

Certification Microsoft - Perform Data Engineering on Microsoft Azure HDInsight (70-775)

CATEGORIE : B

Vue d'ensemble

Domaine(s) d'activité professionnel dans lequel(s) est utilisé la certification :

Transverse : ■ **Tout secteur d'activité sauf artisanat et professions libérales**

Code(s) NAF : —

Code(s) NSF : **326**

Code(s) ROME : **M1802**, **I1401**, **M1805**

Formacode : —

Date de création de la certification : **26/10/2017**

Mots clés : **Big data**, **Azure HDInsight**, **ingénierie de données**, **microsoft**

Identification

Identifiant : **3742**

Version du : **16/10/2018**

Références

Consensus, reconnaissance ou recommandation :

Formalisé :

- [Détail sur le programme de certification Microsoft sur le site de l'éditeur](#)

Non formalisé :

- [Témoignages sur la certification Microsoft](#)
- [Pourquoi passer la certification sur les produits et les technologies Microsoft ?](#)

Norme(s) associée(s) :

—

Descriptif

Objectifs de l'habilitation/certification

Azure HDInsight est un service d'analyse entièrement géré, complet et open source pour les entreprises. HDInsight est un service cloud qui rend facile, rapide et économique le traitement de volumes importants de données. HDInsight prend également en charge un large éventail de scénarios, tels que l'extraction, transformation et chargement (ETL) ; l'entreposage de données ; l'apprentissage automatique ; et IoT.

La certification liée à cet examen est destinée aux ingénieurs de données, aux architectes et experts des données, ainsi qu'aux développeurs qui conçoivent des workflows d'ingénierie Big Data sur HDInsight. Les candidats à cet examen doivent avoir une expérience professionnelle appropriée dans des solutions d'analyse Big Data, et connaître les fonctions et capacités du traitement de données par paquet, du traitement en temps réel et du traitement interactif.

Les candidats qui passent cet examen seront capables d'appliquer l'écosystème de cloud de Microsoft à la conception et à la mise en œuvre de workflows d'ingénierie Big Data et d'utiliser des technologies open source pour l'ajout de valeur stratégique.

Lien avec les certifications professionnelles ou les CQP enregistrés au RNCP

- Certification Microsoft - Conception et mise en oeuvre des solutions Big Data Analytics (70-475)
- Certification Microsoft - Conception et mise en oeuvre des solutions Cloud Data Platform (70-473)

- Certification Microsoft Requête SQL Server 2012/2014

Descriptif général des compétences constituant la certification

Descriptif général des compétences constituant la certification

Les compétences visées et évaluées par la certification «Perform Data Engineering on Microsoft Azure HDInsight » se répartissent en 2 grands axes qui constituent les principales missions de l'ingénieur de données :

- Administrer et fournir des clusters de données HDInsight pour faciliter le traitement de volumes importants de données au sein de l'entreprise, à la fois dans le stockage et dans l'analyse.
- Mettre en œuvre des solutions de traitement pour le Big Data, de différentes natures (par paquet, interactives, temps réel), afin de répondre aux besoins des utilisateurs métiers qui traitent des données dans le cadre de leur fonction

Modalités générales

La certification Microsoft Perform Data Engineering on Microsoft Azure HDInsight est ouverte à tout public. La préparation à cet examen se base sur le suivi d'une formation officielle de 35 heures (incluant support de cours reprenant les concepts théoriques et les labs) ainsi qu'un travail personnel. Il est conseillé de respecter un temps de révision après la formation avant de passer l'examen.

De nombreux outils de préparation existent en complément, comme le MOC-On demand (formation au format vidéo, incluant des labs), le Practice test (test d'entraînement au passage de la certification), et autres ressources en ligne sur les sites de l'éditeur.

