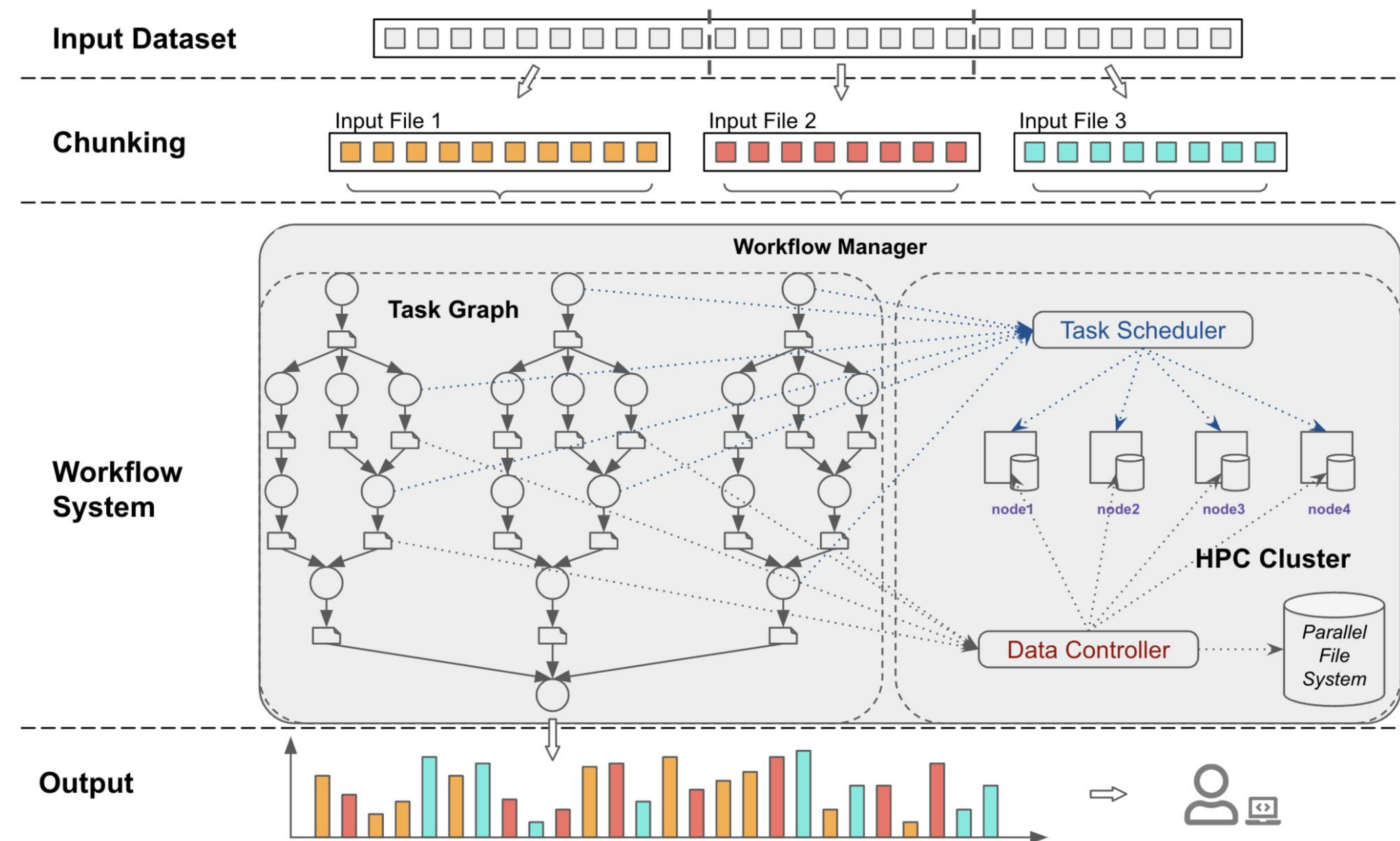


Effectively Exploiting Node-Local Storage for Data-Intensive Scientific Workflows

