计算模式：分散->汇总模式

每个服务器分配到部分数据，把各自处理的结果发送给一台服务器，由其进行汇总处理

中心调度->步骤执行模式

由一个节点作为中心调度管理者，

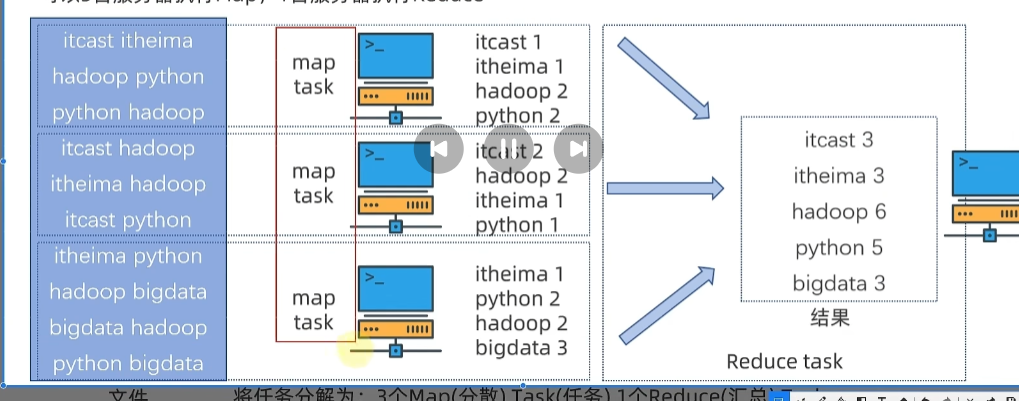
将任务划分为几个步骤

管理者安排每个机器执行任务

第一阶段：各自处理数据：

第二阶段：交换数据

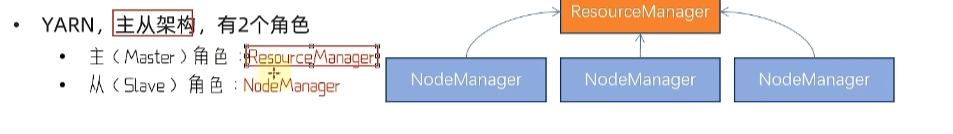
最终阶段，汇总结果



mapreduce是基于yarn运行的，没有yarn无法运行mapreduce程序

yarn管控整个分布式服务器集群的全部资源，整合进行统一调度，给mapreduce分派和管理运行资源，提高资源利用率

yarn核心架构



resourcemanager整个集群资源调度者，负责协调各个程序所需的资源

nodemanager单个服务器的资源调度者，负责调度单个服务器上的资源提供给应用程序使用

开辟的空间，称为容器，nodemanager预先占用这一部分资源，然后创建资源容器，程序运行在容器内，无法突破容器的资源限制

辅助角色：proxyserver代理服务器，保障网页浏览安全性

jobhistoryserver历史服务器，完成日志收集到hdfs文件系统。



提交mapreduce程序到yarn执行

hadoop jar 程序文件 java类名 【程序参数】

hadoop jar /export/server/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar wordcount hdfs://node1:8020/input hdfs://node1:8020/output/wc