

Implementação de Estratégias de Particionamento de Dados em MongoDB para Otimização de Desempenho

Paulo Victor Sobrinho de Jesus

2024

Objetivo:

O objetivo desta atividade é permitir que se apliquem conceitos de dados em um ambiente de banco de dados distribuído. Eles devem projetar um sistema que seja escalável e eficientemente particionado para lidar com um grande volume de dados.

Requisitos Iniciais:

Cada filial possui um grande volume de produtos em seu estoque.

O sistema precisa ser capaz de lidar com milhões de registros de produtos.

A consulta de estoque e atualizações de inventário devem ser rápidas e eficientes.

A escalabilidade do sistema é essencial, pois novas filiais podem ser adicionadas no futuro.

Configuração do ambiente

Para configurar um cluster MongoDB que utiliza replicação e sharding, vamos precisar configurar três tipos principais de serviços ou componentes. Cada um desses serviços desempenha um papel específico na gestão de dados e na garantia de alta disponibilidade e escalabilidade. Aqui estão eles:

Repositório Git: https://github.com/pvsobrinho/puc-minas-nosql-docker-final

1. Config Servers

 Função: Os servidores de configuração armazenam as metadados do cluster de sharding. Estas informações incluem a estrutura e a localização dos dados dentro do cluster, mapeando quais fragmentos (shards) contêm quais dados.

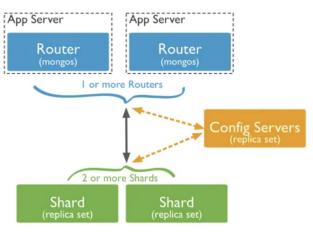
- **Importância**: Eles são cruciais para o funcionamento do cluster porque sem os dados de configuração, o sistema não saberia onde encontrar partes específicas dos dados distribuídos pelo cluster.
- Redundância: Geralmente, é recomendado ter um conjunto de réplicas de servidores de configuração para garantir redundância e alta disponibilidade.

2. Shard Servers

- Função: Cada servidor de shard armazena uma parte dos dados do banco de dados. Em um cluster shardado, os dados são divididos horizontalmente em fragmentos que são distribuídos entre vários servidores de shards.
- Importância: O sharding permite que o banco de dados distribua a carga de dados e as consultas entre vários servidores, melhorando a capacidade de leitura/escrita e a escalabilidade do sistema.
- Configuração: Cada shard pode ser um servidor standalone ou um conjunto de réplicas. A utilização de conjuntos de réplicas é recomendada para garantir a disponibilidade e a durabilidade dos dados.

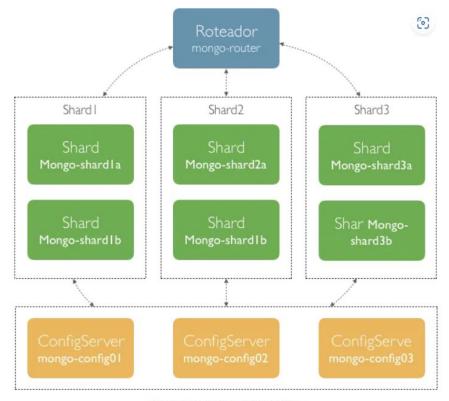
3. Query Routers (mongos)

- Função: O mongos atua como uma interface de consulta entre os clientes do banco de dados e o cluster MongoDB. Ele direciona as operações de consulta para os shards apropriados e agrega os resultados.
- **Importância**: O router de consultas simplifica a complexidade do cliente ao interagir com um cluster distribuído, fazendo parecer que o cliente está lidando com uma única instância de banco de dados.
- Configuração: Em um ambiente de produção, pode haver múltiplos routers de consultas para distribuir e balancear a carga de consultas entre eles.



Tipos de serviços do MongoDb

Vamos criar o cluster com 3 nós replicantes a representação gáfica seria similar a isso:



