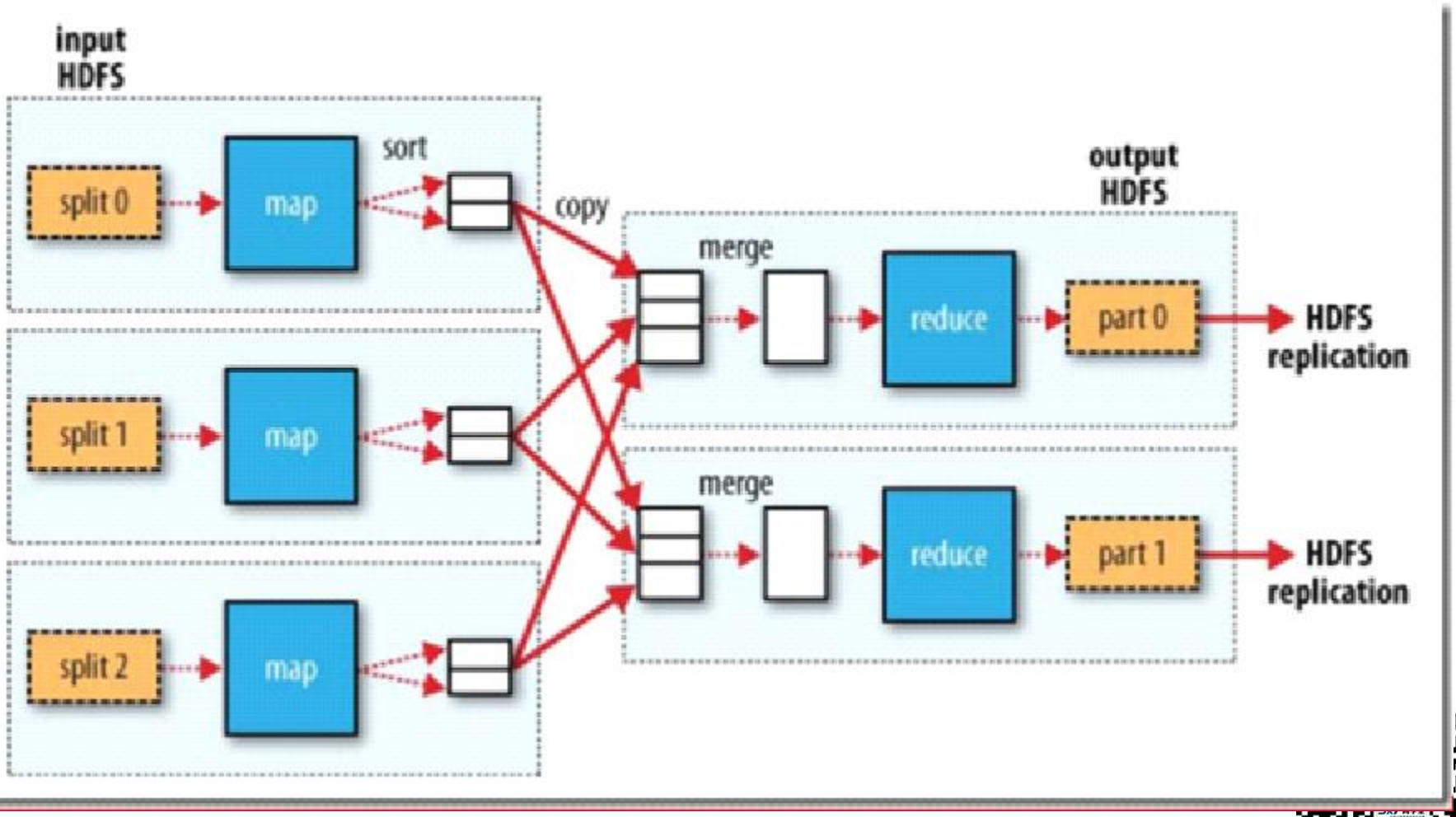
分布式云平台

讲师： 肖斌

