White Paper

NetApp Data Fabric Fundamentals: Building a Data Fabric Today

Joe CaraDonna, Arthur Lent, NetApp
October 2015 | WP-7218

Abstract

Data Fabric is NetApp's vision for the future of data management. A data fabric seamlessly connects different data management environments across disparate clouds into a cohesive, integrated whole.

A Data Fabric enabled by NetApp® helps organizations maintain control and choice in how they manage, secure, protect, and access their data across the hybrid cloud, no matter where it is.

Although the data fabric is constantly evolving, organizations can start taking advantage of it today using NetApp technologies that enable data management and seamless data movement across the hybrid cloud.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

IT professionals today are seeking ways to accelerate innovation by taking advantage of technology trends with cloud, object storage, open source, converged infrastructures, virtualization, flash, containers, and software-defined storage, to name a few.

Hybrid cloud deployment models—a combination of private and public cloud resources—are becoming the new normal for enterprise IT organizations. They offer a wide choice of application environments with a seemingly limitless pool for compute, network, and storage resources. Organizations want the freedom to move applications and workloads to the optimal environment as their needs change and as new options gain traction in the market.

Applications are largely stateless: they can be rapidly spun up and down in various environments. Data, however, is stateful and often served by independent datastores or databases. Data has mass: it takes time to move it where it is needed and resources to store it. Data has value: it is a protected asset for any modern company. Data has temperature: it has different degrees of accessibility at any point in time. All of these properties of data are dynamic, which makes comprehensive data management necessary.

Managing data in hybrid cloud architectures that have evolved into incompatible data silos brings additional challenges, including:

- **Inability to move data.** After an organization's data is in a particular cloud, it is difficult or impossible to move it to a different one.
- **Difficulty managing data consistently.** Each environment has a different set of tools, APIs, and management software that make it difficult to apply consistent policies to data. IT staff must learn how to use all of these tools and applications effectively.
- **Limited choice.** New technologies and services that do not integrate with existing environments are difficult to adopt. As a result, IT is limited in its technology choices, affecting its ability to exploit the capabilities of existing and new environments.
- **Lack of control.** Data is a critical asset for successful organizations. IT must be the stewards of that data no matter where it is. Storing that data in a cloud where there is little visibility into how it is protected and governed can put businesses at risk.

To overcome these challenges, IT needs a secure, seamless way to manage applications and data across clouds, regardless of their underlying data platforms. When clouds are connected, IT is able to draw from the resources of each, move data and applications to new cloud services, and put every workload on the most appropriate platform.

1.1 About This White Paper

The purpose of this paper is to describe how a Data Fabric enabled by NetApp can solve the challenges of data management in today's distributed, ever-changing IT landscape. It is organized into five sections:

- Section 1: Introduction: why a data fabric is necessary
- Section 2: Data Fabric in Action: how enterprises and service providers are using Data Fabric today
- Section 3: Laying the Foundation for a Data Fabric: what it means to have a data-centric view of IT infrastructure
- Section 4: Building Data Fabric Capabilities Up and Out
- Section 5: Conclusion

Readers who seek a broad understanding can read sections 1, 2, and 5 and skip over the technical information in sections 3 and 4.

1.2 Defining the Data Fabric

Data Fabric is NetApp's vision for the future of data management. It enables customers to respond and innovate more quickly because data is free to be accessed where it is needed most. Customers can realize the full potential of their hybrid cloud and make the best decisions for their business. The NetApp Data Fabric vision is continually evolving. Over time, the fabric will expand to cover more environments. With each new weave, the data fabric becomes more textured, more expansive, and more capable.

To fulfill this vision, Data Fabric is also NetApp's technology architecture for hybrid cloud. NetApp products, services, and partnerships help customers seamlessly manage their data across their diverse IT resources, spanning flash, disk, and cloud. IT has the flexibility to choose the right set of resources and the freedom to change them whenever needed.

A true data fabric delivers on five major design principles:

- **Control**. Securely retain control and governance of data regardless of its location: on premises, near the cloud, or in the cloud.
- **Choice**. Choose cloud, application ecosystem, delivery methods, storage platforms, and deployment models, with freedom to change.
- **Integration**. Enable the components in every layer of the architectural stack to operate as one while extracting the full value of each component.
- **Access**. Easily get data to where applications need it, when they need it, in a way they can use it.
- **Consistency**. Manage data across multiple environments using standard tools and processes regardless of where it resides.

When a fabric delivers on these principles, it enables customers to increase efficiency, improve IT responsiveness, and ultimately accelerate innovation.

1.3 Who Needs a Data Fabric

Today, NetApp delivers unified data management across clouds. With a Data Fabric enabled by NetApp, organizations can increase efficiency, improve IT responsiveness, and ultimately accelerate innovation.

Enterprise CEO

With a Data Fabric enabled by NetApp, enterprise CEOs can foster an environment that stimulates innovation. They can improve use of business resources. They can move at the speed of smaller companies without sacrificing the ability to meet business and compliance requirements of their industry. They can have confidence that the organization's data is secure.

Enterprise CIO

When it comes to meeting the needs of the business, CIOs can be challenged by conflicting priorities. They need to maintain existing environments while adopting promising technologies. With a Data Fabric enabled by NetApp, CIOs gain the freedom to make the best decisions for the business, make sure of access to data wherever it is needed, and accelerate innovation with fewer resources.

IT Architects

Heads of IT infrastructure need to satisfy diverse service-level objectives (SLOs). Workloads require different availability, performance, cost, security, and access. With a Data Fabric enabled by NetApp, IT architects can design flexible multiplatform and hybrid cloud application architectures. They can provide security the organization requires with the data access users need.

Dev/Ops

The need for secure, rapid software development is driving fast adoption of hybrid cloud by development organizations (Dev/Ops). Setting up infrastructure for new projects in the cloud is fast and affordable; it gives Dev/Ops the freedom to fail. When costs outstrip advantages, Dev/Ops can bring development environments back to its data centers.

Storage Architects

The Data Fabric extends the reach of storage architects and administrators into cloud architectures. It gives them the opportunity to leverage their existing skills and operational experience to take advantage of new deployment models that are on the leading edge of technology. With new tools at their disposal, they have more freedom to enable their users to be more productive and innovate in new ways.

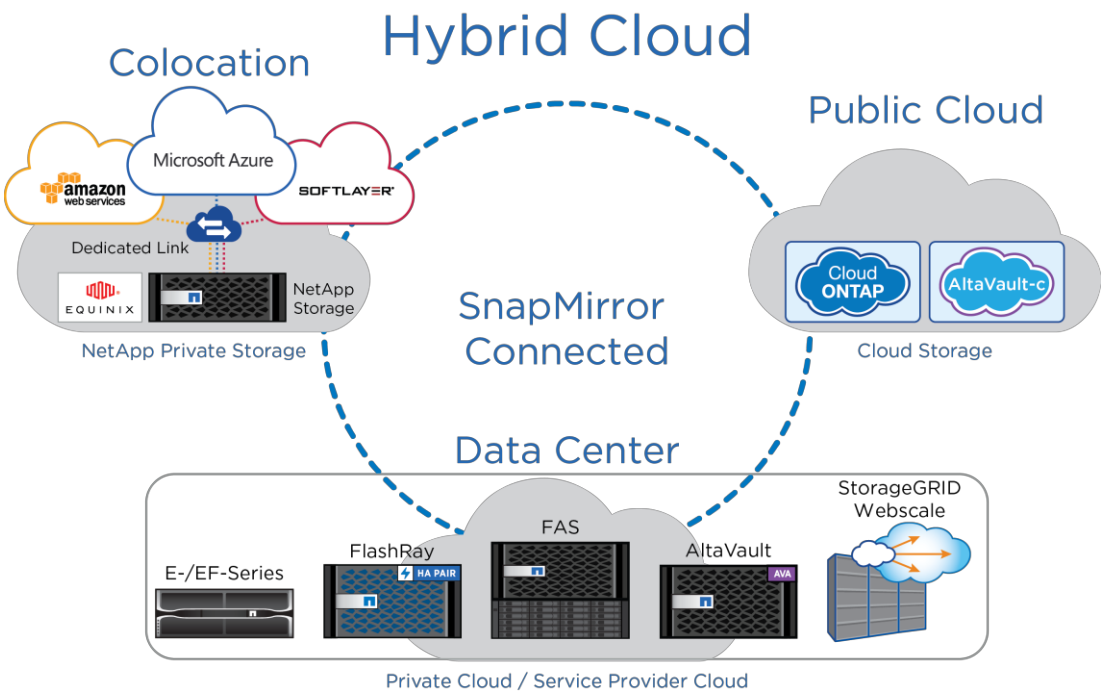
Cloud Service Providers

Cloud service providers (SPs) seek to scale their business up, drive costs down, and onboard customers quickly and easily. With a Data Fabric enabled by NetApp, SPs can build an efficient and dependable infrastructure that scales with demand. Operations are completely automatable, whether the cloud orchestration framework is commercial, open source, or custom built. Customers can be onboarded by enabling them to extend their Data Fabrics to the SP cloud, giving them control of their data while they utilize cloud services.

