Multi-Tenant GPU: Memory Bandwidth Utilization

Peak BW (single): 488 GB/s

High Priority
Low Priority

Bandwidth (GB/s)

Policy Configuration

no_policy(50,50), prefetch_pid_tree(0,0), prefetch_pid_tree(50,50), prefetch_pid_tree(20,80), prefetch_pid_tree(0,20), prefetch_pid_tree(0,40), prefetch_pid_tree(40,40), prefetch_pid_tree(60,60), prefetch_pid_tree(80,80)