Analyzing Streaming Data in Real-time with Amazon Kinesis

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Agenda

- Why Real-Time Analytics?
- What Is Real-Time Data?
- What Real-Time Services Does AWS Offer?
- Common Use Cases
- Deep Dive

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Why Real-Time Analytics?

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

It’s All About the Pace

Batch Processing	Stream Processing
Hourly server logs	Real time metrics
Weekly or monthly bills	Real time spending alerts/caps
Daily web-site clickstream	Real time clickstream analysis
Daily fraud reports	Real time detection

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

A Day in Life

Acknowledgement: Icon created by Iconicbestiary - Freepik.com

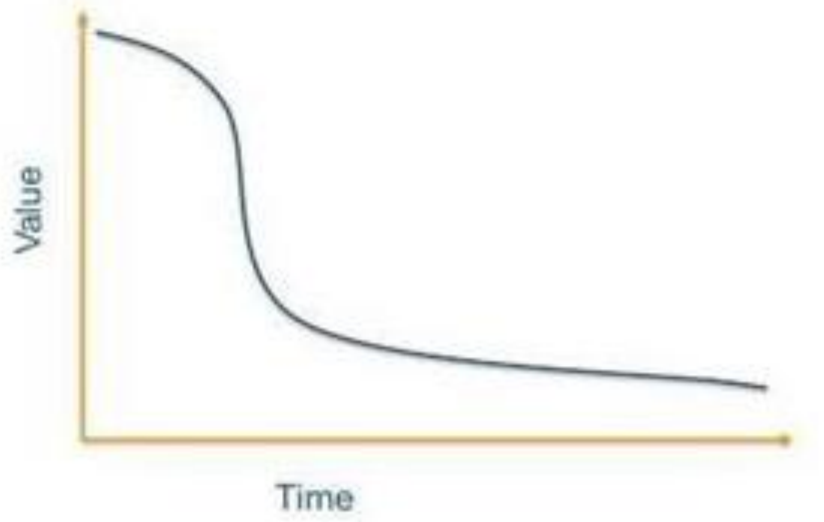
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Data Loses Value Over Time

Ingest data as it is generated

Analyze data in real time to get insights immediately

Deliver data to in seconds instead of hours



© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

What Is Real-Time Data?

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

What Is Real Time Data?

Mobile Apps

Web Clickstream

```
[Wed Oct 11 14:32:52
2000] [error] [client
127.0.0.1] client denied
by server configuration:
/export/home/live/ap/htdo
cs/test
```

