SKILLIA

Catalogue des formations

Innovez, Protégez, Transformez : les formations IA, Cyber et Digital signées Skillia

2025-2026

www.skillia.fr

Sommaire

Nos Valeurs

Une démarche axée sur la qualité

Chez Skillia, nous concevons des formations sur mesure, pensées pour répondre aux spécificités de votre organisation, de votre marché et de votre environnement.

La qualité constitue le pilier central de notre approche : pertinence des thématiques, sélection rigoureuse des experts, méthodes pédagogiques éprouvées, organisation soignée et mise à jour continue de notre offre pour anticiper les évolutions technologiques et les besoins du marché.

L'innovation au service de la formation

Créativité, innovation, écoute, réflexion et rigueur sont au cœur de notre ADN.

Ces valeurs guident chacune de nos actions pour vous offrir des formations à la pointe : des approches pédagogiques modernes, des contenus actualisés et une expérience d'apprentissage stimulante, adaptée aux enjeux d'aujourd'hui.

Une écoute active des entreprises

Chaque entreprise bénéficie d'un interlocuteur dédié, véritable consultant en formation.

Sa mission : analyser vos besoins, recommander les solutions les plus pertinentes en matière de contenu, d'organisation, de budget et de financement, puis vous accompagner tout au long du déploiement.

Il veille à la réussite du projet, à l'atteinte de vos objectifs et à la satisfaction des apprenants.

Pourquoi choisir Skillia ?

Des intervenants d’exception

es formateurs Skillia sont des experts reconnus, dotés de plusieurs années d’expérience dans leurs domaines respectifs et ayant occupé des postes à responsabilité en entreprise.

Ils conçoivent et animent eux-mêmes leurs formations, en s’appuyant sur une double expertise : une solide expérience de terrain et une maîtrise pédagogique éprouvée — véritable gage de qualité et de pertinence.

Une offre en constante évolution

Chez Skillia, l’innovation est continue. Nos programmes de formation sont régulièrement actualisés pour intégrer les dernières avancées technologiques et répondre aux nouveaux défis des entreprises.

Une approche résolument opérationnelle

es formations Skillia sont pensées pour une mise en pratique immédiate. Conçues et animées par des professionnels expérimentés, elles comportent au minimum 60 % d’exercices pratiques et d’études de cas inspirés de situations réelles. L’organisation progressive des modules permet d’acquérir les compétences nécessaires à chaque étape, dans les conditions les plus efficaces.

Une réactivité à chaque étape

Nos consultants en formation sont à votre écoute pour définir rapidement les solutions les plus adaptées à vos besoins : sélection de cours ou de parcours, formations intra ou sur mesure, ingénierie pédagogique et dispositifs d’accompagnement spécifiques à votre entreprise.

Former avant de recruter : un levier gagnant !

La reconversion professionnelle est une approche de recrutement innovante et efficace. Elle consiste à identifier des talents motivés, dotés du potentiel nécessaire, puis à leur offrir un parcours de formation sur mesure pour développer les compétences indispensables à leur futur métier.

Cette stratégie représente une opportunité mutuellement bénéfique : elle permet aux entreprises de répondre à leurs besoins en compétences tout en offrant aux candidats une véritable montée en qualification et une nouvelle trajectoire professionnelle.

Avantages pour l'entreprise

Combler le manque de compétences clés sur le marché

Recruter autrement : des profils compétents, motivés et alignés avec vos valeurs

Être en phase avec les dernières avancées technologiques

Avantages pour les candidats

Une reconversion professionnelle vers des métiers d'avenir

Valoriser son employabilité et répondre aux besoins réels du marché du travail.

Informations Pratiques SKILLIA :

Public visé : Professionnels (salariés, indépendants) et particuliers.

Modalités : Formations disponibles en **distanciel** (classe virtuelle synchrone) ou en **présentiel** (dans les locaux de nos partenaires ou en intra-entreprise).

Délais d'accès : Entre **15 et 30 jours** à compter de la signature de la convention et selon le mode de financement sollicité.

Accessibilité Handicap :

- **En distanciel :** Nos parcours sont adaptables (assistance technique, supports numériques accessibles).
- **En présentiel :** Nous nous assurons avec nos partenaires que les lieux d'accueil respectent les normes d'accessibilité (ERP).
- *Pour toute étude personnalisée de vos besoins spécifiques, contactez notre référent handicap :* ***Skander Maalej****.*

Contact : administration@skillia.fr

Notre catalogue

www.skillia.fr

Data et Big Data

Snowflake Core Architecture & Ingestion Pipelines

Prix : 2 580 € HT

Durée : 4 jours (28 heures)

Public cible : Ingénieurs Data & Développeurs, administrateurs de plateformes analytiques, Data Engineers & Architectes Data, candidats à la certification Snowflake.

Prérequis : Maîtrise du langage SQL (requêtes, jointures, DDL/DML). Connaissance des concepts de bases de données relationnelles et de l'entreposage de données (data warehousing). Une première exposition au cloud (AWS, Azure ou GCP) est un plus. Accès à un compte Snowflake (trial ou entreprise) pour les hands-on labs.

