Amazon OpenSearch Service 機能解説 – 分析編

AWS Black Belt Online Seminar

Takayuki Enomoto

Solutions Architect, Analytics

2023/01

AWS Black Belt Online Seminarとは

- 「サービス別」「ソリューション別」「業種別」などのテーマに分け、アマゾン ウェブ サービス ジャパン合同会社が提供するオンラインセミナーシリーズです
- AWS の技術担当者が、AWSの各サービスやソリューションについてテーマごとに動画を公開します
- 動画を一時停止・スキップすることで、興味がある分野・項目だけの聴講も可能、スキマ時間の学習にもお役立ていただけます
- 以下の URL より、過去のセミナー含めた資料などをダウンロードできます
  - https://aws.amazon.com/jp/aws-jp-introduction/aws-jp-webinar-service-cut/

内容についての注意点

- 本資料では 2023 年 01 月時点のサービス内容および価格について説明しています。最新の情報は AWS 公式ウェブサイト (https://aws.amazon.com/) よりご確認ください
- 資料作成には十分注意しておりますが、資料に記載した情報と AWS 公式ウェブサイトの情報が異なる場合は、AWS 公式ウェブサイトの情報が優先されます
- 価格は税抜の表記です。日本居住のお客様には、別途消費税が請求されます

自己紹介

名前：榎本 貴之 (Enomoto, Takayuki)

所属：アマゾンウェブサービスジャパン
　アナリティクス事業本部
　ソリューションアーキテクト部
　アナリティクスソリューションアーキテクト

経歴：インフラエンジニア @システムインテグレーター
-> インフラエンジニア @ゲーム会社
-> Cloud Support Engineer @AWS
-> **Solution Architect @AWS**

好きなAWSサービス: **Amazon OpenSearch Service**,
　Amazon QuickSight, Amazon Neptune,
　Amazon Kinesis, AWS Config,
　Amazon CloudWatch, **AWS Support**

トピック

1. Amazon OpenSearch Service におけるデータ分析概要
2. データ収集
3. データの可視化・分析
4. データ処理
5. AWS ソリューションの活用
    1. ログ分析
    2. セキュリティ分析

Amazon OpenSearch Service におけるデータ分析概要

OpenSearch

オープンソースの分散型検索・分析スイート

OpenSearch Project によって開発され、Apache 2.0 ライセンスで提供されている

データストア、検索エンジンの **OpenSearch**、
可視化、UI ツールの **OpenSearch Dashboards** から構成されている

セキュリティ、パフォーマンス分析、機械学習など
様々なプラグインによる機能拡張が可能

Amazon OpenSearch Service

OpenSearch を簡単にデプロイ・管理、
スケール可能なフルマネージドサービス

フルマネージド: リソースのデプロイ、
管理に費やす時間を削減

セキュリティ: 認証、認可、暗号化、監査、
およびコンプライアンスのための高度な
セキュリティを維持

データ分析・オブザーバビリティ:
潜在的な脅威を体系的に検出し、機械学習、
アラート、可視化を活用して対処

コスト最適化: 各種リソースを最適化し、
戦略的な作業に注力

Amazon OpenSearch Service 概要

サービス概要については
“AWS Black Belt Online Seminar Amazon OpenSearch Service” を参照のこと

https://pages.awscloud.com/rs/112-TZM-766/images/AWS-Black-Belt_2023_Amazon-OpenSearch-Service-Basic_0131_v1.pdf

OpenSearch がカバーする分析ユースケース

インフラストラクチャと AWS サービスの問題を特定、診断、修正。製品の遅延と安定性を改善

大量のストリーミング データから、セキュアかつ費用対効果の高い方法で洞察を得る

INDUSTRY USE CASES

アプリケーション監視: インフラストラクチャは機能しているか？
レイテンシとエラー率は？
アプリケーションの問題の原因は?

セキュリティ監視: 疑わしい認証アクティビティはないか？
この IP アドレスによってどのデータにアクセスされたか？
侵害の事実はあるか？

ビジネスインサイト: ユーザーが興味を持っているコンテンツ/製品は何か？
最も使用されている機能と使用されていない機能はどれか？
最もアクティブなユーザーとその理由は?

オブザーバビリティ: どのサービスで問題が発生しているか？
リクエスト処理の遅延はどこで発生しているか？

データ分析のワークフロー

- 要件によって順番は前後する
- データ変換、データ格納後に、
  更に追加の変換処理を行うケースもある

データ収集

要件別データ取り込みフロー

リアルタイムデータの取り込み要件がある場合

- エージェント、ストリーム処理を活用したニアリアルタイムなデータ取り込み
- “リアルタイム” ではなく具体的なデータ取り込みのレイテンシ(何秒？何分？)を確認すること

リアルタイムデータの取り込み要件が無い場合

- データ量が少なければ、バッチによる定期取り込み
- データ量が多く少しずつ取り込みたい場合は、エージェントやストリーム処理を活用

リアルタイムなデータ収集の重要性

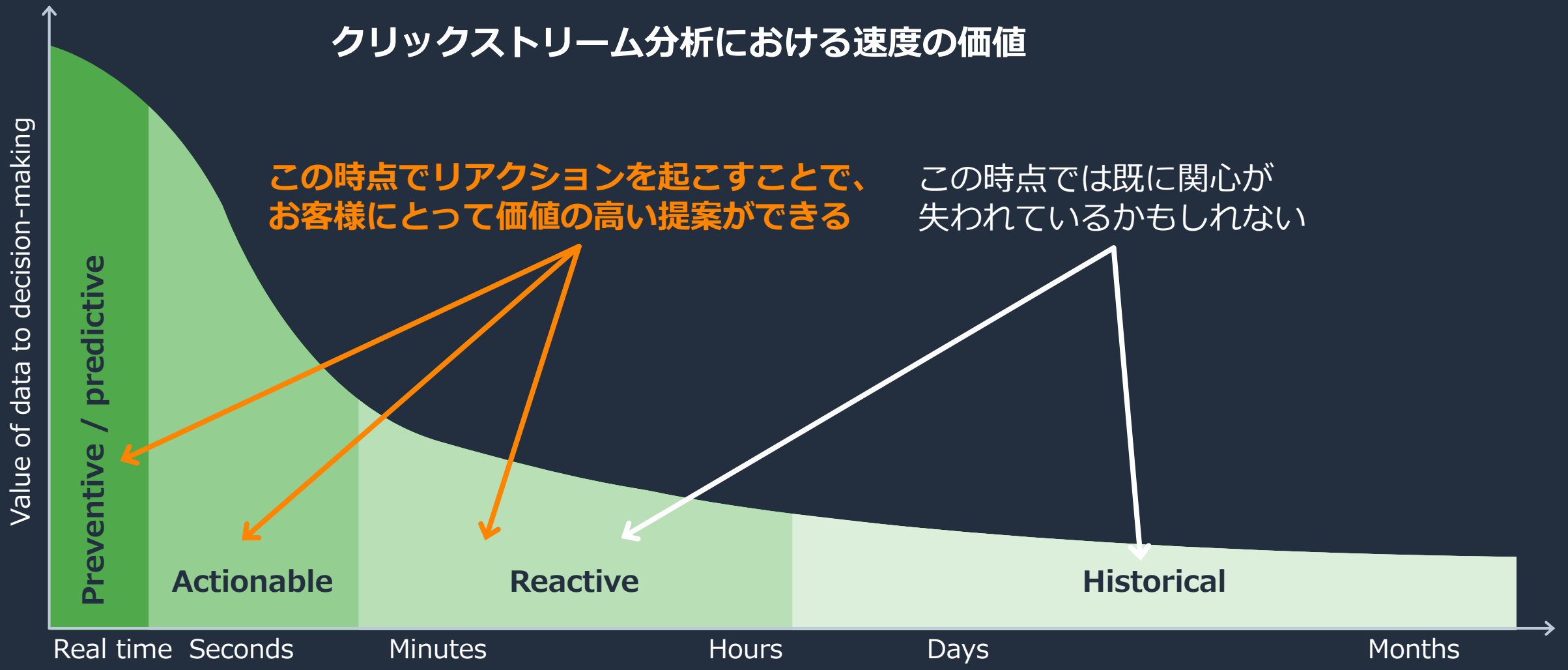
企業が価値を創造するためには, 高速かつ大量に生成される
様々なデータソースから洞察を導き出す必要がある

To create value companies must derive insights from a variety of data sources that are producing data at high velocity and volume

> データ統合の要件として, リアルタイムストリーミング,
> レプリケーション, 仮想化機能が求められている
>
> *Data integration requirements … now demand real-time streaming, replication and virtualized capabilities …*

—Gartner 2019 Planning Guide for Data and Analytics

新しいデータは意思決定における価値が高い



Source: Mike Gualtieri, Forrester, *Perishable Insights*

バッチ処理

一般的に**データストア**に保存されているデータを元に分析、加工を行う

ソース

アプリケーション内の処理に応じて, アプリケーションから随時, または定期的にデータが発生する

保存

アプリケーションサーバーから送信されたデータは, データベースなどのデータストアに保存される

データ処理, 分析

バッチ処理を定期的に実行する. バッチ処理では, データストア内に格納されたデータを使用し処理, 分析を行う

(option) 保存

バッチ処理の結果は必要に応じて再度データストアにロードされる

ストリーム処理

データストリームに**ストリーミングデータ**を一時的に格納し、
処理、 分析を経てからデータストアへ配信を行う

ソース

高速でリアルタイムデータを生成するアプリケーションやデバイス

収集, 集約, 送信

数万規模のデータソースからのデータをリアルタイムで収集し, 必要に応じて集約したうえでデータストリームへ送信する

取込

データストリーム内のレコードは受信した順序で保存され, 設定された期間内であれば無制限に再生(再取得, 再処理)できる

取得, 処理, 配信

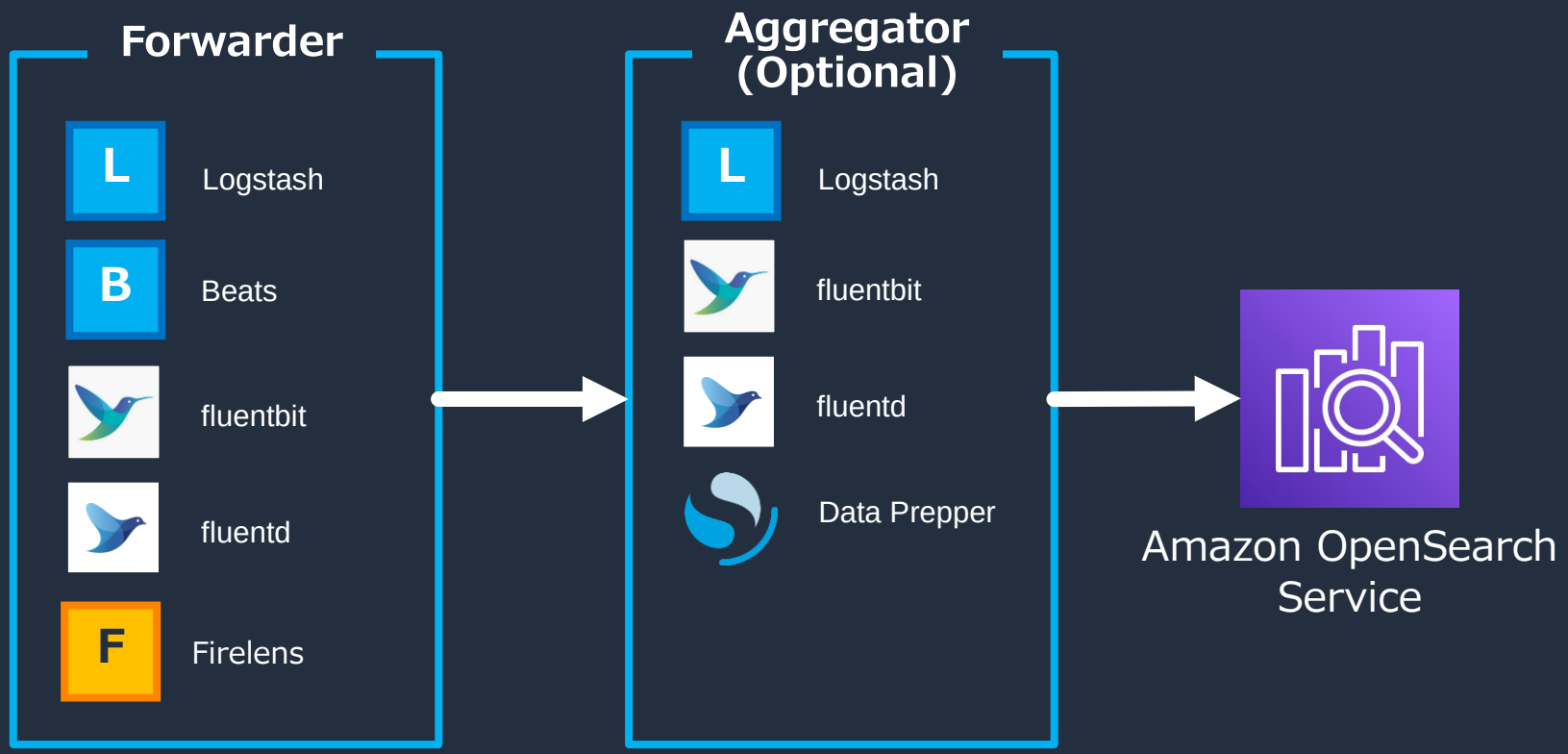
レコードは格納された順番で取り出され, リアルタイム分析やストリーミング ETL に活用される

(Option)保存

処理されたレコードは,必要に応じてデータレイク, データウェアハウス, データベース等の様々なデータストアへ配信される

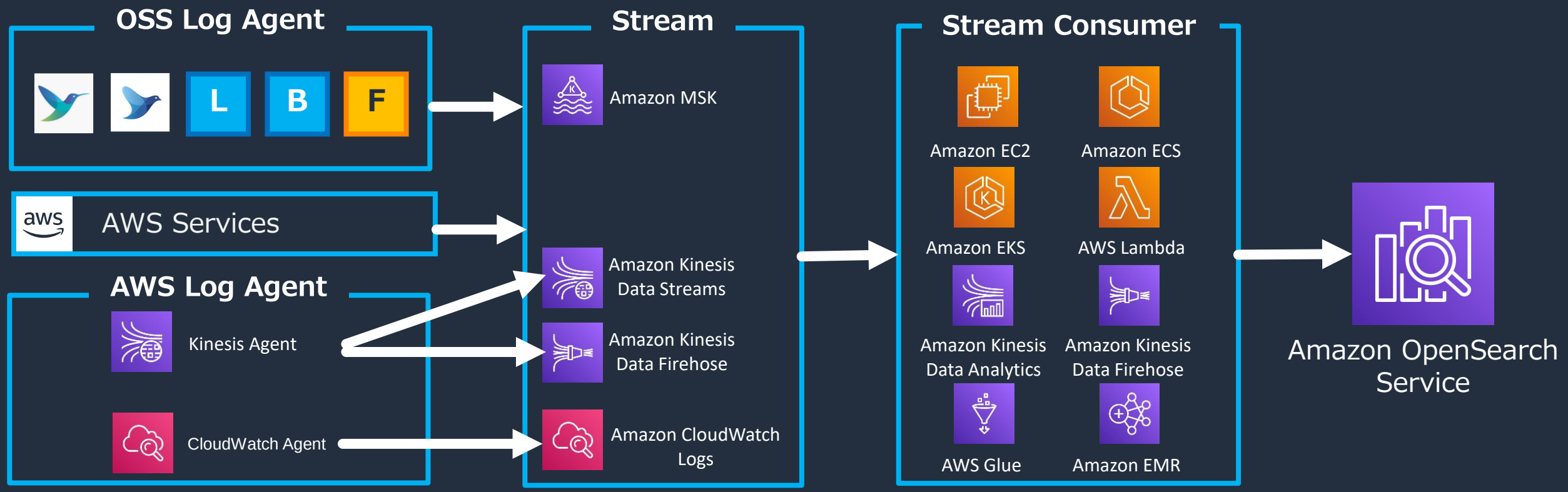
ログ収集エージェントを使用したログ取り込み

- アプリケーション実行環境や OS 上に出力されるログをリアルタイムに OpenSearch に転送
- サーバーにはエージェントソフトウェアをインストールして利用、コンテナ環境では
  ログエージェントをサイドカーコンテナで実行するなどの方法で動作させる
- エージェントからの直接転送も、必要に応じて中継サーバーで集約してからの転送も可能

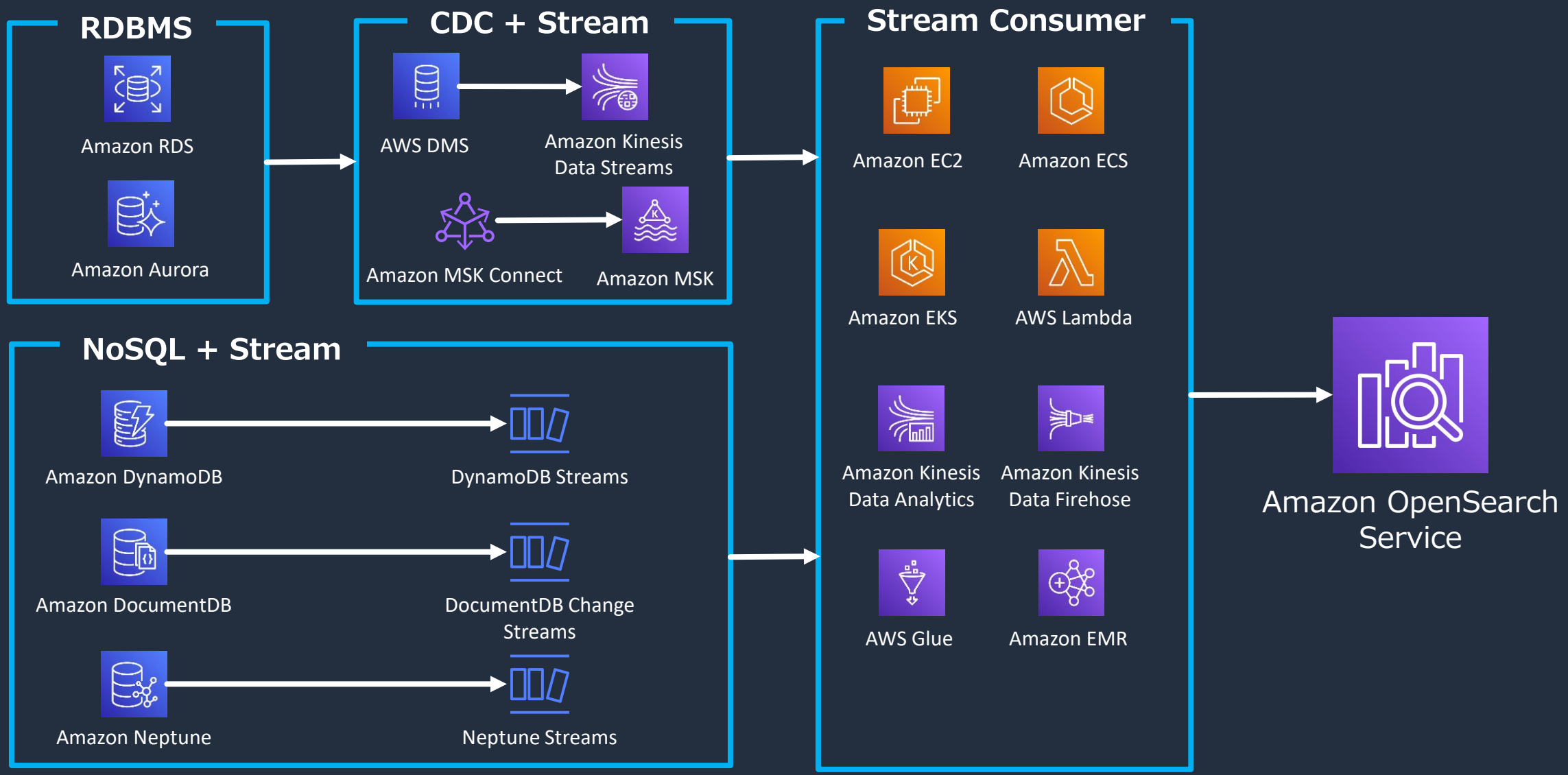


ストリームサービスを活用したログ取り込み

- アグリゲーターの代わりにマネージドなストリームサービスを活用することも可能。運用負荷の削減、ストリーム側でログのバーストを吸収可能といったメリットがある
- 多くの AWS サービスがストリームサービスと統合されており、ログ転送のパイプラインを効率よく実装できる

