White Paper

NetApp Data Fabric Architecture Fundamentals

Joe CaraDonna, Arthur Lent, NetApp

September 2016 | Version 3 | WP-7218

Abstract

Hybrid and multicloud deployment models are the new normal for enterprise IT organizations. With these mixed environments, new challenges around data management emerge.

NetApp's vision for data management is a data fabric that seamlessly connects different clouds, whether they are private, public, or hybrid environments. A data fabric unifies data management across distributed resources to allow consistency and control of data mobility, security, visibility, protection, and access.

In this paper we define the data fabric and its architecture, discuss usage examples, describe deployment models, and reveal how NetApp's data fabric is evolving.

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—combinations of private and public cloud resources—are 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 data stores or databases. Data has value: it is an important asset on which businesses depend. Data has mass: it takes time to move it and resources to store it. 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:

- **Protecting data and addressing security issues.** Wherever an organization's data resides, IT is still responsible for data security, data protection, and data governance to make sure of regulatory compliance.
- **Inability to move data.** After an organization's data is in a particular cloud, it can be 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 new and existing 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 storage systems. 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 in the most appropriate place.

1.1 About This White Paper

The purpose of this paper is to describe how NetApp's data fabric 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: the value of broad ecosystem integration, spanning data center and cloud environments
- **Section 5:** Conclusion

Readers who seek a high-level 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

The 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. NetApp's data fabric vision is continually evolving. Over time, the Data 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, the Data Fabric defines the NetApp® 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 to meet the needs of their applications and the freedom to change them whenever they want.

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 systems, 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 common tools and processes regardless of where it resides.

When a data 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 NetApp's data fabric, organizations can increase efficiency, improve IT responsiveness, and ultimately accelerate innovation.

Enterprise CEOs

NetApp's data fabric enables enterprise CEOs to foster an environment that stimulates innovation. They can improve the use of business resources through agility. Agility enables an organization to move at the speed of smaller companies without sacrificing the ability to meet the business and compliance requirements of their industry. They can have confidence that the organization's data is secure.

Enterprise CIOs

When it comes to meeting the needs of the business, CIOs need to maintain their existing environments and security posture while adopting promising technologies. With NetApp's data fabric, CIOs gain the freedom to make the best decisions for the business by making sure of secure access to data wherever it is needed and accelerating innovation with fewer resources. As hybrid cloud becomes the new normal for IT, CIOs need to protect, manage, and make sure of compliance of their organizations' data no matter where it resides.

IT Architects

Heads of IT infrastructure need to satisfy diverse service-level objectives (SLOs). Workloads require different availability, performance, cost, security, and access. With NetApp's data fabric, IT architects have flexibility in their design of hybrid cloud architectures. They can ensure access control by providing the security that the organization requires with the data access users need.

Application Owners

The need for secure, rapid software development is driving fast adoption of hybrid cloud by application development teams. Setting up infrastructure for new projects in the cloud is fast and affordable; it gives the ability to adopt a DevOps paradigm and gives pilot projects the freedom to fail. When costs outstrip advantages, application owners can bring development and/or production environments back to their 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 NetApp's data fabric, 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 on boarded by enabling them to extend their data fabrics to the SP cloud, giving them control of their data while they utilize cloud services.

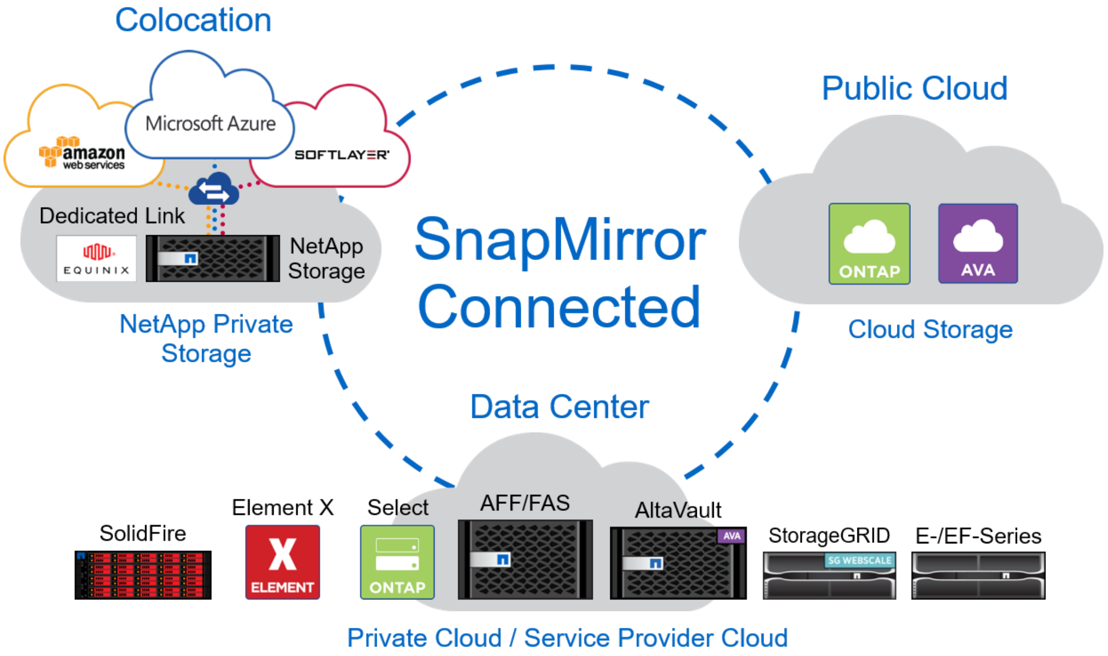
1.4 Deployment Models

An organization may choose a combination of cloud deployment models to define the topology and span of its data fabric.

IT has a wide variety of deployment models at its disposal, including:

- Private cloud in the data center
- Public cloud offerings from service providers
- Hybrid cloud that combines private and public resources
- Multiple clouds composed of two or more combinations of private or public clouds

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 from which to choose.

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

Public clouds are made available by service providers that 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.

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 NetApp SnapMirror® or NetApp 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, NetApp ONTAP® Cloud software can be used to quickly create an ONTAP endpoint in AWS, bringing the value of 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 multicloud 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 ONTAP Cloud 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 colocation 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 connect it to public cloud compute. This deployment model allows for a low-latency hybrid architecture, combining the security of private storage with the elasticity of public cloud compute. IT departments retain control of their data, from its physical location to retention policies and SLOs, and gain the benefit of data mobility across environments through data fabric connectivity.

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 NetApp's data fabric to transform their business in innovative ways. This section provides real-world examples of how service providers and enterprises are using data fabrics today.

2.1 Delivering Highly Scalable, Durable, Secure IaaS

A leading service provider needed to more effectively compete in the growing cloud space. Adding an IaaS model to its dedicated hosting service would enable the SP to further reduce costs to customers and provide new securely managed services.

