Amazon Web Services
Data Engineering Immersion Day

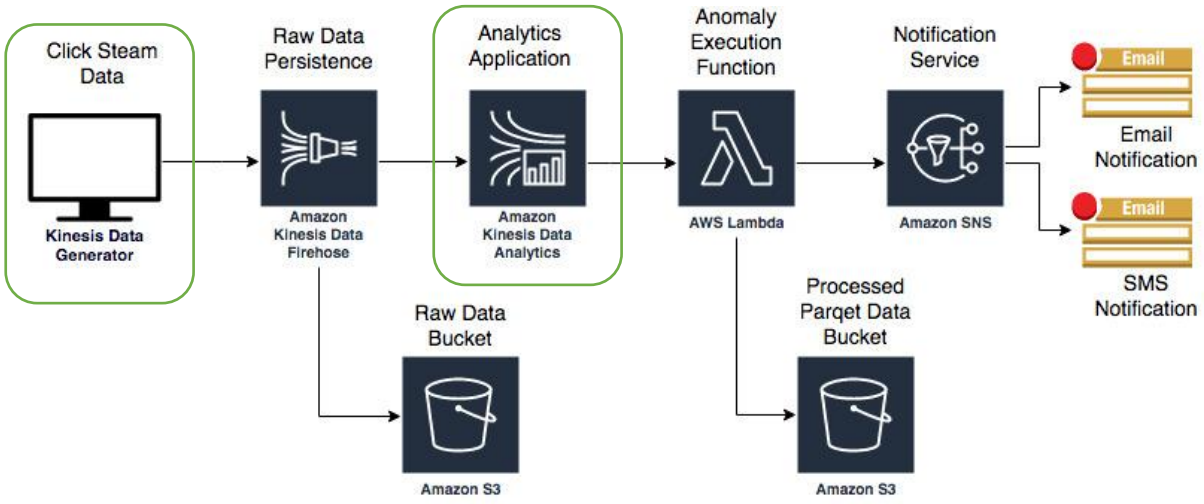
Lab 1 - Prelab. Real-Time Clickstream Anomaly Detection

August 2020

Table of Contents

Introduction

This guide will help you set up the pre-lab environment for the Real-Time Clickstream Anomaly Detection Amazon Kinesis Data Analytics lab.



After you deploy the CloudFormation template, sign into your account to view the following resources:

- Two Amazon Simple Storage Service (Amazon S3) buckets: You will use these buckets to persist raw and processed data.
- One AWS Lambda function: This Lambda function will be triggered once an anomaly has been detected.
- Amazon Simple Notification Service (Amazon SNS) topic with an email and phone number subscribed to it: The Lambda function will publish to this topic once an anomaly has been detected.
- Amazon Cognito User credentials: You will use these user credentials to log into the Kinesis Data Generator to send records to our Amazon Kinesis Data Firehose.

Today, you are attending a formal AWS event, so we provide an AWS account to you. If in the future you might want to perform these labs in your own AWS environment by yourself, suggest you to save this file to your computer for the future reuse. Alternatively, run the lab again by following the online instruction here - https://aws-dataengineering-day.workshop.aws/en/300.html