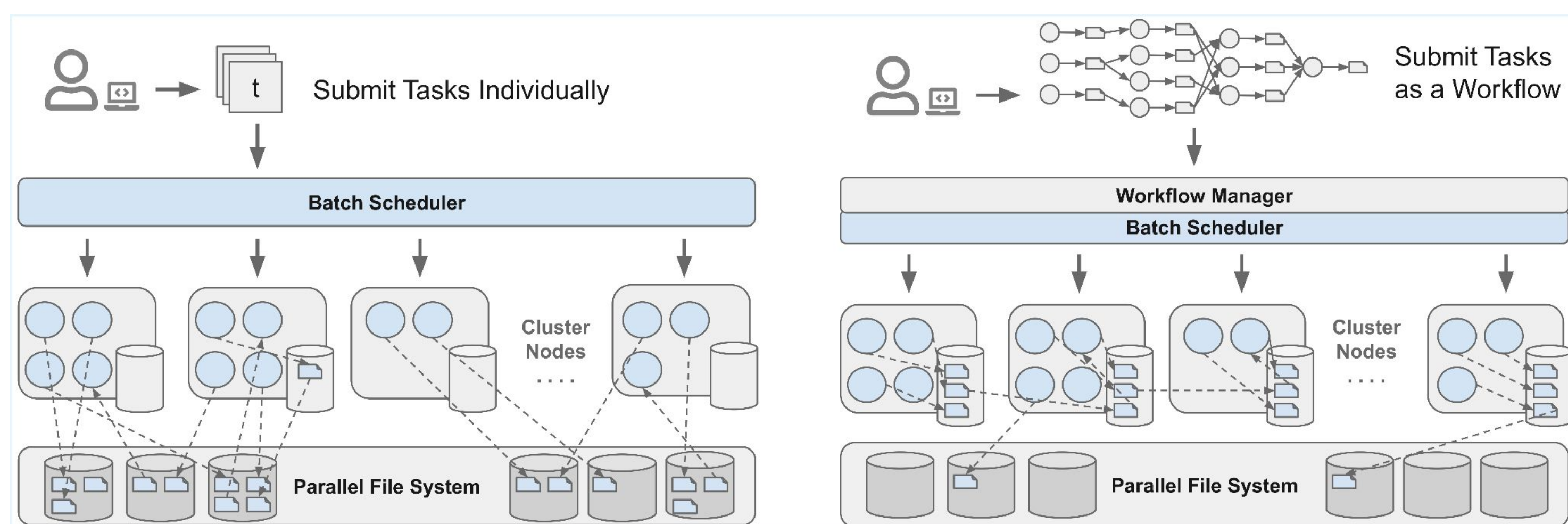
Jin Zhou (jzhou24@nd.edu) University of Notre Dame - <http://ccl.cse.nd.edu/>

Background

Scientific workflows divide large input datasets into small chunks, orchestrate massive numbers of tasks, and execute them on thousands of workers to produce the final results.

