

## **BigData : architecture et technologies**

Pour bien 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., Pythagore F.D. organise une formation «BigData architecture et technologies» **du 14 au 15 février 2018**

Le point essentiel des formations Pythagore F.D. est l'utilisation d'une infrastructure adaptée au BigData : le matériel pédagogique mis à disposition des participants permet de mettre en pratique tous les concepts abordés (clusters de calcul pour MapReduce, baies de stockage distribué, jeux de données pour les tests de performances).

Le programme de la formation est le suivant :

### **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.

### **Indexation et recherche**

Moteurs de recherche.Principe de fonctionnement.

Méthodes d'indexation. Mise en oeuvre avec elasticsearch.

Exemple de Lucene/solr.

Recherche dans les bases de volumes importants.

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.

Les produits : langage de calculs statistiques, R Statistics Language, sas, RStudio.

Ponts entre les outils statistiques et les bases BigData

Outils de calcul sur des volumes importants : storm en temps réel, hadoop en mode batch.

Zoom sur Hadoop : complémentarité de HDFS et MapReduce.

Restitution et analyse : logstash, kibana, elk, pentaho

Présentation de pig pour la conception de tâches MapReduce sur une grappe Hadoop.

---

Pythagore F.D. est un centre de formation, dans les domaines suivants :

- les bases de données et le BigData avec NoSQL, Cassandra, MongoDB, Hadoop, Spark, Storm ...
- la virtualisation et l'orchestration avec xen, kvm, lxc, Docker, et le cloud : cloudstack et openstack, openNebula, cobbler, etc ...

- TCP/IP (IPv6, snmp, Architecture, Sécurité, Administration de réseaux IP, VoIP, ...)
- Unix et Linux, et les applicatifs Apache, Openldap, Squid, Nagios, Zabbix, OCS/GLPI, puppet , chef...
- Développement (langage C, Java, Jee, technologies Jee, JBoss, WebServices, PHP, Perl, Python , ...)

Les formations sont dispensées soit dans les locaux de la société à Paris, soit sur site client, ou à distance en classes virtuelles.

Lieu : centre de formation Pythagore F.D., Paris9è

Pour toute demande d'information, contactez nos conseillers au 33 (0)1 55 33 52 10, ou par mail à l'adresse [pdf@pythagore-fd.fr](mailto:pdf@pythagore-fd.fr), ou sur le site [www.pythagore-fd.fr](http://www.pythagore-fd.fr)