**DAY 1 (29-5-25)**

Intro:

What is the Industrial Revolution?

1. The **Industrial Revolutions** are distinct phases in history that radically changed how humans **produce goods**, **work**, and **live** — each driven by **technological innovation**.
2. In Short it is Known as Revolution which happen based upon the latest tech in that period.

INDUSTRY 1.0 (Late 1700s to Mid-1800s):

* **Steam Engine**
* **Water Wheels**
* **Mechanical Looms**
* **Textile Machines**
* **Iron Production (Blast Furnace)**
* **Canals and Railways**
* **Hand Tools to Machines**

**INDUSTRY 2.0** (Late 1800s to Early 1900s):

* Electricity
* Assembly Line (e.g., Ford Model T)
* Telegraph & Telephone
* Steel Manufacturing
* Internal Combustion Engine
* Light Bulb
* Railroads and Shipping Expansion
* Oil and Chemical Industries

INDUSTRY 3.0 (1970s–2000s):

* Computers & Microprocessors
* Programmable Logic Controllers (PLCs)
* Robotics
* Information Technology (IT)
* Enterprise Resource Planning (ERP) systems
* Internet (early stages)
* Digital Communication (e.g., email)
* Basic Data Storage & Retrieval Systems

INDUSTRY 4.0 (2010s–Now):

* GEN AI
* Metaverse
* Cyber security and Cyber Physical Systems
* Block Chain
* IOT
* BIG DATA
* Quantum Computing
* Cloud Computing
* Prompt Engineering

What is Data Engineering?

Data Engineering focuses on the design, building, and maintenance of systems that collect, store, and process data at scale.

Processes in Data Engineering:

Data Engineering involves the following core processes:

1. **Data Collection** – Getting data from different sources.
2. **Data Ingestion** – Moving raw data into a system (cloud, lake, warehouse).
3. **Data Cleaning & Transformation (ETL/ELT)** – Ensuring data quality, consistency.
4. **Data Storage** – Storing data efficiently using databases or data lakes/warehouses.
5. **Data Pipeline Orchestration** – Automating workflows using tools like Apache Airflow.
6. **Serving Data** – Making it accessible for analysts and data scientists (Power BI).

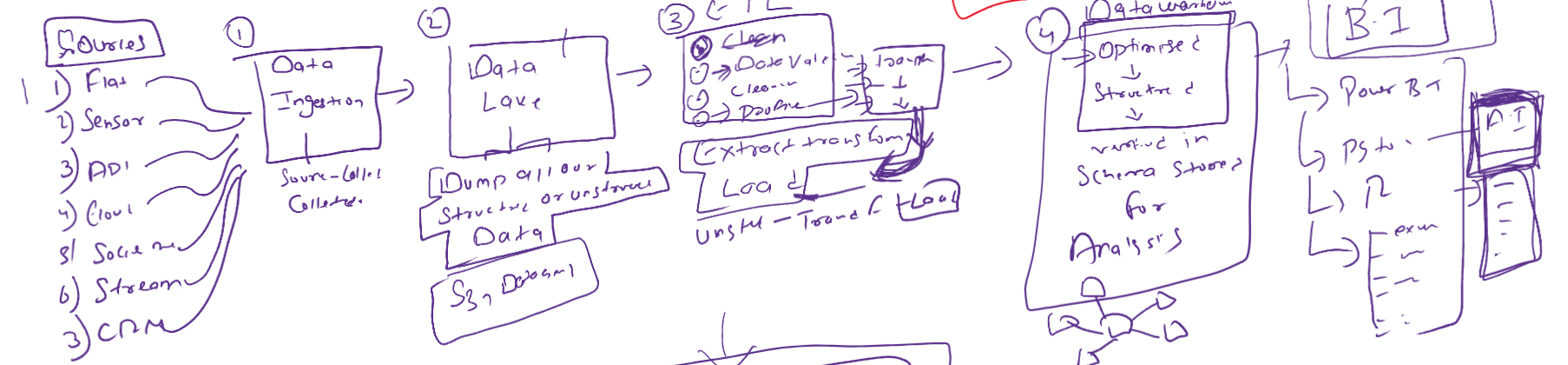
What is Data Science?

Data Science focuses on analysing and interpreting complex data to extract insights, build models, and drive decision-making.

A whiteboard with red text

AI-generated content may be incorrect.

DATA PIPELINE WORKFLOW:



Step 1:

Collect Data from multiple sources like,

* Flat Files (CSV, Excel, etc.)
* Sensors (IoT devices)
* APIs
* Cloud-based data
* Social Media
* Streaming data (real-time sources)
* CRM systems (like Salesforce)

Step 2:

Store the collected Raw Data,

* All Structured and Un-structured Data are stored.
* Stored in Amazon or Azure clouds.

Step 3:

Convert the data into usable Format and need to structure according to it and this process is known as ETL.

* Extract - Pull data from the storage space.
* Transform – clean the data, validate data and prepare for Loading.
* Load - Send the transformed data to the next stage

Step 4:

Storing the cleaned data,

* Transformed data is loaded here.
* Uses schema-based storage for High-speed queries retrieval and Analytical workloads.

Step 5:

Provide insights, dashboards, and reports from the stored data.

Tools which can be used for this process are,

* Power BI
* Tableau
* Excel

A whiteboard with red writing

AI-generated content may be incorrect.

**Data Ingestion**: Importing raw data from multiple sources into a system.

* + *Example*: Bringing data from sensors or websites into a cloud platform.

**Data Lake**: A large storage repository for raw, unstructured, semi-structured, and structured data.

* + *Tools*: AWS S3, Azure Data Lake

**Machine Learning:**

**Machine Learning (ML)** is a subfield of **Artificial Intelligence (AI)** that enables computers to **learn from data** and **make decisions or predictions** without being explicitly programmed for every task.