

Curso AWS Data Analytics for IoT

TSI - IFRN - Campus Parnamirim/RN

Disciplina: Programação para Internet das Coisas (IoT)

Docente: Valério Medeiros

Discente: Marcílio Freitas

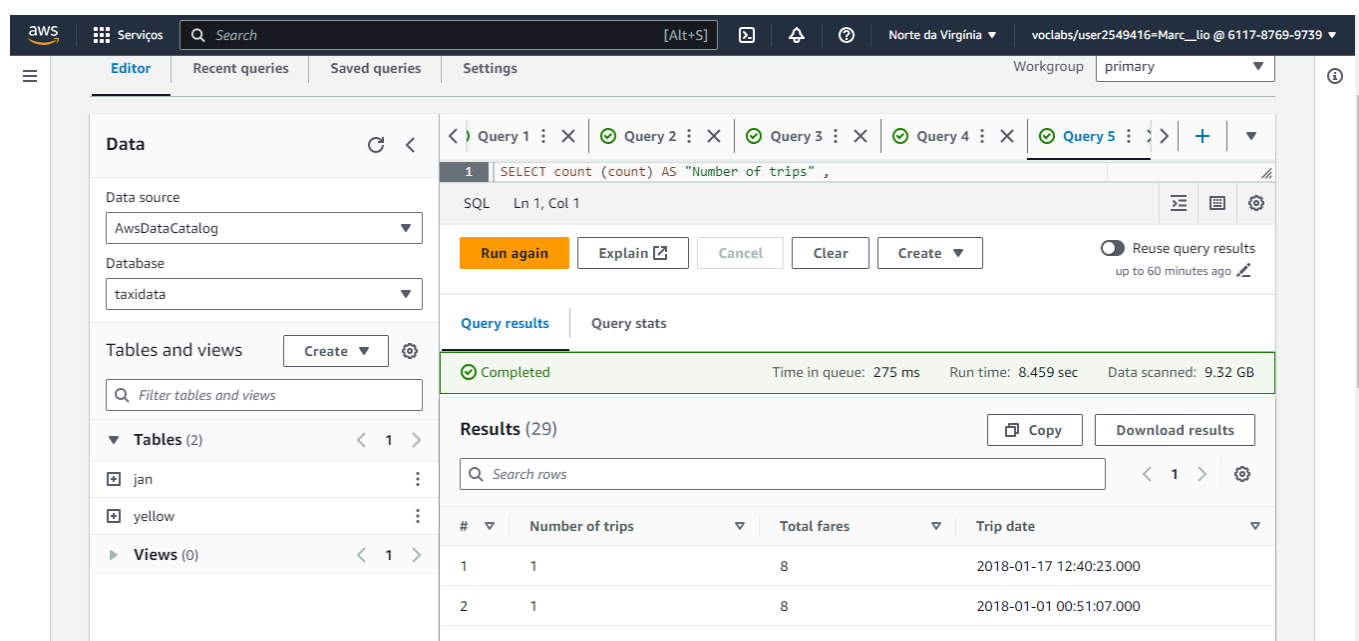
Período: 5º

Atividade

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

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 servira 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.

