Hewlett Packard Enterprise

Technical white paper

HPE StoreOnce Catalyst: A Solution for Enhanced Data Protection

A detailed definition of HPE StoreOnce Catalyst and its functionality

Contents

Executive summary

Among the many features of the Hewlett Packard Enterprise (HPE) StoreOnce backup appliance, HPE StoreOnce Catalyst is perhaps the most important capability to understand. IT professionals struggling with backup performance and looking to implement a simple yet comprehensive data protection solution will find the capabilities of StoreOnce Catalyst to be most compelling. This white paper details the architecture and functionality of StoreOnce Catalyst and explores the benefits of deploying Catalyst as part of a complete HPE data protection solution.

StoreOnce Catalyst definition

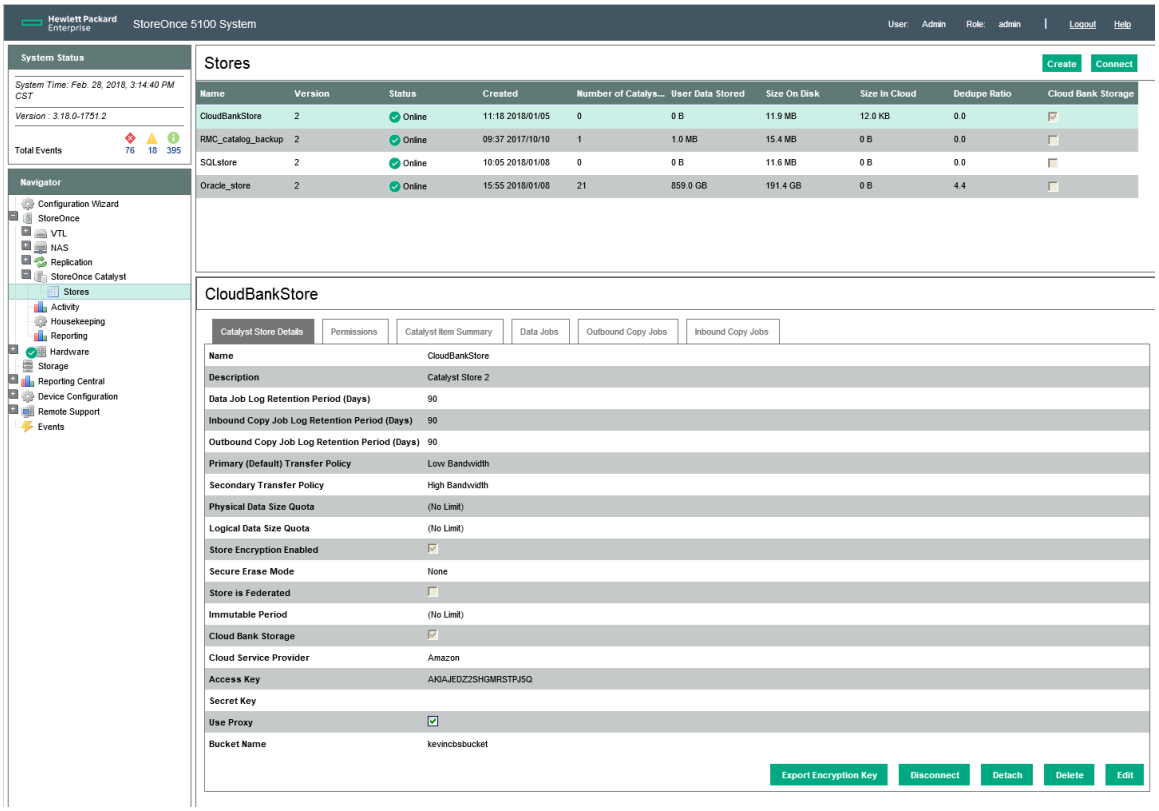

Figure 1. StoreOnce GUI showing Catalyst store details

What is StoreOnce Catalyst? Simply put, Catalyst is a backup protocol developed by HPE that is optimized for disk-based data protection. Data protection applications can use Catalyst to store, replicate, and archive customer data. Users create Catalyst "stores" that reside on StoreOnce appliances or in the cloud. These stores serve as targets for storing data protection jobs. There are four main advantages of using Catalyst to protect data:

- **Simplicity:** Users can create Catalyst stores with just a few clicks in the StoreOnce GUI. Catalyst stores do not require preallocated space; they automatically grow as needed. Authentication, encryption, and cloud archiving can be enabled in seconds.
- **High-performance:** Catalyst stores use low-bandwidth deduplication by default. Low bandwidth shifts the majority of the deduplication process to the Catalyst client, or source-side level. This ensures that only new, unique chunks of data are sent across the network to the Catalyst store. Reduced bandwidth translates to increased performance—less data to write means less time to complete the job. Low-bandwidth deduplication produces significantly better performance over a virtual tape library (VTL), which has no low-bandwidth capability. Catalyst deduplication is on a per-store basis, improving performance by reducing the need to process deduplication across the entire appliance.

