データ整備ツールの セットアップガイド

Ver.1.0

更新履歴

版数	更新日	更新内容
1.0	2025/2/28	初版制定

目次

- ▶ 1.OpenJDKの設定
 - 1.1 OpenJDKのダウンロード
 - 1.2 OpenJDKの初期設定
 - 1.3 OpenJDKの動作確認
- ▶ 2.Pythonの設定
 - 2.1 Pythonのダウンロード
 - 2.2 Pythonのインストール
 - 2.3 Pythonの動作確認
- ▶ 3.ApacheNiFiの設定
 - 3.1 ApacheNiFiのダウンロード
 - 3.2 ダウンロードファイルの検証
 - 3.3 ApacheNiFiのフォルダ構成

- 4.ApacheNiFiのセットアップ、起動、停止方法
 - 4.1 ApacheNiFiのセットアップ
 - 4.2 起動方法
 - 4.3 停止方法
- ▶ 5.ApacheNiFi ユーザID/パスワード変更
 - <u>5.1 ApacheNiFi ユーザID/パスワード変更</u>
- ▶ 6.文字コードの設定
 - 6.1 文字コードの設定
 - 6.2 修正コメントの記載要領
- ▶ 7.データ整備ツールのセットアップ
 - 7.1 ApacheNiFiへの機能追加

ライセンス

MIT License

Copyright (c) 2025 NTT InfraNet

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

商標および知的財産権について

- ・ApacheNiFiは、The Apache Software Foundationの米国およびその他の国における登録商標または商標です。
- ・© 2025 Microsoft Corporation. All rights reserved.

 Microsoft、Windows、およびその他の製品名は、米国およびその他の国における登録商標または商標です。
- ・Pythonは、Python Software Foundationの米国およびその他の国における登録商標または商標です。
- ・OpenJDKは、Oracleおよびその関連会社の米国およびその他の国における登録商標または商標です。
- ・本マニュアルに記載している製品名、システム名、技術用語、規格名などは、一般に各社の登録商標または商標です。また、本マニュアルに記載している技術仕様・規格に関する知的財産権は、それぞれの標準化団体または権利者に帰属します。

前提事項

セットアップ作業を行うには、インターネット接続環境が必要となりますので、事前にご確認ください。



1.1 OpenJDKのダウンロード

手順	画面イメージなど
1)右記URLにWebブラウザでアクセスします。	https://jdk.java.net/archive/
2)Windows/x64欄のzipリンクをクリックします。 ※ご利用環境に適したものをダウンロードしてください。 ※最新のバージョンを利用します。	22.0.1 (build 22.0.1+8) Windows 64-bit zip sha256) 192M Mac/AArch64 64-bit tar.gz (sha256) 187M Mac/x64 64-bit tar.gz (sha256) 189M Linux/AArch64 64-bit tar.gz (sha256) 191M Linux/x64 64-bit tar.gz (sha256) 193M Source Tag jdk-22.0.1-ga
3) ZIPファイルがダウンロードできたことを確認します。	例)openjdk-22.0.1_windows-x64_bin.zip

1.2 OpenJDKの初期設定 (1 / 4)



1.2 OpenJDKの初期設定 (2 / 4)



1.2 OpenJDKの初期設定 (3 / 4)



1.2 OpenJDKの初期設定 (4 / 4)



1.3 OpenJDKの動作確認



2.1 Pythonのダウンロード

手順	画面イメージ	など					
1)右記URLにWebブラウザでアクセスします。	https://www.python.org/downloads/						
2)Windows installer(64bit)のリンクをクリックします。 ※ご利用環境に適したものをダウンロードしてください。	Files						
※バージョンはpython 3.11.9を利用します。	Version	Operating System	Description	MD5 Sum	File Size	GPG	Sigstore
	Gzipped source tarball	Source release		bfd4d3bfeac4216ce35d7a503bf02d5c	25.3 MB	SIG	.sigstore
	XZ compressed source tarball	Source release		22ea467e7d915477152e99d5da856ddc	19.2 MB	SIG	sigstore
	macOS 64-bit universal2 installer	macOS	for macOS 10.9 and later	fa29f456feb6b5c4f52456a8b8ba347b	42.8 MB	SIG	.sigstore
	Windows installer (64-bit)	Windows	Recommended	e8dcd502e34932eebcaf1be056d5cbcd	25.0 MB	SIG	sigstore
	Windows installer (ARM64)	Windows	Experimental	328d93f71cb078965e4cfa2eb2663fa1	24.3 MB	SIG	.sigstore
	Windows embeddable package (64-bit)	Windows		6d9aa08531d48fcc261ba667e2df17c4	10.7 MB	SIG	.sigstore
	Windows embeddable package (32-bit)	Windows		31e7648158376e92a4463aa6f22a78e1	9.6 MB	SIG	.sigstore
	Windows embeddable package (ARM64)	Windows		8611b6aa35483ab1c61d45e0d9f2de0d	10.0 MB	SIG	sigstore
	Windows installer (32 -bit)	Windows		2a1d1ac2d8a0aa847515f9dd121ccbb7	23.8 MB	SIG	sigstore
3) インストーラがダウンロードできたことを確認します。	例)python-3.11.	9-amd6	4.exe				

