1. DATA WAREHOUSES AND DATA MART:

A Data Warehousing (DW) is process for collecting and managing data from varied sources to provide meaningful business insights. A Data warehouse is typically used to connect and analyze business data from heterogeneous sources. The data warehouse is the core of the BI system which is built for data analysis and reporting. It is a blend of technologies and components which aids the strategic use of data. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

Data warehouse system is also known by the following name:

- Decision Support System (DSS)
- Executive Information System
- Management Information System
- Business Intelligence Solution
- Analytic Application
- Data Warehouse

How Data warehouse works?

A Data Warehouse works as a central repository where information arrives from one or more data sources. Data flows into a data warehouse from the transactional system and other relational databases.

Data may be:

1. Structured
2. Semi-structured
3. Unstructured data

The data is processed, transformed, and ingested so that users can access the processed data in the Data Warehouse through Business Intelligence tools, SQL clients, and spreadsheets. A data warehouse merges information coming from different sources into one comprehensive database.

Types of Data Warehouse:

1. Enterprise Data Warehouse:

Enterprise Data Warehouse is a centralized warehouse. It provides decision support service across the enterprise. It offers a unified approach for organizing and representing data. It also provides the ability to classify data according to the subject and give access according to those divisions.

2. Operational Data Store:

Operational Data Store, which is also called ODS, are nothing but data store required 1.2 when neither Data warehouse nor OLTP systems support organizations reporting needs. In ODS, Data warehouse is refreshed in real time. Hence, it is widely preferred for routine activities like storing records of the Employees.

3. Data Mart:

A data mart is a subset of the data warehouse. It specially designed for a particular line of business, such as sales, finance, sales or finance. In an independent data mart, data can collect directly from sources.

Components of Data warehouse :

Load manager: Load manager is also called the front component. It performs with all the operations associated with the extraction and load of data into the warehouse. These operations include transformations to prepare the data for entering into the Data warehouse.

Warehouse Manager: Warehouse manager performs operations associated with the management of the data in the warehouse. It performs operations like analysis of data to ensure consistency, creation of indexes and views, generation of denormalization and aggregations, transformation and merging of source data and archiving and baking-up data.

Query Manager: Query manager is also known as backend component. It performs all the operation operations related to the management of user queries. The operations of this Data warehouse components are direct queries to the appropriate tables for scheduling theexecution of queries.

End-user access tools:

1. Data Reporting
2. Query Tools
3. Application development tools
4. EIS tools,
5. OLAP tools and data mining tools.

Who needs Data warehouse?

- Decision makers who rely on mass amount of data
- Users who use customized, complex processes to obtain information from multipledata sources.
- It is also used by the people who want simple technology to access the data
- It also essential for those people who want a systematic approach for makingdecisions.
- If the user wants fast performance on a huge amount of data which is a necessity for reports, grids or charts, then Data warehouse proves useful.

What Is a Data Warehouse Used For?

Here, are most common sectors where Data warehouse is used:

Airline:

In the Airline system, it is used for operation purpose like crew assignment, analyses of route profitability, frequent flyer program promotions, etc.

Banking:

It is widely used in the banking sector to manage the resources available on desk effectively.Few banks also used for the market research, performance analysis of the product and operations.

Healthcare:

Healthcare sector also used Data warehouse to strategize and predict outcomes, generate patient's treatment reports, share data with tie-in insurance companies, medical aid services, etc.

Public sector:

In the public sector, data warehouse is used for intelligence gathering. It helps government agencies to maintain and analyze tax records, health policy records, for everyindividual.

Telecommunication:

A data warehouse is used in this sector for product promotions, sales decisions and to make distribution decisions.

Advantages of Data Warehouse:

- Data warehouse allows business users to quickly access critical data from some sources all in one place.
- Data warehouse provides consistent information on various cross-functionalactivities. It is also supporting ad-hoc reporting and query.
- Data warehouse helps to reduce total turnaround time for analysis and reporting.
- Data warehouse allows users to access critical data from the number of sources in asingle place. Therefore, it saves user's time of retrieving data from multiple sources.
- Data warehouse stores a large amount of historical data. This helps users to analyze different time periods and trends to make future predictions

Disadvantages of Data Warehouse:

- Creation and Implementation of Data Warehouse is surely time confusing affair.
- Data Warehouse can be outdated relatively quickly
- Difficult to make changes in data types and ranges, data source schema, indexes, and queries.
- The data warehouse may seem easy, but actually, it is too complex for the average users.
- Despite best efforts at project management, data warehousing project scope will always increase.

Data Warehouse Tools:

1. **Mark Logic:**

   MarkLogic is useful data warehousing solution that makes data integration easier andfaster using an array of enterprise features. This tool helps to perform very complex search operations. It can query different types of data like documents, relationships, and metadata.

2. **Oracle:**

   Oracle is the industry-leading database. It offers a wide range of choice of data warehouse solutions for both on-premises and in the cloud. It helps to optimizecustomer experiences by increasing operational efficiency.

3. **Amazon RedShift:**

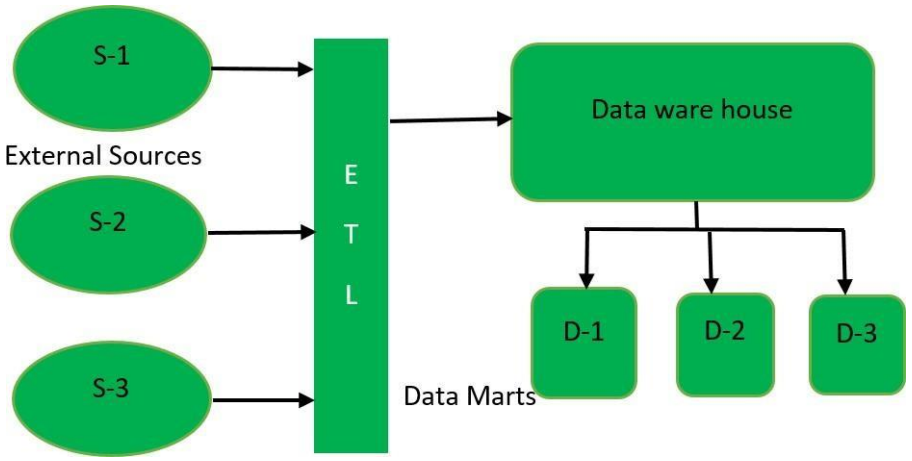
   Amazon Redshift is Data warehouse tool. It is a simple and cost-effective tool to analyzeall types of data using standard SQL and existing BI tools. It also allows running complex queries against petabytes of structured data, using the technique of query optimization.

DATA MART

Data mart is such a storage component which is concerned on a specific department of an organization. It is a subset of the data stored in the data warehouse. Data mart is focused only on particular function of an organization and it is maintained by single authority only, e.g. finance, Marketing. Data Marts are small in size and are flexible.

