



AWS
re:Invent

ANT407 - R

Real-time apps with Amazon Kinesis Data Analytics and Apache Flink

Kamal Lanka

Software Development Manager
Amazon Web Services

Agenda

- Challenges of running streaming applications
- Apache Flink
- Configure and deploy an Apache Flink application using Amazon Kinesis Data Analytics

Challenges of running streaming applications

- **Availability**
 - Much higher requirements on the availability of the streaming applications when compared to traditional batch-based approach
- **State management**
 - Computations of a streaming application often rely on internal state that can be corrupted or even lost when the application fails
- **Scaling**
 - When the load increases, the infrastructure that supports the streaming application must scale to keep the application from becoming overloaded
 - When the load decreases, the infrastructure should scale down to remain cost effective by not provisioning more resources than are needed

Apache Flink

- Apache Flink is an open-source project that is tailored to stateful computations over unbounded and bounded datasets
 - Support for APIs (including Java and SQL) and rich time semantics
 - In-depth state management capabilities
 - Provides exactly-once processing support
 - Well suited for analyzing streaming data with low latency

Deploy and operate Apache Flink application with Kinesis Data Analytics for Java applications

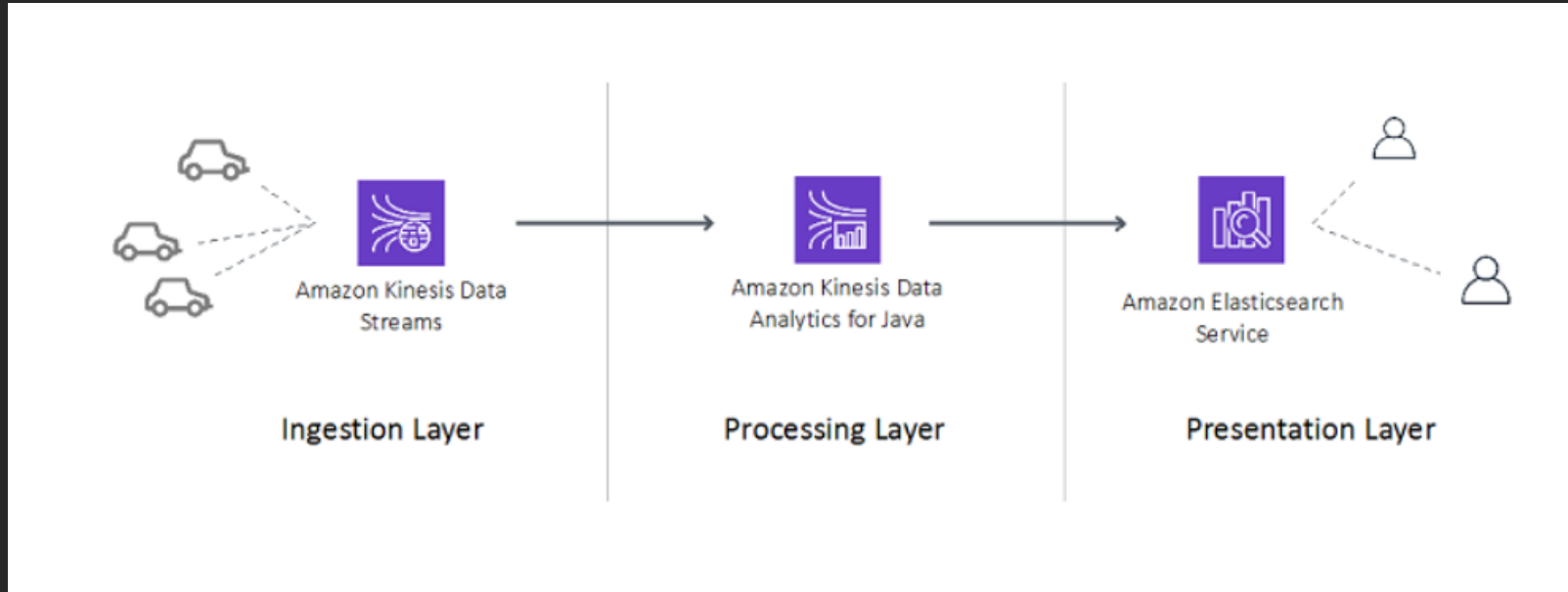
Scenario

- Analyze the telemetry data of a taxi fleet in New York City in near real time to optimize the fleet operation

Goal

- Identify areas that are currently requesting a high number of taxi rides

Architecture



- Ingestion layer: Kinesis data stream serves as a buffer that decouples the producers from the consumers
- Processing layer: Kinesis data analytics for Java processes the data in the Kinesis data stream
- Presentation layer: Persists the processed data into Amazon ES

