Western Digital.

SOLUTION BRIEF

Gain Control of Data Overload in Life Sciences

IBM Spectrum Scale

Highlights

- Increased flexibility by pooling disparate storage into a single pool under a global namespace. Native protocol support for NFS, SMB, and Object for easy integration
- Cost efficiency using software defined storage with off-the-shelf-hardware, and transparent tiering to private and hybrid clouds based on ActiveScale object storage
- Improve storage and resource efficiency by freeing up performance storage tiers through moving infrequently accessed data to ActiveScale using life cycle policies
- Automated management that puts right data in the right place throughout its lifecycle
- Extreme data durability and integrity at petabyte scale with ActiveScale, helping ensures valuable data is protected long-term and always available as an active archive
- Greater IT agility by being able to quickly react, provision, and redeploy storage resources in response to new business requirements

Challenge

- Scaling storage capacity to meet increasing sequencer data volume, and growth related to capturing more annotation data
- Data protection and integrity is critical throughout the information life cycle. Data retention times are increasing to meet regulatory requirements and to support ongoing research and development
- Workload diversity creates a complex mix of file sizes and access patterns, that require both high throughput and low latency
- Matching the right storage technology with the data as it moves through its life cycle. Better leverage cloud models to control cost
- Performance to keep up with real-time instrument output along with the wide variety of applications and users requiring access to the data globally, turning it into useful knowledge

Rapid advances in Life Sciences R&D are making it possible to deliver medical diagnoses and treatment based on a person's genetic makeup. Delivering better, faster and more cost-effective healthcare means Life Sciences organizations must accelerate time to results. Their ability to store, process and analyze high volumes of data with greater efficiency is essential.

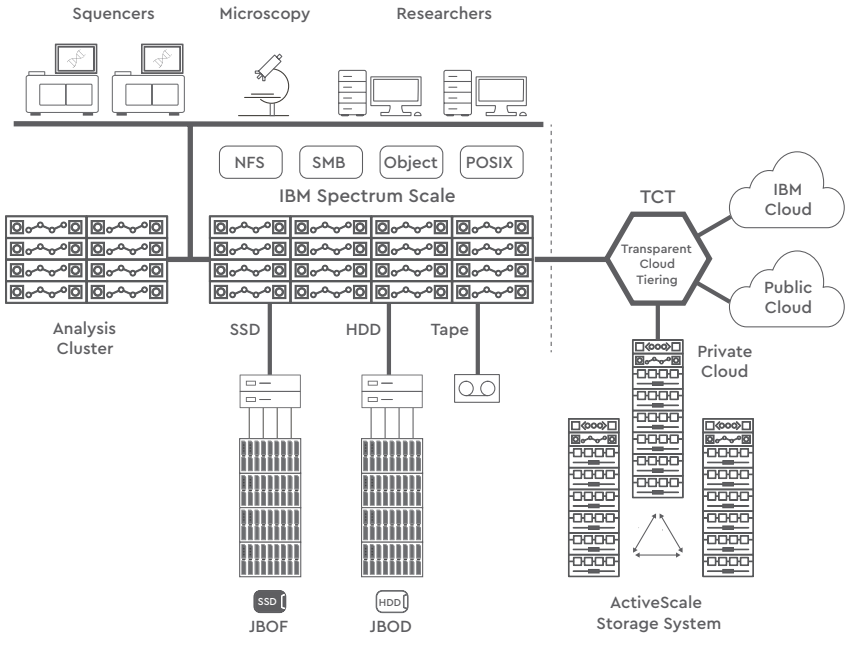
Next-generation genome sequencers are producing more data than ever. Each run produces multiple terabytes of data that need fast analysis and comparison to large stores of genomic data.

New analysis techniques with a greater emphasis on collaboration along with explosive data growth, are causing organizations to consider new approaches for deploying high-performance computing (HPC) and storage infrastructure.

