**Abhiram Basa**

**Day 1 Assessment of Data Engineering - 12/04/23**

**Data Engineering :**

* The process of designing, building and scaling systems that organize data for Analytics.

**4 V’s of Data Engineering:**

* “Volume” says how much data you have.
* “Velocity” explains how fast data is getting to you.
* How different is your data is “Variety”.
* How much you can trust your data is “ Veracity”.

**Data science Blue print**

* The data is gathered from News API, flat files, archives, etc.
* To store the data we have Big data and SQl DB / ware house.
* To process the data, we either use Batch Processing, Stream processing.
* To visualize the data. we can use Power Bi, tableau, etc.

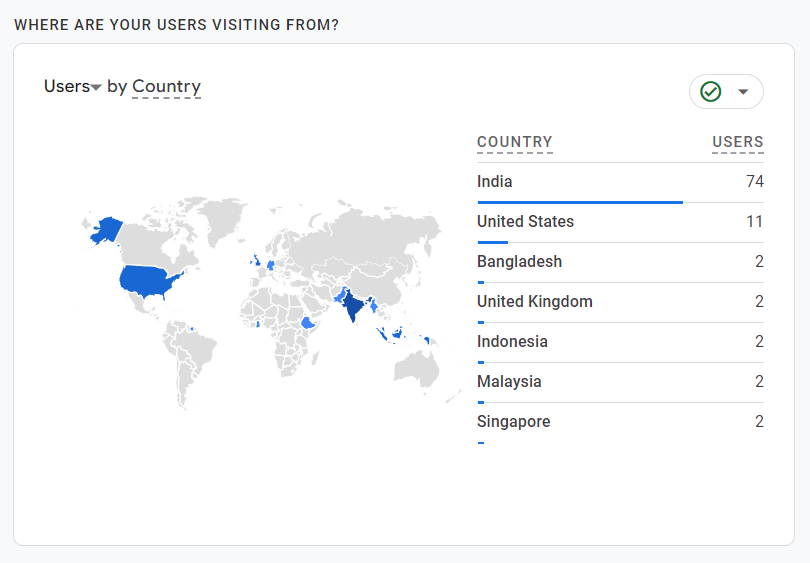
**Batch Processing Example :**

* **Google Analytics** is one the tool which you can check the insights of users visiting your Website.

Insights like

* How many users visited your website?
* Types of devices used to visit your website?
* Country they from?

(see below)



**BIg data tools :**

* Apache spark
* Hadoop
* Azure data bricks.

**Data Warehousing**

Data warehousing is an efficient technique of Data Engineering. It is a type of data management system to support Data Analytics. Lets know what is data ware house,

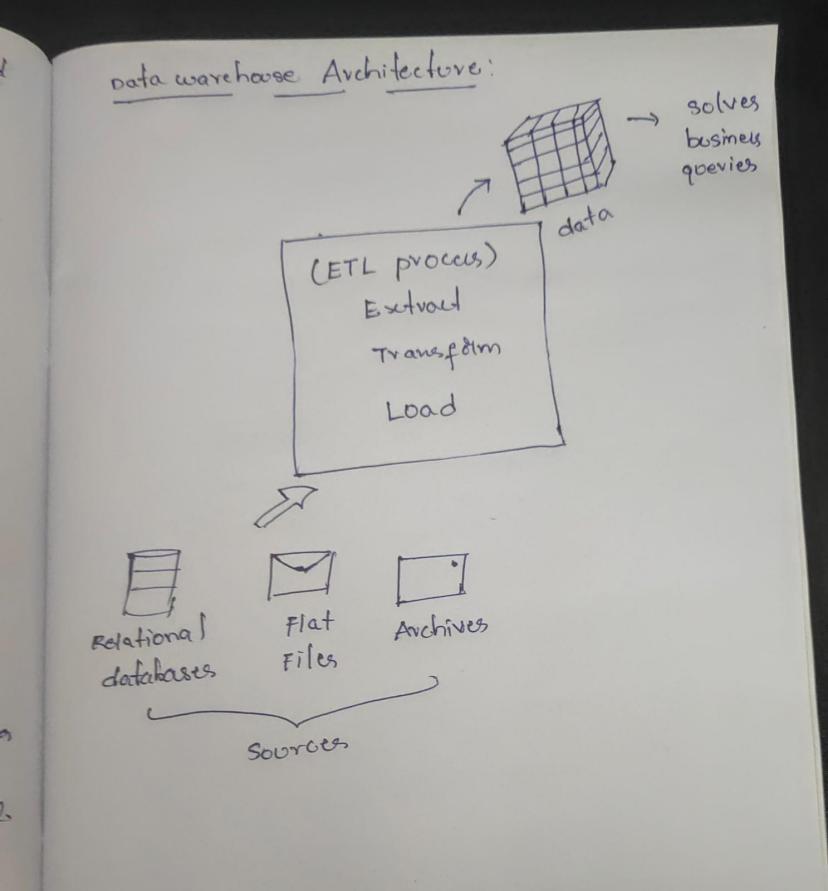
**Data Warehouse :** It is a data storage where huge amount of data is kept, gathered from wide range of resources.

The data can be structured or unstructured format.

**Purpose of Data Warehouse :**

* Used to store large amount of data.
* Used for data analytics.

**Data Warehouse Architecture**

****

First the data is collected from different sources like Relational data bases, flat files,etc.

Then the data is sent to ETL process I.e extract,Transform, load to convert the data into useful information.

* **Extract :** The data is extracted from different sources.
* **Transform**: The data is transformed by removing unnecessary data.
* **Load:** This data is loaded into data marts.

**Operational Database**

Operational database is used to do regular operations of an organization.

**Benefits of OLTP :**

* Simplicity
* Data integrity
* Efficiency
* Fast query processing.

**Disadvantages of OLTP :**

One of the biggest pitfall of OLTP is that it requires **“Instant Updates”**.

Data store contains 2 types of data

1. **Business Data**

This business data contains all the data regarding business insights like product sign ups, newsletter sign ups, etc.

Through this you can increase your product sales.

It is classified into 2

1. **Operational Data**: The data comes from day to day daily activites and is gathered by the Operating systems.

**2.External Data**

All the data gathered from external sources.

2. Business data model

**DSS architectural styles**

You can use the below architectural styles based on the application requirement.

1. OLAP
2. OLTP