



Intégrer et Orchestrer des Données avec Cloud Data Fusion

Lien :

<https://innov-systems.com/formation/integrer-et-orchestres-des-donnees-avec-cloud-data-fusion>

 DURÉE
3 jours (21h)

 RÉFÉRENCE
VSC310

 CATÉGORIE
Google Cloud

OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- ✓ Comprendre les principes d'intégration, de transformation et d'orchestration des données dans le cloud
- ✓ Maîtriser les fonctionnalités avancées de Cloud Data Fusion pour concevoir, déployer et superviser des pipelines complexes
- ✓ Utiliser Wrangler pour nettoyer, transformer et enrichir les données
- ✓ Intégrer des données de sources variées (API, bases de données, stockage cloud, flux en temps réel)
- ✓ Gérer les environnements d'exécution, les performances et la supervision des pipelines
- ✓ Mettre en œuvre la gouvernance des données à travers les métadonnées et la lignée des données
- ✓ Développer des pipelines de streaming et des workflows hybrides en production

POUR QUI ?

- ✓ Data engineers
- ✓ Architectes data
- ✓ Analystes techniques et responsables de l'intégration de données
- ✓ Responsables BI ou Cloud souhaitant industrialiser leurs flux de données



Programme détaillé

1 / INTRODUCTION À L'INTÉGRATION DE DONNÉES DANS LE CLOUD

- Comprendre les enjeux de l'intégration de données multi-sources et hybrides
- Identifier les cas d'usage typiques (migration, synchronisation, préparation analytique)
- Découvrir les outils et solutions d'intégration dans Google Cloud (Dataflow, Dataproc, Data Fusion)
- Positionner Cloud Data Fusion dans l'écosystème data de Google Cloud

2 / DÉCOUVERTE DE CLOUD DATA FUSION ET DE SON ENVIRONNEMENT

- Créer et configurer une instance Cloud Data Fusion
- Explorer l'interface utilisateur, les espaces de travail et les modes d'exécution
- Comprendre l'architecture modulaire (UI, services, pipeline engine)
- Identifier les composants principaux : sources, transformations, sinks, actions

3 / PREMIERS PIPELINES DE TRAITEMENT

- Utiliser Pipeline Studio pour concevoir un pipeline simple
- Ajouter des sources (BigQuery, GCS, fichiers CSV, bases relationnelles)
- Définir des transformations simples et des sorties
- Exécuter, surveiller et visualiser un pipeline en batch

4 / CONSTRUCTION DE PIPELINES COMPLEXES

- Utiliser les branchements, fusions et jointures pour enrichir les données
- Gérer les erreurs et exceptions avec les gestionnaires d'erreurs
- Utiliser les macros et variables d'exécution pour les paramètres dynamiques
- Automatiser les actions avant et après exécution (préparation, notification, nettoyage)

5 / ORCHESTRATION ET PLANIFICATION DES PIPELINES

- Mettre en œuvre des workflows multi-pipelines
- Configurer la planification récurrente et les dépendances d'exécution
- Importer, exporter et versionner les pipelines
- Mettre en place des notifications et alertes automatiques

6 / ENVIRONNEMENT D'EXÉCUTION ET SUPERVISION

- Comprendre les profils de calcul et les approvisionneurs
- Créer et configurer des profils d'exécution adaptés (batch, streaming)
- Surveiller la consommation, les métriques et les logs
- Dépanner les erreurs et optimiser les performances

7 / PRÉPARATION ET TRANSFORMATION DES DONNÉES AVEC WRANGLER

- Découvrir Wrangler et ses fonctionnalités de transformation visuelle
- Nettoyer, normaliser et enrichir des jeux de données
- Créer des directives personnalisées et automatiser les transformations
- Intégrer les transformations Wrangler dans un pipeline Data Fusion

8 / INTÉGRATION DE SOURCES DIVERSES ET PIPELINES DE STREAMING

- Explorer les connecteurs disponibles (BigQuery, Pub/Sub, Cloud Storage, bases SQL, API)
- Intégrer des données en temps réel avec les pipelines de streaming
- Comprendre l'architecture et les bonnes pratiques du streaming sur Data Fusion
- Utiliser les APIs Cloud DLP pour sécuriser et anonymiser les données sensibles

9 / MÉTADONNÉES, LIGNAGE ET GOUVERNANCE DES DONNÉES

- Comprendre les différents types de métadonnées (techniques, métiers, opérationnelles)
- Visualiser et interpréter la lignée des données dans Cloud Data Fusion
- Mettre en place la traçabilité des transformations et des sources
- Intégrer Cloud Data Fusion dans une démarche de gouvernance globale

Approche pédagogique

- ✓ Support Ecrit et Projection
- ✓ Exposés Interactifs, Podcasts et Vidéos
- ✓ Brainstorming et Jeux de Rôle
- ✓ Cas Pratiques et Labs inclus pour leur impact opérationnel
- ✓ Test de Validation des Acquis des Connaissances

Prochaines dates programmées

 28 au 30 Oct. 2026

 Distanciel

 Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Réservation & Renseignements

 **Téléphone** : +212 522 247 210
 **Email** : contact@innov-systems.com
 **Web** : <https://www.innov-systems.com>

Document généré le 25/09/2026 — Réf : VSC310
Innov Systems — Tous droits réservés