

# Big Data, synthèse technique

Cours Synthèse de 2 jours - 14h

Réf : BAG - Prix 2024 : 1 950€ HT

Le prix pour les dates de sessions 2025 pourra être révisé

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 l'apprenant 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

## DÉMONSTRATION

Présenter la 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 : 01/2023

### 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.

## 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.

## 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.

- 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'IA.
- 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.

#### 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.

## LES DATES

---

### CLASSE À DISTANCE

2024 : 10 oct., 12 déc.

2025 : 24 juin, 16 déc.

### PARIS

2024 : 05 déc.

2025 : 17 juin, 09 déc.