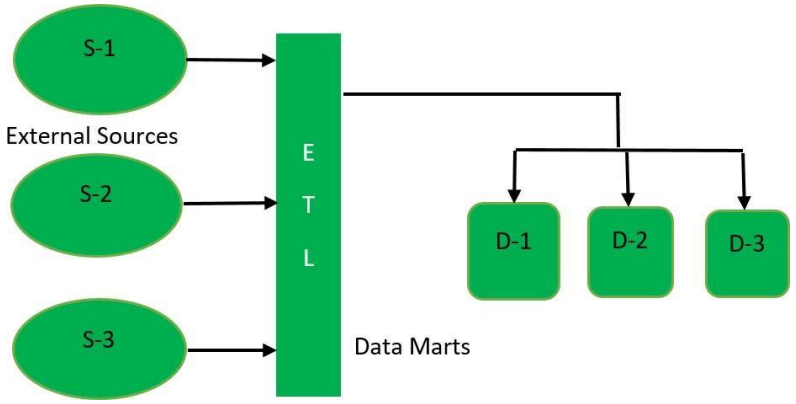
Types of Data Mart:

1. Dependent Data Mart –



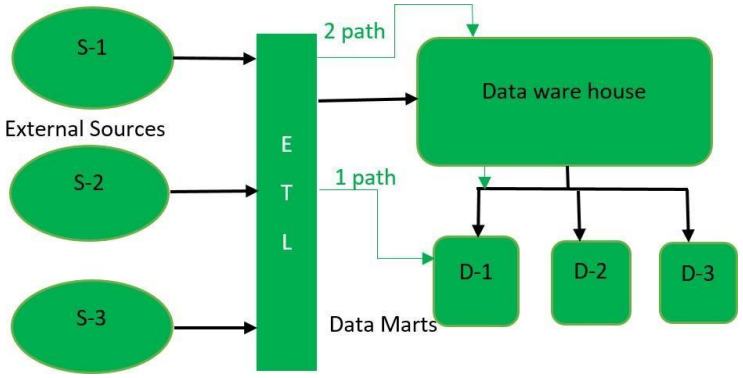
Dependent Data Mart is created by extracting the data from central repository, Datawarehouse. First data warehouse is created by extracting data (through ETL tool) from external sources and then data mart is created from data warehouse. Dependent data mart is created in top-down approach of datawarehouse architecture. This model of data mart is used by big organizations.

2. Independent Data Mart –



Independent Data Mart is created directly from external sources instead of data warehouse. First data mart is created by extracting data from external sources and then data warehouse is created from the data present in data mart. Independent data mart is designed in bottom-up approach of data warehouse architecture. This model of data mart is used by small organizations and is cost effective comparatively.

4. Hybrid Data Mart



This type of Data Mart is created by extracting data from operational source or from data warehouse. 1Path reflects accessing data directly from external sources and 2Path reflects dependent data model of data mart.

Need Of Data Mart:

- Data Mart focuses only on functioning of particular department of an organization.
- It is maintained by single authority of an organization.
- Since, it stores the data related to specific part of an organization, data retrieval from it is very quick.
- Designing and maintenance of data mart is found to be quite cinch as compared to data warehouse.
- It reduces the response time of user as it stores small volume of data.
- It is small in size due to which accessing data from it very fast.

Advantages of Data Mart:

- Implementation of data mart needs less time as compared to implementation of data warehouse as data mart is designed for a particular department of an organization.
- Organizations are provided with choices to choose model of data mart depending upon cost and their business.
- Data can be easily accessed from data mart.
- It contains frequently accessed queries, so enable to analyze business trend.

Disadvantages of Data Mart:

- Since it stores the data related only to specific function, so does not store huge volume of data related to each and every department of an organization like data warehouse.
- Creating too many data marts becomes cumbersome sometimes.

Features of data marts:

Subset of Data: Data marts are designed to store a subset of data from a larger data warehouse or data lake. This allows for faster query performance since the data in the data mart is focused on a specific business unit or department.

Optimized for Query Performance: Data marts are optimized for query performance, which means that they are designed to support fast queries and analysis of the data stored in the data mart.

Customizable: Data marts are customizable, which means that they can be designed to meet the specific needs of a business unit or department.

Self-Contained: Data marts are self-contained, which means that they have their own set of tables, indexes, and data models. This allows for easier management and maintenance of the data mart.

Security: Data marts can be secured, which means that access to the data in the data mart can be controlled and restricted to specific users or groups.

Scalability: Data marts can be scaled horizontally or vertically to accommodate larger volumes of data or to support more users.

Integration with Business Intelligence Tools: Data marts can be integrated with business intelligence tools, such as Tableau, Power BI, or QlikView, which allows users to analyze and visualize the data stored in the data mart.

ETL Process: Data marts are typically populated using an Extract, Transform, Load (ETL) process, which means that data is extracted from the larger data warehouse or data lake, transformed to meet the requirements of the data mart, and loaded into the data mart.

2. KNOWLEDGE MANAGEMENT (KM): CONCEPT, FEATURES AND PROCESS

Knowledge management is a process of acquiring, generating, accumulating and using knowledge for the benefit of the organization to enable it to gain a competitive edge for survival, growth and prosperity in a globalized competitive economy.

According to some management experts, notably Peter F. Drucker, KM is a bad term; in as much as knowledge cannot be managed.

Rather, KM requires conditions for the emergence of a learning organization; which is necessary for generation, sharing and use of knowledge residing in the minds of people.

Features of Knowledge Management

(i) KM is a systematic process; consisting of standardized procedures to collect, store,

distribute and use knowledge. The essence of KM is to get right knowledge to right people, at the right time.

(ii) Knowledge is of two types – explicit and implicit. Explicit knowledge is visible information available in literature, reports, patents, technical specifications, communication with customers, suppliers, competitors etc. It can be embedded in rules, systems, policies and procedures etc. of the organization.

Tacit or implicit knowledge is personal knowledge residing in the minds of people as a result of their personal beliefs, values, perspectives and experience. There is a need for a learning organization for enhancement, sharing and utilization of tacit knowledge.

(iii) KM is a continuous process; as the world economy is dynamic and full of challenges. It requires constant creation of new skills and capabilities and improvement of existing ones.

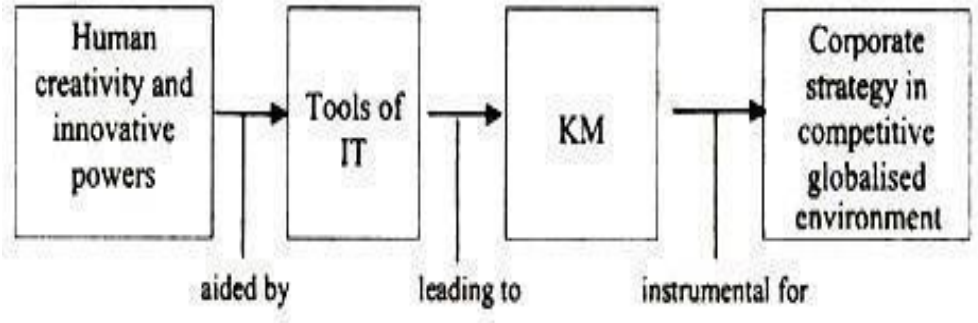
(iv) KM requires whole-hearted support of top management, to provide cultural and technical foundation for the origination and implementation of KM practices.

(v) The objective of KM is improvement in organizational performance; to enable the organization acquire, sharpen and utilize its competitive edge for survival and growthin the global economy of today.

Knowledge Management and Information Technology:

KM is not an outgrowth of IT. Rather, KM requires human skills, creativity and innovative capabilities of people; which are the base of KM. In fact I there are tools of IT like Intranets, Lotus Notes, MS-Exchange etc.; which provide an infrastructure for the free play of human creativity and innovative powers for the formulation of corporation strategy, in a competitive globalized environment.

The above ideas are illustrated with the help of the following diagram:



KM, IT and Corporate Strategy

Knowledge Management IT and Corporate Strategy

An Overview of the Process of KM:

KM broadly consists of the following major steps:

(i) Identification of Knowledge Needs:

The first step in KM is an identification of what type of knowledge is required for the successful designing and implementation of corporate strategy.

(ii) Determination of Knowledge Assets:

The management must identify what are the knowledge assets of the organisation; which basically are competitors, suppliers, governmental agencies, products and processes, technology etc. Management must plan to get maximum returns out of knowledge assets.

(iii) Generation of Knowledge:

Generation of knowledge requires two sources:

(a) Acquisition of knowledge through knowledge assets e.g. knowledge about new products (from competitors), new technologies, social, economic, political changes. It also requires transformation of raw information into knowledge, useful to solve business problems.

(b) Generation of knowledge, by creating conditions for the emergence of a learning organization. This is the most important internal source of knowledge generation which makes tacit knowledge of individuals available for organizational purposes.

(iv) Knowledge Storage:

It includes preserving existing and acquired knowledge in knowledge repositories. (A knowledge repository is an on-line computer-based storehouse of organized information about a particular domain of knowledge).

