COHESITY vmware®

Simplifying Data Protection for VMware Virtual SAN with Cohesity DataPlatform

Table of Contents

Executive summary

VMware has introduced two key new technologies over the past couple of years that have dramatically improved the management and delivery of primary storage resources: Storage Policy-Based Management (SPBM) and Virtual SAN.

Storage Policy-Based Management (SPBM) helps transform storage management for both Virtual SAN and external SAN and NAS storage. SPBM enables administrators and application owners to define service-level policies and associate those policies to applications. The underlying infrastructure automatically ensures the service levels are being met. SPBM turns storage management on its head to go from the legacy, static 'infrastructure-centric' model to a dynamic application-driven model.

VMware Virtual SAN is VMware's enterprise-class primary storage solution for hyperconverged infrastructure. It is designed and capable of supporting all of the vSphere related use cases. Virtual SAN is uniquely embedded in the vSphere Hypervisor. It's multiple flash-optimized based architectures are designed to deliver the highest levels of performance and cost effectiveness to all vSphere virtualized infrastructures — at a fraction of the cost of traditional, purpose-built storage and other less-efficient hyperconverged infrastructure solutions.

Together, SPBM and Virtual SAN have dramatically improved primary storage for vSphere applications. Unfortunately, secondary storage is holding us back. Data protection for vSphere and Virtual SAN is unnecessarily complicated, typically consisting of target dedupe appliances, separate backup software, backup infrastructure, replication software, and archival. Data protection cannot be managed through SPBM and still relies on a bottoms-up, static, infrastructure-centric management model. Typically – applications are backed up once a week with a full backup, and daily using incremental backups. The service level is static, doesn't adapt to individual application requirements, and the RPO and RTOs are long. Finally, secondary storage typically relies on multiple storage silos with many copies of secondary data – one silo for data protection where data is kept unproductive, and additional storage silos for test/dev and analytics.

Cohesity DataPlatform provides the only web-scale platform designed to consolidate all your secondary data and workflows. Cohesity provides a scale-out, globally deduped, highly available storage fabric to consolidate your secondary data, including backups, files, and test / dev copies. Cohesity provides a single, unified solution to simplify data protection, integrate with the public cloud, support test/dev environments, and provide deep visibility into secondary data with built-in analytics.

Cohesity DataPlatform is the ideal secondary storage platform for Virtual SAN. Together, Virtual SAN and Cohesity bring the simplicity of scale-out hyperconverged solutions to both primary and secondary storage. The joint solution delivers the following benefits:

- Web-scale, pay-as-you grow architecture everywhere
- Dynamic, application-centric operations with storage policy-based management everywhere
- Eliminate complexity with a unified platform for end-to-end data protection
- Ensure fast Recovery Points and near-instantaneous Recovery Times
- Lower total cost of ownership for both primary and secondary storage
- Consolidate backups, files, and test/dev copies on a single web-scale platform
- Accelerate application time-to-market with instantaneous provisioning of clones for test/dev
- Get deep visibility into your secondary data with built-in analytics capabilities

This solution paper introduces a comprehensive hyperconverged solution with VMware Virtual SAN and Cohesity DataPlatform. This joint solution provides an organization with a consolidated end-to-end hyperconverged infrastructure solution that delivers a much more cost-effective, dynamic, and high-performance storage fabric through the entire primary and secondary storage stack.

Primary Storage Has Improved Dramatically – But Secondary Storage Remains a Complicated, Fragmented Mess

As we move to a digital world, storage demands continue to explode in many IT environments, with no end in sight. More business models are now being driven by the need to acquire and harness ever-growing mountains of information. According to analyst firms, over the next two years, storage capacity will grow at 40% annual growth.

In today's highly demanding business environments, organizations are placing increased emphasis on the management, accessibility, and availability of mission critical applications and data management. Business critical applications and data management solutions are typically accompanied with rigorous demands for performance, accessibility, and availability service levels that must be satisfied by the infrastructures, systems, and solutions hosting them.

Storage Policy-Based Management and Virtual SAN Solve the Primary Storage Challenge

VMware introduced Virtual Volumes and Virtual SAN to fundamentally solve primary storage challenges in today's virtualized environments. Together, these two technologies introduce two concepts to storage in vSphere environments:

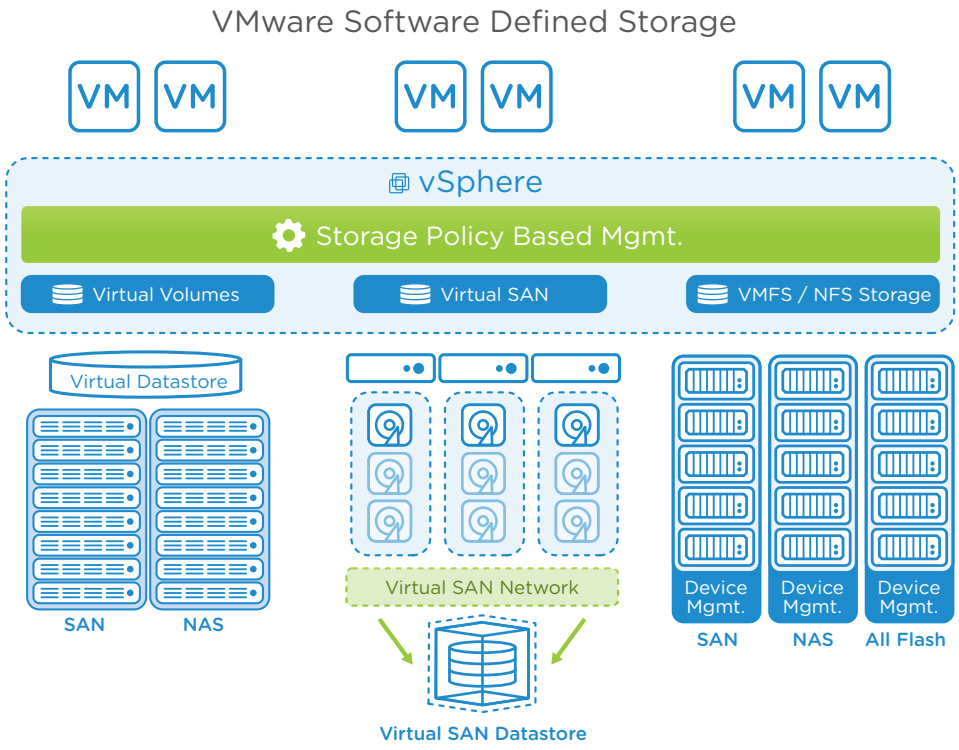


Figure 1: Virtual SAN and Storage Policy-Based Management solve the primary storage challenge

Policy-driven control plane: Storage Policy-Based Management (SPBM)

The policy-driven control plane is the management layer responsible for controlling storage operations. The control plane acts as the bridge between applications and storage infrastructure. The control plane provides a standardized management framework for provisioning and consuming storage across all tiers, whether on external arrays, x86 server storage or cloud storage.

Through SPBM, the storage classes of service, become logical entities controlled entirely by software and interpreted through policies. Defining and making adjustments to these policies enables automating the provisioning process at scale, while dynamically controlling individual service levels over individual virtual machines at any point in time. This makes the SPBM model able to adapt to ongoing changes on specific application requirements. Policies also are the mechanism to automate the monitoring process and to ensure compliance of storage service levels throughout the lifecycle of the application.

The control plane is programmable via public APIs, used to consume and control policies via scripting and cloud automation tools, which in turn enable self-service consumption of storage to application tenants.

Hyperconverged Infrastructure: Virtual SAN

Virtual SAN is VMware's software-defined solution for hyperconverged infrastructure. Seamlessly embedded in the hypervisor, it delivers enterprise class, elastically scalable, high performance shared storage for vSphere VMs that is radically simple to manage and lowers TCO. Virtual SAN pools server-attached HDDs and SSDs to create a distributed shared datastore that abstracts the storage hardware and provides a hyperconverged storage optimized for virtual machines.

From an infrastructure and architecture perspective, VMware Virtual SAN as a primary storage is a proven hyperconverged platform that is capable of complying with the most stringent performance, and availability requirements of any business critical application, and use case supported by vSphere.

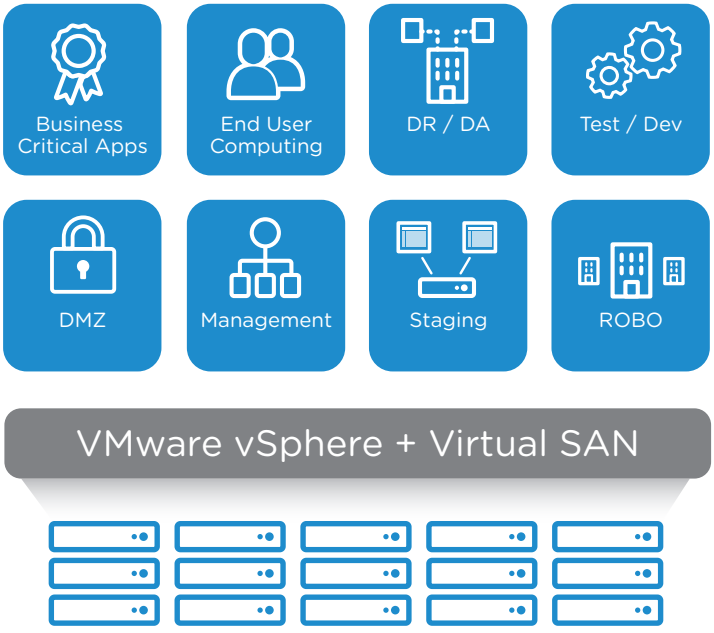


Figure 2: Virtual SAN is the built-in hyperconverged solution for vSphere

But Secondary Storage Is Holding Us Back

SPBM and Virtual SAN have solved many of the storage challenges in primary storage. Unfortunately, the secondary storage segment hasn't kept up. The new primary storage model can be compared to a Ferrari engine – it's dynamic and high-performance – able to keep up with the most demanding and dynamic applications. But running it with the legacy secondary storage infrastructure of yesterday is a bit like mounting a Ferrari engine on a Fiat 500 chassis. In short, the secondary storage infrastructure is holding back the potential of your private cloud as a whole.

Complex and ineffective data protection

Secondary storage starts with data protection: backup, recovery, replication, and archival. Traditional data protection is sustained by a myriad of siloed products and solutions designed with legacy and limited technologies that are individually targeted to different pieces of the data protection landscape. Overall, this is highly complex and risky with the inability to comply with the level of simplification and operational efficiencies demanded by customers for the management of business/mission critical data.

The figure below depicts 3 production Virtual SAN clusters running mission critical multitier enterprise applications. With the existing data management solution for backup and recovery, customers are forced to invest in disparate master and media servers, backup software & appliances, replication software and cloud gateways, etc.

