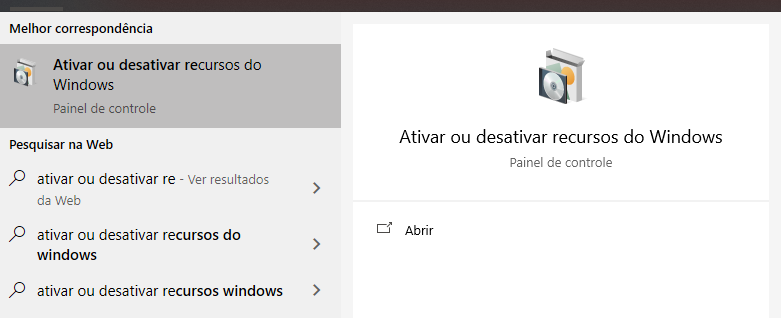
**Instalando Docker**

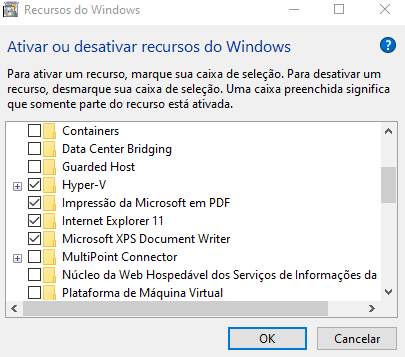
**Requisitos**

Hyper-V ativo no windows.

* No menu iniciar procurer “Ativar ou desativar recursos do windows”



* Marcar HyperV e clicar em OK



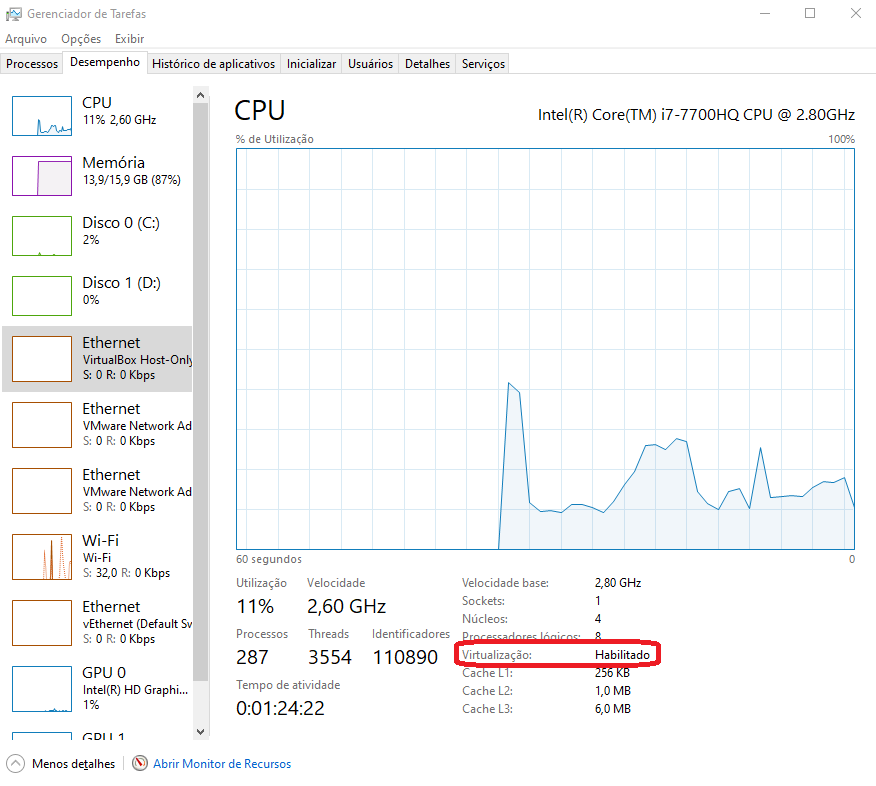
* O windows vai fazer a instalação e irar pedir para reiniciar, confirme a reinicialização.

Processador 64 bit

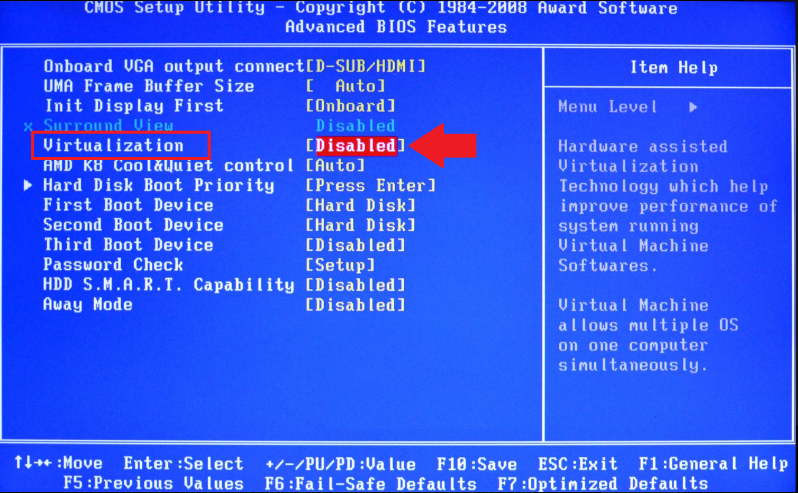
4GB RAM

Virtualização ativada na BIOS

* Para verificar se está ativo acesso o gerenciador de tarefas na aba de desempenho



* Caso esteja desativado, reinicie sua máquina e acesse a bios, geralmente pela Tecla “delete”
* Procure a opção de virtualização, ative, salve e reinicie a máquina



* Esse passo depende muito da placa mãe, caso não consiga encontrar, descubra sua placa mãe e procure na internet o passo a passo específico para ela

**Download**

Windows 10 64-bit: Pro, Enterprise, Education (Build 16299 or later).

<https://download.docker.com/win/stable/Docker%20Desktop%20Installer.exe>

Outros SOs ( Windows 7, Windows 10 Home)

<https://github.com/docker/toolbox/releases/download/v19.03.1/DockerToolbox-19.03.1.exe>

**Instalação**

Seguir passo a passo

**Instalando Kitematic**

Necessário apenas para quem baixou a versão diferente do windows 10 Pro, ele é uma interface gráfica para gerenciamento de containers.

**Download**

<https://github.com/docker/kitematic/releases/download/v0.17.12/Kitematic-0.17.12-Windows.zip>