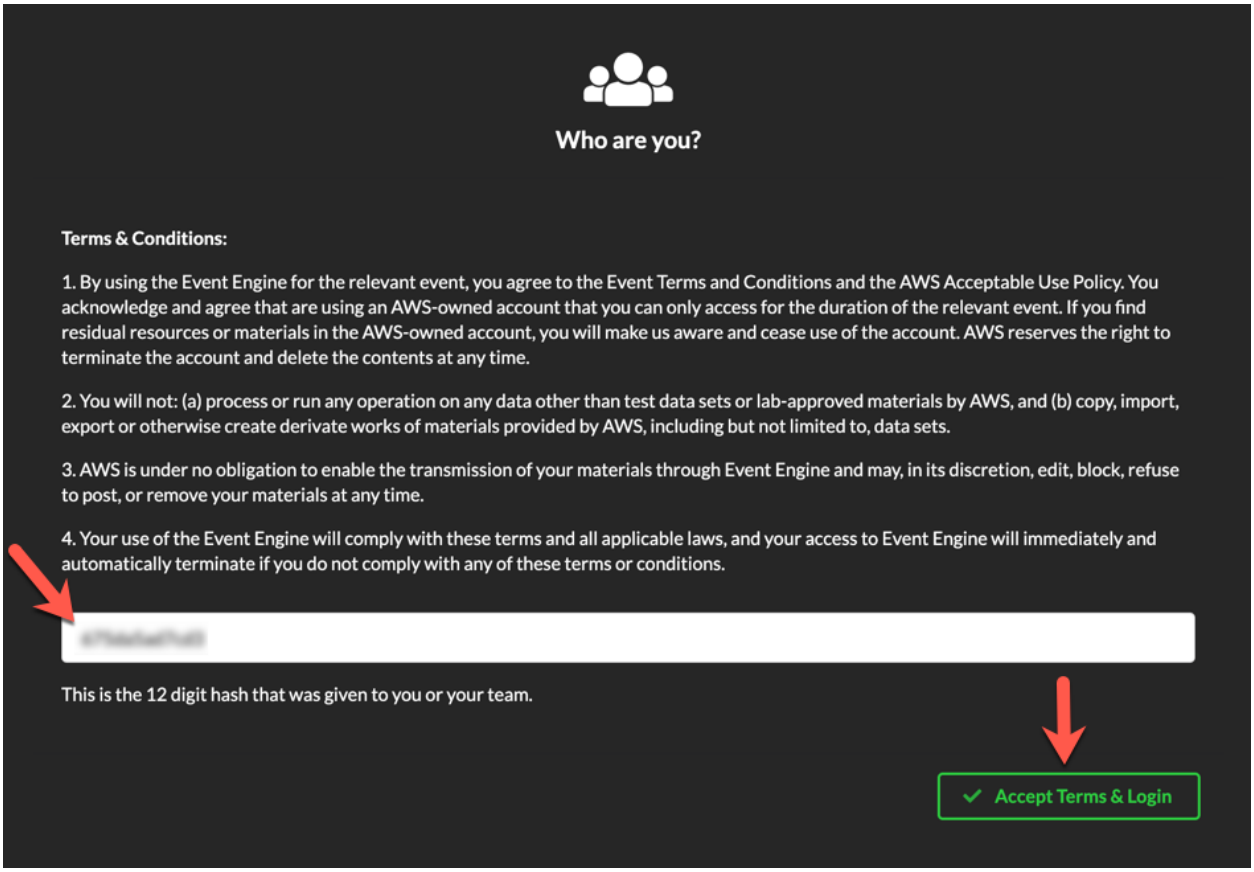
Get Started Using the Lab Environment

Please skip this section if you are running the lab on your own AWS account.

Today, you are attending a formal event and you will have been sent your access details beforehand. If in the future you might want to perform these labs in your own AWS environment by yourself, you can follow instructions on GitHub - https://github.com/aws-samples/data-engineering-for-aws-immersion-day.

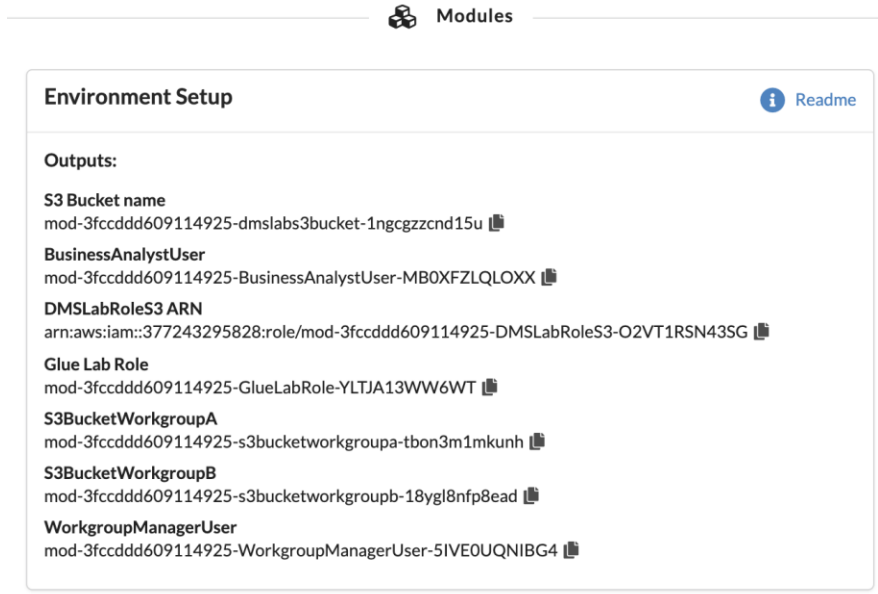
A 12-character access code (or 'hash') is the access code that grants you permission to use a dedicated AWS account for the purposes of this workshop.

1. Go to https://dashboard.eventengine.run/, enter the access code and click Proceed:

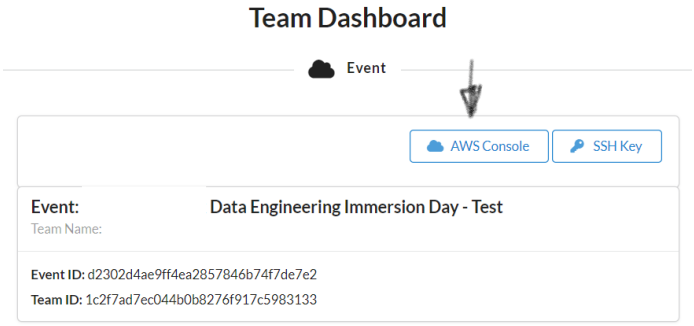


2. On the Team Dashboard web page you will see a set of parameters that you will need during the labs. Best to save them to a text file locally, alternatively you can always go to this page to review them. Replace the parameters with the corresponding values from here where indicated in subsequent labs:

