



# Big Data com Hive e Impala

Acessando o Hive

# Passo a Passo

- ❖ Baixar os arquivos na VM
- ❖ Vamos criar um banco e tabelas e importar os arquivos texto para Hive:
  - ❖ clientes.csv
  - ❖ despachantes.csv
  - ❖ locacao.csv
  - ❖ veículos.csv

# Iniciando

## ❖ Copia Arquivos para HDFS

```
hdfs dfs -mkdir /user/cloudera/locacao  
cd /home/cloudera/Downloads/
```

```
hdfs dfs -put *.csv  
/user/cloudera/locacao
```

```
hdfs dfs -ls /user/cloudera/locacao
```

```
hdfs dfs -cat  
/user/cloudera/locacao/clientes.csv
```

## ❖ Inicia beeline e cria banco de dados

```
beeline  
!connect jdbc:hive2://  
CREATE DATABASE TESTE;
```

# Mais sobre Conexão

```
beeline -u  
jdbc:hive2://localhost:10000/default -  
n usuario -p password
```

# Iniciando

```
SHOW DATABASES;  
OK  
+-----+  
| database_name |  
+-----+  
| default      |  
| teste        |  
+-----+  
DROP DATABASE TESTE CASCADE;
```

# External / Internal

- ❖ External
  - ❖ Hive não controla o arquivo
  - ❖ Arquivo pode ser utilizado por outro recurso
  - ❖ Tipo recomendado
  - ❖ Se excluir a tabela, apenas metadados são excluídos  
(arquivo permanece)
- ❖ Internal (Managed)
  - ❖ Hive controla
  - ❖ Acesso exclusivo pelo Hive
  - ❖ Excluir tabela exclui arquivo físico

# Clientes

```
CREATE DATABASE LOCACAO;  
USE LOCACAO;  
  
CREATE EXTERNAL TABLE CLIENTES (idcliente int, cnh string, cpf  
string, validadecnh date, nome string, datacadastro date,  
datanascimento date, telefone string, status string) row format  
delimited fields terminated by ',' STORED AS TEXTFILE;  
  
LOAD DATA INPATH '/user/cloudera/locacao/clientes.csv' INTO TABLE  
CLIENTES;  
  
SELECT * FROM CLIENTES;
```



# Veiculos

```
CREATE EXTERNAL TABLE VEICULOS (idveiculo int, dataaquisicao date,  
ano int, modelo string, placa string, status string, diaria  
double) row format delimited fields terminated by ',' STORED AS  
TEXTFILE;
```

```
LOAD DATA INPATH '/user/cloudera/locacao/veiculos.csv' INTO TABLE  
VEICULOS;
```

```
SELECT * FROM VEICULOS;
```



# Despachantes

```
CREATE EXTERNAL TABLE DESPACHANTES (iddespachante int, nome  
string, status string, filial string) row format delimited fields  
terminated by ',' STORED AS TEXTFILE;
```

```
LOAD DATA INPATH '/user/cloudera/locacao/despachantes.csv' INTO  
TABLE DESPACHANTES;
```

```
SELECT * FROM DESPACHANTES;
```



# Locacao

```
CREATE EXTERNAL TABLE LOCACAO (idlocacao int, idcliente int,  
iddespachante int, idveiculo int, datalocacao date, dataentrega  
date, total double) row format delimited fields terminated by ','  
STORED AS TEXTFILE;
```

```
LOAD DATA INPATH '/user/cloudera/locacao/locacao.csv' INTO TABLE  
LOCACAO;
```

```
SELECT * FROM LOCACAO;
```



# Propriedades de Tabelas

- ❖ immutable: não permite inserir mais dados
  - ❖ orc.compress: algoritmo de compressão formato orc
  - ❖ skip.hearder.line.count: ignora linhas iniciais (cabeçalhos)
- 
- ❖ CREATE EXTERNAL TABLE VEICULOS (idveiculo int, dataaquisicao date, ano int, modelo string, placa string, status string, diaria double) row format delimited fields terminated by ',' STORED AS TEXTFILE TBLPROPERTIES("skip.hearder.line.count"="2");

