

# Formation Microsoft Fabric Avancée pour Data Professionals

**Durée de la formation : 4 jour(s) / 28 heure(s)**

## OBJECTIFS

- Ingérer et retraiter efficacement des données issues de diverses sources, notamment des fichiers plats et CSV, en utilisant les outils avancés de Microsoft Fabric.
- Maîtriser Python (PySpark et Pandas) pour la transformation, le nettoyage et la consolidation des données au sein de l'environnement Fabric.
- Appliquer les meilleures pratiques en matière de conventions de nommage pour assurer la maintenabilité et la lisibilité des actifs de données.
- Mettre en œuvre des stratégies de déploiement continu (CI/CD) pour les solutions Fabric, en se concentrant sur la migration d'Azure DevOps vers GitHub Actions.
- Optimiser les performances des processus de traitement des données dans Fabric afin d'améliorer la rapidité d'exécution.

## PREREQUIS

Cette formation s'adresse aux professionnels qui :

- Disposent d'une expérience pratique et quotidienne avec Microsoft Fabric et ses composants de base (Lakehouses, Notebooks, Pipelines de données).
- Possèdent une solide expertise dans l'ingestion, la transformation et la modélisation de données.
- Ont des connaissances intermédiaires à avancées en Python, notamment avec les bibliothèques Pandas et une compréhension des concepts de Spark.
- Sont familiers avec les principes du contrôle de version (Git) et des concepts de base du déploiement continu.

## PUBLIC

Cette formation est conçue pour les professionnels des données, les ingénieurs de données et les architectes de solutions qui utilisent Microsoft Fabric au quotidien et souhaitent approfondir leurs compétences en :

- Traitement et consolidation de données complexes.
- Utilisation avancée de Python (PySpark et Pandas) dans Fabric.
- Mise en place de pratiques DevOps pour les projets Fabric.
- Optimisation des performances des workflows de données.

# Formation Microsoft Fabric Avancée pour Data Professionals

Durée de la formation : 4 jour(s) / 28 heure(s)

## PROGRAMME

### Jour 1 : Ingestion et Retraitement Avancé des Données avec Fichiers Plats et CSV

- **Rappels et approfondissements sur l'ingestion de données dans Fabric**
  - Revue des pipelines de données et Dataflows Gen2 : cas d'usage avancés.
  - Gestion des connexions et des sources de données complexes.
- **Ingestion de fichiers plats et CSV dans Lakehouse**
  - Bonnes pratiques d'ingestion pour les données non structurées et semi-structurées.
  - Gestion des schémas dynamiques et de l'inférence de type.
  - Techniques avancées d'ingestion : chargement incrémental, gestion des erreurs.
  - **Exercice pratique** : Ingestion de jeux de données CSV volumineux et hétérogènes.
- **Retraitement et préparation des données avec Dataflows Gen2**
  - Transformations avancées dans Power Query Online.
  - Gestion des données manquantes, doublons, et incohérences.
  - Mise en forme des données pour l'analyse.
  - **Exercice pratique** : Nettoyage et transformation d'un jeu de données CSV complexe en utilisant Dataflows Gen2.

### Jour 2 : Transformation et Consolidation avec Python (Pandas & PySpark) dans les Notebooks

Matin :

- **Introduction avancée aux Notebooks Fabric et environnements Spark**
  - Configuration et gestion des sessions Spark dans Fabric.
  - Utilisation de bibliothèques Python tierces.
- **Manipulation et nettoyage de données avec Pandas**
  - Techniques avancées de nettoyage de données (gestion des chaînes de caractères, dates, expressions régulières).
  - Agrégation et pivotement de données avec Pandas.
  - Jointures et fusions complexes de DataFrames Pandas.
  - **Exercice pratique** : Retraitement d'un jeu de données en utilisant Pandas dans un Notebook Fabric.
- **Transformation et consolidation de données avec PySpark**
  - Exploration des DataFrames PySpark : opérations de base et avancées.
  - Partitionnement et optimisation des données pour Spark.
  - Techniques de jointure et de fusion distribuées avec PySpark.
  - Utilisation de Spark SQL pour des transformations complexes.
  - **Exercice pratique** : Consolidation de plusieurs sources de données (CSV, Parquet) en utilisant PySpark.
- **Conventions de nommage et organisation des actifs Fabric**
  - Définition de conventions de nommage claires pour Lakehouses, Notebooks, Pipelines, etc.”;
  - Structuration des espaces de travail et des dossiers pour la maintenabilité.
  - Gestion des versions des Notebooks et scripts.
  - **Exercice pratique** : Refactorisation d'un projet Fabric existant en appliquant les conventions de nommage.
- **Optimisation des performances des processus de données**
  - Optimisation des requêtes Spark et PySpark (caching, repartitionnement, broadcast joins).
  - Meilleures pratiques pour l'ingestion et la transformation de données volumineuses.
  - Surveillance des performances des pipelines et notebooks.
  - **Exercice pratique** : Identification et résolution des goulots d'étranglement dans un pipeline de données.