1.4 Deployment Models

IT has a wide variety of deployment models at its disposal, including private cloud in the data center, public cloud offerings from service providers, and hybrid cloud that combines multiple clouds. An organization may choose a combination of deployment models to define the topology and span of its data fabric.

Figure 1) Data Fabric deployment models.



Private Cloud

Private clouds can reside in an organization's own data center or be hosted in a remote facility. In either case, the hardware and software architecture choices are the same, and there are plenty to choose from.

Architects choose a commercial or open-source hypervisor and cloud orchestration framework. NetApp products tightly integrate data management capabilities with these ecosystems.

The storage solution choices range from purpose-built to software-defined storage.

Public Cloud

A cloud is called a "public cloud" when the services are rendered over a network that is open for public use.[1] Public clouds are made available by service providers who own and operate their own data centers and infrastructure. Although the largest cloud providers operate at a scale at which they can design proprietary architectures, service providers typically choose from the same options used by enterprises architecting private clouds. Doing so enables the service providers to focus less on infrastructure and more on their core business of service innovation.

[1] https://en.wikipedia.org/wiki/Cloud_computing.

NetApp Service Provider Partnerships

- NetApp integrates and partners with Amazon AWS, Microsoft Azure, and IBM SoftLayer
- 275 NetApp Service Provider Partners around the globe

Service providers utilizing NetApp infrastructure can enable customers to expand their data fabrics to the cloud by offering SnapMirror® or SnapVault® services. This enables the service provider's customer base to efficiently onboard data into the cloud for use with the service provider's services, paving the way for hybrid cloud architectures.

In addition, Cloud ONTAP® can be used to quickly create a Data ONTAP® endpoint in AWS, bringing the value of Data ONTAP data management to cloud storage.

Hybrid Cloud

Hybrid cloud is an IT delivery approach that leverages both on-premises and public cloud resources. NetApp believes a true hybrid cloud is one that customers can build on their terms. It is multi-cloud and multidirectional, integrating any combination of resources that are on premises, near the cloud, and in the cloud.

In its simplest form, a hybrid cloud deployment may consist of a FAS array in the corporate data center and Cloud ONTAP in AWS, connected using the SnapMirror transport to replicate data from one location to the other. This simple architecture establishes a data fabric, enabling application data to be served and managed the same way in both locations.

A hybrid cloud may also connect collocation managed and/or dedicated services with cloud resources. For example, with NetApp Private Storage (NPS), organizations can deploy a private FAS cluster in a colocation facility and use a network exchange to achieve lower latency when connecting to public cloud compute. NPS gives IT control of its data, from physical location to retention policies and SLOs, with data fabric connectivity and enables the use of the elastic compute capabilities of public clouds.

2 Data Fabric in Action

The Data Fabric is real right now and organizations are already benefiting from it. Service providers and enterprises are using a Data Fabric enabled by NetApp to transform their business in innovative ways. This section provides real-world examples of how service providers and enterprises are using their Data Fabrics today.

2.1 Delivering Highly Scalable, Highly Durable IaaS

A leading Service Provider needed to more effectively compete in the growing cloud space. Adding an IaaS model to their dedicated hosting service would enable them to further reduce costs to customers and provide new managed services.

The re-architected platform had to meet the demands of their Enterprise customers for high capacity, performance, durability, and availability. In addition, the infrastructure had to scale to enable them to balance cost with technical capabilities, growing as the business needed.

The Solution Approach

After exploring various architectures, they chose clustered Data ONTAP as the foundation for their IaaS platform.

- Always-on platform. Traditionally a downtime window is required to perform critical maintenance. In a multi-tenant environment a single downtime window doesn't work. Clustered Data ONTAP enables them to perform platform lifecycle management (hardware and software upgrades) non-disruptively.
- Scale. The cloud business is naturally dynamic. The clustered architecture enables the SP to match their cost to revenue, adding nodes to increase capacity as the needs increase.
- Quality of Service. Simplified performance management enables the SP to deliver per tenant SLOs. Customers are charged for the performance they are using, protecting the SP from potential financial losses.
- Replication. Regional availability is a key selling point with Enterprise-class SLA's, which require RPO's measured in minutes while providing data center separation on the continental scale. SnapMirror replication technology allows them to provide a level of disaster recovery at a price their customers can afford.
- A single architecture delivers multiple services and levels of service at differing cost points for customers (e.g., disk, flash). Having a single architecture streamlines the costs of operational overhead and improves margins.
- The flexible nature of the architecture allows the introduction of new services to market more quickly. The ability to support multiple protocols, multiple hypervisors and new applications was possible without significant re-investment.
- Automation and APIs. Cloud customers access their data via a portal. The SP developer teams were able to directly automate the NetApp infrastructure through the API layer.

The Data Fabric Value

By leveraging NetApp's expertise in data management, the Service Provider is able to maintain both the integrity and availability of data that is essential to their brand.

The data fabric allows the SP to focus on differentiated service offerings, revenue growth and customer experience. The single operational framework, enables the SP to easily and efficiently offer multiple services at different price points.

Service Providers utilizing NetApp infrastructure can enable customers to expand their data fabrics to their cloud by offering SnapMirror or SnapVault services. This capability allows for easy onboarding of customer data, paving the way for a hybrid cloud.

2.2 Delivering Multi-Cloud SaaS

Recognizing that its customers would begin moving to a "SaaS first" approach to IT deployment, an Enterprise Resource Planning (ERP) software vendor made a strategic decision to deploy a SaaS delivery model for its software.

To be successful, the company had to meet requirements from a variety of users. Customers required the ability to run their workloads on their chosen cloud. The organization's professional services team required the ability to quickly create dev/test environments for quickly onboarding customers. The IT team needed to preserve the core transactional database upon which the ERP application was running.

Solution Approach

The software vendor understood that it needed to rearchitect its application to get the best economics when running in the public cloud. At the same time, it wanted to preserve the core transactional database underneath the application. Software architects initially chose Amazon's EBS storage platform for the company's Microsoft SQL Server database. Very early on they discovered issues with this approach:

- The costs for scaling were unsustainable. The number of dev/test copies the SQL Server database required resulted in a high consumption of storage capacity on EBS that made EBS cost prohibitive for scaling their SaaS business.
- Onboarding new customers took longer and cost more than planned in the business model. The time required to clone the SQL Server data delayed onboarding new customers and increased new customer acquisition costs.
- Performance and availability SLAs did not satisfy all customers. The performance and availability SLAs that they need to satisfy their largest customers could not be satisfied cost-effectively by AWS alone. They needed the ability to choose among service providers where to host specific customers and to be able to subsequently change the service provider for a given application instance.
- The requirement for 15-minute incremental backups of the production database could not be met.

