

Data warehouse

Ritika

Abstract: Data warehouse is the basically a process where we collect data from all the available sources whether it is online or offline, and to access and analyze data which is further used for various purposes. It can use either top down or bottom up approach. Different layers are available to perform various functions during the process. In data warehousing we transform, clean, extract, load and refresh data.

Keywords: Analyze, top down, bottom up, extract, transform.

Introduction

The term data warehouse was coined by William H. Inmon, who is known as the Father of Data Warehousing. Definition by Ralph Kimball “**A data warehouse is a copy of transaction data specifically structured for query and analysis.**” A data warehouse is basically collection of data from different sources to access and analysis of data, it supports queries and data making. It involves data integration and data cleaning data etc. It uses two approach i.e. top down and bottom up. It uses mainframe server. It also uses data from various online transactions processing applications (OLTP).

Data warehouse architecture

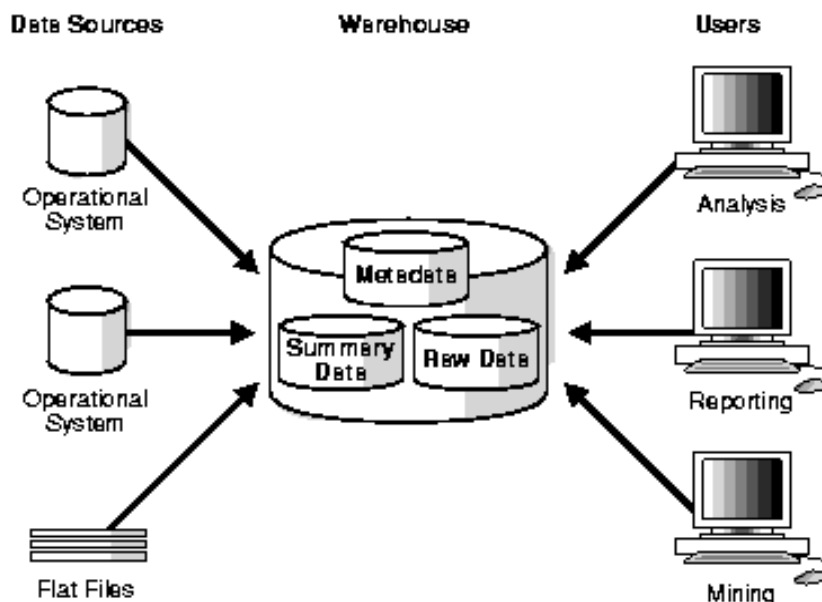
Data warehouses and their architectures vary depending upon the specifics of an organization's situation. Three common architectures are:

- Data Warehouse Architecture (Basic)
- Data Warehouse Architecture (with a Staging Area)
- Data Warehouse Architecture (with a Staging Area and Data Marts)

1) Data Warehouse Architecture (Basic)

Figure shows basic data warehouse architecture; here users access the data which is extracted from different sources through data warehouse.

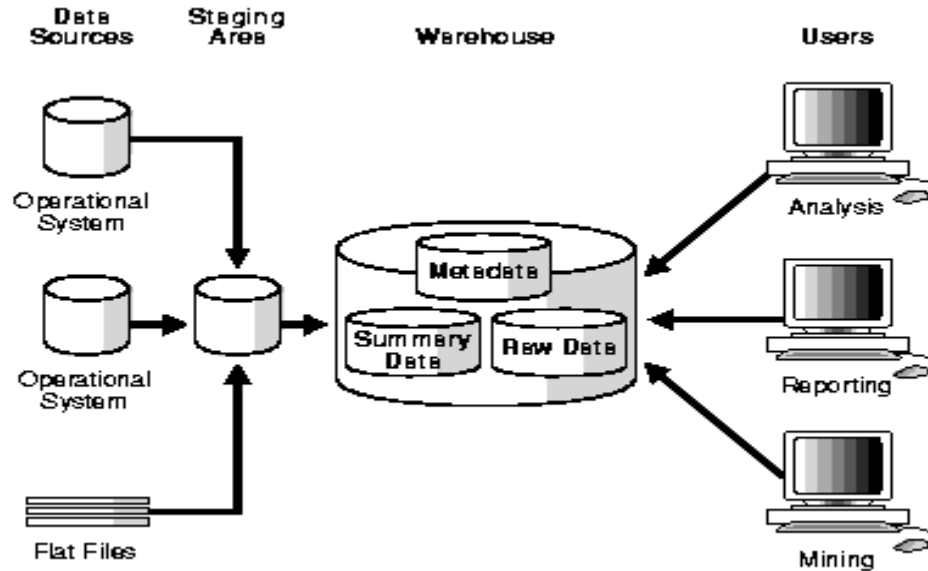
Architecture of a Data Warehouse



2) Data Warehouse Architecture (with a Staging Area)

