Formation Hadoop Cloudera

Code : SYS-HADOOP-100 Durée : 3 jours (21 heures)

Public visé

Cette formation Hadoop s'adresse aux chefs de projets, administrateurs systèmes et à toute personne qui gère ou maintient un système distribué avec Hadoop.

{{< formation-prerequis >}}

Objectifs pédagogiques

- Connaître les principes du framework Hadoop
- Savoir installer et configurer Hadoop
- Maîtriser la configuration et la gestion des services avec Cloudera Manager

Programme

Introduction à Hadoop Cloudera

- Les fonctionnalités du framework Hadoop
- Les différentes versions et distributions (Apache, Cloudera, Hortonworks, EMR, MapR, DSE)
- Architecture et principe de fonctionnement
- Terminologie : NameNode, DataNode, ResourceManager, NodeManager
- Le projet et ses modules (Hadoop Common, HDFS, YARN, Spark, MapReduce, Oozie, Pig, Hive, HBase)

Les outils Hadoop

- Infrastructure : Avro, Ambari, Zookeeper, Pig, Tez, Oozie
- Gestion des données et Sqoop
- Restitution : webhdfs, Hive, Hawq, Mahout, ElasticSearch
- Outils complémentaires : Spark, Storm, BigTop, Zebra, Cascading, Flink, RHadoop, Hama, Chukwa, Kafka

Installation et configuration

- Présentation de Cloudera Manager
- Installation en mode distribué
- Configuration des fichiers : core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml, capacity-scheduler.xml

- Création des users et gestion des droits d'accès
- Gestion de la grappe et exemples en ligne de commandes
- Travaux pratiques : organisation d'une grappe, traitement de données, requêtage SQL avec Impala

Administration d'Hadoop

- Outils : Jconsole, Jconsole Yarn
- Suivi de charges et analyse des journaux
- Gestion des noeuds et accès JMX
- Administration HDFS : présentation des outils de stockage des fichiers, fsck, dfsadmin
- Gestion centralisée de caches avec Cacheadmin

Haute disponibilité

- Mise en place de la haute disponibilité sur distribution Cloudera
- Passage d'un système HDFS en mode HA
- Fédération de cluster Hadoop

Sécurité

- Mécanismes de sécurité et mise en oeuvre pratique : Activation de la sécurité avec Kerberos
- Sécurisation de Yarn avec proxy et Linux Container Executor

Exploitation

- Installation d'une grappe Hadoop et lancement des services
- Supervision des éléments par le NodeManager
- Travaux pratiques : visualisation des alertes, configuration des logs avec log4j

Modalités d'évaluation des acquis

En cours de formation, par des travaux pratiques. En fin de formation, par un questionnaire d'auto-évaluation.

{{< formation-require >}}

{{< formation-seealso >}}