Glossary

- Business Intelligence
- Corporate performance management
- Data Warehousing

Document Version	Format	Date	Audience
V 0.0	Draft	11/April/2007	Internal Team
V 1.0	Review	18/April/2007	Internal Team
V 2.0	Final Release	08/May/2007	Public
V 2.01	Additions	23/May/2007	Public
V 2.02	Additions	04/June/2007	Public

Please note that Digital Viper accepts no responsibility whatsoever for the use of this document by any parties.

This document is a working document in which updates are constantly being added. New terms can be submitted at the following address: info@digitalviper.co.uk

Access permissions

A definition of which resources the members of a group or a user can read, change, or otherwise use. Examples of resources are reports and folders.

Agent

An application that searches the data and sends an alert when a particular pattern is found.

Aggregations

Information stored in a data warehouse in a summarized form.

Instead of recording the date and time each time a certain product is sold, the data warehouse could store the quantity of the product sold each hour, each day, or each week.

Aggregations are used for two primary reasons:

- To save storage space. Data warehouses can get large. The use of aggregations greatly reduces the space needed to store data.
- To improve the performance of business intelligence tools. When queries run faster they take up less processing time and the users get their information back more quickly.

Some data warehouses store both the detailed information and aggregated information. This takes even more space, but gives users the possibility of looking at all the details while still having good query performance when looking at summaries.

Some systems use aggregations for historical data. Perhaps detailed data is kept on-line for a year. After that the detailed data is kept in a less accessible, permanent storage format, and only the aggregated, summary data is kept on-line.

Aggregations are often created as the sum of the individual records. You can also have aggregations for count, distinct count, maximum value, and minimum value.

Aggregation Layer

Data in a data warehouse and a data mart is structured in order to provide the most efficient result when being queried. In order to do this data is held in objects at the highest level of granularity (detail) possible for the report or query. A data warehouse will therefore often have one or more aggregation layers or levels. These layers or levels are created from the detail or atomic data level.

Alert

A message that is sent automatically by a computer system when a certain situation occurs. One of the greatest benefits of data warehousing is the ability to set alerts.
A store manager can be automatically informed when a certain product's sales fall below or rise above a specified range.
A factory manager can be automatically informed when the failure rate of a product exceeds a specified level.
A sales manager can be automatically informed when a member of his staff achieves a personal high level of sales for a time period.
Alerts allow a company to receive critical business information in the quickest possible time.

Analysis Services

Business Intelligence tools included with Microsoft SQL Server 2000.

The same product in Microsoft SQL Server 7.0 was called OLAP Services. The name was changed because the Microsoft SQL Server 2000 version included data mining capabilities as well as the OLAP capabilities.

Anonymous access

A method of accessing resources in which users are not authenticated.

Archive

See deployment archive.

Associate

The action of relating a column to a grouped item. Associating columns is an alternative to grouping. Both associating and grouping suppresses duplicate values in a report layout, but unlike grouped columns, associated columns do not affect the sort of the query results.
Also known as Level Span
Also known as Group Span

Atomic Data

Atomic Data refers to data that is held as its most detailed grain or level. This is usually data that has just been imported or loaded into the datawarehouse.

Atomic Data Store

The atomic data store is the load or detailed level of the data warehouse and often replicates the structure of the source database or system.

Attribute

Additional information included with a dimension, that is not used in defining the levels of the dimension.
Dimensions become more useful when there are many descriptive attributes that can be used for analyzing the data. In Microsoft Analysis Services attributes are used to create member properties and those member properties can be used to create virtual dimensions.

Authentication

The process of verifying the identity of users when they log on. Authentication is usually done by
comparing a user name and password to those in an authorized list in the authentication provider. Users must be authenticated before they can be authorized to use any secured resources.

Authentication provider

A service that authenticates and defines users and the groups of which they are members.

Authorization

See access permissions.

Binary Search

A dichotomizing search with steps in which the sets of remaining items are partioned into two equal parts.

Bit Map

A specialized form of an index indicating the existence or non-existence of a condition for a group of blocks or records. Often used where the number of distinct values is less than 20.

Burst

To create many report outputs by running a single report once.
Example: A report shows outlet sales by region. The report is required by regional managers. The report can be burst to produce one output per region with sales by outlet.
Set up bursting in Cognos Report Studio and enable it in Cognos Connection.

Burst key

The dimension or level of a query in the report specification that is used to create, or burst, a set of report outputs.

Business Intelligence Tools

Software that enables business users to see and use large amounts of complex data.

The following three types of tools are referred to as Business Intelligence Tools:

1. Multidimensional Analysis Software - Also Known As OLAP (Online Analytical Processing) - Software that gives the user the opportunity to look at the data from a variety of different dimensions.
2. Query Tools - Software that allows the user to ask questions about patterns or details in the data.
3. Data Mining Tools - Software that automatically searches for significant patterns or correlations in the data.

Calculated Member

In Microsoft Analysis Services, a member that is calculated from one or more other members using an MDX formula.

Calculated members are often measures, but they can also be members of levels from other dimensions. In an MDX query, a calculated member can be created in the WITH clause.

Canonical model

A data model that represents the inherent structure of data without regard to either individual use or hardware or software implementation.

Cardinality

A property of a relationship that is used to ensure that queries return the correct results. Cardinality describes the association between two query subjects and is set at each end of the
relationship.
Cardinality is expressed by using the following notation:
• 0..1 (zero or one match)
• 1..1 (only one match required)
• 0..n (zero or more matches)
• 1..n (one or more matches required)
The first part of the notation specifies the minimum required matches that must exist between tables: 0 indicates that finding a match is optional, and 1 indicates that at least one row must match. The second part defines the maximum required matches (1=1, n=many).
For example, A and B have an association with one another. The cardinality of 1...1 for table A

means that for each row in table B, there is only one row in table A. The cardinality of 0...n for table B means that for each row in table A, there are zero or many rows in table B.

Cascading prompt

A prompt that uses values from a previous prompt to filter the values in the current prompt or picklist. For example, your report prompts for Countries and cities. By selecting 'United Kingdom' in the first prompt, Manchester and London are shown in the second prompt, but New York, Sydney and Singapore are not shown.

