Tools & technologies overview

GCP cloud(dataflow, dataproc, Airflow, bigquery, GCS, VM, logging etc)

Databrics

DBT

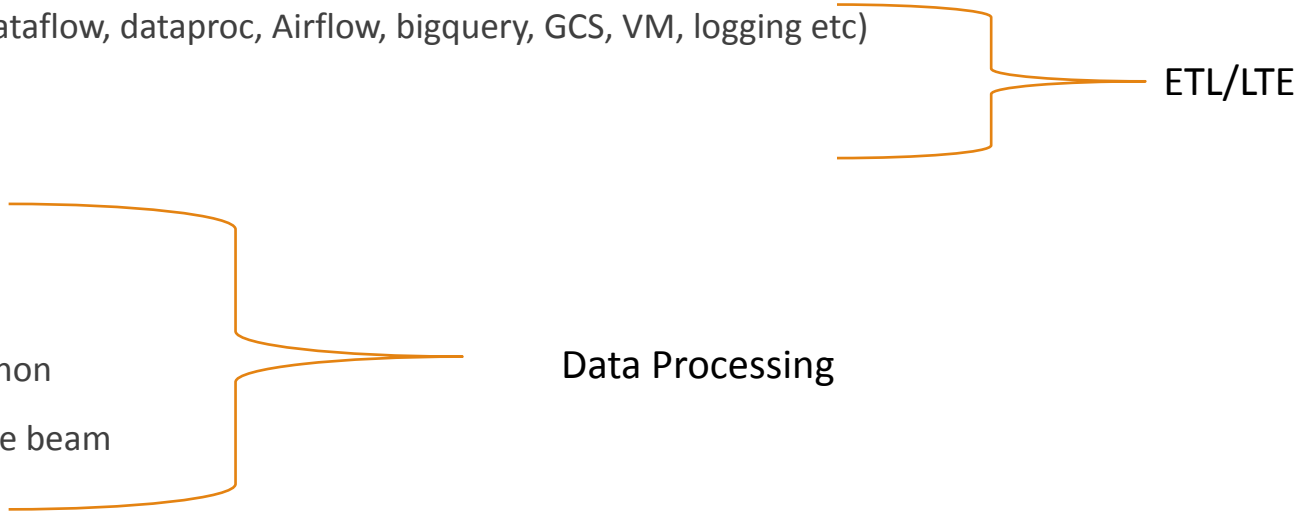
SQL

Python

Advance Python

Spark/Apache beam

Hive



Docker : **Container**: A running instance of an image.

GitHub/Actions : CICD

Terraform : Infrastructure as Code (IaC)

Jira : Fo Ticket monitoring

What is Data?

•**Data** is a collection of information's.

Data is a collection of raw, unorganized facts, figures, or symbols that can be processed, analyzed, and interpreted to derive meaningful information.

Types of data based on Structure

Structured Data

Semi-Structured Data

Unstructured Data

Structured Data

Organized in a fixed schema, typically in rows and columns.

Examples:

- SQL databases (MySQL, PostgreSQL)
- Excel sheets, csv

Semi-Structured Data

Not in traditional table format but has some organization (tags, keys).

Examples:

- JSON, XML, Avro, Parquet
- Logs, API responses, NoSQL data (MongoDB, Cassandra)