To provide dev/test copies for preproduction use, developers started hosting the company's Microsoft SQL Server databases on iSCSI LUNs on Cloud ONTAP, which placed its aggregates on EBS. By leveraging deduplication and other storage efficiency features of Cloud ONTAP, they were able to reduce the EBS storage footprint required by 90%. Using the storage cloning capabilities of Cloud ONTAP, they dramatically reduced the time needed to produce the dev/test database instances that their professional services team needed to onboard additional customers, enabling greater productivity and a faster ramp-up of customers into the new SaaS offering.

To meet the performance and availability SLAs for production customers and to enable choice of public cloud provider for a given instance, the software vendor is now architecting a solution that uses NPS in a colocation facility with connectivity to both AWS and Azure.

Data Fabric Value

This SaaS solution built on a Data Fabric enabled by NetApp offers a common foundational infrastructure for high availability and DR. Rather than depending on application-specific replication mechanisms and other capabilities that vary between AWS and Azure, the company leverages the HA and DR capabilities of clustered Data ONTAP, including 15-minute incremental backups for their production databases.

A single NPS deployment can serve multiple clouds, including AWS, Azure, and SoftLayer. No data movement is required to relocate individual production applications from one public cloud to the other, eliminating the time, cost, and complexity of copying the data.

2.3 Reducing Content Management Platform Sprawl

NetApp has many departmental intranet sites that were built using different content management systems (CMSs), such as Jive, WordPress, Drupal, Joomla, Oracle Content Management platform, and custom Java/HTML/CSS frameworks. Over time, these portals have grown in size and reach. A lack of common standards caused CMS deployments to sprawl across the company, making it difficult to control and manage the data.

Solution Approach

To gain control and stop the sprawl, NetApp IT introduced a cloud-based hosting platform that brings all of the intranet portals into a standard design framework. This platform is built on NPS and AWS Cloud. It uses the open-source WordPress CMS.

Each portal is assigned a midsize AWS EC2 compute with LAMP (Linux, Apache, MySQL, and PHP) stack and a NetApp branded WordPress IT blueprint. The portal contents are stored on a NetApp FAS6290 using the NPS deployment model. The FAS is accessed by the EC2 nodes using NFS-mounted volumes over a 1Gbps direct connect link.

Data Fabric Value

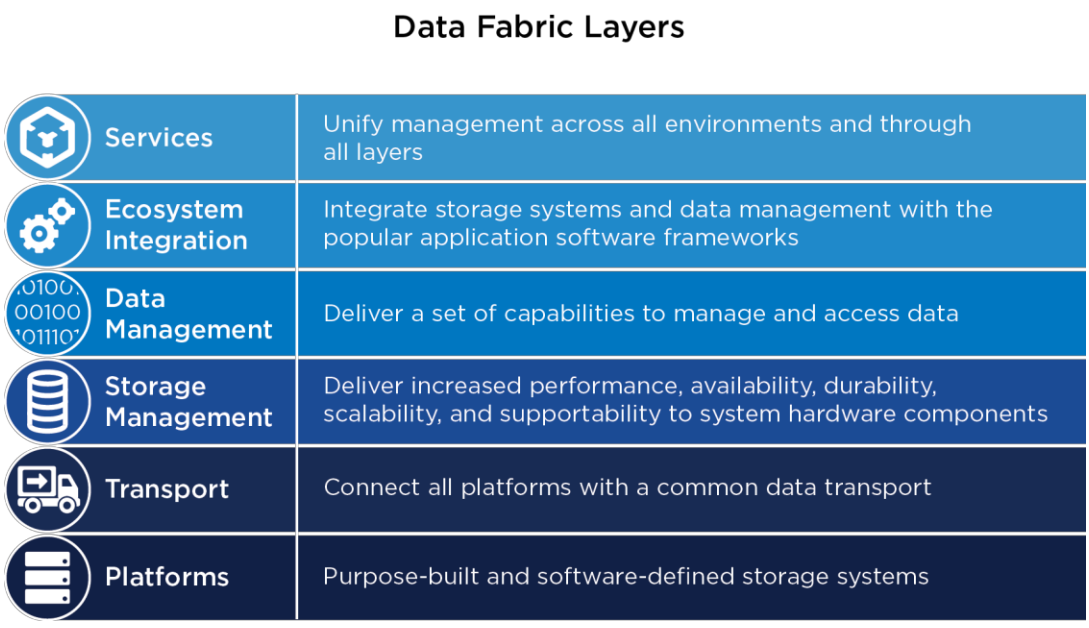
By separating the data from the applications, IT now has full custody of business data on its own storage and is able to comply with NetApp legal policies for data sovereignty, data security, and data lifecycle management. The Data Fabric allows NetApp IT to enforce web portal standards that are easy to use, manage, and support while giving full control of content management to BU users.

To date, NetApp IT has deployed 15 intranet portals using this model and now has a solid foundation upon which to expand.

3 Laying the Foundation for a Data Fabric

Creating a data fabric is about taking a data-centric view of IT infrastructure in every layer of the environment: platform, transport, storage management, data management, ecosystem integration, and services. The overall architecture is comprised of products and solutions that unbind data from underlying systems so that data can be accessed across the fabric. With NetApp, IT architects have many building blocks to choose from at each layer that have been designed with the principles of a true data fabric in mind. In this section we explore the options for laying the foundation for a Data Fabric enabled by NetApp.

Figure 2) Data fabric layers.



3.1 Platform Layer: Establishing the Endpoints of the Data Fabric

The platform layer is composed of the storage systems that are the building blocks for the endpoints of the data fabric. The endpoints vary in function for different workloads and form factor, and they can be purpose-built or software-defined storage systems. The deployment model of the endpoints defines the topology and span of the fabric.

Choice of Endpoints

Clustered Data ONTAP provides the foundation of the Data Fabric enabled by NetApp. It offers multiple deployment options, a large set of data and storage management capabilities, and deep ecosystem integration to enable a system that spans a broad range of environments and use cases.

Clustered Data ONTAP is the native operating system of NetApp purpose-built FAS storage systems. Data ONTAP software-defined storage offerings enable a fabric to span to commodity direct-attached storage (DAS) platforms with Data ONTAP Edge and the public cloud with Cloud ONTAP.

NetApp clustered Data ONTAP brings the same data management capabilities to disparate environments across the hybrid cloud.

Table 1) Available NetApp platform options.

Platform	Purpose-Built Storage Systems	Software-Defined Storage Systems
Clustered Data ONTAP		
FAS	✓	
All Flash FAS	✓	
FlexPod®	✓	
Cloud ONTAP		✓
Data ONTAP Edge		✓
E-Series	✓	
EF-Series all flash	✓	
AltaVault®	✓	✓
StorageGRID® Webscale	✓	✓

Other storage platforms, including third-party arrays, can be added based on workload and business requirements.

Purpose-Built Storage Systems

Purpose-built storage systems provide enterprise-grade storage that is highly scalable, performant, resilient, and available. NetApp systems include these options:

- File workloads: FAS
- Block workloads: FAS, E-Series
- Large object stores: StorageGRID Webscale appliance
- Backup to the cloud: AltaVault

In addition, turnkey converged infrastructures are available with FlexPod, which delivers a full stack with best-in-class components from vendors such as NetApp, Cisco, Microsoft, VMware, Citrix, and RedHat.

Software-Defined Storage Systems

Software-defined storage (SDS) offerings give customers greater flexibility when it comes to platform choice, deployment models, and licensing, without compromising enterprise data management capabilities.

