



Apache
Airflow



Architecture

Where Are We Going?

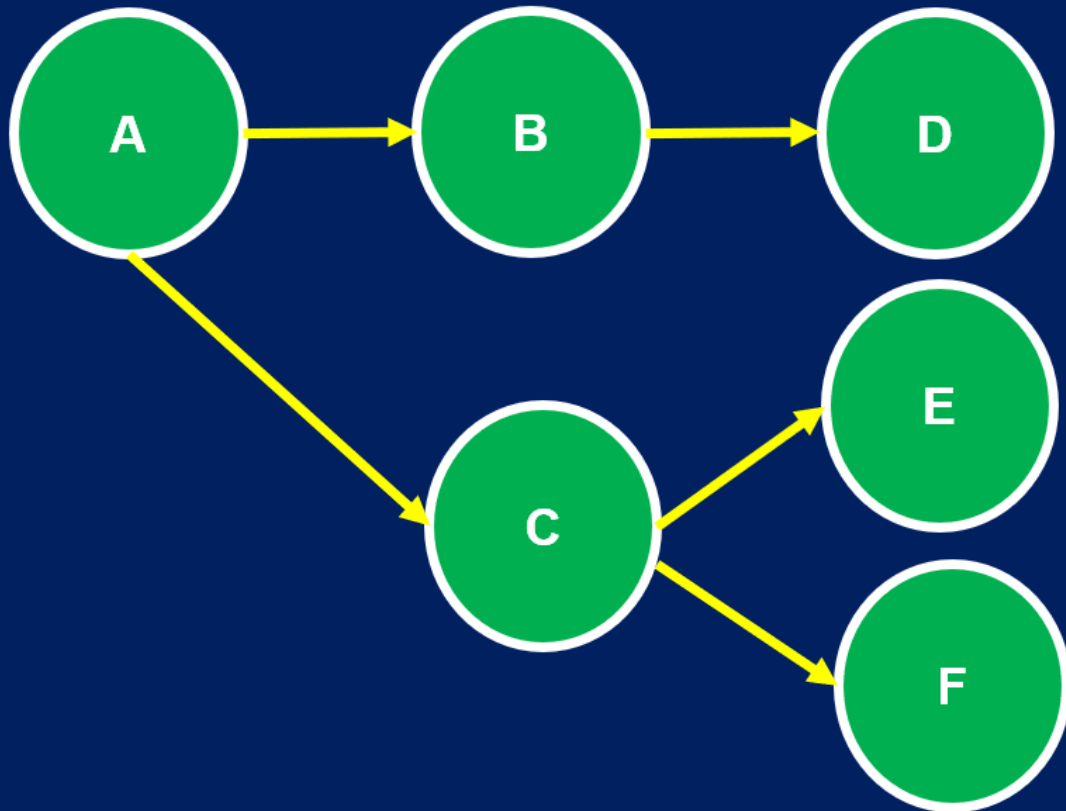
- **What is Airflow in One Slide**
- **Airflow Architecture at 50,000 Feet**
- **Airflow Concepts**
- **Airflow Considerations and Parsimony**

What is Airflow?

