

TP – Pratique containers Dockers/Kafka dans la mise en place d'une chaine de streaming

Contexte : Équipe de 3 au maximum

OBJECTIF

- Utiliser Docker et les containers dockers
- Utiliser un framework de traitement comme spark streaming (installation de l'outil seulement)
- Utiliser un framework d'ingestion de données

On désire mettre en place une configuration de streaming afin de traiter des données reçues en continu

On utilisera le schéma suivant :



Dans un premier temps, on se focalise sur le développement de la chaine avec les besoins suivants.

-Génération de données en continu. On peut utiliser des données générées aléatoirement ou les données de capteurs disponibles sur :

https://drive.google.com/drive/folders/1N6wUKpmaUSFH_3DFS8Coxd4d1ArvibL?usp=drive_link

-Ingestion de ces données en utilisant un outil tel que Kafka.

-Envoi des données ingérées vers un framework de traitement comme spark streaming. On ne développera pas de traitement dans cette partie. On devra par contre relier cette partie avec la partie ingestion.

Proposer les outils à utiliser dans votre chaine. Au minimum, on devra avoir :

- a. Générateur de données
- b. Gestionnaire d'ingestion
- c. Outil de traitement

Les références indiquées dans la section Références seront privilégiées

Une fois votre schéma défini, développer une approche basée sur les containers docker afin de mettre en place cette chaine.

Références

1. <https://thingsolver.com/streaming-analytics-in-banking-how-to-start-with-apache-flink-and-kafka-in-7-steps/>
2. https://ci.apache.org/projects/flink/flink-docs-release-1.1/quickstart/run_example_quickstart.html
3. <https://hands-on-tech.github.io/2018/11/01/kafka-spark-flink-example.html>

Contenu du ppt de présentation: 12 slides max

-Introduction besoin

-Caractérisation des sources et des données

-Architecture des pipelines

-Résultats obtenus

-Problèmes rencontrés

-Compétences acquises

-Conclusion

Une période de questions de 5 min est intégrée dans la présentation de 15 minutes.

