



# PRÉSENTATION DE NIFI

# QU'EST-CE QUE NIFI ?

Apache NiFi est un logiciel de gestion de flux de données automatisé et programmable, conçu pour automatiser le flux de données entre les systèmes. Il fournit une interface utilisateur basée sur le Web pour la création, le monitoring, et le contrôle des flux de données.

# HISTORIQUE DE NIFI

NiFi a été développé par la NSA (National Security Agency) des États-Unis et a été rendu open source en 2014 sous l'égide de l'Apache Software Foundation. Il a été conçu pour être hautement configurable et extensible tout en maintenant la facilité d'utilisation.

# AVANTAGES DE NIFI

- **Facilité d'utilisation :** Interface graphique intuitive pour la gestion des flux.
- **Robustesse :** Supporte le backpressure et garantit la livraison des données.
- **Extensibilité :** Permet l'intégration avec de nombreux systèmes via des processeurs personnalisables.
- **Sécurité :** Supporte l'authentification, l'autorisation, et le chiffrement des données.

# CAS D'UTILISATION DE NIFI

- **Intégration de données** : Collecte et distribution de données entre différents systèmes.
- **Traitement en temps réel** : Permet le traitement des flux de données en temps réel.
- **Automatisation des flux de données** : Automatise les processus de données répétitifs sans codage manuel.
- **IoT (Internet des Objets)** : Gestion des données provenant de dispositifs IoT.

# ARCHITECTURE DE NIFI

# COMPOSANTS PRINCIPAUX DE NIFI

- **Processor:** Effectue des tâches sur les flux de données.
- **Connection:** Lien entre les processeurs pour le transfert des données.
- **Flow Controller:** Gère le flux global des informations.
- **FlowFile Repository:** Stocke les métadonnées des FlowFiles.
- **Content Repository:** Stocke le contenu réel des données.
- **Provenance Repository:** Enregistre les événements qui affectent chaque FlowFile.

# FLUX DE DONNÉES DANS NIFI

1. **Réception des données** : Les données entrent dans Nifi via des processeurs spécifiques.
2. **Traitement des données** : Les processeurs modifient ou analysent les données.
3. **Routage des données** : Les connexions dirigent les données vers les processeurs suivants.
4. **Sortie des données** : Les données sont envoyées à des destinations externes ou stockées.

# CLUSTERISATION ET HAUTE DISPONIBILITÉ

- **Clusterisation** : Plusieurs instances de Nifi travaillent ensemble pour une meilleure performance et fiabilité.
- **ZooKeeper** : Utilisé pour la gestion de la configuration et la coordination des nœuds.
- **RéPLICATION DES DONNÉES** : Assure la synchronisation des données entre les nœuds.
- **Failover** : En cas de défaillance d'un nœud, un autre prend le relais.

# SÉCURITÉ DANS L'ARCHITECTURE NIFI

- **Authentification** : Supporte LDAP, Kerberos pour la vérification des identités.
- **Autorisations** : Contrôle d'accès basé sur les rôles pour les ressources et les actions.
- **Chiffrement** : Chiffrement des données en transit et au repos.
- **Journalisation de la provenance** : Trace toutes les modifications apportées aux données pour l'audit.

# INSTALLATION DE NIFI

# PRÉREQUIS POUR L'INSTALLATION DE NIFI

- Système d'exploitation : Windows, Linux, ou MacOS
- Java 8 ou 11 (Java Runtime Environment ou Java Development Kit)
- Mémoire RAM : Minimum 4GB recommandé
- Espace disque : Minimum 200MB libre pour l'installation
- Accès réseau pour télécharger les fichiers d'installation

# TÉLÉCHARGEMENT DE NIFI

1. Accédez au site officiel Apache Nifi : <https://nifi.apache.org/>
2. Cliquez sur "Download" dans la barre de menu
3. Sélectionnez la version souhaitée
4. Téléchargez le fichier .zip ou .tar.gz correspondant à votre système d'exploitation