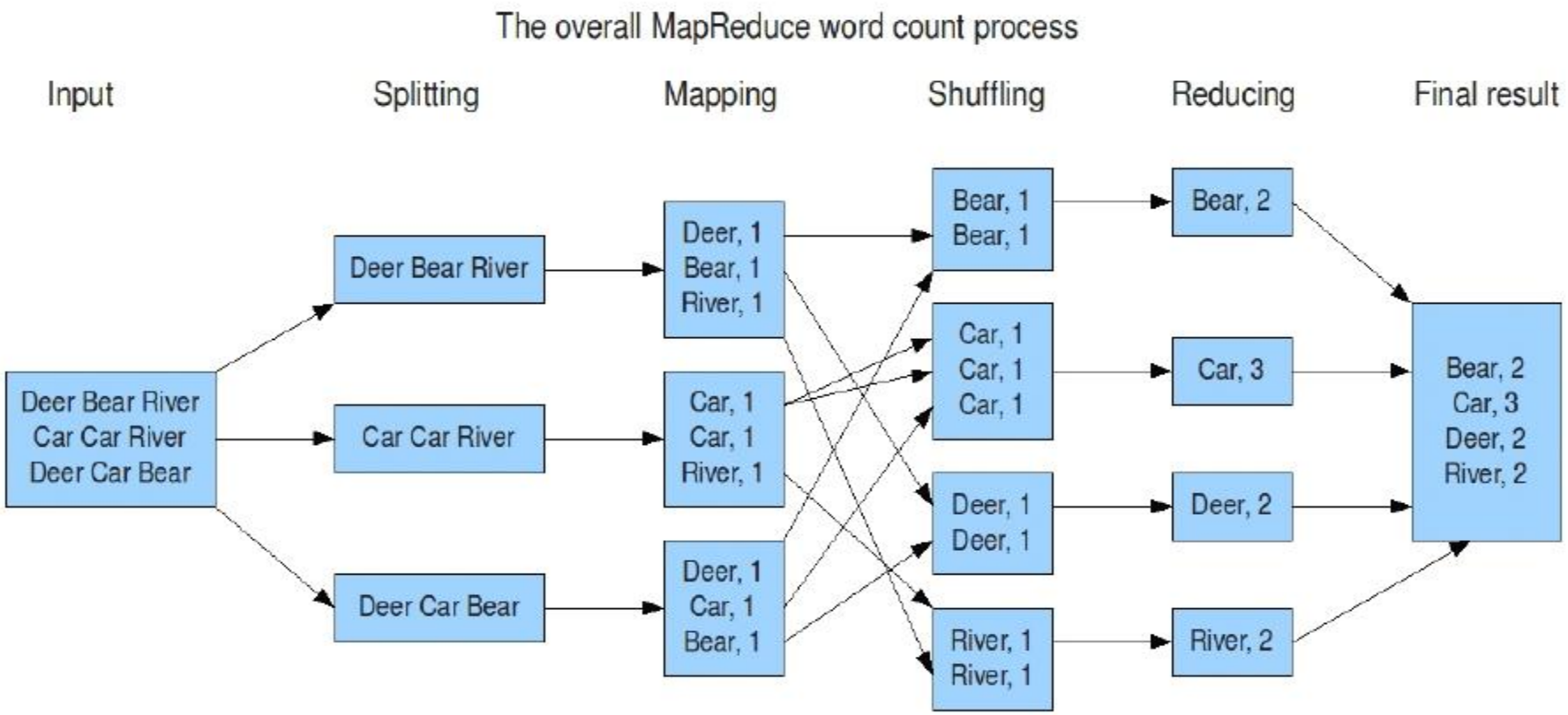
- Hadoop核心组件——MR
  - Hadoop 分布式计算框架（MapReduce）