Cell

A cell is a single point in a cube or dcube (Cognos Planning)
Cubes have cells for all of the possible combinations of points from all of the cube's dimensions.

Cellset

A set of data returned from a cube with an MDX query.
A cellset in the multidimensional world is equivalent to a rowset in the relational database world.

Certificate

A document that identifies someone or something by name and by a reference to a certification authority. In ReportNet, each computer in a distributed installation uses a different certificate.
Certificates are used for encryption and decryption.

Certificate authority (CA)

The component of ReportNet that issues certificates (identification) to each computer on which ReportNet components are installed. You can also use a third-party certificate authority.

Changed Data Capture

In a database replication, changed data capture occurs when only the data that has changed since the previous replication is copied.

Changing Dimensions

A dimension that has level or attribute data that needs to be updated.

Data changes in the fields for levels or attributes is one of the most challenging design issues for multidimensional (star schema) data modeling. This issue is often referred to as the handling of **Slowly Changing Dimensions**. There are three ways outlined by Ralph Kimball (and others) to handle this situation:

Type #1. Change the data in the dimension table.
Type #2. Add new records to the dimension table that contain the new data.
Type #3. Add new fields to the dimension table to contain the values before and after the change.

Various authors (including Ralph Kimball) also suggest the splitting of dimensions into separate dimensions when a consolidated dimension would result in a high level of dimension change.

Clickstream Data

Data regarding web browsing.
Web servers capture a large amount of data in the process of receiving requests for web pages. This data includes page served, time, source of the request, type of browser making request, etc. When analyzed, this data provides information about the behavior of individuals who are browsing the internet. It can help businesses analyze where visitors are coming from, what type of visitors are most likely to buy certain products, what type of web pages are most attractive, etc. This information is essential for analyzing the effectiveness of internet ad campaigns and, in general, for finding ways to improve the effectiveness of internet commerce.

Clickstream data typically requires a significant amount of transformation as it is loaded into a data warehouse. Once in the warehouse it can be used for standard reports, for OLAP, and for data mining.

Cognos Connection

The Cognos portal for Report Studio, Event Studio, Metrics Studio, Analysis Studio and Query Studio.

Common gateway interface (CGI)

A standard that describes how Web servers should access other programs in order to create a document that will appear in a Web browser. For example, Web servers often use CGI programs to process forms.

Condition

An expression that yields a 0 or 1 (Boolean) value. Conditions are normally used in filters, conditional formatting, styles and calculations.

Conformed Dimension

A dimension that is used in more than one cube.
The use of conformed dimensions and shared measures is the primary way a set of data marts can be united into one consolidated data warehouse. The use of common dimensions throughout a number of data marts or warehouses within a business. Ensures a single list is used for each dimension.

Conceptual Schema

A consistent collection of data structures expressing the data needs of the organization. This schema is a comprehensive, base level, and logical description of the environment in which an organization exists, free of physical structure and application system considerations.

Condensation

The process of reducing the volume of data managed without reducing the logical consistency of the data.

Contact

In report distribution, a named email address to which reports can be sent. Contacts are never
authenticated.

Content locale

The code or setting that specifies what language and regional preferences to use for the data that appears in a report.

Content Manager

The ReportNet service that manages the storage of reporting applications, including application-specific security, configuration data, models, reports, and report output. Content Manager is needed to publish models, retrieve or store report specifications, manage scheduling information, and manage the Cognos namespace.

Content store

The database that contains all data that ReportNet needs to operate, such as
• report specifications, published models, and the packages that contain them
• connection information for data sources
• information about the external namespace, and the Cognos namespace itself
• information about scheduling and bursting reports
Design models and log files are not stored in the content store.
The ReportNet service that uses the content store is named Content Manager.

Commit

The explicit or implicit process of confirming the writing of data into an object.

Connector

A symbol used to indicate that one occurrence of data has a relationship with another occurrence of data. Connectors are used in conceptual data base design and can be implemented hierarchically, relationally, in an inverted fashion, or by a network.

.Cooperative Processing

The ability to distribute resources (programs, files and data bases) across the network. Oracle 11g claims to have this ability.

Constraint

A restriction on the possible values that users can enter in a field.
A security specification that denies one or more users the ability to access a model component
or to perform a modeling or authoring task.

Corporate Performance Management tools (CPM)

The grouping of reporting, forecasting, budgeting and financial planning software

Credentials

Information stored about the identity of a ReportNet user, usually a user name and password. You can assign your credentials to someone else so that they can use resources that you are authorized to use. Credentials are created for the components of ReportNet. If a user schedules or programs an action, then credentials must be stored in the content store.

Cube

Also Known As Multidimensional Cube and Powercube
The fundamental structure for data in a multidimensional (OLAP) system.

1. A cube contains dimensions, hierarchies, levels, and measures. Each individual point in a cube is referred to as a cell.
2. The part of a query that defines the levels and order of data and summaries.

Data-Based Knowledge

Knowledge derived from data through the use of Business Intelligence Tools and the process of Data Warehousing.

Most of our knowledge is based on a combination of our experience, perception, and intuition. Business Intelligence and Data Warehousing give us a new kind of knowledge based on data.

Data-based knowledge can have several advantages over experience/intuition-based knowledge:

1. It can be more accurate because it is based on so many detailed facts.
2. It can be more current because the data warehousing and business intelligence tools can so quickly analyze new data.

3. It can be more comprehensive because so many different perspectives are available through the rapid recombination of elements from different dimensions and different levels of the data hierarchy.

4. It can give new insights because there are complex patterns in the data that can be discovered by data mining that would never be detected by human analysis.

5. It can be less subjective because conclusions are tied directly to the physical data.

Data Cleansing

Removing errors and inconsistencies from data being imported into a data warehouse.

Data Dictionary

A document, spreadsheet or software tool for recording the definition of data, the relationship of one category of data to another, the attributes and the keys of groups of data.

Data Driven

Development the approach to development that centers around identifying the commonality of data through a data model and building programs that have a broader scope thn the immediate application.

Data, Information, and Knowledge

Data is the reality that a computer records, stores, and processes.

