

Cursus concerné : Data Engineer

Difficulté : 8.5/10

Description détaillée :

De nos jours, il est possible d'avoir des informations sur les vols dans le monde entier et de traquer en temps réel un avion. Nous pouvons observer ce [site](#) en guise d'exemple. Le but ici est de s'y approcher le plus possible en passant par des API de différentes compagnies aériennes.

Etape	Description	Objectif	Modules / Masterclass / Templates	Conditions de validation du projet
1	Récolte des données	<p>Passer par l'API de Lufthansa pour récupérer des données sur les vols.</p> <p>Vous pouvez tester les différentes routes de l'API de Lufthansa à l'aide de ce lien. Vous pourrez être amené à aller puiser différentes informations comme les codes IATA.</p> <p>Cette étape est importante, vous devez comprendre les données que vous pouvez récupérer et faire un choix des routes à utiliser.</p> <p>Il y a aussi l'API de International Airlines, mais il se peut que vous ayez des soucis pour y accéder..</p>	<p>Utilisation de la librairie requests ou de l'outil Postman (pour tester)</p> <p>Techniques de webscraping</p>	<p>Fichier explicatif du traitement et des différentes données accessible (doc / pdf)</p> <p>Un exemple de données collectées.</p>
2	Architecture des données	<p>Il y a plusieurs options qui s'offrent à nous. Lors de la précédente étape, nous avons observé qu'il y a plusieurs "types" de données. Nous allons qualifier des données de fixe comme les informations sur les aéroports et de variables, les informations sur les vols.</p> <p>Cette diversification de données va conduire à une utilisation de différentes bases de données.</p>	<p>142 - SQL</p> <p>Architecture des données</p> <p>Elasticsearch</p> <p>MongoDB</p> <p>Neo4j</p>	<p>Une base de données relationnelle</p> <p>Diagramme UML</p> <p>Un fichier de requête SQL pour montrer que c'est bien fonctionnel</p> <p>Même rendu mais exemples de requêtes Elastic/Mongo</p>



3	Consommation de la donnée	<p>Ici, le but principal du projet sera de faire un algorithme de prédiction de retard des avions grâce aux données que vous avez précédemment stockées dans vos Bases de données</p> <p>Il faudra aussi mettre en place ML Flow pour pouvoir faire du versioning de Modèles.</p>	<p>DE120</p> <p>ML Flow</p>	<p>Algorithme de Machine Learning</p> <p>ML Flow Fonctionnel</p>
4	Déploiement	<p>Il faut créer une API pour requêter ces différentes bases de données.</p> <p>Faire un conteneur Docker de chaque composant du projet (BDD, API) et faire un docker-compose fonctionnel.</p> <p>Mesurer la dérive des données</p>	<p>FastAPI / Flask</p> <p>Docker</p> <p>Data Drift</p>	<p>API FastAPI</p> <p>Conteneur, Fichier yaml du docker-compose</p>
5	Automatisation et Monitoring	<p>Il faut récupérer les données de l'API Lufthansa selon un rythme bien défini pour l'envoyer aux différents consommateurs de la donnée.</p> <p>Créer un pipeline simple de CI pour déployer des nouveautés</p> <p>Il faudra aussi monitorer l'application en production ainsi que les logs API</p>	<p>Cronjob</p> <p>Airflow</p> <p>Prometheus Grafana</p>	<p>Fichier python du DAG</p>
6	Soutenance	<p>Démonstration de leur appli et explication du raisonnement effectué lors de leur projet.</p>	<p>X</p>	<p>Soutenance Rapport</p>