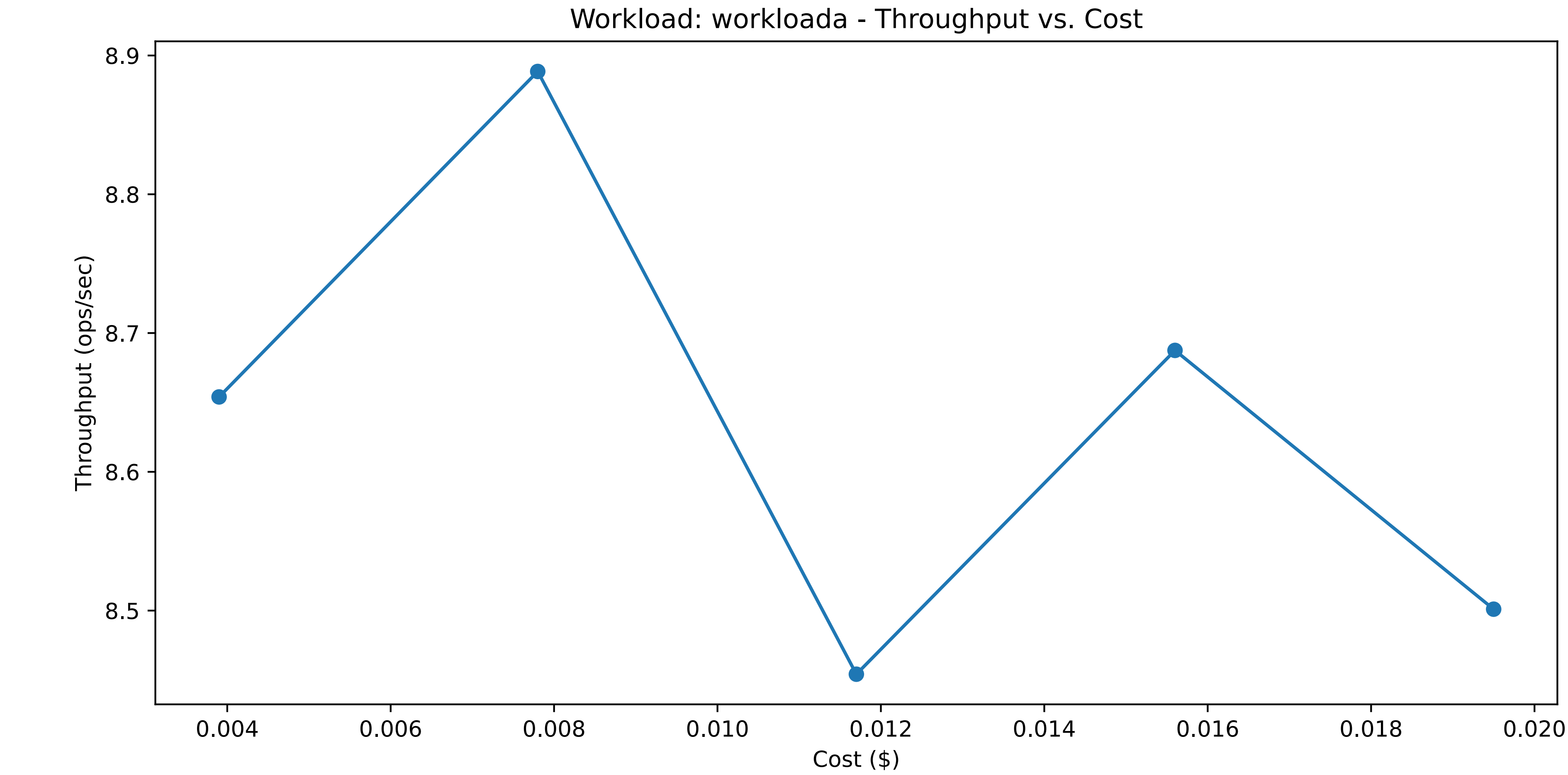
