

アップデートガイド

1. はじめに

本書は、Generative AI FW に於いて、Generative AI FWの最新版へのアップデート手順について記載したものです。

本書の内容に関しては、将来予告なしに変更することがあります。

本書の内容の一部または全部を無断で複製・転載・改編することは禁止します。

1.1. 用語定義について

用語定義については「スタートアップマニュアル（概要編）」をご確認ください。

2. 前提条件

- ・ 本書に記載の手順は全てサーバの管理者ユーザなどの**管理者権限を持つユーザで行う必要があります**。一般ユーザでしかログオンできない環境の場合は以下を実行し、管理者ユーザに昇格させてください。もしくはコマンド実行時に「sudo」を付けてください。

```
1 sudo -i
```

- ・ 本書に記載の手順はサーバにログオンしている状態である必要があります。
- ・ 本書のアップデート手順はセットアップガイドの実施が事前に完了している必要があります。
- ・ 本ガイドの標準のアップデート手順はインターネットに接続できる環境で実施する必要があります。閉域環境下での構築には対応しておりません。閉域環境で構築する場合の手順は後述の「閉域環境でのアップデート」を確認してください。

3. 動作環境

セットアップガイドを確認してください。

4. 事前準備

4.1. OSSイメージ配置ディレクトリ削除

i /opt/nec/genai/oss ディレクトリが存在しない場合は実施不要です。

以下コマンドを実行してOSSイメージ配置ディレクトリを削除します。

```
1 rm -rf /opt/nec/genai/oss
```

5. アップデート手順

i 閉域環境でアップデートする場合は、「閉域環境でのアップデート」の手順を参照してください。

▲ Generative AI FWの最新版が既定で対応しているcotomiはcotomi v3です。アップデートする際はcotomi v3のコンテナイメージのファイルを入手したうえで実施してください。すでにcotomi v3を使用している場合、改めて入手する必要はありません。

5.1. アップデート前のバージョンの確認

アップデート前の環境のバージョンを確認します。

i アップデート前のバージョン番号は後述の「切り戻し手順」で必要になる可能性があります。事前に分かっている場合は再度確認する必要はありません。

バージョンの確認方法は以下の通りです。

i HTTPSに既定では自己証明書を使用しています。その場合「この接続ではプライバシーが保護されませんが、問題ありませんので接続してください。

- ・ チャット画面

1. Webブラウザを立ち上げ以下にアクセスします。アクセス先はGENAI_DOMAINで指定した値になります。

i https://<GENAI_DOMAINの指定値>/

2. 画面右下に記載しているbuild番号を確認します。

- 管理ポータル

1. Webブラウザを立ち上げ以下にアクセスします。アクセス先はGENAI_DOMAINで指定した値になります。

```
https://<GENAI_DOMAINの指定値>/admin
```

2. 画面右下に記載しているbuild番号を確認します。

各バージョンのbuild番号は以下の通りです。

バージョン	build番号
V2.0	チャット画面 : build 202501230406 または build 202502040703 管理ポータル : build 202501211035 または build 202502030913
V2.1	チャット画面 : build 202505210808 または build 202505210837 管理ポータル : build 202505210556 または build 202505210624
V2.1.1	チャット画面 : build 202506230605 管理ポータル : build 202506190411
V2.1.2	チャット画面 : build 202507220707 または build 202507240433 管理ポータル : build 202507230641 または build 202508270531
V2.2	チャット画面 : build 202512112355 または build 202512080254 管理ポータル : build 202512110849 または build 202512110856
V2.2.1	チャット画面 : build 202512112355 または build 202512080254 管理ポータル : build 202601190519 または build 202512110856

5.2. アップデートディレクトリ作成

アップデートモジュールを格納するディレクトリを作成します。

```
1 mkdir /opt/nec/genai_upgrade
```

5.3. アップデートモジュール配置

インストール物件の圧縮ファイルを /opt/nec/genai_upgrade に配置します。

以下コマンドを実行して圧縮ファイルを解凍します。

```
1 cd /opt/nec/genai_upgrade  
2 bash -c "umask 0027; tar --no-same-permissions -zxvf genai.tar.gz"
```

