

# Big Data avec Cassandra

Référence : PYCB010

Durée : 3 jours

Certification : Aucune

## CONNAISSANCES PREALABLES

- Connaissances générales sur les bases de données.

## PROFIL DES STAGIAIRES

- Chefs de projet. • Gestionnaires de bases de données.

## OBJECTIFS

- Connaître les apports de Cassandra, savoir l'installer et le configurer.

## CERTIFICATION PREPAREE

Aucune

## METHODES PEDAGOGIQUES

- Mise à disposition d'un poste de travail par stagiaire
- 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
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émargement

## FORMATEUR

Consultant-Formateur expert Bigdata

## METHODE D'EVALUATION DES ACQUIS

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation de fin de stage adressée avec la facture

## CONTENU DU COURS

### 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

### 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
- Cassandra dans un cloud. Mise en oeuvre avec OpenStack

### Supervision

- OpsCenter : installation, lancement. Utilisation de base
- Supervision avec nodetool cfstats, ou export JMX vers des outils de supervision comme Nagios

### Exploitation

- Sauvegardes. Import/export au format JSON

### **Support Hadoop**

- Principe de MapReduce. Implémentation Hadoop.  
Mise en oeuvre depuis Cassandra

### **Support Spark**

- Description rapide de l'architecture spark. Mise en oeuvre depuis Cassandra
- Execution de travaux Spark s'appuyant sur une grappe Cassandra