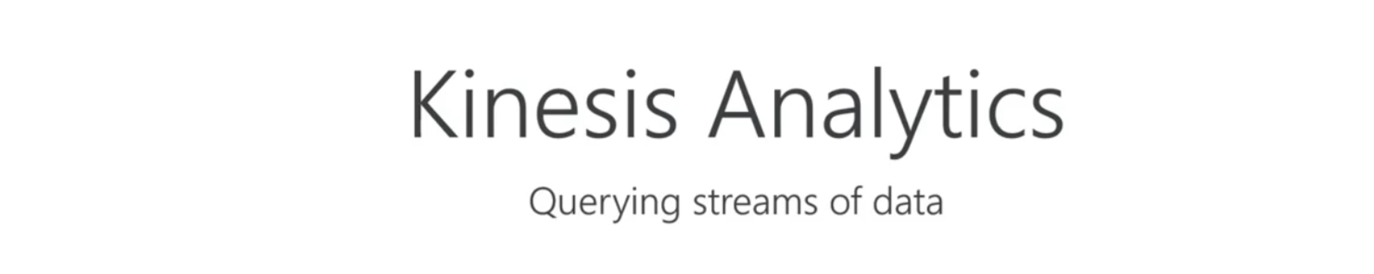
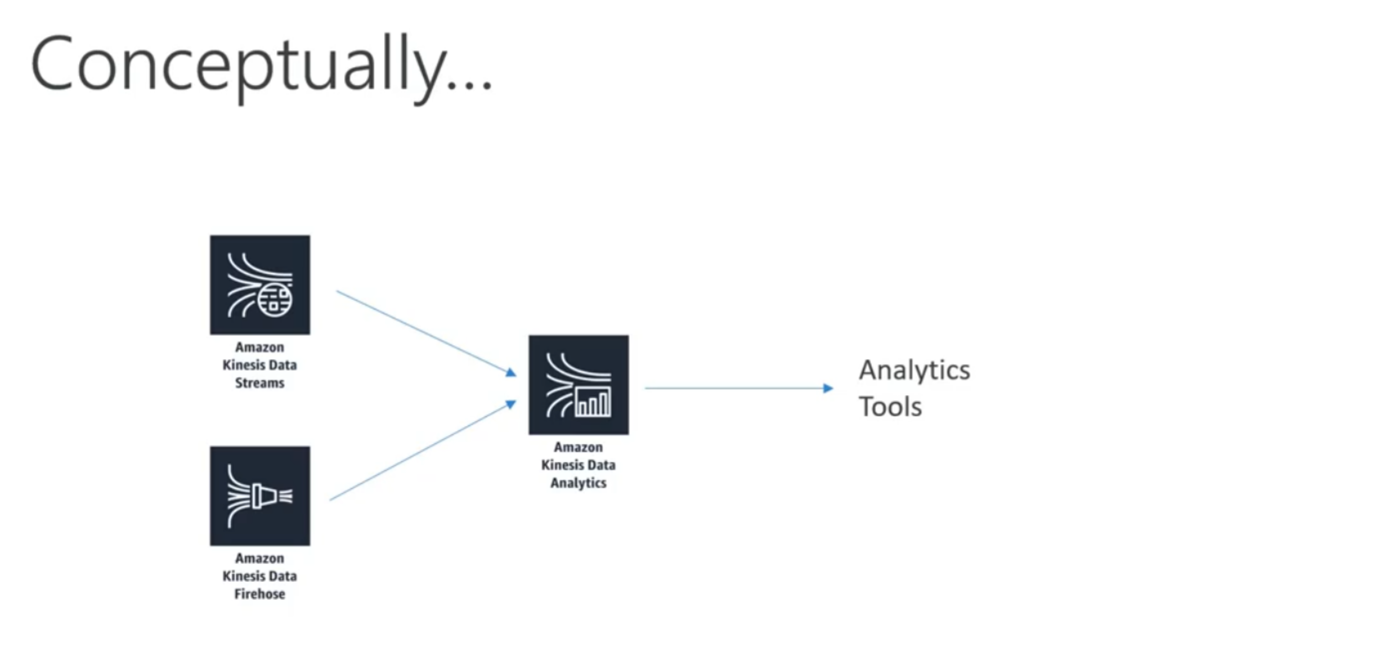
Kinesis Analytics