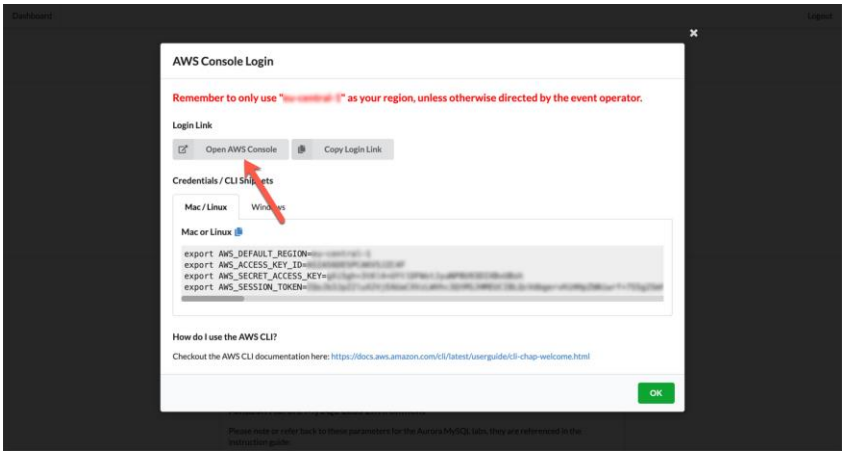
Because you're at a formal event, some AWS resources have been pre-deployed for your convenience, for example you can see a list of parameters on your event dashboard:



3. On the Team Dashboard, please click AWS Console to log into the AWS Management Console:



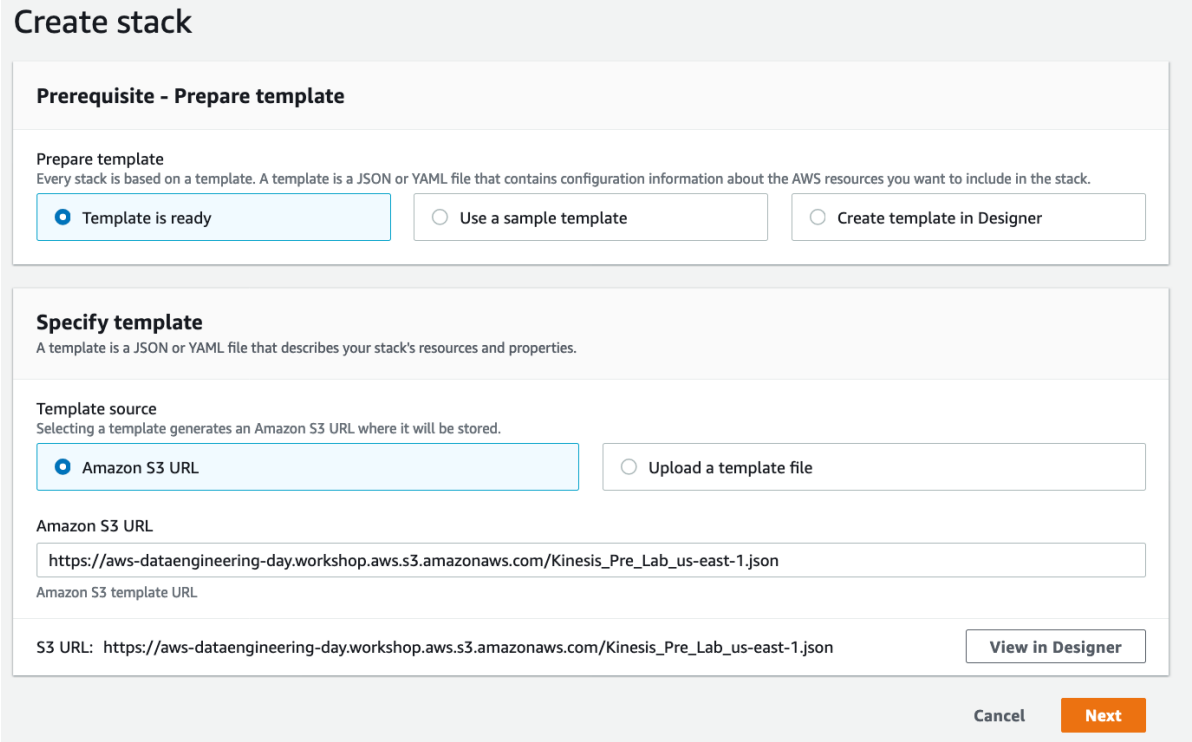
4. Click Open Console. For the purposes of this workshop, you will not need to use command line and API access credentials



