



Big Data com Hive e Impala

Hive com Spark

Execution
Engine



Map Reduce
(padrão)



Tez



Spark

Até 3X
Rápido



Spark

- ◆ Em tempo real / Interativo
- ◆ Compatível com MapReduce
- ◆ Compatível com Hadoop
- ◆ Em memória / disco / híbrido
- ◆ Em cluster
- ◆ Open Source



Hive no Cloudera

- ◇ Pode funcionar com MR ou Spark
 - ◇ A nível de consulta/seção
 - ◇ A nível de servidor
- ◇ Obs: a nível de seção sobrepõe configuração de servidor



Configuração

◇ Spark on Yarn Service (padrão)

Spark On YARN Service

Hive (Service-Wide) ↗

☒ Spark

☐ none



Configuração

- ◊ A nível de seção/consulta:
 - ◊ `set hive.execution.engine;`
 - ◊ `set hive.execution.engine=spark;`
 - ◊ `set hive.execution.engine-mr;`
- ◊ A nível de serviço:
 - ◊ Alterar arquivo de configuração
 - ◊ Reiniciar hive

```
sudo gedit /etc/hive/conf.dist/hive-site.xml  
sudo service hive-server2 stop  
sudo service hive-server2 start
```



Configurações de Servidor

```
<property>  
  <name>hive.execution.engine</name>  
  <value>spark</value>  
</property>
```



Exemplo

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

```
set hive.execution.engine=spark;
```

♦ Reexecutar consulta

```
set hive.execution.engine=mr;
```

