



REFERENTIEL D’EVALUATION ET DE COMPETENCES DE LA CERTIFICATION CREER ET MAINTENIR DES SOLUTIONS D’ANALYSE DE DONNEES EFFICACES ET SECURISEES SUR LE CLOUD AWS

(Titre en anglais : AWS Certified Data Analytics - Specialty)

Créer et maintenir des solutions d’analyse de données efficaces et sécurisées sur le cloud AWS

MODALITES D’EVALUATION

Pour obtenir cette certification, il est proposé au candidat d’évaluer leurs compétences à travers un (1) examen en ligne, supervisé par l’organisme Pearson VUE ou PSI Services, délivrés dans un centre d’examen accrédité (ou via de la surveillance à distance).

L’examen dure environ trois heures (3h) – livres fermés - et comprend une variété de questions appelant différentes formes de réponse* : Questions à choix multiples, Questions à réponses multiples.

*Détaillées à cette URL : https://d1.awsstatic.com/training-and-certification/docs-data-analytics-specialty/AWS-Certified-Data-Analytics-Specialty_Exam-Guide.pdf

Le seuil de réussite est fixé à environ 75% de bonnes réponses, qui correspond à un score de passage de 750 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.

L’examen compte en tout 65 questions mais seulement 50 sont évaluées. L'examen comprend 15 questions non notées qui n'affectent pas le score du candidat. AWS collecte des informations sur la performance des candidats sur ces questions non notées afin d'évaluer ces questions en vue d'une utilisation future. Ces questions non notées ne sont pas identifiées lors de l'examen. Dans le résumé des compétences ci-dessous le pourcentage représente uniquement le contenu scorié. Il en est de même pour le nombre de questions par domaine évalué.

Le contenu des tests est réévalué régulièrement par les équipes Amazon Web Services pour refléter les dernières évolutions des services et de la plate-forme AWS.



RESUME DES ACTIVITES PRINCIPALES :

Compétences mobilisables évaluées	Nature des taches évaluées permettant de valider la compétence	Evaluation		
		% de l'évaluation globale	Modalités d'évaluation	Critères
Collecter les données en fonction des caractéristiques des données, du système de collecte et de ses limites.				
<p>Déterminer les caractéristiques opérationnelles du système de collection des données.</p>	<ul style="list-style-type: none"> ▶ Vérifier que la perte des données se situe dans les limites de tolérance en cas de défaillance. ▶ Évaluer les coûts associés à l'acquisition, au transfert et au provisionnement de données à partir de diverses sources vers le système de collecte (par exemple, mise en réseau, bande passante, coûts de migration ETL/données). ▶ Évaluer les scénarios de défaillance que le système de collecte peut subir et prendre des mesures correctives en fonction de l'impact. ▶ Déterminer la persistance des données à différents points de capture des données. ▶ Identifier les caractéristiques de latence du système de collecte. 	<p>18% de l'évaluation globale de l'examen</p>	<p>Examen en ligne avec une variété de questions (cf. détail plus haut) ** Environ 33 minutes (pour 8 à 10 questions notées) sont consacrées à cette compétence</p>	<p>Examen compensatoire, le taux de bonnes réponses doit être au global de 75% minimum</p>
<p>Sélectionner le système de collection des données adéquat en fonction du volume, de la fréquence et de la source des données.</p>	<ul style="list-style-type: none"> ▶ Décrire et caractériser les caractéristiques de volume et de débit des données entrantes (en continu, transactionnel, par lots). ▶ Adapter les caractéristiques de flux des données aux solutions potentielles. ▶ Évaluer les compromis entre les différents services d'ingestion en tenant compte de l'évolutivité, du coût, de la tolérance aux pannes, de la latence, etc. ▶ Expliquer la capacité de débit de différents types de collecte de données et identifier les goulots d'étranglement. ▶ Choisir une solution de collecte qui répond aux contraintes de connectivité du système de données source. 			



