

SCALE

innovation and discovery.

The power of HP Converged Infrastructure for genomic research

Scalable solutions that simplify data analysis and access; add components as you need them.

Solution brief

HP simplifies the genomics research process

Life sciences research is advancing exponentially, each step bringing us closer to the realization of truly personalized medicine – preventive care and treatments designed specifically for each individual. Healthcare researchers soon expect to be able to use predictive genetic testing to create custom treatment plans for individuals. But research capabilities are only part of the equation; current storage and operating capacities must also evolve to quickly and easily accommodate ever-expanding amounts of data before the goal of personalized medicine can be realized.

Genomic, clinical, and pharmaceutical research generates a staggering amount of data. You require technology solutions that are easy to deploy and manage, easily scalable, minimize risk and cost, and that help effectively transform ever-increasing amounts of data into meaningful results.

That's where HP's Converged Infrastructure approach comes into play. Featuring the fully tested and factory-integrated HP Cluster Platform and the HP X9000 Network Storage System, it allows you to:

- Easily manage and grow your data capacity in a tiered approach, all in a huge single namespace that enables researchers to easily access files without knowing their location.
- Effortlessly support computationally intensive genome sequencing and analysis workloads.
- Seamlessly deploy a complete IT solution with measured outstanding performance for genomic sequencing pipeline.

HP is committed to your success – it's in our genes

For more than 25 years, HP has been advancing research as a trusted IT infrastructure leader, innovator, and partner with leading life sciences facilities like yours. In fact, the first human genome was sequenced in 2003 using HP compute infrastructure.

Today, HP continues to be a trusted IT partner with the leading commercial and public life sciences organizations engaged in genomic research worldwide.

HP has invested in partnerships with leaders in genome sequencing to characterize the workflow and to develop solutions specifically targeted at the unique workloads of genomic research. Most recently, HP collaborated with Partners HealthCare Center for Personalized Genetic Medicine (PCPGM) to conduct a series of tests to validate the performance of the HP X9000 Network Storage System and HP Cluster Platform for Life Sciences running the Illumina genome analysis pipeline software. Test results show that HP StorageWorks X9000 Network Storage System can readily support concurrent processing of a number of sequence runs, offering pay-as-you-grow storage capable of handling exponentially increasing volumes of data without compromising performance.

HP solution workflow

HP applies its 25 years of industry leadership experience to help life sciences professionals meet the most demanding computational and storage needs. Flexible and scalable solutions – powered by HP compute platforms, supported by HP storage, and backed by HP services – enable researchers to drive up success while driving down both risk and cost.

HP's StorageWorks X9000 Network Storage Systems provide the massive scale storage with industry-leading density and costs using scale-out, industry-standard architectures. The HP X9000 systems' input/output performance has been characterized and tuned for each stage of the genomic sequencing pipeline under various genomic sequencing workloads using large sets of DNA samples from ongoing experiments. The single namespace provides the ability to start small and grow seamlessly as your needs increase. With automated policy-driven



data placement features, the HP X9000 greatly reduces storage and administrative costs and data management complexities while providing a high degree of concurrent throughput for downstream analysis. The results have been published in conjunction with PCPGM.

HP is unique in the industry for providing complete end-to-end server, storage, and networking solutions for genomic research that are easy to deploy, simple to manage, and straightforward to service and support. As a life sciences professional, you are then able to focus more on your research and less on designing, deploying, and operating infrastructure. HP has already done all the work for you.

HP Converged Infrastructure helps eliminate silos of information and processes in the research organization to enhance collaboration and communication with internal and external partners. HP Converged Infrastructure is poised to improve life sciences researchers' ability to aggregate and analyze complex genetic and clinical data, helping them to make new discoveries and formulate therapeutic applications that will make personalized medicine a reality.

- **HP Converged Infrastructure** – Delivers new levels of simplicity, integration, and automation; uniting servers, storage, networking, applications, power, and cooling on one highly efficient platform.
- **HP StorageWorks X9000 systems** – Extremely scalable, flexible, and cost-efficient, HP X9000 Network Storage Systems future-proof your investment, delivering excellent performance and a modular storage infrastructure to accommodate unprecedented storage growth and performance for present and future life sciences research needs.
- **HP ProLiant servers** – Available in a wide variety of form factors – including factory integrated

and tested high-performance HP Cluster Platforms – industry-leading HP ProLiant servers and HP BladeSystem cClass deliver exceptional reliability and performance for genome sequencing and analysis workloads

- **HP Networking** – HP Virtual Connect and Flex Fabric optimizes application performance, ease of deployment and management, and increases availability throughout the lifecycle of your network infrastructure.
- **HP services** – As a global IT provider offering comprehensive expertise to align applications, infrastructure, and facilities to your business goals, HP helps you modernize applications and infrastructure to reduce complexity, align business and IT processes, and maximize efficiency.

