⬜ Airflow – Theoretical Assessment

Section A – Basics

1. What is Apache Airflow and why is it used?
2. Define a **DAG**. What does each part of the acronym stand for?
3. Explain the difference between a DAG and a Task.
4. Why should workflows be "Directed Acyclic Graphs" in Airflow?

Section B – Core Concepts

1. Describe the role of the following Airflow components:

   - Webserver
   - Scheduler
   - Metadata Database

2. What is the purpose of the `airflow db init` command?

3. What is the significance of `start_date` and `schedule_interval` in a DAG?

4. What does `catchup=False` do, and when would you use it?

Section C – Operators & Execution

1. What is an Operator? Give two examples.
2. How does Airflow handle task failures and retries?
3. What is XCom and how is it useful?
4. Explain the difference between `BashOperator` and `PythonOperator`.

Section D – Real-World Use

1. Give one real-world example where Airflow can be used for ETL.
2. Why is it recommended to keep DAG scripts lightweight and avoid heavy computations inside them?
3. Why should every DAG have a unique `dag_id`?
4. How does Airflow ensure workflows run in the correct order?