

Big Data - Infrastructure

Cassandra - Mise en oeuvre et utilisation

3 jours (21h00) | ★★★★★ 4/5 | NOSQL-CAS | Évaluation qualitative de fin de stage |
Formation délivrée en présentiel ou distanciel ⁽¹⁾

Formations Informatique > Big Data > Big Data - Infrastructure



À l'issue de ce stage vous serez capable de :

- Connaître les apports de Cassandra
- L'installer et le configurer
- Maîtriser le CQL (Cassandra Query Language)
- Administrer et sécuriser un cluster Cassandra
- Effectuer des calculs distribués avec Spark.

Niveau requis

Avoir des connaissances générales sur les bases de données.

Public concerné

Chefs de projets et gestionnaires de bases de données.

Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence émarginée par demi-journée par les stagiaires et le formateur.

(1) Modalité et moyens pédagogique :

Formation délivrée en présentiel ou distanciel * (e-learning, classe virtuelle, présentiel à distance). Le formateur alterne entre méthodes ** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation). La validation des acquis peut se faire via des études de cas, des quiz et/ou une certification.

Les moyens pédagogiques mis en oeuvre (variables suivant les formations) sont : ordinateurs Mac ou PC (sauf pour les cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel). Environnements de formation installés sur les postes de travail ou en ligne. Supports de cours et exercices.

* Nous consulter pour la faisabilité en distanciel. ** Ratio variable selon le cours suivi.

Programme

Introduction

- Historique
- Fonctionnalités de Cassandra
- Licence
- Format des données
- "Key-value"
- Traitement de volumes importants
- Haute disponibilité
- Système réparti de base de données...

Installation et configuration

- Prérequis
- Plateformes supportées
- Etude du fichier de configuration : conf/cassandra.yaml
- Répertoire de travail et de stockage des données
- Gestion de la mémoire
- Démarrage d'un noeud et test de l'interface cliente cqlsh

CQL (Cassandra Query Language)

- Commandes de base : connexion au système de base de données
- Création de colonnes
- Insertion
- Modification
- Recherche
- Le CQL
- Exécution de scripts
- Comment écrire des requêtes ?
- Approches
- Limitations du CQL

Gestion de la grappe

- Principe et préparation du premier noeud : adresse d'écoute
- Configuration de nouveaux noeuds
- Notion de "Bootstrapping" et "Token"
- Paramètres listen_address et rpc_address
- Réplication
- Topologie du réseau et Endpoint Snitch
- Stratégie de réplication
- Ajout de noeuds et suppression
- Architecture de stockage mémoire et disque dur
- Gestion des tombstones
- Bloom-filter

Exemple de travaux pratiques (à titre indicatif)

- Mise en place d'une configuration de production (multi-datacenters, multi-racks)

Supervision

- OpsCenter
 - Installation
 - Lancement
 - Découverte
- Supervision avec nodetool cfstats
- Principe des accès JMX
- Exports JMX vers des outils de supervision comme Nagios
- Démonstration avec Prometheus et Kibana

Exploitation

- Gestion des noeuds Cassandra
- Sauvegardes, snapshots et export au format JSON
- Principe de cohérence
- "Digest request" et "read repair"

Support Hadoop

- Principe de Map/Reduce
- Implémentation Hadoop

et intégration Hadoop/Cassandra

Support Spark

- Description rapide de l'architecture Spark
- Mise en oeuvre depuis Cassandra

- Exécution de travaux Spark s'appuyant sur une grappe Cassandra

Performance

- Outil de test de performance Cassandra-stress

- Mise en place d'un plan de stress et paramétrage

Modalités d'évaluation des acquis

L'évaluation des acquis se fait :

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation ou une certification (M2i ou éditeur)