

ECOSYSTÈME HADOOP & DISTRIBUTIONS

Cassandra - Big Data

REF : SIBD004

DUREE : 21h

Présentiel Classe virtuelle

PUBLIC

Cette formation s'adresse à des Chefs de projet, gestionnaires de bases de données.

Modalités et délais d'accès : les inscriptions sont fermées 24h avant la 1ère journée de formation.

Accessibilité : Si vous avez des contraintes particulières liées à une situation de handicap, veuillez nous contacter au préalable afin que nous puissions, dans la mesure du possible, adapter l'action de formation.

PREREQUIS

Connaissances générales sur les bases de données.

MODALITES PEDAGOGIQUES

1 poste et 1 support par stagiaire

8 à 10 stagiaires par salle

Remise d'une documentation pédagogique papier ou numérique pendant le stage

La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience

MODALITES D'EVALUATION

Auto-évaluation des acquis par le stagiaire via un questionnaire en ligne

Attestation de fin de stage remise au stagiaire

OBJECTIFS PEDAGOGIQUES

A l'issue de cette formation vous serez capable de :

- Connaître les apports de Cassandra
- Savoir l'installer et le configurer
- Maîtriser le CQL
- Administrer et sécuriser un cluster Cassandra
- Effectuer des calculs distribués avec Spark.

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

CQL

- Commandes de base : connexion au système de base de données, création de colonnes, insertion, modification recherche,
- Le CQL : Cassandra Query Language. Exécution de scripts.
- Comment écrire des requêtes? Approches.
- Limitations du CQL.

Gestion de la grappe

- Principe. Préparation du premier noeud : adresse d'écoute.
- Configuration de nouveaux noeuds. Notion de bootstrapping et de token.
- Paramètres listen_address et rpc_address.
- Réplication : topologie du réseau et EndpointSnitch. Stratégie de réplication. Ajout de noeuds, suppression.
- Architecture de stockage mémoire et disque dur, gestion des tombstones, bloom-filter
- Travaux pratique : Mise en place d'une configuration de production (multi-datacenters, multi-racks)

Supervision

- OpsCenter : installation, lancement et 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.
- Execution d'application 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

Version du : 29/05/2020