

Programmes de formations



Data Flow

Version 17/04/2026

1. Microsoft Fabric – Data Engineer

Public : Data Engineer, Data Analysts, architectes et experts techniques

Durée : 3 jours, adaptable en fonction du besoin

Pré-requis : Connaissance des bases de données relationnelles, en développement Python et SQL et principes d'ingénierie de données.

Méthodologie et Evaluation :

- Interventions sous formes d'exposés
- Mise en application de la théorie via des exercices et mises en situation
- Travaux d'ateliers sur cas réels des participants
- Questionnaire d'évaluation des connaissances
- Support de cours et exercices

Objectifs de la formation :

- Savoir configurer un environnement de travail adapté aux besoins d'un projet analytique
- Comprendre l'impact des paramètres (Spark, Domain, OneLake...) sur la performance
- Adapter les modèles de chargement aux besoins du projet
- Comprendre les différentes méthodes d'ingestion batch et savoir transformer et nettoyer les données
- Maîtriser les outils d'orchestration et automatiser selon les événements
- Comprendre les concepts de traitement de données en temps réel
- Assurer un suivi opérationnel des traitements
- Diagnostiquer rapidement les erreurs
- Intégrer les pratiques DevOps dans un environnement data
- Garantir la sécurité des données

Programme :

Jour 1

- Introduction : Présentation de la plateforme Fabric
 - Comprendre l'écosystème Microsoft Fabric et ses composants clés (OneLake, Power BI, Data Factory, etc.)
 - Identifier les cas d'usage typiques dans l'ingénierie des données
 - Situer Fabric dans une architecture de données moderne
- Configurer les paramètres d'un workspace Microsoft Fabric
 - Créer et configurer un workspace pour les besoins d'un projet data
- Concevoir et mettre en œuvre des modèles de chargement
 - Comprends les enjeux des chargements complets et incrémentaux
 - Préparer les données pour un modèle dimensionnel
 - Définir un modèle de chargement pour le streaming
 - Choisir entre lakehouse, warehouse et eventhouse

Jour 2

- Ingérer et transformer des données batch
 - Choisir le type de stockage
 - Utiliser dataflows, notebooks, KQL et T-SQL
 - Gérer des raccourcis, duplications miroir
 - Ingestion avec pipelines
 - Transformation avec PySpark, SQL
 - Dénormalisation, agrégation
 - Gérer les doublons, données manquantes ou décalées

- Orchestrer les processus
 - Choisir entre pipeline et notebook
 - Créer des planifications et déclencheurs événementiels
 - Intégrer des paramètres et expressions dynamiques dans les notebooks/pipelines

- Présentation de l'ingestion en temps réel
 - Utiliser stockage natif ou raccourcis
 - Traitement avec Eventstreams, Spark structured streaming, KQL

● Jour 3

- Monitorer les ressources Fabric
 - Suivre l'ingestion, la transformation, l'actualisation du modèle
 - Configurer les alertes

- Identifier et résoudre les erreurs
 - Identifier les erreurs dans pipelines, flux, notebooks, Eventhouses, Eventstreams, T-SQL

- Mettre en œuvre la gestion du cycle de vie dans Fabric
 - Présentation de l'intégration avec Git
 - Mettre en œuvre des projets de base de données
 - Créer et configurer des pipelines de déploiement

- Configurer la sécurité et la gouvernance
 - Mettre en œuvre les contrôles d'accès au niveau du workspace
 - Configurer les contrôles d'accès par ressource, ligne, colonne, objet
 - Mettre en place le masquage dynamique
 - Activer et exploiter le logging du workspace

2. Microsoft Azure Databricks

Public : Data Engineer, Data Analysts et experts techniques

Durée : 2 jours, adaptable en fonction du besoin

Pré-requis : Avoir des bases sur Azure, la modélisation en étoile, en développement Python et SQL.

Méthodologie et Evaluation :

- Interventions sous formes d'exposés
- Mise en application de la théorie via des exercices et mises en situation
- Travaux d'ateliers sur cas réels des participants
- Questionnaire d'évaluation des connaissances
- Support de cours et exercices

Objectifs de la formation :

- Comprendre et utiliser les différentes fonctionnalités de Databricks
- Créer et manipuler les différents types de ressources
- Principe et utilisation du Unity Catalog
- Créer et manipuler des notebooks
- La gestion du cycle de vie des tables avec Delta Lake : utiliser les commandes
- Orchestrer des traitements : les workflows
- Produire des analyses à partir de Databricks SQL

Programme :

