**Atividade de Ciências de Dados**

**Aluno:** Heberth Vinícius Amarante Severo

**Descrição do problema:**

Essa é a sexta e última parte da nossa atividade de Web Scraping, iremos inserir os nossos dados em um banco de dados, para esta atividade foi escolhido o MongoDB.

Em cada etapa será explicado o que foi feito, com todas as imagens necessárias para o entendimento da atividade.

**1ª Etapa**

Para o início dessa atividade é necessário fazer o cadastro no *Atlas MongoDB*, para que possamos fazer a inserção dos dados, a Figura 1 mostra a tela de cadastro do *Atlas*.

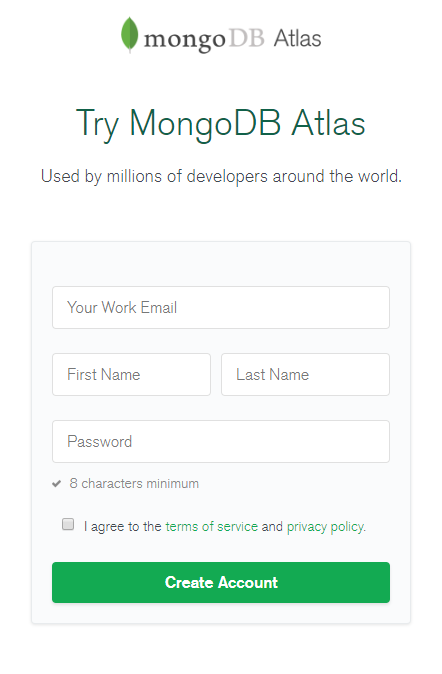


Figura 1 Tela de cadastro no Altas

Após o cadastro temos que escolher um dos tipos para a criação do *cluster*, como estamos no meio acadêmico, iremos escolher a versão gratuita, como mostra a Figura 2.

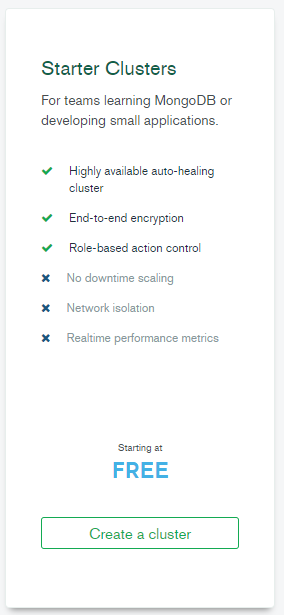


Figura Escolhendo o tipo do cluster

Nossos dados serão armazenados na nuvem, o Atlas nos dá três opções, poderíamos escolher qualquer uma delas, mas iremos escolher o Google Cloud, pois é o único que possui a opção de uma cidade do Brasil, a Figura 3 mostra esse processo.

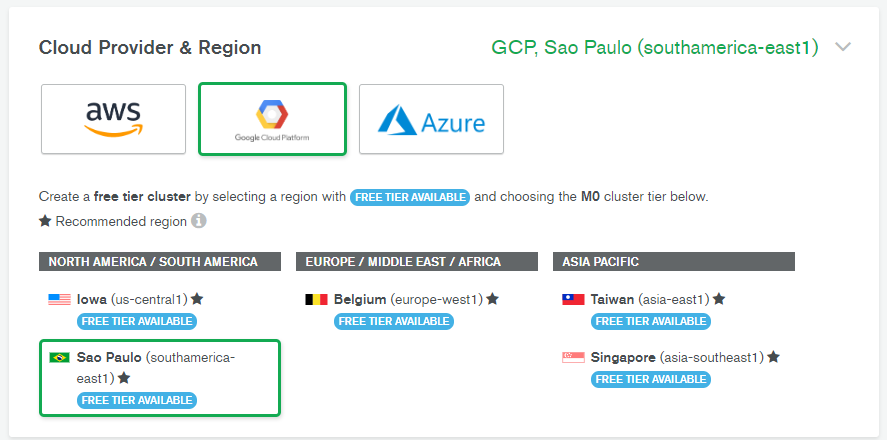


Figura Escolhendo uma nuvem

Antes de começarmos de fato a trabalhar com os dados devemos dar um nome para o nosso *cluster,* a Figura 4 mostra essa parte.

