学位論文

学内のElasticsearchシステムのデータ移行と 冗長化

提出年月日 令和4年X月X日

改定日 令和年月日

指導教員 都築 伸二 教授

入学年度 令和6年

学科名 電子情報工学専攻

論文提出者 祖父江 匠真

内容梗概1

本論文は,筆者が愛媛大学大学院理工学研究科電子情報工学専攻電気電子工学コースに在学中に行った、学内のElasticsearchのデータ移行と冗長化についてまとめたものであり、以下の5章から構成されている。

第1章 緒論

本研究を行うに至った経緯及び、本研究の目的について述べている。

- 第2章 太陽光発電の計測データの補正 ここでは太陽光発電の計測データの補正について述べる。
- 第3章 学内ゾーンにおける Elasticsearch クラスタへのデータ移行

ここでは学内ゾーンにおける Elasticsearch クラスタへの データ移行について述べる。

第4章 サーバーゾーンでのクラスタ構築 ここでは、サーバーゾーンでのクラスタ構築における仮想 環境を使用した事前検証について述べている。

第5章 結論

本研究によって明らかになった事項や今後の研究課題について簡単にまとめている。

目次

内容梗概					
第1章	緒論		1		
第2章	太陽光	発電の計測データの補正	3		
	2.1 節	緒言	3		
	2.2 節	太陽光発電の計測データの問題点について	3		
	2.3 節	日射量の計算式の導出・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	3		
	2.4 節	日射量を計算するプログラムの開発	4		
	2.5 節	実測データと計算データの比較・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	5		
	2.6 節	相互相関によるずれ時間の特定・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	5		
	2.7 節	日射量の計算値の予測精度改善	6		
	2.8 節	pvlib の概要	6		
	2.9 節	実測値と pvlib を使って求めた計算値の比較	7		
	2.10 節	pvlib を用いて計算した日射量を使用した, 相互相関に			
		よるずれ時間の特定・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	8		
	2.11 節	結言	8		
第3章	学内ゾ・	ーンにおける Elasticsearch クラスタへのデータ移行	9		
	3.1 節	緒言	9		
	3.2 節	Elasticsearch の概要	9		
	3.3 節	kibana の概要	10		
	3.4 節	学内ゾーンで稼働している Elasticsearch システムの状況	11		
	3.5 節	データ移行対象の Elasticsearch インデックスについて .	11		
	3.6.節	CO. データの移行手順について	11		

		3.6.1	データのエクスポート	11
		3.6.2	データの重複削除	12
		3.6.3	データのインポート	14
	3.7 節	一度目	$oldsymbol{I}$ のデータ移行で移行できなかった CO_2 データの	
		移行に	こついて	14
	3.8 節	kibana	a によるデータの可視化	15
	3.9 節	LEAF	の運行日誌に関するデータの移行について	16
	3.10 節	LEAF	の運行日誌に関するデータの移行手順について .	19
		3.10.1	データのエクスポート	19
		3.10.2	データのインポート	19
	3.11 節	結言		19
第4章	サーバ・	ーゾーン	ンでのクラスタ構築における仮想環境を使用した事	
	前検証			20
	4.1 節	緒言		20
	4.2 節	サーバ	ビーゾーンで稼働している Elasticsearch システム	
		の状況	2	20
	4.3 節	Docke	r とは	21
		4.3.1	コンテナとは	21
		4.3.2	Docker イメージとは	22
	4.4 節	Docke	r Compose とは	22
	4.5 節	検証環	環境のセットアップ	22
		4.5.1	全て同じバージョンの Elasticsearch を使用した	
			クラスタ構成 (全ノード バージョン 7.17.9)	22
		4.5.2	異なるバージョンの Elasticsearch を使用したク	
			ラスタ構成 (2 ノード バージョン 7.17.9, 1 ノー	
			ド バージョン 7.17.6)	27
	4.6 節	異なる	Elasticsearch クラスタへのノード参加検証	28
	4.7 節	手順		29
		4.7.1	単一ノードで稼働するクラスタ A の構築	29
		4.7.2	クラスタ B の構築	29

		4.7.3	クラスタ	' B への	参加試行	Ī		 		30
	4.8 節	Elastic	esearch の	バージョ	ョンアッ	プ		 		31
	4.9 節	バージ	゚ョンアッ	プ手順.				 		31
		4.9.1	インスト	ール方法	去の特定			 		31
		4.9.2	apt \mathbb{C} \sharp	るバーシ	^ブ ョンア	ップ .		 		31
	4.10 節	kibana	のバージ	ョンア	ップ			 		32
	4.11 節	バージ	゚ョンアッ	プ後の重	协作確認			 		32
	4.12 節	サーバ	ニーゾーン	における	るクラス	タの構築	色	 		32
	4.13 節	結言						 		32
第5章	結論と	今後の記	果題							4 0
謝辞										41
参考文献	ξ									42

第1章

緒論

急速に進化するデータ管理において、Elasticsearch は極めて重要な技術として登場し、データの保存、検索、管理方法に革命をもたらした。当初、大量のデータを効率的に処理するために設計された Elasticsearch は、シンプルなシングルノードシステムから複雑なクラスタ構成へと移行し、データハンドリング技術において大きな進歩を遂げた。この変遷は、様々な分野で堅牢でスケーラブル、かつ冗長性のあるデータ管理システムに対する要求が高まっていることを裏付けている。

データシステムの冗長性、特に Elasticsearch クラスタにおける冗長性は、データの信頼性と可用性を確保する上で重要な要素となっている. 冗長性とは、システムの重要なコンポーネントや機能を二重化することで、信頼性を高め、一点障害のリスクを低減することを指す.

Elasticsearch のノード管理における現在のトレンドは、クラスタ化されたシステムを好む傾向が強まっている。しかし、Elasticsearch の技術的な側面については多くの研究があるが、同じ環境内でシングルノードシステムからクラスタ化されたシステムへデータを移行し、さらに別のネットワークゾーンに新しくクラスタ化されたシステムを構築する際の実際的な課題や戦略について掘り下げた研究はほとんどない。

本研究の目的は、シングルノードの Elasticsearch システムから学内ゾーン内のクラスタ化システムへデータを移行するプロセスを分析することである. 同時に本研究では、この2つのゾーン間のデータ移行を行わずに、サーバゾー

ンに新しい冗長化されたクラスタ化システムを構築する.

第2章では学内ゾーンにおける Elasticsearch ノードから新クラスタへのデータ移行について述べる。第3章では仮想環境を使用してサーバーゾーンの Elasticsearch ノードをシミュレートし、マルチコンテナ Docker アプリケーションのツールである docker-compose を用いてクラスタ構築の実現可能性を検証したことについて述べる。第4章ではサーバーゾーンで既存の Elasticsearch ノードを用いたクラスタ構築について述べる。第5章では結論と課題を述べる。

第2章

太陽光発電の計測データの補正

2.1 節 緒言

本章では太陽光発電の計測データの補正について述べる.

