

ETL et Data Warehouse

Microsoft Azure - Data Engineering

4 jours (28h00) | ★★★★★ 4,6/5 | MSDP203 | Code Certif Info : 109835 | Certification DP-203 (non incluse) | Évaluation qualitative de fin de stage | Formation délivrée en présentiel ou distanciel ⁽¹⁾

Formations Informatique > BI et Outils décisionnels > ETL et Data Warehouse



À l'issue de ce stage vous serez capable de :

- Expérimenter les options de calcul et de stockage pour les workloads de Data Engineering 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 le Data Warehouse
- Transformer les données avec Azure Data Factory ou Azure Synapse Pipelines
- Intégrer les données à partir des notebooks avec Azure Data Factory ou Azure Synapse Pipelines
- Prendre en charge le traitement analytique transactionnel hybride (HTAP) avec Azure Synapse Link
- Assurer la sécurité end-to-end 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.

Niveau requis

Avoir des connaissances sur le Cloud Computing, sur les concepts de base des données et une expérience professionnelle avec des solutions de données. Il est également recommandé d'avoir suivi les formations MSAZ900T00 "Microsoft Azure - Fondamentaux" et "Microsoft Azure Data Fundamentals" ou avoir les connaissances équivalentes.

Public concerné

Professionnels des données, architectes de données, professionnels en Business Intelligence, Data Analysts ou Scientists.

(1) Modalité et moyens pédagogique :

Formation délivrée en présentiel ou distanciel * (e-learning, classe virtuelle, présentiel à distance). Le formateur alterne entre méthodes ** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation). La validation des acquis peut se faire via des études de cas, des quiz et/ou une certification.

Les moyens pédagogiques mis en oeuvre (variables suivant les formations) sont : ordinateurs Mac ou PC (sauf pour les cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel). Environnements de formation installés sur les postes de travail ou en ligne. Supports de cours et exercices.

* Nous consulter pour la faisabilité en distanciel. ** Ratio variable selon le cours suivi.

Partenaire / Éditeur



Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence émarginée par demi-journée par les stagiaires et le formateur.

Programme

Explorer les options de calcul et de stockage pour les workloads de Data Engineering

- Introduction à Azure Synapse Analytics
- Décrire Azure Databricks
- Introduction au stockage Azure Data Lake
- Décrire l'architecture Delta Lake
- Travailler avec des flux de données en utilisant Azure Stream Analytics

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

- Explorer les capacités des pools SQL sans serveur Azure Synapse
- Interroger des données dans le Data Lake à l'aide des pools SQL sans serveur Azure Synapse
- Créer des objets de métadonnées dans les pools SQL sans serveur Azure Synapse
- Sécuriser des données et gérer des utilisateurs dans les pools SQL sans serveur Azure Synapse

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
 - Les méthodes avancées des DataFrames dans Azure Databricks

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

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

Ingérer et charger des données dans le Data Warehouse

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

Transformation des données avec Azure Data Factory ou Azure Synapse Pipelines

- Intégrer des données avec Azure Data Factory ou Azure Synapse Pipelines
- Transformer sans code et à l'échelle avec Azure Data Factory ou Azure Synapse Pipelines

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

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

Sécurité end-to-end avec Azure Synapse Analytics

- Sécuriser un Data Warehouse dans Azure Synapse Analytics
- Configurer et gérer les secrets dans Azure Key Vault
- Mettre en oeuvre des contrôles de conformité pour les données sensibles

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
- Configuration d'Azure Synapse Link avec Azure Cosmos DB
- Interroger Azure Cosmos DB avec des pools :
 - Apache Spark
 - SQL sans serveur

Traitement des flux en temps réel avec Stream Analytics

- Mettre en place 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

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

- Traiter le flux des données avec le flux structuré d'Azure Databricks

Certification (en option)

- Prévoir l'achat d'un voucher en supplément
- Le passage de l'examen se fera (ultérieurement) dans un centre agréé Pearson Vue
- L'examen (en anglais) s'effectuera en ligne

Modalités d'évaluation des acquis

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation ou une certification (M2i ou éditeur)

Les + de la formation

Le support de cours et les Microsoft Labs Online sont en anglais.

Compétences visées

- Concevoir et développer des solutions de stockage et de traitement de données pour l'entreprise
- Configurer et déployer des services de données Cloud tels que des services Blob, des bases de données et des fonctionnalités d'analytique
- Sécuriser la plateforme et les données stockées
- Vérifier que seuls les utilisateurs autorisés peuvent accéder aux données
- Assurer la continuité de l'activité dans les situations exceptionnelles à l'aide des techniques de haute disponibilité et de reprise d'activité après sinistre
- Assurer la supervision des systèmes pour vérifier qu'ils s'exécutent correctement et qu'ils sont économiques.