Case Study

High Performance Cloud Computing
3rd Generation Intel® Xeon Scalable Processors

University at Buffalo Enables Western New York Businesses to Innovate with Dedicated HPC Cluster

The Center for Computational Research's Industry Compute Cluster delivers advanced computing capabilities to a wide range of industries running HPC + AI workloads

Solution Summary:

- Converged HPC + AI cluster supports Western New York commercial research
- 99 Dell EMC PowerEdge Servers with Intel® Xeon® Gold 6330 processors
- Large memory footprint for data-intensive computing
- PCIe Gen 4 supports GPU I/O for AI/ machine learning

University at Buffalo
Center for Computational Research

Executive Summary

The University at Buffalo's (UB) Center for Computational Research (CCR) offers unique opportunities to Western New York's many businesses. CCR provides dedicated converged high-performance computing (HPC) and Artificial Intelligence (AI) capabilities with a new HPC + AI compute cluster. The new cluster, built by Dell Technologies using Intel® Xeon® Gold 6330 processors, enables a large community of customers to develop innovative solutions through simulation, modeling, and machine learning.

Challenge

University-purchased HPC clusters are typically funded to support principal investigators and their teams of faculty, university students, and closely associated researchers. CCR created a unique program specifically designed to enable businesses of Western New York to take advantage of large-scale computing.

Before 2014, the university worked with municipalities and local animation companies that were interested in partnering with the university and its computing capacity. The interest led to the creation of the dedicated Industry Compute Cluster.

"We had these big computers and the ability to do renderings at an exponential pace compared to what companies could do themselves," Adrian Levesque, Head of Industrial Projects and Visualization Services, explained. "But we were limited by what we could offer because many of our resources were purchased with grants for specific types of university research."

Innovative medical technology companies in Western New York, such as Marion Surgical whose VR/AR technology is featured above, take advantage of the supercomputing capabilities at the University at Buffalo's Center for Computational Research (image courtesy Marion Surgical)

"And that put a seed in our thinking to find a way to provide these rich computing resources to industries in our region that could benefit from them," Levesque added.

Around 2014, CCR submitted a proposal to the Empire State Development Corporation, the state's economic development organization, with an opportunity to engage with New York businesses.

"The organization gave us an initial funding for a cluster for four years to test the waters and develop processes for engaging businesses," Dr. Matthew Jones, Director of CCR, said.

The initial Industry Compute Cluster comprised 216 Dell EMC PowerEdge server nodes with 16 cores each of Intel Xeon processor E5-2650v2 and 64 GB of memory, a total of 3,456 cores. From its beginnings, the program grew to support some 37 companies over the next three years, consisting mainly of small to medium businesses and a few international organizations that had a presence in Western New York.

The Industry Compute Cluster is dedicated to businesses with some preemptive nodes offered to faculty when they are not at full capacity. By 2020, with years of success and advancements in computing hardware and computational techniques, CCR approached the state again to upgrade the cluster.

Solution

CCR clients and businesses in the region were engaging in advanced computing methodologies, including machine learning, artificial intelligence, bioinformatics, genomics, large-scale simulations, and others. CCR architects approached the Empire State Development Corporation for an upgraded Industry Compute Cluster to address business needs. The result is a system built by Dell Technologies with 99 Dell EMC PowerEdge servers containing Intel Xeon Gold 6330 processors. Each node hosts 56 cores—a total of 5,544 cores. The system has 3.5 times more cores per node in less than half the number of nodes and eight times, or more, larger memory. To support machine learning and graphic rendering workloads, 16 nodes are built with dual GPUs. 83 nodes contain 512 GB of memory, while 16 support large-memory workloads with a terabyte of RAM. The new cluster went into production in August of 2021.

3rd Gen Intel® Xeon® Scalable Processors Enable Innovation

"One of the big use cases for the previous cluster was large-scale CFD applications," Jones said. "The customer specifically came to us because they did not have that type of memory footprint in house on their engineering systems. "

With scientific computing, greater computing capacity with newer technologies leads to demand for bigger simulations at finer resolutions. Thus, by 2020, the previous cluster memory capacity became an issue, which the new cluster solved.

"The new Intel Xeon processors allow a very large memory footprint to support modern CFD jobs," Jones added.

