



IBM Cognos 8 Data Manager:

Data integration for
the last mile to BI

Contents

- 4 Business Problems**
 BI: Enabling timely response to changing business requirements
 IBM Cognos 8 Data Manager
 Integrating BI into customer environments
- 8 Business Drivers**
 Developing a data warehouse solution
- 22 The Solution**
 Information model delivery: load
 Managing the environment: deployment
 Collaborative projects
 Increased seed to BI deployment
 The benefits of using IBM Cognos 8 Data Manager
 IBM Cognos 8 Data Manager goes beyond integration
- 29 Conclusion**
 Summary

Abstract

Businesses today have a growing number of systems, including legacy ERP, transactional systems, warehouses (relational and OLAP), as well as external systems from suppliers, customers, and partners. For companies to gain a maximum return on their data investment, they must remove the barriers to implementing an integrated architecture, and introduce a framework that supports interoperability among heterogeneous data environments.

As part of this mandate, IT departments must also reduce overhead on operational systems from reporting and analysis needs, consolidate data from disparate data sources, provide access to consistent information across the enterprise, preserve data history, and present data in an understandable way to business users.

Overview

Business intelligence (BI) and performance management (PM) capabilities such as reporting and analysis, scorecarding, event lifecycle management, planning, and budgeting require an integrated architecture to ensure that all data assets are being brought to users in a consistent manner. IBM Cognos® 8 Data Manager creates this data foundation for effective, fact-based business decisions.

IBM Cognos 8 Data Manager provides a seamless data integration environment for business intelligence. The IBM Cognos 8 platform, with its open data access strategy, provides access to any data source including:

- All major relational databases including IBM, Oracle, SQL Server, Teradata, Sybase, and ODBC.
- Widely deployed ERP systems, including SAP R/3 and SAP BW, PeopleSoft, and Siebel.
- Widely used OLAP and dimensionally-aware data sources, including Microsoft® SQL Server® Analysis Services (SSAS), IBM® DB2® UDB, OLAP Server, and Hyperion Essbase.
- Modern data sources, such as XML, Java® Database Connectivity (JDBC), Lightweight Directory Access Protocol (LDAP), and Web Service Definition Language (WSDL).
- Satellite sources, including Microsoft® Excel® files, Microsoft® Access® files, flat files.
- Mainframe sources, including IBM® Virtual storage access method (VSAM), IP Multimedia Subsystem (IMS), IDMS (Integrated Database Management System), COBOL copybooks.
- Content management data, including IBM® FileNet®, EMC Documentum, and OpenSoft.

As an integrated component of IBM® Cognos® 8 BI, IBM Cognos 8 Data Manager provides the capability to extract, transform, and load data into dimensional warehouses ready for IBM Cognos 8 BI and IBM® Cognos® 8 Planning.

In this paper we will discuss the challenges associated with integrating data, the benefits of the IBM Cognos approach, and the process and key factors associated with developing a conformed, dimensional data warehouse solution.

Business Problems

BI: Enabling timely response to changing business requirements

Direct reporting and analysis from existing business systems does not always provide reporting consumers with the functionality and performance that they require. This leaves the business with an inconsistent view of the data across the organization.

Organizations are looking for solutions that will enable them to:

- Integrate data from multiple sources, including transaction systems, warehouses (relational and OLAP), flat, legacy, modern sources, and third-party sources to deliver a consistent enterprise wide view of data.
- Structure the data so that it makes business sense, using dimensions that have a common meaning throughout the organization.
- Perform ad hoc queries against the data without affecting the performance of the underlying applications.
- Scale the solution from a single department to the entire enterprise without compromising data quality, integrity, and consistency.
- Deliver time series analysis over historical data for trend analysis.

Directly reporting against transaction systems does not provide this range of capabilities. Instead, organizations must develop a data warehouse with conformed data to integrate their systems, in order to effectively report and analyze data that drives their business.

IBM Cognos 8 Data Manager

IBM Cognos 8 Data Manager creates data warehouses and data repositories for reporting, analysis, and performance management. Through data extraction, transformation, loading, and dimensional management, IBM Cognos 8 Data Manager delivers dimensional data warehouses that are ready for IBM Cognos 8 BI and IBM Cognos 8 Planning. With the advanced data merging, aggregation, and

transformation capabilities of IBM Cognos 8 Data Manager, data can be merged from multiple sources including traditional legacy files, purchased data (such as mailing lists), and ERP data sources. IBM Cognos 8 Data Manager can transform and deliver data into any design of database schema; however it is best suited to delivering data into dimensional data warehouses and conformed data marts. The IBM Cognos 8 Data Manager dimensional framework allows shared dimensions to be constructed, which can then be shared across all related subject areas. The dimensional framework ensures that individual subject areas can be designed incrementally, but integrated consistently with any existing subject areas within the data warehouse.

IBM Cognos 8 Data Manager is comprised of two main components: a Microsoft® Windows®-based design environment and a multi-platform, server-based engine. The intuitive graphical interface of the design environment makes defining and implementing transformation processes easy. The high-performance transformation engine handles large volumes of data quickly and provides support for all major relational databases.

IBM Cognos 8 Data Manager is integrated with the IBM Cognos 8 platform, and delivers metadata to Framework Manager, allowing the target data warehouse and data repositories to be modeled and used in IBM Cognos 8 BI and performance management solutions. The publishing of IBM Cognos 8 Data Manager tasks into the IBM Cognos 8 BI environment allows data movement tasks, and the delivery and updating of the data warehouse, to be directly managed and integrated into the BI solution.