Airflow is an Open-Source Linux Based Workflow Engine

Airflow = AirBnB + Workflow

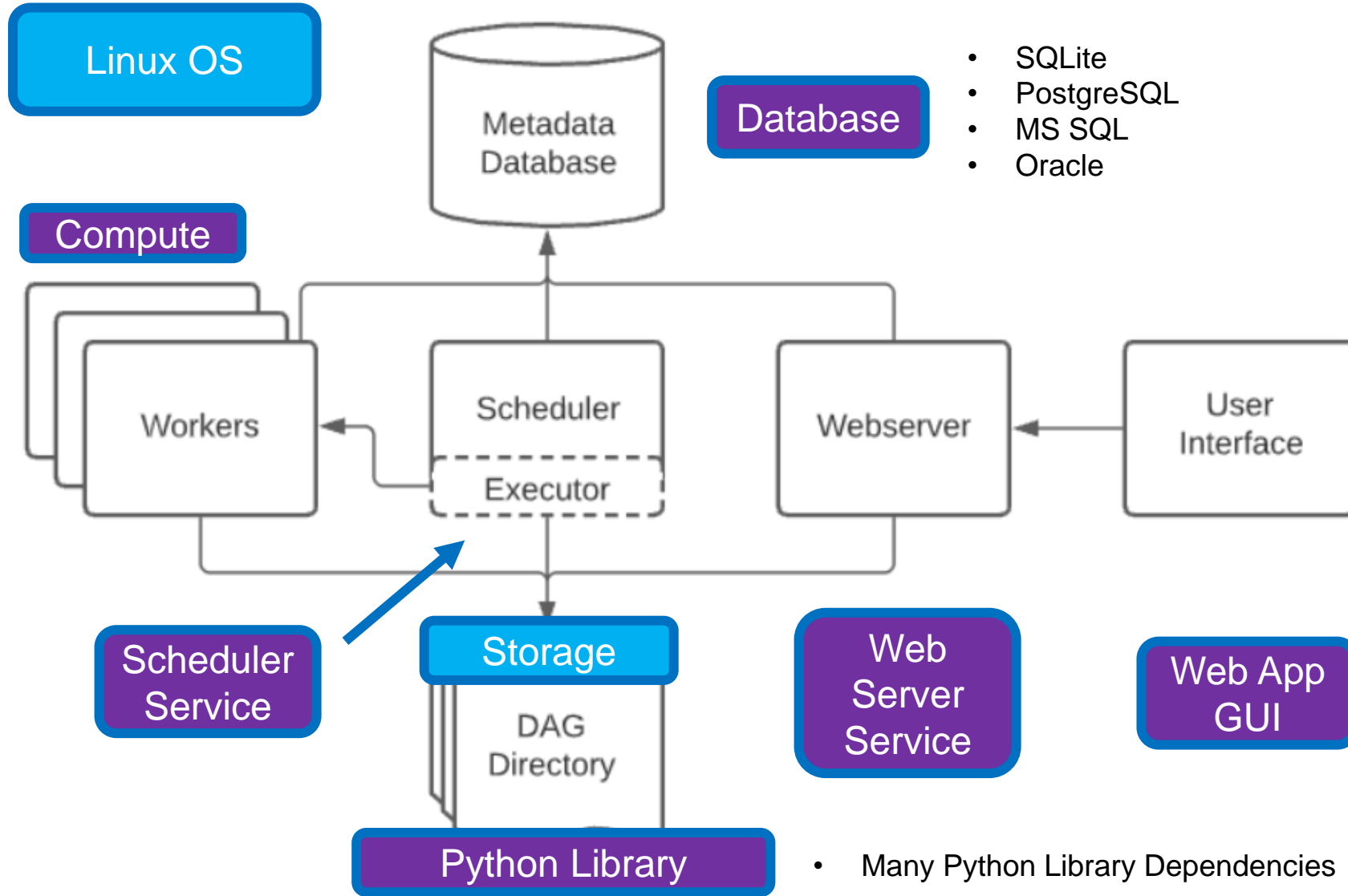
Directed Acyclic Graph



- Vertices (Objects)
- Edges (Relationships)
- Objects and Their Relationships to Other Objects
- Directional (Directed)
- No Loops!!! Acyclic

- **Linux Only**
- **Built on Python**
- **Workflow Engine Service**
- **Web Server & Web App UI**
- **Leverages Python Libraries**

Architecture



➤ **Scale = Complexity**

Concepts

Workloads

Control Flow

User Interface

Executors

Local Executors

- Debug Executor
- Local Executor
- Sequential Executor **Default**

Remote Executors

- Celery Executor
- CeleryKubernetes Executor
- Dask Executor
- Kubernetes Executor
- LocalKubernetes Executor

Security

Security

- Access Control
- API
- Flower
- Kerberos
- Webserver
- Workload
- Secrets

Workloads (BaseOperator)

Operators

Extensibility

- Does Some Work of a Specific Type:
 - BashOperator
 - PythonOperator
 - EmailOperator
 - Database Operators
 - External Service Operators

