



Intelligent Solutions, Inc.

Solve your business puzzles with Intelligent Solutions

**Active Data Warehousing – the Ultimate Fulfillment of
the Operational Data Store**

Written by:

Dr. Claudia Imhoff
Intelligent Solutions, Inc.
P. O. Box 4587
Boulder, CO 80306
Voice: (303) 444-2411
www.IntelSols.com

Introduction

Over the years, data warehousing has gone through a number of evolutions – from a relatively simple reporting database to sophisticated analytical applications such as analyzing customer lifetime values, market basket analyses, potentially defecting customers, fraud patterns, inventory churns, and so on. In all, though, these static sets of data could not give us the most current and recent changes necessary to *ACT* upon the results of Business Intelligence analyses. For example, once we could identify a customer likely to go to a competitor, we still could not view their current situation – e.g., what products does the customer have with the company, is the customer a VIP requiring special treatment, where are they in the sales cycle?

The reason for this lack of insight was that the warehouse was set up to give us static snapshots of data, perhaps as recently as last week. But last week's (or even last night's) data is often not sufficient to react to current situations. Things change rapidly in today's e-business economy and the company with the best set of integrated, current data is the one that will not only survive but will actually thrive in this economy.

Unfortunately most enterprises today do not have any integrated data other than the snapshots found in their data warehouses. This is where the need for the Operational Data Store (ODS) was generated. And fortunately now, you can have integrated data in the static snapshots and in live, current records – an environment in which both types of data and requirements can co-exist. This concept is called the Active Data Warehouse.

To better understand this advance in technology, let's examine the characteristics that make the ODS so very different from the traditional data warehouse. To do this, you must understand the difference between analytical and operational applications (see Figure 1).

We classify the analytical applications as Business Intelligence, noting that they consist of the data warehouse supplying data to the various analytical applications in the data marts. The applications running against these components use decision support interfaces (DSI's) and give us great insight into our customers' demographics, buying habits, profitability, lifetime value, and so on. But insight into customer behavior is not enough. As we stated, you also need the ability to *ACT* upon these findings by having ready access, from anywhere in the enterprise, to integrated, comprehensive and current information about customers, products, inventories, orders, and so on, as well as quick access to some analytical results.

The operational applications are called Business Management components. These applications give the enterprise the ability to take action using its collective intelligence and subject knowledge. They also provide the organization with an enterprise-wide understanding of their situation, which facilitates a transition away from the silo business unit or functional viewpoint. For example, many of us have had the experience where we update our contact information (e.g., new address, phone and fax number) through a web site, only to find ourselves repeating that same information when we talk to a call center representative later that day. This is a prime example of an enterprise with severely isolated systems.

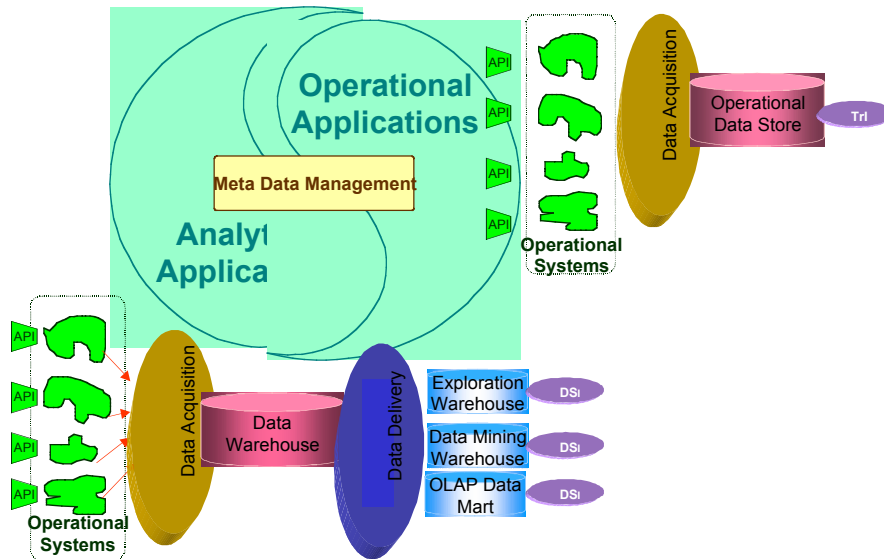


Figure 1 – Analytical and Operational Components

Business management is “where the action is”; it allows our knowledge of the customer to be applied to modify customer behavior and manage customer touches. Business management consists of the operational data store (ODS), the transaction interface (Tri) that provides users with access to the valuable information as well as the ability to update the ODS, and the associated meta data that provides business and technical people with information about the ODS.

