cotomi Appliance Server 初期設定ガイド

第 3 版

2024 年 8 月

商標について

Red Hat、Red Hat Enterprise Linuxは米国およびその他の国におけるRed Hat, Inc.の商標または登録商標です。

Linux は Linus Torvalds 氏の米国およびその他の国における商標または登録商標です。

Microsoft、Windows、Windows Server は米国 Microsoft Corporation の米国およびその他の国における登録商標です。

本書のサンプル画像などで使用している名称は、すべて架空のものです。実在する品名、団体名、個人名とは一切関係ありません。

記載の会社名および商品名は各社の商標または登録商標です。

ご注意

(1) 本書の内容の一部または全部を無断転載することは禁止されています。

(2) 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。

(3) NEC の許可なく複製・改変などを行うことはできません。

(4) 本書の内容および本書を使用した結果について明示的にも黙示的にも一切の保証を行いません。

目次

1 はじめに

本書では、ご購入された cotomi Appliance Server の運用を開始する前に行う初期設定に関して説明します。
初期設定後の使用方法に関してはユーザーズガイドをご参照ください。

本書は、NEC Generative AI Framework v1.3.1 が搭載された cotomi Appliance Server を対象としています。

2 事前に準備が必要なもの

初期設定を始める前に、下記をご用意ください。

2.1 機器

ネットワーク経由でサーバにアクセスできるようになるまでは、装置にコンソールを接続する必要があります。
以下の機器を準備してください。

- ディスプレイ
- キーボード
- マウス

ネットワーク経由でアクセス可能になった以降は、管理端末を使用して設定を行います。
以下のソフトウェアが使用可能な管理端末を準備してください。

- SSH クライアント
- Microsoft Edge ブラウザ

2.2 設定情報

初期設定では以下に示す情報を装置に設定します。事前に設定値を準備してください。

設定	説明
ホスト名(FQDN)	本装置のホスト名(FQDN)。本装置にアクセスしてサービスを利用するクライアントは本装置のホスト名の名前解決が可能でなければなりません。DNS 等に事前に登録しておく必要があります。 (例) cotomi.example.com
IP アドレス	本装置の IP アドレス。上記ホスト名(FQDN)に対応する IP アドレスです。
Red Hat Enterprise Linux 初期ユーザー(cotomi)のパスワード	装置出荷時に設定されているパスワードは装置に添付のスタートアップガイドに記載されています。初期設定で必ず変更を行うので変更後のパスワードを決めておく必要があります。
PostgreSQL 管理者ユーザー(postgres)のパスワード	cotomi ソフトウェアで使用している PostgreSQL 管理者ユーザー(postgres)のパスワードです。初期設定で必ず変更を行うので変更後のパスワードを決めておく必要があります。
Keycloak 管理者ユーザー(keycloak)のパスワード	cotomi Appliance Server では利用者アカウントを管理する ID プロバイダーとして Keycloak を使用しており、本項目は Keycloak 管理者ユーザー(keycloak)のパスワードです。初期設定で必ず変更を行うので変更後のパスワードを決めておく必要があります。
ElasticSearch 管理者ユーザー(elastic)のパスワード	cotomi ソフトウェアで使用している ElasticSearch 管理者ユーザー(elastic)のパスワードです。ログ管理を行う Kibana にログインする時にも使用します。初期設定で必ず変更を行うので変更後のパスワードを決めておく必要があります。
Kibana が ElasticSearch に接続する際のパスワード	cotomi ソフトウェアで使用している Kibana が Elasticsearch との接続と通信に使用するパスワードです。初期設定で必ず変更を行うので変更後のパスワードを決めておく必要があります。
初期管理アカウント(cotomiadmin)のパスワード	管理画面にアクセス可能な初期管理アカウント(cotomiadmin)のパスワードです。初期設定で必ず変更を行うので変更後のパスワードを決めておく必要があります。
API 用クライアント名とパスワード	一般画面の対話を使用せず API に直接アクセスする場合はアクセスするクライアント名とパスワードを決めておく必要があります。
Kibana 閲覧専用ユーザーのユーザー名とパスワード	ログ管理を行う Kibana の閲覧専用ユーザーを作成する場合はそのユーザー名とパスワードを決めておく必要があります。

社内情報登録用のユーザー名、パスワード	「社内情報を探索」で検索する社内情報を登録、更新及び削除する際に使用するユーザー名とパスワードです。初期設定で変更を行うので決めておく必要があります。

2.3 DNS登録

cotomi Appliance Server にアクセスするクライアントはサーバのホスト名(FQDN)を名前解決できる必要があります。事前にホスト名(FQDN)を DNS に登録しておいてください。

DNS を利用できない場合にはサーバを利用する全てのクライアントの hosts にサーバのホスト名(FQDN)を登録するなど、ホスト名の名前解決が可能な状態にしておいてください。

2.4 SSLサーバ証明書

装置に SSL サーバ証明書を配置する必要があるため、外部の認証局などに SSL サーバ証明書の発行を依頼される場合は事前に準備してください。
サーバ証明書のファイルと、それに対応した秘密鍵ファイルが必要です。
証明書には Subject Alternative Name(SAN)が含まれている必要があります。

自己署名証明書を利用される場合も事前に準備しておくとスムーズに設定を行えますが、本書の中で簡易的な作成手順を案内します。

3 初期設定手順

3.1 起動

装置の電源を入れます。

iLO 統合リモートコンソールを利用して、リモートから設定を続ける場合は「iLO6 ユーザーズガイド」を参照してリモートコンソールを使用できるようにしてください。

Express5800/R120j-2M のユーザーズガイドは NEC サポートポータルからコンテンツ ID 3170102819 を検索してください。

3.2　Red Hat Enterprise Linuxの設定

3.2.1　ライセンス契約同意

Red Hat Enterprise Linux のライセンス契約に同意する必要があります。

装置が起動すると初期セットアップ画面となります。

「ライセンス情報(L)」を選択します。ライセンス契約画面が開き、Red Hat Enterprise Linux のライセンス条項が表示されます。使用許諾契約書を確認して、ライセンス契約に同意します チェックボックスを選択します。「完了」をクリックして設定を適用し、初期セットアップ 画面に戻ります。

その後「設定の完了」をクリックするとログイン画面となります。

3.2.2　初期ユーザーのパスワード変更

本装置には Red Hat Enterprise Linux の初期ユーザーが登録されており、パスワードと共に装置に添付されているスタートアップガイドに記載しております。