NetApp offers several SDS solutions, including:

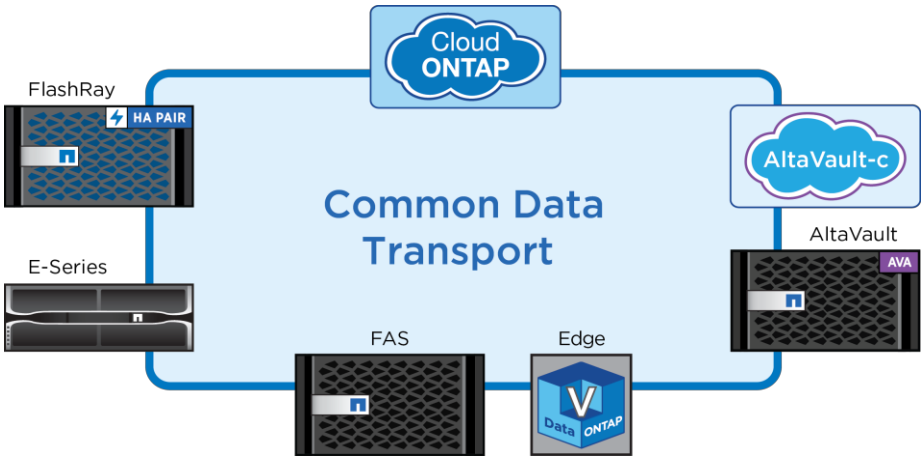
- File or block workloads: Cloud ONTAP or Data ONTAP Edge
- Large object stores: StorageGRID
- Backup to cloud object stores: AltaVault-c or AltaVault-v

Each of these products can be installed on third-party platforms, allowing for deployment in data centers, remote offices, or clouds.

3.2 Transport Layer

The transport layer provides a common data protocol that connects all platforms in the data fabric to seamlessly transfer data in bulk. This transport mechanism enables applications to access data where it is needed most: to a cloud, across storage tiers, or across clusters. Applications are not aware of any data movement.

Figure 3) A common data transport enables data to move seamlessly between platforms.



The Data ONTAP family of products (FAS, Cloud ONTAP, Edge) share a common WAFL® file system format. All other platforms shown in the fabric have their own native file system formats.

Common data transport enables the platforms to interoperate and move data efficiently between them, using SnapMirror replication or SnapVault backup capabilities. Not only does this transport enable interoperability, it also preserves deduplication and compression storage efficiencies, meaning data does not rehydrate when moving from one endpoint to

another. This transport is the fastest way to move data in bulk around the fabric. This is especially important when the fabric spans WANs for hybrid cloud architectures.

A data fabric transport provides cross-cluster data movement where the data can be consumed in a form native to each endpoint. For example, FlashRay® serves a given dataset to clients using its Fibre Channel protocol, but when the data is moved to FAS using SnapMirror, the FAS may serve the dataset to clients using iSCSI without transformation.

Envision the Future: Expanding Common Data Transport Endpoints

Today, common data transport is available only for Data ONTAP endpoints. It is being expanded to include E-Series and AltaVault.

Connectivity with Each Other and with Clouds

Because E-Series platforms plug into the fabric, they not only get the benefit of data exchange with FAS and each other, which is important for managing SLOs and total cost of ownership, but they also become cloud enabled. With common transport, data can be moved efficiently to and from these platforms to the cloud endpoints: Cloud ONTAP or NPS deployed FAS systems.

Efficient Transport, Faster Data Transfer

AltaVault can back up any storage array, but after it becomes a native member of the data fabric, it can leverage SnapVault directly over the common data transport. This integration provides an efficient transport, enabling faster data transfer.

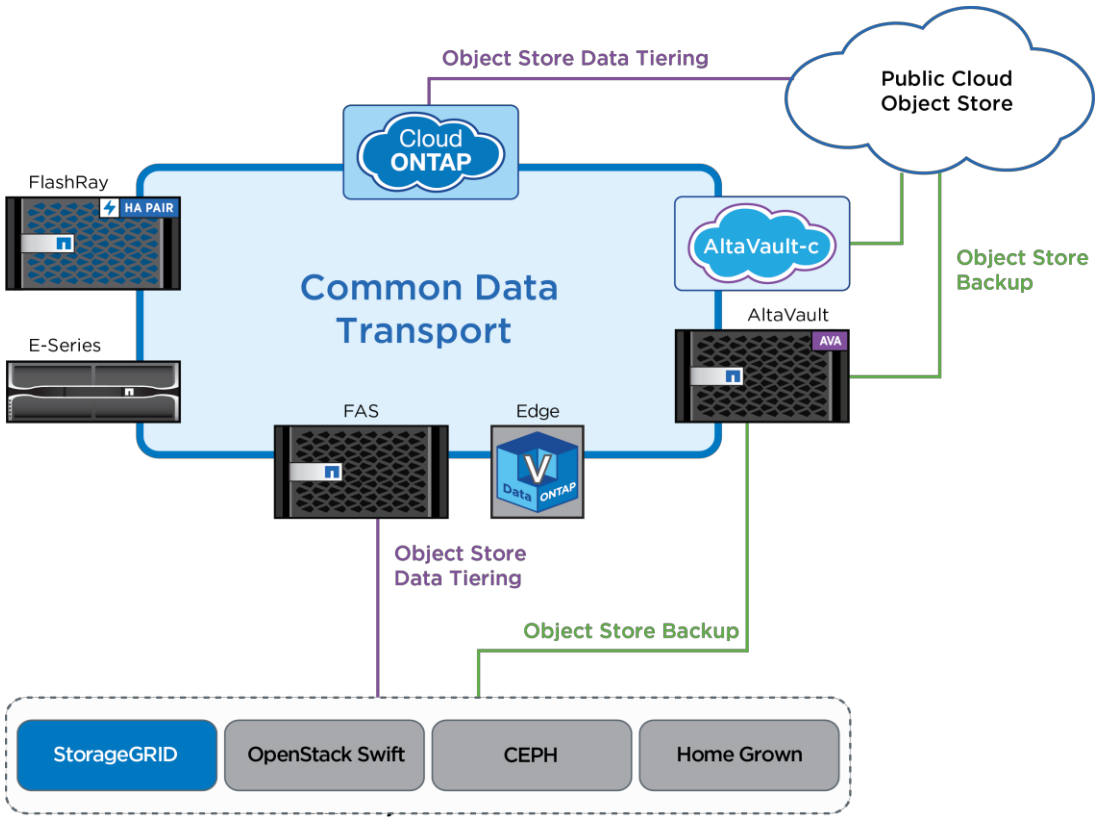
With Data ONTAP as a source, incremental Snapshot® data can be pushed in its native format to AltaVault, preserving storage efficiency. No data rehydration is required. This streamlined payload minimizes both source and destination system load, which, in turn, allows for less application impact, a greater number of concurrent backup and restoration sessions, and improved recovery point objectives (RPOs).

In addition, backing up to AltaVault using the common data transport mechanism allows for an “incremental forever” backup strategy, eliminating the need for periodic full backups. This implementation minimizes backup windows, provides faster data recovery, and reduces the storage capacity requirements on the destination.

Extending the Data Fabric

The more links the fabric has, the stronger it becomes, and the more solutions IT has at hand to address customer problems.

Figure 4) A Data Fabric enabled by NetApp can intelligently and automatically move data to the optimal tier.



Automated Data Tiering

Automated data tiering is the act of automatically and transparently moving active/online data between different classes of storage with different cost and performance, based on well-defined policies. The data remains accessible but experiences higher access latencies when residing on the lower tiers.

Data ONTAP Flash Pool™ aggregates are an example of automated data tiering in which an aggregate consists of a set of SSDs and HDDs. The HDDs are the slower, less costly, high-capacity tier. The SSDs are the faster but more expensive tier. With such a hybrid configuration, Data ONTAP makes sure that the hot data automatically gravitates to the SSD tier, allowing for the highest performance data access.

Automation is an important aspect of tiering because data has temperature, which is a dynamic property. What is hot today may turn cold tomorrow. A human would not be able to keep up by manually migrating data from one storage tier to another. Managing this data with custom management tools creates complexities for the customer and inefficiencies in data motion. Data ONTAP, however, understands the access patterns of the data, enabling it to intelligently and automatically move data to the tier that best suits the observed access patterns. The customer benefits by getting the best of all worlds: SSD-level performance for hot data, HDD prices for capacity, and automatic data placement to exploit both.

Envision the Future: Data Tiering

The same tiering concept can be applied using cloud technology, object stores in particular. In the case of FAS, consider an aggregate consisting of a set of SSDs and an object store bucket. Like with Flash Pool aggregates, the hot data gravitates to the SSDs, while the object store operates as the less costly, deeper, slower capacity tier.

