

15-yarn的原理

-----成都尚学堂-mr-zeng-----

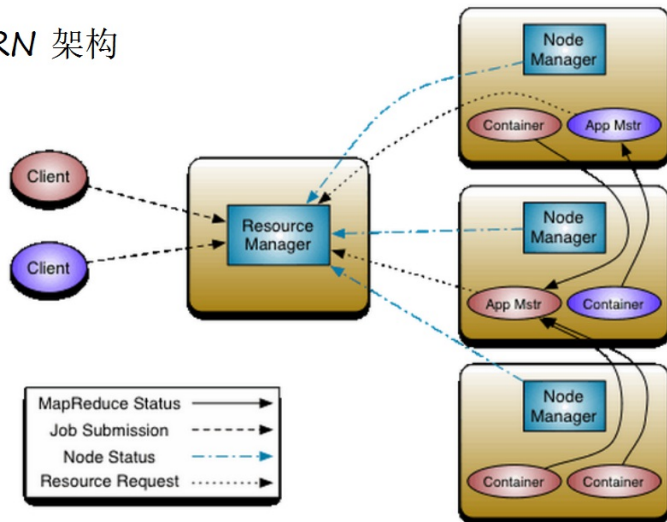
hadoopYARN 概念

YARN是Hadoop 2.0中的资源管理系统，它的基本设计思想是将Hadoop 1.0中的JobTracker拆分成了两个独立的服务：一个全局的资源管理器ResourceManager和每个应用程序特有的ApplicationMaster。

ResourceManager：负责整个系统的资源管理和分配，nodemanager报告资源。

ApplicationMaster：负责单个应用程序（job）的管理，把job分为多个任务交给（container）

YARN 架构



各进程作用

1.ResourceManager (RM)

RM是一个全局的资源管理器，负责整个系统的资源管理和分配。它主要由两个组件构成：调度器（Scheduler）和应用程序管理器（Applications Manager，ASM）。

(1) 调度器

调度器将系统中的资源分配给各个正在运行的应用程序。

(2) 应用程序管理器

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ApplicationMaster、监控ApplicationMaster运行状态并在失败时重新启动它等。

2. ApplicationMaster (AM) :用户提交的每个应用程序job均包含1个ApplicationMaster ,

主要功能包括：

- a) 与RM调度器协商以获取资源（用Container表示）；
- b) 将得到的任务进一步分配给内部的任务；
- c) 与NM通信以启动/停止任务；
- d) 监控所有任务运行状态，并在任务运行失败时重新为任务申请资源以重启任务。

3. NodeManager (NM)

NM是每个节点上的资源和任务管理器，一方面，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；另一方面，它接收并处理来自AM的Container启动/停止等各种请求。

4. 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。

a) 启动yarn就有的进程

ResourceManager

NodeManager

b) 其他进程根据job任务执行-动态的产生