DELL EMC PROTECTPOINT TECHNOLOGY

A Detailed Review

ABSTRACT

This white paper explains how Dell EMC is solving the challenge of ever tightening protection SLAs against the backdrop of exponentially growing data sets. Dell EMC ProtectPoint technology is designed to provide faster, more efficient full backups while eliminating the impact on application servers and reducing overall cost and complexity.

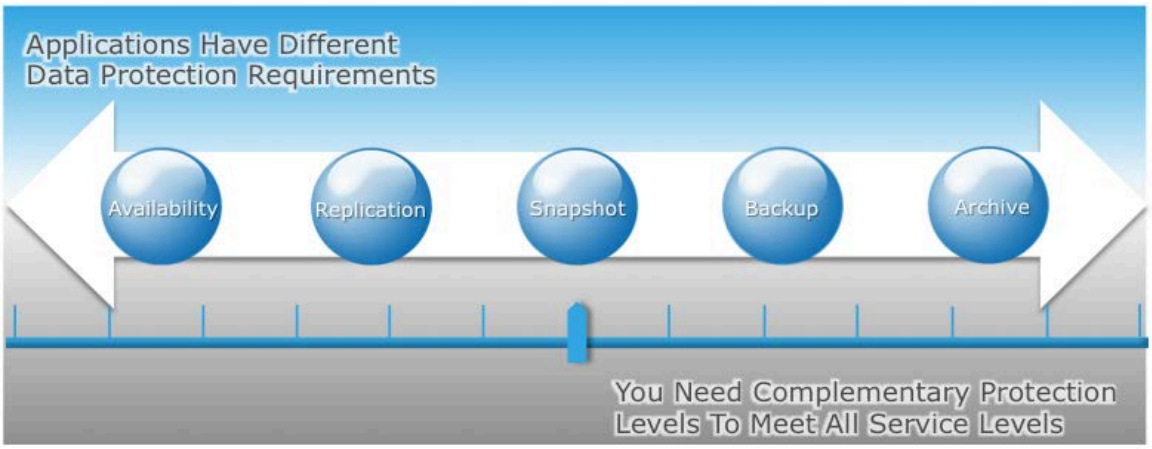
By integrating primary storage and industry leading protection storage (Dell EMC Data Domain deduplication storage systems), ProtectPoint technology eliminates the need for traditional backup applications while still providing the benefits of native full backups. ProtectPoint provides the best of both worlds – the performance of snapshots with the functionality of backups – allowing for non-intrusive data protection.

May 2017

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Data protection can do more than reduce risk and costs. It can drive business agility by enabling you to spend less time worrying whether you can reliably recover your data and more time adding business value. But first, you need the confidence that all of your data is truly protected. Nowhere is the need for protection more crucial than in the protection of mission-critical and high performance applications and their data. Many traditional data protection solutions rely on a one-size-fits-all approach or focus on solving one problem and ignore the others, which leaves gaps in protection. In comparison, Dell EMC delivers a data protection continuum of solutions (FIGURE 1) from availability to archive, which ensures you have the right level of protection for all your data.



The Data Protection Continuum

Many applications are required to be fully operational 24x7x365 and the data for these applications continues to grow. Some applications have particularly challenging performance requirements. At the same time, their RPO and RTO requirements are becoming more stringent. As a result, there is a large gap between the requirement for fast and efficient protection and the ability to meet this requirement without disruption. Traditional backup is unable to meet this requirement, which has led many datacenters to use snapshots for more efficient protection. Unfortunately, when snapshots are used where backups are required, users may lack the protection and functionality required and introduce unnecessary complexity.

EMC ProtectPoint directly addresses this gap in the protection continuum by integrating best in class Dell EMC products to give the performance of snapshots with the functionality of backups. ProtectPoint provides no-compromise data protection by giving the application owner complete control of protection without impacting their application performance.

Audience

The intended audience for this white paper is application, storage and data protection leaders and architects struggling to meet stringent protection requirements for mission critical and high performance applications.

The Application Protection Challenge

The Dell EMC data protection continuum is designed to meet a range of protection requirements across the enterprise. The same application and data often needs protecting from multiple points on the spectrum. However, the data protection continuum fundamentally provides two different classes of solution – availability and protection that are designed to meet different requirements and are therefore implemented and deployed differently:

- Availability solutions (specifically availability, replication and snapshots on the left half of the continuum) are designed to allow short-term recovery in the case of significant specific failures that impact a server, storage array, part or all of a datacenter. Such failures typically need to be responded to with pure application level recovery. In these cases the application needs to be able to collect the replicated data from the primary storage device, and do this without the application itself getting hindered. Only keeping copies online and both on the original primary and a second primary storage array allow instant pickup of an active workload or near instant failover. This also lends itself well to on box snapshot with no detailed catalogue and only a limited history and synchronous or asynchronous replication with no history at all.

