7 Simple Rules For Successful Real-Time Business Intelligence Implementation

BY Jaymin Patel, Provish Consulting Inc.

Real-time data marts and business intelligence solutions are no longer a novelty or luxury. These solutions are not in their infancy stage. They are becoming instrumental in delivering information crucial for organizations to remain competitive in their core businesses. Given the number of tools mushrooming over the past few years touting anything from ready-to-go solutions on day one to analytical ability on-the-fly and on-demand, real-time business intelligence solutions appear to be a piece-of-cake. However, the reality often reveals a different picture than one might imagine!

What challenges can you face when you consider a Real-time Business Intelligence solution for your organization? The most obvious challenges may include the following:

- Ability to provide analytical information with superior performance while refreshing the data marts real-time.
- Justification of cost versus benefits for real-time business intelligence.
- Impact on performance of one or more source applications.
- Ability to perform all of the transformations in the ETL (Extract, Transform Load) process in real-time.

Well, these responses are not incorrect. However, they are only a small part of the total answer. During the implementation of a real-time business intelligence solution, we realize that the biggest challenges include more than just issues around tools or technologies. So what are these challenges?

1. Be careful in defining "real-time".

In absolute terms, there is no such thing as a truly real-time business intelligence solution! Paradoxically, the term real-time business intelligence is not just some industry hype! The definition of "real-time" is relative to an organization's critical information needs. For some organizations, real-time response may be defined as "within-a-minute", while for other organizations working in situations where instantaneous response is required it may be as small as a few milliseconds. For any given scenario, do not forget to include the time it takes to capture changes in the source system, perform ETL processes and cache frequently accessed reports. Every step in this entire cycle introduces overhead in overall response time. Considering the time taken in each step of data integration and presentation the solution will almost always be accurately called "near real-time". However, most such business intelligence solutions usually deal with information needs that are up-to-the minute or delayed by at most a few minutes. For all practical purposes, there is nothing wrong in labeling them "real-time".

The greater challenge lies in managing the expectations of user communities from the beginning of the implementation. It is essential to ensure that everybody is on the same page when defining "real-time".

For the purpose of this discussion, data mart and data warehouse are considered synonymous.

2. Differentiate between *"must-be-real-time", "nice-to-have real-time"* and *"not-at-all-real-time".*

It is difficult to avoid the misconception that every single piece of information in a real-time data mart has to be refreshed in "real-time".

While most of the information delivered by a real-time business intelligence solution may be categorized as absolute real-time, not all pieces of the puzzle need to be laid out in real-time. You can, rather you must, differentiate between subject areas and/or information elements which are required to be available in an absolute real-time manner from those that can be fulfilled with a daily, weekly, or even a monthly refresh frequency.

Consider a Sales Analytics solution where it is essential to report monthly and Year-To-Date sales by employee on a real-time basis. The data integration engine that populates the sales facts into a real-time data mart needs to perform this task as soon as the source application reflects such changes. However, what about the refresh frequency of the employee dimension?

Let's assume that some attributes of the employee dimension are derived from a Human Resource (HR) application that reflects the changes occasionally and that those data elements are not critical for sales analysis. In this scenario, it may make sense to break down the employee dimension into two components. Derive absolutely necessary information (for example, employee's ID and key) real-time from the transaction system and "back-fill" other attributes like middle initial and office location on a nightly basis from the HR application. This kind of an arrangement will reduce the load on the real-time ETL process.

Similarly, if you require some snapshot facts only for monthly reports, build those snapshots only once a month.

A system cannot win a sprint with a fifty-pound sandbag on its back! You need to design a leaner set of data to integrate in real-time. Drawing a line between each refresh frequency and deciding which information should be real-time and which can wait is a key for a successful real-time data mart solution. These requirements need to be tackled early during the analysis and high-level design phase.

3. Don't build a bridge when a boat would do.

The simplest way to capture data from any source application in a real-time fashion is to actually read the source system real-time. However, not many OLTP application managers or architects may not allow that to happen! Hence the choices available to access source application changes are limited.

Most of the real-time integration tools have a real-time data replication engine separate from the batch integration engine. Such engines often come with a separate price tag. This real-time data replication engine leverages database logs in order to do its work. While such tools make life easy for the development team, database native replicators offered by database vendors can prove to be a simpler and cost-effective solution.

Alternatives to such real-time data capture can be designed through database triggers as well. If this option does not tax the source system in terms of performance and maintenance, it is a better candidate than the separate toolsets since both cost and maintenance will be lower. Such triggers can post changes captured in source system into a separate staging area and the real-time integration engine can read continuously from the staging area without affecting the performance of the source systems.

Most likely, the volume and refresh cycles of the data that need to be integrated from various systems can be easily handled by standard ETL refresh strategies including database triggers, replication or some other mechanism. Even if there is a latency of a few minutes in the alternate mechanism, consider them over implementing separate real-time replication tools. The cost and complexity involved in implementing specialized tools solely for the purpose of replicating data may not justify the benefits received by providing the information a few seconds early.

Don't build a bridge when you can cross that river in a boat!