It process and clean your data before you put your data in warehouse. It can be done using various programs available however data warehouse use stage area instead of using programs. Figure shows this architecture.

Architecture of a Data Warehouse with a Staging Area



3) Data Warehouse Architecture (with a Staging Area and Data Marts)

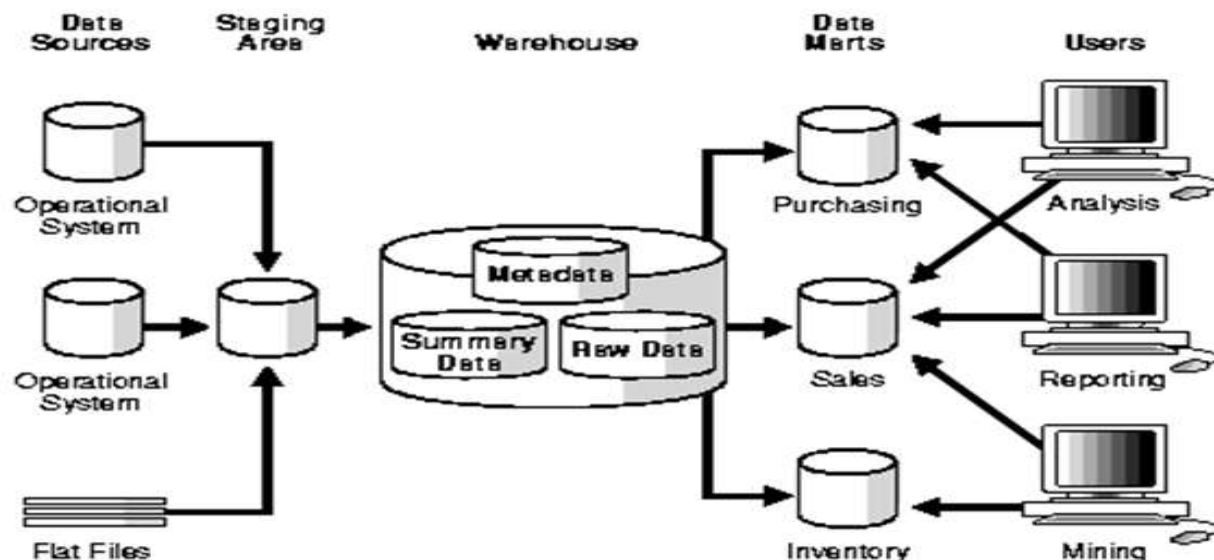
Data warehouse with staging area is quite common but still want to use it for different organization. It can be done by using adding data marts which is designed for particular area.

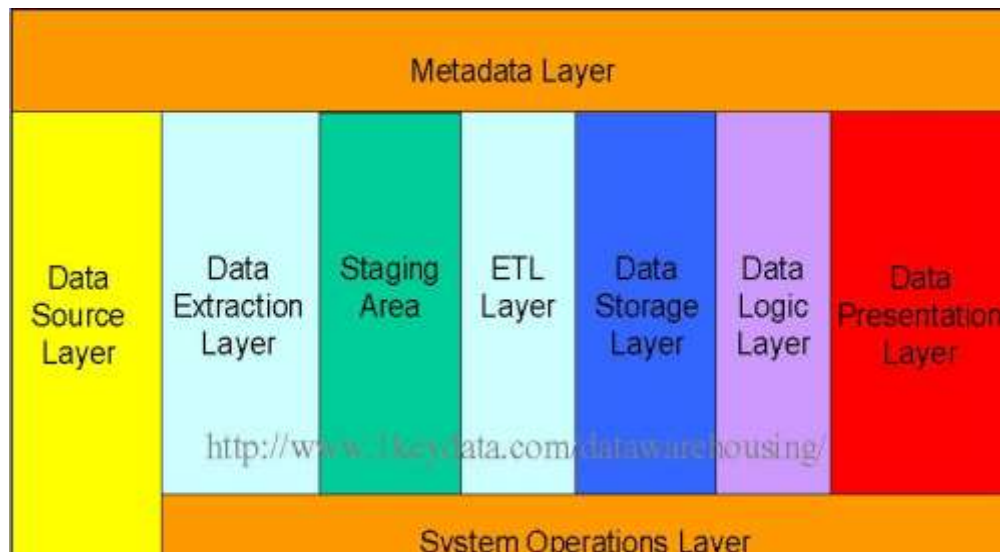
Here a figure shows separation of production, sales and inventories. Here a financial analyst can easily analyze data for different purpose like purchase and sales.

