关于Hive

1. **Hive的shell环境**

hive>show tables;

hive>hive –e ‘select \* from dummy’

生成简单单行表:

echo ‘X’ > /tmp/dummy.txt

hive –e “create table dummy(value string); load data local inpath ‘/tmp/dummy.txt’ overwrite into table dummy”

1. **示例**

**建表：**

Create table records (year string, temperature int, quality int) row format delimited fields terminated by ‘\t’

解析：声明了一个records表，包含三列year，temperature，quality。并且都指明了每一列的数据类型。row format声明的是数据文件的每一行是由制表符分隔的文本，Hive按照这一格式读取数据，每行三个字段，分别对应表中的三列，字段之间以制表符分隔，每行以分隔符分隔。

**加载数据：**

load data local inpath ‘input/ncdc/sample.txt’ overwrite into table records;

这个命令告诉Hive在指定的本地文件放入其仓库目录中，这个操作并不解析文件或者把它存储为内部数据格式，因为Hive并不强制使用任何特定文件格式，文件以原样存储，Hive表存储在本地文件系统中，其中可以在fs.default.name中配置，在Hive仓库目录中，表存储为目录，仓库目录由选项hive.metastore.warehouse.dir控制，默认值是/usr/hive/warehouse。

ls /usr/hive/warehouse/record/

他会输出sample.txt

查询：

hive> select year, MAX(temperature) from records where temperature != 9999 and in (0,1,4,5,9) group by year;

JDBC驱动：

Jdbc:hive2//host:port/db

1. Metastore

Metastore是Hive元数据的集中存放地，metastore包括两部分，服务和后台数据的存储。

默认metastore服务和Hive服务运行在同一个JVM中，它包含一个**内嵌**的以本地磁盘作为存储的Derby数据库实例，称为**内嵌metastore配置**，但每一次只能为每个metastore打开一个Hive会话。

如果需要支持多用户，需要使用一个独立的数据库，被称为**本地metastore配置**。

1. Metastore重要配置属性

|  |  |  |  |
| --- | --- | --- | --- |
| 属性名称 | 类型 | 默认值 | 描述 |
| hive.metastore.warehouse.dir | URI | /user/hive/warehouse | 相对于fs.default.name的目录，托管表就存储在这里 |
| hive.metastore.uris | 逗号分隔的URI | 未设定 | 如果未设置（默认值），则使用当前的metastore,否则连接到由URI列表指定要连接的远程metastore服务器。如果有多个远程服务器，则客户端便以轮询方式连接 |
| Javax.jdo.option.ConnectionURL | URI | Jdbc:derby:;databaseName=metastored b;create=true | Metastore数据库的JDBC URL |
| Javax.jdo.option.ConnectionDriverName | 字符串 | Org.apache.derby.jdbc.EmbeddedDriver | JDBC驱动器的类名 |
| Java.jdo.option.ConnectionUserName | 字符串 | APP | JDBC用户名 |
| Javax.jdo.option.ConnectionPassword | 字符串 | Mine | JDBC密码 |

1. 读时模式和写时模式

在传统数据库里，表的模式是在数据加载时强制确定的。如果在加载时发现书不符合模式，那么则拒绝加载数据，因为在写入数据库时对照模式进行检查，如果不符合模式就拒绝加载，成为写时模式。

Hive对数据的验证是在读取数据的时候，并不在加载数据时进行，成为读时模式。