

# Big Data, BI, NoSQL, SGBD, Data, état de l'art et méthodes

Les données sont au coeur du système d'information. L'efficience de leur organisation et de leur exploitation est une préoccupation majeure pour l'entreprise. A l'heure de la gouvernance des données, du Big Data, du géomarketing, de la BI et du Data Mining, l'expertise métier et la compréhension des solutions techniques sont essentielles. C'est pourquoi ORSYS vous propose des séminaires et cours de synthèse, construits sur l'expérience de nos formateurs experts, pour faire le point sur les technologies et les meilleures pratiques.

## Cycles certifiants

[Cycle certifiant Data Scientist](#) (réf. KDS) .... p.2

## Gestion de données, stockage

[Démarche de gestion de la qualité des données](#) (réf. MQD) .... p.4

[Master Data Management, la gouvernance des données](#) (réf. CYE) .... p.6

[Talend Open Studio for Data Quality, gérer la qualité des données](#) (réf. TDQ) .... p.8

[Big Data, état de l'art](#) (réf. BGA) .... p.10

[Bases de données NoSQL, enjeux et solutions](#) (réf. NSQ) .... p.12

[Réseaux de stockage, SAN/NAS](#) (réf. SAN) .... p.14

[Open Source, base de données](#) (réf. OSC) .... p.16

## Conception de bases de données

[Bases de données et langage SQL pour non-informaticiens](#) (réf. SIF) .... p.18

[Bases de données, modélisation et choix techniques](#) (réf. CBC) .... p.20

[PowerAMC, concevoir une base de données](#) (réf. AMC) .... p.22

[UML 2, modéliser et créer vos bases de données](#) (réf. UMD) .... p.24

[Conception d'une base de données avec Oracle Data Modeler](#) (réf. BDM) .... p.25

## Data Mining, Machine Learning, statistiques

[Big Data, Data Science, Machine Learning, leur impact dans l'entreprise](#) (réf. BMD) .... p.27

[Data Mining : synthèse](#) (réf. DMI) .... p.29

[Modélisation statistique](#) (réf. STA) .... p.31

[Big Data, méthodes et solutions pratiques pour l'analyse des données](#) (réf. BID) .... p.33

## Conception de systèmes d'aide à la décision

[Construire un Data Warehouse](#) (réf. DAW) .... p.35

[Modélisation en étoile, conception et mise en œuvre](#) (réf. AMT) .... p.37

[Social Media Analytics, stratégie et usages](#) (réf. BIM) .... p.39

[Systèmes décisionnels, démarche et solutions](#) (réf. SID) .... p.41

Stage pratique de 12  
jour(s)  
Réf : KDS

## Participants

Data Miner, Data Scientist, analyste, responsable de reporting décisionnel, statisticien, tout acteur ayant à consolider et faire valider ses connaissances pour être reconnu Data Scientist.

## Pré-requis

Connaissances de base en mathématiques / logique (cumul, pourcentage, fonctions conditionnelles).

Prix 2018 : 5660€ HT

## Dates des sessions

PARIS

17 sep. 2018, 22 nov. 2018

## Composition du cycle

### - Modélisation statistique

Réf : STA, Durée : 2 j

### - Data Mining : synthèse

Réf : DMI, Durée : 2 j

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

Réf : BID, Durée : 5 j

### - Master Data Management, la gouvernance des données

Réf : CYE, Durée : 2 j

### - Journée de certification du cycle certifiant Data Scientist

Réf : KSW, Durée : 1 j

## Modalités d'évaluation

Pour les certifications, des examens blancs permettent de mesurer le degré de préparation. Pour les cycles certifiants, une journée d'examen est mise en place.

## 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

# Cycle certifiant Data Scientist

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

La "science des données" nécessite des compétences spécifiques afin de prévoir le comportement des clients ou anticiper des tendances. Ce cycle Data Scientist vous apportera les meilleures pratiques en vigueur, et vous permettra de mettre en œuvre les méthodes et outils destinés à "faire parler les données".

## OBJECTIFS PEDAGOGIQUES

Comprendre le rôle stratégique des données dans l'entreprise et savoir maîtriser leur qualité  
Comprendre et mettre en pratique l'essentiel des modèles statistiques  
Mettre en pratique les techniques de modélisation du Data Mining  
Maîtriser les outils et pratiques d'analyse des données propres au Big Data

### [1\) Présentation de la gouvernance des données](#)

### [2\) L'essentiel des modèles statistiques](#)

### [3\) Data Mining, extraction de connaissances à partir de données brutes](#)

### [4\) Big Data, analyse des données massives](#)

## Méthodes pédagogiques

Le socle des connaissances en statistiques est fondamental. Le stage STA doit être suivi avant le séminaire DMI et le stage BID. Pas d'autres contraintes sur le suivi des cours pendant ce cycle.

## 1) Présentation de la gouvernance des données

- Rôle stratégique des données pour l'entreprise.
- Etat des lieux sur la qualité des données. Définition des notions "Donnée" et "Information".
- Evaluer la qualité des données des entreprises. Réconciliation des données.
- Contrôle de qualité et bonnes pratiques.
- Cycle de vie des données : archivage de base de données. Confidentialité des données.
- Présentation des offres de MDM.

### Travaux pratiques

Arriver à un consensus sur la qualité des données. Lien qualité des données et solvabilité II.

## 2) L'essentiel des modèles statistiques

- Fondamentaux de la statistique descriptive.
- Démarche et modélisation d'une analyse statistique.
- Paramètres de position et de dispersion.
- Analyses prédictives : mise en place d'un modèle de régression linéaire.
- Tests et intervalle de confiance. Valider la précision d'une estimation.
- Panorama des outils (R, SAS...).

### Travaux pratiques

Etude d'une population et dimensionnement d'un échantillon. Modélisation statistique d'un phénomène. Calcul de paramètres de position et de dispersion sur différents échantillonnages.

## 3) Data Mining, extraction de connaissances à partir de données brutes

- Définition et finalité du Data Mining (DM).
- Les méthodes prédictives et descriptives. Les arbres de décisions. Les réseaux de neurones.
- La méthode descriptive du clustering.
- Exemples d'application du DM : le scoring, le géomarketing.
- Méthodologie de projet. Les outils du marché.

### Travaux pratiques

Présentation d'applications du clustering. Mise en œuvre de la méthode de scoring. Cas pratique d'utilisation de géomarketing.

## 4) Big Data, analyse des données massives

- Origines et définition du Big Data : la BI face à la croissance et à la diversité des données.
- Les technologies du Big Data.
- Gérer les données structurées et non structurées.
- Les méthodes d'analyse des données pour le Big Data.
- Data visualisation et cas d'usage concrets.

### Travaux pratiques

Installation d'une plateforme Big Data. Talend For Big Data, implémentation de flux de données. Mise en place d'analyses avec le logiciel R.

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.

Cours de synthèse de 2  
jour(s)

Réf : MQD

## Participants

Data Quality Analysts, chefs de projet qualité des données, urbanistes SI, consultants AMOA SI qualité des données, responsables SI métier, experts méthode qualité, responsables SI.

## Pré-requis

Bonnes connaissances des enjeux de la mise en place d'un MDM dans l'entreprise.

Prix 2018 : 1770€ HT

## Dates des sessions

PARIS

11 oct. 2018, 06 déc. 2018

## Modalités d'évaluation

L'objectif de cette formation étant essentiellement de fournir une synthèse des méthodes et technologies existantes, il n'est pas nécessaire d'avoir recours à une évaluation des acquis.

## 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

# Démarche de gestion de la qualité des données

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

*La gestion de la qualité est nécessaire pour garantir des données fiables pour les SI opérationnel et décisionnel. Vous aborderez les points essentiels pour initier cette démarche : enjeu de data gouvernance, enjeu réglementaire, respect des règles métier, mesure de la qualité et qualification de la chaîne de saisie.*

## OBJECTIFS PEDAGOGIQUES

Comprendre le rôle stratégique de la valeur des données pour l'entreprise  
Comprendre les méthodologies de mesure de la qualité des données  
Mettre en place une démarche de mesure de la qualité des données  
Identifier et corriger les problèmes liés à la qualité des données

### [1\) Introduction](#)

### [2\) Définition de la qualité et cadre méthodologique](#)

### [3\) Démarche gestion de la qualité des données](#)

### [4\) Contrôle qualité et bonnes pratiques](#)

### [5\) Suivi opérationnel de la qualité des données](#)

### [6\) Hausser le niveau de la qualité](#)

### [7\) Synthèse](#)

## 1) Introduction

- Rôle stratégique des données pour l'entreprise.
- Différence entre données et information.
- Les différentes sources de données de l'entreprise.
- Les différentes formes d'exploitation de données.
- Architectures : relationnelle, NoSQL ou Big Data.

## 2) Définition de la qualité et cadre méthodologique

- Définition et mesure de la qualité des données.
- Les méthodologies de la qualité des données.
- Comparaison entre TDQM/DWQ/AIMQ/ORME Data Quality.
- Les principes d'évaluation de la qualité des données des entreprises.
- Synthèse sur les méthodologies qualité.
- Dimensions - définitions et mesures.

### Etude de cas

*Cas fil rouge : un groupe lance une démarche qualité pour la conformité avec Solvabilité 2 et souhaite améliorer la qualité de ses données clients. Organisation générale de la démarche.*

## 3) Démarche gestion de la qualité des données

- La place de la qualité dans la démarche de gouvernance.
- Les acteurs et l'organisation. L'exemple COBIT.
- Mise en œuvre de la démarche Projet.
- Le coût de la non qualité.
- Périmètre de la démarche qualité.
- Niveau d'approche et de granularité.

### Etude de cas

*Les 10 actions à lancer par le comité de gouvernance.*

## 4) Contrôle qualité et bonnes pratiques

- Audit de la qualité.
- Identification des données de faible qualité.
- Collecte et stockage des mesures de la qualité.
- Une approche centralisée de la qualité des données.
- Types de contrôles et outils statistiques.
- Exploitation et évaluation des mesures de la qualité.

## 5) Suivi opérationnel de la qualité des données

- Tableaux de bord qualité.
- Cycle d'amélioration continue.
- Quel rôle pour la gouvernance ?

### Etude de cas

*L'équipe qualité des données du groupe met en place un reporting : définition des indicateurs et méthode d'acquisition.*

## 6) Hausser le niveau de la qualité

- Intervenir en amont de la chaîne.

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.

- Identification des cas atypiques.
- Gestion ciblée des cas de données de faible qualité.
- La réconciliation des données.

#### **Etude de cas**

*Exemple de réconciliation des données dans un groupe suite à l'intégration d'une filiale.*

### 7) Synthèse

- L'offre logicielle : ETL et qualité.
- Synthèse.
- Bibliographie et sites Web.

Cours de synthèse de 2  
jour(s)

Réf : CYE

## Participants

Toute personne ayant à mettre en place une démarche de gouvernance des données d'une entreprise et/ou un projet de Master Data Management.

## Pré-requis

Aucune connaissance particulière.

Prix 2018 : 1770€ HT

## Dates des sessions

### PARIS

20 sep. 2018, 08 nov. 2018  
20 déc. 2018

## Modalités d'évaluation

L'objectif de cette formation étant essentiellement de fournir une synthèse des méthodes et technologies existantes, il n'est pas nécessaire d'avoir recours à une évaluation des acquis.

## 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

# Master Data Management, la gouvernance des données

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

Les enjeux concurrentiels et la nécessité accrue de transparence sur les données impose de disposer d'une méthodologie et d'une architecture de données maîtrisées et alignées sur les métiers. Ce cours vous présentera les enjeux et les méthodes afin d'engager tous les Systèmes d'Information dans cette démarche.

## OBJECTIFS PEDAGOGIQUES

Comprendre le rôle stratégique de la gestion des données pour l'entreprise.  
Comprendre les principes d'architecture des données  
Mettre en place une méthode de gouvernance  
Intégrer la gestion des Master Data dans la démarche  
Savoir identifier les acteurs du MDM et leur positionnement

### 1) Introduction

### 2) La gouvernance des données

### 3) Le Master Data Management

### 4) Cycle de vie des données

### 5) Présentation des offres de MDM

## 1) Introduction

- Enjeux stratégiques des données pour l'entreprise.
- Définition des notions "Donnée" et "Information".
- Les différentes sources de données de l'entreprise.
- Les différentes formes d'exploitation des données.
- Système d'information opérationnel et décisionnel.
- Architectures : Relationnelle/Big Data.

### Echanges

*Echanges sur le rôle stratégique des données pour l'entreprise.*

## 2) La gouvernance des données

- Définition et enjeux de la gouvernance des données.
- L'approche COBIT dans la gouvernance des données.
- Les acteurs de la gouvernance des données.
- Outils de mesure pour l'évaluation de la maturité d'une entreprise.
- Les grands principes de la démarche de gouvernance de données.
- Références et état de l'art.

### Echanges

*Une organisation met en place une démarche de gouvernance des données à l'occasion de la fusion de son SI avec un nouveau SI provenant d'une acquisition. Etude de l'impact sur l'organisation et les SI*

## 3) Le Master Data Management

- Positionnement du Master Data Management dans le Système d'Information d'entreprise.
- Les étapes essentielles de la démarche Master Data Management.
- Présentation des architectures Master Data Management.
- L'administration des données référentielles (Master Data).
- Audit et gestion de la qualité des données.
- Synthèse des bonnes pratiques.
- Le rôle des utilisateurs dans le dispositif MDM.

### Réflexion collective

*Etude de cas d'un SI dépourvu de gestion de Master Data sur lequel on effectue une analyse des besoins en MD. Comparaison de deux approches possibles de mise en œuvre.*

## 4) Cycle de vie des données

- Typologie et volume de données.
- Archivage de base de données.
- Confidentialité des données.
- Gestion des données de test.
- Décommissionnement d'applications.

### Echanges

*Echanges sur la gestion des données et leur cycle de vie.*

## 5) Présentation des offres de MDM

- Les tendances du marché
- MDM dans les suites Informatique /IBM/SAP.
- MDM Microsoft.
- Talend MDM.

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.

- Conclusion.

### **Echanges**

*Echanges sur les stratégies d'approches possibles selon les contextes.*

Stage pratique de 2 jour(s)  
Réf : TDQ

## Participants

Analystes métiers,  
intégrateurs de données,  
gestionnaires de données.

## Pré-requis

Bonnes connaissances  
des bases de données  
relationnelles et du SQL.  
Connaissances de base de  
Talend Open Studio for Data  
Integration.

