

Plan de cours N° : 1146

Durée : 3 jours (21h)

#### **PARTICIPANTS / PRE-REQUIS**

Architecte, chef de projets souhaitant bénéficier des services offerts par la distribution Hadoop fournie par AWS.  
Connaissance des fondamentaux du cloud/

#### **OBJECTIFS PEDAGOGIQUES**

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

#### **MOYENS PEDAGOGIQUES**

Réflexion de groupe et apports théoriques du formateur  
Travail d'échange avec les participants sous forme de réunion-discussion  
Utilisation de cas concrets issus de l'expérience professionnelle  
Validation des acquis par des exercices de synthèse  
Alternance entre apports théoriques et exercices pratiques (en moyenne 30 et 70%)  
Remise d'un support de cours.  
Assistance post-formation d'une durée de 1 an sur le contenu de la formation via notre adresse mail dédiée [formateurs@atp-formation.com](mailto:formateurs@atp-formation.com)

#### **MOYENS PERMETTANT LE SUIVI DE L'EXECUTION ET DES RESULTATS**

Feuille de présence signée en demi-journée,  
Evaluation des acquis tout au long de la formation,  
Questionnaire de satisfaction,  
Attestation de stage à chaque apprenant,  
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.

#### **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.

#### **MOYENS TECHNIQUES DES CLASSES A DISTANCE**

A l'aide d'un logiciel comme Teams, Zoom etc... un micro et éventuellement une caméra pour l'apprenant, suivez une formation 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.

Les participants recevront une convocation avec lien de connexion

Pour toute question avant et pendant le parcours, une assistance technique et pédagogique est à disposition par mail et par téléphone auprès de notre équipe par téléphone au 04.76.41.14.20 ou par mail à [contact@atp-formation.com](mailto:contact@atp-formation.com)

#### **ORGANISATION**

Les cours ont lieu de 9h00-12h30 13h30-17h00.

#### **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.

#### **ACCESSIBILITE**

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.

#### **MISE A JOUR**

01/01/2021

Plan de cours N° : 1146

Durée : 3 jours (21h)

### 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
  - YARN pour les calculs
- Les distributions et leurs caractéristiques
  - Composants de la distribution Hadoop EMR

### Mise en oeuvre

- Démonstrations sur une architecture Hadoop multi-noeuds
- Travaux pratiques
  - 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 évènements 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
- Mise en place d'automates d'optimisation