

# HBase 与 Hive 程序设计

---

## HBase 基本工作原理

---

关系数据库有局限：扩展时性能与可靠性均下降

HBase：针对 HDFS 缺少结构化数据存取的缺陷，提供分布式数据管理系统，是 Google BigTable 的开源实现

HBase 构建于 HDFS 之上，为上层应用提供结构化和半结构化数据访存的能力，可以和 MapReduce 协同工作

数据通过行关键字，列关键字，时间戳进行索引

由一个 MasterServer 和一组子表服务器 Region Server 构成，更新数据时直接与子表服务器交流

查询时使用多级索引

## Hive 基本工作原理

---

Hive 包括数据的存储和查询，提供一个类似 SQL 的执行引擎

Hive 依赖于 HDFS 和 MapReduce

HiveQL 是类似于 SQL 的数据查询语言，提供 shell 接口和网络接口以及 JDBC 接口

Hive 基本的数据模型是表（tables），表可以按照一定规则划分为 partition，而一定范围内的数据按照 hash 的方式存储

Hive 的元数据存储在一般的关系数据库中，如 MySQL