

Compliance is your new **BEST FRIEND**

A well-architected data warehouse infrastructure allows you to re-use required data across the organization.

by Stephen Brobst and Bai Shuo

IT spending in support of regulatory and compliance projects is significant and is expected to increase in the coming years. Gartner figures indicate that 12% of IT project budgeting in 2006 was directly related to compliance and that this figure will likely exceed 15% in 2007 [1]. At first glance, this appears to be a huge strain on an organization's ability to innovate and provide competitive differentiation from within fixed (and often shrinking) IT budgets. However, innovative organizations are turning the compliance liability into an asset through leverage of an enterprise data warehousing infrastructure.

Much of the regulatory compliance burden that organizations are under today is focused on the delivery of reports requiring access to significant amounts of detailed historical data. For example, Basel II stan-



dards for financial risk reporting typically require a minimum of seven years of detailed data and will likely require 11 years in the near future. One bank in the United Kingdom has a Basel II risk reporting

requirement of 25 years of history because of the extended economic cycles associated with real estate and its book of business, which is heavily weighted toward long-term mortgages. Regulatory compliance for

financial reporting within guidelines adhering to the Sarbanes-Oxley (SOX) Act in the United States, its equivalent in Japan (J-SOX) and International Accounting Standards (IAS) also require increasing amounts of detailed data. Summary data is no longer good enough. Moreover, significant retention of historical data is required for trend reporting and audit purposes.

The key to exploiting the regulatory dividend is to leverage the costs associated with compliance into an asset that delivers competitive advantage to the enterprise. By reusing the data required for regulatory reporting to provide analytics that drive better decision making within an organization, significant value can be exploited beyond checkmarking an auditor's compliance box. However, obtaining this value depends on a well-architected data warehouse solution in order to obtain the benefits of data re-use.

The wrong approach

In the rush to "check the box" for regulatory compliance, an organization can easily fall into the trap of data mart deployment without architecting for the potential to re-use data. The primary principle of data mart deployment is to design to the specific needs of a particular business process. Content and organization of information in a data mart is optimized for performance and usability aligned to a distinct purpose.

This type of deployment will sometimes be appropriate for analytic applications with extreme performance requirements, but it can limit the use of data when physical structures or content are overly customized.

The danger in this approach is that each regulatory requirement for reporting results in a new data mart. Information needed across regulatory requirements is not shared, and the data marts are not leveraged to provide analytic capabilities outside the specific regulatory mandates for which they were built. It is not unusual to observe overlap as high as 70% in obligatory data across regulatory reporting requirements.

This scenario increases the cost burden to the organization because information is duplicated in the repositories used to satisfy each regulatory requirement. This causes higher IT investment than is necessary in terms of redundant systems, storage, data movement, resources required for data cleansing and integration, and the people required to manage the overall environment.

At worst, the deployment of many data marts creates multiple sources of truth for regulatory reporting within an organization. This increases organizational cost in terms of effort required for reconciliation of numbers across reports. An enterprise with multiple sources of regulatory reporting is likely to have higher auditing and compliance costs associated with inconsistent data sources. Moreover, the ability to parlay investments in data management into competitive advantage is significantly reduced when a data mart is deployed without adequate consideration given to re-use of the information assets.

The right approach

The high-leverage strategy for regulatory reporting is an enterprise approach to data warehouse deployment in which compliance requirements and advanced analytics to drive business differentiation share information. An enterprise data warehouse (EDW) treats information as a re-usable asset.

In an enterprise approach, the underlying data model is not specific to a particular reporting or analytic requirement. Instead of focusing on a process-oriented design, the repository design is modeled based on data inter-relationships fundamental to the business across processes.

While some summarization and other denormalizations may be undertaken for performance reasons, an EDW design makes best efforts to maintain history at a detailed level with all business relationships intact. This allows for more effective re-use of data because any summarization or denormalization performed in a repository

design must assume some pre-conceived knowledge of business requirements. By keeping history at a detailed level, the information in the EDW can be easily used for additional analytic purposes, some of which may not even be known at the time of deployment for regulatory compliance.

The design goal is to source data into the EDW once and use it for multiple purposes. Store once, use often. Regulatory compliance requirements can be used in this way to fund data provisioning into the data warehouse to support industry-mandated reporting. A fundamental prerequisite for success is having an enterprise logical data model (ELDM) in place as a blueprint for organizing information in the EDW. The ELDM ensures that the organization of information is fit for analytic purposes across multiple domains.

