# COHESITY

# The Essential Guide to Modern Data Management

Protect, Consolidate, and Put Your Data to Work in the Multicloud Era

# Table of Contents

# The Roadblocks to Unlocking the Power of Your Data

Data is a critical element of business success, but in practice, few organizations manage their data as a strategic asset. Many IT teams struggle simply to meet basic SLAs for protection and availability, let alone leverage their data for competitive advantage.

Surveys have revealed an alarming range of issues associated with managing enterprise data that impact both business and IT, including budget overruns, poor customer service, security and compliance exposures, and even sinking morale within overworked operations teams.

The underlying cause of these issues is a phenomenon we call mass data fragmentation, and it has several profound aspects:
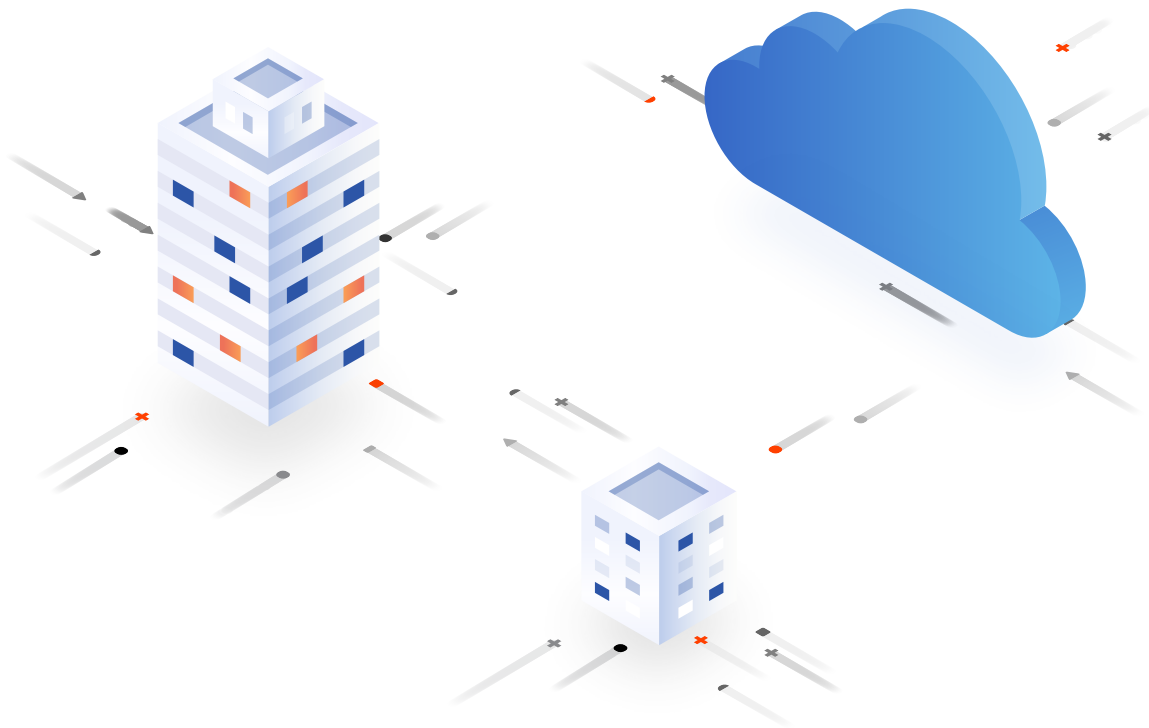


*Mass data fragmentation is one of the most severe impediments to business agility today.*

# Fragmented and Siloed Data Infrastructure

Data has exploded in volume and become scattered across multiple public clouds, data centers, remote offices, and the edge, with little global oversight. In each of these locations, data has become isolated in specialized infrastructure—often from multiple vendors—to manage basic functions such as backup, networking, storage, archiving, disaster recovery, dev/test, and analytics.

To make matters worse, there can be silos within silos. For example, a single backup solution can require several dedicated infrastructure components, such as master and media servers, target storage, deduplication appliances and gateways, in addition to the backup software itself—each of which may hold a copy of a given data source. It is not unusual to find four or more separate such configurations simply to manage backup for different data sources such as virtual machines, physical servers, databases, or containers.

It's worth pointing out that this fragmentation effect is not confined to traditional data centers. Public clouds are an increasingly popular choice to host data and apps, but IT has to deploy supplementary data management functions (typically from different vendors) to handle backups, DR, security and compliance, and so on, since the cloud providers don't automatically provide these services. And this even applies to 'service silos' in a purely SaaS environment, as we shall see later.



*Between multiple geographies, cloud providers, workloads, and data services, most of your organization's data is fragmented, unsecured, and unproductive.*

# Operational Inefficiencies

The complexity created by these infrastructure silos has a considerable knock-on impact on system and operational efficiency. There is typically no sharing of data between components or functions, leading to inefficiencies as multiple redundant copies of the same data are propagated between silos, taking up unnecessary storage space. Likewise, operational efficiency is compromised by the need to manage multiple proprietary systems with different UIs, each of which may require specialist administrators to maintain rather than generalists. Given today's world of 'no downtime,' tighter SLAs, increasing business demands, and slimmer budgets, it is no wonder that IT teams are reporting high levels of stress and even burnout as they grapple with the increasing complexity.

In fact, a recent survey commissioned by Cohesity found that a staggering 40 percent of IT staff time was spent on the installation, configuration and management of data infrastructure.

Data Management as a Service - A Comprehensive Market Study, 2020.

# Dark Data

A more sinister aspect of mass data fragmentation is that most IT teams don't have detailed knowledge about the majority of their data's contents, location, owner, access history, or whether it contains sensitive information. In other words, their data is dark. Petabytes of data are being routinely stored without being classified, indexed, or tracked.

Clearly, this adds considerable risk to the business. How do you prove compliance in handling PII, in order to meet regulations such as GDPR and CCPA? How do you detect anomalous user behavior or programmatic ransomware attacks? And from an operational point of view, how can you optimize expensive storage by deleting or archiving unneeded data when you don't know which items to keep?

# The Unfulfilled Promise of Public Cloud

As mentioned above, public cloud environments aren't necessarily a panacea for the challenges of fragmentation, inefficiency, and dark data. They can exacerbate them. Public clouds can introduce new infrastructure silos for data created by IaaS, PaaS, or SaaS apps, and protecting your cloud data is your responsibility—not the providers'. Inadequate protection can lead to missed SLAs or data loss.

Moreover, an organization's data is never situated entirely in just one cloud. For the majority of enterprises, their footprint straddles multiple public clouds as well as on-premises environments. Some data may need to stay in the data center for compliance reasons, while backup copies or a DR instance may be stored in IaaS. Developers working in another cloud may need to access data to test their applications, while specialist data feeds or big data analyses may come through a SaaS service from yet another provider. The multicloud era is truly upon us: for the majority of enterprises the majority of enterprises, their footprint straddles several public clouds and on-premises environments

# The Lack of Turnkey Data Management Services

There is also growing interest in another aspect of public cloud services—Data Management as a Service, or DMaaS—as a potential solution to these challenges. The idea of applying a SaaS model to a burdensome IT function would seem to offer several advantages, including lower administration effort, more cost predictability, the ability to scale on demand, and access to cloud-based apps to help derive additional value. In the aforementioned survey, it was found that almost 9 out of 10 IT leaders are looking for a turnkey as-a-service solution as an option for their data management.

However, the reality of the status quo does not deliver on the promise. Most SaaS solutions today that purport to offer data management in reality only offer a single service rather than a full range of data management capabilities, such as backup as a service, disaster recovery (DR) services, file and object management, archiving, security, and analytics. These extra services need to be procured from a host of different vendors, leading to the proliferation of service silos and the familiar signs of mass data fragmentation seen with legacy solutions.

This sprawl of data management services into multiple silos raises another important question—is it even possible to continually shore up and train IT staff to rein in your organization's exploding data? At the minimum, it would require a cross-functional, senior-level team across backup, security, storage, cloud, and devops to even begin to control your data, and heavy infrastructure investments to go with it. While your team is already consuming infrastructure and applications as a service, there seems to be no available services to consolidate, protect, and manage your organization's data at scale.

There is a severe lack of solutions that take a holistic approach to managing your company's data in a turnkey fashion—from its generation to its utilization—and you are instead forced to pile on siloed support and services contracts to your already siloed environment, further exacerbating your team's operational complexity.

