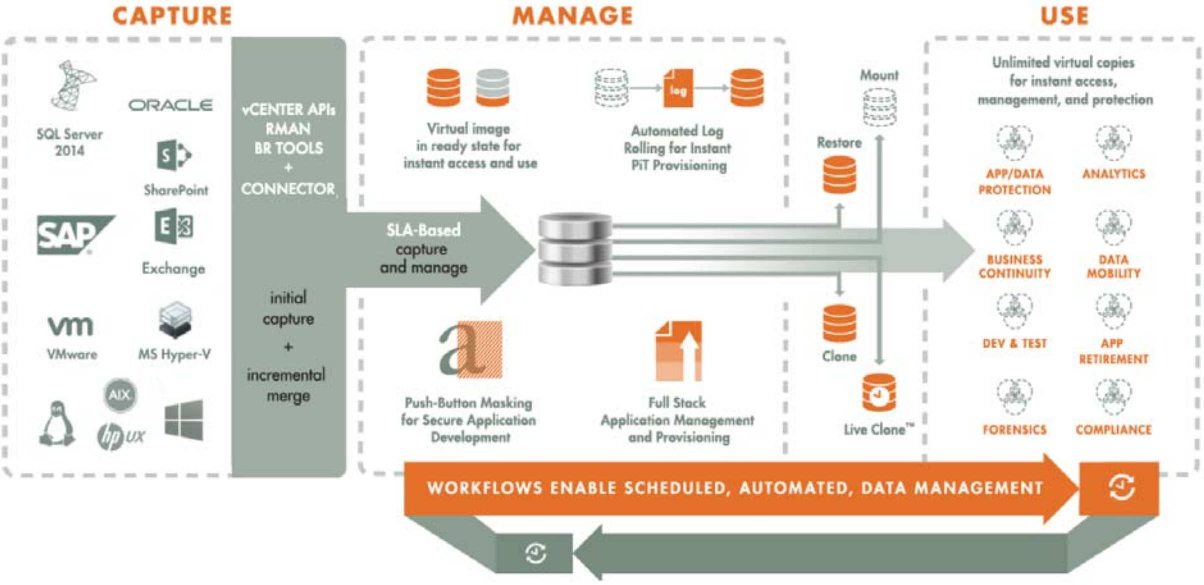
Getting Started with InfoSphere VDP Copy Data Management

IBM InfoSphere VDP 9.0

Contents

1 Introduction

IBM InfoSphere Virtual Data Pipeline (VDP) is a highly scalable copy data management platform that virtualizes application data to improve the resiliency, agility, and cloud mobility of your business. VDP virtualizes data in much the same way other technologies have virtualized servers and networks. IBM InfoSphere VDP enables you to capture data from production systems, manage it in the most efficient way possible, and use virtual or physical copies of the data whenever and wherever they are needed.


CAPTURE
MANAGE
USE
SQL Server 2014
ORACLE
SharePoint
SAP
Exchange
VMware
MS Hyper-V
AIX
UX
vCENTER APIs
RMAN
BR TOOLS
+
CONNECTOR
initial capture
+
incremental merge
SLA-Based capture and manage
Virtual image in ready state for instant access and use
log
Automated Log Rolling for Instant PiT Provisioning
Push-Button Masking for Secure Application Development
Full Stack Application Management and Provisioning
Mount
Restore
Clone
Live Clone™
Unlimited virtual copies for instant access, management, and protection
APP/DATA PROTECTION
ANALYTICS
BUSINESS CONTINUITY
DATA MOBILITY
DEV & TEST
APP RETIREMENT
FORENSICS
COMPLIANCE
WORKFLOWS ENABLE SCHEDULED, AUTOMATED, DATA MANAGEMENT

The IBM InfoSphere Virtual Data Pipeline (VDP)

Application data is captured at the block level, in native format, according to a specified SLA. A golden copy of that data is created, moved, and stored once and is then updated incrementally with only the changed blocks of data in an “incremental forever” model. Unlimited virtual copies of the data can be accessed instantly for use, without proliferating physical copies and taking up additional storage infrastructure.

2 Basic Concepts: Data and Storage

This chapter introduces the concepts of how IBM InfoSphere VDP captures, manages, and accesses data. Understanding these concepts will help you to be successful with IBM InfoSphere copy data management.

InfoSphere VDP Appliances

InfoSphere VDP Appliances capture and manage data locally and can replicate protected data to other InfoSphere VDP Appliances.

VDP Appliance Virtual Appliance

An InfoSphere VDP Appliance is a virtual machine. VDP Appliances are licensed by capacity. An InfoSphere VDP Appliance captures and manages data locally and can replicate protected data to other InfoSphere VDP Appliances.

VDP Appliance for AWS Cloud Appliance

An InfoSphere VDP Appliance for AWS appliance is a virtual machine that is licensed by capacity and resides in the AWS cloud space. An InfoSphere VDP Appliance captures and manages application data in the AWS cloud and can replicate captured data to another location. The underlying engine for VDP Appliance for AWS is VDP Appliance.

Local and Remote Appliances

Multiple InfoSphere VDP Appliances can be joined in any combination of primary and secondary relationships. An exchange of certificates is required to join appliances. Once joined, application data can be replicated between appliances. Local and remote are relative to where you are logged in. The InfoSphere VDP Appliance you are logged into is the local appliance and the other InfoSphere VDP Appliances are considered remote.

Appliance Sharing and Peer Relationships

The InfoSphere VDP Appliance can support joining any number of appliances at any number of sites. InfoSphere VDP Appliances can be joined in sharing or non-sharing mode.

You can join InfoSphere VDP Appliances in two modes:

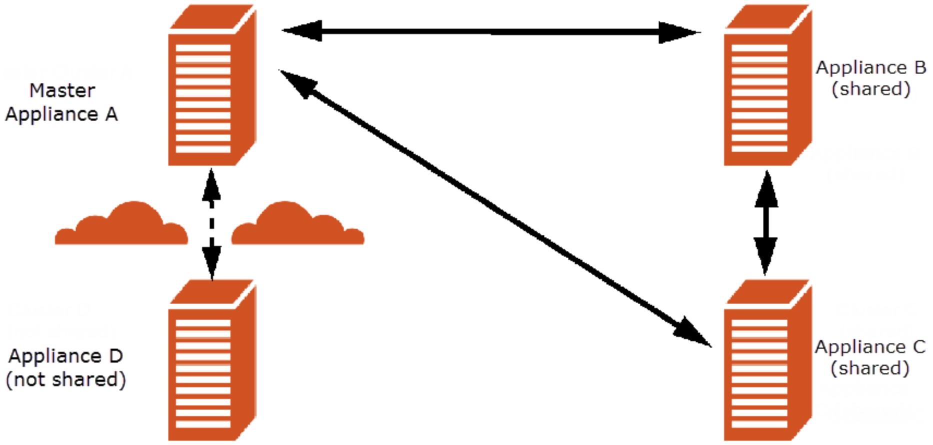
Sharing mode: Use this mode to share:

- Protected applications
- Security settings: Organizations, Users, and Roles
- Policy templates (but not resource profiles)

See Sharing Mode below for details. Section explains how to join appliances in sharing mode.

Non-sharing mode: Use this mode when you want appliances to perform remote backups for each other, for failover, and for dedup-async replication. Non-sharing mode is described in Non-Sharing Mode on page 5. An overview for joining appliances in non-sharing mode is in Non-Sharing Mode on page 5.

You can use both types of relationships within your enterprise. For example, you might use three joined sharing-mode appliances for day-to-day data protection, and a non-shared appliance for off-site disaster recovery.

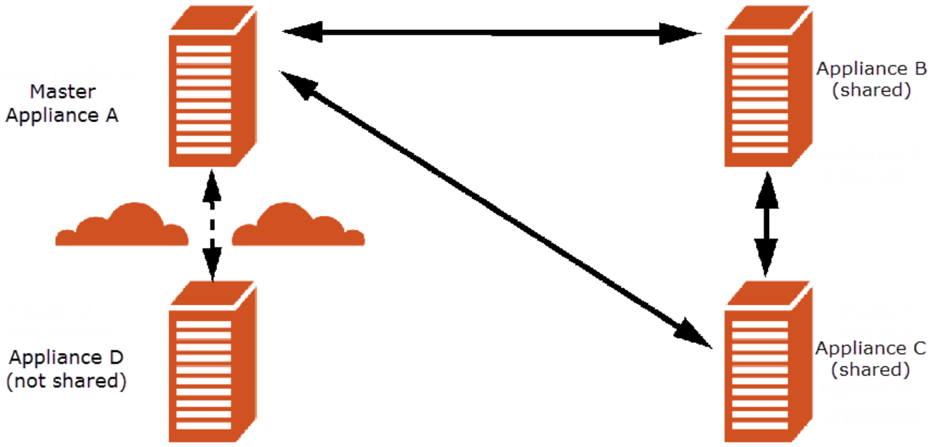


Appliances A, B, and C are Shared for Daily Operations; Remote Appliance D is for Disaster Recovery

Sharing Mode

When InfoSphere VDP Appliance B joins appliance A, the latter becomes the master and appliance B, the secondary.

- There can be only one master appliance in an InfoSphere VDP Appliance domain. There can be any number of secondary appliances.
- Masters are made, not born. The act of joining in shared mode creates the master-secondary relationship. Sharing mode makes a secondary of the appliance that joins. In the diagram above appliance B joined appliance A in sharing mode. At that moment appliance B became the secondary and appliance A became the master. The secondary joins the master.
- A master appliance, which is already in sharing mode (that’s why it’s a master) cannot join any other appliance, because it would become a secondary; once the master, always the master.
- All secondary appliances are peers to each other.
- An InfoSphere VDP Appliance can join any appliance in non-sharing mode. Non-shared appliances are peers to all appliances.
- An InfoSphere VDP Appliance cannot join a secondary appliance. In the diagram above appliance D cannot join appliance C or appliance B. Appliance D can join appliance A in the sharing or non-sharing mode.



Master Appliance A Shares Rules with Secondary Appliances B & C, but Not Remote Appliance D

When InfoSphere VDP Appliance B is joined to appliance A:

- If appliance B has no protected applications, and no users, roles, organizations, and policy templates defined, then appliance B gets users, roles, organizations, and policy templates shared from appliance A, but no applications.
- If appliance B has users, roles, policy templates, and organizations defined, they are deleted and replaced by those shared from appliance A.
- If appliance B has protected applications, they are preserved, but they are protected by the policy templates shared from appliance A.
- If appliances A and B both do not contain protected applications, users, roles, policy templates, and organizations, then when appliance B is joined to appliance A, no obvious changes occur. However, once you create users, roles, organizations, and policy templates on appliance A, they are shared with appliance B.

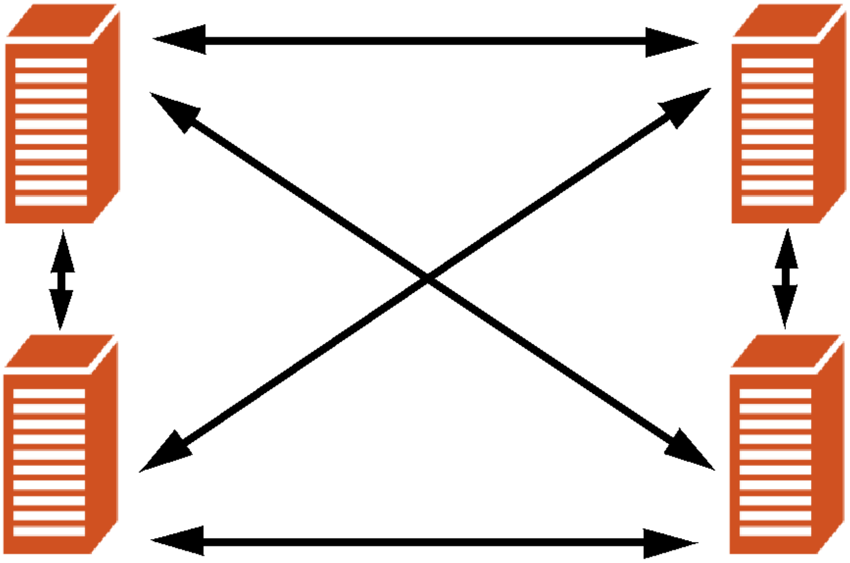
Non-Sharing Mode

Two InfoSphere VDP Appliances that belong to different InfoSphere VDP Appliance domains can be joined in non-sharing mode. You can join:

- One appliance with another appliance to enable replication from the former to the latter.
- One appliance with two appliances to enable different applications to replicate to separate appliances.
- Two appliances with one appliance to let the later act as a target for replication.

Non-sharing mode makes peers of the appliances that are joined. When two appliances join in the non-sharing mode, each appliance can be the destination for remote replication of data from the other appliance.

In the non-sharing mode, each appliance uses its own users, policies, templates, and organizations and protects only the applications assigned to it; peer appliances do not share users, policies, templates, and organizations.



The Master Appliance Shares Users and Policy Templates with Secondary Appliances

The types of replications supported by InfoSphere VDP Appliance are given below. All of these replications can use the appliance setup described above.

- Asynchronous replication of deduped data from the primary appliance to a remote appliance (Dedup-Async Replication, or DAR).
- Remote deduplication replication from the dedup pool of the primary appliance to a remote appliance.
- StreamSnap replication from the primary appliance to a remote appliance.

Storage Pools

An InfoSphere VDP Appliance uses pools of allocated storage to store data. The amount of space to be allocated is based on how data is managed (see Capture Mechanisms on page 19), how much data is involved, the type of data, its change rate, how long it will be retained, and whether or not the data is replicated to another InfoSphere VDP Appliance.

Snapshot Pool

The Snapshot Pool holds the most recent copies of your captured application data. Snapshot Pools retain protected data for short-term retention. Data is instantly accessible and not deduplicated. Policies determine how long data is kept in this pool and when or if data is deduplicated and moved to another pool.

Data that is replicated from a local InfoSphere VDP Appliance to a remote InfoSphere VDP Appliance via a Production to Mirror policy will land in the in the remote InfoSphere VDP Appliance's Snapshot Pool.

For more on the Snapshot Pool, see Snapshot Pool on page 8.

Dedup Pool

The Dedup Pool is used to retain deduplicated copies of snapshot data and to facilitate low-bandwidth replication. You can create a policy that writes VMware VM data directly to this pool, where it bypasses the Snapshot Pool and is deduplicated directly in this pool.

Data that is replicated from a local InfoSphere VDP Appliance to a remote InfoSphere VDP Appliance via a Dedup Backup DR policy will land in the in the remote InfoSphere VDP Appliance's Dedup Pool.

***Note:** Keep space for Dedup Pools separate from space for Snapshot Pools. Dedup operations are I/O intensive. Keeping Dedup Pools separate ensures that dedup operations will not hurt the performance of the Snapshot Pools.*

For more on the Dedup Pool, see Dedup Pool on page 13.

OnVault Pool

An OnVault Pool defines the storage that can be used by a Snapshot to OnVault policy or a Direct to OnVault policy (VMware VMs only). Data is not deduplicated when it is sent to an IBM InfoSphere OnVault Pool. IBM InfoSphere OnVault Pools are used for long-term storage, not for primary data storage. See also: Sending Snapshots to an OnVault Pool's Defined Storage on page 26.

Primary Pool

The Primary Pool, act_pri_pool000, is for InfoSphere VDP Appliance use. It is **not** a storage pool. Do not change the Primary pool or add a second pool unless instructed by IBM InfoSphere Support.

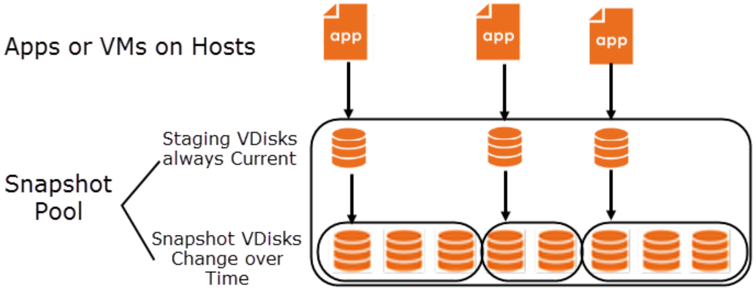
Snapshot Pool

The Snapshot pool (sometimes referred to as the Performance pool) holds "golden copies" of application data at the points in time specified by Service Level Agreement (SLA). Snapshot space is also used to rehydrate images from the deduplication pool. The amount of data consumed is determined by whether an existing snapshot can be used. This section includes:

Staging Disks

A staging disk is a VDisk created when an application is first protected. It is a copy of the production data as of the last backup invoked by the application's SLA. Each staging disk is associated with a number of snapshots on their own snapshot VDisks. The number of snapshots for each application or VM is determined by the SLA frequency of snapshot and retention period.

Because a staging disk is a complete copy of the production application or VM, each staging disk requires as much storage space in the Snapshot Pool as the protected application or VM requires in its production storage. Snapshots made from the staging disk reference the data in the staging disk, so they are much smaller. As subsequent backups change blocks in the staging disk, the original blocks are "pushed" into the snapshot VDisks, so the snapshot appears to have constant content but contains more and more blocks over time.



Apps or VMs on Hosts
app
app
app
Snapshot Pool
Staging VDisks always Current
Snapshot VDisks Change over Time

Virtualized Copy Data on Staging VDisks and Snapshot VDisks in the Snapshot Pool

Growth of Applications

If an application grows from 1TB to 2TB, a new 2TB staging disk is created. The original 1TB disk is preserved until all snapshots that depend on it are expired.

