

retengr

Formation Apache Spark

Durée : 4 jours – Réf : SPARK



- 40 rue des Filatiers 31000
- Organisme de formation : 73 31 08023 31
- SIRET : 812 673 903 00017

Méthode pédagogique

La formation se passe en mode présentiel (face à face), et se compose de 70% de travaux pratiques (Mises en situation, débats, exercices). Le formateur alterne entre théorie et pratique ; chaque participant dispose d'un ordinateur sur lequel il réalisera les exercices proposés.

Une évaluation quotidienne de l'acquisition des connaissances de la veille est effectuée.

Une synthèse est proposée en fin de formation. Une évaluation à chaud sera proposée au stagiaire à la fin du cours.

Un support de cours sera remis à chaque participant comprenant les slides sur la théorie, les exercices et les corrigés de ces derniers. Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

Présentation

La formation se passe en mode présentiel (face à face), et se compose de 70% de travaux pratiques (Mises en situation, débats, exercices). Le formateur alterne entre théorie et pratique ; chaque participant dispose d'un ordinateur sur lequel il réalisera les exercices proposés.

Une évaluation quotidienne de l'acquisition des connaissances de la veille est effectuée.

Une synthèse est proposée en fin de formation. Une évaluation à chaud sera proposée au stagiaire à la fin du cours.

Un support de cours sera remis à chaque participant comprenant les slides sur la théorie, les exercices et les corrigés de ces derniers. Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.



Objectifs

- Concevoir une application avec Spark
- Comprendre le principe de distribution des traitements
- Maîtriser les concepts fondamentaux des et des Resilient Distributed Dataset
- Utiliser les dataframes via Spark SQL
- Utiliser SparkUI afin d'analyser les jobs et tâches de Spark
- Positionner SparkML dans un contexte de data science
- Traiter des données en continu avec Spark Streaming

Audience

Architectes, Développeurs, Chefs de projets, Datamine

Pré-requis

Connaissance d'un langage de programmation

Le formateur

Le formateur est un expert du domaine qui intervient sur le sujet depuis plusieurs années en formation mais aussi en conseil.

Doté d'une grande qualité d'écoute, sa pédagogie et sa compétence technique vous permettront d'acquérir les compétences sur SPARK.

Il saura alterner entre théorie, pratique, et retours d'expérience.

Tarif

1980 euros HT



Programme

Présentation de Spark

- Spark : un besoin de distribuer vos traitements
- Architecture de Spark runtime : driver, executor, master
- Positionner Spark vs Hadoop
- Les langages du framework : Java | Scala | Python | R

RDD : Resilient Distributed Dataset

- RDD : Le composant fondateur du fonctionnement de Spark
- Les partitions : la base de la distribution
- Transformations, actions et directed acyclic Graph
- Manipuler un RDD : Une API riche
- Le cas particulier des Pairs RDD

SparkSQL, Dataframes et Datasets

- Un modèle de programmation haut niveau
Initialisation d'un dataframe
- Manipulation : sélection, tri et fonctions d'agrégation.
- Dataset : une surcouche typée des dataframes
- Comprendre le plan d'exécution d'une requête
- Bonnes et mauvaises pratiques avec SparkSQL

Mise en cluster : Les infrastructures de déploiement

- Les composants d'une exécution Spark : Jobs, stages et tasks
- Un principe important : Data locality
- Distribution des données dans le cadre d'un cluster : les partitions
- Redistribution des données : le shuffle
- Bonnes pratiques et performance



- 40 rue des Filatiers 31000
- Organisme de formation : 73 31 08023 31
- SIRET : 812 673 903 00017

Machine Learning

- Comprendre les principes fondamentaux du Machine Learning
- Apprentissage et création d'un modèle avec SparkML

Spark streaming

- Collecte et traitement des données en continu
- Stream processing avec Spark
- Comprendre le principe du micro-batching

The background is a vibrant, abstract composition. It features large, overlapping organic shapes in shades of purple, yellow, and red. A prominent yellow shape on the left transitions into a red shape on the right. In the upper right, there is a pattern of small yellow dots arranged in a grid that tapers off. At the bottom, there are more organic shapes in light pink and teal. A dashed blue line with a circular loop is visible in the lower-left quadrant.

Vous allez nous adorer si
comme nous vous pensez que...

Une formation doit être au service de la performance du collaborateur et de l'entreprise

Ceci nécessite une quête constante d'excellence de la part de l'organisme formateur avec une adaptation systématique aux enjeux de l'entreprise, la mise à jour régulière des supports de cours et une veille technologique indispensables pour toujours être à la pointe du domaine.



L'expertise technique est aussi importante que les qualités pédagogiques



Nos formateurs sont tous des experts de leur domaine. Mais qu'ont-ils de plus que les autres ? Nous les sélectionnons en plus pour leurs qualités de pédagogue et leurs méthodes d'enseignements. Nous plaçons les qualités pédagogiques au même niveau que l'expertise afin que nos stagiaires tirent le meilleur de leurs formations.

L'excellence naît de l'excellence

Beaucoup de nos clients se classent parmi les leaders de leurs industries respectives ou parmi les start-ups les plus prometteuses. Nous savons que former les collaborateurs de telles entreprises nécessite de prêter attention à chaque détail en prodiguant un accompagnement à la hauteur de l'ambition de nos stagiaires. C'est pourquoi nous savons faire des leaders d'aujourd'hui les champions de demain !



retengr

Faire du leader
d'aujourd'hui, le champion
de demain

www.retengr.com

05.82.95.70.19

36 rue d'Alsace-Lorraine, 31000 Toulouse