Performance for Model: 8B, Seq Len: 4096 on H100 176 0.37 0.42 0.47 0.54 0.56 0.56 8.0 160 0.37 0.42 0.47 0.54 0.56 0.56 -0.7Host Memory (GB) 12 128 144 0.37 0.42 0.47 0.54 0.56 0.56 - 0.6 0.37 0.42 0.47 0.54 0.56 0.56 - 0.5 0.37 0.47 0.54 0.56 0.56 0.42 0.4 96 0.37 0.42 0.47 0.54 0.56 0.56 0.3 80 0.49 0.56 0.50 0.52 0.55 0.56 40 78 30 50 60 70 Device Memory (GB)