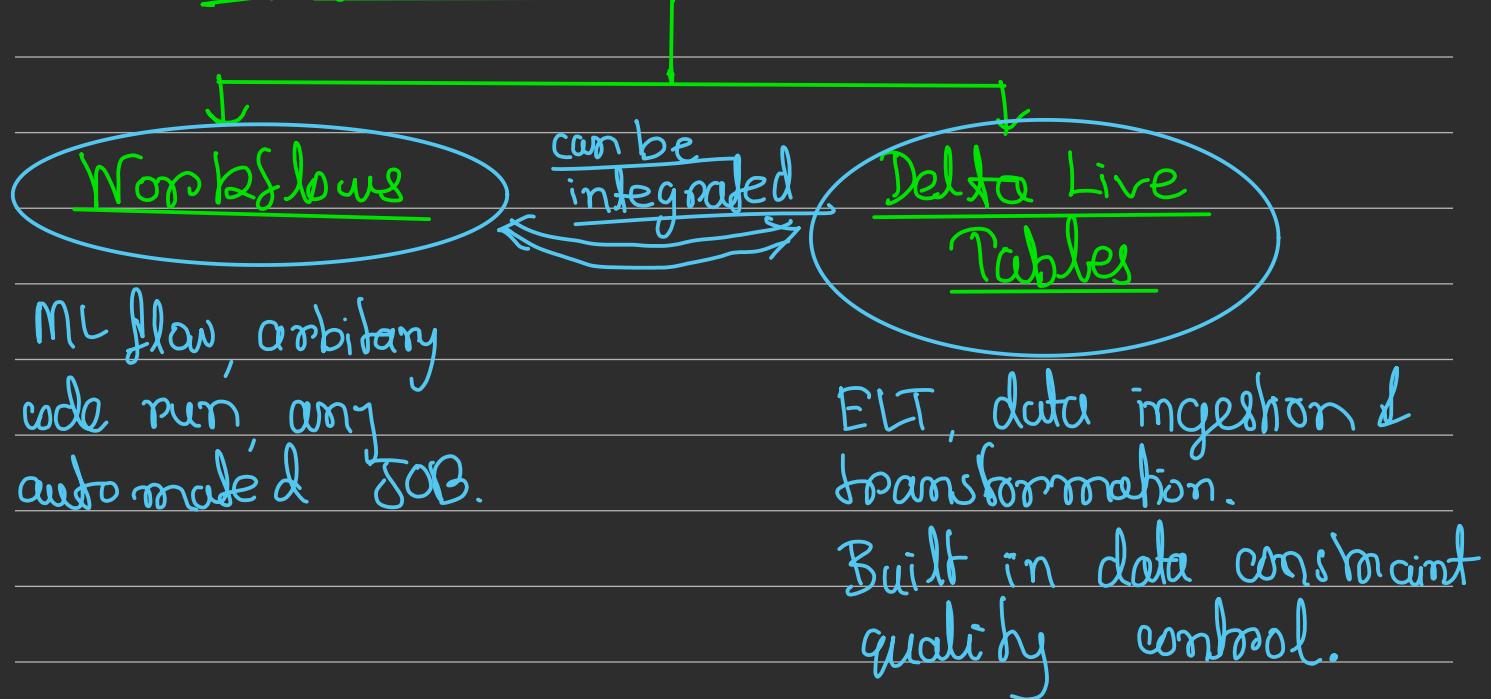


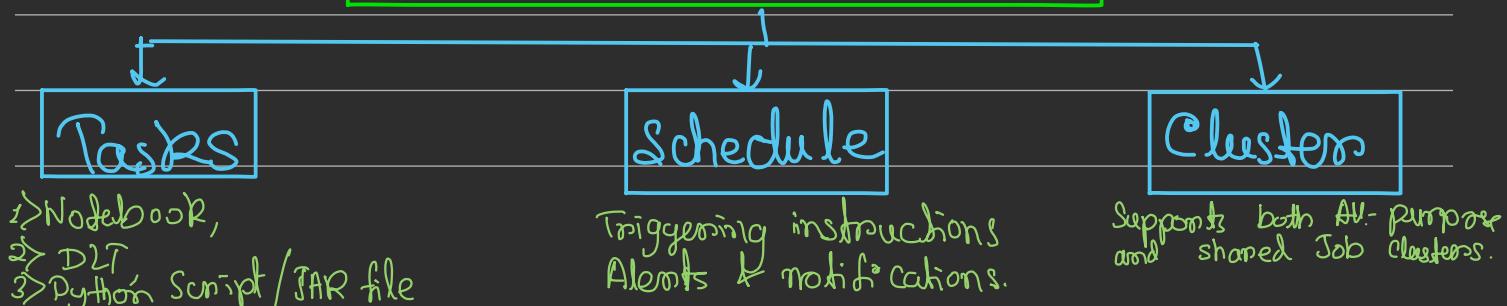
WORKFLOWS :-

- ⇒ Fully-managed (serverless) cloud-based general purpose task orchestration service for the entire lakehouse.
- ⇒ Allows data teams to combine tasks, such as notebooks, SQL, ML models, Python code.

Databricks Workflows 2 Main Task Orchestration Services



Workflow Components



⇒ Shared Job Cluster — recommended for Production env. as it provides flexibility to use same cluster more than once, thus cutting costs and reducing standup time.

⇒ Jobs can be created and run by:-

- (i) CLI ,
- (ii) UI ,
- (iii) invoking Jobs API.

⇒ Job Runs tab contains statistics and metrics of executed jobs.

⇒ Workflows Repair feature helps to rerun failed jobs, and debug failures quickly, by running subset of tasks, or any dependent tasks.

NOTE :-

⇒ In the job workflows, if a DLT pipeline task is set to run continuously based on schedule, it will execute indefinitely, and will never reach a state of Execution Completed. Thus, the tasks dependent on it would never run.

Thus, when DLT is set as a task in Workflow Jobs, no task should be made dependent on the DLT task.

DLT versus Workflow Jobs

Considerations

	Delta Live Tables	Workflow Jobs
Source	Notebooks only	JARs, notebooks, DLT, application written in Scala, Java, Python
Dependencies	Automatically determined	Manually set
Cluster	Self-provisioned	Self-provisioned or existing
Timeouts and Retries	Not supported	Supported
Import Libraries	Not supported	Supported