

Mis à jour le 09/02/2026

S'inscrire

Formation Calista Framework

2 jours (14 heures)

Présentation

Calista est un framework de Qualité de Données (DQ) open source et industriel qui permet de valider des volumes massifs de données au sein d'environnements Big Data.

Notre formation Calista vous permettra de maîtriser l'approche "Data Quality as Code" pour automatiser vos contrôles de données. Vous apprendrez à configurer le moteur, à définir des règles de validation complexes et à intégrer ces tests directement dans vos pipelines ETL ou Spark.

L'objectif de notre formation est de garantir une donnée fiable, intègre et prête pour l'analyse décisionnelle.

Comme toutes nos formations, celle-ci privilégie une approche pratique et opérationnelle basée sur des cas d'usage concrets de production.

Objectifs

- Comprendre la philosophie de Calista et son architecture.
- Mettre en place des profils de données et des règles de contrôle.
- Isoler et traiter les anomalies de données de manière industrielle.
- Intégrer Calista dans une chaîne CI/CD ou un orchestrateur.
- Générer des rapports de qualité et des KPIs actionnables.

Public visé

- Data Engineers
- Data Quality Managers
- Architectes de données

Pré-requis

- Connaissances de base en SQL
- Notions en Scala ou Java
- Familiarité avec l'écosystème Spark/Hadoop

Pré-requis techniques

- 8 Go de RAM au minimum, 16 Go recommandés
- Environnement Docker ou installation locale de Spark
- Un IDE (IntelliJ IDEA de préférence)
- Accès à un terminal (Bash ou PowerShell)

Formation Calista : Qualité de Données & Industrialisation

[Jour 1 - Matin]

Fondamentaux et Architecture

- Comprendre les enjeux de la Data Quality : les 6 dimensions clés
- Présentation de Calista : philosophie "DQ as Code" et écosystème
- Architecture technique : moteur Spark, configurations et métadonnées
- Comparaison vs outils du marché (Great Expectations, Deequ)
- Installation et setup de l'environnement de développement
- Atelier pratique : Initialisation d'un projet Calista et premier scan d'un dataset.

[Jour 1 - Après-midi]

Mise en œuvre des contrôles

- Le Profiling : analyse automatique de la structure et distribution
- Définition des Rules & Checks : types, formats (Regex), nullité, unicité
- Gestion des seuils de tolérance et niveaux d'alerte
- Configuration des sorties (outputs) et logs d'exécution
- Atelier pratique : Création d'un plan de contrôle complet sur un jeu de données métier.

[Jour 2 - Matin]

Fonctions Avancées et Customisation

- Contrôles croisés (Cross-checks) : cohérence inter-sources
- Calcul de KPIs personnalisés et extension du framework
- Gestion et isolation des anomalies (données rejetées)
- Génération de rapports de non-conformité

- Atelier pratique : Réconciliation complexe entre une source SQL et un fichier CSV.

[Jour 2 - Après-midi]

Industrialisation et Monitoring

- Intégration CI/CD (Jenkins, GitLab CI)
- Exploitation des résultats pour le Reporting et Dashboards
- Optimisation des performances sur gros volumes (Spark tuning)
- Stratégies de mise en production et surveillance
- Atelier pratique : Pipeline complet de l'injection au rapport de qualité final.

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.