a Smart Repository Sync:

1. Create a policy that has two Backup operations with different RPOs, in this example we will have 'Backup every hour' and 'Backup to Repository 2' (Figure 5-7).

2. Select the 'Backup to Repository 2' operation and enter into **Trigger on completion of** edit box 'Backup every hour' (without inverted commas). This will result in the 'Backup to Repository 2' operation only being triggered upon completion of the 'Backup every hour' operation.

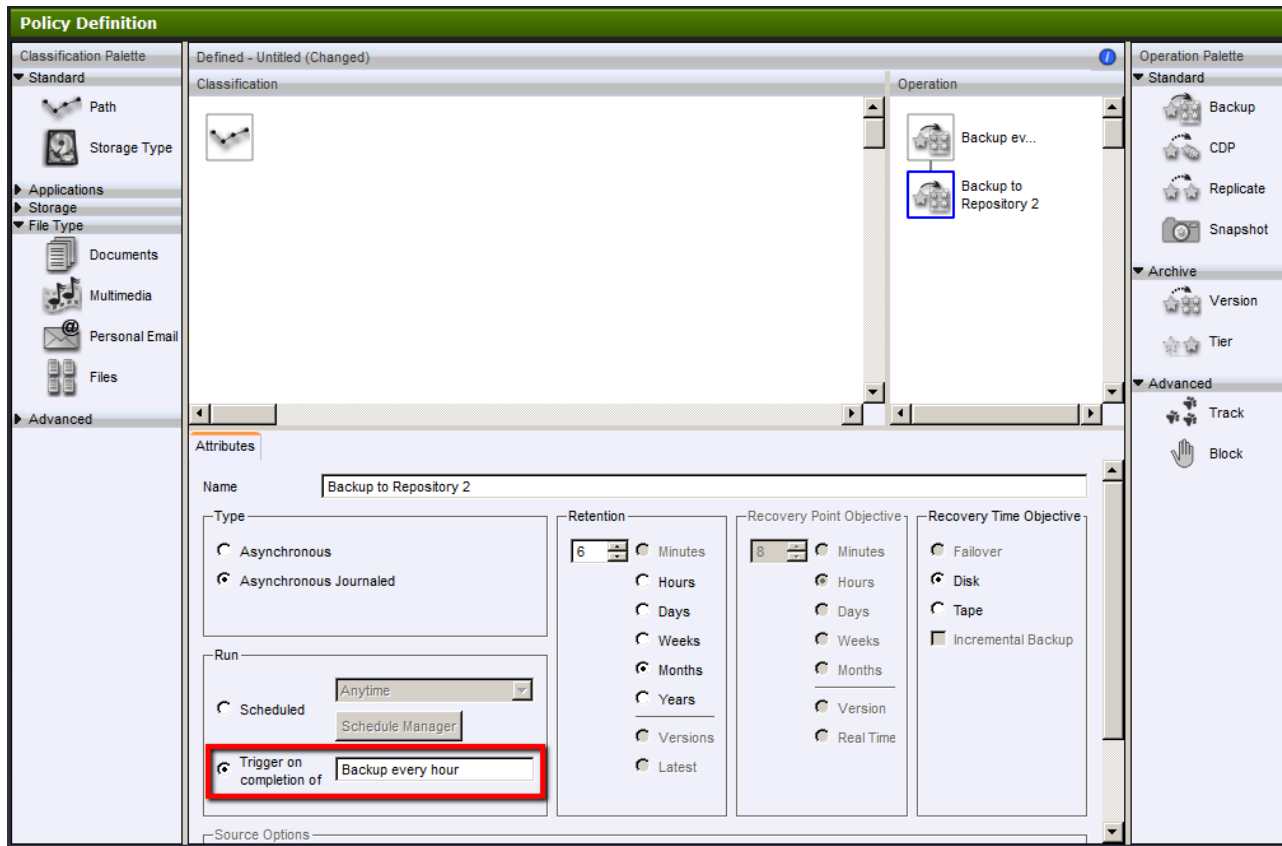


Figure 5-7 Smart repository sync policy

3. Open the **Data Flow** screen.
4. Create a one-to-one topology (from the forwarding node) with three different nodes, two of which are repository nodes. The two repository nodes must be connected together using a batch mover, as shown in Figure 5-8.

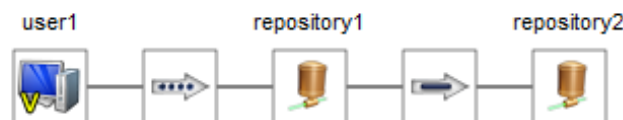


Figure 5-8 Smart repository sync topology

5. Assign the policy created in step 1 to the source node (*user1*), enabling both backup operations

6. Select the first repository (*repository1*) and check only the 'Backup every hour' operation.
7. Select the second repository (*repository2*) and check only the 'Backup to repository 2' operation.



8. Click the **Save and Activate** icon in the tool bar. The **Rules Compilation** dialog displays.
9. Click **Distribute Rules** after the policy is compiled.

On the **Monitor** screen, you can see that the first repository will be syncing with the source node and when complete will create a snapshot. The second repository will not start syncing until the sync to the first repository is complete. Once the first repository has finished syncing the second repository will start to sync with the first repository, not involving the source node at all.

Using **Start on completion of operation**, as opposed to using just RPO for the 'Backup of Repository 2' operation, ensures that the second repository will always be backing up from the first repository's latest snapshot, resulting in a more up to date backup.

Constructing a complex data flow

Data Instance Director is designed to allow data to be filtered at a very granular level according to defined policies, and for data to be either sent directly to a node or routed through one or more intermediate nodes. This section will show two solutions to a common scenario each of which has its own tradeoffs.

Typically a set of policies may be implemented by using parallel (Figure 5-9) or serial (Figure 5-10) data flows. Consider a scenario where the following backup policies are required:

- The entire contents of a source machine need to be copied to a destination machine for disaster recovery purposes.
- Documents from that same source machine need to be copied to a different destination machine for document archiving purposes.