Note: *Windows staging disks up to 2 TB in size are MBR formatted. Those over 2TB are GPT formatted.*

Staging Disks for VMs and Out-of-Band Applications

When you protect a VM or an application, copies of the selected image are put into a dedicated virtual staging disk in the Snapshot pool. VDP creates a snapshot from the image on the staging disk, and stores the snapshot in the snapshot pool for the time specified in the SLA.

Staging disks for backups are allocated from the snapshot pool. The VDisk is thin-provisioned. Each snapshot created of that staging disk also consumes snapshot pool space, the amount depending on the application change rate.

Staging disks are not created or used for In-Band VMware VMs and applications.

An Exception for Direct-to-Dedup Protection for VMware VMs

VMware VMs protected direct-to-dedup do not go through a staging disk because the InfoSphere VDP Appliance can get changed-block information directly from the VMware layer. Hyper-V VMs and all other applications get changed-block information either via Oracle RMAN or the VDP Connector (using an IBM InfoSphere staging disk).

Understanding Snapshot Pool Consumption

The Snapshot Pool contains both the staging disks and the snapshot disks for every protected application or VM, plus any clones and mount images that you make.

The Snapshot Pool holds virtual disks, or VDisks. VDisks and VDisk consumption are explained in VDisks on page 12. Snapshot Pool space is consumed by four different kinds of VDisk:

Staging disks: Staging VDisks, usually called staging disks, hold the IBM InfoSphere golden copy of the application. Staging disks are retained for as long as an application is protected and at least one snapshot exists. See Staging Disks on page 8.

Snapshot VDisks: These are used to preserve the state of staging disks at specific points in time. Snapshots are retained until their expiration time, but the last snapshot will never expire unless the application is unprotected or it is explicitly expired.

Mountable VDisk: Mountable VDisks are mountable images created at restore time from either a snapshot on a snapshot disk or a deduped image that has been rehydrated to a staging disk.

Clone VDisks: Clone disks are full copies of an application’s production data. Clone disks are not automatically expired.

Protecting Data

IBM InfoSphere VDP protects copy data by presenting a staging disk to the host. This staging disk maintains a golden copy of the application data that is protected using IBM InfoSphere VDP snapshots. On Windows, application-consistent backups are made via VSS. Oracle backups on all platforms are application consistent via RMAN interfaces. Whole VM backups are application-consistent if they are configured with vmtools.

Whenever possible, change block tracking is used to minimize backup data movement. Tracking is accomplished with VSS snapshots, Oracle RMAN, and the VDP Connector.

This section describes:

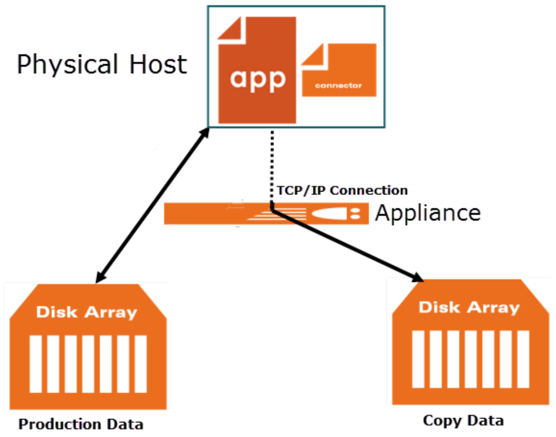
Protecting Physical Hosts on page 10

Protecting Entire VMware and Hyper-V VMs on page 11

Protecting Individual Applications on a VMware or Hyper-V VM on page 11

Protecting Physical Hosts

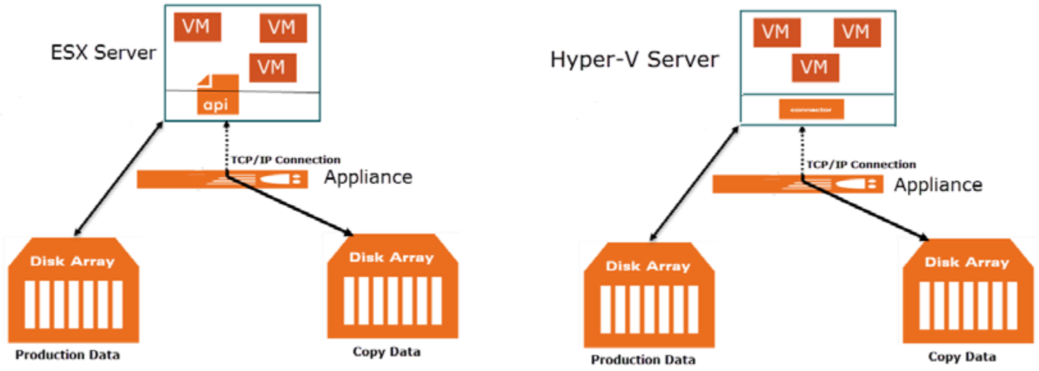
Applications on physical hosts are backed up with the VDP Connector by reading data from the production data and sending this data to the snapshot pool or directly to the dedup pool. Most Windows backups are made in an application-consistent way via VSS, and database backups are application-consistent through database-specific APIs. Change block tracking is typically used to accelerate the process. Applications on non-Windows hosts are low-splash.



Protecting Applications on Physical Hosts

Protecting Entire VMware and Hyper-V VMs

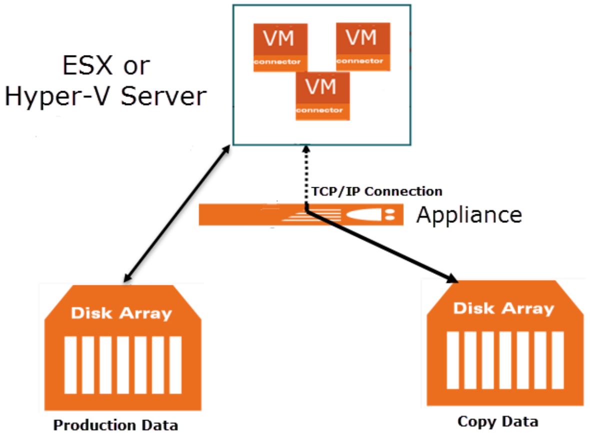
An InfoSphere VDP Appliance can protect entire VMware and Hyper-V VMs Out-of-Band. To protect entire VMware VMs, the InfoSphere VDP Appliance takes advantage of VMware APIs. To protect entire Hyper-V VMs, the VDP Connector is installed on the Hyper-V server.



Protecting Entire VMs

Protecting Individual Applications on a VMware or Hyper-V VM

The VDP Connector is used to protect individual applications on a VM. Once the VDP Connector is installed on a VM, you can create policies to protect individual applications and application groups on the VM.



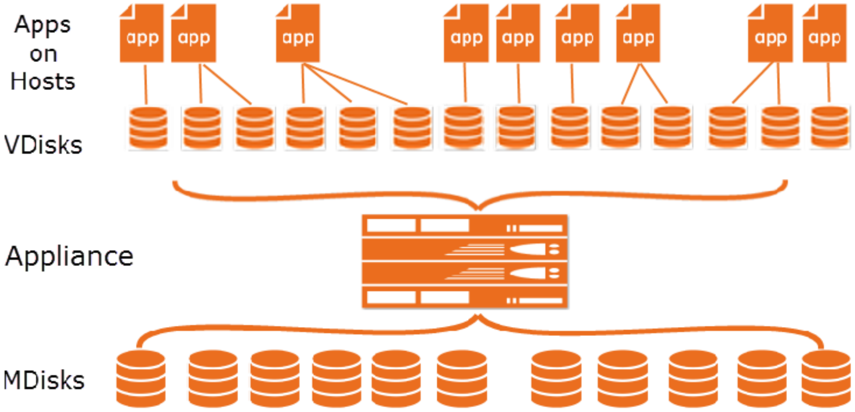
Protecting Applications on a VM

VDisks

InfoSphere VDP Appliances use logical VDisks (virtual disks or volumes) to virtualize data from hosts. VDisks are taken from a pool of managed disks (MDisks) presented to an InfoSphere VDP Appliance from one or more arrays.

From the VDisks, the data can be deduplicated, cloned, mounted, and recovered, presented for test and development work, and manipulated in other tasks. VDisks are created as needed on physical disk arrays.

There is a fixed limit of VDisks per InfoSphere VDP Appliance. As you create protection policies, your InfoSphere VDP Appliance will warn you when a configuration may exceed VDisk limits.



Virtualized Applications on Managed Disks in Your Storage

VDisk Consumption

How Many VDisks Do I Have?

An InfoSphere VDP Appliance creates VDisks as needed from pools of MDisks on the physical disk arrays in the InfoSphere VDP Appliance. The applications and hosts never see the MDisks.

The VDisk limit for the VDP Appliance the VDisk limit varies with the installed capacity license (1000, 3000, or 5000 VDisks). If you have enough VDisks for your needs, but they are growing too large for your existing storage, then you must add storage. If you need more VDisks, then you need another InfoSphere VDP Appliance.

How Many VDisks Do I Need?

In general, each protected application or VM requires one or more VDisks for the staging disk plus the same number more VDisks per snapshot. In addition, note these rules:

- VM-level backups with a snapshot SLA consume one VDisk for each virtual disk in the VM.
- File system backups in a Windows environment consume one VDisk for each protected file system.
- File system backups in a Unix environment consume a VDisk for every 833GB protected times 1+(number of retained snapshots). You can adjust the 833GB value by changing Staging Disk Granularity in Application Advanced Settings, see IVGM Online Help.
- Mounts, LiveClones, and Clones of non-VM applications consume VDisks.
- On Linux and Solaris systems, filesystems and Oracle databases consume one VDisk plus another for every additional 2TB data is being protected.
- Exchange and SQL databases consume one VDisk for every volume that hosts the database.
- Each snapshot of a VDisk consumes one VDisk per snapshot per protected disk.
- Snapshots show peak usage, as new snapshots are created before old snapshots are expired.
- Each VDisk protected by Dedup-Async consumes one additional VDisk for the DAR snapshot and two VDisks on the target appliance.
- After failover and syncback, the failback operation cleans out all the syncback and failover VDisks.

VDisks are thin-provisioned, and can grow over time, as explained in Snapshot Pool on page 8.

Dedup Pool

The local Dedup Pool is used to retain deduplicated copies of the snapshot data for quick access, and to facilitate low-bandwidth replication.

The Dedup Pool has a warning threshold at 80% of capacity. ***When this level has been exceeded, no new dedup related jobs can be run.*** The InfoSphere VDP Appliance sends a warning when dedup pool consumption exceeds the warning level. A remote appliance will reject inbound replication jobs if its own Dedup Pool is above this threshold.

Under normal usage, deduplicated images accumulate over time, and are expired after a set retention period, having been replaced by newer images. The Garbage Collection (GC) feature identifies and reclaims the space used by the expired images. But over time, you may be capturing more or larger images, and the pool can fill up. Ultimately you may need to add more capacity, but before that becomes necessary, review this section carefully to be sure you are following best practices.

Note: *Keep space for Dedup Pools on storage separate from space for Snapshot Pools. Dedup operations are I/O intensive. Keeping Dedup Pools on separate storage ensures that dedup operations will not impede the performance of the Snapshot Pools.*

Understanding Dedup Pool Consumption

The Dedup Pool becomes filled with both current and expired images. GC reclaims space by sweeping away the expired images. An expired snapshot image requires as much space as it did when it was current. In general terms, expired images accumulate in the Dedup Pool at a rate determined by:

Dedup Pool Consumption = Frequency of Snapshots x Snapshot Size

Snapshot Size = Proportion of Changed Blocks (change rate) x Application Size

A 10TB application with a 50% change rate actually generates snapshot images that are 5x bigger than a 100TB application with just a 1% change rate. But if it is protected 5x as often, then they consume the same amount of the Dedup Pool.

Example of Dedup Pool Consumption

Application RPO	Snapshot Size	SLA	Dedup Pool Consumption
Small, Dynamic, Non-Critical Application	50% x 10TB = 5TB	Weekly, Friday, 6pm	5TB/week
Large, Slow-Changing, Critical Application	1% x 100TB = 1TB	Monday, Tuesday, Wednesday, Thursday, and Friday, 6pm	5TB/week

Note: *This is an oversimplification of a complex subject, but the principles illustrated here can help you to understand how to keep your Dedup Pool at top performance.*

Image Retention and Garbage Collection

Those images add up. An application that contributes 30GB per day to the Dedup Pool produces after 30 days close to 1TB of net new data. If these images are retained for 12 months, then after that time it has consumed 12TB of Dedup Pool space. Long retention times can quickly fill the Dedup Pool.

Image retention is how long the image is retained before it is expired. Expired images still consume Dedup Pool, so if those images are retained for only one month, but GC is never run, then the expired images still accumulate.

To reclaim that space, the unneeded images must be expired and then Garbage Collection (GC) must be run. Dedup Pool Consumption grows when the rate of incoming images exceeds the rate at which they are swept out by GC.

To reduce the amount of dedup pool space consumed:

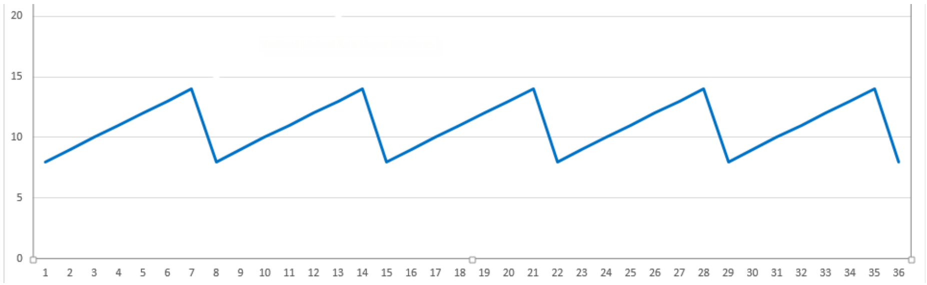
- Retain images no longer than needed, and set GC accordingly.
- Use a mixed retention policy to decrease the number of unnecessary images. For example, retain daily images for one month, weekly images for three months, and monthly images for one year.
- Expiring images or reducing retention rates will not have an immediate affect on the dedup pool. You must run garbage collection and sweep to reclaim the space.
- Run Garbage Collection (GC) and sweep on a regular basis. Once per month is common.

How Dedup Pool Space is Consumed and Reclaimed: Garbage Collection

Space in the Dedup Pool is consumed and reclaimed over time. If the Dedup Pool becomes too full, then there is insufficient space to process new snapshots, and jobs start to fail. Knowing how and when to reclaim space is important for best operation.

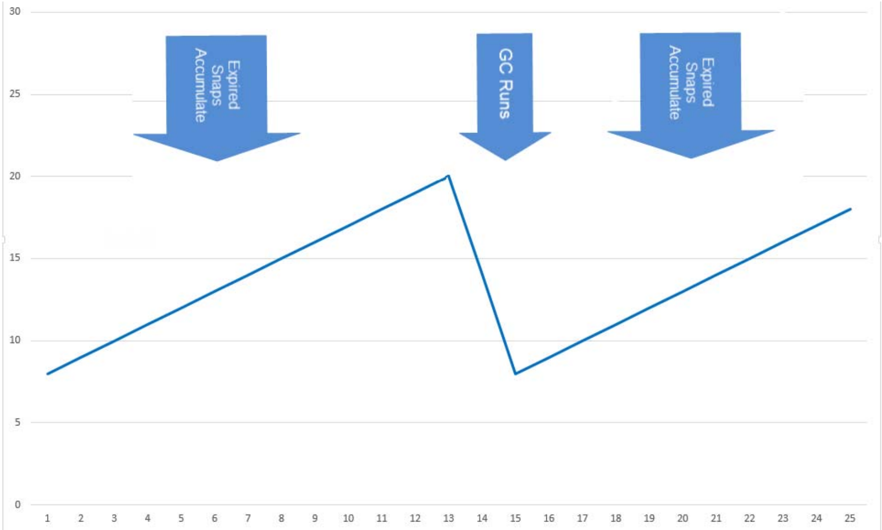
The Simple View

If you had an ideal system with snapshots all the same size being deduplicated and expired at even intervals, you might expect the Dedup Pool consumption over time to look like this:



An Oversimplified and Incorrect View of Dedup Pool Consumption: Snapshots and Expiry

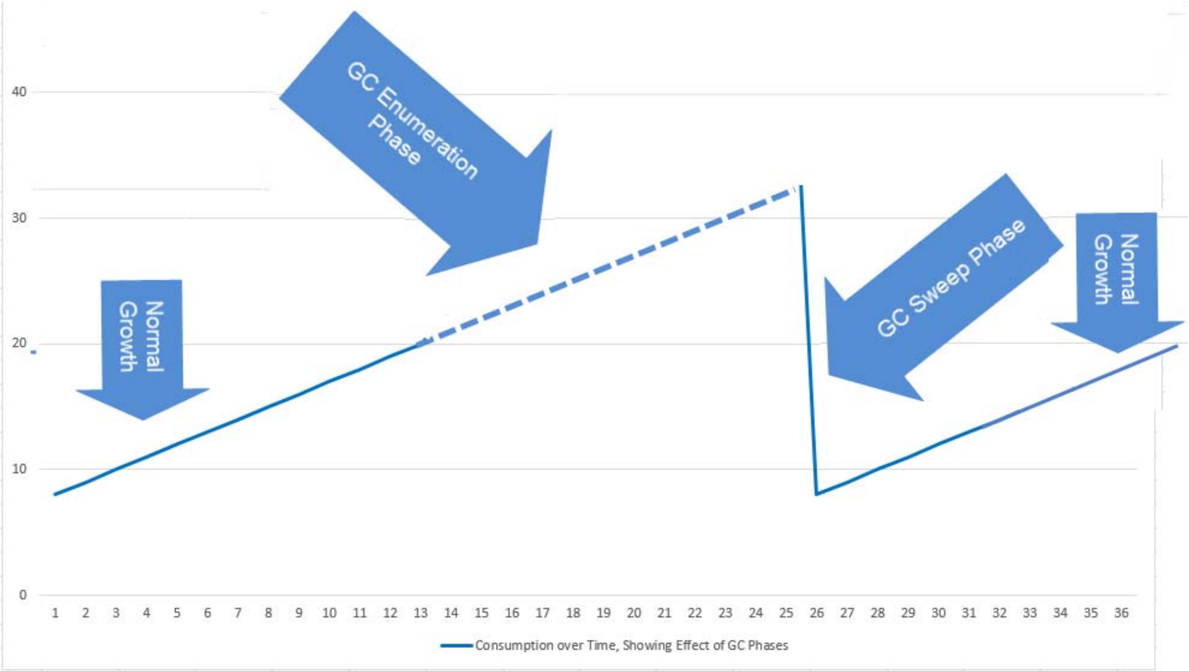
But expiring an image is not enough to reclaim the space. The space is only reclaimed after a Garbage Collection run has enumerated the blocks to keep and then reclaimed the remaining space. So the expired images accumulate until the GC finishes:



A Simplified, More Correct View of Dedup Pool Consumption: GC Reclaims Space

The Detailed View, for When Your Dedup Pool is Filling Up

When GC runs regularly and normally, the image above is enough to understand operations. But the process is a little more complex, and the difference is important when your Dedup Pool is approaching capacity. This is because GC has two phases, the GC Enumeration Phase and the GC Sweep Phase.



