

Hadoop Hortonworks : administration avec Ambari

INFORMATIONS GÉNÉRALES

Type de formation : Formation continue

Éligible au CPF : Non

Domaine : IA, Big Data et Bases de données

Action collective : Non

Filière : Big Data

Rubrique : NoSQL et Hadoop

Code de formation : BD016

PRÉSENTATION

Objectifs & compétences

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 ...)

Public visé

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

€ Tarifs

Prix public : 2010 €

Tarif & financement :

Nous vous accompagnons pour trouver la meilleure solution de financement parmi les suivantes :

Le plan de développement des compétences de votre entreprise : rapprochez-vous de votre service RH.

Le dispositif FNE-Formation.

L'OPCO (opérateurs de compétences) de votre entreprise.

France Travail: sous réserve de l'acceptation de votre dossier par votre conseiller Pôle Emploi.

CPF -MonCompteFormation

Contactez nous pour plus d'information : contact@aston-institut.com

PROGRAMME

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, ...

Les outils Hadoop

Infrastructure/Mise en œuvre : 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

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 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, APIs natives

Exemples en ligne de commandes avec hdfs, yarn, mapred

Lieux & Horaires

Durée : 21 heures

Délai d'accès : Jusqu'à 8 jours avant le début de la formation, sous condition d'un dossier d'inscription complet

Prochaines sessions

Consultez-nous pour les prochaines sessions.

Présentation des fonctions offertes par le serveur http

Travaux pratiques : Organisation et configuration d'une grappe hadoop

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.

Haute disponibilité

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

Travaux pratiques : Passage d'un système HDFS en mode HA

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.

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.

MODALITÉS

Modalités

Modalités : en présentiel, distanciel ou mixte . Toutes les formations sont en présentiel par défaut mais les salles sont équipées pour faire de l'hybride. – Horaires de 9H à 12H30 et de 14H à 17H30 soit 7H – Intra et Inter entreprise.

Pédagogie : essentiellement participative et ludique, centrée sur l'expérience, l'immersion et la mise en pratique. Alternance d'apports théoriques et d'outils pratiques.

Ressources techniques et pédagogiques : Support de formation au format PDF ou PPT Ordinateur, vidéoprojecteur, Tableau blanc, Visioconférence : Cisco Webex / Teams / Zoom.

Pendant la formation : mises en situation, autodiagnostic, travail individuel ou en sous-groupe sur des cas réels.

Méthode

Fin de formation : entretien individuel.

Satisfaction des participants : questionnaire de satisfaction réalisé en fin de formation.

Assiduité : certificat de réalisation.

Validations des acquis : grille d'évaluation des acquis établie par le formateur en fin de formation.