Performance for Model: 8B, Seq Len: 65536 on A100

Host Memory (GB)	30	40	50	60	70	78
176	1088.64	1090.06	1093.25	1095.84	1099.84	1101.16
160	1083.73	1086.33	1081.03	1090.25	1094.64	1096.57
144	1077.13	1075.94	1073.71	1086.30	1089.63	1092.33
128	1071.84	1071.79	1075.27	1079.47	1082.35	1084.00
112	1064.86	1065.60	1069.07	1071.86	1077.43	1078.16
96	1057.69	1059.84	1061.61	1066.01	1069.24	1070.14
80	817.15	847.02	874.06	901.16	926.54	949.63

Device Memory (GB)