- MapReduce设计理念
  - 何为分布式计算。
  - 移动计算，而不是移动数据。

- 计算框架MR

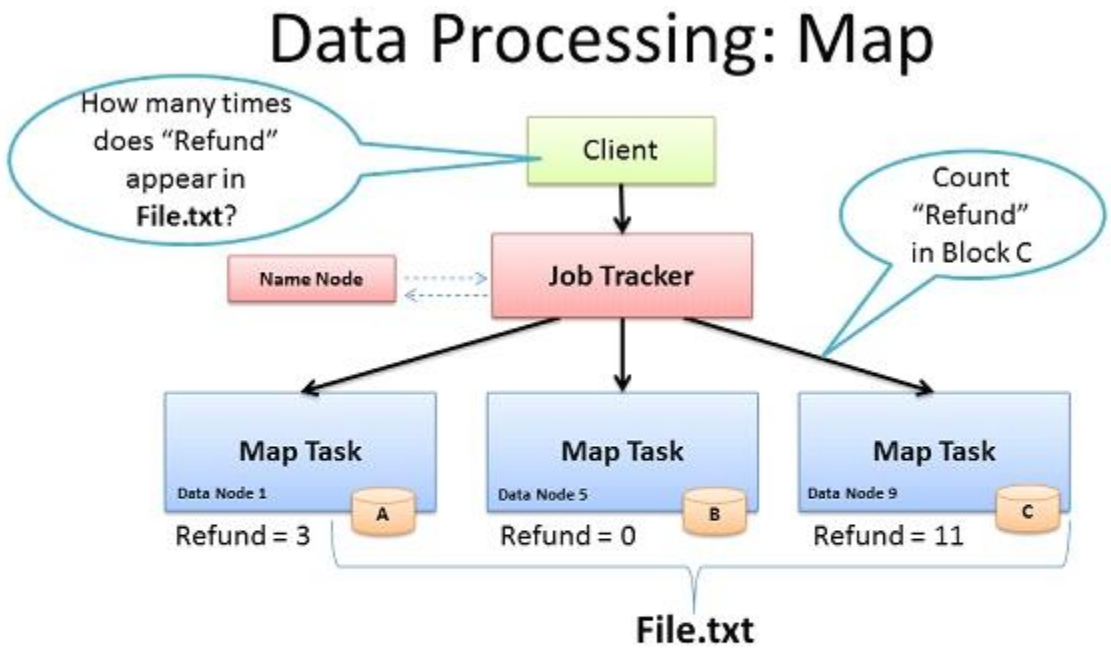


- 计算框架MR


The overall MapReduce word count process
Input
Splitting
Mapping
Shuffling
Reducing
Final result
Deer Bear River
Car Car River
Deer Car Bear
Deer Bear River
Car Car River
Deer Car Bear
Deer, 1
Bear, 1
River, 1
Car, 1
Car, 1
River, 1
Deer, 1
Car, 1
Bear, 1
Bear, 1
Bear, 1
Car, 1
Car, 1
Car, 1
Deer, 1
Deer, 1
River, 1
River, 1
Bear, 2
Car, 3
Deer, 2
River, 2
Bear, 2
Car, 3
Deer, 2
River, 2

- Mapper
  - Map-reduce的思想就是“分而治之”
    - Mapper负责“分”，即把复杂的任务分解为若干个“简单的任务”执行
  - “简单的任务”有几个含义：
    - 数据或计算规模相对于原任务要大大缩小；
    - 就近计算，即会被分配到存放了所需数据的节点进行计算；
    - 这些小任务可以并行计算，彼此间几乎没有依赖关系

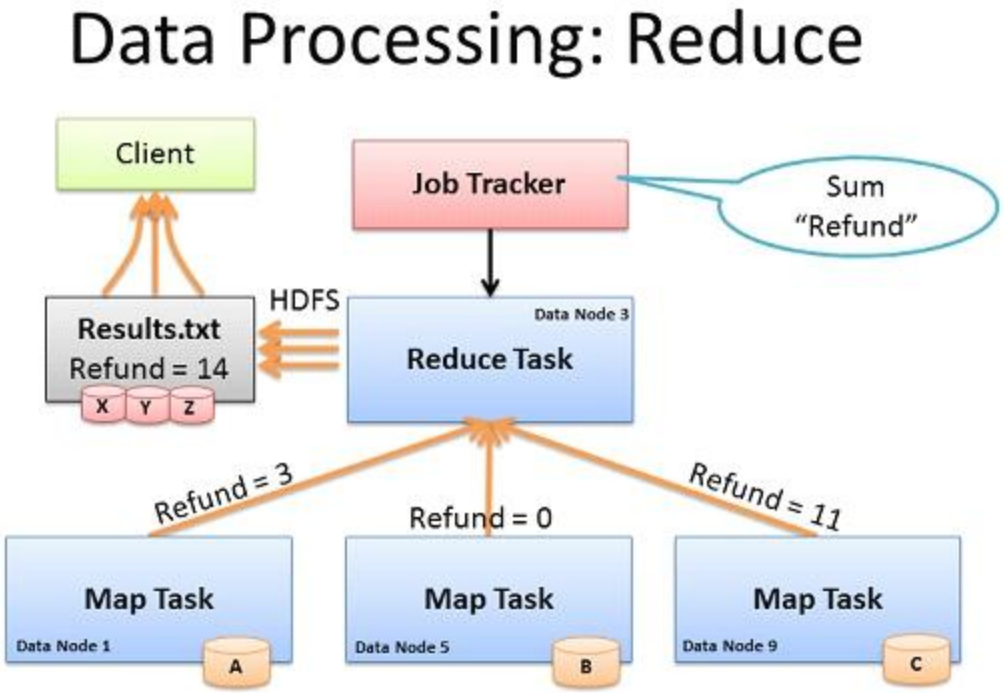
- 计算框架Mapper



- **Map:** "Run this computation on your local data"
- Job Tracker delivers Java code to Nodes with local data

