

Proyecto: Platzi explora el espacio



Primera **EdTech** de la historia en lanzar un satélite al espacio incursionando al mismo tiempo en la exploración espacial.

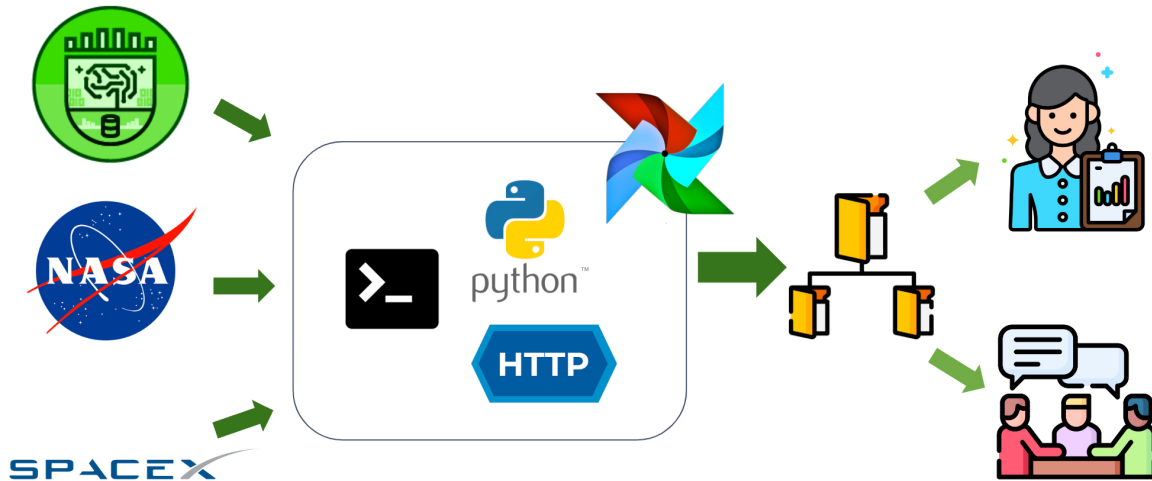
Objetivos

Los equipos de analistas y marketing de Platzi necesitan datos de los estudiantes que han accedido al satélite e información del historial de eventos de SpaceX, por lo tanto necesitamos que nos ayudes a ejecutar las siguientes tareas.

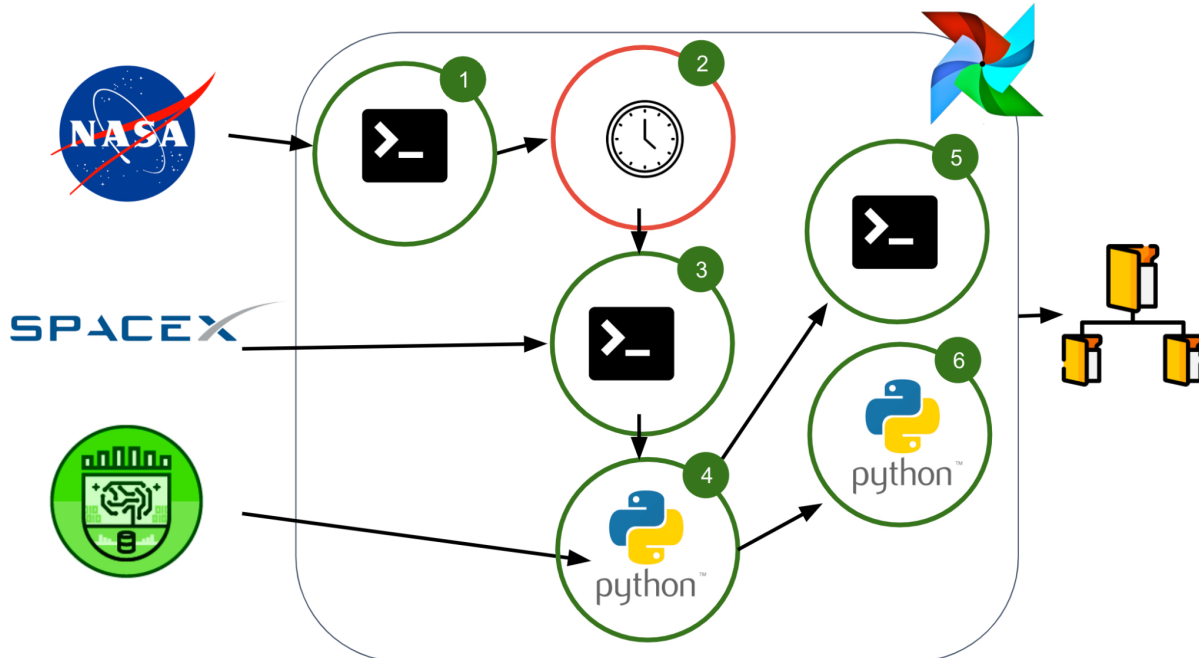
1. Esperar a que la NASA nos dé autorización para acceder a los datos del satélite.
2. Recolectar datos del satélite y dejarlos en un fichero.

3. Recolectar datos de la API de SpaceX y dejarlos en un fichero.
4. Enviar un mensaje a los equipos de que los datos finales están disponibles.

Ejemplo: Workflow



Ejemplo: Grafo



La idea de este proyecto es que apliques el conocimiento adquirido en este curso, y te ánimo a qué investigues nuevas cosas y las apliques en este proyecto. El propósito de este curso es que aprendas a usar Airflow, por lo tanto te voy a proporcionar los comandos bash y scripts de Python que necesitarás para realizar el proyecto:

1) Comando bash para simular la respuesta de la NASA de confirmación:

```
bash_command='sleep 20 && echo "OK" >
/tmp/response_{{ds_nodash}}.txt'
```

2) Comando bash para obtener los datos de SpaceX:

```
bash_command="curl -o /tmp/history.json -L
'https://api.spacexdata.com/v4/history'"
```

3) Script Python para simular la respuesta del satélite:

```
def _generate_platzi_data(**kwargs):
    import pandas as pd

    data = pd.DataFrame({"student": ["Maria Cruz", "Daniel Crema",
    "Elon Musk", "Karol Castrejon", "Freddy Vega"],
    "timestamp": [kwargs['logical_date'],
    kwargs['logical_date'], kwargs['logical_date'],
    kwargs['logical_date']]})

    data.to_csv(f'/tmp/platzi_data_{kwargs['ds_nodash']}.csv',
    header=True)
```

Estos son ejemplos de como puedes simular los datos y creación de ficheros, puedes realizarlo de otras maneras.

Notas:

1) Si quieres visualizar los datos generados por Platzi puedes crear una tarea que ejecute el siguiente comando:

```
bash_command='ls /tmp && head /tmp/platzi_data.csv'
```

2) Te invito a que crees tu propio operador para enviar alertas a los equipos y utilices los templates para notificar la fecha de ejecución {{ds}}

3) Puedes utilizar el `schedule_interval` y fechas que prefieras.

En el repositorio podrás encontrar ayudas en el fichero 11-Proyecto.py para resolver el proyecto, pero te invito a que lo utilices solo cuando estés bloqueado <https://github.com/platzi/curso-airflow>