- **Flexibility and control:** Catalyst stores can do more than just house primary backup data. Backup jobs written to Catalyst stores can be copied to other stores using Catalyst Copy jobs managed by the data protection application. In a disaster recovery configuration, these copies can be sent to other StoreOnce appliances. With the HPE Cloud Bank Storage feature, Catalyst Copy jobs can also be directed to stores in public or private clouds.

  Catalyst also allows the user to decide where deduplication will occur; Catalyst stores can be configured individually for either source-side or target-side deduplication. Expiration and retention settings can be configured within the data protection application, so users can control where the backup copies reside, how many copies are maintained, and when those copies expire and can be deleted.

- **Security:** Catalyst stores are adept at protecting backup data from tampering, through encryption options and in the way StoreOnce handles and processes the data. The Catalyst application programming interface (API) isolates stores from operating system commands, providing protection against ransomware attacks. In addition, malware is incapable of activating within a Catalyst store, because Catalyst does not use standard operating system commands for its operations.

Catalyst technology is supported by many popular data protection applications on the market, such as Veritas NetBackup™ and Backup Exec™, Veeam Backup & Replication, BridgeHead Healthcare Data Management, and Micro Focus Data Protector, as well as HPE Recovery Manager Central. Some of the advanced features of these applications (for example, Veeam Instant VM Recovery and NetBackup Granular Recovery Technology) are only available when using Catalyst as a backup target on a StoreOnce.

With the use of specific plug-in products, Catalyst also has the ability to be used as a backup target by the native backup functions of database applications—specifically, Oracle® Recovery Manager (RMAN), SQL Server® backup, and Backint for SAP HANA®. The integration of Catalyst with these applications allows single pane of glass administration of backups, job copies, and archiving from within the application interfaces, simplifying data protection tasks. The plug-ins also enable database administrators to handle their own backups.

Catalyst comparison to VTL and NAS

StoreOnce appliances have three different backup target options for housing data: VTL, NAS, and Catalyst. To better understand how Catalyst compares to the other two options, review customers' frequently asked questions, which are summarized in Table 1.

Table 1. VTL, NAS, and Catalyst FAQs

Question	VTL	NAS	Catalyst
Will my data protection application support it?	Supported by virtually every app	Supported by virtually every app	Supported by select major apps
How easy is it to deploy?	Must define VTL, drives, cartridges, barcodes, and other options before deployment	Creates shares in seconds	Creates stores in seconds
How does it scale?	Can configure each VTL with thousands of cartridges; can scale up to capacity of appliance	Limited to 1 million files per share, but can configure multiple shares	No limit on objects in store; scales up to capacity of appliance; grows as needed
Can I replicate the data off-site?	Yes	Yes	Yes
Can I control the data from the data protection app?	VTL-to-VTL replication, but cannot be managed by an app—only through StoreOnce GUI	NAS-to-NAS replication, but cannot be managed by an app—only through StoreOnce GUI	Store-to-store copies on the appliance or in the cloud; app manages copies
Can I create multiple replicated copies in different sites?	No—persistent, one-to-one relationship between replicated cartridges	No—persistent, one-to-one relationship between replicated shares	Yes—supports one-to-one or one-to-many copies on remote appliances or in the cloud
Can I use either low-bandwidth or high-bandwidth deduplication?	No, only high bandwidth	No, only high bandwidth	Yes

Catalyst glossary

The following table defines terms related to Catalyst. Each term relates to important aspects of Catalyst operations.