Objectifs pédagogiques

À l'issue de la formation, les participants seront capables de :

- Provisionner et verrouiller l'infrastructure Snowflake : architecture 3 couches, Warehouses, sécurité réseau et chiffrement
- Implémenter le modèle RBAC et la gouvernance des données avec Snowflake Horizon (masquage, Row Access, tagging)
- Construire des pipelines d'ingestion et de transformation : Snowpipe, Streams, Tasks (DAG) et Dynamic Tables
- Maîtriser le FinOps, la résilience (Time Travel, Cloning), l'IA Cortex et les pièges de la certification

Méthodes et moyens pédagogiques : Formation intensive et hands-on : 4 hands-on labs SQL de bout en bout. Format : présentiel ou distanciel. Finalité : préparation à la certification Snowflake.

Modalités d'évaluation

Évaluation en continu à travers les 4 hands-on labs SQL réalisés tout au long de la formation ; parcours orienté préparation à la certification Snowflake.

Accessibilité : Formation accessible aux personnes en situation de handicap, contactez notre référent dédié.

Programme

1. Jour 1 · Architecture & Provisioning de Ressources

- Objectif du Jour 1 : provisionner l'infrastructure et verrouiller les accès au niveau réseau et compte.
- Les 3 couches : Storage, Compute (Virtual Warehouses) et Cloud Services
- Cycle de vie et gestion des Warehouses (Sizing, Multi-cluster, auto-suspend/resume, scaling vertical vs horizontal)
- Création et gestion des ressources : Databases, Schemas (Transient vs Permanent vs Temporary)

2. Jour 1 · Sécurité Réseau & Authentification

- Configuration des Network Policies (autoriser/bloquer des IPs ou VPCs)
- Concepts de sécurité : MFA, SSO et chiffrement des données (clés gérées par Snowflake vs Tri-Secret Secure)
- Hands-on Lab : Déploiement par script SQL d'un environnement complet isolé, sécurisé par une politique réseau, avec affectation de profils de Warehouses distincts

3. Jour 2 · Sécurité & Modèle de Responsabilité (RBAC)

- Objectif du Jour 2 : implémenter le contrôle d'accès, la gouvernance de la donnée et les flux d'entrée/sortie.
- Gestion des Users et cycle de vie des comptes
- Hiérarchie des Roles natifs (ACCOUNTADMIN à PUBLIC) et création de rôles sur-mesure (principe du moindre privilège)
- Attribution fine des privilèges (GRANT / REVOKE)

4. Jour 2 · Gouvernance des Données (Snowflake Horizon)

- Mise en place de Dynamic Data Masking (masquage de colonnes) et Row Access Policies (sécurité à la ligne)
- Tagging d'objets pour le suivi de la sensibilité des données

5. Jour 2 · Ingestion de Données & Connectivité

- Configuration des Stages internes et externes — ingestion Batch via COPY INTO
- Ingestion continue via Snowpipe (et introduction à Snowpipe Streaming)
- Hands-on Lab : Implémentation d'une politique de masquage de données RGPD sur un rôle de développeur et configuration d'un pipeline d'ingestion automatisé

6. Jour 3 · Manipulation de Données Semi-Structurées

- Objectif du Jour 3 : manipuler les formats complexes et automatiser les pipelines de données (ELT / CDC).
- Stockage et requêtage natif du JSON, XML et Parquet via le type VARIANT
- Fonctions d'extraction : notation par point, FLATTEN, LATERAL FLATTEN

7. Jour 3 · Pipelines de données (CDC) & Automatisation

- Capture des changements de données via les Streams
- Orchestration native : création de Tasks, gestion des dépendances (DAGs) et stratégies de Scheduling (CRON / Serverless)
- Utilisation des Dynamic Tables pour simplifier les pipelines déclaratifs
- Hands-on Lab : Création d'un pipeline d'ingestion de logs JSON bruts, traitement automatique des deltas via un Stream, et exécution planifiée via un arbre de Tasks (DAG)

8. Jour 4 · FinOps & Gouvernance des Coûts

- Objectif du Jour 4 : maîtriser les coûts, assurer la reprise après sinistre, utiliser l'IA et valider les pièges de la certification.
- Facturation (Compute, Storage, Cloud Services) — suivi via ACCOUNT_USAGE
- Mise en place de garde-fous : Resource Monitors (niveaux Account et Warehouse)

9. Jour 4 · Observabilité & Alerting

- Configuration du framework d'Alerting natif (CREATE ALERT) pour notifier l'équipe par email / webhook (échec d'une Task, pic de coût…)

10. Jour 4 · Résilience & Cycle de vie de la donnée

- Fonctionnement précis du Time Travel (0 à 90 jours selon l'édition) et du Fail-safe (7 jours non configurables)
- Utilisation du Zero-Copy Cloning pour le clonage instantané d'environnements de Dev / Test

11. Jour 4 · Data Marketplace & IA Cortex

- Partage de données et utilisation du Snowflake Marketplace
- Introduction et cas d'usage des fonctions IA intégrées (Snowflake Cortex LLM & Machine Learning)
- Hands-on Lab : Restauration d'une table supprimée via Time Travel, création d'une alerte FinOps sur dépassement de crédit, et appel d'une fonction d'analyse de sentiment via Snowflake Cortex

Apache Hop — Intégration de données visuelle

Prix : 2 100 € HT

Durée : 2 jours (14 heures)

Public cible : Développeurs, ingénieurs de données.

