

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

RS3744 · 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

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.

COMPÉTENCES ATTESTÉES

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.

- 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

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 originellement 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, POSIXlt).
- 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.

- 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

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/3744/>