

# Exploring the Business Imperative of Real-Time Analytics

by Neil Raden,  
Hired Brains, Inc.

October, 2003

**HIRED BRAINS, INC.**  
**Implementing Business Analytics**

# TABLE OF CONTENTS

<b>INTRODUCTION</b>	<b>3</b>
WHAT IS REAL TIME DATA?	3
WHERE DATA WAREHOUSES FIT	4
ACTIVE DATA WAREHOUSES AND REAL-TIME DATA WAREHOUSES	5
FROM BATCH TO REAL-TIME	5
<b>IS REAL-TIME FOR EVERYONE?</b>	<b>6</b>
<b>CONSIDERATIONS</b>	<b>8</b>
WHERE AND HOW TO LOOK FOR REAL-TIME ANALYTICS OPPORTUNITIES	8
THINGS TO AVOID	9
<b>REAL-TIME ANALYTICS: EXAMPLES</b>	<b>9</b>
DYNAMIC PRICING	10
YIELD MANAGEMENT	11
FRAUD DETECTION	11
PERSONALIZATION	11
"LITTLE" ANALYTICS	12
CALL CENTER	12
RULES ENGINES/AUTOMATED ATTENDANTS	12
<b>CASE STUDIES</b>	<b>13</b>
NORFOLK SOUTHERN RAILROAD	13
CONTINENTAL AIRLINES	14
<b>CONCLUSION</b>	<b>14</b>

## Introduction

Market forces, customer requirements, statutory mandates and, especially in recent years, technology, fuel an endless drive for organizations to improve their operations. Like poor Alice running with the Red Queen, struggling to keep up with the moving background in *Through the Looking Glass*, constant forward motion is necessary just to not fall behind. The emerging reality in the past few years is that equilibrium is just not possible - a true steady state no longer exists. As a result, forward-looking organizations are moving to transform themselves into truly agile enterprises, capable of constant change, not just periodic realignment. In this emerging distributed, collaborative and intensely competitive business environment, the goal is to become truly intelligent enterprises, able to forecast with accuracy, execute with confidence and to make adjustments, large and small, at the speed of business.

This is the imperative for real-time analytics, the ability to use all of the resources at your disposal, including data, to improve your operations and quality of service, at the moment they are called for. If, at the moment (or very soon after) a piece of information is created or modified in an operational system, it is sensed and acted upon by an analytical process, real-time analytics have occurred. No longer the domain of operations research staff behind the scenes, analytics is moving to center stage. Real-time analytics complement real-time operational systems. Static reports of month-old data, endless meetings postulating whether to do this or that, retyping data into spreadsheets - there is no time for that. Agile organizations will need to measure, evaluate and react to events with a closed-loop of telemetry-like information, rules, decisions and triggers, all in real-time.

For some organizations, this flow of data will be an enormous, steady stream. For others, it may be just a trickle and in still others, punctuated bursts. Though the frequency and amplitude of the data flow may vary, what is constant is that much of this data has real value, but a very limited shelf life. The latent value of real-time data is lost if it is not exploited within a very short time. Understanding what data is valuable and how to use it quickly are the most important steps in developing the facility to deal with real-time data. More importantly, developing a facility for this process can lend itself to an enduring competitive advantage. The technology for capturing real-time data is available to everyone. Learning how to apply it effectively is a competitive differentiator.

The purpose of this paper is to explore the emerging need for real-time analytics, examine their key characteristics, and suggest likely opportunities for them. The specifics of designing and implementing real-time analytics are beyond the scope of this paper.

### ***What is Real Time Data?***

Clearly, no data is truly real-time, not in a quantum sense - the moment you view it, it is no longer real-time. But where is the dividing line between real-time and batch? The answer varies by situation, but some rules of thumb, some qualitative, some quantitative, apply:

- For analytical applications, more frequent than daily can be considered real-time, because it crosses the overnight-update barrier
- Whenever operational (source) data can flow to an analytic process without delay, or can be made available with the rest of the previously stored analytic data without a break in service. This is often called "trickle-feed" instead of batch load
- In a practical sense, real time is defined by the service level agreement required of the business process involved and will vary from company to company and from application to application.

Real-time data has some unique characteristics. To an even greater extent than higher latency data, daily or weekly for example, real-time data can only be consumed in combination with other components. Models, history, context and visualization are critical. In many cases, the real-time flow needs to be trapped and handled by rules engines, event brokers and other forms of automated attendants. There isn't time to ponder

over new information as it arrives, in most cases, so it needs to be bundled with a set of added-value features that make it instantly useful.