GC Increases Dedup Pool Consumption Before Reclaiming Space

If your Dedup Pool is approaching capacity, a GC run is very important, but the GC Enumeration Phase may temporarily increase your pool consumption to the point that jobs start to fail. It is important to monitor Dedup Pool consumption and to run GC monthly during times of low activity. With regard to keeping your Dedup Pool from filling up, you need to know these three things:

1. During the GC Enumeration Phase, Dedup Pool consumption continues to increase.
2. The GC Enumeration Phase can take a long time to run, even days, and it must not be interrupted. The Sweep Phase is usually faster, and you can set it to run at intervals such as overnight.
3. If necessary, the GC Enumeration Phase and the GC Sweep Phase can be run at different times.

The Impact of Garbage Collection on Dedup Performance

GC is resource-intensive and time-consuming, sometimes taking days to complete. Initial ingests of new applications cannot run while GC is running.

GC consumes system resources. All jobs except initial ingests continue to run as scheduled, but some of them take longer to complete due to resource competition from GC. A long-running GC can result in a backlog of jobs and even SLA violations from jobs that do not complete within their scheduled time.

GC runs in the Dedup Pool, so Dedup, Direct-to-Dedup, and Remote Dedup jobs are heavily impacted. It is a good idea to reschedule these if a GC is planned, or is running overlong. Snapshot (protection) jobs are unaffected.

A good rule of thumb is to identify the applications that require the most Dedup Pool consumption (see Understanding Dedup Pool Consumption on page 13), and then set GC to run at four times the retention of those images; for example if those applications are typically retained for one week, then run GC monthly.

3 Service Level Agreements

This chapter introduces the concepts of how IBM InfoSphere VDP captures, manages, and accesses data. Understanding these concepts will help you to be successful with IBM InfoSphere copy data management.

Policy Templates and Policies

A Policy Template is a collection of policies. A policy defines:

- The source of the data managed by the policy
- Type of the protection operation
- Frequency of the protection operation
- How long to retain the data
- Whether data is replicated

Multiple policies within a template allow you to create a single template that defines the short term and long term retention of data as well as whether data will be replicated and how long replicated data will be retained.
Policy Templates can be made up of one or more of the following policies:

Production to Snapshot defines when and how often production data will be captured and how many snapshots are retained. Data recovery from the Snapshot Pool is fast because images are not deduplicated and are stored in the local InfoSphere VDP Appliance's Snapshot Pool. Snapshots are meant for short term retention. See Production to Snapshot Policies on page 31 for details.

Snapshot to Dedup Backup defines when to deduplicate snapshot data and how long to retain the deduplicated data. Data in the Dedup Backup Pool is meant for longer term retention. See Snapshot to Dedup Policies on page 31 for details.

Production to Mirror defines how data will be replicated to a Mirror Pool (a Snapshot Pool on a remote InfoSphere VDP Appliance). Data in the Mirror Pool is meant for instant recovery in a disaster recovery scenario. See Snapshot to Dedup Policies on page 31 for details.

When creating a Production to Mirror Policy you have the choice of replicating data via:

- StreamSnap: Used in environments that have high bandwidth networks
- Dedup-Async Replication (DAR): Used in environments where network bandwidth is constrained

For details on replicating data, see Replicating Data on page 25.

Dedup to Remote Dedup defines when to replicate deduplicated data to a remote InfoSphere VDP Appliance's Local Dedup Pool and how long to retain the data in that pool. Data in the remote dedup pool is meant for retention of data in case of a disaster at the local appliance's site. See Dedup to Remote Dedup Policies on page 32 for details.

Production Direct-to-Dedup defines when to deduplicate VMware VMs directly from production data and how long to retain the deduplicated data. Capturing VMware VMs directly to a Dedup Backup Pool is meant for long term retention when instant access from a Snapshot Pool is not required. See Production Direct to Dedup Policies on page 35 for details.

Snapshot to OnVault defines when to send Production to Snapshot data to the storage defined by an IBM InfoSphere OnVault Pool and how long to retain the data. Snapshot to OnVault Policies are meant for long-term retention of data. See OnVault Policies on page 32 for details.

Policy Templates are:

- Created in the Service Level Architect (SLA) service.
- Applied to applications in the Application Manager service.

Resource Profiles

Resource profiles define where application data is retained. They define which pool to use: Snapshot, Dedup, or OnVault. Pools specified in Resource profiles are used along with policy templates to form an SLA for an application. Resource Profiles are:

- Created in the Service Level Architect (SLA) service.
- Applied to applications in the Application Manager service.

Managed Data License

Managed Data License (MDL) is a measure of IBM InfoSphere-protected copy data. The measure of protected data varies by application. For example:

- File Systems: MDL equals the total size of protected files.
- Oracle databases: the total size of protected database files.
- SQL Server, Exchange, and SharePoint: the total size of protected database files, including any log files.

  MDL calculations are detailed in ***IBM InfoSphere Managed Data Licensing***.

4 Data Capture and Replication

This includes:

Capture Mechanisms

An InfoSphere VDP Appliance captures data by making an initial full copy of the data, then making copies of incremental changes. This capability requires the ability to track and capture the changes that occur between capture operations. To track those changes the InfoSphere VDP Appliance uses either The VDP Connector or VMware API Calls.

The VDP Connector

The VDP Connector is used to capture selected applications and for capturing entire Hyper-V VMs. The VDP Connector is a small-footprint, operating system specific, lightweight service that can be installed on either virtual or physical servers.

The VDP Connector provides a more granular data capture capability than what is provided by VMware API calls. It allows you to: Capture selected applications, capture applications that cannot be snapped by VMware, capture Microsoft SQL Server clusters, and Microsoft Exchange Database Availability Groups (DAGs).

Specifically, VDP Connectors:

- Discover applications
- Quiesce applications, for application consistency during capture
- Enables change block tracking for IBM InfoSphere’s incremental forever capture strategy
- Capture and manage transaction logs:
    - Capture database(s) and logs with one policy template
    - Truncate database transaction logs
    - Roll database transaction logs forward for point-in-time recovery when accessing virtual copies
- If multiple applications are resident on a server, a single policy template can be applied to multiple applications.
- Avoids VMware VMs “stun” issues

For Hyper-V Servers, the VDP Connector also enables the capture of entire Hyper-V VMs.

The VDP Connector also enables scripting on the hosts on which it is installed. The VDP Connector (host side) scripts can be invoked for:

- On-demand jobs triggered from the IBM InfoSphere CLI with the -scripts argument.
- Pre- and post-phases of an InfoSphere VDP Appliance Workflow job.

For detailed step-by-step instructions on how use InfoSphere VDP Appliance scripting, see ***Connecting Hosts to InfoSphere VDP Appliances***.

VMware API Calls

An InfoSphere VDP Appliance can take advantage of VMware APIs for data protection (VADP) calls to capture an entire virtual server. Specifically, the API calls:

- Enable change block tracking for IBM InfoSphere's incremental forever capture strategy.
- Quiesce applications for application consistency during capture.

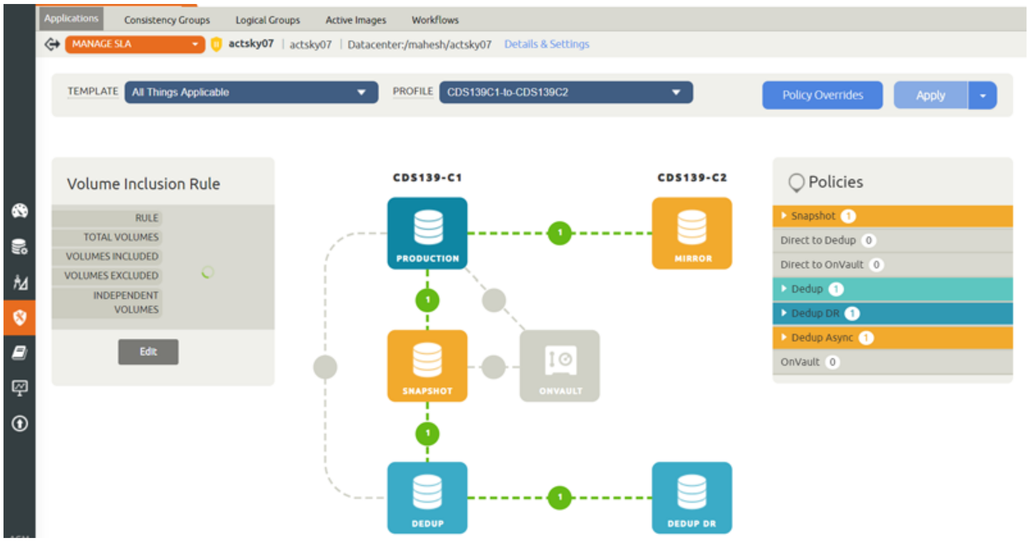
When an entire virtual server is captured, a fully functional virtual server (operating system, applications, and their data) is captured. Having a copy of the entire virtual server guarantees that the data can be accessed fast and without issues. Because the image presented is a fully functional virtual server, if needed, it can be started and run from an InfoSphere VDP Appliance directly and then optionally migrated to a new, permanent location.

Virtual machines and their applications can be grouped and captured with a single Policy Template.

The Stages in Virtualizing an Application or a VM

When you first virtualize an application or a VM, you assign an SLA, and the SLA runs on schedule. Then:

1. The application or VM is running in local Production.
2. According to the SLA settings, the InfoSphere VDP Appliance takes a snapshot image of the production application. The InfoSphere VDP Appliance immediately stores the snapshot image in the Snapshot Pool.
3. Either immediately or at a later time, the InfoSphere VDP Appliance copies the image from the Snapshot Pool to the Dedup Pool or to a Mirror Location or to an OnVault.
4. If the InfoSphere VDP Appliance is joined to other InfoSphere VDP Appliances or to a cloud instance, then the image can be replicated to another appliance’s Remote Dedup DR Pool or to an IBM InfoSphere OnVault for long-term storage.



TEMPLATE All Things Applicable

PROFILE CDS139C1-to-CDS139C2

Policies	
Snapshot	1
Direct to Dedup	0
Direct to OnVault	0
Dedup	1
Dedup DR	1
Dedup Async	1
OnVault	0

Protecting an Application

When Application Protection Takes Effect

Applying an SLA does not immediately protect an application. Protection jobs run on a schedule, according to resource availability. You can also run the job immediately.

- The SLA includes a schedule of when to run the protection job for this application, such as daily between 6 PM and 6 AM, every four hours. If you apply protection to an application at 1 PM today, then the first protection operation will be scheduled for 6 PM today.
- At the scheduled time, the job is assigned a **job slot**. Job slots may be available when the job is scheduled, but some factors can complicate the picture. Job slots are detailed in About Job Slots on page 40.

Changing Protection

You can change an application’s protection at any time. Future backups will occur based on the new template. Existing backups will be retained according to the old template that was in force when they were created.

The Change Tracking Driver

The VDP Connector with its change tracking driver (sometimes called the filter driver) enables efficient incremental backups by tracking changes from the host side. After the first complete backup of a database, the InfoSphere VDP Appliance performs incremental backups by default. If your backups are still always full backups, then check for the following:

- The change tracking driver service is stopped. In this case, restart the change tracking driver service.
- The change tracking driver is incorrectly configured or not installed. In this case, uninstall and then make a full install of the VDP Connector.

Validating Projected Resources Prior to Applying Protection

To avoid impacting the InfoSphere VDP Appliance's performance when applying an SLA policy template to protect applications and VMs, a warning screen informs you of a potential impact to system resources based on the policy settings.

Note: *The calculation of VDisk usage and performance pool usage is directly related to the number of snapshot copies during steady state. The number of snapshot copies during steady state is related to the Recovery Point Objective (RPO) and retention. For example, an RPO of 8 hours and a retention of 3 days means that there will be a total of 9 snapshot copies. This total is subject to the number of days the policy is in effect and the time range defined within the day.*

The Warning screen displays the following information related to applying the configured SLA policy for application protection. It displays projected resource usage with SLA policy changes for the selected applications as well as across the system.

For Selected Applications

VDisks - The expected VDisks usage by the InfoSphere VDP Appliance to protect the selected applications and VMs before **and** after you apply the selected SLA policy. This system resource number is the number of VDisks consumed per volume, including the staging disks.

Snapshot Pools - The expected usage of the performance pools (staging disks and snapshot disks) by the InfoSphere VDP Appliance to protect the selected applications and VMs before **and** after you apply the selected SLA policy. The Snapshot pool holds "golden copies" of application data at the points in time specified by the SLA. This system resource usage is specified as capacity per 1TB based on retention and average change rates.

System-wide

VDisks - The expected VDisks usage across the entire InfoSphere VDP Appliance before **and** after you apply the selected SLA policy for application protection.

Snapshot Pools - The expected performance pool usage across the entries InfoSphere VDP Appliance before **and** after you apply the selected SLA policy for application protection.

The Warning screen can be indicative of the following system resource issues for the specified SLA policy template:

- The SLA policy has more than 14 snapshots.
- VDisk usage with the SLA policy will result in VDisks usage that exceeds the warning level (default of 90%) during steady state.
- Performance pool (staging disks and snapshot disks) usage with the SLA policy will result in a performance pool that exceeds the warning level (default of 90% for the snapshot and primary pools).
- Average dedup pool utilization (7 days) is already at the warning level (75% by default) and there are additional dedup jobs in the queue. This action has the potential of adding more dedup jobs to an already overloaded dedup system.

You can:

- **Cancel** to adjust the SLA policy template in the SLA Architect.
- **Continue** to accept the policy.

If you decide to makes changes to the policy in the SLA Architect, evaluate the frequency of the backup operation and lifetime of the backed up data to see where adjustments can be made to resolve the warning. See the IVGM online help.

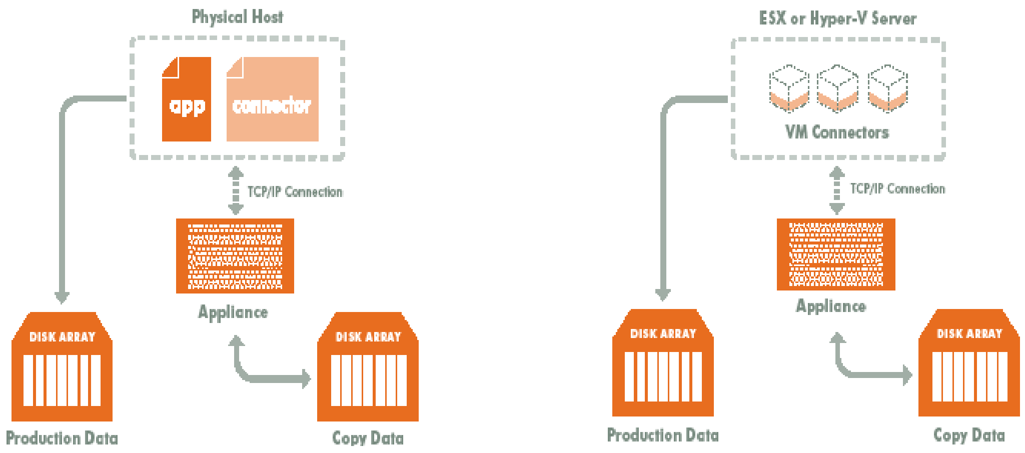
Capture Options

An InfoSphere VDP Appliance allows you to:

- Capture Individual and Groups of Applications on page 23
- Capture Application Data in IBM InfoSphere Consistency Groups on page 23
- Capture a VM's Applications and Boot Volume on page 23
- Capture Entire VMware and Hyper-V VMs on page 24
- Capture Database Logs on page 24

Capture Individual and Groups of Applications

The VDP Connector is used to capture individual and groups of applications on physical and virtual servers.



