Performance for Model: 8B, Seq Len: 16384 on A100 176 - 180 143.05 143.12 143.43 143.37 143.06 143.46 160 -160 143.24 143.51 143.58 143.77 143.04 143.68 - 140 Host Memory (GB) 12 128 144 142.95 143.46 143.71 144.02 144.00 144.26 - 120 TFLOPS/s 143.18 143.25 143.33 143.64 143.28 143.58 100 142.91 142.99 143.33 143.50 144.05 143.74 - 80 96 142.75 142.99 142.54 143.58 143.53 143.36 60 80 137.38 139.46 141.21 143.60 143.56 142.98 60 30 40 50 78 70 Device Memory (GB)