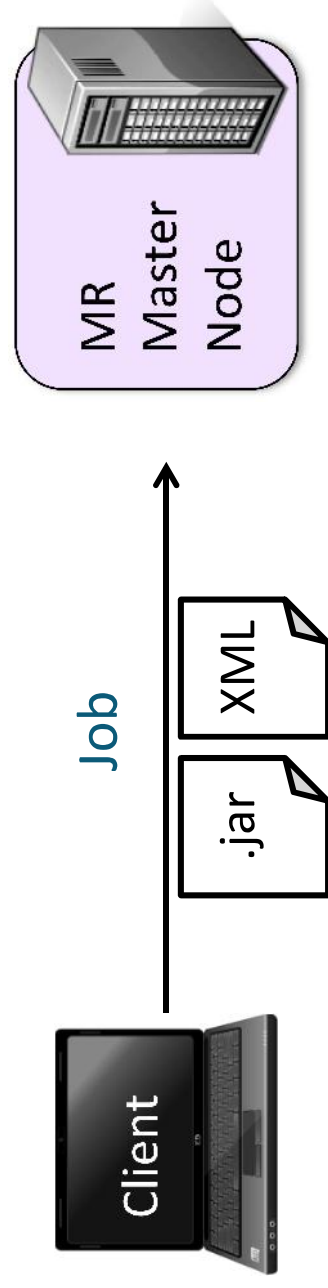


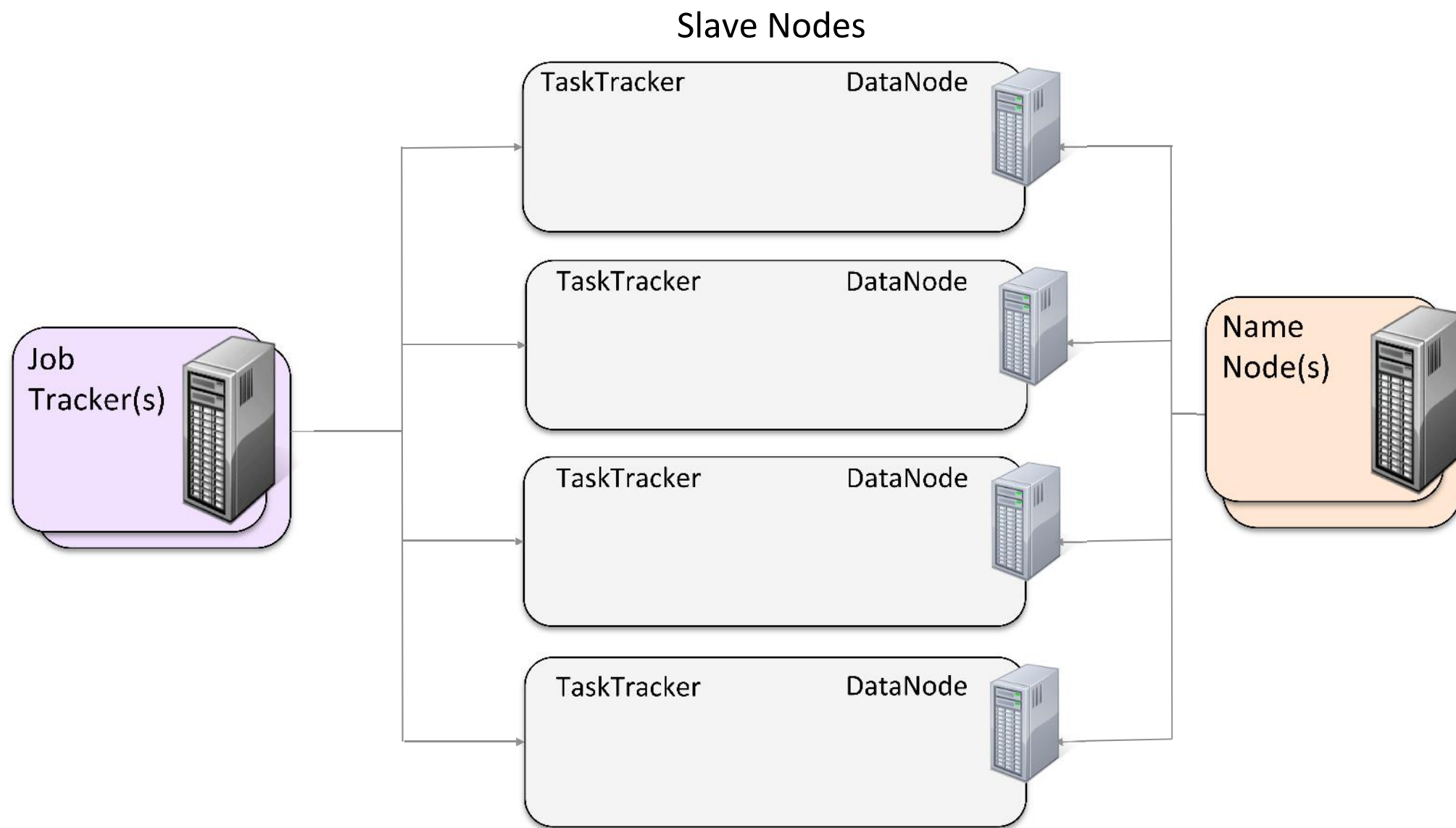
MapReduce Terminology

- **A *job* is a ‘full program’**
 - A complete execution of Mappers and Reducers over a dataset
- **A *task* is the execution of a single Mapper or Reducer over a slice of data**
- **A *task attempt* is a particular instance of an attempt to execute a task**
 - There will be at least as many task attempts as there are tasks
 - If a task attempt fails, another will be started by the JobTracker or ApplicationMaster
 - *Speculative execution* (covered later) can also result in more task attempts than completed tasks

