Performance for Model: 8B, Seq Len: 1024 on H100 18000 176 7521.21 10198.16 8643.37 11714.74 12180.85 12213.49 16000 160 7513.49 8631.06 10187.79 11691.82 12189.38 12201.97 14000 Host Memory (GB) 12 128 144 7511.40 8631.79 10179.51 12214.01 11696.99 12174.80 12000 7512.89 8633.76 10190.58 11694.31 12165.18 12190.37 7510.55 8638.89 10190.88 11687.73 12174.12 12186.33 10000 7520.81 8642.48 10197.25 11708.71 12189.59 12209.68 - 8000 10211.24 10479.70 11056.57 11826.41 12185.55 12193.05 - 6000 70 30 40 50 60 78 Device Memory (GB)