Integrating BI into customer environments

The success of a business intelligence implementation is measured by how well business users understand and access the data. Information has to be presented in a consistent, intuitive way and it has to be simple to understand.

To meet these requirements, the business needs a data warehouse that:

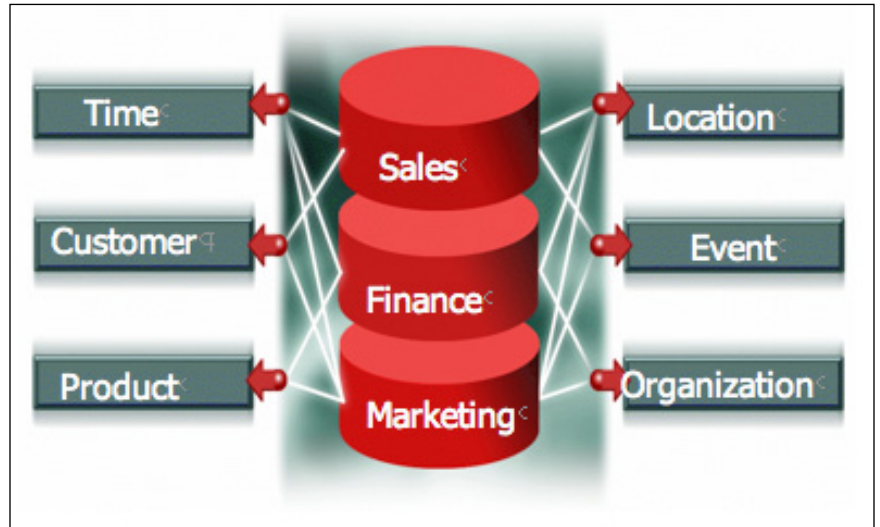
- Makes enterprise information easily accessible and understandable.
- Presents data in a consistent format.
- Is adaptive and resilient to business change.
- Is the foundation for improved decision-making.
- Will be adopted by the business community.
- Is scalable from departmental to enterprise implementations.

There are several components of a data warehouse, including the operational source systems, data staging area, data presentation area, and data access tools. In order to be successful in implementing a data warehouse, you need to understand what these components are, how each of them is used, and how they are to be used together.

The market has reacted strongly to the difficulty of planning and implementing a single, undifferentiated data warehouse for the whole enterprise. This job is overwhelming for most companies to deliver. The future of data warehousing is to deliver data marts that focus on achieving business objectives for departments, allowing a phased-in and prioritized development that will delivery early return on investment.

These data marts are organized in a dimensional framework around conformed dimensions and facts to avoid becoming information silos. This approach has been endorsed by such industry experts as Ralph Kimball, principal of Ralph Kimball Associates, North America's leading provider of dimensional data warehouse design consulting.¹

Dimensional modeling should be used for the presentation phase of the data warehouse and is the only viable technique for delivering data to end users in a data warehouse. The dimensional model is a predictable, standard framework. It packages data in a symmetric format that companies can use, assimilate, and query with high performance. The twin, non-negotiable requirements of a data warehouse are ease in understanding and performance.



Conformed dimensions are the key to delivering a consistent view of data across an organization. For example, the customer dimension should mean exactly the same to a sales manager as it does to an inventory manager or director in finance. Without conforming dimensions, departments, and functional areas will develop different definitions.

Business Drivers

The benefits of developing a data warehouse that has conformed dimensions and facts include:

- Conformed dimensions are developed once and shared among multiple data marts. This saves development resources and reduces the processing required to update data because each mart is not running separately against the source system.
- Each step becomes the building block for the next step. As the solution grows, the effort required for design, modeling, and implementation decreases.
- Solutions built for one department can be reused by other departments.

The benefits to business users include:

- Users can easily understand the data because it is intuitive and presented in a business context.
- The use of conformed dimensions ensures that everyone throughout the company uses the same definitions and shares a single version of the business issues. Decisions across the enterprise are based on consistent information that reflects the entire business – not just individual functions.
- Business-wide confidence in the data due to conformed views.
- As an organization's business and user needs change over time, this approach provides the flexibility to evolve with those changes with minimum disruption to business users.

Developing a data warehouse solution

The following sections outline the process and key factors in developing a data warehouse solution with coordinated data marts.

Data staging area

Extracting source data is the first step. This requires analysis and understanding of source systems and moving the data to the staging area for manipulation. It may also involve some merging of data. The data staging area of the data warehouse is where data is merged, cleansed, and transformed. It is everything in between the source system and the presentation layer. IBM Cognos 8 Data Manager performs the necessary extraction, merging, cleaning, and transformation of data for this phase in developing the data warehouse.

