[AWS Black Belt Online Seminar] Amazon Elasticsearch Service

サービスカットシリーズ

Solutions Architect, Analytics
Makoto Shimura
2020/06/23

AWS 公式 Webinar
https://amzn.to/JPWebinar

過去資料
https://amzn.to/JPArchive

自己紹介

志村 誠

ソリューションアーキテクト

- データ分析・機械学習系サービスを担当
- 好きなサービス
  - Amazon Athena
  - AWS Glue
  - Amazon SageMaker
  - ...and Amazon Elasticsearch Service!!

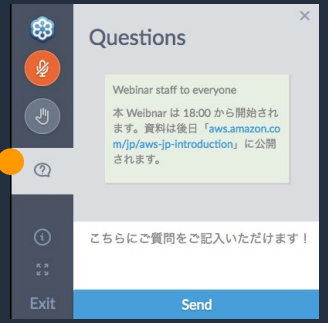
AWS Black Belt Online Seminar とは

「サービス別」「ソリューション別」「業種別」のそれぞれのテーマに分かれて、アマゾン ウェブ サービス ジャパン株式会社が主催するオンラインセミナーシリーズです。

質問を投げることができます！

- 書き込んだ質問は、主催者にしか見えません
- 今後のロードマップに関するご質問は
  お答えできませんのでご了承下さい

① 吹き出しをクリック
② 質問を入力
③ Sendをクリック



Twitter ハッシュタグは以下をご利用ください
#awsblackbelt

内容についての注意点

- 本資料では2020年06月23日時点のサービス内容および価格についてご説明しています。最新の情報はAWS公式ウェブサイト(http://aws.amazon.com)にてご確認ください。

- 資料作成には十分注意しておりますが、資料内の価格とAWS公式ウェブサイト記載の価格に相違があった場合、AWS公式ウェブサイトの価格を優先とさせていただきます。

- 価格は税抜表記となっています。日本居住者のお客様には別途消費税をご請求させていただきます。

- AWS does not offer binding price quotes. AWS pricing is publicly available and is subject to change in accordance with the AWS Customer Agreement available at http://aws.amazon.com/agreement/. Any pricing information included in this document is provided only as an estimate of usage charges for AWS services based on certain information that you have provided. Monthly charges will be based on your actual use of AWS services, and may vary from the estimates provided.

アジェンダ

1. Elasticsearch の概要
2. Amazon Elasticsearch Service (Amazon ES) の概要
3. Amazon ES によるログ分析
4. Amazon ES による検索
5. Amazon ES の運用管理
6. Amazon ES のセキュリティ
7. 料金や制限事項
8. まとめ

aws

アジェンダ

1. Elasticsearch の概要
2. Amazon Elasticsearch Service (Amazon ES) の概要
3. Amazon ES によるログ分析
4. Amazon ES による検索
5. Amazon ES の運用管理
6. Amazon ES のセキュリティ
7. 料金や制限事項
8. まとめ

aws

Elasticsearch

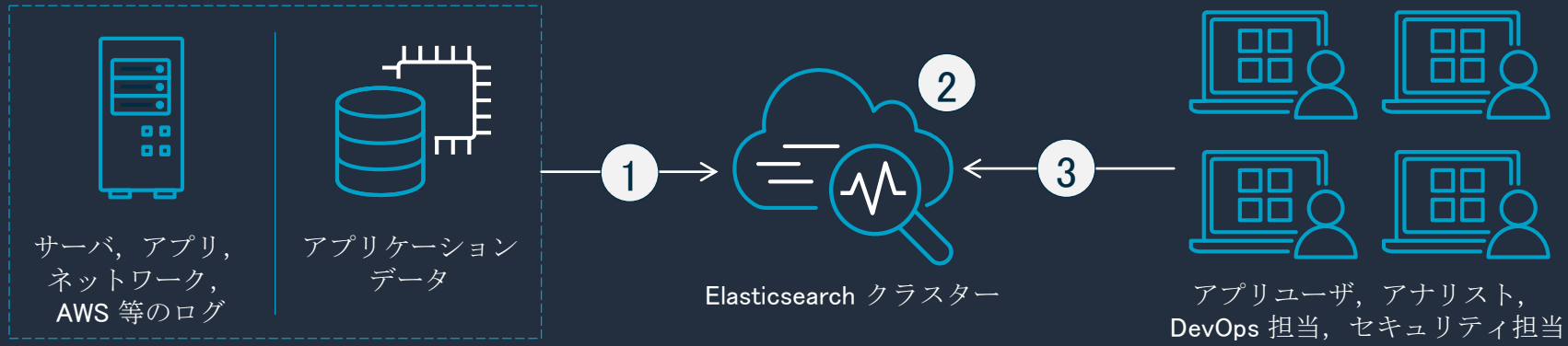
- ログ分析や検索に関する様々なユースケースで利用できる，分散型 RESTful 検索/分析エンジン．2020/6 時点の最新バージョンは 7.8
- コアエンジン部はオープンソースソフトウェアとして提供され，主に Elastic 社によって開発が行われている
- Elasticsearch に付随するソフトウェアとして，データ取り込みの Logstash や Beats，可視化用の Kibana などがある

データ挿入から活用までの流れ

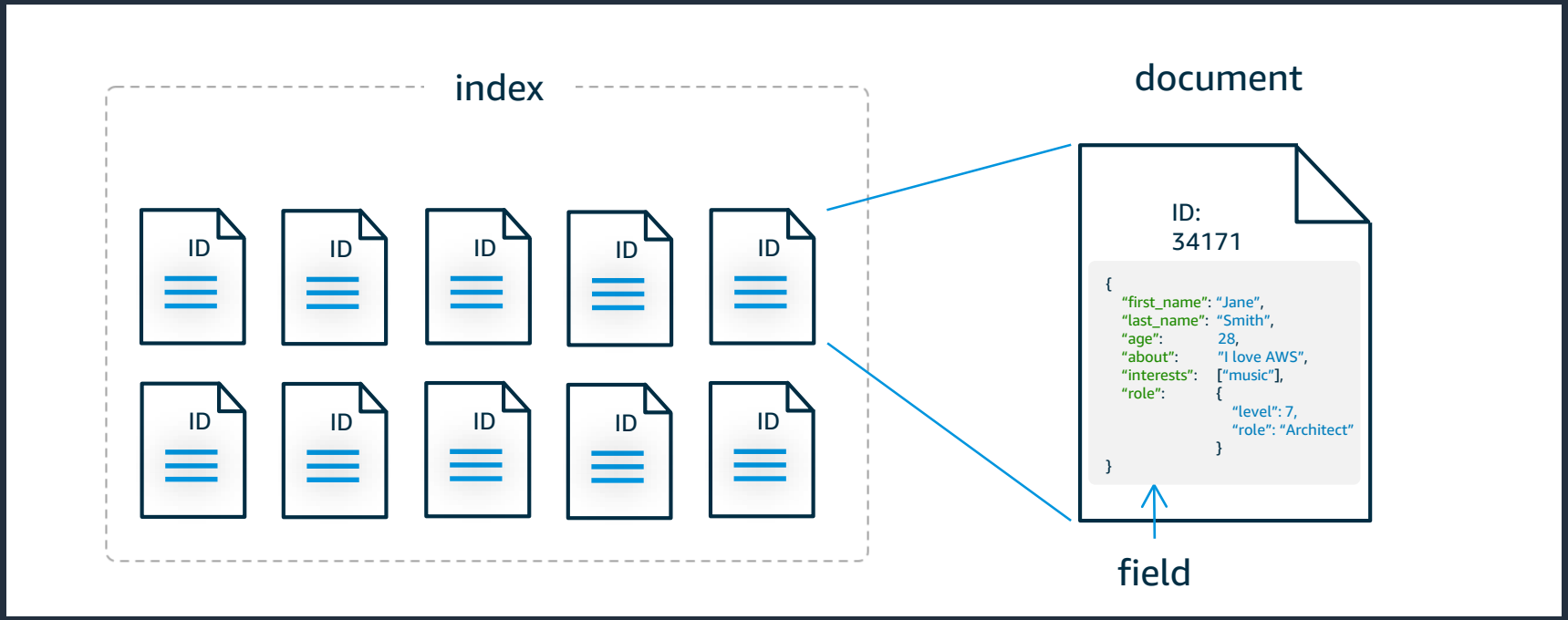
1
JSON 形式のデータを
REST API 経由で送信

2
インデックスに格納された
全てのデータが検索可能

3
REST API 経由でクエリ
複雑な検索・分析条件に対応

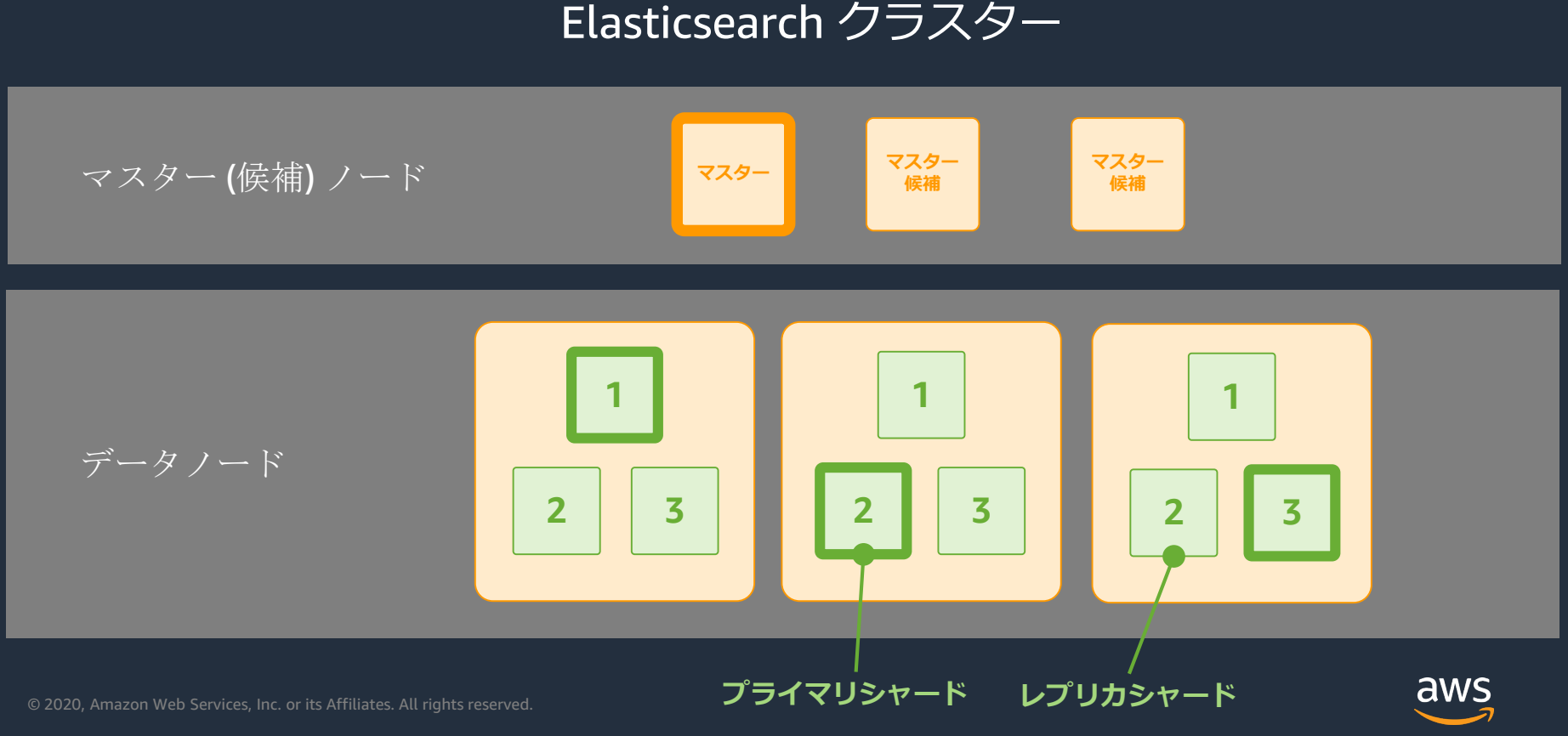


Elasticsearch における論理的なデータの持ち方



*バージョン 6.x までの Elasticsearch には type という概念があったが，7.0 以降では廃止されている

Elasticsearch における物理的なデータの持ち方



アジェンダ

1. Elasticsearch の概要
2. Amazon Elasticsearch Service (Amazon ES) の概要
3. Amazon ES によるログ分析
4. Amazon ES による検索
5. Amazon ES の運用管理
6. Amazon ES のセキュリティ
7. 料金や制限事項
8. まとめ

aws

Amazon Elasticsearch Service (Amazon ES) とは

Amazon Elasticsearch Service (Amazon ES) は，
Elasticsearch と Kibana を簡単に
デプロイ・管理し，スケールさせることが
可能なフルマネージドサービス

Amazon ES における 2 種類の API

設定 API

Amazon ES サービスの API．Amazon ES ドメインやタグ，サービスロール等の操作を行うためのもの

ex. AddTags, DescribeElasticsearchDomains, DeleteElasticsearchServiceRole etc...