The rearchitected platform had to meet the demands of the SP's enterprise customers for high capacity, performance, durability, and availability, while still maintaining their security posture. In addition, the infrastructure had to scale to enable the SP to balance cost with technical capabilities, growing as the business needed.

Solution Approach

After exploring various architectures, the service provider chose clustered Data ONTAP as the foundation for its IaaS platform based on the following factors:

- **Always-on platform.** Traditionally a downtime window is required to perform critical maintenance. In a multitenant environment, a single downtime window does not work. Clustered Data ONTAP enables them to perform platform lifecycle management (hardware and software upgrades) nondisruptively.

- **Scale.** The cloud business is naturally dynamic. The clustered architecture enables the SP to match its 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 SLAs, which require RPOs 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.
- **Single architecture.** A single architecture delivers multiple services and levels of service at differing cost points for customers (for example, disk, flash). Having a single architecture streamlines the costs of operational overhead and improves margins.
- **Flexibility.** 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 reinvestment.
- **Security.** Virtualization within the IaaS cloud solution expands the footprint and threat landscape for security. The SP understood threat vectors such as virtual network, hypervisor, VM-based rootkits, and colocation. The SP implemented the security plan for the cloud-based IaaS solution using logical network segmentation and ONTAP multitenancy techniques.
- **Automation and APIs.** Cloud customers access their data through a portal. The SP developer teams were able to directly automate the NetApp infrastructure through the API layer for automation and management integration.

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 the SP's 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 the service provider 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 Multicloud 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 while maintaining the their security posture. The software vendor's professional services team required the ability to quickly create dev/test environments for quickly onboarding customers. The software vendor's engineering 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 systems 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 needed to satisfy their largest customers could not be satisfied cost-effectively by AWS alone. They needed the ability to choose from 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 ONTAP Cloud, which placed its aggregates on EBS. By leveraging deduplication and other storage efficiency features of ONTAP Cloud, they were able to reduce the EBS storage footprint required by 90%. Using the storage cloning capabilities of ONTAP Cloud, 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 NetApp's data fabric 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 its 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 an open-source CMS.

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

Data Fabric Value

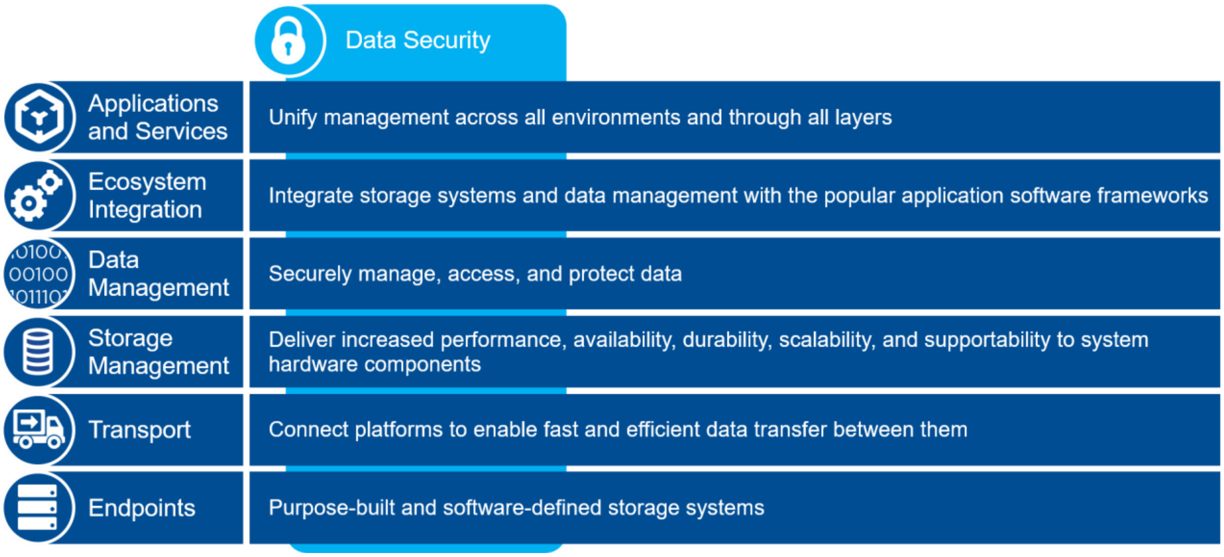
By separating the data from the applications, IT now has full custody of business data on its own storage and is able to comply with NetApp legal policies for data sovereignty, data security, and data lifecycle management. The Data Fabric allows NetApp IT to enforce web portal standards that are easy to use, manage, and support while giving full control of content management to back up 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 across the environment: endpoints, transport, storage management, data management, ecosystem integration, applications, and services, while maintaining an organization’s security posture and framework. The overall architecture is composed 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 from which to choose 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 securing and laying the foundation for NetApp’s data fabric.

Figure 2) Data Fabric layers.



Data Security

Applications and Services	Unify management across all environments and through all layers
Ecosystem Integration	Integrate storage systems and data management with the popular application software frameworks
Data Management	Securely manage, access, and protect data
Storage Management	Deliver increased performance, availability, durability, scalability, and supportability to system hardware components
Transport	Connect platforms to enable fast and efficient data transfer between them
Endpoints	Purpose-built and software-defined storage systems

3.1 Data Security

With the evolution of virtualization, shared infrastructure, and cloud technologies comes an increase in the footprint that organizations must secure. In order to overcome security threats, IT must consider and the data fabric must accommodate governance and data privacy in a way that is appropriate for the service models (that is, IaaS, PaaS, SaaS) being provided. Security capabilities must be considered through every architectural layer of the data fabric.

Understanding Key Security Concepts

Secure Multitenancy (SMT)

Organizations must understand how their service providers segment and isolate each customer in their cloud infrastructure and how that manifests itself across a data fabric. It is also imperative that organizations know how their data is being secured. For example, with secure multitenancy, multiple tenants can share infrastructure. Administrators set the resource limits and QoS levels for each tenant. Tenants have administrative control of their provisioned environment, while remaining isolated from others.

Data in Use, in Motion, and at Rest

Data in use refers to any data being processed by a cloud service provider. It is key to maintaining security throughout the lifecycle because it is crucial to maintaining an organization's security posture. The fact that cloud solutions rely on shared processes and resources introduces the requirement for more diligence. The security criteria that are maintained within an organization must be maintained and possibly enhanced across the data fabric.

Data in motion, also referred to as data in transit or in flight, consists of the state of information or data while being transmitted across the data fabric. When data is in motion, it is more susceptible to being breached or accessed for interception or unauthorized access. Examples of data in transit are users browsing the Internet or accessing a database, migrating data throughout a virtual environment, and accessing third-party or hybrid cloud data from a private environment. Typically the most common solution for data in motion is the use of Secure Sockets Layer (SSL) and Transport Layer Security (TLS), most often seen and referred to as using the HTTPS protocol. In addition, virtual private network (VPN) solutions (typically IPSec based) are utilized.

