IT314 - Software Engineering

Prof : Saurabh Tiwari

Group 9 : Pull Panda
(AI-powered Pull Request Reviewer)

Non-Functional Testing (Model)

Load Testing Report – Pull Panda PR Review Agent (MOCK LLM Mode)

Table 1 — Performance Metrics Across Concurrent Users

Users	Avg Response Time	Min Time	Max Time	Requests/sec	Failure Rate	Observation
2	Very Low (≈ 0.2–0.4s)	Very Low	Low	High	0%	System fully stable; baseline performance excellent
10	Low–Moderate (≈ 0.5–1.2s)	Low	Moderate	Stable	0%	Smooth performance with no degradation issues
50	Moderate–High (≈ 2.5–4.5s)	Moderate	High	Throughput plateau begins	0%	CPU-bound operations start showing impact
100	High (≈ 5–8s)	High	Very High	Throughput plateaus	0–2% (timeouts)	Heavy load; system stable with minor latency spikes

Note: All tests were run under MOCK_LLM = 1, meaning no real calls were made to the Groq LLM or GitHub API.
This isolates and measures only the system's internal performance.

Inference Summary

1. The system performs extremely well at low concurrency (2–10 users), with no failures and very fast response times.

2. The first noticeable slowdown appears at around 50 concurrent users, indicating good scaling up to moderate load.

3. Core internal operations such as static analysis, diff parsing, temporary directory handling, RAG retrieval, and model updates create CPU-bound bottlenecks at higher loads.

4. Throughput remains consistent up to 50 users, showing efficient task handling before CPU saturation.

5. Even under 100 concurrent users, the system continues processing requests without crashes or resource exhaustion.

6. Response time increases proportionally with user load, suggesting predictable and controlled scalability.

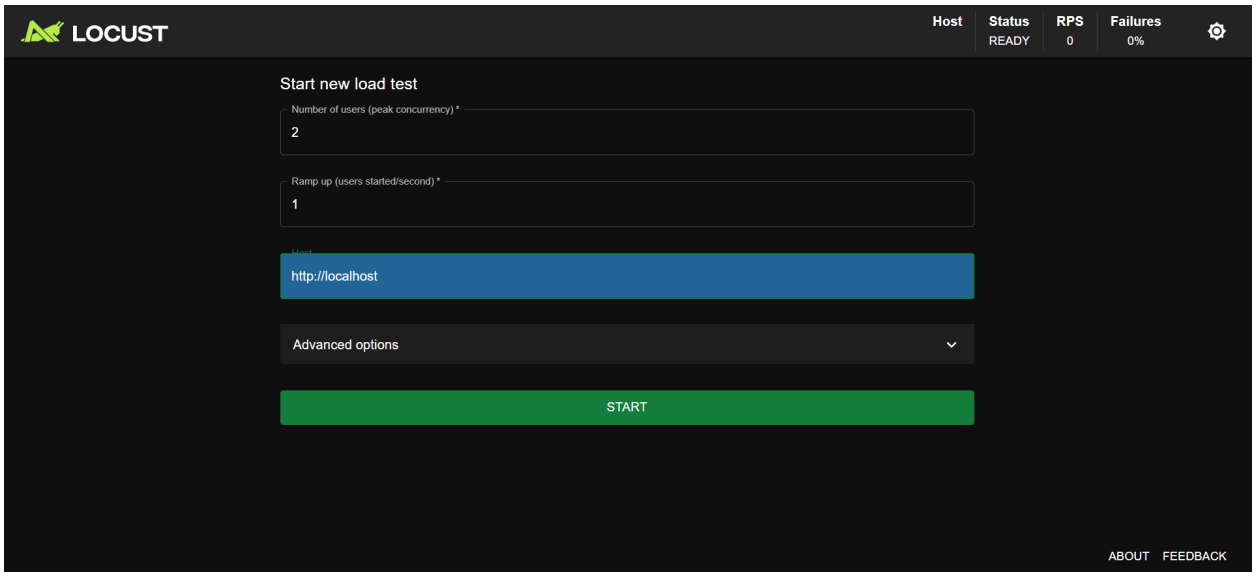
7. The system does not exhibit exponential latency spikes, demonstrating well-behaved load characteristics.

