



## DATA STREAMING - REALISER UN PIPELINE DE TRAITEMENT DES DONNEES EN TEMPS REEL AVEC KAFKA ET KAFKA STREAMS EN MODE IAAS

Le traitement en temps réel de la donnée est une tendance du Big Data et de l'entreprise Data Driven. Traditionnellement, le stockage et le traitement des données se faisait par « batch » induisant un délai entre la production des données et l'utilisation effective de ces données. Le data streaming permet de réduire ce délai permettant ainsi aux applications de répondre plus vite à des événements émis par le système d'information ou arrivant de l'extérieur. Dans certains cas, comme l'IoT par exemple, il n'est pas possible de stocker l'intégralité des données à la source ; il est donc nécessaire d'avoir des solutions permettant le traitement et l'agrégation en temps réel.

### Objectif du stage

L'objectif du stage est de réaliser un pipeline de traitement de flux de données en temps réel dans une démarche complète, allant de l'ingestion des données jusqu'à la visualisation des résultats des traitements en temps réel sur un tableau de bord afin d'obtenir une vision détaillée des capacités et des limites de Kafka et Kafka Streams. L'ensemble des briques techniques devront être installées chez un ou plusieurs providers cloud en mode IaaS.

### Travaux à réaliser

Intégré pour la durée de votre stage à la Taskforce Data, vous serez amené, sous la supervision d'un consultant, à réaliser les travaux suivants :

- Rédiger une note de cadrage du sujet de stage, comprenant :
  - Votre compréhension du stage et ses objectifs
  - Un planning provisionnel et ses jalons clés
- Définir les critères d'évaluation des solutions
- Identifier les briques techniques de l'architecture data streaming nécessaires à la réalisation du pipeline
- Mettre en place les outils de l'architecture sur une infrastructure cloud (IaaS)
- Définir et implémenter les algorithmes de traitement de la donnée en temps réel
- Réaliser des tests de performance pour challenger les capacités des solutions
- Identifier les possibilités et les limites des fonctionnalités de stream processing de Kafka Streams
- Enrichir le séminaire de data streaming de Nexworld

En parallèle, vous participerez à des missions de conseil, en intégrant une équipe de consultants Nexworld, placée sous la responsabilité d'un directeur de mission.

### Outils et technologies (non exhaustif)

Intégré pour la durée de votre stage à la Taskforce Data, vous serez amené, sous la supervision d'un consultant, à réaliser les travaux suivants :

- Kafka, Kafka Streams, Kafka Connect
- Java, Python