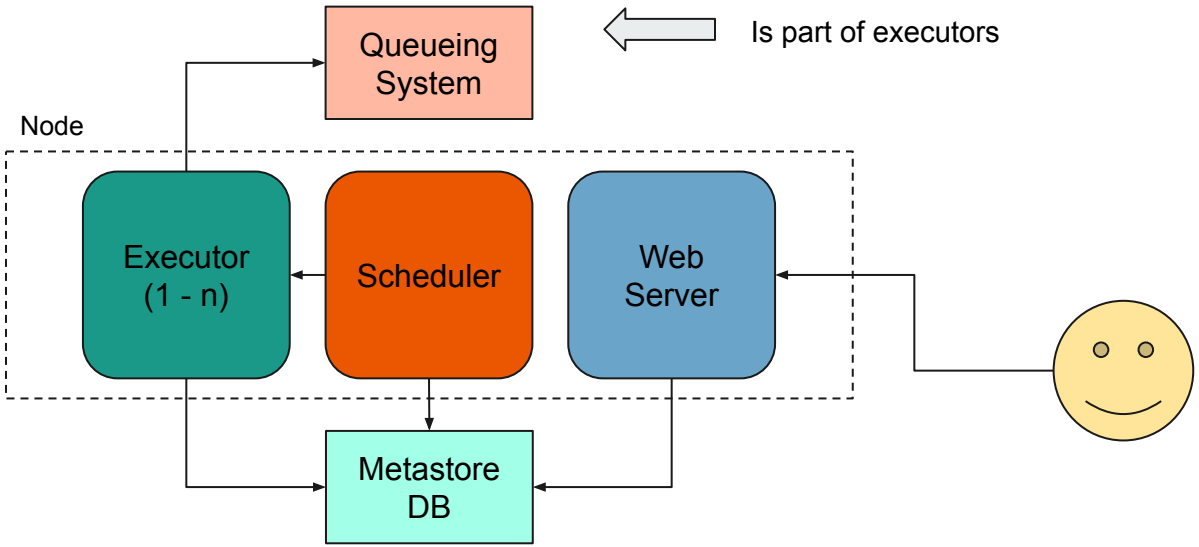
How Airflow works?

