

Business Intelligence Network™ Research Report

Using Operational Business Intelligence for Intra-Day Analysis and Decision Making



Business Intelligence Network • BeyeNETWORK.com

1790 30th Street • Suite 310 • Boulder CO 80301 • (303) 339-7255 • (303) 339-7259 fax

TABLE OF CONTENTS

Executive Summary	1
Introduction	2
Evolution of Operational Business Intelligence	2
Types of Business Intelligence	2
The Three Components of Latency	4
Operational Business Intelligence Applications	6
Operational Business Intelligence Solutions	7
Data Integration and the BI Platform	7
Encapsulation of the Business Model	10
Reporting and Analysis	11
Decision Automation	13
BI as a Service	13
Key Steps in Implementing an Operational BI Solution	13
Case Study: Property/Casualty Insurance Company	15
Organization Background	15
The Business Problem	15
The Operational Business Intelligence Solution	16
Implementation Steps and Advice	19
Summary of Benefits and ROI	20
Summary of Critical Success Factors	21
Future Issues	21
Summary	22
Conclusions	23
Syncsort Business Intelligence Solutions	24
Company Background	24
The SyncSort Foundation	25
DMExpress	25

EXECUTIVE SUMMARY

Integrating BI within daily business operations

Business intelligence (BI) is all about delivering the right information in the right format to the right people at the right time so they can make effective and timely business decisions. BI applications have evolved to focus on *operational BI* – the integration of BI functionality within daily business operations. Operational BI is the use of BI to manage and optimize the day-to-day business with the ability to modify the business intra-day. The goal is a BI solution that reduces latency at one or more levels of the decision cycle – data collection, analysis, and action – and results in better, faster decisions based on up-to-date, accurate, detailed operational data.

Components of an operational BI solution

Components of an operational BI solution include data integration, encapsulation of the business model in tailored business views of the data, easy-to-use yet sophisticated reporting and analysis tools, decision automation technologies, and the ability to implement BI functionality as reusable services.

Key steps to successful solutions

Key steps in implementing operational BI solutions include: Getting top management support, ensuring business ownership of and commitment to the BI solution, ensuring a balance between short- and long-term BI goals, starting off with the right resources, ensuring data quality and integrity, making sure you have the right data structure and metadata, delivering a robust system, and providing comprehensive user support.

In this report, we present an in-depth case study describing the operational BI solution implemented by a property/casualty insurance company to improve the effectiveness of business decisions. This Syncsort customer uses Syncsort DMExpress as the extract, transform, and load (ETL) software for updating the enterprise data warehouse (EDW). DMExpress provides high-speed integration and transformation of operational data into EDW structures.

Better and faster business decisions are a real benefit

The major benefit of the property/casualty insurance company's operational BI solution is better and faster business decisions. Additional benefits are improved customer relations, faster access to high-quality data, and flexibility to adapt to new business requirements.

Operational BI offers a solid foundation for the future

The property/casualty insurance company's operational BI solution provides a solid platform for future enhancements to extend the value beyond the current implementation. Examples include a more flexible EDW infrastructure, retention policies for the EDW, and integration of BI reporting with the enterprise portal.

Operational BI adds business value and enhances agility

Important operational BI trends include creating a centralized BI environment to provide a single view of the business; flexible schema design; on-demand, ad hoc, self-service reporting; integration of reporting and analysis with sophisticated data visualization capabilities; integration of BI with the enterprise portal; increasing focus on advanced, predictive analytics; and encapsulation of BI functionality as a reusable service. The study results highlight the significance of operational BI as a mission-critical business process that adds value and enhances an organization's overall agility.

INTRODUCTION

BI delivers the right information at the right time...

Business intelligence (BI) focuses on delivering the right information in the right format to the right people at the right time so they can make effective and timely business decisions. To do this successfully, you have to know what information you need and when you need it, and then ensure that the business can deliver it.

...to manage and optimize daily operations...

BI services, applications, and technologies have come a long way over the past two decades. Current implementation efforts are targeted at *operational business intelligence* – the integration of BI functionality within daily business operations. The goal is to enable the organization to take advantage of BI to manage and optimize the day-to-day business. Operational BI reduces latency in the decision cycle and extends access to BI capability across the enterprise. Timely access to information about customers, suppliers, products, selling patterns, inventory levels and manufacturing schedules, and other operational business data results in better and faster decisions at all levels of the organization.

...and create a smarter, more agile organization

BI has become a mainstream, mission-critical business process that enables organizations to be smarter about the way they do business. BI also makes organizations more agile and better able to quickly adapt to changing business requirements. Working smarter coupled with increased agility, flexibility, and adaptability are keys to success in staying competitive in today's dynamic, global marketplace.

This report describes the evolution of operational BI, the components of an operational BI solution, and the experiences of four companies that have implemented real-world, operational BI solutions.

EVOLUTION OF OPERATIONAL BUSINESS INTELLIGENCE

TYPES OF BUSINESS INTELLIGENCE

Businesses want more real-time BI

Over the years, business intelligence has evolved from the traditional data warehouse (DW) or data mart, updated overnight or less frequently, to BI solutions that support more real-time data collection, analysis, and decision making. The goal is intra-day analysis of up-to-date information with the ability to make immediate decisions about the business. Today, there are three types of BI applications – strategic, tactical, and operational (see Figure 1). Each supports different business decisions, users, analytical time frames, and data.

Strategic use of BI focuses on achieving long-term, strategic business goals such as increasing revenues, decreasing costs, driving market share, improving customer service, and improving profitability. Executives and business/financial analysts use BI to compare the company's performance in key areas to its long-term goals. This type of analysis can require months or even years of data given its high-level business focus.

Tactical use of BI addresses shorter term, tactical efforts designed to help achieve a strategic objective. Examples of tactical management initiatives are a new marketing campaign for a specific product, introducing a new product, re-pricing an existing product, and changing buying patterns for raw materials. Tactical BI analysis to assess the results of such efforts is valuable to operational, line-of-business (LOB) managers in addition to executives and analysts. Depending on the scope of the initiative, analysis of tactical efforts may use data that is days to a few months in age.

Types of Business Intelligence

	Strategic BI	Tactical BI	Operational BI
Business focus	Achieve long-term business goals	Manage tactical initiatives to achieve strategic goals	Manage and optimize daily business operations
Primary users	Executives & business analysts	Executives, analysts & LOB managers	Analysts, LOB managers, operational users & operational processes
Timeframe	Months to years	Days to weeks to months	Intra-day
Data	Historical data	Historical data	Real-time, low latency & historical data

Sources: Claudia Imhoff, *Intelligent Solutions* and Colin White, *BI Research*

Figure 1

Operational BI integrates BI within business operations

Organizations are now implementing *operational* BI solutions to address a wider range of business decisions by integrating BI capabilities within business operations. Unlike strategic or tactical BI, operational BI focuses on the daily business cycle by incorporating BI in the process of running the day-to-day business.

Operational BI focuses on intra-day

The analytical and action time frame here is intra-day. Examples are monitoring a current campaign or checking order status during the business day with the ability to take immediate action based on analysis of the information. Doing this effectively requires access to detailed, up-to-date operational data (what is happening *now*) in addition to historical data (for trend analysis and context). "Up-to-date" data can mean either real-time data (the actual current value in the operational system) or low-latency data (data that has been updated intra-day). The ultimate in operational BI is real-time BI, where data is analyzed in real-time as it flows into the organization. One example is real-time tracking of inventory to ensure that an online marketing

campaign won't fail because the company cannot deliver the product. Another is real-time tracking of parts availability to support a new manufacturing schedule.

Operational BI brings BI to “the masses”

Operational BI also extends the value and benefits of BI to a broad new set of users across the enterprise. People and processes in operational departments – on the front lines, in the field, in back offices – can now take advantage of BI to improve daily decision making. In addition, operational BI offers the potential to extend access to BI data and analysis outside the organization – to customers, suppliers, partners, etc.