Benefits of the architecture

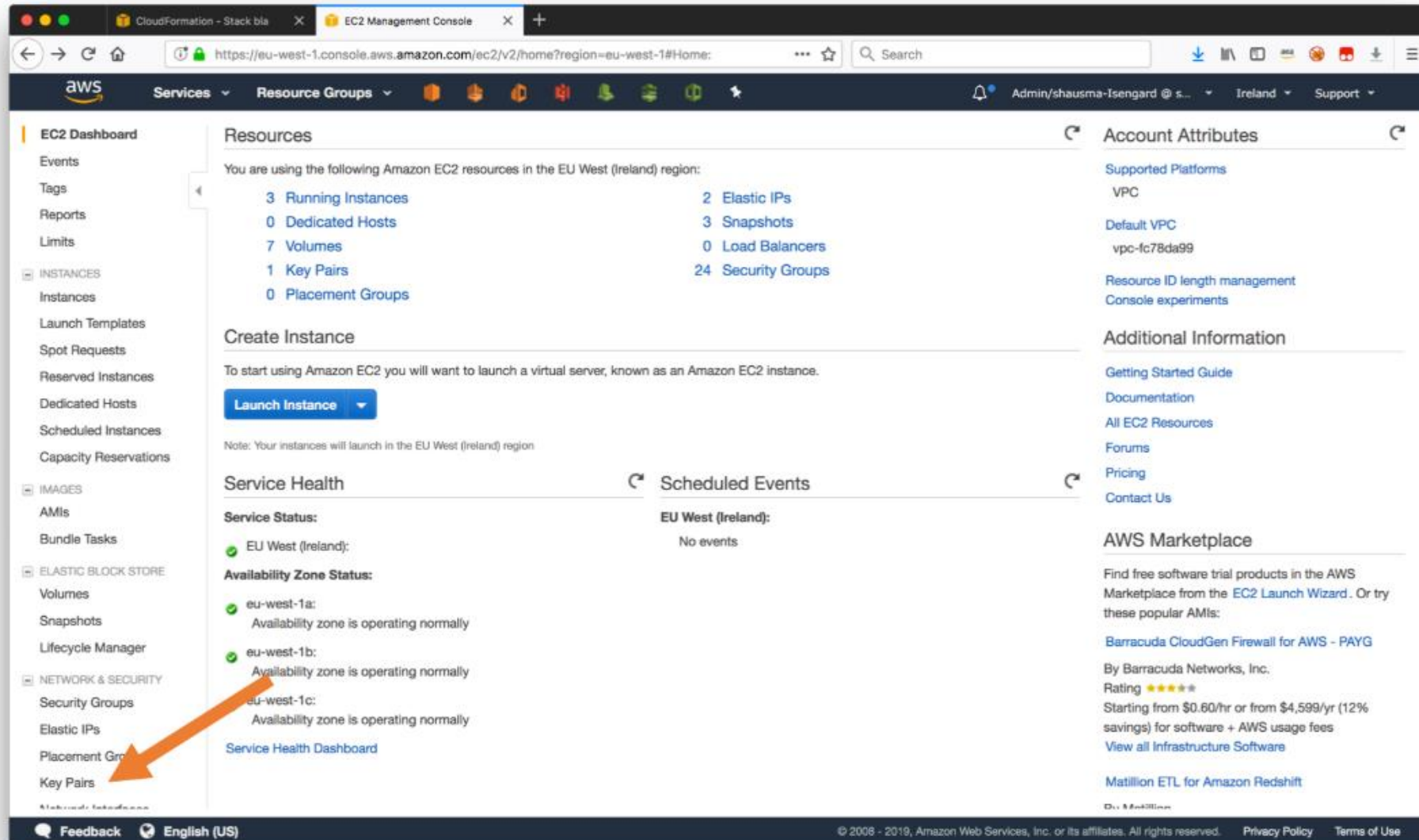
- Separating the different aspects of the architecture into ingestion, processing, and presentation nicely decouples different components
- Components can be scaled independently
- Allows you to experiment and adopt new technologies in the future

Let's kick the tires!

Prerequisites

- Access to an AWS account with **AdministratorAccess**

Create EC2 Key Pair



The screenshot displays the AWS Management Console for the EC2 service in the EU West (Ireland) region. The left-hand navigation pane shows the 'EC2 Dashboard' with various links. An orange arrow points to the 'Key Pairs' link under the 'NETWORK & SECURITY' section. The main content area shows the 'Resources' section with a list of EC2 resources: 3 Running Instances, 0 Dedicated Hosts, 7 Volumes, 1 Key Pairs, 0 Placement Groups, 2 Elastic IPs, 3 Snapshots, 0 Load Balancers, and 24 Security Groups. Below this is the 'Create Instance' section with a 'Launch Instance' button. The 'Service Health' section shows the status of the EU West (Ireland) region and its availability zones. The 'Scheduled Events' section shows no events. The right-hand sidebar contains 'Account Attributes' and 'Additional Information' links.

EC2 Dashboard

- Events
- Tags
- Reports
- Limits
- INSTANCES
 - Instances
 - Launch Templates
 - Spot Requests
 - Reserved Instances
 - Dedicated Hosts
 - Scheduled Instances
 - Capacity Reservations
- IMAGES
 - AMIs
 - Bundle Tasks
- ELASTIC BLOCK STORE
 - Volumes
 - Snapshots
 - Lifecycle Manager
- NETWORK & SECURITY
 - Security Groups
 - Elastic IPs
 - Placement Groups
 - Key Pairs

Resources

You are using the following Amazon EC2 resources in the EU West (Ireland) region:

- 3 Running Instances
- 0 Dedicated Hosts
- 7 Volumes
- 1 Key Pairs
- 0 Placement Groups
- 2 Elastic IPs
- 3 Snapshots
- 0 Load Balancers
- 24 Security Groups

Create Instance

To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

[Launch Instance](#)

Note: Your instances will launch in the EU West (Ireland) region.