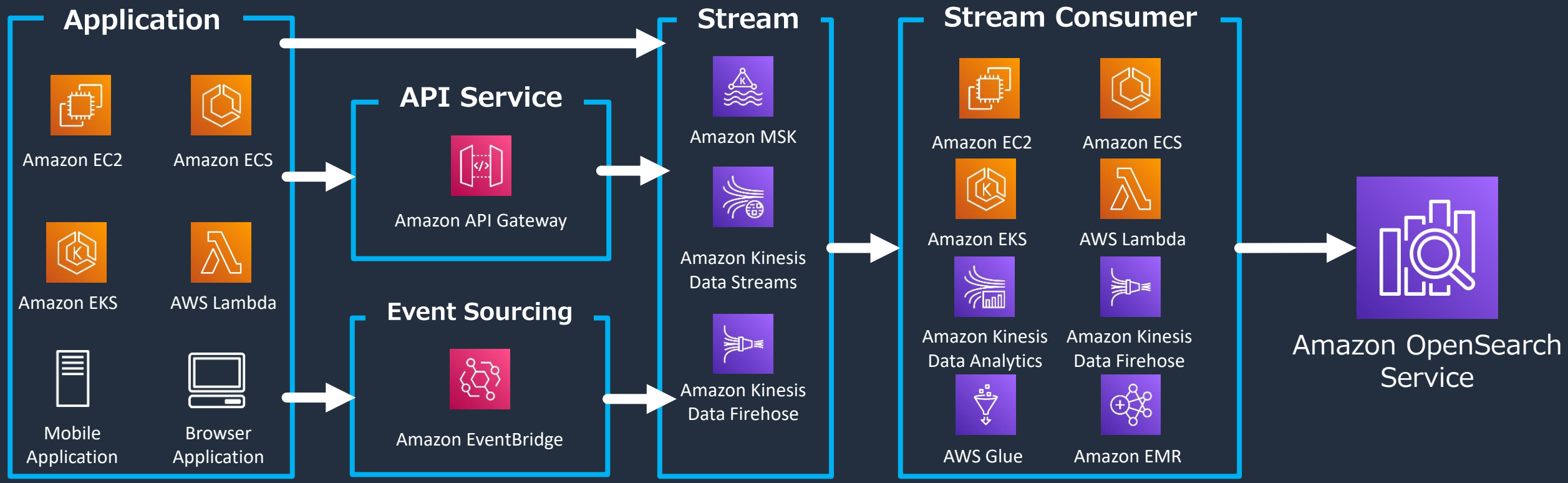


データベース変更情報のニアリアルタイム取り込み



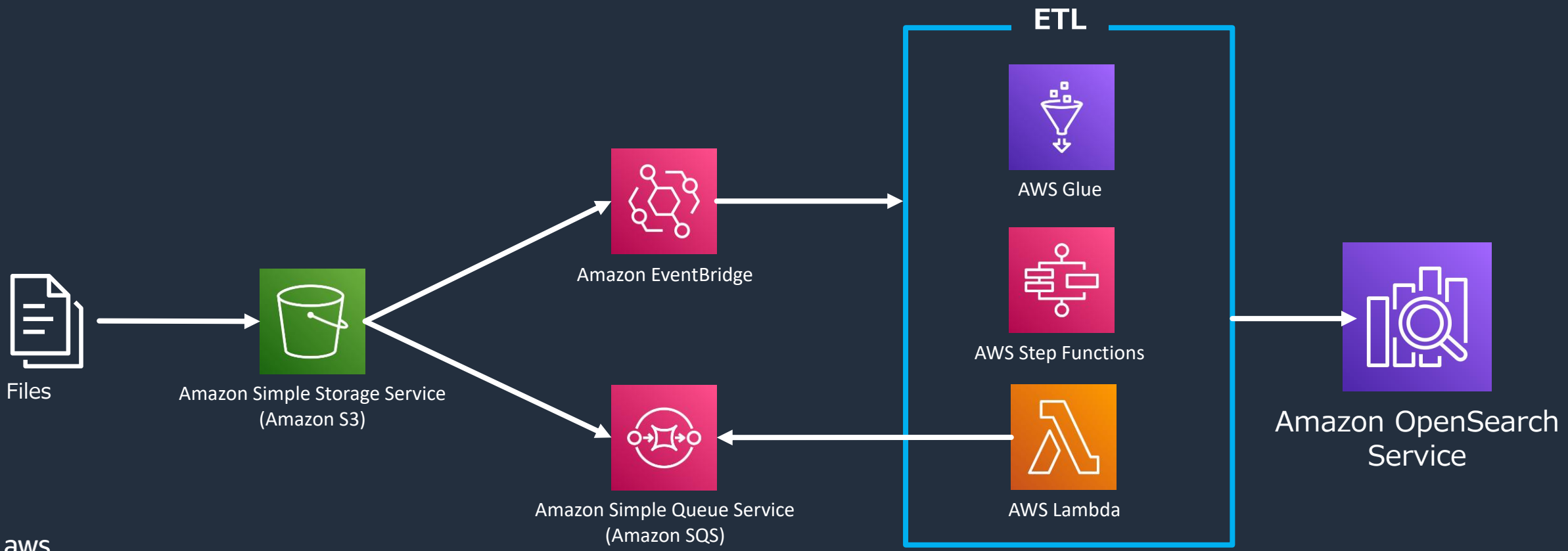
アプリケーションからのリアルタイムデータ取り込み

- アプリケーションから直接ストリームサービスにイベントデータを連携、取り込むことも可能
- API サービスを間に挟んで認証を行ったり、既存のイベントソーシングアーキテクチャーにストリームを追加してイベント分析を行うことも可能



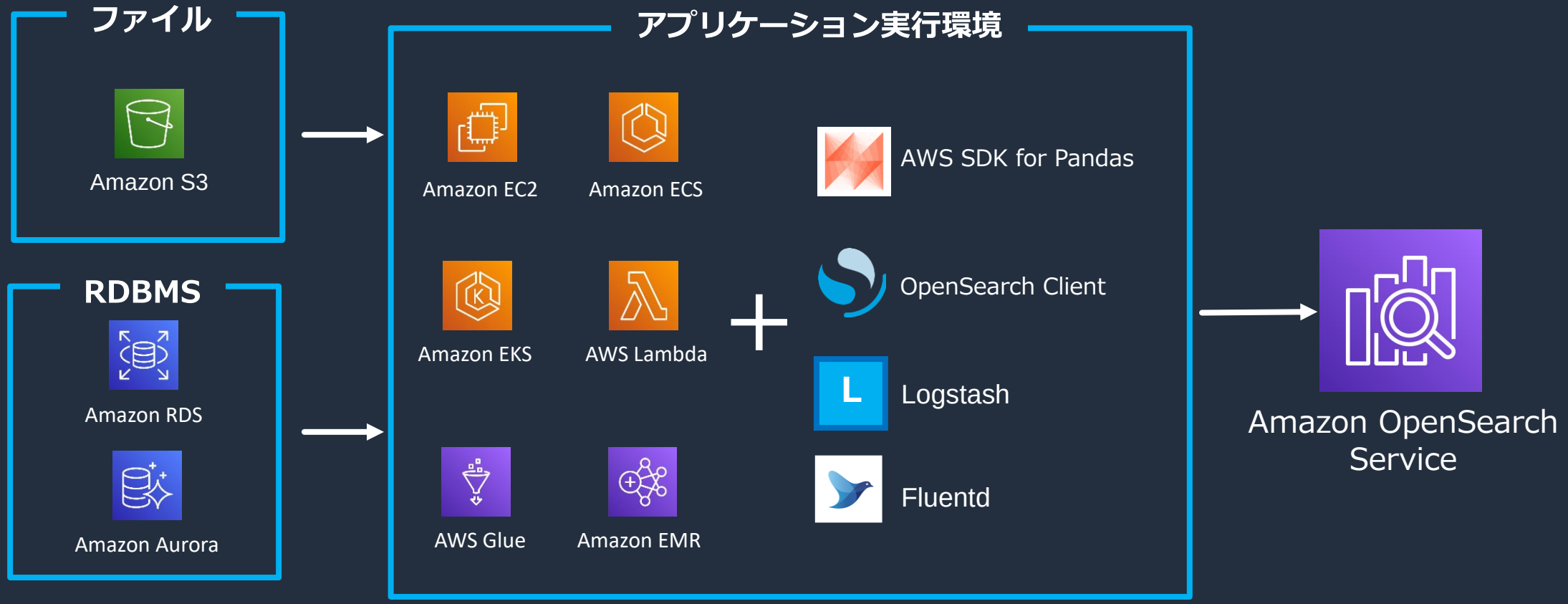
非同期処理によるファイルデータ取り込み

- S3 イベント通知を使用することで、アップロードされたログファイルを非同期的に処理し OpenSearch へ転送
- バッチで取り込む場合と比較して 1 回あたりの処理時間は短いが、一度に多数のオブジェクトがアップロードされると処理効率が落ちる場合がある



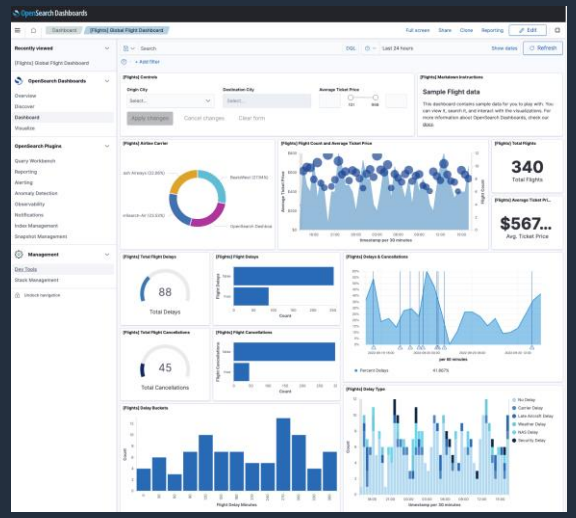
バッチによるデータの取り込み

リアルタイム要件が無い場合や、ソースデータを元にインデックスを全更新するようなケースでは、バッチによる取り込みが選択肢に入ってくる



データの可視化・分析

OpenSearch をサポートするダッシュボードツール



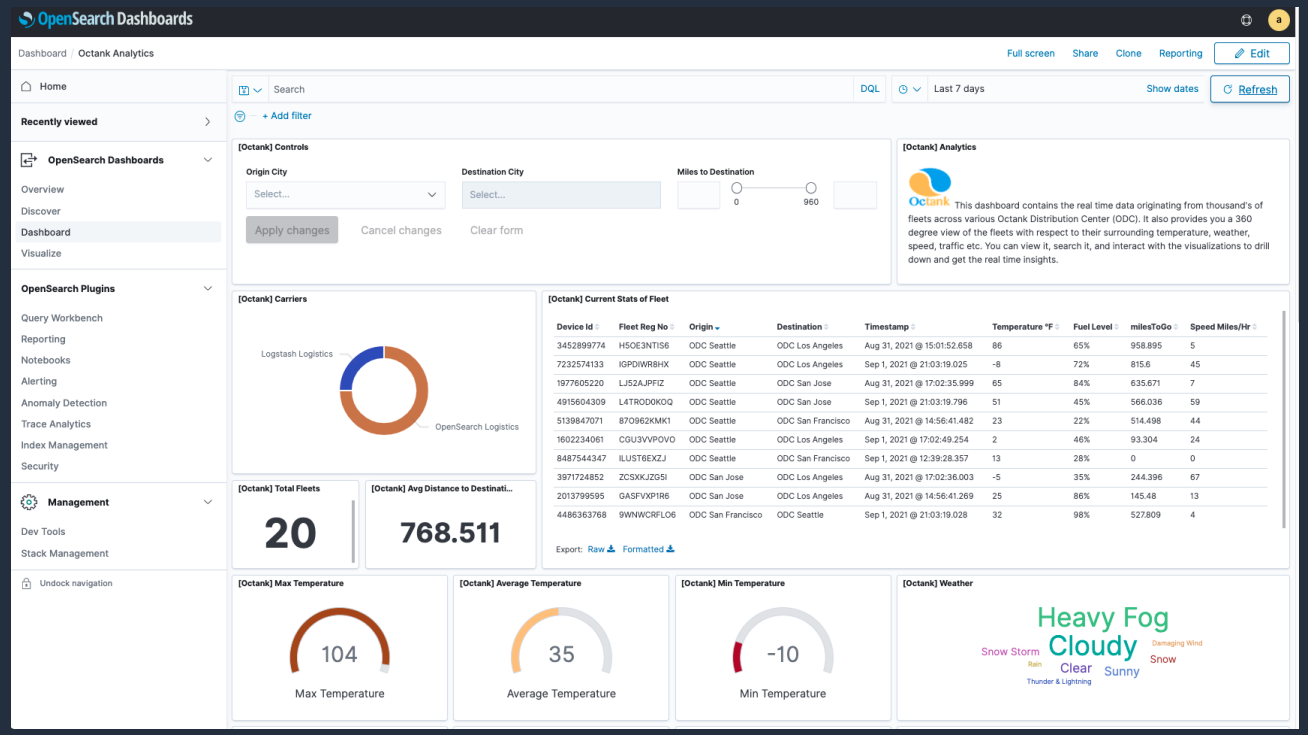
OpenSearch Dashboards

Amazon Managed Service for Grafana

Amazon QuickSight

OpenSearch Dashboards

- OpenSearch にバンドルされている可視化ツール
- シンプルなログの検索、可視化に加えて、アプリケーショントレース分析などの、高度な分析機能も提供
- アラートの設定など、OpenSearch の設定管理機能も付属している
- OpenSearch に格納されたデータを分析する際の最初の選択肢



OpenSearch Dashboards > Index Patterns

- 分析対象の Index グループ。1 つ以上の Index を登録可能
- OpenSearch Dashboards では、Index Patterns に対してログ検索や可視化を行う
- 複数 Index を含める場合は、**logs-*** などワイルドカードを指定する
- Index Patterns は OpenSearch Dashboards 上でのみ利用可能。アプリケーションからの API リクエストなどでは指定不可

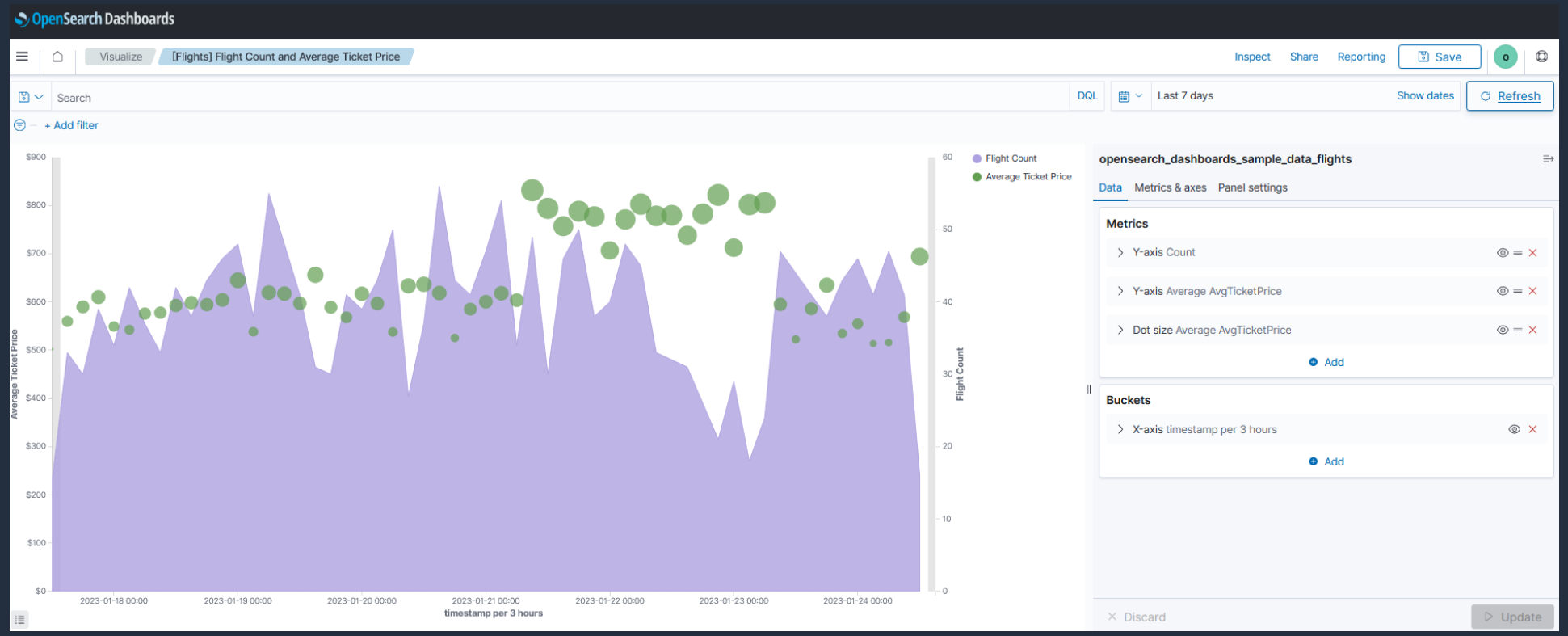
OpenSearch Dashboards > Discovery

- Index Patterns に対する検索を行う機能
- OpenSearch の全文検索機能によるドキュメントの部分一致検索、フィルタなどをサポート。障害調査、セキュリティインシデント調査、問い合わせ検索など様々なケースで利用可能

https://opensearch.org/docs/latest/dashboards/discover/index/

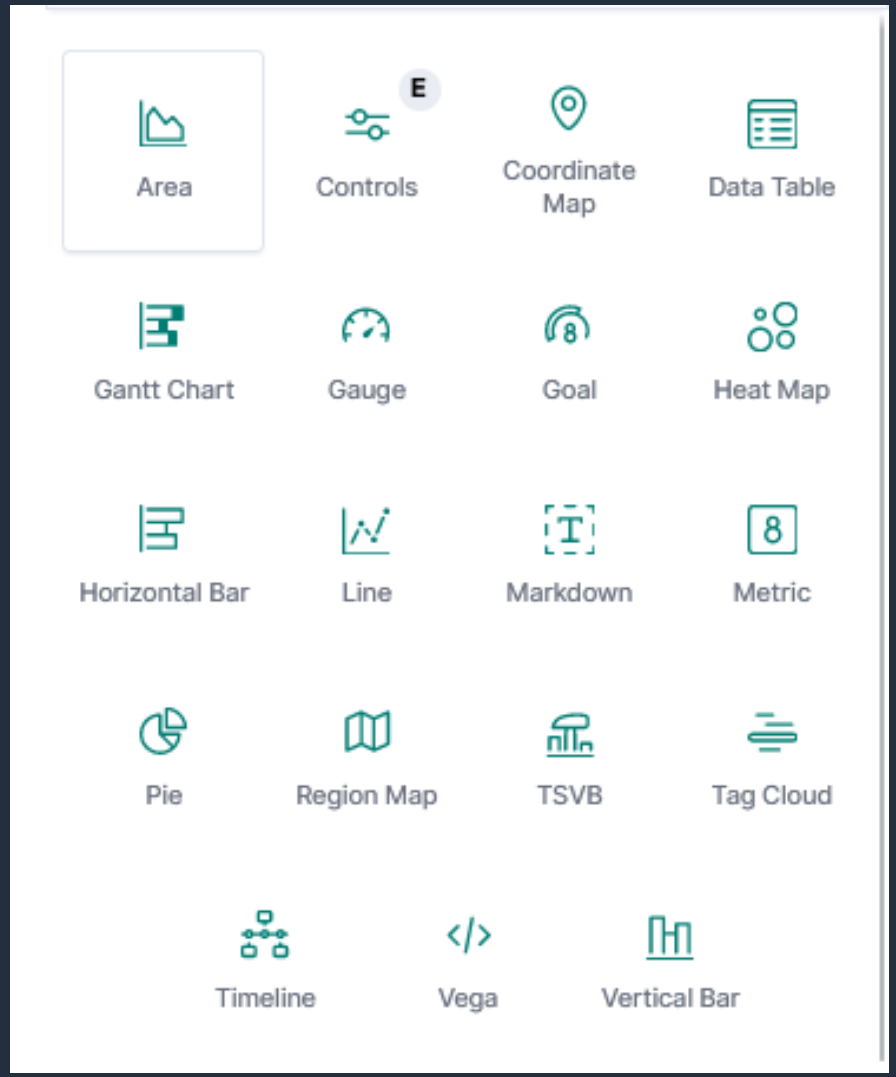
OpenSearch Dashboards > Visualize

- Index Patterns のデータを可視化する機能。
- 作成したグラフは **Visualizations** として保存し、ダッシュボードに張り付け可能



https://opensearch.org/docs/latest/dashboards/visualize/viz-index/

サポートされるビジュアルタイプ



基本的なビジュアルタイプ

- Line Chart(折れ線グラフ)
- Horizontal Bar & Vertical Bar (横棒グラフ & 縦棒グラフ)
- Area (面グラフ)
- Pie (円グラフ)
- Text (マークダウンテキスト)
- Data Table (表)
- Metric (カウント)
- Guage (ゲージチャート)
- Goal (KPI ゲージチャート)
- Heat Map (ヒートマップ)
- Tag Cloud (タグクラウド)
- Gantt Chart (ガントチャート)

