

# CobolとAWS連携 データ視覚化

# 自己紹介

名前:

・NGUYEN TUAN NGHIA (グエン・トUAN・**ギア**)

・[nghiant1@fsoft.com.vn](mailto:nghiant1@fsoft.com.vn)

役割:

・FPTジャパン ソリューションアーキテクト

好みテクノロジー:

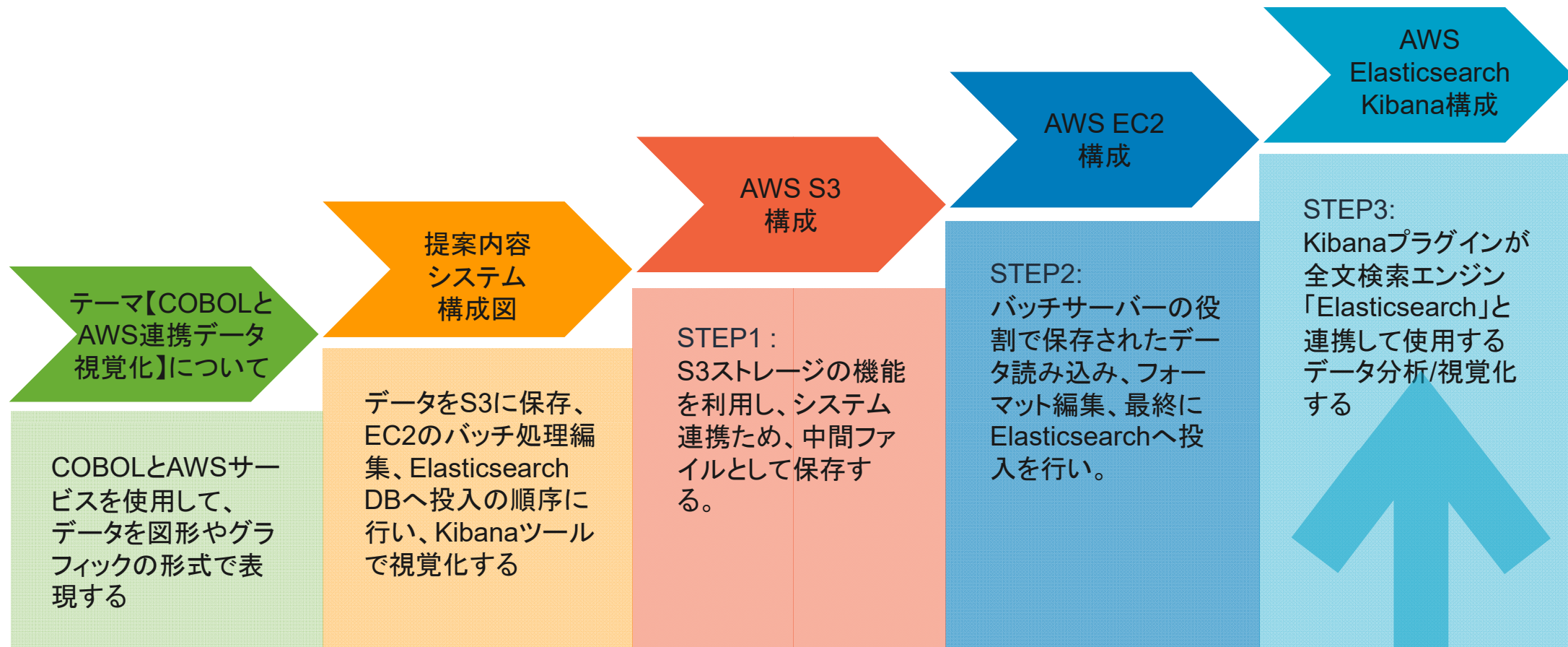
・AWSのS3, EC2, DB, Batch, Elasticsearch, Lucene

