

CS2032 - Cloud Computing (Ciclo 2024-2)

Orquestar pipelines de datos

Semana 15 - Taller 2: Apache Airflow

ELABORADO POR: GERALDO COLCHADO

Con apoyo de Asistente de Cátedra y Laboratorio:

- Sofía García (sofia.garcia@utec.edu.pe)

Contenido

Orquestar pipelines de datos

1. **Objetivo del taller 2**
2. Ejercicio 1: Pipeline de datos complejo
3. Ejercicio propuesto
4. Cierre

Orquestar pipelines de datos

Objetivo del Taller 2

- Aprender a implementar Pipeline de Datos complejos

Contenido

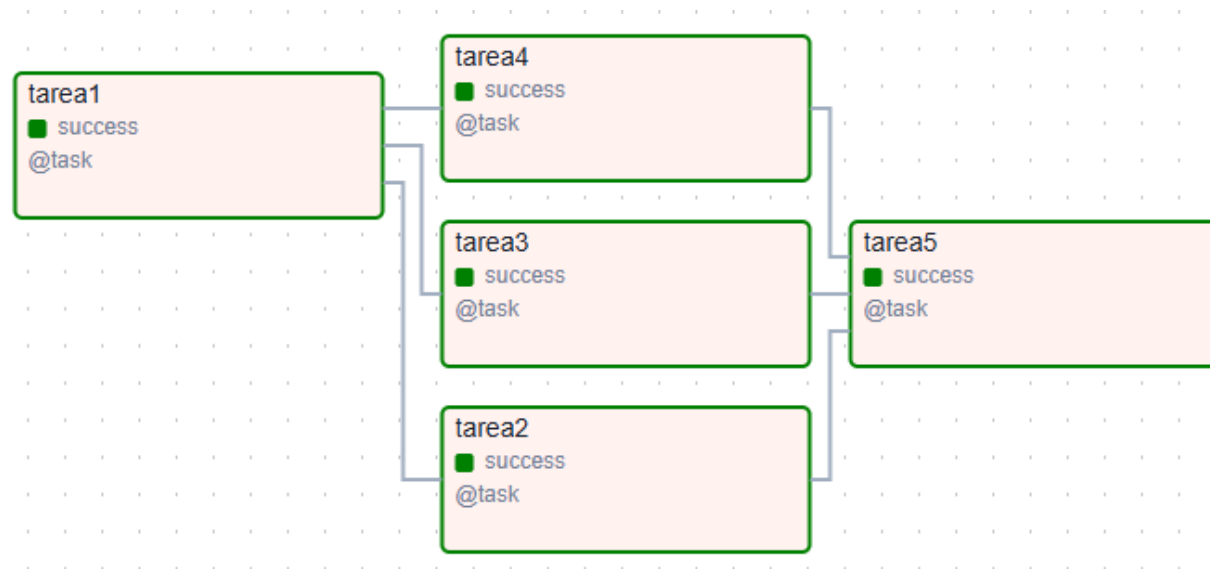
Orquestar pipelines de datos

1. Objetivo del taller 2
2. **Ejercicio 1: Pipeline de datos complejo**
3. Ejercicio propuesto
4. Cierre

Ejercicio 1

Pipeline de datos complejo

Implementar el siguiente Pipeline de datos que se ejecute una vez por día y tenga los tags “complejo” y “diario”



- **Paso 1:** Ingrese a “MV airflow”, verifique que se estén ejecutando los contenedores del airflow.
- **Paso 2:** Ingrese a directorio `/home/ubuntu/airflow/dags/` y cree el archivo “`pipeline_datos_complejo.py`”. Espere aprox. 5 minutos o en su defecto reinicie el Apache airflow en `/home/ubuntu/airflow/` para que se muestre el nuevo DAG.
- **Paso 3:** Prenda el DAG “`pipeline_datos_complejo`” y analice

Contenido

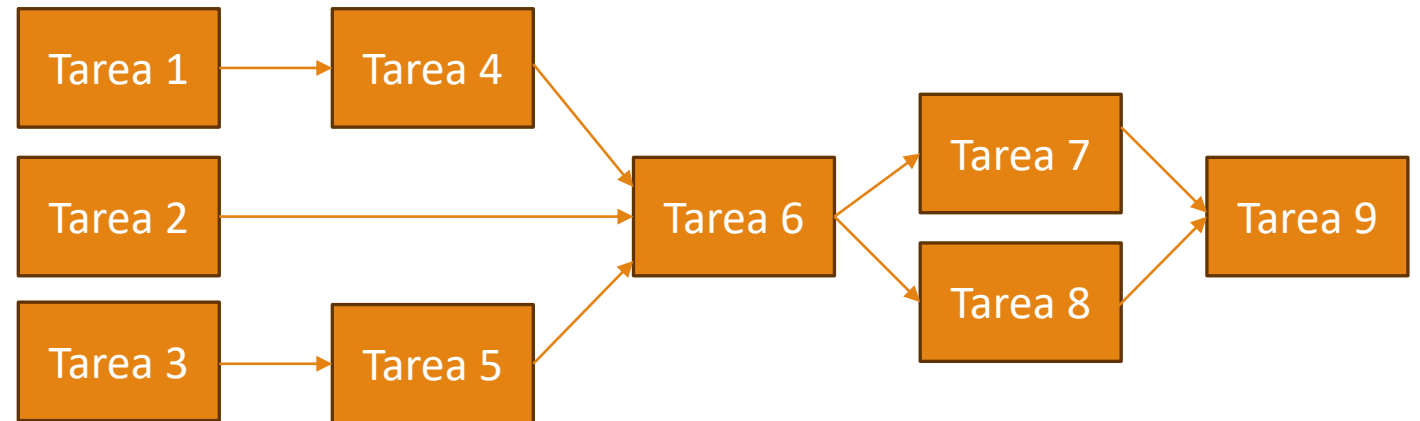
Orquestar pipelines de datos

1. Objetivo del taller 2
2. Ejercicio 1: Pipeline de datos complejo
3. **Ejercicio propuesto**
4. Cierre

Ejercicio propuesto

Pipeline de datos complejo

- Implementar el siguiente Pipeline de datos que se ejecute cada 3 horas y tenga los tags “complejo” y “3horas”
- Suba las evidencias al padlet indicado por el docente



Contenido

Orquestar pipelines de datos

1. Objetivo del taller 2
2. Ejercicio 1: Pipeline de datos complejo
3. Ejercicio propuesto
4. **Cierre**

Cierre:

Orquestar pipelines de datos - Qué aprendimos?

- Implementar Pipeline de Datos complejos

Gracias

Elaborado por docente: Geraldo Colchado