The architecture used to support these two important sets of applications is called the Corporate Information Factory (CIF) (see Figure 2). The CIF is a logical or conceptual architecture that provides an integrated view of enterprise data enabling both business intelligence and business management capabilities. This architecture is a proven roadmap that maximizes the success of enterprise-wide CRM implementations and e-business strategies.

There have been numerous articles written on Business Intelligence and data warehousing in general. This paper focuses on a lesser known but increasingly important component of the CIF, the ODS, its role in business management, and the ability of the Teradata technology – Active Data Warehousing – to handle not only the historical, analytical requirements of the data warehouse but also this new need for actionable information found in the ODS.

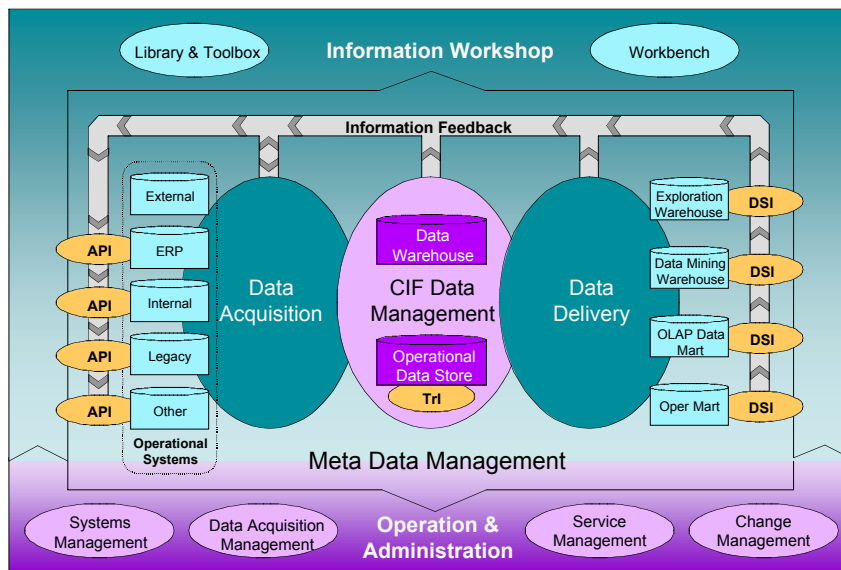


Figure 2 - The Corporate Information Factory

Characteristics of an Operational Data Store

The ODS is a subject-oriented, integrated, current-valued, and volatile collection of detailed data that provides a true enterprise view of data by subject area. Let's look at these defining characteristics of the traditional ODS in more detail:

Subject-oriented – The ODS is organized around major data subjects of interest to the enterprise. The primary purpose of the ODS is to collect, integrate, and distribute current information about the data subject, and to provide an enterprise view of it. The subjects can be any that are important to the organization. For example, a customer-focused ODS will typically house the most current information about a customer as well as information on all recent customer interactions with the organization – including product ownership and summary usage statistics, billing or statement information, summary level contacts, and other related information.

Integrated – The integration characteristic of an ODS is of key importance to e-commerce, CRM and other fast moving business initiatives. The ODS represents an integrated image of a particular profile, whether it is a customer's, a product's or an order's. Information making up this profile can be pulled from any system in the organization, both operational and decision support. While building and refreshing an ODS, the organization integrates all the different sources of operational information into a consistent view within the ODS, which is then used when reacting to a particular

situation or perhaps interacting with the customer across all contact points. As the definitive record and the consolidation point for profiles, the ODS may also provide other systems in the organization with this valuable information. The real value in the ODS is that anyone in the organization can access this integrated view from anywhere in the organization (or outside of it, in the case of your customers or partners), at any time.

Current-Valued – The ODS carries little or no history – much like a typical operational system. Unlike a data warehouse, which is a series of snapshots of information used for strategic analysis, the ODS is a current picture of the subjects in question, and is used for “action”. Note: “Currency” is relative and can be defined differently depending on the subject matter (see the next section). For example, an ODS used by a bank may define current data to include weekly account balance figures carried for one or two months. For example, a customer-focused ODS will usually have some recent historical information, e.g., prior address, closed accounts, recent campaign solicitations. An important point to remember with an ODS is that it should have far less history than the data warehouse, and it should not be considered as a replacement for the warehouse because it cannot facilitate the detailed, historical analysis performed by business intelligence systems.

