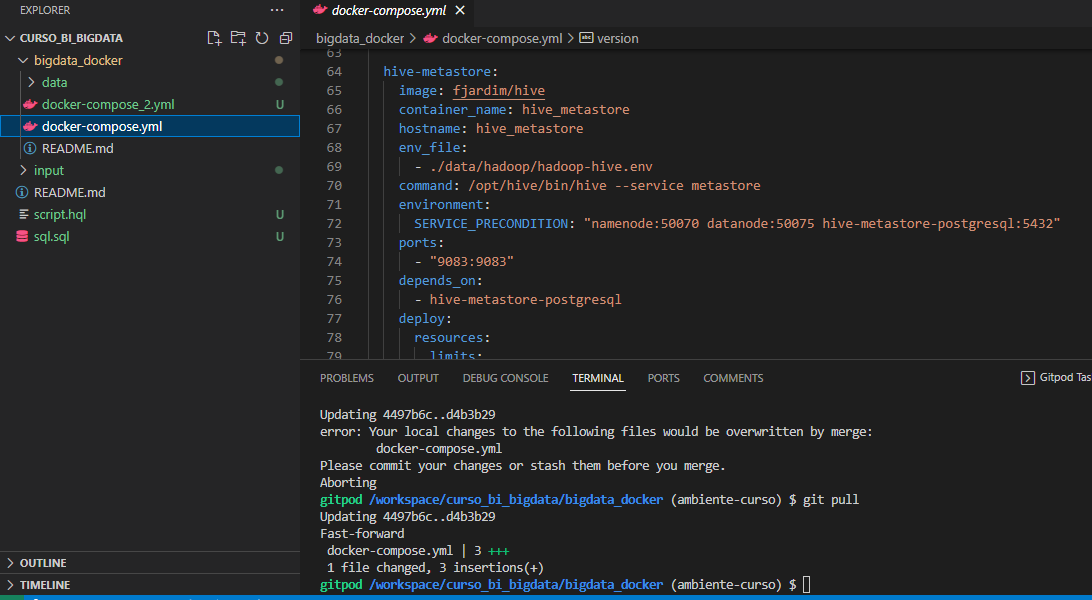
Entrar na pasta do projeto **docker\_bigdata**



Renomear o arquivo docker-compose.yml para docker-compose\_2.yml

No terminal dentro da pasta dar um git pull

**git pull**

**OBS> vai aparecer outro arquivo docker-compose.yml**

Depois iniciar o cluster

Inicializar o Cluster

**docker-compose up –d**

Entrando dentro do container hive-server

**docker exec -it hive-server bash**

Entrando dentro da pasta /input

**cd /input**

**Criar as Pastas**

- categorias - > **mkdir /input/categorias**

- produtos

- clientes

- pedido

- item\_pedido

- cidade

- estado

- parceiro

- subcategoria

- filial

Baixar os Arquivos no GitHub dentro se suas respectivas pastas

Cd /input/categorias

curl -O <https://raw.githubusercontent.com/caiuafranca/dados_curso/main/categoria.csv>

Criando pastas no HDFS

hdfs dfs -mkdir /datalake/raw/categorias



Obs. para cada arquivo baixado será preciso criar uma pasta

Dentro da pasta /input/categorias, vamos enviar o arquivo para o hdfs

**hdfs dfs -copyFromLocal categoria.csv /datalake/raw/categorias**

Verificar se o arquivo está dentro do hdfs

**Hdfs dfs –ls /datalake/raw/categorias**

**Acessar o hive com a string de conexão no terminal:**

beeline -u jdbc:hive2://localhost:10000

**criar a base de dados “desafio”**

* **Create database desafio;**
* **Use desafio;**

**Criar a tabela externa categorias**

CREATE TABLE IF NOT EXISTS desafio.categorias (

    id\_categoria string,

    ds\_categoria string,

    perc\_parceiro string

    )

COMMENT 'Tabela de Categorias'

ROW FORMAT DELIMITED

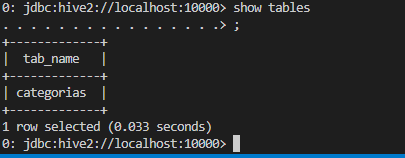
FIELDS TERMINATED BY '|'

STORED AS TEXTFILE

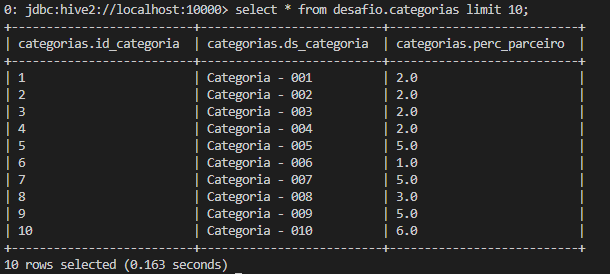
location '/datalake/raw/categorias/'

TBLPROPERTIES ("skip.header.line.count"="1");

**Verificar a criação da tabelas com o camando;**



**Efetuar uma consulta com limit 10 na tabela criada;**



**Este Processo iremos usar para fazer em todas as nossas tabelas;**