Service Health

Service Status:

- EU West (Ireland): ✓

Availability Zone Status:

- eu-west-1a: ✓ Availability zone is operating normally
- eu-west-1b: ✓ Availability zone is operating normally
- eu-west-1c: ✓ Availability zone is operating normally

[Service Health Dashboard](#)

Scheduled Events

EU West (Ireland):

- No events

Account Attributes

[Supported Platforms](#)

- VPC

[Default VPC](#)

- vpc-fc78da99

[Resource ID length management](#)

[Console experiments](#)

Additional Information

- [Getting Started Guide](#)
- [Documentation](#)
- [All EC2 Resources](#)
- [Forums](#)
- [Pricing](#)
- [Contact Us](#)

AWS Marketplace

Find free software trial products in the AWS Marketplace from the [EC2 Launch Wizard](#). Or try these popular AMIs:

- [Barracuda CloudGen Firewall for AWS - PAYG](#)

By Barracuda Networks, Inc.

Rating ★★★★★

Starting from \$0.60/hr or from \$4,599/yr (12% savings) for software + AWS usage fees

[View all Infrastructure Software](#)

[Matillion ETL for Amazon Redshift](#)

[Dr. Matillion](#)

Feedback **English (US)**

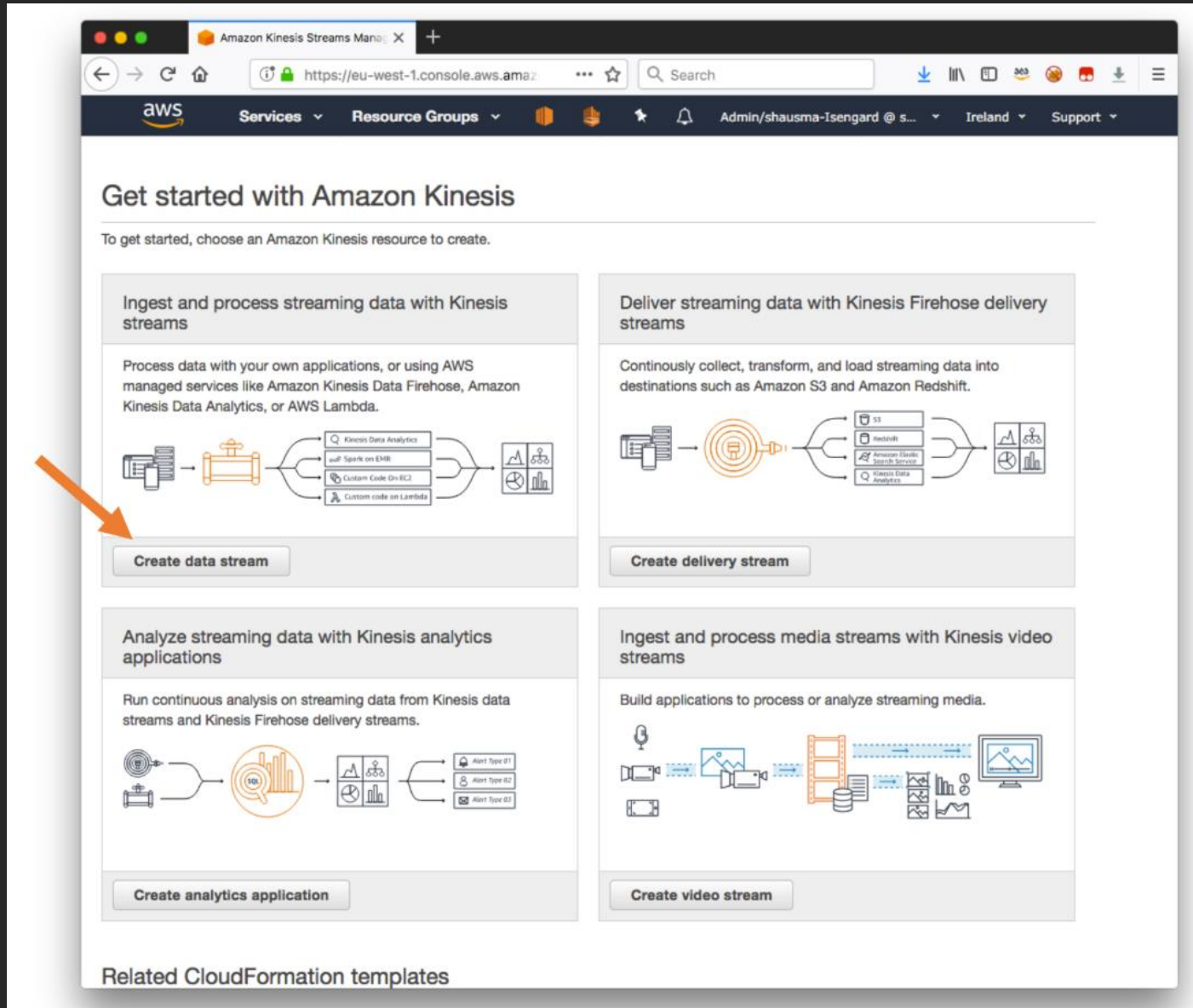
© 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy Policy](#) [Terms of Use](#)

Run AWS CloudFormation

- Execute [AWS CloudFormation](#) template in your own AWS account
 - This steps will take up to 10 minutes
 - Build and generate Kinesis Data Analytics Apache Flink Jar file
 - Creates Amazon ES cluster for presentation layer
 - Provisions an EC2 instance to ingest data
- Navigate to the **Outputs** section of the CloudFormation template and take a note of the outputs. We will need them to complete the subsequent steps.

