Formation Hadoop - Présentation de l'écosystème

Comprendre le rôle de chaque brique logicielle

N° ACTIVITÉ : 11 92 18558 92 | TÉLÉPHONE : 01 85 77 07 07 | E-MAIL : inscription@hubformation.com

Si Internet a révolutionné l'accès à l'information, son utilisation génère une quantité incalculable de nouvelles informations chaque jour... En enrichissant nos profils sur les réseaux sociaux, en rédigeant des billets sur des blogs, en participant à des forums ou en publiant des annonces sur des sites spécialisés, nous contribuons nous-même à créer de l'information. Afin d'exploiter au mieux des volumes sans cesse croissants de données, les moteurs de recherche et autres réseaux sociaux ont dû s'affranchir des outils traditionnels et développer leurs propres solutions. Les travaux de Google ont ainsi été à l'origine de la création du projet Open Source Hadoop, un Framework Java libre destiné à faciliter le stockage, l'exploitation et l'analyse de très grands volumes d'information (on parle ici de péta-octets) hébergés sur des milliers de machines différentes. Ce séminaire a pour objectif de présenter les différents éléments de l'écosystème Hadoop et leur rôle respectif dans un projet Big Data.

Référence	MAG35
Durée	1 jour (7h)
Tarif	850 €HT

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

le 18 septembre 2025
le 12 novembre 2025

PARIS

le 18 septembre 2025
le 12 novembre 2025

VOIR TOUTES LES DATES

Objectifs

| Se repérer dans l'écosystème Hadoop
| Expliquer les différents éléments de l'écosystème Hadoop et leurs rôles respectifs
| Identifier l'architecture des applicatifs Hadoop, les apports et les cas d'usage des solutions Hadoop

Public

| Chefs de projets, développeurs
| Toute personne souhaitant comprendre les mécanismes Hadoop et le rôle de chaque composant

Prérequis

| Connaissances générales des systèmes d'information

Programme de la formation

Introduction
| Rappels sur NoSQL
| Le théorème CAP
| Historique du projet Hadoop
| Fonctionnalités : stockage, outils d'extraction, de conversion, ETL, analyse, ...
| Exemples de cas d'utilisation sur des grands projets
| Les principaux composants : HDFS pour le stockage et YARN pour les calculs
| Les distributions et leurs caractéristiques (HortonWorks, Cloudera, MapR, EMR, Apache, ...)

L'architecture
| Terminologie : NameNode, DataNode, ResourceManager
| Rôle et interactions des différents composants
| Présentation des outils d'infrastructure : ambari, avro, zookeeper
| Présentation des outils de gestion des données : pig, oozie, tez, falcon, sqoop, flume
| Présentation des outils d'interfaçage avec les applications GIS

HUB Formation - 14 avenue de l'Opéra - 75001 Paris
01 85 77 07 07 - www.hubformation.com - SIREN 525 287 785

| Présentation des outils de restitution et requêtage : webhdfs, hive, hawq, impala, drill, stinger, tajo, mahout, lucene, elasticSearch, Kibana
| Les architectures connexes : spark, cassandra

Exemples interactifs
| Démonstrations sur une architecture Hadoop multi-noeuds
| Mise à disposition d'un environnement pour des exemples de calcul
| Travaux pratiques : intégration de données avec Flume, calculs map/reduce avec Pig, Hive, et Impala

Applications
| Cas d'usages de Hadoop
| Calculs distribués sur des clusters Hadoop

Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

| Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
| Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
| Privilégier une connexion filaire plutôt que le Wifi.
| Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
| Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
| Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
| Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
| Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
| Horaires identiques au présentiel.

Accessibilité

Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.
Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.