

**ACCELERATE
BIOPHARMA RESEARCH**
ELIMINATE BOTTLENECKS AND
SPEED-UP TIME TO RESULTS
WITH DDN'S ADVANCED
STORAGE PLATFORMS.

ADVANCED STORAGE PLATFORMS FOR NEXT GENERATION PHARMACEUTICALS WORKFLOWS

Massively Scalable • Highest Performance & Density • Collaboration Centric

According to a recent Gartner survey of Pharmaceutical CIO's, the number one concern cited is datacenter sprawl. This is the unruly decentralization of IT resources across the organization followed by the lack of a comprehensive collaboration strategy.

In the past decade, it was not surprising for each research department to source, build and maintain a custom HPC infrastructure, optimized for a specific in silico simulation study. Hence, having custom HPC siloes for Protein-ligand docking programs like AutoDock[®] and fast shape comparison application like Openeye ROCS[®] was not uncommon. The researchers running these "islands of compute" believed they had dramatically improved their productivity, because the compute resource was temporally close to where the research was occurring. This allowed them to conduct compute intensive experiments, while maintaining total control over the hardware, the application and post run analysis, with minimal corporate intervention. However, there were problems with this approach:

- Research results were only meaningful to the originating group;
- Redundant and overlapping research efforts occurred due to lack of collaboration
- Redundant IT personnel were required to maintaining and up-keeping the decentralized equipment
- Data silos were loosely managed causing incompatible datasets and poor data retention

The noted "islands of compute" strategy is slowly giving way to a new paradigm of datacenter consolidation, fueled, by next generation networking, compute and high performance storage. National Labs and Health Research organizations worldwide have shown that a centralized approach can accelerate all facets of the discovery pipeline, while promoting a culture of collaboration. Pharmaceutical and Biopharmaceutical companies who have already made this transition are benefiting from accelerated times to discover new chemical entities (NCE), new molecular entities (NME) and beneficial compounds - all while fostering a culture of collaboration.

"Preliminary benchmarking shows that performance of the [DataDirect Networks] SFA10K is nominally the same as DAS (local disk) on our compute nodes. Hence, the shared and scalable file-system offered by GRIDScaler[®] [appliance] is a big win in terms of not having to constantly move data around on our network." - Weill-Cornell

As the datacenter consolidates, Systems Architects, Administrators and CIOs are engaging industry experts and OEM's for innovative solutions that not only scale and support hundreds of concurrent clients, but can:

- **Service mixed workloads** - billions of small file or hundreds of terabyte files - concurrently
- **Allow seamless connection of thousands of clients** - Linux[®] and/or Windows[®] based instruments
- **Support charge-back models** - at a department or even user level
- **Independently scale performance and/or capacity** - with minimal cost impact and no downtime
- **Support the federation of immutable data** - across vast geography's and multiple user sites
- **Address electronic-records/electronic-signature (ER/ES) mandates** - by leveraging native metadata and/or file system features.

Since 1998, DataDirect Networks has been recognized by government labs, NASA, The DoE and many world class genomics research facilities, such as: Wellcome Trust Sanger, TGen, and Cornell as the go-to visionary-partner and trusted-advisor for delivering best-in-class, high performance storage solutions. To put that statement in context, DDN is the number one supplier of parallel file solutions to the genomics research community. DDN subject matter experts understand how to accelerate the drug discovery process while mitigating day-to-day risk and they back that up with world-class utilities, support and Enterprise class 24/7 service.

Over the years, DDN has developed strong partnerships with IBM[®], Hewlett Packard[®], Dell[®], Silicon Graphics[®] and other key OEMs to deliver to the pharmaceutical industry an integrated single-vendor solution. We continue to work with our OEM partners to deliver optimized storage solutions that address the varied and demanding needs of the Life Sciences industry. Our storage solutions improve productivity, collaboration and scale to match your growth projections - with predictable performance. DDN can consolidate your storage silos and simplify data-management, all while lowering total cost of ownership.

BUSINESS ADVANTAGES

Industry Leading Value and Performance: The consolidation of the datacenter will require new networks, larger-faster computers and a new commercial-grade HPC storage infrastructure. These new consolidated compute resources are expected to help solve core workflows and business problems including:

- Accelerating large scale in silico simulation studies which will narrow the field of high efficacy NCE/NME and biosimilars
- Facilitate collaboration to share digital intellectual property with all departments, worldwide
- Narrow candidates for In Vitro high-throughput screening

The new integrated digital infrastructure is expected to do all this reliably, at-scale and more efficiently than the “islands of compute” approach. DDN addresses this new centralized paradigm with a family of high performance storage solutions that deliver:

- Proven solutions that are reliable at scale
- Sustained high throughput of digital content to thousands of concurrent users
- Integrated collaboration
- A dense, cost effective storage platform

Reliability at-scale: Pharmaceutical companies have petabytes of legacy NAS and SAN storage platforms, which are not used for in-silico simulation, due to their unpredictable performance (giving rise to “islands of compute”). Embracing a centralized approach for scaling capacity or performance means selecting either scale-up or scale-out architectures. This often results in over provisioning, which is both costly and inefficient at-scale. These two architectural approaches result in unnecessary overhead, due to managing either multiple islands of storage or inefficient budgetary spending, all resulting in lower returns on investment, higher total cost of ownership and environmental complexity.