<p>Sélectionner un système de collection des données qui traite les principales propriétés des données telles que l'ordre, le format et la compression.</p>	<ul style="list-style-type: none"> ▶ Décrire comment capturer les changements de données à la source. ▶ Discuter de la structure et du format des données, de la compression appliquée et des exigences de chiffrement. ▶ Distinguer l'impact de la livraison de données hors service, de la livraison en double des données et des compromis entre le traitement au plus une fois, exactement une fois et au moins une fois. ▶ Décrire comment transformer et filtrer les données pendant le processus de collecte. 			
<p>Organiser le stockage et la gestion des données</p>		<p>22% de l'évaluation globale de l'examen</p>	<p>Examen en ligne avec une variété de questions (cf. détail plus haut) ** Environ 40 minutes (pour 10 à 12 questions notées) sont consacrées à cette compétence</p>	<p>Examen compensatoire, le taux de bonnes réponses doit être au global de 75% minimum</p>
<p>Déterminer les caractéristiques opérationnelles de la solution de stockage pour l'analyse.</p>	<ul style="list-style-type: none"> ▶ Déterminer le(s) service(s) de stockage approprié(s) en fonction du coût par rapport aux performances. ▶ Comprendre les caractéristiques de durabilité, de fiabilité et de latence de la solution de stockage en fonction des besoins. ▶ Déterminer les exigences d'un système pour une cohérence solide par rapport à la cohérence éventuelle du système de stockage. ▶ Déterminer la solution de stockage appropriée pour répondre aux exigences de fraîcheur des données. 			
<p>Déterminer les modèles d'accès et de récupération des données.</p>	<ul style="list-style-type: none"> ▶ Déterminer la solution de stockage appropriée en fonction des modèles de mise à jour (par exemple, en vrac, transactionnel, en micro-lots). ▶ Déterminer la solution de stockage appropriée en fonction des modèles d'accès (par exemple, accès séquentiel ou aléatoire, utilisation continue ou ad hoc). ▶ Déterminer la solution de stockage appropriée pour répondre aux caractéristiques de modification des données (modifications en ajout uniquement par rapport aux mises à jour). ▶ Déterminer la solution de stockage appropriée pour le stockage à long terme par rapport au stockage transitoire. ▶ Déterminer la solution de stockage appropriée pour les données structurées et semi-structurées. ▶ Déterminer la solution de stockage appropriée pour répondre aux exigences de latence des requêtes. 			



Sélectionner la disposition, le schéma, la structure et le format des données appropriés.	<ul style="list-style-type: none">▶ Déterminer les mécanismes appropriés pour répondre aux exigences d'évolution du schéma.▶ Sélectionner le format de stockage de la tâche.▶ Sélectionner les stratégies de compression/codage pour le format de stockage choisi.▶ Sélectionner les stratégies de tri et de distribution des données et la disposition du stockage pour un accès efficace aux données.▶ Expliquer les implications en termes de coûts et de performances des différentes distributions, mises en page et formats de données (par exemple, la taille et le nombre de fichiers)▶ Mettre en œuvre des schémas de formatage et de partitionnement des données pour une analyse optimisée des données.			
Définir le cycle de vie des données en fonction des modèles d'utilisation et des exigences commerciales.	<ul style="list-style-type: none">▶ Déterminer la stratégie pour répondre aux exigences du cycle de vie des données.▶ Appliquer les règles de cycle de vie et de conservation des données à différentes solutions de stockage.			
Déterminer le système approprié pour cataloguer les données et gérer les métadonnées.	<ul style="list-style-type: none">▶ Évaluer les mécanismes de découverte de sources de données nouvelles et de mises à jour.▶ Évaluer les mécanismes de création et de mise à jour de catalogues de données et de métadonnées.▶ Expliquer les mécanismes de recherche et de récupération des catalogues de données et des métadonnées.▶ Expliquer les mécanismes de marquage et de classification des données.			



Mettre en place la solution de traitement des données appropriée				
Déterminer les exigences de solution de traitement des données appropriées.	<ul style="list-style-type: none"> ▶ Comprendre les exigences de préparation et d'utilisation des données. ▶ Comprendre les différents types de sources et de cibles de données. ▶ Évaluer les besoins en performances et en orchestration. ▶ Évaluer les services appropriés en termes de coût, d'évolutivité et de disponibilité. 	24% de l'évaluation globale de l'examen	Examen en ligne avec une variété de questions (cf. détail plus haut) ** Environ 44 minutes (pour 11 à 13 questions notées) sont consacrées à cette compétence	Examen compensatoire, le taux de bonnes réponses doit être au global de 75% minimum
Concevoir une solution pour transformer et préparer les données pour analyse.	<ul style="list-style-type: none"> ▶ Appliquer les techniques ETL/ELT appropriées pour les charges de travail par lots et en temps réel. ▶ Mettre en œuvre des mécanismes de basculement, de mise à l'échelle et de réplication. ▶ Mettre en œuvre des techniques pour répondre aux besoins de concurrence. ▶ Mettre en œuvre des techniques pour améliorer l'efficacité de l'optimisation des coûts. ▶ Appliquer l'orchestration des flux de travail. ▶ Agréger et enrichir les données pour la consommation en aval. 			
Automatiser et opérationnaliser les solutions de traitement des données.	<ul style="list-style-type: none"> ▶ Mettre en œuvre des techniques automatisées pour des workflows reproductibles. ▶ Appliquer des méthodes pour identifier et récupérer les défaillances de traitement. ▶ Déployer des solutions de journalisation et de surveillance pour permettre l'audit et la traçabilité. 			
Procéder à l'analyse et à la visualisation des données				
Déterminer les caractéristiques opérationnelles de la solution d'analyse et de visualisation.	<ul style="list-style-type: none"> ▶ Déterminer les coûts associés à l'analyse et à la visualisation. ▶ Déterminer l'évolutivité associée à l'analyse. ▶ Déterminer la récupération sur incident et la tolérance aux pannes au sein du RPO/RTO. ▶ Déterminer les caractéristiques de disponibilité d'un outil d'analyse. ▶ Évaluer les présentations dynamiques, interactives et statiques des données. 	18% de l'évaluation globale de l'examen	Examen en ligne avec une variété de questions (cf. détail plus haut) ** Environ 33 minutes (pour	Examen compensatoire, le taux de bonnes réponses doit être au global de 75% minimum