- Protection solutions (specifically backup on the right side of the continuum) are designed to protect data long-term against loss due to corruption or damage in other ways, either by software errors, viruses, or user errors. Such events usually entail going back in time hours, days, or weeks to recover the correct data. Similarly, looking back at data retained for analysis also implies longer timescales. In both cases off box off-line protected catalogues and indexes are critical. The protection aspect drives the need to move protected data off of primary storage to protection storage, and the nature of the recoveries means more information about what has been protected is critical to carry out the necessary recoveries.

Therefore, datacenters require various deployment models (FIGURE 2) to meet all availability and protection requirements. In this model the availability function operates within and between primary storage arrays and is typically run by the storage administrator. The protection function operates from and between primary storage and protection storage, and is run by the application administrator with help from the protection administrator. While there is a natural desire to avoid duplication of efforts and resources, the different requirements for each means there is limited value in combining them at this time.

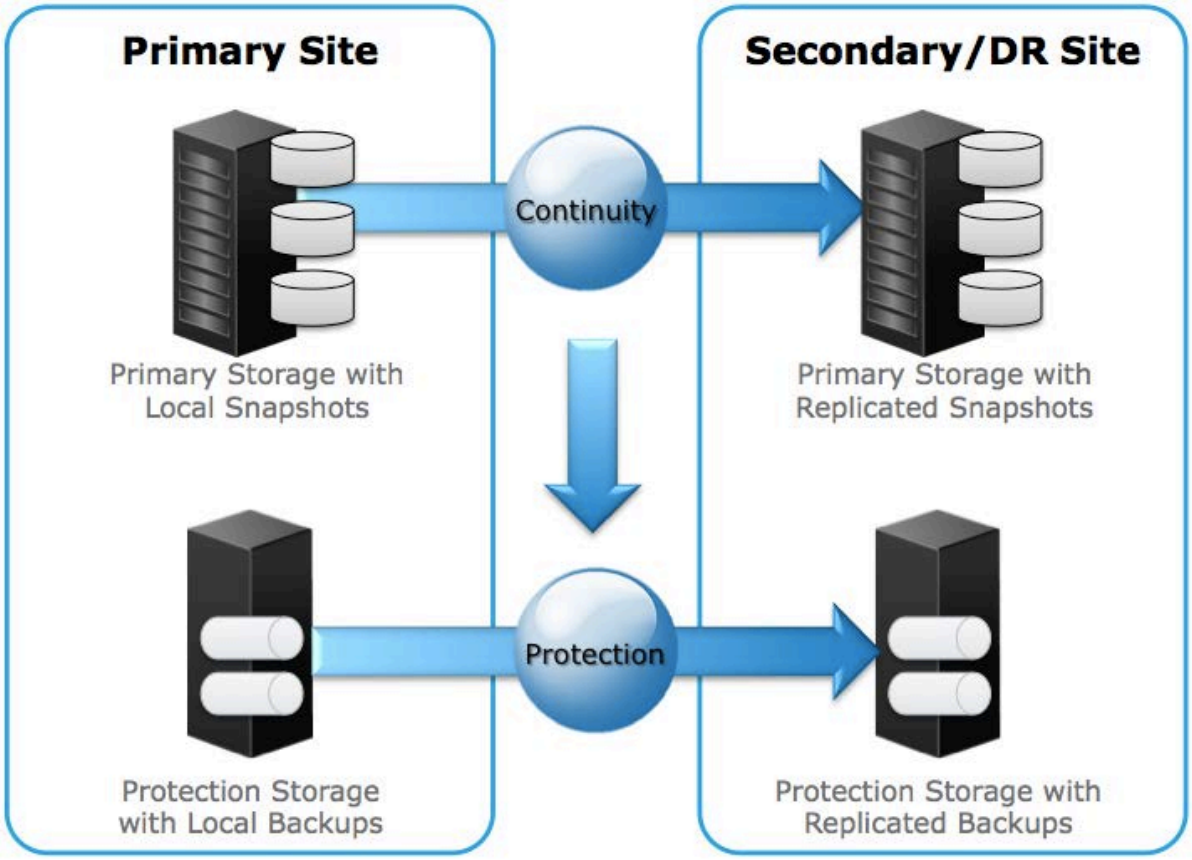


Figure 1. Availability and Protection across Data Centers

Unfortunately today, availability and protection solutions still don't meet the requirements for protecting mission critical applications. With traditional backup, as these environments continue to grow, many enterprises have implemented daily incremental backups with a weekly full backup to reduce the load on the application. Unfortunately the benefit of incremental backups only applies for 6 out of 7 days and comes at the cost of more complicated recovery processes – re-assembling incremental and full backups. Synthetic full backups or incremental forever backups have attempted to improve this, but they come at the cost of additional pressure on the infrastructure at either the time of backup or recovery. Traditional backup approaches impact the application for the duration that it needs to be held in backup mode – both due to the data movement and the backlog of data in the transaction logs of the application. When there is a backup window where an application could be offline or online, but with degraded performance, such approaches were acceptable – even if not ideal.

In addition to unrelenting data growth pushing the limitations of traditional backup, applications increasingly need to be online and fully operational 24x7x365, which makes protecting them even more challenging. Specifically, the always on requirement means the disappearance of the protection window because application performance can no longer afford to be impacted for protection. This requirement is particularly challenging with applications which by their nature have high performance requirements. This has led to a new requirement in the protection space for non-intrusive data protection.

The challenge of performing backups without impacting the application has drawn attention back to the benefits of availability technologies - specifically, the ability for primary storage snapshot, replication, and continuous data protection (CDP) solutions to meet stringent RPO and RTO requirements without application impact. With the pressures on protecting applications, it can be tempting to replace traditional backup with such availability technologies to meet these stringent requirements. However, such an approach comes with a critical and an unacceptable set of limitations when applied to protection:

- Scale: typically limited to 250 snapshots/copies
- Storage efficiency: limited to not duplicating blocks within one snapshot chain, no compression, no global deduplication
- Reliability: corruption cascades through the snapshot set
- Replication: numerous challenges when combining snapshots with replication including replicated corruptions
- Recovery: Complex recovery processes with limited or no native application integration

Due to the limitations, to meet full protection requirements for mission critical applications, snapshots are often blended with less frequent but still impactful traditional backups. Such approaches bring complexities and still don't address the 24x7x365 issue as regular full traditional backups are still present. The real requirement is to meet protection requirements (cost effective retention, corruption protection, and application integration) with the availability benefits (performance and minimal application impact) with a single data protection solution.

ProtectPoint technology: Storage Integrated Data Protection

ProtectPoint technology provides storage-integrated data protection that complements existing Dell EMC data protection and availability solutions and demonstrates the latest proof point of the protection storage architecture. Data proteciton is evolving (FIGURE 2) from the wave of traditional server-centric backup designed around centralized control with backup servers and agents sending data to protection storage. The new wave – modern data management – is designed for self service to empower application owners and storage admins to backup directly from the data source to protection storage. This model is becoming increasingly popular because it puts control directly in the hands of the data owners – rather than relying on a centralized backup team. However, the new self service model requires intelligent oversight on copies and service levels across the business to ensure that data is protected and you maintain consistent service levels across the business. Finally – to tie it all together, both you need intelligent analytics and search that can cross both the traditional and modern approaches ensuring that you optimize operations across all data centers globally.

