DOCAPOSTE Institute

Les architectures et infrastructures pour le Big Data

Présentation

Cette formation vous présente l'architecture et l'infrastructure sous-jacente des projets Big Data. Nous commencerons par définir le cadre d'un projet Big Data, puis expliquerons comment le type et la quantité de données traitées impactent l'architecture au niveau du stockage ou du traitement. Quelles sont les évolutions spécifiques par rapport aux standards (serveurs, SGBD, systèmes de fichiers, etc.) et quels sont les avantages spécifiques des solutions ou produits NoSQL comme Hadoop ?

Apprenez à installer, configurer, sécuriser et surveiller des clusters Hadoop multi-nœuds, offrant ainsi une infrastructure de qualité aux développeurs et aux analystes.

Durée : 21,00 heures (3 jours)
Tarif INTRA : Nous consulter

Objectifs de la formation

- Comprendre les principaux concepts du Big Data ainsi que l'écosystème technologique d'un projet Big Data
- Savoir analyser les difficultés propres à un projet Big Data
- Déterminer la nature des données manipulées
- Appréhender les éléments de sécurité, d'éthique et les enjeux juridiques
- Exploiter les architectures Big Data
- Mettre en place des socles techniques complets pour des projets Big Data

Prérequis

- Avoir une bonne culture générale des systèmes d'information et plus particulièrement
- Avoir des connaissances de base des modèles relationnels, des statistiques et des langages de programmation.

Public

- Chefs de projets

Docaposte Institute
45-47 Boulevard Paul Vaillant Couturier, 94200 IVRY SUR SEINE France
Tél : 01 56 29 70 01 - Email : hello.institute@docaposte.fr
SIRET : 813 058 542 00032 - Code APE : 8559A - N TVA Intra : FR 58 813 058 542
Déclaration d'activité enregistrée sous le numéro 11941117994 auprès du préfet de région d'Ile-de-France

- Architectes
- Développeurs
- Data Scientists ou toute personne souhaitant connaître les outils pour concevoir une architecture Big Data...

Programme de la formation

Définition et contexte spécifique des projets Big Data

- Origines du Big Data
- Les données au cœur des sujets : explosion des données, connexions Big Data et IoT (Internet des objets), données structurées, données semi-structurées, données non structurées et données structurées
- Les limites des architecturales actuelles
- Les définitions des systèmes Big Data
- Principes de fonctionnement
- Différentes offres de marché

Propriété des données, environnement de traitement légal et sécurité

- Sécurité éthique et questions juridiques
- Données personnelles
- Informations confidentielles, interdictions
- Réglementation des Données Numériques par la CNIL
- Accords Nationaux

Impact des choix technologiques liés à l'infrastructure et à l'architecture Big Data

- Architectures décisionnelles "traditionnelles" (Datastores, Data Warehouses, Data Marts, etc.)
- Philosophie des bases NoSQL : Column Family, orienté document, clé-valeur, diagramme
- Plusieurs acteurs (MongoDB, Cassandra, etc.)
- Big Table / Big Query
- Moteur de base de données (Exadata)
- Base de données vectorielle (Sybase IQ)
- Hadoop, système entièrement autonome ?
- Impacts économiques

Mise en œuvre et élaboration d'une stratégie dédiée au Big Data

- Besoins en sujet de Big Data

Docaposte Institute
45-47 Boulevard Paul Vaillant Couturier, 94200 IVRY SUR SEINE France
Tél : 01 56 29 70 01 - Email : hello.institute@docaposte.fr
SIRET : 813 058 542 00032 - Code APE : 8559A - N TVA Intra : FR 58 813 058 542
Déclaration d'activité enregistrée sous le numéro 11941117994 auprès du préfet de région d'Ile-de-France

- Atteindre les impartiaux cabinet au bon droit des conjoncture
- Outils du marché dédiés au Big Data
- Répondre aux attentes d'un collaborateur

