

The Oracle Enterprise Data Warehouse (EDW)

Daniel Tkach

Introduction: Data Warehousing Today

In today's information era, the volume of data in an enterprise grows rapidly. The decreasing costs of processing power and online storage, and the increasing implementation of business applications are key factors that foster this growth. Stored mostly online, the data derived from the company's business transactions is readily available not only for accounting purposes, but also to be leveraged for obtaining competitive advantages.

The problem, however, is that data gathered from different enterprise sources is often inconsistent, redundant, and hard to use effectively for decision-making purposes. In many cases, the business information that can be derived from the wealth of available data is limited by the capacity of the IT organization. The IT staff faces an ever-growing backlog of requests from business users for the implementation of applications and reports. The deployment of these applications could become increasingly complex and time consuming because the data for these applications is dispersed throughout the enterprise, stored in many different formats, and may even reside on many different platforms. In addition, the constant changes in the business environment foster endless business users requests for new information.

The data of the company's daily transactions is stored in operational databases. These databases frequently hold duplicate information. Many companies, for instance, have separate information systems for the sale and support of different product lines. Insurance companies, for example, may have separate systems for life and car insurance; a customer that has both types of insurances is identified by a separate customer record in each system.

Some organizations try to provide business users with direct access to operational databases. This approach is not a good solution because the operational infrastructure components have been designed to support the load put on them by the daily operations. Additional demands caused by decision support queries run the risk of interfering with the systems' mission-critical mainstream functionality. In addition, unscheduled access to data by a large number of users can introduce security concerns, and data that is constantly changing is difficult to use for analysis. Making simple copies of data for query purposes is not a good answer

either because usually data is not correlated across independent application systems, making it difficult to be used for understanding all the important relationships in a business.

Another problem is derived from the fact that part of the operational data may be missing or erroneous, making it inadequate for decision support. For instance, errors in customer zip codes may result in wrong conclusions about the consumer habits of a certain population.

Business Intelligence and Data Warehousing

Business Intelligence is a discipline that enables companies to extract comprehensible and actionable knowledge from their data and use it to make crucial business decisions. The implementation of effective Business Intelligence systems includes building “data warehouses” and “data marts”. Data warehouses are databases populated with cleansed, reconciled and enhanced *historical* data, structured following logical business subject entities such as sales, products, and customers, for the purpose of improving decision-making. Databases that contain such consolidated data are often referred to as “subject-oriented”. Operational databases are normally designed around business processes.

The scope of a data warehouse is usually enterprise-wide, allowing business users to understand, for instance, the problems with company-wide purchases or the complete set of business relationships with a given customer. Data warehouses contain detailed data from business transactions. This low level of granularity allows combining data in many ways, providing a high degree of analytic flexibility.

Data marts are data warehouses that target a specific line of business or department such as sales, marketing, finance, manufacturing, or customer support. Data marts can be built independently or populated from a central enterprise data warehouse. Independent data marts usually contain both detailed and aggregated data. Data marts that are dependent on a data warehouse contain usually only aggregated or summarized data. The detailed data resides in the data warehouse.

Decision support requires the analysis of the state of the business over time. To populate a data warehouse, data is gathered from operational systems and a time stamp is added before loading it into the data warehouse. Thus, data warehouses provide a consolidated historical view of the business that can be used to understand the company’s problems and build predictive models. Most types of decision support require access to historical data – to identify trends and compare the operation of the business from month to month and year to year. Many types of operational systems do not contain historical data because they are designed to reflect the current state of the business. For example, inventory control systems indicate current stock levels and manufacturing information systems indicate the current state of the production systems. This kind of system does not need, in principle, to hold historical data to provide the desired results.

There is additional non-historical data that is used in decision support that is not stored in operational databases. For example, to analyze the sales by sales region, it is necessary to have a complete list of the company's sales offices as well as their grouping by regions. In a data warehouse or in a data mart, this kind of information can be acquired and maintained (in what are called 'dimension tables') for use by analytic tools.

