

Formation : Big data, méthodes et solutions pratiques pour l'analyse des données

Formation pratique - 3j - 21h00 - Réf. IBD

Prix : 2010 € H.T.

★★★★☆ 3,8 / 5

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, le participant 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

Public concerné

Chefs de projet, 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.

Vérifiez que vous avez les prérequis nécessaires pour profiter pleinement de cette formation en faisant [ce test](#).

PARTICIPANTS

Chefs de projet, 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.

Méthodes et moyens pédagogiques

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.

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.

Programme de la formation

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, etc.
- Principales distributions du marché (Hortonworks, Cloudera, MapR, Elastic Map Reduce, Biginsights).
- Installer une plateforme Hadoop.
- Les technologies du data scientist.
- Présentation des technologies spécifiques pour le big data (Tableau, Talend, QlikView, etc.).

Exercice

Installation d'une plateforme big data Hadoop (via Cloudera QuickStart ou autre).

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, etc.).
- Gestion de streaming de données massive (NIFI, Kafka, Spark, Storm, etc.)

Exercice

Implémentation de flux de données massives.

Dates et lieux

CLASSE À DISTANCE
2026 : 10 juin, 23 sep., 25 nov.

PARIS LA DÉFENSE
2026 : 3 juin, 16 sep., 18 nov.

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 formations pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque formation 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 participant 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

Pour toute question ou besoin relatif à l'accessibilité, vous pouvez joindre notre équipe PSH par e-mail à l'adresse psh-accueil@orsys.fr.