初期ユーザーでサインインします。

ログイン後「端末」アプリケーションを起動し、passwd コマンドで初期ユーザーのパスワードを変更します。

```
$ passwd
```

現在のパスワード(current password)と新しいパスワードの入力が促されるので、指示に従ってパスワードを設定します。

注意　必ずパスワードの変更を行ってください。

3.2.3 ネットワーク設定

- ホスト名の設定

「端末」アプリケーションから hostnamectl コマンドでホスト名(FQDN)を設定します。

```
$ sudo hostnamectl set-hostname 〈ホスト名(FQDN)〉
```

「端末」アプリケーションを一旦終了し、再度アプリケーションを起動します。

hostnamectl コマンドでホスト名が変更されたことを確認します。

```
$ hostnamectl status --static
(設定されたホスト名が表示されます)
```

- IP アドレスの設定

Red Hat Enterprise Linux のドキュメントを参照し、ネットワークインタフェースに IP アドレスを設定します。ネットワークボンディングを利用して複数の物理ネットワークインタフェースから冗長性を備えた論理的なネットワークインタフェースを作成することも可能です。

https://access.redhat.com/documentation/ja-jp/red_hat_enterprise_linux/8/html/configuring_and_managing_networking

〈例〉

- 接続の一覧表示

```
$ nmcli connection show
```

- ens15f0 に IP アドレス、デフォルトゲートウェイ、DNS サーバ、自動接続を設定する例

```
$ sudo nmcli connection modify ens15f0 ¥
 ipv4.method manual ¥
 ipv4.addresses 192.0.2.1/24 ¥
 ipv4.gateway 192.0.2.254 ¥
 ipv4.dns 192.0.2.200 ¥
 connection.autoconnect yes
```

3.2.4 時刻同期設定

Red Hat Enterprise Linux は NTP プロトコルによる時刻同期に対応しています。

NTP プロトコルによる時刻同期を利用する場合は、Chrony の設定を行います。

詳細は Red Hat Enterprise Linux のドキュメントを参照してください。

https://access.redhat.com/documentation/ja-jp/red_hat_enterprise_linux/8/html/configuring_basic_system_settings/configuring-time-synchronization_configuring-basic-system-settings

〈例〉NTP サーバ ntp.example.com と同期する例

- Chrony の設定ファイルを更新します(エディタは vi 以外でも構いません)。

```
$ sudo vi /etc/chrony.conf
```

`pool 2.rhel.pool.ntp.org iburst`

の行を削除し、以下の内容を追加し保存します。

`server ntp.example.com iburst`

- chronyd サービスを開始、有効化します。

```
$ sudo systemctl enable --now chronyd
```

- 同期状態を確認します。

```
$ chronyc tracking
...(略)
System time     :  0.000006523 seconds slow of NTP time
...(略)
```

3.2.5 Red Hat カスタマーポータルへの登録

本製品には「Linux サービスセット」契約が含まれており、Red Hat カスタマーポータルに登録が可能です。

手順は NEC サポートポータルでコンテンツ ID 3140001276 を検索し、「[RHEL] Red Hat カスタマーポータル(旧 Red Hat Network) 利用手順」を参照ください。

3.2.6 ファイルシステム追加

アプライアンスサーバーには OS 用のディスク領域(OS ブート専用 SSD ボード)とは別に SSD 4 台が標準で搭載されておりデータ領域として自由に使用できます。

使用する場合は以下のドキュメントを参照し、RAID の構築、ファイルシステムの構築を行ってください。

- RAID の構築

  NEC サポートポータルからコンテンツ ID 3170102819 を検索し「Express5800/R120j-2M ユーザーズガイド」にある「Smart Storage Administrator ユーザーズガイド」を参照してください。
- ファイルシステムの構築

  Red Hat Enterprise Linux のドキュメントを参照してください。

  https://access.redhat.com/documentation/ja-jp/red_hat_enterprise_linux/8/html/managing_file_systems/index

3.3　cotomiソフトウェアの設定

ここまでで Red Hat Enterprise Linux のネットワーク設定が完了しているため、これ以降は管理端末の SSH クライアント、およびブラウザからサーバに接続し設定を進めます。

cotomi ソフトウェアは /opt/nec/llm ディレクトリにインストールされています。
管理端末から SSH で接続し、インストールディレクトリが存在していることを確認してください。

```
$ ls -ld /opt/nec/llm
```

本書は、NEC Generative AI Framework v1.3.1 が搭載された cotomi Appliance Server を対象としています。以下のコマンドで TAG 欄に 1.3.1 が表示されることを確認してください。

```
$ sudo podman images ngf_api_onpre
REPOSITORY                TAG         IMAGE ID      CREATED      SIZE
localhost/ngf_api_onpre   1.3.1       d38267053941  4 weeks ago  1.57 GB
```

3.3.1　GPUの使用準備

以下の 2 つのコマンドを実行し、cotomi ソフトウェアが GPU を使用する準備を行ってください。

```
$ sudo nvidia-ctk system create-device-nodes ¥
 --control-devices ¥
 --load-kernel-modules
```

```
$ sudo nvidia-ctk cdi generate --output=/etc/cdi/nvidia.yaml
```

3.3.2 SSLサーバ証明書の配置

準備しておいた SSL サーバ証明書ファイル(certificate.crt)と秘密鍵ファイル(private.key)をサーバ内の /opt/nec/llm/certs ディレクトリに配置してください(scp や sftp を利用してください)。
※ファイル名が異なる場合は、以降の手順で適宜読み替えてください。

管理端末から SSH で接続し、配置したファイルの権限を変更します。

```
$ sudo chown 0:0 /opt/nec/llm/certs/certificate.crt
$ sudo chmod 440 /opt/nec/llm/certs/certificate.crt
$ sudo chown 0:0 /opt/nec/llm/certs/private.key
$ sudo chmod 440 /opt/nec/llm/certs/private.key
```

SSL 証明書の発行を認証局に依頼せず、自己署名証明書を使用する場合は、本装置で以下のコマンドにより作成することも可能です。簡易的な例ですので、詳細は openssl コマンドのマニュアル等をご確認ください。

(例)　自己署名証明書(有効期限 3653 日)の作成例

```
$ mkdir ~/certs
$ cd ~/certs
$ openssl genrsa -out private.key 2048
$ openssl req -new -key private.key -out server.csr -passin pass: -subj "/CN=cotomi"
$ echo "subjectAltName = DNS:ホスト名" > /tmp/san.txt
$ openssl x509 -req -days 3653 -signkey private.key -in server.csr ¥
 -out certificate.crt -extfile /tmp/san.txt
```

