HPE Data Protector and EMC Data Domain Boost integration

Hewlett Packard Enterprise

Table of contents

The drivers of data growth

Enterprises large and small are embracing technology wherever possible as a catalyst for the creation of new opportunities. While the driving factor at one time was maintaining a competitive edge and consolidating data center infrastructure, today the adoption of new technology is the catalyst for capturing new markets, creating new products, and launching other areas of revenue growth. Regardless of how and why you bring a new technology comes on board, the result is typically the same: more information is created that has to be categorized based on its importance, and then aligned with a suitable backup and recovery process to manage data loss and disaster recovery.

As the rate of technology adoption continues, it deepens our dependency on digital data, which in turn drives unabated data growth. This never-ending cycle results in the need for more storage, greater management and oversight, faster throughput, more stringent governance, and immediate access, as well as having a loss-mitigation strategy. These challenges are amplified by a mobile workforce that quickly adopts new technologies, creating more responsibility for IT professionals and impeding their abilities to deliver revenue-generating solutions to the business.

The need for 24x7 access on a global scale is another challenge in the modern datacenter, requiring the scale and elastic infrastructure necessary to address continually changing user and market demands and provide immediate access to any and all organizational data at any time. The ability to meet these end user expectations can set an organization apart from its competition.

With each item on this long list of challenges and requirements, backup and recovery (a core competency of IT) often competes for physical and human resources, and is often the heaviest casualty of unabated data growth.

How data growth complicates backup and recovery

Managing your information can be more complicated, just due to the sheer rate of growth—doubling every 12 to 18 months. This drives CIOs to continually look for infrastructure features that can minimize data growth, via storage arrays, networking infrastructure, I/O accelerators, etc. And while these approaches have their merit, they often result in pushing the bottleneck to different locations within the datacenter, where investments and advancements have not been applied.

Identifying backup and recovery solutions that can simplify the data protection process and unburden the IT organization can be a challenge, particularly given that data grows and vacillates in stages of importance, relevance, and criticality. For this reason, it is key that IT departments choose the right infrastructure and the right backup and recovery strategy, especially when data exists at scale. At the end of the day, the IT department remains responsible for ensuring the survivability of their organization's information from any kind of event such as disasters, corruption, or system failures (logical, physical and/or virtual). And with so much data to be considered, growing at an unabated rate, one common solution that is often chosen is deduplication.

The effects of deduplication on organizational data

Deduplication is quickly becoming one of the most important storage optimization techniques to combat data growth, and is the process of examining and comparing data that is already being stored to determine if it is identical. Specialized deduplication engines are often implemented based on mathematical algorithms that are unique for each vendor. In practice, the algorithm is applied to data segments and results in what are commonly referred to as a hash values or hash codes. When the comparison of two data segments result in the same hash code, they are determined to be identical and the engine replaces the duplicate with metadata that links it to the original. By processing and comparing chunks of data to isolate and store only the unique segments, deduplication can reduce infrastructure demands by as much as 80 percent, depending upon the workload and structure of the data.

Combating data growth

Keeping data volumes at bay is both an active and time-consuming initiative. Deduplication certainly offloads and automates aspects of this initiative and speaks to why the adoption of deduplication has become so pervasive. But the pervasiveness is not limited to adoption, nor is it a feature/function limited to the primary storage array upon which the organizational data resides. The need for deduplication extends to all places where data resides. More importantly, it adds efficiency to the backup and recovery strategies that are put into place.

Boost backup and recovery with deduplication

Deduplication processes offer a powerful rung in the backup and recovery ladder that allows you to increase the speed of your backup process by reducing the amount of storage space required for both the production data and its related backup sets. By storing a single set of unique data with logical pointers to and from all other duplications of the same data, the overall capacity of the backup infrastructure, is optimized.

Another benefit of deduplication is the impact to the recovery process. The use of pointers to the unique data does not change the restore points. In other words, every point from which the data can be restored simply references the unique data extending the restore process and reducing errors associated with manually managing the way the backup target media handles the deduplication of data.