2.2節 太陽光発電の計測データの問題点について

CSV データなどで保存された太陽光発電の環境データはオフライン環境でファイル書き込みを行っているため、PC の内部時計がずれており、計測データの日時情報が正確な日時とは異なっている.

そこで相互相関を用いて、実測した日射量の時系列データと、計算式により 求まる日射量の時系列データとの時間的遅延の秒数を検出することで、実測 データの計測日時のずれ時間を特定する.

2.3 節 日射量の計算式の導出

任意の緯度経度、日時における日射量 Q は、任意の緯度 ϕ 、経度 λ の地点における任意の日時、太陽高度 α から求めることができる.

まず、次式により元旦からの通し日数 dn に基いて定めた θ を用いて、当該日

の太陽赤緯 δ , 地心太陽距離 $\frac{r}{r^*}$, 均時差 E_q をそれぞれ以下の式により求める.

$$\theta = \frac{2\pi(dn-1)}{365} \tag{2.1}$$

$$\delta = 0.006918 - 0.399912\cos\theta + 0.070257\sin\theta - 0.006758\cos2\theta + 0.000997\sin2\theta - 0.002697\cos3\theta + 0.001480\sin3\theta$$
 (2.2)

$$\frac{r}{r^*} = \frac{1}{\sqrt{1.000110 + 0.034221\cos\theta + 0.001280\sin\theta + 0.000719\cos2\theta + 0.000077\sin2\theta}}$$

日本標準時間から、太陽の時角 h を求める.

$$h = \frac{(\mathbf{日本標準時間} - 12)\pi}{12} + 標準子午線からの経度差 + E_q$$
 (2.5)

 δ , ϕ , h の値が既知となったので α は

$$\alpha = \arcsin(\sin\phi\sin\delta + \cos\phi\cos\delta\cos h) \tag{2.6}$$

として求めることができる.

最後に, Qを

$$Q = 1367\left(\frac{r^*}{r}\right)^2 \sin\alpha \tag{2.7}$$

により求めることができる. また, $1367W/m^2$ は太陽定数である.

2.4節 日射量を計算するプログラムの開発

式 (2.1) ~ 式 (2.7) を用いることで、任意の緯度経度、日時における日射量が求まる.

2.5節 実測データと計算データの比較

Elasticsearch サーバーから取得した日射量データと、計算式から求めた日射量データをプロットしたものを図 2.1 に示す。図 2.1 は Elasticsearch サーバーから取得した 2022 年 6 月 2 日の日射量データと、リサイクル館の緯度経度と日付情報より求めた日射量の値をプロットしている。

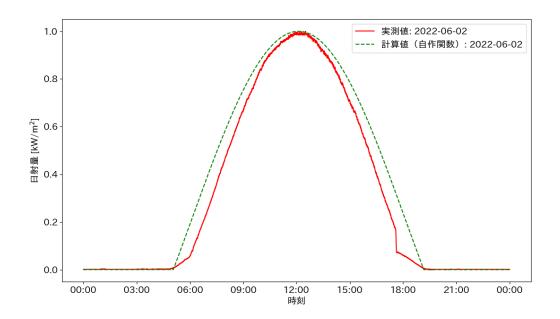


図 2.1 2022 年 6 月 2 日の日射量の実測データと計算データをプロットしたもの

2.6節 相互相関によるずれ時間の特定

今回, 相互相関の計算に使用する実測データの範囲を, 2022 年 6 月 2 日 0 時 0 分から 2022 年 6 月 2 日 23 時 59 分まで期間とする.

2022年6月2日を選定した理由として、図2.1より、2022年6月2日の実測データの概形は計算データの概形と類似していたためである.

実測データの計測日時の情報をもとに、計算データを求め、これらの値を入力として相互相関を計算する.

相互相関の計算結果より、計算値の日時を実測値より 124 秒進めた際に、相互相関の値が最大となった.

しかし、今回使用した実測データには計測日時のずれは殆どないため、計算データの日時を進めていない際に実測データとの相関が最大となるのが正しい.

これは、日射量の計算式の予測精度が低いことが原因であると考えられる.

2.7節 日射量の計算値の予測精度改善

日射量の計算値の予測精度を改善するため、式 (2.1) ~ 式 (2.7) を使った方法ではなく、pvlib というライブラリを使用して、任意の緯度経度と日時における日射量を求めて相互相関を計算する.

2.8節 pvlibの概要

pvlib は、太陽光発電システムの性能シミュレーションや関連するタスクを実行するための関数とクラスのセットを提供する、コミュニティが開発したツールボックスである。

以下は、pvlib の主な特徴である.

- ◆ 太陽位置計算: pvlib は,地球上の任意の場所における太陽の位置を計算する機能を提供する.これは,太陽の方位角や高度角を求めるのに使用される.
- ◆ 大気透過モデル: 大気を通過する太陽放射の量や質を推定するモデルが 含まれている.
- ◆ 太陽光発電システムの性能モデリング:太陽光発電モジュールやインバーターの性能モデルが含まれており、異なる条件下での太陽光発電システムの出力をシミュレートできる.

2.9節 実測値とpvlibを使って求めた計算値の比較

Elasticsearch サーバーから取得した日射量データと、pbliv を用いて求めた日射量データをプロットしたものを図 2.2 に示す。図 2.2 は Elasticsearch サーバーから取得した 2022 年 6 月 2 日の日射量データと、リサイクル館の緯度経度と日付情報を入力として pvlib より求めた日射量の計算値をプロットしている。図 2.1 と比較して、pvlib より求めた日射量が実測データにより近い概形となっていることが分かる。

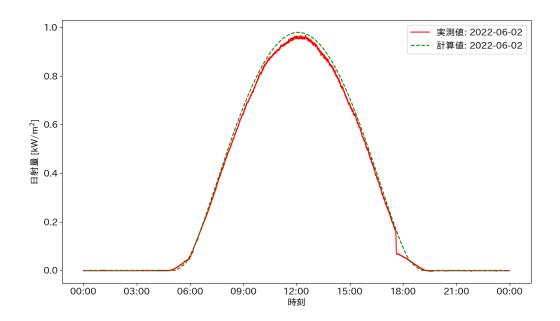


図 2.2 2022年6月2日の日射量の実測値と計算値をプロットしたもの

2.10節 pvlibを用いて計算した日射量を使用した, 相互相関 によるずれ時間の特定

相互相関の計算に使用する実測データの範囲を, 2022 年 6 月 2 日 0 時 0 分から 2022 年 6 月 2 日 23 時 59 分まで期間とする.

実測データの計測日時の情報をもとに、pvlib より計算データを求め、これらの値を入力として相互相関を計算する.

相互相関を求めた結果、計算値の日時を実測値より 74 秒進めた際に、相互相関の値が最大となることが分かった.

式 (2.1) ~ 式 (2.7) を用いて求めた日射量データ用いて相互相関を計算した時と比較して、124 秒から 74 秒へと 50 秒改善した.

2.11節 結言

本章では太陽光発電の計測データの補正について述べた. 次章では学内ゾーンで稼働している Elasticsearch クラスタへのデータ移行について述べる.

