

CASE STUDY:

REAL ESTATE ENTERPRISE MANAGEMENT DATA WAREHOUSE

CLIENT DESCRIPTION:

The client is a large facility and real estate management enterprise with an expansive nationwide portfolio of properties under its management. Historically, as the enterprise expanded and evolved, applications had been acquired to deliver specific functional capabilities, such as enterprise financial management (SAP), lease administration (Strategen), and asset management (Peregrine). Typically, the client had selected best-of-breed solutions to satisfy the requirements for each of these functional areas. This case study evolved from the requirement to combine the information from each of these systems to satisfy changing management and customer reporting requirements. Since these systems were not designed to interoperate with one another, this led to a duplication of essential information in multiple systems in order to create and deliver reports reflecting the required combined enterprise data. This situation created a serious data integrity problem for the client resulting from the time lags in data replication across multiple applications, causing internal users to doubt which systems contained valid, up-to-date information, as well as resulting in extensive efforts to maintain enterprise-wide data synchronization.

The client had also discovered that the built-in reporting capabilities of most applications were ill-equipped to deliver the level of customization required. The client clearly required a solution that could provide the flexibility necessary to integrate data from disparate applications, had the ease-of-use associated with web-based solutions and minimized total cost of ownership.

CTA ENGINEERING OBJECTIVES:

Our approach to data integration for this case centered on the phased application of the web-based query, reporting and analysis tool *WebIntelligence* developed by Business Objects (BOWI). Our first task was to establish the feasibility of the use of BOWI for this application by web-enabling a specific set of mission critical reports from data repositories maintained in two of the applications. In the longer term, CTA was asked to determine whether BOWI could be used for all enterprise reporting requirements, providing a flexible platform for integrating data from multiple data sources and seamlessly merging this information into combined reports. In addition, we were tasked to provide specific and detailed recommendations as to the form and function of a general web-based reporting platform that would act as a data warehouse consolidating the data from diverse applications.

APPROACH

CTA chose to accomplish the objectives by:

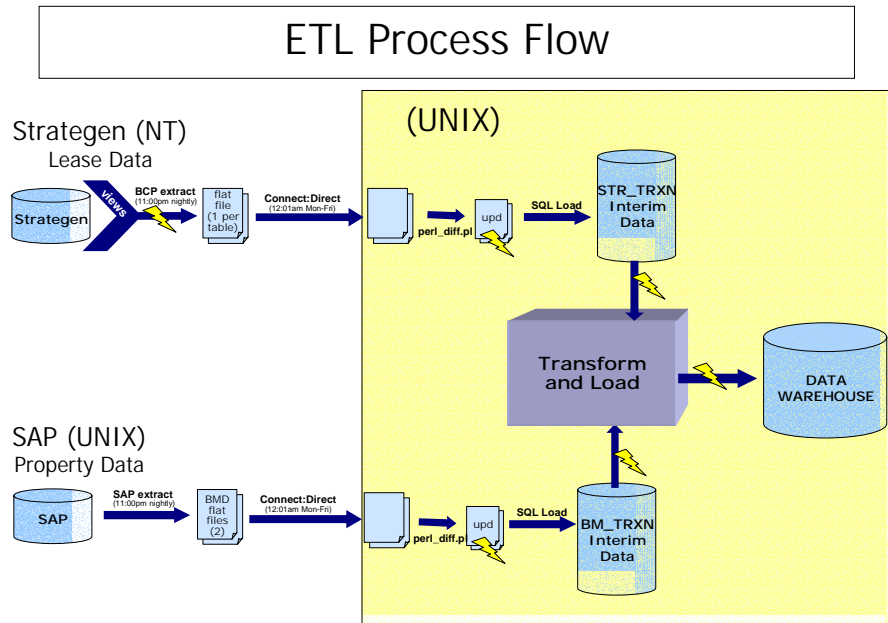
1. Performing an rigorous analysis of current and future reporting requirements and BOWI product capabilities;
2. Demonstrating feasibility of the concept using sample data from SAP and Strategen; and
3. Development of the final systems architecture and data models.

Our proof-of-concept proved the utility of BOWI as a common, web-enabled reporting platform from multiple data sources. We also demonstrated the capability for BOWI reports to perform dynamic filtering and drill-down, much like client-server applications. BOWI also provided the client with ad hoc report creation capabilities. Our proof-of-concept also demonstrated the feasibility of loading both the SAP data and the Strategen data into a single Oracle database. This capability was critical to providing a homogenous reporting environment for all of the client's applications.

Finally, we provided specific and detailed recommendations on the form and function of a general web-based reporting platform that would act as an extensible data warehouse consolidating the data from diverse applications.

