

Bringing Modular Object Storage to Genomics

Challenge Summary

Genomics is a branch of biotechnology focusing on the study of gene sequencing, analysis, and modification of DNA. As such, it has raised the demand for massive data storage, data services, and data placement automation.

Benefits Summary

- High-density, modular, pay-as-you-go, scale-out storage can be used across multiple sites under a single namespace
- Policy-driven data management and data placement delivers extreme efficiency
- Configurable data security
- Turnkey, worry-free deployment and locality services
- Built to lower TCO and data center footprint

Genomics data is soaring rapidly—and so are the challenges of storing, managing, and distributing it. Seagate, SwiftStack, and DC BLOX are working to change that.

To address data center challenges brought on by the growing study of genomics, Seagate is working with SwiftStack and DC BLOX to combine versatile, high-density systems with powerful, flexible software-defined storage. The result? Scalable, manageable, cost-effective, easy-to-deploy object storage.

The Science of Gathering and Storing Genomics Data

Genomics is a branch of biotechnology focusing on the study of gene sequencing, analysis, and modification of DNA. As such, it has raised the demand for massive data storage, data services, and data placement automation.

In response to these challenges caused by massive data growth, **Seagate**, the leading expert in scalable, high-performance, high-capacity, low-cost data storage

solutions, has partnered with software-defined storage leader **SwiftStack**® and data infrastructure provider **DC BLOX**. By combining the versatility and density of Seagate® Exos® E or Exos AP series hardware solutions with the power and flexibility of SwiftStack object software, genomic data customers can benefit from scalable, manageable, cost-effective, easy-to-deploy object storage.

The Challenge

A 2015 Omicsmaps study¹ focused on a collaborative effort across 55 countries and nearly 1000 sequencing centers — including universities, hospitals, and research labs.² The goal of the project? To sequence 100 million to 2 billion human genomes. To do so, they relied on 2500 high-throughput instruments.

Projects like this are proving that the data pool created by genomics is massive, which is why Medical Futurist

estimates that by 2025, 40 exabytes of storage capacity will be required to store the entire globe's human genomic data.³ Which means the storage solution for this data pool will need to be powerful and fully capable of addressing very specific needs.

What Do Genomics Customers Need from Storage?



Scalable, multi-protocol IO under a single name space.

As users continue to access the genomic data pool globally, billions of files will be uploaded in many formats and sizes. Therefore, the storage solution must be elastic, scalable, and able to support different I/O protocols under a single namespace so that the data pool is available to a large number of concurrent researchers and consumers.



Data availability and reliability.

The global, collaborative nature of this effort depends on tens of thousands of researchers and professional data consumers. Therefore, the storage solution must be reliable, consistent, and provide high data availability.



Low-cost deployment and management.

From implementation to deployment to management, the storage solution should help lower TCO through reduced hardware, cost-effective software acquisition, efficient power and cooling, and overall data center management and footprint.



The Seagate Solution

As illustrated in figure 1, Seagate builds flexible, modular solutions that provide different I/O controller modules for multiple enclosures, including the 4U106 and 5U84 enclosures with AP or E series controllers. Offering 1.344PB to 1.6PB of storage in a single chassis, these modular enclosures provide the best customer investment protection thanks to interchangeable components to support current and next-gen HDDs and SSDs.

Seagate Exos AP 5U84 combines a high-density disk array of 84, 3.5-inch disk drives supporting up to 1.344PB of data with dual server nodes. The server nodes powered by Intel Xeon® E5 v4 product family CPU can be deployed in a split-mode or share-nothing host configuration, which provides two SwiftStack storage server nodes per single Exos AP 5U84 system. Each server node supports 128GB

of DDR4 memory and PCI Express slots for adding network cards up to 100GbE. Internal to the AP 5U84, integrated 12Gb/s SAS controllers for each server provide high-speed access independently to each set of 42 drives.

Exos AP 5U84 makes it easy to manage cables, universal ports, self-configuration controls, and standardized zoning so that customers can accelerate deployment and significantly simplify development and testing of storage implementations.

When Seagate Exos AP 4U100 is included in the solution, customers benefit from industry-leading density and storage in a single system, in addition to the features of Exos AP 5U84. The Exos AP 4U100 system's unique design and precision engineering empower intelligent server computing and a whopping 1.6PB of data storage in a single chassis.

