

# Qwen MoE: Cache Hit Rate vs Batch Size

## Multi-Level Caching Effectiveness

