

Hadoop - Présentation de l'écosystème

Comprendre le rôle de chaque brique logicielle

 Présentiel



1 jour (7 h)

Prix inter : 850,00 € HT

Réf.: SEM35

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.

A qui s'adresse cette formation ?



Pour qui

- 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

1 - 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, ...)

2 - 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
- Présentation des outils de restitution et requêtage : webhdfs, hive, hawq, impala, drill, stinger, tajo, mahout, lucene, elasticSearch, Kibana
- Les architectures connexes : spark, cassandra

3 - 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

4 - Applications

- Cas d'usages de Hadoop
- Calculs distribués sur des clusters Hadoop



Les objectifs de la formation

- Se repérer dans l'écosystème Hadoop
- Connaître les différents éléments de l'écosystème Hadoop et leurs rôles respectifs
- Pouvoir comprendre l'architecture des applicatifs Hadoop et savoir quels sont les apports et les cas d'usage des solutions Hadoop



Evaluation

- Cette formation fait l'objet d'une évaluation formative.



Les points forts de la formation

- Un séminaire rythmé basé sur l'alternance de phases théoriques, de démonstrations et de travaux pratiques qui permet aux participants d'avoir une vue d'ensemble d'Hadoop et de son écosystème.
- Un séminaire animé par un spécialiste du Big Data.
- 100% des participants à cette formation se sont déclarés satisfaits ou très satisfaits au cours des 12 derniers mois.



Dates et villes 2026 - Référence SEM35



Dernières places disponibles



Session garantie

Paris

le 18 févr.

le 22 mai

le 18 sept.

le 11 déc.