



# Big Data com Hive e Impala

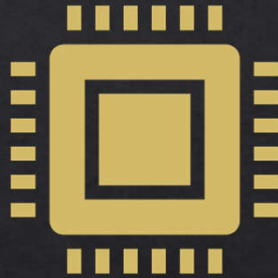
Vetorização

# Otimização

- ✓ Arquivo ORC
  - ✓ Partition
  - ✓ Bucket
- 
- ◇ Nesta seção:
    - ◇ Vetorização
    - ◇ Cost Based Optimization
    - ◇ Engine (Spark)



# Vetorização



Reduz o uso de CPU em  
consultas



Processa blocos de 1024  
linhas por vez



## Como usar

- ◆ Deve utilizar formato ORC
- ◆ Definir parâmetro

```
set hive.vectorized.execution.enabled =  
true;
```



# Exemplo

- ◇ Exemplo, join entre duas tabelas ORC
- ◇ CLIENTES\_ORC JA EXISTE
- ◇ Criar FATOS\_ORC



# Criação de Locacao\_orc

```
CREATE EXTERNAL TABLE LOCACAO_ORC(idlocacao int, idcliente  
int, iddespachante int, idveiculo int, datalocacao date,  
dataentrega date, total double) stored as orc;
```

```
insert overwrite table LOCACAO_ORC select * from locacao;
```



# Teste

```
select loc.data, loc.cacao, loc.total, cli.nome from locacao_orc loc  
join clientes_orc cli on ( loc.idcliente = cli.idcliente);
```

```
set hive.vectorized.execution.enabled = true;
```

◆ Reexecução da consulta

```
set hive.vectorized.execution.enabled = false;
```

