



Amazon Kinesis and Amazon MSK Overview

Analyzing real-time data streams on AWS

Sven Hansen

Solutions Architect – AWS

svenhans@amazon.com

Agenda

Why real-time analytics and data streaming?

Amazon Kinesis overview

Amazon MSK overview

Demo

Data is produced from a large **variety** of sources



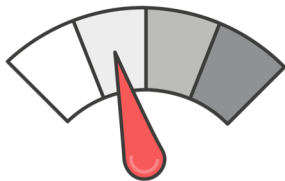
Mobile apps



Web clickstream

```
[Wed Oct 11 14:32:52  
2018] [error] [client  
127.0.0.1] client denied  
by server configuration:  
/export/home/live/ap/htd  
ocs/test
```

Application logs



Metering records

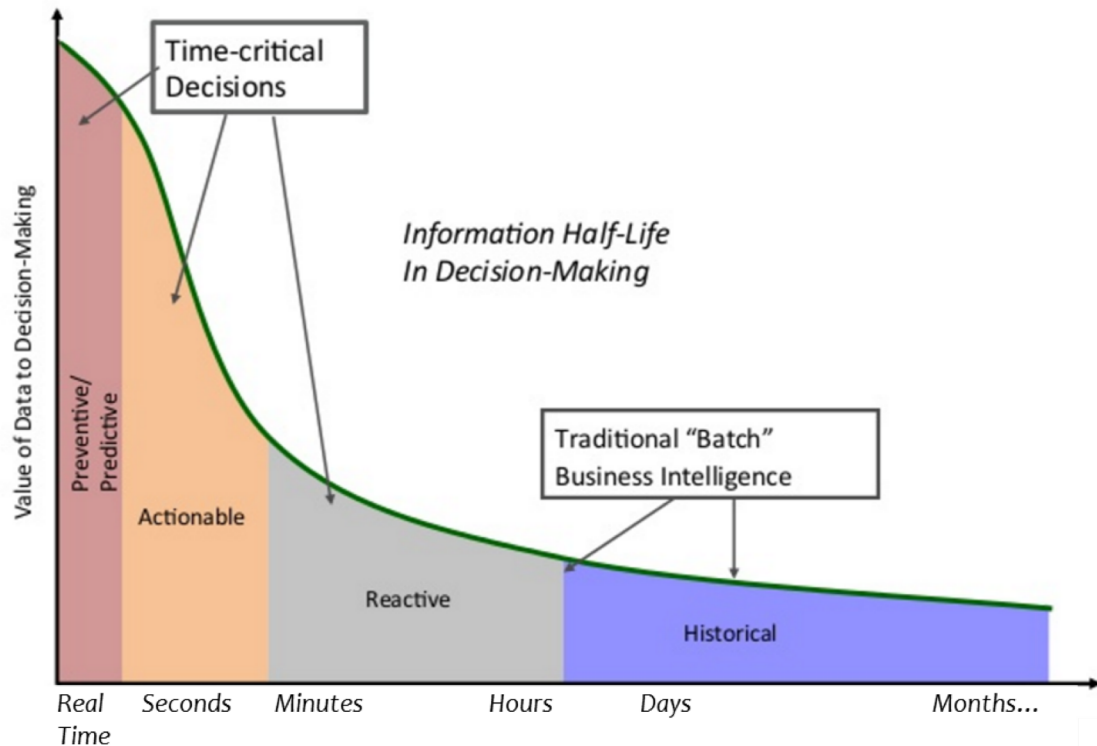


IoT sensors



Smart buildings

The value of data diminishes over time



Source: Perishable insights, Mike Gaultieri, Forrester

Enabling real-time analytics: Considerations

Data streaming technology enables a customer to ingest, process and analyze high volumes of high velocity data from a variety of sources **in real time**



