

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

RS3742 · RS · Publiée

L'essentiel

Code(s) NSF	326
Type d'enregistrement	Enregistrement sur demande
Date d'échéance	31/12/2021

Certificateur(s)

MICROSOFT FRANCE · SIRET : 32773318400516

Résumé de la certification

OBJECTIFS ET CONTEXTE

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.

COMPÉTENCES ATTESTÉES

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

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

- 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

Voies d'accès

PRÉREQUIS À L'ENTRÉE EN FORMATION

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.

Base légale

Date d'échéance de l'enregistrement 31/12/2021

Pour plus d'informations

LIEN VERS LE DESCRIPTIF

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.

- Témoignages sur la certification Microsoft :
- Pourquoi passer la certification sur les produits et les technologies Microsoft ?

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rs/3742/>