

Mis à jour le 30/09/2025

S'inscrire

Formation Apache OpenDAL

3 jours (21 heures)

Présentation

Apache OpenDAL est une couche d'accès aux données unifiée pour les environnements cloud-native. Multi-backends, elle permet d'interagir de manière homogène avec les stockages objets, fichiers et services distribués, tout en offrant performance et extensibilité.

Notre formation OpenDAL vous apprendra à abstraire vos accès aux données, sécuriser les échanges et industrialiser vos pipelines ETL/ELT avec une observabilité complète et une maîtrise des coûts.

Vous saurez exploiter les APIs à options pour un contrôle fin, intégrer OpenDAL dans vos chaînes CI/CD et appliquer des patterns d'optimisation adaptés aux charges Data & Big Data.

À l'issue de la formation, vous serez en mesure de bâtir des pipelines cloud-agnostiques, de fiabiliser vos traitements et de publier des runbooks SRE prêts pour la production.

Comme toutes nos formations, celle-ci couvre la [dernière version stable](#) et ses nouveautés, avec une approche résolument pratique et opérationnelle.

Objectifs

- Maîtriser OpenDAL et ses APIs à options (read/write/list/stat/delete)
- Concevoir des pipelines ETL/ELT cloud-agnostiques et performants
- Renforcer sécurité, résilience et observabilité
- Industrialiser via CI/CD et IaC (Terraform/OpenTofu)
- Optimiser coûts et SLO grâce aux bons patterns

Public visé

- Cloud Engineers
- Data Engineers & équipes Big Data
- Plateformes données, SRE/DevOps

Pré-requis

- Bases cloud (AWS/GCP/Azure) et stockage objet
- Notions de programmation (Rust/Go/Python/Java)
- Connaissances CI/CD et Terraform recommandées

Programme de notre formation Apache OpenDAL

[Jour 1 – Matin]

Fondamentaux & prise en main

- Comprendre Apache OpenDAL : couche d'accès aux données unifiée multi-stockages
- Principes clés : Open Community, Solid Foundation, Fast Access, Extensible Architecture
- Panorama des backends : S3, GCS, Azure Blob, HDFS, POSIX...
- Modèle opérateur / services / couches et gestion des capacités
- Cas d'usage Data & Cloud-native : ingestion, ETL/ELT, feature store, serverless
- Atelier pratique : projet OpenDAL & oli, liste/lecture d'objets S3 via credentials temporaires.

[Jour 1 – Après-midi]

APIs & options : manipuler les données proprement

- API read/write/list/stat/delete et Options-based API (ReadOptions, WriteOptions...)
- Gestion des métadonnées, ranges, multipart et streaming
- Stratégies de résilience : retries, backoff, idempotence
- Multi-langages : Rust, Go, Python, Java (bindings)
- Bonnes pratiques de nommage et préfixes pour data lakes
- Atelier pratique : lecture/écriture streaming et usage d'options (range, type, checksum).

Sécurité & conformité

- Auth : SigV4, Service Accounts, SAS, OIDC
- Chiffrement TLS et au repos, politiques de sécurité fournisseurs
- Gestion des secrets & variables, intégration Vault/KMS
- Cloisonnement comptes/projets et gouvernance
- Journaux & audits des accès
- Atelier pratique : S3 + KMS, rôles assumés, vérification des traces d'audit.

[Jour 2 – Matin]

Performance, fiabilité & industrialisation

- Comprendre latence & throughput ; range requests
- Multipart upload, pipelining, batching, concurrence
- Patterns cache-aside, write-through, read-through
- Tiering & lifecycle ; classes de stockage
- Observabilité : métriques, logs, traces ; bench avec oli
- Atelier pratique : bench multi-régions + tableaux de bord Prometheus/Grafana.

[Jour 2 – Après-midi]

Robustesse : erreurs réelles & tests

- Erreurs réseau, throttling et cas S3 200 + erreur
- Timeouts, circuit-breaker, limiteurs de débit
- Tests de charge et de chaos
- Contracts & golden files
- Politiques de retries par opération
- Atelier pratique : simuler throttling, régler retries/backoff, valider des SLO.

CI/CD & IaC

- Provision Terraform/OpenTofu, secrets CI
- Gabarits 12-Factor : config, blue/green, canary
- Packaging Rust/Go/Python/Java
- Scans sécu & policy as code
- Promotion multi-environnements
- Atelier pratique : pipeline CI provisioning + tests end-to-end OpenDAL.

[Jour 3 – Matin]

Intégration avancée & opérations

- POSIX-like vs object store ; implications parquet/csv/json
- Écosystème Arrow/Polars/Pandas + OpenDAL
- Liens pré-signés, staging, atomic move
- Stratégies multi-cloud/hybride et fallback
- Data contracts & catalogues
- Atelier pratique : mini-pipeline ELT cloud-agnostique avec OpenDAL.

[Jour 3 – Après-midi]

Exploitation & SRE

- Runbook incidents et opérations
- Tableau de bord latence/erreurs/débit
- Alerting : p95, erreurs classées par backend
- Budgets d'erreur, SLO/SLI batch vs interactif
- Optimisation des coûts (listings, small files, multipart)
- Atelier pratique : runbook + SLO & tests d'alerte.

Mise en production & roadmap équipe

- Checklist sécurité, observabilité, performance, limites
- Versioning & compatibilité ; suivi des releases
- Gouvernance préfixes, séparation environnements
- Documentation : guides dev & exploitation
- Plan d'adoption et gestion des risques
- Atelier pratique : revue finale + post-mortem simulé.

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.