

# Knowledge Management

## BCA5001

### Unit-II

**Executive support systems** are intended to be used by the senior managers directly to provide support to non-programmed decisions in strategic management.

These information are often external, unstructured and even uncertain. Exact scope and context of such information is often not known beforehand.

This information is intelligence based –

- Market intelligence
- Investment intelligence
- Technology intelligence

#### Examples of Intelligent Information

Following are some examples of intelligent information, which is often the source of an ESS –

- External databases
- Technology reports like patent records etc.
- Technical reports from consultants
- Market reports
- Confidential information about competitors
- Speculative information like market conditions
- Government policies
- Financial reports and information

#### Features of Executive Information System



## **Advantages of ESS**

- Easy for upper level executive to use
- Ability to analyze trends
- Augmentation of managers' leadership capabilities
- Enhance personal thinking and decision-making
- Contribution to strategic control flexibility
- Enhance organizational competitiveness in the market place
- Instruments of change
- Increased executive time horizons.
- Better reporting system
- Improved mental model of business executive
- Help improve consensus building and communication
- Improve office automation
- Reduce time for finding information
- Early identification of company performance
- Detail examination of critical success factor
- Better understanding
- Time management
- Increased communication capacity and quality

## **Disadvantage of ESS**

- Functions are limited
- Hard to quantify benefits
- Executive may encounter information overload
- System may become slow
- Difficult to keep current data
- May lead to less reliable and insecure data
- Excessive cost for small company

## **What is Artificial Intelligence (AI)?**

In today's world, technology is growing very fast, and we are getting in touch with different new technologies day by day.

Here, one of the booming technologies of computer science is Artificial Intelligence which is ready to create a new revolution in the world by making intelligent machines. The Artificial Intelligence is now all around us. It is currently working with a variety of subfields, ranging from general to specific, such as self-driving cars, playing chess, proving theorems, playing music, Painting, etc.

AI is one of the fascinating and universal fields of Computer science which has a great scope in future. AI holds a tendency to cause a machine to work as a human.



Artificial Intelligence is composed of two words **Artificial** and **Intelligence**, where Artificial defines "*man-made*," and intelligence defines "*thinking power*", hence AI means "*a man-made thinking power*."

So, we can define AI as:

"It is a branch of computer science by which we can create intelligent machines which can behave like a human, think like humans, and able to make decisions."

Artificial Intelligence exists when a machine can have human based skills such as learning, reasoning, and solving problems

With Artificial Intelligence you do not need to preprogram a machine to do some work, despite that you can create a machine with programmed algorithms which can work with own intelligence, and that is the awesomeness of AI.

It is believed that AI is not a new technology, and some people says that as per Greek myth, there were Mechanical men in early days which can work and behave like humans.

### **Why Artificial Intelligence?**

Before Learning about Artificial Intelligence, we should know that what is the importance of AI and why should we learn it. Following are some main reasons to learn about AI:

- With the help of AI, you can create such software or devices which can solve real-world problems very easily and with accuracy such as health issues, marketing, traffic issues, etc.

- With the help of AI, you can create your personal virtual Assistant, such as Cortana, Google Assistant, Siri, etc.
- With the help of AI, you can build such Robots which can work in an environment where survival of humans can be at risk.
- AI opens a path for other new technologies, new devices, and new Opportunities.

## **Goals of Artificial Intelligence**

Following are the main goals of Artificial Intelligence:

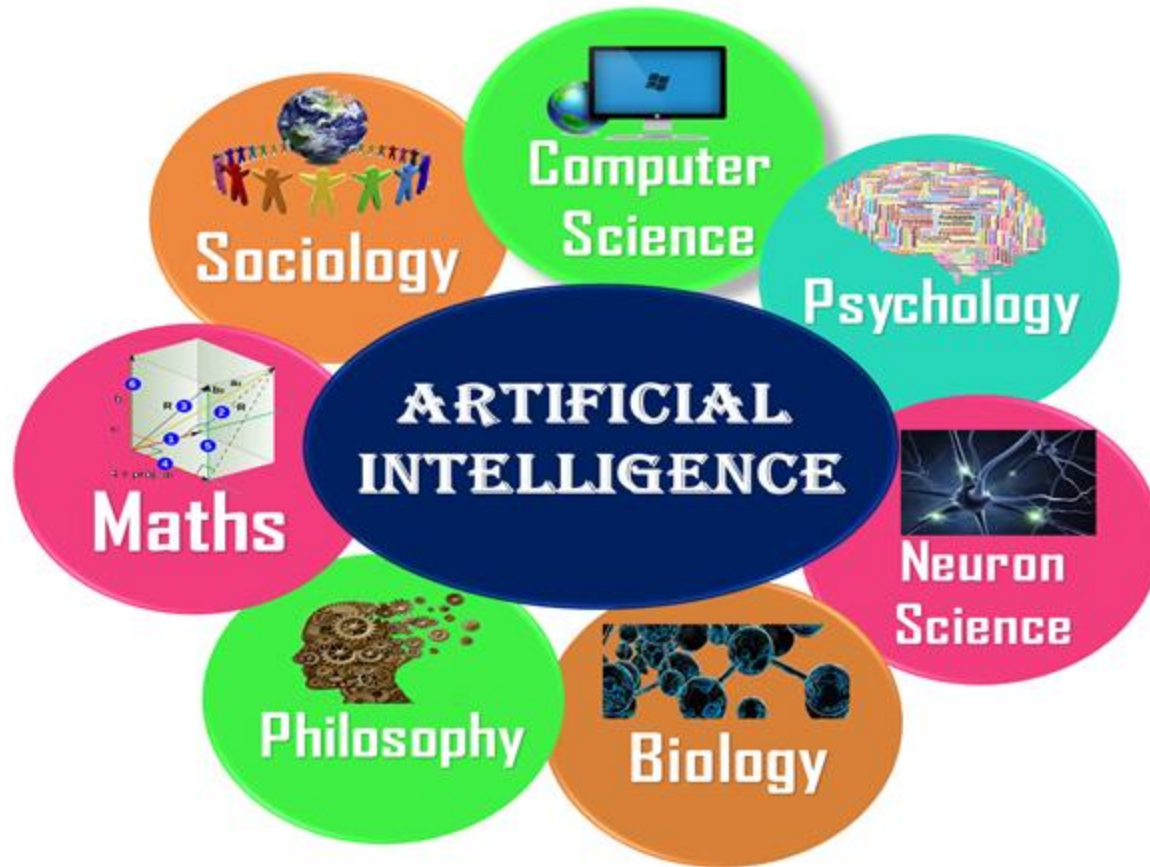
1. Replicate human intelligence
2. Solve Knowledge-intensive tasks
3. An intelligent connection of perception and action
4. Building a machine which can perform tasks that requires human intelligence such as:
  - Proving a theorem
  - Playing chess
  - Plan some surgical operation
  - Driving a car in traffic
5. Creating some system which can exhibit intelligent behavior, learn new things by itself, demonstrate, explain, and can advise to its user.