Data at rest defines the state of information or data while static and not in motion. Examples are data sitting on a database or data backups. Any data that exists in the data fabric must maintain a secure state of integrity and confidentiality when at rest, regardless of where it resides in the fabric. Solutions here typically include the encryption of stored data through self-encrypting disks and other FDE solutions.

Key Management

Key management refers to safeguarding and managing keys in order to manage a secure environment that effectively safeguards an organization's data. Key aspects of the key management solution across a data fabric are maintaining proper access to the keys while securely storing them. In addition, making sure that recovery solutions also exist across the key management solution is paramount in the event that disaster strikes.

The cloud security alliance defines the following key management considerations:

- Random number generation should be constructed as a part of the process
- Cryptographic keys should never be transmitted in the clear, throughout the lifetime of the keys
- Understand that lack of access to the keys results into lack of access to the data
- Key management functions should incorporate separation of duties and segmentation through the data fabric

Aligning Governance, Data Privacy, and Data Sovereignty

Governance, data privacy, and data sovereignty must be aligned. Because data stored in binary or digital form is subject to the conditions of the country in which the data resides, it is imperative that organizations understand the nature of their cloud architecture and interactions. Moreover, it is important to understand the rules, legislation, and regulations that apply to various countries as they apply to personal information, data privacy, and data protection.

Implementing Cloud Security by Service Model

The cloud security alliance model depicts cloud services as operating in three service models:

- Infrastructure as a service (IaaS)
- Platform as a service (PaaS)
- Software as a service (SaaS)

IaaS

Infrastructure as a service allows organizations to provision compute, storage, networks, and additional resources on which IT can deploy and run software: operating systems and applications. The hypervisor of a virtual environment is a key component of an IaaS solution and is the focal point of IaaS security.

Three main areas apply to IaaS security:

- **Hypervisor and virtual machines (VMs).** The number of virtual solutions across the data fabric today increase exposure to VM-based attacks. The impact of these attacks can be exponential because a breach of one VM can affect a number of other VMs in the same hypervisor or physical server. Virtual machine–based rootkits, which directly affect the hypervisor, can propagate attacks to subsequent VMs, applications, and software in the solution. It is critical to protect virtual machines by protecting the hypervisor from traditional attacks and vectors such as DDoS and DNS-based attacks.
- **Virtual network and infrastructure.** Attacks against the virtual network infrastructure are prevalent. These attacks usually target the virtual switch or router that controls the flow of traffic across the fabric. Attacks such as VLAN hopping and manipulation or modification of ARP tables are examples of virtual network vulnerabilities.
- **Management functions.** Colocation, tenancy, and segmentation are the key management functions that pertain to IaaS. Colocation refers to sharing of physical resources. Multiple VMs share compute, storage, memory, and other resources. The sharing of such resources inherently increases the attack surface and therefore increases risk.

  Tenancy or multitenancy makes sure that when different organizations and users share the same applications and hardware in the cloud environment, the information from their respective logical environments is isolated and never shared.

Network segmentation makes sure of logical separation and therefore isolation and reduction of the attack surface while providing key points of visibility.

PaaS

With platform as a service, organizations can deploy applications onto the cloud infrastructure.

Four main areas apply to PaaS security:

- **System and resource isolation.** At all times consumers/tenants in a PaaS solution should be isolated from each other. Administration and monitoring should be segmented and secured in isolation to make sure that configuration or system changes do not affect multiple tenants.
- **User-level permissions.** Apply across the data fabric for every service offered in the PaaS solution. Each service should be isolated and restricted to a level of access or permissions to prevent escalation of privileges and unintended inheritance issues.
- **User access management.** Limits the IT resources a user can access.

  Examples include printers, telephony devices, and data. Single sign-on is often employed across the data fabric to simplify access control.
- **Data loss prevention and protection.** Software that protects against malware, backdoors and trojans, code reviews, education, an established software development lifecycle (SFDC), audits, and controls is essential protections against malicious actions in the data fabric.

SaaS

Software as a service provides applications running on a cloud infrastructure, often on a subscription basis. These applications can be accessed through thin clients or a web browser or integrated with other applications through their APIs.

Three main areas apply to SaaS security:

- **Data segregation** is the clear understanding and segmentation of each user’s or customer’s data regardless of the environment. In today’s cloud environment, it is imperative because segmentation applies at the logical/application level as well as the physical level. Any solutions in the data fabric should make sure that data from various users is segmented.
- **Data access and policies** refers to making sure that data is only accessible by the users/entities who require access to the data and to the extent to which they need that data. Data fabric solutions must make sure that the security postures, including access controls throughout the cloud solutions, are consistent with the organization’s security policies.
- **Web application security** as it applies to SaaS is often referred to as web SecaaS. These security capabilities are essential for managing security for web applications in the cloud:
  - Web URL filtering
  - Web malware scanning
  - Peer-to-peer controls
  - Vulnerability intelligence
  - Central policy control

Web SecaaS can be delivered through multiple deployment solutions throughout the data fabric:

- Proxy methods are the most common methods. These are typically done by browser proxy settings or transparent forwarding by firewalls or other network devices.
- Access gateways typically use an on-premises appliance to forward the web requests to the cloud for web application security services.
- Agents are common with the mobile work force because the mobility aspect increases the dynamics and, in essence, the presence of the web SecaaS solution. In such use cases, the use of agents on endpoints is employed.
- Combination of the web SecaaS solutions is ultimately the best practice approach to sure of complete coverage of web application security services.

3.2 Endpoints Layer

Endpoints

Purpose-built and software-defined storage systems

The endpoints 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 factors: purpose-built or software-defined storage systems. The deployment models for the endpoints define the topology and span of the fabric. In addition, the NetApp solution provides for secure data in motion at the endpoints layer.

Choice of Endpoints

ONTAP software provides the foundation of NetApp's data fabric. 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.

ONTAP is the native data management software of NetApp purpose-built AFF and FAS storage systems. ONTAP software-defined storage offerings enable a fabric to span to commodity direct-attached storage (DAS) with ONTAP Select and to the public cloud with ONTAP Cloud.NetApp ONTAP software brings the same data management capabilities to disparate environments across the hybrid cloud.

Table 1) NetApp system options.

Systems	Purpose-Built Storage Systems	Software-Defined Storage Systems
ONTAP		
FAS	✓	
All Flash FAS (AFF)	✓	
FlexPod®	✓	
ONTAP Cloud		✓
ONTAP Select		✓
SolidFire®	✓	✓
E-Series	✓	
EF-Series all flash	✓	
AltaVault™	✓	✓
StorageGRID® Webscale	✓	✓

ONTAP Cloud implements AES 256-bit software encryption capabilities to secure data at rest on public cloud storage. FAS, SolidFire, and E-Series systems implement hardware-based security with self-encrypting drives.

AltaVault encrypts all data prior to transmitting it over the wire to the target cloud storage. 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.

