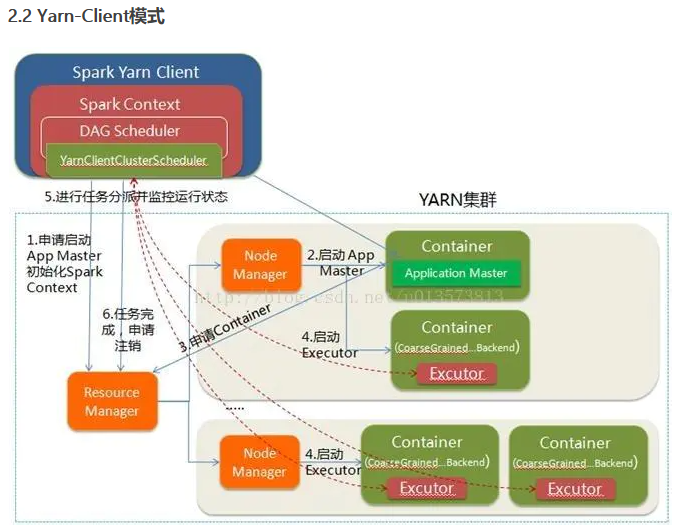
**spark任务的提交流程（yarn）**

spark一般都是部署到yarn上使用的，所以就说y问的最多的就是arn的提交流程，两种模式最大的区别就是driver端的执行位置

**Yarn Client模式**



第一步，Driver端在任务提交的本地机上运行

第二步，Driver启动之后就会和ResourceManager通讯，申请启动一个ApplicationMaster

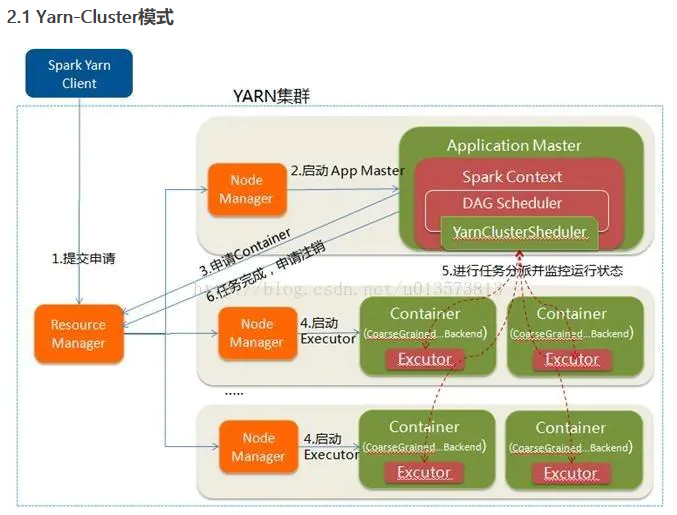
第三步，ResourceManager就会分配container容器，在合适的nodemanager上启动ApplicationMaster，负责向ResourceManager申请Executor内存

第四步，ResourceManager接到ApplicationMaster的资源申请后会分配container，然后ApplicationMaster在资源分配指定的NodeManager上启动Executor进程

第五步，Executor进程启动后会向Driver反向注册，Executor全部注册完成后Driver开始执行main函数

第六步，之后执行到Action算子时，触发一个Job，并根据宽依赖开始划分stage，每个stage生成对应的TaskSet，之后将task分发到各个Executor上执行。

**Yarn Cluster模式**



第一步，在YARN Cluster模式下，任务提交后会和ResourceManager通讯申请启动ApplicationMaster

第二步， 随后ResourceManager分配container，在合适的NodeManager上启动ApplicationMaster，此时的ApplicationMaster就是Driver。

第三步， Driver启动后向ResourceManager申请Executor内存，ResourceManager接到ApplicationMaster的资源申请后会分配container，然后在合适的NodeManager上启动Executor进程

第四步，Executor进程启动后会向Driver反向注册，Executor全部注册完成后Driver开始执行main函数，

第五步，之后执行到Action算子时，触发一个Job，并根据宽依赖开始划分stage，每个stage生成对应的TaskSet，之后将task分发到各个Executor上执行。