Big Data com Hive e Impala

Metadados

Tabelas

show tables

```
+------------------+---+
|     tab_name     |   |
+------------------+---+
| clientes         |   |
| despachantes     |   |
| locacao          |   |
| veiculos         |   |
+------------------+---+
```

Metadados

DESCRIBE CLIENTES

```
+------------------+------------+----------+--+
|     col_name     | data_type  | comment  |  |
+------------------+------------+----------+--+
| idcliente        | int        |          |  |
| cnh              | string     |          |  |
| cpf              | string     |          |  |
| validadecnh      | date       |          |  |
| nome             | string     |          |  |
| datacadastro     | date       |          |  |
| datanascimento   | date       |          |  |
| telefone         | string     |          |  |
| status           | string     |          |  |
+------------------+------------+----------+--+
```

Metadados

DESCRIBE FORMATTED LOCACAO;

col_name	data_type	comment	
# col_name	data_type	comment	
	NULL	NULL	
idlocacao	int		
idcliente	int		
iddespachante	int		
idveiculo	int		
datalocacao	date		
dataentrega	date		
total	double		
	NULL	NULL	
# Detailed Table Information	NULL	NULL	
Database:	locacao	NULL	
Owner:	root	NULL	
CreateTime:	Mon Aug 05 10:49:21 PDT 2019	NULL	
LastAccessTime:	UNKNOWN	NULL	
Protect Mode:	None	NULL	
Retention:	0	NULL	
Location:	hdfs://quickstart.cloudera:8020/user/hive/warehouse/locacao.db/locacao	NULL	
Table Type:	EXTERNAL_TABLE	NULL	
Table Parameters:	NULL	NULL	
	COLUMN_STATS_ACCURATE	true	
	EXTERNAL	TRUE	
	numFiles	1	
	totalSize	3730	
	transient_lastDdlTime	1565027371	
	NULL	NULL	
# Storage Information	NULL	NULL	
SerDe Library:	org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe	NULL	
InputFormat:	org.apache.hadoop.mapred.TextInputFormat	NULL	
OutputFormat:	org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat	NULL	
Compressed:	No	NULL	
Num Buckets:	-1	NULL	
Bucket Columns:	[]	NULL	
Sort Columns:	[]	NULL	
Storage Desc Params:	NULL	NULL	
	field.delim	,	
	serialization.format	,	

Metadados

```
DESCRIBE DATABASE LOCACAO;
```

```
+-----------+-----------+----------------------------------------------------------+-------------+-------------+-------------+--+
| db_name   | comment   |                         location                         | owner_name  | owner_type  | parameters  |  |
+-----------+-----------+----------------------------------------------------------+-------------+-------------+-------------+--+
| locacao   |           | hdfs://quickstart.cloudera:8020/user/hive/warehouse/locacao.db | root        | USER        |             |  |
+-----------+-----------+----------------------------------------------------------+-------------+-------------+-------------+--+
```

Consultado Hcatalog

```
mysql -u root -pcloudera

show databases;
```

```
+---------------------+
|  Database           |
+---------------------+
|  information_schema |
|  cm                 |
|  firehose           |
|  hue                |
|  metastore          |
|  mysql              |
|  nav                |
|  navms              |
|  oozie              |
|  retail_db          |
|  rman               |
|  sentry             |
+---------------------+
```

HIVE

PROF. FERNANDO AMARAL
www.datascientist.com.br

Consultado Hcatalog

```
use metastore;

show tables;
```

```
+---------------------------+
| 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             |
| TBLS                      |
| TBL_COL_PRIVS             |
| TBL_PRIVS                 |
| TXNS                      |
| TXN_COMPONENTS            |
| TYPES                     |
| TYPE_FIELDS               |
| VERSION                   |
+---------------------------+
```

Consultado Hcatalog

◈ Bancos de Dados

```
select * from dbs;
```

```
+-------+----------------------+-----------------------------------------------------------------+----------+------------+------------+
| DB_ID | DESC                 | DB_LOCATION_URI                                                 | NAME     | OWNER_NAME | OWNER_TYPE |
+-------+----------------------+-----------------------------------------------------------------+----------+------------+------------+
|     1 | Default Hive database | hdfs://quickstart.cloudera:8020/user/hive/warehouse            | default  | public     | ROLE       |
|    10 | NULL                 | hdfs://quickstart.cloudera:8020/user/hive/warehouse/locacao.db | locacao  | root       | USER       |
+-------+----------------------+-----------------------------------------------------------------+----------+------------+------------+
```

Consultado Hcatalog

◈ Tabelas

```
select * from TBLS where DB_ID=10;
```

TBL_ID	CREATE_TIME	DB_ID	LAST_ACCESS_TIME	OWNER	RETENTION	SD_ID	TBL_NAME	TBL_TYPE	VIEW_EXPANDED_TEXT	VIEW_ORIGINAL_TEXT	LINK_TARGET_ID
17	1565027304	10	0	root	0	16	clientes	EXTERNAL_TABLE	NULL	NULL	NULL
18	1565027323	10	0	root	0	17	veiculos	EXTERNAL_TABLE	NULL	NULL	NULL
19	1565027346	10	0	root	0	18	despachantes	EXTERNAL_TABLE	NULL	NULL	NULL
20	1565027361	10	0	root	0	19	locacao	EXTERNAL_TABLE	NULL	NULL	NULL

Consultado Hcatalog

◈ Coluna da tabela locação;

```
select * from COLUMNS_V2 where CD_ID = 19;
```

```
+-------+---------+---------------+-----------+-------------+
| CD_ID | COMMENT | COLUMN_NAME   | TYPE_NAME | INTEGER_IDX |
+-------+---------+---------------+-----------+-------------+
|  19   |  NULL   | dataentrega   | date      |           5 |
|  19   |  NULL   | datalocacao   | date      |           4 |
|  19   |  NULL   | idcliente     | int       |           1 |
|  19   |  NULL   | iddespachante | int       |           2 |
|  19   |  NULL   | idlocacao     | int       |           0 |
|  19   |  NULL   | idveiculo     | int       |           3 |
|  19   |  NULL   | total         | double    |           6 |
+-------+---------+---------------+-----------+-------------+
```