

Curso AWS Data Analytics for IoT

TSI - IFRN - Campus Parnamirim/RN

Disciplina: Programação para Internet das Coisas (IoT)

Docente: Valério Medeiros

Discente: Marcílio Freitas

Período: 5º

Atividade

Crie um repositório e um documento markdown respondendo os seguintes itens referente ao curso de dados para IoT.

Introdução

A empresa Amazon possui ferramentas e serviços para web (AWS - Amazon Web Services) que são mundialmente conhecidos. Ela fornece cursos e certificações, sendo em sua maioria pagos ou de acesso para estudantes quando se tem um professor certificado para ministrar os cursos. Esta atividade trata de um dos cursos com acesso para estudantes, o AWS Data Analytics. Um curso voltado para a análise e manipulação de dados muito comuns em ferramentas da AWS. Abaixo será possível ver um breve resumo do que cada capítulo do curso trata.

Quais as soluções AWS estudadas? e o que cada uma atende?

1. Store data in Amazon S3

No capítulo 1 deste curso aprendemos sobre o Amazon S3 (Simple Storage Service), que é o serviço mais básico da AWS que dará início aos nossos estudos e que será a base para os demais projetos. Ele servirá para armazenar e consultar dados através da criação de buckets de dados.

2. Query Data in Amazon Athena

No capítulo 2 deste curso aprendemos a como usar o AWS Amazon Athena e para consultas estruturadas, não estruturadas e semi-estruturadas. Com o Athena é possível criar banco de dados, tabelas, otimizar e executar queries mais complexas.

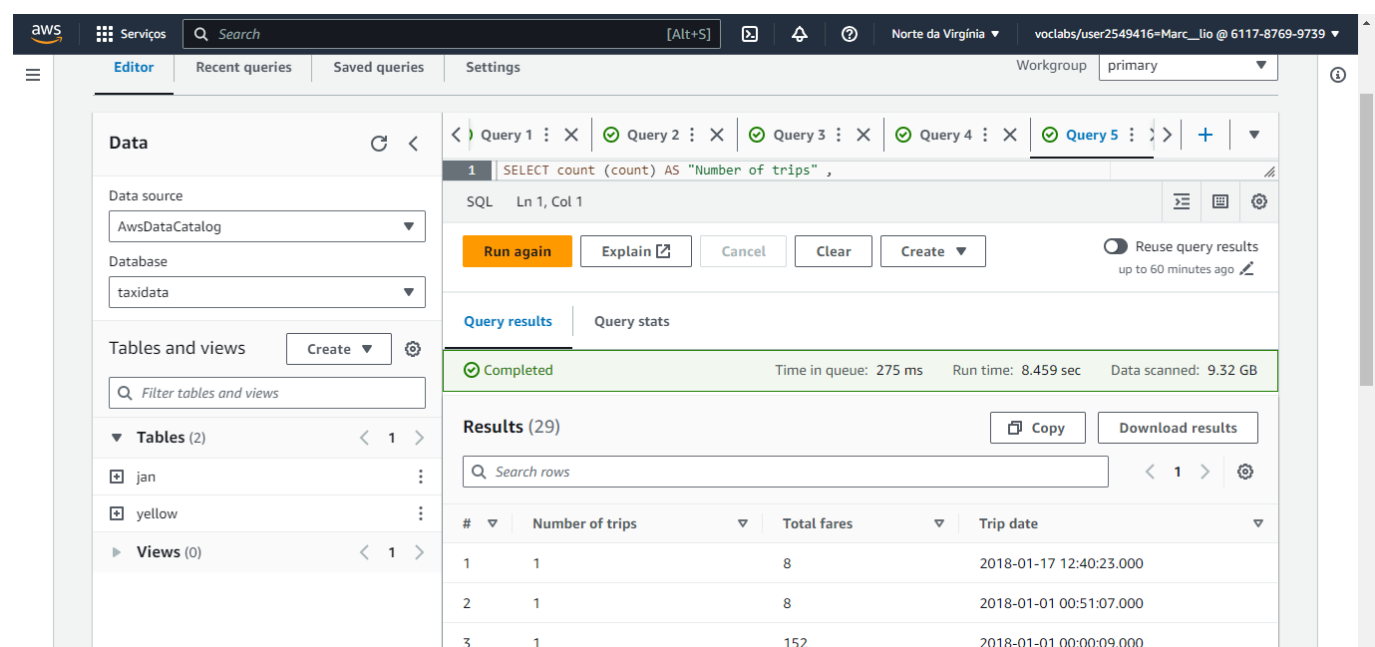


Figura - Lab 2 - Criando banco de dados, tabelas e consultas

3. Query data in Amazon S3 with Amazon Athena and AWS Glue

No capítulo 3 deste curso aprendemos a usar o AWS Amazon Athena junto com o AWS Glue. Ambos possuem o mesmo propósito, porém o Glue vem para identificar os dados através de metadados criando assim um catálogo e um esquema para cada tipo de dados