THE THREE COMPONENTS OF LATENCY

Latency occurs between data collection, analysis, and decision points

An important context for operational BI is *latency* in the decision-making process. On the continuum between when a business event occurs and when an informed action is taken, there are three segments of latency: data latency, analysis latency, and decision/action latency (see Figures 2 and 3).¹

- *Data latency* is the time it takes to collect raw data, prepare it for analysis, and store it where it can be accessed and analyzed.
- *Analysis latency* is the time it takes to access and analyze the data, turn the data into information, apply business/exception rules, and generate alerts if appropriate. Analysis may be done by a user or an application.
- *Decision latency* is the time it takes to review the analysis/alert, decide what action is required based on knowledge of the business, and take action.

Optimize the trade-off between time to action and business value

Operational BI applications seek to *optimize* the decision cycle, taking into consideration the trade-off between time to action and the business value of the action (that is, making a decision today may have greater or lesser value than making the decision next week).¹ Optimizing the decision cycle usually means shortening it. On the other hand, it may not mean minimizing the time lag, or automatically assuming that every decision must be executed in real-time. The key is to define the optimal timeframe, or “right time” for each decision cycle, one that reflects business realities and the trade-offs between risk and cost. It is extraordinarily expensive to create a completely real-time organization. Even if an organization can afford real-time decisions, it may not be necessary or worth the cost.

Operational BI is business- and process-specific

Successful implementation of operational BI requires analysis of each business process to identify opportunities to shorten the decision cycle. Then the organization can assess which BI technologies and products can be applied as part of an operational solution. Manufacturing companies with just-in-time inventories may need data updates at intervals during the day. Financial services companies may need to report exceptions as soon as possible to meet compliance regulations.

Three Components of Latency

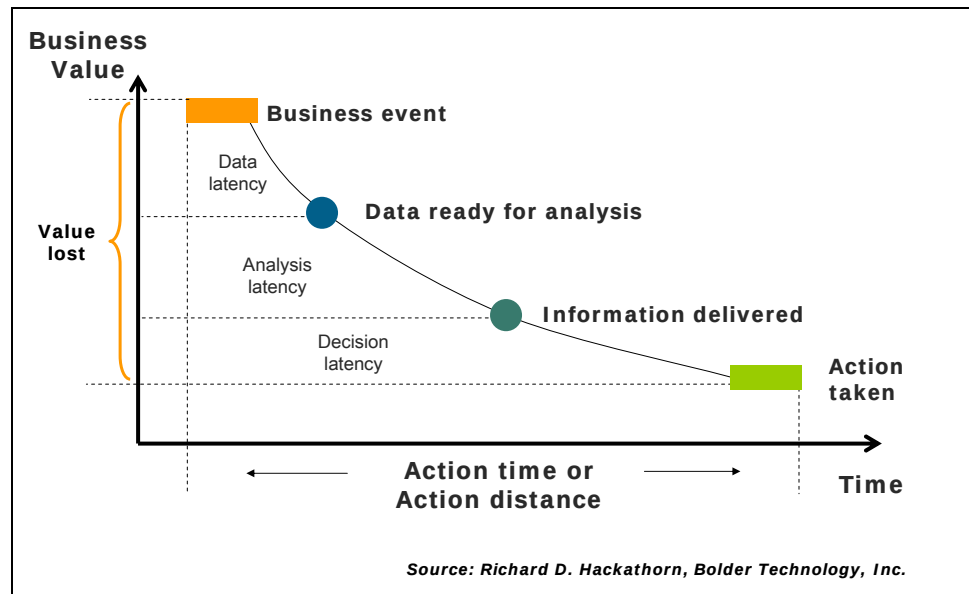


Figure 2. This shows the three components of latency between when a business event occurs and when action is taken. The elapsed time may result in lost business value to the organization.

The Benefit of Reducing Latency

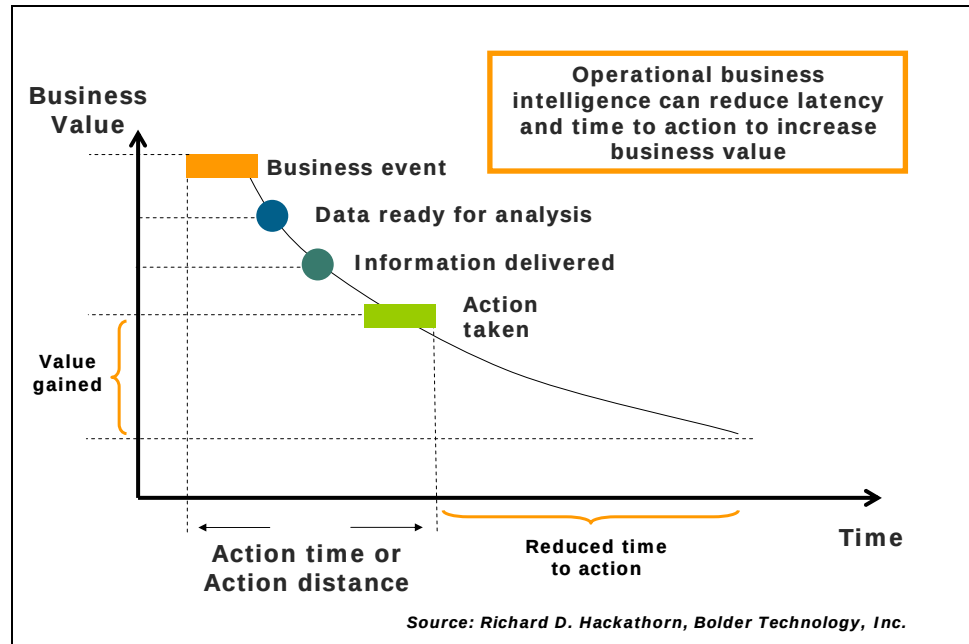


Figure 3. Reducing latency at one or more points in the decision cycle can dramatically increase the business value of the decision.

OPERATIONAL BUSINESS INTELLIGENCE APPLICATIONS

Types of operational BI applications

Operational BI applications fall into several categories²:

- *Operational reporting* – Reporting on business operations is a basic component of any BI solution, providing a structured way to access and view business data, regardless of its currency. Operational BI adds on-demand, ad hoc, self-service reporting with the ability to drill down among other reporting capabilities.
- *Operational analysis and scorecards* – Analysis is another basic BI application, enabling the user to see data in context, analyze it, and create performance metrics and key performance indicators (KPIs). Operational analysis ranges from simple (the ability to drill down from summary data on a report to see selected details) to very complex (data mining, risk assessment, fraud detection, and predictive analysis). Scorecards rate business performance against goals, and can be based on a performance management (PM) methodology.
- *Operational dashboards* – Dashboards visually present business information, including performance metrics, in an easy-to-use format. For example, a dashboard can display KPIs and may have the ability to examine underlying causative factors (e.g., why the organization is on track – or going off track – in achieving its goals). Dashboards can also include an assessment, based on business rules, of potential financial impact, alerts, and options for further analysis.
- *Automated agents* – These proactive components of an operational BI application assist in the analysis and action segments of the decision cycle. Automated agents include alerts (to tell the user an exception or problem has occurred), recommendations (to suggest further steps in analyzing the situation and/or making a decision), and actions (the agent itself takes action).
- *Guided analysis workflows* – Often based on best business practices, guided analysis workflows are designed to help users analyze data and speed the decision process (e.g., helping a user make a decision on a loan application).

Operational BI applications are on-demand, event-driven, or embedded

From an implementation perspective, operational BI applications can be demand-driven, event-driven, or embedded depending on business requirements.

A *demand-driven* operational BI application is accessible to operational users and processes when BI analytical data is needed to make a business decision. A simple example is running a BI report on demand.

An *event-driven* application monitors specific business operations, relates ongoing data events or changes to business rules, and may generate alerts when appropriate. The event-driven BI application may also have the ability to take action based on knowledge embedded in the system.

An *embedded* BI application is one that is an integral component of a business process. An important technological trend here is the use of a service-oriented architecture (SOA) to package BI capabilities as services. BI services can then be called directly by operational processes.