Prix 2018 : 1470€ HT

## Dates des sessions

PARIS

11 oct. 2018, 22 nov. 2018

## 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.

# Talend Open Studio for Data Quality, gérer la qualité des données

> [Big Data](#), [BI](#), [NoSQL](#), [SGBD](#) > [Data](#), [état de l'art et méthodes](#)

Ce cours vous apprendra à exploiter efficacement l'outil Talend Open Studio for Data Quality de manière à évaluer le niveau de qualité des données d'un système d'information. Vous mettrez en œuvre des analyses, vérifierez des règles métiers et définirez des stratégies de correction pour les données erronées.

## OBJECTIFS PEDAGOGIQUES

Se connecter aux sources de données, produire des statistiques, identifier les données à profiler  
Choisir les différents types d'indicateurs et d'analyses adaptés aux données à contrôler  
Mettre en œuvre des analyses complexes permettant de vérifier des règles métiers  
Définir des stratégies de correction des données erronées via des jobs de Talend Data Integration

- 1) [La problématique de la qualité des données](#)
- 2) [Les concepts fondamentaux de TOS for Data Quality](#)
- 3) [Les analyses simples](#)
- 4) [Les analyses avancées](#)
- 5) [Eléments avancés](#)

Méthodes pédagogiques

70 % du temps est consacré à la pratique de l'outil. Chaque participant possède son poste de travail.

## 1) La problématique de la qualité des données

- L'évaluation de la qualité des données d'un système d'information.
- Critères fondamentaux : exhaustivité, précision et intégrité des données.
- Positionnement du produit Talend Open Studio for Data Quality dans la suite Talend.

### Travaux pratiques

Installation du produit, configuration des préférences.

## 2) Les concepts fondamentaux de TOS for Data Quality

- Les métadonnées : connexions aux bases, fichiers délimités et fichiers Excel.
- Présentation des différents types d'analyses.
- Les outils et indicateurs d'aide à la réalisation des analyses.
- L'explorateur de données.

### Travaux pratiques

Effectuer une première analyse de colonne sur des données issues d'un fichier csv, exploitation des résultats obtenus.

## 3) Les analyses simples

- Recherche de doublons, respect de contraintes d'intervalle, format de date, d'email...
- Métriques d'une table, dépendances fonctionnelles entre colonnes.
- Identification des redondances de valeurs.
- Contrôles de cohérence entre clefs étrangères et primaires.
- Utiliser les indicateurs, les modèles, les règles et les fichiers sources.

### Travaux pratiques

Réaliser une analyse de chaque type sur un jeu de données partiellement erronées.

## 4) Les analyses avancées

- Analyse de schéma et de la structure des tables via l'explorateur de données.
- Analyse multitable et multicolonne, respect de règles métiers.
- Recherche et visualisation de corrélation entre colonnes.
- Créer ses propres indicateurs et fichiers sources.
- Gérer les analyses.

### Travaux pratiques

Créer une règle métier complexe impliquant plusieurs tables et l'associer à une tâche. Publier la règle dans la forge Talend.

## 5) Eléments avancés

- Utiliser des variables de contexte.
- Créer des modèles basés sur des expressions régulières.
- Exporter/importer les analyses et les données analysées.
- Corriger les données erronées avec Talend Data Integration.

### Travaux pratiques



- 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.

*Paramétrer les métadonnées et les analyses à l'aide de variables de contexte, exporter les données analysées pour les corriger dans Talend Data Integration.*

Séminaire de 2 jour(s)  
Réf : BGA

## Participants

DSI, directeurs techniques, chefs de projet, architectes, responsables SI.

## Pré-requis

Connaissances de base des architectures techniques.

Prix 2018 : 1910€ HT

## Dates des sessions

### BRUXELLES

04 oct. 2018, 10 déc. 2018

### GENEVE

04 oct. 2018, 10 déc. 2018

### LUXEMBOURG

04 oct. 2018, 10 déc. 2018

### PARIS

06 sep. 2018, 08 nov. 2018  
10 déc. 2018

## Modalités d'évaluation

Les apports théoriques et les panoramas des techniques et outils ne nécessitent pas d'avoir recours à une évaluation des acquis.

## 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

# Big Data, état de l'art

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

*L'accroissement continu des données numériques dans les entreprises et les organismes publics a conduit à l'émergence du Big Data. Ce concept recouvre les questions de stockage et de conservation de vastes quantités de données, mais aussi celles liées au gisement potentiel de valeur que représentent ces masses de données. Ce séminaire présente les problèmes spécifiques du Big Data et les solutions techniques potentielles, de la gestion des données aux différents types de traitement, qui témoignent d'une rupture par rapport aux moyens usuels d'analyse du fait des quantités traitées.*

## OBJECTIFS PEDAGOGIQUES

Découvrir les principaux concepts du Big Data  
Identifier les enjeux économiques  
Evaluer les avantages et les inconvénients du Big Data  
Comprendre les principaux problèmes et les solutions potentielles  
Identifier les principales méthodes et champs d'application du Big Data

### 1) Introduction

#### 2) Big Data : traitements depuis l'acquisition jusqu'au résultat

#### 3) Relations entre Cloud et Big Data

#### 4) Introduction à l'Open Data

### 5) Matériel pour les architectures de stockage

#### 6) Protection des données

#### 7) Méthodes de traitement et champs d'application

#### 8) Cas d'usage à travers des exemples et conclusion

## 1) Introduction

- Les origines du Big Data : un monde de données numériques, l'e-santé, chronologie.
- Une définition par les quatre V : la provenance des données.
- Une rupture : changements de quantité, de qualité, d'habitudes.
- La valeur de la donnée : un changement d'importance.
- La donnée en tant que matière première.
- Le quatrième paradigme de la découverte scientifique.

## 2) Big Data : traitements depuis l'acquisition jusqu'au résultat

- L'enchaînement des opérations. L'acquisition.
- Le recueil des données : crawling, scraping.
- La gestion de flux événementiel (Complex Event Processing, CEP).
- L'indexation du flux entrant.
- L'intégration avec les anciennes données.
- La qualité des données : un cinquième V ?
- Les différents types de traitement : recherche, apprentissage (machine learning, transactionnel, data mining).
- D'autres modèles d'enchaînement : Amazon, e-Santé.
- Un ou plusieurs gisements de données ? De Hadoop à l'in-memory.
- De l'analyse de tonalité à la découverte de connaissances.

## 3) Relations entre Cloud et Big Data

- Le modèle d'architecture des Clouds publics et privés.
- Les services XaaS.
- Les objectifs et avantages des architectures Cloud.
- Les infrastructures.
- Les égalités et les différences entre Cloud et Big Data.
- Les Clouds de stockage.
- Classification, sécurité et confidentialité des données.
- La structure comme critère de classification : non structurée, structurée, semi-structurée.
- Classification selon le cycle de vie : données temporaires ou permanentes, archives actives.
- Difficultés en matière de sécurité : augmentation des volumétries, la distribution.
- Les solutions potentielles.

## 4) Introduction à l'Open Data

- La philosophie des données ouvertes et les objectifs.
- La libération des données publiques.
- Les difficultés de la mise en œuvre.
- Les caractéristiques essentielles des données ouvertes.
- Les domaines d'application. Les bénéfices escomptés.

## 5) Matériel pour les architectures de stockage

- Les serveurs, disques, réseau et l'usage des disques SSD, l'importance de l'infrastructure réseau.

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.

- Les architectures Cloud et les architectures plus traditionnelles.
- Les avantages et les difficultés.
- Le TCO. La consommation électrique : serveurs (IPNM), disques (MAID).
- Le stockage objet : principe et avantages.
- Le stockage objet par rapport aux stockages traditionnels NAS et SAN.
- L'architecture logicielle.
- Niveaux d'implantation de la gestion du stockage.
- Le "Software Defined Storage".
- Architecture centralisée (Hadoop File System).
- L'architecture Peer-to-Peer et l'architecture mixte.
- Les interfaces et connecteurs : S3, CDMI, FUSE, etc.
- Avenir des autres stockages (NAS, SAN) par rapport au stockage objet.

## 6) Protection des données

- La conservation dans le temps face aux accroissements de volumétrie.
- La sauvegarde, en ligne ou locale ?
- L'archive traditionnelle et l'archive active.
- Les liens avec la gestion de hiérarchie de stockage : avenir des bandes magnétiques.
- La réplication multisites.
- La dégradation des supports de stockage.

## 7) Méthodes de traitement et champs d'application

- Classification des méthodes d'analyse selon le volume des données et la puissance des traitements.
- Hadoop : le modèle de traitement Map Reduce.
- L'écosystème Hadoop : Hive, Pig. Les difficultés d'Hadoop.
- OpenStack et le gestionnaire de données Ceph.
- Le Complex Event Processing : un exemple ? Storm.
- Du BI au Big Data.
- Le décisionnel et le transactionnel renouvelés : les bases de données NoSQL. Typologie et exemples.
- L'ingestion de données et l'indexation. Deux exemples : splunk et Logstash.
- Les crawlers Open Source.
- Recherche et analyse : Elasticsearch.
- L'apprentissage : Mahout. In-memory.
- Visualisation : temps réel ou non, sur le Cloud (Bime), comparaison QlikView, Tibco Spotfire, Tableau.
- Une architecture générale du data mining via le Big Data.

## 8) Cas d'usage à travers des exemples et conclusion

- L'anticipation : besoins des utilisateurs dans les entreprises, maintenance des équipements.
- La sécurité : des personnes, détection de fraude (postale, taxes), le réseau.
- La recommandation. Analyses marketing et analyses d'impact.
- Analyses de parcours. Distribution de contenu vidéo.
- Big Data pour l'industrie automobile ? Pour l'industrie pétrolière ?
- Faut-il se lancer dans un projet Big Data ?
- Quel avenir pour les données ?
- Gouvernance du stockage des données : rôle et recommandations, le Data Scientist, les compétences d'un projet Big Data.

## Participants

Direction informatique et fonctionnelle. Responsable informatique, chef de projet, architecte, développeur.

## Pré-requis

Connaissances de base des architectures techniques et du management SI.  
Connaissances de bases de données.

Prix 2018 : 1910€ HT

## Dates des sessions

### BRUXELLES

29 nov. 2018

### GENEVE

29 nov. 2018

### LUXEMBOURG

29 nov. 2018

### PARIS

20 sep. 2018, 29 nov. 2018

## Modalités d'évaluation

Les apports théoriques et les panoramas des techniques et outils ne nécessitent pas d'avoir recours à une évaluation des acquis.

## 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

Les bases de données NoSQL proposent une nouvelle approche répondant à des besoins de volumétrie et de nouveaux types de données. Ce séminaire présente les raisons qui ont mené à ces bases, leur architecture, leurs bénéfices et leurs limites ainsi que les produits existants.

## OBJECTIFS PEDAGOGIQUES

Identifier les différences des BDD SQL et des BDD NoSQL  
Évaluer les avantages et les inconvénients inhérents aux technologies NoSQL  
Analyser les principales solutions du monde NoSQL  
Identifier les champs d'application des BDD NoSQL  
Comprendre les différentes architectures, les modèles de données, les implémentations techniques.

### 1) Introduction au NoSQL

### 2) Le Relationnel et le NoSQL.

### 3) Les mondes du NoSQL

### 4) NoSQL et Big Data

### 5) Choisir et mettre en place

## 1) Introduction au NoSQL

- L'historique du mouvement NoSQL.
- Les différentes approches de gestion de BDD à travers le temps : hiérarchiques, relationnelles, objets, XML, NoSQL.
- Les grands acteurs à l'origine du mouvement NoSQL et du Big Data analytique : Google et Amazon.
- Données structurées, semi-structurées et non-structurées : provenance, typologie...
- Vue synoptique des différents types de moteurs NoSQL du point de vue du modèle de données.
- Le NoSQL, le Big Data et les architectures Cloud: principes d'architecture communs et divergents.
- Les modes de distribution : avec maître et décentralisé.
- Le positionnement du NoSQL au sein du Big Analytics : de l'ère de la transaction à l'ère de l'interaction.

### Réflexion collective

*Pourquoi le NoSQL ? et pourquoi son succès ? Les besoins, l'évolution des architectures, la distribution et l'élasticité, le commodity hardware, quelques scénarios d'utilisation.*

## 2) Le Relationnel et le NoSQL.

- Les bases de données relationnelles : leurs forces et leurs limites.
- Structuration forte des données (schéma explicite) versus Structure souple (schéma implicite) et la modélisation Agile.
- Des qualités ACID aux qualités BASE.
- Théorème CAP (cohérence, disponibilité, tolérance au partitionnement).
- Les différents niveaux de cohérence.
- Le langage SQL, la performance des jointures. L'accès par la clé en NoSQL.
- L'évolution vers le distribué : extensibilité verticale et horizontale.
- Comprendre le NoSQL par le modèle de l'agrégat et de la centralité de la donnée.

### Réflexion collective

*Comment choisir ? (le modèle de l'agrégat versus le modèle relationnel), comment gérer l'interopérabilité ?*

## 3) Les mondes du NoSQL

- Monde du NoSQL à travers ses choix techniques et différentes bases NoSQL libres (du moins structuré au plus structuré).
- L'architecture distribuée : principes, le shared-nothing.
- Disponibilité et cohérence différée : gossip, timestamps, vector clock, règle de majorité, arbre de Merkle.
- Les patterns et les modèles. Comment modéliser et travailler efficacement en NoSQL.
- Différents modèles de connexion avec le client : sur cluster décentralisé, protocoles comme Thrift ou ProtoBuf, REST...
- Les bases orientées clé-valeur et en mémoire : Redis, Riak, Projet Voldemort, Aerospike.
- Les bases orientées documents : Le format JSON. Couchbase Server, MongoDB, Elasticsearch.
- Les bases orientées colonne distribuées pour le Big Data opérationnel : Hadoop, Hbase, Cassandra, Accumulo...
- Les moteurs orientés graphes : Neo4j, OrientDB...
- L'écosystème Hadoop : les différences avec les SGBDR, les relations avec le NoSQL.

### Démonstration

*Démonstrations techniques, du point de vue de développement, de la mise en œuvre et de l'administration, des principaux moteurs NoSQL libres.*

## 4) NoSQL et Big Data

- Big Data analytique : l'écosystème Hadoop.
- Stockage et traitements. Les différentes formes de stockage dans HDFS : SequenceFile, Apache Parquet.

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.

- Différents types de traitements : MapReduce, Graphe orienté acyclique, flux, Machine Learning, de graphes distribué...

- Les architectures Big Data : traitements batch, micro-batch, flux. Architecture Lambda, architecture Kappa.

- Les outils intégrés : Apache Spark, Apache Flink.

- Les outils d'analyses de données : développement pour l'analyse Big Data, les outils du data scientist.

- La pérennité des outils du Big Data analytique devant l'évolution très rapide des frameworks.