Backup and recovery strategies based on deduplication-enabled disk-to-disk (D2D) storage devices are becoming the preferred method in both core IT data centers and associated remote, or "edge," locations. Three factors driving this trend:

• The same data growth effecting the core data center effects edge locations, and the network connections between them cannot always address sending the complete data set with the production-level traffic also in flight.

• The total cost of ownership (TCO) gained from recovery efforts when using deduplication-enabled D2D systems are significantly lower than with traditional tape systems.

• The capacity and cost-savings associated with retaining the backup sets on a deduplicated target are much higher than with traditional disk backup solutions.

The best possible solution in this scenario is to deploy IT infrastructure with components that take full advantage of deduplication—at the primary storage array, in the backup solution, and with the backup target. **HPE's Data Protector software** offers deduplication as a core value of the product, allowing you to choose where the deduplication occurs; either at the edge where the data is created, at backup servers, or completely offloaded to the D2D target.

As organizations continue to ingest multi-vendor solutions through acquisitions, organizational policy, hardware reuse, etc., having support for a heterogeneous storage environment at the backup software layer is key.

A heterogeneous approach

Through the integration of **HPE Data Protector** and EMC Data Domain system, you can implement a heterogeneous approach and use deduplication in your backup and recovery process. EMC Data Domain is a market leader in D2D storage that offers a high-speed deduplication engine known as Data Domain (DD) Boost. By extending Data Protector via integration with the DD Boost, your backup solution can use a common deduplication algorithm at the edge, at the backup server, or target D2D device when using the Data Domain system. The solution allows you to achieve a consistent and faster backup and recovery experience, regardless of the application or data tier.

About EMC Data Domain Boost

EMC Data Domain Boost™ software significantly increases performance, reduces resource requirements, and simplifies management. With DD Boost, parts of the deduplication process are distributed to the backup server or application clients, enabling it to compress and send only unique data segments to a Data Domain system. This enables 50 percent faster backups and reduces network bandwidth requirements by as much as 80-99 percent.

In addition, DD Boost provides advanced load balancing and failover, which further improves throughput and resiliency. With advanced load balancing and failover, you can aggregate multiple transport links on the Data Domain system into a single group so multiple links appear as one to the backup application. The Data Domain system then transparently balances the backup load between links in the group. In addition, the automatic link failover mechanism keeps backup systems operational in case of temporary network glitches.

Data Domain Boost provides advanced integration with the majority of the backup application market and industry leading enterprise applications. For the latest updates on the product please check out **emc.com/data-protection/data-domain.htm**.

About HPE Data Protector

HPE Data Protector enables you to seamlessly protect and harness data from edge of the network to the datacenter and the disaster recovery site—across physical, virtual, and cloud environments. Designed to support the largest and most heterogeneous enterprise environments, a single Data Protector instance can protect thousands of clients and hundreds of billions of files, and run 100,000 sessions per day.

With unique capabilities such as integrated cloud backup, enhanced automated disaster recovery, zero downtime backup, instant recovery, and down-to-the-second application recovery, HPE Data Protector allows you to increase storage efficiency and backup and recovery performance, instantly recover your mission-critical applications, and turn your backup data repository into an information advantage. From a single pane of glass, HPE Data Protector allows you to fully leverage your data asset investment across premise, hybrid, and cloud infrastructure to reduce backup and recovery windows and increase business application performance.

Simple yet powerful snapshot functionality frees you from restrictive backup windows while instant recovery capability enables flexible down-to-the-second recovery through a simple-to-use graphical user interface. One touch setup for off-site backup on the cloud protects data on one of the world's largest and most secure private cloud providing enterprise-class edge-to-cloud protection while reducing CAPEX and OPEX.

