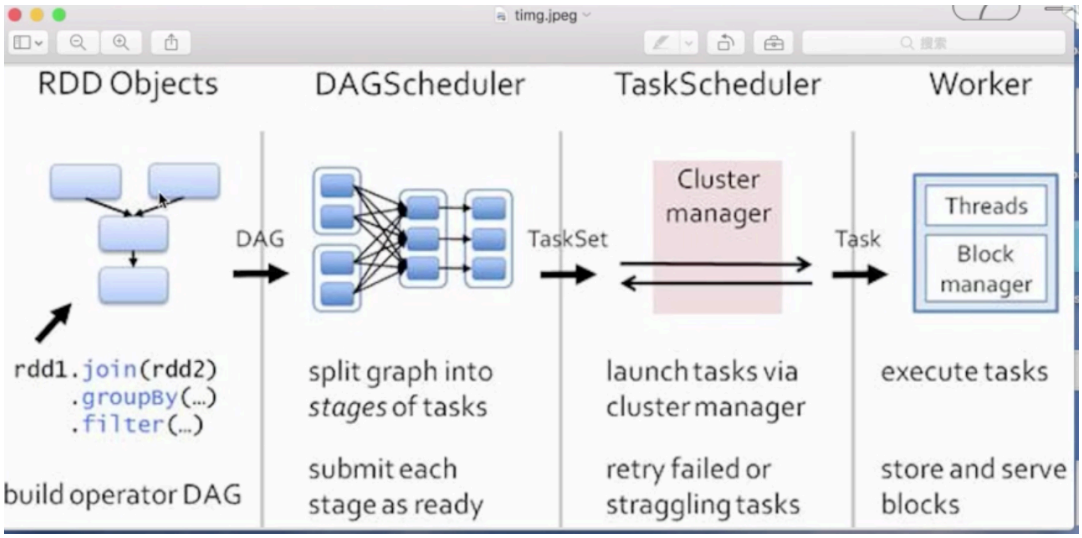
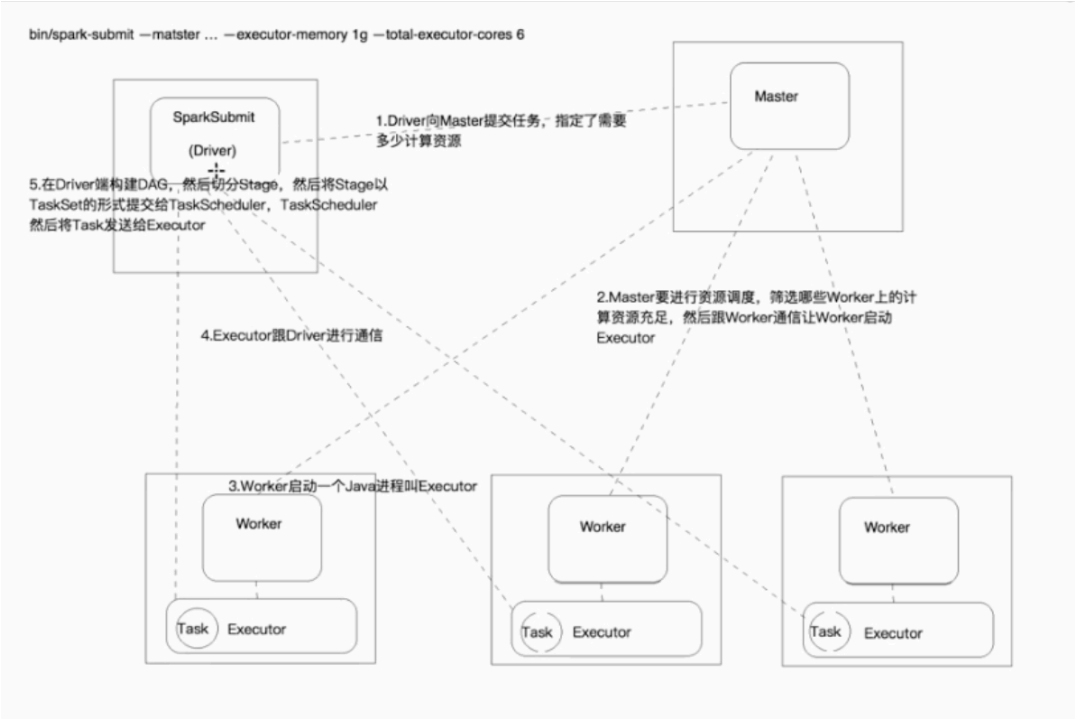
checkpoint 把中间结果 存入hdfs（大概类似于 还原点）
setChekpointDir.