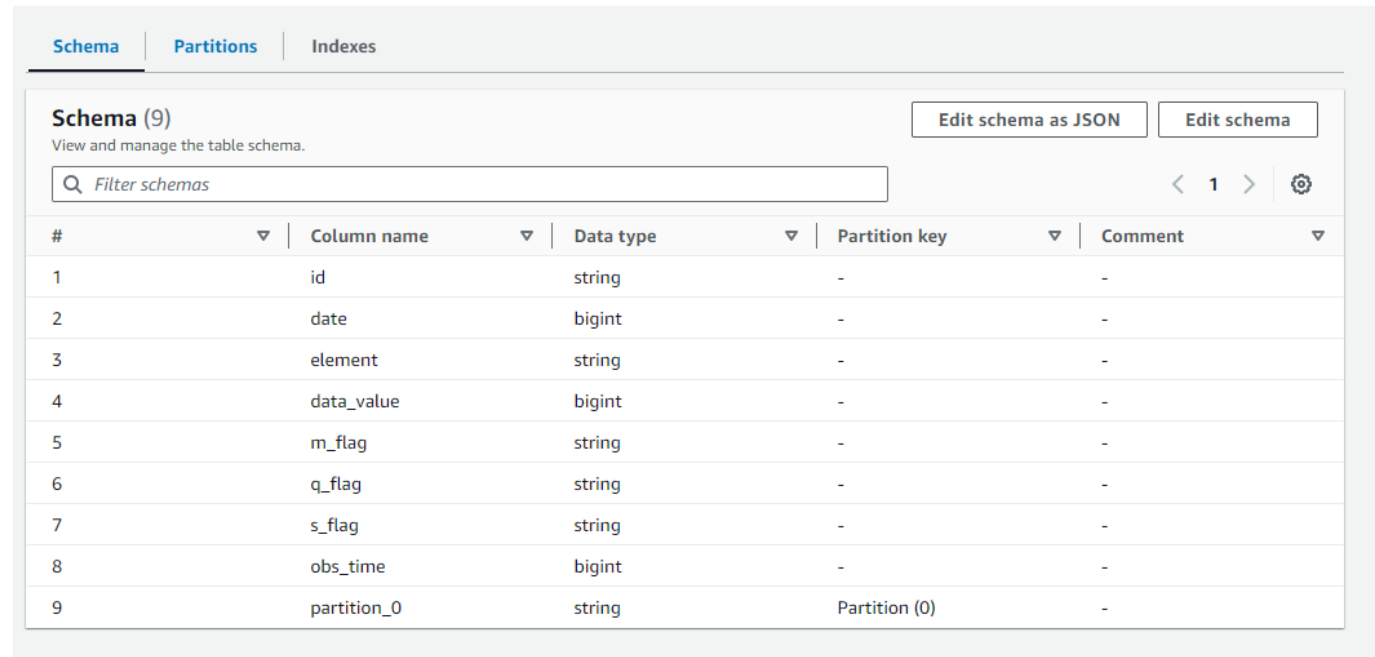
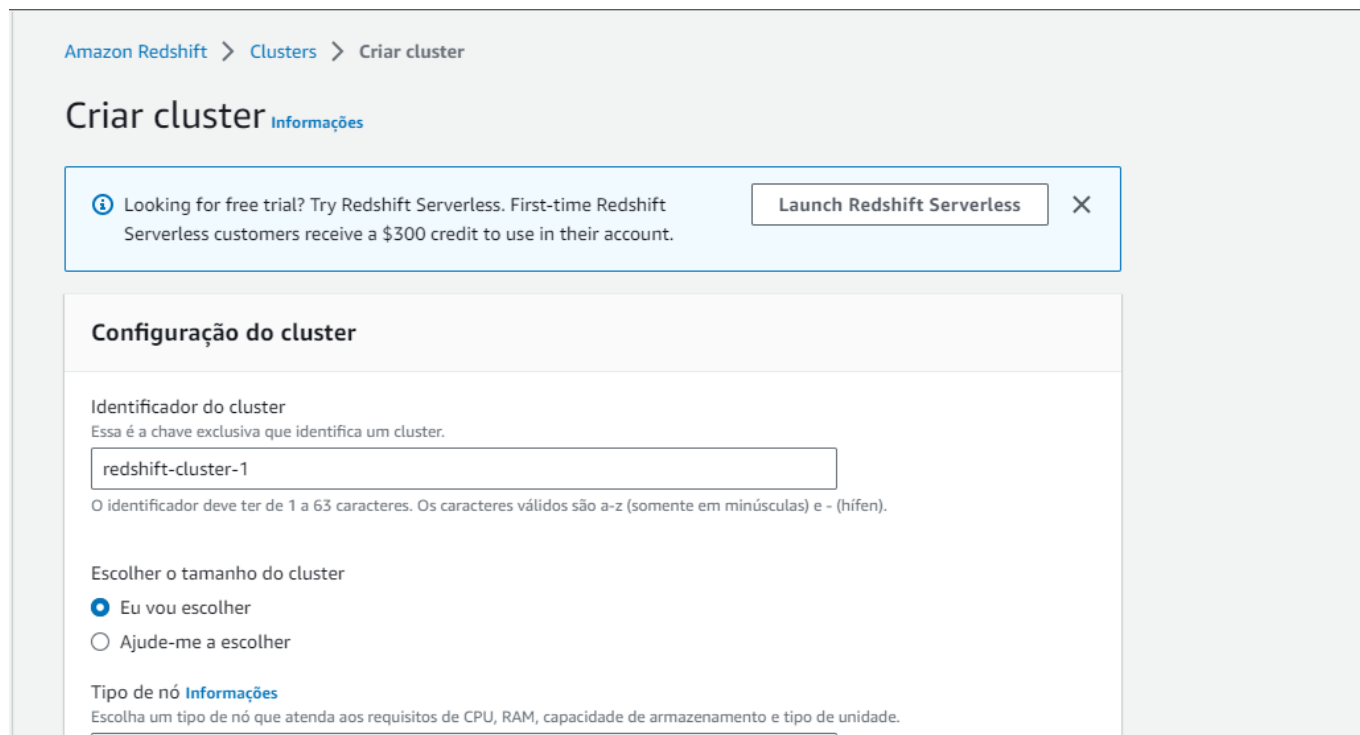


