ByteWise Fellowship – Week 1

Task#5 – Data Loading Date: 16-March-2023

Question#1

**What is historical load?**

Historical load refers to the process of loading historical or past data into a data storage system such as a data warehouse or data lake

This is typically done in situations where there is a need to analyze data that goes back several years or even decades.

**Process of Loading:**

It typically involves

* Data **extraction** from source systems,
* Data **transformation** to a common schema
* Data **loading** into a target system such as a data warehouse.

**Application:**

Historical data can be useful in a variety of ways, such as for trend analysis, identifying patterns and anomalies, and performing historical comparisons. For example, a business might want to analyze sales data from the past 10 years to identify trends and make predictions for future sales.

Question#2

**What is Full Load?**

Full load refers to the process of completely replacing the existing data in a target system with a fresh set of data from source systems

This means that all the data in the target system is deleted and replaced with the new data

**Need of Full Load**

Full load is typically used when there is a need to refresh the entire data set in a target system, such as a data warehouse, to ensure that the most up-to-date information is available

**Working:**  
It is achieved through the process of ETL Pipelines

**Application**

A real-life example of full load in data engineering is when a retail company decides to replace its existing point of sale (POS) system with a new one. Full Load process would be hired to replace all the data of the previous system with a new one

Question#3

**What is incremental load?**

Incremental load refers to the process of selectively loading new or modified data from source systems into a target system, such as a data warehouse, while retaining the existing data that has not changed.

This approach is typically used when there is a large amount of data that needs to be processed, and it is not necessary to refresh the entire dataset.

**Working:**The incremental load process involves identifying the new or modified data in the source system since the last load, extracting it, transforming it to conform to the target schema, and then loading it into the target system. This process can be automated using ETL (extract, transform, load) tools.

**Need:**  
Incremental load is commonly used in situations where source systems generate new data at regular intervals or when there are small updates to existing data. By only processing the new or modified data, the incremental load process can save time and resources compared to the full load process, which involves processing the entire dataset every time.

**Application:**

A real-life example of incremental load in data engineering is when a healthcare organization needs to update its patient records. The healthcare organization receives new patient records and updates to existing patient records on a daily basis from various sources, such as hospitals, clinics, and medical laboratories.