DDN®
STORAGE

INDUSTRY SOLUTION BRIEF

ACCELERATING ACADEMIA

CUTTING-EDGE RESEARCH INITIATIVES BENEFIT FROM HIGH-PERFORMANCE, SCALABLE AND COLLABORATIVE STORAGE SOLUTIONS FROM DDN

ACADEMIC RESEARCH

THE BIG DATA CHALLENGE AND OPPORTUNITY IN ACADEMIA

There may be no field where Big Data is bigger than in academia, particularly academic research institutions' scientific research programs.

Innovative research in applications such as genomics, molecular biology, bioinformatics, disease discovery and treatments, fluid dynamics, marine science, particle physics, computational biology and others is changing the way scientific problems are discovered and solutions developed. But those research advances are a double-edged sword, because along with the exciting breakthroughs come unexpectedly complex demands for greater levels of increasingly sophisticated IT infrastructure to match the Big Data reality of academic research.

How big is the Big Data challenge in academia? Big enough that thinking of data volumes in the multiple hundreds of terabytes is astonishingly conservative. In data centers housing multiple supercomputers to get sufficient processing power to do the kinds of sophisticated modeling and mining needed for breakthrough applications, new ways of thinking have had to evolve to keep up with the Big Data challenge.

One approach is to address the challenge through the development of a new organization to study new ways of collecting and sharing data. The National Science Foundation (NSF) recently proposed the creation of a new organization – the Computation Institute – to examine how technology can make the research process more efficient. The NSF pointed out that new ways of thinking are necessary in order to get the most out of traditional computational methods and IT infrastructure in order to fully mine and share critical data throughout the research process.

But as important as new thinking for leveraging technology may be, another key approach has evolved that may, in fact, be even more directly associated with major improvements in generating better results in academic research – smarter, faster and more efficient storage infrastructure. When a single research project, taking place with even a handful of researchers, requires more storage than multinational corporations required less than a decade ago, it's clear that tremendous improvements in storage for high-performance computing are essential.

ENTER DATADIRECT NETWORKS.

DDN is more than a leading developer of innovative storage hardware for high-performance computing applications such as academic research. It understands the essential elements of software, processor architecture, clustering, high-speed networking and collaborative learning that shape the landscape in academic research.

DDN powers over 60% of the Top100 fastest supercomputers, many of which reside inside the world's most prominent academic research institutions. DDN is a leader in the field of open source file systems, file/block storage appliances and cloud-based storage, and has brought a keen design eye toward such important issues as preserving data center physical space and controlling power consumption problems normally associated with Big Data applications in complex data centers.
Not only that, but DDN understands that there is an essential four-part process that mirrors how
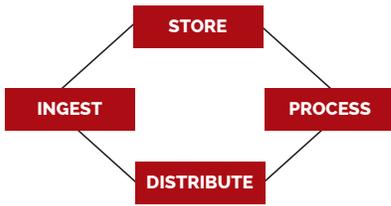
academic research projects work, structuring its solution set accordingly. First, researchers must ingest huge data volumes, often made up of complex data structures and pulled from a myriad of data sources – both at the primary research facility and increasingly from research activities taking place all over the world. Applications like genomics and particle physical introduce massive amounts of data, which must be brought into the IT environment and the extremely large data sets must be mined to extract key data. The emphasis on the ingest phase is the speed of getting the data in and out of the system.

Second, the store and process phases requires all that data to be stored and processed in a way that ensures instantaneous, always-available access – and to do so with massive scalability, strict data protection and cost efficiency.

Finally, the distribute phase of the process may be the most critical one for academic researchers, since research today is intensely collaborative. Data is not always created solely in a single research site, but often in concert with colleagues in other sites – often at other research institutions. Additionally, researchers increasingly are working on projects with small-form-factor devices - from desktops and notebooks to tablets and smart phones. Academic research has been made more powerful and impactful with the ability to share data among multiple research centers and participants, which means IT infrastructure must facilitate data sharing among all nodes in the research lifecycle.

Ingest. Store. Process. Distribute. It's part of the DNA – literally – of DDN's high-speed storage solutions for academia.

