

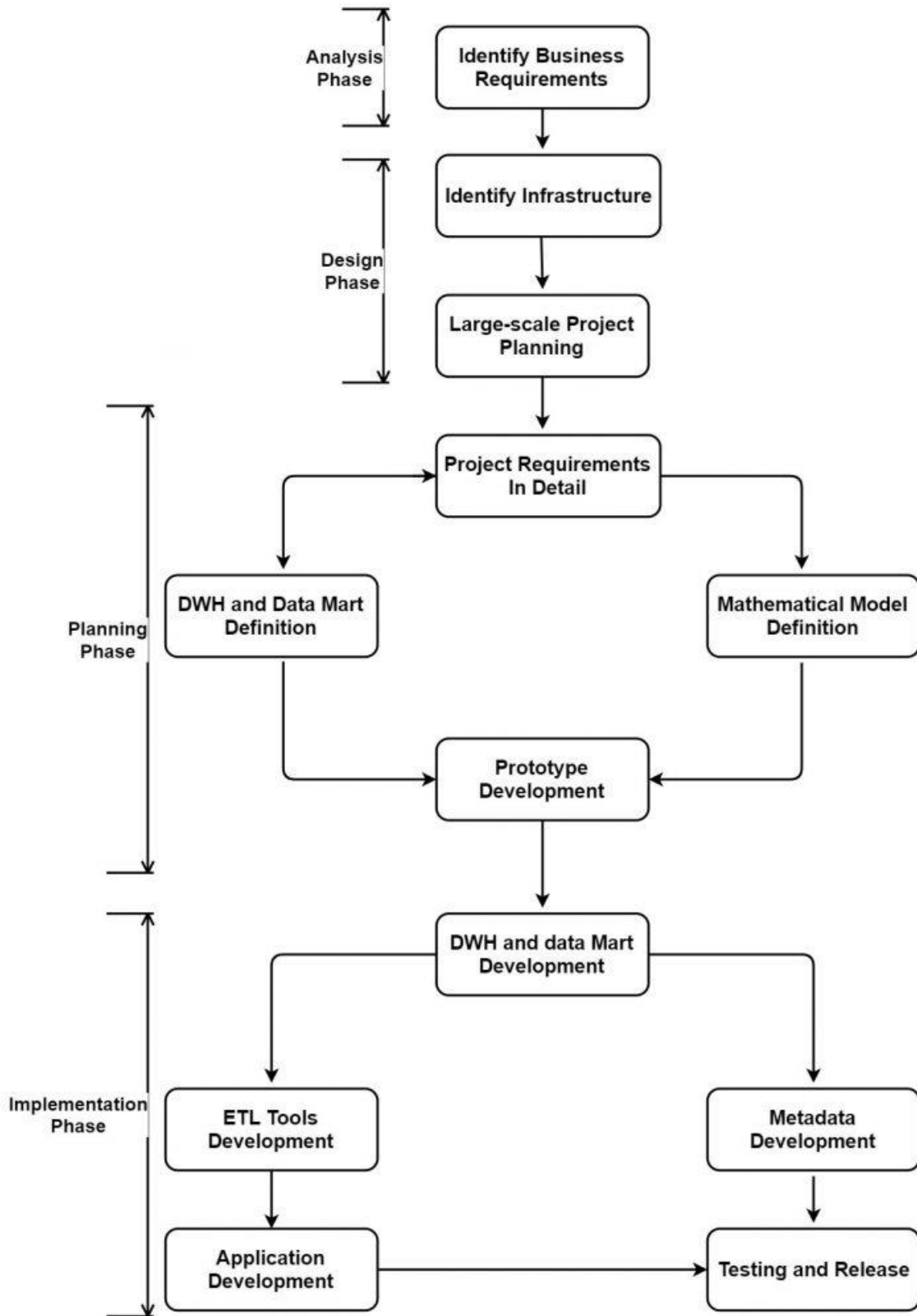
BUISINESS ANALYTICS AND INTELLIGENCE

SOLVED QUESTIONS

Q1) Describe business intelligence system development phases? (6 marks)