## **What Comprises to Artificial Intelligence?**

Artificial Intelligence is not just a part of computer science even it's so vast and requires lots of other factors which can contribute to it. To create the AI first we should know that how intelligence is composed, so the Intelligence is an intangible part of our brain which is a combination of **Reasoning, learning, problem-solving perception, language understanding, etc.**

To achieve the above factors for a machine or software Artificial Intelligence requires the following discipline:

- Mathematics
- Biology
- Psychology
- Sociology
- Computer Science
- Neurons Study
- Statistics



## Advantages of Artificial Intelligence

Following are some main advantages of Artificial Intelligence:

- **High Accuracy with less error:** AI machines or systems are prone to less errors and high accuracy as it takes decisions as per pre-experience or information.
- **High-Speed:** AI systems can be of very high-speed and fast-decision making, because of that AI systems can beat a chess champion in the Chess game.
- **High reliability:** AI machines are highly reliable and can perform the same action multiple times with high accuracy.
- **Useful for risky areas:** AI machines can be helpful in situations such as defusing a bomb, exploring the ocean floor, where to employ a human can be risky.
- **Digital Assistant:** AI can be very useful to provide digital assistant to the users such as AI technology is currently used by various E-commerce websites to show the products as per customer requirement.
- **Useful as a public utility:** AI can be very useful for public utilities such as a self-driving car which can make our journey safer and hassle-free, facial recognition for security purpose, Natural language processing to communicate with the human in human-language, etc.

## Disadvantages of Artificial Intelligence

Every technology has some disadvantages, and the same goes for Artificial intelligence. Being so advantageous technology still, it has some disadvantages which we need to keep in our mind while creating an AI system. Following are the disadvantages of AI:

- **High Cost:** The hardware and software requirement of AI is very costly as it requires lots of maintenance to meet current world requirements.
- **Can't think out of the box:** Even we are making smarter machines with AI, but still they cannot work out of the box, as the robot will only do that work for which they are trained, or programmed.
- **No feelings and emotions:** AI machines can be an outstanding performer, but still it does not have the feeling so it cannot make any kind of emotional attachment with human, and may sometime be harmful for users if the proper care is not taken.
- **Increase dependency on machines:** With the increment of technology, people are getting more dependent on devices and hence they are losing their mental capabilities.
- **No Original Creativity:** As humans are so creative and can imagine some new ideas but still AI machines cannot beat this power of human intelligence and cannot be creative and imaginative.

## What is an Expert System?

**“An Intelligent computer program that uses knowledge to solve problems and such knowledge is only possessed by experts”**

Expert system is a branch of artificial intelligence that makes use of specialized knowledge to solve problems at the level of human expert. Expert system takes the role of a human expert and solves the problem like a specialist. It allows the user to work interactively with the computer in order to develop a variety of decisions. Expert system can provide consistent answers. Expert System holds quality information.

### Components of Expert System:

**Knowledge Base:** It consists of relationships rule, data rule which are generally used by experts to solve different types of problems. It consists of professional knowledge possessed either by an expert or a group of experts.

**Inference Engine:** It contains logic and resolving mechanisms.

**User Interface:** This program allows user to communicate with the expert system.

**Explanation facility:** It provides the user with the logic used by the expert system to solve the query.

**Knowledge Acquisition facility:** A good knowledge base requires efforts of both human expert and a knowledge engineer.

### **What is the need for expert systems?**

Major reasons for the need of expert systems are as follows:

- We cannot afford expert labor since it is expensive and scarce.
- Shortage of experts is also a major issue.
- Experts can handle only a few factors at a time.

### **What are the benefits of expert systems?**

- The cost of expertise per user is reduced.
- Experts are available 24\*7. Thus there is an increased availability of experts.
- Expert system assist new user to think in a way as experienced experts do.
- Experts System is not emotional like human beings.
- Expert system may serve as a strategic tool in several areas.
- The expertise is permanent unlike human experts who may retire, quit or die.

### **What are the business applications of expert system?**

- **Marketing:** It helps an organization to determine discounts, quantity and various factors relating to sale.
- **Manufacturing:** It determines whether a process is correct or incorrect and if it is incorrect, then resolving it and correcting it within time.
- **Accounting:** It helps in making financial decisions and credit authorization decisions.
- **Business:** It helps to run the business smoothly and educate trainees and evaluate performance.

### **What are the limitations of expert systems?**

- Lacks common sense needed in some decision making.
- Cannot make corrective decision in unusual cases.
- Expert system may not be able to explain logic and reasoning behind their decisions.
- It may not be able to perform in changing environment.
- Wrong information stored in the knowledge base may lead to wrong decisions.

### **What is the future of expert system?**

The future of expert system is extremely bright. Expert systems are almost used in all aspects of our life. Whenever there is a need to solve structured problem, the name of expert systems will come first. The experts systems will surely solve all the problems faced by us. Expert system will help us to achieve strategic advantage in life.

## **Difference between OLAP and OLTP in DBMS**

OLAP stands for Online Analytical Processing. OLAP systems have the capability to analyze database information of multiple systems at the current time. The primary goal of OLAP Service is data analysis and not data processing.

OLTP stands for Online Transaction Processing. OLTP has the work to administer day-to-day transactions in any organization. The main goal of OLTP is data processing not data analysis.

### **Online Analytical Processing (OLAP)**

Online Analytical Processing (OLAP) consists of a type of software tool that is used for data analysis for business decisions. OLAP provides an environment to get insights from the database retrieved from multiple database systems at one time.

### **OLAP Examples**

Any type of Data Warehouse System is an OLAP system. The uses of the OLAP System are described below.

- Spotify analyzed songs by users to come up with a personalized homepage of their songs and playlist.
- Netflix movie recommendation system.

### **Benefits of OLAP Services**

- OLAP services help in keeping consistency and calculation.
- We can store planning, analysis, and budgeting for business analytics within one platform.
- OLAP services help in handling large volumes of data, which helps in enterprise-level business applications.
- OLAP services help in applying security restrictions for data protection.
- OLAP services provide a multidimensional view of data, which helps in applying operations on data in various ways.

### **Drawbacks of OLAP Services**

- OLAP Services requires professionals to handle the data because of its complex modeling procedure.
- OLAP services are expensive to implement and maintain in cases when datasets are large.
- We can perform an analysis of data only after extraction and transformation of data in the case of OLAP which delays the system.
- OLAP services are not efficient for decision-making, as it is updated on a periodic basis.