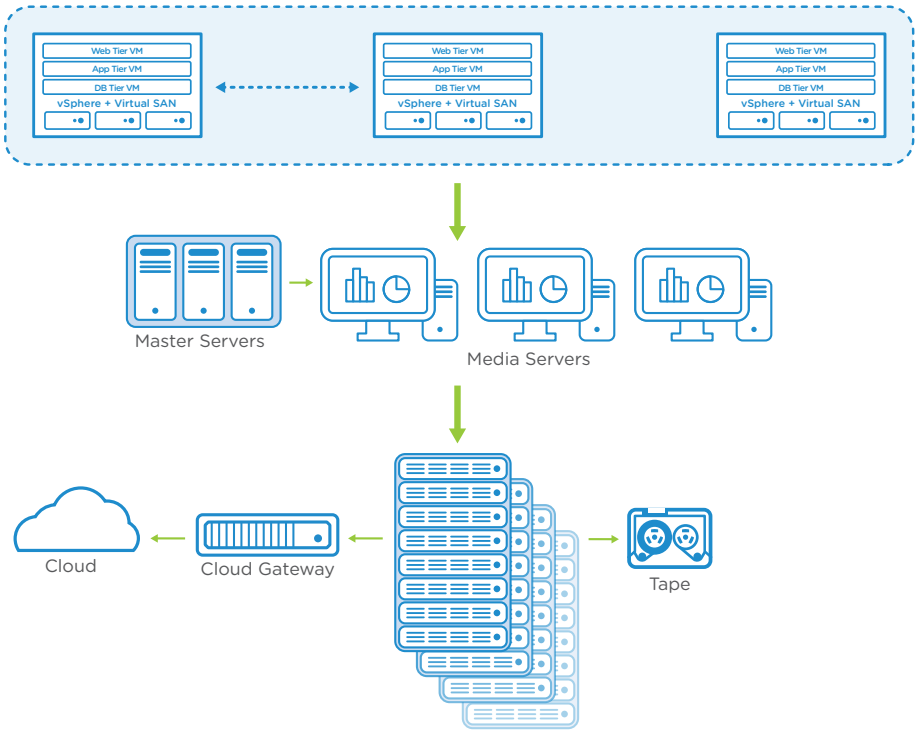


Figure 3: Complex data protection for VMware environments

The data is written to target storage systems by the backup software; which then interfaces with tape and cloud for long term retention needs. Enterprises are forced to invest in multiple tiers and silos of data protection solutions. The infrastructure is brittle and complex, in stark contrast to the radically simple model VMware Hyper-Converged Software with Virtual SAN delivers. Instead, all of these limitations lead to expensive data migrations, forklift upgrades, and complex capacity planning practices to accommodate future growth.

Lack of policy-based management

Legacy data protection solutions are managed in an infrastructure-centric way. Most applications get backed up weekly with a full backup, and daily with an incremental backup. This model is rigid, inflexible, and doesn't adapt well to the individual demands of each application. It's not possible to define a specific RTO, RPO, and retention requirements in a declarative policy framework. In short – it doesn't align with the Storage Policy-Based Management introduced by VMware for primary storage.

Many copies of data required for different use cases

Legacy data protection solutions only offer an insurance policy against failures. The data is kept safe, but does not produce any value. Data sitting on a traditional dedupe appliance cannot be used to support business value. The dedupe appliance is in effect like a non-productive brick.

Consequently, organizations have to make multiple copies of data for different use cases. Data is kept with multiple full backups on dedupe appliances. It is then replicated to an off-site location and taped out for long-term archival.

Even worse – data has to be copied, often to a separate NAS device, for test/dev purposes. And sometimes it is copied to yet another location for analytic-type jobs. This all leads to an ineffective sprawl of data copies across different secondary storage silos.

A good data management strategy needs to 1 – drastically simplify the data protection infrastructure 2 – support a policy-based management framework and tie-in with the VMware SPBM model and 3 – enable the protected data to be productive, typically for use in test/dev environments and analytics.

Introducing VMware Virtual SAN

VMware Virtual SAN Overview

VMware Virtual SAN is radically simple, enterprise-class primary storage solutions that is uniquely embedded in the vSphere hypervisor. Virtual SAN is based on a Flash-optimized architecture, high-performance storage for VMware Hyper-Converged Infrastructures. Virtual SAN delivers high performance, highly resilient shared storage by clustering server-attached flash-based devices (SSD, PCIe, etc.) and magnetic devices (HDDs).

Virtual SAN clusters contain two or more physical hosts that contain either a combination of magnetic disks and flash devices (hybrid configuration) or all flash devices (all-flash configuration) that contribute cache and capacity to the Virtual SAN distributed datastore.

In a hybrid configuration, one flash device and one or more magnetic drives are configured as a disk group. A disk group can have up to seven magnetic drives. One or more disk groups are utilized in a vSphere host depending on the number of flash devices and magnetic drives contained in the host. Flash devices serve as read-and-write cache for the Virtual SAN datastore while magnetic drives make up the capacity of the datastore.

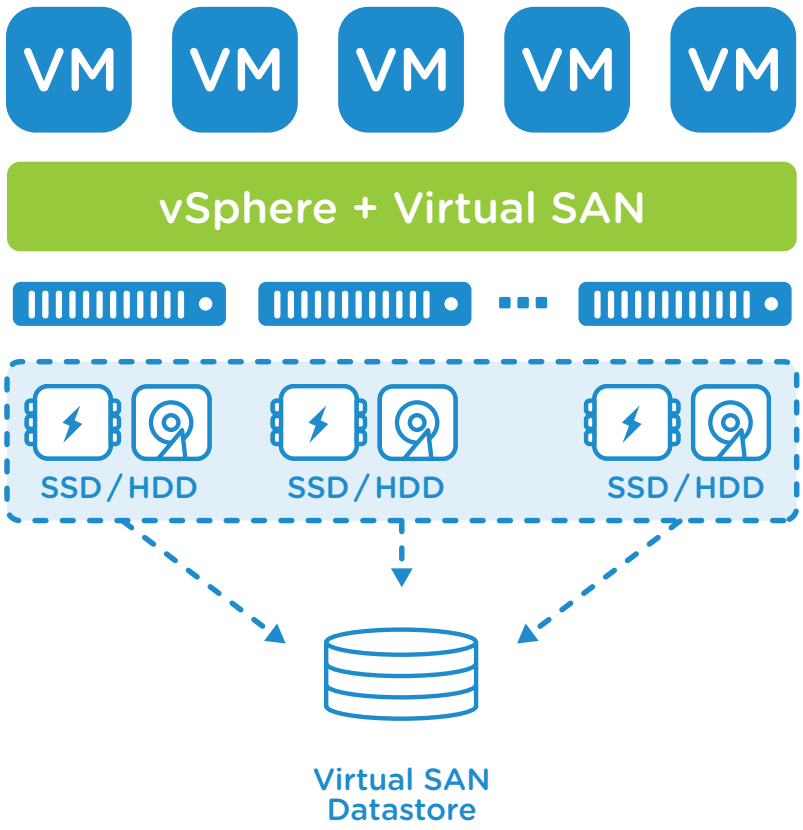


Figure 4: VMware Virtual SAN Overview

By default, Virtual SAN will use 70% of the flash capacity as read cache and 30% as write cache. For all-flash configurations, the flash device(s) in the cache tier are used for write caching only (no read cache) as read performance from the capacity flash devices is more than sufficient.

Two different grades of flash devices are commonly used in an all-flash Virtual SAN configuration: Lower capacity, higher endurance devices for the cache layer and more cost effective, higher capacity, lower endurance devices for the capacity layer. Writes are performed at the cache layer and then de-staged to the capacity layer, only as needed. This helps extend the usable life of the lower endurance flash devices in the capacity layer.

Virtual SAN is a distributed object storage system that uses the vSphere Storage Policy-Based Management (SPBM) framework to deliver centrally managed, application-centric storage services and capabilities. Administrators can specify storage attributes, such as capacity, performance, and availability, as a policy on a per VMDK level. The policies dynamically self-tune and load balance the system so that each virtual machine has the right level of resources.

All this at a fraction of the price of traditional, purpose-built storage arrays. Just like vSphere, Virtual SAN provides users the ability and control to choose from a wide range of hardware options and easily deploy and manage the hardware infrastructure for a variety of IT workloads and use cases.

For more information on VMware Virtual SAN, please refer to:

http://www.vmware.com/files/pdf/products/vsan/VMware_Virtual_SAN_Whats_New.pdf

VMware Virtual SAN Benefits

- End-to-end integration and unified management: VMware Hyper-Converge Software solution provides a single layer of compute, networking and storage software. Management is unified through common tools and interfaces. Features like HA, vMotion, DRS etc. work seamlessly across the VMware stack.

- Streamlined Provisioning and Automation: SLAs can be controlled through ‘virtual machine-level’ policies that can be set and changed on-the-fly. Virtual SAN dynamically self-tunes and load balances, adapting to changes in workload conditions to ensure that each virtual machine has the storage resources it needs, as defined by its policy. The end-to-end vSphere integration and policy-driven approach automates manual tasks and makes management of compute, storage and networking extremely easy.

- Elastic, linear scale out or up: Virtual SAN architecture allows for elastic, linear and non-disruptive scaling. Capacity and performance can be scaled simultaneously by adding new hosts (scale-out); or capacity and performance can be scaled independently by adding new drives to existing hosts (scale-up).

- Lowest TCO: The “grow-as-you-go” scaling approach as well as the ability to use commodity hardware, implies that the overall TCO of the HCI solution is dramatically lower than using standalone hardware or converged systems.

- Choice of deployment options: VMware provides the broadest choice of deployment models with multiple options across vendors. Certified hardware through Virtual SAN Ready Nodes guarantee flexibility while the EVO family of Integrated Systems is designed for simplification of procurement, deployment and management.

Introducing Cohesity DataPlatform – The Only Web-Scale Platform Designed to Consolidate All Your Secondary Data And Workflows.

What is Cohesity DataPlatform?

Cohesity DataPlatform is the only platform designed to consolidate all your secondary data and workflows. Inspired by web scale architectures, Cohesity provides a scale-out, globally deduped, highly available storage fabric to consolidate your secondary data, including backups, files, and test / dev copies.

Cohesity DataProtect is an end-to-end backup and recovery solution that is fully integrated with the platform to provide simple data protection with 15 minute RPOs and instantaneous RTOs.

The platform is deployed on Hyperconverged Nodes, which provide both storage and compute to enable fast data operations to execute directly on the platform.

