**Bike Sharing Demand**

**Description du projet**

Les systèmes de partage de vélos sont un moyen de louer des vélos dans lesquels le processus d'obtention de l'adhésion, de la location et du retour des vélos est automatisé via un réseau de points de vente. Ces systèmes fonctionnent donc comme un réseau de capteurs, qui peut être utilisé pour étudier la mobilité en ville.

**Missions** :

* Analyse et rédaction des spécifications techniques et fonctionnelles détaillées.
* Collection, préparation et stockage des données.
* Récupération du Dataset et les fichiers CSV contenant les informations relatives aux locations horaires de vélos.
* Préparation des données et vérification de non absence et de la qualité des informations.
* Déploiement, configuration et administration du cluster Hadoop.
* Transfert et stockage des fichiers dans HDFS (Hadoop Distributed File System).
* Création des Scripts pour le chargement des données à partir des fichiers CSV dans des tables ORC du Data Warehouse Hive.
* Interrogation et requêtage du Data Warehouse.
* Traitement et analyse exploratoire des données.
* Définition des jeux de données training et test.
* Définition des hypothèses pouvant influencer la demande de location de vélos.
* Exploration et inspection de l’ensemble de données.
* Développement des modèles de classification et de prédiction pour prévoir la demande de location de vélos en combinant les habitudes d’utilisation historiques avec les données météorologiques.
* Entrainement des modèles et mesure de performance.
* Sauvegarde des résultats de traitement et de prédiction sous forme de fichiers CSV.
* Mise en place d’une architecture de streaming Kafka.
* Configuration des serveurs Zookeeper, Kafka et création du topic.
* Développement du programme Producer pour la récupération des données concernant chaque station de vélos et le stockage dans des messages Kafka.
* Développement du programme Consumer pour la sauvegarde et l’affichage de l’état de différentes stations.
* Alimentation en temps réel de la base de données NoSQL MongoDB via les messages Kafka Stream en utilisant un client pour l’insertion des documents sous format JSON.
* Reporting.
* Suivi des résultats de traitement et les rapports de variation de la demande de location de vélos pour chaque heure.
* Visualisation et suivi de la disponibilité et l’état de chaque station.
* Rédaction des rapports et de la documentation technique.
* Démonstrations, rétrospectives et planifications (Framework Scrum).
* Supervision de la plateforme technique (POC).
* Veille technologique.

Environnement technique : Cloudera, Hortonworks, Google Cloud Platform, Ecosystème Hadoop, Spark, Python, Scala, R, Zookeeper, Kafka, MongoDB, QlikView, Tableau, Scrum Agile, Diagramme de Gantt.