Elasticsearch API

Amazon ES 上で動く Elasticsearch クラスタの API．インデックスの作成や削除，ドキュメントの追加，検索クエリの実行等の操作を行うためのもの

ex. _search, _reindex, _cat/indices etc...

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-configuration-api.html
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/aes-supported-es-operations.html

サポートしている Elasticsearch バージョンと API

サポート Elasticsearch バージョン

- 7.4, 7.1
- 6.8, 6.7, 6.5, 6.4, 6.3, 6.2, 6.0
- 5.6, 5.5, 5.3, 5.1
- 2.3
- 1.5

バージョン 7.4 でサポートされている API

- /*index-name*/_close を除く、インデックスパス内のすべてのオペレーション (/*index-name*/_forcemerge や /*index-name*/update/*id* など)
- /_alias
- /_aliases
- /_all
- /_analyze
- /_bulk
- /_cat (/_cat/nodeattrs を除く)
- /_cluster/allocation/explain
- /_cluster/health
- /_cluster/pending_tasks
- /_cluster/settings を使用する複数のプロパティ [4]:
  - action.auto_create_index
  - action.search.shard_count.limit
  - indices.breaker.fielddata.limit
  - indices.breaker.request.limit
  - indices.breaker.total.limit
  - cluster.max_shards_per_node
- /_cluster/state
- /_cluster/stats
- /_count
- /_delete_by_query[1]
- /_explain
- /_field_caps
- /_field_stats
- /_flush
- /_ingest/pipeline
- /_mapping
- /_mget
- /_msearch
- /_mtermvectors
- /_nodes
- /_opendistro/_alerting
- /_opendistro/_ism
- /_opendistro/_security
- /_opendistro/_sql
- /_percolate
- /_plugin/kibana
- /_rank_eval
- /_refresh
- /_reindex[1]
- /_render
- /_rollover
- /_scripts[3]
- /_search[2]
- /_search profile
- /_shard_stores
- /_shrink[5]
- /_snapshot
- /_split
- /_stats
- /_status
- /_tasks
- /_template
- /_update_by_query[1]
- /_validate

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/what-is-amazon-elasticsearch-service.html#aes-choosing-version
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/aes-supported-es-operations.html

aws

サポートしているインスタンスタイプとサイズ

インスタンスタイプ・サイズごとに，最大 EBS サイズや HTTP リクエストペイロードの最大値が異なる．詳細はドキュメントを参照

インスタンスタイプ	ストレージ	対応バージョン	保管時のデータ暗号化	詳細アクセスコントロール	Ultrawarm	その他
C4	EBS	すべて	○	○	○	
C5	EBS	5.1 以降	○	○	○	
I2	内蔵 SSD	すべて	○	○	○	
I3	内蔵 SSD	5.1 以降	○	○	○	
M3	EBS / 内蔵 SSD	すべて	×	×	○	
M4	EBS	すべて	○	○	○	
M5	EBS	5.1 以	○	○	○	
R3	EBS / 内蔵 SSD	すべて	×	×	○	
R4	EBS	すべて	○	○	○	
R5	EBS	5.1 以降	○	○	○	
T2	EBS	T2.micro は 1.5 / 2.3 のみ	×	×	×	ドメインのインスタンス数 10 以下まで

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/aes-supported-instance-types.html
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/aes-limits.html

Amazon ES では Elasticsearch ディストリビューションとして Open Distro を採用

Open Distro
for Elasticsearch

エンタープライズグレードのセキュリティ，アラート、SQL などにより強化された Elasticsearch の Apache 2.0 ライセンスのディストリビューション

aws

Amazon ES の利点

フルマネージド型

APIやAWSコンソールを利用して、わずか数分でクラスターをデプロイできる

柔軟性

データの検索やログの分析が可能

AWS とオープンソースの取り込みツールをサポートする

優れたコスト効率

従量課金制の支払い

運用コストの削減、適切なインスタンスタイプのサイジング

リザーブドインスタンスも選択可能

高可用性

セルフヒーリング、24 時間 365 日のモニタリング、1クリックでマルチ AZ活用、自動バックアップ、AWS サポートの利用、Amazon CloudWatch でのメトリクス取得

