

Architecture lakehouse – Concevoir, implémenter et migrer

3 jours (21 heures)

Perfectionnement

Présentiel ou classe à distance

à partir de

2 625 € HT par personne (TVA en vigueur) **Repas inclus en présentiel**

Contenu mis à jour le 25/09/2026. Document téléchargé le 07/10/2026.

Objectifs

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

- Identifier les différences entre architectures traditionnelles et architecture lakehouse
- Déterminer les concepts et fonctionnalités clés d'un lakehouse, tels que le support des transactions, le "Time Travel" et l'évolution du schéma
- Concevoir les couches architecturales pour le stockage, le calcul, la gestion des métadonnées et la consommation des données
- Evaluer et comparer les solutions lakehouse avec une exigence "future-ready"
- Implémenter une solution lakehouse et migrer vos données.

Les compétences visées par cette formation font l'objet d'une évaluation.

À qui s'adresse cette formation ?

DSI, urbanistes Data, architectes Data, chefs de projets, chefs de produits Data, ingénieurs DevOps / MLOps / DataOps.

Prérequis

Avoir de l'expérience dans la gestion des données d'entreprise.

Programme

1 Pendant votre formation

Architecture lakehouse – Concevoir, implémenter et migrer

Jour 1

Comprendre l'architecture lakehouse

- Importance de l'architecture Data pour les plateformes de données
- Architectures Data traditionnelles (Data Warehouse, Data Lake) et défis face aux nouveaux usages
- Composants clés d'une plateforme de données
- L'architecture lakehouse : le meilleur des deux mondes
- Caractéristiques des lakehouses
- Lakehouses et services Cloud
- Offres du marché :
 - Databricks
 - Delta Lake
 - Apache Hudi
 - Apache Iceberg

Exemples de travaux pratiques (à titre indicatif)

- Analyse de cas d'étude de différentes implémentations / migrations vers des lakehouses dans divers secteurs d'activité
- Analyse des retours d'expérience

Lakehouse : la couche stockage

- Concepts clés de stockage Big Data
- Stockage à base de lignes vs à base de colonnes
- Axes d'optimisation de la performance
- Composants du stockage d'un lakehouse
- Formats de fichiers
- Formats des tables
- Aspects clés de la conception

Exemples de travaux pratiques (à titre indicatif)

- Conception et benchmarking de solutions de stockage lakehouse à base de lignes et à base de colonnes

Jour 2

Lakehouse : les catalogues de données

- Rappel sur les métadonnées (techniques, métier) et la notion d'analytique self-service
- Articulation metastore / catalogue de données

- Caractéristiques d'un catalogue de données :
 - Search / explore / discover
 - Classification
 - Gouvernance
 - Linéage
- Catalogue de données unifié et défi des architectures deux-tiers
- Implémentation d'un catalogue de données avec :
 - Hive Metastore
 - Un Cloud provider

Exemples de travaux pratiques (à titre indicatif)

- Conception et mise en place d'un catalogue de données, comparaison d'alternatives d'implémentation et d'options architecturales

Lakehouse : les moteurs de calcul

- Importance du calcul sur les données dans une architecture lakehouse
- Intérêts d'avoir de multiples moteurs de calcul
- Traitement batch et temps réel unifié
- Performance de BI améliorée
- Outils Open Source de calcul
- Services Cloud
- Plateformes tierces parties
- Aspects clés de conception

Exemple de travaux pratiques (à titre indicatif)

- Réalisation de traitements batch et temps réel unifiés sur une plateforme Open Source et/ou Cloud

Gouvernance et sécurité de l'architecture lakehouse

- Rappels sur les aspects gouvernance et sécurité
- Gestion des métadonnées dans un lakehouse
- Réglementation et conformité
- Impact sur la qualité des données et des modèles de Machine Learning (ML) et d'Intelligence Artificielle (IA)
- Linéage des données et actifs IA
- Partage des données et actifs IA
- Audit et surveillance
- Protection des données
- Manipulation des données sensibles
- Les rôles dans la gouvernance d'un lakehouse :
 - Data Owners
 - Analysts
 - Scientists