Ingestion layer

Create Kinesis Data Stream—taxi-trip-events



The screenshot displays the Amazon Kinesis Streams console interface. The top navigation bar includes the AWS logo, 'Services', 'Resource Groups', and user information. The main heading is 'Get started with Amazon Kinesis'. Below this, a subheading reads 'To get started, choose an Amazon Kinesis resource to create.'.

Four main sections are visible, each with a diagram and a 'Create' button:

- Ingest and process streaming data with Kinesis streams:** The diagram shows data being ingested and then processed by various services including Kinesis Data Analytics, Amazon EMR, Amazon EC2, and AWS Lambda. An orange arrow points to the 'Create data stream' button.
- Deliver streaming data with Kinesis Firehose delivery streams:** The diagram shows data being collected, transformed, and loaded into destinations like Amazon S3 and Amazon Redshift. The button is 'Create delivery stream'.
- Analyze streaming data with Kinesis analytics applications:** The diagram shows data being analyzed and then triggering alerts. The button is 'Create analytics application'.
- Ingest and process media streams with Kinesis video streams:** The diagram shows media being ingested and processed. The button is 'Create video stream'.

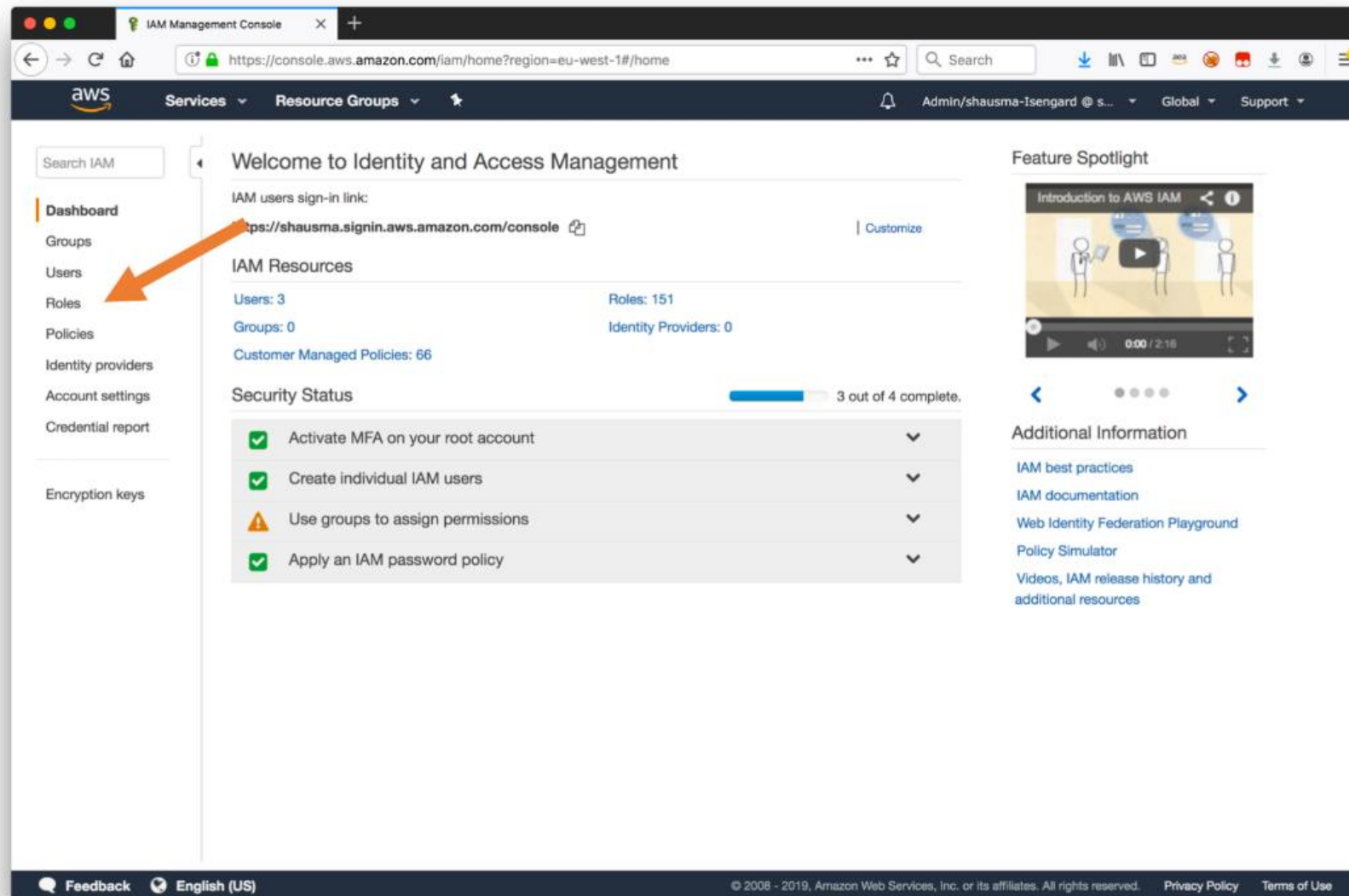
At the bottom, there is a section for 'Related CloudFormation templates'.

