Gowtham SB
www.linkedin.com/in/sbgowtham/   Instagram - @dataengineeringtamil

⚙️ PART 1: How to Install Apache Airflow on Windows (Safest Way)

Airflow doesn’t run **natively** on Windows. So we’ll use **WSL (Windows Subsystem for Linux)** or **Docker** — two safe and official ways.

◇ OPTION 1: Install Airflow on Windows using WSL (Recommended)

☑ Step-by-Step

🧱 Step 1: Install WSL and Ubuntu

In PowerShell (Admin):

```
wsl --install
```

Restart PC and Ubuntu will be installed.

📦 Step 2: Update and Install Python, pip, venv

```
sudo apt update && sudo apt upgrade
sudo apt install python3-pip python3-venv -y
```

🏗️ Step 3: Create and Activate Virtual Environment

```
python3 -m venv airflow_venv
source airflow_venv/bin/activate
```

☁️ Step 4: Set Airflow Environment Variables

```
export AIRFLOW_HOME=~/airflow
```

📥 Step 5: Install Airflow (latest)

```
pip install apache-airflow
```

Optional (with extras like PostgreSQL, Redis):

```
pip install apache-airflow[celery,postgres,redis]
```

🏗 Step 6: Initialize Airflow DB

```
airflow db init
```

👤 Step 7: Create Admin User

```
airflow users create \
    --username admin \
    --firstname Gowtham \
    --lastname SB \
    --role Admin \
    --email admin@example.com \
    --password admin123
```

🚀 Step 8: Start Webserver & Scheduler

In two terminals:

```
# Terminal 1
airflow webserver --port 8080

# Terminal 2
airflow scheduler
```

Then open: http://localhost:8080
Login: `admin` / `admin123`

🔁 PART 2: Schedule a Python Script in Airflow (DAG)

📂 Step 1: Create a DAG File

Create file: `~/airflow/dags/simple_task.py`

```
from airflow import DAG
from airflow.operators.python import PythonOperator
from datetime import datetime, timedelta

def my_task():
    with open("/home/ubuntu/airflow_output.txt", "a") as f:
        f.write(f"Task ran at: {datetime.now()}\n")
    print("Task executed!")

default_args = {
    'owner': 'gowtham',
    'retries': 1,
    'retry_delay': timedelta(minutes=2),
}

with DAG(
    dag_id='my_first_airflow_dag',
    default_args=default_args,
    description='Simple Python print task',
    start_date=datetime(2024, 1, 1),
    schedule_interval='*/2 * * * *',  # Every 2 minutes
    catchup=False,
) as dag:

    task1 = PythonOperator(
        task_id='print_time',
        python_callable=my_task,
    )

    task1
```

☑ Step 2: Wait 1-2 minutes

Then check:

- Airflow UI → DAGs → Enable `my_first_airflow_dag`
- File: `/home/ubuntu/airflow_output.txt`

You'll see:

Task ran at: 2025-05-16 23:58:01

⚠️ OPTION 2: Use Airflow with Docker (If You Prefer Containers)

You can also use the official Airflow Docker setup:
👉 https://airflow.apache.org/docs/apache-airflow/stable/howto/docker-compose/index.html

But for beginners, WSL method is easier and more flexible.

🧠 Summary Table

Step	What You Do
1	Install WSL + Ubuntu
2	Setup virtualenv + Airflow
3	Init DB and create user
4	Create DAG file in `~/airflow/dags`
5	View UI at `localhost:8080`
6	Watch output in `/home/ubuntu/airflow_output.txt`

🎯 Final Tips

- Use **crontab.guru** to generate Airflow `schedule_interval`
- You can deploy more Python scripts in DAGs folder
- DAGs will auto-refresh every 30 seconds in Airflow UI

About the Author

Gowtham SB is a **Data Engineering expert, educator, and content creator** with a passion for **big data technologies, as well as cloud and Gen AI** . With years of experience in the field, he has worked extensively with **cloud platforms, distributed systems, and data pipelines**, helping professionals and aspiring engineers master the art of data engineering.

Beyond his technical expertise, Gowtham is a **renowned mentor and speaker**, sharing his insights through engaging content on **YouTube and LinkedIn**. He has built one of the **largest Tamil Data Engineering communities**, guiding thousands of learners to excel in their careers.

Through his deep industry knowledge and hands-on approach, Gowtham continues to **bridge the gap between learning and real-world implementation**, empowering individuals to build **scalable, high-performance data solutions**.

Socials

🎥**YouTube** - https://www.youtube.com/@dataengineeringvideos

📸**Instagram** - https://instagram.com/dataengineeringtamil

📸**Instagram** - https://instagram.com/thedatatech.in

🤝**Connect for 1:1** - https://topmate.io/dataengineering/

💼**LinkedIn** - https://www.linkedin.com/in/sbgowtham/

🌐**Website** - https://codewithgowtham.blogspot.com

💻**GitHub** - http://github.com/Gowthamdataengineer

💬**Whats App** - https://lnkd.in/g5JrHw8q

📧**Email** - atozknowledge.com@gmail.com

Gowtham SB
www.linkedin.com/in/sbgowtham/ Instagram - @dataengineeringtamil

📱 **All My Socials** - https://lnkd.in/gf8k3aCH