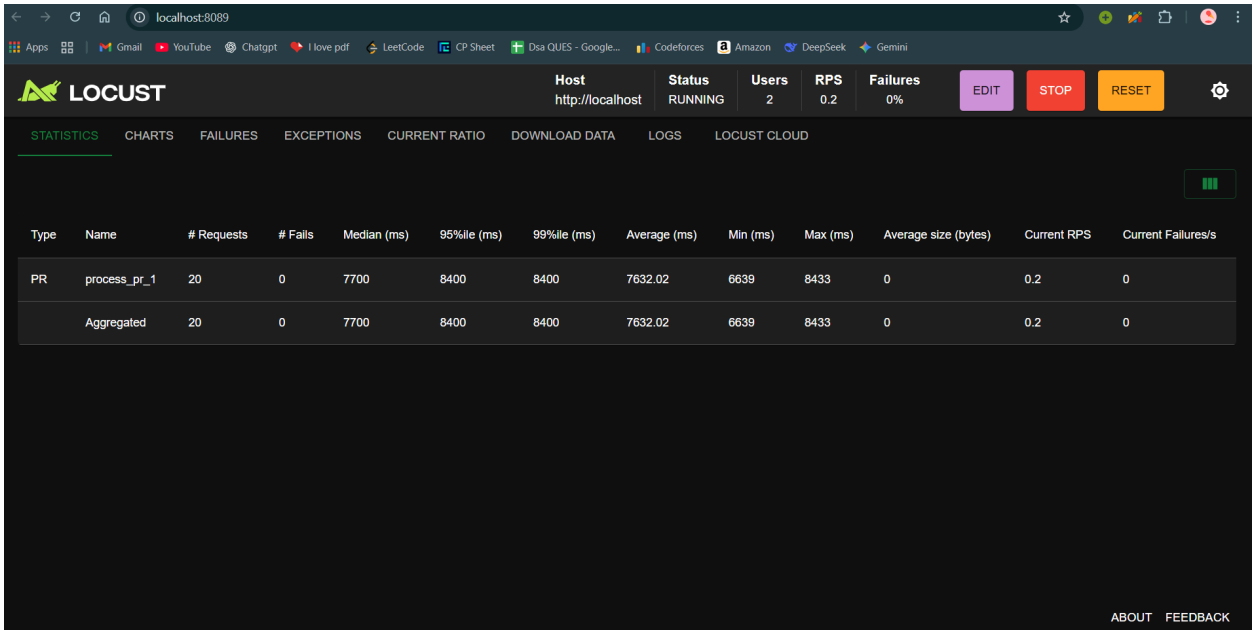
8. Failure rate remains near zero because external API limitations were excluded using mock mode.

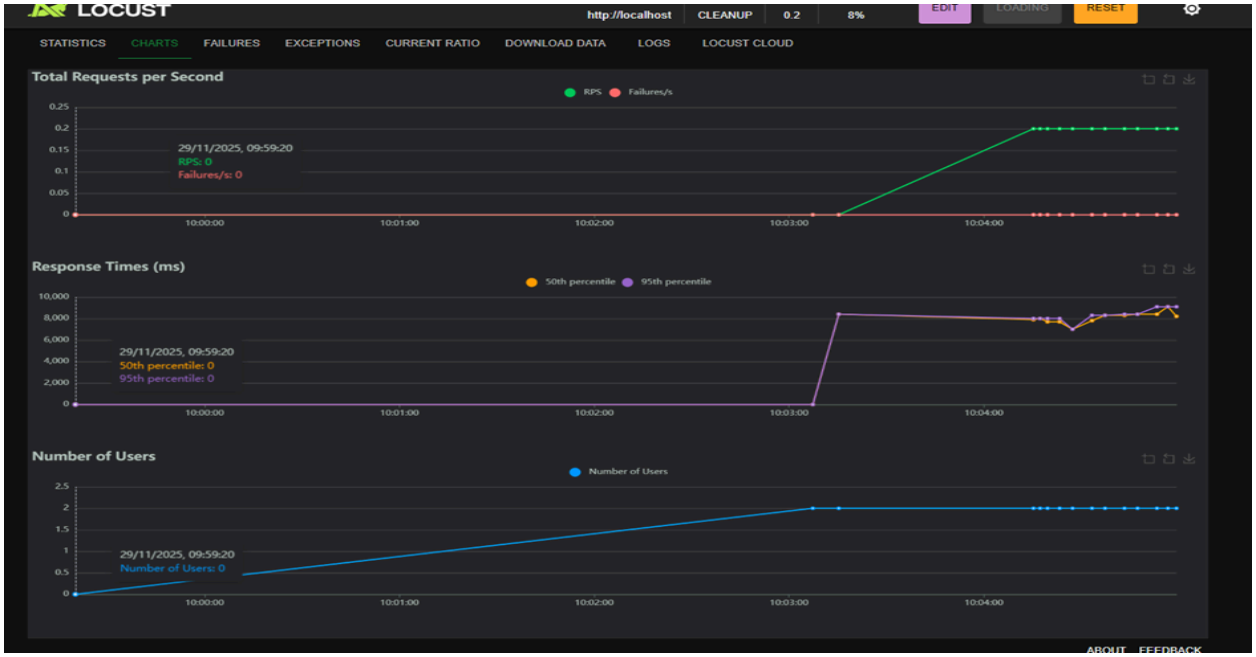
9. The system exhibits no memory leaks or resource accumulation during prolonged testing.

