

BIGDATA : ARCHITECTURE ET TECHNOLOGIES

CODE STAGE : BD001

OBJECTIFS

Comprendre les concepts du BigData et savoir quelles sont les technologies implémentées.

DURÉE

2 jours

PUBLIC

Chefs de projets, architectes, et toute personne souhaitant connaître les outils et solutions pour mettre en place une architecture BigData.

PRÉ-REQUIS

Il est demandé aux participants d'avoir une bonne culture générale sur les systèmes d'information.

PROGRAMME

Module 1 : Introduction

Le besoin : volumes importants de données, traitements optimisés de flux de données au fil de l'eau, liés aux nouvelles technologies et aux nouveaux usages.

Domaines concernés : recherche scientifique, médical, e-commerce, sécurité

Développement des techniques sur différents aspects : stockage, indexation/recherche, calcul.

Définition ETL : Extract Transform Load.

Les acteurs.

Module 2 : Stockage

Caractéristiques NoSQL :

Structure de données proches des utilisateurs, développeurs

Données structurées et non structurées, documents, images, fichiers XML, JSON, CSV,

Les différents modes et formats de stockage.

Stockage réparti : réplication, sharding, gossip protocol, hachage,

Systèmes de fichiers distribués : GFS, HDFS, BigTable,

code2utf('8211',0) Les bases de données.

code2utf('8211',0) Quelques exemples de produits et leurs caractéristiques : cassandra, MongoDB, CouchDB, DynamoDB.

Module 3 : Indexation et recherche

code2utf('8211',0) Moteurs de recherche.

code2utf('8211',0) Principe de fonctionnement.

code2utf('8211',0) Méthodes d'indexation.

code2utf('8211',0) Exemple de Lucene, et mise en oeuvre avec solr.

code2utf('8211',0) Recherche dans les bases de volumes importants :

code2utf('8211',0) Exemples de produits et comparaison : dremel, drill, elasticsearch, MapReduce,

Module 4 : Calcul et restitution, intégration

code2utf('8211',0) Différentes solutions : calculs en mode batch, ou en temps réel, sur des flux de données ou des données statiques.

code2utf('8211',0) Les produits : langage de calculs statistiques, R Statistics Language

code2utf('8211',0) Outils de calcul sur des volumes importants : storm en temps réel, hadoop en mode batch.

code2utf('8211',0) Zoom sur Hadoop : complémentarité de HDMS et MapReduce.

code2utf('8211',0) Evolutions

code2utf('8211',0) Les offres Saas BigData comme Google BigQuery.

code2utf('8211',0) Les limites. Les nouveautés annoncées