The use of computers can be referred to as data processing. At the lowest level data has no significance for people. This lowest level in the perception of reality is sometimes referred to as "raw data".

Information is what a person is able to understand about reality.

Information systems use computers to organize data in such a way that people can understand the results.

Knowledge is what a business uses to make decisions.

The process of organizing information in such a way as to create data-based knowledge is called Data Warehousing. The software products that present this knowledge to users are called Business Intelligence Tools.

The goal of business intelligence and data warehousing - changing data into information and knowledge.

Organizations are gathering and storing more and more data. Every year the amount of data in the world is approximately doubling. This data is of little benefit unless it can be turned into useful information and knowledge.

Information by itself is an inadequate basis for business decisions because the amount of information, like the amount of data, is overwhelming. Business Intelligence Tools are designed to find what is significant - what really adds to our useful knowledge - in the piles of data and information.

Data Mart

AKA: Local Data Warehouse or Datamart

A database that has the same characteristics as a data warehouse, but is usually smaller and is focused on the data for one division or one workgroup within an enterprise.

There are three different (and somewhat contradictory) views of the place of the data mart in the world of data warehousing.

1. The data warehouse gathers all the information from the various legacy systems. Specialized data marts are then created with a subset of the information in the data warehouse. These data marts are easier to use because they only have the particular information the specific user group needs. The use of several data marts also allows the querying load to be spread among several different computers. This can reduce network traffic.

2. Free-standing data marts are created, independent from a data warehouse. The information for the data mart probably comes from just one legacy system. It is quicker and cheaper to build a separate data mart instead of building an enterprise-wide data warehouse with data marts derived from it. The drawback of this solution is that the company's data is not integrated (and thereby violates one of Bill Inmon's original defining characteristics of the data warehouse). If several separate data marts are built using this strategy, they will usually contain data that is duplicated and inconsistent.

3. The data mart is the prototype or the first step of a data warehousing process. An enterprise picks the division or group that would most benefit from data-based knowledge. A data mart is built with that group's data. Additional types of information are added to the data mart as time goes on until it is turned into a data warehouse.

New terminology is often created and developed for marketing purposes. The term 'data mart' probably has a marketing advantage over the term 'data warehouse'. The whole data warehousing process is about creating data-based knowledge and bringing that knowledge to people. A warehouse is a place where things are stored away. A mart is a convenient place to buy something. Most data warehousing professionals (including myself) include ready access to information as a defining characteristic of the term 'data warehouse'. I think, though, that the term 'data mart' captures this sense of data availability more effectively.

Data Migration

The movement of data from one environment or application to another.

This happens when data is brought from a legacy system into a data warehouse. This also occurs when data is moved from a development environment to a test (or stage) and live (or production) environment.

Data Mining

The process of finding hidden patterns and relationships in the data.

Analyzing data involves the recognition of significant patterns. Human analysts can see patterns in small data sets. Specialized data mining tools are able to find patterns in large amounts of data. These tools are also able to analyze significant relationships that exist only when several dimensions are viewed at the same time.

Users can ask data questions using standard queries when they know what they're looking for. Queries can be written for questions like this: "Which of our out-of-town customers have given us the most business in the last year?" **– Standard reporting using a tool such as Cognos Impromptu or ReportNet**

Data mining is needed when the user's questions are more vague or general in nature. Data mining questions would include: "What attributes characterize the customers that gave us the most business in the past year?" **– Exploratory reporting using a tool such as Cognos PowerPlay or Analysis Studio**

Data Quality Assurance

Also Known As: Data Cleansing or Data Scrubbing

The process of checking the quality of the data being imported into the data warehouse.

Data quality assurance is one of the greatest challenges in the process of data warehousing. If the data-based knowledge generated by the data warehouse is to be trusted, the data entered into the warehouse must be complete and accurate - "garbage in, garbage out".

Data quality can be a challenge for several reasons:

The data is being consolidated from a variety of legacy sources that may have differing definitions of key concepts such as "customer" or "profit".
The legacy data was not originally collected for the purpose of decision support so some of the key data might be missing, incomplete, or not as accurate as desired.

There might be times when all the data is not received from one of the legacy systems. This could make comparisons between time periods invalid.

A significant portion of time in the development process should be set aside for setting up the data quality assurance process and implementing whatever data cleansing is needed..

In a production environment, there should be a data quality report generated after each data warehouse import. There should be provision for rolling back an import if data quality testing indicates that the data is unacceptable.

Data source connection

The named information that defines the type of the data source, its physical location, and any signon requirements. A data source can have more than one connection.

Data source

A named set of connections to physical databases. Framework Manager uses the name of the

data source for models. The ReportNet server uses the data source connections to access the
physical databases.

Data Scrubbing

Removing errors and inconsistencies from data being imported into a data warehouse.

Data Transformation

The modification of data as it is moved into or within the data warehouse.

- Data Cleansing - Part of the Process of Data Quality Assurance
- Dimensioning - Organizing the data into the multidimensional structure of one or more star schema(s).
- Normalization - Organizing the data into the normal structure of a relational database.
- Denormalization – Organizing data into a dimensional model from a relational model
- Processing Calculations
- Changing Data Types
- Making the Data More Readable
- Replacing Codes with Actual Values
- Summarizing the Data
- Creating derivations
- Restructuring the data
- Summarizing or aggregating the data.

Data Type

A data type defines the type of data stored in a specific database column, such as date, numeric or character data. Significant differences in data types exist between different platforms' databases.

Data Warehouse

AKA: Datawarehouse or Information Warehouse

A database where data is collected for the purpose of being analyzed.

The defining characteristic of a data warehouse is its purpose.

Most data is collected to handle a company's on-going business. This type of data can be called "operational data". The systems used to collect operational data are referred to as OLTP (On-Line Transaction Processing).

A data warehouse collects, organizes, and makes data available for the purpose of analysis - to give management the ability to access and analyze information about its business. This type of data can be called "informational data". The systems used to work with informational data are referred to as OLAP (On-Line Analytical Processing).

Bill Inmon coined the term "data warehouse" in 1990. His definition is:
"A (data) warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process."

