



Conception et Optimisation des Architectures de Données Modernes

Lien :

<https://innov-systems.com/formation/conception-et-optimisation-des-architectures-de-donnees-modernes>

 DURÉE
4 jours (28h)

 RÉFÉRENCE
BSI429

 CATÉGORIE
Big Data :
développement,
langages et NOSQL

OBJECTIFS DE LA FORMATION

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

- ✓ Comprendre les fondements et l'évolution des architectures de données
- ✓ Identifier les enjeux, avantages et limites des différentes approches (SQL, NoSQL, Big Data, Cloud)
- ✓ Maîtriser les concepts et outils clés pour concevoir, stocker, interroger et analyser efficacement des données massives
- ✓ Expérimenter l'usage conjoint de plusieurs technologies pour répondre à des besoins analytiques complexes

POUR QUI ?

- ✓ Analystes de données
- ✓ Statisticiens
- ✓ Développeurs et ingénieurs logiciels
- ✓ Chefs de projets data
- ✓ Architectes techniques



Programme détaillé

1 / INTRODUCTION AUX ARCHITECTURES DE DONNÉES

- Panorama des enjeux liés à la donnée dans l'entreprise
- Historique des architectures de données et tendances actuelles
- Positionnement des technologies SQL, NoSQL, Big Data et Cloud
- Cas d'usage typiques et secteurs d'application

2 / ACCÈS ET PRÉPARATION DES DONNÉES

- Gestion des fichiers : organisation, formats (CSV, JSON, Parquet, Avro), compression
- Connexion aux bases relationnelles : schéma, index, langage SQL, optimisation de requêtes
- Introduction aux Data Warehouses et Data Lakes
- Utilisation des API et récupération de données via Web Scraping
- Principes du traitement en mémoire et streaming de données

3 / LIMITES DES SYSTÈMES TRADITIONNELS ET ÉVOLUTIONS

- Contraintes des fichiers plats et SGBD relationnels
- Passage à des architectures distribuées : concepts et motivations
- Différences OLTP vs OLAP
- Distribution, réplication et haute disponibilité
- Compréhension du théorème CAP et de la tolérance aux pannes

4 / BASES DE DONNÉES NOSQL ET MOTEURS DE RECHERCHE

- Présentation des familles NoSQL : Key-Value, Document, Column-Family, Graph
- Forces et faiblesses par rapport au relationnel
- Exemple détaillé avec MongoDB et ses requêtes
- Moteurs de recherche et indexation avancée (Elasticsearch, Solr)

- Critères de choix d'une base NoSQL

5 / ARCHITECTURE HADOOP ET TRAITEMENT DISTRIBUÉ

- Genèse et composants de l'écosystème Hadoop
- HDFS : stockage distribué et gestion de blocs
- MapReduce : principe, phases et optimisation
- YARN : gestion des ressources et des jobs
- Commandes de base et outils d'administration (Hue, Resource Manager, History Server)
- Tour d'horizon des distributions Hadoop

6 / INTRODUCTION À SPARK ET SES MODULES

- Spark Core : RDD, transformations, actions, Pair RDD
- Spark SQL : intégration avec des données structurées
- Spark Streaming et Structured Streaming : concepts et cas pratiques
- Optimisation et bonnes pratiques avec Spark
- Comparaison avec MapReduce : avantages et limitations

7 / ÉTUDES DE CAS PRATIQUES

- Traitement de données massives issues de capteurs (courbes de charge, IoT)
- Analyse et visualisation de logs applicatifs
- Pipeline de données combinant API, stockage distribué et traitement analytique
- Comparaison de performances entre SQL et NoSQL sur un même jeu de données

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

 21 au 24 Juil. 2026

 Présentiel - Casablanca

 15 au 18 Sep. 2026

 Distanciel

 10 au 13 Nov. 2026

 Distanciel

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

Réservation & Renseignements

 **Téléphone** : +212 522 247 210

 **Email** : contact@innov-systems.com

 **Web** : <https://www.innov-systems.com>