EBOOK

Paving the Way to a Cloud-Aware Enterprise

Contents

Downsizing Data Centers As Part of a Cloud-First Strategy

STAN COX
Director, NetApp Customer-1

I started my career managing traditional data centers housed in large, brick and mortar facilities with sprawling footprints and dedicated everything. Then came the game changers: hyperscale computing, flash storage, and software-as-a-service (SaaS). Services on demand and visibility to data became an IT imperative.

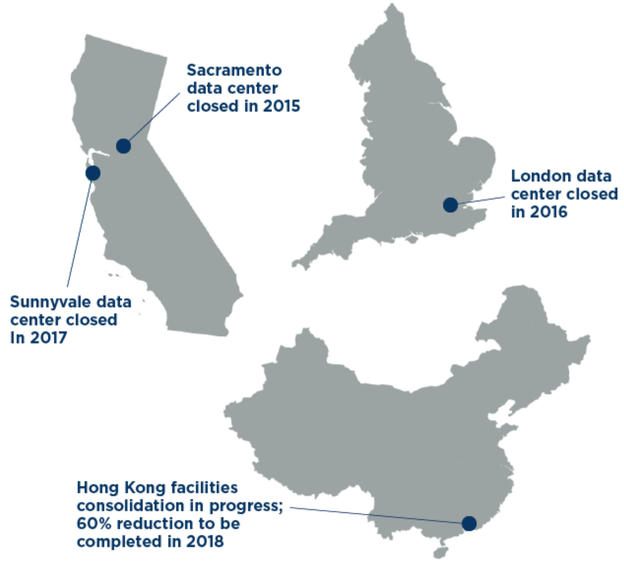
At NetApp, we wanted to take advantage of these game changers as part of our cloud-first strategy. We took a hard look at our data centers and established a strategic roadmap to downsize them where possible. The goal was to lower costs, enhance services, and become more dynamic.

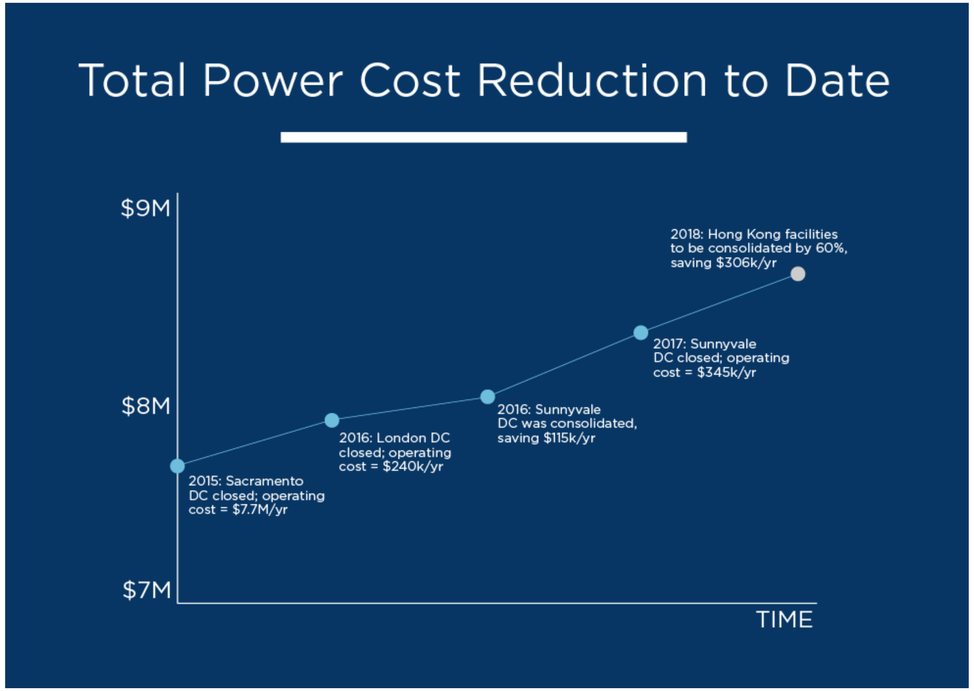
Within a span of four years, our data center consolidation team closed two NetApp data centers in the U.S. (Sacramento and Sunnyvale), another one in the UK (London), and consolidated facilities in Hong Kong by 60% with continued goals to downsize in other facilities.

Sunnyvale Data Center Closure

The latest closure involved the Sunnyvale data center which had served as our tertiary facility for data backups. It was done in less than 18 months (a new record!) under the leadership of an experienced project leader and consolidation team. With the London and Sacramento closures under the leadership of an experienced project leader and consolidation team. With the London and Sacramento closures under their belts, the team easily obtained buy-in from application stakeholders and knew to begin with low-hanging fruit.

First the team identified applications and platforms no longer needed. Some applications were moved to our private cloud, as well as public clouds and SaaS offerings such as Azure to take advantage of services like Microsoft Office 365 and SharePoint. The team consolidated everything else onto denser technology based on the application performance requirements. Non-production lab applications that still required hands-on support were moved to North Carolina; ERP was migrated to our production facility in Oregon; and the remaining noncritical applications were either eliminated or moved to AWS with the data placed on NetApp Private Storage (NPS) residing at a nearby Equinix colocation facility.





NetApp Technology in Action

The team also wanted to jettison hardware no longer needed. At the top of the list was the elimination of tape backups for the archiving of legal, finance, and Sarbanes-Oxley (SOX) compliant data. In its place we chose NetApp AltaVault as the gateway to NetApp StorageGrid Webscale which provides private cloud S3 compliant storage. It works like this: AltaVault provides the local optimized disk cache for applications to store backup data, then the data is transferred to StorageGRID nodes in our North Carolina data center. The cutover off tape backups took about five months, resulted in much faster restores and eliminated a managed tape contract. Today we maintain a small tape library in North Carolina for discovery purposes of archives that still reside on tape. Read more our tape elimination here.

The team used NetApp ONTAP All Flash FAS for the ERP application migration to Oregon as well as the usual contributors for data protection: NetApp Snapshot for local backups and recovery, NetApp SnapMirror for remote disaster recovery and failover, and NetApp SnapVault for short-term archives of critical applications. To manage our storage footprint and gain insight into capacity, volumes, owners, locations and more, the team used NetApp OnCommand Insight and OnCommand Unified Manager.

The closure of our Sunnyvale data center met two crucial goals for NetApp: it supported our cloud-first strategy to implement more cloud and SaaS solutions while aligning to the corporate transformation goal to reduce the real estate footprint in Sunnyvale. It also allowed NetApp IT to maintain budget targets without reducing headcount because of a self-funding approach where savings were directed into open-source licensing and other cloud growth areas.

Key Learnings

Know what's in your data center. We use the Configuration Management Database as our single system of record and have integrated it with

OnCommand Insight. As a result, we have a fully integrated configuration mapping of assets from business services down to the exact storage volume. Knowing what assets are linked to what business services is crucial.

Leverage a program manager and an established data center migration program. It is important to develop a plan before getting started on a data center closure. Designating a program manager and creating a program will keep the process on track.

Expect each closure to get easier with experience. The team's familiarity and knowledge of the data center's infrastructure and environment were crucial to performing a smooth closure. They were able to apply key learnings from previous closures and complete the Sunnyvale closure in record time. ■

In total, the team decommissioned or migrated 264 servers and 59 network devices, eliminating 625 rack units of physical data center space.