Architecture of a Data Warehouse with a Staging Area and Data Marts

Data Warehouse Layers

Data warehouse consists of various layers and relationship between them is given below





- 1) **Data Source Layer:** - this layer extracts data from different sources and put data into ware house. It can be in any format for e.g. plain text, any other type of database, excel file, relational database.
- 2) **Data Extraction Layer:** - in this layer it extracts data from data source layer into data warehouse. It also does some data cleaning before storing it.
- 3) **Staging Area:** - Before any manipulation and transformation of data into data warehouse here data is stored. So it is easier to perform any operation if it is present a one place.
- 4) **ETL Layer:** - In this layer data is transformed and its cleaning is done here. It is most time consuming phase and also various tools are used in this layer.
- 5) **Data Storage Layer:** - here the transformed and cleaned data is stored.
- 6) **Data Logic Layer:** - here the various rules are applied and business rules are stored here.
- 7) **Data Presentation Layer:** - this is the layer where that information is stored which reaches users. This can be in any form like table form, email form or can be in graphical report in browser or an alert which warns users of exceptions.
- 8) **Metadata Layer:** - this is where the information is stored where the data is stored. A tool is used which is metadata tool used to manage metadata.
- 9) **System Operations Layer:** - This layer stores the information about how the data warehouse operates system performance and user access history.

Functions of Data Warehouse Tools and Utilities

- **Data Extraction** – it includes how the data is collected from different sources.
- **Data Cleaning** – it includes finding and correcting errors in data.
- **Data Transformation** – in includes conversion of data from the format from it is extracted to the one which actually stores the data i.e. data warehouse format.
- **Data Loading** – it includes sorting, merging, maintaining integrity and making partitions.
- **Refreshing** – It involves updating data from different sources to warehouse.

Why Data Warehouses Fail

- 1. Focusing On Ideology Rather Than Practicality:** There many theoretical courses available for data warehouse but these courses don't offer practical knowledge
- 2. Making the Process Unnecessarily Complicated:** It is already a very complex process but different companies make it even more complex.
- 3. Lack of Clear Ownership:** - It is such a big project it can be difficult to handle by single person, so the team is divided into groups due to this there are certain problems.
- 4. Not Understanding Proper Protocol:** It doesn't matter in which firm you work you need to understand the protocol of the organization to build successful warehouse.
- 5. Not Fully Understand Project Impact before the Project Starts:** if you don't understand the problem before working on it, it cannot become successful.
- 6. Blindly Sticking To Certain Standards:** Same things and same procedures are applied to both front end and back end or same procedures for different applications leads to failure.
- 7. Bad Project Management:** If any project whether it is big or small it will not be successful if it is not managed properly.

Conclusion

Data warehouse is used to collect data to process and analyze it. Different software's are available to perform these operations which work so efficiently that it works better than the expert. In data ware house it extract data, clean data, transform data. Different layers i.e. Data Source Layer, Data Extraction Layer, Staging Area, ETL Layer, Data Storage Layer, Data Logic Layer, Data Presentation Layer, Metadata Layer, System Operations Layer are used to perform different operations. Different architectures are used where it deals with different problems in different ways.

References

- [1]. <http://www.1keydata.com/datawarehousing/datawarehouse.html>
- [2]. <http://www.tutorialspoint.com/dwh/>
- [3]. http://docs.oracle.com/cd/B10500_01/server.920/a96520/concept.htm
- [4]. <http://datawarehouse4u.info/>
- [5]. <http://searchsqlserver.techtarget.com/definition/data-warehouse>