Application Logs

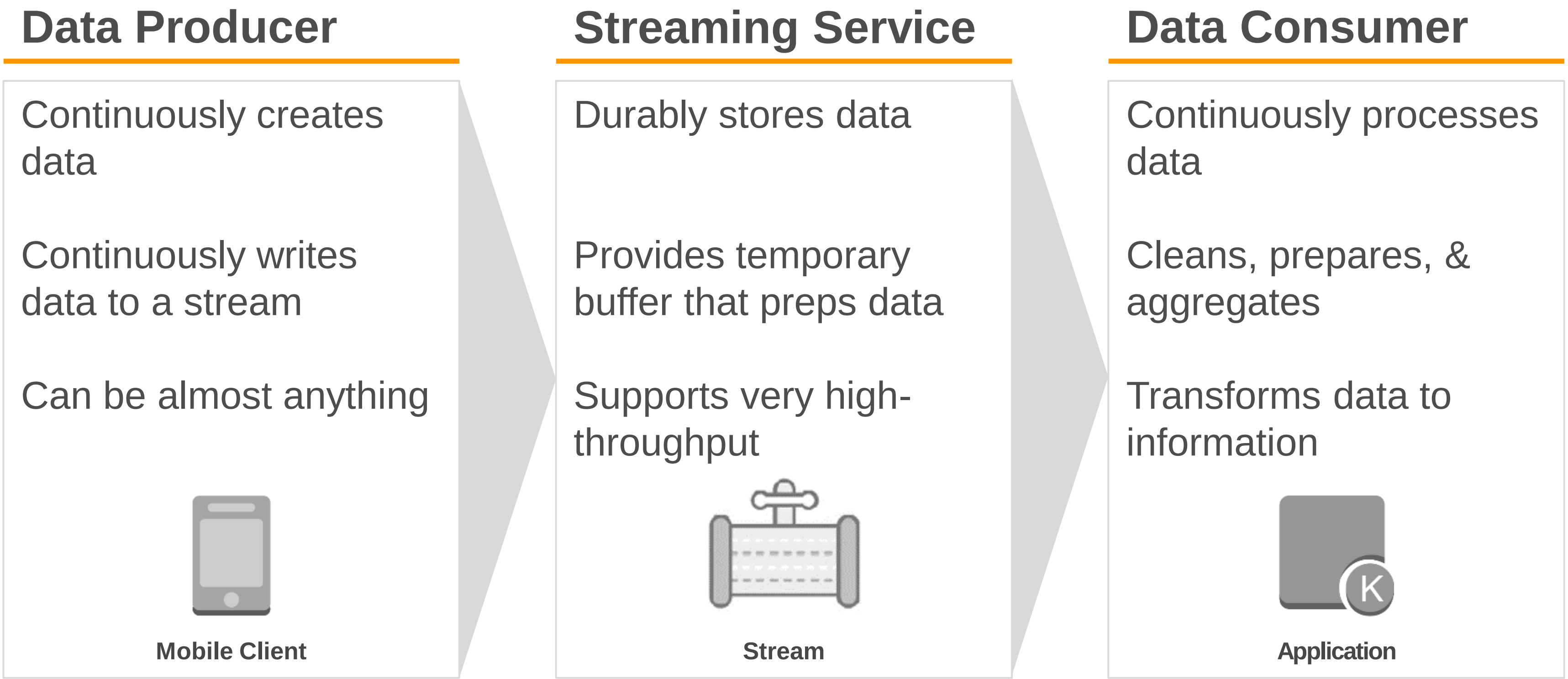
Metering Records

IoT Sensors

Smart Buildings

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Simple Pattern for Streaming Data



Data Producer

Continuously creates data

Continuously writes data to a stream

Can be almost anything

Mobile Client

Streaming Service

Durably stores data

Provides temporary buffer that preps data

Supports very high-throughput

Stream

Data Consumer

Continuously processes data

Cleans, prepares, & aggregates

Transforms data to information

Application

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

What Real-Time Services Does AWS Offer?

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon Kinesis

Amazon Kinesis Data Streams

Build custom applications that process and analyze streaming data

Amazon Kinesis Data Analytics

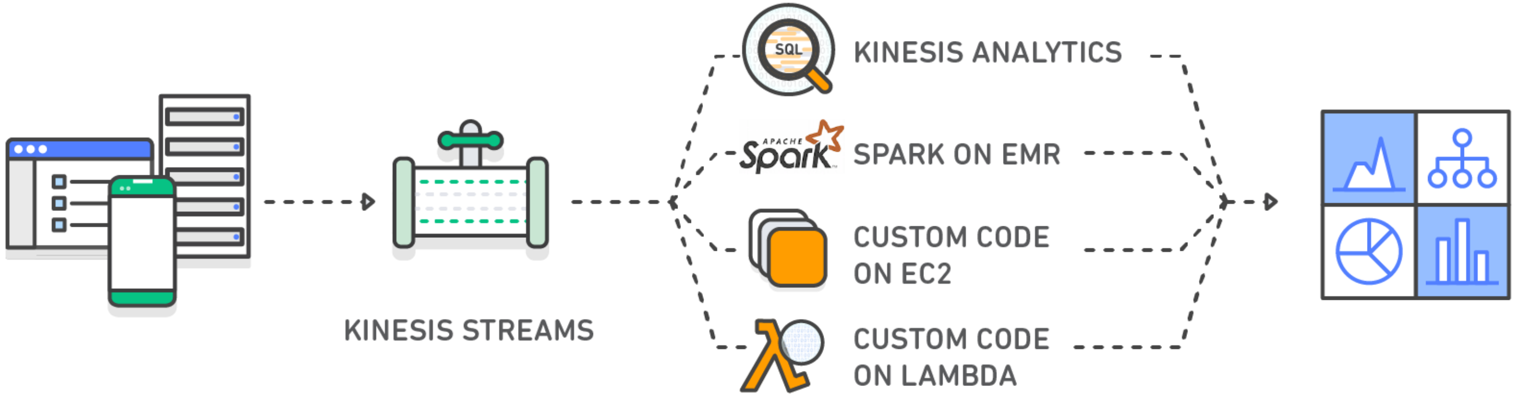
Easily process and analyze streaming data with standard SQL

Amazon Kinesis Data Firehose

Easily load streaming data into AWS

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon Kinesis Data Streams



KINESIS STREAMS

KINESIS ANALYTICS

SPARK ON EMR

CUSTOM CODE ON EC2

CUSTOM CODE ON LAMBDA

Capture and send data to Kinesis Streams

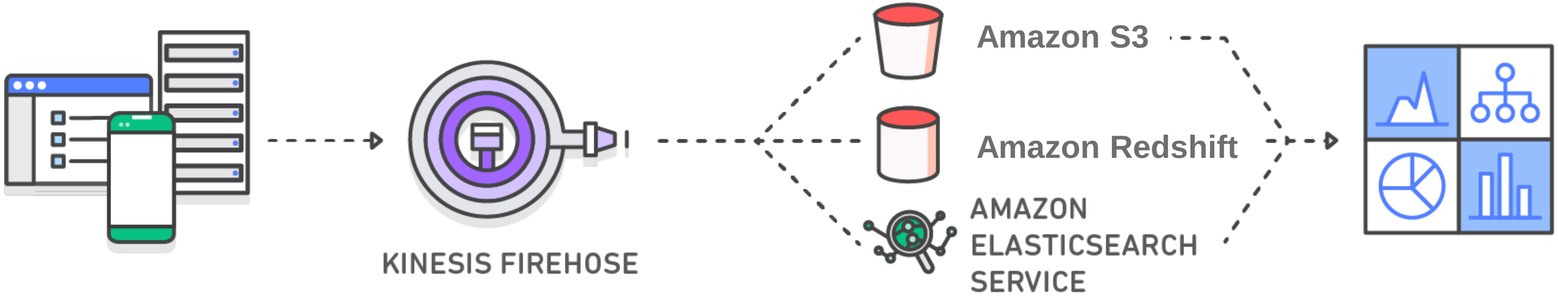
Build custom, real-time applications using Kinesis Analytics, stream processing frameworks like Apache Spark, or your code running on Amazon EC2 or AWS Lambda

