



Amazon Data Firehoseの 入門から実戦まで

2024/11/20 九州AWS若手の会#2

おのやん

- 1.自己紹介
- 2.Amazon Data Firehoseとは？
- 3.実際のユースケース
- 4.IaCテンプレート例
- 5.注意点いろいろ
- 6.まとめ

自己紹介

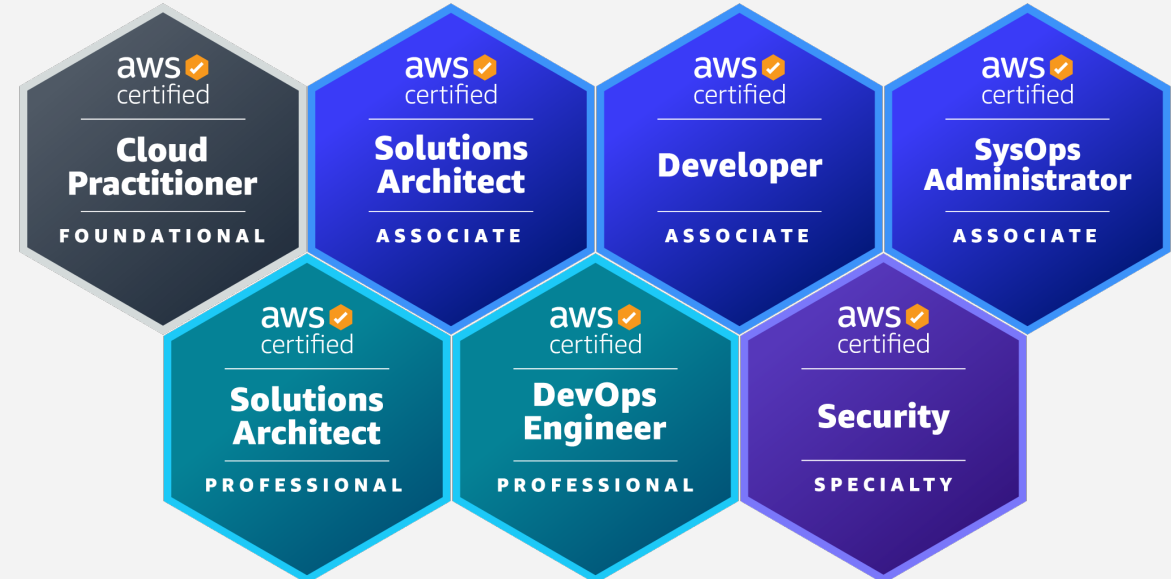


- 部署
 - AWS事業本部 コンサルティング部
- 名前（ニックネーム）
 - 小野山 翔大 (おのやん)
- 出身・住まい
 - 福岡
- 趣味
 - ランニング・バドミントン
 - 特撮