スケーラブル & 高性能

ワンクリックで、スケール、Elasticsearchバージョンのアップグレード、パッチ適用

セキュア

Amazon VPC 内へのデプロイ、Amazon Cognitoによるログイン

HIPAA、FISMA、SOC、PCI、FedRampに準拠

Amazon ES を活用している多くのお客さま

改善につながるインサイト

アプリ API の利用状況，インフラの起動時間，セキュリティ等のログで取得して，改善につなげる

エラーの原因を発見

ログからエラーメッセージやサーバ ID を検索して，起こっている問題の原因を把握

リアルタイムに対応

データをリアルタイムで取得，可視化，そしてダッシュボードにまとめることで，問題に素早く対処できるように

データの検索

アプリケーションデータに関連する検索結果を返す

ログの集約

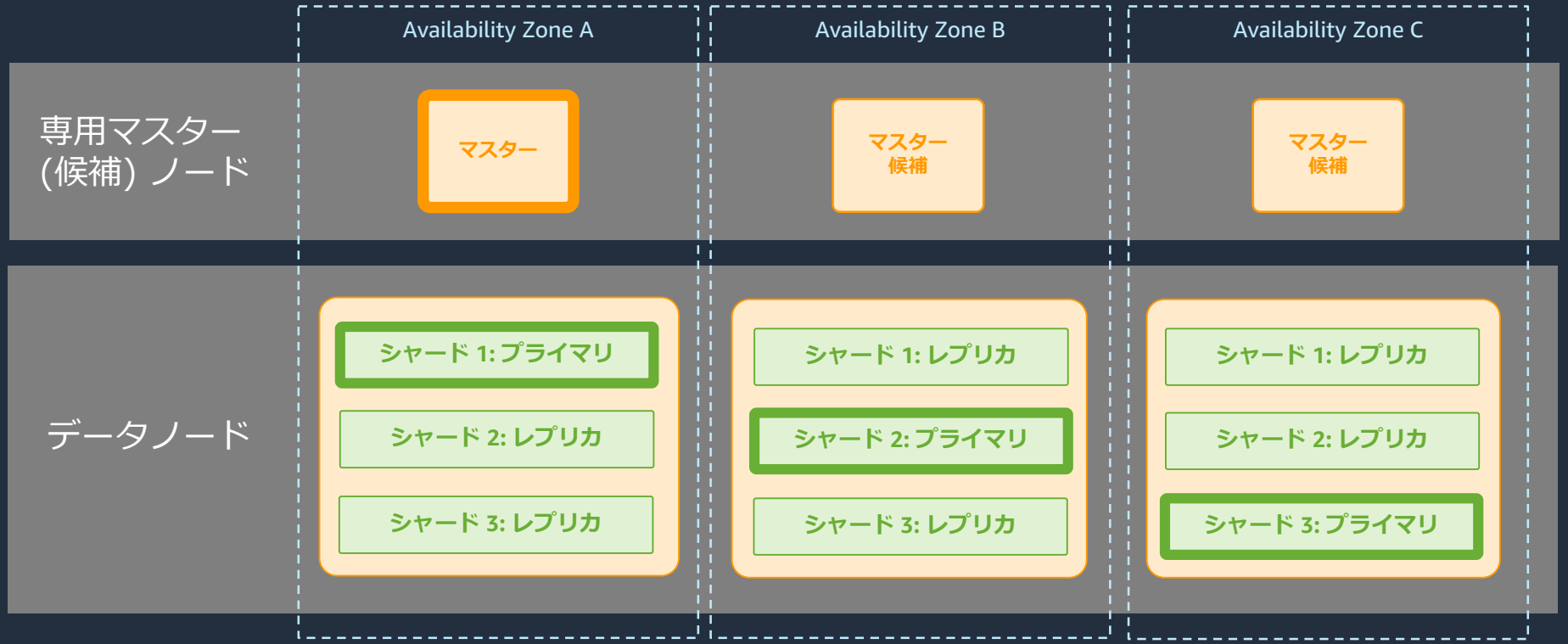
インフラやアプリケーションの情報を集約する単一ダッシュボードの作成

セキュリティ監視

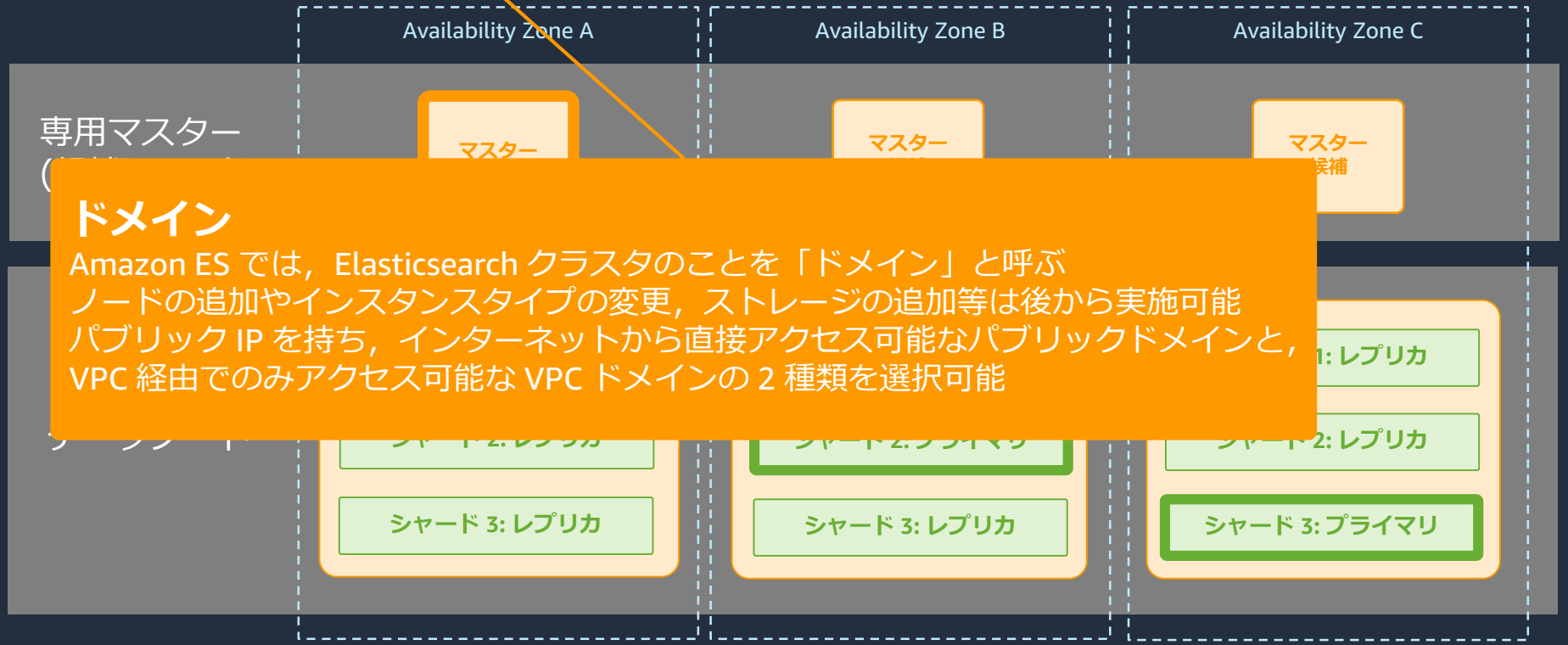
不正利用，DDoS，その他のサイバー攻撃に対応するための SIEM 基盤を構築

aws

Amazon ES ドメインの推奨アーキテクチャ



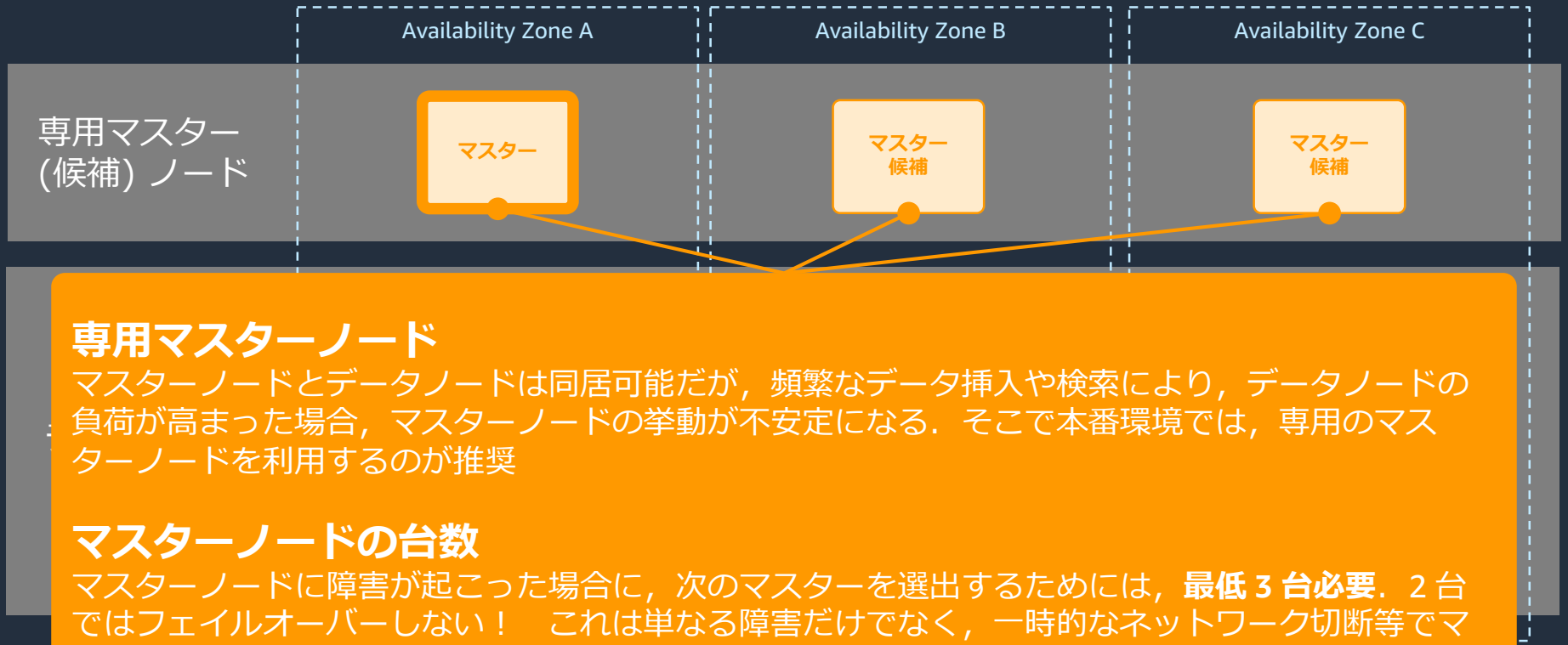
Amazon ES ドメインの推奨アーキテクチャ



ドメイン

Amazon ES では，Elasticsearch クラスタのことを「ドメイン」と呼ぶ
ノードの追加やインスタンスタイプの変更，ストレージの追加等は後から実施可能
パブリック IP を持ち，インターネットから直接アクセス可能なパブリックドメインと，
VPC 経由でのみアクセス可能な VPC ドメインの 2 種類を選択可能

Amazon ES ドメインの推奨アーキテクチャ



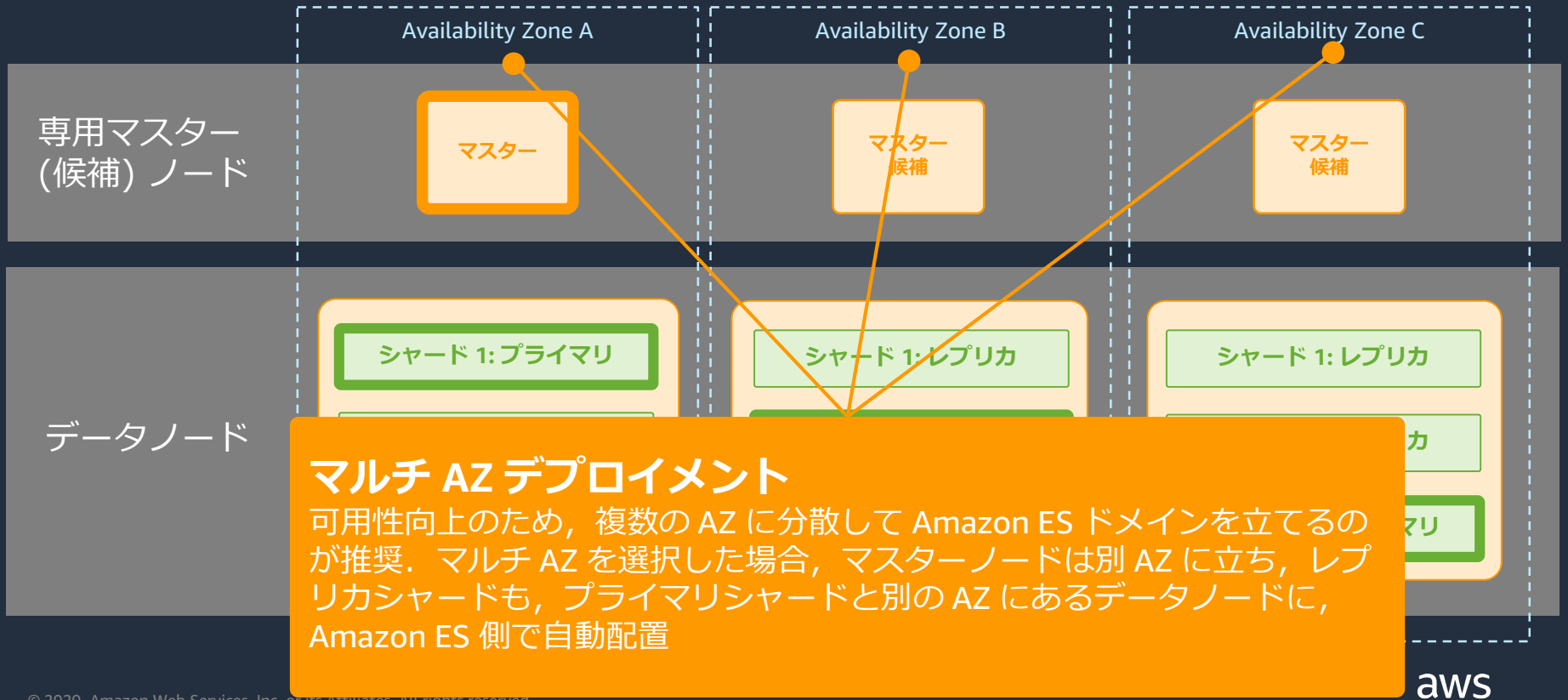
専用マスターノード

マスターノードとデータノードは同居可能だが，頻繁なデータ挿入や検索により，データノードの負荷が高まった場合，マスターノードの挙動が不安定になる．そこで本番環境では，専用のマスターノードを利用するのが推奨

マスターノードの台数

マスターノードに障害が起こった場合に，次のマスターを選出するためには，**最低 3 台必要**．2 台ではフェイルオーバーしない！　これは単なる障害だけでなく，一時的なネットワーク切断等でマスターノードが分割されてしまうこと等も考慮した，分散システムとしての設計思想によるもの

Amazon ES ドメインの推奨アーキテクチャ

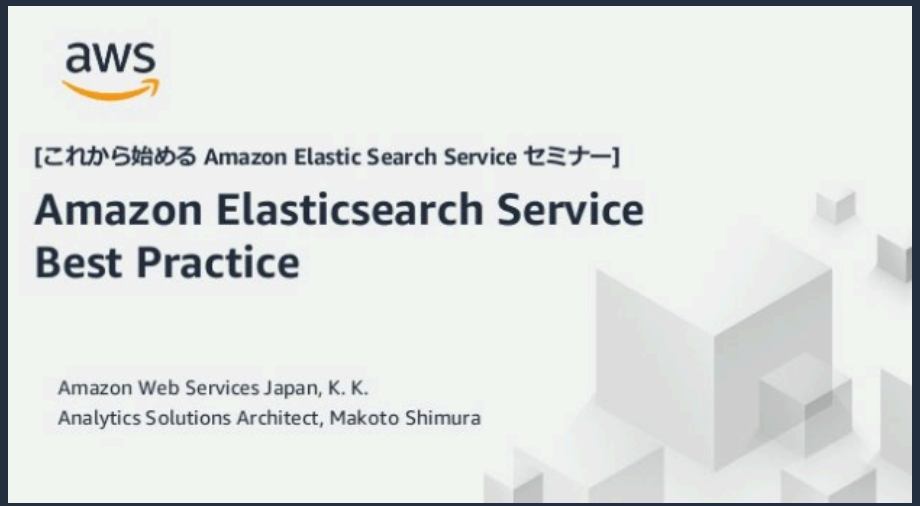


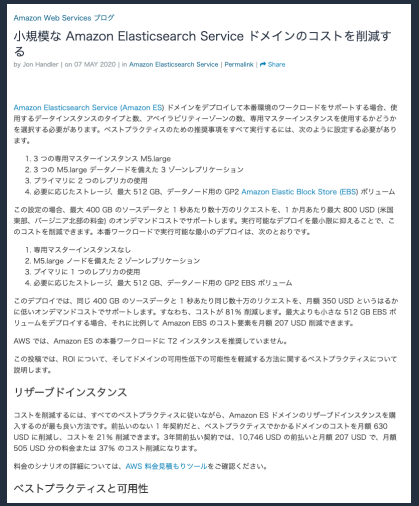
マルチ AZ デプロイメント

可用性向上のため，複数の AZ に分散して Amazon ES ドメインを立てるのが推奨．マルチ AZ を選択した場合，マスターノードは別 AZ に立ち，レプリカシャードも，プライマリシャードと別の AZ にあるデータノードに，Amazon ES 側で自動配置

Amazon ES のベストプラクティス

Amazon ES の運用を効率的に行うためのベストプラクティスがまとまっている
インスタンスタイプとサイズ，インデックスの大きさ，シャード数，ストレージサイズ，
コスト最適化，管理と監視 etc...




