



REFERENTIEL D’EVALUATION ET DE COMPETENCES DE LA CERTIFICATION CONCEVOIR ET METTRE EN ŒUVRE LA GESTION DES DONNEES DANS MICROSOFT AZURE

(Titre en anglais : MICROSOFT AZURE DATA ENGINEER ASSOCIATE)

MODALITES D’EVALUATION

Pour obtenir cette certification, il est proposé au candidat d’évaluer leurs compétences à travers deux (2) examens en ligne, supervisés par l’organisme Pearson VUE, délivrés dans un centre d’examen accrédité (ou via de la surveillance à distance).

Chaque examen dure environ deux (2) heures – livres fermés - et comprend une variété de questions appelant différentes formes de réponse* : Écran actif, Meilleure réponse, Construction de liste, Études de cas, Glisser-déplacer, Zone réactive, Choix multiple, Plusieurs choix de réponse, Réponse courte, Révision des notes, Écran de révision.

*Détaillées à cette URL : <https://www.microsoft.com/en-us/learning/certification-exams.aspx?types=true>

Le seuil de réussite est fixé à environ 70% de bonnes réponses, qui correspond à un score de passage de 700 points (score à l’échelle). Le pourcentage réel varie d’un examen à l’autre. La note de passage est basée sur l’apport d’experts en la matière, le niveau de compétence requis pour être considéré comme compétent dans le domaine du contenu, et la difficulté des questions livrées pendant l’examen. Les pourcentages dans le tableau des compétences évaluées indiquent le poids relatif de chaque sujet principal de l’examen. Plus le pourcentage est élevé, plus les candidats devront répondre à des questions sur cette zone de contenu. La liste des tâches évaluées n’est pas exhaustive et peut couvrir d’autres tâches dans le cadre des compétences évaluées.

Le contenu des tests est réévalué régulièrement par les équipes Microsoft. Le contenu des examens « Azure Data Engineer Associate » a été mis à jour le 31 mars 2020 pour refléter les dernières évolutions des services et de la plate-forme Azure.

COMPETENCES ET MISSIONS DE L’INGENIEUR DE DONNEES AZURE MICROSOFT

Les ingénieurs de données ont à la fois un rôle de conception/recommandation d’une solution de données, et ils sont en même temps responsables des tâches de mise en œuvre relatives aux données qui comprennent : la fourniture de services de stockage de données, l’acquisition et la diffusion des données en contenu et en masse, la transformation des données, la mise en œuvre des exigences de sécurité, la mise en œuvre des politiques de rétention de données, l’identification des goulots d’étranglement de performance et l’accès aux sources de données externes.

Les candidats à ce titre de certification Ingénieur de données ont pour mission d'identifier et de répondre aux exigences de tous les métiers afin de mettre à leur disposition des solutions qui utilisent les services de données Azure. Ainsi, ils doivent être capables de mettre en œuvre des solutions de données qui utilisent les services Azure suivants : Azure Cosmos DB, la base de données Azure SQL, Azure Synapse Analytics, Azure Data Lake Storage, Azure Data Factory, Azure Stream Analytics, Azure Databricks et le stockage Azure Blob.

RESUME DES ACTIVITES PRINCIPALES :

- 1. Mettre en œuvre une solution de données Microsoft Azure** (1^{er} examen)
- 2. Concevoir une solution de données pour Azure** (2^{ème} Examen)

| COMPETENCES MOBILISABLES EVALUEES | | Nature des tâches évaluées permettant de valider la compétence | EVALUATION | | |
|--|---|--|--|---|--|
| 1 ^{ER} EXAMEN - Sur 100% : | | | % de l'évaluation globale | Modalités d'évaluation | Critères |
| Mettre en œuvre des solutions de stockage de données pour conserver l'historique tout en protégeant les données de l'entreprise de manière optimale | | | 40-45% de l'évaluation globale du 1^{er} examen | Examen en ligne avec une variété de questions (cf détail plus haut) ** Environ 50 minutes (pour 20 à 22 questions) sont consacrées à | Le taux de bonnes réponses doit être au global de 70% minimum |
| Implémenter des entrepôts de données non relationnelles qui peuvent prendre différentes formes et requièrent des exigences spécifiques | <ul style="list-style-type: none"> ▶ Installer une solution qui utilise Cosmos DB, Data Lake Storage Gen2 ou le stockage Blob ▶ Mettre en œuvre la distribution de données et les partitions ▶ Définir un modèle de cohérence dans Cosmos DB ▶ Provisionner un entrepôt de données non relationnelles ▶ Fournir un accès aux données pour répondre aux exigences de sécurité ▶ Mettre en œuvre pour la haute disponibilité, la reprise après sinistre et la distribution mondiale | | | | |