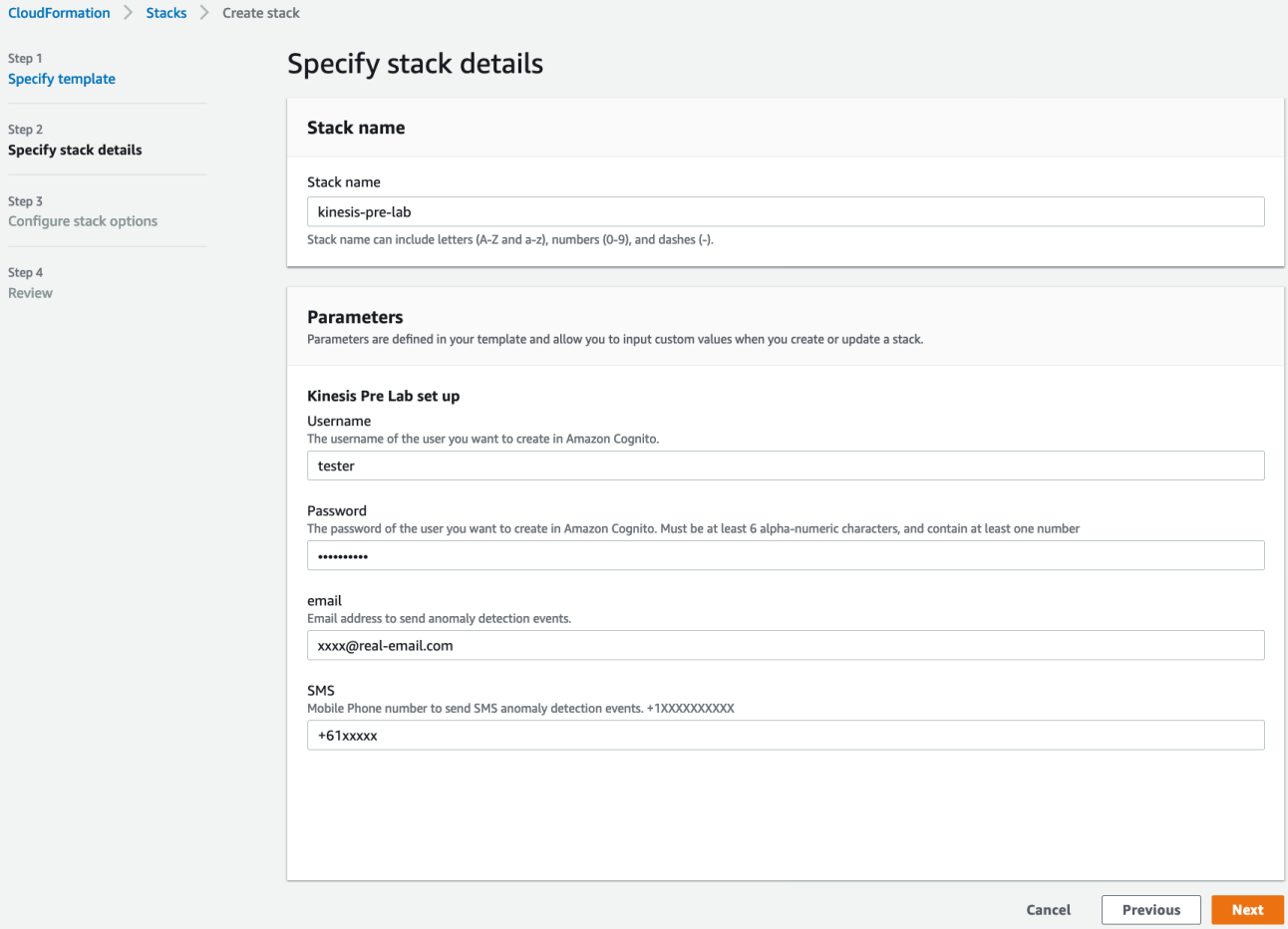
Once you have completed these steps, you can continue with the rest of this lab

CloudFormation Stack Deployment

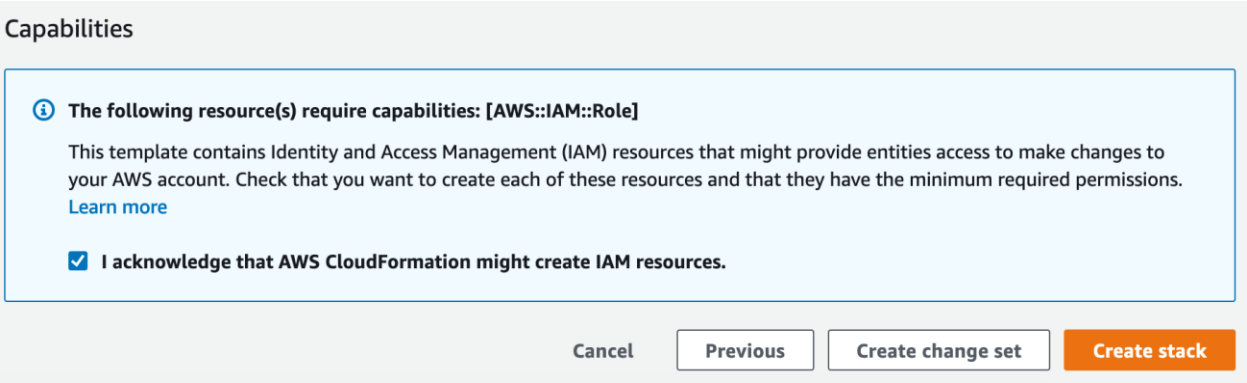
1. Use this link to create a new CloudFormation Stack: https://console.aws.amazon.com/cloudformation/home?region=us-east-1#/stacks/create/template?stackName=kinesis-pre-lab&templateURL=https://aws-dataengineering-day.workshop.aws.s3.amazonaws.com/Kinesis_Pre_Lab_us-east-1.json