Figura - Lab 3 - Criação da tabela csv no schema AWS Glue

4. Analyze Data with Amazon Redshift

No capítulo 4 deste curso aprendemos a usar o AWS Amazon Redshift, que é utilizado para criar clusters de containers de volumes de dados muito grandes, aos quais foi possível carregar e fazer consultas com o uso da ferramenta.



Amazon Redshift > Clusters > Criar cluster

Criar cluster [Informações](#)

Looking for free trial? Try Redshift Serverless. First-time Redshift Serverless customers receive a \$300 credit to use in their account. [Launch Redshift Serverless](#) X

Configuração do cluster

Identificador do cluster
Essa é a chave exclusiva que identifica um cluster.

redshift-cluster-1

O identificador deve ter de 1 a 63 caracteres. Os caracteres válidos são a-z (somente em minúsculas) e - (hifen).

Escolher o tamanho do cluster

☒ Eu vou escolher

☐ Ajude-me a escolher

Tipo de nó [Informações](#)
Escolha um tipo de nó que atenda aos requisitos de CPU, RAM, capacidade de armazenamento e tipo de unidade.

Figura - Criando um cluster com Amazon Redshift

5. Analyze Data with Amazon Sagemaker, Jupyter Notebooks and Bokeh

No Capítulo 5 deste curso aprendemos a usar o AWS Amazon SageMaker, que é utilizado no tratamentos de dados voltados para o uso do machine learning. Além do Sagemaker, usaremos o Jupyter Notebooks, que é um editor para a linguagem Python e a biblioteca Bokeh, ambos para a geração de gráficos dos dados inseridos através do Sagemaker.

```
from bokeh.io import show, output_notebook
from bokeh.models import CategoricalColorMapper, ColumnDataSource, FactorRange
from bokeh.plotting import figure, output_file
fruits = ['Apples', 'Pears', 'Nectarines', 'Plums', 'Grapes', 'Strawberries']
counts = [5, 3, 4, 2, 4, 6]

p = figure(x_range=fruits, height=250, toolbar_location=None, title="Fruit Counts")
p.vbar(x=fruits, top=counts, width=0.9)
p.xgrid.grid_line_color = None
p.y_range.start = 0
output_notebook()
show(p)
```

Figura - Script em Python utilizando a biblioteca Bokeh

You see the following bar chart displayed in the notebook:

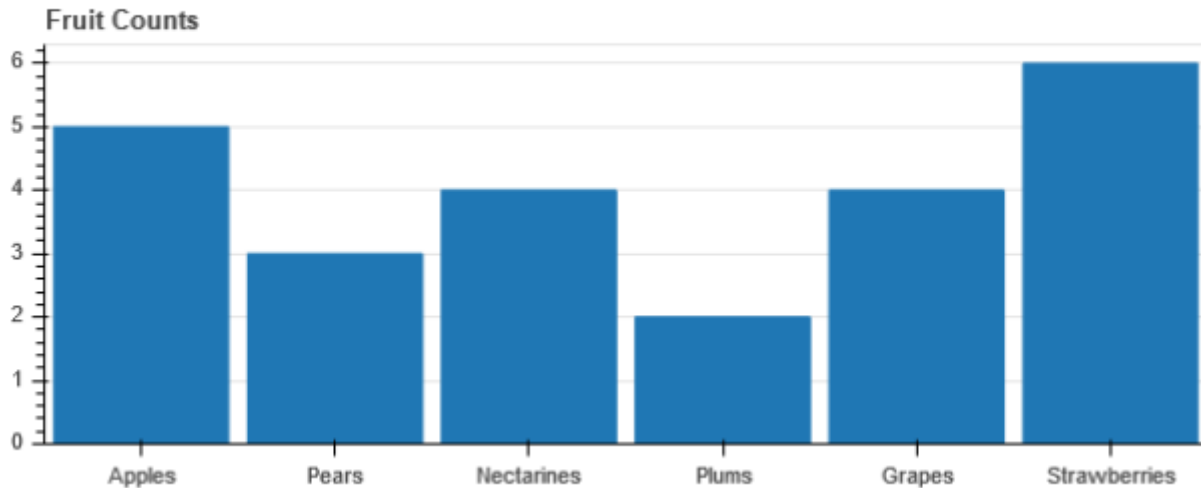


Figura - Gráfico gerado com o Bokeh

6. Automate Loading Data with the AWS Data Pipeline

No capítulo 6 deste curso aprendemos sobre como tratar dados de forma rápida e automatizada usando Pipelines. Com o AWS Data Pipeline é possível fazer o que a empresa chama de ETL (extrair, transformar e carregar) e tudo isso de forma automática. Cria-se um cluster com o AWS Redshift, cria dados com o Pipeline e armazena no S3, e faz consultas com o Redshift.

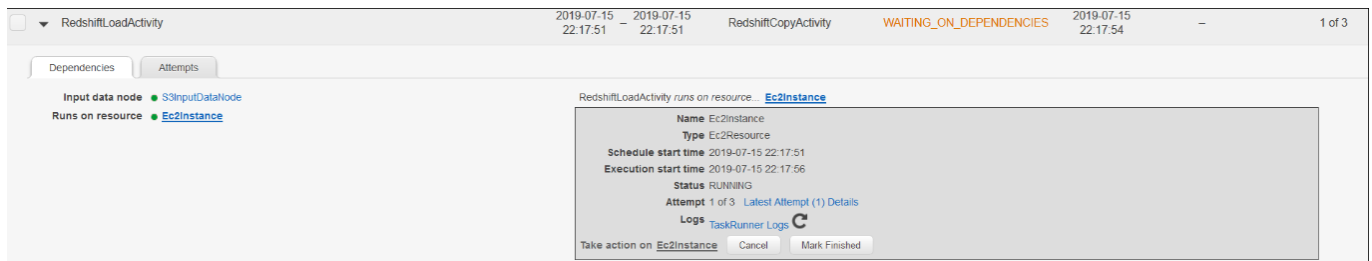


Figura - Lab 6 - Visualizando os logs gerados pelo Pipeline

7. Analyze Streaming Data with Amazon Kinesis Data Firehose, Amazon Elasticsearch Service, and Kibana

No capítulo 7 deste curso, aprendemos sobre como capturar os dados de stream, prepará-los para uma análise e exibir os resultados em gráficos. O Amazon Kinesis Data Firehose é o responsável pela captura dos dados de stream, que são dados que vão usar um dos 5 V's comentados durante o curso: a velocidade.

O Stream possui dados que precisam ser tratados em tempo real, como os vídeos e os logs de servidores web por exemplo. O Firehose também servirá como um injetor de dados que serão recebidos através do Elastic Search, ao qual será a ponte entre o Kinesis e o Kibana, que exibirá os dados em gráficos.

Alguns serviços da Amazon tiveram seus nomes alterados, como é o caso do Elastic Search. Por ser um serviço que possui a patente desse nome, a AWS criou o seu produto "open source" chamado Amazon OpenSearch Service.



▼ Lab 7		Complete All Items	✓
	Lab 7 Introduction Viewed		✓
	Analyze Streaming Data with Amazon Kinesis Firehose, Amazon Elasticsearch and Kibana		

Figura - Lab 7

8. Analyze IoT Data with AWS IoT Analytics



▼ Lab 8		Complete All Items	✓
	Lab 8 Introduction Viewed		✓
	Analyze IoT Data with AWS IoT Analytics		


Figura - Lab 8

 **IoT Analytics** ☆
Collect, preprocess, store, analyze and visualize data of IoT devices

Top features

Quick start Channels Pipelines Datastores Datasets

Figura - IoT Analytics Service

 **IoT Core** ☆
Connect Devices to the Cloud

Top features

Message broker Jobs Authentication and authorization Registry Rules engine

Figura - IoT Core Service

O Capítulo 8 trata da análise de dados para dispositivos de internet das coisas (IoT) o AWS IoT Analytics e o AWS IoT Core. Com essa tecnologia é possível filtrar, transformar e enriquecer dados.