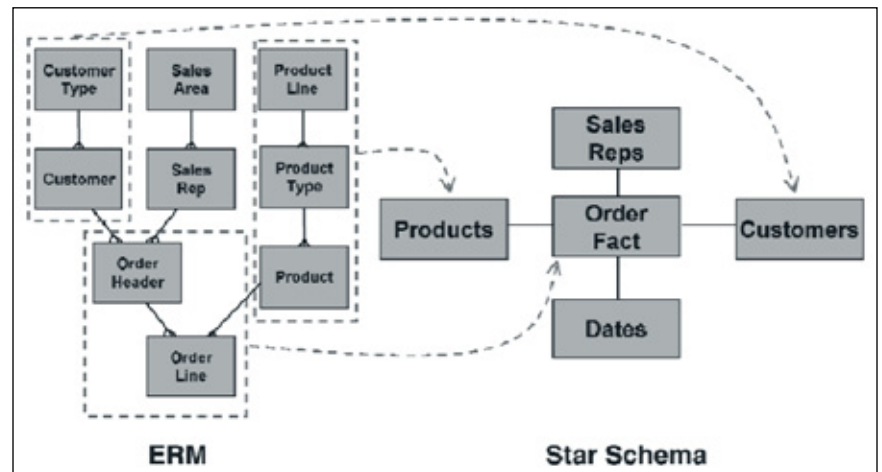
Data presentation

The data presentation area is where the data is organized, stored, and made available for analysis, reporting, and direct querying by report writers and other business users. As far as the business user is concerned, the presentation area is the data warehouse. It is built as a series of conformed data marts. IBM Cognos 8 Data Manager delivers the data in a dimensional schema following industry best practices. The data may be presented, stored, and accessed in dimensional schemas. According to industry best practices, this is the only viable technique for delivering data to data warehouse users.

Dimensional modeling always uses the concepts of facts, also known as measures, and dimensions. Facts are typically numeric values that can be aggregated, and dimensions are groups of hierarchies and descriptors that define the facts. Dimensional models are built by subject area, such as store sales, inventory, claims, etc. Because the different subject areas share some, but not all, dimensions, efficiency in design and operation, as well as consistency, is achieved by conforming the dimensions.

A dimensional model contains two types of tables – fact tables and dimension tables. A fact table is the primary table in each dimensional model. It contains numerical measurements of the business and is normally additive. Fact tables also include a foreign key for each dimension they represent. A dimension table is one member of a set of companion tables to a fact table. Dimension tables contain textual descriptors of the business and provide context to the fact table measures. Each dimension table has a primary key that joins to the fact table.

The numerical measurements of a fact table are joined to a set of dimension tables filled with descriptive elements, often referred to as star schema architecture. A star schema is the simplest style of data warehouse schema, consisting of a few fact tables referencing any number of dimension tables. The facts that the data warehouse helps to analyze are classified along different dimensions. In other words, the fact tables hold the main data, while the usually smaller dimension tables describe each value of a dimension and can be joined to fact tables as needed.



Star schemas facilitate delivery of reporting requirements by making it easier for business users to understand and navigate data.

The dimensional framework

IBM Cognos 8 Data Manager operates with a unique and integral dimensional framework, allowing the design of a series of conformed data marts. The general rules for dimensions can be established once and then shared between data marts within the warehouse and downstream BI tools. The dimensional framework typically defines a number of objects that are organized in the way that business users think (for example, by time, location, product, customer, or sales). These objects describe metadata. Metadata is the structure of the business data. This structure facilitates the navigation and aggregation paths that can be consumed and utilized by the business user community. The framework delivered by IBM Cognos 8 Data Manager is easily extendible, which allows new dimensions to be added as additional data marts are implemented, to satisfy the ever moving business needs.

Referential integrity and validation

Within the IBM Cognos 8 Data Manager dimensional framework, there is a reference explorer that lets the data analyst/developer view and validate hierarchies within dimensions. The reference explorer has a graphical interface so they can quickly see the structure of the hierarchy for a given dimension.

This explorer will also report any discrepancies found in parent/child data relationships, and detect issues around multiple parents, and foster children. This allows issues to be resolved before loading data into the warehouse. This IBM Cognos 8 Data Manager capability ensures the integrity and accuracy of the dimensional data delivered to the warehouse.

As it pertains to a warehouse, referential integrity means that for each foreign key in the fact table an entry exists in the corresponding dimension table. IBM Cognos 8 Data Manager automatically checks for this during the processing of the fact table and lets the system capture any transactional records coming through that do not have an associated dimension row. In this way, IBM Cognos 8 Data Manager ensures that companies build accurate and consistent dimensional information and maintain fact table integrity. This is absolutely critical for BI.

Conformed dimensions

Stovepipe data marts are the bane of any data warehouse project as they fail to offer all-important enterprise-wide views. By using a dimensional framework, companies can implement the data warehouse bus architecture by building conformed dimensions. Based on consistent terms and shared data, conformed dimensions ensure that users in all departments or functional areas approach business issues with the same context.

In a coordinated organization, the customer dimension means exactly the same thing to a sales manager as it does to an inventory warehouse manager or to the vice president of finance. In the absence of conformed dimensions, departments and functional areas tend to develop different definitions, hierarchies, terms, and dimensions for many of the same business measures. This frustrates productivity, hampers decision-making, and produces inconsistencies in reports.

Conformed data marts share dimensions so that decision-making is based on coordinated information. Building conformed dimensions means:

- Less development time as dimensions are built only once.
- Decreased redundancy because data is stored once.
- Shorter time to update because data is loaded once.
- Consistency and accuracy of data.

Multiple star schemas leverage the conformed dimensions to reduce update time and resources. Updates occur once, not multiple times, which accelerates the update process. In addition, conformed dimensions save disk space, reduce redundancy, and ensure that data is consistent from one mart to the next.

A conformed dimensional framework is one of the cornerstones of the IBM Cognos BI solution. It enables BI capabilities to be added in an iterative fashion. As each new data mart is added, report developers simply incorporate any dimensions that the new mart shares with existing ones, saving development time and resources.

Dimensionality and surrogates

An operational system usually contains information about the current state of business for an organization and maintains little or no history. By contrast, information warehouses are expected to hold perhaps five to 10 years of data. This system allows decision-makers to evaluate the performance of the organization from year to year. Evaluating this data is made more difficult with the variations over time, for example, a sales representative who changes divisions. One of the main reasons for warehousing is so you can perform analysis over time. This is not very easy to do directly from an operational system which is one of the reasons why an organization would want a data warehouse.