HP StorageWorks X9000 Network Storage System adapts to your needs

The HP StorageWorks X9000 Network Storage System product family offers flexible, scale-out storage solutions that have been precisely engineered to overcome the challenges of accessing, analyzing, and managing multi-petabyte stores of multiple data types. A low upfront investment allows you to start small and grow over time by independently scaling storage performance and capacity as your needs change. And resources can be reprovisioned or redeployed on the fly to offer a future-proof path to network evolution, without having to rebuild or eliminate current IT infrastructures.

A component of the HP Converged Infrastructure solution, HP X9000 Network Storage Systems address changing file serving needs while reducing ownership cost by enabling simplified management, better resource utilization, and centralized growth and data preservation.

Note: The following configurations are examples only. Your actual configuration will vary, depending on your requirements. Consult with your HP representative for an official recommendation that meets your unique needs.

High-throughput Sequencing Dataflow

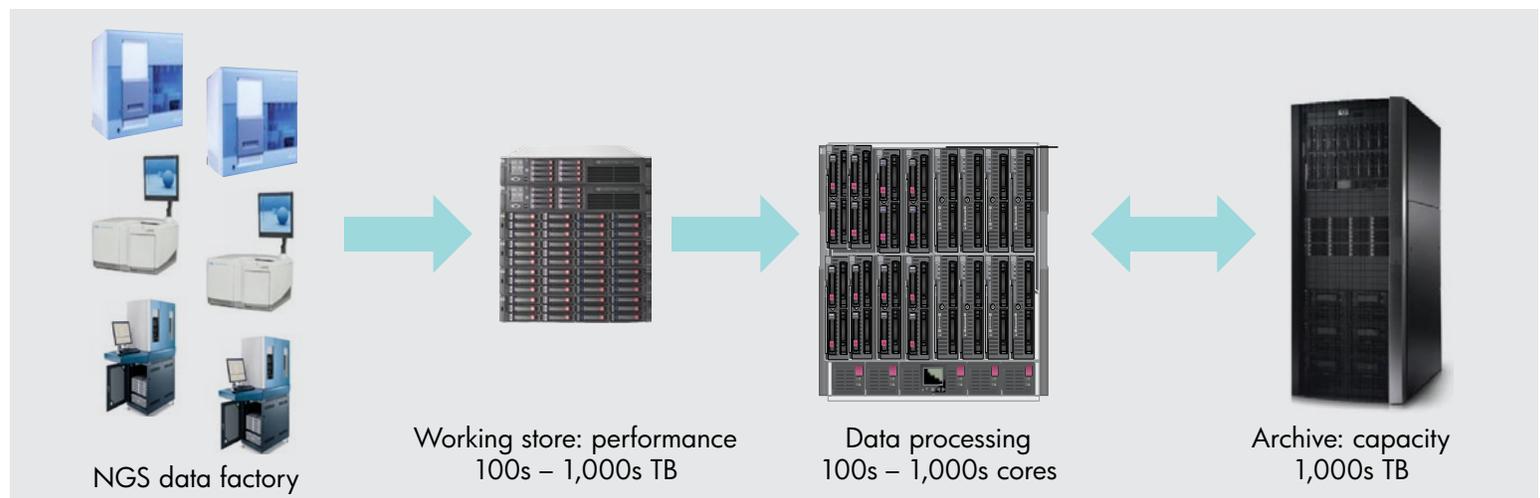
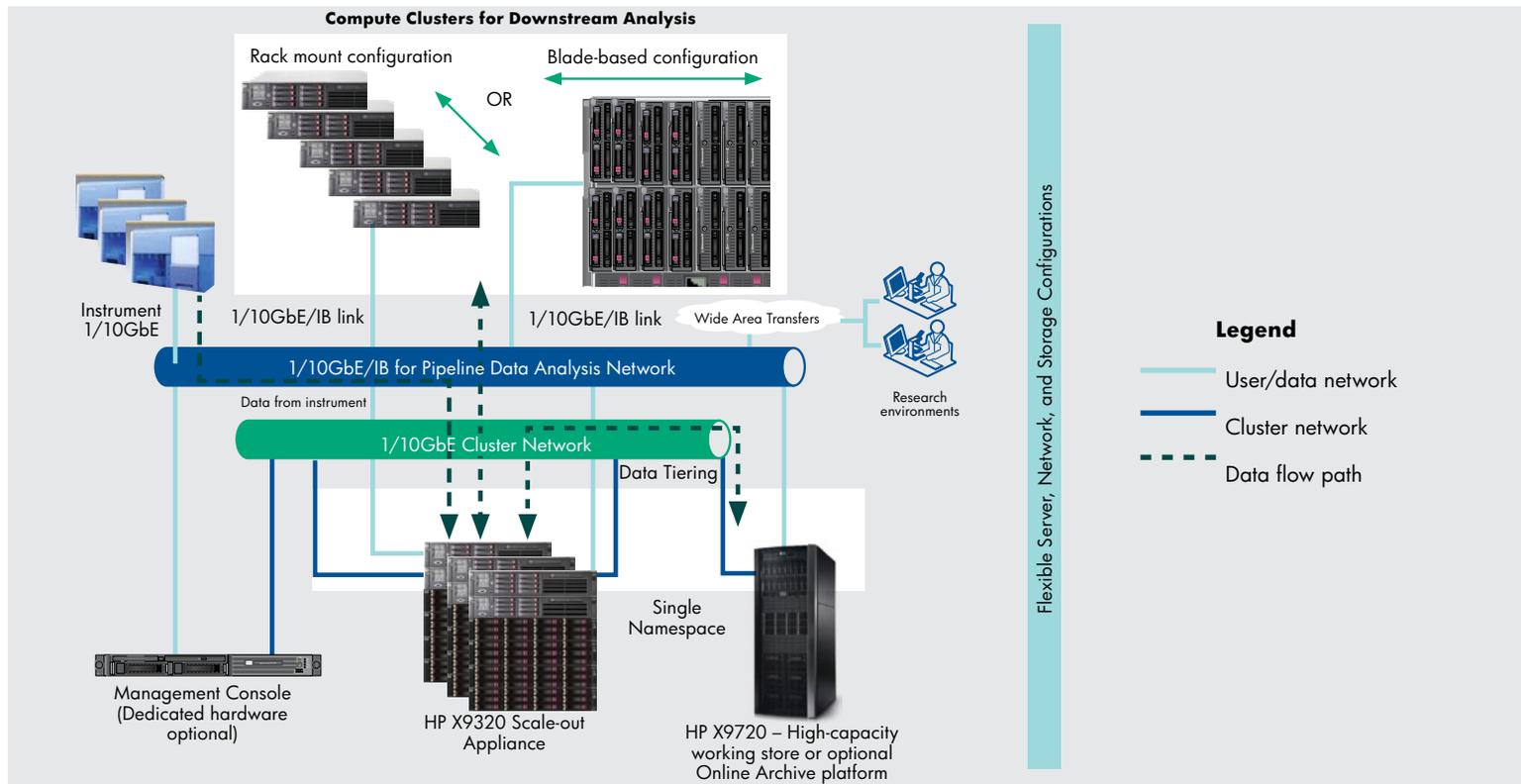
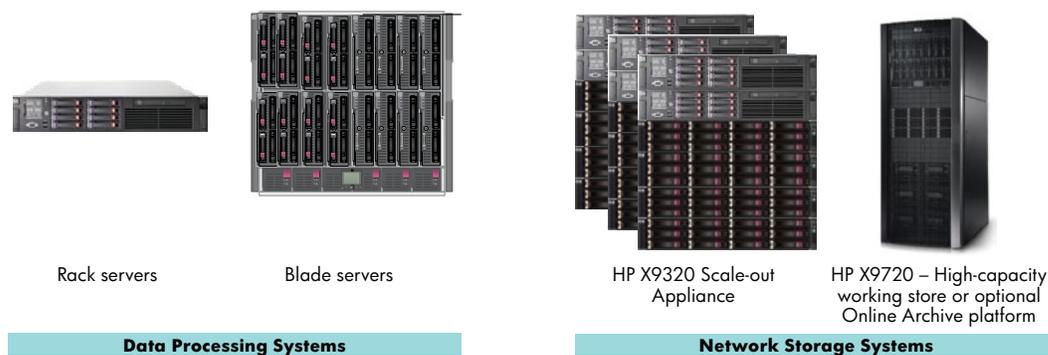


Figure 1. High-throughput sequencing workflow with HP StorageWorks Network Storage System. Output files from sequencing instruments are moved first to a performance-optimized working store – which also supports the subsequent data processing analysis steps – typically on a small compute cluster. Quiescent data for long-term retention is then moved to a capacity-optimized archive.