SYSTEM ARCHITECTURE:

Our proposed solution architecture consisted of the following elements: a process for the extraction, transformation and loading (ETL) of source data, a hardware/software architecture configuration, and data models for both reporting and gathering of enterprise data entities (or meta data).



Through interviews and follow up discussions with the client, CTA developed an effective, daily ETL process. Each night data would be extracted from the source application, using either pre-defined reports or extraction scripts, and automatically transferred to the system containing the data warehouse using Connect:Direct, a third-party secure FTP application. The data would then be automatically loaded into a database and processed into the reporting data entities.

Extraction

The extraction process encompassed the physical extraction of data from the source system, as well as the migration of this data to the staging area of the data warehouse. Only data that was required by the users for reporting, or was used during the transformation and loading processes, were migrated. Extraneous information was not extracted from source systems to maintain the efficiency of the transformation and load processes. Data related to system functionality, such as SYS.ALL_TABLES (part of an Oracle database data dictionary), or tables containing temporary processing data, are both examples of information that is not meaningful to end-users and was not included in the transformation or loading process of the data warehouse.

Transformation

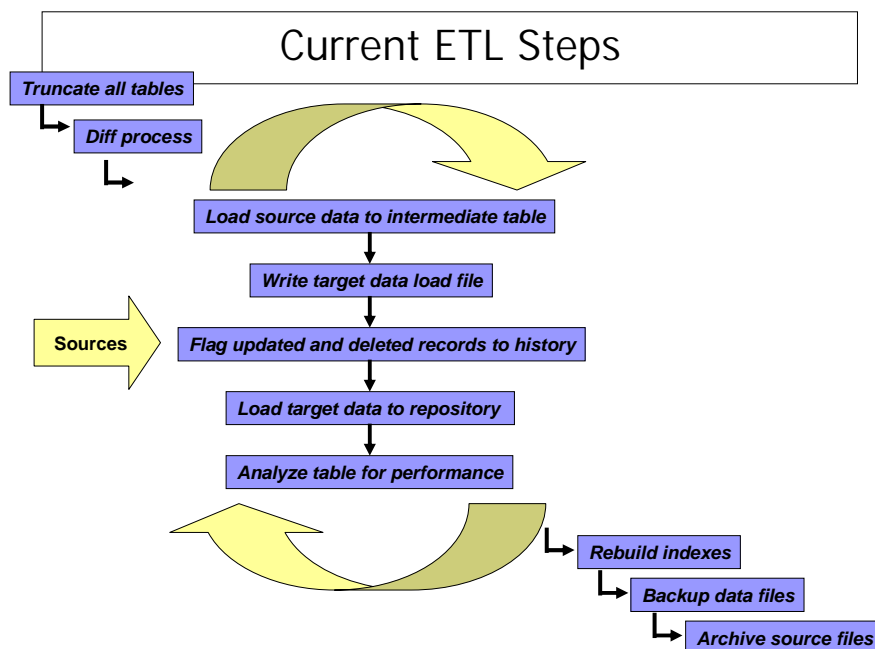
The purpose of transformation was to merge data from multiple sources into a single data model, standardize data values, and to ensure data integrity within the data warehouse. Data transformation would manipulate the data into the format and structure required by the data warehouse. The transformation process would also reduce the number of data elements that were loaded into the data warehouse. The data transformation rules would determine what data elements were loaded. Data elements that were not needed would be excluded from the transformation or load process. The

transformation process would also be used to establish data element relationships and data “aliases”. This process would rely heavily on the use of meta data.

Loading

There were two primary approaches to loading the data warehouse: 1) full refreshes or, 2) incremental refreshes. The full refresh approach began by truncating the tables in the data warehouse and then loading them with all of the required data. The incremental refresh approach identified changes to the source data from the last time the data warehouse was loaded and then inserted or updated rows of data in each table of the data warehouse, as required. Either approach would prevent unwanted data from entering the data warehouse by containing conditions in the load statements.

A full refresh approach worked best when the warehouse was relatively small, or when the process of determining incremental changes required unacceptable trade-offs. An incremental update approach worked best when real-time updates were required, or when the size of the warehouse did not allow ETL to process within the allotted window of time each night. An incremental update approach was also used to allow for tracking of record change history in the warehouse. Each of these approaches was employed in the environment as project requirements evolved.



ACCOMPLISHMENTS

Created a Flexible Data Model

The warehouse data model may be the single most important aspect of the effort. Only a flexible and robust data model would enable the client to integrate legacy systems with modern systems and extend enterprise data as new application systems were sourced. The two primary components of the data model were the reporting data model and the meta data model. While the actual raw data was stored in the reporting data model, it was the meta data repository and its data model structure that allowed the various functional areas within the warehouse to communicate.

Reporting Data Model

The reporting data model represented the actual data stored in the warehouse. It contained both the data and the relationships between the data. In the client warehouse, property information, lease information, financial costs, and personnel responsible for maintaining this data are all examples of entities that existed in the reporting data model.