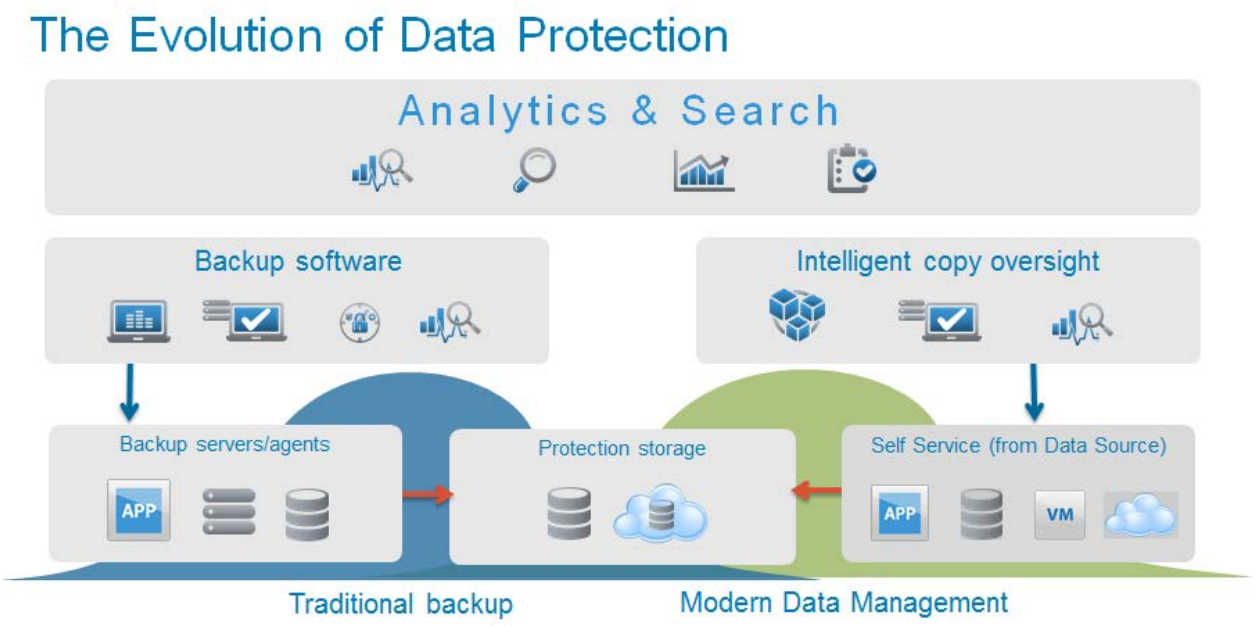


Figure 2. The Evolution of Data Protection

Specifically, ProtectPoint addresses two aspects of data source integration - the integration with primary storage and applications. ProtectPoint is an industry first solution that protects data by copying it directly from its source (primary storage) to the protection storage via the most efficient path and without application impact. To achieve this, ProtectPoint leverages key technologies within the

primary storage and protection storage and introduces new protection software. This protection software is a data protection agent that drives the backup process and supports integration with the application being protected. This agent also enables the application administrator to control his own backup and recovery operations.

ProtectPoint is neither adding application integration to snapshots nor adding snapshot support to backup software. Both of these approaches would bring some benefits, but would not fully address the problems with backup and would inevitably experience many of the limitations of snapshots. Rather, ProtectPoint was designed by decoupling the data plane from the control plane to directly drive the underlying capabilities on the primary and protection storage.

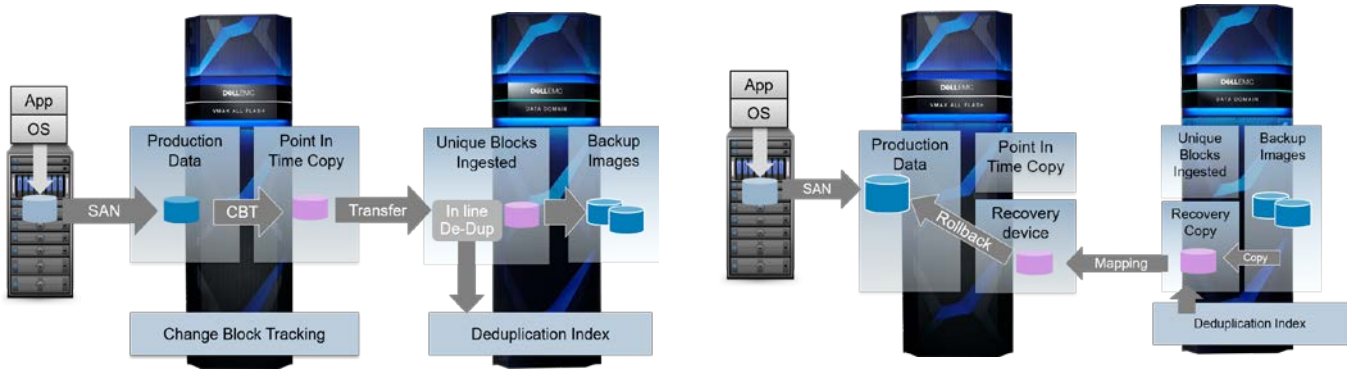
To explore this further, it's helpful to examine ProtectPoint through the lens of the data plane, the control plane, and the management plane.

