**KINESIS**

Kinesis is a aws service which makes it easy to sollect,process and analyze streaming data in real time.It works on real time data such as Al=pplication logsmetrics,website clickstreams and iot telemetry data.

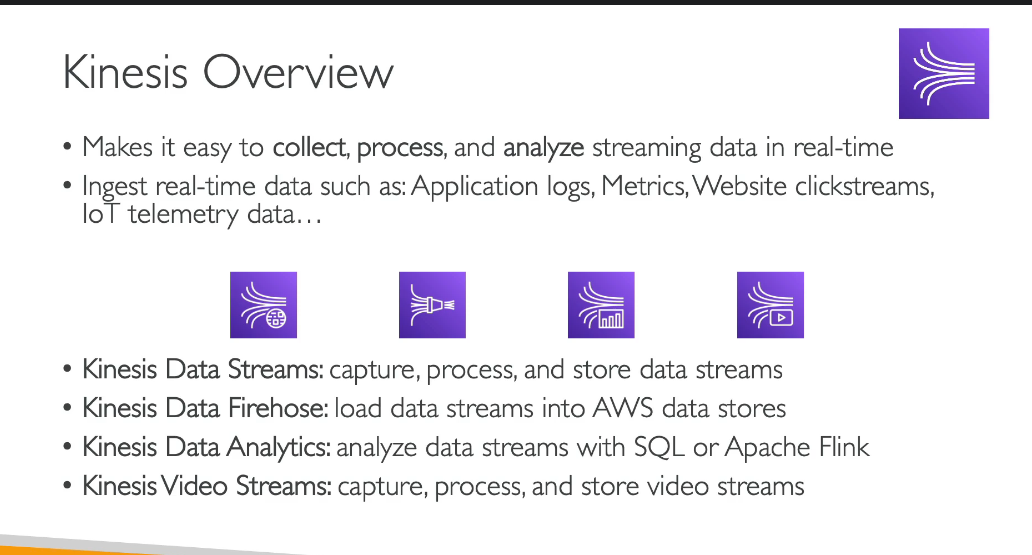
There are four kinds of kinese steams as can be seen below:

1.Kinesis data stream: whichc is used to capture,process and store data streams.

2.Kinesis Data Firehose:which is used to load data streams into Aws data stores

3.Kinesis Data analytics:Whichc is used to analyze data streams with sql and apache flink

4.Kinesisi video streams:which is used to capture, process and store video streams.



**KINESIS DATA STREAM**

Kinesis data stream is a way for you to stream big data into your systems.

Kinesis data stream is made up of many shards.Shard is something that we need to provision ahead of time.

So when we start with kinesis data stream we start with saying I want 6 shards so our data will be distributed across 6 shards

Shards are going to define your capacity in terms of ingestion or consumption

**Producers:**they send data into kinesis data streams.Producers can be applications,client,sdk,kpl,kinesis agent.All producers are going to produce records into our kinesis data stream.

**RECORD:**a record consists of a partition key which tells which shard the data must go to and the blob is the value itself.

**CONSUMER:**once a data is in kinesis data stream it can be consumed by many consumers.

They can be apps,lambda,kinesis data firehose or kinesis data analytics.

Consumer receives a record which tells the partition key and a sequence no which tells which shard the datav was in and the blob.

