

# Qwen MoE: Performance Speedup vs Batch Size

## Prefetching Strategy Effectiveness