Prérequis : Maîtriser les bases du SQL. Une compréhension de base des concepts d'intégration de données serait un atout. Prérequis d'installation : avoir un compte Google avec un accès au Dataflow de Google Cloud service (crédit de 300 $ offert à la première inscription) ; être admin de son poste, avec une DB active (type PostgreSQL).

Objectifs pédagogiques

À l'issue de la formation, les participants seront capables de :

- Maîtriser la conception de flux d'intégration visuelle
- Exploiter les fonctionnalités avancées d'Apache Hop
- Créer des workflows d'intégration complexes

Méthodes et moyens pédagogiques : Stage pratique : 60 % pratique, 40 % théorie. Support de formation distribué au format numérique à tous les participants. Le cours alterne les apports théoriques du formateur, soutenus par des exemples, des séances de réflexion et du travail en groupe. La formation se base sur la dernière version en date, Apache Hop 2.11.

Modalités d'évaluation

Positionnement à l'entrée par un questionnaire d'auto-évaluation (conforme aux critères qualité Qualiopi). En fin de session, un questionnaire à choix multiples permet de vérifier l'acquisition des compétences. Une attestation est remise à chaque stagiaire ayant suivi la totalité de la formation.

Accessibilité : Formation accessible aux personnes en situation de handicap, contactez notre référent dédié.

Programme

1. Introduction à Apache Hop

- Présentation d'Apache Hop et son rôle dans le processus d'intégration
- Comprendre les fondamentaux de l'intégration de données
- Installation et configuration initiale — premiers pas avec Apache Hop
- Exploration de l'interface utilisateur graphique (GUI)

PRATIQUE Création de flux de données simples

2. Les concepts clés d'Apache Hop

- Apprentissage des concepts de base
- Les transformations
- Les étapes
- Les flux de données
- Exploration approfondie des composants de conception de flux
- Utilisation des métadonnées
- Exemples pratiques de manipulation de données avec Apache Hop
- Développement de compétences avancées dans la conception de flux complexes
- Création de flux réutilisables avec Apache Hop

3. Intégration de bases de données avec Apache Hop

- Connexion et extraction de données
- Chargement de données dans des bases de données
- Optimisation des performances
- Ateliers pratiques sur des cas d'utilisation courants
- Gestion des erreurs et récupération dans les flux de base de données
- Utilisation de variables pour une flexibilité accrue

4. Gestion des flux de données

- Introduction aux concepts de gestion des flux de données
- Planification et automatisation des flux avec Apache Hop
- Surveillance et débogage des flux en temps réel
- Gérer efficacement les flux de données
- Création de rapports et de tableaux de bord pour la surveillance
- Gestion des versions des flux

5. Intégration avec d'autres technologies

- Connexion à des services cloud
- Utilisation d'API REST
- Intégration de technologies émergentes dans les flux de données
- Cas d'étude pratique sur l'intégration multi-technologique
- Configuration de connexions sécurisées
- Optimisation des flux

6. Bonnes pratiques et optimisation

- Bonnes pratiques de conception de flux d'intégration
- Optimisation des performances pour des opérations d'intégration rapides
- Sécurité des données dans les flux d'intégration
- Études de cas sur l'application de bonnes pratiques
- Utilisation avancée des options de performance d'Apache Hop

SQL Avancé : Requêtes Complexes et Optimisation

Prix : 1580 € H.T.

Durée : 2 jours

Public cible : Développeurs, data analysts, data engineers ou administrateurs de base de données. Utilisateurs réguliers de SQL souhaitant aller au-delà des requêtes classiques. Chefs de projet ou métiers manipulant des bases de données relationnelles complexes

Prérequis : Maîtrise des bases du langage SQL : SELECT, WHERE, GROUP BY, JOIN. Connaissance d'un SGBD relationnel (PostgreSQL, MySQL, Oracle, SQL Server, etc.). Expérience pratique recommandée sur au moins un outil de requêtage.

Objectifs pédagogiques

À l'issue de la formation, les participants seront capables de :

- Maîtriser les jointures complexes, sous-requêtes et expressions conditionnelles
- Utiliser les fonctions avancées de traitement de données (fenêtres, CTE, agrégats)
- Optimiser les requêtes SQL en analysant leur performance
- Manipuler efficacement des ensembles de données structurées pour des analyses avancées
- Appliquer les bonnes pratiques d'écriture, de lisibilité et de performance SQL

Méthodes et moyens pédagogiques : Travaux pratiques : formation alternant théorie et pratique.