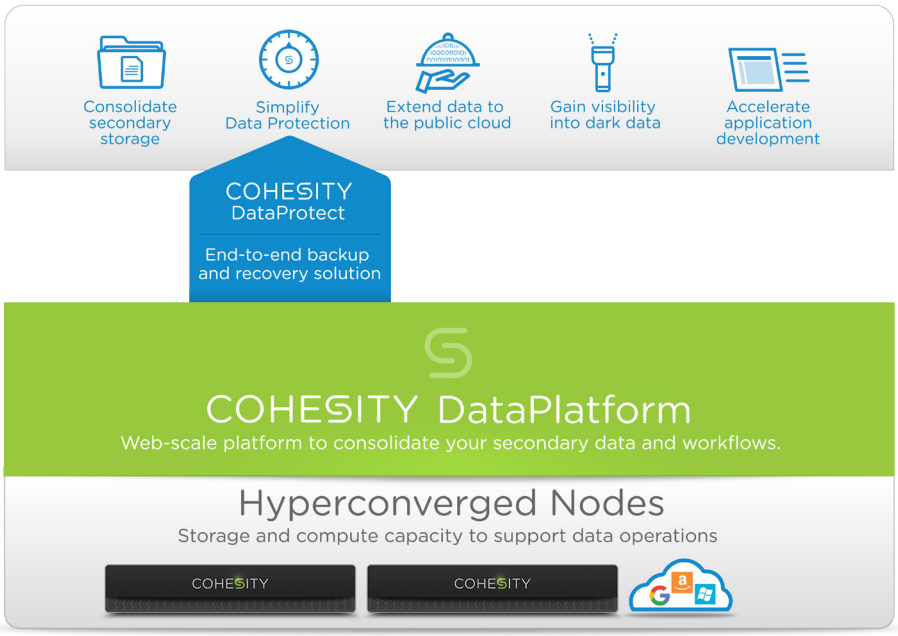


Figure 5: Cohesity DataPlatform and DataProtect

Key Use cases of DataPlatform

Consolidate secondary storage: Eliminate storage silos by consolidating secondary data, including backups, files, and test / dev copies, on a scale-out, globally deduped storage platform. Increase space and cost efficiency, simplify management and capacity planning, and eliminate the need for costly data migrations.

Simplify data protection and recovery: Simplify your data protection infrastructure with an end-to-end backup and disaster recovery solution that is fully converged on the Cohesity platform. Infinite backups. Instant recoveries. Integrated replication for multi-site disaster recovery.

Improve economics with public cloud integration: Extend your data platform to the public cloud for long-term archival, tiering of storage capacity, and disaster recovery. Make use of public cloud economics and flexibility without complicated gateways.

Gain visibility into your dark data: Shine a light on your dark data with real-time analytics on data utilization. Extract valuable insight from your data by running custom queries directly on the Cohesity platform.

Accelerate application development: Release applications faster by instantly cloning and provisioning test and dev environments on the Cohesity platform.

Key features of Cohesity DataPlatform

1. **Scale-out storage fabric:** Replace individual storage appliances with a web-scale platform. Ensure high-availability with no single point of failure, scale performance and capacity linearly, simplify capacity planning, and eliminate expensive forklift upgrades.

2. **Hyper-converged architecture:** Bring compute to your data instead of moving copies of data. Run secondary data workflows including data protection, recovery, test / dev and analytics directly on the Cohesity platform.

3. **Support for backups, files, and test / dev copies:** Consolidate your secondary data on the Cohesity platform with support for NFS and SMB interfaces. Consolidate backups, files, and test / dev copies.

4. **Global deduplication and compression:** Increase storage efficiency with variable-length, global deduplication that spans an entire cluster. Support for both in-line and post-process deduplication.

5. **SnapTree for unlimited snapshots and clones:** Take an unlimited number of snapshots and clones without impacting performance. Each snapshot is instantly available as a fully hydrated copy of the file.

6. **Remote replication for disaster recovery and migrations:** Protect your data off-site and enable disaster recovery / migrations to remote sites, with built-in remote replication. Leverage flexible replication topologies including site-to-site, one-to-many sites, cascaded, and to the cloud.

7. **CloudTier:** Leverage compelling cloud economics with native cloud integration for data tiering. Choose your preferred provider with support for Google Cloud Storage Nearline, Microsoft Azure, and Amazon S3.

8. **CloudArchive:** Replace tape archive and off-site data protection with long-term archival to all the leading cloud providers including Google Cloud Storage, Microsoft Azure, Amazon S3 and Glacier.

9. **Real-time analytics:** Monitor capacity utilization and trends in real-time at the cluster, VM and file level to better plan for future capacity requirements.

10. **Automated global indexing:** Automated global indexing powers Google-like search, enabling instant wildcard searches for any VM, file, or object ingested into the system.

11. **Quality of Service:** Control Quality of Service at a granular level to ensure low-latency, high throughput to mission critical workloads.

12. **Software-based encryption:** Secure your data with software-based encryption and data isolation between tenants.

13. **REST APIs for orchestration:** Orchestrate your data operations with a complete set of REST APIs and associated documentation.

Cohesity DataProtect

Cohesity DataProtect converges all your data protection infrastructure on Cohesity DataPlatform – including target storage, backup, replication, disaster recovery, and cloud tiering. Cohesity DataProtect ensures strict application SLAs by providing fast Recovery Points and near-instaneous Recovery Times, all while cutting data protection costs up to 50%.

Simplify data protection: Simplify your data protection infrastructure with an end-to-end data protection solution that converges backup software, backup infrastructure and target storage on one unified platform.

Improve SLAs: Ensure 15-minute Recovery Points and near-instantaneous Recovery Times

Reduce costs: Cut data protection CAPEX and OPEX by 50% by eliminating the need for the traditional patchwork of data protection solutions.

Key features of Cohesity DataProtect

1. **Integrated backup and recovery solution:** Simplify your data protection infrastructure with an end-to-end backup and recovery solution that is fully converged on Cohesity DataPlatform. Eliminate the need for separate backup software, proxy servers, media servers, replication, disaster recovery and target storage.

2. **Recovery Points of minutes – not hours:** Reduce your Recovery Points to 15 minutes or less by leveraging the high ingest throughput of Cohesity DataPlatform combined with SnapTree for unlimited snapshots and clones.

3. **Instantaneous Recovery Times:** Recover applications instantly by creating a clone of the backup VM and running that clone directly from the Cohesity platform. If needed, the clone can be moved back to primary storage using Storage vMotion.

4. **Fast backups:** Minimize your backup windows by parallelizing backup jobs on the scale-out nodes of the Cohesity Data Platform.

5. **Application-consistent backups:** Perform application-consistent backups with agentless app-consistent backups for virtual environments, and application adapters for physical SQL Server, Windows, and Linux servers.

6. **Full integration with VMware:** Simplify operations with full vSphere integration. vCenter integration provides a full view of the Virtual Machines and enables admins to apply data protection policies on a per-VM basis. Cohesity leverages the vSphere APIs for Data Protection to provide vSphere snapshot-based protection and eliminate the need for in-guest agents.

7. **Policy-based management:** Create policies that specify your application SLA requirements including RPO, retention policies, off-site replication and cloud archival. Assign policies to VMs based on application SLA requirements.

8. **Instant file-level search:** Instantly find your virtual machine and file data with Google-like wild-card search on Virtual Machines and individual files to accelerate recovery times.

9. **VM, file and object-level recovery:** Recover individual VMs, restore files to source VMs, and recover individual application objects for Exchange, SQL, and SharePoint.

10. **Tape archival:** Support integrated tape-out for long-term data archival.

11. **Datastore throttling:** Minimize performance impact on primary storage by modulating ingest rate at the datastore or vCenter level.

Joint Solutions: VMware Virtual SAN + Cohesity DataPlatform

The architecture in Figure below depicts a hyperconverged infrastructure solution with both Virtual SAN and Cohesity DataPlatform. In this example, we show three different VSAN clsuters being protected by a single Cohesity DataPlatform cluster.

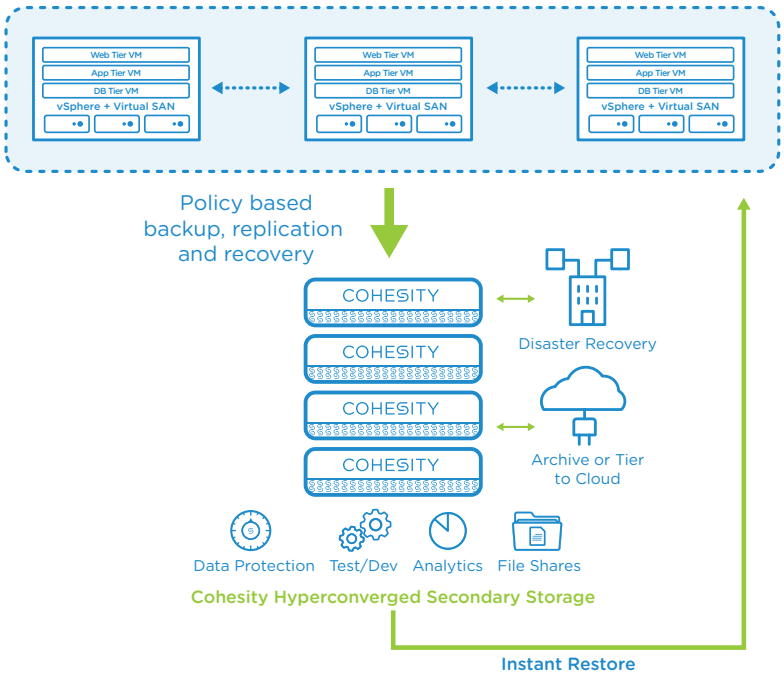


Figure 6: Simpler data protection for Virtual SAN with Cohesity DataPlatform

Benefits of VMware Hyper-Converged Software with Virtual SAN and Cohesity DataPlatform

Web-scale, pay-as-you-grow architecture everywhere

VMware Hyper-Converged Software with Virtual SAN delivers advanced storage services for virtual machines including ash-optimized performance, built-in distributed RAID, Quality of Service (QoS), and storage efficiency through deduplication, compression and erasure coding.

It also integrates natively with vSphere High Availability (HA), Fault Tolerance (FT) and vSphere Replication for asynchronous replication. The solution can handle 100,000 IOPS per host and is designed to meet the requirements for most enterprise applications.

Cohesity complements Virtual SAN with an incremental pay-as-you-grow distributed, web-scale platform for secondary storage. 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. This eliminates the need to invest in hefty upfront capital expenses for anticipated future growth and removes many hidden costs associated with upgrades of traditional scale-up systems.

Dynamic, application-centric operations with storage policy-based management everywhere

Virtual SAN is radically simple storage that streamlines storage provisioning for vSphere environments — storage can be provisioned in just a few clicks in the standard vSphere Web Client. Automation through VM-centric policies ensure SLAs can be set or changed on the fly. Integrated management capabilities simplify monitoring, troubleshooting and reporting. Cohesity's native data protection software complements this simplicity and enables businesses to easily protect data for virtual environments on Virtual SAN; dramatically reducing cost and complexity.

Cohesity DataProtect is tightly integrated with VMware vCenter and can instantly see a full index of the virtual environment, allowing the user to effortlessly choose which virtual machines to protect. The user can set a number

of protection policies with the required service levels, including RPO, retention policies, and off-site protection. Each application receives the level of service it requires by associating these policies with specific VMs. Cohesity DataProtect automatically ensures that each application is protected with the appropriate service levels.

Cohesity DataProtect leverages 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. The Cohesity dashboard also provides an overview of Data Protection jobs, System health, 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.

Eliminate complexity with a unified platform for end-to-end data protection

Cohesity consolidates all the unnecessary silos across the entire lifecycle of data protection (backup, restore, replication, DR, archival, and cloud tiering) for virtualized Virtual SAN environments. The Cohesity platform eliminates fragmentation in data protection by eliminating disparate master & media servers, backup software & appliances, replication software and cloud gateways. All this functionality is integrated in the Cohesity solution that scales with the growing secondary storage needs. This avoids the need to invest in 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.

