

seq16384-hd64-12h-8b | Q=16384, K=16384, head_dim=64, heads=12, batch=8

Kernel	Avg Forward Time (ms)	Relative to Triton Flash
scaled_dot_prod_attention	ERROR	N/A
vectorized_torch	ERROR	N/A
vectorized_torch_compiled	ERROR	N/A
flash_attention_triton	599.06	1.00x