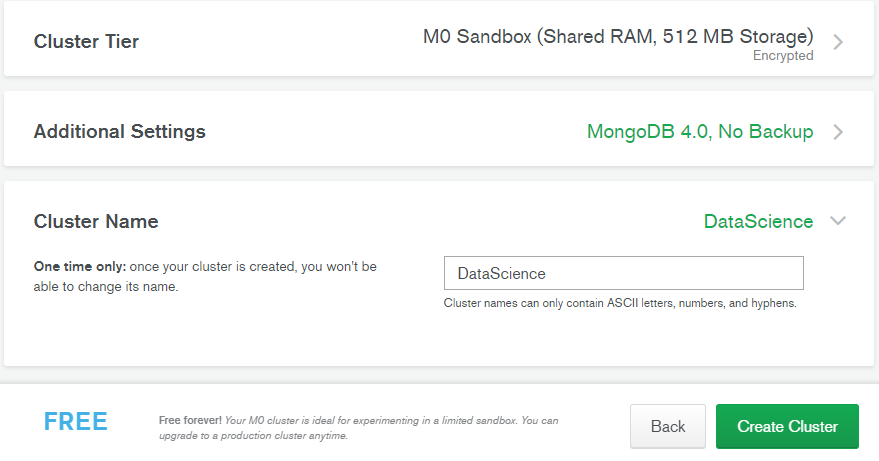


Figura Colocando o nome do cluster

**2ª Etapa**

Nessa etapa da atividade iremos começar a fazer o código no *Colab*, nessa parte iremos fazer todas as importações necessárias, iremos conectar com o Atlas, iremos criar nossa base e por fim iremos criar uma coleção, como mostra a Figura 5.

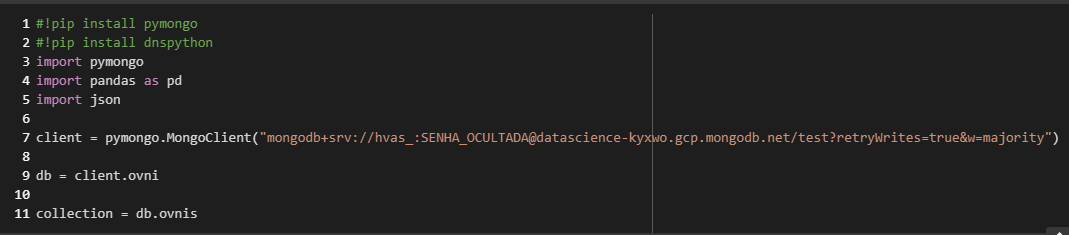


Figura 5 Importação e conexão com Atlas

A linha sete é gerada pelo próprio *Atlas* na parte de conexão, na linha nove criamos o banco com o nome *ovni*, na linha onze estamos criando uma coleção e chamamos ela de *ovnis*.

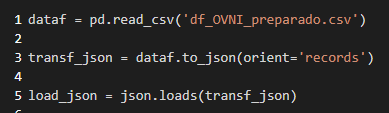


Figura 6 Lendo o csv e transformando em JSON

A Figura 6 é a leitura do nosso arquivo csv e a transformação desse arquivo em *JSON*. A primeira linha é onde está sendo feito a leitura do csv, na terceira linha estamos transformando o csv em um *JSON*, e na linha cinco estamos carregando esse *JSON*.

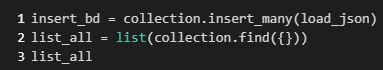


Figura Inserindo o JSON no banco e mostrando os dados

Na Figura 7 estamos inserindo o nosso *JSON* no banco como mostra a primeira linha, na segunda linha estamos listando todos os arquivos do nosso *JSON*. A Figura 8 mostra a saída desse trecho de código com todos os registros.

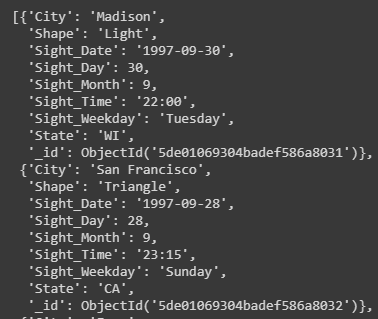


Figura Mostrando todos os registros

A Figura 9 mostra a contagem de registros que nosso arquivo possui e a Figura 10 mostra a saída desse trecho de código.



Figura Contagem dos registros

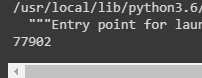


Figura Quantidade de registros

**3ª Etapa**

Nessa etapa irei mostrar como está visualmente o nosso cluster no *Atlas MongoDB*.