Business keys are common to operational systems, for example, invoice number, order number, employee number. They assign uniqueness to data and ensure that related data can be tied together within an operational system. All data warehouse keys must have surrogate keys; companies must not use original production keys. Surrogate keys are artificial, numeric keys that replace the business keys, and they are assigned sequentially as needed to populate a dimension table. Organizations are more than likely pulling data from multiple sources and the surrogate key becomes a unique identifier to a particular current record or historical record

For example, the first customer record is assigned a customer surrogate key with a value of 1; the next customer record is assigned 2, and so on. These keys are then used to join fact tables to the dimension tables. The surrogate keys, smaller in size than business keys and typically an integer insulate the warehouse from operational production system changes; accommodate the merging of data from disparate systems, and assist in the tracking of changes to dimensional information over time. In addition, surrogates enable the technique of slowly changing dimensions (SCDs) that are used to reflect changes in dimensions over time and manage historical data.

IBM Cognos 8 Data Manager shifts the complex process of managing surrogates and SCDs from the data warehouse developer to the ETL tool. Templates automate the maintenance of surrogates and slowly changing dimensions. The members of each dimension have attributes that identify each member and provide other information, for example, product dimension has product, name, color, weight, and size.

A simple dimension table typically has attributes such as a key, a surrogate, and a caption. IBM Cognos 8 Data Manager automatically manages the surrogates in the dimension tables and will also automatically replace the operational source key with the correct surrogate key while processing the fact table. If a company is tracking changes over time, attributes for the dimension and fact table can be set to determine automatically the next surrogate value to be used and places the surrogate key in the fact table.

Surrogate keys are an essential part of a warehouse. IBM Cognos 8 Data Manager automates the creation and management of surrogates for the dimension tables and fact tables.

Automating slowly changing dimensions

A salesperson in a particular company transfers to a different region mid-year. A data mart allows the company to record the move, and reflects the change in the database. Without record of this slowly changing dimension, a year-end revenue summary by region would allocate their entire year's sales to the new regional manager, overstating the new manager's accomplishments and understating the previous manager's performance. Companies that make decisions based on this type of misleading information can end up making incorrect assumptions that can result in costly mistakes.

IBM Cognos 8 Data Manager allows the salesperson to record this move and reflect the change in the database. With slowly changing dimensions, the revenue that the salesperson generated before departure will properly accrue to the previous regional sales manager, and the revenue that is generated after the move will be credited to the new manager.

Over time, certain dimensions – employees, products, and customers – will change, and IBM Cognos 8 Data Manager software is flexible enough to accommodate these changes and produce an accurate view of business performance by creating another dimension record.

Slowly changing dimensions and history preservation

Dimension table attributes change over time and part of the design phase is to identify the appropriate change handling strategy with the business user. Will it be important to track the change and, from the warehouse, analyze the impact of that change? Using surrogate keys, tracking such changes in the warehouse is made possible. IBM Cognos 8 Data Manager automates the tracking of these changes through slowly changing dimensions.

If a company needs to track an attribute, they can set a check box on the dimension table properties; automatically each attribute that needs to be tracked is also checked. Once this is done, whenever a record from the source system comes through with an update and the attribute has changed and tracking has been set, a new dimension record will be created with a new surrogate key. In addition, the previous dimension record will have its current indicator set to 'N' and an end date added. With this type of change, the fact table is untouched. This processing is completely automated with IBM Cognos 8 Data Manager.

In the following example a sales rep has changed regions. To effectively track the change, a second dimension row is created in the warehouse and additional columns are updated to record which dimension row is the latest or current one.

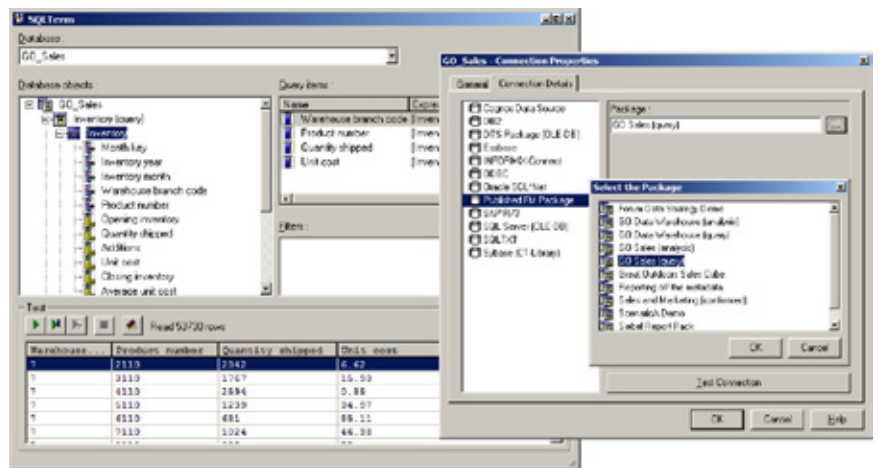
Example

Surrogate key	Sales Rep Number	Name	Region	Start Date	End Date	Curr Indicator
1	AXY12	John Harris	Chicago	050131	050626	N
2	AXY12	John Harris	Boston		050627	Y

This feature is an example of industry data warehousing best practices. It offers two primary benefits. First, it allows users to go back and precisely capture what was happening at a point in the organization's past. Although employees may have moved or sales territories been redrawn, the data warehouse will accurately present information about these slowly changing dimensions as they existed at the specified time. This allows users to derive consistent, repeatable results, solidifying the value of their BI system by preserving history.