2.2 Pythonのインストール (1 / 3)



2.2 Pythonのインストール (2 / 3)

画面イメージなど 手順 4)「Advanced Options」画面が表示されたら、Browseボタンをクリックして、 Python 3.11.9 (64-bit) Setup フォルダーの参照 任意の場所にpythonフォルダを作成後、OKボタンをクリックします。 Python 3.11.9 (64-bit) Setup ※新しいフォルダーの作成ボタンでフォルダを作成できます。 > OneDriveTemp PerfLogs > Program Files > 🚞 Program Files (x86) Recovery > SWSetup フォルダー(F): Browse python 新しいフォルダーの作成(M) キャンセル Cancel windows 5)「Precompile standard library」にチェックを入れて、 「Customize install location」が「4)」で作成したpythonフォルダになって Python 3.11.9 (64-bit) Setup いることを確認後、Installボタンをクリックします。 **Advanced Options** ☐ Install Python 3.11 for all users Associate files with Python (requires the 'py' launcher) Create shortcuts for installed applications Add Python to environment variables Precompile standard library Download debugging symbols Download debug binaries (requires VS 2017 or later) Customize install location C:\python Browse python windows Back Install Cancel

2.2 Pythonのインストール (3 / 3)



2.3 Pythonの動作確認



3.1 ApacheNiFiのダウンロード(1 / 2)



3.1 ApacheNiFiのダウンロード(2 / 2)



3.2 ダウンロードファイルの検証(1 / 2)

画面イメージなど 手順 C:\NiFi>certUtil -hashfile C:\NiFi\nifi-2.0.0-M4-bin.zip SHA256 1) コマンドプロンプトを起動して、下記コマンドを入力をEnterキーを押下します。 SHA256 ハッシュ (対象 C:\NiFi\nifi-2.0.0-M4-bin.zip): fe9b07e2148a2468da4c86ec406308c7965c79b32c1076fca1d2c83b17caf04b <コマンド> CertUtil: -hashfile コマンドは正常に完了しました。 例) certUtil -hashfile "ダウンロード先"¥"ダウンロードファイル名 (拡張子あり) "SHA512 C:\NiFi> 2) 3.1 2)でダウンロードしたファイルと同様のBinaries欄にある SHA-512ボタンをクリックします。 Download Previous releases are available in release archives. Archive downloads are subject to rate limiting. ※バージョンが存在しない場合は、以下URLを使用します。 OpenPGP Project Keys can be used for download verification. NIFI MINIFI REGISTRY FDS URL: Apache NiFi https://downloads.apache.org/nifi/2.0.0/nifi-2.0.0-bin.zip.sha512 NiFi 2.0.0-M4 Release Notes o Released: 2024-07-01 Migration Guidance Source 2.0.0-M4 OpenPGP SHA-256 SHA-512 Binaries OpenPGP SHA-256 SHA-512 OpenPGP SHA-256 SHA-512 OpenPGP SHA-512 SHA-256

3.2 ダウンロードファイルの検証(2 / 2)

手順

3) コマンドプロンプトに表示されているコードと、ブラウザ上に表示されたコードが 一致していることを確認します。

この確認により、ファイルが正しくダウンロードされていることを確認をすることができます。

画面イメージなど

<コマンドプロンプト>

C:\NiFi>certUtil -hashfile C:\NiFi\nifi-2.0.0-M4-bin.zip SHA256 SHA256 ハッシュ (対象 C:\NiFi\nifi-2.0.0-M4-bin.zip): fe9b07e2148a2468da4c86ec406308c7965c79b32c1076fca1d2c83b17caf04b CertUtil: -hashfile コマンドは正常に完了しました。