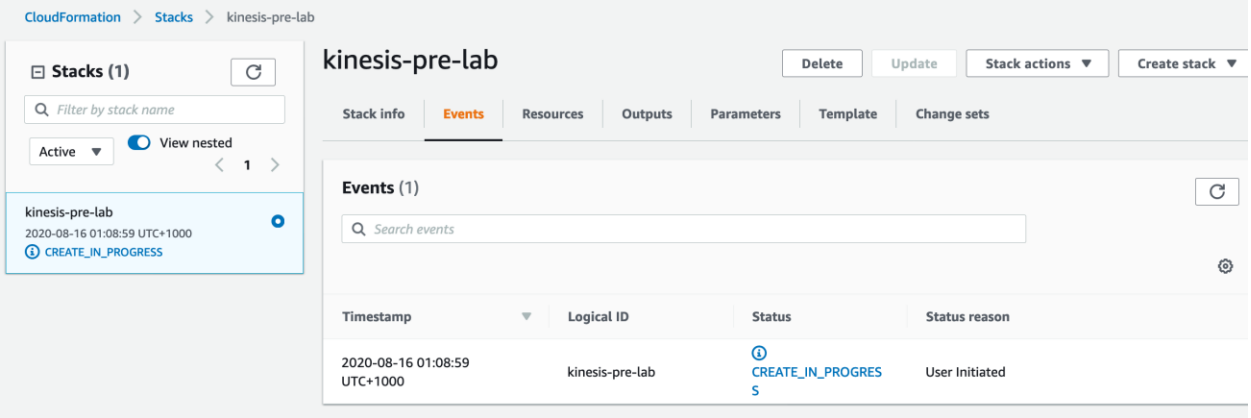
2. Click **Next** at the bottom of the page in as shown in above screenshot**.**
3. **In the Parameters** section**,** fill the following fields as shown in screenshot:
    - **Username:** This is your username to login to the Kinesis Data Generator
    - **Password:** This is your password for the Kinesis Data Generator. The password must be at least 6 alpha-numeric characters and contain at least one number and a capital letter.
    - **Email:** Type an email address that you can access. The SNS topic sends a confirmation to this address.
    - **SMS:** Type a phone number (+1XXXXXXXXX) where you can receive texts from the SNS topic.



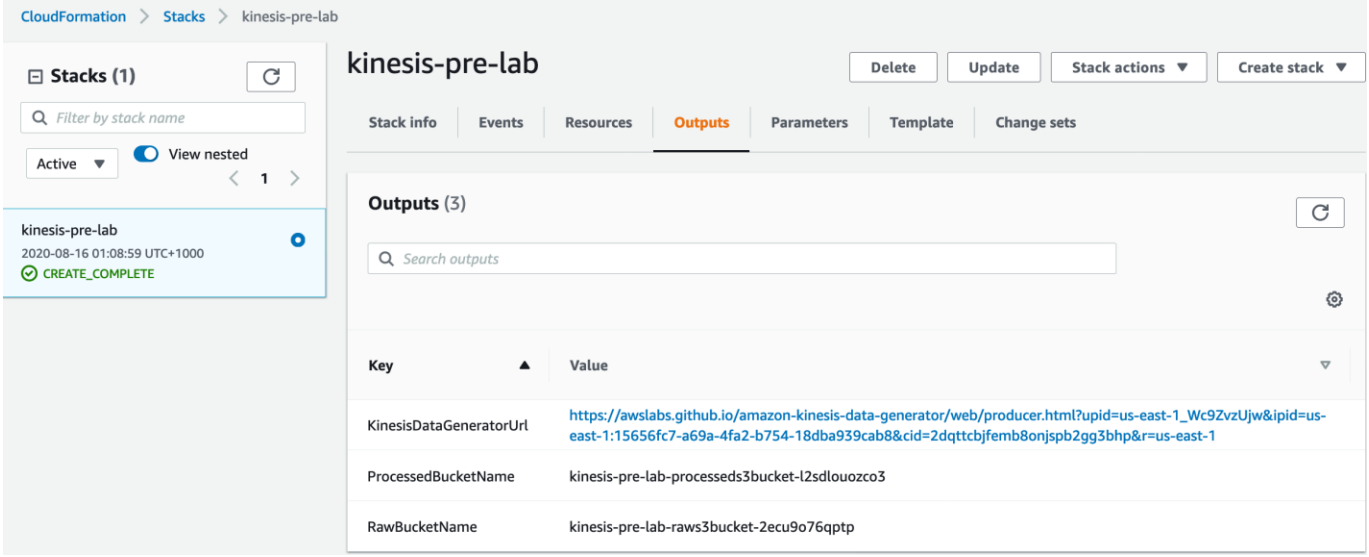
4. Keep the rest of setup as default.
5. In the last **Review kinesis-pre-lab** page, select the check box marked **I acknowledge that AWS CloudFormation might create IAM resources**.



6. Click **Create**. CloudFormation redirects you to your existing stacks.



7. Once your stack is deployed, click the **Outputs** tab to view more information:
    - **KinesisDataGeneratorUrl:** This value is the Kinesis Data Generator (KDG) URL.
    - **RawBucketName –** Store raw data coming from KDG.
    - **ProcessedBucketName –** Store transformed data



Congratulations! You are all done with the CloudFormation deployment.