Business intelligence systems focus not only on populating the data warehouse, but also on providing access to this business information to a wide audience of internal and external users. A new genre of tools is evolving which allow companies to determine and analyze product and customer profitability, perform simulation and strategic planning, and monitor enterprise-wide performance through an enterprise information portal.

Business intelligence systems are not stand-alone systems, but are integrated into the business process, and support closed-loop decision-making. This means that even though data warehouses are separate from operational systems, the analyses provided by data warehousing systems are used to make or modify operational decisions that will in turn create new transactions. The data involved in these transactions is processed and stored in the data warehouse, where it is used for new decision making. For instance, the analysis of the sales trends over a period can be a key factor in deciding how many items to produce. This loop can be closed with human intervention, or automatically through workflow.

Limitations of the Traditional Business Intelligence Approach

Traditional data warehouses do not provide by themselves a complete solution to the needs of business users. Although many data warehouse vendors provide products for building and accessing a data warehouse, these products can require a significant amount of implementation effort because they are often not packaged for specific industries or application areas. For these vendors much of the focus is still on building the data warehouse, rather than accessing it. Many organizations seem to think that if they build a data warehouse and provide users with the right query and reporting tools, the job is done. This misconception caused the failure of many data warehouse projects.

The decision support systems usually require analyzing data from heterogeneous applications that were implemented and maintained separately. Therefore, the data warehouse builders face two options: either start building a centralized data warehouse from scratch and then download line of business or departmental data to data marts, or start by building independent data marts and consolidate them later in a central warehouse.

The first option, building the corporate data warehouse, is a long-range endeavor, and it can take many years before that effort bears useful fruits. The second option yield usable results much faster, but with a restricted scope. In reality the consolidation of the data marts rarely - if ever - happens, and this precludes providing corporate-wide decision support.

Oracle's Complete and Simple Business Intelligence Solution

Oracle Corp. provides a complete and simple solution to providing effective decision support capabilities that gives companies considerable competitive advantages. The complete solution starts with the implementation of the Oracle E-Business Suite. This suite is integrated on the 11i platform, which means that the definitions of all the business objects such as “customer”, “invoice” or “employee” are common across the CRM, ERP, and Exchange applications. This provides an enormous advantage for data consolidation. In addition, Oracle Corp. provides a complete set of Business Intelligence products as part of the E-Business Suite, built on the Oracle Enterprise Data Warehouse (EDW) foundation. This integration makes it easy to deploy a complete decision support infrastructure in a short time.

The Oracle Enterprise Data Warehouse

To help organizations exploit the information hidden in their operational data and analyze it using the right tool set, Oracle provides the Enterprise Data Warehouse (EDW). The Oracle EDW is a packaged data warehouse solution with enterprise-wide data collection capabilities. EDW is an important architectural component of the E-Business Intelligence solution, a set of intelligence products for Financials, Projects, Manufacturing, Human Resources, Marketing, Sales, Customer, and Interaction Center applications. The E-Business Intelligence solution also includes strategic enterprise management applications such as the Oracle Balanced Scorecard and Activity Based Management as well as advanced analytic applications such as Sales Analyzer and Financial Analyzer.

EDW is a “ready-to-run” data warehouse that provides shorter time-to-benefit and reduces the implementation risks and costs usually associated with data warehousing projects. Furthermore, because EDW is packaged software, upgrades and synchronization with the source applications and tools are handled automatically with each new release.

EDW allows companies to pull in data from different sources into a single warehouse with an open schema, organize the data within the warehouse, administer the warehouse and extend it. EDW also allows companies to consolidate their multiple data warehouses or data marts. The end benefit is that all the relevant historical enterprise data is gathered together. This facilitates cross-functional enterprise-wide analysis, which is of utmost importance when analyzing top line growth, identifying operational issues and competing in a global environment. EDW is the underlying data source for:

- **Business Intelligence modules** - For both ERP and CRM, pre-built business intelligence offers a shorter time to benefit empowering the workforce to set objectives, manage exceptions, understand the facts, and make the best possible decisions.
- **Strategic Enterprise Management** - Activity Based Management information can be captured in EDW along with the operational information from the Intelligence

modules. Additionally, the Balanced Scorecard application can help companies implement and manage their strategy and link strategic objectives to key operational data from EDW.

