

# 一页纸精华 | YARN

原创 2016-02-01 牛家浩 中兴大数据

>>>> 这是 **中兴大数据** 第207篇原创文章

要入门大数据，最好的办法就是理清Hadoop的生态系统。中兴大数据公众号将推出“一页纸精华”栏目，将用最精炼的语言，陆续为你介绍Hadoop生态系统的各个组件。本期为你介绍Hadoop统一资源管理框架YARN。

YARN ( Yet Another Resource Negotiator ) 是一个通用的资源管理平台，可为各类计算框架提供资源的管理和调度。

YARN可以将多种计算框架(如离线处理MapReduce、在线处理的Storm、迭代式计算框架Spark、流式处理框架S4等) 部署到一个公共集群中，共享集群的资源。并提供如下功能：

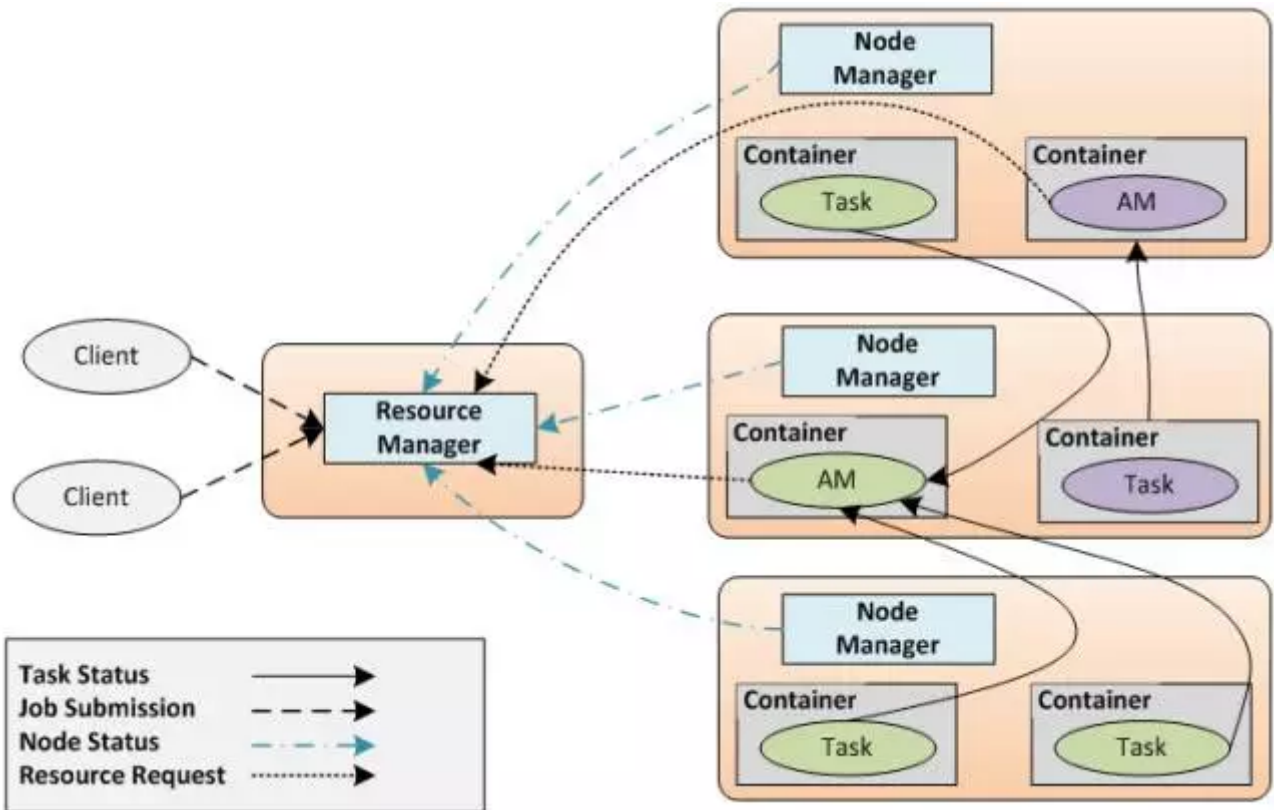
- **资源的统一管理和调度**：集群中所有节点的资源(内存、CPU、磁盘、网络等)抽象为Container。计算框架需要资源进行运算任务时需要向YARN申请Container，YARN按照特定的策略对资源进行调度进行Container的分配。
- **资源隔离**：YARN使用了轻量级资源隔离机制Cgroups进行资源隔离以避免相互干扰，一旦Container使用的资源量超过事先定义的上限值，就将其杀死。

