CODE
7005

TECHNOLOGIES DU NUMÉRIQUE ET SYSTÈME D'INFORMATION / BASES DE DONNÉES - BIG DATA - BLOCKCHAIN

NOUVEAU **POSSIBLE EN INTRA**

APACHE SPARK : DÉVELOPPEMENT ET OPTIMISATION BIG DATA

Face à l'explosion des volumes de données, Apache Spark s'impose pour le traitement distribué, en batch comme en temps réel. Cette formation vise à passer de traitements ponctuels à des pipelines industrialisés en production.

2160 € HT

3 JOURS (21 H.)

VILLEURBANNE
DU 08/06/2027 AU 10/06/2027

COMPÉTENCE PRINCIPALE VISÉE

Concevoir, développer et optimiser des traitements Big Data distribués avec Apache Spark afin de les intégrer dans des environnements de production.



LES + DE LA FORMATION

60 % de la formation sont consacrés à des travaux pratiques et à une étude de cas fil rouge permettant de concevoir un pipeline Big Data complet, depuis l'ingestion des données jusqu'au déploiement en production

PUBLIC

- Cette formation s'adresse aux professionnels de la data qui souhaitent passer de traitements ponctuels à des pipelines de données industrialisés, fiables et reproductibles.

PRÉREQUIS

- Bonne maîtrise de Python ou Scala
- Connaissances de SQL
- Notions de traitement distribué ou de Big Data

OBJECTIFS PÉDAGOGIQUES

- Développer des traitements distribués performants avec Apache Spark et Spark SQL
- Optimiser les performances et l'utilisation des ressources d'un cluster Spark
- Concevoir et déployer des pipelines Big Data robustes intégrés à une plateforme Data moderne

OBJECTIFS PÉDAGOGIQUES

- Développer des traitements distribués performants avec Apache Spark et Spark SQL
- Optimiser les performances et l'utilisation des ressources d'un cluster Spark
- Concevoir et déployer des pipelines Big Data robustes intégrés à une plateforme Data moderne

97,2%
de clients satisfaits*
* enquête réalisée auprès de nos clients en septembre 2025

CONTENU

PARTIE 1 - L'architecture Spark et le traitement distribué

- Les principes fondamentaux de Spark et son architecture
  - Positionnement de Spark dans l'écosystème Big Data
  - Architecture Driver / Executors
  - Gestion des partitions
  - DAG Scheduler
  - Tolérance aux pannes
  - Gestion de la mémoire
  - Exécution distribuée

TRAVAUX PRATIQUES - Installation d'un environnement Spark

- Analyse d'un premier traitement distribué
- Visualisation de l'exécution avec Spark UI

PARTIE 2 - Traitements Big Data performants

- Les pipelines de transformation de données distribués.
  - DataFrames
  - Spark SQL
  - Lecture et écriture de données
  - Formats Parquet, JSON et CSV
  - Jointures distribuées
  - Agrégations
  - Fonctions de fenêtres
  - Catalyst Optimizer

TRAVAUX PRATIQUES - Développement d'un pipeline ETL distribué

- Traitement d'un jeu de données volumineux
- Optimisation d'une requête Spark SQL

PARTIE 3 - Les performances en production

- Goulots d'étranglement et performances des traitements
  - Partitionnement
  - Broadcast Join
  - Cache et Persistence
  - Shuffle
  - Optimisation mémoire
  - Monitoring
  - Spark History Server
  - Analyse des performances

TRAVAUX PRATIQUES - Diagnostic d'un traitement lent

- Optimisation progressive des performances
- Comparaison avant / après optimisation

PARTIE 4 - Industrialiser les traitements Big Data

- Déployer Spark dans une architecture Data moderne
    - Spark et Data Lake
    - Spark et Kafka
    - Structured Streaming
    - Déploiement sur Kubernetes
    - Bonnes pratiques d'industrialisation
    - Monitoring
    - Exploitation en production

TRAVAUX PRATIQUES - Étude de cas fil rouge

- Conception complète d'une plateforme Big Data permettant
    - l'ingestion des données,
    - la transformation distribuée des données,
    - le traitement temps réel,
    - la supervision des données,
    - le déploiement sur une infrastructure Cloud-native.

ÉQUIPE PÉDAGOGIQUE

L'équipe enseignante est composée de plusieurs enseignants-chercheurs de l'INSA Lyon, membres du laboratoire LIRIS, avec une activité de recherche reconnue en sciences des données. Cette équipe dispose d'une forte expertise technico-scientifique, obtenue via des nombreux projets de recherche partenariale, avec des grands groupes français mais également avec des start-up, PME ou ETI.

MOYENS ET MÉTHODES PÉDAGOGIQUES

Alternance d'exposés, illustrations, études de cas, travaux pratiques et échanges avec des experts du domaine Un support de cours sera remis à chacun des participants

PROCHAINE SESSION

VILLEURBANNE : DU 08/06/2027 AU 10/06/2027

Frais pédagogiques individuels : 2 160 € H.T. (* Repas inclus)

L'ouverture de la session est conditionnée par un nombre minimum de participants. Nous consulter pour d'autres dates.

ÉVALUATION ET RÉSULTATS

Évaluation des acquis de la formation

Evaluation des acquis des apprenants par auto-examen. 94.3% des apprenants ont acquis la compétence principale visée. (sur 177 apprenants évalués sur cette thématique depuis 2020)

Évaluation de la satisfaction des participants en fin de formation (Niveau 1 KIRKPATRICK)

4.3 par les participants. (sur 199 participants ayant suivi une formation dans la thématique depuis 2020)

RENSEIGNEMENTS ET INSCRIPTION
Tel : +33 (0)4 72 43 83 93
Fax : +33 (0)4 72 44 34 24
mail : formation@insavalor.fr
Préinscription sur formation.insavalor.fr

Accueil des personnes en situation de handicap nécessitant un besoin spécifique d'accompagnement : nous contacter à l'inscription. Nos locaux sont accessibles aux personnes à mobilité réduite.

Actualisée le 08/10/2026