

HADOOP HORTONWORKS : ADMINISTRATION AVEC AMBARI

CODE STAGE : BD016

OBJECTIFS

Connaître les principes du framework Hadoop et savoir l'installer, le configurer et l'administrer avec Ambari (tableaux de bord, supervision, gestion des services, etc ...)

DURÉE

3 jours

PUBLIC

Chefs de projet, administrateurs et toute personne souhaitant mettre en oeuvre un système distribué avec Hadoop. Les travaux pratiques sont réalisés sur une distribution Hadoop Hortonworks.

PRÉ-REQUIS

Connaissance des commandes des systèmes unix/linux et des bases TCP/IP

PROGRAMME

1. Introduction Les fonctionnalités du framework Hadoop.

Les différentes versions.

Distributions : Apache, Cloudera, Hortonworks, EMR, MapR, DES.

Spécificités de chaque distribution.

Architecture et principe de fonctionnement.

Terminologie : NameNode, DataNode, ResourceManager, NodeManager.

Rôle des différents composants.

Le projet et les modules : Hadoop Common, HDFS, YARN, Spark, MapReduce

Oozie, Pig, Hive, Hbase, 

2. Les outils Hadoop

Infrastructure/Mise en uvre :

Avro, Ambari, Zookeeper, Pig, Tez, Oozie, Falcon, Pentaho

Vue d'ensemble



Gestion des données.

Exemple de sqoop.

Restitution : webhdfs, hive, Hawq, Mahout, ElasticSearch ..

Outils complémentaires:

Spark, SparkQL, SparkMLib, Storm, BigTop, Zebra

de développement : Cascading, Scalding, Flink/Pachyderm

d'analyse : Rhadoop, Hama, Chukwa, kafka

3. Installation et configuration

Trois modes d'installation : local, pseudo-distribué, distribué

Première installation.

Mise en œuvre avec un seul nœud Hadoop.

Configuration de l'environnement, étude des fichiers de configuration :

core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml et capacity-scheduler.xml

Création des users pour les daemons hdfs et yarn, droits d'accès sur les exécutables et répertoires.

Lancement des services.

Démarrage des composants : hdfs, hadoop-daemon, yarn-daemon, etc ..

Gestion de la grappe, différentes méthodes :

ligne de commandes, API Rest, serveur http intégré, APIS natives

Exemples en ligne de commandes avec hdfs, yarn, mapred

Présentation des fonctions offertes par le serveur http

Travaux pratiques :

Organisation et configuration d'une grappe hadoop

4. Administration Hadoop

Outils complémentaires à yarn et hdfs : jConsole, jconsole yarn

Exemples sur le suivi de charges, l'analyse des journaux.

Principe de gestion des nœuds, accès JMX.

Travaux pratiques :

mise en œuvre d'un client JMX

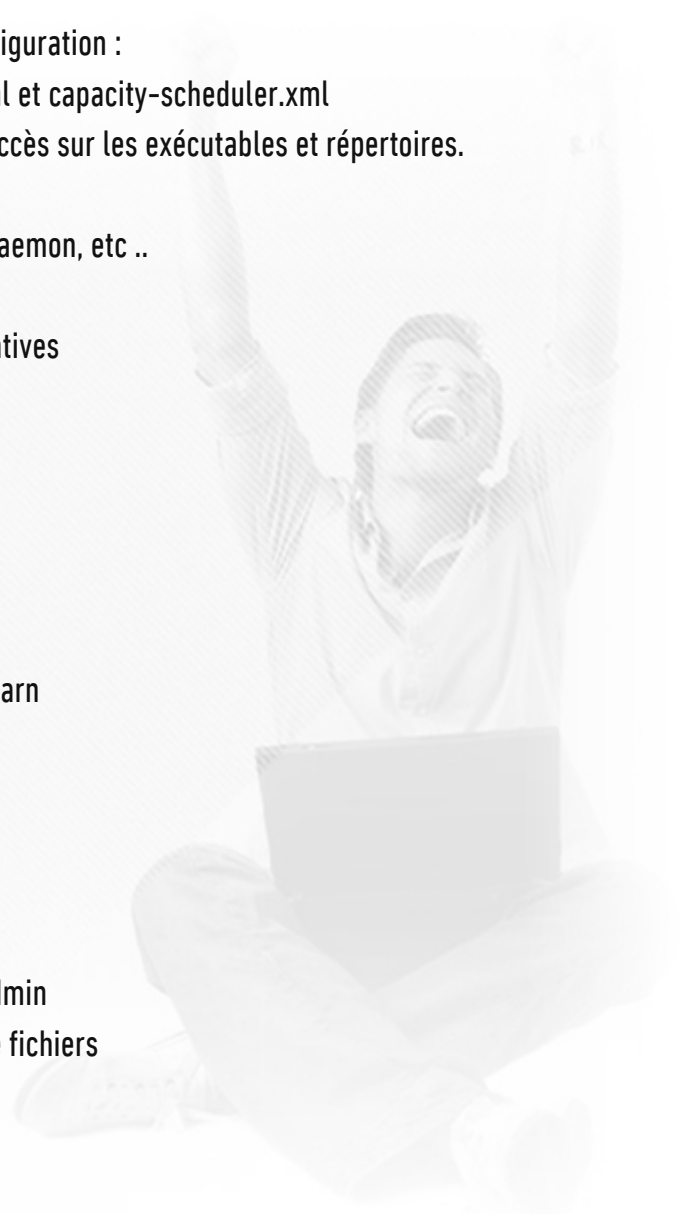
Administration HDFS :

présentation des outils de stockage des fichiers, fsck, dfsadmin

Mise en œuvre sur des exemples simples de récupération de fichiers

Gestion centralisée de caches avec Cacheadmin

Déplacement d'un NameNode. Mise en mode maintenance.



5. Haute disponibilité

Mise en place de la haute disponibilité sur une distribution Ambari.

Travaux pratiques :

Passage d'un système HDFS en mode HA

6. Sécurité

Mécanismes de sécurité et mise en œuvre pratique :

Activation de la sécurité avec Kerberos dans core-site.xml, et dans hdfs-site.xml pour les NameNode et DataNode.

Sécurisation de yarn avec la mise en œuvre d'un proxy et d'un Linux Container Executor.

Travaux pratiques :

Mise en place de la sécurité Kerberos sur une distribution Ambari. Création des utilisateurs. Travaux sur les droits d'accès et les droits d'exécution. Impact au niveau des files Yarn, Oozie et Tez.

7. Exploitation

Installation d'une grappe Hadoop avec Ambari. Tableau de bord. Lancement des services.

Principe de la supervision des éléments par le NodeManager.

Monitoring graphique avec Ambari.

Présentation de Ganglia, Kibana

Travaux pratiques :

Visualisation des alertes en cas d'indisponibilité d'un nœud.

Configuration des logs avec log4j.

