

Mis à jour le 20/03/2026

S'inscrire

Formation Apache Atlas

3 jours (21 heures)

Présentation

Apache Atlas est une solution open source de gouvernance des données conçue pour les environnements Big Data. Elle permet de centraliser les métadonnées, de visualiser le data lineage et de renforcer la maîtrise des actifs data dans des architectures distribuées.

Notre formation Apache Atlas vous permettra de comprendre les principes de catalogage, de traçabilité et de classification des données dans un environnement moderne de gouvernance.

Vous apprendrez à installer, configurer, administrer et automatiser la plateforme, tout en l'intégrant à un écosystème utilisant notamment Hadoop, Hive, Spark ou Ranger.

À l'issue de la formation, vous serez en mesure de structurer un catalogue de données, de mettre en œuvre le lineage, de sécuriser les accès et de déployer une démarche de gouvernance adaptée aux besoins de votre organisation.

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

Objectifs

- Comprendre les concepts fondamentaux de la gouvernance des données.
- Installer et configurer Apache Atlas dans un environnement Big Data.
- Créer et administrer des métadonnées, des entités et des classifications.
- Exploiter le data lineage pour analyser les flux et les impacts.
- Mettre en œuvre la sécurité, la conformité et l'automatisation d'Atlas.

Public visé

- Data engineers
- Data architects
- Administrateurs Big Data
- Profils DevOps data et responsables de la gouvernance technique

Pré-requis

- Connaissances de base des architectures Big Data
- Notions sur les traitements distribués et les métadonnées
- Expérience générale en administration technique ou en environnement data

Programme de formation Apache Atlas

[Jour 1 - Matin]

Fondamentaux de la gouvernance des données avec Apache Atlas

- Comprendre les enjeux de la gouvernance des données
- Introduction à l'écosystème Big Data et rôle d'Apache Atlas
- Maîtriser les concepts clés : metadata, data lineage, classification
- Découvrir l'architecture globale d'Apache Atlas
- Identifier les principaux cas d'usage en entreprise
- Atelier pratique : découverte de l'interface Atlas et navigation dans le catalogue.

[Jour 1 - Après-midi]

Installation, configuration et prise en main

- Étudier les prérequis techniques et l'environnement cible
- Installer Apache Atlas dans un contexte Hadoop
- Configurer les paramètres initiaux de la plateforme
- Prendre en main l'interface utilisateur et les fonctions principales
- Créer les premiers objets de métadonnées
- Atelier pratique : installation et configuration initiale d'un environnement Atlas.

Modélisation et organisation des métadonnées

- Créer des types, des entités et des relations
- Gérer les classifications, les labels et la structuration du catalogue
- Organiser un catalogue de données exploitable par les équipes
- Utiliser les fonctions de recherche et d'exploration
- Appliquer les bonnes pratiques de modélisation métier et technique
- Atelier pratique : modélisation d'un référentiel de métadonnées métier.

[Jour 2 - Matin]

Traçabilité, lineage et analyse d'impact

- Comprendre les principes du data lineage
- Visualiser les flux de données entre systèmes et traitements
- Exploiter l'intégration avec Apache Hive et Apache Spark
- Réaliser une analyse d'impact sur les transformations
- Mettre en place une traçabilité fiable dans une chaîne data
- Atelier pratique : analyse d'un flux de données et visualisation du lineage.

[Jour 2 - Après-midi]

Sécurité, gouvernance avancée et conformité

- Définir des politiques de sécurité autour des métadonnées
- Intégrer Apache Ranger pour la gouvernance des accès
- Gérer les rôles, habilitations et contrôles d'accès
- Répondre aux enjeux de conformité et de RGPD
- Exploiter l'audit et la traçabilité des opérations
- Atelier pratique : mise en place de règles de sécurité et de gouvernance.

API, automatisation et intégration dans l'écosystème data

- Utiliser l'API REST d'Apache Atlas
- Automatiser l'alimentation et la mise à jour des métadonnées
- Intégrer Atlas avec des outils tiers et des workflows existants
- Synchroniser les métadonnées dans plusieurs environnements
- Mettre en place la supervision et la maintenance de la plateforme
- Atelier pratique : automatisation d'actions Atlas via API.

[Jour 3 - Matin]

Optimisation, industrialisation et bonnes pratiques

- Optimiser les performances et la scalabilité de la plateforme
- Gérer les volumes importants de métadonnées
- Définir une stratégie de gouvernance des données durable
- Industrialiser les processus d'administration et de mise à jour
- Étudier des architectures avancées autour d'Atlas
- Atelier pratique : optimisation d'un environnement Atlas existant.

[Jour 3 - Après-midi]

Cas d'usage avancés et gouvernance multi-sources

- Déployer Apache Atlas dans un contexte Data Lake
- Mettre en œuvre une gouvernance sur plusieurs sources de données
- Exploiter Atlas comme data catalog et outil de data discovery
- Favoriser la collaboration entre équipes data, métiers et gouvernance
- Analyser des retours d'expérience et cas d'entreprise
- Atelier pratique : étude de cas complète de gouvernance multi-sources.

Projet fil rouge et mise en situation opérationnelle

- Définir un cas d'usage métier autour d'un catalogue de données
- Mettre en place un catalogue complet avec entités et classifications
- Implémenter le lineage et les mécanismes de traçabilité
- Sécuriser les accès et structurer la gouvernance cible
- Valider les choix d'architecture et restituer les résultats
- Atelier pratique : projet final encadré de déploiement et de gouvernance avec Apache Atlas.

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.