Table 2. Key Catalyst definitions

Term	Description
Catalyst items	Backup data is held as "items" in a Catalyst store. The size and number of items are set in the data protection application. Items are *not* the same as files.
Catalyst server	A Catalyst server is a Catalyst component that runs on the StoreOnce appliance and connects to one or more Catalyst clients.
Catalyst clients	Catalyst clients are applications which connect to StoreOnce using the Catalyst protocol.
Command session	A command session establishes communication between the Catalyst server and Catalyst clients for Catalyst management operations such as item creation or deletion.
Data session	A data session is a backup or recovery job using Catalyst. Multistream jobs use one session per stream.
Low-bandwidth backup	Low-bandwidth backup is a process by which Catalyst performs a portion of the deduplication process on the Catalyst client system so that only unique data is sent to the Catalyst store. This reduces bandwidth and improves backup performance. This is especially helpful in ensuring backups complete within the allotted backup windows over bandwidth-constrained networks.
Catalyst Copy job	A Catalyst Copy job is a copy of a stored Catalyst backup job sent to another Catalyst store. The copy can be sent to a store on the local appliance or on another appliance or to the cloud using Cloud Bank Storage.
Client identifier	A client identifier includes login credentials used by Catalyst clients to access a Catalyst store. This requires client access permission to be enabled for the store.
Client access permission	When the client access permission setting is enabled at the Catalyst store level, a Catalyst client must use a client identifier in order to read to or write from a Catalyst store.

Architecture

The StoreOnce Catalyst architecture uses the classic client-server model to create a relationship between data protection applications and StoreOnce. As outlined in the StoreOnce Catalyst definition section, this relationship allows the application to have complete control over backup, backup copy, and cloud storage jobs. Users can also use the application to recover from any Catalyst job, no matter where it resides, simplifying the recovery process. With other solutions, the application might be aware of backups but might not know about other backup copies, requiring the user to access other interfaces for recoveries. This control, coupled with the ability for users to deploy either low-bandwidth or high-bandwidth deduplication per Catalyst store, gives Catalyst users flexibility in designing their data protection environments.

The following sections detail important Catalyst components and explain how a Catalyst job processes data.

Catalyst API interaction

The Catalyst API manages the relationship between applications and Catalyst stores.

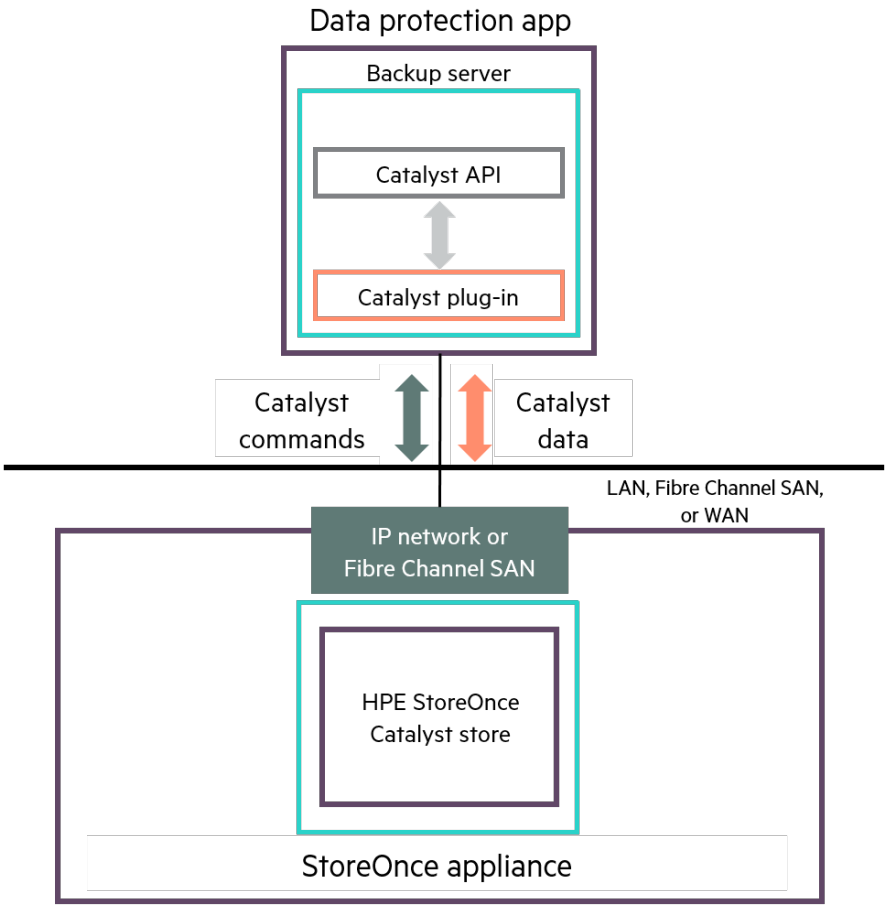

Figure 2. The Catalyst API connects the application with StoreOnce Catalyst stores