# INSTALLATION SUR WINDOWS

1. Extraire le fichier téléchargé (format .zip)
2. Ouvrir le dossier extrait
3. Lancer `run-nifi.bat` pour démarrer Nifi
4. Nifi est accessible via le navigateur à l'adresse `http://localhost:8080/nifi`

# INSTALLATION SUR LINUX

1. Extraire le fichier téléchargé (format `.tar.gz`)
2. Ouvrir un terminal dans le dossier extrait
3. Exécuter la commande: `./bin/nifi.sh start`
4. Nifi est accessible via le navigateur à l'adresse `http://localhost:8080/nifi`

# CONFIGURATION INITIALE DE NIFI

- Configuration de la mémoire : Modifier `conf/bootstrap.conf`
  - `java.arg.2=-Xms512m` : Mémoire initiale
  - `java.arg.3=-Xmx4096m` : Mémoire maximale
- Configuration des logs : Modifier `conf/logback.xml`
- Configuration des ports et interfaces : Modifier `conf/nifi.properties`

# VÉRIFICATION DE L'INSTALLATION

- Ouvrir un navigateur web et entrer l'adresse : <http://localhost:8080/nifi>
- Vérifier que l'interface utilisateur de Nifi s'affiche correctement
- Consulter les logs dans `logs/nifi-app.log` pour d'éventuelles erreurs

# INTERFACE UTILISATEUR DE NIFI

# PRÉSENTATION DE L'INTERFACE UTILISATEUR

L'interface utilisateur de NiFi est conçue pour être intuitive et facile à utiliser. Elle permet de visualiser et de gérer les flux de données en temps réel. L'interface est composée de divers éléments tels que le canevas, les onglets de menu, et les barres d'outils.

# NAVIGATION DANS L'INTERFACE

Pour naviguer dans l'interface de NiFi :

- Utilisez le **canévas** pour ajouter ou modifier des composants de flux.
- Accédez aux différents **onglets** (Summary, Counter, etc.) pour voir les détails des flux.
- Utilisez la barre de **recherche** pour trouver des composants spécifiques.

# PERSONNALISATION DE L'AFFICHAGE

Vous pouvez personnaliser l'affichage dans NiFi par :

- **Zoom avant/arrière** sur le canevas pour ajuster la vue.
- **Organisation des composants** pour une meilleure lisibilité.
- **Modification des couleurs** et des **étiquettes** pour différencier les composants.

# UTILISATION DES ONGLETS PRINCIPAUX

Les onglets principaux de NiFi incluent :

- **Operate** : pour gérer les composants et les connexions.
- **Summary** : pour obtenir un résumé des flux et des performances.
- **Counters** : pour suivre les métriques spécifiques.
- **History** : pour voir l'historique des modifications des flux.

# GESTION DES AUTORISATIONS UTILISATEUR

NiFi permet de gérer les autorisations des utilisateurs via :

- **Policies** : définition des droits d'accès pour les utilisateurs ou groupes.
- **Users/Groups** : gestion des utilisateurs et des groupes.
- **Access Policies** : contrôle des permissions sur les composants et les données.

# CONCEPTS DE BASE DES FLUX DE DONNÉES

# DÉFINITION DES FLUX DE DONNÉES

Les flux de données dans NiFi représentent le mouvement et la transformation des données d'une source à une destination. Ils permettent de collecter, transformer, et distribuer des données en temps réel.

# COMPOSANTS D'UN FLUX DE DONNÉES

- **Processors** : Exécutent des tâches spécifiques sur les données.
- **Connections** : Liens entre les processors, permettant le transfert de données.
- **Flow Controller** : Gère l'ordonnancement et l'allocation des ressources.
- **FlowFile** : Représente une donnée en mouvement à travers NiFi.

# CYCLE DE VIE D'UN FLUX DE DONNÉES

1. **Création** : Début avec la création d'un FlowFile.
2. **Traitement** : Passage à travers divers processors pour transformation.
3. **Routage** : Direction vers différents chemins selon les règles définies.
4. **Fin** : Soit archivage, soit transmission à une autre destination.

