Sql与hive sql

1. hive是什么？

Hive 是一个基于 Hadoop 的数据仓库系统，用于处理和查询存储在 Hadoop 分布式文件系统（HDFS）中的大规模数据集。它提供了一种类似 SQL 的查询语言，称为 HiveQL

1. 数据存储在哪？

数据存储在HDFS中，HDFS是Hadoop 分布式文件系统，是 Hadoop 的存储系统，用于存储大数据

1. 查询是怎么执行的

Hive 将查询转换为 MapReduce 作业来执行，MapReduce是一种编程模型，用于大规模数据集的分布式处理

1. Hive表

Hive 中的数据表类似于关系型数据库中的表，但数据存储在 HDFS 中

分区是用于将表中的数据划分为更小的部分，提高查询效率

桶是进一步细分分区中的数据，提供更高效的数据采样和查询性能。

1. HiveQL与标准SQL的区别

Hive 表的数据存储在 HDFS 上，而不是关系型数据库的行存储

1. SQL是怎么执行的

解析（解析器检查 SQL 语句并拆分成语法单元最后根据语法规则生成一颗解析树）

｜

优化（查询优化器对解析树进行优化比如索引、子查询变联接等等）

｜

生成执行计划（优化器生成并选择出最小代价的执行计划）

｜

执行（数据库引擎根据计划从磁盘或缓存访问数据进行联接再进行过滤排序分组聚合等等最后返回结果）

1. 常见的MySQL支持的数据库引擎

InnoDB：需要事务处理和数据完整性（默认的引擎）

MyISAM：需要快速读取和大量读操作

MEMORY：需要快速访问和临时数据存储

CSV：需要与外部系统交换数据

ARCHIVE：需要长时间存储和压缩数据