4,778
total volumes migrated

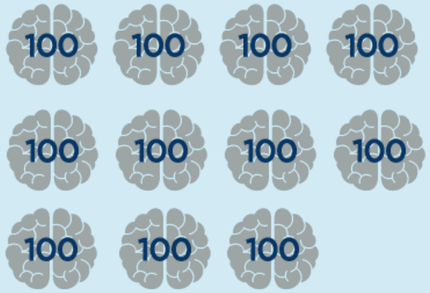
24
filers migrated or decommissioned

2
AltaVault moved to RTP

3
StorageGRIDs remained in Sunnyvale

By closing the Sunnyvale data center, the team moved enough data to store...

...the complete memories of **1,108 human brains,** which can store an estimated 1.25 TB each



...more than 1/20th of data powering **all the Google Maps in the world**

The Business Case for Flash Is Compelling When Looking at Data Center Efficiency

RANDY EGGER
Sr. Manager, NetApp Data Center Services

Technology is evolving faster than anyone anticipated and having a significant impact on our ability to reduce our storage footprint as well as operational costs. We are seeing this evolution play out in our storage lifecycle management strategy, which is undergoing a dramatic upgrade thanks to the benefits of All Flash FAS (AFF).

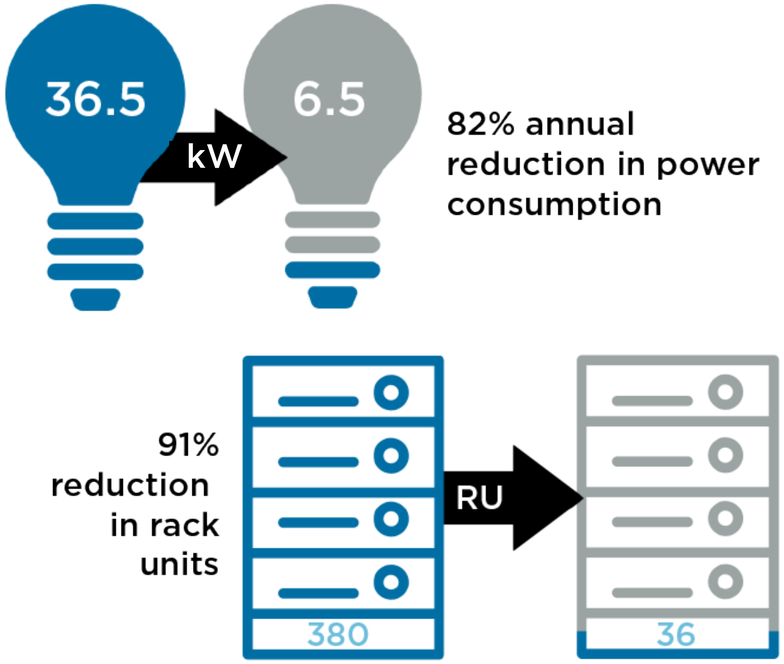
As part of our lifecycle management strategy, we are decommissioning some older FAS nodes running on ONTAP 9 and replacing them with AFF. We are excited to migrate to AFF for a variety for reasons, including improved application performance, higher density, and economies of scale that will help us transition to the future.

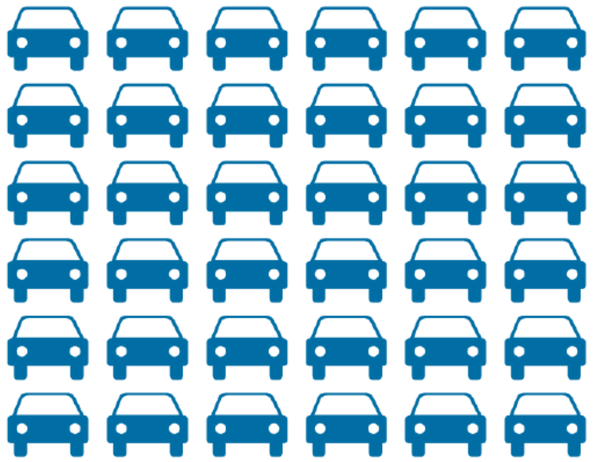
The biggest change we see will be in the provisioning strategy. In the past, we had to overprovision to meet performance requirements. AFF takes that issue off the table. One AFF solid state drive (SSD) equals the I/O performance of multiple shelves of hard disk drive (HDD) or spindle technology. The sizing of the array is much simpler because we can provision based on capacity, not on future performance requirements. The AF700's powerful controllers and incredibly dense drives will dramatically improve the application performance while minimizing the footprint.

Another benefit is the non-disruptive migration made possibly by ONTAP. We anticipate no interruption of service for our end users, something we could not have achieved with older 7-mode technology. To non-disruptively move the applications requires three steps. First, we will add the new A700 nodes to the existing cluster, then we will use ONTAP's Volume-Move feature to non-disruptively migrate all the data from the older nodes to the new AFF nodes. Finally, we will remove the older nodes from the cluster. We will repeat this process to refresh all our existing storage infrastructure without an outage.

A smaller storage footprint reduces data center costs while significantly improving efficiency.

- Power requirements of 36.5kW are being reduced to 6.5kW, an 82% drop. Based on a $0.08 cent/kWh electricity rate and PUE of 1.5, costs will drop from almost $32,000 to just $6,800. Our annual savings would be double or triple if we were in a traditional leased or outsourced facility. We also eliminate the additional costs and





82% or 185 metric tons CO_2 savings—equivalent to taking 36 cars off the road annually

latency issues that come with adding a second cage to house the legacy and/or new hardware during the migration.

- Space requirements are dropping 91%, as we move from 380 rack units (RU) of hardware to 36RU to house the three A700s.
- On the greenhouse gas side, our CO_2 savings are projected to be 185 metric tons, equal to taking 36 cars off the road per year.

Next Steps

We will be upgrading to AFF in phases. Initially, we will be replacing approximately nine racks of FAS equipment with one rack containing three A700 15TB SSDs. Along with the migration, we will introduce 40Gb fiber into our clusters, which should dramatically improve data throughput. The combination of new, faster solid-state drives, faster storage controllers, and faster connections will have an enormous impact on capacity, performance, and efficiency—and the services we delivery to our business customers.

What's Next

The adoption of Flash brings us one step closer to the vision of a data center that fits into a living room, not a football field. AFF plays a major role in achieving this goal by improving performance, simplifying operations, increasing data center efficiency, and streamlining the integration of any future infrastructure changes. A small footprint offers exciting possibilities to reduce IT overhead costs, especially space and power, without sacrificing performance. ■

◀ BACK

NEXT ▶

Automation and Containerized Apps Success with NetApp Trident and RedHat OpenShift & Ansible

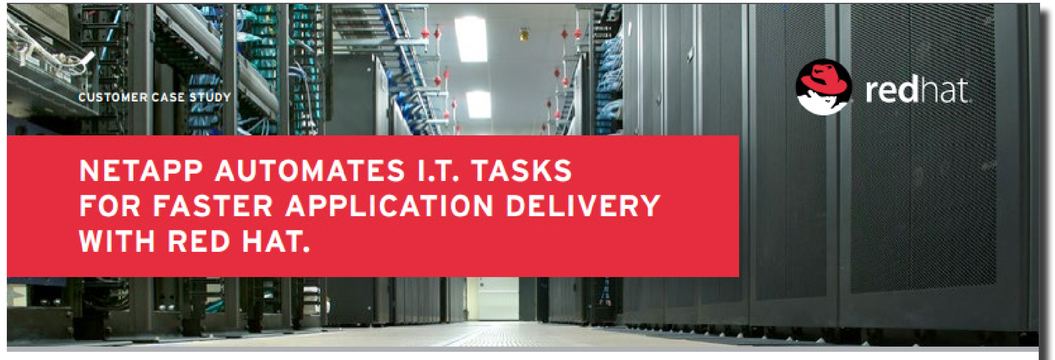
DAVID FOX
Sr. UNIX Systems Engineer, NetApp IT

Over the past few years, NetApp IT has been on a journey to evolve their as-a-service delivery methods and shift to become more business- and service-focused. Automation and containerized applications play a key role. NetApp IT needed a dynamic provisioning solution to support its storage technologies, as well as a fully supported and integrated container platform.

