

谈谈大数据之Hive元数据管理

高级班学员 若泽大数据 2020-05-25 12:13

Hive元数据表

存储Hive版本的元数据表

VERSION表

Field	Type	Null	Key	字段说明
VER_ID	bigint(20)	NO	PRI	ID主键
SCHEMA_VERSION	varchar(127)	NO		Hive版本
SCHEMA_VERSION_V2	varchar(255)	YES		
VERSION_COMMENT	varchar(255)	NO		版本说明

这个表比较简单，就是一张version表，但是如果这张表出现问题，是不能进入Hive-Cli。

DBS表

BDS表数据，该表存储Hive中所有数据库的基本信息。

Field	Type	Null	Key	字段说明
DB_ID	bigint(20)	NO	PRI	数据库ID
DESC	varchar(4000)	YES		数据库描述
DB_LOCATION_URI	varchar(4000)	NO		数据库HDFS路径
NAME	varchar(128)	YES	UNI	数据库名
OWNER_NAME	varchar(128)	YES		数据库所有者用户名
OWNER_TYPE	varchar(10)	YES		所有者角色

DATABASE_PARAMS表

DATABASE_PARAMS该表存储数据库的相关参数，在CREATE DATABASE时候用WITH DBPROPERTIES (property_name=property_value, ...)指定的参数。**DBS和DATABASE_PARAMS这两张表通过DB_ID字段关联。**

Field	Type	Null	Key	字段说明
DB_ID	bigint(20)	NO	PRI	数据库ID
PARAM_KEY	varchar(180)	NO	PRI	参数名
PARAM_VALUE	varchar(4000)	YES		参数值

Hive表和视图相关的元数据表

主要的表有以下几张主TBLS、TABLE_PARAMS、TBL_PRIVS，并且这三张表通过TBL_ID关联。

TBLS表

TBLS:该表中存储Hive表，视图，索引表的基本信息

Field	Type	Null	Key	字段说明
TBL_ID	bigint(20)	NO	PRI	表ID
CREATE_TIME	int(11)	NO		创建时间
DB_ID	bigint(20)	YES	MUL	数据库ID
LAST_ACCESS_TIME	int(11)	NO		上次访问时间
OWNER	varchar(767)	YES		所有者
OWNER_TYPE	varchar(10)	YES		所有者类型
RETENTION	int(11)	NO		保留字段
SD_ID	bigint(20)	YES	MUL	序列化配置信息
TBL_NAME	varchar(128)	YES	MUL	表名
TBL_TYPE	varchar(128)	YES		表类型
VIEW_EXPANDED_TEXT	mediumtext	YES		视图的详细HQL语句
VIEW_ORIGINAL_TEXT	mediumtext	YES		视图的原始HQL语句

TABLE_PARAMS表

TABLE_PARAMS:该表存储表/视图的属性信息

Field	Type	Null	Key	字段说明
TBL_ID	bigint(20)	NO	PRI	表ID
PARAM_KEY	varchar(256)	NO	PRI	属性名
PARAM_VALUE	varchar(32672)	YES		属性值

TBL_PRIVS表

TBL_PRIVS：该表存储表/视图的授权信息

Field	Type	Null	Key	字段说明
TBL_GRANT_ID	bigint(20)	NO	PRI	授权ID
CREATE_TIME	int(11)	NO		授权时间
GRANT_OPTION	smallint(6)	NO		授权选项
GRANTOR	varchar(128)	YES		授权执行用户
GRANTOR_TYPE	varchar(128)	YES		授权者类型
PRINCIPAL_NAME	varchar(128)	YES		被授权用户
PRINCIPAL_TYPE	varchar(128)	YES		被授权用户类型
TBL_PRIV	varchar(128)	YES		权限
TBL_ID	bigint(20)	YES	PRI	表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 表

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。

Field	Type	Null	Key	字段说明
SD_ID	bigint(20)	NO	PRI	存储信息ID
CD_ID	bigint(20)	YES	MUL	字段信息ID
INPUT_FORMAT	varchar(4000)	YES		文件输入格式
IS_COMPRESSED	bit(1)	NO		是否压缩
IS_STOREDASSUBDIRECTORIES	bit(1)	NO		是否以子目录存储
LOCATION	varchar(4000)	YES		HDFS路径
NUM_BUCKETS	int(11)	NO		分桶数量
OUTPUT_FORMAT	varchar(4000)	YES		文件输出格式
SERDE_ID	bigint(20)	YES	MUL	文件输出格式

SD_PARAMS表

SD_PARAMS：该表存储Hive存储的属性信息

Field	Type	Null	Key	字段说明
SD_PARAMS	bigint(20)	NO	PRI	存储配置ID
PARAM_KEY	varchar(256)	NO	PRI	存储属性名
PARAM_VALUE	varchar(32672)	YES		存储属性值

SERDES表

Field	Type	Null	Key	字段说明
SERDE_ID	bigint(20)	NO	PRI	序列化类配置ID
NAME	varchar(128)	YES		属性名
SLIB	varchar(4000)	YES		属性值

Hive表字段相关的元数据表

COLUMNS_V2表

Field	Type	Null	Key	字段说明
CD_ID	bigint(20)	NO	PRI	字段信息ID
COMMENT	varchar(256)	YES		字段注释
COLUMN_NAME	varchar(128)	NO	PRI	字段名
TYPE_NAME	varchar(128)	NO		字段类型
INTEGER_IDX	int(11)	NO		字段顺序

Hive表分分区相关的元数据表

主要涉及PARTITIONS、PARTITION_KEYS、PARTITION_KEY_VALS、PARTITION_PARAMS

PARTITIONS表

PARTITIONS:该表存储表分区的基本信息

Field	Type	Null	Key	字段说明
PART_ID	bigint(20)	NO	PRI	分区ID
CREATE_TIME	int(11)	NO		分区创建时间
LAST_ACCESS_TIME	int(11)	NO		最后一次访问时间
PART_NAME	varchar(767)	YES	MUL	分区名
SD_ID	bigint(20)	YES	MUL	分区存储ID
TBL_ID	bigint(20)	YES	MUL	表ID

PARTITION_KEYS表

PARTITION_KEYS:该表存储分区的字段信息

Field	Type	Null	Key	字段说明
TBL_ID	bigint(20)	NO	PRI	PRI
PKEY_COMMENT	varchar(4000)	YES		分区字段说明
PKEY_NAME	varchar(128)	NO	PRI	分区字段名
PKEY_TYPE	varchar(767)	NO		分区字段类型
INTEGER_IDX	int(11)	NO		分区字段顺序

PARTITION_KEY_VALS表

PARTITION_KEY_VALS:该表存储分区字段值

Field	Type	Null	Key	字段说明
PART_ID	bigint(20)	NO	PRI	分区ID
PART_KEY_VAL	varchar(255)	YES		分区字段值
INTEGER_IDX	int(11)	NO	PRI	分区字段值顺序

PARTITION_PARAMS表

PARTITION_PARAMS:该表存储分区的属性信息

Field	Type	Null	Key	字段说明
PART_ID	bigint(20)	NO	PRI	分区ID
PARAM_KEY	varchar(256)	NO	PRI	分区属性名
PARAM_VALUE	varchar(4000)	YES		分区属性值

Read more