Operational BI adds value to the business

The closer an organization gets to real-time BI, the more event-driven it becomes and the more it will make use of embedded operational BI applications.

The strategic direction is to seamlessly integrate BI functionality into business processes – to narrow the gap between analytical and operational applications by creating a “closed-loop” process. BI analytics become a service available when needed or embedded within an operational process. This recognizes the importance of BI as a mission-critical “operational” system that adds value to the business.

In summary, operational BI is a set of services, applications and technologies for monitoring, reporting on, analyzing, and managing the business performance of an organization’s daily operations.

OPERATIONAL BUSINESS INTELLIGENCE SOLUTIONS

Optimizing the decision cycle with operational BI

Organizations can use operational BI in many ways to optimize the decision cycle. In this section, we discuss the major components of an operational BI solution, key implementation techniques and technologies, and current issues and trends.

DATA INTEGRATION AND THE BI PLATFORM

BI integration techniques are a subset of enterprise data integration

A fundamental requirement for all forms of BI is access to integrated business data from multiple data sources across the enterprise – data about sales, orders, customers, suppliers, partners, products, inventory, etc. Integrating data to support BI applications is a key subset of an enterprise data integration platform (see Figure 4). The data integration component addresses data latency by making disparate data more readily accessible to BI applications and users.

INTEGRATION TECHNIQUES. There are three techniques for integrating data:

Data consolidation physically integrates data

- *Data consolidation* – This *physical* integration approach extracts data from multiple sources and integrates it into a single persistent data store or BI repository. The consolidation process often cleanses and reformats the data during the integration process. The physical repository can be an operational data store (ODS), data warehouse (DW), data mart, or other data repository. The latency of data in the repository depends on how often the data consolidation process is executed.

Data federation virtually integrates data

- *Data federation* – This *virtual* integration approach leaves data in place and provides direct access to live data across multiple sources through a virtual view of the data. The latency of data varies depending on the source.

Data propagation copies data between systems

- *Data propagation* – This approach copies data from one or more source systems to one or more target systems. Data propagation is often used to copy events or data changes between systems. The latter requires changed data capture (CDC) technology that can capture the events or changed data on the source, such as data replication or database triggers. An example is capturing changes to an order management system and copying them to the ODS or DW for BI analysis of open order status, order fill rates, etc. Propagation can be either synchronous (copies

are made within the operational transaction boundaries) or asynchronous. Asynchronous propagation can be scheduled or event-based (near-real-time). Latency of data in the target depends on how often source data is propagated.

Enterprise Data Integration Platform

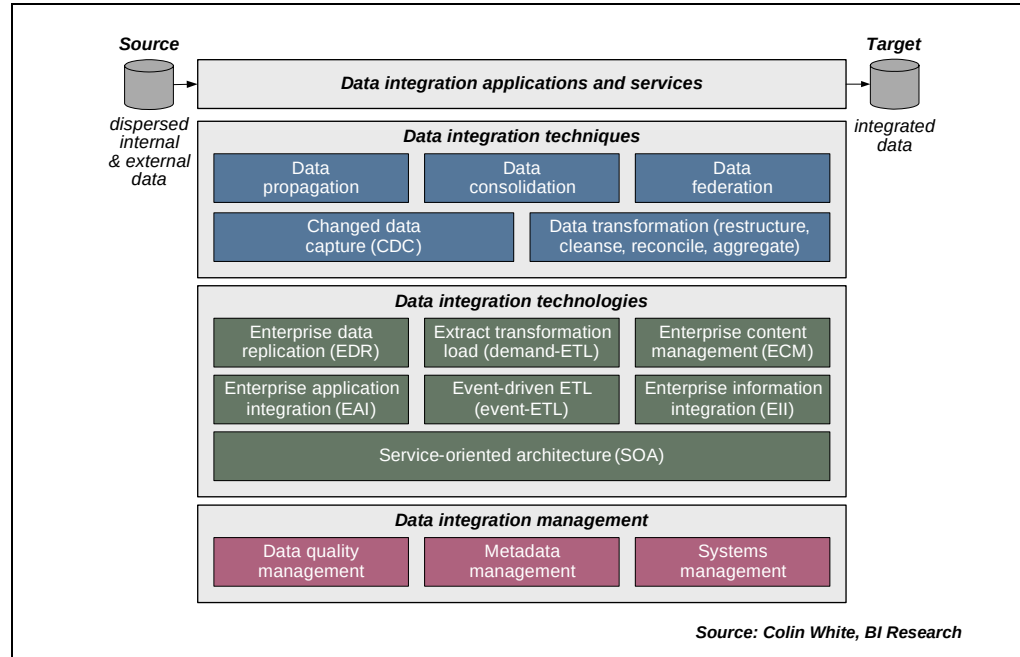


Figure 4. A successful operational BI platform is based on a comprehensive enterprise data integration platform.

BI data integration technologies: ETL, EII, EAI

INTEGRATION TECHNOLOGIES. These techniques are supported by the following data integration technologies:

- *ETL (extract, transform, and load)* – ETL tools extract selected data from operational systems, transform the data, and load the data into the target database. This method of data integration can be either on-demand (batch) or event-driven (online), and supports either data consolidation or data propagation. ETL provides historical and low-latency BI data.
- *EII (enterprise information integration)* – EII tools use data federation to integrate data. The EII tool creates its own metadata layer through which users view and access the data. EII provides real-time access to all types of data – operational, analytic, and even external data.
- *EAI (enterprise application integration)* – EAI tools enable applications to propagate and exchange events, messages, and data. EAI has evolved to support SOA and an enterprise service bus (ESB). Integration between EAI and ETL can give ETL access to EAI capabilities, and integration between EAI and operational BI applications has the potential to help close the loop between operational systems and BI applications. EAI provides real-time and low-latency data.

Select a BI platform to meet operational business needs

THE BI PLATFORM. A successful operational BI platform is based on a comprehensive enterprise data integration platform that takes into consideration operational BI integration needs. In addition to integrated data in the form of a physical data store or virtual view of the data, the BI platform also requires a delivery component to connect the integrated data with BI users and processes. All of these components must be reliable, high-performance, and scalable to support operational BI.

Developing an operational BI platform involves understanding the underlying business processes and systems, clearly identifying data integration requirements and limitations, and deciding which integration implementation tools are appropriate. Sample steps here are:

Evaluate existing resources

- Assess the organization's existing BI capabilities, technical skills, and technologies against what is required to implement a successful operational BI solution. This will inform the organization about its readiness to undertake operational BI and how fast it can implement a solution.

Discover opportunities

- Analyze business processes to identify opportunities to shorten the decision cycle.

Assess latency requirements

- Identify sources of data and the desired level of latency for each type of data. Frequency of updating the BI repository (if real-time access isn't required) depends on business value, cost, volume of data versus time available for updating, and other criteria. Data latency can also be subject to source system limitations; the source may not support changed data capture or may limit the frequency with which data can be extracted.

Understand integration and data quality needs

- Identify what functions need to be part of the data integration process: data profiling, validation, reconciliation, extraction, cleansing, transformation, aggregation, delivery, loading, etc. Another aspect is the complexity of each function. Cleansing and transformation requirements, for example, can be quite sophisticated depending on source data quality. A key here is to identify data quality issues and how to resolve them by cleaning operational systems, creating an enterprise data quality program, etc. As with data latency, data quality requirements may vary depending on business need, and there are similar cost/benefit tradeoffs.

Design the integration solution

- Select the appropriate integration techniques and technologies. Each has strengths and challenges depending on the specific operational and BI environment. (For more information on data integration techniques and technologies, and which to use when, please see The Data Warehousing Institute [TDWI] research report titled "Data Integration: Using ETL, EAI, and EII Tools to Create an Integrated Enterprise" by Colin White of BI Research.³) The solution must obviously support the BI repository in the case of data consolidation and propagation. Repository options include an RDBMS (relational database management system), a multidimensional OLAP (online analytic processing) server, a specialized analytic server, etc.