THE DDN SOLUTIONS FOR ACADEMIC RESEARCH REQUIREMENTS

DDN customers benefit by working with a supplier that combines sophisticated technology with an in-depth understanding of the diverse requirements in academic research. Additionally, DDN offers university research organizations a number of powerful, yet affordable, solutions for different aspects of high-performance storage.

For the **Ingest, Store and Processing** requirements of massively scalable storage, DDN delivers Open Storage Appliances that integrate leading HPC file systems within the storage platform to increase performance, reduce costs and simplify deployment.

 EXAScaler™ is an open file storage appliance, specifically designed for applications requiring massive throughput and highly parallel access to huge data sets. This massive data growth in applications like life sciences, bioinformatics and genomics is literally choking legacy solutions based upon Storage Area Networks (SAN) and Network Attached Storage (NAS) technologies. Many of those SAN and NAS approaches were designed more than 10 years ago, meaning they are no longer sufficient for the kinds of high-performance, massively scalable requirements for academic research environment.

EXAScaler integrates the file system and storage media into a single, hyper-performance appliance solution. This approach combines DDN's widely acclaimed high-performance computing storage technology with the open source Lustre™ File System. EXAScaler appliances are highly scalable to meet the ever-growing needs for higher capacity and faster performance of academic research organizations, and the appliance form factor speeds and facilitates deployment for even the most technically demanding environments, such as the largest number of global, Top500® Supercomputing Sites.

Customers enjoy the benefits of a system that comes pre-configured to support easy deployment of the Lustre file system. Storage hardware, RAID capabilities, operating system, data protection features and Lustre all are factory-integrated into a single platform, for simple, rapid deployment.

GRIDScaler™ is designed to handle the diverse and dynamic demands of today's massive unstructured data generators. It is an extremely versatile solution that can be configured as a parallel file system as well as a Network Attached Storage solutions based on the customer's requirements. With scalable data and metadata technology, the GRIDScaler platform eliminates all bottlenecks to achieve true parallelism and maximum application performance.

GRIDScaler allows for seamless and efficient on-demand expansion by either adding GRIDScaler filer server nodes to easily scale performance or scaling capacity, by adding high-density disk enclosures. It also supports online provisioning that shrinks or grows file systems, as necessary, in order to avoid over-provisioning and unnecessary cost.

It supports intelligent parallel file system clients for both Windows® and Linux® environments and is designed for highly concurrent file/directory access for massively parallel requirements. Striping data across multiple enclosures, then accessing them in parallel, dramatically accelerates application performance, making GRIDScaler well suited for such applications as molecular simulation, fluid dynamics, genomics and computational biology.

For the critical **Distribute** portion of high-performance computing storage, DDN offers a powerful collaboration platform:

WOS® is a turnkey appliance for massively parallel, massively scalable storage built upon a foundation of cloud and object storage. For Big Data applications requiring limitless scalability, secure collaboration and unprecedented throughput, WOS is an ideal solution. Its architecture is designed specifically to encourage safe, secure collaboration among research groups, and its performance enables the kind of performance necessary to support huge data sets, fast throughput and high-integrity content distribution. Since WOS is a true object storage solution, from API to the disk head, with no underlying file system to impact performance, it can achieve the high performance necessary in applications such as life sciences, computer science and bioinformatics.

WOS is an ideal approach for applications requiring fi le store, share and transfer; global collaboration, and content distribution for academic research organizations that are using cloud architecture to improve agility, enhance collaboration and achieve unprecedented cost efficiency. Secure collaboration is promoted by the use of behind-the-firewall security features, as well as easy file access to familiar Web browsers and even by mobile device clients.This means that students and professors not necessarily tethered to a physical data center to contribute, in real time, to the research efforts.

DDN offers three classes of solutions based on the WOS architecture. WOS Access allows distributed users to collaborate at local speeds, using familiar NAS protocols such as NFS and CIFS. It synchronizes gateways across multiple, geographically distributed sites to function as a single federated namespace. For research organizations that act as their own cloud service provider as a private network, WOS Cloud lets them host multiple tenants and supports a number of collaboration features, including commenting and versioning. WOS Share is an ideal solution for students and other individual contributors; it is a personal storage system delivering enterprise-class performance, and lets them securely store, share, and access files, anytime and anywhere from multiple devices without violating IT policies.