- Hadoop计算框架Reducer
  - 对map阶段的结果进行汇总。
  - Reducer的数目由mapred-site.xml配置文件里的项目mapred.reduce.tasks决定。缺省值为1，用户可以覆盖之

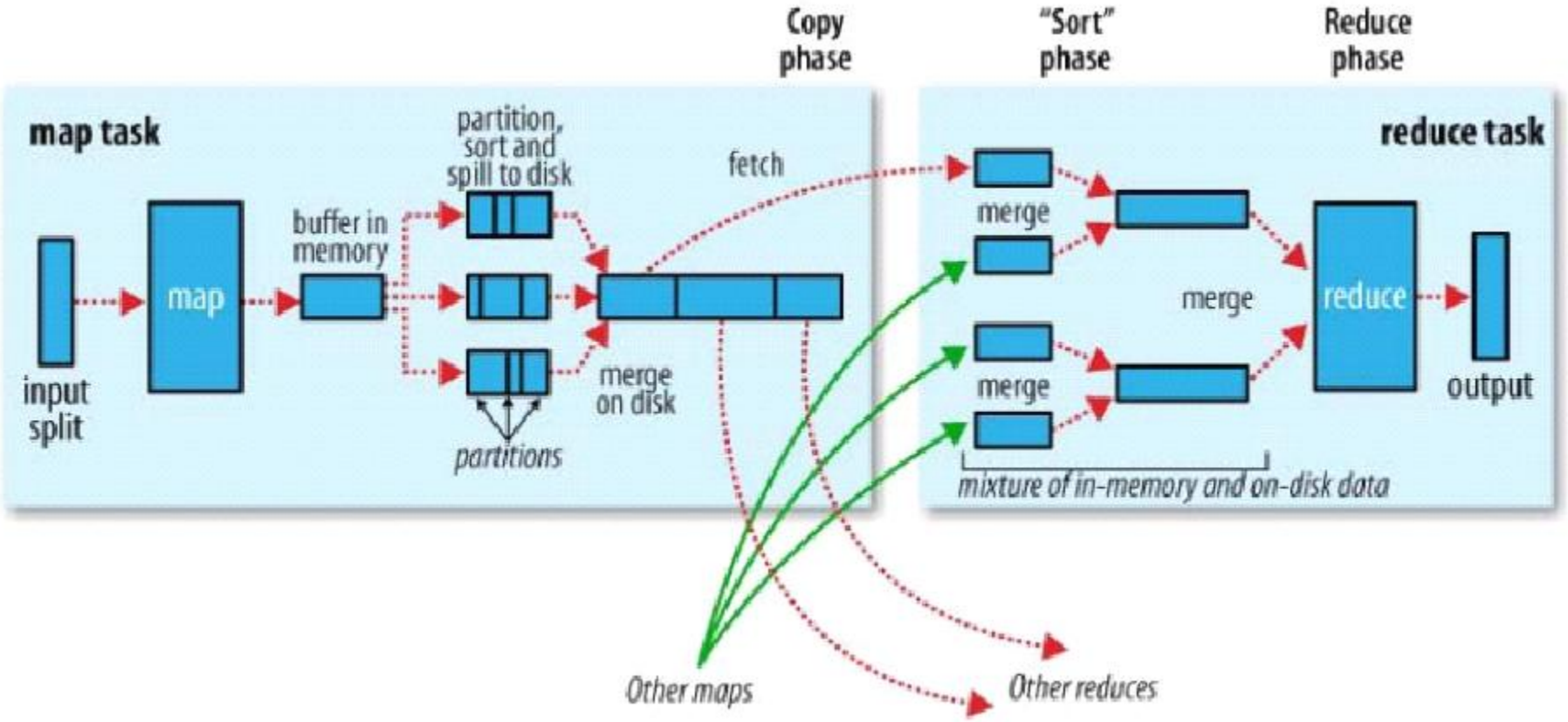
- Hadoop技术框架Reducer



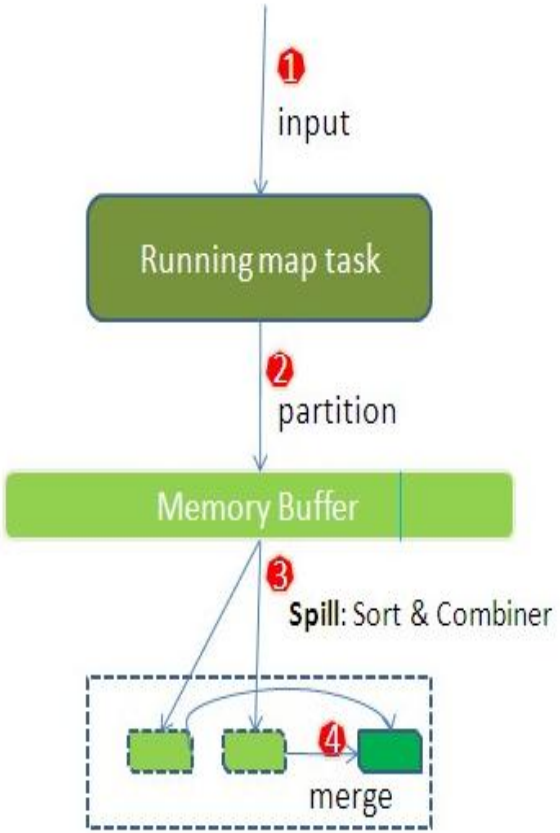
- **Reduce:** "Run this computation across Map results"
- Map Tasks send output data to Reducer over the network
- Reduce Task data output written to and read from HDFS

