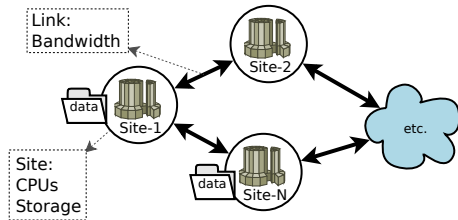


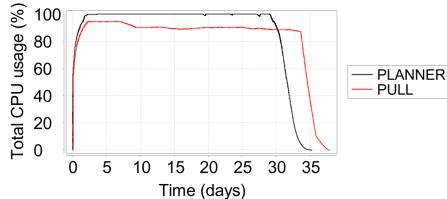
Provenance-aware optimization of workload for distributed data production

New job scheduling approach for data production - global optimization of resource usage {CPU, disk, network bandwidth} (PLANNER)

- ✓ Considers real network structure and data location
- ✓ Adaptive, can deal with loaded (shared) networks and self-discover alternative network path
- ✓ Executed by independent services at each site
- ✓ Scalable to real-sized problems / online planning



Simulation: 51 sites, 36k CPUs, Tier-0/1/2, 2,7 PB of data



Our PLANNER has consistently showed improvements from 7% (large complex network) to 21% (simple network) in makespan over standard scheduling approach used in HENP. Adaptive, can self-adjust to changing infrastructure and take best advantage of sites with poor connectivity