Remember: The ELDM is a blueprint, not a detailed design. Creating the ELDM should be time-boxed to four months or (ideally) less. Modifying packaged data model assets specific to an industry (available from many third-party vendors) by 10% to 20% is much more productive than starting from scratch.

Detailed design for the ELDM should be undertaken incrementally using specific projects such as compliance reporting to provide the funding and detailed requirements for deployment. Having the blueprint in place as a starting point ensures that the requirements for each unique project are realized in a manner that creates data assets that are re-usable across multiple areas of the business.

Competitive advantage

Constant pressure exists within IT organizations to do more with less. Re-use of information with an EDW is a very powerful means to achieve this goal. Data cleansing, integration and provisioning into an information repository typically account for 50% to 70% of the cost of the repository's construction. By sourcing data into a well-designed data warehouse, these costs are incurred just once for the enterprise rather than many times in the case of multiple data

Beyond compliance: Shanghai Stock Exchange case study

The Shanghai Stock Exchange was founded Nov. 26, 1990, and was in operation Dec. 19 of the same year. Its functions include providing the marketplace and facilities for securities trading, formulating business rules related to such trading, accepting and arranging listings, organizing and monitoring securities trading, managing and disseminating market information, and regulating members and listed companies. The Shanghai Stock Exchange is directly governed by the China Securities Regulatory Commission (CSRC).

As of Aug. 31, 2006, the total market capitalization provided through the Shanghai Stock Exchange amounted to nearly 3.8 trillion Ren Min Bi (RMB)—equal to 20% of China's gross domestic product. Tradable market value exceeds 1.13 trillion RMB. The Shanghai Stock Exchange boasts more than:

- > 800 listed companies
- > 200 members
- > 3,700 individual and institutional investors
- > 900 listed securities of many instrument types

Strict regulatory requirements exist to ensure integrity of the marketplace. The Shanghai Stock Exchange has analytics in place to detect and prevent insider trading as well as to verify accuracy in corporate financial reporting. Regulatory requirements mandate accumulation of as many as 20 years of detailed trading history for compliance-related analytics and reporting.

Rather than simply incur such regulatory requirements as a cost, Shanghai Stock Exchange has been able to parlay its compliance investment into significant value for its stakeholders through re-use of data assets in an enterprise data warehouse (EDW).

The normalized repository contains more than 1,000 tables and provides an EDW capability that goes far beyond regulatory analysis and reporting. The current raw data volume is nearly 3TB with incre-



mental acquisition of 2-4GB of new data on a daily basis depending on market volumes for the day.

Data acquisition is particularly challenging because gathering financial data from each listed company on the stock exchange involves nearly 1,000 distinct data sources. Rather than build custom interfaces for each of these feeds, Shanghai Stock Exchange adopted eXtensible Business Reporting Language (XBRL) as a requirement for delivery of financial information from each of its list companies. Leveraging this standard provided for a much more practical data

acquisition strategy. After some minor issues related to the handling of Chinese character strings within the standard, this approach proved to be an extremely scalable method of acquiring information from a huge number of data sources.

In addition to use of the EDW for compliance purposes, Shanghai Stock Exchange performs capacity planning for its trading systems by analyzing detailed data to understand order volumes, peaks and valleys, order sources, and so on by day and time of day. The EDW is also used to design new stock indices such as the SSE 50 and SSE 180. Significant analysis is required to profile performance and volatility of the indices, and these characteristics need to be re-evaluated at least every six months. Marketplace simulations and forward trending use historical data from the EDW to understand market quality and volatility.

Shanghai Stock Exchange has also established an innovation lab that makes use of EDW content to enhance its capabilities for the future. Examples of work performed in the innovation lab include creation of techniques for real-time surveillance related to fraud and insider trading, analysis of appropriateness for trading categorization of stocks and construction of derivative-based index definitions.

— S.B. and B.S.

mart deployments. Competitive advantage can be obtained, because with an EDW the investment in regulatory compliance is leveraged into an enterprise information asset to be used for a variety of strategic purposes.

Little doubt should exist as to the most effective investment when pursuing regulatory compliance. However, implementing an EDW is much easier said than done. An

enterprise approach requires an enterprise data strategy and appropriate data governance within the organization. Success requires organizations to think at an enterprise level rather than in departmental silos. Success also necessitates a scalable approach to deployment of the information repository rather than a data mart mentality with point solutions. As usual, the technology is the

easy part and political considerations will be the biggest challenge to success. **T**

[1] Donald Feinberg. The Future of DW and BI. 2006.

Stephen Brobst is chief technology officer of the Teradata Corporation.

Bai Shuo is chief technology officer of the Shanghai Stock Exchange.