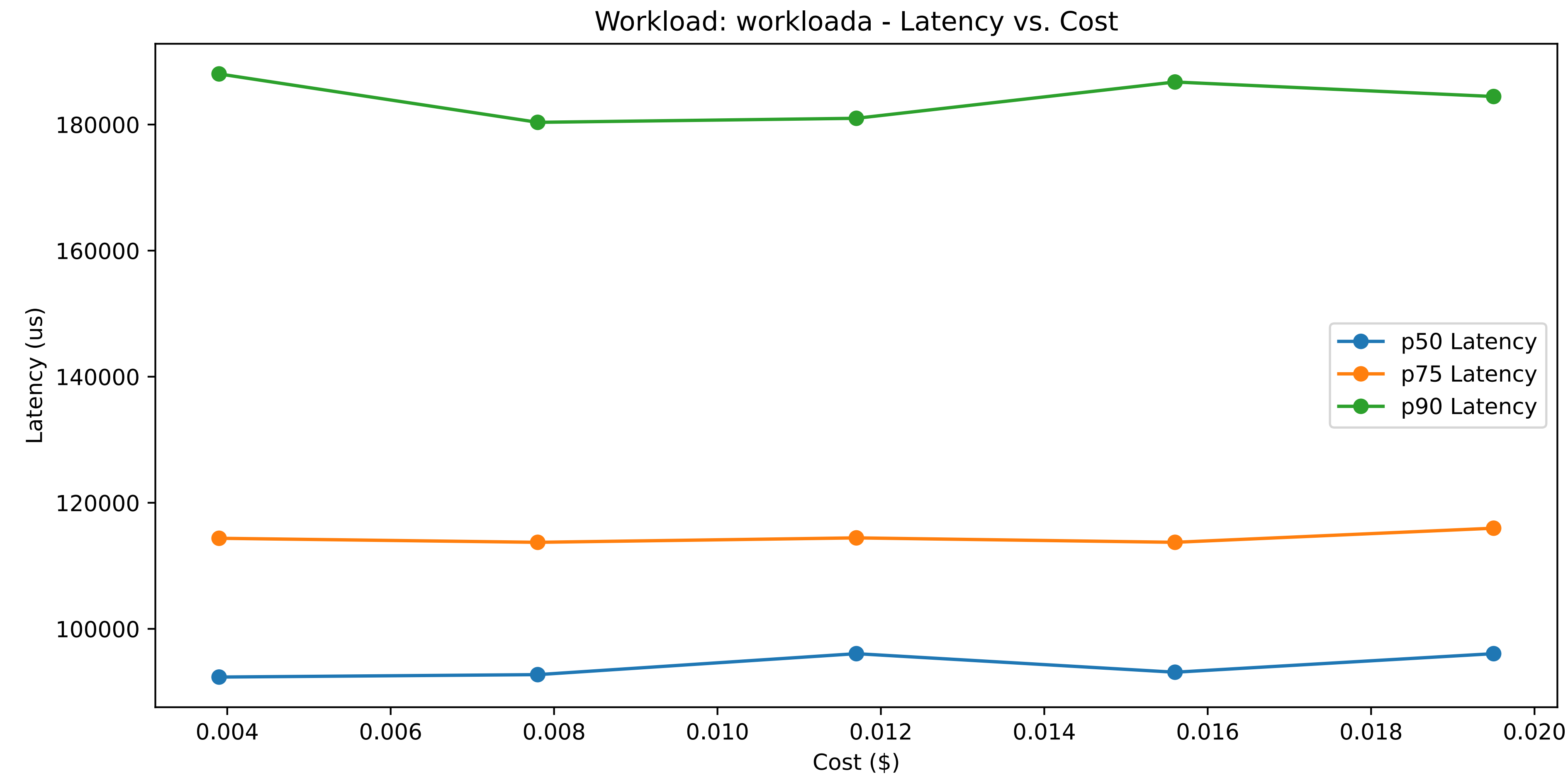