DDN offers a powerfully, efficient alternative to legacy scale-up or scale-out approaches by providing the ability to scale performance and capacity independently. DDN Storage Fusion Architecture® is uniquely suited to adapt to pharmaceutical research compute environments and the unique data storage challenges they present. The SFA Family performs at the highest levels of both sequential and random IOPS. Additionally, SFA enables In-Storage Processing™ for embedding clustered/parallel file-systems directly into the storage – resulting in significant reductions in complexity, latency, and datacenter footprint - delivering best-in-class value.

Sustained viability: DDN storage solutions are designed to not only scale but to accommodate new and emerging media. Our storage systems were engineered to support current and future SATA, SAS and solid-state drives (SSDs) within the same system. Hence, as drives evolve in density and performance, the system administrator can upgrade a legacy DDN system with the latest 4TB SATA drives. Our controllers are also engineered to evolve. It is possible for our legacy customers to swap-out previous generation DDN controllers with current state-of-the-art units with little to no impact on the datacenter – other than a 2x improvement in I/O and bandwidth. Intra generation upgradability and massive overall scalability deliver superior investment protection vs. traditional storage architectures.

Charge back models: Maintaining a detailed record of asset usage is no longer an option, but a necessity for Pharmaceutical companies. In order to gain insight to development costs for the drug pipeline, a detailed accounting of asset usage must be kept. As stated earlier, DDN storage appliances support embedded applications like file-systems, as well as third-party application like iRODS® which can track a coarse (department) to fine (client level) usage model for charge back.

Performance at-scale: Today, DDN is ranked as the #2 Strategic Big Data HW Supplier by CIOs for scalable, dense, powerful (fast) storage solutions. Our Storage Fusion Architecture (SFA™) powers some of the largest pharmaceutical companies in the world. DDN technology is now accelerating drug discovery and synthetic biology pipelines, including: genetic sequencing, bioinformatics, proteomics, computational biology and the development of new molecular entities (NME). We do this at-scale, more efficiently and at a substantially lower price-point than legacy clustered NAS platforms.

Collaboration: Having local and multiple copies of critical data accelerates research and protects digital intellectual property. Collaboration is the new paradigm for Pharmaceutical researchers, which has given rise to applications like SFTP, cloud sharing via Dropbox®, Amazon S3® and even shipping of data loaded drives. There are many downsides to these approaches, such as: access time to data, security, compliance, price, backup, and system firewall concerns.

DDN has developed a family of collaboration solutions that can be integrated into our SFA appliances or as a stand-alone appliance, that resides behind your firewall. Our policy based collaboration solutions simplify the federation of immutable objects (files) down to a file copy.



OUR SOLUTIONS



STORAGE FUSION ARCHITECTURE

As an innovator in massively scalable storage platforms, we provide a number of solutions that increase performance, efficiency and data protection. Designed for multiple classes of structured & unstructured data as well as performance requirements, this unique multi-RAID architecture combines up to 1,680 SATA, SAS and/or SSD drives into a simply managed, multi-petabyte architecture that occupies only two datacenter tiles. When you combine SFA with GRIDScaler® it's possible to have your data reside on the optimal disk technology according to priority and access pattern, while selecting the appropriate RAID level to maximize performance and cost-effectiveness. This single feature alone will accelerate disk intensive, structure-based drug design algorithms like molecule docking, virtual screening (HTS); alignment searches and related algorithms. DDN systems deliver world class QoS through parity generation, real-time data integrity verification and error correction, without impacting application performance. In short, DDN Storage Fusion Architecture is reliable, robust and scales to meet the needs of the Biopharmaceutical workload.

Performance: The flagship SFA12K™ storage engine is almost eight times faster than legacy enterprise storage. With SFA12K, you can consolidate, simplify and leverage industry leading SFA storage performance to satisfy growing IOPS & Bandwidth requirements, with the fewest number of systems. A single system delivers up to 40GB/second of system bandwidth and bandwidth scales with each additional SFA12K system. It's possible to achieve an aggregate bandwidth of 1TB/second in just 25 systems.

Density: SFA is 40% denser than any other storage system. Reclaim your datacenter, resolve space and power limitations with the industry's densest storage platform. Each 4U enclosure houses 84 drives supporting up to 3.36 petabytes¹ of raw storage per rack.