Parallel solution

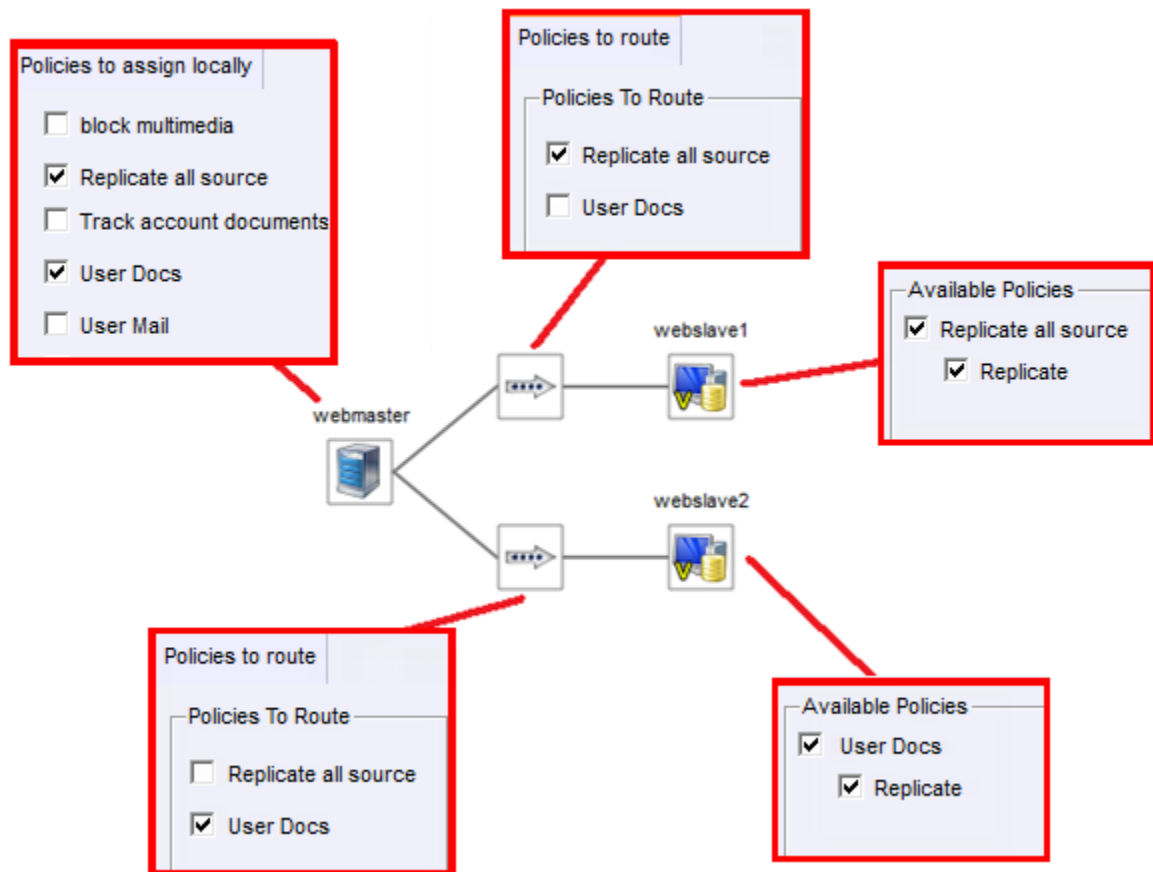


Figure 5-9 Complex data flow, parallel solution

The source node (*webmaster*) has two policies assigned; one for the whole machine and one for the documents. Because there are two movers in parallel, data that is common to both policies (i.e. documents) must be transmitted by *webmaster* twice. This could be considered wasteful as we are doubling the load and bandwidth. However this topology is more fault tolerant since *webslave2* doesn't depend on *webslave1* in order to implement its policy.

However, there are circumstances in which this configuration can be beneficial. For example, you may decide to bandwidth throttle or use a certain scheduling window for the 'Replicate all source' policy to *webslave1* but desire the 'User Docs' policy to be rapidly replicated to *webslave2*.

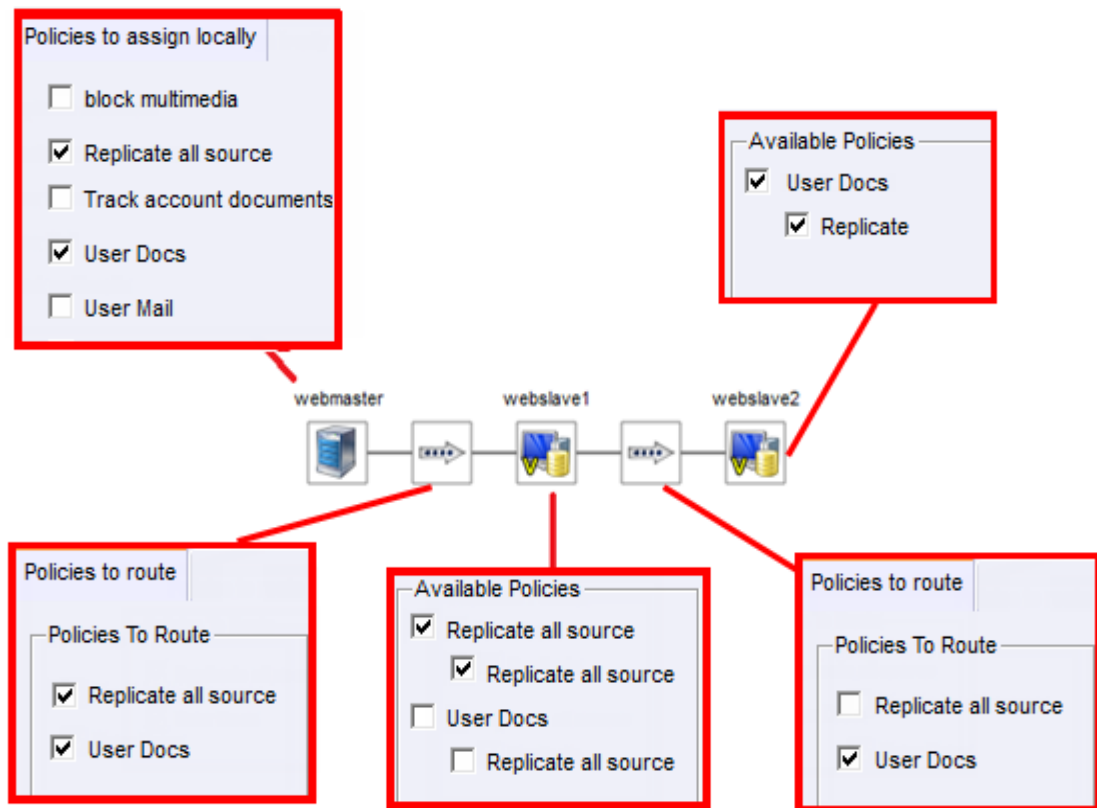


Figure 5-10 Complex data flow, serial solution

In this solution, we apply the same policies as the previous solution, the difference is we route the data going to *webslave2* through *webslave1*. This reduces the data traffic on *webmaster* since it only has to transmit the data once. However the backup of user documents now requires *webslave1* to be up and available on the network.

Command-based task management

This chapter provides information about using the command line to perform tasks such as backing up and restoring node configuration files, and changing a client machine's master node. It also provides customization options for handling alert notifications and for creating blacklists.

- [Reviewing guidelines about running command line tools](#)
- [Backing up, restoring node configuration files: mastersettings.exe](#)
- [Changing a client machine's master node: setconfig.exe](#)
- [Changing the behavior of a Data Instance Director node: diagdata.exe](#)
- [Reducing the repository's size: repomaint.exe](#)
- [Customizing alert notifications](#)
- [Creating a blacklist for stubbed files](#)

Reviewing guidelines about running command line tools

This chapter reviews the usage of the following four command line tools:

- mastersettings.exe
- setconfig.exe
- diagdata.exe
- repomaint.exe

All commands are run from the \bin directory. By default these are located in:

C:\Program Files\Hitachi\HDID\bin on Windows machines

And:

/opt/Hitachi/HDID/bin on Linux machines

We recommend that command line based changes are performed by experienced administrators only.



Note: Information about additional Data Instance Director command line tools and their arguments are available from the Data Instance Director Knowledge Base, and are only to be used with explicit guidance by Customer Support.

