

Formation & Certification • DP-203

Ingénieur en données sur Microsoft Azure

Formation présentielle et en ligne à la demande. Concevez et mettez en œuvre des solutions de gestion et d'analyse de données sur Microsoft Azure : stockage, traitement par lots et en flux, sécurité et optimisation des pipelines de données.

Distributeur officiel Certiport

Formateurs certifiés Microsoft

Préparation à l'examen DP-203

Durée 4 jours (28 h)	Examen DP-203
Modalité Présentiel / Distanciel à la demande	Niveau Intermédiaire

INSCRIPTION / RÉSERVATION



Je m'inscris
maintenant



OBJECTIFS PÉDAGOGIQUES

- Concevoir et implémenter des solutions de stockage Azure adaptées aux workloads analytiques et de streaming.
- Développer des pipelines de données robustes pour l'ingestion, la transformation et le traitement batch/stream.
- Assurer la sécurité, la supervision et l'optimisation des solutions de données dans le cloud.
- Se préparer efficacement à l'examen de certification DP-203.

PUBLIC CIBLE

- Ingénieurs en données/Analystes de données/Administrateurs de bases de données.
- Architectes de solutions cloud/Développeurs souhaitant se spécialiser dans la gestion des données sur Azure.

PRÉREQUIS

- Connaissances de base des concepts de bases de données et de l'architecture cloud.
- Expérience en SQL et en modélisation de données.
- Familiarité avec un langage de programmation (Python, R, etc.).

PROGRAMME DE LA FORMATION – DÉTAILLÉ

Concevoir et implémenter le stockage des données

Implémenter une stratégie de partition

- Implémenter une stratégie de partition pour les fichiers.
- Implémenter une stratégie de partition pour les charges de travail analytiques.
- Implémenter une stratégie de partition pour les charges de travail de streaming.
- Implémenter une stratégie de partition pour Azure Synapse Analytics.
- Identifier quand le partitionnement est nécessaire dans Azure Data Lake Storage Gen2.

Concevoir et implémenter la couche d'exploration des données

- Créer et exécuter des requêtes en utilisant une solution de calcul qui tire parti de SQL serverless et des clusters Spark.
- Recommander et implémenter des modèles de base de données Azure Synapse Analytics.
- Envoyer (push) la traçabilité des données nouvelles ou mises à jour vers Microsoft Purview.
- Parcourir et rechercher des métadonnées dans le catalogue de données Microsoft Purview.

Développer le traitement des données

Ingérer et transformer des données

- Concevoir et implémenter des chargements de données incrémentiels.
- Transformer des données en utilisant Apache Spark.
- Transformer des données en utilisant Transact-SQL (T-SQL) dans Azure Synapse Analytics.
- Ingérer et transformer des données en utilisant des pipelines Azure Synapse ou Azure Data Factory.
- Transformer des données à l'aide d'Azure Stream Analytics.

PROGRAMME DE LA FORMATION – DÉTAILLÉ

Nettoyer les données

- Gérer les données en doublon.
- Éviter les données dupliquées à l'aide de la fonctionnalité Exactly once de Azure Stream Analytics.
- Gérer les données manquantes.
- Gérer les données tardives.
- Fractionner les données.
- Fragmenter du code JSON.
- Encoder et décoder des données.
- Configurer la gestion des erreurs pour une transformation.
- Normaliser et dénormaliser des données.
- Effectuer une analyse exploratoire des données.

Développer des solutions de traitement par lots

- Développer des solutions de traitement par lots en utilisant Azure Data Lake Storage Gen2, Azure Databricks, Azure Synapse Analytics et Azure Data Factory.
- Utiliser PolyBase pour charger des données dans un pool SQL.
- Implémenter Azure Synapse Link et interroger les données répliquées.
- Créer des pipelines de données.
- Mettre les ressources à l'échelle.
- Configurer la taille de lot.
- Créer des tests pour les pipelines de données.
- Intégrer des notebooks Jupyter ou Python dans un pipeline de données.
- Données de lot Upsert.
- Rétablir les données à un état précédent.
- Configurer la gestion des exceptions.
- Configurer la conservation des lots.
- Lire et écrire dans un lac delta.

PROGRAMME DE LA FORMATION – DÉTAILLÉ

Développer une solution de traitement de flux

- Créer une solution de traitement de flux en utilisant Stream Analytics et Azure Event Hubs.
- Traiter des données en utilisant le streaming structuré Spark.
- Créer des agrégats fenêtrés.
- Gérer la dérive de schéma.
- Traiter des données de série chronologique.
- Traiter des données sur plusieurs partitions.
- Traiter au sein d'une même partition.
- Configurer des points de contrôle et un filigranage pendant le traitement.
- Mettre les ressources à l'échelle.
- Créer des tests pour les pipelines de données.
- Optimiser les pipelines à des fins analytiques ou transactionnelles.
- Gestion des interruptions.
- Configurer la gestion des exceptions.
- Transmettre des données Upsert en continu.
- Relire les données de flux archivés.
- Lire et écrire dans un lac delta.

Gérer les lots et les pipelines

- Déclencher des lots.
- Gérer les chargements par lots ayant échoué.
- Valider les chargements par lots.
- Gérer des pipelines de données dans des pipelines Azure Data Factory ou Azure Synapse.
- Planifier des pipelines de données dans des pipelines Azure Data Factory ou Azure Synapse.
- Implémenter le contrôle de version pour les artefacts de pipeline.
- Gérer des travaux Spark dans un pipeline.

PROGRAMME DE LA FORMATION – DÉTAILLÉ

Sécuriser, superviser et optimiser le stockage et le traitement des données

Implémenter la sécurité des données

- Implémenter le masquage des données.
- Chiffrer les données au repos ou en déplacement.
- Implémenter une sécurité au niveau des lignes et des colonnes.
- Implémenter le contrôle d'accès en fonction du rôle (RBAC) Azure.
- Implémenter des listes de contrôle d'accès de type POSIX pour Data Lake Storage Gen2.
- Implémenter une politique de conservation des données.
- Implémenter des points de terminaison sécurisés (privés et publics).
- Implémenter des jetons de ressource dans Azure Databricks.
- Charger un DataFrame avec des informations sensibles.
- Écrire des données chiffrées dans des tables ou des fichiers Parquet.
- Gérer des informations sensibles.

Superviser le stockage des données et le traitement des données

- Implémenter la journalisation utilisée par Azure Monitor.
- Configurer des services de supervision.
- Superviser le traitement des flux.
- Mesurer les performances du déplacement des données.
- Superviser et mettre à jour les statistiques sur les données d'un système.
- Superviser les performances des pipelines de données.
- Mesurer les performances des requêtes.
- Planifier et superviser les tests de pipeline.
- Interpréter les métriques et les journaux Azure Monitor.
- Implémenter une stratégie d'alerte de pipeline.

PROGRAMME DE LA FORMATION – DÉTAILLÉ

Optimiser et résoudre les problèmes de stockage et de traitement des données

- Compacter des petits fichiers.
- Gérer l'asymétrie dans les données.
- Gérer le déversement de données.
- Optimiser la gestion des ressources.
- Optimiser les requêtes en utilisant des indexeurs.
- Optimiser les requêtes en utilisant un cache.
- Résoudre les problèmes d'un travail Spark ayant échoué.
- Résoudre les problèmes d'exécution d'un pipeline ayant échoué, y compris les activités exécutées dans des services externes.