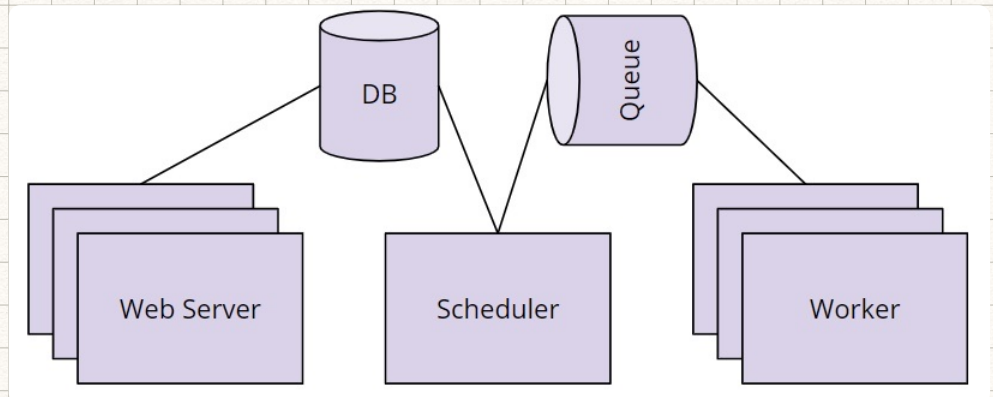# Data Pipelines:

Data Validation → ///// What happens if data goes wrong? ⟹ We need data validation.

What Makes up Airflow?

Apache Airflow consists of:

///// Scheduler: for orquestrating execution of jobs on a trigger or schedule

///// Working Queue: which holds the state of running DAGs and Tasks

///// Worker Processes: that execute operations defined in each DAG.

///// Database: which saves credentials, connections, history and configuration

///// Webserver: which provides a control dashboard for users and maintainers

How Airflow Works?

- The Airflow Scheduler starts DAGs based on time or external triggers.

- Once DAG is started, the Scheduler looks at the steps within the DAG and determines which step can run by looking at their dependencies

- The Scheduler places runnable steps in the queue.

- Workers pick up those tasks and run them.

- Once the worker has finished running the step, the final status of the task is recorded and additional tasks are placed by the scheduler until all tasks are complete.

Once all tasks have been completed, the DAG is complete.