### **Online Transaction Processing (OLTP)**



Online transaction processing provides transaction-oriented applications in a 3-tier architecture. OLTP administers the day-to-day transactions of an organization.

### **OLTP Examples**

An example considered for OLTP System is ATM Center a person who authenticates first will receive the amount first and the condition is that the amount to be withdrawn must be present in the ATM. The uses of the OLTP System are described below.

- ATM center is an OLTP application.
- OLTP handles the ACID properties during data transactions via the application.
- It's also used for Online banking, Online airline ticket booking, sending a text message, add a book to the shopping cart.

### **Benefits of OLTP Services**

- OLTP services allow users to read, write and delete data operations quickly.
- OLTP services help in increasing users and transactions which helps in real-time access to data.
- OLTP services help to provide better security by applying multiple security features.
- OLTP services help in making better decision making by providing accurate data or current data.
- OLTP Services provide Data Integrity, Consistency, and High Availability to the data.

### **Drawbacks of OLTP Services**

- OLTP has limited analysis capability as they are not capable of intending complex analysis or reporting.
- OLTP has high maintenance costs because of frequent maintenance, backups, and recovery.
- OLTP Services get hampered in the case whenever there is a hardware failure which leads to the failure of online transactions.
- OLTP Services many times experience issues such as duplicate or inconsistent data.

### **Difference between OLAP and OLTP**

<b>Category</b>	<b>OLAP (Online Analytical Processing)</b>	<b>OLTP (Online Transaction Processing)</b>
<b>Definition</b>	It is well-known as an online database query management system.	It is well-known as an online database modifying system.

<b>Category</b>	<b>OLAP (Online Analytical Processing)</b>	<b>OLTP (Online Transaction Processing)</b>
Data source	Consists of historical data from various Databases.	Consists of only operational current data.
Method used	It makes use of a data warehouse.	It makes use of a standard database management system (DBMS).
Application	It is subject-oriented. Used for Data Mining, Analytics, Decisions making, etc.	It is application-oriented. Used for business tasks.
Normalized	In an OLAP database, tables are not normalized.	In an OLTP database, tables are normalized (3NF).
Usage of data	The data is used in planning, problem-solving, and decision-making.	The data is used to perform day-to-day fundamental operations.
Task	It provides a multi-dimensional view of different business tasks.	It reveals a snapshot of present business tasks.
Purpose	It serves the purpose to extract information for analysis and decision-making.	It serves the purpose to Insert, Update, and Delete information from the database.
Volume of data	A large amount of data is stored typically in TB, PB	The size of the data is relatively small as the historical data is archived in MB, and GB.
Queries	Relatively slow as the amount of data involved is large. Queries	Very Fast as the queries operate on 5%

<b>Category</b>	<b>OLAP (Online Analytical Processing)</b>	<b>OLTP (Online Transaction Processing)</b>
	may take hours.	of the data.
Update	The OLAP database is not often updated. As a result, data integrity is unaffected.	The data integrity constraint must be maintained in an OLTP database.
Backup and Recovery	It only needs backup from time to time as compared to OLTP.	The backup and recovery process is maintained rigorously
Processing time	The processing of complex queries can take a lengthy time.	It is comparatively fast in processing because of simple and straightforward queries.
Types of users	This data is generally managed by CEO, MD, and GM.	This data is managed by clerks, Forex and managers.
Operations	Only read and rarely write operations.	Both read and write operations.
Updates	With lengthy, scheduled batch operations, data is refreshed on a regular basis.	The user initiates data updates, which are brief and quick.
Nature of audience	The process is focused on the customer.	The process is focused on the market.
Database Design	Design with a focus on the subject.	Design that is focused on the application.

<b>Category</b>	<b>OLAP (Online Analytical Processing)</b>	<b>OLTP (Online Transaction Processing)</b>
Productivity	Improves the efficiency of business analysts.	Enhances the user's productivity.

### **What is a Data Warehouse?**

A Data Warehouse (DW) is a relational database that is designed for query and analysis rather than transaction processing. It includes historical data derived from transaction data from single and multiple sources.

A Data Warehouse provides integrated, enterprise-wide, historical data and focuses on providing support for decision-makers for data modeling and analysis.

A Data Warehouse is a group of data specific to the entire organization, not only to a particular group of users.

It is not used for daily operations and transaction processing but used for making decisions.

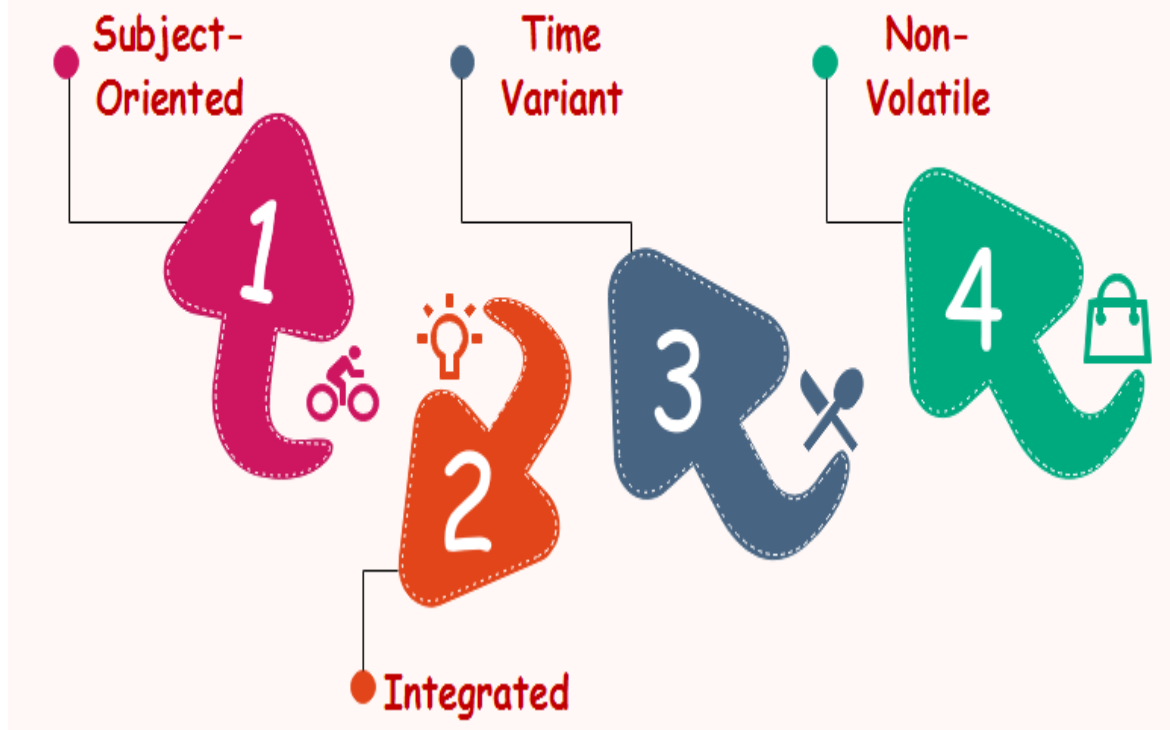
A Data Warehouse can be viewed as a data system with the following attributes:

- It is a database designed for investigative tasks, using data from various applications.
- It supports a relatively small number of clients with relatively long interactions.
- It includes current and historical data to provide a historical perspective of information.
- Its usage is read-intensive.
- It contains a few large tables.