THE DATA PLANE

The data plane carries the data from source to destination. With ProtectPoint, the data plane (FIGURE 4) is the connection between primary storage to the Data Domain system. Since ProtectPoint leverages primary storage change block tracking technology, it minimizes data sent on the data plane. When a backup is triggered, unlike a traditional backup application, the primary storage knows exactly what has changed since the last backup and only has to send those unique blocks across the network. The direct data movement from primary to protection storage eliminates the local area network impact by isolating all data traffic to the SAN. In addition, unlike other backup mechanisms that consume valuable host side resources on the primary storage, ProtectPoint data movement is handled by separate resources of the primary storage that are dedicated to protection workflows.

ProtectPoint technology is very different from snap and replication solutions thanks to the efficient way the data is processed and stored by the protection storage system. One of the benefits of leveraging Data Domain protection storage is its industry leading inline deduplication technology. When the Data Domain system receives the changed blocks from the primary storage, it will segment the incoming data stream and uniquely identify each segments and compares each segments to all previously stored data to determine if it's unique. If the segment is unique, it's compressed inline and stored on the Data Domain system. However, if the segment is not unique, the system will simply use a pointer and will not store the segment again. After the data is ingested and deduplicated, the Data Domain system then creates a new full independent backup image. This backup image is independent from all previous backups, but is deduplicated against all other known data, which enables 10 to 30x reduction in storage requirements, but still enables simple recovery. In addition, as with all data on a Data Domain system, ProtectPoint backups are protected against data integrity issues by the Data Domain Data Invulnerability Architecture with continuous fault detection and self-healing, which ensures data remains recoverable throughout its lifecycles on the Data Domain system.

Depending on the Primary Storage technology in question, ProtectPoint uses one of two approaches for the Data Plane. Both approaches are carefully architected to ensure best in class protection. In the case of ProtectPoints' support of VMAX^3 or VMAX All Flash, ProtectPoint leverages the underlying technology of SnapVX and Fast.X. With the support of XtremIO, ProtectPoint leverages the underlying technology of RecoverPoint. In Both cases, the choice of integration is intended to ensure the simplest overall deployment of end to end protection for the application. The same underlying technology is used to support both the availability and protection segments of the continuum.



