



---

## BIG DATA : Architecture et Infrastructure

BIG DATA : Architecture et Infrastructure

durée : 3 jour(s)

code formation : BAI

---

Description :

Pré-requis :

---

Programme :

### Les questions clés à se poser lors de la mise en place d'un projet Big Data

- | Quels sont les différents cas d'usage de l'application ?
- | Quelles sont les sources de données ?
- | Les mesures de qualité et de performance ?
- | Quel est le cycle de vie de la donnée ? Par quelles transformations passe-t-elle ? Comment la gouverner ?
- | Comment gérer les différentes charges de travail et héberger des projets différents sur la même infrastructure ?
- | Quelles sont les principaux modèle d'architecture d'un SI Big Data ? C'est quoi le "Data lake" ?
- | Comment dimensionner l'infrastructure et mesurer la scalabilité du système ?
- | C'est quoi le cycle de vie d'un projet Big Data ?
- | Quelles sont les compétences nécessaires pour réussir dans la mise en place d'un projet Big Data ?
- | Comparaison des principales distributions d'Hadoop : Apache Hadoop vs HortonWorks vs Cloudera vs MapR

### Architecture Big Data

- | Les limites des architectures classiques
- | Avantages des nouvelles architectures BIG DATA
- | Structures différentes de données (structurées, semi-structurées et non structurées)
- | Stockage de grande qualités de données à moindre coût et Scalabilité en termes de stockage
- | Performances élevées du traitement de données massives sous différentes latences (Batch ou en Streaming) et scalabilité en termes de traitements
- | Haute disponibilité
- | Agilité du système et son impact positif sur son évolution et sa maintenance
- | Impact sur l'entreprise et sa transformation en Client-Centric
- | Le Data Lake : une nouvelle philosophie pour le stockage et le traitement de la donne
- | Architecture du Data Lake et centralisation des données
- | Cas d'usages basés sur le Data Lake
- | Le modèle "en couches" de données
- | Data Lake vs Data Warehouse

- | La qualité de la donnée (DataQuality) dans un système Big Data
- | Causes du problème de la qualité de données dans les systèmes Big Data
- | Mise en place de la Gouvernance Applicative pour assurer la qualité de la donnée
- | Retour d'expérience sur la mise en place de projets de gouvernance de la donnée chez les clients grands comptes
- | La qualité des traitement dans un système Big Data
- | Importance de la supervision des chaînes de traitements
- | Exploitation efficaces des logs et détection des anomalies en temps réelle
- | Mise en place de tableaux de bord pour la visualisation de la qualité des traitements en temps réelle
- | Types d'architecture Big Data selon les exigences temporelles
- | Architectures Batch
- | Architectures Streaming
- | Lambda Architectures

## **Tour d'horizon sur les technologies BIG DATA classés par cas d'usage**

- | Limites des outils et des paradigmes classiques qui ont mené à l'apparition des technologies BIG DATA
- | Historique des technologies BIG DATA : GOOGLE'S GFS & MAPREDUCE
- | Principaux frameworks du BIG DATA: HADOOP, SPARK, STORM, ETC.
- | Classification des outils du BIG DATA selon leurs cas d'usage
- | Stockage et traitement de la donnée
- | Le système de fichiers distribué de Hadoop
- | Mapreduce
- | YARN
- | Spark
- | Les bases de données et la gestion des données
- | Définition du NoSQL
- | NoSQL Vs SGBD Relationnels
- | Classification des bases de données NoSQL selon leurs types et leurs cas d'usages
- | La sérialisation
- | Avro
- | JSON
- | Parquet
- | Le management et le monitoring
- | Ambari
- | HCatalog
- | ZooKeeper
- | Oozie
- | L'analyse et la visualisation des données
- | Spark MLIB
- | Mahout
- | Hadoop Streaming
- | Pig MapReduce
- | Hadoop Image Processing Interface (HIPI)
- | Recherche

| Elasticsearch

| Solr

| Le transfert des données

| Sqoop

| Flume

| DistCP

| Storm

| Kafka

| La sécurité et le contrôle d'accès

| Kerberos

| Ranger

| Sentry

| Knox

| Solutions BIG DATA sur le cloud

| Retour d'expérience sur les frameworks et les outils BIG DATA utilisés : Leurs avantages, leurs limites et leurs évolutions