Reducing data latency can be challenging

As organizations drive to reduce the latency of BI data, they are moving data from operational systems into the DW more often than once a day or on a continuous

basis. However, data integration can be easier said than done as the frequency of updating the BI data repository increases. Traditional BI environments have relied primarily on overnight batch ETL to update a DW. As companies move toward operational BI, the concept of the overnight batch window for updating the DW may be replaced by the need to trickle-feed data changes into the BI system during the day. The more often BI data is updated, the less time there is to cleanse and transform data. This pushes the envelope for high-speed data integration solutions. It also means the data repository must be able to effectively handle a mixed workload of both ad hoc queries/reports and frequent data updates.

Many companies choose to reduce the latency of data in stages as they gain experience with the significant issues involved in delivering low-latency or real-time data to BI applications. An example here is a company that initially loaded its EDW in batch mode. Most data was loaded on a daily basis with some coming in weekly or monthly depending on the source system. As the company gained experience with the data, it began to move to more real-time updates where appropriate, depending on business need and availability of source data. The evolution from batch to more frequent updates provided an important learning process.

Ensure detailed data is available when needed for analysis

Another direction in this area is increasing the granularity of data in the DW to allow users to drill down to lower levels of detail for analysis. Summarized data is valuable but not sufficient to support many operational BI activities. The key here is to decide how granular data needs to be in each case. As with latency and data quality, the granularity required will vary depending on business need and cost/benefit tradeoffs.

ENCAPSULATION OF THE BUSINESS MODEL

Creating a business view of the operational data

Another requirement for an operational BI solution is a layer of abstraction over the operational data that represents and encapsulates the business model of the data – a “business view” of the data. This layer creates the business taxonomy for the underlying data that users can understand and use when requesting reports and analyzing data. Elements of the taxonomy can include standard definitions of business terms and information about where the data comes from. Many organizations create multiple business views tailored to the needs of specific BI user groups or functions – sales, marketing, finance, distribution, customer service, etc. These views are created in the data repository, the EII tool, or the BI reporting and analysis tool depending on the specific operational BI implementation.

Central BI repository with virtual data marts

In traditional BI environments, business views are often implemented as separate, physical data marts. The benefit is better performance on a smaller set of tailored data. However, data marts are typically updated from the DW. Every time you move data, you increase latency.

The trend we see here is toward an operational BI architecture with a single, central BI repository or DW – a “single version of the truth” – with virtual business views (virtual data marts) created within the DW itself. Organizations find this approach reduces the cost of technology, storage, and maintenance and administration, and results in higher quality data.

Schema flexibility

Another trend is to build as much flexibility as possible into the design of the DW schema to accommodate unexpected, unpredictable business requirements. The goal is increased agility as companies swallow acquisitions, create new products with unanticipated features, and deal with changing government and compliance regulations without having to dramatically change the DW structure.

Taking advantage of existing master data

Many organizations are implementing master data management (MDM) applications to formalize and centralize the definition and maintenance of reference data about key business entities – people (customers, employees, suppliers, partners), things (products, assets), and places (regions, countries, territories). Integrating MDM with BI enables BI applications to take advantage of existing corporate master data. Thus, MDM applications can be the source of master data and the business model around the master data for operational BI applications in the same way that business transaction systems are the source of detailed business data. (For more information on MDM, please refer to “Master Data Management: Creating a Single View of the Business,” a research report by Claudia Imhoff and Colin White published by the Business Intelligence Network.)⁴

REPORTING AND ANALYSIS

Tailor reporting and analysis to user needs and skills

The reporting and analysis component of the operational BI solution delivers operational BI data to users. Therefore, it is important to understand these users, their skills, the types and complexity of BI applications they need, and how and where they work. The organization can then tailor the solution to the user.

Because operational BI focuses on managing the daily business, it must be readily available and easy to use by a wide variety of users throughout the organization – the retail buyer in the store, the call center operator, or the salesperson on the road in addition to executives and managers. Here are some of the ways vendors and customers are accomplishing this.

Spreadsheet-like interface for reports

- Operational reporting tools can have a familiar spreadsheet-like interface and enable users to move BI data into common user formats for further analysis (e.g., Excel spreadsheets or Adobe PDF documents).

On-demand, ad hoc, self-service reports

- Operational reports are on-demand, ad hoc, and self-service. On-demand means the reports are available whenever users need them. No one needs to request the report from IT. Ad hoc means that reports/queries can be unpredictable and still perform well. Self-service means the user does not need help from IT to create, modify, or manipulate reports. Some BI tools can deliver reports with embedded report-manipulation functionality and detailed data so the user doesn't need separate software installed to view, manipulate, drill down, etc.

On-demand and event-driven analysis

- Analysis can be on-demand or event-driven depending on the user's needs. An operational dashboard could be continuously updated based on source system events, or updated periodically on demand.

Sophisticated data visualization

- Reporting and analysis are integrated with sophisticated data visualization capabilities, including geographic information system (GIS) and mapping.

Access to BI through the enterprise portal

- Reporting and analytical applications are accessible via an enterprise portal, providing integrated security and sign-on. Here again, users may not need any software installed to view, manipulate, drill down, etc. This is particularly important for mobile users and external users.

Avoid the issue of multiple answers to the same question

- For operational BI solutions implemented via EII, which combines various sources of data in a real-time fashion, including a “currency date” on reports and analysis (the date and time the data was retrieved) avoids confusion if two users have the same report but different data.

Predictive analysis

A future goal of many organizations is to move beyond first-generation, reactive BI applications (knowing what happened in the past or even what is happening now) to gain enhanced business value from the ability to better predict what will happen in the future using predictive analytics (see Figure 5).

Moving to Predictive BI Applications

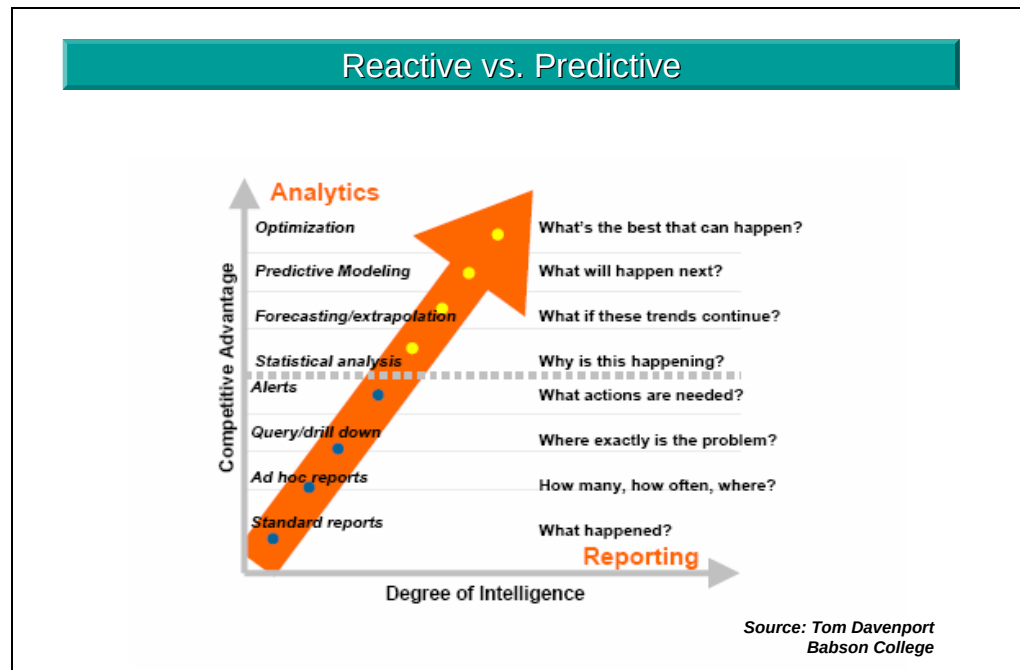


Figure 5.