In Figure 2, a backup server with a data protection application installed on it also has a Catalyst plug-in installed. The plug-in uses the Catalyst API, developed by HPE using the OpenStorage Technology (OST) industry standard. The API translates application commands into Catalyst commands to move backup data from the backup server to the Catalyst store on a StoreOnce appliance. Catalyst provides flexibility because it can transmit over Ethernet or Fibre Channel networks, allowing the application to write data over the LAN, to a Fibre Channel SAN, or across a WAN link.

Note

With Micro Focus Data Protector, the Catalyst plug-in functionality is built into the application; no plug-in installation is necessary.

The API linkage allows the application to be able to create and manage backups, copies, and archives, no matter where they reside. The Catalyst plug-in also exchanges backup metadata with the Catalyst store during low-bandwidth deduplication. This exchange allows the plug-in to handle a portion of the deduplication in the low-bandwidth process.

Catalyst deduplication process in detail

Figure 3 details what occurs during the low-bandwidth Catalyst job process.

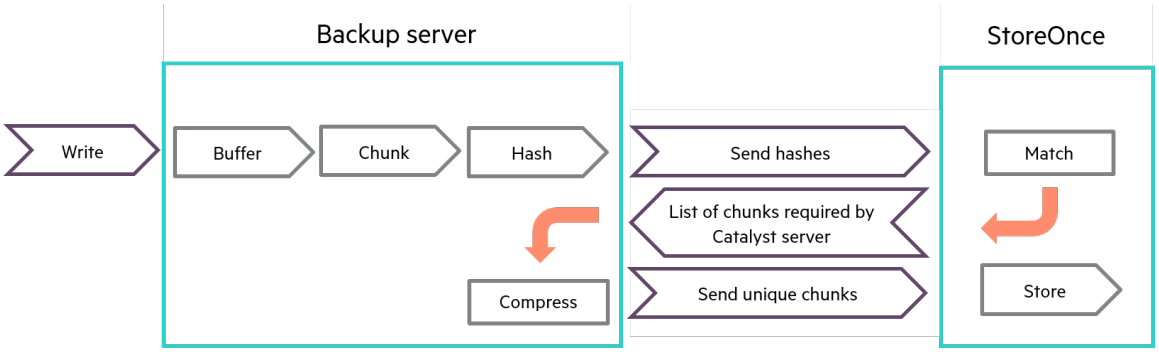

Figure 3. Catalyst data handling process—low bandwidth

When data is sent from the backup clients to the backup server, the data enters the write buffer and is then divided into chunks using the StoreOnce deduplication algorithm. These chunks are then catalogued with entries for the deduplication hash table. These hashes are then forwarded to the StoreOnce appliance, which compares the hashes against hashes for previously written chunks to find matches. Any hashes that do not have matches are returned to the backup server, providing the list of unique chunks that need to be sent to the Catalyst store. The unique chunks are then typically compressed and sent to the Catalyst store. With some data protection applications, disabling compression improves performance. In those instances, Catalyst compression can be disabled.

In a high-bandwidth Catalyst job, the backup client data is only buffered at the backup server and then sent as-is to the StoreOnce appliance. All the deduplication happens in the StoreOnce appliance. This process is illustrated in Figure 4.

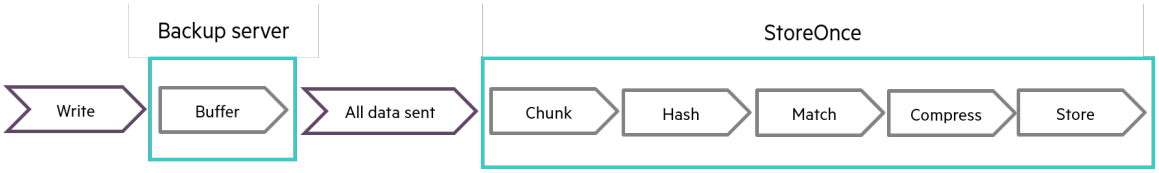

Figure 4. Catalyst data handling process—high bandwidth

Catalyst Copy

HPE endorses the 3-2-1 data protection strategy, which mandates the following:

- Keeping three copies of data
- Using two different types of media to store copies
- Storing one copy off-site

A backup job provides two of the three copies (the original and the backup) and satisfies the requirement for copies on two different types of media. To get the third copy and satisfy the off-site requirement, StoreOnce offers Catalyst Copy.