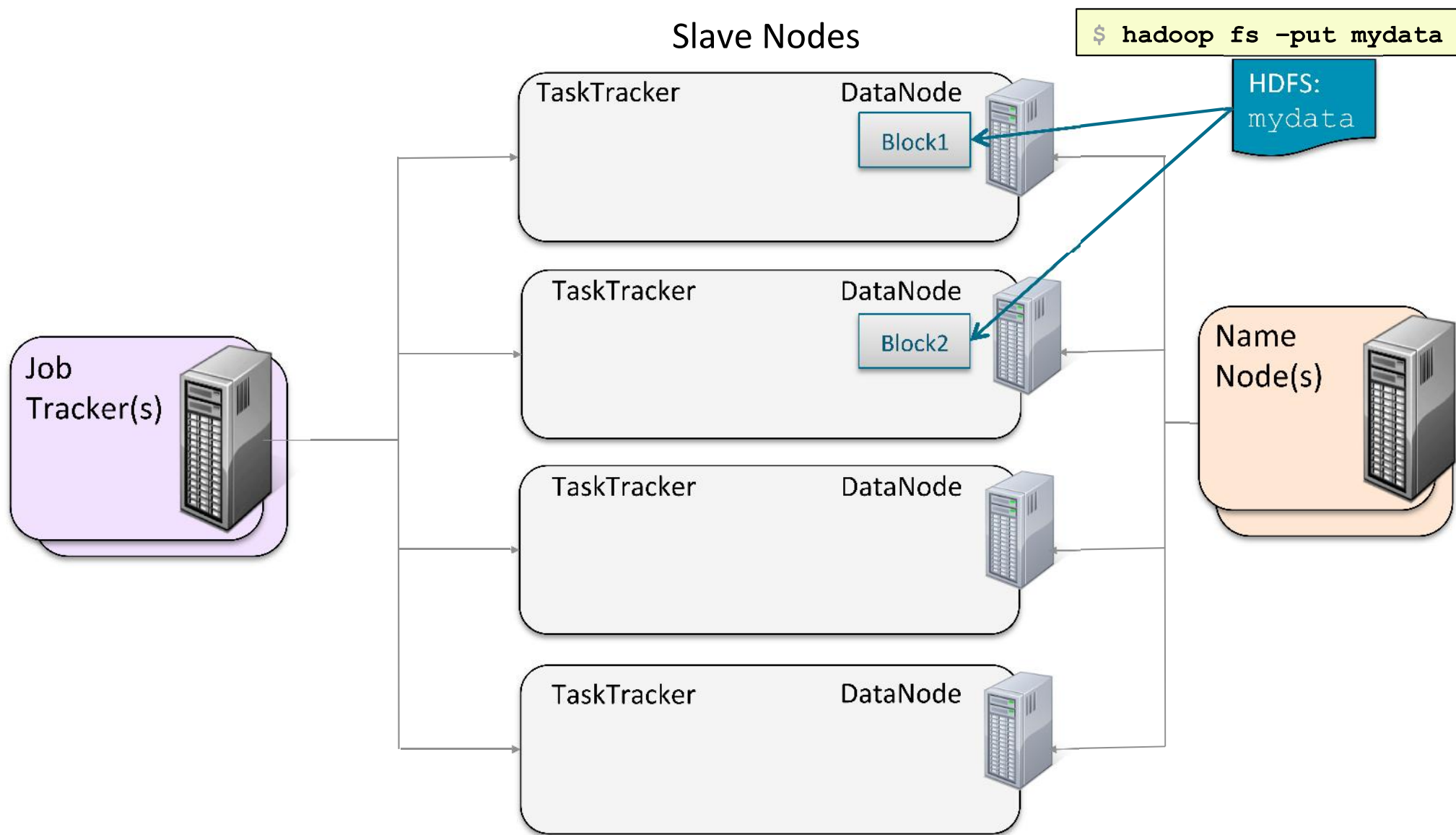
Submitting A Job