The 4 phases of BI system development are as follow:

- Analysis phase
- Design phase
- Planning phase
- Implementation phase



Let discuss this one by one in detail:

- Analysis phase: The first step is requirement gathering which involves recognition of requirements. Requirement gathering is a most important phase of Business Intelligence system. We can get the clear view of organization workflow through working employee and staff. The overall cost and benefits are cleared in this phase.
- Design phase: This is the second phase of Business Intelligence system development. The whole development plan is decided in this phase. This phase has two sub-phases which are
- Identification of required infrastructure: The existing infrastructure information should be evaluated. The decision-making processes are studied for deciding information requirement.
- Large-scale project planning: For project planning, the project management methodologies are used which recognize the different development stages of the project.
- Planning phase: In the planning phase, different functions such as mathematical model, data warehouse are defined. For implementing the mathematical model, make sure that it is clearly defined and data is available. This phase also verifies the efficiency of the algorithm and whether it solving the particular problem or not. The entire project requirements are clearly described in this phase.
- Implementation Phase: This phase consists of 5 sub-phases:
 1. Data Warehouse and Data Mart Development: In this phase, the Data Warehouse and Data Mart are created which responsible for providing data to the different BI system.
 2. Metadata Development: Metadata can be defined as it is a data that describe other data. In this phase, Metadata is created for the stored data in Data Warehouse.
 3. ETL Tools Development: Function of ETL tool is to extract data from the source, transform it and load it into the data warehouse.
 4. Application Development: This phase is focus towards application development for BI for the organization.
 5. Testing and Release: This is the last phase where software is tested and released to the user.

Q2)What is ad-hoc query analysis? (2 marks)

- An ad-hoc analysis is the discipline of analyzing data on an as needed or requested basis. It allows the user to use reports and analysis as per their need. The word ad-hoc means as the occasion requires.
- The ad-hoc analysis is used to create reports that do not create yet. It can also be used to get more current data for the existing parts covered by a static report.
- Example: If a business has large sales database and one wants a sale outcome related to a specific scenario then by using software application which takes input as the analytical question about business data, one can build a single report which runs once and gives a unique result.

Q3) Name any two standard business intelligence tools available. (2 marks)

The business tools that are available are RDBMS and ETL tools.

- RDBMS:
 1. Netezza: Netezza is developed to decrease the query times and hasten it. It provides a quick and nearly correct answer to the complex business questions.
 2. DB2 Connector: The main function of the DB2 connector is to provide connectivity and metadata integration for other data sources which can be relation database etc.
 3. Sybase IQ: It is column based and relational database system specifically used in business intelligence as well as in data warehouse.
- ETL Tools:
 1. Informatica: It is used for abstraction of data from sources then loading into the target storage. It makes easy to develop the data warehouses as well as data marts.
 2. Data Stage: It contains the collection of windows based graphical tools that are very helpful in the construction of ETL tools. This tool also makes a connection to the data stage server and help in the managing of metadata associated with it.
 3. Ab initio: It has the good graphical user interface which is based on parallel processing tool that used for ETL data management and its analysis. It provides high data potency.
 4. SAS: It mainly provides multithreaded and multiprocessing data extraction that reduces the time for operations and data transfer. It also eliminates the less accurate data that reduce the overall cost.

OR

Q4 Define following terms with examples. (6 marks)

- **Data**
- **Information**
- **Knowledge**

Data:

1. Data is formatted in a specific way and can exist in different forms such as numbers, picture, text, etc. Data is meaningless if it is not processed.
2. Most organization gather data daily in the form of matrices. Data may contain unfiltered or general information.
3. Different activities in BI such as sales transactions, production where data is generated and later it is stored.

Information:

1. Information is anything that is communicated. Information itself is a data that handled in a way which looks meaningful to others who get it.
2. For example, if your boss tells you details about how the sales of your company have decreased in previous 3 months then you can use this information for increasing your sales in upcoming months.

