

## Ingénierie de données sur Microsoft Azure

Durée: 4 Jours    Réf de cours: M-DP203    Méthodes d'apprentissage: Classe à distance

### Résumé:

**Cette formation est délivrée en synchrone à distance tout en garantissant l'accès à un environnement d'apprentissage complet!**

Cette formation vous permettra de découvrir 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 plate-forme de données Azure.

Vous découvrirez les technologies de calcul et de stockage de base qui sont utilisées pour construire une solution analytique, et explorerez ensuite comment concevoir des couches de service analytiques et se concentrer sur les considérations d'ingénierie des données pour travailler avec des fichiers sources.

Vous apprendrez à explorer de manière interactive les données stockées dans des fichiers dans un lac de données, les différentes techniques d'ingestion qui peuvent être utilisées pour charger des données à l'aide de la fonctionnalité Apache Spark présente dans Azure Synapse Analytics ou Azure Databricks, ou comment ingérer à l'aide d'Azure Data Factory ou des pipelines Azure Synapse.

Vous aborderez également les différentes façons de transformer les données à l'aide des mêmes technologies que celles utilisées pour l'acquisition des données et apprendront à surveiller et à analyser les performances des systèmes analytiques afin d'optimiser les performances des charges de données ou des requêtes émises sur les systèmes.

Enfin, vous comprendrez l'importance de la mise en œuvre de la sécurité pour garantir la protection des données au repos ou en transit.

#### Les plus :

\*\*L'environnement des labs d'apprentissage reste accessible après la formation durant 365 jours. Ceux-ci peuvent être pratiqués de façon illimitée.

\*\*Cette formation est éligible au Compte Personnel de Formation (CPF) sur [moncompteformation.gouv.fr](http://moncompteformation.gouv.fr)

Mise à jour : 26.11.2022

### Public visé:

Cette formation s'adresse aux professionnels des données, aux architectes de données et aux professionnels de la veille stratégique qui souhaitent se familiariser avec l'ingénierie des données et la création de solutions analytiques à l'aide des technologies de plateforme de données qui existent sur Microsoft Azure.

Le public secondaire de cette formation sont les Data Analyst et les Data Scientist qui travaillent avec des solutions analytiques construites sur Microsoft Azure.

### Objectifs pédagogiques:

- A l'issue de la formation, les participants seront capables de :
  - Transformer les données avec Azure Data Factory ou Azure Synapse Pipelines
- Explorer les options de calcul et de stockage pour les charges de travail d'ingénierie des données dans Azure
  - Intégrer les données des ordinateurs portables avec Azure Data Factory ou Azure Synapse Pipelines
- Concevoir et mettre en œuvre la couche de service
  - Optimiser les performances des requêtes avec des pools SQL dédiés dans Azure Synapse
- Exécuter des requêtes interactives à l'aide de pools SQL sans serveur
  - Prendre en charge le traitement analytique transactionnel hybride (HTAP) avec Azure Synapse Link
- Explorer, transformer et charger des données dans l'entrepôt de données à l'aide d'Apache Spark.
  - Effectuer un traitement de flux en temps réel avec Stream Analytics
- Effectuer l'exploration et la transformation des données dans Azure Databricks

### Pré-requis:

Les participants doivent avoir une connaissance du cloud computing et des concepts de données de base et une expérience professionnelle des solutions de données.

- M-DP900 - Microsoft Azure Data Fondamentaux

### Test et certification

Ce cours prépare à l'examen DP-203 (non inclus) qui permet d'atteindre le titre de **certification Microsoft Certified Azure Data Engineer Associate**(Certification éditeur).

\*\*\*

■ M-AZ900 - Microsoft Azure Fundamentals pour les professionnels  
IT

Cette formation mène également à la Certification professionnelle inscrite au Répertoire Spécifique de France Compétences RS5307 "Concevoir et mettre en œuvre la gestion des données dans Microsoft Azure". Plus de détails sur :  
<https://www.francecompetences.fr/recherche/rs/5307/>

---

## 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
- Description d'Azure Databricks
- Introduction au stockage Azure Data Lake
- Décrire l'architecture Delta Lake
- Travailler avec des flux de données à l'aide d'Azure Stream Analytics

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

- Combiner le traitement en continu 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 accélérer les requêtes et les charges de travail.

