



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](mailto: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

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

---

“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

the StorageGRID, which is completely separate from your compute resources,” Beedanagari explains.

### 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](http://netapp.com/us/products/data-management-software/object-storage-grid-sds.aspx)

✉ [NETAPP.COM/CONTACT](mailto:NETAPP.COM/CONTACT)

+1 877 263 8277



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](http://www.netapp.com). #DataDriven

© 2018 NetApp, Inc. All Rights Reserved. NETAPP, the NETAPP logo, and the marks listed at [netapp.com/TM](http://netapp.com/TM) are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners. CSS-7021-0418