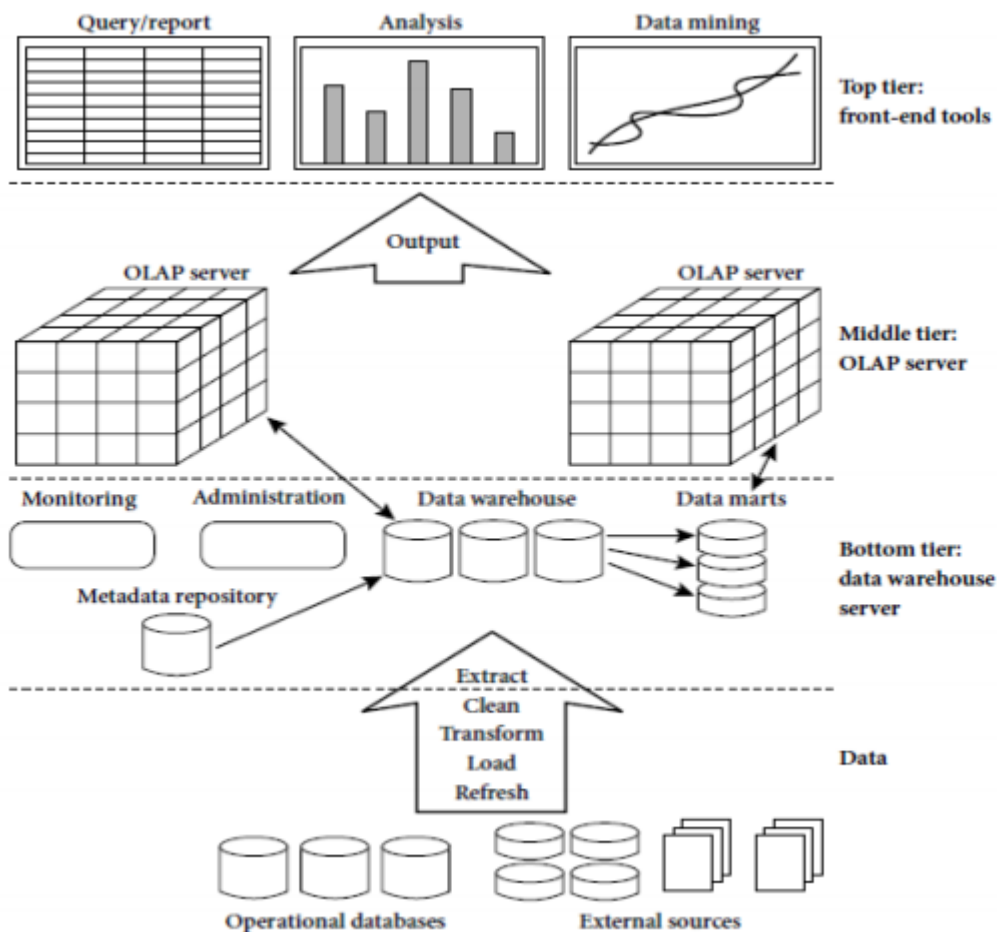


Q.1	Define Data warehouse?
	<ul style="list-style-type: none"> – A decision support database that is maintained separately from the organization’s operational database – A consistent database source that bring together information from multiple sources for decision support queries – Support information processing by providing a solid platform of consolidated, historical data for analysis
Q.2	What is Data warehouse? Explain.
	<ol style="list-style-type: none"> 1. DW is a store of information organized in a unified data model. 2. Data collected from a number of different sources Finance, billing, website logs, personnel, 3. Purpose of a data warehouse (DW): support decision making 4. Easy to perform advanced analysis <ul style="list-style-type: none"> • Solution: new analysis environment (DW) where data are <p>Subject oriented (versus function oriented)</p> <p>Integrated (logically and physically)</p> <p>Time variant (data can always be related to time)</p> <p>Stable (data not deleted, several versions)</p> <p>Supporting management decisions (different organization)</p> <ul style="list-style-type: none"> • Data from the operational systems are <p>Extracted</p> <p>Cleansed</p> <p>Transformed</p> <p>Aggregated</p> <p>Loaded into the DW</p> <ul style="list-style-type: none"> • A good DW is a prerequisite for successful BI
Q.3	Explain A Three Tier Data Warehouse Architecture



Tier-1: The bottom tier is a warehouse database server that is almost always a relational database system. Back-end tools and utilities are used to feed data into the bottom tier from operational databases or other external sources (such as customer profile information provided by external consultants). These tools and utilities perform data extraction, cleaning, and transformation (e.g., to merge similar data from different sources into a unified format), as well as load and refresh functions to update the data warehouse. The data are extracted using application program interfaces known as gateways. A gateway is supported by the underlying DBMS and allows client programs to generate SQL code to be executed at a server. Examples of gateways include ODBC (Open Database Connection) and OLEDB (Open Linking and Embedding for Databases) by Microsoft and JDBC (Java Database Connection). This tier also contains a metadata repository, which stores information about the data warehouse and its contents.

Tier-2: The middle tier is an OLAP server that is typically implemented using either a relational OLAP (ROLAP) model or a multidimensional OLAP. OLAP model is an extended relational DBMS that maps operations on multidimensional data to standard relational operations. A multidimensional OLAP (MOLAP) model, that is, a special-purpose server that directly implements multidimensional data and operations.

Tier-3: The top tier is a front-end client layer, which contains query and reporting tools, analysis tools, and/or data mining tools (e.g., trend analysis, prediction, and so on).

Q.4	Data Warehouse Models
	There are three data warehouse models

1. 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.
- An enterprise data warehouse may be implemented on traditional mainframes, computer superservers, or parallel architecture platforms. It requires extensive business modeling and may take years to design and build.

2. 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 confined to specific selected subjects. For example, a marketing data mart may confine its subjects to customer, item, and sales. The data contained in data marts tend to be summarized.
- Data marts are usually implemented on low-cost departmental servers that are UNIX/LINUX- or Windows-based. The implementation cycle of a data mart is more likely to be measured in weeks rather than months or years. However, it may involve complex integration in the long run if its design and planning were not enterprise-wide.
- Depending on the source of data, data marts can be categorized as independent or dependent. Independent data marts are sourced from data captured from one or more operational systems or external information providers, or from data generated locally within a particular department or geographic area. Dependent data marts are sourced directly from enterprise data warehouses.

3. 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 database servers.