COHESITY

Cohesity Data Protection White Paper

Table of Contents

Executive summary

Cohesity enables enterprises to take control of their increasingly complex storage environments through a hyperconverged secondary storage infrastructure. The Cohesity Data Platform can consolidate multiple use cases such as data protection, test/dev, file services and analytics onto a single web-scale platform. This paper specifically focuses on the business and technical benefits of the Cohesity Data Platform for the data protection use case. It is intended for IT professionals interested in learning more about Cohesity's technology differentiation and advantages it offers for data protection - (i) Pay-as-you-grow scalable architecture (ii) Consolidation of data protection silos (iii) Simplification of operational management (iv) Improvement in RTO/RPO objectives and (v) Leveraging economics of public cloud services.

Challenges in data protection fueling the next wave of innovation

A robust data protection strategy must meet the following criteria: (i) Provide protection from accidentally deleted files, application crashes, data corruption, and viruses (ii) Retain data for long periods of time to satisfy compliance and regulatory requirements (iii) Rapidly recover from unplanned downtime due to site disasters and maintain business continuity.

Over the last two decades, we have experienced tremendous innovations in data protection technology. Each innovation made huge advances in backup & recovery SLAs at a lower cost per unit of capacity when compared to previous generations of products (Figure 1).

With Cohesity Data Protection, organizations have the flexibility to choose how they want their data protected, managed, and stored. Whether using Cohesity's native data protection capabilities or leveraging its highly efficient, web-scale storage capabilities in conjunction with third-party data protection software, organizations benefit from complete data protection consolidation.

Key Benefits

- Simplify data protection with integrated backup and recovery running directly on the Cohesity Data Platform
- Instantly search-and-recover data with the Cohesity Indexing Engine
- Scale data protection across the enterprise with Cohesity's web-scale Data Platform architecture

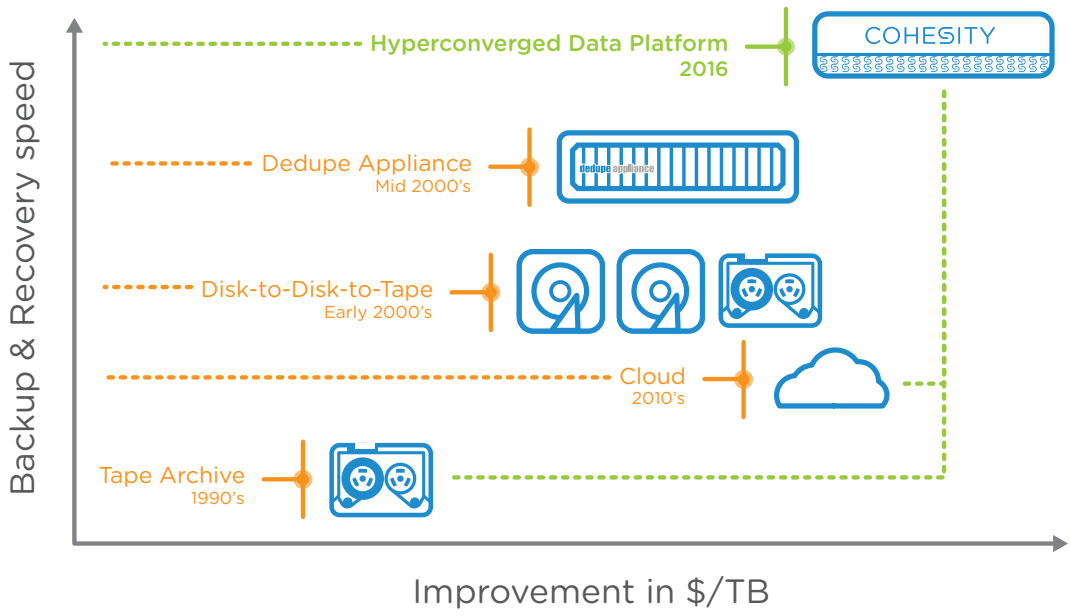

Figure 1: Evolution of data protection technologies. Cohesity is driving innovation through a Hyperconverged Data Platform that improves the Data protection metrics of speed and economics.

Originally data protection focused on backing up data from internal hard drives to tape drives and libraries. With advances in disk densities and inevitable reduction in costs, backup broadened to include hard disk drives (HDD) as the backup target. HDDs also improved the granularity of allowing individual items to be recovered thereby improving restore times and processes. The next wave of disk-based backup introduced a tiered backup architecture that combined disk and tape leading to Disk-to-Disk-to-Tape (D2D2T). The continued growth of disk as a backup target offered new opportunities to exploit the advantages of disk over tape leading to deduplication and compression technologies driving efficiencies in physical storage. Vendors began packaging purpose built backup appliances to further simplify deployment and maintenance. In the last few years, external cloud based storage as a backup target has entered the scene driving cost optimization efforts through a shift from upfront capital investment to an operational pay-as-you-grow model. Cohesity provides the next revolutionary wave of innovation through a hyperconverged storage architecture that delivers significant improvement in speed and SLA metrics at a substantially lower $/TB.