5.4. アップデートスクリプト実行

アップデートスクリプトを実行します。

① アップデートスクリプトでは以下を自動で実施します。そのため、個別に事前で実施していただく必要はありません。

- Generative AI FWのバックアップ
- 現環境の設定ファイルの引継ぎ (引継ぎ対象 : genai.env)
- genaiサービス停止

① 最新版では自動的にcotomi v3がセットアップされます。

① アップデート前の時点で、セットアップガイドの「プロキシ設定・証明書の設定」を実施済みの場合、アップデート時に同設定が自動的に再適用されます。ただし、以下の注意事項がありますのでご確認ください。

- RHEL リポジトリへのアクセスが必要です。V2.0.2以前のバージョンからアップデートする場合は、リポジトリが設定されていない可能性があります。事前にセットアップガイドの「RHEL リポジトリの有効化」の手順を実施してください。

- ・ squidの設定ファイルが上書きされます。本製品以外の用途で Squid を使用している場合は、手順完了後にセットアップガイドの「Squid の設定更新」に従って設定を再適用してください。

5.4.1. LLMモデルのコンテナイメージの配置

1. cotomi v3のLLMのコンテナイメージのファイルを以下に格納します。

▲ 以下の条件を満たす場合は本手順は実施不要です。

- ・ ファイル名が「vllm-openai-v0.11.2-cotomi-fp8-v3.0.1.tar.gz」のファイルが既に格納されている
- ・ LLMなし構成で構築・セットアップをしており、cotomi v3のセットアップが不要

```
1 /opt/nec/genai/models/cotomi
```

5.4.2. 削除対象コンテナイメージの確認

初めに削除対象のOSSのコンテナイメージに問題ないか確認します。

1. バージョンアップスクリプトを実行します。引数にKeycloakの管理者ユーザのパスワードを指定します。

```
1 cd /opt/nec/genai_upgrade/genai/operation
2 bash genai_verup.sh <Keycloakの管理者ユーザのパスワード>
```

2. 削除対象のコンテナイメージが以下のように表示されます。

i IMAGE_IDの値やREPOSITORY:TAGの数値に関しては以下の画面と一致しない場合がありますが問題ありません。

2.

```
root@vm-pp-test-094 operation]# bash genai_verup.sh keycloak
The OSS container images that will be deleted by upgrade are as follows.
-----
IMAGE_ID      REPOSITORY:TAG
0ae1c7c9841d  docker.io/mongodb/mongodb-community-server:7.0.20-ubuntu2204
2ff0239b7235  docker.io/library/postgres:16.9
3389325d940e  docker.io/bitnami/etcd:3.5.21
0623d5d38cc6  docker.io/keycloak/keycloak:26.2.4
0068c8d73dbd  docker.io/vllm/vllm-openai:v0.8.5.post1
b05483cc4dbb  docker.io/library/nix:1.27.5-alpine-slim
1ee9779d3b90  docker.io/elastic/elasticsearch:8.16.1
3b8c02bc2641  docker.io/apache/apisix:3.11.0-debian
-----
Are you sure you want to 2.1.2 upgrade Generative AI FW? [y/N]:
```

3. REPOSITORY:TAGに表示されているコンテナイメージを削除して問題ないか確認してください。

- i REPOSITORY:TAGに記載があるコンテナイメージを同一サーバに同居する別製品が使っていた場合、影響を受ける可能性があります。ただし問題があるのは以下の**全ての条件を満たす場合のみ**になります。基本的には削除して問題ありません。
 - Generative AI FWサーバにPodmanを使用するソフトウェアを同居させている場合（詳細は同居しているソフトウェアの動作環境をご確認ください）。
 - 使用しているコンテナイメージのバージョンがリビジョン番号含めて完全に一致している場合（バージョンの詳細は同居しているソフトウェアの動作環境をご確認ください）
 - インターネット通信が不可能な環境（閉域環境など）で使用している場合（インターネット通信が可能な場合は、誤って削除された場合でも次回起動時に新たにコンテナイメージを取得するため問題ありません）、もしくは使用しているコンテナイメージのバージョンが古過ぎてDocker Hubに存在しない場合（コンテナイメージが存在するかはDocker Hubにて確認してください）
- ・ 上記のすべての条件に合致した場合、OSSのコンテナイメージを削除することは推奨しておりません。後述の「OSSコンテナイメージを削除しない場合」に進んでください。それ以外は「OSSコンテナイメージを削除する場合」に進んでください。
- ・ 削除対象のコンテナイメージ一覧は後述の「OSSコンテナイメージを個別に削除する場合」で使用する場合があります。「OSSコンテナイメージを削除しない場合」の手順を実施する場合は表示された情報を保存しておくことを推奨します。