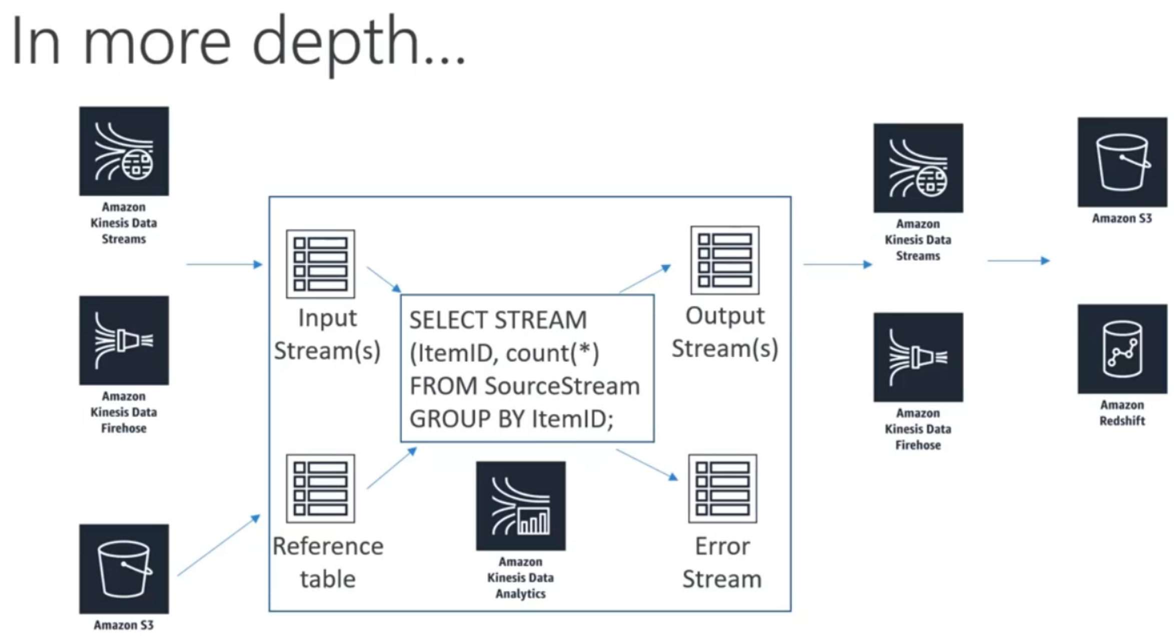
Kinesis Analytics

Querying streams of data



Conceptually...

Amazon Kinesis Data Streams
Amazon Kinesis Data Firehose
Amazon Kinesis Data Analytics
Analytics Tools

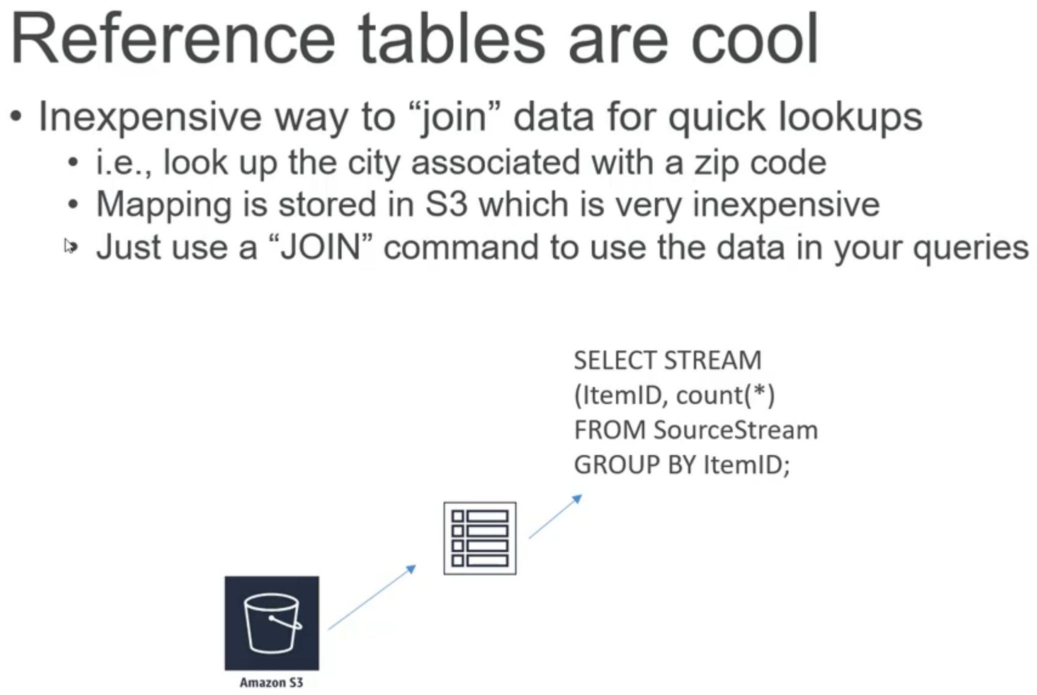


In more depth...

```
SELECT STREAM
(ItemID, count(*)
FROM SourceStream
GROUP BY ItemID;
```