One common metaphor used to depict how people will interact with real-time data is the dashboard or cockpit, a form of visualization. The concept is that data flows from operations and is measured and displayed in dials and gauges for the worker and executive to watch and take appropriate action. Dashboards are a useful metaphor to describe the use of data in a flow rather than batch. However, dashboards are just one possible application, and the metaphor should not be taken too literally. There are some subtleties that are masked by the accessible visual display:

1. Gauges rarely display raw data at all. Consider a speedometer. The actual raw data is the number of revolutions of a wheel, data that is converted to mph by a *model*, in this case, a model that uses a constant value for the circumference of the tire. This illustrates why real-time data needs to be manipulated by models before it can be understood.
2. Second, unlike drivers, and especially pilots, workers are not fixed in a seat looking at the dashboard. Real-time data has to be evaluated by machine processes and its presentation customized for the work process it is meant to support. Real-time analytics do not allow for iterative exploration of alternatives - there isn't time. The capture of real-time data without real-time analytics is like drinking from a fire hose.
3. A dashboard needs to be connected to decision process. The action to take is based on rules, either applied manually by inspecting the dashboard (sales are not on target for the quarter, time to call the sales managers in) or through the use of a rules engine (manufacturing sequences cannot be reconfigured for back-ordered parts, execute contingent sourcing). Once a decision of this "automated attendant" is reached, "closing the loop" requires output routed to an event handler that is programmed to react to the conclusion of the rules engine. Not many data warehouses have this infrastructure in place.
4. Instruments lack history, they are typically instantaneous representations of a continuous process. Models, evaluation and context are the domain of business intelligence.

As appealing as the dashboard metaphor may be, remember that real-time analytics cannot exist without the integrated data and historical context that a data warehouse provides.

### ***Where Data Warehouses Fit***

Integrating real-time information with persistent, historical facts and reference data is Real-Time Data Warehousing, an absolutely essential part of real-time analytics. Simply flowing instantaneous data to waiting devices is nothing new. Industrial process control, mechanical telephone switches, even the thermostat in your house have all done this for decades. What separates real-time analytics from these process automata is the idea of *context*. In a process control stream, the system may be aware of the "boiler" but it probably does not know who the manufacturer is, what the maintenance history is, how long it has been in service (in point of fact, some state-of-the-art process control systems do have this information, but they are examples of real-time analytics in action). A useful example of the need for context in real time is how an airline decides to re-route customers with connections arriving on a delayed flight, reserving the few remaining seats for the "best" customers and making other offers based on loyalty, lifetime value, policies and the current flight situation. Not only does this require real-time data (multiple agents making these decisions simultaneously, with each decision reflected in the data as it occurs), it must join this information with historical information about the customers and apply rules against detailed and derived attributes about those customers. Only through the use of rich contextual data, the kind kept in data warehouses, can real-time flows of information be lined up with all that the organization already knows. This is much more robust decision-making than simple process control, this is business intelligence.

No organization can actually view, evaluate and act on all of the detailed data (theoretically) available to it. Rather, the data needs to be managed by back-office processes, such as transformation, cleansing, summarizing, modeling and even presentation. Companies that have implemented real-time analytics, described later in this paper, have all recognized that low-latency data is the input, but the models and

triggers that they built are the real assets. In most cases, implementing these systems involves substantial reworking of existing work processes and, in some cases, the creation of new products, markets and industries. It takes a huge commitment, it is not an out-of-the-box effort.

In almost every case, however, leveraging existing data integration projects, especially data warehouses, is key to successful implementations. Certain companies, like Teradata, already provide leading-edge solutions, like the Active Data Warehouse, which fits perfectly with real-time analytics.

### **Active Data Warehouses and Real-Time Data Warehouses**

It is important to make a distinction between Real-Time Data Warehousing (RTDW) and Active Data Warehousing (ADW). Though 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. RTDW concepts include physical modifications to the database schema and the database environment, movement of data across the enterprise, ETL 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 an architecture or a methodology, rather, it speaks to the role the data warehouse plays in the enterprise. The 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, such as CRM, ERP, SCM (Supply Chain Management) and any other system operating on-line. For example, an ADW may be part of a fraud detection application, providing quick historical analysis or pattern matching upon request to an on-line banking application. This ability to communicate in real-time, with or without human intervention, is a bold step, vastly different than the current practice of simply sourcing data for ad hoc queries, OLAP and reporting. Timeliness is only one aspect of ADW. Also inherent are critical things like scalability, availability, and manageability features to enable responding quickly to ever-changing workload dynamics.