#### **Démonstration**

*Démonstrations de l'usage d'une plate-forme intégrée de Big Data analytique comme Apache Spark.*

### **5) Choisir et mettre en place**

- Le type d'utilisation propice au NoSQL. À quels usages correspondent les bases NoSQL.

- Les choix matériels.

- Qu'est-ce qu'un modèle de données NoSQL ?

- Comment aborder la migration ?

- Les impacts sur le développement client. Comment développer efficacement avec des bases NoSQL ?

- Quels outils de supervision et comment les choisir ?

- Quelle est la complexité administrative et la courbe d'apprentissage ?

- Cas d'utilisation dans des entreprises existantes.

- Et les performances ? Quelques benchmarks ?

- Qu'est-ce que NewSQL ?

#### **Réflexion collective**

*Quel est l'intérêt de déployer des moteurs NoSQL dans le contexte de chacun, et quel moteur NoSQL choisir ?*

Séminaire de 2 jour(s)  
Réf : SAN

## Participants

Chef de projet, architecte, responsable de production, storage manager.

## Pré-requis

Connaissances de base des problématiques de stockage de données en entreprise.

Prix 2018 : 1910€ HT

## Dates des sessions

PARIS

20 sep. 2018, 29 nov. 2018

## Modalités d'évaluation

Les apports théoriques et les panoramas des techniques et outils ne nécessitent pas d'avoir recours à une évaluation des acquis.

## 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

# Réseaux de stockage, SAN/NAS

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

Ce séminaire présente les principes fondamentaux nécessaires à la mise en place et à l'administration d'un réseau de stockage. Il aborde les points clés de toutes les phases d'un projet de mise en œuvre d'un SAN/NAS : recueil des besoins en stockage des applications de l'entreprise, choix d'un fournisseur, définition de l'architecture, migration des applications sur le SAN, et enfin administration. Les principaux aspects du cours concernent aussi bien les aspects techniques qu'organisationnels. Les évolutions technologiques sont évaluées du point de vue de leur utilisation par l'entreprise.

## OBJECTIFS PEDAGOGIQUES

Découvrir les points clés de la gestion d'un réseau de stockage  
Comprendre les différents composants d'un réseau de stockage  
Définir un type de stockage pour une application déterminée  
Gérer un projet de mise en place d'un SAN/NAS  
Organiser l'administration d'un réseau de stockage

### 1) Quel stockage pour quelle application ?

### 2) Les composants d'un réseau de stockage

### 3) Les composants matériels d'un réseau de stockage

### 4) Composants logiciels d'un réseau de stockage

### 5) Protocole Fibre Channel

### 6) Mise en œuvre

### 7) Organisation de l'administration

## 1) Quel stockage pour quelle application ?

- Le stockage vu comme un nouveau paradigme.
- Les besoins en stockage des applications.
- En quoi le SAN et le NAS sont-ils différents l'un de l'autre ?
- Quelles sont les applications candidates à la migration sur le SAN ?
- Quelles sont les applications destinées au NAS ?
- Les données des SGBDR sont-elles intégrées sur un NAS ?
- Quelles applications pour le SAN et pour le NAS ?
- Avantages et inconvénients du SAN et NAS.
- Qu'est-ce qu'un ROI, un TCO ?
- Peut-on justifier le coût d'un réseau de stockage ?

## 2) Les composants d'un réseau de stockage

- Les bases du protocole Fibre Channel.
- Pourquoi passer du SCSI vers la Fibre Channel ?
- Description des limitations du SCSI.
- En quoi la Fibre channel représente-t-elle un progrès par rapport au SCSI ?
- Terminologie Fibre Channel.
- Choix et construction d'un réseau de stockage ISCSI.
- Evolution des matériels Ethernet vers le 10 Gb, architecture spécifique et configuration pour l'ISCSI.
- L'écart prix/performance entre Fibre Channel et ISCSI selon les architectures et matériel utilisés.
- Architectures de sauvegarde : client de sauvegarde traditionnel. Lanfree et Serverfree Backup.

## 3) Les composants matériels d'un réseau de stockage

- Câbles : description des différents types monomodes et multimodes, et distances maximales.
- GBICS : les convertisseurs optiques et la nouvelle génération (SFP).
- HBA : les Host Bus Adapters et le double chemin d'accès permettant le fail-over, load balancing.
- Hub : bande passante partagée, mode 'bypass'. Topologies à base de Hubs.
- Switch : bande passante agrégée. Protocoles d'optimisation de chemin FSPF et limites. Monitorer les inter-switch links.
- Topologies réseaux/boucle arbitrée et switched Fabric.
- Pourquoi choisir un directeur plutôt que de nombreux switches départementaux ?
- Baie de stockage : comparaison des principales architectures internes, grands principes de fonctionnement.
- En quoi les dernières évolutions répondent-elles au besoin des applications ?
- Tête de NAS : caractéristiques principales d'une tête de NAS, architecture de convergence SAN/NAS.

## 4) Composants logiciels d'un réseau de stockage

- Business Continance Volume : mode de fonctionnement, contraintes de cohérence des données, types d'utilisation.
- Adéquation des applications avec les BCV.
- Snapshots : fonctionnement, avantages et inconvénients.
- Snapclones : description du fonctionnement.
- Continuous Data Protection : gestion du retour à un état antérieur des volumes.

ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

- Réplication de données entre baies : mécanismes en mode synchrone/asynchrone, choix dans les environnements.
- Jusqu'à quelle distance peut-on répliquer les données ?
- Haute disponibilité. Déduplication des données. Impacts sur les performances des applications.
- Nouvelles technologies asynchrones. Virtualisation du stockage.
- Ajouter la flexibilité et la mobilité de vos données sur un site ou entre sites.

## 5) Protocole Fibre Channel

- Couches Fibre Channel : liste de leurs fonctions.
- Classes de service : notion de Buffer to Buffer, crédit et principales utilisations.
- Convention de nommage et d'adressage (WWN).
- Conversion en adresse interne.
- Boucle arbitrée : description de la topologie et de la procédure d'arbitrage.
- Fabrique : description d'un réseau logique.
- Comment instaurer la redondance dans le SAN ?
- Norme FCoE (Fibre Channel over Ethernet), adaptateur CNA (Converged Network Adapter).

## 6) Mise en œuvre

- L'inventaire et le recueil des besoins.
- Les critères pour la sélection des machines candidates.
- Evaluer la volumétrie nécessaire.
- Faut-il calculer la charge en IO ? Jusqu'où doit-on détailler les composants de l'architecture ?
- Le cahier des charges : les éléments à intégrer.
- Le choix du fournisseur : les pièges à éviter.
- Démarche pour définir l'architecture du SAN/NAS.
- La migration : les différentes options pour la migration.

## 7) Organisation de l'administration

- Administration in-band et outband : administration en dehors du réseau de stockage et dans le Fibre Channel.
- Dispositifs d'alerte : Snmp, Http, Telnet, call-home.
- Quand faut-il mettre en place un système de surveillance ? Les problèmes de sécurité.
- Ajout de serveurs ou de volumes : Zoning, Lun Masking, Persistent Binding.
- Monitorer les performances du SAN.
- Mener un audit de performances, éléments clés à surveiller en fonction des composants.
- Modes de répartition des données. Analyse des conséquences de la baisse de densité d'accès.
- Notions de qualité de service.
- Constituer une équipe de Storage Managers : définir les processus et les procédures, Service Level Objectives.
- La gestion du stockage par attribut et les différents niveaux de service.

Cours de synthèse de 2  
jour(s)

Réf : OSC

## Participants

Directions informatique et  
fonctionnelle. Responsables  
informatiques, chefs de projet,  
architectes, développeurs.

## Pré-requis

Connaissances de base des  
architectures techniques  
et du management SI.  
Connaissances de bases de  
données.

Prix 2018 : 1770€ HT

## Dates des sessions

### AIX

15 nov. 2018

### ANGERS

20 déc. 2018

### BORDEAUX

27 sep. 2018

### DIJON

20 déc. 2018

### GRENOBLE

27 sep. 2018

### LILLE

15 nov. 2018

### LIMOGES

20 déc. 2018

### LYON

27 sep. 2018

### MONTPELLIER

15 nov. 2018

### NANCY

20 déc. 2018

### NANTES

15 nov. 2018

### ORLEANS

20 déc. 2018

### PARIS

27 sep. 2018, 29 nov. 2018

### RENNES

15 nov. 2018

### ROUEN

20 déc. 2018

### SOPHIA-ANTIPOLIS

15 nov. 2018

### STRASBOURG

15 nov. 2018

### TOULOUSE

27 sep. 2018

### TOURS

20 déc. 2018

## Modalités d'évaluation

L'objectif de cette formation  
étant essentiellement de  
fournir une synthèse des  
méthodes et technologies  
existantes, il n'est pas

# Open Source, base de données

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

Ce cours de synthèse présentera les différentes catégories de bases de données Open Source : les SQL et les NoSQL. Vous comprendrez les raisons de l'évolution de ces bases, vous saurez identifier leurs forces et leurs faiblesses et les mettre en perspective avec les solutions propriétaires.

## OBJECTIFS PEDAGOGIQUES

Identifier les différences entre les SGBD Open Source SQL et NoSQL, et les SGBDR propriétaires. Evaluer les caractéristiques, les forces et les faiblesses des différentes catégories de SGBD. S'approprier les principales solutions SGBD Open Source. Identifier les champs d'applications en fonction du type de SGBD Open Source SQL et NoSQL, et les SGBDR propriétaires. Comprendre les différentes architectures, les modèles de données, les implémentations techniques.

### [1\) Introduction aux bases de données Open Source](#)

### [2\) Logique des bases de données relationnelles](#)

### [3\) Les SGBDR Open Source](#)

### [4\) SGBD NoSQL versus SGBDR](#)

### [5\) Les SGBD NoSQL Open Source](#)

### [6\) Solutions alternatives propriétaires](#)

## 1) Introduction aux bases de données Open Source

- L'historique du modèle relationnel et des systèmes de gestion de bases de données relationnelles (SGBDR).
- Les premiers SGBDR et les principaux acteurs.
- Un changement d'importance : le Big Data.
- Une nouvelle approche de SGBD : le NoSQL.
- L'importance de l'Open Source dans les SGBD.

## 2) Logique des bases de données relationnelles

- Rappel des théories sous-jacentes aux bases de données relationnelles.
- Histoire des modèles de données.
- Typologie des SGBD et objectifs des SGBD relationnels.
- Fonctionnalités : règles de Codd, transactions, indexation, traitement côté serveur.
- Architectures avancées : réplication, haute disponibilité, distribution.

### Démonstration

*Fonctionnalités principales des SGBDR, mise en place d'une réplication et de la haute disponibilité avec MySQL*

## 3) Les SGBDR Open Source

- Principales solutions de SGBDR Open Source du marché.
- MySQL, SGBDR rapide et robuste.
- PostgreSQL, SGBDR Open Source le plus complet.
- MariaDB, transfuge de MySQL.
- Forces et faiblesses de ces solutions : architectures, stockage, performances, typage, rigidité et résilience.

### Démonstration

*Illustration des différences fonctionnelles entre MySQL, PostgreSQL et MariaDB, en matière de syntaxe SQL et de possibilité de supervision.*

## 4) SGBD NoSQL versus SGBDR

- La structuration forte des données (schéma explicite) versus structure souple (schéma implicite) et la modélisation Agile.
- Des qualités ACID aux qualités BASE.
- Le théorème CAP.
- Les langages déclaratifs de requête.
- Le modèle de l'agrégat et de la centralité des données.

### Démonstration

*Présentation des méthodes de modélisation pour le relationnel et le NoSQL (Cassandra).*

## 5) Les SGBD NoSQL Open Source

- Vue des différents types de moteurs NoSQL du point de vue du modèle de données.
- Bases orientées "clé/valeur" : Redis, Riak, etc.
- Bases orientées "colonne" : Cassandra, HBase, etc.
- Bases orientées "document" : MongoDB, Couchbase, Elasticsearch, etc.
- Bases orientées "graphe" : Neo4J, etc.
- Le NoSQL, le Big Data et les architectures Cloud : principes d'architectures communs et divergents.



nécessaire d'avoir recours à une évaluation des acquis.

## 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.

- Les offres purement Cloud : Azure, Amazon RDS, Google, BigTable.
- Positionnement du NoSQL au sein du Big Data Analytics.

### Démonstration

*Démonstrations de MongoDB, Redis, Couchbase et Cassandra.*

## 6) Solutions alternatives propriétaires

- Principaux SGBDR propriétaires et leurs caractéristiques (performance, fiabilité...).
- Les différences : langage de commande, contrôle de transactions, organisation des objets de BDD, etc.
- Modèle économique : licence, maintenance, support, sécurité, fonctions avancées...
- Propriétaire ou Open Source : une rivalité présentée en chiffres.
- Champs d'applications SGBDR Open Source versus propriétaires.

Stage pratique de 3 jour(s)  
Réf : SIF

## Participants

Chargé de reporting ou d'analyse, assistant(e), toute personne ayant des besoins d'interrogation ou de mises à jour simples d'une base de données avec le langage SQL.

## Pré-requis

Aucune connaissance particulière.

Prix 2018 : 1270€ HT

## Dates des sessions

### AIX

24 sep. 2018, 26 nov. 2018

### ANGERS

24 sep. 2018, 26 nov. 2018

### BORDEAUX

10 sep. 2018, 03 déc. 2018

### BRUXELLES

03 déc. 2018

### DIJON

10 sep. 2018, 03 déc. 2018

### GENEVE

03 déc. 2018

### GRENOBLE

10 sep. 2018, 03 déc. 2018

### LILLE

24 sep. 2018, 26 nov. 2018

### LIMOGES

24 sep. 2018, 26 nov. 2018

### LUXEMBOURG

03 déc. 2018

### LYON

10 sep. 2018, 03 déc. 2018

### MONTPELLIER

24 sep. 2018, 26 nov. 2018

### NANCY

24 sep. 2018, 26 nov. 2018

### NANTES

24 sep. 2018, 26 nov. 2018

### ORLEANS

24 sep. 2018, 26 nov. 2018

### PARIS

20 août. 2018, 03&17 sep. 2018

01&15 oct. 2018, 05&19 nov. 2018

03&17 déc. 2018

### RENNES

24 sep. 2018, 26 nov. 2018

### ROUEN

24 sep. 2018, 26 nov. 2018

### SOPHIA-ANTIPOLIS

24 sep. 2018, 26 nov. 2018

### STRASBOURG

24 sep. 2018, 26 nov. 2018

### TOULOUSE

10 sep. 2018, 03 déc. 2018

### TOURS

26 nov. 2018

# Bases de données et langage SQL pour non-informaticiens

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

*Cette formation "découverte" vous permettra de comprendre les bases de données dites relationnelles et leurs principes de fonctionnement. Vous manipulerez le langage SQL pour interroger les données d'une base. Vous vous familiariserez également avec des requêtes plus avancées pour analyser les informations.*

## OBJECTIFS PEDAGOGIQUES

Comprendre le principe et le contenu d'une base de données relationnelle  
Créer des requêtes pour extraire des données suivant différents critères  
Réaliser des requêtes avec des jointures, pour restituer les informations de plusieurs tables  
Utiliser des calculs simples et des agrégations de données  
Combiner les résultats de plusieurs requêtes

### 1) Introduction aux bases de données

### 2) Extraire les données d'une table

### 3) Interroger les données de plusieurs tables

### 4) Ordonnement et statistiques

### 5) Présenter et trier les données

### 6) Utiliser des sous-requêtes

## Méthodes pédagogiques

*Nombreux exercices progressifs d'extraction de données sur base exemple. Formation commune à toutes les bases relationnelles (Oracle, SQL Server, DB2, PostGreSQL, MySQL, Access, SQL Lite...).*

## 1) Introduction aux bases de données

- Qu'est-ce qu'une base et un serveur de base de données ?
- Lire un modèle relationnel.
- Composition d'une table. Notion de colonne et de types.
- Clé primaire et unicité.
- Notion d'intégrité référentielle.
- Outil pour interroger une base.

