



# Formação Engenheiro de Dados

Pig: Comandos

# Carregar dados do HDFS

```
dimensaocliente = LOAD '/user/cloudera/dimensaocliente/' USING  
PigStorage(',') as  
(chavecliente:int, idcliente:int, cliente:chararray,  
estado:chararray, sexo:chararray, status:chararray,  
datainiciovalidade:datetime, datafimvalidade:datetime);
```

# Exibir

Dump dimensaocliente;

(193,193,Gláuber Guedella,RS,M,Silver,2019-07-05T00:00:00.000-07:00,)

(194,194,Gláucia Canela,RO,F,Silver,2019-07-05T00:00:00.000-07:00,)

(195,195,Godinho ou Godim Fogaça,RN,M,Silver,2019-07-05T00:00:00.000-07:00,)

(196,196,Godinho ou Godim Jácome,RN,M,Silver,2019-07-05T00:00:00.000-07:00,)

# Exibir Dados

```
dim10 = limit dimensaocliente 10;  
dump dim10;  
  
(193,193,Gláuber Guedella,RS,M,Silver,2019-07-05T00:00:00.000-07:00,)  
(194,194,Gláucia Canela,RO,F,Silver,2019-07-05T00:00:00.000-07:00,)  
(195,195,Godinho ou Godim Fogça,RN,M,Silver,2019-07-05T00:00:00.000-07:00,)  
(196,196,Godinho ou Godim Jácome,RN,M,Silver,2019-07-05T00:00:00.000-07:00,)  
(197,197,Godofredo Mascareñas,SP,M,Silver,2019-07-05T00:00:00.000-07:00,)  
(198,198,Godofredo Quiroga,BA,M,Silver,2019-07-05T00:00:00.000-07:00,)  
(199,199,Gomes Hurtado,AC,M,Silver,2019-07-05T00:00:00.000-07:00,)  
(200,200,Gonçalo Figueiró,PI,M,Silver,2019-07-05T00:00:00.000-07:00,)  
(201,201,Graça Tabosa,RN,F,Silver,2019-07-05T00:00:00.000-07:00,)  
(202,202,Greice Lameirinhas,PI,F,Silver,2019-07-05T00:00:00.000-07:00,)
```

# Exibir Schema

```
describe dimensaocliente;  
  
dimensaocliente: {chavecliente: int,idcliente: int,cliente:  
chararray,estado: chararray,sexo: chararray,status:  
chararray,datainiciovalidade: datetime,datafimvalidade: datetime}
```

# Agrupar

```
clienteporstatus = GROUP dimensaocliente by status;
```

```
(Platinum, { (247, 247, Joana Ataí, GO, F, Platinum, 2019-07-05T00:00:00.000-07:00, ), (256, 3, Adélio Lisboa, SE, M, Platinum, 2019-07-05T00:00:00.000-07:00, ), (172, 172, Flor Vilanova, CE, M, Platinum, 2019-07-05T00:00:00.000-07:00, ), (230, 230, Ibijara Botelho, RR, F, Platinum, 2019-07-05T00:00:00.000-07:00, ), (6, 6, Adriana Guedelha, RO, F, Platinum, 2019-07-05T00:00:00.000-07:00, ) })
```

# Junção

```
juncaocliente = JOIN fatovendas by chavecliente, dimensaocliente  
by chavecliente;
```

```
(899,7,251,5,17149,1,2268.99,2268.99,862.22,251,1,Adelina  
Buenaventura,RJ,M,Gold,2019-07-05T00:00:00.000-07:00,)
```

```
(898,7,251,4,17149,1,6892.2,6892.2,2412.27,251,1,Adelina  
Buenaventura,RJ,M,Gold,2019-07-05T00:00:00.000-07:00,)
```

```
(901,7,251,10,17149,1,114.75,114.75,33.28,251,1,Adelina  
Buenaventura,RJ,M,Gold,2019-07-05T00:00:00.000-07:00,)
```

# Junção

Left join:

```
juncaocliente = JOIN fatovendas by chavecliente LEFT OUTER,  
dimensaocliente by chavecliente;
```

Chave Composta:

```
juncaocliente = JOIN fatovendas by (CHAVE1,CHAVE2,CHAVE3),  
dimensaocliente by CHAVE1,CHAVE2,CHAVE3);
```

# Filtro

```
dimensaoclientePlatinum = FILTER dimensaocliente BY status ==  
'Platinum';
```

(230,230,Ibijara Botelho,RR,F,Platinum,2019-07-05T00:00:00.000-07:00,)

(247,247,Joana Ataí,GO,F,Platinum,2019-07-05T00:00:00.000-07:00,)

(256,3,Adélio Lisboa,SE,M,Platinum,2019-07-05T00:00:00.000-07:00,)

(172,172,Flor Vilanova,CE,M,Platinum,2019-07-05T00:00:00.000-07:00,)

(6,6,Adriana Guédelha,RO,F,Platinum,2019-07-05T00:00:00.000-07:00,)

# Dividir

```
SPLIT dimensaocliente into dimensaoGold if status=='Silver',  
dimensaoPlatinum if status=='Platinum';
```

```
(230,230,Ibijara Botelho,RR,F,Platinum,2019-07-05T00:00:00.000-07:00,)  
(247,247,Joana Ataí,GO,F,Platinum,2019-07-05T00:00:00.000-07:00,)  
(256,3,Adélio Lisboa,SE,M,Platinum,2019-07-05T00:00:00.000-07:00,)  
(172,172,Flor Vilanova,CE,M,Platinum,2019-07-05T00:00:00.000-07:00,)  
(6,6,Adriana Guédelha,RO,F,Platinum,2019-07-05T00:00:00.000-07:00,)
```

# Ordenar

```
dimensaoclienteorder = ORDER dimensaocliente BY cliente ASC;
```

```
(1,1,Adelina Buenaventura,RJ,M,Silver,2019-07-05T00:00:00.000-07:00,2019-07-05T00:00:00.000-07:00)
```

```
(251,1,Adelina Buenaventura,RJ,M,Gold,2019-07-05T00:00:00.000-07:00,)
```

```
(252,2,Adelino Gago,RJ,M,Gold,2019-07-05T00:00:00.000-07:00,)
```

```
(2,2,Adelino Gago,RJ,M,Silver,2019-07-05T00:00:00.000-07:00,2019-07-05T00:00:00.000-07:00)
```

```
(255,5,Adolfo Patrício,PE,M,Gold,2019-07-05T00:00:00.000-07:00,)
```

# Persistir Dados

```
STORE dimensaoclientePlatinum INTO '/user/cloudera/pig/' USING  
PigStorage(',',');
```