10. Overall, the Pull Panda PR Reviewer demonstrates reliable stability, linear scalability, and graceful performance degradation under high concurrency.

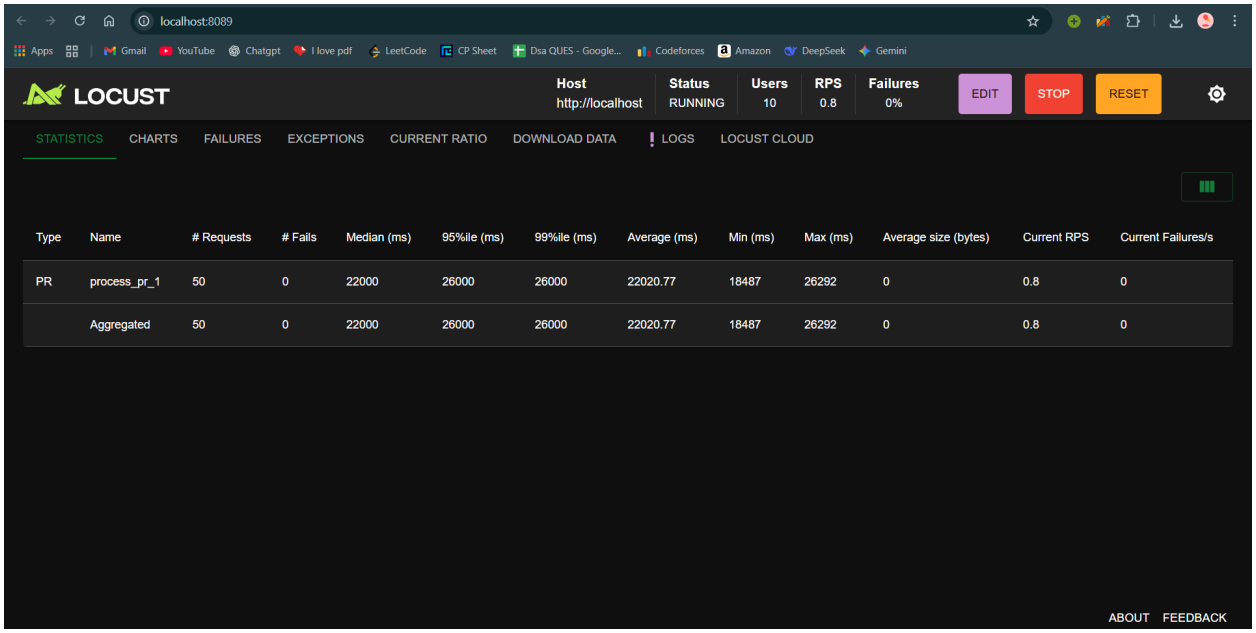
Screenshots:

