Data ONTAP 8.0 Supports Dramatic Growth in Ground-Breaking Research Data for JCVI

NetApp®
Go further, faster

J. Craig Venter™
INSTITUTE

CUSTOMER PROFILE

The J. Craig Venter Institute (JCVI) is a world leader in genomic research. The organization's success depends on research, publication of its findings, and securing grants and contracts to fund further research. Formed through the merger of several affiliated organizations, JCVI employs approximately 350 people in Rockville, Maryland, and San Diego, California.

THE CHALLENGE

Wanted: reliable data storage
The science of genomics deals with staggering amounts of data—which double every year, on average. For JCVI, high-performance computing solutions and a reliable IT storage infrastructure are mission critical.

"Data growth has slowed in the past year or so," says Eddy Navarro, computer systems manager for storage at JCVI. "But this could just be the calm before the storm. New sequencing technologies could increase our storage requirements by an order of magnitude. We have to make sure we're ready to meet today's demands—and handle any future data deluge—with high-performance computing solutions and a massive, reliable IT storage infrastructure."

Reliability is key. "If we lose data, we can regenerate it from the biological samples, but that takes time, and publication and grant deadlines don't wait," explains Navarro. "We need a storage system we can count on to keep the data safe and readily available."

Supporting world-class research
Several years ago, after the merger of several of its affiliated not-for-profit institutes into one unified organization, JCVI found itself with multiple EMC and NetApp storage systems. Navarro was tasked with creating a network-attached storage (NAS) system while keeping costs low and maximizing the value of existing investments.

"During the consolidation effort we found significant overlap in the data each organization stored, and once we removed duplicate data, there wasn't a need to keep all of the existing storage arrays," says Navarro. What the organization did need was more capacity to store its most important data—genetic structure sequences generated by machines that analyze biological samples.

Impressed by the NetApp environment's flexibility and efficiency, JCVI had standardized on NetApp NAS to store this information. Navarro realized he could add capacity to this NAS cost-effectively by implementing the NetApp V-Series open storage controller to act as a front end for the EMC arrays. "The V-Series enabled us to get value from idle arrays by making them part of

> "FlexVol has always helped us make the most of storage capacity by rapidly allocating space to projects. Now with the larger aggregate size of Data ONTAP 8, we can rest easy knowing that we're not going to run up against space constraints in the middle of an analysis."

Eddy Navarro
Computer Systems Manager for Storage, JCVI

our NetApp NAS," says Navarro. "Most importantly, it lets us manage them using NetApp data protection and management features, such as Snapshot™, SnapMirror®, and FlexVol."

Sizing up the solution
With the new infrastructure of NetApp FAS6070 storage systems and V-Series, Navarro and his team found their jobs had become much simpler, except for one thing. "As data volumes continued to grow, we occasionally had to play a shell game, moving data from one part of the system to another in order to find large blocks of free space," explains Navarro. While this did not affect performance for JCVI's scientists and other data users, it did demand hands-on management and team resources. Worse, every month or two, the team would have to take some blocks of data offline for a few hours during a weekend.

"With multiple teams drawing on the same data set, we had trouble finding a weekend when nobody was planning to run an analysis on it," explains Navarro. "Plus, I'd rather not have to ask my team to work on weekends."

The problem arose because operating systems at the time only supported data volumes that aggregated up to 16TB of data.

When NetApp unveiled the Data ONTAP 8.0 7-Mode operating system, which supports 64-bit aggregates and provides enhanced flexibility and storage efficiency, Navarro was eager to implement it. To boost performance and capacity across the infrastructure, he replaced four NetApp FAS6070 storage controllers with a single NetApp FAS6080 cluster running Data ONTAP 8.0 and new 450GB 15K rpm SAS drives.

THE SOLUTION

Making sure of availability of large and growing data volumes
At JCVI, genomic sequencing machines take biological samples, sequence the genetic structure, and place that data into Oracle® and Sybase databases. Specialized applications analyze and store the data in NetApp NAS file shares on a NetApp FAS6080 cluster in an active-active configuration, for access by a blade server farm using NFS to provide a high-performance computing grid. The NetApp FAS6080 cluster supports MySQL in addition to the file shares.

