Programme de cours
Ingénierie des données sur Microsoft Azure (remplacera DP-200 et DP-201)

Durée :	4 jours soit 28 heures	Référence :	DP-203T00

Public visé :

Cette formation est destinée aux professionnels de la donnée, architectes de données qui souhaitent en savoir plus sur l'ingénierie des données et la création de solutions analytiques à l'aide des technologies de plate-forme de données qui existent sur Microsoft Azure.
Cette formation peut également être destinée, aux analystes de données et data scientists qui travaillent avec des solutions analytiques basées sur Microsoft Azure.

Pré-requis :

Pour suivre cette formation, les apprenants doivent :

- Avoir suivi le cours AZ-900T00 "les fondamentaux de Microsoft Azure" ou en posséder les connaissances équivalentes.
- Avoir suivi le cours DP-900T00 " Microsoft Azure Data Fundamentals" ou en posséder les connaissances équivalentes.

Objectifs pédagogiques :

A l'issue de la formation, les apprenants auront acquis les compétences suivantes :

- Explorez les options de calcul et de stockage pour les charges de travail d'ingénierie des données dans Azure
- Exécuter des requêtes interactives à l'aide de pools SQL sans serveur
- Effectuer l'exploration et la transformation des données dans Azure Databricks
- Explorer, transformer et charger des données dans le Data Warehouse à l'aide d'Apache Spark
- Ingérer et charger des données dans l'entrepôt de données
- Transformez les données avec Azure Data Factory ou Azure Synapse Pipelines
- Intégrer les données des notebooks avec Azure Data Factory ou Azure Synapse Pipelines
- Prise en charge du traitement analytique transactionnel hybride (HTAP) avec Azure Synapse Link
- Assurer la sécurité de bout en bout avec Azure Synapse Analytics
- Effectuez un traitement de flux en temps réel avec Stream Analytics
- Créez une solution de traitement de flux avec Event Hubs et Azure Databricks

Modalités pédagogiques :

Session dispensée en présentiel ou téléprésentiel, selon la modalité inter-entreprises ou intra-entreprises sur mesure.
La formation est animée par un(e) formateur(trice) durant toute la durée de la session et présentant une suite de modules théoriques clôturés par des ateliers pratiques validant l'acquisition des connaissances. Les ateliers peuvent être accompagnés de Quizz.
L'animateur(trice) présente la partie théorique à l'aide de support de présentation, d'animation réalisée sur un environnement de démonstration.
En présentiel comme en téléprésentiel, l'animateur(trice) accompagne les participants durant la réalisation des ateliers.

Moyens et supports pédagogiques :

Cadre présentiel
Salles de formation équipées et accessibles aux personnes à mobilité réduite.
- Un poste de travail par participant
- Un support de cours numérique ou papier (au choix)
- Un bloc-notes + stylo
- Vidéoprojection sur tableau blanc
- Connexion Internet
- Accès extranet pour partage de documents et émargement électronique

Cadre téléprésentiel
Session dispensée via notre solution iClassroom s'appuyant sur Microsoft Teams.
- Un compte Office 365 par participant
- Un poste virtuel par participant
- Un support numérique (PDF ou Web)
- Accès extranet pour partage de documents et émargement électronique

Modalités d'évaluation et suivi :

Avant
Afin de valider le choix d'un programme de formation, une évaluation des prérequis est réalisée à l'aide d'un questionnaire en ligne ou lors d'un échange avec le formateur(trice) qui validera la base de connaissances nécessaires.
Pendant
Après chaque module théorique, un ou des ateliers pratiques permettent la validation de l'acquisition des connaissances. Un Quizz peut accompagner l'atelier pratique.
Après
Un examen de certification si le programme de formation le prévoit dans les conditions de l'éditeur ou du centre de test (TOSA, Pearson Vue, ENI, PeopleCert)
Enfin
Un questionnaire de satisfaction permet au participant d'évaluer la qualité de la prestation.

Description / Contenu

Module 1 : Explorer les options de calcul et de stockage pour les charges de travail d'ingénierie des données :

- Introduction à Azure Synapse Analytics
- Décrire Azure Databricks
- Introduction au stockage Azure Data Lake
- Décrire l'architecture du lac Delta

- Travailler avec des flux de données à l'aide d'Azure Stream Analytics

Atelier : Explorer les options de calcul et de stockage pour les charges de travail d'ingénierie des données

- Combinez le streaming et le traitement par lots avec un seul pipeline
- Organiser le lac de données en niveaux de transformation de fichiers
- Indexer le stockage du lac de données pour l'accélération des requêtes et de la charge de travail

Module 2 : Exécuter des requêtes interactives à l'aide des pools SQL sans serveur Azure Synapse Analytics :

- Explorez les capacités des pools SQL sans serveur Azure Synapse
- Interroger les données dans le lac à l'aide des pools SQL sans serveur Azure Synapse
- Créer des objets de métadonnées dans les pools SQL sans serveur Azure Synapse
- Sécurisez les données et gérez les utilisateurs dans les pools SQL sans serveur Azure Synapse

Atelier : Exécuter des requêtes interactives à l'aide de pools SQL sans serveur

- Interrogez les données Parquet avec des pools SQL sans serveur
- Créer des tables externes pour les fichiers Parquet et CSV
- Créer des vues avec des pools SQL sans serveur
- Accès sécurisé aux données dans un lac de données lors de l'utilisation de pools SQL sans serveur
- Configurer la sécurité du lac de données à l'aide du contrôle d'accès basé sur les rôles (RBAC) et de la liste de contrôle d'accès

Module 3 : Exploration et transformation des données dans Azure Databricks :

