

Nome: Vanessa de Andrade Formiga

Problema:

Nesta última parte do trabalho, tinha como problema a ser resolvido realizar o armazenamento do pipeline dos dados e armazenar esses dados utilizando o mongoDb.

Solução:

Para executar essas partes do trabalho teve a sua divisão das tarefas em etapas.

Primeira Etapa:

Na primeira etapa a atividade que precisava ser feita era a criação de uma no mongoDb Atlas, para criar essa conta precisou colocar o email, nome, sobrenome, senha e aceitar os termos e por último clicar em Create Account.

A imagem 1 mostrar o resultado obtido na criação da conta no mongoDb.



Try MongoDB Atlas

Used by millions of developers around the world.

A screenshot of the MongoDB Atlas account creation form. It features a light gray background with white input fields. The fields are labeled "Your Work Email", "First Name", "Last Name", and "Password". Below the password field, there is a checkmark icon followed by the text "8 characters minimum". Further down, there is a checkbox icon followed by the text "I agree to the terms of service and privacy policy." At the bottom of the form is a prominent green button with the text "Create Account" in white.

Your Work Email

First Name Last Name

Password

✓ 8 characters minimum

☐ I agree to the [terms of service](#) and [privacy policy](#).

Create Account

imagem 1

Segunda etapa:

Depois de criar a conta precisa selecionar a opção gratuita para iniciar a Clusters. A imagem 2 mostra o resultado obtido.

Starter Clusters

For teams learning MongoDB or developing small applications.

- ✓ Highly available auto-healing cluster
- ✓ End-to-end encryption
- ✓ Role-based action control
- ✗ No downtime scaling
- ✗ Network isolation
- ✗ Realtime performance metrics

Starting at

FREE

Create a cluster

imagem 2

Terceira etapa:

A atividade desta etapa era escolher o provedor e a região, e foram escolhidos o AWS e o estado da Virginia. A imagem 3 demonstrar como ocorreu a escolha.

Cloud Provider & Region

AWS, N. Virginia (us-east-1) ▼



Create a **free tier cluster** by selecting a region with **FREE TIER AVAILABLE** and choosing the **M0** cluster tier below.

★ Recommended region ⓘ

imagem 3

Quarta etapa:

A imagem 4 mostrar como ocorreu a escolha da cluster com a memória de 512Mb.

Cluster Tier **M0 Sandbox (Shared RAM, 512 MB Storage)** Encrypted ▼

Base hourly rate is for a MongoDB replica set with **3 data bearing servers**.

Shared Clusters for development environments and low-traffic applications

Tier	RAM	Storage	vCPU	Base Price
✓ M0 Sandbox	Shared	512 MB	Shared	Free forever
M0 clusters are best for getting started, and are not suitable for production environments.				
100 max connections Low network performance 100 max databases 500 max collections				
M2	Shared	2 GB	Shared	\$9 / MONTH
M5	Shared	5 GB	Shared	\$25 / MONTH

imagem 4

Quinta etapa:

A imagem 5 mostra a opção de não escolher fazer o backup no mongoDb.

Additional Settings **MongoDB 4.0, No Backup** ▼

Turn on Backup (M2 and up)

☐ NO

[See Backup Solutions for Paid Clusters \(M2+\)](#)

imagem 5

Sexta etapa:

A imagem 6 mostrar como realizou a definição do nome para a cluster, o nome escolhido foi Nuforc.

Cluster Name **Nuforc** ▼

One time only: once your cluster is created, you won't be able to change its name.

Nuforc

Cluster names can only contain ASCII letters, numbers, and hyphens.

imagem 6

Sétima etapa:

A imagem 7 mostrar a criação da cluster.

FREE

Free forever! Your M0 cluster is ideal for experimenting in a limited sandbox. You can upgrade to a production cluster anytime.

Back

Create Cluster

Oitava etapa:

imagem 7

A imagem 8 demonstrar a inicialização da cluster.

