Performance for Model: 8B, Seq Len: 512 on H100 176 0.35 0.40 0.47 0.54 0.56 0.57 8.0 160 0.35 0.40 0.47 0.54 0.56 0.57 -0.7Host Memory (GB) 12 128 144 0.35 0.40 0.47 0.54 0.56 0.57 - 0.6 0.35 0.40 0.47 0.54 0.56 0.57 - 0.5 0.35 0.47 0.54 0.56 0.57 0.40 0.4 96 0.35 0.40 0.47 0.54 0.57 0.57 0.3 80 0.47 0.51 0.48 0.54 0.56 0.57 40 50 78 30 60 70 Device Memory (GB)