Arquitetura do cluster a ser criado

Configurando o Docker e MongoDB

1º passo camos criar as instancias e Shared. Os comandos do passo a passoe estão contidos nas imagens.

```
Windows PowerShell X + V - - - X

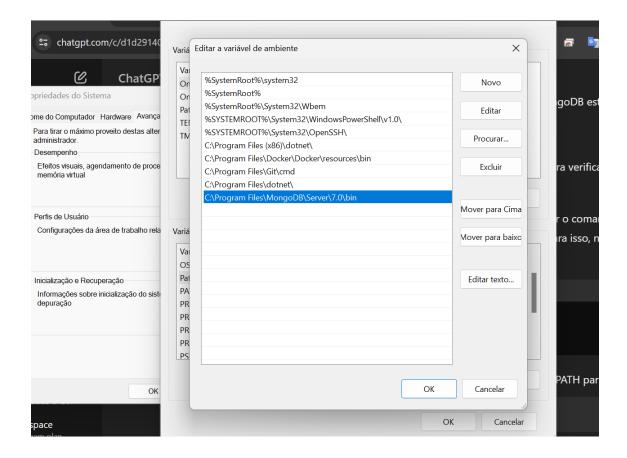
O Windows PowerShell
Copyright (C) Microsoft Corporation. Todos os direitos reservados.

Instale o PowerShell mais recente para obter novos recursos e aprimoramentos! https://aka.ms/PSWindows

PS C:\puc-teste\nosql> docker-compose up -d
time="2024-06-04T20:17:11-03:00" level=warning msg="C:\puc-teste\nosql\\docker-compose.yml: `version` is obsolete"
[i-] Running 0/5
- shard2_1 Pulling
- configsvr1 Pulling
- shard1_1 Pulling
- mongos Pulling
- mongos Pulling
- shard1_2 Pulling
```

Após este processo precisamos obter a ID do container

A depender dos eu ambiente de desenvolvimento talvez seja necessário configurar a variável de ambiente para usar o shell do mongo para reconhecimento de algumas palavras reservadas e variáveis



Programa em Pyhton usa a instancia do Docker mongos.

Precisamos também instalar o cliente na maquina Docker. Mesmo que já tenha instalado em seu computador. Instalar dentro da maquina Docker nos trará versatilidade para manipular a instancia do banco de dados dentro do próprio Docker

```
root@b0dd4ed61ab7:/# apt-get update
apt-get install -y mongodb-clients
Ign:1 http://repo.mongodb.org/apt/ubuntu jammy/mongodb-org/7.0 InRelease
Get:2 http://security.ubuntu.com/ubuntu jammy-security InRelease [129 kB]
Get:3 http://repo.mongodb.org/apt/ubuntu jammy/mongodb-org/7.0 Release [2090 B]
Get:4 http://security.ubuntu.com/ubuntu jammy-security/universe amd64 Packages [1085 kB]
Get:5 http://repo.mongodb.org/apt/ubuntu jammy/mongodb-org/7.0 Release.gpg [866 B]
Get:6 http://archive.ubuntu.com/ubuntu jammy InRelease [270 kB]
Get:7 http://repo.mongodb.org/apt/ubuntu jammy/mongodb-org/7.0/multiverse amd64 Packages [40.9 kB]
Get:8 http://security.ubuntu.com/ubuntu jammy-security/multiverse amd64 Packages [44.7 kB]
Get:9 http://security.ubuntu.com/ubuntu jammy-security/main amd64 Packages [1858 kB]
Get:10 http://archive.ubuntu.com/ubuntu jammy-security/restricted amd64 Packages [2395 kB]
Get:11 http://archive.ubuntu.com/ubuntu jammy-backports InRelease [127 kB]
Get:13 http://archive.ubuntu.com/ubuntu jammy-backports InRelease [17.5 MB]
36% [13 Packages 2240 kB/17.5 MB 13%]
```

Agora podemos testar as conexões das instancias/ shareds etc..

PS C:\puc-teste\nosql> Test-NetConnection -ComputerName localhost -Port 27019 ComputerName : localhost RemoteAddress : ::1 : 27019 RemotePort InterfaceAlias : Loopback Pseudo-Interface 1 SourceAddress TcpTestSucceeded : True PS C:\puc-teste\nosql> Test-NetConnection -ComputerName localhost -Port 27018 ComputerName : localhost RemoteAddress RemotePort : 27018 InterfaceAlias : Loopback Pseudo-Interface 1 SourceAddress TcpTestSucceeded : True

PS C:\puc-teste\nosql> Test-NetConnection -ComputerName localhost -Port 27021

ComputerName : localhost
RemoteAddress : ::1
RemotePort : 27021
InterfaceAlias : Loopback Pseudo-Interface 1
SourceAddress : ::1
TcpTestSucceeded : True