After data has been written to a Catalyst store, the data protection application can be configured to copy backup jobs in the original store to another Catalyst store. These copies are always created in low-bandwidth mode so that Catalyst Copy jobs only write unique chunks to the target Catalyst store. This is an efficient alternative to disk storage system replication, because the primary storage no longer has to handle the overhead to move data off-site for disaster recovery purposes.

Catalyst Copy jobs can be written to multiple appliances for further redundancy; they can also be written to the cloud using Cloud Bank Storage.

HPE Cloud Bank Storage

The Cloud Bank Storage feature uses the Catalyst protocol to allow copies of backup jobs to be written to an object store in the cloud. These copies can be used for long-term retention, archival, and disaster recovery purposes. Configuring a Cloud Bank Storage store is as simple and quick as configuring a regular Catalyst store. The only difference is that the store is linked to a customer-configured object store in Amazon Web Services® or Microsoft® Azure®, or in a Scality private cloud configuration. To the data protection application, it is presented as just another Catalyst store, but the data is actually written to the object store during the Catalyst Copy process. Cloud Bank Storage should only be used for Catalyst Copy jobs, not for direct backups.

Cloud Bank Storage is designed to minimize both the time needed to archive data in the cloud and the costs associated with cloud archival. As with all Catalyst Copy jobs, data is sent in low-bandwidth mode to improve write performance. Plus, with fewer data blocks stored in the cloud, the per-gigabyte charges from the cloud provider are kept in check.

Recovery is optimized as well. Cloud Bank Storage recoveries only require retrieving Catalyst items not already stored on the StoreOnce appliance.

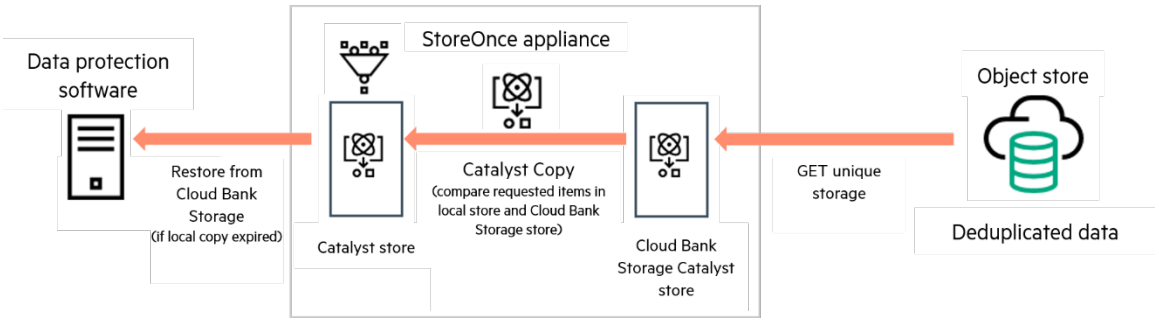

Figure 5. Cloud Bank Storage recovery process

In Figure 5, the data protection application requests a recovery from the Cloud Bank Storage Catalyst store. On the StoreOnce appliance, the metadata of the needed Catalyst items in the cloud is compared to the Catalyst items already stored locally in other Catalyst stores. Only the unique items in the cloud are retrieved and combined with the local items to produce the complete recovery. This significantly reduces the time needed to recover from the cloud. Another benefit of Cloud Bank Storage is reduced costs, because cloud providers typically charge premium prices to recover data from object stores.

Additional features of Cloud Bank Storage include:

- Detaching, which allows the user to switch a Cloud Bank Storage store to read-only mode. The store is preserved in the cloud but cannot be altered, protecting it against accidental deletion. Stores can be reattached later if needed. Detaching also allows “cold data,” or data that must be maintained but is rarely accessed, to be moved to lower-cost cloud storage. Detach functionality is enabled by a separate capacity-based license.