Backing up, restoring node configuration files: mastersettings.exe

Backing up node configuration files

You can back up the master node's configuration files to a zipped archive file using the mastersettings command. During the procedure, only the master node's settings are zipped up; other information such as repositories and their metadata stores, media indexes, and log messages are excluded.

We recommend that you back up the master node's configuration files once a week. You can set up a backup policy on a recurring schedule that runs the mastersettings CLI tool as a pre-script, then backs up the resulting zip file.

Options and usage

Options

Abbreviated Argument	Full Argument	Description
-h	--help	Displays command line parameter usage instructions.
-e	--export <zipfile>	Zips up all configuration files to <zipfile> If no path is specified, the file will be created in the \$Data Instance Director_HOME\runtime directory. If no .zip suffix is specified, then it will be appended. The file is created in zip format and may be read with commonly available tools such as WinZip.
-i	--import <zipfile>	Restores all master node configuration files from <zipfile> The CofioHub service must be stopped prior to importing.
-c	--config <cfgfile>	Used with --export and --import. Specifies a configuration file used to control what files are included and excluded in the archive. By default, the cfgfile name mastersettings.cfg is assumed. (See the section, About the mastersettings.cfg file .)
-p	--preclear True False	Used when importing a zip file. Indicates whether existing master node configuration files are to be deleted prior to restoring them from the archive.

Usage example

Backing up the master node:

```
mastersettings.exe -e mastercfgfiles.zip
```

About the mastersettings.cfg file

mastersettings.cfg is an XML file that contains information indicating which Data Instance Director files are to be zipped up (exported) or deleted prior to a restore (imported). This file is located in the following directory:

```
$HDID_HOME$db\config
```

The mastersettings.cfg file may contain information as described in the following table.

Item Name	Details
Path	The path to be backed up. The path specified is relative to the \$Data Instance Director_HOME directory.
Depth	The depth, defined as a value, within the path to be backed up: -1 (default) backs up all files in all subdirectories. 1 backs up only the files in the path. 2 backs up all files in the path and within subdirectories of the path (but not the subdirectories themselves).
Export Mask	Controls which files in the path are backed up. May contain the following entries: (If no export mask is specified then all files will be matched and exported).
	NameMask A file name mask such as *.cfg.
	MatchAction Indicates how to process matching files: Either CopyFile or IgnoreFile
	NonMatchAction Indicates how to process non-matching files: Either CopyFile or IgnoreFile
Import Mask	Controls what is cleared from the path prior to a restore. May contain the following entries: (If no import mask is specified then all files will be matched and deleted prior to the restore unless the option --preclear false is specified at the command line).
	NameMask A file name mask such as *.cfg.
	MatchAction Indicates how to process matching files: Either DeleteFile or IgnoreFile
	NonMatchAction Indicates how to process non-matching files: Either DeleteFile or IgnoreFile

Examples

The following XML tells mastersettings.exe to export all files in and below the main Data Instance Director configuration directory:

```
<item name="Entry" argtype="list" >
  <item name="Path" argtype="single" >
    <value type="string" >db/config</value>
  </item>
  <item name="Depth" argtype="single" >
    <value type="int32" >2</value>
  </item>
</item>
```


The following XML tells mastersettings.exe not to export any files from the media manager events directory, but to delete all events from that directory prior to a restore. This will ensure that all events that are logged prior to an import are deleted, because they will not be relevant after the master settings have been restored.

```
<item name="Entry" argtype="list" >
  <item name="Path" argtype="single" >
    <value type="string" >db/mediamanager/events</value>
  </item>
  <item name="Export Mask" argtype="list" >
    <item name="MatchAction" argtype="single" >
      <value type="string" >IgnoreFile</value>
    </item>
  </item>
  <item name="Import Mask" argtype="list" >
    <item name="MatchAction" argtype="single" >
      <value type="string" >DeleteFile</value>
    </item>
  </item>
</item>
```

Restoring node configuration files

You can use the mastersettings.exe command line utility to restore a master node's configuration files from a previously created zip file.

To restore the master node configuration files:

1. Install a new master on a machine with the same IP address as the previous master.
2. If the repository node is separate from the master, run the setconfig -m command to point the repository to the new master and re-authorize it. (This also clears any SSL certificates for the 'old' master.)
3. Locate the zip file from the **Restore** screen and restore it to the master.
4. Stop the CofioHub service on the master.
5. Rerun the setconfig -m command on the repository node to point to the master. (This also clears any SSL certificates from the 'interim' master.)
6. Import the archived master settings using the mastersettings -i option.
7. Start the CofioHub service on the master.



Note: If you did not install a new master on a machine with the same IP address as the previous master, then you must run the `setconfig -m` command for all nodes to specify the new master IP. Complete this task before proceeding to step 8. For command details, see the following section, [Changing a Client Machine's Master node](#).

8. To finish the restoration, complete the following:

- Redistribute rules.
- Resynchronize policies.
- If the master contains a media manager for tape backups, import each tape from the **Storage** screen's **Media** tab.

Changing a client machine's Master node: setconfig.exe

If you need to change a client machine's Master node, you can do so by running the setconfig.exe utility from the command line interface.

The following is information about the most commonly used options and command usage.

Options

Abbreviated Argument	Full Argument	Description
-h	--help	Displays command line parameter usage instructions.
-m	--master <IP address>	Set the IP address of this node's master. This option is used with -f
-f	--force	Forcibly set the node's Master. This option is used with -m if the node in question is already connected to an authorized master.
-x	--fixed <IP address>	Sets the External-facing address of the Master when backing up over the Internet. This is the IP address that will be entered when installing Internet connected nodes.
-l	--list	Lists the node's settings: node/master name and IP addresses, master, firewall and encryption state and capabilities installed.
--filterrulescheck		Check the filter rules for the file system are up-to-date and triggers a resync if they needed updating.
--deletesslcert <node name>		Delete the SSL certificate for the local node/specified node.
--firewall true		Provides the same behavior as selecting 'Internet connected node' during the software installation.

Usage example

Setting the Master node's IP address:

```
setconfig.exe --master 192.156.34.0
```

Changing the behavior of a Data Instance Director node: diagdata.exe

diagdata.exe is a CLI tool that is used to aid diagnostic testing of Data Instance Director. It provides the following functions:

- stopping and starting of Data Instance Director services
- resetting of Data Instance Director system state
- raising and lowering of trace level for diagnostics
- collection of Data Instance Director state information to aid investigation of issues

The following is information about the most commonly used options and command usage.

Options

Abbreviated argument	Full argument	Description
-? -h	--help	Display all options.
-f	--full	Produce a full dump of the Data Instance Director configuration on the system, including all trace files. Results are placed in a ZIP file named: diag_<node name>_<YYYYMMDDhhmm>.zip in the directory: <Data Instance Director install>\runtime This ZIP file may be requested by Hitachi support when investigating complex end user issues.
-m	--mini	Similar to -f but only saves a recent portion of all trace files.
-t	--trace	Sets the trace level of the node. Valid arguments are: TraceAlways – logs only salient events, minimizing trace file size. This is the default level. The trace output increases in detail with the following levels: TraceTesting, TraceNormal, TraceVerbose upto TraceDebug – logs everything and generates trace files on disk of up to 4GB (configurable). Trace files are placed in TRA files in the directory: <Data Instance Director install>\runtime\trace
--stop <argument>		Stops the requested Data Instance Director service. Options are all, hub, filter name or filters.
--start <argument>		Initiates Starts the requested Data Instance Director service. Options are all, hub, filter name or filters.