Challenges of data streaming



Difficult to setup



Tricky to scale



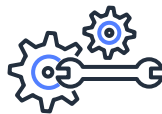
Hard to achieve high availability



Integration requires development



Error prone and complex to manage



Expensive to maintain

Streaming real-time data with AWS

Easily collect, process and analyze data streams in real time

Easy to use

Elastic

High availability
and durability

Seamless integration
with AWS services

Fully managed

Pay for what you use

Streaming data with AWS

Easily collect, process, and analyze data streams in real time



Amazon
Kinesis Data
Streams

Capture and store
data streams



Amazon
Kinesis Data
Analytics

Analyze data
streams in real time



Amazon
Kinesis Data
Firehose

ETL streaming into
streams, data lakes
and warehouses

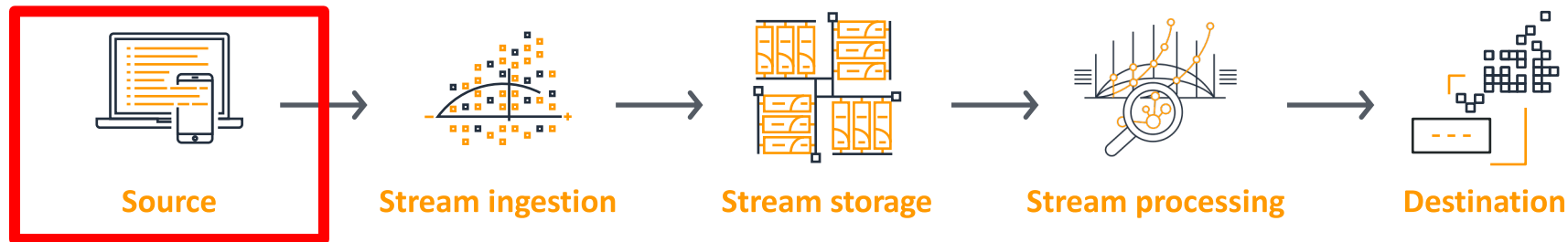


Amazon Managed
Streaming for Kafka

Capture and store
data streams

Enabling real-time analytics

Data streaming technology enables a customer to ingest, process and analyze high volumes of high velocity data from a variety of sources **in real time**



Enabling real-time analytics

Data streaming technology enables a customer to ingest, process and analyze high volumes of high velocity data from a variety of sources **in real time**



Stream Ingestion

Data from tens of thousands of data sources can be written to a single stream



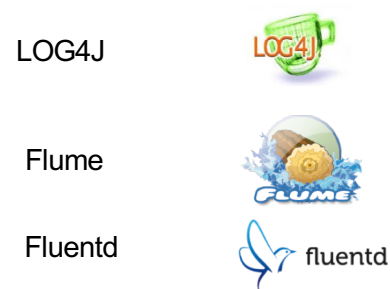
AWS Toolkits/Libraries



AWS Service Integrations



3rd Party Offerings



* Amazon DMS includes 8 on-premise databases, 1 Azure database, 5 RDS/Aurora database types, and S3

Enabling real-time analytics

Data streaming technology enables a customer to ingest, process and analyze high volumes of high velocity data from a variety of sources **in real time**



Stream Storage

Data is stored in the order it was received for a set duration of time, and can be replayed indefinitely during this time.