Load processed data to any data store, send real-time alerts, feed live dashboards, and more

- Easy administration and low cost
- Build real-time application with framework of choice
- Secure, Durable storage

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon Kinesis Data Firehose



Capture and send data to Kinesis Firehose

Kinesis Firehose prepares and loads the data continuously to the destinations you chose from among S3, Redshift, Amazon Elasticsearch Service, and Kinesis Analytics

Analyze streaming data using your favorite BI tools

- Zero administration and seamless elasticity
- Direct-to-data store integration
- Serverless, continuous data transformation

splunk> NEW

AWS re:Invent

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

Stream Data To Amazon Kinesis

Automatic ingestion

- Amazon VPC Flow Logs
- AWS CloudTrail Event Logs
- Amazon CloudWatch Logs
- AWS IoT events
- Amazon Pinpoint

Easy setup

As a proxy:

- Amazon API Gateway
- Elastic Load Balancing

For change data capture:

- Amazon DynamoDB
- Amazon RDS

Write your own

- Amazon Kinesis Agent
- Amazon Kinesis Producer Library
- AWS SDKs
- Apache LOG4J™
- fluentd

Just a sample… many more ways stream data to Amazon Kinesis

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Integrate With Your Current Solution

VOLTDB

splunk>

Qubole

elastic

APACHE STORM™
Distributed • Resilient • Real-time

databricks

DATADOG

fluentd

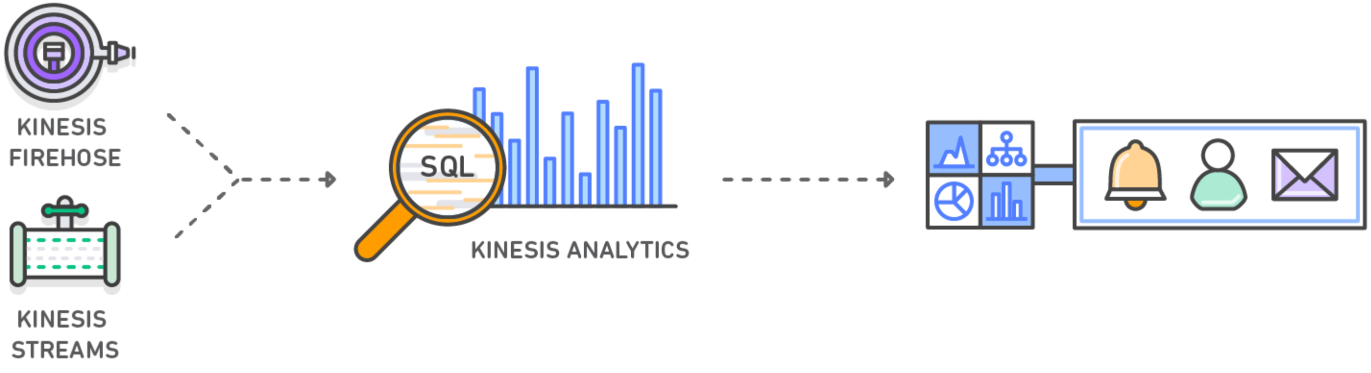
FLUME

mongoDB

Flink

MEMSQL

Amazon Kinesis Data Analytics



Capture streaming data with Kinesis Streams or Kinesis Firehose

Run standard SQL queries against data streams

Kinesis Analytics can send processed data to analytics tools so you can create alerts and respond in real-time

- Continuous anomaly detection
- Continuous time series analysis
- Continuous filtering
- Continuous aggregation
- Continuous enrichment

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon Kinesis Data Analytics Applications

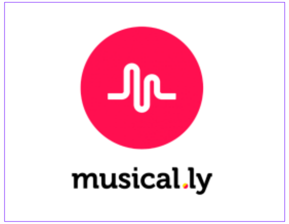
Connect to streaming source

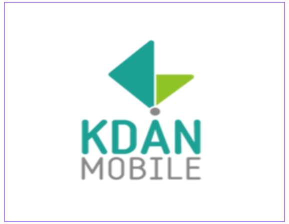
Easily write SQL code to process streaming data

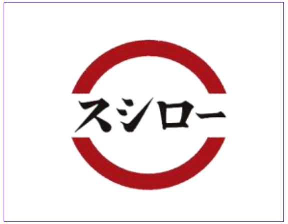
Continuously deliver SQL results

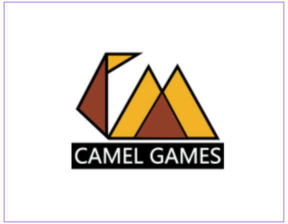
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