### Jour 3 : Bonnes Pratiques, Optimisation et Sécurité

- **Comprendre les Enjeux de la Sauvegarde dans Spark/Fabric**
  - Vue d'ensemble : Pourquoi les sauvegardes sont-elles souvent un goulot d'étranglement ?
  - Rappel sur Delta Lake et Parquet : Avantages de ces formats.
- **Stratégies d'Optimisation des Écritures (Sauvegardes)**
  - Gestion de la taille des fichiers : Problème des "petits fichiers" et des fichiers "trop grands". Comment viser une taille de

# Formation Microsoft Fabric Avancée pour Data Professionals

**Durée de la formation : 4 jour(s) / 28 heure(s)**

fichier optimale.

- Contrôle du nombre de partitions d'écriture : Utilisation de `.coalesce()` et `.repartition()` pour ajuster le nombre de fichiers de sortie.
- Partitionnement physique (`.partitionBy()`) : Importance pour les performances de lecture ultérieures.
- Compaction des fichiers (`OPTIMIZE`) et Z-Ordering (pour Delta Lake) : Fusion et ordonnancement pour les lectures.
- Exercice Pratique : Optimisation des Sauvegardes sur un Volume Important de Données
- **Enrichissement des Données : Concepts et Applications**
- `merge` (Jointures) : Explication et types de jointures (`inner`, `left`, etc.) pour combiner horizontalement des DataFrames.
- `cut` (Segmentation / Binning) : Regrouper des valeurs numériques continues en catégories discrètes.
- `pivot` / `unpivot` (Pivot / Dé-pivot) : Remodeler les données (lignes en colonnes, colonnes en lignes).
- Concepts clés de manipulation de base : Renommer, ajouter/supprimer, filtrer, sélectionner/réordonner des colonnes, gérer les valeurs nulles, opérations sur chaînes et dates.
- **Exercice pratique** : Enrichissement et Segmentation de Données Clients
- **Exercice pratique** : Pivotelement/Dé-pivotelement et Nettoyage Avancé

## Jour 4 : Sécurité et CI/CD Simplifié

- **Gestion des Rôles et des Autorisations dans Fabric**
- Rôles d'espace de travail (`Admin`, `Member`, `Contributor`, `Viewer`) et leurs niveaux d'accès.
- Autorisations au niveau des éléments (Items) : Permissions granulaires sur les Lakehouses, Notebooks, Dataflows.
- Modèle de permission et bonnes pratiques (moindre privilège, groupes de sécurité Azure AD).
- **Exercice Pratique** : Configuration des Permissions et Rôles (Simulé/Explication)
- **Sécurité au Niveau des Lignes (RLS) et des Colonnes (CLS)**
- RLS (Row-Level Security) : Explication, cas d'usage (filtrage par région), implémentation via le SQL Analytics Endpoint.
- CLS (Column-Level Security) : Explication, cas d'usage (masquage d'informations sensibles), implémentation via le masquage dynamique des données en SQL.
- **Exercice pratique** : Implémentation de RLS et CLS (Principalement par le SQL Endpoint)
- **Audit et Conformité**
- Journaux d'audit Fabric : Importance pour le suivi des activités et la conformité.
- Accès aux journaux : Où les trouver et comment les utiliser.
- Conformité réglementaire : Comment Fabric aide à répondre aux exigences (GDPR, HIPAA)
- **Introduction au Déploiement Continu (CI/CD)**
- Concept du CI/CD : Rappel des principes (intégration/livraison/déploiement continu) et leur importance.
- Intégration Git dans Fabric : Comment Fabric s'intègre avec Azure DevOps Repos ou GitHub.
- Vue d'ensemble d'un pipeline CI/CD pour Fabric : CI (validation du code, tests) et CD (déploiement automatisé entre environnements).
- Outils : Discussion sur Azure DevOps Pipelines ou GitHub Actions.
- Transition Azure DevOps vers GitHub Actions (Focus sur l'explication) : Similitudes, différences, et démonstration conceptuelle (sans exécution pratique)