・IoTのraspberry pi , arduino

・Cobol, JP1, JCL, Shell

・その他色々

# 今日の流れ

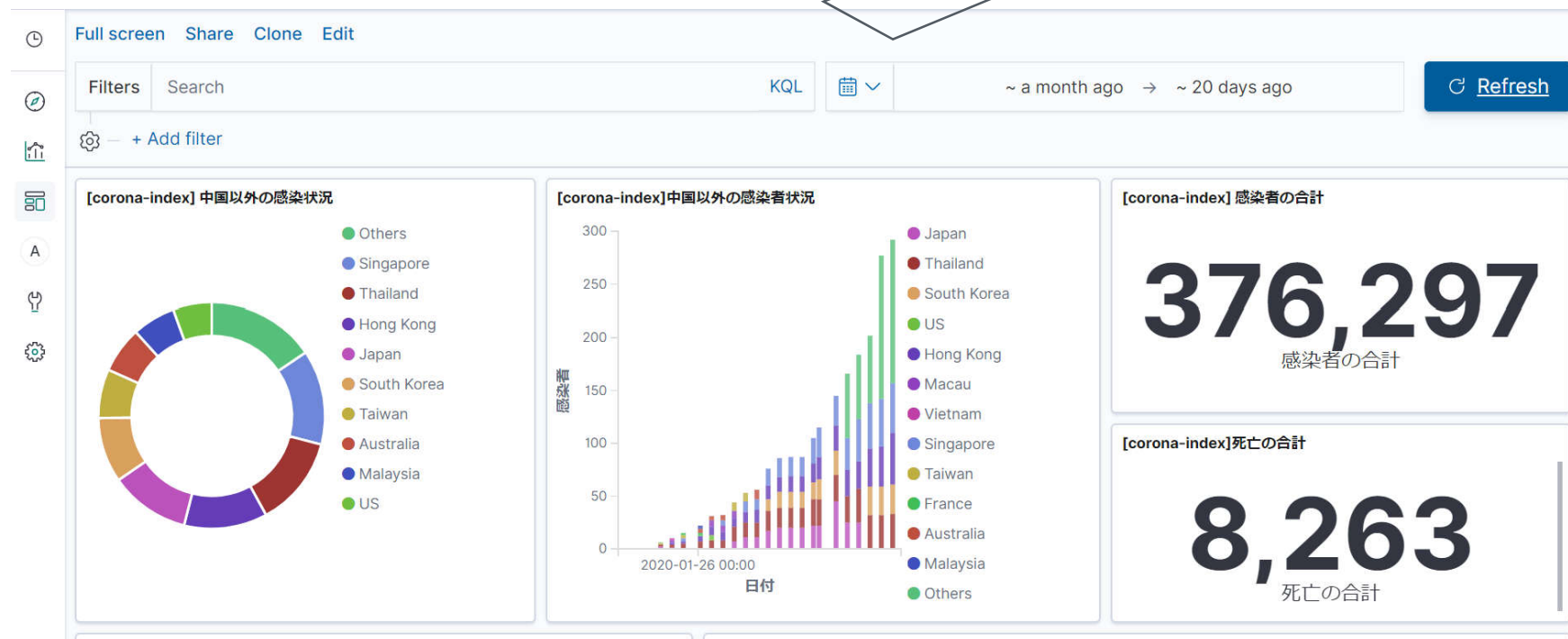




# 1. テーマ:「CobolとAWS連携データ視覚化」

最近コロナウイルスが世界中に広がり、WHOなどから公開された統計データに基づいてAWSでCobolとElasticsearch/Kibanaを使用して、データ分析を視覚化する。

Sno	Date	Province/State	Country	Last Update	Confirmed	Deaths	Recovered
1	2020-01-22T12:00:00	Anhui	China	2020-01-22T12:00:00	1	0	0
2	2020-01-22T12:00:00	Beijing	China	2020-01-22T12:00:00	14	0	0
3	2020-01-22T12:00:00	Chongqing	China	2020-01-22T12:00:00	6	0	0