| | | | | | |
|--|---|---|---|--|--|
| | Implémenter des entrepôts de données relationnelles pour permettre l'analyse des données métiers structurées et pour mettre en place des politiques de prise de décision centrées sur la donnée. | <ul style="list-style-type: none"> ▶ Configurer des pools « élastiques » (partagés) ▶ Configurer la géoréplication ▶ Fournir un accès aux données pour répondre aux exigences de sécurité ▶ Mettre en œuvre la haute disponibilité, la reprise après sinistre et la distribution globale ▶ Implémenter la distribution de données et les partitions pour Azure Synapse Analytics ▶ Installer PolyBase | | cette compétence | |
| | Gérer la sécurité des données | <ul style="list-style-type: none"> ▶ Mettre en oeuvre le masquage des données ▶ Crypter les données au repos et en mouvement | | | |
| Gérer et développer le traitement des données en vue de l'optimisation pour l'analyse des données | | | | | |
| | Développer des solutions de traitement des données par lots pour améliorer l'efficacité du traitement, en définissant des priorités et en automatisant les jobs | <ul style="list-style-type: none"> ▶ Développer des solutions de traitement par lots à l'aide de Data Factory et Azure Databricks ▶ Ingérer des données à l'aide de PolyBase ▶ Implémenter le runtime d'intégration pour Data Factory ▶ Implémenter l'activité de copie dans Azure Data Factory ▶ Créer des services et des ensembles de données liés ▶ Créer des pipelines et des activités ▶ Paramétrer des flux de données de mappage dans Azure Data Factory ▶ Créer et planifier des déclencheurs ▶ Mettre en oeuvre des clusters, des blocs-notes, des travaux et une mise à l'échelle automatique Azure Databricks ▶ Ingérer des données dans Azure Databricks | 25-30% de l'évaluation globale du 1er examen | Examen en ligne avec une variété de questions (cf détail plus haut) ** Environ 32 minutes (pour 12 à 14 questions) sont consacrées à cette compétence | Le taux de bonnes réponses doit être au global de 70% minimum |
| | Développer des solutions de streaming pour gérer les flux importants de données | <ul style="list-style-type: none"> ▶ Configurer l'entrée et la sortie ▶ Sélectionner les fonctions de fenêtrage appropriées ▶ Implémenter le traitement des événements à l'aide de Stream Analytics | | | |

| Surveiller et optimiser les solutions de données pour garantir les performances et la prise de décision au service des différents métiers utilisateurs des données | | | | | |
|--|---|---|---|--|--|
| | Surveiller le stockage des données pour anticiper les défaillances | <ul style="list-style-type: none"> ▶ Surveiller les sources de données relationnelles et non relationnelles ▶ Mettre en place la surveillance des outils de stockage utilisés : Blob, Data Lake Storage, ▶ Mettre en place la surveillance de la base de données SQL ▶ Implémenter la surveillance des outils d'analyse tels que Azure Synapse Analytics, Cosmos DB ▶ Configurer les alertes Azure Monitor ▶ Implémenter l'audit à l'aide d'Azure Log Analytics | 30-35% de l'évaluation globale du 1er examen | Examen en ligne avec une variété de questions (cf détail plus haut) ** Environ 40 minutes (pour 15 à 18 questions) sont consacrées à cette compétence | Le taux de bonnes réponses doit être au global de 70% minimum |
| | Surveiller le traitement des données pour assurer la continuité de service | <ul style="list-style-type: none"> ▶ Surveiller les pipelines Data Factory ▶ Surveiller Azure Databricks, Stream Analytics ▶ Configurer les alertes Azure Monitor ▶ Implémenter l'audit à l'aide d'Azure Log Analytics | | | |
| | Optimiser les solutions de données Azure en utilisant les outils à disposition pour mettre à disposition des données fiables | <ul style="list-style-type: none"> ▶ Résoudre les goulots d'étranglement du partitionnement des données ▶ Optimiser le stockage Data Lake ▶ Optimiser les outils d'analyse Stream Analytics, Azure Synapse Analytics ▶ Optimiser le paramétrage de la base de données SQL ▶ Gérer le cycle de vie des données | | | |