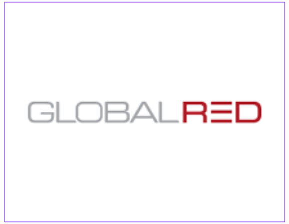
Amazon Kinesis Customers

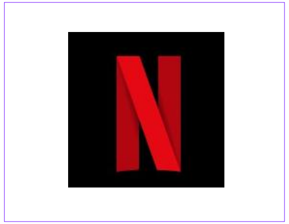


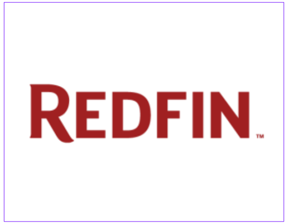


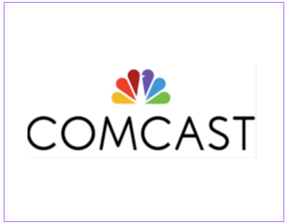


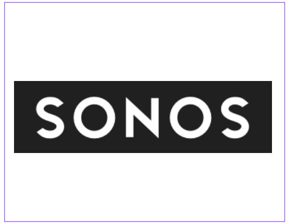


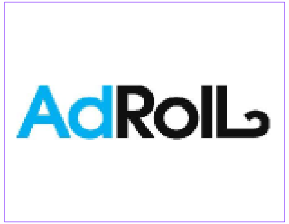


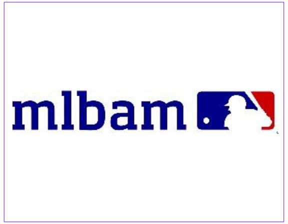


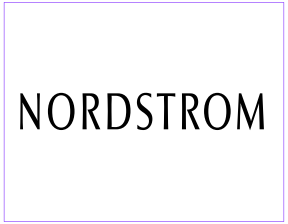












Common Use Cases

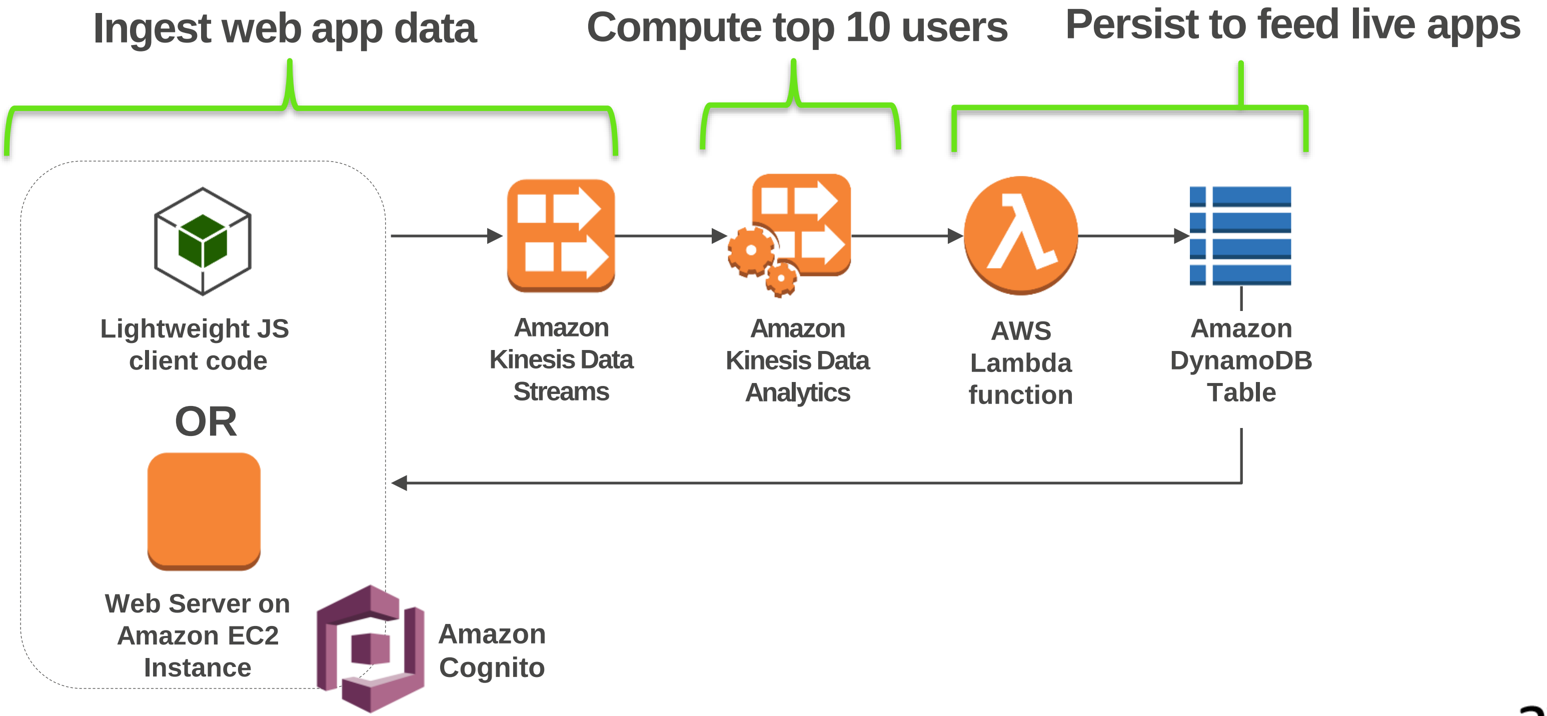
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Three Common Scenarios