Sensors

- Wait Operator which can be:
 - Time Based
 - Event Based

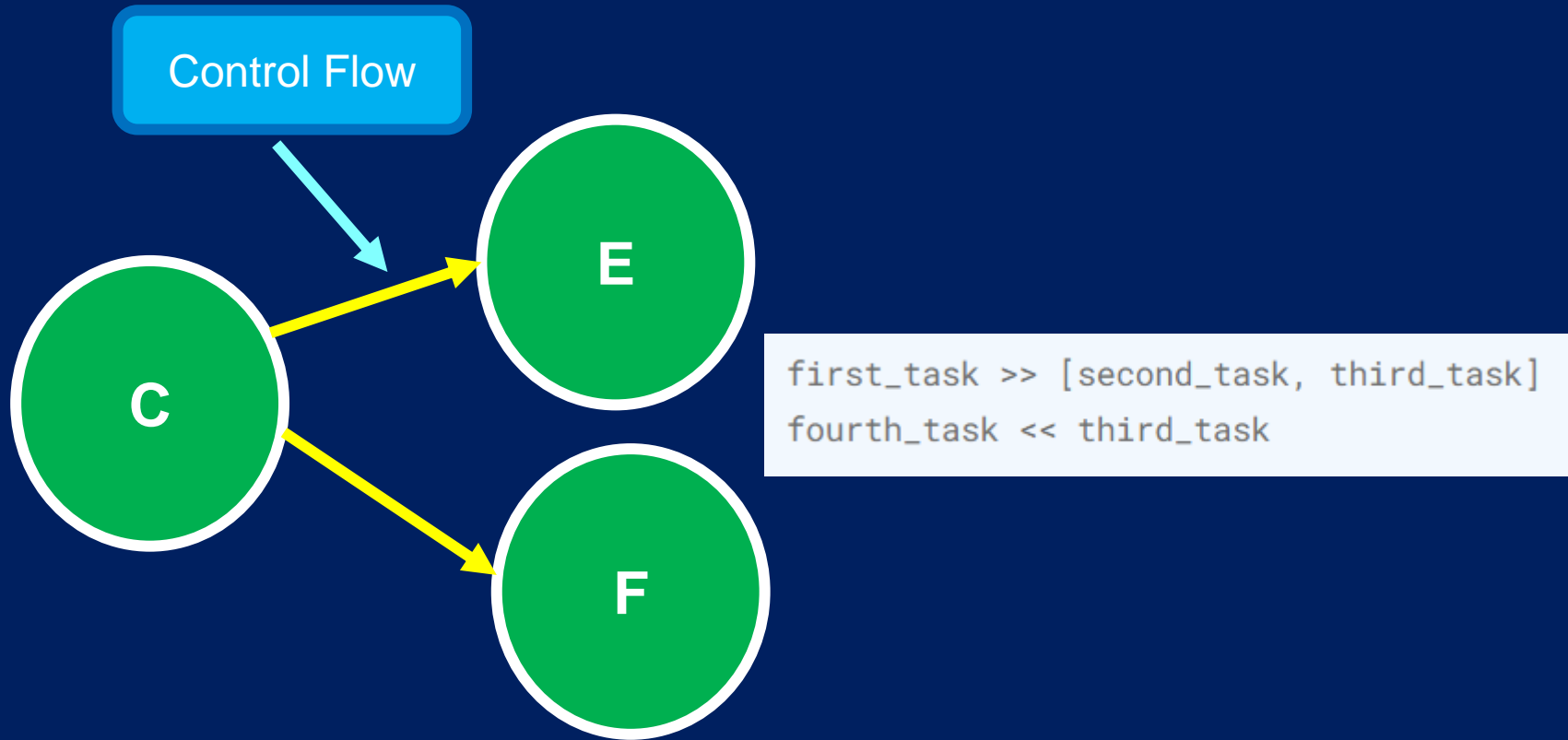
TaskFlow

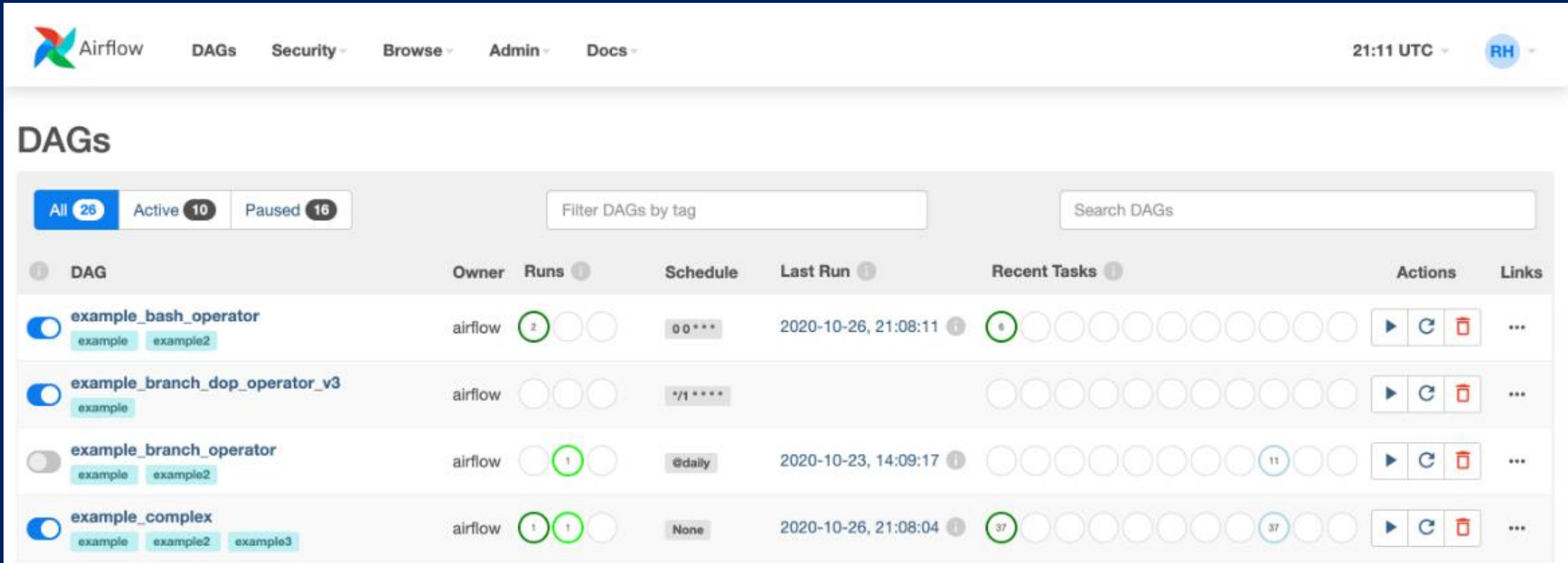
- Custom Python function packaged up as a Task.

```
from airflow.decorators import task
from airflow.operators.email import EmailOperator

@task
def get_ip():
    return my_ip_service.get_main_ip()
```

Control Flow





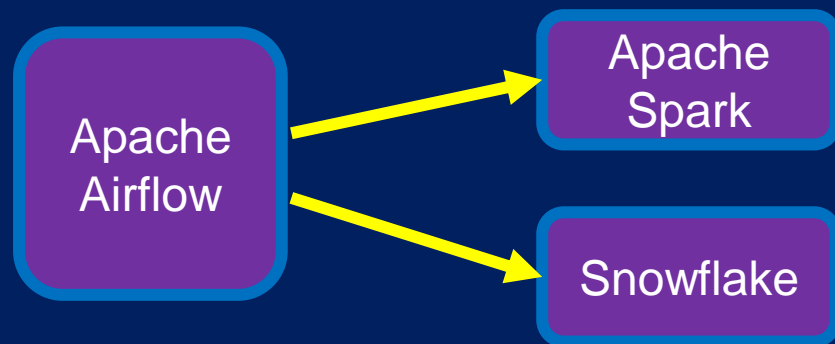
The screenshot shows the Apache Airflow web interface. At the top, there's a navigation bar with the Airflow logo, a menu (DAGs, Security, Browse, Admin, Docs), the current time (21:11 UTC), and a user profile (RH). Below the navigation bar, the main heading is "DAGs". Underneath, there are filters for "All" (26), "Active" (10), and "Paused" (16). There are also input fields for "Filter DAGs by tag" and "Search DAGs". The main content is a table listing DAGs with columns: DAG, Owner, Runs, Schedule, Last Run, Recent Tasks, Actions, and Links. The table contains four rows of example DAGs.

DAG	Owner	Runs	Schedule	Last Run	Recent Tasks	Actions	Links
<input checked="" type="checkbox"/> example_bash_operator example example2	airflow	<div><div>2</div><div></div><div></div></div>	00***	2020-10-26, 21:08:11	<div><div>6</div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	<div><div>▶</div><div>↺</div><div>🗑️</div></div>	...
<input checked="" type="checkbox"/> example_branch_dop_operator_v3 example	airflow	<div><div></div><div></div><div></div></div>	*1****		<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	<div><div>▶</div><div>↺</div><div>🗑️</div></div>	...
<input type="checkbox"/> example_branch_operator example example2	airflow	<div><div></div><div>1</div><div></div></div>	@daily	2020-10-23, 14:09:17	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	<div><div>▶</div><div>↺</div><div>🗑️</div></div>	...
<input checked="" type="checkbox"/> example_complex example example2 example3	airflow	<div><div>1</div><div>1</div><div></div></div>	None	2020-10-26, 21:08:04	<div><div>37</div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	<div><div>▶</div><div>↺</div><div>🗑️</div></div>	...

