

Sun Business Ready HPC for Life Sciences



Highlights

- Scalable efficiency — optimized compute, storage, networking, and software components deliver maximal application performance
- Dramatically reduced complexity — fully integrated hardware and software eases deployment and use
- Breakthrough economics — Sun Open Network Systems architecture provides industry-leading price/performance
- Increased productivity — run more simulations and computational jobs, get more work done faster
- Rapid deployment — factory integration and professional services available for faster, easier installation
- Reduced risk — preintegration and factory testing decrease installation issues

Target applications

- Genomics
- Early-stage drug discovery
- Molecular structure and dynamics prediction
- Systems biology
- High-resolution imaging
- Open laboratory data management



Life sciences research is in the middle of a revolution — the open-science revolution. New instruments rapidly generate terabytes of information that enable unprecedented insights into everything from genomes to environmental impacts. The holistic approach to systems biology is fueling the growth of interdisciplinary research, and community-developed application software is democratizing that research, empowering any scientist, anywhere, to come up with the next big discovery.

The open-science revolution requires solutions that scale to meet exploding compute and storage requirements. Sun Business Ready HPC for Life Sciences meets these challenges with a scalable High Performance Computing (HPC) infrastructure that is capable of supporting this massive growth of open-science data. Sun Open Network Systems architecture enables new levels of scalable performance through optimized systems, innovative storage and file systems, high-speed networking, and a comprehensive software and management stack in an integrated and easy-to-use platform.

As a result, researchers can execute life sciences workflows with any mix of high-throughput, high-performance,

large-memory, and I/O-intensive applications. This enables organizations to easily adapt to the unpredictable demands of research, ensuring that investments made today will support future workload- and application-driven challenges.

Life sciences research organizations around the world benefit by increasing throughput for bioinformatics and virtual screening, accelerating molecular modeling and quantum mechanics simulations, improving database and file access, and simplifying data archiving. Sun Open Network Systems deliver unsurpassed versatility, performance, system integration, and manageability while reducing acquisition and operating costs by up to 75%.¹

Accelerating insight for life sciences research

The Sun vision for life sciences research is to provide open, integrated, end-to-end HPC and data management solutions that accelerate the pace of scientific inquiry and qualitatively improve the insights it delivers.

Genomics

Next-generation sequencing instruments are inundating researchers with massive amounts of data that must be processed, assembled, sequenced, annotated, and then used in broader GWAS or phylogenetic studies. The advancement of genomic research is tightly tied to the efficient automation of storage and processing of data.



Sun Blade 6048 modular system

Sun Storage 7000 Unified Storage Systems and the Sun Lustre Storage System deliver easily managed, scalable capacity that can be extended with HSM to make almost limitless amounts of data directly accessible for later reprocessing or analysis. Sun Grid Engine prioritizes critical computational tasks and optimizes overall throughput by ensuring that high-performance, memory-intensive applications run on the appropriate compute nodes or clusters. This combination enables scientists to focus on genomics instead of on computers that support it.

Early-stage drug discovery

Virtual screening and cheminformatics quickly eliminate “uninteresting” compounds during early-stage drug discovery, so more resources can be focused on promising

candidates. The recent development of flexible models for virtual screening has turned 2D database matching into sophisticated 3D analysis of many conformations, significantly increasing high-throughput computing and I/O requirements.

Sun solutions improve these workflows by efficiently distributing independent tasks to every CPU available and eliminating storage bottlenecks — no matter how large the HPC cluster grows.

Molecular structure and dynamics prediction

Uncovering the time-varying 3D structure of biological molecules and predicting interactions are critical steps in the drug discovery pipeline. Molecular modeling and quantum mechanics use thousands of CPUs, terabytes of memory, and GB/sec of I/O to identify the structure of a virus or the bonding of ligands and proteins.



