

Programme de formation

DP-203: Data Engineering on Microsoft Azure

(Préparation certification Microsoft DP-203)

Mis à jour le 09/02/2024

DESCRIPTION DE LA FORMATION :

Les ingénieurs de données de Azure intègrent, transforment et consolident les données provenant de divers systèmes de données structurées et non structurées dans des structures qui conviennent à l'élaboration de solutions analytiques. Cette formation vous aidera pour y parvenir en adoptant les meilleurs pratiques.

OBJECTIFS PEDAGOGIQUES :

A l'issue de cette formation, les participants seront en capacité de :

- Démarrer avec l'ingénierie des données sur Azure
- Créer des solutions d'analytique données avec des pools SQL serverless Azure Synapse
- Appliquer l'Engineering données avec des pools Azure Synapse Apache Spark
- Utiliser des entrepôts de données avec Azure Synapse Analytics
- Transférer et transformer des données avec des pipelines Azure Synapse Analytics
- Utiliser des solutions de traitement transactionnel et analytique hybrides avec Azure Synapse Analytics
- Implémenter une solution de streaming de données avec Azure Stream Analytics
- Gérer les données dans toute l'entreprise

MÉTHODES PÉDAGOGIQUES :

- Cette formation sera principalement constituée de théorie et d'ateliers techniques qui permettront d'être rapidement opérationnel.
- Support : un support de cours officiel Microsoft sera remis aux participants au format électronique.
- Evaluation : Les acquis sont évalués tout au long de la formation par le formateur (Prérequis évalués avant la formation, questions régulières, travaux pratiques, QCM ou autres méthodes).
- Formateur : le tout animé par un consultant-formateur expérimenté, nourri d'une expérience terrain, et accrédité Microsoft Certified Trainer.
- Satisfaction : à l'issue de la formation, chaque participant répond à un questionnaire d'évaluation qui est ensuite analysé en vue de maintenir et d'améliorer la qualité de nos formations.
- Suivi : une feuille d'émargement par demi-journée de présence est signée par chacun des participants.
- Cette formation peut être dispensée en format inter-entreprises ou intra-entreprise sur demande et en mode présentiel comme en distanciel.

PROGRAMME DE FORMATION :

Découvrir l'engineering données dans Azure

- Identifier les tâches courantes de l'engineering données.
- Décrire les concepts courants de l'engineering données.
- Identifier les services Azure pour l'engineering données.

Découvrir Azure Data Lake Storage Gen2

- Décrire les principaux avantages et fonctionnalités d'Azure Data Lake Storage Gen2.
- Activer Azure Data Lake Storage Gen2 dans un compte Stockage Azure.
- Comparer Azure Data Lake Storage Gen2 et le Stockage Blob Azure.
- Décrire les situations dans lesquelles Azure Data Lake Storage Gen2 s'intègre aux étapes du traitement analytique.
- Utiliser Azure Data Lake Storage Gen2 dans les charges de travail analytiques courantes.

Découvrir Azure Synapse Analytics

- Identifier les problèmes métier traités par Azure Synapse Analytics adresse.
- Décrire les fonctionnalités principales d'Azure Synapse Analytics.
- Déterminer quand utiliser Azure Synapse Analytics.

Utiliser un pool SQL serverless Azure Synapse pour interroger des fichiers dans un Data Lake

- Identifier les fonctionnalités et les cas d'usage des pools SQL serverless dans Azure Synapse Analytics.
- Interroger des fichiers CSV, JSON et Parquet avec un pool SQL serverless.
- Créer des objets de base de données externes dans un pool SQL serverless.

Utiliser des pools SQL serverless Azure Synapse pour transformer des données dans un Data Lake

- Utiliser une instruction CREATE EXTERNAL TABLE AS SELECT (CETAS) pour transformer des données.
- Encapsuler une instruction CETAS dans une procédure stockée.
- Inclure une procédure stockée de transformation de données dans un pipeline.

Créer une base de données lake dans Azure Synapse Analytics

- Comprendre les concepts et les composants de la base de données lake.
- Décrire les modèles de base de données dans Azure Synapse Analytics.
- Créer une base Datalake.

Sécuriser les données et gérer les utilisateurs dans les pools SQL serverless Azure Synapse

- Choisir une méthode d'authentification dans les pools SQL serverless Azure Synapse ;
- Gérer les utilisateurs dans les pools SQL serverless Azure Synapse ;
- Gérer les autorisations utilisateur dans les pools SQL serverless Azure Synapse.

