# Genomic Breakthroughs Powered by HPC Technology

Oregon State University uses HPC technology to fuel groundbreaking scientific genome research to sustain the earth's ecosystems.

Customer profile

Oregon State University

Scientific Research | United States

"It used to take about 10 years to fully sequence a seawater sample, now it takes about less than a week to analyze and sequence all of the DNA in a sample."

**Christopher Sullivan**
Assistant Director of Biocomputing at OSU's Center for Genome Research and Biocomputing

## Business needs

At Oregon Status University (OSU), smart technology choices are enabling smarter scientific discoveries that are transforming the future. Fueled by HPC solutions from Dell Technologies, OSU's Center for Genome Research and Biocomputing is accelerating scientific breakthroughs. Computations that used to take 10 years and now take one week.

## Solutions at a glance

- Dell PowerEdge R7525 servers with 2x AMD EPYC 7662 processors each
- PowerSwitch Z9100 and NVIDIA networking

## Business results

- Advances sophistication of scientific breakthroughs
- Accelerates the speed of scientific research
- Increases productivity and processing capabilities
- Reduces operating costs
- Enables 20,000 jobs processed each day
- Extends hardware life

Innovative researchers at Oregon State University (OSU) are using technology to progress science in areas that help safeguard life. Through HPC technology, The Center for Genome Research and Biocomputing (CGRB) at OSU runs about 20,000 jobs a day to process massive data sets and enable more sophisticated scientific breakthroughs that help humankind. Analysis that used to take 10 years now takes about one week, accelerating the pace of scientific research.

In collaboration with Dell Technologies, OSU recently upgraded its high-performance compute (HPC) technology to achieve more processing capacity, increased memory and lower operating costs. With this upgrade, scientific researchers can do more, faster, paving the way for life-changing breakthroughs in science.

Built on a rich, 150-year heritage, OSU is an international public research university focused on building a smarter, healthier and more prosperous future. OSU engages with students and staff from more than 100 countries and all 50 states. OSU's collaboration with Dell Technologies underpins the latest technology upgrade at CGRB using a Dell High Performance Computing solution with PowerEdge R7525 servers with AMD EPYC™ 7662 processors, PowerSwitch Z9100 and NVIDIA high-speed networking.

## Doing more science requires HPC

Scientific researchers live in a data world. Leveraging colossal amounts of data helps improve scientific precision, new discoveries and meaningful outcomes. But capturing, processing and storing this data for repeated use poses novel challenges. OSU, in collaboration with technology innovators like Dell Technologies, excels at solving these challenges to fuel next-level scientific breakthroughs.

Scientific researchers work with large quantities of data and make sense of it. However, there is "never enough" data to support everything that needs to be done. With climate change research, for example, petabytes of data are not enough. As a result, innovative HPC technology in tailored configurations is the must-have capability that drives meaningful research.

A petabyte is equal to 2 to the 50th power of bytes. Some estimates indicate that a petabyte would fill 20 million tall filing cabinets or 500 billion pages of standard printed text.[1]

> "People come to us to work on things that "cannot" be done. We use computers to solve them. Each time we need to handle a new quantity or type of data, OSU rebuilds the infrastructure to accommodate it and process it in a way that is meaningful in terms of time and output."
>
> **Christopher Sullivan**
> Assistant Director of Biocomputing at OSU's Center for Genome Research and Biocomputing

## An ocean of possibilities for earth's future

HPC technology has helped power an earth-sustaining initiative with OSU's Hatfield Marine Science Center and is supported by the United States, Brazil, France, and Japan to monitor plankton in the ocean. As an essential linchpin to life on earth, plankton are powerful oxygen producers, accounting for up to 50% of the atmospheric oxygen. Plankton also play a crucial role in the food web and help sustain the earth's ecosystems. Monitoring plankton production is an important mission for the earth's future.[1]

Using a technology-equipped boat, a biological sample is taken by shining a laser in the water and generating videos. That laser can identify billions of points across a flow cell that has changed in color. That data is then digitized and geospatially mapped. Researchers receive a file that has billions of entries in it as the boat generates 100 terabytes of data per week. This volume of data requires tailored computing infrastructure to handle, analyze and store it. At this magnitude of data, it does not scale in the cloud.

1 https://www.nature.com/articles/483S17a, https://education.nationalgeographic.org/resource/plankton-revealed

"It used to take about 10 years to fully sequence a seawater sample, now it takes about less than a week to analyze and sequence all of the DNA in a sample," says Sullivan. "That's how technology has progressed science," Sullivan observes. "We use accelerators like GPUs to increase the speed of processing and this has helped genome research tremendously, especially as new applications are written," he continues.

## Forest sounds help endangered owls

A stunning example of the power of research coupled with technology uses sound to identify different species of owls in the forest. Sound data is collected, converted into spectrograms and then classified via GPUs. These classifications are done on an epic level, managing enormous volumes on lots of little GPUs. "Before our GPU, it was like pushing 10 gallons of water through a straw," shares Sullivan.

How has the owl research helped? As an endangered species, the spotted owl population continues to decline despite 30 years of protection. The sound-based research uncovered that barred owl population increases were the primary reason for the spotted owl declines. When barred owls were removed, spotted owl populations began recovering.[2]

As data patterns become apparent – as in the owl research – algorithms are developed, published and shared. These algorithms are always in dynamic development and help other scientists with new and expanded discoveries to accelerate knowledge and learning.

## COVID breakthroughs

In genome research, sequencing is used to determine the entire DNA of a cell type. Because of the numerous COVID variants, each time OSU receives a COVID sample, it must be sequenced. Like other research projects, COVID is a very, very large puzzle. "We are sequencing everything in the sample, using assembly and alignment and many different techniques to figure out how things separate and how things go together," comments Sullivan.

These techniques are being used throughout OSU and the research community with the compute aspect of research. They help to solve complex problems such as COVID that require massive amounts of data and novel HPC approaches.

## More threads, more memory, more GPU

The genome space needs a large memory footprint, a high number of threads and fast GPU performance. When a genome is assembled, it requires only 1-4 threads, but all of the available memory. When a genome is aligned, it uses a lot of threads but has lower memory. DNA methylation, or knowledge graphs, must be performed on a GPU.

"Technology must serve many at the same time. Therefore technology must be compartmentalized by workloads and by hardware. OSU users know how to do this and it's enabling breakthrough outcomes with our next-level HPC technology from Dell Technologies," shares Sullivan.

2 https://today.oregonstate.edu/news/removal-barred-owls-slows-decline-iconic-spotted-owls-pacific-northwest-study-finds