Second, users can see all values or changes over time. This capability lets users uncover longer-term trends that may affect the business. Incomplete historical information can lead to improper assumptions and compromise the quality of users' decisions.

Transactional systems typically archive everything except the most recent year or two years of data and do not provide access to supporting details. A data warehouse allows users to explore history over several years. This gives users unique perspectives on an issue's current state. This trend-analysis capability allows companies to track the impact of decisions over time.



IBM Cognos 8 Data Manager supports SCDs with an approach that completely automates the management of SCDs, significantly reducing development and maintenance time. In addition, it allows for future requests from business users that were not identified in the design phase to be implemented in a matter of minutes.

Loading dimensional history and late-arriving facts

Some transactional systems store a form of history about a dimension. For example, most HR systems must track all changes to an employee over time. This history needs to be loaded into the data warehouse on initial load and maintained as if it had been trickle-fed from the beginning of its history. IBM Cognos data integration software enables this by allowing the assignment of an effective date to the source attributes of the template. It creates all appropriate surrogates and effective date ranges while the records are being loaded into the dimension table.

Once the data warehouse stores dimensional history, another issue arises quickly. If there are a number of records in the dimension table for a given business key value, then the logical step is to take the current value. What happens if that record comes in late?

Normally this would require complex coding to check for the correct dimension record based on the transaction date from the fact record. IBM Cognos 8 Data Manager automates this process by using the dimensional framework to load the appropriate amount of history into memory and then assign the correct surrogate, based on the effective date ranges of the dimension records and the transaction date of the fact record.

There is innate flexibility in having a solution that does not require the skills of an experienced SQL expert to manage and maintain this complexity. IBM Cognos 8 Data Manager automates these processes quickly, which makes them easy to develop and maintain. IBM Cognos 8 Data Manager enables a user to make simple changes, instead of having to sort through hundreds of lines of code, or require a working knowledge of native SQL.

Data aggregation

When reviewing the performance of their business, decision-makers first look at totals, averages, counts, and maximum values to get a big picture view. With data marts, this is often done more efficiently through summary or aggregate tables. In the dimensional framework, aggregation attributes can be set for a metric or derivation at any level within a dimension table. The time to create summary tables and maintain them is significantly reduced.

Through the dimensional framework, IT can:

- Aggregate data to any level with the hierarchy of any dimension.
- Aggregate data simultaneously over a number of levels (multi-level aggregation), a number of dimensions (multi-dimensional aggregation), or both.
- Exclude detail data from the output to provide compact summary data collections.
- Include or exclude individual levels.

For example, companies can include every conceivable combination of summary data of in-depth business analysis, or just a high-level summary for management reporting.

Automated hierarchy delivery capabilities

IBM Cognos 8 Data Manager allows the construction of different types of hierarchy presenting particular views of a business dimension. IBM Cognos 8 Data Manager provides capabilities to work with balanced, unbalanced and ragged hierarchies.

Flattening and balancing

Many source systems have hierarchy data that contains variable depth hierarchies. SQL-based and OLAP reporting do not handle these parent/child structures very well, typically requiring that they be flattened to a known number of levels. This is a particular issue for reporting of financial data where cost center and cost element structures are common. IBM Cognos 8 Data Manager provides automated capabilities to flatten and balance data structures when the source data for a hierarchy contains parent/child relationships. The process of flattening the data creates a set of records with each record containing the details of a single branch of a tree. Balancing refers to taking an unbalanced tree, a tree whose leaf nodes are not all at the same level, and creating the required nodes to balance it. You perform this by taking each leaf node not at the lowest level and creating successive nodes beneath it down to the lowest level. These new nodes share the attributes of the original node.

Flattening and balancing improve support for the construction of dimension tables that are to be consumed by IBM Cognos 8 BI reporting and by any BI.

Example: HR Systems

Several associates hold Director titles. Some of the directors have direct reports and others do not have any reports. For more information, see the example below. For reporting, you may want to balance the hierarchy by taking everything down to the bottom level, so that in a situation where the customer is drilling in, up, or down, the results will make more sense.

Director	Director	Director
Report 1	Report 1	
Report 2	Report 2	
Report 3		

Recursive hierarchy levels

IBM Cognos 8 Data Manager allows a hierarchy to contain a mixture of normal and recursive levels. It is possible to attach a level to the bottom and/or top of a recursive level, even link two recursive levels together. By default, a dimension build on a recursive hierarchy will deliver a flattened version of the data, that is, the leaf level and all parent levels in a row.

Granularity of the data warehouse

Decision-makers will need to see transaction details, not just high-level summaries, to solve a business problem. Dimensional models and data marts can contain both summary and granular data. This allows appropriate calculations and summaries to be delivered to the warehouse. IBM Cognos 8 Data Manager provides the ability to perform calculations which will be delivered to the data warehouse, as well as deliver the necessary summary data to aid reporting performance. IBM Cognos 8 Data Manager extracts granular data from source systems and uses it to populate the dimensional model in the data warehouse. Decision-makers can easily access transaction-level detail and gain a micro-view of the business issues at hand. Summary or aggregate data complements the granular level detail data and provide improved performance for common queries.

Offering detailed data in the data warehouse takes pressure off the transaction system. Rather than querying the production system every time they need to perform detailed analysis, decision-makers can simply query the data warehouse for the information.

