

# Programme de formation

## DP-203: Data Engineering on Microsoft Azure

(Préparation certification Microsoft DP-203)

### 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 :

- Explorer Azure Synapse Analytics, Azure Databricks, Azure Data Lake Storage et Azure Stream Analytics
- Utiliser un pool SQL serverless Azure Synapse pour interroger des fichiers dans un lac de données
- Créer une base de données lake dans Azure Synapse Analytics
- Sécuriser les données et gérer les utilisateurs dans les pools SQL serverless Azure Synapse
- Utiliser Apache Spark et Delta Lake dans Azure Databricks
- Analyser des données avec Apache Spark dans Azure Synapse Analytics
- Intégrer des pools SQL et Apache Spark dans Azure Synapse Analytics
- Utiliser les bonnes pratiques en matière de chargement de données dans Azure Synapse Analytics
- Ingérer des données à l'échelle du pétaoctet avec un pipeline Azure Synapse ou Azure Data Factory
- Intégrer des données à un pipeline Azure Data Factory ou Azure Synapse
- Effectuer une transformation à grande échelle sans code avec un pipeline Azure Data Factory ou Azure Synapse
- Orchestrer le déplacement et la transformation des données dans le pipeline Azure Data Factory ou Azure Synapse
- Planifier un traitement transactionnel et analytique hybride avec Azure Synapse Analytics
- Implémenter Azure Synapse Link avec Azure Cosmos DB
- Sécuriser un entrepôt de données dans Azure Synapse Analytics
- Configurer et gérer des secrets dans Azure Key Vault
- Implémenter des contrôles de compatibilité pour les données sensibles
- Activer une messagerie fiable pour des applications Big Data avec Azure Event Hubs

## 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 :

### Connaitre 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.

### Explorer Azure Databricks

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

### Introduire Azure Data Lake Storage

- Déterminer à quel moment il convient d'utiliser Azure Data Lake Storage Gen2.
- Créer un compte de stockage Azure en utilisant le portail Azure.
- Comparer Azure Data Lake Storage Gen2 et le Stockage Blob Azure.
- Explorer les phases du traitement du Big Data à l'aide d'Azure Data Lake Store.
- Lister les plateformes open source prises en charge.

### Démarrer avec Azure Stream Analytics

- Comprendre les flux de données et le traitement des événements.
- Démarrer avec Azure Stream Analytics.

### Utiliser un pool SQL serverless Azure Synapse pour interroger des fichiers dans un lac de données

- Identifier les fonctionnalités et 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.

### 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 de données de lac.

### **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.

### **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.

### **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.

### **Intégrer des pools SQL et Apache Spark dans Azure Synapse Analytics**

- Décrire les méthodes d'intégration entre les pools SQL et Spark dans Azure Synapse Analytics.
- Comprendre les cas d'utilisation de l'intégration des pools SQL et Spark.
- S'authentifier auprès d'Azure Synapse Analytics.
- Transférer des données entre des pools SQL et Spark dans Azure Synapse Analytics.
- S'authentifier entre des pools Spark et SQL dans Azure Synapse Analytics.
- Intégrer des pools SQL et Spark dans Azure Synapse Analytics.
- Externaliser l'utilisation des pools Spark dans l'espace de travail Azure Synapse.
- Transférer des données en dehors de l'espace de travail Synapse à l'aide de l'authentification SQL.
- Transférer des données en dehors de l'espace de travail Synapse à l'aide du connecteur PySpark.
- Transformer des données dans Apache Spark et les réécrire dans un pool SQL au sein d'Azure Synapse Analytics.

### **Suivre les bonnes pratiques en matière de chargement de données dans Azure Synapse Analytics**

- Comprendre les objectifs du chargement des données.
- Expliquer le chargement des méthodes dans Azure Synapse Analytics.
- Gérer les fichiers de données sources et les mises à jour de base de données unique.
- Configurer des comptes de chargement de données dédiées.
- Gérer l'accès simultané à Azure Synapse Analytics.
- Implémenter la gestion de la charge de travail.

### **Ingérer les données à l'échelle du pétaoctet avec un pipeline Azure Synapse ou Azure Data Factory**

- Répertorier les méthodes d'ingestion de Data Factory.
- Décrire les connecteurs Data Factory.
- Comprendre les considérations relatives à la sécurité de l'ingestion des données.

### **Intégrer des données à un pipeline Azure Data Factory ou Azure Synapse**

- Comprendre Azure Data Factory.
- Décrire les modèles d'intégration de données.
- Expliquer le processus de fabrique de données.
- Comprendre les composants Azure Data Factory.
- Gérer la sécurité Azure Data Factory.
- Configurer Azure Data Factory.
- Créer des services liés.
- Créer les jeux de données.
- Créer des activités et pipelines de fabrique de données.
- Gérer un runtime d'intégration.

### **Effectuer une transformation à grande échelle sans code avec un pipeline Azure Data Factory ou Azure Synapse**

- Expliquer les méthodes de transformation Data Factory.
- Décrire les types de transformation Data Factory.
- Déboguer le flux de données de mappage.

### **Orchestrer le déplacement et la transformation des données dans le pipeline Azure Data Factory ou Azure Synapse**

- Comprendre le flux de contrôle d'une fabrique de données.
- Utiliser des pipelines de fabrique de données.
- Déboguer des pipelines de fabrique de données.
- Ajouter des paramètres aux composants de fabrique de données.
- Intégrer un notebook dans des pipelines Azure Synapse.
- Exécuter des packages de fabrique de données.

### **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.
- Analyser les données liées à l'aide de Synapse SQL.

### **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.

### **Configurer et gérer des secrets dans Azure Key Vault**

- Explorer l'utilisation correcte d'Azure Key Vault.
- Gérer l'accès à un coffre Azure Key Vault.

- Explorer la gestion des certificats avec Azure Key Vault.
- Configurer une solution de génération de clés de module de sécurité matériel.

#### Implémenter des contrôles de compatibilité pour les données sensibles

- Planifier et implémenter la classification des données dans Azure SQL Database.
- Comprendre et configurer la sécurité au niveau des lignes et le masquage dynamique des données.
- Comprendre l'utilisation de Microsoft Defender pour SQL.
- Découvrir comment fonctionne le Registre Azure SQL Database.

#### Activer une messagerie fiable pour des applications Big Data avec Azure Event Hubs

- Créer un hub d'événements à l'aide d'Azure CLI.
- Configurer des applications pour envoyer ou recevoir des messages par le biais d'Event Hub.
- Évaluer les performances d'Event Hub à partir du portail Azure.

#### 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 requis, 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