Subject-oriented - Data that gives information about a particular subject instead of about a company's on-going operations.
Integrated - Data that is gathered into the data warehouse from a variety of sources and merged into a coherent whole.

Time-variant - All data in the data warehouse is identified with a particular time period.

Non-volatile - Data is stable in a data warehouse. More data is added, but data is never removed. This enables management to gain a consistent picture of the business.

Data Warehousing

The process of visioning, planning, building, using, managing, maintaining, and enhancing data warehouses and/or data marts.

Whether we're building a data warehouse, a data mart, or both, we are taking part in a complex, on-going process. The emphasis in the data-based knowledge business needs to be kept on the process.

Actions inherent in data warehousing:
Data Replication
Data Transformation
Data Quality Assurance
Data Storage
Metadata Storage
Data Mart Population

Database Link (Dblink)

A connectivity path between two separate instances in Oracle. Allows queries to pass and retrieve data from an external database to an internal database.

Database Management System (DBMS)

The software that is used to store, access, and manage data.
There are two main types of Database Management Systems used for business intelligence and data warehousing - specialized Multidimensional Database Management Systems (MDBMS) and the more widely used general purpose Relational Database Management Systems (RDBMS).

Decision Support System (DSS)

A computer system designed to assist an organization in making decisions.
The Decision Support Systems and Enterprise Information Systems of the 1980's and early 1990's were forerunners of today's Business Intelligence Tools.

Density or Dense

The level of population of data within a table.
One of the primary challenges of storing multidimensional data is the degree of sparsity that is often encountered. When many dimensions are considered with a fine grain of detail, most of the cells will be empty.

Deployment

The process of moving a ReportNet application (reports, models, and so on) to a different instance of ReportNet. For example, you often create reports in a test environment and then deploy them to production.
When you deploy an application, you export, transfer, and import it.

Deployment archive

A file used for deployment. A deployment archive contains the data from the content store that is
being moved.

Deployment specification

A definition of what packages to move (deploy) between source and target environments, the deployment preferences, and the archive name. You use deployment specifications for import and for export

Derivation

The creation of a data item from another data item by using a calculation or structured condition.

Detail Data

See Atomic Data

Detail level / layer

See Atomic Data Store
The detail level can also refer to the most detailed level in a OLAP hierarchy either within the data warehouse or an OLAP tool such as Cognos PowerPlay or Analysis Studio.

Dimension

One of the perspectives that can be used to analyze the data in an OLAP cube or a fact table.

When you are browsing the data in a cube, you can view the data from the perspective of different combinations of dimensions.

For a Sales database, the dimensions could include Product, Time, Store, and Promotion.

Dimensions contain one or more hierarchies, which have levels for drilling up and drilling down in the the cube. When a dimension has just one hierarchy (which is quite common), people often refer to the dimension itself having levels.

Dimension Table

A dimension table is a table holding qualitative (descriptive) data about a logical structural area of the business such as products or customers. A dimension table is normally used within a star schema and links directly to one or more fact tables.

The dimension table has a primary key which is used to connect it to the fact table.

The dimension table has one field for each level of each hierarchy contained in the dimension. The data values in these fields become the members of each of the dimension's levels.

Each field within a dimension table describes individual characteristics of the dimension.

If there are multiple hierarchies in the dimension, there is one level field for each distinct level in each of the hierarchies. If the hierarchies share some levels in common, they are represented by a single field For Calendar and Fiscal hierarchies in a Time dimension, the level fields could be Fiscal Year, Calendar Year, Fiscal Quarter, Calendar Quarter, Month, and Day.

For the Product dimension table, some of the attribute fields could include Description, Product Number, Product Type, Department, Package Size, Weight, Shelf Length, etc.

The dimension tables in a star schema are intentionally de-normalized. The level fields and the attribute fields contain data that is duplicated in many of the records. This normally does not add a significant amount to the amount of storage space needed in the database, because the overall size of each dimension table is very small when compared to the size of the fact table.

Dimensioning

The process of transforming data into a multidimensional (or star) schema.

DDL

Data definition language
DDL refers to SQL that describes the definition of the database and objects within it. For example when a new table is created, the SQL used to create it can be referred to as DDL.

DML

Data Manipulation Language
DML refers to SQL that is used to manipulate or transform data.

Drill Down and Drill Up

The ability to move between levels of a hierarchy when exploring data using reporting tools.
Drill Down - Changing the view of the data to a greater level of detail.
Drill Up - Changing the view of the data to a higher level of aggregation.

Multidimensional analysis (OLAP) tools organize the data in two primary ways: in multiple dimensions and in hierarchies.

Drilling down and drilling up allow an analyst to move down and up the hierarchies to see how the information at the various levels is related.

Data Transformation Services (DTS)

An ETL tool provided as a part of Microsoft SQL Server.
DTS was first released with SQL Server 7.0. It provides a design environment for creating data transformation applications. DTS has now been replaced by SSIS in SQL Server 2005.

EDI

Electronic Data Interchange is a standard format for exchanging business data.
Encryption the transformation of data from a recognizable format to a form unrecognizable without the algorithm used for the transformation.

EIS

Enterprise Information System
AKA: Decision Support System (DSS)
A computer system that presents a summary of a company's important data.

Entry

A piece of data in the content store. You use the portal tool to view, manipulate, and apply security to entries. Report results and specifications, folders, shortcuts, models, and packages areall entries.

ETL (Extract, Transform, and Load)

ETL refers to the process of getting data out of one data store (Extract), modifiying it (Transform), and inserting it into a different data store (Load).

Expression

A mathematical or logical construction of code that is used in calculations and filters to derive values from one or more columns in a database. An expression may be as simple as a reference to column, or as complex as a long series of mathematical expressions that includes many operators, parameters, and database functions. A condition is also an expression as is a calculation.

Fact table

A table containing quantitative data.
Within a dimensional model (star schema) a fact table will contain:

1. A foreign key (a pointer to a row in a dimension table)
2. The fields storing the individual facts (or measures) - such as number, amount, or price.

The fact table should be designed to represent the required level of granularity by report users.
The more detailed a fact table (granular), the more foreign keys and facts the larger the table will be (in memory size).

