

Sessions garanties disponibles

Formation Pig, Hive et Impala avec Hadoop

Réf. SF33387

formation Pig, Hive et Impala avec Hadoop

4 j Durée	3 000 € HT Inter / participant	7 400 € HT Intra (≤12 pers.)	9 Sessions à venir
--------------	-----------------------------------	---------------------------------	-----------------------

Objectifs pédagogiques

- ✓ Apprendre à manier Pig, Hive et Impala avec Hadoop
- ✓ Savoir utiliser Impala pour explorer, trier, filtrer et analyser les données
- ✓ Savoir explorer Pig, l'utiliser pour traiter les données basiques et complexes
- ✓ Utiliser efficacement la complémentarité des outils
- ✓ Analyser les données relationnelles, textuelles et sémantiques avec Hive

Public visé & prérequis

Public visé Architectes techniques, Développeurs	Prérequis Connaissances basiques en Hadoop et Big Data, en gestion de données et SQL
---	---

Programme détaillé

Chapitre introductif

Les problématiques du Big Data
Retour sur l'architecture MapReduce
Le processus ETL
Hadoop : solutions apportées et manques
Retour sur le système de fichiers distribués Hadoop (HDFS)
L'environnement d'Hadoop

Exploration de l'outil Apache Pig

Pig : définition, caractéristiques et rayon d'action
Les cas d'utilisation de Pig
Le langage Pig Latin : caractéristiques et mise en œuvre
Démarrer avec Pig

Utilisation de Pig pour traiter des données basiques

Connaître les types et les caractéristiques de données simples
Charger les données et définir les champs
Gérer la sortie des données
Techniques de tri et de filtrage des données récoltées
Utiliser les principales fonctions de traitement

Utilisation de Pig pour traiter des données complexes

Les différents formats de stockage
Connaître les types et les caractéristiques des données complexes et emboîtées
Grouper les données et utiliser la fonction built-in
Programmer des itérations de traitement de données groupées

Utilisation avancée de Pig

Effectuer des combinaisons d'ensembles de données
Exécuter des opérations sur des groupes de données
Paramètres avancés
Utiliser des macros et des fonctions utilisateurs (UDF)
Utiliser Pig avec d'autres langages

Résolution de problèmes et optimisation

Méthodes de résolution de problèmes
Utiliser l'UI web d'Hadoop pour le trouble shooting
Méthodes de débogage par échantillonnage de données
Monitoring des performances

Exploration de l'outil Apache Hive

Hive : définition, caractéristiques et rayon d'action
Le modèle de stockage de données de Hive
Hive et Pig : concurrence et complémentarités
Le langage de requête HiveQL
Démarrer avec Hive

Utilisation de Hive pour l'analyse de données relationnelles

Les bases et tableaux de données sous Hive
Connaître les types de données et leurs caractéristiques
Les formats de données dans Hive
Méthodes d'assemblage de données et fonctions de built-in

Gestion des données avec Hive

Construire des bases de données et tableaux de gestion Hive
Utiliser des tableaux autogérés
Stocker le résultat des requêtes
Sécuriser l'accès aux données

Analyse de données textuelles et études sémantiques

Les principes du traitement de données textuelles
Utiliser les fonctions String
Principes et applications du « Opinion Mining »

Optimisation et utilisation avancée

Mettre en œuvre les bonnes pratiques pour la performance des requêtes
Paramétrer les requêtes
Contrôler l'exécution des tâches
Partitionnement des données, bucketing et indexation
Utiliser des scripts pour transformer les données
Mettre en œuvre des fonctions utilisateurs (UDF)

Exploration du moteur de requêtes Impala

Impala : définition, caractéristiques et rayon d'action
Impala, Pig et Hive : concurrence et complémentarités
Impala dans le monde des bases de données relationnelles
Exemples d'utilisations du Shell Impala

Utilisation d'Impala pour l'analyse de données

Utiliser la syntaxe Impala
Connaître les types de données et leurs caractéristiques

Techniques de tri et de filtrage des données récoltées
Méthodes d'assemblage de données
Optimiser les performances

Conclusion

Prochaines sessions

DATE	LIEU	STATUT
03/11/2026	Lyon	Sur confirmation
03/11/2026	à distance	Sur confirmation
03/11/2026	Grenoble	Sur confirmation
03/11/2026	Toulouse	Sur confirmation
03/11/2026	Aix-en-Provence	Sur confirmation
01/12/2026	Paris	Sur confirmation

Pourquoi choisir Sparks Formation ?

20 ans d'expertise IT, des formateurs experts et des sessions garanties contre l'annulation.

20 ans
d'expertise

1 068+
formations

11 178
apprenants formés/an

+67
Net Promoter Score

Avis des participants

Note moyenne des participants de la filière Bases de données & Stockage : 4,9/5 — 27 avis

Sébastien D. — Tout était parfait. J'ai appris plein de nouvelles notions que je mettrai à coups sûrs, en œuvre dans mes activités quotidiennes.

Nicolas E. — M. Romefort est un formateur d'excellente qualité : pédagogue, professionnel et parfaitement préparé. Il explique chaque sujet avec soin et prend le temps de répondre à toutes les...

Votre formateur

Yassine E. — Formateur Data Engineering & Big Data

Yassine intervient sur des formations portant sur le data engineering et le Big Data. Parmi les formations qui lui sont confiées figurent notamment « Spark » et « Cassandra »....

Ils nous font confiance

Cette année notre filière Bases de données & Stockage a accompagné plus de 122 entreprises dont :

CEGEDIM GROUPE LDLC PROXIAD AXE SEINE AKANEA AGRO SOFTWARE AKANEA DEVELOPPEMENT GIE AG2R GIE LA MONDIALE GROUPE
HARDIS GROUP MAINCARE SOLUTIONS VEOLIA PROPTE SIEGE +112 autres

Financement & certifications

Cette formation est éligible aux financements OPCO et FNE-Formation. Nos équipes vous accompagnent dans le montage de votre dossier.



Réservez votre place dès maintenant

20 ans d'expertise · 1 000+ formations · NPS +67 · réponse sous 24h ouvrées

0 805 950 800

demande@sparks-formation.com