理论根据：**GFS**(Google File System)和**Map****-Reduce**

**权威指南第二版**

**第一章**

术语：LAMP(Linux+Apache+Mysql+PHP)

核心：**HDFS**(文件系统Hadoop Distributed FileSystem)、**MapReduce**(计算架构)和**HBase**(结构化数据处理，分布式、按列存储的数据库)，相关的项目有**Hive(**分布式、按列存储的数据仓库**)**、**Pig(**数据流语言和运行环境，用以检索非常大的数据集**)**、**ZooKeeper(**分布式、可用性高的协调服务**)**等。

MapReduce和RDBMS比对：

前者适合以批处理的方式分析整个数据集，一次写多次读；后者适用于点查询和更新(其中，数据集已经被索引)。

第二章

page40