# Data Management for the Modern Multicloud Era

You are not alone in facing the above challenges. There has been almost no fundamental innovation from the data management industry in decades. Basic techniques for performing backup and recovery, disaster recovery, cloud mobility, NAS storage, and copy data management have evolved only incrementally and within proprietary vendor silos. Until recently, IT teams have had no option but to perpetuate the traditional hardware-centric approach which only adds more complexity, risk, and cost rather than solving underlying issues. Moreover, most of these old solutions follow the legacy transaction model—you are made to pay a fat license fee up front, and then made to pay a recurring support and services fee, and repeat the same cycle after three years when your hardware is already obsolete, overburdened, and out of capacity.

Enterprises require a fundamentally new approach in order to break the cycle of technical debt, liberate IT and business teams from outdated architectures, and free trapped data to become an asset instead of a costly liability.

The breakthrough idea here at Cohesity is to leverage the approach taken by the hyperscale software giants such as Amazon, Google, and Facebook—who manage exabytes of the world's consumer data highly efficiently and scalably—and apply the same architectural principles to managing enterprise data, while providing an array of consumption and deployment choices to the customer.
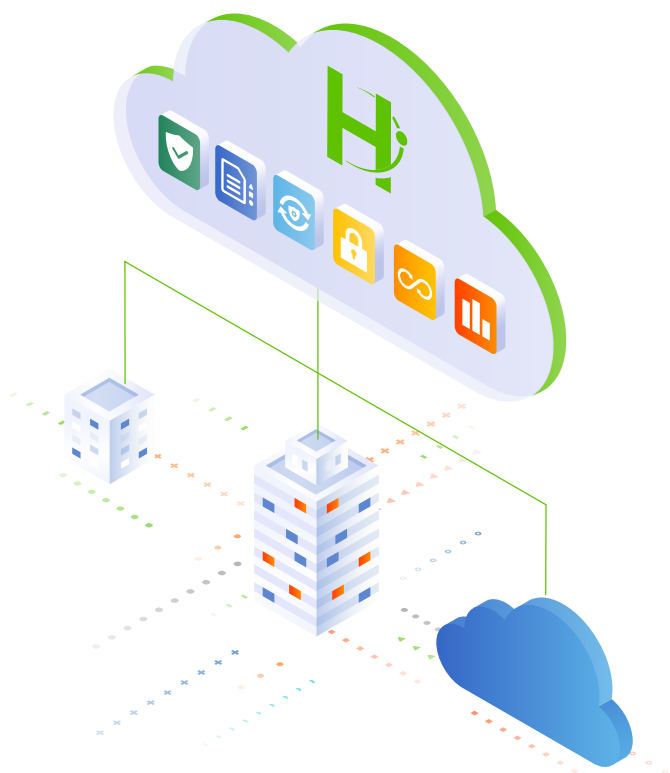
Cohesity simplifies enterprise data management with a single software-defined multicloud platform to consolidate all your data across data center, edge, and cloud, a unified management system, and simple, powerful services that you can easily deploy or consume.

# Cohesity – A Radical New Approach to Unlocking the Power of Your Data

The secret to the hyperscalers' success lies in their architectural approach, which has three major components: a distributed file system—a single platform—to store data across locations, a single logical control plane through which to manage it, and the ability to run and expose services atop this platform to provide new functionality through a collection of applications.

The Cohesity solution takes this same hyperscale architectural approach and adapts it to the specific needs of enterprise data management. A single software-defined multicloud platform to consolidate all your data across data center, edge, and cloud, a unified management system, and simple, powerful services that you can easily deploy or consume.

# Helios – A Multicloud Data Platform to Protect, Consolidate, and Manage All Your Organization's Data

Cohesity Helios enables the consolidation of all your data—from servers, databases, tape, cloud, data center storage, and applications—onto a single software-defined platform, eliminating the need for the complex jumble of fragmented infrastructure. Data is shared rather than siloed, stored efficiently rather than wastefully, and made visible rather than kept in the dark—simultaneously addressing the problem of mass data fragmentation while allowing both IT and business teams to holistically leverage its value.

| Backup and Recovery | File and Object | Disaster Recovery | Security and Compliance | Development and Test | Analytics and Insights |
| --- | --- | --- | --- | --- | --- |

## Cohesity Helios

| Data Center | Edge | Public Cloud |
| --- | --- | --- |

*Cohesity Helios enables all of your organization's critical data to be protected and made available to apps and services to derive additional value.*

Cohesity Helios provides unlimited scale, delivering the ability to harbor and index unlimited data—structured or unstructured, to always keep you a step ahead of your organization's exponential data growth. It gives you the ability to automatically store, index, protect, and make productive any new data added by your company.

Helios is also designed for flexible deployment and operations. Data within Helios is secure from inside out, with in-flight and at-rest encryption, immutability, WORM, role-based access, and multifactor authentication keeping your data safe. Helios also provides ML-powered security recommendations to streamline operations and defend against ransomware. The Helios platform also helps you control your costs and rein in your data volume by eliminating costly data copies with zero-cost clones.

## Key Benefits of a Unified Platform for Enterprise Data
Across data centers, edge sites, and public cloud environments.

**Unlimited Scale**
Hyperscale architecture easily keeps pace with your business and data growth.

**Always On**
Gain peace of mind whether managing yourself, or consuming as SaaS.

**Enterprise-grade Security**
In-flight and at-rest encryption, immutability, WORM, role-based access, and multifactor keeps your data safe.

**Built-in Machine Learning**
ML-powered recommendations to streamline operations and defend against ransomware.

**Optimized for Efficiency**
Eliminate costly data copies with zero-cost clones.

**Easy to Manage**
See and control your data across all locations from a single UI.

Next, Helios enables all aspects of enterprise data to be managed globally through a single GUI. This dramatically simplifies operations by replacing multiple specialized consoles with a single global dashboard that allows policies to be set and actions to be taken in one place—whether managed by you, or consumed as a service. Built-in machine learning reduces administrative burden by providing proactive actionable recommendations, what-if analyses, anomaly detection, and health checks that would otherwise go unnoticed or require additional manual effort. Uniquely, Helios also enables global search across all data/workloads/locations, cutting through previously disconnected silos to provide a single view into your entire data estate. This allows operators to quickly locate objects of interest, as well as audit and remediate data that needs attention.

The patented design of Cohesity Helios allows all data management functions—including backup and recovery, disaster recovery, long-term archival, file and object services, test data management, security and compliance, and analytics—to be run and managed in the same software environment at scale, whether in the public cloud, on-prem, or at the edge. But the platform's elegance goes even beyond, by allowing third-party apps and services to coexist with the platform's own. Cohesity users can simply download and install certified applications from the Cohesity Marketplace and immediately run them against data previously captured and indexed on the platform. The use-cases are endless—eDiscovery, compliance checks, analytics, data vulnerability scanning and much more can be conducted in the same environment, under the control of the same UI, with access to the full consolidated data estate in a single pass. Such an approach is an efficiency breakthrough compared to segmenting and copying data subsets to new instances of infrastructure to perform similar functions.



*Cohesity helps you protect, consolidate, and manage your global data estate - across clouds, workloads, and consumption models.*

> "
>
> By leveraging Cohesity's software-defined platform for data management, we substantially enhanced our operational capabilities at AutoNation, improved reliability, and lowered our TCO.
>
> **Adam Rasner,** Vice President Technology Operations, AutoNation
>
> **AutoNation**

# Under The Hood – SpanFS: The Infinitely Scalable Distributed File System Powering Helios

At the heart of Helios is a fully distributed, shared-nothing file system. Inspired by web-scale principles, Cohesity SpanFS, is meticulously architected to address the challenge of mass data fragmentation. To effectively consolidate data, enterprises need a file system that is simultaneously able to handle the requirements of multiple use cases. In order to meet modern data management requirements, Cohesity SpanFS provides the following:

