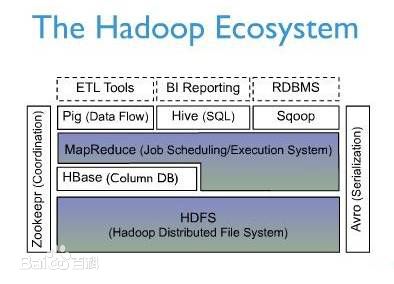
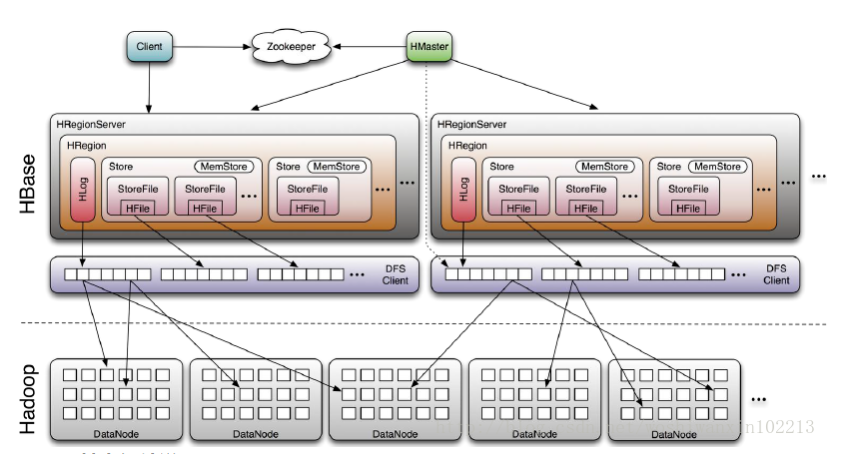
# Hbase存储

1. **为什么选择Hbase存储数据**

HBase – Hadoop Database，是一个高[可靠](https://baike.baidu.com/item/%E5%8F%AF%E9%9D%A0)性、高性能、面向列、可伸缩的[分布式存储系统](https://baike.baidu.com/item/%E5%88%86%E5%B8%83%E5%BC%8F%E5%AD%98%E5%82%A8%E7%B3%BB%E7%BB%9F)，利用HBase技术可在廉价PC Server上搭建起大规模[结构](https://baike.baidu.com/item/%E7%BB%93%E6%9E%84)化[存储](https://baike.baidu.com/item/%E5%AD%98%E5%82%A8)集群。HBase位于结构化[存储](https://baike.baidu.com/item/%E5%AD%98%E5%82%A8)层，Hadoop HDFS为HBase提供了高可靠性的底层存储支持，Hadoop MapReduce为HBase提供了高性能的计算能力，Zookeeper为HBase提供了稳定服务和failover机制。



HBase基于HDFS分布式存储系统，基本原理流程图如下：



HBase是一个构建在HDFS上的分布式列存储系统，为典型的key/value系统，主要用于海量结构化数据存储。从逻辑上讲，HBase将数据按照表、行和列进行存储。  
与Hadoop一样，HBase目标主要依靠横向扩展，通过不断增加廉价的商用服务器，来增加计算和存储能力。

**2、HBase表的特点**  
大：一个表可以有数十亿行，上百万列；  
无模式：每行都有一个可排序的主键和任意多的列，列可以根据需要动态的增加，同一张表中不同的行可以有截然不同的列；  
面向列：面向列（族）的存储和权限控制，列（族）独立检索；  
稀疏：空（null）列并不占用存储空间，表可以设计的非常稀疏；  
数据多版本：每个单元中的数据可以有多个版本，默认情况下版本号自动分配，是单元格插入时的时间戳；  
数据类型单一：HBase中的数据都是字符串，没有类型

1. **Hbase与MySQL的迁移**

Sqoop(发音：skup)是一款开源的工具，主要用于在Hadoop(Hive)与传统的数据库(mysql、postgresql...)间进行数据的传递，可以将一个关系型数据库*（例如 ： MySQL ,Oracle ,Postgres等）*中的数据导进到Hadoop的HDFS中，也可以将HDFS的数据导进到关系型数据库中。