# GESTION DES ERREURS DANS LES FLUX DE DONNÉES

- **Back Pressure** : Mécanisme de régulation pour gérer la surcharge.
- **Data Provenance** : Suivi des données pour identifier et corriger les erreurs.
- **Bulletins** : Notifications en cas d'erreurs dans les processors.
- **Logging** : Enregistrement détaillé des événements pour le débogage.

# PROCESSORS ET LEURS CONFIGURATIONS

# TYPES DE PROCESSORS DANS NIFI

- **GetFile**: Lit des fichiers depuis un système de fichiers.
- **PutFile**: Écrit des fichiers vers un système de fichiers.
- **FetchFile**: Récupère des fichiers à partir d'une source externe.
- **ExecuteSQL**: Exécute des requêtes SQL sur une base de données.
- **InvokeHTTP**: Envoie des requêtes HTTP à des services externes.

# CONFIGURATION DE BASE D'UN PROCESSOR

- **Name:** Nom du processor.
- **Scheduling Period:** Fréquence d'exécution du processor.
- **Penalization Period:** Temps de pénalité si le processor échoue.
- **Yield Duration:** Temps d'attente avant de réessayer après un échec.
- **Run Schedule:** Définit combien de fois le processor doit être exécuté.

# PARAMÈTRES COMMUNS DES PROCESSORS

- **Input Directory/Path:** Chemin d'entrée pour les fichiers.
- **Output Directory/Path:** Chemin de sortie pour les fichiers.
- **File Filter:** Filtre pour sélectionner des fichiers spécifiques.
- **Success/Failure Relationship:** Définit comment gérer les succès ou les échecs.
- **Back Pressure Object Threshold:** Seuil pour la pression arrière.

# UTILISATION DES PROCESSORS POUR LE TRAITEMENT DES DONNÉES

- **Collecter des données:** Utilisation de `GetFile` pour lire des données.
- **Transformer des données:** Utilisation de `ExecuteSQL` pour transformer des données via SQL.
- **Distribuer des données:** Utilisation de `PutFile` pour distribuer des données transformées.
- **Interactions externes:** Utilisation de `InvokeHTTP` pour interagir avec des APIs externes.

# CONNECTIONS ENTRE PROCESSORS

# TYPES DE CONNECTIONS

NiFi permet de créer différents types de connections entre les processors pour gérer le flux de données :

- **Standard** : Transfert direct de données.
- **Self-looping** : Connection d'un processor à lui-même pour réutiliser des données.
- **Remote Process Group (RPG)** : Pour connecter à des instances NiFi distantes.

# CONFIGURATION DES CONNECTIONS

Pour configurer une connection dans NiFi :

1. Sélectionnez le processor source.
2. Cliquez sur l'icône de connection.
3. Tracez la ligne vers le processor de destination.
4. Configurez les propriétés de la connection (nom, conditions de flux, etc.).

# GESTION DES FILES D'ATTENTE

Les files d'attente dans NiFi permettent de gérer les données en attente entre deux processors :

- **Visualisation** : Surveillez la file d'attente via l'interface utilisateur.
- **Gestion** : Possibilité de vider ou de visualiser le contenu de la file.
- **Taille** : Configurable pour limiter le nombre de données en attente.

# PRIORISATION DES FLUX

La priorisation des flux dans NiFi se fait à travers :

- **Priorité des files d'attente** : Définir des stratégies de priorisation (FIFO, LIFO, Priorité attribuée).
- **Back Pressure** : Mécanisme pour gérer la pression en arrière quand les files d'attente atteignent un seuil spécifique.
- **Penalization** : Temporisation des flux pour gérer les ressources.

# GROUPES DE PROCESSUS

# DÉFINITION DES GROUPES DE PROCESSUS

Un groupe de processus dans NiFi est une collection de processeurs, de contrôleurs de services, et d'autres éléments qui sont regroupés pour organiser et gérer le traitement des données de manière collective. Chaque groupe peut encapsuler une logique de traitement spécifique.

