

Business Intelligence (Apache Superset)

目次

はじめに	1
1. BI ツール	1
2. Apache Superset のインストール	1
2.1. Python 仮想環境の作成	1
2.2. 仮想環境へのインストール	3
3. Superset の初期設定	5
3.1. 依存ライブラリ入替 (Flask フレームワーク他)	5
3.2. 環境の初期化	6
3.3. Superset 開発用サーバの起動	6
3.4. ログイン	7
3.5. 日本語化	8
4. 機能、使い方	9
4.1. データベースに SQLite3 を使う場合の設定	9
4.2. 使い方の手順	10
4.3. ユーザの追加とロール/許可の付与	10
4.4. データベースの登録	11
4.5. データセットの登録	12
4.5.1. SQL Lab-SQL Editor	12
4.5.2. データ：時間	13
4.6. チャートの作成	15
4.7. ダッシュボード	18
5. BI/Apache Superset の用途について	19

Business Intelligence (Apache Superset)

はじめに

BI(Business Intelligence)という言葉は DSS(意思決定支援システム)と共に 1980 年代には既に使われていて、基盤製品の一つに Adabas というデータベースとそれを自然言語(英語)で操作する Natural がありました。この DBMS はデータを複合条件で検索できる優れたもので、RDB や SQL よりも時期的に先行して広まりました(RDB と同じ用途で使われ、2020 年代にも生き残っています)。しかし、そこから BI が日本で一般化したかというところでもありません。当時の大型汎用機は 32 ビット CPU に主記憶容量が最大 1 ギガバイト、24 ビットアドレッシングと現代から見れば非力なものでしたが、現代と比べればデータ量が少なかったこともあり、アプリ(キー項目を固定した API を使っていました)の処理速度が問題になることはなく、むしろ検索条件の複雑さのわりに応答が早いという印象でした。

BI は ETL、DWH(Data Ware House)と一緒に語られることが多いのですが、既に存在するデータを加工・表示するツールとして使えばプログラムレスでデータ利用環境を作ることができます。

1. BI ツール

BI ツールも ETL ツール/フレームワークと同様に沢山のオープンソースによる開発プロジェクトがあります¹。企業が開発の中核となっている帳票定義や出力管理までできる製品レベルのものもありますが、企業色が少なく開発体制がオープンなプロジェクト Apache Superset (Airbnb 社が開発して OSS に提供、Apache License 2.0=商用利用可) ver2.0.0 を Windows10 で利用する例を説明します。

2. Apache Superset のインストール

公式サイト²では「Apache Superset は認証関連の暗号化処理が OS に依存しており Linux の各種ディストリビューションと Mac OS X がサポート対象」としていますが、Windows での「試し方」が記載されているのでその手順に沿って作業をすれば開発環境を作ることができます。

2.1. Python 仮想環境の作成

Apache Superset 2.0.0 は Python 3.8 (または 3.7) でないと依存ライブラリがインストールできないのでプロジェクト固有の環境(仮想環境)にインストールして Python のバージョンや依存ライブラリ入れ替えが他のプロジェクトに影響しないようにします。

(1) Python 3.8 のインストール

ダウンロードサイト³にあるバージョン 3.8 の最新のインストーラを使います。

異なるバージョンを共存させることが可能なので、カスタムインストールで新しいディレクトリをインストール先にします。バージョンの使い分けは Windows の場合は py コマンドで行います。

<インストール例>

```
C:¥Tools¥Python
├─Python310
└─Python38
```

¹ GitHub #BI <https://github.com/topics/bi>

² Superset インストールと構成 <https://apache.github.io/superset/installation.html>

³ Python Releases for Windows <https://www.python.org/downloads/windows/>

Business Intelligence (Apache Superset)

(2) venv コマンド

以下の手順で、仮想環境を作ります。

- ① プロジェクトのディレクトリを作成 (名前は適宜に...)
- ② 作ったディレクトリに移動する
- ③ Python の venv モジュールを実行 (パラメータは仮想環境の設定を格納するディレクトリ名)

> py -3.8 -m venv 仮想環境設定ディレクトリ名 (先頭のピリオド . は Linux では隠しファイルになります)

※ Python の Windows 用起動コマンド py のオプションで使用するバージョン 3.8 を指定します

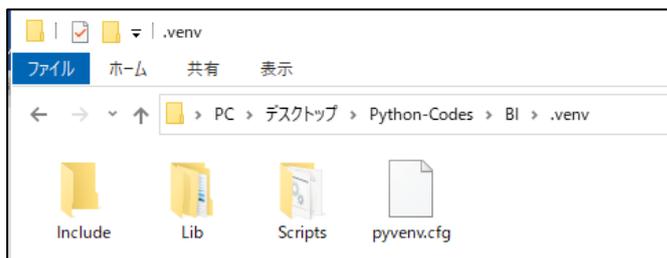
・パラメータで指定したディレクトリは新しく作られ、仮想環境用の資材が格納されます

【仮想環境作成例】

コマンドプロンプト

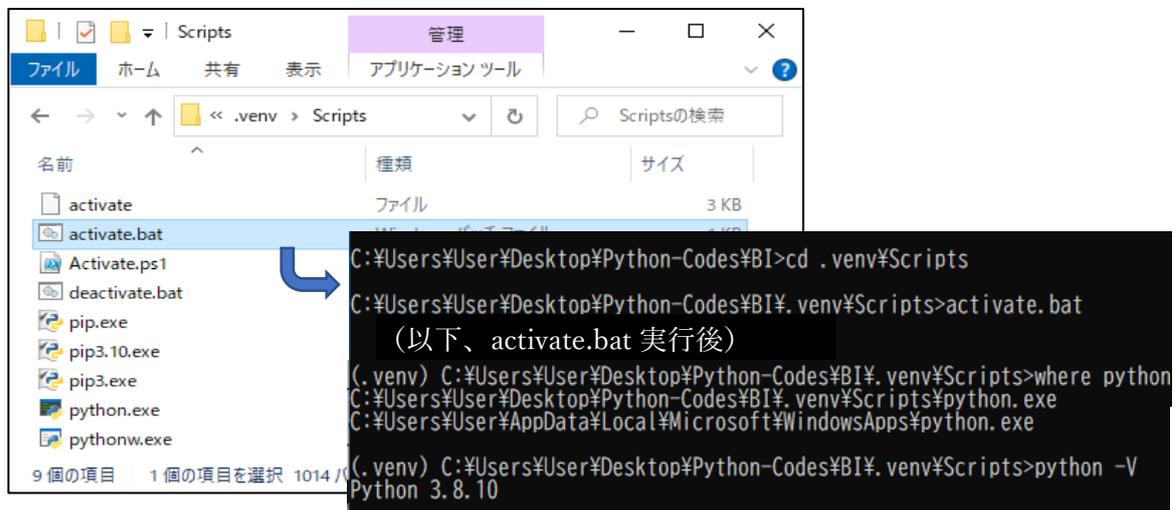
```
C:¥Users¥User>cd C:¥Users¥User¥Desktop¥Python-Codes
C:¥Users¥User¥Desktop¥Python-Codes>mkdir BI ①プロジェクト作成 (名前は適宜)
C:¥Users¥User¥Desktop¥Python-Codes>cd BI ②プロジェクトに移動
C:¥Users¥User¥Desktop¥Python-Codes¥BI>py -3.8 -m venv .venv ③v3.8 のvenv を実行
C:¥Users¥User¥Desktop¥Python-Codes¥BI>
```