第3章

学内ゾーンにおけるElasticsearch クラスタへのデータ移行

3.1 節 緒言

本章では学内ゾーンで稼働している Elasticsearch クラスタへのデータ移行 について述べる.

3.2節 Elasticsearchの概要

Elasticsearch は、分散処理に対応した全文検索エンジンである。主な特徴は、以下の通りである。

- 高速な検索性能: ビッグデータなどの巨大で複雑なデータの集合にも対応可能
- 部分一致検索が可能: 検索キーワードの一部に一致するドキュメントも 検索可能
- ほぼリアルタイムの検索: ドキュメントにインデックスを付けてから検索可能になるまで約1秒程度
- スケーラビリティ: サーバー数を増やすことで、検索性能と処理能力を 拡張可能

これらの特徴から、Elasticsearch は、以下のような用途に適している。

- ログ分析: Web サイトやアプリケーションのログから、アクセス状況や エラー情報を分析する
- セキュリティインテリジェンス: ネットワークやシステムから、セキュ リティ脅威を検知する
- ビジネス分析: 顧客データや販売データから、トレンドや傾向を分析 する

3.3 節 kibanaの概要

Kibana は、Elasticsearch に保存されたデータを可視化するためのツールで ある。主な特徴は、以下の通りである。

- 直感的な操作性: ドラッグ&ドロップで簡単に可視化を作成できる
- 豊富な可視化機能: グラフ、表、地図など、さまざまな可視化機能を 提供
- 高度なフィルタリング機能:条件を指定して、データを詳細に絞り込む ことができる

これらの特徴から、Kibana は、以下のような用途に適している。

- ログ分析: Web サイトやアプリケーションのログから、アクセス状況や エラー情報を可視化する
- セキュリティインテリジェンス: ネットワークやシステムから、セキュ リティ脅威を可視化する
- ビジネス分析: 顧客データや販売データから、トレンドや傾向を可視化 する

3.4節 学内ゾーンで稼働している Elasticsearch システムの 状況

学内ゾーンでは、133.71.106.168で単一ノードの Elasticsearch と、133.71.106.170、133.71.106.141、133.71.106.136の Elasticsearch ノードによって構成された Elasticsearch クラスタが稼働している。

133.71.106.168 の Elasticsearch には, CO_2 濃度監視システムによって計測されたデータと LEAF の運行日誌に関するデータが保存されている.

3.5 節 データ移行対象のElasticsearch インデックスについて

133.71.106.168 で稼働している単一ノードの Elasticsearch に保存された CO_2 データと LEAF の運行日誌に関するデータを、学内ゾーンで稼働している Elasticsearch クラスタへ移行する.

3.6 節 CO_2 データの移行手順について

CO₂のデータ移行を行う上で、タイムスタンプと部屋番号の組み合わせが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る。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インサートする。

3.6.1 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 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, co2 という文字列を含むインデックスのデータのみをエクスポートした.

3.6.2 データの重複削除

重複データの削除はSQLite データベースを用いて行った.

SQLite は、軽量で自己完結型のデータベースエンジンです。これは、多く のアプリケーションに組み込まれており、以下の特徴を持っています:

軽量: SQLite は非常に小さく、リソースの少ない環境でも動作します。例 えば、組み込みシステムやモバイルアプリケーションなどです。自己完結型: SQLite はサーバーレスで、設定や管理が不要です。これは、データベースが 単一のファイルとして存在し、外部の依存関係がないことを意味します。トラ ンザクション: SQLite は ACID トランザクションをサポートしており、デー タの整合性を保ちます。言語のサポート: 多くのプログラミング言語で使用で きます。幅広い用途: デスクトップ、モバイル、ウェブアプリケーションなど、 さまざまな環境で利用されています。フリーかつオープンソース: SQLite は パブリックドメインに属し、誰でも自由に使用、変更、配布できます。SQL標 準準拠: 多くの SQL 標準機能をサポートしていますが、全てではありません。

SQLite は、その単純さと効率性から、広く使われているデータベースエン ジンの一つです。特に、大規模なデータベースシステムのオーバーヘッドが不 要な場合や、独立したアプリケーションでの使用に適しています。

SQLite データベースはリレーショナルデータベースの一種であり、複合主 キーを使って複数のテーブルカラムの組み合わせを一意の識別子として扱う ことができる、これにより、同じ組み合わせのデータを重複して挿入しようと した場合、データベースエンジンがコンフリクトエラーを発生させ、重複デー タの挿入を阻止する. そのため, 今回の重複データ削除には適していると判断 した.

今回使用した SQLite データベースでは、部屋番号 (number) とタイムスタン プ (utctime) を一意のキーとして設定した. 以下のリスト 4.1, リスト 3.2 に示 すように、移行元の ElasticSearch サーバーに保存されている co2 インデックス のドキュメントは、フィールドのメンバーが統一されておらず、一部センサー 情報が存在しない場合がある、そのため、データの挿入時にコンフリクトエ ラーが発生した場合は、既存のレコードと挿入しようとしたレコードを比較し、 既存レコードの値が NULL であるカラムにおいて、挿入しようとしているレ Faculty of Engineering

コードの値が非 NULL である場合には、既存レコードのカラムの値を更新する ようにした。これにより、重複データ削除時に一部センサー情報などが欠けて しまう問題を解決した.

```
_source フィールドのメンバー数が少ないドキュメ
Listing 3.1
ント
{
 "_index": "co2_e411",
 "_type": "_doc",
  "_id": "nEi2nnoB2—iFXnrMOobM",
 "_score": 1,
 "_source": {
     "utctime": "2020-10-09T05:09:06+00:00",
     "number": "E411",
     "PPM": "481",
     "data": "Thingspeak"
  }
}
          _source フィールドのメンバー数が多いドキュメント
Listing 3.2
{
 "_index": "co2_e411",
 "_type": "_doc",
  "_id": "YKBqU4QBugDzeydA2gyi",
  "_score": 1,
 "_source": {
     "RH": 26.98,
     "PPM": 423,
     "JPtime": "2022-11-06T22:45:30.080925",
     "ip": "172.23.68.19/16",
     "utctime": "2022-11-06T13:45:30.080895",
```

```
"TEMP": 24.47,
      "index_name": "co2_e411",
      "ms": "",
      "number": "E411"
  }
}
```

3.6.3 データのインポート

重複データ削除後のデータが保存された SQLite テーブルからすべてのレ コードを読み出して、ターゲットの ElasticSearch サーバーに移行した.

その際, python の elasticsearch ライブラリを使用し, co2_modbus という名 前のインデックスに保存した.

3.7節 一度目のデータ移行で移行できなかった CO₂ データの 移行について

実装したデータ移行プログラムを使用して 133.71.201.197 から 133.71.106.141 の ElasticSearch サーバーへ CO₂ データを移行したのが 2023 年 5 月中旬頃で あり、CO2 濃度監視システムを開発、運用している高木君が、移行先である 133.71.106.141 の ElasticSearch サーバーに対してラズベリーパイから CO₂ デー タのインサートを行うよう対応したのが2023年7月中旬であったため,2023 年 5 月中旬から 2023 年 7 月中旬までの間の約 2ヶ月間の CO₂ データが移行先 の ElasticSearch サーバーに移行出来ていなかった. そこで、追加の移行作業を 行った.