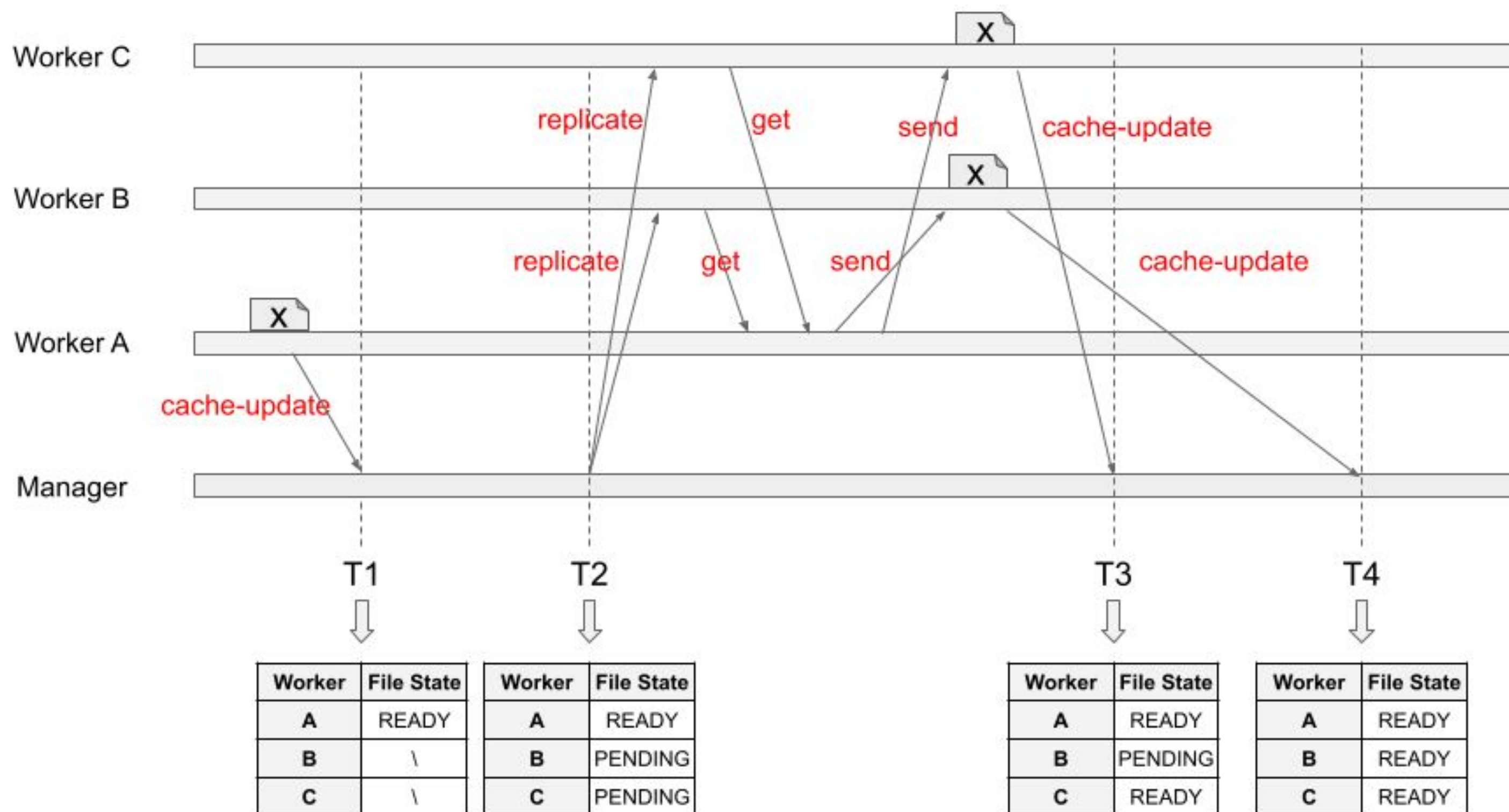


Node-local storage (NLS) has emerged as a viable alternative to the parallel filesystem (PFS) when running data-intensive scientific workflows, alleviating I/O contention.