Graphic rendering continues to be an important workload on the cluster. And clients are using machine learning as they design new products and offer advanced analytics. 16 GPU nodes with the Intel Xeon Gold 6330 processors support these types of compute-intensive workloads. The 3rd Gen Intel Xeon Scalable processors with Intel® Advanced Vector Extensions 512 (Intel® AVX512) have twice the double-precision FLOPS compared to the older system with Intel AVX. The new system also integrates PCIe Gen 4, offering the fastest PCIe I/O available for the GPUs used in 16 of its nodes.[1]

"CPU scaling is much bigger. It's night and day compared to the older system," Jones concluded.

Supporting clients with advanced computing hardware, the University at Buffalo CCR also offers programmatic expertise. This includes helping companies get set up with startup programming to entrepreneur programming, and helping with codes specific to life sciences, and genomics and the types of data they deal with.

"We have a very rich group of computational scientists and software experts," Levesque explained. "They are experienced with Intel software tools for tuning and optimization."

"I think oneAPI will greatly streamline our work," Jones added.

"So, all these different touchpoints in terms of hardware, software capabilities, and tools build the assets that appeal to advanced manufacturing or advanced life sciences, whether it's medical devices or health informatics," Levesque concluded.

Result

With over seven years of experience working with industry, CCR has built a strong portfolio of diverse companies that draw upon its computing power. But they also continue to market to the New York business community that new resources are available. And the university continues to work directly with municipalities.

"We basically have a B2B business approach to let companies of Western New York know what the University at Buffalo can do for them," Tracy Krawczyk-Schiedel, Director of Marketing and Communications at UB Business and Entrepreneur Partnerships, said.

Managing Municipal Stormwater

Alan Rabideau, professor in the Department of Civil, Structural and Environmental Engineering in the School of Engineering and Applied Sciences, has a long history of using supercomputers to simulate and analyze municipal water systems.

"Traditionally, I've worked in groundwater modeling," explained Dr. Rabideau. "When we started the CCR, an early project was with the Environmental Protection Agency to perform large-scale groundwater modeling and calibration, plus many other things."

The EPA has its own Stormwater Management Model, (SWMM pronounced "swim"). It is used across the country by cities to model systems that meet the requirements of the Clean Water Act. SWMM is a complex model using partial differential equations and other calculations to account for topography, weather, piping networks, and other aspects of water runoff. Dr. Rabideau works with the Buffalo Sewer Authority to model, study, and design solutions to combined sewer overflows within the city of Buffalo.

"Our models are large and complex. We are looking at the traditional system of collecting runoff in pipe networks, but also at green infrastructure, such as rain gardens and porous pavements, which capture or divert runoff before it reaches the pipe systems. While piping network methods are well understood, green approaches still need a lot of study and data. We routinely run thousands of different simulations at high resolution. The new cluster will provide scalability to analyze even more scenarios in parallel."

Digital Biology Research

An important biotech/digital biology community resides in Western New York. One of these companies that has engaged with the CCR is Marion Surgical, a company that uses virtual reality (VR) and augmented reality (AR) to realistically teach surgeons complex procedures.

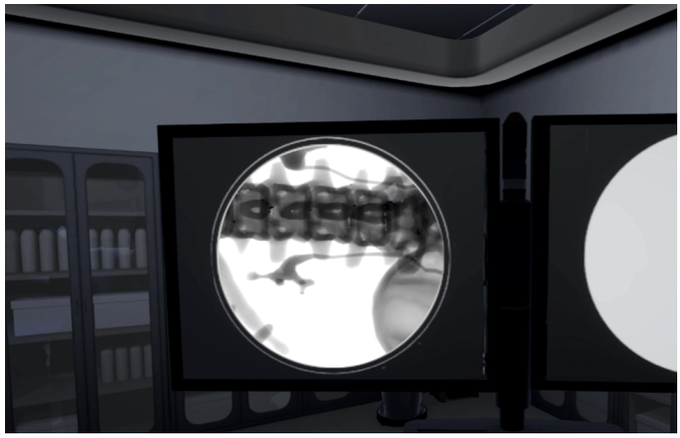
"Our system with VR goggles and a proprietary haptic feedback robot creates realistic and immersive kidney and abdominal surgery simulations," Ben Sainsbury, CEO of Marion Surgical, explained. "We provide an interactive environment to train clinicians on the actual feel of surgical devices as they penetrate different tissues, such as skin, fat, muscle, and different layers of kidney."

