

Objectifs

Cette formation intensive de 4 jours permettra aux stagiaires de se familiariser avec Spark et de découvrir les principales librairies de la solution (Streaming, SQL, Machine Learning, GraphX).

Cette formation laisse une large place aux exercices pratiques: les stagiaires apprendront notamment à analyser des données Twitter ou encore à construire leur premier moteur de recommandation.



Modalités d'entrée en formation

La date limite de l'inscription est de 5 jours ouvrés avant le début de la formation.

Objectifs pédagogiques et d'évaluation

Cette formation intensive de 4 jours permettra aux stagiaires de se familiariser avec Spark et de découvrir les principales librairies de la solution (Streaming, SQL, Machine Learning, GraphX).

Cette formation laisse une large place aux exercices pratiques: les stagiaires apprendront notamment à analyser des données Twitter ou encore à construire leur premier moteur de recommandation.

Méthodes pédagogiques

- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques et de réflexions

Parcours pédagogique

Introduction

- Présentation de SPARK dans l'écosystème big data

Spark Fondamentaux

- présentation des RDD et des fonctions de type MAP, REDUCE et ACTION
- explication d'un pgm de comptage de mots d'un fichier texte
- exercice : à partir de l'exemple précédent, analyse de tweets par pays

Déploiement Spark

- utilisation de SPARK-SUBMIT pour lancer un traitement en batch
- exercice : déployer un pgm de streaming

Spark SQL

- présentation des datasets et des dataframes et des fonctions de bases associées à ces objets

- exercice : exploration de données via le notebook Zeppelin et les datasets

Spark Machine Learning

- mise en place d'un moteur de recommandation (algorithme ALS)
- mise en place d'un modèle de clustering (algorithme K-MEANS)
- exercice : mise en place de d'un moteur de recommandation pour des films

Spark Streaming

- présentation du streaming et des commandes associées
- exercice : faire un wordcount sur un flux réseau

Spark GRAPHX

- présentation des graph et de l'API de Pregel
- exercice : analyse d'un jeu de données de l'aviation civile américaine

Spark Performance

- présentation des notions de shuffling, de data locality, de persistance et de broadcast des données

Modalités de suivi

Modalités d'évaluation :

Les modalités d'évaluation s'articulent autour d'exercices pratiques, d'un questionnaire à choix multiples et d'un projet opérationnel :

- Les compétences les plus conceptuelles et théoriques font l'objet uniquement de questionnaires à choix multiples ;

Digital Services School

Data . AI . Generative AI

- Les compétences les plus pratiques font l'objet :

o d'exercices qui permettent de valider que l'apprenant sait les mettre en application. Ainsi, ces compétences sont évaluées à la fois par des questionnaires à choix multiples et des travaux pratiques ;

o d'un projet opérationnel permettant de mettre en pratique et d'articuler l'ensemble des compétences

Exercices pratiques

Les exercices pratiques sont réalisés pendant l'action de formation.

Le résultat attendu est la validation de la réalisation intégrale et complète de chaque exercice.

Finir l'exercice pratique permet de valider que la compétence évaluée est acquise

Questionnaire à choix multiples

Les questions permettent à l'apprenant de valider qu'il a bien assimilé les compétences techniques.

Pour chaque question :

- 4 réponses possibles

- 1 seule bonne réponse

Obtenir un score supérieur à la moyenne permet de valider que la compétence évaluée est acquise

Projet opérationnel

Au début de la formation, les candidats recevront une information sur le projet à réaliser, la constitution des groupes et le planning de réalisation du projet. Le thème du projet sera communiqué pendant la formation lorsque les candidats auront abordé les compétences nécessaires à sa réalisation.

Méthodes et modalités d'évaluation

Une note globale égale à 70% de la note obtenue au QCM + 30% de la note obtenue au projet pédagogique supérieure à 10/20

Accomplissement des exercices pratiques de chacun des modules

Durée

28.00 Heures **4** Jours

Descriptif détaillé RNCP / RS - France Compétences

Maîtriser le fonctionnement de l'écosystème Spark

