

Stage pratique de 5 jour(s)  
Réf : BID

## Participants

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

## Pré-requis

Connaissances de base des modèles relationnels, des statistiques et des langages de programmation. Connaissances de base des concepts de la Business Intelligence.

Prix 2019 : 3090€ HT

## Dates des sessions

### AIX

02 déc. 2019, 17 fév. 2020  
25 mai 2020

### ANGERS

09 déc. 2019, 03 fév. 2020  
11 mai 2020

### BORDEAUX

02 déc. 2019, 10 fév. 2020  
25 mai 2020

### BRUXELLES

25 nov. 2019, 06 jan. 2020  
11 mai 2020

### DIJON

09 déc. 2019, 09 mar. 2020  
08 juin 2020

### GENEVE

09 déc. 2019, 02 mar. 2020  
15 juin 2020

### GRENOBLE

25 nov. 2019, 03 fév. 2020  
11 mai 2020

### LILLE

02 déc. 2019, 03 fév. 2020  
11 mai 2020

### LIMOGES

09 déc. 2019, 10 fév. 2020  
25 mai 2020

### LUXEMBOURG

09 déc. 2019, 02 mar. 2020  
15 juin 2020

### LYON

02 déc. 2019, 09 mar. 2020  
08 juin 2020

### MONTPELLIER

25 nov. 2019, 09 mar. 2020  
08 juin 2020

### NANCY

09 déc. 2019, 10 fév. 2020  
25 mai 2020

### NANTES

02 déc. 2019, 03 fév. 2020  
11 mai 2020

### NIORT

09 déc. 2019, 03 fév. 2020  
11 mai 2020

### ORLEANS

25 nov. 2019, 02 mar. 2020  
15 juin 2020

### PARIS

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

Ce stage 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 PEDAGOGIQUES

Comprendre les concepts et l'apport du Big Data par rapport aux enjeux métiers  
Comprendre l'écosystème technologique nécessaire pour réaliser un projet Big Data  
Acquérir les compétences techniques pour gérer des flux de données complexes, non structurés et massifs  
Implémenter des modèles d'analyses statistiques pour répondre aux besoins métiers  
Appréhender un outil de data visualisation pour restituer des analyses dynamiques

### 1) Comprendre les concepts et les enjeux du Big Data

### 2) Les technologies du Big Data

### 3) Gérer les données structurées et non structurées

### 4) Technique et méthodes Big data analytics

### 5) Data visualisation et cas d'usage concrets

### 6) Conclusion

## Exercice

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.

## 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.

## 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).

## 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.

## 4) Technique et méthodes Big data analytics

- Machine Learning, une composante de l'intelligence artificielle.
- Découvrir les trois familles : Régression, Classification et Clustering.
- La préparation des données (data preparation, feature engineering).
- Générer des modèles en R ou Python.
- Ensemble Learning.
- Découvrir les outils du marché : Jupyter Notebook, Dataiku, Amazon Machine Learning...

## Exercice

Mise en place d'analyses avec une des outils étudiés.

## 5) Data visualisation et cas d'usage concrets

- Définir le besoin de la data visualisation.
- Analyse et visualisation des données.

23 sep. 2019, 07&21 oct. 2019  
 04&18 nov. 2019, 02&16 déc. 2019  
 06&20 jan. 2020, 03&17 fév. 2020  
 02&16 mar. 2020, 06&20 avr. 2020  
 11&25 mai 2020, 15&22 juin 2020  
 06 juil. 2020, 17 août. 2020

## REIMS

09 déc. 2019, 02 mar. 2020  
 15 juin 2020

## RENNES

25 nov. 2019, 17 fév. 2020  
 25 mai 2020

## ROUEN

09 déc. 2019, 02 mar. 2020  
 15 juin 2020

## SOPHIA-ANTIPOLIS

02 déc. 2019, 03 fév. 2020  
 11 mai 2020

## STRASBOURG

02 déc. 2019, 17 fév. 2020  
 25 mai 2020

## TOULON

09 déc. 2019, 17 fév. 2020  
 25 mai 2020

## TOULOUSE

02 déc. 2019, 10 fév. 2020  
 25 mai 2020

## TOURS

09 déc. 2019, 10 fév. 2020  
 25 mai 2020

- Peut concerner tous les types de données dans la DataViz ?
- Les outils DataViz du marché.

## Exercice

*Installation et utilisation d'un outil de Data Visualisation pour constituer des analyses dynamiques.*

## 6) Conclusion

- Ce qu'il faut retenir.
- Synthèse des bonnes pratiques.
- Bibliographie.

## Modalités d'évaluation

L'évaluation des acquis se fait tout au long de la session au travers des multiples exercices à réaliser (50 à 70% du temps).

## 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.

## 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.

- A 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.