

# What's the REAL real-time opportunity?

## The lines between operational and analytical data have begun to blur.

by Neil Raden

The following is an excerpt from "Exploring the Business Imperative of Real-Time Analytics" (Neil Raden, Hired Brains, Inc., October 2003).

**THE CLASSIC DEFINITION** of a data warehouse is a database or collection of databases that houses data culled from multiple "operational" systems and provides an enduring repository of information to support analytical needs.

There has always been a distinction between operational processing—those processes involved in actually running the minute transactions of the business—and analytical processing—the sorts of things that are more contemplative or informational in nature. The boundaries between these processes have always been clear. Going forward, however, those lines are starting to blur.

Here are some axioms about data warehousing that no longer hold up:

- Data warehouses do not support operational reporting—those reports that can be generated wholly within a single operational system, such as accounts payable.
- The flow of information is one-way, from operational systems to data warehouses.
- While data flows into operational systems in near real time, data warehouse updates happen at discrete intervals, such as nightly, weekly or monthly.
- Operational queries are simple and fast while analytical queries are complex and slow.
- Operational systems support process-oriented events; data warehouses support informational events.

- It is impossible to perform load balancing for analytical and operational processing on the same platform simultaneously.
- Analytical work takes time; therefore there is no need for real-time data.

The breakdown of these axioms, through the relentless onslaught of technology and economics, is erasing the distinction between operational and analytical processing. The cost of computing resources—CPU, memory and storage—has collapsed. At the same time, the capacity of these components has continued to double every two years just as Moore's Law predicted. What was unthinkable only five years ago is now commonplace, especially in terms of real-time analytics.

### Why get real?

Why the rush to real time? A major impetus is the effect of external factors. Because of ATMs, banks had to come up with a process to manage account balances in real time, not overnight. The growth of e-commerce also created a huge demand to have real-time enabled back-office systems that can communicate with each other. The gradual retirement of many legacy systems and replacement with enterprise software, such as enterprise resource planning and customer relationship management, broke through the overnight batch processing model and moved organizations to a lower-latency model.

Once the operational systems went to real time, it was only a matter of time before the business processes they supported needed to monitor more up-to-date data, hence the push toward real-time data warehouses.

Obviously, lower latency in all systems is desirable, but how do you know when it's time to make the investment in a real-time data warehousing environment? Unless there is a business process that can take the output from a real-time data warehouse and act on it, there is no real need for one.

The subject matter of the data warehouse can also provide clues. At the heart of almost every query is a fact, a metric, some numeric value. The metrics are always a proxy of some phenomena being measured. If the conditions being monitored and measured change slowly enough within a given time period, say hourly, then up-to-the-second data is clearly not needed. However, when those phenomena can be materially changed with a single transaction, then real time is essential.

But a business's ability (or inability) to react within a certain time frame can reverse this decision. For example, continuous replenishment is a win-win for manufacturers and retailers, but if distribution operates on a two-week replenishment cycle, it doesn't make sense to optimize the inventory with every sale.

The most important thing to consider is what purpose a data warehouse serves in your organization. For years, the presumption was that interactive tools, such as ad hoc queries, online analytical processing (OLAP) applications and even exotic types of business intelligence (BI) tools like data mining were the driving forces behind the effort. As it turns out, these activities play a somewhat smaller role than anticipated.

There are two important functions that truly leverage the data warehouse. The first is reporting. People in organizations have shown that they rely on standard, repeatable, reliable presentations of data. Making it as useful, current and painless as possible is the goal. That three major BI vendors released new reporting products in the past

year is evidence that the message is being heard. Enterprise reporting tools are starting to be viewed as BI tools of choice, especially those that can provide authoring, security, administration and distribution services efficiently.

The second area has yet to fully emerge, but this is where real time will be prominent. The move to hybrid systems, a complete merger of operational and analytical processing, will finally pull the data warehouse onto center stage. In a few years, it will be difficult to know whether a data warehouse is part of a process "work step" or not. The only reason that most current operational systems are not requesting data from data warehouses is that it was not considered possible before. That was before active data warehouses.

### Are you active?

It is important to make a distinction between real-time data warehousing (RTDW) and active data warehousing (ADW). Although they are similar, they are not the same.