Managing Individual or Groups of Applications

Installing the VDP Connector on a physical server or VM allows you to create a single Policy Template to capture all applications on the server or several Policy Templates to capture groups of applications.

Capture Application Data in IBM InfoSphere Consistency Groups

A consistency group is enabled by the VDP Connector. As the name implies, consistency groups ensure consistent point-in-time capture and recovery across multiple applications on the same host. To achieve application consistency, members of a consistency group are quiesced and captured together via a single Policy Template.

If IBM InfoSphere's Database Log Backup option (Microsoft SQL Server and Oracle only) is enabled on a Snapshot policy, then all databases captured by the Policy Template in which the Snapshot policy resides can be recovered to the same point-in-time. Recovery and rolling forward of the logs (for databases) in a group is performed via the IBM InfoSphere user interface with a single action.

In addition to making capture and recovery operations easy and fast, consistency groups consume fewer system resources (VDisks).

Capture a VM's Applications and Boot Volume

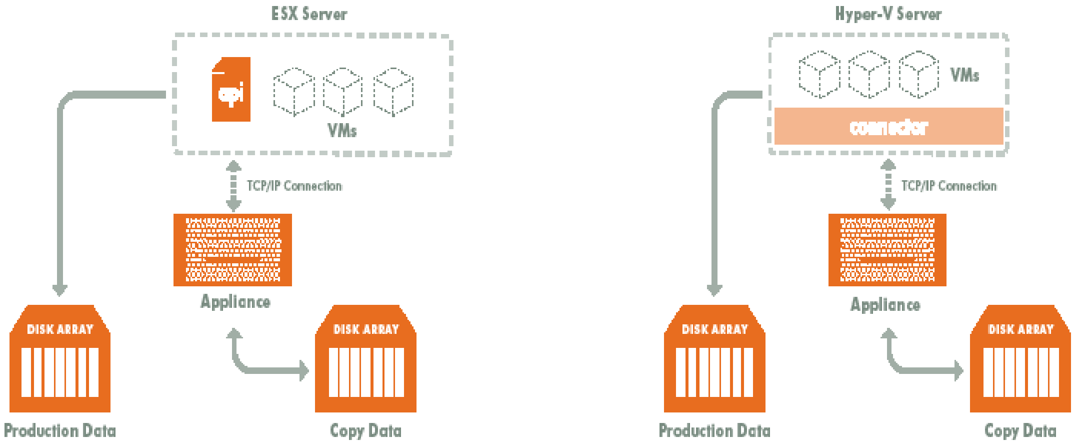
When managing applications on VMs you have the option of also capturing the VM's boot volume. When a VM's boot volume is captured along with its applications, an image can be presented that is a fully functional VM. The image can then be migrated to a new, permanent location if needed.

***Note:** VDP Appliance for AWS cannot capture VM boot volumes.*

Capture Entire VMware and Hyper-V VMs

To capture entire VMware VMs, the InfoSphere VDP Appliance takes advantage of VMware APIs. To capture entire Hyper-V VMs, the InfoSphere VDP Appliance uses an VDP Connector installed on the Hyper-V server.

***Note:** VDP Appliance for AWS cannot capture VMs.*



Managing Entire VMs

***Note:** An InfoSphere VDP Appliance can be on the same ESX server as the VMs it captures.*

When an entire virtual server is captured, a fully functional virtual server (operating system, applications and their data) is captured. Having a copy of the entire virtual server guarantees that the data can be accessed fast and without issues. Because the image presented is a fully functional virtual server, it can be migrated to a new, permanent location if needed. Capturing whole virtual servers allows groups of virtual servers and their applications to be protected with a single Policy Template.

Capture Database Logs

Database log capture is enabled in a Snapshot policy's Advanced Options. It enables a single Snapshot policy to capture logs for Microsoft SQL Server databases, Oracle databases, and consistency groups that contain Microsoft SQL Server databases or Oracle databases. The frequency at which database logs are captured is defined separately from that of the database. For example, a database can be captured every day and its logs captured every hour.

The frequency of database log backup is set in minutes, and the frequency at which logs are captured must not exceed the frequency at which its associated database is captured. For example, if a database capture frequency is every 24 hours, the log file capture frequency must be less than every 24 hours.

Log retention is defined separately from the retention of the Snapshot policy. Having a separate retention period allows you to use logs in conjunction with copies of the database stored in both the Snapshot and Dedup pools. For example, if a database's Snapshot data is kept for three days and its Dedup data kept for seven days, you can define log retention to span all seven days. In this example, a single captured database image can be selected and its logs can be rolled forward over the seven day period.

Database logs are not deduplicated, and regardless of how many logs are captured during a specified log retention period, a database's captured logs are staged to a single VDisk in the IBM InfoSphere Snapshot pool. To conserve space in the Snapshot pool, you can use an advanced setting to instruct the database to compress its logs.

Capture System State

IBM InfoSphere by default will capture the system state when capturing data from Windows/Linux based on-premises or cloud based virtual machines and physical machines. This allows for easy conversion of the captured machine into a cloud native machine across AWS, GCP and Azure. This functionality is central to IBM InfoSphere's cloud mobility suite of features that provides you the flexibility to decouple your physical and virtual machines from infrastructure and open the doors for mobility to the cloud, across clouds, and back to the data center.

Replicating Data

Replication of copy data to remote storage protects the data in the event of disaster at the primary site and reduces the amount of storage required at the primary site. The goal of replication is to get your data back in situations of data loss and impact to your production systems due to issues such as a hardware failure, software issues, or a site event. Data replication also supports the creation of remote copies of Test/Dev, QA, and Analytics data. Data can be replicated from one InfoSphere VDP Appliance to a second (remote) appliance or to the cloud for recovery, disaster recovery, or test/development purposes.

Your SLA templates determine the method, schedule, and frequency of how data replication to a remote site is to be performed. The SLA template defines how to move and store data efficiently to the remote InfoSphere VDP Appliance. Data replication is controlled by the individual template policies:

- Dedup Backup policies use an IBM InfoSphere proprietary replication engine to replicate data to a second InfoSphere VDP Appliance. Dedup Backup is efficient for the long-term storage (weeks to months) of captured and deduplicated data to a remote InfoSphere VDP Appliance. Dedup Backup is intended to retain data for a medium to long-term retention period (for example, 3 months to a year). In addition, Second Hop Replication policies allow you to replicate dedup backup data to a second location operating as the second-leg in a multi-hop configuration of joined InfoSphere VDP Appliances. For details see Dedup Backup to Dedup DR Replication on page 26.
- Production to Mirror policies protect your application or VM data against a site failure by having a full copy of that data mirrored to a remote production site. Applications are kept up-to-date and can be re-started at a moment’s notice at the remote site by accessing data from the remote DR copy. Data mirroring can be considered as access optimized replication to a remote site. For details see Production to Mirror Policy Replication on page 26.
- Snapshot to OnVault policies use an HTTPS connection to send data to storage defined by an IBM InfoSphere OnVault Pool. Data sent to storage defined by an IBM InfoSphere OnVault Pool is not deduplicated. The compress option is on by default in IBM InfoSphere OnVault Pools. For details see Sending Snapshots to an OnVault Pool’s Defined Storage on page 26.

Dedup Backup to Dedup DR Replication

Dedup Backup replication is efficient for long-term storage of captured and deduplicated data to a remote InfoSphere VDP Appliance. Data replicated using a Dedup Backup policy is transmitted from the local InfoSphere VDP Appliance dedup pool to the dedup pool managed by another InfoSphere VDP Appliance. The need for and number of Dedup DR copies to retain on a second InfoSphere VDP Appliance for long-term data recovery (LTDR) is driven by offsite retention requirements for the data.

Dedup Backup uses a Dedup Backup to Dedup DR policy. Dedup Backup replication is incremental, globally deduplicated, and compressed and encrypted in flight. The Dedup Backup replication process begins after the deduplication process completes. A proprietary deduplication-aware replication protocol enables the transmission of only the globally unique blocks, which minimizes the bandwidth required to move data between InfoSphere VDP Appliances.

Blocks are compressed and encrypted in flight for the most efficient bandwidth utilization. Deduplication and compression optimize the data set for transport between sites, eliminating the cost of WAN optimization.

Dedup Backup replication also provides an added benefit of allowing data to be replicated to a remote site, and then from that remote site to a second remote site. This feature is referred to as multi-hop. You can use this process to perform on-demand replication of a backup dedup image to a remote InfoSphere VDP Appliance that is part of a multi-hop configuration of joined InfoSphere VDP Appliances.

Sending Snapshots to an OnVault Pool’s Defined Storage

The Snapshot to OnVault policy allows you to send snapshot data to a location defined by an IBM InfoSphere OnVault Pool. A schedule within the policy is used to send the most recent snapshot taken by the Policy Template’s Production to Snapshot policy to the location defined by the IBM InfoSphere OnVault Pool. IBM InfoSphere OnVault Pool storage is typically used for long-term retention. For details on the OnVault Pool, see OnVault Pool on page 7.

When sending data to a storage defined by an IBM InfoSphere OnVault Pool, an HTTPS connection is used to ensure data security over the network. The OnVault Pool’s compression option is on by default to minimize network traffic.

Data sent to the IBM InfoSphere OnVault Pool storage is not deduplicated. However, after the initial ingest of the full snapshot, only the changes to data are sent to the OnVault Pool. This is the same incremental forever model used by other IBM InfoSphere policies.

When accessing data in an IBM InfoSphere OnVault Pool’s storage:

- InfoSphere VDP Appliances can create clones.
- InfoSphere VDP Appliances can mount data, but because data will first be copied to the snapshot pool then mounted, it is not recommended.
- LiveClones cannot be created.

Production to Mirror Policy Replication

Production to Mirror policies provide a means to replicate a copy of the application or VM data to a target InfoSphere VDP Appliance and to have data access without a restore window, providing for very low RTO. As needed, you have the ability to perform a failback to the production site with an identical set of data that is mirrored between the local and remote InfoSphere VDP Appliances.

StreamSnap

StreamSnap facilitates high-availability by allowing you to keep a remote copy of an application's storage and configuration up-to-date and ready for a failover scenario. When a StreamSnap-managed application fails, you mount a failover image of the application from the remote site. When the problem has been resolved, then you can restore the syncback image to the local site with the latest changes and then failback the application to the production site.

StreamSnap replicates data snapshots that are not deduplicated to a remote InfoSphere VDP Appliance over a high quality bandwidth IP network, which can provide RPOs as low as one hour.

- For VMware VMs, snapshot replication is streamed to the second InfoSphere VDP Appliance in parallel. Streaming of a VMware VM is performed to avoid waiting until the local snapshot job completes before initiating replication.
- For non-VMware VM applications, snapshot replication occurs after the local snapshot job is completed.

Note: *StreamSnap replication and local snapshots are integrated to avoid the creation of double snapshots. The InfoSphere VDP Appliance allows you to maintain multiple local snapshots and store local images in the Dedup pool for long-term retention.*

Production to Mirror policies that use StreamSnap replication are tied to a specific Production to Snapshot policy. They use the schedule and frequency settings of their associated Production to Snapshot policy.

You can retain snapshot images from multiple available points in time at the remote site by applying retention in a StreamSnap policy. When retaining snapshot images at the remote InfoSphere VDP Appliance, a new snapshot image will be created at the remote appliance with an expiration date determined by the policy settings. Each remote snapshot image supports all operations available with a local snapshot image when accessed from the Application Manager.

StreamSnap replication requires a reliable network connection to replicate data snapshots to the remote InfoSphere VDP Appliance. The bandwidth required on the network connection is directly related to the application size (initial copy) and amount of change (for incremental updates).

Dedup-Async (DAR)

Dedup-Async Replication (DAR) allows you to keep a remote copy of an application's data up-to-date and ready to be used in a failover scenario, facilitating high-availability and redundancy. When a DAR-managed application fails for any reason, you can mount a failover image of the application at the remote site. When the problem has been resolved, then you can incrementally reverse replicate the changes made at the DR site to the primary site and then failback the application to the production site

DAR is an IBM InfoSphere-proprietary form of replication where initially a full copy of the application data is replicated to a target Snapshot pool on the second InfoSphere VDP Appliance or when a VM is replicated to the datastores of an ESX server. Dedup-Async replication sends deduplicated and compressed data over the network at a fraction of the bandwidth required for traditional replication technologies.

Once the Dedup-Async job takes a snapshot, it deduplicates the data, replicates the deduplicated data to another InfoSphere VDP Appliance, rehydrates that data on the second InfoSphere VDP Appliance, and updates the full copy of data on the second InfoSphere VDP Appliance to provide the flexibility of instant access at the remote location. This ensures that a full, up-to-date copy of data is ready and available on the second IBM InfoSphere site.

Because the data is deduplicated before it is replicated, DAR requires less network bandwidth than StreamSnap replication but it does require additional IBM InfoSphere system resources.

Note: *Production to Mirror policies that use DAR make snapshots of their own. They do not use a snapshot created by another Production to Snapshot policy.*

5 Capturing Applications Overview

This chapter presents high-level descriptions of the processes used to capture an application:

Before You Begin

Add hosts that host applications using the Domain Manager service. For detailed, application-specific instructions see ***Connecting Hosts to InfoSphere VDP Appliances***. Detailed, application-specific instructions on capturing applications and VMs are in the IVGM online help.

Adding Hosts

The first step in capturing an application is to add the host on which the application(s) reside.

From this page you can add servers that host applications and hypervisors that host VMs. Individual VMs are discovered via the Application Manager service.

Discovering Applications and VMs

After a host is added, use the Application Manager service to discover applications on physical servers, VMs on hypervisors, and applications on VMs. You will be prompted to select which of the discovered hosts or hypervisors you want to discover applications or VMs.

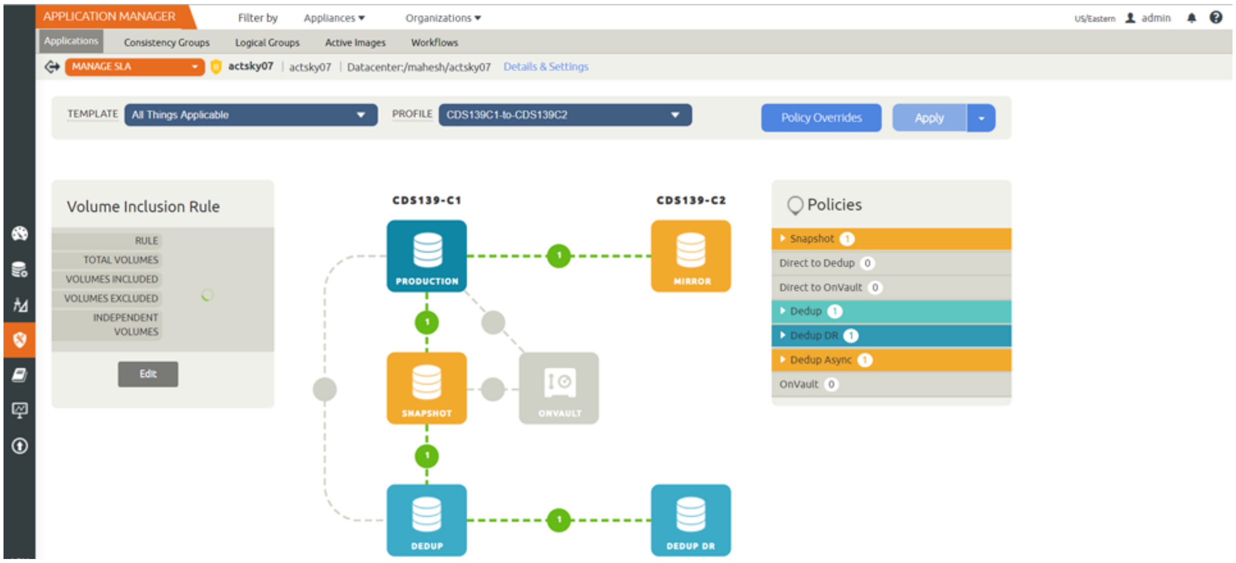
Creating Policy Templates

Policy templates are made up of one or more policies. Policy templates provide a high-level wrapper for an end-to-end definition of capturing application data. For example, if you need to capture an image as a local snapshot and replicate that image off to another InfoSphere VDP Appliance, the policy template will contain both the local snapshot and remote off site policies. Once the policy template is created, create the individual policies that comprise the SLA template.

Creating Policies

Policies define how often to capture an application, how long to retain the captured application and when applicable, where and how to replicate the captured application's data.

The green arrows in the SLA Architect represent the policies within a template that control data flow to the various pools.



TEMPLATE All Things Applicable

PROFILE CDS139C1-to-CDS139C2

Green Arrows Represent Policies In a Policy Template

A Policy allows you define whether its schedule will run:

- **Within a Window**: A period of time in which jobs are allowed to start.
- **Continuous**: Defines when its first job can start but as the name implies, allows subsequent jobs to run at a frequency without regard to any time boundary.

