



bio1NFORM

The integrated informatics news source

volume 7, number 36
September 22, 2003

Netezza Brings New Tera-Scale Data Warehouse Architecture to the Bioinformatics Market

FRAMINGHAM, Mass.-based startup Netezza began shipping its data warehouse technology — which packs a server, storage, and database into a single rack-based appliance — a year ago to business intelligence customers. Now, the company is setting its sights on the bioinformatics research community with a “biology aware” version of the system.

What makes the young firm so certain it can go head-to-head with Oracle, IBM, Teradata, and the other vendors in the highly competitive market? According to Bill Blake, senior vice president of product development at Netezza, “performance improvements of 10-50 times over those systems at half the cost” — not to mention new functionality that integrates genomic data types and NCBI’s Blast within the system so that users can perform all their database searches within the database using SQL.

Blake said that Netezza (pronounced Net-ee-zza) has developed a data warehouse architecture where “the processing for a database search occurs as close as possible to where the data sits on the disk.” Each rack in the company’s Netezza Performance Server — or NPS — includes a host Intel/Linux server and several hundred blade-like combinations of a microprocessor, a disk, and an optimized disk controller that includes logic for database operations. The company calls these disks Snippet

Processing Units, or SPUs, because each one processes a small piece of a database query. This parallel arrangement offers dramatically improved performance over traditional relational databases, Blake said, where “for the database query to be processed, much of the data has to be moved over the local area network and into the host.”

For life science researchers, access to bioinformatics-friendly data types should provide additional performance improvements, Blake said. The company designed character large objects that can accommodate the four letters of nucleotides or the 20 letters of amino acids rather than store them as strings, like traditional relational database systems do. “We want to make the database capable of doing the work that today is done with flat files and supercomputers right inside the relational database, rather than just store the end result,” he said. The Blast capability, Blake said, was added because it’s “the most common similarity check” for bioinformatics.

Blake moved to Netezza in June of last year from Compaq, where he was vice president of the company’s high-performance technical computing business, and a key player in deals that placed Compaq computers at the heart of the human genome sequencing effort — both on the public side, with the Whitehead and Sanger Institutes, and the private side, with Celera. Blake has already continued on next page



www.genomeweb.com

The GenomeWeb Intelligence Network

© 2003 GenomeWeb, LLC. All rights reserved. Copying, photocopying or duplicating this publication in any form other than as permitted by agreement with GenomeWeb, LLC is prohibited and may constitute copyright infringement subject to liability up to \$100,000 per infringement. For photocopy permission, back issues, bulk distribution or site licenses, please contact +1.212.269-4747 or sales@genomeweb.com



Netezza...

continued from page 1

drawn upon his connections in the life science community to line up the J. Craig Venter Science Foundation as a beta customer for the biology-enabled NPS. Blake said the architecture "lets them take some of the work they were doing on Oracle systems and some of the work they were doing on their supercomputing farms and merge it into one system."

Marshall Peterson, CTO of the

Venter Foundation, was unavailable for comment last week. Blake said that Netezza does not yet have benchmarks for bioinformatics searches performed on the database, but noted that complex queries in other application areas have shown performance improvements of 10-50 percent.

Netezza, however, is not the only database provider touting embedded Blast search capabilities. Oracle's new 10g database also includes a version of NCBI Blast [*BioInform 09-15-03*], and the company has a much larger

customer base — not to mention marketing muscle — than Netezza. Blake said he hasn't seen Oracle's Blast-enabled database yet, but was confident that the Netezza architecture will give NPS the advantage. "I don't expect them to be at the same performance level," he said.

Netezza will also compete against its larger competitors on cost: A one-rack 1.5 terabyte NPS has a list price of \$622,000 — less than half the price of an equivalent Oracle, IBM, or Teradata system, Blake said.

— BT