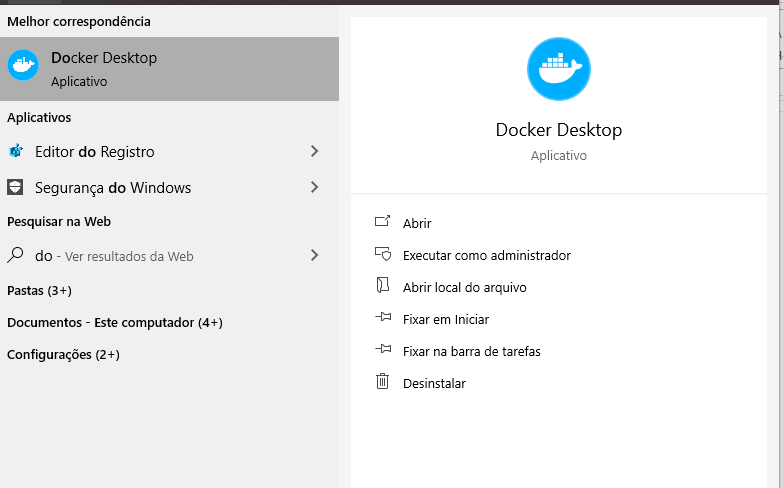
**Instalação**

Extrair em um diretorio do windows

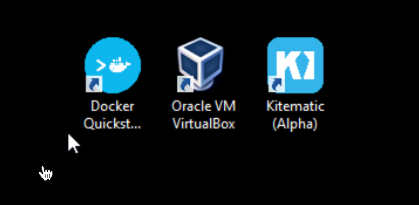
**Iniciando com docker**

1. Inicializar o docker

No Windows 10 pro Docker Desktop

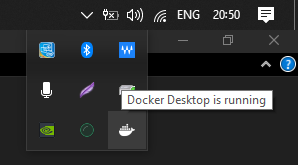


Nos outros SOs Docker Quickstart

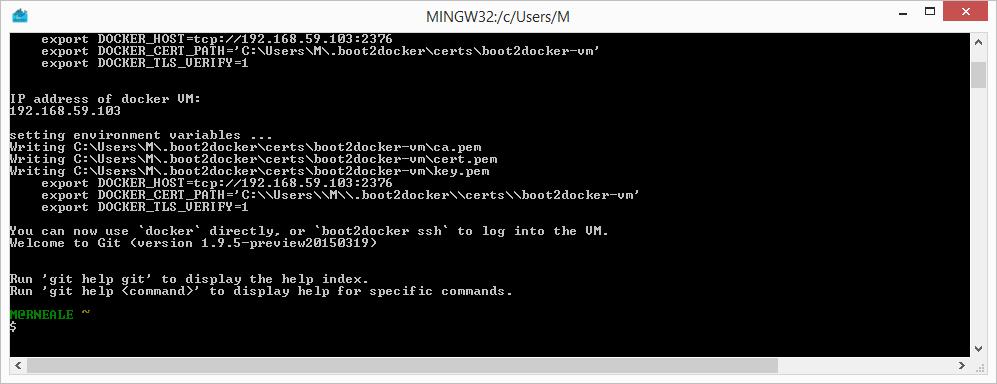


1. Verificar se foi inicializado

No windows 10 Pro é possivel ver na barra do windows

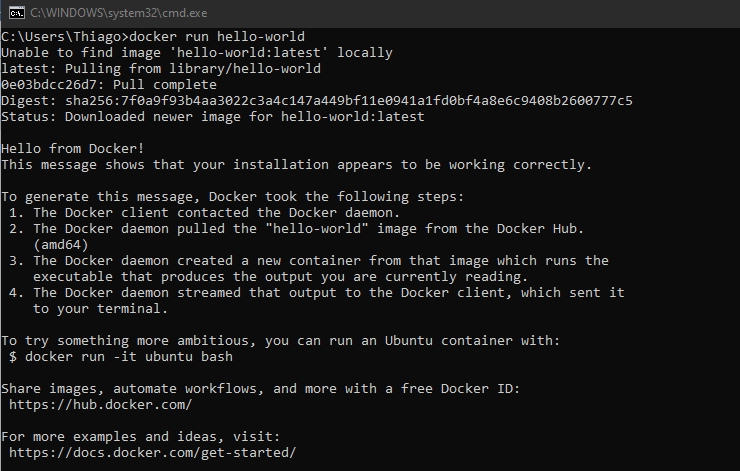


Nos outros SOS abre um terminal

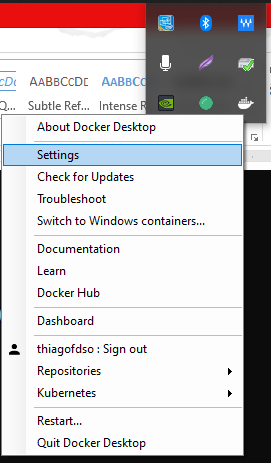
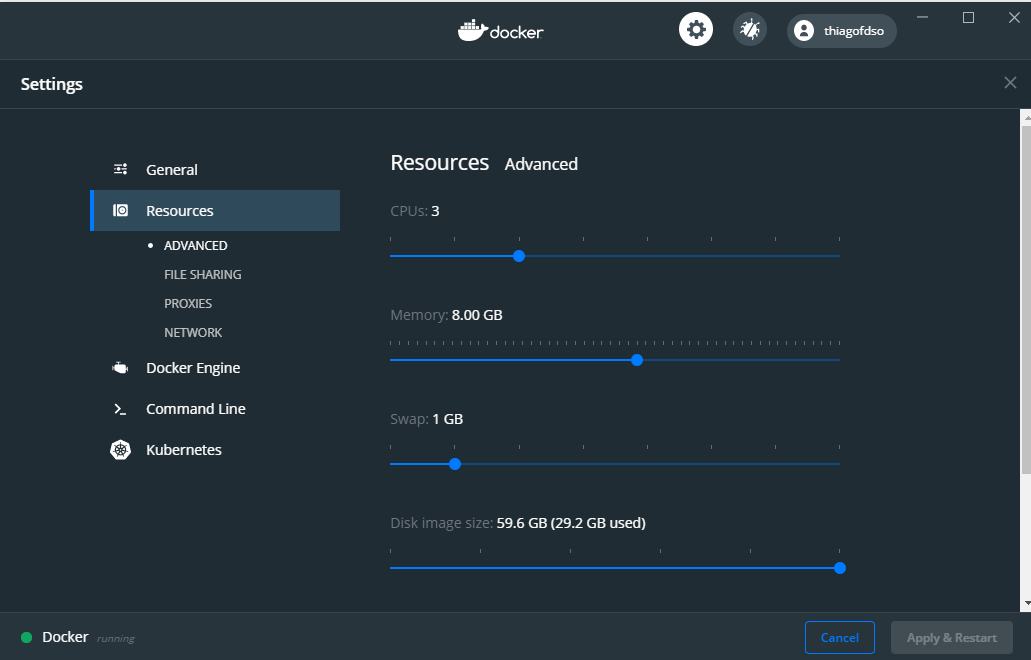


1. Testando um hello-world

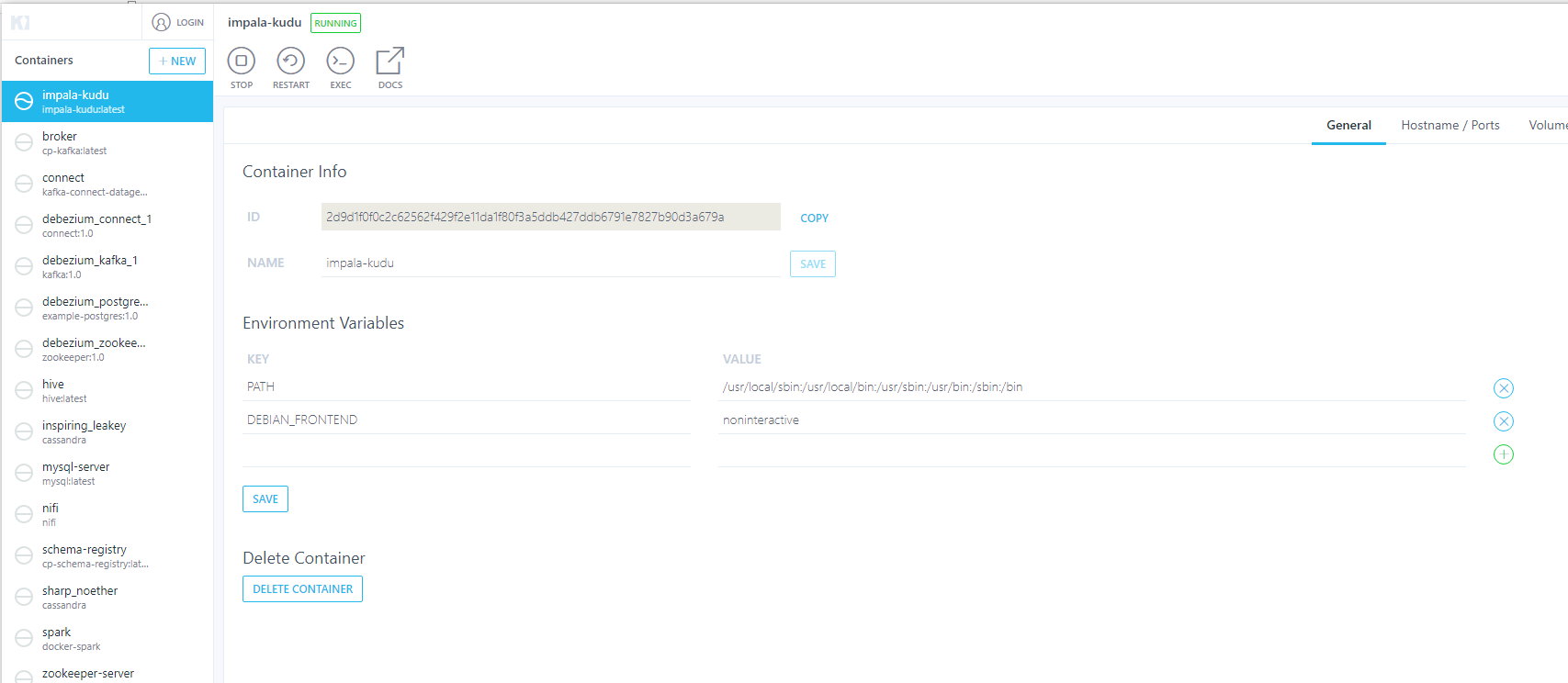
docker run hello-world



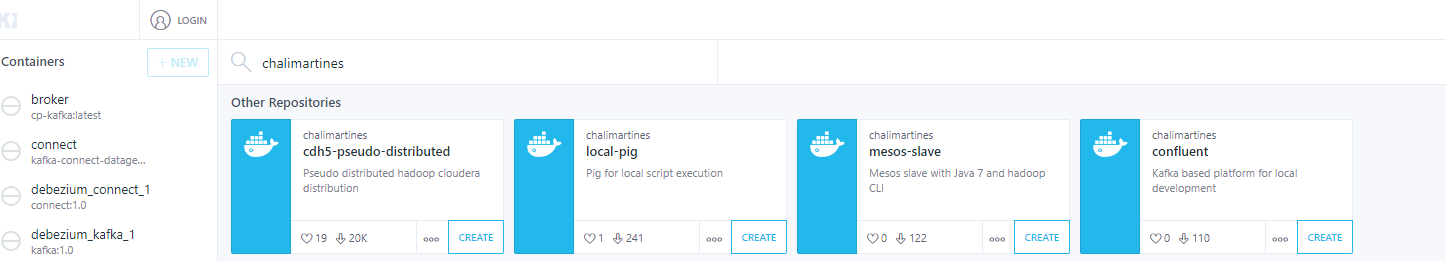
1. Aumentando a quantidade de memória, o ideal é 4GB, mas pode ser aumentado.