【venv で作られた設定ディレクトリ (上記例の場合 .venv) の中身】



(2) 仮想環境の起動コマンド

venv が作った 環境設定ディレクトリ¥Scripts の中の activate.bat をコマンドプロンプトから実行すると環境変数の PATH に 仮想環境設定ディレクトリ¥Script が追加され、プロンプトの表示が変わり、Python バージョン 3.8 の仮想環境が起動します。



Business Intelligence (Apache Superset)

2.2. 仮想環境へのインストール

(1) pip の最新化

以下のコマンドで仮想環境のセットアップツールを最新化します。

```
> py -m pip install --upgrade setuptools pip
```

※ 新しいものがあると、インストールして既存のモジュールを置き換えられます

【実行例：表示されるバージョン等は環境により異なります】

```

(.venv) C:\Users\User\Desktop\Python-Codes\BI>python -V
Python 3.8.10

(.venv) C:\Users\User\Desktop\Python-Codes\BI>py -m pip install --upgrade setuptools pip
Requirement already satisfied: setuptools in c:\users\user\Desktop\Python-Codes\BI\venv\lib\site-packages (56.0.0)
Collecting setuptools
  Using cached setuptools-65.4.1-py3-none-any.whl (1.2 MB)
Requirement already satisfied: pip in c:\users\user\Desktop\Python-Codes\BI\venv\lib\site-packages (21.1.1)
Collecting pip
  Using cached pip-22.2.2-py3-none-any.whl (2.0 MB)
Installing collected packages: setuptools, pip
  Attempting uninstall: setuptools
    Found existing installation: setuptools 56.0.0
    Uninstalling setuptools-56.0.0:
      Successfully uninstalled setuptools-56.0.0
  Attempting uninstall: pip
    Found existing installation: pip 21.1.1
    Uninstalling pip-21.1.1:
      Successfully uninstalled pip-21.1.1
Successfully installed pip-22.2.2 setuptools-65.4.1

(.venv) C:\Users\User\Desktop\Python-Codes\BI>
  
```

(2) Apache Superset のインストール

仮想環境で Apache Superset のインストールを行います。

```
> py -m pip install apache-superset
```

途中、python-geohash のインストールで以下のようなエラーが出たらマイクロソフト社のサイト⁴から Microsoft C++ Build Tools をダウンロードしてインストールします。

```

Running setup.py install for python-geohash ... error
error: subprocess-exited-with-error

× Running setup.py install for python-geohash did not run successfully.
  exit code: 1
  → [14 lines of output]
    running install
    C:\Users\User\Desktop\Python-Codes\BI\venv\lib\site-packages\setuptools\command\install.py:34: SetuptoolsDeprecationWarning: setup.py install is deprecated. Use build and pip and other standards-based tools.
      warnings.warn(
    running build
    running build_py
    creating build
    creating build\lib.win-amd64-cpython-38
    copying geohash.py -> build\lib.win-amd64-cpython-38
    copying quadtree.py -> build\lib.win-amd64-cpython-38
    copying jgrid.py -> build\lib.win-amd64-cpython-38
    copying jpiarea.py -> build\lib.win-amd64-cpython-38
    running build_ext
    building 'geohash' extension
    error: Microsoft Visual C++ 14.0 or greater is required. Get it with "Microsoft C++ Build Tools": https://visualstudio.microsoft.com/visual-cpp-build-tools/
    [end of output]

note: This error originates from a subprocess, and is likely not a problem with pip.
error: legacy-install-failure
  
```

⁴ Microsoft C++ Build Tools <https://visualstudio.microsoft.com/ja/visual-cpp-build-tools/>

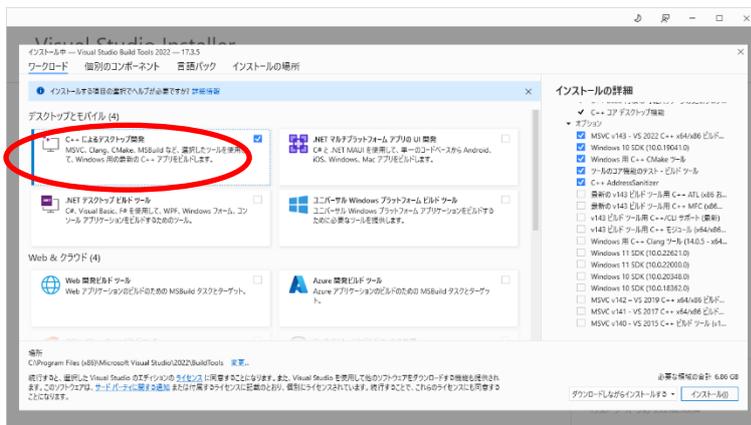
Business Intelligence (Apache Superset)

【Microsoft C++ Build Tools ダウンロードサイト】



【Microsoft C++ Build Tools のインストール画面】

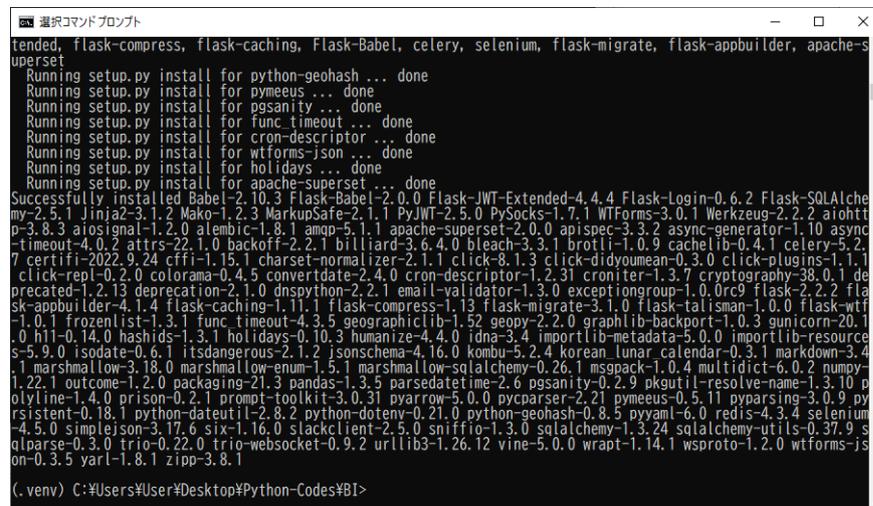
「C++によるデスクトップ開発」を選択します。(画面右に出るオプションはデフォルトのまま)



※ インストールは数十分かかりますが、以下の画面が出て完了です



※ ビルドツールのインストールが終わったら再度 `py -m pip install apache-superset` を実行します



※ この画面で apache-superset のインストールが終了です

Business Intelligence (Apache Superset)

3. Superset の初期設定

Apache Superset (以下、Superset) のインストールが終わると仮想環境設定ディレクトリに実行ファイルが格納されますが、更にライブラリの入替等や管理 DB の初期化が必要です。

3.1. 依存ライブラリ入替 (Flask フレームワーク他)