How the Airflow's components work together

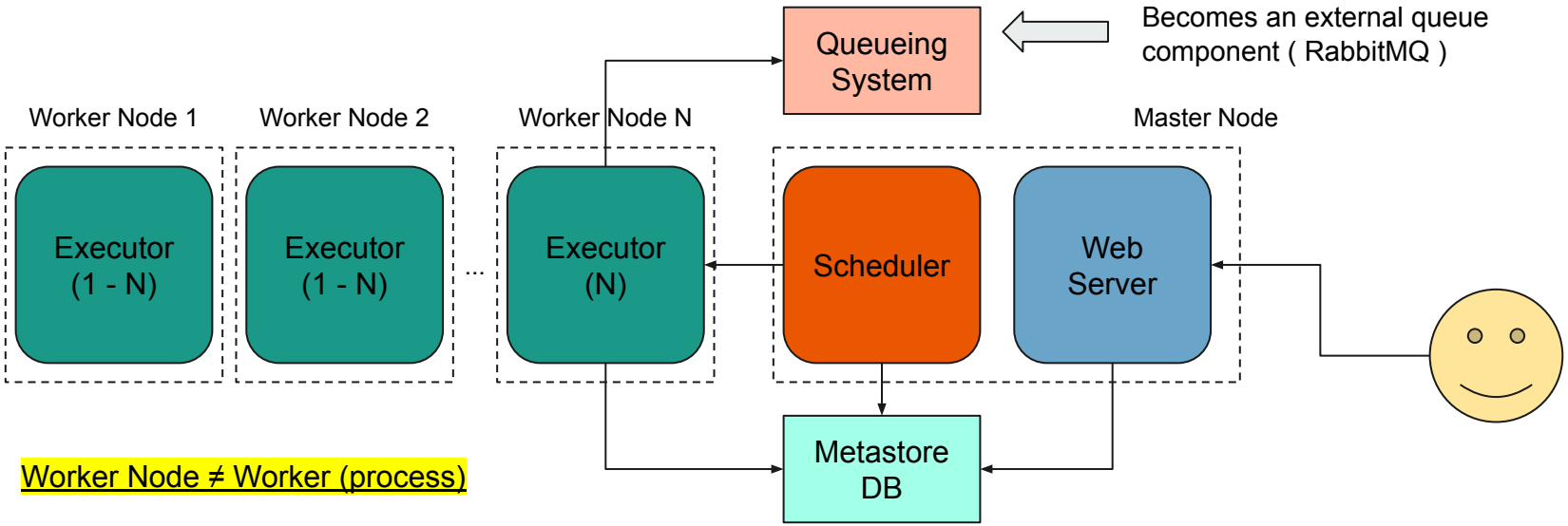
A Quick Remember

- Web Server
  - Airflow's UI to see the status of your jobs and a lot more information (we will see later).
- Scheduler
  - Responsible for scheduling your jobs.
- Executor
  - Message queuing process tightly bound to the Scheduler and determines the worker processes that actually execute each task.
- Workers
  - Processes that execute the tasks, determined by the executor.
- Metadatabase
  - A database where all the metadata related to your jobs are stored.

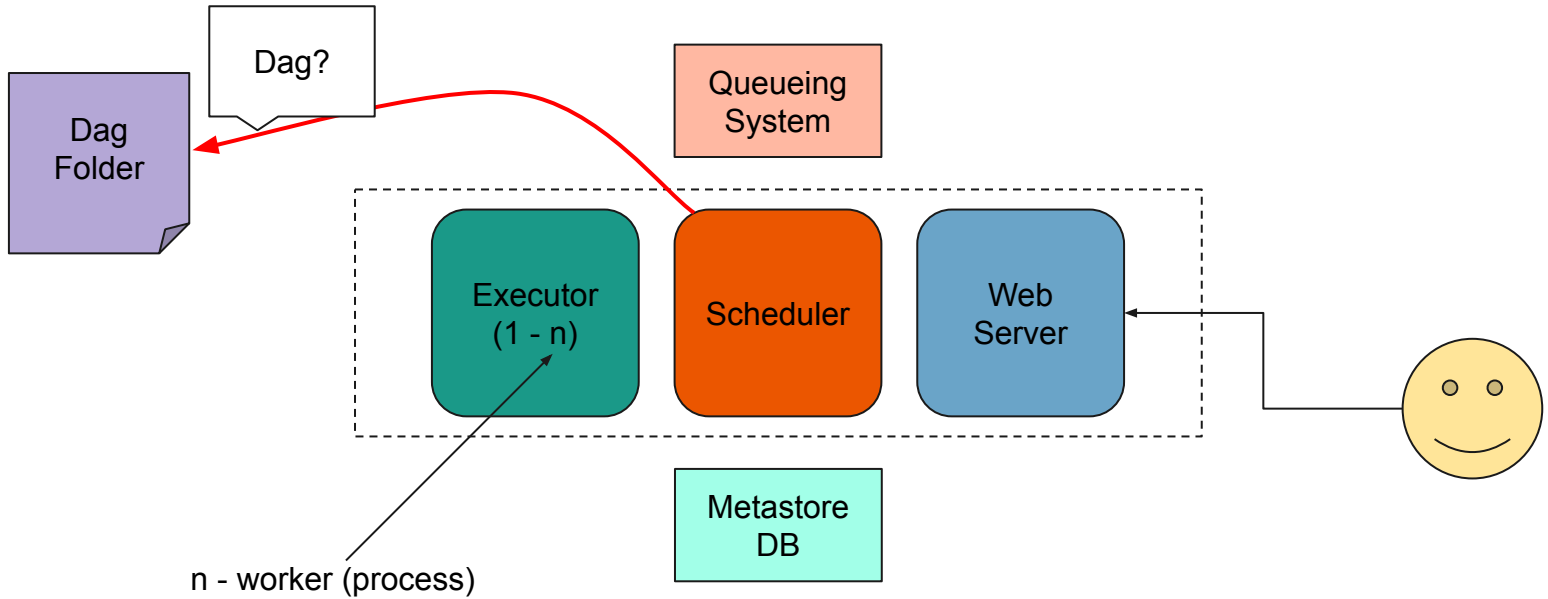
Architecture Overview (Single Node)



