Technology Assessment

Cohesity SmartFiles: Efficiently Tackles Unstructured Data

Amita Potnis

IDC OPINION

The need for a simple, flexible, and intelligent file-based storage solution is on the rise. Traditional file systems often suffer from lack of ease of management, scalability, and in-built analytics to monitor system health and resource usage. Modern file systems address these issues for customer success in ways traditional providers cannot. IDC believes that the market is seeing a growing need of intelligent file system that is cost effective and offers several enterprise features and capabilities such as data visibility and control, security, and predictive and policy-based data tiering. As enterprises undertake digital transformation, they will also be looking to modernize their IT infrastructures. As the variety of workloads consuming data and the endpoints where data is generated continue to increase, there is a need for end-to-end integrated solutions that can address some key requirements. In recent times, several vendors such as Cohesity have brought forward cost-effective, intelligent file systems to support various workloads. Such solutions are expected to see greater adoption and market share over time. The most common concerns for IT administrators contending with data growth while pursuing business strategy include:

- **Aligning IT projects.** Rightsizing infrastructure for applications with large amounts of unstructured data is a strategy to future proof IT and ensure cost effectiveness. Integrated solutions that span the needs of a core-edge-cloud strategy are especially beneficial to organizations as they offer ease of adoption and management.
- **Cost and efficiency.** Organizations can keep costs in check by adopting new storage technologies such as software-defined storage (SDS), cloud strategies, and other solutions. The key benefits of SDS — ease of scalability and management, better economics, and hardware flexibility — are very appealing to enterprises that are refreshing their existing IT infrastructure. While technology advancements in compute, storage, and networking technologies are enabling SDS to handle a broader set of workloads, there are still workloads and customer types for which it is not the best choice. Solutions that offer data visibility across storage silos can allow infrastructure resource optimization and thus realize cost benefits as well.
- **Mitigating risk.** To keep business running and to prevent financial losses, it is imperative that data be secured by implementing appropriate data protection and recovery technologies and integrated cybersecurity, staying compliant to data governance regulations, and offering application-based quality of service (QoS). Any organization that considers data as its intellectual property (IP) will be especially concerned with the risk of data loss as it directly impacts revenue, trust, and safety.
- **Analytics and governance.** Data collected and stored over an extended period is a gold mine for driving insights. Depending upon the type of data, it may need to be stored for specific periods of time as per regulatory requirements. Data durability, visibility, and control across deployment locations, therefore, is an important consideration.

June 2020, IDC #US46538520

IN THIS STUDY

In this study, IDC will assess Cohesity's scale-out storage solution and how customers can benefit from it.

SITUATION OVERVIEW

Cohesity, which originally targeted its market entry points as data protection and secondary storage, has evolved to take a much broader view of helping organizations extract value from their data whether that data is on premises or in the cloud. This often involves using backup data sets that would otherwise lie fallow for additional use cases such as analytics, dev/test, and staging – all on the same platform.

Over the past year or so, Cohesity has transformed itself from an appliance vendor to a software-only vendor. Thus its platform is now capable of being software defined, giving it broad deployment possibilities: on premises, in the cloud, or in conjunction with third parties. We believe that the Cohesity architecture is a key strength of the product, allowing for wide extensibility with third-party products, thereby creating a symbiotic ecosystem around the product. This ecosystem is a critical success factor for any solution that purports to be a platform.

Adding to its success, on April 9, 2020, Cohesity announced that it has secured a $250 million investment in a Series E round of funding. This round brings the total investment in the company to more than $650 million and puts a $2.5 billion valuation on the company. The round included investments from more than 10 different investors, including both new and existing venture investors and IT industry heavyweights HPE and Cisco. Cohesity expects to use this funding to fuel its go-to-market growth, especially outside North America, and expand product development.

The section that follows looks at the company's DataPlatform and latest developments around SmartFiles.