Set up the Amazon Kinesis Data Generator

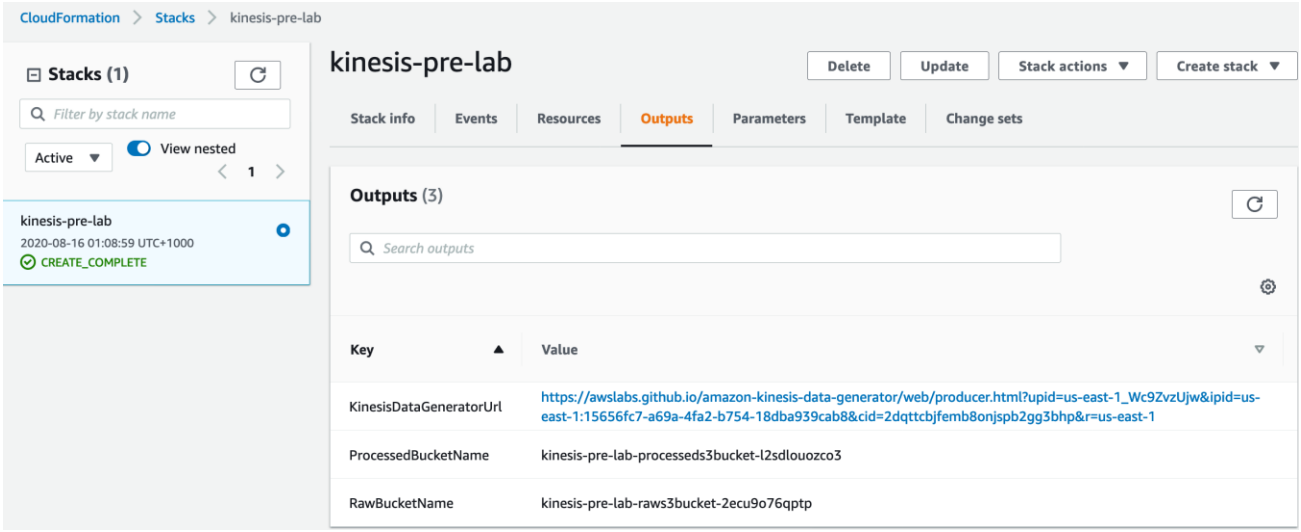
On the **Outputs** tab, notice the **Kinesis Data Generator URL**. Navigate to this URL to login into the Amazon Kinesis Data Generator (Amazon KDG).

The KDG simplifies the task of generating data and sending it to Amazon Kinesis. The tool provides a user-friendly UI that runs directly in your browser. With the KDG, you can do the following tasks:

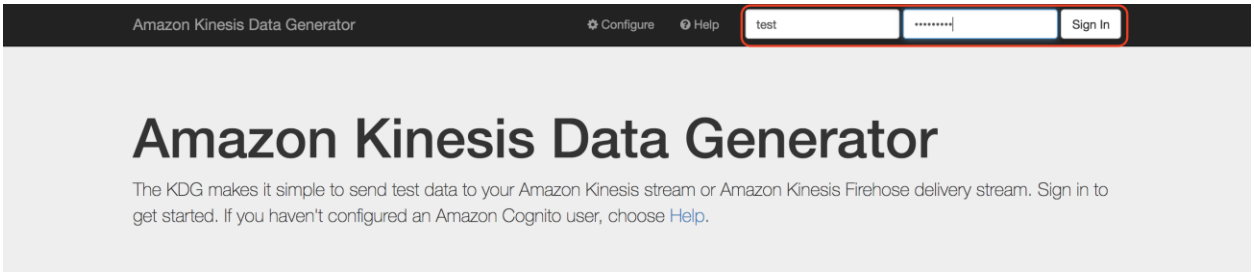
- Create templates that represent records for your specific use cases
- Populate the templates with fixed data or random data
- Save the templates for future use
- Continuously send thousands of records per second to your Amazon Kinesis stream or Firehose delivery stream

Let's test your Cognito user in the Kinesis Data Generator.

