

Formation : Big data, synthèse technique

Cours de synthèse - 2j - 14h00 - Réf. BAG

Prix : 2020 CHF H.T.



Cette synthèse vous présentera les enjeux et les apports du big data ainsi que les technologies disponibles pour sa mise en œuvre. Vous suivrez les étapes d'un projet de données massives depuis l'installation d'une plateforme big data, l'ingestion et le traitement des données, jusqu'à la visualisation des résultats.

Objectifs pédagogiques

À l'issue de la formation, le participant sera en mesure de :

- ✓ Découvrir les concepts clés du big data
- ✓ Comprendre l'écosystème technologique d'un projet big data
- ✓ Évaluer les techniques de gestion des flux de données massives
- ✓ Implémenter des modèles d'analyses statistiques pour répondre aux besoins métiers
- ✓ Découvrir les outils de data visualisation

Public concerné

Dataminers, chargés d'études statistiques, développeurs, chefs de projet, consultants en informatique décisionnelle.

Prérequis

Connaissances de base des modèles relationnels, des statistiques et des langages de programmation. Connaissances de base des concepts de la business intelligence.

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

Méthodes et moyens pédagogiques

Démonstration

Présenter la plateforme Hadoop et ses composants de base, utiliser un ETL pour gérer les données, créer des modèles d'analyse et dashboards.

PARTICIPANTS

Dataminers, chargés d'études statistiques, développeurs, chefs de projet, consultants en informatique décisionnelle.

PRÉREQUIS

Connaissances de base des modèles relationnels, des statistiques et des langages de programmation.
Connaissances de base des concepts de la business intelligence.

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.

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 clés et les enjeux du big data

- Les origines du big data.
- La valeur de la donnée : un changement d'importance.
- La donnée en tant que matière première.
- Les chiffres clés du marché dans le monde et en France.
- Les enjeux du big data : ROI, organisation, confidentialité des données.

Démonstration

Présentation d'une architecture big data.

2 Technologies du big data

- Architecture et composants de la plateforme Hadoop 2.
- Les modes de stockage (NoSQL, HDFS).
- Fonctionnement de MapReduce et Yarn...
- Principales distributions Hadoop : Hortonworks, Cloudera, MapR...
- Les technologies : Spark, Storm, Databrick, Machine Learning Azure...
- Démarche d'installation d'une plateforme Hadoop.
- Présentation des technologies spécifiques pour le big data (Talend, Tableau, QlikView...).

Démonstration

Installation d'une plateforme big data complète.

3 Traitement des données big data

- Fonctionnement de Hadoop Distributed File System (HDFS).
- Importer des données vers HDFS.
- Traitement des données avec PIG.
- Requêtes SQL avec HIVE.
- Création de flux de données massives avec un ETL.

Démonstration

Implémentation de flux de données massives.

4 Méthodes d'analyse et traitements des données pour le big data

- Les méthodes d'exploration.
- Segmentation et classification.
- machine learning, estimation et prédiction.
- Le temps réel, l'intelligence artificielle.
- L'implémentation des modèles.

Démonstration

Présentation de l'environnement Spark, Jupyter Notebook, R Notebook et Shiny. Mise en place d'analyses de machine learning avec le langage R, Python et Scala.

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.

5 Data visualisation, représenter des données de façon visuelle

- Principales solutions du marché.
- Aller au-delà des rapports statiques.
- La data visualisation et l'art de raconter des chiffres de manière créative et ludique.
- Mesurer l'e-réputation, la notoriété d'une marque, l'expérience et la satisfaction clients...

Démonstration

Présentation et utilisation d'un outil de data visualisation pour constituer des analyses dynamiques.

6 Conclusion

- Les conditions du succès.
- Synthèse des meilleures pratiques.
- Bibliographie.

Dates et lieux

CLASSE À DISTANCE

2026 : 2 avr., 26 mai, 8 sep., 17 nov.