### **From Batch to Real-Time**

In its original formulation, the concept of data warehousing architecture was based on the bedrock of batch processing. The whole concept was to build, incrementally, a repository of integrated, historical information to support an enterprise's needs for reporting and analysis. There was no sense of urgency. First came the extract, cleanse, transform, load, index, aggregate, subset and a host of other back-office activities. These steps took a considerable amount of time, especially when source data was scattered in often dozens of homegrown applications. This limited the up time and lengthened the update cycle time so severely, that most early data warehouses were only refreshed once per month. The widespread implementation of ERP and other enterprise packages, along with the emergence of ETL tools, alleviated much of that over time by reducing the number of source systems. As a countervailing effect, though, the data warehouses grew larger and larger, and the update cycles remained considerably long. As capabilities, tools and expertise improved, these refresh cycles dropped to weekly and, eventually, to daily, which is the norm today. But moving from a daily cycle to less-than-daily is a quantum leap. In a batch-oriented architecture, it is difficult to imagine how to shorten this cycle any further. The data warehouse often depends on the daily batch cycling of other operational systems, a problem that has to be dealt with before any real-time processing can occur. A batch process implies taking the data warehouse out of production for a period of time during the working day, which is somewhere between inconvenient and untenable in most organizations. In intra-day updates, there is no time to quiesce the database while it is being refreshed.

The technology to make the leap from batch-oriented to continuous, real-time update of a data warehouse is already available. It requires, however, that we take a look at our methodologies and processes closely. The

technical components -- databases, ETL tools, etc. -- are mostly up to the task already, but most data warehouses are still based on high-latency approaches.

## Is Real-Time For Everyone?

Is real-time for everyone? The answer is clearly no. For some organizations, there may be a pressing need to implement this capability, but culture or crisis can obscure it. In some cases, there may be more critical needs elsewhere and no resources to apply to real-time. There are those who resist any technology innovation unconditionally, whether it's good for the organization or not. The goal of this paper is not to convince the reader that moving to real-time is the highest priority, but rather, to describe how and why it can be applied. Other than an unspecified desire to make things faster, what are the motivations to move in this direction? Are there pent-up and/or emerging sets of business applications that require integrating multi-source historical and near-real time data? How do you know if your organization needs these capabilities? 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.

It is easy to find business reporting or analytical processes that might be greatly enhanced with more timely data, but how do we know for sure? Would it really matter if brand managers had sales data in 15-second increments during in-store promotions? Is it possible for an organization to react quickly enough? Does it matter if the continuous replenishment system can recalculate the shipments every 10 minutes if the truck only leaves once per day? Clearly, some requirements for more timely data need to be evaluated for bona fide business value. But there are many applications, where the ability to include up-to-the-second analytics is extremely useful (see Table 1). What is important to understand is that the value in merging operational and analytical processes in real-time has to be found in the business processes that leverage it. In other words, don't necessarily look to your existing applications for candidates, the best application for the technology might very well be the things you could never do before.

INDUSTRY/AUDIENCE	TYPE OF ANALYTICS
Financial Services	Credit card fraud Risk Management (Market, credit, ops)
Energy	Forward price projection Dynamic triggering
Retail	Pricing optimization, replenishment, markdown and inventory management
Government	CDC monitors Rx for epidemic outbreak; Anti-terrorism port traffic
Consumer Products	Promotional effectiveness Spot promotions
Web Commerce B2B or B2C	Targeted incentives at touch point Collaborative filtering/customization
Finance	Instant financial close Continuous planning
Executive	Real-time dashboards BAM
Healthcare	Service level optimization

**Table 1: Real-time Analytics By Sample Industry or Audience**

Let's look at how real-time analytics power some of the solutions in Table 1.

**Government:** The volume of information flowing from port traffic is enormous and no individual or group of individuals could correlate the data and spot patterns in real-time and alert authorities to possible terrorist activities. The assistance of automated attendants, pattern recognition algorithms and event brokers are essential for dealing with information at this volume and speed. This is real-time analytics in action.

**Consumer Products:** A consumer products company with two thousand sales agents needs to place and price products at outlets with precision, offering on-the-spot discounts and rebates, program payments to distributors and deal with back orders and a myriad of other "small" issues that happen in the course of a business day. Promotional effectiveness is the Holy Grail of this industry, where companies spend billions each year and are never really sure if it helps or not. The ability to assist the field force in making the best decisions, while constantly adjusting the big picture to achieve system-wide optimization of resources and revenues is a problem that can only be solved in real-time.

**Web Commerce:** B2B or B2C commerce over the web is the perfect venue for customization, using tools like collaborative filtering<sup>1</sup>, cluster analysis<sup>2</sup>, product affinity and other forms of personalization to modify the experience for each interaction.