- **Analysis, Planning, Simulation and Mining tools** - Using one or a combination of Oracle's analytic tools including Discoverer, Sales Analyzer, Financial Analyzer and Data Mining, users can perform sophisticated ad-hoc query, analysis, simulation and data mining as well as run advanced analytical applications focused on strategic analysis and decision-making that enhance long-term shareholder value.
- **Custom Applications** - Companies can use EDW as the underlying source to build analytical applications specific to their business needs.

EDW Highlights

EDW provides a robust and open architecture that allows flexible customization of the packaged software framework. It is based on an enterprise-wide repository that allows the data warehouse to evolve together with the company's business.

Because the data schema of EDW is open, it can serve as a warehouse to different ERP vendors and legacy data sources because:

- Many third-party Online Analytical Processing (OLAP) tools can access the data stored in EDW through the Common Warehouse Metadata interface.
- EDW's flexible data warehouse schema can support different business models, because it allows schema configuration without having to regenerate the warehouse.
- New versions and upgrades to EDW do not require regeneration of the warehouse.
- Data cleansing, data alignment, and error checking procedures are integrated into EDW to ensure data integrity in the warehouse.
- EDW can either be installed on the same server as the Oracle E-Business Suite or on another machine.
- EDW is deployed on Oracle8i, the industry-leading platform for data warehousing.

EDW and the Oracle E-Business Suite

The transactional data created by the applications of the Oracle E-Business Suite can be loaded into EDW without any coding effort. EDW shares many common administration elements of the E-Business Suite applications infrastructure in order to minimize the management effort. EDW includes pre-built data collection programs - programs that extract data from transaction systems for Oracle Applications Release 10.7, 11.0 and 11i. EDW collection packs are available to support different entities across many functional areas. In addition:

- EDW collection packs understand Oracle Application flexfield structures and translate them into data warehouse dimensions.

- EDW uses a variety of standard Oracle Applications features to provide installation, upgrade, administration, management, and security of the warehouse.
- EDW collection pack releases will be synchronized with future releases of Oracle Applications, thereby providing the ability to analyze data from newer releases of Oracle E-Business Suite applications as they become available.

EDW Infrastructure for Enterprise-Wide Reporting and Analysis

EDW provides a single warehouse infrastructure supporting reporting and analysis requirements of different classes of users - executives, line managers and business analysts. EDW is a common, open repository that is the foundation for Oracle's strategic and operational analytical product areas, as well as for custom analytical applications built with Oracle Express or other third party tools.

Oracle Business Intelligence System (BIS) products are available for all major applications of the Oracle E-Business Suite, such as Purchasing, Financials, Projects, Operations, Process Manufacturing, Human Resources, Marketing, Sales, Customer and Interaction Center. These products support day-to-day performance management by providing real-time performance information and exception notifications through Oracle Workflow.

The Oracle BIS products enable users to address proactively business performance issues and opportunities. By focusing both on the internal and external facing applications, these Business Intelligence products are designed not only to help companies run more efficiently but also to enable their growth by providing key information that the companies can use to gain and retain profitable customers. These products include:

- Strategic Enterprise Management (SEM) products. These are analytical applications that support advanced management practices and methodologies such as Activity Based Management and Oracle Balanced Scorecard. These products help executives make decisions related to the strategic direction and goals of their organizations, align strategy with operations, and monitor the execution of the strategy with an enterprise wide scope.
- Custom analytical applications such as sales analysis and forecasting, or other applications that requires sophisticated modeling or what-if analysis, can leverage the power of Oracle Express to rapidly build and deploy powerful e-business analytical applications to be deployed over the web. These applications can access EDW directly, thus enabling the IT organization to focus on the end-user requirements because the data extraction, cleansing, and transformations are already performed when the data is loaded into EDW. This approach is of tremendous value to customers who wish to develop their own analytical applications but have been forced to spend much of the time, effort, and money in designing and building the data warehouse.

