Apache Airflow Cheatsheet

By Dhandapani Yedappalli Krishnamurthi   Sep 11, 2025

Apache Airflow Basics Cheatsheet

1. Installation

```
# Install Airflow with pip
pip install apache-airflow

# Initialize metadata DB
airflow db init

# Start webserver (default: http://localhost:8080)
airflow webserver --port 8080

# Start scheduler
airflow scheduler
```

2. Core Concepts

- **DAG** → Workflow definition
- **Task** → A single unit of work
- **Operator** → Template for a task
- **Task Instance** → Execution of a task
- **Scheduler** → Decides when to run
- **Executor** → Runs tasks (Local, Celery, Kubernetes)

---

3. Common Operators

```python
from airflow.operators.bash import BashOperator
from airflow.operators.python import PythonOperator


# BashOperator
task1 = BashOperator(
    task_id="print_date",
    bash_command="date"
)


# PythonOperator
def my_func():
    print("Hello from Python!")


task2 = PythonOperator(
    task_id="run_python",
    python_callable=my_func
```

```
)
```

---

4. DAG Example

```python
from airflow import DAG
from airflow.operators.bash import BashOperator
from datetime import datetime

with DAG(
    dag_id="demo_dag",
    start_date=datetime(2025, 9, 1),
    schedule_interval="@daily",
    catchup=False
) as dag:

    task1 = BashOperator(
        task_id="print_date",
        bash_command="date"
    )

    task2 = BashOperator(
        task_id="say_hello",
        bash_command="echo 'Hello, Airflow!'"
    )
```

```
    task1 >> task2   # Set dependency
```

5. Scheduling

- `@once` → run only once
- `@hourly` → every hour
- `@daily` → once a day
- `@weekly` → once a week
- `@monthly` → once a month
- Cron expressions also supported (e.g., `"0 12 * * *"` → run daily at 12 PM).

6. DAG Dependencies

```
task1 >> task2        # task1 before task2
task1 << task2        # task1 after task2
task1 >> [task2, task3]  # task1 before task2 & task3
```

7. Useful Airflow CLI Commands

```
airflow dags list              # List all DAGs
airflow dags trigger demo_dag    # Trigger DAG manually
airflow dags state demo_dag run1 # Check run state
airflow tasks list demo_dag      # List tasks in a DAG
airflow tasks test demo_dag task1 2025-09-10
```

8. Best Practices

Use **catchup=False** to avoid backfilling unless needed

Keep DAGs small & modular

Use **XComs** for sharing data between tasks

Add retries & alerts `(retries=3, retry_delay=timedelta(minutes=5))`

Store configs in **Variables & Connections**

With this cheatsheet, you can:

- Install & start Airflow
- Create your first DAG
- Run tasks manually or on schedule
- Monitor in the UI