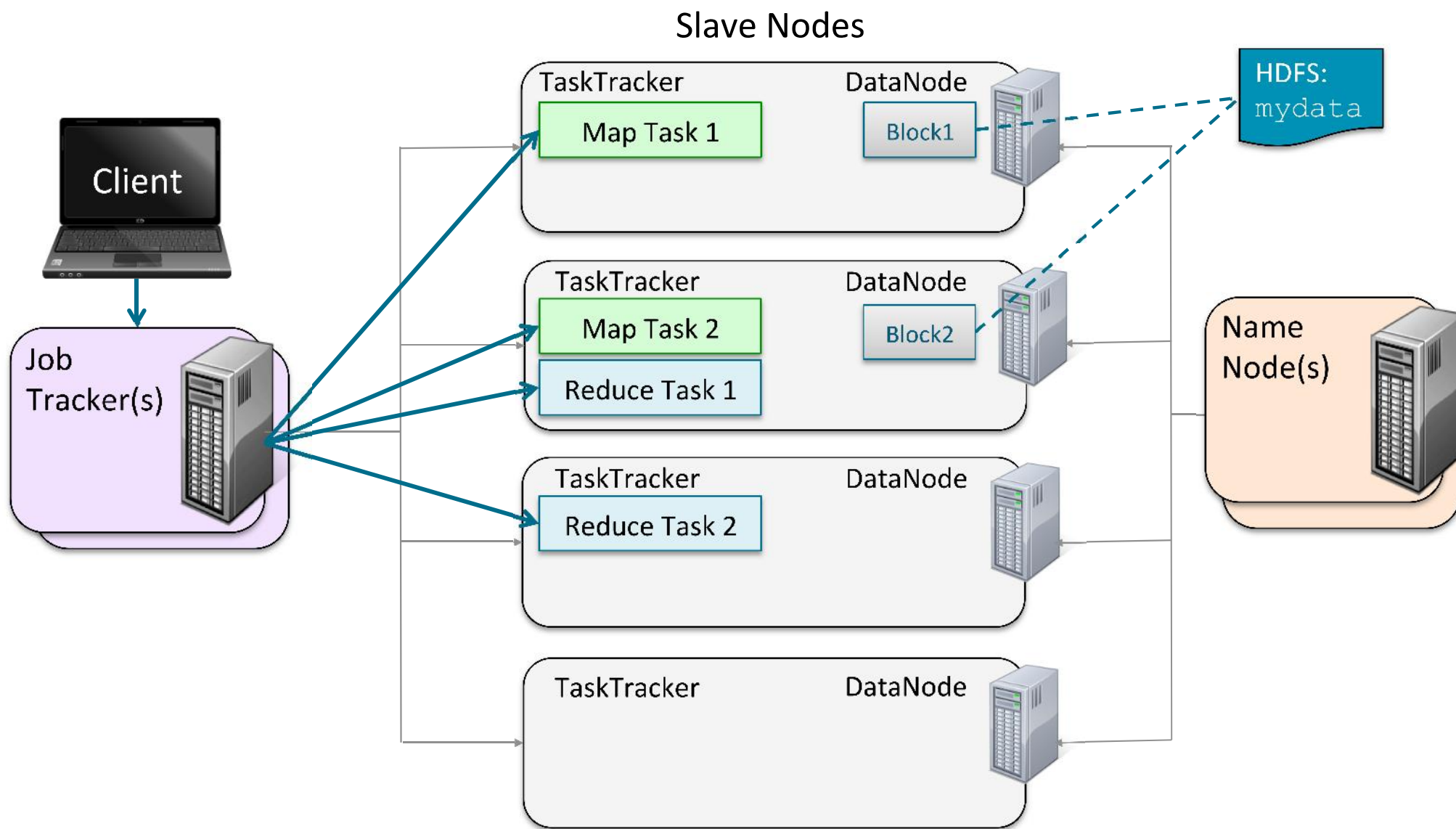
A MapReduce v1 Cluster