Backup Data Flow VMAX^3 with Data Domain

Restore Data Flow VMAX^3 with Data Domain

Figure 3. Backup and Restore workflows with VMAX3

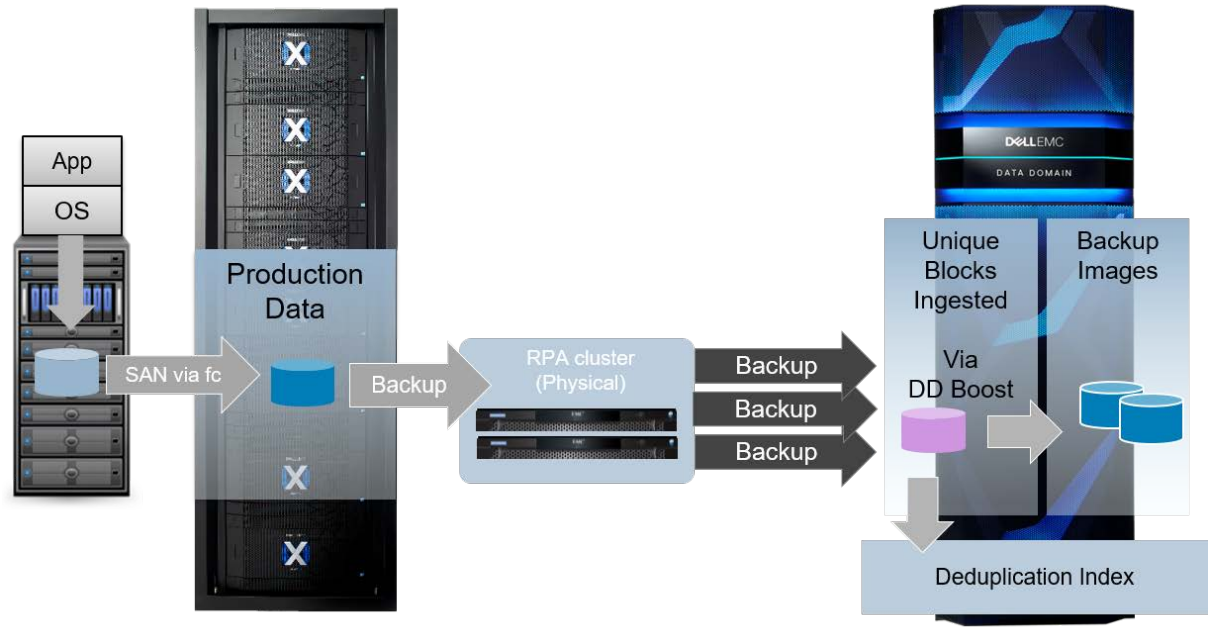


Figure 4. ProtectPoint Data Plane for XtremIO

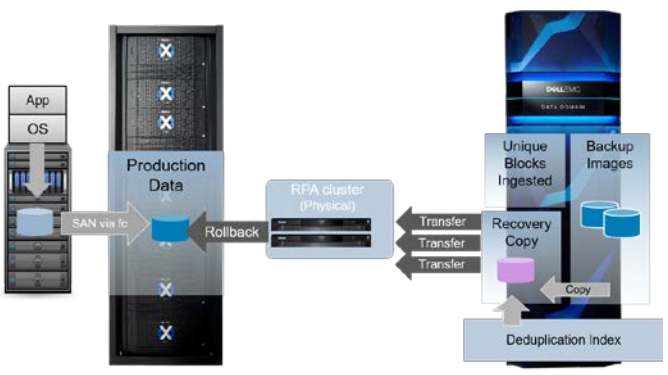
THE CONTROL PLANE

While the data plane carries out the data movement and processing, the control plane coordinates each of the steps along with other related activities.

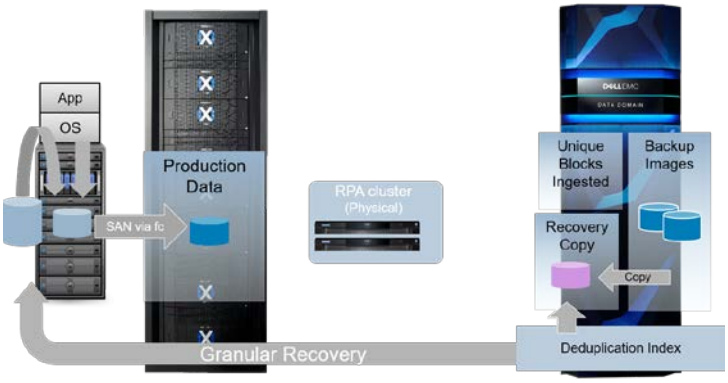
With ProtectPoint, the control plane is governed by two key functions within the ProtectPoint agent that runs on the application server being protected. First, is the application layer that supports or controls the application and file system integration. Second is the ProtectPoint controller, which controls the processes described in the data plane section above.