Explique três exemplos de atividades que você realizou no laboratório prático?

A prática deste capítulo propõe a captura de logs de um servidor web em dois navegadores diferentes enquanto se navega entre as suas páginas, utilizando o Amazon Kinesis Data Firehose Service, injetando esses dados no Amazon Elastic Search Service, e preparando a visualização desses dados no Kibana.

O aws fornece instâncias de servidores e uma delas é a demo do kibana. Nas suas propriedades de rede podemos pegar o seu IPV4 público e fazer o login com as credências fornecidas pelo capítulo atual do curso. Após o login, utilizamos textos que informam o método http utilizado e dados do tipo json, para excluir logs antigos do servidor web e acrescentar as novas pesquisas. Esses dados contêm os campos e valores aos quais serão montados nos gráficos do Kibana. No Kibana, escolhemos as cores, qual o tipo de gráfico e os campos/valores que serão exibidos.

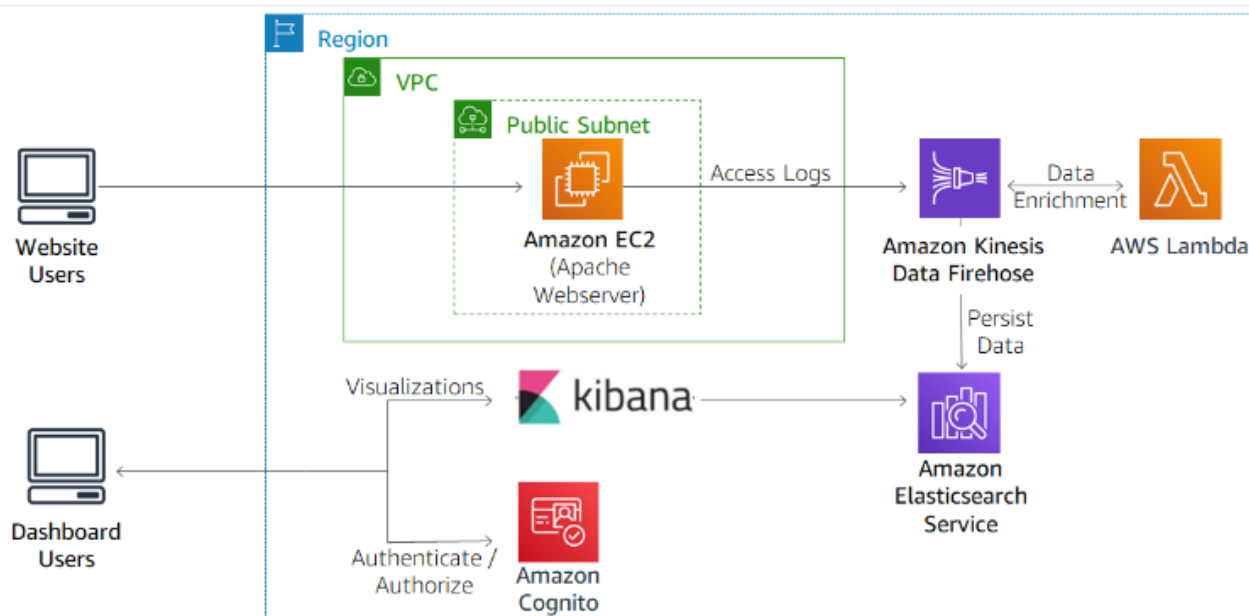


Figura - Cenário Lab 7

2. Capítulo 8

A prática deste capítulo propõe a criação de canais, armazenamentos de dados, tratamentos dos dados e a configuração dos dados para que possamos fazer consultas que filtrem, transformem e enriqueçam os dados coletados de um dispositivo de IoT e exibi-los com a ajuda do Python no sistema operacional Windows, Linux ou MacOs, além de carregar os dados utilizando o protocolo de mensageria para IoT, o Mqtt.