Using NetApp Trident, a fully automated persistent storage provisioner for Red Hat OpenShift, IT can very quickly spin up compute resources for containerized applications while doing the same for storage. These two products together means IT can deliver full-featured stacks in minutes or even seconds, instead of days or weeks.

The team has also used RedHat Ansible Automation to update the network time protocol (NTP) clients on thousands of servers to eliminate hundreds of manhours of manual effort. The team found that Ansible is well-suited to templating the configuration files and prevents human errors that inevitably occur in highly iterative manual tasks.

View the next few pages for a preview of the RedHat case study, or click here to download.



NETAPP AUTOMATES I.T. TASKS FOR FASTER APPLICATION DELIVERY WITH RED HAT.

NetApp
Data Driven

NetApp, a data management technology company, sought to improve the speed and efficiency of its IT service delivery. The company automated manual, repetitive management processes and streamlined storage provisioning using Red Hat Ansible Automation and Red Hat OpenShift Container Platform. As a result, NetApp has dramatically reduced delays and human errors—eliminating hundreds of hours of manual work—and accelerated application delivery times from weeks to minutes.

SOFTWARE AND SERVICES

Red Hat® Ansible® Engine

Red Hat OpenShift® Container Platform

Red Hat Enterprise Linux®

"This project had occupied dozens of staff for weeks, with many more weeks to go. With Ansible, we completed it in two days."

DAVID FOX
SENIOR UNIX ENGINEER, NETAPP

DATA MANAGEMENT TECHNOLOGY

9,000 EMPLOYEES
45 OFFICES WORLDWIDE

BENEFITS

- Achieved more efficient and accurate IT infrastructure maintenance
- Reduced application delivery time from weeks to minutes
- Gained support for evolution to modern, cloud-based application architecture

facebook.com/redhatinc
@redhat
linkedin.com/company/red-hat

redhat.com

"With Red Hat OpenShift, we can very quickly spin up compute resources for containerized applications, and we can do the same for storage with NetApp Trident. Using them together, we can deliver full-featured stacks in minutes or even seconds, instead of days or weeks."

DAVID FOX
SENIOR UNIX ENGINEER, NETAPP

IMPROVING SERVICE EFFICIENCY

NetApp delivers data management solutions that help organizations use information to its full potential. The company's globally distributed corporate IT environment includes four datacenters with 58PB of storage, production footprints in both Amazon Web Services (AWS) and Microsoft Azure, more than 5,300 servers—75% of which are virtualized—and 560 enterprise applications.

To meet a strategic directive to make the enterprise more lean and agile, NetApp's corporate IT team continuously seeks to improve the efficiency of IT services and application delivery—for example, by adopting a DevOps software delivery approach and a hybrid cloud approach to infrastructure services.

However, this team faced significant challenges, such as the high level of time and effort required to manage infrastructure and delays in application delivery due to a lack of dynamic storage provisioning.

Similar to other IT organizations, NetApp was also hindered by repetitive, manual processes. For example, updating the network time protocol (NTP) clients of NetApp's thousands of servers manually took about 5 minutes per machine or 25,000 total minutes. Traditional approaches, such as using scripts to iterate across servers with a stream editor, could not effectively account for high variations in configurations. As a result, this simple maintenance task would require 415 employee hours, or 52 8-hour work days, with high risk of inconsistency. Manual provisioning of persistent storage also prevented NetApp from achieving its goal of delivering software using a container- and microservices-based approach.

To overcome these challenges, NetApp needed a dynamic provisioning solution to support its in-house storage technologies, as well as a fully supported and integrated container platform. "We were looking for a scalable, consistent, and programmatic approach to eliminate manual work and human error," said David Fox, senior UNIX engineer at NetApp.

AUTOMATING INFRASTRUCTURE AND PROCESSES

As an active open source participant and a long-time user of Red Hat software, NetApp decided to pilot Red Hat Ansible Engine, part of Red Hat Ansible Automation, in its NTP update project to eliminate repetitive manual tasks in configuration management. Ansible Engine is an agentless automation platform based on a powerful yet human-readable language that communicates with existing systems, applications, and tools. NetApp's corporate IT team has used Ansible Engine to automate dozens of previously manual processes and continue to streamline its processes.

"Red Hat Ansible Automation was well-suited to templating the configuration files we needed for our NTP update to speed and automate the project. Once we saw what it could do, we started to see automation opportunities everywhere," said Fox.

To address its dynamic storage provisioning challenge, NetApp combined Trident, an open source project it hosts and leads, with Red Hat OpenShift Container Platform, a solution for advanced application creation and delivery automation across private, public, and hybrid infrastructures. Trident lets applications that are provisioned and managed using the Kubernetes container orchestration engine—included in OpenShift—use storage resources on demand.

SPEEDING AND SIMPLIFYING I.T.

ACCELERATED INFRASTRUCTURE MAINTENANCE

NetApp streamlined the iterative configuration file changes required for the NTP project with Ansible Engine, cutting the work from weeks to days.

"This project had occupied dozens of staff for weeks, with many more weeks to go," said Fox. "With Ansible, we completed it in two days: one day to write the new template and playbook, then one day to run it and restart the services. If we need to make standardized changes in our IT environment, we use Ansible to save hundreds of work hours."

Ansible Engine also eliminates the human errors that inevitably occur in highly iterative manual tasks. "If I make an error in my Ansible Playbook, I'll find and correct it when I test my code, then Ansible will run its delivery process, without opportunities for new errors to be introduced," said Fox. "It really turns infrastructure maintenance into something more like a software application. It's infrastructure as code."

REDUCED DELIVERY TIME FROM WEEKS TO MINUTES

Eliminating the bottleneck created by manual processes for persistent storage provisioning has helped NetApp dramatically accelerate application delivery.

"With Red Hat OpenShift, we can very quickly spin up compute resources for containerized applications, and we can do the same for storage with NetApp Trident," said Fox. "Using them together, we can deliver full-featured stacks in minutes or even seconds, instead of days or weeks."

GAINED SUPPORT FOR BUSINESS EVOLUTION

Over the past five years, NetApp's IT team has been on a journey to evolve their as-a-service delivery methods and shift to become more business- and service-focused. Automation and containerized applications play a key role in redirecting the resources reclaimed as a result of these efforts to more proactive, strategic IT initiatives.

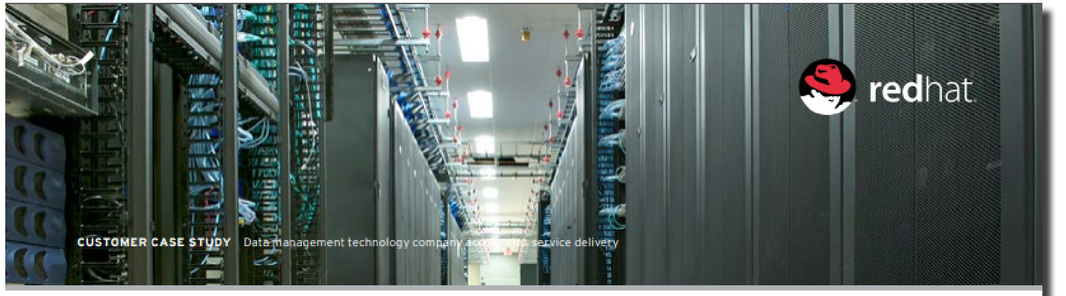
"Red Hat technology is helping us revolutionize our hybrid cloud approach to build cloud capabilities with microservices running containers, through OpenShift with NetApp Trident, as well as API [application programming interface] provisioning," said Fox.