- Knowledge
1. Knowledge can be defined as the information about the particular topic that you get either by studying that topic or by your past experience.
 2. It can be a combination of different facts and statements which are exits true.

Q5 What is BI dashboard. (2marks)

- BI dashboard is BI software interface which used to track KPIs, metrics as well as key data points related to business or a particular process.
- BI dashboards provide smooth and real-time visibility to the user. The dashboard is the first page which user sees when he logs into BI software. We can customize Dashboard look and interface for desktop, mobile users.
- The aim of dashboards is to help business individuals to take knowledgeable decisions by allowing organizations to collect, analyze, develop dashboards and make reports on their valuable and business-driving data.

Q6 What is the purpose of reporting in BI? (2 marks)

- It is the process of providing reports to end users or organization by using BI software. Reporting referred to collecting data from sources and presenting that data to the user in a format which user can understand and analyze.
- BI reporting is preconfigured component into BI software. It helps the organization in strategic decision making as well as short and long-term tactical.

Q7 Explain characteristic of decision support system. (6 marks)

Following are the main characteristic of decision support system:

- Flexibility
- Effectiveness
- Integration in the decision-making process
- Mathematical models
- Organizational role

Let discuss this one by one in detail:

- Flexibility: For accommodating changes a decision support system must need flexibility. Also, it needs to be user-friendly with a good graphical user interface for displaying extracted information. The web browser is normally used for communication with different users.
- Effectiveness: DSS improves the effectiveness of decisions and not the efficiency through which the decision is taken. Decision support system is intended to enhance the accuracy, quality, and effectiveness of set decisions.
- Integration in the decision-making process: Decision support system should help different knowledge workers which may be in semi-structured and unstructured decision processes

in the same domain. Also, decision makers can include their preferences and competencies in DSS.

- Mathematical models: Decision support systems make use of mathematical models for better decision making. The main feature of using mathematical model is that it transforms the data into knowledge.
- Organizational role: Decision support system has the capacity to encourage the interaction between different departments of an organization. In an organization, DSS keep the record of analysis and data that responsible for decision making.

Q8 List types of decision support system and explain anyone in detail. (4 marks)

Types of Decision Support System are

- Communication-driven DSS
- Data-driven DSS
- Document-driven DSS
- Knowledge-driven DSS
- Model drove DSS

Let discuss one of them in detail:

- Communication driven DSS: IT emphasizes communication, collaboration and shared decision making. It enables two people to share information with each other in a group or to individuals. In an organization, it helps in a team meeting by allowing communication between each participant. Client-server technology is used in communication-driven DSS. Example: Messenger, webinar, net-meeting system.

OR

Q9 Define Decision Support System and explain its structure. (6 marks)

- Decision Support System is a computer-based application which collects, analyses business data for better decision making.
- Decision Support System is a set of hardware and software applications which forms the backbone of one's decision-making process. It helps to make perfect decisions. Decision support system supports the decision making. This system can be used by one user or many users. It is location independent which means it is accessible from multiple locations. It is very beneficial for decision makers.
- Structure of Decision Support System:

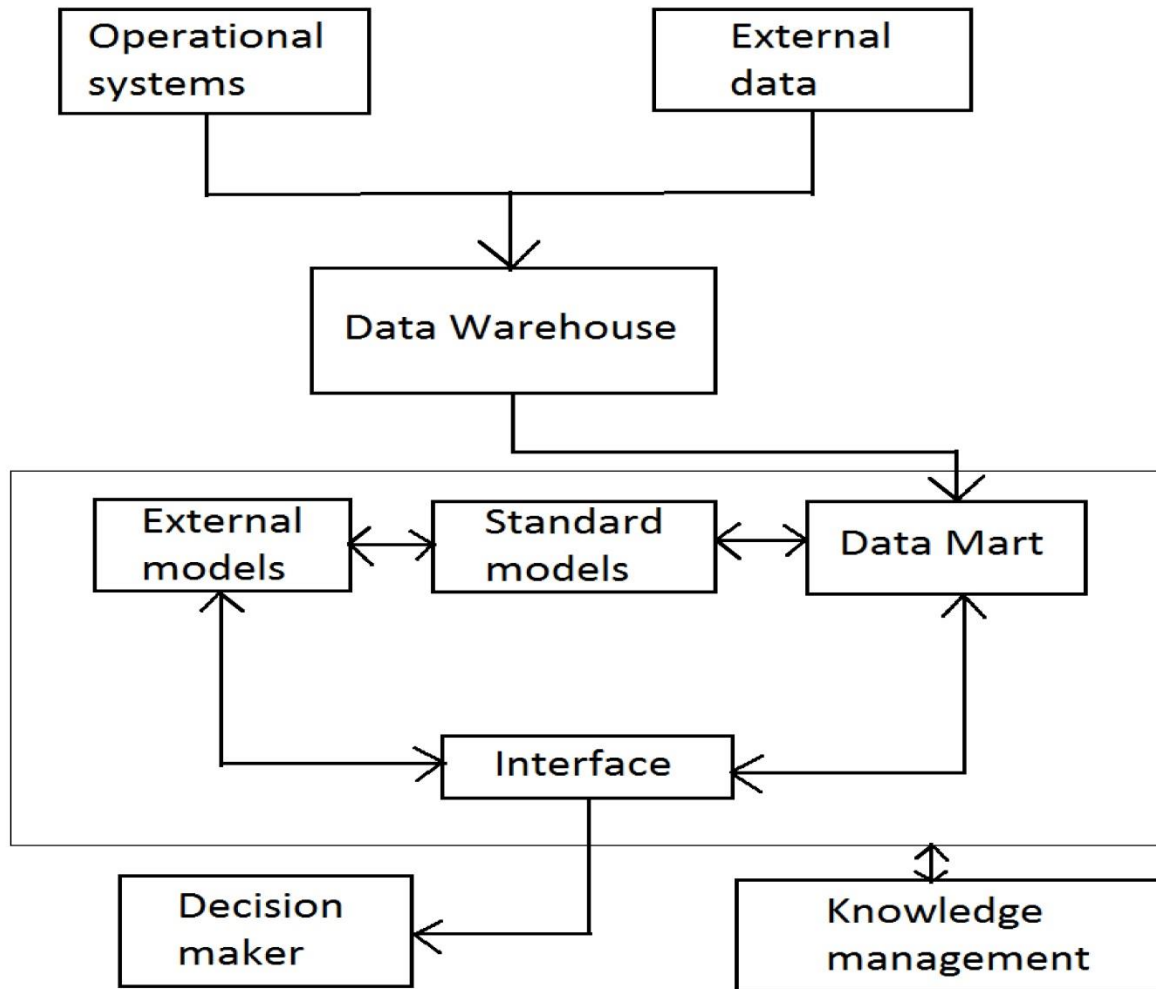


Fig. Structure of Decision Support System

Let discuss all one by one in detail:

- **Data Management:** In this, the data management module is linked with the data warehouse that corresponds to the central data warehouse storage. Data management module contains database and many times data mart act as a database.
- **Model management:** Model management module contains set of mathematical models. It is a simple model and very useful in the decision-making process. It can be integrated with more complex external models. It helps the knowledge workers for the evolution of ad-hoc models.
- **Interaction:** The module which is responsible for the interaction takes its input from users during the graphical user interface of web browser and returns the mined data with knowledge prepared by the system.
- **Knowledge management:** This module allows the decision makers to show different shared knowledge and it is interdependent to the knowledge management integral system.