Processing layer

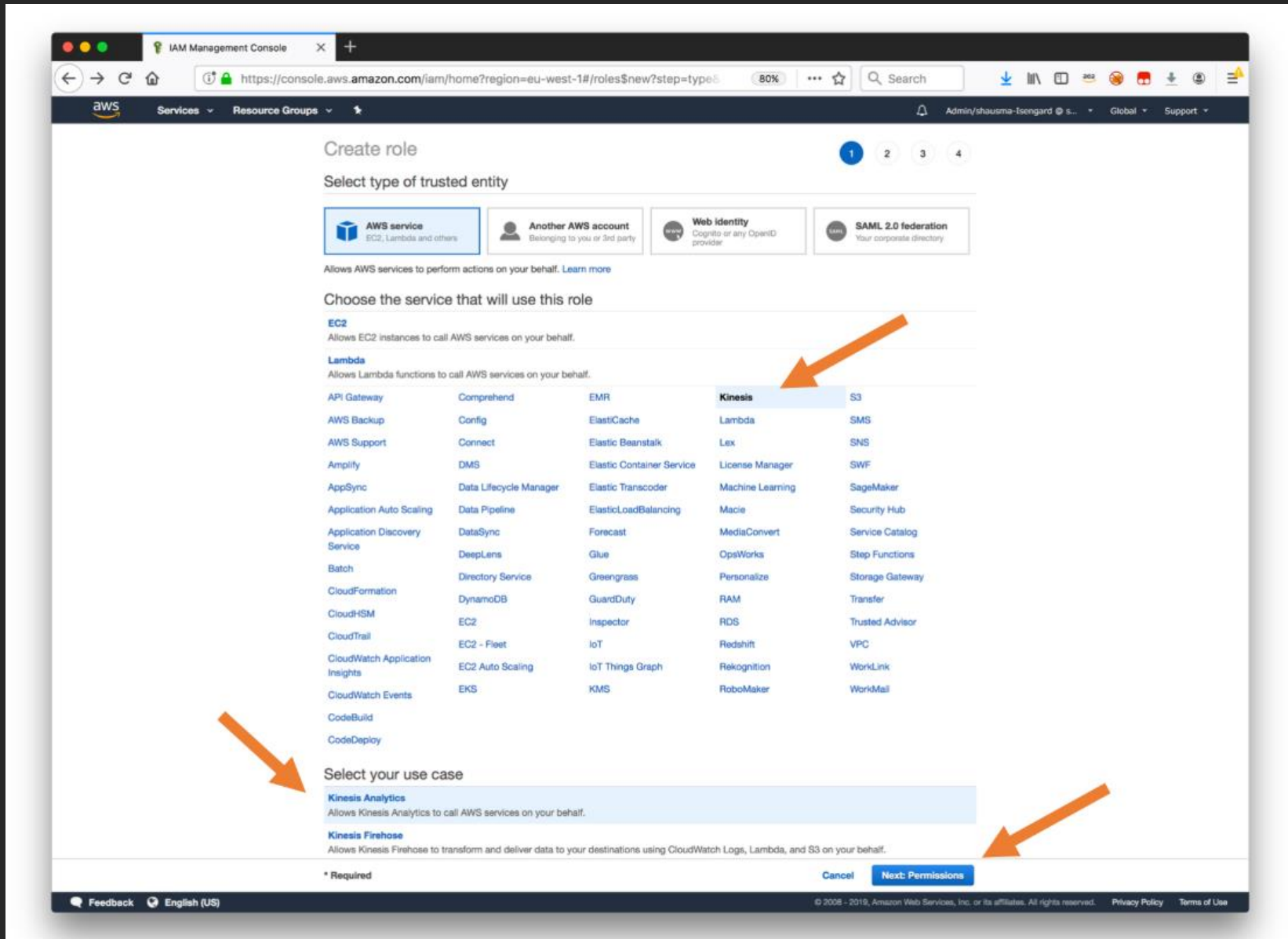
Steps

1. Create IAM role to call services
2. Create Kinesis Data Analytics application

Step 1: Create IAM role to call services



Step 1: Create IAM role to call services



Step 1: Create IAM role to call services

The screenshot shows the AWS IAM Management Console interface for creating a new role. The page is titled 'Create role' and is at step 2 of a 4-step process. The first step, 'Attach permissions policies', is active. Below the title, there is a section 'Attach permissions policies' with a sub-header 'Choose one or more policies to attach to your new role.' and a 'Create policy' button. A table of available policies is displayed, with 'AdministratorAccess' selected. An orange arrow points to the 'Next: Tags' button at the bottom right of the page.

Create role

1 2 3 4

▼ Attach permissions policies

Choose one or more policies to attach to your new role.

Create policy

Filter policies Search Showing 588 results

	Policy name	Used as	Description
<input checked="" type="checkbox"/>	AdministratorAccess	Permissions policy (5)	Provides full access to AWS services an...
<input type="checkbox"/>	AlexaForBusinessDeviceSetup	None	Provide device setup access to AlexaFor...
<input type="checkbox"/>	AlexaForBusinessFullAccess	None	Grants full access to AlexaForBusiness r...
<input type="checkbox"/>	AlexaForBusinessGatewayExecution	None	Provide gateway execution access to AI...
<input type="checkbox"/>	AlexaForBusinessNetworkProfileServicePolicy	None	This policy enables Alexa for Business to...
<input type="checkbox"/>	AlexaForBusinessReadOnlyAccess	None	Provide read only access to AlexaForBu...
<input type="checkbox"/>	AllowStorageGatewayAssumeBucketAccessR...	Permissions policy (1)	Allow storage gateway to access: shaus...
<input type="checkbox"/>	AmazonAPIGatewayAdministrator	None	Provides full access to create/edit/delete...