MOVING TOWARDS SELF-SERVICE HYBRID CLOUD

After its success with Ansible and OpenShift, NetApp is continuing to evolve its private cloud environment by using Red Hat OpenStack® Platform to offer self-service provisioning for its application developers and IT operations teams.

"No one at NetApp sat down and said, let's use Ansible, OpenShift, and OpenStack together, but these tools have unique capabilities that help us solve real problems," said Fox. "We have separate initiatives in a larger effort to increase velocity and efficiency, but it just happened that these tools fit our needs very, very well."

NetApp's corporate IT team is also evaluating using Red Hat Enterprise Linux Atomic Host as a foundation for containerized applications, as well as Red Hat CloudForms® for hybrid infrastructure management, to further streamline and speed its application and service delivery.



ABOUT NETAPP

NetApp is the data authority for hybrid cloud. The company empowers customers to simplify and integrate data management across cloud and on-premise environments to accelerate digital transformation. Together with its partners, NetApp provides a full range of hybrid cloud data services to help global organizations unleash the full potential of their data to expand customer touchpoints, foster greater innovation, and optimize their operations.

ABOUT RED HAT

Red Hat is the world's leading provider of open source software solutions, using a community-powered approach to provide reliable and high-performing cloud, Linux, middleware, storage, and virtualization technologies. Red Hat also offers award-winning support, training, and consulting services. As a connective hub in a global network of enterprises, partners, and open source communities, Red Hat helps create relevant, innovative technologies that liberate resources for growth and prepare customers for the future of IT.

facebook.com/redhatinc
@redhat
linkedin.com/company/red-hat

NORTH AMERICA
1 888 REDHAT1

EUROPE, MIDDLE EAST, AND AFRICA
00800 7334 2835
europe@redhat.com

ASIA PACIFIC
+65 6490 4200
apac@redhat.com

LATIN AMERICA
+54 11 4329 7300
info-latam@redhat.com

redhat.com
#f12421_0618

NEXT ▶

NetApp IT Insider Tips on ONTAP Management Using Ansible

EZRA TINGLER
Sr. Storage Engineer, NetApp IT

Most, if not all IT shops, are continually looking for opportunities to reduce costs, deliver efficiency, and be prudent with their time. Automation with custom scripts is a common approach. Yet maintaining these scripts is cumbersome and time consuming.

As data management experts, NetApp leans on tools like RedHat Ansible to automate software provisioning, configuration management, and application deployment. NetApp IT relies on some of these NetApp valuable Ansible modules to automate their storage configuration tasks. For example, we have used the Ansible ONTAP modules to reduce Day 0 build times from a 2-day manual process to an automated, 10-minute process.

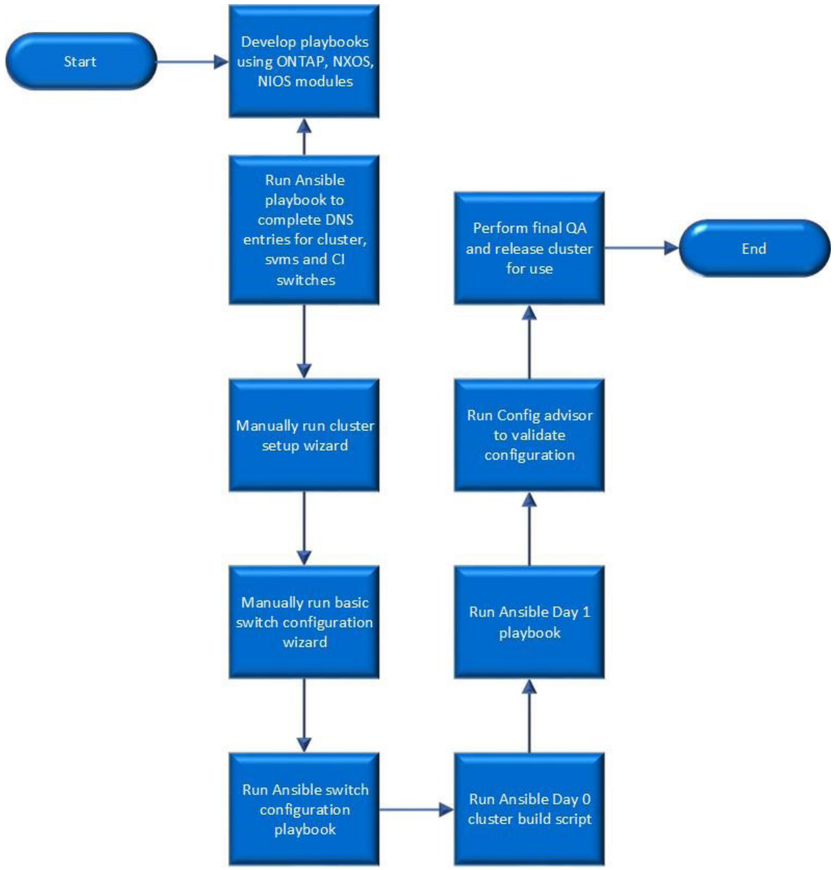
Using Ansible modules for ONTAP, NXOS (Cisco) and NIOS (Infoblox) we have successfully completed a full Day 0 through Day 1 deployment for the following tasks.

Day 0	Day 1	Switch	Infoblox

• Rename the cluster nodes
• Rename the any existing interfaces to match the new node names
  • Cluster interface and node management interfaces
• Rename the root aggregates to match the new node names
• Install licenses
• Configure the service processors
• Set the flow control to none on all 10g ports
• Create broadcast domains and assigns the proper ports
• Create interface groups and add the correct ports
• Create VLANs
• Create failover groups and assign the correct ports
• Create backup interfaces (intercluster)
• Create user roles
• Create users
• Set the raid scrub schedule
• Create aggregates
• Disable aggregate snapshots on all aggregates
• Create cluster peers
• Configures NTP
• Configures SNMO
• Configures CDP
• Configures web services
• Configures autosupport
	• Build a data SVM
	• Create volumes
	• Create Export policies
	• Create LUNs
		• Configure SNMP
		• Configure Traps
		• Install Current RCF file
			• Assign IP addresses and create DNS entries
			  • Cluster
			  • Management
			  • Switches
			  • SVMs

◀ BACK NEXT ▶

For those of you that like to visualize the process, here's how it works:



Here is a sample portion of a playbook used to rename the root aggregates.

```
- hosts: localhost
  gather_facts: no
  connection: local
  tasks:
    # Rename the root aggregates to match the new node names.
    - name: rename aggregates
      na_ontap_aggregate:
        state: present
        service_state: online
        from_name: "{{ item.name }}"
        name: "{{ item.rename }}"
        username: "{{ netapp_username }}"
        password: "{{ netapp_password }}"
        hostname: "{{ netapp_hostname }}"
      with_list:
        - { name: 'durlabdevclu01n01a_aggr0', rename: 'durlabdevclu01n01a_aggr_vol0' }
        - { name: 'durlabdevclu01n01b_aggr0', rename: 'durlabdevclu01n01b_aggr_vol0' }
```

Now that you can see the process and a sample of the playbook, it's time to make it happen. Get out there and start writing playbooks of your own. Automate some tasks and save time for other more worthwhile ventures.