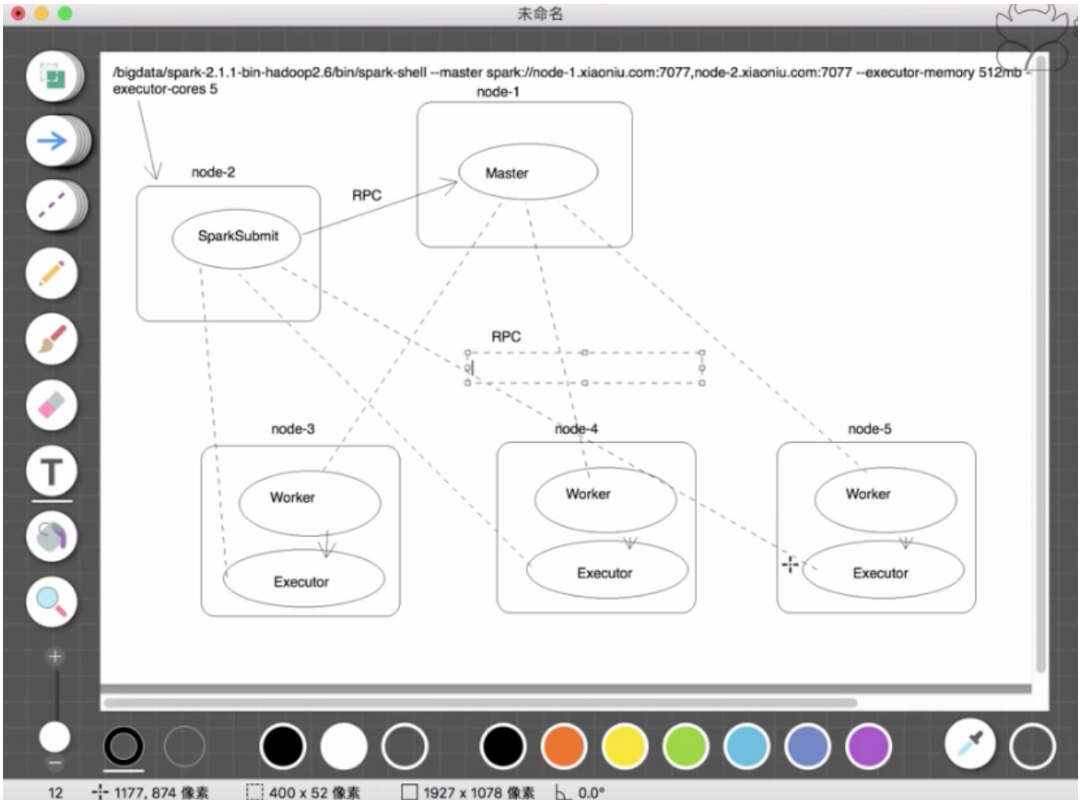


从RDD Object 到 TaskScheduler 都是在driver端
worker下的execute执行任务，被提交到具体的执行端

图二：



spark 对比yarn



跟yarn类比：

Master负责资源调度（决定在哪些Worker上启动executor）、监控Worker —> ResourceManager

Worker负责启动执行任务的进程（Executor），并且监控Executor，并且将当前机器的信息通过心跳汇报给Master —> NodeManager

Executor负责执行计算任务的 —> YarnChild

SparkSubmint负责向Master提交任务并申请资源，然后该任务下的Executor跟SparkSubmit进行通信，监控Executor —> AppMaster