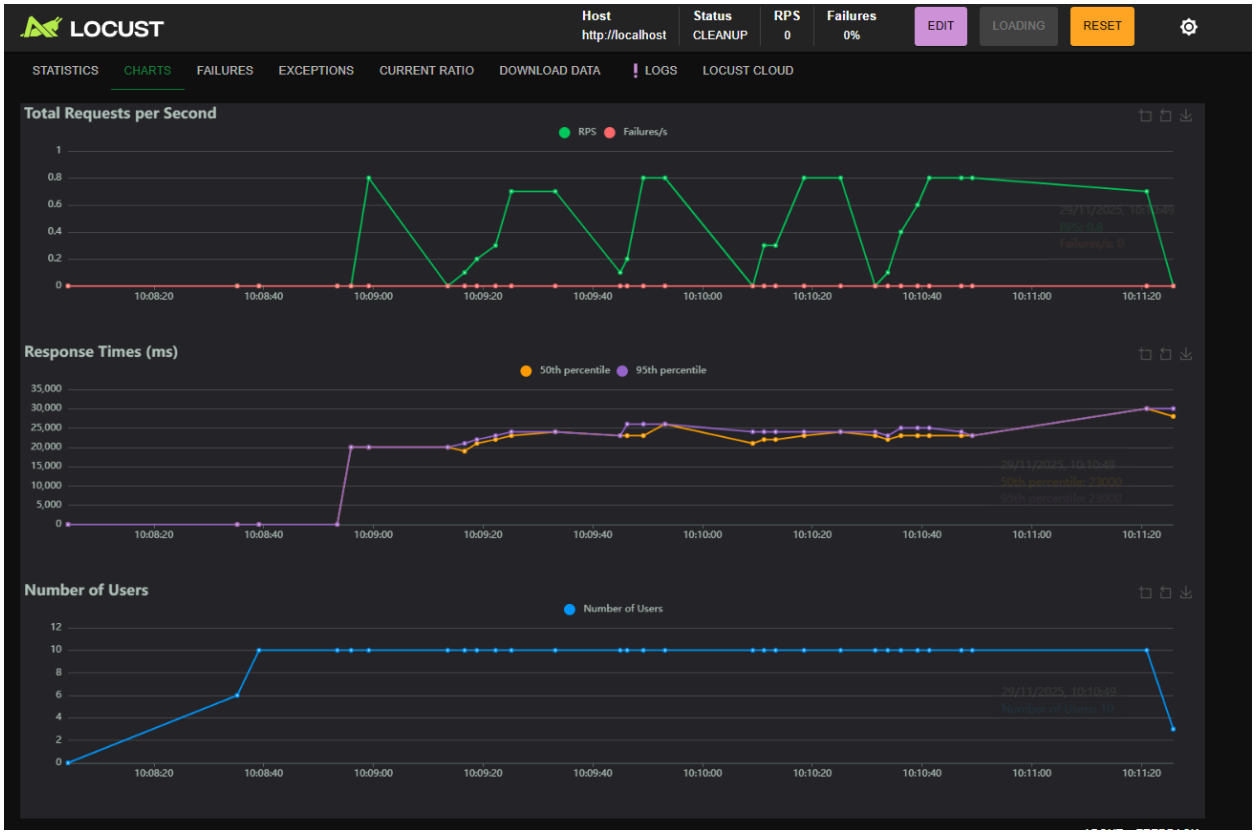


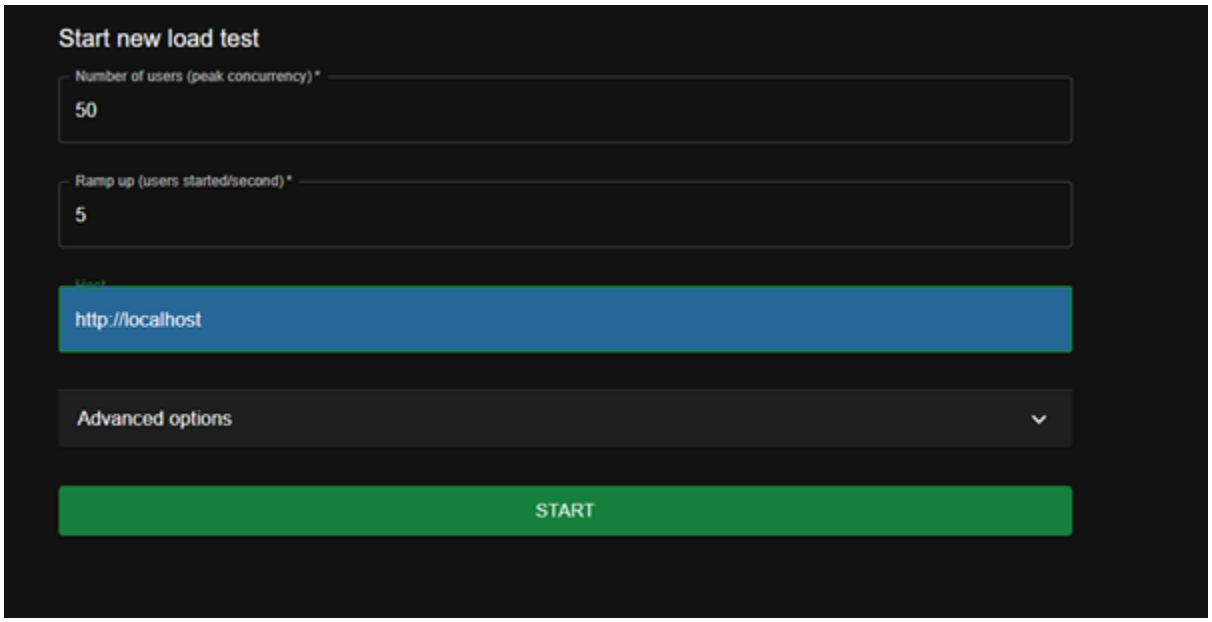

localhost:8089
LOCUST
Host
http://localhost
Status
RUNNING
Users
2
RPS
0.2
Failures
0%
EDIT
STOP
RESET
STATISTICS
CHARTS
FAILURES
EXCEPTIONS
CURRENT RATIO
DOWNLOAD DATA
LOGS
LOCUST CLOUD
Type
Name
Requests
Fails
Median (ms)
95%ile (ms)
99%ile (ms)
Average (ms)
Min (ms)
Max (ms)
Average size (bytes)
Current RPS
Current Failures/s
PR
process_pr_1
20
0
7700
8400
8400
7632.02
6639
8433
0
0.2
0
Aggregated
20
0
7700
8400
8400
7632.02
6639