# Formation Microsoft Fabric Avancée pour Data Professionals

## Durée de la formation : 4 jour(s) / 28 heure(s)

### Prise en compte du handicap

Pour les personnes en situation de handicap, afin de nous permettre d'organiser le déroulement de la formation dans les meilleures conditions possibles, contactez-nous via notre formulaire de contact ou par mail (formation@access-it.fr) ou par téléphone (0320619506). Un entretien avec notre référente handicap pourra être programmé afin d'identifier les besoins et aménagements nécessaires.

### Modalités et moyens Pédagogiques, techniques et d'encadrement mis en œuvre

Répartition théorie/pratique : 45%/55%. Cette formation se compose d'une alternance d'apports théoriques, de travaux pratiques s'articulant autour d'une application fil rouge, de démonstrations, de phases d'échanges entre participants et de synthèses de la part du formateur.

Formation accessible à distance de n'importe où et n'importe quand, via un ordinateur type PC disposant d'une connexion à Internet à haut débit (ADSL ou plus).

Pour assurer un démarrage dans les meilleures conditions au premier jour de la formation, notre service logistique se met systématiquement en relation, en amont, avec vous afin de réaliser un test de validation technique et de vous présenter l'environnement de formation.

Pendant toute la durée de la formation, le stagiaire dispose d'une assistance technique et pédagogique illimitée, par e-mail, avec un délai de prise en compte et de traitement qui n'excède pas 24h. En complément, le stagiaire peut planifier un rendez-vous pédagogique avec un formateur expert afin d'échanger sur des éléments de la formation.

La durée de la formation affichée sur cette page est une durée estimée qui peut varier en fonction du profil du stagiaire et de ses objectifs (notamment s'il souhaite valider sa formation par le passage d'un examen de certification).

Durant la formation, le formateur prévoit :

Des démonstrations organisées en modules et en séquences découpées le plus finement possible, en suivant le programme pédagogique détaillé sur cette page ;

Des énoncés et corrigés de travaux pratiques à réaliser tout au long de la formation ;

Des travaux pratiques sont proposés ; la plateforme prévoit l'environnement technique nécessaire à la réalisation de l'ensemble des travaux pratiques ;

Le formateur valide les connaissances acquises après chaque TP ;

Il est proposé un ou plusieurs livres numériques faisant office d'ouvrage(s) de référence sur le thème de la formation.

### Validation et sanction de la formation

Une attestation mentionnant les objectifs, la nature et la durée de l'action et les résultats de l'évaluation des acquis de la formation sera remise au stagiaire à l'issue de sa formation par courrier électronique.

A la demande, il sera délivré un certificat de réalisation.

### Type de formation

Professionalisante ayant pour objectif le perfectionnement, l'élargissement des compétences

### Moyens permettant de suivre l'exécution de l'action

Le contrôle de la présence des stagiaires sera assuré par la vérification de l'assiduité des participants. Le stagiaire signera une feuille de présence par demi-journée de formation. Celle-ci sera également signée par le formateur.

### Modalité d'évaluation des acquis

Durant la formation, le participant est amené à mettre en pratique les éléments du cours par la réalisation de travaux pratiques réalisés sur PC.

La validation des acquis du stagiaire est faite par le formateur à la fin de chaque atelier. Cette validation individuelle est possible du fait du faible nombre de participants par session de formation (6 personnes maximum).