RTDW refers to the technical aspects of a data warehouse that updates as data is presented to it. Key concepts include physical modifications to the database schema and the database environment; movement of data across the enterprise; extraction, transformation and load processes; modification of downstream processes, especially alerts; creation of extracts, cubes and data marts; and the whole new methodology for designing and implementing RTDWs.

The concept of ADW, on the other hand, is in the realm of work, not technology. In other words, ADW doesn't necessarily define architecture or methodology; rather, it speaks to the role the data warehouse plays in the enterprise. A true ADW is an active participant in the portfolio of applications plugged into the messaging pipelines, listening to and responding to operational systems behind and beyond the firewall. It is an integral part of the real-time hum, no longer relegated to a

passive role of serving up reports and queries to a small audience.

ADW will almost always involve an element of RTDW, but what makes an ADW "active" is that it participates, in real time, with other systems operating online. For example, an ADW may be part of a supply-chain sourcing application, switching suppliers in real time after analyzing risk factors based on an unexpected change in the production schedule.

This ability to communicate in real time, with or without human intervention, is a bold step, vastly different from the current practice of simply sourcing data for ad hoc queries, OLAP and reporting. Timeliness is only one aspect of ADW. Also inherent are features such as scalability, availability and manageability that enable quick responses to ever-changing workload dynamics.

### Looking for opportunities?

In evaluating opportunities where real-time analytics can provide measurable value, I've found that the best opportunities are those situations where the organization's response to a set of variables can be well defined, automatic and reflexive. Paradoxically, there are not many situations in business where we can find well-defined, reflexive, automatic processes. That is precisely why the need for real-time analytics is becoming more urgent. But for the time being, until the technology matures and you gain some experience with it, reach for the low-risk situations, the low-hanging fruit.

What types of decision-making are enhanced by real-time analytics?

Generally, they are time-sensitive decisions that also have one or more of the following characteristics:

- High risk or high cost (a customer could be immediately lost or gained)
- Numerous, often conflicting, constraints (the revenue potential must be quickly weighed against the cost of obtaining it)
- Potential for optimization through injection of context data (breadth, depth and the history to enable a more comprehensive, on-the-spot decision)
- Significant competitive advantage (earlier identification of anomalies or opportunities to more quickly limit exposure or optimize gain)
- Increased business efficiency (consistent, optimal execution of frequent actions which individually seem insignificant but collectively have high value/cost)

When applying real-time analytics, it is best to avoid those instances that involve experimenting with new ways of doing business, at least for the time being. To do so involves a great deal of risk (although in certain situations, this may be perfectly acceptable).

Implementing real-time analytics requires the integration of a number of technologies that are not interoperable off

What was unthinkable  
only five years ago  
is now commonplace,  
especially in terms of  
real-time analytics.



the shelf. There are no established best practices. There is a very shallow experience base, and it will take practitioners a little time to come up to speed.

Implementing new business practices is risky enough. Compounding the risk with new technology is not advisable unless:

- As an early adopter, you have developed some facility with it, or
- The opportunity is so strategic that the risk is understood.

Companies that take bold steps can reap large rewards. Some have been involved with process re-engineering for a long time and have created a corporate culture that is very adaptive and not at all risk-averse. If this describes your company, real-time analytics can be an important support mechanism for increased competitive advantage.

Is real time for everyone? Clearly, the answer is no. Other than an unspecified desire to make things faster, what are the motivations to move in this direction? As always, your decisions should put your business first. If you lack business processes that can react to real-time analytics, you'll need to establish those first. Our industry is susceptible to the "tail wagging the dog"; that is, new waves of technology, rather than business reasons and some good old common sense, sometimes drive decisions about investments. But if you have very low-latency business processes in place already, it is likely that you can enhance them.

Are there pent-up and/or emerging sets of business applications that require integrating multi-source historical and near-

real-time data? One thing to keep in mind is that it isn't the availability of data itself that determines whether a low-latency approach is appropriate; it's the application of the data and the business it supports. **T**

*Neil Raden, noted author and speaker, is the founder of Hired Brains, Inc., a provider of consulting, systems integration and implementation services.*

To download the full text of "Exploring the Business Imperative of Real-Time Analytics" by Neil Raden, visit the Library section of [teradata.com](http://teradata.com).



Teradata Professional Services  
1/2 page ad to go here