**Amazon Kinesis
Data Streams**

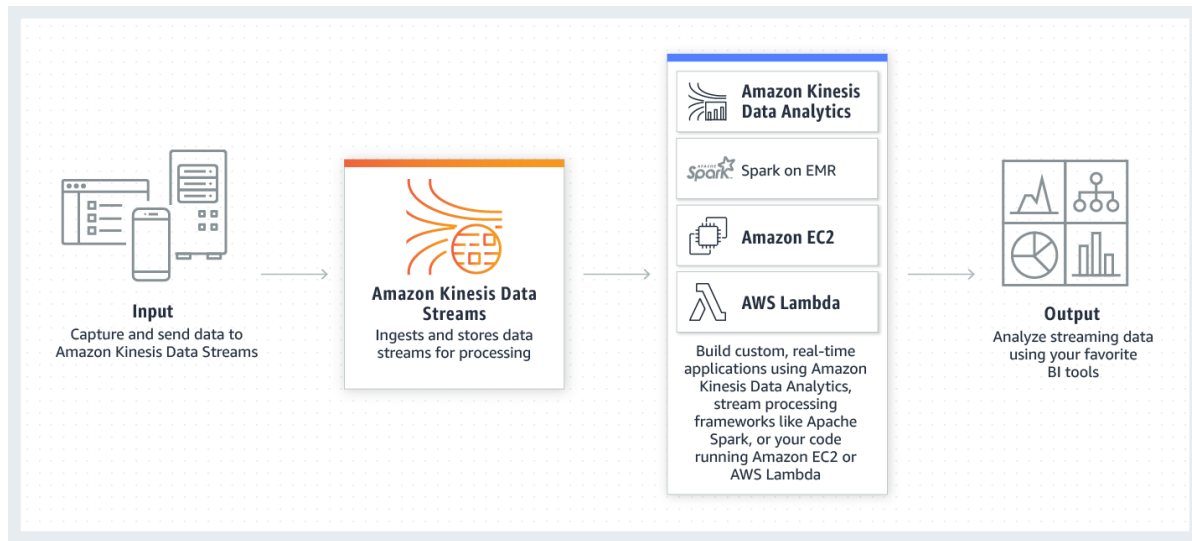


**Amazon
Managed
Streaming for
Kafka**



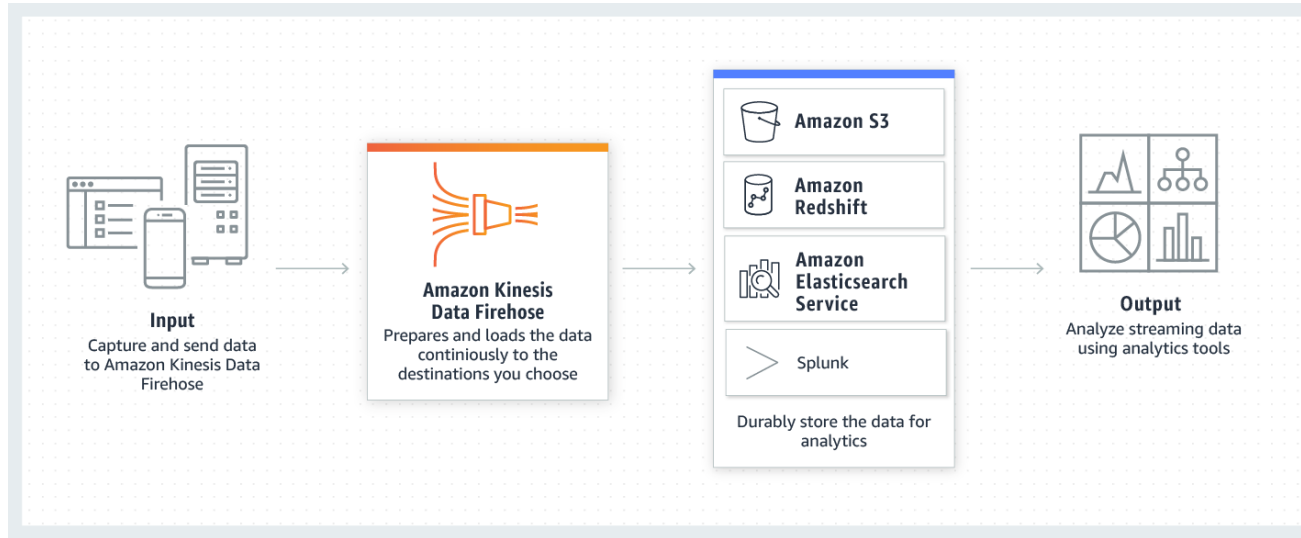
**Amazon Kinesis
Data Firehose**

Amazon Kinesis Data Streams



- Easy administration and low cost
- Real-time, elastic performance
- Secure, durable storage
- Available to multiple real-time analytics applications

Amazon Kinesis Data Firehose



- Zero administration and seamless elasticity
- Direct-to-data store integration
- Serverless continuous data transformations
- Near real-time

Enabling real-time analytics

Data streaming technology enables a customer to ingest, process and analyze high volumes of high velocity data from a variety of sources **in real time**



Stream Processing

Records are read in the order they are produced enabling real-time analytics or streaming ETL