**Unlimited scalability**

**Global indexing and search**

**Strictly consistent**

**Multi-tenancy with QoS**

**Multiprotocol (NFS / SMB / S3)**

**COHESITY SpanFS®**

**Sequential and random IO**

**Global dedupe**

**Multicloud**

**SnapTree® limitless snaps & clones**

**Automated tiering**

**Self healing**
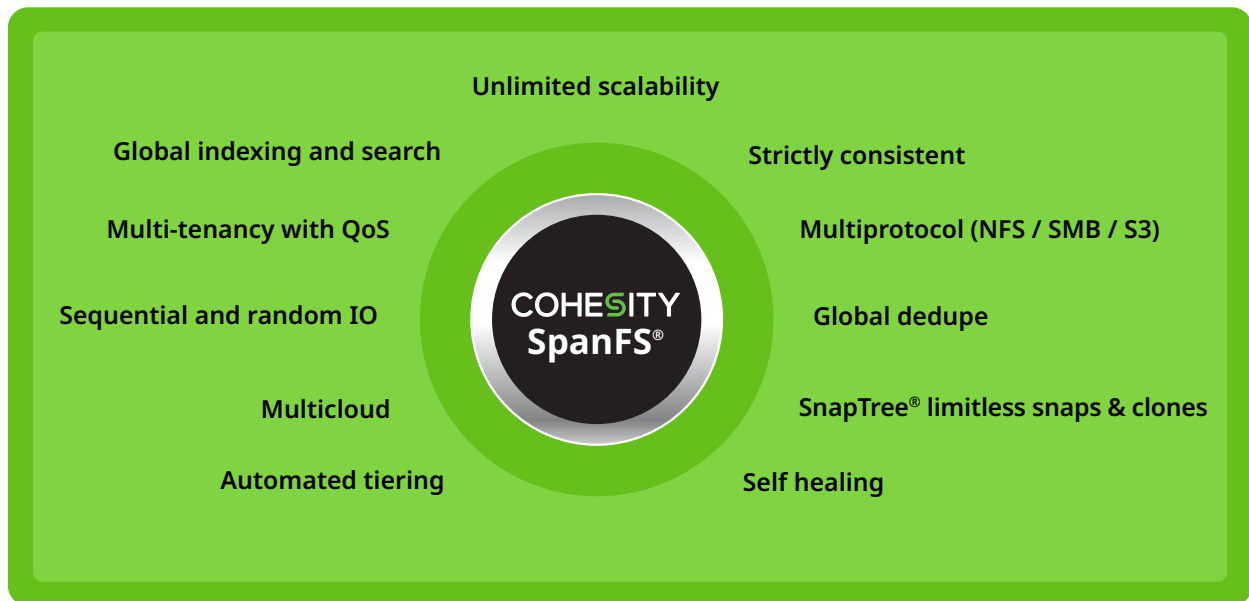
*Cohesity SpanFS consolidates data silos across locations and workloads by uniquely exposing globally distributed NFS, SMB, and S3 protocols on a single platform.*

| SpanFS Attribute | Implication |
|---|---|
| Unlimited Scalability | Start with as little as three nodes and grow limitlessly on-prem or in the cloud. |
| Strictly Consistent | Ensure data resiliency with strict consistency across nodes within a cluster. |
| Multiprotocol | Support traditional NFS and SMB-based applications as well as modern S3-based applications. Read and write to the same data volume with simultaneous multiprotocol access. |
| Global Dedupe | Significantly reduce data footprint by deduplicating across data sources, clouds, and workloads with global variable-length deduplication. |
| Unlimited Snapshots and Clones | Create and store an unlimited number of snapshots and clones with significant space savings and no performance impact. |
| Self-Healing | Auto-balance and auto-distribute workloads across a distributed architecture. |
| Automated Tiering | Automatic data tiering across SSD, HDD, and cloud storage for achieving the right balance between cost optimization and performance. |
| Multicloud | Native integrations with leading public cloud providers for archival, tiering, replication. Protect cloud-native applications. |
| Sequential and Random IO | High I/O performance by auto-detecting the IO profile and placing data on the most appropriate media. |
| Multitenancy with QoS | Native ability to support multiple tenants with QoS support, data isolation, separate encryption keys, and role-based access control. |
| Global Indexing and Search | Rapid global search powered by global indexing of file and object metadata. |

"

At the end of the day, we trust Cohesity to protect our most valuable data assets. Now we not only receive files with video and stats from our own major and minor league operations, but also additional files from MLB. This has caused multifold data growth every year, and since this information is used to make franchise decisions and affects if we win on the field, the ability to easily access data quickly at any moment was imperative.

**David Woolley,** Senior Director, Network Operations, SF Giants

# Delivered the Way You Want

Similar to how public cloud ushered in a new paradigm with infrastructure as a service, and then platform as a service, data management should be available to be consumed as a fully managed software service, as a simple line item on your monthly cloud bill.

Just like with any other cloud service, data management as a service helps IT teams make the transition from CapEx to OpEx, immediately lowering up-front costs and freeing up IT resources to be more productive. And the benefits cascade into the long-term—restructuring your IT team to be more efficient, putting the focus on developer needs while empowering ops, and finally, making your data more productive.

Cohesity delivers the most comprehensive data management solution that includes backup as a service, and is rapidly growing to encompass DR as a service, file and object services, and data governance services, all secure, all available—all on a single platform, available in any form factor—managed by you in your data center, as a turnkey SaaS offering from Cohesity, as well as a fully managed service from Cohesity-powered partners.



*Cohesity offers customers the flexibility to consume data management services in the way they want—managed by their IT team, through a service provider partner, or as a SaaS service.*

# Solving Your Core Data Management Challenges

So far, we have laid out the myriad challenges that enterprises face in trying to rein in their data. And we have also laid out the case for a unified platform approach as a potential solution to this problem. But what are the key services that a platform needs to deliver to provide enterprises with real control of their data? The path to building and executing a successful data strategy for an enterprise starts with protecting its data, and then moving to consolidating it, and then putting it to work.

## Backup and Recovery

Cohesity DataProtect is a backup and recovery service that runs directly on Helios and can be consumed as a cloud service or deployed and managed by the customer. DataProtect provides a modern data protection solution for a wide range of data sources and environments across on-prem, cloud, and edge.

DataProtect takes advantage of all the underlying platform benefits when it comes to scale, performance, and efficiency. In addition, the distributed metadata structure and Cohesity SnapTree combine to provide unique benefits to backup and recovery operations.

DataProtect provides these core benefits:

- A single, hyperscale solution that eliminates multiple legacy point solutions
- A broad range of supported data sources including leading hypervisors, relational databases, NoSQL, Kubernetes, and SaaS applications
- The ability to instantly restore files, objects, databases, and VMs at scale with parallelized recovery and fully hydrated snapshots
- Immutable backups with write once read many capabilities (WORM) for ransomware protection

| | | | | |
|---|---|---|---|---|
| **Virtual** | vmware vSphere | NUTANIX | Microsoft Hyper-V | kubernetes |
| **Databases** | ORACLE | Microsoft SQL Server | mongoDB | cassandra |
| **Applications** | Microsoft 365 | Exchange | SAP S/4 HANA | Epic |
| **Storage** | PURESTORAGE | NetApp | CISCO | ISILON Hewlett Packard Enterprise |
| **Physical** | Linux | Windows Server | AIX | solaris |

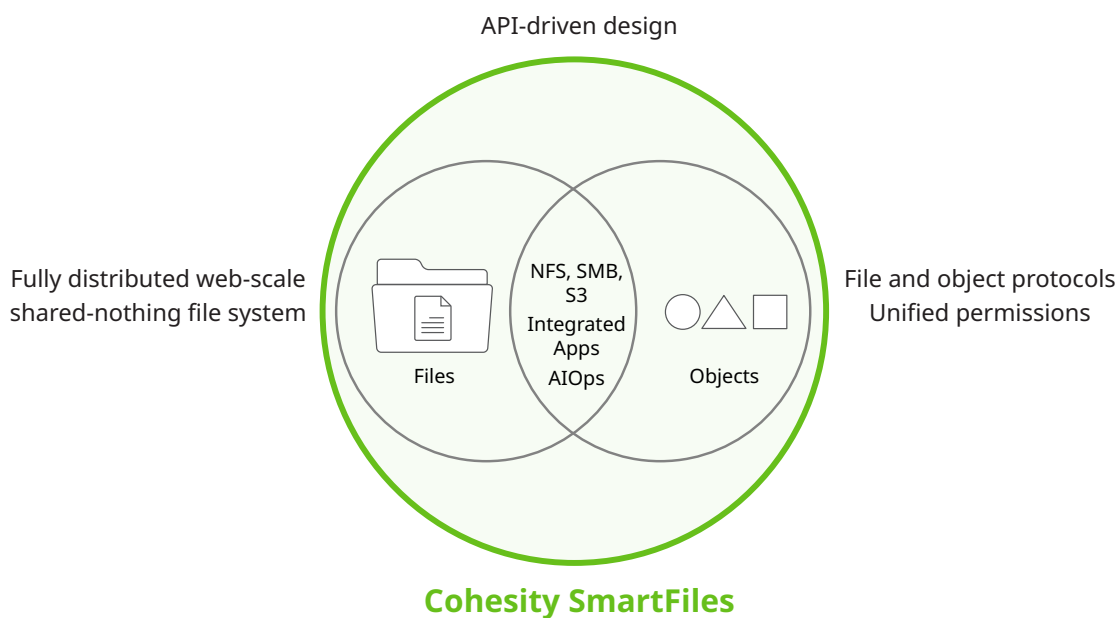| Workload | Type | Supported Data Sources |
|---|---|---|
| Virtual | Hypervisors | VMware<br><br>Microsoft Hyper-V<br><br>Nutanix AHV |
| Physical | Physical | Windows, RHEL, CentOS, AIX, Solaris |
| | NAS | Isilon, NetApp ONTAP, Pure Storage FlashBlade |
| | Storage Snapshot Provider | Pure Storage FlashBlade, Cisco HyperFlex, HPE Nimble |
| Relational and Modern Databases | Relational Databases | Oracle<br><br>Microsoft SQL Server<br><br>SAP HANA |
| | Distributed Databases | NoSQL and Hadoop |
| SaaS | Microsoft 365 | Exchange Online<br><br>SharePoint Online<br><br>OneDrive |
| Cloud-native | Containers | Kubernetes |
| | Public Cloud | Amazon Web Services<br><br>Microsoft Azure<br><br>Google Cloud |
| | VMware on Hyperscale | VMware Cloud on AWS<br><br>Google Cloud VMware Engine |