Streaming Ingest-Transform-Load	Deliver data to analytics tools faster and cheaper
Continuous Metric Generation	Compute analytics as the data is generated
Actionable Insights	React to analytics based off of insights

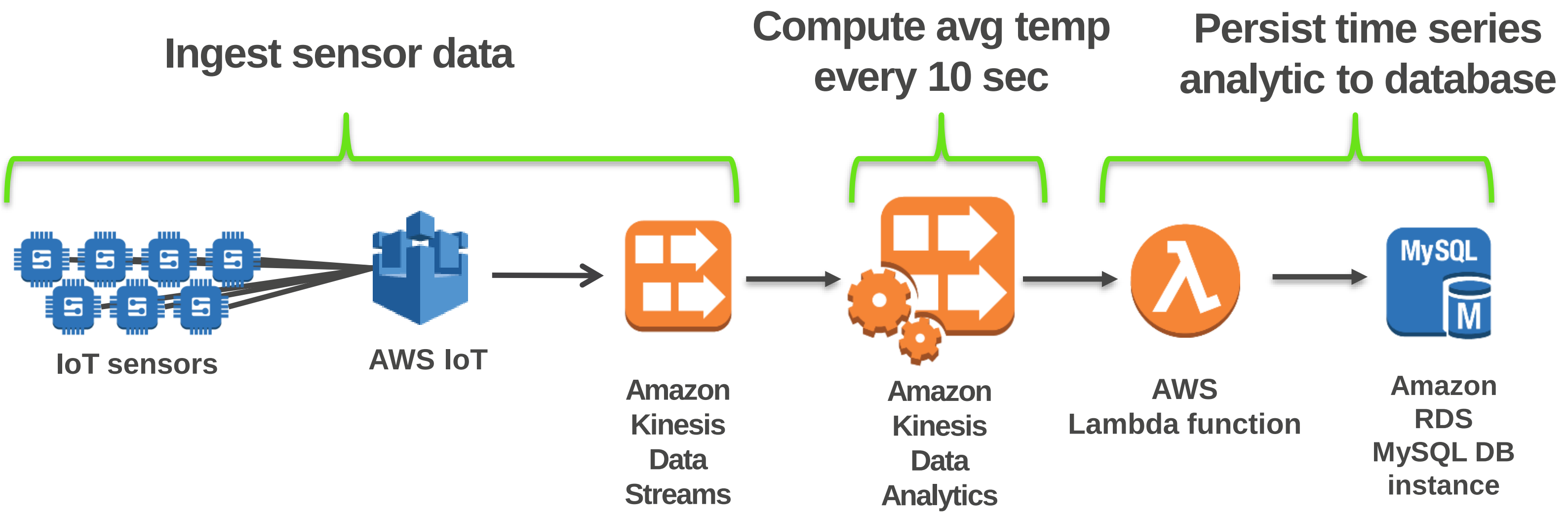
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Web Analytics and Leaderboards


Ingest web app data
Compute top 10 users
Persist to feed live apps
Lightweight JS client code
OR
Web Server on Amazon EC2 Instance
Amazon Cognito
Amazon Kinesis Data Streams
Amazon Kinesis Data Analytics
AWS Lambda function
Amazon DynamoDB Table