EXOS Enterprise Storage Platform with Versatility and Modularity

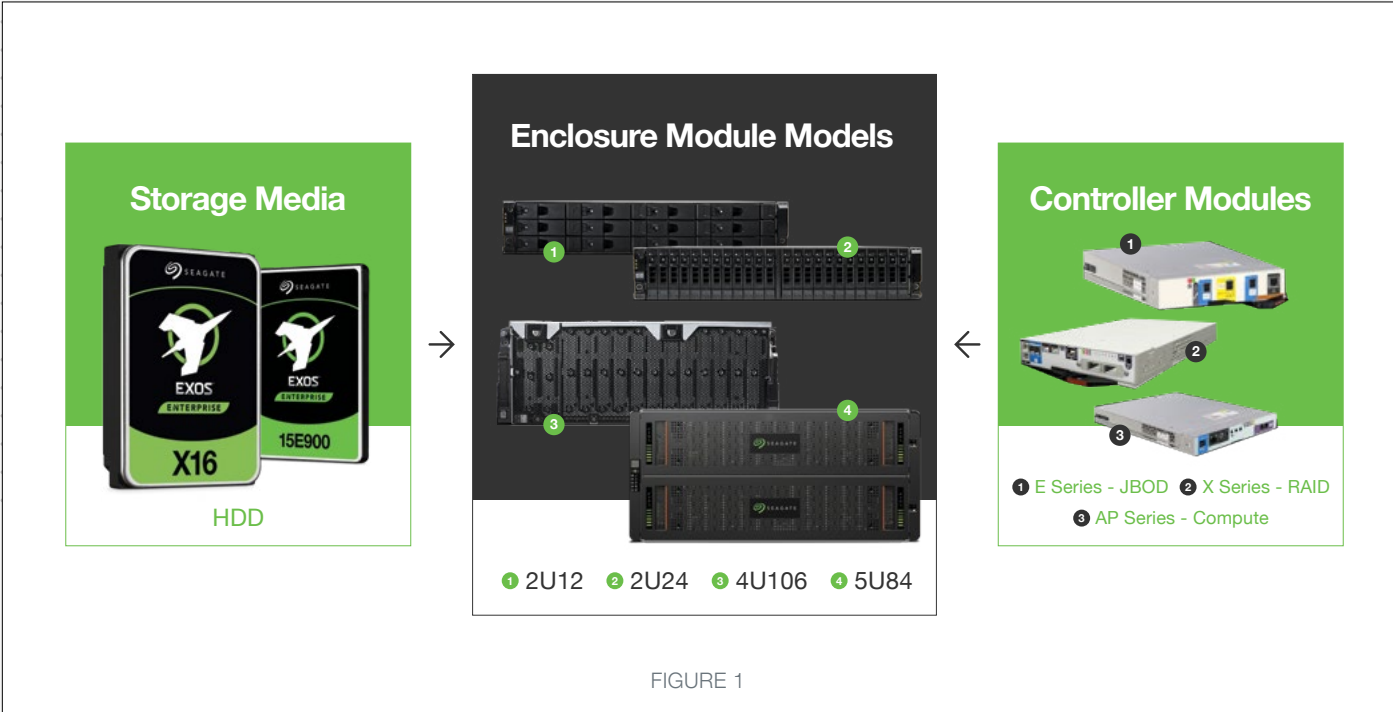


FIGURE 1



The SwiftStack and DC Blox Solution

Policy-driven data management and placement.

SwiftStack Storage Cluster is an on-premises software-defined storage platform. At its core is OpenStack Object Storage (code named “Swift”) running on two Intel Xeon-based Seagate servers. OpenStack Swift is designed to

operate on industry-standard x86 server hardware and scale out to store billions of files and petabytes without any single point of failure. Included in SwiftStack Storage Cluster software are data management services that drive:

- Proxy services to balance the front-end I/O loads and support I/O protocols, including NFS/SMB/HTTP/S3/RESTful API
- User access management
- Containerized metadata index
- Data integrity and consistency
- Configurable data protection with erasure code and replication
- Policy-driven data placement and service that store and distribute data across multiple sites and locations

Policy-driven data management and placement offers users a stress-free, set-and-forget style of storage system

administration. Figure 2 shows an architectural overview of SwiftStack Storage Cluster software.

