

Performance for Model: 8B, Seq Len: 65536 on RTX3090