Before we delve into the details of the Cohesity data protection solution, it is important to understand the metrics across which these incumbent technology advances fall short in meeting the rigors of data protection. IT executives increasingly recognize that the key driver of exponential data growth is redundant copies of primary corporate data created by the various point-solution tools used to protect, share, and analyze information. IDC quantifies this as the copy data ratio which is measured as the total data in the environment over the total amount of production data. The increase in this ratio for enterprises stems from data residing in backup storage systems, in disaster recovery environments, in test and development clusters, and in archival for long term retention; all of which quickly multiplies across all applications. Gaining visibility and control over these data copies is becoming paramount to gaining operational efficiency and agility. Figure 2 is an illustration of the data protection architecture that is common across enterprises today. The challenge is that the different tiers of storage and backup solutions work in siloed hardware and software form factors with their corresponding data management and protection software. The solutions are mostly passive in nature, becoming expensive insurance policies that sit idle until a restore request is received. Also, in many cases customers only test these data backup copies infrequently and long gaps between testing increase the chance of issues being found when data needs to be actually recovered – at which point it may be too late.

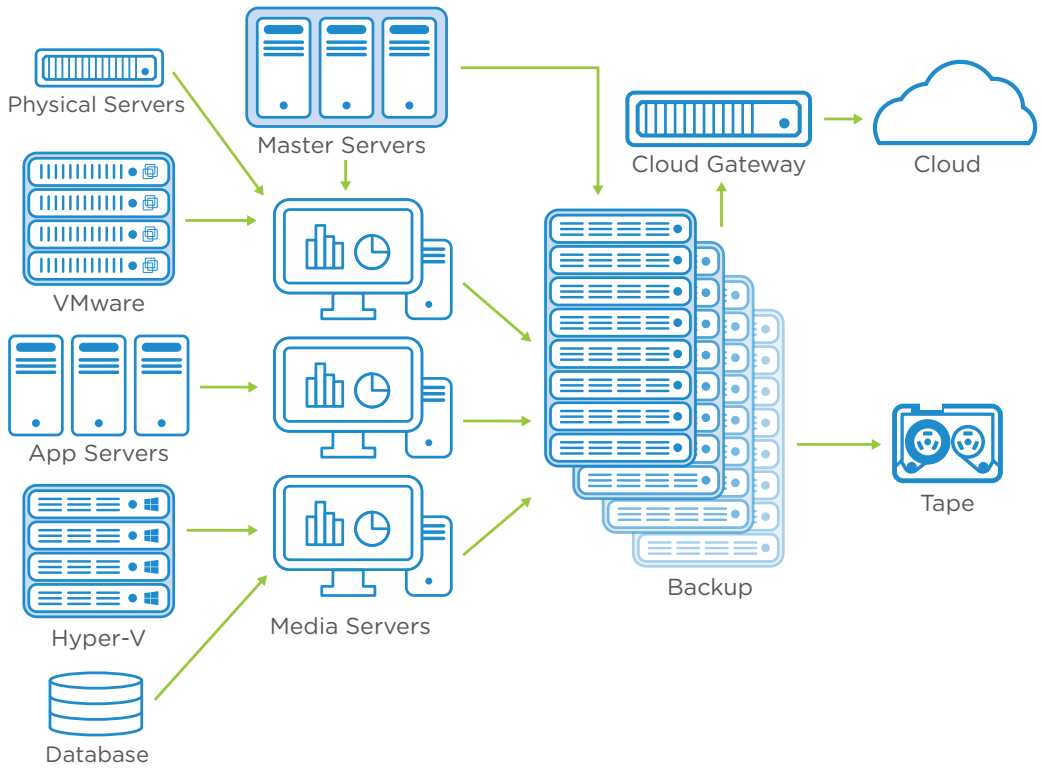

Figure 2: Challenges with current data protection solutions include but are not limited to multiple siloed hardware and software, heterogeneous management tools, islands of storage subsystems, and complexity of operational maintenance.

When we look further at the incumbent data protection solutions through the lens of speed and the associated costs, the assessment is as follows:

Speed of Backup & Recovery measured against delivery on SLAs

The 24/7 nature of enterprise is leading customers to ensure Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO) are reduced to a minimum. In spite of the high investment across various products, enterprises are still incurring downtime leading to several million dollars in penalties and lost productivity. This business impact is exacerbated by loss of customer and employee confidence if the data protection plan for critical services does not achieve its objectives. In trying to reduce RTO and RPO; customers and vendors create expensive, complex and one-off solutions that require ongoing maintenance to deliver on the "always-on-enterprise" precept.