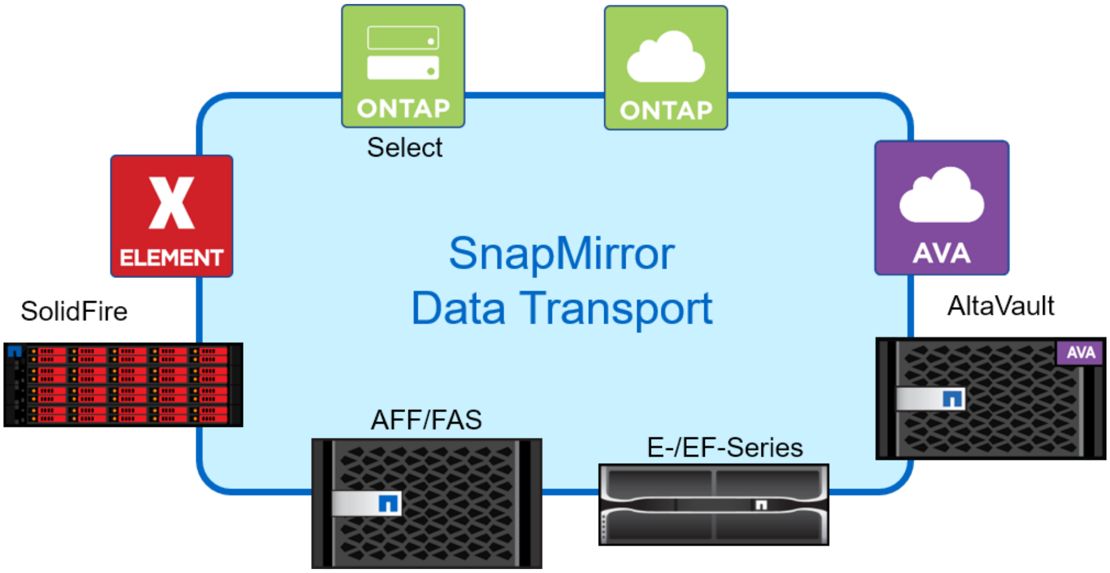
Other storage systems, including third-party arrays, can be added based on workload and business requirements.

3.3 Transport Layer

Transport | Connect platforms to enable fast and efficient data transfer between them

The ability to move data across the data fabric requires a transport mechanism to establish lines of communication between the endpoints. NetApp SnapMirror® is the primary transport for the data fabric. The SnapMirror protocol connects endpoints in the data fabric for seamless transfer of data in bulk. It enables applications to move data nondisruptively where it is needed most: to a cloud, across storage tiers, or across clusters. Applications are unaware of any data movement.

Figure 3) SnapMirror data transport enables data to move seamlessly among endpoints.



The ONTAP family of products (FAS, AFF, ONTAP Cloud, ONTAP Select) share a common WAFL® file system format. All other endpoints in the fabric have their own native data formats.

The SnapMirror transport enables the endpoints to interoperate and move data efficiently between them using SnapMirror replication or SnapVault backup capabilities. Not only does this transport enable interoperability, when used between ONTAP endpoints it also preserves deduplication and compression 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 allows the secure movement of data from one endpoint to another and enables the data to be accessed and consumed in a form native to each endpoint.

Envision the Future: Expanding SnapMirror Data Transport Endpoints

Today, SnapMirror data transport is available for ONTAP and AltaVault endpoints. It is being expanded to include E-/EF-Series and SolidFire.

SnapMirror data transport allows data to be accessed and consumed in a form native to each endpoint. For example, SolidFire serves a given dataset to clients using its iSCSI protocol, but after the data is moved to FAS using SnapMirror, the FAS may serve the dataset to clients using iSCSI or Fibre Channel.

Secure Cloud Connectivity

When endpoints plug into the fabric, they not only get the benefit of data exchange with FAS and each other, which is important for managing service level objectives (SLOs) and total cost of ownership, but they also become cloud enabled. With the SnapMirror transport, data can be moved efficiently to and from these endpoints to the cloud endpoints: ONTAP Cloud or FAS systems deployed as NetApp Private Storage.

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 ONTAP Cloud in multiple AWS regions. SnapMirror replicates data from the centralized data warehouse to the various ONTAP Cloud endpoints. Users access reports locally in their region using a business analytics web portal. When demand for reports is fulfilled, the ONTAP Cloud instances are torn down.

NetApp follows a strict data governance policy, and hosting business data on public cloud storage met requirements due to ONTAP Cloud 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.

Efficient Transport, Faster Data Transfer, and Security

AltaVault itself can back up any storage array. By leveraging SnapVault backup capabilities for data transport, AltaVault data transfers are more efficient and faster.

With ONTAP as a source, incremental NetApp Snapshot® data can be pushed in its native format to AltaVault. 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 SnapMirror 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.

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 users experience higher access latencies when residing on the lower tiers.

NetApp 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, 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 might 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. 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: Object Store Data Tiering

The data tiering concept can be applied to cloud technology, object stores in particular. In the case of AFF, consider an aggregate consisting of a set of SSDs and an object store bucket. As with Flash Pool aggregates, the hot data gravitates to the SSDs, while the object store operates as the inexpensive, deeper, slower capacity tier. (See Figure 4.)

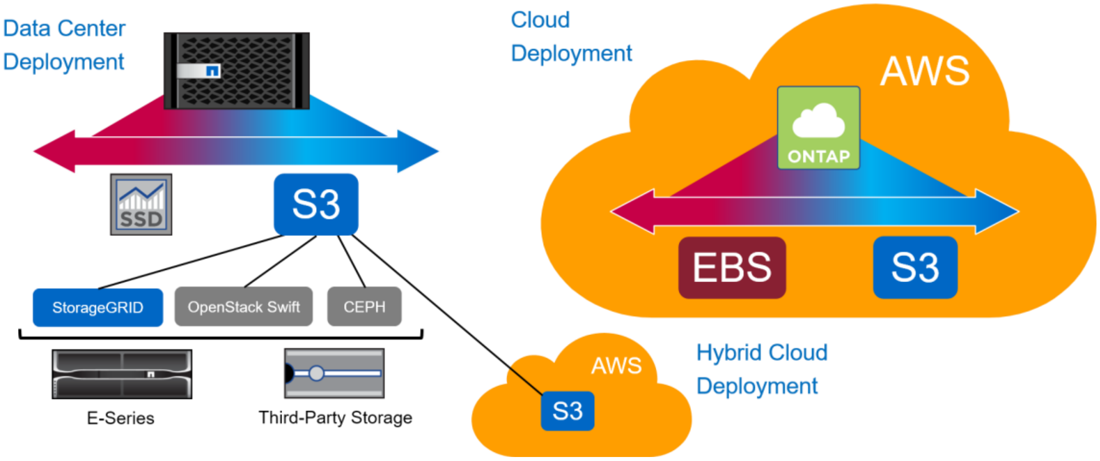
The object store can take many forms in the data center, including StorageGRID, OpenStack Swift, or CEPH. The object service may be backed by various physical storage systems, including E-Series and third-party DAS systems.