Ensure fast Recovery Points and near-instantaneous Recovery Times

Cohesity's patented SnapTree technology for snapshots allows 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 leverage the Distributed Redirect-on-write (DRoW) technique to keep track of changes by writing the changed data to new blocks.

Leveraging SnapTree, each VM can be backed up incrementally with an RPO of less than 15 minutes. Each incremental backup is kept as a fully hydrated snapshot in SnapTree.

Virtual Machines can be recovered instantly by creating a clone of the backup VM and running that clone directly on the Cohesity platform. If needed, the clone can be moved back to primary storage using Storage vMotion.

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 accessing backup data with a simple text-based search and restore.

This restore can also place VMs, files and objects in the original Virtual SAN source location further reducing the burden associated with managing restore processes.

Lower total cost of ownership for both primary and secondary storage

Virtual SAN reduces CapEx needs by leveraging inexpensive, industry-standard server components and by providing storage efficiency features such as deduplication. Virtual SAN also lowers OpEx by automating and managing storage and compute resources from the same virtualized platform. Virtual SAN can lower TCO by up to 50% and deliver all-flash solutions for as low as half the price of competitive hybrid Hyper-Converged infrastructure systems.

Cohesity DataPlatform enables users to consolidate their backup data on a scale-out, globally deduped storage platform with compression and zero-cost snapshots and clones. This approach increases space and cost efficiency, ensures high-availability with no single point of failure, scales performance and capacity linearly, and eliminates expensive forklift upgrades.

Cohesity DataProtect further reduce CapEx by providing an end-to-end backup and recovery solution integrated with the platform, eliminating the need for separate backup software licenses and associated infrastructure.

Cohesity also has cloud connectivity to either tier or archive data to the public cloud, enabling users to benefit from compelling economics of the public cloud.

Consolidate backups, files, and test/dev copies

Virtual SAN is designed to provide primary storage for vSphere VMs, but doesn't support file access and is not optimized for secondary storage use cases. Cohesity DataPlatform is the perfect complement to Virtual SAN. It provides NFS and SMB interfaces to consolidate not only backup data, but also files and test/dev copies, thereby eliminating the need for dedupe appliances and expensive NAS devices.

Accelerate application time-to-market with instantaneous provisioning of clones for test/dev

In legacy approaches, protected data isn't productive, because it is sitting in a dedupe appliance that doesn't support read/write activities directly on the protected data. Cohesity DataPlatform makes your protected data productive. Snapshots and clones can be taken instantly and VMs can be run directly from the platform.

This is particularly useful in test/dev use cases. An instant, zero cost clone can be taken from a backup image, and presented to any Virtual SAN cluster for rapidly spinning up test/dev environments. Developers gain instant access for dynamic development, QA, and staging of their applications.

Built-in analytics capabilities

The Analytics capabilities built into the Cohesity Data Platform unlock the vast potential of the backed up data sets. These capabilities ensure that the data doesn't have to be moved out of the cluster but can be analyzed in-place using the Elastic Search and MapReduce framework that are integral to the Platform.

All the data ingested into the system is automatically indexed – including the VMs and the files within the VMs. The indexing engine, powered by Elastic Search can immediately provide deep storage-level information such as storage, file-level and VM-level metrics.

- **Google-like search:** By indexing all the content upon initial ingestion, users have access to Google-like search for any VM or file in the Cohesity platform.
- **Storage Metrics:** Storage utilization, available capacity and data growth trend analytics.
- **File-level Metrics:** Comprehensive file-level information such as file-type and user access history to gain better understanding of how storage is being used in a particular environment.
- **VM-level Metrics:** Dashboard to show storage consumption by application, and data change rate that coupled with a predictive engine anticipates future storage needs]

However, the capabilities go well beyond VM and file-level information exposed through pre-built reports. The platform makes it possible to run the most complex MapReduce computations within the platform without requiring external compute. All this power is unleashed through Cohesity Analytics Workbench (AWB). Users can run complex computational tasks for such diverse use cases as eDiscovery, Compliance and Threat Analytics. The platform QoS capabilities ensure that the AWB workloads can be prioritized as needed and run in the background.

The analytics capabilities bring compute to data and as a result extracts value from a data asset that is sitting idle most of the time.

Deploying Cohesity Data Management Platform for VMware Virtual SAN Infrastructures

The Cohesity C2000 Appliance Series delivers a Hyper-Converged secondary storage with a web-scale architecture and standards-based hardware to deliver pay-as-you-go simplicity with the power to store, protect and analyze the secondary data in one place.

The appliance initial configuration starts with a single 2U block of 96TB raw storage capacity for data protection, which converges scale-out backup with global dedupe, instant recovery, replication, and archival to tape and cloud. The deployment, and use of the Cohesity Data Management Platform with Virtual SAN is detailed below.

VMware Virtual SAN cluster setup

Deploying VMware Virtual SAN is a radically simple process. This enterprise-class storage solution for virtual machines can be prepared and configured with just a few mouse clicks. As shown in the image below, the first step is to turn on Virtual SAN on an eligible vSphere Cluster.

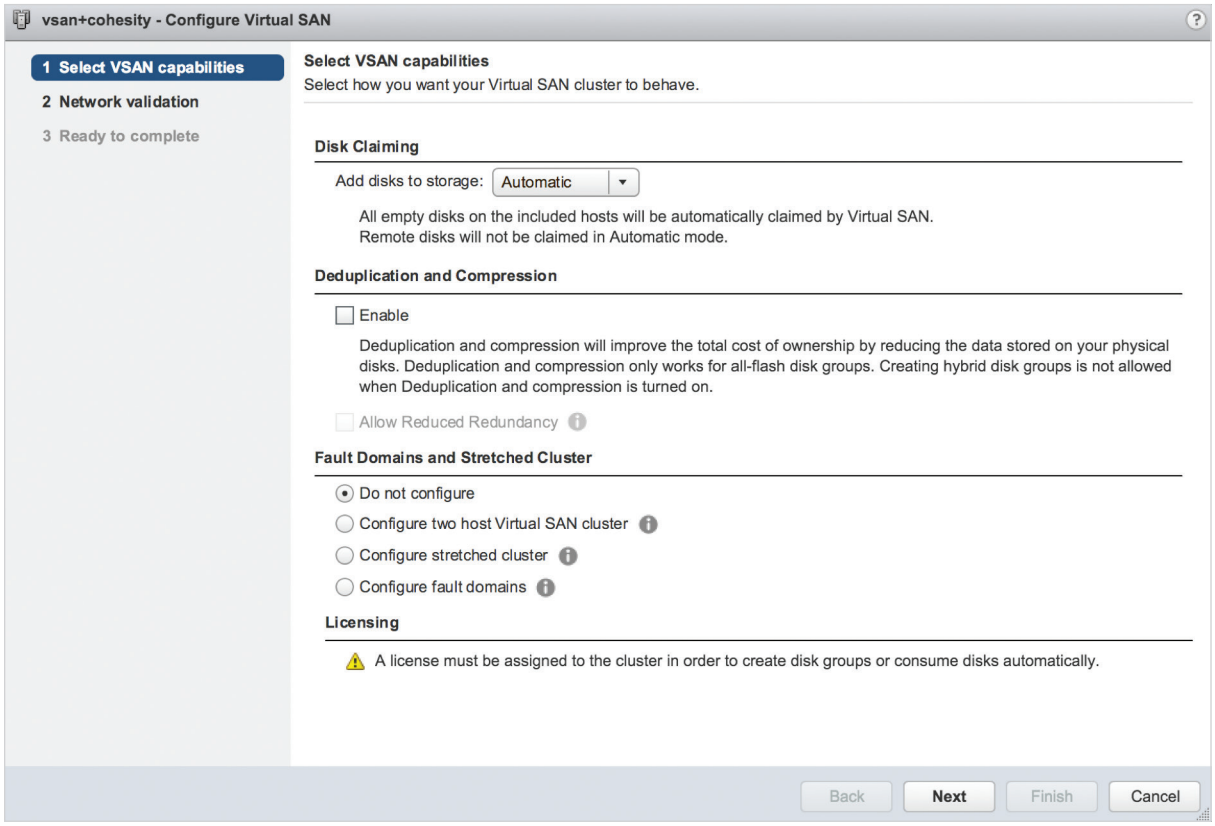


Figure 7: Configuring Virtual SAN

Virtual SAN provides a choice of hardware options for customers – it is critical to ensure that the hardware is supported by Virtual SAN by checking the VMware Compatibility Guide. Detailed instructions on configuration of Virtual SAN can be found on VMware documentation pages.

http://www.vmware.com/products/virtual-san/resources.html

Storage Policy Based Management (SPBM)

vSphere utilizes a policy driven, virtual machine-centric control plane to dynamically allocate and compose the storage services on Virtual SAN. Storage Policy-Based Management or SPBM is the native storage policy framework for vSphere.

To streamline and simplify Virtual SAN's operational model, VM Storage Policies, a component of the SPBM framework, are created to determine storage capacity, performance, availability, fault tolerant methods (space efficient or performance based), and others throughout the lifecycle of a virtual machine.

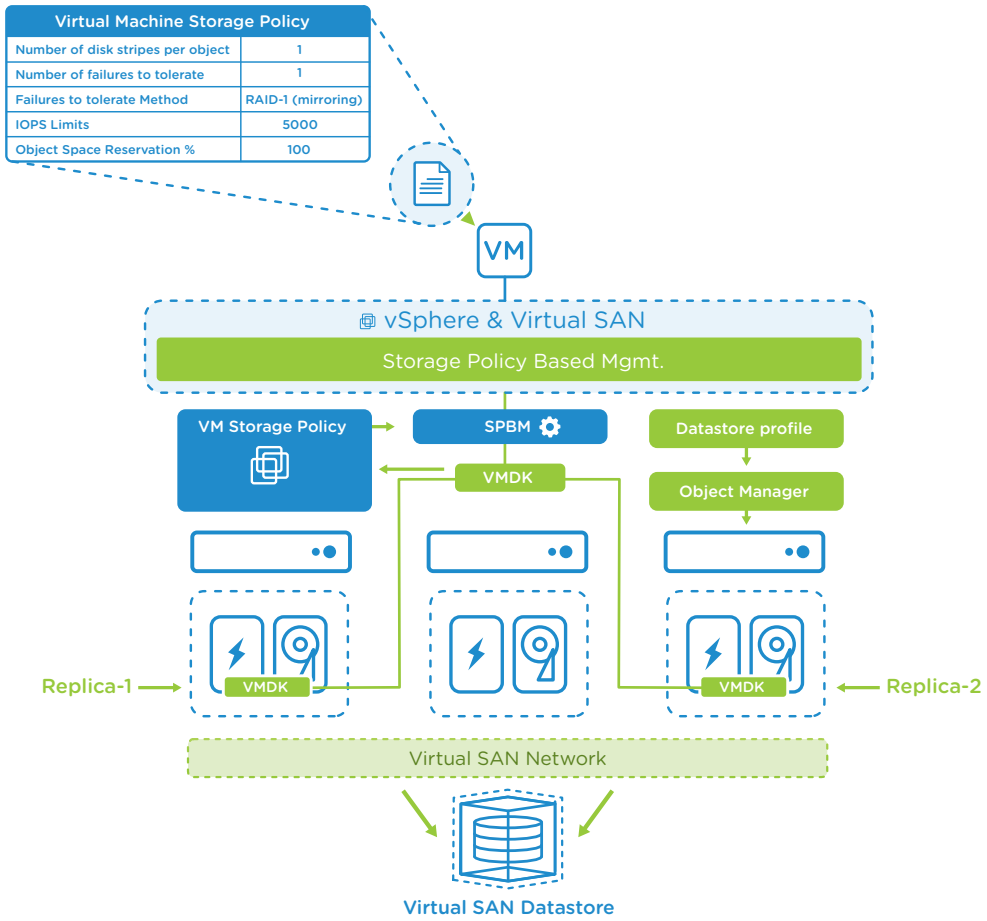


Figure 8: Storage Policy-Based Management for Virtual SAN

All of the VM-centric features, and data services in Virtual SAN are all accessible under the VM Storage Policy component. Virtual SAN comes with a built-in VM Storage Policy (Virtual SAN Default Storage Policy) with predefined availability and data distribution settings. The features and data services configuration of the Virtual SAN default storage policy are illustrated in the figure below.