1. On the **Outputs** tab, click the **KinesisDataGeneratorUrl**.



2. Sign in using the **username** and **password** you entered in the CloudFormation console.



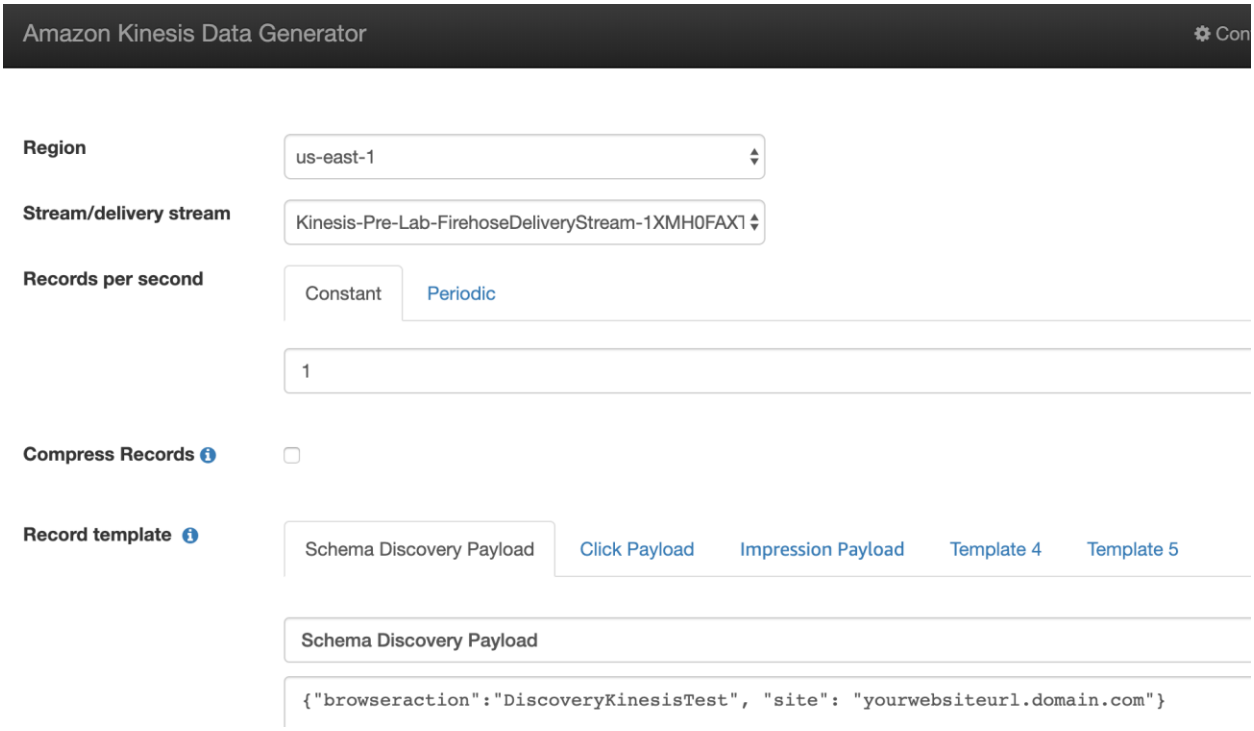
3. After you sign in, you should see the KDG console. You need to set up some templates to mimic the clickstream web payload.

- Create the following three templates. Copy the tab name highlight in bold letter and value as json string, refer screenshot:

Schema Discovery Payload
{"browseraction":"DiscoveryKinesisTest", "site": "yourwebsiteurl.domain.com"}
Click Payload
{"browseraction":"Click", "site": "yourwebsiteurl.domain.com"}
Impression Payload
{"browseraction":"Impression", "site": "yourwebsiteurl.domain.com"}

- Change Region to **US-EAST-1** and select a created Firehose Delivery Stream from the dropdown.
- Set **Records per second** to **1**

Your Amazon Kinesis Data Generator console should look similar to this example.

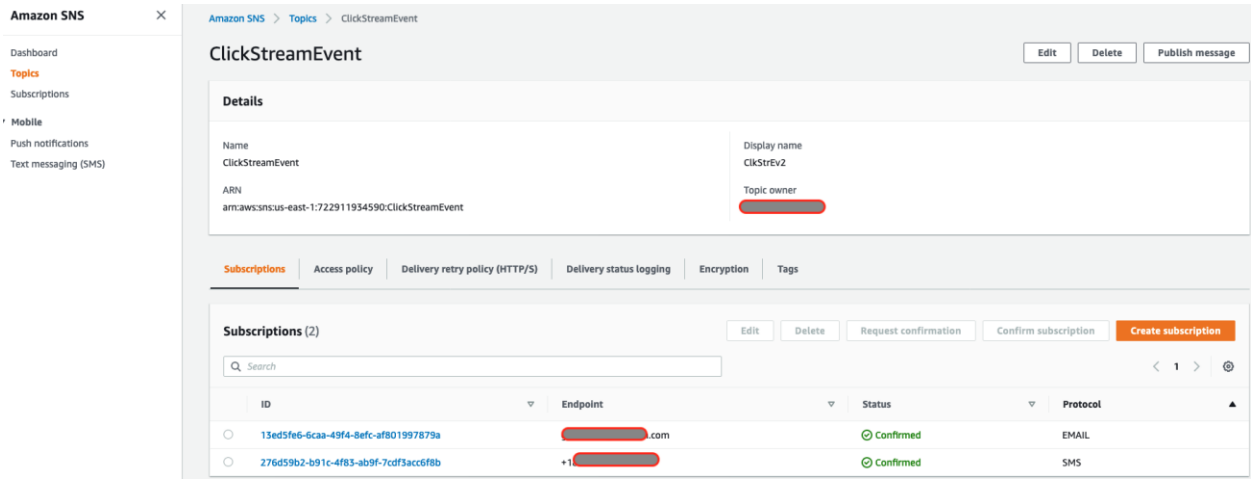


Don't click on Send Data yet, leave this browser tab open, we will do that during the main lab.