**Executive:** BAM (Business Activity Monitoring) is a generic term for real-time interaction with business processes. Watching data flow by is of no value. Real-time analytics must be applied. Consider tickertape. No one watching tickertape is just looking at data, unless they have other tools at their disposal, like a trader's workstation for example. They have a model in their heads. What stocks are they watching, what movements are important to them, in what combination with other stocks are these movements compelling, what triggers or collars have they decided on to manage their portfolio? No real-time data is useful without models.

The power of Teradata's ADW is to provide the environment where rich, deep data, integrated from many sources, transformed and cleansed and stretching back into years of history can be updated nearly instantaneously and where queries to support real-time analytics can be executed. The context for decision-making is provided by the ADW. Specialized tools, such as rules engines, can react to conclusions drawn from the ADW.

Figure 1 depicts, at a very high level, how an ADW might participate, in real time, with operational systems. In this example, it both provides information instantaneously to other modules, such as a Pricing Engine, as well as receives and integrates data in real-time from rapidly changing sources, such as updated demand forecasts and completed transactions. One could argue that the "analytics" are mostly performed by the Demand Forecast and Pricing engines, but in fact, those engines could be implemented within the ADW itself. In either case, it is likely that calculated parameters from the ADW are crucial to the process and would include metrics that only the ADW could provide, such as those based on:

- Historical context, trend analysis, continuity of reference data such as discontinued product ID's
- Integrated information that derives from multiple, conflicting source systems whose consistency is the domain of the data warehouse
- Detailed data, at the transaction or sub-transaction level, whose attributes cannot be aggregated and must be queried at the detail level, such as credit card transactions above or below a certain amount as a constraint.

---

<sup>1</sup> Collaborative: Filtering accumulates a database of consumers' product preferences, and then uses them to make customer-tailored recommendations for products. The user's preference can be either explicit votes or implicit usage.

<sup>2</sup> Cluster Analysis is collection of statistical methods that is used to assign cases to groups (clusters). Group members share certain properties in common, from which predictions are made about their behavior.

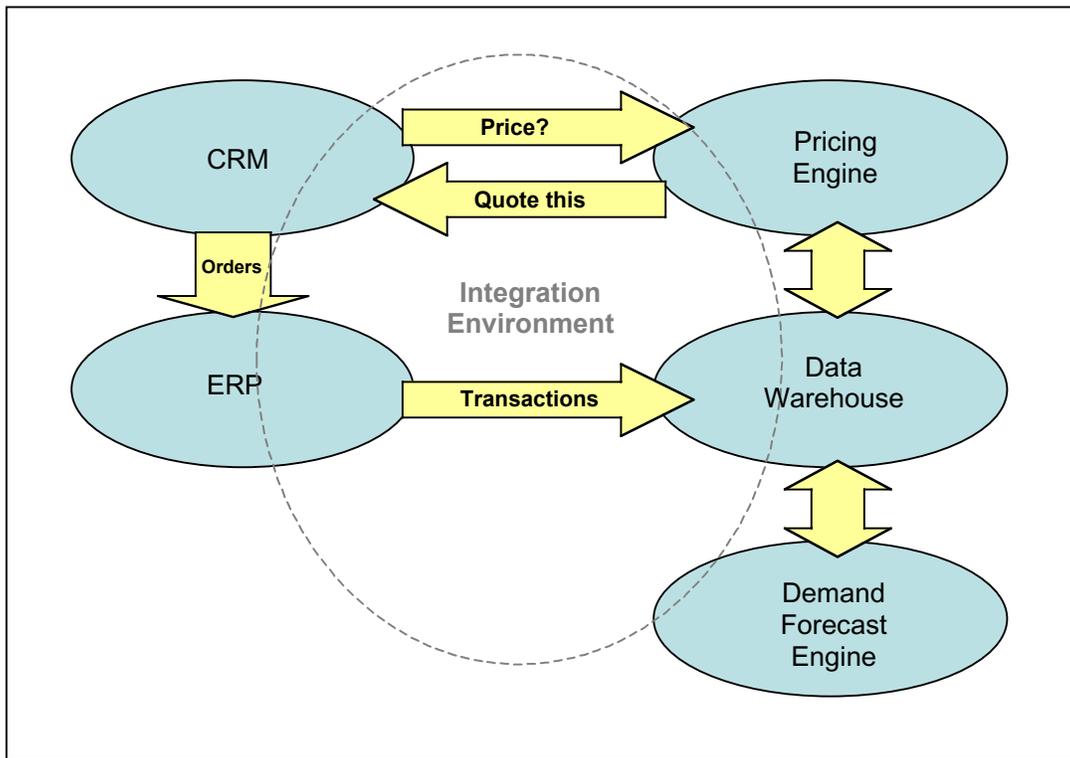


Figure 1: An Active Data Warehouse in Action

## Considerations

### ***Where and How to Look for Real-Time Analytics Opportunities***

In evaluating opportunities where real-time analytics can provide measurable value, at this point in time, 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, which 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, these are 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, history to enable a more comprehensive decision on-the-spot)
- Significant competitive advantage could result (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)