4. 以下の確認には**必ず「n」を入力して一度アップデートを中止**します。X.X.Xにはアップデート後のバージョン番号が記載されています。

```
1 Are you sure you want to X.X.X upgrade Generative AI FW? [y/N]:
```

5.4.3. OSSコンテナイメージを削除する場合

削除対象のコンテナイメージを全て削除してアップデートを行います。以下の手順を実施してください。

1. アップデートスクリプトを実行します。引数にKeycloakの管理者ユーザのパスワードを指定します。

```
1 cd /opt/nec/genai_upgrade/genai/operation
2 bash genai_verup.sh <Keycloakの管理者ユーザのパスワード>
```

2. 削除対象のコンテナイメージが再度表示されます。X.X.Xにはアップデート後のバージョン番号が記載されています。問題ないことを確認のうえ「y」を入力してアップデートを実行してください。

```
1 Are you sure you want to X.X.X upgrade Generative AI FW? [y/N]:
```

なお、アップデートスクリプトを実行中下記のようなログが表示され処理が進まないことがあります。しばらくすると処理が進みますので問題ありません（5分から10分程度進まない可能性があります）。

```
1 =====
2 fluent-package Installation Script
3 =====
4 This script requires superuser access to install rpm packages.
5 You will be prompted for your password by sudo.
```

3. アップデートが正常に完了すると、以下メッセージが表示されます。

```
1 ...
2 version upgrade succeeded.
```

5.4.4. OSSコンテナイメージを削除しない場合

OSSのコンテナイメージを残して、アップデートを行います。以下の手順を実施してください。

i 本手順ではOSSコンテナイメージを削除しないため、ディスクにデータが残ります。データは残りますがコンテナ自身は起動していないため、意図せずOSSが動作することはありません。

1. 必ず「--skip-uninstall-oss」オプションを指定してアップデートスクリプトを実行します。引数にKeycloakの管理者ユーザのパスワードを指定します。

```
1 cd /opt/nec/genai_upgrade/genai/operation
2 bash genai_verup.sh <Keycloakの管理者ユーザのパスワード> --skip-uninstall-oss
```

2. 削除対象のコンテナイメージが再度表示されます。「y」を入力してアップデートを実行してください。X.X.Xにはアップデート後のバージョン番号が記載されています。問題ないことを確認のうえ「y」を入力してアップデートを実行してください。

```
1 Are you sure you want to X.X.X upgrade Generative AI FW? [y/N]:
```

i アップデートスクリプトを実行中下記のようなログが出ることがありますが、しばらくすると処理が進みますので問題ありません。

```
1 =====
2 fluent-package Installation Script
3 =====
4 This script requires superuser access to install rpm packages.
5 You will be prompted for your password by sudo.
```

3. オプションを指定した場合、ターミナルに以下メッセージが表示されます。

```
1 Container image deletion was skipped.
```

4. アップデートが正常に完了すると、以下メッセージが表示されます。

```
1 ...
2 version upgrade succeeded.
```

5. fluentd がすでにインストールされている環境の場合、環境構築スクリプト実行時のメッセージに以下が表示されます。メッセージに従い、バックアップファイル(/opt/nec/genai/config/backup/fluentd/fluentd.conf)の内容で必要な設定をfluentdの設定ファイル(/etc/fluent/fluentd.conf)にマージしてください。

```
1 Fluentd is already installed.
2 The pre-installation configuration file is located at /opt/nec/genai/config/backup/fluentd/fluentd.conf.
3 Merge it into the fluentd configuration file (/etc/fluent/fluentd.conf) if necessary.
```

i . メッセージが表示されない場合、マージ作業は必要ありません。
. アップデートでは既存のfluentdの設定ファイルを上書きしています。マージする際は既にある設定は消さないようにしてください。

