University at Buffalo, Center for Computational Research

EMC, Dell, and IBRIX team to deliver a high-performance computing solution to enable scientific research and discovery

In the seven years since its inception, the University at Buffalo's Center for Computational Research (CCR) has emerged as one the world's foremost supercomputing research facilities in computational science—an emerging discipline that unites computer science and mathematics with research in biology, chemistry, physics, and other applied and engineering fields.

Well known for high-performance computing and high-end visualization, its research activities encompass everything from computational chemistry, solid-state physics, and life sciences to earthquake engineering, computational fluid dynamics, data mining, and virtual reality simulations.

Today, nearly 140 research groups covering nearly 40 different departments across the campus, as well as a few dozen companies and institutions in the Buffalo-Niagara region, rely on the CCR to provide the human and supercomputing resources necessary to further their work and enable discovery. CCR bioinformatics researchers also conduct extensive genomics research including DNA and protein sequence analysis, database search, gene expression analysis, biological pathway analysis, statistical computing, and inferring phylogenies.

The CCR recently deployed a state-of-the-art, high-performance computing (HPC) solution to cost-efficiently address a steadily rising demand for computational resources including sequencing, assembling and classifying, and annotating. This innovative compute farm and clustered storage solution brings together products and expertise from industry leaders EMC, Dell, IBRIX, Myricom, and Force10 to provide the power and linear scalability necessary to support the center's computationally intensive applications and expanding user community. The development of this cyberinfrastructure enables the efficient collection, management, organization, analysis, and visualization of data.

> "EMC, Dell, and IBRIX provided a single system that gives us the ability to enable discovery in a wide sphere of influence. No single system is going to be optimal for all scientific problems, but it is our belief that this system is meeting the vast majority of our users' needs in a very efficient and very effective fashion."
>
> **Dr. Russ Miller**
> **Director, Center for Computational Research**

"Given the user demand and funds available for this project, we are convinced that we have made an intelligent choice in the acquisition of a significant resource," says Dr. Russ Miller, Director of the Center for Computational Research and UB Distinguished Professor of Computer Science and Engineering. "We now have a scalable high performance computing and data solution that our computational scientists and system administrators can operate efficiently; the systems works can be utilized and maintained in a straightforward fashion and has been seamlessly integrated into our grid environment."

Optimizing supercomputing capabilities with an integrated storage solution

A more cost-efficient and easily scaled alternative to expensive supercomputers, CCR chose to invest in an HPC cluster containing 800 new Dell PowerEdge servers with two backend Gigabit Ethernet and bidirectional connects from Force10 and Myricom. Able to meet the demands of CCR's high-performance computing applications, this system can deliver up to 10 TFLOPs of computing power with sustained performance anticipated at around 7 TFLOPS.

CCR's new EMC® SAN consists of three high-performance, highly reliable EMC CLARiiON® CX series storage arrays which provide a total capacity of 30 terabytes and are managed by EMC Navisphere® storage software. High performance file sharing is delivered through IBRIX Fusion™, an advanced parallel file system software product. This combined hardware and software solution provides CCR's HPC environment with a centralized, easily managed, and scaled, single-image pool of storage that allows different applications to quickly access the same data.

Combining the best of both worlds, CCR's integrated storage systems are as easy to monitor and manage as a single large array, yet they provide a greater level of redundancy to ensure non-stop operations should a node fail. In addition, this configuration provides the flexibility to quickly add storage as required without reconfiguring the environment or disrupting applications with downtime.

SAN-friendly IBRIX segment servers enhance the solution by dramatically expanding connectivity into and out of the EMC storage platforms to 48 connections. Within this highly optimized and integrated HPC environment, exceptional read rates of 2.3 gigabytes per second and about 1.1 gigabytes per second for data writes have been achieved. High I/O throughput means applications can run faster for better performance. It also means fewer bottlenecks due to idle processors waiting for I/O to complete as well as better productivity because more data can be processed in less time. These improvements are all vitally important to this compute and data-intensive research facility.

"Combining the IBRIX software solution with EMC storage platforms has provided us with a much more efficient way of getting data into and out of the compute system than any of our other systems," says Dr. Miller.

Collective partner expertise has streamlined deployment and laid the groundwork for future research expansion including parallel algorithms, grid computing, image processing, and computational crystallography. Collaborating on the design and implementation of CCR's comprehensive new HPC environment, Dell, EMC, IBRIX, Myricom, and Force10 have been credited with facilitating a more rapid rollout and, as a result, a faster return on investment. Daily meetings with the project coordinator and regular contact with EMC and IBRIX technicians helped ensure that installation timelines were met.

"This is a large and complex system and the installation went very smoothly," says Miller. "As we continue to do work in highly data-intensive areas we certainly expect the demand to grow significantly. The way we've architected this system, there are no constraints or restrictions in terms of expansion."

Meeting diverse research needs with the right combination of compute and storage resources

Ranging from bioinformatics, structural genomics, computational chemistry and fluid dynamics to medical and scientific imaging, data mining, and database development—virtually all the work supported by the center now takes advantage of the new high-performance cluster and EMC integrated storage system. This dynamic environment can support the needs of a wide variety of users—some of whom have capability computing requirements; others who have capacity computing needs; some whose work is very I/O-driven and data-intensive; and others whose compute requirements fall somewhere in between.

"EMC, Dell, and IBRIX provided a single system that gives us the ability to enable discovery in a wide sphere of influence," says Miller. "No single system is going to be optimal for all scientific problems, but it is our belief that this system is meeting the vast majority of our users' needs in a very efficient and very effective fashion."

Since deployment, user feedback has been very positive, both from researchers inside the university as well as some of CCR's external collaborators. The level of performance, which has shown a sustained threefold to fourfold increase over

CCR's next most powerful machine, as well as the amount of data that users can work with and store have been cited as key benefits.

On the operational side, system administrators have found the system to be a reliable, scalable, and straightforward to manage next generation machine.

"The operation, support, and utilization of the system is moving along very smoothly. Our users and system administrators are pleased, price/performance is excellent, and so is the return on investment. I would recommend this solution without hesitation," says Dr. Miller.