Amazon Web Services ブログ

小規模な Amazon Elasticsearch Service ドメインのコストを削減する

by Jon Handler | on 07 MAY 2020 | in Amazon Elasticsearch Service | Permalink | Share

Amazon Elasticsearch Service (Amazon ES) ドメインをデプロイして本番環境のワークロードをサポートする場合、使用するデータインスタンスのタイプと数、アベイラビリティーゾーンの数、専用マスターインスタンスを使用するかどうかを選択する必要があります。ベストプラクティスのための推奨事項をすべて実行するには、次のように設定する必要があります。

1. 3 つの専用マスターインスタンス M5.large
2. 3 つの M5.large データノードを備えた 3 ゾーンレプリケーション
3. プライマリに 2 つのレプリカの使用
4. 必要に応じたストレージ、最大 512 GB、データノード用の GP2 Amazon Elastic Block Store (EBS) ボリューム

この設定の場合、最大 400 GB のソースデータと 1 秒あたり数十万のリクエストを、1 か月あたり最大 800 USD (米国東部、バージニア北部の料金) のオンデマンドコストでサポートします。実行可能なデプロイを最小限に抑えることで、このコストを削減できます。本番ワークロードで実行可能な最小のデプロイは、次のとおりです。

1. 専用マスターインスタンスなし
2. M5.large ノードを備えた 2 ゾーンレプリケーション
3. プライマリに 1 つのレプリカの使用
4. 必要に応じたストレージ、最大 512 GB、データノード用の GP2 EBS ボリューム

このデプロイでは、同じ 400 GB のソースデータと 1 秒あたり同じ数十万のリクエストを、月額 350 USD というはるかに低いオンデマンドコストでサポートします。すなわち、コストが 81% 削減します。最大よりも小さな 512 GB EBS ボリュームをデプロイする場合、それに比例して Amazon EBS のコスト要素を月額 207 USD 削減できます。

AWS では、Amazon ES の本番ワークロードに T2 インスタンスを推奨していません。

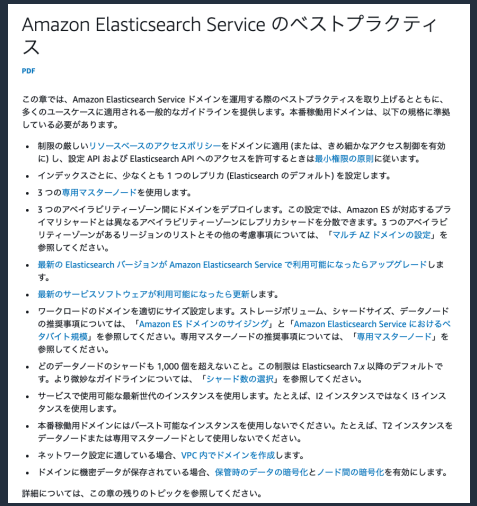
この投稿では、ROI について、そしてドメインの可用性低下の可能性を軽減する方法に関するベストプラクティスについて説明します。

リザーブドインスタンス

コストを削減するには、すべてのベストプラクティスに従いながら、Amazon ES ドメインのリザーブドインスタンスを購入するのが最も良い方法です。前払いのない 1 年契約だと、ベストプラクティスでかかるドメインのコストを月額 630 USD に削減し、コストを 21% 削減できます。3年間前払い契約では、10,746 USD の前払いと月額 207 USD で、月額 505 USD 分の料金または 37% のコスト削減になります。

料金のシナリオの詳細については、AWS 料金見積もりツールをご確認ください。

ベストプラクティスと可用性


Amazon Elasticsearch Service のベストプラクティス

PDF

この章では、Amazon Elasticsearch Service ドメインを運用する際のベストプラクティスを取り上げるとともに、多くのユースケースに適用される一般的なガイドラインを提供します。本番稼働用ドメインは、以下の規格に準拠している必要があります。

- 制限の厳しいリソースベースのアクセスポリシーをドメインに適用 (または、きめ細かなアクセス制御を有効に) し、設定 API および Elasticsearch API へのアクセスを許可するときは最小権限の原則に従います。
- インデックスごとに、少なくとも 1 つのレプリカ (Elasticsearch のデフォルト) を設定します。
- 3 つの専用マスターノードを使用します。
- 3 つのアベイラビリティーゾーン間にドメインをデプロイします。この設定では、Amazon ES が対応するプライマリシャードとは異なるアベイラビリティーゾーンにレプリカシャードを分散できます。3 つのアベイラビリティーゾーンがあるリージョンのリストとその他の考慮事項については、「マルチ AZ ドメインの設定」を参照してください。
- 最新の Elasticsearch バージョンが Amazon Elasticsearch Service で利用可能になったらアップグレードします。
- 最新のサービスソフトウェアが利用可能になったら更新します。
- ワークロードのドメインを適切にサイズ設定します。ストレージボリューム、シャードサイズ、データノードの推奨事項については、「Amazon ES ドメインのサイジング」と「Amazon Elasticsearch Service におけるペタバイト規模」を参照してください。専用マスターノードの推奨事項については、「専用マスターノード」を参照してください。
- どのデータノードのシャードも 1,000 個を超えないこと。この制限は Elasticsearch 7.x 以降のデフォルトです。より微妙なガイドラインについては、「シャード数の選択」を参照してください。
- サービスで使用可能な最新世代のインスタンスを使用します。たとえば、I2 インスタンスではなく I3 インスタンスを使用します。
- 本番稼働用ドメインにはバースト可能なインスタンスを使用しないでください。たとえば、T2 インスタンスをデータノードまたは専用マスターノードとして使用しないでください。
- ネットワーク設定に適している場合、VPC 内でドメインを作成します。
- ドメインに機密データが保存されている場合、保管時のデータの暗号化とノード間の暗号化を有効にします。

詳細については、この章の残りのトピックを参照してください。

https://www.slideshare.net/AmazonWebServicesJapan/20200414-amazon-elasticsearch-service-best-practice
https://aws.amazon.com/jp/blogs/news/reducing-cost-for-small-amazon-elasticsearch-service-domains/
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/aes-bp.html

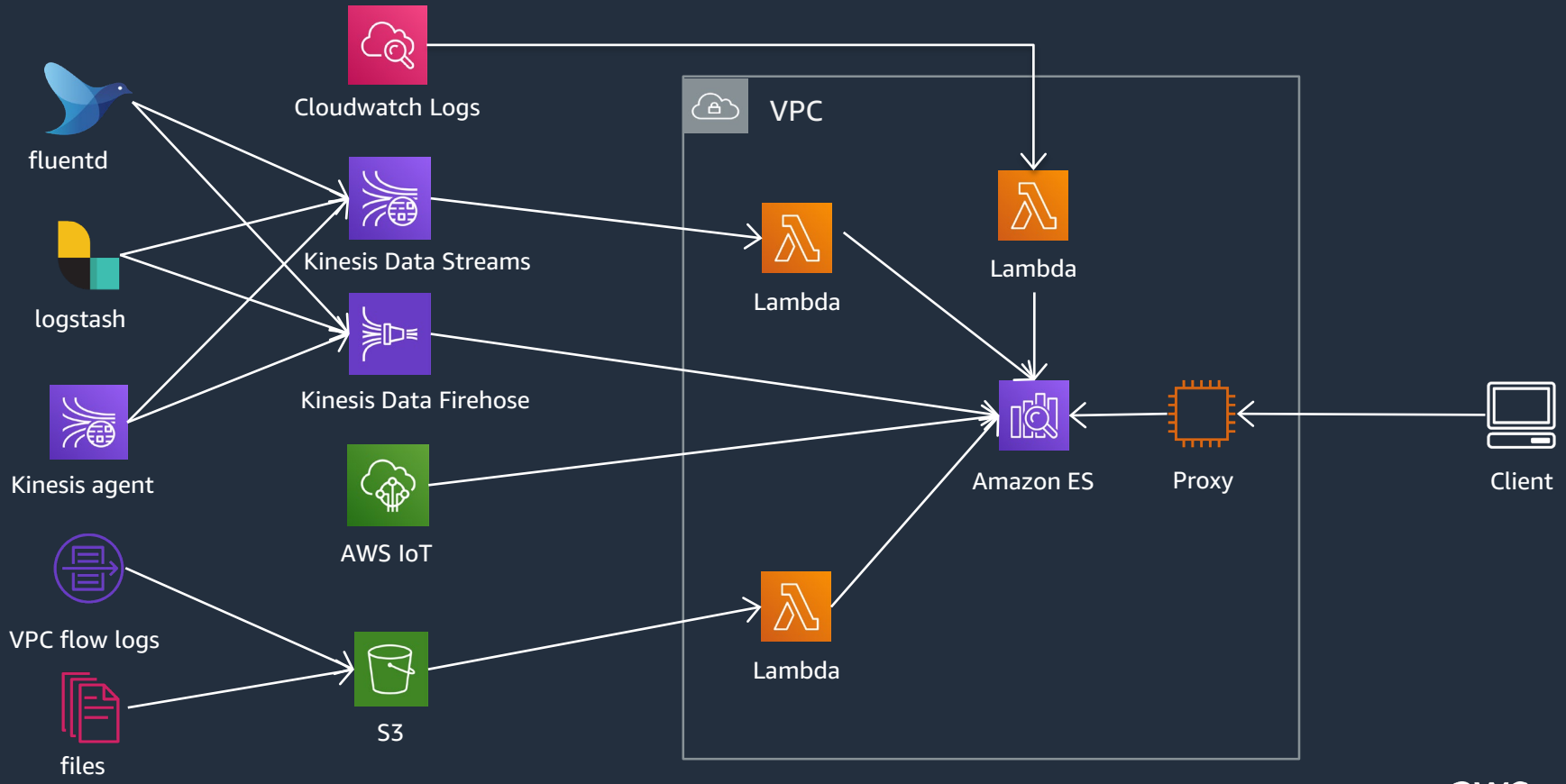
aws

アジェンダ

1. Elasticsearch の概要
2. Amazon Elasticsearch Service (Amazon ES) の概要
3. Amazon ES によるログ分析
4. Amazon ES による検索
5. Amazon ES の運用管理
6. Amazon ES のセキュリティ
7. 料金や制限事項
8. まとめ

