Pythagore F.D.

Apprendre à apprendre

CB052

Environnement R, analyse de données

Durée: 3 jours

1980 €	26 au 28 février	17 au 19 septembre
	2 au 4 mai	14 au 16 novembre

Public:

Chefs de projet, data scientists, statisticiens, développeurs souhaitant comprendre les apports de R pour l'analyse des données, et savoir l'intégrer à un environnement Hadoop.

Objectifs:

Connaître les principales fonctions statistiques de R, et savoir utiliser des programmes R dans un environnement BigData, en s'appuyant sur le système distribué hdfs.

Connaissances préalables nécessaires:

Notions de calculs statistiques

Programme:

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. Expressions.
Manipulations de nombres, vecteurs, tableaux, matrices.listes, etc ..

Tableaux et matrices : Déclaration, dimensionnement, indexation.
Opérations de base : produit de tableaux, transposition, produits de matrices.
Matrices : équations linéaires, inversion, valeur propre, vecteur propre, déterminant, moindre carré, ...

Pythagore F.D.

Apprendre à apprendre

CB052

Environnement R, analyse de données

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