Kinesis



Amazon
Kinesis Data
Analytics

SQL



Flink



Kinesis Client Library
+
Connector Library

AWS Services



AWS Lambda

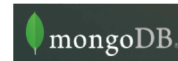


Amazon EMR

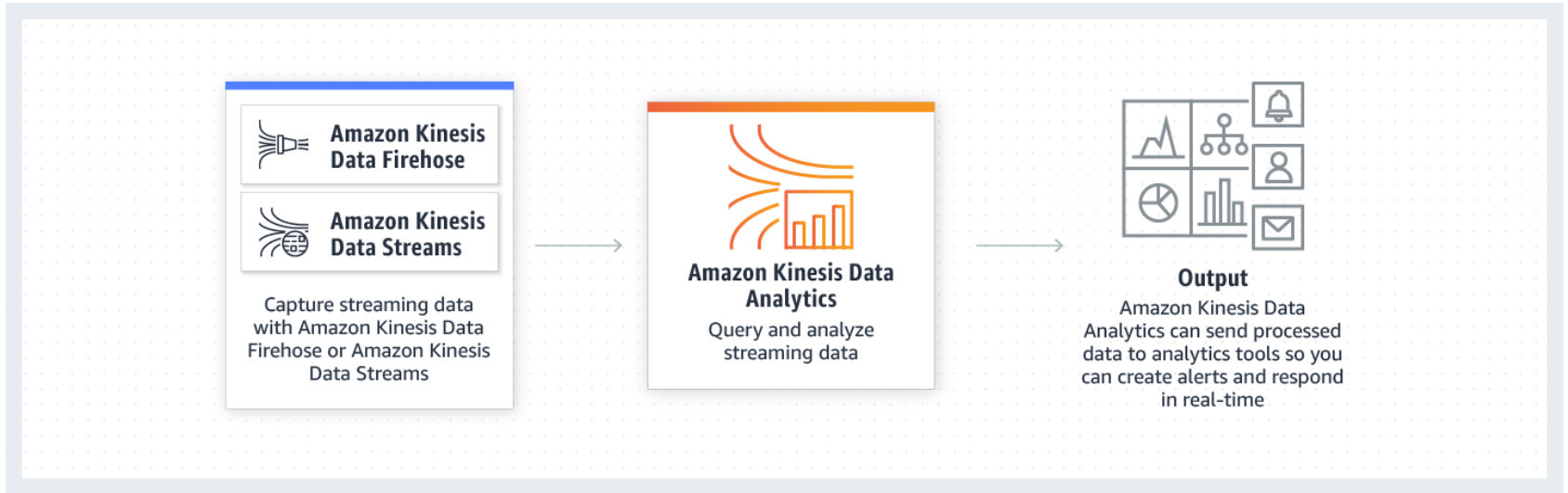
3rd party



Apache Spark



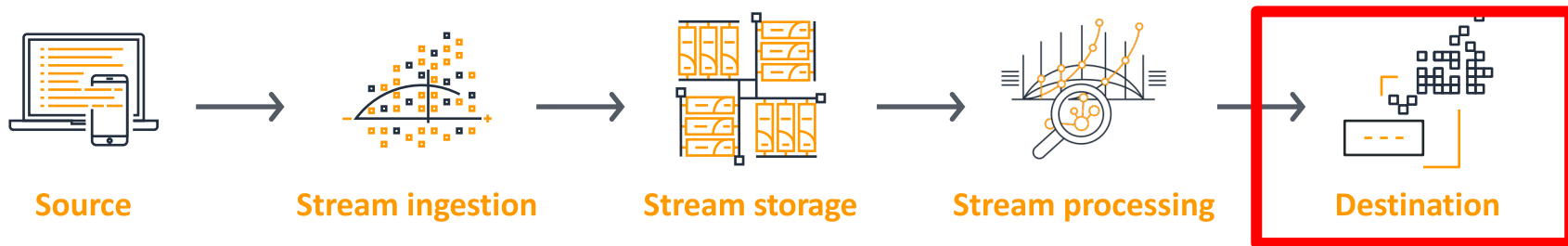
Amazon Kinesis Data Analytics



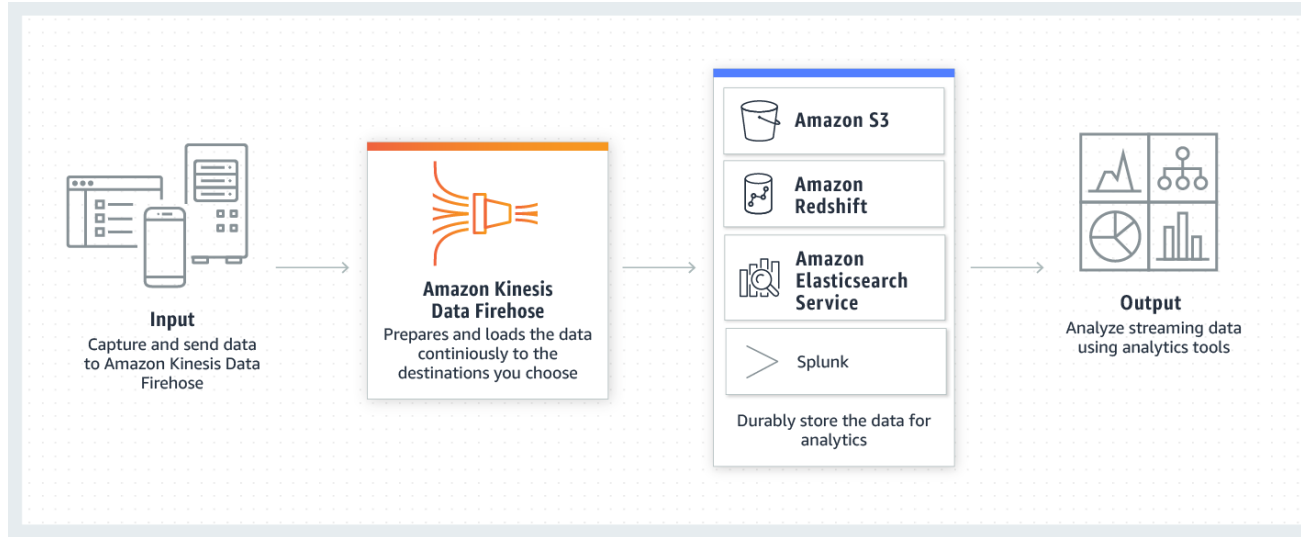
- Interact with streaming data in real-time using SQL or integrated Java applications