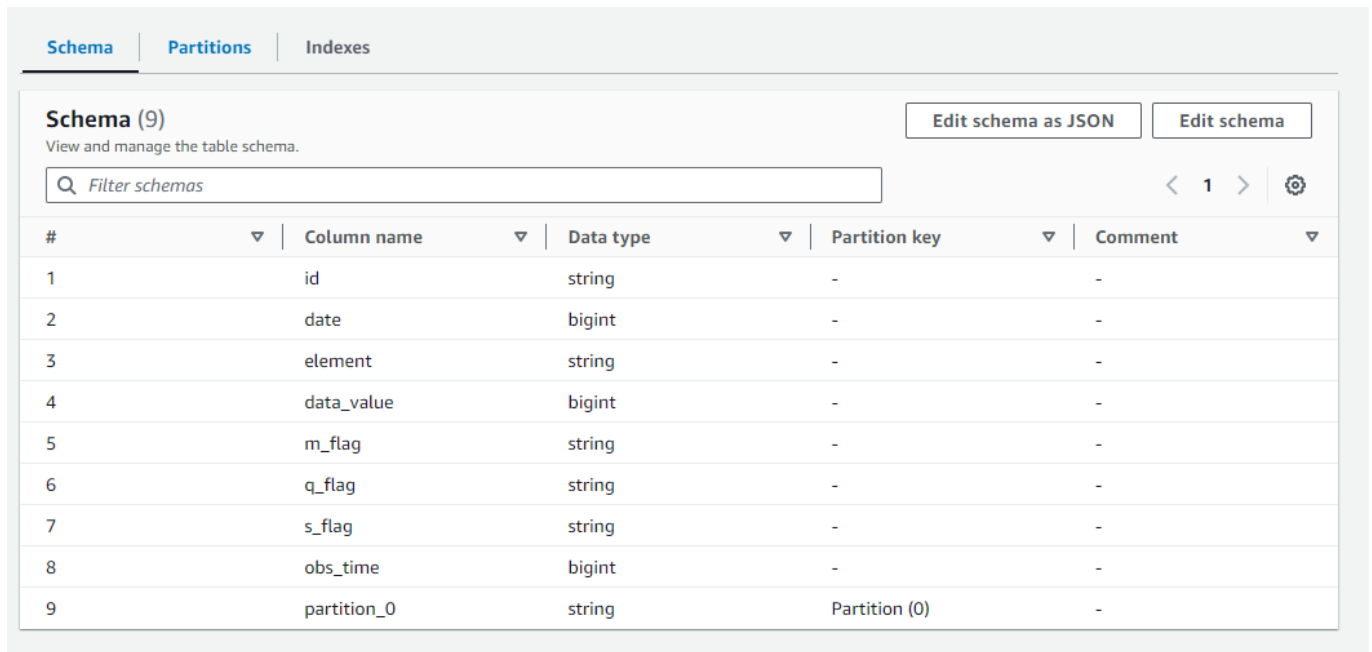


The screenshot displays the AWS Athena console interface. On the left, the 'Data' sidebar shows the 'Data source' as 'AwsDataCatalog' and the 'Database' as 'taxidata'. Below this, 'Tables and views' are listed, including 'jan' and 'yellow'. The main area shows a SQL query: `SELECT count (count) AS "Number of trips" ,`. The query status is 'Completed', with a 'Time in queue' of 275 ms, a 'Run time' of 8.459 sec, and 'Data scanned' of 9.32 GB. The 'Results (29)' section shows a table with 3 rows and 4 columns: '#', 'Number of trips', 'Total fares', and 'Trip date'. The data is as follows:

#	Number of trips	Total fares	Trip date
1	1	8	2018-01-17 12:40:23.000
2	1	8	2018-01-01 00:51:07.000
3	1	152	2018-01-01 00:00:09.000