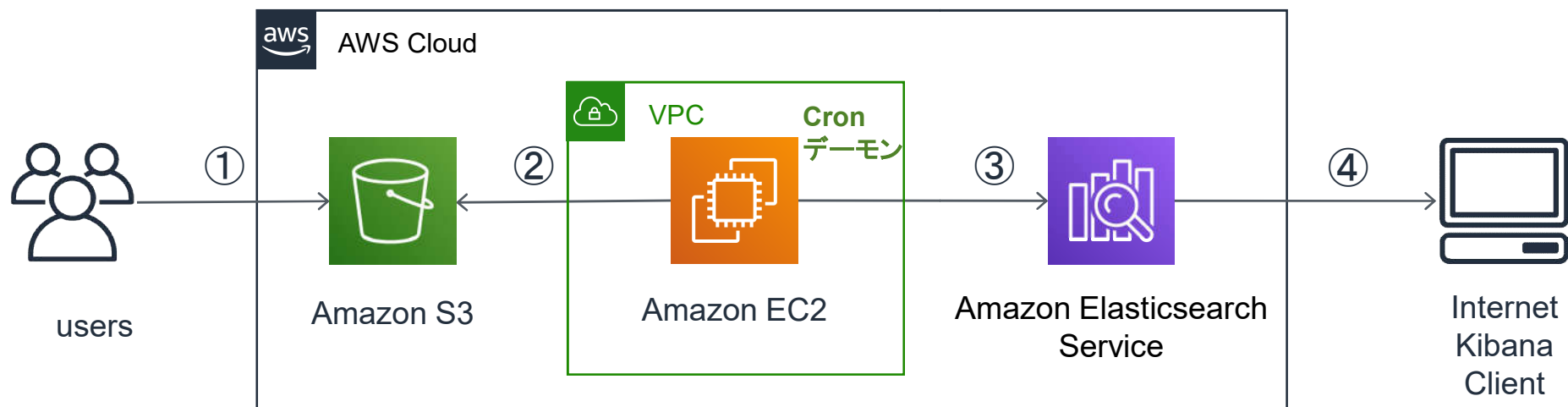


## 2. 提案内容

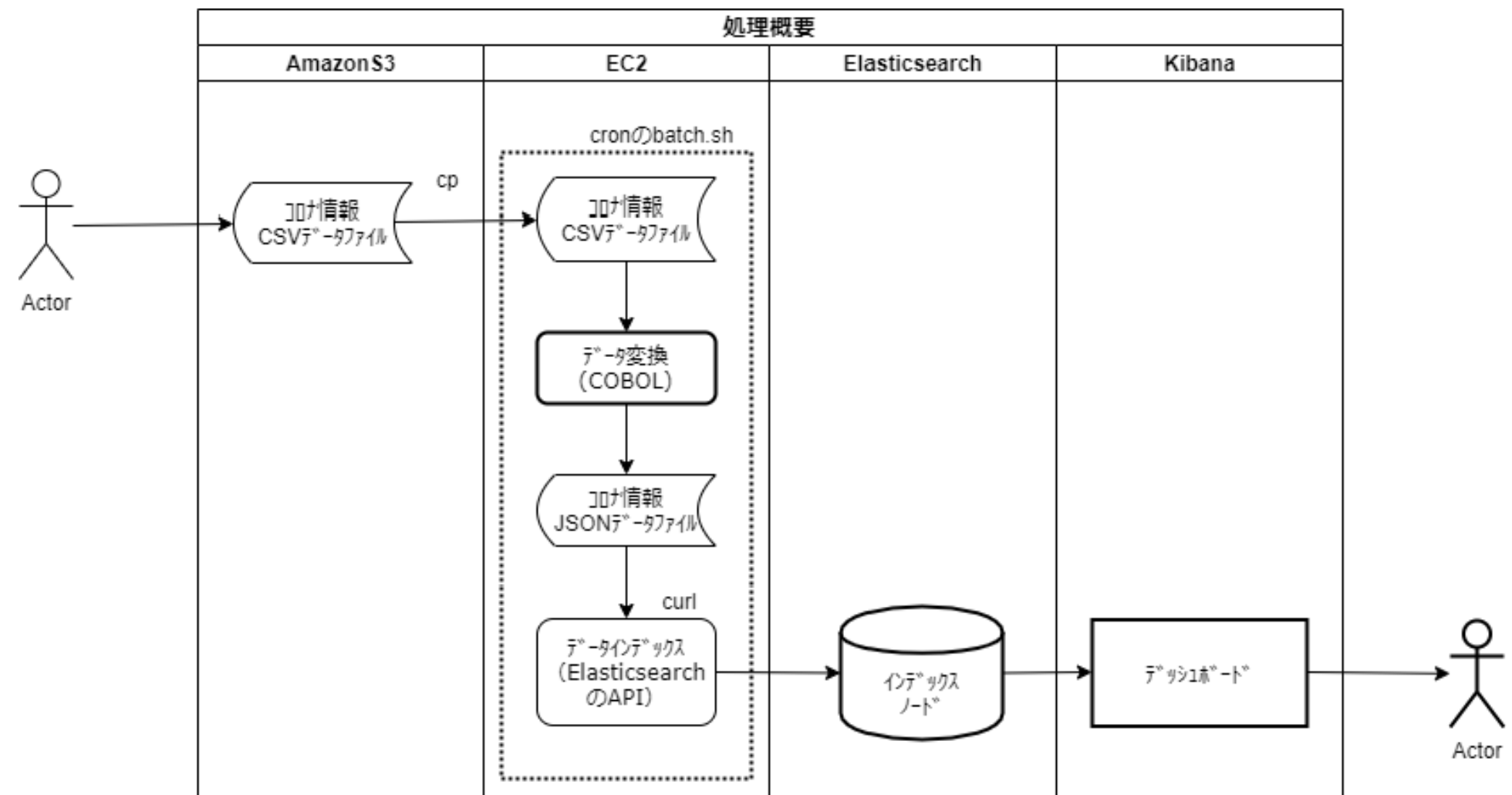
- ① ユーザーCSVデータをAmazon S3にアップロードする。
- ② EC2サーバーに, shellスクリプトを配置し、cronを日次に行う。