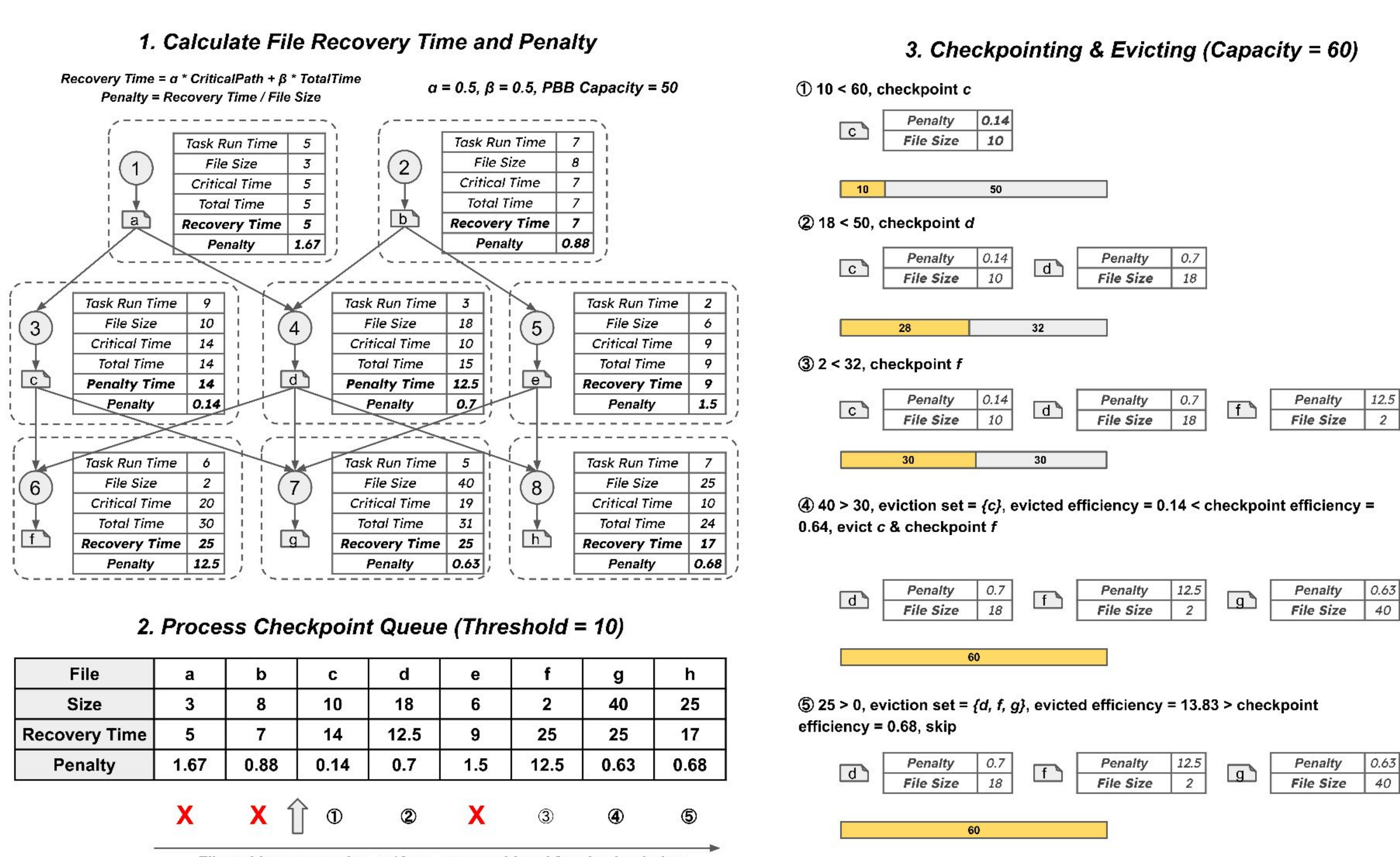


Fault Tolerance

Replication: Each file is replicated to other workers.



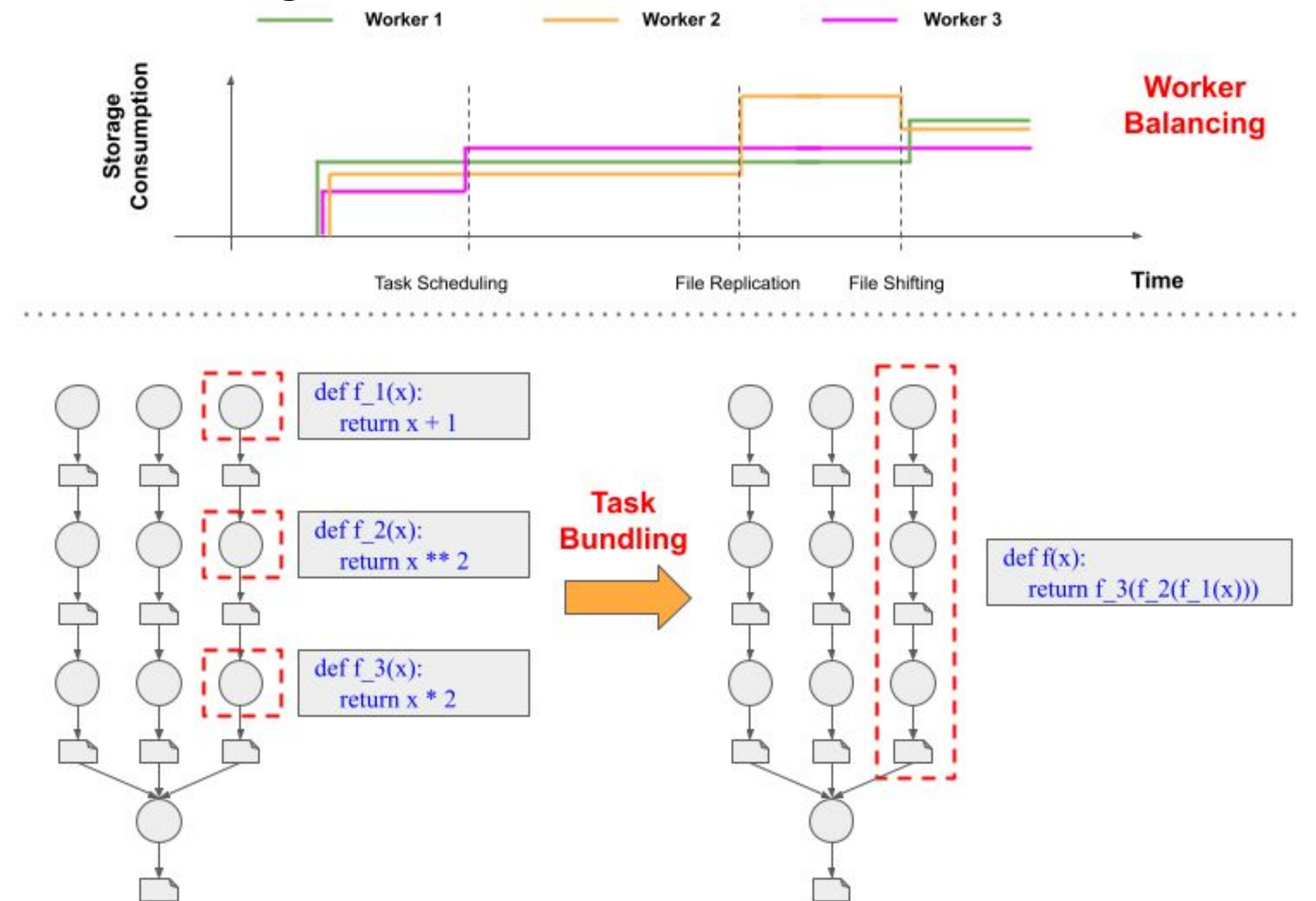
Checkpointing: Some selected files are persisted to the PFS.