(v) Knowledge Distribution:

It is a process which allows members of the organization to have an access to the collectiveknowledge of the organization.

(vi) Knowledge Utilization:

It requires embedding knowledge in products, processes, procedures etc. of the organization. Best utilization of knowledge takes place when managers utilize knowledge in organizational decision making. A learning organization creates conditions for sharing and utilizing knowledge in organizational contexts.

(vii) Feedback on Knowledge Management

Feedback on KM implies evaluating the significance of knowledge assets. It also includes impact of KM on organizational performance; and devising techniques for betterment ofKM in future.

Significance of Knowledge Management

(i) Building and Sharpening Competitive Edge:

KM enables a corporation to build and sharpen its competitive edge, for survival and growth in the competitive globalized economy. In fact, KM aided by IT tools enables a corporation to

design and implement most appropriate corporate strategies.

(ii) Betterment of Human Relations:

KM is basically built on the knowledge generated, shared and utilized through a learning organization. There is no doubt that learning organization provides the foundation on which the building of KM could be built. A learning organization through facilitating interaction among people of the organisation, leads to betterment of human relations; which is a very big permanent asset an organisation can boast of to possess.

(iii) Improvement in Organizational Efficiency:

KM provides knowledge which can be embedded in organizational processes. It makes knowledge available for decision-making purposes. Thus it helps to improve organizational efficiency, resulting in reduced costs and increased profits, for the organization.

(iv) Enhancement of Human Capital Capabilities

KM-its concept and practices – motivate people to enhance their intellectual capabilities, resulting in new skills, improvement of existing skills etc. Thus not only does KM enhance the intellectual elements of people; but also indirectly prevents depreciation of human capital.

(v) Enhancement of Enterprise Goodwill:

Initiation and practices of KM help an enterprise enhance its goodwill in the global market; enabling it to acquire more success and prosperity.

3. TYPES OF DECISIONS

The characteristics of decisions faced by managers at different levels are quite different. Decisions can be classified as structured, semi structured, and unstructured. Unstructured decisions are those in which the decision maker must provide judgment, evaluation, and insights into the problem definition. Each of these decisions is novel, important, and nonroutine, and there is no well-understood or agreed-on procedure for making them.

Structured decisions, by contrast, are repetitive and routine, and decision makers can follow a definite procedure for handling them to be efficient. Many decisions have elements of both and are considered semi structured decisions, in which only part of the problem has a clear-cut answer provided by an accepted procedure. In general, structured decisions are made more prevalently at lower organizational levels, whereas unstructured decision making is more

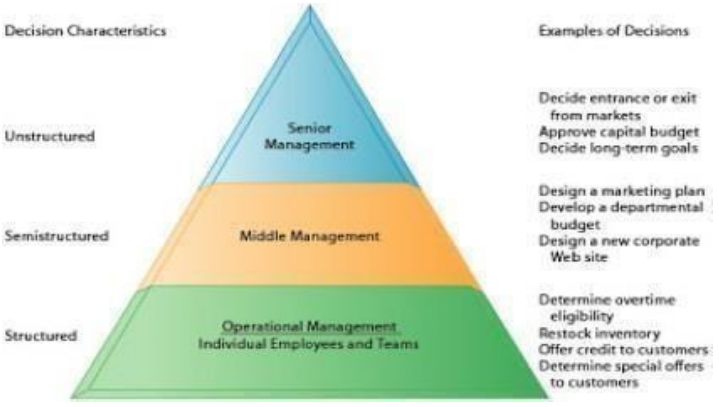
common at higher levels of the firm.

Senior executives tend to be exposed to many unstructured decision situations that are open ended and evaluative and that require insight based on many sources of information and personal experience. For example, a CEO in today's music industry might ask, "Whom should we choose as a distribution partner for our online music catalog— Apple, Microsoft, or Sony?" Answering this question would require access to news, government reports, and industry views as well as high-level summaries of firm performance. However, the answer would also require senior managers to use their own best judgment and poll other managers for their opinions.

Middle management and operational management tend to face more structured decision scenarios, but their decisions may include unstructured components. A typical middle level management decision might be "Why is the order fulfillment report showing a decline over the last six months at a distribution center in Minneapolis?" This middle manager could obtain a report from the firm's enterprise system or distribution management system on order activity and operational efficiency at the Minneapolis distribution center. This is the structured part of the decision. But before arriving at an answer, this middle manager will have to interview employees and gather more unstructured information from external sources about local economic conditions or sales trends. Rank-and-file employees tend to make more structured decisions. For example, a sales account representative often has to make decisions about extending credit to customers by consulting the firm's customer database that contains credit information. In this case the decision is highly structured, it is a routine decision made thousands of times each day in most firms, and the answer has been preprogrammed into a corporate risk management or credit reporting system.

The types of decisions faced by project teams cannot be classified neatly by organizational level. Teams are small groups of middle and operational managers and perhaps employees assigned specific tasks that may last a few months to a few years.

Their tasks may involve unstructured or semi structured decisions such as designing new products, devising new ways to enter the marketplace, or reorganizing sales territories and compensation systems.



SYSTEMS FOR DECISION SUPPORT

There are four kinds of systems used to support the different levels. We introduced some of these systems in Management *information systems (MIS)* provide routine reports and summaries of transaction- level data to middle and operational-level managers to provide

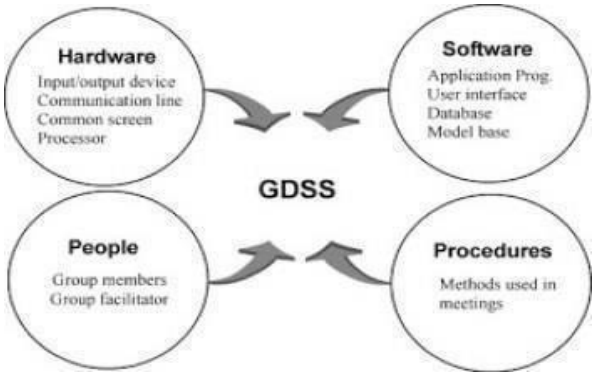
answers to structured and semi structured decision problems.

1. Decision-support systems (DSS) are targeted systems that combine analytical models with operational data and supportive interactive queries and analysis for middle managers who face semi structured decision situations.

2.Executive support systems (ESS) are specialized systems that provide senior management making primarily unstructured decisions with a broad array of both externalinformation (news, stock analyses, industry trends) and high-level summaries of firm performance. The purpose of ESS to help the C- level managers to focus on the information that really affect the overall profitability and success of the firm. The leading methodology for understanding the really important information needed by the firm's executive is calledthe Balanced Score Card Method,a frame work for operationalizing the firm's strategic planby focusing on measurable

outcomes on four dimensions of firm performance. Financial, business process, customer, learning and growth. Performance on each dimension is measured using KPI's.

3. Group decision-support systems (GDSS) are specialized systems that provide a group electronic environment in which managers and teams can collectively make decisions and design solutions for unstructured and semi structured problems. GDSS guided meetings takes place in a conference room with special software and hardware tools to facilitate group decision making. It makes possible to increase the meeting size and increase in productivity. Because individuals contribute simultaneously at the same time rather than one at a time.