"Data Warehouse is a subject-oriented, integrated, and time-variant store of information in support of management's decisions."

### **Characteristics of Data Warehouse**

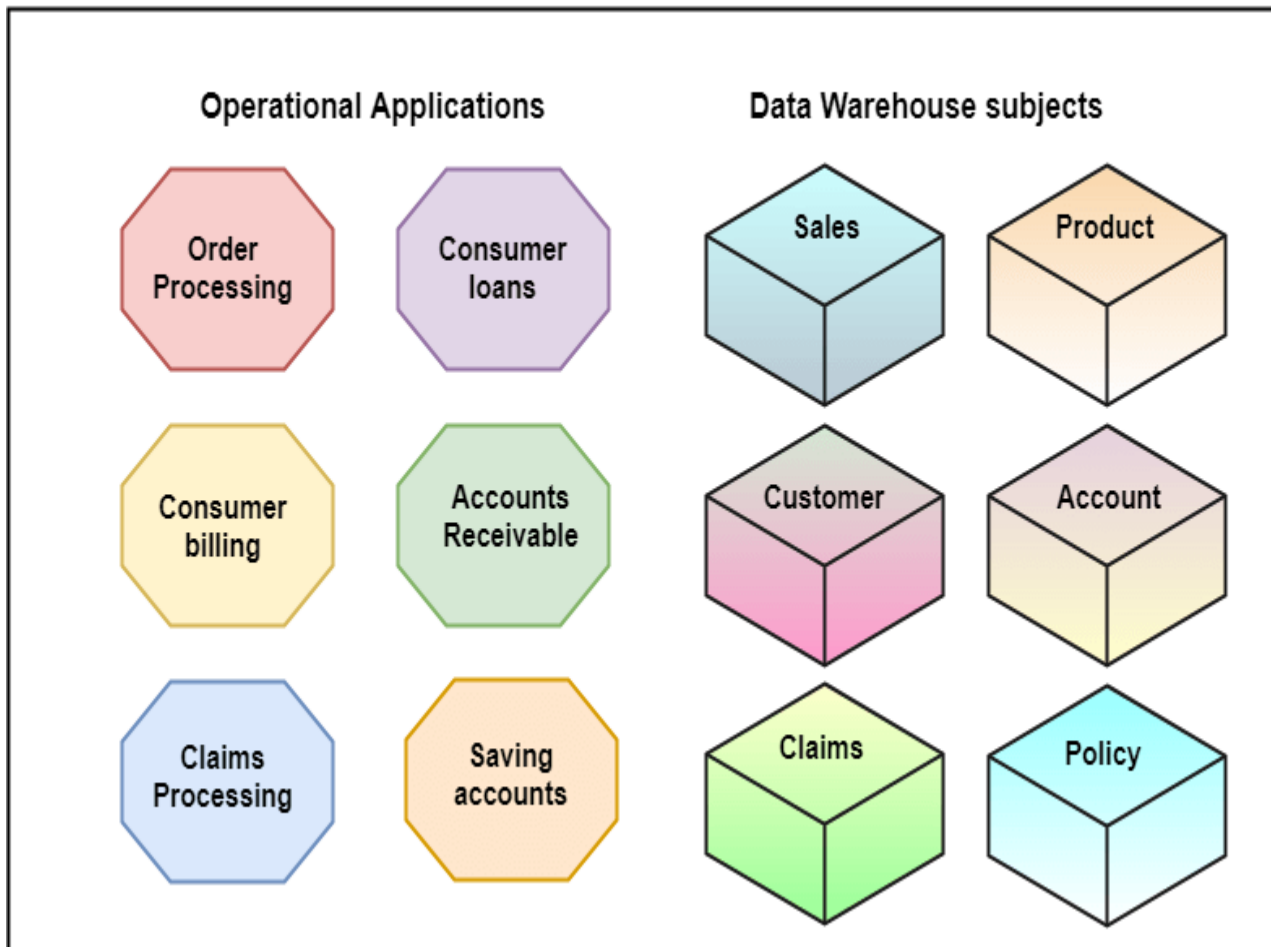
## The key features of Data Warehouse are:



### **Subject-Oriented**

A data warehouse target on the modeling and analysis of data for decision-makers. Therefore, data warehouses typically provide a concise and straightforward view around a particular subject, such as customer, product, or sales, instead of the global organization's ongoing operations. This is done by excluding data that are not useful concerning the subject and including all data needed by the users to understand the subject.