In organizations, there has to be a low-latency process in place to absorb the outputs of a real-time analysis. For example, by sifting through sales orders, trend analysis and predictive analysis, it may be possible to optimize, at any instant, the palletizing and loading of products for shipment. However, if the communication to the dock operates in a loop of 6-8 hours, it is wasted effort to flood the channel with this stream of information. In this case, where your own internal processes dictate the latency, the alternatives are to attack the process first, if there is a business-driven reason to do so, or to apply the technology only to the extent that it facilitates a smooth closing of the loop. Implementation of successful real-time analytics is typically accompanied by process re-engineering, the total result of which is better decisions, cost savings and new business opportunities.

When implementing a real-time analytic process, it is not uncommon to uncover underlying or subordinate business and analytical processes that must be addressed before the candidate real-time analytic application can be implemented. For instance, a retail marketing department may want to create a real-time analytic application to provide an extra-value service to loyalty program customers at the checkout counter. However, if current checkout processes can't sustain the additional burden, it will be futile to develop the real-time application until the existing checkout processes are modified or streamlined to accommodate the additional service.

### ***Things to Avoid***

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 (though 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-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 who take bold steps can reap large rewards. Some companies 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.

Once your organization gains some experience with the discipline and finds the right balance of technology and human interaction, new opportunities will arise. To start, it is best to reduce the variables and focus on workflows that are already well documented and understood, and operate well already. Try to find situations like that and apply this technology to enhance them.

## **Real-Time Analytics: Examples**

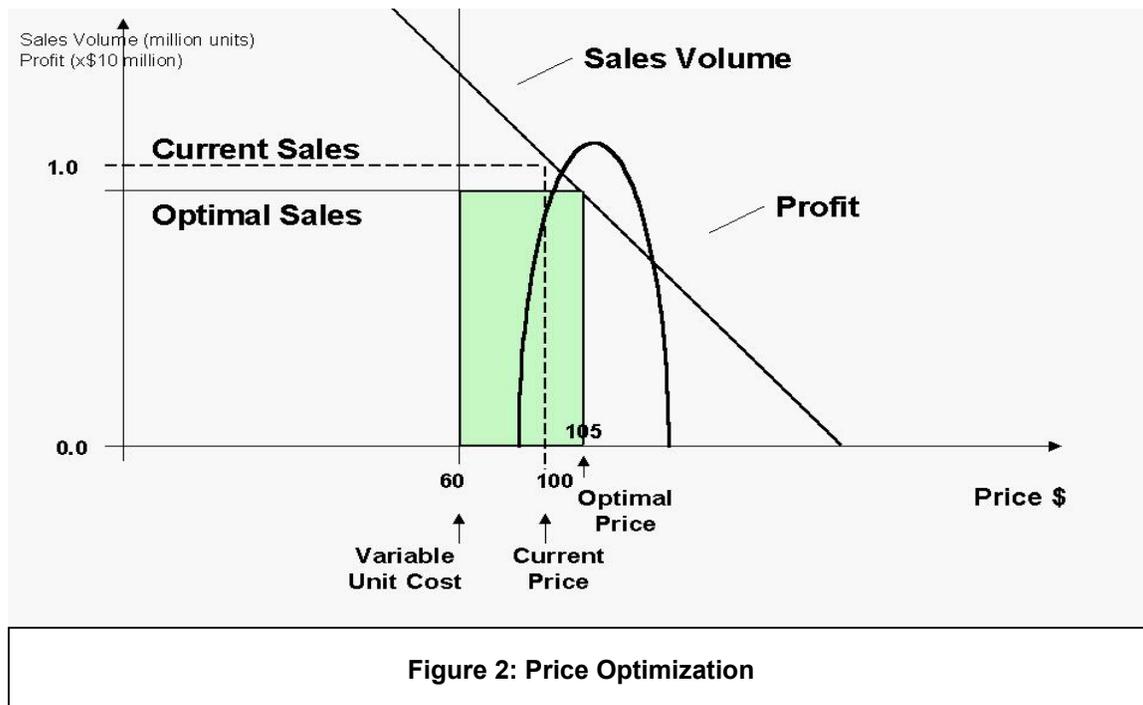
In order to add some real-life examples to the discussion of real-time, this section will describe the operation of a number of existing real-time analytical processes. In most cases, these are implemented as purely proprietary processes, such as airline Yield Management, or as industry-specific applications operating in a stovepipe manner. It is not the current architectural implementation that is of interest, but rather, the business requirements they serve. In most cases, they all would benefit from re-platforming in an ADW environment, where they undoubtedly will move in time.

