Performance for Model: 1B, Seq Len: 16384 on H100 56 410.62 424.31 448.10 448.79 449.10 449.03 447.15 650 48 410.33 424.30 446.85 447.74 448.78 448.52 448.93 600 Host Memory (GB) 32 40 410.22 - 550 424.29 446.11 447.97 450.06 448.44 448.74 200 TFLOPS/s 449.80 448.25 410.26 422.96 446.87 449.14 449.29 450 450.91 411.49 424.89 446.32 449.43 448.54 449.96 400 16 409.95 424.69 445.56 448.44 449.89 448.77 450.72 - 350 16 20 28 32 24 64 78 Device Memory (GB)