

Throughput (tokens/s)

1F1B TP FSDP LoRAPP  
(a) TinyLlama-1.1B

(b) Llama-2-7B

(c) Llama-2-13B