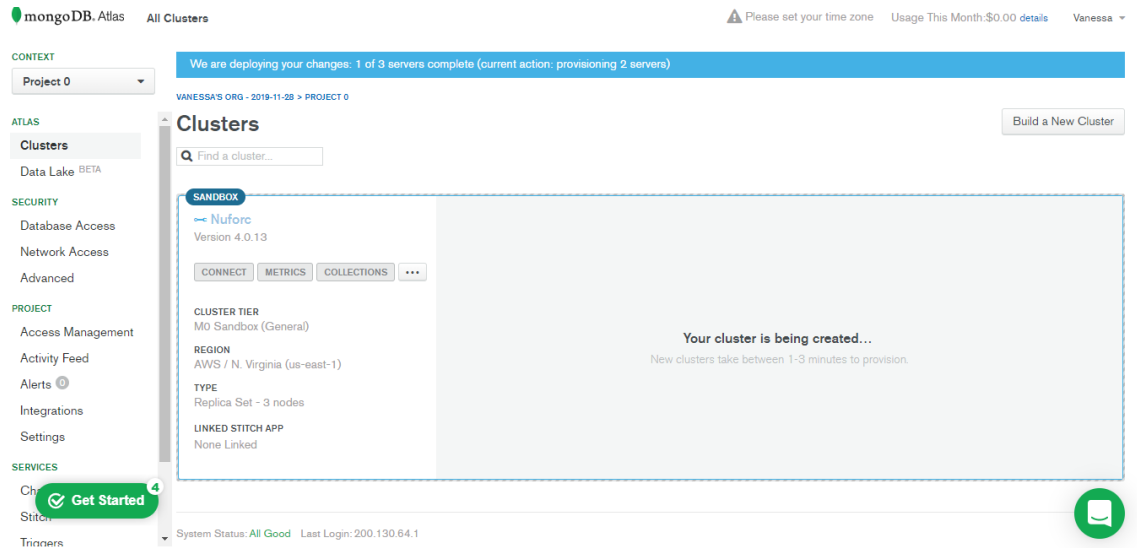


imagem 8

A nona etapa:

Na nona etapa precisava configurar a conexão do Database Access. Para configurar o banco de dados com nome do usuário e senha. O resultado desta criação está sendo demonstrada nas imagens 9,10 e 11.

Database Access



imagem 9

Add New User

Choose Method

PASSWORD **CERTIFICATE**

SCRAM Authentication

SCRAM is MongoDB's default authentication method.

dataScience

e.g. new-user_31

.....

SHOW

Autogenerate Secure Password

User Privileges

Atlas admin

Read and write to
any database

Only read any
database

Select Custom
Role

[Add Default Privileges](#)

☐ Save as temporary user

Cancel

Add User

imagem 10

Mostra o nome do usuário e a senha

Database Access

MongoDB Users

MongoDB Roles

+ ADD NEW USER

User Name	Authentication Method	MongoDB Roles	Actions
dataScience	SCRAM	atlasAdmin@admin	EDIT DELETE

imagem 11

Decima etapa:

Fazer a configuração do acesso a internet, e nesta opção precisa permitir conexão para todos os ips. O resultado está sendo exibido na imagens 12,13 e 14.

Network Access

IP Whitelist

Peering

Private Endpoint

IP Address

Comment

Status

Click here to add your first IP address

+ ADD IP ADDRESS



Whitelist an IP address

Configure which IP addresses can access your cluster.

[Learn more](#)

imagem 12

Add Whitelist Entry

Add a whitelist entry using either CIDR notation or a single IP address. [Learn more](#).