# CRÉATION D'UN GROUPE DE PROCESSUS

Pour créer un groupe de processus dans NiFi :

1. Cliquez droit sur l'espace de travail.
2. Sélectionnez "Create Process Group".
3. Nommez le groupe et spécifiez sa localisation.
4. Cliquez sur "Create" pour finaliser.

# CONFIGURATION DES GROUPES DE PROCESSUS

Configuration d'un groupe de processus inclut :

- Définir les paramètres de flux de données.
- Configurer les contrôleurs de services utilisés.
- Ajuster les paramètres de planification et les stratégies de file d'attente.

# UTILISATION DES GROUPES DE PROCESSUS POUR L'ORGANISATION DU FLUX DE DONNÉES

Les groupes de processus permettent de :

- Structurer logiquement les flux de données.
- Isoler les fonctionnalités pour la maintenance et le débogage.
- Réutiliser des configurations dans différents contextes ou projets.

# GESTION DES PERMISSIONS DANS LES GROUPES DE PROCESSUS

Dans NiFi, la gestion des permissions pour les groupes de processus permet :

- De contrôler l'accès aux données et aux configurations.
- D'assigner des rôles spécifiques aux utilisateurs ou groupes.
- De sécuriser le traitement des données en limitant les actions possibles.

# CONTRÔLEURS DE SERVICES

# DÉFINITION DES CONTRÔLEURS DE SERVICES

Les contrôleurs de services dans NiFi sont des composants qui gèrent des services partagés utilisés par plusieurs éléments de flux de données. Ils fournissent des fonctionnalités centralisées comme la connexion à des bases de données, la mise en cache, ou la génération de données.

# TYPES DE CONTRÔLEURS DE SERVICES

- **Controller Service d'accès aux données** : Permet la connexion à des sources de données externes.
- **Controller Service de mise en cache** : Stocke des données temporairement pour améliorer les performances.
- **Controller Service de générateur de données** : Crée des données simulées pour les tests.

# CONFIGURATION DES CONTRÔLEURS DE SERVICES

1. Accéder à l'interface utilisateur de NiFi.
2. Naviguer vers le panneau de configuration.
3. Sélectionner le contrôleur de service à configurer.
4. Configurer les propriétés spécifiques au type de service.

# GESTION DES CONTRÔLEURS DE SERVICES

- **Activer/Désactiver** : Permet de contrôler l'état du service.
- **Modifier** : Permet de changer la configuration du service.
- **Supprimer** : Retire le service de l'environnement NiFi.

# UTILISATION DES CONTRÔLEURS DE SERVICES DANS LES FLUX DE DONNÉES

- Associer un contrôleur de services à des processeurs ou des groupes de processus.
- Configurer les processeurs pour utiliser les services du contrôleur.
- Utiliser les fonctionnalités du service dans le traitement des flux de données.

# TEMPLATES ET LEUR UTILISATION

# IMPORTATION DE TEMPLATES

Pour importer un template dans NiFi, suivez ces étapes :

1. Accédez à l'interface utilisateur de NiFi.
2. Cliquez sur l'icône "Operate" dans le panneau de gauche.
3. Sélectionnez l'option "Upload template".
4. Choisissez le fichier XML du template à importer.
5. Confirmez l'importation pour ajouter le template à votre instance NiFi.

# CRÉATION DE TEMPLATES

Pour créer un template dans NiFi :

1. Construisez votre flux de données en utilisant les processeurs et connexions nécessaires.
2. Sélectionnez les composants du flux que vous souhaitez inclure dans le template.
3. Faites un clic droit sur la sélection et choisissez "Create Template".
4. Donnez un nom et une description au template.
5. Cliquez sur "Create" pour enregistrer le template.