4. A picture is worth a thousand words (especially when you are refreshing live).

The key imperative of a real-time business intelligence solution lies in delivering information "real-time". What better way is there than to provide graphical visualization. Due to increased awareness about the importance of data visualization in business intelligence, analytic dashboards are no longer considered a "nice-to-have" tool. While numeric and statistical information remain essential to completely represent the facts, dashboards are the perfect tool for providing a bird's eye view of the enterprise or a particular functional area.

Consider these two representations of the same information:

Sales Facts	May-04
Actual month-to-date Sales (USD)	87,816,300
Projected Sales for the month (USD)	*112,585,000*
Most optimistic Forecast (USD)	*140,731,250*
Most pessimistic Forecast (USD)	*84,438,750*
Totally unacceptable (USD)	*<= 45,034,000*

Month-to-Date Sales as % of Projection
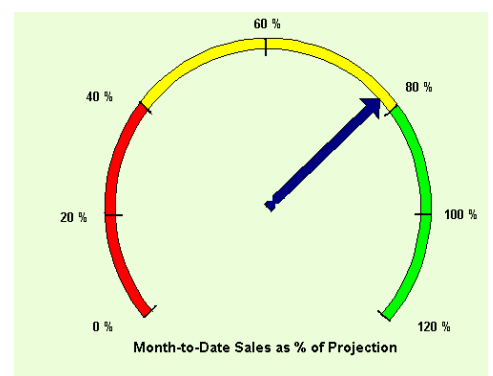
While both presentations are useful, the second one definitely delivers the overview instantly. When the sales facts are changing near real-time, the dashboard visualization would be easier to interpret compared to plain statistics. For accurate presentation, users can be given "click-through" access to the detail information from the dashboard dial.

5. Placing the information right.

Most real-time business intelligence initiatives start with an objective of delivering the right information at the right time. That correctly describes the importance of the timeliness and correctness of information. However, where should that timely and correct intelligence be delivered?

Even though a real-time business intelligence solution may deliver up-to-the minute information through traditional reports and dashboards, it may not be enough. For a senior sales vice president who barely has time to log into his laptop to access the information when he is on the go visiting his key customers, information is as good as not available. For such users, the solution cannot be called "real-time" in the truest sense unless it is also delivered to the right place where it is easily accessible.

Delivering the right information to the right place at the right time has never been easier with the latest communication gadgets and technologies. Besides the traditional desktop reporting interface, the real-time solution can deliver dashboard reports or exception alerts to PDAs or similar web-enabled devices. Gone are the days when "Information at your fingertips" was the punch line, the new mantra is "Information in front of your eyeballs "!

While technology is readily available, the art remains in defining the critical exceptions and providing concise alerts or selective information based on the users' differing information needs. Attention needs to be given to such analysis during the requirements gathering sessions with users at various levels in the organization. Such consideration will reveal personalization as well as security requirements based on each user's roles and responsibilities within the organization.

6. Can you please all the people all the time?

Data mining and satisfying ad-hoc analysis are key features of the traditional data warehouse. At the same time both these types of analyses impact database performance. For a real-time solution this is a primary challenge because the solution has to perform both the data refresh as well as deliver reports in real-time. If a user hits the database with a random query joining all dimensions to fact tables for the last five years of data while the ETL process is performing key real-time updates, the database performance will deteriorate. Remember that the limits of the data mining and ad-hoc analyses need to be defined during the requirements phase.

If you contemplate the objectives of data mining and ad-hoc analyses, it is obvious that neither of the two require up-to-the minute information. When users are interested in researching the trends over a few years or finding the correlation between two discrete attributes or even investigating a particular instance, they can utilize a snapshot as of yesterday or even last month. A pragmatic approach would be to maintain a separate environment that delivers information for such uses from a back-up of the data mart database instance and not the real-time data mart solely. Without this arrangement, it would be a challenge to please all of the users all the time!

7. What prevails in a real-time solution? Operational Intelligence or Analytical Intelligence?

The biggest myth about business intelligence solutions is that they fulfill analytical reporting needs while the core transactional applications deliver operational information. In reality, a typical business intelligence solution must satisfy both kinds of needs with equal ease. Real-time business intelligence is no exception. Dashboards and alerts are not meant only for senior management. In a typical organization, the majority of the analytical reporting and research initiatives are performed by tech-savvy middle managers. At the same time they are also responsible for monitoring the performance of their sub-units on a real-time basis. The slightest deviation from the plan can result in a large impact on the bottom line. These middle managers are the heaviest users for both the operational and analytical reporting capabilities.

With real-time changes reflected into business intelligence solutions, a whole new platform is open to the users where they can access up-to-the minute operational information. This is superior to traditional business intelligence solutions where users have to wait for a day or more in order to get the same information as the source application contained. Careful consideration to operational reporting requirements during the requirements gathering phase is essential to deliver the highest return on your real-time business intelligence solution investment.

After all is said and done, the solution delivery team often tends to focus more on tools and technologies. Some of the above areas tend to get overlooked during the requirements and design phases. The next time your organization needs real-time business intelligence, take the seven challenges outlined above Into consideration.