Data Protector is the key component of the HPE Adaptive Backup and Recovery (ABR) suite which leverages operational analytics and data insight to enable high performance and tuned backup environment for today's dynamic data-driven enterprise. The ABR suite also includes HPE Storage Optimizer for analyzing, classifying and managing data based on its value and HPE Backup Navigator for optimizing the backup environment by using operational analytics and insight to identify protection gaps, run rapid root-cause analysis and allow IT to plan for future storage resources. Together, this suite provides our customers with a 360 degree view of their backup environments, allowing them to constantly tune and adapt to deliver peak results.

Industry collaboration boosts customer satisfaction

Organizational IT leadership typically base their solution deployment decisions on a number of criteria, one of which is the degree of flexibility that a chosen solution may offer the existing infrastructure. The goal is to make the solution work with and for the infrastructure, not the reverse. As a core IT function, HPE Data Protector helps you coordinate and orchestrate your backup and recovery operations, from how, to where, to what performance features you use.

A key question for many IT professionals is how to minimize the amount of data backed up and transferred over the network. HPE Data Protector software provides its own native deduplication algorithm; the same algorithm used in the **HPE StoreOnce** device. While the deduplication feature of Data Protector addresses most user requirements, there are some instances where the use of existing infrastructure features is more advantageous to the customer. By working with industry partners, such as EMC, HPE is able to remain committed to addressing market challenges with solutions that address critical challenges. We can also enable you to continue running in heterogeneous environments, with integration and support for the leaders in the data center ecosystem.

HPE Data Protector's integration with DD Boost technology enables customers to solve a market challenge by offering more choices to solve the backup and recovery requirements of their organization. The integration allows users to specify a Data Domain target when setting up the backup configuration and also to determine where the deduplication could occur.

Using Data Domain's deduplication algorithm, available via DD Boost, HPE Data Protector supports deduplication in three locations: client-side, backup server, or target device. Each of these, along with their advantages and considerations, are explained below:

Client-side deduplication

The term "client-side" is a form of deduplication that is enabled at the point the data is created and before it is transmitted over the network to the target backup device. With EMC DD Boost, HPE Data Protector can be configured for client-side deduplication to make sure that only unique data is used during backup. An example of this implementation can be seen in Figure 1.
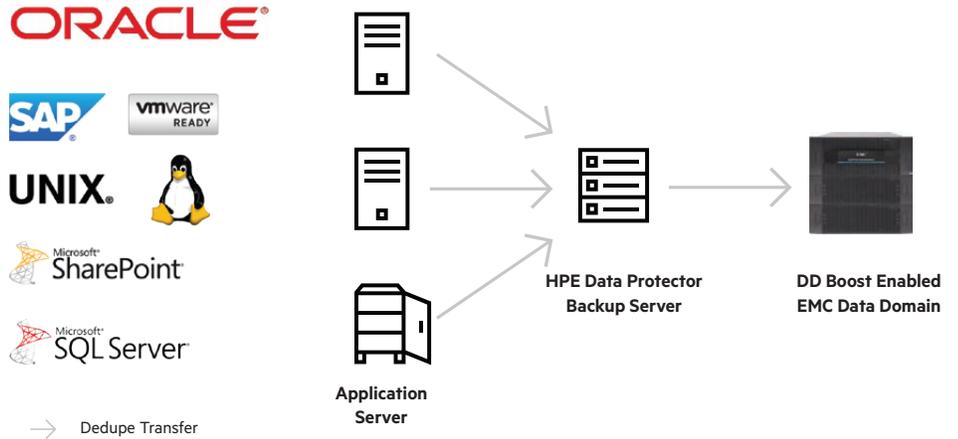

Figure 1. Client-side deduplication implementation

Features

• Data deduplication at the point of data creation via a DD Boost-enabled HPE Data Protector media agent installed on the application server

• Works in conjunction with EMC Data Domain Boost's distributed segment processing, which distributes the deduplication process between the DD Boost-enabled media agent and the Data Domain system

• Client-side deduplication is supported with Microsoft Windows and HP-UX operating systems

• HPE Data Protector's implementation results in single-click configuration without device selection per backup client

• HPE Data Protector can leverage EMC Data Domain Boost as a high performance deduplication backup target

Benefits

• Higher efficiency in network utilization by only sending unique data to the Data Domain system