Analyser des données avec Apache Spark dans Azure Synapse Analytics

- Identifier les principales fonctionnalités d'Apache Spark.
- Configurer un pool Spark dans Azure Synapse Analytics.
- Exécuter du code pour charger, analyser et visualiser des données dans un notebook Spark.

Transformer des données avec Spark dans Azure Synapse Analytics

- Utiliser Apache Spark pour modifier et enregistrer des dataframes.
- Partitionner les fichiers de données pour améliorer les performances et la scalabilité.
- Transformer des données avec SQL.

Utiliser Delta Lake dans Azure Synapse Analytics

- Décrire les principales fonctionnalités de Delta Lake.
- Créer et utiliser des tables Delta Lake dans un pool Spark Synapse Analytics.
- Créer des tables de catalogue Spark pour les données Delta Lake.
- Utiliser des tables Delta Lake pour la diffusion en continu de données.
- Interroger des tables Delta Lake à partir d'un pool SQL Synapse Analytics.

Analyser les données dans un entrepôt de données relationnelles

- Concevoir un schéma pour un entrepôt de données relationnel.
- Créer des tables de faits, de dimension et de mise en lots.
- Utiliser SQL pour charger des données dans des tables d'entrepôt de données.
- Utiliser SQL pour interroger des tables d'entrepôt de données relationnelles.

Charger des données dans un entrepôt de données relationnel

- Charger des tables de mise en lots dans un entrepôt de données.
- Charger des tables de dimension dans un entrepôt de données.
- Charger des dimensions de temps dans un entrepôt de données.
- Charger des dimensions variables à évolution lente dans un entrepôt de données.
- Charger des tables de faits dans un entrepôt de données.
- Effectuer des optimisations postchargement dans un entrepôt de données.

Gérer et superviser les activités d'un entrepôt de données dans Azure Synapse Analytics

- Mettre à l'échelle des ressources de calcul dans Azure Synapse Analytics.
- Mettre en pause le calcul dans Azure Synapse Analytics.
- Gérer des charges de travail dans Azure Synapse Analytics.
- Utiliser Azure Advisor pour passer en revue des recommandations.
- Utiliser des vues de gestion dynamique pour identifier et résoudre les problèmes de performance des requêtes.

Sécuriser un entrepôt de données dans Azure Synapse Analytics

- Découvrir les options de sécurité réseau pour Azure Synapse Analytics.
- Configurer un accès conditionnel.
- Configurer l'authentification.
- Gérer l'autorisation par le biais de la sécurité au niveau des colonnes et des lignes.
- Gérer les données sensibles avec le masquage dans Dynamic Data.
- Implémenter le chiffrement dans Azure Synapse Analytics.

Créer un pipeline de données dans Azure Synapse Analytics

- Décrire les concepts de base des pipelines Azure Synapse Analytics.
- Créer un pipeline dans Azure Synapse Studio.
- Implémenter une activité de flux de données dans un pipeline.
- Lancer et superviser des exécutions de pipeline.

Utiliser des notebooks Spark dans un pipeline Azure Synapse

- Décrire l'intégration du notebook et du pipeline.
- Utiliser une activité de notebook Synapse dans un pipeline.
- Utiliser des paramètres avec une activité de notebook.

Planifier un traitement transactionnel et analytique hybride avec Azure Synapse Analytics

- Décrire les modèles de traitement transactionnel et analytique hybride.
- Identifier les services Azure Synapse Link pour HTAP.

Implémenter Azure Synapse Link avec Azure Cosmos DB

- Configurer un compte Azure Cosmos DB pour utiliser Azure Synapse Link.
- Créer un conteneur avec le magasin analytique activé.
- Créer un service lié pour Azure Cosmos DB.
- Analyser les données liées à l'aide de Spark et de Synapse SQL.

Implémenter Azure Synapse Link pour SQL

- Comprendre les concepts clés et les fonctionnalités d'Azure Synapse Link pour SQL.
- Configurer Azure Synapse Link pour Azure SQL Database.
- Configurer Azure Synapse Link pour Microsoft SQL Server.

Découvrir Azure Stream Analytics

- Comprendre les flux de données.
- Comprendre le traitement des événements.
- Comprendre les fonctions de fenêtre.
- Utiliser Azure Stream Analytics.