- Build fully managed and elastic stream processing applications

Enabling real-time analytics

Data streaming technology enables a customer to ingest, process and analyze high volumes of high velocity data from a variety of sources **in real time**

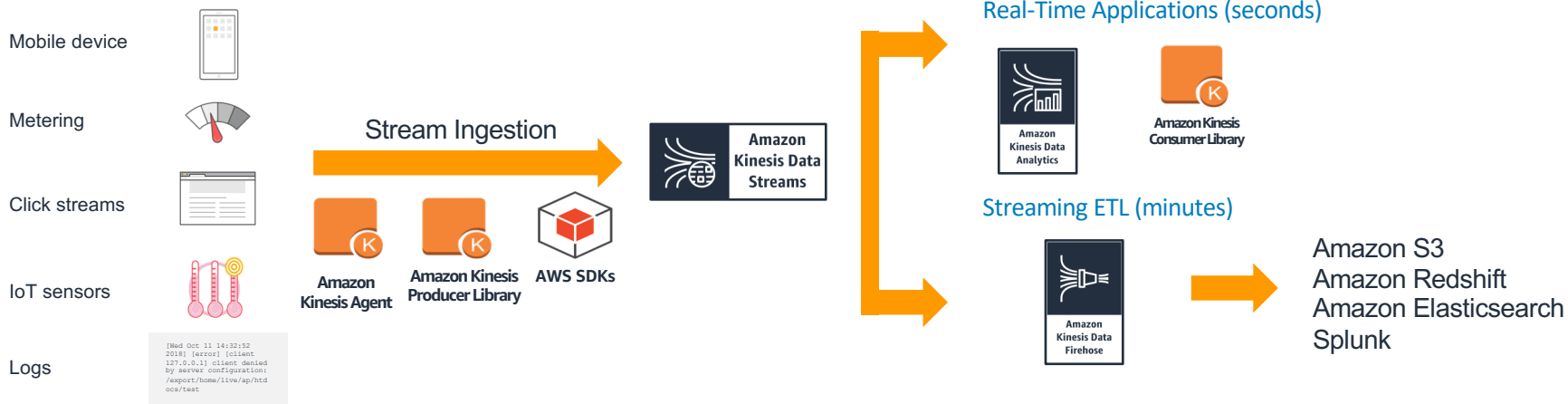


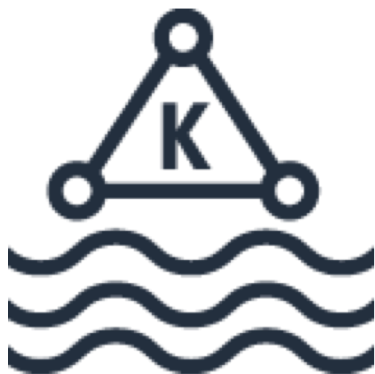
Amazon Kinesis Data Firehose



- Zero administration and seamless elasticity
- Direct-to-data store integration
- Serverless continuous data transformations
- Near real-time