• Potential reduction in the time required for the backup, especially on high latency and limited bandwidth connections between core data centers and remote locations

• Improved backup and replication Service Level Agreement (SLA) definitions based on lower network traffic and predictable backup sets at the point at which the data is created

• Same deduplication algorithm as the target, via DD Boost, results in a standardized approach tailored to the specific backup needs and infrastructure capabilities

Backup Server deduplication

When bandwidth considerations are out-weighed by the need for all resources to support the workload of the client-side system, the backup server can handle the deduplication process. In this architecture, the raw backup set is transmitted from the source to the backup server where the DD Boost-enabled Data Protector media agent is running. On the backup server , the Data Domain deduplication algorithm is used to verify that only unique data is transferred to the Data Domain system. An example of this implementation can be seen in Figure 2.
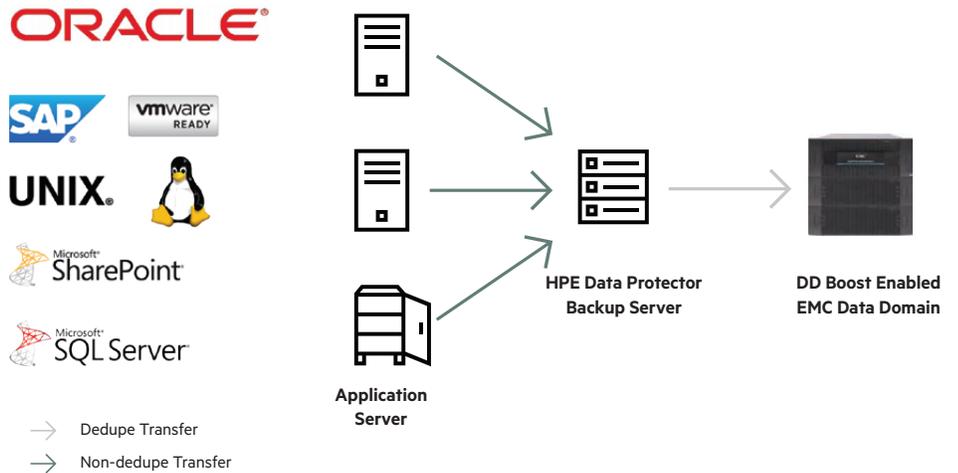

Figure 2. Backup server deduplication implementation.

Features

- Data deduplication at the point where the backup process is managed via a DD Boost-enabled HPE Data Protector media agent installed on the backup server

- Works in conjunction with EMC Data Domain Boost distributed segment processing, which distributes the deduplication process between the DD Boost-enabled media agent and the Data Domain system

- Backup server deduplication preserves the physical resources at the source for resource intensive workloads

- HPE Data Protector can leverage EMC Data Domain Boost as a high performance deduplication backup target

Benefits

- Higher efficiency in network utilization by only sending unique data to the Data Domain system

- Potential reduction in the time required for the backup, especially on high latency and limited bandwidth connections between the backup server and the target backup device

- Improved performance of the client-side infrastructure where resources are allocated for the production workload and deduplication is offloaded to the backup server

- Same deduplication algorithm as the target, via DD Boost, results in a standardized approach tailored to the specific backup needs and infrastructure capabilities