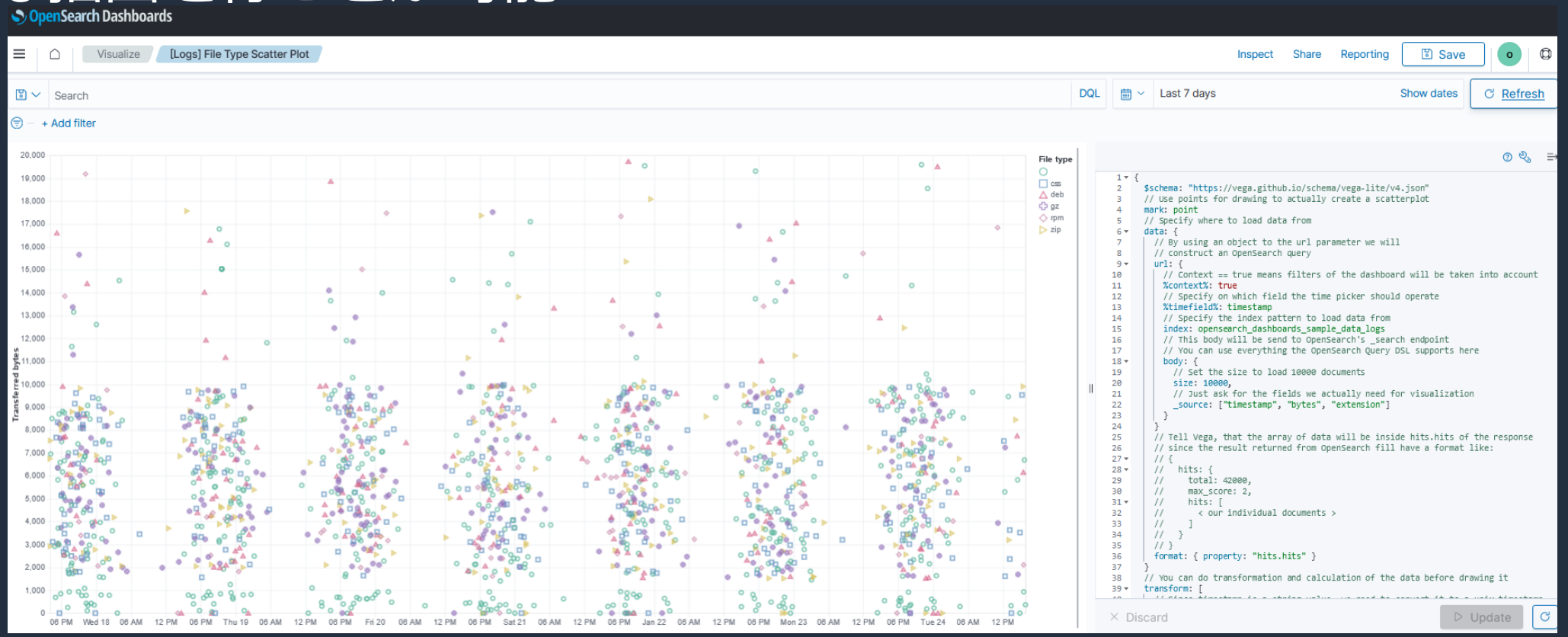
OpenSearch 固有のビジュアルタイプ

- Coordinate Map
- Region Map
- Vega(カスタムビジュアル)
- Timeline (時系列データ分析)
- TSVB (Time Series Visual Builder)
- Control (コントロール)

https://opensearch.org/docs/latest/dashboards/visualize/viz-index

Vega

- カスタムビジュアルを作成できる機能
- Vega-lite と呼ばれる JSON 形式の設定から、散布図などの複雑な描画を行ことが可能



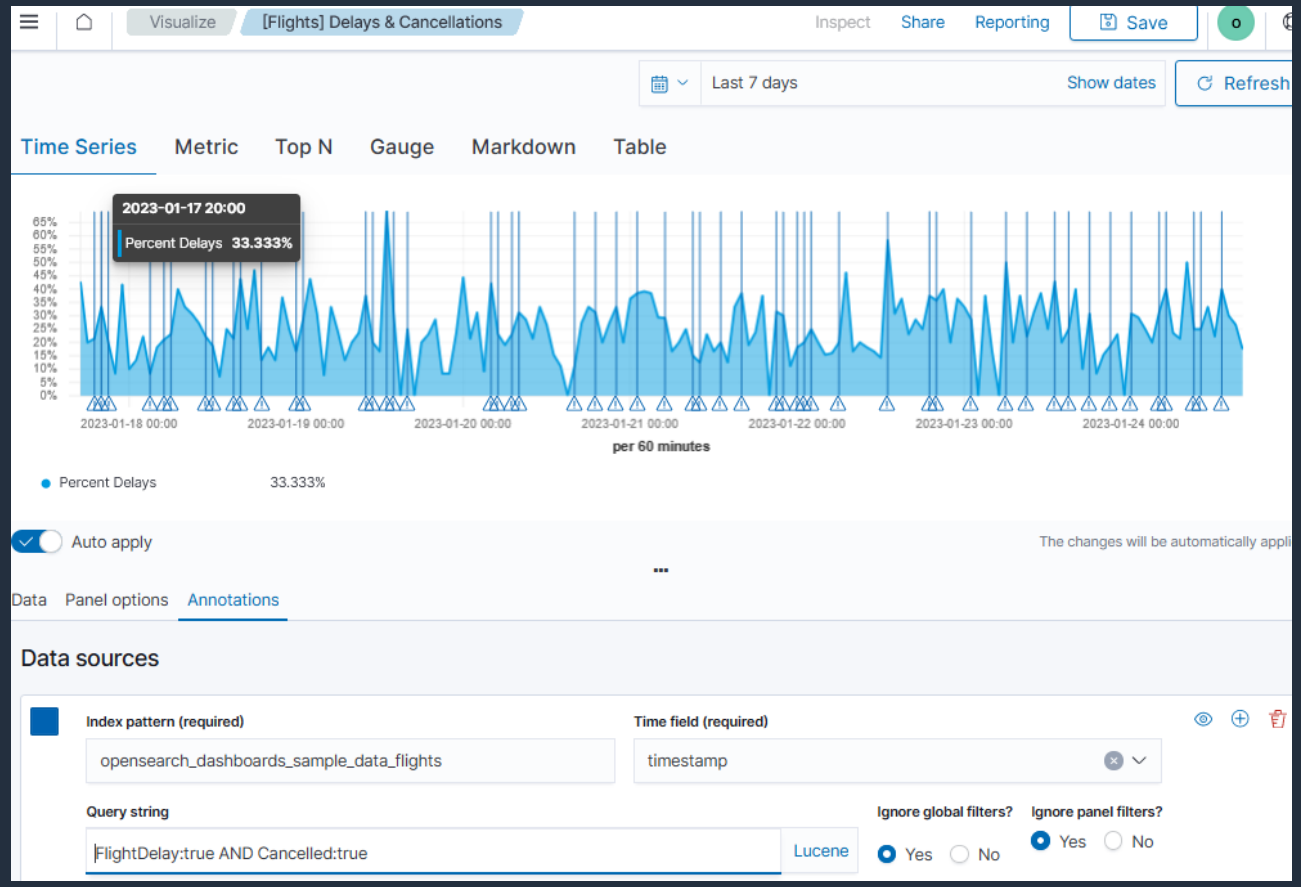
https://vega.github.io/vega-lite/examples/
https://opensearch.org/docs/latest/dashboards/visualize/viz-index#vega

TSVB (Time Series Visual Builder)

- 時系列データを可視化するための
  ビルダーツール

- 同一データから複数タイプの
  ビジュアル生成が可能

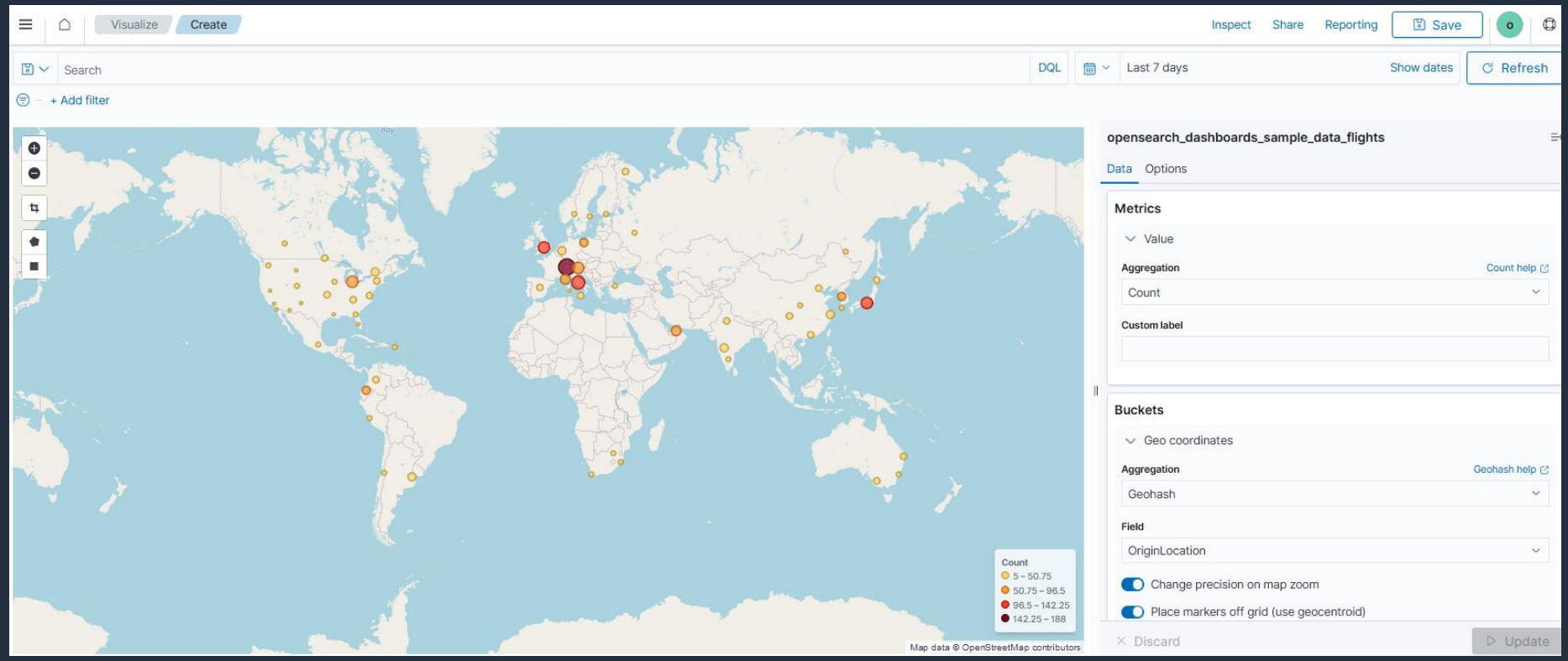
- アノテーションや閾値などを
  グラフに追加することが可能



https://opensearch.org/docs/latest/dashboards/visualize/viz-index#tsvb

Coordinating Map

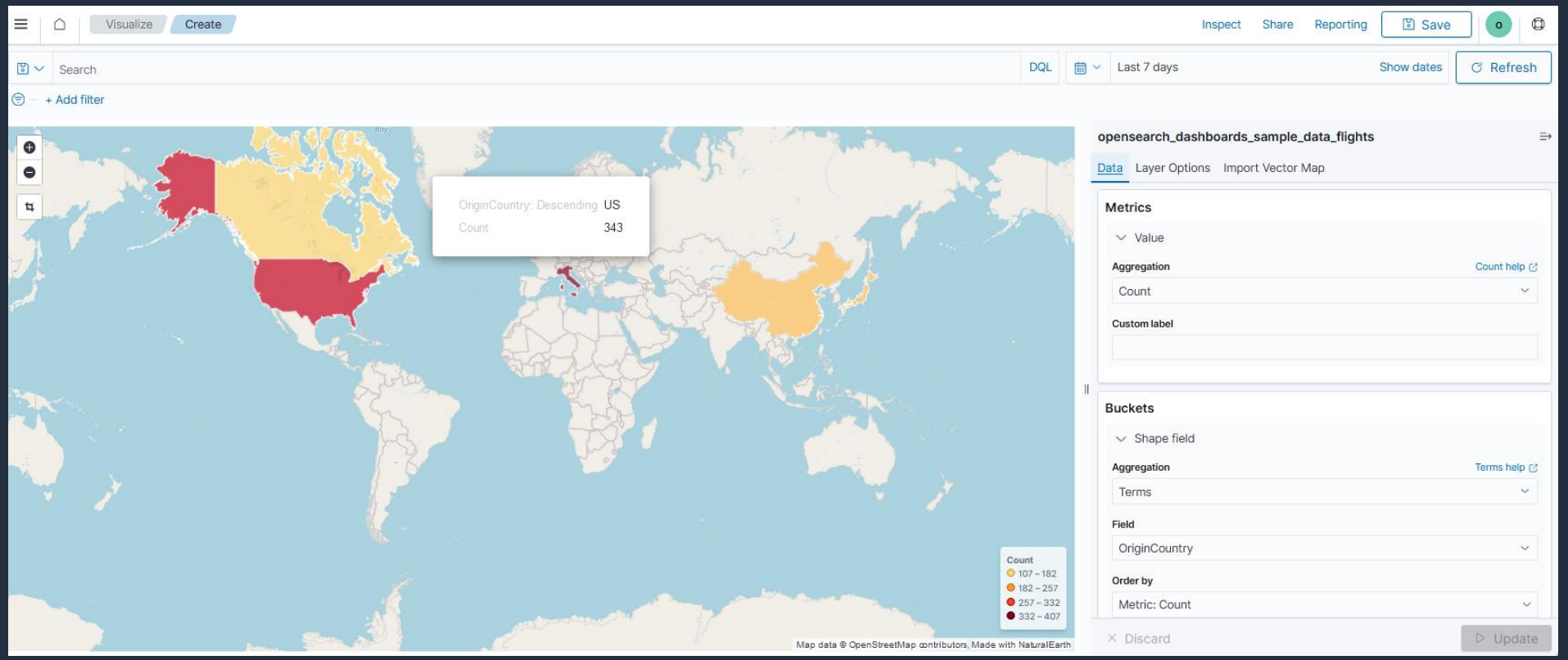
- データセットに含まれる緯度と経度の値を元にした集計結果を地図上にプロットする
- フライトデータから空港の発着数をカウントするな場合などに利用



https://opensearch.org/docs/latest/dashboards/visualize/viz-index#maps

Region Map

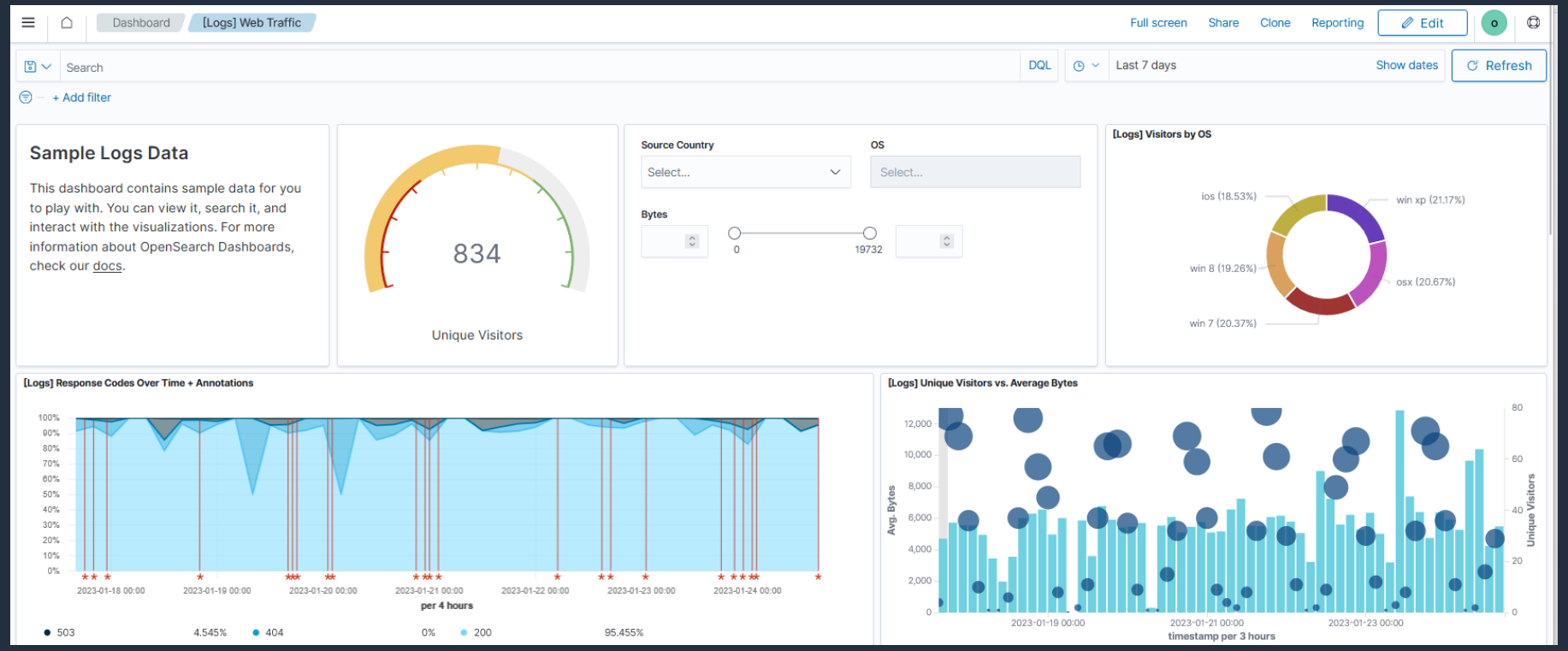
- 地域、国、州、都道府県などの地理的領域に関する統計データを可視化
- 国別の売上高などを可視化するようなケースで利用可能



https://opensearch.org/docs/latest/dashboards/visualize/viz-index#maps

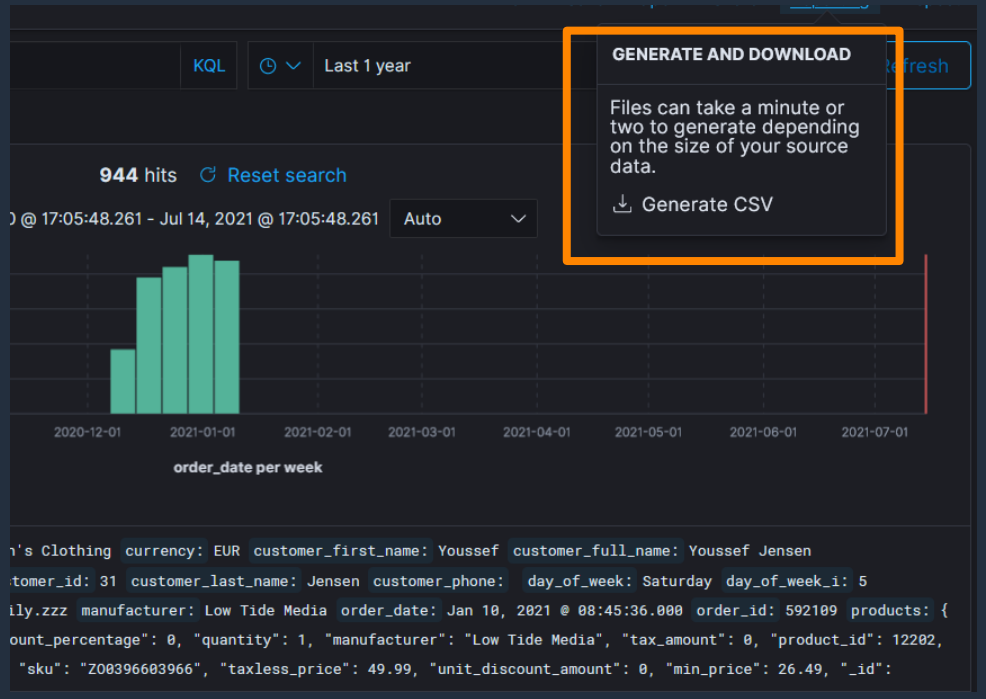
OpenSearch Dashboards > Dashboard

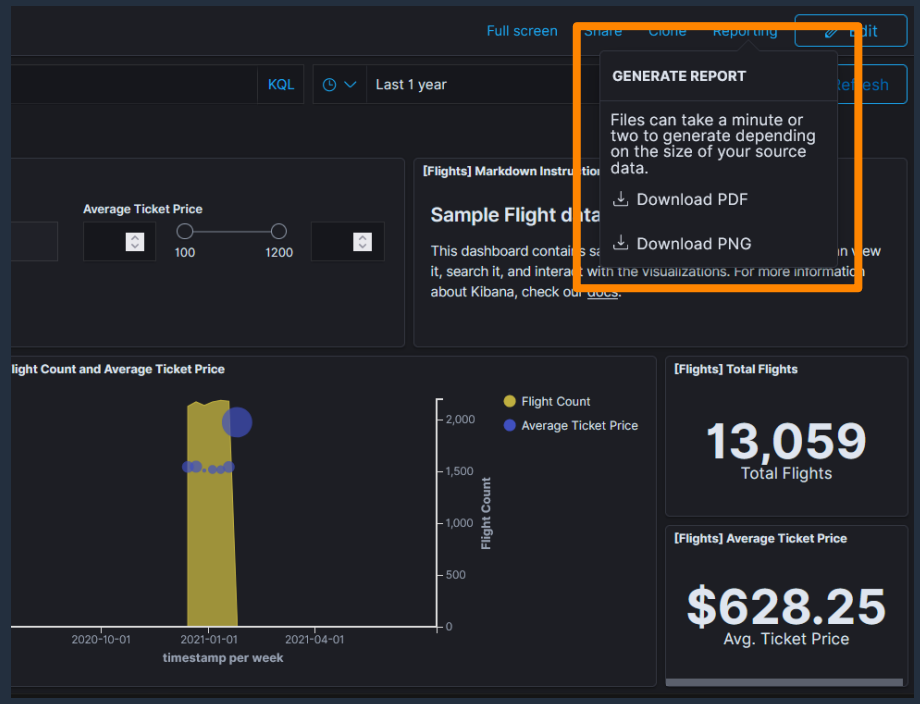
- 複数の Visualization を配置したダッシュボード
- Control Visualization を配置することで、利用者がリストやスライダーで簡単にデータを絞り込むことが可能。OpenSearch のクエリが記載できれば、検索バーでより高度な絞り込みも可能。