Meta Data Model

Simply defined, meta data is "data about data." Our premise was that client data should be considered similar to any other client asset with intrinsic value. Similarly then, as the client maintained knowledge about assets, the client also needed to maintain knowledge about the data in the data warehouse. Questions such as, where does the data come from, when does it get loaded and transformed, who is responsible for maintaining it, and most importantly, how does it relate to the other data in the warehouse were all answered by the meta data. As reporting requirements evolved and new systems were integrated, the meta data model provided the necessary translation between older data and new data, enabling the integration of not only diverse systems but also data that existed in different chronological timeframes.

The meta data repository model was the place where information about operational source systems (such as SAP and Strategen); the data warehouse; the ETL processes; the business views; reports; and operational statistics would be maintained. The meta data repository would reduce the delivery speed of new or modified reports through the use of standardized integration templates. These templates would aid future developers as they estimated the impact of proposed changes in the warehouse environment. This type of functionality would be critical for any company looking to manage a data warehouse over time.

Development and Use of Standardized Integration Templates

A data warehouse collects data from the operational systems of a business. Since business rules and data structures are unique to any given operational system, each new data source would need to be interfaced into the data warehouse system. Standardized integration templates were used to meet this challenge. The first step in this process was to define a standardized, generic integration process for new sources. To make this process repeatable, a new source template was constructed that specified a set of data to collect during the discovery phase for any new source. The information captured using this template became the foundation for the functional and physical requirements for that source. This information also enabled the client to identify whether new or merged systems could be readily integrated with existing systems, and provided a guide to the necessary steps required for integration.

Once a new source came online, the next step was to develop a mechanism to maintain the established interface to the source system as the source system changed. Understanding the impact of a major change to a business operational system required careful analysis of both the operational system and the data warehouse. Capturing this information in the integration templates reduced the amount of time spent by the development team in manually analyzing the impact of changes. Used in conjunction with the meta data repository, the integration templates reduced the cost of development and the time needed to capture the data transformation rules, data sources, data structures and the context use of the data in both systems. In addition, the likelihood of development errors was significantly reduced as all impacted programs were readily identifiable.

Standard templates were also created to provide a strategy for the addition of new reports. These templates identified the data elements and specific criteria used in the data warehouse configuration and were used to determine the extent of modifications or enhancements to existing ETL processes for new data entities. The resulting document, known as a "report specification" included all of the information needed to build the report, such as a list of data elements and report formatting considerations. This served as the documented requirements set for each report.

Leveraged the Features of Business Objects' WebIntelligence

The client's previous investment in the Business Objects' WebIntelligence product was leveraged as a common reporting tool across the organization and provides the following capabilities:

- Users and customers can view pre-built reports and build ad hoc reports through a graphical report interface using a web browser across the enterprise network.
- Business Objects tools can be used to control the installation and configuration of BOWI applications across the network, as well as manage user access to these applications.
- The web-based architecture is scalable and highly adaptive to new requirements. WebIntelligence provides high availability, good performance, and the ability to deploy under both UNIX and Windows NT operating environments.
- BOWI can be deployed using multiple security options, on intranet, extranet or mobile platforms. WebIntelligence works with NT Authentication, SSL, DMZ and Proxy servers.

BENEFITS TO CLIENT:

Throughout our client's history of systems development, the primary emphasis had been given to the operational systems (such as SAP, Strategen, etc.) and the data they process. The fundamental requirements of operational systems are different from reporting systems and data warehouses; operational systems are based on transactional accountability, whereas reporting systems must support a high degree of query performance and flexibility.

The data warehouse served as a much needed platform to combine the data not only from legacy systems, but also new applications. The data warehouse provided a general reporting platform to serve as a foundation for the consolidation of enterprise data needed to generate reports. This foundation also supported the addition of future reporting requirements and additional data sources, and it standardized technologies and repeatable processes. This provided our client with a long-term flexible, scalable and maintainable reporting solution and decision support system.

In summary, CTA provided the client with a robust and flexible web-based reporting solution using the Business Objects – WebIntelligence (BOWI) platform and application. By implementing BOWI as their corporate reporting tool, the client can provide its users wide accessibility to corporate data, under a flexible and scalable architecture, and unify the source of "true" data within the organization.

CTA delivered a project plan that enabled the client to implement BOWI in a low risk incremental manner that also provided a flexible path for expansion. Our recommendation to implement a data warehouse using standardized integration templates and a flexible data model has provided the client with a stable platform for developing and deploying future enterprise-level reports. The data warehouse has enabled the client to leverage rapidly changing technology, embrace best-of-breed products, and integrate legacy systems into a powerful information engine delivering timely and relevant reporting capabilities for their customers.