Performance for Model: 8B, Seq Len: 4096 on H100 176 0.37 0.42 0.48 0.55 0.57 0.57 8.0 160 0.37 0.42 0.48 0.55 0.57 0.57 -0.7Host Memory (GB) 12 128 144 0.37 0.42 0.48 0.55 0.57 0.57 - 0.6 0.57 0.37 0.42 0.48 0.55 0.57 - 0.5 0.37 0.42 0.48 0.55 0.57 0.57 0.4 96 0.37 0.42 0.48 0.55 0.57 0.57 -0.30.54 0.57 0.53 0.54 0.56 0.57 40 60 78 30 50 70 Device Memory (GB)