Merging data

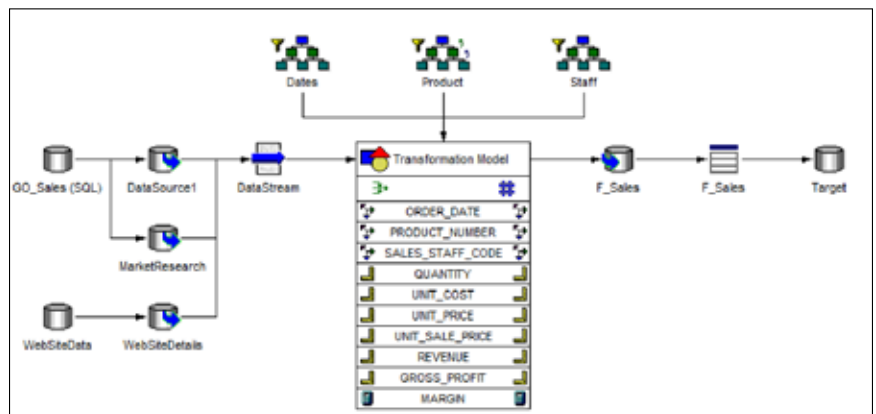
Merging data from different data sources may lead to duplicate values from the different databases. IBM Cognos data integration software allows the specification of how to merge the metrics for these values. IBM Cognos 8 Data Manager allows the developer to simply select a checkbox for the quick and efficient merging of data. The user can select the sum of the values, the maximum or minimum value, to count the values, to average the values, or the first value that occurs. This capability shortens the creation and maintenance time for the data warehouse.

Data transformation

Missing, incomplete, or inaccurate data impairs the quality of downstream BI, performance management solutions, and the resulting business decisions.

For consistent, high-quality data, IBM Cognos 8 Data Manager provides functionality that formats and integrates source data before it is stored in the data warehouse. This process might involve any number of functions, including:

- Restructuring data files, records, and fields.
- Decide what data you want in your data warehouse as part of your ETL process.
- Decoding and translating field values to enhance data.
- Improving data readability.
- Validating data.
- Calculating new values from one or more source columns.
- Simplifying data.
- Changing data types.



Input data stream transformation

The graphical design environment in IBM Cognos 8 Data Manager displays the data flows of a build and allows for direct access to build object properties, enabling rapid prototyping and deployment.

The transformation process can also include complex error correction and warning systems to ensure data integrity. Once established, the ETL process can run automatically according to the desired schedule.

IBM Cognos 8 Data Manager allows the aggregation of all data in a single pass, rather than using multi-pass steps to achieve aggregation. The sophisticated star join engine merges transactional data from multiple sources and relates it to the multiple hierarchies defined in the dimensional framework.

The Solution

Information model delivery: load

Once the source data has been transformed, IBM Cognos data integration software loads it into the target database.

IBM Cognos 8 Data Manager supports delivery of dimensional information to any appropriate storage or access platform. It allows organizations to mix and match relational and OLAP targets to choose the best technology for the situation. Organizations can partition information between databases and access tools according to specific requirements. Flexible partitioning also lets the organization send data to multiple tables or targets at the same time within one fact build, when different departments of an organization need to be provided with different data summaries. For example, data about Texas could be sent to the North America data mart, while, simultaneously, data about England could be sent to the UK data mart.

With IBM Cognos data integration, you can apply different updating rules to different tables, depending on the nature of the component data. By tailoring the data loading process to the data, information is updated more quickly, and with less demand on the source system. Tables defined as static contain data that changes infrequently. Therefore, they require refreshing on an ad hoc basis only. Tables that require more frequent refreshing can be treated differently, according to the characteristics of their data. With this flexibility, data updates can mean a complete refresh, incremental updates, or maintenance of a slowly changing dimension.

Managing the environment: deployment

Once dimension and fact tables have been designed and created, IT can automate basic data warehouse processes for efficient functioning. Automating the extracting, transforming, and loading of data means IT can execute these activities in a timely manner to maximize the resources of the computing environment. Managing a data warehouse requires coordinated tasks. Users develop JobStreams with IBM Cognos 8 Data Manager through a graphical interface. The JobStream coordinates groups of builds, processing instructions, conditions, and SQL into an operational process. JobStreams can be scheduled like any other program supporting a remote and unsupervised approach to the process.

A JobStream has nodes to automate data extraction, data transformation, data loading, exception/error handling, and logging/notification. These tasks include coordinating fact and dimension builds, data staging, cleaning data prior to data mart creation, pre- and post-processing SQL, different arrival rates of source data, and partitioning tasks to use multiple CPUs. These tasks, or job nodes, can be performed in sequence or parallel. Conditional nodes can dictate the next step in the process. Notifications can be sent out via e-mail or notes can be written to the log about the status of each job node.

The Data Movement Service allows users to publish IBM Cognos 8 Data Manager fact builds, dimension builds and JobStreams into the IBM Cognos 8 BI environment as data movement tasks. You can then integrate the data movement tasks with events, batch reporting, and email, allowing central management of the production environment. This process provides the ability to execute ETL processes and then run batch reports as a single IBM Cognos 8 job.

Collaborative projects

Most data warehouses are designed and developed by a team of developers. Tasks are assigned to individuals; however, they must also share work-in-progress and common components of the project. Furthermore, there may be many projects that all adhere to a set of components, testing groups, and production teams.

IBM Cognos 8 Data Manager lets development teams manage their projects with an external source code control system. Using the source code control application, they can allow check-out/ins, locking/unlocking, versioning, and allow each developer to work individually while being part of a larger network of developers.

IBM Cognos 8 Data Manager also provides the ability to package components for sharing among projects—all in a graphical user interface. This process allows the developer to select a component to package. With IBM Cognos 8 Data Manager, the component, along with all of its underlying components, are automatically included in the package. The package can then be shared with another project and imported into the environment.

