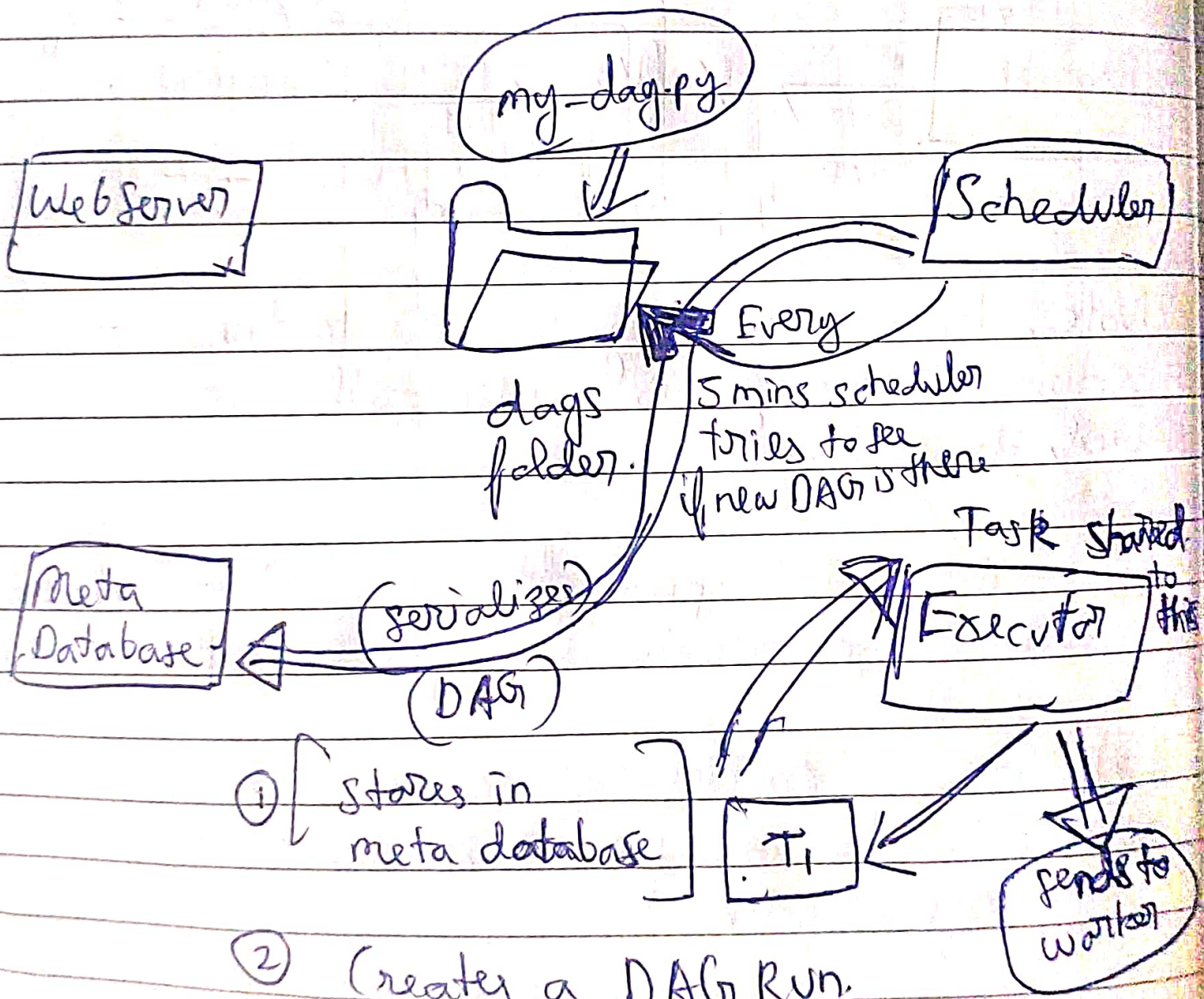
(23rd Dec 2025)

(Airflow)

(How does Airflow work?)



② Creates a DAG Run.

③ If it has multiple ~~DAGs~~ Tasks to run, it creates Task instances.

Limitations of Airflow

Airflow is an orchestrator, not a data processor, it should not do heavy computation.

Bad Practice :-

- Large spark jobs inside Python Operators

Correct Usage :-

- Airflow triggers work, Actual work run in.
  1) Spark
  2) EMR
  3) Kubernetes jobs.

Also always use airflow to delegate tasks to systems. We can use airflow to spin up a spark streaming job in EMR cluster.

(Airflow = control plane (orchestration)
Spark Streaming = data plane)