Organizational Level	Decision Type	Type of Decision-Support System	Examples
Senior management	Unstructured	Executive support systems (ESS)	Decide entrance or exit from markets Approve capital budget Decide long-term corporate objectives
Middle management/ project teams	Semistructured	Management information systems (MIS)	Allocate resources to managers and departments
		Decision-support systems (DSS) Group decision-support systems(GDSS)	Design a new corporate Web site Develop a marketing plan Design a departmental budget
Operational management/ project teams Employees	Semistructured Structured	Decision-support systems (DSS) Management information systems (MIS) Group decision-support systems (GDSS)	Evaluate employee performance Restock inventory Routine credit decisions Determine special offers to customers

Trends in Decision Support and Business Intelligence

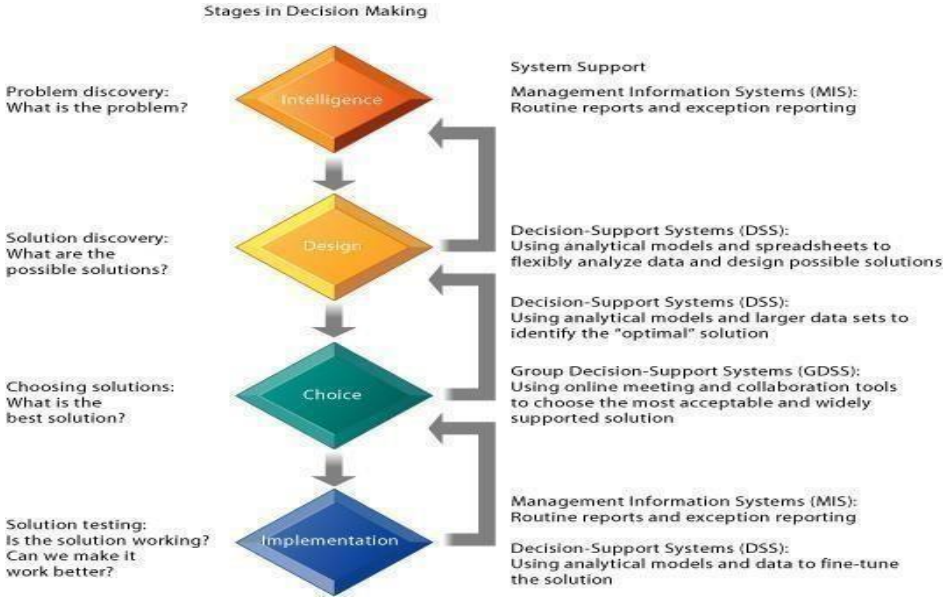
Systems supporting management decision making originated in the early 1960s as early MIS that created fixed, inflexible paper-based reports and distributed them to managers on a routine schedule. In the 1970s, the first DSS emerged as standalone applications with limited data and a few analytic models. ESS emerged during the 1980s to give senior managers an overview of corporate operations. Early ESS were expensive, based on custom technology, and suffered from limited data and flexibility.

The rise of client/server computing, the Internet, and Web technologies has made a major impact on systems that support decision making. Many decision-support applications are now delivered over corporate intranets. We see six major trends:

- **Detailed enterprise-wide data.** Enterprise systems create an explosion in firmwide, current, and relatively accurate information, supplying end users at their desktops with powerful analytic tools for analyzing and visualizing data.
- **Broadening decision rights and responsibilities.** As information becomes more widespread throughout the corporation, it is possible to reduce levels of hierarchy and grant more decision-making authority to lower-level employees.
- **Intranets and portals.** Intranet technologies create global, company-wide networks that ease the flow of information across divisions and regions and delivery of near real- time data to management and employee desktops.
- **Personalization and customization of information.** Web portal technologies provide great flexibility in determining what data each employee and manager sees on his or her desktop. Personalization of decision information can speed up decision making by enabling users to filter out irrelevant information.
- **Extranets and collaborative commerce.** Internet and Web technologies permit suppliers and logistics partners to access firm enterprise data and decision-support tools and work collaboratively with the firm.
- **Team support tools.** Web-based collaboration and meeting tools enable project teams, task forces, and small groups to meet online using corporate intranets or extranets. These new collaboration tools borrow from earlier GDSS and are used for both brainstorming and decision sessions.

4. STAGES IN THE DECISION-MAKING PROCESS

Making decisions consists of several different activities. Simon (1960) describes four different stages in decision making: intelligence, design, choice, and implementation.



Intelligence consists of discovering, identifying, and understanding the problems occurring in the organization—why is there a problem, where, and what effects is it having on the firm. Traditional MIS that deliver a wide variety of detailed information can help identify problems, especially if the systems report exceptions.

Design involves identifying and exploring various solutions to the problem. Decision support systems (DSS) are ideal in this stage for exploring alternatives because they possess analytical tools for modeling data, enabling users to explore various options quickly.

Choice consists of choosing among solution alternatives. Here, DSS with accessextensive firm data can help managers choose the optimal solution. Also group decision support systems can be used to bring groups of managers together in an electronic online environment to discussdifferent

solutions and make a choice.

Implementation involves making the chosen alternative work and continuing to monitor how well the solution is working. Here, traditional MIS come back into play by providing managers with routine reports on the progress of a specific solution. Support systems can range from full-blown MIS to much smaller systems, as well as project- planning software operating on personal computers.

In the real world, the stages of decision making described here do not necessarily follow a linear path. You can be in the process of implementing a decision, only to discoverthat your solution is not working. In such cases, you will be forced to repeat the design, choice, or perhaps even the intelligence stage.

For instance, in the face of declining sales, a sales management team may strongly support a new sales incentive system to spur the sales force on to greater effort. If paying the sales force, a higher commission for making more sales does not produce sales increases, managers would need to investigate whether the problem stems from poor product design, inadequate customer support, or a host of other causes, none of which would be "solved" by a new incentive system.

4. BUSINESS INTELLIGENCE

Business intelligence combines business analytics, data mining, data visualization, data tools and infrastructure, and best practices to help organizations make more data-driven decisions. In practice, you know you've got modern business intelligence when you have a comprehensive view of your organization's data and use that data to drive change, eliminate inefficiencies, and quickly adapt to market or supply changes. Modern BI solutions prioritize flexible self-service analysis, governed data on trusted platforms, empowered business users, and speed to insight

Business Intelligence is a set of processes, architectures, and technologies that convert rawdata into meaningful information that drives profitable business actions. It is a suite of software and services to transform data into actionable intelligence and knowledge.

BI has a direct impact on organization's strategic, tactical and operational business decisions. BI supports fact-based decision making using historical data rather than assumptions and gut feeling.

BI tools perform data analysis and create reports, summaries, dashboards, maps, graphs,and charts to provide users with detailed intelligence about the nature of the business.

Why is BI important?

- Measurement: creating KPI (Key Performance Indicators) based on historic dataIdentify and set benchmarks for varied processes.
- With BI systems organizations can identify market trends and spot business problemsthat need to be addressed.
- BI helps on data visualization that enhances the data quality and thereby the quality of decision making.
- BI systems can be used not just by enterprises but SME (Small and Medium Enterprises)

How Business Intelligence systems are implemented?

step 1) Raw Data from corporate databases is extracted. The data could be spread across multiple systems heterogeneous systems.

step 2) The data is cleaned and transformed into the data warehouse. The table can belinked, and data cubes are formed.

Step 3) Using BI system the user can ask quires, request ad-hoc reports or conduct any other analysis.

Examples of Business Intelligence System used in Practice

Example 1:

In an Online Transaction Processing (OLTP) system information that could be fed into product database could be

- add a product line
- change a product price

Correspondingly, in a Business Intelligence system query that would beexecuted for the product subject area could be did the addition of new product line or change in product price increase revenues

In an advertising database of OLTP system query that could be executed

- Changed in advertisement options
- Increase radio budget

Correspondingly, in BI system query that could be executed would be how many new clients added due to change in radio budget.

In OLTP system dealing with customer demographic data bases data that could be fed would be

- increase customer credit limit

- change in customer salary level

Correspondingly in the OLAP system query that could be executed would be can customer profile changes support support higher product price

Example 2:

A hotel owner uses BI analytical applications to gather statistical information regarding average occupancy and room rate. It helps to find aggregate revenue generated per room.

It also collects statistics on market share and data from customer surveys from each hotel to decides its competitive position in various markets.

By analyzing these trends year by year, month by month and day by day helps management to offer discounts on room rentals.

Example 3:

A bank gives branch managers access to BI applications. It helps branch manager to determine who are the most profitable customers and which customers they should work on.

The use of BI tools frees information technology staff from the task of generating analytical reports for the departments. It also gives department personnel access to a richer data source.

