

Mis à jour le 26/08/2026

S'inscrire

Formation OpenLineage

2 jours (14 heures)

Présentation

OpenLineage est un standard ouvert pour la traçabilité des données, la lineage et l'observabilité des pipelines. Avec son modèle d'objets Run, Job et Dataset, ses facets extensibles et ses intégrations avec les outils de data engineering, OpenLineage aide à fiabiliser les chaînes de traitement modernes.

Notre formation OpenLineage vous permettra d'aller au-delà de la découverte du standard pour maîtriser les enjeux concrets de l'instrumentation et du suivi des traitements de données en environnement réel.

Vous apprendrez à comprendre le modèle d'objet, à exploiter les facets, à structurer les événements de lineage et à relier OpenLineage aux pratiques de data observability et de gouvernance.

À l'issue de la formation, vous serez en mesure de concevoir une stratégie de traçabilité cohérente, de mieux diagnostiquer les dépendances entre jobs, d'améliorer la qualité des métadonnées et de renforcer la lisibilité des flux de données.

Cette formation aborde également l'écosystème OpenLineage, les bonnes pratiques d'intégration, les usages orientés production et les méthodes de mise en œuvre adaptées aux équipes Data Engineer.

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

Objectifs

- Comprendre le rôle d'OpenLineage dans la traçabilité des traitements de données.

- Maîtriser le modèle conceptuel basé sur Run, Job et Dataset.
- Identifier et exploiter les facets pour enrichir les événements de lineage.
- Structurer une approche d'observabilité des pipelines adaptée à un contexte de production.
- Mettre en place des bonnes pratiques d'instrumentation et d'intégration avec l'écosystème data.
- Analyser les dépendances, les incidents et la qualité des métadonnées pour fiabiliser les flux.

Public visé

- Data Engineer
- Ingénieur Data
- Architecte Data
- Responsable Data Platform
- Développeur ETL

Pré-requis

- Connaissances de base en data engineering et en orchestration de pipelines.
- Compréhension des notions de métadonnées, de flux de traitement et de dépendances entre jobs.
- Pratique d'un environnement cloud ou d'une plateforme data de type Airflow, dbt ou équivalent.
- Lecture aisée de formats structurés comme JSON et compréhension des échanges API.

Pré-requis techniques

- Un ordinateur récent avec suffisamment de mémoire pour exécuter les outils de manipulation de données et les exemples fournis.
- Une connexion Wi-Fi ou Ethernet stable pour accéder aux ressources de formation et aux environnements distants.
- Un navigateur à jour et un éditeur de code compatible, par exemple VS Code.
- Selon les ateliers, un environnement local avec Java ou Python 3.x installé, ainsi que Git.
- En distanciel, un micro et un casque fonctionnels pour les échanges et les travaux guidés.

Programme de notre formation OpenLineage

[Jour 1 - Matin]

Fondations et concepts clés

- Comprendre le rôle d'OpenLineage dans la lineage data et l'observabilité.
- Identifier les problèmes résolus par un standard ouvert de traçabilité des traitements.
- Découvrir le périmètre fonctionnel, les intégrations et les cas d'usage principaux.
- Maîtriser le vocabulaire de base, run, job, dataset et événements.
- Relier OpenLineage aux enjeux de gouvernance, de qualité et de diagnostic.
- Atelier pratique : cartographier un pipeline existant et repérer les informations de lineage à capturer.

[Jour 1 - Après-midi]

Modèle d'objet et événements

- Explorer le modèle objet OpenLineage et la logique de relation entre entités.
- Comprendre la structure d'un event et les informations minimales attendues.
- Analyser la distinction entre métadonnées statiques et métadonnées liées à l'exécution.
- Étudier les principes d'identification, de nommage et de cohérence des objets.
- Découvrir les notions de backend et de transport des événements vers une cible.
- Atelier pratique : construire manuellement un événement OpenLineage simple à partir d'un scénario de traitement.

[Jour 2 - Matin]

Facets et enrichissement

- Comprendre le rôle des facets comme extension du modèle de base.
- Différencier les run facets, job facets et dataset facets.
- Découvrir les usages des facets pour enrichir les métadonnées de production.
- Apprendre à lire une spécification de facet et à interpréter sa structure.
- Relier les facets aux besoins de gouvernance, d'audit et d'observabilité.
- Atelier pratique : analyser plusieurs facets et choisir celles à utiliser pour un cas métier donné.

[Jour 2 - Après-midi]

Intégration et mise en production

- Étudier les stratégies d'instrumentation des jobs et pipelines.
- Comprendre les bonnes pratiques d'intégration avec les outils de data engineering.
- Structurer une démarche de déploiement orientée production et exploitation.
- Aborder les contrôles de qualité, le suivi des dépendances et le diagnostic des incidents.
- Identifier les points de vigilance pour fiabiliser la collecte et l'usage des métadonnées.
- Atelier pratique : concevoir un plan d'intégration OpenLineage pour une architecture data cible.

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.