Modalités d'évaluation

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

Programme

1. Rappels et bonnes pratiques

- Requêtes imbriquées, jointures internes et externes
- Clauses GROUP BY, HAVING, ORDER BY avancées
- Astuces pour écrire du SQL lisible et maintenable

2. Sous-requêtes et expressions conditionnelles

- Sous-requêtes corrélées vs non corrélées
- Utilisation avancée de CASE, COALESCE, NULLIF
- Requêtes avec EXISTS / NOT EXISTS

3. Fonctions d'agrégation et analytiques

- Fonctions d'agrégation conditionnelle
- Fonctions analytiques (fenêtres) : ROW_NUMBER(), RANK(), LAG(), LEAD(), etc.
- PARTITION BY, ORDER BY dans les fonctions de fenêtre
- Applications : classement, cumul, comparaison temporelle

4. Common Table Expressions (CTE)

- Introduction aux CTE (WITH) : lisibilité et modularité
- CTE récursives pour explorer des hiérarchies ou graphes
- Comparaison avec les sous-requêtes imbriquées classiques

5. Manipulation d'ensembles complexes

- UNION, INTERSECT, EXCEPT
- Traitement de données temporelles (différences de dates, regroupements par périodes)
- Requêtes dynamiques et pivotements de données (PIVOT / UNPIVOT ou alternatives)

6. Optimisation et performance

- Comprendre les plans d'exécution (EXPLAIN / ANALYZE)
- Index, statistiques, tri et filtrage efficace

- Détection des goulots d’étranglement et surcharge mémoire
- Bonnes pratiques : éviter les requêtes inefficaces, limiter la charge serveur

7. Atelier pratique

- Étude de cas à partir d’une base relationnelle métier (ex : ventes, RH, support, finance)
- Résolution de requêtes complexes à plusieurs étapes
- Comparaison de solutions et optimisation collaborative

SQL Analytique : Fonctions de Fenêtre et Traitement Avancé des Données

Prix : 1880 € H.T.

Durée : 2 jours

Public cible : Data analysts, data scientists, contrôleurs de gestion, business analysts. Développeurs ou ingénieurs souhaitant approfondir l'analyse de données avec SQL. Toute personne manipulant régulièrement des bases de données pour produire des indicateurs ou des reporting

Prérequis : Maîtrise des bases du SQL : SELECT, WHERE, GROUP BY, JOINS. Expérience pratique avec un outil de requêtage ou un SGBD relationnel (PostgreSQL, SQL Server, MySQL, Oracle...). Connaissances élémentaires en statistiques ou reporting métier appréciées

Objectifs pédagogiques

À l'issue de la formation, les participants seront capables de :

- Utiliser les fonctions analytiques (fenêtres) pour produire des analyses complexes
- Maîtriser les opérations de cumul, classement, variation, comparaison temporelle
- Segmenter dynamiquement les données avec PARTITION BY, ORDER BY, LAG, LEAD, etc.
- Appliquer SQL à des cas concrets d'analyse métier (KPI, churn, comportement utilisateur)
- Optimiser les requêtes analytiques pour les rendre lisibles et performantes

Méthodes et moyens pédagogiques : Travaux pratiques : formation alternant théorie et pratique.

Modalités d'évaluation

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

Programme

1. Rappels SQL pour l'analyse

- Clauses fondamentales : GROUP BY, HAVING, filtres multiples
- Jointures internes / externes pour enrichir les jeux de données
- Sous-requêtes vs Common Table Expressions (CTE)

2. Introduction aux fonctions de fenêtre (window functions)

- Structure : OVER(), PARTITION BY, ORDER BY
- Différences avec les agrégats classiques
- Cas d'usage : moyenne mobile, cumul, classement dans un segment

3. Fonctions analytiques courantes

- ROW_NUMBER(), RANK(), DENSE_RANK() : classement et détection de doublons
- LAG(), LEAD() : comparaison entre lignes consécutives
- FIRST_VALUE(), LAST_VALUE() : repérage de bornes dans une série
- NTILE(), PERCENT_RANK(), CUME_DIST() : segmentation en quantiles

4. Analyse temporelle et séries

- Calculs sur périodes glissantes : moyenne sur 7 jours, évolution mensuelle
- Études d'évolution entre deux dates (delta, variation, croissance)
- Fenêtres mobiles : RANGE vs ROWS
- Cas d'usage : churn, comportements utilisateurs, suivi de cohortes

5. Études de cas métiers

- Analyse du chiffre d'affaires cumulé par client et par mois
- Classement des meilleurs vendeurs par zone / période
- Détection de récurrence ou d'événements inhabituels
- Analyse de la rétention ou de la fréquence d'utilisation

6. Optimisation et lisibilité

- Structuration des requêtes analytiques complexes
- Utilisation de CTE pour segmenter les étapes d'analyse
- Lecture d'un plan d'exécution pour diagnostiquer une requête lente
- Bonnes pratiques de rédaction pour les dashboards automatisés

Apache Spark – Niveau Avancé : Optimisation, Streaming et Traitements Scalables

Prix : 2100 € H.T.

