



Recap

Let's review what we've learned so far



Summary

- Apache Airflow is a platform to programmatically author, schedule and monitor workflows.
- 5 core components: the Scheduler, the Metadatabase, the Webserver, the Worker and the Executor.
- 4 key concepts: DAG (graph of operators), Operator (define your tasks), Task (Instance of an operator) and Task Instance (Specific run of task DAG + TASK + POINT IN TIME), Workflow.
- A Queue is internally used in standalone mode. In distributed mode an external queue like RabbitMQ will be required as we'll see later.
- We did a quick tour of Airflow UI and CLI.
- The Scheduler is the heart of Apache Airflow, it schedules DagRuns and TaskInstances.
- Installing Airflow is really easy, don't forget to initialize your DB before using it.



What's next?

Now you know the basics, let's create our FIRST WORKFLOW :)