Volatile – The data in the ODS will change frequently and these changes are typically reflected as updates to the existing fields in a record, not snapshots of whole records as in the warehouse. Changes to information in the operational systems will be reflected as changes in the ODS as well. Some types of information, such as account ownership, order status changes, customer touch records, product usage records, and contact information can change quite frequently. In many cases, the ODS can be updated directly by the users and customers, adding to its volatility. New records might be added directly into the ODS at the same time that their new product information is placed into the business operations systems. The customer ODS must be designed to handle these frequent updates and changes with ease and with appropriate referential integrity protocols.

Detailed – The customer ODS carries mostly low level, detailed data for all customer profile information but may have some summarized information about customer contacts and products or services. The summary data that exists in the ODS is different from the summarized data found in the warehouse. The summary data in the ODS is dynamic in nature rather than static as in the warehouse. That is, summaries in the ODS may be calculated at the time of request rather than being pre-calculated and stored in the database.

Refreshment Speed of the ODS

There is an additional characteristic of the ODS that is very relevant to the CRM and e-commerce world – the speed at which the ODS is refreshed. Your organization has some choices in terms of the currency of information and the frequency with which that information is updated. For example, a customer may log onto your web site and enter his new address, phone and fax number. The new customer contact information must be updated in the customer ODS almost immediately, within a few seconds, upon its entry into the operational environment. This type of ODS is labeled a **Class I**. The class I ODS is used when the information must be very accurate and up-to-date at all times, e.g., the customer service representative talking to a customer must see his most current information no matter where it was initially entered or changed.

A **Class II** ODS is a little more relaxed, using store and forward techniques for data update rather than performing synchronous updates. A class II ODS receives updates every half hour or hour. The customer's summary web purchases may be updated into the customer ODS every 30 minutes. Because the service representative only uses this information to get a feel for the customer's product interest, going instead to the product-centric ordering systems for his transaction details, class II updates are satisfactory for these summaries. Because the information is not as current as in a class I, the integration of the ODS with the other systems may not be as difficult as for a class I. There is also some trade-off between update frequency and integration. The faster the update time, then the less time there is to perform complicated or extensive integration routines.

A **Class III** ODS is typically updated in batches, most often on a daily basis. A customer's product preferences may be updated into the ODS only once every day or so. Information currency requirements are not nearly as robust when organizations build a class III ODS. Because the product preferences are used to understand cross-sell recommendations, and because they do not change that frequently, class III updates work well.

The fourth type of ODS, **Class IV**, is a special case where information provided to the ODS comes not only from the operational systems but also from the data warehouse or specific data marts. The information from the data warehouse or data mart is transferred into the ODS only periodically, usually in a scheduled fashion. This aspect of the Corporate Information Factory, the interaction between the data warehouse / data marts and the operational data store, is often overlooked or ignored.

Class IV interaction calls for small amounts of pre-aggregated or pre-analyzed data to flow from the strategic decision support environment into the ODS for use with more tactical applications. For example, the corporation may determine the lifetime value of its customers through an extensive analysis of customer data. The results of that analysis – the lifetime value score for each customer – is then updated in the customer's record within the ODS so that employees have ready access to this key strategic CRM data while performing operational tasks.

Once the strategic results are stored in the ODS, online real-time support of important strategic information is possible. In doing so, the data warehouse and data marts can be said to support online high performance access of data when that access is needed.

For the Active Data Warehouse architecture, the integration of the data warehouse and ODS is much simpler than if these two components were built in separate environments. Because there is a single instance of the overall database and architecture, movement of data between and among the various Corporate Information Factory components is much cleaner. When appropriate, the same set of reference data may be used by all components rather than be replicated or recreated over and over.

Figure 3 is a depiction of the conceptual architecture of the Active Data Warehouse infrastructure. Notice that the overlap in the middle is where the common dimensions, calculations, reference data, and so on, may reside, to be used by all components.