shellスクリプトがCSV→JSONフォーマットの変換を行う。

- ③ ②で変換したデータをElasticsearchへ投入する。
- ④ Kibanaを使用してダッシュボードにデータを表示する。

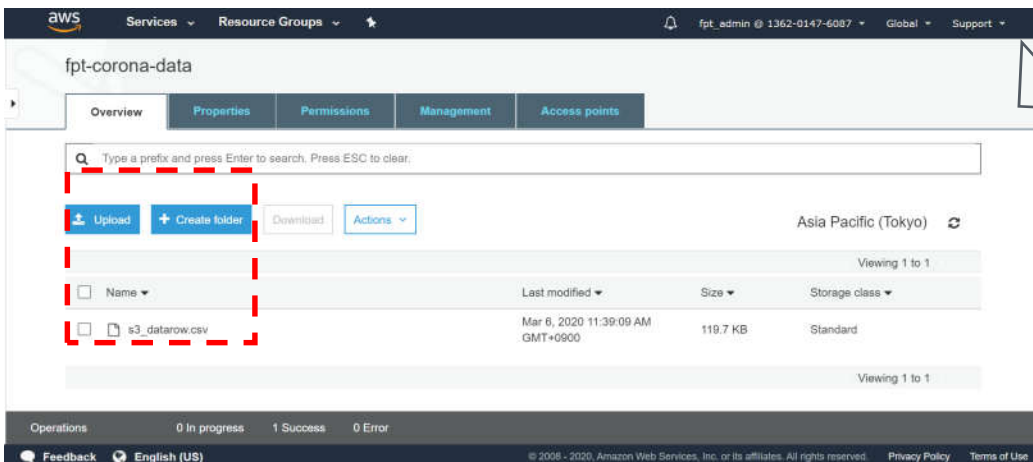
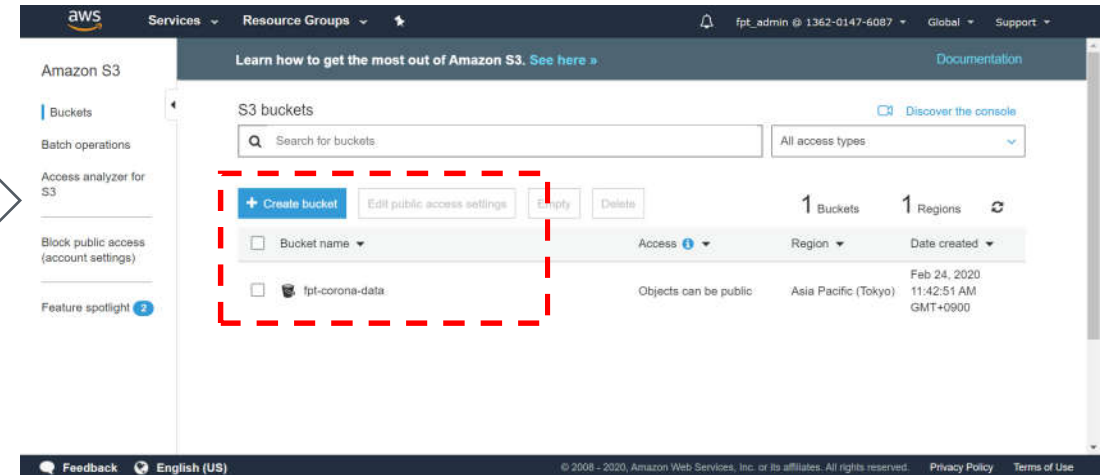
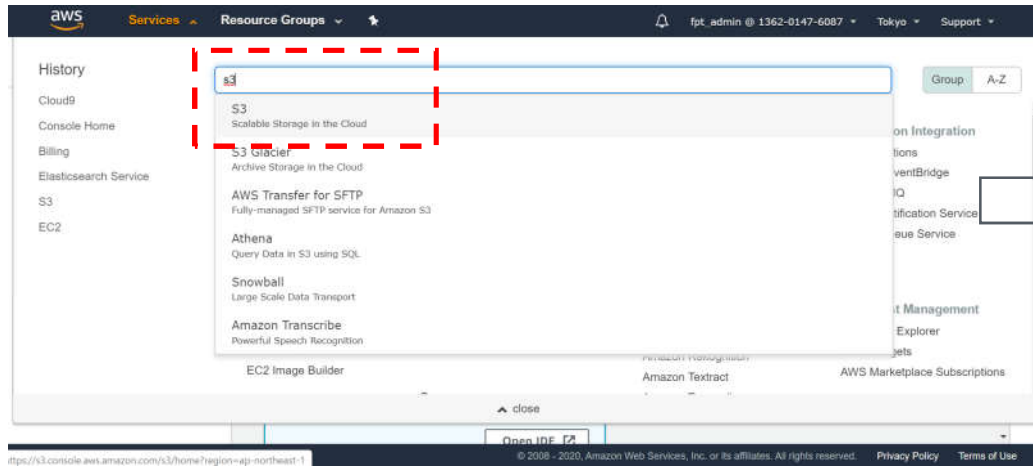