Superset は Flask というフレームワークを使った Web アプリケーションで、Flask (及び依存ライブラリ) のバージョンが合わないとエラーが発生して動作しません。Superset のインストールで Flask をはじめとする依存ライブラリは最新版がインストールされますが、一部のライブラリはバージョンを戻す必要があるかもしれません⁵。この資料は、以下の組合せで実行しています。

```
(. venv) C:\Users\User\Desktop\Python-Codes\BI\.venv\Scripts>superset --version  
Python 3.8.10  
Flask 2.1.3  
Werkzeug 2.0.3
```

<設定時に置換したライブラリ>

Superset 2.0.0 を Python 3.8.10 の環境にインストールした際に発生したエラーと入れ替えたライブラリは以下のものです。今後も依存ライブラリ側の更新により旧バージョンへの固定が必要なものが発生する可能性があり、また一方で Superset 側の上位バージョンへの対応も進んでいます。

[エラー事象]

```
ModuleNotFoundError: No module named 'werkzeug.wrappers.etag'
```

[対象ライブラリと入れ替えコマンド]

```
pip uninstall -y Werkzeug  
pip install Werkzeug==2.0.3
```

[エラー事象]

```
TypeError: _init_() got an unexpected keyword argument 'unbound_message'
```

[対象ライブラリと入れ替えコマンド]

```
pip uninstall -y flask  
pip install flask==2.1.3
```

[エラー事象]

```
ModuleNotFoundError: No module named 'wtforms.ext'
```

[対象ライブラリと入れ替えコマンド]

```
pip uninstall -y wtforms  
pip install wtforms==2.3.0
```

⁵ Superset2.0.0 の依存ライブラリ (随時更新されます)

<https://github.com/apache/superset/blob/master/requirements/base.txt>

Business Intelligence (Apache Superset)

3.2. 環境の初期化

Superset は利用（管理）者の認証情報等をデータベースで管理しています。以下のコマンドで初期化を行います。

- ① データベースの初期化 > `superset db upgrade`
- ② 管理者アカウントの登録 > `set FLASK_APP=superset` …Flask が呼出すライブラリの指定
> `superset fab create-admin`

Username [admin]: ~ Email [admin@fab.org]: までは、Enter キーでデフォルト値にした例です。
Password: は入力必須ですが非表示で、確認と合わせて 2 回入力が求められます。

```

C:\Users\User\Desktop\Python-Codes\BI\venv\Scripts>flask fab create-admin
Username [admin]:
User first name [admin]:
User last name [user]:
Email [admin@fab.org]:
Password:
Repeat for confirmation:

-----
WARNING
-----
A Default SECRET_KEY was detected, please use superset_config.py to override it.
Use a strong complex alphanumeric string and use a tool to help you generate
a sufficiently random sequence, ex: openssl rand -base64 42
-----

```

(中略)

```

EXPLORE_FORM_DATA_CACHE_CONFIG: It is recommended to use `RedisCache`, `MemcachedCache` or another dedicated caching
backend for production deployments
2022-10-11 14:10:18,777:WARNING:superset.utils.cache_manager:Falling back to the built-in cache, that stores da
ta in the metadata database, for the following cache: `EXPLORE_FORM_DATA_CACHE_CONFIG`. It is recommended to us
e `RedisCache`, `MemcachedCache` or another dedicated caching backend for production deployments
No PIL installation found
2022-10-11 14:10:18,980:INFO:superset.utils.screenshots:No PIL installation found
Recognized Database Authentications.
Admin User admin created.
(.venv) C:\Users\User\Desktop\Python-Codes\BI\venv\Scripts>

```

- ③ サンプルデータの登録（任意） > `superset load examples` …数分かかります
- ④ デフォルトの役割、許可を設定 > `superset init`

※ ここまでの作業で Superset のサーバを起動できる状態になります

3.3. Superset 開発用サーバの起動

Python 仮想環境で環境変数 FLASK_APP に“superset”が設定された状態であれば、superset run で開発用のサーバが起動します。デフォルトのポート番号は 8088 (-p オプションで変更が可能) です。

[確認] プロンプトに仮想環境名 (.venv) 等が表示されていない状態のとき、activate を実行します

> <Python 仮想環境設定ディレクトリ>\Scripts\activate.bat

[確認] 環境変数 FLASK_APP が設定されていなかったら、“superset”を設定します

> `set FLASK_APP=superset`

開発用サーバをポート 8088 で起動

> `superset run -p 8088 --with-threads --reload --debugger`

Business Intelligence (Apache Superset)

開発時の環境立ち上げをバッチファイルで表すと、以下のようになります。

このスクリプト例では、②のバッチファイルを実行すると環境設定⇒サーバ起動が連続して行われ(数十秒かかります)、①を実行すると Python 仮想環境が立ち上がって操作できる状態のウインドが表示されます。

① Python 仮想環境立ち上げ

【ファイル名：Superset-venv.bat】

```
rem i. 仮想環境立ち上げスクリプトを同一プロセス内(call)で実行
rem ii. Superset 用の環境変数を設定する
rem iii. パラメータ無で実行された場合は、cmd /k で開いたウインドを残す
rem -----
call <Python 仮想環境設定ディレクトリ>%Scripts%activate.bat
set FLASK_APP=superset

if "%1"==" " cmd /k
```

② Superset サーバ起動

```
rem i. 同一ディレクトリにある仮想環境立ち上げバッチを同一プロセス内で実行
rem ii. Superset サーバを実行する
rem iii. サーバ起動後も cmd /k で開いたウインドを残す
rem -----
call .%Superset-venv.bat dummypara
superset run -p 8088 --with-threads --reload --debugger
cmd /k
```

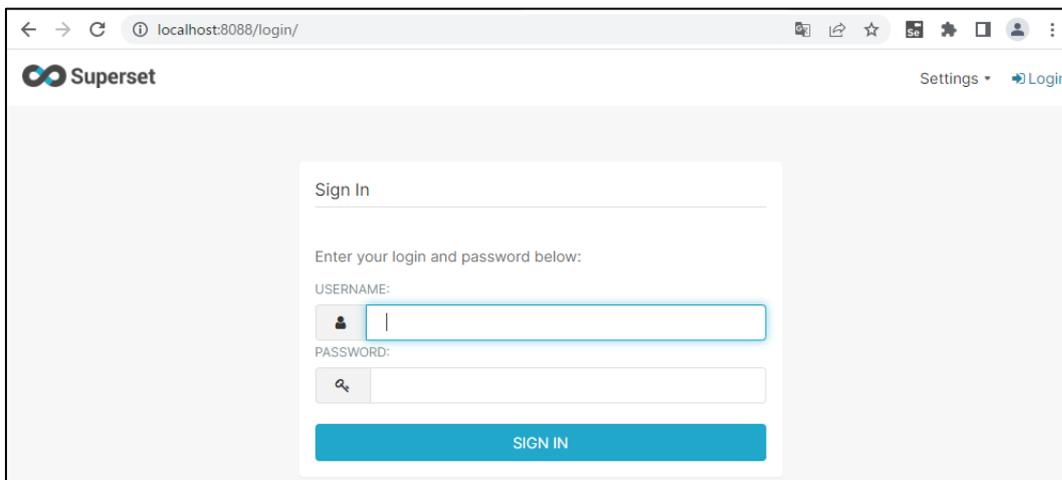
3.4. ログイン

Superset の起動が完了したら、ブラウザから起動ポートにアクセスしてログインします。

サーバがデフォルトのポートで起動しているホストでログインする場合は、アドレスは以下のようになります。

<http://localhost:8088>

ログインには、環境の初期化で登録した管理者アカウントを使います。



※ 初期状態では言語の選択等は行えません

Business Intelligence (Apache Superset)