Abbreviated argument	Full argument	Description
--deltrace		Deletes the contents of <Data Instance Director install>\runtime\trace
--reset		Resets the state of the Data Instance Director node and flushes cache and logs. Can only be run if Data Instance Director services are stopped (see --stop above).

Usage examples

Stopping the Data Instance Director services on any node:

```
diagdata.exe --stop all
```

Setting the trace level of the Data Instance Director services on any node:

```
diagdata.exe --trace TraceAlways
```

Reducing the repository's size: repomaint.exe

If you need to reduce the size of the repository and free unused disk space, you can do so by running the executable file `repomaint.exe` from the command line interface.

We recommend running `repomaint.exe` once excessive unused space has accumulated in the repository; for example, after a large set of data has been manually deleted, a large number of snapshots have expired, or a number of stores became inactive and were deleted.

The following is information about the most commonly used options and command usage.

Options

Abbreviated Argument	Full Argument	Description
-h	--help	Displays command line parameter usage instructions.
-g	--storegroup	Use -g followed by the Storage Group name. If no -g argument is provided, then the default Storage Group, StorageGroup01, is used.
-r	--reclaim <argument>	Scans the repository when no cleanup process is currently running. If any error condition exists, it is reported. We recommend completing a scan and correcting any reported errors before running <code>repomaint.exe</code> with the --compact option.
--compact <argument>		Analyzes the repository, reorganizes the stored data, then deletes the unused space.

Usage examples

Scanning the repository:

```
repomaint.exe --reclaim -g StorageGroup01
```

Compacting the repository:

```
repomaint.exe --compact -g StorageGroup01
```



Note: Before running `repomaint.exe` with the --compact argument, unmount the repository:

1. Launch Data Instance Director.
2. From either the **Monitor** or **Storage** screen, right click on the storage group and select **Unmount**.

After `repomaint` has completed, remount the repository so it will shrink to fit the data it contains.

Customizing alert notifications

If your site needs to receive alert notifications based on real-time or audit log events, then you can create a custom configuration file with instructions for which script/program Data Instance Director should run as an event handler, and the arguments to pass to it.

Custom event handlers are added as checkboxes in the **Notification Method** group box displayed on the **Add Notification Condition** dialog (Figure A-6-1) accessed via the **Log Manager** screen:

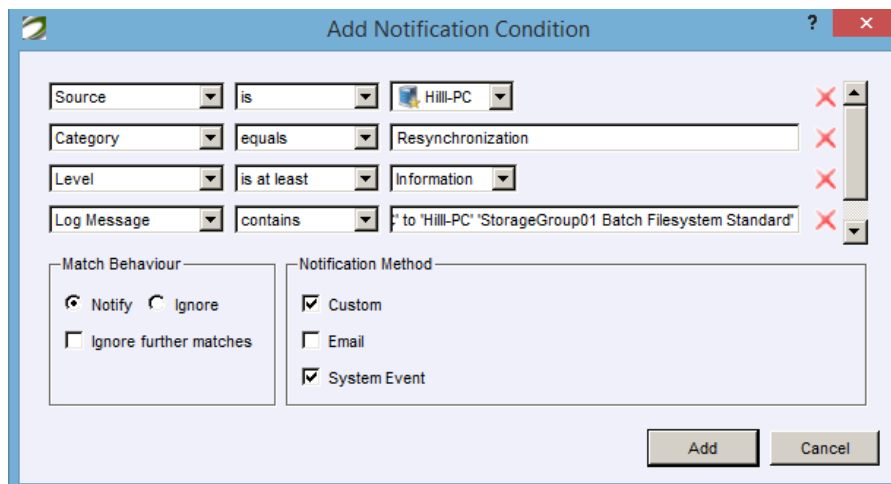


Figure A-6-1 Add notification condition dialog

By default, all Data Instance Director notification handler configuration files are stored relative to the installation directory in:

`\db\config\notification`

The path to the handler executable must be relative to the `\bin` directory where Data Instance Director is installed. By default this is:

Windows Installations:

`C:\Program Files\Hitachi\HDID\bin`

Linux Installations:

`/opt/Hitachi/HDID/bin` for

The built-in **Email** and **System Event** notification methods are configured by `email.cfg` and `systemlog.cfg` and implemented by `nehemail.exe` and `neheventlog.exe` respectively (we recommend using the 'neh' prefix to identify notification event handlers).

The first parameter passed to the handler is always the path to a log message file (described in the next section). You can also specify your own additional arguments for Windows and Linux that are stored separately, so the same configuration file can be reused on both operating systems.

The custom handler is specified in the format shown in the following example (customeventhandler.cfg):

```
<cofioxmllist>
  <!-- text for custom notification method checkbox -->
  <item name="HandlerName" argtype="single" >
    <value type="string" >Custom</value>
  </item>

  <!-- default state of custom notification method checkbox -->
  <item name="Default" argtype="single" >
    <value type="bool" >False</value>
  </item>

  <!-- handler executable or script name -->
  <item name="HandlerExecutable" argtype="list" >
    <item name="WinExecutable" argtype="single" >
      <value type="string" >nehcustom.exe</value>
    </item>
    <item name="UnixExecutable" argtype="single" >
      <value type="string" >nehcustom</value>
    </item>
  </item>

  <!-- arguments 2, 3, ..., n passed to custom handler -->
  <item name="WinHandlerArgs" argtype="list" >
    <item name="HandlerArg" argtype="single" >
      <value type="string">windows custom arg</value>
    </item>
  </item>
  <item name="UnixHandlerArgs" argtype="list" >
    <item name="HandlerArg" argtype="single" >
      <value type="string">unix custom arg</value>
    </item>
  </item>
</cofioxmllist>
```


Any text printed to the standard output device by the handler will be included as an attachment to an error log message with the words:

"<handler name> notifier failed. (Attachment 1) *** Attachment count 1 ***"

The exit code of the handler is ignored.

About the log message file

The log message file name is passed as the first parameter to all NEH handlers. It consists of a list of value pairs named:

NotificationDetailFieldName – a string identifying the parameter

and

NotificationDetailFieldValue – a value for the parameter

By reading these name–value pairs, the handler is able to analyze the log entry that caused it to be invoked and respond accordingly.

Log message files have a variable number of fields and may contain additional fields, such as a SessionID or a Data Source. The following table lists common fields.

GUI heading	Log file name	Description
 000008934 (padlock icon)	Immutable	The log message will not be cleared when erasing the logs.
Sequence Number	SequenceNumber	All logs have an ascending sequence number.
Master Date	MasterDate	A 32 bit timestamp of when the master received the log message.
Local Date	LocalDate	A 32 bit timestamp of when the log message was sent.
Log Origin	Source	The node from which the log message originated.
Category	Category	The log message category.
Level	Level	A value between 1 and 4 with 1 being detail and 4 being error.
 Custom notifier (paperclip icon)	Attachments	The text of any attachment to the log message.
Log Message	Log	The body of the log message.
acknowledged 18/09/2014 (green highlight)	Acked	Indicates if the log message is acknowledged (0 or 1).

