

## Task 5

### **What is Historical Load?**

Historical load in data engineering refers to the process of loading historical data into a data warehouse or data lake. This type of data load is done when there is a need to analyze past data over a long period of time, usually for trend analysis, forecasting or reporting purposes.

Historical data can come from a variety of sources, including legacy systems, historical backups, and data archives. This data is often unstructured and may require extensive cleaning and transformation before it can be loaded into a data warehouse or data lake.

The historical load process typically involves extracting the data from its source, transforming it into a suitable format for analysis, and loading it into the data warehouse or data lake. The data may be loaded in batches or all at once, depending on the volume of data and the available resources.

Historical load is an important aspect of data engineering as it allows organizations to gain insights into long-term trends and patterns that can inform decision-making. By analyzing historical data alongside real-time data, businesses can make informed decisions about their operations, identify areas for improvement, and develop data-driven strategies for growth.

### **What is Full Load?**

Full load is a data integration process that involves loading all the data from the source system into the target system. This means that all the existing data in the target system is replaced with the data from the source system. Full load is typically used for the initial load of data into a system or when a significant change has occurred in the source system that requires a complete refresh of the target system.

During a full load, all the data from the source system is extracted, transformed, and loaded into the target system. This process can be time-consuming and resource-intensive, especially for large datasets. However, it ensures that the target system has the most up-to-date and accurate data from the source system.

Full load is often used in conjunction with incremental loads, which involve only loading new or changed data into the target system since the last load. This approach helps to minimize the amount of data that needs to be processed and loaded, which can be more efficient and faster than a full load.

### **What is Incremental Load?**

Incremental load is a process of adding new or updated data to an existing data set. Unlike full load, which replaces the entire data set, incremental load only updates the changes or

additions to the data. This is typically done by comparing the source system's data with the target data set and identifying the new or updated data.

Incremental loads are useful when dealing with large data sets, as updating the entire data set can be time-consuming and resource-intensive. It is commonly used in situations where data changes frequently and needs to be updated in near-real-time, such as stock prices, social media feeds, or weather data.

The incremental load process typically involves the use of tools or scripts that automate the process of identifying and updating the new or changed data. These tools can also include features such as error handling and logging to ensure the integrity and accuracy of the data. Incremental load can be scheduled to run at regular intervals, such as daily or hourly, depending on the frequency of data changes.

Overall, incremental load is a key component of data integration and data warehousing, as it helps keep data up-to-date and accurate while minimizing the time and resources required for data updates.