Astronomer Exam Guide

Airflow Fundamentals

Introduction

👋 Welcome to the Airflow Fundamentals Certification Exam study guide! We are very excited you have decided to get certified with us. This guide will give you an overview of the certification exam to help you determine how to study and when to take it. This guide covers the following sections:

- Preparation Expectations
- Exam Details
- Recommended Training and Resources
- Exam Topics
- Sample Questions

This guide covers the latest version of the exam, which was last updated on September 1, 2023. If you have any questions or want to talk to our team, reach out to us at: academy@astronomer.io.

Preparation Expectations

To pass the Airflow Fundamentals Certification, you must demonstrate an understanding of Apache Airflow's core concepts, such as architecture, DAGs, the task lifecycle, and the scheduling process. You should be comfortable recommending use cases, architectural needs, settings, and design choices for data pipelines. You should be able to trigger, debug, and retry DAGs (and their associated tasks) and use the correct views in the UI to monitor them.

At a minimum, it's recommended that you have:

- At least three months of experience using Airflow
- Completed the Airflow 101 Astronomer Academy course

Exam Details

- Format: 75 Multiple-choice questions
- Time Allotted: 60 Minutes
- Passing Score: 70% (53 correct out of 75)
- Cost: $150 USD
- Language: English
- Important Notes:
    - Once enrolled in the exam, you will have 30 days to complete it. After 30 days, the exam expires and must be purchased again.
    - Once the exam is completed (both the exam itself and any other modules), our badge vendor, Credly, will issue your digital badge to the email associated with your Astronomer Academy account.

Recommended Training and Resources

This guide is intended to be only one of the many resources you can use to prepare for the certification exam. We also recommend leveraging the following resources:

- Astronomer Academy: A large catalog of free Airflow courses taught by the Astronomer experts behind the project.
    - Airflow 101 Learning Path: A curated learning path that guides you through the foundational skills and knowledge you need to start with Apache Airflow.

Exam Topics

The Airflow Fundamentals Certification Exam covers a variety of topics about Airflow. The exam randomizes questions from a pool of over 90+ questions that are categorized by topic. This means that you may encounter a different set of questions each time you take the exam, but all topics will be covered. Use the learning outcomes below to guide your study and prepare for the exam.

Topic 1: Airflow Use Cases

- Given a specific scenario, identify if Airflow is an applicable solution.

Topic 2: Airflow Concepts

- Identify which folder the Airflow Scheduler parses when searching for new DAG files.
- Identify what an Airflow provider is.
- Identify what a DAG run is.
- Identify the role of a worker in Airflow.
- Identify which programming language Airflow primarily uses.
- Identify the purpose of an XCom.
- Identify the purpose of a DAG.
- Identify the purpose of the `default_args` DAG parameter.
- Identify the default time zone of an Airflow instance.
- Identify the role of an executor in Airflow.
- Identify the core architectural components of Airflow.
- Identify the typical journey of a task.
- Identify what happens when two DAGs share the same `dag_id`.
- Identify optional and non-optional DAG parameters.
- Identify what each of the task lifecycle stages does.
- Identify valid ways to define a DAG in Airflow.

Topic 3: Dependencies

- Identify the purpose of task-level dependencies in Airflow.
- Identify where DAG dependencies are set up in Airflow.
- Compare and contrast DAG task dependency relationships for equivalency.
- Match DAG task dependency graphs to their equivalent DAG dependency code.

Topic 4: Airflow CLI

- Identify the purpose of specific Airflow CLI commands:
  - `airflow tasks test`
  - `airflow db init`
  - `airflow info`
  - `airflow tasks test`
  - `airflow config list`
  - `airflow cheat-sheet`
  - `airflow variables`
  - `airflow users`
  - `airflow standalone`
  - `airflow version`
- Identify the impact of using the `airflow tasks test` Airflow CLI command with a DAG that has an XCom.

Topic 5: Airflow UI

- Identify the most helpful Airflow UI view to use for real-world scenarios.
  - Grid view
  - Graph view
  - Gantt view
  - DAGs view
  - Landing times view
  - Tree view
  - Calendar view
- Identify the default time for DAGs to appear in the Airflow UI
- Identify the result of deleting a DAG using the Airflow UI.
- Identify the purpose of core Airflow UI components (e.g., The Last Run Column)
- Given a specific Airflow UI, identify solutions to common issues.