Jour 1

- INTRODUCTION
 - Comprendre les enjeux et spécificités des plateformes cloud modernes
- Notions sur Azure : les grands principes du cloud
 - Présentation et intérêt d'une plateforme cloud
- Databricks : vue d'ensemble
 - Présentation d'Azure Databricks : les concepts
 - Le Data Lakehouse : ETL et Datawarehouse
 - Gouvernance des données : Concept de Unity Catalog
 - Avantages de Databricks
- Les espaces de travail
 - Les ressources Azure Databricks
 - Créer un espace de travail
 - Connecter l'espace de travail à des sources de données externes
- Les différents types de ressources : clusters & SQL Warehouse
 - Présentation

- Les différents types de Calculs
- Création d'un cluster
- Gestion du cluster
- Le Unity Catalog
 - Créer un catalogue
 - Se connecter à des données à partir du Unity Catalog
 - Ajouter une table dans le schéma du catalogue Définir des autorisations sur une table du catalogue
 - Focus : se connecter à des bases de données SQL
 - Autre méthode pour ajouter des données au catalogue

Jour 2

- Utilisation des bases de données
 - Présentation des principaux connecteurs
 - Opérations sur les tables
 - Partage des connexions et gestion des transactions
- Transformer les données
 - Présentation du tMap
 - Jointure, filtrage et mutliplexage
 - Gestion des rejets
- Les Notebooks
 - Objectifs et intérêts des Notebooks
 - Le principe des cellules
 - Les commandes
 - Les Magic Commands
 - Utiliser les cellules et visualiser un résultat
 - Export et partage des Notebooks
- Principe de Spark & Delta
 - Présentation et prise en main du Spark SQL
 - Concepts de Delta Lake
 - Gestion du cycle de vie des tables :
 - Merge delta
 - Delete/insert
- Orchestration des traitements
 - Appel de notebooks depuis Azure data factory et configuration
 - Utilisation des workflows databricks
 - Paramétrage et configuration des notebooks databricks depuis un workflow
- Databricks SQL
 - Databricks SQL : Administration
 - L'Editeur SQL
 - Créer et exécuter une requête
 - Partager le résultat d'une requête

- Options disponibles sur les résultats d'une requête
 - Exporter
 - Créer des visualisations
 - Utiliser des filtres d'affichages
 - Paramétrer le résultat d'une requête

3. Data Build tool

Public : Data Scientists, Data Analysts, Data Engineers

Durée : 2 jours, adaptable en fonction du besoin

Pré-requis : Avoir des connaissances SQL, connaître les principes relationnels

Méthodologie et Evaluation :

- Interventions sous formes d'exposés.
- Mise en application de la théorie via des exercices et mises en situation
- Travaux d'ateliers sur cas réels des participants
- Questionnaire d'évaluation des connaissances
- Support de cours et exercices

Objectifs de la formation :

- Comprendre Dbt et son rôle dans le Modern Data Stack
- Démarrer un projet Dbt from scratch
- Maîtriser Dbt de bout en bout
- Industrialiser et mettre en production un projet Dbt

Programme :

Jour 1

- Introduction
 - Introduction à Dbt & définition d'un ELT
 - Rôle de dbt dans nos projets data et comparaison avec les outils classiques
 - Comprendre l'architecture « médaillon »
 - Organiser et structurer un projet Dbt
 - Installation de l'outil en local
- Structure d'un projet
 - Comprendre les modèles et les sources dans Dbt
 - Explorer les types de matérialisation (vue, table, modèle éphémère, modèle incrémental, etc.)
 - Comprendre les macros et leur utilité
 - Explorer les hooks et leurs cas d'usage
 - Présentation du lineage de données générées automatiquement par Dbt

Jour 2

- Industrialisation de Dbt : Mise en production d'un projet Dbt avec Git et Prefect
 - Présentation de Git, rappel des fondamentaux du versionning
 - Présentation des orchestrateurs et de leur positionnement dans l'écosystème data
 - Architecture de Prefect : services clés et opérateurs

- Pour aller plus loin avec Dbt
 - Découvrir les différents types de tests disponibles
 - Utiliser la documentation intégrée et visualiser le lignage des modèles
 - Bonnes pratiques de Dbt
 - Debugging, performance & optimisation Dbt
 - Dbt dans cloud (vs Dbt Core)

4. Oracle Data Integrator (ODI) 12c – niveau 1

Public : Data engineer et experts techniques

Durée : 3 jours, adaptable en fonction du besoin

Pré-requis : Connaissance du langage SQL

Méthodologie et Evaluation :

- Interventions sous formes d'exposés.
- Mises en situation avec mises en pratique sur une machine virtuelle via des supports de cours et des exercices
- Etude de cas
- Questions/Réponses

