Basic Airflow architecture

Primarily intended for development use, the basic Airflow architecture with the Local and Sequential executors in an excellent starting point for understanding the architecture of Apache Airflow

There are a few components to note:

- Metadata Database: Airflow uses a SQL database to store metadata about the data pipelines being ru. In the diagram above, this is represented as Postgres which is extremely popular with Airflow. Alternate database supported with Airflow include MySQL.
- Web Server and Scheduler: The Airflow web server and Scheduler are separate processes run (in this case) on the local machine and interact with the database mentioned above.
- The Executor is shown separately above, since it is commonly discussed within Airflow and in the documentation, but in reality it is NOT a separate process, but run within the Scheduler.
- The Worker(s) are separate processes which also interact with the other components of the Airflow architecture and the metadata repository.
- airflow.cfg is the Airflow configuration file which is accessed by the Web Server, Scheduler, and Workers.
- DAGs refers to the DAG files containing Python code, representing the data pipelines to be run by Airflow. The location of these files is specified in the Airflow configuration file, but they need to be accessible by the Web Server, Scheduler, and Workers.

Ref

- https://airflow.apache.org/docs/apache-airflow/stable/start.html

---

Plugin Backlinks: .

From:
http://jace.link/ - **Various Ways**

Permanent link:
http://jace.link/open/basic-airflow-architecture

Last update: **2021/01/28 02:06**