OpenSearch Dashboards > Reporting

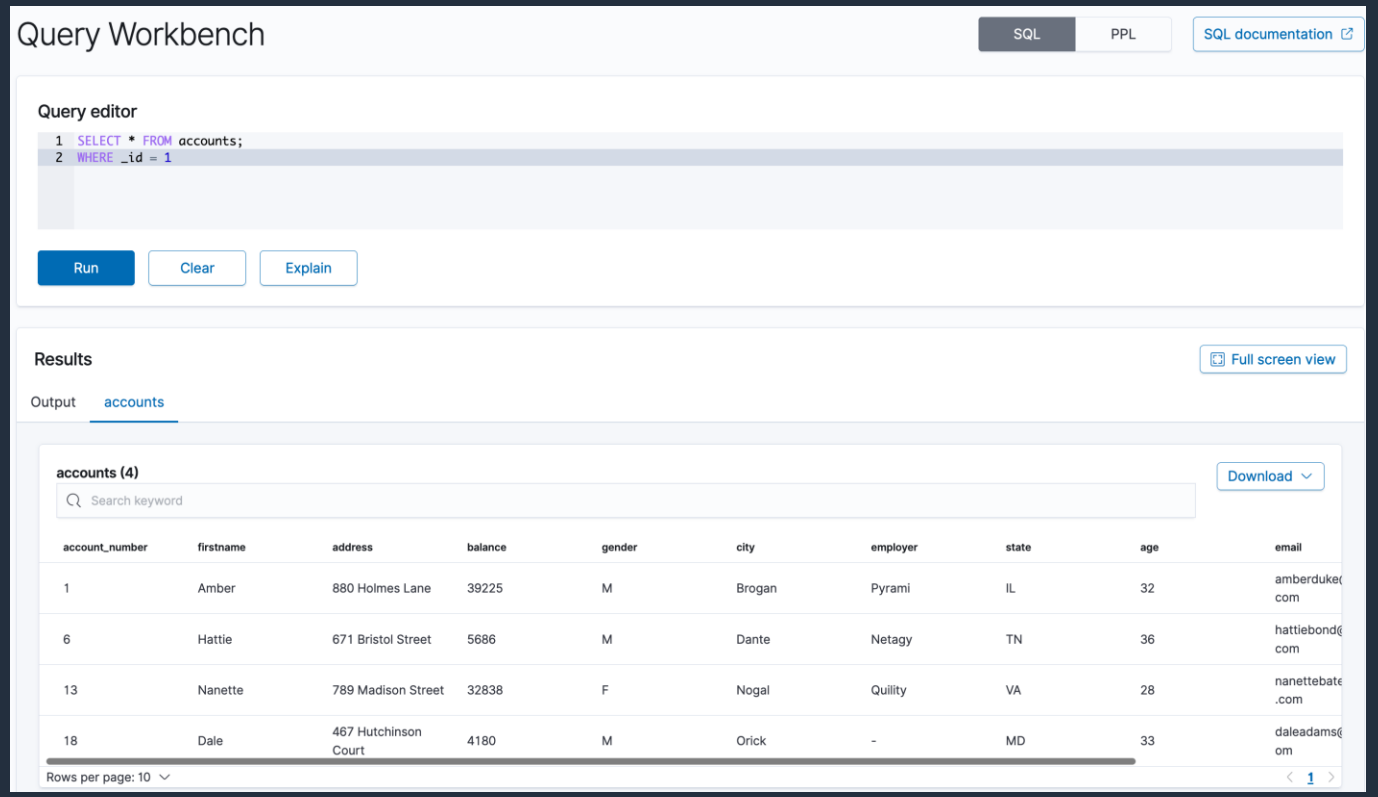
- ダッシュボードおよびドキュメントの検索結果をダウンロードする機能
- ダッシュボードは PDF または PNG で、Discover の検索結果は CSV で提供
- ダウンロード用のファイルを定期作成する機能も提供





OpenSearch Dashboards > Query Workbench

- SQL、Piped Processing Language (PPL) を使用し、データの検索や集計、分析を行うためのツール
- 実行結果はダウンロード可能



SQL

- SELECT 文のみをサポートしており、データの書き込みと削除はできない
- 独自に JOIN やサブクエリなど複雑な処理もサポートしている。
  - これらの処理は OpenSearch コアエンジン外、SQL モジュールで処理されるため、通常の検索と比べてパフォーマンスは低下する点に注意
- Query Workbench、API や CLI、JDBC Driver、ODBC Driver など複数のインターフェイスに対応

```
SELECT
  a.account_number, a.firstname, a.lastname,
  e.id, e.name
FROM accounts a
JOIN employees_nested e
 ON a.account_number = e.id
```

```
SELECT a1.firstname, a1.lastname, a1.balance
FROM accounts a1
WHERE a1.account_number IN (
  SELECT a2.account_number
  FROM accounts a2
  WHERE a2.balance > 10000
)
```

https://docs.aws.amazon.com/ja_jp/opensearch-service/latest/developerguide/sql-support.html

https://opensearch.org/docs/latest/search-plugins/sql/sql/index/

PPL (Piped Processing Language)

- PPL とは、パイプ | でコマンドを繋いで処理を記述する言語

- 検索だけでなく、フィールドの値をパースし複数のフィールドに分割するなど複雑な操作も可能

```
search source=accounts | eval doubleAge = age * 2 | fields age, doubleAge;
```

age	doubleAge
32	64
36	72
28	56
33	66

```
os> source=accounts | parse email '.+@(?<host>.+)' | fields email, host ;
fetched rows / total rows = 4/4
```

email	host
amberduke@pyrami.com	pyrami.com
hattiebond@netagy.com	netagy.com
null	null
daleadams@boink.com	boink.com

https://opensearch.org/docs/latest/search-plugins/sql/ppl/functions/
https://docs.aws.amazon.com/ja_jp/opensearch-service/latest/developerguide/ppl-support.html

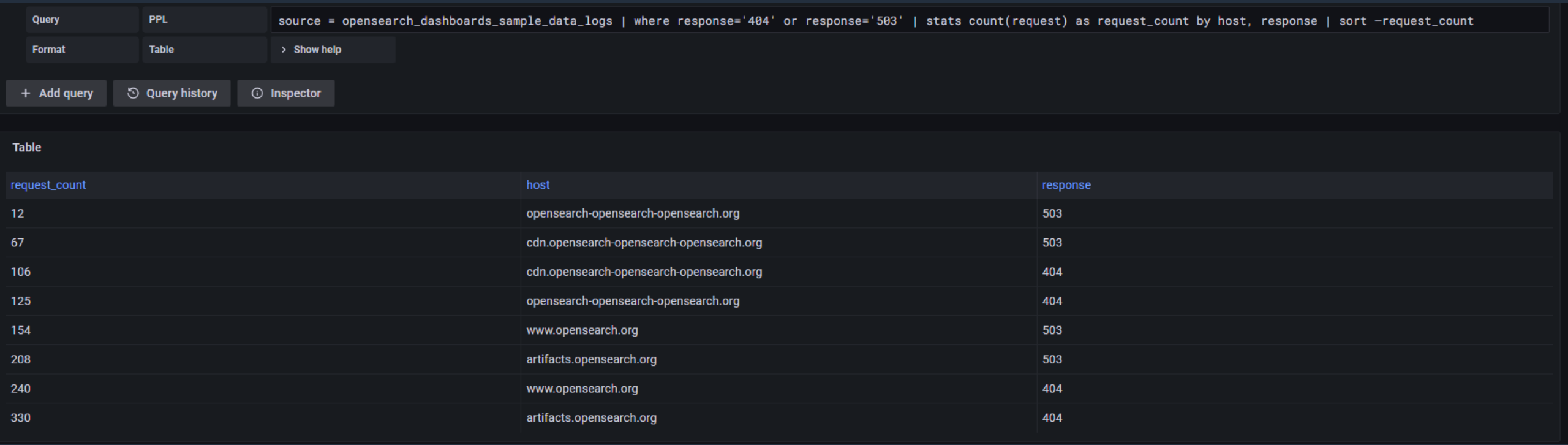
Amazon Managed Grafana (AMG)

拡張性、安全性、高可用性を
備えたフルマネージドな
Grafana サービス

Grafana は、ポピュラーな
オープンソースの
分析プラットフォーム。
保存されている場所に関係
なく、複数のオープンソース、
クラウド、サードパーティの
データソースから得られる
メトリクスのクエリ、可視化、
アラートを可能にする

Amazon Managed Service for Grafana による可視化

- OpenSearch のインデックスをグラフやテーブル形式で可視化
- Lucene Query に加えて PPL も利用可能
- VPC 内の OpenSearch ドメインもデータソースとしてサポート



request_count	host	response
12	opensearch-opensearch-opensearch.org	503
67	cdn.opensearch-opensearch-opensearch.org	503
106	cdn.opensearch-opensearch-opensearch.org	404
125	opensearch-opensearch-opensearch.org	404
154	www.opensearch.org	503
208	artifacts.opensearch.org	503
240	www.opensearch.org	404
330	artifacts.opensearch.org	404

https://docs.aws.amazon.com/grafana/latest/userguide/using-opensearch-in-AMG.html

Amazon QuickSight

- クラウドネイティブに作られたサーバーレス分析サービス
- 高速、スケーラブル、堅牢なセキュリティ
- 大規模利用に適した従量課金制のコスト体系

Built for scale

Developer focused

ML-powered insights for end users

Enterprise ready

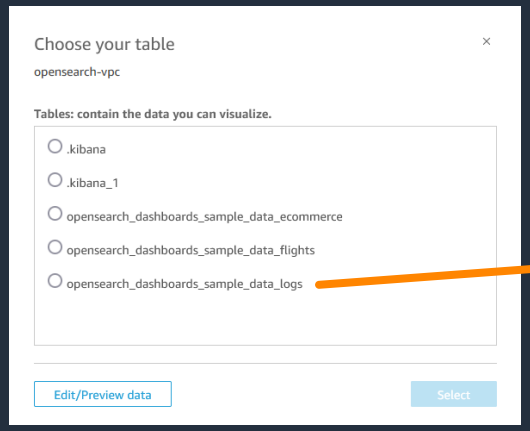
Scalable Pricing

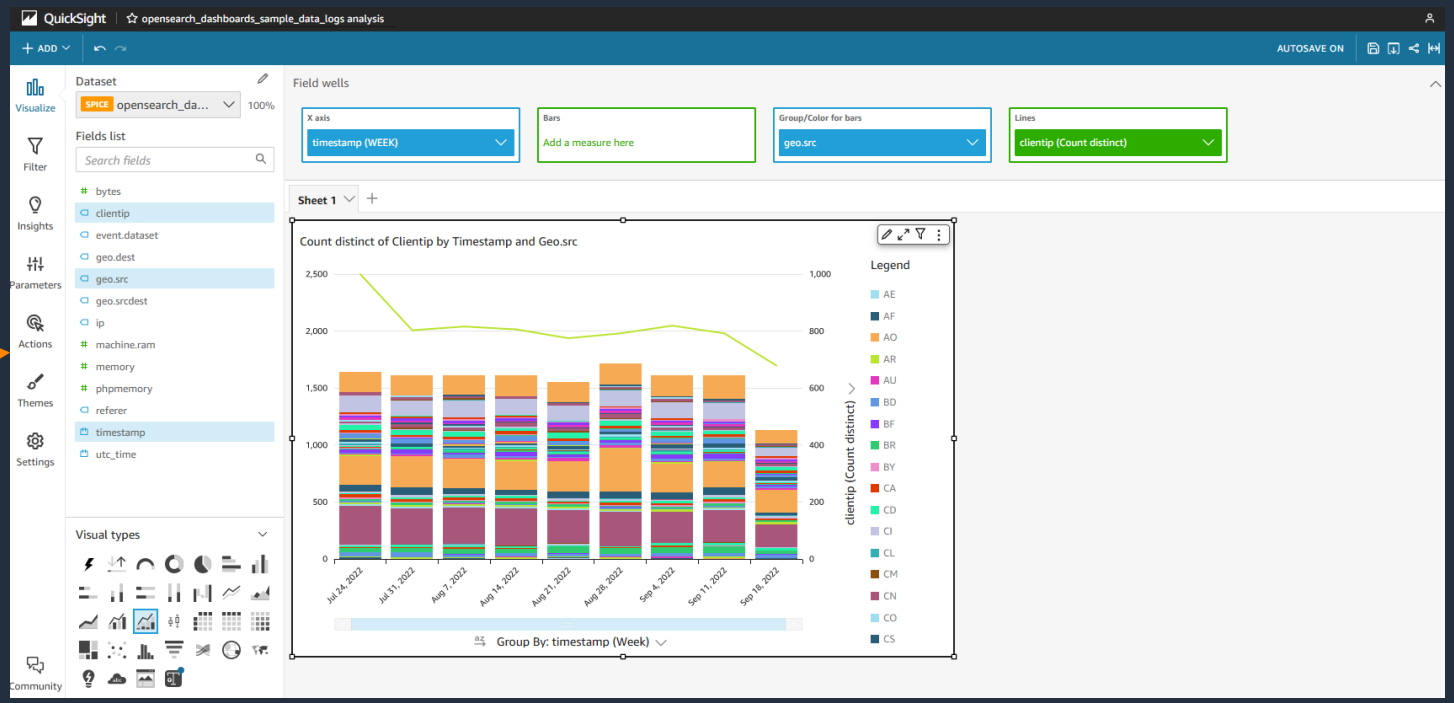
Try it free Today @

Quicksight.AWS

Amazon QuickSight による可視化

- OpenSearch インデックスのデータを使った可視化・分析が可能
- VPC 内の OpenSearch ドメインもデータソースとしてサポート
- 一部利用できないビジュアルタイプが存在する
- text 型や object 型など, QuickSight でサポートされていないデータ型のフィールドは利用不可



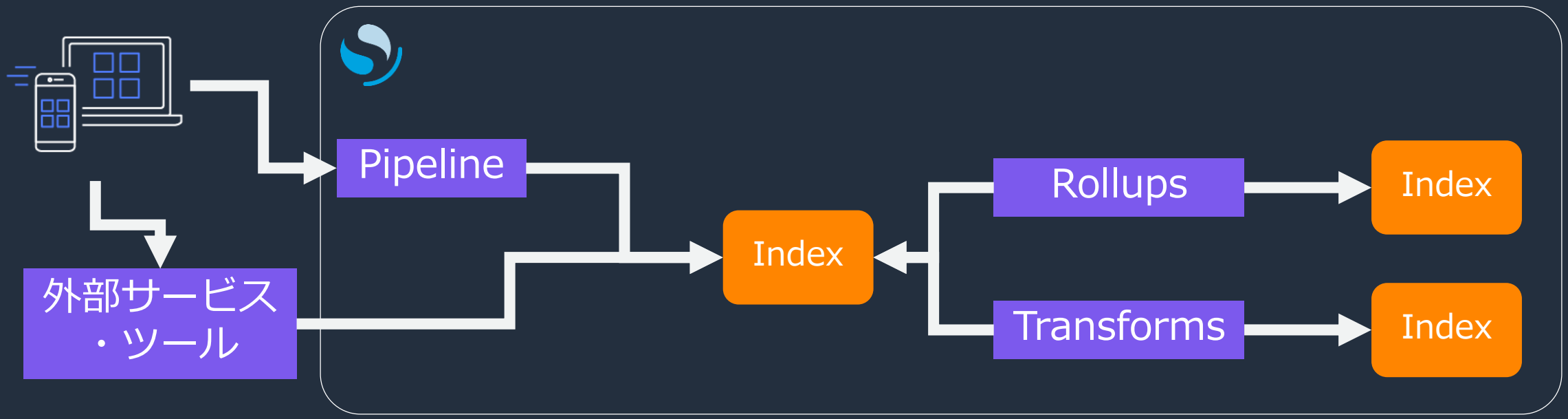


データ変換

OpenSearch におけるデータ変換

- OpenSearch には Pipeline という機能があり、データ投入時に簡易的な変換処理を行うことができる。また Index Rollups や Index Transforms といった機能を活用することで、OpenSearch 上で完結したデータ変換処理を行うことも可能

- Amazon Kinesis Data Firehose や AWS Glue などの AWS サービスやその他サードパーティツールを OpenSearch 外で実行する方法も有る。より柔軟なデータの処理が行える他、データ処理にかかる負荷を外部にオフロードできるメリットもある



Pipeline

小規模な変換処理をデータ取り込み前に行うための機能。
以下のような変換処理をサポート

- フィールドの追加、削除、変換、フィールド名の置き換え
- JSON 文字列のパース
- 正規表現のサポート
- 文字列の置き換え、分割
- Painless Script を利用したカスタム処理

https://opensearch.org/docs/latest/opensearch/rest-api/ingest-apis/index/

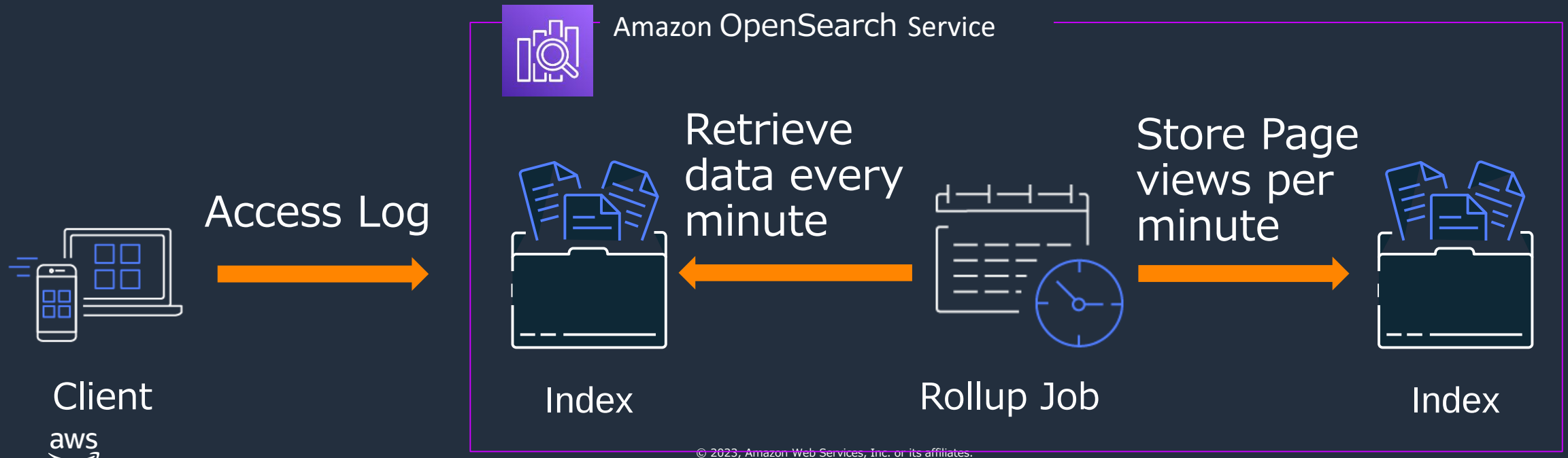
Pipeline で使用可能な処理(一部抜粋)