6. OSSコンテナイメージをどうしても削除したい場合のみ、次の「OSSコンテナイメージを個別に削除する場合」を実施してください。ディスクに残してもよい場合は実施不要です。

5.4.4.1. OSSコンテナイメージを個別に削除する場合

OSSのコンテナイメージを個別に削除する手順は以下の通りです。

i . 本手順は削除してよいOSSが明確であり、コンテナイメージのデータをディスクから削除したい場合のみ実施してください。明確でない場合は、同居する別製品への影響を考慮し、削除せず残すことを推奨いたします。

。OSSコンテナイメージが他の製品で使用されていないか確認する明確な手段はございません。削除する場合は、お客様の判断で実施してください。

1. 前述の「削除対象コンテナイメージの確認」で保存した一覧情報のうち削除したいコンテナイメージのREPOSITORY:TAG列の値で以下を実行します。

```
1 podman images <REPOSITORY:TAG列の値>
```

実行結果例は以下の通りです。指定したREPOSITORY:TAGのコンテナイメージが存在するか確認します。

```
1 REPOSITORY          TAG          IMAGE ID      CREATED      SIZE
2 docker.io/mongodb/mongodb-community-server 7.0.20-ubuntu2204 10009356e87d 44 hours ago 1.25 GB
```

2. 削除して問題ない場合は**REPOSITORY:TAGの値**を使って以下を実行し削除してください。

```
1 podman rmi <REPOSITORY:TAG列の値>
```

実行例は以下の通りです。

```
1 podman rmi docker.io/mongodb/mongodb-community-server:7.0.20-ubuntu2204
```

i 削除対象のコンテナイメージを利用するコンテナが存在している場合コンテナイメージは削除できません。Generative AI FWで使用しているコンテナはアンインストール時に削除されています。削除できない場合は指定が誤っているか別の製品で使用している可能性があります。

5.5. アップデート後のバージョンの確認

アップデート後の環境のバージョンを確認します。

確認対象	バージョン
チャット画面	build 202512112355 または build 202512080254
管理ポータル	build 202601190519または build 202512110856

バージョンの確認方法は以下の通りです。

i HTTPSに既定では自己証明書を使用しています。その場合「この接続ではプライバシーが保護されません」と注意が表示されますが、問題ありませんので接続してください。

- ・チャット画面

1. Webブラウザを立ち上げ以下にアクセスします。アクセス先はGENAI_DOMAINで指定した値になります。

```
i https://<GENAI_DOMAINの指定値>/
```

2. 画面右下に記載しているbuild番号を確認します。

- ・管理ポータル

1. Webブラウザを立ち上げ以下にアクセスします。アクセス先はGENAI_DOMAINで指定した値になります。

```
i https://<GENAI_DOMAINの指定値>/admin
```

2. 画面右下に記載しているbuild番号を確認します。

5.6. アップデート後の追加対応（外部LLMのセットアップ）

アップデートで外部LLMに関する設定が引き継いでいないため、外部LLM機能を使用していた場合は再度セットアップが必要です。設定方法については運用ガイドの「LLMの設定変更」を参照ください。

外部LLMを使用しない場合は本セットアップは不要です。

5.7. アップデート後の追加対応 (Web検索機能のセットアップ)

- Web検索機能を使うには別途検索エンジンサービスとの契約が必要です。契約はお客様ご自身で行う必要があります。
 - 利用可能な外部の検索エンジンサービスは、セットアップガイドをご確認ください。
 - V2.1.2以降からのアップデートの場合は本手順は実施不要です。アップデート前の設定が引き継がれています。
- V2.1.2より前のバージョンからのアップデートの場合のみ実施してください。

Web検索機能を使用するためにはセットアップが必要です。

設定方法についてはセットアップガイドの「拡張対話 (Web検索) 機能のセットアップ」を参照ください。

5.8. アップデート後の追加対応 (ハルシネーション対策機能のセットアップ)

