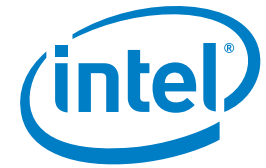


## CASE STUDY

Intel® Xeon® processor 5600 and 7500 family  
Enterprise Server  
High-Performance Computing  
Energy Efficiency, Environment, and Performance



# Mining for Cancer's Causes

## BC Cancer Agency doubles application performance with energy-efficient Intel® Xeon® processor 5670-based supercomputer

Who gets cancer and why? Scientists are using one of the world's most energy-efficient high-performance computing (HPC) systems—installed at the BC Cancer Agency's Michael Smith Genome Sciences Centre (GSC) in Vancouver, British Columbia—to help answer those questions. GSC uses the Intel® Xeon® processor 5600 family for computational performance and the Intel® Xeon® processor 7520 for high-speed storage. The solution provides eight times the processing power and 10 times the I/O performance of GSC's previous infrastructure, and researchers using it say their work can help lead to breakthroughs in understanding, preventing, and treating cancer and other genetic diseases.



"We were looking for a processor that had good characteristics in both memory-intensive and processing-intensive workload environments, and we wanted to get as many cores in an energy-efficient package as we could. The six-core Intel® Xeon® processor 5670 fit our needs very well."

— Greg Stazyk,  
Senior Manager of Research Systems  
Genome Sciences Centre

### CHALLENGES

- **Analyze rapidly growing data.** GSC expanded its genome sequencing capacity and now produces a petabyte of sequencing data annually. GSC's dual- and quad-core clusters couldn't keep up with the need to mine the data.
- **Fit into an existing data center.** GSC needs a flexible HPC system that meets floor space and power constraints and supports the Province of British Columbia's commitment to sustainability. The system must handle both memory- and processing-intensive workloads.

### SOLUTIONS

- **Intel Xeon processor 5600 family for high-efficiency performance.** GSC selected an IBM System x iDataPlex\* solution with 420 nodes, each populated with two six-core Intel Xeon processors 5670.
- **Intel Xeon processor 7500 family to accelerate I/O.** GSC uses the Intel Xeon processor 7520 in its IBM file system servers and storage subsystems, accessing primarily open source databases such as MySQL\* and PostgreSQL\*.

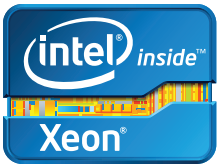
### IMPACT

- **Leadership in energy-efficient performance.** GSC's iDataPlex installation ranks as the world's 20th most energy-efficient supercomputer<sup>1</sup> and the 178th most capable one.<sup>2</sup> The 53.6 teraflop system fits into six racks, providing four times more cores than the previous 1U blade infrastructure in half the space. Its positions on the Green500 and Top500 lists enhance GSC's reputation in both science and sustainability.
- **Scientific progress.** Using the system's high-performance architecture, applications run twice as fast and GSC can run four times more applications. Scientists are better able to discover variations involved in developing and potentially treating cancer and other genetic diseases.

### Inundated with Data

GSC is a world leader in genomics and bioinformatics research. Hundreds of research teams at GSC, across Canada, and around the world use GSC's HPC resources to assemble and map genes and mine genomic data as they seek to unlock the genetic origins of disease and health. GSC's ABySS\* genomic assembly tool is widely used at other sites.

GSC received a grant from the Canadian Foundation for Innovation (CFI) to expand its genome sequencing capacity and capability, bringing its production of sequencing data to one petabyte a year—a fivefold jump over 2009. In addition to GSC's own genome research, other scientists often send samples to be analyzed, further raising the demand for HPC resources.



## GSC can run four times more applications, and the applications run twice as fast

"With our expanded ability to sequence more genomes at greater depth, the research challenge is to effectively analyze and mine the vast amounts of new data that can help pinpoint variations involved in cancer," says Greg Stazyk, senior manager of research systems at GSC. "We're being inundated with large volumes of new data, so this HPC environment is really essential."

