Throughput-Latency characterization
meta-llama_llama-3.1-70b (TP=4)
Batch Size: 8
Dataset: wildchat