Apache Kafka : streaming de données

Référence : **BDDAPA002** Durée : **3 jours (21 heures)** Certification : **Aucune**

CONNAISSANCES PRÉALABLE

- Avoir des notions de messagerie

PROFIL DES STAGIAIRES

- Développeurs, data engineers

OBJECTIFS

- Déployer et configurer Kafka
- Produire et consommer des messages
- Gérer les topics, partitions et réplicas
- Assurer la fiabilité et la résilience des flux

CERTIFICATION PRÉPARÉE

- Aucune

MÉTHODES PÉDAGOGIQUES

- Mise à disposition d'un poste de travail par stagiaire
- Remise d'une documentation pédagogique numérique pendant le stage
- La formation est constituée d'apports théoriques interactif, démonstrations en direct, travaux dirigés, ateliers collaboratif, mises en situation professionnelles, projet fil rouge, études de cas réels, d'ateliers d'optimisation et de résilience, d'ateliers de supervision et de diagnostic
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émargement par demi-journée de formation

FORMATEUR

- Consultant-formateur expert Kafka

MÉTHODES D'ÉVALUATION DES ACQUIS

- Validation des prérequis envoyée en amont de la formation. En l'absence de réponse d'un ou plusieurs participants, un temps sera consacré en ouverture de session pour prendre connaissance du positionnement de chaque stagiaire sur les objectifs pédagogiques évalués
- Validation progressive des compétences tout au long de la formation via exercices pratiques, quiz techniques, études de cas, travaux de configuration et d'administration, résolution d'incidents simulés et restitution des travaux réalisés dans le cadre du projet fil rouge
- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

ACCESSIBILITÉ DE LA FORMATION

- Si vous avez besoin d'un aménagement pour suivre votre formation dans de bonnes conditions, nous vous invitons à contacter notre référent handicap au 01 71 19 70 30 ou à l'adresse referent.handicap@edugroupe.com. Un échange en amont nous permettra d'étudier avec vous les adaptations possibles.

CONTENU DU COURS

1. JOUR 1 - Comprendre Kafka et mettre en œuvre les premiers flux (7h)

2. Comprendre l’architecture Apache Kafka (2h)

- Introduction au streaming de données
- Architecture orientée événements
- Cas d’usage Kafka
- Brokers
- Topics
- Partitions
- Producteurs et consommateurs
- ZooKeeper et architecture KRaft
- Écosystème Kafka
- 💡 *Mises en pratique : Découverte d’une plateforme Kafka ; Analyse de l’architecture d’un cluster ; Identification des composants et de leurs interactions*

3. Installer et configurer Apache Kafka (2h)

- Installation de Kafka
- Configuration des brokers
- Gestion des ports et services
- Architecture mono-nœud et cluster
- Paramètres fondamentaux
- Gestion du stockage
- Outils d’administration
- 💡 *Mises en pratique : Installation d’un broker Kafka ; Configuration d’un environnement de test ; Vérification du bon fonctionnement du cluster*

4. Produire et consommer des messages (3h)

- Producteurs Kafka
- Consommateurs Kafka
- Sérialisation des données
- Formats JSON, Avro et Protobuf
- Gestion des offsets
- Groupes de consommateurs
- Bonnes pratiques de développement
- 💡 *Mises en pratique : Création de producteurs ; Développement de consommateurs ; Échange de messages temps réel ; Gestion des groupes de consommation*

5. JOUR 2 - Gestion des données et résilience des flux (7h)

6. Administrer les topics, partitions et réplicas (3h)

- Création et administration des topics
- Partitionnement
- Réplication
- Facteur de réplication
- Rééquilibrage
- Stratégies de distribution
- Gestion du stockage
- 💡 *Mises en pratique : Création de topics ; Configuration de partitions ; Mise en œuvre de la réplication ; Analyse des performances selon différentes configurations*

7. Assurer la fiabilité et la résilience des flux (2h)

- Acknowledgements
- Idempotence
- Exactly Once Semantics
- Gestion des pannes
- Tolérance aux fautes
- Réplication inter-brokers
- Stratégies de reprise
- 💡 *Mises en pratique : Simulation de pannes ; Analyse du comportement du cluster ; Configuration des mécanismes de fiabilité*

8. Superviser et administrer Kafka (2h)

- Outils d'administration Kafka
- Métriques de supervision
- Journaux d'exécution
- Monitoring
- Gestion des alertes
- Diagnostic des incidents
- 💡 *Mises en pratique : Analyse des métriques ; Identification de problèmes ; Mise en place d'indicateurs de supervision*

9. JOUR 3 - Streaming avancé et projet de synthèse (7h)

10. Intégrer Kafka dans une architecture de données (2h)

- Kafka Connect
- Connecteurs sources et cibles
- Intégration avec bases de données
- Intégration Cloud
- Cas d'usage Data Engineering
- Architecture événementielle
- 💡 *Mises en pratique : Mise en œuvre d'un connecteur ; Intégration avec une base de données ; Analyse d'un flux complet*

11. Découvrir Kafka Streams et les traitements temps réel (2h)

- Principes de Kafka Streams
- Traitement continu
- Agrégations
- Jointures
- Fenêtres temporelles
- Cas d'usage analytiques
- 💡 *Mises en pratique : Création d'un traitement de flux ; Transformation de données en temps réel ; Analyse des résultats*

12. Projet fil rouge : mise en œuvre d'une plateforme de streaming complète (3h)

- Déploiement d'un cluster Kafka
- Création des topics
- Production et consommation de messages
- Configuration de la réplication
- Mise en œuvre des mécanismes de fiabilité
- Supervision des flux
- Analyse des performances
- 💡 *Mises en pratique : Réalisation d'un projet complet de streaming de données ; Déploiement et administration d'une plateforme Kafka ; Gestion d'incidents simulés ; Restitution des travaux et débriefing collectif*