移行方法は以下のとおりである.

- 1. まず、2023年5月中旬に移行した際の全移行データの中で最も最新の utctime フィールドの値を検索する.
 - 検索した結果、2023年5月中旬に移行した際の全移行データの中で 最も最新の utctime は「2023-05-16T05:48:30.081305」であった.

- 2. 次に、移行先 ElasticSearch サーバーに対してラズベリーパイからイン サートされた全データの中で最も古い utctime フィールドの値を検索 する.
 - 検索した結果、ラズベリーパイからインサートされた全データの中 で最も古い utctime は「2023-07-20T07:15:39.314008」であった.
- 3. 前回の CO₂ データの移行は 2023 年 5 月中旬頃に行ったため, 2023 年 5月1日0時0分0秒以降の utctime を持つドキュメントを、移行元 ElasticSearch サーバーのインデックス名に co2 という文字列を含むイ ンデックスから elasticdump [1] ライブラリを使用してローカルマシン にエクスポートする.
- 4. 部屋番号 (number) とタイムスタンプ (utctime) の組み合わせがユニー クになるようにエクスポートしたデータをフィルタリングする.
- 5. 更に、1 と 2 で得られた utctime の範囲に含まれる utctime を持つドキュ メントのみになるようフィルタリングする.
- 6. フィルタリング後のデータを移行先 ElasticSearch サーバーにバルクイ ンサートする.

kibana によるデータの可視化 3.8節

計 2 回の CO_2 データを移行した後の co2 modbus インデックスについて, 横 軸をタイムスタンプ (utctime) 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 3.1 ~ 図 3.3 に示す.

2回目の CO₂ データの移行によって, 2023年5月中旬から 2023年7月中旬 までの期間とその前後の期間において、図 3.1 ~ 図 3.3 より、連続的にデータ が変化していることが目視で確認できるので、データ移行は正常に出来たと判 断できる.

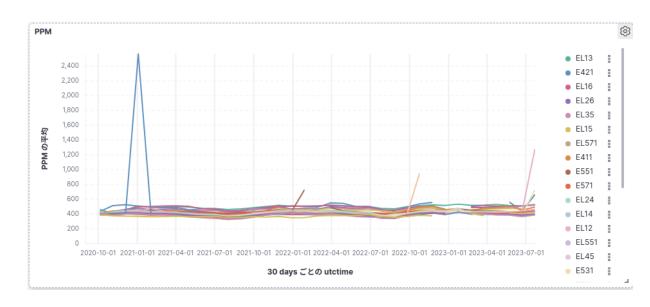


図 3.1 co2_modbus \mathcal{O} PPM

LEAFの運行日誌に関するデータの移行について 3.9 節

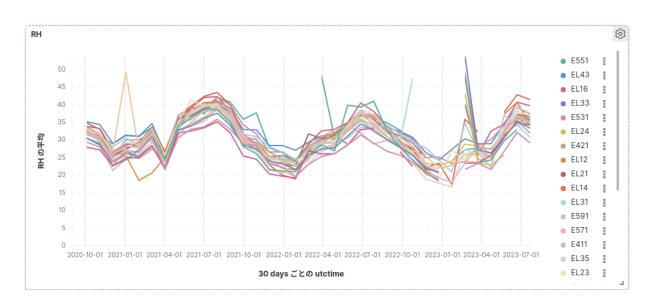
LEAF の運行日誌に関するデータが保存されたインデックスは以下の2つで ある.

- movement_diary
- movement_diary01

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.

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、 以下に movement_diary と movement_diary 01 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.



 $\boxtimes 3.2$ co2_modbus \mathcal{O} RH

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。

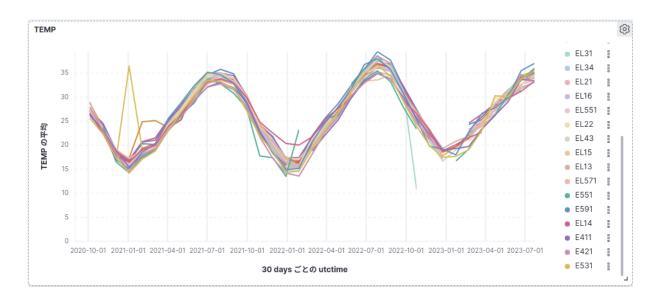


図 3.3 co2_modbus \mathcal{O} TEMP

- 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.

movement_diary と movement_diary01 のドキュメントの違いより、movement_diary01 は movement_diary のもつ情報量を全て保持しており、その上で 追加のフィールドを持っていることから、移行するのは movement_diary01 イ ンデックスのみで十分であることが分かった.

LEAFの運行日誌に関するデータの移行手順について 3.10 節

3.10.1 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 ライブラリを使用して、movement_diary01 インデックス の全ドキュメントを JSON 形式でエクスポートした.

3.10.2 データのインポート

python の elasticsearch ライブラリを使用し、移行先の Elasticsearch に movement_diary01 という名前のインデックスを作成して、エクスポートしたデータ を全てインサートした.

3.11 節 結言

本章では学内ゾーンで稼働している Elasticsearch クラスタへのデータ移行 について述べた。

次章ではサーバーゾーンでのクラスタ構築における仮想環境を使用した事 前検証について述べる。

第4章

サーバーゾーンでのクラスタ構築に おける仮想環境を使用した事前検証

4.1 節 緒言

本章では、サーバーゾーンでのクラスタ構築における仮想環境を使用した 事前検証ついて述べる。

4.2 節 サーバーゾーンで稼働している Elasticsearch システムの状況

サーバーゾーンでは、133.71.201.197で単一ノードの Elasticsearch が稼働しており、リサイクル館の太陽光パネルの計測データが保存されている.

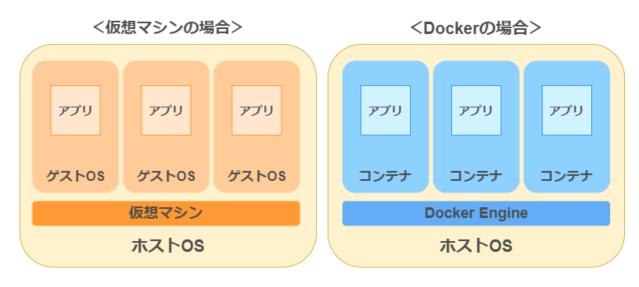
133.71.201.197 の Elasticsearch のバージョン 7.17.6 であり、学内ゾーンで稼働している Elasticsearch クラスタに参加している Elastisearch ノードのバージョンは 7.17.9 である.

本研究室では現在、バージョン 7.17.9 の Elasticsearch を採用しているため、サーバーゾーンで構築しようとしているクラスタの Elasticsearch のバージョンも、学内ゾーンで稼働している Elasticsearch クラスタと同様、バージョン 7.17.9 を採用する.