Architectures distribuées

- Problématiques et objectifs
- Des conjoncture cohérentes, disponibles et tolérantes aux pannes ?
- Les architectures lourdement parallèles
- L'ouverture aux traitements complexes (datamining, intention learning, etc.)
- Paradigmes de calculs distribués
- Les bases NoSQL et le calcul distribué
- Qualité des données (Dataquality)
- Liens entre infrastructure et qualité des données
- Pas de qualité, pas d'analyse
- Les 4 V
- Bases à chaud et à froid
- Les apports d'un outil de Dataquality
- Pourquoi utiliser un ETL ?
- Illustration via Talend Data Integration
- Analyser les données en les fusionnant avec les données internes
- Le Master Data Management (MDM)

Préparation et visage du cluster Hadoop

- Principes de fonctionnement de Hadoop Distributed File System (HDFS)
- Principes de fonctionnement de MapReduce
- Design « type » du cluster

Installation d'une plateforme Hadoop

- Type de déploiement
- Installation d'Hadoop
- Installation de divers composants (Hive, Pig, HBase, Flume...)
- Différences parmi les distributions Cloudera, Hortonworks et MapR

Gestion d'un cluster Hadoop

- Gestion des nœuds du cluster Hadoop
- Les TaskTracker, JobTracker dans MapReduce
- Gestion des services via les schedulers
- Gestion des logs

Docaposte Institute
45-47 Boulevard Paul Vaillant Couturier, 94200 IVRY SUR SEINE France
Tél : 01 56 29 70 01 - Email : hello.institute@docaposte.fr
SIRET : 813 058 542 00032 - Code APE : 8559A - N TVA Intra : FR 58 813 058 542
Déclaration d'activité enregistrée sous le numéro 11941117994 auprès du préfet de région d'Ile-de-France

Gestion des données pour HDFS

- Import de conjoncture externes (fichiers, bases de conjoncture relationnelles) enthousiasme HDFS
- Manipulation des fichiers HDFS

Configuration avancée

- Gestion des autorisations et de la sécurité
- Reprise sur échec d'un name node (MRV1)
- Haute disponibilité d'un NameNode (MRV2/YARN)

Monitoring et optimisation

- Monitoring (Ambari, Ganglia...)
- Benchmarking/profiling d'un cluster
- Les outils Apache GridMix, Vaidya
- Taille des blocs
- Autres options de tuning (maniement de la compression, visage mémoire...)

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

- Apports didactiques pour apporter des connaissances communes.
- Mises en situation de réflexion sur le thème du stage et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 60 % / 40 %.
- Supports de cours fournis au format papier et/ou numérique.
- Ressources documentaires en ligne et références mises à disposition par le formateur.
- Pour les formations en présentiel dans les locaux de Softeam, les stagiaires sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau blanc

Docaposte Institute
45-47 Boulevard Paul Vaillant Couturier, 94200 IVRY SUR SEINE France
Tél : 01 56 29 70 01 - Email : hello.institute@docaposte.fr
SIRET : 813 058 542 00032 - Code APE : 8559A - N TVA Intra : FR 58 813 058 542
Déclaration d'activité enregistrée sous le numéro 11941117994 auprès du préfet de région d'Ile-de-France

ou paperboard. Un carnet de notes est offert. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

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

En amont de la formation :

- Recueil des besoins : permet de récolter des informations sur le stagiaire (profil, formation, attentes particulières, ...).
- Auto-positionnement des stagiaires afin de mesurer le niveau de départ.

Tout au long de la formation :

- Évaluation continue des acquis via des questions orales, exercices / projet fil rouge, des QCM, des cas pratiques et mises en situation.

A la fin de la formation :

- Auto-positionnement des stagiaires afin de mesurer l'acquisition des compétences.
- Evaluation du formateur des compétences acquises par les stagiaires.
- Questionnaire de satisfaction à chaud : permet de connaître le ressenti des stagiaires à l'issue de la formation.
- Questionnaire de satisfaction à froid : permet d'évaluer les apports réels de la formation et leurs mises en application au quotidien.

Accessibilité

Cette formation est **accessible** aux personnes en situation de handicap, consultez-nous pour plus d'informations.

Docaposte Institute
45-47 Boulevard Paul Vaillant Couturier, 94200 IVRY SUR SEINE France
Tél : 01 56 29 70 01 - Email : hello.institute@docaposte.fr
SIRET : 813 058 542 00032 - Code APE : 8559A - N TVA Intra : FR 58 813 058 542
Déclaration d'activité enregistrée sous le numéro 11941117994 auprès du préfet de région d'Ile-de-France