Total cost of ownership

Organizations have invested in multiple solutions across backup, replication, DR, and archival to create and manage multiple copies of their data. This leads to expensive data migrations, forklift upgrades, and sub-optimal capacity planning strategies to accommodate future growth. In addition to the complexities associated with managing the independent hardware and multiple software elements, data protection solutions often have different procurement methods and licensing terms, making it extremely difficult for organizations to budget and procure. Even as customers are juggling with all these elements, they are also looking at cloud services as an extensible part of an overall storage solution because of the promise of economic and scalability advantages.

Introducing Cohesity next generation Hyperconverged solution for Data Protection

Cohesity was founded on the core vision of eliminating the fragmentation in data protection and putting an end to the above mentioned shortcomings that plague incumbent solutions. Architected and designed from the ground up to be the world's most efficient, flexible solution for enterprise data, the Cohesity Data Platform couples commercial off-the-shelf (COTS) hardware with intelligent, extensible software (Figure 3). The platform supports multiple protocols such as NFS, SMB and REST APIs. The integrated hardware and software solution enables organizations to spend less time worrying about how to retrofit their legacy solutions with future needs, and more time focusing on the core functions of the enterprise. The Cohesity Data Platform provides the following benefits for data protection:

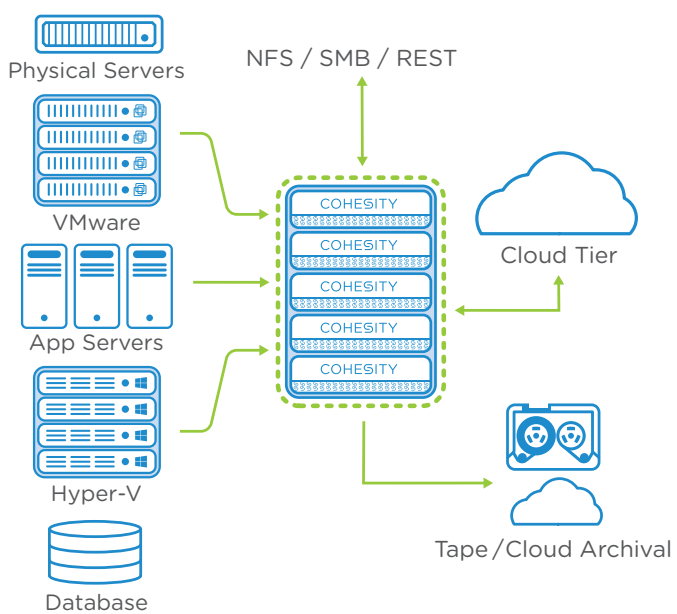

Figure 3: Cohesity Data Protection for virtual and physical environments with integrated, policy-driven backup and recovery. The solution also supports long term data retention of least-used data to external tape and public cloud services such as Google Cloud Storage Nearline, Microsoft Azure and Amazon S3/Glacier.

Pay-as-you-grow scalable architecture with maximum data reduction

Cohesity provides incremental pay-as-you-grow yet limitless growth through a distributed, web-scale platform called OASIS™ (Open Architecture for Scalable Intelligent Storage). This eliminates the need to invest in hefty upfront capital expenses for anticipated future growth and removes many hidden costs associated with forklift upgrades of traditional scale-out systems. With the ability to add or remove hardware nodes at any time, the Cohesity cluster automatically scales up or down by rebalancing the data and its associated metadata to ensure redundancy. The OASIS platform coupled with Cluster wide global data deduplication and compression ensures that maximum data reduction goals are achieved while backups are readily available for instant recovery.

Consolidate data protection appliances and media servers

Cohesity consolidates all the unnecessary silos across the entire lifecycle of data protection (backup, restore, replication, DR and failover/failback orchestration) for both virtualized and physical environments. The Cohesity platform eliminates this fragmentation by eliminating disparate Master & Media servers, Backup software & appliances, Replication software and Cloud gateways. The necessary functionality of all these is integrated in the Cohesity solution and scales with the growing secondary storage needs. This avoids the need to invest layered backup software and backup target solutions. Additionally, the consolidation decreases the hardware footprint in the data center leading to lower spend on space, power and cooling requirements.

Simplify operational management

Cohesity's native data protection software enables businesses to easily protect data for virtual and physical environments, dramatically reducing cost and complexity for businesses. The Cohesity Data Protection software is tightly integrated with VMware vCenter for businesses to instantly see a full index of the virtual environment and choose which virtual machines to protect. These virtual machines can then be protected with easy-to-use protection profiles that can be customized based on SLA, retention period, or application group. The process leverages available VMware APIs for Data Protection (VADP), eliminating the need to install in-guest agents across the virtualized infrastructure. As new virtual machines are added, they are auto discovered and included in the protection policy that meets the desired SLAs. Easy and intuitive policy administration through a single pane of glass across an entire global datastore provides a plug-and-play experience for managing the daily operations. In addition, Cohesity provides out-of-the-box connectors for Oracle, SQL and other widely used enterprise applications to backup and restore at the database, object or item level. Customers also have the option to supply custom backup scripts that leverage native APIs (such as Oracle RMAN, rsync for physical server, dumpbin for MySQL) to move the backup file onto the Cohesity platform. Such broad support for virtual and physical environments has the benefit of having Cohesity schedule, report, replicate, and archive the backup datasets across heterogeneous applications.