アップデート前のバージョンですでにセットアップ済みの場合は本手順は実施不要です。

ハルシネーション対策機能を使用するためにはセットアップが必要です。

設定方法についてはセットアップガイドの「ハルシネーション対策機能のセットアップ」を参照ください。

5.9. アップデート後の追加対応 (拡張対話 (ファイル添付) 機能のセットアップ)

LLMに画像ファイルを添付して対話したい場合、セットアップが必要になります。

設定方法についてはセットアップガイドの「拡張対話 (ファイル添付) 機能のセットアップ」を参照ください。

本機能を使用するには条件があります。詳細は上記セットアップガイドの説明をご確認ください。

5.10. アップデート後の追加対応 (APIキー)

- V2.1.1以降からのアップデートの場合は本手順は実施不要です。APIキーはアップデート前の値が引き継がれています。
- V2.1.1より前のバージョンからのアップデートの場合のみ実施してください。

アップデートでAPIキーが初期化されているため再度セットアップが必要です。

APIキーの更新は管理ポータル画面から行います。詳細は「管理ポータル操作ガイド (設定画面編)」の「APIキー情報」をご確認ください。

5.11. アップデート後の正常性確認

アップデートが正常にできているかどうかを管理ポータル画面、チャット画面、APIの観点で確認していきます。

正常性確認手順は「正常性確認ガイド」をご確認ください。

5.12. アップデートディレクトリ削除

アップデートモジュールを格納したディレクトリを削除します。配下にアップデート時のバックアップデータも含まれます。

- ▲ アップデートディレクトリの削除した場合、配下にあるアップデート前のバックアップデータも削除されるため後述の切り戻し手順の実施ができなくなります。削除する場合は事前に削除して問題ないか確認してください。

```
1 rm -rf /opt/nec/genai_upgrade
```

以上でアップデートの手順は完了になります。

6. 切り戻し手順

アップデート後に不具合が発生し、前バージョンに切り戻すは以下の手順を参照ください。

6.1. Generative AI FW のアンインストール

アップデート後のアンインストールガイドの手順に従って、Generative AI FWをアンインストールします。

6.2. Generative AI FW のインストール

構築ガイドに従って、アップデート前と同じバージョンのGenerative AI FWを構築します。

インストール物件がない場合は「アップデート前のバージョンの確認」で確認したbuild番号をもとにサポート窓口までご相談ください。

❶ 累積パッチはリストア後の再アップデート時に適用するため本時点での適用は不要です。

6.3. アップデート前のデータでリストア

6.3.1. リストアコマンドの確認

アップデート前のバージョンによってはリストアコマンドが配置されていない可能性があります。

以下のファイルが存在するか確認してください。

```
1 /opt/nec/genai/operation/genai_restore.sh
```

存在する場合は次に進んでください。

存在しない場合は以下のファイルを/opt/nec/genai/operation配下にコピーします。

```
1 cp /opt/nec/genai_upgrade/genai/setup/v2_scripts/genai_restore.sh /opt/nec/genai/operation
```

6.3.2. リストア実行

1. アップデート時のバックアップフォルダを確認します。以下のフォルダ配下にある日時のフォルダがバックアップフォルダになります。

```
1 ls /opt/nec/genai_upgrade/backup
2 20250620_021122
```

2. リストアを実行します。引数には上記で確認したバックアップフォルダを指定します。

```
1 cd /opt/nec/genai/operation
2 bash genai_restore.sh /opt/nec/genai_upgrade/backup/XXXXXXXX_XXXXXX
```

3. リストアに成功すると以下のメッセージが表示されます。バックアップ対象のリストアが成功していることを確認してください。

```
1 all restore succeeded.
```

4. リストアコマンドの確認でリストアコマンドが存在しなかった場合のみ、アップデート前の状況と揃えるために以下のファイルは削除してください。

```
1 rm -f /opt/nec/genai/operation/genai_restore.sh
```

切り戻し後のアップデートは本ガイドの最初から再度実施してください。

7. 閉域環境でのアップデート

閉域（インターネット通信不可）環境下のサーバー上でアップデートする場合、本手順を実施する必要があります。閉域環境ではない環境下で構築する場合は、本手順は不要です。

