

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、前回に引き続き 133.71.106.168 から 133.71.106.141 への ElasticSearch サーバー間のデータ移行と kibana を用いた可視化結果について報告する。また、移行元 ElasticSearch サーバーにあった CO₂ データ以外のデータの移行についても報告する。

2 CO₂ データの移行作業

前回の報告書では、移行元 ElasticSearch サーバーの CO₂ データについて、JPtime が 2023 年より以前のドキュメントを 2022_co2 という名前のインデックスに保存し、JPtime が 2023 年のドキュメントを 2023_co2 という名前のインデックスに保存した。しかし、インデックスを分けることで 2023 年以前と 2023 年のデータを kibana で同じグラフにプロットして確認することが難しい可能性があることと、ElasticSearch サーバー間のデータ移行が正しく完了したか kibana で可視化することによって確認することが出来ないという問題があったため、今回は JPtime の値に関わらず co2 という名前のインデックスに保存した。保存先のインデックスが co2 であることを除くと、データの移行手順は前回のものと同様の手順で行った。

3 kibana によるデータの可視化

移行後の co2 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とし、縦軸を PPM, RH, TEMP としてそれぞれプロットしたものを図 1 ~ 図 3 に示す。

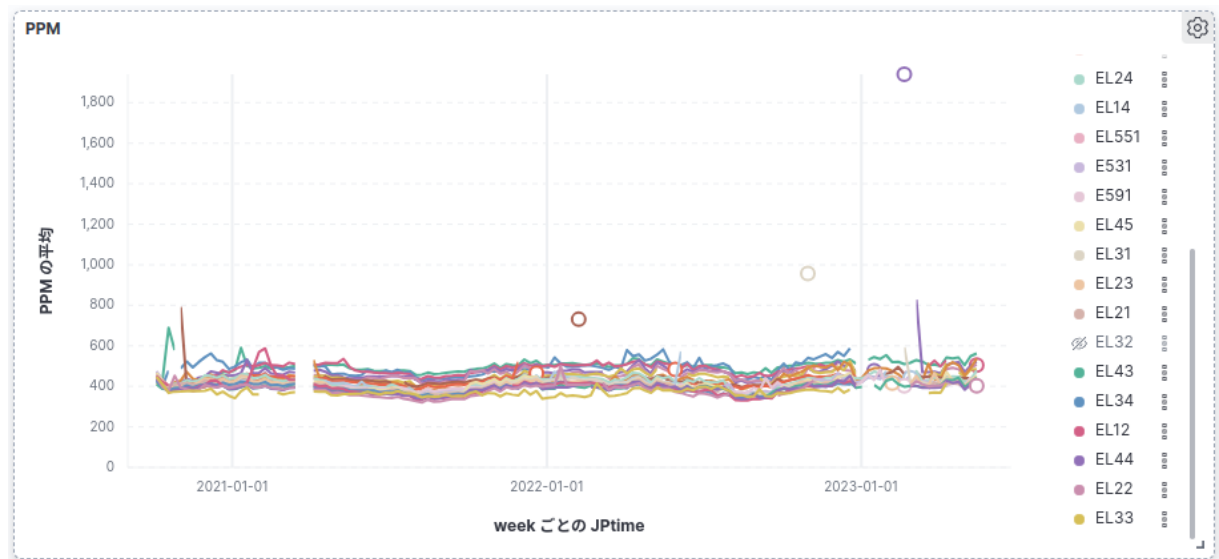


図 1: co2 の PPM

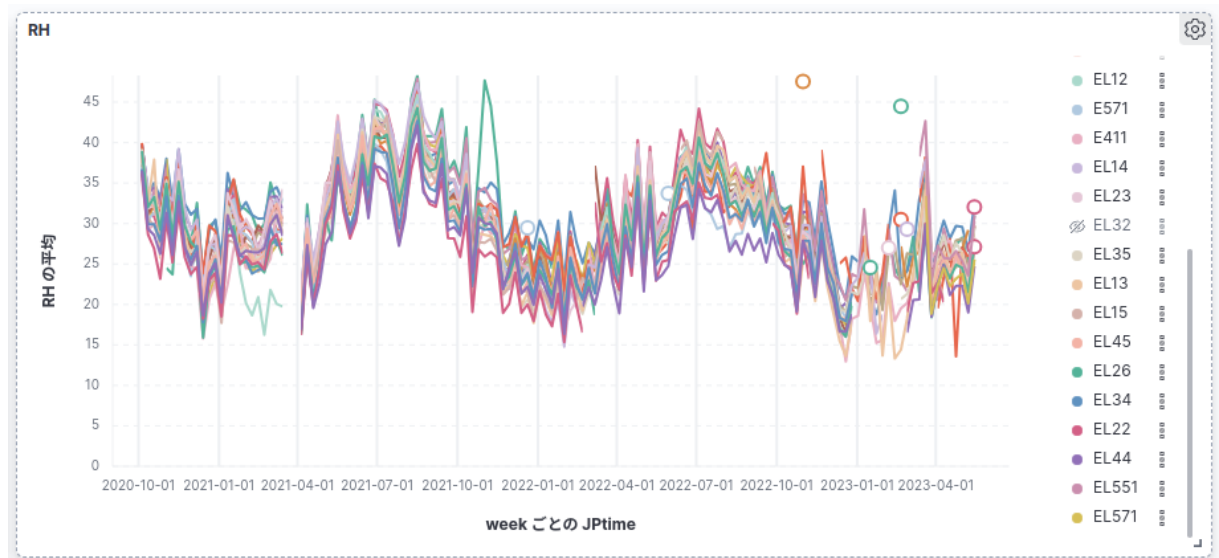


図 2: co2 の RH

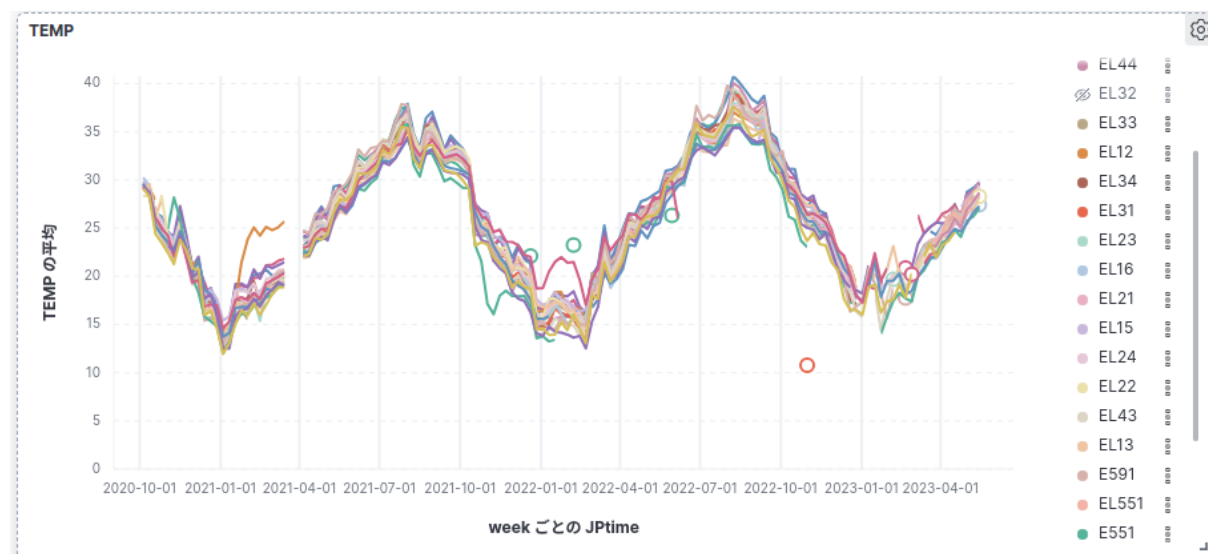


図 3: co2 の TEMP

図 1 ~ 図 3について, 2022 年と 2023 年の境目でデータが連続的に変化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 CO₂ データ以外のデータの移行について

移行元の Elasticsearch サーバーにはインデックス名に co2 という文字列を含むインデックス以外に以下のインデックスがあった.

- movement_diary
- movement_diary01
- temp2
- temp3
- test

これらのインデックスのデータ移行は, 同名のインデックスを移行先の Elasticsearch サーバーに作成して, 作成したインデックスにデータを挿入することで行った.

ただし, test インデックスは移行先の Elasticsearch サーバーに既に同名のインデックスが作成されていたので, 133.71.106.168.test というインデックス名にした.