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Monitoring IoT Devices



Ingest sensor data

Compute avg temp every 10 sec

Persist time series analytic to database

IoT sensors

AWS IoT

Amazon Kinesis Data Streams

Amazon Kinesis Data Analytics

AWS Lambda function

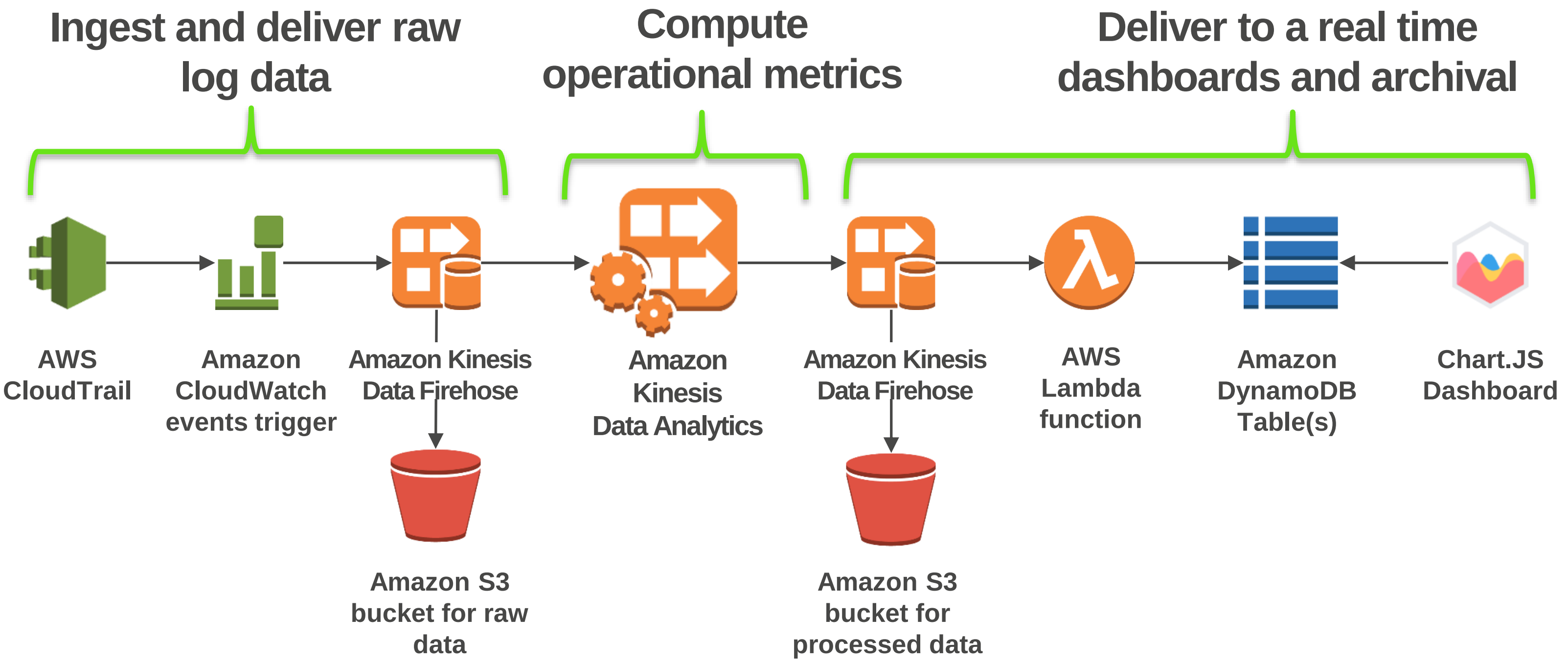
Amazon RDS MySQL DB instance

AWS re:Invent

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

Analyzing CloudTrail Event Logs


Ingest and deliver raw log data
Compute operational metrics
Deliver to a real time dashboards and archival
AWS CloudTrail
Amazon CloudWatch events trigger
Amazon Kinesis Data Firehose
Amazon S3 bucket for raw data
Amazon Kinesis Data Analytics
Amazon Kinesis Data Firehose
Amazon S3 bucket for processed data
AWS Lambda function
Amazon DynamoDB Table(s)
Chart.JS Dashboard