そこで、Docker、Docker Compose を使用して、異なるバージョンである 7.17.6 と 7.17.9 の Elasticsearch ノードをクラスタリングすることが可能かど うかを確認するために実施した.

4.3 節 Dockerとは

Docker は、軽量で独立したコンテナ型仮想環境用のプラットフォームである。 従来の仮想化では、VMWare などの仮想化ソフトウェアを用いて、ホスト OS 上にゲスト OS を構築する形式だった. しかし、Docker はホスト OS 上にゲス ト OS なしで独立したコンテナ型の仮想環境として構築される. Docker コンテ ナを利用する場合は、Docker Engine をインストールすることでコンテナの立 ち上げ、停止、削除といった操作を行うことができる.



仮想マシンと Docker **の違い** [2] 図 4.1

コンテナとは 4.3.1

コンテナは、アプリケーションとそのすべての依存関係(ライブラリ、実行 環境など)をカプセル化した軽量な実行単位である. Docker の場合、コンテナ の作成には Docker イメージが必要となる.

Docker イメージとは 4.3.2

Docker イメージとは、Docker コンテナを作成するためのテンプレートであ り、Docker イメージの中には、Docker コンテナの実行に必要な Linux ファイ ルシステムとメタ情報を含む.

Linux ファイルシステムというのは、/ ディレクトリ以下の /etc /bin /sbin /usr などのディレクトリ階層およびファイルである.

Docker では、コンテナとして動かしたいアプリケーションが必要とする、最 小限のファイルを Docker イメージの中に入れる.

さらに、そのアプリケーションを動かすために必要なデフォルトのコマンド や引数の指定、外に公開するポート番号の情報などの情報がある。これらをメ タ情報として、同じく Docker イメージの中に入れられる

Docker イメージは Docker Hub やその他のレジストリで共有されており、こ れらのサービスから取得することが可能である.

今回は Elasticsearch の開発元である Elastic 社が提供している Elasticsearch の Docker イメージを使って検証を行う.

4.4 節 Docker Compose とは

Docker Compose は、複数のコンテナを定義し、実行するためのツールであ る. これは YAML ファイルを使用して設定され、複数のコンテナで協調して動 作するアプリケーションの開発を単純化する.

検証環境のセットアップ 4.5 節

4.5.1 全て同じバージョンの Elasticsearch を使用したクラスタ構成 (全ノード バージョン 7.17.9)

> Listing 4.1 に 7.17.9 バージョンの Elasticsearch のみを使用してクラスタを 構築した時の docker-compose.yml をを示す.

第4章 サーバーゾーンでのクラスタ構築における仮想環境を使用した事前検証 25

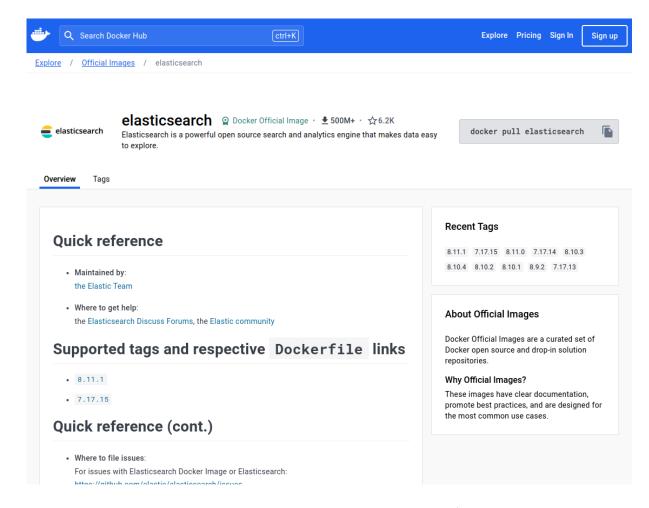
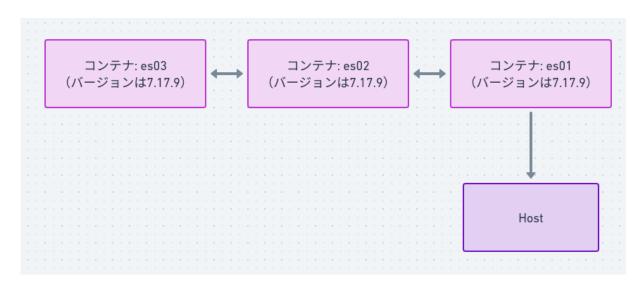


図 4.2 Elasticsearch の Docker イメージ

Listing 4.1 全て同じバージョンの Elasticsearch を使用したクラスタを構成する docker-compose.yml

```
- cluster.name=es-docker-cluster
     - discovery.seed_hosts=es02,es03
      - cluster.initial_master_nodes=es01, es02, es03
    ports:
     -9200:9200
    networks:
     - elastic
  es02:
    image: docker.elastic.co/elasticsearch/elasticsearch:7.17.9
    container_name: es02
    environment:
      - node.name=es02
      - cluster.name=es-docker-cluster
      - discovery.seed_hosts=es01,es03
      - cluster.initial_master_nodes=es01, es02, es03
    networks:
     - elastic
  es03:
    image: docker.elastic.co/elasticsearch/elasticsearch:7.17.9
    container_name: es03
    environment:
      - node.name=es03
      - cluster.name=es-docker-cluster
      - discovery.seed_hosts=es01,es02
      - cluster.initial_master_nodes=es01, es02, es03
    networks:
     - elastic
networks:
  elastic:
    driver: bridge
```

また、図 4.3 に Listing 4.1 の docker-compose.yml を図で表現したものを示す.



docker-compose.yml を図で表現したもの \mathbf{Z} 4.3

Listing 4.1 の docker-compose.yml ファイルで記述している内容について説 明する.

サービスの定義

• es01, es02, es03: and bar = andる. 各ノードは異なるコンテナとして定義されている. es01, es02, es03 はそれぞれ異なるコンテナ名で、Elasticsearch の異なるインスタンスを 実行する.

各ノードの設定

- image: 使用する Docker イメージ. ここでは Elasticsearch の 7.17.9 バー ジョンを使用している.
- container_name: コンテナに割り当てられる名前.
- environment: 環境変数の設定. Elasticsearch のクラスタ設定を含む.

- ports: ホストマシンとコンテナ間のポートマッピング. 例えば, '9200:9200'はホストマシンの 9200 ポートをコンテナの 9200 ポート にマッピングする.
- networks: コンテナ間通信のためのネットワーク設定. ここでは elastic ネットワークが使用されている.

ボリュームとネットワークの設定

 networks: デフォルトのドライバである bridge ドライバを使用する elastic ネットワークを定義している. これにより、異なるコンテナが相 互に通信できるようになる.

この設定により、Elasticsearch の 3 ノードを含むクラスタが Docker 上で動作するようセットアップされる.

以降, 説明を簡単にするため, docker-compose.yml を図で表現したもののみを掲示する.