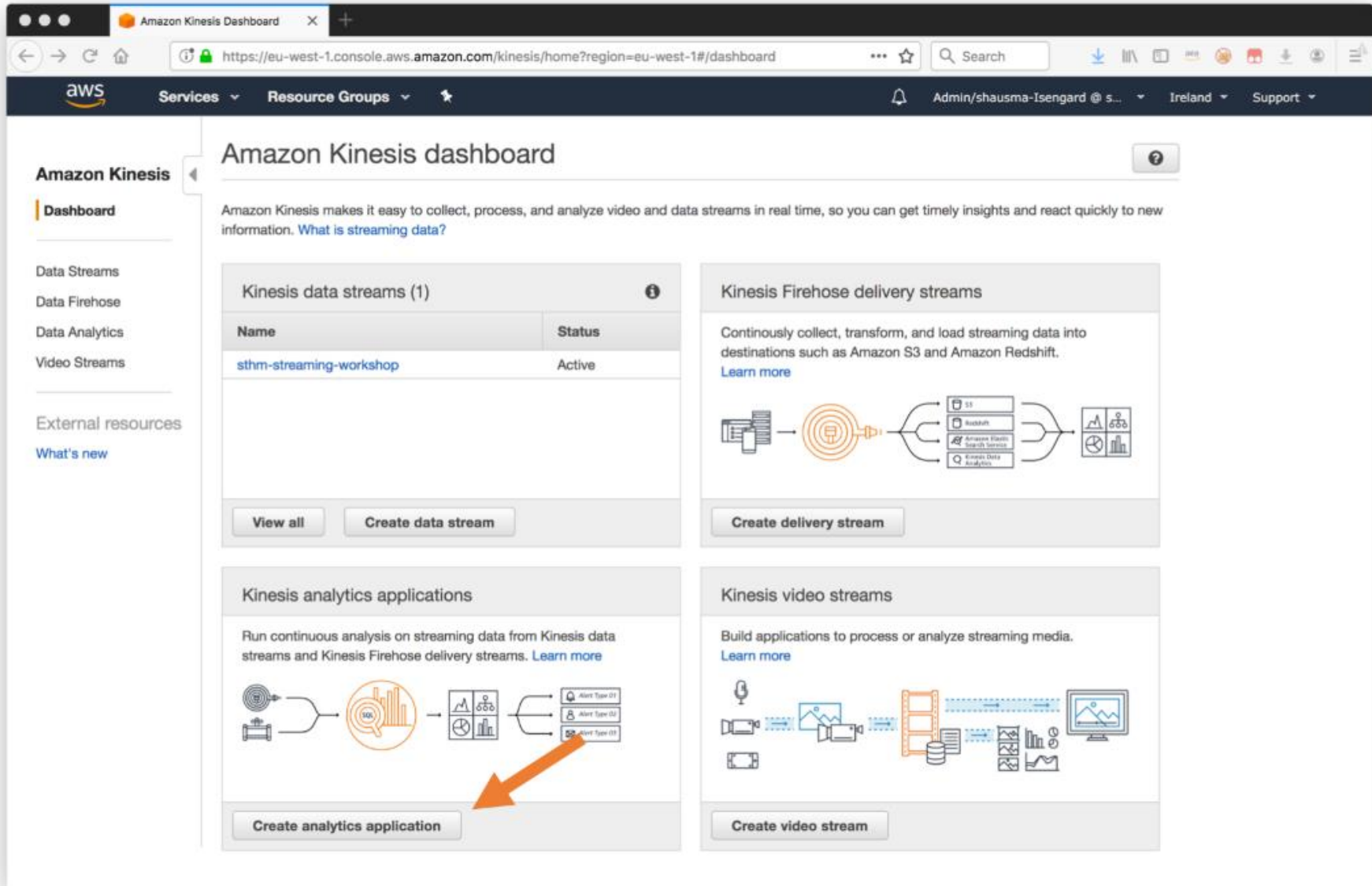
► Set permissions boundary

* Required

Cancel Previous **Next: Tags**

Feedback English (US) © 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Step 2: Create Kinesis Data Analytics application



The screenshot displays the Amazon Kinesis dashboard in a web browser. The browser's address bar shows the URL `https://eu-west-1.console.aws.amazon.com/kinesis/home?region=eu-west-1#/dashboard`. The dashboard header includes the AWS logo, navigation links for Services, Resource Groups, and a user profile for Admin/shausma-Isengard. A left-hand sidebar lists navigation options: Dashboard (selected), Data Streams, Data Firehose, Data Analytics, Video Streams, External resources, and What's new.

The main content area is titled "Amazon Kinesis dashboard" and includes a brief description of Kinesis. It is divided into four primary sections:

- Kinesis data streams (1)**: A table listing existing data streams. The table has two columns: "Name" and "Status". One stream, "sthm-streaming-workshop", is listed with a status of "Active". Below the table are buttons for "View all" and "Create data stream".
- Kinesis Firehose delivery streams**: A section explaining how to continuously collect, transform, and load streaming data into destinations like Amazon S3 and Amazon Redshift. It includes a diagram showing data flow from a source through a Firehose delivery stream to various storage and analytics destinations. A "Create delivery stream" button is at the bottom.
- Kinesis analytics applications**: A section for running continuous analysis on streaming data. It features a diagram showing data from a stream being processed by an analytics application and then triggering alerts. An orange arrow points from the "Create analytics application" button to the diagram. A "Create analytics application" button is at the bottom.
- Kinesis video streams**: A section for building applications to process or analyze streaming media. It includes a diagram showing video data being processed and analyzed. A "Create video stream" button is at the bottom.

