

Certification Microsoft - Analyzing Big Data with Microsoft R (70-773)

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 : **M1805**, **I1401**, **M1802**

Formacode : —

Date de création de la certification : **05/10/2017**

Mots clés : **Langage R**, **microsoft**, **ingénierie de données**, **Big data**

Identification

Identifiant : **3744**

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 :

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

Norme(s) associée(s) :

—

Descriptif

Objectifs de l'habilitation/certification

La certification liée à cet examen est destinée aux candidats de profil scientifique ou des analystes de données qui doivent traiter et analyser des ensembles de données plus importants que la mémoire en utilisant le langage R. R est un langage de programmation et un logiciel libre dédié aux statistiques et à la science des données. En validant ce titre de certification, les professionnels démontrent qu'ils sont en mesure de créer et activer une analyse d'un grand nombre de données, et montrer comment les utiliser dans des environnements Big Data, tels que Hadoop ou le cluster Spark, ou encore une base de données SQL Server.

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

Les compétences visées et évaluées par la certification « Analyse de Big Data avec Microsoft R » se répartissent en 4 grands champs d'intervention qui constituent les principales responsabilités de l'analyste de données.

Public visé par la certification

Tous publics

Lire et explorer le Big Data pour prendre en compte tous les formats de fichiers de données pouvant servir à l'entreprise dans l'analyse de sa performance

Traiter le Big Data, pour fournir aux directions métiers, aux opérationnels ou à la direction de l'entreprise les analyses servant à la prise de décision

Construire des modèles prédictifs avec ScaleR pour développer la stratégie de l'entreprise

Utiliser le R Server dans différents environnements, pour s'adapter aux spécificités des systèmes propriétaires ou en open source, et nombreuses sont les entreprises qui utilisent des systèmes d'information hybrides

Modalités générales

La certification Microsoft - Analyzing Big Data with Microsoft R est ouverte à tout public. La préparation à cet examen se base sur le suivi d'une formation officielle de 35 heures (incluant un support de cours qui reprend 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

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 analystes et développeurs ont pour mission de restituer les données de l'entreprise et surtout de les analyser avec des outils adaptés ; un rôle de plus en plus stratégique pour faciliter les choix de la direction.

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 pré-requis 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.

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>

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 « Analyse de Big Data avec Microsoft R » se répartissent en 4 grands champs d'intervention qui constituent les principales responsabilités de l'analyste de données. **Les questions sur lesquelles les candidats sont évalués à l'examen portent sur les thèmes décrits dans la liste, mais ne s'y limitent pas. :**

Lire et explorer le Big Data pour prendre en compte tous les formats de fichiers de données pouvant servir à l'entreprise dans l'analyse de sa performance

Lire les données avec R Server, notamment lire les formats de fichiers de données supportés tels que les fichiers texte, SAS et SPSS, convertir les données, identifier les compromis entre les fichiers types, traiter les données de source qui ne peuvent être lues originairement par le serveur R Server.

Résumer les données, créer des formules et Calculer les tableaux croisés et les statistiques univariées

Visualiser les données : créer des visualisations personnelles avec différents outils

Traiter le Big Data, pour fournir aux directions métiers, aux opérationnels ou à la direction de l'entreprise les analyses servant à la prise de décision

Traiter les données avec rxDataStep et traiter les dates (POSIXct, POSIXt).

Procéder à des transformations complexes qui utilisent des fonctions de transformation

Gérer les ensembles de données avec des fonctions de tri, de déduplication, de fusion, etc.

Certificateur(s)

- MICROSOFT

Centre(s) de passage/certification

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

Traiter les données texte en utilisant les paquets RML

Construire des modèles prédictifs avec ScaleR pour développer la stratégie de l'entreprise

Utiliser rxLinMod, rxGlm et rxLogit pour estimer les modèles linéaires et choisir entre différents types de sélections de variables automatiques, tels que la recherche gourmande, la notation répétée et le sous-produit de la formation

Utiliser rxDTree, rxDForest, et rxBTrees pour construire des modèles de partitionnement, et sélectionner les paramètres qui affectent les biais et la variance, tels que le raccourcissement, le taux d'apprentissage et la profondeur de l'arbre

Générer des prédictions et des résidus avec rxPredict

Évaluer des modèles et des paramètres de configuration

Créer des modèles supplémentaires en utilisant les paquets RML tels que des régression linéaires et logistiques

Utiliser le R Server dans différents environnements, pour s'adapter aux spécificités des systèmes propriétaires ou en open source, et nombreuses sont les entreprises qui utilisent des systèmes d'information hybrides

Utiliser des contextes de calculs différents pour exploiter le R Server efficacement : identifier quel contexte de calcul utiliser pour différentes tâches ; identifier et utiliser les sources de données appropriées pour différentes sources de données et contextes de calcul

Optimiser les tâches en utilisant les contextes de calcul locaux : Identifier et exécuter les tâches qui ne peuvent être exécutées que dans le contexte de calcul local, profiler dans différents contextes de calcul.

Choisir quand procéder aux calculs dans la base de données par rapport à quand procéder aux calculs hors de la base de données, identifier les limites des calculs dans la base de données, communiquer de façon efficace les propriétés de performance aux administrateurs et architectes SQL (SQL Server Profiler).

Mettre en place des flux de travail d'analyse dans l'écosystème Hadoop et dans Spark

Déployer des modèles prédictifs destinés à SQL Server et Azure Machine Learning

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

N/A

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.

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>