- Hadoop计算框架Shuffler
  - 在mapper和reducer中间的一个步骤
  - 可以把mapper的输出按照某种key值重新切分和组合成n份，把key值符合某种范围的输出送到特定的reducer那里去处理
  - 可以简化reducer过程

- Hadoop计算框架Shuffler



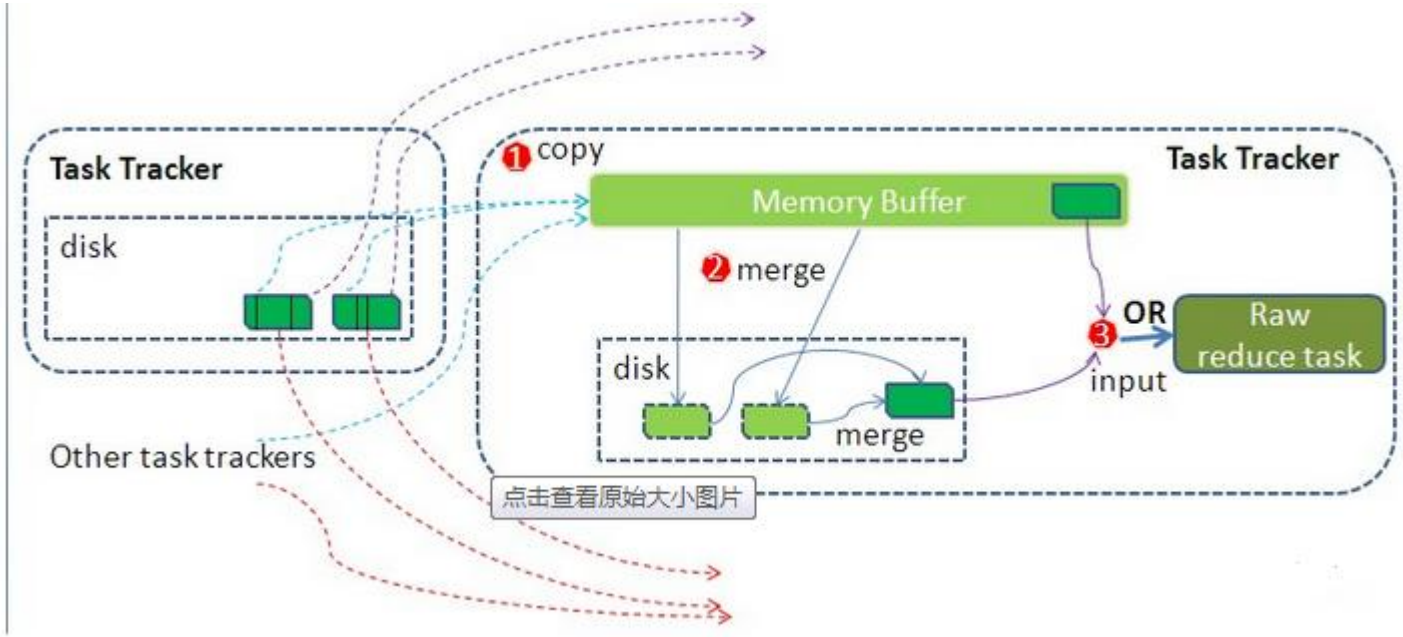
- Hadoop计算框架shuffle过程详解
  - 每个map task都有一个内存缓冲区（默认是100MB），存储着map的输出结果
  - 当缓冲区快满的时候需要将缓冲区的数据以一个临时文件的方式存放到磁盘（Spill
  - 溢写是由单独线程来完成，不影响往缓冲区写map结果的线程（spill.percent，默认是0.8）
  - 当溢写线程启动后，需要对这80MB空间内的key做排序(Sort)