Load Balancing

Worker Balancing: a) workers with more available disk are prioritized for task scheduling & file replication; b) proactively shift files from overloaded workers to more available ones.

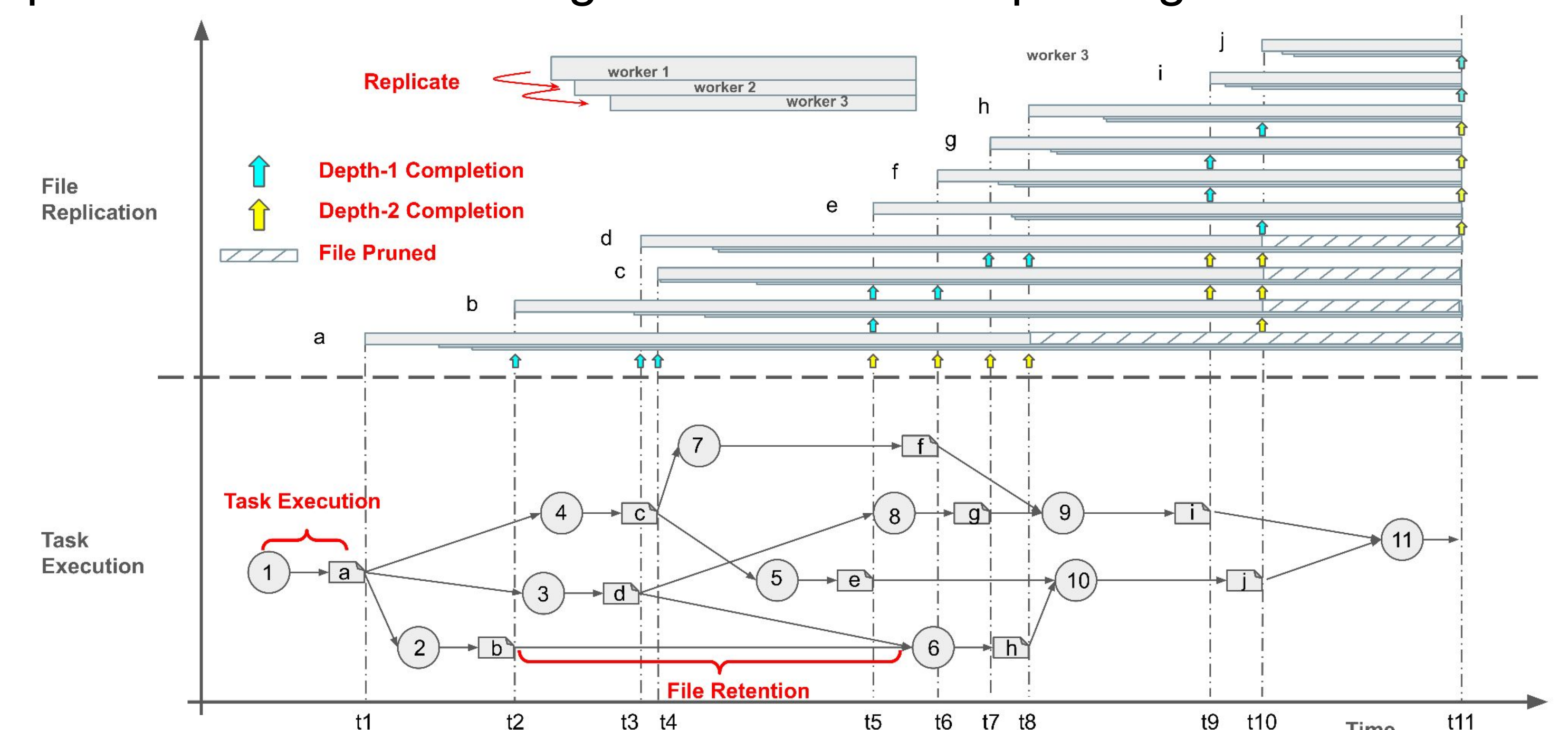
Task Bundling: bundle chained tasks to reduce transient files.