Marion Surgical engineers use the CCR cluster to build their simulations. They can use data from patient scans of the surgical targets and their surgical robot to create individualized simulations for each patient. Calculations use neural networks and visual simulation on the cluster's GPU and non-GPU PowerEdge server nodes. Doctors then can practice with the system, gaining unique experience, before the actual surgery.

"For example," Sainsbury added, "a patient with spina bifida presents a unique situation of how the surgeon can access the kidney. We can build a unique 3D model of the patient from individual CT scans and integrate the model with our simulation so the surgeon can both feel what he's doing while watching the surgical device move through the model."

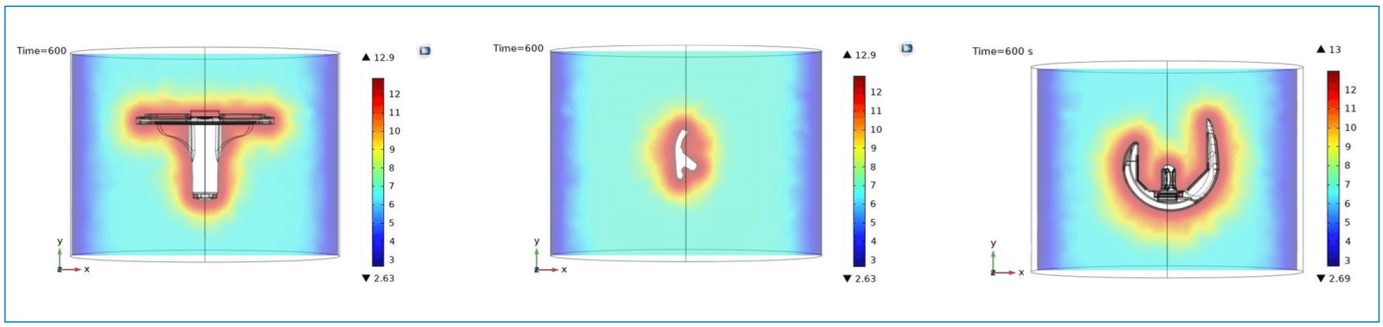
Simulating Treatment for Implants

Garwood Medical Devices, another customer, uses the CCR cluster to develop BioPrax*, a novel treatment for patients with metal implants. Bacterial biofilm infection is a common complication for implantable devices. Garwood's approach attacks infection around the implant using DC current to generate electrolysis that results in pH adjustments and agents that attack infection. Garwood is collaborating with the University at Buffalo's Mark Ehrensberger, associate professor of biomedical engineering and Priyanshu Vishnoi, doctoral candidate in chemical and biological engineering, to develop computational models that can simulate the treatment on the cluster with unique patient data and implant composition. They are working with the Food and Drug Administration and the Department of Defense to optimize their research for clinical application and develop more advanced use cases.



View through VR goggles (courtesy of Marion Surgical)

Garwood's BioPrax* device uses DC current to provide protection from infections associated with devices such as the femoral component (images courtesy Garwood Medical Devices)



The University at Buffalo's Industry Compute Cluster enables computational studies that simulate the BioPrax treatment of orthopedic implants (Images courtesy Priyanshu Vishnoi)

Additional Work

Bioinformatics and genomics work by local companies is also run on the CCR, supporting treatments that battle the SARS-CoV-2 virus and Covid-19.

Solution Summary

The University at Buffalo's Center for Computational Research engages with industry in Western New York, supporting innovative research and development with a dedicated compute cluster from Dell Technologies. The system is built on Intel Xeon Gold 6330 processors, meeting the needs of its customers with high performance, large memory capacity, and fast I/O. The CCR supports more than 500 customers statewide, who are innovating solutions across a wide range of industries.

Where to Get More Information

Learn more about the University at Buffalo.

Find out more about the CCR and its Industry Compute Cluster.

Explore the capabilities of the 3rd Generation Intel Xeon Scalable processors with integrated Intel® Deep Learning Boost capabilities for accelerated AI inferencing.

Solution Ingredients

- Converged HPC + AI cluster supports Western New York commercial research
- 99 Dell EMC PowerEdge Servers with Intel Xeon Gold 6330 processors
- Large memory footprint for data-intensive computing
- PCIe Gen 4 supports GPU I/O for AI/machine learning

[1] https://www.intel.com/content/www/us/en/architecture-and-technology/avx-512-overview.html

Printed in USA 012022/RJMJ/RL/PDF ♻ Please Recycle 349495-001US