Some projects can grow very large with hundreds of data loads and supporting components. Finding specific components in large projects and seeing the impact of changes becomes an arduous task. IBM Cognos data integration provides a utility to navigate the project. The Navigator allows developers to select one component and see all other components that use it, allowing component impact analysis. The Navigator will also allow the developer to search by name for matching components.

Usually the last task most development teams complete is documentation. Documentation of a data warehouse can be difficult because it requires documenting all the data flows, transformation, and targets. IBM Cognos 8 Data Manager automates this process by creating an HTML document of the project, complete with hyperlinks and comments from every developer.

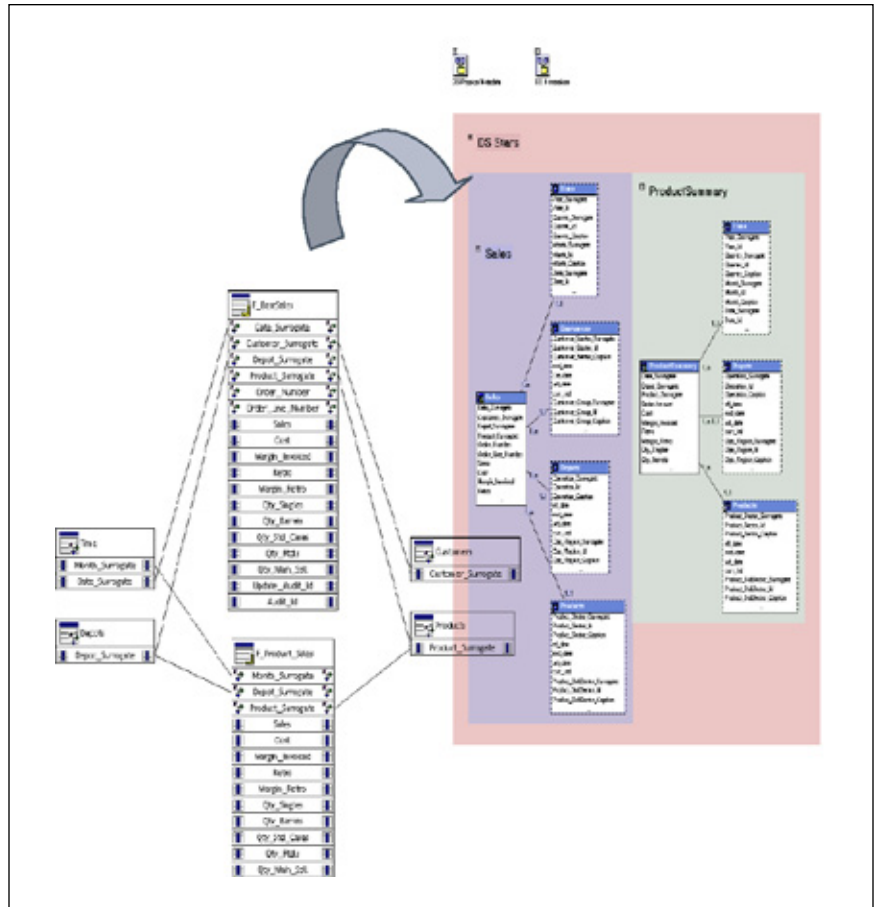
Increased speed to BI deployment

The IBM Cognos 8 Data Manager model describes additional value over the target tables and columns that have been delivered to the warehouse. The additional value can be viewed by the dimensional framework, star schema groupings, tables and joins, measures, descriptions, and business names.

The dimensional framework contains

- Dimension tables & hierarchies
- Keys (surrogate, business, levels)
- Time variance
- Conformed model, which is important for multi fact joins

The IBM Cognos 8 Data Manager model can be exported for consumption by IBM Cognos 8 Framework Manager, which is the metadata modeling environment for IBM Cognos 8 BI.



The modeling capabilities of Framework Manager.

This model includes the full metadata for the warehouse along with all joins and cardinality (or elements in the data set). The dimensional metadata, meaning the hierarchical structures, are also passed to the metadata environment.

In addition, IBM Cognos 8 Data Manager can leverage the IBM Cognos 8 BI metadata model that spans all data and applications. This lets companies extend their IBM Cognos 8 platform by integrating IBM Cognos planning and budgeting data with their reporting and analysis.

The benefits of using IBM Cognos 8 Data Manager

Using IBM Cognos 8 Data Manager capabilities for data warehouse creation, as part of an overall business intelligence solution, offers application developers several key benefits including:

- A dimensional framework—Building conformed dimensions and having the ability to manage the tasks associated with dimensionality provides consistency and reusability of dimensions. It delivers a consistent view of the business.
- Automatic dimensional management—IBM Cognos data integration automates surrogate key generation and management and automatically maintains links between dimension and fact tables.
- Automatic Slowly Changing Dimension management— Point and click functionality automatically creates and maintains slowly changing dimensions.
- Automated hierarchy management— Includes the ability to manage recursive hierarchies, automatically balance and flatten hierarchies and deliver segments of large hierarchies.
- Flexibility—The transformation model is built around logical fact tables that map source extractions and target deliveries. Sources are not tied directly to targets, therefore, sources or delivery targets may be changed with minimal impact on other aspects of a job.
- Powerful transformation— Overcome all the challenges of transforming data for the warehouse.
- Declarative approach—The intuitive graphical interface eliminates code writing and maintenance. This delivers results quickly and reduces development and maintenance time.