Boundaries between BI and PM are blurring

In this area, the boundaries between operational BI and performance management (PM) have begun to blur. PM has expanded from its original focus on financial performance management to encompass “operational performance management” (OPM) in the areas of sales and marketing, human resources, operations, etc. As BI applications become more operational and predictive in nature, there will be an increasing overlap in focus.

Recognize the value of access to unstructured data

Another trend is access to and integration of unstructured data content within the BI solution. This has two aspects. One is that a vast majority of existing data in most organizations is unstructured and there is potential analytical value to be unlocked if the data can be accessed in an efficient and meaningful way. The second aspect is enabling BI users to capture text comments to provide context for BI data. An

example is a call center operator entering notes to document the results of a phone dialog with a customer. The operational BI solution must be able to effectively process the unstructured text data.

Integrate BI with search capabilities

A related trend is integrating search capabilities within the operational BI solution to facilitate analysis of both structured and unstructured data. Approaches here include the use of search engines instead of EII to access data, and the use of business taxonomy in the form of faceted search to drive the search engine.

DECISION AUTOMATION

Automated BI agents for alerts, recommendations, and actions

We have already discussed the use of automated (event-driven) agents to assist in the analysis and action segments of the decision cycle. These agents include alerts, recommendations, and actions. The key is building knowledge about the business into an automated analysis and decision process so the agents can help users analyze and make decisions that impact the daily business. Thus, event-driven BI monitors the health and well-being of the organization. When exceptions occur, the BI application alerts someone and may include recommendations about what action should be taken, or the application may make a decision automatically.

BI AS A SERVICE

BI is an on-demand, embedded, reusable service

Encapsulating BI functionality as a reusable service that can be called as required by any application or embedded in a business process is a key component of operational BI.

SOA to build BI services internally

One option for this is to implement an in-house service-oriented architecture (SOA), enabling the organization to create BI functions as Web services (although the use of Web services is not a prerequisite). An example is encapsulating a customer-scoring process and embedding it in a customer service application. When a customer phones in, the customer service application runs the embedded scoring process and gives the result to the customer service rep who can then use the score as a guide to handling the contact.

SaaS to outsource BI applications

Another option is to outsource BI applications using a SaaS (software as a service) approach. Here, an organization uses a third-party, commercial software computing service to run a business application. An example is calling a third-party tool to analyze external marketing data as part of an evaluation of a marketing campaign.

KEY STEPS IN IMPLEMENTING AN OPERATIONAL BI SOLUTION

Critical success factors

How does an organization implement an operational BI solution? Several steps are important to success and to ensure acceptance, adoption, and active use of the BI solution as a platform for making better business decisions.

Top management support is always important*Get top management support*

Customers consistently stress the critical importance of securing visible and active top management support for the BI effort.

**Business ownership
of the BI solution***Ensure business ownership of and commitment to the BI solution*

Business involvement and collaboration in the development of an effective BI solution is very important. Establishing a partnership between business and IT drives acceptance and usage of the BI applications.

**Organize to balance
short- and long-term
goals***Ensure a balance between short- and long-term BI goals*

It is also important to effectively balance short and long-term BI goals, and ensure that long-term strategic goals drive specific EDW project efforts.

**Ensure adequate
resources for
development and
delivery***Start off with the right resources*

Choosing the right resources in the beginning goes a long way toward ultimate success. We discussed above the need for ongoing support, commitment, and participation from top management and business users. It is also important to have the requisite IT resources, in addition to business support, to design, implement, and deliver an effective operational BI solution.

**Data quality drives
acceptance of BI***Ensure data integrity and quality*

Giving users the quality of information they need to make good decisions is a critical requirement. BI content must be accurate, informative, and easy to use to ensure that users make proper business decisions based on the information presented.

**Build flexibility into
the data structure***Make sure you have the right data structure and metadata*

Making good decisions about data structure and metadata management is another key to success. One example is designing flexible BI data structures to accommodate unexpected and unpredictable business requirements. Another aspect of this is careful mapping of data to the business model to create effective business views of the data. Business views are abstractions of the appropriate data elements, using familiar terminology, for a particular type of end-user analysis.

**User support drives
effective use of
EDW***Deliver a robust system and provide comprehensive user support*

A critical success factor is giving users a system they want to use because it delivers what they need. It is also important to ensure that people have the right tools and are effectively using the operational BI solution.

CASE STUDY: PROPERTY/CASUALTY INSURANCE COMPANY

ORGANIZATION BACKGROUND

Property and casualty insurance and services

This organization is a U.S. property and casualty insurance company. The company has expanded beyond its original single-state focus to become a national insurer licensed in a majority of states. The company has fewer than 1,000 employees, offers its products through independent agents, and sells primarily to small- and medium-sized businesses.

For this case study, we interviewed the Enterprise Data Warehouse manager and members of his development and implementation team within the Information Systems (IS) group.

THE BUSINESS PROBLEM

Inflexible, “stovepipe” systems and applications

In late 2003, the vice president of IS decided that the company was falling behind in two key areas: data management and Web development. The company had an existing operational data store (ODS) that had been built over many years. This had started as an effort to make mainframe data available to a client/server environment. Subsequently, the ODS was enhanced to provide data from multiple sources. However, it had grown incrementally in a “stovepipe” fashion with no overarching vision of its objectives, architecture, or evolution. And there were no standard data definitions among the stovepipes. Now, when the business side requested support for new requirements, the cost of modifying the ODS was far higher than it should be for two reasons: the inflexibility of the ODS architecture and a lack of data management expertise within IS.

On the Web side, the organization had outsourced the development of point solutions from a variety of vendors. As a result, every new Web application had a different underlying architecture, essentially replicating the stovepipe problems of the ODS environment. This “many-to-many” environment had geometrically increased the possible points of failure in the system and the cost of maintenance.

Increasing competitive pressure

The VP was also concerned about the growing importance of data mining and other business intelligence (BI) analytics that she saw competitors adopting. The company would have to support these capabilities in the future to be competitive.

New enterprise data warehouse platform for BI

After researching how the company could change in these two areas to build in flexibility and move forward, IS staff came to the conclusion that it would be too costly to “fix” the existing ODS. Instead, the company decided to design and implement a new enterprise data warehouse (EDW) and acquire the appropriate skills and training to accomplish this. The EDW would provide a single view of the business and the foundation for future BI analysis and reporting.

Need to integrate data from new source systems

An important requirement was the ability to easily integrate data from new source systems. For example, the company goal is to minimize the number of policy and claims systems it supports. However, as the company adds new lines of business, acquires other companies, and/or gets licensed in new states, there may be a need to integrate additional existing policy or claims systems within the data-warehouse environment. As the EDW manager states, “In this industry, IT doesn’t lead the business. The business tells us what we need to do and we find a way to make it happen. As a result, we have to accommodate whatever source system comes along.”

Enterprise portal and application server for Web apps

IS staff also recognized the need to develop an enterprise Web portal and an enterprise application server on which to build a “best practices” architecture for Web applications.

In 2004 the company hired additional staff to design and build the first release of the new EDW and enterprise Web portal, and develop a plan to phase out the old ODS and Web applications.

THE OPERATIONAL BUSINESS INTELLIGENCE SOLUTION

An Enterprise Data Warehouse on an Oracle/Sun platform

THE ENTERPRISE DATA WAREHOUSE. The EDW is a 40GB Oracle 9i database running on Sun Solaris servers. There are two components of the EDW, each with its own schema within the EDW database. One defines the atomic-level warehouse (individual transactions stored in third-normal form) and the other defines the Data Mart (facts and conformed dimensions stored in a star schema). As described below, the atomic-level warehouse is the basis for the Data Mart.

The decision to implement a single, centralized EDW had important benefits. First, it ensured the existence of a “single version of the truth,” a single business view of the data. The company saw this as key to providing the greatest business value with minimal maintenance effort. Second, other organizations that had initially implemented multiple physical data marts were now experiencing significant data-consistency issues; many were undertaking extensive and expensive data consolidation efforts. The insurance company wanted to avoid this by starting out with one, integrated BI data source.