Cohesity DataPlatform

Cohesity's solid footprint in the data protection space has always been underpinned by its core DataPlatform product. Cohesity's DataPlatform is a software-defined storage scale-out NAS offering that consolidates all secondary data including backups, archives, and test and development copies into a single platform. At the core of Cohesity's DataPlatform is a distributed file system called SpanFS with support for NFS, SMB, and S3 data access. The solution's native cloud tiering capability can tier data to Microsoft Azure, AWS, Google Cloud, and other S3 compatible options.

The solution supports deduplication across the entire cluster to reduce an organization's data footprint in secondary storage. Snapshots, erasure coding, compression, and variable length sliding window are other data efficiency features supported by DataPlatform. The solution also supports management through a simple policy-driven UI, multitenancy, role-based access control, QoS, and real-time and custom reporting. The solution allows enterprises to increase visibility into existing data sets and drive valuable insights using global indexing and a Google-like search capability.

Cohesity has two additional flavors of this solution for the edge and the cloud called DataPlatform Edge and DataPlatform Cloud Edition. These solutions extend the key functionality and policy-based data mobility from core to edge and cloud.

More recently, the DataPlatform announced several additional capabilities as detailed in the section that follows.

SmartFiles: Empowering DataPlatform

Announced in late 2019, SmartFiles was introduced as a part of the DataPlatform to enable enterprises to find greater value from their data sets. SmartFiles adds several capabilities such as integrated cybersecurity, native content search, file auditing, and automated and policy-driven data tiering into DataPlatform.

Cohesity's SmartFiles is a part of its DataPlatform, offering integrated applications to bring the following capabilities to enterprises:

- **Integrated cybersecurity.** SmartFiles utilizes ML-based capabilities to identify anomalous system behavior in order to detect potential cyberattacks and protect data from cyberattacks. Antivirus support is an integrated feature in SmartFiles, and enterprises can take advantage of quick installation and avoid additional ICAP infrastructure costs. Encryption, immutable file system, and WORM capabilities protect data loss and thus business reputation.
- **Data tiering.** Enterprises can transparently move cold data to Cohesity using SmartFiles policies with no additional infrastructure or software cost or management overhead. This results in cost savings by freeing up valuable performant resources on primary NAS but keeping data accessible through symbolic links.
- **File auditing.** Storage or compliance administrators can quickly understand whether there are any anomalous file access patterns and file usage on their NAS shares. File audit logs capture the access and usage patterns, and these logs can be searched to understand detailed data creation, modification, access, or deletion patterns. This capability flags abnormal user activity and entities.
- **Search.** SmartFiles allows users to search data across storage silos, looking within VMs, files, and backups, and across datacenters, remote sites, and multiple clouds. This increased visibility of data, despite deployment location and tier, allows enterprises to gain actionable insights from their data sets in an expedient manner.
- **Data efficiency.** SmartFiles supports a variety of data efficiency features such as data deduplication, compression, small file optimization, and cloud volume deduplication (across storage silos).

FUTURE OUTLOOK

Industry Focus and Initial Success

Currently, Cohesity's primary revenue stream is from the backup business, but the company confirms that SmartFiles is seeing increased traction. Today Cohesity has several customers with petabyte-scale data in a variety of industries across the world including finance, healthcare, government, professional services, education, and research.

IDC believes that Cohesity's DataPlatform and SmartFiles can provide benefits in the following ways:

- **Infrastructure optimization.** Policy-based and intelligent data tiering allows for cost-optimized utilization of infrastructure.

- **Simplified management.** Increased visibility into data, access patterns across storage silos, and public clouds allows enterprises to utilize their time more efficiently for other high-value tasks like extracting insights from data.
- **Flexibility.** Being a software-defined offering, enterprises can deploy the solution on a choice of hardware or cloud service providers including leading public cloud providers like AWS.
- **Support for newer workloads.** Cohesity's DataPlatform and SmartFiles have been deployed to support public sector Internet of Things (IoT) initiatives such as "Smart Cities" projects.

