

ElasticSearch サーバーのデータ移行について

祖父江匠真

1 概要

今回は, 133.71.201.197 から 133.71.106.141 への ElasticSearch サーバー間のリサイクル館の太陽光パネルの測定データ移行と kibana を用いた可視化結果について報告する. また, 133.71.106.168 の ElasticSearch サーバーにあった movement_diary インデックスと movement_diary01 インデックスのドキュメントについて調査した結果も報告する.

2 データ移行手順について

データ移行を行う上で, ローカルマシンに JSON 形式でダンプしていたデータと, 133.71.201.197 の ElasticSearch サーバー上に存在するデータとの間で重複しているデータが一部存在しており, この重複データを取り除いた上でデータ移行を行う必要があった.

そこで一度, 移行元の ElasticSearch サーバーのデータをローカルマシンにエクスポートして, 重複データを取り除いた上で, 移行先の ElasticSearch サーバーにデータをアップロードした.

2.1 データのエクスポート

移行元の ElasticSearch サーバーのデータのローカルマシンへのエクスポートには, elasticdump [1] ライブラリを使用して JSON 形式でエクスポートした. その際, pcs_recyclekan という名前のインデックスのデータをエクスポートした.

2.2 データの重複削除

重複データの削除は, 予めローカルマシンに JSON 形式でダンプしていたデータと, elasticdump ライブラリを使用して JSON 形式でエクスポートしたデータを, utctime フィールドの値がユニークになるようにフィルタリングすることで行った.

2.3 データのインポート

重複データ削除後のデータを保存した JSON ファイルを読み出して、移行先の ElasticSearch サーバーにアップロードした。

その際、python の elasticsearch ライブラリを使用し、133.71.201.197 の ElasticSearch サーバーと同名の pcs_recyclekan という名前のインデックスに保存した。

3 kibana によるデータの可視化

移行後の pcs_recyclekan インデックスに保存されたデータを kibana を用いて可視化した。

横軸をタイムスタンプ (Jptime) とし、縦軸を日射量 (solarIrradiance(kw/m²)) としてプロットしたものを図 1 に示す。



図 1: 133.71.106.141 の pcs_recyclekan

次に、移行元である 133.71.201.197 の ElasticSearch サーバーの pcs_recyclekan インデックスに保存されたデータを図 2 に示す。



図 2: 133.71.201.197 の pcs_recyclekan

図 1, 図 2について, 2022 年 8 月以降のグラフの概形が一致していることが確認出来るので, データ移行作業は正常に行うことが出来たと判断できる.

4 movement_diary のデータについて

movement_diary と movement_diary01 は一部のフィールドを除いてほとんど同じデータを保有しているため, それぞれのインデックスのドキュメント数の比較と, タイムスタンプ情報を格納するフィールドをインデックス間の比較を行うことで, movement_diary が不要なインデックスであるかを調査した.

まず, movement_diary と movement_diary01 のドキュメント数を調べたところ, 同じ 142 件であった.

次に, movement_diary と movement_diary01 でタイムスタンプ情報を持つフィールドである dt_S フィールドの値同士を比較した.

ただし, dt_S フィールドが null のドキュメントが一部存在するので, その場合は inspection フィールドの値同士を比較した.

比較した結果, dt_S フィールドと inspection フィールドの両方が null である 3 件のドキュメントを除いて他全てのドキュメントは dt_S フィールドもしくは inspection フィールドの値が movement_diary と movement_diary01 の間で一致した.

dt_S フィールドと inspection フィールドの両方が null だった 3 件のドキュメントについても, 全てのフィールドにおいて, movement_diary01 のドキュメントが movement_diary のドキュメントの持つ情報を持っていたので, これらの調査結果から movement_diary インデックスは movement_diary01 インデックスで代替でき, 削除して良いインデックスであると判断した.

5 まとめ

今回は, 133.71.201.197 から 133.71.106.141 への Elasticsearch サーバー間のリサイクル館の太陽光パネルの測定データ移行と kibana を用いた可視化結果について報告した.

移行元のデータと移行後のデータを kibana で可視化して目視で比較したところ, 2022 年 8 月以降のグラフの概形が一致していることが確認出来たので, データ移行作業は正常に行うことが出来たと判断できる.

また, 133.71.106.168 の Elasticsearch サーバーにあった movement_diary インデックスと movement_diary01 インデックスのドキュメントについて調査した結果を報告した.

調査結果より, movement_diary01 インデックスは movement_diary インデックスの持つ全ての情報を保持しており, movement_diary インデックスは削除して良いインデックスであると判断した.

次週は, 133.71.201.197 の Elasticsearch サーバーにある pcs_recyclekan という名前のインデックス以外のインデックスについて調査を行い, 必要であればデータ移行を行う.

参考文献

- [1] Ferron H, "ElasticDump ", <https://github.com/elasticsearch-dump/elasticsearch-dump>, 参照 June 19,2023.