HP Solution Architecture for Next-Generation Sequencing



HP's Scalable Architecture Elements



Component

Processing	HP ProLiant Rack Mount or BladeServer high-performance compute cluster running pipeline software. Incremental addition of compute power as necessary.
Storage	HP X9320 high-performance appliance. Flexible network options to suit different data center needs. High density, cost-optimized HP X9720 appliance for large content work stores as well as long-term archives. Single namespace for easier management across multiple appliances.
Network	Flexible networking options. Includes 1GbE/10GbE or Infiniband options. Flex fabric with blades for change-ready architectures.

Figure 2. A network configuration example that implements the workflow illustrated in Figure 1. Several HP X9320 scale-out network attached storage appliances serve as the main performance working store, with an HP X9720 providing additional working store as well as archive within a single namespace over multiple data tiers. Depending on performance requirements, data processing and file system cluster networks can be 1 or 10 gigabit, or Infiniband networks.

HP StorageWorks X9000 Network Storage Systems provide answers to increasingly complex issues. A single namespace offers ease of data access and scalability, as well as seamless data tiering and archiving, with balance in performance and capacity.

Equip your research team with the tools to make groundbreaking advances in genomic technology. Ultimately, HP StorageWorks X9000 Network Storage Systems lead to faster results, more productive researchers, and improved time to discovery.

The flexible solution architecture shown in Figure 2 permits solutions at all scales to be implemented according to particular requirements for performance, capacity, and growth.

• **Entry-level solutions** can be built with a few rack or blade ProLiant servers with the latest in multi-core CPUs, more than 8 GB per core memory, and cost-effective

• embedded 10 Gbit networking. The HP X9320 Network Storage System, starting at 21 TB, offers an excellent working store foundation for storage solutions that can be easily scaled in a single namespace.

• **Mid- to large-scale solutions** start with HP BladeServer Cluster Platform with 8, 16, 24, or more compute server nodes with integrated cluster management software from HP. Depending on scale and throughput requirements, the working store can include several HP X9320s, and one or more HP X9720 NAS appliances. These systems target aggregate sequencing throughput of 500 to 1000s of gigabases per week.

• **Massive scale solutions** incorporate additional HP X9720 appliances for multi-petabyte data stores, large capacity archives, or remote replication for data security.

Both disk-to-disk and tape-based backup solutions can be deployed with solutions at any scale. HP Solution Architects provide expert assistance in designing and recommending solutions best suited to individual requirements.

“With the advent of large-scale microarray gene expression analysis and next-generation sequencing technologies – including a growing number of Illumina sequencers – Partners HealthCare Center for Personalized Genetic Medicine experienced dramatic increases in the need to manage growing volumes of data. Our existing system was hampered by the need to adjust file systems manually and move data to handle new growth.

HP helped us formulate a solution based on the HP X9000 Network Storage System. The X9000 is an easy to use, expandable, and flexible storage platform that allows us to address our data management needs without having to manage multiple file systems, and it eliminates the need for manual data movement.”

Brent Richter, Director, Enterprise IS, Partners HealthCare Systems

“HP’s X9000 Network Storage System offers a rich portfolio of solutions to large-scale storage and data management challenges in next-generation sequencing and related areas of life sciences bioinformatics. The ability to scale both capacity and performance within a single 16-petabyte namespace allows customers to deploy and seamlessly expand storage and file systems for explosive data growth – the hallmark of life sciences in research and the clinic. Mix-and-match architectural options under a single management console ensure ease of use and access to data for everything from high-performance computing to long-term archival storage to remote replication for high availability.”

Les Fox, Distinguished Technologist, HP

HP solution in action

Exploding data storage demands threatened to hinder the progress of biomedical research at PCPGM’s academic medical centers of Massachusetts General Hospital, Brigham and Women’s Hospital. To overcome these challenges, PCPGM’s Enterprise Research Infrastructure and Services (ERIS) worked with HP to conduct a series of throughput and scaling tests of the HP StorageWorks X9000 Network Storage System with an XC3000BL compute cluster running the Illumina v1.5.1 pipeline software.

Together, HP and PCPGM were able to demonstrate that a small compute cluster with HP X9720 Network Storage System was able to process a large volume of data output with no evidence of input-output bottlenecks. Test results showed that the HP X9720 offers pay-as-you-grow storage that’s capable of handling exponentially increasing volumes of data without compromising performance.

Resources

To learn more about the HP StorageWorks X9000 Network Storage System for health and life sciences, visit www.hp.com/go/healthandlifesciences.

Share with colleagues

