Glossary

**Ad Hoc Query:**
Any query that cannot be determined prior to the moment the query is issued. A query that consists of dynamically constructed SQL, which is usually constructed by desktop-resident query tools.

**Aggregation:**
It is the process of summarizing or combining data to answer a query.

**Bulk Data Transfer**
A software-based mechanism designed to move large data files. It supports compression, blocking and buffering to optimize transfer times.

**Central Data Warehouse**
A database created from operational extracts that adheres to a single, consistent, enterprise data model to ensure consistency of decision-support data across the corporation. It is a style of computing where all the information systems are located and managed from a single physical location.

**Change Data Capture**
The process of capturing changes made to a production data source. Change data capture is typically performed by reading the source DBMS log. It consolidates units of work, ensures data is synchronized with the original source, and reduces data volume in a data warehousing environment.

**Data Cleansing**
It is the process of cleaning or removing errors, redundancies and inconsistencies in the data that is being imported into a data mart or data warehouse. It is part of the quality assurance process.

**Data Mart**
Data mart has a structure similar to a data warehouse, but is typically smaller and is focused on a more limited area. Multiple, integrated data marts are sometimes referred to as an Integrated Data Warehouse. Data marts may be used in place of a larger data warehouse or in conjunction with it. They are typically less expensive to develop and faster to deploy and are therefore becoming more popular with smaller organizations.

**Data Migration**
This term represents the transfer of data from one platform to another. This may include conversion from existing file structure and/or operating environment to another.

**Data Mining**
The process of researching data marts and data warehouses to detect specific patterns in the data sets. Data mining may be performed on databases and multi-dimensional data cubes with ad hoc query tools and OLAP software. The queries and reports are typically designed to answer specific questions of interest to uncover trends or hidden relationships in the data.

**Extraction, Transformation and Loading (ETL) Tool**
Software that is used to extract data from a data source like a operational system or data warehouse, modify the data and then load it into a data mart, data warehouse or multi-dimensional data cube.

**Legacy System**
Older systems developed on platforms that tend to be one or more generations behind the current state-of-the-art applications. Data marts and warehouses were developed in large part due to the difficulty in extracting data from these system and the inconsistencies and incompatibilities among them.
**Meta Data**
Information in a data mart or warehouse that describes the tables, fields, data types, attributes and other objects in the data warehouse and how they map to their data sources. Meta data is contained in database catalogs and data dictionaries.

**Non-Volatile Data**
Data that is static or that does not change. In transaction processing systems the data is updated on a continual regular basis. In a data warehouse the database is added to or appended, but the existing data seldom changes.

**Normalization**
It is the process of eliminating duplicate information in a database by creating a separate table that stores the redundant information.

**Online Analytical Processing (OLAP)**
The process employed by multi-dimensional analysis software to analyze the data resident in data cubes. There are different types of OLAP systems named for the type of database employed to create them and the data structures produced.

**Open Database Connectivity (ODBC)**
It is a database standard developed by Microsoft and the SQL Access Group Consortium that defines the “rules” for accessing or retrieving data from a database.

**Structured Query Language (SQL)**
A standard programming language used by contemporary relational database management systems.

**Data Quality Assurance**
It is the process of checking the quality of the data being imported into the data warehouse.

**Schema**
It is the logical organization of data in a database.

**XML (eXtensible Markup Language)**
It is a method of sharing data between disparate data systems, without needing a direct connection between them.

**Staging area**
It is a place where data in transit is placed, usually coming from the legacy environment prior to entering the ETL layer of processing.

**Operational data**
Data used to support the daily processing a company does.

**Operational data store (ODS)**
A hybrid structure designed to support both operational transaction processing and analytical processing.

**Index**
It is the portion of the storage structure maintained to provide efficient access to a record when its index key item is known.

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

**Snapshot**
A database dump or the archiving of data, from a database at some moment of time.