

CHALLENGE

The Texas Advanced Computing Center (TACC) required a high-speed, scalable research data repository that was friendly to both their computer center and their budget. The solution needed to not only meet current needs, but also be able to scale for the next generation of scientific research.

SOLUTION

Together with DataDirect Networks, Dell provided a highly scalable turnkey system using Dell PowerEdge servers with Intel® Xeon® processors to support the sophisticated TACC compute and visualization systems.

BENEFITS

- Massively scalable storage provides 1.2 PB of capacity, with a file system that can potentially support up to 10 PB.
- High performance, including up to 6 GB/sec of aggregate file system performance and over 1 GB/sec of per-client performance, helps support supercomputing sites worldwide across a wide range of scientific disciplines.
- Rack-dense, energy-efficient solution helps minimize physical footprint and power consumption.

THE SCIENCE OF SCALABILITY

The Texas Advanced Computing Center, a leading scientific computation facility at The University of Texas at Austin, tackles the research data explosion using high-performance Dell™ PowerEdge™ servers and massively scalable storage from DataDirect Networks.

Outfitted with some of the most powerful computation systems in the open science world, the Texas Advanced Computing Center (TACC) at The University of Texas at Austin is helping lead a class of research institutions that champion computer simulation as a cornerstone of the scientific discovery process—an element on par with theory and experimentation. But large-scale simulations and data analysis tools have challenged TACC with controlling the scientific data explosion. To help prepare for the next generation of scientific data resolution, TACC turned to Dell and DataDirect Networks to provide a scalable, cost-effective research storage solution designed for the petabyte era.

BUILDING MASSIVELY SCALABLE STORAGE

Because TACC runs several of the world's largest supercomputers, including the Ranger system ranked as the world's eighth-fastest clustered computer on the June 2009 TOP500 Supercomputing Sites list,¹ the need for a scalable data repository that could deliver petascale performance was critical when selecting storage for the project. Dubbed Corral, the TACC platform incorporates eight Dell PowerEdge 1950 servers, each configured with one quad-core Intel Xeon processor at 2.3 GHz, 16 GB of RAM, and Community Enterprise Operating System (CentOS) 5.3; eight Dell PowerEdge 2950 servers, each configured with two quad-core Intel Xeon processors at speeds ranging from 2.3 GHz to 3.2 GHz, 16–32 GB of RAM, and CentOS 5.3; and 1.2 PB of storage using the S2A9900 storage system from DataDirect Networks. The massive platform—four times larger than any other current data-collection resource on the U.S. National Science Foundation's distributed research network, the TeraGrid—is designed to effortlessly handle the challenges and opportunities of data-driven science.

Chris T. Jordan, a senior operating systems specialist in the Advanced Systems Group at TACC, notes that Corral complements the TACC system portfolio by enabling users to gain additional insights from the systems already in place. For example, users can access all of the Corral storage capabilities from the Ranger or

¹For more information, visit www.top500.org.

“Dell and DataDirect Networks have always been a good choice for us to deploy and manage very scalable data sets. The solution delivered by Dell is designed to grow as we grow—and the price is right to get the most out of our budget.”

—Chris T. Jordan
Senior operating systems specialist
in the Advanced Systems Group at TACC
June 2009

Lonestar high-performance computing (HPC) systems and from the Spur or Stallion visualization systems.

Able to support over 1,200 hard drives behind a single storage system, the S2A9900 storage used in Corral is well suited for scale-out file systems. The open source Lustre file system uses the Dell PowerEdge servers as file system gateways and aggregates the disks in the one-to-many S2A9900 systems to deliver high-speed parallel file access to TACC researchers and supercomputing systems. This solution can scale to over 10 PB of capacity in a single file system to help satisfy the demands of the world's largest supercomputing systems. “Dell and DataDirect Networks have always been a good choice for us to deploy and manage very scalable data sets,” says Jordan. “The solution delivered by Dell is designed to grow as we grow—and the price is right to get the most out of our budget.”

OPTIMIZING PERFORMANCE FOR DATA-DRIVEN SCIENCE

The Corral solution is highly optimized to help serve the diverse performance requirements of many research disciplines, including climate research, high-energy physics, space and cosmological research, disease studies, and more. It is designed

to deliver up to 6 GB/sec of aggregate file system performance and over 1 GB/sec of per-client performance for demanding single-stream workloads. Additional storage and gateway building blocks can be added to the Lustre namespace to increase capacity, performance, or both.

More than just a storage repository, the Corral platform enables TACC to compute and visualize data in the system. It is used worldwide by supercomputing facilities requiring greater performance than what is available from traditional network attached storage systems. DataDirect Networks also packages high-performance Dell PowerEdge servers with its award-winning storage platforms and the Lustre file system in its ExaScaler™ parallel file storage system.


MAXIMIZING SPACE AND ENERGY EFFICIENCY

Because large data sets often lead to large data storage complexities, DataDirect Networks has developed a series of advanced capabilities to help minimize the impact of petabyte-scale storage. With its innovative 4U, 60-drive storage enclosure, the S2A9900 is designed to store up to 600 hard drives in a single data center rack. And by using 2 TB Serial ATA (SATA) drives, a single S2A9900

system can store up to 2.4 PB total, and as much as 1.2 PB in a single rack.

The S2A9900 also features high levels of energy efficiency. The high-density, 60-drive storage enclosure is designed to significantly reduce the number of fans, power supplies, and I/O modules compared with systems that use low-density storage enclosures. This reduction in components also helps reduce overall power consumption—which in turn helps both reduce costs and support environmentally conscious IT practices.

DESIGNING FOR THE FUTURE

The high-performance solution from Dell and DataDirect Networks was delivered and running in less than two weeks, and researchers are now enjoying the benefits of a massive storage repository that helps provide high-performance, cost-effective access to scientific data. “Corral provides online storage at the petabyte scale—it's all online, accessible, and high-speed so that researchers can store and use much more data as part of their computation or visualization,” says Jordan. Dell and DataDirect Networks have implemented a system at TACC that can grow in nearly any direction to help meet its research needs both now and in the future—enabling TACC to corral the next generation of scientific discovery data. 

MORE

ONLINE

DELL.COM/PowerSolutions

QUICK LINKS

Dell PowerEdge servers:
DELL.COM/PowerEdge

DataDirect Networks:
www.ddn.com