HOW UNIVERSITY RESEARCH INITIATIVES BENEFIT: REAL-WORLD EXAMPLES

Data growth was a huge challenge for the Tokyo Institute of Technology, the largest academic research institute in Japan dedicated to technological and scientific research. Increased project loads drive huge volumes of data growth, and the need to collaborate reliably and in real-time creates a potential
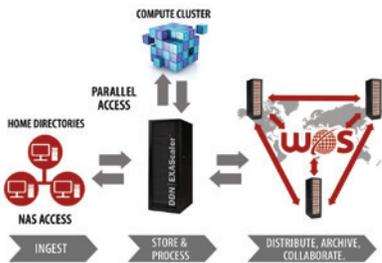
bottleneck. The institute needed to store high-resolution microscopy, genomic data, molecular and macroscopic biological imagery and other biologic data, as well as supporting materials for meetings and presentations. Additionally, the institute's academy of computational life sciences wanted a dedicated cloud storage service to provide secure file sharing and storage for its researchers and students. DDN configured a solution based on its WOS cloud, utilizing three WOS servers and three high-performance storage systems. Files are now uploaded by teachers and students from their PCs and personal devices to the cloud storage through the WOS cloud client; about 600 users access the WOS cloud, and all researchers at disparate sites have real-time data access, facilitating collaboration and discovery.

Looking to leverage sophisticated research in such areas as low-level physics analysis, molecular simulations, weather modeling and fluid dynamics, the University of South Florida needed higher-performance, affordable and flexible storage solutions to handle the huge growth in data volumes and complexity in their laboratories. USF turned to DDN for a significant upgrade on the university's legacy parallel file system environment, which was straining badly under the weight of massive data growth. Additionally, data loss was becoming a serious problem for their research team – clearly an unacceptable situation. USF selected GRIDScaler which allowed them to consolidate storage assets, improve throughput, increase resiliency and support rolling software upgrades for maximum uptime.

Wellcome Trust Sanger Institute, a charitably funded genomic research center in Hinxton, U.K., selected DDN technology to help power its groundbreaking genetics research. Sanger relies on EXAScaler for its Wellcome Trust Sanger UK10K project, which aims to sequence the DNA of 10,000 individuals.

EXAScaler is uniquely suited to scale-out performance, meeting the challenges of sequence alignment and bioinformatics processing, while also scaling-up to cost & space-efficiently manage petabytes of genomic research data. The DDN solution will also support the arrival of next-generation Illumina® gene sequencing systems at The Wellcome Trust Sanger. Its performance, scalability and ease of management keep up with strain demanding research initiatives place on IT infrastructure.

At the Texas Advanced Computing Center (TACC) – located at the University of Texas – large-scale simulations are a way of life. But when Big Data demands, driven by one of the world's fastest supercomputer clusters, became too big of a drain on the center's computing resources, TACC turned to DDN. A data repository, delivering petascale-class performance, was needed to handle seemingly unimaginable levels of computing and storage performance for data-driven scientific research. TACC uses SFA™ products to complement the center's existing infrastructure, which is a critical tool in helping researchers mine important data volumes and glean key insights from systems already in place.

SUMMARY

For academic research organizations, fields such as genomics, medical research, molecular biology and other applications hold tremendous potential to change the world – but only if they can successfully access, mine and share the mountains of data being created. Teachers and their students need a new level of IT infrastructure that provide storage capacity, performance and access that matches the raw processing power of the supercomputer typically on their sites. More and more often, those in the field of academia are turning to DDN's cutting-edge storage solutions – such as EXAScaler, GRIDScaler and WOS – to tame and fully leverage this seemingly runaway Big Data phenomenon, and turn it into a strategic asset supporting breakthroughs that improve the lives of countless people around the world. Throughout the scientific discovery process, DDN solutions are used to ingest, store and distribute previously unheard-of volumes of raw data.

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.

DDN® STORAGE

BOTTOM LINE RESULTS.
▰▰▰▰▰▰▶ ACCELERATED.

SALES@DDN.COM | +1.800.837.2298