The object store can take many forms in the data center, including StorageGRID, OpenStack Swift, CEPH, or even a homegrown implementation. The object service may be backed by various physical storage platforms, including E-Series and generic DAS systems.

Data tiering can be applied to public cloud storage as well. Cloud ONTAP will use the same technology to automatically tier data between AWS EBS storage and S3.

With Data ONTAP supporting object storage as a native tier, object stores become tightly woven into the fabric. This level of system integration allows customers flexibility in choosing storage architectures, SLO, and price points, while unifying data management.

StorageGRID Webscale

StorageGRID offers software-defined object storage, which can be mixed and matched into a single cluster.

Depending on how it is used, StorageGRID can be an integrated component of the data fabric, a standalone object fabric, or both.

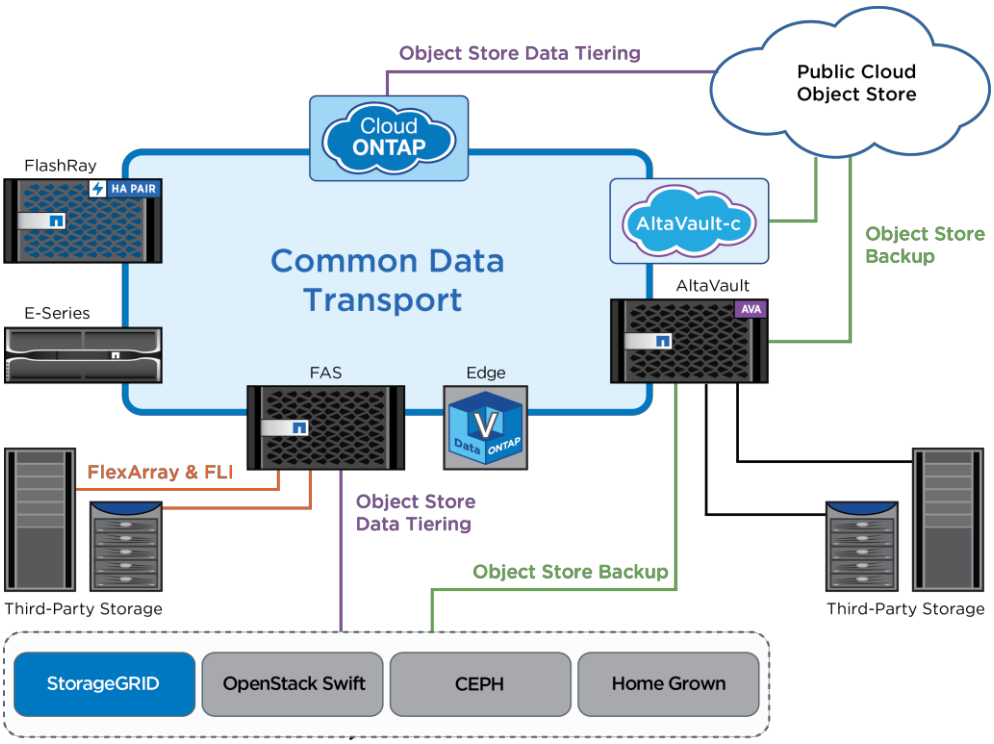
In the Data Fabric diagram shown in Figure 5, StorageGRID is an integrated part of the Data Fabric serving as a target for FAS object store data tiering and AltaVault backups.

In addition to those use cases, StorageGRID is an object-based fabric in its own right. Cloud applications can store and retrieve objects directly over its S3 or CDMI protocols. StorageGRID is a true multisite architecture and has advanced object management capabilities, such as geo-dispersed erasure coding, single global namespace, and policy-based object management and movement.

Extending the Data Fabric Further to Third-Party Platforms

Although the NetApp Data Fabric is optimized for NetApp products, it is also open to third-party products, as shown in Figure 5.

Figure 5) A Data Fabric enabled by NetApp can expand to third-party arrays.



Storage Array Investment Protection

Organizations can leverage investments in third-party arrays by bringing them into their data fabric using FlexArray®, a capability of clustered Data ONTAP. FlexArray enables third-party arrays to receive the benefits of native NetApp systems, including storage efficiency, multiprotocol support, common data transport connectivity, cloud enablement, and uniform data management processes and tools.

With this approach, IT can set up a disaster recovery fabric endpoint using SnapMirror to replicate from a third-party system to a FAS array or Cloud ONTAP.

In addition, customers may choose to move data that resides on third-party systems to native NetApp FAS storage. This action is known as foreign LUN import (FLI), which migrates the data to a NetApp storage pool. FLI allows for array migration, cloud migration, or data tiering for better SLO management.

Commodity Direct-Attached Storage

Direct-attached storage (DAS) systems can connect to the Data Fabric using Data ONTAP Edge to manage the storage. Data ONTAP Edge software installs on the server, creates a storage pool, and provides its full set of data management capabilities to applications and ecosystems.

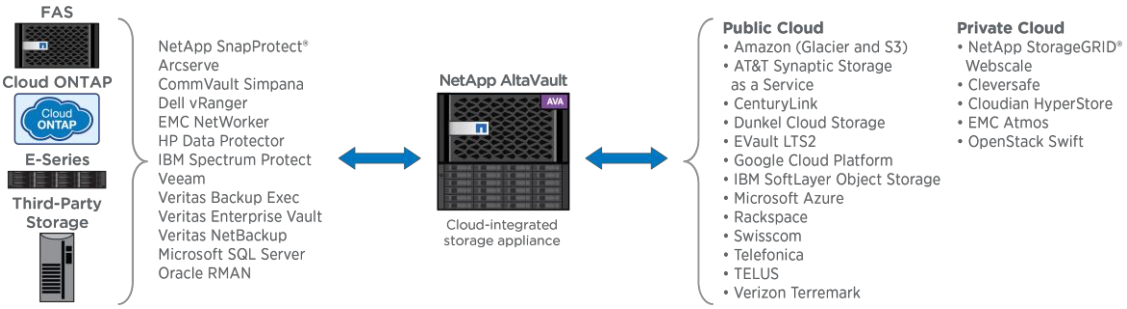
Open Source Software-Defined Storage

E-Series systems may be used for high-performance, cost-effective enterprise DAS with open-source SDS offerings, including CEPH and OpenStack Cinder. StorageGRID can provide an object store for OpenStack Swift.

Back Up and Restore Third-Party Arrays

An AltaVault appliance can back up any storage array to any cloud and supports a wide variety of backup software. After the data has been backed up into the Data Fabric, it can be restored to any storage array within the fabric.

Figure 6) AltaVault can back up any array to any cloud, with support for a wide variety of backup software options.



Object Repositories

The Data Fabric supports multiple use cases for object repositories, including a target for AltaVault backups and for object store data tiering. For these use cases, either Data ONTAP or AltaVault manages the data stored in the repository. The implementation and configuration of the repository itself are flexible. For on-premises deployments, the software that implements the object protocol can be NetApp StorageGRID, open-source solutions (OpenStack Swift, CEPH), or even homegrown solutions that meet the AWS S3 specification. The back-end physical storage for the repository can be a NetApp storage array or any third-party storage server. For cloud deployments, Cloud ONTAP will support object store data tiering to AWS S3.

Expanding Endpoints to More Clouds

The multi-cloud endpoints not only provide customers with a choice of environments, but also enable them to avoid cloud vendor lock-in, protect their assets in the event a particular cloud is compromised, and manage their assets in a consistent and seamless way regardless of geographical location.

The NetApp Data Fabric supports a wide array of virtualized environments and clouds today and will continue to expand to more clouds.

A Case for Multi-Cloud Endpoints

CodeSpaces.com was a startup that offered code repository services in the cloud. It was committed exclusively to the AWS cloud platform, using EC2 for compute, EBS and S3 for primary storage, and Glacier for backups.

Hackers acquired the CodeSpaces AWS account credentials and systematically deleted all of the company's cloud resources. Because CodeSpaces did not own or control any of its data, the security breach automatically put a multimillion-dollar startup out of business.