Where applicable, SLA Template Policies allow you to define the rules for determining whether or not a data protected by a policy meets your requirements. The InfoSphere VDP Appliance automatically calculates and sets default SLA Compliance settings. Default settings are based on whether the policy is set to windowed or continuous, the policy type, and IBM InfoSphere-recommended best practices. The default settings calculated will meet the needs of most users.

If data is being protected according to your needs, then it is considered to be in compliance.

The SLA Templates are made up of the following types of policies:

Production to Snapshot Policies on page 31

Snapshot to Dedup Policies on page 31

Dedup to Remote Dedup Policies on page 32

OnVault Policies on page 32

Production to Mirror Policies on page 33

Production Direct to Dedup Policies on page 35

Production to Snapshot Policies

Recent data is accessed the most frequently. Because of this, you only need to retain two or three snapshots of data. Older, less frequently accessed data can be rehydrated and accessed from the dedup pool. You can schedule a Snapshot policy schedule that occurs during a specific frequency and time window or on a continuous basis. The minimum recommended frequency for a Snapshot policy is 1 hour (local RPO).

Note: *When creating a snapshot policy for a database you have the option of also capturing its log files at a specified frequency. See Database Log Protection in an SLA Policy for background information.*

Note: *Snapshot to Dedup and Dedup to Remote Dedup policies associated with a snapshot policy should be configured with the same window start time as the snapshot policy.*

Snapshot to Dedup Policies

A Snapshot to Dedup policy schedule must be coordinated with its associated Production to Snapshot policy schedule. A dedup policy will not run if there is not a snapshot to deduplicate. For example, a policy template with a Snapshot policy that runs once a week and a dedup policy that runs three times a day would have twenty deduplication jobs per week with nothing to deduplicate.

The best practice for a Snapshot to Dedup Backup policy recommends deduplicating one snapshot per day. If you are taking multiple snapshots per day, when the Snapshot to Dedup Backup policy runs it will deduplicate the most recently completed snapshot. The other snapshots will expire before they are deduplicated.

In addition, if you set the window for a dedup policy too narrow, a snapshot could complete outside of the dedup policy's window. In most cases it is advisable to set a dedup policy's window from 19:00 to 18:59. This ensures that the dedup policy will run, even for those rare occasions when the snapshot finishes later than usual.

Note: *Snapshot to Dedup and Dedup to Dedup DR policies associated with a Production to Snapshot policy should be configured with the same window start time as the Production to Snapshot policy.*

For example, to create a Snapshot to Dedup Backup policy schedule that runs as soon as there is something to deduplicate, set:

- Schedule type of Windowed (default)
- Dedup on these days: Everyday except Never
- The window to open and close as needed. Typically set to 19:00 to 18:59
- The frequency to Once Per Window
- The desired retention time (for example, retain for 14 days)

Note: *With this configuration, when multiple Production to Snapshot images are available, only the most recently completed snapshot will be deduplicated. This ensures that regardless of what time a snapshot job finishes, the dedup job will start in the next available scheduler slot.*

Dedup to Remote Dedup Policies

Remote Dedup replication uses a Dedup to Remote Dedup policy. Remote Dedup replication is efficient for long-term storage of captured and deduplicated data to a remote InfoSphere VDP Appliance. Data replicated using a Dedup to Remote Dedup policy is transmitted from the local InfoSphere VDP Appliance dedup pool to the dedup pool managed by another InfoSphere VDP Appliance. The minimum recommended frequency for a Dedup to Remote Dedup policy is 4 hours, with a recommended best practice of 24 hours (remote RPO).

Remote Dedup replication is incremental, globally deduplicated, and compressed and encrypted in flight. The Remote Dedup replication process begins after the deduplication process completes. A proprietary deduplication-aware replication protocol enables the transmission of only the globally unique blocks, which minimizes the bandwidth required to move data between InfoSphere VDP Appliances.

To create a Dedup to Remote Dedup policy schedule that will replicate a single deduplicated image in a day, and will run as soon as there is something to replicate, set:

- Schedule type of Windowed (default)
- Replicate on these days: Everyday except Never
- The window to open and close as needed. Typically set to 19:00 to 18:50
- The frequency to Once per Window
- The desired retention time (for example, retain for 14 days)

Note: *This ensures that regardless of what time the Snapshot to Dedup policy finishes, the Dedup to Remote Dedup policy will start in the next available scheduler slot.*

OnVault Policies

OnVault policies allow you to send data to object storage (an IBM InfoSphere OnVault Pool). A schedule within the policy is used to send the most recent data to object storage. After the initial ingest of data, an OnVault capture operation follows IBM InfoSphere’s incremental forever data capture process.

When sending data to storage defined by the IBM InfoSphere OnVault Pool, an HTTPS connection is used to ensure data security over the network. The IBM InfoSphere OnVault Pool’s compression option is on by default to minimize network traffic.

Data sent to storage defined by an IBM InfoSphere OnVault Pool is not deduplicated.

When accessing data in an IBM InfoSphere OnVault Pool’s defined storage location:

- All InfoSphere VDP Appliances can create clones.
- All InfoSphere VDP Appliances can mount data, but because data will first be copied to the snapshot pool then mounted, it is not recommended.
- LiveClones cannot be created.

IBM InfoSphere allows you to create a two types of OnVault policies:

- IVGM users can create **Snapshot To** OnVault **Policies** that allow them to capture data in an IBM InfoSphere Snapshot Pool on any InfoSphere VDP Appliance and then protect the data in the snapshot pool to object storage defined by an IBM InfoSphere OnVault Pool.
- IVGM users can create Direct To OnVault Policies that allow them to capture VMs in their production environment and protect them directly to object storage defined by an IBM InfoSphere OnVault Pool.

Snapshot To OnVault Policies

To create a Snapshot to OnVault policy schedule that will, once a day, within a defined window, send the most recent snapshot data to object storage defined by an IBM InfoSphere OnVault Pool, set:

- Vault on these days: Everyday
- The window to open and close as needed. Typically set to 19:00 to 18:50
- The desired retention time (for example, retain for 3 years)

Direct To OnVault Policies

To create a Direct to OnVault policy schedule that will, once a day, within a defined window, send the most recent incremental updates directly to storage defined by an IBM InfoSphere OnVault Pool, set:

- On these days: Everyday
- The window to open and close as needed. Typically set to 19:00 to 18:50
- The desired retention time (for example, retain for 3 years)

Production to Mirror Policies

Best practices for Production to Mirror policies depend on which replication option you plan to use. This section outlines the policy best practices for the StreamSnap and Dedup Async, replication methods offered by the InfoSphere VDP Appliance.

Production to Mirror: StreamSnap Replication Policies

Production to Mirror policies that use StreamSnap replication are tied to a specific snapshot policy. They use the schedule and frequency settings of the associated snapshot policy in the template.

Note: *Before creating a StreamSnap replication policy, you must first create a snapshot policy.*

StreamSnap replicates data snapshots to a remote InfoSphere VDP Appliance without deduplication, over a high quality network, which can provide RPOs as low as one hour.

- For VMware VMs, snapshot replication is streamed to the second InfoSphere VDP Appliance in parallel to the snapshot being copied. Streaming of a VMware VM is performed to avoid waiting until the local snapshot job completes before initiating replication.
- For non-VMware VM applications, snapshot replication occurs after the local snapshot job is completed.

Note: *StreamSnap replication and local snapshots are integrated to avoid the creation of double snapshots. Each InfoSphere VDP Appliance allows you to maintain multiple local snapshots and store local images in the Dedup pool for longterm retention.*

StreamSnap replication:

- Achieves Recovery Point Objectives (RPOs) as short as one hour. The StreamSnap replication policy relies on the associated Production to Snapshot policy for RPO and the other advanced snapshot settings. A StreamSnap policy can point to any Snapshot policy with frequency of 1 hour or longer (remote RPO).
- Uses an existing IP network to replicate data.
- Replicates data that is not conducive to deduplication (for example, data that is compressed or encrypted). Such data includes: images, videos, and encrypted databases.
- Replicates large amounts of data to remote users (for example, test and development environments).
- Retains multiple point-in-time snapshot images at the remote site, with retention behavior being driven by the settings in the StreamSnap policy.
- More efficient when replicating a large single dataset (such as a large database) than deduplication.
- Makes fail-over to a host on the remote site simple.
- Enables incremental reverse replication (syncback) to the local InfoSphere VDP Appliance.
- Compresses and encrypts replicated data to the second InfoSphere VDP Appliance. You can disable compression if the data is already compressed (for example, for images and videos).

Note: *StreamSnap jobs run for non-DB, DB, and DB+Log types. To perform on-demand log replication of the database logs to a remote InfoSphere VDP Appliance, select the database in Application Manager, then select Replicate Logs.*

When you apply the SLA template to an application or VM in the Application Manager, System Monitor will record the results of the StreamSnap job and it will appear as a single job. Once replication is complete, two jobs appear in System Monitor with a Succeeded status; one for the Snapshot job and one for the StreamSnap job (see StreamSnap Job Error Handling). If there is a job failure, either for the StreamSnap job or the Snapshot job, two job entries appear to identify which job was successful.

Switching from DAR to StreamSnap

You can switch from Dedup-Async to StreamSnap without having to re-ingest and transmit a new full image. The existing local and remote snapshots will be used for the StreamSnap replication.

If an application is protected with a Production to Mirror template that uses DAR, you can replace that application’s template with a Production to Mirror template that uses StreamSnap replication.

When you replace the application’s Production to Mirror template, a confirmation message is presented.

As the message states, the action is irreversible, will take time and system resources (depending on the amount of data) and the first backup with the new template will be an incremental backup.

Replacing a Production to Mirror template that uses StreamSnap with one that uses DAR replication is not supported.

Production to Mirror: Dedup-Async Replication (DAR) Policies

Production to Mirror policies that use Dedup Async replication (DAR) take snapshots of their own. They do not use snapshots created by other policies.

Once the Production to Mirror policy takes a snapshot, it deduplicates the data, replicates the deduplicated data to another InfoSphere VDP Appliance, rehydrates that data on the second InfoSphere VDP Appliance, and updates the full copy of data on the second InfoSphere VDP Appliance. This ensures that a full, up-to-date copy of data is ready and available on the second IBM InfoSphere site.

Because the data is deduplicated before it is replicated, Dedup Async replication is optimized to require less network bandwidth than the other replication options (such as StreamSnap replication), but it does require additional IBM InfoSphere system resources to deduplicate data.

Dedup-Async Replication:

- Achieves Recovery Point Objectives (RPOs) of 24 hours with 12 and 8 hour RPOs possible. The minimum recommended frequency for a Dedup Async policy is 4 hours (remote RPO).
- Replicates data that is can be efficiently deduplicated.
- Uses an existing IP network to replicate data.
- Minimizes bandwidth requirements for replication.
- Replicates repeatedly at intervals determined by the Dedup-Async policy.
- Makes disk management transparent.
- Replicates VMware VMs to a datastore (optional).
- Makes fail-over to a host on the remote site simple.
- Makes syncback to the local InfoSphere VDP Appliance simple.

Dedup Async replication allows you to define a frequency (Every) for the Production to Mirror policy. IBM InfoSphere's best practice is to set a frequency of (Every) of 24 hours.

Note: *A Production to Mirror policy job will be queued as soon as it is saved. Before saving a Production to Mirror policy consider its impact other data operations; especially for the initial ingest of data.*

Production Direct to Dedup Policies

Production Direct to Dedup policies can only be applied to VMware VMs. Direct to dedup takes advantage of VMware's change block tracking capabilities and writes deduplicated captured data directly to the IBM InfoSphere dedup pool. Data protected with Direct to Dedup policies do not consume IBM InfoSphere VDisks.

For example to create a Direct to Dedup policy schedule that will capture a single image in a day, set:

- Schedule type of Windowed (default)
- Dedup on these days: Everyday except Never
- The window to open and close as needed. Typically set to 19:00 to 07:00
- The frequency to Once per Window
- The desired retention time (for example, retain for 14 days)

Resource Profiles

A resource profile specifies the storage media for captured application and VM data. The policy and the resource profile that make up the SLA dictate the type of application data capture to perform and where to store the captured application data (which pool of disks can be used). Resource Profiles define which Snapshot Pool (if needed) will be used and/or to which remote InfoSphere VDP Appliance data will be replicated.

In addition to policy templates and policies, you also create resource profiles in the SLA Architect. Resource profiles define where to store data. Data can be stored:

- Local: The InfoSphere VDP Appliance that the resource profile is created for.
- Remote: The InfoSphere VDP Appliance used for remote deduplication or replication. This remote appliance must be an appliance that is already paired to the selected local InfoSphere VDP Appliance.

  Note: *You can configure the Remote field only when one or more remote InfoSphere VDP Appliances are configured on the selected local InfoSphere VDP Appliance.*

- OnVault: Object storage defined by an IBM InfoSphere OnVault storage pool.

Note: *You can use the OnVault Pool option only if the InfoSphere VDP Appliance has defined a OnVault storage pool.*

Resource profiles are applied to applications in the Application Manager and the resource profiles work in tandem with policy templates:

- A policy template that does not include a replication policy must be applied to an application along with a resource profile that only stores data locally.
- A policy template that includes a replication policy must be applied to an application along with a resource profile that stores data either on another InfoSphere VDP Appliance or to object storage defined by an IBM InfoSphere OnVault storage pool.

You define a resource profile for any InfoSphere VDP Appliance that has been added to IVGM.

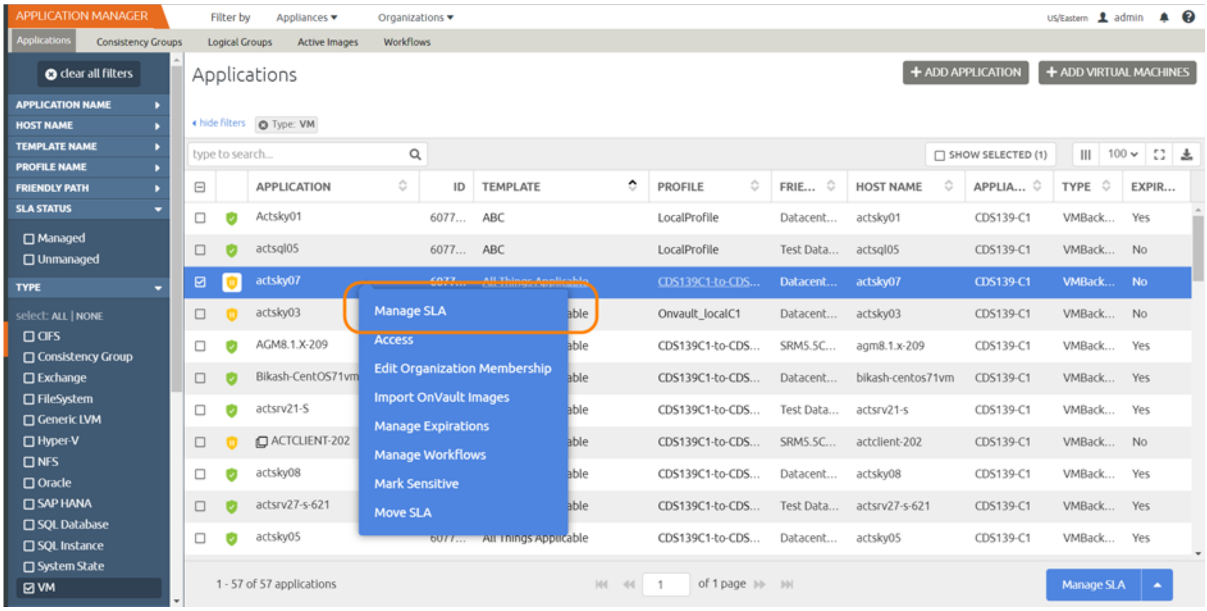
Applying Policy Templates and Resource Profiles

After defining at least one policy template and one resource profile, use the Application Manager service to apply them to an application or VM.