ホスト名には本装置のホスト名(FQDN)を指定してください。カレントディレクトリに証明書ファイル(certificate.crt)と秘密鍵ファイル(private.key)が作成されます。

3.3.3 パラメーターの設定

3.3.3.1 GPU枚数の設定

cotomi Appliance Server には GPU 数の異なる複数のモデルがあります。出荷時には GPU が 1 個の状態の設定ファイルとなっているため、GPU 数が 2 個の場合には変更が必要です。

搭載されている GPU の個数を以下のコマンドで確認します。

```
$ sudo nvidia-smi -L
```

「GPU 0:」で始まる行だけが出力された場合は GPU の個数は１個です。

「GPU 0:」で始まる行と「GPU 1:」で始まる行の２行が出力された場合は GPU の個数は２個です。

GPU の個数が２個の場合は以下の２点の変更を行ってください。

１個の場合は次の「3.3.3.2 パラメーターの更新」に進んでください。

(1) /etc/systemd/system/llm.service ファイルの変更

```
$ sudo vi  /etc/systemd/system/llm.service
```

（エディタは vi 以外でも構いません）

`ExecStartPre=/usr/bin/nvidia-modprobe -c 0`

の行に以下のように「 -c 1」を追加して保存します。

`ExecStartPre=/usr/bin/nvidia-modprobe -c 0 -c 1`

(2)　/opt/nec/llm/.env ファイルの更新

```
$ sudo vi /opt/nec/llm/.env
```

（エディタは vi 以外でも構いません）

`TENSOR_PARALLEL_SIZE=1`

の行の「1」を以下のように「2」に変更します。

`TENSOR_PARALLEL_SIZE=2`

3.3.3.2 パラメーターの更新

/opt/nec/llm/.env ファイルに定義されている下記のパラメーターの該当行を更新します。

（エディタは vi 以外でも構いません）

```
$ sudo vi /opt/nec/llm/.env
```

- ホスト名

  本装置のホスト名(FQDN)を指定します。

  (例)

```
HOST=cotomi.example.com
```

- IP アドレス

  本装置の IP アドレスを指定します。

  (例)

```
IP=10.8.1.11
```

- SSL サーバ証明書、秘密鍵のファイル名

  SSL サーバ証明書のファイル名が certificate.crt ではない場合、または、秘密鍵ファイルのファイル名が private.key ではない場合は、そのファイル名を指定してください。

  ※ファイル名が certificate.crt と private.key の場合は変更不要です。

  (例)

```
NGINX_CERTIFICATE=./certs/<SSL サーバ証明書のファイル名>
NGINX_KEY=./certs/<秘密鍵ファイルのファイル名>
```

- ElasticSearch 管理者ユーザーのパスワード

  準備しておいた ElasticSearch 管理者ユーザーのパスワードを指定します。

  (例)

```
ELASTIC_PASSWORD=oupun-LX
```

- Kibana が ElasticSearch に接続する際のパスワード

  準備しておいた Kibana が ElasticSearch に接続する際のパスワードを指定します。

  (例)

```
KIBANA_PASSWORD=oupun-LX
```

3.3.3.3 Proxyサーバの URL設定

Proxy 環境下に本装置を設置する場合は、以下の設定を行います。Proxy 環境下にない場合や外部にアクセスを行わない場合、この設定は不要です。

```
$ sudo vi /etc/sysconfig/llm
```

(エディタは vi 以外でも構いません)

```
#HTTP_PROXY=<http://proxy.example.com:8080>
#HTTPS_PROXY=<http://proxy.example.com:8080>
#NO_PROXY=127.0.0.1,localhost,ui,api,ui-manage,es01,basaran,vllm
```

上記のようにコメントアウトされた 「HTTP_PROXY」、「HTTPS_PROXY」、「NO_PROXY」の行の行頭の'#'を削除しコメントアウトを解除します。「HTTP_PROXY=」の後に HTTP 用の Proxy サーバの URL を、「HTTPS_PROXY=」の後に HTTPS 用の Proxy サーバの URL を指定します。

例)

```
HTTP_PROXY=http://proxy.example.com:8080
HTTPS_PROXY=http://proxy.example.com:8080
NO_PROXY=127.0.0.1,localhost,ui,api,ui-manage,es01,basaran,vllm
```

3.3.4 Keycloak管理者、初期管理アカウント、Keycloakクライアント設定、PostgreSQL管理者パスワードの設定

- 一時的なサービスの起動

PostgreSQL と Keycloak 管理者のパスワードを変更するため、一時的に cotomi ソフトウェアの LLM サービスを開始します。

```
$ sudo systemctl start llm
```

- Keycloak 管理者(keycloak)、初期管理アカウント(cotomiadmin)のパスワード変更、Keycloak クライアント設定

管理端末からブラウザで以下の URL にアクセスします。

`https://<ホスト名(FQDN)>/keycloak/admin`

Keycloak 管理者のユーザー名は `keycloak` です。

以下のユーザー名、初期パスワードでログインします。

- `Username: keycloak`
- `Password: keycloak`

(1) 左側メニュー(≡)から「Users」をクリックし、ユーザー一覧を表示します。

(2) ユーザー一覧の中から Keycloak 管理者ユーザーである「keycloak」をクリックします。

(3) 表示された画面の「Credentials」タブをクリックし、「Reset Password」ボタンをクリックします。

(4) 表示されたダイアログに変更後の新しいパスワードを入力します。

(5) 「Temporary」のスイッチは「Off」にします。

(6) 「Save」ボタンを押すと確認があるので「Reset password」ボタンを押してパスワード変更を確定

します。

続いて、初期管理アカウント(cotomiadmin)のパスワードを変更します。

(7) 左側メニュー(≡)の上部にあるドロップダウンリストの Realm が「master」等「LLMService」以外になっている場合は Realm を「LLMService」に切り替えます。

(8) メニューの「Users」をクリックし、ユーザー一覧を表示します。

(9) ユーザー一覧の中から初期管理アカウントである「cotomiadmin」をクリックします。

(10) 表示された画面の「Credentials」タブをクリックし、「Reset Password」ボタンをクリックします。

(11) 表示されたダイアログに変更後の新しいパスワードを入力します。

(12) 「Temporary」のスイッチは「Off」にします。

(13) 「Save」ボタンを押すと確認があるので「Reset password」ボタンを押してパスワード変更を確定します。

続いて、Keycloak のクライアント設定を行います。

(14) 左側メニュー(≡)の上部にあるドロップダウンリストの Realm が「master」等「LLMService」以外になっている場合は Realm を「LLMService」に切り替えます。

