Pipeline automatisé d’extraction, de traitement et de publication de recettes :

Étapes pour réaliser ce projet Contexte On a un pipeline avec 3 agents (ou microservices) qui communiquent via Kafka :

* Agent 1 : Extraction (extraction des recettes depuis des sites web via Selenium)
* Agent 2 : Traitement (analyse, traduction, enrichissement via API Deepsek + GCP API)
* Agent 3 : Publication (envoi des recettes traitées vers un système blog) Kafka assure la communication asynchrone et découple ces agents.

**Agent 1 : Extraction avec Selenium**

• Automatiser la navigation sur des sites source.

• Extraire les données brutes (titre, ingrédients, étapes, images…).

• Traiter les données ( upload image vers imagekit

• Envoyer les données extraites sous forme Excel via messages Kafka vers un topic “raw\_recipes”.

**Agent 2 : Traitement via API Deepsek et GCP**

• Consommer les messages Kafka du topic “raw\_recipes”.

• Traiter les données :

o convert excel a un page html via API Deepsek.

• Publier les résultats enrichis dans un autre topic Kafka “processed\_recipes”.

**Agent 3 : Publication**

• Consommer le topic “processed\_recipes”.

• Charger ou publier les recettes traitées :

o Stockage dans une base de données (BigQuery, Firestore).

o Publication sur un blog via une API Blogger.

• Log des opérations.