

R R – Big Data Spark

REF : SIDS204

DUREE : 7h

Présentiel Classe virtuelle

PUBLIC

Chefs de projets statistiques, chargés d'étude, analystes ...

Modalités et délais d'accès : les inscriptions sont fermées 24h avant la 1ère journée de formation.

Accessibilité : Si vous avez des contraintes particulières liées à une situation de handicap, veuillez nous contacter au préalable afin que nous puissions, dans la mesure du possible, adapter l'action de formation.

PREREQUIS

Connaissance des bases de la théorie statistique,

Avoir suivi la formation R niveau 1 ou avoir une utilisation avancée de R.

MODALITES PEDAGOGIQUES

8 à 10 personnes maximum par cours

1 poste de travail par stagiaire

Remise d'une documentation pédagogique papier ou numérique pendant le stage

La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience

MODALITES D'EVALUATION

Auto-évaluation des acquis par le stagiaire via un questionnaire en ligne

Attestation de fin de stage remise au stagiaire

OBJECTIFS PEDAGOGIQUES

L'objectif de la journée est de:

- développer les compétences de l'utilisation du langage R dans une architecture Big Data

Pour cette journée, nous étudierons les environnements Spark et H2O.

PROGRAMME

- Introduction : Data Mining vs Machine Learning vs Deep Learning
- Principes de Spark, utilisation dans R avec SparkR
- Principe de H2O, utilisation dans R avec Sparkling Water : R + H2O
- Mise en place d'un projet de Machine Learning dans cet environnement
- Cas pratique

Version du : 12/01/2023