Syncsort’s DMEExpress is the ETL tool

The company uses Syncsort’s DMEExpress 2.3 on Windows as the extract, transform, and load (ETL) software for updating the EDW. DMEExpress enables the EDW to integrate data from several source systems:

- *Internal production systems* – Two primary source systems are internal Oracle production databases running an integrated policy/claims system and a second claims system.
- *Third-party systems* – Two source systems are hosted offsite by a third-party vendor in an IBM DB2 database on an AS/400 platform. Here, two policy systems receive feeds from one other claims system. The vendor sends the insurance company data from these systems nightly as a flat file which is then moved into an Oracle staging area.

The company plans to upgrade to DMEExpress 3.0 to take advantage of enhanced search capabilities for impact analyses and an improved interface for viewing data.

DMExpress reduced time to rebuild EDW to 7 hours from 4 days

The latter enables an analyst to view both a data layout and an actual data file within DMExpress, and to overlay one on the other, to research ETL process problems.

WEEKLY REFRESH. Currently, the company creates a full refresh of the EDW weekly. Data as of the close of business on Wednesday is available the following Monday. This schedule is based on the fact that the original ETL software took 40 to 60 hours just to transform the source data and load the EDW. Because the company wanted refreshed data available Sunday night, it had to start the process the previous Thursday. After the company replaced the original ETL software with DMExpress in 2005, that same process now takes only seven hours, a dramatic improvement. Here are the steps involved:

- EDW staff requests a copy of the internal Oracle production data and a copy of the Oracle staging area (with the data from third-party systems) as of the close of business Wednesday. IS makes a backup of the data and restores it to the EDW load area. The next step is to execute a set of Oracle materialized views that perform an initial data extraction and transformation for the EDW. By Thursday noon the data is available to DMExpress.
- The DMExpress ETL process then runs against the materialized views and builds a complete new version of the atomic-level EDW data. A second DMExpress process builds the Data Mart from the atomic data.
- EDW staff validates the EDW data using control totals.
- The new version of the EDW is then copied to the production EDW server Sunday night.
- The EDW with data current through the previous Wednesday is available Monday morning.

Moving to an incremental update

The company hasn't yet changed its weekly refresh schedule even though DMExpress has greatly reduced the ETL processing window. One consideration is the fact that there are other steps before and after DMExpress that extend the overall time to refresh the EDW beyond the seven hours it takes DMExpress. However, users would like less latency in the EDW, especially at month- and quarter-end. Some users want the data refreshed on a daily basis. The current goal is to work toward a daily incremental update. The next step in achieving that goal would be to cut off on Friday with new data (still a full refresh) available Monday morning to reduce data latency. This is now possible with DMExpress. After that, moving to an *incremental* update would further reduce the time required to update the EDW because only changes would be processed (see "Future Issues" below for more discussion of this.)

DMExpress will improve the quality of offsite data

OFFSITE DATA QUALITY. One problem with the offsite systems is data quality. The insurance company is not privy to the code that extracts the data, and the "once-removed" offsite claims system is essentially a black box. The third-party vendor has service-level agreements in place and does not want the insurance company to access the data directly in case performance issues arise. In the future, the company plans to have the vendor mirror the data nightly on a separate AS/400 partition. The company can then control the data coming from these systems by using DMExpress to grab the data directly from the mirrored partition without affecting performance.

Both internal and external access to reporting

EDW REPORTING. All of the currently available reports represent pre-defined queries with parameters the end-user can specify. Reports do not feature drill-down capability because key detailed data are included on the reports.

For internal users, Business Objects is the primary reporting tool. A critical report is the experience report used by underwriters to monitor their book of business – written premium policies, loss ratios, etc. – broken down by agent, by region, and other dimensions. This information is updated weekly, as described earlier, and the underwriters typically use it on a monthly basis.

The EDW also provides several reports that address issues on the claims side. One is the claims severity report to identify trends in medical claim costs, such as whether costs are rising or declining and types of accidents by industry.

Recently, the EDW group extended access to the EDW to its independent agencies. Each agency can run its own performance report to assess the book of business it has with the insurance company. A key measure is the comparison of premium dollars in to claims dollars out. This external access is supported by the BEA WebLogic Server platform and custom Java code. This approach does not require an agency to have or install any client software in order to access the reports. These reports are also available internally using the same Web interface.

Enhancements coming Q1 2007

THE NEXT RELEASE. The EDW group is developing the next release of the BI solution and will roll it out in the first quarter of 2007. Enhancements fall into three areas:

- Development of virtual data marts to support ad hoc queries and additional user groups, including Actuarial. Each virtual data mart represents a different business view of the data, yet all are based on the same “single version of the truth” contained in the EDW.
- A redesigned star schema to improve usability and performance.
- Expanded end-user support from an EDW data-delivery team to ensure effective use of the EDW.

The EDW group views this major release as the springboard that will transform the EDW into a key strategic asset.

Virtual data marts for ad hoc queries

AD HOC ACCESS. The EDW group is building the virtual data marts (separate business “views” of EDW data) as universes within the Business Objects environment. Each universe provides an abstraction of the appropriate data elements, using familiar terminology, for a particular type of end-user analysis. The goal is to support ad hoc queries for research and analysis while minimizing the risk of users creating inaccurate reports.

Three universes are currently in pilot mode with five users: a policy universe (for policy analysis of premiums, credit usage, losses), an accident-year universe (to support claims analysis based on the year an accident occurred), and a claims universe (to slice and dice claims data). The plan is to make nine universes available to over 50 users for ad hoc analysis and reporting in the first quarter of 2007. Included are the company’s actuaries, “who have by far the most demanding data needs in the company,” states the EDW manager. The actuarial group will use the

EDW for all of its research and analysis. “If our actuaries trust the data, everyone else will, too.”

Improved star schema

SCHEMA CHANGES. The current design of the EDW star schema is not as denormalized as it could be, creating both usability and performance problems. According to the EDW data architect, “We got the center of the star right, but are revising the dimensions to minimize joins and make the data structure easier to understand. The EDW will only have dimension tables that can define the granularity of the fact table.” The result will be a “best practices” star schema design that provides a strong foundation for the virtual data marts.

User education and training is critical to success

DATA DELIVERY SUPPORT. A very important component of the overall BI solution for this company is its EDW data delivery team. In fact, three of the eight people in the EDW group focus on this area. As the EDW manager puts it, “We have invested a lot in the quality and consistency of our EDW data. We don’t want the value of that undermined by a poorly supported user.” The data delivery team will provide several levels of support to ensure users understand the EDW and can use it effectively. These include education and training classes (to learn about EDW data and how to use Business Objects), release notes (a version of the EDW metadata in business language identifying sources of data and cross-referencing EDW fields with existing ODS data), user group meetings, and monthly tip sheets that focus on specific Business Objects functionality.

While Business Objects is the preferred BI tool, others are in use and the data delivery team needs to learn and support all of these different tools.

IMPLEMENTATION STEPS AND ADVICE

Focus on tools, skills, data quality, business involvement

The IS department encountered many initial issues along the EDW development path. These include decisions about what toolsets to use and support, the level of technical experience needed versus what was available, and how to deal with data quality problems.

In late 2004 and early 2005, the company ramped up to invest in the EDW with the goal of creating a “healthy” data-warehouse environment. One step was establishing a corporate data-quality program. Another was getting people on the business side involved. A third was hiring an EDW manager to oversee the overall effort. As a result, through 2005, the EDW effort really took hold. IS implemented several successful releases and data quality was much better than it ever had been.

The newest release of the EDW, as described above, is slated for Q1 2007. After that, the EDW group will implement a series of technical infrastructure enhancements. (See “Future Issues” below.)

ADVICE. The EDW manager offered this advice to others based on his company’s experiences.

Business ownership of the BI solution is key

- *Ensure business ownership of the BI solution* – Business involvement and collaboration in the development of an effective BI solution is very important. Establishing a partnership between business and IT drives acceptance and usage of the BI applications. In this organization, the EDW team has worked hard to

understand the business requirements. The flip side is to help the business understand technical requirements so they can fully support that aspect of the endeavor.