Sun Blade X6275 blade server

Sun solutions accelerate molecular modeling and quantum mechanics simulations with fast x64 processors, high-memory bandwidth, low-latency communications, and high-bandwidth parallel file systems and have been shown to scale to over 150x performance on 192 cores.

Systems biology

Understanding an organism is about more than understanding each of its subsystems. Systems biology uses multidisciplinary research to gain insights into relationships between everything from genomes to the environment by focusing on the network of explicit and inferred connections between



Sun Storage 7000 Unified Storage System

subsystems. Computationally, this requires the construction of modular frameworks that link simulations running at different physical and temporal scales, so changes effected in one subsystem can be understood everywhere.

Sun solutions are designed to address the full range of systems biology applications, from those that run on single nodes to those that link tens of applications and thousands of CPUs. Versatile Sun servers efficiently support compute-intensive applications and communication-intensive frameworks that link them together.

High-resolution imaging

New experimental science is being driven by imaging systems that generate up to 5 TB per day. To gain insight from this data, it must be stored, analyzed, and archived in an automated fashion so scientists can focus on interpreting results. A single instrument typically supports several research teams with different analysis algorithms, so a versatile computing infrastructure is needed.





Sun StorageTek tape libraries for archiving of life sciences data

Sun Open Network Systems have the versatility to run different styles of image processing and analysis, from high-throughput processing of volumetric time series to high-performance analysis of terabyte data sets. Sun Open Storage Systems handle the most demanding high-resolution imaging requirements with NAS services that expand to over 250 TB; Sun Lustre Storage Systems that scale to 250 million files, petabytes of storage, and over 100 GB/sec of I/O; and the Sun Storage and Archiving Solution for HPC, which uses HSM, disk, and tape technology to provide almost limitless, long-term, cost-effective storage for multi-TB data sets.

Open laboratory data management

The increasingly collaborative world of life sciences research means that discovery teams might be distributed across dozens of academic, government, corporate, and contract research groups around the world. Each group generates its own data and analyzes it with vendor-supplied programs, making it difficult to track data, reprocess it, and share results.

Sun HPC and storage solutions are open-standards based, and Sun works with application partners to promote open scientific storage formats so key insights can be generated by anyone, anywhere, at any time, instead of being tied to specific instruments and software. These open scientific standards simplify the management and use of data over extended periods of time.

Efficient and scalable application performance

With a host of world-record benchmarks, Sun Fire rackmount servers and Sun Blade systems, in combination with Sun Open Storage products, are especially suited to delivering application performance and scalability. Sun's innovative servers combine the power and communication bandwidth needed for quantum mechanics and high-resolution imaging; the I/O speed needed for bioinformatics, virtual screening, and cheminformatics; and the memory size needed for sequence assembly and GWAS analysis.

Sun's latest storage products, including Sun Storage 7000 Unified Storage Systems, take full advantage of Flash technology to deliver high-performance storage with up to 75% lower cost than traditional storage solutions. For environments with immense data storage requirements, the Sun Fire™ X4540 storage

server, the first server designed for open-storage environments, offers the highest throughput rate of 3 Gb/sec and 48 TB in 4U of rack space — twice the capacity, in half the space, at half the cost of the competition.

Making HPC for life sciences easy

Sun Business Ready HPC Solutions are designed to be easy to deploy, use, and manage. You can expect out-of-the-box performance, ease of use, and reliability from these factory-built and -tested systems. By leveraging the Sun Compute Cluster and Sun Open Storage Systems, you can quickly deploy infrastructure for life sciences research that delivers breakthrough economics, with the power to complete more research and bring new products to market faster.

Sun Professional Services provides the flexibility to customize solutions to specific application requirements. Sun offers a range of services to help enable successful deployments, including:

- Professional services, including project management and field installation support services for ongoing maintenance.
- Sun HPC Quick Start Services make it easy for you to adopt, implement, and optimize HPC solutions, helping to decrease deployment time by up to 80%.