Le suivi de la formation officielle n'est pas obligatoire et ne valide pas l'obtention du titre de certification, mais elle est toutefois recommandée puisqu'elle met en pratique les connaissances à acquérir

Liens avec le développement durable

Aucun

Public visé par la certification

Tous publics

Valeur ajoutée pour la mobilité professionnelle et l'emploi

Pour l'individu

Une étude de rémunérations éditée par le cabinet Robert Half en 2017 indique que près de 90% des data scientists (tout profil ingénieur de données) sont approchés au moins une fois par mois afin de leur proposer de nouvelles opportunités d'emploi, plus de 50% sont contactés chaque semaine et 30% plusieurs fois par semaine.

Grâce à la création par l'éditeur de badges numériques, chaque professionnel certifié peut afficher son expertise via son CV, son blog ou ses réseaux professionnels. L'affichage des compétences via ce badge est un vecteur de recrutement grâce à l'octroi de services supplémentaires.

<https://www.microsoft.com/fr-fr/learning/badges.aspx>

Pour l'entité utilisatrice

Les données, structurées ou non structurées (big data), jouent un rôle prédominant au sein des entreprises qui les capturent, les gèrent et cherchent à les analyser : les ingénieurs de données (qu'ils aient une orientation analyste, architecte ou développeur) ont pour mission de restituer les données de l'entreprise et surtout de les analyser avec des outils adaptés et ultra spécialisés; un rôle de plus en plus stratégique pour faciliter les choix de la direction et qui requiert des connaissances pointues.

D'après une étude fin 2016 du Gartner, Microsoft SQL Server était numéro un des leaders de solutions de bases de données, soit la base de données la plus utilisée par les entreprises au monde. Ainsi, détenir un titre de certification SQL constitue une excellente opportunité en terme d'employabilité.

Témoignages sur la certification Microsoft :

<https://www.microsoft.com/fr-fr/learning/certification-success-stories.aspx>

Pourquoi passer la certification sur les produits et les technologies Microsoft ?

<https://www.microsoft.com/fr-fr/learning/certification-testimonials.aspx>

<https://sgbd.developpez.com/actu/105895/Qui-est-leader-des-bases-de-donnees-en-2016-Microsoft-est-le-numero-un-d-apres-l-etude-d-octobre-2016-de-Gartner/>

Pour les entités dont l'activité commerciale est de revendre, intégrer ou préconiser les solutions Microsoft SQL et Azure, l'intégration d'un professionnel certifié lui permet de satisfaire les prérequis techniques du programme de partenariat Microsoft, d'assurer aux clients utilisateurs une expertise et un savoir-faire, et d'acquérir ainsi un réel avantage concurrentiel.

Evaluation / certification

Pré-requis

Cette certification ne requiert pas de prérequis en matière d'âge ou de niveau scolaire.

Il est toutefois recommandé d'avoir une expérience dans la programmation via R, de posséder des connaissances sur les méthodes statistiques courantes et les meilleures pratiques d'analyse des données. La connaissance du système d'exploitation Windows et ses fonctionnalités principales ainsi que des bases de données relationnelles est un plus.

Compétences évaluées

Les compétences visées et évaluées par la certification «Perform Data Engineering on Microsoft Azure HDInsight » se répartissent en 2 grands domaines qui constituent les principales responsabilités de l'ingénieur de données ; **à noter que la compétence dans la mise en œuvre de solutions de traitement est beaucoup plus étoffée et donc comporte plus de questions d'évaluation :**

Administrer et fournir des clusters de données HDInsight pour faciliter le traitement de volumes importants de données au sein de l'entreprise, à la fois dans le stockage et dans l'analyse.