Pipeline で使用可能な処理を Processor と呼ぶ。
Processor の一覧は _nodes API より確認可能

- データ変換(文字列を配列に変換、データ型の変換など)
- 文字列操作(大文字小文字変換など)
- フィールド操作(特定フィールドの削除、フィールドの分割、特定フィールドから別のフィールドを生成など)
- 配列操作(要素の追加、ソート)
- バイナリファイル(PDF, PPT, XLS など)からの文字抽出

Index Rollups

- 時系列データを丸めたサマリを作成する機能。サマリは別インデックスに保存
- グラフ表示や集計処理の時間短縮に有用
- 定期実行、差分集計をサポート
  - AM 1:00 に前日の 0:00 から当日の 0:00 までのデータを集計するといった遅延実行にも対応



Index Rollups > 活用例 > 要件と課題

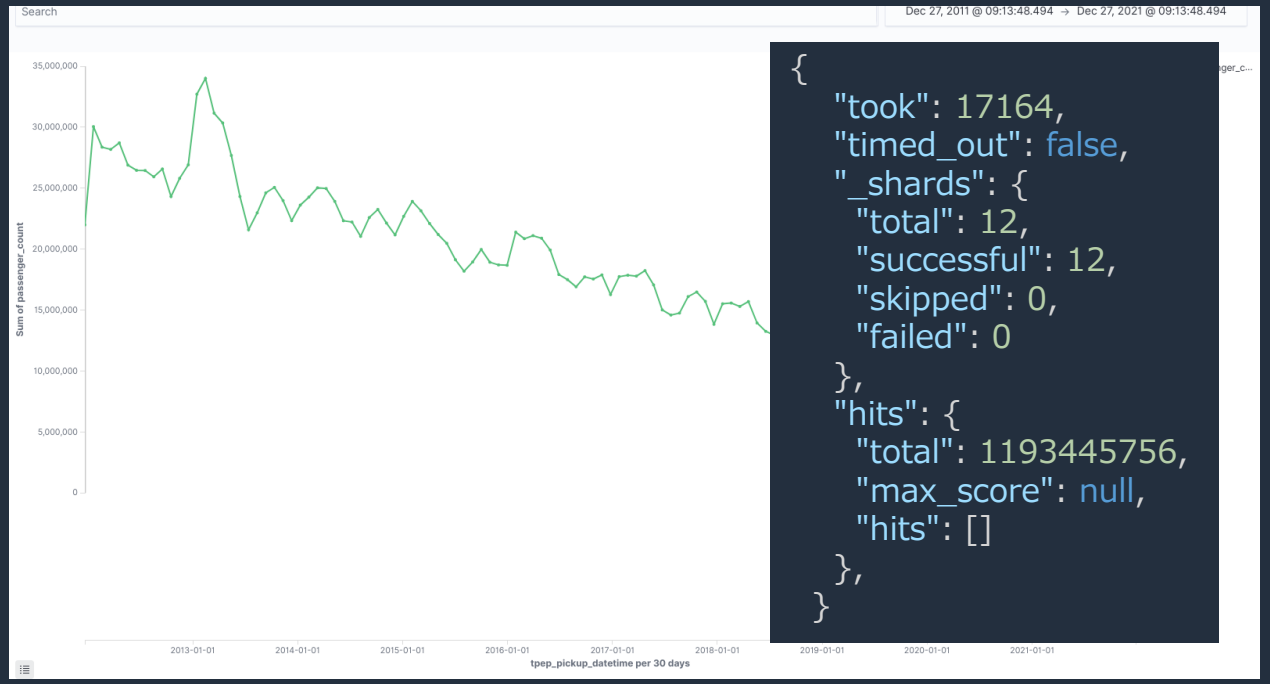
要件:タクシー乗降データから日次の乗降客数を集計し、グラフで可視化する

課題: 対象のデータサイズが大きいため、全期間のデータに対して集計を行った場合、
ダッシュボードの表示速度が低下する

対象 Index

index	pri.store.size
taxi-2009	26.7gb
taxi-2010	27.8gb
taxi-2011	27.6gb
taxi-2012	28.7gb
taxi-2013	28.4gb
taxi-2014	24.1gb
taxi-2015	23.1gb
taxi-2016	22gb
taxi-2017	16.4gb
taxi-2018	14.6gb
taxi-2019	13.2gb
taxi-2020	3.8gb

集計所要時間(17.1 秒)
＊ i3.16xlarge.search x 3 node 構成



Index Rollups > 活用例 > 対策と結果

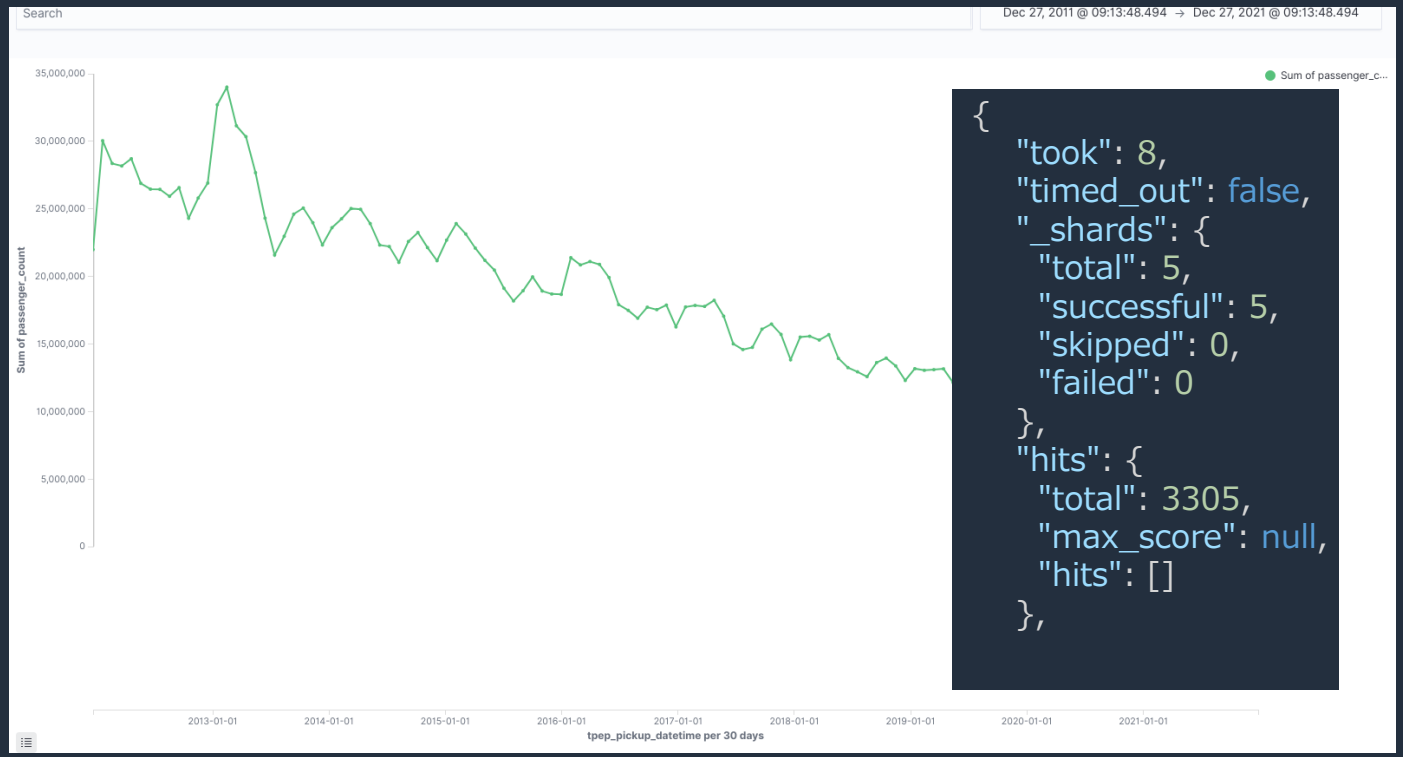
対応: 乗降客数のフィールドをロールアップジョブで日次集計

結果: ロールアップされたインデックスを可視化に使用することで、
グラフの表示速度が大幅に改善

対象 Index

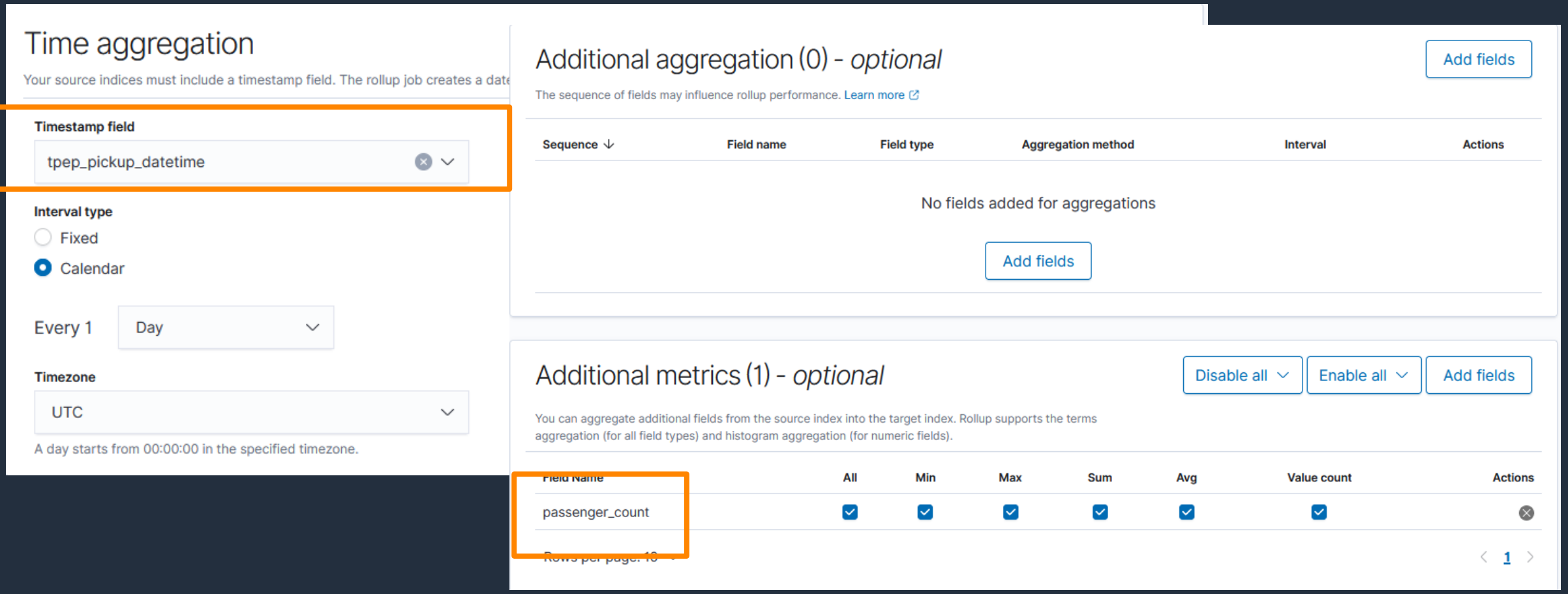
index	pri.store.size
daily_passenger_count	870.2kb

集計所要時間(8 ミリ秒)
i3.16xlarge.search x 3 node 構成



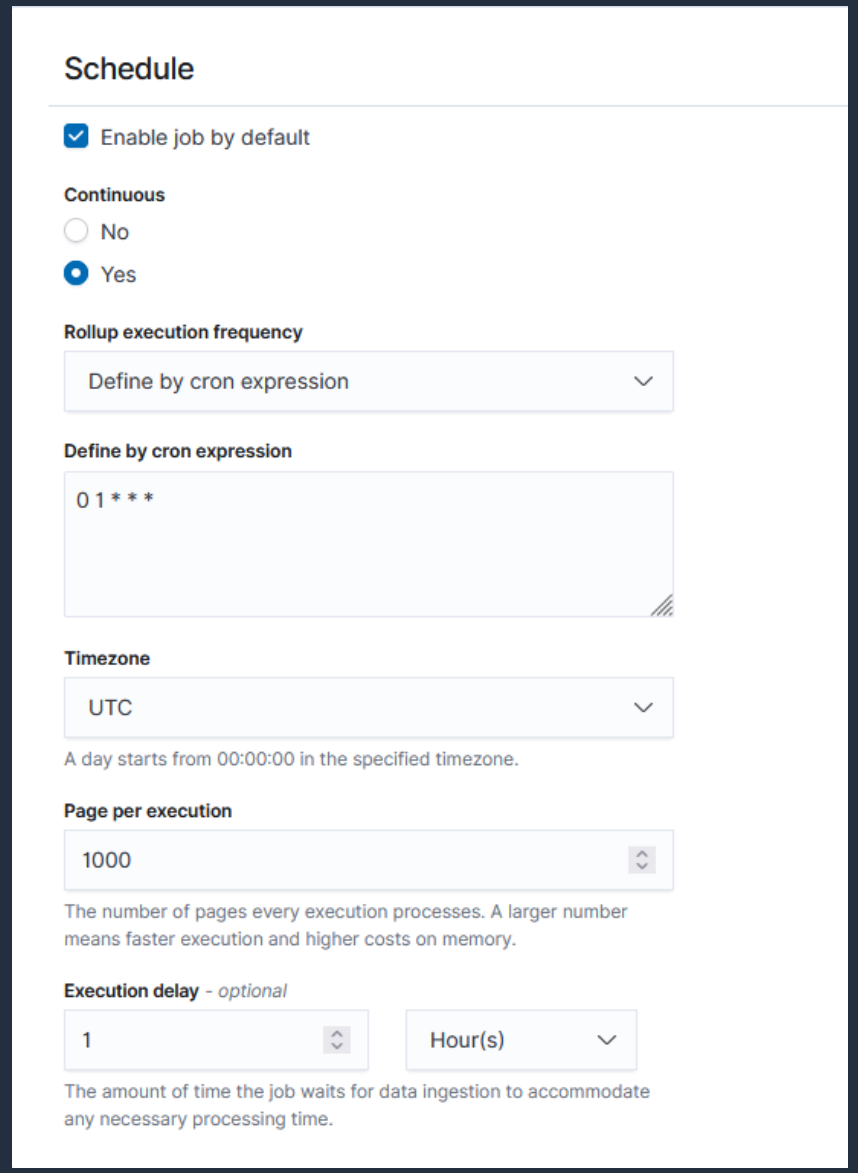
Index Rollups > 設定方法

タイムスタンプフィールド、集計対象のフィールド、集計に用いる統計を指定する



Index Rollups > スケジュール実行

- 定期実行により、一定間隔で集計処理を自動で実行することが可能
- 増分処理に対応しているため、ストリームデータに対するニアリアルタイムな集計処理を行うことが可能
- 遅延設定を有効化することで、集計タイミングと集計対象のウィンドウをコントロール可能
  - 例: 実行タイミングを毎日 1 時, 遅延を 1 時間に設定することで, 1:00 に前日の 0:00 から翌日の 0:00 までの集計が実行される



Index Rollups > 集計結果のフィルタリング

- Rollup 時に任意の集計軸を追加することが可能
- ここで指定したフィールドは、Rollup されたインデックスを使用して集計を行う際のフィルタ処理に使うことが可能

Additional aggregations (4)

Sequence ↓	Field name	Aggregation method	Interval
1	customer_gender	terms	-
2	geoip.city_name	terms	-
3	geoip.region_name	terms	-
4	day_of_week	terms	-

Rows per page: 10

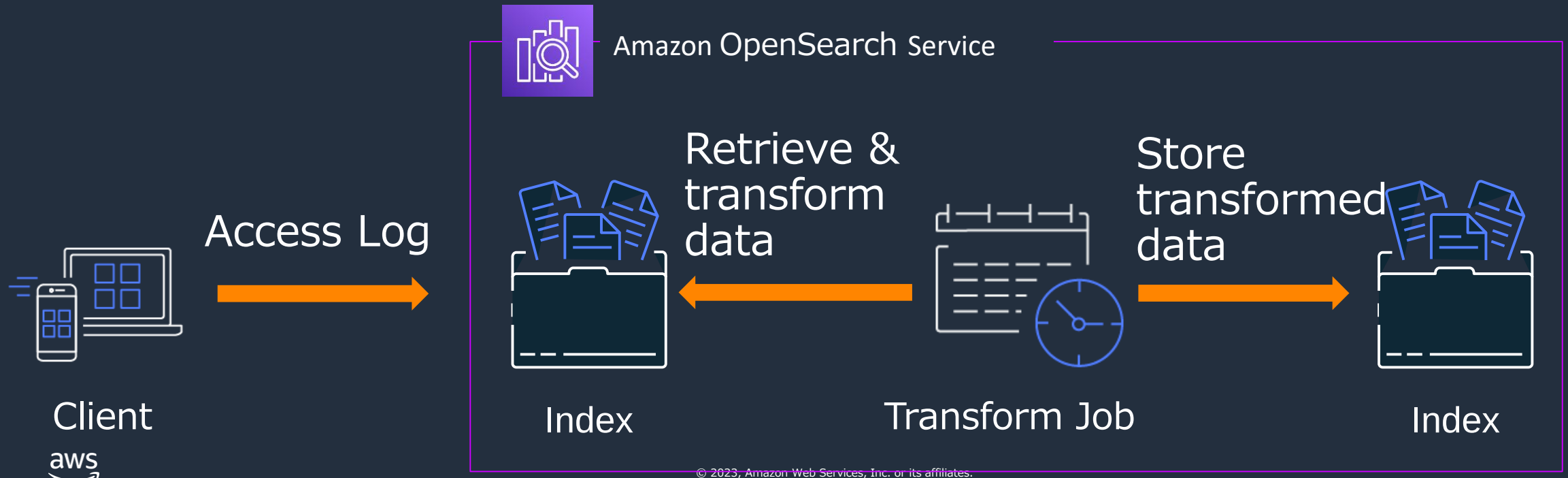
1

Index Rollups > 制約

- Rollup されたインデックスは集計処理でのみ利用可能
  - “size”: 0 の指定が必須
- 追加の集計で指定しなかったフィールドを指定してのデータフィルタリングは不可
- Rollup されたインデックスを再度 Rollup することはできない

Index Transforms

- ある Index に対する集計結果を別の Index に保存する機能
- 時系列データに加えて、**非時系列データ**の集計もサポート
- 定期実行、差分集計をサポート
- **多段集計された結果の可視化**を実現。事前集計を行うことでダッシュボードの表示速度の改善や、アクセス時の負荷低減といったメリットが得られる



Index Transforms > 活用例 > 要件と課題

要件

ショッピングサイトの購買履歴から、ユニークなユーザーごとの月額購入額の分布を作成したい

- 月額購入額1000円未満のユーザー数
- 1000 円以上 10000 円未満のユーザー数
- 10000円以上購入したユーザー数

課題

バッチ処理であれば以下のように段階的な処理で実装できるが、OpenSearch Dashboards の可視化では、そのような段階処理をサポートしていない

① ユーザーごとの月額売り上げ合計を aggregation で算出

② ①の結果を別の Index に保存し、range で分類

サンプルデータ

```
GET opensearch_dashboards_sample_data_ecommerce/_search?filter_path=hits.hits.fields{  "size": 1,  "fields": ["order_date", "customer_id", "taxless_total_price"],  "_source": false}
```


```
{
  "hits" : {
    "hits" : [
      {
        "fields" : {
          "order_date" : [
            "2022-01-10T09:28:48.000Z"
          ],
          "taxless_total_price" : [
            36.98
          ],
          "customer_id" : [
            "38"
          ]
        }
      }
    ]
  }
}
```

Index Transforms > 活用例 > 対策

対策: Index Transform を使用しユーザーごとの月次購入額の合計を事前集計

Select fields to transform

Original fields with sample data

Viewing sample data from index opensearch_dashboards_sample_data_ecommerce*

52 columns hidden

customer_id	order_date	taxless_total_price
38	01/10/22 6:28PM	36.98
20	01/10/22 6:59AM	53.98
26	01/10/22 7:32AM	199.98
22	01/10/22 7:58AM	174.98
38	01/03/22 12:48PM	80.98
22	01/03/22 6:44AM	71.98
7	12/27/21 6:27PM	45.98
52	01/10/22 11:19AM	138.96
17	01/02/22 9:59AM	88.96
5	01/10/22 12:41PM	171.96

Rows per page: 10 1 2 3 4 5

Transformed fields preview based on sample data

This fields preview displays only the first 10 results of your transform job.

