

Analyse des données IoT avec Hadoop

1. Introduction

Ce système d'information a pour objectif d'analyser les données collectées par les capteurs IoT afin d'extraire des corrélations significatives entre les différentes métriques environnementales (température, humidité, pollution, etc.). Le système utilise une infrastructure Hadoop, reposant sur une architecture maître-esclave, pour gérer efficacement de grandes quantités de données et fournir des analyses précises.

2. Architecture du système

Le système s'appuie sur les composantes principales suivantes :

1. **HDFS (Hadoop Distributed File System) :**
 - Gère le stockage distribué et tolérant aux pannes des données collectées.
 - Organise les fichiers dans des DataNodes, sous la supervision d'un NameNode (le maître).
2. **MapReduce :**
 - Permet le traitement distribué des données en deux phases principales : **map** (partitionner et organiser les données) et **reduce** (agréger et analyser les données).

Schéma général :

- **Entrée :** Données IoT contenant des métriques comme température, humidité, PM2.5, niveau sonore, et usage d'eau.
 - **Traitement :** Des jobs MapReduce identifient les corrélations entre métriques par localisation.
 - **Sortie :** Résultats exportés sous forme de rapports ou fichiers analytiques.
-

3. Fonctionnalités clés

1. Collecte et stockage des données :

- Les capteurs IoT envoient périodiquement des données.
- Les données sont enregistrées dans HDFS via des fichiers de type CSV.

2. Analyse des données :

- Utilisation de **SensorMapper** et **SensorReducer** pour exécuter les tâches MapReduce :
 - **SensorMapper** organise les métriques par localisation.
 - **SensorReducer** calcule des corrélations entre plusieurs métriques environnementales.

3. Export et visualisation des résultats :

- Export des résultats sous format CSV pour intégration dans d'autres outils.
- Compatibilité avec des outils comme Tableau ou Power BI pour la création de tableaux de bord interactifs.

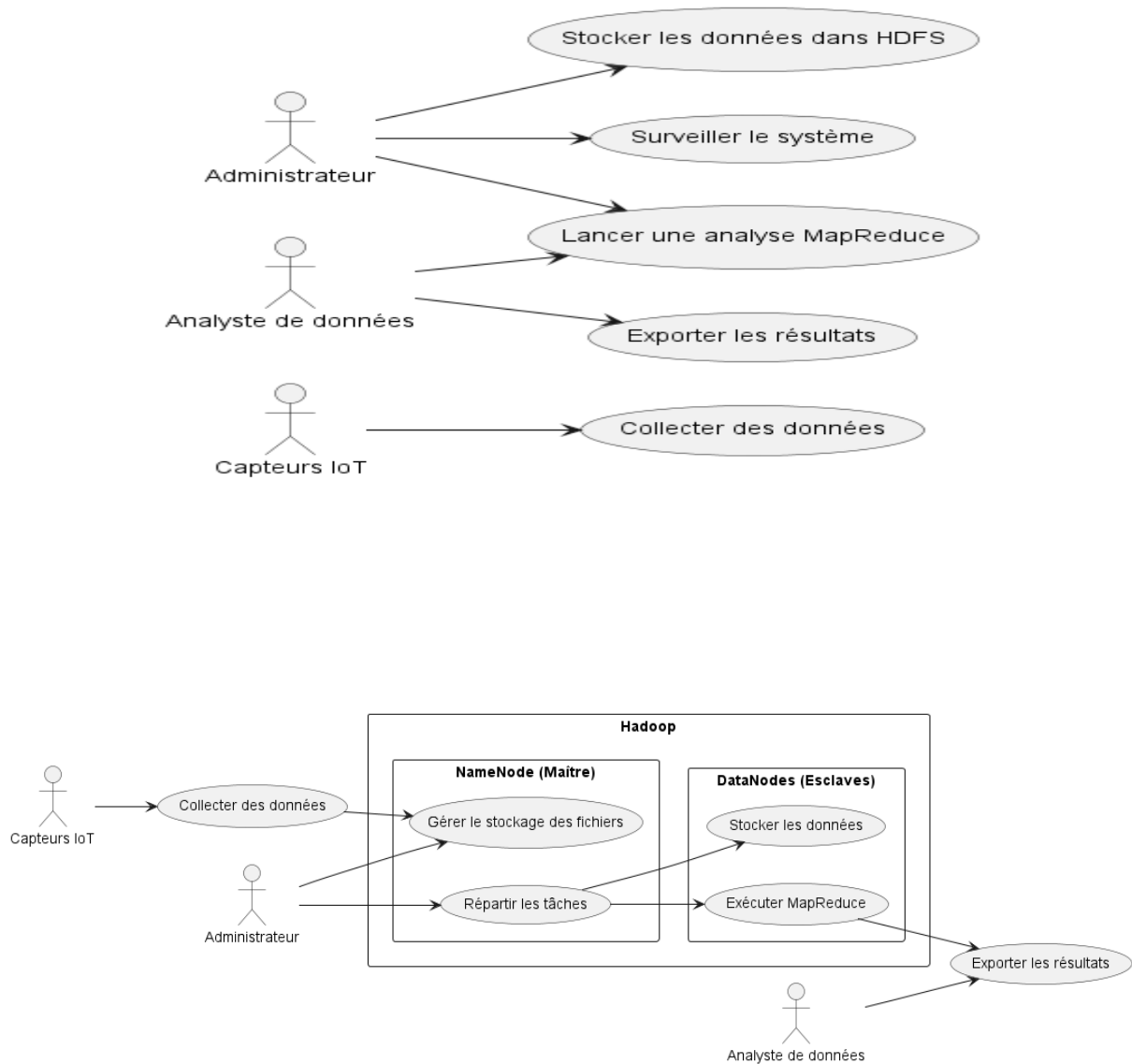
4. Surveillance du système :

- Surveillance en temps réel des nœuds NameNode et DataNodes via les outils natifs de Hadoop.
- Gestion automatique des pannes et réaffectation des tâches en cas de défaillance d'un nœud.