Four types of BI users

Following given are the four key players who are used Business Intelligence System:

1. **The Professional Data Analyst:**

   The data analyst is a statistician who always needs to drill deep down into data. BIsystem helps them to get fresh insights to develop unique business strategies.

2. **The IT users:**

   The IT user also plays a dominant role in maintaining the BI infrastructure.

3. **The head of the company:**

   CEO or CXO can increase the profit of their business by improving operationalefficiency in their business.

4. **The Business Users"**

   Business intelligence users can be found from across the organization. There are mainly two types of business users

   1. Casual business intelligence user
   2. The power user.

The difference between both of them is that a power user has the capability of working with complex data sets, while the casual user need will make him use dashboards to evaluate predefined sets of data.

Types of BI Tools and Software

- **Spreadsheets:** Spreadsheets like Microsoft Excel and Google Docs are some of the most widely used BI tools.
- **Reporting software:** Reporting software is used to report, organize, filter, and display data.
- **Data visualization software:** Data visualization software translates datasets into easy-to-read, visually appealing graphical representations to quickly gain insights.
- **Data mining tools:** Data mining tools "mine" large amounts of data for patterns using things like artificial intelligence, machine learning, and statistics.
- **Online analytical processing (OLAP):** OLAP tools allow users to analyze datasets from a wide variety of angles based on different business perspectives.

Advantages of Business Intelligence

Here are some of the advantages of using Business Intelligence System:

1) **Boost productivity**

With a BI program, It is possible for businesses to create reports with a single click thus saves lots of time and resources. It also allows employees to be more productive on their tasks.

2) **To improve visibility**

BI also helps to improve the visibility of these processes and make it possible to identify any areas which need attention.

3) **Fix Accountability**

BI system assigns accountability in the organization as there must be someone who should own accountability and ownership for the organization's performance against its set goals.

4) **It gives a bird's eye view:**

BI system also helps organizations as decision makers get an overall bird's eye view through typical BI features like dashboards and scorecards.

5) **It streamlines business processes:**

BI takes out all complexity associated with business processes. It also automates analytics by offering predictive analysis, computer modeling, benchmarking and other methodologies.

6) **It allows for easy analytics.**

BI software has democratized its usage, allowing even nontechnical or non-analysts users to collect and process data quickly. This also allows putting the power of analytics from the hand's

many people.

BI System Disadvantages

1. **Cost:**

Business intelligence can prove costly for small as well as for medium-sized enterprises. The use of such type of system may be expensive for routine business transactions.

2. **Complexity:**

Another drawback of BI is its complexity in implementation of data warehouse. It can be so complex that it can make business techniques rigid to deal with limited use

Like all improved technologies, BI was first established keeping in consideration the buying competence of rich firms. Therefore, BI system is yet not affordable for many small and medium size companies.

3. **Time Consuming Implementation**

It takes almost one and half year for data warehousing system to be completely implemented. Therefore, it is a time-consuming process.

5. OLAP (ONLINE ANALYTICAL PROCESSING)

OLAP (online analytical processing) is a computing method that enables users to easily and selectively extract and query data in order to analyze it from different points of view. OLAP business intelligence queries often aid in trends analysis, financial reporting, sales forecasting, budgeting and other planning purposes. For example, a user can request that data be analyzed to display a spreadsheet showing all of a company's beach ball products sold in Florida in the month of July, compare revenue figures with those for the same products in September and then see a comparison of other product sales in Florida in the same time period.

How OLAP systems work

- To facilitate this kind of analysis, data is collected from multiple data sources andstored in data warehouses then cleansed and organized into data cubes.

- Each OLAP cube contains data categorized by dimensions (such as customers, geographic sales region and time period) derived by dimensional tables in the datawarehouses.
- Dimensions are then populated by members (such as customer names, countriesand months) that are organized hierarchically.
- OLAP cubes are often pre-summarized across dimensions to drastically improve query time over relational databases.

OLAP Architecture

OLAP architecture is a multi-layered approach to building data analysis systems that are optimized for OLAP processing. At its core, OLAP architecture is designed to facilitate multidimensional data modeling, fast query performance, and flexible data analysis. The key components of OLAP architecture include:

Data Source Layer: Responsible for storing the raw data that will be analyzed by the OLAP system.

Data Access Layer: Provides a way for the OLAP system to access the data stored in the data source layer.

OLAP Engine Layer: Responsible for processing queries, performing data aggregation, and generating reports.

Front-end Layer: Provides the user interface for interacting with the OLAP system and visualizing the data.

OLAP architecture can be implemented in a variety of ways, depending on the specific needs and requirements of the organization. Some OLAP systems are standalone solutions, while others are integrated with existing data management systems. OLAP architecture can also be customized to support different types of OLAP processing, such as MOLAP, ROLAP, or HOLAP.

OLAP architecture is an important consideration for organizations that need to implement OLAP processing capabilities. OLAP architecture offers the benefit of streamlining data analysis processes and improving efficiency, thereby enabling organizations to make data-driven decisions more efficiently.

Types of OLAP Servers

We have four types of OLAP servers −

- Relational OLAP (ROLAP)
- Multidimensional OLAP (MOLAP)
- Hybrid OLAP (HOLAP)
- Specialized SQL Servers

1. Relational OLAP

Instead of using a data cube, relational online analytical processing (ROLAP) allows data engineers to perform multidimensional data analysis on a relational database. In other words, data

engineers use SQL queries to search for and retrieve specific information based on the required dimensions. ROLAP is suitable for analyzing extensive and detailed data. However, ROLAP has slow query performance compared to MOLAP.ROLAP servers are placed between relational back-end server and client front-end tools. To store and manage warehouse data, ROLAP uses relational or extended-relational DBMS.

ROLAP includes the following −

- Implementation of aggregation navigation logic.
- Optimization for each DBMS back end.
- Additional tools and services.

2. Multidimensional OLAP

MOLAP uses array-based multidimensional storage engines for multidimensional views of data. With multidimensional data stores, the storage utilization may be low if the data set is sparse. Therefore, many MOLAP server use two levels of data storage presentation to handle dense and sparse data sets. Multidimensional online analytical processing (MOLAP) involves creating a data cube that represents multidimensional data from a data warehouse. The MOLAP system stores precalculated data in the hypercube. Data engineers use MOLAP because this type of OLAP technology provides fast analysis.

3. Hybrid OLAP

Hybrid OLAP is a combination of both ROLAP and MOLAP. It offers higher scalability of ROLAP and faster computation of MOLAP. HOLAP servers allows to store the large data volumes of detailed information. The aggregations are stored separately in MOLAP store.

4. Specialized SQL Servers

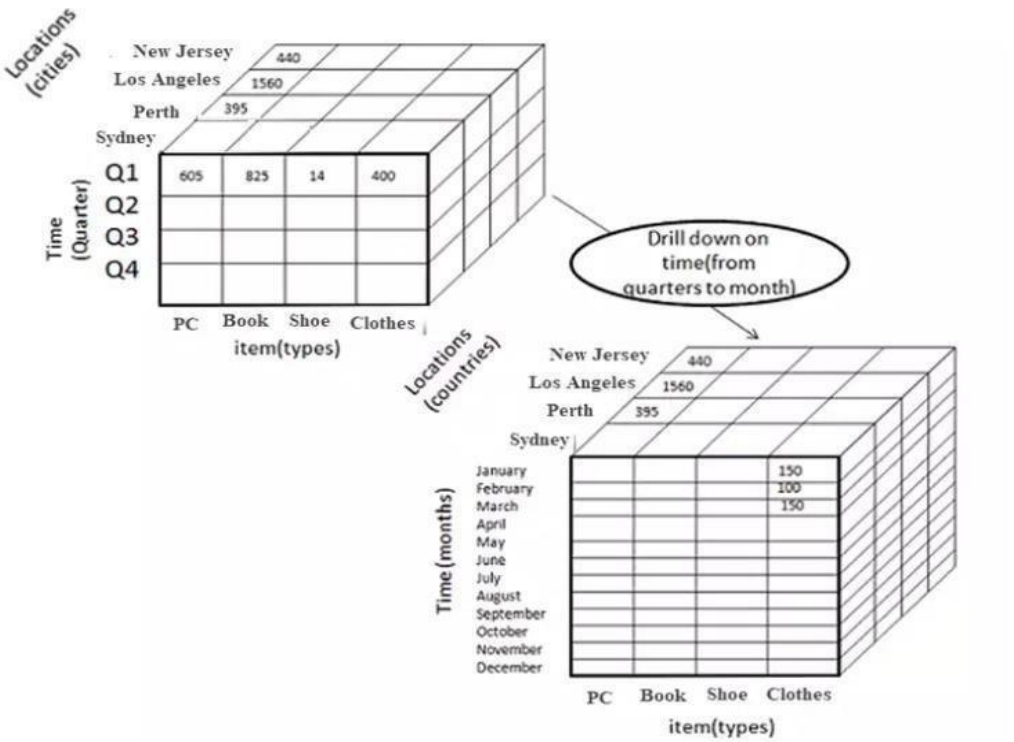
Specialized SQL servers provide advanced query language and query processing support for SQL queries over star and snowflake schemas in a read-only environment.