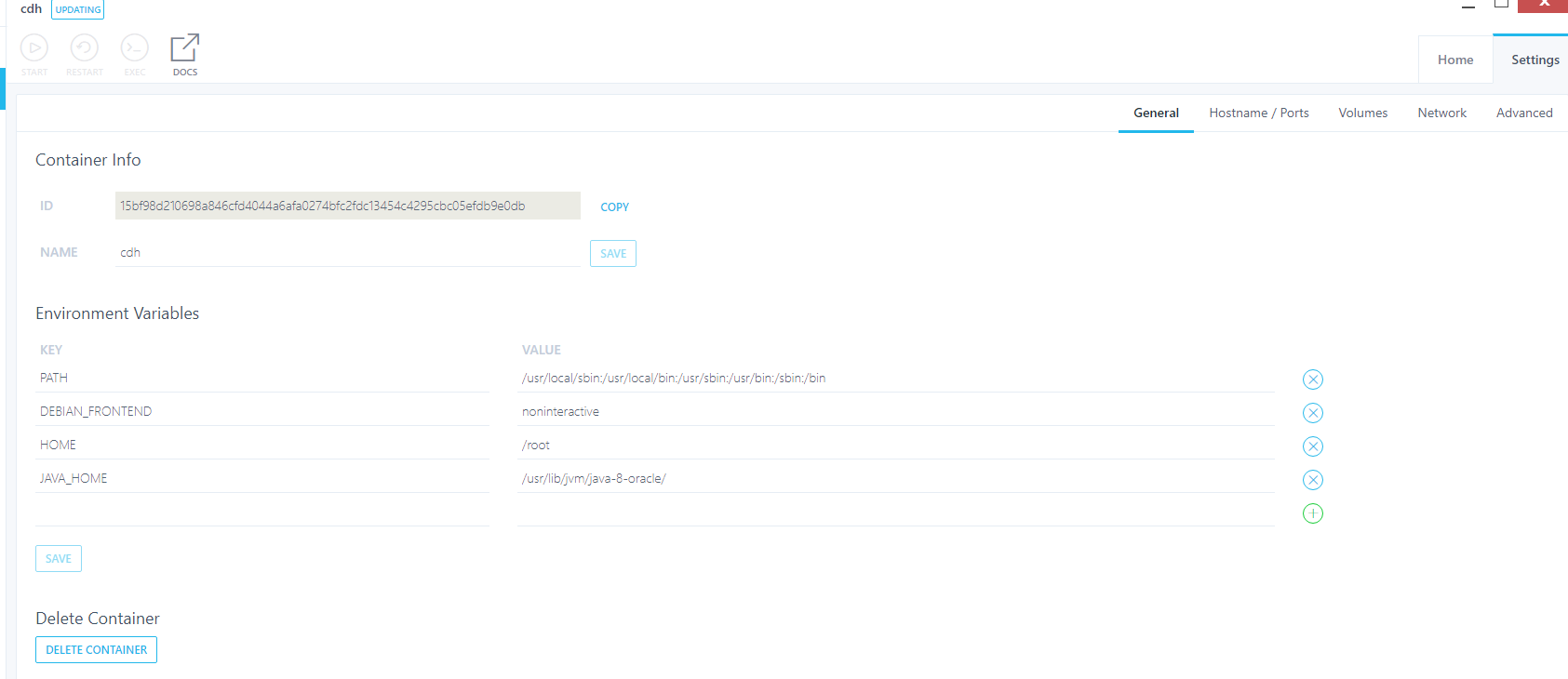
1. Inicie o Kitematic pelo arquivo Kitematic.exe na pasta onde foi extraido.

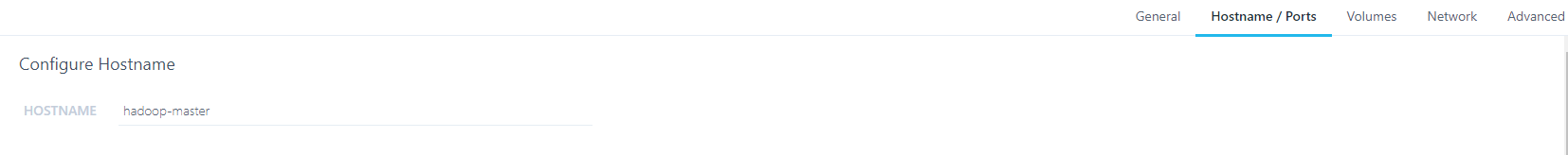


1. Clique em New e procure por prestodev

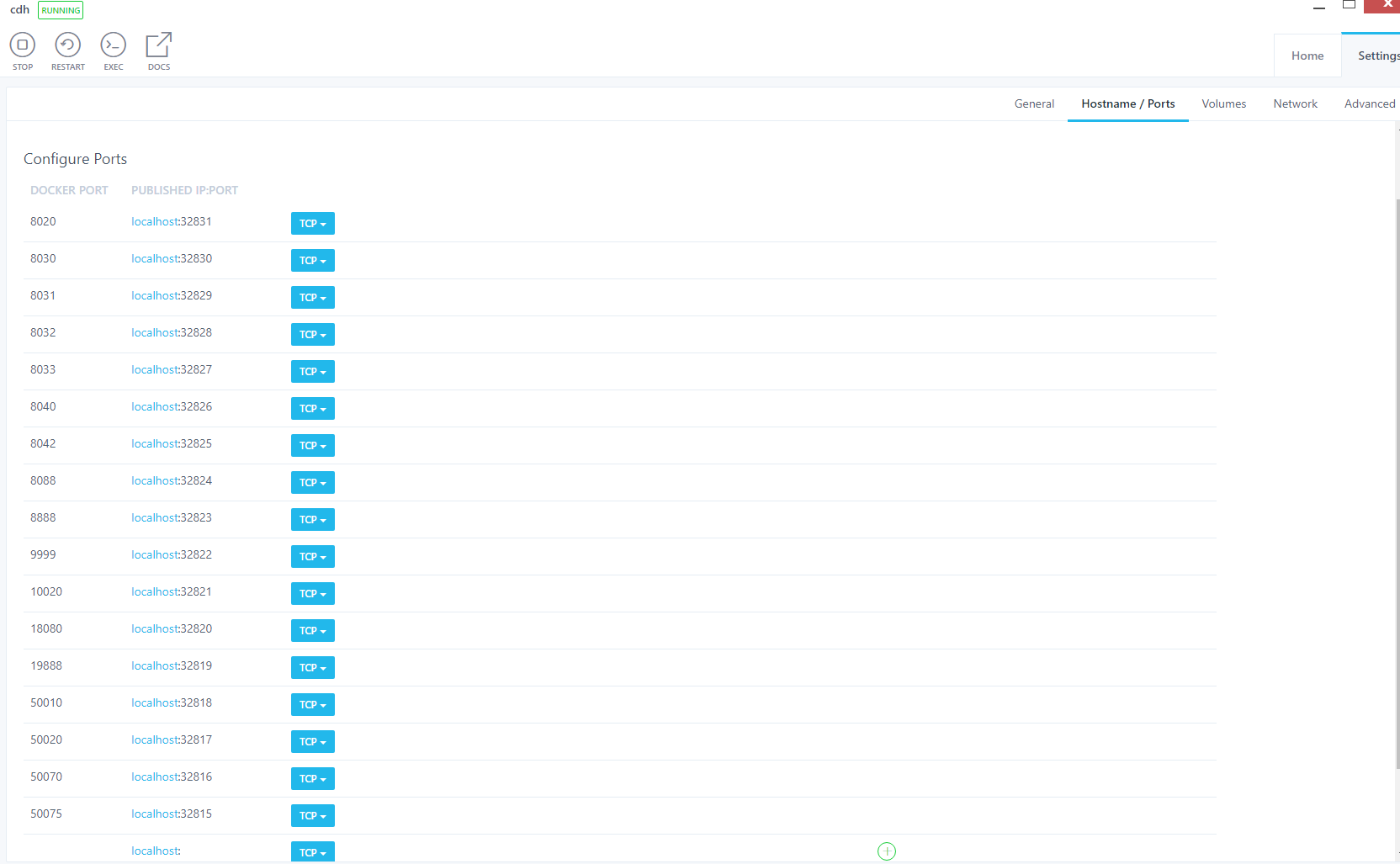


1. Clique em CREATE nan opção cdh5.15-hive
2. O docker irá baixar e inicializar o container, você pode alterar configurações clicando em settings, para efeito de seguir o restante do tutorial altere em General o Name para CDH e em Hostname altere para hadoop-master, salve após cada alteração.

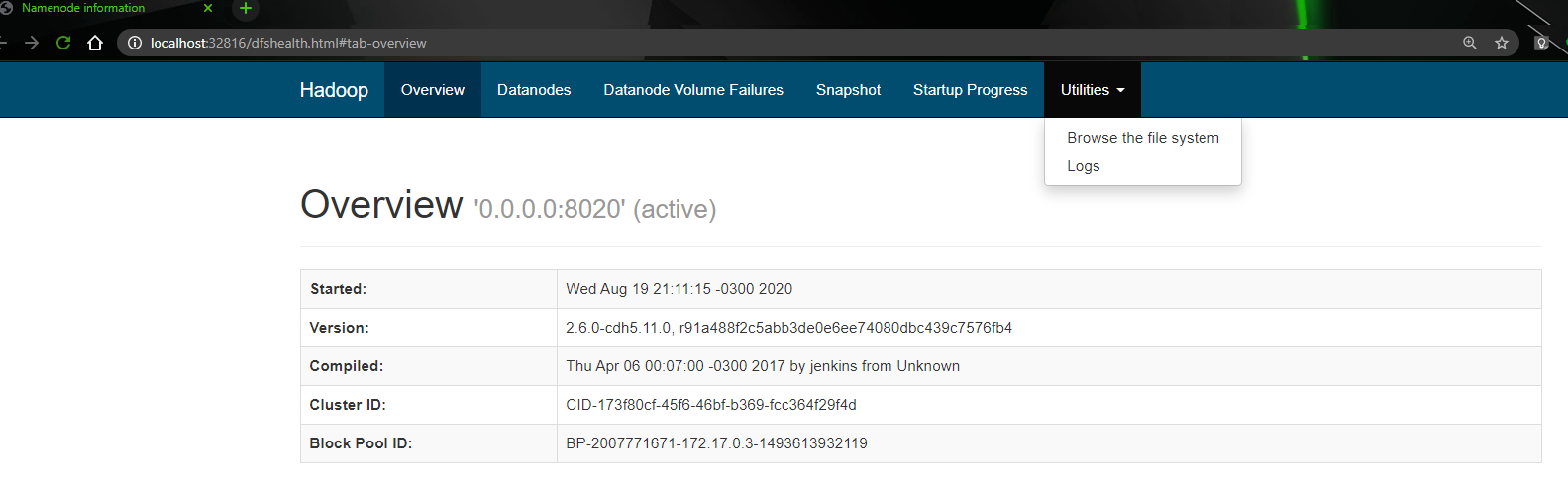




1. Na guia Hostname / Ports é exibido o mapeamento de portas do container para sua máquina, por padrão o Kitematic automaticamente configure portas, deve se ter cuidado ao usar vários containers para não ter conflito.

