Wall Street & Technology

Xtreme Data Debuts Fast Database for Analytics

By **Penny Crosman (mailto:pcrosman@cmp.com)**
Jul 27, 2009
URL: **http://www.wallstreetandtech.com/showArticle.jhtml?articleID=218600647**
(http://www.wallstreetandtech.com/showArticle.jhtml;jsessionid=HLTWXDRXOHEZSQSNDLPSKHSCJUNN2JVN?
articleID=218600647)

In a new solution intended to give Wall Street firms a way to quickly analyze ever-increasing volumes of market messages, **Xtreme Data (http://www.xtremedata.com)** has taken its FPGA (field programmable gate array) hardware accelerator chips, which are used by Activ Financial in its high-speed market data messaging product, and built a data analytics appliance around them called **dbX (http://www.xtremedata.com/dbx)** . dbX, its maker says, can perform ordinary SQL queries against thousands or millions of rows of data at the speed of light, using hardware-based parallel programming. The FPGA replaces an AMD or Intel chip in a server and "spoofs" the system it's in (so far in this case, an HP Proliant server) into thinking it's a regular CPU, yet the company has found it processes SQL queries about 15 times faster than a six-core Intel or AMD chip.

"We've created an appliance that's purpose built for ad hoc analysis of large data sets," says Geno Valente, vice president. "People want to be able to ask questions of data, look at the results of their queries, have an 'aha' moment, then ask more questions. Typically IT departments ask you to predefine the query so they can build a box that will get you the answer you want, but as soon as you want to ask a different question, a new box needs to be built. Analysts and researchers want unrestricted access." This new product puts Xtreme Data into direct competition with Netezza, Oracle, and Teradata.

Most SQL processing engines are software-based. Xtreme Data, in contrast, has built a SQL querying appliance by using hardware definition language to put the SQL engine on a microchip. Therefore, "all the instructions are processed at light speed with electrical signals on the silicon chip, instead of in software, and thus you eliminate all the configuration issues that come with an Oracle, Sybase or other database," says Justin S. Magruder, president and CEO of consultancy Noetic Partners. "You can eliminate massive amounts of latency and process thousands or millions of instructions in parallel. We're seeing the ability to process billions of rows in a fraction of the time it would take us to do the same thing with a software-based database."

A quant or algo trader, Magruder says, could ask a question about a pattern he's seeing in the market by querying the NYSE Euronext TAQ database, which contains more than 30 years of tick-

level history from U.S. markets for all issues listed on the exchange, and get a response in seconds or minutes instead of hours or days. At one large Wall Street firm, where Magruder worked in the advanced research and quantitative strategies group, "We were writing and running very fast code with flat files that could optimize speed, but we were still getting results in days because we could only process 6-20 threads at a time in software, whereas in hardware we could crush through the rows of data more quickly — 1,000 rows at a time across millions of rows," he says. "This gives an opportunity for risk managers, algo traders, and quants looking for patterns an opportunity to get answers to questions more quickly."

The dbX database appliance is a box the size of a small refrigerator with only a power cord and Ethernet cable running out the back, according to Magruder.

Typically, the downside of an appliance is that certain functions are baked into hardware and therefore hard to customize or upgrade. However, Magruder says, "There are certain things that make sense to put into hardware because they're never going to change," such as basic SQL processes.

On Wall Street, this appliance could be used for decision support, running algo routines; pre-trade and post trade compliance, finding signals that would indicate a trader is about to do something that's not compliant, compliance reporting and possibly real time risk monitoring dashboards; and performance, investor, and client reporting. According to Valente, it's low-cost: the solution runs about $20,000 per terabyte — including computing, storage, networking, database software and a year of maintenance and support — whereas competitors are in the $60,000 per terabyte range, he says. He also says the FPGA accelerator helps save more than a kilowatt of energy per server; "it's the greenest database out there," he says.