Native YARN Framework layer for services and beyond

——YARN-5079

# 1.概况，核心JIRA

当前YARN中仅提供low-level API，给API可以支持任意类型的应用，类似于MapReduce

Apache Tez,Apache Spark,Apache REEF,Apache Twill及Apache Helix及其他框架为用户或者程序开发者提供higher level APIs以构建自己的应用。在services sider，Apache Slider实现的功能与上述框架类似。

YARN-5079（源于Apache Slider）为YARN提供higher level APIs来创建应用，简化Docker集群的创建及维护成本，其执行流程图如下所示：



执行步骤如下所示：

1. SliderClient，向RM提交创建Docker应用的请求，请求中包括应用的参数和定义的组件
2. RM接收到请求后，分配AM Container，与NM交互
3. 分配到Container的NM，在Container启动SliderAppMaster
4. SliderAppMaster启动后，接管Docker 应用的启动和管理，首先根据组件的定义参数，向RM请求需要的资源，然后通过DockerProviderService，从应用定义中初始化docker container的执行脚本，最后向NM发送Container launch请求
5. NM接收到RM的Docker Launch的请求后，调用LCE中DockerLinuxContainerRuntime，通过Docker Daemon启动Docker Container

YARN-5461本质上是将Apache Slider的源代码打入yarn-application模块，同时增加slider的脚步到yarn的执行文件中，涉及到slider-core模块的源码不再分析。

# 2.RESOLVED JIRA

**YARN-5505**, Slider AM可配置模块provider，Apache Slider中实现的ProviderService需要实现Slider Agent(python模块)。该JIRA实现了docker provider(DockerProviderService)，不使用agent，直接从NM获取信息及执行命令，如下所示：



在Slider Core模块中增加了DockerProviderService，根据参数—provider docker来选择使用，当前仅支持这种方式，默认是agent（直接启动进程，目前不支持）。

**YARN-5079** <=> SLIDER-875，为了支持app-of-apps，将SLIDER-875的patch打入，如下所示：



YARN-5735，创建app时设置lifetime



**YARN-5909**，删除了Slider AM中所有的agent相关代码，删除的类：

AgentRestOperations，接受Slider Agent心跳的REST

AgentClientProvider

AgentLaunchParameter

AgentProviderFactory

AgentProviderService

AgentUtils

Command | CommandResult | ComponentCommandOrder |CommandReport

ComponentInstanceState

ComponentTagProvider

ContainerState

HeartbeatMonitor，接受Agent心跳

该JIRA后，yarn native services仅支持Docker应用

**YARN-5944**，当RM宕机后，native services AM重复连接am知道max-wait-ms超时，直到AM停止工作。但是对于long running services，需要保证RM宕机后AM不停服，如下所示：



**YARN-5079**，当前已经使用docker provider，发布quicklinks export的发生也要变化，如下所示：



**YARN-4793|YARN-6192**，为Slider提供一个REST Server，与Slider Router作用相同，具体代码不再分析

# 3.UNRESOLVED JIRA

**YARN-6023**, ProviderUtils.updateServiceRecord调用时仅在ServiceRecord中设置一个IP（yarn：ip），在YARN-4757(YARN-DNS)中需要读取yarn:ip然后创建DNS 记录：

**YARN-6161**，当前YARN Native Service中没有agent机制，运行docker containers不需要端口分配，但是对于非docker-less及agent-less需要使用随机分配的端口。

**YARN-6167**，当NM失去与RM的连接后，Schedulder会kill掉该NM的运行containers。但是对于某个应用来说，NM仅是临时丢失。Kill containers的决策应该交由AM来完成。

**YARN-6168**，当RM重启导致竞争关系。当RM和AM都宕机，NMs和app Containers仍然在运行。当重启RM后，会重启AM并注册给RM，在NM reported所有正在运行的containers之前，AM将新启动的containers汇报给RM。AM仅被告知部分已有的containers。

**YARN-6185**，将SLIDER-1199打入yarn-native-service，来支持blacklisting nodes。当某个role在一个节点失败次数超过阈值，这个节点不再放置该role。但是不可避免RM重新选择node。如果将这个节点放入blacklist中，可以解决这个问题，如下所示：