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

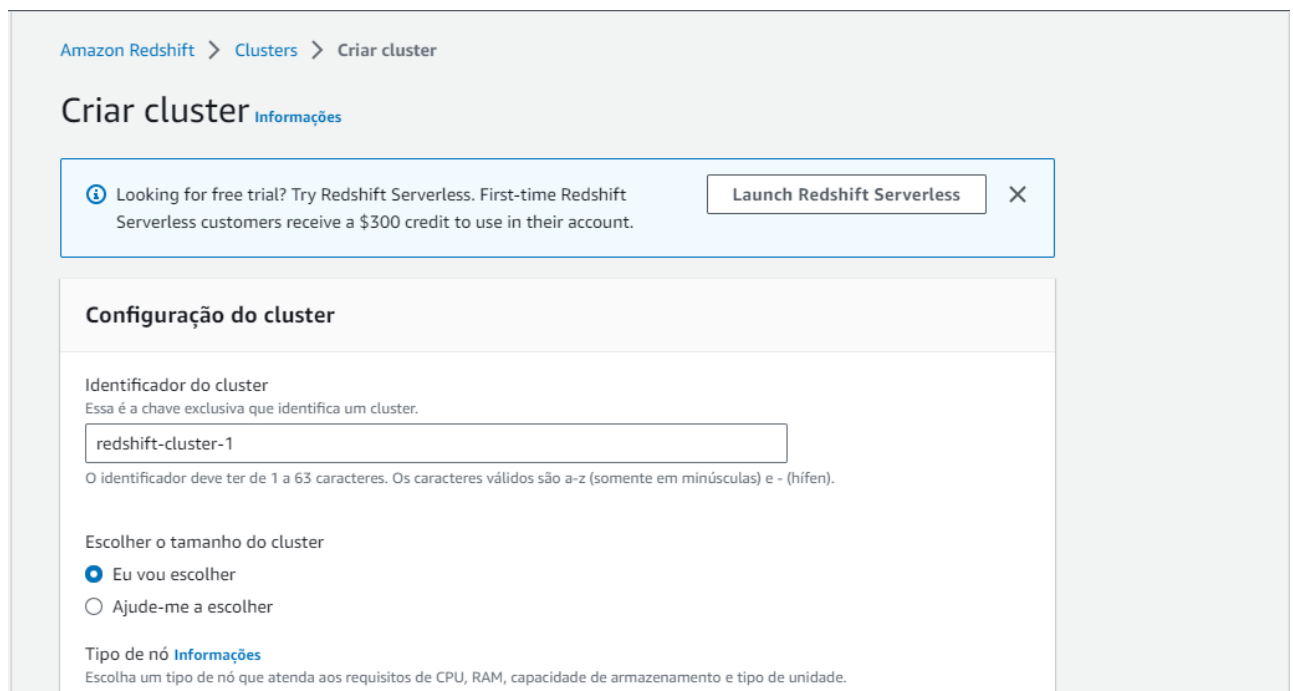


The screenshot shows the AWS Glue console interface for managing a schema. At the top, there are tabs for 'Schema', 'Partitions', and 'Indexes'. The 'Schema' tab is active, displaying 'Schema (9)' with buttons for 'Edit schema as JSON' and 'Edit schema'. Below this is a search bar labeled 'Filter schemas'. The main area contains a table with the following columns: '#', 'Column name', 'Data type', 'Partition key', and 'Comment'. The table lists 9 columns: 'id' (string), 'date' (bigint), 'element' (string), 'data_value' (bigint), 'm_flag' (string), 'q_flag' (string), 's_flag' (string), 'obs_time' (bigint), and 'partition_0' (string, marked as a Partition key). The 'partition_0' column is the only one designated as a partition key.

#	Column name	Data type	Partition key	Comment
1	id	string	-	-
2	date	bigint	-	-
3	element	string	-	-
4	data_value	bigint	-	-
5	m_flag	string	-	-
6	q_flag	string	-	-
7	s_flag	string	-	-
8	obs_time	bigint	-	-
9	partition_0	string	Partition (0)	-

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 server para criar clusters de containers de volumes de dados muito grandes, aos quais foi possível carregar e fazer consultas com o uso da ferramenta.



The screenshot shows the 'Criar cluster' (Create cluster) page in the Amazon Redshift console. The breadcrumb navigation is 'Amazon Redshift > Clusters > Criar cluster'. The main heading is 'Criar cluster' with a link to 'Informações'. A blue banner at the top promotes 'Redshift Serverless' with a 'Launch Redshift Serverless' button. Below this is the 'Configuração do cluster' (Cluster configuration) section. It includes a field for 'Identificador do cluster' (Cluster identifier) with the value 'redshift-cluster-1' and a note that it must be 1 to 63 characters, lowercase, and hyphenated. There are two radio buttons for 'Escolher o tamanho do cluster' (Choose cluster size): 'Eu vou escolher' (I will choose) which is selected, and 'Ajude-me a escolher' (Help me choose). At the bottom, there is a section for 'Tipo de nó' (Node type) with a link to 'Informações' and a note to choose a node type that meets CPU, RAM, storage, and unit requirements.