Organize to balance short- and long-term goals

- *Establish an organizational structure that ensures a balance between short and long term BI goals* – To effectively balance short and long-term BI goals, the insurance company has dedicated a full-time team to EDW development and implementation, and manages the effort as a high-level program rather than as a series of IT projects (see “Critical Success Factors” below).

SUMMARY OF BENEFITS AND ROI

Improved access to data for both internal and external users

BENEFITS. The company has realized significant benefits from its operational BI solution.

- *Faster access to high-quality data* – The creation of a single, integrated data source for all analysis and reporting has dramatically improved data quality, improved performance, and reduced the effort required to generate business reports.
- *Improved customer relations* – The quality of the data in the EDW has reduced negative feedback from agents who claim “that’s not the right number,” and eliminated the need for internal people to then track down the “right” number among multiple data sources. In addition, the availability of the EDW for agency performance reporting offers added value to these customers.
- *Better business decisions* – Ad hoc query and reporting will support key analysts who make critical decisions that can have a huge impact on the bottom line. Examples are setting reserves and prices. An insurance company needs good assumptions about how it expects losses to develop over time. The better the company is at risk assessment, the better the business decisions will be.
- *Flexibility to adapt to new business requirements* – The DMExpress software helps here in two areas. High-speed ETL processing gives the company the opportunity to reduce the latency of its EDW data in the future. And the ability to access a wide variety of data sources will facilitate the incorporation of new data sources in the EDW.

Comprehensive support for key business decisions

In summary, the EDW has become a flexible service platform on which to build future BI applications for both internal and external users.

ROI. The company does not have a formal program for measuring ROI for the EDW. Rather, the EDW is viewed as a basic business requirement for growth and continued competitiveness.

SUMMARY OF CRITICAL SUCCESS FACTORS

The EDW manager mentioned several factors critical to the success of the BI effort.

Data quality drives acceptance of EDW

- *Data quality* – Having a single version of the truth that maps to a single, accepted set of business definitions has been an important foundation for acceptance of the EDW. The EDW group regularly reconciles with finance the four key numbers on which the business depends. If there are any discrepancies, IS and the business side work together to identify the problem and find a solution. “Defining those four metrics and reconciling them has led to a lot of trust in the EDW. The finance stamp of approval means a great deal to us,” stated the EDW manager.

The business drives the BI solution

- *Business collaboration and partnership* – An EDW business support team, made up of business users, meets weekly. This group deals with data quality issues, helps IS define and refine business and functional requirements for EDW releases, provides testing resources, and helps manage expectations for the EDW within the business community. The fact that this team is empowered to make decisions about the EDW is a clear example of the fact that the business drives the BI effort and not the reverse.

User support drives effective use of EDW

- *Comprehensive user support* – The data delivery team is key to ensuring that people are using the right tools and effectively using the data to its full capacity.

Balancing short- and long-term goals

- *Program-level management of the EDW* – The EDW effort is managed and overseen at the program level within the company. This puts the EDW on the same level as other programs (e.g., customer relationship management, master data management, and Web strategies) and separates it from the IT project portfolio planning process, where priorities can change over time. This ensures that long-term strategic goals drive specific EDW project efforts.

FUTURE ISSUES

Future enhancements to the operational BI solution include the following.

Build a more flexible EDW infrastructure

Move toward a more flexible EDW infrastructure – The company plans to “retire” the old ODS in 2007 while expanding use of the EDW. It is not clear yet whether this will lead to more frequent loading of the EDW from source systems. There is a desire among the business units for daily updates, but the business case is still being developed. One potential limitation to daily updates is how frequently source systems can provide data. The company is building its systems to support incremental updates and may go to daily updates. For incremental updates, DMExpress would detect changes to source systems using its full-file-compare capability, and then push the changes to the EDW. As stated above, this would reduce the time and effort required to update the EDW.

Componentize the ETL process

Another goal is to more fully “componentize” the overall ETL process. The current process splits ETL functions as follows. Creating the materialized views is an extract/partial transform (“ET”) process. DMExpress then completes the transformation process and loads the EDW (a “TL” process). Eliminating the

materialized views and giving DMExpress a full ETL role will eliminate the duplicate transformation (“T”) processes and greatly simplify the EDW maintenance environment.

According to the EDW manager, “IS has to be able to respond to whatever we are asked to do in the future, and we are working to implement an infrastructure that can deliver this. We don’t know for sure what competitive developments will occur in the insurance business. Today, customers expect one- to two-day turnaround cycles for their insurance information. But as Web use increases, we may need operational systems that better support much faster access and processing. In summary, our EDW strategy cannot lead the business somewhere. We must have the flexibility to follow on the heels of the business strategy as closely as possible.”

Assess data retention to offset growing volume

Develop retention policies for the EDW – The EDW manager is predicting 500% growth in the volume of data in the EDW over the next year with the addition of monthly snapshots. “We need to get more sophisticated in developing retention policies. We want to keep only the data we need.”

Integrate BI with the enterprise Web portal

Integrate Business Objects reporting with the enterprise portal – Currently, internal users do not access EDW reports through the enterprise Web portal. Upgrading to the latest version of Business Objects (XI Release 2) in early 2007 will enable the company to integrate Business Objects within its federated portal architecture. Advantages include:

- Better security – This version of Business Objects offers better management of virtual/multiple data marts built on a centralized EDW.
- Authentication tie-in – The insurance company will assess options such as simplified or single sign-on.
- Compliance with basic portal standards – Plusses here include the ability to deploy Business Objects through the portal and the ability to define Business Objects as another content type in the portal environment.

SUMMARY

Flexibility to handle new business requirements

This company has come a long way in three years. Arriving relatively late to the party enabled it to learn from early adopters and avoid potential missteps in implementing a strategic BI solution. Key success factors have been the emphasis on close and ongoing collaboration with the business side, and taking the time to build a flexible BI platform that can support new and changing business requirements.

Staying competitive in a dynamic industry

The company plans to move forward and enhance this platform with support for more frequent incremental updates to the EDW and tighter integration of EDW reporting with the enterprise Web portal. The goal is a BI platform that can stand up to competitive pressures in an increasingly dynamic industry.

CONCLUSIONS

Using BI to manage and optimize the daily business

Operational business intelligence highlights the state-of-the-art in developing BI solutions to support timely, effective business decisions. Operational BI has the potential to extend the reach of BI services, applications, and technologies to all users and business decisions across the enterprise. Operational BI optimizes the BI decision cycle by reducing latency between data collection, analysis, and action. It also optimizes overall business processing by integrating business transaction systems and BI applications. The goal is to take advantage of BI in managing daily business operations.

Implementing an operational BI solution is not without its challenges. As organizations reduce data latency, increase data granularity, extend the reach of BI to operational users and processes, and enhance the sophistication of BI applications, the performance, scalability, and reliability of the BI platform and IT infrastructure becomes ever more critical. They must be able to effectively handle larger volumes of data, many more concurrent users, and a mixed workload of both complex, ad hoc queries and data loading.

Centralized BI for a single view of the business

The property/casualty insurance company has implemented a centralized EDW. Unique departmental or functional requirements are handled as business views within the EDW environment, not as separate, physical data marts. This reduces the number of times data is moved, and the costs associated with synchronizing copies of data. Creating a single view of the business, a single version of the truth, is a key business goal.

Extending the reach and sophistication of BI processing

Other emerging operational BI developments with significant potential benefits are:

- Support for on-demand, ad hoc, self-service operational reports so users can get what they need when they need it without help from IT.
- Integration of operational BI reporting and analysis with sophisticated data visualization capabilities to help users better understand business data.
- Delivery of operational BI reporting and analysis through an enterprise portal to enable all users – both internal and external – to gain the advantage of BI.
- Advanced analytics to better predict what will happen in the future.
- Encapsulation of BI functions as common, reusable services. Business processes, new or old, tap into existing knowledge about the business, reducing redundancy and enhancing both the efficiency and effectiveness of the organization as a whole.