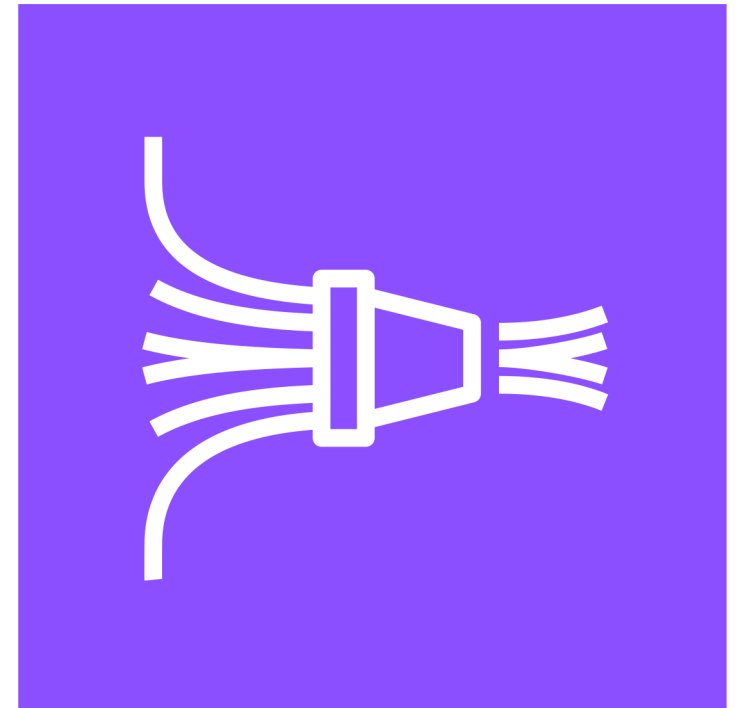
- 2023年4月 クラスメソッド 新卒入社

- 2024年7月 2024 AWS Jr.Champion

- お客様のAWS環境構築を技術的に支援




- **AWS フルマネージドサービス**
 - ストリーミングデータを収集・変換
 - 他のAWS サービスなど、
分析用データストアに配信可能
- 旧 Amazon Kinesis Data Firehose
- ストリーミングデータを
S3バケット（バケツ）に流す
消火ホース（Firehose）のイメージ