Durée : 3 jours

Public cible : Data engineers, développeurs ou data scientists utilisant Spark en production. Architectes ou tech leads responsables de pipelines de données distribués. Équipes travaillant sur des projets Big Data avec des contraintes de performance, de volume ou de temps réel

Prérequis : Pratique régulière de Spark (PySpark ou Scala). Maîtrise des DataFrames, transformations de base, Spark SQL. Connaissances en cluster, format de données, et environnement distribué

Objectifs pédagogiques

À l'issue de la formation, les participants seront capables de :

- Optimiser les performances de leurs jobs Spark : partitionnement, cache, joins, formats
- Structurer des pipelines complexes et robustes
- Gérer des traitements en streaming temps réel avec Spark Structured Streaming
- Implémenter des workflows de machine learning distribué avec Spark MLlib
- Déployer Spark efficacement dans un cluster (YARN, Kubernetes, Databricks…)

Méthodes et moyens pédagogiques : Travaux pratiques : formation alternant théorie et pratique.

Modalités d'évaluation

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

Programme

1. Architecture approfondie de Spark

- Analyse détaillée : DAG, stages, tasks, shuffles
- Fonctionnement du Catalyst Optimizer et du Tungsten Engine
- Paramètres critiques : mémoire, nombre d'exécuteurs, partitions, parallélisme

2. Optimisation des performances

- Partitionnement personnalisé (repartition, coalesce, partitionBy)
- Joins performants : broadcast, sort-merge, bucketed joins
- Cache/persist : usages et limites
- Choix des formats de données : Parquet, ORC, Delta
- Utilisation de explain() et Spark UI pour profiler un job

3. Gestion avancée des données

- Manipulation de fichiers volumineux / structurés
- Lecture/écriture optimisées : compression, schema evolution, merge
- Traitement de données hiérarchiques (nested structures, arrays)

4. Spark Structured Streaming avancé

- Fenêtrage temporel, agrégations par fenêtre
- Gestion de l'état (stateful operations)
- Stratégies de déclenchement (triggering)
- Tolérance aux pannes : checkpointing, exactly-once
- Intégration Kafka ⟷Spark : lecture, écriture, offset management

5. Machine Learning distribué avec MLlib

- Pipelines de traitement : VectorAssembler, StringIndexer, StandardScaler
- Modèles supervisés : régression, classification, arbres de décision
- Évaluation, grid search, cross-validation en mode distribué
- Sauvegarde/rechargement de modèles Spark MLlib

- Différences Spark MLlib vs scikit-learn / TensorFlow

6. Structuration de projets Spark

- Bonnes pratiques de structuration de code (fonctions, scripts modulaires)
- Déploiement en production : packaging avec spark-submit, intégration CI/CD
- Variables d'environnement, configuration dynamique
- Exécution sur Databricks, EMR, Azure Synapse ou Kubernetes

7. Atelier final

- Conception et exécution d'un pipeline complet :
- Ingestion Kafka ou fichiers
- Traitement batch + streaming
- Pipeline ML distribué
- Export dans stockage optimisé
- Profiling et tuning

Apache Kafka – Niveau Avancé : Administration, Optimisation et Streaming Temps Réel

Prix : 2180 € H.T.

Durée : 3 jours

Public cible : Administrateurs, data engineers, développeurs et architectes techniques utilisant Kafka en production. Équipes responsables de la scalabilité, sécurité, et haute disponibilité des clusters Kafka. Professionnels souhaitant concevoir et optimiser des architectures de streaming distribuées complexes

Prérequis : Bonne connaissance des concepts de base Kafka (topics, partitions, producteurs, consommateurs). Expérience pratique avec Kafka en environnement de développement ou production. Notions d'administration système et de réseaux recommandées

Objectifs pédagogiques

À l'issue de la formation, les participants seront capables de :

- Installer et configurer un cluster Kafka scalable et résilient
- Gérer la sécurité : authentification, autorisation, chiffrement
- Optimiser la production et la consommation pour des performances maximales
- Mettre en œuvre Kafka Connect pour l'intégration avec d'autres systèmes
- Développer des applications streaming avancées avec Kafka Streams
- Surveiller et dépanner un cluster Kafka en production

Méthodes et moyens pédagogiques : Travaux pratiques : formation alternant théorie et pratique.

Modalités d'évaluation

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

Programme

1. Architecture avancée et administration de cluster

- Installation multi-nœuds, configuration de brokers, Zookeeper / KRaft
- Réplication, ISR (In-Sync Replicas), stratégies de partitionnement
- Gestion des logs, configuration de la rétention, nettoyage des topics
- Gestion des quotas et contrôle du débit

2. Sécurité dans Kafka

- Authentification : SSL, SASL (PLAIN, SCRAM, Kerberos)
- Autorisation : ACLs (Access Control Lists)
- Chiffrement des données en transit et au repos
- Meilleures pratiques et audit de sécurité

3. Performance et optimisation

- Paramètres de production et consommation (batch.size, linger.ms, fetch.min.bytes)
- Compression des messages (gzip, snappy, lz4)
- Gestion des délais, backpressure, et gestion des erreurs
- Tuning du GC, configuration mémoire JVM pour brokers