The control plane carries out the sequencing that provides one of the most critical benefits of the ProtectPoint – eliminating the backup impact on the application being protected, and provides 20x faster backup than a traditional backup solution. With ProtectPoint, the application only needs to be in backup mode for the moment that the backup is triggered, which is just the time it takes to create a point-in-time bookmark. Since this is a fairly simple process and decoupled from the actual movement of data, the application only needs to be in backup mode for a brief instant. For mission-critical applications, this is vitally important, as the longer the application is in backup mode the more IOs are queued in the logs and the heavier the impact on the application when exiting backup mode.

The control plane coordinates all the activities for the backup and recovery (full and granular) workflows. To control backup



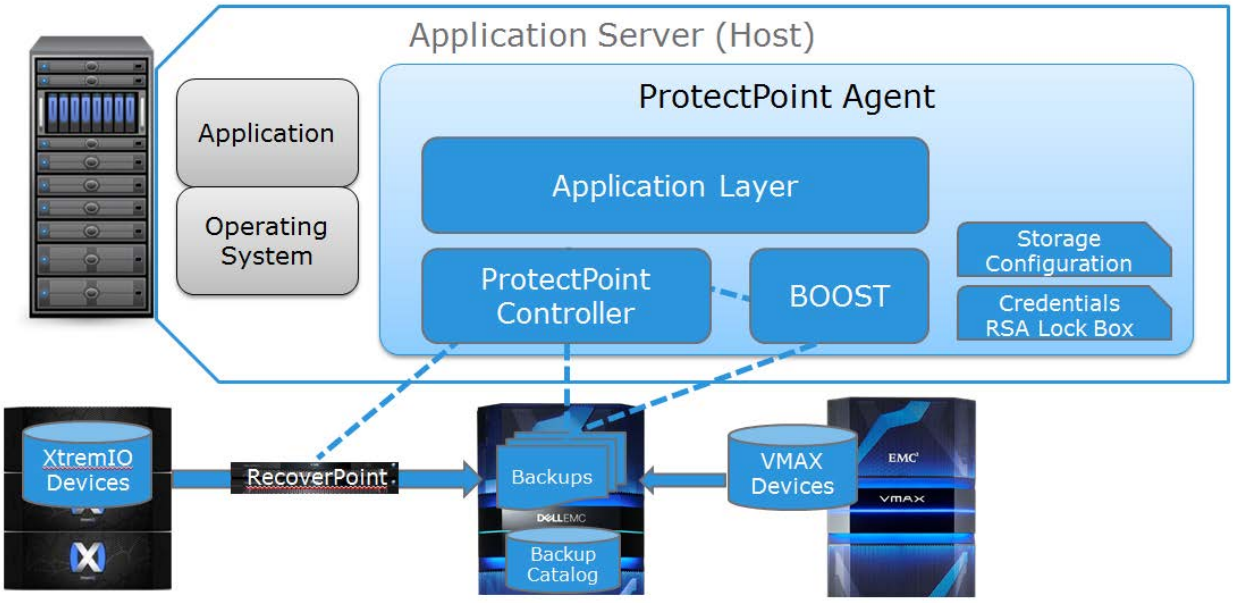
Full Restore Data Flow XtremIO with Data Domain



