

# Inference Latency vs. Number of Servers (DistillBERT)