## Dynamic Pricing

Dynamic pricing is a generic term to describe any process that adjusts prices at or near the point of sale. The concept behind dynamic pricing is to, obviously, maximize profits, however, there are often conflicting constraints. For example, offering a returning frequent customer an instant 50% off coupon will actually lower the revenue of that single sale, but the incentive is applied for a host of different reasons to maximize overall profit. The whole concept of promotions and advertising is based on the idea of maximizing profits over time, or of reaching certain volumes to reap economies of scale, not necessarily boosting the margin of each transaction (this is why these activities seem to sometimes defy logic). Dynamic pricing models can be almost infinitely complex, such as an airline yield management system, or relatively simple, based on the application of pricing schedules and evaluation of criteria at the moment of sale.

Regardless of the specific case, the concept behind dynamic pricing is the same, naturally: to maximize profit by making the best possible pricing decisions at the moment they are needed. Consider for a moment a simple business. In the graph below, we depict a demand curve against which we can plot various combinations of sales and prices. Assuming variable costs per unit are reliable, it's easy to see that, in this case, the optimal price is actually higher than the current price, even though it will generate fewer sales. The reason is that the total profit is greater at the lower volume.

There is another, more subtle lesson that can be derived from this simple model. Notice the shape of the profit curve, an inverted parabola. While all of the other curves in the model are flat, this one is tall and narrow. The correct interpretation is that being away from the optimal price puts you, very quickly, on a slippery slope, the farther you are away from the apex, the slipperier it gets. In other words, the farther off the mark your pricing is, the quicker will be your acceleration away from maximum profit.



Dynamic pricing is a perfect example of the need for rich, historical, contextual information, that can only be sourced from a data warehouse, enhanced with instantaneous data streaming into systems in the normal course of business. The ability to bridge the operational and analytical environments in the ADW is a crucial factor in building real-time analytics.

## ***Yield Management***

Yield management is a type of dynamic pricing and is an umbrella term for an approach for maximizing revenue in service industries that are limited by capacity, such as airlines and the hospitality industry. Using an overused phrase, the whole point of yield management is to provide the right service to the right customer at the right time at the right price. A fixed inventory and instantaneous "spoilage" characterize these problems. Hotels cannot add or drop rooms, with an attendant change in inventory cost and an airline cannot add or drop seats at will from a given aircraft. In addition, pricing in these examples is fairly elastic, which allows for a fair amount of tinkering. British Airways realized this almost thirty years ago when they began to implement one of the very early successful yield management systems, the Origin and Destination system. It has been rewritten and re-platformed a few times since, but its basic concept remains the same.

The starting point of a yield management system is forecasting demand. This is a continuous process, with each transaction being taken into account, continuously. The reason the ADW is needed is that tickets are not sold all at once. If they were, customer behavior would not be a factor. The tradeoffs occur when a manager is faced with the option of accepting an early reservation from a customer who wants low price, or waiting to see if a higher paying customer will show up.

Demand is seasonal (so we need lots of historical data). Unlike forecasting shipments of mayonnaise in cases by month, the yield management problem is extraordinarily complex. In formal terms, it is best described as a non-linear, stochastic, mixed-integer mathematical program that requires data such as passenger demand patterns, cancellations, group reservations, cargo load, and other estimates. Solving this problem would require approximately 250 million decision variables! For the sake of feasibility and time, implementations of this problem have been reduced to three smaller ones: overbooking, discount allocation and traffic management. If an airline only had point-to-point routes with no connecting passengers, traffic management would not be an issue in this discussion because for a given flight, overbooking level and discount allocation problems alone could be used to maximize revenue.

Yield management is a mature discipline and very rich in theory, but there is only space to describe its most common features. In summary, yield management is an exotic application, extremely costly to design and implement, creating a barrier to entry to all but the largest concerns. Real-time analytics and the ADW will lower that barrier by allowing smaller and less well-endowed firms the ability to develop these kinds of processes piecemeal, over time.