Diagram

Description automatically generated

**SOME PROPERTIES OF KINESIS DATA STREAMS**

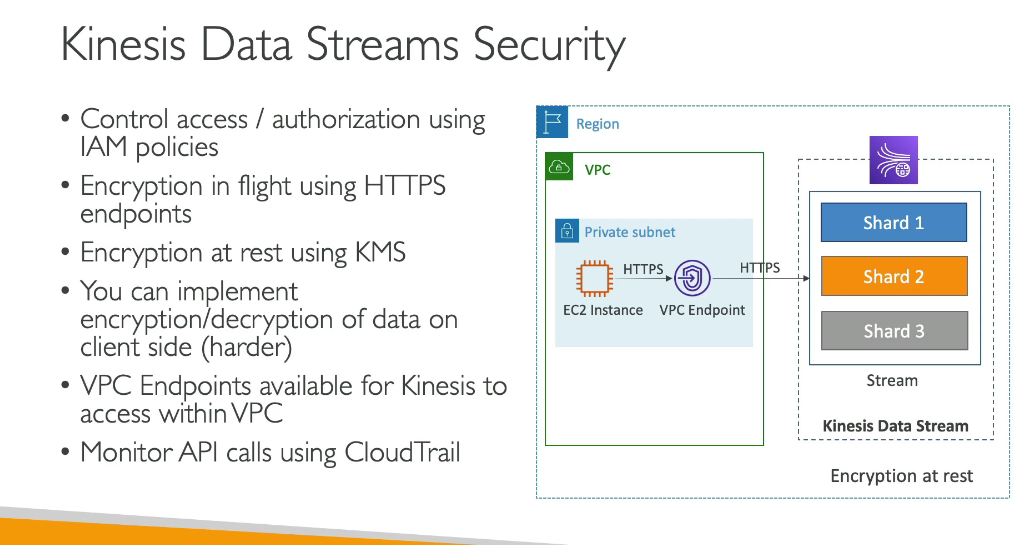
**Text, letter

Description automatically generated**

KINESIS **DATA STREAMS-CAPACITY MODES**

**Text

Description automatically generated**

****

**KINESIS PRODUCERS**

A kinesis producer will put data records into the data streams.The api used to put record into the kinesis data stream is though a PutRecord API.We can use batching with putrecords api to reduce the cost and increase the throughput.

Text, letter

Description automatically generated

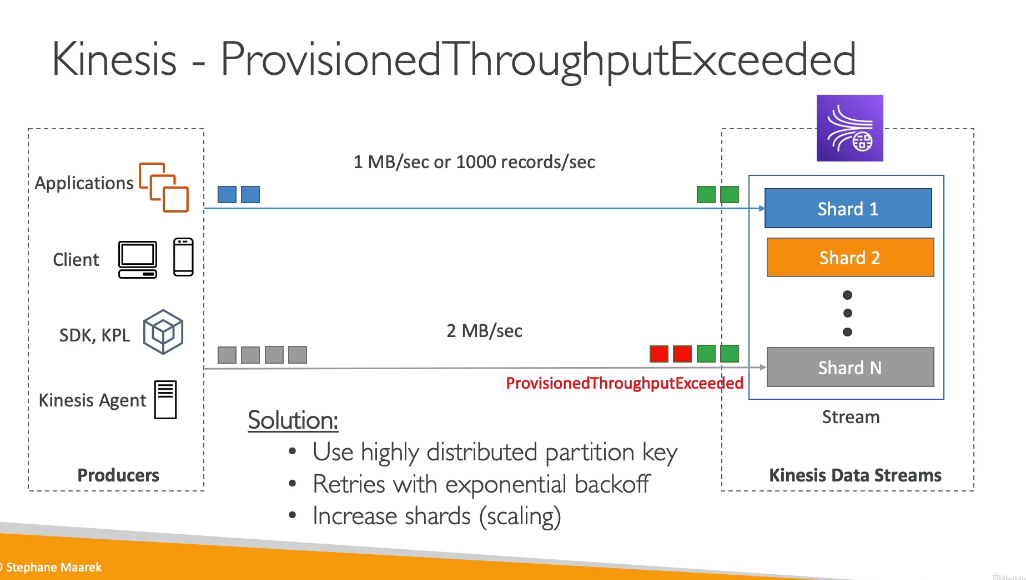
**HOW KINESIS PRODU CERS WORK**

**Diagram

Description automatically generated**

We have a device and each has a device id which will go through a hash function and will be distributed to a shard function.

We need to keep in mind that we need to have a highly distributed partition of the device ids so that we avoid the hot partitioning in this case.



**KINESIS CONSUMERS**

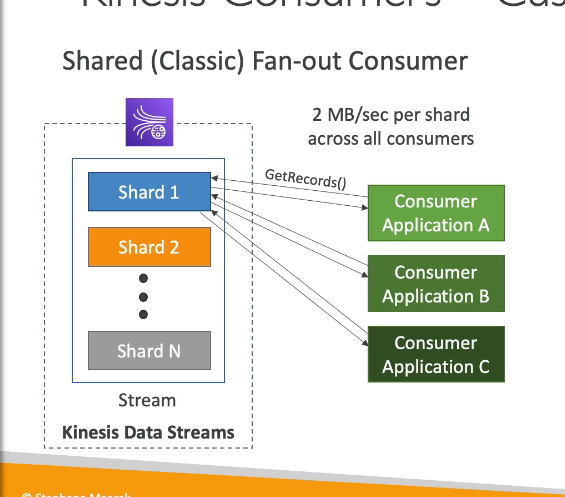
Consumers take data from the kinesis data streams.

