

# AWS : BigData avec Hadoop EMR

## INFORMATIONS GÉNÉRALES

**Type de formation :** Formation continue

**Éligible au CPF :** Non

**Domaine :** Systèmes et Réseaux

**Action collective :** Non

**Filière :** Cloud computing

**Rubrique :** AWS

**Code de formation :** VC203

## € Tarifs

**Prix public :** 1900 €

### 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)

## PRÉSENTATION

### Objectifs & compétences

Savoir mettre en oeuvre les techniques de calcul distribué avec Hadoop EMR

### Public visé

Architectes, chefs de projets souhaitant bénéficier des services offerts par la distribution Hadoop fournie par AWS

### Pré-requis

Connaissance des fondamentaux du cloud

## Lieux & Horaires

**Durée :** 21 heures

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

## PROGRAMME

Présentation AWS Hadoop EMR

Rappels des services de base : EC2 (Elastic Compute Cloud), VPC (Virtual Private Cloud), S3 (Simple Storage Service), EBS (Elastic Block Storage), RDS (Relational Database Service)

Caractéristiques du calcul distribué et du service EMR (Amazon Elactis MapReduce)

La distribution Hadoop EMR

Historique du projet hadoop

Les fonctionnalités : stockage, outils d'extraction, de conversion, ETL, analyse, ...

Exemples de cas d'utilisation sur des grands projets.

Les principaux composants :HDFS pour le stockage et YARN pour les calculs.

Les distributions et leurs caractéristiques.

Composants de la distribution Hadoop EMR

Mise en œuvre

Démonstrations sur une architecture Hadoop multi-noeuds.

TP : mise en place d'une configuration de base avec HBase.

Cycle de fonctionnement.

Transferts par s3.

Envoi des travaux.

Visualisation des résultats.

Suivi des travaux avec Hue.

Interactivité avec les Notebooks : jupyter.

Configuration des composants de la distribution.

Exploitation

Gestion des événements avec Events.

Contrôle du réseau et des VPC.

Automatisation de l'exécution sur AWS depuis un poste local.

Suivi distant des travaux.

Journalisation, visualisation des logs.Utilisation de la ferme aws en débordement d'une ferme locale.

Sécurité : mise en place d'une configuration de sécurité.

Liens avec IAM.

Optimisation

Analyse des performances.

Déploiement d'une configuration avec Spark.

Evaluation des performances par rapport à une ferme locale.

Comparaison des coûts.

## Prochaines sessions

Consultez-nous pour les prochaines sessions.

Mise en place d'automates d'optimisation.

## MODALITÉS

### Modalités

**Modalités :** en présentiel, distanciel ou mixte . Toutes les formations sont en présentiel par défaut mais les salles sont équipées pour faire de l'hybride. – Horaires de 9H à 12H30 et de 14H à 17H30 soit 7H – Intra et Inter entreprise.

**Pédagogie :** essentiellement participative et ludique, centrée sur l'expérience, l'immersion et la mise en pratique. Alternance d'apports théoriques et d'outils pratiques.

**Ressources techniques et pédagogiques :** Support de formation au format PDF ou PPT Ordinateur, vidéoprojecteur, Tableau blanc, Visioconférence : Cisco Webex / Teams / Zoom.

**Pendant la formation :** mises en situation, autodiagnostic, travail individuel ou en sous-groupe sur des cas réels.

### 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.