Figura - Criando um cluster com Amazon Redshift

Busca de grande volumes de dados utilizando clusters de containers


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

6. Automate Loading Data with the AWS Data Pipeline

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

▼ Lab 7

Complete All Items ✓

 Lab 7 Introduction

Viewed

✓


 Analyze Streaming Data with Amazon Kinesis Firehose, Amazon Elasticsearch and Kibana

Figura - Lab 7

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 Searsh, 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.

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

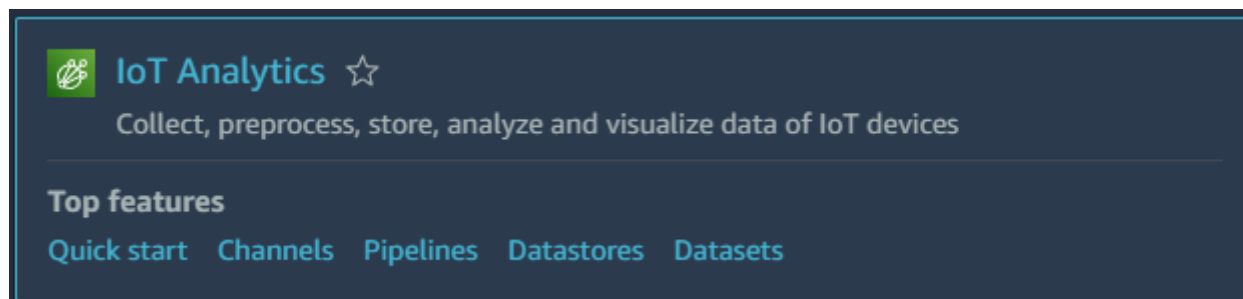


Figura - IoT Analytics Service

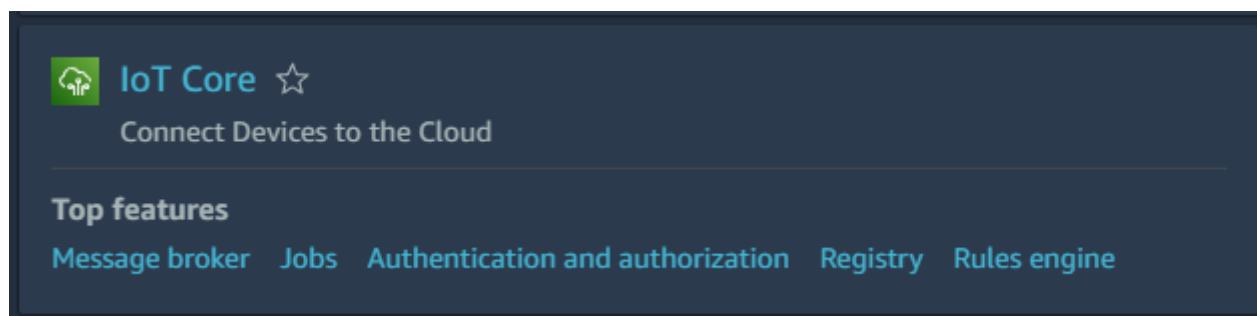


Figura - IoT Core Service

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?

1. Capítulo 7

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.