3.5. 日本語化

Superset は日本語のリソースが同梱されていますが、設定ファイルを書き換えないと選択メニューも表示されず使えません。

【Superset 設定ファイルのパス】

<Python 仮想環境設定ディレクトリ>%Lib¥site-packages¥superset¥config.py

〔変更前〕

```

312 # -----
313 # Babel config for translations
314 # -----
315 # Setup default language
316 BABEL_DEFAULT_LOCALE = "en"
317 # Your application default translation path
318 BABEL_DEFAULT_FOLDER = "superset/translations"
319 # The allowed translation for you app
320 LANGUAGES = {
321     "en": {"flag": "us", "name": "English"},
322     "es": {"flag": "es", "name": "Spanish"},
323     "it": {"flag": "it", "name": "Italian"},
324     "fr": {"flag": "fr", "name": "French"},
325     "zh": {"flag": "cn", "name": "Chinese"},
326     "ja": {"flag": "jp", "name": "Japanese"},
327     "de": {"flag": "de", "name": "German"},
328     "pt": {"flag": "pt", "name": "Portuguese"},
329     "pt BR": {"flag": "br", "name": "Brazilian Portuguese"},
330     "ru": {"flag": "ru", "name": "Russian"},
331     "ko": {"flag": "kr", "name": "Korean"},
332     "sk": {"flag": "sk", "name": "Slovak"},
333     "sl": {"flag": "si", "name": "Slovenian"},
334     "nl": {"flag": "nl", "name": "Dutch"},
335 }
336 # Turning off i18n by default as translation in most languages are
337 # incomplete and not well maintained.
338 LANGUAGES = {}
339

```

〔変更後〕

- の部分を書き換えます。①デフォルト言語設定 BABEL_DEFAULT_LOCALE を ja(日本語)に、
②LANGUAGE を空オブジェクトに書き換えているのを無効にします。

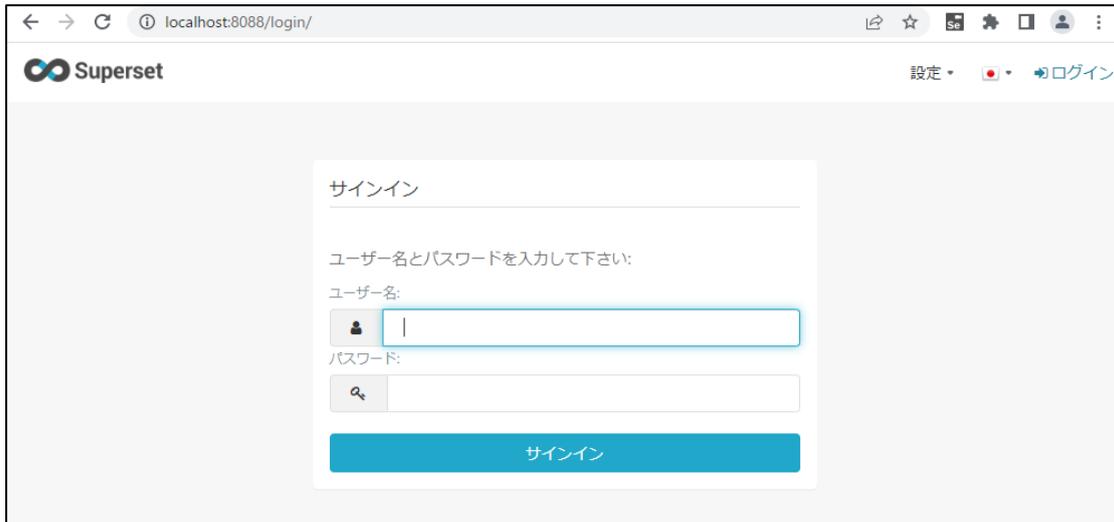
```

312 # -----
313 # Babel config for translations
314 # -----
315 # Setup default language
316 BABEL_DEFAULT_LOCALE = "ja"
317 # Your application default translation path
318 BABEL_DEFAULT_FOLDER = "superset/translations"
319 # The allowed translation for you app
320 LANGUAGES = {
321     "en": {"flag": "us", "name": "English"},
322     "es": {"flag": "es", "name": "Spanish"},
323     "it": {"flag": "it", "name": "Italian"},
324     "fr": {"flag": "fr", "name": "French"},
325     "zh": {"flag": "cn", "name": "Chinese"},
326     "ja": {"flag": "jp", "name": "Japanese"},
327     "de": {"flag": "de", "name": "German"},
328     "pt": {"flag": "pt", "name": "Portuguese"},
329     "pt BR": {"flag": "br", "name": "Brazilian Portuguese"},
330     "ru": {"flag": "ru", "name": "Russian"},
331     "ko": {"flag": "kr", "name": "Korean"},
332     "sk": {"flag": "sk", "name": "Slovak"},
333     "sl": {"flag": "si", "name": "Slovenian"},
334     "nl": {"flag": "nl", "name": "Dutch"},
335 }
336 # Turning off i18n by default as translation in most languages are
337 # incomplete and not well maintained.
338 LANGUAGES = {}
339

```

Business Intelligence (Apache Superset)

Superset を再起動すると画面右上に言語選択が表示され、日本語リソースが登録されている部分は日本語で表示されます。



4. 機能、使い方

Supersrt はレポート作成機能を開発中のようですが、ver 2.0.0 には含まれていません。このバージョンで実行できるのは、各種のデータソースから複数の表やグラフを表示するダッシュボードを作成する機能です。データソースは SQLite3、CSV ファイル以外を使う場合はそれぞれのドライバーが必要になるので別途インストールをしてください。また、SQLite3 を使う場合は Superset 設定ファイルの修正が必要になります。

4.1. データベースに SQLite3 を使う場合の設定

SQLite3 はドライバが標準で入っていて小規模な試行等に便利なのですが、Superset 設定ファイルに使えなくする設定が入っています（セキュリティ上の理由による）。

PREVENT_UNSAFE_DB_CONNECTIONS を False にしてこの設定を無効に（使えるように）することができます。

```

1202 # Optional prefix to be added to all static asset paths when rendering the UI.
1203 # This is useful for hosting assets in an external CDN, for example
1204 STATIC_ASSETS_PREFIX = ""
1205
1206 # Some sqlalchemy connection strings can open Superset to security risks.
1207 # Typically these should not be allowed.
1208 PREVENT_UNSAFE_DB_CONNECTIONS = False # デフォルトは True
1209
1210 # Path used to store SSL certificates that are generated when using custom certs
1211 # Defaults to temporary directory.
1212 # Example: SSL_CERT_PATH = "/certs"
1213 SSL_CERT_PATH: Optional[str] = None
1214

```

Business Intelligence (Apache Superset)

