DeepSpace Technologies B.V

CAPSTONE PROJECT
(confidential)

You're a brilliant rocket scientist driven by a deep passion for exploring the mysteries of space working for DeepSpace Technologies. Your boss[1], an ambitious and visionary woman, approaches you with an intriguing assignment. She tasks you to create a new data pipeline using Apache Airflow to gather crucial information about rocket launches happening worldwide. The objective is to track launches from the last 60 days and those scheduled for the future by making use of The SpaceDevs API.

https://lldev.thespacedevs.com/2.2.0/launch

Your boss recognizes your expertise in rocket science and data analysis, making you the ideal candidate for this mission. She explains that building a robust system will allow you to collect comprehensive information about rocket launches, process the data, and derive valuable insights.

With this information, you connect to the API and embark on an exciting journey to explore the data. As you delve into the API, you uncover a treasure full of details about rocket missions. The API becomes your gateway to the world of space exploration, revealing insights about rockets ids, name, mission names, launch statuses, country information, launch service providers, and their respective types.

Moving forward, you dive into the exciting task of creating an operator specifically designed to assess whether a rocket launch is scheduled for today. Your boss emphasizes the importance of this operator as it will act as a critical checkpoint in the pipeline. If there are no rocket launches scheduled for the day, your system must halt further execution to avoid processing unnecessary or outdated data. By including this

---

[1] because despite being a laureated scientist, you still have a boss

operator, you ensure that your pipeline remains efficient and focused on the most relevant and current information.

As you construct the data pipeline, you meticulously design a framework to store the extracted data efficiently. To ensure streamlined access and optimize analysis, you choose to organize the data using parquet files to store daily files with relevant information over the launches.

Some question that are still in the air[2]:

- How do we make sure the API is actually working on a given day?
- How do you pass the response of the API call to our pipeline?
- Do we need to store anything locally?
- How do you tell API that you need data for only one date?

Recognizing the importance of seamless collaboration and accessibility, your boss suggests storing the organized data in a secure and scalable cloud storage solution. This collaborative environment fosters knowledge exchange and propels the boundaries of space exploration.

Furthermore, to unlock the full potential of your analysis, your boss recommends integrating the data pipeline with BigQuery, a powerful analytical database. By loading the daily rocket launches data into a BigQuery table, you gain access to advanced querying capabilities and scalable processing.

Additionally, your boss presents an optional step: integrating a PostgreSQL database into the system. This flexible storage solution caters to the preferences of team members who favor working with PostgreSQL. The inclusion of this option fosters collaboration and ensures seamless data access, regardless of the analytical tools individuals choose to employ.

As you approach the culmination of the assignment, you reflect on your boss's guidance and vision. Despite her occasionally demanding nature, her direction has propelled you to develop a robust data pipeline using Apache Airflow. By extracting pertinent information from the provided API, storing it efficiently, and harnessing advanced tools like local storage, BigQuery and PostgreSQL, you can now gain a comprehensive overview of rocket launches happening worldwide. This achievement allows you to push the boundaries of scientific exploration and unravel the mysteries of our vast universe.

Good Luck!

To the Infinity…and beyond 🚀

---

[2] pun not intended