

Performance for Model: 1B, Seq Len: 131072 on RTX5090

