

The Real Time Data Management Imperative

By Ken Rugg and Mark Palmer

By the time you finish reading this article, an automated trading system on Wall Street will have evaluated the state of the stock market 210 million times, executed over 100,000 electronic stock trades, and may account for 10 million dollars. This system captures every trade made on every stock in the world - 750,000 of them, and executes mathematics-based trading strategies that identify opportunities to buy and sell stocks. Those algorithms evaluate questions like: Is the real-time price of IBM above its average of the last 10 seconds? 10 minutes? How have HP and Sun performed over this interval? Has any one firm bought or sold an unusual number of shares of any of this group of stocks? Every year, this system collects 2 terabytes of market data in real time. The system has been called a "sub-atomic real time data warehouse" by its creators.

But this is just one system in an overwhelming trend toward real-time program trading. Today, over 25 percent of the trade volume is generated programmatically; by 2008, 40 percent of the world's volume will be traded programmatically.

The data management implications of this trend present radically new challenges for database architects: how do you design a 10 terabyte database that handles streaming event data that changes 50,000 times a second, support complex, concurrent real-time queries, has a proactive trigger-like model, and operates with near-zero latency?

The answer is real time data management and it's not only critical to program trading; it will be a critical part of your future as well. Here's why...

The New Technology Landscape and the Business Driver to Real Time

EBay, Amazon, Wal-Mart, JetBlue, iTunes, and Google are making dinosaurs out of their competition by changing the rules with on-line, real time business processes. They recognize that the ubiquitous computing connectivity we possess can be leveraged to resounding advantage. And while these companies innovate by leveraging connectivity and efficiency, we're at the beginning of another technology sea change that may yield the next set of incumbent-bashing business models. Radio Frequency Identification (RFID) will allow businesses to gather real time data inexpensively, all the time, in real time. What innovative entrepreneur will tap RFID's power and fully automate a business, and dramatically cut cost in an industry that's still thinking bar codes are "good enough?"

Business leaders have begun to recognize the power of connectivity and real time data collection and issued a tidal wave of mandates & regulations to force a shift to real time: FAA security regulations after 9/11, RFID for real-time supply chains, dramatically increased auditing requirements of Sarbanes-Oxley, HIPAA and Gramm-Leach-Bliley, and FDA mandates to track drug and blood "pedigree" to ensure public health. The list grows every day.

So technology change lies ahead; but what kind of change?

What's Changing?

The style of processing required, and the data volume, velocity, and rate of change present new challenges for applications to effectively leverage real time data.

Let's start with the style of processing. Traditional relational databases process sets of data. That's fine when you have a bunch of fish in a pond, and you want to know how many trout, goldfish and minnows are in it. Queries are designed, and reports run nightly, or monthly. Real time systems, on the other hand, process streams of data - like salmon swimming up stream. Multiple streams are processed at the same time - the movement of bears relative to salmon is important too. "Reports" must be able to run in milliseconds, continuously, against all data streams, to help the salmon avoid the bear's mouth. Real time systems require new, event processing technology. Real time databases, for example, have unique "physics" of storing, querying, and managing event streams that is radically different than set-oriented data processing. The new style of computing is called Complex Event Processing (CEP). CEP helps detect patterns in the multiple streams of data, and enables organizations to act on those patterns, in real time.

How much data? A real time tick-by-tick feed will generate a few terabytes of tick data per

year. It's becoming more of the norm to store and process this type of volume.

How quickly does the data change? In the program trading application discussed earlier, 5,000-50,000 change events occur every second! The second generation RFID standard for readers stipulates a single reader must handle 1,800 reads a second, and a typical distribution center will have 20-100 readers creating event streams simultaneously.

Finally, the time to make decisions is also being compressed. Something that used to take minutes or hours is now expected to be done almost instantaneously.

A new type of event oriented computing is coming into the mainstream. At the same time, the volume and rate of change of data is steadily increasing while the time to process the data is getting shorter. This combination of changes is driving a shift in computing architecture - a sea-change toward a new real time computing platform.

6 Steps to Real Time

Real time is not brand-new; it's just becoming more commonplace. And it's not coming that quickly, so you still have time to prepare for the new age. Here's a simple "6 step action plan" to learn the lessons in industries that have already made the shift to get you started on your real time journey:

1. **Employ Complex Event Processing (CEP).** CEP is technology designed to process multiple streams of simple data events with the goal of identifying the meaningful events within those streams. Learn how to process this new streaming event data by reading the seminal book on the topic: *The Power of Events*, by David Luckham from Stanford University.
2. **Cache data In-Memory.** Simply analyzing raw event data is useless without access to reference data that helps inform decisions. In a trading system, "reference data" might be the groups of securities that represent a portfolio. In an RFID system, "reference data" might be the warehouse management database. In either case, reference data must be accessible at in-memory speed, and real-time data caching technology makes this possible.
3. **Exploit Distributed Shared Memory Grids.** "Distributed, transactional memory grids" have been around for a while. They allow large sets of data to be shared and distributed, so event processing can proceed at in-memory speeds for real time automated trading decisions to be made. Integrated with a database, shared memory grids enable real time data access.
4. **Implement EDA & SOA - at the Same Time.** Event-driven architecture can be service-oriented, and vice-versa. Financial trading systems are event driven, and they are more easily maintained if they are service-oriented. As more and more systems require real time data views, services help snap-in existing the real time platform with the enterprise.
5. **Be Analytical, Be Curious, and Be Bold.** Real time is bold change that can create dinosaurs out of your competition. But simply making old decisions faster isn't a bold business move. The business processes that will make the big impact are ones you probably aren't even thinking of right now. In his book *"Heads Up"*, Kenneth G. McGee presents a methodology for selecting which among your business processes are the best candidates for real time. Be curious about new opportunities! Be creative! Be bold!
6. **Iterate, Embrace & Extend.** Only about 20 percent of the world's existing IT systems will become real time, so plan for real time systems to live with old time systems for a long time. So CEP systems must have a safe, scalable way to be connected to existing systems that provides enterprise qualities of service.

By embracing the new era of real time opportunity and boldly attacking the unique opportunity we have, not only will you avoid the bear, but maybe you'll start gobbling up your competitors as well.

About the Authors:

Ken Rugg is vice president for ObjectStore, and Mark Palmer is vice president of marketing with ObjectStore