Sun Open Storage Solutions

When next-generation sequencers generate 1 TB of data per day, proteomic databases exceed 10 TB, and molecular dynamics simulations create hundreds of gigabytes of results, it is clear that HPC storage is crucial for life sciences research. Sun Open Storage Solutions for HPC use standard components and Sun-developed open-source software to deliver high-capacity, high-performance, high-reliability, simplified management as well as exceptional value to life sciences customers.



Sun Grid Engine
qmon Job Submission GUI

Costly proprietary solutions lock you in and limit your ability to scale your infrastructure or adapt to changing business needs. Sun Open Storage Solutions for HPC combine open-source software with industry-standard hardware, enabling you to reduce your reliance on high-priced systems and to save up to 75% on storage costs. Solid-state drive (SSD) disks also dramatically increase performance while cutting power consumption by up to 80% compared to spinning disks. They're just what you need to reduce space and energy costs and improve scalability. Sun's offerings include:

- **Flash-based storage** — The latest Sun Open Storage solutions, including Sun™ Storage 7000 Unified Storage Systems, take advantage of Flash technology and ZFS™ File System to deliver high performance at up to 75% lower cost than traditional storage solutions. Sun has also fully integrated Flash technology into its new Sun Blade™, x64, and CoolThreads™ technology-based servers to help customers boost application performance without adding more servers or more high-speed disk drives to their HPC environment. Flash technology can be used with HPC applications to accelerate hot data files, alleviate the need to overbuy disks for performance, and reduce the need for installing large memory pools in compute nodes.
- **Network storage** — For home directories and shared scratch on small HPC clusters, Sun Storage 7000 Unified Storage Systems offer tremendous performance, scaling, reliability, and ease of management at a fraction of the cost of traditional NAS filers. Sun Unified Storage Systems utilize Hybrid Storage Pools and ZFS File System to provide high-performance NFS storage with industry-leading throughput and price per gigabyte of capacity.
- **Parallel storage** — For scalable environments with thousands of cores that need fast I/O for high-throughput bioinformatics and high-performance computational chemistry simulations, the Sun Lustre Storage System uses an easy-to-deploy, modular architecture to scale from 1 GB/sec to 100 GB/sec of performance and from tens of terabytes to petabytes of capacity. The Sun Lustre Storage System provides superior throughput for applications, as well as access to a single global name space. Based on the Lustre file system and Sun Open Storage disk subsystems, the Sun Lustre Storage System can deliver I/O performance in excess of 100 GB/sec.
- **Archive storage** — To address the escalating cost of storing large volumes of HPC data, Sun has a complete line of tape and archive products and software solutions for archival management. Sun StorageTek™ tape drives and libraries provide power-efficient and cost-effective media for storing large volumes of data. Built around Sun StorageTek™ Storage Archive Manager software, the Sun Storage and Archive Solution for HPC provides automated, policy-based archiving and on-demand, transparent file retrieval — so the cost of building and managing large data repositories can be reduced through low power consumption, low cost media, and reduced effort for data management and operations.

Learn More

To learn more about Sun Business Ready HPC Solutions, go to: sun.com/hpc.

Sun and life sciences

New instrumentation and the open-science revolution enable new levels of analytic and holistic research, from genomes to cells to environmental impacts. Terabytes of data need to be stored, analyzed, and archived every day, and new applications and integrative frameworks need to be generated.

Sun's Business Ready HPC offerings deliver no-compromise versatility that supports life sciences workflows with any mix of high-throughput, high-performance, large-memory, and I/O-intensive applications. These systems leverage Sun's innovative open-systems architecture, open-source software, and high levels of systems integration to deliver unsurpassed performance and value. Sun HPC and storage systems are easy to deploy, manage, and use, so life sciences researchers can focus on generating insight instead of managing computers.

¹ 75% reduction based on Sun Storage 7000 Unified Storage Systems comparisons to NetApp filers.