## ***Fraud Detection***

Real-time analytics can significantly reduce fraud exposure and the effects of fraud. For example, credit card companies have created sophisticated analytic environments to detect fraud early and take specific actions to minimize the impact. It is not uncommon event to receive a call from your credit card company when you have credit transactions that fall outside the range of expected patterns, or to experience an inconvenient rejection. Being able to correctly identify fraud, but not offend a well-intentioned valuable customer is a critical necessity. Rejecting a valid transaction due to ineffective fraud detection could offend a high-value customer, but not detecting a truly fraudulent transaction could cost the credit card company thousands of dollars. Weighing all of the potential fraud indicators quickly within the unique context of each customer is an obvious opportunity for real-time analytics.

## ***Personalization***

If we anthropomorphize business transactions, then we can say that personalization is the process of making each customer feel special. In practice, personalization really means customizing each customer touch-point based on the known characteristics of the customer, an association of those characteristics with certain behaviors and, lastly, an "offer" that aligns those expected behaviors with a desired (on our part) outcome. If these transactions occur more than a few times a day, it is beyond the capability of any person to manage these variables without some assistance.

One could argue that a CRM system and a "personalization engine" should be sufficient for this purpose. In case after case, it has been shown that only the data warehouse has a reliable set of information, including historical contexts, reference data, integrated information from many sources and an available platform for nearly instantaneous analytical work. But a CRM system that does not, by definition, include a data warehouse, cannot, even with a personalization engine, perform the bulk of the requirements because it lacks the historical and enterprise perspective.

Once again, it is the ability to join the operational flow of data with the context of history and rich reference data that makes the active data warehouse an indispensable part of real-time analytics.

### **"Little" Analytics**

Yield Management, Fraud Detection and Personalization are all well-known examples of advanced analytics. There are requirements for emerging real-time analytics on a far less complex scale: they are a necessary part of the "M" initiatives: CPM (Corporate Performance Management), BPM (Business Process Measurement) and BAM (Business Activity Monitoring). The latter two in particular imply a real-time analysis, decision and action environment. "Little" analytics may seem insignificant singularly. The measurements themselves may not be very exotic. In fact, they are likely to be mundane. As one brand manager with a consumer products company said, "Our success will not be based on developing a grand strategy once a year, it will depend entirely on making thousands of small decisions every day, and making the right ones." In the CPG industry, those decisions can be: dealing with allocation of scarce stocks, promotional effectiveness, revising routes, managing royalties, reacting to competitive intelligence, local advertising. In the past, these decisions were made by formula; broad strokes that were often ineffective. The inefficiency and cost implications of that approach are no longer viable, especially in high-volume, low-margin businesses.

### **Call Center**

90% of the activity in a call center concerns issues of that business day. Call center software (part of CRM) uses analytics to route calls, help customer service reps with scripts based on the type of call, measures volumes and allocates resources and a host of other useful functions. If the data available to the call center software is a day old, the usefulness of the application can be reduced by, roughly, 90%. This argues for a real-time set of data, which should be available through the actual call center operations software. However, like all other metrics, the right set of data is usually a combination of current day facts, data from multiple systems that surround the call center, such as warranty, order entry, billing, etc., as well as the sort of historical context and reference data only a data warehouse can offer to resolve the issue or inquiry on the first call. This argues strongly for an ADW approach.

Call center real-time analytics are a rapidly growing field because near-real-time facts combined with broader data are the bedrock for this kind of decision support.

There are some complications, though. Travelocity, for example, uses their own internal CRM systems, but their on-line system for looking up and booking is SABRE, which is not their own. Integration of outsourced online systems in real-time is tricky. Nevertheless, by taking certain shortcuts, such as calculating the LTV (Lifetime Value) of a customer in the ADW ahead of time, they can route calls from customers with high LTV's to more experienced agents. Likewise, using real-time analytics from the ADW, they can make more consistent decisions on-the-spot, giving agents the authority to, say, issue gift certificates, refunds, coupons, etc.

### **Rules Engines/Automated Attendants**

Strictly speaking, rules engines and automated attendants are not part of real-time analytics. The analytics engines are used to drive models and metrics, or to formulate decisions. The rules engines and attendants actually implement the decisions. In a real-time, closed-loop environment, though, they all work so closely together that the boundaries are not perfectly circumscribed.

A rules engine can mean extensive, hand-coded application code or it can mean a package designed for automating the bulk of the development and maintenance efforts. The effect is pretty much the same - a system for applying business "rules" to business "objects." Most commercial packages today are designed to run in a J2EE or .Net environment, come with connectors to the major databases and often enterprise applications and typically generate code in Java, HTML, XML or JSP, among others. It is vitally important, of course, that the rules engine connects cleanly with the operational environment, so that it can recognize events as they are presented, evaluate them quickly and pass messages to the other listening processes, such as automated attendants.