Target-side deduplication

When bandwidth considerations are out-weighed by the need for all resources to support both the client-side system workload and the backup server handling competing workload demands, the deduplication process can be offloaded to EMC Data Domain, the target device. In this architecture, the raw backup set is transmitted from the source to the backup server, and from backup server to the Data Domain system, without consideration for duplicate data in the backup set. On the target device, the same Data Domain deduplication algorithm is used to make sure only the unique data is stored in the system. An example of this implementation can be seen in Figure 3.
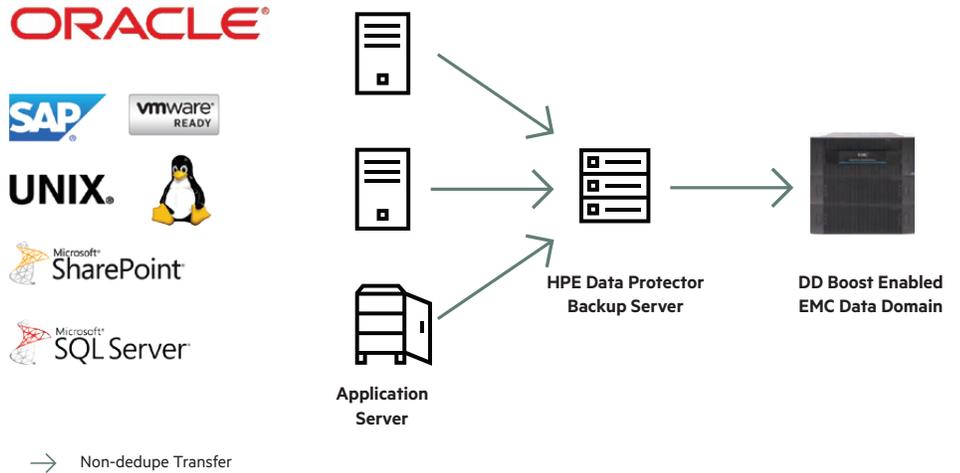

Figure 3. Target-side deduplication implementation

Features

- Inline deduplication at the point where the backup set is stored and managed as part of the data protection life cycle

- Target-side deduplication preserves the physical resources at the source and the backup server for resource intensive parallel workloads

Benefits

- The solution provides higher efficiency in the resource utilization of the backup systems by offloading the deduplication effort to the Data Domain system

- Improved performance of the client-side infrastructure where resources are allocated for the production workload

Data Protector Managed Replication

Increasing frequency of catastrophic events like hurricanes, floods, fire, etc. have raised the urgency to have disaster recovery (DR) procedures. One of the most crucial steps for DR is to have a copy of the data at a remote site. To improve reliability of disaster recovery and meet stringent recovery time objectives (RTO) imposed by the business, organizations are increasingly replicating backups to create this offsite copy of their critical data. Reducing the amount of backup and archive data replicated through deduplication and compression reduces the network bandwidth required, and makes replication over existing networks economically viable.
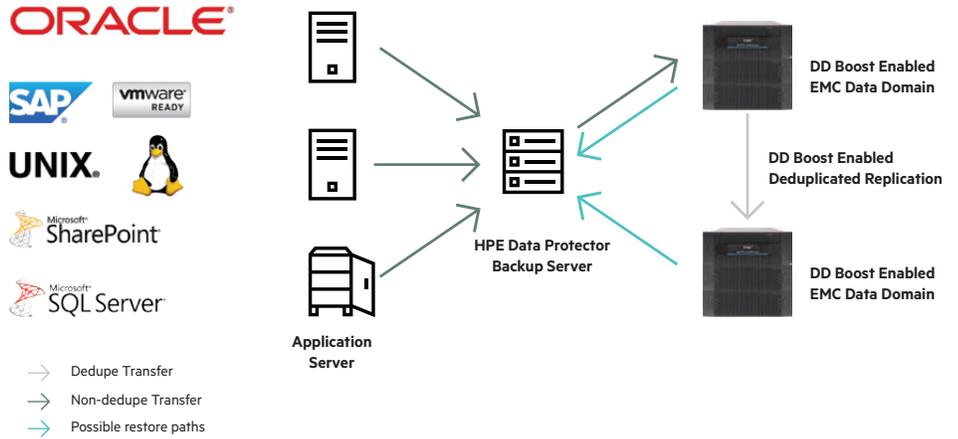

Figure 4. Data Protector managed replication

Features

- Managed file replication using DD Boost allows the backup software to control the replication on a per file basis. When integrated with DD Boost, the backup software's users can configure policies to selectively replicate the individual backup image or dataset to another system after completion of the backup

- Managed file replication directly transfers a backup image from one Data Domain system to another, one at a time on request from the backup software