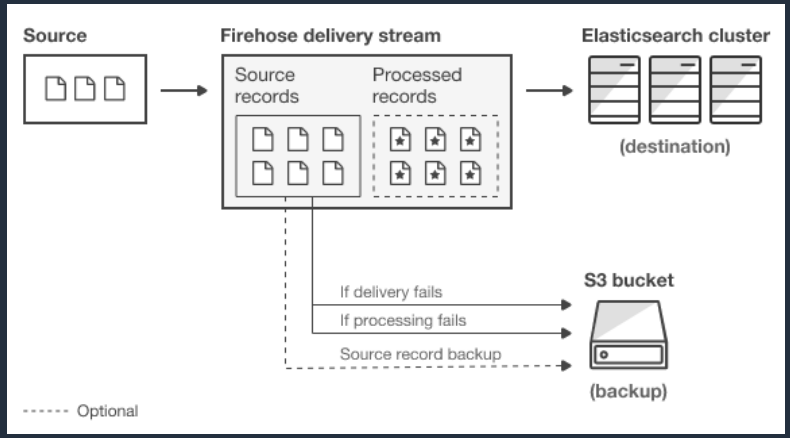
aws

Amazon ES を使ったログ分析



Amazon ES へのデータ投入

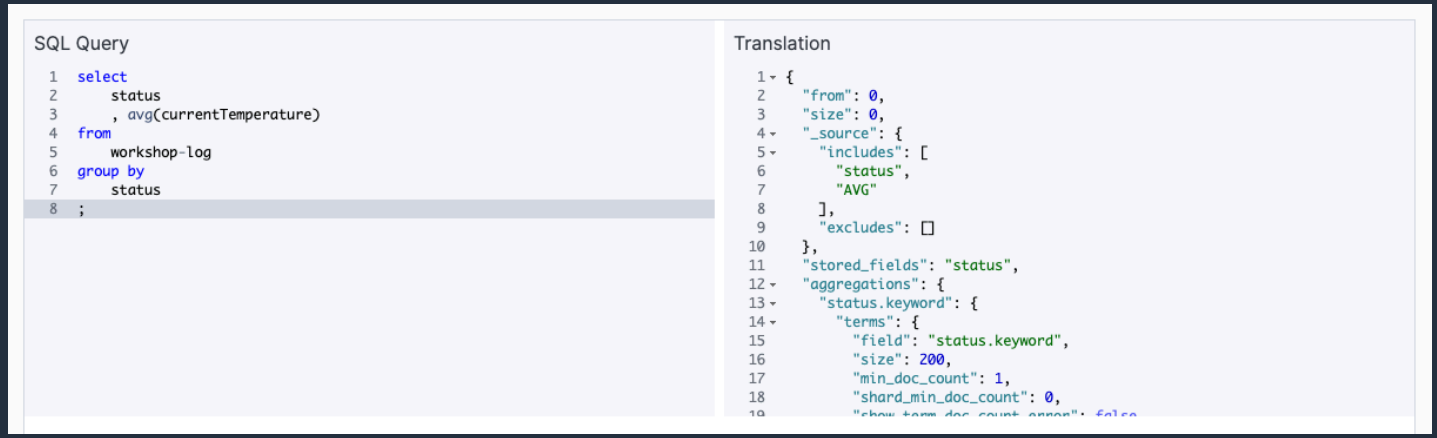
- ログ分析の場合，Amazon Kinesis Data Firehose 経由でストリームデータを，直接 Amazon ES に投入するのが定番のやり方．パブリックドメインだけでなく，**VPC ドメイン**の Amazon ES にも対応．また**別アカウントの Firehose** からの投入も可能
- Cloudwatch Logs サブスクリプションにより，Lambda を経由したストリームデータ投入も行える
- Lambda 等から Elasticsearch の Bulk API を叩いて，S3 上のデータを定期的にインポートするといったことも可能



https://aws.amazon.com/jp/about-aws/whats-new/2019/10/amazon-kinesis-data-firehose-adds-cross-account-delivery-to-amazon-elasticsearch-service/
https://aws.amazon.com/jp/blogs/news/ingest-streaming-data-into-amazon-elasticsearch-service-within-the-privacy-of-your-vpc-with-amazon-kinesis-data-firehose/
https://docs.aws.amazon.com/ja_jp/AmazonCloudWatch/latest/logs/CWL_ES_Stream.html
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-aws-integrations.html

SQL インターフェースのサポート

- 通常の Elasticsearch 検索 API ではなく，SQL を使用したデータ分析が可能．JDBC ドライバーが提供されており，BI ツールなどからも利用できる
- Kibana に SQL Workbench が実装されており，SQL で直接分析が可能．SQL を Search API の構文に変換する機能も実装済
- CLI から Amazon ES に認証付きで SQL クエリを実行できる SQL CLI もある



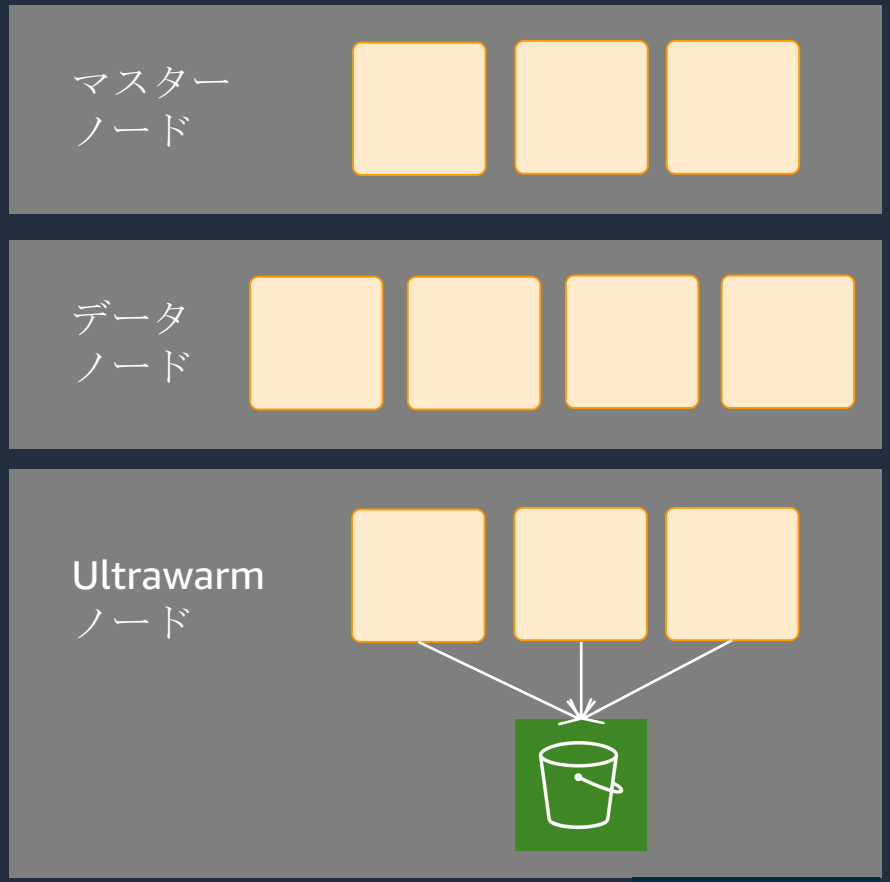
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/sql-support.html
https://opendistro.github.io/for-elasticsearch-docs/docs/sql/
https://opendistro.github.io/for-elasticsearch-docs/docs/sql/cli/

大規模ログ分析のための Ultrawarm ノード

- 大規模ログ分析を低コストで実現するための，Amazon ES オリジナルの新しいノードタイプ
- Ultrawarm ノードは，S3 に保持したデータに対してクエリを実行
- Ultrawarm ノードに移動されたインデックスは，読み取り専用となる
- インデックスごとに API を呼んで hot/warm の移動を行う

  POST _ultrawarm/migration/my-index/_warm
  POST _ultrawarm/migration/my-index/_hot

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/ultrawarm.html



Ultrawarm の特徴

大量のログ
データのストア

Elasticsearch のスナップショットとレプリカを排除した形で，S3 にデータを保存．また使用したストレージぶんの料金だけを支払い．I3 インスタンスと比較して最大 90% のコスト削減

対話型の
ログ分析と可視化

既存の検索 API と完全な互換性を持ち，通常のデータノードと Ultrawarm ノードにまたがった分析が可能．Kibana でインタラクティブな分析を行うのに最適化されている

高い性能

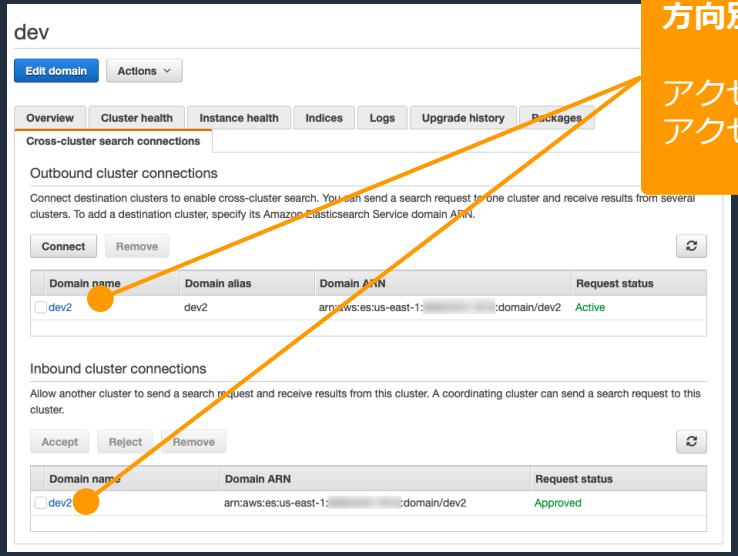
多層の詳細なキャッシュ，アダプティブなプリフェッチ，クエリエンジンの最適化により，高速なパフォーマンスが得られる

キャッシュされていない場合でも，従来の HDD ベースインスタンスより最大 2 倍高速

aws

クラスター横断でのクエリ実行

- 複数のドメインを横断した形で，検索クエリが実行可能
- 用途毎に最適なサイズのドメインを持ち，他ドメインのデータが必要なときのみ横断クエリを実行
- Kibana のダッシュボードも，他ドメインデータを参照して作成可能
- 自己管理の Elasticsearch クラスターは非対応．その他注意点はドキュメントを参照



方向別にコネクションを設定

アクセスしにいく先のドメイン
アクセスしにくるドメイン

```
GET dev2:jpdocs/_search
{
  "query": {
    "match_all": {}
  }
}
```

検索 API の実行時にドメインエイリアスを追加
GET DOMAIN_ALIAS/INDEX_NAME/_search

https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/cross-cluster-search.html
https://opendistro.github.io/for-elasticsearch-docs/docs/security-access-control/cross-cluster-search/

アジェンダ

1. Elasticsearch の概要
2. Amazon Elasticsearch Service (Amazon ES) の概要
3. Amazon ES によるログ分析
4. Amazon ES による検索
5. Amazon ES の運用管理
6. Amazon ES のセキュリティ
7. 料金や制限事項
8. まとめ

日本語全文検索用のプラグイン

Kuromoji プラグイン (analysis-kuromoji)

- 日本語形態素解析用のプラグイン
- 単語分割，品詞タグ付け，基本形抽出等の機能を持つ
- _analyze API で形態素解析の結果を確認可能

ICU プラグイン (analysis-icu)

- Unicode 正規化用のプラグイン
- 「㍱ → キログラム」「① → 1」など

https://www.atilika.com/ja/kuromoji/
https://www.elastic.co/guide/en/elasticsearch/plugins/current/analysis-kuromoji-analyzer.html
http://site.icu-project.org/
https://www.elastic.co/guide/en/elasticsearch/plugins/current/analysis-icu.html

aws

ユーザー辞書ルールによるカスタム辞書の使用

- Kuromoji プラグインによる形態素解析の際に，固有名詞やドメイン用語などを正しく判別するために，カスタム辞書を登録したい場合がある
- Elasticsearch 7.4 以降では，user_dictionary_rules を使用することで，テンプレートマッピング内にカスタム辞書を記述可能に
- インデックス作成時にのみ設定可能で，変更するためにはインデックスの再作成が必要

```
PUT my_index
{
  "mappings" : {
    "properties" : {
      "content" : {
        "type" : "text",
        "analyzer": "my_analyzer"
      }
    }
  },
  "settings": {
    "index": {
      "analysis": {
        "tokenizer": {
          "my_kuromoji_tokenizer": {
            "type": "kuromoji_tokenizer",
            "mode": "search",
            "user_dictionary_rules": [
              "高輪ゲートウェイ,高輪 ゲートウェイ,タカナワ ゲートウェイ,カスタム名詞"
            ]
          }
        }
      }
    }
  }
}
```


aws

カスタムパッケージによるファイルベースの辞書の使用

- より大規模な辞書を登録する際に，S3 に置いた辞書やシノニム用のファイルを，パッケージとして Amazon ES ドメインに読み込む機能
- 一度作成したパッケージは，複数の Amazon ES ドメインに適用可能
- インデックス作成時にのみ設定可能で，変更にはインデックスの再作成が必要
- あくまで既存プラグイン用のファイルを利用できるもので，プラグインの追加は非対応