ADD CURRENT IP ADDRESS

ALLOW ACCESS FROM ANYWHERE

Whitelist Entry:

0.0.0.0/0

Comment:

Optional comment describing this entry

☐ Save as temporary whitelist

Cancel

Confirm

imagem 13

Network Access

IP Whitelist

Peering

Private Endpoint

+ ADD IP ADDRESS

You will only be able to connect to your cluster from the following list of IP Addresses:

IP Address	Comment	Status	Actions
0.0.0.0/0 (includes your current IP address)		<div>Pending</div>	<div>EDIT</div> <div>DELETE</div>

imagem 14

Clusters

[Build a New Cluster](#)

SANDBOX

Nuforc
Version 4.0.13

[CONNECT](#) [METRICS](#) [COLLECTIONS](#) [...](#)

CLUSTER TIER
M0 Sandbox (General)

REGION
AWS / N. Virginia (us-east-1)

TYPE
Replica Set - 3 nodes

LINKED STITCH APP
None Linked

Operations R: 0 W: 0
100.0/s
0

Logical Size 0.0 B
512.0 MB max
0.0 B

Connections 0
500 max
0

Enhance Your Experience
For dedicated throughput, richer metrics and enterprise security options, upgrade your cluster now!
[Upgrade](#)

Click "...". and then press Load Sample Dataset.

imagem 15

Load Sample Dataset

We've created a sample dataset to help you test features on Nuforc.

[Sample Dataset](#)

Size: ~350 MB

Please confirm that you want to load this sample dataset.

[Cancel](#)[Load Sample Dataset](#)

imagem 16

Decima primeira etapa:

Nesta etapa realizou a conexão com driver e depois seleccionar a conexão string only, copiou essa conexão e colou no código do colab.

Connect to Nuforc

✓ Setup connection security

✓ Choose a connection method

Connect

1 Choose your driver version

DRIVER

Python

VERSION

3.6 or later

2 Add your connection string into your application code

Connection String Only

Full Driver Example

mongodb+srv://dataScience:<password>@nuforc-sflsg.mongodb.net/test?re

Copy

Replace **<password>** with the password for the **dataScience** user.
When entering your password, make sure that any special characters are [URL encoded](#).

Having trouble connecting? [View our troubleshooting documentation](#)

imagem 17

Decima segunda etapa:

A tarefa realizada definida nesta etapa foi a execução do tutorial do python mongoDb para conhecer como funciona o pymongo e o mongoDb. Para utilizar o módulo PyMongo, precisar fazer a instalação deste módulo usando o comando `!pip install dnspython` e depois de fazer essa instalação precisa reiniciar o colab, e restart runtime, para estabelecer conexão de dados com o banco de dados.

Decima terceira etapa:

Para fazer essa tarefa teve com operação a ser realizada foi a importação do pymongo, e json. E a instalação do mongo e do dnspyt.

```
import pandas as pd
import pymongo
#!pip install mongo
#!pip install dnspyt
import dns
import json
```

Decima quarta etapa:

Para realizar a conexão definiu uma variável o comando ocorreu a conexão, mais o nome do banco de dados e a senha.

```
myclient = pymongo.MongoClient("mongodb+srv://dataScience:h7sixIgapVJs9zp@nuforc-2yhd7.azure.mongodb.net/test?retryWrites=true&w=majority")
# bancoDados = myclient.test
```

Imagem 18

```
Database(MongoClient(host=['nuforc-shard-00-02-2yhd7.azure.mongodb.net:27017', 'nuforc-shard-00-01-2yhd7.azure.mongodb.net:27017', 'nuforc-shard-00-00-01-2yhd7.azure.mongodb.net:27017'], document_class=dict, tz_aware=False, connect_pool=True, read_preference=ReadPreference.PRIMARY, ssl=False, uri='mongodb+srv://dataScience:h7sixIgapVJs9zp@nuforc-2yhd7.azure.mongodb.net/test?retryWrites=true&w=majority'))
```

Decima quinta etapa:

A imagem 19 mostra como realizou a criação do banco de dados denominado ovni.

```
# #cria um banco de dados
# bancoDados = myclient.test
# print(myclient)

#4.cria um banco de dados com nome ovni
bancoDados = myclient.ovni
print(bancoDados)
```