- Hadoop计算框架shuffle过程详解

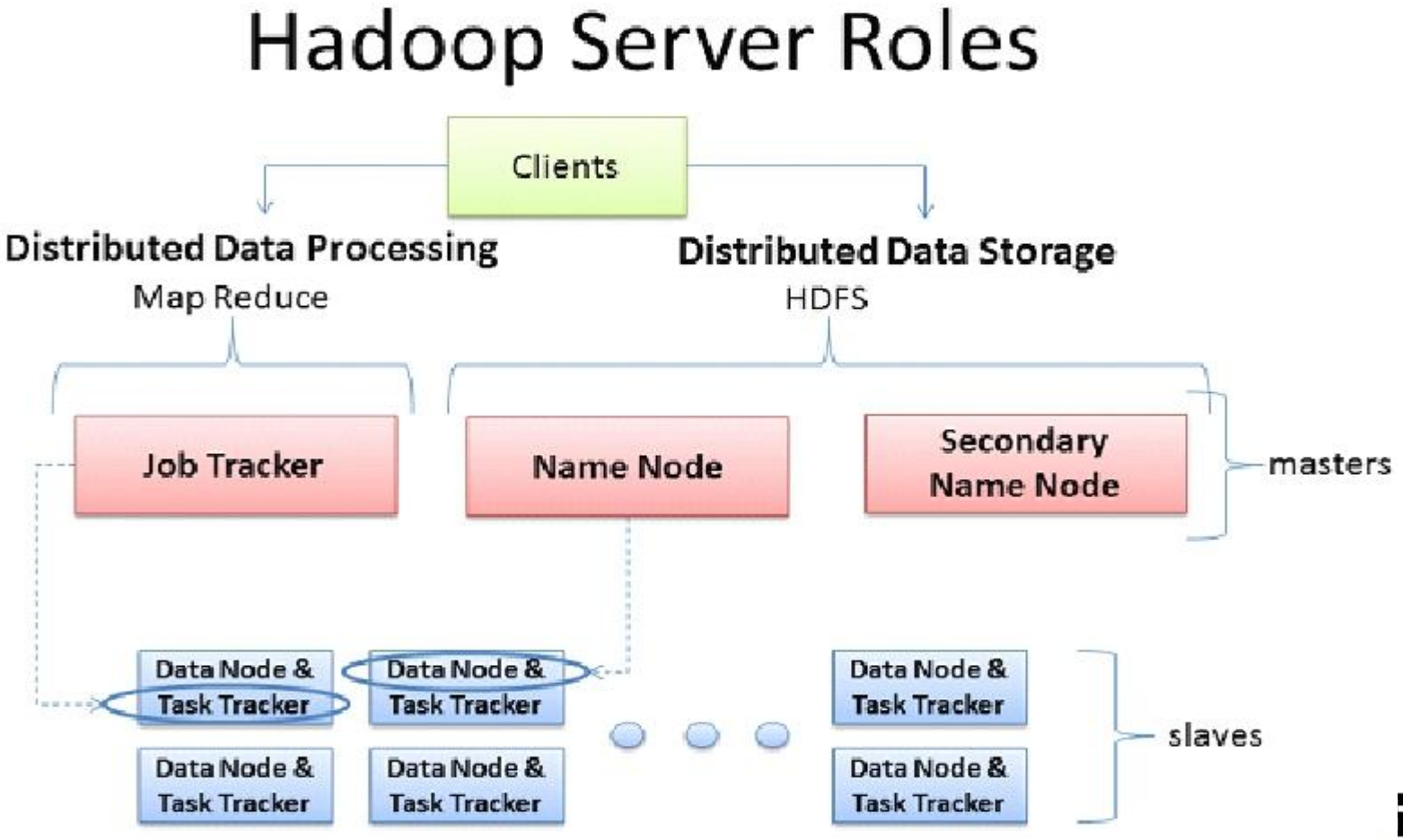
  - 假如client设置过Combiner，那么现在就是使用Combiner的时候了。将有相同key的key/value对的value加起来，减少溢写到磁盘的数据量。（reduce1，word1，[8]）。
  - 当整个map task结束后再对磁盘中这个map task产生的所有临时文件做合并（Merge），对于“word1”就是像这样的：{“word1”, [5, 8, 2, …]},假如有Combiner,{word1 [15]}，最终产生一个文件。
  - reduce 从tasktracker copy数据
  - copy过来的数据会先放入内存缓冲区中，这里的缓冲区大小要比map端的更为灵活，它基于JVM的heap size设置
  - merge有三种形式：1)内存到内存 2)内存到磁盘 3)磁盘到磁盘。merge从不同tasktracker上拿到的数据，{word1 [15，17，2]}
  - 参考博客http://langyu.iteye.com/blog/992916?page=3#comments