# File and Object Services

Another major way in which Cohesity addresses mass data fragmentation is by folding in direct support for file and object services. Cohesity SmartFiles not only provides general file and object services, but also includes multiple data reduction technologies to maximize the amount of data stored in each raw TB of storage hardware.

All data on Cohesity is indexed for fast search results. This applies to searching file metadata as well as file contents. These search results can then be used for data-driven decisions. SmartFiles search capabilities can be used to shed light on dark data or to drive compliance and governance requirements.

Additionally, SmartFiles uniquely takes advantage of third-party apps from the Cohesity Marketplace to support platform-native data-centric use cases. This includes antivirus protection, file audit with the ability to detect anomalous file accesses, and content search inside of files. As the foundation for a data-centric approach, such search results can be used to drive eDiscovery, compliance, and other business processes.

There are dozens of file and object workloads and use-cases that SmartFiles is designed to address, including corporate file shares, Splunk cold buckets, videos, archives, and security footage.

API-driven design

Fully distributed web-scale
shared-nothing file system

NFS, SMB, S3
Integrated Apps
AIOps

Files

Objects

File and object protocols
Unified permissions

**Cohesity SmartFiles**

*Cohesity SmartFiles goes beyond traditional NAS capabilities by integrating files and objects from disparate sources on to a single  web-scale platform that delivers industry-leading space and cost efficiencies.*

## Multicloud, Multi-Protocol

SmartFiles provides multiprotocol file and object services with unified permissions via NFS, SMB, and S3. This combination of protocols can be used to facilitate hybrid cloud deployment. It is built on the aforementioned SpanFS filesystem for scale, performance, fast ingest, and support for limitless snapshots and clones. File permissions across NFS and SMB environments are mapped for transparent file access from either protocol. Files and objects can be accessed simultaneously, and all data is always accessible via the S3 protocol, regardless of the original file protocol.

SmartFiles is not bound to proprietary hardware and has the ability to run on a wide selection of approved hardware platforms. SmartFiles also runs in the cloud with all the functionality of on-premises deployments.

## Policy-based Management

Policies are used to automate file and object services. Machine learning is also utilized to assist with management tasks. One example is predictive planning for capacity management. There is also support for data migration from existing filers. The data from the existing filer can be backed up and the completed backup can be presented to users and applications as a storage volume. Operations can be automated through RESTful management APIs. Cold data on costly tier 1 storage can be automatically moved by policies to a cost-optimized tier or to the cloud. Policies can be utilized to ensure consistent management across multiple clusters and locations using a single management pane. Cloud or on-prem. Core or edge. Physical or virtual. Data movement is transparent to applications.
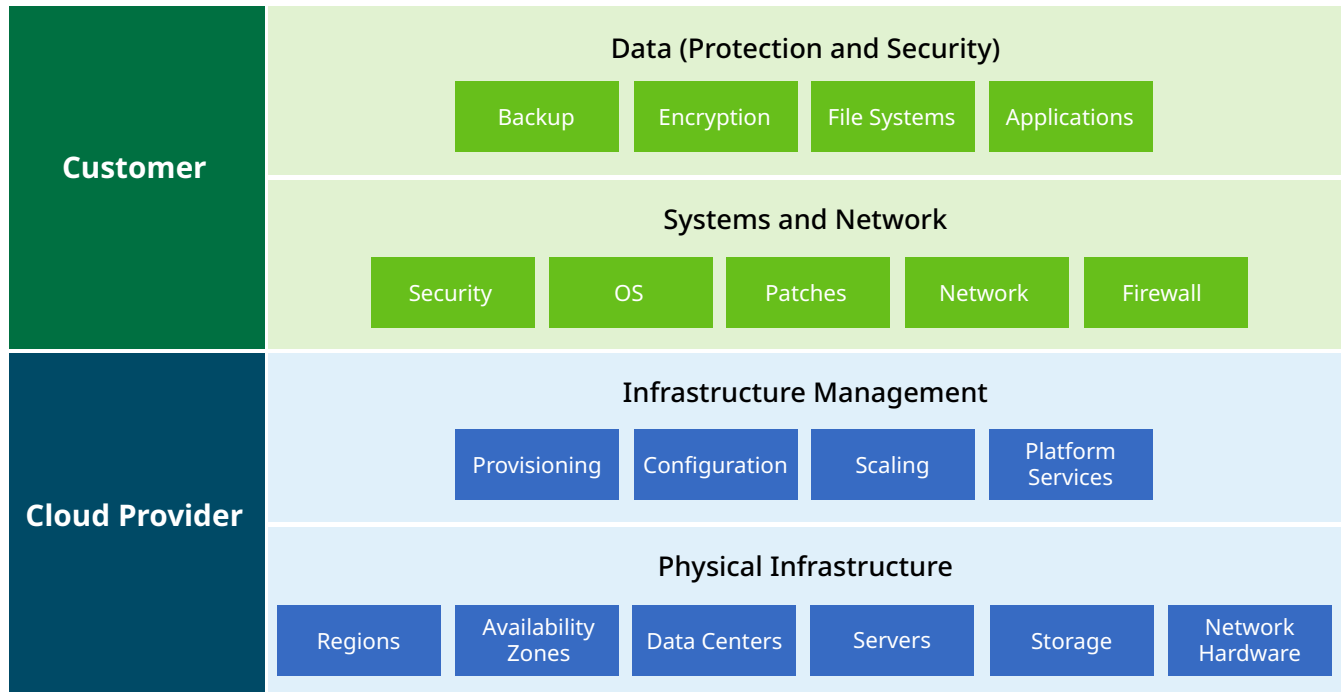
## Integrated Cybersecurity and Anti-ransomware

SmartFiles integrates cybersecurity with a multilayered security approach to prevent, detect, and analyze threats:

- **Prevent**. Immutable file system, software encryption, over the wire encryption, multi factor authentication, DataLock (WORM), and adherence to FIPS 140-1 and 140-2 standards.
- **Detect**. Detection of infected files through the use of apps and anti-compromise tools
- **Analyze**. Observe user and access-pattern analytics for security and compliance

# Cloud: Native Hybrid and Multicloud Integration

Most IT teams are accelerating cloud adoption to cater to an increasing number of use cases and seeing benefits in time to market and agility. However, cloud adoption can contribute to mass data fragmentation by generating data in more locations and in different silos and formats. A common misconception is that cloud providers take care of everything for you, including backing up your data. This is far from the truth, as all major cloud providers employ a shared responsibility model.

| Customer | Data (Protection and Security) | | | |
| --- | --- | --- | --- | --- |
| | Backup | Encryption | File Systems | Applications |
| | Systems and Network | | | |
| | Security / OS / Patches / Network / Firewall | | | |
| Cloud Provider | Infrastructure Management | | | |
| | Provisioning / Configuration / Scaling / Platform Services | | | |
| | Physical Infrastructure | | | |
| | Regions / Availability Zones / Data Centers / Servers / Storage / Network Hardware | | | |

*Public cloud providers are not responsible for the security, integrity, and availability of your data.*

A shared responsibility model implies that they'll take care of physical cloud infrastructure, as in their data centers and servers. But when it comes to your data, it's almost solely your responsibility to secure, manage, and back up your data. If data was accidentally deleted or if a bad actor employs the latest ransomware scheme to get a hold of your cloud account, you may be out of luck if you didn't properly back up your cloud data.