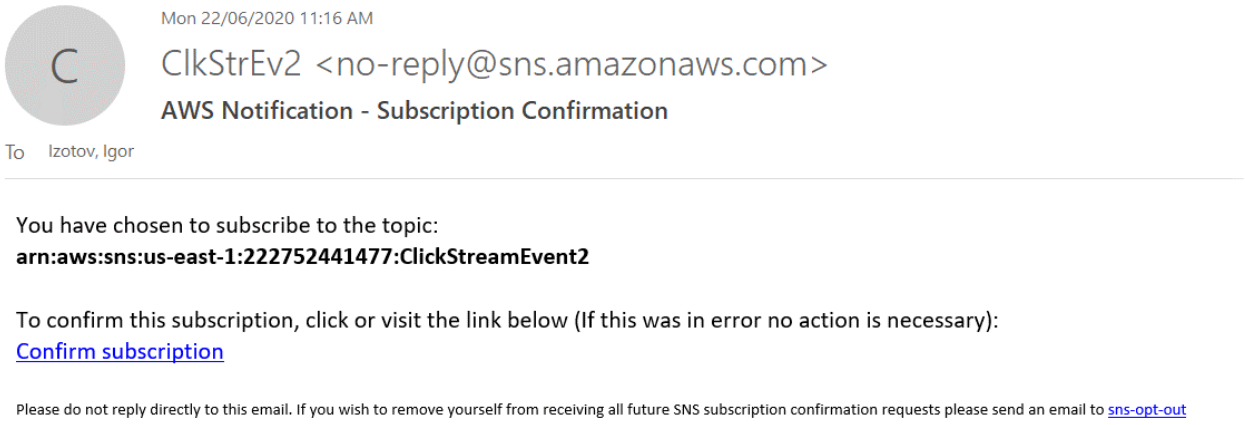
Set up Email and SMS Subscription

1. Navigate to Amazon SNS Topics by following this link: https://console.aws.amazon.com/sns/v3/home?region=us-east-1#/topics

2. Click the topic name. The Topic details screen appears listing the e-mail/SMS subscription as pending or confirmed.



3. Check your inbox for a subscription confirmation email from no-reply@sns.amazonaws.com, click **Confirm subscription** to confirm



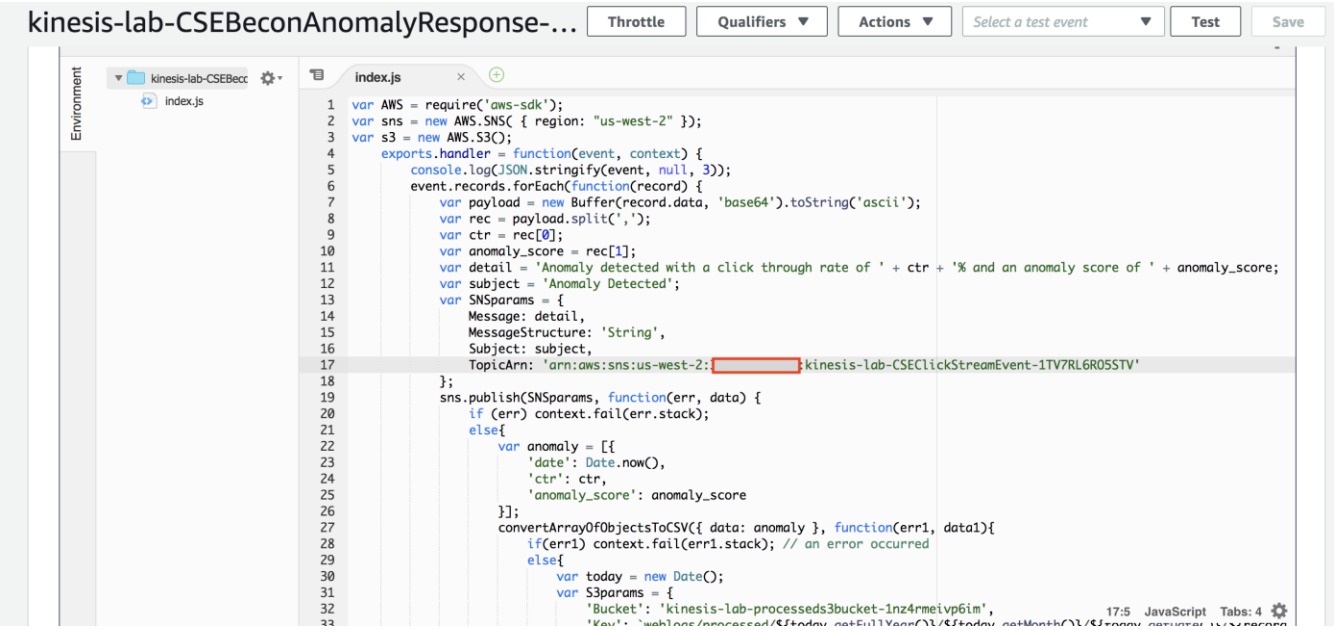
Note: If you can't locate the request confirmation email, make sure to check your email junk folder.

Review AWS Lambda Anomaly function:

The Lambda function has been deployed to your AWS account by the CloudFormation template at the start. You just need to spend few minutes to observe the code and understand the action behind the lambda trigger:

1. In the console, navigate to **CSEBeconAnomalyResponse** AWS Lambda function by following the link: https://console.aws.amazon.com/lambda/home?region=us-east-1&#/functions/CSEBeconAnomalyResponse?tab=configuration

2. Scroll down to code section.



3. Review the code in the Lambda code editor. Notice the TopicArn value matches the SNS topic ARN from the previous step.

You've completed the pre-lab. Please proceed to Kinesis main lab.