

Cold execution time components of ML inference benchmarks
NVIDIA A100, n=100, median execution time

