

Formation Hadoop, Développement

Présentation

Pour faire face à un volume croissant de données à analyser, et fournir aux internautes des résultats de recherche pertinents, le framework Hadoop voit le jour au milieu des années 2000.

Devant le succès rencontré par la solution, Hadoop passera chez Cloudera qui créera alors la distribution Hadoop la plus utilisée à ce jour.

Durée : 21,00 heures (3 jours)

Tarif INTRA : Nous consulter

Objectifs de la formation

- Connaître les différents éléments de l'écosystème Hadoop et leurs rôles respectifs.
- Comprendre l'architecture des applicatifs Hadoop.
- Comprendre les apports et cas d'usage des solutions Hadoop.

Prérequis

- Avoir des connaissances en matière de développement d'applications d'entreprise, d'intégration de systèmes commerciaux et/ou de conception, d'interrogation et de rapports de bases de données.

Public

- Développeurs de logiciels,
- Administrateurs de bases de données,
- Analystes BI,
- Architectes logiciels,
- Chefs de projet.

Programme de la formation

Architecture Hadoop

- Histoire de Hadoop - Facebook, Dynamo, Yahoo, Google



- Le noyau Hadoop
- Architecture YARN, Hadoop 2.0

Système de fichiers distribués Hadoop (HDFS)

- Clusters HDFS - NameNodes, DataNodes et clients
- Métadonnées
- Administration en ligne

MapReduce

- Traitement et génération de grands ensembles de données
- Fonctions Map
- Programmation de MapReduce à l'aide de SQL / Bash / Python
- Traitement parallèle
- Failover

Entreposage de données avec Hive

- Synthétisation des données
- Requêtes ad-hoc
- Analyser de grands ensembles de données
- HiveQL (langage de requête de type SQL)
- Intégration avec les bases de données SQL
- Analyse des n-grammes

Traitement parallèle avec Pig

- Évaluation parallèle
- Interface du langage de requête
- Algèbre relationnelle

Extraction de données avec Mahout

- Regroupement ou Clustering
- Classification
- Filtrage collaboratif par lots

Recherche avec Elastic Search

- Concepts de recherche Elastic

- Installation, importation des données
- Démonstration de l'API, exemples de requêtes

Stockage de données structurées avec HBase

- Big Data : échelle de taille d'un grand ensemble de données.
- Optimiser l'accès en lecture/écriture en temps réel

Base de données multi-master Cassandra

- Le modèle de données de Cassandra
- Cohérence éventuelle
- Quand utiliser Cassandra

Redis

- Modèle de données Redis
- Quand utiliser Redis

MongoDB

- Modèle de données MongoDB
- Installation de MongoDB
- Quand utiliser MongoDB

Kafka

- L'architecture de Kafka
- Installation
- Exemple d'utilisation
- Quand utiliser Kafka

Architecture Lambda

- Concept
- Hadoop + Intégration du traitement des flux
- Exemples d'architecture

Big Data dans le Cloud

(Si le temps le permet)

- Amazon Web Services

- Concepts : modèle de paiement à l'utilisation
- Amazone S3, EC2, EMR
- Plateforme Google Cloud
- Google Big Query

Organisation

Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes.

Moyens pédagogiques et techniques

- Alternance d'apports théoriques et de mises en pratique contextualisées.
- Progression pédagogique structurée, de la compréhension à la mise en œuvre opérationnelle.
- Pédagogie immersive favorisant une mise en application immédiate des compétences.
- Formation interactive et participative s'appuyant sur des ateliers pratiques, études de cas et échanges collectifs (+ 60 % du temps de formation.)
- Séquences pédagogiques rythmées et évaluations formatives régulières pour renforcer l'ancrage des apprentissages.
- Supports de cours, ressources documentaires et références mis à disposition par le formateur

Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation

En amont de la formation *(évaluation en ligne dans l'extranet apprenant)*

- Recueil des besoins et des attentes des participants.
- Auto-positionnement permettant d'évaluer le niveau initial des apprenants.

Tout au long de la formation

- Évaluations continues réalisées par le formateur travers des questions orales, exercices, QCM, études de cas et mises en situation pratiques.

À l'issue de la formation (*évaluations en ligne dans l'extranet apprenant*)

- Auto-positionnement afin de mesurer l'acquisition des compétences visées.
- Validation de l'acquisition des compétences par le formateur
- Questionnaire de satisfaction « à chaud » à l'issue de la formation
- Questionnaire de satisfaction "à froid", 45 jours après la formation afin d'évaluer l'ancrage des acquis et leur mise en application en situation professionnelle.

NB : dans le cadre d'une Action collective, chaque stagiaire bénéficiaire sera contacté par un prestataire choisi par l'Opco Atlas afin d'évaluer « à chaud » la qualité de la formation suivie.

Accessibilité

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.