- Multireader functionality, which allows multiple StoreOnce systems to connect to a Cloud Bank Storage store in read-only mode, even if they do not have a Cloud Bank Storage license. This provides disaster recovery capability by simply using another StoreOnce appliance in a remote site. The user can create a new Cloud Bank Storage store on the remote StoreOnce system and link it to the object store. The store recognizes the new link and the cloud data is automatically available in read-only mode. This is also helpful for testing and verification purposes—data can be retrieved from the store by another StoreOnce appliance without impacting the primary backup configuration.

Catalyst integrated with a data protection application

Figure 6 details how Catalyst integrates with a data protection application. The application is configured to use a Catalyst store as a backup target. The store is created on the StoreOnce appliance and is presented to a backup server in the application configuration using either Ethernet or Fibre Channel. The backup server has the Catalyst client plug-in installed on it.

Note

Although Figure 6 shows backup data streaming from backup clients to a backup server, this is not typical of Catalyst database plug-in configurations. In most cases, the Catalyst database plug-in is installed on the database server (which would typically replace the backup server in Figure 6), and backups go directly from the database server to the StoreOnce system.

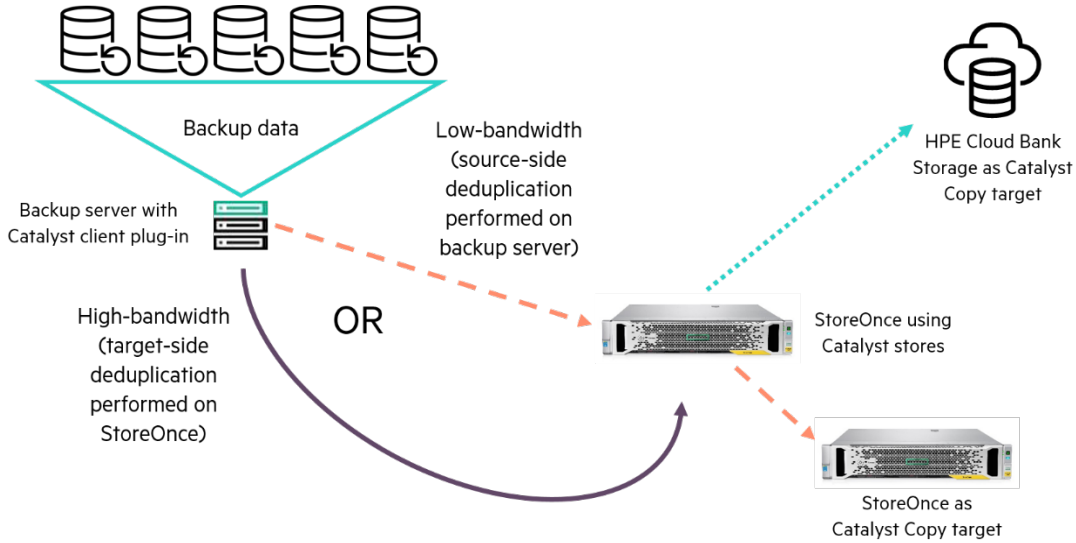

Figure 6. Catalyst integrated with data protection application

As stated previously, backup jobs written to Catalyst stores are low bandwidth by default, ensuring that backups finish quicker while saving significant storage space. However, there might be cases in which users do not want source-side deduplication. For example, if an Oracle administrator wants to use the Catalyst RMAN plug-in to back up databases, they might not want even the slight overhead caused by the low-bandwidth deduplication to occur on their production database server. Catalyst stores can be configured on StoreOnce to use high-bandwidth deduplication. In that instance, the Catalyst RMAN plug-in would send all of the backup data from the Oracle server to StoreOnce and allow the deduplication to occur on the appliance itself.

Catalyst integrated with HPE Recovery Manager Central

For those who want to incorporate snapshot-based data protection, HPE Recovery Manager Central (RMC) can create snapshot jobs of volumes on HPE 3PAR StoreServ storage systems and use Catalyst to store the snapshot data on a StoreOnce appliance. All RMC jobs are managed through the RMC appliance, a virtual machine that is created during the installation process. RMC appliances can be installed in a VMware or Microsoft Hyper-V® environment.

RMC data protection with Catalyst

RMC integrates with Catalyst using HPE RMC Express Protect, which allows data to be written from RMC snapshot jobs to a StoreOnce Catalyst store. These writes are always low-bandwidth Catalyst jobs, with the RMC appliance functioning as the "backup server" for handling the deduplication. Snapshots are mounted on the RMC appliance, which can communicate with a Catalyst store over either Ethernet or Fibre Channel. The snapshot data is then written to the Catalyst store in low-bandwidth mode.