| COMPETENCES MOBILISABLES EVALUEES | | Nature des tâches évaluées permettant de valider la compétence | EVALUATION | | |
|--|---|--|---|--|--|
| 2EME EXAMEN - Sur 100% : | | | % de l'évaluation globale | Modalités d'évaluation | Critères |
| Concevoir des solutions de stockage de données Azure pour garantir le stockage des données et leur traitement qui répondent aux besoins de l'entreprise | | | 40-45% de l'évaluation globale du 2nd examen | Examen en ligne avec une variété de questions (cf détail plus haut) ** Environ 50 minutes (pour 20 à 22 questions) sont consacrées à cette compétence | Le taux de bonnes réponses doit être au global de 70% minimum |
| | Recommander une solution de stockage de données Azure en fonction des exigences | <ul style="list-style-type: none"> ▶ Choisir la bonne solution de stockage de données pour répondre aux besoins techniques et exigences commerciales ▶ Choisir le type de distribution de partition | | | |
| | Concevoir des entrepôts de données cloud non relationnelles pour stocker et traiter les différents formats de données non structurées | <ul style="list-style-type: none"> ▶ Concevoir la distribution des données et les partitions ▶ Conception à l'échelle (y compris multi-région, latence et débit) ▶ Concevoir une solution qui utilise Cosmos DB, Data Lake Storage Gen2 ou le stockage Blob ▶ Sélectionner l'API Cosmos DB appropriée ▶ Concevoir une stratégie de reprise après sinistre ▶ Anticiper et paramétrer la haute disponibilité | | | |
| | Concevoir des entrepôts de données cloud relationnelles pour stocker et traiter les données structurées en vue d'analyses décisionnelles | <ul style="list-style-type: none"> ▶ Concevoir la distribution des données et les partitions ▶ Conception à l'échelle (y compris multi-région, latence et débit) ▶ Concevoir une solution qui utilise SQL Database et Azure Synapse Analytics ▶ Concevoir une stratégie de reprise après sinistre ▶ Anticiper et paramétrer la haute disponibilité | | | |
| Concevoir des solutions de traitement des données | | | 25 -30% de l'évaluation globale du 2nd examen | Examen en ligne avec une variété de questions | Le taux de bonnes réponses doit être au global de |
| | Concevoir des solutions de traitement par lots pour automatiser le maximum d'opérations et définir les priorités | <ul style="list-style-type: none"> ▶ Concevoir des solutions de traitement par lots qui utilisent Data Factory et Azure Databricks | | | |

| | | | | | |
|--|--|---|---|---|--|
| | | <ul style="list-style-type: none"> ▶ Identifier la méthode optimale d'ingestion de données pour une solution de traitement par lots ▶ Identifier où le traitement doit avoir lieu, comme à la source, à la destination, ou en transit ▶ Identifier la logique de transformation à utiliser dans le flux de données de mappage dans Azure Data Factory | | (cf détail plus haut) ** | 70% minimum |
| | Concevoir des solutions de traitement en temps réel en étudiant les outils à disposition | <ul style="list-style-type: none"> ▶ Concevoir le traitement en temps réel à l'aide de Stream Analytics et Azure Databricks ▶ Concevoir et fournir des ressources de calcul | | Environ 32 minutes (pour 12 à 14 questions) sont consacrées à cette compétence | |
| Concevoir la sécurité et la conformité des données pour prévenir les pertes et répondre aux exigences du RGPD | | | | | |
| | Définir les moyens de sécurité pour accéder aux données source | <ul style="list-style-type: none"> ▶ Planifier des points de terminaison sécurisés (privés / publics) ▶ Choisir le mécanisme d'authentification approprié, comme les clés d'accès, l'accès partagé, les signatures (SAS) et Azure Active Directory (Azure AD) | 25 -30% de l'évaluation globale du 2nd examen | Examen en ligne avec une variété de questions (cf détail plus haut) ** | Le taux de bonnes réponses doit être au global de 70% minimum |
| | Concevoir la politique de sécurité et des normes de données, en répondant aux différentes étapes-clés | <ul style="list-style-type: none"> ▶ Concevoir le cryptage des données pour les données au repos et en transit ▶ Concevoir le cadre de l'audit des données et le masquage des données ▶ Définir le cahier des charges pour la confidentialité des données et la classification des données ▶ Concevoir une politique de conservation des données ▶ Planifier une stratégie d'archivage ▶ Planifier la purge des données en fonction des besoins de l'entreprise | | | |