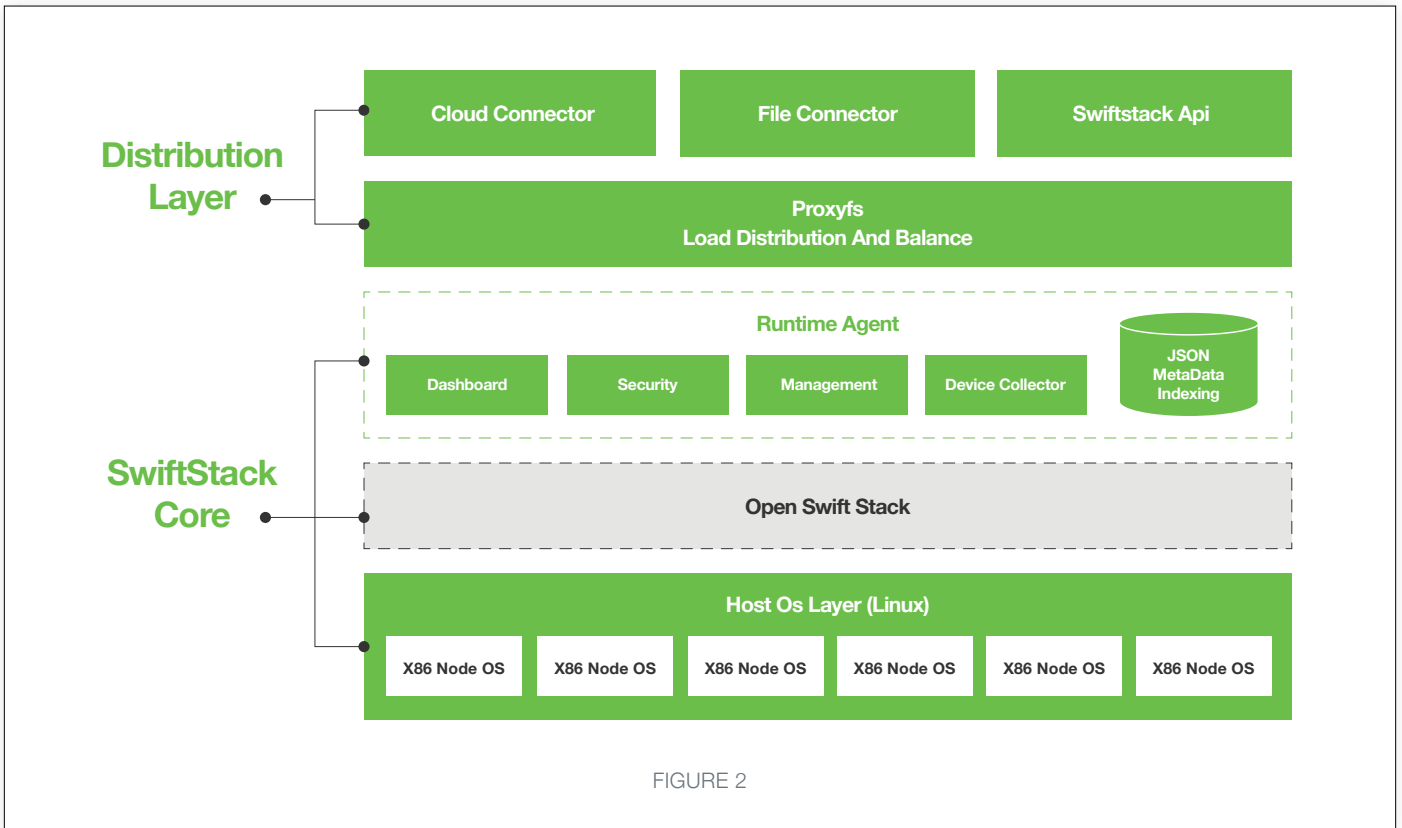


FIGURE 2



Reliable services that connect data pools globally.

DC BLOX brings years of experience in building data centers and implementing data center services. As shown in figure 3, they connect global collaboration data pools through Megaport®, a network

technology that balances customer data I/O loads at desired network performance with economic cost and high-level network security. These services enable data centers with:

- Fast access to data center colocation where the customer needs it
- Rapid turn-key hyperscale storage implementation for collaboration data pools, backup/recovery, and archiving across multiple locations
- High network performance for secure, speedy connectivity