A hybrid cloud data tiering deployment will support the AFF in the customer's data center, and AWS S3 object stores in the public cloud.

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

With 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, SLOs, and price points, while unifying data management.

Figure 4) With object store data tiering, the Data Fabric can intelligently and automatically move data to the optimal tier, whether it be SSD or object media.



StorageGRID Webscale

StorageGRID offers secure 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 AFF/FAS object store data tiering, SolidFire backups, and AltaVault backups.

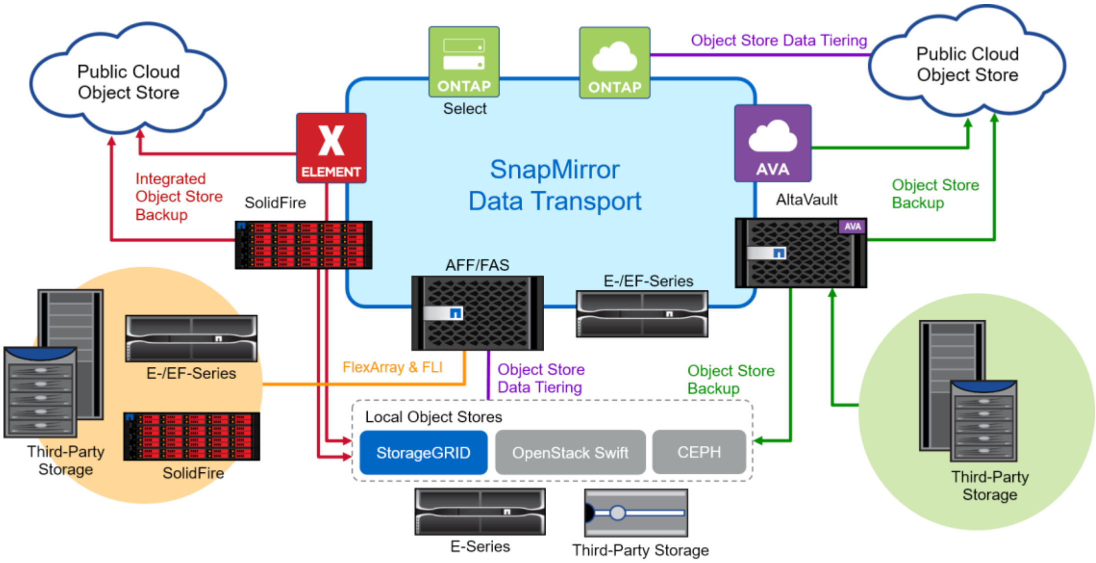
In addition to those use cases, StorageGRID is an object-based fabric in its own right. Cloud applications can store and retrieve objects directly over its S3, CDMI, or Swift protocols. StorageGRID is a true multisite architecture and has advanced object management capabilities, such as geographically-dispersed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

Further Extending the Data Fabric to Third-Party Endpoints

Although NetApp's data fabric is optimized for NetApp products, it is open to third-party products, as shown in Figure 5. Organizations can extend NetApp's data fabric in many ways, including:

- Protecting investments in existing third-party storage arrays
- Bringing data fabric capabilities to commodity, direct-attached storage systems
- Leveraging open-source software-defined storage offerings
- Backing up to public or private cloud object stores
- Integrating with object stores

Figure 5) NetApp's data fabric endpoint integration expands to third-party systems.



Storage Array Investment Protection

Organizations can leverage investments in third-party arrays by bringing them into their data fabric using NetApp FlexArray® software, a capability of ONTAP software. FlexArray enables third-party arrays to receive the benefits of native NetApp systems, including storage efficiency and security, multiprotocol support, SnapMirror 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 ONTAP Cloud. In addition, customers may choose to move data that resides on third-party systems to native NetApp AFF/FAS storage. This action is known as foreign LUN import (FLI), which migrates the data to a NetApp storage pool. FLI allows array migration, cloud migration, or data tiering for better SLO management.

Envision the Future: Expanding ONTAP FlexArray

Today, FlexArray supports NetApp E-/EF-Series systems, as well as systems from EMC and HDS. We are evolving the Data Fabric capabilities to extend FlexArray support to SolidFire systems.

Commodity Direct-Attached Storage Integration

Direct-attached storage (DAS) systems can connect to the data fabric using ONTAP Select (file or block), SolidFire Element X (block), or StorageGRID (object) to manage the underlying physical storage. ONTAP Select and StorageGRID install on the server as a virtual machine, while Element X installs on bare metal. After the software is installed on the server, it provides its full set of data management capabilities to applications and ecosystems.

SolidFire Element X enables service providers to scale their block services as business demands increase. It also enables them to leverage commodity hardware components and offer multiple quality of service (QoS) levels.

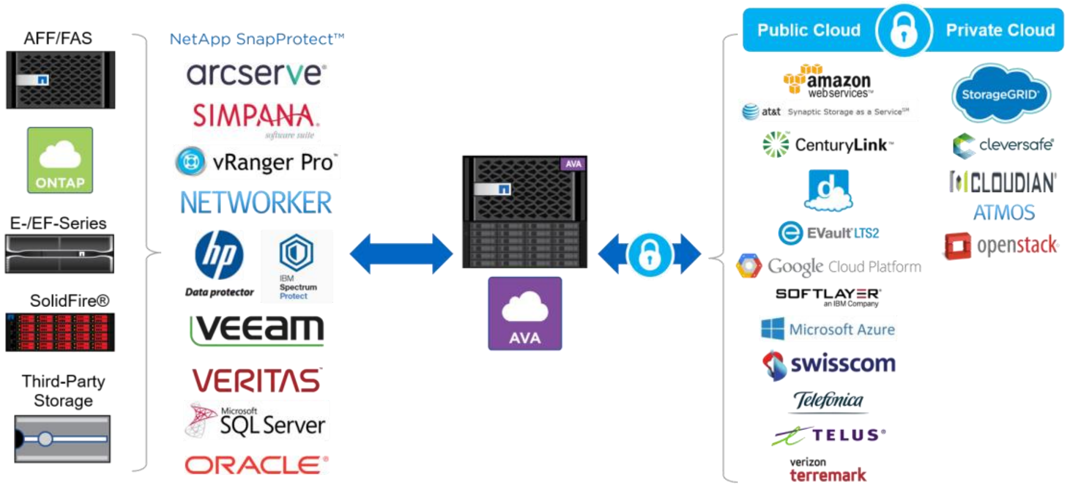
Open-Source Software-Defined Storage Integration

E-Series systems may be used as enterprise DAS, a high-performance, scalable, cost-effective, secure alternative to internal DAS. With E-Series as a physical foundation, open-source SDS offerings such as CEPH and OpenStack Cinder can be layered on top. By adding StorageGRID capabilities, E-Series can also provide an object store for OpenStack Swift. In both of these scenarios, security can be enhanced by leveraging the SafeStore encryption capability of E-Series.