Topic 6: DAG Scheduling

- Identify the purpose of each DAG scheduling parameter:
  - `catchup`
  - `start_date`
  - `end_date`
  - `schedule_interval`
- Identify which tools/commands make it possible to backfill DAGs when the `catchup` parameter is set to `false`
- Identify the default value for the `start_date` parameter
- Identify the valid values a DAG can accept for its `schedule_interval` parameter
- Given a specific DAG scheduling goal (e.g., Schedule a DAG every day at 2 PM), identify the correct DAG scheduling parameter values to accomplish the goal.
- Given specific DAG scheduling parameter values, identify if a specific scheduling goal will be accomplished

Topic 7: DAG Runs

- Given a specific DAG scheduling scenario or DAG code, identify the number of DAG runs that will occur.

Topic 8: Debugging

- Given a specific Airflow issue or completed DAG code, identify the cause of it.

Topic 9: XComs

- Identify the purpose of each XCom method:
  - `xcom push`
  - `xcom pull`
- Identify the limitations of using XComs

Topic 10: Operators

- Identify what a Transfer Operator does
- Identify what a Sensor Operator does
- Identify the purpose of the Airflow `PythonOperator`

Topic 11: Best Practices

- Given specific DAG code, identify ways to improve the code using Airflow best practices.

Topic 12: Connections

- Identify the different ways to create an Airflow connection
- Identify the correct way to create an Airflow connection in a `.env` file
- Given a specific Airflow connection string, identify the connection ID

Topic 13: Tasks

- Given DAG code, identify the number of tasks that will run when the DAG is scheduled.

Topic 14: Sensors

- Identify the default timeout value of a sensor
- Identify the mode to use in a DAG when the DAG's poke_interval parameter value is set to specific durations.
- Given the code for a sensor in a DAG, identify if the sensor is properly configured to accomplish a specific goal.

Topic 15: Variables

- Identify the purpose of an Airflow variable
- Identify the types of data that can be stored in a variable
- Identify the correct way to define a variable with a specific value in Airflow
- Identify how to fetch the value of an Airflow variable in specific formats (e.g., JSON)
- Given a specific variable name, Identify if a variable will be visible in the Airflow UI

Sample Questions

To help give you a sense of the types of questions that will be asked on the exam, we are providing five sample questions. These questions are modified versions of similar questions you might find on the exam. The answer key can be found at the end of this section.

Sample Exam Questions 1 - DAG Scheduling

What would be the Cron value of the `schedule_interval` parameter of a DAG if it needed to be triggered every two hours but only on weekends?

a. '0 */2 * * 6,7'
b. '0 2 * * 6,7'
c. '0 */2 2 * 0,1'
d. 0 0/2 * * 6,7'

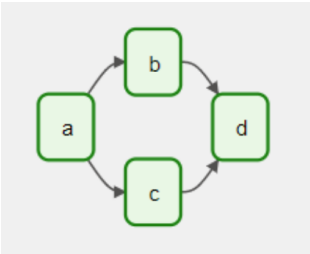
Sample Exam Question 2 - Airflow Concepts

What is the role of the Airflow scheduler?

a. To execute tasks.
b. To both trigger scheduled workflows and submit tasks to the executor to run.
c. To define how tasks are executed and on which system.
d. To define the interval of when a task is expected to be executed.

Sample Exam Question 3 - Dependencies

What task dependency relationship results in the following DAG?



a. a >> b >> c >> d
b. a >> bc >> d
c. a >> [b,c,d]
d. a >> [b,c] >> d

Sample Exam Question 4 - Debugging

Examine the following DAG:

```
from airflow import DAG
from airflow.operators.python_operator import PythonOperator
from datetime import datetime

with DAG(
   'example_dag',
   schedule_interval= '@daily',
   catchup=False
):

  task_1 = PythonOperator(
     python_callable=lambda: print("Task 1 executed.")
  )

  task_2 = PythonOperator(
     python_callable=lambda: print("Task 2 executed.")
  )

  task_1 >> task_2
```

Which of the following are issues with this DAG? (select all that apply)

a. The DAG is missing a `start_date` parameter
b. Both tasks are missing a task_id
c. The DAG has a cycle
d. The tasks are not assigned to the dag

Sample Exam Question 5 - DAG Runs

Assume a DAG is set to run daily but is paused on 2023/05/11 at 08:00 UTC. The DAG was then unpaused on 2023/05/15 at 10:00 UTC. How many DAG runs will occur if the DAG catchup parameter value is set to False?

a. 0
b. 1
c. 4
d. 5

Answer Key

1. A
2. B
3. D
4. A & B
5. B