What are OLAP operations?

Business analysts perform several basic analytical operations with a multidimensional online analytical processing (MOLAP) cube.

Drill down

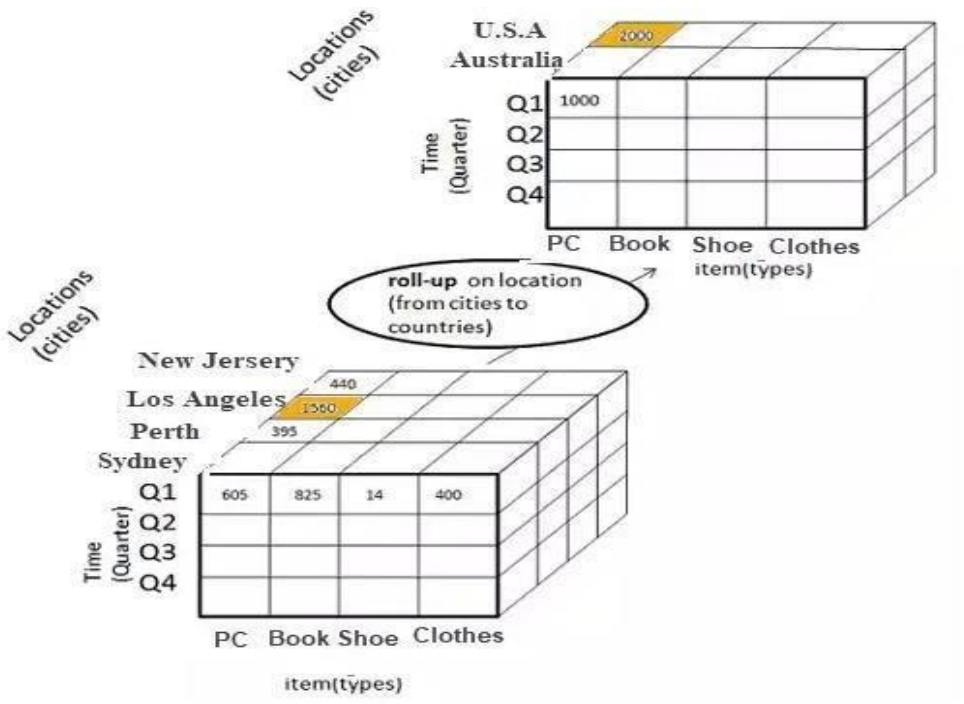
Drill down is the opposite of the roll-up operation. Business analysts move downward in the concept hierarchy and extract the details they require. For example, they can move from viewing sales data by years to visualizing it by months.



Drill down into the time dimension

Roll up

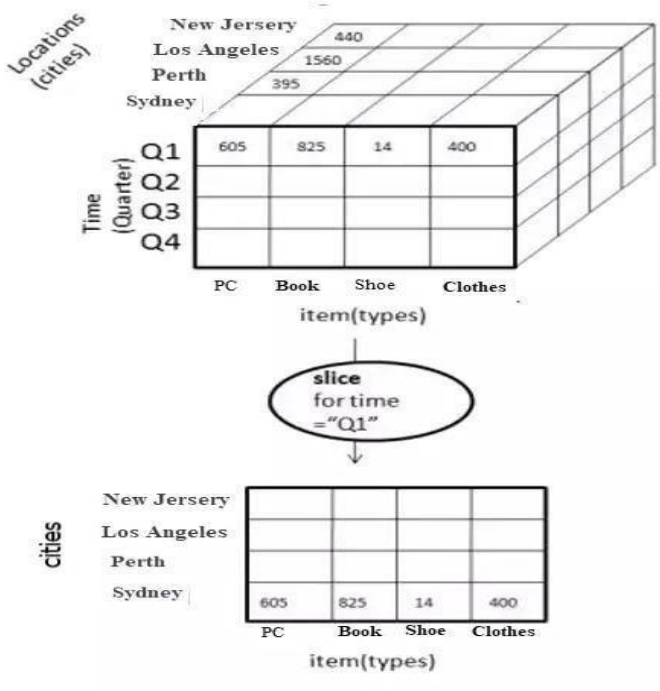
In roll up, the online analytical processing (OLAP) system summarizes the data for specific attributes. In other words, it shows less-detailed data. For example, you might view product sales according to New York, California, London, and Tokyo. A roll-up operation would provide a view of the sales data based on countries, such as the US, the UK, and Japan.



Roll up the location dimension

Slice

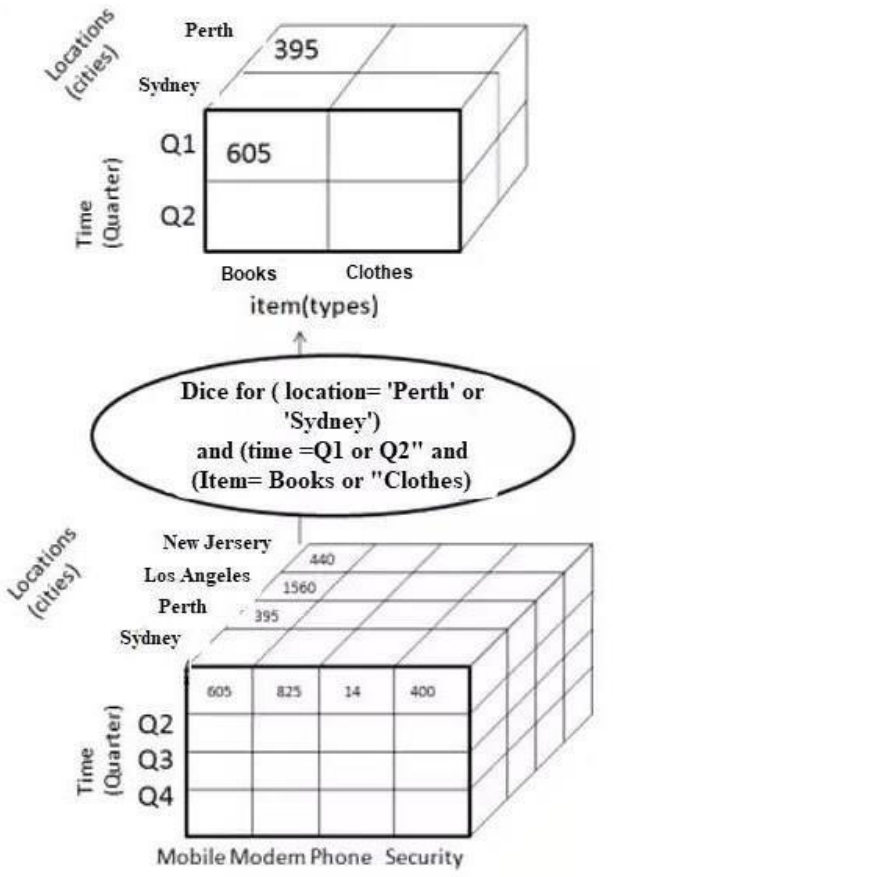
Data engineers use the slice operation to create a two-dimensional view from the OLAP cube. For example, a MOLAP cube sorts data according to products, cities, and months. By slicing the cube, data engineers can create a spreadsheet-like table consisting of products and cities for a specific month.