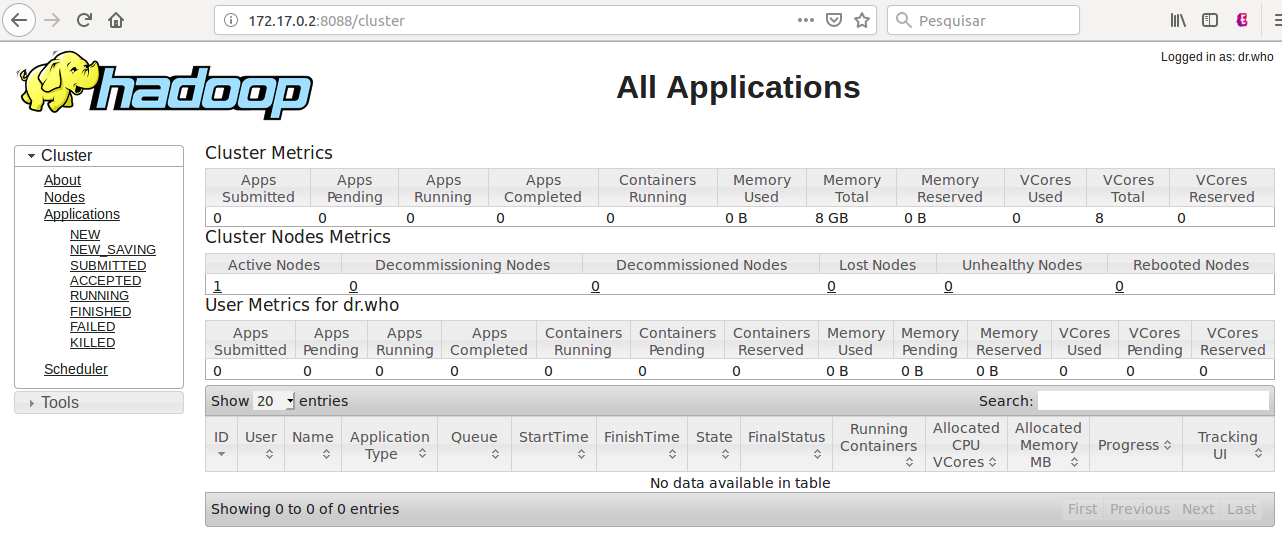


1. Ao clicar na opção localhost:32816 por exemplo abrimos no navegador o serviço referente a porta 50070, o serviço do Hadoop.

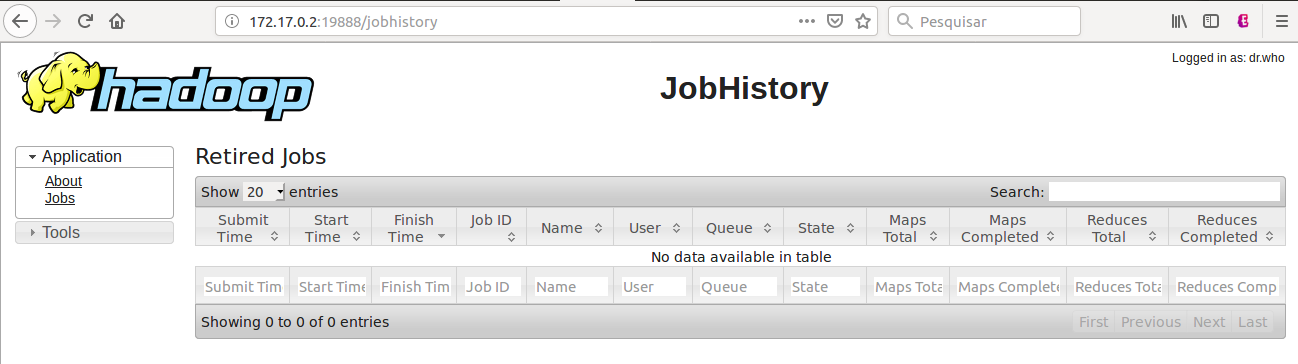


1. Você pode testar os demais serviços da distribuição nos endereços abaixo, usando as portas mapeadas acima, caso ela não esteja você pode adicionar:

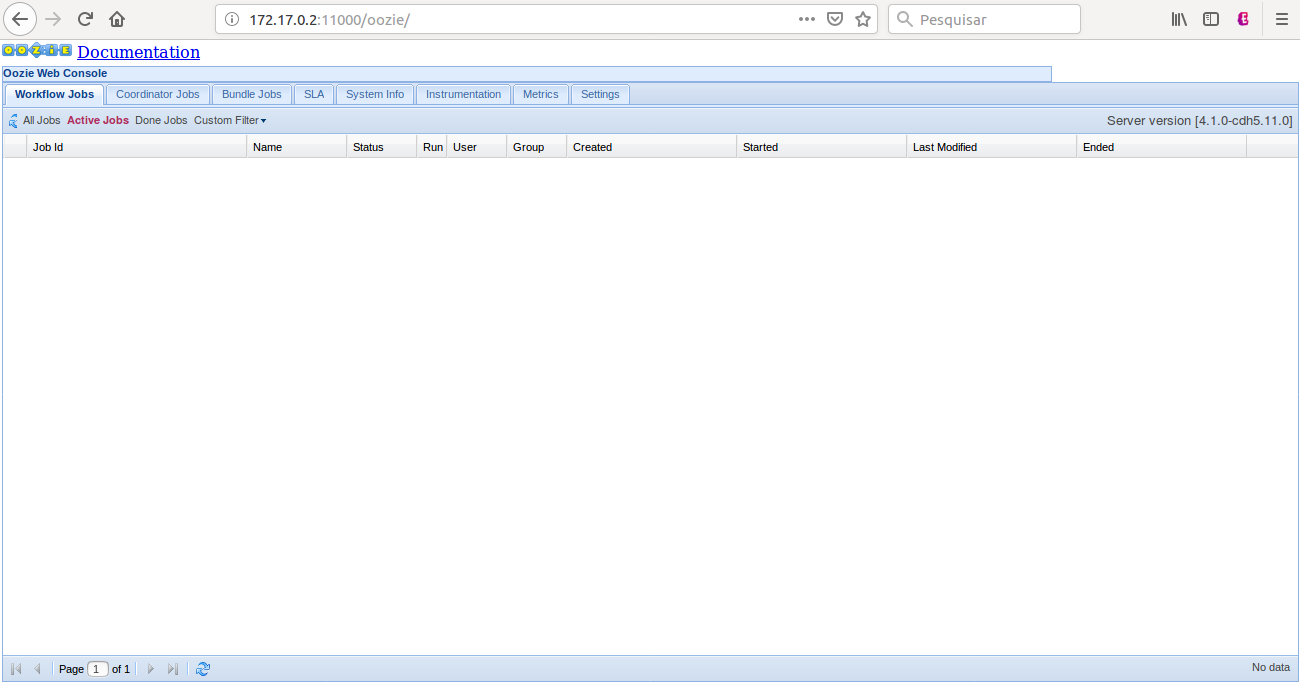
* resource manager: [http://<IP>:8088](http://localhost:8088/) -> [http://<IP>:32824](http://localhost:8088/)



