



Big Data - Analyse des obstacles

Responsable : **Andreani Quentin**

Ce document décrit les obstacles potentiels rencontrés dans l'architecture technique proposée pour notre solution Big Data. L'objectif est d'identifier et d'anticiper ces défis pour mieux y faire face et optimiser l'efficacité de la mise en œuvre de la solution.

1. Collecte des données

1.1 Fiabilité des capteurs IoT

Les capteurs IoT installés sur les véhicules pourraient rencontrer des problèmes techniques, notamment en termes de fiabilité et de précision des données collectées. Les défauts de capteurs peuvent entraîner des données incorrectes ou manquantes, ce qui pourrait affecter l'exactitude de notre analyse.

1.2 Connectivité IoT

Il peut y avoir des défis liés à la transmission des données en temps réel à un serveur central. Des problèmes tels que la connectivité intermittente, les délais ou les problèmes de bande passante peuvent entraver la collecte des données.

Pour plus d'information concernant la partie IOT veuillez vous référer au document de spécification technique de l'équipe en charge de l'IOT.

2. Stockage des données

2.1 Coût du stockage Cloud

Le coût du stockage des données sur Google Cloud Platform peut être important, surtout si le volume de données collectées augmente rapidement. Il faudra mettre en place une stratégie de gestion des coûts afin d'éviter des dépenses excessives.

2.2 Performance de la base de données

L'efficacité de Google BigQuery à traiter des pétaoctets de données est prouvée, mais la performance pourrait diminuer si le volume de données est extrêmement élevé ou si des requêtes complexes sont exécutées simultanément. Pour plus d'information concernant la partie IA veuillez vous référer au document de spécification technique de l'équipe en charge de l'IA.

3. Traitement des données

3.1 Capacité de traitement en temps réel

Même si Google Dataflow est capable de traiter des flux de données en temps réel, des retards peuvent survenir si le volume des données entrantes est très élevé.

3.2 Complexité du Machine Learning

L'élaboration de modèles d'apprentissage automatique précis et efficaces pour analyser les données et faire des prédictions peut être un défi. Cela nécessite des compétences spécialisées et une compréhension approfondie des données.

4. Visualisation des données

4.1 Gestion de la surcharge d'informations

La visualisation efficace de grandes quantités de données peut s'avérer difficile. Il est important de créer des tableaux de bord qui présentent les informations de manière claire et compréhensible, sans surcharger l'utilisateur avec trop de détails.

5. Sécurité des données

5.1 Gestion de l'encryption

Bien que l'encryption de bout en bout soit essentielle pour la sécurité, elle peut aussi ralentir le traitement des données. Il faudra trouver un équilibre entre la sécurité des données et la performance du système.

5.2 Gestion des accès

La mise en place d'une gestion des droits d'accès efficace et sûre peut s'avérer complexe, en particulier si de nombreuses personnes ont besoin d'accéder aux données à des fins diverses.

5.3 Respect des réglementations sur la confidentialité des données

Il sera crucial de veiller à ce que toutes les procédures de collecte, de stockage et de traitement des données respectent les réglementations sur la confidentialité des données, telles que le RGPD (Règlement général sur la protection des données) en Europe.