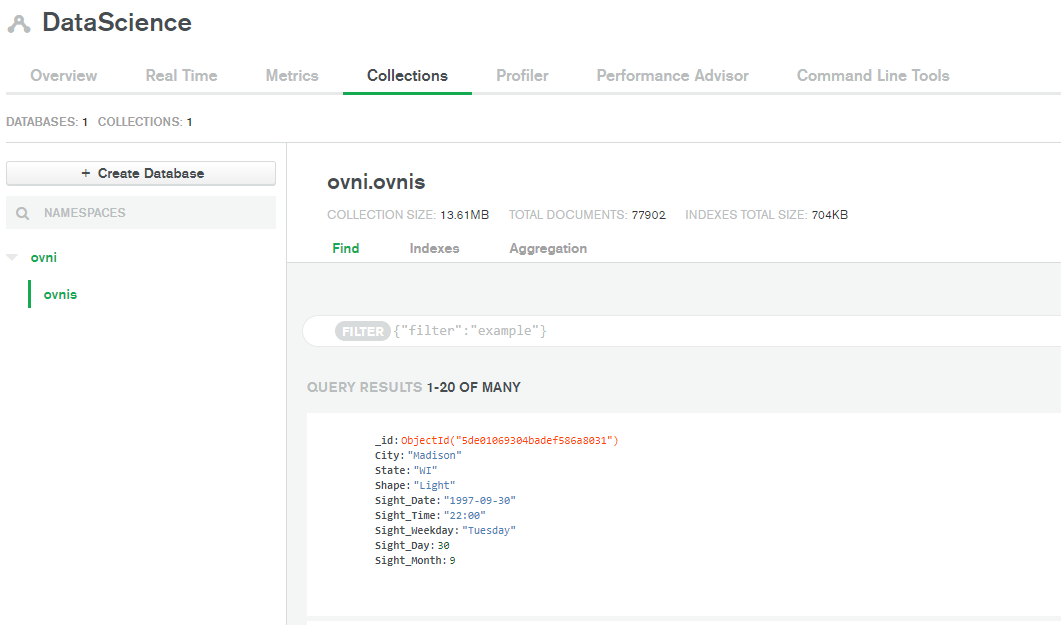


Figura Cluster no Atlas MongoDB

**4ª Etapa**

Nessa etapa iremos mostrar todos os registros do banco, mas eles devem estar ordenados pelo Shape, a Figura 12 mostra esse trecho de código.



Figura Listando os registros e ordenando por Shapes

A Figura 12 mostra a saída desse trecho de código.



Figura Saída do trecho de código anterior

**5ª Etapa**

Nessa etapa iremos mostrar os dados onde ocorreram mais registros ordenados de forma decrescente, a Figura 14 mostra o código e a Figura 15 mostra a saída desse trecho de código.



Figura Mostrando as ocorrencias ordenadas por registros

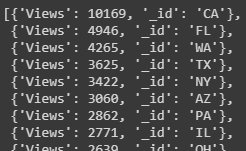


Figura Saída do trecho anterior

**6ª Etapa**

Nessa etapa iremos mostrar os dados que foram registrados da cidade de Phoenix, a Figura 16 mostra o trecho de código para essa etapa.



Figura Registros da cidade de Phoenix

A Figura 17 mostra a saída gerada por esse trecho de código.

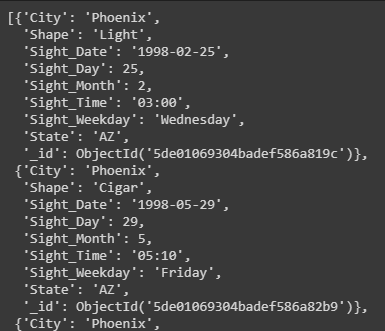


Figura Registros de Phoenix

**7ª Etapa**

Nessa etapa iremos buscar os registros feitos no estado da Califórnia, mas dessa vez iremos ocultar o id desses registros, a Figura 18 mostra o trecho de código para a realização dessa etapa.



Figura Registros da Califórnia sem o id

A Figura 19 mostra a saída desse trecho de código.

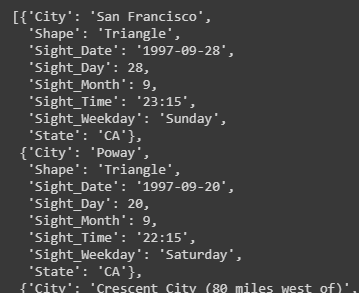


Figura Registros da Califórnia

**Observação**

Este relatório estará no GitHub, cujo link está a seguir: <https://github.com/heberth-hvas/icd_5>

**Referências**

**Google Colab:** <https://colab.research.google.com/notebooks/welcome.ipynb>

**Documentação Pandas:** <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.append.html>

**Atlas MongoDB:** <https://www.mongodb.com/atlas-signup-from-mlab?utm_source=mlab.com&utm_medium=referral&utm_campaign=mlab%20signup&utm_content=blue%20sign%20up%20button>

**Tutorial W3Schools:** <https://www.w3schools.com/python/python_mongodb_getstarted.asp>

**Pandas DataFrame.to\_json:** <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.to_json.html>

**Python JSON:** <https://docs.python.org/3/library/json.html>

**ORDERBY JSON:** <https://docs.mongodb.com/manual/reference/operator/meta/orderby/>

**Collection Aggregate JSON:** <https://docs.mongodb.com/manual/reference/method/db.collection.aggregate/>