

Yarn 入门概述

Yarn 产生的背景

yarn产生之前，hadoop主要使用MRv1，在这里我们了解一下MRv1的一些局限性：

1. **扩展性差**：在MRv1中，JobTracker同时具备了资源管理和作业控制两个功能，在成为了系统的最大一个瓶颈，严重限制了Hadoop的集群扩展性。
2. **可靠性差**：MRv1采用了master/slave结构，容易出现单节点故障问题，即master节点出现了故障将会导致整个集群瘫痪不能使用
3. **资源利用低**：MRv1采用了基于槽位的资源分配模型，槽位是一种粗粒度的资源划分单位，通常一个任务不会用完槽位的所有资源，且其他任务不能够使用这些空闲资源。此外Hadoop将槽位划分为Map Slot和Reduce Slot两种，且不允许资源共享，通常会导致一个槽位资源紧张另外一个闲置。（比如当启动时候）
4. **无法支持多种计算框架**：MRv1不能满足内存计算框架，流式计算框架、迭代计算框架等并存。

Yarn概念

为了克服MRv1以上的几个缺点，MRv2的核心从mapreduce转到了资源管理系统Yarn，YARN 实际上是一个弹性计算平台，它的目标已经不再局限于支持 MapReduce 一种计算框架，而是朝着对多种框架进行统一管理的方向发展。相比于“一个集群一个计算框架”，共享集群的好处是：

1. **资源利用高**：共享集群模式可以通过多个框架共享资源，使得集群中的资源能够更加充分的利用。
2. **运维成本低**：只需要少量管理员即可完成多个框架的管理。
3. **数据共享**：不需要跨集群传输数据，不同集群可以同时使用共同的数据。

下图为Yarn轻量级弹性计算平台：

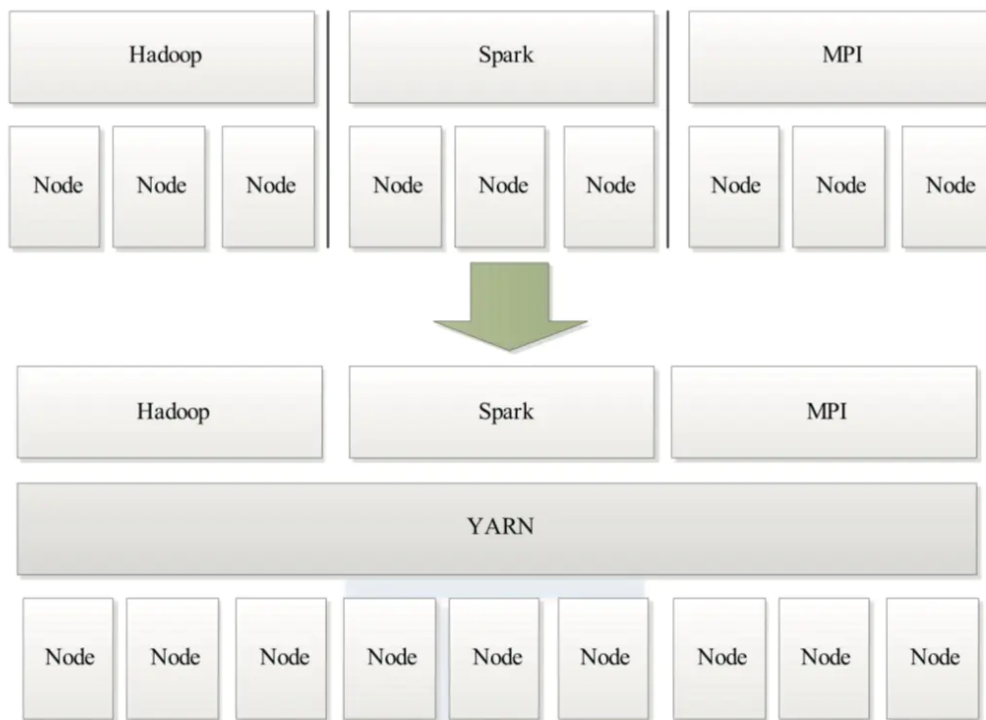


图 2-2 以 YARN 为核心的弹性计算平台的基本架构

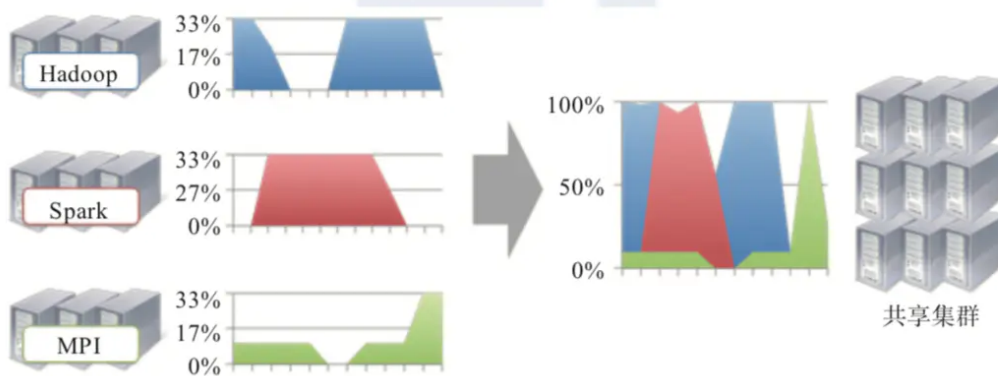


图 2-3 共享集群模式使得资源利用率提高