

Hadoop : stockage avec HBase

Détails

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

Public

- Administrateurs
- Chefs de projets
- 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, interface avec des jobs MapReduce

Architecture

- HBase master node, Region Master, liens avec les clients HBase
- Présentation du rôle de Zookeeper

Installation

- Choix des packages
- Installation et configuration dans le fichier conf/hbase-site.xml
- Installation en mode 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
 - désactiver une table ou l'effacer : disable (enable), drop, ...
- Programmation de scripts
- Gestion des tables : principe des filtres
- Mise en œuvre de filtres de recherche, paramètres des tables
- Présentation des espaces de nommage

Cluster HBase

- Fonctionnement en mode distribué
- Première étape : fonctionnement indépendant des démons (HMaster, HRegionServer, Zookeeper)
- Passage au mode distribué : mise en oeuvre avec HDFS dans un environnement distribué

Travaux pratiques sur un exemple de tables réparties

- Mise en œuvre des splits

Programmation

- Introduction, les APIs (REST, Avro, Thrift, Java, Ruby, Spark...)
- Principe des accès JMX
- Démonstration avec Prométheus
- Exemple de programmation d'un client Java
- Gestion des tables
- Lien avec map/reduce

Travaux pratiques

- Lancement et écriture de programmes Spark

Modalités

- **Type d'action** :Acquisition des connaissances
- **Moyens de la formation** :Formation présentielle – 1 poste par stagiaire – 1 vidéo projecteur – Support de cours fourni à chaque

stagiaire

- **Modalités pédagogiques** : Exposés – Cas pratiques – Synthèse
- **Validation** : Exercices de validation – Attestation de stages