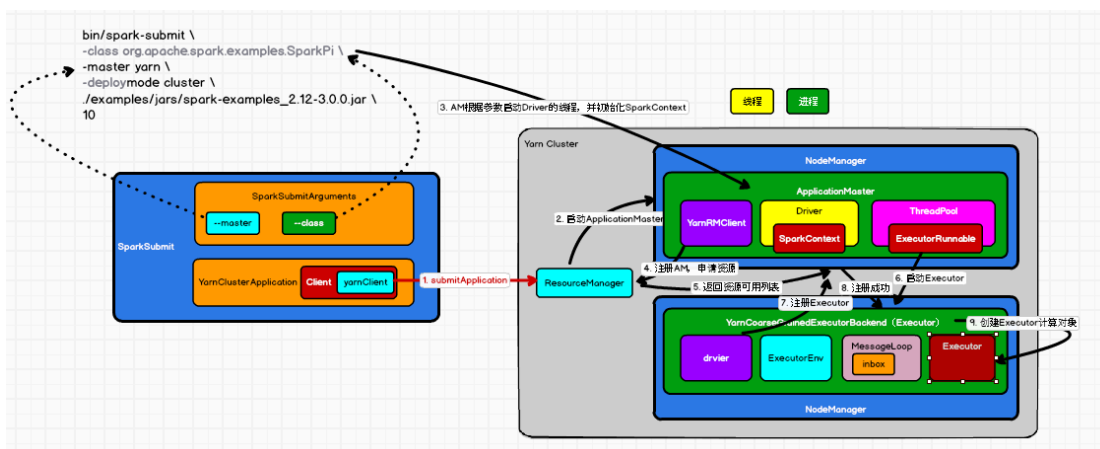


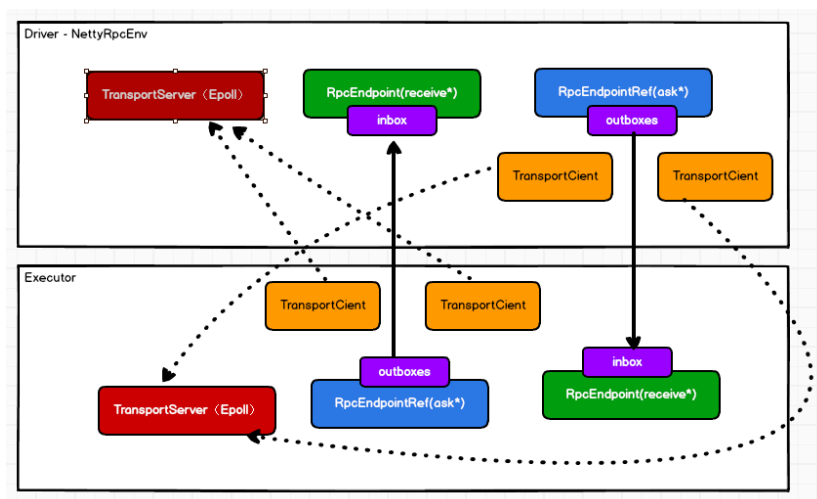
1. 环境准备 (Yarn 集群)

Driver , Executor



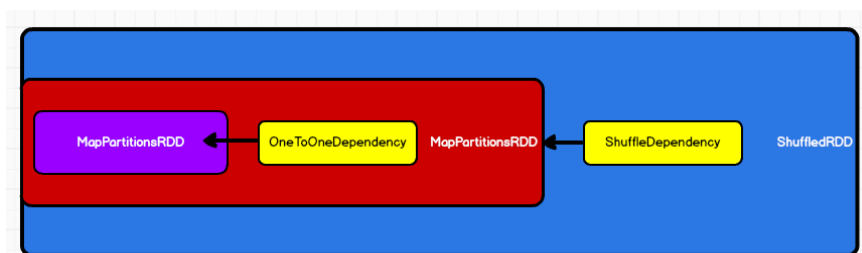
2. 组件通信 (Socket)