Imagem 19

```
Database(MongoClient(host=['nuforc-shard-00-02-2yhd7.azure.mongodb.net:27017', 'nuforc-shard-00-01-2yhd7.azure.mongodb.net:27017', 'nuforc-shard-00-00-01-2yhd7.azure.mongodb.net:27017'], document_class=dict, tz_aware=False, connect_pool=True, read_preference=ReadPreference.PRIMARY, ssl=False, uri='mongodb+srv://dataScience:h7sixIgapVJs9zp@nuforc-2yhd7.azure.mongodb.net/test?retryWrites=true&w=majority'))
```

Decima sexta etapa:

A imagem 20 mostra como realizou a criação da coleção denominado ovnis.

```
#5.cria uma coleção com nome de ovnis
collections = bancoDados.ovnis
```

Imagem 20

Decima sétima etapa:

A imagem 21 mostra a inserção da coleção usando o arquivo df_OVNI_preparado.

```
#leitura do arquivo csv
#6.inserir da coleção criada todos os registros do df_OVNI_preparado
df = pd.read_csv("df_OVNI_preparado.csv")
I
```

Imagem 21

Decima oitava etapa:

A imagem 22 mostra como ocorreu a transformação em json.

```
docs = json.loads(df.iloc[:,0:].to_json(orient='records'))
df.shape[0]
collections.insert_many(docs)
```

Imagem 22

A imagem 23 mostrar como ocorreu a contagem das ocorrências em ovnis, e retornou a quantidade de 77902.

```
#!curl ipecho.net/plain
#7.utilizando as funções do pymongo
#1. contar e mostrar quantos documentos há na coleção ovnis
ovnis = collections.count()
```

Imagem 23

77902

Imagem 24

A imagem 25 demonstrar como realizou o resgate dos registros e a ordenação por Shape.

```
#2. resgatar todos os documentos(registros) da coleção ovnis e order por tipo (shape)
myCursor = list(collections.find().sort("Shape"))
```

Imagem 25

```
[{'City': 'West Manchester',
  'Shape': 'Changing',
  'Sight_Date': '09/18/1997',
  'Sight_Day': 18,
  'Sight_Month': 9,
  'Sight_Time': '04:15:00',
  'Sight_Weekdays': 'Thursday',
  'State': 'OH',
  '_id': ObjectId('5de0141ee8ba993d39d03f35')},
{'City': 'Louisville',
  'Shape': 'Changing',
  'Sight_Date': '11/08/1997',
  'Sight_Day': 8,
  'Sight_Month': 11,
  'Sight_Time': '19:30:00',
  'Sight_Weekdays': 'Saturday',
  'State': 'KY',
  '_id': ObjectId('5de0141ee8ba993d39d03ff9')}]
```

Imagem 26

A imagem 27 demonstra o resultado obtido para mostra as ocorrências por estados.

```
#3.verificar quantas ocorrências existentes por estado
estado = list(collections.aggregate([{'$group':{'Views':{'$sum':1}, '_id':'$State'}}, {"$sort" : {"Views": -1}}]))
```

Imagem 27

```
[{'Views': 10169, '_id': 'CA'},
 {'Views': 4946, '_id': 'FL'},
 {'Views': 4265, '_id': 'WA'},
 {'Views': 3625, '_id': 'TX'},
 {'Views': 3422, '_id': 'NY'},
 {'Views': 3060, '_id': 'AZ'},
 {'Views': 2862, '_id': 'PA'},
 {'Views': 2771, '_id': 'IL'},
 {'Views': 2620, '_id': 'OH'}]
```

Imagem 28

A imagem 29 mostra como ocorreu a busca de todas as ocorrências por Phoenix.

```
#4. buscar todas as ocorrências da cidade Phoenix
city = list(collections.find({'City': 'Phoenix'}))
```