クラスタの起動には、docker compose up -d コマンドを使用する.

docker compose up -d コマンドを実行した後, curl コマンドを使用してクラスタに参加しているノードを一覧表示した結果を図 4.4 に示す.

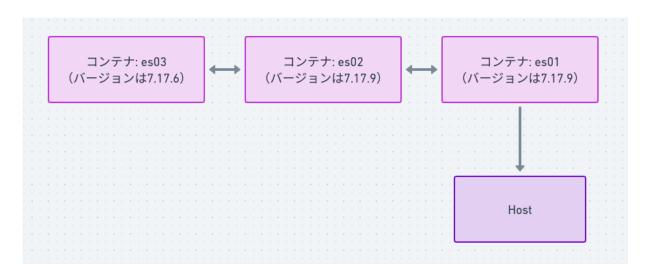
図 4.4 クラスタに参加しているノードを一覧表示した結果

図 4.4 より、3 つのノード (es01, es02, es03) すべてが正常にクラスタに参加できていることが確認できる.

異なるバージョンの Elasticsearch を使用したクラスタ構成 (2 4.5.2 **ノード バージョン 7.17.9, 1 ノード バージョン 7.17.6)**

図 4.3 の docker-compose.yml の es03 のコンテナが使用する Docker イメー ジを変更して, es03 のノードで使用する Elasticsearch のバージョンを 7.17.9 か ら 7.17.6 に変更する.

図 4.5 に変更後の docker-compose.yml を図で表現したものを示す.



24.5変更後の docker-compose.yml を図で表現したもの

変更後, docker compose up -d コマンドを実行してクラスタを起動する. クラスタの起動後、curl コマンドを使用してクラスタに参加しているノード を一覧表示した結果を図 4.6 に示す.

図 4.6 より、バージョンが 7.17.9 である 2 つのノード (es01, es02) のみが正 常にクラスタに参加できていることが確認できる.

また、Elasticsearch 起動時に出力されたログを確認したところ、図 4.7 に示 すように、クラスタに参加できなかった es03 のコンテナで Elasticsearch がエ ラーログを出力して終了していることが分かった.

```
$ curl -X GET "localhost:9200/_cat/nodes?v=true&pretty"
ip heap.percent ram.percent cpu load_1m load_5m load_15m node.role master name
172.22.0.4 20 14 7 1.86 1.32 1.14 cdfhilmrstw - es02
172.22.0.3 18 14 8 1.86 1.32 1.14 cdfhilmrstw * es01
```

図 4.6 クラスタに参加しているノードを一覧表示した結果

図 $4.7 ext{ es} 03$ のログ

4.6 節 異なる Elasticsearch クラスタへのノード参加検証

サーバーゾーンでのクラスタ構築において、リサイクル館の太陽光パネルの 計測データを保存している Elasticsearch ノードを新たなノードとして構築し たクラスタに参加できるか、Docker を用いて検証した.

4.7節 手順

単一ノードで稼働するクラスタ A の構築 4.7.1

まず、docker-compose を用いて単一ノード (コンテナ名は es04) でクラスタ 以後このクラスタをクラスタ A と呼ぶを構築する. 以後このクラスタをクラ スタ A と呼ぶ.

図 4.8 にクラスタ A の構築の際に使用した docker-compose.yml を図で表現 したものを示す.

docker-compose を用いてノードを起動した後、クラスタの情報について問い 合わせた結果を図 4.9 に示す.

クラスタの情報について問い合わせた後、Docker コンテナを停止してノー ドをシャットダウンした.

4.7.2クラスタBの構築

次に、クラスタ A に使用したノードとは別の 3 ノード (コンテナ名はそれぞ \mathbf{h} es01, es02, es03) でクラスタを構築する. 以後このクラスタをクラスタ B と 呼ぶ.

図 4.10 にクラスタ B の構築の際に使用した docker-compose.yml を図で表現 したものを示す.

docker-compose を用いて3つのノードを起動した後、クラスタの情報につい て問い合わせた結果を図 4.11 に示す.

図 4.9, 4.11 より, クラスタ A とクラスタ B はそれぞれ異なるクラスタ ID を 付与されたことが分かる.

クラスタの起動後、クラスタに参加しているノードの一覧を取得した結果を 図 4.12 に示す.

図 4.12 より, es01, es02, es03 ノードが全てクラスタ B に参加できているこ とが分かる.

クラスタに参加しているノードの一覧を取得した後、全ての Docker コンテ ナを停止してノードを全てシャットダウンした.

4.7.3 クラスタBへの参加試行

次に、図 4.10 の docker-compose.yml に対して、クラスタ A のノード (es04 コンテナ) を追加し、合計 4 ノードでのクラスタ B の起動を試みる.

図 4.13 に、合計 4 ノードでクラスタ B の起動を試みた際に使用した docker-compose.yml を図で表現したものを示す.

クラスタの起動後、クラスタに参加しているノードの一覧を取得した結果を 図 4.14 に示す.

図 4.14 より, クラスタ A のノードがクラスタ B に参加できていないことが分かる.

es04 コンテナ(クラスタ A のノード)で出力されたログの一部を図 4.15 に示す.

図 4.15 には、異なるクラスタ ID を持つクラスタにノードが参加することは禁止されており、これを行うためにはインデックスやドキュメント情報などが格納されているデータパス配下のフォルダ、ファイルを削除する必要があると書かれている.

以上の検証結果から、既に稼働しているノードを別のクラスタに新しいノードとして参加させることは出来ないことが分かった.

したがって、リサイクル館の太陽光パネルの計測データが保存された Elasticsearch ノードをクラスタに参加させるには以下の 2 通りの方法が考えられる.

- リサイクル館の太陽光パネルの計測データが保存された Elasticsearch ノードのバックアップを取り、ノードに保存されたインデックスやド キュメントのデータを削除した上で、CO₂ データなどが保存されたクラ スタに新しいノードとして参加させる
- CO₂ データなどが保存されたクラスタとは別で, サーバーゾーンに新た にクラスタを構築する. クラスタの構築にはリサイクル館の太陽光パネ ルの計測データが保存された Elasticsearch ノードが所属するクラスタ を使用する.

4.8 節 Elasticsearch のバージョンアップ

3章の検証結果より、異なるバージョンの Elasticsearch ノードでクラスタを構築することは出来ないため、リサイクル館の太陽光パネルの計測データを保存している Elasticsearch (133.71.201.197) をバージョンアップする必要がある。そこで、133.71.201.197 にインストールされた Elasticsearch のバージョンアップを行う。

4.9節 バージョンアップ手順

4.9.1 インストール方法の特定

バージョンアップを行うためには、133.71.201.197 の UbuntuPC にどのように Elasticsearch をインストールしたか特定する必要がある.

図 4.16 に, apt によってインストールされたパッケージの中に elasticsearch という文字列を含むパッケージが存在するか調べた結果を示す. 図 4.16 より, apt によってインストールされたことが分かった.