Backing Up to Cloud Object Stores

An AltaVault appliance can securely back up any storage array to any cloud and supports a wide variety of backup software. It’s a modern day replacement for tape solutions that provides the economic and geolocation advantages of cloud implementations while the value of encrypted data in flight to the object store and at rest. After 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 Store Integration

The Data Fabric supports multiple use cases for object stores, including a target for AltaVault backups, direct volume backup/restore, and object store data tiering. For these use cases, ONTAP, SolidFire, or AltaVault manages the data stored in the repository. The implementation and configuration of the repository itself are flexible. The AWS S3 object store is the most commonly deployed public cloud object store. Many NetApp products support other public cloud object stores as well. For on-premises deployments, the software that implements the object protocol can be NetApp StorageGRID or open-source solutions such as OpenStack Swift and CEPH. The back-end physical storage for the repository can be a NetApp storage array or any third-party storage server.

Expanding Endpoints to More Clouds

The multicloud 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.

NetApp's data fabric supports a wide array of virtualized environments and clouds today and will continue to expand to more clouds.

A Case for Multicloud Endpoints

CodeSpaces.com was a startup that offered code repository services in the cloud. It 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 instantly put a multimillion-dollar startup out of business.

Multicloud 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, ONTAP Cloud could have been used with SnapVault to copy the data to a service provider offering NetApp ingest backup as a service.

3.4 Storage Management Layer

Storage Management

Deliver increased performance, availability, durability, scalability, and supportability to system hardware components

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, Helix, 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 applications they serve.

Scale-Out Architecture

A scale-out storage architecture allows IT departments to expand their storage capabilities as their needs change. Traditional architectures require planned downtime in order to complete these system upgrades or expansions.

ONTAP Data Management Software

ONTAP software enables IT departments to scale both vertically and horizontally without service disruption, including:

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

Third-Party Storage Arrays

When plugged into the data fabric using ONTAP FlexArray, third-party storage arrays can also be upgraded or expanded without disruption.

SolidFire

The SolidFire scale-out architecture allows for linear, predictable performance gains as nodes are nondisruptively added to the cluster. Data is automatically redistributed in the background across all nodes in the cluster, maintaining balance as the system grows.

The native QoS capabilities allow for delivery of firm performance SLAs for applications, workloads, and tenants across the entire infrastructure—an important capability for cloud service providers and enterprises building private clouds.

StorageGRID Webscale

StorageGRID Webscale is a massively scalable object storage solution. Its unique software-defined architecture supports billions of objects and tens of petabytes of storage spanning numerous locations in a single namespace. StorageGRID Webscale may be seamlessly expanded in size and capability with nondisruptive node addition and software upgrades.

Quality of Service

Quality of Service (QoS) capabilities enable shared infrastructure to deliver a specific level of performance to tenants and applications without impact from noisy neighbors.

SolidFire QoS permits the specification of an allowed minimum, maximum, and burst level of performance to enable the consistent, predictable, and guaranteed performance needed for each deployed application.

When applications predictably get the resources they require, the need for constant monitoring dissipates, and IT operations are simplified.

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 ONTAP software and available across the data fabric.

Table 2) Storage efficiency technologies inherent in ONTAP software.

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.

3.5 Data Management Layer

Data Management

Deliver a set of capabilities to securely manage and access data

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

- Data management efficiencies
- 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 post-processing 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.

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 different endpoints of 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. When used between ONTAP endpoints, 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

NetApp’s data fabric allows numerous options for managing backup and restore operations. Organizations have the choice of using:

- AltaVault to back up any storage array to any cloud object repository. Beyond its deployment flexibility, AltaVault offers ingest rates of more than 9TB per hour, encryption with customer controlled keys, and local caching to minimize RTO of recent data. For additional cost efficiency, AltaVault supports tiering across multiple object store types (for example, AWS S3 and Glacier).
- SnapVault to back up Snapshot copies from any primary storage endpoint with SnapMirror data transport connectivity. 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, ONTAP Select with SnapVault can be used to easily and efficiently back up data from each remote office to the data center.
- SolidFire systems with integrated backup and restore functionality to manage native backup and restore operations based on Snapshot, compatible with S3 and Swift object stores. This native backup integration eliminates the need for and cost of third-party backup software.
- ONTAP systems supporting Network Data Management Protocol (NDMP), providing an open standard for network-based backup of network-attached storage (NAS). NDMP minimizes coding needed for different applications by providing standard commands for backing up and restoring file servers. For NetApp customers, NDMP increases the speed and efficiency of NAS data protection, because data can bypass backup servers and be written directly to tape storage.

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 system.

Table 4) NetApp data access protocols by system.

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

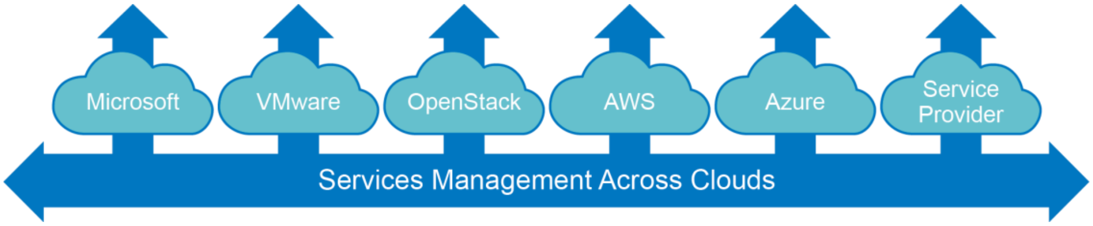
As the various endpoints integrate with the SnapMirror Data Fabric transport, data can move between them and be served by the protocols native to each system.

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 range of application ecosystems and cloud service providers, NetApp has developed a technology stack.

Figure 7) NetApp's data fabric 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

Ecosystem Integration
Integrate storage systems and data management with the popular application software frameworks

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.

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

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

VMware Integration

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 NetApp's 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 NetApp's data fabric
- vRealize Operations (vROps) Management Pack, which extends vROps with metrics and analytical views specific to NetApp for ONTAP storage systems
- Qualification of NPS solutions to run attached to the 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 replicate their data fabric from 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 ONTAP software in a private cloud architecture

OpenStack Integration

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, secure, and more scalable.

NetApp OpenStack block storage (Cinder) integrations include:

- **Storage provisioning and data management.** Specific NetApp drivers for ONTAP, SolidFire, and E-/EF-Series systems enable NetApp storage provisioning and data management capabilities while offering full disk encryption (FDE).
- **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 through FDE.
- **Enhanced persistent instance creation with copy offload.** The NetApp Cinder drivers for 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 Mitaka.

Containers

With the availability and scale of cloud platforms, application architectures are evolving to capitalize on their advantages. This often means the adoption of a service-oriented architecture or a more modern microservices architecture. Both of these encapsulate the functionality of a portion of the application into a component which can be independently developed, deployed, scaled, and even replaced nondisruptively. Such components are known as containers.

