Hive：数据仓库工具，可以将hadoop下的原始结构化数据变成Hive表，支持一种几乎和SQL完全相同的语言HiveQL，但不支持更新、索引、事务等，Hive可以将HiveSQL映射成MapReduce作业

Hive起源自Facebook由Jeff Hammerbacher领导的团队，是构建在hadoop上的数据仓库框架，2008年Facebook将Hive项目贡献给Apache基金会

Hive组件与体系架构：用户接口有shell，thrift，web等，Thrift服务器，元数据库MetaStore（Derby，MySQL），解析器，hadoop

Hive安装：

内嵌模式：元数据保持在内嵌的Derby数据库模式，只允许一个会话连接

本地独立模式：在本地安装MySQL，把元数据放到MySQL内

远程模式：元数据放置在远程的MySQL数据库

Hive内嵌模式，下载解压后配置：

修改hive-config.sh文件

export JAVA\_HOME=java安装路径

export HIVE\_HOME=hive安装路径

export HADOOP\_HOME=hadoop安装路径

复制hive-default.xml为hive-site.xml并修改：

hive.metastore.warehouse.dir:HDFS的数据目录，默认/user/hive/warehouse

hive.exec.scratchdir:HDFS的临时文件目录，默认/tmp/hive-${user.name}

改变hive安装目录的所有者为hadoop集群用户

复制hive安装目录/config/hive-log4j.properties.template为hive-log4j.properties文件并修改：

org.apache.hadoop.metrics.jvm.EventCounter为org.apache.hadoop.log.metrics.EventCounter

Hive独立模式，在内嵌模式基础上配置：

安装MySQL并启动服务

在MySQL中创建Hive的账号并授予足够权限一般为all privileges，使用Hive账号登录MySQL创建Hive的数据库用于存放Hive的元数据

本地安装MySQL客户端，配置hive-site.xml文件，指定使用本地MySQL数据库的协议账号口令等，将MySQL的jdbc连接包复制到Hive安装目录的lib目录下

Hive远程模式配置：

修改hive-site.xml文件设置hive.metastore.local为false，指定使用远程MySQL数据库的协议账号口令等

HiveQL的语言标准大约相当于SQL92标准，Hive shell可以查看或临时设置配置Hive的参数，但只对当前会话有效

Hive可以像传统数据库一样使用JDBC或者ODBC连接，需要使用hive –service hiveserver命令启动Hive的Thrift Server，项目使用JDBC或者ODBC连接Hive需要将Hive和Hadoop的lib下的所有包复制到项目的类路径下

Hive的WEB的浏览器UI访问为<http://Hive>所在的主机的IP地址:9999/hwi/（端口默认9999）

Hive的数据存放在HDFS的warehouse目录下，一般每个表对应一个子目录，Hive桶会对应HDFS的warehouse目录下的一个子目录，Hive桶与Reduce相对应，本地的/tmp目录存放Hive的日志和执行计划