There are NetApp modules for ONTAP, Element OS and E-Series which are fully supported and documented. Module documentation can be found here, and there are always good people in thePub if you have any questions. Good luck! ■

Introducing Platform-as-a-Service (PaaS) in NetApp IT

EDUARDO RIVERA
IT Sr. Manager, Storage Services

The introduction of PaaS to the NetApp portfolio offers exciting possibilities for NetApp IT storage. Integrating this feature set into our enterprise IT strategy is one of our top priorities. A cloud-based, automated, self-service infrastructure will enable IT to rapidly deliver platforms to help shorten development cycles and adapt to changing environments.

The Challenge

Speed is the currency of application development today. IT is leveraging technology to shorten development cycles so the business can respond as quickly as possible to the changing business environment. Enter platform as a service (PaaS).

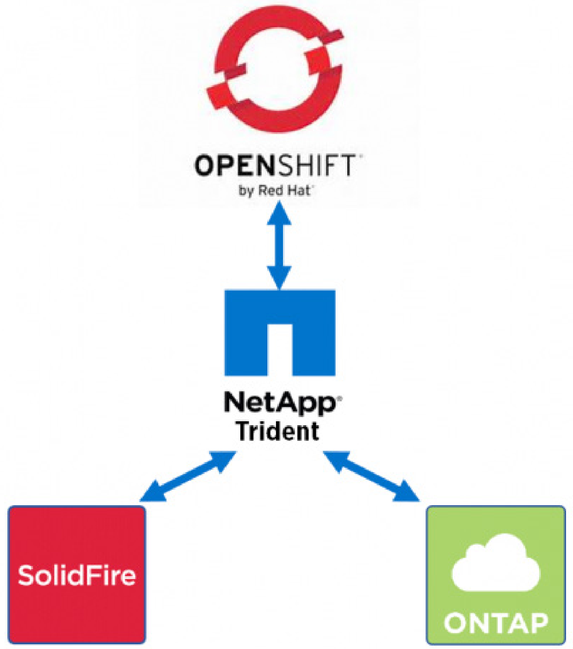
Our objective was to develop push-button deployment of an entire application development environment, such as GitLab or Jenkins, that includes the application code, database, components, network interfaces, and a storage environment. A developer could visit a self-service portal and select a platform from a menu. By expanding the number of testing environments, we can help shorten the development lifecycle. In turn, the business applications team could roll out new features faster.

The Solution

The PaaS environment was architected using OpenShift as the orchestrator of the technology stack. OpenShift automatically provisions the requested app stack with everything needed in containers. This architecture enables the data to persist when the container stops.

To translate the commands from OpenShift to the storage cluster, NetApp developed Trident, an open-source, Kubernetes storage orchestrator available on GitHub. Trident can simultaneously talk to and provision ONTAP® or SolidFire storage based on the application requirements. In our case, Trident would leverage ONTAP for NAS and SolidFire for block storage. This entire process would be transparent to the user.

In this use case, Trident is the owner and administrator of storage. OpenShift creates a ticket for a storage volume, prompting Trident to talk to the ONTAP or SolidFire cluster, then automatically configure, export, and mount the requested volume. In a traditional environment, this manual process would take days or weeks working with various teams to setup



◀ BACK

NEXT ▶

network configuration, create the necessary virtual machines, provision the storage resources, and install and configure the application stack. With human interaction engineered out of the process, it takes seconds.

Application stacks with multiple components, big data stacks, and static content that needs to be shared on a common file system, like GitLab, can use NAS storage. Block is suitable for application stacks like MongoDB, Hadoop, and Elasticsearch, where the components can share data among themselves as part of an application layer cluster and don't require a shared file system.

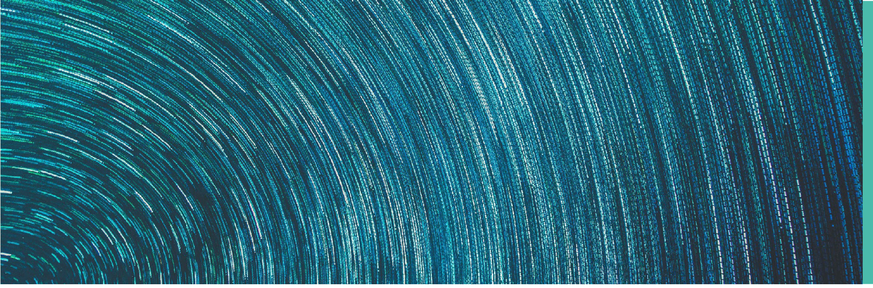
Storage in PaaS delivery

On a strategic level, the PaaS architecture is a shift in the way NetApp IT provisions storage that goes hand in hand with the software-driven, cloud-aware apps of the future. It fits in our efforts to modernize the data center by simplifying the infrastructure layer and managing it as part of an advanced application feature set or architecture.

This architecture takes advantage of the features of the underlying array to provide enterprise-grade storage. This includes leveraging the cloud and orchestration tools to integrate the storage into the service delivery process. By automating platform deployment, NetApp's developers should benefit from faster development iterations—from days to hours to minutes.

The PaaS deployment is changing roles within IT. As storage management is pushed higher into the app stack, IT business owners and engineers will need to become fluent in managing the entire app stack, including storage, business continuity, and disaster recovery. The role of the storage team is also shifting from managing hardware to software. This trend is only the beginning of the departure from the traditional ways of doing things in IT.

Both ONTAP and SolidFire will play a significant role in the modern data center as we expand our service delivery options using the cloud. They are just a first step in our evolution to a next-generation, cloud-based application architecture. ■



Data is Changing and the Demand for StorageGRID is Insatiable

KEN LEE
Sr. Storage Engineer, NetApp IT

In terms of modern-day IT, the amount and type of data received is different from the past. Historically, data has been structured, transaction-oriented, and constantly updated. At the same time, the way applications interact with data has been stateful, atomic (database) transactions. The data growth that many companies are seeing today, however, is a different type of data: unstructured data that sprawls, it's kept forever, seldomly updated, and is meant to be accessed by multiple application services across geographical boundaries.

Increased Appetite for Object Storage

While growth is happening with all kinds of data, NetApp IT has witnessed an increased appetite for object storage. Our consumption of StorageGRID storage has grown 400% over the past 18 months and it's on track to become one of the primary storage solutions within our organization. While StorageGRID is not a replacement for traditional storage protocols such as NFS/iSCSI/FC, it is a great complement to these technologies and plays an important role as it can be replicated across different data centers and offers simple web services interfaces.

Why I Like StorageGRID

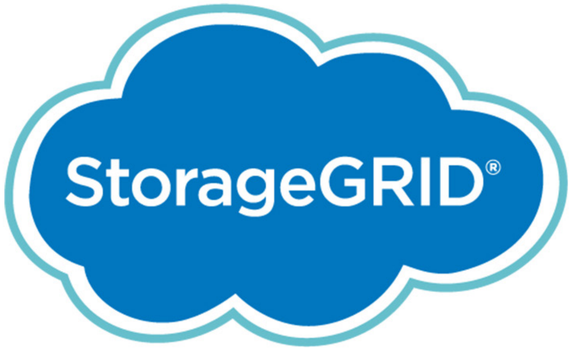
I am a big fan of object storage. I recently worked on a project to streamline the way customers upload core dump files to NetApp technical support for troubleshooting. Since the core files contain images of what was in memory at the time of failure, the size of the files has grown exponentially over time. It is not uncommon to have a 200GB+ core file.

When combining the growing size of core files with the growth of NetApp systems in the field, it results in a large unstructured data pool. This made the new upload process for core files an ideal candidate for StorageGRID.

