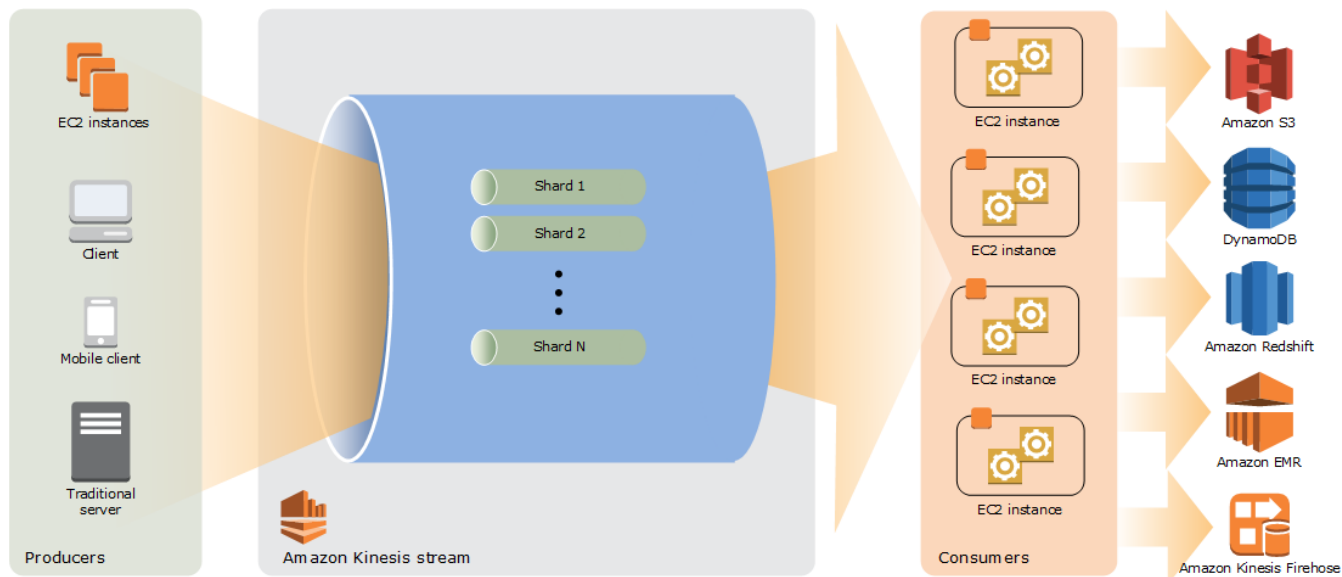


학습열기

학습열기



https://docs.aws.amazon.com/ko_kr/streams/latest/dev/key-concepts.html

나레이션

이 그림은 Kinesis Data Streams의 high level 아키텍처입니다. 생산자는 지속적으로 Kinesis Data Streams에 데이터를 푸시하며, 소비자는 실시간으로 데이터를 처리합니다. 소비자는(Amazon EC2 application 또는 Amazon Kinesis Data Firehose)는 Amazon DynamoDB, Amazon Redshift 또는 Amazon S3 등의 AWS Service를 사용하여 결과를 저장할 수 있습니다.

AWS EMR(Elastic Map Reduce)