PS C:\puc-teste\nosql> Test-NetConnection -ComputerName localhost -Port 27017

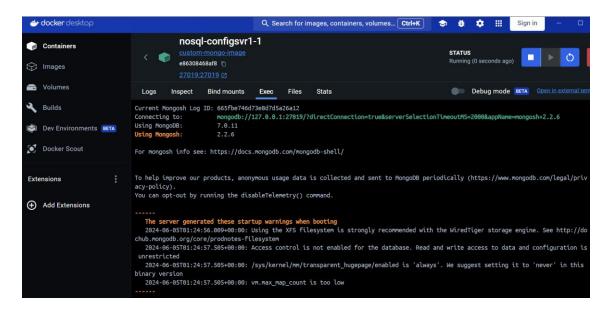
ComputerName : localhost
RemoteAddress : ::1
RemotePort : 27017
InterfaceAlias : Loopback Pseudo-Interface 1
SourceAddress : ::1
TcpTestSucceeded : True

Caso enfrente problemas pode remover e recriar (Esta etapa da próxima imagem é Opcional)

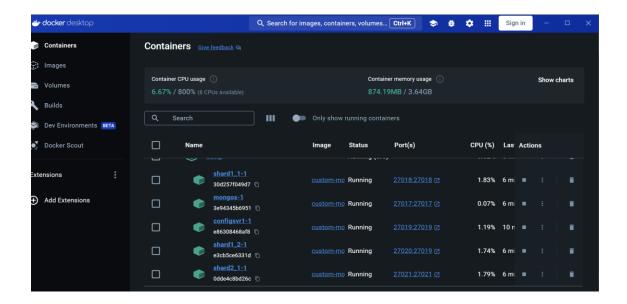
Uma outra alternativa é recriar os containeres fora do Docker desktop (Também é uma etapa opcional ou uma alternativa para configuração do ambiente)

```
| Start a build | Start a bui
```

Instalamos o mongosh e podemos manipular os scripts do Mongo através dele diretamente na interface Docker.



Pronto. Agora temos todas as instancias rodando



No inicio deste documento explica o papel de cada instância destas da imagem acima.

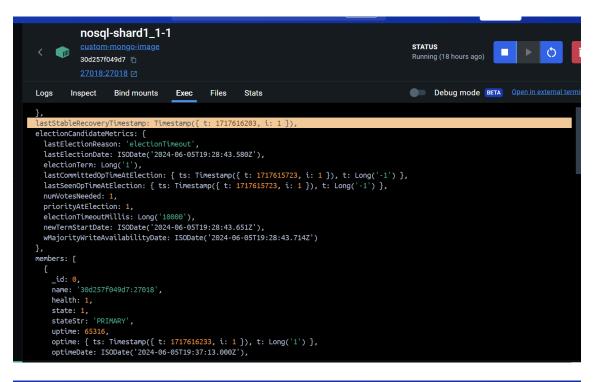
O mongos atua como uma interface de consulta que interage com os clientes externos. Ele direciona as operações para os shards apropriados e agrega os resultados das consultas. Para adicionar coleções, criar índices, distribuir dados, e realizar consultas em um ambiente sharded, você interage através do mongos, que gerencia a complexidade do cluster de shards nos bastidores.

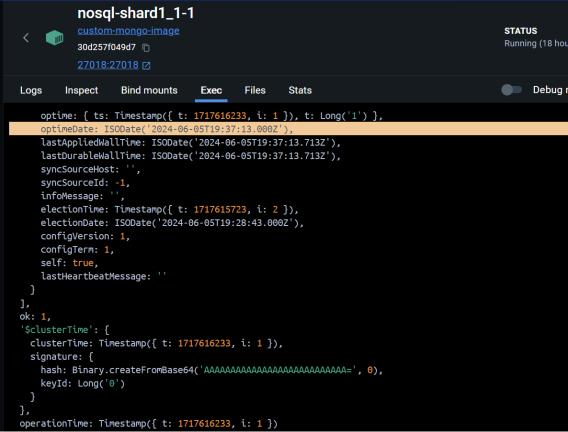
E finalmente neste momento vamos configurar a ligação dos nós.