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/custom-packages.html

kNN による高速な最近傍探索

- kNN は，ベクトル空間内の最も近い k 個の点を，高速に探すための手法
- いわゆる類似検索の用途で使用される
- Amazon ES では，nmslib というライブラリに含まれる hnsw という高速・高精度な近似近傍探索アルゴリズムを使用

```
PUT my-index
{
  "settings": {
    "index.knn": true
  },
  "mappings": {
    "properties": {
      "my_vector1": {
        "type": "knn_vector",
        "dimension": 2
      },
      "my_vector2": {
        "type": "knn_vector",
        "dimension": 4
      }
    }
  }
}
```

```
POST _bulk
{ "index": { "_index": "my-index", "_id": "1" } }
{ "my_vector1": [1.5, 2.5], "price": 12.2 }
{ "index": { "_index": "my-index", "_id": "2" } }
{ "my_vector1": [2.5, 3.5], "price": 7.1 }
{ "index": { "_index": "my-index", "_id": "3" } }
{ "my_vector1": [3.5, 4.5], "price": 12.9 }
{ "index": { "_index": "my-index", "_id": "4" } }
{ "my_vector1": [5.5, 6.5], "price": 1.2 }
{ "index": { "_index": "my-index", "_id": "5" } }
{ "my_vector1": [4.5, 5.5], "price": 3.7 }
{ "index": { "_index": "my-index", "_id": "6" } }
{ "my_vector2": [1.5, 5.5, 4.5, 6.4], "price": 10.3 }
{ "index": { "_index": "my-index", "_id": "7" } }
{ "my_vector2": [2.5, 3.5, 5.6, 6.7], "price": 5.5 }
{ "index": { "_index": "my-index", "_id": "8" } }
{ "my_vector2": [4.5, 5.5, 6.7, 3.7], "price": 4.4 }
{ "index": { "_index": "my-index", "_id": "9" } }
{ "my_vector2": [1.5, 5.5, 4.5, 6.4], "price": 8.9 }
```

```
GET my-index/_search
{
  "size": 2,
  "query": {
    "knn": {
      "my_vector2": {
        "vector": [2, 3, 5, 6],
        "k": 2
      }
    }
  }
}
```

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/knn.html
https://opendistro.github.io/for-elasticsearch-docs/docs/knn/

アジェンダ

1. Elasticsearch の概要
2. Amazon Elasticsearch Service (Amazon ES) の概要
3. Amazon ES によるログ分析
4. Amazon ES による全文検索
5. Amazon ES の管理
6. Amazon ES のセキュリティ
7. 料金や制限事項
8. まとめ

Index State Management (ISM) でインデックス管理の自動化

- ISM 機能により，日/週/月単位で新しく作られるインデックスの管理を自動化．従来 Curator で行う必要があったものを，Amazon ES 側で設定可能に
- Kibana 上でインデックス管理ポリシーを設定・管理
- _template API と併用すること で，新しいインデックスにポリシーを自動で反映


```
PUT _template/index_policy_templete
{
  "index_patterns": ["access-log-*"],
  "settings": {
    "opendistro.index_state_management.policy_id": "my_policy"
  }
}
```

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/ism.html
https://opendistro.github.io/for-elasticsearch-docs/docs/ism/
https://curator.readthedocs.io/en/latest/index.html

ロギング

Elasticsearch API のログ

- 以下の 3 種類のログを Cloudwatch に出力
  - インデックススローログ（ドキュメントの追加・削除・更新）
  - 検索スローログ（検索クエリ）
  - エラーログ（WARN, ERROR, FATAL, DEBUG の例外）
- スローログは，ログレベルごとにログを吐き出す閾値を設定可能
- デフォルトでは無効．有効にすると Cloudwatch 利用料金が別途必要

設定 API のログ

- 通常の AWS サービスと同様，API コールのログを Cloudtrail に出力

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-createupdatedomains.html#es-createdomain-configure-slow-logs
https://aws.amazon.com/jp/blogs/news/viewing-amazon-elasticsearch-service-error-logs/
https://aws.amazon.com/blogs/database/analyzing-amazon-elasticsearch-service-slow-logs-using-amazon-cloudwatch-logs-streaming-and-kibana/

aws

Amazon ES の監視項目

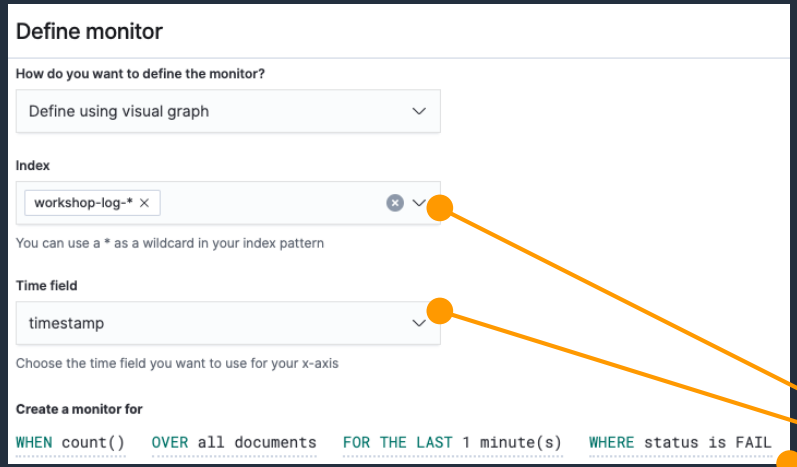
公式ドキュメントには，アラームを設定するべき CloudWatch メトリクスがまとめられている．代表的な項目は以下の通り

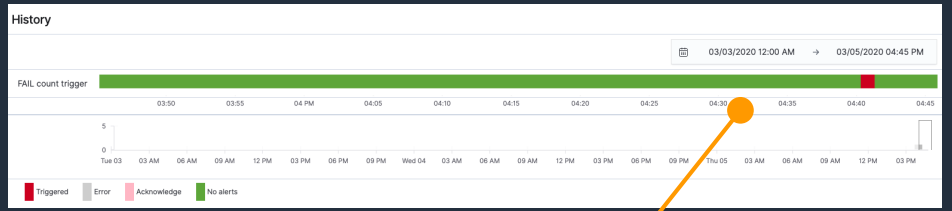
メトリクス	値	基準	説明
CPUUtilization	>= 80%	15 分間隔 3 回以上連続	データノードの CPU リソースが足りない恐れあり．インスタンスのスケールアップ or スケールアウトを検討
JVMMemoryPressure	>= 80%	5 分間隔 3 回以上連続	データノードのメモリエラーの恐れあり．インスタンスのスケールアップ or スケールアウトを検討
ClusterStatus.yellow	>= 1	1 分間隔 1 度生じたら	1 つ以上のレプリカシャードがノードに割りあってられていない．クラスタ状態を確認
FreeStorageSpace	-	1 分間隔 1 度生じたら	ディスク容量不足の恐れあり．XXX には，各ノードのストレージ容量の 25% の値を MB 単位で記入
MasterCPUUtilization	>= 50%	15 分間隔 3 回以上連続	専用マスターノードのリソースが足りない恐れあり．インスタンスのスケールアップを検討
MasterJVMMemoryPressure	>= 80%	15 分間隔 1 度生じたら	専用マスターノードのリソースが足りない恐れあり．インスタンスのスケールアップを検討

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-managedomains.html
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/cloudwatch-alarms.html

検索 API ベースのアラート

- Elasticsearch のクエリを指定して，一定の閾値を超えた場合に，重要度を指定してアラートを上げることが可能
- Slack, Amazon SNS, webhook URL を指定して通知を飛ばす
- アラートの実行履歴を Kibana 上で確認可能





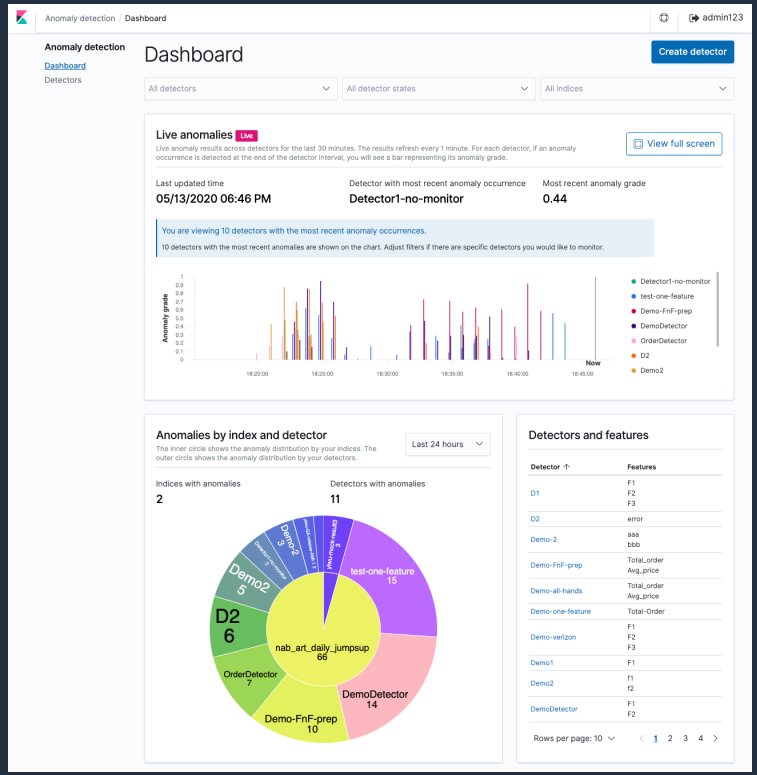
過去のアラート履歴を可視化

インデックスパターンやフィールド，クエリとその閾値を指定

https://aws.amazon.com/jp/blogs/news/setting-alerts-in-amazon-elasticsearch-service/
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/alerting.html

Anomaly Detection をトリガーにしたアラートの実行

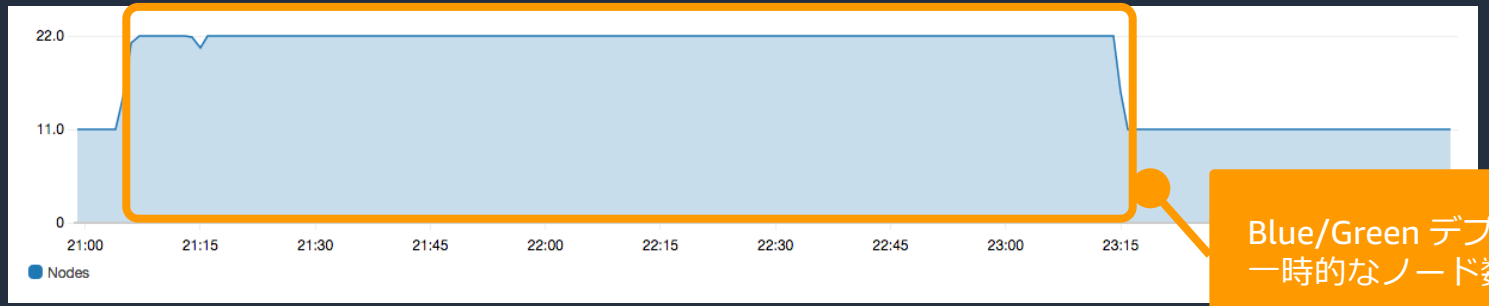
- Random Cut Forest アルゴリズムを用いた，時系列の異常検知機能
- Kibana 上でインデックスやフィールド，メトリクスを指定して複数の Detector を作成
- フィールドの average() や max() 以外にも，クエリ構文で任意のメトリクスを作成することが可能
- アラート機能と連携して，異常な値が出たら SNS 経由でアラートを飛ばせる



https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/ad.html
https://opendistro.github.io/for-elasticsearch-docs/docs/ad/
https://www.semanticscholar.org/paper/Robust-Random-Cut-Forest-Based-Anomaly-Detection-on-Guha-Mishra/ecb365ef9b67cd5540cc4c53035a6a7bd88678f9