EDW Architecture

EDW supports the organization of data according to multi-dimensional models. These models allow analyzing the business data from multiple perspectives. From a conceptual standpoint, the business data downloaded to EDW is organized – that is, stored and indexed– as a multi-dimensional “cube”. Each edge of this cube represents a dimension – a data identifier, variable, or attribute– such as “time”, “product” or “channel”. Dimensions represent a mechanism for slicing the data in the cube, as shown in Fig.1.

The power of the multi-dimensional model resides in the analytical capabilities that it provides, such as pivoting (rearranging dimensions), aggregations (sum, count), comparisons (ratios, shares) and sequencing (percentiles, top/bottom). The ability to perform sophisticated analysis such as modeling, simulation, and time-series comparisons are natural extensions of multi-dimensional modeling. Finally, this model provides a structured paradigm for end-users to navigate and analyze business data.

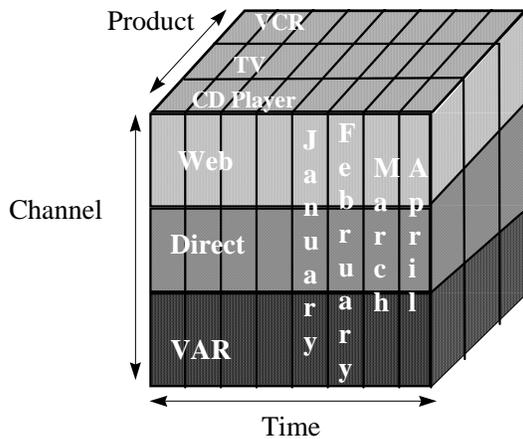


Figure 1: Multi-Dimensional Analysis

Translating this conceptual model into physical design, the cube containing the business data is a physical entity in the warehouse that can be either a Fact Table or a Summary Table. A fact table stores data at a granular (detailed) level; a summary table stores aggregated data. The data in the fact or the summary table can be viewed across different dimensions.