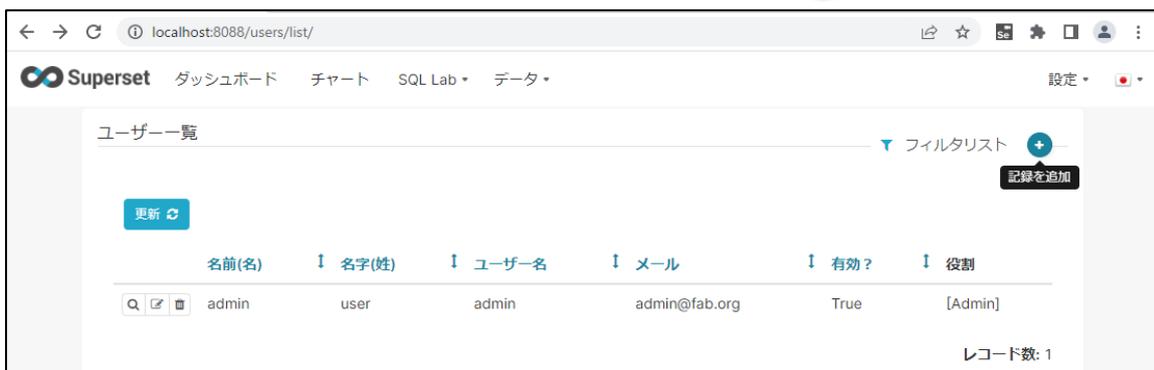
4.2. 使い方の手順

Superset は以下のメニューから機能を作成します。利用者の役割により使える機能が異なります。

- ① データ-データベース …データベースへの接続文字列を登録
- ② データ-データセット / SQL Lab-SQL Editor …表やグラフの基になるテーブルやクエリを登録
- ③ チャート …表やグラフの作成条件を登録します
- ④ ダッシュボード …表やグラフを配置する画面構成や自動更新間隔を設定

4.3. ユーザの追加とロール / 許可の付与

環境を複数人で共用し、開示範囲の制限や役割分担が必要であれば個人毎の利用者登録が必要になります。画面右上メニュー「設定」⇒「ユーザー一覧」を選び、を押下します。



※役割（ロール）は Admin、Public、Alpha、Gamma、granter、sql_lab が予め用意されていて、利用者は複数の役割を持つことができます

メニュー「設定」⇒「ロール一覧」から各ロールが持つ**許可**が確認できます

Alpha : データセットを追加および変更できます

Gamma : ダッシュボードを作成することはできません

Public ; 予め与えられた許可はありません。個別に追加する必要があります

[ユーザの追加画面から「保存」すると...]

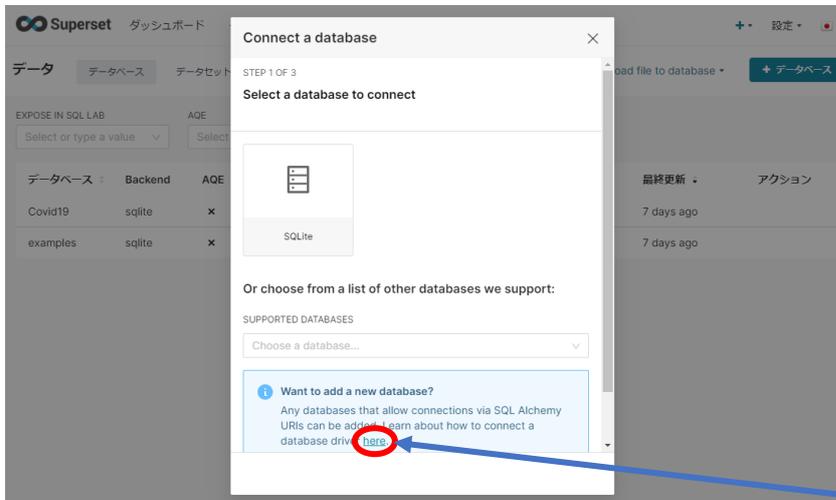


※ユーザー名でログインできるようになります。役割で与えられた許可にない操作はメニューに表示されていても実行すると「An error occurred while …」や「…中に問題が発生しました: Xxx.gl」といった通知が表示されます

Business Intelligence (Apache Superset)

4.4. データベースの登録

メニュー「データ」→「データベース」を選択し、画面右上の **+ データベース** を押下します。



SQLite は初期状態から選択可能です。それ以外の DBMS を使いたい場合は、リンク [here](#) へ SQLite を選択した場合は、以下の画面に切り替わるので...

- ・ BASIC タブで、表示名と URI に DB ファイルのパスを設定します⁶
- ・ ADVANCED タブ（下図-右）は、CSV ファイルのアップロード機能を使いたい場合のみ Security 項目の Allow data upload をチェックします

⁶ <使用データ>以降の説明は「ETL抽出-変換-書出 (Python,petl)」Extract-Transform-Load.pdf で作ったデータベースを前提に使用します。テーブルの内容等は当該資料を参照してください

Business Intelligence (Apache Superset)

4.5. データセットの登録

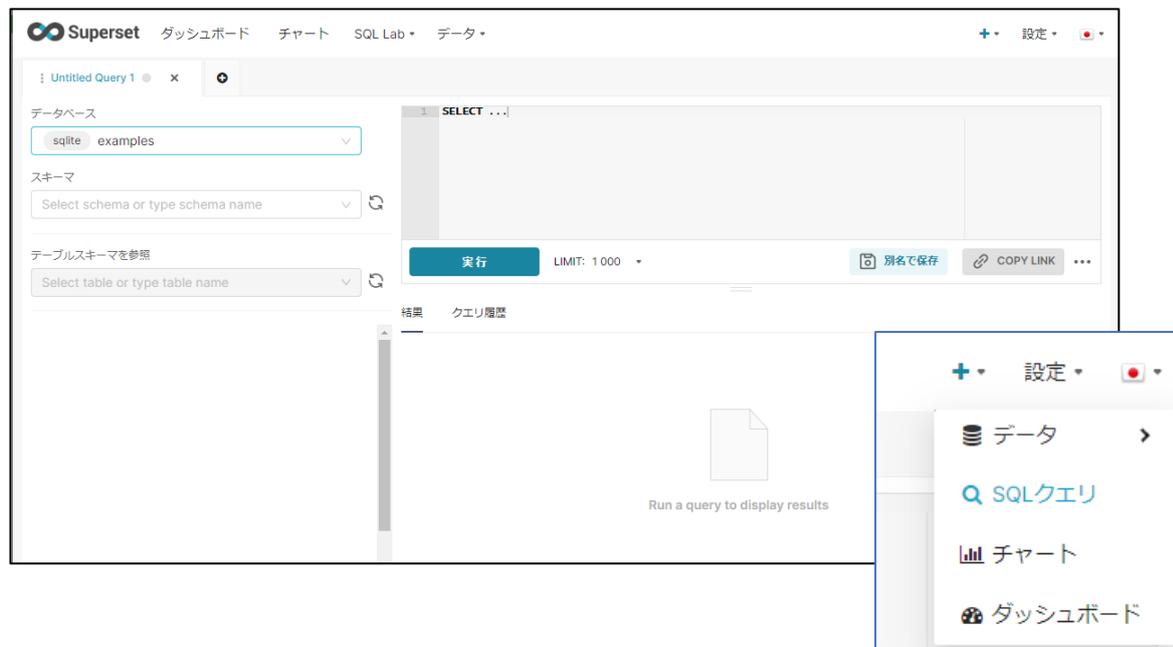
チャート（グラフや表）はデータセットに定義されたデータから作られます。データセットはRDBのテーブル/ビュー（Physical）かSQL（Virtual）を指定します。