To apply a policy template and a resource profile to an application or a VM:

1. From the IVGM left-hand navigation, click the **Application Manager** icon. The Applications page opens.

2. Right-click the VM or application that you want to manage, and then choose Manage SLA from the drop-down list.



Applications

	APPLICATION	ID	TEMPLATE	PROFILE	FRIE...	HOST NAME	APPLIA...	TYPE	EXPIR...
□	Actsky01	6077...	ABC	LocalProfile	Datacent...	actsky01	CDS139-C1	VMBack...	Yes
□	actsql05	6077...	ABC	LocalProfile	Test Data...	actsql05	CDS139-C1	VMBack...	No
☑	actsky07	6077...	All Things Applicable	CDS139C1-to-CDS...	Datacent...	actsky07	CDS139-C1	VMBack...	No
□	actsky03		able	Onvault_localC1	Datacent...	actsky03	CDS139-C1	VMBack...	No
□	AGM8.1.X-209		able	CDS139C1-to-CDS...	SRM5.5C...	agm8.1.x-209	CDS139-C1	VMBack...	Yes
□	Bikash-CentOS71vm		able	CDS139C1-to-CDS...	Datacent...	bikash-centos71vm	CDS139-C1	VMBack...	Yes
□	actsrv21-S		able	CDS139C1-to-CDS...	Test Data...	actsrv21-s	CDS139-C1	VMBack...	Yes
□	ACTCLIENT-202		able	CDS139C1-to-CDS...	SRM5.5C...	actclient-202	CDS139-C1	VMBack...	No
□	actsky08		able	CDS139C1-to-CDS...	Datacent...	actsky08	CDS139-C1	VMBack...	Yes
□	actsrv27-s-621		able	CDS139C1-to-CDS...	Test Data...	actsrv27-s-621	CDS139-C1	VMBack...	Yes
□	actsky05	6077...	All Things Applicable	CDS139C1-to-CDS...	Datacent...	actsky05	CDS139-C1	VMBack...	Yes

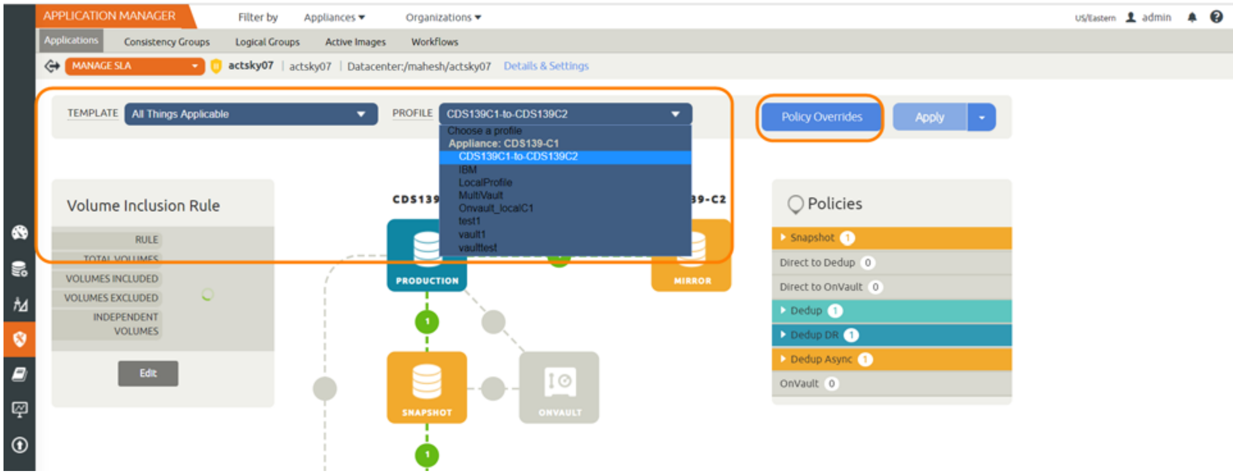
1 - 57 of 57 applications   1 of 1 page

3. The Manage SLA window appears. From the Manage SLA window, choose from the Template and Profile drop-down lists:

   - Template - An existing SLA template that includes policies to define the snapshot/deduplication/replication of the application data.

   - Profile - An existing SLA resource profile that defines the resources used to store the data of the application as snapshot/deduplicated/replicated images.

   The Application Manager service shows a view of the SLA status assigned to the selected applications in the SLA policy map.



actsky07 | actsky07 | Datacenter:/mahesh/actsky07

TEMPLATE: All Things Applicable

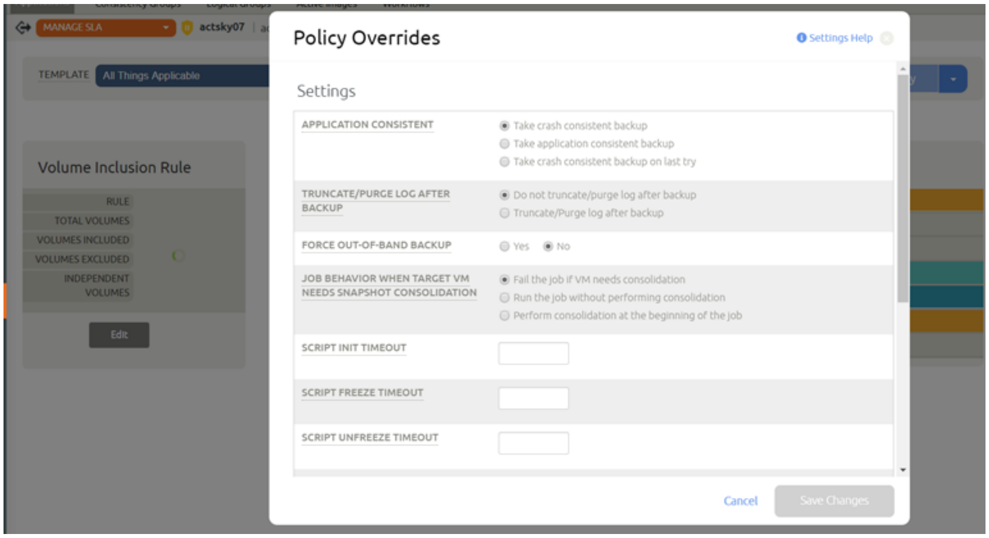
PROFILE: CDS139C1-to-CDS139C2

4. From the SLA Template page, apply **Policy Overrides** as needed before applying the SLA:

   - **Details and Settings**: Apply application-specific settings when managing a VM. Application settings may be useful or required in certain circumstances. See Configuring Application Settings for Protecting VM Data for details.

   - **Policy Overrides**: Override specific policy settings previously configured in the selected SLA template. Advanced settings may be useful or required in certain circumstances. See Configuring Advanced Policy Settings for an SLA Policy for details.

Note: *You can override policy settings in the Application Manager only if the policy template Allow Overrides on Policy Settings parameter has been set to Yes.*



Policy Overrides

Settings

APPLICATION CONSISTENT
● Take crash consistent backup
○ Take application consistent backup
○ Take crash consistent backup on last try

TRUNCATE/PURGE LOG AFTER BACKUP
● Do not truncate/purge log after backup
○ Truncate/Purge log after backup

FORCE OUT-OF-BAND BACKUP
○ Yes ● No

JOB BEHAVIOR WHEN TARGET VM NEEDS SNAPSHOT CONSOLIDATION
● Fail the job if VM needs consolidation
○ Run the job without performing consolidation
○ Perform consolidation at the beginning of the job

SCRIPT INIT TIMEOUT

SCRIPT FREEZE TIMEOUT

SCRIPT UNFREEZE TIMEOUT

5. In the upper right corner, click **Apply** to apply the SLA template and resource profile and the success message box appears.

6. Click **Save Changes** when you complete modifying the settings. The Success message box appears.

   The application is not captured until the scheduled job runs according to the window defined in the SLA template. For example, if at 10:00 am you assign a template that has hours of operation from 2:00 am to 5:00 am, then the first job will not start until the InfoSphere VDP Appliance has an available job slot after 2:00 am.

7. If you want to review the details of the managed application and/or modify any of the application-specific settings, click Details & Settings. The Application Details and Settings dialog box includes:

   - Application-specific information associated with the VM, such as application type, host name, host IP address, path, operating system, InfoSphere VDP Appliance, and appliance IP address.

   - Application-specific settings that you can modify for the VM (see Configuring Application Settings for Protecting VM Data for details).

Note: *To reset one or more application settings back to its default state, click the check box to the left of the selection you want to reset, or click Select options that will revert back to default to reset all application selections back to their default state.*

---

8. After you modify application settings, click **Save Changes**.

9. The job will run when the SLA window opens. To run a job immediately, see Running an On-Demand Capture from IVGM.

Scheduled Jobs

Jobs run according to the schedule assigned in their SLA Template Policies. If you try to run many resource-intensive jobs simultaneously, then some will have to wait for the resources to come available. In very bad situation, they may have to wait so long that an SLA Violation occurs.

It is better to stagger the more resource-intensive jobs like initial snapshots and deduplication jobs over time rather than to have them all compete for resources at the same moment. For example, instead of snapping all VMs, file systems, and databases at 6:00pm on weekdays, consider snapping one type of application on the hour, another type at 10 minutes after the hour, another type at 20 minutes after the hour, and so on.

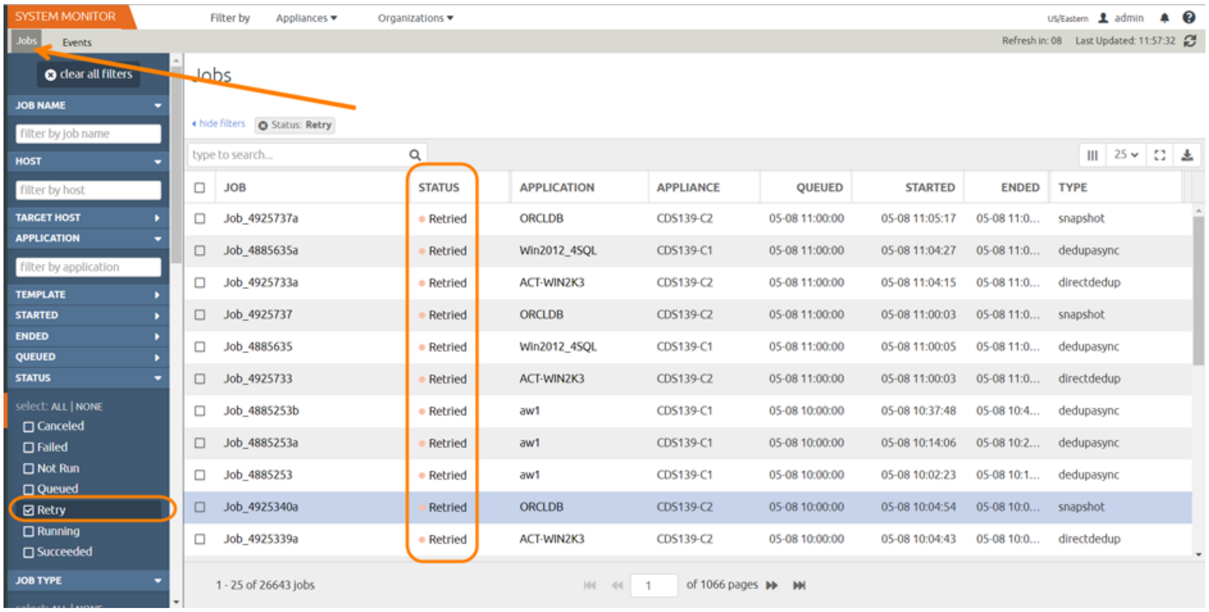
It is not necessary to deduplicate snapshots as soon as they are captured. Once the snapshot is taken, the data is safe in the Snapshot Pool; it can be deduplicated at a slower time such as at night.

The initial snapshot of an application or a VM is the largest and most time-consuming snapshot it will ever get because every bit of data is new. When you add a new application or VM, perform an on-demand snapshot at an off-peak time for the first snapshot and then schedule an SLA Template Policy for all future snaps.

Relaunching Failed Jobs

All scheduled jobs are automatically re-launched if they fail. The number of retries depends on the configuration value that is set in the InfoSphere VDP Appliance.

You can view the relaunched jobs from **System Monitor** > **Jobs** with the job status as 'retry'. To view the details of a relaunched job, double-click the job. See the IVGM online help for more information on the IBM InfoSphere System Monitor.



☐	JOB	STATUS	APPLICATION	APPLIANCE	QUEUED	STARTED	ENDED	TYPE
☐	Job_4925737a	Retried	ORCLDB	CDS139-C2	05-08 11:00:00	05-08 11:05:17	05-08 11:0...	snapshot
☐	Job_4885635a	Retried	Win2012_4SQL	CDS139-C1	05-08 11:00:00	05-08 11:04:27	05-08 11:0...	dedupasync
☐	Job_4925733a	Retried	ACT-WIN2K3	CDS139-C2	05-08 11:00:00	05-08 11:04:15	05-08 11:0...	directdedup
☐	Job_4925737	Retried	ORCLDB	CDS139-C2	05-08 11:00:00	05-08 11:00:03	05-08 11:0...	snapshot
☐	Job_4885635	Retried	Win2012_4SQL	CDS139-C1	05-08 11:00:00	05-08 11:00:05	05-08 11:0...	dedupasync
☐	Job_4925733	Retried	ACT-WIN2K3	CDS139-C2	05-08 11:00:00	05-08 11:00:03	05-08 11:0...	directdedup
☐	Job_4885253b	Retried	aw1	CDS139-C1	05-08 10:00:00	05-08 10:37:48	05-08 10:4...	dedupasync
☐	Job_4885253a	Retried	aw1	CDS139-C1	05-08 10:00:00	05-08 10:14:06	05-08 10:2...	dedupasync
☐	Job_4885253	Retried	aw1	CDS139-C1	05-08 10:00:00	05-08 10:02:23	05-08 10:1...	dedupasync
☐	Job_4925340a	Retried	ORCLDB	CDS139-C2	05-08 10:00:00	05-08 10:04:54	05-08 10:0...	snapshot
☐	Job_4925339a	Retried	ACT-WIN2K3	CDS139-C2	05-08 10:00:00	05-08 10:04:43	05-08 10:0...	directdedup

1 - 25 of 26643 jobs

1 of 1066 pages

The List of Retried Jobs in the System Monitor

On-Demand Jobs

The great majority of jobs run on schedule according to their SLAs, but for upcoming maintenance windows, software upgrades, and for the first snapshot of a new application, you want to ensure that you have a successful copy of the data created before you start your scheduled maintenance task. These cases call for an on-demand job.

About Job Slots

The InfoSphere VDP Appliance manages jobs by assigning *job slots*. The InfoSphere VDP Appliance reserves a pool of slots for each category of jobs, plus an pool of unreserved slots.

Before starting a job, the InfoSphere VDP Appliance checks whether a slot corresponding to the job's category is available to run the job. When a reserved slot is not available because all the slots of that category are running jobs, the InfoSphere VDP Appliance checks whether an unreserved slot is available. If an unreserved slot is available, the job is started.

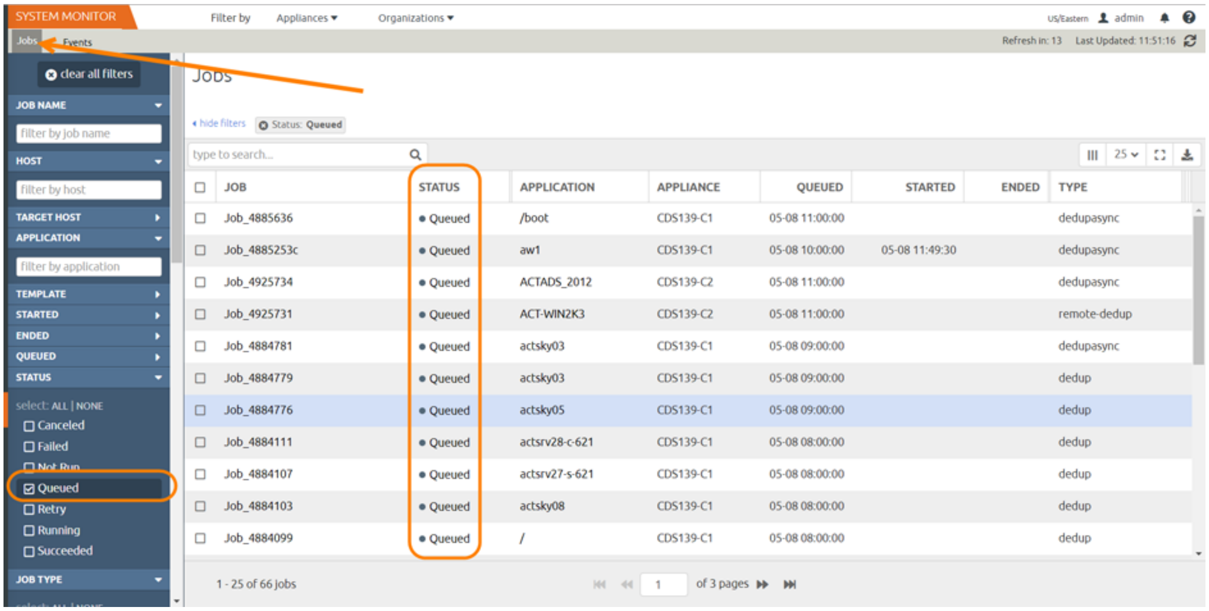
Queuing of On-Demand Backup Jobs

The InfoSphere VDP Appliance supports queuing of on-demand jobs to provide you with the flexibility to create your images without concern for the number of on-demand job slots available to start the job. The queued on-demand job remains in the queued state until an on-demand job slot is available.

When an on-demand slot opens, the job progresses to the running state. This sequence occurs in the order that the job was submitted. If an on-demand job fails, the InfoSphere VDP Appliance will attempt to run the next job in the queue. On-demand jobs use different job slots than scheduled jobs, so scheduled jobs may run before queued jobs.

While an on-demand job is in a queued state you can cancel the job or cancel protection for the application. The on-demand job will then appear in the job history table as a canceled job. The start time of the job and the end time of the job will be the time that the cancel request or the cancellation of application protection was acknowledged.

You can view the queued jobs from **System Monitor > Jobs**.



