

BIGDATA AVEC CASSANDRA

CODE STAGE : BD010

OBJECTIFS

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

DURÉE

3 jours

PUBLIC

Chefs de projet, gestionnaires de bases de données.

PRÉ-REQUIS

Connaissances générales sur les bases de données.

PROGRAMME

Module 1 : Introduction

- Historique, fonctionnalités de Cassandra, licence
- Format des données, »key-value », traitement de volumes importants, haute disponibilité, système réparti de base de données, ...

Module 2 : Installation et configuration

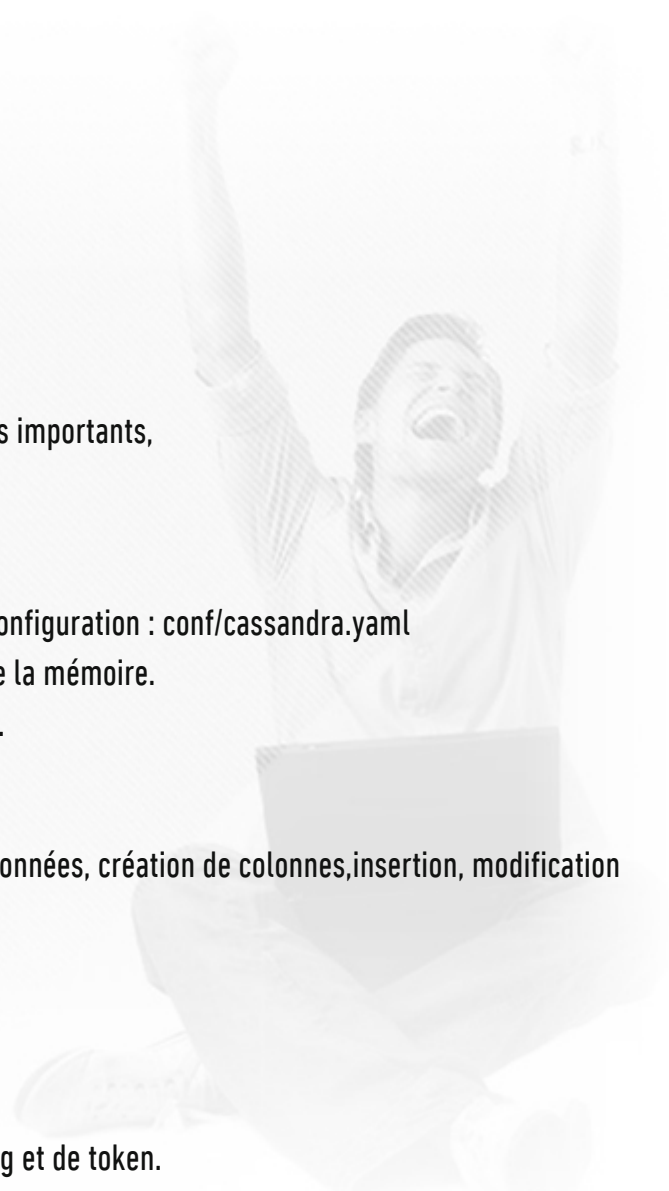
- Prérequis. Plate-formes supportées. Etude du fichier de configuration : conf/cassandra.yaml
- Répertoire de travail, de stockage des données, gestion de la mémoire.
- Démarrage d'un noeud et test de l'interface cassandra-cli.

Module 3 : 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.

Module 4 : Clusters

- Principe.Préparation du premier nœud : 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.

Module 5 : Exploitation

- Sauvegardes. Import/export au format JSON. Supervision avec `nodetool cfstats`, ou export JMX vers des outils de supervision comme Nagios.

Module 6 : Support Hadoop

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