C:\NiFi>

<ブラウザ>

fe9b07e2148a2468da4c86ec406308c7965c79b32c1076fca1d2c83b17caf04b

3.3 ApacheNiFiのフォルダ構成

フォルダ階層	各フォルダの概要
¥NIFI-2.0.0 —assets —bin —conf —content_repository —database_repository —docs —extensions —flowfile_repository —lib —logs —nar_repository —provenance_repository —python —run —state —work	[assets] NiFiが使用するリソースファイルを格納される場所です。 [bin] NiFiの起動やステータス確認用のバッチファイルなどが保存されている場所です。 [conf] NiFiのコンフィグファイルなどが保存されている場所です。 [content_repository] システム内のすべてのFlowFilesのコンテンツを保持する場所です。 [database_repository] H2 データベースリポジトリの保存場所で、フロー設定履歴を追跡する内部データベースの設定の定義ファイルがあります。 [docs] 各ドキュメントが保存されている場所です。 [extensions] JAVA製力スタムプロセッサを保存する場所です。 [flowfile_repository] 現在システムに存在する各FlowFileのメタデータの「先行書き込みログ」が保存される場所です。 [lib] narライブラリが保存される場所です。 [logs] 標準ログファイルが保存される場所です。 [nar_repository] カィファイルが格納される場所です。 [provenance_repository] 各FlowFileの履歴が格納される場所です。 [python] Python製のカスタムプロセッサを保存する場所です。 [state] コンポーネントの状態を保存する場所です。 [state] コンポーネントの状態を保存する場所です。 [work] Python製のカスタムプロセッサの実行環境のことです。データ整備ツールの機能群として使用されてます。

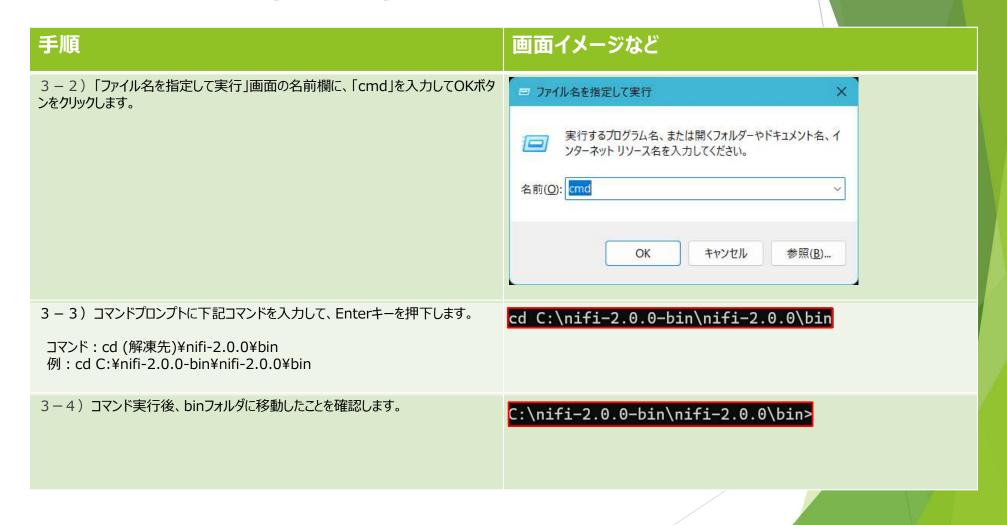
4.1 ApacheNiFiのセットアップ

手順	画面イメージなど
1) ダウンロードしたZIPファイルを任意の場所に解凍します。	nifi-2.0.0-bin.zip
2) 「(解凍先)¥nifi-2.0.0¥conf¥nifi.properties」を ダブルクリックして開きます。	
3) 45行目辺りに以下の記述があるので、下記のとおり修正して保存します。 修正前) #nifi.python.command=python3	######################################
修正後)nifi.python.command=python	nifi.python.command=python nifi.python.framework.source.directory=./python/framework
※「=」の右側は、「2.2 Pythonのインストールの4)」で指定した フォルダ内にあるpython.exeを指定します。※Pythonカスタムプロセッサを有効にするために、この行のコメントを外してPythonを呼び出すために使用するコマンドに設定します。	

4.2 起動方法(1 / 4)



4.2 起動方法(2 / 4)



4.2 起動方法(3 / 4)