- Eliminates multiple methods of uploading files via a single StorageGRID platform
- Utilizes StorageGRID multipart upload API capability for performance and the restartability of large files. If a transmission of any part fails, we retransmit that part without affecting other parts and without involving the customer
- Uploads of large objects through a URL (i.e. http/https browser with no other plug-ins or programs required)

400%

growth in consumed StorageGRID capacity in the past 18 months



- Handles large files of static (unchanging) data with minimal data management via StorageGRID ILM rules (i.e. data retention policies)
- Uses metadata to automatically cleanup any orphans inadvertently lost in the transmission process

Other Use Cases for StorageGrid

Another important reason for the rise of object storage within NetApp is the rise of cloud-aware applications. Cloud-aware applications tend to require large data repositories of unstructured data accessed via stateless protocols (e.g. https). As our application teams develop cloud-aware applications they are coding these applications to natively use object storage by default. StorageGRID provides the platform with which cloud-aware apps can interact more naturally, becoming an important player in NetApp IT's journey to the cloud.

Data archiving is another driver for object storage. Today we use StorageGRID as the repository for artifacts such as Docker image repositories, application backup data, and archiving of data analysis results. StorageGRID's information lifecycle management policies allow administrators to provide different retention policies for different types of data; it also dictates how this data is replicated across a StorageGRID global deployment.

It seems like every day we discover new use cases for object storage inside NetApp IT. As we move into the future, the focus on developing cloud applications and services will continue to demand stateless and unstructured data repositories. For these reasons, StorageGRID is an attractive solution. ■

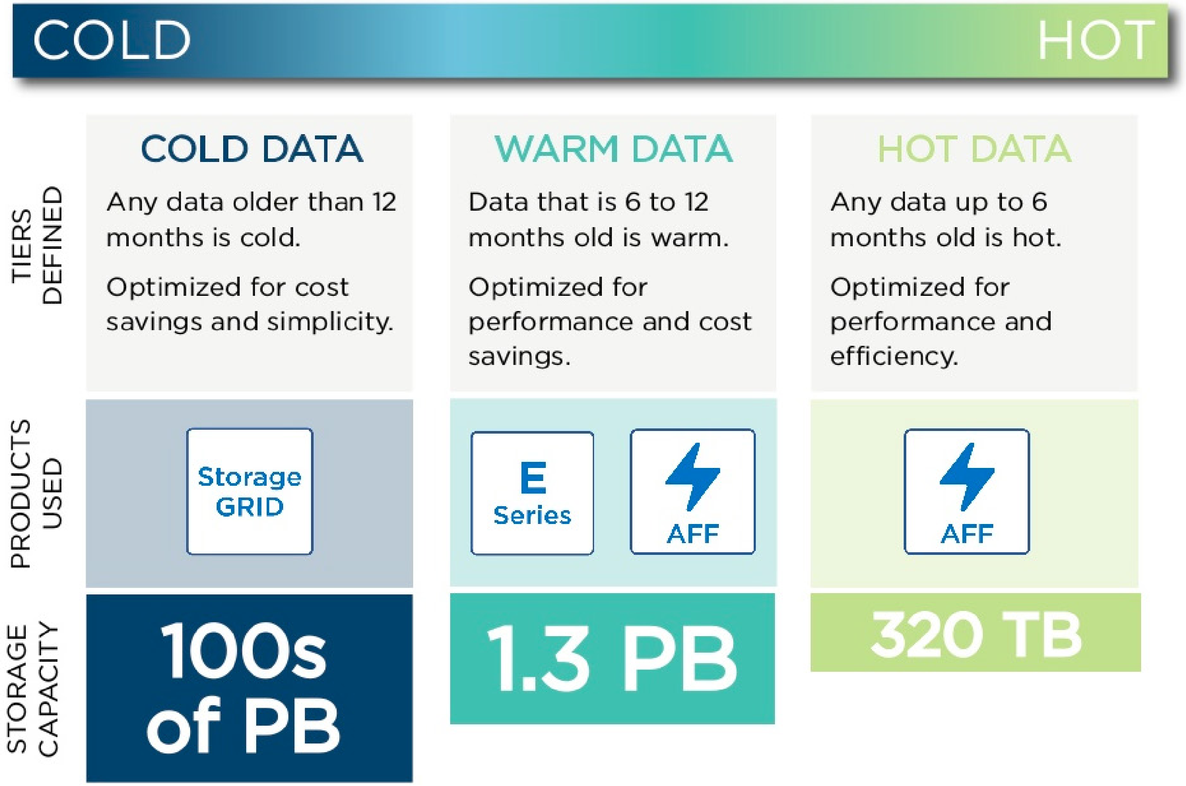
Improving Efficiency with a Tiered Archive Approach

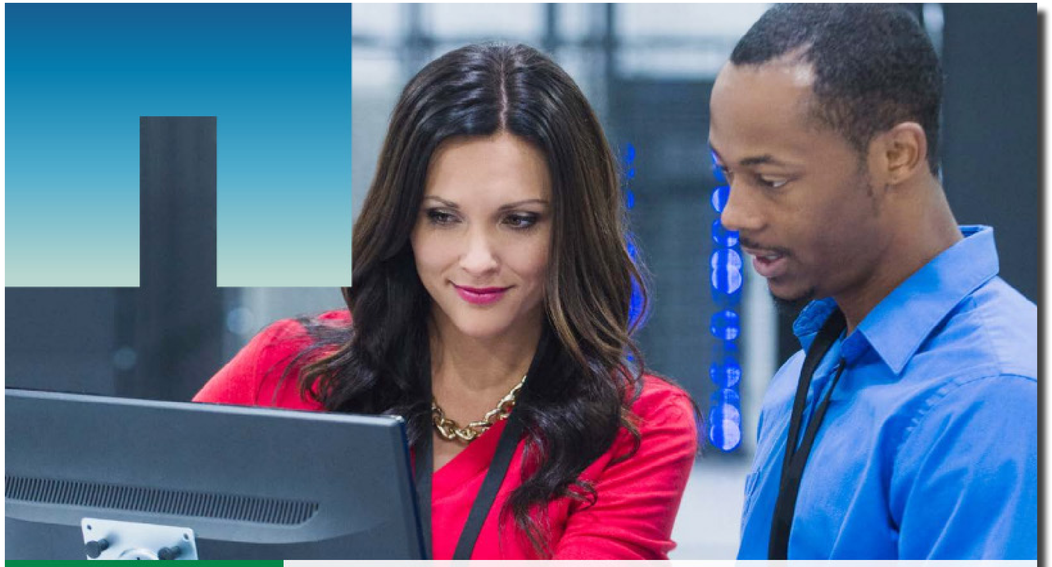
PRAVEEN BEEDANAGARI
Big Data Platform Architect, NetApp

In 2012, NetApp began developing an analysis solution using Apache Hadoop on hybrid flash storage, and the NetApp Hadoop Infrastructure team has seen other companies take a similar approach since then. However, NetApp ActiveIQ, which uses big data and cloud to deploy machine learning algorithms trained to provide information about a customer's install base of NetApp products, helped them to discover that the rising rate of data growth would make the one-tier solution too expensive.

NetApp expanded its one-tier system by adding an archive tier for "cold" data and a more responsive tier for "hot data," utilizing StorageGRID Object Storage, NetApp AFF Storage, and NetApp E-Series Hybrid Flash Storage to realize a 300% improvement in operational efficiency, a 66% reduction in storage space for the most active data and a 40% reduction in licensing fees for NetApp IoT data.

Visit our customer story page to learn about NetApp IT's approach and results, and to download the case study.





SUCCESS STORY
Service Provider

NETAPP IT | PROBLEM SOLVED

Using a three-tier system, NetApp saves on costs while improving the efficiency of its archival solution.

