

09

## Mão na massa: Incrementando a coleção clima\_consolidado

Chegou a hora de você executar o que foi visto na aula! Para isso, execute os passos listados abaixo.

1) Baixe [aqui \(https://s3.amazonaws.com/caelum-online-public/799-data-science-preparacao-e-exploracao-dos-dados/01/Mescla\\_Colecoes\\_v2\\_Util.zip\)](https://s3.amazonaws.com/caelum-online-public/799-data-science-preparacao-e-exploracao-dos-dados/01/Mescla_Colecoes_v2_Util.zip), a nova versão do código responsável por mesclar as coleções `ufo.ufos` e `dbclima.clima`, o `Mescla_Colecoes_v2.py`. Dentro do zip você também encontrará a biblioteca de utilidades, `Util.py`.

2) Para restaurar a base de dados e coleção do MongoDB, primeiramente disponibilize o seu servidor, por exemplo:

```
start /b mongod --dbpath D:\MongoDB\data\db --logpath D:\MongoDB\log.txt --oplogSize 50 --smallfiles
```

3) Baixe [aqui \(https://s3.amazonaws.com/caelum-online-public/799-data-science-preparacao-e-exploracao-dos-dados/01/clima bson\)](https://s3.amazonaws.com/caelum-online-public/799-data-science-preparacao-e-exploracao-dos-dados/01/clima bson), o arquivo `clima.bson`, e restaure a coleção através deste arquivo, por exemplo:

```
mongorestore --db dbclima --collection clima --drop "D:\MongoDB\bkp_clima_curso2\dbclima\clima.bson"
```

4) Baixe [aqui \(https://s3.amazonaws.com/caelum-online-public/799-data-science-preparacao-e-exploracao-dos-dados/01/ufos bson\)](https://s3.amazonaws.com/caelum-online-public/799-data-science-preparacao-e-exploracao-dos-dados/01/ufos bson), o arquivo `ufos.bson`, e restaure a coleção através deste arquivo, por exemplo:

```
mongorestore --db ufos --collection ufo --drop "D:\MongoDB\bkp_ufo\ufos\ufo.bson"
```

5) Pelo próprio terminal, você pode contar quantos documentos há em cada coleção:

```
mongo dbclima --eval "db.clima.count()" --quiet
mongo ufos --eval "db.ufo.count()" --quiet
```

6) A coleção `clima_consolidado` só deve estar com os documentos trazidos do **Kaggle**, que possuem o atributo `cor`. Logo, remova os documentos que não possuem esse atributo, executando o seguinte comando, também pelo terminal:

```
mongo dbclima --eval "db.clima_consolidado.remove ( { cor : { $exists : 0 } } )" --quiet
```

Com isso feito, você pode executar a rotina.

7) Caso a rotina seja executada com sucesso, a coleção `clima_consolidado` será incrementada. Pelo próprio terminal, você pode contar novamente quantos documentos há em cada coleção:

```
mongo dbclima --eval "db.clima.count()" --quiet
mongo dbclima --eval "db.clima_consolidado.count()" --quiet
mongo ufos --eval "db.ufo.count()" --quiet
```

8) Com a coleção crescendo, crie um índice único, que só valerá para os documentos que tenham posição. Através do **Robo 3T**, na base de dados **dbclima**, execute:

```
db.clima_consolidado.createIndex (
    { posicao : 1 } ,
    {
        name : "ix_clima_consolidado_posicao" ,
        unique : true,
        partialFilterExpression: {
            posicao: { $exists: true }
        }
    }
)
```

9) Restaure a nova coleção **clima**, com mais documentos, que você pode baixar [aqui](https://s3.amazonaws.com/caelum-online-public/799-data-science-preparacao-e-exploracao-dos-dados/01/clima/clima.bson) (<https://s3.amazonaws.com/caelum-online-public/799-data-science-preparacao-e-exploracao-dos-dados/01/clima/clima.bson>), por exemplo:

```
mongorestore --db dbclima --collection clima --drop "D:\MongoDB\bkp_clima_curso3\dbclima\clima.bson"
```

10) Agora, execute novamente a rotina **Mescla\_Colecoes\_v2.py**. E terminada a execução, pode ver os novos tamanhos das coleções. Execute, pelo terminal:

```
mongo dbclima --eval "db.clima.count()" --quiet
mongo dbclima --eval "db.clima_consolidado.count()" --quiet
mongo ufos --eval "db.ufos.count()" --quiet
```

11) Para ver se a base está coerente, verifique se os últimos 5 documentos da coleção **clima** possuem as mesmas posições do que os 5 últimos documentos da coleção **clima\_consolidado**:

```
db.clima.find( { } ,
    { _id : 0 , posicao : 1}
).sort ( { posicao : -1 }). limit (5)

db.clima_consolidado.find( { } ,
    { _id : 0 , posicao : 1}
).sort ( { posicao : -1 }). limit (5)
```