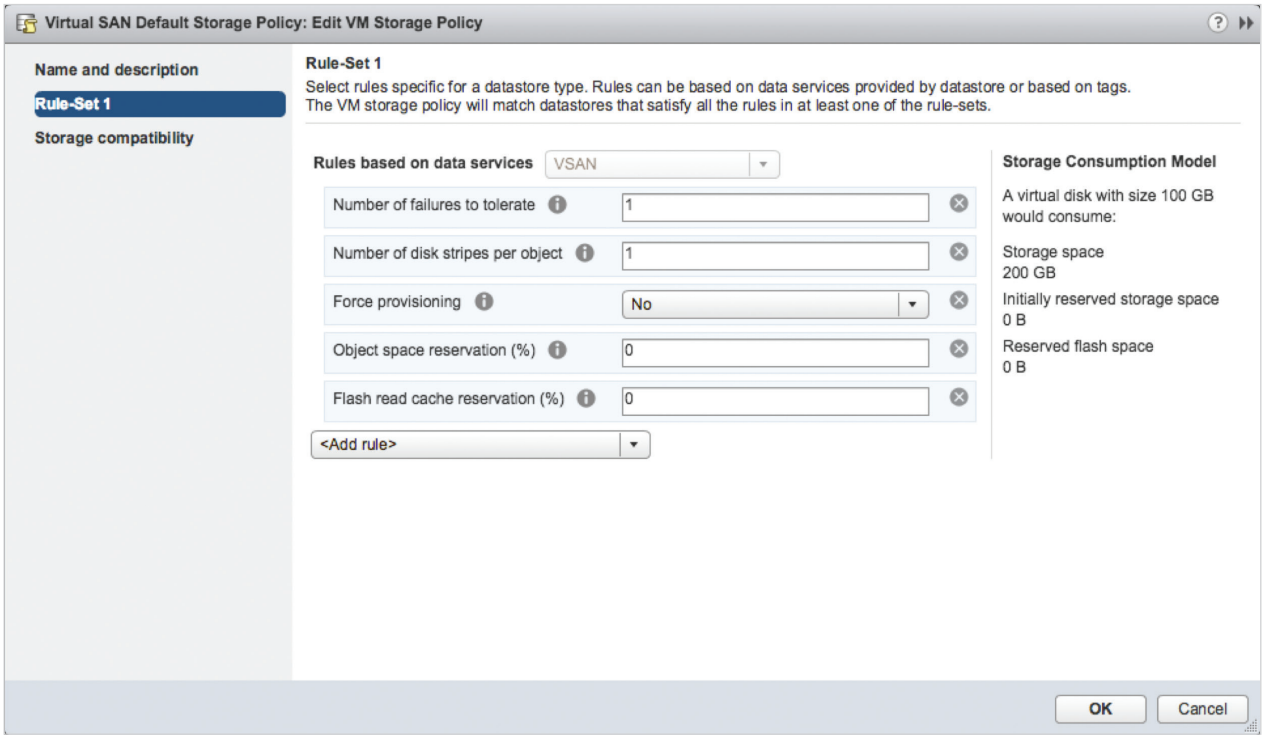


Figure 9: Storage policies for Virtual SAN

VM Storage Policies are created and assigned to virtual machines and their individual disks at any time during provisioning or post provisioning operations. Any workload deployed on to the Virtual SAN datastore will be assigned the systems default policy if one is not individually assigned.

Cohesity cluster setup

The configuration and deployment of the Cohesity cluster involved three main steps:

Front

Front

I. The Cohesity CS2000 Series appliances are physically racked and stacked

II. Redundant 10GbE network ports from each node are connected up to 2 separate top of rack switches

III. Cohesity software is installed on the nodes using the Cohesity install ISO image

As the Cohesity CS2000 Series appliances are powered on and the Cohesity software gets initialized, each of the appliances broadcasts a pre-programmed node name over the network that is utilized to setup and form the Cohesity cluster.

Network Switching Requirements

I. All of the IPMI interfaces on the Cohesity CS2000 Series appliances that will participate in the same Cohesity cluster need to be configured on the same network IP subnet and VLAN.

II. All data interfaces on the Cohesity CS2000 Series appliances that will participate in the same Cohesity Cluster also need to be on the same network IP subnet and VLAN.

III. All switch ports connected to IPMI interfaces of the Cohesity CS2000 Series appliances are required to be configured as access ports.

IV. All switch ports connected to data interfaces of the Cohesity CS2000 Series appliances are required to also be configured as access ports.

V. Validate that Multicast (L2) has been enabled on the physical network devices.

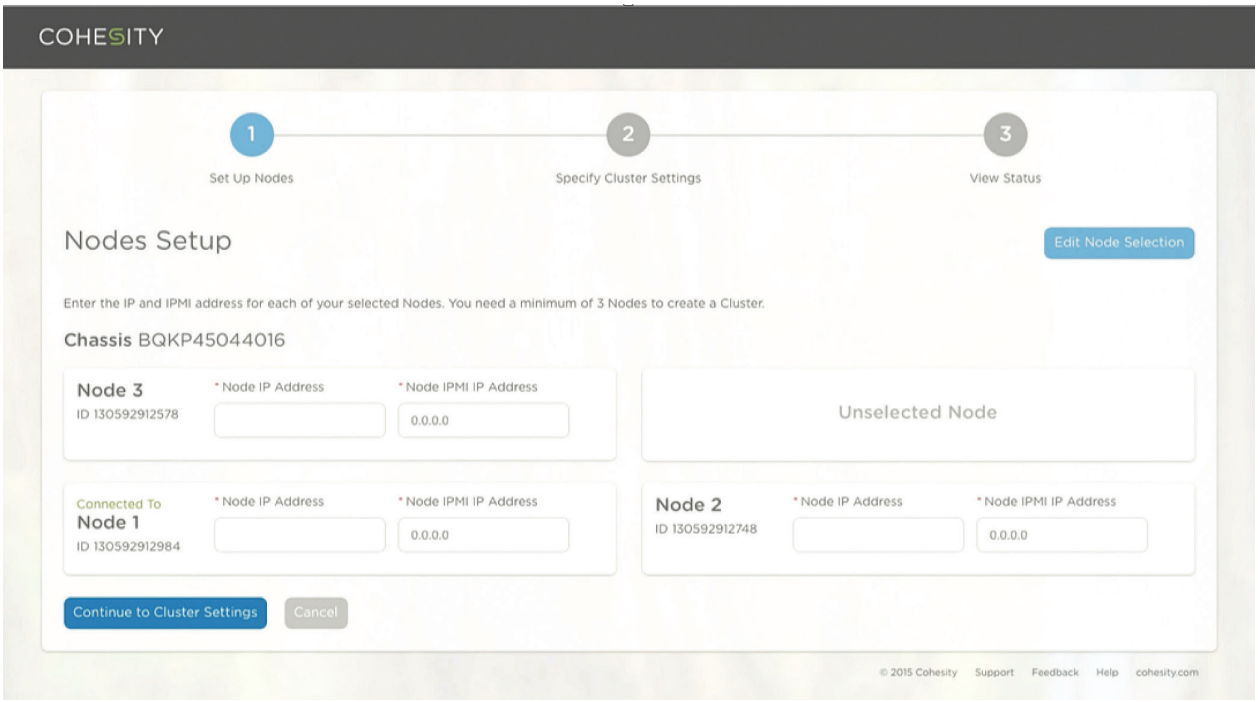


Figure 10: Network configs for Cohesity

Cohesity Protection Policy

Cohesity Protection Policies provide a reusable set of settings that define how and when Virtual Machines are protected, replicated and archived. A Protection Job uses the schedules and settings defined in the selected Policy to determine when Snapshots of the virtual machine are captured, replicated and archived.

Cohesity DataProtect provides and set of built-in system policies (Gold, Silver and Bronze) with different snapshot frequencies, retention policy and Service Level Agreements (SLA). In addition, administrators have the ability to create custom policies and defined their own custom policies and SLAs.

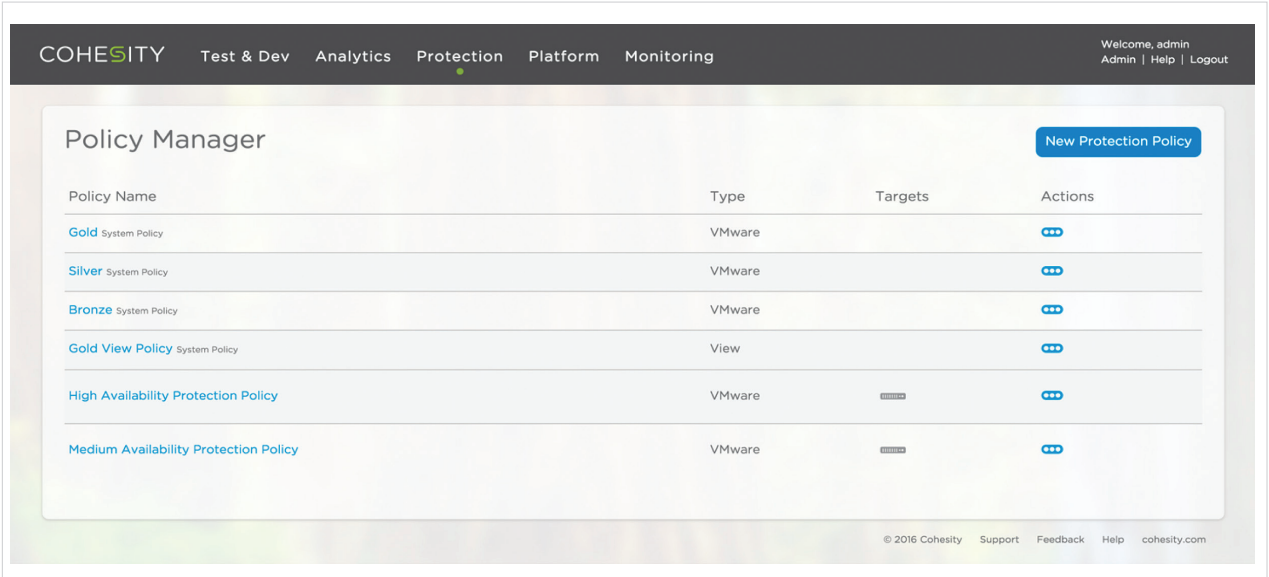


Figure 11: Cohesity protection policies

Cohesity default policies artifacts and their definitions

Gold	Silver	Bronze
Snapshot every hour	Snapshot every 6 hours	Snapshot every day
Snapshots retained for one year	Snapshots retained for 180 days	Snapshots retained for 180 days
Job Priority is high	Job Priority is Medium	Job Priority is Low
SLA is 10 minutes for regular backups and 120 minutes for Full backups	SLA is 30 minutes for regular backups and 180 minutes for Full backups	SLA is 60 minutes for regular backups and 240 minutes for Full backups

Cohesity's Protection Policy framework is conceptually very similar to the vSphere Storage Policy-Based Management framework utilized by Virtual SAN for storage infrastructure management, consumption, and operating model.

