TECHNOLOGIES OF ERP

DATA WAREHOUSE

Data Warehouse Introduction

A data warehouse is a collection of data marts representing historical data from different operations in the company. This data is stored in a structure optimized for querying and data analysis as a data warehouse. Table design, dimensions and organization should be consistent throughout a data warehouse so that reports or queries across the data warehouse are consistent. A data warehouse can also be viewed as a database for historical data from different functions within acompany.

The term Data Warehouse was coined by Bill Inmon in 1990, which he defined in the following way: "A warehouse is a subject-oriented, integrated, time-variant and non- volatile collection of data in support of management's decision making process".

Other important terminology

Enterprise Data warehouse: It collects all information about subjects (*customers, products, sales, assets, personnel*) that span the entire organization

Data Mart: Departmental subsets that focus on selected subjects. A data mart is a segment of a data warehouse that can provide data for reporting and analysis on a section, unit, department or operation in the company, e.g. sales, payroll, production.

Decision Support System (DSS) : Information technology to help the knowledge worker (executive, manager, and analyst) makes faster & betterdecisions

Drill-down: Traversing the summarization levels from highly summarized data to the underlying current or old detail

Metadata: Data about data. Containing location and description of

warehouse system components: names, definition, structure…

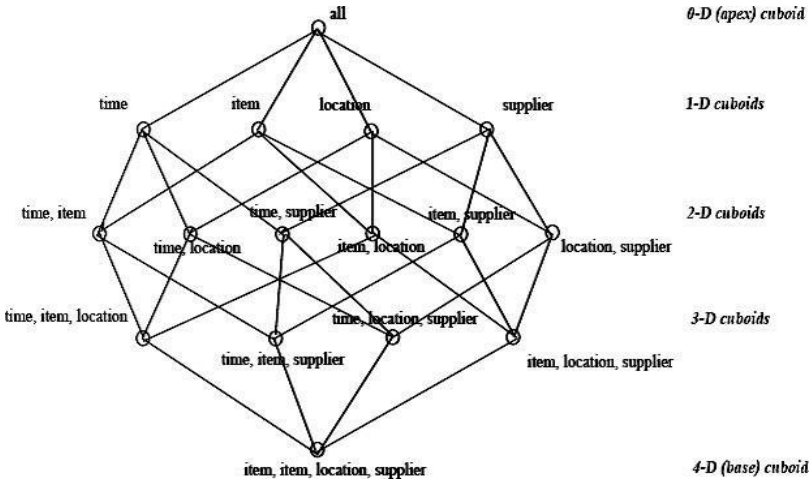
Data Warehouse Characteristics

A data warehouse can be viewed as an information system with the following attributes:

– It is a database designed for analyticaltasks--It_s content is periodicallyupdated

Multidimensional Data Model.

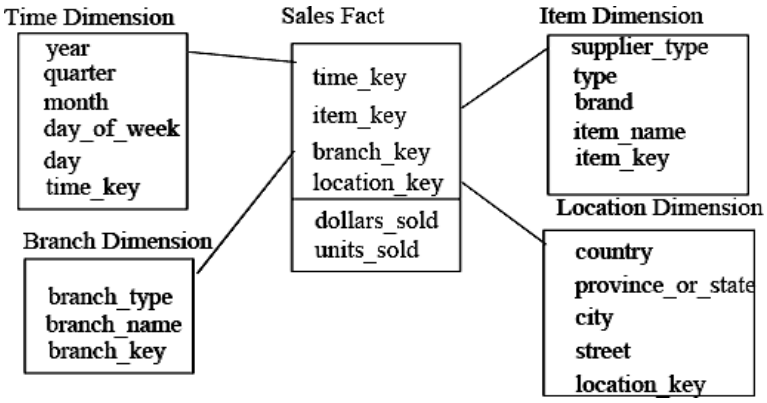
The most popular data model for data warehouses is a multidimensional model. This model can exist in the form of a star schema, a snowflake schema, or a fact constellation schema. Let's have a look at each of these schema types.