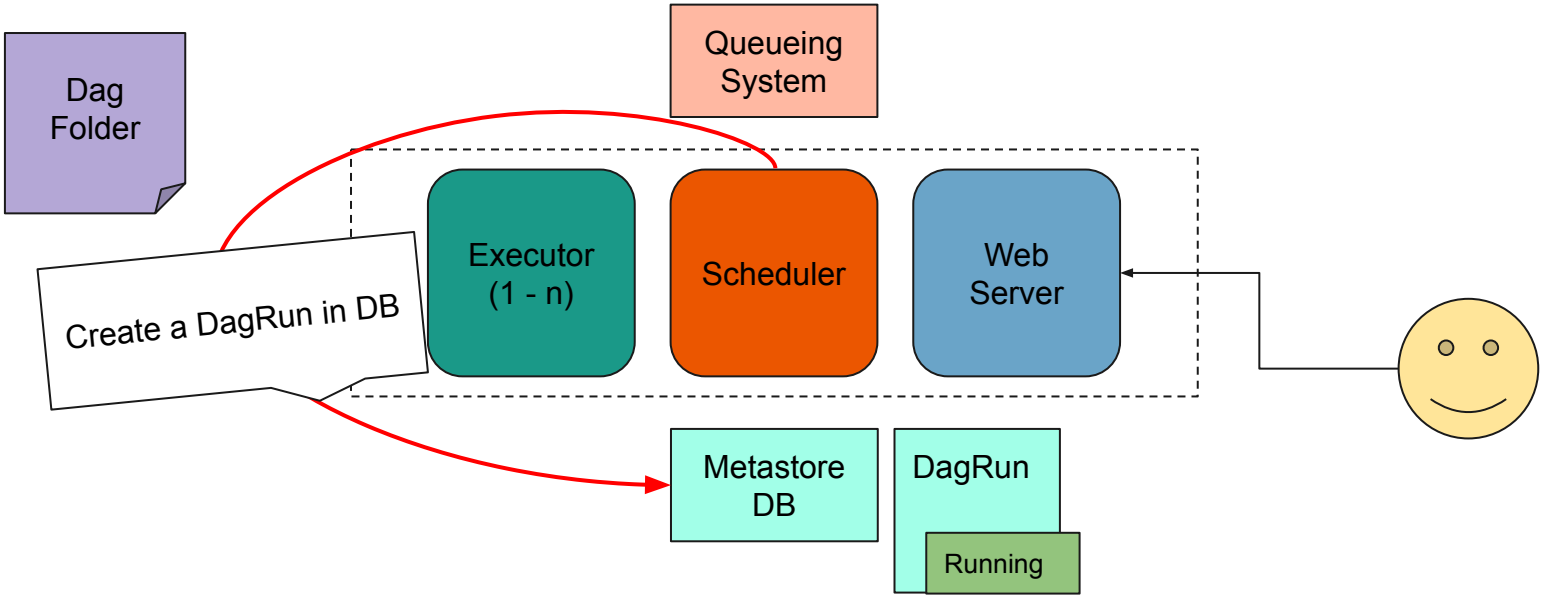
Architecture Overview (Multi Nodes)



