**Feuille de route - Automatisation du projet avec AWS et Mage.AI**

**Objectif**

Automatiser l’ensemble du workflow de collecte, transformation et stockage des données issues de requêtes API et de web scraping, en utilisant **Mage.AI**, **AWS (S3, Elasticsearch, Lambda, EC2)** et **Docker**. Les données seront visualisées via **Power BI ou Streamlit**.

**Étapes détaillées**

**1. Collecte des données (Scraping & API) ✅ (*OK, déjà prêt !*)**

* Exécuter le notebook Jupyter pour récupérer et transformer les données sous forme de **DataFrame pandas**.
* Générer un fichier JSON stocké en local.
* ✅ **Status** : Fait.

**2. Stockage automatique des données sur AWS S3 ✅ (*Bucket déjà créé !*)**

* **Écrire un script Python** pour envoyer automatiquement le fichier JSON vers **AWS S3**.
* Ajouter le script d’upload dans un pipeline **Mage.AI** pour une exécution automatique.
* **Tester l’upload manuel** avec **boto3** avant d’intégrer dans Mage.AI.
* ✅ **Status** : En cours.

**3. Automatisation du pipeline avec Mage.AI et Docker**

* Définir un **pipeline Mage.AI** pour :
  1. Déclencher l’exécution du script de scraping.
  2. Transformer les données et les préparer.
  3. Envoyer automatiquement le fichier JSON sur S3.
* Intégrer **Mage.AI** avec **Docker** pour garantir une exécution stable.
* Tester et valider que le pipeline fonctionne correctement.

**4. Indexation des données dans AWS Elasticsearch**

* Déployer un cluster **AWS OpenSearch (Elasticsearch)**.
* **Écrire un script Python** qui récupère les fichiers depuis S3 et les indexe dans Elasticsearch.
* Ajouter ce script dans le pipeline **Mage.AI**.
* Vérifier l’indexation des données et tester quelques requêtes Elasticsearch.

**5. Automatisation complète avec AWS Lambda et EC2**

* **Déployer une fonction AWS Lambda** qui surveille S3 et déclenche l’indexation vers Elasticsearch.
* **Configurer une instance EC2** pour exécuter les tâches Mage.AI de façon autonome.
* Tester le déclenchement automatique du pipeline Mage.AI via **AWS EventBridge**.

**6. Visualisation des données (Power BI ou Streamlit)**

* Connecter **Power BI** ou **Streamlit** à Elasticsearch pour visualiser les données.
* Construire des dashboards pour exploiter les données collectées.
* Tester et valider l’ensemble de la chaîne de collecte → stockage → indexation → visualisation.

**Prochaines actions immédiates 🚀**

✅ **Écrire et tester le script d’upload sur S3** (étape 2 en cours). ➡️ **Créer et configurer le pipeline Mage.AI pour automatiser l’upload**.

Une fois validé, on passe à **l’intégration d’Elasticsearch** !

**Dis-moi si cette feuille de route est bien alignée avec ton objectif !** 😉