JOB	STATUS	APPLICATION	APPLIANCE	QUEUED	STARTED	ENDED	TYPE
Job_4885636	● Queued	/boot	CDS139-C1	05-08 11:00:00			dedupasync
Job_4885253c	● Queued	aw1	CDS139-C1	05-08 10:00:00	05-08 11:49:30		dedupasync
Job_4925734	● Queued	ACTADS_2012	CDS139-C2	05-08 11:00:00			dedupasync
Job_4925731	● Queued	ACT-WIN2K3	CDS139-C2	05-08 11:00:00			remote-dedup
Job_4884781	● Queued	actsky03	CDS139-C1	05-08 09:00:00			dedupasync
Job_4884779	● Queued	actsky03	CDS139-C1	05-08 09:00:00			dedup
Job_4884776	● Queued	actsky05	CDS139-C1	05-08 09:00:00			dedup
Job_4884111	● Queued	actsrv28-c-621	CDS139-C1	05-08 08:00:00			dedup
Job_4884107	● Queued	actsrv27-s-621	CDS139-C1	05-08 08:00:00			dedup
Job_4884103	● Queued	actsky08	CDS139-C1	05-08 08:00:00			dedup
Job_4884099	● Queued	/	CDS139-C1	05-08 08:00:00			dedup

1 - 25 of 66 jobs — 1 of 3 pages

Queued Job List in the System Monitor

Maintaining Performance When Adding New Applications

If your system has been performing acceptably and then you add new applications, performance may suffer for a short time. This is because IBM InfoSphere change block tracking recognizes new data and protects it even when it is only a small part of a large application. This means the system is optimized to process many changed blocks every day.

A new application requires a lot more resources for the initial capture, because it is all new data to the IBM InfoSphere system.

For best results when adding new applications:

- When you add a new application, protect it for the first time using an on-demand job during a period of light load. This will prevent the resource-intensive initial ingest job from interfering with other jobs.
- When adding multiple new applications or VMs, try to stagger the initial protection jobs for each new application over time, to prevent all of the new data from being ingested simultaneously. Do this by assigning SLAs that run at different times. You can also use the on-ramp job slots feature to minimize disruption.
- Separate the initial protection job in time from the dedup or Mirror job. Once an application snapshot has been taken, the deduplication or Mirror job can run some hours later when the system load is lighter.
- When you need to add additional applications, check your MDL. If your managed data is close to or over your licensed capacity, contact your IBM InfoSphere representative to ensure continued high performance.
- Consistency Groups can be an efficient way to protect multiple applications with similar needs; see Capture Application Data in IBM InfoSphere Consistency Groups on page 23
- Be aware of your existing SLAs and try not to schedule snapshot jobs simultaneously with the snapshot jobs for very large or dynamic applications.

Working with Preserved Snapshot and Dedup Images

You can display a list of preserved/discarded snapshot and dedup images in the Images section of the Domain Manager.

- From the Preserved Images window you can:
  - View a Preserved Images history to see how many snapshot and dedup images were preserved over a period of a week or a month.
  - Select a single snapshot or dedup image and access this image in the Restore window of the Application Manager.
  - Expire a single image or multiple preserved images.
- From the Discarded Images window you can see a summary of images that have been expired without processing over the past day, week, or month.

You can also monitor Image Preservation status for snapshot and dedup images:

- The Dashboard displays an Image Preservation status under the System Health widget to inform you of the general status of preserved snapshot images and dedup images.
- The System Monitor displays alerts and warnings specific to image preservation to inform you of different activities that occur when the InfoSphere VDP Appliance attempts to preserve images.

This section includes the following topics:

- Viewing Preserved Images History on page 42
- Working with Preserved Snapshot and Dedup Images on page 42
- Viewing Discarded Images on page 43
- Monitoring Preserved Image Status in the VDP Desktop Dashboard on page 43
- Alerts and Warnings for Image Preservation in System Monitor on page 44

For details on how to modify or disable the application priority settings for preserved snapshots jobs and/or local dedup jobs, see the IVGM online help.

Viewing Preserved Images History

You can view a graph that shows how many images were in a preserved state on each day over a selected time period (which can be an interval of either Last Week or Last Month). Data is logged in the Preserved Images History graph on an hourly basis.

Working with Preserved Snapshot and Dedup Images

You can choose from the list of preserved snapshot and dedup images and perform the following actions on those images from the Domain Manager:

- Select from the list of preserved images and navigate to that image in the Application Manager.
- Expire one or more selected snapshot or dedup images.

Note: *When expiring snapshots, the amount of space reclaimed may be less than the amount of space consumed by that snapshot. This is due to the common block reference between snapshots. To ensure maximum space reclamation, expire the oldest snapshot first.*

Viewing Discarded Images

Preserved images will automatically be expired and discarded when pool space or VDisk count reaches the warning threshold levels set in the Domain Manager. In this case, images are expired based on application priority and age. Images for applications with lower priority will be expired ahead of applications with higher priority. Within a priority level, older images will be expired before newer images.

From the Discarded Images window, you can see a summary of images that have been expired over the past day, week (7-day interval), or month (30-day interval) along with the reason for discarding the image. This window also includes images that have been manually expired prior to processing.

Monitoring Preserved Image Status in the VDP Desktop Dashboard

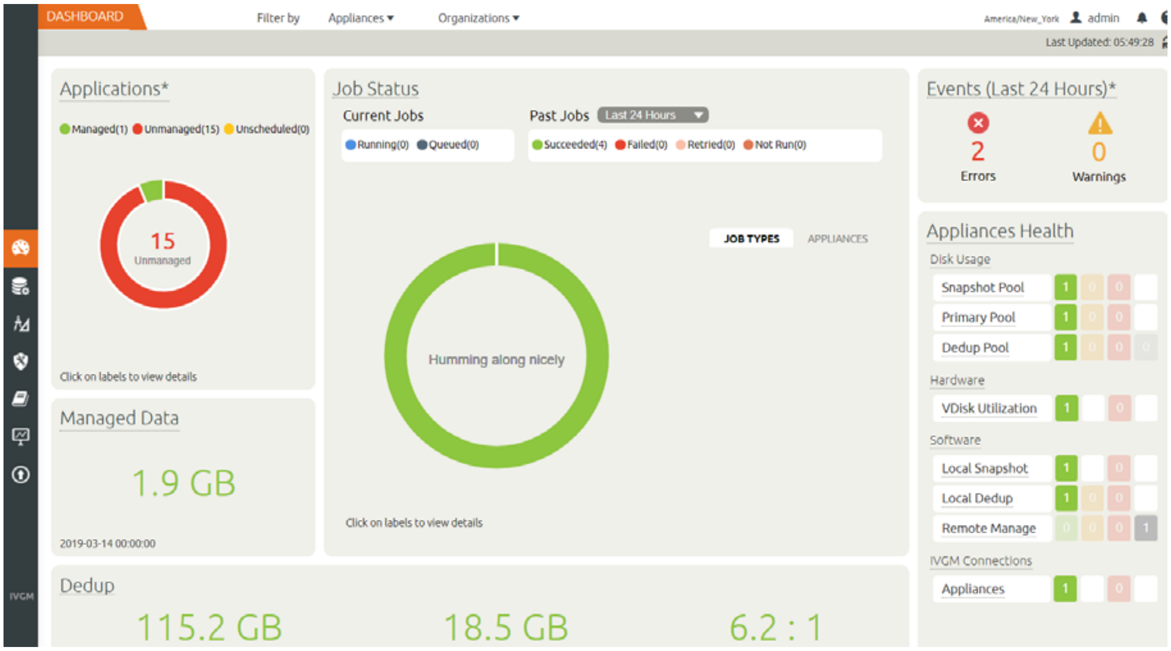
The Dashboard includes Image Preservation status under the System Health widget to inform you of the general status of preserved snapshot images and dedup images.

> ***Note:** See IVGM Online Help for additional details about the IBM InfoSphere Dashboard and the System Health widget.*

Image Preservation states include:

- Image Preservation is enabled, and one of the following conditions exists:
  - Green = No images are being preserved beyond their expiration date.
  - Yellow = At least one image is being preserved beyond expiration but no images are more than 7 days beyond expiration (yellow).
  - Red = At least one image is being preserved beyond expiration and at least one image is more than 7 days beyond expiration (red).
- Image Preservation Mode is not enabled (gray).

The Image Preservation History button opens a chart that shows how many images were preserved on each day over the selected time period (which can be an interval of either Last Week or Last Month). This is the same chart that appears from the Preserved Images tab in the Domain Manager (see Viewing Preserved Images History on page 42).



Viewing Preserved Images Status in the Dashboard

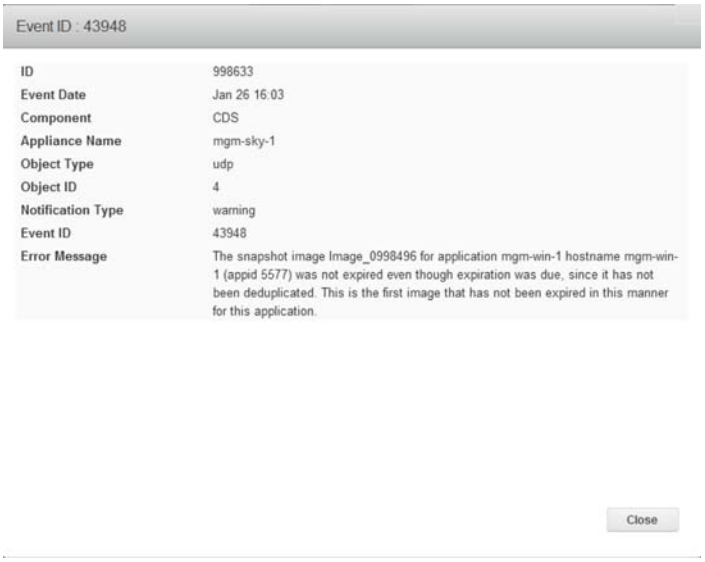
Alerts and Warnings for Image Preservation in System Monitor

This section outlines the various alerts and warnings related to image preservation.

Warning Level Alert: First Time the Snapshot Expiration Window is Reached

A Warning level alert is generated (and posted to the event log) the first time a snapshot that is eligible for expiration is held for pending processing. A similar Warning level Alert also occurs for remote deduplication of local dedup images.

This Warning level alert is generated for the first snapshot for each application that has its expiration deferred. When the count of deferred expirations for an application goes to zero, the Warning alert trigger is reset. The next time there is a dedupable snapshot image that is held by the InfoSphere VDP Appliance an alert will be posted again. An example of this particular Warning alert message is shown below:



Event ID : 43948

ID	998633
Event Date	Jan 26 16:03
Component	CDS
Appliance Name	mgm-sky-1
Object Type	udp
Object ID	4
Notification Type	warning
Event ID	43948
Error Message	The snapshot image Image_0998496 for application mgm-win-1 hostname mgm-win-1 (appid 5577) was not expired even though expiration was due, since it has not been deduplicated. This is the first image that has not been expired in this manner for this application.

Warning Level Alert: Snapshot Image Expired Because Threshold Limit Exceeded

When an application has preserved snapshots, and a dedupable snapshot is expired because the InfoSphere VDP Appliance has exceeded the threshold limits (such as VDisk count or pool capacity), a Warning level alert indicating this condition is posted. This warning will be logged only for the first snapshot expired due to this situation. The same requirement applies to local dedup images with regard to remote dedup replication.

Below is a summary of the Warning thresholds for VDisk and storage pools as specified in the Domain Manager:

- The Warning threshold for VDisks usage is 90%. The VDisk limit for the VDP Appliance varies with the installed capacity license (1000, 3000, or 5000 VDisks).
- The Warning level is 80% for the snapshot pool and 75% for the dedup pool. The default value is 90% for the snapshot and primary pools.

Daily Warning Level Event: Deferred Expirations for Snapshots

A daily warning level event is generated when there are deferred expirations for snapshots. This daily warning includes a count of images for which expiration was deferred because these images are all candidates for deduplication. An example of such an daily warning level event is shown below:

```
The number of images not expired awaiting further processing is 2 images (2 snapshots, 0
dedups) from 1 unique applications. 2 snapshots and 0 dedups were added in the last period
of 24 hours.
```

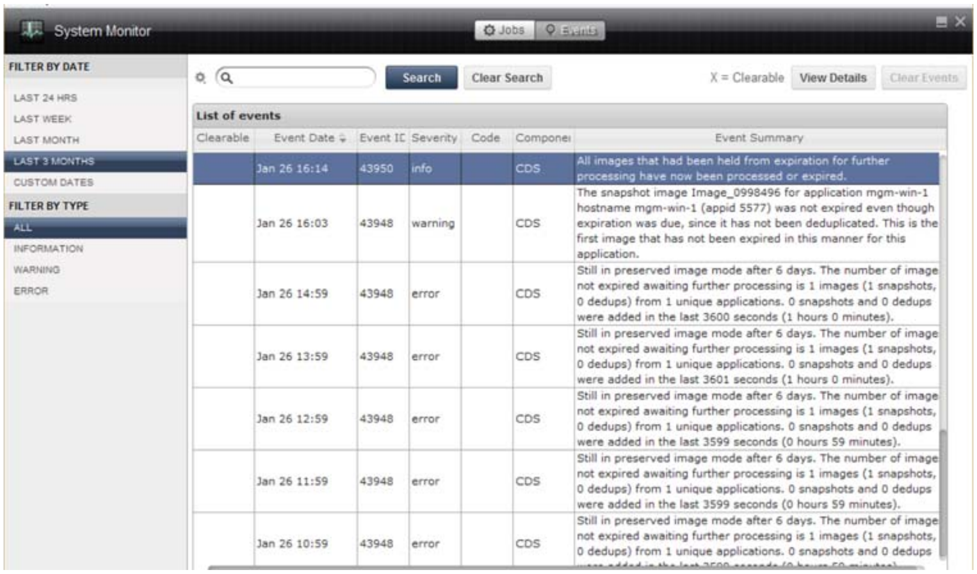
Daily Warning Level Event: Deferred Expirations Because Threshold Limit Exceeded

A daily warning level event is posted when a number of images that had deferred expirations were expired because the InfoSphere VDP Appliance has exceeded the threshold limits (such as VDisk count or pool capacity). The message includes a count of images expired in this fashion. An example of such an event is shown below:

```
The number of images awaiting further processing that had to be discarded is 5 images (3
snapshots, 2 dedups) from 3 unique applications in the last period of 24 hours.
```

Warning Level Alert: All Preserved Images Have Been Processed

When the number of preserved images drops to zero, the following alert is posted in the System Monitor:



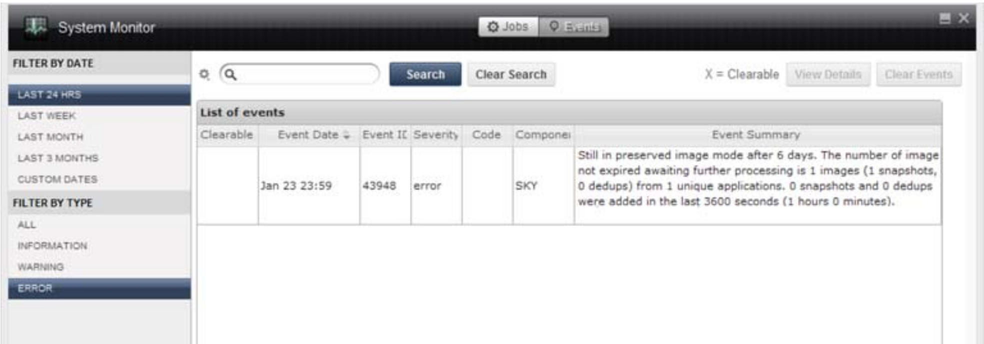
List of events

Clearable	Event Date	Event IC	Severity	Code	Componei	Event Summary
	Jan 26 16:14	43950	info		CDS	All images that had been held from expiration for further processing have now been processed or expired.
	Jan 26 16:03	43948	warning		CDS	The snapshot image Image_0998496 for application mgm-win-1 hostname mgm-win-1 (appid 5577) was not expired even though expiration was due, since it has not been deduplicated. This is the first image that has not been expired in this manner for this application.
	Jan 26 14:59	43948	error		CDS	Still in preserved image mode after 6 days. The number of image not expired awaiting further processing is 1 images (1 snapshots, 0 dedups) from 1 unique applications. 0 snapshots and 0 dedups were added in the last 3600 seconds (1 hours 0 minutes).
	Jan 26 13:59	43948	error		CDS	Still in preserved image mode after 6 days. The number of image not expired awaiting further processing is 1 images (1 snapshots, 0 dedups) from 1 unique applications. 0 snapshots and 0 dedups were added in the last 3601 seconds (1 hours 0 minutes).
	Jan 26 12:59	43948	error		CDS	Still in preserved image mode after 6 days. The number of image not expired awaiting further processing is 1 images (1 snapshots, 0 dedups) from 1 unique applications. 0 snapshots and 0 dedups were added in the last 3599 seconds (0 hours 59 minutes).
	Jan 26 11:59	43948	error		CDS	Still in preserved image mode after 6 days. The number of image not expired awaiting further processing is 1 images (1 snapshots, 0 dedups) from 1 unique applications. 0 snapshots and 0 dedups were added in the last 3599 seconds (0 hours 59 minutes).
	Jan 26 10:59	43948	error		CDS	Still in preserved image mode after 6 days. The number of image not expired awaiting further processing is 1 images (1 snapshots, 0 dedups) from 1 unique applications. 0 snapshots and 0 dedups

System Monitor -- Event Page with All Preserved Images Have Been Processed Event

Weekly Error Level Event: Images Deduplicated or Remotely Replicated After 7 Days

When there are images that have not been deduplicated or remotely replicated for a period of 7 days, a weekly error level event of severity Error is raised. When the 7th day is reached an alert will be generated to inform you that the InfoSphere VDP Appliance has been in Preserve Image Mode for 7 days.