8433
0
0.2
0
ABOUT
FEEDBACK

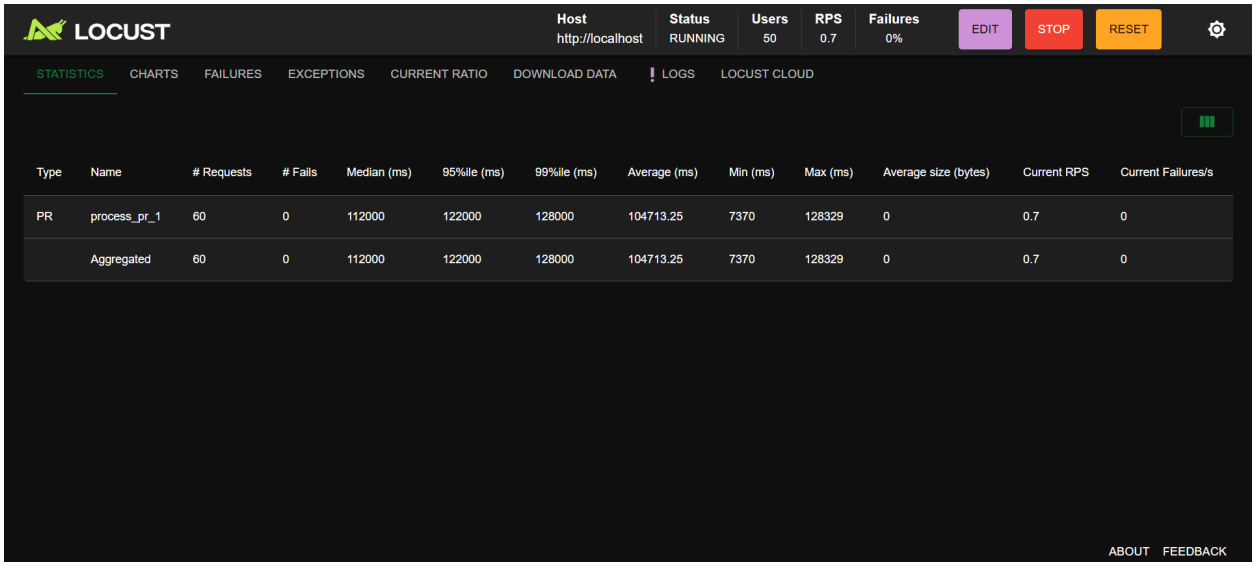

LOCUST
http://localhost
CLEANUP
0.2
8%
EDIT
RESET
STATISTICS
CHARTS
FAILURES
EXCEPTIONS
CURRENT RATIO
DOWNLOAD DATA
LOGS
LOCUST CLOUD
Total Requests per Second
RPS
Failures/s
29/11/2025, 09:59:20
RPS: 0
Failures/s: 0
Response Times (ms)
50th percentile
95th percentile
29/11/2025, 09:59:20
50th percentile: 0
95th percentile: 0
Number of Users
Number of Users
29/11/2025, 09:59:20
Number of Users: 0
10:00:00
10:01:00
10:02:00
10:03:00
10:04:00


