Traitement de données Big Data en temps réel avec Spark et Storm

Lien : https://innov-systems.com/formation/traitement-de-donnees-big-data-en-temps-reel-avec-spark-et-storm

DURÉE
4 jours (28h)

RÉFÉRENCE
BSI22

CATÉGORIE
Big Data : développement, langages et NOSQL

Innov Systems

OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- Comprendre les fondamentaux du développement d'applications Big Data en temps réel
- Evaluer les outils Spark et Storm
- Utiliser le système distribué de calcul en temps réel Storm et Spark

POUR QUI ?

- Concepteurs
- Développeurs
- Architectes

Programme détaillé

1 / L'architecture temps réel

- Traitements temps réel
- Les architectures LAMDA
- Les architectures KAPPA
- Les architectures SMACK

2 / Introduction à apache Kafka

- L’historique du projet Kafka
- L’intérêt de Kafka dans le Big Data
- Fonctionnalités, principe de fonctionnement de Kafka
- Les composants du système : brokers, topics, consumers, producers…
- Les fichiers journaux de Kafka
- Les schémas Avro
- Liaison avec Zookeeper
- Les différentes APIs

3 / L’architecture de Storm

- Définition de l'environnement de développement
- Création de projets basés sur Storm
- Les composants primitifs : Spouts et Bolts
- Différents types de flux
- Le modèle de données (clé, valeur)
- Etude des composants d'un cluster Storm : Master node "Nimbus", Worker node, Zookeepers
- Comprendre le cycle de vie du déploiement d’une topologie
- Positionnement par rapport à un cluster Hadoop

Innov Systems

4 / Fiabilité de traitement des messages Storm

- Le mécanisme de suivi des messages
- Cycle de vie d'un message
- Utiliser différents langages pour le développement de services (Java, Python, Clojure…)
- Définir la fiabilité avec l’API Storm
- Assurer la fiabilité d’une application : méthodes et stratégies

5 / L'architecture d'apache Spark

- Présentation Spark, origine du projet
- Les solutions Spark apportées au Big Data
- Environnement et outils de Spark
- Langages supportés
- Comparaison avec l'environnement Storm
- Les différents modules de Spark
- Différents types d'architecture (Standalone, Apache Mesos ou Hadoop YARN)

6 / Traiter les données en temps réel avec spark streaming

- Fonctionnement et concepts de base
- La notion de "DStream"
- Apports, principe de fonctionnement (RDD, DataFrames, Data Sets)
- Travailler avec les RDD : Créer des RDD, Opérations principales avec les RDD
- Accumulateurs et variables broadcastées
- Principe de fonctionnement
- Principales sources de données
- Comparaison avec Apache Storm

7 / Optimisation de spark

- Gestion des variables partagées
- Données broadcastées

Inov Systems

- Accumulateurs
- Méthodes et outils d’optimisation des performances

8 / Les autres acteurs du marché

- L'architecture d'Apache Apex
- Comparaison entre Samza et Storm
- Comparaison entre Apex et Flink
- L'intégration Spark Streaming à l'aide de Talend

Approche pédagogique

- Support Ecrit et Projection
- Exposés Interactifs, Podcasts et Vidéos
- Brainstorming et Jeux de Rôle
- Cas Pratiques et Labs inclus pour leur impact opérationnel
- Test de Validation des Acquis des Connaissances

Prochaines dates programmées

19 au 22 Jan. 2027	Présentiel - Casablanca
16 au 19 Mars 2027	Présentiel - Casablanca
11 au 14 Mai 2027	Présentiel - Casablanca
06 au 09 Juil. 2027	Présentiel - Casablanca
07 au 10 Sep. 2027	Présentiel - Casablanca

Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Innov Systems

Réservation & Renseignements

Téléphone : +212 522 247 210
Email : contact@innov-systems.com
Web : https//www.innov-systems.com

Innov Systems — Tous droits réservés

Innov Systems