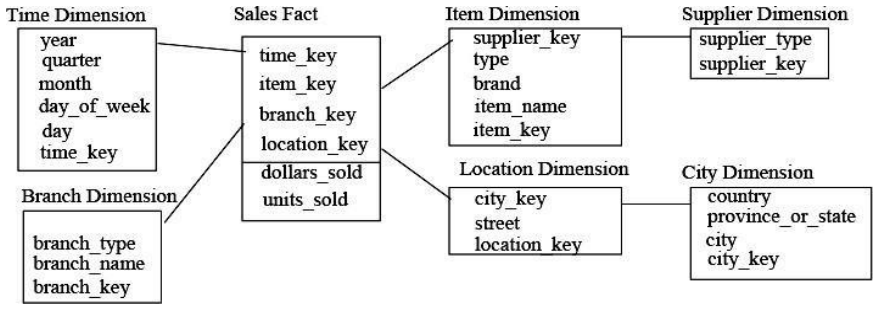
Stars,Snowflakes,and Fact Constellations:Schemas for Multidimensional Databases

Star schema: The star schema is a modeling paradigm in which the data warehouse contains (1) a large central table (fact table), and (2) a set of smaller attendant tables (dimension tables), one for each dimension. The schema graph resembles a starburst, with the dimension tables displayed in a radial pattern around the central facttable.

Figure Star schema of a data warehouse for sales.



- **Snowflake schema**: The snowflake schema is a variant of the star schema model, where some dimension tables are normalized, thereby further splitting the data into additional tables. The resulting schema graph forms a shape similar to a snowflake.
    - Figure Snowflake schema of a data warehouse for sales.



Data warehouse architecture-Steps for the Design and Construction

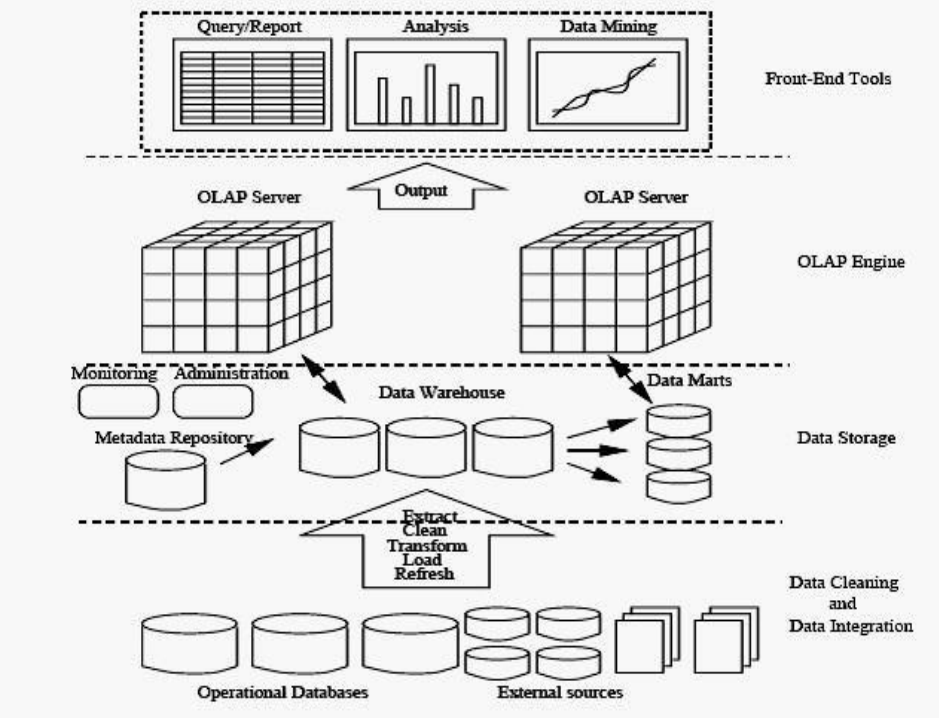
.

The Design of a Data Warehouse: A Business Analysis Framework

Four different views regarding the design of a data warehouse must be considered: the top-down view, the data source view, the data warehouse view, the business query view.

- The top-down view allows the selection of relevant information necessary for the datawarehouse.
- The data source view exposes the information being captured, stored and managed by operationalsystems.
- The data warehouse view includes fact tables and dimensiontables
- Finally the business query view is the Perspective of data in the data warehouse from the viewpoint of the enduser.

Three-tier Data warehouse architecture



The bottom tier is ware-house database server which is almost always a relational database system. The middle tier is an OLAP server which is typically implemented using either a Relational OLAP (ROLAP) model, (2) a Multidimensional OLAP (MOLAP) model. The top tier is a client, which contains query and reporting tools, analysis tools, and/or data mining tools

(e.g., trend analysis, prediction, and soon).

From the architecture point of view, there are three data warehouse models: the enterprise warehouse, the data mart, and the virtual warehouse.
Enterprise warehouse: An enterprise warehouse collects all of the information about subjects spanning the entire organization. It provides corporate-wide data integration, usually from one or more operational systems or external information providers, and is cross-functional in scope. It typically contains detailed data as well as summarized data, and can range in size from a few gigabytes to hundreds of gigabytes, terabytes, or beyond.
Data mart: A data mart contains a subset of corporate-wide data that is of value to a specific group of users. The scope is connected to specific, selected subjects. For example, a marketing data mart may connect its subjects to

customer, item, and sales. The data contained in data marts tend to be summarized. Depending on the source of data, data marts can be categorized into the following twoclasses:

Virtual warehouse: A virtual warehouse is a set of views over operational databases. For efficient query processing, only some of the possible summary views may be materialized. A virtual warehouse is easy to build but requires excess capacity on operational databaseservers.

