**Bytewise DE Task**

**BY: Mohammad Yaqoob**

**Date: March/16/2023**

**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 for analysis and reporting purposes. This involves extracting data from various sources such as databases, files, and APIs, transforming it into a format suitable for analysis, and loading it into a data storage system.

Historical data is typically stored in a time-series format, with each record representing a specific point in time. This allows for analysis of trends and patterns over time, as well as comparison of current data with historical data.

Historical load is important in data engineering because it enables organizations to gain insights into past performance, identify trends, and make informed decisions based on historical data. It also enables organizations to meet compliance requirements by maintaining a record of historical data.

* What is Full Load

Full load in data engineering refers to the process of loading all the data from source systems into a target system, typically a data warehouse or a data lake. This is typically done when setting up a new data storage system or when there has been a major change in the data source that requires a full refresh of the data.

Full load involves extracting all the data from the source systems, transforming it into a format suitable for storage and analysis, and then loading it into the target system. This process can be time-consuming and resource-intensive, especially if the amount of data to be loaded is large. It is also important to ensure that the data loaded into the target system is accurate and complete.

Full load is typically followed by incremental loads, where only new or changed data is loaded into the target system. Incremental loads are faster and less resource-intensive than full loads and are used to keep the target system up to date with the latest data from the source systems.

* What is Incremental Load

Incremental load in data engineering refers to the process of loading only the new or changed data into a data storage system, such as a data warehouse or a data lake. This approach is used to keep the target system up to date with the latest data from the source systems, without having to reload all the data every time.

The incremental load process typically involves comparing the data in the source system with the data already loaded in the target system, to identify the new or changed data. The new or changed data is then extracted from the source system, transformed into a format suitable for storage and analysis, and loaded into the target system.

The benefits of incremental load include reduced resource consumption, faster processing times, and the ability to keep the target system up to date with the latest data. It also reduces the risk of errors or inconsistencies that can occur when reloading large amounts of data.

Incremental load can be scheduled to run at regular intervals, such as daily or hourly, to ensure that the target system is always up to date with the latest data from the source systems.