(15) メニューの「Clients」をクリックし、クライアント一覧を表示します。

(16) クライアント一覧の中から一般対話画面用の「general-client」をクリックします。

(17) 表示された画面の「Settings」タブをクリックし、「Access Settings」にある「Valid redirect URIs」に「https://<ホスト名(FQDN)>/api/auth/callback/keycloak」を設定します。

(18) 「Save」ボタンを押して「Valid redirect URIs」変更を確定します。

(19) さらに、クライアント一覧の中から管理者画面用の「admin-client」をクリックします。

(20) 表示された画面の「Settings」タブをクリックし、「Access Settings」にある「Valid redirect URIs」に「https:// <ホスト名(FQDN)>/admin/api/auth/callback/keycloak」を設定します。

(21) 「Save」ボタンを押して「Valid redirect URIs」変更を確定します。

- PostgreSQL 管理者のパスワード変更

(1) PostgreSQL コンテナに接続します。

```
$ sudo podman exec -it postgres bash
```

(2) PostgreSQL に管理者ユーザーで接続します。

```
# psql -h localhost -U postgres
```

(3) 以下のように入力し管理者ユーザーのパスワードを変更します。

※「(新しいパスワード)」の部分に変更後のパスワードを指定してください

```
postgres=# ALTER ROLE postgres with PASSWORD '(新しいパスワード)';
postgres=# exit
```

(4) PostgreSQL コンテナから出ます。

```
# exit
```

- サービスの停止

一時的に起動していたサービスを停止します。

```
$ sudo systemctl stop llm
```

- パラメーターの更新

/opt/nec/llm/.env ファイルに定義されている PostgreSQL 管理者ユーザーのパスワードを更新します。
（エディタは vi 以外でも構いません）

```
$ sudo vi /opt/nec/llm/.env
```

➢ PostgreSQL 管理者ユーザーのパスワード

前の手順で変更した PostgreSQL 管理者ユーザーのパスワードを指定します。

(例)

```
PG_PASSWORD=postgres
```

3.3.5 サービスの起動・終了と自動起動設定

cotomi ソフトウェアの LLM サービスを開始します。

```
$ sudo systemctl start llm
```

起動していることを確認します。

```
$ sudo systemctl status llm
...(略)
   Active: active (running) since ...
...(略)
```

サービス起動後、チャットの機能が使えるようになるまでに 1 分程度時間を要します。それまではチャットはエラーとなるため、しばらく待ってから使用を開始してください。

サービスを停止する場合は以下のコマンドを実行します。

```
$ sudo systemctl stop llm
```

OS 起動時に自動的にサービスを起動したい場合はサービスの有効化を行います。

```
$ sudo systemctl enable llm
```

3.3.6 API用 Basic認証クライアント登録

cotomi ソフトウェアの一般画面を使用せず API に直接アクセスする場合は、API に対して basic 認証を有効化し、アクセス可能なクライアント(*1)を制限します。

*1: 本節の文脈におけるクライアントとは、API にアクセスする Web アプリケーションやコマンドラインといったユーザーインタフェースとなるプログラムを示します。一般画面や管理画面にログイン時に指定するアカウントとは異なります。

この手順はサービスを起動した状態で実施してください。

- アクセスするクライアントと対応するパスワードの登録

“client” という名前のクライアントとそのパスワードを登録する例

```
$ sudo -i
# cd /opt/nec/llm/nginx
# htpasswd -c .htpasswd client
New password: <パスワードを入力>
Re-type new password: <もう一度パスワードを入力>
Adding password for user client
# ls -la /opt/nec/llm/nginx/
total 12
drwxr-xr-x. 2 root root 41 Nov 21 06:28 .
drwxr-xr-x. 21 root root 4096 Nov 21 05:58 ..
-rw-r--r--. 1 root root 45 Nov 21 06:27 .htpasswd
-rw-r--r--. 1 root root 1963 Nov 21 06:07 nginx.conf
# exit
```

- 登録されているクライアントのパスワード変更

“client” という名前のクライアントのパスワードを変更する例

```
$ sudo -i
# cd /opt/nec/llm/nginx
# htpasswd .htpasswd client
New password: <パスワードを入力>
Re-type new password: <もう一度パスワードを入力>
Updating password for user client
# ls -la /opt/nec/llm/nginx/
total 12
drwxr-xr-x. 2 root root 41 Nov 21 06:28 .
drwxr-xr-x. 21 root root 4096 Nov 21 05:58 ..
-rw-r--r--. 1 root root 45 Nov 21 06:27 .htpasswd
-rw-r--r--. 1 root root 1963 Nov 21 06:07 nginx.conf
# exit
```

- 登録されているクライアントの削除

“client” という名前のクライアントを削除する例

```
$ sudo -i
# cd /opt/nec/llm/nginx
# htpasswd -D .htpasswd client
Deleting password for user client
# ls -la /opt/nec/llm/nginx/
total 12
drwxr-xr-x. 2 root root 41 Nov 21 06:28 .
drwxr-xr-x. 21 root root 4096 Nov 21 05:58 ..
-rw-r--r--. 1 root root 45 Nov 21 06:27 .htpasswd
-rw-r--r--. 1 root root 1963 Nov 21 06:07 nginx.conf
# exit
```

- 確認

API アクセスを許可するクライアントのみがファイルに登録されているか確認します。

```
$ sudo cat /opt/nec/llm/nginx/.htpasswd
client:$apr1$uMNRRyLj$bRQ1FJsGmvDxAydFqBpUI1
...
```

登録されている<クライアント名>/<パスワード>で API にアクセスできるか確認します。

“client” という名前のクライアントで確認する例

```
$ curl -k -i -u <クライアント名>:<パスワード> ¥
 https://<ホスト名(FQDN)>/apiauth/api/v1/chat ¥
 -H "Content-Type: application/json" ¥
```

```
 -d '{"user_id": "api_client", "messages":
 [{"role": "user", "content": "Hello"}], "model":"necllm"}'
HTTP/1.1 200 OK
Server: nginx/1.25.2
Date: ...
```