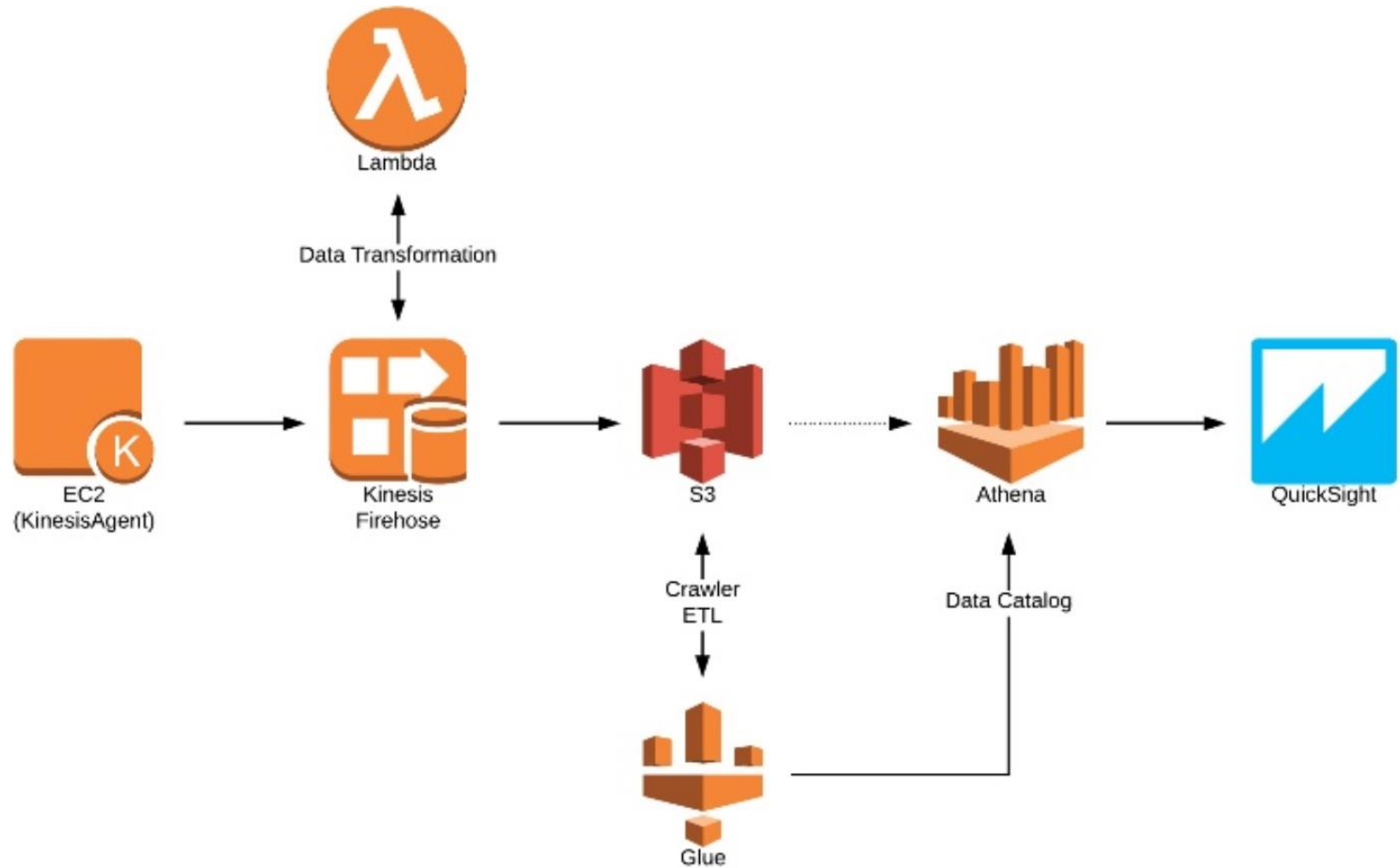
AWS 빅데이터 분석 Service



출처 : 아마존 웹 서비스(Amazon Web Service)

AWS 도구

AWS 빅데이터 Demo 아키텍처



Kinesis Firehose

Amazon Kinesis Data Firehose

완전 관리형 Service, 서버리스
스트리밍 데이터를 손쉽게 캡처
S3, Redshift, ES 등으로 데이터 업로드
Lambda를 이용한 데이터 전처리
데이터 처리량에 대응한 자동 확장



데이터 수집 단계

서버에서 발생하는 데이터를 실시간 수집
Kinesis Data Firehose를 이용하여 클릭스트림 데이터를 S3에 수집
Kinesis Agent로 데이터를 수집해 Kinesis Data Firehose로 전송
Lambda를 이용한 전처리