Multi-cloud endpoints could have helped CodeSpaces protect its data, its business, and its customers. For example, AltaVault could have backed up CodeSpaces data to a secondary cloud of its choice and then easily recovered the hacked data. Alternatively, Cloud ONTAP could have been used with SnapVault to copy the data to a service provider offering NetApp ingest backup as a service.

3.3 Storage Management Layer

The storage management layer of the data fabric includes the technologies that deliver high availability, scale-out architecture, and storage efficiency to hardware components. Automatic data tiering within the storage system is also managed by technologies associated with this layer.

High Availability and Durability

High availability is a measure of data accessibility. A highly available system is expected to deliver high levels of data service and automatically and transparently adapt in the event of partial system failures of disks, networks, or nodes.

Storage durability measures the probability of data loss in the event of failure.

Storage systems use a variety of techniques to manage durability and availability, including RAID, erasure coding, replication, and clustered node failover with multipathing.

Availability and durability ratings vary depending on storage system architecture, so it is important to understand whether the rated measures for a given environment and storage tier meet the requirements of the application.

Scale-Out Architecture

Planned downtime events account for a large majority of the downtime that IT departments experience. A scale-out storage architecture is one that lets you expand your storage capabilities as needs change.

Clustered Data ONTAP can scale both vertically and horizontally without service disruption, allowing the following operations:

- Expand flash capacity to increase performance
- Add high-density drives to increase raw capacity
- Scale up to create a higher end storage array
- Scale out horizontally, adding storage nodes, to distribute workloads

Third-party storage arrays can gain these same benefits when plugged into the fabric using Data ONTAP FlexArray capabilities.

For object storage, StorageGRID Webscale is massively scalable. Its unique software-defined architecture supports billions of objects and tens of petabytes of storage spanning numerous locations in a single namespace.

Storage Management Efficiency

Maximizing storage utilization; reducing data center space, power, and cooling requirements; simplifying data management; and accelerating storage processes are all benefits of storage efficiency technologies inherent in clustered Data ONTAP and available across the Data Fabric.

Table 2) NetApp Data ONTAP storage efficiency technologies.

Technology	Description
Shared storage HA	Provides a single copy of data, saving up to three copies of replication (in FAS systems).
Flash Pool	Combines solid-state disks (SSDs) and traditional hard disk drives (HDDs); offers automated storage tiering between different storage media.
RAID	Allows safe, highly durable data storage on even the lowest cost SATA disks.

3.4 Data Management Layer

The data management layer makes it possible for IT to deliver consistent data services across all environments. It includes:

- Data management efficiencies
- Security technologies
- Replication and backup technologies
- Data access protocols

Data Management Efficiency

Table 3) NetApp data management efficiencies.

Technology	Description
Snapshot copies	Provides near-instantaneous point-in-time copies that protect data with no performance impact, using minimal storage space.
Deduplication	Automatically removes duplicate data blocks.
Data compression	Offers inline or postprocessing compression.
Thin provisioning	Allocates space for LUNs and volumes on demand, instead of reserving them up front, resulting in physical capacity being consumed only when needed to store unique new data.
Cloning	Allows instant replication of data files, LUNs, and volumes as transparent, virtual copies without requiring additional storage at the time of creation.

Security Technologies

A defense-in-depth strategy relies on layers of defenses to protect systems and data. Such a strategy requires that all components across the hybrid cloud environment support a consistent security architecture and policy.

Example security mechanisms include:

- **Secure data at rest.** Cloud ONTAP implements AES 256-bit software encryption capabilities to secure data at rest on public cloud storage. FAS and E-Series systems implement hardware-based security with self-encrypting drives.
- **Secure data in flight and at rest.** AltaVault encrypts all data prior to transmitting it over the wire to the target cloud storage.

- **Control encryption keys.** All Data ONTAP and AltaVault encryption capabilities enable users to have end-to-end control of their security policies. Users manage their encryption keys with their own KMIP-compliant key management servers.
- **Control data access.** Role-based access controls (RBAC) are used for managing a set of actions that a user or administrator may perform in a given environment.
- **Multi-tenancy.** Clustered Data ONTAP enables multiple tenants to share the resources of a cluster with its storage virtual machine capabilities. Cluster administrators set the resource limits and QoS levels for each tenant. Tenants have administrative control of their slice of the system, while remaining isolated from others.

A Case for Security: Providing Secure, Global Access to Corporate Data

NetApp users in offices around the globe need secure access to corporate business data for quarterly reporting and analytics. The centralized data warehouse had become inadequate. Report demand is bursty in nature, which often resulted in access bottlenecks.

To provide globally distributed users secure access to corporate data, NetApp IT deploys Cloud ONTAP in multiple AWS regions. SnapMirror replicates data from the centralized data warehouse to the various Cloud ONTAP endpoints. Users access reports locally in their region using a business analytics web portal. When demand for reports is fulfilled, the Cloud ONTAP instances are torn down.

NetApp follows a strict data governance policy, and hosting business data on public cloud storage met requirements due to Cloud ONTAP encryption capabilities. IT controls and manages the encryption keys for the data residing on cloud storage. The data is secured at rest and permanently inaccessible when the cloud resources are torn down.

Replication Technologies

Replication technologies are used to protect against data loss of any kind. Organizations use NetApp SnapMirror to satisfy strict RPO and RTO requirements while controlling costs and improving operational processes. SnapMirror is a single replication solution that can be used across the Data Fabric.

SnapMirror is ideal for moving data across regions and clouds, because it is a resilient IP-based protocol that is optimized for data transfer over WANs. A new SnapMirror session performs a baseline copy in which all data in a volume is replicated to the destination. Incremental updates are performed continuously, leveraging Snapshot copies to minimize data transfer, sending only the blocks that have changed. Storage efficiencies remain intact, where data that has been compressed or deduplicated on the source stays that way over the wire and at the destination.

Backup and Restore

As previously mentioned, AltaVault can back up any storage array to any cloud object repository.

Alternatively, SnapVault can be used with the common data transport to back up Snapshot copies from any endpoint in the Data Fabric. SnapVault enables data stored on multiple systems to be backed up to a central, secondary system quickly and efficiently as read-only Snapshot copies.

For example, customers can use Data ONTAP Edge on third-party servers to expand their fabric to remote offices and use SnapVault to easily and efficiently back up data from each remote office to the data center.

Data Access Protocols

An extensive data fabric must provide data access in ways that meet the needs of today's applications. The protocols supporting data access vary depending on the platform.

Table 4) NetApp data access protocols by system.

System	Data Access Protocols
Clustered Data ONTAP	NFS, CIFS, iSCSI, FC
StorageGRID	S3, CDMI, NFS gateway services, Swift (coming soon)
AltaVault backup software interface	NFS, SMB
E-Series	FC, iSCSI

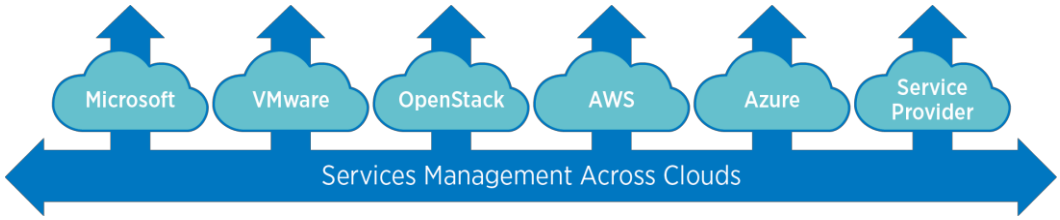
As the various platforms become data fabric enabled, common data transport facilitates the movement of data between them and enables data to be served by the protocols native to each system.

4 Building Data Fabric Capabilities Up and Out

Increasingly, IT organizations do not have specialists dedicated to managing either the server, networking, or storage infrastructures. Instead, they have IT generalists managing the entire infrastructure. Traditional approaches of monitoring application servers, networking, and storage infrastructure, typically with different vendor-provided tools, do not fit the new data center paradigm.

To address this challenge and to enable organizations to choose from a broad ecosystem of technology and cloud service providers, NetApp has developed a technology stack.