Uma maneira de verificar se os shareds estão funcionando corretamente é através deste comando. Também é importante que eles estejam definidos como PRIMARY

```
nosql-shard1_1-1
                                                                                                             STATUS
                                                                                                             Running
              30d257f049d7 🗇
              27018:27018
 Logs
           Inspect
                       Bind mounts
                                         Exec
                                                   Files
                                                             Stats
s.status()
 set: 'rs-shard-1',
 date: ISODate('2024-06-05T19:37:17.380Z'),
 myState: 1,
 term: Long('1'),
 syncSourceHost:
 syncSourceId: -1,
heartbeatIntervalMillis: Long('2000'),
majorityVoteCount: 1,
 writeMajorityCount: 1,
 votingMembersCount: 1,
 writableVotingMembersCount: 1,
 optimes: {
   lastCommittedOpTime: { ts: Timestamp({ t: 1717616233, i: 1 }), t: Long('1') },
   lastCommittedWallTime: ISODate('2024-06-05T19:37:13.713Z'),
   readConcernMajorityOpTime: { ts: Timestamp({ t: 1717616233, i: 1 }), t: Long('1') }, appliedOpTime: { ts: Timestamp({ t: 1717616233, i: 1 }), t: Long('1') }, durableOpTime: { ts: Timestamp({ t: 1717616233, i: 1 }), t: Long('1') },
   lastAppliedWallTime: ISODate('2024-06-05T19:37:13.713Z'),
   lastDurableWallTime: ISODate('2024-06-05T19:37:13.713Z')
 lastStableRecoveryTimestamp: Timestamp({ t: 1717616203, i: 1 }),
 electionCandidateMetrics: {
```





Pronto. Finalizada esta etapa agora podemos ir para o nosso programa em python. O arquivo inicial de configuração o Docker composse e DockerFile estão no repositório GIT. Abaixo uma inical que foi utilizada para fazer a configuração acima versão do arquivo:

Arquivo DockerCompose:

```
version: '3.8'
services:
 configsvr1:
  image: custom-mongo-image
  command: ["mongod", "--configsvr", "--replSet", "rs-config", "--port",
"27019"]
  ports:
   - 27019:27019
  volumes:
   - ./data/config/configsvr1:/data/db
 shard1 1:
  image: custom-mongo-image
  command: ["mongod", "--shardsvr", "--replSet", "rs-shard-1", "--port",
"27018"]
  ports:
   - 27018:27018
  volumes:
   - ./data/shard1 1:/data/db
 shard1 2:
  image: custom-mongo-image
  command: ["mongod", "--shardsvr", "--replSet", "rs-shard-1", "--port",
"27019"]
  ports:
   - 27020:27019
```

```
volumes:
   - ./data/shard1_2:/data/db
 shard2 1:
  image: custom-mongo-image
  command: ["mongod", "--shardsvr", "--replSet", "rs-shard-2", "--port",
"27021"]
  ports:
   - 27021:27021
  volumes:
   - ./data/shard2_1:/data/db
 mongos:
  image: custom-mongo-image
  command: ["mongos", "--configdb", "rs-config/configsvr1:27019", "--
port", "27017"]
  ports:
   - 27017:27017
Arquivo DockerFile:
FROM mongo:latest
RUN apt-get update && apt-get install -y wget gnupg
# Adicionar a chave pública do MongoDB
```

RUN wget -qO - https://www.mongodb.org/static/pgp/server-4.4.asc | apt-key add -

Adicionar o repositório do MongoDB à lista de fontes

RUN echo "deb [arch=amd64,arm64] https://repo.mongodb.org/apt/ubuntu focal/mongodb-org/4.4 multiverse" | tee /etc/apt/sources.list.d/mongodb-org-4.4.list

Atualizar a lista de pacotes e instalar o cliente mongo RUN apt-get update && apt-get install -y mongodb-org-shell

Aplicação em Python

Arquivo de conexão com banco de dados:

Aplicação onde tem as operações de CRUD da loja:

```
Current File ∨ ▷ 🌣
□ Project ∨
                                 db_connection.py
                                                                                            manage_inventory.py ×

∨ □ puc-mongo-final C:\Users\v

1

                                    from pymongo.collection import Collection
       > 🗀 .idea
80
                                     db = get_database()
         db_connection.py
                               6 def create_branch(branch_name): 1usage new*
7 db.create_collection(branch_name)
        🕹 Dockerfile
                                9 def add_product(branch_name, sku, name, price, stock): 1usage new*
     > 
Scratches and Consoles
                                             "name": name,
                                         }
                                         collection.insert_one(product)
寥
                                     def query_product(branch_name, sku): 1 usage new *
Ð
                                         return collection.find_one({"sku": sku})
   def update_stock(branch_name, sku, new_stock): 1usage new*
       collection.update_one( filter: {"sku": sku}, update: {"$set": {"stock": new_stock}})
   # as chaamdas via aplicação seriam desta forma em um ambiente onde pode executar uma grande quantidade de chamadas
```

add_product(branch_name: "Filial_1", sku: "1234", name: "Teclado Mecânico", price: 150.00, stock: 30)
print(query_product(branch_name: "Filial_1", sku: "1234"))
update_stock(branch_name: "Filial_1", sku: "1234", new_stock: 25)

create_branch("Filial_1")