- Stewards
- Architects
- Admins
- Ingénieurs ML / BI

Exemples de travaux pratiques (à titre indicatif)

- Définition et mise en oeuvre de politiques de gouvernance et de sécurité d'un service lakehouse sur le Cloud

Jour 3

Conception et implémentation d'une plateforme lakehouse

- Rappel des meilleures pratiques
- Comprendre le besoin, collecte des exigences
- Etude de l'existant
- Comprendre la vision et la stratégie Data de l'organisation
- Définir les solutions possibles
- Réaliser des interviews et ateliers
- Choisir la meilleure architecture
- Spécifier le lakehouse :
 - Ecosystème des données
 - Scalabilité et performance
 - Contrôle des coûts et optimisation
- Optimisation des opérations
- Gouvernance et sécurité
- Futures tendances et impact de l'IA

Exemples de travaux pratiques (à titre indicatif)

- Conception et implémentation d'une plateforme Data à partir des exigences collectées auprès du client demandeur (jeu de rôle)

Migration vers une plateforme lakehouse

- Introduction à la migration de données
- Importance de la préparation des données
- Evaluation de l'infrastructure existante et des capacités de données
- Planification de la migration
- Création d'un plan de migration détaillé
- Choix des outils et des technologies adaptés à la migration
- Définition des phases de la migration :
 - Préparation
 - Migration
 - Validation

- Déploiement
- Techniques de migration
- Migration Lift and Shift
- Migration par restructuration des données (replatforming)
- Approche incrémentielle vs migration Big Bang
- Gestion des données durant la migration
- Techniques de synchronisation des données
- Validation et qualité des données pendant la migration
- Gestion des métadonnées
- Aspects sécuritaires et de conformité
- Assurer la sécurité des données pendant la migration
- Conformité réglementaire et implications légales
- Post-migration
- Tests post-migration pour garantir l'intégrité et la performance
- Optimisation post-migration et ajustements

Exemples de travaux pratiques (à titre indicatif)

- Atelier de planification d'une migration d'un entrepôt de données vers un lakehouse
- Mise en oeuvre de la migration dans un environnement Open Source ou Cloud

Financement

Plusieurs dispositifs de financement sont accessibles via les OPCO (Opérateurs de Compétences), organismes agréés par le ministère du Travail dont le rôle est d'accompagner, collecter et gérer les contributions des entreprises au titre du financement de la formation professionnelle.

Pour plus d'information, une équipe de gestionnaires SKOLAE Formation spécialisée vous accompagne dans le choix de vos formations et la gestion administrative.

[Nous contacter pour en savoir plus →](#)

Modalités pédagogiques

Formation délivrée en présentiel ou distanciel* (blended-learning, e-learning, classe virtuelle, présentiel à distance).

Le formateur alterne entre méthode** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation).

Variables suivant les formations, les moyens pédagogiques mis en oeuvre sont :

- Ordinateurs Mac ou PC (sauf pour certains cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel)
- Environnements de formation installés sur les postes de travail ou en ligne
- Supports de cours et exercices

En cas de formation intra sur site externe à M2I, le client s'assure et s'engage également à avoir toutes les ressources matérielles pédagogiques nécessaires (équipements informatiques...) au bon déroulement de l'action de formation visée conformément aux prérequis indiqués dans le programme de formation communiqué.

* nous consulter pour la faisabilité en distanciel

** ratio variable selon le cours suivi

Accessibilité aux personnes en situation de handicap

Détails de la session

Du 2 février 2027 au 4 février 2027 (3 jours / 21 heures)

À distance

Du 2 février 2027 au 4 février 2027

À distance

- 2 février 2027 9h00 – 16h00
- 3 février 2027 9h00 – 16h00
- 4 février 2027 9h00 – 16h00