

Hadoop - Architecture et administration de clusters

3 jours - 21 heures

Code formation : ADHSYS0654



adhara.fr

Objectifs

Connaitre les différents composants d'un cluster Hadoop et savoir dimensionner une solution répondant aux besoins des services. Savoir mettre en oeuvre les différentes façons de déployer Hadoop, et opérer les outils d'administration et de surveillance pour maintenir un cluster en condition opérationnelle.

Participants

Architectes et urbanistes SI, administrateurs systèmes. Toute personne souhaitant mettre en oeuvre un système Hadoop.

Prérequis

Connaissance systèmes Unix/Linux.

Pédagogie

La pédagogie est basée sur le principe de la dynamique de groupe avec alternance d'apports théoriques, de phases de réflexion collectives et individuelles, d'exercices, d'études de cas et de mises en situations observées. Formation / Action participative et interactive : les participants sont acteurs de leur formation notamment lors des mises en situation car ils s'appuient sur leurs connaissances, les expériences et mettront en oeuvre les nouveaux outils présentés au cours de la session.

Remarques

Certification

Profil de l'intervenant

Consultant-formateur expert. Suivi des compétences techniques et pédagogiques assuré par nos services.

Moyens techniques

Encadrement complet des stagiaires durant la formation. Espace d'accueil, configuration technique des salles et matériel pédagogique dédié pour les formations en centre. Remise d'une documentation pédagogique papier ou numérique à échéance de la formation.

Méthodes d'évaluation des acquis

Un contact téléphonique est systématiquement établi avec le stagiaire ou la personne chargée de son inscription afin de définir le positionnement. Si besoin, un questionnaire est adressé pour valider les prérequis en correspondance et obtenir toute précision nécessaire permettant l'adaptation de l'action. Durant la formation, des exercices individuels et collectifs sont proposés pour évaluer et valider les acquis du stagiaire. La feuille d'émargement signée par demi-journée ainsi que l'évaluation des acquis sont adressées avec la facture.

Programme

Introduction aux technologies Big Data

Stockage et traitement de données massives : problèmes et solutions
Panorama des technologies NoSQL, bases de données distribuées et en colonnes

Hadoop - Architecture et administration de clusters

3 jours - 21 heures

Code formation : ADHSYS0654



adhara.fr

Présentation de l'écosystème Hadoop

Coût, performance et évolutivité : promesses et gains effectifs

Les composants logiciels majeurs : Zookeeper, HDFS, HBase

Les différentes utilisations : comme solution de stockage, d'ETL, de traitement batch

Hadoop par la pratique

Notions de base : HDFS et programmation Map/Reduce

Design d'un cluster Hadoop

Topologie : les différents serveurs et leurs rôles

Configuration matérielle

Les différentes distributions Hadoop

Choix des briques logicielles selon l'usage du cluster

Dimensionnement

Provisionnement et déploiement

Déploiement manuel

Outils d'automatisation (Chef/Puppet, Whirr/Pallet)

Installation d'Hive, Pig, Hbase

Configuration et paramétrage

Administration et opération

Gestion des données (backup, localisation et réplication)

Gestion des jobs et schedulers

Monitoring du cluster

Ajout et décommission de nœuds

Benchmarks, tuning et optimisation

Résolution de problèmes, login et debugging

Intégration au SI

Stratégies et étapes d'intégration

Les différentes couches d'abstraction selon le public utilisateur

Connection aux bases de données relationnelles via Sqoop et JDBC

Ingestion de données via Flume

Interfacer avec les services avals

Hadoop et ses compléments

Forces et faiblesses de la plateforme selon les cas d'utilisation

Alternatives et compléments

Comment intégrer Hadoop à Storm, Cassandra, Mongo, Giraph ...

Déploiement à grande échelle

Hadoop sur le cloud : l'offre d'Amazon, Elastic MapReduce

Hadoop chez Facebook, LinkedIn, Orbitz...