In addition to running in your data center and edge sites, Cohesity's software-defined platform runs on major cloud providers including Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP), in addition to being available as a cloud service. It provides native cloud integration without the costs or complexity of bolt-on gateways.

*Cohesity Helios is a truly ubiquitous platform—running on any cloud, datacenter, and edge.*

## Cloud Backup

While backup appears to be the simplest on-ramp to cloud for many organizations, traditional backup-to-cloud can be complex, expensive, and time-consuming. This is why the cornerstone of Cohesity DataProtect service is simplicity. Available as a cloud service, the solution reduces infrastructure management overheads, while also eliminating the complexities of CapEx-heavy procurement cycles. DataProtect service backs up both on-prem as well as cloud data, with the flexibility of restores in the cloud or back in your data center. The service also allows for zero-cost clones to be leveraged in tandem with other cloud services, such as analytics.

## Cloud Archival

With multiple providers and cloud regions to choose from, it is very easy to replace tape backups and archives with Cohesity to simplify operations and reduce costs. Cohesity CloudArchive—standard within DataProtect—provides native integration to all major cloud providers using their cloud-native APIs as well as industry-standard S3-compliant APIs. This allows on-prem data to be automatically and efficiently archived to cloud storage for long-term retention (LTR) based on policy-based backups and data retention rules. Cohesity supports a wide range of cloud storage options from Amazon S3, Amazon Glacier, to Microsoft BLOBs, to Google Nearline and Coldline enabling you to best match your costs, redundancy, and retrieval requirements.
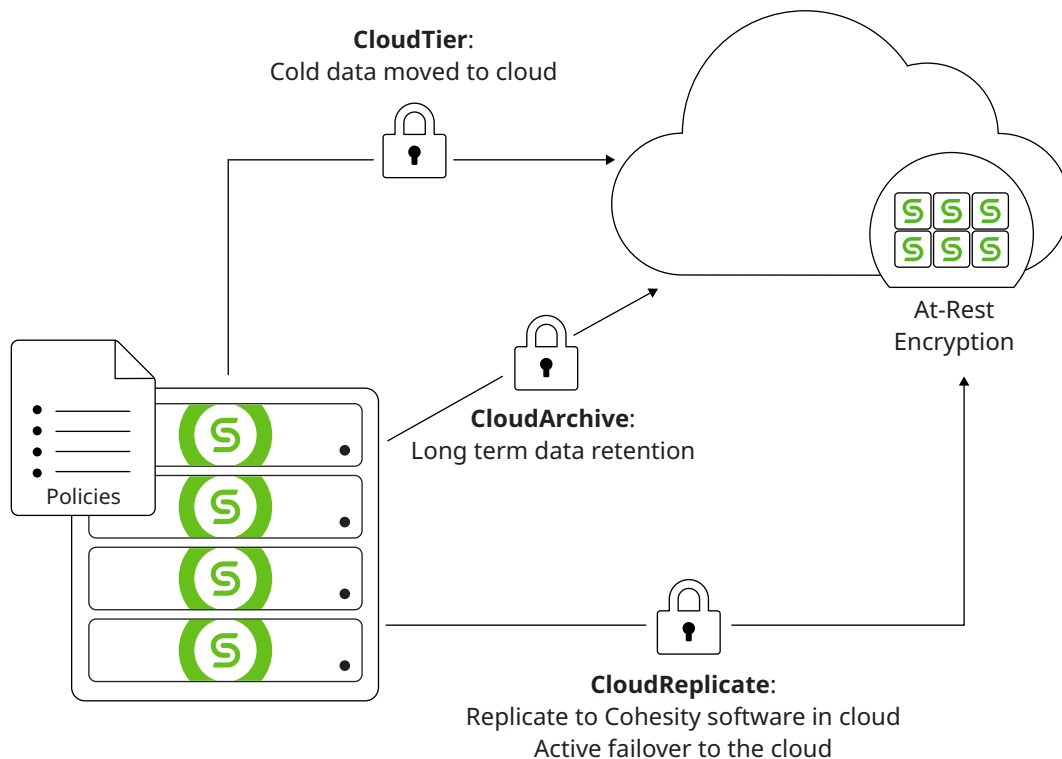
> "
>
> For a film production environment, it was key to quickly and easily retrieve and restore files, and to seamlessly archive to the cloud. In our business, if we cannot restore data for our production teams fast enough, then that may mean we miss production deadlines or our film isn't launched on Netflix when we want. The Cohesity solution was key to our bottom line.
>
> **Tyson Clark,** Technical Director, Air Bud Entertainment

## Intelligent Tiering

Cohesity intelligently tiers cold data into the public cloud as a seamless extension of local data. Low-cost public cloud storage is treated as just another tier, and the platform's policy-based waterfall model moves cold data to the cloud periodically. Data temperature is determined based on last access time and is configurable. Of course, data is encrypted in-flight as well as at-rest.

**CloudTier**:
Cold data moved to cloud

At-Rest Encryption

Policies

**CloudArchive**:
Long term data retention

**CloudReplicate**:
Replicate to Cohesity software in cloud
Active failover to the cloud

*Cohesity automatically tiers data to cloud or high-performant storage based on usage and access patterns.*

## Cloud Dev/Test

In today's world of continuous development, high-quality data in the hands of your developer leads to faster ship, higher-quality applications, quicker iterations, and an overall competitive edge.

Developers and test teams need access to high-quality test data in order to build software. However, provisioning test data fast enough is a critical challenge for many of today's organizations. It takes enterprises multiple days, often weeks, and multiple approval cycles to provide relevant test data to developers and test teams. Consequently, a lack of rapid access to high-quality data prolongs both development and test cycles. This challenge in provisioning test data to developers is principally an architectural one. Mass data fragmentation has led to inefficiencies between IT and software development teams.

Cohesity solves these challenges directly and enables IT operators to make their backup data productive by instantly and securely provisioning quality data to development and test teams through zero-cost clones. By being able to repurpose backed up data as test data, Cohesity reduces data sprawl, costs, and accelerates the development of high-quality applications within Cohesity's runtime environment. Once an organization has already gathered data on the Cohesity platform through backup and recovery or the use of SmartFiles, agile dev/test is a key example of getting more out of your existing data under management.

Combined with automated and on-demand data refresh, Cohesity helps development and test teams with high-quality, relevant data.

Cohesity accelerates application development and enterprise productivity by consolidating previously disparate solutions: test data management, and backup and recovery. By using one platform and repurposing backups for test data, Cohesity helps:

- **Reduce development roadblocks** and allows more accurate testing earlier in the release cycle
- **Simplifies infrastructure and access** to production-quality test data, without impacting production
- **Makes security and compliance simple** with data masking, and delivers greater visibility and control on data

**Data Mobility / Application Migration**

Most organizations are choosing to implement both hybrid and multicloud strategies. This is a good starting point, but your hybrid or multicloud implementations would be limited and siloed if you were not able to easily move and exchange data between all your providers and on-prem.

Cohesity's software-defined platform accelerates the adoption of hybrid and multicloud. Data can be stored and unified on one platform providing a central place to manage all your data. Data mobility is made easy using our fast and efficient Cohesity replication engine between all your cloud providers and on-prem clusters.

# Mitigating Business Risk

On paper, cloud's pay-as-you-go model is a great fit for disaster recovery where resources typically sit idle for months on end waiting for an outage to happen or to be spun up for the occasional six month or annual test. In reality, using the cloud for disaster recovery is much harder in practice due to different network environments, different data formats, and complex manual process steps required to move and spin up workloads in the cloud.

> "
>
> 34% of organizations have experienced downtime as a result of not being able to failover to a cloud-based disaster recovery copy.
>
> **Mass Data Fragmentation in the Cloud: [Global Market Study](#), by Vanson Bourne**

Keeping your business running should not create yet another silo—it should fit seamlessly in your broader IT strategy, and needs to encompass the concepts of availability, continuity, and instant recovery from Day 1.

And Cohesity is the only provider that can deliver. When an outage strikes, you want your solution to recover instantly, regardless of your cloud or application, and you want it to be automated and requiring minimal intervention and additional cost. The Cohesity solution takes the same 3-pronged approach to continuity—a multicloud business continuity platform, with instant recovery capabilities, that is deeply automated and cost-effective.