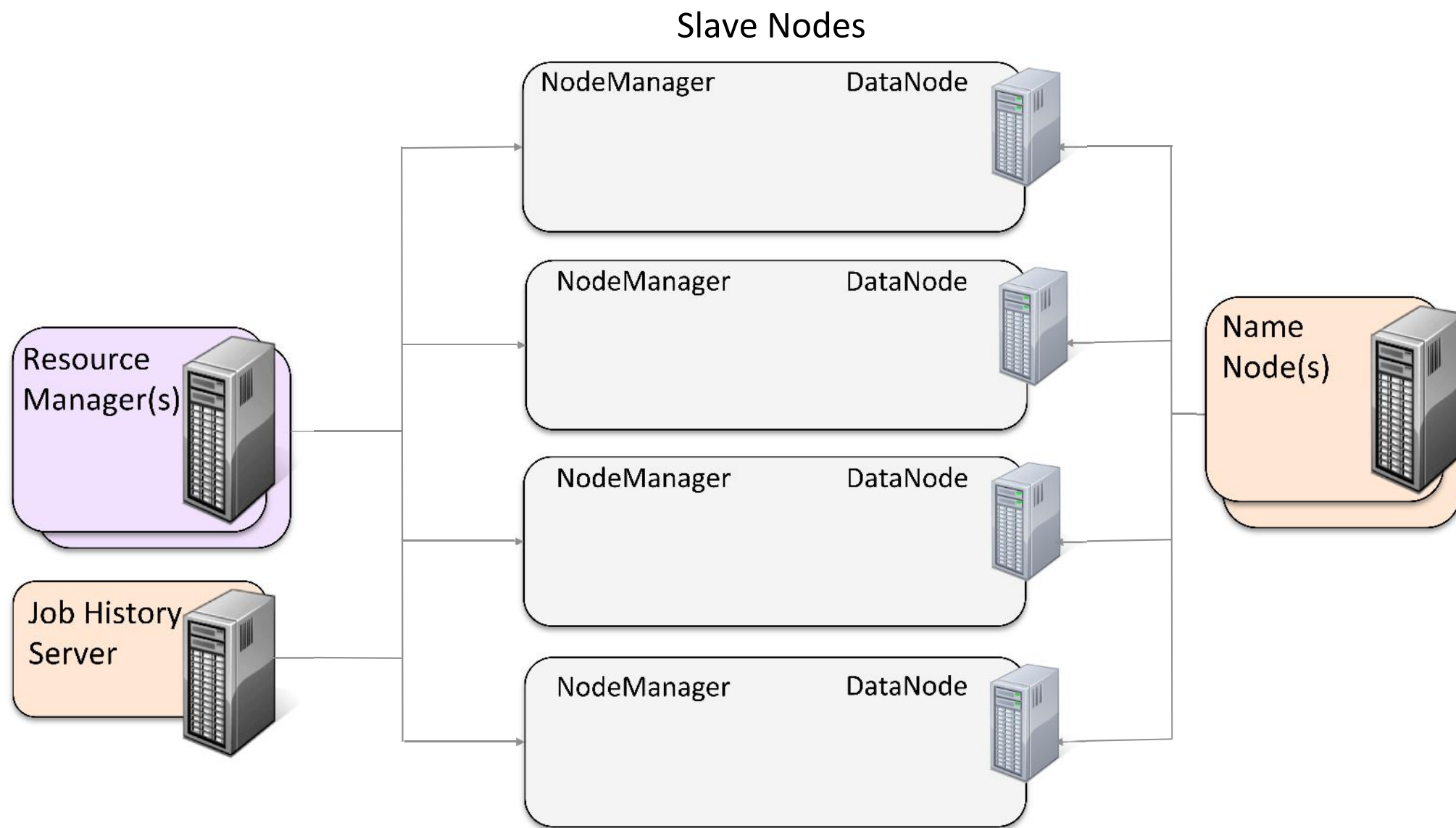
Running a Job on a MapReduce v1 Cluster (1)



