Apache Atlas, DataHub & OpenMetadata : Déployer un Catalogue de Données d'Entreprise

Lien : https://innov-systems.com/formation/apache-atlas-datahub-openmetadata-deployer-un-catalogue-de-donnees-dentreprise

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI95

CATÉGORIE
Big Data : fondamentaux et architecture

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Comprendre le rôle d’un data catalog en entreprise
- Maîtriser Apache Atlas, DataHub et OpenMetadata
- Mettre en place une gouvernance des données efficace
- Centraliser et organiser les métadonnées d’entreprise
- Implémenter le data lineage de bout en bout
- Améliorer la découverte et l’accès aux données
- Intégrer les catalogues dans les architectures data modernes
- Renforcer la conformité et la sécurité des données
- Automatiser la collecte des métadonnées
- Faciliter la collaboration entre équipes data
- Déployer un data catalog en environnement production
- Étudier des cas concrets de gouvernance data

POUR QUI ?

- Data engineers
- Data architects
- Data analysts
- Data scientists
- Data stewards
- Data governance managers
- BI developers
- DevOps engineers
- Cloud engineers
- Compliance officers
- CDO (Chief Data Officers)
- Toute personne impliquée dans la gouvernance et la gestion des données d’entreprise

Innov Systems

Programme détaillé

1 / Introduction aux catalogues de données d’entreprise

- Rôle du data catalog dans la gouvernance des données
- Enjeux de discovery, lineage et conformité
- Cas d’usage en data-driven enterprise

2 / Présentation des solutions de data catalog

- Apache Atlas : gouvernance data native Hadoop
- DataHub : plateforme moderne de metadata management
- OpenMetadata : catalogue open source cloud-native

3 / Architecture d’un data catalog moderne

- Métadonnées techniques, business et opérationnelles
- Centralisation et indexation des assets data
- API et intégration avec l’écosystème data

4 / Collecte et ingestion des métadonnées

- Connecteurs vers bases de données et data lakes
- Extraction automatique des schémas
- Synchronisation continue des métadonnées

5 / Data lineage et traçabilité

- Suivi des flux de données end-to-end
- Impact analysis des transformations
- Visualisation des pipelines de données

Innov Systems

6 / Gouvernance des données

- Classification des données sensibles
- Politiques d’accès et ownership
- Gestion des rôles data stewardship

7 / Recherche et discovery des données

- Moteur de recherche de datasets
- Tags, glossaires et documentation métier
- Amélioration de l’accessibilité des données

8 / Intégration avec l’écosystème data

- Connexion avec Spark, Airflow, dbt et BI tools
- Intégration avec data lakes et data warehouses
- APIs pour automatisation et extension

9 / Sécurité, conformité et audit

- Gestion des accès aux métadonnées
- Conformité RGPD et réglementations
- Traçabilité des usages des données

10 / Déploiement et bonnes pratiques

- Installation et configuration en production
- Adoption organisationnelle du data catalog
- Études de cas en entreprise

Innov Systems

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

23 au 27 Nov. 2026	Distanciel
22 au 26 Fév. 2027	Présentiel - Casablanca
19 au 23 Avr. 2027	Présentiel - Casablanca
14 au 18 Juin 2027	Présentiel - Casablanca
09 au 13 Août 2027	Présentiel - Casablanca

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

Innov Systems — Tous droits réservés