### Exercice

*Investigation dans la base à la recherche de tables, colonnes et clés.*

## 2) Extraire les données d'une table

- Qu'est-ce qu'une requête d'extraction ?
- Lister les valeurs à retourner.
- La clause WHERE pour filtrer les données.
- L'absence de valeur (marqueur NULL).
- Retourner des lignes sans doublon (DISTINCT).
- Opérateurs de restriction (BETWEEN, IN, LIKE...).

### Exercice

*Interrogations de plusieurs tables sur des critères différents.*

## 3) Interroger les données de plusieurs tables

- Principe des jointures : restituer les informations à partir de plusieurs tables.
- Jointure interne. Jointure externe.
- La jointure "naturelle"... et ses difficultés.
- Opérateurs ensemblistes (UNION, INTERSECT...).
- Introduction aux vues : principe et intérêt.

### Exercice

*Réaliser des requêtes avec des jointures et des opérateurs ensemblistes.*

## 4) Ordonnement et statistiques

- Trouver des valeurs agrégées (MIN, MAX, AVG, SUM...).
- Calculer des agrégats relatifs avec GROUP BY.
- Filtrer les valeurs agrégées avec HAVING.
- Mélanger agrégats et détails avec OVER.
- Ordonnement des résultats avec RANK, ROW\_NUMBER.

### Exercice

*Réaliser des requêtes utilisant des calculs simples et agrégats. Sous-totaux et numération.*

## 5) Présenter et trier les données

- Présenter les données des colonnes avec des alias.
- Conversion d'un type à un autre.

## 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.

- Effectuer des choix à l'aide de l'opérateur CASE.
- Trier les données avec ORDER BY.
- Opérations sur les chaînes de caractères, sur les dates.

### Exercice

*Utilisation de fonctions permettant d'améliorer la présentation du résultat des requêtes.*

## 6) Utiliser des sous-requêtes

- Qu'est-ce qu'une sous-requête ?
- Les différentes typologies de résultats.
- Sous-requêtes liste et opérateurs IN, ANY/SOME et ALL.
- Sous-requêtes corrélées.

### Exercice

*Ecriture de requêtes incluant des sous-requêtes de différentes formes.*

Stage pratique de 4 jour(s)

Réf : CBC

## Participants

Développeurs, architectes, responsables informatiques, chefs de projets, décideurs informatiques.

## Pré-requis

Aucune connaissance particulière. La connaissance d'un langage informatique est toutefois un plus.

Prix 2018 : 2490€ HT

## Dates des sessions

### AIX

25 sep. 2018, 27 nov. 2018

### BORDEAUX

18 sep. 2018, 20 nov. 2018

### BRUXELLES

09 oct. 2018, 11 déc. 2018

### GENEVE

09 oct. 2018, 11 déc. 2018

### GRENOBLE

18 sep. 2018, 20 nov. 2018

### LILLE

25 sep. 2018, 27 nov. 2018

### LUXEMBOURG

09 oct. 2018, 11 déc. 2018

### LYON

18 sep. 2018, 20 nov. 2018

### MONTPELLIER

25 sep. 2018, 27 nov. 2018

### NANTES

25 sep. 2018, 27 nov. 2018

### PARIS

04 sep. 2018, 20 nov. 2018

### RENNES

25 sep. 2018, 27 nov. 2018

### SOPHIA-ANTIPOLIS

25 sep. 2018, 27 nov. 2018

### STRASBOURG

25 sep. 2018, 27 nov. 2018

### TOULOUSE

18 sep. 2018, 20 nov. 2018

## 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

# Bases de données, modélisation et choix techniques

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

*Cette formation apporte une synthèse complète des connaissances aujourd'hui indispensables en matière de bases de données, du processus de conception à son utilisation effective. Elle analyse les architectures et domaines d'application, évalue l'offre du marché et examine les démarches pratiques de mise en œuvre.*

## OBJECTIFS PEDAGOGIQUES

Comprendre la logique des SGBD relationnels  
Réaliser le modèle conceptuel d'une base de données à l'aide d'UML  
Passer du modèle conceptuel au modèle logique  
Comprendre la normalisation et dénormalisation d'un schéma  
Passer du modèle logique au modèle physique, et l'implémenter avec SQL

### 1) Logique des SGBD

### 2) Techniques de modélisation

### 3) Modèle logique de données

### 4) Implémentation avec SQL

### 5) Paramètres clés pour le choix d'un serveur

## 1) Logique des SGBD

- Rappel des théories sous-jacentes aux bases de données relationnelles.
- Modèles de données (hiérarchique, réseau...).
- Typologie des SGBD. Objectifs des SGBD relationnels.
- Fonctionnalités (règles de Codd, transactions, indexage, sauvegardes, restaurations).
- Les architectures.

### Travaux pratiques

*Création d'un utilisateur, attributions de droits, transaction, interrogation d'une base de données.*

## 2) Techniques de modélisation

- Le modèle conceptuel.
- Analyse des besoins.
- Concepts de base (entités, associations...).
- Dépendances fonctionnelles.
- Niveau conceptuel avec Merise et UML.
- La démarche.
- Prise en compte des règles métier et contraintes.

### Travaux pratiques

*Etudes de plusieurs cas réels, modélisations Merise et UML à l'aide d'un outil.*

## 3) Modèle logique de données

- Le modèle logique. La structure des tables et des contraintes.
- Modèle relationnel (concepts, notations, terminologie).
- Du conceptuel au logique.
- Modélisation des arborescences.
- Les formes normales (de la 1ère à la 5ème, BCNF).
- Dénormalisation.
- Calculs de volumétrie.

### Travaux pratiques

*Etude de cas réels, exercices de normalisation et dénormalisation, génération de schémas à l'aide d'un outil.*

## 4) Implémentation avec SQL

- Les transactions.
- Normes SQL (SQL1, SQL2, SQL:1999, SQL:2003).
- Du logique au physique.
- Qualité des données.
- Vues et vues matérialisées.
- Droits d'accès et rôles.
- Procédures cataloguées et déclencheurs.
- Interfaces avec les langages de programmation.

### Travaux pratiques

*Création et manipulation d'une base, Reverse Engineering, mesures de performances (démonstrations).*

## 5) Paramètres clés pour le choix d'un serveur

- Etude des paramètres des serveurs SQL.
- Composants techniques (disques et interfaces réseau).
- Modules applicatifs (outils d'administration...).

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.

- Solutions de haute disponibilité, sécurité des données.
- Critères techniques.
- Coût global d'exploitation.
- Offre du marché.

## 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.

Stage pratique de 2 jour(s)

Réf : AMC

## Participants

Concepteur et administrateur de base de données, analyste développeur.

## Pré-requis

Aucune connaissance particulière. La connaissance d'un langage informatique est toutefois un plus.

## Dates des sessions

## 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

# PowerAMC, concevoir une base de données

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

*Pendant ce stage, vous découvrirez les fonctionnalités majeures de PowerAMC. Vous apprendrez à construire, formaliser et maintenir les modèles utilisés pour concevoir une base de données (modèle orienté objet - UML, modèle conceptuel et physique, dictionnaire...).*

## OBJECTIFS PEDAGOGIQUES

Etablir un dictionnaire de données dans PowerAMC  
Implémenter et vérifier un modèle conceptuel de données dans PowerAMC  
Paramétrer les règles de transformation du modèle conceptuel au modèle physique de données  
Générer le modèle physique des données avec PowerAMC et analyser son résultat  
Pratiquer le Reverse Engineering avec PowerAMC

### 1) Introduction

### 2) Identification des données

### 3) Modélisation sémantique des données

### 4) Modélisation physique des données

### 5) Reverse Engineering

### 6) Conclusion

## Travaux pratiques

*Les travaux pratiques sont répartis pour moitié entre la conception des modèles (sur l'étude de cas) et leur représentation avec PowerAMC.*

## 1) Introduction

- Présentation de PowerAMC.
- Les fonctionnalités principales.
- L'interface et les menus.

### Travaux pratiques

*Prise en main de PowerAMC.*

## 2) Identification des données

- Apprendre à élaborer un dictionnaire des données.
- La description des données et les contrôles.
- Comment définir les attributs ?

### Travaux pratiques

*Analyse de documents et identification des données contenues. Saisir dans PowerAMC les attributs retenus et les règles associées.*

## 3) Modélisation sémantique des données

- Le modèle orienté objet (MOO - Diagramme de classe UML).
- Identification des classes, des associations, des contraintes.
- La démarche pour élaborer un modèle.
- La normalisation : le rôle des formes normales dans la compréhension des données.
- Les principaux concepts proposés par PowerAMC.
- Les vérifications proposées.
- Le Modèle Conceptuel des Données (MCD).
- Présentation des concepts du MCD par analogie et différence avec le MOO.

### Travaux pratiques

*Construire un diagramme de classe à partir du dictionnaire des données. Saisir le résultat avec PowerAMC. Effectuer les contrôles proposés. Générer le MCD avec PowerAMC et analyser le résultat.*

## 4) Modélisation physique des données

- Présentation des règles de transformation d'un modèle sémantique en modèle physique des données.
- Les principales règles d'optimisation.
- Comment PowerAMC réalise la génération d'un MPD ?
- La génération des scripts SQL correspondants.

### Travaux pratiques

*Transformer le modèle sémantique en modèle physique. Générer le MPD avec PowerAMC et analyser le résultat.*

## 5) Reverse Engineering

- Présentation.
- Le Reverse Engineering avec PowerAmc.
- Fonctions pour créer un MPD à partir d'une base existante.
- Comment générer un MOO ou un MCD à partir d'un MPD existant ?

bien assisté à la totalité de la session.

### **Travaux pratiques**

*Ajouter des nouveaux attributs à une base de données. Régénérer les modèles amont.*

## **6) Conclusion**

- Présentation des fonctions de traitement de PowerAMC.
- Les autres diagrammes.

Stage pratique de 2 jour(s)

Réf : UMD

## Participants

Architectes, chefs de projet, analystes, concepteurs, développeurs, responsables méthodes.

## Pré-requis

Connaissances de base d'UML.

## Dates des sessions

## 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.

# UML 2, modéliser et créer vos bases de données

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

Cette formation vous expliquera comment utiliser à bon escient le diagramme de classes UML afin de concevoir une base de données relationnelle normalisée. Les étapes menant d'un diagramme de classes à sa traduction en langage SQL sont étudiées à l'aide de cas réels.

## OBJECTIFS PEDAGOGIQUES

Concevoir une base de données avec UML.

Maîtriser la transformation d'un modèle de classes UML en un modèle relationnel.

Identifier et utiliser les principaux modeleurs du marché.

### 1) Les bases UML

#### 2) Analogies avec les modèles conceptuels

#### 3) Modélisation avec le diagramme de classes d'UML

### 4) Le niveau logique

#### 5) Le niveau physique

## 1) Les bases UML

- Les diagrammes d'UML.
- Du modèle entité-association à la notation UML.
- Niveaux d'abstraction couverts par UML.

## 2) Analogies avec les modèles conceptuels

- Associations binaires (de one-to-one à many-to-many).
- Identification des classes.
- Associations réflexives.
- Héritages.

### Travaux pratiques

Comparaisons avec les formalismes entité-association (Barker, IDEF1X, ER, Merise).

## 3) Modélisation avec le diagramme de classes d'UML

- Les outils du diagramme de classes.
- Ecueils des associations n-aires.
- Intérêts des classes-association.
- Historisation.
- Agrégation pour l'identification relative.
- Apports du langage OCL.
- Intérêts des rôles.

### Travaux pratiques

Etudes de cas réels, modélisation à l'aide d'outils du marché (Power AMC, MEGA, Objecteering, Rational Rose, Win'Design).

## 4) Le niveau logique

- Mappings des classes et des associations (déduction des clés).
- Cas des agrégations et des héritages.
- Identification alternative.
- Utilisation du diagramme de classes.
- Profil UML.
- Normalisation d'un modèle logique.

### Travaux pratiques

Transformation de modèles conceptuels en modèles relationnels. Validation de schémas (normalisation). Utilisation des outils du marché.

## 5) Le niveau physique

- Processus de génération de tables et de clés.
- Génération de scripts SQL.
- Programmation avec SQL des règles métiers (contraintes et déclencheurs).
- Reverse engineering.

### Travaux pratiques

Génération de scripts SQL pour Oracle et SQL-Server. Contraintes et indexation.



Stage pratique de 2 jour(s)

Réf : BDM

## Participants

Architecte, chef de projet, analyste, concepteur/développeur, responsable méthode.

## Pré-requis

Connaissances de base en informatique. La connaissance de SQL pour Oracle (cours réf. OSL) et des principes de modélisation (cours réf. CBC ou réf. UMD) est un plus.

