

AIRFLOW

1. What is Apache Airflow?
2. Is Apache Airflow an ETL tool?
3. How do we define workflows in Apache Airflow?
4. What are the components of the Apache Airflow architecture?
5. What are Local Executors and their types in Airflow?
6. What is a Celery Executor?
7. How is Kubernetes Executor different from Celery Executor?
8. What are Variables (Variable Class) in Apache Airflow?
9. What is the purpose of Airflow XComs?
10. What are the states a Task can be in? Define an ideal task flow.
11. What is the role of Airflow Operators?
12. How does airflow communicate with a third party (S3, Postgres, MySQL)?
13. What are the basic steps to create a DAG?
14. What is Branching in Directed Acyclic Graphs (DAGs)?
15. What are ways to Control Airflow Workflow?
16. Explain the External task Sensor.
17. What are the ways to monitor Apache Airflow?
18. What is TaskFlow API? and how is it helpful?
19. How are Connections used in Apache Airflow?
20. Explain Dynamic DAGs.
21. What are some of the most useful Airflow CLI commands?
22. How to control the parallelism or concurrency of tasks in Apache Airflow configuration?
23. What do you understand by Jinja Templating?
24. What are Macros in Airflow?
25. What are the limitations of TaskFlow API?
26. How is the Executor involved in the Airflow Life cycle?
27. List the types of Trigger rules.
28. What are SLAs?
29. What is Data Lineage?
30. What is a Spark Submit Operator?
31. What is a Spark JDBC Operator?
32. What is the SparkSQL operator?
33. Difference between Client mode and Cluster mode while deploying to a Spark Job.
34. How would you approach if you wanted to queue up multiple dags with order dependencies?
35. What if your Apache Airflow DAG failed for the last ten days, and now you want to backfill those last ten days' data, but you don't need to run all the tasks of the dag to backfill the data?
36. What will happen if you set 'catchup=False' in the dag and 'latest_only = True' for some of the dag tasks?
37. What if you need to use a set of functions to be used in a directed acyclic graph?
38. How would you handle a task which has no dependencies on any other tasks?
39. How can you use a set or a subset of parameters in some of the dags tasks without explicitly defining them in each task?
40. Is there any way to restrict the number of variables to be used in your directed acyclic graph, and why would we need to do that?

50 Airflow Question for Data Engineers - https://www.projectpro.io/article/airflow-interview-questions-and-answers/685?source=post_page-----6b61d3b8135d-----

50 Airflow Question for Data Engineer - <https://medium.com/@shaloomathew/top-interview-questions-of-apache-airflow-6b61d3b8135d>