# GESTION DES TEMPLATES

Gérer les templates dans NiFi :

1. Accédez à l'onglet "Templates" dans l'interface utilisateur.
2. Ici, vous pouvez voir tous les templates importés et créés.
3. Pour chaque template, vous avez des options pour le télécharger, le visualiser ou le supprimer.
4. Utilisez ces options pour gérer efficacement vos templates selon vos besoins.

# UTILISATION DES TEMPLATES DANS LES FLUX DE DONNÉES

Pour utiliser un template dans un flux de données NiFi :

1. Naviguez vers l'onglet "Components Toolbar".
2. Sélectionnez l'icône "Template" qui ressemble à un morceau de puzzle.
3. Faites glisser le template désiré sur votre espace de travail.
4. Le flux de données du template sera automatiquement créé sur votre canevas.

# PARTAGE DE TEMPLATES

Partager un template dans NiFi :

1. Exportez le template en faisant un clic droit dessus et en sélectionnant "Download template".
2. Le template sera téléchargé sous forme de fichier XML.
3. Partagez ce fichier XML avec d'autres utilisateurs de NiFi.
4. Ils peuvent importer le template en suivant les étapes d'importation mentionnées précédemment.

# SÉCURITÉ DANS NIFI

# AUTHENTIFICATION ET AUTORISATION

NiFi utilise un mécanisme d'authentification pour contrôler l'accès des utilisateurs. Les méthodes incluent:

- Authentification par mot de passe
- Authentification via Kerberos
- Authentification via OAuth

Les autorisations sont gérées via:

- Les politiques d'accès
- Les groupes d'utilisateurs
- Les droits spécifiques sur les composants et les données

# CHIFFREMENT DES DONNÉES

NiFi supporte le chiffrement des données en transit et au repos:

- **TLS/SSL** pour les données en transit
- **EncryptConfig** pour les données au repos

Configuration via `nifi.properties` pour le chiffrement des données sensibles.

# CONFIGURATION DES POLITIQUES D'ACCÈS

Les politiques d'accès dans NiFi définissent qui peut voir ou modifier les ressources. Configuration via:

- L'interface utilisateur de NiFi
- Le fichier `authorizers.xml`

Les politiques peuvent être appliquées à des niveaux tels que:

- Process group
- Controller services
- Provenance data

# AUDIT ET TRAÇABILITÉ

NiFi propose des fonctionnalités d'audit complètes:

- Logs détaillés des actions des utilisateurs
- Provenance des données qui trace chaque événement de données

Accessible via:

- L'onglet "Provenance" dans l'interface utilisateur
- Les fichiers de log `nifi-app.log` et `nifi-user.log`

# SÉCURITÉ DU RÉSEAU ET DES COMMUNICATIONS

NiFi assure la sécurité du réseau via:

- Configuration de TLS/SSL pour sécuriser les communications
- Support de multiples interfaces réseau pour la séparation du trafic

Paramètres réseau configurables dans `nifi.properties`.

# MONITORING ET GESTION DES PERFORMANCES

# SURVEILLANCE DE L'ÉTAT DES FLUX

NiFi offre une vue en temps réel de l'état des flux de données grâce à son interface utilisateur. Cette surveillance permet de visualiser le parcours des données, les files d'attente, et l'état des processeurs.

# UTILISATION DES TABLEAUX DE BORD INTÉGRÉS

NiFi fournit des tableaux de bord intégrés pour suivre diverses métriques telles que le débit, la latence, et les erreurs. Ces tableaux de bord sont accessibles via l'interface utilisateur principale et peuvent être personnalisés selon les besoins de surveillance.

# CONFIGURATION DES ALERTES ET NOTIFICATIONS

NiFi permet de configurer des alertes et notifications pour surveiller les événements critiques dans les flux de données. Les utilisateurs peuvent définir des seuils qui, une fois franchis, déclenchent des alertes envoyées par email ou d'autres moyens.

# ANALYSE DES MÉTRIQUES DE PERFORMANCE

NiFi collecte des données de performance telles que le débit, la latence, et le taux d'erreur. L'analyse de ces métriques aide à identifier les goulets d'étranglement et les problèmes de performance dans les flux de données.

