Interview Questions:

1. **What is an orchestration tool and why is its use?**
    a. Every system is becoming complex with each passing day. Therefore, setting the same system pipeline in every environment is not an option. Orchestration tools help integrate multiple processing units and link them together. These tools also help in monitoring these units and trigger alerts in case of failures.

2. **Why do we need Airflow?**
    a. Airflow is an open-source workflow management tool. It allows users to programmatically create workflows (data pipelines), schedule them, monitor them with web UI and also trigger alerts during failures.

3. **What is DAG in Airflow?**
    a. DAG, or Directed Acyclic Graph, is a data structure that helps maintain the task dependency at any given time. Airflow uses DAG for maintaining tasks relations to ensure that tasks are executed in an expected order.

4. **How can task relations be applied in a DAG?**
    a. To apply tasks dependencies in a DAG, all tasks must belong to the same DAG. Now, relations can be given using the up_stream() and down_stream() methods. Airflow also provides bit wise operators such as >> and << to apply the relations. Bit wise operators are easy to use and help to easily understand the task relations.

5. **What are the operators in Airflow?**
    a. Airflow is a workflow management tool. A workflow consists of various tasks that can perform processing at various other tools such as MySQL, S3, Hive and Shell scripts. Airflow cannot provide a single platform to support all processing engines, instead it provides a way to integrate and connect these engines. Operator is a generic way to connect those processing engines and perform the tasks.

6. **How can we create Big data ETL jobs using Airflow?**
    a. Generally, Big data ETL jobs include data migration jobs such as getting data from mysql or any relational database, perform some transformations on it and then moving the data to Hadoop tables such as Hive.

b. Airflow provides sqoop operators, spark operators, and hive operators, so Airflow can be used to invoke any of the Big data tasks and Airflow can also sequence and monitor the jobs. In this way, Airflow is useful in Big data ETL jobs.

7. **Mention the features of Airflow web UI.**
    a. Airflow web UI has many features that are very useful for any workflow management tools.
        i. It allows you to view the schedule of all the jobs and their historical status.
        ii. It allows to view DAG and their task relation dependency.
        iii. At any time, you can also see the actual code used to run the pipelines.
        iv. It supports custom executions using web UI and CRUD operations on DAG.

8. **What are the different executors supported in Airflow?**
    a. Every task needs to be executed. Executors are components that determine how these tasks get executed.
    b. Airflow supports various executors such as Sequential, Celery, Debug, and Local.
    c. Different techniques are used to scale the number of processes using Celery, Dask, and local executors.
    d. For the debugging purpose, a debug executor is used and Sequential is used to run all tasks sequentially with no parallelism.
    e. It also supports kubernetes as an Executor engine.

9. **What are the alternatives to Airflow?**
    a. Some alternatives to Airflow are Oozie, Azkaban, and Luigi.
    b. Oozie is mostly used for Big data pipelines as Oozie is distributed and well integrated with the Hadoop environment.
    c. Azkaban and Luigi have capabilities similar to those of Airflow and can create and schedule workflows.

10. **How does Airflow maintain the DAG and tasks status and historical data?**
    a. For maintaining any results and data, any tool will have some sort of database storage. Airflow also uses DB. It uses SQLite by default, and there is a way to use MySQL and Postgres DB for more performance improvisation.

11. **How can you notify workflow failures in Airflow?**
    a. Airflow supports alert triggers on workflow failures. You can provide email addresses and set the flags for alerting to such emails in case of workflow failures.

b. Notifications can also be sent on other platforms like Slack and Hipchat.

12. How can custom Operators be created in Airflow?

a. Airflow has a BaseOperator as Base Class for all operators created in Airflow. For creating a custom operator, you have to extend the Operator class and implement its abstract methods. The execute method needs to be designed for the newly created Operator class.