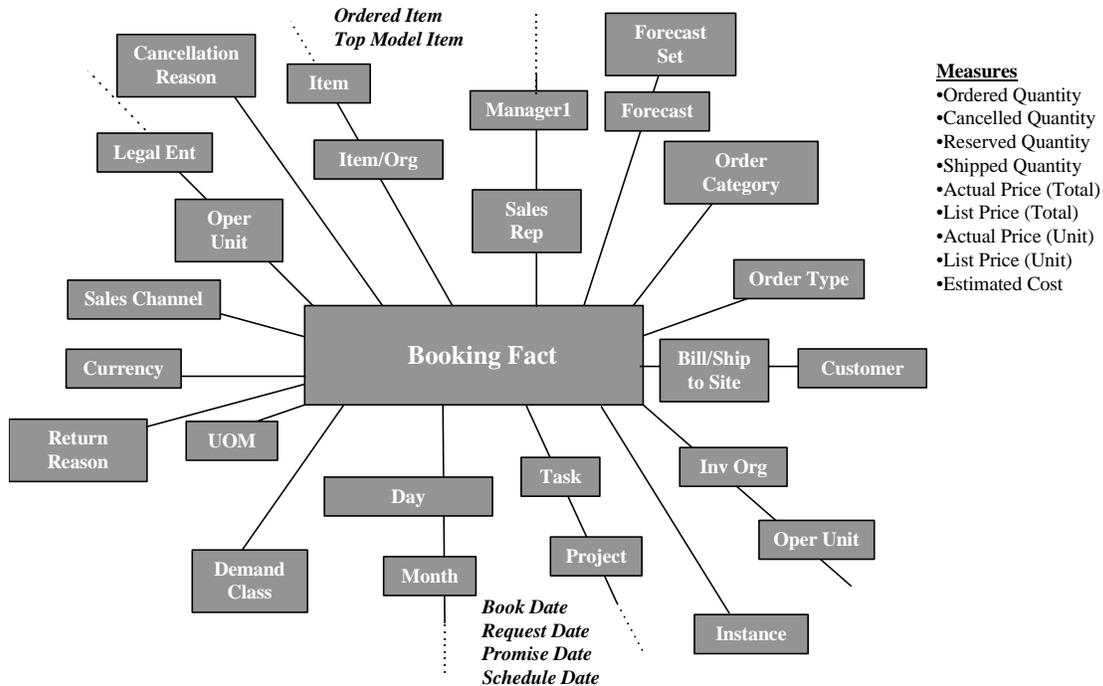


Figure 2: Sales Order Bookings Fact Table

EDW has dozens of prepackaged fact tables and summary tables, which can be analyzed by different dimensions. Figure 2 shows the Sales Order Bookings Fact table, the key measures (data elements) it contains and the dimensions they can be analyzed by.

EDW Architectural Components

EDW is built using the Oracle Warehouse Builder (OWB). OWB is an integrated system for building a complete end-to-end data warehouse solution. OWB provides EDW with facilities to design, map data sources, and generate the code to create all the fact tables and dimensions. EDW metadata is stored in the OWB repository. EDW has three main architectural components:

- **Runtime Warehouse**
This is the system that houses the data warehouse and it contains the fact tables, summary tables and dimensions. Data from multiple sources is cleansed, aligned and brought into this system. The customers may choose to install physically the runtime warehouse in the same database instance as the Oracle E-Business Suite applications, or may create a separate dedicated database instance.
- **OWB Repository:**
This repository contains the metadata needed for generating the runtime warehouse. It has knowledge about the various environments and sources from which the data is

extracted, the definition of the runtime warehouse, and how the fact and dimension tables should be populated.

- **Source System Components:**

EDW contains pre-built collection packs that extract data from the Oracle E-Business Suite transactional systems and populate the fact and dimension tables. The collection packs can extract data from Oracle Applications release 10.7, 11.0 and E-Business Suite 11i, as shown in Fig.3. A set of related product offerings called Oracle Integrator for SAP, PeopleSoft, etc. facilitate multi-source integration.

Data Visualization

Each BIS Intelligence application that accesses EDW includes pre-built reports and workbooks. Customers can extend and change these reports, and also create new reporting and analytical applications that use the data stored in EDW. Other data visualization components include Oracle Portal and the Oracle Balanced Scorecard.

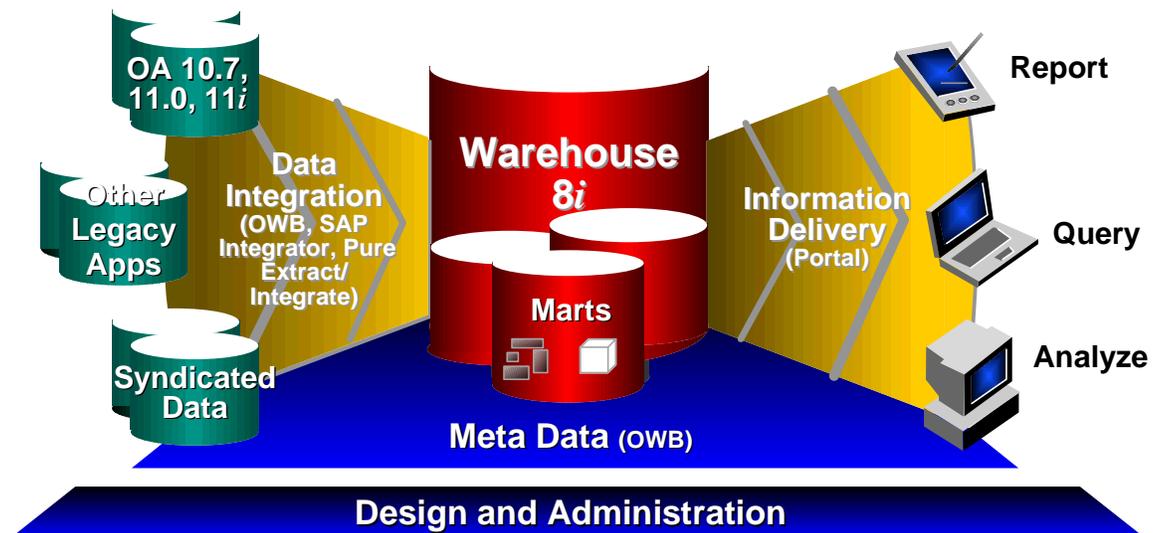


Figure 3: EDW Architecture.