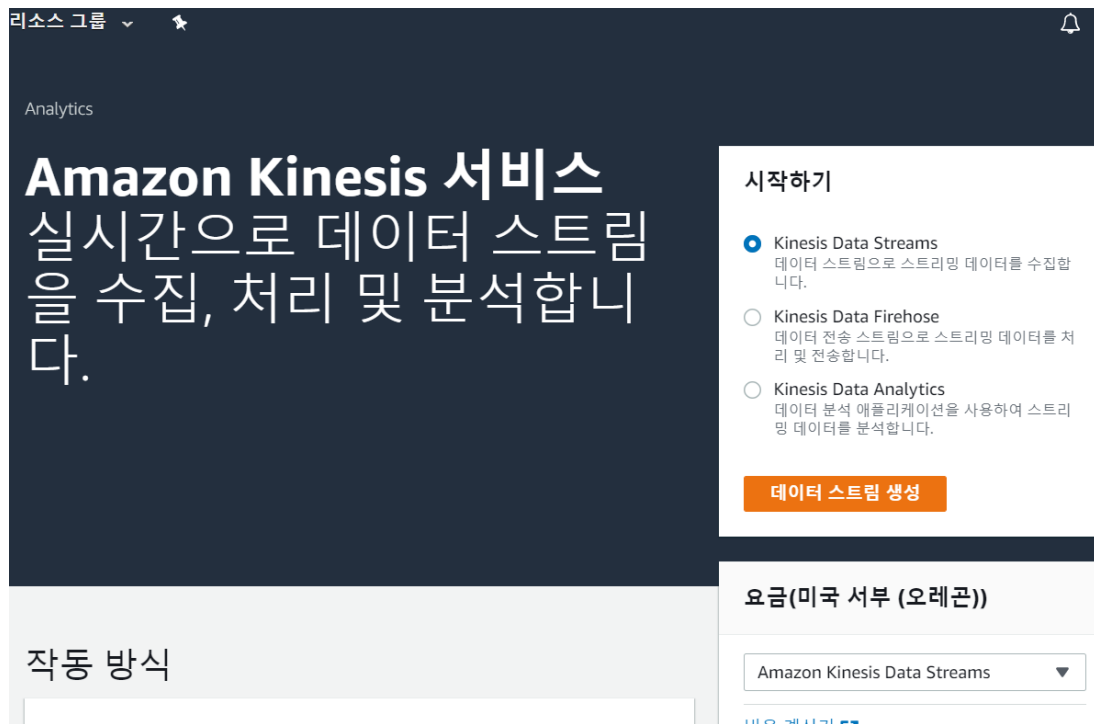
Kinesis Firehose

Kinesis를 사용하기 위해 리소스 선택 가능

먼저 Data streaming 저장용으로 사용할 Firehose를 선택하여 스트림을 생성하고 적재

Firehose는 스트리밍 데이터를 데이터 레이크 혹은 스토어/분석도구에 쉽고 안정적으로 로드 가능

Firehose로 전송 후 60초내 S3/Redshift/Elasticsearch/splunk로 로드 가능



Kinesis Firehose

Kinesis Data Firehose - Create delivery stream

Step 1: Name and source

Step 2: Process records

Step 3: Choose a destination

Step 4: Configure settings

Step 5: Review

New delivery stream

Delivery streams load data, automatically and continuously, to the destinations that you specify. Kinesis Data Firehose resources are not covered under the [AWS Free Tier](#), and **usage-based charges apply**. For more information, see [Kinesis Data Firehose pricing](#). [Learn more](#)

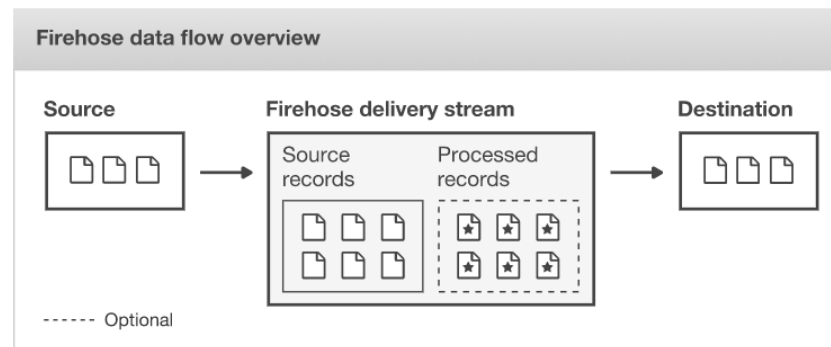
Delivery stream name

Enter a delivery stream name

Acceptable characters are uppercase and lowercase letters, numbers, underscores, hyphens, and periods.

Choose a source

Choose how you would prefer to send records to the delivery stream.



Kinesis Firehose

Data source 선택 (직접 PUT 호출 또는 Kinesis Data Stream 사용)

Source

To learn about enabling server-side encryption (SSE), see [Data Protection in Amazon Kinesis Data Firehose](#).

- ☒ Direct PUT or other sources
Choose this option to send records directly to the delivery stream, or to send records from AWS IoT, CloudWatch Logs, or CloudWatch Events.
- ☐ Kinesis Data Stream

Server-side encryption for source records in the delivery stream

You can use AWS Key Management Service (KMS) to create and manage customer keys (CMKs) and to control the use of encryption across a wide range of AWS services and in your applications.