Graphical user interface, text, application, email

Description automatically generated

The consumers can be aws lambda,kinesis data analytics,kinesis data firehose,custom consumer,kinesis client library.

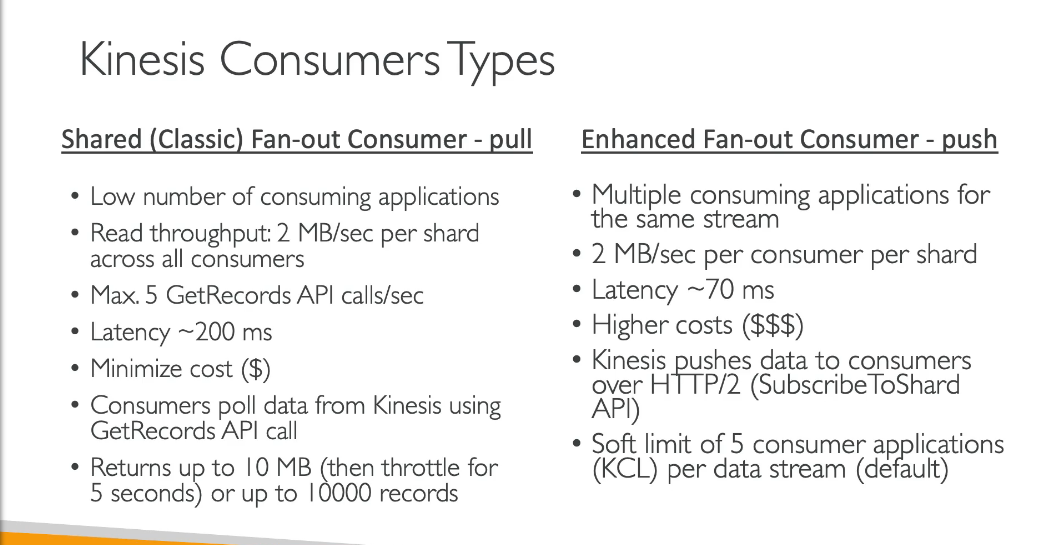
**Shared(classic) fan out consumer**

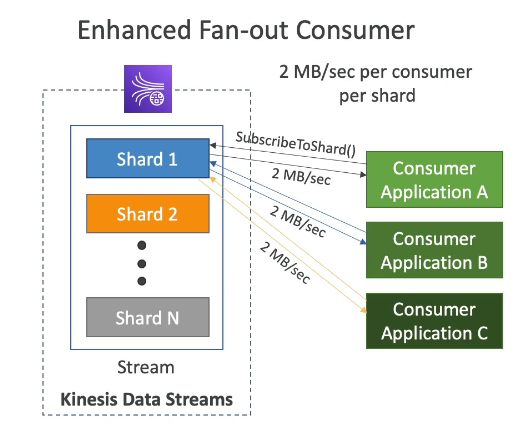
****

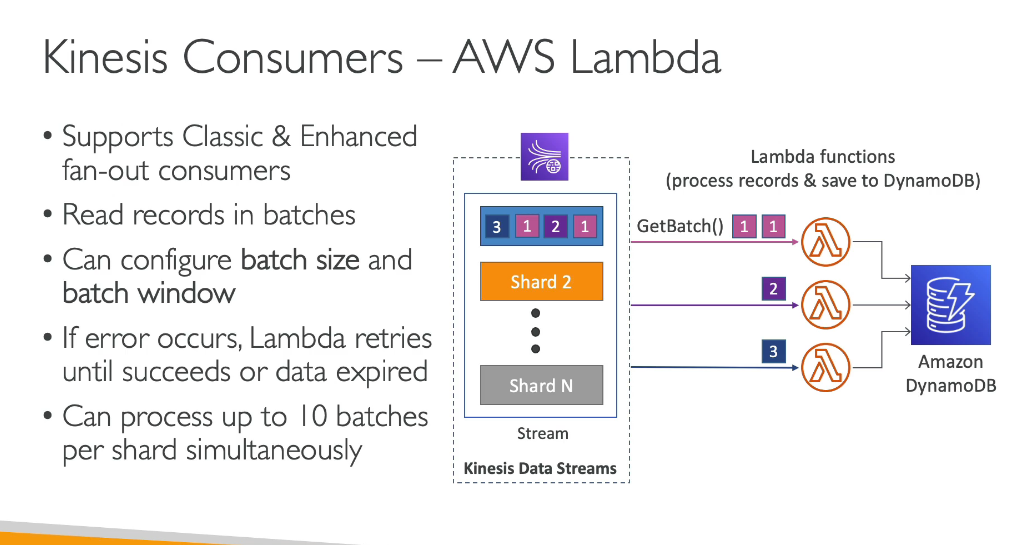
In this we can have a limited number of shards and the number of applications that can access the shards.Here we had 2MB/sec per shard across all consumers.This is considered to be a pull modwel