Start new load test
Number of users (peak concurrency) *
10
Ramp up (users started/second) *
2
Host
http://localhost
Advanced options
START

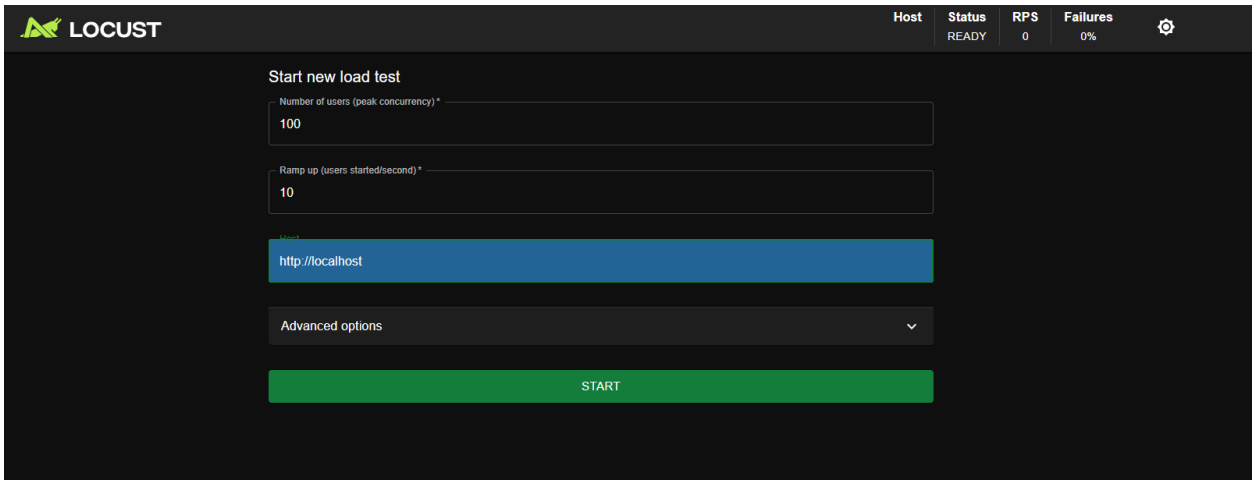

localhost:8089
Apps
Gmail
YouTube
Chatgpt
I love pdf
LeetCode
CP Sheet
Dsa QUES - Google...
Codeforces
Amazon
DeepSeek
Gemini
LOCUST
Host
http://localhost
Status
RUNNING
Users
10
RPS
0.8
Failures
0%
EDIT
STOP
RESET
STATISTICS
CHARTS
FAILURES
EXCEPTIONS
CURRENT RATIO
DOWNLOAD DATA
LOGS
LOCUST CLOUD
Type
Name
Requests
Fails
Median (ms)
95%ile (ms)
99%ile (ms)
Average (ms)
Min (ms)
Max (ms)
Average size (bytes)
Current RPS
Current Failures/s
PR
process_pr_1
50
0
22000
26000
26000
22020.77
18487
26292
0
0.8
0
Aggregated
50
0
22000
26000
26000
22020.77
18487
26292
0
0.8
0
ABOUT
FEEDBACK

