Apache Pig : Développement de scripts

Plan de cours N° : 1224
Durée : 2 jours (14h)

PARTICIPANTS / PRE-REQUIS

Chef de projet, data scientist, développeur souhaitant utiliser Pig pour l'analyse de données.

Connaissance de Java ou Python, des bases Hadoop, et notions de calculs statistiques.

OBJECTIFS PEDAGOGIQUES

Comprendre le fonctionnement de Pig. Savoir développer des requêtes en latin, pour effectuer des transformations sur des données, des analyses de données, intégrer des données de différents formats.

MOYENS PEDAGOGIQUES

Réflexion de groupe et apports théoriques du formateur
Travail d'échange avec les participants sous forme de réunion-discussion
Utilisation de cas concrets issus de l'expérience professionnelle
Validation des acquis par des exercices de synthèse
Alternance entre apports théoriques et exercices pratiques (en moyenne 30 et 70%)
Remise d'un support de cours.
Assistance post-formation d'une durée de 1 an sur le contenu de la formation via notre adresse mail dédiée formateurs@atp-formation.com

MOYENS PERMETTANT LE SUIVI DE L'EXECUTION ET DES RESULTATS

Feuille de présence signée en demi-journée,
Evaluation des acquis tout au long de la formation,
Questionnaire de satisfaction,
Attestation de stage à chaque apprenant,
Positionnement préalable oral ou écrit,
Evaluation formative tout au long de la formation,
Evaluation sommative faite par le formateur ou à l'aide des certifications disponibles.

MOYENS TECHNIQUES EN PRESENTIEL

Accueil des stagiaires dans une salle dédiée à la formation, équipée d'ordinateurs, d'un vidéo projecteur d'un tableau blanc.

MOYENS TECHNIQUES DES CLASSES A DISTANCE

A l'aide d'un logiciel comme Teams, Zoom etc... un micro et éventuellement une caméra pour l'apprenant, suivez une formation en temps réel et entièrement à distance. Lors de la classe en ligne, les apprenants interagissent et communiquent entre eux et avec le formateur.
Les formations en distanciel sont organisées en Inter-Entreprise comme en Intra-Entreprise. L'accès à l'environnement d'apprentissage (support de cours, labs) ainsi qu'aux preuves de suivi et d'assiduité (émargement, évaluation) est assuré.
Les participants recevront une convocation avec lien de connexion
Pour toute question avant et pendant le parcours, une assistance technique et pédagogique est à disposition par mail et par téléphone auprès de notre équipe par téléphone au 04.76.41.14.20 ou par mail à contact@atp-formation.com

ORGANISATION

Les cours ont lieu de 9h00-12h30 13h30-17h00.

PROFIL FORMATEUR

Nos formateurs sont des experts dans leurs domaines d'intervention
Leur expérience de terrain et leurs qualités pédagogiques constituent un gage de qualité.

ACCESSIBILITE

Les personnes atteintes de handicap souhaitant suivre cette formation sont invitées à nous contacter directement, afin d'étudier ensemble les possibilités de suivre la formation.

MISE A JOUR

01/01/2021

Apache Pig : Développement de scripts

Plan de cours N° : 1224
Durée : 2 jours (14h)

Introduction

- Le projet Apache Pig
  - Fonctionnalités et versions
- Présentation de Pig dans l'écosystème Hadoop
  - Chaîne de fonctionnement
- Comparatif avec l'approche Hive ou Spark

Mise en oeuvre

- Rappels sur les commandes HDFS
- Prérequis techniques
  - Configuration de Pig
- Travaux pratiques
  - Les différents modes : interactif ou batch
  - Principe de l'exécution de scripts Pig Latin avec Grunt

Base latin

- Modèles de données avec Pig
- Intégration Pig avec MapReduce
- Les requêtes Latin
  - Chargement de données
  - Instructions
- Ordres de bases
  - LOAD
  - FOREACH
  - FILTER
  - STORE
- Travaux pratiques
  - Création d'un ETL de base
  - Contrôle d'exécution

Transformations

- Groupements, jointures, tris, produits cartésiens
- Transformation de base de la donnée
- Découpages
  - Découpages sur filtres

Analyse de la donnée

- Echantillonages
  - Filtres
  - Rangements avec rank et dense
- Calculs
  - min/max
  - sommes
  - moyennes
- Travaux pratiques
  - Traitements de chaînes de caractères
  - Traitement de dates

Intégration

- Formats d'entrées/sorties
  - Interfaçage avro, Json
- Travaux pratiques
  - Chargement de données depuis HDFS vers Hbase
  - Analyse de données Pig/Hbase
  - Restitution Json

Extensions

- Extension du PigLatin
- Création de fonctions UDF en java
- Intégration dans les scripts Pig
- Travaux pratiques
  - Utilisation de Pig Latin depuis des programmes Python
  - Exécution de programmes externes, streaming

31, av. du Granier
38240 Meylan
Tél. 04 76 41 14 20
Fax 04 76 41 38 45

SARL au capital de 120 000€ - 390 032 050 RCS GRENOBLE
SIRET 390 032 050 00010 - APE 8559A - TVA Intracom. FR03390032050