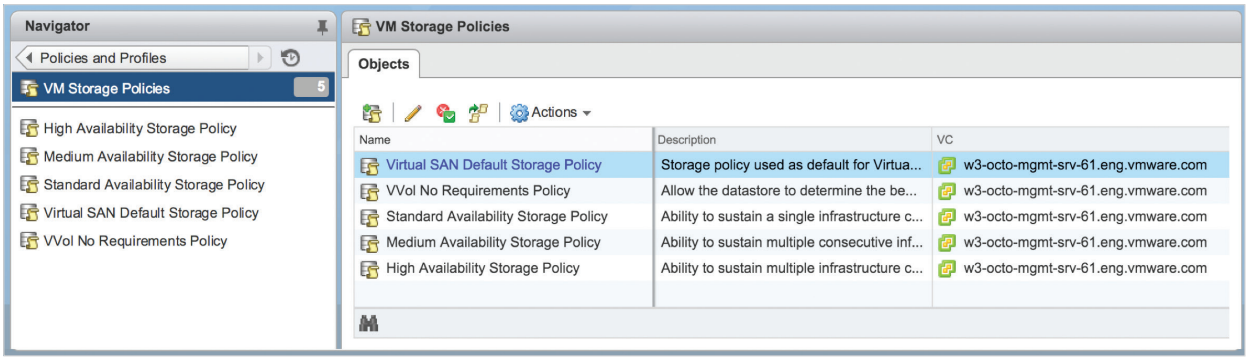


Figure 12: VM storage policies in SPBM

vCenter Server Discovery

Cohesity DataProtect can protect any source that is registered and supported on the platform. Registering a vCenter Server as a source that is running or managing a Virtual SAN infrastructure can be done on the 'Register Sources' UI page. This ensures that a vCenter Server is registered with Cohesity DataProtect.

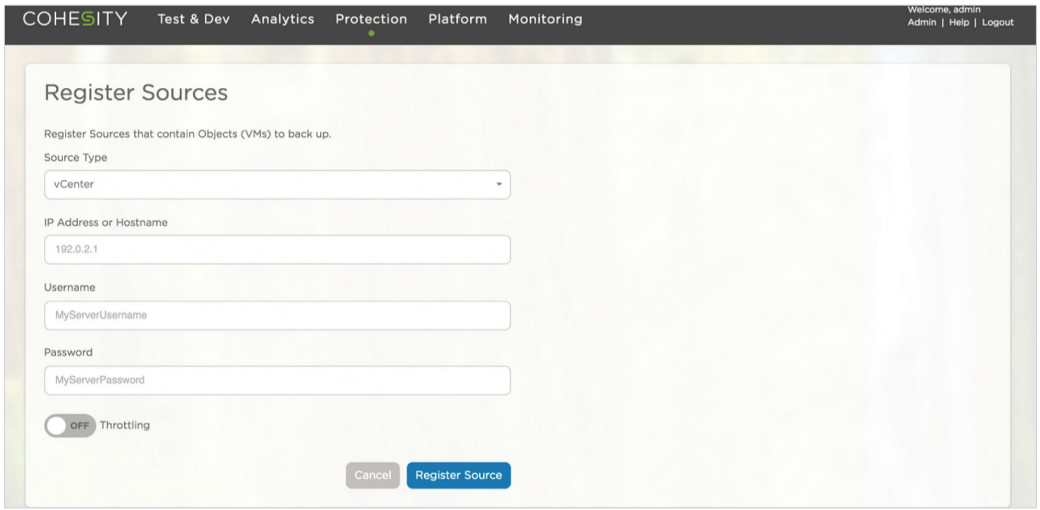


Figure 13: vCenter Server discovery

Storage Class of Service to Manage SLAs

In order to ensure that performance of the primary datastore is impacted minimally, as part of the data protection capabilities, Cohesity DataProtect offers Data Throttling mechanism that modulates the backup ingest performance over the production workloads at the vCenter Server or datastore level.

This is of particular value with multiple Virtual SAN clusters which require management for how and when data is pulled from the primary datastore to secondary infrastructure.

This is done by setting the latency thresholds prior to starting new Data Protection job runs and by monitoring the I/O response times associated with currently running Data Protection jobs. The Cohesity Cluster uses the statistics from Storage IO control (SIOC) to calculate the observed latency of datastore. The two available settings are:

- Latency Threshold for throttling the new tasks of Protection Job Runs—If the observed latency of a datastore is greater than the specified Latency Threshold, the Cohesity cluster throttles the processing of new Backup tasks using that datastore.

- Latency Threshold for throttling the currently running tasks of Protection Job Runs—If the observed latency of a datastore is greater than the specified Latency Threshold, the Cohesity cluster throttles any currently running Backup tasks using that datastore.

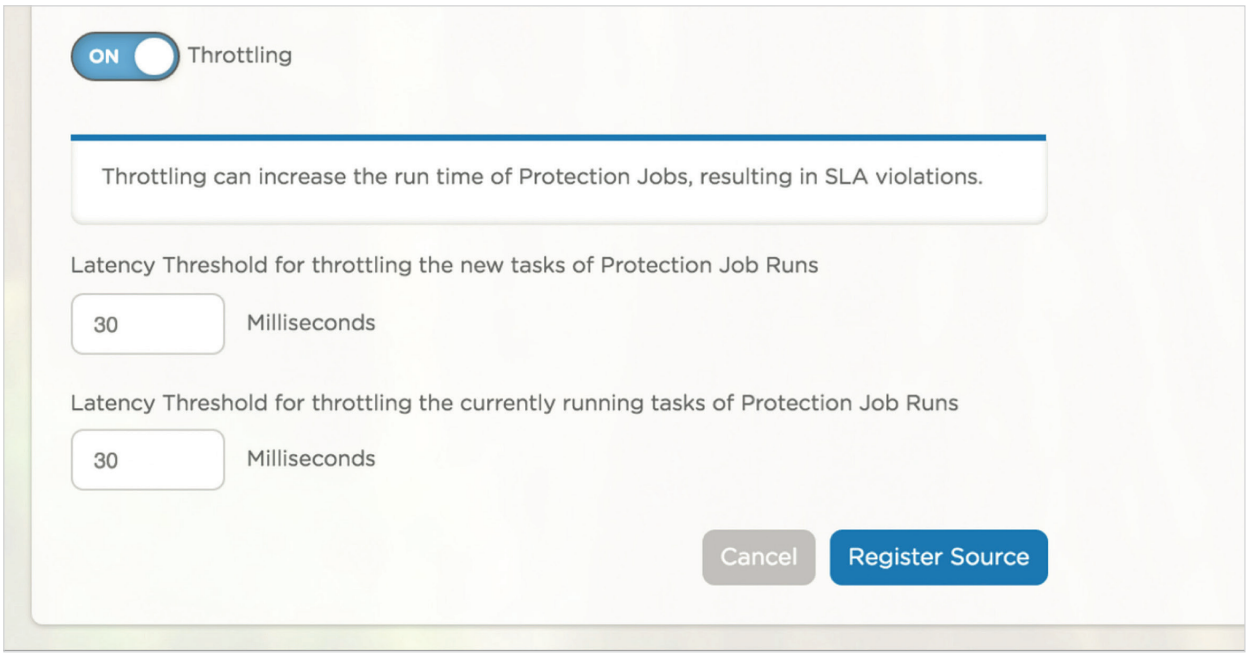


Figure 14: Storage class of service

Instant restores of a VM, File or entire vCenter environment using Cohesity SnapTree

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.

Because SnapTree is implemented on a distributed filesystem, every node sees the same nested structure of the chain with a fixed depth which provides the bene t of taking a large number of fully hydrated snapshots. Keeping the snapshots fully hydrated improves the recovery times (RPO & RTO) of any snapshot because it does not incur the time penalty of traversing the entire chain of changes.