Storage Consumption

File Pruning: Stale files are cleaned to release disk space

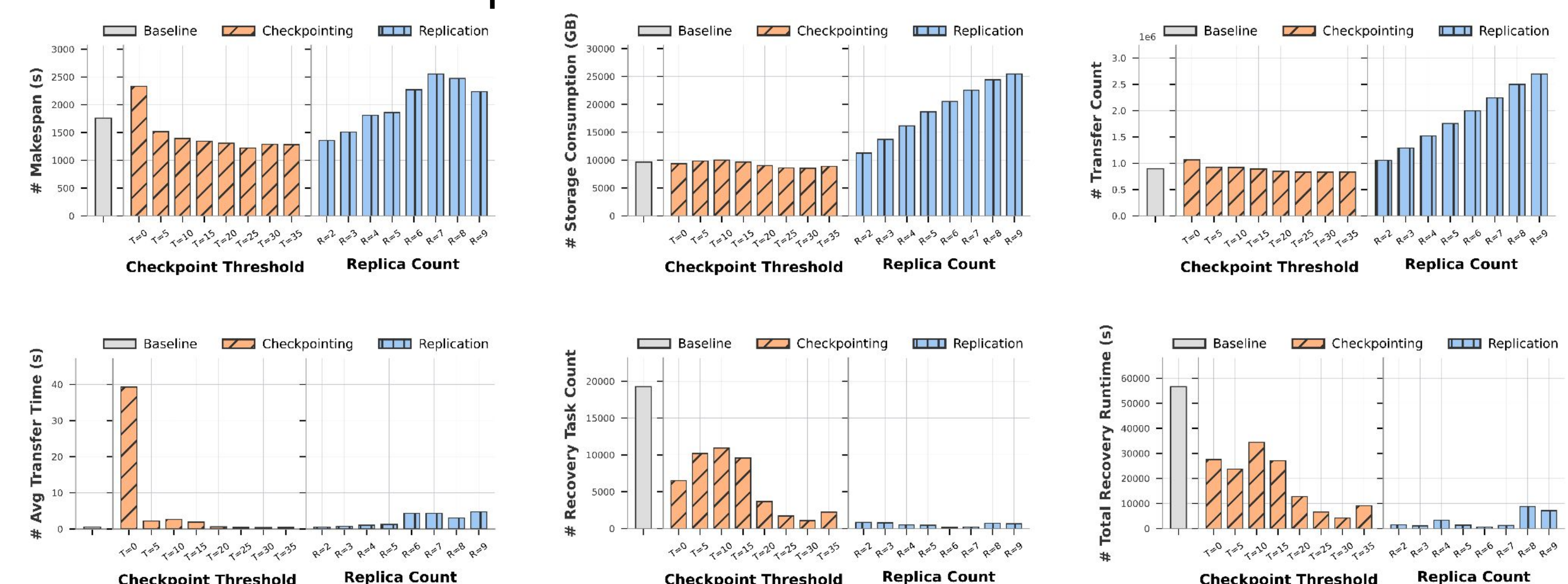
LSF Scheduling: Tasks that can consume larger inputs are prioritized for scheduling to accelerate file pruning.



Results

Setup: 250K tasks on 960 cores, killing worker per 90s

Replication reduces recovery tasks, while **checkpointing** shortens the makespan.



The storage consumption is reduced by 93.68%.