3.3.7 ログ管理 Kibana 閲覧専用ユーザーの作成

ログ管理 Kibana の閲覧専用ユーザーを作成する場合はこの手順を実施してください。

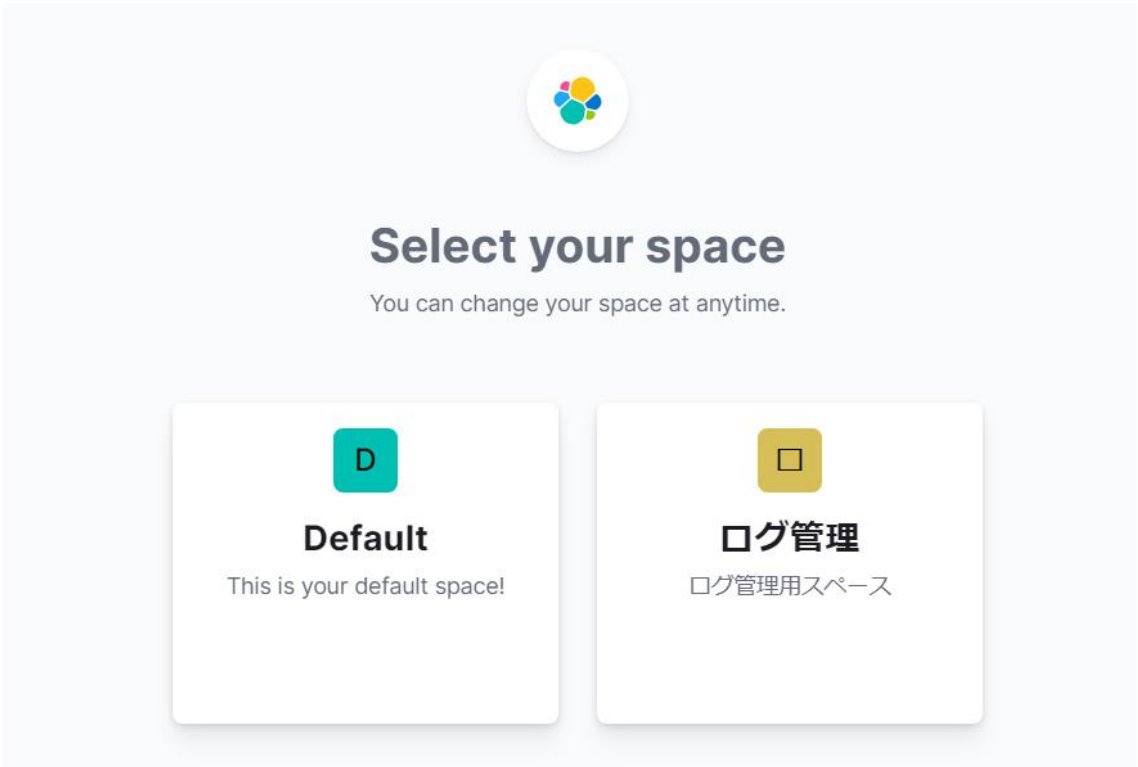
この手順はサービスを起動した状態で実施してください。

Kibana 画面はブラウザから https://<ドメイン名(FQDN)>/kibana でアクセスできます。

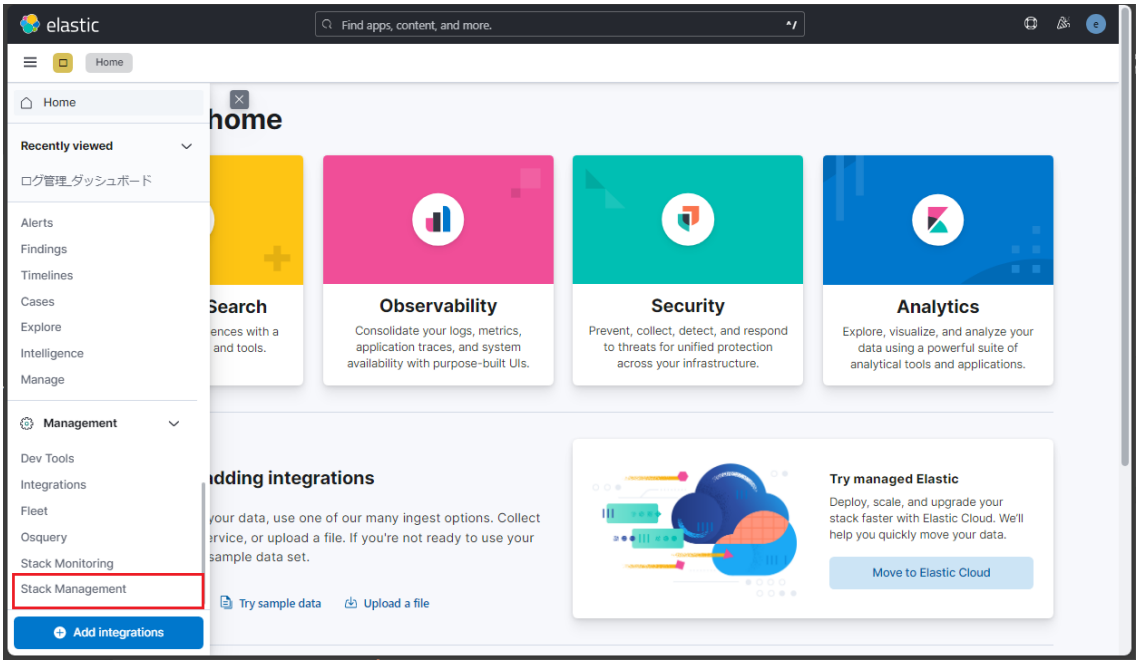
- Username：elastic
- Password：ElasticSearch 管理者ユーザーのパスワード(前の手順により変更済みのもの)