Improve SLAs

Cohesity's patented SnapTree™ technology for snapshots allows businesses to take a large number of clones at any time interval with uncapped retention policies, without ever affecting performance or consuming additional space. Snapshots built on SnapTree™ technology allow businesses to protect their data as often as they want, without any performance impact or space overhead. Snapshots leverage the Redirect-on-write (RoW) technique to keep track of changes by writing the changed data to new blocks. Each of these snapshot clones is fully hydrated so that businesses can achieve fast RTO and near-continuous RPO objectives. Data protection is further enhanced through an indexing engine that rapidly indexes an entire vCenter cluster and all its associated metadata. This has the benefit of easily mining backup data with a simple

Tribune

Tribune Media, a global media company, had separate data management, protection, and storage solutions that resulted in integration challenges and management complexity. Tribune achieved the following benefits by deploying Cohesity Data Protection

- Global data deduplication that curbs data growth and reduces footprint by over 10x
- Single pane of glass management that saves over 80% in time spent managing storage
- Unlimited snapshots and near-instant search-and-recover capabilities shrinking RTO and RPO windows

> "Cohesity gives us an easy way to manage, store, and extract value from our data — all from a single platform that scales as we need it. Cohesity is dramatically simplifying the way we do business."
>
> – David Giambruno, CIO and Senior VP of IT

text-based search and restore. This restore can also place the file(s) in the original source location further reducing the burden associated with managing restore processes. Data can also be replicated across Cohesity clusters in multiple sites to offer protection from disaster scenarios. This is all achieved with a single click user interface, drastically reducing the time and overhead when compared with traditional data protection solutions.

Leverage favorable economics of public cloud services

The Cohesity solution supports long term data retention of seldom-used data to external tape and public cloud services such as Google Cloud Storage Nearline, Microsoft Azure and Amazon S3/Glacier. Customers can leverage the public cloud as an extension of the on-prem Cohesity infrastructure in one of two ways (i) Cloud Archival – archiving the older local snapshots in the Cohesity cluster to cloud for long-term retention. As restores from archivals are typically few and far apart, the cost associated to recover data from the cloud can be kept under check. (ii) Cloud Tiering – using cloud as an extension to Cohesity's built-in storage tiers provides the ability to algorithmically decide when to down-tier or up-tier the data between Cohesity cluster and the Cloud.

Cohesity Deployment Options for Data Protection

Cohesity delivers these benefits through a scalable integrated hardware and software solution for data protection that achieves economies of consolidation and scale. With the Cohesity Data Platform, organizations have the flexibility to choose how they want their data protected, managed, and stored. You can deploy the Cohesity solution in either of two modes (Figure 4): (i) Using Cohesity's native data protection hardware and software capabilities or (ii) Leveraging its highly efficient, web-scale storage capabilities as a Backup target in conjunction with third-party data protection software (such as Veeam, Commvault, NetBackup). In either mode, you can buy what you need today and scale up as your data grows with the benefit of leveraging the latest hardware technology in the market.

