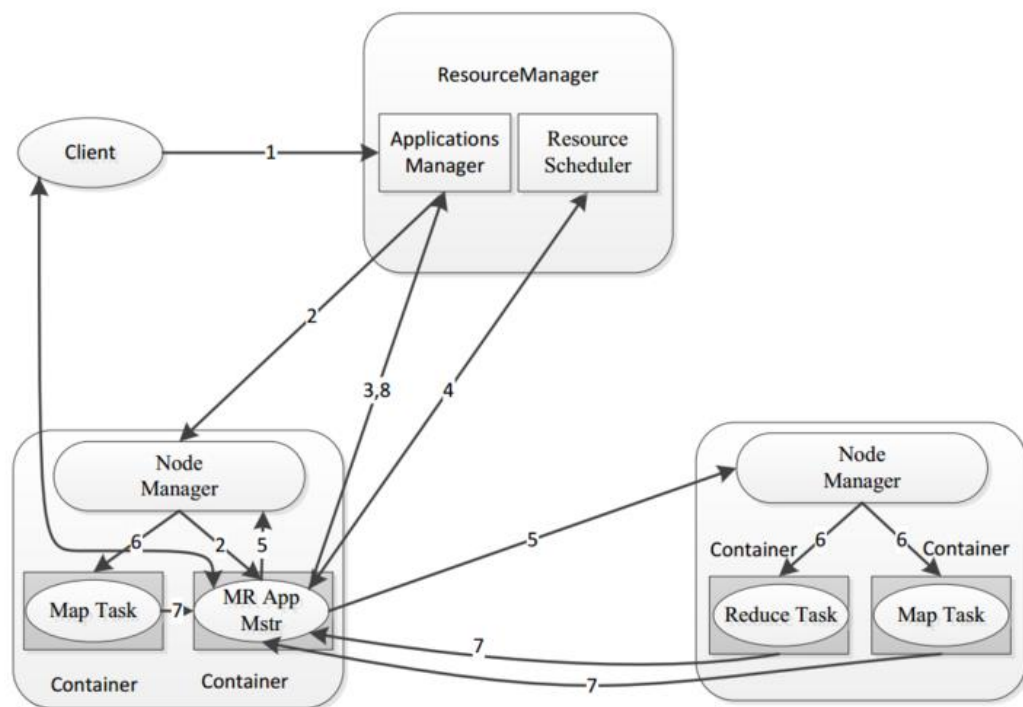


MapReduce 框架结构

一个完整的 mapreduce 程序在分布式运行时三类实例进程：

- 1、MRAppMaster：负责整个程序的过程调度及状态协调
- 2、MapTask：负责 map 阶段的整个数据处理流程
- 3、ReduceTask：负责 reduce 阶段的整个数据处理流程



图：MapReduce 2.0 运行流程图

- ◆ 一种分布式计算模型，解决海量数据的计算问题
- ◆ MapReduce将整个并行计算过程抽象到两个函数
 - Map(映射): 对一些独立元素组成的列表的每一个元素进行指定的操作，可以高度并行。
 - Reduce(化简): 对一个列表的元素进行合并。
- ◆ 一个简单的MapReduce程序只需要指定map()、reduce()、input和output，剩下的事由框架完成。

