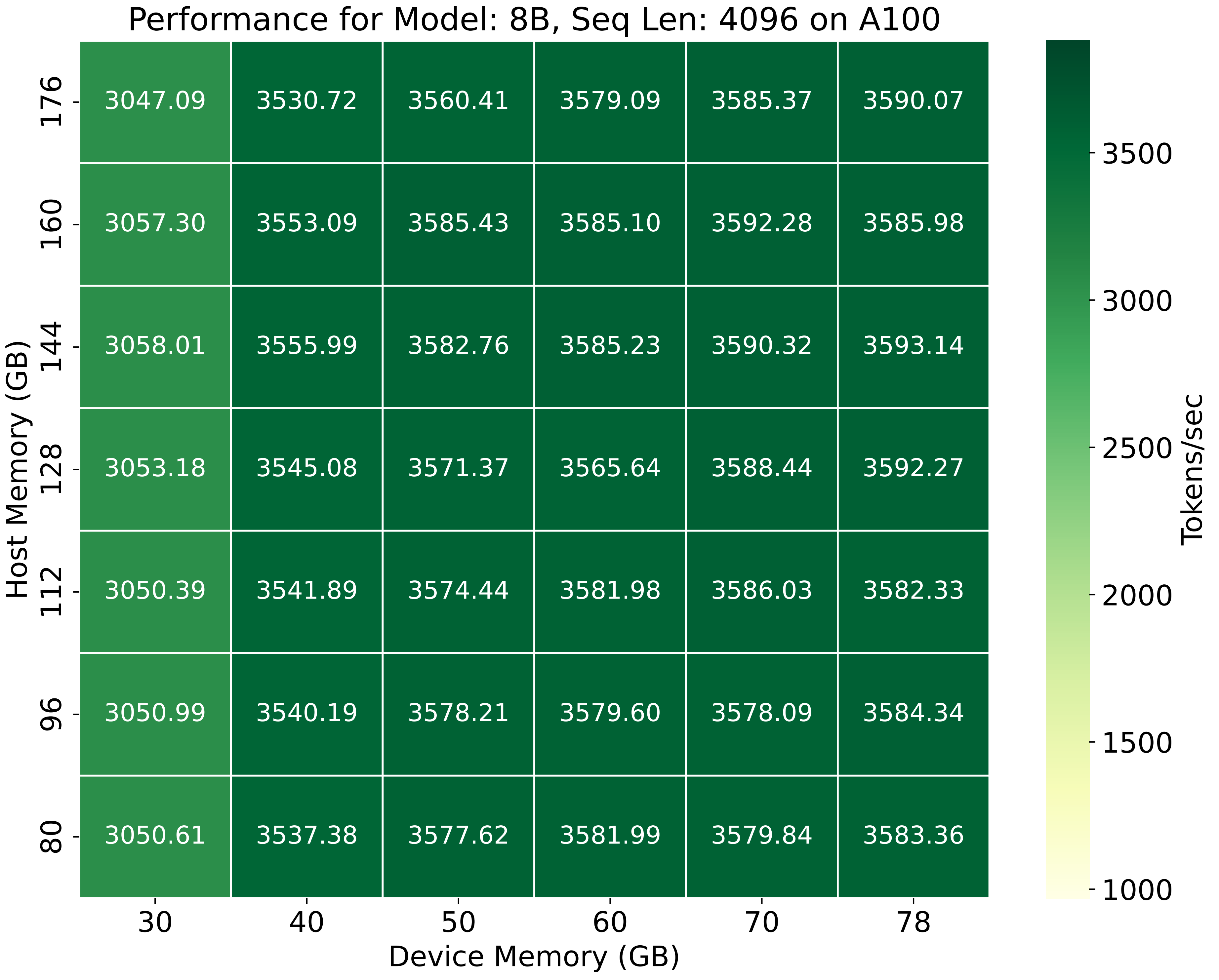

Performance for Model: 8B, Seq Len: 4096 on A100

Host Memory (GB) \ Device Memory (GB)	30	40	50	60	70	78
176	3047.09	3530.72	3560.41	3579.09	3585.37	3590.07
160	3057.30	3553.09	3585.43	3585.10	3592.28	3585.98
144	3058.01	3555.99	3582.76	3585.23	3590.32	3593.14
128	3053.18	3545.08	3571.37	3565.64	3588.44	3592.27
112	3050.39	3541.89	3574.44	3581.98	3586.03	3582.33
96	3050.99	3540.19	3578.21	3579.60	3578.09	3584.34
80	3050.61	3537.38	3577.62	3581.99	3579.84	3583.36