Imagem 29

```
{'Sight_Time': '15:35:00',
 'Sight_Weekdays': 'Thursday',
 'State': 'AZ',
 '_id': ObjectId('5de0141fe8ba993d39d16bd3')},
{'City': 'Phoenix',
 'Shape': 'Other',
 'Sight_Date': '06/07/2017',
 'Sight_Day': 7,
 'Sight_Month': 6,
 'Sight_Time': '08:20:00',
 'Sight_Weekdays': 'Wednesday',
 'State': 'AZ',
 '_id': ObjectId('5de0141fe8ba993d39d16c38')},
{'City': 'Phoenix',
 'Shape': 'Fireball',
 'Sight_Date': '07/26/2017',
 'Sight_Day': 26,
 'Sight_Month': 7,
 'Sight_Time': '04:20:00',
 'Sight_Weekdays': 'Wednesday',
 'State': 'AZ',
 '_id': ObjectId('5de0141fe8ba993d39d16cbb')},
{'City': 'Phoenix',
 'Shape': 'Oval',
 'Sight_Date': '07/06/2017',
 'Sight_Day': 6,
 'Sight_Month': 7,
 'Sight_Time': '21:25:00',
 'Sight_Weekdays': 'Thursday',
 'State': 'AZ',
 '_id': ObjectId('5de0141fe8ba993d39d16da6')},
{'City': 'Phoenix',
```

```

'Sight_Month': 12,
'Sight_Time': '22:00:00',
'Sight_Weekdays': 'Wednesday',
'State': 'CA'},
{'City': 'Roseville',
'Shape': 'Triangle',
'Sight_Date': '12/20/2000',
'Sight_Day': 20,
'Sight_Month': 12,
'Sight_Time': '05:45:00',
'Sight_Weekdays': 'Wednesday',
'State': 'CA'},
{'City': 'Woodland',
'Shape': 'Light',
'Sight_Date': '12/19/2000',
'Sight_Day': 19,
'Sight_Month': 12,
'Sight_Time': '23:30:00',
'Sight_Weekdays': 'Tuesday',
'State': 'CA'},
{'City': 'Sebastopol',
'Shape': 'Circle',
'Sight_Date': '12/17/2000',
'Sight_Day': 17,
'Sight_Month': 12,
'Sight_Time': '18:00:00',
'Sight_Weekdays': 'Sunday',
'State': 'CA'},
{'City': 'Palo Alto',
'Shape': 'Disk',
'Sight_Date': '12/15/2000',
'Sight_Day': 15,
'Sight_Month': 12,
'Sight_Time': '12:00:00',
'Sight_Weekdays': 'Friday',
'State': 'CA'}

```

Imagem 30

A imagem 31 demonstra como ocorreu a ocorrência por estados da Califórnia.

```

#5. Buscar as ocorrências do estado da Califórnia e ocultar o id de cada documento (registro).
ocorrencias_states = list(collections.find({'State': 'CA'}, {'_id': 0} ))

```

Imagem 31

```
[{'City': 'San Francisco',  
  'Shape': 'Triangle',  
  'Sight_Date': '09/28/1997',  
  'Sight_Day': 28,  
  'Sight_Month': 9,  
  'Sight_Time': '23:15:00',  
  'Sight_Weekdays': 'Sunday',  
  'State': 'CA'},  
{ 'City': 'Poway',  
  'Shape': 'Triangle',  
  'Sight_Date': '09/20/1997',  
  'Sight_Day': 20,  
  'Sight_Month': 9,  
  'Sight_Time': '22:15:00',  
  'Sight_Weekdays': 'Saturday',  
  'State': 'CA'},  
{ 'City': 'Crescent City (80 miles west of',  
  'Shape': 'Disk',  
  'Sight_Date': '09/19/1997',  
  'Sight_Day': 19,  
  'Sight_Month': 9,  
  'Sight_Time': '22:00:00',  
  'Sight_Weekdays': 'Friday',  
  'State': 'CA'},  
{ 'City': 'Delilah Lookout (Sequoia N F)',  
  'Shape': 'Disk',  
  'Sight_Date': '09/17/1997',
```

Imagem 32

Link Para o Github:

<https://github.com/vanessaformiga/Nuforc/tree/master/dadosMongoDB>

Referências:

https://www.w3schools.com/python/python_mongodb_getstarted.asp

<https://colab.research.google.com/>

<https://www.mongodb.com/>

<https://docs.mongodb.com/>

<https://api.mongodb.com/python/current/index.html>