



# BigData, architecture et technologies

Mise à jour nov. 2023

**Durée** 2 jours (14 heures )

« Délai d'accès maximum 1 mois »

25 nov.-26 nov.

07 oct.-08 oct.

28 oct.-29 oct.

26 août-27 août

Nantes / Rennes : 1155 € HT

Brest / Le Mans : 1155 € HT

Certification : NON

## OBJECTIFS PROFESSIONNELS

- Comprendre les concepts du BigData et connaître les technologies associées.

## PARTICIPANTS

- 

## PRE-REQUIS

- PUBLIC : Chefs de projets, architectes, et toute personne souhaitant connaître les outils et solutions pour mettre en place une architecture BigData.
- PREREQUIS : bonne culture générale sur les systèmes d'information.

## MOYENS PEDAGOGIQUES

- Réflexion de groupe et apports théoriques du formateur
- Travail d'échange avec les participants sous forme de
- Utilisation de cas concrets issus de l'expérience professionnelle
- Validation des acquis par des questionnaires, des tests d'évaluation, des mises en situation et des jeux pédagogiques.
- Remise d'un support de cours.

## MODALITES D'EVALUATION

- Feuille de présence signée en demi-journée,
- Evaluation des acquis tout au long de la formation,
- Questionnaire de satisfaction,
- Positionnement préalable oral ou écrit,
- Evaluation formative tout au long de la formation,
- Evaluation sommative faite par le formateur ou à l'aide des certifications disponibles,
- Sanction finale : Certificat de réalisation, certification éligible au RS selon l'obtention du résultat par le stagiaire

## MOYENS TECHNIQUES EN PRESENTIEL

- Accueil des stagiaires dans une salle dédiée à la formation, équipée d'ordinateurs, d'un vidéo projecteur d'un tableau blanc et de paperboard. Nous préconisons 8 personnes maximum par action de formation en présentiel

## MOYENS TECHNIQUES DES CLASSES EN CAS DE FORMATION DISTANCIELLE

- A l'aide d'un logiciel comme Teams, Zoom etc... un micro et éventuellement une caméra pour l'apprenant,
- suivez une formation uniquement synchrone en temps réel et entièrement à distance. Lors de la classe en ligne, les apprenants interagissent et communiquent entre eux et avec le formateur.
- Les formations en distanciel sont organisées en Inter-Entreprise comme en Intra-Entreprise.
- L'accès à l'environnement d'apprentissage (support de cours, labs) ainsi qu'aux preuves de suivi et d'assiduité (émargement, évaluation) est assuré. Nous préconisons 4 personnes maximum par action de formation en classe à distance

## ORGANISATION

- Les cours ont lieu de 9h à 12h30 et de 14h à 17h30.

## PROFIL FORMATEUR

- Nos formateurs sont des experts dans leurs domaines d'intervention
- Leur expérience de terrain et leurs qualités pédagogiques constituent un gage de qualité.

## A L'ATTENTION DES PERSONNES EN SITUATION DE HANDICAP

- Les personnes atteintes de handicap souhaitant suivre cette formation sont invitées à nous contacter directement, afin d'étudier ensemble les possibilités de suivre la formation.

## Programme de formation

### Introduction (03h30)

- Le besoin : volumes importants de données, traitements optimisés de flux de données au fil de l'eau (nouvelles technologies et nouveaux usages)
- Domaines concernés : recherche scientifique, médical, e-commerce, sécurité...
- Développement des techniques sur différents aspects : stockage, indexation/recherche, calcul
- Définition de ETL : Extract Transform Load
- Les acteurs aujourd'hui

### Evolutions (01h30)

- Les offres Saas BigData, type Google BigQuery
- Les limites actuelles. Les avancées annoncées

### Stockage (05h00)

- 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...
- Les bases de données
- Quelques exemples de produits et leurs caractéristiques : cassandra, MongoDB, CouchDB, DynamoDB...

### Indexation et recherche (04h15)

- Moteurs de recherche
- Principe de fonctionnement
- Méthodes d'indexation
- Exemple de Lucene, et mise en oeuvre avec solr
- Recherche dans les bases de volumes importants
- Exemples de produits et comparaison : dremel, drill, elasticsearch, MapReduce...

### Calcul et restitution, intégration (03h00)

- 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
- Outils de calcul sur des volumes importants : storm en temps réel, hadoop en mode batch
- Zoom sur Hadoop : complémentarité de HDFS et MapReduce