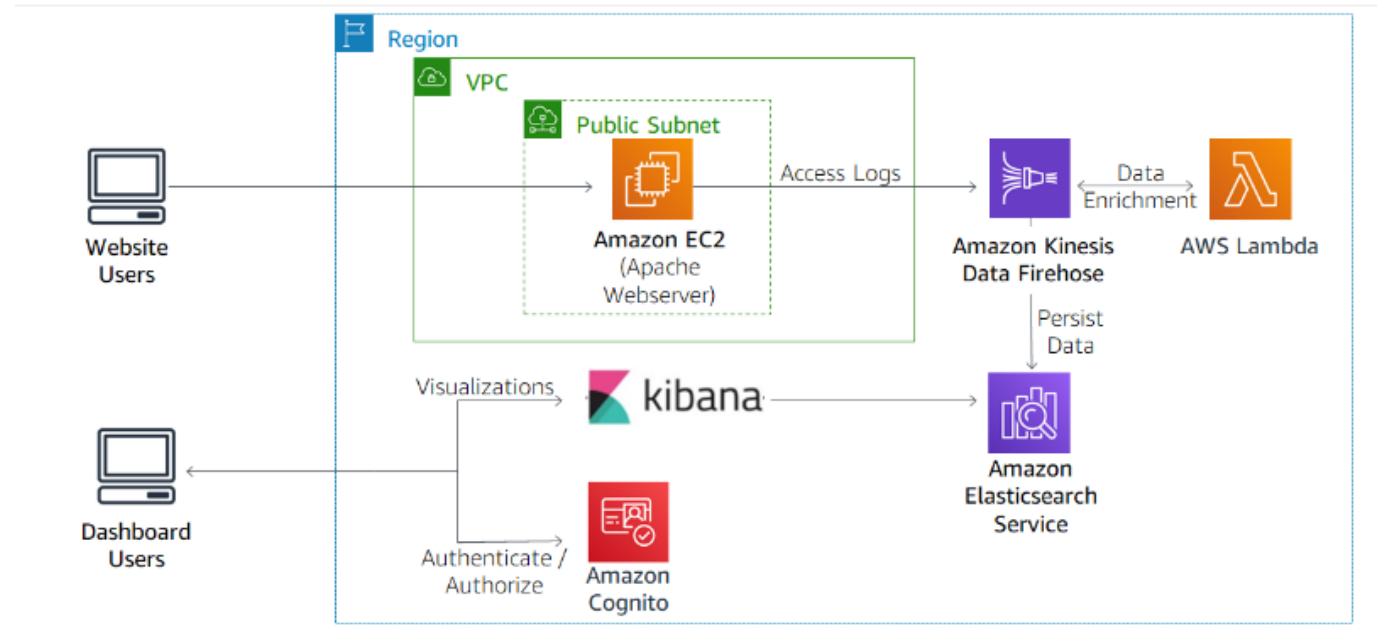


Figura - Cenário Lab 7

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.