- 閲覧専用ユーザーの作成

1. [Select your space] で[ログ管理]を選択します。

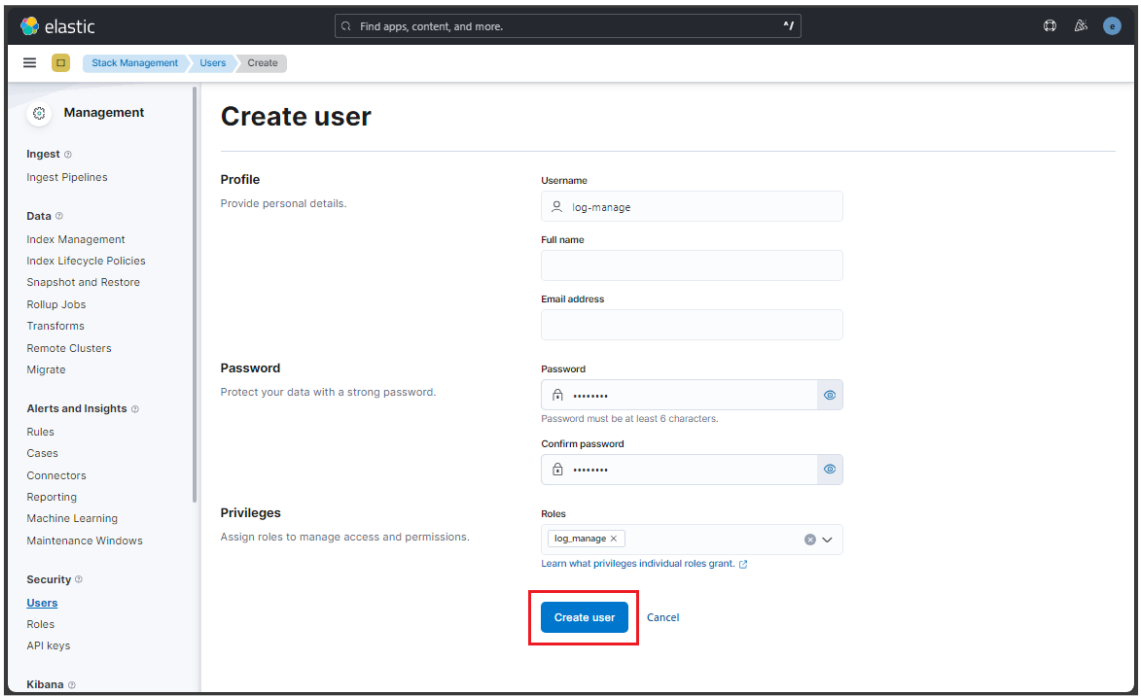


2. メニュー(≡) > [Management] > [Stack Management] > [Users] と進み、[Create user] を選択します。



3. 閲覧専用ユーザーのユーザー情報、パスワード、Privileges の Roles には「log_manage」を設定して [Create user] をクリックします。

