

TaskTracker

A [TaskTracker](#) is a node in the cluster that accepts tasks - Map, Reduce and Shuffle operations - from a [JobTracker](#).

Every [TaskTracker](#) is configured with a set of *slots*, these indicate the number of tasks that it can accept. When the [JobTracker](#) tries to find somewhere to schedule a task within the [MapReduce](#) operations, it first looks for an empty slot on the same server that hosts the [DataNode](#) containing the data, and if not, it looks for an empty slot on a machine in the same rack.

The [TaskTracker](#) spawns a separate JVM processes to do the actual work; this is to ensure that process failure does not take down the task tracker. The [TaskTracker](#) monitors these spawned processes, capturing the output and exit codes. When the process finishes, successfully or not, the tracker notifies the [JobTracker](#). The [TaskTrackers](#) also send out heartbeat messages to the [JobTracker](#), usually every few minutes, to reassure the [JobTracker](#) that it is still alive. These message also inform the [JobTracker](#) of the number of available slots, so the [JobTracker](#) can stay up to date with where in the cluster work can be delegated.