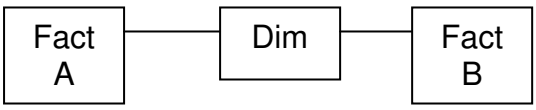
A detail level fact table should store facts at the lowest level of detail possible. A lower (more summarized level) of granularity can be created later in aggregated fact tables to ensure the speed and efficiency of commonly run reports.

Factless Fact

A fact table without any metrics in it. Only foreign keys of surrounding dimensions are held. The fact is only created to facilitate the joining of dimension tables.

Fan Trap

AKA 'Chasm Trap'
The use of data from two fact tables with a central dimension table within one query.



If one entry from the dimension table is used and fact table B has 5 facts the result is 5 rows in the resultant data set.
If the query is then linked to Fact B that has 10 rows relating to the single dimensional item, the resultant data set will be 50 rows. This happens because the result set of the first two tables (5 rows) is multiplied by the entries in the factB table (10 rows) as there are now effectively 5 entries in the dimension table.

This error can be avoided by using a stitch query either built into the reporting tool or using stitch SQL commands such as UNION or INTERSECT.

Flat File

A collection of records containing no data aggregates, nested repeated data items, or groups of data items.

Functional Decomposition

The division of operations into hierarchical functions that form the basis for procedures.

Gateway

An extension of a Web server program that transfers information from the Web server to another
server. Gateways are often CGI programs, but may follow other standards such as ISAPI and NSAPI.

Global Index

An index that spans table partitions

Glyphs

An image of a character in a font. Letters are glyphs, but in most cases, the term is used in discussions of non-alphabetic writing
systems.

Granularity

The level of detail of the data within a table. This is most often used to refer to fact (measure) data.

Group

1. Security: A list of users or other groups that you use as a single object for setting access permissions. Groups are normally created in either Access Manager or Cognos Connection. Users are authenticated as members of the groups they belong to, unless the groups have also been defined as roles. Users can choose one or more roles when they log on, to change what data they have authorization for.
2. Reporting & SQL: Grouping is the action of organizing and sorting common values of a column or query item together. Headers and footers are often created for grouped items.

Heterogeneous

Multiple or different or both. This term is often used to describe the data sources for a data warehouse.

Heuristic

The mode of analysis in which the next step is determined by the results of the current step of analysis.

Hierarchy

Organization of qualitative data into a logical tree structure.

Dimensions can have one or more hierarchies. A Time dimension, for example, could have a Calendar hierarchy and a Fiscal hierarchy. Hierarchies contain levels, which organize data into a logical structure.

It is the combination of a multidimensional with a hierarchical view in Business Intelligence Software that allows users to grasp large amounts of data. If each member in a level has 5 to

10 children that are members at the next lower level, the user has a better chance of understanding the significance of the data.

Moving between the levels of a hierarchy is called drilling up and drilling down.

HTTPS

A secure version of HTTP that incorporates secure sockets layer (SSL) security. HTTPS is designed to transmit individual messages securely. The URL-accessed method is https://. The default HTTPS port number is 443. This default number is assigned by the Internet Assigned Numbers Authority. Cognos products use HTTPS and SSL to encrypt and transmit passwords securely over the Internet.

Hybrid OLAP (HOLAP)

A combined use of Relational OLAP (ROLAP) and Multidimensional OLAP (MOLAP).

In HOLAP, the source data is usually stored using a ROLAP strategy and aggregations are stored using a MOLAP strategy. This combination usually results in the least amount of storage space and the fastest cube processing.

Hyper-Cube

Also Known As Cube and Multidimensional Cube
A cube with more than three dimensions.
A cube is an object with three dimensions. A hyper-cube is a cube-like structure with more than three dimensions. In the world of OLAP, hyper-cubes are nearly always simply referred to as cubes.

Index / Indexing / Indices

The method of using a sorting algorithm to data within one or more columns to facilitate fast value location for sorting, joining, filtering or navigation of data values within large data sets.

Information

Data that is structured and labeled for human understanding and use.

Instance

A set of values representing a specific entity belonging to a particular entity type.
Also – A database when described by Oracle.

Iterative Process / Analysis

The mode of processing or analysis in which the next step of processing depends on the results obtained by the existing step in execution. This method of development is often used when creating reports.

Job

A group of reports and other jobs in ReportNet that you run and schedule as a batch.

Job step

The smallest part of a job that can be run separately. Usually, a job step is a report. A job step can also be another job.

Join

A join is a link allowing two query subjects to be used in a single query. The most commonly used join in a data warehouse is between a primary key on a dimension table to a foreign key in a fact table.
See also relationship and cardinality.

Key

A unique identifier for a row in a table.
A table can have more than one key. Keys are also used to create relationships.
For example, the CUST-ID column is the **primary key** for the CUSTOMERS table because there can be only one customer ID for each customer.
Foreign keys are columns that establish relationships between tables. For example, the CUST-ID column in the ORDERS table is the **foreign key** that joins the ORDERS table to the CUSTOMERS table. Each order in the ORDERS table can be associated to only one customer.

Knowledge

The assimilation and understanding of information by individuals.

Latency

Often used to mean any delay or waiting that increases real or perceived response time beyond the response time desired. (ie query time in databases or in websites)

Layout

In reporting, layout defines the appearance of the report, including formatting, style, and design. In report specifications, layout is also the name of an element that defines how the data returned by queries is presented.

Legacy System

A computer system that has been used within a company normally for a long period of time. The term 'Legacy System' is often used when the old system is to replaced by a new system or when data from the old system is to be used by a new system such as a data warehouse. The data in these systems is usually mutually incompatible and sometimes inaccurate. One of the biggest challenges of the data warehousing process is to bring data out of the variety of systems where it currently is located and organize it so it all fits together in the data warehouse.

Level

A group of query items that must contain a key query item, so that each member within the group is unique.
Levels may contain other non-key query items. Levels are parts of dimensions and hierarchies.
For example, a geographical dimension might contain levels for country, region, and city.

Link

See Join

Local Cube

A cube contained in a file.

Microsoft Analysis Services (OLAP Services) provides the ability to take all or a subset of a server cube and create a local cube file. The local cube can be used to analyze OLAP data while the user is disconnected from the network.

