







**Apache Airflow:**

Apache Airflow is an **open-source workflow orchestration tool** used to **schedule, monitor, and manage data pipelines**.

* **Workflow orchestration** → means arranging tasks (like extracting data, transforming it, loading into a database, training a model, etc.) in the correct order and making sure they run automatically.

**Key Points about Airflow:**

* **Created by Airbnb**, now part of the Apache Software Foundation.
* Uses **DAGs (Directed Acyclic Graphs)** → a set of tasks with dependencies showing the order of execution.
* **Python-based** → Workflows are written in Python scripts.
* Provides a **web UI** to track jobs, see logs, retry failed tasks, and monitor progress.
* Integrates with many systems (databases, cloud storage, Hadoop, Spark, etc.).
* Commonly used in **ETL pipelines, data engineering, machine learning workflows, and DevOps automation**.

Example: If you want to

1. Extract sales data from an API,
2. Transform it using Python,
3. Load it into a database,
4. Then send a report email

Airflow can schedule and automate all these steps daily, and if one step fails, it can retry or alert you.