

Formation BIG DATA « Architecture et Infrastructure »

Formation Informatique / Réseaux et Sécurité / Sécurité



Le BIG DATA (littéralement Megadonnées) désigne la propriété de données recueillies en masse et leur traitement par des méthodes technologiques spécifiques permettant de mieux les visualiser pour obtenir des indicateurs permettant des prises de décisions stratégiques

OBJECTIFS

- Maîtriser les aspects fondamentaux de BIG DATA
- Maîtriser des éléments clés liés aux architectures des SI
- Pouvoir proposer des architectures répondant aux besoins du client
- Comprendre les différents cas d'usage et les outils correspondants
- Avoir une vision globale des différents outils et Framework
- Mettre en pratique les connaissances sur des problématiques réelles

PUBLIC

- Architecte SI, Développeur, Business analyste, Consultant BI, ingénieur d'étude, Concepteur, Chef de projet, Manager

PRE-REQUIS

- Aucune connaissance sur les technologies Big Data n'est requise
- Des connaissances en architecture SI et en conception de Bases de données serait un plus

PROGRAMME

PROGRAMME DE LA FORMATION

Jour 1

1. Les questions clés à se poser lors de la mise en place d'un projet Big Data
1. Quels sont les différents cas d'usage de l'application ?
2. Quelles sont les sources de données ?
3. Les mesures de qualité et de performances ?
4. Quel est le cycle de vie de la donnée ? Par quelles transformations passe-t-elle ? Comment la gouverner ?
5. Comment gérer les différentes charges de travail et héberger des projets différents sur la même infrastructure ?
6. Quelles sont les principaux modèles d'architecture d'un SI Big Data ? C'est quoi le "Data Lake" ?
7. Comment dimensionner l'infrastructure et mesurer la scalabilité du système ?
8. C'est quoi le cycle de vie d'un projet Big Data ?
9. Quelles sont les compétences nécessaires pour réussir dans la mise en place d'un projet Big Data ?
10. Comparaison des principales distributions d'Hadoop : Apache Hadoop vs HortonWorks vs Cloudera vs MapR

II. Architecture Big Data

1. LIMITES DES ARCHITECTURES CLASSIQUES

2. AVANTAGES DES NOUVELLES ARCHITECTURES BIG DATA

- a. Structures différentes de données (structurées, semi-structurées et non structurées)
- b. Stockage de grandes quantités de données à moindre coût et Scalabilité en termes de stockage
- c. Performances élevées du traitement de données massives sous différentes latences (Batch ou en Streaming)
- d. et scalabilité en termes de traitements
- d. Haute disponibilité
- e. Agilité du système et son impact positif sur son évolution et sa maintenance
- f. Impact sur l'entreprise et sa transformation en Client-Centric

3. LE DATA LAKE : UNE NOUVELLE PHILOSOPHIE POUR LE STOCKAGE ET LE TRAITEMENT DE LA DONNEE

- a. Architecture du Data Lake et centralisation des données
- b. Cas d'usages basés sur le Data Lake
- c. Le modèle « en couches » de données
- d. Data Lake vs Data Warehouse

4. LA QUALITE DE LA DONNEE (DATAQUALITY) DANS UN SYSTEME BIG DATA

- a. Causes du problème de la qualité de données dans les systèmes Big Data
- b. Mise en place de la Gouvernance Applicative pour assurer la qualité des données
- i. Identification des actions clés à mettre en place pour garantir la qualité de la donnée
- ii. Retour d'expérience sur la mise en place de projets de gouvernance de la donnée chez des clients grands comptes

5. LA QUALITE DES TRAITEMENTS DANS UN SYSTEME BIG DATA



A retenir

Durée : **3 jours** soit 21h.
Réf. **BGDTAI**

☎ 01 42 93 52 72

Dates des sessions

Paris
25/09/2019
27/11/2019

Cette formation est également proposée en formule **INTRA-ENTREPRISE.**



Inclus dans cette formation



Coaching Après-COURS

Pendant 30 jours, votre formateur sera disponible pour vous aider. CERTyou s'engage dans la réalisation de vos objectifs.

