

IBM Linux on POWER solution aids Princeton Molecular Biology lab.



Overview

■ **The Challenge**

Improve responsiveness to researchers by implementing a powerful, scalable server that could support a growing database and compute-intensive bioinformatics applications

■ **The Solution**

IBM @server® pSeries® 630 server, with two high-performance IBM POWER4+™ 64-bit processors and 4GB of memory running SUSE LINUX Enterprise Server 8

■ **The Benefit**

Increased the speed of database queries by up to ten times and increased system availability, while reducing administrative requirements and providing a clear upgrade path

At the forefront of bioinformatics

The Department of Molecular Biology is a Princeton University center of research in the life sciences. In the department, approximately 100 graduate students, 130 postdoctoral fellows and 42 faculty and associated faculty pursue goals in diverse areas of research. The department includes a number of autonomous, specialized laboratories. The laboratory of Ihor Lemischka, a professor of molecular biology, focuses on stem cell research. “The main thing that differentiates us from stem cell groups at other universities is that we place more

emphasis on bioinformatics and gene expression analysis,” says Jason Hackney, a graduate student in Princeton’s Department of Molecular Biology. As a result, the Lemischka lab requires a significant amount of computing power to run its programs and experiments.

Microarrays and the need for speed

As part of its stem cell research, the Lemischka lab has always collected and stored substantial amounts of data. Previously, the lab focused its efforts on gathering deoxyribonucleic acid (DNA) sequence information, which consists of base pairs of nucleotides that are made up of adenine (A), guanine (G), cytosine (C) and thymine (T). However, in early 2002, the lab began moving toward storing more than just sequence data to also storing microarray data. Laurie Kramer, a staff member of the Department of Molecular Biology, explains, “Microarrays are used in molecular biology to determine gene expression, and you get lots of data off each microarray.” As a result, the Lemischka lab faced the prospect of

a rapidly expanding database and longer and longer wait times for database queries. And this poor performance not only slowed down experiments in the Lemischka lab, it also caused long wait times for scientists from partner universities, who relied on the lab's Web site for research data.

The Lemischka lab's existing database and Web site were running on two aging Dell servers, which were powered by the Linux operating system. The servers lacked the processing speed and storage capacity that the lab required to efficiently run queries on and house microarray data. To increase responsiveness to researchers' needs, the lab sought a new solution that could provide the necessary storage and processing capacity, high availability and a clear roadmap for growth. Says Hackney: "I'm envisioning that this group will need more processing power in the near future. Additionally, one of my primary concerns was finding a server that would be fairly easy to administer."

A powerful, open solution

Once grant money was secured for a new server, Hackney researched the lab's options. He narrowed the choices down to two solutions—one from IBM and one from Sun—that met the lab's performance and budgetary requirements. The Lemischka team chose the IBM solution due to its upgradability, reputation for reliability and ability to support the open-standards-based Linux operating system. "We've had good luck with Linux, and it's relatively easy to administer, so we wanted to continue using it," reports Hackney. "I also thought that the IBM solution was priced competitively and showed a good upgrade path," he adds.

The Lemischka lab implemented an IBM **@server** pSeries 630 Model 6E4 server, which can provide enterprise-class reliability, availability and serviceability. It runs SUSE LINUX. The p630 includes two high-performance POWER4+ 64-bit processors and 4GB of memory. In addition to BLAST, which is a sequence-alignment program from the National Center for Biotechnology Information (NCBI), the p630 runs an Apache Web server, and MySQL and PostgreSQL databases. "With respect to microarray programs, we're also using several

applications that I've written, as well as Xcluster, which is a bioinformatics application written by Gavin Sherlock," relates Hackney.

Using the p630, researchers in the Lemischka lab can now quickly perform bioinformatic analysis of microarray data. The highly resilient server also enables the lab to efficiently provide a group of partner universities, which are located around the world, with 24x7 access to its data. The Web site has a data repository (for microarray and sequence data) and information about individual genes that are expressed in stem cells. The lab plans to add collaborative features to the Web site in the future.

According to Douglas Welsh, a lecturer and senior professional technical staff member in Princeton's Department of Molecular Biology, setting up the p630 server was unexpectedly simple. "This was one of the easiest system installs that I've ever been involved with," states Welsh. "The hardware and software on the p630 are well integrated. It's easy to install most of the necessary drivers and software components, because they're already there. You don't have to spend time tracking

them down.” Hackney attributes the smooth set up to a combination of the solid engineering of the server along with its compatibility with Linux. “It’s very nicely designed for installing exactly what we wanted,” says Hackney. The Lemischka lab also signed a maintenance contract with IBM for its p630.

Why IBM?

Based on the prevalence of open applications in its environment and the existing skill set of its support team, a Linux solution was the natural choice for the Lemischka lab. And IBM was one of the few vendors that offered a pure Linux solution in a high-performance, reliable, easy-to-install-and-maintain environment. Although technical performance wasn’t the lab’s only consideration: service was also a pivotal factor in the decision. “I work with another team in the Department of Molecular Biology that uses an IBM RS/6000® based Beowulf Linux cluster, and the service and support from IBM for that solution has been first rate,” reports Welsh. “Software support and support for hardware problems is transparent. I call for help one day, and a technician has the issue fixed the next day,” he says. Given Welsh’s experience, the team from the Lemischka lab knew it could rely on an IBM solution.

Isolating encouraging results

Since implementing the p630 Linux on POWER™ solution, response times to database queries in the Lemischka lab have improved dramatically, helping to increase the productivity of researchers. Hackney estimates that queries are, at minimum, four times quicker on the p630. “It’s hard to get exact numbers because network overhead affects the results, but on longer queries it’s at least four times faster and sometimes as much as ten times faster,” reports Hackney.

The p630 has also proven easier to administer than the lab’s old solution. According to Hackney, he now spends less than ten minutes a week performing administrative tasks on the server. Previously, he was often sidetracked from his research to fix hardware issues. He also notes that since installation in September 2003, the server has only been down once for less than ten minutes, to upgrade a Linux kernel for security reasons. No hardware failures have occurred.

Looking ahead, the Lemischka lab plans to take advantage of advanced features on its flexible p630 server. For example, it may use logical partitions (LPAR) to isolate the database, Web and bioinformatic analysis operating environments.

This could help further increase operating efficiency without the cost of adding new hardware. The addition of another pair of processors to the server is also on the horizon. “That’s part of the upgrade path, and it will be necessary given the amount of data we’re generating,” says Hackney.

“I’ve been involved with IBM in the 20 years that I’ve been in this department, and we have a very close relationship that includes support for our research, hardware grants and collaboration with IBM scientists.”

—Douglas Welsh, lecturer and senior professional technical staff member, Princeton Department of Molecular Biology

For more information

To learn more about this solution, please visit:

ibm.com/eserver



© Copyright IBM Corporation 2004

IBM Global Services
Route 100
Somers, NY 10589
U.S.A.

Printed in the United States of America
05-04
All Rights Reserved

The eServer logo, IBM, the IBM logo, POWER, POWER4+, pSeries and RS/6000 are trademarks of International Business Machines Corporation in the United States, other countries, or both.

Other company, product and service names may be trademarks or service marks of others.

References in this publication to IBM products and services do not imply that IBM intends to make them available in all countries in which IBM operates.

This customer story is based on information provided by the Princeton University Department of Molecular Biology and illustrates how one organization uses IBM products. Many factors may have contributed to the results and benefits described; IBM does not guarantee comparable results elsewhere.