### Varied Workloads, Flexible Architecture

Backed by funding from CFI and the Province of British Columbia, GSC set out to acquire a flexible, high-performance system that could meet the power and space constraints of its existing data center, handle workloads that Stazyk says are "extremely varied," and help researchers keep up with the explosive data growth. The IBM iDataPlex system and Intel Xeon processor 5670 fit the bill. Both are optimized for energy-efficient computing, and the Xeon processor's six-core design and expanded memory bandwidth make it a great fit for GSC's mix of workloads. GSC calls the new supercomputer Genesis\*.

GSC optimizes processing performance and throughput by configuring Genesis with InfiniBand\* Quad Data Rate (QDR) interconnects, 20 terabytes of RAM, and 20 TB of flash memory. The operating system is CentOS\* Linux\* 5.4. Genesis connects to 1 petabyte of storage and uses the Intel

Xeon processor 7520 to enhance the performance of its IBM General Parallel File System\* storage servers. Stazyk says GSC now enjoys eight times more processing power and 10 times more I/O performance than the previous infrastructure provided. Genesis fits into six racks, taking up just half the space of the previous 1U blade systems.

### Advancing the Study of Cancer

With Genesis, scientists at GSC and other institutions are using genome sequencing data to advance the understanding of cancer. "Sequencing just gets you the raw information," Stazyk says. "Data analysis is where you can make discoveries that can lead to the recognition of proteins which could eventually be targeted by anti-cancer treatments, or provide more information for targeted clinical treatment options. With this system, we're in a better position to make those discoveries. Our applications have a wide mix of performance characteristics, and this system handles them all."

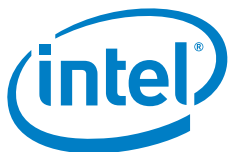
Going forward, Stazyk's team plans to virtualize with Intel® Virtualization Technology (Intel® VT) to further improve flexibility. "In the research community, we run a wide range of applications, and some may have been written for operating systems that we don't natively run on the cluster," he explains. "Virtualization will give us greater flexibility to run those applications and minimize the development work that might otherwise

### SPOTLIGHT ON BC CANCER AGENCY AND MICHAEL SMITH GENOME SCIENCES CENTRE

The BC Cancer Agency provides a comprehensive program of cancer control and health care delivery for the province of British Columbia, Canada. GSC is a multidisciplinary research department within the Cancer Agency. GSC offers sophisticated tools and expertise that enable scientists to better examine the inner workings of human and other genes, including specialized machines for sequencing genomes and analyzing genetic material. The center began operation at the BC Cancer Agency in 1999.

be needed."

Find a solution that is right for your organization. Contact your Intel representative or visit the Reference Room at [www.intel.com/references](http://www.intel.com/references).



1 <http://www.green500.org/lists/2010/11/top/list.php?from=1&to=100>

2 <http://www.top500.org/lists/2010/11>

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Intel may make changes to specifications, product descriptions, and plans at any time, without notice.

Intel® Virtualization Technology requires a computer system with an enabled Intel® processor, BIOS, and virtual machine monitor (VMM). Functionality, performance, or other benefits will vary depending on hardware and software configurations. Software applications may not be compatible with all operating systems. Consult your PC manufacturer. For more information, visit <http://www.intel.com/go/virtualization>.

This document and the information given are for the convenience of Intel's customer base and are provided "AS IS" WITH NO WARRANTIES WHATSOEVER, EXPRESS OR IMPLIED, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NON-INFRINGEMENT OF INTELLECTUAL PROPERTY RIGHTS. Receipt or possession of this document does not grant any license to any of the intellectual property described, displayed, or contained herein. Intel® products are not intended for use in medical, lifesaving, life-sustaining, critical control, or safety systems, or in nuclear facility applications.

© 2011, Intel Corporation. All rights reserved. Intel, the Intel logo, and Intel Xeon are trademarks of Intel Corporation in the U.S. and other countries.

\*Other names and brands may be claimed as the property of others.