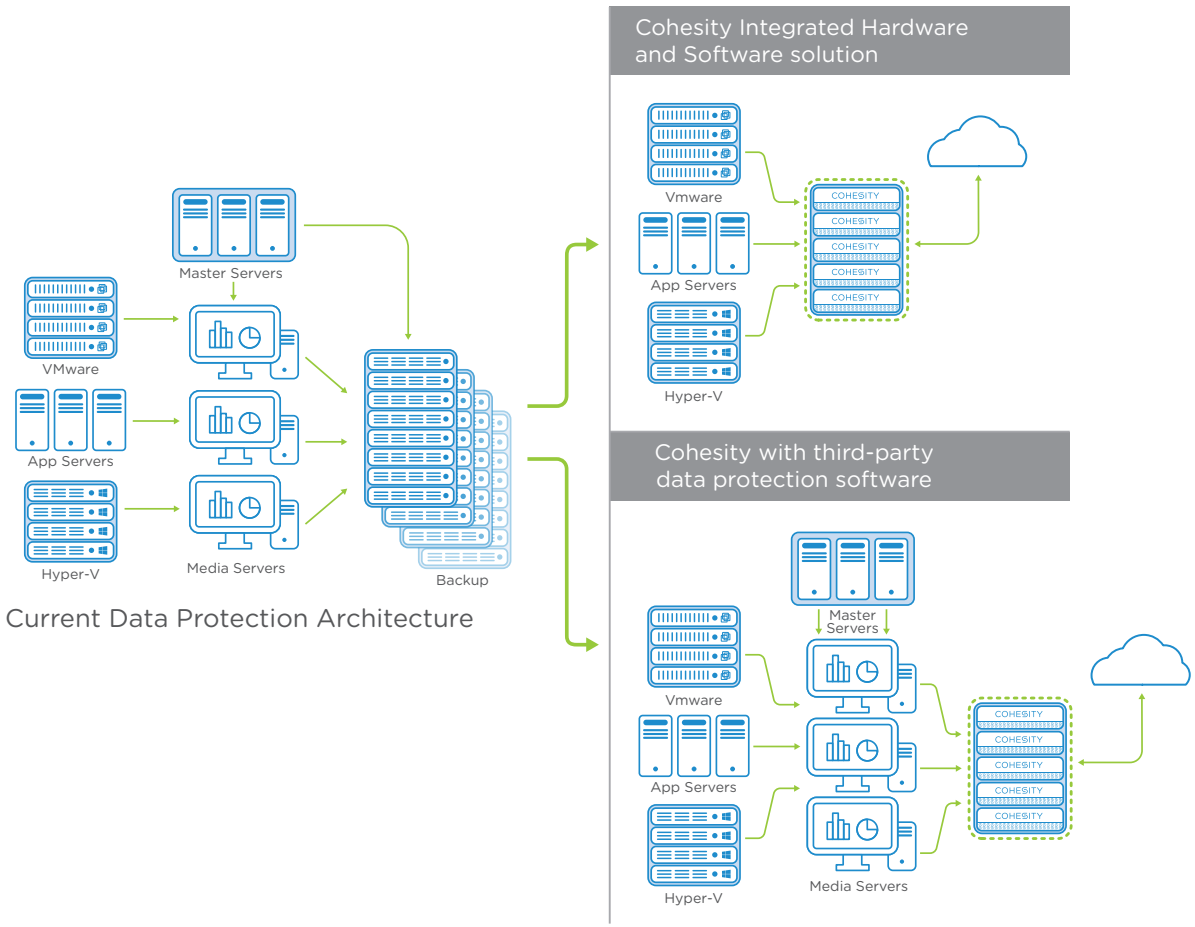

Figure 4: Cohesity Data Protection deployment has two options (i) Cohesity Integrated Hardware and Software solution (ii) Cohesity as backup target storage with third-party data protection software

Cohesity Data Protection Technical Advantages

Cohesity Data Protection is available as an integrated appliance. The hardware is preloaded with Cohesity OASIS[1], the flagship operating environment that provides the foundation for complete data consolidation. The specific modules handling the data protection aspects are detailed below (Figure 5):

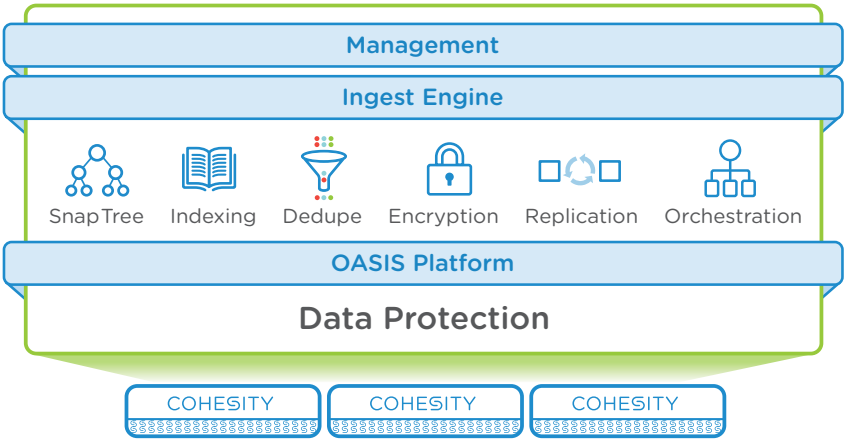

Figure 5: Cohesity Data Protection Architecture

Cohesity OASIS web-scale Platform

Cohesity OASIS is the operating environment that provides the foundation for data protection. Customers can start with as low as a three-node cluster and scale out one-node-at-a-time, to accommodate future capacity and performance requirements. A Cohesity node is a physical CPU with disk storage that includes the Cohesity Data Protection software. OASIS operates on file units called chunks that have a data and metadata component. The metadata includes the access control information, the mapping from files to chunks, and the current locations of the chunks. Several multi-threaded processes also control system-wide activities such as chunk management, garbage collection of orphaned chunks, and chunk migration between the nodes. This process also ensures that the system can self-heal from underlying hardware component failures. Periodic communication among the cluster nodes are used to exchange messages and distribute the workload uniformly across all the Data protection jobs. For recovery, the platform also has the ability to mount the virtual machines on the Cohesity cluster itself to provide Instant Restore capability. Storage vMotion is then used in the background to perform the live migration of the VM from the Cohesity storage to the production storage. This achieves the benefit of being able to perform backup and recovery operations much faster than traditional file system architectures.