次に、apt でインストール可能な elasticsearch のバージョンを一覧表示した 結果を図 4.17 に示す。図 4.17 にターゲットである 7.17.9 が含まれているため、apt を使用してバージョンアップできることが確認できた。

4.9.2 apt によるバージョンアップ

まず, sudo systemctl stop elasticsearch.service コマンドを実行して elasticsearch ノードをシャットダウンする.

次に, sudo apt install elasticsearch=7.17.9 コマンドを実行して elasticsearch パッケージをバージョンアップする.

elasticsearch をバージョンアップ後, sudo systemctl start elasticsearch コマンドを実行して Elasticsearch ノードを起動する.

ノードの起動後、Elasticsearch のバージョンを確認した結果を図 4.18 に示す。図 4.18 より、Elasticsearch のバージョンが 7.17.9 にバージョンアップ出来たことが確認できた。

kibana のバージョンアップ 4.10 節

kibana も elasticsearch と同様, apt を使用してインストールされていた ため, sudo systemctl stop kibana.service コマンド, sudo apt install kibana=7.17.9 コマンド, sudo systemctl start kibana コマンドをそれぞ れ実行して、kibana のバージョンアップも行った.

バージョンアップ後の動作確認 4.11 節

Elasticsearch のバージョンアップ後、太陽光パネルの計測データが Elasticsearch に保存されているか kibana 上で確認した結果を図??に示す.

図 4.19 より、 バージョンアップ後の Elasicsearch ノードを起動した 14:22:00 以降にドキュメントがインサートされていることが確認できた.

4.12節 サーバーゾーンにおけるクラスタの構築

4.13 節 結言

本章では、サーバーゾーンでのクラスタ構築における仮想環境を使用した 事前検証について述べた.

次章ではサーバーゾーンでのクラスタ構築について述べる。

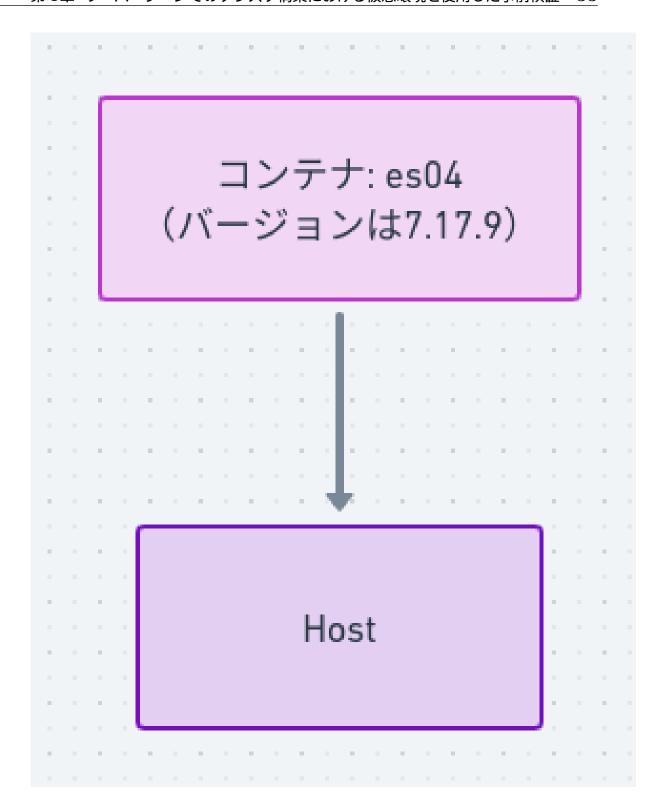


図 4.8 クラスタ A の構築の際に使用した docker-compose.yml を図で表現したもの

```
sofue@sofue-DAIV-DGX750:~/apps/clustering-different-es-ver$ curl -XGET http://localhost:9200/
{
    "name" : "es04",
    "cluster_unid" : "VbMebhfLQ0yQSlEx2nPFhg",
    "version" : {
        "number" : "7.17.9",
        "build_flavor" : "default",
        "build_tlavor" : "default",
        "build_hash" : "ef48222227ee6b9e70e502f0f0daa52435ee634d",
        "build_date" : "2023-01-31T05:34:43.305517834Z",
        "build_snapshot" : false,
        "lucene_version" : "8.11.1",
        "minimum_wire_compatibility_version" : "6.8.0",
        "minimum_index_compatibility_version" : "6.0.0-beta1"
    },
    "tagline" : "You Know, for Search"
}
```

図 4.9 クラスタの情報について問い合わせた結果

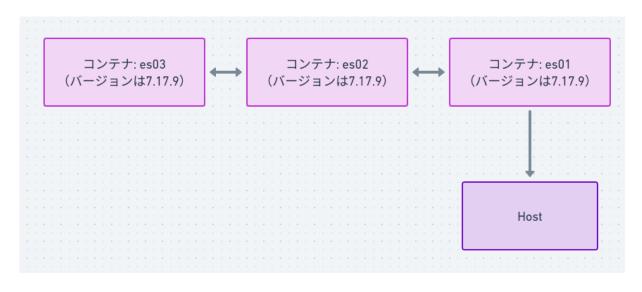


図 4.10 クラスタ B の構築の際に使用した docker-compose.yml を図で表現したもの

```
sofue@sofue-DAIV-DGX750:~/apps/clustering-different-es-ver$ curl -XGET http://localhost:9200/
{
    "name" : "es01",
    "cluster_name" : "docker-cluster",
    "cluster_uuid" : "XNQVSIDyTkuytnT_rKQt4w",
    "version" : {
        "number" : "7.17.9",
        "build_flavor" : "default",
        "build_type" : "docker",
        "build_hash" : "ef48222227ee6b9e70e502f0f0daa52435ee634d",
        "build_date" : "2023-01-31T05:34:43.305517834Z",
        "build_snapshot" : false,
        "lucene_version" : "8.11.1",
        "minimum_wire_compatibility_version" : "6.8.0",
        "minimum_index_compatibility_version" : "6.0.0-beta1"
    },
    "tagline" : "You Know, for Search"
}
```

図 4.11 クラスタの情報について問い合わせた結果

図 4.12 クラスタBの起動後、クラスタに参加しているノードの一覧を取得した結果

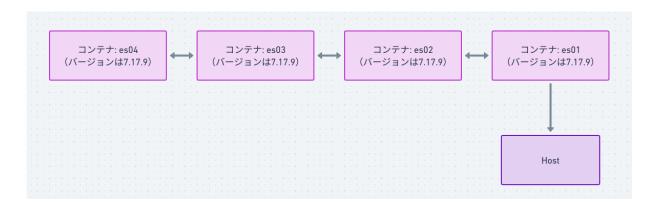


図 4.13 合計 4 ノードでクラスタ B の起動を試みた際に使用した docker-compose.yml を図で表現したもの

```
sofue@sofue-DAIV-DGX750:~/apps/clustering-different-es-ver$ curl -X GET "localhost:9200/_cat/nodes?v&pretty" ip heap.percent ram.percent cpu load_1m load_5m load_15m node.role master name 172.22.0.4 18 21 5 0.75 1.24 1.02 cdfhilmrstw - es02 172.22.0.2 16 21 5 0.75 1.24 1.02 cdfhilmrstw - es01 172.22.0.5 33 21 5 0.75 1.24 1.02 cdfhilmrstw * es03
```