List of events

Clearable	Event Date	Event IC	Severity	Code	Componei	Event Summary
	Jan 23 23:59	43948	error		SKY	Still in preserved image mode after 6 days. The number of image not expired awaiting further processing is 1 images (1 snapshots, 0 dedups) from 1 unique applications. 0 snapshots and 0 dedups were added in the last 3600 seconds (1 hours 0 minutes).

System Monitor -- Event Page with Alert That Appliance Has Been in Preserve Mode for x Days

6 Accessing Data

This chapter describes the various ways in which you can access your captured data:

For detailed, application-specific, step-by-step instructions on how to access data, refer to the IVGM online help.

Mounts

The IBM InfoSphere mount function provides instant access to data without moving data. There are two options for mounting data:

- The **standard mount** presents and makes application data available to a target server as a file system, not as an application. This is useful if an application is corrupt, lost, or if an application server is being replaced. In such cases you can mount an image and copy the application files from the mounted image to their original location on the application server.

  Note: *VDP Appliances can use a standard mount to access data that resides in an IBM InfoSphere OnVault.*

- **Application aware mounts** allow you to mount captured Microsoft SQL and Oracle databases as virtual applications. This allows you to quickly bring a database on line without having to actually move the data and without having to manually configure a new instance of the database.

  Application aware mounts are particularly useful in test and development environments where multiple copies of a database must be quickly brought on line.

  Data presented as an application aware mount can be captured like any other application. Once the application aware mounted application data is captured, it too can be can be mounted as an application aware mount.

  The capture, application mount, capture sequence can be repeated to any depth. By default, the sequence is restricted to five generations of the original database.

Clones

Use the clone function to create an independent copy of a data set. The most common uses are: application development and testing, data audit for compliance, data warehousing, e-discovery, and user acceptance testing. Physical server or VM application-consistent data sets can be copied to a separate storage location anywhere in your environment. Like any other VM, a VM clone can be migrated to a new storage location.

LiveClones

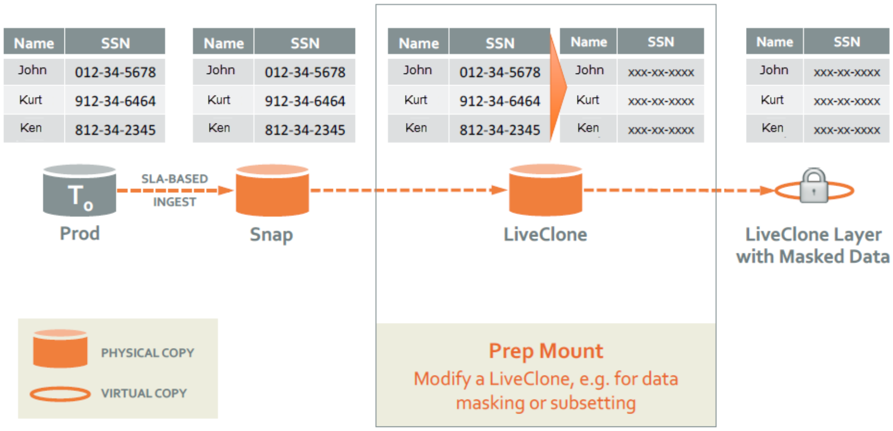
The LiveClone is similar to the Clone function, however, unlike a Clone, a LiveClone can be updated on demand or according to a schedule. When an updated copy of the data is available, a LiveClone allows an independent copy of a data to be mounted. This allows teams such as development and test to ensure they are working on the latest set of data without having to manually manage the data.

Workflows

Within the context of an InfoSphere VDP Appliance, a Workflow refers to the automation of access to copy data in a scheduled or prescribed fashion. While SLA Policy Templates govern the automated capture of production data and its management as copy data within the Virtual Data Pipeline, Workflows automate the access to this data.

Steps are defined within a Workflow to perform a series of tasks on a schedule or on demand. This includes creating and refreshing LiveClones, data masking, persistent mounts, and non-persistent processing mounts for tasks such as tape-out, database integrity checks, and ETL loads. Workflows are also used by administrators to provide simplified and secured self-service data access to end users such as database administrators and application developers.

The following image provides a high-level description of a Workflow that creates a LiveClone from production data, then scrubs the LiveClone of sensitive information, before mounting the LiveClone to a work environment.



Name	SSN
John	012-34-5678
Kurt	912-34-6464
Ken	812-34-2345

Name	SSN
John	012-34-5678
Kurt	912-34-6464
Ken	812-34-2345

Name	SSN
John	012-34-5678
Kurt	912-34-6464
Ken	812-34-2345

Name	SSN
John	xxx-xx-xxxx
Kurt	xxx-xx-xxxx
Ken	xxx-xx-xxxx

Name	SSN
John	xxx-xx-xxxx
Kurt	xxx-xx-xxxx
Ken	xxx-xx-xxxx

Workflow With Scrubbed Data

Restores

The Restore function reverts the production data to a specified point in time. Restores and clones are the only data access operations that actually move data. Typically restore operations are performed to restore an application to a valid state after a massive data corruption or storage array failure. The amount of time required to complete a restore operation depends on the amount of data involved.

Granular Restores

If the catalog option is enabled for virtual machines/file system applications, then IBM InfoSphere will index the files/folders and the associated metadata in to a global searchable index. This allows for granular search and recovery of files/folders.

1 Glossary of Terms

A | B | C | D | E | F | G | H | I-K | L | M-N | O | P-Q | R | S-T | U | V | W-Z

	Term	Definition
A	**App or Application**	An app or application is a data resource that can be discovered and protected by an InfoSphere VDP Appliance. Examples include Oracle or SQL databases, Exchange databases, network or local file systems or parts of file systems, virtual or physical machines, and so on.
	Appliance	An “appliance” is the generic term for an IBM InfoSphere virtual server. Virtual IBM InfoSphere appliances are referred to as VDP Appliances.
	Application Aware mount	See Virtual Application.
	Application Manager	The Application Manager service in the IBM InfoSphere VDP - Global Manager and in the VDP Desktop is used to discover applications, application data, and virtual machines, and to apply protection templates and resource profiles to them.
B	**Baseboard System Identification**	The base board system identifier (BBSID) is an arbitrary unique number that becomes part of a unique suffix for a node’s World Wide Node Number and World Wide Port Number. These both must be unique within a fabric. The BBSID is also used by IBM InfoSphere to generate a unique ID for IBM InfoSphere SecureConnect access.
	BBSID	Baseboard System Identification
C	**CBT**	Changed Block Tracking.
	Changed Block Tracking	Changed block tracking is the process of comparing the golden snapshot to incremental point in time snapshots in order to identify changed data that must be preserved.
	CLI	Command line interface.
	Clone	The clone function creates an independent copy of a data set. A virtual server or physical server data set can be copied from any application-consistent point in the system to a separate storage location anywhere in the environment.
	Clone VDisks	Clone VDisks, are the part of a Snapshot pool that contains full copies of an application’s production data.
	Cloud Migration	Cloud migration/data migration enables organizations to manage their data proactively and stage it as required across diverse, distributed infrastructure. The ability to provide secure, network-optimized replication allows you to efficiently move data wherever it is needed.

	Term	Definition
	Cloud Vaulting	Cloud vaulting enables data vaulting across diverse infrastructure, including the use of targets in customer-owned data centers and field locations, private or service provider clouds, or public clouds like Amazon, Azure, and GCP.
	Connector	See VDP Connector.
	Consistency Group	A group of storage resources protected as a single entity by an IBM InfoSphere appliance.
	Copy Data Virtualization	Copy Data Virtualization is the IBM InfoSphere data management model — capture data and process it in a virtual data pipeline to create a single golden master copy that is incrementally updated according to a service level agreement (SLA) and is used to generate a virtual copy of any application data from any point in time for any authorized use.
D	**DAR**	See Dedup Async Replication (DAR).
	Dedup	Deduplication is a storage technology and a process that reduces the amount of storage space consumed by data by removing redundant data.
	Dedup Async Replication (DAR)	Dedup Async™ Replication is a unique, proprietary IBM InfoSphere technology that keeps a remote copy of production data always up-to-date and ready for data recovery. Also see Dedup Backup Replication and StreamSnap.
	Dedup Backup Replication	An IBM InfoSphere proprietary deduplication-aware replication protocol used for replication of captured images from one InfoSphere VDP Appliance to a second and optionally to a third for long-term storage. Also see StreamSnap, and Dedup Async Replication (DAR).
	Domain Manager	The Domain Manager service in the IBM InfoSphere VDP - Global Manager and in the VDP Desktop controls the organizations and users that have access to an InfoSphere VDP Appliance, identify hosts that IBM InfoSphere can protect, and manages the resources on which copy data resides.
E	**Enumeration**	Enumeration is the first phase Garbage Collection. Enumeration analyzes the catalog of deduplicated data to determine what data must be kept and what can be deleted. Enumeration is followed by Sweep.
	External Snapshot Pool	IBM InfoSphere has extended its Virtual Data Pipeline to use and manage external snapshot pools using with IBM Storwize/SVC and Pure FlashStorage. You can use the array native snapshots for IBM InfoSphere's snapshot pool, gaining the storage arrays' performance, connectivity, and availability.
F	**Failback**	Failback is the recovery process used when a primary system or data is restored to operation after a Failover. Also see Syncback.
	Failover	The process of using a secondary system, usually hardware, to replace a primary system that fails during operation. Also used to describe the data copied when a failover occurs. See Failback and Syncback.
	Fibre Channel	Fibre channel is a high-speed network technology commonly running at 2, 4, 8 or 16-gigabit per second that is used primarily to connect data storage devices.
	Filter Driver	The mechanism used by the VDP Connector for Changed Block Tracking.
G	**Garbage Collection**	Garbage collection it the two-phase process of reclaiming space in the dedup pool. Enumeration, the first phase, is the selection of data that can be deleted. Sweep is the removal of the unneeded data.
	GC	See Garbage Collection.
H	**Host**	A server with managed or manageable applications.

	Term	Definition
	Hyper-V	Microsoft’s virtual machine platform is a native hypervisor that can create virtual machines on x86-64 systems.
I-K	**IBM InfoSphere VDP - Global Manager**	The IBM InfoSphere VDP - Global Manager (IVGM) provides a web-based interface to manage multiple InfoSphere VDP Appliances, including day-to-day copy data operations. IVGM is the successor to the VDP Desktop.
	Immutability	You can use policy settings to make a backup image immutable. An immutable image cannot be expired by any user until it reaches a date set in the policy.
	iSCSI	The Internet Small Computer System Interface works on top of the Transport Control Protocol (TCP) and allows the SCSI command to be sent end-to-end over local-area networks (LANs), wide-area networks (WANs) or the Internet.
	IVGM	See IBM InfoSphere VDP - Global Manager.
L	**LAN-Free**	LAN-free is a network architecture in which application data is protected using a a shared, central storage device without sending the data over the local area network (LAN).
	LiveClone	An independent clone of a captured image that consumes full storage resources and can be mounted to a host. It can be refreshed incrementally from another captured image, allowing very fast and efficient data refreshes for ETL and test & development purposes. A LiveClone can also be mounted for direct modification to support operations such as data masking.
M-N	**Managed Data License (MDL)**	IBM InfoSphere’s Copy Data Virtualization licensing. It is based on the amount of source data managed.
	Managed Disk	A SCSI Disk presented by a RAID controller and managed by the InfoSphere VDP Appliance. The Managed Disk is not visible to host systems on the SAN.
	Managed disk group (MDiskgrp or MDG)	A collection of Managed Disks that jointly contain all the data for a specified set of Virtual Disks.
	MDisk	These are disks presented to and managed by the IBM InfoSphere solution.
	Mount	The mount function is the most frequently used data access method, as it directly leverages the virtual copies of data stored on an InfoSphere VDP Appliance. By eliminating the data movement from the process, data sets of any size can be accessed instantly on any server.
	Multi-Hop replication	Replication, usually Dedup Backup Replication, is the process that replicates data from a “source” InfoSphere VDP Appliance to a “remote” InfoSphere VDP Appliance, and then from the remote appliance to a third InfoSphere VDP Appliance.
O	**OnVault**	The InfoSphere VDP Appliance vaults data to selected cloud storage according to a defined OnVault policy. Supported cloud storage platforms include Amazon S3, Google Nearline Storage, IBM Cloud Object Storage, and Microsoft Azure. Users manage and pay for their own cloud storage directly with the provider
P-Q	**Performance pool**	The Snapshot pool.
	Policy	A policy defines when data will be captured, how long it will be retained, and where it will be replicated.
	Policy template	A collection of policies that, together, define when to perform a snapshot, when to perform dedup activity that creates an image, and how long to retain the image.
	PSRV	The platform service is a component of IBM InfoSphere VDP software that coordinates other VDP services and functions.

	Term	Definition
R	**RD**	See Resiliency Director.
	ReadyVM	The CLI shorthand term for replicating a VMware VM to an ESX datastore. This is an asynchronous replication mechanism in which the data is directly replicated onto the datastore volumes that are configured for the remote VM. This lets you use an existing or a new virtual machine as the replication target.
	Report Manager	The report manager is an optional stand alone software package that reports on data protection and recovery operations.
	Resiliency Director	Resiliency Director is an optional product that works with InfoSphere VDP Appliances to create and manage data that are part of disaster recovery services.
	Resource Profile	A resource profile specifies if, and which, Snapshot pool is used by IBM InfoSphere and/or to which remote InfoSphere VDP Appliance data will be replicated. A resource profile is paired with policy templates to protect a specific application by the Application Manager.
	Restore	The restore function reverts the production data to look exactly as it did at the time of the data collection point. Typical use cases for restore would be to recover an entire server or application to a valid state after a massive data corruption or storage array failure.
	RM	Report Manager.
	RPO	A recovery point objective is the maximum period in which data might be lost from an IT service due to a major incident. See RTO.
	RTO	The recovery time objective is a period of time and a service level within which a business process must be restored after a disruption in order to avoid a break in business continuity. See RPO.
S-T	**Service Level Agreement**	An IBM InfoSphere service level agreement is the linkage of a single policy template that defines when to perform actions, and a resource template that defines what storage resources are used by the actions.
	SideBand	See LAN-Free
	SLA	See Service Level Agreement.
	Snapshot	A snapshot is the process that captures and stores the state of a Snapshot VDisk as a Snapshot VDisk.
	Snapshot pool	The snapshot pool holds “golden copies” of application data for short-term retention. Data is instantly accessible and not deduplicated. Policies determine how long data is kept in the pool and when data is deduplicated and moved to another pool. The snapshot pool contains Staging VDisk, Snapshot VDisk, and Clone VDisks.
	Snapshot VDisk	A Snapshot VDisk is part of a Snapshot pool that preserves the state of Staging VDisk at specific points in time. Snapshots are retained according to a predefined protection policy.
	Staging VDisk	A Staging VDisk is part of a Snapshot pool that contains the IBM InfoSphere golden copy of an application. It is retained for as long as an application is protected.
	StreamSnap	Direct replication of incremental snapshots from a local snapshot pool to a remote pool, supporting a much lower RPO compared with Dedup-Async replication. StreamSnap is used in high-quality, high-bandwidth IP networks. StreamSnap keeps a full virtual copy of the application on the remote side, available for immediate failover, test failover, or mount operations.

	Term	Definition
	Sweep	Sweep is the second phase Garbage Collection. It copies data that must be preserved from discrete areas of the dedup store into a new, contiguous area of the dedup store to both defragment storage media and to free areas for reuse. See Garbage Collection and Enumeration.
	Syncback	Syncback is the process that verifies data that has failed over to be valid before a Failback. Also see Failover.
	System Monitor	A service within the IBM InfoSphere VDP - Global Manager and in the VDP Desktop the monitors the process of jobs.
U	**UDS**	Universal data system
V	**VDisk**	Also referred to as a *volume*. See Virtual disk.
	VDP	See Virtual Data Pipeline™.
	VDP Appliance™	VDP Appliance is a robust virtual appliance built on IBM InfoSphere's patented Virtual Data Pipeline™ (VDP). VDP Appliance offers deployment flexibility and range. As a virtual appliance, VDP Appliance can be deployed in minutes at any site across an organization's locations and environments.
	VDP Connector	VDP Connector is a lightweight service that may be run on physical or virtual appliances. It discovers and captures individual applications virtual and physical machines and servers so they can be replicated.
	VDP Desktop	Legacy software for controlling the configuration and operation of individual virtual VDP Appliances.
	Virtual Application	You can mount captured Microsoft SQL and Oracle databases as "Application Aware mounts", fully functional replicas of the original database. This allows you to quickly bring a database online without having to manually configure a new instance of the database and actually move the data into it.
	Virtual Data Pipeline™	IBM InfoSphere's data virtualization platform helps companies take control of their data, delivering greater resilience and agility while enabling secure mobility of data to and from the cloud. By virtualizing data, creating a single "gold copy" available for instant access and use, IBM InfoSphere frees application data from underlying infrastructure, enabling transformational change in IT and the business.
	Virtual disk	These are disks presented to applications by the IBM InfoSphere solution that appear to host systems attached to the storage area network as a SCSI disk. Each VDisk is associated with one I/O group.
	VM	Virtual machine. IBM InfoSphere supports both VMware and Hyper-V instances.
W-Z	**Workflow**	IBM InfoSphere Workflows automate access to captured data. Workflows can run according to a schedule or on demand. Workflows present captured data as a LiveClone, a virtual application, or as just the application data.