手順	画面イメージなど
3 – 5) コマンドプロンプトに下記コマンドを入力して、Enterキーを押下します。 コマンド: nifi.cmd start	C:\nifi-2.0.0-bin\nifi-2.0.0\bin>nifi.cmd start
3 - 6) コマンドプロンプトに以下のようなメッセージが表示されていることを確認します。 <メッセージ> INFO [main] org.apache.nifi.bootstrap.Command Application Process [xxxx] started	2025-02-03 14:45:20,263 INFO [main] org.apache.nifi.bootstrap.Command Application Process [2440] started 2025-02-03 14:45:20,267 INFO [main] org.apache.nifi.bootstrap.Command Bootstrap Process Running
※xxxxは起動の都度、変化します。 INFO [main] org.apache.nifi.bootstrap.Command Bootstrap Process Running	

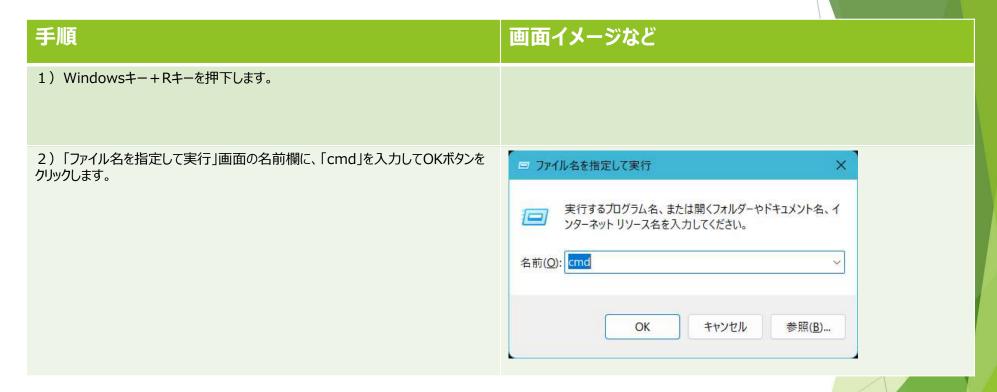
4.2 起動方法(4 / 4)

手順	画面イメージなど
4 - 1)テキストに以下コマンドを入力して、保存する。	
<コマンド> set "filePath=%~dp0¥nifi.cmd" start cmd /k "%filePath% start"	
4 - 2) 「4 - 1)」で保存したファイルを「nifi.cmd」と同様のフォルダ内に格納し、 実行する。	™ nifi.cmd
	nifi.sh
	Nifi-env.cmd ■ Nifi-env.cmd Nifi-env
	nifi-env.sh Surun.bat
	■ Turi.Dat
4-3) コマンドプロンプトに以下のようなメッセージが表示されていることを確認します。	2025-02-03 15:07:37,550 INFO [main] org.apache.nifi.bootstrap.Command Application Process [19332] started 2025-02-03 15:07:37,553 INFO [main] org.apache.nifi.bootstrap.Command Bootstrap Process Running
<メッセージ> INFO [main] org.apache.nifi.bootstrap.Command Application Process [xxxx] started ※xxxxは起動の都度、変化します。 INFO [main] org.apache.nifi.bootstrap.Command Bootstrap Process Running	

4.3 停止方法

手順	画面イメージなど
1) 起動したコマンドプロンプトをアクティブにして、CTRL+Cキーを押下します。	
2)「バッチジョブを終了しますか(Y/N)?」のメッセージが表示されたら、「Y」を入力してEnterキーを押下します。 ※すぐに表示されない場合がありますが、表示されるまで待つ様にお願いします。 ※停止に時間がかかった場合に「NiFi has not finished shutting down after 20 seconds. Killing process.」メッセージが表示される場合がありますが、「Y」を入力してEnterキーを押下で問題ありません。	2025-02-03 15:07:37,550 INFO [main] org.apache.ロ2025-02-03 15:07:37,553 INFO [main] org.apache.ロバッチ ジョブを終了しますか (Y/N)?