4. Diagramme de cas d'utilisation (Use Case)

Le diagramme UML illustre les interactions entre les acteurs principaux et les fonctionnalités du système.

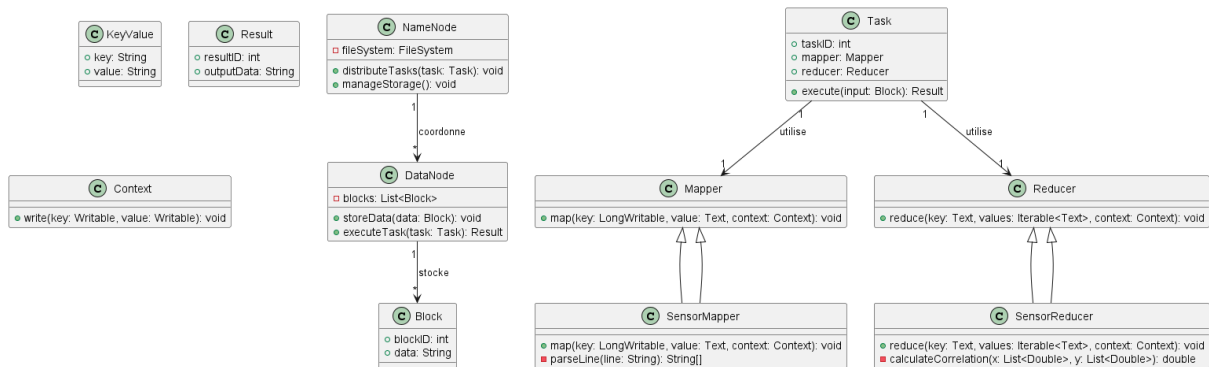
- **Acteurs** : Administrateur système, Analyste IoT, Capteur IoT.
- **Cas d'utilisation** :
 - Collecter des données IoT.
 - Lancer une analyse MapReduce.
 - Exporter les résultats.
 - Surveiller l'état du système.



5. Diagramme de classes

Le diagramme de classes structure les éléments principaux du système :

1. **Capteur :**
 - Représente les capteurs IoT et leurs caractéristiques (emplacement, type de métrique).
2. **FichierDonnees :**
 - Modélise les fichiers de données enregistrés dans HDFS.
3. **SensorMapper et SensorReducer :**
 - Implémentent les phases MapReduce (mapper et reducer).
 - Incluent des fonctions spécifiques comme `map()` et `reduce()`.
4. **ResultatAnalyse :**
 - Contient les données d'analyse après le traitement, avec des fonctions pour export et visualisation.



Avantages spécifiques du calcul de la corrélation

- **Simplicité mathématique :** La corrélation de Pearson est facile à calculer et fournit un seul nombre (-1 à 1) qui résume la relation entre deux variables.
- **Intuitif :**
 - Une **corrélation positive** (proche de 1) : Lorsque la température augmente, l'humidité augmente aussi.
 - Une **corrélation négative** (proche de -1) : Lorsque la température augmente, l'humidité diminue.
 - Une **corrélation proche de 0** : Les deux variables ne sont pas liées.

a) Qualité de l'air et santé

- **Confort thermique :** Une température élevée combinée à une forte humidité peut rendre l'environnement inconfortable (indice de chaleur élevé).
- **Risque de maladies :** Certaines conditions, comme l'humidité élevée dans un environnement chaud, favorisent la propagation de moisissures et de maladies respiratoires.

- **Prédictions climatiques** : Les données de corrélation peuvent être utilisées pour prédire des épisodes d'inconfort thermique.

b) Agriculture

- **Impact sur les cultures** : Les plantes sont sensibles à la température et à l'humidité. Une forte corrélation pourrait indiquer des conditions optimales ou critiques pour la croissance.
- **Gestion des serres** : Les corrélations aident à ajuster les conditions de température et d'humidité dans des environnements contrôlés.

c) Énergie et climatisation

- Les systèmes de climatisation ou de chauffage peuvent être optimisés en comprenant comment la température et l'humidité varient ensemble.
- Cela permet de réduire les coûts d'énergie tout en maintenant un environnement confortable.

d) Modélisation météorologique

- **Prédiction des précipitations** : Les relations entre température et humidité peuvent aider à prévoir la probabilité de pluie ou de formation de brouillard.
- **Alertes météorologiques** : Comprendre la corrélation permet de détecter des anomalies (par exemple, sécheresses, vagues de chaleur).

6. Conclusion

Ce système d'information s'appuie sur une infrastructure robuste et scalable grâce à Hadoop. Il permet de traiter efficacement des données massives issues des capteurs IoT et d'extraire des corrélations significatives. Ces analyses sont précieuses pour améliorer la prise de décisions dans des domaines tels que la gestion environnementale, l'urbanisme ou l'optimisation des ressources.

Ce projet peut facilement être étendu avec des fonctionnalités avancées, comme l'analyse en temps réel ou l'intégration de modèles de machine learning pour des prévisions précises.