Columns

order_date_date_histogram_month	customer_id_terms	sum_taxless_total_price
12/01/21 9:00AM	10	1977.3515625
12/01/21 9:00AM	11	2485.4453125
12/01/21 9:00AM	12	5052.0625
12/01/21 9:00AM	13	5170.12890625
12/01/21 9:00AM	14	1926.640625
12/01/21 9:00AM	15	4901.3125

Index Transforms > 活用例 > 結果

結果: 事前集計結果を利用した分類が可能、可視化も実現

```
GET
transform_opensearch_dashboards_sample_da
ta_ecommerce/_search?filter_path=aggregatio
ns.
order_date_date_histogram_month.buckets
{
  "size": 0,
  "aggs": {
    "order_date_date_histogram_month": {
      "date_histogram": {
        "field":
"order_date_date_histogram_month",
        "interval": "month"
      },
      "aggs": {
        "sum_taxless_total_price": {
          "range": {
            "field": "sum_taxless_total_price",
            "ranges": [
              { "from": 0, "to": 1000 },
              { "from": 1000, "to": 10000 },
              { "from": 10000}
            ]
(...)
```

```
{
  "aggregations" : {
    "order_date_date_histogram_month" : {
      "buckets" : [
        {
          "key" : 1638316800000,
          "doc_count" : 46,
          "sum_taxless_total_price" : {
            "buckets" : [
              {
                "key" : "0.0-1000.0",
                "from" : 0.0, "to" : 1000.0,
                "doc_count" : 0
              },
              {
                "key" : "1000.0-10000.0",
                "from" : 1000.0, "to" : 10000.0,
                "doc_count" : 45
              },
              {
                "key" : "10000.0-*",
                "from" : 10000.0,
                "doc_count" : 1
              }
            ]
          }
        },
```

Index Transforms > 活用例 > 解説

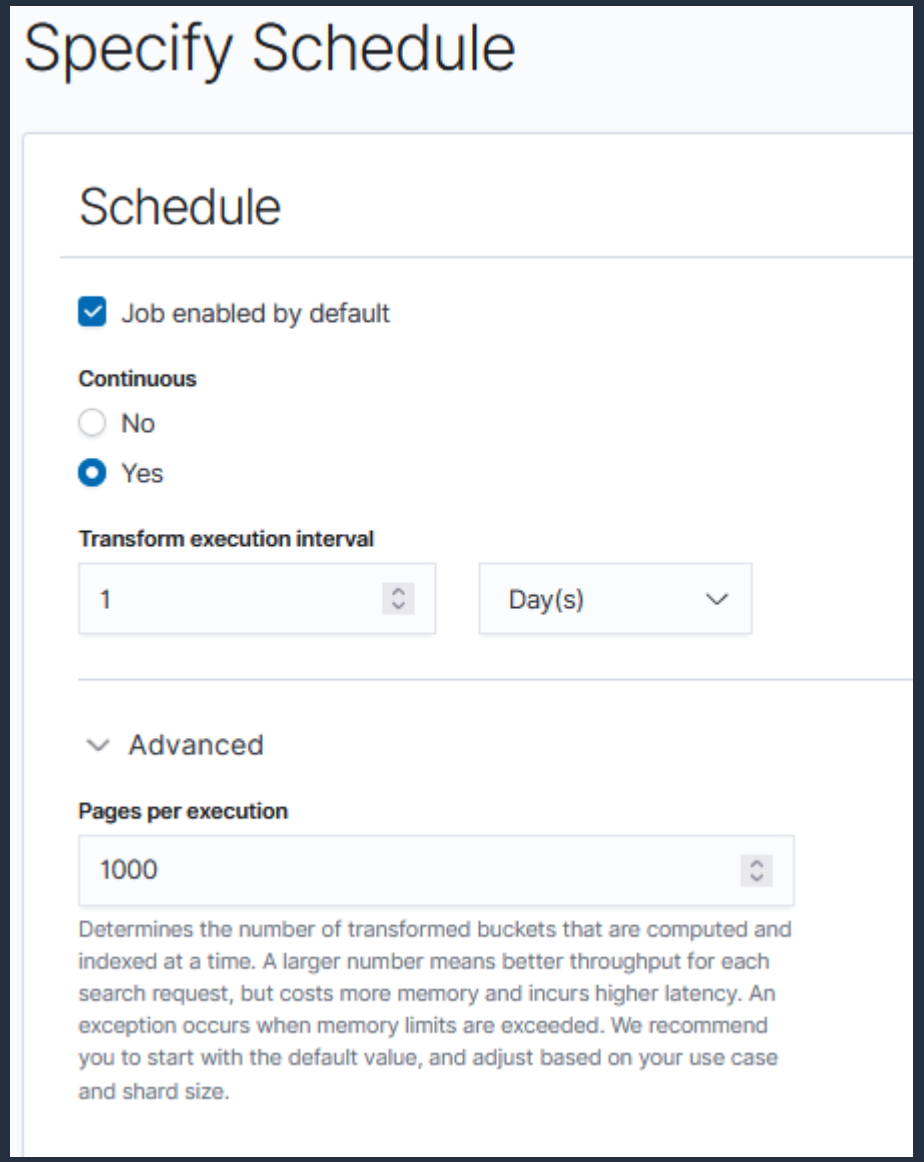
解説: Index Transform による個々の集計結果は document として格納されるため、aggregation による再集計が可能

```
GET transform_opensearch_dashboards_sample_data_ecommerce/_search?filter_path=hits.hits._source

{
  "hits" : {
    "hits" : [
      {
        "_source" : {
          "transform._id" : "opensearch_dashboards_sample_data_ecommerce",
          "transform._doc_count" : 32,
          "order_date_date_histogram_month" : 1638316800000,
          "customer_id_terms" : "10",
          "sum_taxless_total_price" : 1977.3515625
        }
      },
      {
        "_source" : {
          "transform._id" : "opensearch_dashboards_sample_data_ecommerce",
          "transform._doc_count" : 67,
          "order_date_date_histogram_month" : 1638316800000,
          "customer_id_terms" : "18",
          "sum_taxless_total_price" : 4331.796875
        }
      },
      (...)
```

Index Transforms > 定期実行

- 定期実行により、一定間隔で集計処理を自動で実行することが可能
- 増分処理に対応しているため、ストリームデータに対するニアリアルタイムな集計処理を行うことが可能
- Index Rollups にあった遅延実行は非サポート
- 同じ軸での集計結果は上書きされるようにドキュメント ID が設計されているため、仮に同じデータが二回処理されたとしても、集計結果自体が重複して保存されることは無い



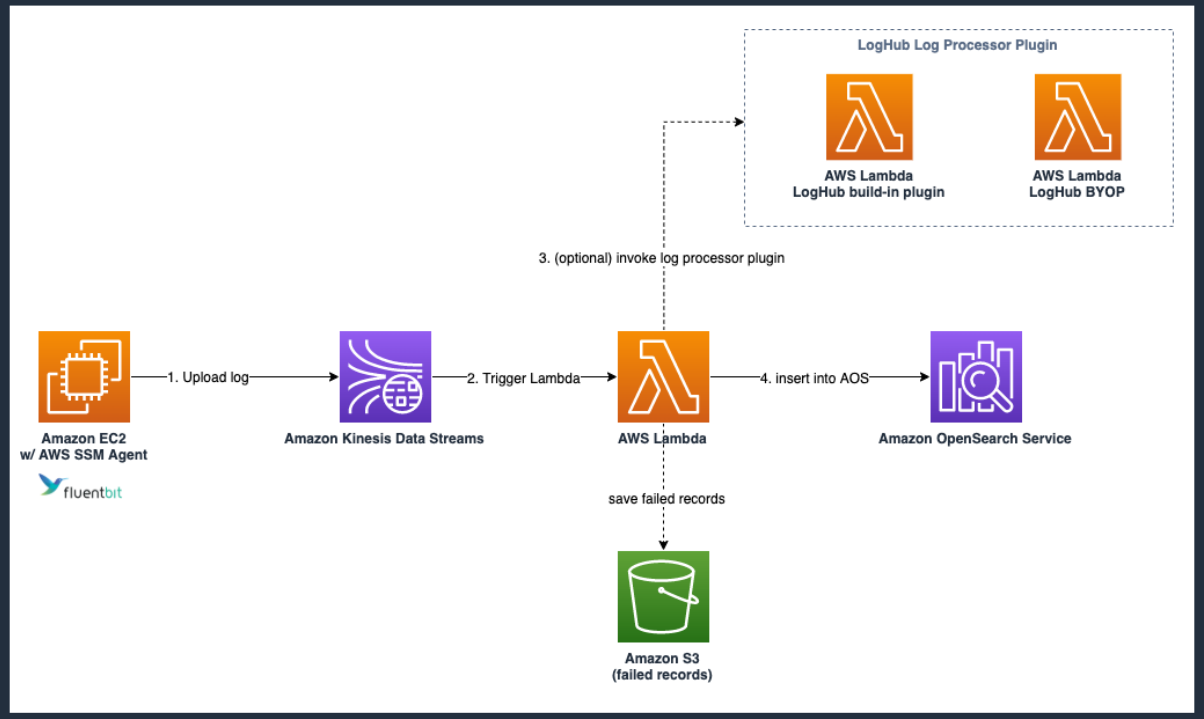
AWS ソリューションの活用

AWS ソリューションの活用
- ログ分析 -

ログ取り込みパイプライン構築時の課題

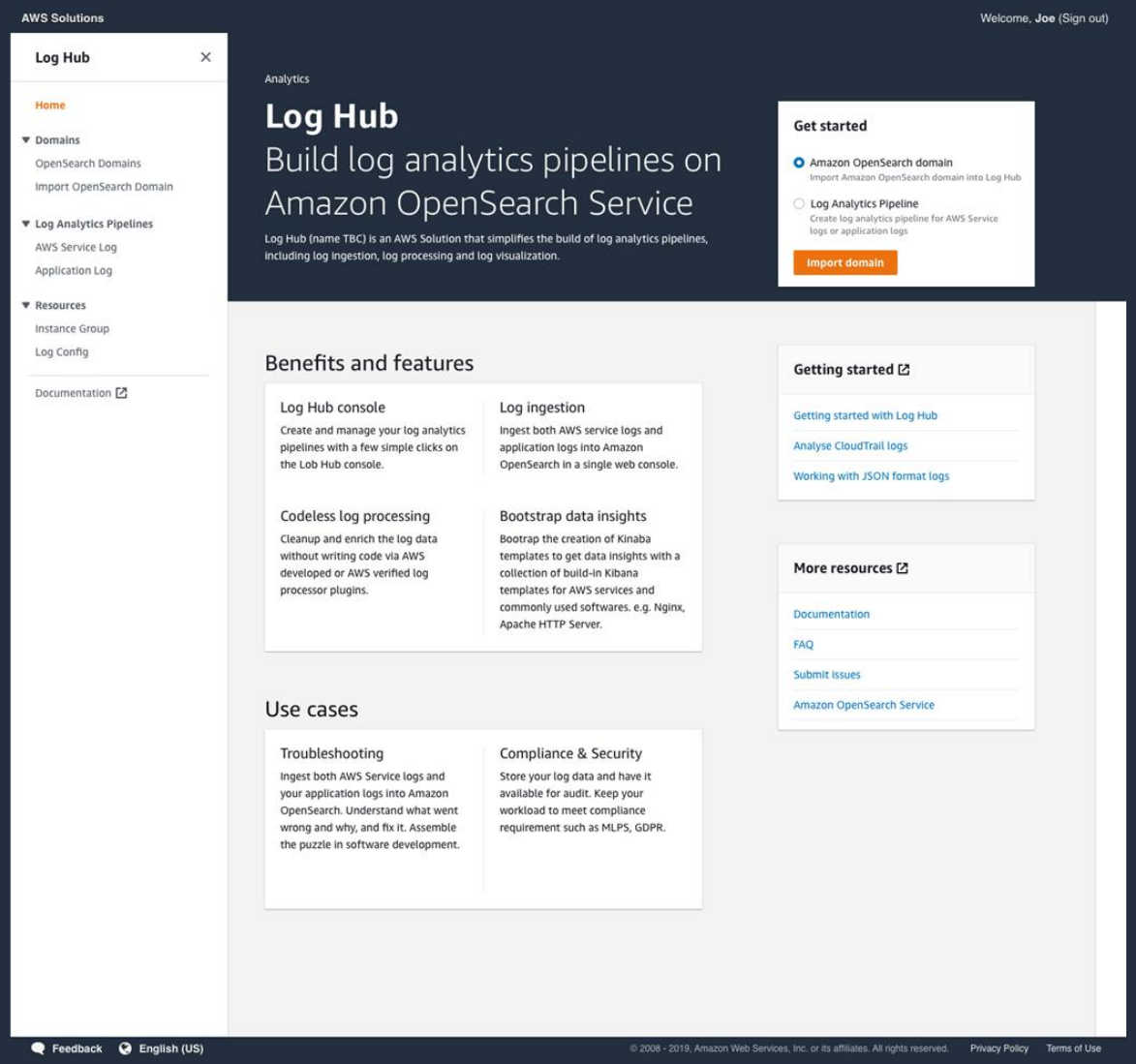
- データエンジニア、経験者の不足
- 学習曲線 (Amazon OpenSearch Service, Amazon Kinesis…)
- 様々なログ、一貫性のないフォーマット
- ログが分散しており、統一管理されていない
- 問題をどのように分析するか、またどの指標(メトリクス)を確認すべきかが分からない

nginx ログ取り込みパイプラインの構築例



1. ログ格納用の Kinesis Data Streams および S3 バケットの作成
2. 全ての EC2 インスタンスにログ収集用のログエージェントをインストール
3. IAM ロールとインスタンスプロファイルを作成し、Kinesis 関連の権限を付与
4. EC2 インスタンスに IAM ロールを関連付け
5. EC2 インスタンスにログインし、ログエージェントの設定ファイルを修正。データソース、フォーマット、宛先リソースを記載。
6. データ変換、データ配信、エラーハンドリングを行う Lambda Function を実装
7. OpenSearch 内のログライフサイクルを設定
8. OpenSearch Dashboards 上でダッシュボードを作成

Log Hub ソリューション

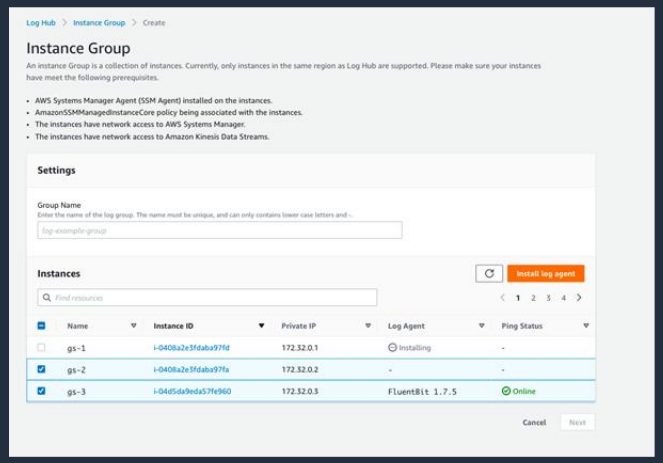


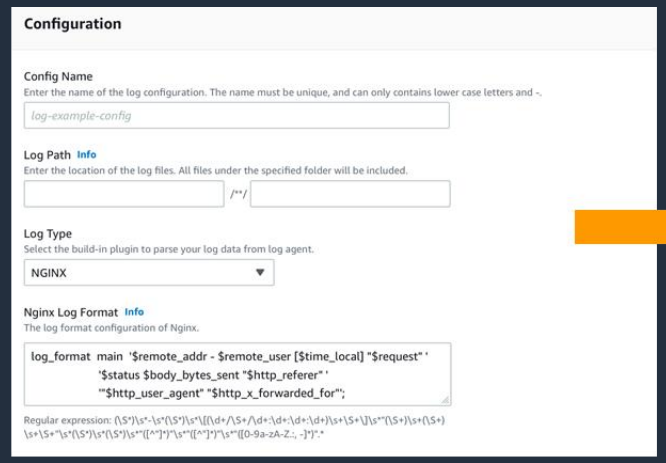
AWS CloudFormation テンプレートにより、迅速なログ分析パイプラインのソリューションを展開