## Dates des sessions

## 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

# Conception d'une base de données avec Oracle Data Modeler

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

*Oracle SQL Developer Data Modeler est une application développée en Java permettant la modélisation de bases de données SQL. Ce stage, après une introduction à la conception de bases de données, vous apprendra à installer et configurer l'outil ainsi que les différentes étapes pour modéliser vos bases de données.*

## OBJECTIFS PEDAGOGIQUES

Découvrir la modélisation des bases de données avec l'outil Oracle SQL Developer Data Modeler  
Créer des modèles Entité-Association (Entity Relationship)  
Développer ou importer des modèles relationnels  
Générer des scripts SQL (DDL)  
Gérer des versions de schémas

### [1\) Introduction à la conception de bases de données](#)

### [2\) Prise en main d'Oracle Data Modeler](#)

### [3\) Le modèle logique d'Oracle Data Modeler](#)

### [4\) Le modèle relationnel d'Oracle Data Modeler](#)

### [5\) Le modèle physique d'Oracle Data Modeler](#)

### [6\) Les étapes à suivre](#)

### [7\) Autres fonctionnalités](#)

## Travaux pratiques

*Utilisation de l'outil d'Oracle SQL Developer Data Modeler.*

## 1) Introduction à la conception de bases de données

- Les théories sous-jacentes aux bases de données relationnelles.
- Les typologie des Systèmes de Gestion de Base de Données Relationnels (SGBDR).
- Les objectifs des SGBDR.
- De la modélisation à la conception de bases de données.
- Les approches de conception (Top-Down et Bottom-Up).

### Réflexion collective

*Echanges en groupe pour comprendre ce qu'est la modélisation des données et la conception d'une base de données.*

## 2) Prise en main d'Oracle Data Modeler

- L'historique d'Oracle Data Modeler.
- Les principaux outils concurrents.
- Le processus d'installation.
- La partie configuration.
- Avantages et inconvénients.

### Exercice

*Prise en main de l'outil.*

## 3) Le modèle logique d'Oracle Data Modeler

- Le modèle logique de données.
- Les domaines.
- Les types personnalisés.
- Entités, attributs, identifiants, héritages, relations et vues.
- Les différentes notations.
- Le modèle multidimensionnel.
- Le forward et le reverse engineering.

### Exercice

*Composer un modèle logique de données. Création d'un glossaire.*

## 4) Le modèle relationnel d'Oracle Data Modeler

- Les tables.
- Les colonnes.
- L'index.
- Les clés étrangères.
- Les schémas et les sous-vues.

### Echanges

*Traduire un modèle logique. Modifier un modèle relationnel existant.*

ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

## 5) Le modèle physique d'Oracle Data Modeler

- Les liens avec le niveau relationnel.
- Les sites SGBD.
- Les séquences.
- Les espaces logiques.
- Les rôles.
- Les schémas.

### **Exercice**

*Traduire un modèle relationnel. Générer un modèle physique.*

## 6) Les étapes à suivre

- La construction d'un modèle conceptuel normalisé.
- La validation du schéma relationnel.
- La génération de scripts SQL.

### **Exercice**

*Application à un cas réel. Création du modèle logique puis création des tables.*

## 7) Autres fonctionnalités

- Les conventions de nommage (Naming Standards).
- Les importations (DDL, Data Dictionary et JDBC).
- La génération de rapports (RTF, HTML et PDF).
- La gestion de versions (Subversion et Pending Changes).

### **Exercice**

*Création d'un rapport dans Word. Importation d'un schéma.*

Séminaire de 2 jour(s)  
Réf : BMD

## Participants

Dirigeants d'entreprise, directeurs financiers, responsables métier, responsables maîtrise d'ouvrage, DSI, responsables informatiques, consultants, auditeurs, informaticiens.

## Pré-requis

Aucune connaissance particulière.

**Prix 2018 : 1910€ HT**

## Dates des sessions

### PARIS

24 sep. 2018, 13 déc. 2018

## Modalités d'évaluation

Les apports théoriques et les panoramas des techniques et outils ne nécessitent pas d'avoir recours à une évaluation des acquis.

## 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.

# Big Data, Data Science, Machine Learning, leur impact dans l'entreprise

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

*La révolution numérique est rendue possible par l'explosion du volume des données échangées au sein et en dehors des entreprises. Cette révolution a permis l'apparition de nombreux nouveaux usages et nous n'en sommes qu'au début. Sans entrer dans le détail de la technique ni des aspects architecturaux ou des modèles de données, ce séminaire vous présentera une vision d'ensemble de l'impact que peuvent avoir le Big Data, la Data Science et le Machine Learning sur les nouveaux modèles d'entreprise.*

## OBJECTIFS PEDAGOGIQUES

Découvrir les fondamentaux du Big Data, de la Data Science et du Machine Learning  
Evaluer en quoi les données prennent une position prédominante dans les nouveaux business models  
Comprendre le rôle de la Data Science au sein de l'organisation et de la gouvernance de l'information  
Présenter les enjeux des usages que représentent le machine learning, le deep learning et les impacts organisationnels

Justifier l'ouverture du SI vers l'extérieur, tant en collecte de données qu'en fourniture

### 1) Historique et contexte de l'explosion des usages autour des données

### 2) Le Big Data, la Data Science, la Machine Learning, l'IA

### 3) La gestion du patrimoine informationnel et le pilotage de la valeur

### 4) Les acteurs du Big Data et les enjeux organisationnels associés

### 5) Le Big Data étendu

### 6) Les notions pour aller plus loin et quelques pistes pour se lancer

## 1) Historique et contexte de l'explosion des usages autour des données

- La révolution numérique est en route ... poussée par un tsunami informationnel !
- Quelles sont les grandes briques fonctionnelles et techniques de la gestion du patrimoine informationnel ?
- Qu'est ce que le Big Data ? Définition et périmètre général.
- Un peu d'histoire sur les origines et l'explosion du Big Data.
- Le Big Data n'est pas une mode et s'inscrit dans la révolution numérique comme le carburant de l'innovation.
- Le Big Data : les grands usages actuels.
- Quelques grandes notions technologiques à retenir et comment surnager parmi tous ces acronymes techniques.
- Quelle cohabitation entre les différentes générations de système de gestion des données ?
- Quel futur pour toutes ces solutions et usages ? Quel est le niveau de maturité à date ?

### Echanges

*Réflexion collective sur la révolution numérique.*

## 2) Le Big Data, la Data Science, la Machine Learning, l'IA

- Créer et améliorer les facultés de l'entreprise grâce aux Big Data.
- Les grandes briques applicatives au service du Big Data dans l'entreprise.
- De l'analyse statistique à la Data Science : quels profils, pour quels bénéfices ?
- Les principales déclinaisons de l'Intelligence Artificielle.
- L'Intelligence Artificielle dans l'analyse prédictive.
- Le Machine Learning ou la capacité d'apprentissage par la machine.
- Apprendre par le traitement de masse de l'information recueillie : le Deep Learning.
- L'Intelligence Artificielle, le Machine Learning et le Deep Learning : quels enjeux pour nos sociétés ?

### Echanges

*Réflexion collective sur l'évolution de le Big Data, de la Data Science, du Machine Learning et l'Intelligence Artificielle.*

## 3) La gestion du patrimoine informationnel et le pilotage de la valeur

- La place des données dans la révolution numérique.
- La place de la donnée dans les modèles d'entreprise disruptifs.
- La valeur des données au coeur des enjeux métiers.
- Peut-on connaître et piloter la valeur de son patrimoine informationnel ?
- Quel est le risque d'infobésité ? Comment s'en prémunir ?
- Les opportunités de monétisation de l'information.
- Comment installer une culture de la donnée dans l'entreprise ?
- Le poids et le coût de la qualité des données du patrimoine géré par l'entreprise.
- L'apport de valeur par les données touche tous les métiers de nos clients.

### Etude de cas

*Etude de cas sur l'analyse de la valeur de la donnée.*

• 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.

#### 4) Les acteurs du Big Data et les enjeux organisationnels associés

- Organisation, rôles et gouvernance : les nouveaux modèles à l'heure du Big Data.
- Pourquoi les silos organisationnels sont un frein majeur à l'exploitation des données et comment y remédier ?
- La Data Science : adapter les profils de statisticiens aux nouveaux enjeux.
- Les spécialistes des nouveaux domaines et leurs méthodes de travail.
- Les thématiques de mise en œuvre de la gouvernance des données.
- Les acteurs de la gouvernance de la donnée.
- Les acteurs en charge de la qualité de la donnée.
- Quelles sont les grandes tendances réglementaires autour de la gestion des données et comment les aborder ?

##### **Démonstration**

*Démonstration de l'importance de la gouvernance des données.*

#### 5) Le Big Data étendu

- La place de l'entreprise dans son écosystème de données.
- Les données à la frontière de l'entreprise (DMP).
- Les données fournies par des partenaires : réseaux sociaux...
- L'Internet des Objets (IoT) ou comment rendre son entreprise sensible aux nouveaux usages.
- L'Open Data et ses apports dans l'analyse de données : nouvelles données, enrichissement, croisement de données.
- L'émergence des Chatbots ou l'automatisation de fonctions d'interface clients : quels impacts dans nos organisations ?
- L'expérience client augmentée, levier de la transformation digitale des entreprises.
- La fourniture de données comme nouvelle source de revenus.
- Les données, carburant de l'innovation métier.

##### **Démonstration**

*Démonstrations sur l'innovation basée sur le Big Data, la Data Science et le Machine Learning.*

#### 6) Les notions pour aller plus loin et quelques pistes pour se lancer

- Quels sont les risques à ne pas prendre en compte la gestion du patrimoine informationnel ?
- Les grandes tendances autour de la Data pour les trois années à venir.
- Les solutions technologiques : Big Data et Machine Learning.
- Mode d'emploi pour lancer une initiative Big Data.
- Les techniques du Big Data : tour d'horizon.
- Les métiers, fonctions et rôles autour de la donnée : la synthèse et les différents niveaux de mise en œuvre.
- Pourquoi les entreprises se lancent-elles dans le Big Data ?

Séminaire de 2 jour(s)  
Réf : DMI

## Participants

Responsables Infocentre, responsables marketing, statisticiens, responsables informatiques, chefs de projets et experts décisionnels. Utilisateurs et gestionnaires métiers de bases de données.

## Pré-requis

Connaissances de base en analyse décisionnelle. Connaissances de base en statistiques.

Prix 2018 : 1910€ HT

## Dates des sessions

### PARIS

04 oct. 2018, 06 déc. 2018

## Modalités d'évaluation

Les apports théoriques et les panoramas des techniques et outils ne nécessitent pas d'avoir recours à une évaluation des acquis.

## 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

# Data Mining : synthèse

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

*Le Data Mining apporte des solutions pour l'extraction de connaissances à partir de données brutes. Ce séminaire, alliant principes théoriques et expériences pratiques, vous propose une synthèse sur le Data Mining, ses applications, ses apports pour l'entreprise, ainsi que sur les méthodes et outils les plus utilisés.*

## OBJECTIFS PEDAGOGIQUES

Comprendre les principes et la finalité du Data Mining (DM)  
Identifier les principales techniques du DM et leur cas d'utilisation  
Mettre en oeuvre sur un cas simple les méthodes de scoring et de géomarketing  
Découvrir les méthodes prédictives et les méthodes descriptives du DM  
Connaître les principales étapes d'un projet Data Mining

### [1\) Le Système d'Information Décisionnel \(SID\)](#)

### [2\) Comprendre le Data Mining \(DM\)](#)

### [3\) Les techniques du Data Mining](#)

### [4\) La méthode descriptive du Clustering](#)

### [5\) Exemples d'application du DM](#)

### [6\) Les données de l'entreprise](#)

### [7\) Méthodologie de projet Data Mining](#)

### [8\) Panorama des outils](#)

## 1) Le Système d'Information Décisionnel (SID)

- Les enjeux du SID : besoins, domaines d'application.
- Architecture type d'un SID, état de l'art.
- Elaboration des informations décisionnelles.
- Conception d'un SID : étapes, optimisation, organisation des données, dictionnaires.

## 2) Comprendre le Data Mining (DM)

- Définition et finalité du Data Mining (DM).
- Quel lien entre le DM et les statistiques, quelle dépendance entre le DM et l'informatique ?
- Différence entre DM et OLAP ?
- Les attentes des entreprises, les réponses du DM.

## 3) Les techniques du Data Mining

- Les différentes familles du DM.
- Les méthodes prédictives et les méthodes descriptives.
- Analyse factorielle, typologique. La classification...
- Les arbres de décisions, les réseaux de neurones...
- Classification des techniques de DM.

## 4) La méthode descriptive du Clustering

- Définition et méthodologie.
- Les critères pour structurer les données à classer.
- Evaluation et validation des classes obtenues.
- Les différentes sous-familles du Clustering.

### Exemple

*Présentation d'applications du Clustering.*

## 5) Exemples d'application du DM

- Le scoring : définition, finalité, méthodologie.
- Le géomarketing : définition, finalité, méthodologie.

### Exemple

*Mise en oeuvre de la méthode de scoring. Cas pratique d'utilisation de géomarketing.*

## 6) Les données de l'entreprise

- Rappel de la problématique des données du SI.
- Qualité des données et administration des données.
- Processus de collecte et d'exploration.
- Création d'agrégats et de nouvelles données.
- Transformation des données.

## 7) Méthodologie de projet Data Mining

- Définition du problème métier à résoudre et des objectifs à atteindre.
- Inventorier, décrire et classer les données.
- Concevoir et alimenter la base Data Mining.
- Explorer, segmenter des entités analysées.

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.

- Etablir et appliquer les modèles d'analyse.
- Itérer, déployer auprès des utilisateurs.
- Maintenir le modèle et le logiciel associé.

## 8) Panorama des outils

- Principaux outils du marché : SAS, R, IBM SPSS...
- Zoom sur l'outil SAS et sur l'ETL Powercenter.
- Quels critères de choix pour ce type d'outils ?

Stage pratique de 2 jour(s)

Réf : STA

## Participants

Responsables Infocentre, responsables marketing, responsables Qualité, utilisateurs et gestionnaires métiers de bases de données, futurs Data Scientist.

## Pré-requis

Connaissances de base en mathématiques et statistiques ou connaissances équivalentes à celles apportées par le stage "Statistiques descriptives, introduction" (réf. UES). Connaissances de base Excel.

Prix 2018 : 1470€ HT

## Dates des sessions

### PARIS

17 sep. 2018, 22 nov. 2018

## 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.

# Modélisation statistique

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

Ce stage présente l'essentiel de la modélisation statistique. Il vous permettra de comprendre leur rôle dans le monde de l'analyse décisionnelle, du Big Data et du Data Mining, ainsi que les mécanismes qui permettent de transformer et d'affiner des données pour en tirer des informations métiers utiles.

