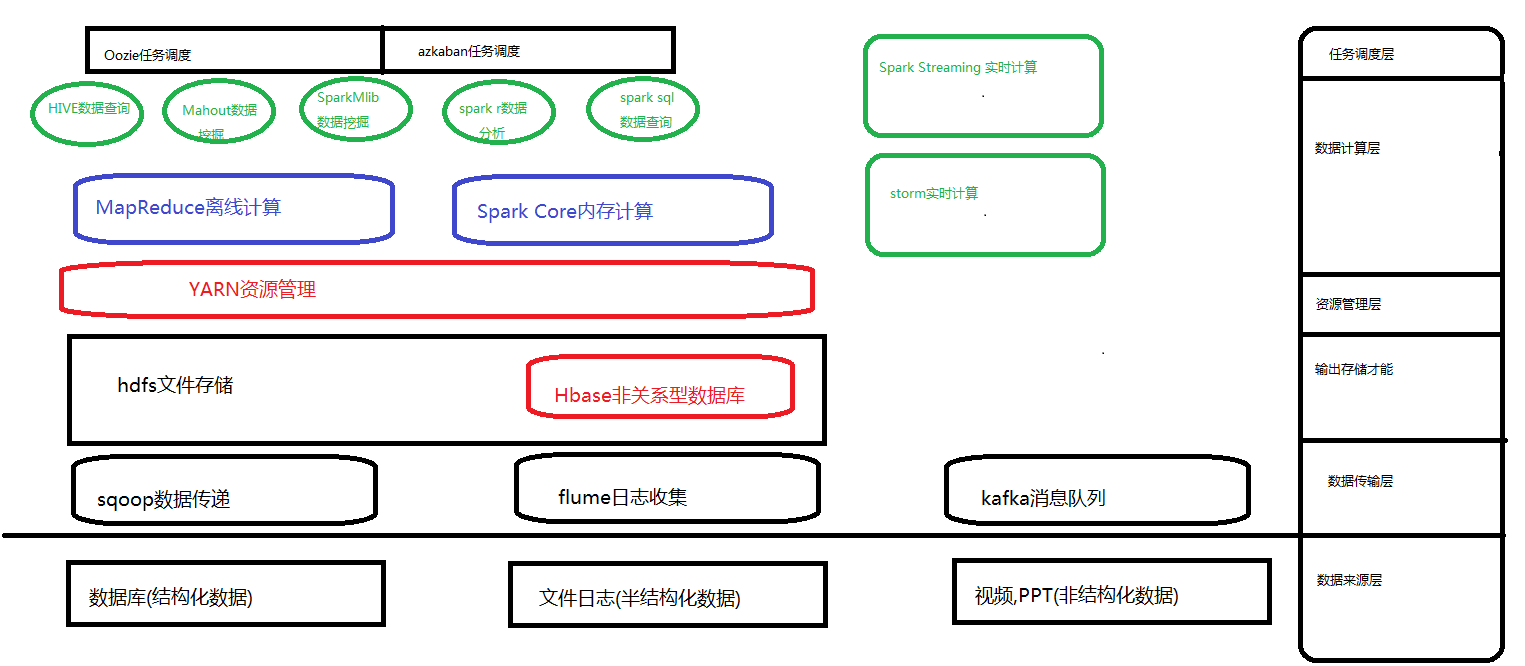
hadoop

1.大数据的特点:数据海量,传输速度,数据格式多样化,数据的价值

2.hadoop是一个分布式的框架,可以创建多个副本,可以对其进行扩展,可以并行操作,拥有自动容错功能.分为apache和cloudare两个版本.

3.hadoop分为四个部分,HDFS:分布式文件系统,相当于一个磁盘,MAPREDUCE:是一个离线的并行计算的框架,YARN:作业调度,资源分配,common:其他模块.

4.整个大数据的结构



5.在hdfs上运行hadoop

格式化namenode

配置hdfs-env.sh,hdfs-site.xml

启动namenode

sbins

启动datanode

6.在yarn上运行hadoop