Slicing a location dimension in a hierarchy level of cities

Dice

Data engineers use the dice operation to create a smaller sub cube from an OLAP cube. They determine the required dimensions and build a smaller cube from the original hypercube.



A dice operation with the choice of time and location dimensions

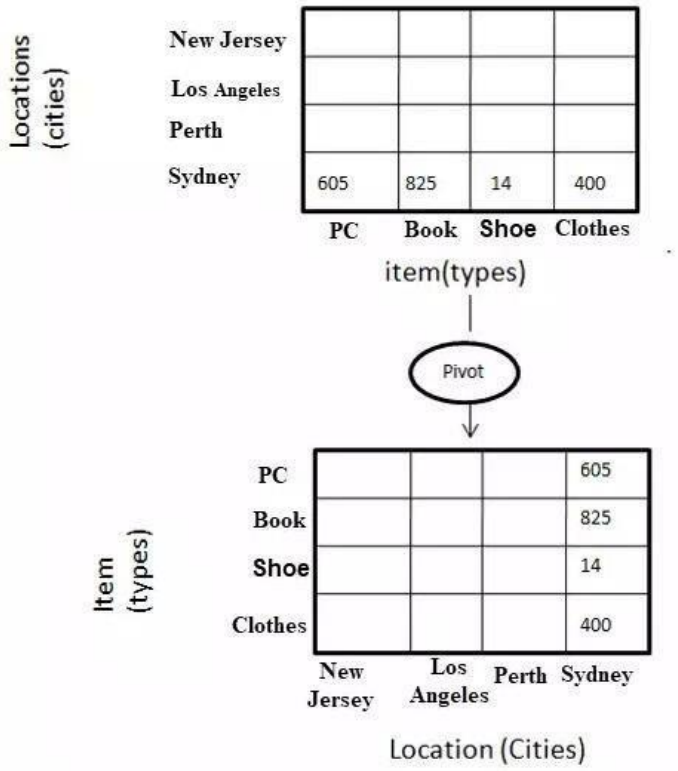
Pivot

The pivot operation involves rotating the OLAP cube along one of its dimensions to get a different perspective on the multidimensional data model. For example, a three-dimensional OLAP cube has the following dimensions on the respective axes:

- X-axis—product
- Y-axis—location
- Z-axis—time

Upon a pivot, the OLAP cube has the following configuration:

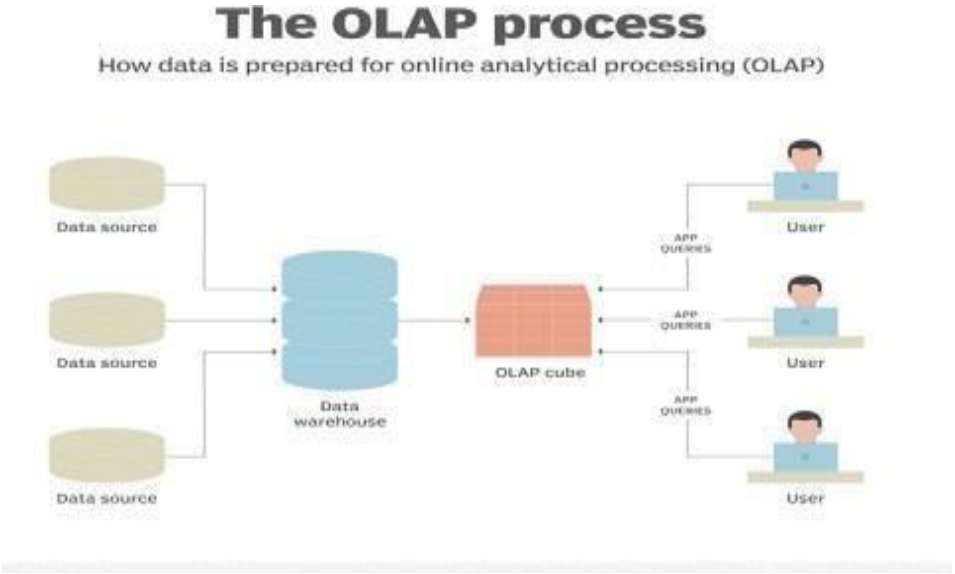
- X-axis—location
- Y-axis—time
- Z-axis—product



Pivot example on a single dimension of the cube

OLAP software then locates the intersection of dimensions, such as all products sold in the Eastern region above a certain price during a certain time period, and displays them. The result is the "measure"; each OLAP cube has at least one to perhaps hundreds of measures, which are derived from information stored in fact tables in the data warehouse.

OLAP begins with data accumulated from multiple sources and stored in a data warehouse. The data is then cleansed and stored in OLAP cubes, which users run queries against.



➢ Association Rule Mining

Association analysis is useful for discovering interesting relationships hidden in large datasets. The uncovered relationships can be represented in the form of association rules or sets of frequent items. Given a set of transactions, find rules that will predict the occurrenceof an item based on the occurrences of other items in the transaction Market- Basket transactions

TID Items

1 Bread, Milk

2 Bread, Diaper, Beer, Eggs

3 Bread, Milk, Diaper, Coke

Implication means co-occurrence, not causality! Example of Association Rules

{Beer}-{Diaper}

{Milk, Bread}-{Eggs, Coke}

{Beer, Bread}-{Milk}

Support Count ( ) – Frequency of occurrence of a item set.

Here ({Milk, Bread, Diaper})=2

Frequent Item set – An item set whose support is greater than or equal to minus threshold.

Association Rule – An implication expression of the form X -> Y, where X and Y are any

2 item sets.

Example: {Milk, Diaper}-

>{Beer}Rule Evaluation Metrics

– Support(s) –

The number of transactions that include items in the {X} and {Y} parts of the rule as percentage of the total number of transaction. It is a measure of how frequently the collection of items occurs together as a percentage of all transactions.

Support = (X+Y) ÷ total –

It is interpreted as fraction of transactions that contain both X and Y. Confidence(c) – It is the ratio of the no of transactions that includes all items in {B} as well as the no of transactions that includes all items in {A} to the no of transactions that includes all items in
{A}.

Conf(X=>Y) = Supp(X Y) ÷ Supp(X) –

It measures how often each item in Y appears in transactions that contains items in Xalso.

Lift (l) –The lift of the rule X=>Y is the confidence of the rule divided by the expected confidence, assuming that the item sets X and Y are independent of each other. The expected confidence is the confidence divided by the frequency of {Y}.

Lift(X=>Y) = Conf(X=>Y) ÷ Supp(Y) –

Lift value near 1 indicates X and Y almost often appear together as expected, greater than 1 means they appear together more than expected and less than 1 means they appear less than expected. Greater lift values indicate stronger association

- Example – From the above table, {Milk, Diaper}=>{Beer}

  s= ({Milk, Diaper, Beer}) ÷|T|

  = 2/5

  = 0.4

- c= (Milk, Diaper, Beer) ÷(Milk, Diaper)

  = 2/3

  = 0.67

- l= Supp({Milk, Diaper, Beer}) ÷Supp({Milk, Diaper})*Supp({Beer})

  = 0.4/ (0.6*0.6)

  = 1.11

The Association rule is very useful in analyzing datasets. The data is collected using bar- code scanners in supermarkets. Such databases consist of a large number of transaction records which list all items bought by a customer on a single purchase. So the manager could know if certain groups of items are consistently purchased together and use this datafor adjusting store layouts, cross-selling, promotions based on statistics.