Superset 2.0では「データセットを追加」からはSQLの登録ができないので、SQLの関数やその他の機能を使いたい場合はメニューのSQL Lab-SQL Editorで作成します。



4.5.1. SQL Lab-SQL Editor

SQLの編集画面は、画面右上の+メニューの「SQLクエリ」からも呼び出せます。



Business Intelligence (Apache Superset)

4.5.2. データ：時間

チャートの可視化方式に Time-series ~ という名前が付いているものは時系列で表示するもので、この方式を使う場合は対象データに「時間」を持っている必要があります。データは DATETIME 型と認識されれば年月日時分秒の全てに有効データを持っている必要はありません。

時間項目は Superset がある程度自動的に認識することが可能で、例えば DBMS により DATE 型や TIMESTAMP 型を時間として扱えますが、SQLite のように日付、時刻を文字列型として管理する DBMS を使うや SQL で TIMESTAMP のデータを導出する場合は明示的に項目を指定する必要があります。

The screenshot shows the Apache Superset interface. On the left, the 'dailyTrace' table schema is visible with columns like Date, newconf, deathsc, severec, reqcare, and various count_* columns. The main area shows a SQL query:

```

1 SELECT
2   DATETIME(REPLACE(Date, '/', '-')) as 日付
3   , CAST(newconf as INTEGER) as 新規感染
4   , CAST(deathsc as INTEGER) as 死者
5   , CAST(severec as INTEGER) as 重症者
6 FROM dailyTrace
    
```

The results table shows data for the dates 2020-01-16, 2020-01-17, and 2020-01-18. A warning message indicates that only 1000 rows are returned due to a dropdown limit.

日付	新規感染	死者	重症者
2020-01-16 00:00:00	1	NULL	NULL
2020-01-17 00:00:00	0	NULL	NULL
2020-01-18 00:00:00	0	NULL	NULL

(画面右上 SQL 部分拡大)

クエリで DATETIME 関数を使っていますが、Superset は結果を文字列として認識します。また、パラメータの形式は ISO 8601 形式(yyyy-mm-dd)にする必要があります。

```

1 SELECT
2   DATETIME(REPLACE(Date, '/', '-')) as 日付
3   , CAST(newconf as INTEGER) as 新規感染
4   , CAST(deathsc as INTEGER) as 死者
5   , CAST(severec as INTEGER) as 重症者
6 FROM dailyTrace
    
```

(元データ：dailyTrace)

Date	newconf	deathsc	severec	reqcare	count_first_shot_general
2020/01/16	1	NULL	NULL	NULL	NULL
2020/01/17	0	NULL	NULL	NULL	NULL
2020/01/18	0	NULL	NULL	NULL	NULL

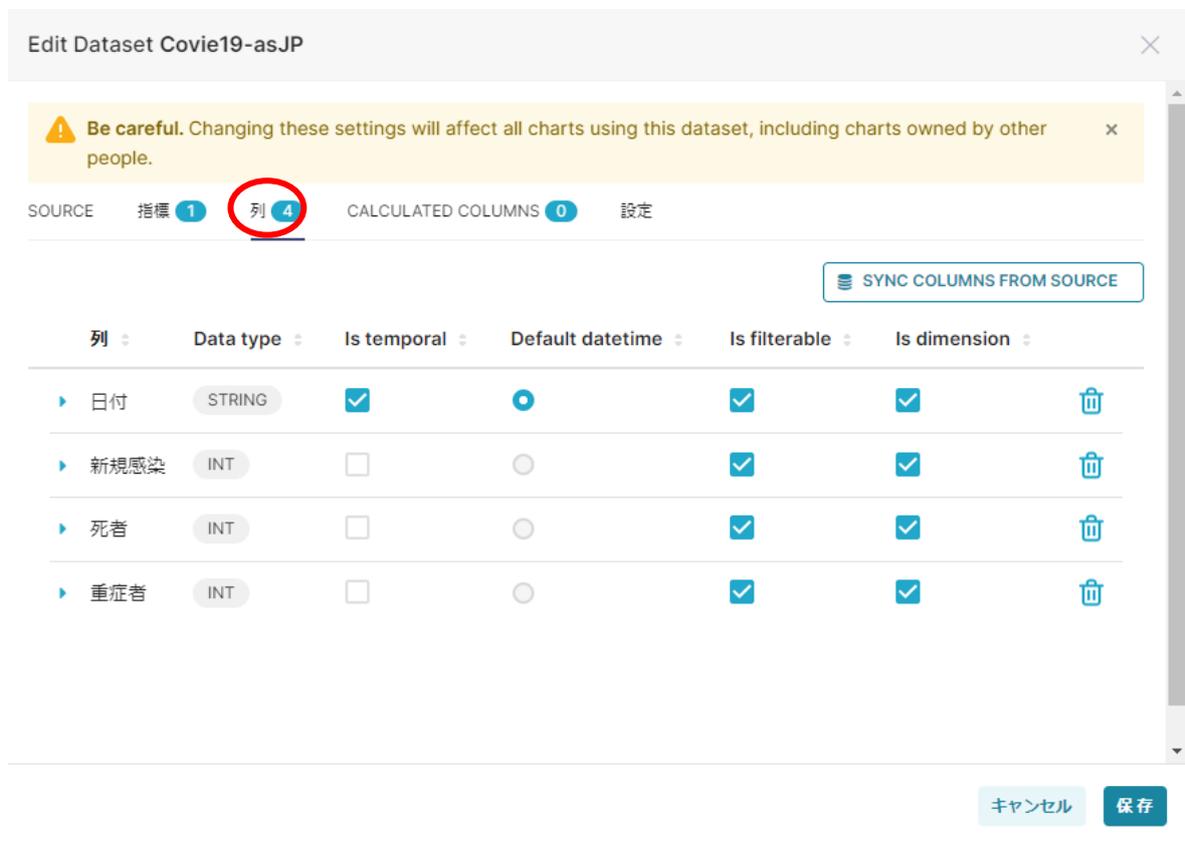
Business Intelligence (Apache Superset)

SQLを「別名で保存」で保存後、データ-データセットから編集を選びます。

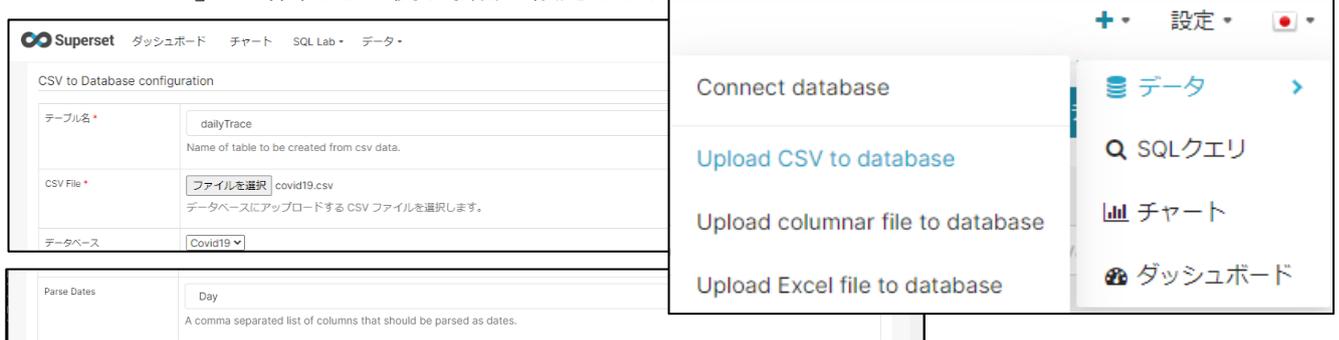


【列の属性設定-DATETIME】

画面のリンク「列」を選択し、時間として使う項目の Is temporal をチェック⇒Default datetime をチェックします。これで DATETIME として扱うことができるようになります。



