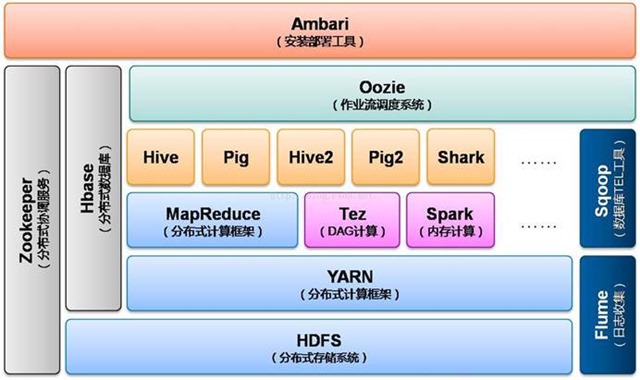
**一、Hadoop架构及其生态**



蓝色部分：hadoop核心模块HDFS、YARN和MapReduce

粉色部分：基于YARN的第三方分布式内存计算框架

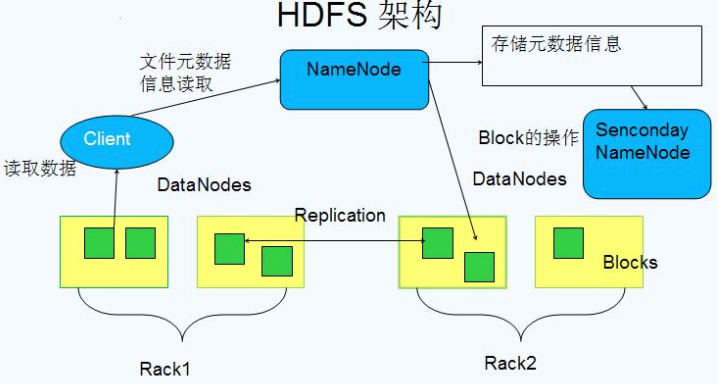
黄色部分：分布式计算工具(调用底层分布式计算框架，如MP等)

蓝色部分：数据导入工具(sqoop导入sql，flume导入日志)

灰色部分：独立模块(zk分布式协调，Hbase基于HDFS存储的列式数据库)

**二、HDFS**

**1 HDFS 1.x版**



NameNode

·元数据维护：文件树、文件和block关系、block和datanode关系

·持久化元数据：文件树镜像文件fsimage、文件操作日志文件edits

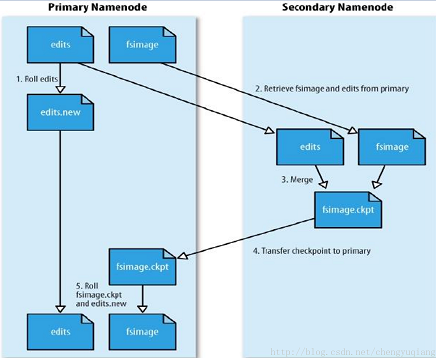
·接受客户端对文件系统访问(文件元数据访问、文件数据block重定向)

DataNode：

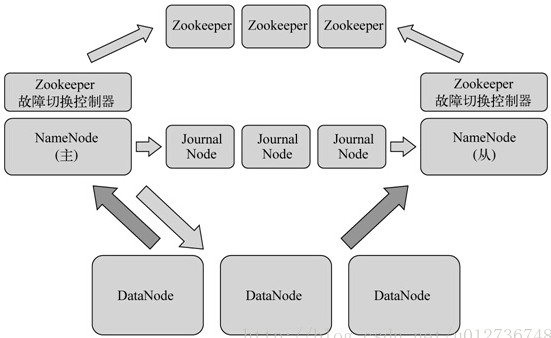
·存储文件内容block(默认128MB)