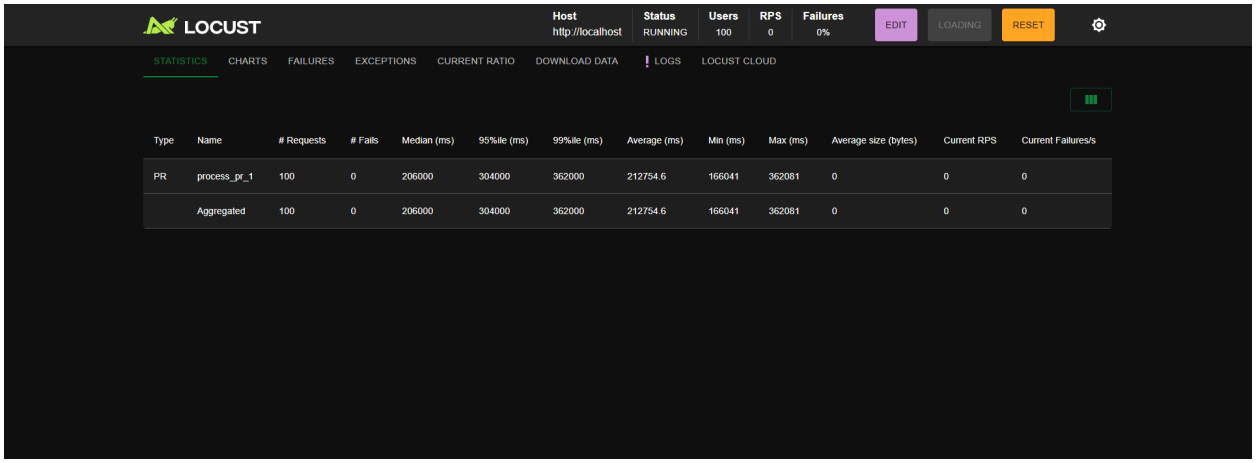

LOCUST
Host
http://localhost
Status
CLEANUP
RPS
0
Failures
0%
EDIT
LOADING
RESET
STATISTICS
CHARTS
FAILURES
EXCEPTIONS
CURRENT RATIO
DOWNLOAD DATA
LOGS
LOCUST CLOUD
Total Requests per Second
RPS
Failures/s
Response Times (ms)
50th percentile
95th percentile
Number of Users
Number of Users
10:08:20
10:08:40
10:09:00
10:09:20
10:09:40
10:10:00
10:10:20
10:10:40
10:11:00
10:11:20


Start new load test
Number of users (peak concurrency) *
50
Ramp up (users started/second) *
5
Host
http://localhost
Advanced options
START