GUI heading	Log file name	Description
-	BuildNumber	The version of Data Instance Director on the node that sent the log message.

Testing

Use the sendlog CLI in the Data Instance Director\bin directory to create logs to test that an NEH functions properly. For example, running the following command will create an information-level log message in the 'Test' category, with the message 'Test log message'.

```
sendlog -c "Test" -l 2 -m "Test log message"
```

Log message files only exist for the duration of the handler execution, after which they are deleted. To view the contents of a log message file, create an event handler that dumps the file to standard output. In Windows the following command could be used:

```
more < %1
```

A log message entry will then appear on the **Log Manager** screen with an attachment containing the above batch command, followed by the XML formatted log message that triggered it.

Creating a blacklist for stubbed files

If needed, you can specify the files and directories that you want to exclude from being stubbed/removed, and you can block specific programs from reading and unstubbing stubbed files. The `sourcestubdaemon.cfg` file lets you create such exclusions:

- The Stubbing Blacklist prevents files and directories from being stubbed. This resides within the XML block that is headed with the tag `<item name="StubbingBlacklist" argtype="list">`.
By default, `desktop.ini` files, `.lnk` files and the Windows directory are not stubbed.
- The Stub Access Blacklist prevents programs from accessing stubbed files. This resides within the XML block that is headed with the tag, `<item name="StubAccessBlacklist" argtype="list">`
By default, no programs are blacklisted.



Note: The `Sourcestubdaemon.cfg` file makes use of QT regular expressions. For additional information, refer to the following site:

<http://qt-project.org/doc/qt-5.1/qtcore/qregexp.html>

Usage examples

Stubbing Blacklist entry preventing text files from being stubbed or removed:

```
<item name="Entry" argtype="list" >
  <item name="Category" argtype="single" >
    <value type="string" >File</value>
  </item>
  <item name="Path" argtype="single" >
    <value type="string" >.*\.txt</value>
  </item>
  <item name="Actions" argtype="single" >
    <value type="string" >Stub, Remove</value>
  </item>
</item>
```

Stubbing Blacklist entry preventing a file, `projections.doc`, from being stubbed:

```
<item name="Entry" argtype="list" >
```

```

<item name="Category" argtype="single" >
    <value type="string" >File</value>
</item>
<item name="Path" argtype="single" >
    <value type="string" >.*\\projections\.doc </value>
</item>
<item name="Actions" argtype="single" >
    <value type="string" >Stub</value>
</item>
</item>

```

Stubbing Blacklist entry preventing the \Program Files directory from being stubbed:

```

<item name="Entry" argtype="list" >
    <item name="Category" argtype="single" >
        <value type="string" >File</value>
    </item>
    <item name="Path" argtype="single" >
        <value type="string" %programfiles%\\.* </value>
    </item>
    <item name="Actions" argtype="single" >
        <value type="string" >Stub</value>
    </item>
</item>

```

StubAccessBlacklist preventing notepad from accessing stubbed files:

```
<item name="Entry" argtype="list" >
  <item name="Process" argtype="single" >
    <value type="string" >C:\\WINDOWS\\system32\\notepad.exe</value>
  </item>
</item>
```


Troubleshooting

This chapter provides guidelines for how to troubleshoot issues that might occur with Microsoft Exchange Mailbox and HCP nodes.

- ❑ [Troubleshooting the Exchange Mailbox](#)
- ❑ [Troubleshooting HCP](#)
- ❑ [Troubleshooting a down machine](#)
- ❑ [Restoring a Hyper-V Virtual machine](#)
- ❑ [Troubleshooting Hitachi NAS](#)
- ❑ [Troubleshooting Hitachi VSP](#)
- ❑ [Troubleshooting Oracle Database snapshotting on VSP](#)

Troubleshooting the Exchange Mailbox

Node creation

Failed to authorize node <name> (Invalid username/password)

Exchange authorization credentials have to be specified when authorizing the node, even though they are not validated by the rules compiler. Incorrect credentials will only be highlighted when the rules are actually run. An error will be logged in the **Log Manager** screen if there is a problem with the supplied credentials when the rules are distributed (see Logs section below).

Reauthorize the Exchange node by locating it in the **Unauthorized Nodes** list on the **Node Manager** screen. Ensure that valid credentials are entered when authenticating the node.

Rules compilation

Compiler error 10256

Exchange proxy nodes can only implement *Exchange Mailbox* policies.

Deselect all other policies assigned to the Exchange proxy node in the **Data Flow** screen and recompile the rules.

Logs

Failed to initialize rules: [...] (0x8007052e) [...]

"Failed to initialize rules: failed to open active directory object '<LDAP path>' (0x8007052e) Logon failure: unknown user name or bad password"

The Agentless node credentials supplied when the node was authorized are incorrect.

Verify the credentials by locating the Exchange node in the **Authorized Nodes** list on the **Node Manager** screen and selecting **Edit Node** from the context menu. Recompile and redistribute the rules after any changes have been made.

Rules resolved to no user mailboxes

The rules for the proxy node do not resolve to any user mailboxes. It may be that the node doesn't have any specific rules, or that the specified groups do not contain users, or that no mailboxes were identified on the system.

Review the **Attributes** tab for the *Exchange Mailbox* in the **Classification** workspace on the **Policy Definition** screen.

No mailboxes to synchronize [...]

"No mailboxes to synchronize for full synchronization of '<node>' to '<node>' '<store>', check requested users and groups"

The Exchange Mailbox classification defined within the selected policy does not resolve to mailbox enabled accounts.

Verify the following on the **Attributes** tab for the *Exchange Mailbox* in the **Classification** workspace on the **Policy Definition** screen:

- user mailboxes are included by providing the user's domain email address
- groups are included by the correct display name
- those groups contain the one or more members.

Failed to initialize rules: [...], database may have moved [...]

"Failed to initialize rules: Failed to find database for user '<user display name>', database may have moved or invalid permissions to access object"

The database for the specified user was not found. Either the database moved during initialization or the 'backupmanager' account does not have permissions to access the Active Directory object.

To resolve this, try the following:

- Validate the object exists in the Active Directory database
- Check impersonation permissions
- Retry the operation which caused the failure.

User without mailbox

If the rules specify a user mailbox which does not exist, this will not result in any error log being produced; that user mailbox will simply be ignored during the backup. Ensure that mailboxes are correctly specified.

Failed to initialize rules: [...] (0x80072030) [...]

"Failed to initialize rules: failed to open active directory object at '<LDAP path>' (0x80072030) There is no such object on the server."

The impersonation permissions for the 'backupmanager' account are incorrect. Check the permissions for the backupmanager in Active Directory.

The account does not have permission to impersonate the requested user

"Exception while performing a full synchronization of '<policy>' to '<node>' '<destination>'. The account does not have permission to impersonate the requested user."

The impersonation permissions for the 'backupmanager' account are incorrect. Check the permissions for the backupmanager in Active Directory.

Connection did not succeed. Try again later.

Check that the database is mounted and is available.

Sender had problems while performing a full synchronization of [...]

The attachment to the log message contains the following text:

Attachment:
=====

Unknown
=====

EWS Exception SSL_ERROR_SSL
error:14090086:SSL
 routines:SSL3_GET_SERVER_CERTIFICATE:certificate verify failed

This log entry is generated if there is no valid SSL certificate or there is a Self-Signed SSL Certificate.