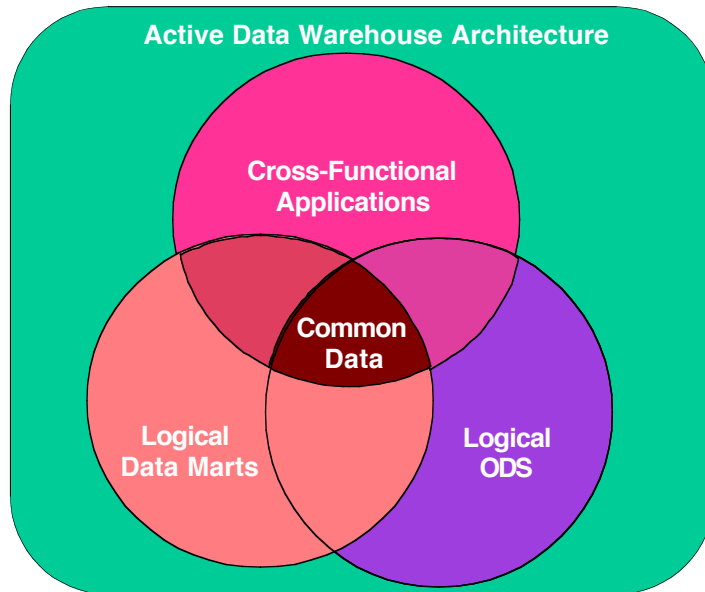


Figure 3 – Active Data Warehouse Architecture

Active Data Warehouses have advantages in terms of efficiency and productivity. The common data is created, maintained and updated only once, thus improving its quality and reliability.

Common ODS Applications

Let's look at a number of ways that the Operational Data Store can be used in different environments. One example is a CRM-oriented ODS. A customer-focused ODS supports functions in the customer touch zone, i.e., when the customer interacts with the enterprise in some fashion, by supplying the integrated, current customer data. Any employee who touches the customer (e.g., phone call, bill, collection notice, email, and so on) should have access to this valuable information. While the ODS concept is not new, many organizations are just beginning to realize its potential as a prime facilitator of their CRM strategy. Confusion still reigns in some organizations as to just what constitutes a true ODS, and as to which systems fall into the ODS classification (thus providing true enterprise level business management capabilities). The banking industry was one of the first to recognize the need for integrated customer data from the individual account systems and to provide access to a profile of the customer and all products and services owned or used by that customer. The first of these systems appeared in the late 1970's. The term ODS had not been coined when these first applications were built, but the concepts embodied in these efforts were clear predecessors to both CRM and the ODS. In keeping with the early start down the CRM path, Financial Services, in general, are still considered thought leaders in the CRM environment – both technically and strategically.

As another example, let's look at the insurance industry. Many insurers are building an ODS to consolidate their overall policies and manage their coverages. The ability to call up a complete and current customer's portfolio (including all the policies owned by the customer, the claims made against each policy, the statuses of the claims, and so on) is invaluable to the insurer – in determine the value of the customer, the discounts and other incentives to offer to retain the customer, and in tracking each individual agent's performance. As an added bonus, many insurers have discovered that they can use their ODS to determine their exposure in a specific geographical area when a disaster strikes. For example, if a hurricane approaches southern Florida, the insurer can use its ODS to determine how many policies it has covering property in that part of Florida, thus enabling the insurer to act more quickly to the impending claims.

A final example is in the e-commerce area. When you log onto a web site, have you ever wondered about the rationale for what banner ads appear? Companies with little or no insight into their web visitors simply slap up one of any number of banner ads, hoping that one of these will catch your eye and interest you. This is similar to the mass marketing or "spray and pray" techniques traditionally used by brick and mortar companies. Other companies seem to always have an appropriate ad or coupon that seems just right for you. What is the second group doing differently? How do they know what interests you?

The answer is that the second group of companies likely has a sophisticated environment that captures your click stream (i.e., every click you performed on your last visit to their web site), analyzes where you came from (the referring web site), the areas you visit in the company's web site, the items you showed an interest in, the ones you put in your shopping basket and then took back out, and the items that you did eventually purchase. All these transactions give the company insight into you – your interests, your temptations, your method of shopping, and so on. By analyzing this invaluable information in a variety of ways, the company can determine the appropriate banner ad to display when you next visit their web site. The trick is – they only have about a tenth of a second to recognize you, determine the appropriate banner ad, and then get it displayed on your screen. Certainly this cannot be done if this involves scouring and analyzing all your previous click stream data to determine the appropriate ad. That could take several minutes or longer and you will be long gone before the site even comes up.