Improving Efficiency with a Tiered Archive Approach

Solutions that tap into the Internet of Things (IoT) often generate a rush of data from thousands of remote devices. To harness the true power of IoT, companies must access the knowledge hidden within that mass of data. However, dynamic growth in IoT technology makes it difficult to find a flexible, efficient, and cost-effective archival solution that can deliver results now and in the future.

300%
improvement
in operational efficiency

66%
reduction
in storage space for the most active data

NETAPP.COM/CONTACT

NetApp

> "One of the great things about NetApp StorageGRID is you can distribute it to different data centers, and it will look like one single pool of storage."
>
> Praveen Beedanagari
> Big Data Infrastructure Architect, NetApp

That's one of the goals of the NetApp Hadoop Infrastructure team, which includes big data infrastructure architect Praveen Beedanagari. The team is constantly challenging NetApp® systems to be more efficient and cost effective. They recently helped to engineer a solution that saves 66% of storage space and 40% of licensing fees for NetApp IoT data.

"What's exciting is that almost any company with an IoT solution could get the same benefits our team is getting," Beedanagari says about the new NetApp three-tier backup solution.

THE HIDDEN COST OF IOT

When NetApp started tracking hardware assets with IoT solutions, developers knew that they were opening the flood gates on a constant rush of data flowing in from hundreds of thousands of devices. But with the rising volume of data, the demand for storage and analysis is reaching new heights every year.

"We have an installed base of close to 400,000, and each one is sending a lot of data—we get close to 7 billion data points a day," Beedanagari explains.

But the true power of IoT doesn't come from capturing remote data—it comes from understanding the data. That analysis is where IoT can generate a hidden cost. NetApp needed a more powerful way to both store and analyze its IoT data to make sure that the data is always available.

THE INITIAL APPROACH

In 2012, NetApp began developing an analysis solution using Apache Hadoop on hybrid flash storage. The NetApp Hadoop Infrastructure team has seen other companies take a similar approach since then, but they found that the rising rate of data growth would make the one-tier solution too expensive.

"In NetApp Active IQ, we double our data roughly every 16 months. We are at the point of 3PB with archiving, just in the last 4 years. The team connected to one storage unit with four server nodes. So, if we needed storage, our only option was to grow that block—even if we didn't need compute resources," Beedanagari says. "As we looked at licensing costs for cloud data management, we knew that something had to change. Any IoT company out there will have the same issue—they'll be spending a lot of money for archivable data."

To manage the swelling waves of data, NetApp sought a more flexible solution.

THE THREE-TIERED SOLUTION

NetApp expanded its one-tier system by adding an archive tier for "cold" data and a more responsive tier for "hot" data. The solution's scripts automatically migrate older data, keeping the most current data easily accessible. Any data up to six months old is considered "hot," and is managed on NetApp All-Flash FAS, which can store 320TB and is optimized for performance and efficiency. Warm data (6 to 12 months) is housed on NetApp E-Series hybrid flash storage and NetApp AFF storage devices, which form a data lake that can store 1.3PB, and are optimized for performance and cost savings. Anything older than this, "cold" data, is moved to a NetApp StorageGRID® solution, which can archive hundreds of

NetApp

NEXT ▶

petabytes across multiple sites, and is optimized for cost savings and simplicity.

Many companies have defined their own three-tiered solutions, but NetApp wanted to do more than archive its data—it wanted to tailor each tier to optimize overall efficiency and cost.

A 66% SPACE SAVINGS

According to Beedanagari, "When most companies build a multitier solution, they build it on servers—which is pretty wasteful, and the licensing is expensive, so you have a huge footprint and a lot of licensing."

The first tier of the NetApp solution offers all-flash performance with deduplication and compression technology that reduces the overall space demands by 66%—even on this tier of "hot" data.

A 40% REDUCTION IN LICENSING COSTS, WITH ADDED EFFICIENCY

"When you are tying up storage on a licensed server, you're tying up the license," Beedanagari says. "Our team started copying anything older than one year from the data lake into NetApp StorageGRID, and that's reducing our licensing costs by at least 40%."

At the same time, the data is still accessible. "One of the great things about NetApp StorageGRID is you can distribute it to different data centers, and it will look like one single pool of storage," Beedanagari says. The solution uses NetApp StorageGRID in two data centers that stay in sync, so data remains available even if one center is offline. Users—and Hadoop—see and write to only one target, so performance and licensing are not affected by the second location.

The NetApp StorageGRID solution also gives operation teams less to worry about. According to Beedanagari, "With NetApp StorageGRID, I'm reducing every 100 nodes to 33 nodes. So now, our team can focus on only 33 of those 100 original nodes."

These benefits grow as the solution is implemented across multiple environments. With StorageGRID, a company's development, test, and quality assurance environments can all read from a single source of data. This keeps all systems current and saves the storage, licensing, and time required to update duplicate datasets on multiple environments.

"And if 10 different clusters need to be built with the data, you are not touching your production data or production compute resources 10 times—you are only touching the StorageGRID, which is completely separate from your compute resources," Beedanagari explains.

> "We looked into options for object storage, and NetApp StorageGRID fit right in with the S3 protocol support. S3 makes it possible to use an archiving or any other solution integrated with the Hadoop or big data technologies."
>
> Praveen Beedanagari
> Big Data Infrastructure Architect, NetApp

NetApp

AN OPEN ALTERNATIVE THAT CAN GROW AND EVOLVE

Ongoing NetApp development indicates that this solution can become even more streamlined and powerful in the future. Because NetApp StorageGRID supports the S3 protocol, it is free to integrate other solutions without being blocked by proprietary protocols. "We looked into options for object storage, and NetApp StorageGRID fit right in with the S3 protocol support. S3 makes it possible to use an archiving or any other solution integrated with the Hadoop or big data technologies," Beedanagari says.

Apache Hadoop offers a powerful and versatile open-source platform for big data analysis, and many companies use it to develop in-house solutions. But if data storage is not optimized, home-grown solutions can quickly fill up storage space and generate expensive licensing fees.

Beedanagari explains, "This solution just gives us more flexibility. And we can automate with NetApp offerings that are now optimized for Hadoop workflows. With the testing and development our team has done, we've also improved performance and optimized processes. At the same time, we're maintaining open-source accessibility to grow into the future."

Companies need to know how they can tap into their data—and their solutions—both now and in the future.

"As people look to the future, and want to hold onto their data, object storage and NetApp StorageGRID is the right place for it," says Beedanagari. "You never know the value the data might have tomorrow that you don't know about today."

SOLUTION COMPONENTS

NETAPP PRODUCTS

- StorageGRID Object Storage
- NetApp AFF Storage
- NetApp E-Series Hybrid Flash Storage

LEARN MORE

netapp.com/us/products/data-management-software/object-storage-grid-sds.aspx

NETAPP.COM/CONTACT +1 877 263 8277

NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of applications and data across cloud and on-premises environments to accelerate digital transformation. Together with our partners, we empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize their operations. For more information, visit www.netapp.com. #DataDriven

NEXT ▶

NetApp IT Expands Business Intelligence with a Modern Analytic Platform and Azure NetApp Files

APARNA BADRINATH
IT Enterprise Architect, NetApp

Business intelligence (BI) is undeniably crucial. Data is king, and business leaders want to quickly and securely tap into it to gain competitive insight, go to market faster, and make the best possible tactical and strategic decisions. The key to success is real time data analytics.

At NetApp, real time data analytics capability is limited because of our traditional data warehouse (DWH) approach using ETL (Extract Transform Load) processes that pull data out of one database, places it into another, and perhaps another, before reports can be generated. It can take days to weeks to ingest data from the multiple sources to report historical operational results.