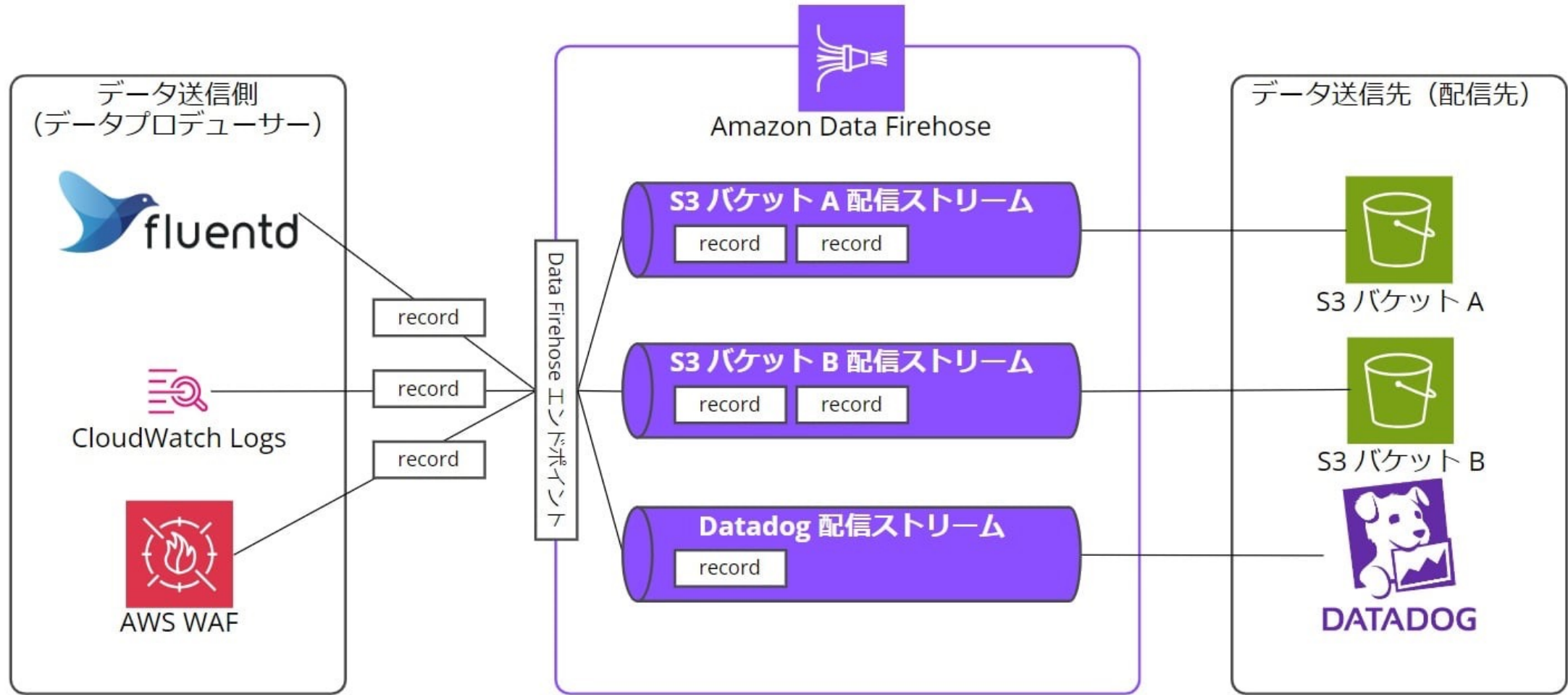
「firehose」の検索結果

サービス

-  **Amazon Data Firehose** ☆
低コストで、あらゆるデータ、あらゆる規模のリアルタイムストリーミング ETL
-  **Kinesis** ☆
リアルタイムストリーミングデータとの連携
-  **AWS End User Messaging** ☆
複数のコミュニケーションチャンネルにわたって顧客を引き付ける