# OPTIMISATION DES PERFORMANCES DES FLUX

Pour optimiser les performances, NiFi permet d'ajuster la configuration des processeurs, de gérer la priorisation des files d'attente et d'utiliser des groupes de pression. L'optimisation régulière des flux assure une gestion efficace des ressources et améliore le traitement des données.

# DÉBOGAGE DES FLUX DE DONNÉES

# IDENTIFICATION DES ERREURS DANS LES FLUX DE DONNÉES

- Identifier les erreurs visuellement dans l'interface utilisateur de NiFi.
- Utiliser les indicateurs de file d'attente pour voir les données en attente.
- Vérifier les bulletins pour les avertissements et les erreurs.
- Examiner le statut des processeurs pour détecter les défaillances.

# UTILISATION DES LOGS POUR LE DÉBOGAGE

- Accéder aux logs via l'interface de NiFi ou les fichiers sur le serveur.
- Rechercher des erreurs ou des exceptions spécifiques dans les logs.
- Utiliser les timestamps pour corrélérer les événements dans les logs avec les erreurs dans les flux.
- Filtrer les logs pour se concentrer sur un flux ou un processeur spécifique.

# TECHNIQUES DE DÉBOGAGE PAS À PAS

- Activer le mode de débogage pas à pas pour suivre le traitement des données.
- Utiliser des points d'arrêt pour arrêter le flux à des étapes clés.
- Inspecter les attributs et le contenu des fichiers à chaque étape.
- Modifier les paramètres des processeurs pour tester différentes configurations.

# GESTION DES EXCEPTIONS ET DES ERREURS

- Configurer des processeurs spécifiques pour gérer les erreurs (ex. PutEmail pour les alertes).
- Utiliser la file d'attente de connexion pour reprendre ou rediriger les données en erreur.
- Définir des politiques de reprise sur erreur pour automatiser les réponses.
- Documenter les erreurs communes et les solutions dans la documentation du projet.

# INTÉGRATION AVEC DES Systèmes EXTERNES

# CONNECTEURS POUR BASES DE DONNÉES

NiFi fournit des connecteurs spécifiques pour intégrer différentes bases de données :

- **JDBC** pour les bases de données relationnelles
- **PutMongo** pour MongoDB
- **PutHBase** pour HBase
- **PutCassandra** pour Cassandra

# INTÉGRATION AVEC DES SYSTÈMES DE FICHIERS (HDFS, S3)

NiFi permet l'intégration avec des systèmes de fichiers distribués et des services de stockage en cloud :

- **PutHDFS** pour écrire dans Hadoop HDFS
- **FetchHDFS** pour lire depuis HDFS
- **PutS3Object** pour stocker des données sur Amazon S3
- **FetchS3Object** pour récupérer des données depuis S3

# UTILISATION DE L'API REST POUR L'INTÉGRATION

NiFi peut interagir avec des systèmes externes via l'API REST :

- Utilisation de **InvokeHTTP** pour envoyer des requêtes HTTP/HTTPS
- Configuration de méthodes GET, POST, PUT, DELETE
- Gestion des paramètres et des headers de la requête
- Traitement des réponses HTTP

# CONFIGURATION DES PROCESSEURS POUR LES SERVICES WEB

Configurer des processeurs dans NiFi pour consommer des services web :

- **InvokeHTTP** pour les appels API
- Paramétrage des URL, headers, et méthodes HTTP
- Gestion des réponses et redirection des flux de données
- Utilisation de **HandleHttpResponse** et **HandleHttpRequest** pour créer des services web

# GESTION DES FLUX DE DONNÉES VERS DES APPLICATIONS EXTERNES

NiFi facilite la gestion et le routage des flux de données vers des applications externes :

- Utilisation de **RouteOnAttribute** pour diriger les données selon des critères spécifiques
- **MergeContent** pour agréger des données avant envoi
- **UpdateAttribute** pour modifier des attributs de flux avant transmission
- Support de divers protocoles et formats de données