**Enhanced fan-out consumer**

**This is meant to be a push model.**in this each of the consumer will get acapacity of 2mb/sec per shard per consumer as can be seen below.





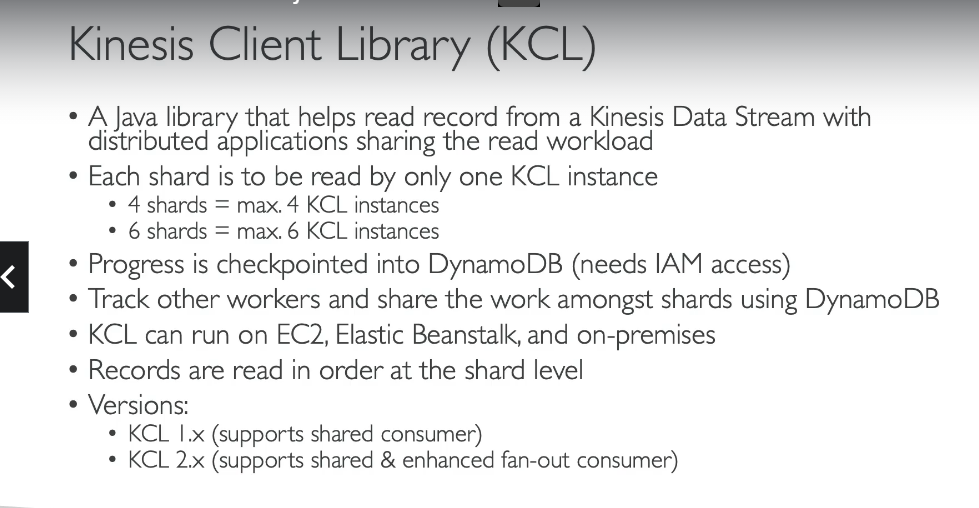


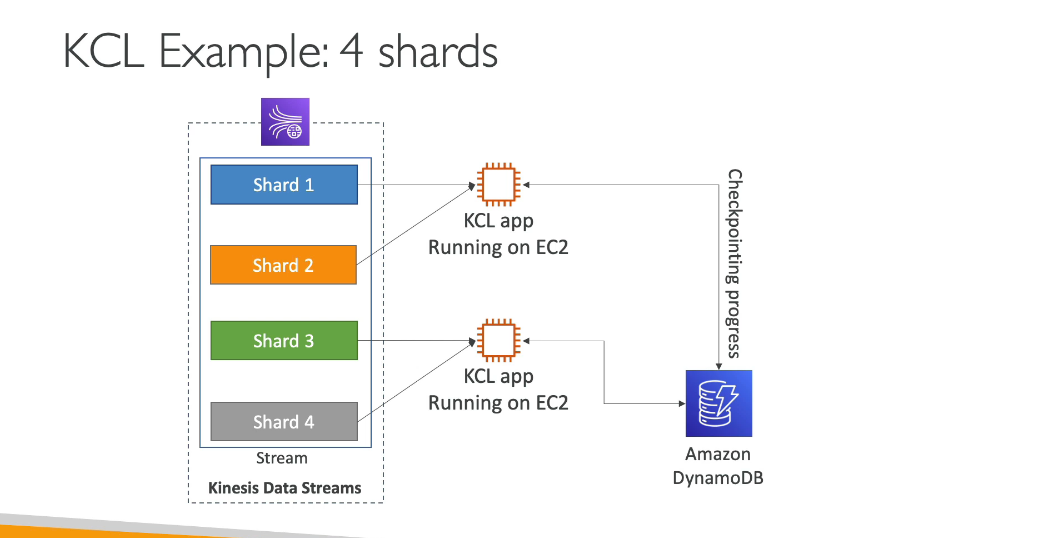
**KINESIS CLIENT LIBRARY**

A kinesis client library is a java library that helsp read records from the kinesis data stream with distributed applications sharing the read workload.

