

# Prefill stays compute-bound even at 95% reuse (Qwen3-Coder-30B-A3B, H20, seqlen=64,000)