- ☐ Enable server-side encryption for source records in delivery stream

▼ How to send source records to Kinesis Data Firehose

After creating the delivery stream, send source records using the Kinesis Data Firehose PUT API or the Amazon Kinesis Agent.

Kinesis Data Firehose PUT APIs

Use the Kinesis Data Firehose `PutRecord()` or `PutRecordBatch()` API to send source records to the delivery stream. [Learn more](#)

Amazon Kinesis Agent

The Amazon Kinesis Agent is a stand-alone Java software application that offers an easy way to collect and send source records to Kinesis Data Firehose. [Learn more](#)

AWS IoT

Create AWS IoT rules that send data from MQTT messages. [Learn more](#)

CloudWatch Logs

Use subscription filters to deliver a real-time stream of log events. [Learn more](#)

Kinesis Firehose

샘플 프로젝트 다운로드

```
git clone https://github.com/aws-labs/aws-nodejs-sample.git
```

SDK 및 종속성 설치

aws-nodejs-sample 디렉터리에서 다음을 실행합니다.

```
npm install
```

액세스 키 구성

~/.aws/credentials(Windows 사용자의 경우

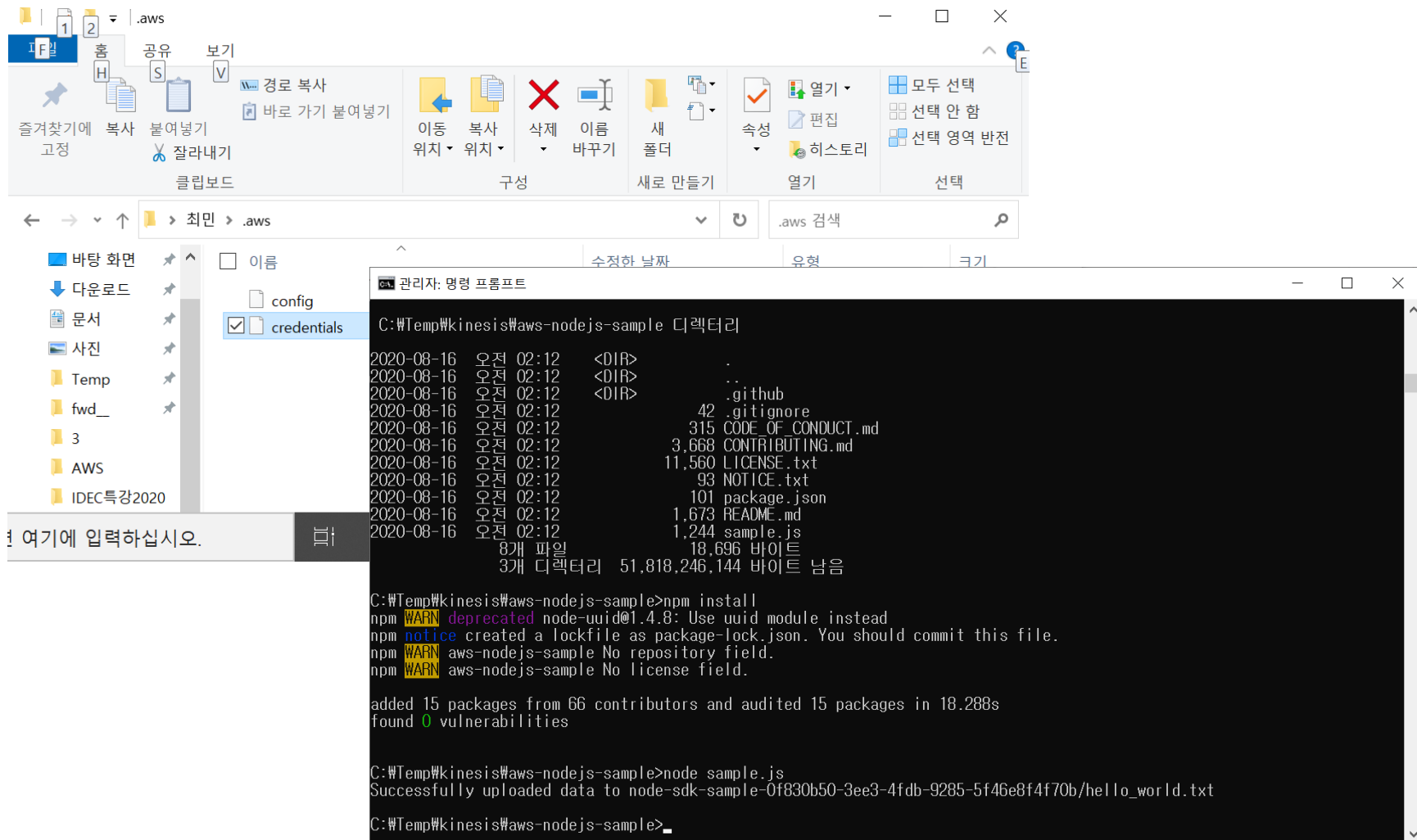
C:\Users\USER_NAME\.aws\credentials)에 자격 증명 파일을 생성하고
밑줄 친 값을 자체 값으로 대체한 후 다음 라인을 저장합니다.

```
aws_access_key_id = YOUR_ACCESS_KEY_ID
```

```
aws_secret_access_key = YOUR_SECRET_ACCESS_KEY
```


