

Projet 8 : Analyse de données financières en temps réel

Mots clés

Finance, analyse temps réel, streaming, visualisation de données, systèmes distribués, aide à la décision

Contexte

Le secteur boursier est depuis plusieurs années, centré sur les problématiques big-data : des volumes de données de plus en plus importants, et la nécessité de traiter l'information avec des temps de latences de plus en plus faibles. De ce fait, il est devenu impératif de développer des systèmes résilients, extrêmement performants et permettant à des humains d'exploiter la donnée, facilitant ainsi la prise de décision, décision pouvant mettre en jeu des montants financiers considérables.

Objectifs

Les étudiants devront réaliser un système temps réel d'analyse de données financières. Ils devront choisir les sources de données adéquates, construire l'intégralité de l'ETL, définir et calculer un certain nombre de métriques d'intérêt, déployer les infrastructures nécessaires pour le processing et le stockage, et proposer une solution de visualisation appropriée. Ils pourront, de manière conjointe au calcul de données statistiques, intégrer un module d'analyse de sentiments, en se basant sur les articles de presse spécialisés.

Le projet devra être réalisé en utilisant les outils traditionnellement employés dans les environnements big-data (cluster kafka/spark, stockage distribué, streaming...), pour rendre leur système scalable, résilient et performant. L'infrastructure pourra être déployée sur une solution cloud ou bien sur une solution on-premise.

Travail demandé

- Définir les sources de données boursières, les plus adaptées à du traitement en temps réels
- Définir des grandeurs statistiques utiles à l'aide à la prise de décision dans le domaine boursier
- Définir l'architecture logiciel du système de bout en bout (de l'acquisition de la donnée à la visualisation)
- Mettre en place une infrastructure de calcul distribuée (cluster spark)
- Développer la chaîne de traitement de la donnée (mise en place du schema registry, calcul des kpi en temps réel, storage)
- Sélectionner et mettre en place un système de stockage approprié (rdbms, nosql ou stockage distribué)
- Développer un système de visualisation type dashboard financier, mettant en avant les kpi calculés
- Intégrer un module d'analyse de sentiment basé sur les données issues de la presse financière.

- Mettre en place un pipeline de ci/cd
- Livrer un outil facilement déployable

Logiciels et langages utilisés

- Logiciels : Spark, Kafka, Avro, Docker, framework js (vue/angular/react), Git
- Stockage : Rdms, nosql ou hdfs
- Langages : Scala, js