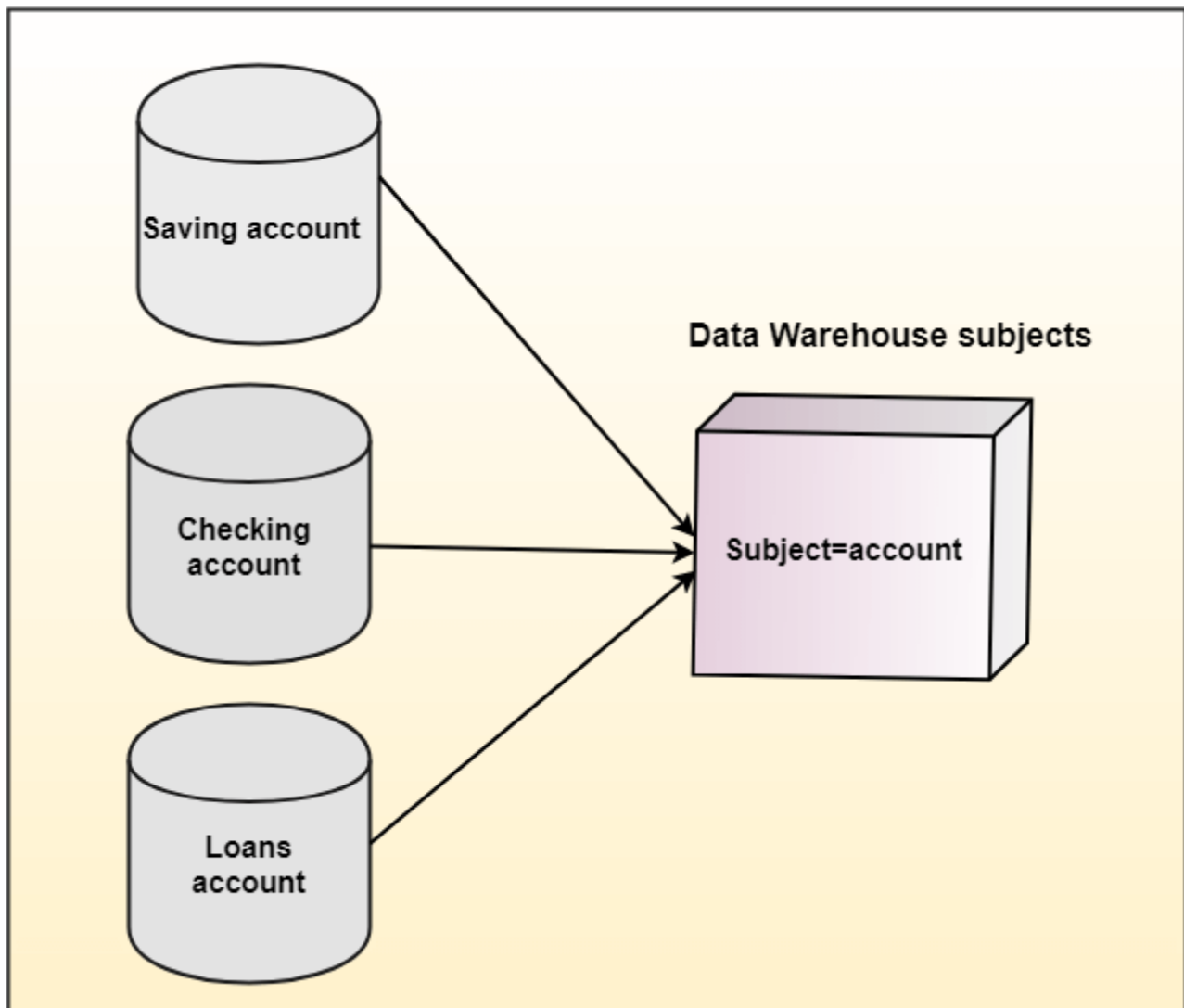
## Data Warehouse is Subject-Oriented



### Integrated

A data warehouse integrates various heterogeneous data sources like RDBMS, flat files, and online transaction records. It requires performing data cleaning and integration during data warehousing to ensure consistency in naming conventions, attributes types, etc., among different data sources.

## Data Warehouse is Integrated



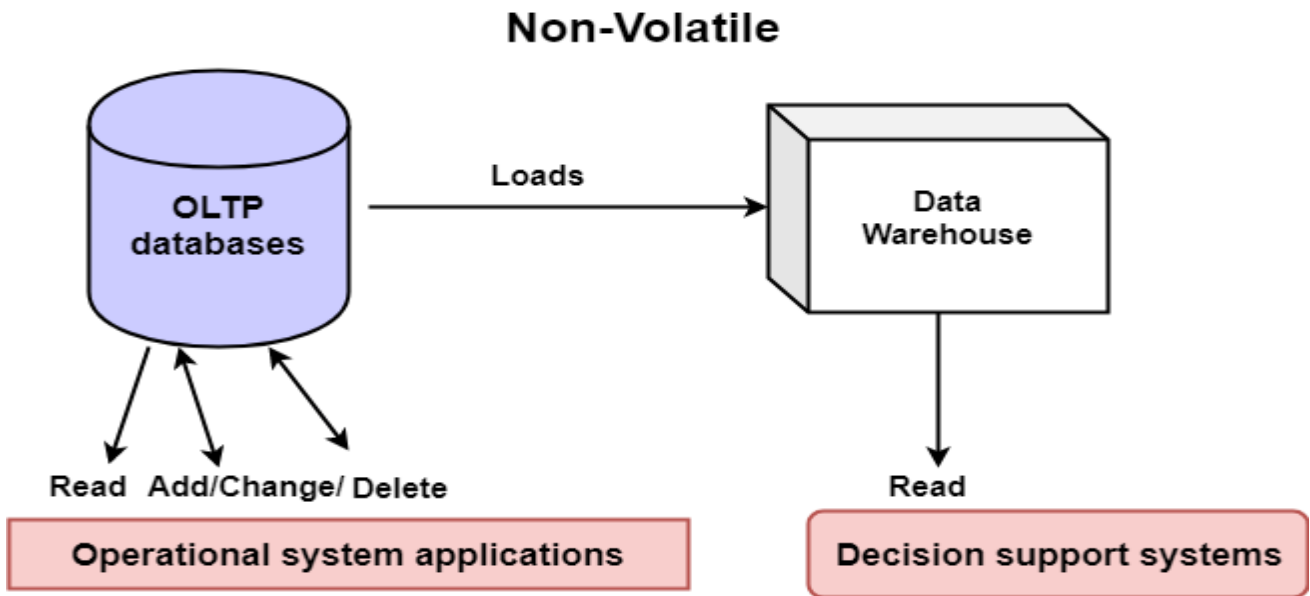
### Time-Variant

Historical information is kept in a data warehouse. For example, one can retrieve files from 3 months, 6 months, 12 months, or even previous data from a data warehouse. These variations with a transaction system, where often only the most current file is kept.



### Non-Volatile

The data warehouse is a physically separate data storage, which is transformed from the source operational RDBMS. The operational updates of data do not occur in the data warehouse, i.e., update, insert, and delete operations are not performed. It usually requires only two procedures in data accessing: Initial loading of data and access to data. Therefore, the DW does not require transaction processing, recovery, and concurrency capabilities, which allows for substantial speedup of data retrieval. Non-Volatile defines that once entered into the warehouse, and data should not change.





## History of Data Warehouse

The idea of data warehousing came to the late 1980's when IBM researchers Barry Devlin and Paul Murphy established the "Business Data Warehouse."

In essence, the data warehousing idea was planned to support an architectural model for the flow of information from the operational system to decisional support environments. The concept attempt to address the various problems associated with the flow, mainly the high costs associated with it.

In the absence of data warehousing architecture, a vast amount of space was required to support multiple decision support environments. In large corporations, it was ordinary for various decision support environments to operate independently.

## Goals of Data Warehousing

- To help reporting as well as analysis
- Maintain the organization's historical information
- Be the foundation for decision making.