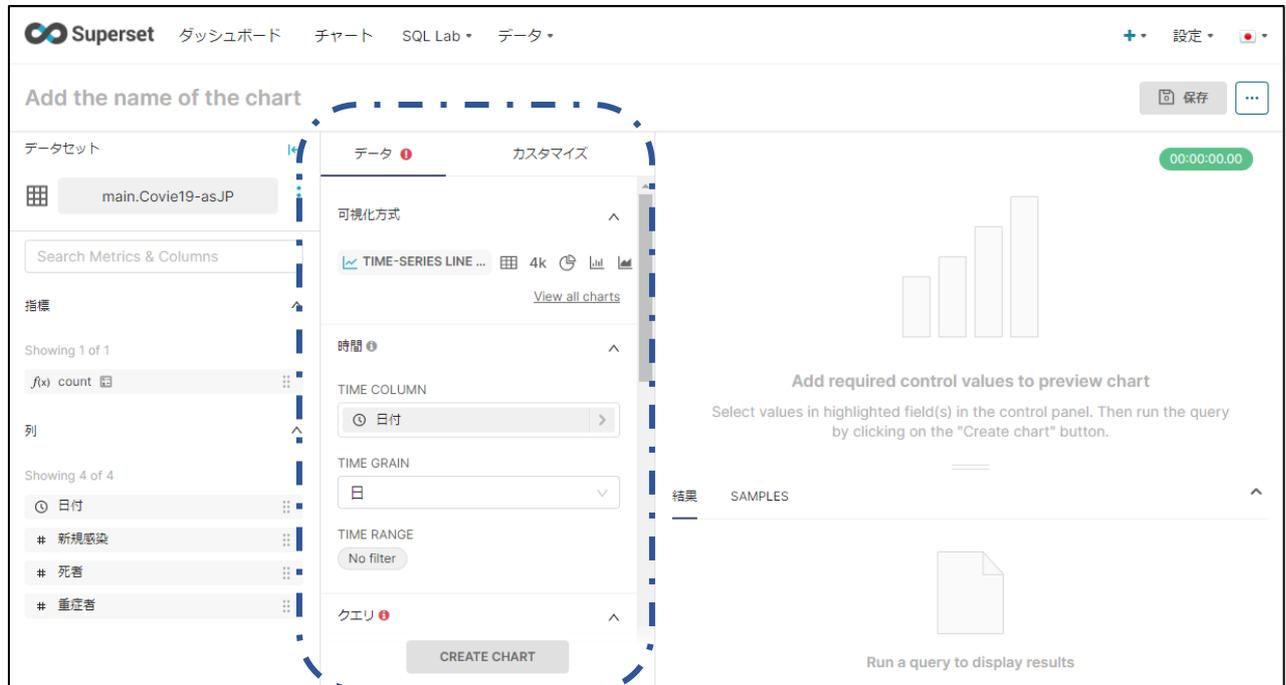
DATETIME 型に設定するもう一つの方法は、CSV ファイルからのアップロード機能を使う際に「Parse Dates」に時間として使う項目を設定します



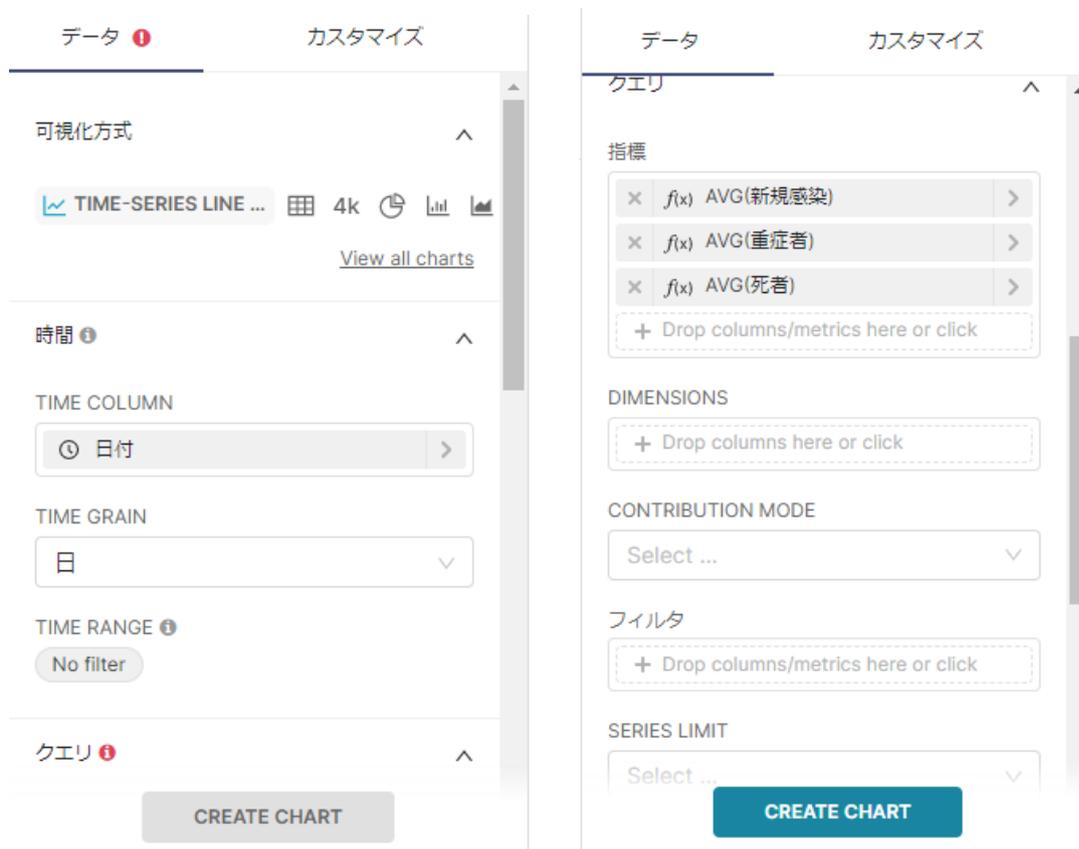
Business Intelligence (Apache Superset)

4.6. チャートの作成

メニューに「チャート」がありますが、データセットの一覧表示から目的のデータセットの名前のリンクを開くとチャート（グラフ、図、他）の作成ができます。



可視化方式でチャートの形式を選び「指標」に表示したい項目を追加すると「CREATE CHART」が活性化し、グラフが表示できます。（TIME COLUMN は DATETIME 項目が自動設定されるはず）



Business Intelligence (Apache Superset)



【可視化方式：テーブル】単純に表示するには「RAW RECORDS」を選び列を指定します。

データ カスタマイズ

可視化方式

テーブル 4k

時間

TIME COLUMN: 日付

TIME GRAIN: 日

TIME RANGE: No filter

クエリ

UPDATE CHART

データ カスタマイズ

QUERY MODE: AGGREGATE RAW RECORDS

列

- 日付
- # 新規感染
- # 死者
- # 重症者

フィルタ

ORDERING: Select ...

