

Big Data, méthodes et solutions pratiques pour l'analyse des données

Cours Pratique de 3 jours - 21h

Réf : IBD - Prix 2024 : 2 280€ HT

Le prix pour les dates de sessions 2025 pourra être révisé

Ce cours vous permettra de comprendre les enjeux et les apports du Big Data ainsi que les technologies pour le mettre en œuvre. Vous apprendrez à intégrer des volumétries massives de données structurées et non structurées via un ETL, puis à les analyser grâce à des modèles statistiques et des dashboards dynamiques.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

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

Mettre en place une plateforme Hadoop et ses composants de base, utiliser un ETL pour gérer les données, créer des modèles d'analyse et dashboards.

LE PROGRAMME

dernière mise à jour : 11/2022

1) Comprendre les concepts et les enjeux du Big Data

- Origines et définition du Big Data.
- Les chiffres clés du marché dans le monde et en France.
- Les enjeux du Big Data : ROI, organisation, confidentialité des données.
- Un exemple d'architecture Big Data.
- Aspect éthique et juridique de la gestion des données.
- La sécurité des données.

2) Les technologies du Big Data

- Description de l'architecture et des composants de la plateforme Hadoop.
- Les modes de stockage (NoSQL, HDFS).
- Principes de fonctionnement de MapReduce, Spark, Storm...
- Principales distributions du marché (Hortonworks, Cloudera, MapR, Elastic Map Reduce, Biginsights).
- Installer une plateforme Hadoop.
- Les technologies du datascientist.
- Présentation des technologies spécifiques pour le Big Data (Tableau, Talend, Qlikview...).

Exercice : Installation d'une plateforme Big Data Hadoop (via Cloudera QuickStart ou autre).

FINANCEMENT

Ce cours fait partie des actions collectives Atlas.

PARTICIPANTS

Chefs de projets, architectes, développeurs, data scientists ou toute personne souhaitant connaître les outils pour concevoir une architecture 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.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

3) Gérer les données structurées et non structurées

- Principes de fonctionnement de Hadoop Distributed File System (HDFS).
- Importer des données externes vers HDFS.
- Réaliser des requêtes SQL avec HIVE.
- Utiliser PIG pour traiter la donnée.
- Le principe des ETL (Talend...).
- Gestion de streaming de données massive (NIFI, Kafka, Spark, Storm...)

Exercice : Implémentation de flux de données massives.

LES DATES

CLASSE À DISTANCE

2024 : 04 nov.

2025 : 19 mars, 25 juin, 15 oct.,
17 déc.

PARIS

2024 : 28 oct.

2025 : 12 mars, 18 juin, 08 oct.,
10 déc.

LYON

2025 : 19 mars, 25 juin, 15 oct.,
17 déc.