ESSENTIAL GUIDANCE

Market Expectations

Organizations of all sizes continually face the challenges of managing data growth. As organizations embrace Big Data, social, mobile, and virtualization, storage needs will continue to grow. The rate of data growth may vary depending upon the organization, but it is evident that scale-out storage solutions are suited well to manage data growth in these environments.

When looking for the right scale-out solution, organizations should ensure that the solution scales not just from a hardware perspective but also from a manageability, file volume, and file size standpoint that align with the price/performance requirements of the particular use case. Data optimization techniques (deduplication, compression, thin provisioning, etc.), as well as data management (metadata, indexing, analytics, etc.), are also essential considerations. Data security, efficient search, and auditing are important capabilities to consider.

Organizations should work with suppliers to create a solution that best fits the requirements. Organizations should also evaluate the strategic role scale-out storage will play through the various delivery models that may best suit their IT and business requirements.

Considering Cohesity

Over the past year or so, Cohesity has transformed itself from an appliance vendor to a software-only vendor. With its software-defined approach, the company has broad deployment possibilities: on premises, in the cloud, or in conjunction with third parties. This is especially important in the world of unstructured data where data growth is exponential and the requirements to store and retain data long term in a compliant and secure fashion drive choice of appropriate infrastructure.

IDC believes that the Cohesity architecture is a key strength of the product, allowing for wide extensibility with third-party solutions, thereby creating a symbiotic ecosystem around the product. This ecosystem is a critical success factor for any solution that purports to be a platform.

IDC encourages potential customers to consider Cohesity DataPlatform and SmartFiles for their secondary storage NAS requirements to realize ease of use, cost efficiency, and several other benefits.

LEARN MORE

Related Research

- *Red Hat Summit 2020 Kicks Off the Trade Show Season — Virtually, Of Course* (IDC #US46265620, May 2020)

- *The "AI Plane": An Interoperable Framework for Artificial Intelligence Infrastructure Stacks* (IDC #US46283420, May 2020)
- *Classifying Artificial Intelligence Workloads* (IDC #US46070820, April 2020)
- *What Are the Key Considerations When Choosing a Solution for Long-Term Retention?* (IDC #US46167520, March 2020)
- *Long-Term Unstructured Data Retention Trends* (IDC #US46163920, March 2020)
- *Video Surveillance Needs Massively Scalable Storage, AI, and Analytics Capabilities* (IDC #US46163420, March 2020)
- *NVIDIA Acquires SwiftStack to Enable AI at Scale* (IDC #IcUS46113820, March 2020)
- *IDC's Worldwide Software-Defined Infrastructure Taxonomy, 2020* (IDC #US45427920, February 2020)
- *IDC's Worldwide Enterprise Storage Systems Taxonomy, 2020* (IDC #US45949020, February 2020)
- *Worldwide File- and Object-Based Storage Forecast, 2019-2023* (IDC #US45706819, December 2019)
- *IDC MarketScape: Worldwide Object-Based Storage 2019 Vendor Assessment* (IDC #US45354219, December 2019)
- *Data Analytics for Unstructured Data: A Growing Workload on Object Storage* (IDC #US45693019, December 2019)
- *IDC MarketScape: Worldwide Object-Based Storage 2018 Vendor Assessment* (IDC #US42665518, June 2018)

Synopsis

This IDC study assesses Cohesity's scale-out storage solution and how customers can benefit from it.

"Simplicity of deployment and management, data and resource analytics, reporting and logging, policy-based, and predictive data tiering as well as integrated cybersecurity are important features that organizations seek when considering a solution for unstructured data," said Amita Potnis, research director, Content Infrastructure. "Cohesity's DataPlatform has brought to market SmartFiles that enables organizations to take care of several of these capabilities and enhance their infrastructure."

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com