Performance for Model: 8B, Seq Len: 8192 on H100 176 0.40 0.43 0.49 0.56 0.57 0.57 8.0 160 0.40 0.43 0.49 0.56 0.57 0.57 -0.7Host Memory (GB) 12 128 144 0.40 0.43 0.49 0.56 0.57 0.57 - 0.6 0.57 0.40 0.43 0.49 0.56 0.57 - 0.5 0.40 0.43 0.49 0.56 0.57 0.57 0.4 96 0.40 0.43 0.49 0.56 0.57 0.57 -0.30.55 0.56 0.56 0.56 0.57 0.57 60 78 40 50 30 70 Device Memory (GB)