A la fin de la formation, le stagiaire a donc atteint les objectifs fixés par la formation.

En complément, pour les stagiaires qui le souhaitent, certaines formations peuvent être validées officiellement par l'éditeur en passant un examen de certification.

Access it étant centre d'examen ENI, les examens peuvent être réalisés sur demande à distance ou dans nos locaux de Villeneuve d'Ascq.

Les candidats à la certification doivent produire un travail personnel important en vue de se présenter au passage de l'examen, le seul suivi de la formation ne constitue pas un élément suffisant pour garantir un bon résultat et/ou l'obtention de la certification

### Assistance Post-Formation

Toute personne ayant suivi une formation avec Access it bénéficie d'une assistance post-formation d'une durée de 1 mois. Ce nouveau service d'accompagnement permet aux stagiaires rencontrant des difficultés dans la mise en œuvre des connaissances acquises de solliciter l'aide de nos formateurs sur des aspects relatifs aux programmes de formation suivis. Pour en bénéficier, il suffit de se rendre sur la page contact de notre site web et de remplir le formulaire. Une réponse est apportée par mail ou par téléphone dans un délai de 48 heures.

### Centre d'Examen

Notre centre de formation agréé Qualiopi bénéficie de l'agrément d'ENI et ICDL pour les certifications informatiques. C'est pour nos clients la garantie de pouvoir suivre des formations préparant à des certifications professionnelles.

### Aide à l'orientation

Pour chacune des grandes thématiques couvertes par notre offre de formation, nous proposons via nos spécialistes un rendez-vous physique ou téléphonique qui via un diagnostic permettra aux personnes souhaitant être accompagnées dans le choix d'un programme ou dans la définition d'un parcours de formation une orientation vers les programmes les plus adaptés à leurs besoins et à leur niveau.

### Aspects Pratiques

Dès leur inscription, les participants sont contactés par nos services qui s'assurent que les débits internet constatés sur le lieu depuis lequel ils souhaitent se former sont suffisants pour suivre la formation dans des conditions optimales.

À l'occasion de cet appel, nos experts s'assurent également qu'ils disposent du matériel nécessaire pour suivre la formation (PC Portable, webcam, Micro-casque..).

Avant le début de la formation, les participants reçoivent un lien leur permettant d'accéder à la classe virtuelle et leurs identifiants personnels de connexion. Un aide à l'utilisation de la solution de visioconférence utilisée leur est également proposée.

Le jour de la formation, ils se connectent à la classe virtuelle depuis leur navigateur internet. Ils voient et entendent le formateur ainsi que les autres participants et peuvent à tout moment communiquer avec eux.

Ils participent aux échanges et réalisent les ateliers dans les mêmes conditions que s'ils étaient en salle. Grâce à nos outils de prise en main à distance, les formateurs peuvent à tout moment prendre la main sur leurs postes pour les aider ou vérifier leurs TP.

Tout au long de la formation, les participants peuvent bénéficier de l'assistance immédiate de nos experts en composant le numéro qui leur a été communiqué avant la formation.

Des bilans intermédiaires ont lieu en présence des participants du formateur et du référent pédagogique d'Access it afin de vérifier l'état d'avancement de la session, les difficultés rencontrées et permettre d'éventuels actions correctives.

### Bénéfices pour les participants

Se former depuis leur lieu de travail ou leur domicile,

Accéder sans se déplacer à la qualité d'une formation délivrée par un formateur consultant ayant une expérience probante sur le sujet animé.

Bénéficier à distance de la richesse d'une formation interentreprises : échanges avec le formateur et les autres participants, partages d'expériences, ateliers pratiques...

Pouvoir se former en toutes circonstances et notamment en cas d'imprévu.

### Bénéfices pour l'entreprise

Optimiser ses budgets en limitant les frais de déplacement et d'hébergement.

Proposer à tous ses collaborateurs, quelle que soit leur situation géographique, des formations de qualité (en Inter comme en Intra).

Limiter les temps de déplacement.

Proposer davantage de choix dans les formations à des collaborateurs peu mobiles.

Assurer la montée en compétences de ses collaborateurs quelles que soient les circonstances