100%
SATISFACTION
GARANTIE

Votre garantie 100% SATISFACTION

Notre engagement 100% satisfaction vous garantit la plus grande qualité de formation.

- a. Importance de la supervision des chaînes de traitements
- b. Exploitation efficaces des logs et détection des anomalies en temps réelle
- c. Mise en place de tableaux de bord pour la visualisation de la qualité des traitements en temps réelle

6. TYPES D'ARCHITECTURES BIG DATA SELON LES EXIGENCES TEMPORELLES

- a. Architectures Batch
- b. Architectures Streaming
- c. Lambda Architectures

Jour 2

III. Tour d'horizon sur les technologies Big Data classées par cas d'usage

1. LIMITES DES OUTILS ET DES PARADIGMES CLASSIQUES QUI ONT MENE A L'APPARITION DES TECHNOLOGIES BIG DATA

2. HISTORIQUE DES TECHNOLOGIES BIG DATA : GOOGLE'S GFS & MAPREDUCE

3. PRINCIPAUX FRAMEWORKS DU BIG DATA : HADOOP, SPARK, STORM, ETC.

4. CLASSIFICATION DES OUTILS DU BIG DATA SELON LEURS CAS D'USAGE :

- a. Stockage et traitement de la donnée
 - i. Le système de fichiers distribué de Hadoop (HDFS)
 - ii. MapReduce
 - iii. YARN
 - iv. Spark
- b. Les bases de données et la gestion des données
 - i. Définition du NoSQL
 - ii. NoSQL Vs SGBD Relationnels
 - iii. Classification des bases de données NoSQL selon leurs types et leurs cas d'usages
- c. La sérialisation
 - i. Avro
 - ii. JSON
 - iii. Parquet
 - d. Le management et le monitoring
 - i. Ambari
 - ii. HCatalog
 - iii. ZooKeeper
 - iv. Oozie
- e. L'analyse et la visualisation des données
 - i. Spark MLIB
 - ii. Mahout
 - iii. Hadoop Streaming
 - iv. Pig et MapReduce
 - v. Hadoop Image Processing Interface (HIPI)
- f. Recherche
 - i. Elasticsearch
 - ii. Solr
 - g. Le transfert des données
 - i. Sqoop
 - ii. Flume
 - iii. DistCP
 - iv. Storm
 - v. Kafka
- h. La sécurité et le contrôle d'accès
 - i. Kerberos
 - ii. Ranger
 - iii. Sentry
 - iv. Knox

5. SOLUTIONS BIG DATA SUR LE CLOUD

6. RETOUR D'EXPERIENCE SUR LES FRAMEWORKS ET LES OUTILS BIG DATA UTILISES : LEURS AVANTAGES, LEURS LIMITES ET LEURS EVOLUTIONS

Jour 3

IV. Cas pratiques et études de cas

Horaires, Planning et Déroulement de cette formation

Horaires

- Formation de 9h00 (9h30 le premier jour) à 17h30.
- Deux pauses de 15 minutes le matin et l'après-midi.
- 1 heure de pause déjeuner

DEROULEMENT

- Les horaires de fin de journée sont adaptés en fonction des horaires des trains ou des avions des différents participants.
- Une attestation de suivi de formation vous sera remise en fin de formation.
- Cette formation est organisée pour un maximum de 14 participants.

PROCHAINES FORMATIONS

[Réussir la Certification Gestion de Projet PMP du PMI](#)

[Réussir la Certification PRINCE2 Foundation](#)

[Réussir les Certifications PRINCE2 Foundation et PRINCE2 Practitioner](#)

[Réussir la Certification ITIL Foundation](#)

[Réussir la Certification Agile certifié SCRUM Master](#)

[Réussir les Certifications TOGAF Certified et TOGAF Foundation](#)

Retrouvez cette formation sur notre site :

[Formation BIG DATA « Architecture et Infrastructure »](#)