Figure: A recommended approach for data warehouse development

Data warehouse Back-End Tools and Utilities

The ETL (Extract Transformation Load) process-In this section we will discussed aboutthe 4 major process of the data warehouse.

EXTRACT

Some of the data elements in the operational database can be reasonably be expected to be useful in the decision making, but others are of less value for that purpose. For this reason, it is necessary to extract the relevant data from the operational database before bringing into the data warehouse. Many commercial tools are available to help with the extraction process. **Data Junction** is one of the commercial products.

TRANSFORM- Transformation process deals with rectifying any inconsistency (ifany).

One of the most common transformation issues is _Attribute Naming Inconsistency‘. It is common for the given data element to be referred to by different data names in different databases. Employee Name may be EMP_NAME in one database, ENAME in the other. Thus one set of Data Names are picked and used consistently in the data warehouse. Once all the data elements have right names, they must be converted to common formats. The conversion may encompass the following:

- Characters must be converted ASCII to EBCDIC or viseversa.
- Mixed Text may be converted to all uppercase forconsistency.
- Numerical data must be converted in to a commonformat.
- Data Format has to bestandardized.
- Measurement may have to convert. (Rs/$)
- Coded data (Male/ Female, M/F) must be converted into a commonformat.

CLEANSING

Information quality is the key consideration in determining the value of the

information. The developer of the data warehouse is not usually in a position to change the quality of its underlying historic data, though a data warehousing project can put spotlight on the data quality issues and lead to improvements for the future. It is, therefore, usually necessary to go through the data entered into the data warehouse and make it as error free as possible. This process is known as **DataCleansing**.

LOADING--Loading often implies physical movement of the data from the computer(s) storing the source database(s) to that which will store the data warehouse database, assuming it is different. This takes place immediately after the extraction phase. The most common channel for data movement is a high-speed communication link. Ex: Oracle Warehouse Builder is the API from Oracle, which provides the features to perform the ETL task on Oracle DataWarehouse.

Benefits of a Data Warehouse

1. **Delivers enhanced business intelligence**

By having access to information from various sources from a single platfor like a business's processes, forinstance, market segmentation, sales, risk, inventory, and financial management.

2. **Saves times**

A data warehouse standardizes, preserves, and stores data from distinct sources, aiding the consolidation and integration of all the data. Since critical data is available to all users, it allows them to make informed decisions on key aspects. In addition, executives can query the data themselves with little to no IT support, saving more time and money.

3.Enhances data quality and consistency

A data warehouse converts data from multiple sources into a consistent format. Since the data from across the organization is standardized, each department will produce results that are consistent. This will lead to more accurate data, which will become the basis for solid decisions.

4. Generates a high Return on Investment (ROI)

Companies experience higher revenues and cost savings than those that haven't invested in a data warehouse.

5.Provides competitive advantage

Data warehouses help get a holistic view of their current standing and evaluate opportunities and risks, thus providing companies with a competitive advantage.

6.Improves the decision-making process

Data warehousing provides better insights to decision makers by maintaining a cohesive database of current and historical data. By transforming data into purposeful information, decision makers can perform more functional, precise, and reliable analysis and create more useful reports with ease.

7.Enables organizations to forecast with confidence

Data professionals can analyze business data to make market forecasts, identify potential KPIs, and gauge predicated results, allowing key personnel to plan accordingly.

8.Streamlines the flow of information

Data warehousing facilitates the flow of information through a network connecting all related or non-related parties.

- Data warehouses are designed to perform well with aggregate queries running on large amounts ofdata.
- The structure of data warehouses is easier for end users to navigate, understand and query against unlike the relational databases primarily designed to handle lots oftransactions.
- Data warehouses enable queries that cut across different segments of a company's operation. E.g. production data

could be compared against inventory data even if they were originally stored in different databases with different structures.

- Queries that would be complex in very normalized databases could be easier to build and maintain in data warehouses, decreasing the workload on transaction systems.
- Data warehousing is an efficient way to manage and report on data that is from a variety of sources, non uniform and scattered throughout acompany.
- Data warehousing is an efficient way to manage demand for lots of information from lots ofusers.
- Data warehousing provides the capabilityto analyze large amounts of historical
- data for nuggets of wisdom that can provide an organization with competitive advantage.