**Relatório de envio dos dados coletados do site *nuforc.org* no banco de dados mongodb**

Aluno: Daniel Barros Candeira

1. **Introdução**

O relatório tem por objetivo demonstrar as etapas percorridas para enviar os dados coletados do site *nuforc.org* entre 07 de setembro de 1997 e 07 de agosto de 2017 para o banco de dados mongodb, o site nuforc contém relatos de “visualização” de OVNIs – Objetos Voadores Não Identificados, a fim de obter as informações relevantes.

1. **Problema**

Nessa etapa será realizado o envio de dados para a plataforma em nuvem, tendo como atividades, carregar o arquivo df\_OVNI\_preparado.csv, converter o dataframe para o formato JSON, enviar os dados para o banco de dados na plataforma em nuvem, contar quantos documentos foram criados no banco, todos os documentos (registros) da coleção ovnis e ordenar por tipo (shape), verificar quantas ocorrências existem por estado, buscar todas as ocorrências da cidade *Phoenix*, buscar as ocorrências do estado da Califórnia e ocultar o id de cada documento (registro).

1. **Resolução**

A fim de realizar os passos identificados na seção 2, foi desenvolvido um script utilizando a linguagem de programação **Python**, e a biblioteca **Pandas**, **json** e **pymongo** que servem respectivamente para lidar com dados de forma estatística e com base/arquivos de dados volumosos, converter/ler arquivos JSON, e realizar a comunicação com banco de dados.

**1º Passo:**

Antes de iniciar a escrita do código, é realizada a configuração do ambiente em nuvem, que realiza a conexão do banco de dados. Primeiro criamos uma conta gratuita (**Fig 1**), após criar a conta, realizamos as configurações do cluster (**Fig 2**), em sequência realizamos as configurações da comunicação da aplicação com o baco de dados (**Fig 3**), acessamos a url de conexão (**Fig 4**), e por fim configuramos o acesso ao banco com IP’s (**Fig 5**).

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 1. Página de cadastro.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 2. Configurando o cluster.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 3. Configurando a conexão com o banco de dados.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 4. Acessando os dados para conexão com o banco de dados.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 5. Adicionando itens a lista de acesso.

**2º Passo:**

Realizamos então a instalação do *dnspython* (**linha 1, Fig 6**) e a reinicialização do runtime do **Colab** (‘Ambiente de programação’), após isso são importadas as bibliotecas a serem utilizadas (**Fig 6)**.

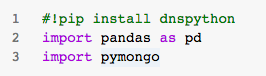


Fig 6. Importando as bibliotecas.

Após isso iniciamos a conexão com o banco de dados (**Fig 7**), criamos o banco (**Fig 8**) e a coleção (**Fig 9**).



Fig 7. Realizando conexão com banco de dados.



Fig 8. Criando banco de dados.



Fig 9. Criando coleção.

Realizamos a importação do dataframe (**Fig 10**) e convertemos o dataframe para o formato JSON usando a biblioteca **json**, e enviamos os dados para o banco (**Fig 11**).



Fig 10. Importando dataframe.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 11. Convertendo o dataframe em JSON e inserindo os dados no banco de dados.

**3º Passo:**

O terceiro passo é contar quantos documentos foram criados no banco de dados (**Fig 12**)**.**

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 12. Contando documentos.

**4º Passo:**

O quarto passo é selecionar todos os documentos (registros) da coleção ovnis e ordenar por tipo (shape), como é demonstrado na **Fig 13**.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 13. Selecionando todos os documentos e ordenando por *shape*.

**5º Passo:**

No quinto passo verificamos quantas ocorrências foram registradas por estado (**Fig 14**).

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 14. Verificando ocorrências por estados.

Uma imagem contendo interior

Descrição gerada automaticamente

Fig 15. Dados retornados da consulta.

**6º Passo:**

No sexto selecionamos todas as ocorrências que ocorreram na cidade *Phoenix* (**Fig 16**).

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 16. Selecionando ocorrências da cidade de Phoenix.

**7º Passo:**

No sétimo e último passo selecionamos todas as ocorrências que ocorreram no estado da Califórnia e escondemos o id dos documentos retornados na consulta (**Fig 17**).

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Fig 17. Selecionando ocorrências que ocorreram e Califórnia.

**Github**:<https://github.com/dan-candeira/Nuforc-Project/tree/master/5.10-DadosMongodb>

**Referências**

<http://www.dba86.com/docs/mongo/3.2/tutorial/project-fields-from-query-results.html>

<https://stackoverflow.com/questions/23116330/mongodb-select-count-group-by>

<https://stackoverflow.com/questions/50639768/colaboratory-virtual-instances-ip-range>