Amazon Kinesis Data Streams
Amazon Kinesis Data Firehose
Amazon S3
Input Stream(s)
Reference table
Amazon Kinesis Data Analytics
Output Stream(s)
Error Stream
Amazon Kinesis Data Streams
Amazon Kinesis Data Firehose
Amazon S3
Amazon Redshift

about pic we saw the data is going via amazon kinesis data stream , kinesis data firehouse , amazon s3 bucket is reference table if it got any error stream means broken are some missing values . after processing data it will go to the next process going to kinesis data stream to s3 bucket



Reference tables are cool

- Inexpensive way to “join” data for quick lookups
  - i.e., look up the city associated with a zip code
  - Mapping is stored in S3 which is very inexpensive
  - Just use a “JOIN” command to use the data in your queries

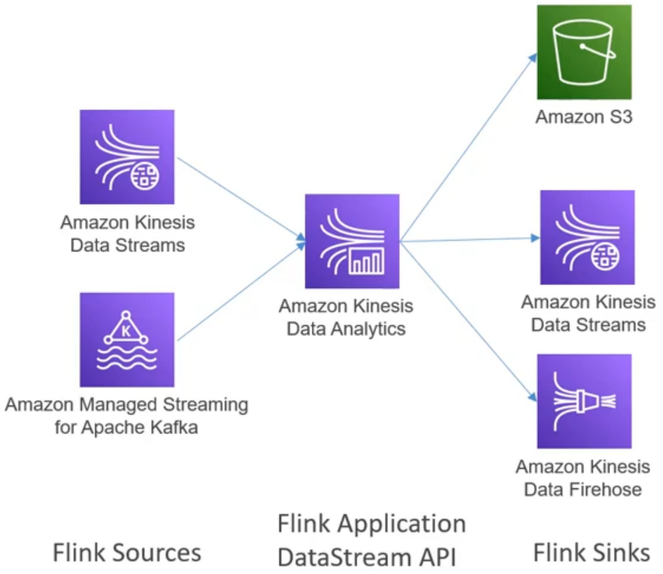
```
SELECT STREAM
(ItemID, count(*)
FROM SourceStream
GROUP BY ItemID;
```

Kinesis Data Analytics + Lambda

- AWS Lambda can be a destination as well
- Allows lots of flexibility for post-processing
  - Aggregating rows
  - Translating to different formats
  - Transforming and enriching data
  - Encryption
- Opens up access to other services & destinations
  - S3, DynamoDB, Aurora, Redshift, SNS, SQS, CloudWatch

Kinesis Data Analytics for Apache Flink

- Formerly Kinesis Data Analytics for Java
  - Kinesis Data Analytics always used Flink under the hood
  - But now supports Scala as well as Java
- Flink is a framework for processing data streams
- Kinesis Data Analytics integrates Flink with AWS
  - Instead of using SQL, you can develop your own Flink application from scratch and load it into KDA via S3
- Serverless