Q10 What are the advantages and disadvantages of decision support system. (4 marks)

The advantages of Decision Support System are:

- **Improve Interpersonal Communication:** Interpersonal communication include communication between decision makers. Also, it is useful for communication-driven and group of Decision Support System. Model-driven DSS provides a way for sharing facts and analyses. Improved data accessibility is the primary motivation for making data-driven Decision Support System.
- **Increase Decision Maker Satisfaction:** DSS decrease dissatisfaction of decision makers by developing consciousness among decision makers that by use of higher quality data they can make better decisions.
- **The decrease in Cost:** Proofs proved that decision support system is very useful for overall cost reduction. Cost can be reduced by minimizing infrastructure and technology expenses.
- **Save Time:** There is a record of decreased decision cycle time for all department of Decision Support System. With the decrease in time, the employee efficiency is increased.
- **Enhance Decision making Efficiency:** The decision support system enhance the effectiveness of better decision making. The better analysis can improve the decision-making process.
- **Improve Organizational Control:** For performance monitoring and ad-hoc querying, the transaction data made accessible by data-driven decision support system which will enhance the management understanding of business operations.
- **Improve Data Accuracy and Data Sharing:** If the Data of higher quality is available in data source system and developers have the capability to incorporate it in single common data source then managers of organization can take a single version of truth of all operations of the organization. The aim of decision support system is to improve the data accessibility and shareability.

Disadvantages of Decision Support System are:

- **The assumption of Relevance:** If Decision Support System becomes common in organizations, then there is a chance that managers of organization can use them inappropriately. Providing proper Training to the managers is the solution to this problem.
- **False Belief in Objectivity:** It is an overstatement that people who are using DSS are more objective and rational than managers who are not using computerized decision support.
- **Information Overload:** The excessive information load can also be a problem is DSS system.
- **Overemphasize Decision Making:** Implementing DSS can reinforce rational view and overemphasize decision processes and decision making.
- **Status Reduction:** Managers worries about a status reduction of the organization so they prefer clerical work.

- **Transfer of Power:** This is more related with decision automation systems as compared to Decision Support System. It advocates building computerized DSS.
- **Unanticipated Effects:** Some Decision Support System reduces the skill required for taking a decision as well as some might overload decision-makers with data and decrease the effectiveness of decision making.
- **False Belief in Objectivity:** Managers may not be more focus though they are using DSS. It is an exaggeration that people who are using DSS are more objective as well as rational as compared to managers who are not using computerized decision support.

Q11 Explain data warehouse terms ROLAP, MOLAP, HOLAP, DOLAP (6 marks)

- **ROLAP:** Relational Online Analytical Processing (ROLAP) is online analytical processing (OLAP) which analyzes the given data by using multidimensional data models. ROLAP can deal with huge amount of data. It can generate SQL queries. The processing time of ROLAP is slow.
- **MOLAP:** Multidimensional Online Analytical Processing (MOLAP) is online analytical processing (OLAP) which make indexing directly into the multidimensional database. Data stored in the multidimensional array is processed by MOLAP. It provides speed and user responsiveness, therefore, it is preferred by the user. MOLAP provide performance and have the ability to perform complex operations rapidly.
- **HOLAP:** Hybrid Online Analytical (HOLAP) is the combination of both relational Online Analytical Processing (ROLAP) and Multidimensional Online Analytical Processing (MOLAP). HOLAP combine data capacity of ROLAP and processing capability of MOLAP. It stores the data in both relational and multidimensional database and uses whichever is more suited to the desired type of processing.
- **DOLAP:** Desktop Online Analytic Processing (DOLAP) is desktop-based OLAP technology. Data is stored in cubes on the desktop. It is user-friendly and provides good query processing performance. It is easy to deploy as compared to all OLAP approaches.

OR

Q12 Explain term “Subject-oriented” regarding Data Warehouse design with suitable example. (6 marks)

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: A data warehouse can be used to analyze a particular subject area. For example, "sales" can be a particular subject. A data warehouse is subject oriented because it provides information around a subject rather than the organization's ongoing operations. These subjects can be product, customers, suppliers, sales, revenue, etc. A data warehouse does not focus on the ongoing operations, rather it focuses on modelling and analysis of data for decision making.

Integrated: A data warehouse integrates data from multiple data sources. For example, source A and source B may have different ways of identifying a product, but in a data warehouse, there will be only a single way of identifying a product.

