

# Executive Summary – Stress Test Results

---

## Objective

Evaluate the scalability of Dummy JSON APIs under ramping user load until SLA thresholds break (p95 > 1200 ms or > 2 % failures).

## Response Time Trends

- Median ≈ 11.8 s, p99 ≈ 84 s → latency increased sharply beyond 200 threads.
- Only ~25 % of requests completed < 500 ms.
- Tail latency growth indicates backend queuing or thread starvation.

## Throughput (RPS)

- Peaked around 4 transactions per second.
- Throughput dropped rapidly after SLA breach as errors escalated.

## Error Breakdown

Error Type	% of Errors	Cause
Assertion Failed	58 %	Responses exceeded SLA thresholds
Socket Exception	41 %	Server connections dropped under load
HTTP 400	1 %	Minor functional issues

## Bottleneck Hypothesis

- Fetch products List (81 % fail)** → Likely DB or I/O delay
- Login (86 % fail)** → Authentication lock or session contention.
- Add to Cart (64 % fail)** → Business-logic processing lag.

## Tuning Recommendations

- Increase server thread pool and connection limits.
- Optimize database queries and enable caching for read-heavy endpoints.
- Use asynchronous auth calls or token reuse.
- Enable HTTP keep-alive and tune timeout thresholds.
- Add horizontal scaling / load balancing once latency stabilizes.