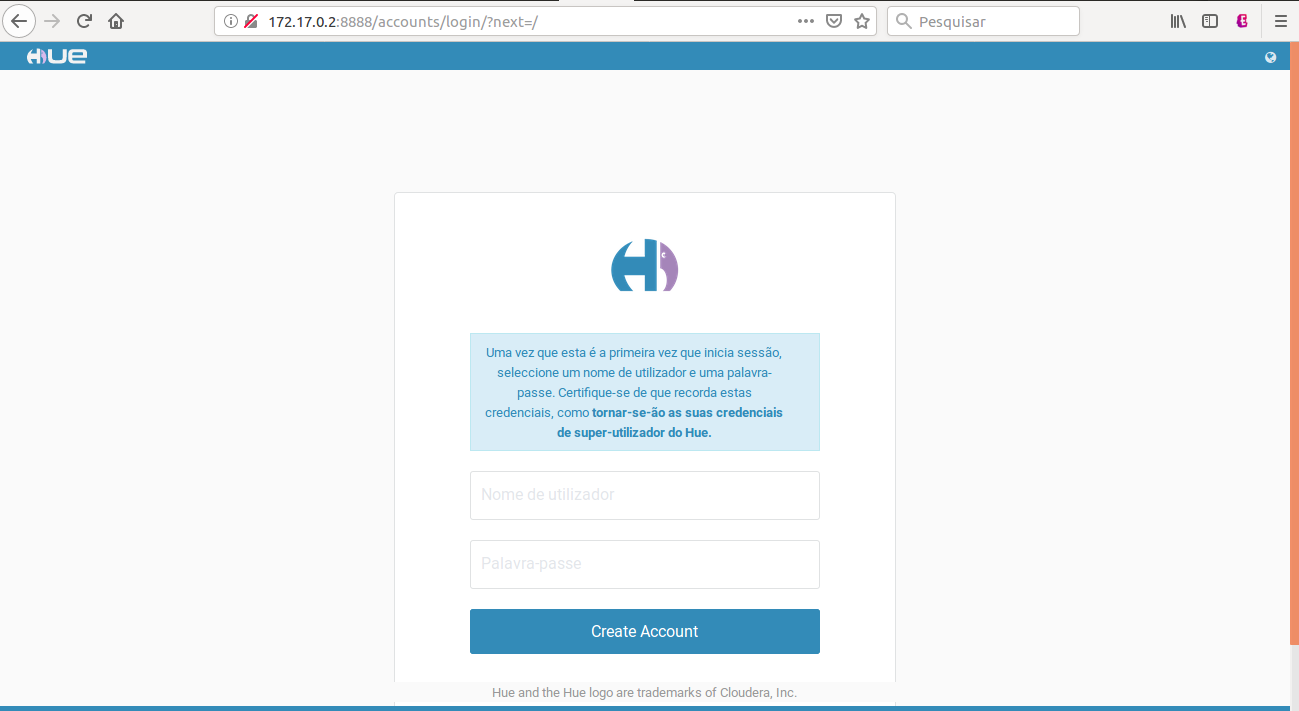
* job history server: [http://<IP>:19888](http://localhost:19888/) -> [http://<IP>:32819](http://localhost:8088/)



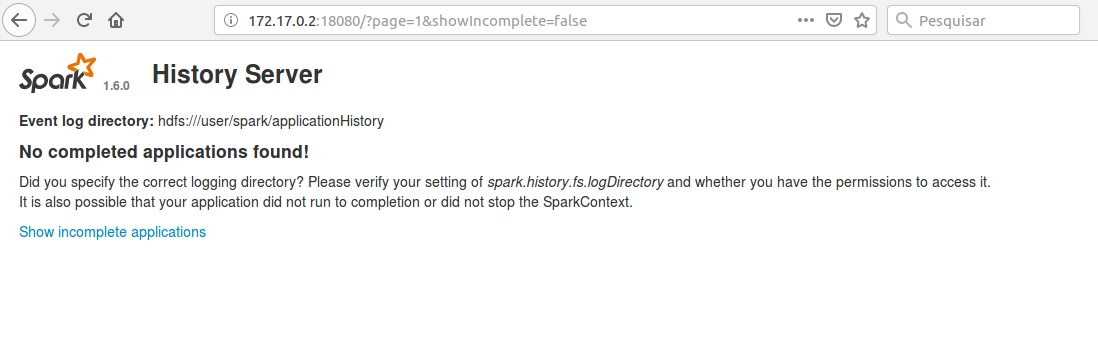
* oozie console: [http://<IP>:11000](http://localhost:11000/)



* hue: http://<IP>:8888 -> [http://<IP>:32823](http://localhost:8088/)



* spark history server: [http://<IP>:18080](http://localhost:18080/) -> [http://<IP>:3282](http://localhost:8088/)0

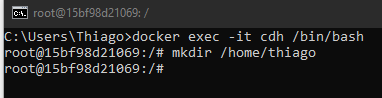


**Trabalho Prático – Hadoop e MapReduce**

1. Abaixo commandos que vai precisar para acessar o container e transferir arquivos, no windows 10 pro pode ser executado de qualquer terminal, nos outros executar na console do quickstart

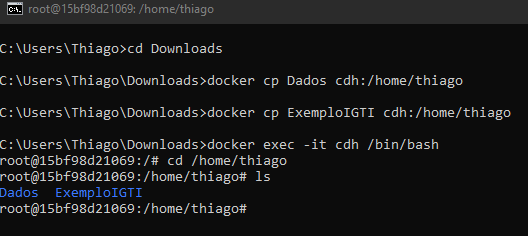
* Acessar o shell do container e executar comandos no container

docker exec -it <nomecontainer> /bin/bash

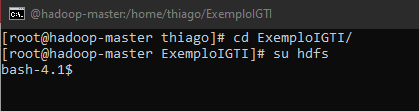


* Copiar arquivos para o container

docker cp <arquivo ou diretorio de origem> <nome container>:<caminho de destino>

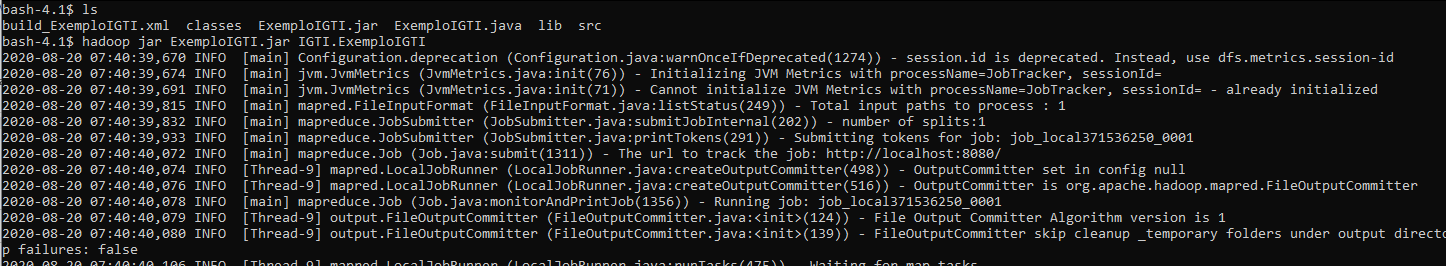


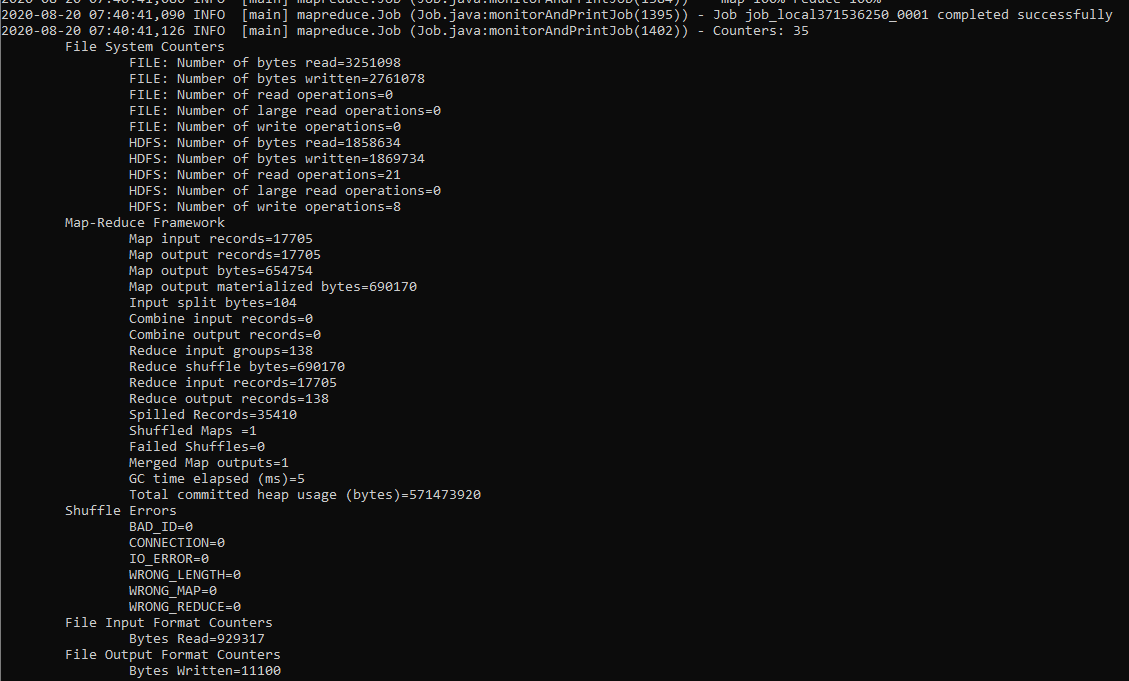
1. Nesse container específico o usuario root não possui permissao no hdfs, use o comando abaixo para mudar para o usuario hdfs.  
   su hdfs