### 3. システム構成図



## 4. AWS S3構成

S3サービスとは安くて高い信頼性、容量が無制限。



今回は、単純にファイルの前処理をS3に格納する。  
最初にbucket( fpt-corona-data )を作成して、パソコンの  
ロカルからデータ( s3\_datarow.csv )をアップロードして、  
システム連携の中間ファイルとして保存しておく。

## 5. AWS EC2構成

EC2サービスでは、バッチサーバーの役割で、定期的にScheduleを起動して、データの変換、Elasticsearchへ投入する。

### 1. Schedule処理

cronのUNIX系の標準的常駐プログラム(デーモン)で、設定したスケジュールに従って指定されたJOBプログラムを起動する。MS WindowのScheduleTaskと同様な機能。

例: 5分1回バッチ処理を実行のCron設定:

```
1  
2 [ec2-user@ ~]$ crontab -l  
3 */5 * * * * /home/ec2-user/corona/cronjob/batch.sh >> /home/ec2-user/corona/cronjob/batch.log 2>&1  
4 [ec2-user@ ~]$
```

上記のcrontabのコマンド作成は: <https://crontab-generator.org/>を使用し、作成してくれた。



## 5. AWS EC2構成(続く)

### 2. JOB

shellスクリプトを使って、S3データコピー、COBOLプログラムでのデータ変換、最後に変換されたデータをElasticsearch投入コマンドの実行順序に一括処理する。

```
1  #!/bin/bash
2
3  PATH="/usr/local/bin:/usr/bin:/bin"
4  S3_DATA_PATH="s3://fpt-corona-data"
5  EC2DATA_PATH="/home/ec2-user/corona/cobol"
6  FDDOMAIN="https://search-corona-5sbkgdsencyyork.ap-northeast-1.es.amazonaws.com"
7
8  # 10 Start job
9  echo "start run corona job: "$(date)
10
11 # 20 Copy data file from S3
12 aws s3 cp $S3_DATA_PATH/s3_datarow.csv $EC2DATA_PATH/ec2_datarow.csv
13
14 # 30 Convert CSV to ES-JSON
15 cd $EC2DATA_PATH
16 cobcrun CZZ1230
17
18 # 40 Elasticsearch index
19 curl -X POST -H "Content-Type: application/x-ndjson" \
20     $FDDOMAIN/_bulk?pretty \
21     --data-binary @$EC2DATA_PATH/ec2_dataconv.json
22
23 # End job
24 echo "end corona job. "$(date)
```

1. 環境変数など定義

2. S3からEC2へデータコピー

3. COBOLのCSVからJSON変換プログラム呼び出し

4. Elasticsearchデータ投入コマンド実行

## 5. AWS EC2構成(続く)

### 3. OpenCobol

メインフレームコボルと比べて、OpenCobolにはTRIM機能のようなユーティリティ機能がたくさん設計されている。

今回は、OpenCobolを使用してcsvファイルを読み取り、Elasticsearchのjsonに変換する。

入力 : ec2\_datarow.csv

Sno	Date	Province/State	Country	Last Update	Confirmed	Deaths	Recovered
1	2020-01-22T12:00:00	Anhui	China	2020-01-22T12:00:00	1	0	0
2	2020-01-22T12:00:00	Beijing	China	2020-01-22T12:00:00	14	0	0
3	2020-01-22T12:00:00	Chongqing	China	2020-01-22T12:00:00	6	0	0