Amazon ES の設定変更

- Amazon ES では，ドメインの更新時に Blue/Green デプロイプロセスが行われる．具体的にはインスタンスタイプの変更，マルチ AZ の有効化/無効化，バージョンアップグレード等．アクセスポリシーを変更するだけの場合は，通常 Blue/Green デプロイは発生しない
- Blue/Green デプロイ時には，一時的にクラスターに倍のノードが追加され，シャードの移動も発生するため，専用マスターノードおよびデータノードに大きな負荷がかかる
- そのため大きな設定変更は，負荷の低い深夜帯等に実施するのを推奨



https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-managedomains.html#es-managedomains-multiaz

Amazon ES のスナップショット

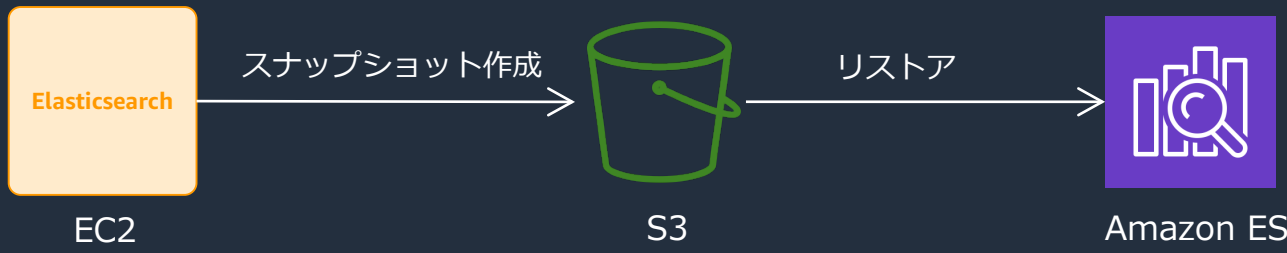
- スナップショットは，クラスターのバックアップ
- Amazon ES には，以下の 2 種類のスナップショットが存在する
- 基本的には自動スナップショットで足りるが，異なる Amazon ES ドメインにデータ移行したい場合は，手動での取得が必要

種類	用途	説明
自動スナップショット	• バックアップ	• バージョン 5.3 以降の場合，1h ごとにスナップショットを取得し，14 日間保持（ES 5.1 以前は 1 日ごと） • 追加課金なし
手動スナップショット	• バックアップ • データ移行	• Elasticsearch 自体の API を叩いて，手動で S3 に対してスナップショットを作成 • S3 利用料金がかかる

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-managedomains-snapshots.html

Elasticsearch on EC2 を Amazon ES に移行

1. Elasticsearch on EC2 で _snapshot/repos API を用いて，S3 にスナップショットリポジトリを登録し，スナップショットを作成
2. Amazon ES ドメインを新しく作成し，同じ S3 バケットをスナップショットリポジトリとして登録
3. Amazon ES ドメイン側から，_restore API を用いてスナップショットを復元



https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-managedomains-snapshots.html
https://aws.amazon.com/jp/premiumsupport/knowledge-center/migrate-amazon-es-domain/

アジェンダ

1. Elasticsearch の概要
2. Amazon Elasticsearch Service (Amazon ES) の概要
3. Amazon ES によるログ分析
4. Amazon ES による検索
5. Amazon ES の運用管理
6. Amazon ES のセキュリティ
7. 料金や制限事項
8. まとめ

Amazon ES における認証と認可

2 つのレイヤー

1. IAM ベースの Amazon ESドメインに対するアクセスポリシー
2. Open Distro ベースの，ドメイン内サブリソースに対する詳細なアクセス権限管理

1.1 アイデンティティベースのアクセスポリシー

IAM プリンシパルに対して付与
リソースはドメイン，インデックス， Elasticsearch API の単位で制御可能

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "es:Describe*",
        "es:List*"
      ],
      "Effect": "Allow",
      "Resource": "*"
    }
  ]
}
```

```
"arn:aws:es:region:aws-account-id:domain/domain-name",
"arn:aws:es:region:aws-account-id:domain/domain-name/*",
"arn:aws:es:region:aws-account-id:domain/domain-name/test-index",
"arn:aws:es:region:aws-account-id:domain/domain-name/test-index/_search"
```

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-ac.html
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-configuration-api.html

1.2 Amazon ES ドメインへのリソースベースアクセスポリシー

- ドメインごとに個別で設定．アイデンティティベースのポリシーと併せて制御
- 競合するポリシーに対しては，常に Deny が勝る．指定がない場合はデフォルト Deny

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Principal": {
      "AWS": [
        "arn:aws:iam::987654321098:user/test-user"
      ]
    },
    "Action": [
      "es:ESHttp*"
    ],
    "Condition": {
      "IpAddress": {
        "aws:SourceIp": [
          "192.0.2.0/24"
        ]
      }
    },
    "Resource": "arn:aws:es:us-west-1:987654321098:domain/test-domain/*"
  }]
}
```

リソースベース

アイデンティティベース

	Allow	Deny	指定なし
Allow	Allow	Deny	Allow
Deny	Deny	Deny	Deny
指定なし	Allow	Deny	Deny

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-ac.html
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-configuration-api.html

1.3 Amazon ES ドメインへの IP ベースアクセスポリシー

リソースベースポリシーのサブセット．パブリックドメインでのみ利用可能．
Amazon ES に署名なしでのリクエストを許可（VPC ドメインの場合は，セキュリティグループで制御）