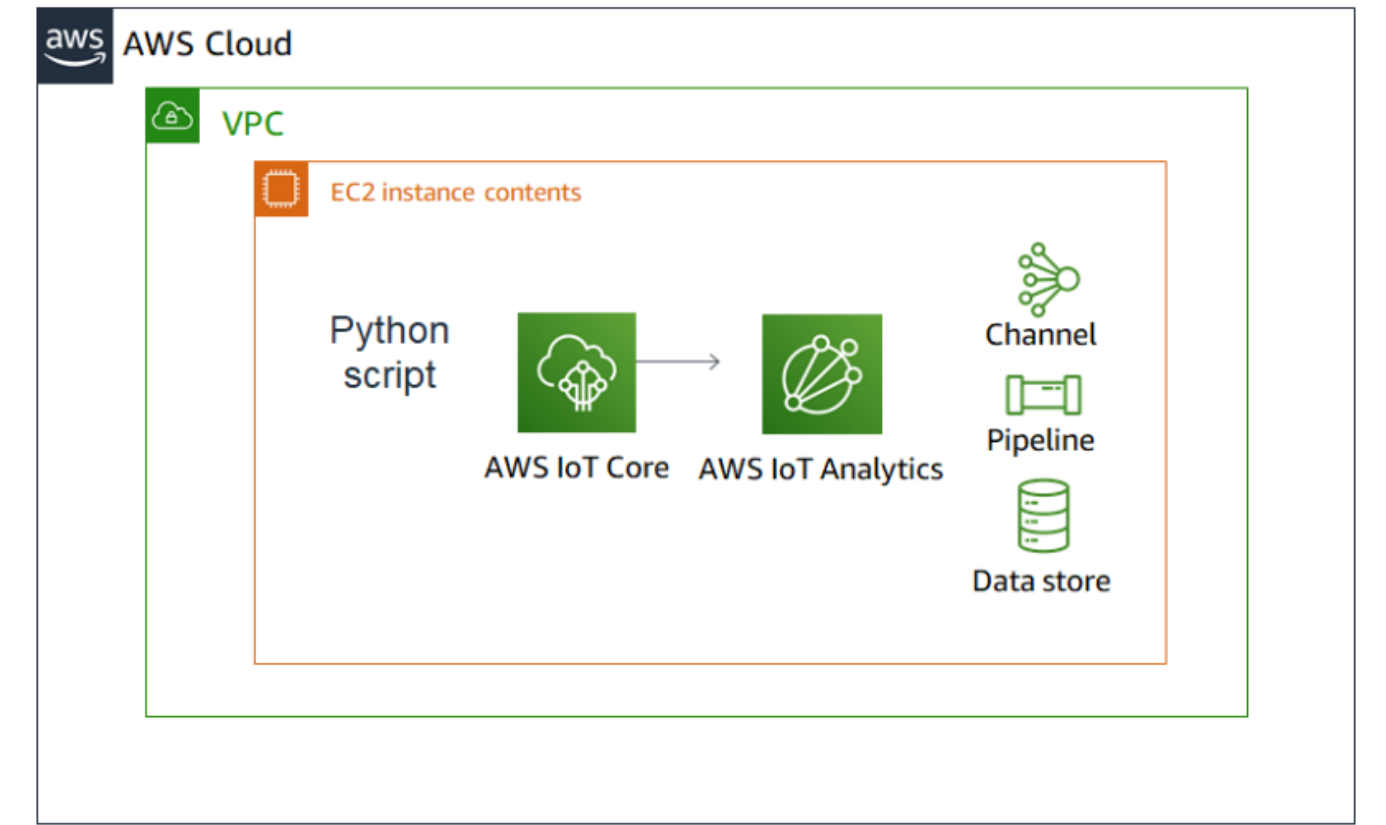


Figura - Cenário Lab 8

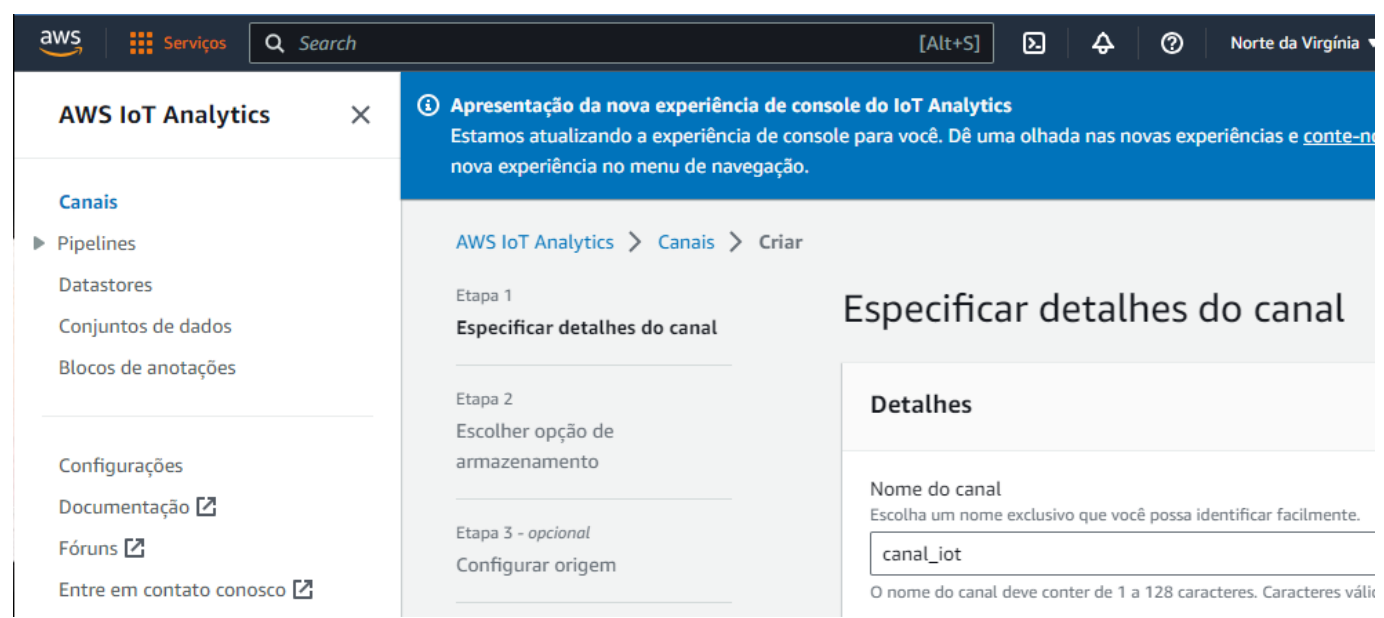


Figura - Criando um canal

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.