## Multicloud Failover and Failback

The Helios multicloud data platform runs directly on major cloud providers. This provides a single platform to unify, move, failover, and failback data between on-prem and cloud for disaster recovery. Data can be recovered and failed over from an active Cohesity cloud instance, and restored and spun up from archived data in cloud storage. And Cohesity CloudSpin—also standard with DataProtect—provides a quick and seamless way to bridge the compatibility gap and convert on-prem VM formats from VMware and Hyper-V into cloud-native formats for Amazon EC2 and Azure VMs.
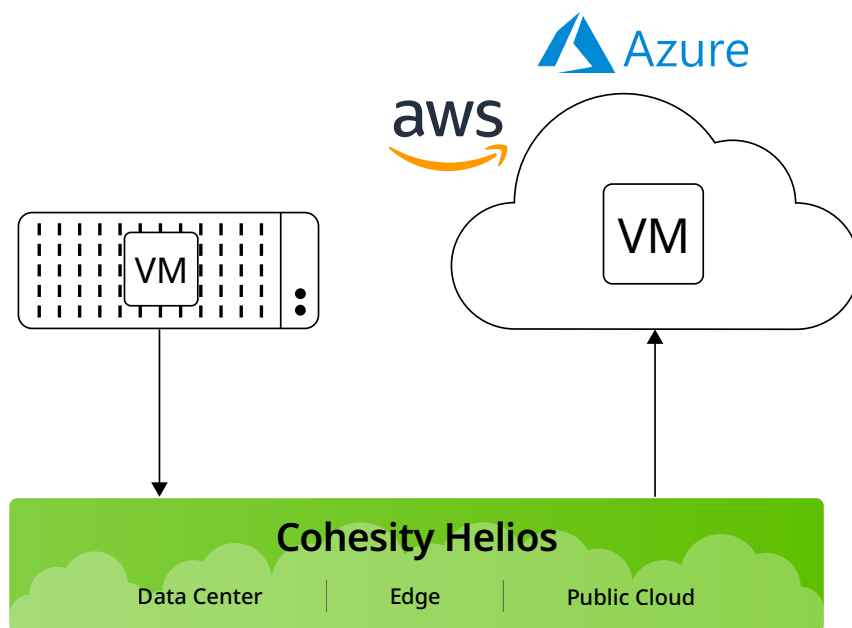


*Cohesity can quickly convert VMs to cloud formats for failover and archival.*

## Instant Mass Restore

Now that the platform can help you recover any app from any cloud, the next step is to do it instantaneously. Cohesity SnapTree's B+ tree structure, combined with the ability to parallel ingest or recover, gives users the ability to recover hundreds of files, objects, and VMs instantly. Unlike legacy and other modern backup solutions, Cohesity uniquely maintains fully hydrated snapshots that can be restored within minutes. With Cohesity, recovery is rapid. Instead of waiting for the restore to complete, Cohesity mounts the data using NFS and SMB for the application to access it, while the restore happens in the background.

**Disaster Recovery Automation**

In case of an outage, your DR strategy needs to be independent of human intervention. Cohesity SiteContinuity ties all your DR plan components together with an easy-to-use automation and orchestration engine that removes manual steps and complexity from the disaster recovery process. It simplifies business continuity and disaster recovery (BCDR) operations with a unified policy framework that protects your applications and data—across tiers, service levels, and environments—both on-prem and in cloud. Its converged journal-based continuous data protection and automated DR saves your IT staff from managing disparate infrastructure silos and continually spinning up new virtual machines. The unified policy framework of SiteContinuity gives you the flexibility to restore your critical applications and data to any point in time and location. You can choose to recover data from years ago or just seconds before the unexpected happened.
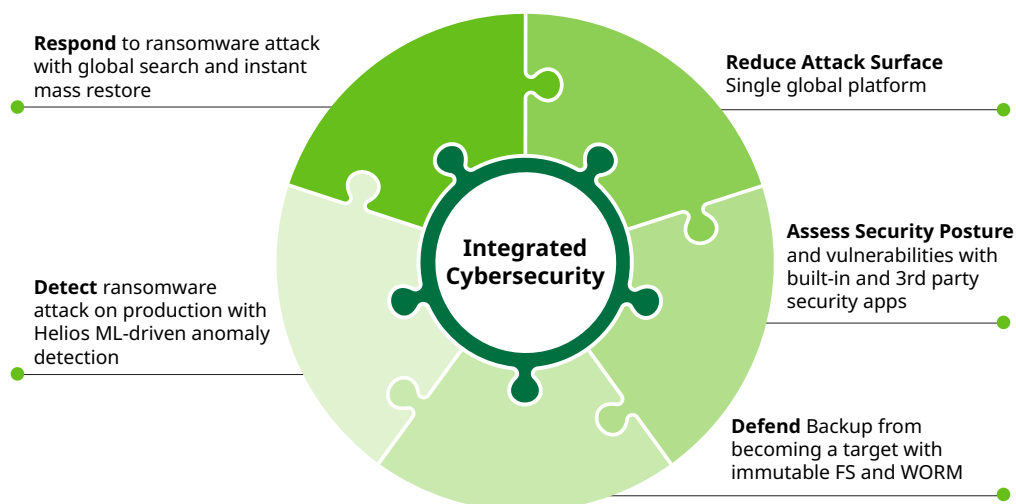
# Defeating Ransomware

The average cost to remediate a ransomware attack has risen to over $750,000 in 2020. With an increasingly distributed workforce and larger attack footprint due to growing data volume, ransomware protection needs to be front and center in your cloud and data strategy to ensure your business stays up and running.

Cohesity's integrated security offers a holistic approach to:

1. Protecting your backup data against ransomware attacks

2. Leveraging data to discover cyber vulnerabilities and detect attacks in progress

3. Reducing downtime with instant restore at scale

This unique set of capabilities is delivered through a combination of platform attributes combined with the ability to run Marketplace applications and is a great example of the value of taking a holistic *platform* approach to data management.

**Respond** to ransomware attack with global search and instant mass restore

**Reduce Attack Surface**
Single global platform

**Integrated Cybersecurity**

**Assess Security Posture** and vulnerabilities with built-in and 3rd party security apps

**Detect** ransomware attack on production with Helios ML-driven anomaly detection

**Defend** Backup from becoming a target with immutable FS and WORM

*Cohesity is a fortress for your enterprise data—built from the ground up with a security-first approach.*

**Reduce the Attack Surface**: Cybersecurity is one of the key tenets of Cohesity's architecture, starting with a single software-defined data management platform that consolidates multiple point solutions and reduces data footprint. This helps enterprises to reduce their attack surface.

**Cyber Exposure and Vulnerability Management**: Cohesity uniquely allows backup administrators to scan backup snapshots on-demand or on a scheduled basis to uncover cyber exposures and vulnerabilities in their production environment without impacting performance or overhead. Today, this capability is delivered through the Cohesity Marketplace application built in conjunction with Tenable, a cyber exposure prevention expert.

> "
>
> A modern backup and recovery solution that works and provides ransomware protection from edge to cloud.
>
> **Jonathan Mancini,** Supervisor, Datacenter Operations, City of Hamilton
>
> Hamilton

**Defend Backup Against Ransomware Attack**: Backup data is a prime target for attackers. But Cohesity's immutable file system SpanFS and WORM stops malware from modifying or deleting backups from the get-go. The industry's first WORM for backup creates time-bound protection to the snapshot that even your administrator or security officer role can't overrule.

**Machine Learning-Powered Detection and Actionable Recommendations**: In the case that attackers hold your primary environment as a hostage, Cohesity is not just able to detect a potential attack, but also point to the last known clean snapshot to initiate a response.

**Rapid Response**: With unlimited fully hydrated snapshots, combined with a distributed architecture, global actionable search, and live mount, Cohesity uniquely allows IT to instantly recover hundreds of files, objects, and VMs rapidly. You can bounce back from even the most severe attacks, instantly and effortlessly.