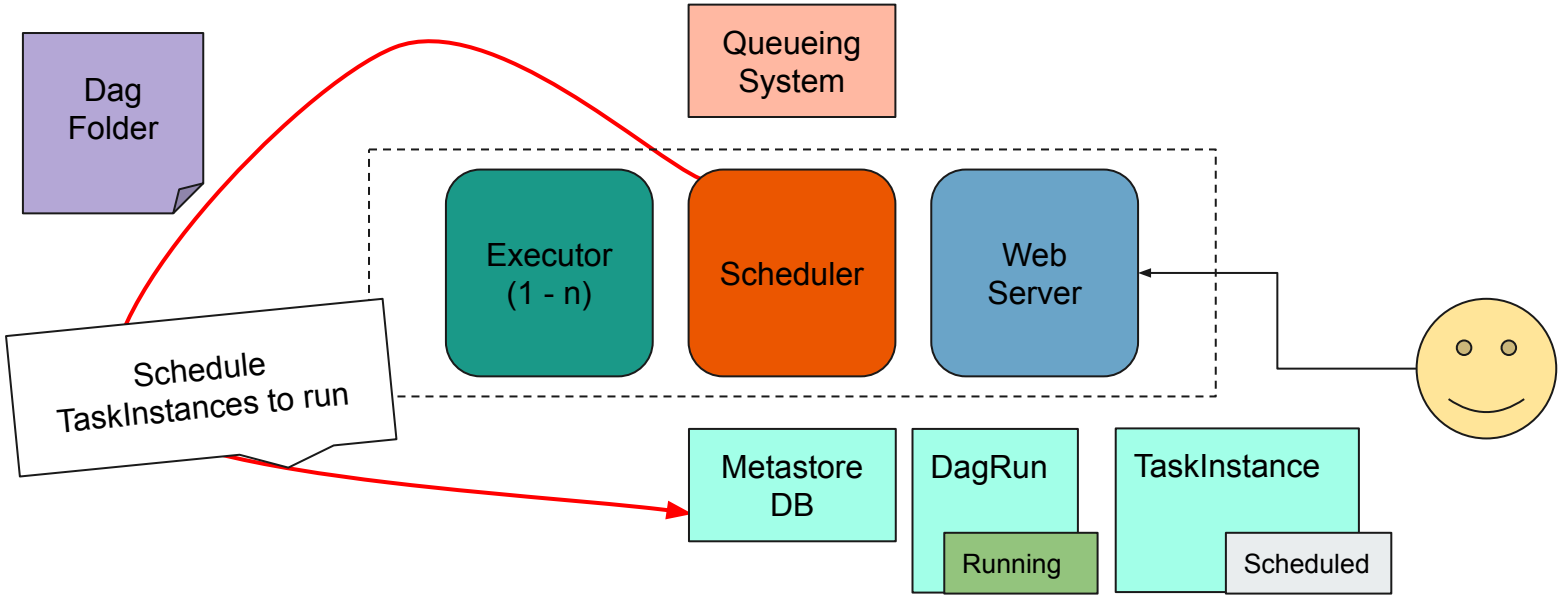
How Your Work Gets Done?