Granular Restore Data Flow XtremIO with Data Domain

operations, the ProtectPoint controller has the necessary configuration data and credentials to provide connectivity and authentication to the primary and protection storage. The ProtectPoint agent stores the credentials in an RSA secure lockbox. In addition, the agent stores configuration data - mapping the LUNs on the primary storage to the storage devices on the Data Domain system - to orchestrate backup (the transfer of changed blocks and creation of backup images) and recovery operations. The ProtectPoint control

plane ensures that these ProtectPoint backup and recovery operations seamlessly coexist with traditional primary storage availability workflows. In addition, ProtectPoint backups are recorded in a catalog on the Data Domain system along with the backup data. As shown here, a common agent is used for control, no matter if it's for VMAX^3 or XtremIO - as the differences between the two mechanisms are largely invisible.



VMAX^3 & XtremeIO to Data Domain using the universal ProtectPoint Agent

Figure 5. ProtectPoint Control Plane

THE MANANAGEMENT PLANE

Finally, the management plane enables the interaction with various administrators to configure, monitor and manage the environment. With ProtectPoint, the management plane supports the application owner, the storage administrator, and the backup or protection administrator to use their appropriate user interface to carry out tasks for ProtectPoint. Separating the management plane from the other data plane and control plane allows the provisioning of the most appropriate information and control to each of these administrators, which allows them to do their jobs more effectively and work together.

For the application owner, the ProtectPoint management plane provides full control of backup and recovery operations directly from their native utilities and interfaces. This includes full recovery of a LUN or set of LUNs as well as instant access to ProtectPoint backups for simplified granular recovery. Full recovery at the cost of the differential provides application owners and database administrators 10x faster recovery. In addition, ProtectPoint enables application owners to delete or expire previous backups.

For the storage administrator, the management plane allows the initial configuration and ongoing maintenance of the primary and protection storage infrastructure for multiple application owners. The management plane provides access to configure the storage infrastructure using their native storage management tools.

For the backup or data protection administrator, ProtectPoint provides centralized monitoring, analysis, and reporting capabilities through Dell EMC Data Protection Advisor (DPA). DPA provides dashboard views of ProtectPoint KPIs and will monitor and report on the progress of the data movement from primary storage to protection storage, providing insight into health, performance, and other key environment metrics. DPA also monitors protection policy compliance, providing at-a-glance visibility into gaps in expected application protection levels. And, should issues arise in the environment, DPA identifies and troubleshoots backup failures and generates automatic alerts.

For managers who have to keep track of SLO compliance standards, Dell EMC Enterprise Copy Data Management (eCDM) leverages ProtectPoint as an enabling technology. This allows IT Managers to configure and manage ProtectPoint workflows for superior copy data management between primary and protection storage. eCDM enables self-service data management with global oversight to maximize efficiency, streamline operations and ensure consistent service level compliance.