To resolve this issue, locate the exchangeaccs.cfg file within the following directory on the Exchange proxy node:

```
<install dir>\db\config\exchangeaccs.cfg
```

Then, enter the following code within the `<xml:root>` tag of exchangeaccs.cfg:

```
<item name="NoAuthenticateSSLCertificate" argtype="single">  
    <value type="bool">true</value>  
</item>
```

Troubleshooting HCP

Node creation

Couldn't connect to server or HTTP status 503 Service Unavailable

These messages might appear when authorizing an HCP node. Check that the Management API is enabled in the Tenant Console on HCP for that tenant.

HTTP status 403 Forbidden

This message might appear when authorizing an HCP node or might be logged when tiering or stubbing. Check all the details are correct including username, password and tenant.

Couldn't resolve host name

This message might appear when authorizing an HCP node or might be logged when tiering or stubbing. Check that the HCP can be contacted; for example, by pinging the tenant and/or the IP address.

If system level credentials were provided then ensure that the 'Allow HCP system-level users to manage me and search across my space' checkbox is checked on the HCP's Tenant's Overview screen.

Logs

413 Request Entity Too Large

This message might be logged when tiering or stubbing. Check on HCP that the namespace has not exceeded its quota.

In operation

Files, email bodies, or attachments are not being stubbed/removed.

Check the logs for errors/warnings in the 'Stubbing' category.

Check that the source is connected to the repository node that was used to tier the data.

If an email message is moved, then it may reset the time before it's stubbed/removed.

After a file is opened, the file's access time is not modified.

By default, Windows Vista/Server 2008 (and later) do not update a file's access time. If you plan to stub files based on the **Was not accessed in the last** criterion of the **Tier** operation's **Stub/Remove Source Item Options**, then ensure that the file access timestamp will be correctly updated. This one-time change is applied by modifying a registry key on the client node.

To set the file access time handling:

1. Locate the registration entries (.reg) file
 <installation path>\HDID\db\config\enableLastAccessTimeUpdate.reg
 on the client node:
2. Double-click the .reg file to modify the registry.
3. Restart the client node.



Notes:

- Even after applying enableLastAccessTimeUpdate.reg, when files are opened by Microsoft Notepad, the file's accessed time might not be modified. In this case, consider the types of files that are identified for stubbing and the associated stubbing criteria.
 - If your operating system is other than Windows XP or Windows 2003, you can undo the setting change made by enableLastAccessTimeUpdate.reg by double-clicking disableLastAccessTimeUpdate.reg. This file is also located in the <installation path>\HDID\db\config folder.
-

Troubleshooting a down machine

The machine from which I've backed up using HDID has a damaged OS

You can restore a full machine backup by using the Hitachi Data Instance Director disaster recovery software. Refer to the document, *Hitachi Data Instance Director Disaster Recovery Reference*, for information about how to create an ISO image, and how to use the image to restore a designated destination machine from an existing backup.

Restoring a Hyper-V virtual machine

I need to restore a single Hyper-V virtual machine.

Restoring all Virtual Machines (non-cluster):

1. Shut down the originating Virtual Machine.
2. Stop the Hyper-V Virtual Machine Management service.
3. Restore the entire snapshot.

Restoring a Single Virtual Machine (Cluster):

When restoring a single Virtual Machine, either the entire Virtual Machine can be restored, or the Virtual Hard Disk can be restored and a new Virtual Machine can be created from the restored Virtual Hard Disk.

To restore an entire Virtual Machine:



Note: The following procedure cannot be used to restore individual Hyper-V Virtual Machine snapshots. Before you begin, check the following:

- On the Hyper-V machine, shut down the originating Virtual Machine.
 - Some versions of Hyper-V may require stopping the **Virtual Machine Management Service**.
-

1. Identify the relevant virtual machine files within the Data Instance Director snapshot. (The .xml, .gui, and .bin files use a unique string to identify the virtual machine.)
2. Within the Windows PowerShell, translate the string name to the virtual machine name, using the applicable command:
 - Server 2008 R2:
`Get-WmiObject -Namespace root\virtualization -class msvm_computersystem | select elementname, name | ft -auto`
 - Server 2012 R2:
`Get-VM | select VMName, VMId`
3. Delete the originating Virtual Machine within the **Hyper-V Manager** (not in the **Failover Cluster Manager**).
4. Check the following:
 - The directory and XML file which have the name identified in step 1.
 - The Virtual Hard Disk (.vhd, or .vhdx) file which needs to be restored, and which is located within the snapshot.
 - Any file that is prepended with the Virtual Machine name.
5. Restore to the original location with the **Differentiate file name** option unchecked.
By default, the Virtual Machine should be re-added to the Hyper-V Manager. The process can take up to an hour.

Troubleshooting Hitachi NAS

Node authorization

Authorization Failure dialog: Failed to authorize/deauthorize node [...]

If a dialog is displayed stating that the HNAS node failed to authorize, with the additional error “The API call failed.” Check that the **Admin EVS IP** address entered on the **Agentless / Hardware Node** authorization dialog is correct. The correct IP address can be found in the **Server Status Console** pane on the **HNAS Home / SMU Administration / Managed Servers** web page.

Directory clone

Handler call failed: [...] cannot overlap with source directory

The error “Handler call failed: Since destination directory needs to be on same file system as source, and, therefore root directory cannot be a source. Or source directory tree overlaps with specified destination directory. Or Source directory doesn't exist. Or source is not a directory. Or destination is not a directory.” is logged when trying to do a **Snapshot Type** with **Directory (using Directory Clone)** selected.

You cannot take a snapshot of the root of the file system using Directory Clone.

Replication

API call failed

An error containing “API call failed” is logged when attempting to replicate.

Check that PowerShell scripts can be executed locally on the machine by running “Get-ExecutionPolicy” in PowerShell. If the output is “Restricted” or “AllSigned” then run “Set-ExecutionPolicy RemoteSigned”.

Troubleshooting Hitachi VSP

Snapshotting

Handler call failed: [...] failed to create snapshot pair [...]

Hitachi VSP has a limit of 1024 snapshots per LDEV. When this limit is reached, no more snapshots can be created. If you set an overly aggressive RPO combined with long retention times then this limit can be quickly reached.

Handler call failed: [...] snapshots that are in a different pool

Hitachi VSP will only allow one pool to be used for snapshots (S-VOL) per primary volume (P-VOL). This error message indicates that the P-VOL in question already has a pool assigned for snapshots, and this is not the same pool as defined in the policy.

Handler call failed: refreshLogicalDevices Failed:: Failed run process "Permission denied with the Resource Group [raidcom.exe get ldev -ldev_id 2503 -l2500]"

This error message occurs where the user account does not have permission to access the LDEV because the LDEV is not in the resource group the user account can access. This can be confirmed by running raidcom manually:

```
C:\HORCM\etc>raidcom.exe get ldev -ldev_id 2503 -l0
raidcom: [EX_EGPERM] Permission denied with the Resource Group
```


Troubleshooting Oracle Database snapshotting on VSP

OracleHandler [...]: Oracle Database Files and Redo in the [...]