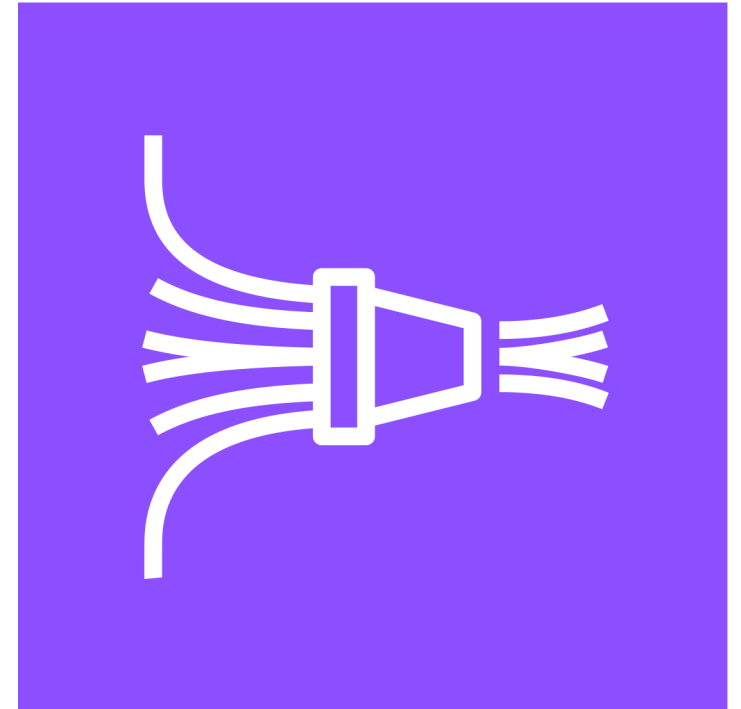
コンソールで新アイコンが適用されている 数少ないサービス（当方観測範囲内）

Amazon Data Firehoseとは

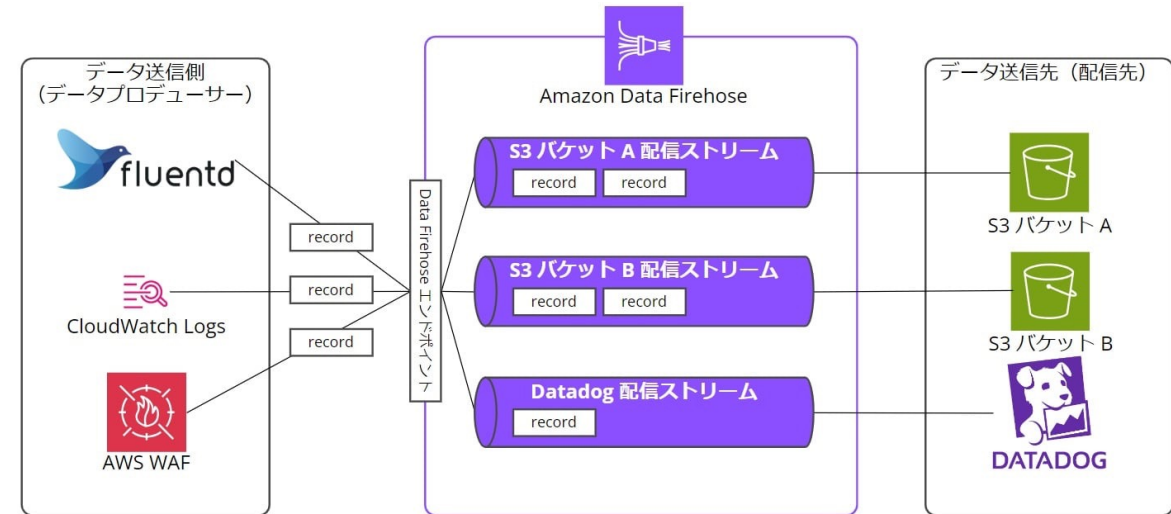


<https://dev.classmethod.jp/articles/introduction-2024-amazon-data-firehose/>

- **1つのストリーム**を構成単位とする
 - 「Data Firehose ストリームを作成」
- **バッファサイズとバッファ間隔**をストリームごとに設定
 - バッファサイズ：特定のサイズに達したらデータストアに保存
 - バッファ間隔：特定の期間が経過したらデータストアに保存



- ストリームデータをデータストアに保存する際のつなぎ
 - CloudWatchログをS3に保存したい！
 - 外部SaaSのログをS3に保存したい！
- バッファリングのサイズや時間について実装が不必要に
 - シンプルな分、保存ファイル名の変更が効かないなどの制約はある



<https://dev.classmethod.jp/articles/introduction-2024-amazon-data-firehose/>


```
1 #=====#
2 # Data Firehose
3 #=====#
4
5 AWSTestFirehoseStream:
6   Type: AWS::KinesisFirehose::DeliveryStream
7   Properties:
8     DeliveryStreamName: !Sub ${SystemName}-${EnvName}-firehose-stream
9     DeliveryStreamType: DirectPut
10    ExtendedS3DestinationConfiguration:
11      BucketARN: !GetAtt AWSTestBucket.Arn
12      BufferingHints:
13        IntervalInSeconds: 60
14        SizeInMBs: 5
15      CompressionFormat: GZIP
16      CustomTimeZone: Asia/Tokyo
17      RoleARN: !GetAtt AWSTestFirehoseRole.Arn
18      Prefix: !Sub "${timestamp:yyyy}/${timestamp:MM}/${timestamp:dd}/"
19      ErrorOutputPrefix: !Sub errors/${timestamp:yyyy}/${timestamp:MM}/${timestamp:dd}/${firehose:error-output-type}/
20      ProcessingConfiguration:
21        Enabled: true
22        Processors:
23          - Type: AppendDelimiterToRecord
24            Parameters:
25              - ParameterName: Delimiter
26                ParameterValue: "\\n"
27      CloudWatchLoggingOptions:
28        Enabled: true
29        LogGroupName: !Ref AWSTestFirehoseCloudWatchLogGroup
30        LogStreamName: !Ref AWSTestFirehoseCloudWatchLogStream
31
32 AWSTestFirehoseCloudWatchLogGroup:
33   Type: AWS::Logs::LogGroup
34   Properties:
35     LogGroupName: !Sub /aws/kinesisfirehose/${SystemName}-${EnvName}-firehose-stream
36     RetentionInDays: 7
```

IaC経由で構築時：Data Firehose ログストリームの作成が必要

- CloudWatch ロググループだけだと構築に失敗する
- CloudWatch ログストリームも明示的に作成が必要

```
#=====#
# CloudWatch ロググループ
#=====#

AWSTestFirehoseCloudWatchLogGroup:
  Type: AWS::Logs::LogGroup
  Properties:
    LogGroupName: !Sub /aws/kinesisfirehose/${SystemName}-${EnvName}-firehose-stream
    RetentionInDays: 7

#=====#
# CloudWatch ログストリーム
#=====#

AWSTestFirehoseCloudWatchLogStream:
  Type: AWS::Logs::LogStream
  Properties:
    LogGroupName: !Ref AWSTestFirehoseCloudWatchLogGroup
    LogStreamName: !Sub ${SystemName}-${EnvName}-firehose-stream-logstream
```

