



Big Data com Hive e Impala

CBO

CBO

- ◇ Cost-Based Optimization
- ◇ Coleta e usa estatísticas para melhorar a performance de consultas
- ◇ Tenta otimizar o plano de execução da consultas
- ◇ Trade off: overhead pela coleta de dados



Como usar

♦ Definir propriedades:

```
set hive.cbo.enable=true;  
set hive.compute.query.using.stats=true;  
set hive.stats.fetch.column.stats=true;  
set hive.stats.fetch.partition.stats=true;
```

♦ Se as tabelas já haviam sido criadas, é preciso calcular suas estatísticas:

```
ANALYZE TABLE LOCACAO_ORC COMPUTE STATISTICS;
```

```
ANALYZE TABLE LOCACAO_ORC COMPUTE STATISTICS FOR COLUMNS;
```



Exemplo

```
select sum(total) from locacao_orc
```

```
set hive.cbo.enable=true;  
set hive.compute.query.using.stats=true;  
set hive.stats.fetch.column.stats=true;  
set hive.stats.fetch.partition.stats=true;
```

```
ANALYZE TABLE LOCACAO_ORC COMPUTE STATISTICS;
```

```
ANALYZE TABLE LOCACAO_ORC COMPUTE STATISTICS FOR COLUMNS;
```

```
select sum(total) form locacao_orc
```



Exemplo

```
set hive.cbo.enable=false;  
set hive.compute.query.using.stats=false;  
set hive.stats.fetch.column.stats=false;  
set hive.stats.fetch.partition.stats=false;
```

