



Big Data com Hive e Impala

Arquitetura

Componentes



MetaStore



HiveServer2



MySQL Server



Hive Client

Hcatalog

- ❖ Metadados
- ❖ Não está no HDFS: MySQL, PostgreSQL ou Oracle
- ❖ Padrão: MySQL

```
+-- Tables_in_metastore
| BUCKETING_COLS
| CDS
| COLUMNS_V2
| COMPACTION_QUEUE
| COMPLETED_TXN_COMPONENTS
| DATABASE_PARAMS
| DBS
| DB_PRIVS
| DELEGATION_TOKENS
| FUNCS
| FUNC_RU
| GLOBAL_PRIVS
| HIVE_LOCKS
| IDXS
| INDEX_PARAMS
| MASTER_KEYS
| METASTORE_DB_PROPERTIES
| NEXT_COMPACTION_QUEUE_ID
| NEXT_LOCK_ID
| NEXT_TXN_ID
| NOTIFICATION_LOG
| NOTIFICATION_SEQUENCE
| NUCLEUS_TABLES
| PARTITIONS
| PARTITION_EVENTS
| PARTITION_KEYS
| PARTITION_KEY_VALS
| PARTITION_PARAMS
| PART_COL_PRIVS
| PART_COL_STATS
| PART_PRIVS
| ROLES
| ROLE_MAP
| SDS
| SD_PARAMS
| SEQUENCE_TABLE
| SERDES
| SERDE_PARAMS
| SKewed_COL_NAMES
| SKewed_COL_VALUE_LOC_MAP
| SKewed_STRING_LIST
| SKewed_STRING_LIST_VALUES
| SKewed_VALUES
| SORT_COLS
| TABLE_PARAMS
| TAB_COL_STATS
```



Execution Engine



Map Reduce
(padrão)



Tez



Spark



Schema on Read

Metadados definidos
na leitura dos dados

Tabelas



APONTAM PARA
ARQUIVOS NO HDFS



PODEMOS ESTAR EM
DIFERENTES
FORMATOS



UMA MESMO
ARQUIVO PODE
“PERTENCER” A
DIFERENTES TABELAS



UM ARQUIVO PODE
SER MANIPULADO
POR OUTRA
APLICAÇÃO



APAGAR TABELA
APAGA METADADOS,
NÃO ARQUIVOS*

Banco de Dados

- ❖ Conjunto de tabelas definidos em um schema
- ❖ Não é um arquivo proprietário como um SGBD
- ❖ Um mesmo arquivo pode estar em diferentes tabelas em diferentes banco de dados
- ❖ Fisicamente em /user/hive/warehouse

Configurações

- ❖ Configurações ativas: Arquivo `hive-site.xml`
 - ❖ Local `/etc/hive/conf.dist/`
-
- ❖ No Shell do beeline usar comando `set` (efeito na sessão)
`set hive.default.fileformat = Orc`
`set hive.default.fileformat.managed=Orc`

Configurações

- ❖ Configuration Variables
- ❖ Metastore Configuration Variables
- ❖ Configuration Variables do Hadoop
- ❖ Run Time Information