5.1 ApacheNiFi ユーザID/パスワード変更(1)



5.1 ApacheNiFi ユーザID/パスワード変更(2)

手順	画面イメージなど
3) コマンドプロンプトに下記コマンドを入力して、Enterキーを押下します。 コマンド: cd (解凍先)¥nifi-2.0.0¥bin 例: cd C:¥NiFi¥nifi-2.0.0¥bin	cd C:\nifi-2.0.0-bin\nifi-2.0.0\bin
4)コマンド実行後、binフォルダに移動したことを確認します。	C:\nifi-2.0.0-bin\nifi-2.0.0\bin>
5)コマンドプロンプトに下記コマンドを入力して、Enterキーを押下します。 コマンド: dir	C:\nifi-2.0.0-bin\nifi-2.0.0\bin>dir

5.1 ApacheNiFi ユーザID/パスワード変更(3)



5.1 ApacheNiFi ユーザID/パスワード変更(4)

6.1 文字コードの設定方法

手順	画面イメージなど
1) ApacheNiFiをインストール後、「ProcessorInspection.py」に2点の変更を行います。 ※作業する際は、ApacheNiFiを停止後に作業をしてください。	フォルダ: (解凍先) ¥nifi-2.0.0¥python¥framework ProcessorInspection.py
2) 赤字で記載している内容に修正します。(get_module_string_constants)	変更前)
ソースコード) def get_module_string_constants(module_file: str) -> dict: with open(module_file, encoding="utf-8") as file:	<pre>def get_module_string_constants(module_file: str) -> dict: with open(module file) as file:</pre>
	変更後) def get_module_string_constants(module_file: str) -> dict: with open(module_file, encoding="utf-8") as file:
3) 赤字で記載している内容に修正します。(get_processor_class_nodes)	変更前)
ソースコード) def get_processor_class_nodes(module_file: str) -> list: with open(module_file, encoding="utf-8") as file:	<pre>def get_processor_class_nodes(module_file: str) -> list: with open(module file) as file:</pre>
	変更後) def get_processor_class_nodes(module file: str) -> list: with open(module_file <mark>, encoding="utf-8")</mark> as file:

6.2 修正コメントの記載要領

手順 画面イメージなど 1)「ProcessorInspection.py」の修正を行う場合、 例) ライセンス表記の後に、以下の変更内容で追記してください。 # Licensed to the Apache Software Foundation (ASF) under one or more # contributor license agreements. See the NOTICE file distributed with # this work for additional information regarding copyright ownership. 例) # The ASF licenses this file to You under the Apache License, Version 2.0 # (the "License"); you may not use this file except in compliance with # Modified <修正した年> by <変更した方の名前> # the License. You may obtain a copy of the License at # - Modified Python custom processor to support UTF-8 encoding http://www.apache.org/licenses/LICENSE-2.0 # Unless required by applicable law or agreed to in writing, software # distributed under the License is distributed on an "AS IS" BASIS, # WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. # See the License for the specific language governing permissions and # limitations under the License. # Modified 2025 by TARO # - Modified Python custom processor to support UTF-8 encoding

7.1 ApacheNiFiへの機能追加(1/3)



7.1 ApacheNiFiへの機能追加(2/3)

手順	画面イメージなど
3) ApacheNiFiでPythonカスタムプロセッサが使用できるか確認します。 ApacheNiFiにを起動してログインします。 ※ログイン方法は、別紙「データ整備ツールの利用者向けマニュアル」をご参照く ださい。	
4) Prosessorアイコンをキャンバス上にドラッグアンドドロップして、「Add Processor」画面のFilter typesに「2)」で追加したプロセッサいづれかを入力し、リストに表示されるかを確認します。	Add Processor Type ↑ Version Tags CheckinvalidGeometry 1.0.0 beta, GeoDataFrame, python, Pyth
5) NiFi Flowのキャンバスに、対象のPythonカスタムプロセッサを追加します。 ※プロセッサ追加の手順は、別紙「データ整備ツールの利用者向けマニュアル」を 参照してください。 ※選択肢に表示されるまで時間を要する場合があります。	CheckInvalidGeometry CheckInvalidGeometry 1.0.0 org.apache.nifi - python-extensions In 0 (0 bytes) 5 min Read/Write 0 bytes / 0 bytes 5 min Out 0 (0 bytes) 5 min Tasks/Time 0 / 00:00:00.000 5 min

7.1 ApacheNiFiへの機能追加(3 / 3)

手順

6) 追加・変更時は、Pythonの仮想環境の作成と依存関係のある サードパーティのダウンロードが開始されます。

進捗状況は▲ にカーソルを合わせることで確認可能で、メッセージ①⇒ メッセージ②の順に変化し、両メッセージが表示されなくなれば使用可能です。 ※追加・変更にはインターネット接続が必要です。

画面イメージなど



<メッセージ①: ランタイムの初期化>

'Processor' is invalid because Initializing runtime environment for the Processor.

<メッセージ②:依存関係のあるサードパーティのダウンロード>

 'Processor' is invalid because In the process of downloading third-party dependencies required by the Processor.