DistCp

DistCp (distributed copy) is a tool used for large inter/intra-cluster copying. It uses MapReduce to effect its distribution, error handling and recovery, and reporting. It expands a list of files and directories into input to map tasks, each of which will copy a partition of the files specified in the source list. Its MapReduce pedigree has endowed it with some quirks in both its semantics and execution. The purpose of this document is to offer guidance for common tasks and to elucidate its model.