5. ANALYTIC FUNCTION

An analytic function is a function provided by the relational database that performs an analytical task on a result set. An analytic function in a query returns, with each row in the result set, a calculation from a group of rows. The groups of rows can be ordered and partitioned.

For example, you can use analytic functions to retrieve the following results:

- The rank of a record, for example the rank of retail stores by the sales amount for last month.
- Moving sum or average, for example the average volume of sales in a three-month period.
- Displaying the same information in different contexts, for example this quarter's sales and last quarter's sales.
- Relative calculations, for example the difference between the sales of this quarter and the highest sales amount ever.

Following are some examples of analytic functions:

- Aggregation-like functions: SUM, COUNT, AVG, STDDEV, MEDIAN, VARIANCE
- Order-based functions: RANK, PERCENT_RANK, DENSE_RANK, LEAD, LAG, FIRST_VALUE, ROW_NUMBER

The generic syntax and examples of analytic functions are given to help you understand how analytic functions can be used.

Function (arguments) OVER ([<PARTITION BY clause>] [<ORDER BY clause>] [<ROW or RANGE clause>])

Part of analytic function statement	Description
Function (arguments)	The name and arguments of the function that define the calculation.
OVER (	OVER signals that this is an analytic function. The OVER clause defines the data over which you want to do the calculation. It has three optional clauses.
PARTITION BY clause	The grouping over which the calculation is applied.
ORDER BY clause	The order of the results to be used in the calculation.
ROW or RANGE clause)	The interval of records used for calculation.

The **PARTITION BY clause** lets you define the groups of data over which the function will be calculated. For example:

SELECT employee_id, department, COUNT(employee_id) OVER (PARTITION BY department) FROM employee_table

This query returns, for each employee, the employee's department and the count of the number of employees in each department. The count is returned with every row (employee) in the result set.

employee_id	department	count
1	Marketing	2
2	Marketing	2
3	Sales	3
4	Sales	3
5	Sales	3

The **ORDER BY clause** lets you define the order in which the rows are used when applying the calculation. For example:
SELECT employee_id, salary, RANK () OVER (ORDER BY salary)

This query returns, for each employee, the employee's salary and the employee's overall rank by salary.

employee_id	salary	rank
3	3000	1
2	5000	2
5	6000	3
4	7000	4
1	7200	5

The **ROW or RANGE clause** lets you define a window or interval of ordered rows to take into account when calculating the function on a given row. For example:

SELECT employee_id, salary, SUM(salary) OVER (ORDER BY salary ROWS between unbounded preceding and current row)

This query returns, for each employee, the employee's salary and the sum of salaries starting from the lowest salary up to and including the current employee's salary. The results are ordered by salary. The sum for the last row represents the sum of salaries for all employees.

employee_id	salary	sum
3	3000	3000
2	5000	8000
5	6000	14000

4	7000	21000
1	7200	28200

The following example uses **both the PARTITION BY and ORDER BY clauses:**

SELECT employee_id, department, salary, RANK() OVER (PARTITION
BY department ORDER BY salary)

This query returns, for each employee, the employee's department, salary, and rank within the department, ordered by salary within the department.

employee_id	department	salary	rank
2	Marketing	5000	1
1	Marketing	7200	2
3	Sales	3000	1
5	Sales	5000	2
4	Sales	7000	3

The following example uses **all three clauses:**

SELECT employee_id, department, salary, SUM (salary) OVER (PARTITION
BY department ORDER BY salary ROWS between unbounded preceding and current row)

This query returns, for each employee, the employee's department and salary, and the sum of salaries within the department starting from the lowest salary in the department up to and including the current employee's salary. The rows are ordered by salary within each department.

employee_id	department	salary	sum
2	Marketing	5000	5000
1	Marketing	7200	12200
3	Sales	3000	3000
5	Sales	5000	8000
4	Sales	7000	15000

Analytic Functions is defined as a function that is locally given by the convergent power series. The analytic function is classified into two different types, such as real analytic function and

complex analytic function. Both the real and complex analytic functions are infinitely differentiable. Generally, the complex analytic function holds some properties that do not generally hold for real analytic function.

Types of Analytic Function

Analytic Functions can be categorised into two different types, which are similar in some ways, but it has some different characteristics. The two types of analytic functions are:

- Real Analytic Function
- Complex Analytic Function

Real Analytic Function

A function "f" is said to be a real analytic function on the open set D in the real line if for any x0 ∈ D, then we can write:

$$f(x) = \sum_{n=0}^{\infty} a_n(x - x_0)^n = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3 + \dots$$

where the coeffienets a0, a1, a2, … are the real numbers and also the series is convergent to the function f(x) for x in the neighbourhood of x0.

Complex Analytic Function

A function is said to be a complex analytic function if and only if it is holomorphic. It meansthat the function is complex differentiable.

Properties of Analytic Function

- The limit of a uniformly convergent sequence of analytic functions is also an analytic function
- If f(z) and g(z) are analytic functions on U, then their sum f(z) + g(z) and product f(z).g(z)are also analytic
- If f(z) and g(z) are the two analytic functions and f(z) is in the domain of g for all z, then their composite g(f(z)) is also an analytic function.
- The function f(z) = 1/z (z≠0) is analytic
- Bounded entire functions are constant functions
- Every nonconstant polynomial p(z) has a root. That is, there exists some z0 such that p(z0) = 0.
- If f(z) is an analytic function, which is defined on U, then its modulus of the function |f(z)| cannot attains its maximum in U.
- The zeros of an analytic function, say f(z) are the isolated points unless f(z) is identicallyzero
- If F(z) is an analytic function and if C is a curve connecting two points z0 and z1 in the domain of f(z), then ∫C F'(z) = F(z1) – F(z0)
- If f(z) is an analytic function defined on a disk D, then there is an analytic function F(z) defined on D such that F′(z) = f(z), called a primitive of f(z), and, as a consequence, ∫C f(z)dz =0; for any closed curve C in D.

- If f(z) is an analytic function and if z0 is any point in the domain U of f(z), then the function, [f(z)-f(z0)]/[z – z0] is analytic on U as well.
- If f(z) is an analytic function on a disk D, z0 is a point in the interior of D, C is a closed curve not passing through z0, then W = (C, z0)f(z0) = (1/2π i)∫C [f(z)]/[z – z0]dz, where W(C; z0) is the winding number of C around z

Difference between OLAP and OLTP in DBMS

S.No.	OLAP	OLTP
1	OLAP stands for Online analytical processing.	OLTP stands for online transaction processing.
2	It includes software tools that help in analyzing data mainly for business decisions.	It helps in managing online database modification.
3	It utilizes the data warehouse.	It utilizes traditional approaches of DBMS.
4	It is popular as an online database query management system.	It is popular as an online database modifying system.
5	OLAP employs the data warehouse.	OLTP employs traditional DBMS.
6	It holds old data from various Databases.	It holds current operational data.
7	Here the tables are not normalized.	Here, the tables are normalized.
8	It allows only read and hardly write operations.	It allows both read and write operations.
9	Here, the complex queries are involved.	Here, the queries are simple.

Difference between Data Warehouse and Data Mart

S.N0.	Data Warehouse	Data Mart
1.	It is a centralised system.	It is not a centralised system.
2.	Data warehouse defines a top-down model.	Data mart defines a bottom-up model.
3.	Slightly denormalization is involved in data warehouses.	Highly denormalization is involved in data mart.
4.	It is tough to build a data warehouse.	It is easy to build a data mart.
5.	Fact constellation schema is preferred in data warehouses.	Star schema and snowflake schema are preferred in data mart.
6.	It is more flexible as compared to the data mart.	It is not flexible.
7.	It is data-oriented in behaviour.	Data mart is project-oriented in behaviour.
8.	They mostly have longer life spans.	It has a shorter life span.
9.	Here we get the data in a detailed format.	Here we get the summarised version of data.
10.	The data warehouse is huge in size.	It is smaller as compared to the data warehouse.