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "*"
      },
      "Action": [
        "es:ESHttp*"
      ],
      "Condition": {
        "IpAddress": {
          "aws:SourceIp": [
            "192.0.2.0/24"
          ]
        }
      },
      "Resource": "arn:aws:es:us-west-1:987654321098:domain/test-domain/*"
    }
  ]
}
```

この指定の場合，指定 IP レンジからのリクエストは，すべて認証なしで実行可能
curl の実行やブラウザからの Kibana へのアクセスがスムーズに実行可能に

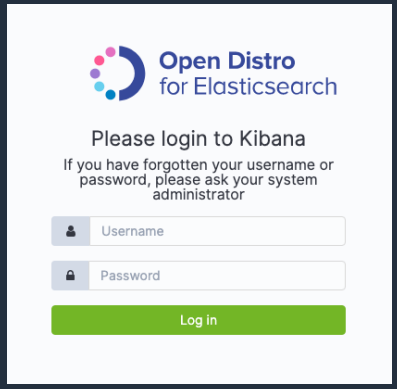
Kibana 内部での詳細な挙動については，次の詳細なアクセス権限管理で実現

この指定を行わない場合，HTTP リクエストを行う際には，必ず AWS の署名つきで送らないとエラーになる

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-ac.html
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-configuration-api.html
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-request-signing.html

2.1 Kibana による詳細なアクセス権限の管理

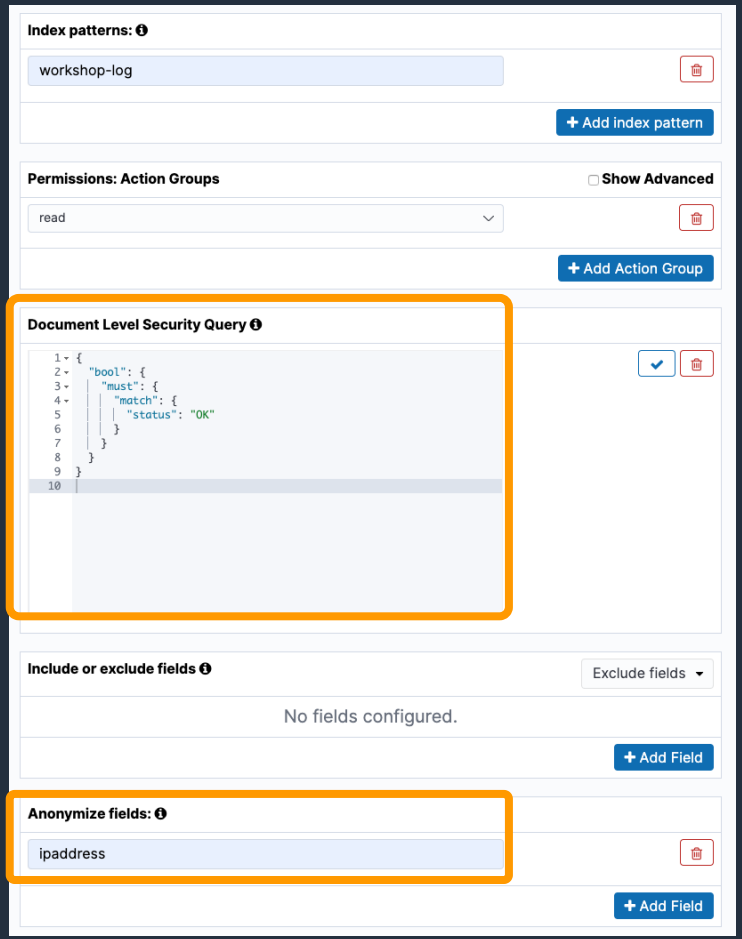
- ドメイン作成時に，詳細なアクセス権限管理の有効 / 無効を設定可能（ドメイン構築後に変更はできない）
- ユーザー管理は，以下の 2 つから選択
  - Amazon ES の内部ストア
  - Amazon Cognito 連携による IAM 制御
- ユーザーアカウントは，Kibana ログイン時に使用



https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/fgac.html

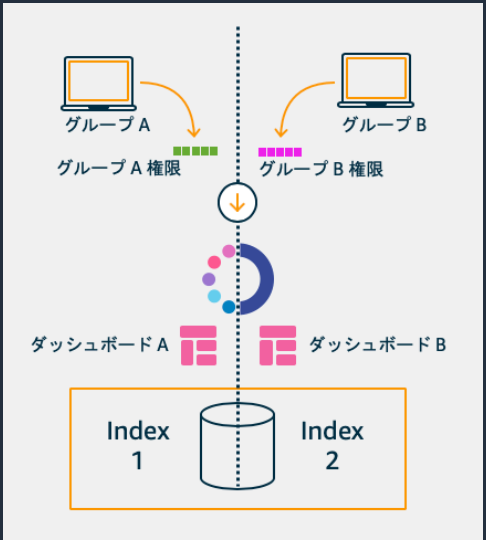
2.2 ドメイン内リソースに対する詳細な権限管理

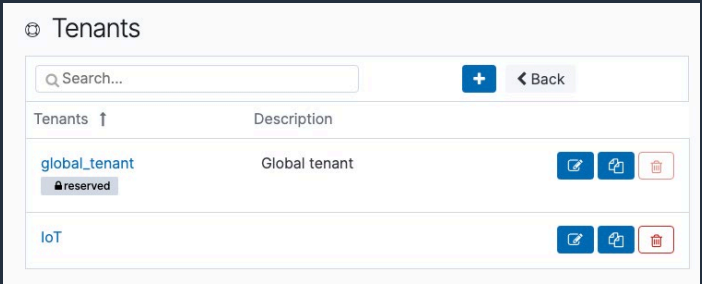
- 複数レベルでのアクセス権限管理が可能
  - インデックスレベル
  - ドキュメントレベル
  - フィールドレベル
- プリセットの権限セットも多数
  - cluster_monitor
  - kibana_all_read
  - manage_snapshots etc...
- フィールドのマスキングにも対応



2.3 Kibana のマルチテナンシー

Kibana の各コンポーネント（Index pattern, Visualize, Dashuboard etc...）を管理するための単位．部署ごとに Kibana テナントを分けることで，安全に複数部署のユーザーで Kibana を利用可能に





https://opendistro.github.io/for-elasticsearch-docs/docs/security-access-control/multi-tenancy/#add-tenants

暗号化

保管時のデータ暗号化

下記のデータに対して暗号化の有無を指定可能．有効化した場合，AWS KMS の暗号化キー用いた暗号化が行われる

- インデックス
- Elasticsearch ログ
- スワップファイル
- アプリケーションディレクトリのその他全てのデータ
- 自動スナップショット

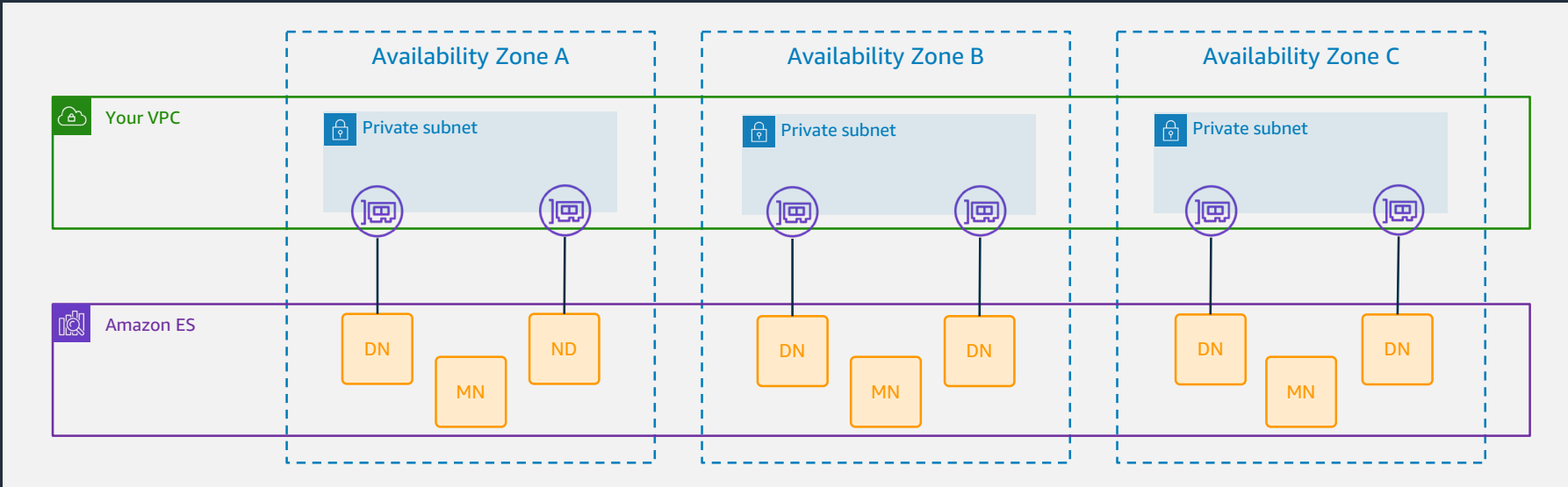
転送時のデータ暗号化

Amazon ES ドメインへのアクセス，およびノード間の通信について暗号化の有無を指定可能

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/encryption-at-rest.html
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/ntn.html

VPC から Amazon ES にプライベート接続

- Amazon ES ドメインへのアクセスを，VPC からクローズドな形で実現可能．VPC のセキュリティグループを利用したアクセス制御が可能．Kibana へのアクセスにはProxy や Client VPN が必要となる
- 各サブネットには，AZ に割り当てられたデータノード数の 3 倍の IP アドレスが必要
- VPC 接続が必要かどうかは，ユースケースに応じて判断



https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-vpc.html
https://aws.amazon.com/jp/blogs/news/set-access-control-for-amazon-elasticsearch-service/

アジェンダ

1. Elasticsearch の概要
2. Amazon Elasticsearch Service (Amazon ES) の概要
3. Amazon ES によるログ分析
4. Amazon ES による検索
5. Amazon ES の運用管理
6. Amazon ES のセキュリティ
7. 料金や制限事項
8. まとめ

Amazon ES の料金

インスタンス料金（オンデマンドおよび RI に対応）

- インスタンス利用料
- EBS 利用料（EBS Volumes インスタンス選択時のみ）

Ultrawarm 料金（オンデマンドのみ）

- インスタンス利用料
    - ultrawarm1.medium.elasticsearch: 0.279USD/時
    - ultrawarm1.large.elasticsearch: 3.144USD/時
- マネージドストレージ利用料: 0.026USD/GB/月

そのほかの料金

- データ転送料金（Amazon ES の in/out データに対する通常の転送料金．ノード間通信は無料）
- スナップショットの料金（自動スナップショットは無料，手動スナップショットは通常の S3 利用料が発生）

https://aws.amazon.com/elasticsearch-service/pricing/?nc1=h_ls
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/aes-ri.html

Amazon ES の制限

クラスタとインスタンスの制限

そのほか，インスタンスタイプ・サイズごとに EBS ストレージのサイズ，HTTP リクエストペイロードの最大サイズ制限あり

項目	制限	説明
クラスタあたりの最大データノード数	40	最大 200 まで上限緩和可能．ただし T2 インスタンスは 10
クラスタあたりの最大 Ultrawarm ノード数	45	
専用マスターノードの最大数	5	T2 インスタンスは，データノード数 10 以下の時のみ利用可
リージョンあたりの最大ドメイン数	100	

Ultrawarm ストレージの制限

インスタンスタイプ	最大ストレージ
ultrawarm1.medium.elasticsearch	1.5TiB
ultrawarm1.large.elasticsearch	20TiB

そのほかの制限

- Elasticsearch の Java プロセスの 最大ヒープサイズ: 32GiB
- ドメインのアクセスポリシーの最大サイズ: 100KiB

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/aes-limits.html

Elasticsearch の機能やプラグインにおけるバージョン要件

機能	Elasticsearch の最小バージョン
VPC サポート	すべてのドメインで含まれる
ドメインへのすべてのトラフィックに HTTPS を要求する	
マルチ AZ のサポート	
専用マスターノード	
カスタムパッケージ	
エラーおよびスローログの発行	5.1
Curator CLI のサポート	
保管時のデータの暗号化	
Kibana の Cognito 認証	
Elasticsearch のインプレースアップグレード	
時間単位の自動スナップショット	5.3
ノード間の暗号化	6.0
Java 高レベル REST クライアントのサポート	
アラート	6.2
SQL	6.5
クラスター間検索	6.7
きめ細かなアクセス制御	
UltraWarm	6.8
Index State Management	6.8
KNN	7.1
異常検出	7.4

プラグイン	Elasticsearch の最小バージョン
ICU Analysis	すべてのドメインで含まれる
Japanese (kuromoji) Analysis	
Phonetic Analysis	2.3
Seunjeon 韓国語分析	5.1
Smart Chinese Analysis	
Stempel Polish Analysis	
Ingest Attachment Processor	
Ingest User Agent Processor	
Mapper Murmur3	
Mapper Size	5.3
Ukrainian Analysis	
Open Distro for Elasticsearch Alerting	6.2
Open Distro for Elasticsearch SQL	6.5
Open Distro for Elasticsearch Security	6.7
Open Distro for Elasticsearch の Index State Management	6.8
Open Distro for Elasticsearch KNN	7.1
Open Distro for Elasticsearch の異常検出	7.4

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/aes-features-by-version.html
https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/aes-supported-plugins.html

aws

Elasticsearch バージョンのアップグレード

移行元 バージョン	移行先 バージョン	補足
7.x	7.x	
6.8	7.x	• 7.0 には多くの変更点が含まれる．インプレースアップグレードを開始する前に，6.8 ドメインのスナップショットを手動で作成し，それをテスト用の 7.*x* ドメインで復元してテストを行うことを推奨 • 7.0 ではインデックスの type として _doc のみを許容する形に変更されたので，6.8 のインデックスの形式がこれに沿っている必要がある．また Elasticsearch 7.x のデフォルトシャード数は 1 だが，Amazon ES のデフォルトは 5 のまま
6.x	6.x	
5.6	6.x	• バージョン 6 *x* で作成されたインデックスでは，複数のマッピングタイプがサポートされなくなった．ただしバージョン 5.*x* で作成されたインデックスは，6.*x* クラスターへの復元時に複数のマッピングタイプをサポート
5.x	5.6	

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/es-version-migration.html

aws

アジェンダ

1. Elasticsearch の概要
2. Amazon Elasticsearch Service (Amazon ES) の概要
3. Amazon ES によるログ分析
4. Amazon ES による検索
5. Amazon ES の運用管理
6. Amazon ES のセキュリティ
7. 料金や制限事項
8. まとめ

aws

まとめ

- Amazon Elasticsearch Service は，Elasticsearch ディストリビューションの Open Distro をベースとしたマネージドサービス
- Kinesis, Cloudwatch, IAM 等のさまざまなサービスと連携して，AWS 上でデータ分析や検索アプリケーションを非常に簡単に構築可能
- VPC 内のドメイン，保存・転送時の暗号化，IAM + Open Distro ベースの詳細な権限管理などによる，セキュアな形での Elasticsearch の運用が可能
- 分散システムであり運用コストの高い Elasticsearch を，マネージドサービスとしてデプロイ・管理できるように

aws

Amazon ES をこれから始めてみたいという方向けに

Amazon Web Services ブログ

Amazon Elasticsearch Service Intro Workshop を公開しました！- 基本的な使い方から最新アップデートまで 2 時間で体験

by AWS Japan Staff | on 08 MAY 2020 | in General | Permalink | Share

こんにちは、アナリティクスソリューションアーキテクトの志村です。本日公開した、Amazon Elasticsearch Service (Amazon ES) の初心者向けワークショップについてご紹介します。

Amazon ES は 2015 年にリリースされた、オープンソースの Elasticsearch を大規模かつ簡単でコスト効率の良い方法を使用してデプロイ、保護、実行する完全マネージド型サービスです。ストリームデータの分析を行いたい、全文検索エンジンを構築したい、といったときに手軽にご利用いただけます。ただ実際に Amazon ES を試そうとしたときによく当たるのが、ログ分析に適したストリームデータを生成するのが意外に面倒ということです。また権限管理周りをきちんと設定しないと、ただしく Kibana や Elasticsearch API にアクセスしたり，ログを挿入したりもできません．今回ご紹介する Amazon Elasticsearch Service Intro Workshop は、こうした問題を解消しながら、ハンズオン形式で Amazon ES をお試しいただけるような構成となっています。

ワークショップの構成

ワークショップで構築する仕組みは、以下の図のようになっています。Amazon Kinesis Data Generator (KDG)というデータ生成ツールを使って、Amazon Kinesis Data Firehose 経由でデータを Amazon ES に挿入します。挿入したデータを Kibana で分析・可視化します。KDG については、Amazon Kinesis Data Generatorを使用してストリーミングデータソリューションをテストするというブログ記事を参照ください。

https://github.com/aws-samples/amazon-elasticsearch-intro-workshop/tree/master/jp
https://aws.amazon.com/jp/blogs/news/amazon-elasticsearch-service-hands-on/

参考資料

公式ドキュメント

https://docs.aws.amazon.com/ja_jp/elasticsearch-service/latest/developerguide/what-is-amazon-elasticsearch-service.html

Open Distro ドキュメント

https://opendistro.github.io/for-elasticsearch-docs/

Amazon Elasticsearch Service Best Practice

https://www.slideshare.net/AmazonWebServicesJapan/20200414-amazon-elasticsearch-service-best-practice

小規模な Amazon Elasticsearch Service ドメインのコストを削減する

https://aws.amazon.com/jp/blogs/news/reducing-cost-for-small-amazon-elasticsearch-service-domains/

Amazon Elasticsearch Service Intro workshop

https://github.com/aws-samples/amazon-elasticsearch-intro-workshop/tree/master/jp

Q&A

お答えできなかったご質問については
AWS Japan Blog 「https://aws.amazon.com/jp/blogs/news/」にて
後日掲載します。

AWS の日本語資料の場所「AWS 資料」で検索

https://amzn.to/JPArchive

AWS Well-Architected 個別技術相談会

毎週”W-A個別技術相談会”を実施中

- AWSのソリューションアーキテクト(SA)に対策などを相談することも可能
- 申込みはイベント告知サイトから

(https://aws.amazon.com/jp/about-aws/events/)

AWS イベント で[検索]

ご視聴ありがとうございました

AWS 公式 Webinar
https://amzn.to/JPWebinar

過去資料
https://amzn.to/JPArchive