Docker abstracts container instantiation, decoupling them from the underlying operating system, helping facilitate the movement to distributed, container-based microservices. This application architecture presents new challenges for accessing persistent data from many different locations simultaneously.

To address these challenges, NetApp provides a Docker volume plug-in that allows the management and attachment of persistent storage devices to containers across multiple hosts. This removes the burden of data management from the application and simplifies microservice components while providing enterprise-class storage performance, efficiency, and flexibility. Today, the plug-in supports NFS and iSCSI protocols for ONTAP systems, with support for additional endpoints coming in the future.

Integrating ONTAP Cloud Endpoints

OnCommand Cloud Manager is the primary portal for establishing ONTAP Cloud endpoints in public clouds and managing cloud resources in the data fabric. As we expand the reach of the data fabric customers will be able to use a single Cloud Manager portal to manage data across their multicloud data fabric in a uniform way. Today Cloud Manager and ONTAP Cloud are available for AWS and Microsoft Azure.

Cloud Manager 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 ONTAP Cloud 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.

4.2 Applications and Services Layer

Applications & Services

Unify management across all environments and through all layers

Data Fabric integrated applications and services from NetApp and its partners deliver high-value security solutions that leverage the foundational capabilities of the underlying fabric.

Applications take the form of IT-deployed and -managed software products, which can be deployed on premises or in IaaS clouds. Services are capabilities that are consumed over the network (typically using a SaaS delivery model) with subscription-based, consumption-based, or pay-per-use payment methods.

The solutions provided by the mix of Data Fabric integrated applications and services include data analytics, data discovery, fabric monitoring and reporting, data protection, data movement among all storage endpoints, copy data management, data governance and compliance, data access management, and others.

This layer helps manage the Data Fabric effectively by providing consistency across endpoints, sites, and environments. As the data fabric evolves, the solutions to manage a data-centric approach will evolve with it.

Analytics of Unstructured Data in Hybrid Cloud Environments

Data centers are generating a lot of unstructured data that contains valuable information that can be mined using data analytics. IT has the choice of managing its own analytics platform (for example, Hadoop) in the data center or public cloud or leveraging a public cloud service such as AWS Elastic Map Reduce (EMR). Data Fabric supports both options today.

Hadoop

For managed Hadoop systems, the NetApp NFS connector lets organizations swap out the Hadoop Distributed Filesystem (HDFS) for NFS or run NFS alongside HDFS. NFS Connector works with MapReduce for compute or processing and supports other Apache projects, including HBase (columnar database) and Spark (processing engine compatible with Hadoop). As a result, NFS Connector can support many types of workloads—batch, in-memory, streaming, and more. The connector is applicable to multiple deployment models, whether it be ONTAP on premises, near the cloud with NPS, or in the cloud with ONTAP Cloud.

AWS Elastic Map Reduce

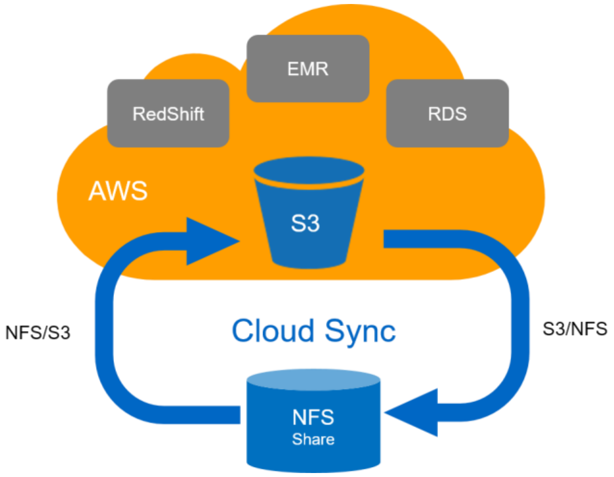
AWS offers an Elastic Map Reduce (EMR) service for organizations that do not want to install and manage their own analytics engine. EMR loads data exclusively from AWS S3 buckets. This limitation presents challenges for hybrid cloud environments. Tools such as the rsync utility and the homegrown scripts used with it to perform these tasks are often inefficient and cumbersome.

How do organizations efficiently and securely transfer unstructured data residing in local (on-premises) file shares to the public cloud, run the analytics, and retrieve the results? How do they deal with file/object format transformation? How do they keep files or objects in sync as they change?

IT needs a simple, automated way to get data into AWS S3 buckets, run the desired cloud service, and get the data back to wherever it is needed.

With the NetApp Data Fabric Cloud Sync service, organizations can extract value from unstructured file data using the native cloud services of AWS.

Figure 8) Extract value from unstructured file data using the native cloud services of AWS with Cloud Sync.



Cloud Sync rapidly moves datasets to and from AWS, converts file-based NFS datasets to the S3 object format, and can launch various cloud services, including EMR or RDS.

Cloud Sync supports any NFS v3 file share (NetApp or third party). It effectively manages a large number of files by rapidly traversing directory structures and transferring them to AWS in parallel. After the baseline is copied, Cloud Sync's continuous sync capability synchronizes the dataset incrementally as changes occur, minimizing the time it takes to transport new data and get results.

After data is transferred, Cloud Sync automatically triggers the AWS cloud analytics service and returns the results back to the original location of the data, whether on premises or in the cloud.

Cloud-Integrated Data Protection for Applications and File Shares

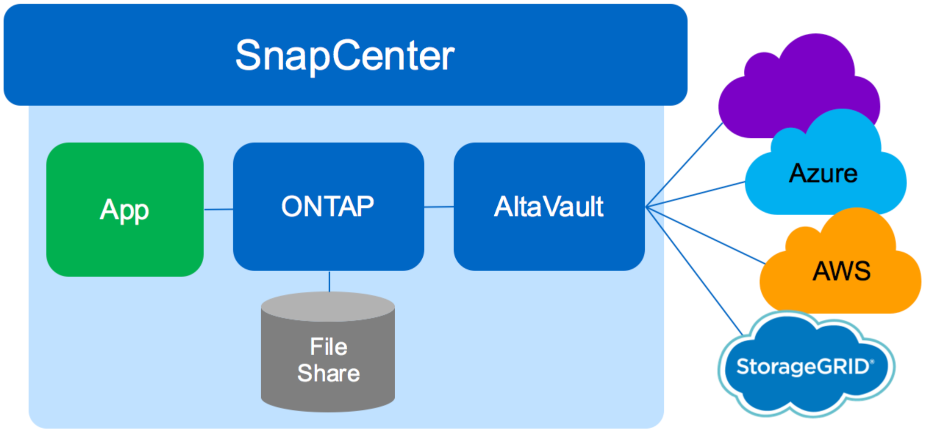
IT organizations today are leveraging object stores to implement their backup and archive strategies. They tap into cloud economics while preserving investments in existing backup infrastructure and meet or improve backup and recovery SLAs. Using cloud storage for data protection can lower the cost per gigabyte for applications and file shares, provide increased performance, and offer web scalability.