Objectifs de la formation :

- Comprendre l'architecture générale d'ODI 12c.
- Appréhender le studio de développement ODI 12c.
- Développer des mappings (flux d'alimentation).
- Utiliser des fonctions, variables et traitements.
- Développer avec XML et JSON

Programme :

Jour 1

- Introduction à ODI 12c
 - Présentation générale d'ODI12c et de son architecture technique.
 - Présentation environnement et revue des différents objets ODI 12c
- Développer des mappings avec ODI 12c
 - Développement de mapping avec utilisation des différents composants proposés
 - Composants Aggrégats, Filtre, tri, dataset, distinct, fractionnement, jointure, pivot, set, filtres sous-requêtes, Look up, mapping réutilisables
 - Mise en pratique

Jour 2

- Utiliser l'environnement d'ODI 12c et les outils
 - Utilisation du SCD
 - Mise en pratique
 - Exploitation du multicibles et de la parallélisation
 - Mise en pratique
 - Revue des différents outils utilisables dans les packages
- Utiliser des fonctions, variables et traitements

- Théorie et utilisations avec ODI
- Etude de cas

Jour 3

- Développer avec XML et JSON
 - Utiliser XML et mise en pratique
 - Utiliser JSON et mise en pratique
- Questions/Réponses

5. xDI Semarchy

Public :

- Data Engineer et expert technique,

Durée : 2 jours

Pré-requis : Avoir des bases en SQL et en modélisation de données.

Méthodologie et Evaluation :

- Quiz
- Cas pratiques
- Transfert d'outils

Objectifs de la formation :

- Concevoir et développer des mappings
- Développer des variables et des sous-variables
- Comprendre et connaître les process

Programme :

Jour 1

Introduction

- Présentation générale de Semarchy xDI 2025.1.
- Présentation du Designer et des différents objets.

Développer avec xDI : Les mappings

- Développement de mapping avec utilisation des différents composants présentés
- Filtre, jointure, tri, agrégats, template, distinct
- Mise en pratique

Jour 2

Présentation (suite)

- Complément mapping : Multi-cible, parallélisation
- Les process
- Les fonctions utilisateurs, variables, sous-process
- Fichier structure variable du XML, JSON

Développer avec xDI : Les process

- Développement de process

- Mise en pratique
- Présentation (fin)
- Semarchy Analytics

6. Talend

Public : Data Engineer, Data Analysts et experts techniques

Durée : 3 jours, adaptable en fonction du besoin

Pré-requis : Avoir de bonnes connaissances en modélisation relationnelle en SQL

Méthodologie et Evaluation :

- Interventions sous formes d'exposés
- Mise en application de la théorie via des exercices et mises en situation
- Travaux d'ateliers sur cas réels des participants
- Mini-projet le 3^{ème} jour
- Questionnaire d'évaluation des connaissances
- Support de cours et exercices

Objectifs de la formation :

- Concevoir et développer des jobs dans l'application ETL Talend Open Studio
- Optimiser les jobs développés par l'utilisation de contextes et jeux de données
- Réaliser des transformations plus complexes en utilisant variables, expressions et jointures
- Exécuter et déboguer un job, tracer les statistiques d'exécution

Programme :

Jour 1

- Présentation de Talend Open Studio
 - Présentation des ETL
 - Pré-requis et installation
 - Philosophie du produit
 - Prise en main de l'interface
- Job Designer
 - Business Modeler et Job Designer
 - Composants de base
 - Génération de code et exécution d'un job
- Utiliser des composants entrée/sortie (métadonnées)
 - Configuration des métadonnées, mise à jour
 - Utilisation des contextes

Jour 2

- Utilisation des bases de données
 - Présentation des principaux connecteurs
 - Opérations sur les tables
 - Partage des connexions et gestion des transactions

- Transformer les données
 - Présentation du tMap
 - Jointure, filtrage et mutliplexage
 - Gestion des rejets

Jour 3

- MINI PROJET : Mise en place d'un DWH
 - Chargement des tables de dimensions à partir des sources
 - Construction d'une dimension dates à partir de zéro
 - Réconciliation de données de faits avec les dimensions et chargement dans l'entrepôt
 - Ordonnancer l'ensemble des traitements

- Récapitulatif : Les bonnes pratiques
 - Types de flux
 - Lisibilité des jobs
 - Documenter

- Compléments
 - Générer et paramétrer des logs
 - Gestion de la volumétrie
 - Exporter des jobs
 - Orchestration