An automated attendant is just another way of saying the analysis has been done and the resulting information passed to a system for further action, rather than a human. For example, the rules engine may tell the event handler, "wake up the BI software, have it run a 2-brand share report on these two entities, attach this Press Release, and email it to everyone of List 222, marked, 'Urgent: Price Event.'"

Humans are not capable of responding to every relevant event. Dashboards are used to summarize information and are presented to people for an overview. When individual events in real-time need attention, rules engines and automated attendants are applied to the extent that they can be. Those decisions which require further analysis must be dealt with using more sophisticated real-time analytics.

## Case Studies

Some Teradata customers have already employed successful real time analytics using the ADW. The latest product release made huge advances toward enabling these organizations to squeeze the latency out of their analytics and unlock real value in their data warehouse investments.

### **Norfolk Southern Railroad**

Because it is possible to mask the actual workings of real-time analytics from the people who benefit from them, to increase the usability of systems by making the underlying complexity disappear, it is possible to apply the technology in customer-facing situations. Real-time analytics are not just for your employees but will, in many ways, be used extensively for reaching audiences far beyond the firewall. Norfolk Southern Railroad is doing exactly that. Shippers have a critical need to understand the location and expected arrival time of their shipments and Norfolk Southern provides over 1,500 external customers (and 1,000 internal ones) web-based access to information no more than one hour old.

Prior to the implementation of the ADW, customers called in via telephone and operators retrieved information directly from the OLTP systems. The growing business and the desire to cut costs without reducing services led to a change. The decision was made to move the application to a data warehouse to combine information from multiple systems, to reduce the load on the operational systems and to provide the scalability to service thousands of users. This happened in stages, but once the hourly data load became operational, the system was truly capable of providing real-time analytics.

From the external customers' point of view, this information is already extremely useful. In the future, shippers may wish to pull this data as a feed and apply their own models and analytics to it for a wide variety of uses such as dynamic scheduling, precision pricing of shipping costs, etc. Norfolk Southern has already begun using the data for their own internal pricing and optimization analysis.<sup>3</sup>

---

<sup>3</sup> Norfolk Southern reference from [www.teradata.com](http://www.teradata.com)

## Continental Airlines

Continental Airline's Alicia Acebo was recently interviewed by Information Week about the active data warehouse environment for which they won a best practice award from The Data Warehouse Institute. Following is an excerpt from the article that describes the real-time nature of their implementation:

*"As user demands for more up-to-date analysis have increased, Acebo's team has adapted the data warehouse to operate on a near-real-time basis. The mainframe and Cobol tools that originally handled data transformation and loading chores have been replaced by a faster network of Windows servers and custom-built C++ tools. Now users analyze flight operations and reservations data that's only seconds old. "Business users at Continental are very savvy. They always want more [information], and they want it faster," Acebo says.*

*The data warehouse's real-time architecture and automated data-transformation capabilities are two of the best practices that brought Continental the award. The system's design simplifies combining data from different operational areas, making it easy to get a single view of a customer, for example. Standard data definitions developed by a steering committee and used throughout Continental make such cross-functional analysis easier. And revenue and profitability are factored into any decision, no matter how minor."*<sup>4</sup>

## Conclusion

In all organizations, private and public sector, profit and not-for-profit, small and large, performance management and process control will continue without the implementation of real-time analytics. Many decisions based on "classic" analysis will occur naturally with a great deal of latency. The movement to real-time enterprises has no effect in these areas. On the other hand, each industry has its own specific set of business requirements, any of which already are, or soon will be, driven to a low-latency effort simply due to market pressure. In addition, creative entities will find ways to use the technology to get a leg up, which will force others to adopt the same approach. Deciding when and how to deploy the infrastructure for real-time analytics is crucial.

Another fundamental change in the way data warehouses and business intelligence in general will be used is that active data warehousing will distribute real-time analytics to people and processes far beyond the current audience of analysts and report-viewers. As the data warehouse takes its place at the table with the rest of the operational systems, the investment in data warehousing will be leveraged by all facets of the organization, its suppliers, customers, in short, all stakeholders. Real-time analytics will be integrated with operational processes in such a way that they will become transparent.

The technology to implement ADW, real-time analytics and closed-loop decision support systems is available now. Organizations will decide to deploy these technologies at different rates, but for those that see real-time analytics as a strategic move, expect to see rapid deployment. The time to plan for it, to become informed and to plot your strategy is now.

---

<sup>4</sup> <http://www.informationweek.com/story/showArticle.jhtml?articleID=12802974>