This warning message is logged during backup operations when the data and redo files of the Oracle database are located in the same directory. See the section entitled “Snapshotting an Oracle database on Hitachi block ” for advice on locating Oracle files.

OracleHandler [...]: normal or high redundancy used; revert

This warning message is logged during backup operations, followed by the error message:

“Handler call failed: [...] normal or high redundancy used; revert not possible with one mirror”

if you attempt to revert a database running on ASM in normal or high redundancy mode. See “Snapshotting an Oracle database on Hitachi block ” for advice on ASM Diskgroup support for reverting Oracle databases.

OracleHandler [...]: mount for ‘Oracle DB: [...]’ finished [...]

This information message is logged when an Oracle database is mounted. The attachment lists the destination paths where the Oracle metadata files have been placed.

Handler call failed: [...] RAC Lock Database failed, check if [...]

This error message is logged, during online snapshot operations, for all but one node in a RAC environment. Only one RAC node will succeed to lock the database prior to online snapshotting, all other nodes in the RAC will fail. Check that the operation has succeeded on exactly one of the RAC nodes.

Handler call failed: [...] ASM Diskgroups [...] mounted

This error message is logged if you attempt to mount an Oracle database snapshot to location where one is already mounted. You cannot mount a second snapshot of the database to a location where one is already mounted.

Checks to perform when an online backup or mount fails

In the event of a failure, Data Instance Director attempts to revert the operations performed up to the point of that failure, however this cannot be guaranteed to be successful. For this reason we describe here what should be checked and if necessary performed manually:

If an Oracle online backup fails, then check that the operations described in the attachments to the log entry 'backupOnline Worksteps' are reverted. It is especially important that the step 'endOnlineBackup' is performed. If not then execute the steps described in the attachment to the log entry 'endOnlineBackup'.

If an Oracle offline backup fails then check that the database is started. If not then start the database.

If a mount fails then check if the status of the snapshot is mounted. If so, then perform an un-mount. In addition to this please check that an eventual ASM Diskgroup (asmcmd lsdg) is un-mounted and removed from the resource group (asmcmd umount <ASM_DISKGROUP>; srvctl remove diskgroup -g <ASM_DISKGROUP>). Do not try to issue a mount to another server prior a successful un-mount operation.



This glossary defines the special terms used in this document. Click the desired letter below to display the glossary entries that start with that letter.

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
---	-------------------	-------------------	-------------------	-------------------	---	---	-------------------	-------------------	---	---	---	-------------------	-------------------	---	-------------------	-------------------	---	-------------------	-------------------	-------------------	---	-------------------	---	---	---	---

A

Archive

A copy that is created for long-term retention.

Asynchronous replication

Transactions are held in memory before being sent over the network. If the network is unavailable then transactions are written to disk and sent to the destination machine when the connection is re-established. Asynchronous replication is optimal for connections with sporadic efficiency.

Asynchronous journaling

Transactions are written to disk and also placed in a journal log file, to protect against data loss in the event of a system failure. Transactions from the log file are sent to the destination machine.

B

Backup

A copy that is created for operational and disaster recovery.

Bandwidth throttling

Used to control when and what proportion of available network bandwidth is used by Data Instance Director for replication.

Batch backup

A process by which the repository is updated periodically using scheduled resynchronizations. This method involves a scan of the source machine's file system, but only the changed bytes are transferred and stored. This method is useful for data that does not change often, such as data contained on the operating system disk. Linux based source nodes are *only* able to perform batch backups. See CDP below.

Blocking

Data Instance Director can be used to block access to files or directories on a given source machine. Only source nodes that are running on a Windows OS can be blocked.

Warning: Applying blocking policy to data classifications such as OS Type may lock up the operating system and require the assistance of product support to resolve.

C

Continuous Data Protection (CDP)

A method of capturing the state of a file system in near real time. CDP shares much of the functionality of Live Backup, except that RPO is measured in minutes, data is retained for a much shorter period of time and is not indexed by the MDS. Typically, CDP and Live Backup are used in conjunction. CDP is only supported on source nodes running the Microsoft Windows operating system.

D

Data flow

Identifies the data sources, movers and destinations participating in a backup, along with interconnection paths between them. Policies are assigned to each node to determine what type of data is backed up.

Deduplication

A method of reducing the amount of storage space that your organization requires, to archive data, by replacing multiple instances of identical data with references to a single instance of that data.

Destination node

A machine that is capable of receiving data for the purposes of archiving. This machine might be the Data Instance Director Repository, a mirror, a tape library, cloud storage or tiered storage.

F

Forest

The collective noun used to describe a group of tree structures without a common root node.

G

Groups

Within the Node Manager screen, you can assign multiple machines to one or more groups. Then, within the Data Flow screen, you can assign policies the nodes within these groups en-mass.

H

Hitachi Content Platform (HCP)

An object storage system recommended for large-scale archiving. Data Instance Director keeps only the file metadata in its repository, while HCP stores the data.

L

License key

A unique, alphanumeric code that is associated with the unique machine ID that is generated during the Data Instance Director installation. The license key must be activated in order to use the software.

Live backup

A backup technique that avoids the need for bulk data transfers by continuously updating the repository with changes to the source file system. This is similar to CDP but with longer retention periods and RPOs being available. Live backups perform byte level change updates whereas batch backups perform block level change updates.

M

Master node

The machine that controls the actions of all other nodes within the Data Instance Director network. Only one Master Node may be deployed in a network.

Metadata Store (MDS)

Records metadata that describes items that are held in repositories, tape libraries and HCP. The MDS supports indexing of stored data, thus enabling fast searches when locating data for restoration.

Mover

Defines the type of data movement operation to be performed between source and destination nodes, during the creation of a data flow. Batch movers perform block level data transfers on a scheduled basis, whereas continuous movers perform byte level data transfers on a near-continuous basis. Mirror movers are a specific type of continuous mover used to create uni- and bi-directional mirrors.

O

Outlook form

When viewing stubbed emails within Microsoft Outlook, a custom Outlook form must be installed that allows Data Instance Director to retrieve the attachment from HCP. The Outlook Form must be installed within an Organizational level Exchange Public Folder within an Exchange Public Database, so it can be accessible by all Outlook users.

P

Policy

A configurable data protection objective that is mapped to machines or groups, and to the data management agents that implement the policy. Multiple policies can be assigned to a single node.

R

Recovery Point Objective (RPO)

The frequency at which a backup will occur. This governs the point in time to which data can be recovered should a restore be needed.

Repository

A destination node that stores data from one or more source nodes. The Data Instance Director Repository supports live backup, batch backup, archiving, and versioning policies.

S

Source node

Any node (server, workstation or virtual machine) that hosts data to be protected by Data Instance Director. The source node has an Active Data Change Agent, which is responsible for monitoring the host file system and performing the relevant actions defined by the policies. Nodes need to be configured as a source node if they need to transfer locally stored data to a destination node, or implement data tracking, blocking and auditing functions. A node can be both a source and destination simultaneously.

Stubbing

The process by which the contents of a file, email body, or email attachment are replaced by a reference or link that points to the copy of the data that is stored on the HCP.

T

Tiering

A process in which data is taken from the repository and placed into HCP, at established backup cycles.