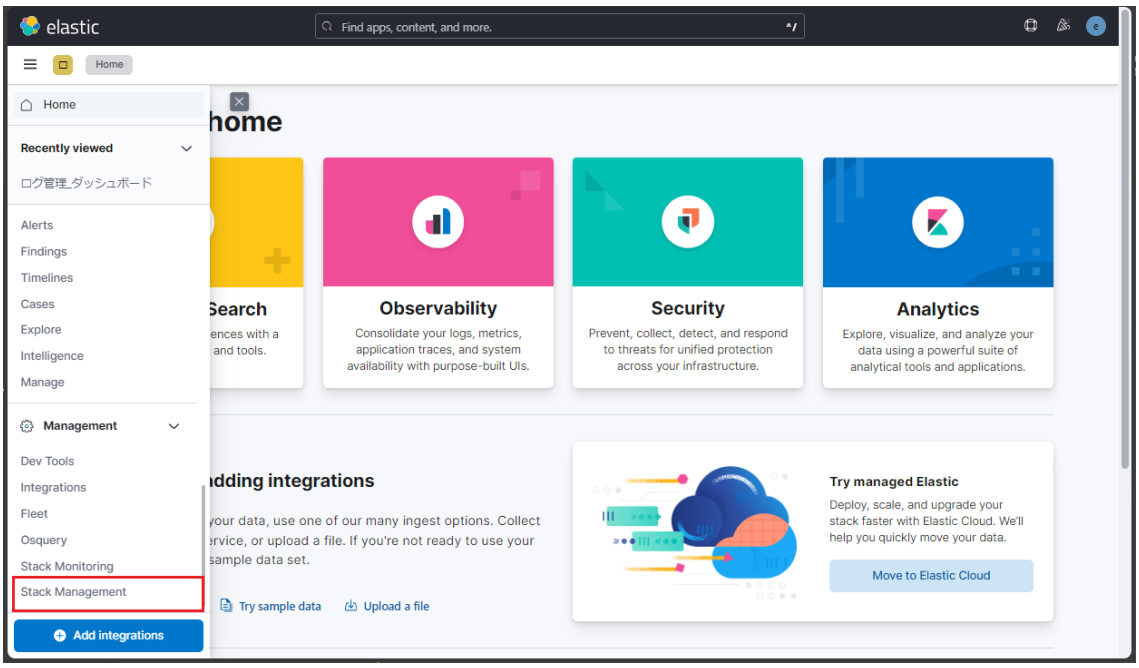
※「log_manage」Role は本製品出荷時にあらかじめ定義されています



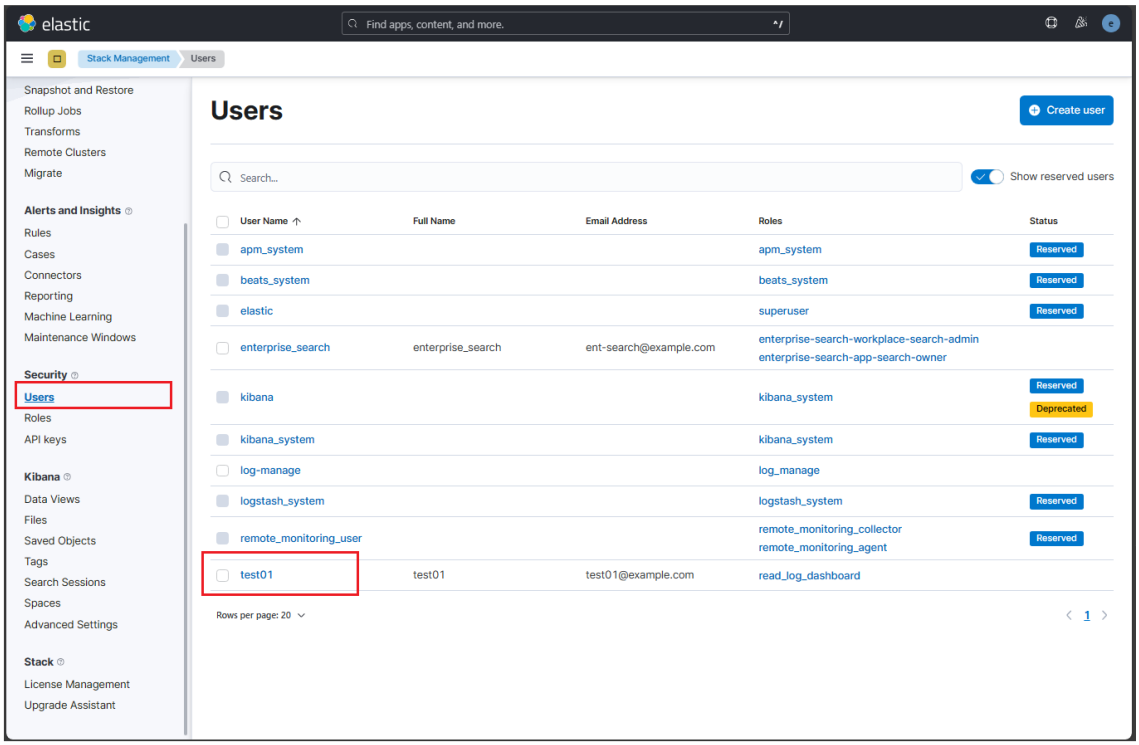
- 閲覧専用ユーザーのパスワード変更

作成済みの閲覧専用ユーザーのパスワードを変更する場合は、以下の手順を実施してください。

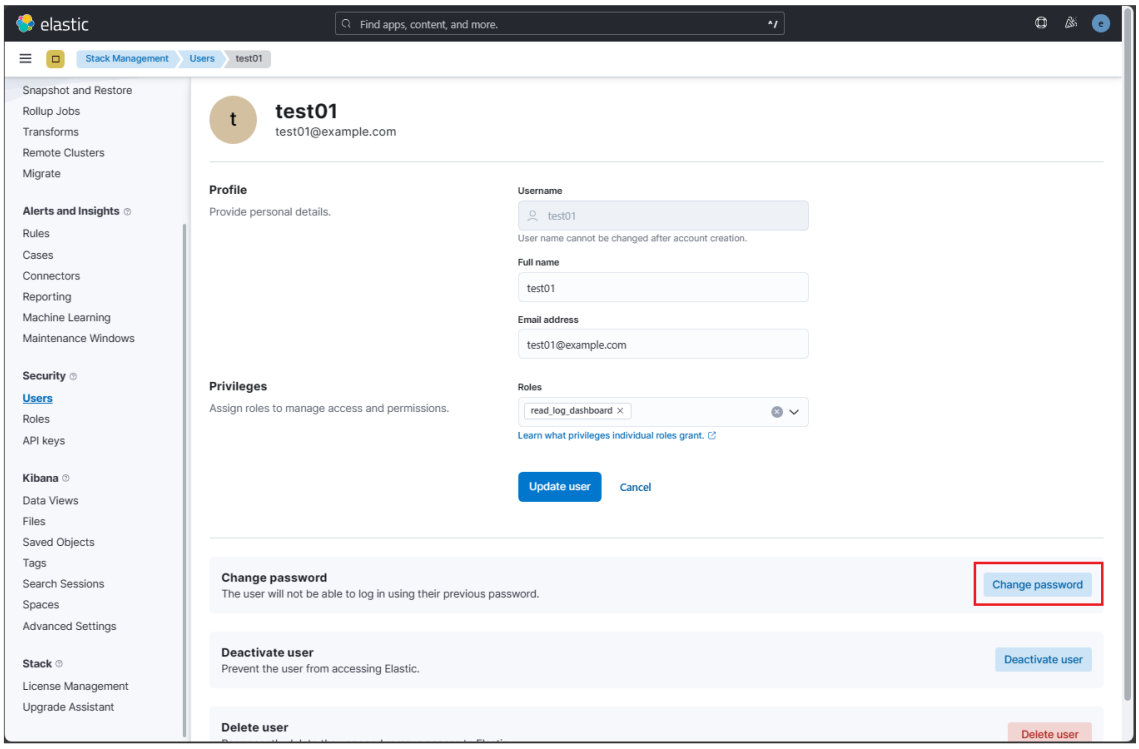
1. メニュー(≡) > [Management] > [Stack Management] > [Users] と進み、対象ユーザーを選択します。



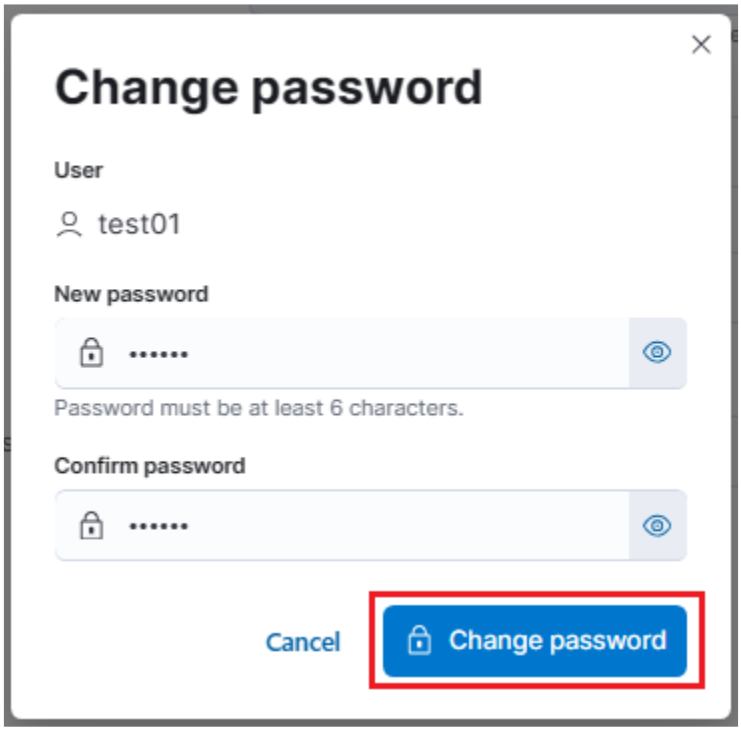
(以下では test01 ユーザーを対象例としています)



2. [Change password] を選択します。



3. 新しいパスワードを設定し、 [Change password] をクリックします。



3.3.8 社内情報登録用ユーザー、パスワードの設定

検索インデックス生成機能は「社内情報を探索」で検索する社内情報を登録、更新及び削除することがで

きる機能です。社内情報の登録、更新及び削除は共有ディレクトリを介して行いますが、ここでは共有ディレクトリへ接続するためのユーザー名、パスワードを設定します。

この手順はサービスを起動した状態で実施してください。ただし、途中サービスを停止します。

/opt/nec/llm/wsgidav.yaml ファイルに定義されているユーザー名のパスワードを更新します。

```
$ sudo cd /opt/nec/llm
$ sudo vi wsgidav.yaml
(...前のセクションは省略)
simple_dc:
    # Access control per share.
    # These routes must match the provider mapping.
    # NOTE: Provider routes without a matching entry here, are inaccessible.
    user_mapping:
        '*':  # default (used for all shares that are not explicitly listed)
            'llm_user':                ←デフォルトユーザ名:'llm_user'を変更
                password: '9PW8cQdD'   ←デフォルトパスワード:'9PW8cQdD'を変更
                # Optional: passed to downstream middleware as environ["wsgidav.auth.roles"]
                roles: ['editor']
        '/pub': true  # Pass true to allow anonymous access
```

