

Mis à jour le 08/04/2025

S'inscrire

Formation Apache NiFi : Automatisez vos flux de données

2 jours (14 heures)

Présentation

Notre formation Apache NiFi vous permettra de facilement traiter et distribuer des données de manière performante.

En effet, proche d'un ETL, Apache NiFi est un outil qui a prouvé son efficacité, par delà le fait qu'il est hautement configurable afin que les utilisateurs bénéficient d'une livraison garantie, d'une faible latence, d'un débit élevé, d'une priorisation dynamique et d'une fonctionnalité permettant de modifier les flux en cours d'exécution.

Naturellement orienté vers les composants, le paradigme de NiFi est le flow-based programming (FBP). Des messages pouvant atteindre plusieurs giga-octets peuvent ainsi être traités de façon automatisée au sein de votre "usine de données". Très productive, son interface intuitive permet de créer des flux de données par un simple glisser-déposer sans avoir besoin d'écrire la moindre ligne de code.

De plus, le programme est open-source et gratuit. Les documentations faisant référence à Apache NiFi sont nombreuses. Les utilisateurs du logiciel profitent de reportings personnalisés, d'une interface transparente pour le traitement des données avec la technologie DAG et d'un prototypage rapide.

NiFi supporte aussi les protocoles sécurisés comme SSL, HTTPS et SSH. Notre formation Apache NiFi se basera sur la dernière version du logiciel en date à savoir [Apache NiFi 2.3](#).

Objectifs

- Savoir installer et configurer Apache NiFi
- Comprendre l'architecture et les fondements du traitement des données avec Apache NiFi
- Maîtriser la gestion de flux de données
- Intégrer Kafka avec NiFi

Public visé

Développeurs, Professionnels du Big Data, Architectes, Administrateurs systèmes, DevOps

Pré-requis

- Connaissances de base des systèmes Unix
- Connaissances de base en Java

Programme de la formation Apache NiFi

Introduction à NiFi et à l'Écosystème de Données

- Vue d'ensemble de l'écosystème Big Data et rôle de NiFi
- Cas d'usage et positionnement d'Apache NiFi
- Principes fondamentaux du traitement des flux de données (Flow-Based Programming)
- Introduction aux concepts clés de NiFi : FlowFile, Processeur, Connexion, Attributs

Installation, Configuration et Prise en Main de NiFi

- Procédure détaillée d'installation et de configuration initiale de NiFi (Standalone, basée sur Unix)
- Navigation et utilisation de l'interface utilisateur (UI) de NiFi
- Comprendre les états des processeurs et des connexions
- Ajouter, connecter et configurer ses premiers processeurs (GenerateFlowFile, LogAttribute)
- Introduction essentielle au NiFi Expression Language (EL)
- Introduction aux Controller Services : Rôle et gestion
- Comprendre les Execution Engines (Traditional vs Stateless - Conceptuel)

Manipulation de Base des Données et des Flux

- Travailler avec les attributs : UpdateAttribute, ExtractText
- Gestion des fichiers : GetFile, PutFile, ListFile
- Routage de base des FlowFiles : RouteOnAttribute
- Introduction aux Processeurs orientés "Record"
- Configuration des Record Readers/Writers (CSV, JSON)
- Traitement de base des enregistrements : UpdateRecord, SplitRecord
- Traitement spécifique JSON : EvaluateJsonPath, SplitJson
- Gestion simple des erreurs et relation 'failure'

Intégration Approfondie avec les Systèmes Externes (Focus Kafka & API)

- Rappel des concepts clés d'Apache Kafka (Topic, Partition, Broker, Consumer Group)

- Intégration NiFi-Kafka (Consommation) : Configurer et utiliser ConsumeKafkaRecord
- Intégration NiFi-Kafka (Production) : Configurer et utiliser PublishKafkaRecord
- Bonnes pratiques pour l'intégration NiFi-Kafka (Clés, Group ID, Formats)
- Interaction avec les API REST (Client) : InvokeHTTP
- Exposer une API simple avec NiFi : ListenHTTP, HandleHttpRequest/HandleHttpResponse
- Interaction avec les bases de données (SQL) : QueryDatabaseTableRecord, PutDatabaseRecord

Administration, Surveillance et Optimisation des Flux NiFi

- Surveillance de l'état de NiFi via l'UI (Dashboard, Status, Bulletins)
- Analyse de la Provenance des Données (Data Provenance) : Suivi et investigation
- Techniques de débogage avancées et gestion des erreurs (Files d'attente, stratégies de retry)
- Gestion des utilisateurs et des accès : Politiques de base (Users, Groups, Politiques globales)
- Utilisation des Templates pour la réutilisation des flux
- Introduction aux Parameter Contexts pour la gestion des configurations
- Optimisation des performances : Back Pressure, Yield Duration, Concurrent Tasks
- Introduction aux Reporting Tasks pour le monitoring externe
- Introduction aux Flow Analysis Rules pour la validation des flux

Gestion Avancée, Sécurité et Écosystème NiFi

- Introduction à NiFi Registry : Rôle et architecture
- Gestion des versions des flux avec NiFi Registry (Enregistrer, Importer, Déployer)
- Intégration avec un Schema Registry (ex: Confluent) pour la gestion centralisée des schémas
- Gestion de la sécurité : Concepts clés (Authentification, Autorisation)
- Gestion des certificats SSL/TLS pour la communication sécurisée (Controller Service StandardSSLContextService)
- Introduction à la NiFi Command Line Interface (CLI) pour l'administration et l'automatisation
- Gestion du Clustering dans NiFi : Concepts et avantages (Haute Disponibilité, Scalabilité)
- (Optionnel) Introduction à MiNiFi pour la collecte de données à la source (Edge/IoT)

Pour aller plus loin

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.