

Cahier des Charges - AutoOptimize Initiative

Réalisé par : Amine BENAZZOUZ

1.Introduction

1.1Objectif du Projet

Le projet AutoOptimize Initiative vise à optimiser les performances de vente dans le secteur automobile en utilisant des analyses en temps réel et des traitements batch pour fournir des informations exploitables aux vendeurs. Deux workflows distincts seront implémentés pour répondre aux besoins en temps réel et historiques.

1.2Technologies Utilisées

- Python
- Kafka (pour le streaming en temps réel)
- Spark (pour le traitement des données)
- Elasticsearch et Kibana (pour le stockage et la visualisation des données)
- ETL (Extract, Transform, Load)
- Docker (pour la gestion des conteneurs)
- Hadoop File System (HDFS) pour le Data Lake
- SQL Server (pour le stockage des données historiques)
- Power BI (pour la Business Intelligence)
- Etc...

2. Workflow en Temps Réel

2.1 Objectifs

- Analyser en temps réel les performances des vendeurs.
- Fournir des informations immédiates sur les vendeurs les plus performants, les transactions en cours et les revenus générés.
- Utiliser Python, Kafka, Spark, Elasticsearch et Kibana pour mettre en œuvre le workflow en temps réel.

2.2 Fonctionnalités Principales

- Streaming des données de vente en temps réel à l'aide de Kafka.
- Traitement des données en temps réel avec Spark pour l'analyse des performances.
- Stockage des données dans Elasticsearch pour une récupération rapide.
- Visualisation des résultats à l'aide de tableaux de bord interactifs dans Kibana.

3. Workflow de Traitement Batch

3.1 Objectifs

- Analyser les données historiques pour des insights approfondis sur les ventes et les emplacements géographiques.
- Investir de manière stratégique dans la région africaine en analysant les tendances historiques.
- Utiliser ETL, Docker, HDFS, SQL Server et Power BI pour implémenter le workflow de traitement batch.

3.2 Fonctionnalités Principales

- Extraction des données historiques à partir de sources variées.
- Transformation des données via un processus ETL pour l'analyse.
- Stockage des données dans un Data Lake basé sur HFS.
- Utilisation de Docker pour l'environnement de conteneurs.
- Chargement des données traitées dans SQL Server.
- Création de rapports et tableaux de bord interactifs dans Power BI.

4. Sécurité et Gouvernance des Données

4.1 Sécurité

- Mise en œuvre de l'authentification et de l'autorisation pour l'accès aux données.
- Chiffrement des données en transit et au repos.
- Gestion des clés de manière sécurisée.

4.2 Gouvernance des Données

- Respect des règles GDPR pour la protection des données personnelles.
- Documentation et suivi des modifications apportées aux données.
- Rôles définis avec des autorisations appropriées pour l'accès aux données.

5. Planification du Projet

5.1 Phases du Projet

- Conception et Planification
- Analyse de besoins

- Mise en Place de l'Environnement
- Développement du Workflow en Temps Réel
- Tests et Optimisation du Workflow en Temps Réel
- Développement du Workflow de Traitement Batch
- Tests et Optimisation du Workflow de Traitement Batch
- Mise en Production et Formation

5.2 Livrables Attendus

- Environnement de développement opérationnel.
- Workflow en temps réel fonctionnel avec des tableaux de bord Kibana.
- Workflow de traitement batch fonctionnel avec des rapports Power BI.
- Documentation complète du projet.

6. Suivi et Maintenance

6.1 Suivi

- Surveillance régulière des performances du système.
- Analyse continue des données pour les mises à jour futures.

6.2 Maintenance

- Correction des erreurs et bogues.
- Mises à jour pour s'aligner sur les nouvelles exigences.

7. Approbation du Cahier des Charges

Ce cahier des charges est approuvé par: EL OUTMADI Abderrahim