## Need for Data Warehouse

Data Warehouse is needed for the following reasons:



1) **Business User:** Business users require a data warehouse to view summarized data from the past. Since these people are non-technical, the data may be presented to them in an elementary form.

2) **Store historical data:** Data Warehouse is required to store the time variable data from the past. This input is made to be used for various purposes.

3) **Make strategic decisions:** Some strategies may be depending upon the data in the data warehouse. So, data warehouse contributes to making strategic decisions.

4) **For data consistency and quality:** Bringing the data from different sources at a commonplace, the user can effectively undertake to bring the uniformity and consistency in data.

5) **High response time:** Data warehouse has to be ready for somewhat unexpected loads and types of queries, which demands a significant degree of flexibility and quick response time.

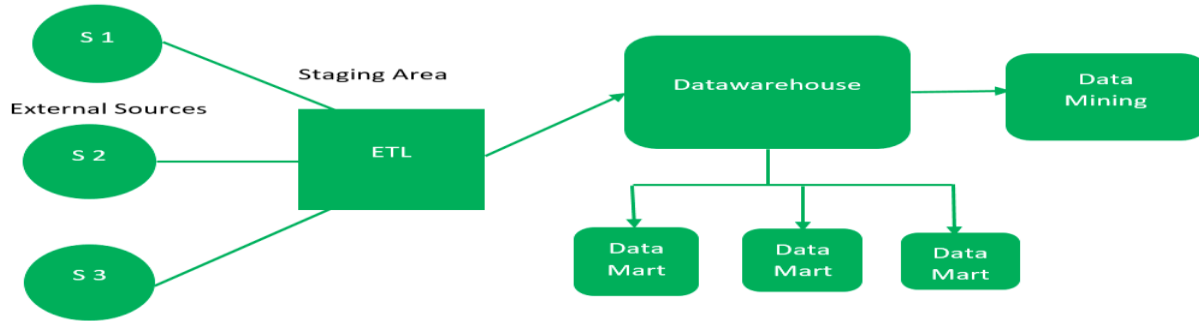
### **Benefits of Data Warehouse**

1. Understand business trends and make better forecasting decisions.
2. Data Warehouses are designed to perform well enormous amounts of data.
3. The structure of data warehouses is more accessible for end-users to navigate, understand, and query.
4. Queries that would be complex in many normalized databases could be easier to build and maintain in data warehouses.
5. Data warehousing is an efficient method to manage demand for lots of information from lots of users.
6. Data warehousing provide the capabilities to analyze a large amount of historical data.

### **Data Warehouse Architecture**

A **data-warehouse** is a heterogeneous collection of different data sources organised under a unified schema. There are 2 approaches for constructing data-warehouse: Top-down approach and Bottom-up approach are explained as below.

#### **1. Top-down approach:**



The essential components are discussed below:

1. **External Sources** –

External source is a source from where data is collected irrespective of the type of data. Data can be structured, semi structured and unstructured as well.

2. **Stage Area** –

Since the data, extracted from the external sources does not follow a particular format, so there is a need to validate this data to load into datawarehouse. For this purpose, it is recommended to use **ETL** tool.

- **E(Extracted):** Data is extracted from External data source.
- **T(Transform):** Data is transformed into the standard format.
- **L(Load):** Data is loaded into datawarehouse after transforming it into the standard format.

3. **Data-warehouse** –

After cleansing of data, it is stored in the datawarehouse as central repository. It actually stores the meta data and the actual data gets stored in the data marts. **Note** that datawarehouse stores the data in its purest form in this top-down approach.

4. **Data Marts** –

Data mart is also a part of storage component. It stores the information of a particular function of an organisation which is handled by single authority. There can be as many number of data marts in an organisation depending upon the functions. We can also say that data mart contains subset of the data stored in datawarehouse.

5. **Data Mining** –

The practice of analysing the big data present in datawarehouse is data mining. It is used to find the hidden patterns that are present in the database or in datawarehouse with the help of algorithm of data mining.

This approach is defined by **Inmon** as – datawarehouse as a central repository for the complete organisation and data marts are created from it after the complete datawarehouse has been created.

### **Advantages of Top-Down Approach –**

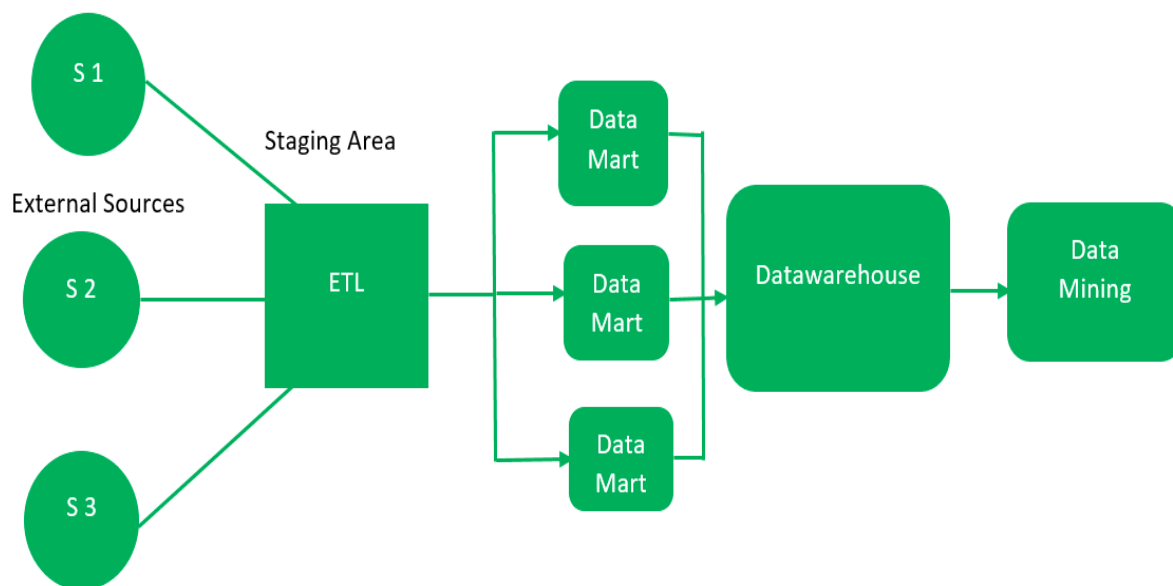
1. Since the data marts are created from the datawarehouse, provides consistent dimensional view of data marts.
2. Also, this model is considered as the strongest model for business changes. That's why, big organisations prefer to follow this approach.
3. Creating data mart from datawarehouse is easy.
4. Improved data consistency: The top-down approach promotes data consistency by ensuring that all data marts are sourced from a common data warehouse. This ensures that all data is standardized, reducing the risk of errors and inconsistencies in reporting.
5. Easier maintenance: Since all data marts are sourced from a central data warehouse, it is easier to maintain and update the data in a top-down approach. Changes can be made to the data warehouse, and those changes will automatically propagate to all the data marts that rely on it.
6. Better scalability: The top-down approach is highly scalable, allowing organizations to add new data marts as needed without disrupting the existing infrastructure. This is particularly important for organizations that are experiencing rapid growth or have evolving business needs.
7. Improved governance: The top-down approach facilitates better governance by enabling centralized control of data access, security, and quality. This ensures that all data is managed consistently and that it meets the organization's standards for quality and compliance.
8. Reduced duplication: The top-down approach reduces data duplication by ensuring that data is stored only once in the data warehouse. This saves storage space and reduces the risk of data inconsistencies.
9. Better reporting: The top-down approach enables better reporting by providing a consistent view of data across all data marts. This makes it easier to create accurate and timely reports, which can improve decision-making and drive better business outcomes.
10. Better data integration: The top-down approach enables better data integration by ensuring that all data marts are sourced from a common data warehouse. This makes it easier to integrate data from different sources and provides a more complete view of the organization's data.