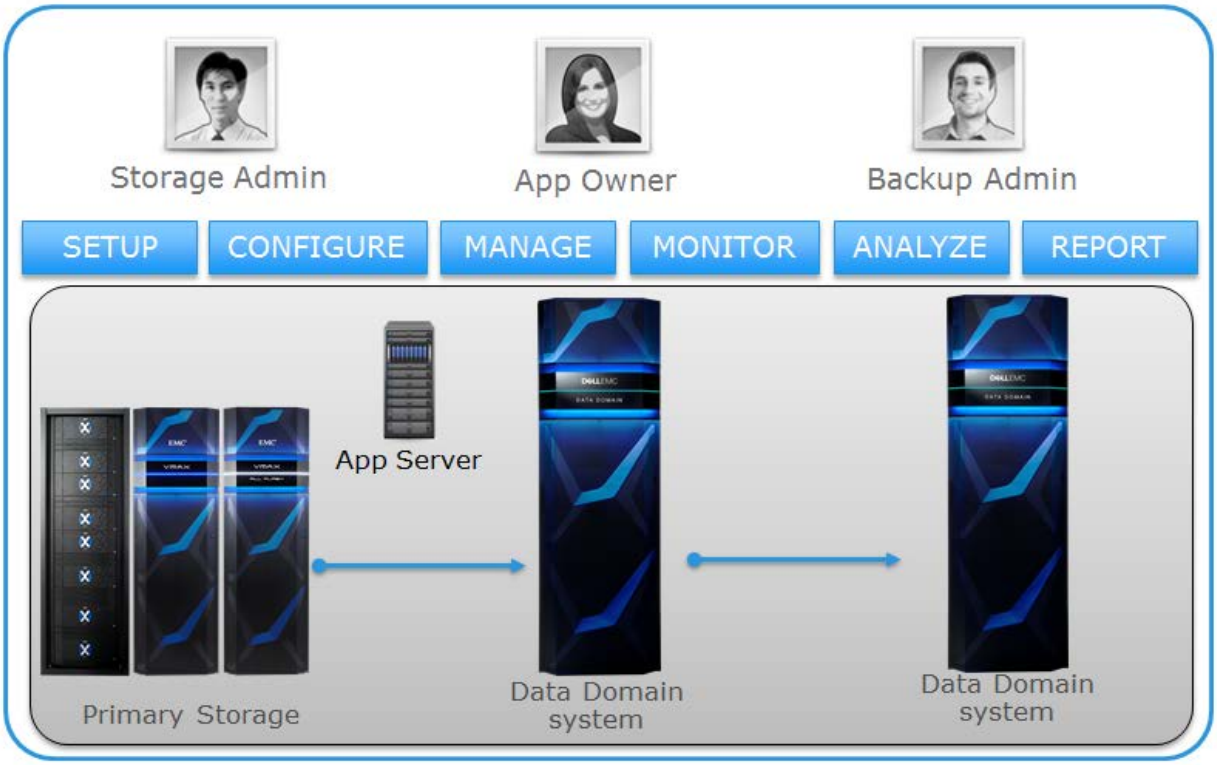


Figure 6. ProtectPoint Management Plane

SUMMARY

ProtectPoint provides the best of both worlds of data protection. Specifically the benefits of backup from Dell EMC, which includes:

- Cost effective retention with Data Domain deduplication that reduces backup storage requirements by 10 to 30x.
- Corruption protection that ensures data remains recoverable via the Data Domain Data Invulnerability Architecture provides.
- Native application integration with agents that empower application owners to control their own backups and gain application consistent backups for simplified recovery.

In addition, ProtectPoint, provides the benefits of snapshots including:

- No impact on the application server
- An RPO of minutes or hours
- Minimal RTO

This enables ProtectPoint to provide the functionality of backups with the performance of snapshots.

Overall, with ProtectPoint, you can reduce the time, cost, and complexity of managing application backups (FIGURE 7). First, you will eliminate the backup impact on the application server(s) with non-intrusive data protection, since no data flows through the application server and the application will only be momentarily paused for a backup. This ensures you will maintain consistent application performance, but still gain application consistent backups for simple recovery. Next, you'll finally be able to meet stringent protection SLAs without sacrificing anything. Since only changed blocks are sent directly across the SAN and all backups are stored in native formats – you'll gain much 20x faster backup, 10x faster recovery and instant access to protected data for simplified granular recovery. Finally, you'll be able to do all of this with a greatly streamlined infrastructure –reducing overall cost and complexity. ProtectPoint is simple, efficient and requires no additional infrastructure.

Figure 7. The Benefits of ProtectPoint

ProtectPoint technology addresses the key challenges of protecting large mission-critical applications that are particularly rigorous on those workloads commonly found on high-end enterprise class storage arrays. ProtectPoint uniquely addresses these challenges because it was designed from the ground up for efficient application protection, rather than relying on an existing backup application or native array snapshot replication. With this industry unique solution, it is now possible to protect applications efficiently without making any compromises.