Amazon Kinesis Data Streams

Amazon Managed Streaming for Apache Kafka

Amazon Kinesis Data Analytics

Amazon S3

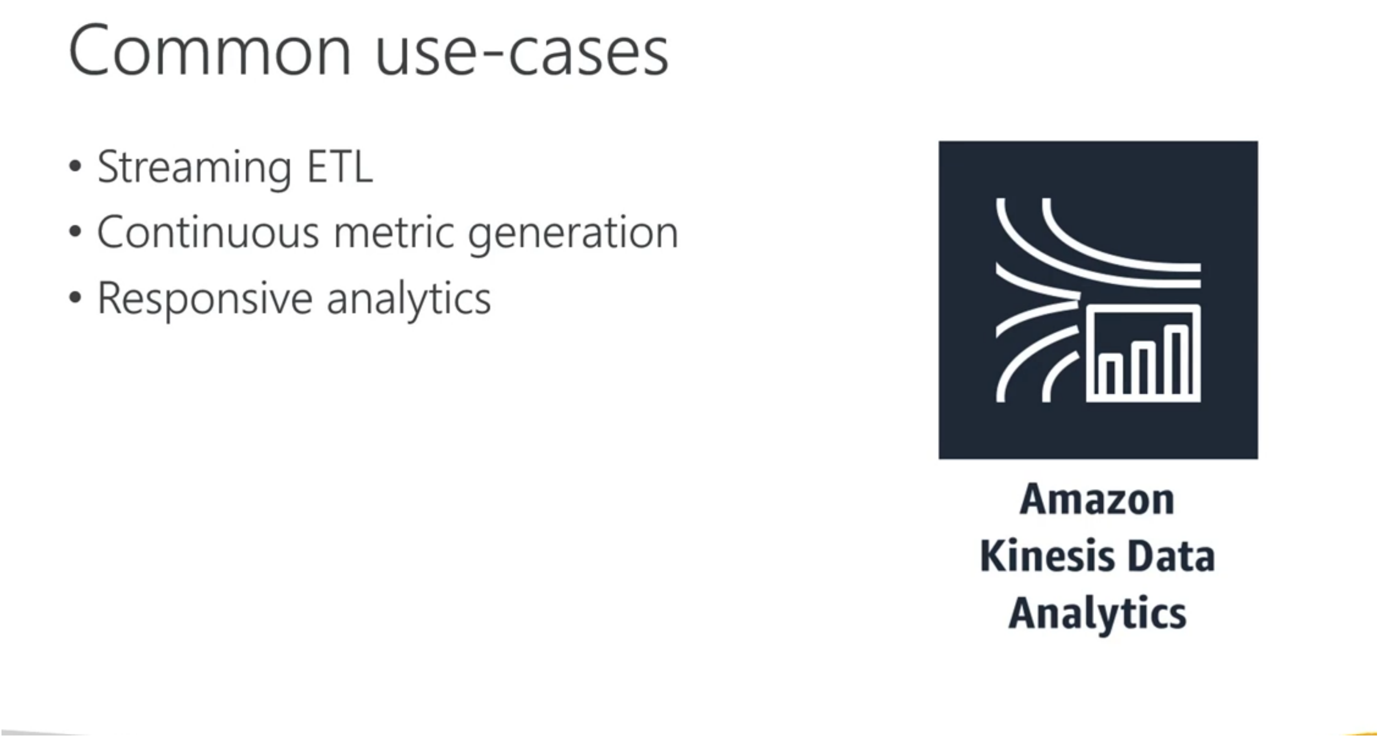
Amazon Kinesis Data Streams

Amazon Kinesis Data Firehose

Flink Sources

Flink Application DataStream API

Flink Sinks



Common use-cases

- Streaming ETL
- Continuous metric generation
- Responsive analytics

Amazon Kinesis Data Analytics

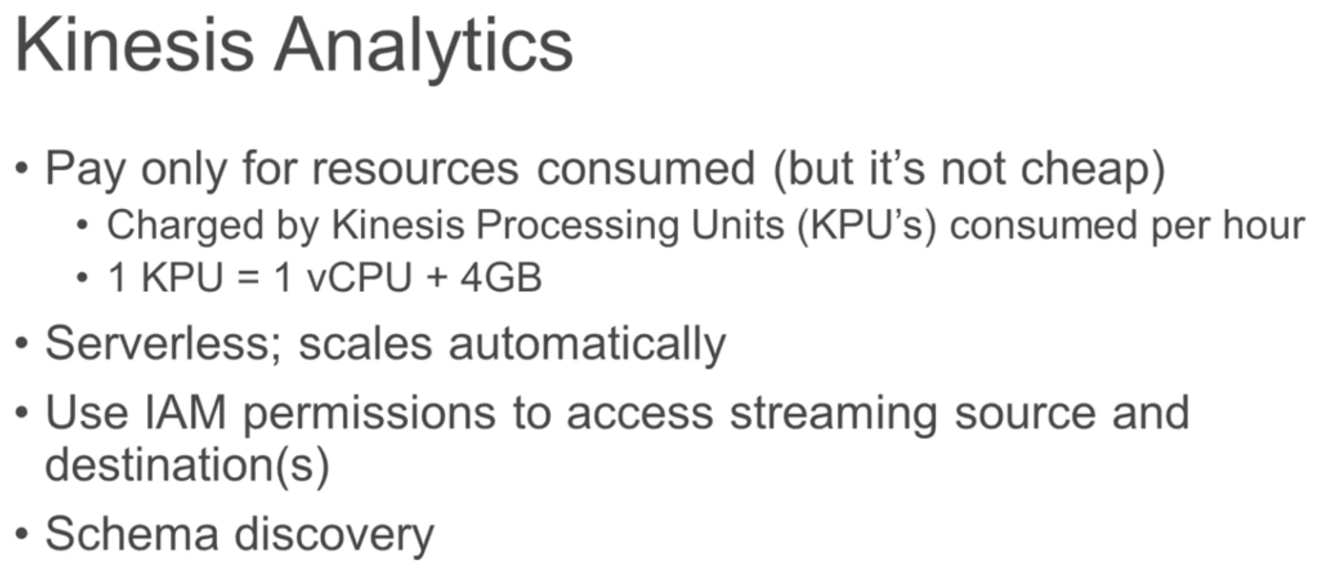
stremang Etl : for example you create a app continuous iot sensors in a kinesis data stream organise the data normalise the data detect the data it was repeated or not and send the data to s3 bucket

