



Formation

# Data modeling

[DataTipsLearning.com](https://datatipslearning.com)

**Tarif :**  
**1200 € HT**

**Durée :**  
**2 Jours**  
**(14h)**

**Lieu :**  
**PARIS/**  
**Distancielle/**  
**Votre Local**

## Contact :

[contact@datatipslearning.com](mailto:contact@datatipslearning.com)

+33 0751386021

Réf. DTPSL-DMO-2022

[S'inscrire](#)

## Plus de détails :

Objectifs

Public / Prérequis

Programme détaillé

Méthode pédagogique

Les bases de données jouent un rôle important dans presque toutes les applications que nous utilisons, le choix et la conception correcte est un élément essentiel de tout développement, car l'évolutivité et les performances dépendent d'une conception bien choisie.

Depuis les années 1980, presque toutes ces bases de données sont des bases de données relationnelles. Avec le big data, de différents modèles de données sont à notre disposition tel que le NoSQL, datalake et Lakehouse.

# Objectifs

- Présentation des données et des bases de données.
- Comprendre SQL.
- Différencier OLTP et OLAP.
- Comprendre les diagrammes de relation d'entité.
- Apprendre à normaliser les données.
- Implémentation d'une base de données Azure SQL.
- Conception d'une base de données NoSQL.
- Implémentation d'une base de données Azure Cosmos DB.
- Apprendre à utiliser la modélisation dimensionnelle.
- Comprendre le schéma en étoile et le schéma Snowflake.
- Présentation de la couche de stockage Data Lake.
- Comprendre Delta Lake et son architecture.
- Présentation de Lakehouse.
- Apprendre la modélisation Data Vault.
- Implémentation de ETL avec Azure Data Factory.

# Public

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

# Prérequis

- Pas de connaissances particulières.

# Programme détaillé

- I. Introduction aux bases de données
  1. Présentation des bases de données relationnelles
  2. Introduction au langage de requête structuré SQL
  3. La conception de la base de données
  4. Comprendre la théorie relationnelle
  5. Les clés

## 6. Types de charge de travail OLTP/OLAP

### II. Analyse d'entité

1. Le scope project & product
2. Comprendre les diagrammes de relation d'entité
3. Les entités
4. Les relations
5. Création d'un diagramme modèle entité-association (ERD)
6. Contexte d'un ERD

### III. Normalisation des données

1. Quand utiliser la normalisation comme stratégie de conception
2. Empêcher la redondance
3. Les étapes de normalisation
4. Une approche alternative à la normalisation des données
5. Intégration de résultats séparés
6. Diagramme entité-relation

### IV. Provisionnement et implémentation d'une base de données Azure SQL

1. Comprendre les types de données SQL Server
2. Quantifier un data model
3. Provisionner une base de données Azure SQL
4. Connexion à la base de données
5. Langage de définition des données
6. Insertion de données
7. Indexage

### V. Conception d'une base de données NoSQL

1. Comprendre le big data
2. Comprendre les clusters Big Data
3. Utilisation de Cosmos DB
4. Bases de données clé-valeur
5. Autres bases de données NoSQL

### VI. Provisionnement et implémentation d'une base de données Azure Cosmos DB

1. Provisionnement d'une base de données Cosmos DB
2. Création d'un conteneur
3. Charger des documents dans un conteneur
4. Paramètres du conteneur Cosmos DB
5. Importation de données à l'aide de l'outil de migration de données Azure Cosmos DB

### VII. Modélisation dimensionnelle

1. Contexte de la modélisation dimensionnelle
2. Comprendre la modélisation dimensionnelle
3. Étapes de la modélisation dimensionnelle

4. Conception des dimensions
  5. Conception des tables Fact
  6. Utilisation d'un data Warehouse Kimball vs data Marts
- VIII. Les avantages des data Warehouses par rapport aux data Lakes
1. Distinction entre les data Warehouses et data Lakes
  2. Comprendre les opportunités du cloud computing moderne
  3. Explorer les avantages de l'IA et du ML
- IX. Présentation de la couche de stockage Data Lake
1. Configuration du stockage cloud Big Data
  2. Organisation d'un data Lake
  3. Implémenter un data model dans votre Data Lake
- X. Comprendre Delta Lake
1. Comprendre comment Delta Lake rend possible le Lakehouse
  2. Présentation de Delta Lake
  3. L'architecture Data Lakehouse
  4. Créer une table Delta Lake
  5. Modification des données dans une table Delta Lake existante
  6. Utiliser time travel
  7. Effectuer des upserts
  8. Comprendre les niveaux d'isolement
  9. Comprendre le contrôle de la concurrence
  10. Nettoyer les ressources Azure
- XI. Provisionnement et implémentation d'un pool SQL Azure Synapse
1. Présentation de Synapse Analytics
  2. Créer un espace de travail Synapse Analytics
  3. Création d'un pool SQL dédié
  4. Implémentation de tables dans les pools Synapse SQL
  5. Comprendre la gestion de la charge de travail
  6. Utiliser PolyBase pour charger des données
  7. Connexion à et utilisation d'un pool SQL dédié
- XII. Modélisation Data Vault
1. Contexte de la modélisation de Data Vault
  2. Concevoir des tables Hub
  3. Concevoir des tableaux Link
  4. Concevoir des tables satellites
  5. Utilisation des clés de hachage
  6. Conception d'une structure Data Vault
  7. Conception de Business Vault
  8. Implémentation de Data Vault
- XIII. Conception et implémentation d'un Data Lake avec Azure Storage
1. Contexte des data Lakes
  2. Modélisation d'un data Lake

3. Utilisation de différents formats de fichiers
4. Choisir la bonne taille de fichier
5. Provisionner un compte de stockage Azure
6. Créer un système de fichiers de data Lake
7. Création de plusieurs comptes de stockage

#### XIV. Implémentation de ETL avec Azure Data Factory

1. Présentation de Azure Data Factory
2. Présentation des principaux composants de Azure Data Factory
3. Utilisation de l'activité copy
4. Implémentation d'un data flow
5. Exécuter du code SQL depuis Data Factory

## 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.