Déployer des clusters HDInsight : sélectionner un type de cluster approprié sur la base des considérations de charge de travail, personnaliser un cluster en utilisant des actions de script, fournir un cluster en utilisant les outils Azure CLI, fournir un cluster en utilisant les modèles Azure Resource Manager (ARM) et PowerShell, gérer les disques gérés

Déployer et sécuriser les clusters HDInsight multi-utilisateurs avec différents rôles ; gérer les utilisateurs, les groupes et les permissions par le biais d'Apache Ambari, de PowerShell et d'Apache Ranger

Recevoir des données pour le traitement par lots et interactif et stocker les données

Certificateur(s)

- MICROSOFT

Centre(s) de passage/certification

- Centres Pearson VUE - accessibles depuis <http://www.pearsonvue.com/microsoft/locate/>

Configurer des clusters HDInsight

Gérer et déboguer les tâches HDInsight, surveiller et gérer les alertes et effectuer des actions prédictives

Pour répondre aux besoins des utilisateurs métiers qui traitent des données dans le cadre de leur fonction, les ingénieurs de données sont évalués sur la mise en œuvre des solutions de traitement pour le Big Data, de différentes natures (par paquet, interactives, temps réel). Ils ont besoin pour cela de connaître et choisir les outils les plus appropriés en fonction des résultats recherchés

Mettre en œuvre des solutions de traitement par paquet pour le Big Data

Utiliser Hive et Apache Pig pour le traitement ainsi qu'un format de stockage approprié tel que Apache Parquet, ORC, Text et JSON

Concevoir des solutions ETL par paquet pour le Big Data avec Spark et identifier les goulets d'étranglement de requêtes en utilisant le graphe de requête Spark SQL

Opérationnaliser Hadoop et Spark

Mettre en œuvre des solutions de traitement interactives pour le Big Data

Exécuter des requêtes en utilisant Spark SQL

Procéder à l'analyse exploratoire des données en utilisant Spark SQL : Utiliser Jupyter et Apache Zeppelin pour visualiser et développer de petits cadres de données Spark pour la modélisation.

Utiliser Interactive Hive : connecter les outils BI aux clusters Interactive Hive

Procéder à l'analyse exploratoire des données en utilisant Hive, avec requête et visualisation interactives

Procéder au traitement interactif en utilisant Apache Phoenix sur HBase : choisir entre Hive, Spark et Phoenix sur HBase pour le traitement interactif

Mettre en œuvre des solutions de traitement en temps réel pour le Big Data

Créer des applications de transmission en utilisant DStream API

Visualiser les données de streaming dans un tableau de bord en temps réel PowerBI

Créer des applications de streaming structurées dans Spark

Créer des clusters Storm pour des tâches en temps réel

Créer des solutions qui utilisent Kafka

Construire des solutions qui utilisent HBase et Identifier les cas d'utilisation de HBase dans HDInsight

Niveaux délivrés le cas échéant (hors nomenclature des niveaux de formation de 1969)

Non Applicable

La validité est Permanente

Possibilité de certification partielle : non

Matérialisation officielle de la certification :

Les candidats reçoivent la notification de réussite/échec quelques minutes après la fin de l'examen et accède à un relevé de notes détaillé : son transcript est disponible à tout moment sur un portail sécurisé dédié.

Plus d'informations

Statistiques

Chaque année près de 10 000 professionnels passent les Certifications Microsoft. Plus de 1000 professionnels se préparent et passent au moins un examen en Business Intelligence et Big Data à partir des solutions Microsoft, et 200 candidats ont déjà choisi de valider leur expertise sur la certification Microsoft Azure HDInsight.

Autres sources d'information

Grâce à la création par l'éditeur de badges numériques, chaque professionnel certifié peut afficher son expertise via son CV, son blog ou ses réseaux professionnels. L'affichage des compétences via ce badge est un vecteur de recrutement grâce à l'octroi de services supplémentaires.

<https://www.microsoft.com/fr-fr/learning/badges.aspx>

Témoignages sur la certification Microsoft :

<https://www.microsoft.com/fr-fr/learning/certification-success-stories.aspx>

Pourquoi passer la certification sur les produits et les technologies Microsoft ?

<https://www.microsoft.com/fr-fr/learning/certification-testimonials.aspx>