

Amazon Web Services Data Engineering Immersion Day

Lab 1 - Prelab. Real-Time Clickstream Anomaly Detection August 2020

Table of Contents

Introduction	. 2
Get Started Using the Lab Environment	, 3
CloudFormation Stack Deployment	. 5
Set up the Amazon Kinesis Data Generator	. 7
Set up Email and SMS Subscription	, 9
Review AWS Lambda Anomaly function:	10

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 -<u>https://aws-dataengineering-day.workshop.aws/300.html</u>

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 - <u>https://github.com/aws-samples/data-engineering-for-aws-immersion-day</u>.

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:

	Who are you?	
Т	Terms & Conditions:	
1 au re te	1. By using the Event Engine for the relevant event, you agree to the Event Terms and Conditions and the AWS Acce acknowledge and agree that are using an AWS-owned account that you can only access for the duration of the relev residual resources or materials in the AWS-owned account, you will make us aware and cease use of the account. A terminate the account and delete the contents at any time.	eptable Use Policy. You vant event. If you find WS reserves the right to
2 e:	2. You will not: (a) process or run any operation on any data other than test data sets or lab-approved materials by A export or otherwise create derivate works of materials provided by AWS, including but not limited to, data sets.	WS, and (b) copy, import,
3 to	3. AWS is under no obligation to enable the transmission of your materials through Event Engine and may, in its disc to post, or remove your materials at any time.	retion, edit, block, refuse
4 ar	4. Your use of the Event Engine will comply with these terms and all applicable laws, and your access to Event Engin automatically terminate if you do not comply with any of these terms or conditions.	e will immediately and
	4756dathath	
т	This is the 12 digit hash that was given to you or your team.	↓ _
		Accept Terms & Login

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:

Environment Setup	Readme
Outputs:	
S3 Bucket name mod-3fccddd609114925-dmslabs3bucket-1ngcgzzcnd15u 🌓	
BusinessAnalystUser mod-3fccddd609114925-BusinessAnalystUser-MB0XFZLQLOXX 🏢	
DMSLabRoleS3 ARN arn:aws:iam::377243295828:role/mod-3fccddd609114925-DMSLabRoleS3-O2VT1RSN43SG 🍺	
Giue Lab Role mod-3fccddd609114925-GlueLabRole-YLTJA13WW6WT 🍺	
S3BucketWorkgroupA mod-3fccddd609114925-s3bucketworkgroupa-tbon3m1mkunh р	
S3BucketWorkgroupB mod-3fccddd609114925-s3bucketworkgroupb-18ygl8nfp8ead 順	
WorkgroupManagerUser mod-3fccddd609114925-WorkgroupManagerUser-5IVE0UONIBG4 🎼	

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

Team Dashboard



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

AWS Console Login	
Remember to only use " "as your region, unless otherwise directed by the event operator. Login Link	
Credentials / CLI Shin ets Mac/Linux Wind op Mac or Linux (#) Image: Clinux (#)	
euport MS_DEFAULT_RECOM- euport MS_SECRET_ACCESS_KEY- euport MS_SECRET_ACCESS_KEY- euport MS_SESSION_TOKEN-	
How do I use the AWS CLI? Checkout the AWS CLI documentation here: https://docs.aws.amazon.com/cli/intest/userguide/cli-chap-welcome.html	
OK	

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

CloudFormation Stack Deployment

 Use this link to create a new CloudFormation Stack: <u>https://console.aws.amazon.com/cloudformation/home#/stacks/create/review?stackN</u> <u>ame=kinesis-pre-lab&templateURL=https://s3.amazonaws.com/aws-dataengineering-day.workshop.aws/Kinesis_PreLab.yaml</u>
 Quick create stack

emplate		
emplate URL		
ttps://s3.amazonaws.com	s-dataengineering-day.workshop.aws/Kinesis_PreLab.yaml	
tack description		
upporting elements for th	nesis Analytics click stream lab	
the state of the second		
таск пате		
tack name		

- 2. 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 (+1XXXXXXX) where you can receive texts from the SNS topic.

Parameters	
Parameters are defined in your template and allow you to input custom values when you create or update a stack.	
Kinesis Pre Lab set up	
Username	
The username of the user you want to create in Amazon Cognito.	
Password	
Password The password of the user you want to create in Amazon Cognito. Must be at least 6 alpha-numeric characters, and contain at least one number	
Password The password of the user you want to create in Amazon Cognito. Must be at least 6 alpha-numeric characters, and contain at least one number	
Password The password of the user you want to create in Amazon Cognito. Must be at least 6 alpha-numeric characters, and contain at least one number	
Password The password of the user you want to create in Amazon Cognito. Must be at least 6 alpha-numeric characters, and contain at least one number email	R.
Password The password of the user you want to create in Amazon Cognito. Must be at least 6 alpha-numeric characters, and contain at least one number email Email address to send anomaly detection events.	1
Password The password of the user you want to create in Amazon Cognito. Must be at least 6 alpha-numeric characters, and contain at least one number mail Email address to send anomaly detection events.	Ę
Password The password of the user you want to create in Amazon Cognito. Must be at least 6 alpha-numeric characters, and contain at least one number email Email address to send anomaly detection events.	Ę
Password The password of the user you want to create in Amazon Cognito. Must be at least 6 alpha-numeric characters, and contain at least one number email Email address to send anomaly detection events.	R
Password The password of the user you want to create in Amazon Cognito. Must be at least 6 alpha-numeric characters, and contain at least one number email Email address to send anomaly detection events. SMS Mobile Phone number to send SMS anomaly detection events. +1XXXXXXXXXX	R
Password The password of the user you want to create in Amazon Cognito. Must be at least 6 alpha-numeric characters, and contain at least one number email Email address to send anomaly detection events. SMS Mobile Phone number to send SMS anomaly detection events. +1XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	Ψ

3. In the **Capabilities**, select the check box marked **I acknowledge that AWS CloudFormation might create IAM resources**.

