# Why Airflow?

Airflow is meant to process multiple data pipelines. When you have a workflow that you need to monitor, and that workflow can have 100s of instances. This is where airflow can help you track the progress and monitor its failures for each instance.

In a nutshell, Airflow is useful when you have multiple data pipelines to schedule, monitor and track.

# What is Airflow?

Airflow is an Orchestration Framework

Airflow is the best Orchestration Framework, tool for scheduling and executing your data pipelines.

Airflow can help you

Features:

Dynamic – Airflow is coded in Python, so anything you can do in Python you can do in Airflow

Scalable – You can execute any number of tasks as you want provided you use the right executer

User interface – Beautiful Interface

Extensible – you don’t have to wait for the community to create some plugin that you need in your project. You can create your own plugin and use it.

Core Components

Webserver – Webserver is the User Interface where you can see all your DAGs and monitor them. Flask with Gunicorn server

Scheduler – Daemon in charge of scheduling the tasks

Meta-store – Database typically MySQL or Postgres to store metadata about dags and runs

Executor – The executor to be used to schedule the tasks, how your tasks will be executed on which system.

Worker – The process or subprocess executing the task

Concepts

DAGs – Directed Acyclic Graphs

* DAG is basically your data pipeline