- Décrire Azure Databricks
- Lire et écrire des données dans Azure Databricks
- Utiliser des DataFrames dans Azure Databricks
- Travailler avec les méthodes avancées DataFrames dans Azure Databricks

Atelier : Exploration et transformation des données dans Azure Databricks

- Utilisez DataFrames dans Azure Databricks pour explorer et filtrer les données
- Mettre en cache un DataFrame pour des requêtes ultérieures plus rapides
- Supprimer les données en double
- Manipuler les valeurs de date/heure
- Supprimer et renommer les colonnes DataFrame
- Agréger les données stockées dans un DataFrame

Module 4 : Explorer, transformer et charger des données dans le Data Warehouse à l'aide d'Apache Spark :

- Comprendre l'ingénierie Big Data avec Apache Spark dans Azure Synapse Analytics
- Ingérer des données avec des blocs-notes Apache Spark dans Azure Synapse Analytics
- Transformez les données avec les DataFrames dans les pools Apache Spark dans Azure Synapse Analytics
- Intégrer les pools SQL et Apache Spark dans Azure Synapse Analytics

Atelier : Explorer, transformer et charger des données dans le Data Warehouse à l'aide d'Apache Spark

- Exécuter l'exploration de données dans Synapse Studio
- Ingérer des données avec des blocs-notes Spark dans Azure Synapse Analytics
- Transformez des données avec des DataFrames dans des pools Spark dans Azure Synapse Analytics
- Intégrer les pools SQL et Spark dans Azure Synapse Analytics

Module 5 : Ingérer et charger des données dans l'entrepôt de données :

- Utiliser les bonnes pratiques de chargement de données dans Azure Synapse Analytics
- Ingestion à l'échelle du pétaoctet avec Azure Data Factory

Atelier : Ingérer et charger des données dans le Data Warehouse

- Réalisez une ingestion à l'échelle du pétaoctet avec Azure Synapse Pipelines
- Importer des données avec PolyBase et COPIER à l'aide de T-SQL
- Utiliser les bonnes pratiques de chargement de données dans Azure Synapse Analytics

Module 6 : Transformer les données avec Azure Data Factory ou Azure Synapse Pipelines :

- Intégration de données avec Azure Data Factory ou Azure Synapse Pipelines
- Transformation sans code à grande échelle avec Azure Data Factory ou Azure Synapse Pipelines

Atelier : Transformer des données avec Azure Data Factory ou Azure Synapse Pipelines

- Exécutez des transformations sans code à grande échelle avec Azure Synapse Pipelines
- Créer un pipeline de données pour importer des fichiers CSV mal formatés
- Créer des flux de données cartographiques

Module 7 : Orchestrer le mouvement et la transformation des données dans Azure Synapse Pipelines

- Orchestrer le mouvement et la transformation des données dans Azure Data Factory

Atelier : Orchestrer le mouvement et la transformation des données dans Azure Synapse Pipelines

- Intégrer les données des notebooks avec Azure Data Factory ou Azure Synapse Pipelines

Module 8 : Sécurité de bout en bout avec Azure Synapse Analytics :

- Sécurisez un entrepôt de données dans Azure Synapse Analytics
- Configurer et gérer les secrets dans Azure Key Vault
- Mettre en œuvre des contrôles de conformité pour les données sensibles

Atelier : Sécurité de bout en bout avec Azure Synapse Analytics

- Infrastructure sécurisée prenant en charge Azure Synapse Analytics
- Sécurisez l'espace de travail Azure Synapse Analytics et les services managés
- Sécuriser les données de l'espace de travail Azure Synapse Analytics

Module 9 : Prise en charge du traitement analytique transactionnel hybride (HTAP) avec Azure Synapse Link :

- Concevoir un traitement transactionnel et analytique hybride à l'aide d'Azure Synapse Analytics
- Configurer Azure Synapse Link avec Azure Cosmos DB
- Interroger Azure Cosmos DB avec les pools Apache Spark

- Interroger Azure Cosmos DB avec des pools SQL sans serveur

Atelier : Prise en charge du traitement analytique transactionnel hybride (HTAP) avec Azure Synapse Link

- Configurer Azure Synapse Link avec Azure Cosmos DB
- Interroger Azure Cosmos DB avec Apache Spark pour Synapse Analytics
- Interrogez Azure Cosmos DB avec un pool SQL sans serveur pour Azure Synapse Analytics

Module 10 : Traitement de flux en temps réel avec Stream Analytics :

- Activer une messagerie fiable pour les applications Big Data à l'aide d'Azure Event Hubs
- Travailler avec des flux de données à l'aide d'Azure Stream Analytics
- Ingérer des flux de données avec Azure Stream Analytics

Atelier : Traitement de flux en temps réel avec Stream Analytics

- Utilisez Stream Analytics pour traiter les données en temps réel des Event Hubs
- Utilisez les fonctions de fenêtrage de Stream Analytics pour créer des agrégats et générer des sorties vers Synapse Analytics
- Mettre à l'échelle le travail Azure Stream Analytics pour augmenter le débit via le partitionnement
- Repartitionner l'entrée du flux pour optimiser la parallélisation

Module 11 : Créer une solution de traitement de flux avec Event Hubs et Azure Databricks :

- Traiter les données de streaming avec le streaming structuré Azure Databricks

Atelier : Créer une solution de traitement de flux avec Event Hubs et Azure Databricks

- Explorez les principales fonctionnalités et utilisations du streaming structuré
- Diffusez des données à partir d'un fichier et écrivez-les dans un système de fichiers distribué
- Utilisez des fenêtres coulissantes pour agréger des morceaux de données plutôt que toutes les données
- Appliquer un filigrane pour supprimer les données obsolètes
- Connectez-vous aux flux de lecture et d'écriture Event Hubs