Pythagore F.D.

*Apprendre à apprendre*

**CB001**

# BigData Architecture et technologies

**Durée:** 2 jours

| | | |
|---|---|---|
| 1130 € | 15 au 16 février<br>4 au 5 mai | 13 au 14 septembre<br>6 au 7 décembre |

## Public:

Chefs de projets, architectes, développeurs, data-scientists, et toute personne souhaitant connaître les outils et solutions pour concevoir et mettre en oeuvre une architecture BigData.

## Objectifs:

Comprendre les concepts essentiels du BigData, et les technologies implémentées. Savoir analyser les difficultés propres à un projet BigData, les freins, les apports, tant sur les aspects techniques que sur les points liés à la gestion du projet.

## Connaissances préalables nécessaires:

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

## Programme:

Introduction :
- 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.
- Quelques éléments d'architecture.
- L'écosystème du BigData : les acteurs, les produits, état de l'art.
- Cycle de vie des projets BigData.
- Emergence de nouveaux métiers : Datascientists, Data labs, ...

Stockage :
- Caractéristiques NoSQL : adaptabilité, extensibilité,
structure de données proches des utilisateurs, développeurs
- Les types de bases de données : clé/valeur, document, colonne, graphe.
- 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,
- Quelques exemples de produits et leurs caractéristiques : Cassandra, MongoDB, CouchDB, DynamoDB, Riak, Hadoop, HBase, BigTable, ...
- Qualité des données, gouvernance de données.

Pythagore F.D.

*Apprendre à apprendre*

**CB001**

# BigData Architecture et technologies

| | | |
|---|---|---|
| Indexation et recherche | : | Moteurs de recherche.Principe de fonctionnement.<br>Méthodes d'indexation. Mise en oeuvre avec elasticsearch.<br>Exemple de Lucene/solr.<br>Recherche dans les bases de volumes importants.<br>Exemples de produits et comparaison : Dremel, Drill, ElasticSearch, MapReduce, |
| Calcul et restitution, intégration | : | Différentes solutions : calculs en mode batch, ou en temps réel, sur des flux de données ou des données statiques.<br>Les produits : langage de calculs statistiques, R Statistics Language, sas, RStudio.<br>Ponts entre les outils statistiques et les bases BigData<br>Outils de calcul sur des volumes importants : storm en temps réel, hadoop en mode batch.<br>Zoom sur Hadoop : complémentarité de HDFS et MapReduce.<br>Restitution et analyse : logstash, kibana, elk, pentaho<br>Présentation de pig pour la conception de tâches MapReduce sur une grappe Hadoop. |