

HOW NCGR MAKES PROGRESS



GENETIC RESEARCH FOR HUMAN HEALTH

The National Center for Genome Research (NCGR) was founded in 1994 as an independent, non-profit institute for discovery-driven research in computational biology, medicine, and bioinformatics—that is, the application of databases, algorithms, computational, and statistical techniques to the field of molecular biology.

Specializing in genetic sequencing (determining the individual arrangements of components of RNA and DNA that compose a specific gene), NCGR has since developed Web-based information resources in collaboration with research communities around the world.

PROCESSING MIND-BOGGLING AMOUNTS OF DATA

The evolution of NCGR into a sequencing center in its own right marked the start of a new era for the organization.



CHALLENGE

Implementing a highly scalable, specialized data warehousing solution to support genetic sequencing

SOLUTION

Progress® DataDirect® for SequeLink demonstrated up to 20 times better performance than other solutions in providing JDBC connectivity

BENEFIT

Saved two–three months time for two developers to natively enable NCGR's application to access the data warehouse platform

“It catapulted us into the next realm” explains Neil Miller, informatics team leader at NCGR. “Our database needs grew very rapidly.”

Genome sequencing can generate mind-boggling amounts of data. The IT team at NCGR quickly realized that the database platform they currently had in place—a Sybase ASE relational database running on a Sun Sparc infrastructure—would not be up to handling the task entirely on its own.

NCGR decided to partner with Kognitio in optimizing that vendor’s WX2. But, while WX2 seemed the perfect fit as a data warehouse and analysis solution for NCGR’s data processing requirements, it lacked a way to connect to the organization’s own application, Alpheus.

“The software written to pipeline all this data into the database is written in Java,” says Kathy Myers, Sr. Database Administrator at NCGR. “We either had to rewrite the entire application to enable ODBC connectivity—which the WX2 allowed—or else find a third-party JDBC connectivity solution that could handle the extremely large amount of data we’d be throwing at it.”

“We needed the glue to hold, and DataDirect supplied that glue.”

*John Utsey
Systems Manager
NCGR*

A CLEAR PERFORMANCE FRONT-RUNNER

The team decided to go with a third-party solution. Finding one capable of effectively handling the projected workload, however, proved easier said than done. The Illumina Genome Analyzers that NCGR uses in its genome sequencing projects are expected to generate tens-to-hundreds of terabytes of data over the next year.

Our first choice of a JDBC connectivity solution was made strictly on the basis of price,” says Miller. “It failed miserably, but led us to write a very trimmed-down test case that gave us a rows-per-second metric, which we were then able to use in quickly evaluating different solutions.”

That test eventually revealed a clear front-runner in terms of performance: Progress® DataDirect® SequeLink®, which performed 20 times faster than the initial product.

Using an “n-tier” deployment option, SequeLink provides all the benefits of single, universal client components while extending access to virtually all data in an enterprise. The SequeLink Server for ODBC Socket used by NCGR effectively turns the SequeLink Server into an application server. This option allowed NCGR to access virtually any ODBC-compliant database—in this case WX2—from within the SequeLink environment, but through a scalable n-tier architecture.

FROM DAYS TO HOURS

Installation and configuration of the ODBC Socket is quite simple; it involves simply installing the ODBC Socket plus the specific ODBC driver and any prerequisite software for the target data source on the server platform of choice—in NCGR’s case, Linux running on the Sparc platform. NCGR’s team reports that the entire process of getting the solution up and running took approximately one day.

“From the point of view of our application, the connection was immediate and seamless,” Miller recalls. “We required a little help from Kognitio to actually connect to WX2.”

DBA Kathy Myers relates that Progress DataDirect support was instrumental in tracking down the problem and expediting the install.

“They were right there without delay,” she says. “They very quickly pinpointed the issue as being on the WX2 side so we could get a patch from Kognitio. The fact that we didn’t have to jump through any crazy hoops to get the solution working was key.”

Miller points out that, from the development side, NCGR saved considerable time and resources that would have otherwise gone into rewriting the Alpheus application.

“We were looking at a good two or three months for two developers to enable pure ODBC connectivity in our app,” he says. “It was going to be a big deal.”

He adds that the type of performance SequeLink delivers is absolutely essential for the extreme data-intensive projects NCGR undertakes.

“At that order of magnitude, the number of rows we can get per second becomes extremely important,” he says. “It can make a data processing pipeline run in four hours versus three days.”

But John Utsey says that it was the stability of the DataDirect solution that was of the most critical importance.

“We’d been under a lot of pressure because of the time we’d lost trying to use the original third-party solution,” he says. “We needed the glue to hold, and DataDirect supplied that glue. It held, performed very well, and made us happy.”

PROGRESS SOFTWARE

Progress Software Corporation (NASDAQ: PRGS) is a global software company that enables enterprises to be operationally responsive to changing conditions and customer interactions as they occur. Our goal is to enable our customers to capitalize on new opportunities, drive greater efficiencies, and reduce risk. Progress offers a comprehensive portfolio of best-in-class infrastructure software spanning event-driven visibility and real-time response, open integration, data access and integration, and application development and management—all supporting on-premises and SaaS/cloud deployments. Progress maximizes the benefits of operational responsiveness while minimizing IT complexity and total cost of ownership.

WORLDWIDE HEADQUARTERS

Progress Software Corporation, 14 Oak Park, Bedford, MA 01730 USA
Tel: +1 781 280-4000 Fax: +1 781 280-4095 On the Web at: www.progress.com

For regional international office locations and contact information, please refer to the Web page below:
www.progress.com/worldwide

Progress, DataDirect, SequeLink, and Business Making Progress are trademarks or registered trademarks of Progress Software Corporation or one of its affiliates or subsidiaries in the U.S. and other countries. Any other trademarks contained herein are the property of their respective owners. Specifications subject to change without notice.

© 2009-2010 Progress Software Corporation and/or its subsidiaries or affiliates. All rights reserved.
Rev. 06/10 | 6525-128928