## OBJECTIFS PEDAGOGIQUES

Dimensionner un échantillon de population  
Calculer des paramètres de position et de dispersion (médiane, étendue, quantile, écart-type)  
Exploiter les paramètres statistiques pour comprendre une série de données  
Valider la précision d'une estimation, à l'aide des intervalles de confiance  
Découvrir des outils comme R et Excel pour la mise en œuvre des modèles étudiés

### [1\) Rappels des fondamentaux de la statistique descriptive](#)

### [2\) Démarche et modélisation d'une analyse statistique](#)

### [3\) Paramètre de position et de dispersion](#)

### [4\) Tests et intervalle de confiance](#)

### [5\) Panorama des outils](#)

## 1) Rappels des fondamentaux de la statistique descriptive

- Définition de la statistique descriptive.
- Analyse d'une population.
- Méthodes d'échantillonnage.
- Variables qualitatives et quantitatives.
- Effectifs et calcul des fréquences.
- Effectifs cumulés croissants et décroissants.
- Représentation graphique des variables qualitatives et quantitatives.

### Etude de cas

*Application pratique sur excel d'analyses statistiques et interprétation*

## 2) Démarche et modélisation d'une analyse statistique

- Statistique descriptive.
- Phase d'apprentissage.
- Statistique prédictive pour estimer et anticiper.
- Modélisation statistique d'un phénomène.

## 3) Paramètre de position et de dispersion

- Mode, valeur modale, valeur la plus probable.
- Moyenne d'une population (ou d'un échantillon).
- Médiane, partager une série numérique.
- Etendue, différence entre valeurs extrêmes.
- Utiliser les quantiles.
- Ecart-Type, calculer la dispersion d'un ensemble de données.
- Calcul de la variance et de la covariance.

### Etude de cas

*Calcul de paramètres de position et de dispersion sur différents échantillonnages et comparaisons des résultats.*

## 4) Tests et intervalle de confiance

- Lois statistiques et intervalle de confiance.
- Tests statistiques courants (Test de Student, Analyse de variances,  $\chi^2$ ).
- Valider la précision d'une estimation. Amplitude de l'intervalle.

### Etude de cas

*Exercices sur le logiciel R.*

## 5) Panorama des outils

- Zoom sur le logiciel Open Source "R".
- Initiation au logiciel Open Source "R".

### Travaux pratiques

*Utilisation de packages pour faire les analyses statistiques.*

- 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.



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 2018 : 3030€ HT

## Dates des sessions

### AIX

10 sep. 2018, 19 nov. 2018

### ANGERS

10 sep. 2018, 19 nov. 2018

### BORDEAUX

17 sep. 2018, 03 déc. 2018

### BRUXELLES

10 sep. 2018, 26 nov. 2018

### DIJON

17 sep. 2018, 03 déc. 2018

### GENEVE

10 sep. 2018, 26 nov. 2018

### GRENOBLE

17 sep. 2018, 03 déc. 2018

### LILLE

10 sep. 2018, 19 nov. 2018

### LIMOGES

10 sep. 2018, 19 nov. 2018

### LUXEMBOURG

10 sep. 2018, 26 nov. 2018

### LYON

17 sep. 2018, 03 déc. 2018

### MONTPELLIER

10 sep. 2018, 19 nov. 2018

### NANCY

10 sep. 2018, 19 nov. 2018

### NANTES

10 sep. 2018, 19 nov. 2018

### ORLEANS

10 sep. 2018, 19 nov. 2018

### PARIS

27 août. 2018, 10&24 sep. 2018

08&22 oct. 2018, 12&19 nov. 2018

10 déc. 2018

### RENNES

10 sep. 2018, 19 nov. 2018

### ROUEN

10 sep. 2018, 19 nov. 2018

### SOPHIA-ANTIPOLIS

10 sep. 2018, 19 nov. 2018

### STRASBOURG

10 sep. 2018, 19 nov. 2018

### TOULOUSE

17 sep. 2018, 03 déc. 2018

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

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

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.

## TOURS

10 sep. 2018, 19 nov. 2018

### 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.

- Analyse et visualisation des données.
- 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.

Cours de synthèse de 3  
jour(s)

Réf : DAW

## Participants

Responsables Infocentre,  
responsables informatiques,  
responsables des études,  
architectes de systèmes  
d'information, chefs de projets  
fonctionnels et techniques.

## Pré-requis

Bonnes connaissances en  
gestion de bases de données.  
Connaissances de base en  
analyse décisionnelle.

Prix 2018 : 2290€ HT

## Dates des sessions

PARIS

03 sep. 2018, 26 nov. 2018

## Modalités d'évaluation

L'objectif de cette formation  
étant essentiellement de  
fournir une synthèse des  
méthodes et technologies  
existantes, il n'est pas  
nécessaire d'avoir recours à  
une évaluation des acquis.

## 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

# Construire un Data Warehouse qualité des données et performance du SI décisionnel

> [Big Data](#), [BI](#), [NoSQL](#), [SGBD](#) > [Data](#), état de l'art et méthodes

*Le Data Warehouse est au cœur de l'analyse décisionnelle. Au travers d'une démarche structurée et pragmatique, vous découvrirez les meilleures méthodes et outils pour élaborer un entrepôt de données à partir des besoins métier, l'alimenter et le rendre fiable et évolutif. Vous découvrirez également quels sont les rôles clés dans ce type de projet, et quel est l'impact sur l'architecture SI et la qualité du référentiel de données d'entreprise. Un cours centré sur l'expérience pratique qui vous donnera aussi une première approche de la modélisation "en étoile".*

## OBJECTIFS PEDAGOGIQUES

Comprendre les enjeux stratégiques d'un SI décisionnel et ses avantages  
Repérer les différentes couches de l'organisation d'un système décisionnel  
Modéliser Data Warehouse au cœur du système décisionnel  
Identifier les étapes essentielles dans la démarche de construction d'un Data Warehouse  
Maîtriser les rôles et les livrables sur un projet de construction de Data Warehouse  
Avoir une vue complète et synthétique des solutions BI du marché

### 1) Le Data Warehouse, finalité et principes

### 2) L'architecture d'un SI décisionnel d'entreprise

### 3) Les principes de la modélisation Data Warehouse

### 4) La démarche de construction d'un Data Warehouse

### 5) Organisation du projet, les acteurs et les livrables

### 6) Les outils dans le domaine du décisionnel

### 7) Synthèse

## 1) Le Data Warehouse, finalité et principes

- Les enjeux stratégiques d'un SI décisionnel.
- Les raisons techniques et culturelles qui conduisent au Data Warehouse.
- Définition du Data Warehouse selon Bill Inmon.
- Les solutions apportées par l'architecture technique et fonctionnelle du Data Warehouse.
- Caractéristiques des données du SI décisionnel
- L'Infocentre et le SI décisionnel.
- Présentation des différentes approches Data Warehouse et Infocentre, leurs avantages et inconvénients.

## 2) L'architecture d'un SI décisionnel d'entreprise

- Les différentes couches de l'organisation d'un Data Warehouse.
- La collecte et l'intégration des données.
- L'Operational Data Store et la Data Staging area.
- La couche de présentation, le portail décisionnel.
- Les moteurs d'analyse OLAP : (MOLAP) et/ou relationnel OLAP (ROLAP).
- Les techniques d'analyse "data mining" : méthodes prédictives, méthodes descriptives.
- Croissance du volume et de la nature des données, les enjeux du Big Data.
- La documentation des données du DW : notions de référentiel de données.
- Comment le DW fiabilise la gestion du référentiel de données (MDM).
- La gestion des flux : la capture des données sources, les règles de transformation.

### Exemple

*Présentation d'exemples de projets d'analyse décisionnelle divers.*

## 3) Les principes de la modélisation Data Warehouse

- Les modèles relationnels opérationnels et dénormalisés.
- Les modèles hybrides.
- Les modèles génériques.
- Comprendre le modèle en étoile, sa finalité.
- Comprendre les notions de faits et d'analyse. Les hiérarchies d'axe d'analyse.
- Le modèle en flocon.
- La problématique des dimensions à évolution.
- Gestion des agrégats et de la stabilité du périmètre fonctionnel.
- Quelle approche favorise les informations de détail ou les agrégats ? Bonnes pratiques, questions à poser au métier.

### Réflexion collective

*Construction collective et enrichissement d'un modèle de données en étoile, suivant plusieurs études de cas données. Elaboration du questionnaire à proposer pour recueillir le besoin utilisateur.*

## 4) La démarche de construction d'un Data Warehouse

- Identifier le périmètre fonctionnel candidat. Déterminer l'objectif et les événements de gestion à suivre.

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.

- Estimer la volumétrie du périmètre.
- Analyse fonctionnelle, recueil des besoins utilisateurs.
- Conception de l'architecture technique détaillée.
- Etablir une démarche générique de mise en œuvre.
- Les apports d'une démarche itérative, le contenu d'une itération.
- Première itération ou projet pilote, bien le choisir. Rôle du sponsor, de la MOA, de la MOE, impact sur l'organisation.
- L'administration et le suivi de la solution opérationnelle.

#### **Echanges**

*Présentation de la démarche de conception d'un périmètre fonctionnel du Data Warehouse.*

### 5) Organisation du projet, les acteurs et les livrables

- Le rôle fondamental du sponsor ou promoteur.
- Le comité de pilotage.
- Rôle de l'équipe fonctionnelle, du groupe de projet utilisateurs : valider le design de l'environnement utilisateur.
- Le transfert de compétences vers les utilisateurs finaux par l'équipe fonctionnelle : formation et documentation.
- L'équipe technique, les architectes.
- Les principaux livrables d'un projet décisionnel.

#### **Echanges**

*Présentation des livrables et de leur responsable suivant les étapes de la démarche.*

### 6) Les outils dans le domaine du décisionnel

- Les dernières évolutions techniques des SGBDR dans le domaine du décisionnel.
- Panorama et typologie des solutions BI du marché. Les offres en mode SaaS.
- Les solutions de reporting : SSRS, IBM Cognos, SAS, BusinessObjects... La mise en œuvre des outils de requête.
- Les outils d'analyse OLAP côté serveur et côté client : utilisation, évolutivité, approche DataMart, temps de réponse.
- Les solutions d'analyse data mining : SAS Enterprise Miner, IBM, OBI data mining. Exigences et points forts.
- Les solutions ETL : IBM, Informatica, Oracle, SAP, Talend...
- Les outils de modélisation relationnelle : possibilités et limites.

#### **Exemple**

*Présentation des possibilités de divers outils BI.*

### 7) Synthèse

- Les tendances d'évolution des systèmes décisionnels.
- Bonnes pratiques pour la modélisation.
- Recommandations pour l'organisation du projet Data Warehouse.

Stage pratique de 3 jour(s)

Réf : AMT

## Participants

Maîtres d'ouvrage et maîtres d'œuvre, responsables des systèmes décisionnels, responsables informatiques, responsables des études, architectes de systèmes d'information, chefs de projets.

## Pré-requis

Connaissances de base de l'analyse décisionnelle et des SGBD relationnelles.

Prix 2018 : 1660€ HT

## Dates des sessions

### PARIS

17 sep. 2018, 15 oct. 2018  
26 nov. 2018

## 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

# Modélisation en étoile, conception et mise en œuvre

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

Ce cours, construit autour de nombreux cas pratiques, vous donnera une idée précise de la démarche de modélisation en étoile dans le cadre du projet Data Warehouse. Vous verrez pourquoi elle est l'expression même du besoin de la maîtrise d'ouvrage et comment elle permet de faire converger la vision des opérationnels, des analystes et des pilotes sur les activités de l'entreprise.

## OBJECTIFS PEDAGOGIQUES

Conduire des interviews pour recueillir les besoins d'analyse auprès des métiers

Définir les critères de qualité d'un Data Warehouse

A partir d'un cahier des charges d'analyse, identifier les dimensions et les faits à placer dans le modèle

Concevoir et décrire un macro-modèle en étoile

Optimiser le modèle en étoile en vue de la prise en compte de la volumétrie et des problématiques de reporting

Décrire une architecture de chargement des données dans le modèle en étoile décrit

### [1\) Introduction et rappels](#)

### [2\) Les architectures en réponse aux besoins décisionnels](#)

### [3\) Principes et définitions de base sur la modélisation en étoile](#)

### [4\) Conception du modèle en étoile](#)

### [5\) Optimisation fonctionnelle du modèle en étoile](#)

### [6\) Replacer la modélisation dans le cadre du projet décisionnel](#)

### [7\) Optimisation physique du modèle](#)

### [8\) Alimentation du modèle en étoile](#)

### [9\) Restitution des informations d'un modèle en étoile](#)

### [10\) Conclusion](#)

## 1) Introduction et rappels

- Qu'est-ce qu'un système d'information décisionnel ?
- Evolution des exigences de décision dans le contexte actuel.
- Infocentres, SIAD, EIS, Data Warehouse, définition et positionnement.
- Comprendre la finalité de l'approche Data Warehouse.

## 2) Les architectures en réponse aux besoins décisionnels

- Les composants principaux, Data Warehouse, ODS ou "staging area", datamarts.
- Les architectures proposées par Kimball et Inmon. Avantages et inconvénients.
- Positionnement du modèle en étoile dans le Data Warehouse selon l'architecture.
- Les phases du cycle de vie d'un Data Warehouse.
- Les critères de qualité d'un Data Warehouse.
- La notion de métadonnée, de référentiel.

### Réflexion collective

*Définition des critères de qualité d'un Data Warehouse.*

## 3) Principes et définitions de base sur la modélisation en étoile

- Rappels sur la modélisation des bases de données opérationnelles.
- Différences entre OLTP et OLAP.
- Entités, attributs, cardinalités, formes normales.
- Le principe de la dénormalisation pour concevoir un modèle en étoile.
- Comprendre les notions de fait, dimension et axe d'analyse.
- Les alternatives de modélisation : modèle en flocon, en galaxie.
- Les règles et bonnes pratiques de modélisation en étoile. Proposition alternative de Kortink et Moody.

### Etude de cas

*A partir d'un cahier des charges d'analyse, identifier les dimensions et faits principaux d'un modèle.*

## 4) Conception du modèle en étoile

- Organisation et synthèse des interviews utilisateur pour le recueil du besoin.
- Compréhension et identification des processus métiers à modéliser.
- Choix des dimensions d'analyse.
- Création de hiérarchies dans les dimensions.
- Identification des mesures et croisements avec les dimensions.
- Définition de la granularité de l'analyse.
- Définition des règles d'agrégation.
- Utilisation d'outils de modélisation.

### Exercice

*A partir d'objectifs fournis par la MOA, réaliser un macro-modèle, en reliant les dimensions.*

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.

## 5) Optimisation fonctionnelle du modèle en étoile

- Gestion de l'évolution des référentiels et du changement des nomenclatures.
- Gestion des dimensions à évolution lente et rapide.
- Les clés de substitution.
- Gestion de la qualité, fiabilité des données.
- Gestion du contexte non renseigné ou inconnu.
- Les dimensions dégénérées.

