GPT-3 MoE 8x350M All-to-All Straggler Effect ECDF: 1x8 V100 GPUs

ECDF

$\mu = 1.17$ms
$median = 0.64$ms
$P95 = 0.95$ms

Delay (ms)