# 起源

2006年，Google发表了一篇论文《BigTable：A Distributed Storage System for Structured Data》，介绍了一种数据模型——多维稀疏排序模型。

2007年，Powerset公司基于这篇文章研发了BigTable的开源版本，即HBase。

2008年，HBase成为Apache的顶级项目。

注：Google三驾马车：分布式文件系统GFS（开源实现：HDFS），分布式计算框架MapReduce（开源实现：Hadoop MapReduce），分布式结构化数据存储BigTable（开源实现：HBase）。

# 概述

Apache HBase是一种Key/Value系统，它运行在HDFS之上。**和Hive不一样，Hbase的能够在它的数据库上实时运行，而不是运行MapReduce任务**。Hive被分区为表格，表格又被进一步分割为列簇。列簇必须使用schema定义，列簇将某一类型列集合起来(列不要求schema定义)。例如，"message"列簇可能包含："to", "from" "date", "subject"和"body"。每一个 key/value对在Hbase中被定义为一个cell，每一个key由row-key，列簇、列和时间戳。在Hbase中，行是key/value映射的集合，这个映射通过row-key来唯一标识。Hbase利用Hadoop的基础设施，可以利用通用的设备进行水平的扩展。

## 特点

HBase通过存储key/value来工作。

它支持四种主要的操作：增加或者更新行，查看一个范围内的cell，获取指定的行，删除指定的行、列或者是列的版本。

版本信息用来获取历史数据(每一行的历史数据可以被删除，然后通过Hbase compactions就可以释放出空间)。

虽然HBase包括表格，但是schema仅仅被表格和列簇所要求，列不需要schema。Hbase的表格包括增加/计数功能。

## 限制

HBase查询是通过特定的语言来编写的，这种语言需要重新学习。类SQL的功能可以通过Apache Phonenix实现，但这是以必须提供schema为代价的。另外，Hbase也并不是兼容所有的ACID特性，虽然它支持某些特性。最后但不是最重要的——为了运行Hbase，Zookeeper是必须的，zookeeper是一个用来进行分布式协调的服务，这些服务包括配置服务，维护元信息和命名空间服务。

## Hive

Hive和Hbase是两种基于Hadoop的不同技术–Hive是一种类SQL的引擎，并且运行MapReduce任务，Hbase是一种在Hadoop之上的NoSQL 的Key/vale数据库。当然，这两种工具是可以同时使用的。就像用Google来搜索，用FaceBook进行社交一样，Hive可以用来进行统计查询，HBase可以用来进行实时查询，数据也可以从Hive写到Hbase，设置再从Hbase写回Hive。

# 原理

# 应用

Hbase非常适合用来进行**大数据**的**实时查询**。Facebook用Hbase进行消息和实时的分析。它也可以用来统计Facebook的连接数。