EDW Data Flow

In order to move data from the transactional system to the data warehouse and to ensure that it is suitable for analysis, a three-step process is required - Extraction, Transformation and Transportation (ETT). The extraction process involves pulling the right operational data from the transactional tables. Some of the operational data might not be suitable for business analysis the way it was gathered, therefore business rules are applied to the data during the transformation process to prepare it

for analysis. The extracted and transformed data is then moved to the warehouse in the transportation process.

EDW uses a push-pull strategy to move data between the source systems and the runtime warehouse. The collection packs that reside on the source systems periodically push data into the staging tables in the runtime warehouse. From these staging tables, the runtime warehouse pulls data to load the fact and dimension tables. The push-pull strategy allows for maximum flexibility when consolidating multiple data sources, because:

- Multiple source systems can push data to the staging tables
- Different source systems can push data at different times

The flow of data in EDW is shown in Fig.4. The collection programs extract data from Oracle E-Business Suite applications using database views. The loader engine in turn is responsible for resolving foreign keys, data alignment, and data integrity such as duplicate rows entering the systems. The collection programs perform both the extractions and the initial transformations. The data is transported between source and target databases using database links. The data in the fact tables can then be summarized across different sets of dimension levels using Materialized Views, an important high performance database capability of Oracle8i and 9i.

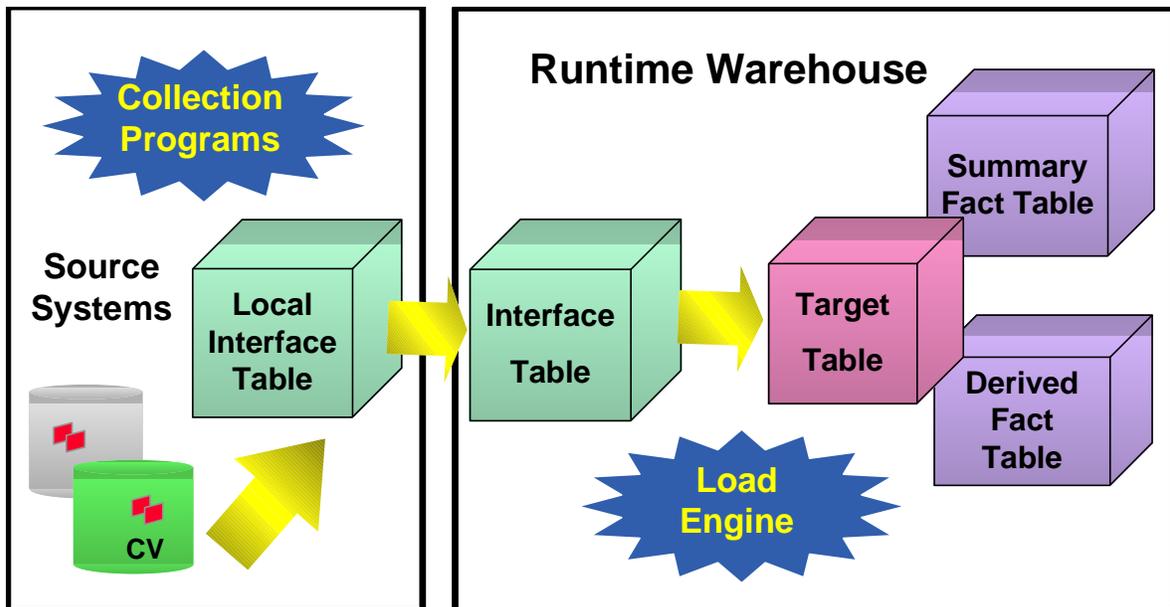


Figure 4: Data Flow

EDW Metadata

Data warehousing tools support the design of the warehouse as well as the extraction, cleansing and movement of data, together with analysis and business intelligence reporting. These tools use metadata — information that describes the characteristics and origin of data — to drive the various steps of the data warehousing process. These tools require the definition of a common information model (a standard representation), and a repository or catalog with supporting services for metadata.

