

## Hadoop : stockage avec HBase

Mettre en oeuvre le stockage Hadoop avec Hbase

### Détails

- Code : DB-HDS
- Durée : 2 jours ( 14 heures )

#### Public

- Chefs de projets
- Administrateurs
- Toute personne souhaitant stocker des données avec Hbase

#### Pré-requis

- Connaissance des principes de base Hadoop et des bases de données

#### Objectifs

- Comprendre le fonctionnement de HBase
- Savoir mettre en place une configuration distribuée

### Programme

#### Introduction

- Rappels rapides sur l'écosystème Hadoop
- Les fonctionnalités du framework Hadoop
- Le projet et les modules : Hadoop Common, HDFS, YARN, Spark, MapReduce
- Présentation HBase
- Historique
- Lien avec HDFS
- Format des données dans HBase
- Définitions : table, région, ligne, famille de colonnes, cellules, espace de nommage, ...
- Fonctionnalités : failover automatique, sharding, requêtage

#### Architecture

- HBase master node, Region Master, liens avec les clients HBase
- Haute disponibilité
- Consistance des données
- Présentation du rôle de Zookeeper

#### Installation

- Choix des packages
- Installation en mode distribué
- Mise en oeuvre avec HDFS dans un environnement distribué
- Test de connexion avec hbase shell
- Travaux pratiques : Interrogations depuis le serveur http intégré

#### HBase utilisation : shell

- Présentation des différentes interfaces disponibles
- Travaux pratiques avec hbase shell

- Commandes de base, syntaxe, variables, manipulation des données : create, list, put, scan, get
- Manipulation des tables : disable (enable), drop, ...
- Programmation de scripts
- Gestion des tables : principe des filtres
- Mise en oeuvre de filtres de recherche, paramètres des tables
- Présentation des espaces de nommage

#### Cluster HBase

- Fonctionnement en mode distribué
- Fonctionnement indépendant des démons (HMaster, HRegionServer, Zookeeper)
- Gestion de la consistance
- Mise en évidence
- Outils d'exploitation : hbck, hfile, ...
- Travaux pratiques : mise en oeuvre des splits sur un exemple de tables réparties. regionsplitter

#### Performances

- Travaux pratiques avec hbase pe et hbase ltt
- Mesures : mise en place de prométhéus

#### Programmation

- Introduction, les APIs (REST, Avro, Thrift, Java, Ruby, Spark...)
- En exploitation : principe des accès JMX
- Supervision automatisée
- Démonstration avec Prométhéus
- Requêtage : exemple de programmation d'un client Java
- Gestion des tables
- Travaux pratiques : requêtage avec Spark

### Modalité

- Stage pratique en présentiel
- Stage pratique en distanciel

- Nombre de stagiaires minimum : 4
- Nombre de stagiaires maximum : 10

### Méthodes pédagogiques

- Exposés
- Cas pratiques

### Profils des intervenants

- Toutes nos formations sont animées par des consultants-formateurs expérimentés et reconnus par leurs pairs.

### Modalités d'évaluation

- Evaluation des acquis de la formation par le biais de cas pratiques et/ou mises en situation.
- Attestation de formation remise à chaque participant.

### Démarche qualité

- Questionnaire d'évaluation de satisfaction à chaud complété par chaque participant à l'issue de la formation.

### Moyens pédagogiques

- Salle équipée de PC (1 poste par stagiaire), vidéo-projecteur.
- Espace de pause.

Dernière mise à jour le 19/01/2022