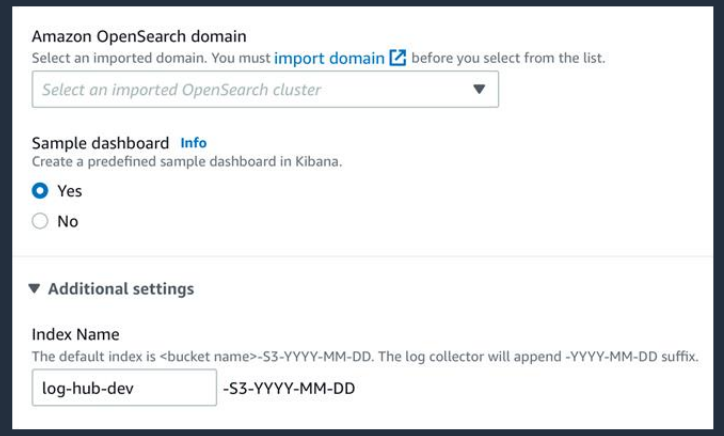
設定は独自の管理コンソールで完結

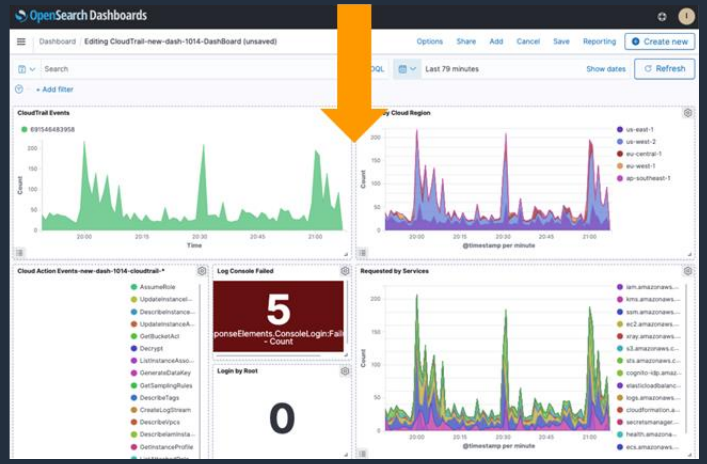
既存の OpenSearch ドメインを利用可能

Log Hub を使用した Nginx ログ分析基盤の構築





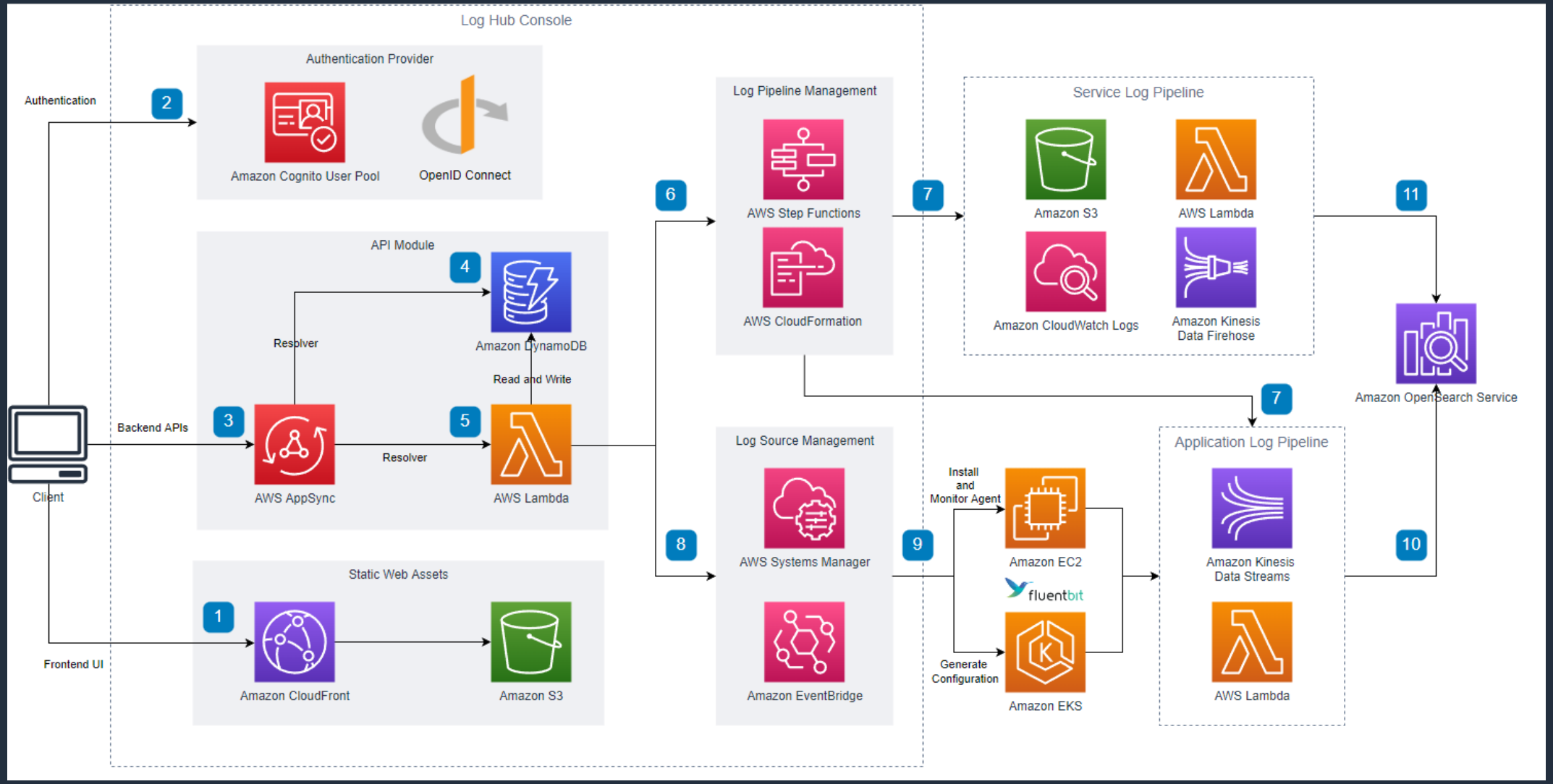




1. Log Hub コンソールでマウスをクリックしてマシンを選択し、ログエージェントを自動的にインストール
2. Log Hub コンソールで、事前に設定されたログエージェント設定を選択し、ログエージェントに自動的に適用。
3. Log Hub コンソールで、宛先の OpenSearch ドメインを指定
4. OpenSearch Dashboards で事前に設定されたダッシュボードにアクセス、分析を開始

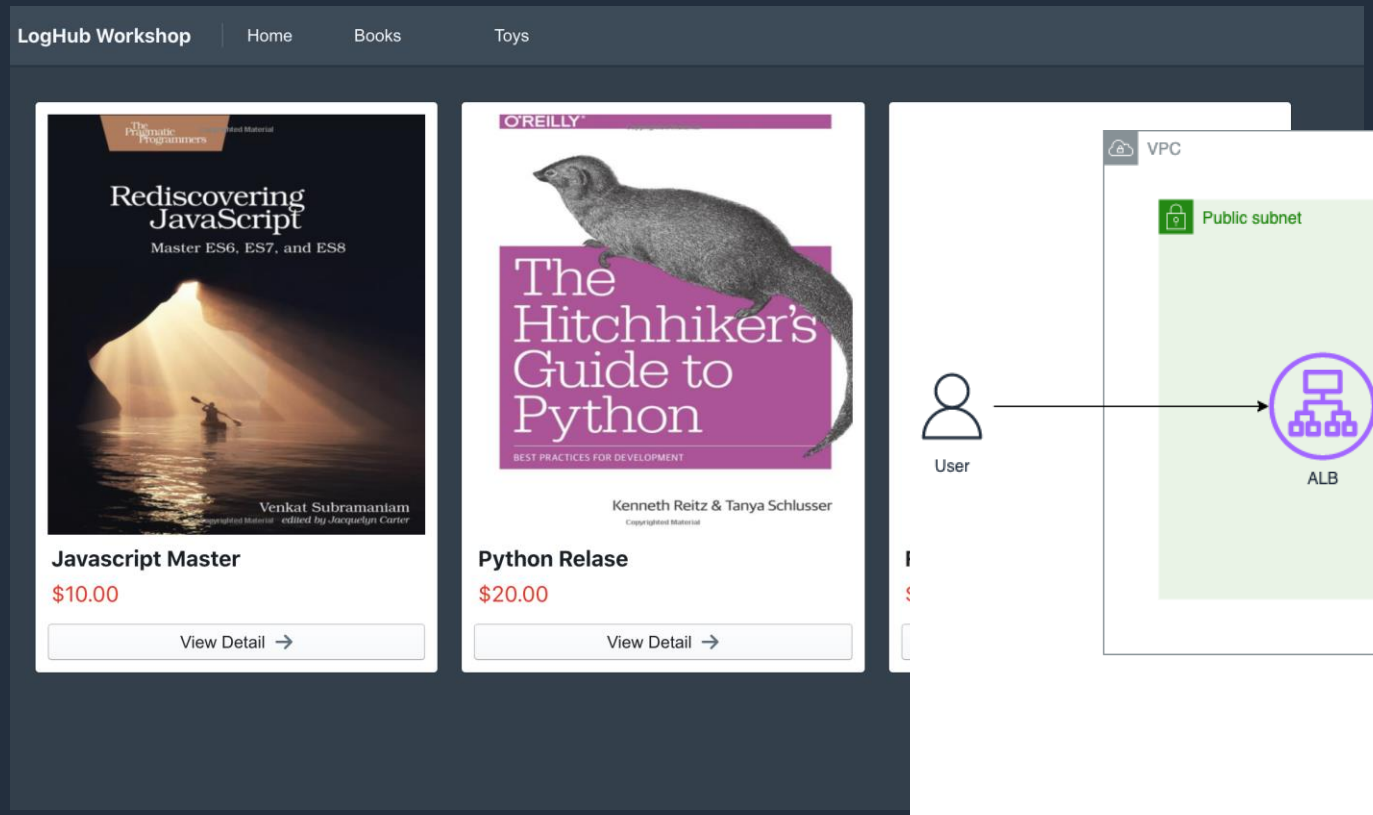
* **平均所要時間 = 20 分**

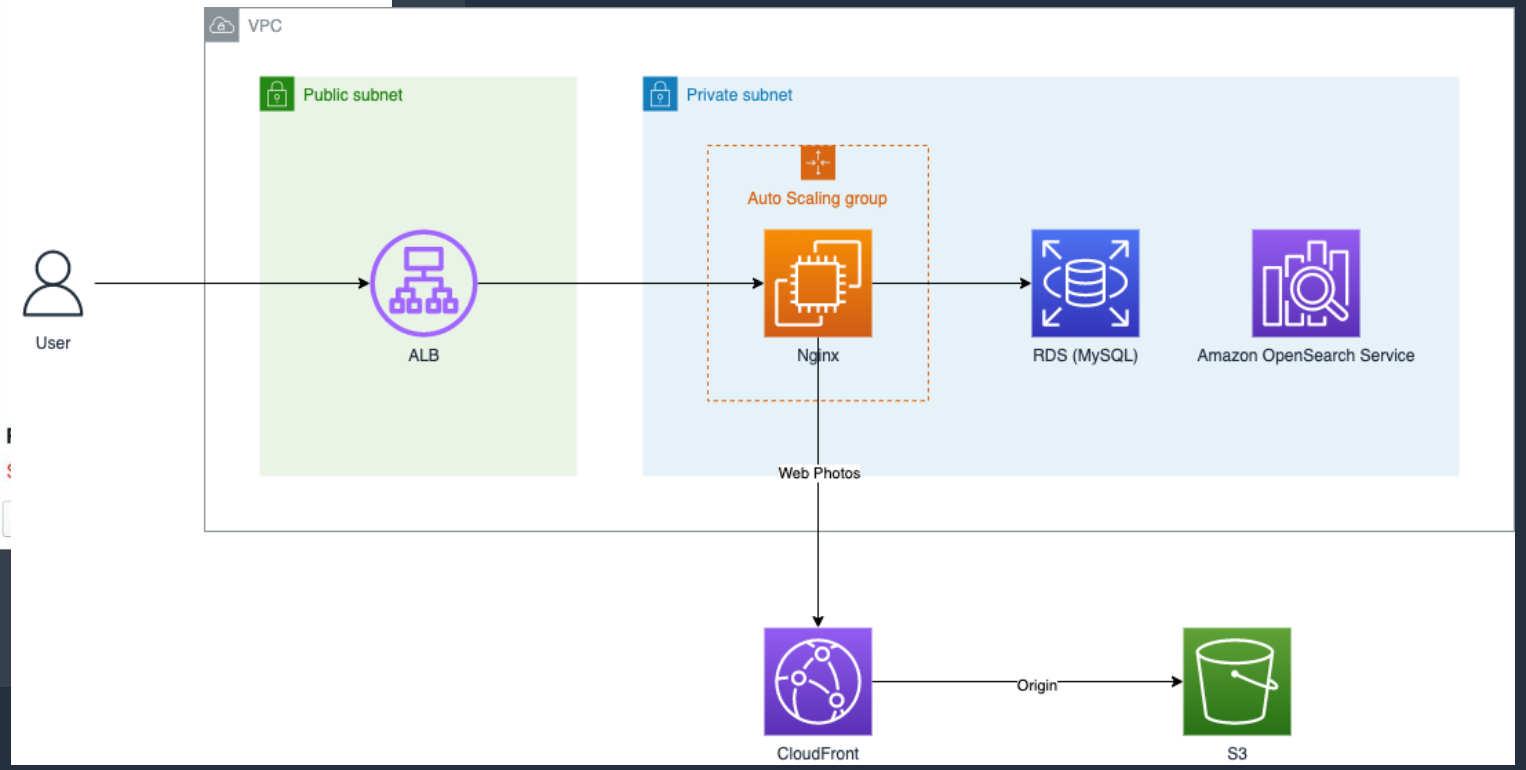
Log Hub アーキテクチャー



https://awslabs.github.io/log-hub/en/implementation-guide/architecture/

Log Hub Workshop





https://awslabs.github.io/log-hub/en/workshop/introduction/

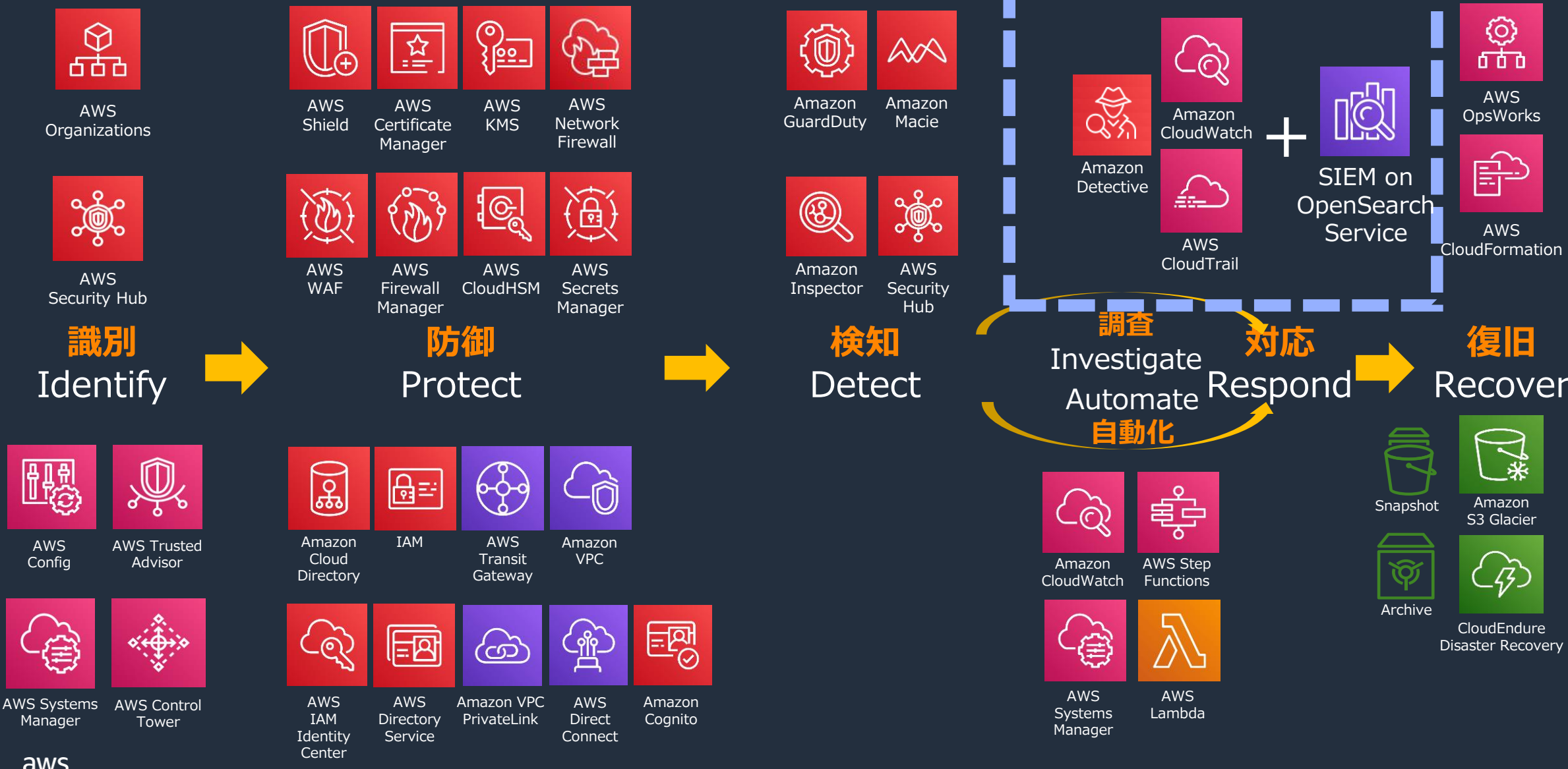
AWS ソリューションの活用
- セキュリティ分析 -

SIEM とは

Security Information and Event Management

セキュリティ機器、ネットワーク機器、その他のあらゆる機器のデータを収集及び一元管理をして、相関分析によって脅威検出と**インシデントレスポンス**をサポートするためのソリューション

AWS セキュリティサービス



インシデントレスポンスにおける調査とは？

インシデントレスポンスでは脅威の検出後に調査が必要

調査の目的

攻撃成否/被害の有無の判断(トリアージ)

根本原因の特定

被害規模や漏洩情報の特定

⇒ 調査結果に基づいて復旧対応

迅速に調査ができれば対応完了までの時間を短縮して、被害を最小化できる

例: S3 バケットに関するイベントの検出と分析・調査

検出

アラートの受信「S3 バケットがパブリックオープンになりました」

⇒ Amazon GuardDuty/Amazon Macie/Config Rules 等で自動検出が可能

分析と調査

- 誰がいつ、何の目的で設定変更をしたのか？
- 正規ユーザーによる正しいアクティビティか？
- パブリックオープンによりデータ漏洩等の影響はあったか？等々

⇒ **GuardDuty、CloudTrail、Config、S3 アクセスログ、**

送信元 IP の地理情報、アクティビティ履歴、等々を分析して判断

⇒ **目的の違い、複雑な条件分岐、人間による判断が伴うため全ての自動化は困難**

インシデントレスポンスにおける SIEM とその必要性

ログ調査の課題

- 脅威のアラート生成が複数に分散すると管理が難しい
- 調査対象が広範囲になりがちで多数のログの収集や分析で時間を要する

SIEMで解決

SIEM on Amazon OpenSearch Service

AWS サービスのセキュリティ監視をするためのスクリプトやダッシュボードのサンプル。
日本の セキュリティ/アナリティックス のソリューションアーキテクトが中心となって開発。
オープンソースソフトウェアとして公開。テンプレートは無料で利用可能。

■特徴

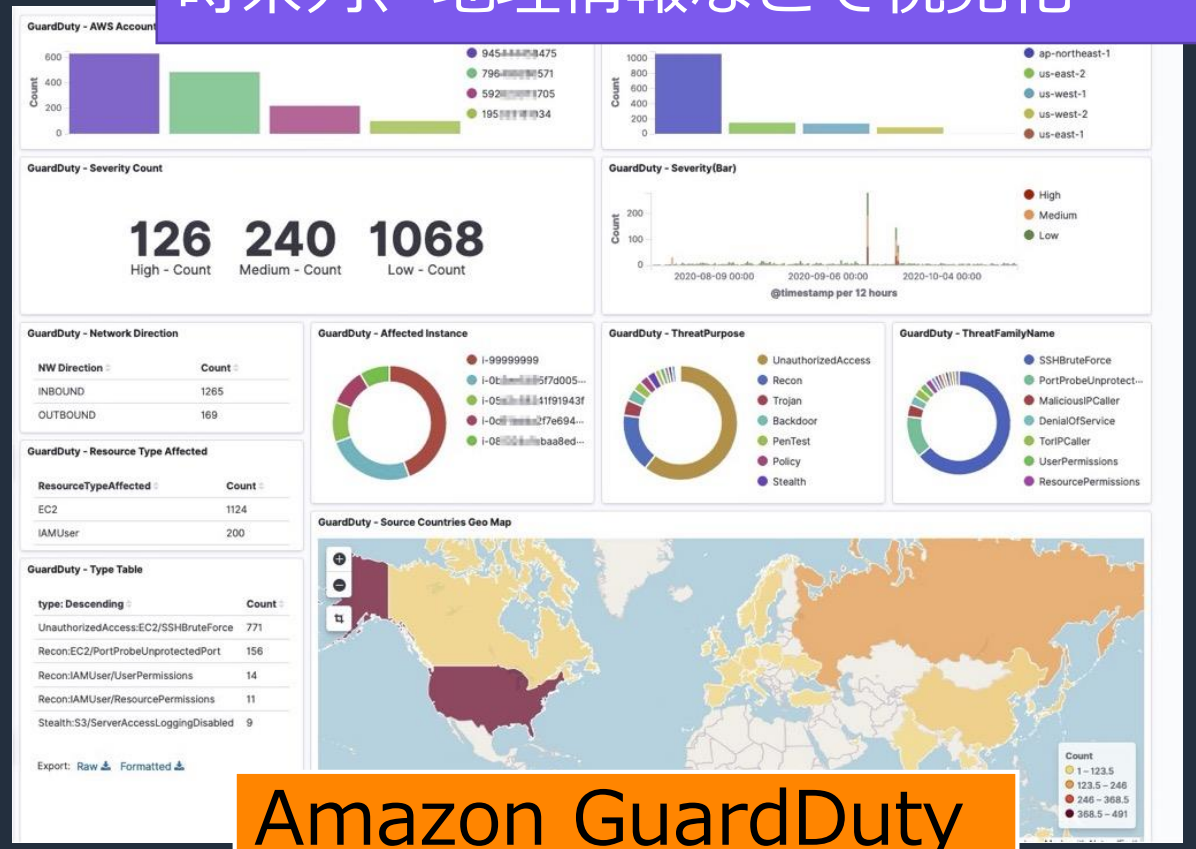
- マネージドサービスとサーバーレスのみで構成
- マルチアカウント・マルチリージョン対応
- AWS サービス専用の**ログ正規化**、**ダッシュボード**
- 脅威インテリジェンスによる**ログのエンリッチメント new!**
- **Amazon Security Lake**、**AWS Control Tower** との連携 **new!**
- AWS CloudFormation/AWS CDK によるデプロイ。約30分で完了

https://github.com/aws-samples/siem-on-amazon-opensearch-service

AWS サービス専用ダッシュボードの例

ログをセキュリティ分析の観点から
時系列、地理情報などで視覚化

AWS CloudTrail



Amazon GuardDuty

ログの正規化 (ETL)

