



Big Data com Hive e Impala

ORC

ORC

- ◇ Optimized Row Columnar
- ◇ Formato binário, comprimido
- ◇ Orientado a coluna
- ◇ Arquivo até 75% menor
- ◇ Otimizado para Análise
- ◇ Vem se tornando o padrão em Data Lakes Hadoop

Compressão: linha vs coluna

veiculos.idveiculo	veiculos.dataaquisicao	veiculos.ano	veiculos.modelo	veiculos.placa	veiculos.status	veiculos.diaria
3	2019-01-10	2019	BMW 750i xDrive	LWJ2931	Disponivel	1450.0
7	2019-01-06	2019	BMW 750i xDrive	LWJ2935	Disponivel	1450.0
9	2019-01-27	2019	BMW 750i xDrive	LWJ2937	Disponivel	1450.0
12	2019-01-27	2019	BMW 750i xDrive	LWJ2940	Disponivel	1450.0
18	2019-01-24	2019	BMW 750i xDrive	LWJ2946	Disponivel	1450.0
23	2019-01-13	2019	BMW 750i xDrive	LWJ2951	Disponivel	1450.0
26	2019-01-05	2019	BMW 750i xDrive	LWJ2954	Disponivel	1450.0
28	2019-01-28	2019	BMW 750i xDrive	LWJ2956	Disponivel	1450.0
30	2019-01-12	2019	BMW 750i xDrive	LWJ2958	Disponivel	1450.0



Leitura: linha vs coluna

```
SELECT MODELO, DIARIA FROM LOCACAO
```

Linhas: todo os dados são carregados, pois estão fisicamente em um mesmo arquivo

Colunas: apenas modelo e diária são carregados, pois estão fisicamente em arquivos separados

veiculos.idveiculo	veiculos.dataaquisicao	veiculos.ano	veiculos.modelo	veiculos.placa	veiculos.status	veiculos.diaria
3	2019-01-10	2019	BMW 750i xDrive	LWJ2931	Disponivel	1450.0
7	2019-01-06	2019	BMW 750i xDrive	LWJ2935	Disponivel	1450.0
9	2019-01-27	2019	BMW 750i xDrive	LWJ2937	Disponivel	1450.0
12	2019-01-27	2019	BMW 750i xDrive	LWJ2940	Disponivel	1450.0
18	2019-01-24	2019	BMW 750i xDrive	LWJ2946	Disponivel	1450.0
23	2019-01-13	2019	BMW 750i xDrive	LWJ2951	Disponivel	1450.0
26	2019-01-05	2019	BMW 750i xDrive	LWJ2954	Disponivel	1450.0
28	2019-01-28	2019	BMW 750i xDrive	LWJ2956	Disponivel	1450.0
30	2019-01-12	2019	BMW 750i xDrive	LWJ2958	Disponivel	1450.0



Criando arquivos ORC

- ◇ Load não faz transformação de dados
 - ◇ Carregar dados para tabela em formato original: ex: csv
 - ◇ Criar tabela com mesma estrutura com formato ORC
 - ◇ Inserir dados da tabela original para a tabela em formato ORC

Clientes

```
CREATE EXTERNAL TABLE  
CLIENTES_ORC(idcliente int, cnh string,  
cpf string, validadecnh date, nome  
string, datacadastro date,  
datanascimento date, telefone string,  
status string) stored as orc;
```

```
insert overwrite table CLIENTES_ORC  
select * from clientes;
```