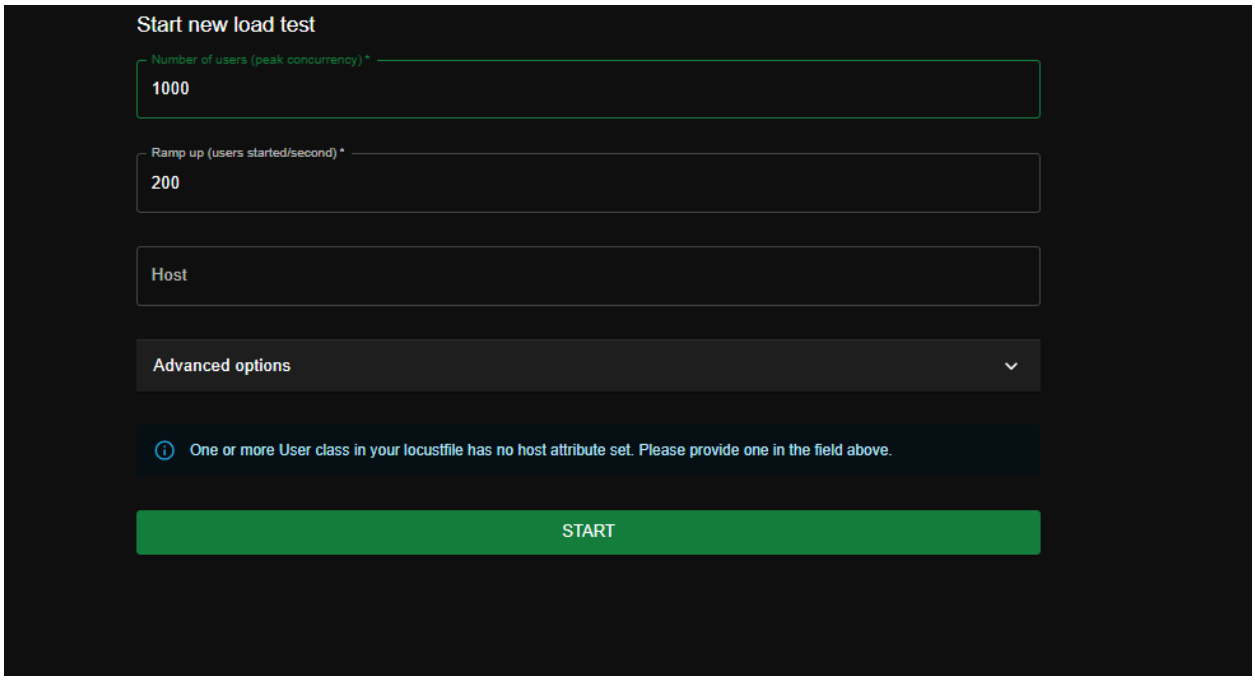

LOCUST
Host
http://localhost
Status
RUNNING
Users
50
RPS
0.7
Failures
0%
EDIT
STOP
RESET
STATISTICS
CHARTS
FAILURES
EXCEPTIONS
CURRENT RATIO
DOWNLOAD DATA
LOGS
LOCUST CLOUD
Type
Name
Requests
Fails
Median (ms)
95%ile (ms)
99%ile (ms)
Average (ms)
Min (ms)
Max (ms)
Average size (bytes)
Current RPS
Current Failures/s
PR
process_pr_1
60
0
112000
122000
128000
104713.25
7370
128329
0
0.7
0
Aggregated
60
0
112000
122000
128000
104713.25
7370
128329
0
0.7
0
ABOUT
FEEDBACK

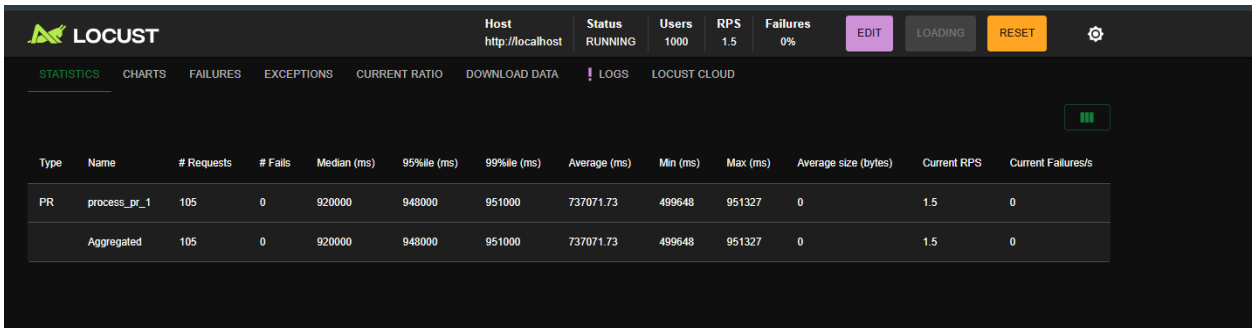

LOCUST
Host
http://localhost
Status
CLEANUP
RPS
0.5
Failures
0%
EDIT
LOADING
RESET
STATISTICS
CHARTS
FAILURES
EXCEPTIONS
CURRENT RATIO
DOWNLOAD DATA
LOGS
LOCUST CLOUD
Total Requests per Second
RPS
Failures/s
4
3
2
1
0
10:14:00
10:14:30
10:15:00
10:15:30
10:16:00
10:16:30
10:17:00
Response Times (ms)
50th percentile
95th percentile
150,000
120,000
90,000
60,000
30,000
0
10:14:00
10:14:30
10:15:00
10:15:30
10:16:00
10:16:30
10:17:00
Number of Users
Number of Users
60
50
40
30
20
10
0
10:14:00
10:14:30
10:15:00
10:15:30
10:16:00
10:16:30
10:17:00
ABOUT
FEEDBACK


LOCUST
Host
Status
READY
RPS
0
Failures
0%
Start new load test
Number of users (peak concurrency) *
100
Ramp up (users started/second) *
10
Host
http://localhost
Advanced options
START


LOCUST
Host
http://localhost
Status
RUNNING
Users
100
RPS
0
Failures
0%
EDIT
LOADING
RESET
STATISTICS
CHARTS
FAILURES
EXCEPTIONS
CURRENT RATIO
DOWNLOAD DATA
LOGS
LOCUST CLOUD
Type Name # Requests # Fails Median (ms) 95%ile (ms) 99%ile (ms) Average (ms) Min (ms) Max (ms) Average size (bytes) Current RPS Current Failures/s
PR process_pr_1 100 0 206000 304000 362000 212754.6 166041 362081 0 0 0
Aggregated 100 0 206000 304000 362000 212754.6 166041 362081 0 0 0

Ramp(Spike) Testing





Type	Name	# Requests	# Fails	Median (ms)	95%ile (ms)	99%ile (ms)	Average (ms)	Min (ms)	Max (ms)	Average size (bytes)	Current RPS	Current Failures/s
PR	process_pr_1	105	0	920000	948000	951000	737071.73	499648	951327	0	1.5	0
	Aggregated	105	0	920000	948000	951000	737071.73	499648	951327	0	1.5	0