Cohesity Ingest Engine

The Ingest Engine ensures that data is optimally placed onto the SSD or spinning disk tier that best suits the profile of the incoming data stream. This is called the Tier-Optimized Write Scheme (TOWS) wherein spinning disks that prefer sequential I/O write data out-of-place; while SSDs are used for random I/O such that the writes are in-place and committed straight away. This engine provides the basis to interoperate with VMware VADP, physical servers and application connectors; to provide end-to-end data protection for customer environments. In order to ensure that performance of the primary datastore is impacted minimally, the Ingest Engine offers Adaptive data throttling that modulates the backup ingest performance over the production workloads at the vCenter or Datastore level.

Cohesity SnapTree™ for managing data copies

In legacy storage solutions, snapshots (of a file system at a particular given point in time) form a chain, tracking the changes made to a set of data and form the basis for organizing and storing copies of data. Every time a change is captured, a new link is added to the chain. As these chains grow with each and every snapshot, the time it takes to retrieve data on a given request grows because the system must re-link the chain to access that data. Cohesity's patented SnapTree™ technology creates a tree of pointers that limits the number of hops it takes to retrieve blocks of data, regardless of the number of snapshots that have been taken. SnapTree uses a B+ tree data structure such that access to any point in the tree takes a fixed number of hops no matter how many snapshots there are, without having to rebuild any chain linkage. Because

[1]For full details of the Platform architecture, please refer to http://13554-presscdn-0-95.pagely.netdna-cdn.com/library/2015/10/Cohesity-Architecture-WhitePaper.pdf

SnapTree is implemented on a distributed file system, every node sees the same nested structure of the chain with a fixed depth independent of where the actual data is stored in the cluster. Keeping the snapshots fully hydrated improves the recovery times of any snapshot from t0 to tn because it does not incur the time penalty of traversing the entire chain of changes (Figure 6). This capability is available with the Integrated Cohesity data protection.

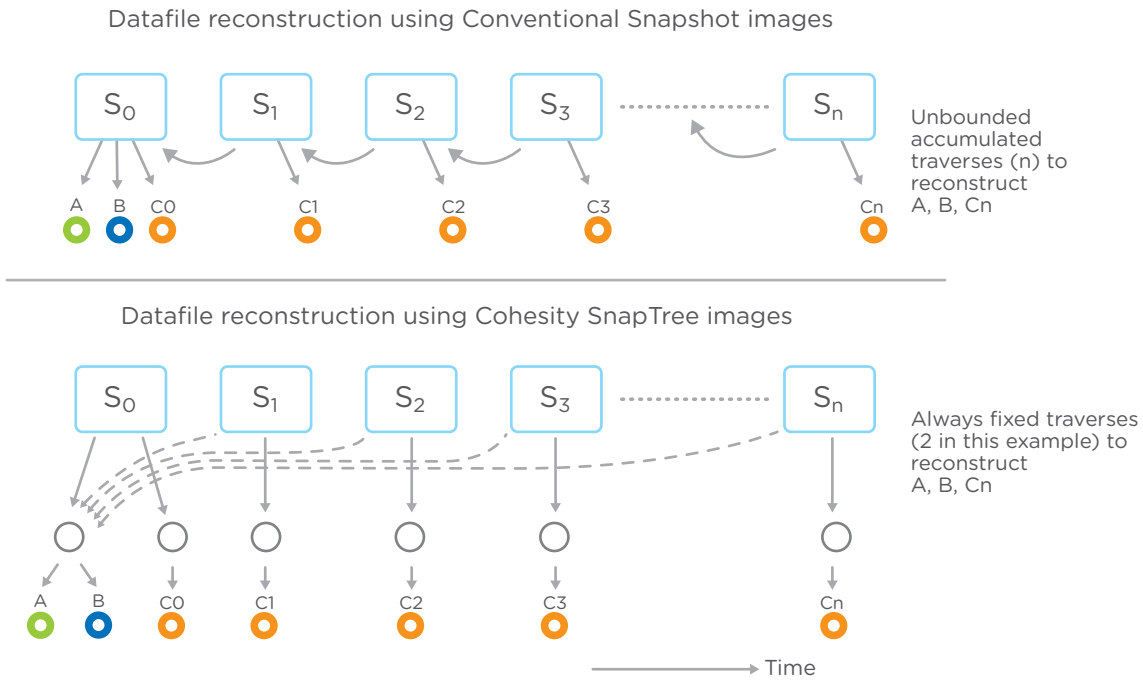

Figure 6: Cohesity SnapTree technology helps create large number of snapshots without incurring the penalty of traversing the entire snapshot chain seen in traditional snapshot architecture