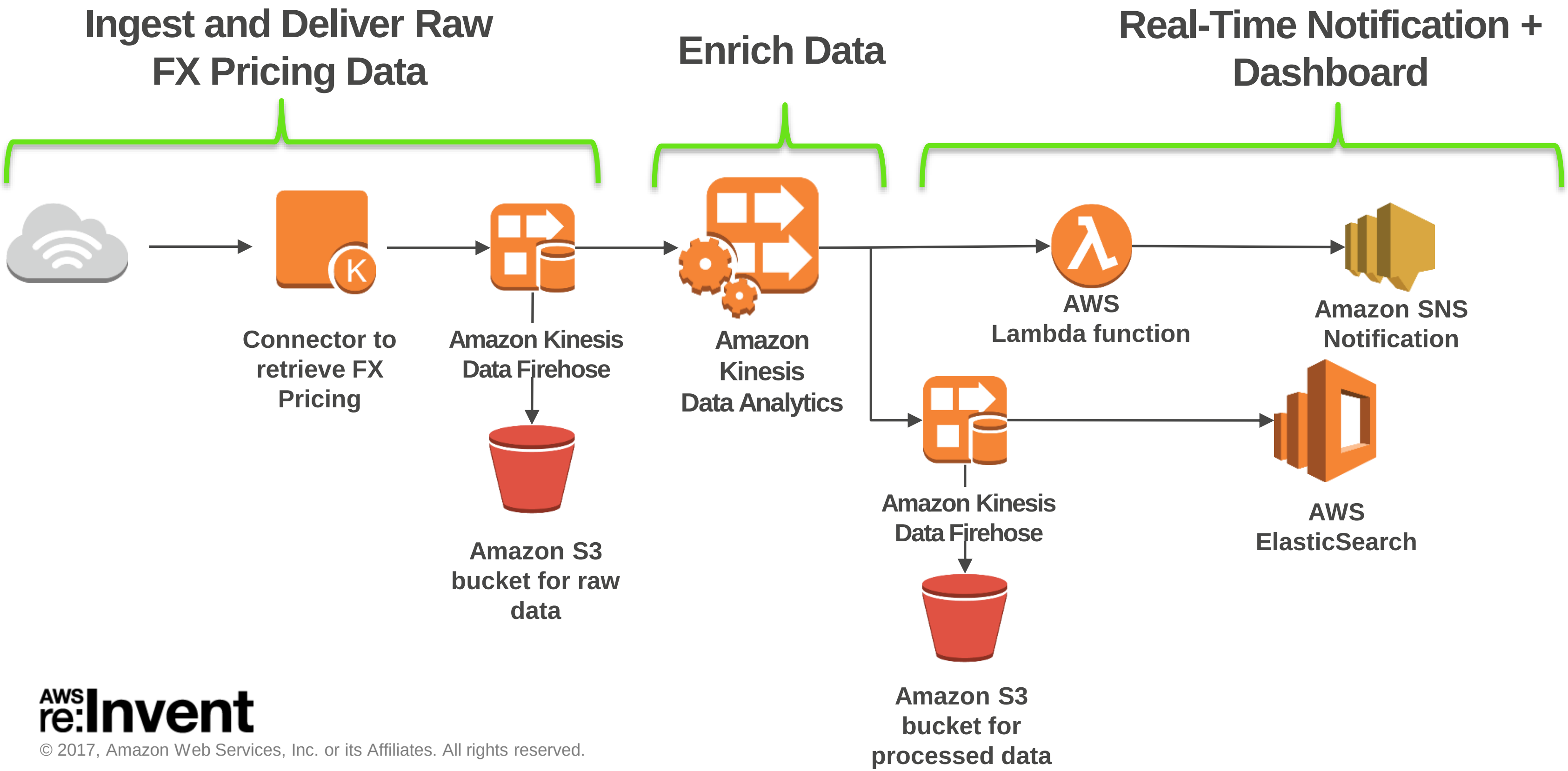
AWS re:Invent

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Analyzing FX Pricing in Near Real Time

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

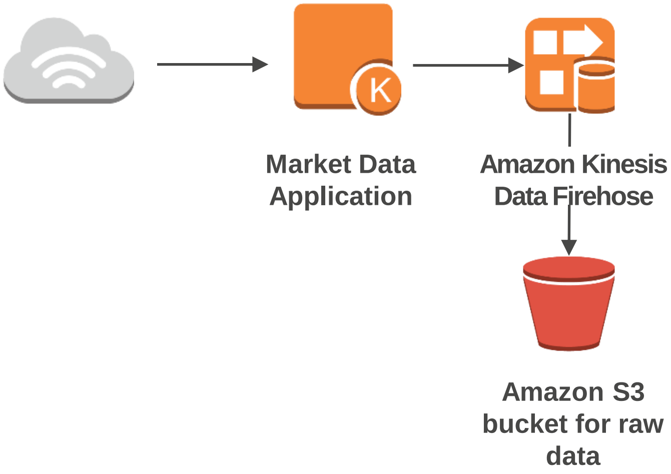
Analyzing FX Pricing in Near Real-Time



Ingest and Deliver Raw FX Pricing Data
Enrich Data
Real-Time Notification + Dashboard
Connector to retrieve FX Pricing
Amazon Kinesis Data Firehose
Amazon S3 bucket for raw data
Amazon Kinesis Data Analytics
AWS Lambda function
Amazon SNS Notification
Amazon Kinesis Data Firehose
Amazon S3 bucket for processed data
AWS ElasticSearch

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Ingest and Deliver FX Pricing

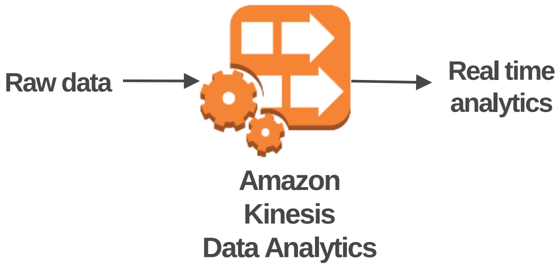


- An application is connected to a Market Data provider delivering FX pricing, every 30 seconds
- Pricing are sent in real time (to near real time) to Kinesis Data Firehose or Streams
- Each pricing in JSON is in below format

```
{
  "ask": 212.2553,
  "timestamp": 1515484223,
  "symbol": "XAGZAR",
  "bid": 211.6336,
  "price": 211.9445
}
```

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Compute Pricing Metrics in Near Real-Time



Raw data

Amazon Kinesis Data Analytics

Real time analytics

Compute metrics using SQL in real time like:

- Max, Min and Average price for the past hour
- Previous Price and Price Change

```
{
  "min": 212.2553,
  "timestamp": 1515484223,
  "symbol": "XAGZAR",
  "max": 211.6336,
  "average": 211.9445
  "change": 1.1954
  "previous_price ": 210.7491
  "price ": 211.9445
}
```

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

How Do I Write Streaming SQL? Easy!

Streams (in memory tables)

```
CREATE OR REPLACE STREAM DESTINATION_SQL_STREAM(
    "@timestamp" BIGINT,
    symbol VARCHAR(8),
    min_price DOUBLE,
    max_price DOUBLE,
    avg_price DOUBLE,
    previous_price DOUBLE,
    new_price DOUBLE
    change DOUBLE
);
```

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

How Do I Write Streaming SQL? Easy!

