`Please use '!option handwritten true' to enable handwritten`

Horizontal Scaling Performance Analysis

Scaling Performance Analysis

Three Instances

CPU Usage: 35%

Response Time: 90ms

Throughput: 1400 RPS

Two Instances

Throughput: 950 RPS

Response Time: 95ms

CPU Usage: 45%

Single Instance Baseline

Throughput: 500 RPS

Response Time: 100ms

CPU Usage: 75%

Scaling Efficiency
2 instances: 95% efficiency
3 instances: 93% efficiency
Near-linear scaling achieved

Scaling Metrics
=
Based on load testing
Using Azure Kubernetes Service
--
Green: Improved
Yellow: Neutral
Red: Degraded