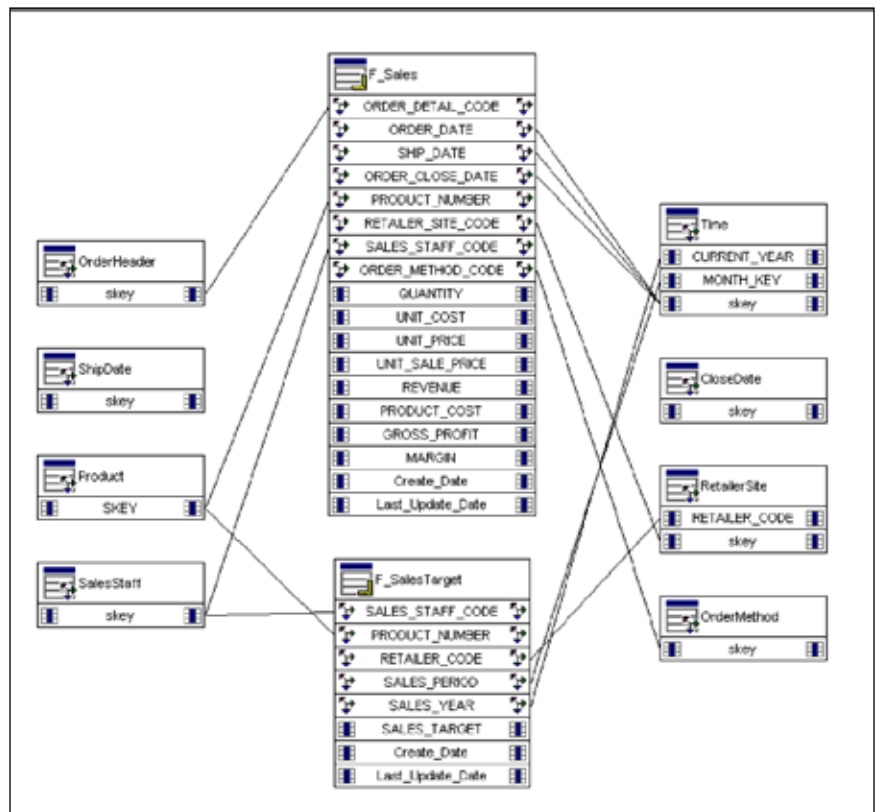
- Reusability – Conformity of dimensions within the dimensional framework lets dimensions be reused throughout data warehouse project.
- One product with IBM Cognos 8 BI – Metadata is exchanged between IBM Cognos data integration to the IBM Cognos 8 BI metadata environment. In addition, any metadata modeled first in the IBM Cognos 8 BI metadata environment can be leveraged within IBM Cognos 8 Data Manager and treated as a data source.
- Integration with application/scheduling architecture – The cross-platform, script-driven engine in IBM Cognos data integration software lets IT embed IBM Cognos integration technology into different applications and scheduling architectures across UNIX and NT environments.
- Programmability – Full command line access delivers flexible transformation package integration. Scripts can take advantage of runtime variables.

IBM Cognos 8 Data Manager goes beyond integration

There are many integration tools on the market that provide basic table-to-table data movement, can read from any source, can transform, can aggregate, provide filtering, and all the basic capabilities for most integration solutions. Like the best of those tools, IBM Cognos 8 Data Manager extracts data from data sources including transaction systems, warehouses (relational and OLAP), flat, legacy or modern sources. It then performs transformation, aggregation, and filtering operations on this data before loading the data to a target – staging tables or a star schema – in a relational database.

The primary differentiator of IBM Cognos 8 Data Manager is that it has a multidimensional model at its core. It was designed specifically to build dimensional data marts.

The designer and core engine work in terms of fact and dimension deliveries, not in terms of arbitrary table movement. Since it starts from a premise of dimensionality, IBM Cognos data integration software creates data warehouses, data marts or conformed marts that most closely resemble how a business and its employees operate.



A conformed dimension, or coordinated mart

The shared dimensional framework of IBM Cognos data integration allows IT to design a series of coordinated data marts.

Conclusion

Summary

A data warehouse with a coordinated data mart approach for BI capabilities gives users the ability to make coordinated, high-quality business decisions. The IBM Cognos approach to data warehousing provides a framework that delivers BI capabilities such as reporting, analysis, scorecarding, dashboarding, and event lifecycle management – all in a single product, on a single, proven architecture.

The end result is a comprehensive solution for enterprise-ready business intelligence with the shortest possible time to success. That means better, faster results for IT and report developers, increased satisfaction for decision-makers, and improved corporate performance.



© Copyright IBM Corporation 2009

IBM Canada
3755 Riverside Drive
Ottawa, ON, Canada K1G 4K9

Produced in Canada
March 2009
All Rights Reserved.

About IBM Cognos BI and Performance Management

IBM Cognos business intelligence (BI) and performance management solutions deliver world-leading enterprise planning, consolidation and BI software, support and services to help companies plan, understand and manage financial and operational performance. IBM Cognos solutions bring together technology, analytical applications, best practices, and a broad network of partners to give customers an open, adaptive and complete performance solution. Over 23,000 customers in more than 135 countries around the world choose IBM Cognos solutions.

For further information or to reach a representative: www.ibm.com/cognos

Request a call

To request a call or to ask a question, go to www.ibm.com/cognos/contactus.

An IBM Cognos representative will respond to your enquiry within two business days.

IBM, the IBM logo and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

Any reference in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

Endnotes

- 1 Ralph Kimball, *The Data Warehouse Toolkit: A Complete Guide to Dimensional Modeling*, John Wiley & Sons, 2002.