HDFS一一　　Hadoop分布式文件系统，GFS的Java开源实现，运行于大型商用机器集群，可实现分布式存储。  
MapReduce一一　　一种并行计算框架，Google MapReduce模型的Java开源实现，基于其写出来的应用程序能够运行在由上千个商用机器组成的大型集群上，并以一种可靠容错的方式并行处理T级别及以上的数据集。

Zookeepe一一　　一分布式协调系统，Google Chubby的Java开源实现，是高可用的和可靠的分布式协同(coordination)系统，提供分布式锁之类的基本服务，用于构建分布式应用。

Hbase一一　基于Hadoop的分布式数据库，Google BigTable的开源实现，是一个有序、稀疏、多维度的映射表，有良好的伸缩性和高可用性，用来将数据存储到各个计算节点上。  
Hive一一　是为提供简单的数据操作而设计的分布式数据仓库，它提供了简单的类似SQL语法的HiveQL语言进行数据查询。  
Cloudbase一一　　基于Hadoop的数据仓库，支持标准的SQL语法进行数据查询。  
Pig一一   大数据流处理系统。建立于Hadoop之上为并行计算环境提供了一套数据工作流语言和执行框架。  
Mahout一一    基于HadoopMapReduce的大规模数据挖掘与机器学习算法库。  
Oozie一一  　　MapReduce工作流管理系统。

Sqoop—一    数据转移系统，是一个用来将Hadoop和关系型数据库中的数据相互转移的工具，可以将一个关系型数据库中的数据导入Hadoop的HDFS中、也可以将HDFS的数据导入关系型数据库中。  
Flume一一　　一个可用的、可靠的、分布式的海量日志采集、聚合和传输系统。  
Scribe一一　　Facebook开源的日志收集聚合框架系统。