Kinesis Firehose

Kinesis Firehost PUT API 사용



The screenshot shows a Windows File Explorer window with the address bar set to `.aws`. The left sidebar shows the navigation pane with `바탕 화면`, `다운로드`, `문서`, `사진`, `Temp`, `fwd_`, `3`, `AWS`, and `IDEC특강2020`. The main pane shows the `credentials` file selected. Below the File Explorer, there is a text input field with the placeholder text "여기에 입력하십시오." and a button with a magnifying glass icon.

Overlaid on the bottom right is a terminal window titled "관리자: 명령 프롬프트". The terminal shows the following commands and output:

```
C:\Temp\kinesis\aws-nodejs-sample 디렉터리
2020-08-16 오전 02:12 <DIR> .
2020-08-16 오전 02:12 <DIR> ..
2020-08-16 오전 02:12 <DIR> .github
2020-08-16 오전 02:12 42 .gitignore
2020-08-16 오전 02:12 315 CODE_OF_CONDUCT.md
2020-08-16 오전 02:12 3,668 CONTRIBUTING.md
2020-08-16 오전 02:12 11,560 LICENSE.txt
2020-08-16 오전 02:12 93 NOTICE.txt
2020-08-16 오전 02:12 101 package.json
2020-08-16 오전 02:12 1,673 README.md
2020-08-16 오전 02:12 1,244 sample.js
2020-08-16 오전 02:12 8개 파일 18,696 바이트
3개 디렉터리 51,818,246,144 바이트 남음

C:\Temp\kinesis\aws-nodejs-sample>npm install
npm WARN deprecated node-uuid@1.4.8: Use uuid module instead
npm notice created a lockfile as package-lock.json. You should commit this file.
npm WARN aws-nodejs-sample No repository field.
npm WARN aws-nodejs-sample No license field.

added 15 packages from 66 contributors and audited 15 packages in 18.288s
found 0 vulnerabilities

C:\Temp\kinesis\aws-nodejs-sample>node sample.js
Successfully uploaded data to node-sdk-sample-0f830b50-3ee3-4fdb-9285-5f46e8f4f70b/hello_world.txt

C:\Temp\kinesis\aws-nodejs-sample>
```

Kinesis Firehose

Kinesis Firehose PUT API 사용

Lambda Service 사용 여부 선택

레코드 포맷 변환 여부 선택

Transform source records with AWS Lambda

To return records from AWS Lambda to Kinesis Data Firehose after transformation, the Lambda function you invoke must be compliant with the required record transformation output model. [Learn more](#)

Data transformation

- ☒ Disabled
☐ Enabled

Convert record format

Data in Apache Parquet or Apache ORC format is typically more efficient to query than JSON. Kinesis Data Firehose can convert your JSON-formatted source records using a schema from a table defined in [AWS Glue](#). For records that aren't in JSON format, create a Lambda function that converts them to JSON in the [Transform source records with AWS Lambda](#) section above. [Learn more](#)

Record format conversion

- ☒ Disabled
☐ Enabled

If record format conversion is enabled, Kinesis Data Firehose can deliver data to Amazon S3 only. Record format conversion will be configured using the OpenX JSON SerDe. For other options use the [AWS CLI](#).