Time-Variant: Historical data is kept in a data warehouse. For example, one can retrieve data from 3 months, 6 months, 12 months, or even older data from a data warehouse. This contrasts with a transactions system, where often only the most recent data is kept. For example, a transaction system may hold the most recent address of a customer, where a data warehouse can hold all addresses associated with a customer.

Non-volatile: Once data is in the data warehouse, it will not change. So, historical data in a data warehouse should never be altered.

Q13 Differentiate between OLTP and OLAP. (4 marks)

OLTP	OLAP
1) OLTP stands for Online Transaction Processing.	OLAP stands for Online Analytical Processing.
2) Data comes from different OLTP databases.	It is the original source of data.
3) Database transactions are short.	Database Transactions are long.
4) In OLTP, the processing time of the transaction is less.	In OLAP, the processing time of the transaction is more.
5) Queries are simple.	Queries are complex.
6) It is market orientated.	It is customer orientated.
7) Tables are normalized	Tables are not normalized.
8) OLTP Insert, Update, Delete data from the database.	For analyzing, it extracts data which helps in decision making.
9) It is very fast.	Depending on data, it is slow.

Q14 Data Warehouse project cycle.

After the tools and team personnel selections are made, the data warehouse design can begin. The following are the typical steps involved in the data warehousing project cycle.

- Requirement Gathering
- Physical Environment Setup
- Data Modeling
- ETL

- OLAP Cube Design
- Front End Development
- Report Development
- Performance Tuning
- Query Optimization
- Quality Assurance
- Rolling out to Production
- Production Maintenance
- Incremental Enhancements

Each page listed above represents a typical data warehouse design phase, and has several sections:

- **Task Description:** This section describes what typically needs to be accomplished during this particular data warehouse design phase.
- **Time Requirement:** A rough estimate of the amount of time this particular data warehouse task takes.
- **Deliverables:** Typically at the end of each data warehouse task, one or more documents are produced that fully describe the steps and results of that particular task. This is especially important for consultants to communicate their results to the clients.

Q15 Business Analysis Framework

The business analyst get the information from the data warehouses to measure the performance and make critical adjustments in order to win over other business holders in the market. Having a data warehouse offers the following advantages –

- Since a data warehouse can gather information quickly and efficiently, it can enhance business productivity.
- A data warehouse provides us a consistent view of customers and items, hence, it helps us manage customer relationship.
- A data warehouse also helps in bringing down the costs by tracking trends, patterns over a long period in a consistent and reliable manner.

To design an effective and efficient data warehouse, we need to understand and analyze the business needs and construct a **business analysis framework**. Each person has different views regarding the design of a data warehouse. These views are as follows –