Pumps (continuous query)

```
CREATE OR REPLACE PUMP "STREAM_PUMP" AS
INSERT INTO "DESTINATION_SQL_STREAM"
SELECT STREAM "COL_timestamp” * 1000,
      "ticker_symbol",
     MIN("price") OVER W1 as min_price,
     MAX("price") OVER W1 as max_price,
     AVG("price") OVER W1 as avg_price,
     FIRST_VALUE("price")OVER R1 as previous_price,
      "price" as new_price,
      "price" – FIRST_VALUE("price") OVER R1 as CHANGE
FROM "SOURCE_SQL_STREAM_001”
WINDOW
 W1 AS (PARTITION BY "ticker_symbol" RANGE INTERVAL '1' HOUR PRECEDING),
 R1 AS (PARTITION BY "ticker_symbol" ROWS 1 PRECEDING)
```

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

How Do We Aggregate Streaming Data?

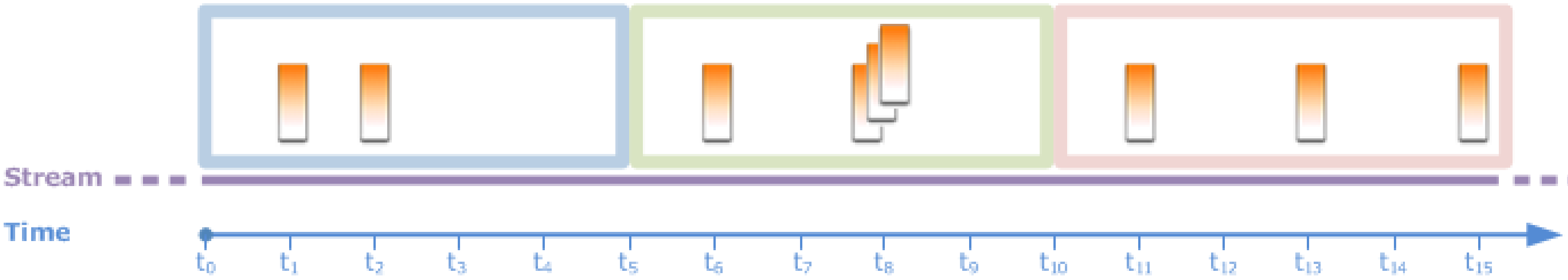
- Aggregations (count, sum, min,…) take granular real time data and turn it into insights
- Data is continuously processed so you need to tell the application when you want results

Windows!

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

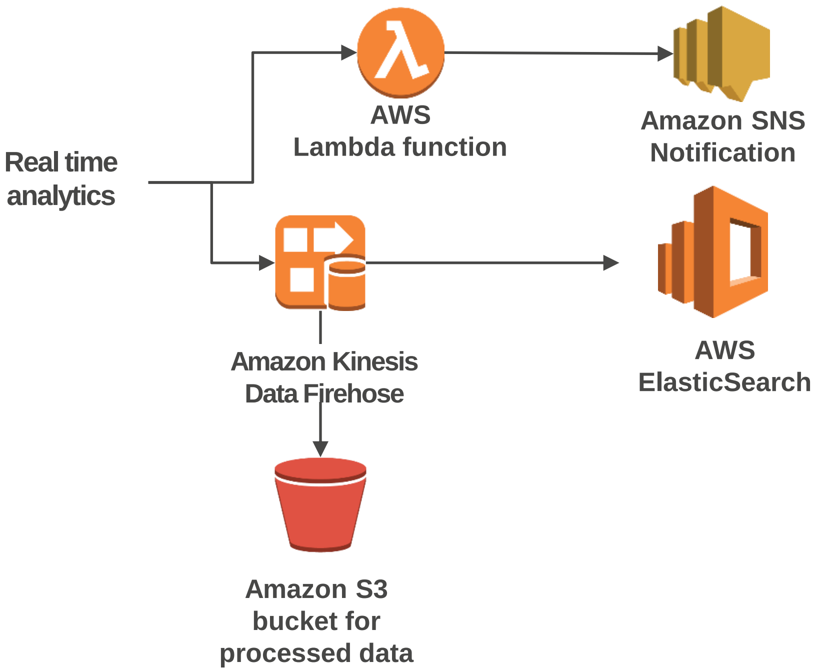
Window Types

- Sliding, tumbling, and custom windows
- Tumbling windows are fixed size and grouped keys do not overlap



© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Real-Time Notification & Dashboard



- Use Kinesis Data Firehose to archive processed data in S3
- Use AWS Lambda to generate notification
- Open source or other tools to visualize the data

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Where To Go Next?

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Getting Started

- Kinesis Home Page: https://aws.amazon.com/kinesis/
- Kinesis Blog Post: https://aws.amazon.com/kinesis/blog-posts/
- Getting Started Page: https://aws.amazon.com/kinesis/getting-started/

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

THANK YOU!

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.