JCVI's NAS augments that cluster with a V3070 open storage controller running Data ONTAP 7.3 and managing EMC arrays that store 60TB of data for scientific project file shares, as well as a VMware® virtual desktop

datastore. By leveraging the Data ONTAP deduplication functionality, Navarro is able to achieve space efficiency savings of over 60% for virtual desktop images. SnapMirror is used to migrate data between the FAS6080 and V3070 systems, both at the Rockville location. The result is a seamless NAS managed through NetApp Data ONTAP capabilities, including FlexVol and Snapshot technologies.

"We'd been using FlexVol from the outset, because it helps us maximize the amount of data we can store in a given amount of space," says Navarro. "Now with Data ONTAP 8.0, we can create larger aggregates and as a result reduce the number of aggregates needed to manage a given amount of data." The team leverages NetApp Operations Manager to help monitor the arrays for capacity utilization and manage the performance of the system, and NetApp My AutoSupport™, a feature embedded within Data ONTAP, which unobtrusively collects data from the NetApp systems and sends alerts to NetApp Services.

In addition, My AutoSupport enables JCVI to monitor system uptime and storage efficiency. "My AutoSupport has been a tremendous time saver," says Navarro.
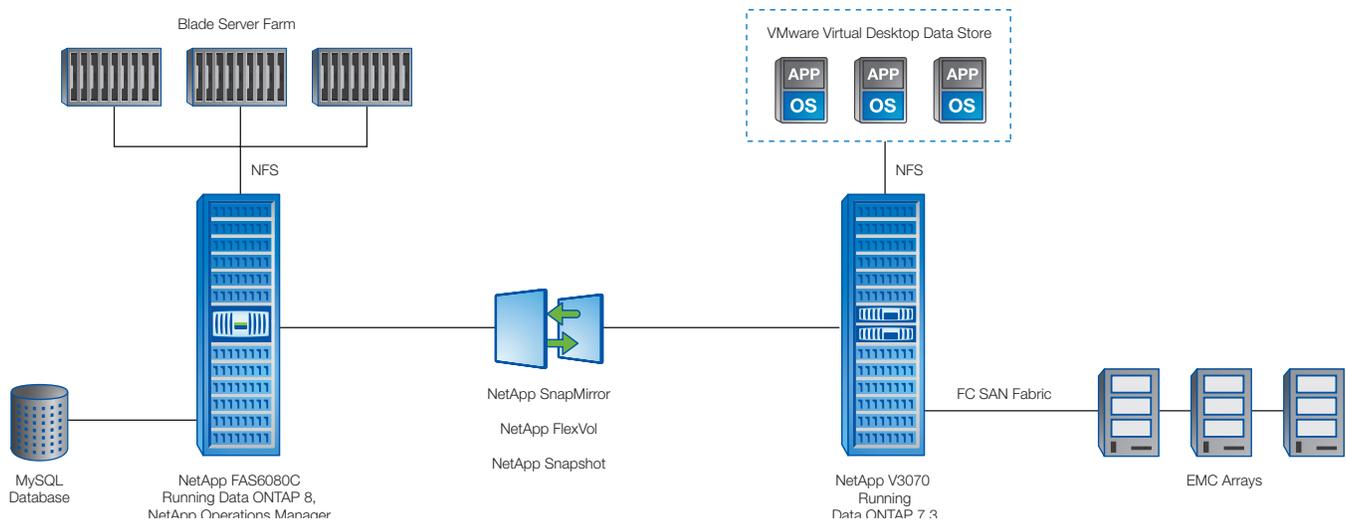
Figure 1) JCVI's Rockville data center storage infrastructure.

Diagram labels:
- Blade Server Farm
- NFS
- NetApp FAS6080C Running Data ONTAP 8, NetApp Operations Manager
- MySQL Database
- NetApp SnapMirror
- NetApp FlexVol
- NetApp Snapshot
- VMware Virtual Desktop Data Store
- APP / OS (×3)
- NFS
- NetApp V3070 Running Data ONTAP 7.3
- FC SAN Fabric
- EMC Arrays

"By collecting and analyzing data from NetApp storage systems, it gives us insight into the storage efficiencies we are achieving—and provides early warning of a potential failure."

