

# Data Engineering on Microsoft Azure (Microsoft DP-203)

## Cours officiel DP-203, préparation à l'examen

Cours Pratique de 4 jours - 28h

Réf : AZR - Prix 2024 : 2 630€ HT

Avec cette formation, vous découvrirez les modèles et les pratiques d'ingénierie des données dans le cadre de solutions analytiques en temps réel et par lots utilisant les technologies de la plateforme de données Azure (à savoir, les technologies de calcul et de stockage de base utilisées pour construire une solution analytique, concevoir des couches de service analytiques, se concentrer sur les considérations d'ingénierie des données pour travailler avec des fichiers sources, explorer de manière interactive les données stockées dans des fichiers du data lake, etc.).

### PARTICIPANTS

Professionnels des données, architectes de données et professionnels BI. Analystes de données et data scientists qui travaillent avec des solutions analytiques basées sur Microsoft Azure.

### PRÉREQUIS

Bonnes connaissances du cloud et des concepts de base de données. Avoir une expérience avec des solutions de données. Avoir suivi les cours officiels : AZ-900 (Réf. MKV) et DP-900 (Réf. MZP).

### COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils sont agréés par l'éditeur et sont certifiés sur le cours. Ils ont aussi été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum trois à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

### MODALITÉS D'ÉVALUATION

Évaluation des compétences visées en amont de la formation.

Évaluation par le participant, à l'issue de la formation, des compétences acquises durant la formation.

Validation par le formateur des acquis du participant en précisant les outils utilisés : QCM, mises en situation...

À l'issue de chaque stage, ITTCERT fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Les participants réalisent aussi une évaluation officielle de l'éditeur. Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le participant a bien assisté à la totalité de la session.

### MOYENS PÉDAGOGIQUES ET TECHNIQUES

Les ressources pédagogiques utilisées sont les supports et les travaux pratiques officiels de l'éditeur.

### MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

### ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante [psh-accueil@orsys.fr](mailto:psh-accueil@orsys.fr) pour étudier au mieux votre demande et sa faisabilité.

## OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Explorer les options de calcul et de stockage pour les charges de travail d'ingénierie des données dans Azure

Concevoir et mettre en œuvre la couche de service

Comprendre les considérations relatives à l'ingénierie des données

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

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

Effectuer l'exploration et la transformation des données dans Azure Databricks

Ingérer et charger des données dans l'entrepôt de données

Transformer les données avec Azure Data Factory ou Azure Synapse Pipelines

Optimiser les performances des requêtes avec des pools SQL dédiés dans Azure Synapse

Analyser et optimiser le stockage de l'entrepôt de données

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

Sécurité de bout en bout avec Azure Synapse Analytics

Effectuer un traitement de flux en temps réel avec Stream Analytics

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

Créer des rapports à l'aide de l'intégration de Power BI avec Azure Synapse Analytics

Effectuer des processus d'apprentissage automatique intégrés dans Azure Synapse Analytics

## MÉTHODES PÉDAGOGIQUES

Animation de la formation en français. Support de cours officiel au format numérique et en anglais. Bonne compréhension de l'anglais à l'écrit.

## CERTIFICATION

La réussite de l'examen permet d'obtenir la certification "Microsoft Certified: Azure Data Engineer Associate".

## LE PROGRAMME

---

dernière mise à jour : 01/2024

### 1) Explorer les options de calcul et de stockage pour les charges de travail analytiques

- Introduction à Azure Synapse Analytics.
- Décrire Azure Databricks.
- Introduction au stockage Azure Data Lake.
- Décrire l'architecture du data lake.
- Travailler avec des flux de données à l'aide d'Azure Stream Analytics.

*Travaux pratiques : Combiner le streaming et le traitement par lots avec un seul pipeline.*

*Organiser le data lake en niveaux de transformation de fichiers. Indexer le stockage du data lake pour l'accélération des requêtes et de la charge de travail.*

## 2) Concevoir et implémenter la couche de service

- Concevoir un schéma multidimensionnel pour optimiser les charges de travail analytiques.
- Savoir transformer sans code à grande échelle avec Azure Data Factory.
- Remplir les dimensions qui changent lentement dans les pipelines Azure Synapse Analytics.

*Travaux pratiques : Concevoir un schéma en étoile pour les charges de travail analytiques. Remplir les dimensions à évolution lente avec Azure Data Factory et mapper les flux de données.*

## 3) Considérations d'ingénierie des données pour les fichiers source

- Concevoir un entrepôt de données moderne à l'aide d'Azure Synapse Analytics.
- Sécuriser un entrepôt de données dans Azure Synapse Analytics.

*Travaux pratiques : Gestion des fichiers et sécurisation des fichiers stockés dans un data lake Azure.*

## 4) Exécuter des requêtes interactives à l'aide de pools SQL sans serveur ASA

- Explorer 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.
- Sécuriser les données et gérer les utilisateurs dans les pools SQL sans serveur Azure Synapse.