- 正規化とは複数種類のログで同じ意味を持つフィールドに同一のフィールド名を付与
- 正規化により複数ログを効率的に検索が可能となる
- 正規化は **Elastic Common Schema (ECS)** 準拠

分析例) GuardDutyで脅威検知したインシデントの関連ログを複数ログから抽出
正規化処理されてない例 (検索式は概念的なもの)

```
CloudTrailに(recipientAccountId:111111111111 AND sourceIPAddress:198.51.100.1 ) OR
VPCFlowLogsに(account_id:111111111111 AND srcaddr:198.51.100.1 ) OR
GuardDutyに(accountId:111111111111 AND
 (service.action.awsApiCallAction.remoteIpDetails.ipAddressV4:198.51.100.1 OR
  service.action.portProbeAction.portProbeDetails.remoteIpDetails.ipAddressV4:  198.51.100.1))
```

正規化処理をした場合

```
3つのログ同時 ( cloud.account.id:111111111111 AND source.ip:198.51.100.1 )
```

正規化に対応している AWS サービスログ

- サポートされているログについてはデプロイ後すぐに分析可能
- 最新のサポートされているログについては Supported Log Types を参照
- 他のログについても順次対応予定

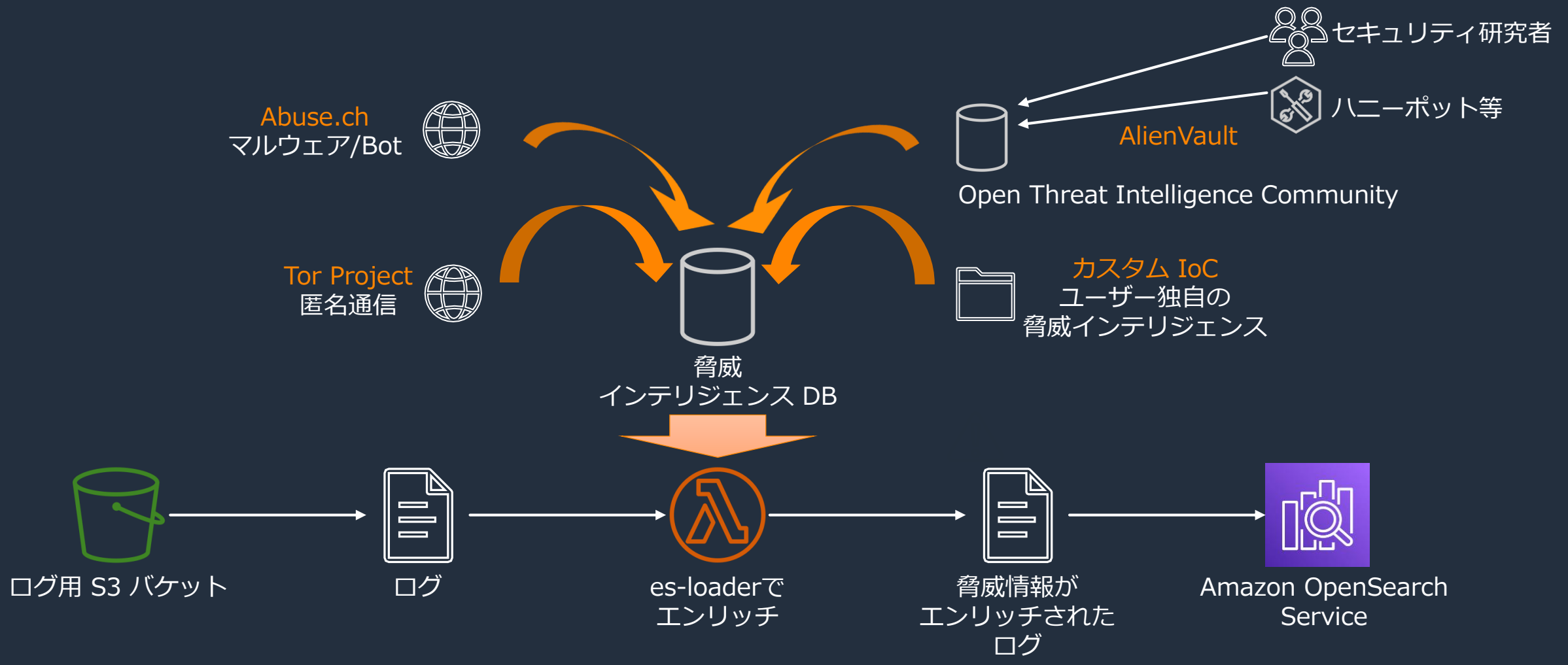
- AWS CloudTrail
- AWS Network Firewall
- Amazon VPC Flow Logs
- Amazon GuardDuty
- Amazon Inspector
- Amazon Security Hub
  - GuardDuty
  - Amazon Macie
  - Amazon Inspector
  - AWS IAM Access Analyzer
- Amazon Trusted Advisor
- AWS CloudHSM
- AWS Config / Config Rules
- Linux on EC2 (/var/log/message, /var/log/secure)
- Windows Server on EC2 (System/Security)
- AWS Directory Service (Microsoft AD)
- Amazon SFx for Windows File Server
- Amazon WorkSpaces

- Elastic Load Balancing
  - Application Load Balancer
  - Network load balancer
  - Classic Load Balancer
- AWS WAF
  - AWS WAF
  - AWS Classic WAF
- Amazon CloudFront
  - Standard access log
  - Real-time log
- Amazon Relational Database Service
  - MySQL/MariaDB/PostgreSQL
- Amazon Managed Streaming for Apache Kafka
- Amazon ElastiCache Service for Redis
- Amazon Managed Streaming for Apache Kafka
- Amazon Elastic Container Service
- Amazon S3 access log
- Route53 Resolver DNS query log
- AWS Client VPN

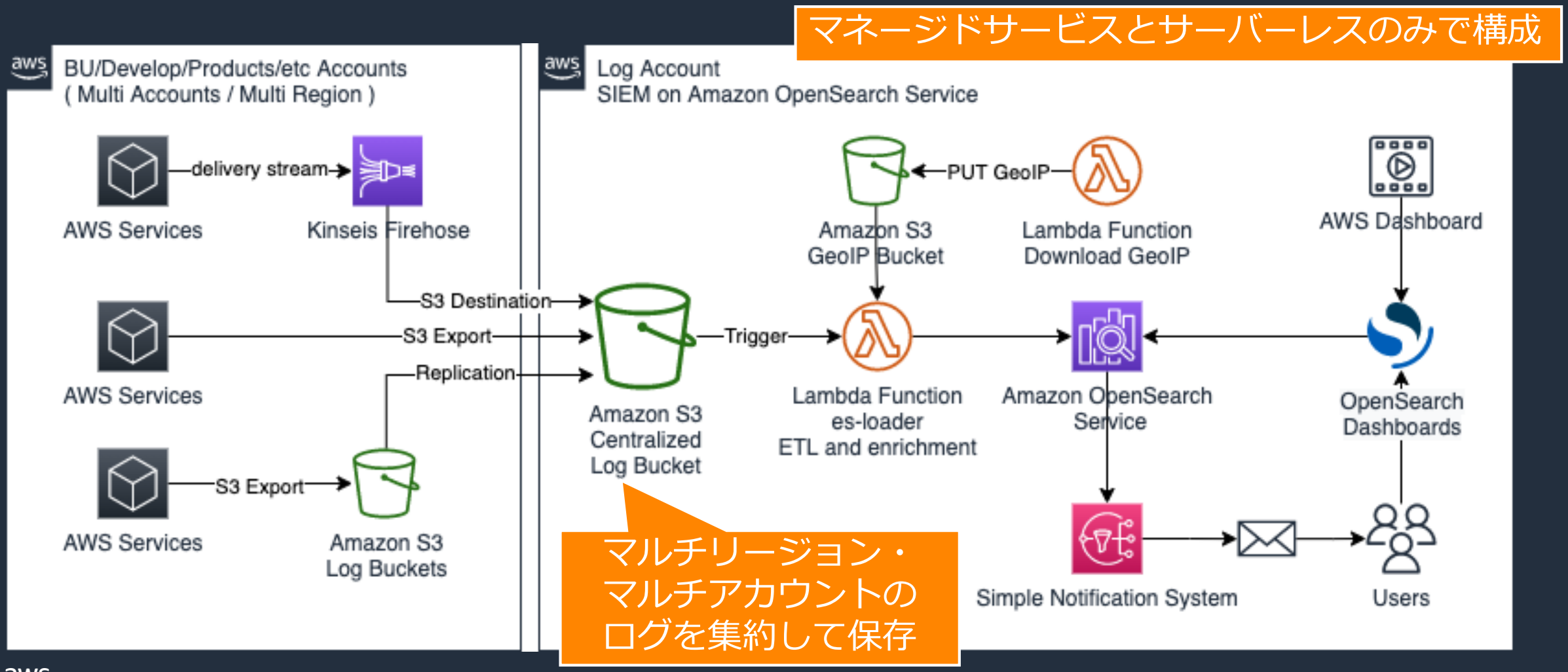
https://github.com/aws-samples/siem-on-amazon-opensearch-service#supported-log-types
https://github.com/aws-samples/siem-on-amazon-opensearch-service/issues

脅威インテリジェンスによるエンリッチメント

悪性のIPアドレスとドメイン情報を自動で収集して、ログに脅威情報をエンリッチ (付与)

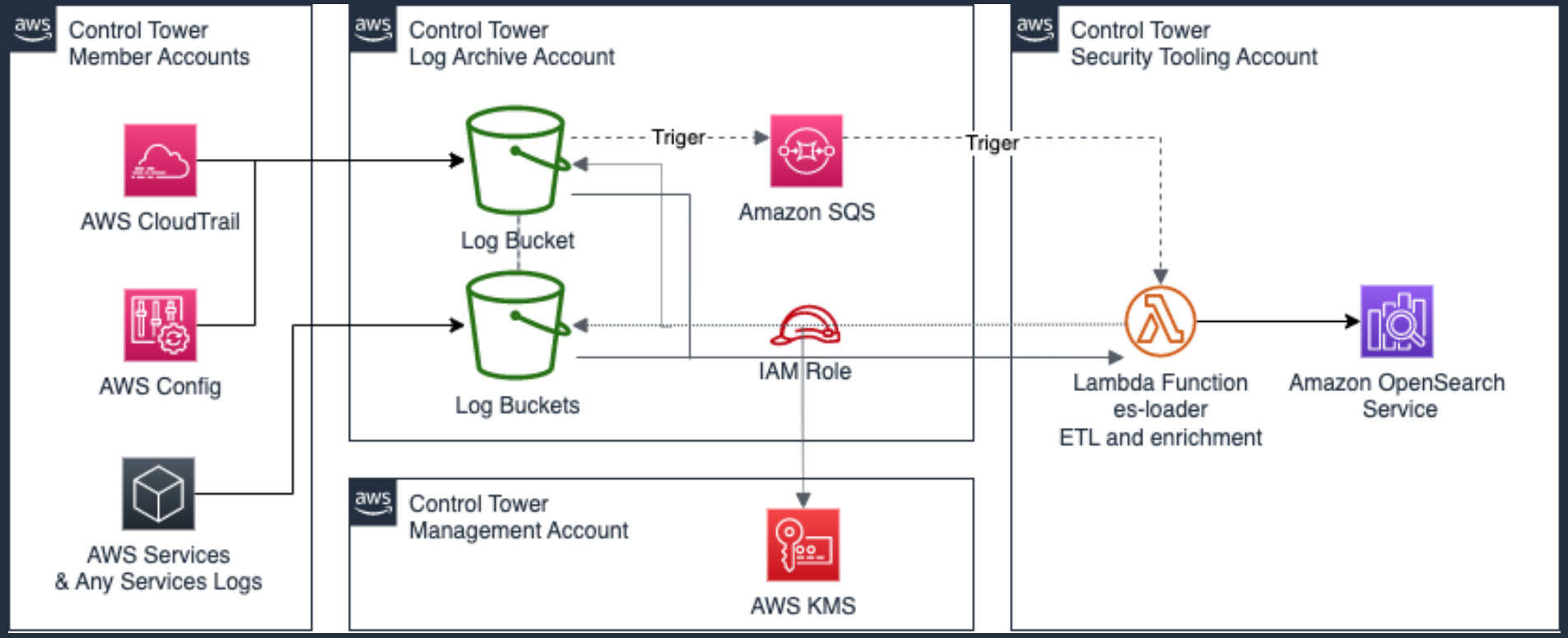


SIEM on OpenSearch Service のアーキテクチャ



AWS Control Tower との連携

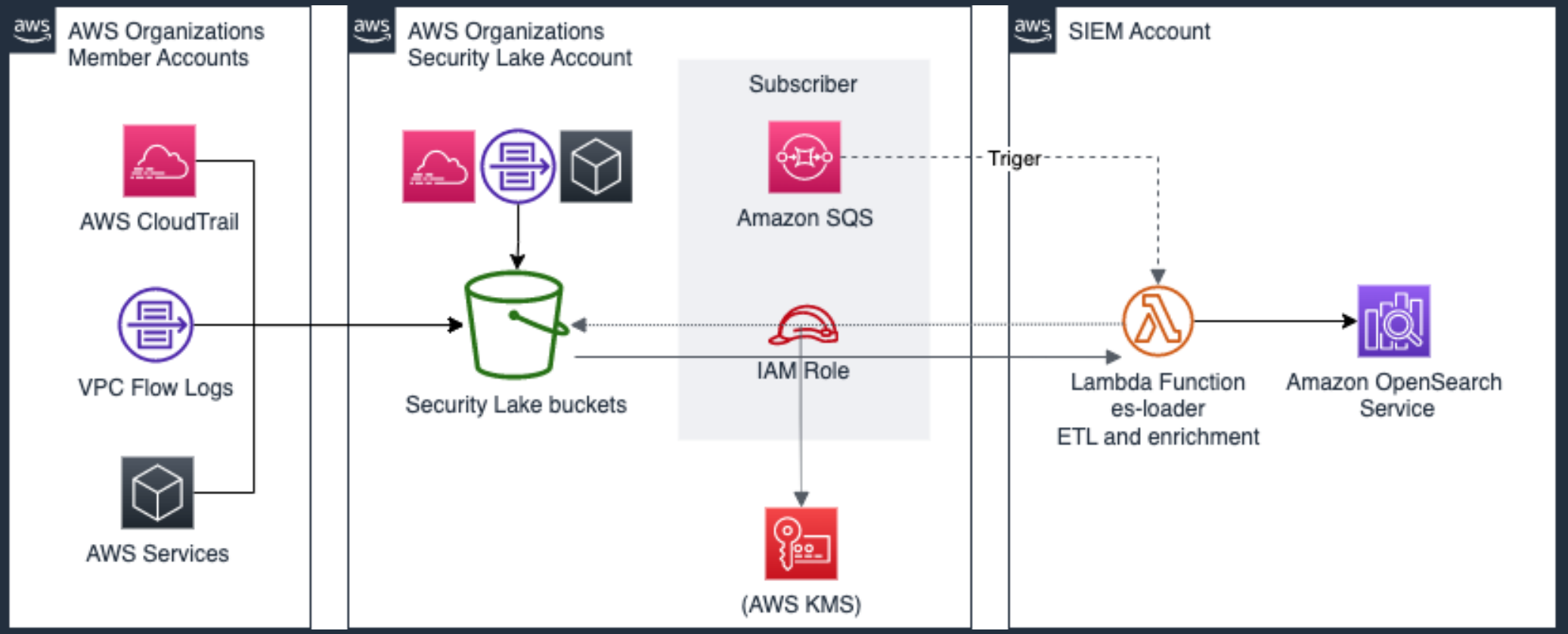
Control Tower の Log Archive アカウントのログを SIEM on OpenSearch に直接に取り込むことが可能



※ SIEM on OpenSearch は Audit アカウントまたは一般のアカウントにデプロイ

Amazon Security Lake との統合

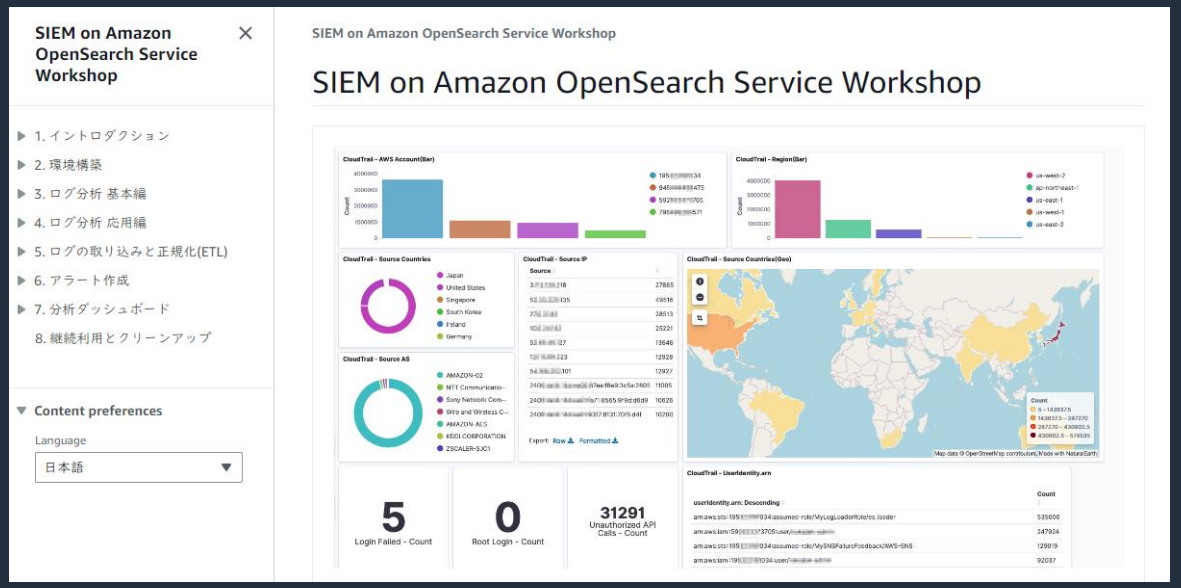
Security Lake で OCSF に正規化されたログを直接取り込み、複数ログを分析。
ECS で正規化されたログとも相関分析が可能



SIEM on Amazon OpenSearch Service Workshop

以下を実際に体験できるワークショップを提供

- AWS サービスのログの取り込み
- Dashboards の一般的な使い方 (ダッシュボード作成、アラート設定)
- サンプルログを使ったインシデント調査。以下のログ調査から根本原因、被害規模(漏洩情報)を特定
  - Amazon GuardDuty
  - AWS CloudTrail
  - Amazon VPC Flow Logs
  - Amazon Macie
  - Amazon Inspector
  - Linux SSH のログ



https://catalog.us-east-1.prod.workshops.aws/workshops/60a6ee4e-e32d-42f5-bd9b-4a2f7c135a72/ja-JP

その他補足事項

リファレンス

よくある質問:
https://aws.amazon.com/jp/opensearch-service/faqs/

トラブルシューティング:
https://docs.aws.amazon.com/ja_jp/opensearch-service/latest/developerguide/handling-errors.html

ナレッジセンター:
https://aws.amazon.com/jp/premiumsupport/knowledge-center/#Amazon_OpenSearch_Service

料金:
https://aws.amazon.com/jp/opensearch-service/pricing/

本資料に関するお問い合わせ・ご感想

技術的な内容に関しましては、有料のAWSサポート窓口へ
お問い合わせください

https://aws.amazon.com/jp/premiumsupport/

料金面でのお問い合わせに関しましては、カスタマーサポート窓口へ
お問い合わせください（マネジメントコンソールへのログインが必要です）

https://console.aws.amazon.com/support/home#/case/create?issueType=customer-service

具体的な案件に対する構成相談は、後述する個別相談会をご活用ください

ご感想はTwitterへ！ハッシュタグは以下をご利用ください
#awsblackbelt

その他コンテンツのご紹介

ウェビナーなど、AWSのイベントスケジュールをご参照いただけます

https://aws.amazon.com/jp/events/

ハンズオンコンテンツ

https://aws.amazon.com/jp/aws-jp-introduction/aws-jp-webinar-hands-on/

AWS 個別相談会

AWSのソリューションアーキテクトと直接会話いただけます

https://pages.awscloud.com/JAPAN-event-SP-Weekly-Sales-Consulting-Seminar-2021-reg-event.html

Thank you!