### Echanges

*Décrire l'impact d'un changement donné à partir d'un modèle proposé. Optimisation du modèle pour son évolution.*

## 6) Replacer la modélisation dans le cadre du projet décisionnel

- Présentation de la méthode Kimball et Inmon pour l'organisation du projet.
- Les acteurs et livrables du projet.
- Recueil des besoins métier. Formalisation des exigences techniques et d'organisation.
- Identification des priorités et du périmètre pilote.
- Modélisation des informations.
- Choix de l'infrastructure. Implémentation et recette.
- Déploiement et maintenance du modèle.
- Gestion des historiques.

### Jeu de rôle

*Conduite d'interview de recueil de besoin pour l'analyse.*

## 7) Optimisation physique du modèle

- Gestion de la performance des requêtes.
- Estimation de l'espace disque requis pour le modèle.
- Limitation de la taille occupée par une dimension.
- Agrégation directe de certains éléments dans les tables.
- Dimensions techniques pour assurer la traçabilité des faits.

### Exercice

*Estimations de volumétrie moyenne sur quelques cas d'analyse.*

## 8) Alimentation du modèle en étoile

- Contraintes des systèmes opérationnels sources.
- Rôle des ODS dans l'alimentation.
- L'organisation des traitements dans la DSA (Data Staging Area).
- Les différents types d'alimentation (delta, stock, complète).
- Les étapes, les règles et les prérequis de l'alimentation.
- Gestion des rejets.
- Gestion des sources différentes pour l'alimentation d'une dimension ou d'un fait.
- ETL, les solutions d'alimentation disponibles sur le marché.

### Exercice

*Sur une étude de cas, proposer une architecture de chargement : ODS / Staging Area.*

## 9) Restitution des informations d'un modèle en étoile

- Les différents types d'outils au service de la restitution.
- Le marché des outils de restitution.
- Optimisation du modèle pour l'exploration des données.
- Optimisation des index.
- Utilisation du partitionnement des tables.

### Echanges

*Présentation de bonnes pratiques pour optimiser le modèle en vue du reporting.*

## 10) Conclusion

- Ce qu'il faut retenir.
- Les pièges à éviter.
- Pour aller plus loin.

Cours de synthèse de 2  
jour(s)

Réf : BIM

## Participants

Directeurs marketing, directeurs du digital, DSI, responsables organisation ou études, chefs de projet fonctionnel, responsables Business Intelligence.

## Pré-requis

Aucune connaissance particulière.

Prix 2018 : 1470€ HT

## Dates des sessions

PARIS

27 sep. 2018, 17 déc. 2018

## Modalités d'évaluation

L'objectif de cette formation étant essentiellement de fournir une synthèse des méthodes et technologies existantes, il n'est pas nécessaire d'avoir recours à une évaluation des acquis.

## 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

# Social Media Analytics, stratégie et usages

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

Tous les jours, des millions d'internautes partagent leurs expériences sur les médias sociaux. L'enjeu pour les entreprises est d'intégrer cet échange entre leur marque et les internautes dans les processus de l'entreprise, afin de produire un véritable avantage concurrentiel. Dans ce contexte, l'adoption d'une démarche de « Social Data Analytics », permettant de définir les bons indicateurs et décider, nécessite une vision éclairée de son propre niveau de maturité, les bonnes clés pour s'aligner avec la stratégie de son entreprise, et une connaissance des freins à sa mise en place.

## OBJECTIFS PEDAGOGIQUES

Comprendre les changements et les enjeux du Social Business  
Savoir cadrer une stratégie BI pour le Social Business  
Connaître les étapes clés de mise en oeuvre de la BI pour le Social Media  
Savoir mesurer la performance de cette stratégie

### 1) Comprendre les enjeux du Social Business

#### 2) Cadrer la stratégie de Business Intelligence pour le Social Business

#### 3) Mettre en oeuvre la stratégie BI pour le Social Business

### 4) Cas concret d'usage de KPI et tableaux de bord Social Business

#### 5) Synthèse et compléments

## 1) Comprendre les enjeux du Social Business

- Définition du Social media et de ses usages.
- Compréhension et définition des médias sociaux, Social CRM, Social Analytics.
- Social media, ou l'empowerment du consommateur.
- Panoramas et chiffres clés du Social Business dans le monde et en France.
- Evolutions et impacts sur le Social Business.
- Analyse du contexte : de la relation à l'interaction Web via des technologies complexes.
- Le Paid Media, le Owned media et le Earned media, définitions et bénéfices.
- Typologies d'usages du Social Business dans les entreprises.
- Les applications du Social Business au Marketing, aux RH, R&D, innovation et ventes.
- Les enjeux du Social Business : image, trafic, conversion et fidélisation. Les enjeux de son pilotage.
- Le social media, un risque à anticiper.

## 2) Cadrer la stratégie de Business Intelligence pour le Social Business

- Les facteurs clés du succès de la Business Intelligence pour le Social Business.
- Les principaux risques à éviter sur les objectifs, l'apport de valeur, l'organisation, les impacts IT, l'implication.
- Les principales étapes et la méthodologie. Questions clés et facteurs de réussite.
- Diagnostiquer la maturité dans l'entreprise et le changement.
- Les questions clés pour établir un diagnostic, les différents stades de maturité.
- Identifier les freins à l'adoption d'une stratégie Social Business.
- Définir les objectifs métiers et les usages cibles, lier les objectifs du Social Business aux objectifs Business.
- Exemple d'une Roadmap Social Business.
- Piloter la stratégie et mettre en place une organisation adaptée. Les différents modes d'organisation.
- Zoom sur les équipes consacrées aux médias sociaux.
- Construire un plan de veille technologique, objectifs et bonnes pratiques.
- Benchmark des solutions du marché, les bonnes questions à se poser.
- Respecter la réglementation (vie privée, CNIL). Les données personnelles et les réseaux sociaux.

## 3) Mettre en oeuvre la stratégie BI pour le Social Business

- Définir la tactique en mode intégré ou spécifique.
- Vue d'ensemble de l'approche intégrée.
- Panorama des solutions du marché (outils de Social Media Management, de Social CRM).
- Vue d'ensemble de l'approche spécifique, de ses avantages et des contraintes.
- Panorama des technologies : représentation d'une plateforme Big Data, outils de Data Visualisation (DataViz).
- Etablir la méthode de benchmark des solutions.
- Comparaison des deux approches. Analyse des avantages et inconvénients de chaque méthode.
- Mettre en oeuvre de manière agile l'approche retenue. Mise en oeuvre d'un Proof of Concept, de l'idée à l'action.
- Zoom sur la conduite du changement. Principes, règles et actions du changement.
- Maintenir et optimiser la stratégie. Bonnes pratiques globales et points spécifiques à chaque méthode.

## 4) Cas concret d'usage de KPI et tableaux de bord Social Business

- Pourquoi mettre en place des indicateurs de performance ? Les différents indicateurs (de résultat, de performance...).

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.

- Mesurer l'e-réputation et la notoriété de sa marque. Une frontière qui s'amenuise entre réputation et e-réputation.
- Les indicateurs de mesure de l'e-réputation : mentions, analyse sentimentale, reach/exposition, Share of Voice (SOV).
- Comment mesurer les différents indicateurs et quand les utiliser ?
- Mesurer le ROI des influenceurs et l'efficacité des programmes Social Business.
- Les influenceurs, qui sont-ils ? Leurs différentes catégories. Comment bien les choisir, stratégie à adopter.
- Audience Growth rate, engagement, partage, influence... Un exemple concret de calcul du ROI des réseaux sociaux.
- Optimiser le ROI des campagnes marketing, facteurs à prendre en compte et bonnes pratiques.
- Identifier l'apport de valeur ajoutée des différents canaux digitaux. Les plateformes sociales préférées des marketeurs.
- Mesurer l'impact des médias sociaux sur le cycle et le processus de vente des produits ou services.
- Mesurer l'expérience et la satisfaction client, optimiser le parcours client.

## 5) Synthèse et compléments

- Opportunités du social media.
- Perspectives pour demain.
- Pour aller plus loin.



Séminaire de 3 jour(s)  
Réf : SID

## Participants

Responsables informatiques, responsables des études, architectes SI, consultants et chefs de projet décisionnel, autres chefs de projets fonctionnels et techniques.

## Pré-requis

Aucune connaissance particulière.

Prix 2018 : 2610€ HT

## Dates des sessions

### PARIS

25 sep. 2018, 27 nov. 2018

## Modalités d'évaluation

Les apports théoriques et les panoramas des techniques et outils ne nécessitent pas d'avoir recours à une évaluation des acquis.

## 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.

# Systemes décisionnels, démarche et solutions

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

Aujourd'hui, 80% des dirigeants attribuent une importance stratégique à l'extraction de la valeur de l'information par le machine learning et l'analyse. Pourtant, 43% des entreprises ne dégagent que peu d'avantages concrets de leurs informations. Devant ces constats, et face au levier de croissance que représentent les données, les dirigeants souhaitent en tirer parti en intégrant les technologies type Big Data, à leur décisionnel existant. Tout l'enjeu est une intégration cohérente pour servir au mieux l'analyse des données dans un souci de liberté, de réactivité et d'opportunités.

## OBJECTIFS PEDAGOGIQUES

Comprendre la valeur ajoutée, les enjeux et les principes des systèmes décisionnels  
Mélanger la modélisation décisionnelle classique et le mode découverte  
Mettre en œuvre une démarche pour concevoir un référentiel des données d'entreprise  
Intégrer le Big Data dans le SID pour bâtir la Data Factory  
Suivre pas à pas les étapes pour piloter son projet SID  
Faire les bons choix d'outils en ayant un panorama global du marché

### 1) Finalité et principes des SID

### 2) Démarche de conception du SID, impact du mode découverte

### 3) Construire les référentiels

### 4) Optimiser l'accès aux données

### 5) Mesurer la valeur du SID

### 6) Big Data en mode industriel

### 7) Etat de l'art des outils décisionnels

### 8) Opportunité et création de valeur pour l'entreprise

### 9) Organisation de la gouvernance : équipes - process - données

## Exemple

Un exemple complet de mise en œuvre d'un système d'information décisionnel sera traité.

## 1) Finalité et principes des SID

- Positionnement du SID par rapport au Big Data et au SIO (Système d'Information Opérationnel).
- Tendances et évolutions du décisionnel. Data Lake versus Datawarehouse.
- Nouveaux enjeux : valorisation de l'information, corrélation rapide.
- L'architecture d'un SID.
- Les choix stratégiques d'architecture : intégration du Data Lake.
- Nouvel équilibre entre modélisation préalable et analyse dynamique.

## 2) Démarche de conception du SID, impact du mode découverte

- La typologie universelle des demandes sur un SID autour du pilotage et de l'analyse comportementale prédictive.
- Maîtriser le processus de conception des datamarts et des datalabs.
- Comment passer de datamarts en silo à un SI ouvert à la BI découverte et au Big Data ?
- Cohérence des modèles en étoile : multidimensionnel. Analyse en mémoire contre modèle en étoile. Modèle du datalab.
- Dénormalisation NoSQL versus dénormalisation décisionnelle classique.
- Différence entre analyse multidimensionnelle et prédictive.
- Eviter le foisonnement d'agrégats et indicateurs en réutilisant les développements.
- Concevoir une architecture mixte BI classique-BI découverte-Analytique dynamique.

### Etude de cas

Proposer une démarche de conception à partir de besoins d'analyse.

## 3) Construire les référentiels

- Comment faciliter l'accès aux informations ? Réconcilier MOA, MOE autour du référentiel.
- Construire le référentiel d'entreprise. Les dimensions d'analyse et les indicateurs partageables.
- Construire une architecture couvrant toutes les étapes, du pilotage à l'analyse des comportements.
- Construire les dictionnaires pour le SID, utiliser les métadonnées pour gérer la cohérence.

### Etude de cas

Déploiement des méthodes proposées sur des exemples.

## 4) Optimiser l'accès aux données

- Comment répondre aux besoins et définir les réponses informatiques optimales ?
- Le multidimensionnel, ROLAP, MOLAP ou Hybride : les critères de choix, les évolutions futures.
- Le multidimensionnel In-Memory.
- Organiser son Data Lake. Construire les Datalabs.
- Appliquer une normalisation décisionnelle à vos modèles en étoile.

• 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.

## 5) Mesurer la valeur du SID

- Faire de son SID un levier de la stratégie d'entreprise.
- Réunir les cas d'usage en process décisionnels.
- Urbanisation du SID : éviter trop de traitements et des couches sémantiques surchargées.
- Identifier les zones éligibles au Cloud Computing.
- Cartographier son SID pour lier finalité d'usage et données utilisées pour le RGPD.
- Définir les critères d'efficacité d'un SID.
- Piloter la valeur des données.

## 6) Big Data en mode industriel

- Principaux types de cas d'usage.
- Problème de l'industrialisation des projets Big Data.
- Check-list des recommandations.
- Analytique – prédictif temps réel et streaming (CEP Complex Event Processing).

## 7) Etat de l'art des outils décisionnels

- Lier ou fusionner son Data Lake et son Datawarehouse pour créer la Data Factory.
- Panorama des suites décisionnelles : SAS, Microsoft, SAP B.O...
- Degré d'intégration du mode découverte, de l'analytique et de la data visualisation.
- ETL. Outils multidimensionnels. Déploiement Web.
- Big Data intégré au SID. BD NoSQL. BD NewSQL. Cohabitation entre les différentes Bases de Données.
- Analyse en mémoire. Appliance ou Commodity Hardware.
- Positionnement de l'écosystème HADOOP dans l'analyse du comportement client.
- Basculer le SID sur une BD NoSQL, NewSQL ou intégrer les approches ?
- Combiner une solution agile de data-discovery et des capacités d'industrialisation de la BI.

### Etude de cas

*Déterminer votre trajectoire d'évolution vers une architecture intégrée.*

## 8) Opportunité et création de valeur pour l'entreprise

- Evaluer la valeur ajoutée pour l'entreprise et la conduite du changement utile.
- Gérer son portefeuille de projets, les prioriser. Critères de lotissement.
- Spécificités d'un projet décisionnel et d'un projet Big Data.
- Techniques d'analyse de besoin en décisionnel : pièges à éviter.
- Comment évaluer la complexité, la maturité des besoins ?
- Conduire le passage du décisionnel classique au mixte classique-analytique en mode libre.

