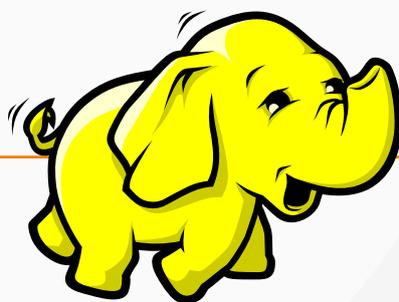




Plan de Formation **Hadoop**



Description de la formation



Bien comprendre le vocabulaire et le rôle de chaque brique de Hadoop

Cette formation Hadoop vous invite à faire le point sur les différents éléments de l'écosystème Hadoop et leurs rôles respectifs. Egalement, elle vous aidera à comprendre l'architecture des applicatifs hadoop et à savoir quels sont les apports et les cas d'usage des solutions hadoop.

La formation Hadoop Administration Cloudera vous apportera aussi les compétences et connaissances nécessaires pour

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



Hadoop

A qui s'adresse cette formation ?

Public

Ce cours Hadoop cible principalement les chefs de projets, développeurs, et toute personne souhaitant comprendre les mécanismes Hadoop et le rôle de chaque composant.



Hadoop

Contenu du cours

Introduction à l'écosystème Hadoop

- Rappels sur NoSQL. Le théorème CAP.
- Historique du projet hadoop
- Les fonctionnalités : stockage, outils 'extraction, de conversion, ETL, analyse, ...
- Exemples de cas d'utilisation sur des grands projets.
- Les principaux composants :
HDFS pour le stockage et YARN pour les calculs.
Les distributions et leur caractéristiques (HortonWorks, Cloudera, MapR, GreenPlum, Apache, ...)

Hadoop

Contenu du cours

L'architecture d'Hadoop

- Terminologie : NameNode, DataNode, ResourceManager
- Rôle et interactions des différents composants
- Présentation des outils d'infrastructure : ambari, avro, zookeeper;
- de gestion des données : pig, oozie, tez, falcon, pentaho, sqoop, flume;
- d'interfaçage avec les applications GIS;
- de restitution et requêtage : webhdfs, hive, hawq, impala, drill, stinger, tajo, mahout, lucene, elasticSearch, Kibana
- Les architectures connexes : spark, cassandra

Hadoop

Contenu du cours

Exemples interactifs avec Hadoop

- Démonstrations sur une architecture Hadoop multi-noeuds.
- Mise à disposition d'un environnement pour des exemples de calcul

Travaux Pratiques :

- Recherches dans des données complexes non structurées.

Applications

Cas d'usages de hadoop.

- Les infrastructures hyperconvergées avec les appliances hadoop
- Calculs distribués sur des clusters hadoop

Hadoop

Contenu du cours

Hadoop Cloudera

- Les fonctionnalités du framework Hadoop
- Les différentes versions
- Distributions : Apache, Cloudera, Hortonworks, EMR, MapR, DSE
- 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...

Hadoop

Contenu du cours

Les outils Hadoop

- Infrastructure / Mise en œuvre : Avro, Ambari, Zookeeper, Pig, Tez, Oozie,
- Vue d'ensemble
- Gestion des données
- Exemple de sqoop
- Restitution : webhdfs, Hive, Hawq, Mahout, ElasticSearch...
- Outils complémentaires : Spark, SparkQL, SparkMLib, Storm, BigTop, Zebra
- Outils complémentaires de développement : Cascading, Scalding, Flink
- Outils complémentaires d'analyse : RHadoop, Hama, Chukwa, Kafka

Hadoop

Contenu du cours

Installation et configuration

Présentation de Cloudera Manager

Installation en mode distribué

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 exécutable 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, API natives

Exemples en ligne de commandes avec hdfs, yarn, mapred

Présentation des fonctions offertes par le serveur HTTP

Hadoop

Contenu du cours

Travaux Pratiques :

- Organisation et configuration d'une grappe Hadoop avec Cloudera Manager
- Traitement de données
- Requêtage SQL avec Impala

Administration d'Hadoop

- Outils complémentaires à Yarn et hdfs : Jconsole, Jconsole Yarn
- Exemples sur le suivi de charges, l'analyse des journaux
 - Principe de gestion des noeuds, accès JMX
- Administration HDFS : présentation des outils de stockage des fichiers, fsck, dfsadmin
- Mise en oeuvre sur des exemples simples de récupération de fichiers
- Gestion centralisée de caches avec Cacheadmin

Hadoop

Contenu du cours

Haute disponibilité

- Mise en place de la haute disponibilité sur une distribution Cloudera
- Passage d'un système HDFS en mode HA
- Explication et démonstration d'une fédération de cluster Hadoop

Sécurité

- Mécanismes de sécurité et mise en œuvre pratique : Activation de la sécurité avec Kerberos dans core-site.xml
- Sécurisation de Yarn avec la mise en œuvre d'un proxy et d'un Linux Container Executor

Exploitation

- Installation d'une grappe Hadoop. Lancement des services.
- Principe de la supervision des éléments par le Node Manager

Travaux Pratiques :

- Visualisation des alertes en cas d'indisponibilité d'un nœud
- Configuration des logs avec log4j.



Knowledge forges Empires

Nos Contacts



empire-training.tn



+216 55 826 628
+216 74 201 616



16 Rue D'Athènes, Sfax



Contact@empire-training.tn