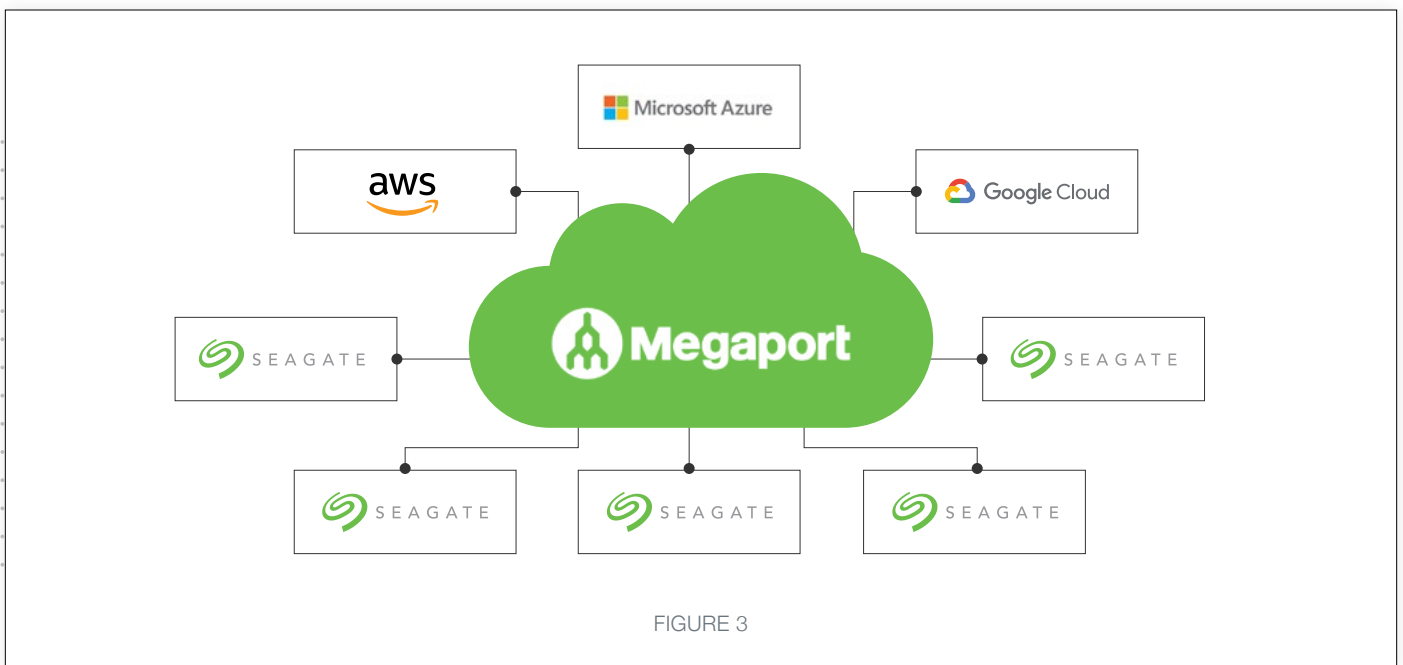


FIGURE 3

The Total Solution

Figure 4 illustrates the joint solution by Seagate, Swiftstack, and DC BLOX. In this solution, data is generated by genomic applications across multiple sites and locations. After initial processing, the applications ingest data into the genomic data pool through selected protocols, such as NFS, SMB, HTTP, or S3. The data in this pool is further distributed and analyzed by the collaboration task forces working on the project.

The collaboration data pool is hosted on the data storage systems comprised of SwiftStack storage running on an X86-server platform connected to Seagate Exos E or Exos AP series—both of which provide high-density disk arrays with performance that matches or exceeds the application requirement. Data center operator DC BLOX installs, connects, manages, and monitors the data center operation.



SwiftStack Storage Cluster (the software-defined storage stack) manages the data services and drives customer data placement. This solution architecture, with its built-in public cloud interfaces such as

AWS S3 and Swift, can also connect on-premises storage or private cloud to public cloud services such as AWS S3, Google Cloud, and Microsoft Azure.