Local Data Warehouse

See Data Mart

Locale

A code that is used to set
• the language or dialect used for browsers, report text, and so on
• the regional preferences, such as formats for time, date, money, money expressions, and time of day
In ReportNet, you can specify a locale for the product interface (product locale) and for the data
in the report (content locale). A locale is also stored to record what locale an author used to create a report specification or a Framework Manager project.

Lockup

The event that occurs when update is done against a data base record and the transaction has not yet reached a commit point.

Logging

The automatic recording of events.

Many-to-many relationship

A join between two query objects (such as tables or views) where many keys on one query object links to many entries on the other query object. This form of query will produce very high volume query result.

Maximum Transaction Arrival Rate

MTAR is the rate of arrival of transactions at the moment of peak period processing.

MDX (Multidimensional Expressions)

The querying language for OLAP cubes.

MDX has some similarities to SQL, but has many unique features. The following query returns a cellset with the names of the store regions on the columns, the names of product families on the rows, and the profit displayed in the cells:

```
select

[Stores].[Region].Members on columns,

[Products].{Product Family].Members on rows

from SalesCube

where ([Measures].[Profit])
```

Measure

A numeric value stored in a fact table and in an OLAP cube.
Sales Count, Sales Price, Cost, Discount, and Profit could all be measures.

Member

One of the data points for a level of a hierarchy of a dimension.
Some of the members of the Month level of the Time dimension are January, February, March, and April.

Member Property

An attribute of a level that is available for OLAP querying.

Meta Data

Also Known As: Meta data or Meta-data
Data that describes the structure or content of data.
Metadata includes the following:

- A description of tables and fields in a data warehouse, including data types and the range of acceptable values.
- A similar description of tables and fields in the source databases, with a mapping of fields from the source to the warehouse.
- A description of how the data has been transformed, including formulae, formatting, currency conversion, and time aggregation.
- Any other data that is needed to support and manage the operation of a data warehouse or operational system.

Metric

See measure

Migration

The movement of data, applications or database between separate environments, applications or databases.

Model

A business presentation of the structure of the data from one or more databases. A model describes data objects, structure, and grouping, as well as relationships and security.
A model, called a design model, is created and maintained in Framework Manager. The design model or a subset of the design model must be published to the ReportNet server as a package for users to create and run reports.

Multidimensional Analysis

Also Known As: OLAP (On-Line Analytical Processing)

The process of analysis that involves organizing and summarizing data in a multiple number of dimensions.

A spreadsheet is a two-dimensional analysis tool if you regard the rows and columns as dimensions. It could be regarded as three dimensional if you also used sheets.

Multidimensional data is also organized hierarchically, allowing users to "drill down" for more detailed information, "drill up" to see a broader, more summarized view, and "slice and dice" to dynamically change the combinations of dimensions that are being viewed.

Multidimensional Cube

Also known as Cube or Powercube

Multidimensional Online Analytical Programming (MOLAP)

OLAP that stores data and aggregations in a multidimensional database structures.

Multidimensional Schema

See Star Schema

Namespace

In security, a collection of user accounts and user groups from an authentication provider.
In XML, namespaces uniquely identify element types and attributes. An XML namespace is defined by a URI (Uniform Resource Identifier) whose purpose is to name the namespace, not necessarily to identify a location from which to obtain information.
In Framework Manager, namespaces uniquely identify query items, query subjects and so on. You import different databases into separate namespaces to avoid duplicate names.

Natural Join

A join in which the redundant logic components generated by the join are removed.

Non-Volatile

Data that does not change.

Data is stable in a data warehouse. More data is added, but data is never removed. This enables management to gain a consistent picture of the business. Non-volatility is one of the original defining characteristics of a data warehouse.

Normalization

The process of organizing data in accordance with the rules of a relational database.

A fully normalized database is usually the most efficient design for an On-Line Transaction Processing System. A data warehouse, with its emphasis on efficient retrieval of data, often benefits from some intentional de-normalization. See the discussion of the Star Schema.

Normalization is a step-by-step process of removing redundancies and dependencies of attributes in a data structure. The condition of the data at completion of each step is described as a "normal form." Thus, when normalizing we talk about data as being in the first normal form, the second normal form, etc. Normalization theory identifies normal forms up to at least the fifth normal form, plus an adjunct form known as Boyce-Codd Normal Form (BCNF). The first three forms are sufficient to meet the needs of warehousing

OLAP

OLAP is the most widely used term for multidimensional analysis software. The term "On-Line Analytical Processing" was developed to distinguish data warehousing activities from "On-Line Transaction Processing" - the use of computers to run the on-going operation of a business.

In its broadest usage the term "OLAP" is used as a synonym of "data warehousing". In a more narrow usage, the term OLAP is used to refer to the tools used for multidimensional Analysis.

OLAP Browser

A tool used for multidimensional (OLAP) browsing such as Cognos PowerPlay, Cognos ReportNet and Cognos Series8

OLAP Services

Business Intelligence tools included with Microsoft SQL Server 7.0.

OLAP Services was extended and renamed as Analysis Services in SQL Server 2000. This service remains as Analysis Services in SQL Server 2005.

OLTP (OnLine Transaction Processing)

A system designed to handle a single row of data at a time such as a booking or sales system. These systems are highly normalized to ensure that as little memory as possible is used and the system is as fast as possible.

Operational Data Store (ODS)

A store of all data from which Data Warehouses and marts can draw data. An Operational Data Store is structured to hold data at the same granularity and structure as source systems. An Operational Data Store allows those developing Data Warehouses and Data Marts the option of 'cherry picking' data rather than having to store all data at the atomic level within the Data Warehouse. An Operational Data Store is often used in conjunction with an Enterprise Application Interface system.

Orphan

An orphan is a category within a hierarchy that does not have a parent category. This terms is most often referred to in Cognos PowerPlay Transformer (The OLAP modeling tool for Cognos PowerPlay)

Overflow

The condition in which a record or a segment cannot be stored in its home because the address is already occupied or full.

Package

A container for models, reports, and so on. Modelers create packages in Framework Manager to publish models to the ReportNet server.