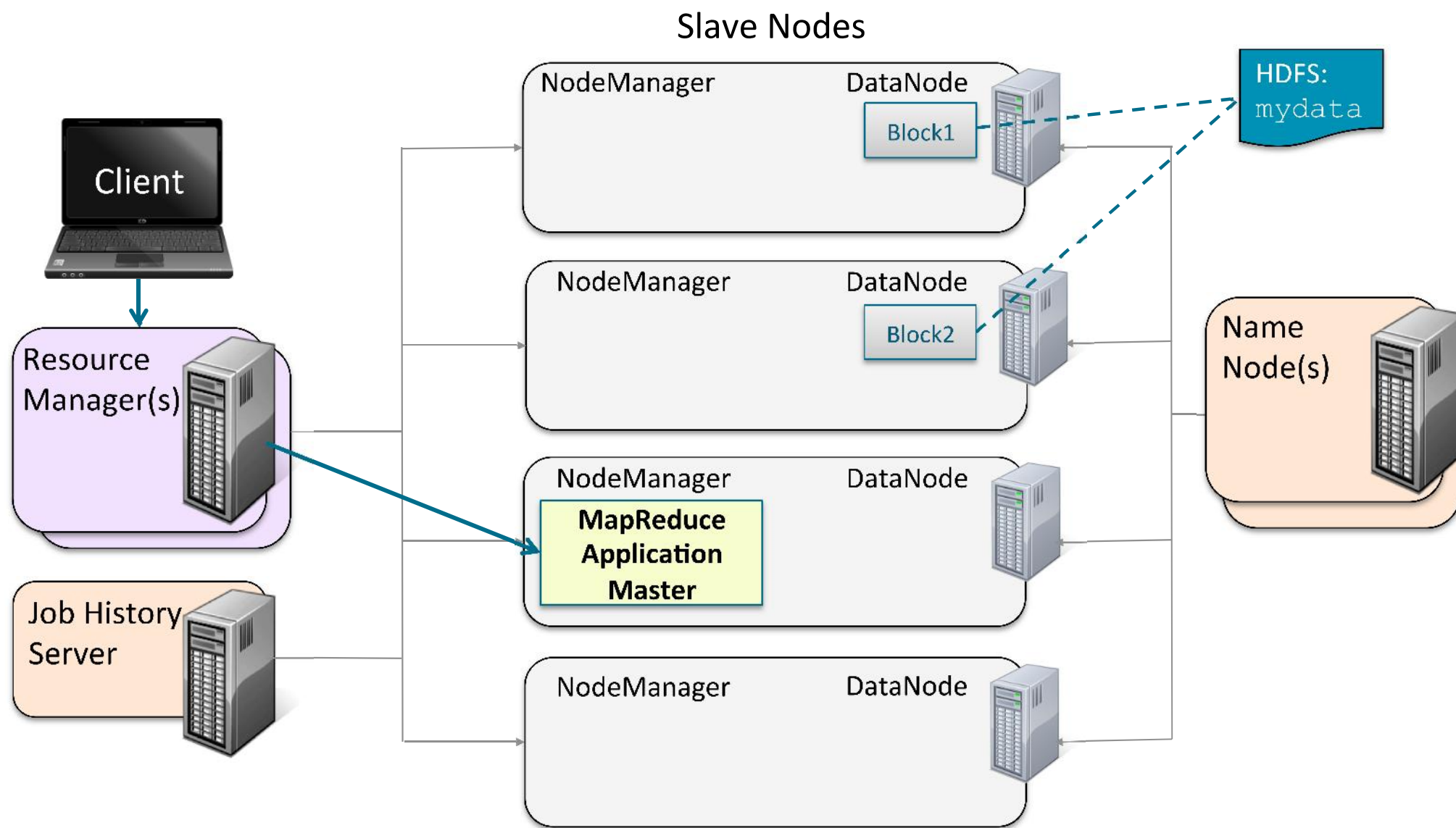
Running a Job on a MapReduce v1 Cluster (2)