7. Programmation SQL

Public : développeur, administrateur, expert technique

Durée : 3 jours, adaptable en fonction du besoin

Pré-requis : Connaissances de base en programmation

Méthodologie et Evaluation :

- Interventions sous formes d'exposés
- Mise en application de la théorie via des exercices et mises en situation
- Travaux d'ateliers sur cas réels des participants
- Questionnaire d'évaluation des connaissances
- Support de cours et exercices

Objectifs de la formation :

- Comprendre le modèle relationnel et identifier les objets d'une base
- Interroger et classer les données d'une table à l'aide d'opérateurs simples
- Restituer les données de plusieurs tables grâce aux jointures et aux sous-requêtes
- Regrouper les données et réaliser des calculs à des fins d'analyse
- Manipuler les données

Programme :

Jour 1

- Algèbre relationnel et norme SQL
 - Les principes du modèle relationnel.
 - Historique des SGBDR et du langage SQL.
 - Entités, associations et MCD.
 - Types de données et expression des valeurs
- Théorie des bases de données
 - Eléments du SGBD.
 - Tables, clefs et formes normales.
 - Entités, associations et MCD.
 - Contraintes de colonnes et de tables.
- Interrogation des données monotabulaire
 - Structure d'une requête
 - Filtre WHERE, prédicats et logique ternaire (NULL).
 - Opérateurs et expressions SQL.
 - Fonctions prédéfinies.
- Interrogation multitabulaire

- Combiner des résultats avec les opérations ensemblistes (UNION, INTERSECT, EXCEPT).
- Présentation de la jointure.
- Imbrication des requêtes.
- Utilisation des CTE et découverte des requêtes récursives.
- Les sous-requêtes corrélées et l'opérateur EXISTS.

Jour 2

- Analyse de données
 - Fonctions d'agrégation (AVG, SUM, COUNT...).
 - Création de sous-ensembles (clause GROUP BY).
 - Filtrage des agrégats avec la clause HAVING.
 - Découverte des fonctions d'analyse de données (LEAD, LAG...) et rangements (RANK, ROW_NUMBER, NTILE...).
- Mise à jour des données
 - Les ordres INSERT, UPDATE, DELETE, TRUNCATE.
 - Mises à jour simples.
 - Mises à jour à l'aide de sous-requêtes.
 - Mise à jour des données à travers les vues.
- Introduction au SQL Procédural
 - Les dialectes SQL des principaux éditeurs.
 - Notions basiques sur les déclencheurs.
 - Introduction aux procédures stockées.
 - Les UDF ou fonctions "utilisateur".

Jour 3

- Historique et contexte de l'explosion des usages autour des données
 - Organisation, rôles et gouvernance : les nouveaux modèles à l'heure du Big Data.
 - Pourquoi les silos organisationnels sont-ils un frein majeur à l'exploitation des données et comment y remédier ?
 - La Data Science : adapter les profils de statisticiens aux nouveaux enjeux.
 - Les spécialistes des nouveaux domaines et leurs méthodes de travail.
 - Les thématiques de mise en œuvre de la gouvernance des données.
 - Les acteurs de la gouvernance de la donnée.
 - Les acteurs en charge de la qualité de la donnée.
 - Quelles sont les grandes tendances règlementaires autour de la gestion des données et comment les aborder ?
- Le Big Data étendu
 - La place de l'entreprise dans son écosystème de données.
 - Les données à la frontière de l'entreprise (DMP).

- Les données fournies par des partenaires : réseaux sociaux...
- L'Internet des objets (IoT) ou comment rendre son entreprise sensible aux nouveaux usages.
- L'Open Data et ses apports dans l'analyse de données : nouvelles données, enrichissement, croisement de données.
- L'émergence des chatbots ou l'automatisation de fonctions d'interface clients : quels impacts dans nos organisations ?
- L'expérience client augmentée, levier de la transformation digitale des entreprises.
- La fourniture de données comme nouvelle source de revenus.
- Les données, carburant de l'innovation métier.
- La gestion du patrimoine informationnel et le pilotage de la valeur
 - Quels sont les risques à ne pas prendre en compte dans le cas de la gestion du patrimoine informationnel ?
 - Les grandes tendances autour de la Data pour les trois années à venir.
 - Les solutions technologiques : Big Data et Machine Learning.
 - Mode d'emploi pour lancer une initiative Big Data.
 - Les techniques du Big Data : tour d'horizon.
 - Les métiers, fonctions et rôles autour de la donnée : la synthèse et les différents niveaux de mise en œuvre.
 - Pourquoi les entreprises se lancent-elles dans le Big Data patrimoine géré par l'entreprise.
 - L'apport de valeur par les données touche tous les métiers de nos clients.