EDW has been designed with a very flexible and extensible architecture defined by metadata that is stored in the Oracle Warehouse Builder repository using the Metadata Loader. Metadata is used by EDW to describe the collection process, the resulting fact and dimension tables, the rules for aggregation, and the hierarchies. This means that business analysts and data warehouse administrators can manage the structure of the warehouse through metadata definitions using the functionality provided by the Oracle Warehouse Builder.

Oracle is a key contributor to the Common Warehouse Metamodel (CWM), an industry standard for metadata interchange, developed through the Object Management Group (OMG), and created for data warehousing. The CWM is designed for extensibility to support rapidly evolving metadata requirements, and in conjunction with supporting tools developed by Oracle, it addresses the need of interoperability with other vendor solutions based on the CWM standard.

EDW Common dimensions

EDW provides a comprehensive native set of dimensions, called “common dimensions” which span functional areas. For example, common dimensions include time, organization, geography, item, trading partner, and many more. A dimension may have multiple levels organized into a hierarchy to facilitate rollups and drill-downs. EDW provides several dimensions with more than one simultaneous hierarchy to accommodate multiple functional perspectives. As an example, Fig. 5 shows four different hierarchies for the Time dimension.

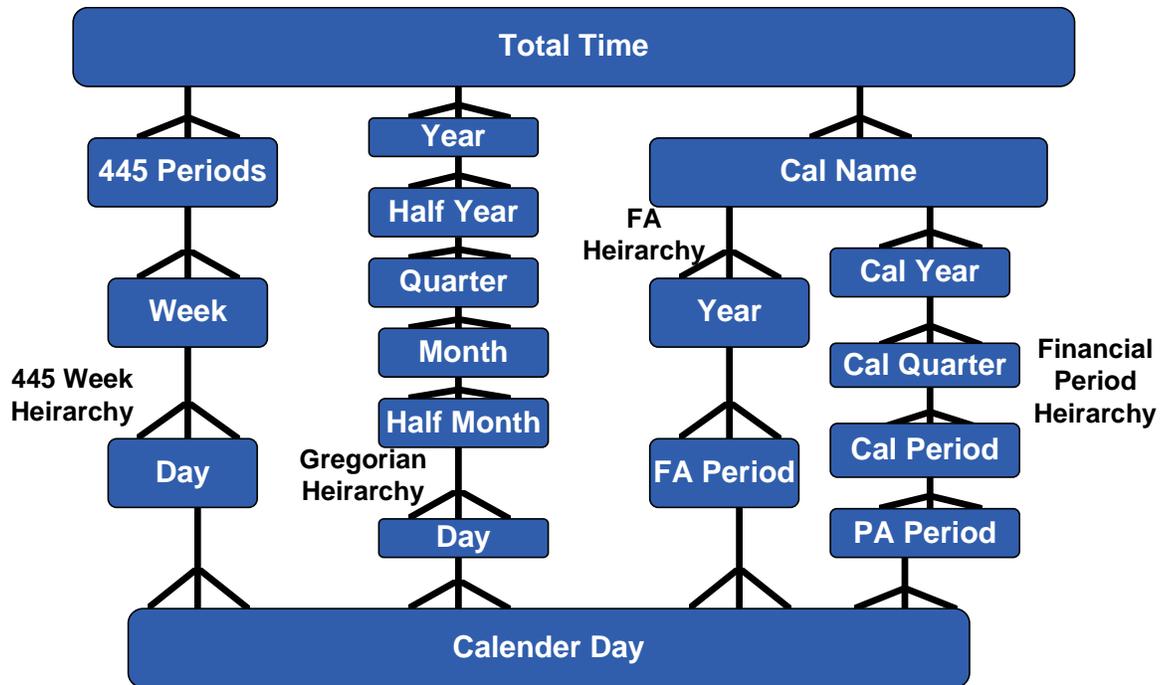


Figure 5: The Time Dimension with three different hierarchies

These dimensions are of great value during cross-functional analysis, as data from different sources can be viewed along a common dimension. For example, we can investigate seemingly unrelated items across the time dimension such as backlog, order volume, promotion analysis, inventory position, supply chain velocity and sales force competency to understand if there is a relationship among these factors that we were not aware of, and that can help us understand why revenue targets have been missed.

