

Mis à jour le 20/10/2026

S'inscrire

Formation Cloud Data Fusion

2 jours (14 heures)

Présentation

Cloud Data Fusion est un service Google Cloud permettant de créer des pipelines de données visuels, évolutifs et adaptés aux besoins des entreprises. Avec son interface de conception, ses connecteurs et Wrangler pour la préparation des données, il facilite la construction de flux ETL entre des sources variées et vos plateformes analytiques.

Notre formation Cloud Data Fusion vous permettra de structurer vos traitements et de maîtriser les étapes essentielles de la conception à l'exploitation de pipelines de données.

Vous apprendrez à configurer les connexions, à nettoyer les données avec Wrangler et à combiner jointures, filtres et agrégations pour alimenter BigQuery depuis vos fichiers et référentiels.

À l'issue de la formation, vous serez en mesure de déployer des pipelines paramétrables, de planifier leurs exécutions, de contrôler la qualité des résultats et de diagnostiquer les erreurs grâce aux journaux et métriques disponibles.

Cette formation aborde également les comptes de service et les droits d'accès, les ressources de calcul et les bonnes pratiques de transfert entre environnements afin de fiabiliser vos flux tout en maîtrisant les coûts.

Comme toutes nos formations, celle-ci vous présentera **la dernière version stable** de la technologie et ses nouveautés.

Objectifs

- Concevoir un pipeline ETL visuel adapté à un besoin d'intégration de données
- Préparer et transformer des données avec Wrangler pour alimenter une destination analytique

- Configurer les connexions et les comptes de service nécessaires aux traitements
- Paramétrer et planifier des pipelines avec des macros et des arguments d'exécution
- Diagnostiquer les erreurs et ajuster les ressources à partir des journaux et métriques disponibles

Public visé

- Data Engineer
- Développeur ETL
- Consultant décisionnel
- Architecte de données

Pré-requis

- Pratique du SQL pour interroger et agréger des données
- Connaissance des traitements ETL et des principes d'intégration de données
- Première expérience de Google Cloud et de sa console d'administration
- Notions de Cloud Storage et de BigQuery pour le stockage et l'analyse de données

Pré-requis techniques

- Ordinateur avec 8 Go de mémoire vive recommandés et un navigateur Google Chrome ou Mozilla Firefox à jour
- Connexion Internet stable autorisant l'accès à la console Google Cloud et à l'interface de formation
- Compte nominatif donnant accès à un projet Google Cloud dédié, avec facturation active et quotas de calcul disponibles
- Instance Cloud Data Fusion accessible et autorisations nécessaires sur Cloud Storage et BigQuery préparées avant la session
- Comptes de service et accès à Cloud Logging et Cloud Monitoring configurés pour les exercices

Programme de notre formation Cloud Data Fusion

[Jour 1 - Matin]

Architecture et premiers flux de données

- Positionnement de Cloud Data Fusion dans une architecture d'intégration et distinction entre traitements batch et temps réel
- Organisation des instances, des namespaces et des pipelines dans la console et le Studio
- Cycle d'exécution des pipelines et rôle d'Apache Spark dans les traitements distribués
- Configuration des connexions, des sources et des destinations avec les plugins disponibles
- Distinction entre agent de service et compte d'exécution, droits d'accès et principe du moindre privilège
- Atelier pratique : Explorer l'instance préparée, configurer une source Cloud Storage et exécuter un premier chargement vers BigQuery.

[Jour 1 - Après-midi]

Préparation et transformation des données

- Exploration d'un échantillon dans Wrangler et limites de la préparation visuelle pour les pipelines batch
- Construction de recettes avec des directives de nettoyage, de conversion des types et de traitement des valeurs manquantes
- Définition des schémas d'entrée et de sortie et vérification de leur cohérence entre les étapes
- Enrichissement des flux par jointures, filtres, déduplication et agrégations
- Validation et Preview du pipeline, contrôle des résultats et traitement des enregistrements invalides
- Atelier pratique : Nettoyer un fichier de ventes, l'enrichir avec un référentiel BigQuery et vérifier les données chargées dans une table cible.

[Jour 2 - Matin]

Paramétrage et automatisation des pipelines

- Construction de pipelines réutilisables avec des macros sur les propriétés compatibles des plugins
- Organisation des préférences et des arguments d'exécution, héritage et priorités des valeurs
- Paramétrage des chemins sources, des tables cibles et des dates de traitement
- Déploiement et planification des traitements, dépendances et déclencheurs selon l'état des pipelines
- Transfert entre environnements par export et import JSON, vérification des plugins et recreation des déclencheurs
- Atelier pratique : Paramétrer le pipeline de ventes pour plusieurs fichiers, programmer son exécution et déclencher un pipeline de consolidation après sa réussite.

[Jour 2 - Après-midi]

Supervision et fiabilisation des traitements

- Lecture de l'historique des exécutions et des métriques pour contrôler les volumes et les durées
- Analyse des journaux dans Cloud Logging et suivi des indicateurs dans Cloud Monitoring
- Diagnostic des erreurs de permissions, de schéma, de connexion et de ressources
- Ajustement des profils de calcul, du parallélisme et des ressources selon les volumes et les coûts
- Stratégies de reprise, prévention des doublons et nettoyage des ressources de formation
- Atelier pratique : Diagnostiquer un échec provoqué, corriger le pipeline, vérifier sa relance sans doublons et comparer les métriques avant et après ajustement.

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.