



Formation Azure Data Factory

DataTipsLearning.com

Tarif :
1600 € HT

Durée :
3 Jours
(21h)

Lieu :
PARIS/
Distancielle/
Votre Local

Contact :

contact@datatipslearning.com

+33 0751386021

Réf. DTPSL-AZDFC-2022

[S'inscrire](#)

Plus de détails :

Objectifs

Public / Prérequis

Programme détaillé

Méthode pédagogique

La plate-forme de données Azure fournit un certain nombre de services de données pour les bases de données, le stockage de données et l'analyse. L'utilisation de ces services permet de créer une solution d'analyse moderne, sécurisée et évolutive. Azure Data Factory est un service d'intégration de données pour des flux de travail complexes. Il s'agit d'un service cloud géré sans serveur, conçu pour ces projets hybrides complexes d'ETL, d'ELT et d'intégration de données, utilisé pour créer et planifier des pipelines qui peuvent ingérer et transformer des données à partir de magasins de données disparates.

Objectifs

- Présentation le Big Data et Azure Data Factory.
- Présentation les blocs de construction d'Azure Data Factory.
- Apprendre à créer des pipelines Azure Data Factory.
- Transformer les données avec des activités Azure Data Factory.
- Apprendre à déplacer des données avec data copy.
- Utiliser des activités de d'orchestration et flux de contrôle.
- Gérer un processus de déploiement avec Azure DevOps.
- Apprendre à monitorer et debugger des pipelines de données.
- Sécuriser les données et les pipelines.

Public

- Data Engineer
- Data Architectes
- Data Scientist
- Data Analyst

Prérequis

- Connaissances des bases d'Azure.

Programme détaillé

- I. Introduction Data Analytics
 1. Big Data Analytics avec Microsoft Azure
 2. Présentation Azure Data Factory
 3. Azure Data Factory v1 vs. Azure Data Factory v2
 4. Data Integration avec Azure Data Factory
- II. Création d'Azure Data Factory
 1. Présentation d'Azure data platform
 2. Création d'un pipeline ADF à l'aide de Copy Data
 3. Création d'un pipeline ADF à l'aide de PowerShell
 4. Création d'un pipeline ADF avec Python
 5. Utilisation des templates pour créer des pipelines ADF

III. Ingestion avec Azure Data Factory

1. Architecture de l'ingestion
2. Les blocs de construction d'Azure Data Factory
3. Linked services
4. Triggers
5. Datasets
6. Pipelines
7. Activités
8. External Compute Activities
9. Internal Activities
10. Output Constraints
11. Web Activities
12. Itération et activités conditionnelles
13. Parameter-Driven Pipelines
14. Azure Data Factory Processing Patterns

IV. Data Movement

1. Copy Activity
2. Connecteurs supportés
3. Les performances de l'activité de Copy

V. Data Transformation

1. Data Transformation
2. HDInsight Transformation
3. Azure Machine Learning
4. Azure Data Lake
5. ETL vs ELT
6. Azure Databricks
7. Procédure stockée
8. Activité personnalisée

VI. Orchestration et flux de contrôle

1. Les expressions
2. Les fonctions
3. Les activités
4. Utilisation des paramètres et des fonctions built-in
5. Utilisation des activités de métadonnées et de procédure stockée
6. Utilisation des activités ForEach et Filter
7. Des activités de branching et Chaînage des activités au sein d'un pipeline
8. Utilisation des activités Lookup, Web, et Execute Pipeline
9. Création d'un event-based triggers

VII. Configuration d'un Data Warehouse cloud

1. Chargement de données dans Azure Synapse avec Azure Data Factory
2. Suspendre/reprendre un pool Azure SQL à partir d'Azure Data

Factory

3. Chargement de données dans Azure Synapse avec bulk load
4. Utilisation SQL on-demand

VIII. Azure Data Lake

1. Configuration d'Azure Data Lake Storage Gen2
2. Connecter Azure Data Lake à Azure Data Factory
3. Création de pipelines Big Data avec Azure Data Lake et Azure Data Factory

IX. Big Data, HDInsight et Databricks

1. Configuration d'un cluster HDInsight
2. Traitement du Big Data avec Apache Spark
3. Traitement des données d'Azure Data Lake avec HDInsight et Hive

X. Intégration avec MS SSIS

1. Exécution de packages SSIS à partir d'ADF

XI. Migration de données avec Azure Data Factory

1. Copier des données de S3 vers Azure Blob
2. Copier des données de S3 vers ADLS
3. Copier des données de Google Cloud vers Azure Data Lake
4. Copier des données de Google BigQuery vers Azure Data Lake
5. Migrer des données de Google BigQuery vers Azure Synapse
6. Déplacer des données vers Dropbox

XII. Azure Services Intégration

1. Déclencher un traitement de données avec Logic Apps
2. Utilisation de l'activité web pour appeler un logic app
3. Utilisation des Azure Functions
4. Transformer et préparer vos données via Azure Databricks

XIII. Gestion des processus de déploiement avec Azure DevOps

1. Configuration d'Azure DevOps
2. Déployer des fonctionnalités dans la branche master
3. Créer un CI/CD d'ADF
4. Publication des modifications dans ADF
5. Création d'un pipeline Azure pour CD

XIV. Monitoring et dépannage des pipelines de données

1. Monitoring des exécutions de pipeline et des runtimes d'intégration
2. Investigation des pipelines en failure et exécution en mode débogage
3. Rerunning des activités
4. Configurer des alertes pour les exécutions Data Factory

XV. Sécurité

1. Sécurisation des Data Credentials
2. Data Encryption in Transit
3. Data Encryption at Rest
4. Scénario hybride
5. Stockage des informations d'identification dans Azure Key Vault
6. Sécurité avancée avec Managed Service Identity MSI

Méthode pédagogique

La formation se compose d'une partie théorique, et également une partie pratique représentant 60% de de la formation.

La partie pratique contient plusieurs exercices sous forme de notebook Databricks avec les corrections, avec aussi un projet à la fin de la formation comme simulation d'une prod.

Chaque jour, une évaluation rapide des connaissances est effectuée avant de commencer les nouvelles parties de la formation.

A la fin, une synthèse globale est délivré aux stagiaires, renforcé par un projet prod.

Finalement, une évaluation QCM est proposée.

Un support de cours sera remis à chaque stagiaire comprenant les slides, les exercices et les corrigés et un git du projet prod.

Une feuille de présence est fournie en fin de formation avec une certificat de complétion de formation pour chaque stagiaire.

Le formateur est un Data Engineer expert, qui intervient sur le sujet depuis plusieurs années en formation mais aussi en conseil.