

Execution time components of ML inference benchmarks
NVIDIA A100, n=20, median execution time