Module 2 : Conception et mise en œuvre de la couche de service

- Concevoir un schéma multidimensionnel pour optimiser les charges de travail analytiques.
- Transformation sans code à l'échelle avec Azure Data Factory
- Remplissage de dimensions à évolution lente dans les pipelines Azure Synapse Analytics

Lab 2 : conception et mise en œuvre de la couche de service

- Conception d'un schéma en étoile pour les charges de travail analytiques.
- Remplissage de dimensions à évolution lente avec Azure Data Factory et mappage des flux de données

Module 3 : Considérations d'ingénierie des données pour les fichiers sources

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

Lab 3 : considérations relatives à l'ingénierie des données

- Gestion des fichiers dans un lac de données Azure
- Sécurisation des fichiers stockés dans un lac de données Azure

Module 4 : Exécution de requêtes interactives à l'aide des pools SQL sans serveur Azure Synapse Analytics

Module 7 : Ingérer et charger des données dans l'entrepôt de données

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

Lab 7 : ingérer et charger des données dans l'entrepôt de données

- Effectuez une ingestion à l'échelle du pétaoctet avec Azure Synapse Pipelines.
- Importer des données avec PolyBase et COPY à l'aide de T-SQL
- Utiliser les meilleures pratiques de r données dans Azure Synapse Analytics

Module 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 l'échelle avec Azure Data Factory ou Azure Synapse Pipelines

Lab 8 : transformation de données avec Azure Data Factory ou Azure Synapse Pipelines

- Exécution de transformations sans code à l'échelle avec Azure Synapse Pipelines
- Créer un pipeline de données pour importer des fichiers CSV mal formatés
- Créer des flux de données mappés

Module 9 : Orchestrer le mouvement et la transformation des données dans Azure Synapse Pipelines

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

Lab 9 : 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 10 : Optimiser les performances des requêtes avec des pools SQL dédiés dans Azure Synapse

- Optimiser les performances des requêtes de l'entrepôt de données dans Azure Synapse Analytics.
- Comprendre les fonctions de

Module 13 : Sécurité de bout en bout avec Azure Synapse Analytics

- Sécurisation d'un entrepôt de données dans Azure Synapse Analytics
- Configuration et gestion des secrets dans Azure Key Vault
- Mise en œuvre de contrôles de conformité pour les données sensibles

Lab 13 : sécurité de bout en bout avec Azure Synapse Analytics

- Sécurisation de l'infrastructure de support d'Azure Synapse Analytics
- Sécurisation de l'espace de travail et des services gérés Azure Synapse Analytics
- Sécurisation des données de l'espace de travail Azure Synapse Analytics

Module 14 : Traitement des flux en temps réel avec Stream Analytics

- Activation d'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

Lab 14 : traitement de flux en temps réel avec Stream Analytics

- Utiliser Stream Analytics pour traiter les données en temps réel provenant des Event Hubs.
- Utiliser les fonctions de fenêtrage de Stream Analytics pour créer des agrégats et les transmettre à Synapse Analytics.
- Faites évoluer le job Azure Stream Analytics pour augmenter le débit grâce au partitionnement.
- Repartitionner l'entrée du flux pour optimiser la parallélisation.

Module 15 : Créer une solution de traitement de flux avec Event Hubs et Azure Databricks

- Traitement des données en continu avec le streaming structuré d'Azure Databricks

Lab 15 : Créer une solution de traitement de flux avec Event Hubs et Azure Databricks

- Explorer les principales caractéristiques et utilisations du streaming structuré.
- Streamer des données à partir d'un fichier et les écrire sur un système de fichiers distribué.
- Utiliser des fenêtres glissantes pour agréger des morceaux de données plutôt que toutes les données.

- Explorer les fonctionnalités des pools SQL sans serveur Azure Synapse
- Interroger les données dans le lac à l'aide des pools SQL sans serveur Azure Synapse
- Créer les 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

Lab 4 : Exécution de requêtes interactives à l'aide de pools SQL sans serveur

- Interroger des 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
- Sécuriser l'accès aux données dans un lac de données en utilisant des pools SQL sans serveur
- Configurer la sécurité du lac de données en utilisant le contrôle d'accès basé sur les rôles (RBAC) et la liste de contrôle d'accès.

Module 5 : Explorer, transformer et charger des données dans l'entrepôt de données avec Apache Spark

- Comprendre l'ingénierie du big data avec Apache Spark dans Azure Synapse Analytics
- Ingérer des données avec Apache Spark notebooks dans Azure Synapse Analytics
- Transformer les données avec DataFrames dans les pools Apache Spark dans Azure Synapse Analytics.
- Intégrer des pools SQL et Apache Spark dans Azure Synapse Analytics

Lab 5 : Explorer, transformer et charger des données dans l'entrepôt de données à l'aide d'Apache Spark

- Exploration des données dans Synapse Studio
- Ingérer des données avec les notebooks Spark dans Azure Synapse Analytics
- Transformer des données avec des DataFrames dans des pools Spark dans Azure Synapse Analytics.
- Intégrer des pools SQL et Spark dans Azure Synapse Analytics

Module 6 : Exploration et transformation des données dans Azure Databricks

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

développement d'entrepôt de données d'Azure Synapse Analytics

Lab 10 : Optimisation des performances des requêtes avec des pools SQL dédiés dans Azure Synapse

- Comprendre les fonctionnalités de développement d'Azure Synapse Analytics
- Optimiser les performances des requêtes de l'entrepôt de données dans Azure Synapse Analytics
- Améliorer les performances des requêtes

Module 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.