1. Essa maquina nao possui o apt-get entao precisa ser enviado o jar diretamente para ela. No caso está na pasta ExemploIGTI que irei compartilhar. Executar o comando:

hadoop jar ExemploIGTI.jar IGTI.ExemploIGTI

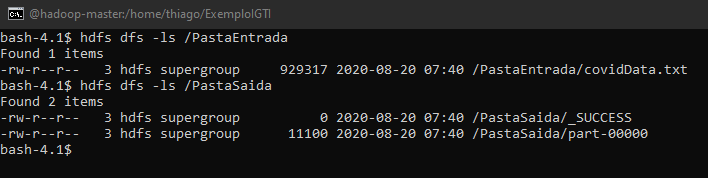




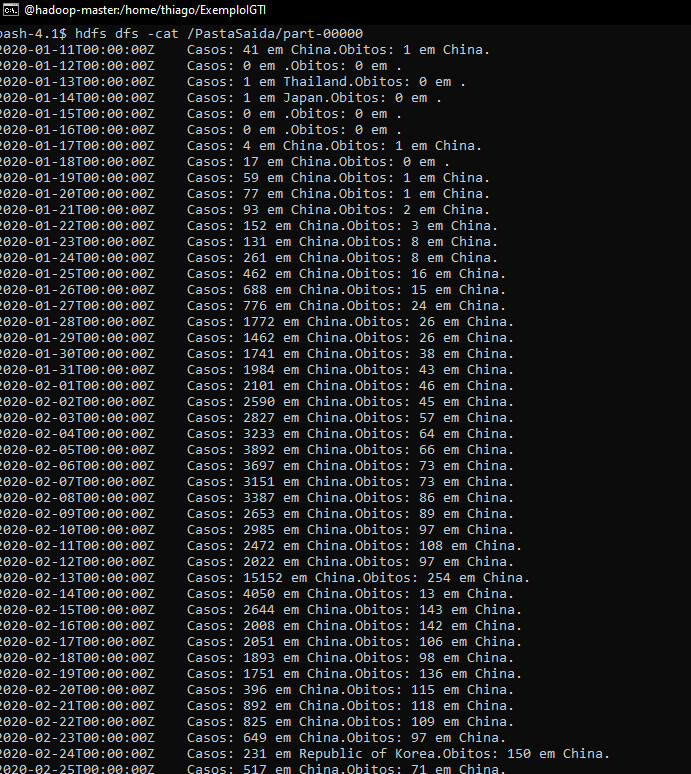
1. Verificando os arquivos no HDFS:

hdfs dfs -ls /PastaEntrada

hdfs dfs -ls /PastaSaida



hdfs dfs -cat /PastaSaida/part-00000



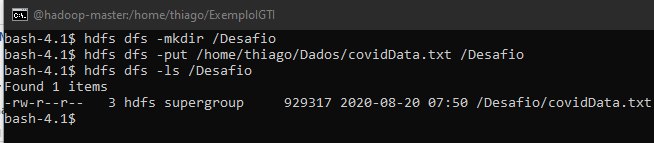
**Desafio - Hive**

1. Criando diretorio e adicionando arquivo

hdfs dfs -mkdir /Desafio

hdfs dfs -put /home/thiago/Dados/covidData.txt

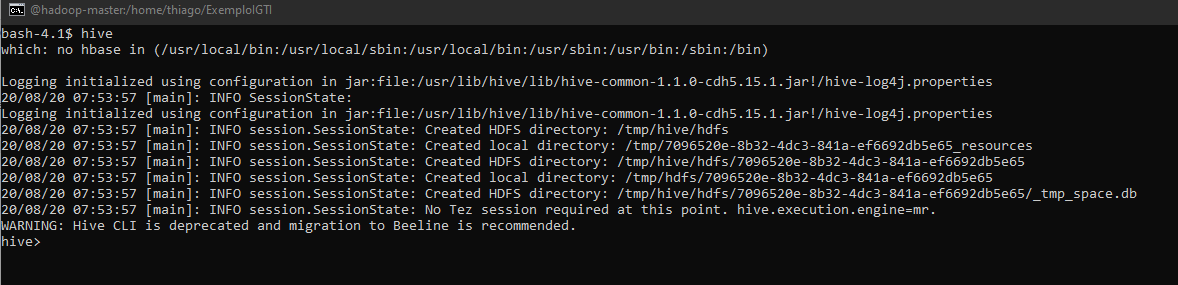
hdfs dfs -ls /Desafio



1. Execute o hive

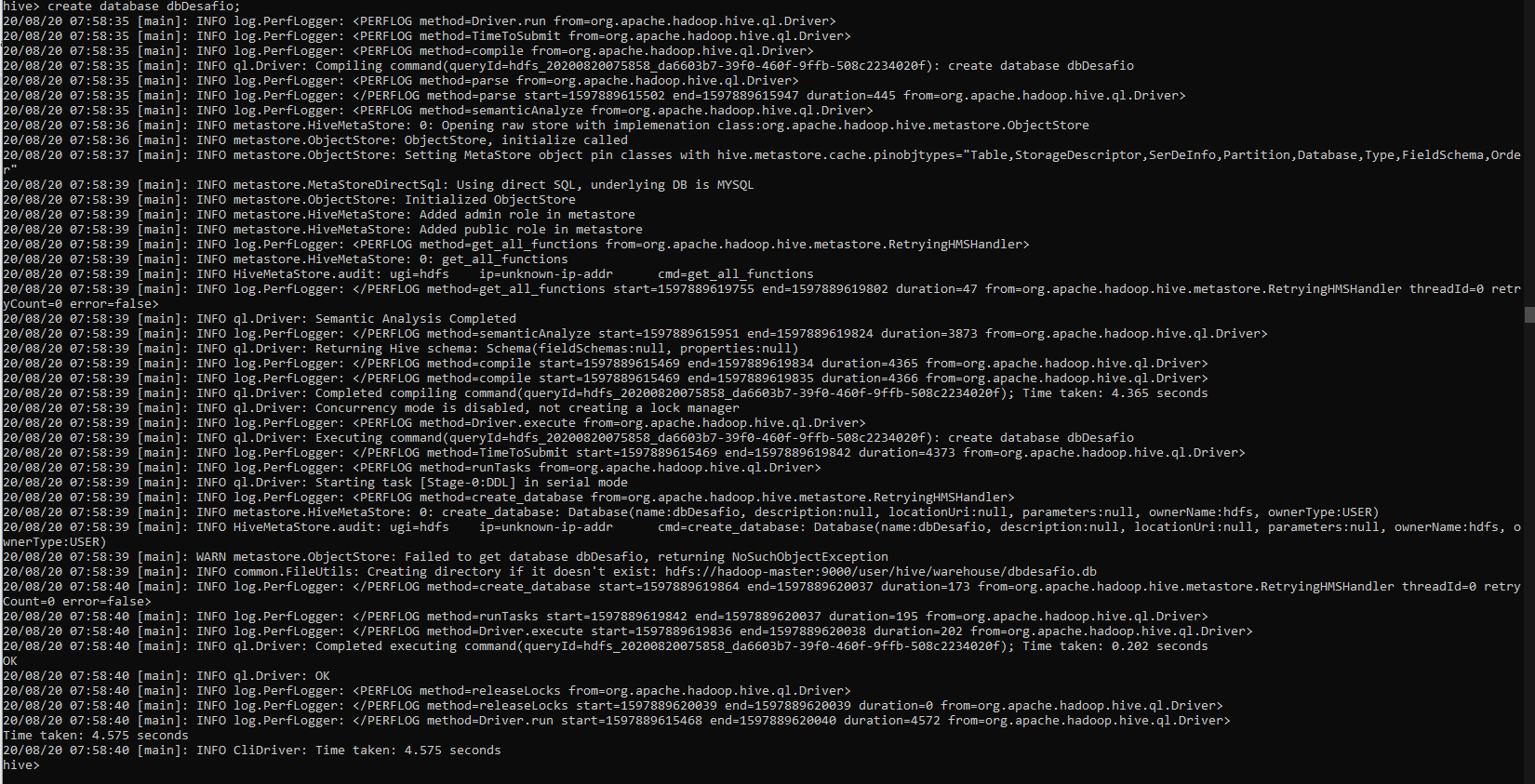
Use o comando abaixo para reduzir o numero de logs