EDW Integration with the Oracle E-Business Suite

EDW defines several points of integration with the Oracle E-Business Suite. It includes built-in procedures, called collection programs that extract data from the E-Business Suite applications. In addition, EDW shares many common administration modules with the applications to minimize the management effort required to keep the data warehouse up to date when the organization changes.

Collection Programs

One of the most salient features of EDW are the pre-built collection programs. These programs, designed by the functional area experts, pull data from the various transactional tables and transform it into data structures optimized for multi-

dimensional query access. As newer releases of the Oracle E-Business Suite become available, the collection programs are also upgraded so that both the transaction system and the analysis system are kept in synchronization. Each release's collection can co-exist with prior collections, so that multiple releases of the applications of the Oracle E-Business Suite can feed the common data warehouse, allowing for non-disruptive growth of the application base.

Integration of Multiple Currencies

EDW provides functions for easy handling of multiple currencies that may appear in the transaction systems of global enterprises. EDW captures the transaction currency and the Set of Books currency from the transaction systems. In addition, EDW converts the transaction and Set of Books currency to a global warehouse currency thereby allowing global financial analysis at both the detail and summary level.

Flexfield Support

The Oracle E-Business Suite applications include a user-defined flexible data structure called "flexfields" to allow customization without programming. EDW is sensitive to the installation specific setups including the flexfields defined in the E-Business Suite. Each flexfield attribute can be mapped into EDW as either a dimension or a measure.

Administration and Management

The EDW collection programs are run periodically to extract data from the source systems and stage it in the warehouse interface tables. Because EDW is implemented using the Oracle E-Business Suite framework, these programs are scheduled using the concurrent manager of the E-Business Suite applications. Loading the runtime warehouse from the interface tables is also scheduled using the concurrent manager, and monitored through the concurrent request screens.

Oracle E-Business Suite Compliant Install

EDW is delivered through the normal E-Business Suite installation mechanism. This allows users to easily choose where they would like to locate the physical data warehouse – coexisting with the same instance of the applications or on a separate instance. An important benefit of this approach is that the data warehouse follows the standard E-Business Suite applications management framework. This allows for easily applying fixes and upgrades when they become available.

Security

EDW uses the Oracle E-Business Suite Security model to provide both warehouse administration functions and data access functions. Users may have one or more responsibilities, each of which has security attributes. For ETT functions, a user has privileges to collect and load a certain set of data, depending on the assigned responsibilities. For OLAP access, users may only access authorized data. EDW provides fine grain access control that can restrict the access of a user to specific data in the warehouse.

Conclusions

Business intelligence tools and techniques produce decision support information that can provide a sustainable competitive advantage for companies operating in the e-Business environment. But no matter how good the tools and techniques are, their output is worthless if they are not fed by quality data. Business intelligence systems are ruled by the “garbage in – garbage out” principle: it is not possible to obtain meaningful insights from data that is inconsistent, incomplete, duplicated, and hard to integrate.

Data warehouses are repositories of clean, consistent data that provide an excellent base for business intelligence. The problem of building a data warehouse, however, is far from trivial, when the architects and data analysts have to start from scratch defining the appropriate data structures, maintaining them over time, and populating the warehouse with data from diverse sources.

The Oracle EDW provides users with a data warehouse infrastructure that is ready to be implemented and used in a very short time. EDW is fully integrated with the applications of the Oracle E-Business Suite; therefore, it leverages the customers’ investment in the implementation of the applications by providing the users with a complete and simple business intelligence environment. This environment helps enterprises to achieve significant cost reduction, a better understanding of the market and the competition, and to make major improvements in customer retention and satisfaction. Thus, EDW enables enterprises to compete fast and effectively in the demanding environment of the new economy.