## 9) Organisation de la gouvernance : équipes - process - données

- Différents acteurs et rôles respectifs. Nouvelle relation entre métiers et IT.
- Leviers de la BI découverte, du Big Data et de la Datascience.
- Positionner le décisionnel au sein de l'entreprise. Organiser la gouvernance, la cohérence globale des données.
- Créer une organisation cohérente et pluridisciplinaire.
- Préserver l'autonomie des utilisateurs. Gérer la réactivité.
- Intégrer les métiers dans le pilotage de la valeur : données et cas d'usages.
- Organiser la Data Factory. Administrer les composants du SID.
- Garantir la fiabilité des données et des informations.
- Mettre en place le processus de la gestion de la qualité et de la véracité.
- Définir les contrôles qualité minimum. Définir le phasing des contrôles.
- Impact de la réglementation RGPD sur la sécurité d'accès aux données.

Stage pratique de 2 jour(s)

Réf : STA

## Participants

Responsables Infocentre, responsables marketing, responsables Qualité, utilisateurs et gestionnaires métiers de bases de données, futurs Data Scientist.

## Pré-requis

Connaissances de base en mathématiques et statistiques ou connaissances équivalentes à celles apportées par le stage "Statistiques descriptives, introduction" (réf. UES). Connaissances de base Excel.

Prix 2018 : 1470€ HT

## Dates des sessions

PARIS

17 sep. 2018, 22 nov. 2018

## 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.

# Modélisation statistique

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

Ce stage présente l'essentiel de la modélisation statistique. Il vous permettra de comprendre leur rôle dans le monde de l'analyse décisionnelle, du Big Data et du Data Mining, ainsi que les mécanismes qui permettent de transformer et d'affiner des données pour en tirer des informations métiers utiles.

## OBJECTIFS PEDAGOGIQUES

Dimensionner un échantillon de population  
Calculer des paramètres de position et de dispersion (médiane, étendue, quantile, écart-type)  
Exploiter les paramètres statistiques pour comprendre une série de données  
Valider la précision d'une estimation, à l'aide des intervalles de confiance  
Découvrir des outils comme R et Excel pour la mise en œuvre des modèles étudiés

[1\) Rappels des fondamentaux de la statistique descriptive](#)

[2\) Démarche et modélisation d'une analyse statistique](#)

[3\) Paramètre de position et de dispersion](#)

[4\) Tests et intervalle de confiance](#)

[5\) Panorama des outils](#)

## 1) Rappels des fondamentaux de la statistique descriptive

- Définition de la statistique descriptive.
- Analyse d'une population.
- Méthodes d'échantillonnage.
- Variables qualitatives et quantitatives.
- Effectifs et calcul des fréquences.
- Effectifs cumulés croissants et décroissants.
- Représentation graphique des variables qualitatives et quantitatives.

### Etude de cas

*Application pratique sur excel d'analyses statistiques et interprétation*

## 2) Démarche et modélisation d'une analyse statistique

- Statistique descriptive.
- Phase d'apprentissage.
- Statistique prédictive pour estimer et anticiper.
- Modélisation statistique d'un phénomène.

## 3) Paramètre de position et de dispersion

- Mode, valeur modale, valeur la plus probable.
- Moyenne d'une population (ou d'un échantillon).
- Médiane, partager une série numérique.
- Etendue, différence entre valeurs extrêmes.
- Utiliser les quantiles.
- Ecart-Type, calculer la dispersion d'un ensemble de données.
- Calcul de la variance et de la covariance.

### Etude de cas

*Calcul de paramètres de position et de dispersion sur différents échantillonnages et comparaisons des résultats.*

## 4) Tests et intervalle de confiance

- Lois statistiques et intervalle de confiance.
- Tests statistiques courants (Test de Student, Analyse de variances,  $\chi^2$ ).
- Valider la précision d'une estimation. Amplitude de l'intervalle.

### Etude de cas

*Exercices sur le logiciel R.*

## 5) Panorama des outils

- Zoom sur le logiciel Open Source "R".
- Initiation au logiciel Open Source "R".

### Travaux pratiques

*Utilisation de packages pour faire les analyses statistiques.*

- 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.

Séminaire de 2 jour(s)  
Réf : DMI

## Participants

Responsables Infocentre, responsables marketing, statisticiens, responsables informatiques, chefs de projets et experts décisionnels. Utilisateurs et gestionnaires métiers de bases de données.

## Pré-requis

Connaissances de base en analyse décisionnelle. Connaissances de base en statistiques.

Prix 2018 : 1910€ HT

## Dates des sessions

### PARIS

04 oct. 2018, 06 déc. 2018

## Modalités d'évaluation

Les apports théoriques et les panoramas des techniques et outils ne nécessitent pas d'avoir recours à une évaluation des acquis.

## 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

# Data Mining : synthèse

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

*Le Data Mining apporte des solutions pour l'extraction de connaissances à partir de données brutes. Ce séminaire, alliant principes théoriques et expériences pratiques, vous propose une synthèse sur le Data Mining, ses applications, ses apports pour l'entreprise, ainsi que sur les méthodes et outils les plus utilisés.*

## OBJECTIFS PEDAGOGIQUES

Comprendre les principes et la finalité du Data Mining (DM)  
Identifier les principales techniques du DM et leur cas d'utilisation  
Mettre en oeuvre sur un cas simple les méthodes de scoring et de géomarketing  
Découvrir les méthodes prédictives et les méthodes descriptives du DM  
Connaître les principales étapes d'un projet Data Mining

### [1\) Le Système d'Information Décisionnel \(SID\)](#)

### [2\) Comprendre le Data Mining \(DM\)](#)

### [3\) Les techniques du Data Mining](#)

### [4\) La méthode descriptive du Clustering](#)

### [5\) Exemples d'application du DM](#)

### [6\) Les données de l'entreprise](#)

### [7\) Méthodologie de projet Data Mining](#)

### [8\) Panorama des outils](#)

## 1) Le Système d'Information Décisionnel (SID)

- Les enjeux du SID : besoins, domaines d'application.
- Architecture type d'un SID, état de l'art.
- Elaboration des informations décisionnelles.
- Conception d'un SID : étapes, optimisation, organisation des données, dictionnaires.

## 2) Comprendre le Data Mining (DM)

- Définition et finalité du Data Mining (DM).
- Quel lien entre le DM et les statistiques, quelle dépendance entre le DM et l'informatique ?
- Différence entre DM et OLAP ?
- Les attentes des entreprises, les réponses du DM.

## 3) Les techniques du Data Mining

- Les différentes familles du DM.
- Les méthodes prédictives et les méthodes descriptives.
- Analyse factorielle, typologique. La classification...
- Les arbres de décisions, les réseaux de neurones...
- Classification des techniques de DM.

## 4) La méthode descriptive du Clustering

- Définition et méthodologie.
- Les critères pour structurer les données à classer.
- Evaluation et validation des classes obtenues.
- Les différentes sous-familles du Clustering.

### Exemple

*Présentation d'applications du Clustering.*

## 5) Exemples d'application du DM

- Le scoring : définition, finalité, méthodologie.
- Le géomarketing : définition, finalité, méthodologie.

### Exemple

*Mise en oeuvre de la méthode de scoring. Cas pratique d'utilisation de géomarketing.*

## 6) Les données de l'entreprise

- Rappel de la problématique des données du SI.
- Qualité des données et administration des données.
- Processus de collecte et d'exploration.
- Création d'agrégats et de nouvelles données.
- Transformation des données.

## 7) Méthodologie de projet Data Mining

- Définition du problème métier à résoudre et des objectifs à atteindre.
- Inventorier, décrire et classer les données.
- Concevoir et alimenter la base Data Mining.
- Explorer, segmenter des entités analysées.

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.

- Etablir et appliquer les modèles d'analyse.
- Itérer, déployer auprès des utilisateurs.
- Maintenir le modèle et le logiciel associé.

## 8) Panorama des outils

- Principaux outils du marché : SAS, R, IBM SPSS...
- Zoom sur l'outil SAS et sur l'ETL Powercenter.
- Quels critères de choix pour ce type d'outils ?

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 2018 : 3030€ HT

## Dates des sessions

### AIX

10 sep. 2018, 19 nov. 2018

### ANGERS

10 sep. 2018, 19 nov. 2018

### BORDEAUX

17 sep. 2018, 03 déc. 2018

### BRUXELLES

10 sep. 2018, 26 nov. 2018

### DIJON

17 sep. 2018, 03 déc. 2018

### GENEVE

10 sep. 2018, 26 nov. 2018

### GRENOBLE

17 sep. 2018, 03 déc. 2018

### LILLE

10 sep. 2018, 19 nov. 2018

### LIMOGES

10 sep. 2018, 19 nov. 2018

### LUXEMBOURG

10 sep. 2018, 26 nov. 2018

### LYON

17 sep. 2018, 03 déc. 2018

### MONTPELLIER

10 sep. 2018, 19 nov. 2018

### NANCY

10 sep. 2018, 19 nov. 2018

### NANTES

10 sep. 2018, 19 nov. 2018

### ORLEANS

10 sep. 2018, 19 nov. 2018

### PARIS

27 août. 2018, 10&24 sep. 2018

08&22 oct. 2018, 12&19 nov. 2018

10 déc. 2018

### RENNES

10 sep. 2018, 19 nov. 2018

### ROUEN

10 sep. 2018, 19 nov. 2018

### SOPHIA-ANTIPOLIS

10 sep. 2018, 19 nov. 2018

### STRASBOURG

10 sep. 2018, 19 nov. 2018

### TOULOUSE

17 sep. 2018, 03 déc. 2018

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

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

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.

## TOURS

10 sep. 2018, 19 nov. 2018

### 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.

- Analyse et visualisation des données.
- 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.



Cours de synthèse de 2  
jour(s)  
Réf : CYE

## Participants

Toute personne ayant à mettre en place une démarche de gouvernance des données d'une entreprise et/ou un projet de Master Data Management.

## Pré-requis

Aucune connaissance particulière.

Prix 2018 : 1770€ HT

## Dates des sessions

### PARIS

20 sep. 2018, 08 nov. 2018  
20 déc. 2018

## Modalités d'évaluation

L'objectif de cette formation étant essentiellement de fournir une synthèse des méthodes et technologies existantes, il n'est pas nécessaire d'avoir recours à une évaluation des acquis.

## 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

# Master Data Management, la gouvernance des données

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

Les enjeux concurrentiels et la nécessité accrue de transparence sur les données impose de disposer d'une méthodologie et d'une architecture de données maîtrisées et alignées sur les métiers. Ce cours vous présentera les enjeux et les méthodes afin d'engager tous les Systèmes d'Information dans cette démarche.

## OBJECTIFS PEDAGOGIQUES

Comprendre le rôle stratégique de la gestion des données pour l'entreprise.  
Comprendre les principes d'architecture des données  
Mettre en place une méthode de gouvernance  
Intégrer la gestion des Master Data dans la démarche  
Savoir identifier les acteurs du MDM et leur positionnement

### 1) Introduction

### 2) La gouvernance des données

### 3) Le Master Data Management

### 4) Cycle de vie des données

### 5) Présentation des offres de MDM

## 1) Introduction

- Enjeux stratégiques des données pour l'entreprise.
- Définition des notions "Donnée" et "Information".
- Les différentes sources de données de l'entreprise.
- Les différentes formes d'exploitation des données.
- Système d'information opérationnel et décisionnel.
- Architectures : Relationnelle/Big Data.

### Echanges

*Echanges sur le rôle stratégique des données pour l'entreprise.*

## 2) La gouvernance des données

- Définition et enjeux de la gouvernance des données.
- L'approche COBIT dans la gouvernance des données.
- Les acteurs de la gouvernance des données.
- Outils de mesure pour l'évaluation de la maturité d'une entreprise.
- Les grands principes de la démarche de gouvernance de données.
- Références et état de l'art.

### Echanges

*Une organisation met en place une démarche de gouvernance des données à l'occasion de la fusion de son SI avec un nouveau SI provenant d'une acquisition. Etude de l'impact sur l'organisation et les SI*

## 3) Le Master Data Management

- Positionnement du Master Data Management dans le Système d'Information d'entreprise.
- Les étapes essentielles de la démarche Master Data Management.
- Présentation des architectures Master Data Management.
- L'administration des données référentielles (Master Data).
- Audit et gestion de la qualité des données.
- Synthèse des bonnes pratiques.
- Le rôle des utilisateurs dans le dispositif MDM.

### Réflexion collective

*Etude de cas d'un SI dépourvu de gestion de Master Data sur lequel on effectue une analyse des besoins en MD. Comparaison de deux approches possibles de mise en œuvre.*

## 4) Cycle de vie des données

- Typologie et volume de données.
- Archivage de base de données.
- Confidentialité des données.
- Gestion des données de test.
- Décommissionnement d'applications.

### Echanges

*Echanges sur la gestion des données et leur cycle de vie.*

## 5) Présentation des offres de MDM

- Les tendances du marché
- MDM dans les suites Informatique /IBM/SAP.
- MDM Microsoft.
- Talend MDM.

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.

- Conclusion.

### **Echanges**

*Echanges sur les stratégies d'approches possibles selon les contextes.*

Stage pratique de 1 jour(s)  
Réf : KSW

## Participants

Candidats à la certification Data Scientist.

## Pré-requis

Avoir suivi l'intégralité des stages du cycle certifiant Data Scientist.

Prix 2018 : 550€ HT

## Dates des sessions

### PARIS

05 oct. 2018, 05 nov. 2018  
14 déc. 2018

## Modalités d'évaluation

Pour les certifications, des examens blancs permettent de mesurer le degré de préparation. Pour les cycles certifiants, une journée d'examen est mise en place.

## 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

# Journée de certification du cycle certifiant Data Scientist

> [Big Data, BI, NoSQL, SGBD](#) > [Data, état de l'art et méthodes](#)

## OBJECTIFS

*Cette journée est destinée à valider les connaissances acquises au cours du cycle certifiant Data Scientist. Les compétences validées portent sur la modélisation statistique pour réaliser la fouille des données (Data Mining) et sur les connaissances connexes en matière de "Data Quality" et de technologies "Big Data".*

### 1) Présentation des épreuves

### 2) Epreuve écrite : savoir

### 3) Epreuve écrite : savoir- faire

## Certification

*Dans le cadre de cette journée, les participants auront une étude de cas à analyser et des conclusions à choisir sur les questionnaires proposés.*

## 1) Présentation des épreuves

- Rappels sur les points essentiels vus pendant le cycle certifiant et sur le rôle du Data Scientist.
- Présentation des épreuves écrites et orales (documents ressources, durée, plan de passage etc.).
- Echanges avec l'intervenant.

## 2) Epreuve écrite : savoir

- Questionnaire de connaissances sur les principaux stages du cycle certifiant.
- Questionnaire à choix multiple.

## 3) Epreuve écrite : savoir- faire

- Prise de connaissance de l'étude de cas (une problématique d'entreprise).
- Questionnaire à partir de l'étude de cas.

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.