hive --hiveconf hive.root.logger=WARN,console



1. Criar um *database* chamado dbDesafio, por meio do comando create database. Crie uma tabela chamada DadosCovid.

**create database dbDesafio;**

  
**use dbDesafio;**

**CREATE TABLE DadosCovid(**

**dataOcorrencia String,**

**siglaPais String,**

**descPais String,**

**regiao String,**

**novosCasos int,**

**casosAcumulados int,**

**novosObitos int,**

**obitosAcumulados int)**

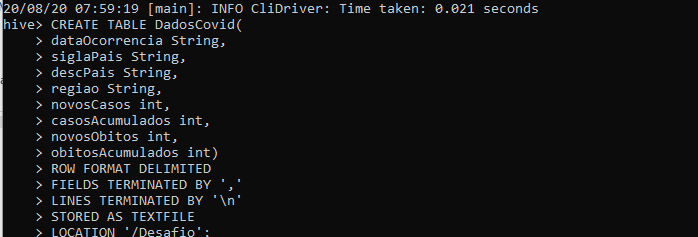
**ROW FORMAT DELIMITED**

**FIELDS TERMINATED BY ','**

**LINES TERMINATED BY '\n'**

**STORED AS TEXTFILE**

**LOCATION '/Desafio';**

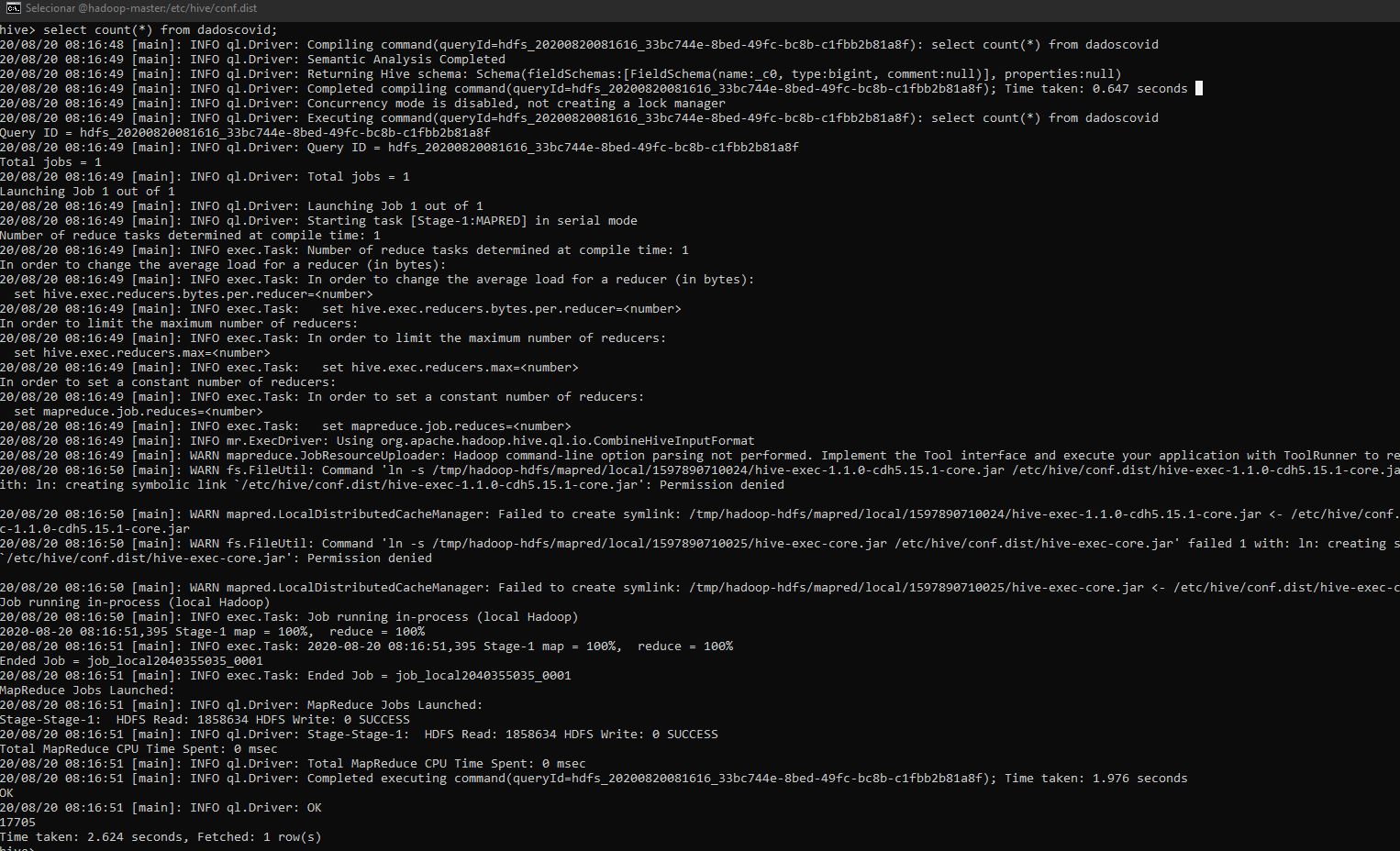


1. Faça a importação dos dados que estão no HDFS para a nova tabela, usando o comando LOAD DATA INPATH.

**LOAD DATA INPATH '/Desafio/covidData.txt' INTO TABLE DadosCovid;**

1. Execute uma sentença SQL que conte todos os registros da tabela DadosCovid.

**select count(\*) from dadoscovid;**

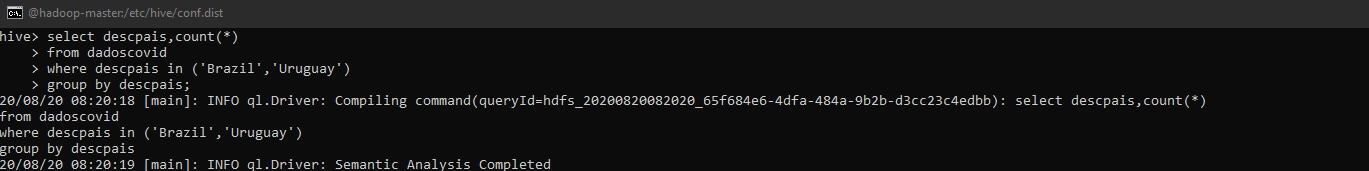


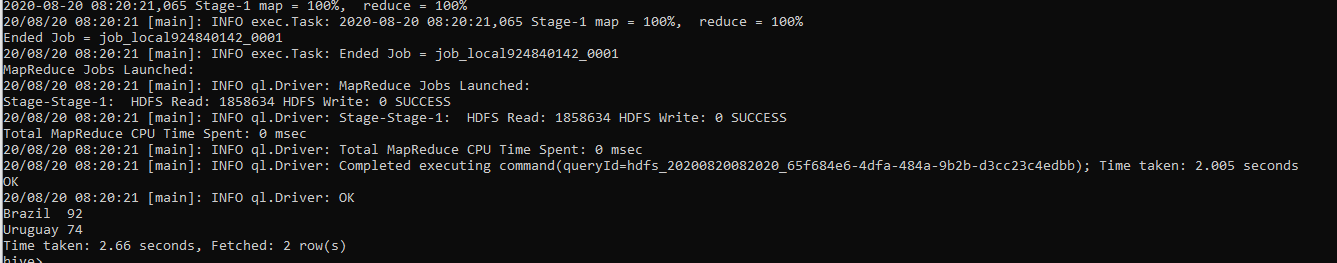
1. Execute uma sentença SQL que verifique quantas comunicações os países Uruguay e Brazil fizeram **cada um** durante o período de apuração do arquivo **select descpais,count(\*)**

