DST Airlines

Program : Data Engineer

Difficulty : 8.5/10

Description:
Nowadays, it is able to get information about flights all over the world and to track an aircraft in real time. We can look at this site as an example. The aim here is to get as close as possible to it by going through the APIs of different aviation companies.

Steps	Description	Goal	Courses / Masterclass / Templates	Conditions of validation
1	Collecting data	Use the Lufthansa API to retrieve flight data. You can test the different routes in the Lufthansa API using this link. You will have to retrieve various information such as IATA codes (you could do this by web scraping) This step is important, you need to understand the data you can retrieve and make a choice of routes to use. There is also the API of International Airlines, but you may have problems using it.	You will need to use the requests library or you can use the Postman tool. Webscraping (Selenium, Beautiful Soup)	Explanatory file of the treatment and the different data accessible (doc / pdf) An example of collected data.
2	Data modeling	There are several options available to us. In the previous step, we observed that there are several "types" of data. We will qualify data as **fixed** data such as airport information and **variable** data such as flight information. This diversification of data will lead to the use of different databases.	142 - SQL Elasticsearch 143 - MongoDB Neo4j	A relational database UML Diagram A file who creates and queries the SQL database . Same files for a Elastic/Mongo/Neo4j DataBase
3	Data consumption	Implement a DashBoard using Dash to view your relational database data. Here you will need to select the values that you feel are relevant as a Power BI DashBoard. Here, there is no use of Machine Learning on this data (you can propose ideas to your mentor). This is	Dash Plotly FastAPI	Dash API FastAPI Docker, docker-compose file

DataScientest.com
Agrément organisme de formation 11755665975
+09 80 80 79 49
2 place de Barcelone, 75016 Paris

		appropriate data for a reporting tool, instead we will have to make an API to query these different databases. Indeed, with the database on Neo4j, we could represent our flight from one airport to another. Make a Docker container of each component of the project(BDD,API) and make a functional docker-compose.	Docker	
4	Automation of flows	The data must be retrieved from the Lufthansa API at a defined rate and sent to the various consumers of the data.	Airflow	Python file for Airflow
5	Defense	Demonstrate their application and explain the reasoning behind their project.	X	Defense Documentation

DataScientest.com
Agrément organisme de formation 11755665975
+09 80 80 79 49
2 place de Barcelone, 75016 Paris