·接受客户端对文件内容访问

Secondary NameNode：整合NameNode对文件系统的操作碎片

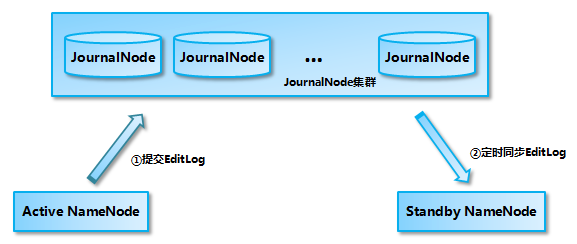


**2 HDFS高可用(2.x版)**

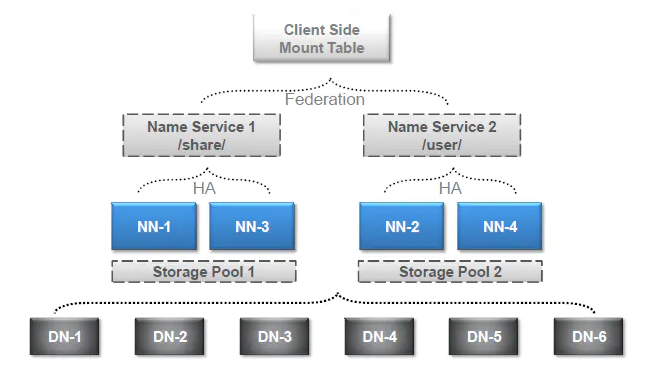


ZK故障切换控制器：ZKFailoverController作为独立的进程运行，对NN主从切换

JournalNode：QJM，取代Secondary NameNode

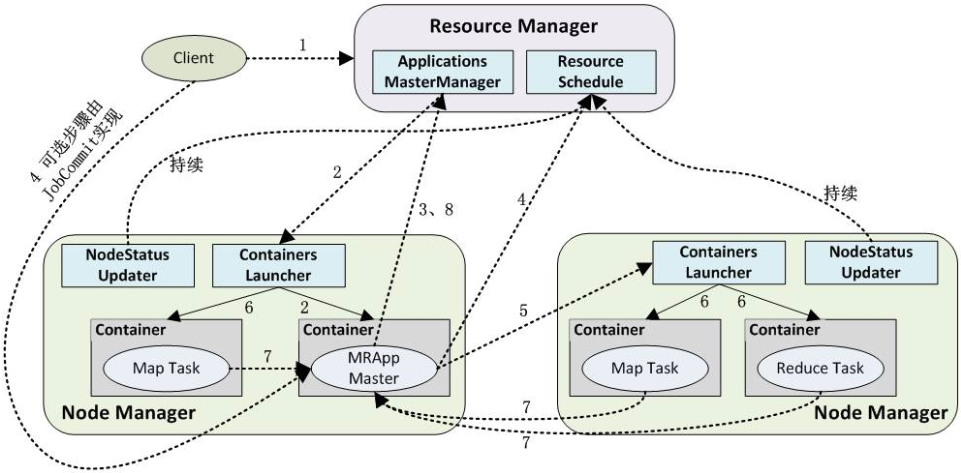


**3 HDFS联邦**



**三、YARN**

**1组织架构**



RM：负责对NM上的资源进行统一管理和调度

·应用管理器：接受应用提交、申请容器并启动AM、跟踪容器状态

·资源调度器：根据应用需求，结合集群机器资源状况，为应用分配容器

NM：定时向RM汇报资源占用情况和各个容器运行状态、接收AM对容器的使用

AM：负责申请Task容器(与RM通信)、启动Task并监控Task状态(与NM通信)

**2工作流程**

1用户向YARM提交程序

2 RM为应用分配容器，并要求该容器启动AM

3 AM向RM注册

4 AM向RM申请Task容器

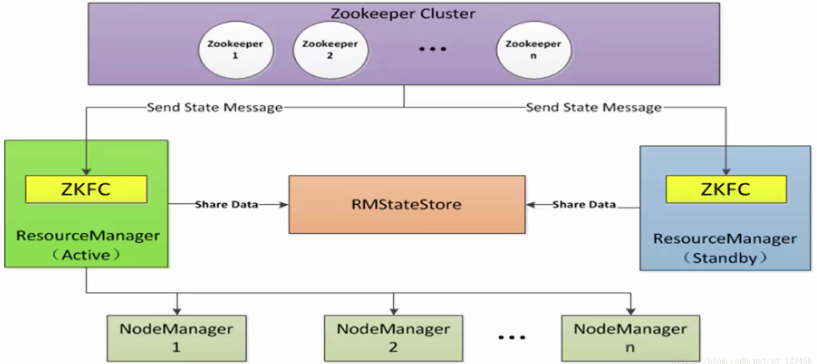
5 AM与NM通信，要求启动任务

6启动申请到容器中的Task任务

7各个Task向AM汇报自己状态和进度

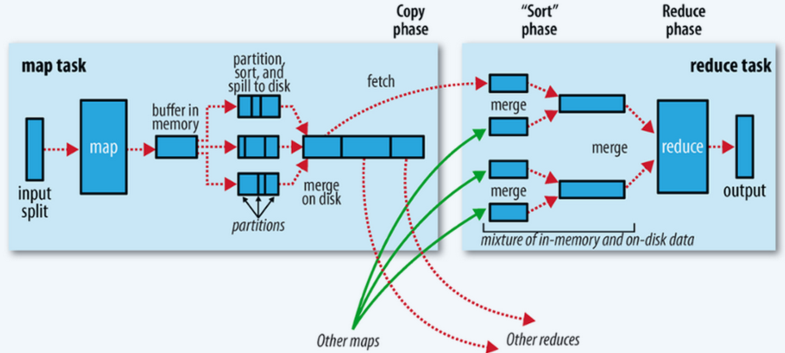
8 AM运行完成后向RM注销

**3高可用**



**四、MapReduce**

**1计算架构**



输入分片：即split逻辑分片，行不能被分割(HDFS物理块中行可能被分割)

split计算公式：max(minSize, min(goalSize, blockSize))，默认同hdfs块即128M

map引擎：将行数据转化为用户期望key,value形式

shuffle过程：map引擎与reduce引擎中间的部分

reduce引擎：将所有key, value转化为用户期望输出

输出分片：将reduce输出数据存入HDFS物理块

2 Shuffle过程

Collect阶段：对map输出

Spill阶段

Merge阶段

Copy阶段

Sort阶段