4. Kafka Connect et intégrations

- Présentation de Kafka Connect et connecteurs disponibles (bases, fichiers, cloud)
- Configuration et gestion de connecteurs source et sink
- Transformation simple des données dans les pipelines
- Déploiement et monitoring des connecteurs

5. Kafka Streams avancé

- Concepts clés : topologies, KStream, KTable, GlobalKTable
- Opérations stateless vs stateful

- Fenêtrage, agrégations, joins, transformations personnalisées
- Tolérance aux pannes, sauvegarde d'état (state stores)
- Monitoring et debugging d'applications Kafka Streams

6. Monitoring et maintenance

- Outils de monitoring : Kafka Manager, Confluent Control Center, Prometheus, Grafana
- Logs, métriques JMX, alerting
- Diagnostic des problèmes : lag, déséquilibre des partitions, ISR, erreurs réseau
- Sauvegarde, migration, mises à jour du cluster

7. Cas pratiques et ateliers

- Mise en place d'un cluster Kafka multi-nœuds sécurisé
- Déploiement de connecteurs source/sink avec transformation des données
- Développement d'une application Kafka Streams avancée
- Analyse et résolution de problèmes de performance et de stabilité

Apache Hadoop – Niveau Avancé : Administration, Optimisation et Écosystème

Prix : 2200 € H.T.

Durée : 3 jours

Public cible : Administrateurs systèmes et data engineers responsables d'un cluster Hadoop. Architectes Big Data et équipes techniques en charge de projets Hadoop complexes. Développeurs avancés souhaitant optimiser et gérer des traitements distribués à grande échelle

Prérequis : Maîtrise des concepts fondamentaux Hadoop et Big Data. Expérience pratique avec HDFS et MapReduce. Connaissances en administration Linux et en réseaux

Objectifs pédagogiques

À l'issue de la formation, les participants seront capables de :

- Installer, configurer et administrer un cluster Hadoop multi-nœuds
- Optimiser les performances de HDFS et des jobs MapReduce
- Maîtriser YARN et la gestion des ressources dans Hadoop
- Utiliser et administrer les composants de l'écosystème Hadoop (Hive, HBase, Oozie, etc.)
- Assurer la sécurité, la haute disponibilité et la résilience du cluster

Méthodes et moyens pédagogiques : Travaux pratiques : formation alternant théorie et pratique.

Modalités d'évaluation

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

Programme

1. Architecture avancée de Hadoop

- Détails de HDFS : NameNode HA, Federation
- Gestion des DataNodes, balancement de charge
- Fonctionnement de YARN : ResourceManager, NodeManager, Scheduler
- Cycle de vie des applications et gestion des ressources

2. Installation et administration d'un cluster Hadoop

- Installation multi-nœuds (distribution Apache ou Cloudera/Hortonworks)
- Configuration avancée des fichiers XML (core-site, hdfs-site, yarn-site)
- Démarrage, arrêt, mise à jour du cluster
- Surveillance via JMX, logs, outils (Ambari, Cloudera Manager)

3. Optimisation des performances

- Gestion des fichiers et blocs dans HDFS, taille de bloc optimale
- Compression et formats de fichiers adaptés (Parquet, ORC)
- Optimisation des jobs MapReduce et des ressources YARN
- Tuning du Garbage Collector, JVM, et paramètres réseau

4. Sécurité et gestion des accès

- Authentification Kerberos et intégration LDAP
- Contrôle d'accès basé sur les rôles (RBAC) et permissions HDFS
- Chiffrement des données au repos et en transit
- Audit et traçabilité des accès

5. Écosystème Hadoop avancé

- Hive : optimisation des requêtes, partitionnement, bucketing
- HBase : architecture, administration, scalabilité
- Oozie : orchestration de workflows Hadoop
- Autres outils : Sqoop, Flume, ZooKeeper

6. Haute disponibilité et résilience

- Configuration NameNode HA et failover automatique
- Backup et restauration des métadonnées
- Gestion des pannes DataNodes et récupération
- Stratégies de sauvegarde et reprise d'activité

7. Atelier pratique

- Mise en place d'un cluster Hadoop multi-nœuds en mode pseudo-distribué avancé
- Configuration et tuning d'un job MapReduce complexe
- Mise en œuvre de la sécurité Kerberos et gestion des utilisateurs
- Orchestration d'un workflow Hive/Oozie

Apache Hop – Niveau Débutant : Orchestration Visuelle de Flux de Données

Prix : 1780 € H.T.

Durée : 2 jours

Public cible : Data engineers, développeurs ETL/ELT et intégrateurs de données débutants sur Apache Hop. Chefs de projet ou architectes souhaitant découvrir l'orchestration visuelle de pipelines. Toute équipe technique comparant Hop à d'autres outils d'intégration (Talend, Pentaho, Airflow…)

Prérequis : Connaissances de base en SQL et en architecture ETL/ELT. Notions de manipulation de données (fichiers plats, JSON, bases relationnelles). Java JDK 8+ installé et environnement Hop Studio configuré

Objectifs pédagogiques

À l'issue de la formation, les participants seront capables de :

- Installer et configurer Apache Hop Studio
- Concevoir, tester et déployer des pipelines (transformations) et des workflows (orchestration)
- Utiliser efficacement les composant clés pour sources, cibles et traitements
- Gérer métadonnées, variables, paramètres et réutilisabilité
- Automatiser, superviser et optimiser des flux de données simples à complexes

Méthodes et moyens pédagogiques : Travaux pratiques : formation alternant théorie et pratique.