An Example Architecture





Apache Managed Streaming for Kafka

Getting started with Amazon MSK is easy

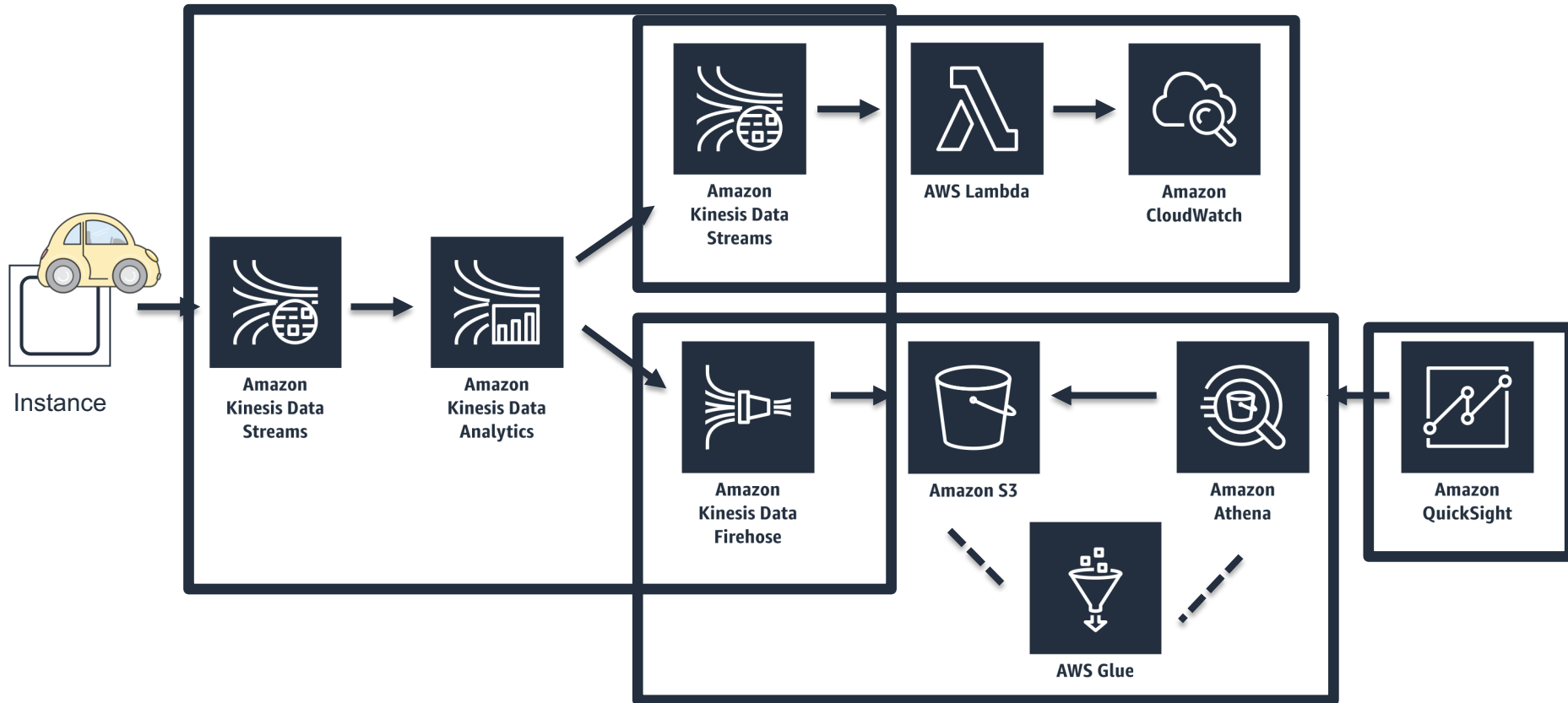


- Fully compatible with Apache Kafka v1.1.1 and v2.1.0
- AWS Management Console and AWS API for provisioning
 - Clusters are setup automatically
 - Provision Apache Kafka brokers and storage
 - Create and tear down clusters on-demand
- Deeply integrated with AWS services

Analyzing NYC Taxi Trips



Application architecture



Demo



Questions?