Output: ec2\_dataconv.json

```
1 { "index" :  
2   {  
3     "_index" : "corona-index"  
4   }  
5 }  
6 {  
7   "Date": "2020-01-22T12:00:00",  
8   "Province": "Anhui",  
9   "Country": "China",  
10  "LastUpdate": "2020-01-22T12:00:00",  
11  "Confirmed": 1,  
12  "Deaths": 0,  
13  "Recovered": 0  
14 }
```

Source code: <https://github.com/pkm0306pkm/corona>

## 6. AWS Elasticsearch

Amazon Elasticsearch Service は無料のオープンソースのソフトウェアで大規模なデータコレクションをインデックス化、管理、検索、視覚化できます。

1. Index: ドキュメントのコレクションであり。リレーショナルデータベース(RDS)のデータベースと同様。
2. Document: JSON形式で定義されたフィールドのコレクションであり。

Elasticsearch indexのマッピング定義:

```
1  curl -X PUT -H "Content-Type: application/json" /corona-index/_index -d '{
2  {
3      "mappings":{
4          "properties":{
5              "Date":{ "type":"date"},
6              "Country": { "type":"text"},
7              "Province": { "type":"text"},
8              "LastUpdate":{ "type":"date"},
9              "Confirmed":{ "type":"double"},
10             "Deaths":{ "type":"double"},
11             "Recovered":{ "type":"double"}
12         }
13     }
14 }'
```

1. index定義

2.フィールド定義

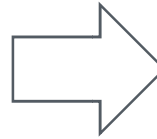
## 6. AWS Elasticsearch (操作)

### 1. データ一括投入API:

```
18 # 40 Elasticsearch index
19 curl -X POST -H "Content-Type: application/x-ndjson" \
20     $FDDOMAIN/_bulk?pretty \
21     --data-binary @$EC2DATA_PATH/ec2_dataconv.json
22
```

### 2. データ検索API:

```
26 curl -X POST -H "Content-Type: application/json" \
27     $FDDOMAIN/corona-index/_search?pretty -d '
28 {
29     "query":{
30         "query_string":{
31             "query":"China"
32         }
33     },
34     "size": 1
35 }'
```



```
46 {
47     "hits" : {
48         "total" : {
49             "value" : 835,
50             "relation" : "eq"
51         },
52         "max_score" : 0.9140898,
53         "hits" : [
54             {
55                 "_index" : "corona-index",
56                 "_type" : "_doc",
57                 "_id" : "LJP5m3ABtQIsqp-p28_f",
58                 "_score" : 0.9140898,
59                 "_source" : {
60                     "Date" : "2020-01-22T12:00:00",
61                     "Province" : "Guangdong",
62                     "Country" : "China",
63                     "Confirmed" : 26,
64                     "Deaths" : 0,
65                     "Recovered" : 0
66                 }
67             }
68         ]
69     }
70 }
```

### 3. その他API(create, update, delete, count...):

<https://www.elastic.co/guide/en/elasticsearch/reference/current/docs.html>



## 7. Kibana

KibanaはElasticsearch で稼働して、以下の手順にデータを可視化する。

1. Managementタブ: Elasticsearch のデータを操作するため、index patternを作成する。
2. Visualizeタブ: インデックスからデータの視覚化を作成する。
3. Dashboardタブ: リアルタイムでの視覚化のグラフ、およびマップを提供する。



## 8.デモ



## 9. 参考リンク

### 1. Cron設定

<https://www.express.nec.co.jp/linux/distributions/knowledge/system/crond.html>

### 2. ElasticSearch設定

<https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-gsg.html>

### 3. Kibana使用

<https://www.elastic.co/guide/en/kibana/current/tutorial-visualizing.html>

### 4. FileBeat, LogStack, Elasticsearch, Kibana

<https://medium.com/@quoeamaster/use-case-coronavirus-covid-19-investigation-and-visualisation-with-elasticsearch-598df7849e82>

### 5. Cloud9 設定

[https://docs.aws.amazon.com/ja\\_jp/cloud9/latest/user-guide/setting-up.html](https://docs.aws.amazon.com/ja_jp/cloud9/latest/user-guide/setting-up.html)

### 6. Corona-19データ: <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset/data>



ご清聴ありがとうございました