Driver => Executor

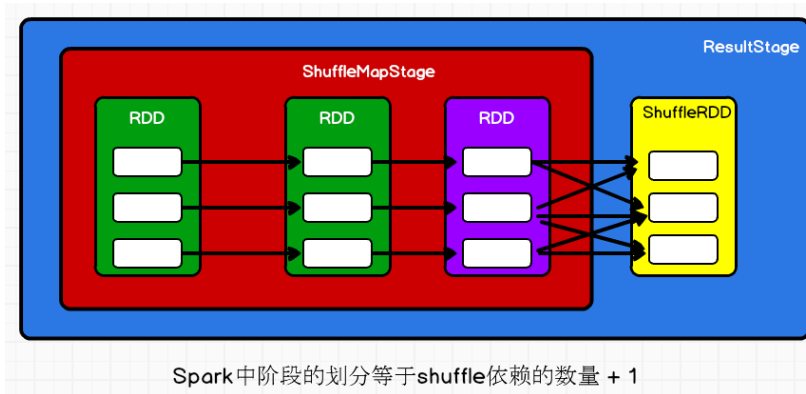


3. 应用程序的执行

(1) RDD 依赖

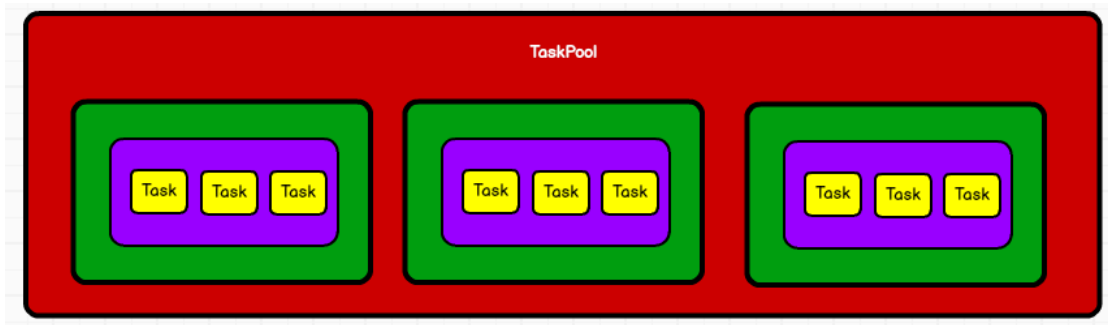


(2) 阶段的划分

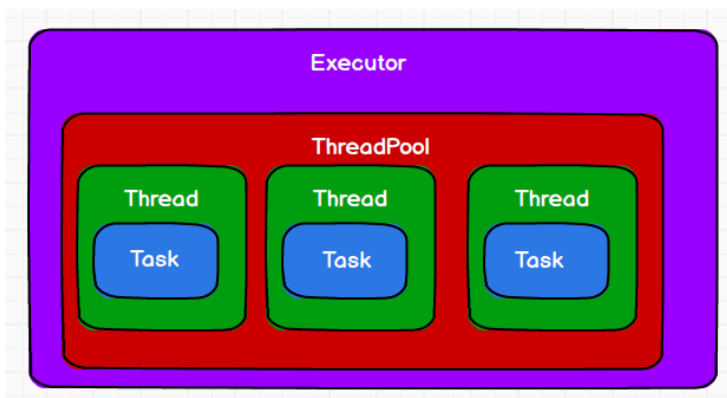


(3) 任务的切分

(4) 任务的调度



(5) 任务的执行



4. Shuffle

(1) Shuffle 的原理和执行过程

- shuffleWriterProcessor (写处理器)
- ShuffleManager: Hash(早期) & Sort (当前)

处理器	写对象	判断条件
SerializedShuffleHandle	UnsafeShuffleWriter	1. 序列化规则支持重定位操作（java 序列化不支持，KRYO 支持） 2. 不能使用预聚合 3. 如果下游的分区数量小于或等于 16777216
BypassMergeSortShuffleHandle	BypassMergeSortShuffleWriter	1. 不能使用预聚合 2. 如果下游的分区数量小于等于 200（可配）
BaseShuffleHandle	SortShuffleWriter	其他情况

(2) Shuffle 写磁盘

(3) Shuffle 读取磁盘

5. 内存的管理

(1) 内存的分类

(2) 内存的配置