図 4.14 合計 4 ノードでクラスタの起動を試みた後, クラスタに参加しているノードの一覧を取得した結果

```
es04 | "Caused by: org.elasticsearch.cluster.coordination.CoordinationStateRejectedException: This node previously joined a cluster with UUID [VMebhfLQ8yOSLEX2nPhg] and is now trying to join a different cluster with UUID [XNOYSIDYRhythT_rKQt4w]. This is forbidden and usually indicates an incorrect discovery or cluster bootstrapping configuration. Note that the cluster UUID persists across restarts and can only be changed by deleting the contents of the node's data paths [] which will also remove any data held by this node.", es04 | "at org.elasticsearch.cluster.coordination.JoinHelper.lambda$new$8(JoinHelper.java:213) ~[elasticsearch-7.17.9.jar;7.17.9]", es04 | "at org.elasticsearch.xpack.security.transport.SecurityServerTransportInterceptor$profileSecuredRequestHandler$1.doRun(SecurityServerTransportInterceptor.java:341) ~{?:?}", es04 | "at org.elasticsearch.common.util.concurrent.AbstractRunnable.run(AbstractRunnable.java:26) ~[elasticsearch-7.17.9]", es04 | "at org.elasticsearch.transport.SecurityServerTransportInterceptor$profileSecuredRequestHandler.messageReceived(SecurityServerTransportInterceptor.java:417) ~{?:?}", es04 | "at org.elasticsearch.transport.RequestHandlerRegistry.processMessageReceived(RequestHandlerRegistry.java:67) ~[elasticsearch-7.17.9.jar;7.17.9]", es04 | "at org.elasticsearch.transport.InboundHandler$1.doRun(InboundHandler.java:272) ~[elasticsearch-7.17.9.jar;7.17.9]", es04 | "at org.elasticsearch.common.util.concurrent.ThreadContext$ContextPreservingAbstractRunnable.doRun(ThreadContext.java:777) ~ [elasticsearch-7.17.9.jar;7.17.9]", es04 | "at org.elasticsearch.common.util.concurrent.AbstractRunnable.run(AbstractRunnable.java:26) ~[elasticsearch-7.17.9.jar;7.17.9]", es04 | "at org.elasticsearch.common.util.concurrent.AbstractRunnable.run(AbstractRunnable.java:26) ~[elasticsearch-7.17.9.jar;7.17.9]", es04 | "at java.util.concurrent.ThreadPoolExecutor.yun(ThreadPoolExecutor.java:1444) ~{?:?}", es04 | "at java.lang.Threa\d.run(Thread.java:1589) ~{?:?}"] }
```

図 4.15 es04 コンテナのログ

```
matsubara@iris:~$ apt list --installed | grep elasticsearch

WARNING: apt does not have a stable CLI interface. Use with caution in scripts.

elasticsearch/stable,now 7.17.6 amd64 [installed,upgradable to: 7.17.15]
```

図 4.16 apt によって elasticsearch がインストールされたか調べた結果

```
matsubara@iris:~$ apt list elasticsearch -a
Listing... Done
elasticsearch/stable 7.17.15 amd64 [upgradable from: 7.17.6]
elasticsearch/stable 7.17.14 amd64
elasticsearch/stable 7.17.13 amd64
elasticsearch/stable 7.17.11 amd64
elasticsearch/stable 7.17.10 amd64
elasticsearch/stable 7.17.10 amd64
elasticsearch/stable 7.17.8 amd64
elasticsearch/stable 7.17.8 amd64
elasticsearch/stable 7.17.5 amd64
elasticsearch/stable 7.17.5 amd64
elasticsearch/stable 7.17.5 amd64
elasticsearch/stable 7.17.5 amd64
...
elasticsearch/stable 7.0.1 amd64
elasticsearch/stable 7.0.1 amd64
```

図 4.17 apt でインストール可能な elasticsearch のバージョンを一覧表示した結果

```
● ● ●

matsubara@iris:~$ curl -u takenaka:takenaka -s -XGET http://localhost:9200/ | grep number
"number" : "7.17.9",
```

図 4.18 ノードの起動後, Elasticsearch のバージョンを確認した 結果

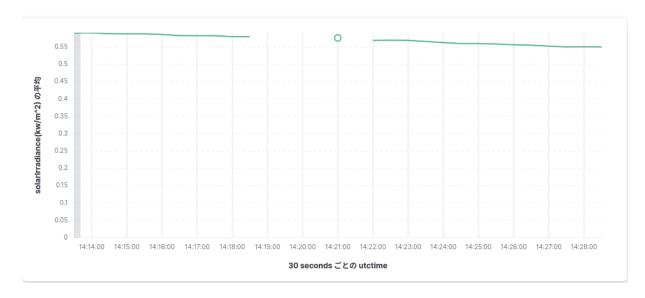


図 4.19 Elasticsearch のバージョンアップ後, 太陽光パネルの計 測データが保存されているか kibana 上で確認した結果

第5章

結論と今後の課題

本研究では、単一ノードの Elasticsearch システムから学内ゾーン内のクラス タ化システムへデータを移行するプロセスを分析し、さらにサーバゾーンに新 しい冗長化されたクラスタ化システムを構築する検証を行った.

データ移行プロセスの検証では、CO2 データと LEAF の運行日誌に関するデータの移行を行った。CO2 データの移行では、重複データの削除が必要であったが、SQLite データベースを用いた手法で対応した。LEAF の運行日誌に関するデータの移行では、同じ名前のインデックスを移行先の ElasticSearch サーバーに作成して、データをインサートすることで行った。

サーバーゾーンでのクラスタ構築の検証では、Docker、Docker Compose を使用した事前検証を行った. 異なるバージョンの Elasticsearch (7.17.6 と 7.17.9)を使用したクラスタの構築を試したが、正常に構築できないことを確認した.

学内ゾーンとサーバーゾーンでそれぞれ稼働しているクラスタごとに kibana が存在しており、本研究室で管理する Elasticsearch に保存されたデータを一元 的に管理、閲覧することが出来ないので、kibana の統合による一元管理が出来る状態を目指す.

謝辞

本研究を行うにあたり、終始、懇切丁寧な御指導と適切な御助言を賜りました本学工学部電気電子工学科通信システム工学研究室の都築伸二教授に深甚なる感謝の意を表します。

参考文献

[1]	Elasticsearch B.\							
	"Install	Elasticsearch	with	Docker	_			
	Elasticsearch	Guide	[7.17]		Elastic ",			
	https://www.elastic.co/guide/en/elasticsearch/reference/7.17/							
	参照 Nov 20,2023	3.						

[2] RAKUS Developers Blog, "Docker とは一体何なんだ?【初心者向け】 - RAKUS Developers Blog — ラクス エンジニアプログ ", https://techblog.rakus.co.jp/entry/20221007/docker, 参照 Nov 20,2023.