7.1. 前提条件

- ・閉域環境での構築にあたり、インストールするサーバとは別にインターネット通信が可能な事前準備用の別サーバが必要です。事前準備手順と構築手順を行うサーバのOSとバージョンは同じである必要があります（Windowsには対応していません）

7.2. 事前準備用サーバ動作環境

❶ 事前準備用サーバを用意できない場合はサポート窓口までご相談ください。

OS	セットアップガイドの動作環境と同じ
ディスク空き容量	var配下（Podman、コンテナイメージ）：40GB以上 opt配下（OSSイメージファイル、パッケージファイル）：60GB以上
コンテナ管理	セットアップガイドの動作環境と同じ

その他	以下がインストールされていない環境 <ul style="list-style-type: none"> • podman-compose • nvidia-container-toolkit
-----	--

7.3. 事前準備手順

① インターネット通信が可能な事前準備用の別サーバ上で本手順を実施してください。

1. 管理者アカウントでOSにログオンします。一般ユーザでしかログオンできない環境の場合は以下を実行し、管理者ユーザに昇格させてください。

```
1 sudo -i
```

2. podmanインストール

```
1 dnf install -y podman
```

3. /opt/nec/genai/setupフォルダを作成します。以下のコマンドを実行します。

```
1 mkdir -p /opt/nec/genai/setup
```

4. インストール物件の圧縮ファイルを /opt/nec/genai_upgrade に配置します。以下コマンドを実行して圧縮ファイルを解凍します。

```
1 cd /opt/nec/genai_upgrade
2 bash -c "umask 0027; tar --no-same-permissions -zxvf genai.tar.gz"
```

5. 物件内のgenai/setup/download_oss.shを/opt/nec/genai/setupフォルダに配置します。
6. 閉域環境構築に必要なOSSイメージファイルおよびパッケージをダウンロードします。

```
1 cd /opt/nec/genai/setup/
2 bash download_oss.sh <ダウンロード先ディレクトリパス>
```

8. 以下のOSSイメージファイルがダウンロードされていることを確認します。

① ダウンロードされるOSSイメージファイル

```
apisix.tar.gz
elasticsearch.tar.gz
etcd.tar.gz
keycloak.tar.gz
mongodb-community-server.tar.gz
nginx.tar.gz
postgres.tar.gz
vllm-openai.tar.gz
```

※閉域環境構築に必要な上記OSSイメージファイルのほか、パッケージ（rpmファイル群）も含まれています

9. インストールした物件はインストールするサーバに転送するため圧縮します。

```
1 cd <ダウンロード先ディレクトリパス>
2 tar -zcvf oss.tar.gz ./*
```

7.4. アップデート手順

① 閉域（インターネット通信不可）環境下のアップデートするサーバー上で本手順を実施します。

1. 管理者アカウントでOSにログオンします。一般ユーザでしかログオンできない環境の場合は以下を実行し、管理者ユーザに昇格させてください。

```
1 sudo -i
```

2. OSSイメージファイル配置先ディレクトリを作成します。

```
1 mkdir -p /opt/nec/genai/oss
```

3. 事前準備手順でダウンロードしたOSSイメージファイルおよびパッケージを/opt/nec/genai/oss配下に配置し、解凍します。

```
1 mv <配置ディレクトリパス>/oss.tar.gz /opt/nec/genai/oss
2 cd /opt/nec/genai/oss
3 tar -zxvf oss.tar.gz
4 rm -f oss.tar.gz
```

4. /opt/nec/genai/oss配下が以下の構成になっていることを確認します。

```
1 /opt/nec/genai
2 |— oss
3 |   |— apisix.tar.gz
4 |   |— elasticsearch.tar.gz
5 |   |— etcd.tar.gz
6 |   |— keycloak.tar.gz
7 |   |— mongodb-community-server.tar.gz
8 |   |— nginx.tar.gz
9 |   |— postgres.tar.gz
10 |  |— vllm-openai.tar.gz
11 |  └─ パッケージ (rpmファイル群)
```

5. 以降の手順は通常のアップデート手順と同じになります。先述のアップデート手順の「アップデート前のバージョンの確認」から順次手順を実施してください。