**from dadoscovid**

**where descpais in ('Brazil','Uruguay')**

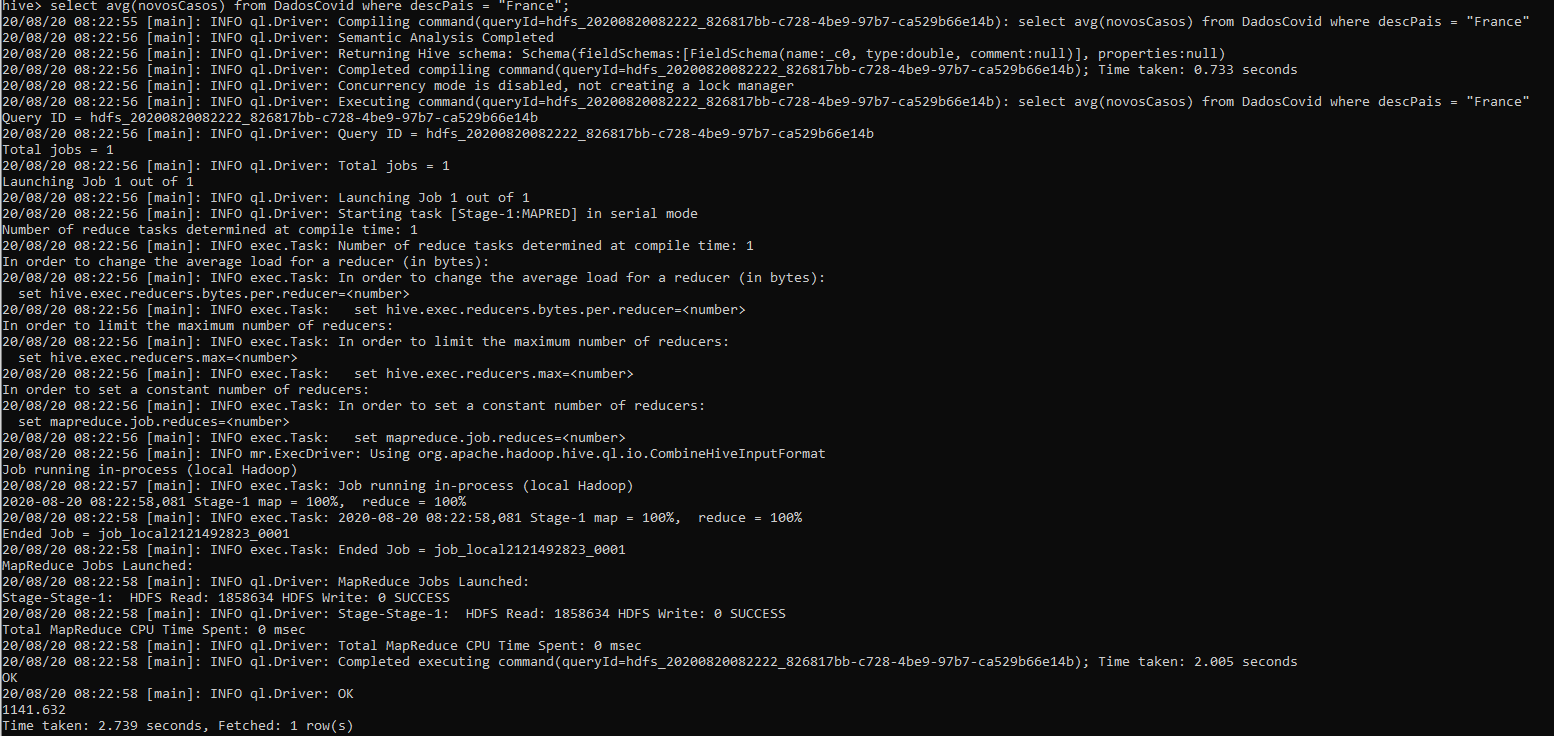
**group by descpais;**





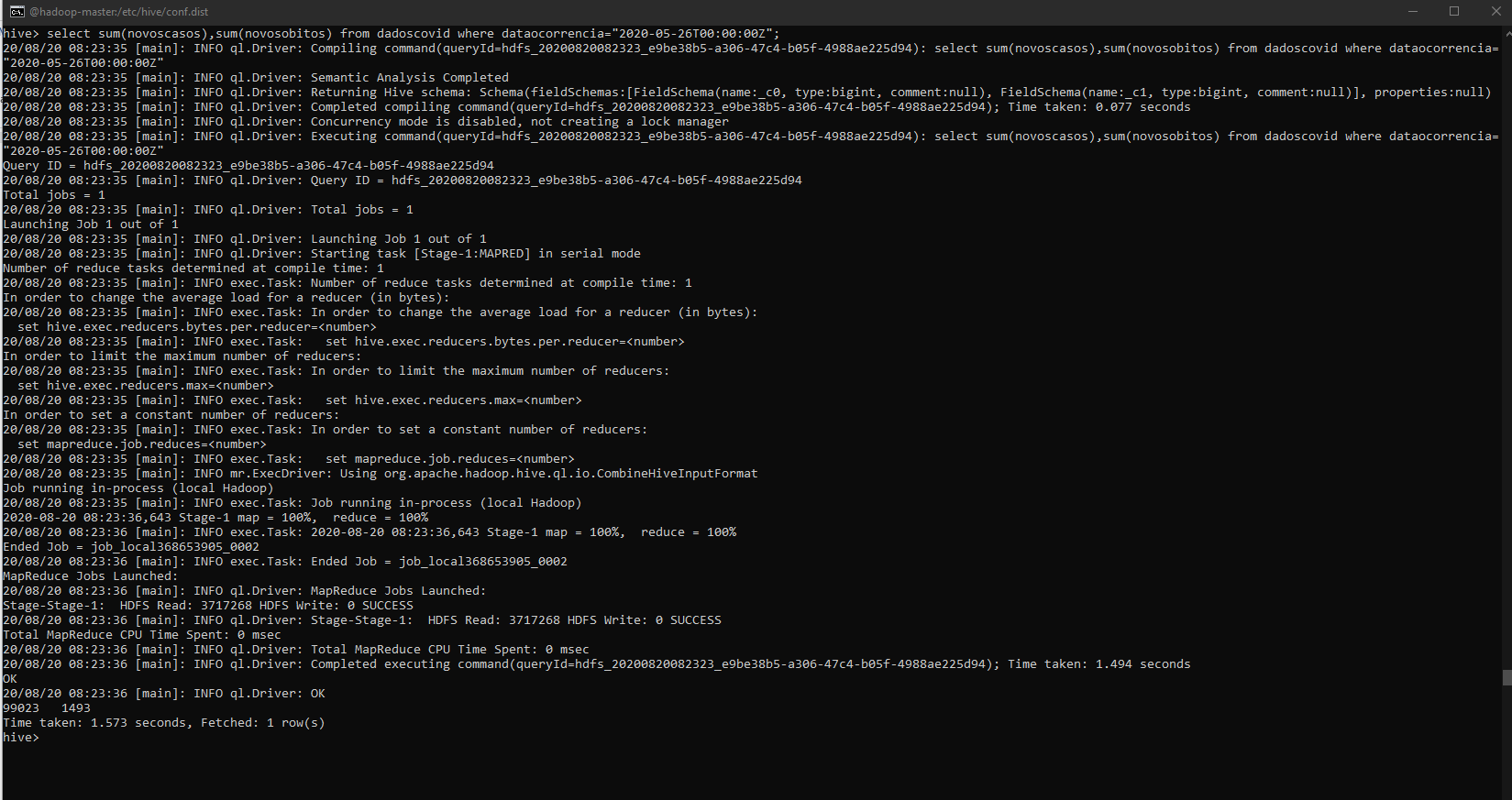
10. Execute a seguinte sentença: select avg(novosCasos) from DadosCovid where descPais = “France”;

**select avg(novosCasos) from DadosCovid where descPais = "France";**



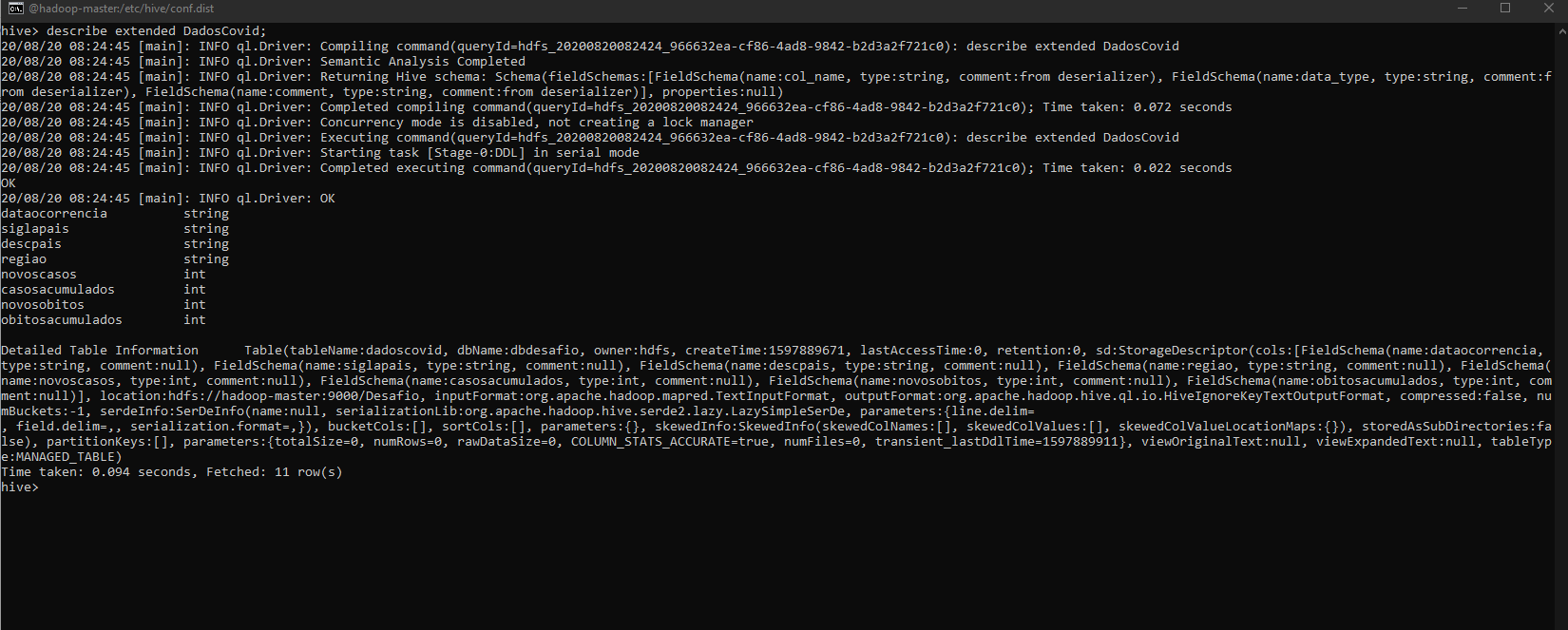
11. Execute uma sentença que apure quantos novos casos e quantos novos óbitos foram comunicadas no dia 26/05/2020, considerando todos os países.

**select sum(novoscasos),sum(novosobitos) from dadoscovid where dataocorrencia="2020-05-26T00:00:00Z";**



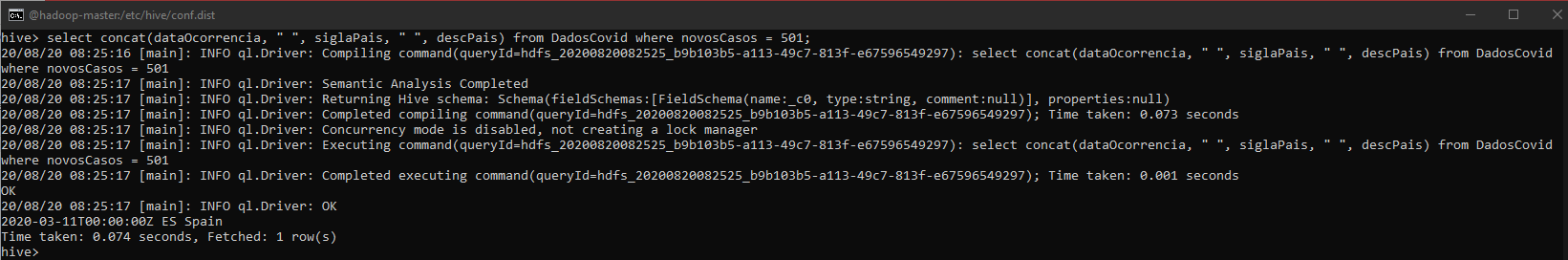
12. Execute o seguinte comando: describe extended DadosCovid

**describe extended DadosCovid;**



13. Execute a sentença: select concat(dataOcorrencia, “ “, siglaPais, “ “, descPais) from DadosCovid where novosCasos = 501;

**select concat(dataOcorrencia, " ", siglaPais, " ", descPais) from DadosCovid where novosCasos = 501;**



14. Execute a sentença: select região, count(1) from DadosCovid group by regiao order by regiao;

**select regiao, count(1) from DadosCovid group by regiao order by regiao;**

