

Plan de cours N° : 1120

Durée : 3 jours (21h)

PARTICIPANTS / PRE-REQUIS

Chef de projet, administrateur et toute personne souhaitant mettre en oeuvre un système distribué avec Hadoop.
Connaissance des commandes des systèmes Unix/Linux et des bases TCP/IP.

OBJECTIFS PEDAGOGIQUES

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

MOYENS PEDAGOGIQUES

Réflexion de groupe et apports théoriques du formateur
Travail d'échange avec les participants sous forme de réunion-discussion
Utilisation de cas concrets issus de l'expérience professionnelle
Validation des acquis par des exercices de synthèse
Alternance entre apports théoriques et exercices pratiques (en moyenne 30 et 70%)
Remise d'un support de cours.
Assistance post-formation d'une durée de 1 an sur le contenu de la formation via notre adresse mail dédiée formateurs@atp-formation.com

MOYENS PERMETTANT LE SUIVI DE L'EXECUTION ET DES RESULTATS

Feuille de présence signée en demi-journée,
Evaluation des acquis tout au long de la formation,
Questionnaire de satisfaction,
Attestation de stage à chaque apprenant,
Positionnement préalable oral ou écrit,
Evaluation formative tout au long de la formation,
Evaluation sommative faite par le formateur ou à l'aide des certifications disponibles.

MOYENS TECHNIQUES EN PRESENTIEL

Accueil des stagiaires dans une salle dédiée à la formation, équipée d'ordinateurs, d'un vidéo projecteur d'un tableau blanc.

MOYENS TECHNIQUES DES CLASSES A DISTANCE

A l'aide d'un logiciel comme Teams, Zoom etc... un micro et éventuellement une caméra pour l'apprenant, suivez une formation en temps réel et entièrement à distance. Lors de la classe en ligne, les apprenants interagissent et communiquent entre eux et avec le formateur.

Les formations en distanciel sont organisées en Inter-Entreprise comme en Intra-Entreprise. L'accès à l'environnement d'apprentissage (support de cours, labs) ainsi qu'aux preuves de suivi et d'assiduité (émargement, évaluation) est assuré.

Les participants recevront une convocation avec lien de connexion

Pour toute question avant et pendant le parcours, une assistance technique et pédagogique est à disposition par mail et par téléphone auprès de notre équipe par téléphone au 04.76.41.14.20 ou par mail à contact@atp-formation.com

ORGANISATION

Les cours ont lieu de 9h00-12h30 13h30-17h00.

PROFIL FORMATEUR

Nos formateurs sont des experts dans leurs domaines d'intervention
Leur expérience de terrain et leurs qualités pédagogiques constituent un gage de qualité.

ACCESSIBILITE

Les personnes atteintes de handicap souhaitant suivre cette formation sont invitées à nous contacter directement, afin d'étudier ensemble les possibilités de suivre la formation.

MISE A JOUR

01/01/2021

Plan de cours N° : 1120

Durée : 3 jours (21h)

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, Tez, Pig, Hive, HBase...

Les outils Hadoop

- Infrastructure/mise en oeuvre
 - Avro, Ambari, Zookeeper, Pig, Tez, Oozie
 - Vue d'ensemble
 - Gestion des données
 - Exemple de sqoop
- Restitution
 - Webhdfs, hive, Hawq, Mahout, ElasticSearch...
- Outils complémentaires de traitement
 - Spark, SparkQL, Spark/ML, Storm, BigTop, Zebra
- Outils complémentaires de développement
 - Cascading, Scalding, Flink
- Outils complémentaires d'analyse
 - RHadoop, Hama, Chukwa, Kafka

Installation et configuration

- Installation en mode distribué
- Configuration du cluster et de l'environnement
- Etude des fichiers de configuration
 - core-site.xml / hdfs-site.xml / mapred-site.xml / yarn-site.xml / 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...
- 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

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 noeuds
- Principe des accès JMX
 - Démonstration avec Prometheus
- 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
 - Déplacement d'un NameNode
 - Mise en mode maintenance
 - Gestion de la file d'attente, paramétrage, capacity-scheduler

Haute disponibilité

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

Plan de cours N° : 1120

Durée : 3 jours (21h)

Sécurité

- Mécanismes de sécurité et mise en oeuvre pratique de la sécurité avec Kerberos
- 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 de traitement YARN
- Sécurisation de YARN avec les Linux Container Executor

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 noeud
 - Configuration des logs avec log4j