Hadoop计算框架shuffle过程详解



- MapReduce的 Split大小
  - max.split(100M)
  - min.split(10M)
  - block(64M)
  - max(min.split,min(max.split,block))

- MapReduce的架构



- MapReduce的架构
  - 一主多从架构
  - 主 JobTracker:
    - 负责调度分配每一个子任务task运行于TaskTracker上，如果发现有失败的task就重新分配其任务到其他节点。每一个hadoop集群中只一个 JobTracker,一般它运行在Master节点上。
  - 从TaskTracker:
    - TaskTracker主动与JobTracker通信，接收作业，并负责直接执行每一个任务，为了减少网络带宽TaskTracker最好运行在HDFS的DataNode上

- MapReduce安装
  - Mapred-size.xml

mapred.job.tracker	主机名和端口	local	jobtracker 的 RPC 服务器运行的主机名和端口。如果设置为默认值 local，则当运行一道作业时，jobtracker 也在同一进程内(此时，不必启动 MapReduce 守护进程)
mapred.local.dir	用逗号分隔的目录名称	${hadoop.tmp.dir}/mapred/local	MapReduce 存储作业中间数据的目录列表。当作业结束时数据会被清空
mapred.system.dir	URI	${hadoop.tmp.dir}/mapred/system	当一道作业运行时，与存储共享文件的 fs.default.name 相关的目录
mapred.tasktracker.map.tasks.maximum	int	2	任一时刻 tasktracker 上运行 map 任务的数量
mapred.tasktracker.reduce.tasks.maximum	int	2	任一时刻在 tasktracker 上运行的 reduce 任务的数量
Mapred.child.java.opts	String	-Xmx200m	用来启动 tasktracker 子进程，运行 map 和 reduce 任务的 JVM 选项。此属性可以在每道作业上设置，这对为 debug 设置 JVM 属性比较有用

- MR安装-----Hadoop配置有关文件

表 9-1：Hadoop 配置文件

文件名	格式	描述
hadoop-env.sh	bash 脚本	在运行 Hadoop 的脚本中使用的环境变量
core-site.xml	Hadoop 配置 XML	Hadoop 核心[1]的配置，例如 HDFS 和 MapReduce 中很普遍的 I/O 设置
hdfs-site.xml	Hadoop 配置 XML	HDFS 后台程序设置的配置：名称节点，第二名称节点和数据节点
mapred-site.xml	Hadoop 配置 XML	MapReduce 后台程序设置的配置：jobtracker 和 tasktracker
masters	纯文本	记录运行第二名称节点的机器(一行一个)的列表
slaves	纯文本	记录运行数据节点和 tasktracker 的机器(一行一个)的列表
hadoop-metrics.properties	Java 属性	控制 Hadoop 怎么发布 metrics(参见第 10 章)的属性
log4j.properties	Java 属性	系统日志文件的属性、名称节点审计日记和 tasktracker 子进程(参见第 5 章)的日志的属性