DA - III

1. No SQL -
2. A No-SQL (not only SQL) database provides a mechanism for storage and retrieval of data that is modeled in document based format or key value pair other than the traditional tabular relations used in relational databases.

B- No-SQL is schema-free.

C- Relational databases are not designed to **scale and change easily** to cope up with the needs of the modern industry to be **frequently stored and accessed**. No-SQL uses **distributed architecture and works on multiple processors to give high performance**.

1. No-SQL may not provide atomicity, consistency, isolation, durability (ACID) properties but guarantees eventual consistency, basically available, soft state (BASE/refreshing), by having a distributed and fault-tolerant architecture.
2. Key Value -

- Data is stored in key/value pairs.

- It is designed in such a way to handle lots of data and heavy load.

- Key-value pair storage databases store data as a hash table where each key is unique, and the value can be a JSON, BLOB, string, etc.

- It is one of the most basic types of NoSQL databases.

- This kind of NoSQL database is used as a collection, dictionaries, associative arrays, etc.

- Key value stores help the developer to store schema-less data. They work best for shopping cart contents.

1. Document Based -
2. Column Family Store/Wide Column Store -
3. Types of Data Analystics -
4. Descriptive -

- Descriptive analysis **involves summarizing and describing** data to **provide a better understanding of what has happened in the past**. The key features of descriptive analysis are:

- It focuses on **describing** the data rather than **explaining** it

- It is used to **gain insights into historical trends and patterns**

- It is often used to **create dashboards** **and reports to communicate insights to stakeholders**

Example: A retail store analyzes its sales data to understand which products are selling the most, which regions are performing well, and which promotions are driving sales.

1. Diagnostic Analysis -

- Diagnostic analysis **involves examining data to determine why something happened in the past**. The key features of diagnostic analysis are:

- It focuses on **identifying the root cause of a problem**

- It is used to **gain insights into why a particular outcome occurred.**

- Data mining, Data discovery, Correlations

- It is often used **to identify the factors that led to a particular trend or pattern.**

Example: A hospital analyzes its patient data to understand why the readmission rate for a particular disease is higher than expected. The analysis may reveal that patients are not receiving proper follow-up care after being discharged.

1. Prescriptive Analysis -

- Prescriptive analysis involves using data to provide recommendations on what action should be taken in the future. The key features of prescriptive analysis are:

- It focuses on using data to make decisions that will impact the future

- It is used to identify the best course of action to achieve a particular outcome

- It often involves the **use of predictive models to forecast the outcomes of different actions**

Example: A bank uses predictive models to identify which customers are most likely to default on their loans. The bank can then use this information to develop a targeted intervention strategy to prevent defaults from occurring.

- This is like ‘What I need to do’ - Recommend action or strategies based on insights from historical data.

1. Predictive Analysis -

Predictive analysis is a type of analysis in data analystics that **involves using statistical and machine learning algorithms to analyze historical data and make predictions about future events.**

- **Uses historical data**: Predictive analysis uses historical data to identify patterns and relationships that can be used to predict future events.

- **Statistical and machine learning techniques**: Predictive analysis involves using statistical and machine learning techniques to develop models that can predict future outcomes based on historical data. Helps make data-driven decisions: Predictive analysis provides insights that can help organizations make data-driven decisions about future events and trends.

- **Continuous learning**: Predictive models can be refined and improved over time by incorporating new data and feedback.

Example: An e-commerce company uses predictive analysis to forecast customer demand for a particular product. By analyzing historical sales data and other relevant factors such as pricing and promotions, the company can develop a predictive model that accurately forecasts future demand. This allows the company to optimize inventory levels, reduce waste, and improve customer satisfaction by ensuring that popular products are always in stock.

- it predicts 'what is likely to happen'

1. Two types of Visualization -
2. Info Graphics - **visual representation of data or information rapidly.**

- Best for telling a premeditated story and offer subjectivity.

- Best for guiding the audience to conclusions and point out relationships.

- **Created manually for one specific dataset.**

- It is used for Marketing content, Resumes, Blog posts, and Case studies, etc.

1. Data visualization -

- study of representation of data in visual format.

- Best for allowing the audience to draw their own conclusions, and offer objectivity

- **Ideal for understanding data at a glance automatically generated for arbitrary datasets.**

- **It is used for Dashboards, Scorecards, Newsletters, Reports, and Editorials etc.**

Difference - Info for subjective analysis(tells a planned story to guide the audience to draw conclusions) and Data visualization is objective, let the audience draw conclusions from it.

An info can contain data visualizations but not the other way round.

1. Purpose of data visualization

* Data visualization helps analysts and stakeholders to communicate insights and findings in a clear and concise way. By presenting complex data in a visual format, data visualization can help to simplify complex concepts and make it easier for stakeholders to understand the insights and findings.
* Identify trends and patterns: Data visualization helps analysts to identify trends and patterns that might be difficult to see in a table or spreadsheet format. By representing data visually, analysts can quickly spot trends and patterns that might not be immediately apparent in a traditional data format.
* Explore data and generate hypotheses: Data visualization helps analysts to explore data and generate hypotheses about the relationships between different variables. By visualizing data in different ways, analysts can uncover hidden relationships and identify areas for further analysis.

1. Analysis and Reporting -
2. CAP in Brewer’s theorem -

A - **Consistency refers to the property that all nodes in a distributed system see the same data at the same time. In other words, when a client reads data from the system, it should receive the most up-to-date version of that data, even if it is being accessed from different nodes at the same time.**

**Availability refers to the property that the system is always available to respond to requests, even in the face of failures. This means that if one node in the system fails or goes offline, the system should still be able to respond to requests from other nodes.**

**Partition tolerance refers to the ability of the system to continue to function even if network partitions occur. A network partition occurs when a subset of nodes in the system are unable to communicate with the rest of the nodes due to network failures or other issues. Partition tolerance means that the system can continue to operate despite these partitions, and that data will eventually be reconciled when the partitions are resolved.**

No SQL Properties -

1.Volume

The need to horizontal scaling made organizations to

move from serial to distributed parallel processing where

big data is fragmented and processed using clusters of

commodity machines.

2.Velocity

Queries have to be read and written in real time. Random

bursts in web traffic slows down the response for every

user in relational databases.

3.Variability

Uncommon data is frequent.

Special attributes lead to sparse matrix.

ALTER TABLE can not be executed when transactions

are happening continuously.

Advantages -

Can be used as Primary or Analytic Data Source

Big Data Capability

No Single Point of Failure

Easy Replication

Disadvantages -

No standardization rules

Limited query capabilities