YARN是对Mapreduce V1重构得到的。其设计思想类似于单机程序中多线程并行编程在分布式中的扩展。

YARN可以看成是一个云操作系统，由一个ResourceManager和多个NodeManager组成，它负责管理所有NodeManger上多维度资源，并以Container(启动一个Container相当于启动一个进程)方式分配给应用程序启动ApplicationMaster(相当于主进程中运行逻辑) 或运行ApplicationMaster切分的各Task(相当于子进程中运行逻辑)。

## YARN体系架构

YARN架构如下图所示：



YARN总体上是Master/Slave结构，主要由ResourceManager、NodeManager、ApplicationMaster和Container等几个组件构成。

- **ResourceManager(RM)**：负责对各NM上的资源进行统一管理和调度。将AM分配空闲的Container运行并监控其运行状态。对AM申请的资源请求分配相应的空闲Container。主要由两个组件构成：调度器和应用程序管理器：
  - i. 调度器(Scheduler)：调度器根据容量、队列等限制条件（如每个队列分配一定的资源，最多执行一定数量的作业等），将系统中的资源分配给各个正在运行的应用程序。调度器仅根据各个应用程序的资源需求进行资源分配，而资源分配单位是Container，从而限定每个任务使用的资源量。
  - ii. 应用程序管理器(Applications Manager)：应用程序管理器负责管理整个系统中所有应用程序，包括应用程序提交、与调度器协商资源以启动AM、监控AM运行状态并在失败时重新启动等。
- **NodeManager (NM)**：NM是每个节点上的资源和任务管理器。它会定时地向RM汇报本节点上的资源使用情况和各个Container的运行状态；同时会接收并处理来自AM的Container