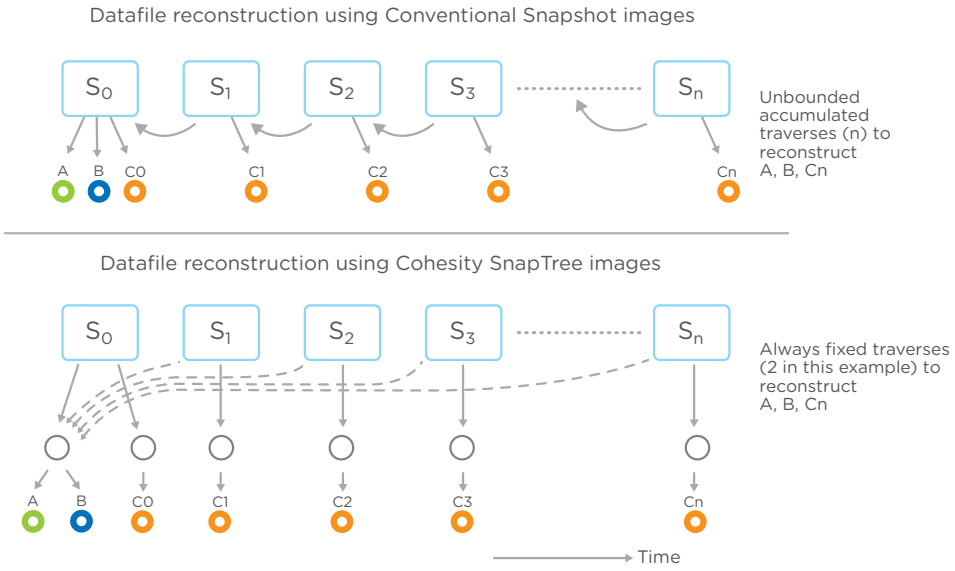
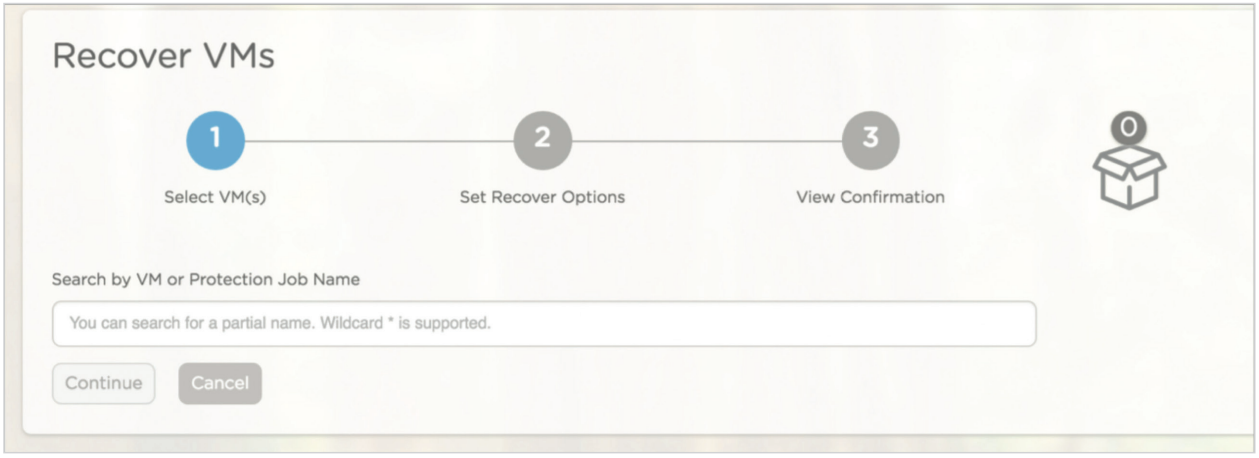
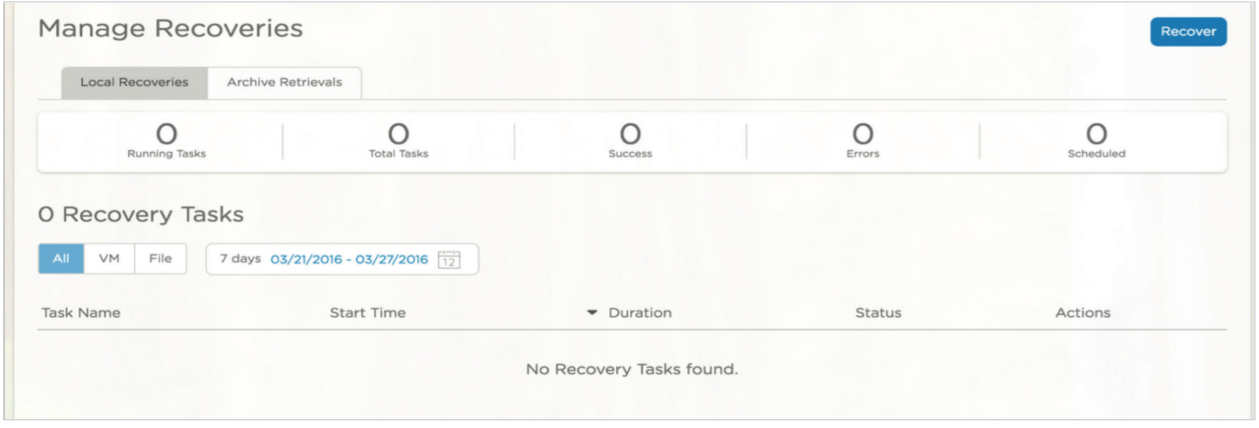


Figure 15: Cohesity SnapTree technology

Recovery Process

From the Manage Recoveries page on the Cohesity UI, the following recovery actions can be started - Recover a VM from Snapshots created earlier by Protection Jobs, Recoveries files from inside Snapshots that were created by Protection Jobs, Download VMX file of the last Snapshot of a Protection Job. The user also has the ability to use wildcard searches to recover a VM or a file in a VM. The Manage Recoveries page also provides an overview of the Recovery tasks that have already been created.





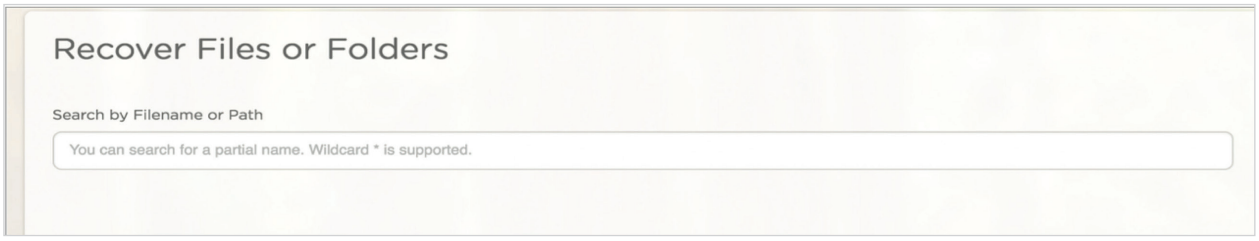


Figure 16: Cohesity recovery process

Replicating data across Cohesity clusters to protect against DR scenario

Organizations can achieve enterprise-level resiliency with site-to-site replication between Cohesity clusters. Replication can be configured at multiple granularity levels for maximum flexibility including Cluster-wide (all data on a single Cohesity cluster is replicated to one or more clusters), or replication job-level or Cohesity View level.

Replication makes copies of snapshots created by Protection Jobs located in a local Cohesity cluster and stores them in a remote Cohesity cluster. Replication pairing is done on the Platform > Replication Setup page.

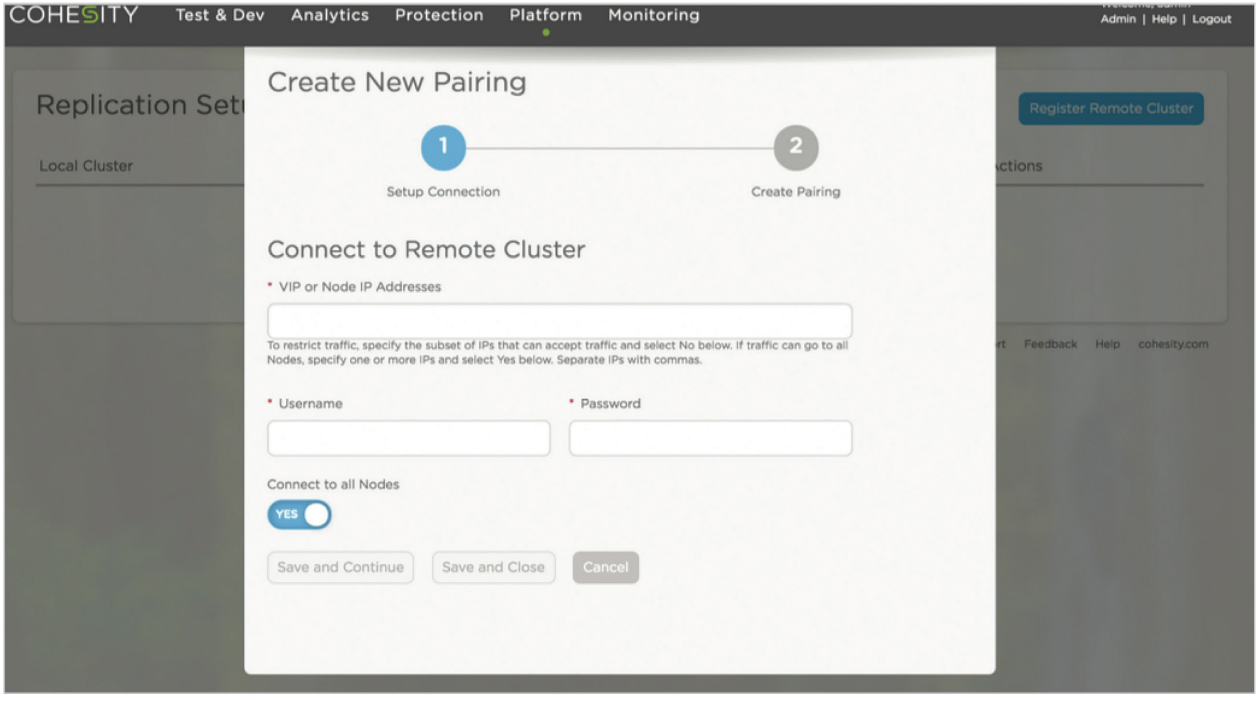


Figure 17: Cohesity recovery process

Data Protection for Archival

Cohesity DataPlatform 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-premise Cohesity infrastructure in one of two ways:

- CloudArchive – archiving the older local snapshots in the Cohesity cluster to cloud for long-term retention.
- CloudTier – using cloud as an extension to Cohesity's built-in storage tiers. This provides the ability to algorithmically decide when to tier the data between Cohesity cluster and the Cloud.

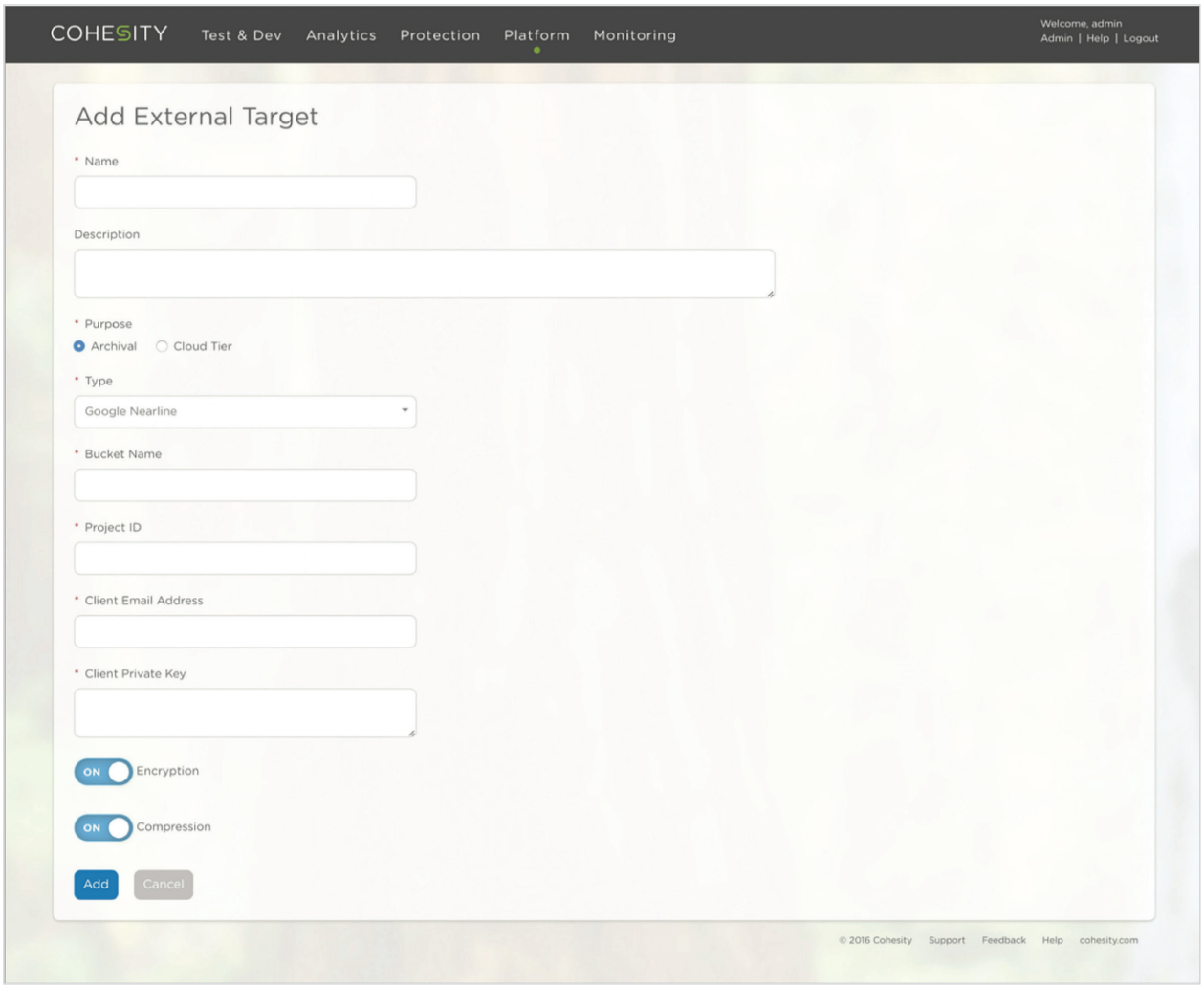


Figure 18: Cohesity CloudTier and CloudArchive

Ability to clone VMs and run on Cohesity DataPlatform for test/dev

Cohesity's platform empowers developers to instantiate the latest backup of the production application stack and run it directly off the Cohesity clusters. Cohesity DataPlatform provides Instant, zero-space clones capabilities that enable businesses to quickly spin up environments from a backup without any capacity overhead.