EXOS + SwiftStack + DC Blox for Genomic Data Pool

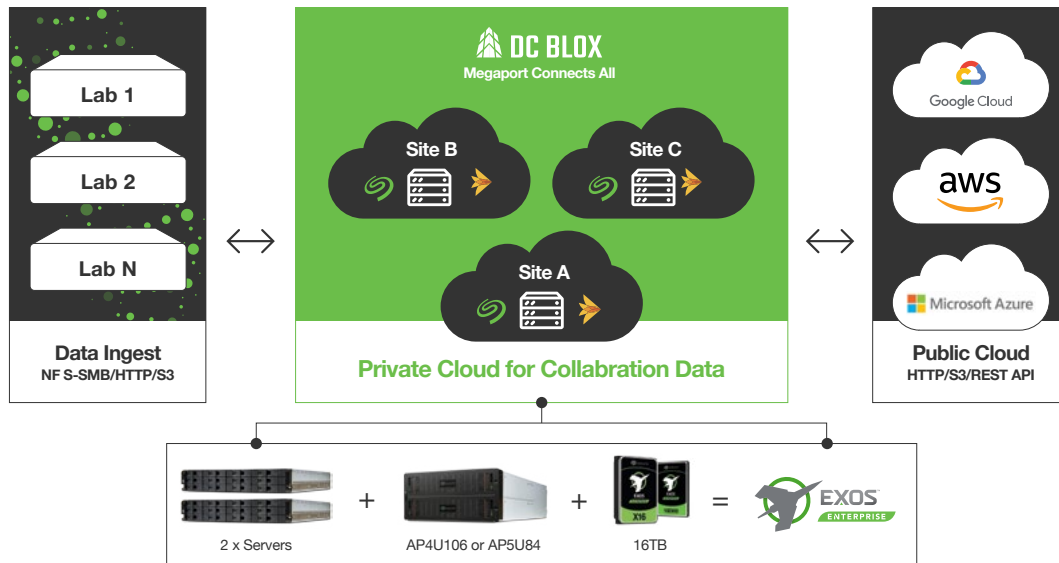


FIGURE 4



High-density media and modular storage offer pay-as-you-go, scale-out storage that can be used across multiple sites under a single namespace for optimal capacity and efficiency



High-performance storage satisfies global genomic collaboration workloads



Low-cost data storage helps lower TCO



Data security and data protection are configurable at user discretion






Turn-key, worry-free deployment and locality services for the implementation, management, and monitoring of data storage infrastructure



Policy-driven data management and data placement delivers extreme efficiency

Total Solution Benefits



Partner Name	Component Name	Component Option
	Exos E or AP 5U84/4U100	Intel Xeon server models
		High-density HDD and modular enclosure
		High performance SAS controllers
	SwiftStack Storage Cluster	Base OS + SwiftStack Storage Cluster Software
	Data Center (DC) Services	DC management and Megaport network connection with speed options

In Conclusion

Genomics data is soaring rapidly—and so are the challenges of storing, managing, and distributing it. This joint solution aims to solve these challenges by offering:

- Seamless, pay-as-you-go, modular scalability with best cost and capacity
- Set-and-forget storage administration by policy-driven data management and data placement
- Turn-key implementation of hyperscale data storage for genomic data growth
- Great savings on initial Capex and Opex, along with optimal power, cooling, and equipment footprint



Ready to
Learn More?

Visit us at www.seagate.com

About SwiftStack

SwiftStack is a leader in designing SDS software to empower enterprises to harness the power of object storage in their own data centers. SwiftStack enables enterprise users to manage, deploy, scale, upgrade, and monitor single-and multi-site clusters using the OpenStack Swift object storage engine plus additional capabilities— including LDAP and Active Directory integration, CIFS/NFS access with the SwiftStack File system Gateway, and 24x7 enterprise support. Learn more at www.swiftstack.com.

About DC Blox

DC Blox builds data center infrastructures to serve locally, connect globally, and help businesses achieve their digital transformation goals. Learn more at www.dcblox.com.

EXOS AP 5U84 data sheet

www.seagate.com/files/www-content/datasheets/pdfs/exos-ap-5u84-DS2013-3-2002US-en_US.pdf

seagate.com

© 2020 Seagate Technology LLC. All rights reserved. Printed in USA. Seagate, Seagate Technology and the Spiral logo are registered trademarks of Seagate Technology LLC in the United States and/or other countries. Nytro is either a trademark or registered trademark of Seagate Technology LLC or one of its affiliated companies in the United States and/or other countries. All other trademarks or registered trademarks are the property of their respective owners. When referring to drive capacity, one gigabyte, or GB, equals one billion bytes and one terabyte, or TB, equals one trillion bytes. Your computer's operating system may use a different standard of measurement and report a lower capacity. In addition, some of the listed capacity is used for formatting and other functions, and thus will not be available for data storage. Actual data rates may vary depending on operating environment and other factors. Seagate reserves the right to change, without notice, product offerings or specifications. SB000.0-0000US, Month 2020



SEAGATE