Capabilities				
() The following resource(s) require capabilities: [AWS::IAM::Role]				
This template contains Identity and Access Management (IAM) resources that might provide entities access to make changes to your AWS account. Check that you want to create each of these resources and that they have the minimum required permissions. Learn more 🖸				
I acknowledge that AWS CloudFormation might create IAM resources.				
Cancel	Create change set Cre	ate stack		

4. Click **Create stack**. CloudFormation redirects you to your existing stacks.

CloudFormation > Stacks > kinesis-pre-la				
🖃 Stacks (1)	kinesis-pre-lab	Delete	Update Stack actions	▼ Create stack ▼
Q Filter by stack name	Stack info Events Resources O	tputs Parameters Template	Change sets	
kinesis-pre-lab 2020-08-16 01:08:59 UTC+1000 () CREATE_IN_PROGRESS	Events (1) Q. Search events			C
	Timestamp v Logical ID	Status	Status reason	۵
	2020-08-16 01:08:59 UTC+1000 kinesis-pre-	lab CREATE_IN_PROGRES S	User Initiated	

- 5. 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

CloudFormation > Stacks > kinesis-pre-la	ab	
🖸 Stacks (1)	kinesis-pre-lab	Delete Update Stack actions V Create stack V
Q Filter by stack name	Stack info Events	Resources Outputs Parameters Template Change sets
Active View nested		
kinesis-pre-lab	Outputs (3)	C
O CREATE_COMPLETE	Sector outputs	0
	Key 🔺	Value
	KinesisDataGeneratorUrl	https://awslabs.github.io/amazon-kinesis-data-generator/web/producer.html?upid=us-east-1_Wc92vzUjw&ipid=us- east-1:15656fc7-a69a-4fa2-b754-18dba939cab8&cid=2dqttcbjfemb8onjspb2gg3bhp&r=us-east-1
	ProcessedBucketName	kinesis-pre-lab-processeds3bucket-l2sdlouozco3
	RawBucketName	kinesis-pre-lab-raws3bucket-2ecu9o76qptp

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**.

CloudFormation > Stacks > kinesis-pre-la	b	
🖸 Stacks (1)	kinesis-pre-lab	Delete Update Stack actions ▼ Create stack ▼
Q Filter by stack name	Stack info Events	Resources Outputs Parameters Template Change sets
Active View nested		
	Outputs (3)	
kinesis-pre-lab 2020-08-16 01:08:59 UTC+1000 C ORFATE COMPLETE	Q Search outputs	
		۲
	Key 🔺	Value v
	KinesisDataGeneratorUrl	https://awslabs.github.io/amazon-kinesis-data-generator/web/producer.html?upid=us-east-1_Wc9ZvzUjw&ipid=us- east-1:15656fc7-a69a-4fa2-b754-18dba939cab8&cid=2dqttcbjfemb8onjspb2gg3bhp&r=us-east-1
	ProcessedBucketName	kinesis-pre-lab-processeds3bucket-l2sdlouozco3
	RawBucketName	kinesis-pre-lab-raws3bucket-2ecu9o76qptp

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

Amazon Kinesis Data Generator	🖨 Configure	🛛 Help	test	••••••	Sign In	
Amazon K	inesis Data	Ge	enerato	or		
The KDG makes it simple to send t get started. If you haven't configure	est data to your Amazon Kinesis stra d an Amazon Cognito user, choose	eam or Am Help.	nazon Kinesis Firehos	e delivery stream. Sig	gn in to	

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 **the appropriate region** where you launched the CloudFormation stack 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.

Amazon Kinesis Data G	enerator	🗘 Con
Region	us-east-1	
Stream/delivery stream	Kinesis-Pre-Lab-FirehoseDeliveryStream-1XMH0FAX1 \$	
Records per second	Constant Periodic	
	1	
Compress Records ()		
Record template 1	Schema Discovery Payload Click Payload Impression Payload Template 4 Template 5	
	Schema Discovery Payload	
	{"browseraction":"DiscoveryKinesisTest", "site": "yourwebsiteurl.domain.com"}	

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

Amazon SNS ×	Amazon SNS > Topics > ClickStreamEvent		
Dashboard Topics	ClickStreamEvent		Edit Delete Publish message
Subscriptions	Details		
Push notifications Text messaging (SMS)	Name ClickStreamEvent ARN amawssnsus-past-1/722911934590/ClickStreamEvent	Display name ClkStrEv2 Topic owner	
	Subscriptions Access policy Delivery retry policy (HTTP/S)	Delivery status logging Encryption Tags	
	Subscriptions (2) Q. Search	Edit Delete	Request confirmation Confirm subscription Create subscription < 1 >
	ID v	Endpoint \bigtriangledown	Status $ abla Protocol abla $
	O 13ed5fe6-6caa-49f4-8efc-af801997879a	.com	Confirmed EMAIL
	O 276d59b2-b91c-4f83-ab9f-7cdf3acc6f8b	+1	⊘ Confirmed SMS

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



You have chosen to subscribe to the topic: arn:aws:sns:us-east-1:222752441477:ClickStreamEvent2
To confirm this subscription, click or visit the link below (If this was in error no action is necessary): <u>Confirm subscription</u>
Please do not reply directly to this email. If you wish to remove yourself from receiving all future SNS subscription confirmation requests please send an email to sns-opt-o

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:

- In the console, navigate to CSEBeconAnomalyResponse AWS Lambda function by following the link: <u>https://console.aws.amazon.com/lambda/home -</u> /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.