Présentiel/distanciel

NOUVEAUTÉ

Introduction au Big Data (Hadoop, Spark)

Objectifs pédagogiques

- Comprendre les enjeux liés à la gestion et au traitement de données massives.
- Découvrir les concepts et l'architecture des frameworks Hadoop et Spark.
- Savoir déployer et utiliser un cluster Big Data simple.
- Mettre en oeuvre des traitements distribués avec Spark.
- Identifier les cas d'usage et les limites des solutions Big Data.

Prérequis

Maîtriser les bases de la programmation et connaître les fondamentaux ds bases de données.

Public concerné

Développeurs, data engineers et analystes souhaitant décrouvrir les écosystèmes Big Data.

Programme détaillé

Fondamentaux du Big Data (3h30)

- Les 5V du Big Data (Volume, Vélocité, Variété, Véracité, Valeur).
- Exemple de volume de données sur un cas réel (logs, IoT, réseaux sociaux...).
- Travaux pratiques : Identifier des cas d'usage Big Data et catégoriser les besoins.

Architecture et écosystème Hadoop (10h30)

- HDFS, NameNode, DataNode, YARN, MapReduce.
- Lecture et traitement de fichiers dans HDFS.
- Travaux pratiques : Charger et traiter un fichier CSV volumineux avec un traitement MapReduce simple.

Introduction à Apache Spark (3h30)

- Architecture de Spark, différences avec Hadoop, RDD vs DataFrame.
- Manipulation de jeux de données avec Spark DataFrame.
- Travaux pratiques : Filtrage, jointure et agrégation de données avec PySpark.

ETL distribué avec Spark (10h30)

- Bonnes pratiques d'un ETL Big Data, formats optimisés (Parquet, ORC).
- Création d'un pipeline ETL simple (extraction JSON, transformation, stockage au format Parquet).
- Travaux pratiques : Développement d'un pipeline Spark complet pour intégrer et transformer plusieurs sources de données hétérogènes.

Introduction à l'écosystème Big Data avancé (7h00)

- Présentation de Hive, HBase, Kafka et Airflow.
- Interrogation d'un fichier à l'aide de HiveQL.
- Travaux pratiques : Exploration libre ou mini-projet guidé avec l'écosystème Big Data.

Délais d'accès à la formation

Les inscriptions sont possibles jusqu'à 48 heures avant le début de la formation.
Dans le cas d'une formation financée par le CPF, ENI Service est tenu de respecter un délai minimum obligatoire de 11 jours ouvrés entre la date d'envoi de sa proposition et la date de début de la formation.

Modalités et moyens pédagogiques, techniques et d'encadrement

La formation peut être suivie uniquement en présentiel dans l'un des campus de l'ENI Ecole Informatique (Nantes, Rennes, Niort, Quimper). Le stagiaire intègre une promotion de l'ENI Ecole Informatique d'environ 25 à 28 stagiaires max par session.
Chaque stagiaire dispose d'un poste de travail adapté aux besoins de la formation et des supports de cours au format numérique.
Il alterne apports théoriques, démonstrations et mise en pratique au travers d'exercices, de travaux pratiques et de cas concrets, réalisés individuellement ou en groupe.

Modalités d'évaluation des acquis

Une évaluation des acquis du stagiaire est réalisée en fin de formation.

Moyens de suivi d'exécution et appréciation des résultats

Feuille de présence, émargée par demi-journée par chaque stagiaire et le formateur.
Évaluation qualitative de fin de formation, qui est ensuite analysée par l'équipe pédagogique de l'ENI Ecole Informatique.
Attestation de fin de formation remise au stagiaire.

Qualification du formateur

La formation est animée par un professionnel de l'informatique et de la pédagogie, recruté sur la base de son expérience professionnelle dans le domaine enseigné.
Nos formateurs sont pour la plupart ENI Certified Trainers. Ils assurent une veille technologique permanente.

Accessibilité de la formation

ENI Service s'engage en faveur de l'accessibilité pour les personnes en situation de handicap (PSH). Toutes nos formations sont ainsi accessibles aux PSH. Pour en savoir plus, nous vous invitons à consulter la page "Accueil des personnes en situation de handicap" de notre site internet.