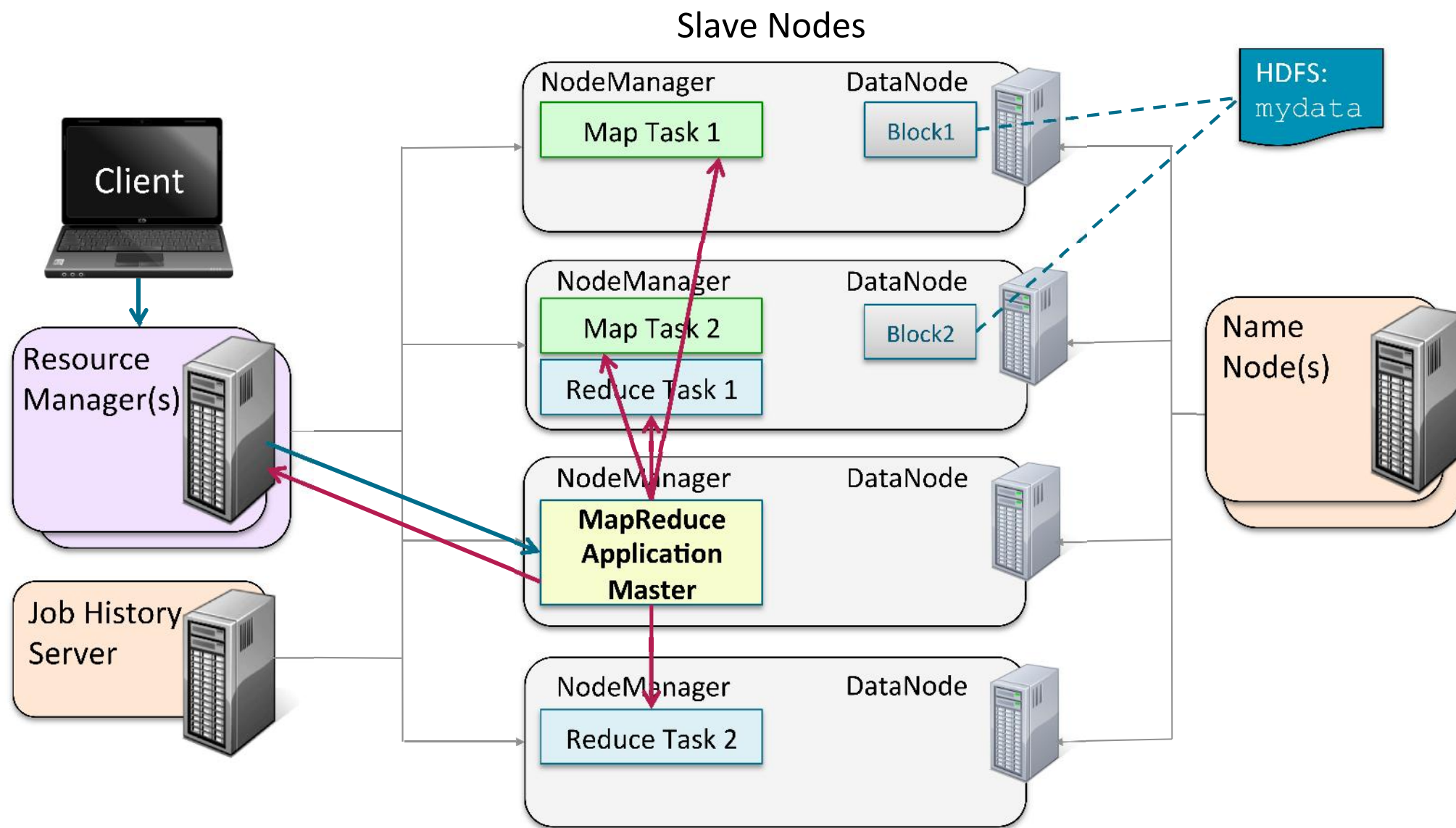
A MapReduce v2 Cluster



Running a Job on a MapReduce v2 Cluster (1)



Running a Job on a MapReduce v2 Cluster (2)



Job Data: Mapper Data Locality

When possible, Map tasks run on a node where a block of data is stored locally

Otherwise, the Map task will transfer the data across the network as it processes that data

