

MapReduce, appliquer le paradigme dans tous les langages et le transposer sous Hadoop

tutorat en option

Réf : 4TP - Prix 2023 : 90€ HT

Cette formation digitale a pour objectifs de comprendre le paradigme MapReduce, ses spécificités et ses finalités. Elle s'adresse à tous les développeurs et administrateurs système. La pédagogie s'appuie sur un auto-apprentissage séquencé par actions de l'utilisateur sur l'environnement à maîtriser. Une option de tutorat vient renforcer l'apprentissage (option non disponible actuellement).

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Connaître la théorie du paradigme MapReduce (la fonction map() et la fonction reduce())

Analyser des journaux Apache (access log) dans plusieurs langages (Java, Scala, Python, JavaScript, PHP et Bash)

Connaître les fonctionnalités MapReduce au sein de la plateforme Big Data Hadoop

Faire fonctionner un MapReduce codé en C++ avec Hadoop Pipes

Comprendre les fonctions MapReduce plus complexes

Combiner les fonctions MapReduce avec Sqoop, MySQL et Oozie

PÉDAGOGIE ET PRATIQUES

Pédagogie active mixant pratique, théorie, gamification et partages. Un chef de projet coordonne les échanges et un service technique est dédié au support de l'apprenant. La formation est diffusée au format SCORM et accessible en illimité pendant 1 an.

ACTIVITÉS DIGITALES

Démonstrations, cours enregistrés, partages de bonnes pratiques, quiz, fiches de synthèse.

LE PROGRAMME

dernière mise à jour : 06/2022

1) Présentation du paradigme MapReduce

- Histoire du paradigme MapReduce
- Principe général
- Fonction map() en détails
- Fonction reduce() en détails

2) MapReduce dans différents langages

- Analyse des journaux d'accès : Cas d'un MapReduce simple
- Développement de notre MapReduce en Java
- Développement de notre MapReduce en Scala
- Développement de notre MapReduce en Python
- Développement de notre MapReduce en JavaScript
- Développement de notre MapReduce en PHP
- Développement de notre MapReduce en Bash

3) MapReduce dans Hadoop

- Implémentation des MapReduce sous YARN
- Adaptation de notre code Java pour Hadoop

PARTICIPANTS

Tous les développeurs et administrateurs système.

PRÉREQUIS

Aucune connaissance particulière.

COMPÉTENCES DU FORMATEUR

Les experts qui ont conçu la formation et qui accompagnent les apprenants dans le cadre d'un tutorat sont des spécialistes des sujets traités. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

La progression de l'apprenant est évaluée tout au long de sa formation au moyen de QCM, d'exercices pratiques, de tests ou d'échanges pédagogiques. Sa satisfaction est aussi évaluée à l'issue de sa formation grâce à un questionnaire.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : documentation et support de cours, exercices pratiques d'application et corrigés des exercices, études de cas ou présentation de cas réels. ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Une attestation de fin de formation est fournie si l'apprenant a bien suivi la totalité de la formation.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

- Suivi de l'évolution du job sous Hadoop
- Hadoop streaming : Lancement de nos autres MapReduce grâce à des exécutables
- Hadoop Pipes : Exécution de MapReduce en C++

4) Exemples avancés de MapReduce

- Compter le nombre de visiteurs uniques par page
- Opérer une jointure de données
- Utiliser une clé composite
- Réaliser un tri secondaire (secondary sort)

5) Plus loin grâce à Hadoop

- Compteurs
- Journaux
- Importation/exportation des données de MySQL grâce à Sqoop
- Réalisation des workflows de MapReduce grâce à Oozie