



escola
britânica de
artes criativas
& tecnologia

Profissão: Analista de dados



COMPUTAÇÃO EM NUVEM I



GUIA DA AULA 4



Consulte o AWS Athena



Acompanhe aqui
os temas que
serão tratados
na videoaula

- Introdução
- Funcionamento
- Preço
- Atividade



Introdução

O [AWS Athena](#) é um serviço de computação distribuída e sem servidor que atua como um motor de consulta (*query engine*). O serviço implementa na infraestrutura computacional da AWS o projeto *open source* [Presto](#).

A combinação do AWS Athena e o AWS S3 forma uma poderosa base de dados para processamento de grandes volumes de dados de baixo custo e alta latência, ideal para *data lakes*.



Funcionamento

1. Transforma uma *query* SQL em código [Spark](#);
2. Processa os objetos armazenados no AWS S3 com o código Spark gerado;
3. Salva o resultado da *query* SQL como um objeto em um *bucket* no AWS S3;
4. Retorna o resultado da *query* SQL no console.



Preço

O AWS Athena cobra por consulta. O preço atual é de 9,00 USD por *terabyte* escaneado (49,65 BRL aproximadamente). Você sempre deve consultar o preço na página web do serviço no *link* <https://aws.amazon.com/pt/athena/pricing/?nc=sn&loc=3>.



Atividade

- Criar um *bucket* no AWS S3 para armazenar os resultados;
- Configurar o AWS Athena para utilizar o *bucket* criado;
- Criar uma tabela (com SQL) utilizando o *bucket* da aula 3, item 3.3;
- Executar uma consulta SQL.



Código SQL para criar a tabela:

```

CREATE EXTERNAL TABLE iris (
  sepal_length double,
  sepal_width double,
  petal_length double,
  petal_width double,
  specie string
)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
WITH SERDEPROPERTIES ('separatorChar'=',')
LOCATION 's3://<bucket-name>/'
  
```



Códigos SQL para executar consultas:

```

SELECT * FROM iris LIMIT 10;

SELECT
  especie,
  round(avg(sepal_length),      4) as avg_sepal_length,
  round(avg(sepal_width),       4) as avg_sepal_width,
  round(avg(petal_length),      4) as avg_petal_length,
  round(avg(petal_width),       4) as avg_petal_width,

FROM iris
GROUP BY especie;
  
```