Lowest Total Cost of Ownership (TCO) in the Industry: 50% Lower TCO than any legacy enterprise storage solution. SFA is a smarter choice with petascale storage support. Choosing SFA12K provides your datacenter with industry leading performance, capacity, density and DDN's unique In-Storage Processing™ technology - SFA12K-E, today supporting two parallel-file systems and iRODS®.

In-Storage Processing™ The SFA Operating System utilizes Open Source KVM virtualization software and the SFA OS hypervisor to enable applications to run inside the storage device. SFA12K can support up to twelve virtual machines. Today, DDN has ported two unique file-systems which result in our GRIDScaler and EXAScaler® embedded appliances and iRODS for automation, logging and building a next generation data management cyber-infrastructure.



PARALLEL FILE SYSTEM SOLUTIONS

GRIDScaler®: This powerful parallel storage solution combined with our SFA, supports LAN, SAN and NAS including NFS/CIFS and can ingest data from up to 2,000 clients. GRIDScaler's advanced data protection suite and replication features ensure data integrity and high availability for the life of your digital assets. With GRIDScaler, high performance applications like aligners and search algorithms, as well as high capacity applications like de novo aligners, are supported and accelerated on the same platform and within a common set of files and storage. This performance sharing mechanism allows your IT department to deploy fewer systems and reduce up-front acquisition costs and TCO.

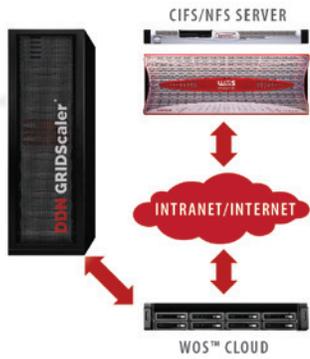
EXAScaler®: This highly scalable turnkey solution that provides Linux-based applications near wire-speed data transfer capabilities and unbounded capacity. Built upon best-of-breed open source file system technology and industry leading SFA Storage, DDN has carefully engineered this turnkey solution to reduce the complexity of deploying high performance computing storage and to scale with your evolving application requirements.



COLLABORATION

WOS® (Web Object Scaler) nodes are self-contained appliances, configured with disk storage, CPU and memory resources. Each node is preconfigured with the WOS software and Gigabit Ethernet network interfaces. A private peer-to-peer cloud architecture can be created out of any node that has Internet Protocol (IP) connectivity up to 64 other nodes, regardless of their physical location.

Another variation is WOS Bridge™, designed to enable collaboration for GRIDScaler appliances. Functionally, WOS Bridge is a 1U server that enables a cloud tier interface for GRIDScaler appliances, either to offload data to WOS or to enable collaboration between GRIDScaler and distributed WOS Access applications and/or users. As with WOS, data is replicated from GRIDScaler to WOS on a peer-to-peer basis. All content remains completely under your control, secure and behind your firewall.



- Integrate with WOS Access from GRIDScaler
- Collaborate between geographically distributed teams
- Seamlessly migrate data from GRIDScaler to a private cloud environment on WOS
- Aligned genomics data on GRIDScaler is replicate to remote sites via WOS Bridge
- Genomic files are safely backed-up to your WOS cloud for disaster recovery

DATA ANALYTICS

Data analytic applications from SAS®, SAP® and Vertica™ have assured pharmaceutical companies that the information extracted was relevant, accurate and consistent. Today, SAS running on GRIDScaler has yielded the fastest runtime ever recorded for SAS Grid™.

“IO intensive SAS Grid workloads have demonstrated excellent performance characteristics utilizing this storage appliance. The choice of shared file system and storage is a critical component of high-performance SAS Grid deployments.” - SAS

The art of data analytics provides a new way to access, manage and analyze massive datasets more efficiently. Pharmaceutical companies, through data analytics, can transform biomedical data into clinical insights - or analyze the supply chain to manage risk or convert digital intellectual property from NME into targeted drugs, making drug trials even more successful.

ABOUT DDN®

DataDirect Networks (DDN) is the world's leading big data storage supplier to data-intensive, global organizations. For more than 15 years, DDN has designed, developed, deployed and optimized systems, software and solutions that enable enterprises, service providers, universities and government agencies to generate more value and to accelerate time to insight from their data and information, on premise and in the cloud. Organizations leverage the power of DDN technology and the deep technical expertise of its team to capture, store, process, analyze, collaborate and distribute data, information and content at largest scale in the most efficient, reliable and cost effective manner. DDN customers include many of the world's leading financial services firms and banks, healthcare and life science organizations, manufacturing and energy companies, government and research facilities, and web and cloud service providers. For more information, visit our website www.ddn.com or call 1-800-837-2298.