CS 327E Class 11

April 27, 2020

Announcements

- [Milestone 12](#) and [Presentation Schedule](#)

- GCP credits

- CIS survey

1) An Airflow `DAG` represents a _____ ?

A. A single task
B. Dependency between tasks
C. Task input data
D. Workflow

2) How is a `DAG` implemented in Airflow?

A. Standard SQL file
B. Jinja template file
C. Standard Python file
D. Beam Python file

3) How is the unit of work for a task executed?

A. By a single `Operator`
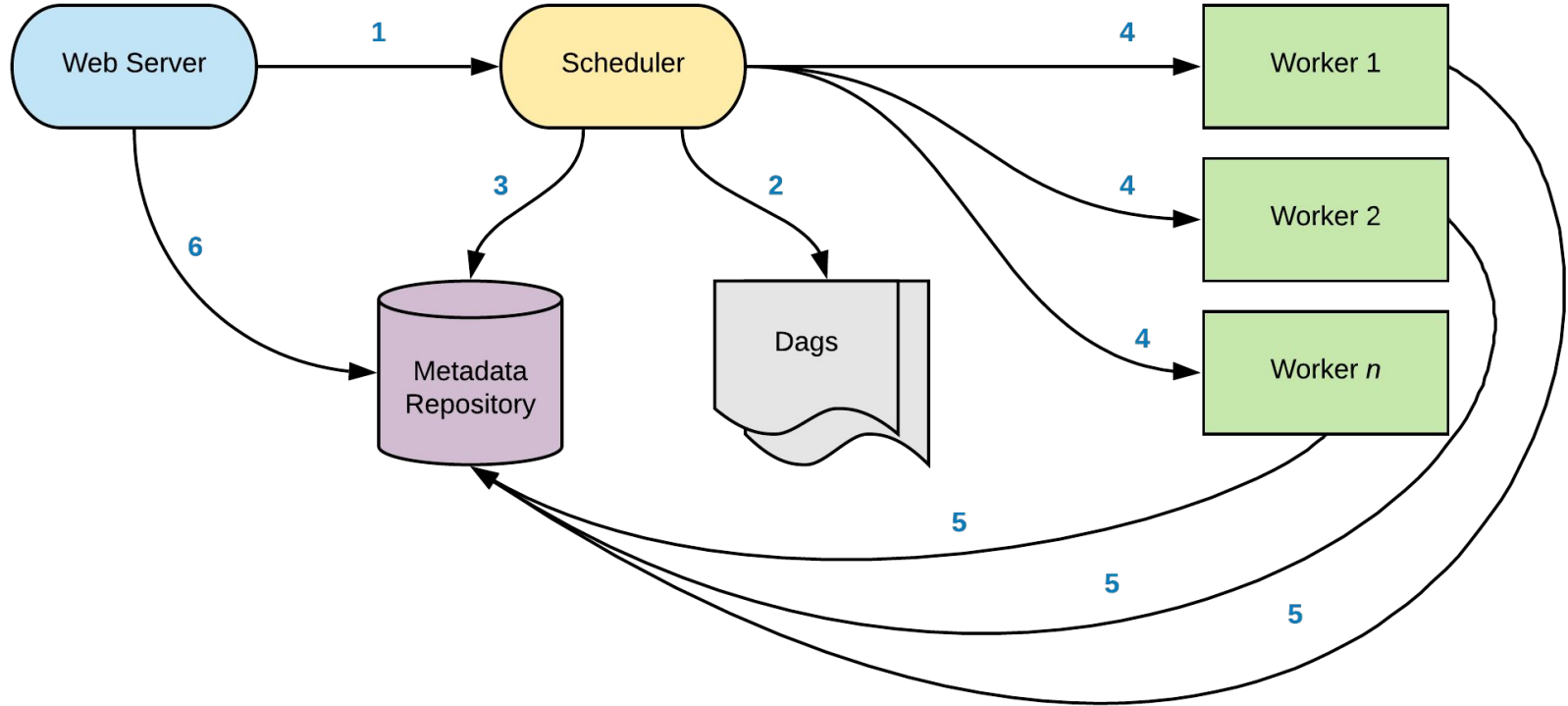B. By multiple `Operators`
C. By a `Dag`
D. By a `Subdag`

4) What does the statement `t1 >> t2` mean where `t1` and `t2` each represent a task?

A. `t1` runs first followed by `t2`
B. `t2` runs first followed by `t1`
C. `t1` and `t2` run in parallel
D. `t1` and `t2` do not have an ordering

5) A `trigger rule` specifies the conditions under which a task can _____.

A. run
B. sleep
C. abort
D. finish

Airflow Architecture

Airflow Basic Commands

```
airflow list_dags

airflow list_tasks <dag> --tree

airflow test <dag_name> <task_name> <yyyy-mm-dd>

airflow backfill <dag_name> -s <yyyy-mm-dd> -e <yyyy-mm-dd>

airflow clear <dag_name> -s <yyyy-mm-dd> -e <yyyy-mm-dd>
```

Airflow Setup on Jupyter Notebook

https://github.com/cs327e-spring2020/snippets/wiki/Airflow-Setup

Hands-on Exercise

Pull down latest snippets:

```
git clone https://github.com/cs327e-spring2020/snippets.git
or git pull origin master
```

We'll be using these files:

```
country1.py - country4.py
college_workflow1.py - college_workflow2.py
Student_beam_dataflow2.py
Takes_beam_dataflow.py
Teacher_beam_dataflow.py
```

Practice Problem 1

Run and fix `college_workflow1.py`

Practice Problem 1

Run and fix `college_workflow1.py`

What was the cause of the error?

A. Syntax error
B. Logic error
C. All of the above
D. None of the above

Practice Problem

Run and fix `college_workflow2.py`

Practice Problem

Run and fix `college_workflow2.py`

What was the cause of the error?

A. Syntax error
B. Logic error
C. All of the above
D. None of the above