SERVER PAGINATION

UPDATE CHART

日付	新規感染	死者	重症者
2022-10-19	36110	46085	113
2022-10-18	43555	46014	125
2022-10-17	42294	45946	116
2022-10-16	15388	45890	118
2022-10-15	20884	45870	121

Business Intelligence (Apache Superset)

【可視化方式：棒グラフ (BAR CHART)】 グループングしてランキングする例

DIMENSIONS で指定した項目の値毎に「指標」の項目が集計されます。

```

1 SELECT
2   DATETIME(REPLACE(Date, '/', '-')) as 日付
3   , CAST(newconf as INTEGER) as 新規感染
4   , CAST(deathsc as INTEGER) as 死者
5   , CAST(severec as INTEGER) as 重症者
6   , CASE
7     WHEN count_fourth_shot_general > 0 THEN 4
8     WHEN count_third_shot_general > 0 THEN 3
9     WHEN count_second_shot_general > 0 THEN 2
10    WHEN count_first_shot_general > 0 THEN 1
11    ELSE 0
12  END as 接種回数
13  , CASE
14    WHEN count_fourth_shot_general > 0 THEN
15      count_fourth_shot_general - count_fourth_shot_general_wo_deceased
16    WHEN count_third_shot_general > 0 THEN
17      count_third_shot_general - count_third_shot_general_wo_deceased
18    WHEN count_second_shot_general > 0 THEN
19      count_second_shot_general - count_second_shot_general_wo_deceased
20    WHEN count_first_shot_general > 0 THEN
21      count_first_shot_general - count_first_shot_general_wo_deceased
22    ELSE CAST(deathsc as INTEGER)
23  END as 接種後死者
24 FROM dailyTrace
  
```



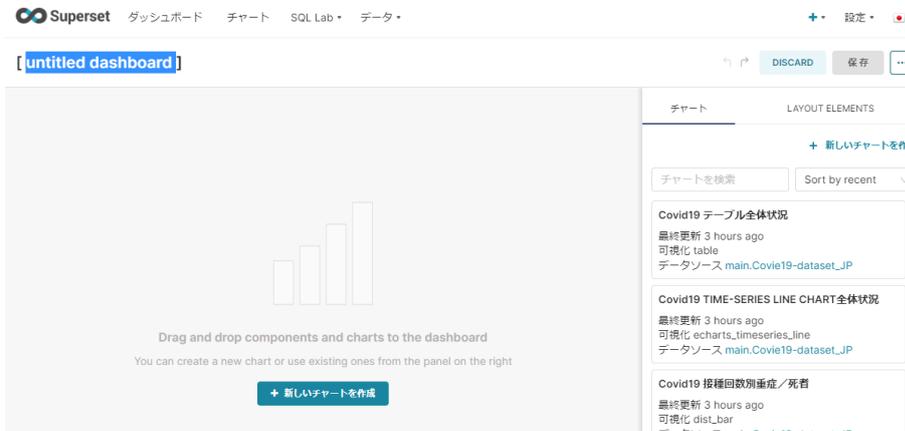
【注意】厚労省等のオープンデータを加工してグラフを作成していますが目的は DIMENSIONS の使い方を例示するため、以下理由により内容に意味はありません。厳密な情報が欲しい場合はより詳細なデータと分析が必要です

- ・新規感染や重症者、死者が感染した際のワクチン接種の有無、回数、時期を関係づけていません (接種開始時期と感染時期がおおよそ一致しているだけです)
- ・死者は死因を特定していません

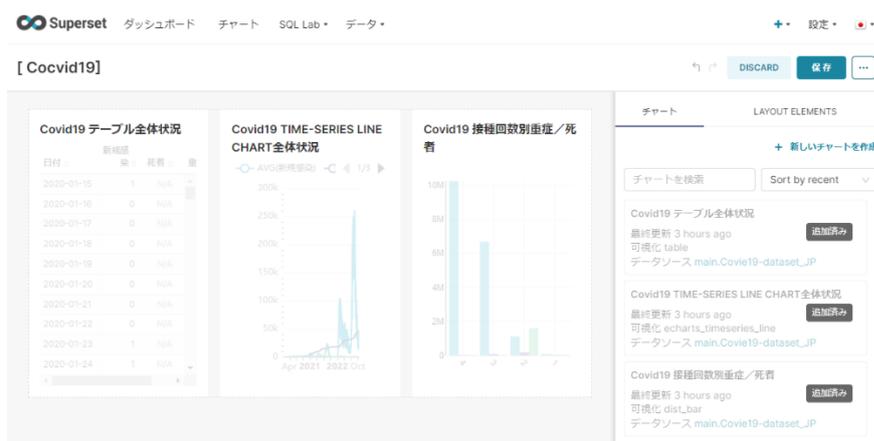
Business Intelligence (Apache Superset)

4.7. ダッシュボード

メニュー「ダッシュボード」を選択し、画面右上の **+ ダッシュボード** を押下します。
画面左上の[untitled dashboard]に名前を入れて、



画面右に表示されている作成済のチャート（前項参照）をドラッグ&ドロップしてダッシュボードに配置できます。



「LAYOUT ELEMENTS」から区切り線をドラッグしてダッシュボードの構成を決めてチャートを配置できます。



Business Intelligence (Apache Superset)

5. BI/Apache Superset の用途について

海外では BI に分類されるオープンソースが活発に開発されています。その中で Apache Superset は個人向けの旅行・宿泊施設のレンタルに関わる企業により開発されたので利用者の属性（性別、住所、年齢...）や移動先、季節、使用金額等の情報をグラフや世界地図へのマッピングで分析できるように作られています。一般に使われている表計算ソフトにも同様の機能がありますが、大量のデータを処理したり共用するには ETL や DWH と合わせてこのような BI ツールが必要になります。

これに比べて国内企業の場合は対象にする市場サイズが小さく顧客が均質で「はじめに」で書いたように DSS や BI による分析が必要でなかったのかもしれませんが、近いものに売れ筋/死に筋、バンドル効果等の商品分析や情報系がありますが、基幹システムのログ集計+ α にとどまり積極的に経営戦略や DSS 的な“分析”のための情報を集めている企業は少ないと推測します。

BI ツールとして Superset を紹介しましたが、ログを使った統計的な処理は既に AI のツールが出ていたので情報分析に使うというのは今更感があります。ただ、Superset のダッシュボードを情報共有に使える今まで使っていたファイルのような漏出の危険がなく、即時性を持って共有できるようになります。システム開発プロジェクトでいえば、お客様への定例報告資料を徹夜で作る作業が要らなくなり、リアルタイムの進捗データで状況の精査と言いつくを考える作業に集中できるようになるかもしれません。

以上