Figure 7) The Data Fabric enabled by NetApp integrates vertically and horizontally.

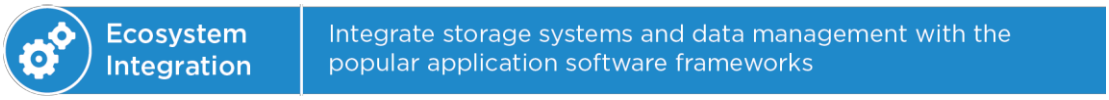


The lowest layer of this technology stack includes APIs that enable automation and management integration with both third-party products and custom-built tools and workflows. As a result of these ecosystem integrations:

- Enterprises do not need to be specialized on NetApp. They just need to understand their ecosystem products.
- Enterprises eliminate the need for a lot of homegrown tools by using NetApp tools and frameworks.

- Enterprises can protect applications (including databases, SharePoint, and Exchange) without having to understand the data protection mechanisms of those applications.
- Enterprises can tie their legacy systems into the modern infrastructure without having to change processes.

4.1 Ecosystem Integration Layer



NetApp is making significant investments in integrating our innovative data management capabilities with the virtualization and cloud management tools on which our customers rely.

NetApp contributes to OpenStack to support customers that choose to leverage its open-source management ecosystem.

OnCommand® Cloud Manager simplifies the creation and management of Cloud ONTAP instances in the cloud.

These management plane integrations are built using the same underlying APIs that are available to customers and partners.

VMware

The NetApp strategy for ecosystem integration with VMware is to enable self-service through VM granular data management that storage administrators can safely expose to VM and application owners.

Initially, NetApp added functionality such as rapid VM cloning to the native VMware vSphere user interface (UI). Today, deep integration with VMware means that the data management functionality of the NetApp Data Fabric can be seamlessly leveraged through vSphere native management APIs and UIs. These native integrations include:

- SnapMirror integration with VMware Site Recovery Manager for automation of DR testing and DR failover/failback scenarios
- VAAI array integration with VMware vStorage APIs to enable offloading of cloning and other storage-intensive operations
- VVols integration, which enables service-level aware provisioning and management of virtual disks and VMs and virtual disk granular data management integration between vSphere and the NetApp Data Fabric
- vRealize Operations (vROps) Management Pack, which extends vROps with metrics and analytical views specific to NetApp for clustered Data ONTAP storage systems
- Qualification of NPS solutions to run attached to VMware public cloud platform, vCloud Air

Microsoft Windows Server and Azure Cloud Integrations

Microsoft technologies for the private cloud are based on Microsoft Hyper-V and System Center Virtual Machine Manager (VMM). For the public cloud, the technologies are based on the Microsoft Azure cloud platform. To connect multiple private and public clouds into a single IT infrastructure, Microsoft offers Azure Site Recovery (ASR) for hybrid cloud disaster recovery. ASR enables virtual machines to move between Hyper-V servers within a single site and between multiple sites in secondary private cloud data centers as well as Azure cloud data centers.

Organizations that require replication between their private and Azure public clouds for Azure site recovery can replication their Data Fabric SAN to NPS.

NetApp integrations with Microsoft are enabled by the combination of protocol-level capabilities and industry-standards-based storage management integration with Microsoft System Center. In the management domain, these enable the Microsoft System Center administrator to:

- Automate workflows using System Center Virtual Machine Manager and System Center Orchestrator for workflow automation
- Monitor server and storage availability and capacity for Microsoft Windows Server Hyper-V VMs
- Isolate problems using System Center Operations Manager alerts and health explorer views
- Enable high availability and load balancing of management servers with OpsMgr Management Server Resource Pool
- Leverage Windows Azure Pack (WAP) to provide service provider capabilities on top of clustered Data ONTAP in a private cloud architecture

OpenStack

There is significant open-source innovation around cloud platforms, and OpenStack is the leading open-source cloud platform. NetApp integrates OpenStack into the data fabric to make deployment of cloud services simpler, faster, and more scalable.

NetApp OpenStack block storage (Cinder) integrations include:

- **Storage provisioning and data management.** Specific NetApp drivers for Data ONTAP, E-Series, and EF-Series systems enable NetApp storage provisioning and data management capabilities.
- **Storage service catalog capabilities.** NetApp Cinder drivers allow IT to create a catalog of storage capabilities that meet diverse application and tenant needs for efficiency, performance, availability, and protection.
- **Enhanced persistent instance creation with copy offload.** The NetApp Cinder drivers for clustered Data ONTAP make use of NetApp cloning technology to quickly and efficiently create multiple virtual machines from Glance images.

With shared file systems underpinning much of the total storage shipped worldwide, NetApp is driving the addition of a shared file system service known as Manila.

NetApp Commitment to OpenStack

NetApp is a charter member of the OpenStack Foundation and has been a contributor since 2011. NetApp has made significant contributions in storage-related functionality to releases of OpenStack spanning from Essex to Liberty.

Containers

Designing applications as a collection of microservices housed in containers is an important new trend. The rapid adoption of container technology is driving a new set of data management and infrastructure requirements, especially for applications that benefit from leveraging advanced storage functionality. To meet this need, Docker and its ecosystem partners are going beyond stateless use cases and creating more robust deployment capabilities.

Flocker container management software from ClusterHQ enables operations teams to directly attach a storage volume to a container and then migrate the container, along with its data, between servers. Container volumes are no longer tied to a single host; instead, they can follow the container as it is moved between hosts. By decoupling the container from the volume, Flocker enables a new set of use cases where applications with persistent data can fully realize the benefits of a containerized deployment method.

Today, NetApp supports Flocker in conjunction with an OpenStack environment. In the near future NetApp will also be releasing a native integration that will not require OpenStack for Flocker use with clustered Data ONTAP.

Cloud Manager

OnCommand Cloud Manager is the primary portal for establishing Cloud ONTAP endpoints in public clouds and managing cloud resources in the data fabric. It has simplified wizards for common workflows such as creation, resource provisioning, and replication. Hybrid clouds may be constructed by simply dragging and dropping FAS systems to Cloud ONTAP systems and establishing a SnapMirror session for data replication.

In addition, OnCommand Cloud Manager exports a RESTful API set so that higher level tools can automate operations.

Envision the Future: Expanding the Reach of the Data Fabric

Cloud Manager and Cloud ONTAP are available for AWS today and will support additional clouds as we expand the reach of the Data Fabric. With support for Microsoft Azure coming soon, customers will be able to use a single Cloud Manager portal to manage data across their multi-cloud data fabric in a uniform way. In addition, AltaVault will be added to Cloud Manager's responsibilities, so cloud-resident backup services can easily be instantiated and policies established for Cloud ONTAP volumes.

Application Integration

The NetApp SnapDrive®, SnapManager®, SnapCenter®, and Snap Creator® family of products empowers self-service data management by DBAs and SharePoint and Exchange application owners. They enable customers to leverage the data management capabilities of clustered Data ONTAP and the replication capabilities of the data transport to automate critical enterprise application lifecycle tasks, including:

- Simplifying storage layout, planning, backup, and restore operations
- Reducing application server data recovery times from hours to minutes leveraging NetApp Snapshot, SnapMirror, and SnapVault technologies
- Providing clone lifecycle management to accelerate deployment of new releases and new applications

In addition to traditional on-premises deployment models, these application integrations support NPS and Cloud ONTAP deployment models as well as hybrid cloud solutions where the application data is transported by SnapMirror or SnapVault across sites, clouds, and deployment types.

Fully integrated solutions include:

- Microsoft SQL Server
- Microsoft SharePoint

- Microsoft Exchange
- Oracle databases
- SAP systems

In addition, Snap Creator and SnapCenter provide a framework for extending these capabilities to be leveraged for additional commercial databases and applications as well as for custom databases and applications.

A Case for Application Integration: Addressing SaaS Security

NetApp actively uses hundreds of software-as-a-service (SaaS) applications to run its business, and NetApp IT is continuously evaluating new SaaS products. IT security governance dictates that SaaS vendors have security certifications (for example, SOC1, SOC2), have periodic audit and review processes in place, and can demonstrate the ability to enforce needed security controls using the latest technologies. Most of the small to midsize SaaS providers do not qualify and therefore fail IT security screening.