Modalités d'évaluation

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

Programme

1. Introduction à Apache Hop

- Présentation d'Apache Hop et rôle dans l'intégration de données
- Fondamentaux de l'ETL/ELT et cas d'usage
- Installation, configuration initiale et premiers pas dans l'interface graphique (GUI)
- Création d'un premier pipeline simple

2. Concepts clés d'Apache Hop

- Transformations vs Workflows
- Étapes (steps) et flux de données (hops)
- Métadonnées partagées et gestion des repositories
- Exemples pratiques : manipulation de données, filtrage, enrichissement
- Création de pipelines réutilisables

3. Intégration de bases de données

- Connexion JDBC et extraction de données (Table Input, Table Output)
- Chargement dans une base cible et optimisation des performances
- Ateliers pratiques sur cas courants (batch insert/update)
- Gestion des erreurs, transactions et reprise sur incident
- Variables et paramètres pour pipelines dynamiques

4. Gestion et supervision des flux

- Planification et automatisation via Hop Server ou scheduler externe
- Débogage en temps réel, points d'arrêt, aperçu de lignes
- Création de rapports et tableaux de bord de supervision
- Versioning des pipelines et bonnes pratiques de lifecycle

5. Intégration avec d'autres technologies

- Connexion à services cloud (S3, Azure Blob, Google Storage)
- Appels à des API REST et traitement JSON/XML

- Exemples de pipelines multi-technologie (Kafka, Hadoop, Spark)
- Sécurisation des connexions (SSL, authentification)
- Optimisation et tuning des flux hétéroclites

6. Bonnes pratiques et optimisation

- Principes de conception de pipelines robustes et maintenables
- Astuces pour accélérer les traitements et réduire la latence
- Sécurité des données dans Hop : masquage, chiffrement, accès
- Études de cas : application des bonnes pratiques sur un projet réel
- Utilisation avancée des options de performance (parallelisme, clustering)

Elasticsearch – Niveau Avancé : Performance, Sécurité et Scalabilité

Prix : 2180 € H.T.

Durée : 3 jours

Public cible : Administrateurs, data engineers et architectes chargés de clusters Elasticsearch en production. Développeurs et DevOps souhaitant optimiser, sécuriser et faire évoluer un environnement Elastic. Équipes en charge de cas d'usage critiques : logs à fort volume, analytics temps réel, recherche d'entreprise

Prérequis : Maîtrise des fondamentaux d'Elasticsearch (indexation, requêtes DSL, agrégations). Expérience pratique avec un cluster mono- ou multi-nœuds. Connaissances de base en administration Linux et en réseaux

Objectifs pédagogiques

À l'issue de la formation, les participants seront capables de :

- Concevoir et dimensionner un cluster Elasticsearch scalable et résilient
- Optimiser les performances d'indexation et de recherche pour de gros volumes de données
- Mettre en place des pipelines d'ingestion robustes et sécurisés
- Assurer la haute disponibilité et la reprise après sinistre (HA/DR)
- Sécuriser l'accès et garantir la conformité (authentification, chiffrement, audit)
- Superviser et diagnostiquer finement l'état du cluster

Méthodes et moyens pédagogiques : Travaux pratiques : formation alternant théorie et pratique.

Modalités d'évaluation

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

Programme

1. Architecture distribuée et dimensionnement

- Topologie de cluster : nœuds master, data, ingest, coordinating
- Stratégies de sharding et de réplication selon le volume et la criticité
- Dimensionnement CPU, mémoire, stockage, I/O
- Migration et extension de cluster en production

2. Performance et tuning avancé

- Optimisation de l'indexation : bulk, refresh interval, merge policy
- Choix des formats de stockage : compression, doc_values, store
- Tuning des requêtes : caches, circuit breaker, préférences de nœuds
- Analyse de performance via _nodes/stats, _cluster/stats et Profiler API

3. Ingest Pipelines & Transformations

- Configuration d'Ingest Pipelines : processors (grok, date, geoip, script)
- Enrichissement des documents à l'ingestion
- Utilisation de Logstash et Beats en complément
- Gestion des erreurs et des documents DLQ (dead letter queue)

4. Sécurité et gouvernance

- Configuration X-Pack Security : utilisateurs, rôles, permissions
- Authentification : LDAP, Active Directory, SAML, OAuth
- Chiffrement TLS inter-nœud et client-nœud
- Audit Logging et conformité RGPD

5. Haute Disponibilité et Résilience

- Backup et snapshots automatisés avec Snapshot API
- Stratégies de restauration et PRA/PCA
- Multi-cluster et cross-cluster replication (CCR)