Tracking

Data Instance Director can be used to track access to files or directories on any given source machine. Each time a file is accessed, a log message is generated that lists the type of access and the user who accessed the file. Only source nodes that are running on a Windows OS can be tracked.

V

Versioning

Whenever a file is saved, any changes, are recorded. For instance, when applied to office documents, every time a document is saved, the previous version is retained and indexed. This allows a history of a file to be generated, showing how the file changed over time.

A

- Age Classification
 - usage for, 4-30
- Agentless nodes
 - creating policy types for, 3-38
- Alert notifications
 - creating a configuration file for, 3-138, 6-11
- Attributes
 - of authorized nodes, 4-6
 - of unauthorized nodes, 4-6
- Authorized nodes
 - attributes of, 4-6
 - icons used for, 4-3
- Azure
 - backing up file system files to, 3-77

B

- Backup operation
 - usage for, 4-33
- Badges
 - used for node states, 4-80
- Bandwidth usage
 - setting, 4-60
- Block
 - creating policy types for, 3-20
- Block attribute, 4-42

C

- CDP
 - creating policy types for, 3-2
- CDP operation
 - usage for, 4-34
- Clearing
 - error logs, 4-94
- CLI
 - using mastersettings.exe, 6-2
 - using setconfig.exe, 6-7, 6-8, 6-10, 6-15

D

- Data Flow
 - reverting, 4-71
- Data Flow screen
 - overview of, 4-53
- Data Flow window
 - using to configure mirror replication, 3-18
- Data Flows
 - constructing complex types, 5-10
- Data Instance Director
 - installing, 2-4
 - licensing, 2-18

- uninstalling, 2-20
- upgrading, 2-15
- Deduplication
 - Overview of, 1-8
- Disconnected node
 - indicator for, 4-81
- Documents Classification
 - usage for, 4-24

E

- Elements
 - using in the Data Flow screen, 4-56
- Error logs
 - clearing, 4-94
- Exchange Mailbox Classification
 - usage for, 4-21
- Exporting
 - logs, 4-94

F

- File Archiving
 - creating policy types, 3-8
- File Versioning
 - creating policy types, 3-8
- Files Classification
 - usage for, 4-27
- Forwarding
 - stubbed e-mail information, 2-25
- Topologies, 5-6

G

- Group Classification
 - usage for, 4-30
- Groups
 - Overview of, 1-3

H

- HCP
 - backing up file system files to, 3-77
- HCP indexing
 - prerequisites, 2-2
- Hitachi Block classification
 - usage for, 4-23
- Hitachi NAS classification
 - usage for, 4-23
- Hyper-V Classification
 - usage for, 4-20

I

- Icons

- used for authorized nodes, 4-3
- used for node states, 4-80
- used for unauthorized nodes, 4-3

Installation

- Microsoft cluster, 2-11
- Overview for Data Instance Director, 2-4

L

Licensing

- Data Instance Director, 2-18
- Microsoft cluster environment, 2-19

Log Manager screen

- overview of, 4-87

Logs

- exporting, 4-94

M

Macros

- usage for, 4-42

Metadata Store

- Overview of, 1-11

Microsoft cluster

- installation, 2-11

Microsoft Cluster

- licensing, 2-19

Microsoft Outlook form

- uninstalling, 2-25

Microsoft Outlook Form

- setting up for stubbed e-mails, 2-21

Mirror Destination

- overview of, 4-62

Mirror replication

- configuring using the Data Flow window, 3-18

Monitor screen

- overview of, 4-77

MS SQL Server Classification

- usage for, 4-19

Multimedia Classification

- usage for, 4-26

N

Node Manager

- overview of screen, 4-2

Notification Management

- overview of, 4-95

O

Oracle database

- snapshotting on Hitachi block storage, 3-127

Oracle Database classification

- usage for, 4-22

OS Type Classification

- usage for, 4-28

Owner Classification

- usage for, 4-28

P

Path Classification

- usage for, 4-16

Paths

- using macros to define, 4-42

Personal Email Classification

- usage for, 4-26

Policies

- Overview of, 1-2

Policy Classification

- refining using AND, 4-12

- refining using OR, 4-12

Policy Definition

- overview of screen, 4-9

Policy Types

- creating for agentless nodes, 3-38

- creating for Block, 3-20

- creating for CDP, 3-2

- creating for file archiving, 3-8

- creating for file backup to Azure, 3-77

- creating for file backup to HCP, 3-77

- creating for file versioning, 3-8

- creating for HNAS snapshotting, 3-119

- creating for Microsoft Exchange mailboxes, HCP, 3-63

- creating for MS Exchange Server backups, 3-27

- creating for MS Hyper-V backups, 3-27

- creating for MS SQL backups, 3-27

- creating for replication, 3-13

- creating for snapshotting Oracle on block storage, 3-127

- creating for tape library backup, 3-56

- creating for Track, 3-24

- creating for VSP or HUS VM replication, 3-98

- creating for VSP or HUS VM replication & snapshot, 3-110

- creating for VSP or HUS VM snapshotting, 3-88

Prerequisites

- HCP indexing, 2-2

Process Classification

- usage for, 4-31

R

Remote Path Classification

- usage for, 4-31

Replicate operation

- usage for, 4-35

Replication

- creating policy types, 3-13

Reports screen

- overview of, 4-98

Restore screen

- overview of, 4-119

- Reverting
 - a Data Flow, 4-71
- Rules Compiler
 - overview of, 4-73

S

- Schedule Manager
 - overview of, 4-47
- Setting up
 - Microsoft Outlook form, 2-21
- Smart Repository Sync
 - setting up, 5-8
- Snapshot operation usage for, 4-36
- Storage screen
 - overview of, 4-106
- Storage Type Classification
 - usage for, 4-17
- Stubbed e-mail information
 - about forwarding, 2-25

T

- Tape Library
 - creating a policy type for, 3-56
- Tier operation
 - usage for, 4-39
- Topologies
 - Forwarding, 5-6
 - Many-to-One, 5-2
 - One-to-Many, 5-4
- Track
 - creating policy types for, 3-24

- Track attribute, 4-41
- Transfer modes
 - Overview of, 1-5
- Troubleshooting
 - logs for Exchange, 7-2
 - logs for HCP, 7-5
 - node creation for Exchange, 7-2
 - node creation for HCP, 7-5, 7-7, 7-8
 - Oracle Database snapshotting on VSP, 7-11
 - rules distribution for Exchange, 7-2

U

- Unauthorized nodes
 - attributes of, 4-6
 - icons used for, 4-3
- Uninstalling
 - Data Instance Director, 2-20
 - Microsoft Outlook form, 2-25
- Upgrading
 - Data Instance Director, 2-15
- User Classification
 - usage for, 4-29

V

- Version operation
 - usage for, 4-38
- VMWare ESX Classification
 - usage for, 4-20
- VSP snapshotting
 - Prerequisites, 2-2

Hitachi Data Systems

Corporate Headquarters

2845 Lafayette Street
Santa Clara, California 95050-2639
U.S.A.
www.hds.com

Regional Contact Information

Americas

+1 408 970 1000
info@hds.com

Europe, Middle East, and Africa

+ 44 (0)1753 618000
info.emea@hds.com

Asia Pacific

+852 3189 7900
hds.marketing.apac@hds.com



MK-93HDID001-01