

# Bigdata : architecture et technologies

## INFORMATIONS GÉNÉRALES

**Type de formation :** Formation continue

**Éligible au CPF :** Non

**Domaine :** IA, Big Data et Bases de données

**Action collective :** Non

**Filière :** Big Data

**Rubrique :** Fondamentaux

**Code de formation :** BD001

## PRÉSENTATION

### Objectifs & compétences

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

### Public visé

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

### Module 3 : Indexation et recherche

- 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,

### Module 4 : 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
- Outils de calcul sur des volumes importants : storm en temps réel, hadoop en mode batch.
- Zoom sur Hadoop : complémentarité de HDFS et MapReduce.

## € Tarifs

**Prix public :** 1270 €

### Tarif & financement :

Nous vous accompagnons pour trouver la meilleure solution de financement parmi les suivantes :

**Le plan de développement des compétences de votre entreprise :** rapprochez-vous de votre service RH.

**Le dispositif FNE-Formation.**

**L'OPCO** (opérateurs de compétences) de votre entreprise.

**France Travail:** sous réserve de l'acceptation de votre dossier par votre conseiller Pôle Emploi.

**CPF -MonCompteFormation**

Contactez nous pour plus d'information : [contact@aston-institut.com](mailto:contact@aston-institut.com)

## 📍 Lieux & Horaires

**Durée :** 14 heures

### Délai d'accès :

Jusqu'à 8 jours avant le début de la formation

## 📅 Prochaines sessions

Consultez-nous pour les prochaines sessions.

- Evolutions
  - Les offres Saas BigData comme Google BigQuery.
  - Les limites.
- Les nouveautés annoncées

## MODALITÉS

### Modalités

Jusqu'à 8 jours avant le début de la formation, sous condition d'un dossier d'inscription complet

### Méthode

**Fin de formation :** entretien individuel.

**Satisfaction des participants :** questionnaire de satisfaction réalisé en fin de formation.

**Assiduité :** certificat de réalisation.

**Validations des acquis :** grille d'évaluation des acquis établie par le formateur en fin de formation.