Performance for Model: 8B, Seq Len: 2048 on H100 176 0.35 0.41 0.48 0.54 0.56 0.56 8.0 160 0.35 0.41 0.48 0.54 0.56 0.56 -0.7Host Memory (GB) 12 128 144 0.35 0.41 0.48 0.54 0.56 0.56 - 0.6 0.35 0.41 0.48 0.54 0.56 0.56 - 0.5 0.35 0.41 0.48 0.54 0.56 0.56 0.4 96 0.35 0.41 0.48 0.54 0.56 0.56 -0.380 0.48 0.56 0.49 0.51 0.55 0.56 40 50 60 78 30 70 Device Memory (GB)