Unstructured Data

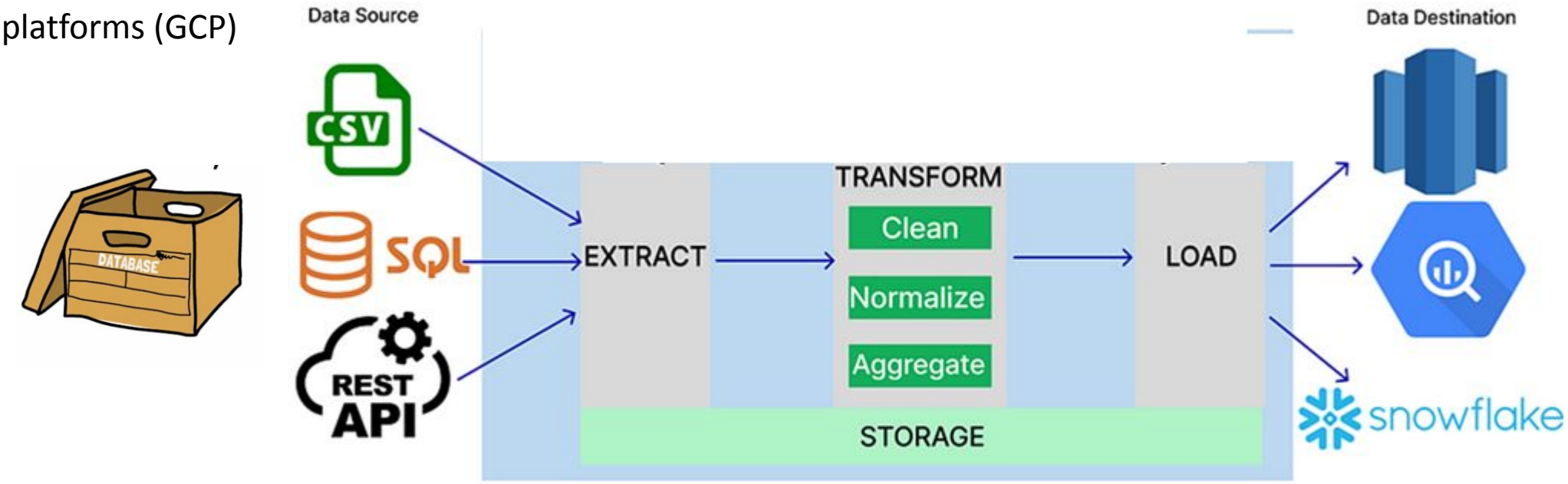
No predefined format or organization.

Examples:

- Images, videos, PDFs
- Emails, social media posts, audio

What is Data Engineering?

- Data Engineering = Building pipelines + managing data architecture
- Focuses on:
  - Data ingestion
  - ETL (Extract, Transform, Load)/ LTE
  - Data Warehousing
  - Big Data tools
  - Cloud platforms (GCP)



What is Database?

A database is an organized collection of data, typically stored electronically, that can be easily accessed, managed, and updated

What is SQL?

SQL IS A STANDARD LANGUAGE FOR ACCESSING AND MANIPULATING DATABASES.

SQL STANDS FOR STRUCTURED QUERY LANGUAGE

Name	Fathers Name	Age	DOB	Contact

Oracle DB setup

Installation

SQL is divided into the following

Data Definition Language (DDL)

Data Manipulation Language (DML)

Data Retrieval Language (DRL)

Transaction Control Language (TCL)

Data Control Language (DCL)

SQL is divided into the following

DDL -- create, alter, drop, truncate, rename

DML -- insert, update, delete

DRL -- select

TCL -- commit, rollback, savepoint

DCL -- grant, revoke

DDL (Data Definition Language)

- **CREATE** – Used to create a new database object (like a table, view, or database).
- **ALTER** – Modifies an existing database object, such as adding or deleting columns in a table.
- **DROP** – Deletes an existing database object permanently.
- **TRUNCATE** – Removes all rows from a table, but keeps the table structure intact.
- **RENAME** – Changes the name of an existing database object.

DML (Data Manipulation Language)

- **INSERT** – Adds new records (rows) into a table.
- **UPDATE** – Modifies existing records in a table.
- **DELETE** – Removes existing records from a table.

DRL (Data Retrieval Language)

SELECT – Retrieves data from one or more tables in a database.

TCL (Transaction Control Language)

- **COMMIT** – Saves all changes made during the current transaction permanently to the database.
- **ROLLBACK** – Undoes changes made in the current transaction (since the last COMMIT or to a SAVEPOINT).
- **SAVEPOINT** – Sets a point within a transaction to which you can later roll back,

  without affecting the entire transaction.

DCL (Data Control Language)

- **GRANT** – Gives specific privileges (like SELECT, INSERT) to users on database objects.
- **REVOKE** – Removes previously granted privileges from users.

CREATE TABLE SYNTAX

Create table <table_name> (col1 datatype1, col2 datatype2 …coln datatypen);

Ex:

SQL> create table student (no number (2), name varchar (10), marks number (3));