Hive是什么：

* Hive是一个SQL解析引擎，他将SQL语句转译成MapReduce作业并在Hadoop上执行。
* Hive表是HDFS的一个文件目录，一个表名对应一个目录名，如果有分区的话，则分区值对应子目录名称。

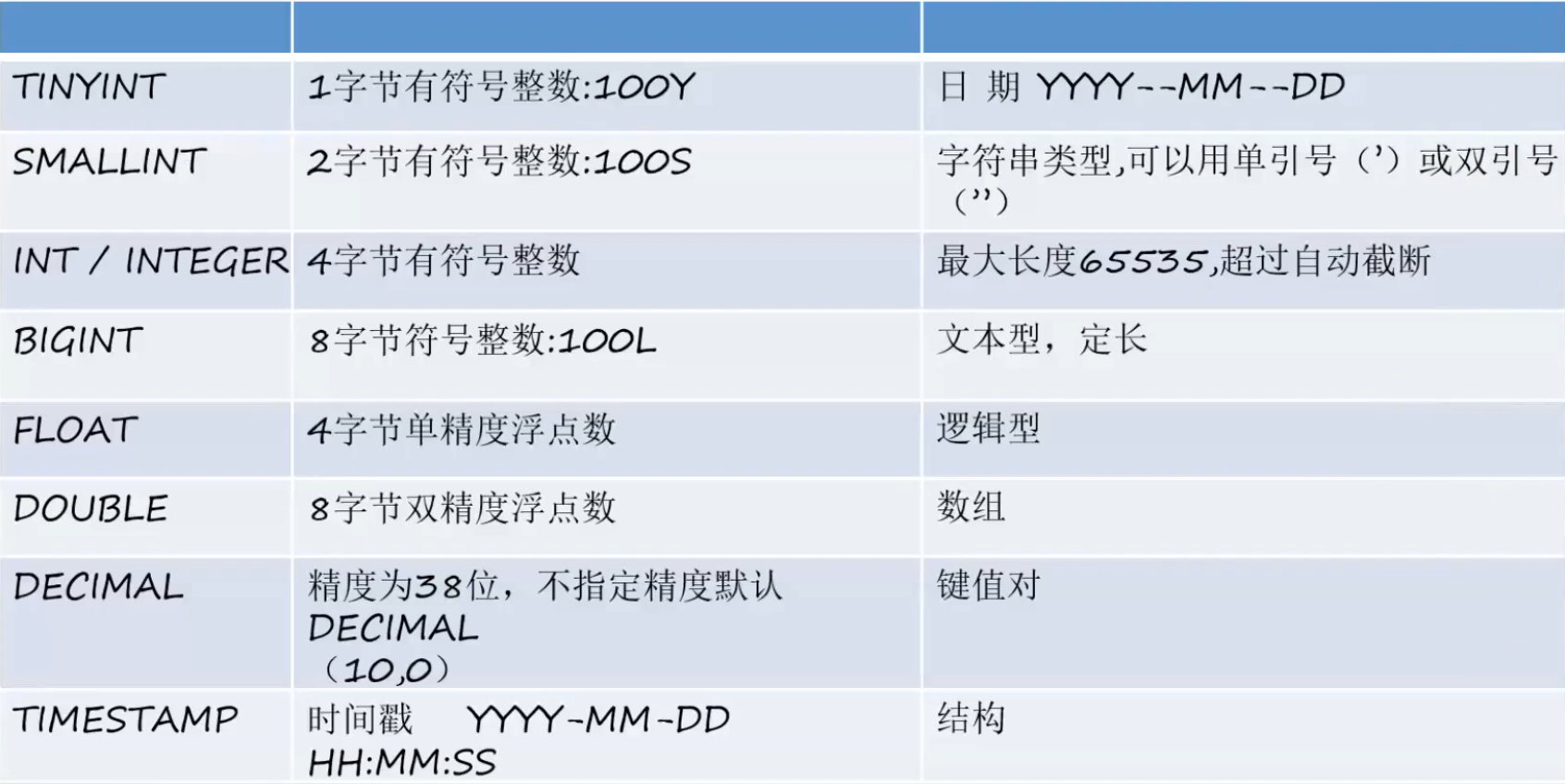
Hive的特点：

* 支持索引，加快数据查询
* 不同的存储类型，比如纯文本文件，HBase中的文件
* 将元数据保存在关系数据库中，大大减少了在查询过程中执行语义检查的时间
* 可以直接使用存储在hadoop文件系统中的数据。
* 内置大量用户函数UDF来操作时间，字符串和其他数据挖掘工具，支持用户扩展UDF函数来完成内置函数无法实现的操作。
* 类SQL的查询方式，将SQL查询转换为MapReduce的Job在Hadoop集群上执行。

Hive访问方式：

* Command-line shell:Hive自己的shell操作
* Trift/JDBC：封装了Thrift，java应用程序，可以通过指定的主机和端口连接到另一个进程中运行的Hive服务器
* ODBC客户端：ODBC驱动允许支持ODBC协议的应用程序连接到Hive
* WUI接口：通过浏览器访问Hive
* Driver：解释器，编译器，优化器，完成HQL查询语句从词法分析，语法分析，编译，优化及查询计划的生成。生成的查询计算存储在HDFS，并随后由MapReduce调用执行。

Hive数据类型：



常用SQL语法：