Ingérer des données de streaming à l'aide d'Azure Stream Analytics et Azure Synapse Analytics

- Décrire les scénarios d'ingestion de flux courants pour Azure Synapse Analytics.
- Configurer les entrées et les sorties pour une tâche Azure Stream Analytics.
- Définir une requête pour ingérer des données en temps réel dans Azure Synapse Analytics.
- Exécuter une tâche pour ingérer des données en temps réel et consommer ces données dans Azure Synapse Analytics.

Visualiser des données en temps réel avec Azure Stream Analytics et Power BI

- Configurer une sortie Stream Analytics pour Power BI.
- Utiliser une requête Stream Analytics pour écrire des données dans Power BI.
- Créer une visualisation de données en temps réel dans Power BI.

Découvrir Microsoft Purview

- Déterminer si Microsoft Purview est adapté à vos besoins de découverte et de gouvernance des données.
- Décrire comment les fonctionnalités de Microsoft Purview fonctionnent pour assurer la découverte et la gouvernance des données.

Découvrir les données approuvées avec Microsoft Purview

- Parcourir, rechercher et gérer les ressources du catalogue de données.
- Utiliser des ressources de catalogue de données avec Power BI.
- Utiliser Microsoft Purview dans Azure Synapse Studio.

Découvrir les artefacts de données de catalogue à l'aide de Microsoft Purview

- Décrire la classification des ressources dans Microsoft Purview.

Gérer les ressources Power BI avec Microsoft Purview

- Enregistrer et numériser un locataire Power BI.
- Utiliser les fonctions de recherche et de navigation pour rechercher des ressources de données.
- Décrire les détails du schéma et le suivi de la traçabilité des données de ressources de données Power BI.

Intégrer Microsoft Purview et Azure Synapse Analytics

- Cataloguer les ressources de base de données Azure Synapse Analytics dans Microsoft Purview.
- Configurer l'intégration de Microsoft Purview dans Azure Synapse Analytics.
- Rechercher dans le catalogue Microsoft Purview à partir de Synapse Studio.
- Suivre la traçabilité des données dans les activités de pipelines Azure Synapse Analytics.

Explorer Azure Databricks

- Provisionner un espace de travail Azure Databricks.
- Identifier les principales charges de travail et entités pour Azure Databricks.
- Décrire les concepts clés d'une solution Azure Databricks.

Utiliser Apache Spark dans Azure Databricks

- Décrire les éléments clés de l'architecture Apache Spark.
- Créer et configurer un cluster Spark.
- Décrire les cas d'usage pour Spark.
- Utiliser Spark pour traiter et analyser les données stockées dans des fichiers.
- Utiliser Spark pour visualiser les données.

Utiliser Delta Lake dans Azure Databricks

- Décrire les principales fonctionnalités de Delta Lake.
- Créer et utiliser des tables Delta Lake dans Azure Databricks.
- Créer des tables de catalogue Spark pour les données Delta Lake.
- Utiliser des tables Delta Lake pour la diffusion en continu de données.

Utiliser des entrepôts SQL dans Azure Databricks

- Créer et configurer des entrepôts SQL dans Azure Databricks.
- Créer des bases de données et des tables.
- Créer des requêtes et des tableaux de bord.

Exécuter des notebooks Azure Databricks avec Azure Data Factory

- Décrire comment les notebooks Azure Databricks peuvent être exécutés dans un pipeline.
- Créer un service lié Azure Data Factory pour Azure Databricks.
- Utiliser une activité Notebook dans un pipeline.
- Passer des paramètres à un notebook.

PRÉREQUIS :

Les candidats doivent avoir une expertise en matière d'intégration, de transformation et doivent savoir consolider divers systèmes de données structurés et non structurés dans un outils adapté à la création de solutions d'analyses. Il faut également une bonne connaissance des langages tels que SQL, Python ou Scala et comprendre l'architecture des données.

Les candidats doivent avoir suivi la formation AZ-900 Azure Fundamentals et DP-900 Data Fundamentals ou avoir un niveau d'expérience équivalent.

Un niveau d'anglais B1 est recommandé, retrouvez les niveaux sur ce lien : [Classification des niveaux de langue](#)

PRE-CERTIFICATION :

Cette formation ouvre la porte à la certification Microsoft « DP-203 – Data Engineering on Microsoft Azure ».

DUREE : 4 jours (28 heures)

INTERLOCUTEURS : Data Engineers, Data Scientists

NIVEAU : Intermédiaire