HAC201 – Code Green: Hacking on Amazon Sustainability Data Initiative datasets

Workshop Duration: 2 hours

Welcome to Code Green! This workshop is a supplement to the Hacking on Amazon Sustainability Data Initiative datasets Hack-a-thon.

You will be building a simple API to expose you to working with ASDI datasets and AWS services, but it could be expanded with additional sources of data and more sophisticated algorithms to help select the most sustainable location for any kind of event.

The authors of this workshop wanted to give you a head start at playing with sustainability data. Please use this as a baseline to get up and running with a basic application, then play with the data, what interesting things can you uncover? What other metrics could you look at when determining if a city is the best city for our fictious event? What other ways can you leverage this dataset?

Introduction:

The year is 2030, and MLD—Major League DeepRacer—has exploded, and is now the most popular sporting event in America. A series of identical stadiums have been built in different parts of the country to host races. A focus on sustainability has swept the country as well, and sustainability is an important consideration in where championship races are held.

Today we're going to create an API for use by DeepRacer officials when choosing race locations. It queries an ASDI data to select a city where the race will have the least environmental impact. The competition committee has decided to look for sites where the predicted average daily temperature is closest to 23.0 degrees Celsius.

ASDI Data:

You'll be using NOAA's Global Historical Climatology Network Daily (GHCN-D) dataset to estimate temperature. The GHCN contains measurements from hundreds of sensors worldwide. Read more about the data and its ASDI properties here. Find information about the data schema in the NOAA documentation here.

Requirements and Expectations:

- *Participants use their own AWS accounts to run the lab with IAM Admin permissions. Please see a facilitator if you do not have access to an account.*

- *Basic Knowledge of AWS services (Amazon S3, Amazon Athena, AWS Lambda, and Amazon API Gateway)*
- *Comfortable working on the AWS Console and configure AWS services*
- *Working knowledge of SQL Queries*

Github repository:

This workshop has several supporting documents:

- PDF/Instructional Labs to walk through:
    - Lab-guide-Start-Here.pdf – this document
    - Lab-guide-Section-1-S3.pdf
    - Lab-guide-Section-2-Athena.pdf
    - Lab-guide-Section-3-APIGW&Lambda.pdf
    - Lab-guide-Section-4-S3-web.pdf
    - Lab-guide-End-Here-Improvements-and-References.pdf
- Text files with text to cut/paste into the console:
    - iam.json.txt
    - lambda-code.py.txt
    - sql-statements.sql.txt
    - test-event-lambda.json.txt
    - index.html
- CSV files with sample data to provide city location information
    - stadiums-with-stations_global.csv
- Optional: Cloud Formation Template of this workshop
    - completed-workshop.cfn.json

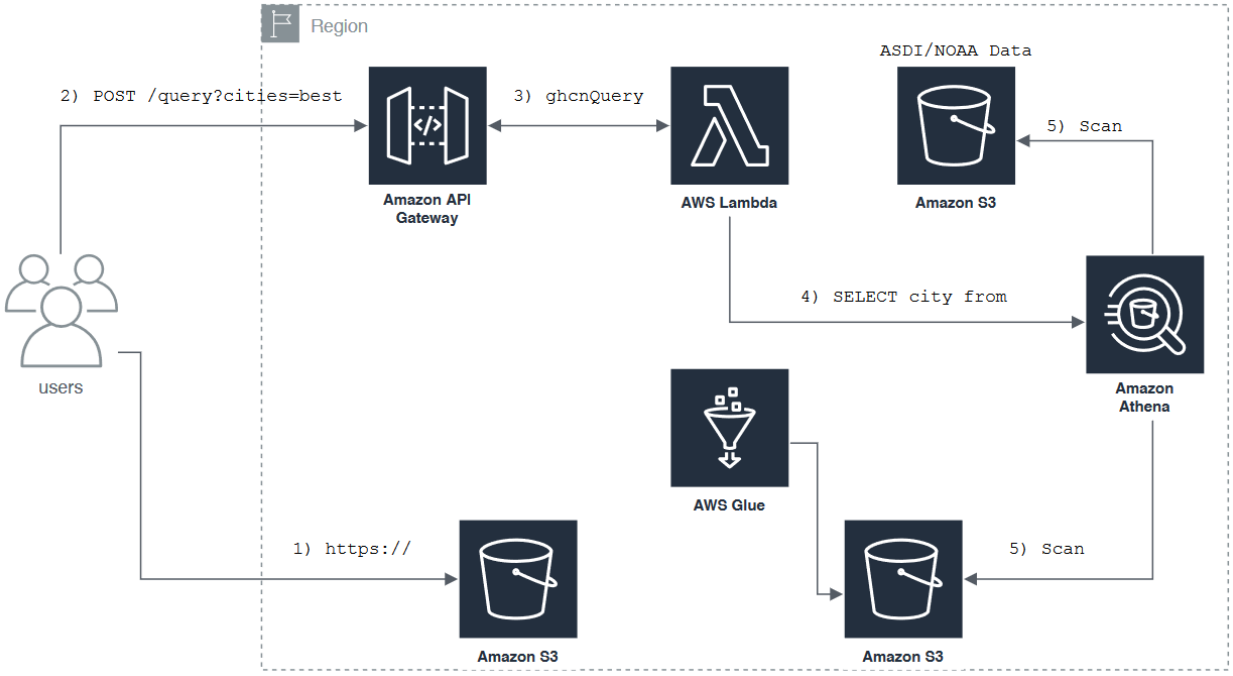
Instructions:

- *Begin from Exercise 1 and progress through rest of the exercises in order as successful completion of previous exercise is necessary to continue through the next one*
- *As a best practice, as you create new objects in AWS, keep track of what you have created (names of things like S3 buckets, database names, etc) in a notebook or text editor. This will make it easier to refer to these objects in other sections.*
- *Login to the AWS console, ensure N. Virginia (us-east-1) is selected as AWS region*
- *Ensure your AWS account has service limits/instance limits to launch the below AWS services.*
- *Ensure you have IAM permission to launch and configure the following AWS services:*
    - *AWS IAM to setup and configure IAM roles, policies*
    - *Amazon S3*
    - *Amazon Athena*
    - *Amazon API Gateway*
    - *AWS Lambda*
    - *AWS Glue*
- *Recommended to login as an IAM admin user to run the lab so you're not tweaking IAM policies for each AWS service*

High-level workflow

In this workshop, you will be using various services to create an application that will enable end-users to query an ASDI dataset, to determine the city with the temperature closest to 23 degrees Celsius. The process will be following these high-level steps:

1) Create a new AWS S3 bucket that will act as the query repository and hold the city location data
2) Creating Athena tables to query the data set
3) Create an Amazon Lambda function to query Amazon Athena
4) Create an AWS API Gateway endpoint to front end Amazon Lambda
5) Build static S3 Website to interface with API Gateway



CloudFormation Template

A CloudFormation template which builds the resources described in this workshop is included. **You do not need to run the template as part of the workshop.** *The template is available for those want the recommendation API and do not wish to work through the lab guides.*

To use the template, complete section 1 of the Lab-guide-Section-1-S3.pdf including creating an S3 bucket and uploading the stadiums-with-stations_global.csv file into it. You will pass this bucket name into the template as parameter. The template is available in the Github repo or at this url: https://s3.amazonaws.com/code-green-asdi/templates/completed-workshop.cfn.json

Launch the template using the AWS CLI or Management Console. Refer to the CloudFormation documentation for details on how to launch a CloudFormation template. Once the stack completes, your API endpoint will be listed in the Outputs.