

Plan de cours N° : 1144

Durée : 3 jours (21h)

PARTICIPANTS / PRE-REQUIS

Chef de projet informatique, data scientist, statisticien, développeur souhaitant comprendre les apports de R pour l'analyse des données et savoir l'intégrer à un environnement Hadoop.

Notions de calculs statistiques.

OBJECTIFS PEDAGOGIQUES

Connaître les principales fonctions statistiques du langage R. Savoir utiliser des programmes R dans un environnement Big Data en s'appuyant sur le système distribué HDFS (Hadoop Distributed File System).

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

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

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° : 1144

Durée : 3 jours (21h)

Présentation R

- Le projet R Programming
- Calculs statistiques et génération de graphiques
- Points forts de R Programming
- Besoins du BigData
- Positionnement R programming par rapport à Hadoop

Mise en oeuvre de R

- Travaux pratiques
Installation et tests sur une plate-forme CentOS
- Utilisation de R en mode commande
- Commandes de base
Syntaxe
- Opérations de base
Expression
- Manipulations de nombres, vecteurs, tableaux, matrices, listes...

Tableaux et matrices

- Déclaration, dimensionnement, indexation
- Opérations de base
Produit de tableaux, transposition, produits de matrices
- Matrices
Equations linéaires, inversion, valeur propre, vecteur propre, déterminant, moindre carré...

Liste et DataFrames

- Définitions
Cas d'utilisation
- Attachement, détachement
- Chargement d'un dataframe
- La fonction scan

Statistiques

- Distributions embarquées
Uniforme, normale, poisson, exponentielle
- Calculs statistiques
Modèles statistiques
- Affichage en graphes, histogrammes

Import/export

- Formats texte, csv, xml, binaire, largeur fixe, images (jpeg, png)
Encodage. Filtrage
- Importation SQL
Importation depuis un socket réseau
- Travaux pratiques
Importation de données géodésiques et export au format Json

Intégration Hadoop

- Association de la puissance du calcul distribué fourni par les outils hadoop
- Différents moyens d'intégration
sparkR, RHbase, RHDFS, RHadoop, rmr2 pour utiliser le système distribué hdfs depuis R, pour accéder à HBase depuis les programmes en R.
- Transformation d'un dataframe R en un dataframe Spark
- Travaux pratiques avec Hadoop

Fonctions spécifiques

- Définition de nouvelles fonctions
Appels. Passage d'argument
- Construction d'une bibliothèque
- Diffusion, installation avec R CMD INSTALL.

Evolutions

- Les acteurs
IBM avec BigInsights, Revolution R avec ScaleR