Step 2: Create Kinesis Data Analytics application

Amazon Kinesis Analytics

Services Resource Groups

Admin/shausma-Isengard @ s... Ireland Support

Amazon Kinesis

Dashboard

Data Streams

Data Firehose

Data Analytics

Video Streams

External resources

What's new

Kinesis Analytics - Create application

Kinesis Analytics applications continuously read and analyze data from a connected streaming source in real-time. To enable interactivity with your data during configuration you will be prompted to run your application. Kinesis Analytics resources are not covered under the [AWS Free Tier](#), and **usage-based charges apply**. For more information, see [Kinesis Analytics pricing](#).

Application name* sthm-streaming-workshop

Description

Run on

☐ SQL

☒ Apache Flink 1.6

Access permissions

Create or choose IAM role with the required permissions. Kinesis Analytics will assume this role when you later execute your code. [Learn more](#)

Access permissions* ☐ Create / update IAM role **kinesis-analytics-sthm-streaming-workshop-eu-west-1**

☒ Choose from IAM roles that Kinesis Analytics can assume

IAM role* sthm-streaming-workshop

[View sthm-streaming-workshop in IAM](#)

Only IAM roles with the [required trust policy](#) attached are available for selection.

* Required

Cancel Create application

Feedback English (US)

© 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Step 2: Create Kinesis Data Analytics application

Amazon Kinesis

- Dashboard
- Data Streams
- Data Firehose
- Data Analytics**
- Video Streams
- External resources
- What's new

Configure application

In this section, you specify your application's code location, IAM role, application properties, snapshots, and monitoring.

Code location

Choose the S3 bucket that contains your application's JAR file. Your application code is reloaded from the S3 bucket every time you update your application using this page. To update other application settings without reloading the application code, use the AWS CLI. [Learn more](#)

Amazon S3 bucket* [Refresh](#) [Create new](#)

[View tech-summit-02-artifactbucket-vfzls1gu26no in Amazon S3](#)

Path to Amazon S3 object*

Specify the path and object name of the JAR file containing your application code.

[View specified object in Amazon S3](#)

Access to application resources

Access permissions*

- ☐ Create / update IAM role `kinesis-analytics-sthm-streaming-workshop-eu-west-1`
- ☒ Choose from IAM roles that Kinesis Analytics can assume

IAM role* [Refresh](#)

[View sthm-streaming-workshop in IAM](#)

Only IAM roles with the [required trust policy](#) attached are available for selection.

► Properties
► Snapshots
► Monitoring

* Required

[Cancel](#) [Update](#)

Feedback English (US)

© 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy Policy](#) [Terms of Use](#)

Step 2: Create Kinesis Data Analytics application

- Expand the Properties section
 - Enter Group ID : FlinkApplicationProperties
 - Add two key/value pairs:
 - InputStreamName with the name of the Kinesis stream you've created earlier, e.g., taxi-trip-events
 - ElasticsearchEndpoint with the correct Amazon ES https endpoint that can be obtained from the Output section of the CloudFormation template under ElasticsearchEndpoint

Step 2: Create Kinesis Data Analytics application

- Finally, press the blue Update button at the bottom of the page to update the properties of the application
- Once the update has completed, press Run on the resulting page and confirm that you want to run the application by choosing Run again

Ingest data

Ingest data into a Kinesis Data Stream

- Connect to the EC2 instance via browser SSH
- Once the connection has been established, start ingesting events into the Kinesis data stream by executing the jar file that has already been downloaded to the EC2 instance
 - `java -jar amazon-kinesis-replay-1.0-SNAPSHOT.jar -streamRegion us-east-1 -speedup 1000 -streamName taxi-trip-events`

Visualizing data

Analyze the data

- Navigate to the Kibana dashboard—the URL can be obtained from the Output section of the CloudFormation template under KibanaDashboard
- The Kibana dashboard contains a heatmap and a line graph
 - The heatmap visualizes locations where taxis are currently requested
 - The line graph visualizes the average trip duration between to these two airports

