

Performance for Model: 8B, Seq Len: 1024 on RTX3090

