# Hadoop大数据开发与项目实战

## 第9章 Hadoop数据仓库Hive

Hive是基于Hadoop的一个数据仓库工具，可以将结构化数据文件映射为一张数据库表，并提供简单的SQL查询功能，可以将SQL语句转换为MapReduce任务运行。

Hive支持数据类型两大类：基本数据类型和复杂数据类型（集合数据类型）。

**基本数据类型（4种）：**

①tinyint/smallint/int/bigint

②float/double

③boolean

④string

**复杂数据类型（3种）：**

①Array

②Map

③Struct

Metastore是Hive元数据的集中存放地，默认和Hive绑定，部署在同一个JVM中，默认使用内嵌Derby数据库作为存储引擎，元数据存储在Derby中。（内嵌数据库没有办法为Hive开启多个实例）

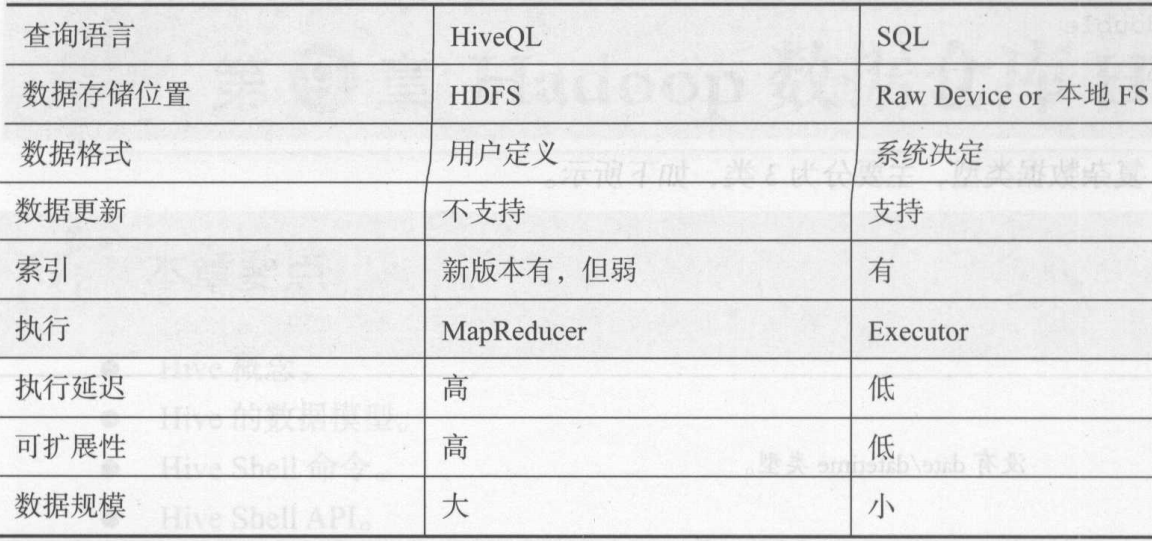
Derby引擎缺点：一次只能打开一个会话。

MySQL作为外置存储引擎，多用户同时访问。

Hive没有专门的数据存储格式，也没有为数据建立索引。

Hive所有数据都存储在HDFS中，包含以下数据模型：Table、External Table、Partition、Bucket。

**Hive与传统数据库的对比**

****

Hive的系统架构组成分成四个部分：用户接口部分、存放元数据的数据库，解释器、编译器，存放数据的HDFS系统。

