Performance for Model: 1B, Seq Len: 16384 on A100 - 180 80 85.60 85.68 85.53 85.55 85.72 85.61 - 160 85.58 85.57 85.62 85.59 85.78 85.63 85.64 85.41 85.61 -140 Host Memory (GB) 48 TFLOPS/s 85.30 85.43 85.59 85.51 85.62 85.52 85.56 85.56 85.52 100 32 85.43 85.54 85.46 85.49 85.51 85.59 85.43 85.50 85.41 - 80 16 85.43 85.25 85.29 85.42 85.40 84.92 85.53 85.42 85.42 60 20 24 28 30 40 50 60 70 78 Device Memory (GB)