Cohesity Indexing Engine for Metadata Management

OASIS and SnapTree work in conjunction with an Indexing Engine that enables rapid search-and-recover capabilities. This allows users to quickly find and restore files stored within higher-level data objects such as VMs. This Indexing Engine automatically and rapidly indexes an entire vCenter cluster and all its associated metadata. As virtual machines become protected, Cohesity's Indexing Engine cracks open the underlying files and indexes the metadata; enabling extremely fast, wild-card search results that are then used for instantaneous granular restores. As organizations scale-out, this Indexing Engine spans across all nodes in the cluster, leveraging the aggregate power of all CPUs and available memory to rapidly recover files or virtual machines. This greatly improves the RTO and RPO objectives compared to traditional data protection architectures. This capability is available with the Integrated Cohesity data protection.

Cohesity Global Deduplication for Data Reduction

Data deduplication is a storage efficiency feature that frees up storage capacity by eliminating redundant data blocks. Different vendors implement deduplication at a file-level and/or a block-level of different sizes, which only works well across a single storage pool or within a single object (e.g. application or VM). Cohesity leverages a unique, variable-length data deduplication technology that spans an entire cluster, resulting in significant savings across a customer's entire storage footprint (Figure 7). With variable-length deduplication, the size is not fixed. Instead, the algorithm divides the data into chunks of varying sizes based on the data characteristics. The chunk size is varied in real time in response to the incoming data which results in greater data reduction than fixed-size deduplication. The efficiency benefit of variable-length deduplication compounds over time as additional backups are retained. Cohesity also allows customers to decide if their data should be deduplicated in-line (when the data is written to the system) or post-process (after the data is written to the system) to optimize the backup protection jobs against backup time windows.

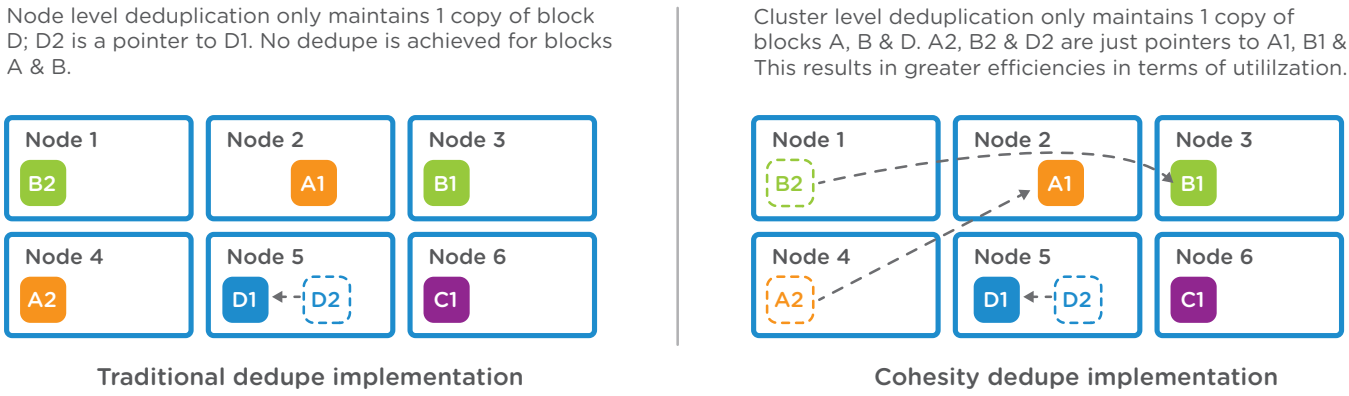

Figure 7: Cohesity's Global deduplication across all nodes in a cluster results in less storage consumed compared to just node level deduplication used in several other in market data protection solutions

Cohesity Encryption Engine

OASIS also provides encryption of data at rest and in transit over the network with AES 256-bit encryption to secure data. Encryption in flight is applicable to data that is replicated to a remote Cohesity cluster or when data is tiered/archived to the cloud from the Cohesity platform. This ensures that data stored on the Cohesity cluster is protected well from malicious attacks.

Cohesity Replication and DR Orchestration

The backups that are generated in one Cohesity cluster can be replicated to one or more target Cohesity clusters on multiple sites. The ability to replicate backups from primary Cohesity cluster to multiple secondary Cohesity clusters across various geographical sites, helps facilitate disaster recovery needs. The technology supports (i) One-to-one model - A single production cluster can back up to a disaster recovery site (ii) One-to-many model - A single production cluster can backup to multiple disaster recovery sites (iii) Many-to-one model - Multiple Cohesity clusters can back up to a remote cluster (iv) Many-to-many model - Multiple Cohesity clusters can backup to multiple remote clusters across several sites. This technology works in conjunction with global data deduplication to greatly reduce storage requirements across several sites and as such the network bandwidth required for replication of data for DR purposes. Replication is also augmented through DR workflow orchestration to provide customers the ability to seamlessly discover and connect to virtualized infrastructure on the secondary site. This ease of use enables customers to perform frequent testing to ensure disaster recovery predictability.