Modern Data Analytics

The traditional DWH and ETL approach is not responsive enough for our newly formed Cloud Data Services (CDS) business unit at NetApp. This team delivers NetApp's public and multi-cloud solutions that enable customers to access modern data management applications and services like real-time data analytics. They needed a modern analytics platform to conduct real-time aggregation of data sources with unstructured data.

NetApp IT partnered with the CDS team to implement a new, self-service BI platform using the Microsoft Azure cloud services for compute capacity and Azure NetApp Files for storage (part of the NetApp Cloud Volume Service portfolio). The new analytics platform uses Direct Data Mapping architecture and will provide the team the ability to:

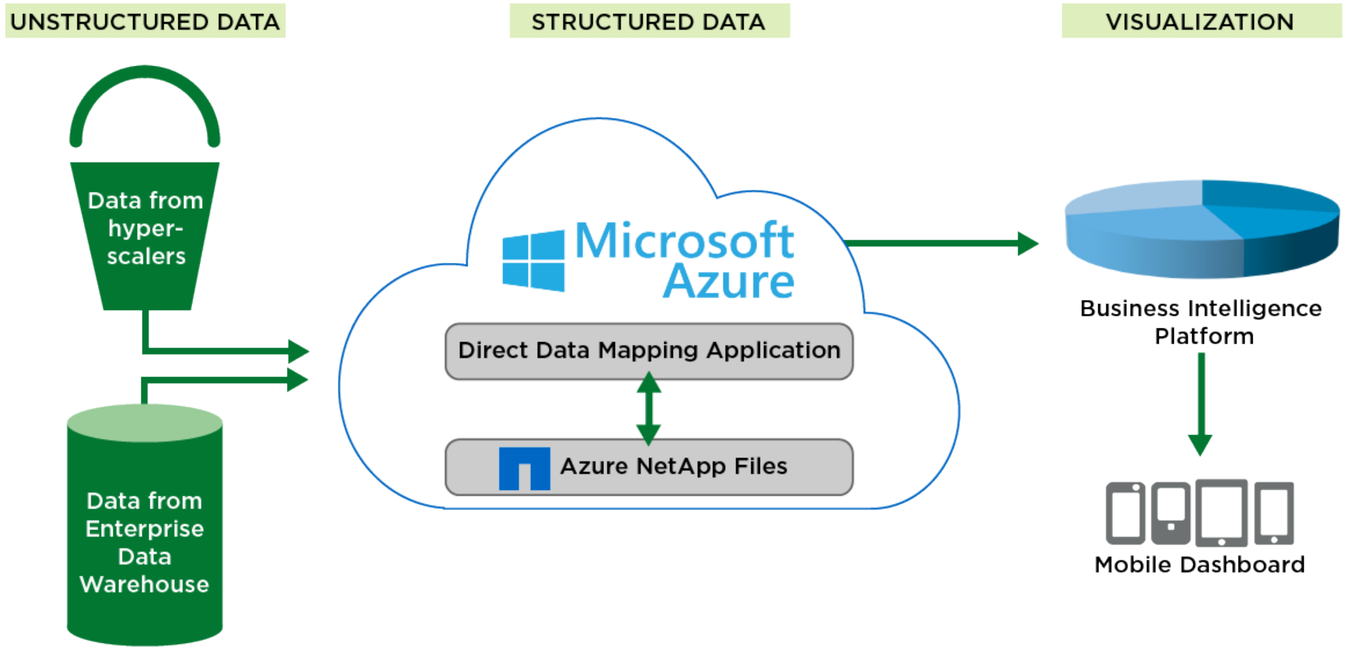
- Make decisions with up-to-date information from multiple databases
- Ingest data, in real time, from cloud-based sources
- Analyze recent (and historical) data—when and how they need it
- Drag and drop report creation
- View a mobile dashboard for manaement level reporting

NetApp Cloud Volumes

While the new analytic platform could have been deployed on premises, we chose to deploy at MS Azure to take advantage of their elastic, pay-as-you-go compute capacity to quickly grow and shrink services as needed. Using cloud-based services allows the team to better leverage machine learning and predictive analytics to drive business growth.

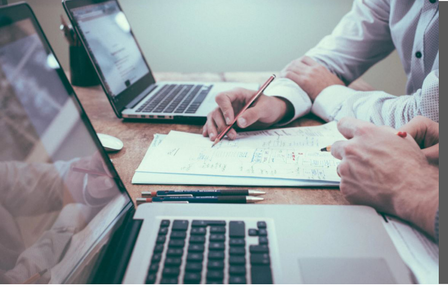
◀ BACK

NEXT ▶



To ensure the data management solution was equally flexible and elastic, the environment uses Azure NetApp Files. This cloud native file storage solution ensures secure control of data and is built on proven NetApp ONTAP performance, scalability, and security. It is a fast, simple, and metered way to get Network File Shares (NFS) for cloud workloads and integrates well with cloud add-on services such as analytics, databases, and other native services.

Our transformation to a modern analytic platform for actionable business intelligence has begun and is important to proving the fit of native cloud solutions our future IT architecture. The establishment of this environment is expected to spur demand for advanced and sophisticated decision-support while encouraging startup-like behaviors within our CDS business unit. ■



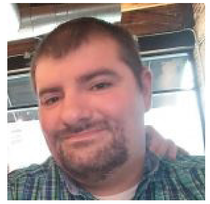
Authors

Aparna Badrinath
Aparna is a member the IT Enterprise (Information) Architecture team and works with NetApp business teams to ensure application capability in the enterprise and the cloud. This team maps applications to needed capabilities and reference architectures so organizations can most effectively achieve its current and future business objectives.

Praveen Beedanagari
Praveen Beedanagari is an SAP HANA and big data infrastructure architect at NetApp. Praveen focuses on deploying efficient Hadoop/Spark clusters by reducing data duplication across multiple clusters.

Stan Cox
Stan Cox is the Director of Customer 1, the team who provides strategic oversight of IT Infrastructure at NetApp and are responsible for early adoption and integration of NetApp products to run IT. Stan and team design, manage, and operate NetApp's storage, global data centers, hosting, and cloud strategy and delivery.

Randy Egger
As a data center lead, Randy is responsible for NetApp's IT global data center space and power capacity planning. He has local accountability and responsibility for the operations of the NetApp Hillsboro Data Center (HDC) and was heavily involved in its planning, development, and construction. Randy has more than 23 years of experience in high tech, with a focus on data center implementations and projects.



David Fox
David Fox is a Senior UNIX Systems Engineer in NetApp IT. He is responsible for managing the IT UNIX environment, including implementing IT automation using Ansible, maintaining the use of Docker in the production environment, and leveraging OpenStack. He previously worked as a systems administrator and infrastructure engineer at IBM.

Ken Lee
As one of the senior storage engineers on the Customer-1 team inside IT, Ken plans, engineers, builds and runs NetApp products and services in support of the corporation's enterprise applications. Ken has over 20 years of experience across a wide range of disciplines including DBA, SAP Basis Admin, AIX System Admin, and Enterprise Data Protection and Disaster Recovery.

Eduardo Rivera
Eduardo and his team are responsible for the design, architecture, deployment, and roadmap for all data storage services in NetApp IT. His team is responsible for storage lifecycle management, data protection, and storage innovation. He is a key contributor to the NetApp Customer-1 program, under which his team installs and runs NetApp products upon release and provides feedback on product performance.

Ezra Tingler
Ezra Tingler is a Senior Storage Engineer in NetApp's corporate IT team. In this role, he is a member of Customer-1, which acts as the first adopter of NetApp's products and services. He advocates for automation to improve IT delivery and efficiency.

◀ BACK