Each shard is to be read by onle one KCL instance.

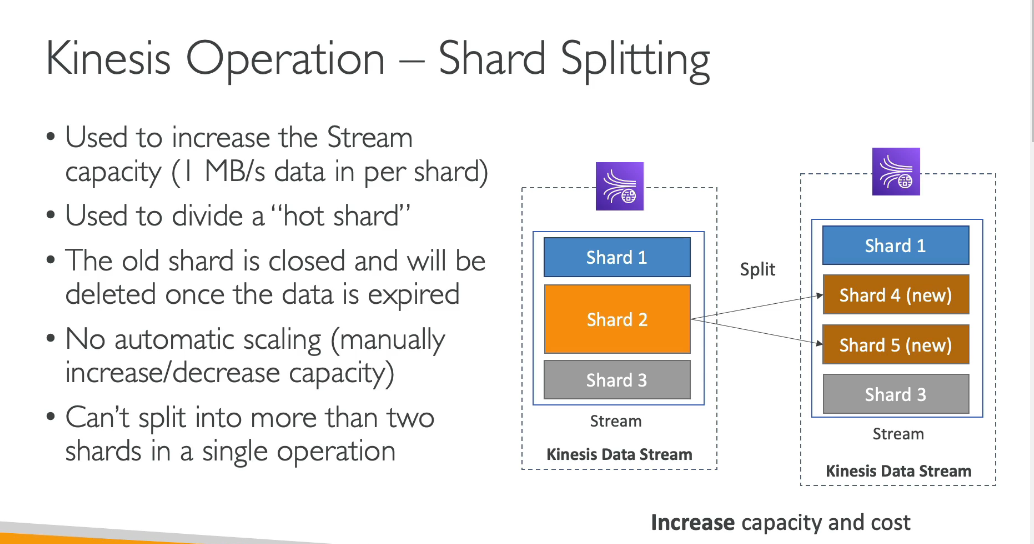
The progress is checkpointed into dynamodb.





Diagram

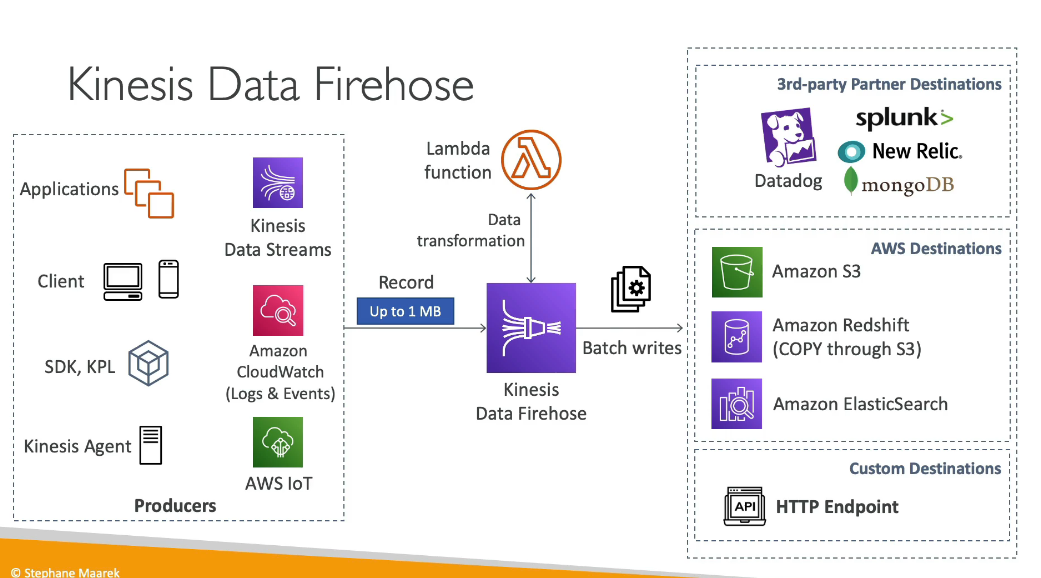
Description automatically generated

**KINESIS SCALING:SHARD SPLITTING**

**Graphical user interface, diagram, application

Description automatically generated**

**KINESIS DATA FIREHOSE**

****

**A** kinesis data firehose can take data from the below shoen producers and then it can process that data suppose using an aws lambda function and after that it can write the data to a destination without needing any code to be written by us.