Operational BI: a foundation for improving the business

The property/casualty insurance company stresses the significance of operational BI as an integral business system that adds value to the business. In addition, the operational BI solution described in the case study provides a solid BI platform going forward. This company plans to enhance its strategic use of operational BI to gain distinctive advantage within its industry. The goal is to work smarter, become

more agile, and effectively adapt to changing business requirements in an increasingly competitive environment.

References

- ¹ Hackathorn, Richard. "The BI Watch: Real-Time to Real-Value," *DM Review*, January, 2004.
- ² Imhoff, Claudia and White, Colin. "A Guide to Operational Business Intelligence: What It Is and How to Use It," seminar presented at the DAMA International Symposium in Boston, March, 2007.
- ³ White, Colin. "Data Integration: Using ETL, EAI, and EII Tools to Create an Integrated Enterprise," research report published by The Data Warehousing Institute, October, 2005.
- ⁴ Imhoff, Claudia and White, Colin. "Master Data Management: Creating a Single View of the Business," research report published by the Business Intelligence Network, September, 2006

SYNCSORT BUSINESS INTELLIGENCE SOLUTIONS

COMPANY BACKGROUND

High-performance software for sorting, backup, and data transformation

Syncsort Incorporated develops high-performance data management software for mainframe, UNIX, Windows, and Linux environments. SyncSort is the company's high-performance commercial sort and data manipulation utility. DMExpress builds on the power of SyncSort to provide additional extract, transform, and load (ETL) functionality. Backup Express is Syncsort's data protection/backup and recovery solution. A privately-held company headquartered in Woodcliff Lake, NJ, Syncsort has been in business for more than 35 years. Today, over 12,000 sites worldwide and 90 of Fortune 100 companies use Syncsort products. See www.syncsort.com for more information.

SyncSort originally an IBM mainframe product

The company debuted as a vendor of sorting and data manipulation software with the original SyncSort product on IBM mainframes. SyncSort's fast performance greatly reduced the time to process large volumes of data and led to high penetration in this market. According to Syncsort, the product is still used on more than half of IBM mainframes. The current release of SyncSort runs on UNIX, Windows, and Linux in addition to the mainframe. SyncSort provides record filtering, reformatting and summarization for file and data conversion, and generates reports.

DMExpress focuses on ETL functionality

The company introduced DMExpress in early 2004 to target ETL processes in data-intensive applications. DMExpress is the evolution of SyncSort for UNIX, Windows, and Linux. The current version of DMExpress is Release 3.

THE SYNC SORT FOUNDATION

Speed is key

Some background on SyncSort is helpful to understand the foundation for DMExpress. SyncSort was designed to sort, filter, reformat, and convert huge amounts of data very quickly. Over 30 years, the company received a number of patents on the sorting, aggregation, and pattern-matching algorithms that make SyncSort so fast. SyncSort also optimizes performance by exploiting the underlying hardware and operating-system architecture, and executing operations in parallel where possible.

Call-level detail, Web logs, credit card transactions, barcode scans

Most SyncSort customers are in industries that typically process very-large data volumes. For example, the company states that every major telephone company in the world uses SyncSort somewhere, usually to handle the processing of call-detail records. Telephone switches collect hundreds of gigabytes of call data every day. The data is not organized in a way that can be easily analyzed, yet the telephone company wants to use the data to send bills and analyze network traffic. SyncSort reorganizes and filters the detailed data collected by the switch and loads it into a database for analysis and processing.

Another example is heavily-used Web sites such as AOL, Expedia, and Travelocity. The log files for these sites are enormous because every time a site visitor clicks the mouse, a record is written on the Web server. Each site wants to know how many visitors it had, which page was most often requested, and other analytic information. SyncSort applies its pattern-matching, counting, and aggregation algorithms to extract only required data from the detailed data records (which may contain extensively verbose information, e.g., URLs) and move the data to an environment where it can be analyzed. The issue here is that the customer cannot simply dump a huge volume of raw data records into a database because the loading process takes too long, and much of the data is not relevant. Other examples include credit card transactions and supermarket transactions based on barcode scans.

Derivation of changes to data

Customers also use SyncSort to effect changed-data capture, that is, identify changes that have occurred between two points in time. For example, AT&T receives an enormous amount of new and revised customer and inventory data on a daily basis. The company needed to be able to compare one day of AT&T's Local Network Inventory data with another day of data, and then output the difference between the two files. A SyncSort application was created to compare the "yesterday" and "today" files and update, add, and delete the records as appropriate. Performance becomes critical because SyncSort has to process up to 200% of the data in order to find changes, which might be only 1% of the data. This same approach can be used to find only new records or missing records, eliminate duplicate records, etc. in a large volume of data.

DMEXPRESS

ETL functionality to move, transform, and aggregate data fast

Advancing SyncSort's capabilities, DMExpress is designed to provide additional high-speed ETL functionality. DMExpress is aimed at the market for data movement and transformation in data-warehousing (DW) and business-intelligence (BI) environments. The company's strategy here focuses on the speed, performance, and scalability of the DMExpress ETL solution.

Heterogeneous operating platforms, sources, and targets

DMExpress extracts data from the source, transforms it on the DMExpress server, and loads it into the target. DMExpress runs on HP-UX, IBM AIX, SunOS, Linux, and Windows. DMExpress supports 64-bit processing on Intel and compatible platforms.

Sources and targets can be a relational database management system (RDBMS) (IBM DB2, Microsoft SQL Server, Oracle, Red Brick, Sybase, Teradata), Microsoft Access and Excel, flat files, structured files, gzip compressed format, pipes, and data in memory. DMExpress also accepts XML as a data source. DMExpress handles multinational encoding with support for multibyte text and Unicode. In addition, target data partitioning enables DMExpress to automatically partition data from an input file on the target by size, field values, or conditions.

Extensive data transformation

Data transformation capability is flexible and includes:

- *Source-level operations* – All types including conversion of database tables to/from flat files.
- *Record-level operations* – Aggregate, join, sort, merge, copy, filter, reformat, etc. Aggregates include total, count, average, minimum and maximum on multiple dimensions. Multiple rows on the source can be consolidated on the target and vice versa.
- *Field-level operations* – Data type and format conversions; pattern matching; arithmetic, string, date-time, conditional operations.

The DMExpress Expression Builder enables the user to define ever more sophisticated business rules using pre-defined functions and user-defined variables. The ability to set and retrieve values in any number of user-defined variables supports complex processing and transformation.

In every release, Syncsort has improved DMExpress's high-speed join, aggregate, and transformation processing to better handle changed-data-capture requirements and very large volumes of data.

The Job Editor manages metadata and applications

A major advantage for DMExpress is the Windows-based Job Editor. SyncSort is essentially a product for the systems programmer, someone comfortable writing commands at the command line. In contrast, the Job Editor is a visual management tool that builds, tests, runs, and monitors DMExpress applications. An important feature of the Job Editor is the metadata layer for storing record layouts, business rules, transformation definitions, run history, and data statistics. The Job Editor also supports graphical representation of record structure (to build and compare source and target data layouts), and data sampling (the ability to display the first few records of the file). The Job Editor supports multiple languages, including French, German, and Japanese in addition to English.

Flexible job execution

DMExpress jobs can be composed of multiple tasks, including subjobs, or jobs defined within other jobs. All of the components are reusable. The administrator can specify where and when each job should run. Execution can be immediate or scheduled on any supported server. For example, from the Job Editor, the administrator can point to a UNIX server and run a DMExpress job there. Direct

flow editing improves speed and performance by eliminating the writing of intermediate physical files.

The Job Editor includes the ability to automatically create scripts in several programming languages for (1) executing jobs from within a script if appropriate, and (2) integrating DMEExpress functionality in other applications. Thus, a DMEExpress job can become a service invoked by another application. DMEExpress can integrate with any software from the command line, and can invoke any software from the command line. DMEExpress can also run SyncSort applications for backward compatibility.