Parallelism

Parallelism is the splitting up and processing of an individual query in parallel. Parallelism is used in Oracle databases after v8.1.

Parsing

The algorithm that translates syntax into meaningful machine instructions. Parsing determines the meaning of statements issued in the data manipulation language.

Partition / Partioning

A segmentation technique in which data is divided into physically different units.

Passport

Information about the authentication of a user. ReportNet creates a passport each time a user logs on. The passport is retained until the session ends, either when the user logs off, or after a specified period of inactivity. A passport is a reference to a set of credentials maintained on the ReportNet server, not the credentials themselves.

Performance

The length of time from the moment a request is issued until the first of the results of the request are received. Performance can be affected by hardware, software, schema design, indexing, report configuration, network speed among others.

Portal

A Web site or page that provides a single presentation and a single starting point for a set of information. Also, the Cognos component that runs the Cognos portal site. Cognos Web products may use a Cognos portal or may be integrated with other portals.

Privilege

The right or capability to perform an action.

Product locale

The code or setting that specifies what language and regional settings to use for parts of the product interface, such as menu commands.

Project

A set of models, packages, and related information for administration, and for sharing model information. You create projects in Framework Manager

Prompt

A report element that asks for parameter values before the report is run.

PivotTable Services

The tools for client access to Microsoft's Analysis Services (OLAP Services).

Private Dimension

In Microsoft Analysis Services, a dimension that is restricted in use to one particular cube. Shared (conformed) dimensions are very useful in creating a unified data warehousing structure. You can create a dimension once and use it in several different cubes.

Processing

The term processing can be used as a synonymous term for ‘working’ in that the PC or server is working on a job such as a query or a cube build.

Query

A description of the data to be retrieved from a database. Queries identify what columns or query items to include or show in the result set, and what operations should be performed as the data is retrieved. Queries are expressed in a formal language, such as SQL, or by using ReportNet XML specifications. In ReportNet, you use tools such as Query Studio and Report Studio to create and execute queries.

Query item

A representation of a column of data in a data source. It contains a reference to a database

column, a reference to another query item, or a calculation. Query items may appear in a model or in a report.

Query subject

One of the types of objects inside a model. A query subject can be
• defined as a collection of references to items of other query subjects
• expressed as an SQL expression that represents selected query items, which will be retrieved from objects such as tables, synonyms, views, and so on Query subjects contain query items. Query subjects may be part of folders in the model. The query subject is the basis of a query in Report Studio and in report specifications.

Record

A set of values of data organized by their relation to a common key.

Recovery

The restoration of the database to an original position or condition, often after major damage to the physical medium.

Recursive Statement

A recursive statement is a query that uses a self join (a link back to the original object) to navigate a parent and child relationship between rows within the same table. In procedural languages a loop would normally be used.

Redundant relationships

Redundant relationships are those that return the same result set. For example, query subject A is joined to query subject B both directly and through query subject C. Because both of these relationships return the same result set, you do not need them both. While this is not a problem in terms of query results, you may want to simplify the project by deleting the redundant relationships.
• non-redundant relationships
Non-redundant relationships are those that return different result sets. Unintended result sets may be returned if the query chooses a different path than the one intended by the user, for example, the path between query subject A and query subject B may return a different result set than the path that includes query subject C.

Referential Integrity

the facility of a DBMS or Data Warehouse application to ensure the validity of a predefined relationship. Most often used in a data warehouse to ensure that each fact and dimension entry has a corresponding primary and foreign key.

Refresh

The process of updating a data warehouse. A refresh is often used to describe a process which completely clears a data warehouse or table and recreates values within it based upon more recent source data.

Relational Database Management System (RDBMS)

A Database Management System based on relational theory.
Most modern Database Management Systems (Oracle, Sybase, Microsoft SQL Server) are relational databases. These databases support a standard language - SQL (Structured Query Language).

Relational On-Line Analytical Processing (ROLAP)

OLAP that stores data and aggregations in a relational database.

Replication

The physical copying of data from one database to another.

In data warehousing replication takes place as data is moved from the on-line transaction processing system into the data warehouse. Replication also takes place if one or more data marts is being populated with the same data from the data warehouse or originating source system.

Heterogeneous replication occurs when the source and the target database are not the same database management system.

Bi-directional replication is the ability to copy data in both directions between two databases.

In synchronization all the data stored in the database is replicated.

Relationship

A connection that explains how the data in one table relates to the data in another. When you create a relationship, you define the cardinality of each end of the relationship.
For example, a one-to-many relationship between table A and table B means that for each row in table A, there can be 1 or more row matches in table B. Therefore, table B has a many-to-one relationship with table A.
See also join, and cardinality.

Report specification

An XML representation of the queries, prompts, layouts, and styles in a report. You create report
specifications by using Cognos Report Studio or Query Studio, or by writing your own report specifications in XML.

Response file

An ASCII file that contains the setup data that is needed to automate an installation. In Cognos
ReportNet, the response file automates only the process of copying files with the default configuration.

Role

A special group that users can choose when they log on to change what groups they are authenticated as members of, so that they change what data they have authorization for.

Roll Up

See Aggregation

Round trip safety

A process in which Japanese characters can be converted between their Shift-JIS data encoding to Unicode and back to Shift-JIS encoding. ReportNet contains a utility to help you customizeround-trip encoding of Japanese characters.

Scale, Scalable, and Scalability

Pertains to the ability of a computer system, database or reporting system to operate efficiently with larger quantities of data or to be increased in size to handle larger requests, more request or to satisfy requests faster.

Scalability is often discussed in situations when multiple processors are joined together. The system scales well (or is scalable) if doubling the number of processors also doubles the speed at which the system performs its tasks. The extra work involved in coordinating larger systems usually prevents them from being fully scalable - so that going from one to two processors would increase the total speed by less than a factor of two.

Schema

The logical organization and representation of data in a database.

Security provider

A service that authenticates and defines users and the groups of which they are members.

Session

The time during which an authenticated user is logged on to ReportNet.

Shared Dimension

See conformed dimension
In Microsoft Analysis Services, a dimension used by more than one cube.
In general, a dimension that is used by more than one cube is called a conformed dimension.