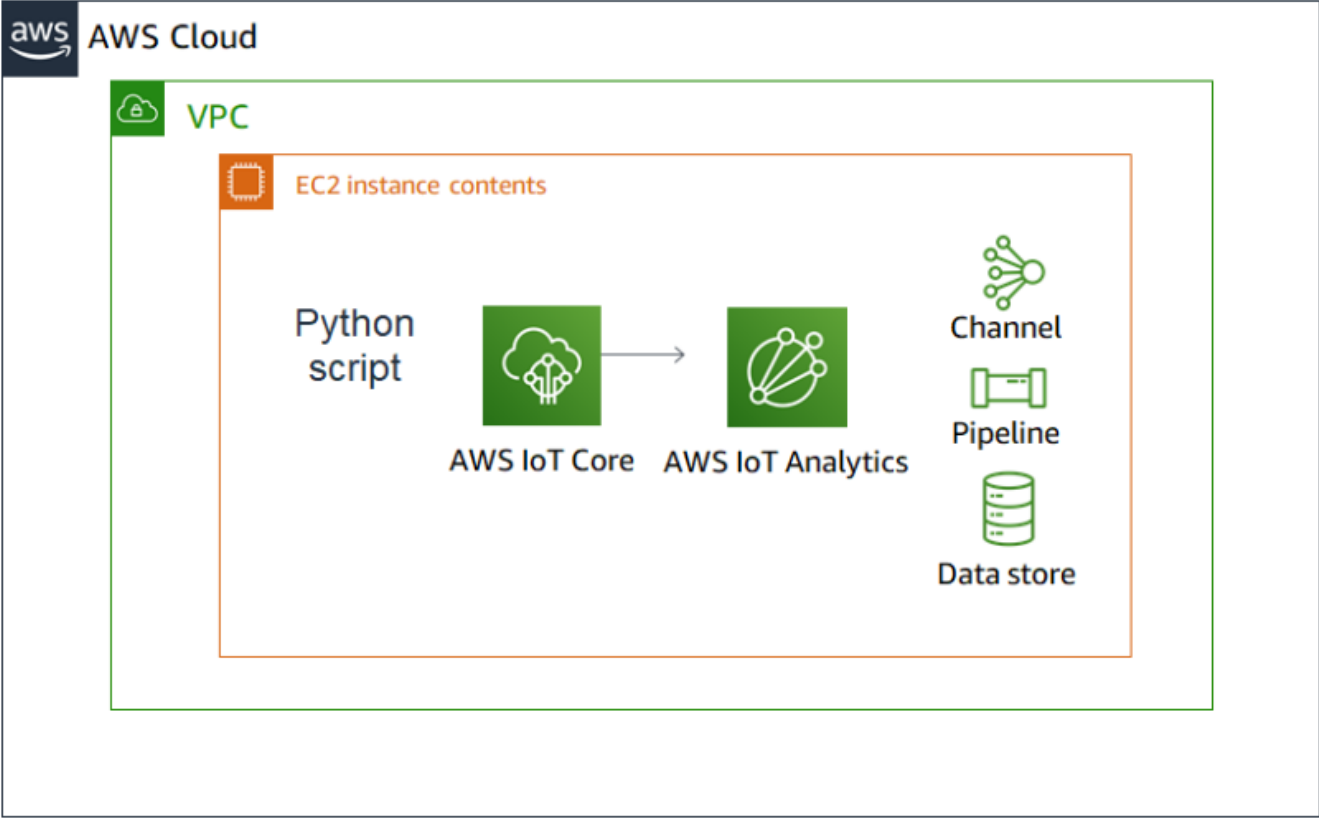


Figura - Cenário Lab 8

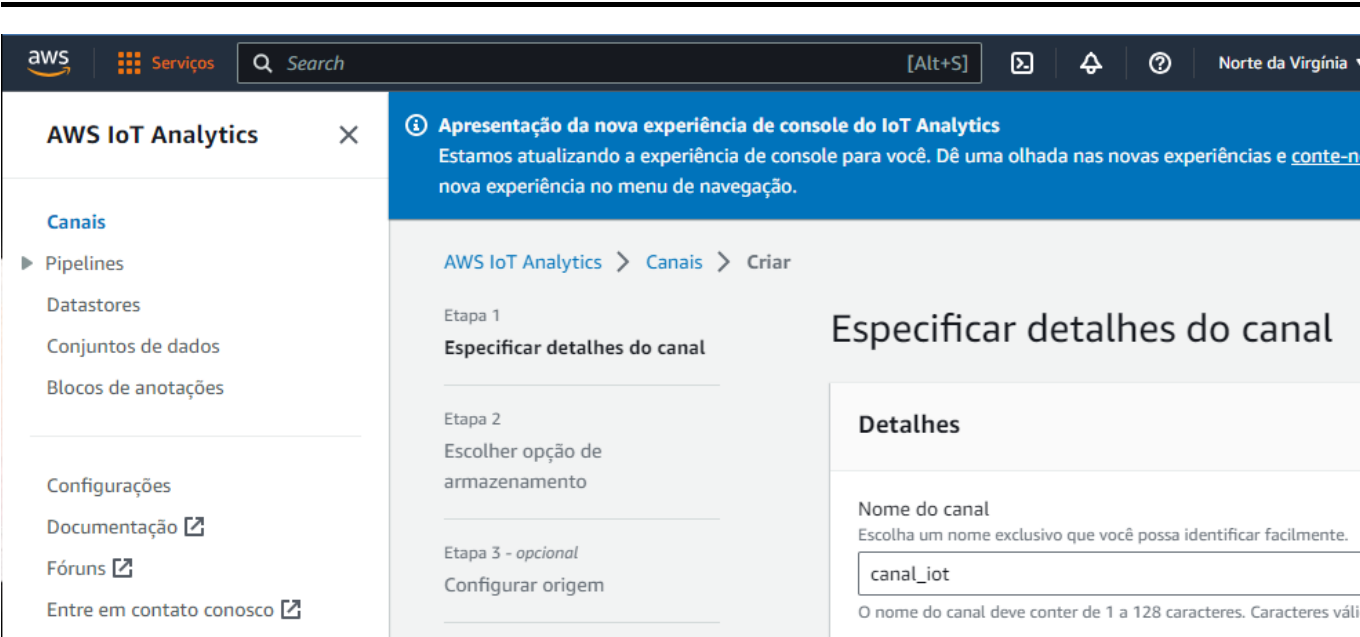


Figura - Criando um canal

3. Capítulo 1

A prática deste capítulo propões a criação de um bucket, que é uma estrutura na AWS onde os dados são carregados, analisados, consultados, criados, entre outros. Baixamos um arquivo csv chamado lab 1 no qual contém dados, carregamos ele no bucket do S3 e realizamos consultas nesses dados.