### **Disadvantages of Top-Down Approach –**

1. The cost, time taken in designing and its maintenance is very high.
2. Complexity: The top-down approach can be complex to implement and maintain, particularly for large organizations with complex data needs. The design and implementation of the data warehouse and data marts can be time-consuming and costly.
3. Lack of flexibility: The top-down approach may not be suitable for organizations that require a high degree of flexibility in their data reporting and analysis. Since the design of the data warehouse and data marts is pre-determined, it may not be possible to adapt to new or changing business requirements.

4. Limited user involvement: The top-down approach can be dominated by IT departments, which may lead to limited user involvement in the design and implementation process. This can result in data marts that do not meet the specific needs of business users.
5. Data latency: The top-down approach may result in data latency, particularly when data is sourced from multiple systems. This can impact the accuracy and timeliness of reporting and analysis.
6. Data ownership: The top-down approach can create challenges around data ownership and control. Since data is centralized in the data warehouse, it may not be clear who is responsible for maintaining and updating the data.
7. Cost: The top-down approach can be expensive to implement and maintain, particularly for smaller organizations that may not have the resources to invest in a large-scale data warehouse and associated data marts.
8. Integration challenges: The top-down approach may face challenges in integrating data from different sources, particularly when data is stored in different formats or structures. This can lead to data inconsistencies and inaccuracies.

## 2. Bottom-up approach:



1. First, the data is extracted from external sources (same as happens in top-down approach).
2. Then, the data go through the staging area (as explained above) and loaded into data marts instead of datawarehouse. The data marts are created first and provide reporting capability. It addresses a single business area.
3. These data marts are then integrated into datawarehouse.

This approach is given by **Kinball** as – data marts are created first and provide a thin view for analyses and datawarehouse is created after complete data marts have been created.

### **Advantages of Bottom-Up Approach –**

1. As the data marts are created first, so the reports are quickly generated.
2. We can accommodate more number of data marts here and in this way datawarehouse can be extended.
3. Also, the cost and time taken in designing this model is low comparatively.
4. Incremental development: The bottom-up approach supports incremental development, allowing for the creation of data marts one at a time. This allows for quick wins and incremental improvements in data reporting and analysis.
5. User involvement: The bottom-up approach encourages user involvement in the design and implementation process. Business users can provide feedback on the data marts and reports, helping to ensure that the data marts meet their specific needs.
6. Flexibility: The bottom-up approach is more flexible than the top-down approach, as it allows for the creation of data marts based on specific business needs. This approach can be particularly useful for organizations that require a high degree of flexibility in their reporting and analysis.
7. Faster time to value: The bottom-up approach can deliver faster time to value, as the data marts can be created more quickly than a centralized data warehouse. This can be particularly useful for smaller organizations with limited resources.
8. Reduced risk: The bottom-up approach reduces the risk of failure, as data marts can be tested and refined before being incorporated into a larger data warehouse. This approach can also help to identify and address potential data quality issues early in the process.
9. Scalability: The bottom-up approach can be scaled up over time, as new data marts can be added as needed. This approach can be particularly useful for organizations that are growing rapidly or undergoing significant change.
10. Data ownership: The bottom-up approach can help to clarify data ownership and control, as each data mart is typically owned and managed by a specific business unit. This can help to ensure that data is accurate and up-to-date, and that it is being used in a consistent and appropriate way across the organization.

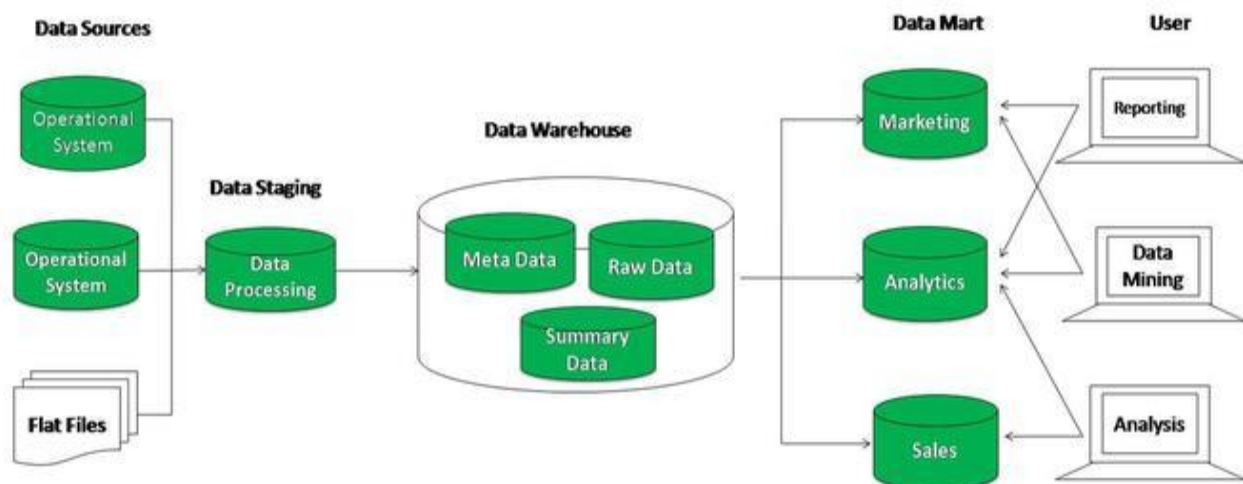
### **Disadvantage of Bottom-Up Approach –**

1. This model is not strong as top-down approach as dimensional view of data marts is not consistent as it is in above approach.
2. Data silos: The bottom-up approach can lead to the creation of data silos, where different business units create their own data marts without considering the needs of other parts of the organization. This can lead to inconsistencies and redundancies in the data, as well as difficulties in integrating data across the organization.
3. Integration challenges: Because the bottom-up approach relies on the integration of multiple data marts, it can be more difficult to integrate data from different sources and ensure consistency across the organization. This can lead to issues with data quality and accuracy.
4. Duplication of effort: In a bottom-up approach, different business units may duplicate effort by creating their own data marts with similar or overlapping data. This can lead to inefficiencies and higher costs in data management.