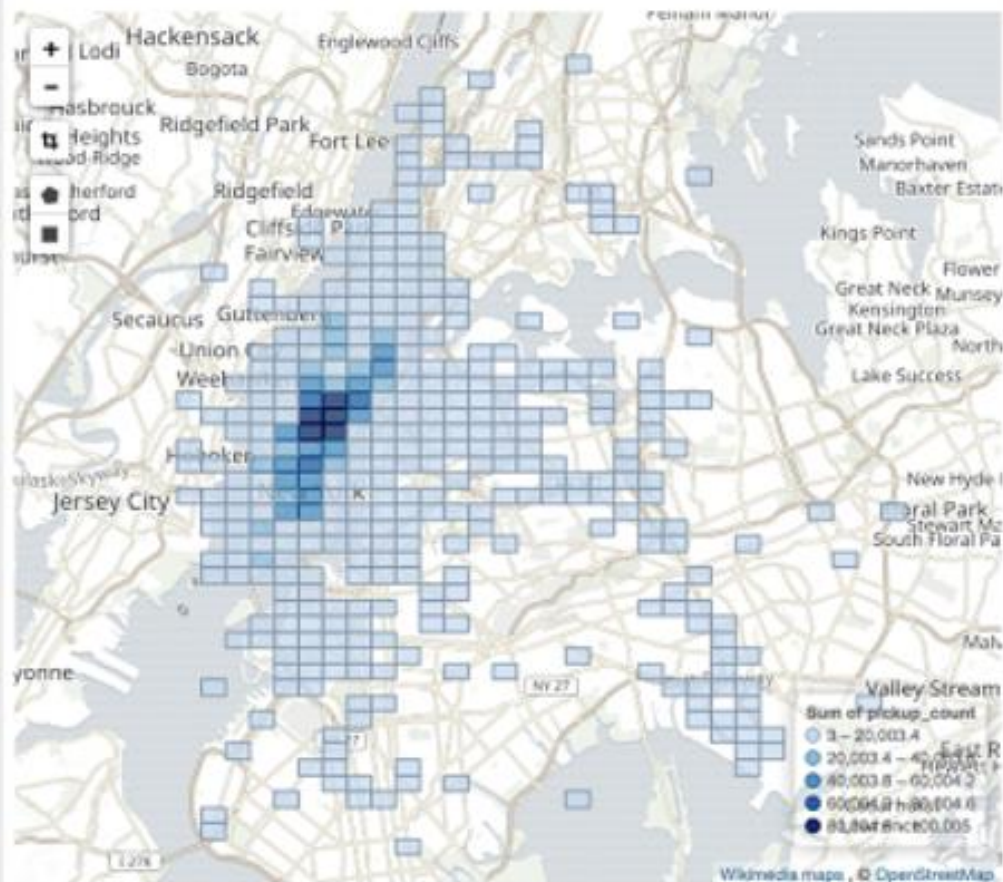
Search... (e.g. status:200 AND extension:PHP)

Options

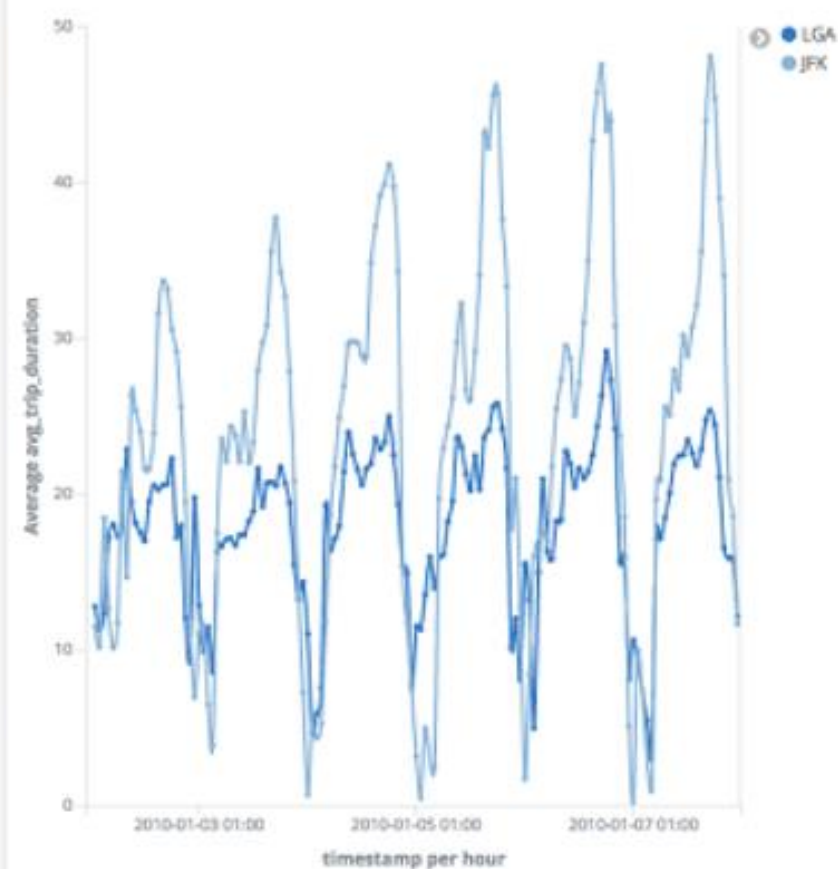


Add a filter +

pickup_location



avg_trip_duration



Processing layer Java code walkthrough

Let's step back for a moment and review what you just did

- You created a fully managed, highly available, scalable streaming architecture
- You ingested and analyzed up to 10k events per second
 - Try to imagine what it would have taken you to build something similar from scratch

Cleanup

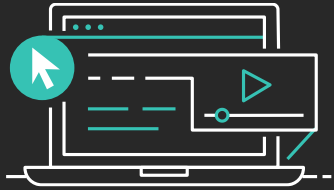
Cleanup

- Delete the Kinesis Data Analytics Java app
- Delete the Kinesis stream
- Delete the IAM role you created earlier
- Delete the CloudFormation stack. If it fails, delete the S3 bucket and delete the stack again.

Q&A

Learn big data with AWS Training and Certification

Resources created by the experts at AWS to help you build and validate data analytics skills



New free digital course, Data Analytics Fundamentals, introduces Amazon S3, Amazon Kinesis, Amazon EMR, AWS Glue, and Amazon Redshift



Classroom offerings, including Big Data on AWS, feature AWS expert instructors and hands-on labs



Validate expertise with the **AWS Certified Big Data - Specialty** exam or the new **AWS Certified Data Analytics - Specialty** beta exam

Visit aws.amazon.com/training/paths-specialty/

Thank you!



Please complete the session
survey in the mobile app.