Cohesity Management for Data Protection

Data protection workflow in the Cohesity platform is managed through a single user interface (Figure 8). The dashboard provides an overview of Data Protection jobs, System health, Alerts and Storage utilization of all Cohesity clusters under management. Users also have the option of creating custom reports or using the REST API interface to integrate with existing IT Management and Monitoring tools.

GS1, global logistics provider, had data growth rates of 30% with stringent retention requirements. With a limited time window for completing backups, they were using expensive primary storage to accomplish this task. By leveraging Cohesity, they were able to

- Achieve faster backup and restore that has shrunk RTO and RPO windows by over 90%
- Single pane of glass to manage consolidated secondary storage

"Faced with the complex task of managing massive amounts of data that must be backed up and secured, we are excited about Cohesity's vision for converging data onto a unified, scale-out solution. Cohesity provides simpler data protection and powerful insight, which is what we want from our next-generation solution for storage and recovery. We look forward to scaling this environment as we need, one node at a time, so we can spend less time managing our data infrastructure." – Sase Janki, SVP of Technology

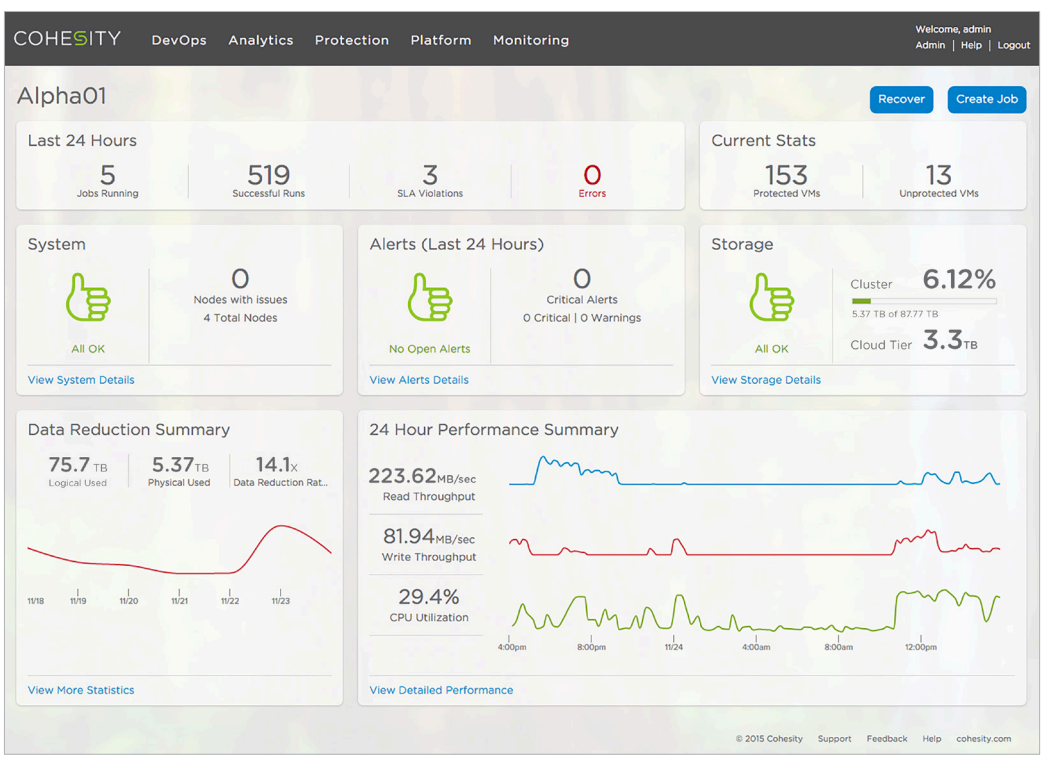

Figure 8: View of Cohesity User Interface and Data Protection

Cohesity is powering customers into a new era of data protection

The Cohesity data management platform is helping customers usher in new technology to achieve their data protection business objectives. Customers such as Tribune Media and GS1 had disparate data management, protection, and storage solutions that resulted in integration challenges and management complexity. With Cohesity, they have been able to shrink their RTO and RPO windows by over 90% and have greatly reduced the time spent on managing the infrastructure. In conclusion, organizations can use the integrated data protection capabilities of the Cohesity Data Platform for physical and virtual environments to consolidate disparate hardware and software elements. It is also possible to leverage the highly efficient, web-scale storage capabilities of Cohesity in conjunction with third-party data protection software. The move to Cohesity Data platform can provide a robust data protection strategy that delivers improvement in backup/recovery speed and better economics than incumbent solutions.

For more information about Cohesity, please visit www.cohesity.com