Here is where the interplay between the ODS and the various analytical applications comes in. Figure 4 shows the traditional data flow in this scenario. Several types of analytical applications (data mining, exploration and traditional OLAP types of marts) are used to perform the hard-core analytical functions on the massive click stream and customer data. These applications work in concert, the results of one perhaps feeding an analysis in another to determine your preferences, buying history, responses to different campaigns (e.g., banner ads, coupons, or other types of incentives)

For each visitor to the web site, the analyses performed in Figure 4 determine the banner ad ids appropriate for each visitor. Once these ids are identified for each visitor, they are stored in the ODS. When you return to the web site, your identification is sent to the ODS, a process then looks up the banner ad id for you specifically, sends the id to the content server and – presto – the ad appears on your screen in sub second time!

Every time you log on, the process picks the next id in the listing and another appropriate ad appears.

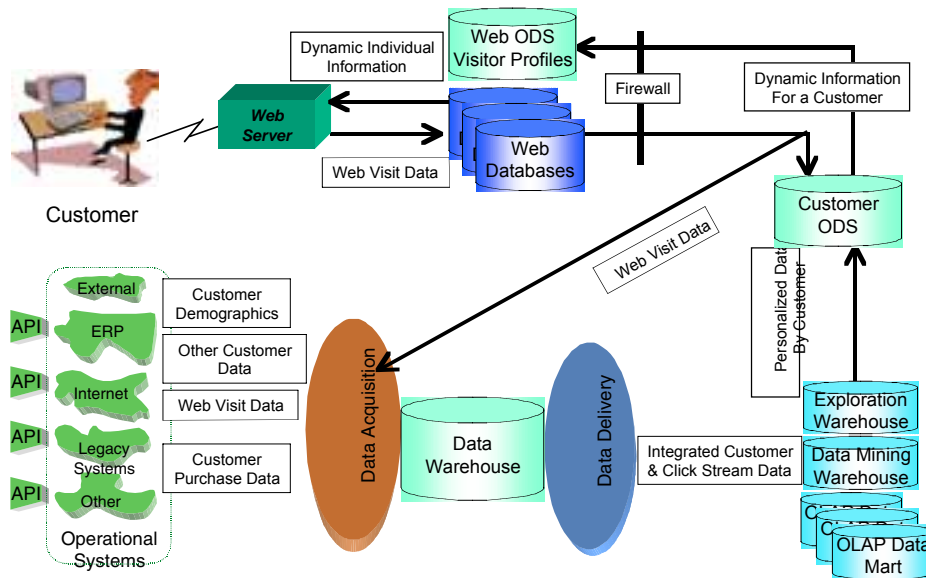


Figure 4 – The Interplay of CIF Components

Periodically, say once a month, your click stream and other data are analyzed again and a new set of banner ad ids is loaded into the ODS for your next visits. Now imagine how easy this would be if you had all the components in a single technology environment. Again, you begin to see the importance of the Active Data Warehouse architecture. A Teradata customer that provides web profiling and analytics, is reaping the benefits from this scenario. Their Active Data Warehouse and relationship optimization application help their customers gain deeper insight into consumer profiles and behaviors in addition to providing them with intelligence regarding brands, products, and services. The payoff for this company was increased campaign effectiveness by leading customers to a desired conclusion and increased revenues from new service creation generated from their enhanced analytics.

Building the Ideal ODS

The ODS is simply a database that houses all data associated with a particular subject – like customer, order, policy, claim, etc. The organizations that realize the most value from this component of the CIF are those that utilize this information in unique and innovative ways. These organizations build comprehensive ODSs with valuable information. However, in building these components, they also take several additional steps:

- They integrate this database with many other Corporate Information Factory structures such as specific data marts.
- They build intuitive and user-friendly access mechanisms to let business users view and modify the information.
- They combine these ODSs with toolkits, which allow the users to make the most of the information found in them.
- They pay close and on-going attention to the quality of the information contained in the ODS itself.

Any discussion of the ODS begins with the depth and breadth of information found within the database, but continues with a focus on some of the additional capabilities that must be combined with the ODS to facilitate a true enterprise view. As indicated, the ODS provides the comprehensive source for integrated, current information needed throughout the enterprise. As such, it is imperative that this database have the ability to store and maintain a broad range of information and interact with other systems in the organization to provide a good deal of functionality. An ODS should not be a “black hole of information”; instead it should be the “nerve center” of an organization’s activities. Thus far, we have described the general characteristics of an ODS, distinguished this database from others in the Corporate Information Factory, and provided an overview of several examples of ODSs.