Laboratoire 11 : analyse et optimisation du stockage de l'entrepôt de données

- Vérifier si les données et l'utilisation de l'espace sont asymétriques
- Comprendre les détails du stockage en colonnes
- Étudier l'impact des vues matérialisées
- Explorer les règles pour les opérations à journalisation minimale

Module 12 : Prendre en charge le 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 des pools Apache Spark
- Interroger Azure Cosmos DB avec des pools SQL sans serveur

Lab 12 : prise en charge du traitement analytique transactionnel hybride (HTAP) avec Azure Synapse Link

- Configuration d'Azure Synapse Link avec Azure Cosmos DB
- Interroger Azure Cosmos DB avec Apache Spark pour Synapse Analytics
- Interroger Azure Cosmos DB avec un pool SQL sans serveur pour Azure Synapse Analytics

- Appliquer le filigrane pour supprimer les données périmées
- Connexion aux flux de lecture et d'écriture des Event Hubs

Module 16 : Créer des rapports en utilisant l'intégration de Power BI avec Azure Synapse Analytics

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

Lab 16 : Créer des rapports à l'aide de l'intégration de Power BI avec Azure Synapse Analytics

- Intégrer un espace de travail Azure Synapse et Power BI
- Optimiser l'intégration avec Power BI
- Améliorer les performances des requêtes avec les vues matérialisées et la mise en cache des ensembles de résultats
- Visualisation des données avec SQL serverless et création d'un rapport Power BI

Module 17 : Exécuter des processus d'apprentissage automatique intégrés dans Azure Synapse Analytics

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

Lab 17 : exécution de processus d'apprentissage automatique intégré dans Azure Synapse Analytics

- Créer un service lié à Azure Machine Learning
- Déclencher une expérience Auto ML à l'aide de données provenant d'une table Spark
- Enrichir les données à l'aide de modèles formés
- Afficher les résultats de prédiction à l'aide de Power BI

## DataFrames dans Azure Databricks

Lab 6 : exploration et transformation des données dans Azure Databricks

- Les DataFrames dans Azure Databricks pour explorer et filtrer les données.
- Mettre en cache un DataFrame pour accélérer les requêtes ultérieures
- Supprimer des données dupliquées
- Manipuler des valeurs de date et d'heure
- Supprimer et renommer des colonnes de DataFrame
- Agréger des données stockées dans un DataFrame

---

### Méthodes pédagogiques :

Suivi de cette formation à distance en synchrone depuis un site client équipé ou les locaux de Global Knowledge.

Une fiche explicative est adressée en amont aux participants pour leur permettre de vérifier leur installation technique et de se familiariser avec la solution technologique utilisée.

L'accès à l'environnement d'apprentissage (support de cours officiel, labs), ainsi qu'aux feuilles d'émargement et d'évaluation est assuré.

---

### Autres moyens pédagogiques et de suivi:

- Compétence du formateur : Les experts qui animent la formation sont des spécialistes des matières abordées et ont au minimum cinq ans d'expérience d'animation. Nos équipes ont validé à la fois leurs connaissances techniques (certifications le cas échéant) ainsi que leur compétence pédagogique.
- Suivi d'exécution : Une feuille d'émargement par demi-journée de présence est signée par tous les participants et le formateur.
- En fin de formation, le participant est invité à s'auto-évaluer sur l'atteinte des objectifs énoncés, et à répondre à un questionnaire de satisfaction qui sera ensuite étudié par nos équipes pédagogiques en vue de maintenir et d'améliorer la qualité de nos prestations.

Délais d'inscription :

- Vous pouvez vous inscrire sur l'une de nos sessions planifiées en inter-entreprises jusqu'à 5 jours ouvrés avant le début de la formation sous réserve de disponibilité de places et de labs le cas échéant.
- Votre place sera confirmée à la réception d'un devis ou ""booking form"" signé. Vous recevrez ensuite la convocation et les modalités d'accès en présentiel ou distanciel.
- Attention, si cette formation est éligible au Compte Personnel de Formation, vous devrez respecter un délai minimum et non négociable fixé à 11 jours ouvrés avant le début de la session pour vous inscrire via [moncompteformation.gouv.fr](http://moncompteformation.gouv.fr).

Accueil des bénéficiaires :

- En cas de handicap : plus d'info sur [globalknowledge.fr/handicap](http://globalknowledge.fr/handicap)
- Le Règlement intérieur est disponible sur [globalknowledge.fr/reglement](http://globalknowledge.fr/reglement)