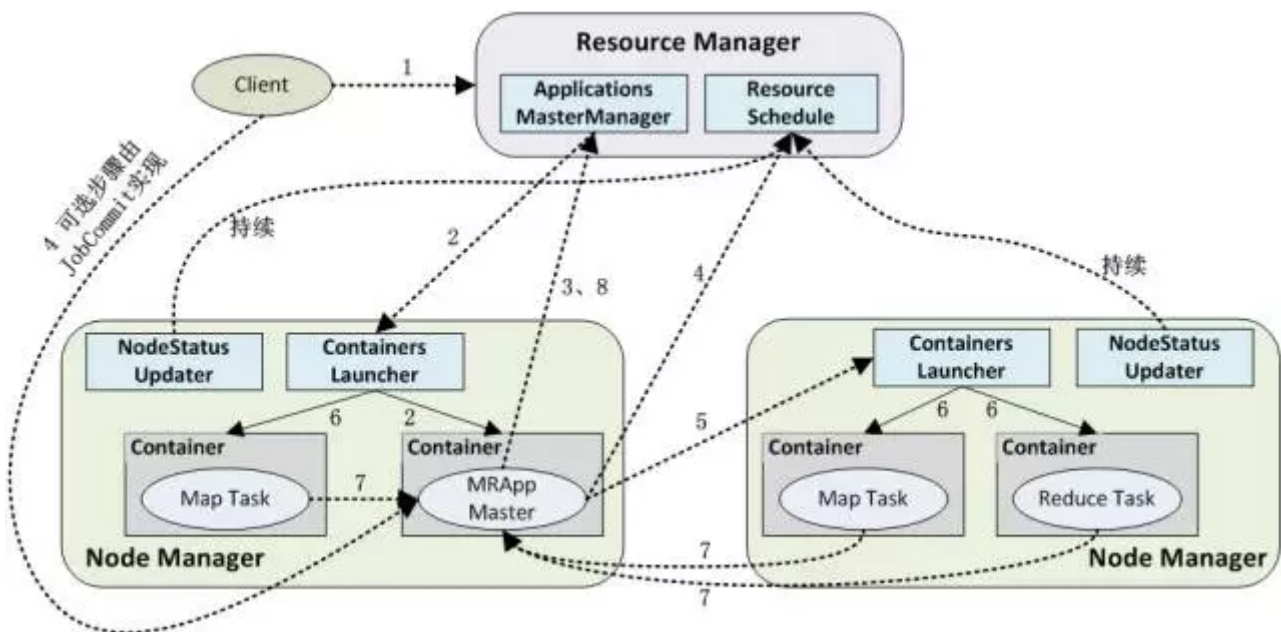
启动/停止等请求。

- **ApplicationMaster (AM)**：用户提交的应用程序均包含一个AM，负责应用的监控，跟踪应用执行状态，重启失败任务等
- **Container**：Container是YARN中的资源抽象，它封装了某个节点上的多维度资源，如内存、CPU、磁盘、网络等，当AM向RM申请资源时，RM为AM返回的资源便是用Container表示的。YARN会为每个任务分配一个Container且该任务只能使用该Container中描述的资源。

## YARN应用工作流程

如下图所示用户向YARN中提交一个应用程序后，YARN将分两个阶段运行该应用程序：

- 启动AM，如下步骤1~3；
- 由AM创建应用程序为它申请资源并监控它的整个运行过程，直到运行完成，如下步骤4~7。



YARN应用工作流程图

1. 用户向YARN中提交应用程序，其中包括AM程序、启动AM的命令、用户程序等；
2. RM为该应用程序分配第一个Container，并与对应的NM通信，要求它在这个Container中启动AM；

3. AM首先向RM注册，这样用户可以直接通过RM查看应用程序的运行状态，然后它将为各个任务申请资源，并监控它的运行状态，直到运行结束，即重复步骤4~7；
4. AM采用轮询的方式通过RPC协议向RM申请和领取资源；
5. 一旦AM申请到资源后，便与对应的NM通信，要求它启动任务；
6. NM为任务设置好运行环境(包括环境变量、JAR包、二进制程序等)后，将任务启动命令写到一个脚本中，并通过运行该脚本启动任务；
7. 各个任务通过某个RPC协议向AM汇报自己的状态和进度，以让AM随时掌握各个任务的运行状态，从而可以在任务失败时重新启动任务；
8. 应用程序运行完成后，AM向RM注销并关闭自己。

## YARN资源调度模型

YARN提供了一个资源管理平台能够将集群中的资源统一进行管理。所有节点上的多维度资源都会根据申请抽象为一个个Container。

YARN采用了双层资源调度模型：

- RM中的资源调度器将资源分配给各个AM：资源分配过程是异步的。资源调度器将资源分配给一个应用程序后，它不会立刻push给对应的AM，而是暂时放到一个缓冲区中，等待AM通过周期性的心跳主动来取；
- AM领取到资源后再进一步分配给它内部的各个任务：不属于YARN平台的范畴，由用户自行实现。

YARN目前采用的资源分配算法有三种。但真实的调度器实现中还对算法做了一定程度的优化。

1. 先来先调度FIFO：先按照优先级高低调度，如优先级相同则按照提交时间先后顺序调度，如提交时间相同则按照队列名或ApplicationID比较顺序调度。
2. 公平调度FAIR：该算法的思想是尽可能地公平调度，即已分配资源量少的优先级高。
3. 主资源公平调度DRF：该算法扩展了最大最小公平算法使之能够支持多维资源，算法是配置资源百分比小的优先级高。



**中兴通讯大数据团队广招贤才啦！**

## 招聘信息

**招聘岗位：** WEB软件开发工程师

**任职要求：**

1. 全日制重点本科及以上学历，硕士两年以上的工作经验，本科三年以上相关经验。
2. 熟练掌握JAVA语言。
3. 熟练掌握WEB开发技术（HTML/CSS/JAVASCRIPT等），了解较前沿的WEB技术（如HTML5等）。
4. 能够使用主流的WEB框架和组件（Struts,Ajax,jQuery,YUI等）进行设计和开发。
5. 良好的沟通协作能力，能阅读一般的英文技术文章。
6. 有WEB应用项目、中大型网站开发经验者优先。

**主要职责：**

1. 负责智慧城市等政企应用软件的开发、设计工作。
2. 参与和用户的沟通、交流等工作。

**简历请投至：** xie.haibo@zte.com.cn



喜欢这篇文章？向“中兴大数据”要更多



↑ 扫描指纹识别二维码 ↑