- The backup software keeps track of copies, allowing easy monitoring of replication status and recovery from multiple copies

Benefits

- Backups are protected from Data Domain failure at the primary location

- Backups can be restored from either the primary or replicated backup target

Data Protector Automated Replication Synchronization

To protect against full site loss at a location, Data Protector provides Automated Replication Synchronization utilizing Data Domain Boost Replication. This allows a backup server at a DR location to be immediately aware of backups that have been replicated to that location and to be capable of restoring those backups in the event of the complete failure of the original site.
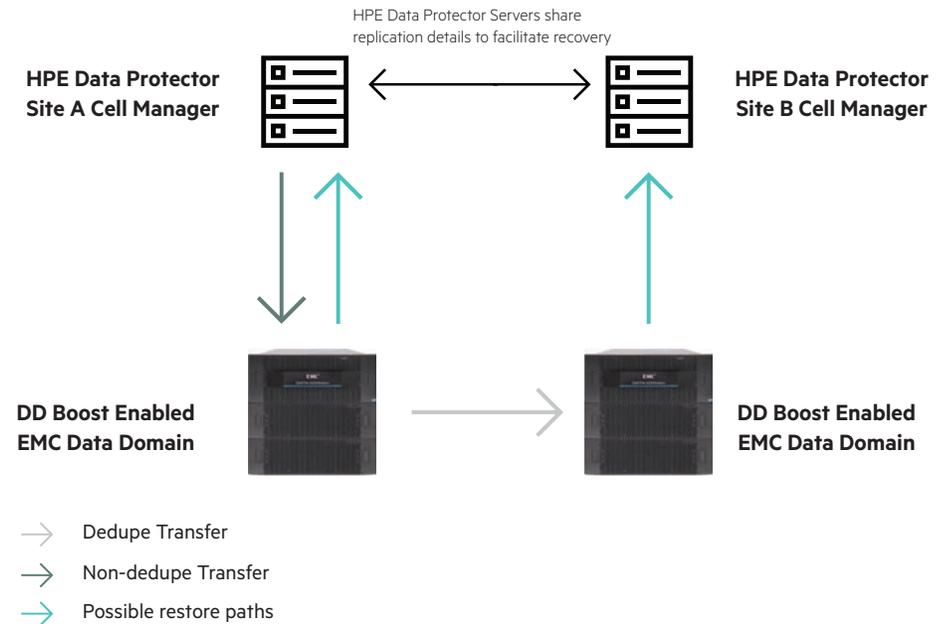

Figure 5. Data Protector automated replication synchronization

Features

• Data protector servers share replication information to facilitate recovery

• Managed file replication directly transfers a backup image from one Data Domain system to another, one at a time on request from the backup software

• The backup software keeps track of copies, allowing easy monitoring of replication status and recovery from multiple copies

Benefits

• Backups are protected from full site failure

• Disaster recovery is easily implemented

• Business continuance in disaster situations

Summary

HPE and EMC have teamed up to bring together **HPE Data Protector** and EMC Data Domain Boost to enable you to deploy a common deduplication approach at the source, the backup server, or the target when combined with an EMC Data Domain system. With this approach, you experience both flexibility in choosing your backup infrastructure and should see dramatic improvements in the performance and reliability of the backup and recovery process, with operational and disaster recovery strategies. In addition, by extending HPE Data Protector with EMC DD Boost, you benefit from a best of breed multi-vendor data protection experience that delivers an advanced approach to tackle the toughest backup and recovery challenges— protecting a diverse, highly dynamic, and heterogeneous data center environment.

Adopting this combined solution provides the following benefits:

- The ability to connect HPE Data Protector and an EMC Data Domain system to create a backup and recovery infrastructure based on a common deduplication engine
- The choice as to where the deduplication occurs: client-side, backup server, or target device
- The ability to leverage HPE Data Protector's object consolidation and object copy features to further preserve data stored on Data Domain systems

Please visit **hpe.com/software/dataprotector** to learn more.

Hewlett Packard Enterprise