

Hadoop - Stockage avec HBase

Introduction

- Rappels rapides sur l'écosystème Hadoop
- Fonctionnalités du framework Hadoop
- Le projet et les modules
 - Hadoop Common
 - HDFS
 - Yarn
 - Spark
 - MapReduce
- Architecture
- Présentation HBase
- Historique
- Lien avec HDFS
- Format des données
- 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
- Rôle de ZooKeeper

Installation

- Choix des packages
- Installation et configuration dans le fichier conf/hbase-site.xml
- Démarrage en mode Standalone start-hbase
- Test de connexion avec HBase Shell
- Installation en mode distribué
- 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 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
- Mise en oeuvre avec HDFS dans un environnement distribué
- Tables réparties : mise en oeuvre des splits

Programmation

- Introduction
- Les APIs
 - REST
 - Avro
 - Thrift
 - Java
 - Ruby...
- Utilisation d'un client Java
- Gestion des tables
- Lien avec MapReduce
- Principes des accès JMX
- Exemple d'un client JMX