# Managing It All Centrally
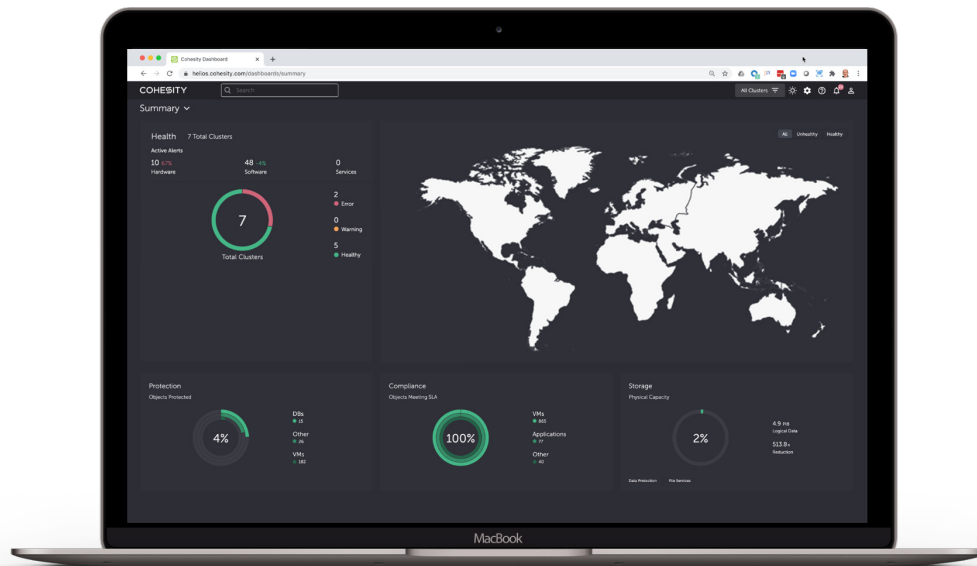


*Cohesity enables IT professionals to get deep visibility and control across their entire data estate.*

Whether on-prem or in the cloud, deployed or consumed as a service, Helios lets you centrally view, operate, and manage your entire data estate through one simple UI. This is a single dashboard that replaces multiple consoles and the need to manage, monitor, and operate multiple clusters. It allows users to manage policies and clusters globally, and proactively detect and avoid downtime with built-in machine learning.

Cohesity provides global access and control of data across locations, and delivers predictive and actionable insights to help meet SLAs and improve business continuity. In this way, Cohesity understands that tackling mass data fragmentation doesn't stop with data consolidation. It's critical to streamline data control and management, as well.
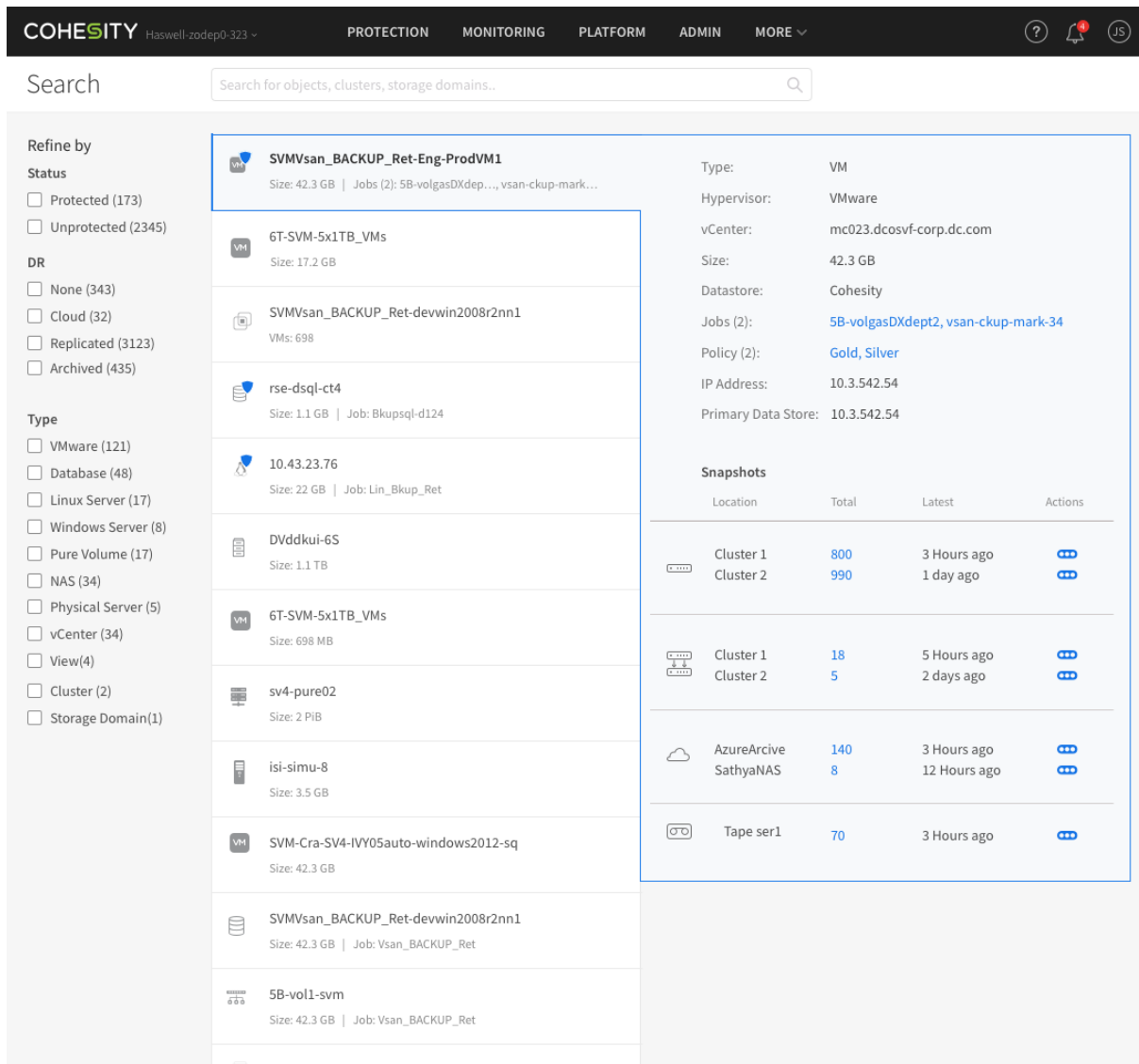
*Cohesity indexes all your data and makes it searchable, no matter the source.*

The collection of metadata from distributed Cohesity instances also makes Global Actionable Search possible, allowing organizations to search and recover any data across any cluster, irrespective of its physical location—including native cloud snapshots. Search returns results instantaneously but also offers an actionable recommendation that puts users in the desired workflow, giving them the keys to take action without leaving the page they're on.

# Putting Your Data to Work

Instead of creating multiple copies and shipping data around, Cohesity makes backup and other unstructured data productive by allowing you to run Cohesity-developed, as well as third-party applications directly on the Cohesity cluster. The philosophy of bringing compute to the data empowers IT operators to derive insights from previously dormant data to meet security, compliance, and analytics requirements. Cohesity's unique architecture allows applications to access and use the data within each Cohesity cluster, providing visibility and insights into the data that previously was dark.

Users can download Cohesity and third-party applications directly from Cohesity Marketplace or build their own apps using the Cohesity App SDK to address their unique business requirements.

The overarching benefit of Marketplace and apps is that the value of your data—and your ROI—grows with time. New applications continuously allow users to unlock more value from their data, in the form of insights as an example, or more effectively secure data in the face of evolving threats and compliance regulations.

*The Cohesity Marketplace features data-centric apps from leading SaaS vendors,*
*purpose-built to run on Cohesity.*

## Analytics

Cohesity Helios is feature-rich when it comes to reporting and analytics. The platform is natively capable of generating over 20 different reports on a variety of data center health vectors such as data capacity, SLA performance, backup status, and much more. Strong business outcomes are powered by strong business intelligence, and there are several apps available on the Cohesity Marketplace that generate analytics and reports in a variety of consumable formats, on demand, or on schedule.

## Data Governance

Cohesity provides a data classification and governance solution that makes the process of defining classification policies and discovering data both simple and automatic. Cohesity's solution incorporates the ability to choose a pre-built compliance template, for GDPR, as an example, or search and define custom classification policies directly from the same Cohesity user interface.

Cohesity's solution works across multiple data sources across multiple clusters, and with rapid search, complex and large global environments can automatically meet evolving compliance requirements. After data discovery, users have options to control data access, movement, alerting, tracking, deletion, and redaction.