Signon

A user ID and password, and sometimes other information that is required for a user to access a database

Slice, Slicer, Slicing

The limiting of a cellset to data for a single member from a particular dimension.

Slicing in MDX is similar to filtering in a relational database.

Slice and Dice

The ability to move between different combinations of dimensions when viewing data with an OLAP browser such as Cognos PowerPlay or Analysis Studio.

Multidimensional analysis tools organize the data in two primary ways: in multiple dimensions and in hierarchies. Slicing and dicing refers to the ability to combine and re-combine the dimensions to see different slices of the information on the fly (instantly).

Slowly Changing Dimensions (SCD)

A dimension that has levels or attributes that are changing on an occasional basis.
SEE Changing Dimensions.

SQL (Structured Query Language)

The standard language for accessing relational databases.
This language differs between Oracle and SQL Server but the main structure and methods remain the same.

Snow flaking / Snowflake

Normalization applied to the dimension tables of a star schema.

The star schema is a very simple database design, which clearly presents the multidimensional character of the data and allows for rapid querying of the data in a data warehouse. In snow flaking, some of the fields of the dimension tables are split off into separate tables. This achieves a higher level of normalization, but makes the database design more complex and can reduce the performance and ease of use for Business Intelligence Tools. By snow flaking a star schema the schema becomes closer in structure to an OLTP database.

Sparsity and Density, Sparse and Dense

The degree to which the cells of a cube are filled with data.

One of the primary challenges of storing multidimensional data is the degree of sparsity that is often encountered. When many dimensions are considered with a fine grain of detail, most of the cells will be empty.

It is not uncommon for large cubes to have data in fewer than one in a million cells. Expressed numerically, that cube would have a density of less than .0001%.

Star Schema (Business Definition)

A method of organizing information in a data warehouse that allows the business information to be viewed from many perspectives.

The star is a picture of the way the data is being stored. The basic factual information is in the middle of the star. The points of the star represent various perspectives from which the factual information can be viewed. In Figure 1, the Shipments can be grouped and viewed from the perspective of any of the five points of the star.

Star Schema (Technical Definition)

A database design that consists of a fact table and one or more dimension tables. Each of the dimension tables has a single field primary key which has a one-to-many relationship with a foreign key in the fact table. The star schema is an intentional simplification of the database design that would be achieved by following the standard rules of normalization. The dimension tables are often flattened, to allow for more efficient querying (see 'snow flaking').

Stitch Query

A query that joins two or more resultant data sets from separate queries as though they are the result of one query. Oracle: UNION / MINUS / UNION ALL / INTERSECT

Eg

```
  SELECT
        CUSTOMER_NO,
        GROSS_PROFIT,
        GROSS_SALE,
        TRANSACTION_DATE
  FROM D_FAC_TRANSACTIONS_2002
UNION ALL
  SELECT
        CUSTOMER_NO,
        GROSS_PROFIT,
        GROSS_SALE,
```

```
    TRANSACTION_DATE
FROM D_FAC_TRANSACTIONS_2003
```

Structured Query Language (SQL)

See SQL

Summary

In reporting, summaries are aggregate values that are calculated for all the values of a particular level or dimension. Examples of summaries include total, minimum, maximum, average, and count.

Summary Tables

See aggregation level/layer/table
Tables used to store summarized or aggregated data.

Surrogate Key

A system-generated artificial primary key value, which allows the maintenance of historical records in the Data Warehouse more effectively.

Synchronization

The copying of all data in a database replication. This job ensures that data in two or places has the exactly the same structure and content.

Synonym

A database object in which a table is referred to by a new object name. A synonym is often used for the representation of tables form other databases that have been accessed via a dblink.

Table

A database object in which data is stored in columns and rows.

Tablespace

A volume of data blocks in one or more data files in which database objects are housed or located.

Tabular model

An object that you can use to select query items for a report. A tabular model defines a simple list query. In the XML code for report specifications, tabular model is represented by the tabularModel element.

Tabular reference

An object that you use to reference the tabular object of another query within the same report. In report specifications, it is represented by the tabularReference element.

Tabular set

An object that merges rows from two tabular objects, usually tabular models. A tabular set produces a single result set using union, intersect or except (minus) set operators.
In report specifications, tabular set is represented by the tabularSet element.

Tabular SQL

An object that you use to specify a SQL statement that returns a result set. The SQL statement may be expressed by using syntax recognized by the native database, or by using Cognos SQL syntax. In report specifications, tabular SQL is represented by the tabularSQL element.

Time-variant data

Data that is identified with a particular time period.
Time-variant is one of the original defining characteristics of a data warehouse.

Top down methodology

Involves in building a data warehouse first and then building data marts.

Uniform Resource Locator (URL)

The global address of documents and other resources on the World Wide Web. The first part of the address indicates what protocol to use, and the second part specifies the IP address or the domain name where the resource is located. A sample URL is http://www.digitalviper.co.uk/default.html, where http is the protocol, www.mysite.com is the domain name, and default.html is the document.

Universe

The reporting meta data layer for Business Objects Business Intelligence tools

User

An entry that uniquely identifies an account for a person or computer. Users are created and maintained in a third-party security provider, not ReportNet. Users are used for authentication and to record preferences. General information, such as first and last names, passwords, IDs, locales, and email addresses, is stored in the authentication provider. Other information, such as the location of personal folders or preferred formats for viewing reports, is stored in ReportNet.

Virtual Cube

The term used in Microsoft's Analysis Services (OLAP Services) for a cube that is created from portions of one or more base cubes.

Virtual Dimension

The term used in Microsoft's Analysis Services (OLAP Services) for a dimension that is created from one or more member properties in another dimension.

Virtual Directory

A directory within a website that sources its contents from a file directory on the underlying server. This term is used within Microsoft IIS.

Widow category

(see orphan category)

XML

(eXtensible Markup Language)
A method of sharing data between disparate data systems, without needing a direct connection between them. XML is often distributed using SOAP packets. It should be noted

that due to the volume of meta data within an XML file, these files are condisered unsuitable for high volumes of data such as sales or stock.

XML for Analysis Services

An XML schema that can be used to communicate with a Microsoft Analysis Server.

Copyright and Use