[Cancel](#)

[Previous](#)

[Next](#)

Kinesis Firehose

Kinesis Firehose PUT API 사용

PUT API를 통해 취득된 데이터의 저장소 선택 S3 버킷 정보 설정

Choose a destination

[Learn more](#)

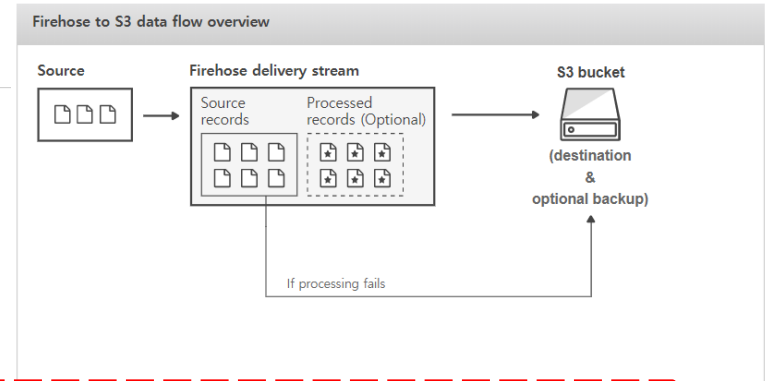
Destination

- ☒ Amazon S3
Object storage built to store and retrieve any amount of data from anywhere.
- ☐ Amazon Redshift
An enterprise-level, petabyte scale, fully managed data warehousing service.
- ☐ Amazon Elasticsearch
An open-source search and analytics engine for use cases such as log analytics, real-time application monitoring, and click stream analytics.
- ☐ HTTP Endpoint
A way to deliver data to your custom destination.
- ☐ Third-party service provider
Choose from a list of third-party service providers.

S3 destination

Choose a destination in Amazon S3 where your data will be stored. Amazon S3 is object storage built to store and retrieve any amount of data from anywhere.

[Learn more](#)



S3 bucket

node-sdk-sample-0f830b50-3ee3-4fdb-9285-5f46e8f4f70b



Create new

in S3 console

node-sdk-sample-0f830b50-3ee3-4fdb-9285-5f46e8f4f70b

photo-collection2

By default, Kinesis Data Firehose appends the prefix "YYYY/MM/dd/HH" (in UTC) to the data it delivers to Amazon S3. You can override this default by specifying a custom prefix that includes expressions that are evaluated at runtime.

If your custom prefix doesn't include expressions, Kinesis Data Firehose uses your prefix and appends "YYYY/MM/dd/HH". If your custom prefix includes a Kinesis Data Firehose random string or timestamp expression, Kinesis Data Firehose doesn't append "YYYY/MM/dd/HH". [Learn more](#)

Backup S3 bucket prefix - optional

Enter a prefix


S3 error prefix

Kinesis Firehose


Kinesis Firehose PUT API 사용

데이터 압축이나 암호화 관련 설정, IAM 권한 설정

Configure settings

Configure buffer, compression, logging, and IAM role settings for your delivery stream. [Learn more](#) 

S3 buffer conditions

Kinesis Data Firehose buffers incoming records before delivering them to your S3 bucket. Record delivery will be triggered once either of these conditions has been satisfied. [Learn more](#) 

Buffer size

 MiB


Enter a buffer size between 1 - 128 MiB.

Buffer interval

 seconds

Enter a buffer interval between 60 - 900 seconds.

S3 compression and encryption

Kinesis Data Firehose can compress records before delivering them to your S3 bucket. Compressed records can also KMS master key. [Learn more](#) 

S3 compression

- ☒ Disabled
- ☐ GZIP
- ☐ Snappy
- ☐ Zip
- ☐ Hadoop-Compatible Snappy

S3 encryption

- ☒ Disabled
- ☐ Enabled

Tags - optional

You can add tags to organize your AWS resources, track costs, and control access. [Learn more](#) 

Key

Value - optional


Remove tag

Add tag

You can add 49 more tag(s)

Permissions

IAM role

Kinesis Data Firehose uses this IAM role for all the permissions that the delivery stream needs. To specify different roles for the different permissions, use the API or the CLI. [Learn more](#) 

- ☒ Create or update IAM role **KinesisFirehoseServiceRole-test1-ap-northeast-2-1597518238505**
This creates the role or updates it if it already exists, adds the required policies to it, and enables Kinesis Data Firehose to assume it.
- ☐ Choose existing IAM role
The role that you choose must have policies that include the permissions that Kinesis Data Firehose needs.

Cancel

Previous

Next

AWS Glue

AWS Glue

AWS Glue는 완전관리형 ETL(추출, 변환, 로드) Service로, 효율적인 비용으로 간단하게 원하는 데이터를 분류, 정리, 보강하고 다양한 데이터 스토어 간에 이동