次に, 上記のインデックスに保存されているデータについて説明する.

temp2, temp3 は TIME, TEMP, HUMI フィールドを持ったドキュメントが格納されており, temp2 のドキュメント数は約 150 件, temp3 は約 60 件であった. test は

Jptime, PPM, RH, TEMP, ip, number, utctime フィールドを持ったドキュメントを格納しているインデックスであり、ドキュメント数は95件であった。これらの情報から、temp2, temp3, test インデックスはCO₂データの収集の研究の中で動作確認目的に作成されたインデックスではないかと考えられる。

movement_diary と movement_diary01 は運行日誌に関する情報を格納したインデックスであり、それぞれ以下のような構造のドキュメントが保存されている。

Listing 1: movement_diary のドキュメントデータ

```
1  {
2    "_index": "movement_diary",
3    "_type": "_doc",
4    "_id": "OL65VHwBr1DnHOWC_d0U",
5    "_score": 1.0,
6    "_source": {
7      "pageNo": "1",
8      "type": "運行日誌",
9      "wether": "雨",
10     "driver": "仲村泰明",
11     "passenger": 2,
12     "destination": "新田高校・松山聖稜高校",
13     "dt_S": "2018-05-08T08:26:00",
14     "battery_S": 99,
15     "cruisingdistance_S": null,
16     "odometer_S": 15,
17     "out_charge": "していない",
18     "etc": "使用していない",
19     "etc_section": "",
20     "etc_budget": "",
21     "dt_R": "2018-05-08T11:50:00",
22     "battery_R": 94,
23     "cruisingdistance_AC": null,
24     "odometer_R": 30.0,
25     "meter": 15.0,
26     "start_charg_time": "2018-05-08T11:52:00",
27     "remark": "感じなかった",
28     "inspection": null,
29     "break_rest_Be": null,
30     "break_rest_Af": null,
31     "tire_rest_Be": null,
32     "tire_rest_Af": null,
33     "inspection_battery": null,
34     "charging_rate": null,
35     "health": null
36   }
37 }
```

Listing 2: movement_diary01 のドキュメントデータ

```

1  {
2    "_index": "movement_diary01",
3    "_type": "_doc",
4    "_id": "oMmqbnwBr1DnHOWCgH90",
5    "_score": 1.0,
6    "_source": {
7      "pageNo": "1",
8      "type": "運行日誌",
9      "wether": "雨",
10     "driver": [
11       "仲村泰明",
12       null,
13       null
14     ],
15     "destination": [
16       "新田高校",
17       "松山聖稜高校"
18     ],
19     "dt_S": "2018-05-08T08:26:00",
20     "battery_S": 99,
21     "cruisingdistance_S": null,
22     "odometer_S": 15,
23     "out_charge": "していない",
24     "charge_place": "",
25     "etc": "使用していない",
26     "etc_section": "",
27     "etc_budget": "",
28     "dt_R": "2018-05-08T11:50:00",
29     "battery_R": 94,
30     "cruisingdistance_AC": null,
31     "odometer_R": 30.0,
32     "meter": 15.0,
33     "start_charg_time": "2018-05-08T11:52:00",
34     "remark": "感じなかった",
35     "inspection": null,
36     "break_rest_Be": null,
37     "break_rest_Af": null,
38     "tire_rest_Be": null,
39     "tire_rest_Af": null,
40     "inspection_battery": null,
41     "charging_rate": null,
42     "health": null,
43     "battery_rate": 5,
44     "battery_rate_distance": 3.0
45   }

```

以下に movement_diary と movement_diary01 のドキュメントの違いを列挙する。

1. driver フィールド:

- movement_diary のドキュメントでは, driver フィールドは文字列である.
- movement_diary01 のドキュメントでは, driver フィールドは配列で, その中に文字列と 2 つの null 値が含まれている.

2. “destination” フィールド:

- movement_diary のドキュメントでは, “destination” フィールドは単一の文字列である.
- movement_diary01 のドキュメントでは, “destination” フィールドは配列で, その中に 2 つの文字列が含まれている.

3. “charge_place” フィールド:

- movement_diary のドキュメントには, “charge_place” フィールドは存在しない.
- movement_diary01 のドキュメントでは, “charge_place” フィールドが追加されているが, その値は空文字列である.

4. “battery_rate” フィールド:

- movement_diary のドキュメントには, “battery_rate” フィールドは存在しない.
- movement_diary01 のドキュメントでは, “battery_rate” フィールドが追加されており, その値は数値である.

5. “battery_rate_distance” フィールド:

- movement_diary のドキュメントには, “battery_rate_distance” フィールドは存在しない.
- movement_diary01 のドキュメントでは, “battery_rate_distance” フィールドが追加されており, その値は数値である.

5 まとめ

今回は, Elasticsearch サーバー間でのデータ移行と kibana を用いた可視化結果について報告した. すべての CO₂ データを co2 という名前のインデックスに格納して kibana で可視化した結果, 2022 年と 2023 年の境目でデータが連続的に変化していることが確認出来たので, データ移行は正しく行えたと判断した.

また, 移行元 Elasticsearch サーバーにあった CO₂ データ以外のデータの特徴と移行作業についても報告した.

参考文献

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