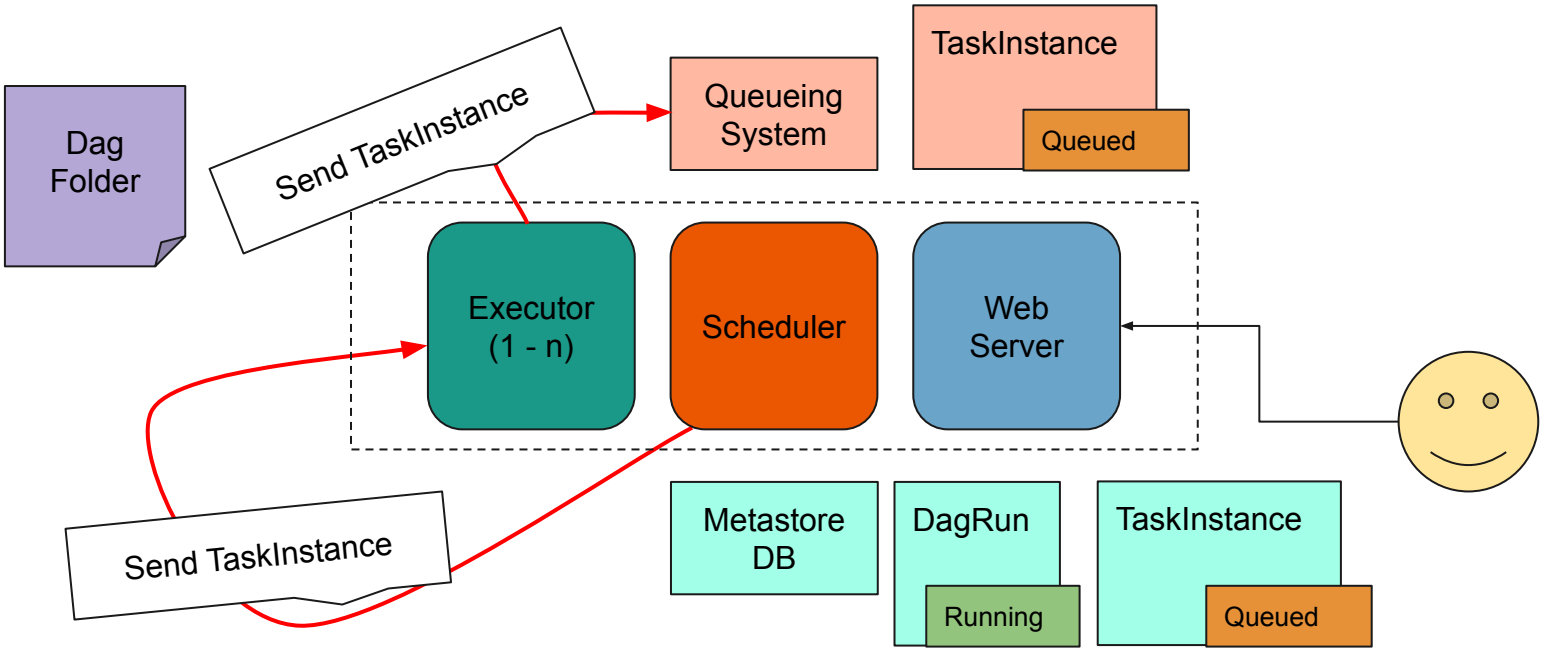
How Your Work Gets Done?



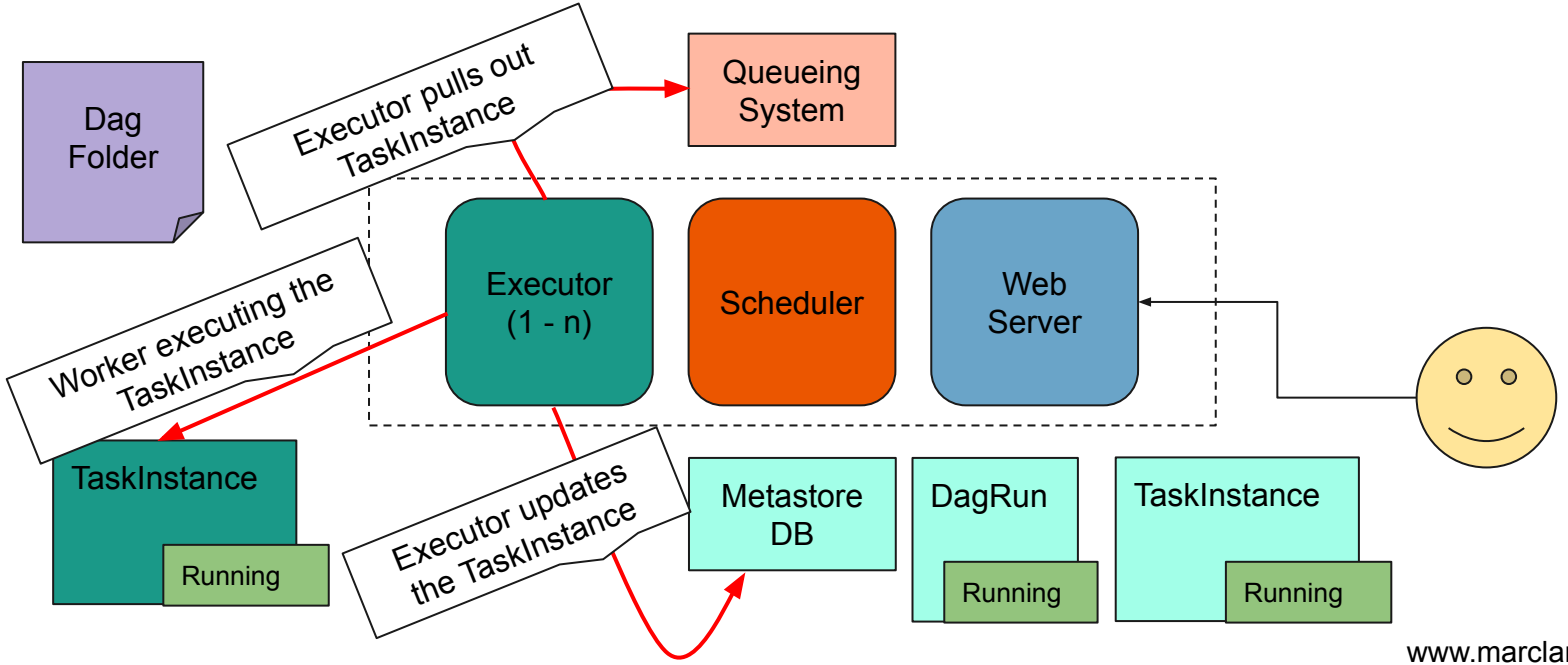
How Your Work Gets Done?



