

Cold vs. Warm Latency - phi3, AWS CPU

