

133.71.201.197 の Elasticsearch サーバーのデータの移行について

祖父江匠真

1 概要

今回は, 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, 133.71.106.141 の Elasticsearch サーバーにデータ移行を行った.

2 133.71.201.197 の Elasticsearch サーバーにあるインデックスについて

図 1 に 133.71.201.197 の Elasticsearch サーバーにあるインデックスの一覧を示す.

インデックス管理							
インデックス管理ドキュメント							
インデックス データストリーム インデックステンプレート コンポーネントテンプレート							
Elasticsearch インデックスを個々に、または一斉に更新します。 詳細情報							
<input type="checkbox"/> <input checked="" type="checkbox"/> ローアップインデックスを含める <input type="checkbox"/> <input checked="" type="checkbox"/> 非表示のインデックスを含める							
<input type="text" value="検索"/> <input type="button" value="ライフサイクルステータス"/> <input type="button" value="ライフサイクルフェーズ"/> <input type="button" value="インデックスを再読み込み"/>							
<input type="checkbox"/> 名前	ヘルス	ステータス	プライマリ	レプリカ	ドキュメント数	ストレージサイズ	データストリーム
<input type="checkbox"/> pcs_test	● yellow	open	1	1	7715	1.4mb	
<input type="checkbox"/> pcs_recyclekan	● yellow	open	1	1	12995555	2.6gb	
<input type="checkbox"/> leaf_load	● yellow	open	1	1	3123	867.8kb	
<input type="checkbox"/> pcs_log	● yellow	open	1	1	1551	206.1kb	
<input type="checkbox"/> pcs_test4	● yellow	open	1	1	9436	3.6mb	
<input type="checkbox"/> leaf	● yellow	open	1	1	5070	1.1mb	
<input type="checkbox"/> leaf_grid	● yellow	open	1	1	1948	201.7kb	
ページごとの行数: 10 1							

図 1: 133.71.201.197 の Elasticsearch サーバーにあるインデックスの一覧

これらのインデックスが保存しているデータについて説明する.

- pcs_test
 - 恵村君がプログラムの検証目的で使用しているインデックス
- pcs_recyclekan

- リサイクル館の太陽光発電に関するデータを保存しているインデックス
- leaf_load
 - leaf のデータが保存されているインデックス
- pcs_log
 - リサイクル館の太陽光発電に関するデータを Elasticsearch にインサートする Python プログラムのログ情報を保存しているインデックス
- pcs_test4
 - 恵村君がプログラムの検証目的で使用しているインデックス
- leaf
 - leaf のデータが保存されているインデックス
- leaf_grid
 - leaf のデータが保存されているインデックス

pcs_test インデックスと pcs_test4 インデックスに関しては、検証用途で使っているものであるため、今回の移行対象からは除外し、pcs_log, leaf, leaf_load, leaf_grid インデックスのみを移行対象とした。

3 データ移行手順について

データ移行手順について、まず移行元の 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、移行先の Elasticsearch サーバーにデータをインサートした。

3.1 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_log, leaf, leaf_load, leaf_grid 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。

3.2 データのインポート

ダンプした JSON ファイルを作成した Python プログラムで読み込んで、Python の elasticsearch ライブラリを用いて移行先の Elasticsearch サーバーにインサートした。

移行先の Elasticsearch サーバーにおけるインデックス名については、133.71.201.197 の Elasticsearch サーバーと同名のインデックスに保存した。

4 データ移行が正常に行えたか確認

図 2 に移行元の Elasticsearch サーバーの leaf という文字列を含むインデックスのドキュメント数をカウントしたものを、図 3 に移行先の Elasticsearch サーバーの leaf という文字列を含むインデックスのドキュメント数をカウントしたものを示す。

図 2 と図 3 より、ドキュメント数が一致していることからデータ移行が正常に行えたと判断できる。

<input type="checkbox"/> 名前	ヘルス	ステータス	プライマリ	レプリカ	ドキュメント数
<input type="checkbox"/> leaf_load	● yellow	open	1	1	3123
<input type="checkbox"/> leaf	● yellow	open	1	1	5070
<input type="checkbox"/> leaf_grid	● yellow	open	1	1	1948

図 2: 133.71.201.197 の Elasticsearch サーバーの leaf という文字列を含むインデックスのドキュメントのカウント結果

<input type="checkbox"/> 名前	ヘルス	ステータス	プライマリ	レプリカ	ドキュメント数
<input type="checkbox"/> leaf_load	● green	open	1	1	3123
<input type="checkbox"/> leaf	● green	open	1	1	5070
<input type="checkbox"/> leaf_grid	● green	open	1	1	1948

図 3: 133.71.106.141 の Elasticsearch サーバーの leaf という文字列を含むインデックスのドキュメントのカウント結果

次に、図 4 に移行元の Elasticsearch サーバーの pcs_log インデックスのドキュメント数をカウントしたものを、図 5 に移行先の Elasticsearch サーバーの pcs_log インデックスのドキュメント数をカウントしたものを示す。

図 4 と図 5 より、ドキュメント数が一致していることからデータ移行が正常に行えたと判断できる。

<input type="checkbox"/> 名前	ヘルス	ステータス	プライマリ	レプリカ	ドキュメント数
<input type="checkbox"/> pcs_log	● yellow	open	1	1	1535

図 4: 133.71.201.197 の Elasticsearch サーバーの pcs_log インデックスのドキュメントのカウント結果

<input type="checkbox"/> 名前	ヘルス	ステータス	プライマリ	レプリカ	ドキュメント数
<input type="checkbox"/> pcs_log	● green	open	1	1	1535

図 5: 133.71.106.141 の Elasticsearch サーバーの pcs_log インデックスのドキュメントのカウント結果

5 まとめ

今回は, 133.71.201.197 の Elasticsearch サーバーにある pcs_log, leaf, leaf_load, leaf_grid インデックスを 133.71.106.141 の Elasticsearch サーバーにデータ移行して, 各インデックスのドキュメント数を比較することでデータ移行が正常に行えたことを報告した.

pcs_log インデックスは, 継続して 133.71.201.197 の Elasticsearch サーバーにデータがインサートされているので, 恵村君が 133.71.106.141 の Elasticsearch サーバーの pcs_log インデックスにデータをインサートするようシステムの対応を行った時点で, 追加のデータ移行を行う必要がある.

今回は, 以前 CO₂ データを移行した際に Elasticsearch の Bulk API を使用してローカルからデータをインサートしたことが原因で, ラズベリーパイ上で実行しているデータインサート用のプログラムからインサートが出来なくなった問題について, 以前データ移行を行った pcs_recyclekan インデックスについても Bulk API を使用しており, 将来同じ問題が発生する可能性があるため, 解決方法を調査する.

参考文献

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