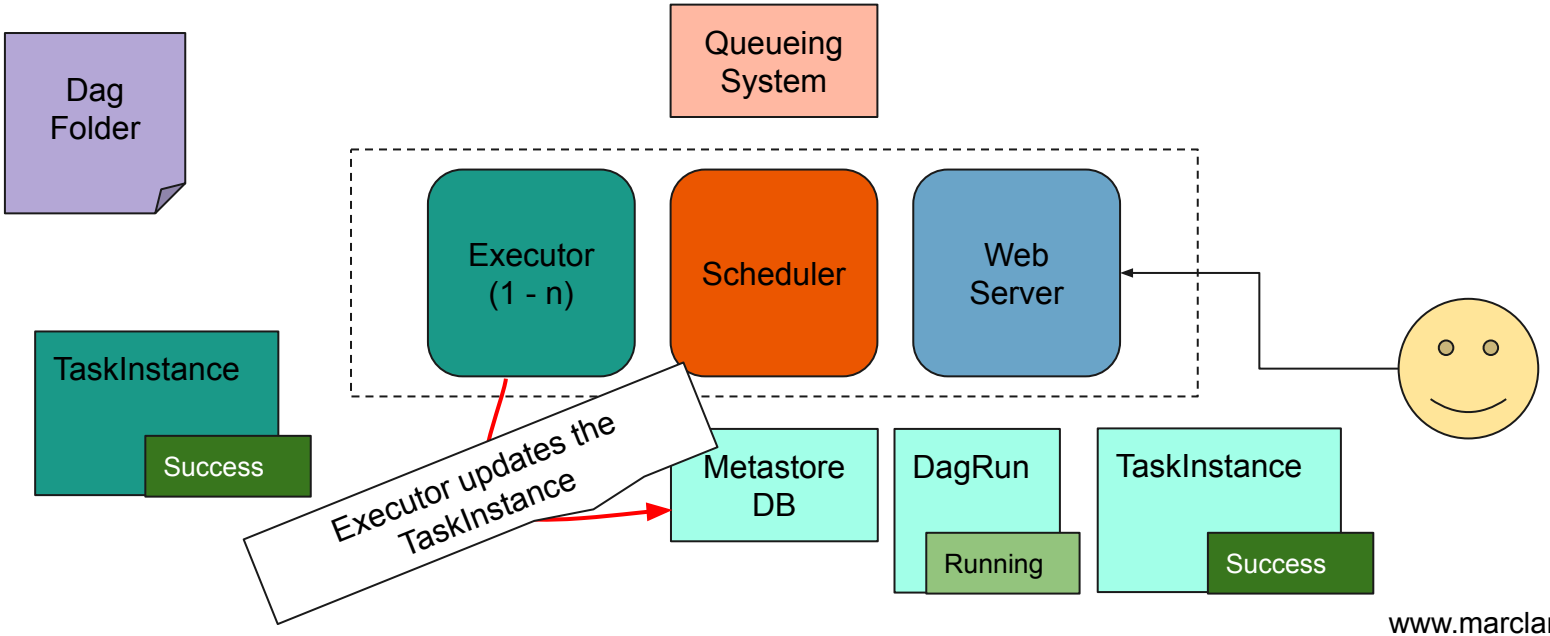
How Your Work Gets Done?