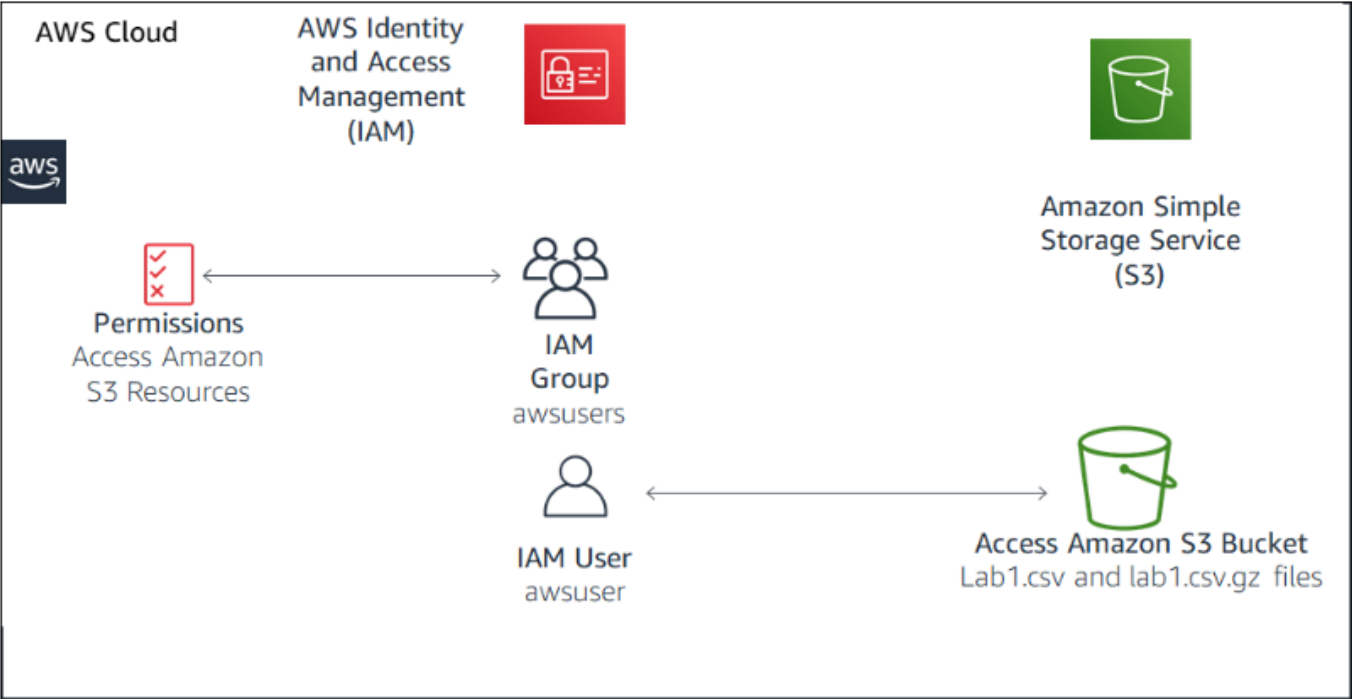


Figura - Cenário Lab 1

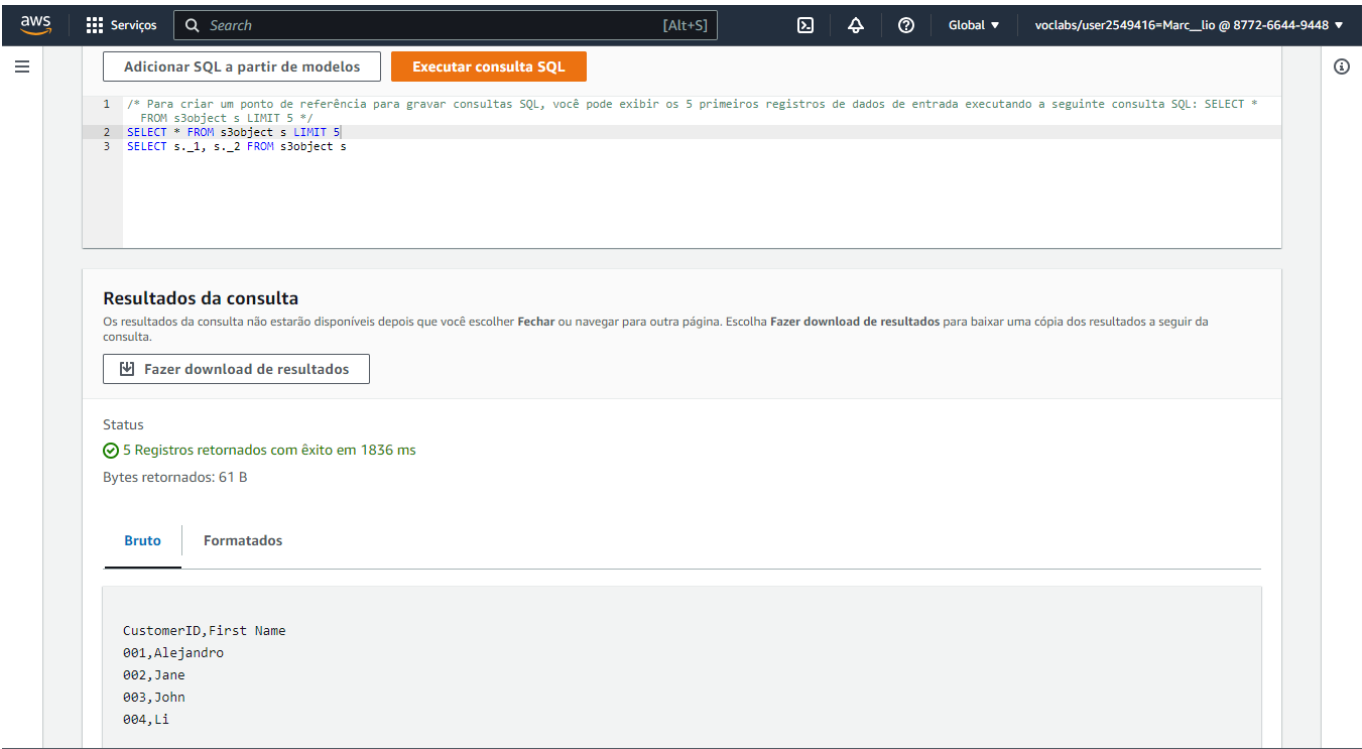


Figura - Carregando um objeto e fazendo consultas - Amazon S3 Service

Quais as principais lições apreendidas do curso?

Com o curso de análise de dados da AWS é possível perceber que existem diversos tipos de dados e várias formas de analisar os mesmos. As ferramentas da AWS apresentadas nesse curso mostram algumas dessas análises e tipos de dados: análise mais rápida (stream e Kinesis), análise mais robusta, de volume (Redshift e Athena), dados do tipo json, logs, entre outros.

Muitas das vezes pra quem apenas cria um gráfico que possa mostrar alguns dados não se pensa em porque esses dados precisam ser exibidos, não vê como um investidor na bolsa, um dono de uma empresa de produção de carros, um gestor de uma fabrica de tecidos, etc. Sendo essa análise mais rápida, com uma quantidade maior de dados e sendo estes com uma finalidade escolhida a dedo por analistas de dados, faz uma grande diferença na hora de uma produção ou de uma compra de insumos por exemplo.

Dados são coletados e gerados a todos os momentos e eles precisam ser analisados. O melhor uso das ferramentas ou mesmo do SGBD com o SQL/NoSQL pode mudar muito a vida de grandes empresas e empresários.

Conclusão

A atividade propôs conhecer algumas das ferramentas da AWS para trabalhar com dados, seja de IoT, Machine Learning ou Big Data. Não apenas a Amazon e empresas de sistemas proprietários fornecem acesso a cursos ou tecnologias que manipulam dados. Um outro exemplo seria o Apache, que possui muitos ferramentas para o mesmos fins.

Referências

Referência: AMAZON. AWS Academy Data Analytics [50734]. 2023. Disponível em: <https://awsacademy.instructure.com/courses/50734>. Acesso em: 25 jul. 2023.

Referência: AMAZON. AWS Data Analytics. 2023. Disponível em: <https://aws.amazon.com/pt/training/learn-about/data-analytics/>. Acesso em: 26 jul. 2023.