Diagram, Teams

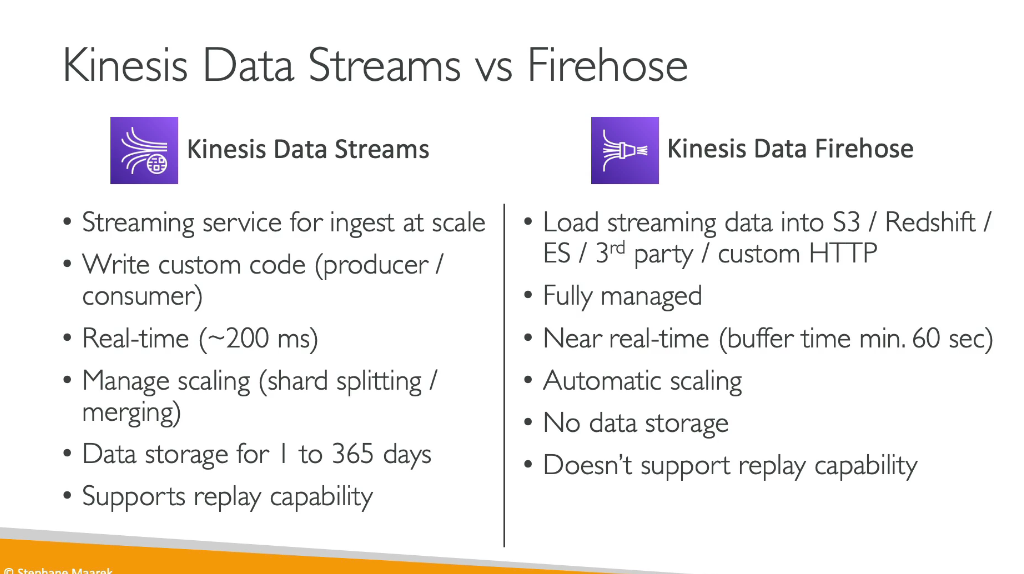
Description automatically generated

Once the data is sent to all these places then then it can be stored into aws s3 as a backup or the failed data can be sent to the s3 bucket.

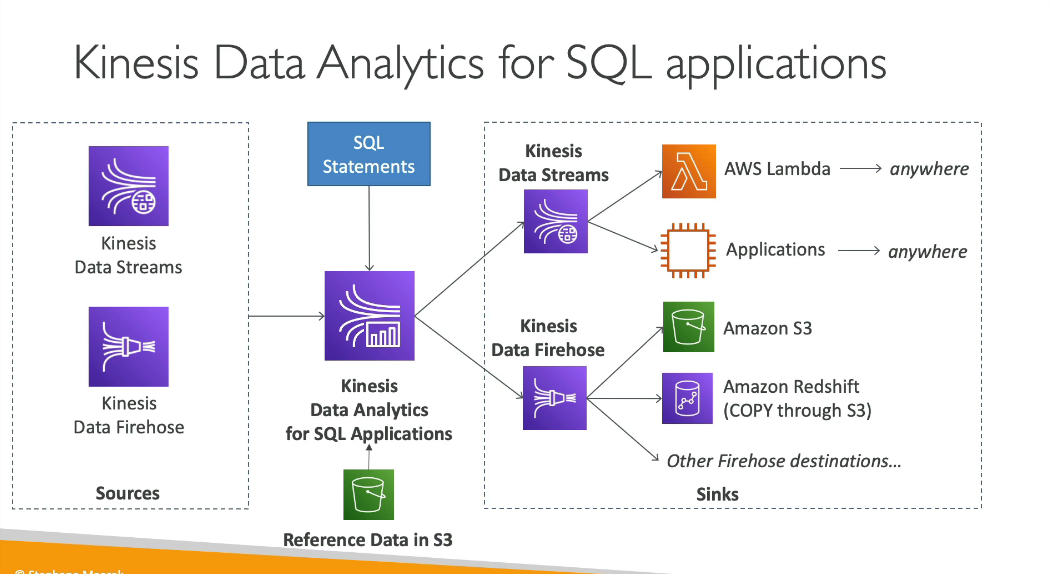
Graphical user interface, text, application, email

Description automatically generated

**WHEN TO USE A KINESIS DATA STREAMA ND WHEN TO USE A KINESIS DATA FIREHOSE**



**KINESIS DATA ANALYTICS FOR SQL APPLICATION**

****