- Maintenance sans interruption : rolling upgrade, allocation filtering

6. Scripting, Painless et fonctions avancées

- Écriture de scripts Painless pour scoring et transformations
- Scripted fields et runtime fields
- Utilisation de Stored Scripts et Script Templates
- Cas pratiques : calculs dynamiques, filtres géospatiaux, déduplication

7. Monitoring et alerting

- Intégration avec Elastic Stack Monitoring, Prometheus et Grafana
- Configuration de métriques clés et seuils d'alerte
- Visualisation des indicateurs de performance et de santé du cluster
- Analyse des logs de cluster et détection d'anomalies

8. Atelier pratique

- Mise en place d'un cluster multi-nœuds avec HA et sécurité
- Tuning d'un index volumineux et optimisation de requêtes réelles
- Création d'un ingest pipeline pour un cas métier (logs ou analytics)
- Simulation de sinistre et restauration via snapshots

Power BI – Niveau Débutant

Prix : 1580 € H.T.

Durée : 2 jours

Public cible : Analystes métier, responsables reporting, consultants ou toute personne souhaitant créer rapidement des tableaux de bord interactifs. Développeurs et data engineers souhaitant intégrer Power BI dans leur flux de données. Décideurs et managers souhaitant comprendre les possibilités de la plateforme

Prérequis : Connaissances de base en Excel (tableaux, formules simples, graphiques). Notions élémentaires de bases de données relationnelles et SQL. Ordinateur avec Power BI Desktop installé

Objectifs pédagogiques

À l'issue de la formation, les participants seront capables de :

- Se connecter à différentes sources de données (fichiers, bases, services cloud)
- Préparer et transformer les données avec Power Query
- Modéliser les données (relations, hiérarchies, mesures de base)
- Créer des visualisations interactives et construire un rapport unifié
- Publier, partager et rafraîchir un tableau de bord sur Power BI Service

Méthodes et moyens pédagogiques : Travaux pratiques : formation alternant théorie et pratique.

Modalités d'évaluation

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

Programme

1. Introduction à Power BI et son écosystème

- Présentation de la suite Power Platform (Desktop, Service, Mobile, Report Server)
- Cas d'usage BI self-service vs BI traditionnelle
- Architecture et flux de travail : data-connect →model →report →publish

2. Connexion et préparation des données (Power Query)

- Connexion à des sources variées : Excel, CSV, SQL Server, SharePoint, APIs
- Transformation : filtres, colonnes calculées, fusion, appariement, pivot/unpivot
- M-language : aperçu des formules et étapes de requête
- Bonnes pratiques de nettoyage et documentation des requêtes

3. Modélisation de données

- Concepts de modèle en étoile : tables factuelles, tables de dimensions
- Définition des relations et cardinalités
- Création de hiérarchies et de tables de dates
- Introduction au DAX : mesures vs colonnes calculées

4. Visualisations et rapports

- Galerie de visuels : tables, matrices, graphiques barres, lignes, cartes, KPI
- Personnalisation : filtres, slicers, bookmarks, boutons d'action
- Conception de pages interactives et navigation entre vues
- Accessibilité et storytelling : mise en page, couleurs, bonnes pratiques UX

5. Power BI Service et collaboration

- Publication depuis Desktop vers Power BI Service
- Création d'espaces de travail, applications et apps Power BI
- Partage, permissions, et gestion des accès
- Configuration du rafraîchissement des jeux de données (gateways)

6. Introduction aux fonctionnalités avancées

- Q&A (requêtes en langage naturel)
- Alertes et abonnements aux rapports
- Introduction aux visuels personnalisés du marketplace
- Aperçu de Power BI Mobile

7. Atelier pratique

- Construction d'un tableau de bord de reporting métier (ventes, finance, RH…)
- Import, transformation, modélisation et mise en forme finale
- Publication et configuration du rafraîchissement automatique

Le Sur-Mesure

Avec notre équipe pédagogique, nos formateurs experts et nos consultants dédiés, nous co-construisons des formations personnalisées pour développer l'engagement et la performance de vos collaborateurs.

La conception de formations sur mesure

Votre consultant Skillia organise un entretien de cadrage avec l'expert-formateur pressenti afin de définir précisément les objectifs de la formation et d'évaluer les besoins de préparation. Spécialiste du domaine concerné, l'expert analyse les spécificités du projet, le profil des participants et élabore, avec vous, un cahier des charges détaillé garantissant la pertinence du dispositif.

Des solutions e-learning personnalisées

Vous souhaitez une formation 100 % sur mesure, composée de modules e-learning et de vidéos adaptées à votre métier ?

Skillia met à votre disposition son savoir-faire et ses outils de production pour concevoir des formations digitales performantes, alliant interactivité, efficacité et accompagnement personnalisé par des tuteurs experts.

La gestion de projets de grande envergure

Qu'il s'agisse de former plusieurs équipes, sur différents sites ou dans plusieurs langues (français, anglais…), Skillia dispose de l'expertise nécessaire pour piloter l'ensemble du projet : audit, conception de programmes, construction de parcours, ainsi que coordination technique et logistique.

SKILLIA

Le spécialiste de la formation professionnelle