continuous metric generation ; for example you create a game leader broad to continuous update the first and some use places

read the data from kinesis data strem and analysis the data in kinesis data anyalytics then data will be wrote on dynamo db

check the website uniq users for every five minutes and wrote the data in redshift for after analysis

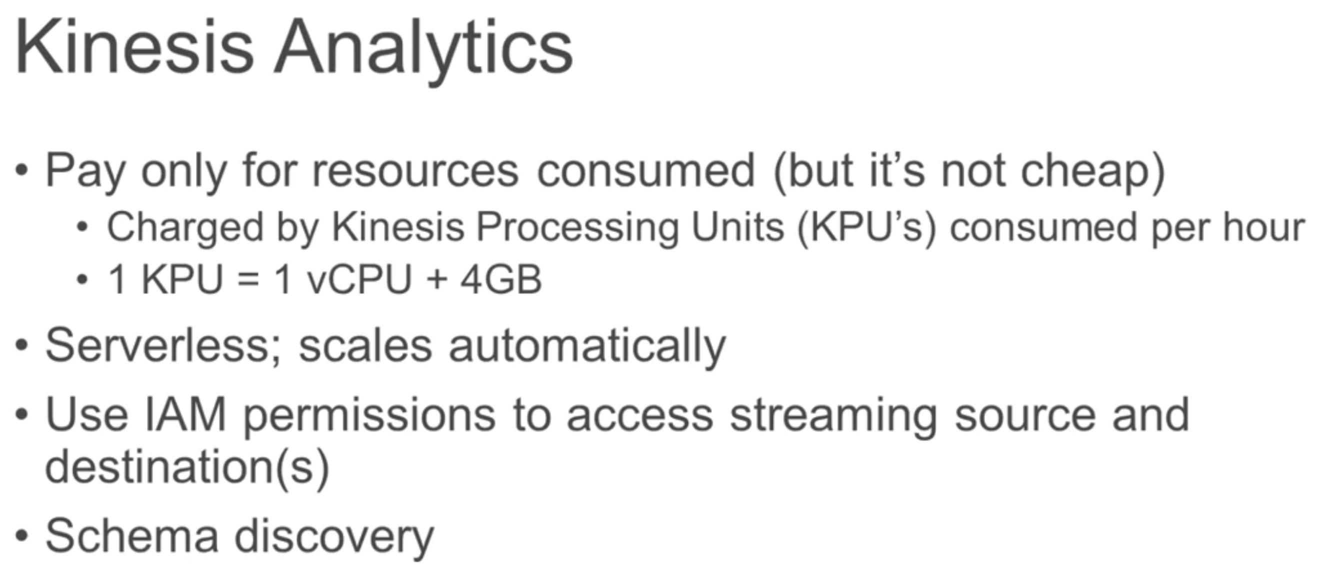
responsive analytics : check the api success rate for payment request and send data to could watch



Kinesis Analytics

- Pay only for resources consumed (but it's not cheap)
  - Charged by Kinesis Processing Units (KPU's) consumed per hour
  - 1 KPU = 1 vCPU + 4GB
- Serverless; scales automatically
- Use IAM permissions to access streaming source and destination(s)
- Schema discovery

we can also use anomaly detection



Kinesis Analytics

- Pay only for resources consumed (but it’s not cheap)
  - Charged by Kinesis Processing Units (KPU’s) consumed per hour
  - 1 KPU = 1 vCPU + 4GB
- Serverless; scales automatically
- Use IAM permissions to access streaming source and destination(s)
- Schema discovery