	<ul style="list-style-type: none"> ▶ Transformez les exigences de performance en une approche de visualisation appropriée (pré-calcul et consommation de données statiques par rapport à la consommation de données dynamiques). 		8 à 10 questions notées) sont consacrées à cette compétence	
Sélectionner la solution d'analyse de données appropriée pour un scénario donné.	<ul style="list-style-type: none"> ▶ Évaluer et comparer les solutions d'analyse. ▶ Sélectionner le bon type d'analyse en fonction du cas d'utilisation du client (streaming, interactif, collaboratif, opérationnel). 			
Sélectionner la solution de visualisation des données appropriée pour un scénario donné.	<ul style="list-style-type: none"> ▶ Évaluer les capacités d'output (résultats) d'une solution d'analyse donnée (métriques, KPI, tableau, API). ▶ Choisir la méthode appropriée pour la livraison des données (par exemple, Web, mobile, e-mail, blocs-notes collaboratifs). ▶ Choisir et définir le calendrier d'actualisation des données approprié. ▶ Choisir les outils appropriés pour différentes exigences d'actualisation des données (par exemple, Amazon Elasticsearch Service vs Amazon QuickSight vs ordinateurs portables Amazon EMR). ▶ Comprendre les capacités des outils de visualisation pour les cas d'utilisation interactifs (par exemple, hiérarchisation vers le bas, exploration et pivot). ▶ Mettre en œuvre le mécanisme d'accès aux données approprié (par exemple, en mémoire ou en accès direct). ▶ Implémenter une solution intégrée à partir de plusieurs sources de données hétérogènes. 			
Assurer la sécurité des données utilisées				
Sélectionner les mécanismes d'authentification et d'autorisation appropriés.	<ul style="list-style-type: none"> ▶ Mettre en œuvre des méthodes d'authentification appropriées (par exemple, accès fédéré, SSO, IAM). ▶ Mettre en œuvre des méthodes d'autorisation appropriées (par exemple, stratégies, ACL, autorisations au niveau de la table/colonne). ▶ Mettre en œuvre des mécanismes de contrôle d'accès appropriés (par exemple, groupes de sécurité, contrôle basé sur les rôles). 	18% de l'évaluation globale de l'examen	Examen en ligne avec une variété de questions (cf. détail plus haut) **	Examen compensatoire, le taux de bonnes réponses doit être au global de 75% minimum



Appliquer des techniques de protection et de cryptage des données.	<ul style="list-style-type: none">▶ Déterminer les besoins de cryptage et de masquage des données.▶ Appliquer différentes approches de chiffrement (chiffrement côté serveur, chiffrement côté client, AWS KMS, AWS CloudHSM).▶ Mettre en œuvre des mécanismes de chiffrement au repos et en transit.▶ Mettre en œuvre des techniques d'obfuscation et de masquage des données.▶ Appliquer les principes de base de rotation des clés et de gestion des secrets		Environ 33 minutes (pour 8 à 10 questions notées) sont consacrées à cette compétence	
Appliquer des contrôles de gouvernance et de conformité des données.	<ul style="list-style-type: none">▶ Déterminer les exigences de gouvernance et de conformité des données.▶ Comprendre et configurer la journalisation des accès et des audits dans les services d'analyse de données.▶ Mettre en œuvre des contrôles appropriés pour répondre aux exigences de conformité.			