During the Cohesity Cloning process, new virtual machines files (such as VMDKs) are created from snapshots and stored in a View on the Cohesity Cluster.

This View becomes the Datastore for the newly restored virtual machines files. For low RTO, developers can instantiate any point-in-time snapshot and run directly on the Cohesity platform; but in order to meet the desired performance and SLAs, users can storage vMotion the virtual machines to the primary storage resources that is provided by any Virtual SAN cluster in their infrastructure.

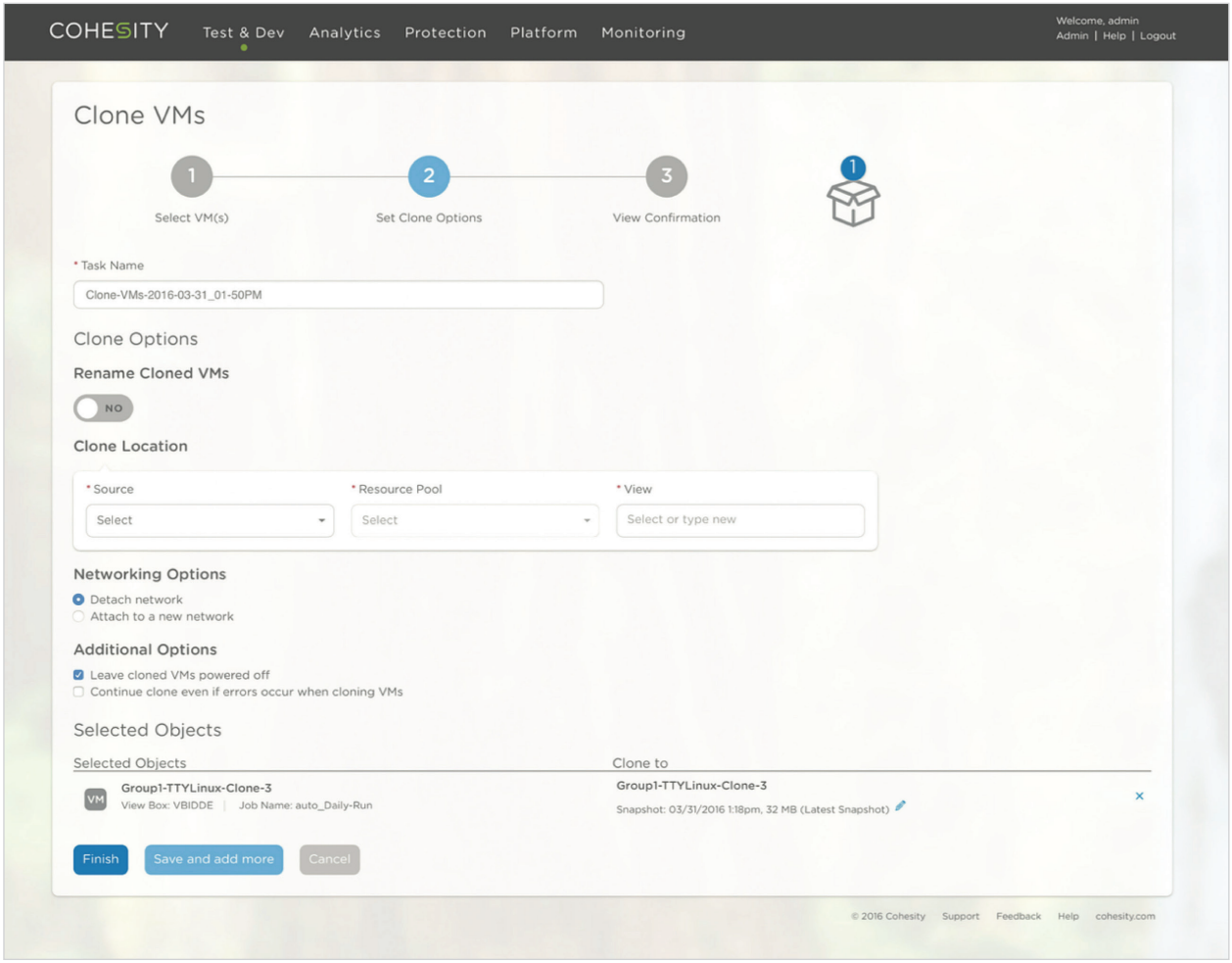


Figure 19: Cohesity cloning for test/dev

Conclusion

Cohesity DataPlatform and DataProtect are a web-scale platform to consolidate secondary data, data protection, replication, cloud tiering, and test/dev workflows. It is designed to meet the operational efficiencies, security, scalability requirements of the most demanding data management services in the enterprise. Unlike traditional data management solutions that require external media and management servers in order to backup, archive, and secure data, with Cohesity all the data management functions and services are performed directly on Cohesity DataPlatform.

The Cohesity platform enables the centralization and management of backed up, archive data (Dark Data). With Cohesity, organizations can intelligently use and secure the integrity of stored and archived data through the use of the platforms analytics capabilities.

Virtual SAN is a cost-effective and high-performance primary storage platform that is rapidly deployed, easy to manage, and fully integrated into the industry-leading VMware vSphere platform. Cohesity complements VMware Virtual SAN primary storage resources and provides scalable data protection for backup, archival and disaster recovery required for mission-critical enterprise applications or any running application on VMware Virtual SAN.

Together, Virtual SAN and Cohesity DataPlatform offer:

- **Pay as You Grow Scaling** - Modular growth capabilities enable easy scaling on both Virtual SAN and Cohesity solutions so you only buy what you need when you need it. Simply add more storage when you need more high performance primary storage, and add a single node or 2U block of 4 nodes to the Cohesity cluster when you need additional secondary storage capacity.

- **Plug and Play Installation & Policy driven Management** - Both Virtual SAN and Cohesity are designed to be turnkey, eliminating the need for timely and expensive professional services. Both solutions leverage intuitive management consoles and offer non-disruptive upgrades. Once primary workloads are running on Virtual SAN, policies can be set up in minutes on the Cohesity Data Platform that leverages VMware native VADP APIs for data protection. The Cohesity solution dramatically lowers RPOs and RTOs.

- **App-Specific storage class of Service** - While Virtual SAN is tuned to focus the fastest flash resources on your top tier applications, Cohesity follows up with application specific QoS policies for the backup services in your secondary storage. Combined, you are ensured that you are getting the best bang for your flash investment across your entire datacenter.

- **Fast RPOs and instantaneous RTOs** – Cohesity provides sub-15 minute RPOs and instantaneous RTOs leveraging SnapTree zero-cost snapshots and clones. Cohesity also provides cluster-wide VM, file and object-level search for rapidly identifying recovery targets.

- **Data Reduction** - Virtual SAN implements inline and post-process data reduction to enable effective storage reduction across a wide range of mixed workloads. Cohesity also offers both inline and post process dedupe and compression, configurable on a per workload basis.

- **Consolidation & Cost Savings** - With intelligent use of flash in Virtual SAN and Cohesity, you can eliminate storage sprawl and consolidate your data onto the best in class all- flash tier and the most efficient hyperconverged secondary storage platform. The consolidation enabled by Virtual SAN and Cohesity's approach will reduce datacenter footprint to achieve lower total cost of ownership. In addition, Cohesity offers native integration with all the leading cloud providers for data tiering and archival.

- **Support for backups, files, and test/dev copies** – Cohesity complements Virtual SAN with support for backups, files, and test/dev copies. This further eliminates storage sprawl by eliminating NAS appliances specifically used for file serving or test/dev environments.

- **Accelerate application time to market** – Cohesity makes secondary data productive. An instant, zero cost clone can be taken from a backup image, and presented to any Virtual SAN cluster for rapidly spinning up test/dev environments. Developers gain instant access for dynamic development, QA, and staging of their applications.

- **Built-in analytics** – Cohesity automatically indexes all the content ingested into the system, enabling Google-like search to rapidly indentify individual VMs or files. In addition, Cohesity provides detailed reporting on capacity and data utilization. Finally, the platform supports running MapReduce jobs directly on the platform for in-place, custom analytics on the data.

Authors

Rawlinson Rivera, VMware, Principal Architect in office of the CTO for Storage and Availability

Blog: http://blogs.vmware.com/virtualblocks/
http://www.punchingclouds.com/

Twitter: @PunchingClouds

Rawlinson specializes in cloud enterprise architectures, Hyper-Converged Infrastructures (HCI). Primarily focused on Software-Defined Storage products and technologies such as Virtual SAN, vSphere Virtual Volumes, vSphere API for I/O Filtering as well as storage related technologies and solutions for OpenStack and Cloud-Native Applications. Rawlinson serves as a trusted adviser to VMware's customers primarily in the US.

Rawlinson is among the few VMware Certified Design Experts (VCDX#86) in the world, and author of multiple books based on VMware and other technologies. He is the owner and main author of virtualization blog punchingclouds.com.

Gaetan Castelein, Cohesity, Head of Product Marketing

Twitter: @gcastelein1

Gaetan is head of Product Marketing for Cohesity, driving industry awareness for the Cohesity vision. Prior to Cohesity, Gaetan ran Product Marketing and Product Management at VMware in the Storage and Availability Business Unit. Gaetan has extensive experience working with vSphere, availability, and storage products.

Sai Mukundan, Cohesity, Product Management

Twitter: @saikmuk

Sai is in the Product Management team at Cohesity, specifically driving product integration for Cohesity in the file services and public cloud aspects. Prior to Cohesity, Sai worked at Microsoft in Product Strategy and Marketing for the StorSimple product line, and at Symantec in Product Management for the Storage and Availability business unit.