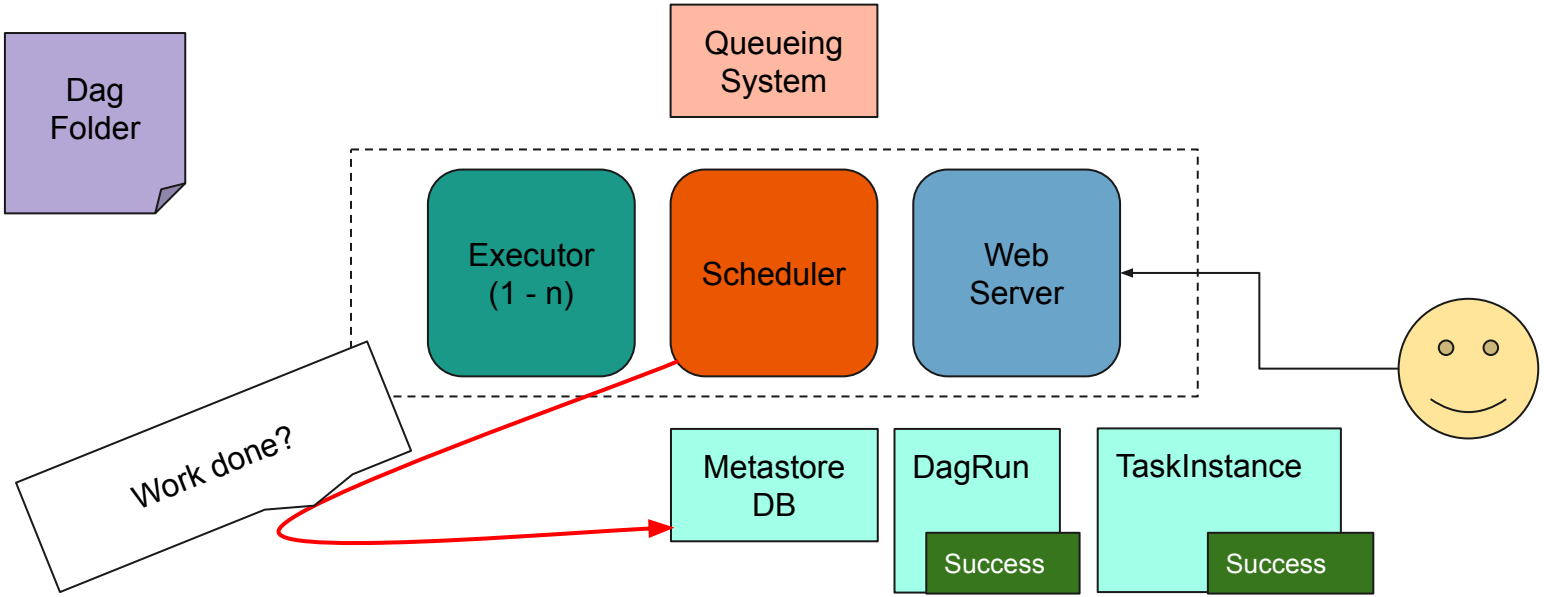
How Your Work Gets Done?