Diverse Datasets and Workloads

Data formats and performance characteristics vary greatly depending on the application workload. Performance requirements can range from extremely demanding for the processing of genome data, to less demanding for collaboration and archiving reference data. Life Sciences R&D environments typically deploy fast parallel file systems using standard file interfaces, while large microscopy images are best served using objects.

IBM Spectrum Scale with TCT + ActiveScale Life Sciences Architecture



Life Sciences architecture example using an ActiveScale object storage system

Solution

IBM Spectrum Scale™ with Transparent Cloud Tiering and the Western Digital ActiveScale object storage system. The combined solution provides extreme scalability with support for flash accelerated performance and automatic policy-based storage tiering of flash through disk to tape. Native protocol access allows clients to seamlessly access all data from a shared global namespace. Transparent cloud tiering enables users to easily create and manage a private or hybrid cloud using the ActiveScale system to help reduce storage costs. Data that does not require high performance can be transparently moved to ActiveScale object storage.

Legacy storage architectures lack the flexibility, scalability and accessibility needed modern workloads at petabyte scale. Data is often silo'd by architectural limitations even though total system capacity is sufficient. This often leads to redundant cost and higher management complexity.

Next generation storage architectures based on modern technologies and storage constructs can eliminate the performance bottlenecks, complexity, and cost of traditional storage systems.

Breakthrough Storage Efficiency

IBM Spectrum Scale with Transparent Cloud Tiering is software-defined storage for applications that demand high performance and shared access to a common set of local or remote data. It provides extreme scalability with support for flash accelerated performance and automatic policy-based tiering across flash, spinning disk and tape media. Native protocol access allows clients to seamlessly access all data via a shared global namespace. A highly available cluster can range from two nodes to thousands in order to meet the most demanding Life Sciences workload requirements.

Private and Hybrid Cloud Storage Made Easy

Transparent Cloud Tiering enables users to integrate Western Digital's ActiveScale object storage to create a private or hybrid cloud to help reduce storage costs. Data that does not require high performance storage can be transparently moved to ActiveScale using life cycle policies. This is ideal for data that needs to remain on-premises or has a higher recall or access frequency that would make off-premises cloud storage too expensive.

IBM Spectrum Scale combined with ActiveScale object storage allows Life Sciences and other industries gain control of data growth with:

- **Unified storage** that supports a diverse set of Life Sciences applications and workloads where performance, reliability, and availability of data are essential to the business. Native protocol support for NFS, SMB, Object and POSIX allow the solution to seamlessly integrate into existing environments.
- **Improve storage efficiency** by pooling redundant isolated storage resources under a single global namespace. Free up performance storage tiers by transparently moving infrequently accessed data to ActiveScale using automated life cycle policies.
- **Accelerate workflow and collaboration** to translate genomic research into insights that can contribute to better patient care. Spectrum Scale provides a global name space across all storage tiers, while ActiveScale offers fast access vs. tape and with lower management complexity making it easier for teams to collaborate boosting overall productivity.
- **Extreme data durability and integrity at petabyte** scale helps ensure valuable data is protected and always available. ActiveScale delivers up to 19 nines durability and site-level fault tolerance in a multi-site configuration. Robust data integrity checks occur automatically and transparently protecting long-term archives; each object can tolerate up to 1000 bit-errors without data loss.
- **Easy to Install, and manage** by virtue of being software defined storage. ActiveScale systems are also easy to deploy – simply add power and network connections and it is ready to go. The system self-protects and heals requiring significantly less IT intervention compared to traditional storage systems.

Conclusion

The scale of the data challenge faced by Life Sciences and genomics organizations is daunting. Sequencers are outputting extreme amounts of data that needs to be quickly stored, processed, analyzed and kept for long periods of time. The software defined storage solution from IBM and Western Digital deliver the necessary performance, scale, and efficiency to address these challenges with fewer resources and lower cost than traditional approaches.

To learn more about ActiveScale visit: www.wdc.com/dc-systems

Western Digital.

5601 Great Oaks Parkway
San Jose, CA 95119, USA
US (Toll-Free): 800.801.4618
International: 408.717.6000

www.westerndigital.com