

Formation Scala pour la Data Science : Fondamentaux et Préparation Spark

Présentation

Cette formation de trois jours est conçue pour introduire les Data Scientists à Scala, en mettant l'accent sur son utilisation dans un environnement Spark. Le premier jour couvre les bases de Scala, y compris sa syntaxe et la manipulation de données sur Databricks. Le deuxième jour explore les paradigmes fonctionnels et impératifs, ainsi que la modélisation des données et la gestion des erreurs. Le troisième jour prépare les participants à utiliser Scala avec Spark, en structurant des notebooks et en travaillant sur un projet final de pipeline data science. Chaque journée inclut des exercices pratiques pour renforcer les compétences acquises et préparer les participants à des projets réels.

Durée : 21,00 heures (3 jours)

Tarif INTRA : [Nous consulter](#)

Objectifs de la formation

- Acquérir les bases de Scala pour une utilisation efficace avec Spark.
- Comprendre les spécificités du langage par rapport à Python, notamment dans le contexte data science.
- Être capable d'écrire du code Scala propre, idiomatique, prêt à être utilisé avec Spark.

Prérequis

Connaissance solide de Python ou un langage du type Java, notamment dans un contexte notebook et data science

Public

Le public visé pour cette formation comprend principalement les data scientists, les chefs de projet et les développeurs.



Cette formation est également adaptée aux statisticiens, consultants Big Data, et data analysts.

Programme de la formation

Jour 1 – Introduction à Scala pour Data Scientists

1. **Contexte et positionnement de Scala**
 - Historique et cas d'usage
 - Intérêt de Scala dans un environnement Spark
 - Scala vs Python pour la data science
2. **Syntaxe de base Scala**
 - Variables, types, expressions
 - Fonctions : déclaration, expression lambda
 - Structures de contrôle (if, match, for, while)
3. **Collections et manipulation de données**
 - List, Set, Map, Tuple
 - Immutabilité, transformation avec map, filter, flatMap
 - Comparaison avec les outils Python : pandas, list comprehension, etc.
4. **Premiers pas sur Databricks (support de la formation)**
 - Notebooks Scala
 - Chargement de fichiers simples (CSV, JSON)
 - Affichage de données tabulaires

Exercices pratiques : mini-notebooks de manipulation de données (prétraitement, parsing, regroupement)

Jour 2 – Paradigmes et structuration du code Scala

1. **Fonctionnel vs impératif**
 - Réduction des effets de bord
 - Fonctions pures, immutabilité, fonctions d'ordre supérieur
 - Illustration sur des transformations de données
2. **Pattern matching avancé**
 - Cas d'usage sur des structures de données complexes (JSON, vecteurs, etc.)
 - Comparaison avec les structures conditionnelles Python
3. **Modélisation avec classes et objets**
 - case class pour modéliser des données typées

- companion object, méthodes utilitaires
- Introduction à la POO en Scala
- 4. **Gestion fonctionnelle des erreurs**
 - Option, Some, None dans le prétraitement
 - Try, Success, Failure dans les chargements et conversions de données

Exercices pratiques : pipeline de préparation de données (nettoyage, enrichissement, typage)

Jour 3 – Scala pour la data science avec Spark (préparation)

1. **Structuration de notebooks Scala pour Spark**
 - Organisation du code (fonctions réutilisables, séparations logiques)
 - Bonnes pratiques de structuration des notebooks
2. **Lecture/écriture de données**
 - Chargement de données CSV, JSON, Parquet
 - Préparation des données avant modélisation
 - Export de résultats typés
3. **Interopérabilité et pratique mixte Scala/Python (si utile)**
 - Notebooks mixtes via %python, %scala
 - Cas d'usage : Scala pour les prétraitements, Python pour les modèles
4. **Projet final – Atelier Scala data science (sans Spark API avancée)**
 - Objectif : simuler un pipeline data science complet (chargement, nettoyage, enrichissement)
 - Structuration idiomatique Scala pour une future utilisation avec Spark

Exercice final : pipeline Scala préparatoire à un traitement Spark distribué

Organisation

Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes. Riches de leur expérience sur le sujet, ils sauront accompagner vos collaborateurs dans leur montée en compétences.

Moyens pédagogiques et techniques

- Alternance d'apports théoriques et de mises en pratique contextualisées.
- Progression pédagogique structurée, de la compréhension à la mise en œuvre opérationnelle.
- Pédagogie immersive favorisant une mise en application immédiate des compétences.
- Formation interactive et participative s'appuyant sur des ateliers pratiques, études de cas et échanges collectifs (+ 60 % du temps de formation.)
- Séquences pédagogiques rythmées et évaluations formatives régulières pour renforcer l'ancrage des apprentissages.
- Supports de cours, ressources documentaires et références mis à disposition par le formateur

Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation**En amont de la formation** *(évaluation en ligne dans l'extranet apprenant)*

- Recueil des besoins et des attentes des participants.
- Auto-positionnement permettant d'évaluer le niveau initial des apprenants.

Tout au long de la formation

- Évaluations continues réalisées par le formateur à travers des questions orales, exercices, QCM, études de cas et mises en situation pratiques.

À l'issue de la formation *(évaluations en ligne dans l'extranet apprenant)*

- Auto-positionnement afin de mesurer l'acquisition des compétences visées.
- Validation de l'acquisition des compétences par le formateur
- Questionnaire de satisfaction « à chaud » à l'issue de la formation
- Questionnaire de satisfaction "à froid", 45 jours après la formation afin d'évaluer l'ancrage des acquis et leur mise en application en situation professionnelle.

Accessibilité

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.