次に以下のコマンドで、ユーザ名とパスワードを設定したファイルを新しいコンテナイメージに反映します。

```
$ sudo podman cp wsgidav.yaml ngf_tool:/root/vectorIndexer/conf/wsgidav.yaml
$ sudo podman commit ngf_tool ngf_tool-2
```

新しいコンテナイメージが登録されていることを確認し、現在動作している ngf_tool の IMAGE ID を取得します。

```
$ sudo podman images
REPOSITORY TAG IMAGE ID CREATED SIZE
localhost/mylogstash latest 046fa73cb0d2 4 weeks ago 966 MB
localhost/ngf_tool-2 latest 23367634e850 2 minutes ago 10.6 GB ←新しいコンテナイメージ
localhost/ngf_tool 1.3.1 211fb49de3ff 3 weeks ago 10.6 GB ←現在動作しているngf_tool
docker.elastic.co/kibana/kibana 8.8.1 33e5d657f95c 10 months ago 935 MB
docker.elastic.co/elasticsearch/elasticsearch 8.8.1 8552a239ee1e 10 months ago 1.34 GB
docker.elastic.co/logstash/logstash 7.6.0 799d4bde3bdd 4 years ago 830 MB
```

上記の例では「211fb49de3ff」が現在動作している ngf_tool の IMAGE ID です。

次のコマンドでサービスを停止しコンテナを停止します。

```
$ sudo systemctl stop llm
```

現在動作している ngf_tool のコンテナイメージを IMAGE ID を使って削除します。

```
$ sudo podman rmi <現在動作している ngf_tool のコンテナイメージを IMAGE ID>
```

(例) 以下は実行例ですので、適切に IMAGE ID を置き換えて実行してください。

```
$ sudo podman rmi 211fb49de3ff
```

現在動作しているコンテナイメージが削除され、新しいコンテナイメージが存在することを確認します。

```
$ sudo podman images
REPOSITORY TAG IMAGE ID CREATED SIZE
localhost/mylogstash latest 046fa73cb0d2 4 weeks ago 966 MB
localhost/ngf_tool-2 latest 23367634e850 2 minutes ago 10.6 GB
docker.elastic.co/kibana/kibana 8.8.1 33e5d657f95c 10 months ago 935 MB
docker.elastic.co/elasticsearch/elasticsearch 8.8.1 8552a239ee1e 10 months ago 1.34 GB
docker.elastic.co/logstash/logstash 7.6.0 799d4bde3bdd 4 years ago 830 MB
```

新しいコンテナイメージをリネームします。

```
$ sudo podman tag ngf_tool-2:latest ngf_tool:1.3.1
```

新しいコンテナイメージが登録されていることを確認します。

```
$ sudo podman images
REPOSITORY TAG IMAGE ID CREATED SIZE
localhost/mylogstash latest 046fa73cb0d2 4 weeks ago 966 MB
localhost/ngf_tool-2 latest 23367634e850 2 minutes ago 10.6 GB
localhost/ngf_tool 1.3.1 23367634e850 3 weeks ago 10.6 GB ←新しいコンテナイメージ
docker.elastic.co/kibana/kibana 8.8.1 33e5d657f95c 10 months ago 935 MB
docker.elastic.co/elasticsearch/elasticsearch 8.8.1 8552a239ee1e 10 months ago 1.34 GB
docker.elastic.co/logstash/logstash 7.6.0 799d4bde3bdd 4 years ago 830 MB
```

次のコマンドでサービスの再起動を行います。

```
$ sudo systemctl start llm
```

3.3.9 ログ管理設定追加

「NEC Generative AI Framework オンプレミス ログ管理設定追加ガイド」を参照し、ログ管理設定を行ってください。

3.3.10 各画面のアドレス

以上で初期設定は終了です。

各画面には以下のアドレスでアクセスできます。

- 管理画面

  https://<ホスト名(FQDN)>/admin

  アカウント管理や禁止ワード設定を行うことができます。

  初期管理アカウント(cotomiadmin)や、新規に作成する管理者アカウントでアクセスできます。

  詳細は「NEC Generative AI Framework 管理画面 ユーザーズガイド」を参照ください。

- Keycloak 管理画面

  https://<ホスト名(FQDN)>/keycloak/admin

  利用者アカウントを管理する ID プロバイダーの機能を使用できます。

  Keycloak 管理者ユーザー(keycloak)でアクセスできます。

- ログ管理 Kibana 画面

  https://<ホスト名(FQDN)>/kibana

  グラフや表、フィルター操作機能を使ってログ管理を行うことができます。

  ElasticSearch 管理者ユーザー、新規に作成した閲覧専用ユーザーでアクセスできます。

  詳細は「NEC Generative AI Framework ログ管理画面 ユーザーズガイド」を参照ください。

- 一般画面

  https://<ホスト名(FQDN)>/

  チャットやコールログ要約を行うことができます。

  新規に作成する一般ユーザーアカウント、管理者アカウントでアクセスできます。

  詳細は「NEC Generative AI Framework 一般画面 ユーザーズガイド」を参照ください。

- 検索インデックス作成ツールの URL

  https://<ホスト名(FQDN)>/indexer/

  検索インデックス生成機能は、「社内情報を探索」で検索する社内情報を登録、更新及び削除することができる機能です。

  詳細は「NEC Generative AI Framework 検索インデックス作成機能 ユーザーズガイド」を参照ください。

- API について

  API に関する詳細は「NEC Generative AI Framework API リファレンス」の「オンプレ版」を参照ください。

4 その他

4.1 ESMPRO/ServerAgentServiceインストール

ESMPRO/ServerAgentService (Linux 版)は装置を監視するソフトウェアです。

本装置にインストールする場合には、75MB 以上の空き容量が必要です。

Starter Pack に格納されている「ESMPRO/ServerAgentService インストレーションガイド(Linux 編)」を参照して、ESMPRO/ServerAgentService をインストールしてください。

ESMPRO/ServerAgentService (Linux 版)がインストールされているか確認するには、次のコマンドを実行してください。

```
$ sudo rpm -qa | grep Esmpro-Provider
```

次のように、Esmpro-Provider パッケージが表示された場合、インストール済みであることを意味します。

```
Esmpro-Provider-"バージョン情報"
```

4.2 SSLサーバ証明書の更新について

SSL サーバ証明書を更新する場合は、サービスを一時的に停止した状態で 「3.3.2 SSL サーバ証明書の配置」を実施し、サービスを開始してください。