As a result, NetApp and other enterprises cannot adopt many SaaS offerings even if they offer the most compelling features, cost efficiency, and support.

To address this issue, NetApp is actively working with SaaS vendors to develop support for NPS deployments hosting the business data. The SaaS platform would have the flexibility to use cloud storage or NPS storage for the business data.

By separating the business and application data, NetApp IT is able to adopt a broader range of SaaS services while using the data fabric to take full custody of its business data. Security policies are enforced and data lifecycle managed using standard data management operational processes across environments.

Backup Management Integration

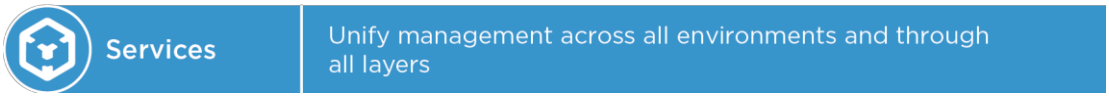
A key use case for data transport is backup and recovery. Backup and recovery is a distinct specialty from storage. Many enterprises have a dedicated teams and infrastructures for their enterprise backup architectures. For NetApp data transport to gain rapid adoption in performing backup and recovery for existing on-premises deployments, it is critical that it be able to fit in with the customer's chosen enterprise backup management architecture. To accomplish this, NetApp has partnered with most of the major enterprise backup management software vendors to provide integrated management of the Data Fabric:

- CommVault Simpana software is fully integrated with NetApp Snapshot technology for primary, vault, and mirror data. It also includes deduplication-aware replication and tiering across storage and cloud repositories. Simpana is also supported in NPS deployment configurations.
- Symantec NetBackup integration with Replication Director enables a similar value proposition for NetBackup users.

Envision the Future: Backup Management

We are working with many other enterprise backup management vendors to provide similar integration capabilities with Snapshot, SnapMirror, and SnapVault and expect to be significantly broadening the set of partner integrations in this space.

4.2 Services Layer



The services layer helps manage the Data Fabric effectively by providing consistency across platforms, sites, and environments. As the Data Fabric evolves, the tools to manage a data-centric approach will evolve with it. NetApp continually optimizes services management tools to bring them into alignment with other cloud management platforms and our Data Fabric vision.

Cloud management platforms are used for orchestration across clouds and the Data Fabric. Typically, these include a range of capabilities, including:

- Automation and self-service mechanisms
- Transparency for governance
- Metering and billing
- Workload optimization across public, private, and hybrid cloud environments

Today all NetApp products are automatable using APIs. Data ONTAP has native NetApp Manageability SDKs, which serve as a foundation for higher level management services. NetApp OnCommand Unified Manager interfaces with Data ONTAP NetApp Manageability SDKs and itself presents an API for programmatic monitoring of multiple nodes and clusters within the Data Fabric.

NetApp OnCommand data management software gives IT organizations the tools they need to standardize data management across resources in their Data Fabric. It interoperates with clustered Data ONTAP to simplify data management of on-premises and hybrid cloud environments. It plays a central role in helping IT meet storage SLAs, minimize risks, and boost performance.

Optimize Workloads Across the Data Fabric

OnCommand Insight provides a cross-domain view of performance metrics, including application performance, datastore performance, virtual machine performance, and storage infrastructure performance. It analyzes tier assignments and enables load-balancing of an organization's entire application portfolio across the Data Fabric.

OnCommand Insight also helps improve application performance and drive up the efficiency levels of existing storage resources so that organizations can maximize resource investment. It lets IT administrators manage storage as an end-to-end service and to integrate storage into the entire IT service delivery chain.

IT can improve services management by using NetApp foreign LUN import technology to move the data that resides on third-party systems to native NetApp FAS storage in a NetApp storage pool for array migration, cloud migration, or data tiering.

Automate Workflows Across the Data Fabric

Although orchestration solutions prove to be beneficial for end-to-end automation, they lack a comprehensive storage component to meet customer process needs. As seen in other domains, such as monitoring, an expert storage solution is required in order to address storage automation requirements.

NetApp OnCommand Workflow Automation (WFA) bridges the gap between data center orchestration solutions and customer requirements for storage automation. WFA is an automation framework for storage services. Using it, customers can realize the capabilities of the data fabric by automating processes including provisioning, migration, replication, decommissioning, and cloning. WFA enables:

- Cross-environment management, including conversions from one virtualized environment to another
- Cross-environment monitoring of capacity, performance, showbacks, and billbacks
- Cross-environment capacity planning

The automated workflows defined using WFA can be invoked in a variety of ways:

- From the native WFA GUI offering one-click execution by operators
- From data center orchestration platforms such as VMware vRealize Orchestrator, Microsoft System Center Orchestrator, and OpenStack
- From custom-built automation platforms using WFA RESTful APIs

Workflow Automation and Hybrid Cloud

With WFA, a single workflow can automatically launch a set of Cloud ONTAP instances across AWS regions, set up SnapMirror sessions between them and FAS in the data center, replicate data, and create NFS and/or iSCSI exports to make the data available to applications running in EC2.

Migrate Virtual Machines Across Hypervisors and Clouds

To migrate VMs across hypervisors and clouds, data must be independent from the hypervisor or cloud in which the application that is using it is running. Traditional techniques can take hours or days of downtime because of the amount of data that is involved.

To perform the necessary data format conversions with minutes of downtime, organizations can use NetApp OnCommand Shift to convert VMs between VMware ESX/ESXi and Microsoft Hyper-V. OnCommand Shift leverages the data management and rapid cloning and block sharing capabilities of clustered Data ONTAP.

Proactively Monitor the Data Fabric

AutoSupport™ checks the health of NetApp systems enabled with AutoSupport on a continual basis across any cloud. It is integrated with the My AutoSupport service, a suite of web-based applications hosted on the NetApp Support site and accessible through a web browser. Using the data from AutoSupport, My AutoSupport proactively identifies storage infrastructure issues through a continuous health-check feature and automatically provides guidance on remedial actions that help increase uptime and avoid disruptions to business.

5 Conclusion

As the landscape of enterprise IT rapidly changes to include a wider range of technologies and geographically dispersed applications and data, new data management challenges arise.

The Data Fabric enabled by NetApp is a collection of integrated technologies and services designed to deliver high-value data management solutions, which are greater than the sum of their parts. Platforms (physical and software defined) establish the datastore endpoints in various environments. A common data transport links them together. Management tools, APIs, and ecosystem integration make them easy to use collectively. Layered services provide overall visibility and control of business operations.

The multi-cloud capabilities of the data fabric provide organizations with a choice of environments for running their applications. This flexibility enables a wider range of services to choose from to meet application needs and business requirements. Assets can be protected, and access maintained in the event a particular cloud is compromised. Cloud vendor lock-in can be avoided. All while managing data in a consistent and seamless way regardless of geographical location.

A Data Fabric enabled by NetApp can be applied to address real data management problems today, enabling IT organizations to safely span data centers and clouds, while retaining ownership and control of their data.

You can get started by laying your foundation with clustered Data ONTAP and grow the capabilities and endpoints of your data fabric from there. A wide variety of deployment models are available, ranging from FlexPod converged infrastructure for data centers to Cloud ONTAP and NetApp Private Storage for clouds.

NetApp is committed to evolving the data fabric to fulfill our vision of delivering the ultimate data management paradigm. Deeper product and ecosystem integration and wider endpoint coverage will continue, giving you more choice over application environments and delivery models. Data management services will increase to give you greater visibility and control over your data regardless of where it physically exists. We're not building it alone. NetApp is also actively working with partners to leverage their offerings to jointly bring solutions to market.

NetApp has the right technology, the right business model, and the right vision, enabling us to be at the forefront of delivering unified data management across clouds. As technologies and solutions continue to evolve, customers will be empowered to expand their Data Fabric to take advantage of new capabilities and meet new business needs.

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

NetApp
www.netapp.com