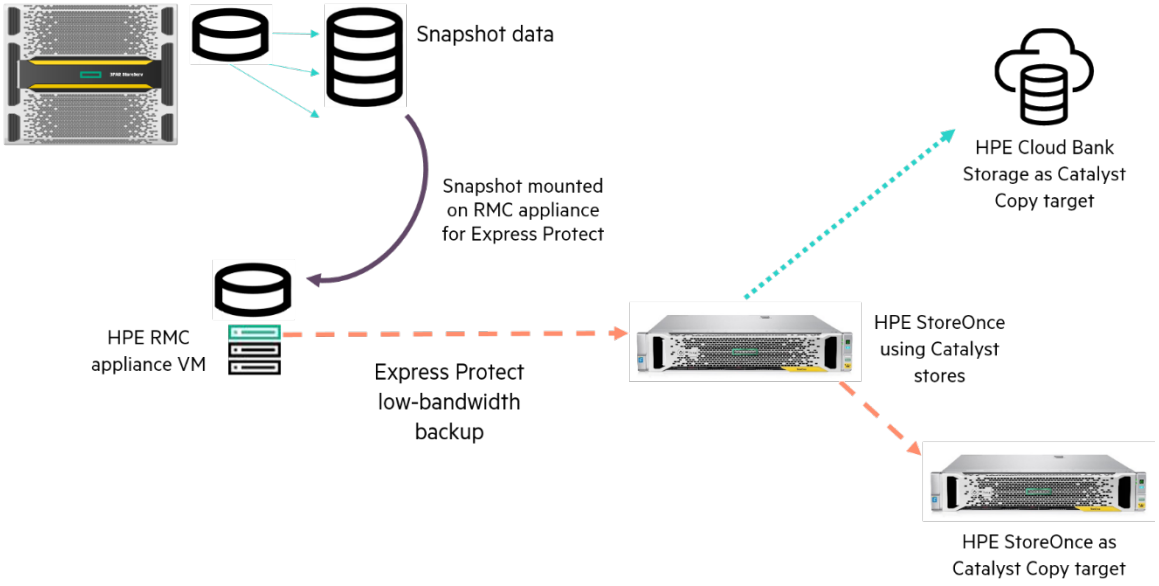

Figure 7. Catalyst integrated with RMC

After the snapshot data is backed up to the Catalyst store, RMC can also be configured to create Catalyst Copy jobs to copy backup data to another StoreOnce system or to write data to Cloud Bank Storage.

RMC recovery with Catalyst

The HPE RMC Express Restore feature offers RMC users the option to recover faster from an older Express Protect job if the RMC snapshot for that recovery point is no longer available. With Express Restore, RMC compares the blocks between the older Express Protect job and the most recent one. It then uses the snapshot associated with the newer Express Protect and creates a "child" read/write snapshot, swapping out the unique blocks from the older job. The new snapshot can then be mounted on the backup client to recover files. Express Restore enables much faster recovery because only the unique blocks have to be recovered from the Catalyst store.

Conclusion

HPE StoreOnce Catalyst was created to simplify data protection, improve backup performance while saving backup storage space, and meet the requirements for properly protecting customer data, as outlined in the 3-2-1 protection strategy. Catalyst reduces administration tasks by allowing the data protection application to manage Catalyst backup, copy, and recovery operations. Catalyst database plug-ins let database administrators be in charge of their own backups, using the native backup functionality they understand. Catalyst low-bandwidth deduplication reduces the amount of data sent over the network in a backup job, which results in faster backups. Catalyst Copy and HPE Cloud Bank Storage give users multiple options for creating copies of their backups for proper redundancy and disaster recovery capability. StoreOnce Catalyst adds simplicity, flexibility, and high performance to a data protection environment.

Resources

StoreOnce Support Matrix
hpe.com/Storage/StoreOnceSupportMatrix

Protecting Data from Ransomware with HPE StoreOnce Catalyst
h20195.www2.hpe.com/V2/GetDocument.aspx?docname=a00042003enw

Learn more at HPE StoreOnce Data Protection Backup Appliances

hpe.com/us/en/storage/storeonce.html

Sign up for updates

a00045521ENW, April 2018