

## BigData : concevoir et gérer un projet

Durée: 2 jours

1160 €

24 au 25 mai

13 au 14 septembre

6 au 7 décembre

### Public:

Chefs de projet, et toute personne souhaitant comprendre et pratiquer les méthodes, outils de conception d'un projet BigData

### Objectifs:

Comprendre le contexte spécifique, le vocabulaire et les enjeux du Big Data afin de bien positionner les fondations d'un tel projet tant sur le plan humain, technique, sécurité, financier et juridique.

### Connaissances préalables nécessaires:

Il est demandé aux participants d'avoir une bonne culture générale sur les systèmes d'information (systèmes, réseaux, architectures cloud), de maîtriser un langage de développement objet.

### Programme:

Définition et contexte spécifique des projets BigData : Panorama technologique et enjeux socio-économiques  
Propriété de la donnée, environnement juridique du traitement, sécurité  
Impact des choix technologiques en matière d'analyse et de visualisation de données.  
L'essentiel du BigData : calcul distribué, données non structurées.  
Besoins fonctionnels et caractéristiques techniques des projets.  
La valorisation des données.  
Le positionnement respectif des technologies de cloud, BigData et noSQL, et les liens, implications.

Propriété de la donnée, environnement juridique du traitement, sécurité : Sécurité éthique et enjeux juridiques Les données personnelles Les informations sensibles, interdites de collecte La CNIL régule les données numériques Les accords intra-pays

## BigData : concevoir et gérer un projet

- Technologies, compétences et métiers** : Ranger les technologies dans les bonnes cases (Hadoop, MapReduce, Pig, Hive, Impala, Spark, Elasticsearch, etc.) Différences de savoir-faire pour le chef de projet avec la BI «traditionnelle» Léger zoom sur Hadoop Les métiers « informatiques » concernés par le Big Data (administrateur, développeur, analyste, data scientist, etc.) Quel élargissement des compétences pour les administrateurs et les développeurs Hadoop ? Synthèse : exemple d'une architecture Big Data en production et positionnement des métiers La journée type du chef de projet Big Data Conséquences financières des choix techniques
- Comprendre et traiter les spécificités d'un projet Big Data au sens organisationnel, méthodologique, technologique, économique, juridique et humain** : Quelle démarche méthodologique et quelles étapes clé du chantier ? Les prérequis à réunir et points de vigilance à surveiller ? Comment gouverner et piloter le chantier Big Data ? Quels sont les contributeurs internes à mobiliser ? Les profils et compétences nécessaires à intégrer ? La connaissance de la question
- Composer et piloter une équipe Big Data** : Hadoop dans le SI : Processus d'adoption Comment conserver les ressources humaines « rares » dans le contexte du Big Data ? Datalab : une équipe, un lieu, une approche Comment élaborer un Datalab ?
- Manager le processus de mise en place d'un projet Big Data** : Gouvernance des Données et « Culture Données » Définir une véritable stratégie Données Partager les enjeux Investir sur des initiatives ciblées et porteuses de valeur pour les métiers Rompre avec les visions en silos Promouvoir une « Culture Données » Savoir communiquer (marketing digital)

## BigData : concevoir et gérer un projet

- Identifier les besoins et le type de données à traiter avec les métiers (use cases) : À travers différents cas d'école piochés sur l'Open Data, nous effectuons des use cases permettant de définir les indicateurs clés à déterminer La viabilité de chaque effort est bien sûr mesurée par la détermination du ROI
- La collecte et le stockage des données : Quelles données ? Quelles sources ? L'importance de la qualité des données (data quality) Exemple de traitement avec un ETL dédié Big Data Résumé : les différentes phases de la collecte dans un projets
- L'exploitation des données : Données structurées, semi structurées et non structurées La réconciliation avec le référentiel interne La question du Master Data Management Big Data ou Smart Data ?
- L'analyse des données : Définition de l'analyse statistique La Datascience La place du Data scientist dans un projet Big Data Datamining L'implémentation d'indicateurs à destination du décisionnel Résumé : les différentes phases et formes de l'analyse dans un projet
- La visualisation des données (Dataviz) : Ce que les statistiques ne disent pas Les objectifs de la visualisation Quels graphes pour quels usages ? Représentation de données complexes (encodage visuel, visualisation interactive) Savoir communiquer sur les analyses de données (Data Storytelling)
- Piloter et maîtriser les risques des projets Big Data : Différentes méthodes pour piloter un datalab : - Business Driven - Lean - Scrum
- Tests et analyses de performances : Stratégie pour le code de débogage MapReduce Test local du code en utilisant LocalJobRunner Écriture et utilisation de fichiers journaux