Apps on the Cohesity Marketplace also help in the masking of test data so developers can build apps on high-fidelity data without having to worry about PII exposure.
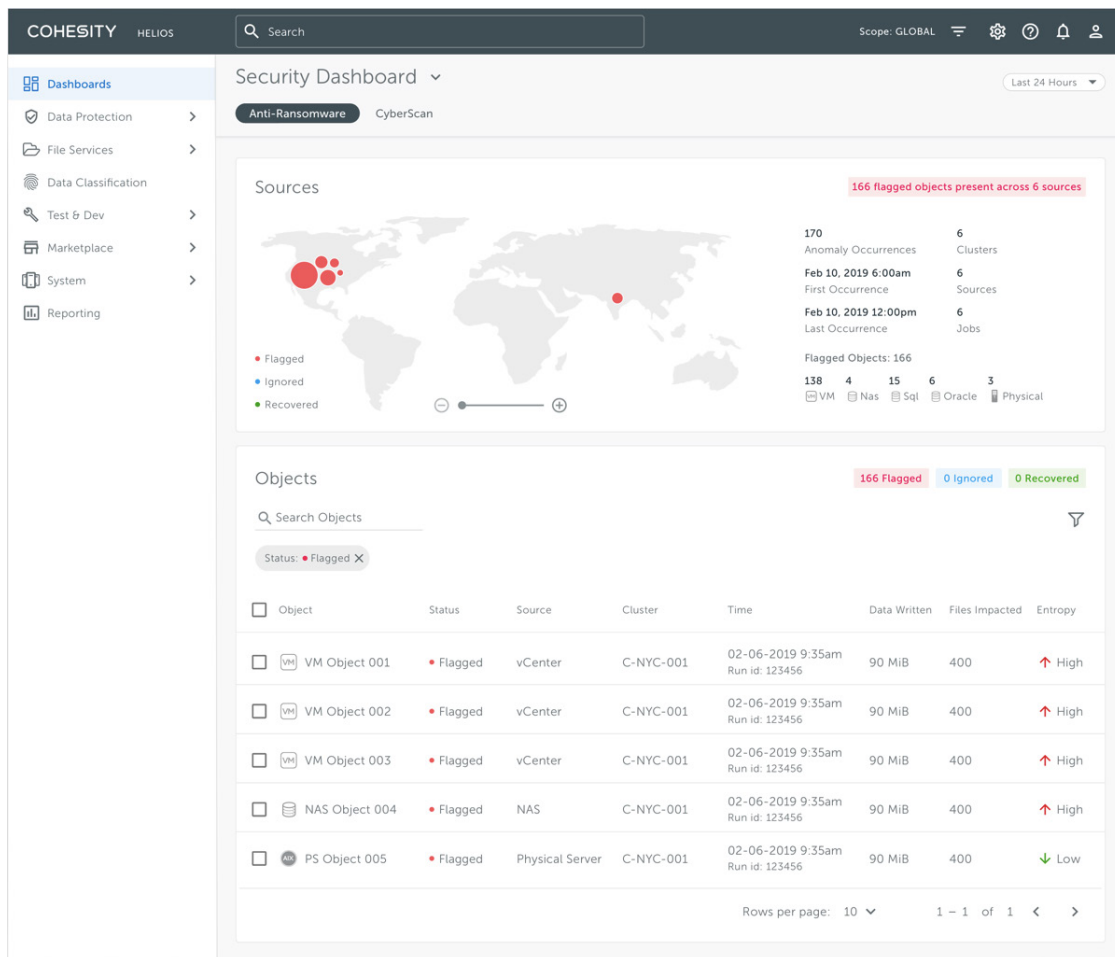
*Cohesity provides security and compliance audits and recommendations for all your data—
on schedule or on demand.*

## Alerts and Automation

Whether the Cohesity user is a data center admin, business continuity engineer, sec-ops, or an SRE, service automation and automated alerting are pivotal to the speed of the entire IT department. With that in mind, Cohesity has a full suite of integrations with toolchains for any user that wants to automate core data management operations for better SLA performance and a more agile IT team. Cohesity's list of integrations include Powershell, Ansible, VMware vRealize Automation, Hashicorp Terraform, PagerDuty, VMware Cloud Director, ServiceNow, and more.

PowerShell



Python



Ansible



vRealize Automation



ServiceNow



vCloud Director



PagerDuty



Terraform

*Cohesity has integrations with almost all leading toolchains and platforms, and can be invoked or leveraged through its REST API.*

Based on an API-driven architecture, Cohesity empowers every member of the IT team to spend less time with infrastructure management and instead help accelerate app development.

## Vulnerability Scanning and Antivirus

The Cohesity Marketplace offers several solutions—in the form of apps—from leading data security vendors. These apps can provide comprehensive vulnerability scanning and threat detection for your entire data estate. These apps have UI-level integration within Cohesity, and allow several personas within an IT team to access and control data security policies right from within the Cohesity UI.

The Marketplace also offers antivirus apps that can protect files resident on Cohesity. This helps customers avoid the hassles of porting over data, deploying an ICAP server, and adding vulnerabilities in the process, and instead, shift to running antivirus on the same platform that houses all their data.

## Auditing and Compliance

With growing regulatory mandates and focus on giving more control to people over their data, the need to make data easier to find and determine its level of sensitivity or risk is increasingly important. Fragmentation of data across multiple silos makes it difficult to meet complex and ever-evolving compliance requirements. With Cohesity, customers can implement a data classification and governance solution that makes the process of defining classification policies and discovering data both simple and automatic.

# Turnkey Data Management as a Service

Cohesity is unique in the industry when it comes to offering a comprehensive data management portfolio as a service. This is out of the realm of possibility with legacy technologies and is made possible only with Cohesity's multicloud, holistic approach to data management.

Cohesity Data Management as a Service (DMaaS) is a SaaS offering that customers of all sizes can choose to consume straight from AWS. Cohesity brings SaaS benefits to turnkey data management—including the switch to 0-CapEx, eliminating overprovisioning, and redeploying IT staff to more strategic projects. Starting with backup and recovery, the offerings will quickly expand to cover disaster recovery, archiving, file and object services, dev/test provisioning, data governance, security, and much more. Just like Cohesity consolidated siloed legacy infrastructure, DMaaS consolidates myriad cloud services—all onto a single solution. But that's not all—customers will now gain the benefits of using Cohesity DMaaS services alongside industry-leading cloud services on AWS all on the same platform, enabling them to quickly and easily put their data to work. Use-cases include feeding your data into ETL engines (with AWS Glue), warehouses (such as AWS Redshift), machine-learning frameworks (AWS SageMaker), and privacy services (AWS Macie).

 Data Management as a Service from Cohesity bridges the chasm between infrastructure/platform, and the analytics/machine learning environment that you are running in the public cloud. To be able to put your data to good use, it's not enough that it's available and resilient, but also needs to be controllable, visible, and secure. And Cohesity makes all that available as a cloud service on AWS, and combined with your associated platform services such as data warehouses and data lakes on one side, and your analytics platform and developers on the other, you have yourself a fully optimized, fully operational data engine.

Cohesity DMaaS allows you to instantly access an enterprise-class data platform that is used by the world's largest firms across sectors—financial, retail, pharmaceutical, consumer electronics, and more. As a result, your IT team can focus more on development, and more meaningful and productive tasks.

> "
> After testing DataProtect delivered as a Service we can see how the offering can expand and simplify data management at regional/edge locations for us and give IT staff the ability to focus on other tasks. From a strategic standpoint, there is tremendous potential in a comprehensive DMaaS solution that addresses a range of use cases from disaster recovery, to security, to compliance, and provides the ability to use cloud services to geteven more value from data.
>
> **Oded Haner,** SVP of Technology. Franklin Templeton
>
> FRANKLIN TEMPLETON
> INVESTMENTS

# Building a Secure Foundation for Continuous Innovation with Data

Today, IT organizations face unprecedented demands to not simply support business operations efficiently, but to also act as a source of innovation and competitive advantage. We believe that mass data fragmentation is the most significant roadblock to digital transformation and that more effective management of data is key to enabling IT to deliver against those expectations.

Cohesity has built a unique solution based on the same architectural principles employed by cloud hyperscalers managing consumer data, but optimized for the enterprise world. The unique capabilities of Cohesity Helios allow all data management functions and workloads—including backup and recovery, target storage, DR, archiving, file and object services, cloud tiering, dev/test provisioning, and data analytics—to be run and managed in a software-defined environment across any cloud, rather than in isolated silos. All of these functions can be managed and operated within its beautiful UI or with its rich APIs, leveraging deep automation and a unified policy engine. It makes the IT administrator's job that much more enjoyable and easier. Fundamentally, Cohesity helps curtail the damaging impacts of mass data fragmentation on your business and begins to get your data to work for you.

Available in customer-managed deployments, partner-managed offerings, or as a SaaS solution, Cohesity helps you take control of your data, build data resilience and compliance, and helps your IT team become more productive to your business outcomes. Cohesity is an essential piece of the data pipelines of the world's most successful companies.

**If this paper has piqued your interest, we encourage you to learn more at www.cohesity.com or speak to one of our representatives at www.cohesity.com/contact**