*Travaux pratiques : Exécuter des requêtes interactives à l'aide de pools SQL sans serveur Azure Synapse Analytics (ASA) : interroger 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, etc.*

## 5) Explorer, transformer et charger des données dans l'entrepôt de données

- 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.
- Transformer 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.

*Travaux pratiques : Explorer, transformer et charger des données dans le data warehouse à l'aide d'Apache Spark.*

## 6) Explorer et transformer 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.

*Travaux pratiques : 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 des colonnes DataFrame, etc.*

## 7) Ingérer et charger des données dans l'entrepôt de données

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

*Travaux pratiques : Réaliser 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.*

## 8) Transformer les données avec Azure Data Factory ou Azure Synapse Pipelines

- Intégrer des données avec Azure Data Factory ou Azure Synapse Pipelines.

- Transformer sans code à grande échelle avec Azure Data Factory ou Azure Synapse Pipelines.

*Travaux pratiques* : Exécuter 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.

#### 9) Utiliser Azure Synapse Pipelines

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

*Travaux pratiques* : Intégrer les données des notebooks avec Azure Data Factory ou Azure Synapse Pipelines.

#### 10) Optimiser les performances des requêtes avec des pools SQL

- Optimiser les performances des requêtes d'entrepôt de données dans Azure Synapse Analytics.

- Comprendre les fonctionnalités de développement d'entrepôt de données d'Azure Synapse Analytics.

*Travaux pratiques* : Comprendre les fonctionnalités de développement d'Azure Synapse Analytics. Optimiser les performances des requêtes d'entrepôt de données dans Azure Synapse Analytics. Améliorer les performances des requêtes.

#### 11) Analyser et optimiser le stockage de l'entrepôt de données

- Analyser et optimiser le stockage de l'entrepôt de données dans Azure Synapse Analytics.

*Travaux pratiques* : Vérifier les données faussées et l'utilisation de l'espace. Comprendre les détails du stockage du magasin de colonnes. Étudier l'impact des vues matérialisées. Explorer les règles pour les opérations minimalement enregistrées.

#### 12) Prendre en charge du traitement analytique transactionnel hybride (HTAP)

- 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.

*Travaux pratiques* : Prise en charge du traitement analytique transactionnel hybride (HTAP) avec Azure Synapse Link.

#### 13) Sécuriser de bout en bout avec Azure Synapse Analytics

- Sécuriser 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.

*Travaux pratiques* : Sécuriser l'infrastructure de support Azure Synapse Analytics. Sécuriser l'espace de travail et des services gérés Azure Synapse Analytics. Sécuriser les données de l'espace de travail Azure Synapse Analytics.

#### 14) Effectuer des traitements de flux en temps réel avec Stream Analytics

- Ingérer des flux de données avec Azure 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.

*Travaux pratiques* : Utiliser Stream Analytics pour traiter les données en temps réel des Event Hubs. Utiliser 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, etc.

### 15) Créer une solution de traitement de flux avec Event Hubs et Azure Databricks

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

*Travaux pratiques* : Créer une solution de traitement de flux avec Event Hubs et Azure

*Databricks* : explorer les principales fonctionnalités et utilisations du streaming structuré, diffuser des données à partir d'un fichier et les écrire dans un système de fichiers distribué, etc.

### 16) Créer des rapports à l'aide de Power BI avec Azure Synapse Analytics

- Créer des rapports avec Power BI en utilisant son intégration avec Azure Synapse Analytics.

*Travaux pratiques* : Intégrer un espace de travail Azure Synapse et Power BI. Optimiser l'intégration avec Power BI. Améliorer les performances des requêtes avec des vues

matérialisées et la mise en cache des résultats. Visualiser les données avec SQL sans serveur et créer un rapport Power BI.

### 17) Processus d'apprentissage automatique intégrés dans Azure Synapse Analytics

- Utiliser le processus d'apprentissage automatique intégré dans Azure Synapse Analytics.

*Travaux pratiques* : Exécuter des processus d'apprentissage automatique intégrés dans Azure

*Synapse Analytics* : créer un service lié Azure Machine Learning. Déclencher une expérience

Auto ML à l'aide des données d'une table Spark. Enrichir les données à l'aide de modèles entraînés. Diffuser les résultats avec Power BI.

## LES DATES

---

#### CLASSE À DISTANCE

2024 : 02 juil., 01 oct., 17 déc.

#### PARIS

2024 : 25 juin, 24 sept., 10 déc.

#### LYON

2024 : 17 déc.

#### AIX-EN-PROVENCE

2024 : 02 juil., 01 oct., 17 déc.

#### BORDEAUX

2024 : 02 juil., 01 oct., 17 déc.

#### LILLE

2024 : 02 juil., 01 oct., 17 déc.

#### NANTES

2024 : 02 juil., 01 oct., 17 déc.

#### SOPHIA-ANTIPOLIS

2024 : 02 juil., 01 oct., 17 déc.

#### STRASBOURG

2024 : 02 juil., 01 oct., 17 déc.

#### TOULOUSE

2024 : 02 juil., 01 oct., 17 déc.