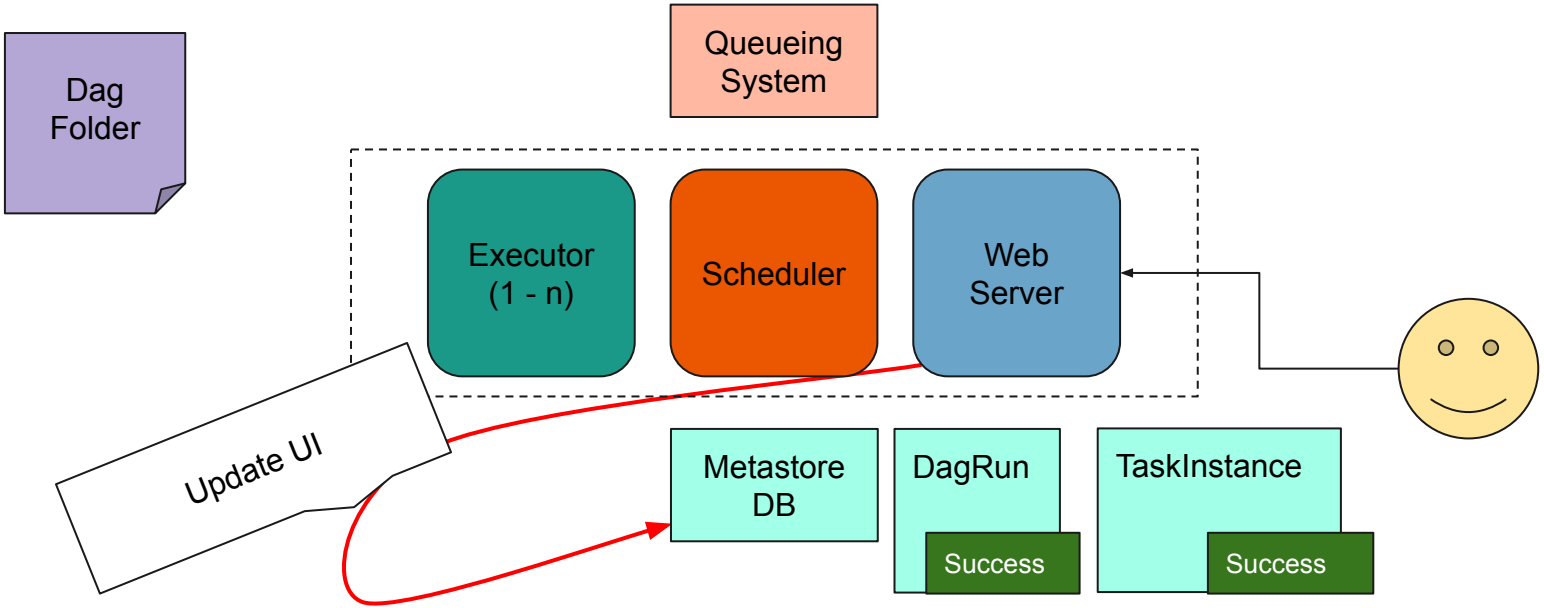
How Your Work Gets Done?



How Your Work Gets Done?



How Your Work Gets Done?



Sum up

1. The Scheduler reads the DAG folder.
2. Your DAG is parsed by a process to create a DagRun based on the scheduling parameters of your DAG.
3. A TaskInstance is instantiated for each Task that needs to be executed and flagged to “Scheduled” in the metadata database.
4. The Scheduler gets all TaskInstances flagged “Scheduled” from the metadata database, changes the state to “Queued” and sends them to the executors to be executed.
5. Executors pull out Tasks from the queue (depending on your execution setup), change the state from “Queued” to “Running” and Workers start executing the TaskInstances.
6. When a Task is finished, the Executor changes the state of that task to its final state (success, failed, etc) in the database and the DAGRun is updated by the Scheduler with the state “Success” or “Failed”. Of course, the web server periodically fetch data from the metadatabase to update the UI.