In the next section of our paper, we discuss the interface needed for the ODS and then examine the Active Data Warehouse describing how it satisfies the ODS portion of the CIF.

The ODS Transaction Interface

The transaction interface is the primary method for users to access the data stored in the ODS. Unlike the data warehouse and data mart access tools, the primary purpose of the transaction interface is not to issue a query and bring back all the records fitting the characteristics specified therein. Instead, it is designed to pull one record at a time, e.g., one customer profile and all associated information, and display this information in an easily readable format. The searching mechanism described below is an exception to this general characteristic; good searches bring back any record that is a possible match to the search criteria. The user then examines the returned records and chooses the appropriate one. Additionally, there are some other situations where the users of the ODS have business questions that require the transaction interface to return a list of items matching some criteria. One insurance company uses its ODS to return lists of insured properties to use in determining its risk exposure in coastal areas during hurricane season.

Also, unlike the cleansed data stored in the data warehouse, the users may directly update the ODS. Because the ODS provides an extensive profile, it is sometimes the best (or only) place an organization has to store new profile data elements, such as income, investment preferences, etc. As we know, it is quite common for old product applications to keep a minimal amount of customer information – only

that needed to open the account or deliver the product to the customer. In cases like these, the product applications are hard to modify, and the ODS becomes the only place to store the extensive information that the organization wants to collect and use in its enterprise-wide strategies. This usually means that the organization must build the transaction interface to allow users to directly update the ODS, rather than building the ODS as read-only. Direct update capability means that the transaction interface must facilitate the addition, change, and (possibly) deletion of information as well as its retrieval. Detailed process models and their associated Create, Retrieve, Update, and Delete matrices, which specify the business rules applied to the 'CRUD' activities for each element of information, must be included with the development of the transaction interface.

The transaction interface design will be heavily dependent on the business processes that drive the use of the ODS. Customer service may require a different interaction with the ODS than sales, marketing or finance. The transaction interface may be required to provide different functions and views of data to different parts of the organization.

Also inherent in the transaction interface is its ability to integrate with other applications. As indicated above, other systems get and receive information from the ODS. Middleware is an essential technology component that works in conjunction with the transaction interface to facilitate this transfer of information between applications. Middleware or message brokering software connects different systems and allows them to share data. It acts as the go-between for diverse platforms and systems, providing a common link to many systems. In an environment where multiple systems have to exchange data, middleware can facilitate the process. Rather than writing an application interface from the ODS to each legacy system, an application interface from each legacy system to the ODS, and an application interface from each legacy system to the others, middleware requires only one application interface per system - to and from the middleware.

Characteristics of an Active Data Warehouse Environment

What is Active Data Warehousing? It is a technology that provides an integrated, consistent repository of data to drive both strategic and tactical or operational decision support within an organization. Given that, what would be the ideal characteristics of this technology supporting not only the ODS for tactical decision-making but also the data warehouse and data mart environment supporting the entire strategic decision-making capabilities? Let's start with some obvious ones:

- The workload in such an environment would consist of traditional complex decision support queries but must expand to support the short, quick queries used in the tactical decision-making scenario. Therefore an Active Data Warehouse must accommodate a new set of business requirements. New requirements means new service levels in terms of performance, scalability and availability of the environment.
 - For example, the decision support queries submitted by a marketer may be used to derive patterns in customer buying

habits, models of customer demographics, determine customer profitability, and so on. Meanwhile, tactical queries may be needed to determine the best offer or banner ad for a customer, determine the availability of a product, or alter a campaign based on current results. To accommodate this, it may mean that restraints may have to be put onto the longer running analytical queries to guarantee the tactical query's performance.

- Active Data Warehousing means large volumes of data and, therefore, scalability becomes critical and absolutely required to provide the large amounts of detailed data needed to understand business events. Scalability also means the ability to support the concurrent queries described above. Therefore, the technology should provide the means to referee the contending actions and balance the conflicting needs of these various types of users.
- The availability (and thus the reliability) is perhaps the most distinguishing characteristic of the technology to support both tactical and strategic queries. Traditional data warehouses do not usually have to be functional 24X7X365. Not so for the ODS functions! As a result, the Active Data Warehouse must never go down, must never be inaccessible, or the business simply cannot operate.
- Because the requirement for data freshness is far more stringent in the Active Data Warehouse than in traditional data warehouses, there is a need for a much more sophisticated data acquisition mechanism.
 - Data Acquisition must be performed much closer to the time a business event took place. Ideally, the acquisition mechanism will provide a continuous feed of new or changed data into the environment without blocking access to the very tables being updated.
 - As we mentioned in the description of the different classes for the ODS, the timing of data acquisition will vary by application. However, a data freshness characterized by being only a few minutes old would be the norm. Therefore, an Active Warehouse must be able to handle large volumes of changing data with ease.
- As the need for decision-making expands from strategic to tactical, it makes sense that the environment would evolve even more to event-based activities. This opens up a need for database triggers based on a chain of action and reaction. Triggers can be quite useful because they can automatically initiate certain actions when specific conditions are reached.