SnapCenter® integrates with data fabric endpoints to provide cloud-integrated data protection services for applications and file shares. With SnapCenter, organizations can easily create and manage durable, cost-effective, secure private cloud archives at web scale.

Data Protection for File Shares

SnapCenter enables admins to seamlessly manage Snapshot copy–based data protection operations for Windows and UNIX file shares. Snapshot copy–based backups may be stored on secondary ONTAP systems or in any public/private cloud object store supported by AltaVault.

Figure 9) SnapCenter works across the data fabric to provide cloud-integrated data protection services for applications and file shares.



The ONTAP and AltaVault endpoints are integrated using the SnapMirror data protocol, allowing for fast and efficient data transfer between them. This integration allows endless incremental backups after creation of the first complete baseline.

Application Integration

SnapCenter enables self-service data management by DBAs and application owners. It empowers customers to leverage the data management capabilities of 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, with advanced catalog and index and search capabilities
- Reducing application server data recovery times from hours to minutes leveraging NetApp Snapshot, SnapMirror, and SnapVault technologies
- Providing clone lifecycle management accelerate deployment of new releases and new applications with tight RBAC control for copy: creation, refresh, expiration, and split

In addition to traditional on-premises deployment models, these application integrations support NPS and ONTAP Cloud 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 application solutions include IBM DB2, Microsoft SQL Server, MySQL, and Oracle databases.

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

Envision the Future: Application-consistent backups to cloud

SnapCenter can manage cloud integrated data protection operations for file shares today. In the future, it will manage application integrated backups (application-consistent snapshots) for cloud object stores.

In addition, SnapCenter capabilities will expand to include SolidFire storage endpoints.

SaaS Data Protection Services

SaaS applications relieve IT from having to manage applications and IT infrastructure, but they do not free them from being responsible for protecting the data that these applications contain or produce.

Cloud service agreements often contain clauses that the customer is responsible for securing and protecting its data in the cloud.

Cloud service agreements often contain clauses that the customer is responsible for securing and protecting its data in the cloud

“Although Salesforce does maintain back up data and can recover it, it’s important to regularly back up your data locally so that you have the ability to restore it to avoid relying on Salesforce backups to recover your data.

The recovery process requires [sic] is time and resource consuming and typically involves an additional fee.”

Salesforce Knowledge Article 000004037
Retrieved September 2016

“Account Compromised
If you can't find your messages in All Mail, Spam, or Trash, or by performing a search, then they've been permanently removed from Gmail, possibly deleted by someone else. We regret that we are unable to recover messages that have been permanently deleted. If you're concerned that someone may have gained access to your address, we suggest that you take the following measures to protect your account.”

https://support.google.com/a/answer/112445
Google Apps Administrator Help
Retrieved September, 2016

Cloud providers do not offer guarantees or SLAs regarding data recovery or the corruption of data that is stored in their services. If application/user access or natural disaster results in data corruption or deletion, the cloud provider is not liable.

Therefore, to mitigate risk to business continuity, it is just as important for IT administrators to maintain backup copies of their SaaS resident data as they do with primary data in their data centers.

NetApp’s data fabric data protection services enable backup of Microsoft Office 365 data. This data protection solution secures data by performing automatic daily backups of Exchange, SharePoint, and OneDrive for Business. Administrators can control the archive period and secondary destinations to meet regulatory compliance guidelines.

All content in O365 instances is protected by restoring Exchange and SharePoint components—including individual items—without native O365 rollback burdens. Backups are stored either in the data center or in the public cloud using ONTAP destination endpoints or object stores supported by AltaVault (for example, StorageGRID, OpenStack Swift, public cloud).

Data protection service benefits include:

- Helps retain control of sensitive data as users, folders, and mailboxes are moved to Office 365
- Provides a choice of deployment model, archiving length, and backup window
- Enables business continuity with fault-tolerant data protection
- Simplifies administration

Envision the Future

NetApp's data protection service will include support for additional SaaS offerings such as Google Apps and Salesforce.

Copy Data Management Across the Data Fabric

Capacity requirements for secondary copies are growing much more quickly than requirements for primary data. Secondary copies are used for backup, archive, test/dev, disaster recovery, analytics, and so on. To add to complexity, these copies are managed by separate tools.

NetApp and its partners offer copy data management (CDM) solutions built on the Data Fabric to address these challenges in a hybrid cloud environment. When the secondary copies of data can be accessed by cloud compute, there is almost infinite capacity available for processing highly intermittent workloads.

NetApp technologies provided by the Data Fabric include:

- Array-based Snapshot copies allow virtual copies to be quickly and efficiently created for any use case.
- SnapMirror allows copies of data to be efficiently replicated to different locations and easily accessed where and when they are needed.
- FlexClone allows a disaster recovery copy of data to be made immediately available for test/dev or data analytics.

While NetApp's latest SnapCenter software offers copy data management capabilities, our partners offer storage- and copy-efficient solutions that provide the highest data protection service levels available today.

For example, Commvault IntelliSnap backup and recovery software is fully integrated with NetApp Snapshot technologies for primary/vault/mirror data, dedupe-aware replication, and tiering across storage and cloud repositories.

Veeam builds on NetApp data fabric capabilities through AltaVault and FlexPod integration to enable efficient, flexible back and recovery operations for all applications in vSphere and Hyper-V environments.

Catalogic's copy management platform, ECX, is designed to work across the data fabric to automate the creation and use of copy data: Snapshot copies, vaults, clones, and replication. Catalogic enterprise catalog management allows organizations to maintain a centralized catalog for all copies. ECX allows storage administrators to create automated workflows for a number of use cases, including recovery, disaster recovery, test/dev and DevOps, and big data/analytics.

Optimize Workloads across the Data Fabric

NetApp OnCommand Insight provides a cross-domain view of performance metrics, including application performance, data store performance, virtual machine performance, and storage infrastructure performance. It analyzes tier assignments and enables load balancing of an organization’s entire application portfolio across all of the endpoints of 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 a native NetApp AFF or FAS managed 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 ONTAP Cloud instances across AWS regions, set up SnapMirror sessions between them and ONTAP in the data center, replicate data, and create NFS and/or iSCSI exports to make the data available to applications running in EC2.

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.

NetApp's data fabric 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. Storage platforms (physical and software-defined) establish the data store endpoints in various environments. A SnapMirror data transport links them together. Management tools, APIs, and ecosystem integration make them easy to use collectively. Data-fabric integrated applications and services provide overall visibility and control of business operations.

The multicloud 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 of this can be done while managing data in a secure way regardless of geographical location.

NetApp's data fabric 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 ONTAP data management software and growing the capabilities and endpoints of your data fabric from there. A variety of deployment models are available, ranging from FlexPod converged infrastructure for data centers to ONTAP Cloud 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.

Copyright Information

Trademark Information