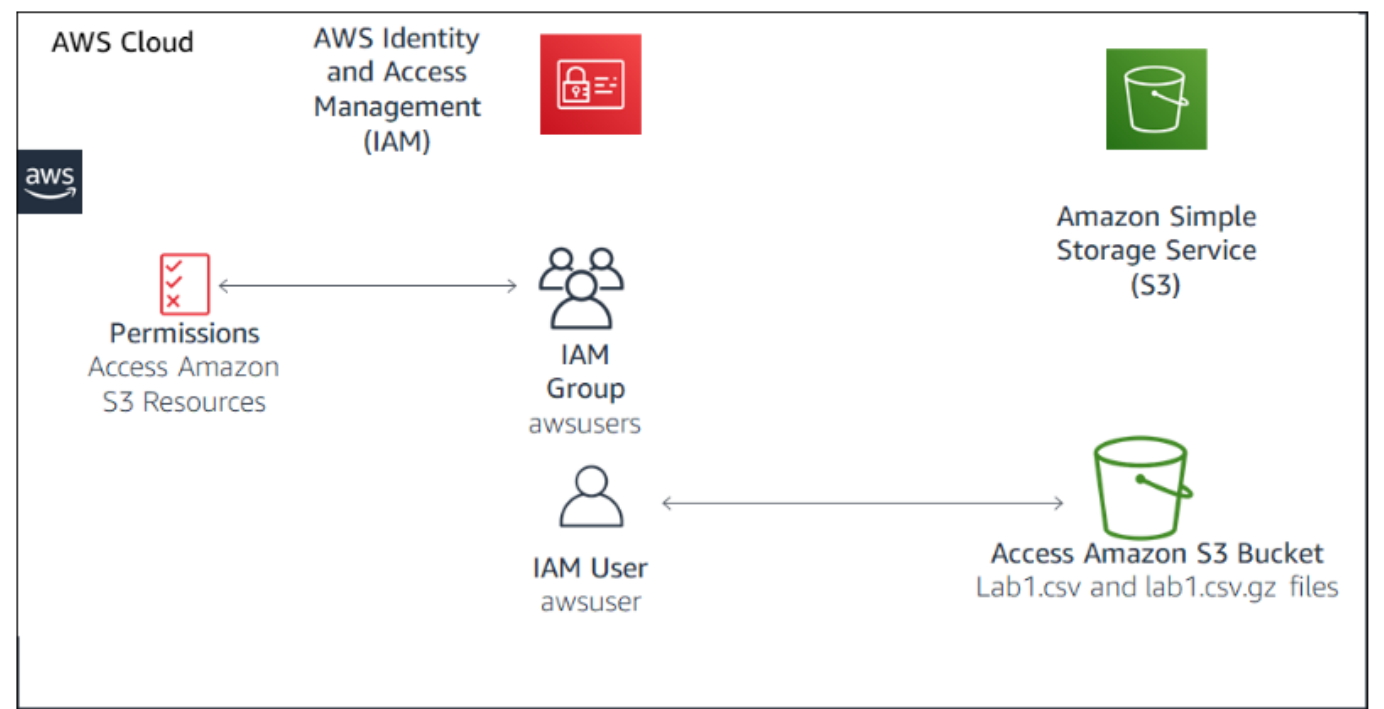


Figura - Cenário Lab 1

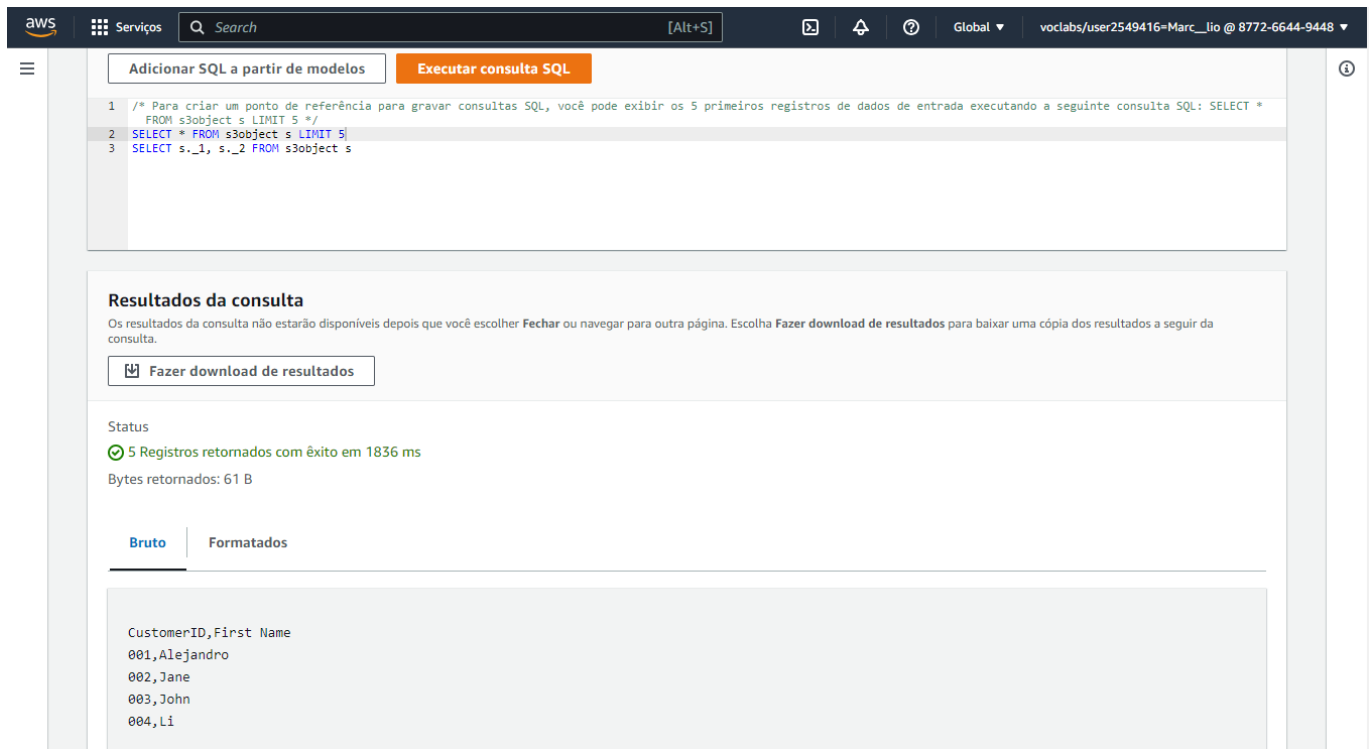


Figura - Carregando um objeto e fazendo consultas - Amazon S3 Service

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 mesmo. 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.

Referências

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