Given these characteristics, what benefits should you expect to gain? Besides the obvious benefits of better performance, availability, and scalability of this technology, there are a few less obvious ones:

- The elimination of both latency of action and data redundancy is a great advantage.
 - Latency of action is defined in terms of the time to study the results of a particular strategic query to the time it takes to act upon those results. With a single environment in which both the tactical and strategic data are co-located, this latency is almost zero. The integration of Business Intelligence with the related Business Management function is the ultimate goal
 - Because there is a single environment, there is no need to replicate or duplicate data in physically distinct and separate environments for strategic and tactical decision-making.
- Active Data Warehousing yields a seamless infrastructure. The technology logically incorporates a fully functioning ODS as well as the traditional data warehouse and marts in a single physical platform . This means that the components are easier to develop, maintain, sustain, and enhance. It also means that the environment is far more flexible in terms of its usage, changes to the underlying database, and additions to the existing data.

Active Data Warehousing and ODS – Summary

The ODS is a key component of your technology environment. The ODS provides business management capabilities to the organization. It is the vehicle by which analyses performed in the data warehouse and data marts are made “actionable” and distributed for use by the entire enterprise

For example, for a CRM initiative, a customer-focused ODS supports all activity in the customer touch zone by supplying integrated customer data throughout the enterprise. Another example is a consolidated set of policy information for an insurer. For an e-Commerce company, the ODS can coordinate all web interactions including personalization, orders and other Internet transactions.

Architecturally, the ODS works in conjunction with the data warehouse and data marts by providing data into and receiving analytical results from these components. This is the critical process that makes the Business Intelligence of your organization actionable.

The Active Data Warehouse simplifies the overall construction and maintenance of the Corporate Information Factory by creating physical or logical components in a single instance of the database thus giving you:

- Maximum flexibility with a minimum of effort. Because many of the components are logical constructs, the reuse of the data along with the ability to quickly create new applications is a significant advantage.
- An efficient environment to maintain and enhance. Because there is one physical environment, it is a simplified process to enhance or change the existing applications and CIF components.

- An elimination of data latency and redundancy. Because many of the components are logical in nature, the time it takes traditional environments to extract data from the warehouse, format it for various data mart usage and then deliver the data to the data mart locations is eliminated. In addition, access to the current (ODS) as well as historical (data warehouse and marts) data is easily carried out with minimal delay in performance.

This technology not only creates a responsive Business Intelligence environment through the integration of the data warehouse and associated marts, but now also supports the critical characteristics of the actionable piece of the CIF architecture – the Operational Data Store used for Business Management. This is a strategically significant technological breakthrough and one that should be seriously considered for any enterprise embracing the Corporate Information Factory architecture.

About Intelligent Solutions, Inc.

Intelligent Solutions was founded in 1992 by Dr. Claudia Imhoff, and is headquartered in Boulder, Colorado. ISI is a leading provider of business intelligence and customer relationship management consulting, education, and seminars to business and government organizations throughout the world. Using its Corporate Information Factory architecture as a guide, the company specializes in developing business intelligence systems using data warehouses, operational data stores and data marts. More information on Intelligent Solutions, Inc., its services and its educational offerings is available on the Web at www.intelsols.com, or by calling (303) 444 – 2411.

For more information about Teradata:

Teradata, a division of NCR, offers technology solutions that power award-winning data warehouses built by the world's most successful companies. Teradata active warehousing brings analytic capabilities into operational environments, extending a company's capabilities. To learn more about these implementations from Teradata customers, visit www.TeradataMagazine.com where you can read customer articles and subscribe to Teradata Magazine. Additional case studies are available at www.teradatalibrary.com (select "Success Stories"), and white papers and other information are available at www.teradata.com.

Teradata is a registered trademark of NCR Corporation.