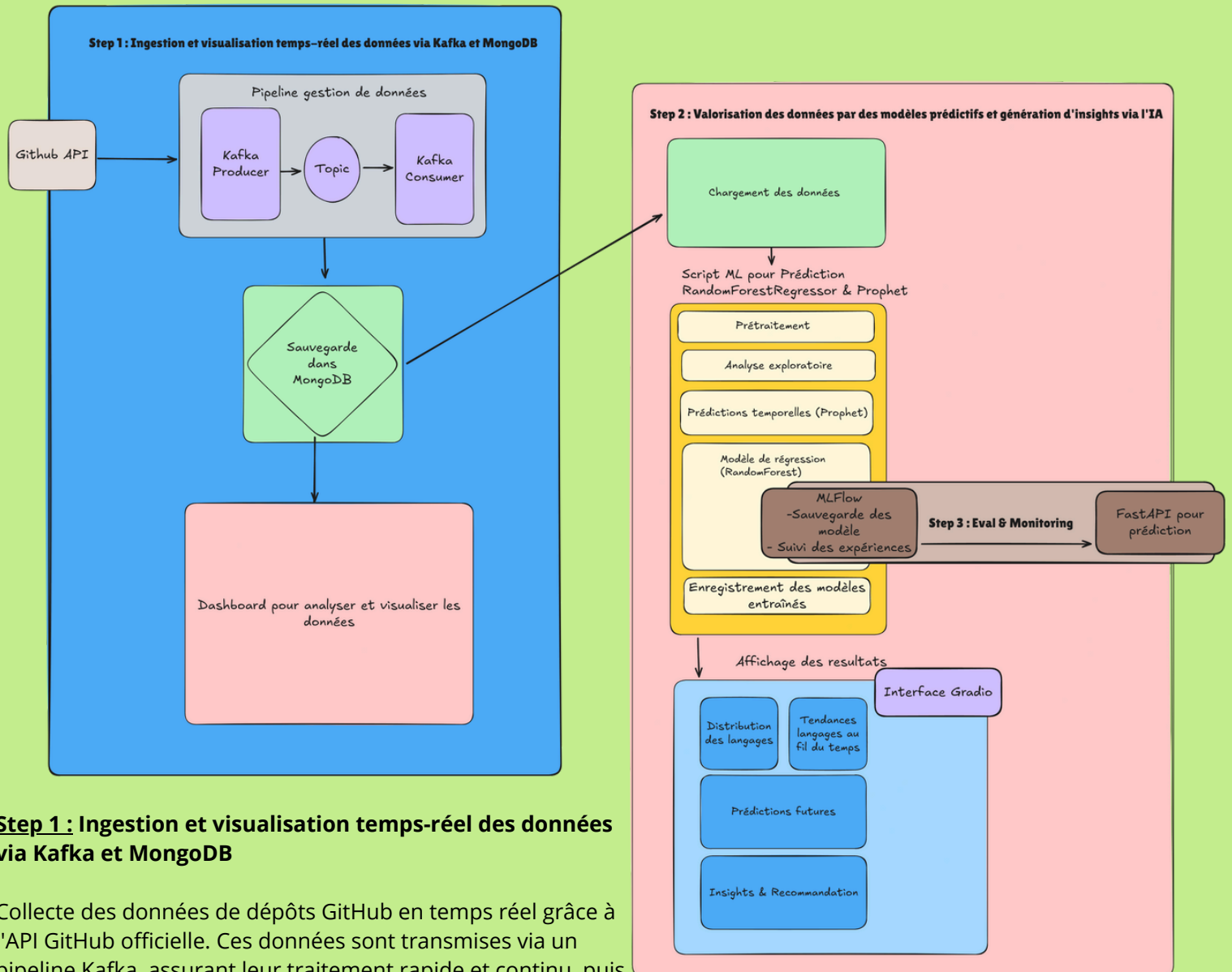


# Projet BigData & MLOps



## Step 1 : Ingestion et visualisation temps-réel des données via Kafka et MongoDB

Collecte des données de dépôts GitHub en temps réel grâce à l'API GitHub officielle. Ces données sont transmises via un pipeline Kafka, assurant leur traitement rapide et continu, puis stockées dans MongoDB. Un dashboard interactif permet alors aux utilisateurs de visualiser et explorer immédiatement les données collectées.

## Step 2 : Valorisation des données par des modèles prédictifs et génération d'insights via l'IA

Les données stockées dans MongoDB sont récupérées pour un prétraitement approfondi et une analyse exploratoire poussée. Des modèles prédictifs sont développés : Prophet pour anticiper les tendances temporelles des langages de programmation, et RandomForest pour prédire l'engagement des dépôts GitHub. Les résultats sont présentés sous forme d'insights et recommandations au travers d'une interface utilisateur intuitive (Gradio).

## Step 3 : Évaluation, monitoring des modèles et déploiement temps-réel avec MLflow et FastAPI

Les modèles entraînés sont suivis, versionnés et sauvegardés grâce à MLflow, garantissant leur reproductibilité et permettant une gestion efficace des expériences. Un service REST basé sur FastAPI est mis en place pour exposer ces modèles sous forme d'une API temps-réel, permettant ainsi une intégration facile dans d'autres applications ou services tiers.