5. Lack of enterprise-wide view: The bottom-up approach can result in a lack of enterprise-wide view, as data marts are typically designed to meet the needs of specific business units rather than the organization as a whole. This can make it difficult to gain a comprehensive understanding of the organization's data and business processes.
6. Complexity: The bottom-up approach can be more complex than the top-down approach, as it involves the integration of multiple data marts with varying levels of complexity and granularity. This can make it more difficult to manage and maintain the data warehouse over time.
7. Risk of inconsistency: Because the bottom-up approach allows for the creation of data marts with different structures and granularities, there is a risk of inconsistency in the data. This can make it difficult to compare data across different parts of the organization or to ensure that reports are accurate and reliable.

## Popular Data Warehouse Tools

A data warehouse is a Data management system that is used for storing, reporting, and data analysis. It is the primary component of business intelligence and is also known as an enterprise data warehouse. Data Warehouses are central repositories that store data from one or more heterogeneous sources. Data warehouses are analytical tools built to support decision-making for reporting users across many departments. Data warehouse works to create a single, unified system of truth for an entire organization and store historical data about business and organization so that it could be analyzed and extract insights from it.



### *Data Flow through Warehouse Architecture*

Previously, organizations had to build lots of infrastructure for data warehousing but today, cloud computing technology has amazingly reduced the efforts as well as the cost of building data warehousing for businesses. Data warehouses and their tools are moving from physical data

centers to cloud-based data warehouses. Many large organizations still operate data through the traditional way of data warehousing but clearly, the future of the data warehouse is in the cloud. The cloud-based data warehousing tools are fast, efficient, highly scalable, and available based on pay-per-use.

There are various cloud-based Data Warehousing tools available. So, it becomes difficult to select top Data Warehouse tools according to our project requirements. Following are the top 8 Data Warehousing tools:

### **1. Amazon Redshift:**

Amazon Redshift is a cloud-based fully managed petabytes-scale data warehouse By the Amazon Company. It starts with just a few hundred gigabytes of data and scales to petabytes or more. This enables the use of data to accumulate new insights for businesses and customers. It is a relational database management system (RDBMS) therefore it is compatible with other RDBMS applications. Amazon Redshift offers quick querying capabilities over structured data by the use of SQL-based clients and business intelligence (BI) tools using standard ODBC and JDBC connections. Amazon Redshift is made around industry-standard SQL, with additional practicality to manage massive datasets and support superior analysis and reporting of these data. It helps to work quickly and easily along with data in open formats, and simply integrates with and connects to the AWS scheme. Also query and export data to and from the data lake. No alternative cloud data warehouse tool makes it straightforward to query data and writes data back to the data lake in open formats. It focuses on simple Use and Accessibility. MySQL and alternative SQL-based systems are one in all the foremost well-liked and simply usable interfaces for database management. Redshift's easy query-based system makes platform adoption and acclimatization a light breeze. It is incredibly quick once it involves loading data and querying it for analytical and reporting functions. Redshift features a massively parallel processing (MPP) design that permits loading data at a very high speed.

### **2. Microsoft Azure:**

Azure is a cloud computing platform that was launched by Microsoft in 2010. Microsoft Azure is a cloud computing service provider for building, testing, deploying, and managing applications and services through Microsoft-managed data centers. Azure is a public cloud computing platform that offers Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). The Azure cloud platform provides more than 200 products and cloud services such as Data Analytics, Virtual Computing, Storage, Virtual Network, Internet Traffic Manager, Web Sites, Media Services, Mobile Services, Integration, etc. Azure facilitates simple portability and genuinely compatible platform between on-premise and public Cloud. Azure provides a range of cross-connections including virtual private networks (VPNs), caches, content delivery networks (CDNs), and ExpressRoute connections to improve usability and performance. Microsoft Azure provides a secure base across physical infrastructure and operational security. Azure App offers a completely managed web hosting service that helps in building web applications, services, and Restful APIs. It offers a variety of plans to meet the requirements of any application, from small to globally scaled web applications. Running virtual machines or containers in the cloud is one of the most popular applications of Microsoft Azure.



### **3. Google BigQuery:**

BigQuery is a serverless data warehouse that allows scalable analysis over petabytes of data. It's a Platform as a Service that supports querying with the help of ANSI SQL. It additionally has inbuilt machine learning capabilities. BigQuery was declared in 2010 and made available for use there in 2011. Google BigQuery is a cloud-based big data analytics web service to process very huge amount of read-only data sets. BigQuery is designed for analyzing data that are in billions of rows by simply employing SQL-lite syntax. BigQuery can run advanced analytical SQL-based queries beneath big sets of data. BigQuery is not developed to substitute relational databases and for easy CRUD operations and queries. It is oriented for running analytical queries. It is a hybrid system that enables the storage of information in columns; however, it takes into the NoSQL additional features, like the data type, and the nested feature. BigQuery is a better option than Redshift since we have to pay by the hour. BigQuery may also be the best solution for data scientists running ML or data mining operations since they deal with extremely large datasets. Google Cloud also offers a set of auto-scaling services that enables you to build a data lake that integrates with your existing applications, skills, and IT investments. In BigQuery, most of the time is spent on metadata/initiation, but the actual execution time is very small.

### **4. Snowflake:**

Snowflake is a cloud computing-based data warehousing built on top of the Amazon Web Services or Microsoft Azure cloud infrastructure. The Snowflake design allows storage and computes to scale independently, thus customers can use and pay money for storage and computation individually. In Snowflake data processing is simplified: Users will do data blending, analysis, and transformations against varied forms of data structures with one language, SQL. Snowflake offers dynamic, scalable computing power with charges primarily based strictly on usage. With Snowflake, computation and storage are fully separate, and also the storage value is that the same as storing the data on Amazon S3. AWS tried to handle this issue by introducing Redshift Spectrum, which allows querying data that exists directly on Amazon S3; however, it's not as seamless as Snowflake. With Snowflake, we can clone a table, a schema, or perhaps a database in no time and occupying no extra space. This is often because the cloned table creates pointers that point to the kept data, however, not the actual data. In alternative words, the cloned table solely has data that's completely different from its original table.