

Formation Apache NiFi, Automatisation et orchestration des flux de données

Présentation

Cette formation de 3 jours permet de maîtriser Apache NiFi, une solution open source de gestion et d'automatisation de flux de données (Dataflow). Grâce à son interface visuelle et sa flexibilité, NiFi permet de collecter, transformer, router et monitorer des données en temps réel ou différé. La formation met l'accent sur les bonnes pratiques, la sécurité, la scalabilité et l'intégration dans des architectures Big Data ou Cloud.

Durée : 21,00 heures (3 jours)

Tarif INTRA : Nous consulter

Objectifs de la formation

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

- Comprendre le rôle et l'architecture d'Apache NiFi
- Concevoir et configurer des flux de données automatisés (Dataflows)
- Manipuler les processeurs, contrôleurs, filtres et routes
- Gérer les erreurs, les files d'attente, les priorités et les performances
- Sécuriser, versionner et superviser les flux
- Intégrer NiFi dans un écosystème data (Kafka, HDFS, DB, API REST, Cloud...)

Prérequis

- Connaissances générales en traitement de données (ETL, API, formats comme JSON, XML, CSV)
- Notions en réseau, HTTP et bases de données
- Connaissance d'un environnement Linux recommandée

Public

- Développeurs ou ingénieurs Data
- Architectes SI ou Big Data
- Administrateurs de flux ETL / ELT



- Ingénieurs DevOps / Plateforme
- Toute personne en charge de la gestion des flux de données

Programme de la formation

Jour 1 : Introduction et prise en main de NiFi

Module 1 : Présentation d'Apache NiFi

- Origine (NSA → Apache), cas d'usage typiques
- Architecture générale (FlowFile, FlowFile Processor, Repository, etc.)
- Installation de NiFi (mode standalone)

Module 2 : Prise en main de l'interface graphique

Module 3 : Création de premiers flux simples

- Collecte de fichiers locaux (GetFile, ListFile, FetchFile)
- Transformation simple (ReplaceText, UpdateAttribute)
- Routage conditionnel (RouteOnAttribute, RouteOnContent)

Module 4 : Monitoring et files d'attente

- Construire un pipeline qui lit un fichier CSV, transforme des colonnes, et route les lignes selon une règle métier.

Jour 2 : Connecteurs, transformation et contrôle des flux

Module 5 : Connecteurs de sources et cibles

- Bases de données relationnelles (JDBC, QueryDatabaseTable)
- Web API (InvokeHTTP, REST JSON)
- Kafka, MQTT, HDFS, FTP, Email

Module 6 : Manipulation des formats de données

- JSON, XML, CSV avec ConvertRecord, EvaluateJsonPath, JoltTransformJSON
- Utilisation des Record Processors et schémas Avro

- Gestion des erreurs, rejets, timeout
- Traitement par lot, priorisation et parallélisme
- Variables, templates et paramètres de configuration
- Introduction à NiFi Registry (versionnement des flux)

TP

- Consommer une API REST, extraire des données JSON, les stocker en base de données avec logs d'erreurs

Jour 3 : Sécurité, supervision et intégration avancée**Module 7 : Supervision, alertes et audit**

- Logs, provenance, tracking
- Performance des processeurs, files d'attente

Module 8 : Sécurité

- HTTPS, utilisateurs, rôles et autorisations
- Chiffrement de bout en bout, configuration sécurisée

Module 9 : Déploiement avancé

- Mode cluster / haute disponibilité
- Intégration avec Apache Kafka, Hadoop, Elasticsearch, Cloud (AWS S3, Azure Blob)
- Bonnes pratiques et patterns d'architecture

TP final

- Créer un flux complet avec collecte de fichiers via SFTP → transformation JSON → API → base de données, avec supervision et gestion des erreurs

Organisation

Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes.

Riches de leur expérience sur le sujet, ils sauront accompagner vos collaborateurs dans leur montée en compétence.

Moyens pédagogiques et techniques

- Alternance d'apports théoriques et de mises en pratique contextualisées.
- Progression pédagogique structurée, de la compréhension à la mise en œuvre opérationnelle.
- Pédagogie immersive favorisant une mise en application immédiate des compétences.
- Formation interactive et participative s'appuyant sur des ateliers pratiques, études de cas et échanges collectifs (+ 60 % du temps de formation.)
- Séquences pédagogiques rythmées et évaluations formatives régulières pour renforcer l'ancrage des apprentissages.
- Supports de cours, ressources documentaires et références mis à disposition par le formateur

Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation**En amont de la formation** *(évaluation en ligne dans l'extranet apprenant)*

- Recueil des besoins et des attentes des participants.
- Auto-positionnement permettant d'évaluer le niveau initial des apprenants.

Tout au long de la formation

- Évaluations continues réalisées par le formateur travers des questions orales, exercices, QCM, études de cas et mises en situation pratiques.

À l'issue de la formation *(évaluations en ligne dans l'extranet apprenant)*

- Auto-positionnement afin de mesurer l'acquisition des compétences visées.
- Validation de l'acquisition des compétences par le formateur
- Questionnaire de satisfaction « à chaud » à l'issue de la formation

- Questionnaire de satisfaction "à froid", 45 jours après la formation afin d'évaluer l'ancrage des acquis et leur mise en application en situation professionnelle.

Accessibilité

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.