

主要用于处理与元数据库的交互，Hive 的元数据信息通常存储在关系型数据库中，常用 MySQL数据库作为元数据库管理。

Hive的元数据信息在MySQL数据中有57张表,关联到字段方式：**DB_ID->TBL_ID->SD_ID->CD_ID**，数据库-》表-》存储-》字段。

Hive数据库相关的元数据表（DBS、DATABASE_PARAMS）

DBS **-- 存储Hive中所有数据库的基本信息**

元数据表字段	说明	示例数据
DB_ID	数据库ID	1
DESC	数据库描述	测试库
DB_LOCATION_URI	数据库HDFS路径	hdfs://mycluster/user/hive/warehouse
NAME	数据库名	
OWNER_NAME	数据库所有者用户名	
OWNER_TYPE	所有者角色	USER

DATABASE_PARAMS **--该表存储数据库的相关参数，在CREATE DATABASE时候用**

元数据表字段	说明	示例数据
DB_ID	数据库ID	2
PARAM_KEY	参数名	createdby
PARAM_VALUE	参数值	lxw1234

Hive表和视图相关的元数据表

主要有TBLS、TABLE_PARAMS、TBL_PRIVS，这三张表通过TBL_ID关联。

TBLS 该表中存储Hive表、视图、索引表的基本信息。

元数据表字段	说明	示例数据

TBL_ID	表ID	1
CREATE_TIME	创建时间	1436317071
DB_ID	数据库ID	2，对应DBS中的DB_ID
LAST_ACCESS_TIME	上次访问时间	1436317071
OWNER	所有者	admin
RETENTION	保留字段	0
SD_ID	序列化配置信息	86，对应SDS表中的SD_ID
TBL_NAME	表名	student
TBL_TYPE	表类型	MANAGED_TABLE、EXTERNAL_TABLE、INDEX_TAB VIRTUAL_VIEW
VIEW_EXPANDED_TEXT	视图的详细HQL语句	select `lxw1234`.`pt`, `lxw1234`.`pcid` from `liuxiaowen`.`lxw1234`
VIEW_ORIGINAL_TEXT	视图的原始HQL语句	select * from lxw1234

TABLE_PARAMS 该表存储表/视图的属性信息。

元数据表字段	说明	示例数据
TBL_ID	表ID	1
PARAM_KEY	属性名	totalSize、numRows、EXTERNAL
PARAM_VALUE	属性值	970107336、21231028、TRUE

TBL_PRIVS 该表存储表/视图的授权信息

元数据表字段	说明	示例数据
TBL_GRANT_ID	授权ID	1
CREATE_TIME	授权时间	1436320455
GRANT_OPTION		0
GRANTOR	授权执行用户	liuxiaowen
GRANTOR_TYPE	授权者类型	USER

PRINCIPAL_NAME	被授权用户	username
PRINCIPAL_TYPE	被授权用户类型	USER
TBL_PRIV	权限	Select、Alter
TBL_ID	表ID	22，对应TBLS表中的TBL_ID

Hive文件存储信息相关的元数据表

主要涉及SDS、SD_PARAMS、SERDES、SERDE_PARAMS。由于HDFS支持的文件格式很多，而建Hive表时候也可以指定各种文件格式，Hive在将HQL解析成MapReduce时候，需要知道去哪里，使用哪种格式去读写HDFS文件，而这些信息就保存在这几张表中。

SDS 该表保存文件存储的基本信息，如INPUT_FORMAT、OUTPUT_FORMAT、是否压缩等。TBLS表中的SD_ID与该表关联，可以获取Hive表的存储信息。

元数据表字段	说明	示例数据
SD_ID	存储信息ID	1
CD_ID	字段信息ID	21，对应CDS表
INPUT_FORMAT	文件输入格式	org.apache.hadoop.mapred.TextInputFormat
IS_COMPRESSED	是否压缩	0
IS_STOREDASSUBDIRECTORIES	是否以子目录存储	0
LOCATION	HDFS路径	hdfs://namenode/hivedata/warehouse/ut.db/t_lxw
NUM_BUCKETS	分桶数量	5
OUTPUT_FORMAT	文件输出格式	org.apache.hadoop.hive ql.io.HiveIgnoreKeyTextOutputFormat
SERDE_ID	序列化类ID	3，对应SERDES表

SD_PARAMS 该表存储Hive存储的属性信息，在创建表时候使用

元数据表字段	说明	示例数据
SD_ID	存储配置ID	1

PARAM_KEY	存储属性名	
PARAM_VALUE	存储属性值	

SERDES 该表存储序列化使用的类信息

元数据表字段	说明	示例数据
SERDE_ID	序列化类配置ID	1
NAME	序列化类别名	
SLIB	序列化类	org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe

SERDE_PARAMS 该表存储序列化的一些属性、格式信息,比如：行、列分隔符

元数据表字段	说明	示例数据
SERDE_ID	序列化类配置ID	1
PARAM_KEY	属性名	field.delim
PARAM_VALUE	属性值	

Hive表字段相关的元数据表

COLUMNS_V2 该表存储表对应的字段信息

元数据表字段	说明	示例数据
CD_ID	字段信息ID	1
COMMENT	字段注释	
COLUMN_NAME	字段名	pt
TYPE_NAME	字段类型	string
INTEGER_IDX	字段顺序	2

Hive表分区相关的元数据表

主要涉及PARTITIONS、 PARTITION_KEYS、 PARTITION_KEY_VALS、 PARTITION_PARAMS

PARTITIONS 该表存储表分区的基本信息。

元数据表字段	说明	示例数据
PART_ID	分区ID	1
CREATE_TIME	分区创建时间	
LAST_ACCESS_TIME	最后一次访问时间	
PART_NAME	分区名	pt=2015-06-12
SD_ID	分区存储ID	21
TBL_ID	表ID	2

PARTITION_KEYS 该表存储分区的字段信息

元数据表字段	说明	示例数据
TBL_ID	表ID	2
PKEY_COMMENT	分区字段说明	
PKEY_NAME	分区字段名	pt
PKEY_TYPE	分区字段类型	string
INTEGER_IDX	分区字段顺序	1

PARTITION_KEY_VALS 该表存储分区字段值。

元数据表字段	说明	示例数据
PART_ID	分区ID	2
PART_KEY_VAL	分区字段值	2015-06-12
INTEGER_IDX	分区字段值顺序	0

***PARTITION_PARAMS* 该表存储分区的属性信息。**

元数据表字段	说明	示例数据
PART_ID	分区ID	2
PARAM_KEY	分区属性名	numFiles、 numRows
PARAM_VALUE	分区属性值	15、 502195

参考文章：

<https://www.cnblogs.com/frankdeng/p/9126088.html>