Protecting data with Snapshot copies

Snapshot technology creates point-in-time copies for all JCVI data every four hours. Scientists use these copies to access information they've accidentally deleted or corrupted. "Even if we had a major systems problem, we'd only lose a few hours of work," explains Navarro. "Our second line of defense is a weekly backup to disk, and third, a monthly backup to tape, which we send off-site."

BUSINESS BENEFITS

The larger volume capacity of Data ONTAP 8.0 helps JCVI make sure that its NetApp storage system continues to deliver cost-efficient, highly available, and reliable storage, even as the amounts of data stored increase dramatically.

Delivering scientists data when they need it

Navarro has never had complaints about performance of the NetApp infrastructure.

Instead, the challenge is to make sure that researchers don't run out of storage while they're running a multiday analysis. "FlexVol has always helped us make the most of storage capacity by rapidly allocating space to projects," says Navarro. "Now with the larger aggregate size of Data ONTAP 8.0, we can rest easy knowing that we're not going to run up against space constraints in the middle of an analysis. We no longer need to optimize aggregate layouts, which used to take about 10% of our time."

Providing fast and easy access to deleted data

The larger aggregate size of Data ONTAP 8.0 also supports rapid data recovery. "We create frequent Snapshot copies, and our users can recover deleted data themselves from Snapshot copies within about 10 minutes," says Navarro. "This is important for supporting scientist productivity, but previously we sometimes had to make space within a volume by reducing retention or frequency of Snapshot copies. With Data ONTAP 8.0, we've been able to increase retention and frequency, helping make sure staff can quickly find and recover the data they need from a Snapshot copy, without having to wait for us."

Storage efficiencies help research funds go further

By replacing multiple storage devices with a single FAS6080 cluster, Navarro more than doubled storage capacity while also reducing the system's footprint of space and power requirements.

According to Navarro, capacity on several JCVI storage arrays would have gone unused if he hadn't been able to make them part of the organization's NetApp solution. "The V-Series enabled us to preserve our existing investment and expand our NetApp system at minimal cost," he says. The V-Series also supports JCVI's tiered storage model for ongoing cost savings. "We've pushed terabytes of data off expensive Fibre Channel storage onto less expensive SATA disk, saving tens of thousands of dollars," says Navarro.

In addition to saving money on physical infrastructure, the NetApp solution has enabled Navarro to make the most of his staff, managing over half a petabyte of storage with just three people. "The boost we get from supporting larger aggregates using Data ONTAP 8.0 has allowed us to

"The boost we get from supporting larger aggregates using Data ONTAP 8.0 has allowed us to focus on more critical projects and provide better service to researchers."

Eddy Navarro
Computer Systems Manager for Storage, JCVI

focus on more critical projects and provide better service to researchers," comments Navarro.

Navarro faces the future with more confidence now that JCVI has upgraded to Data ONTAP 8.0 and the FAS6080 cluster. "Even when the amount of data grows exponentially, we can count on NetApp to deliver flexible, efficient storage, so that our scientists have the data they need to make the most of our funding," says Navarro. "This means JCVI can continue to lead efforts to conquer disease and improve the environment through genomics research."

SOLUTION COMPONENTS

NetApp products
Data ONTAP 8.0 7-Mode

FAS6080 storage system

V3070 open storage controller

NetApp SnapMirror

NetApp Snapshot

NetApp FlexVol®

Operations Manager

NetApp services
NetApp Professional Services for implementation

NetApp My AutoSupport

Protocols
NFS

CIFS

iSCSI

Environment
Red Hat and SuSE Linux®

Microsoft® Windows®

Sun Solaris

Mac® OS X

NetApp

www.netapp.com

NetApp creates innovative storage and data management solutions that deliver outstanding cost efficiency and accelerate performance breakthroughs. Discover our passion for helping companies around the world go further, faster at *www.netapp.com*.