Data Firehose => S3で保存したGZIP圧縮ファイル：

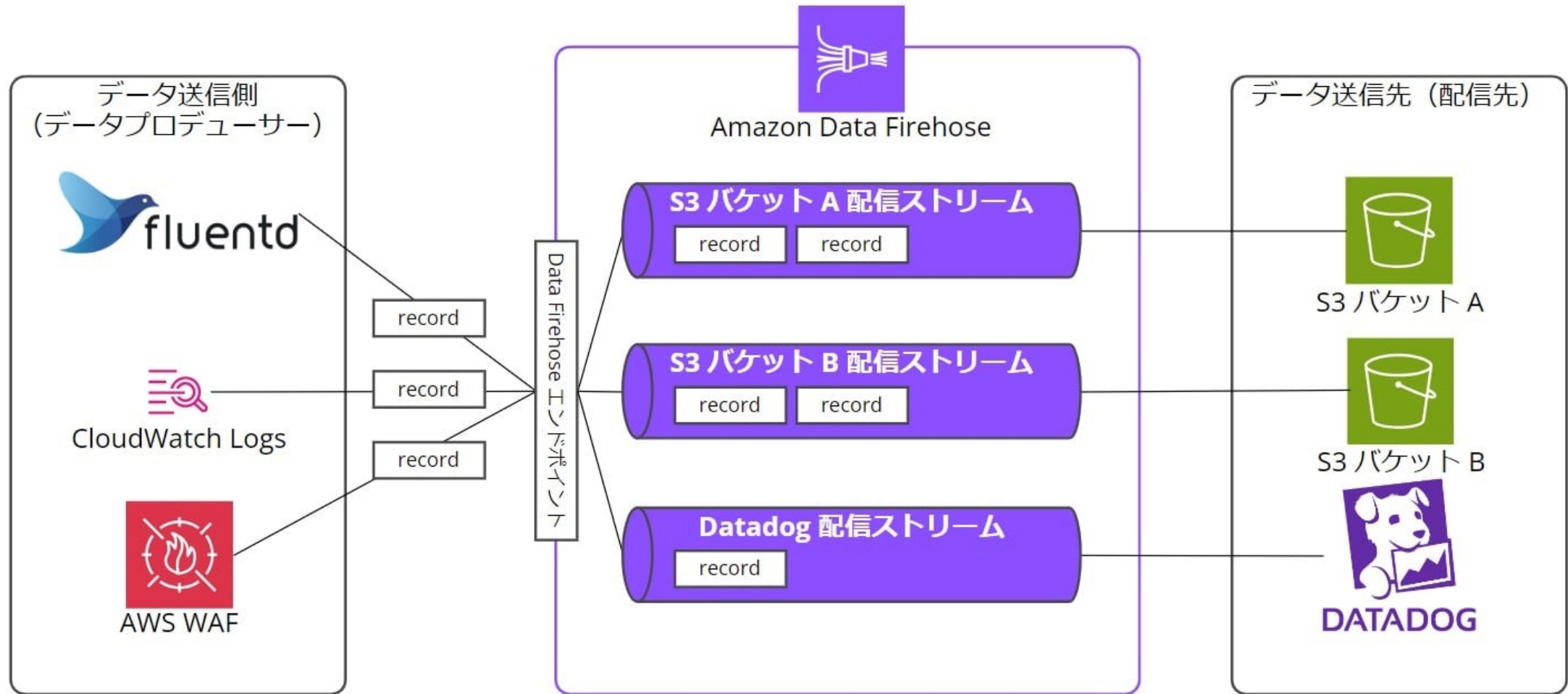
ブラウザ経由でDLすると勝手に解凍されるが、ファイルの拡張子は変わらない

- PCでファイルの内容を確認するときに引っかけがち



```
onoyama.shodai@ ~ % cd Downloads
onoyama.shodai@ Downloads % file aws-test-sample-log.gz
aws-test-sample-log.gz: JSON data
```

Amazon Data Firehoseとは：おかわり



<https://dev.classmethod.jp/articles/introduction-2024-amazon-data-firehose/>

ファイル名や圧縮形式にそこまで制約がなく
ストリーミングデータをデータストアへシンプルに配信したいなら

Data Firehoseは大いに採用の価値あり！！

classmethod