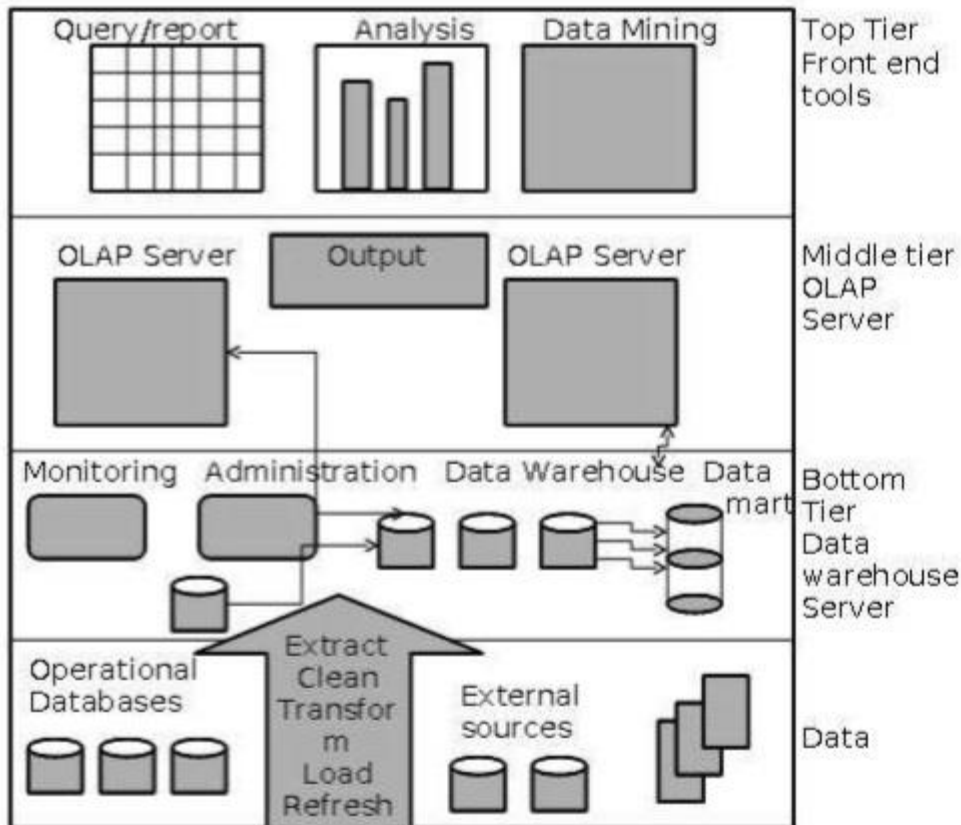
- **The top-down view** – This view allows the selection of relevant information needed for a data warehouse.
- **The data source view** – This view presents the information being captured, stored, and managed by the operational system.
- **The data warehouse view** – This view includes the fact tables and dimension tables. It represents the information stored inside the data warehouse.
- **The business query view** – It is the view of the data from the viewpoint of the end-user.

Q16 Explain Three-Tier Data Warehouse Architecture

Generally a data warehouses adopts a three-tier architecture. Following are the three tiers of the data warehouse architecture.

- **Bottom Tier** – The bottom tier of the architecture is the data warehouse database server. It is the relational database system. We use the back end tools and utilities to feed data into the bottom tier. These back end tools and utilities perform the Extract, Clean, Load, and refresh functions.
- **Middle Tier** – In the middle tier, we have the OLAP Server that can be implemented in either of the following ways.
 - By Relational OLAP (ROLAP), which is an extended relational database management system. The ROLAP maps the operations on multidimensional data to standard relational operations.
 - By Multidimensional OLAP (MOLAP) model, which directly implements the multidimensional data and operations.
- **Top-Tier** – This tier is the front-end client layer. This layer holds the query tools and reporting tools, analysis tools and data mining tools.

The following diagram depicts the three-tier architecture of data warehouse –



Q17 Describe Data Warehouse Models.

From the perspective of data warehouse architecture, we have the following data warehouse models –

- Virtual Warehouse
- Data mart
- Enterprise Warehouse

Virtual Warehouse

The view over an operational data warehouse is known as a virtual warehouse. It is easy to build a virtual warehouse. Building a virtual warehouse requires excess capacity on operational database servers.

Data Mart

Data mart contains a subset of organization-wide data. This subset of data is valuable to specific groups of an organization.

In other words, we can claim that data marts contain data specific to a particular group. For example, the marketing data mart may contain data related to items, customers, and sales. Data marts are confined to subjects.

Points to remember about data marts –

- Window-based or Unix/Linux-based servers are used to implement data marts. They are implemented on low-cost servers.
- The implementation data mart cycles is measured in short periods of time, i.e., in weeks rather than months or years.
- The life cycle of a data mart may be complex in long run, if its planning and design are not organization-wide.
- Data marts are small in size.
- Data marts are customized by department.
- The source of a data mart is departmentally structured data warehouse.
- Data mart are flexible.

Enterprise Warehouse

- An enterprise warehouse collects all the information and the subjects spanning an entire organization
- It provides us enterprise-wide data integration.
- The data is integrated from operational systems and external information providers.
- This information can vary from a few gigabytes to hundreds of gigabytes, terabytes or beyond.