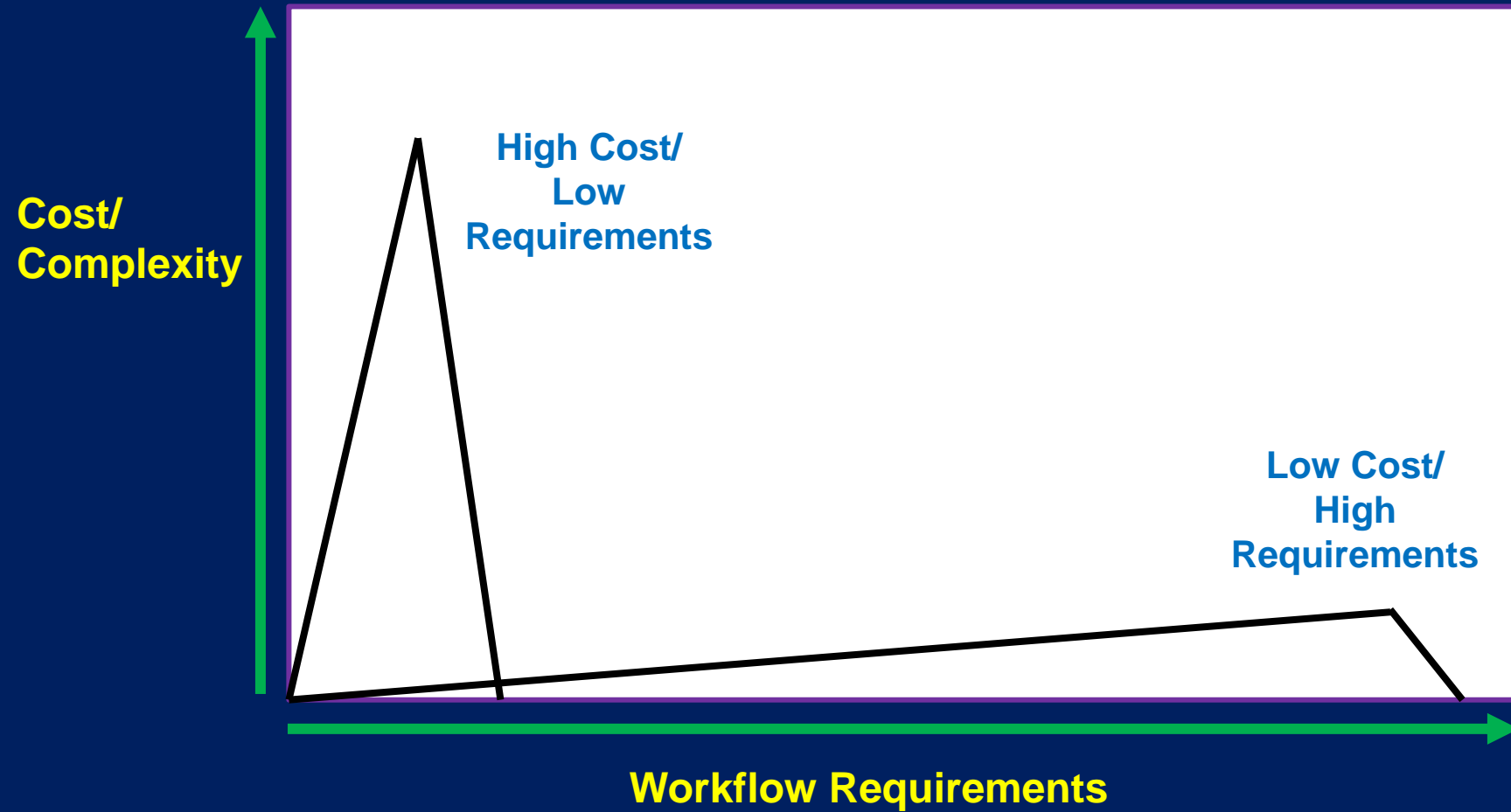
Does Not Support Development

Airflow Considerations

- **Each DAG Must End Before It Can Be Executed Again**
- **Does Not Support Streaming Data**
- **Using Airflow to Execute Jobs External to the Airflow Environment Adds Complexity and Cross Dependencies**



- Your Goal is to Find the Optimal Balance



- **What is Airflow in One Slide**
- **Airflow** **Thank You!** **50,000 Feet**
- **Airflow Concepts**
- **Airflow Considerations and Parsimony**