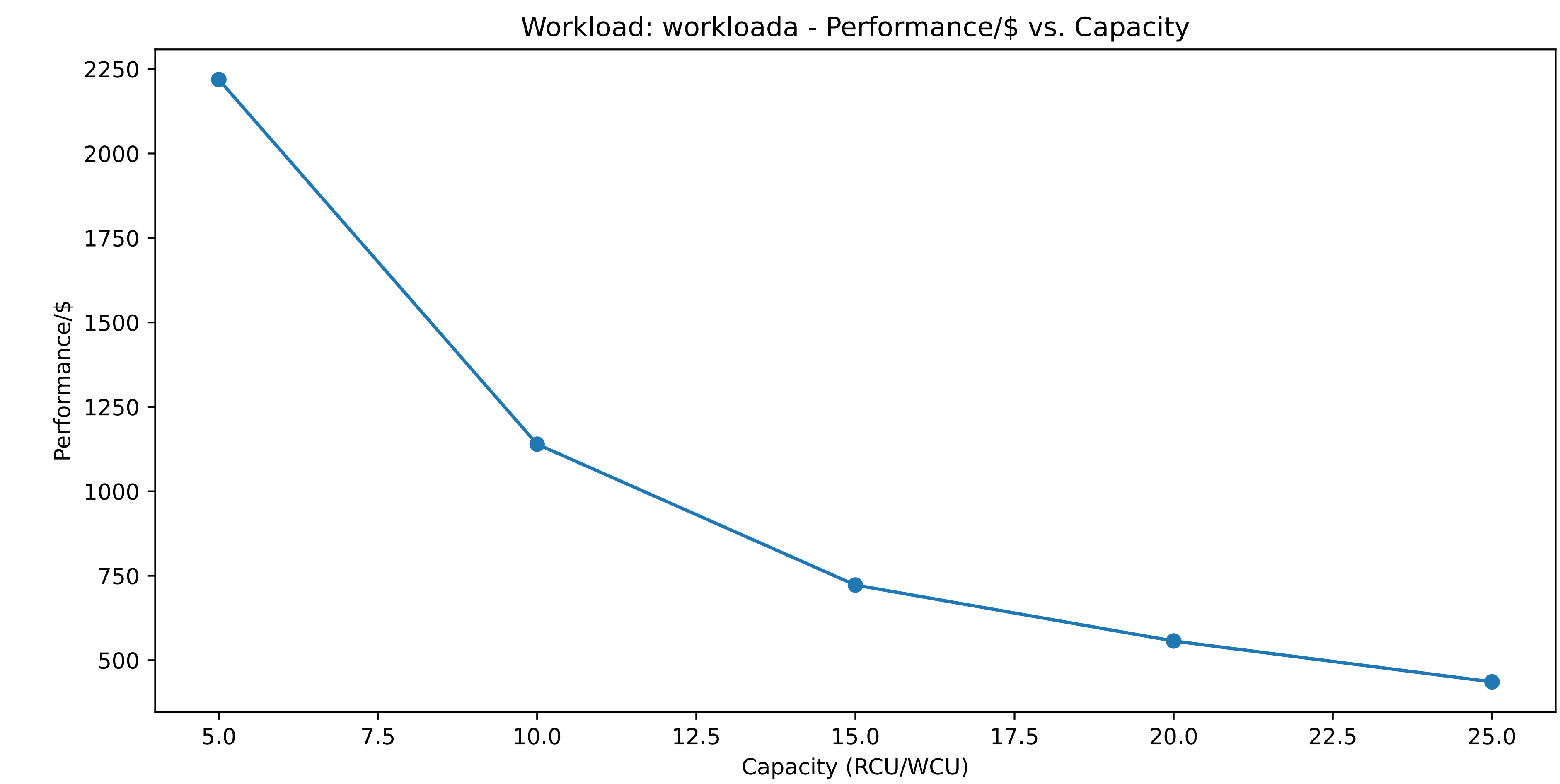
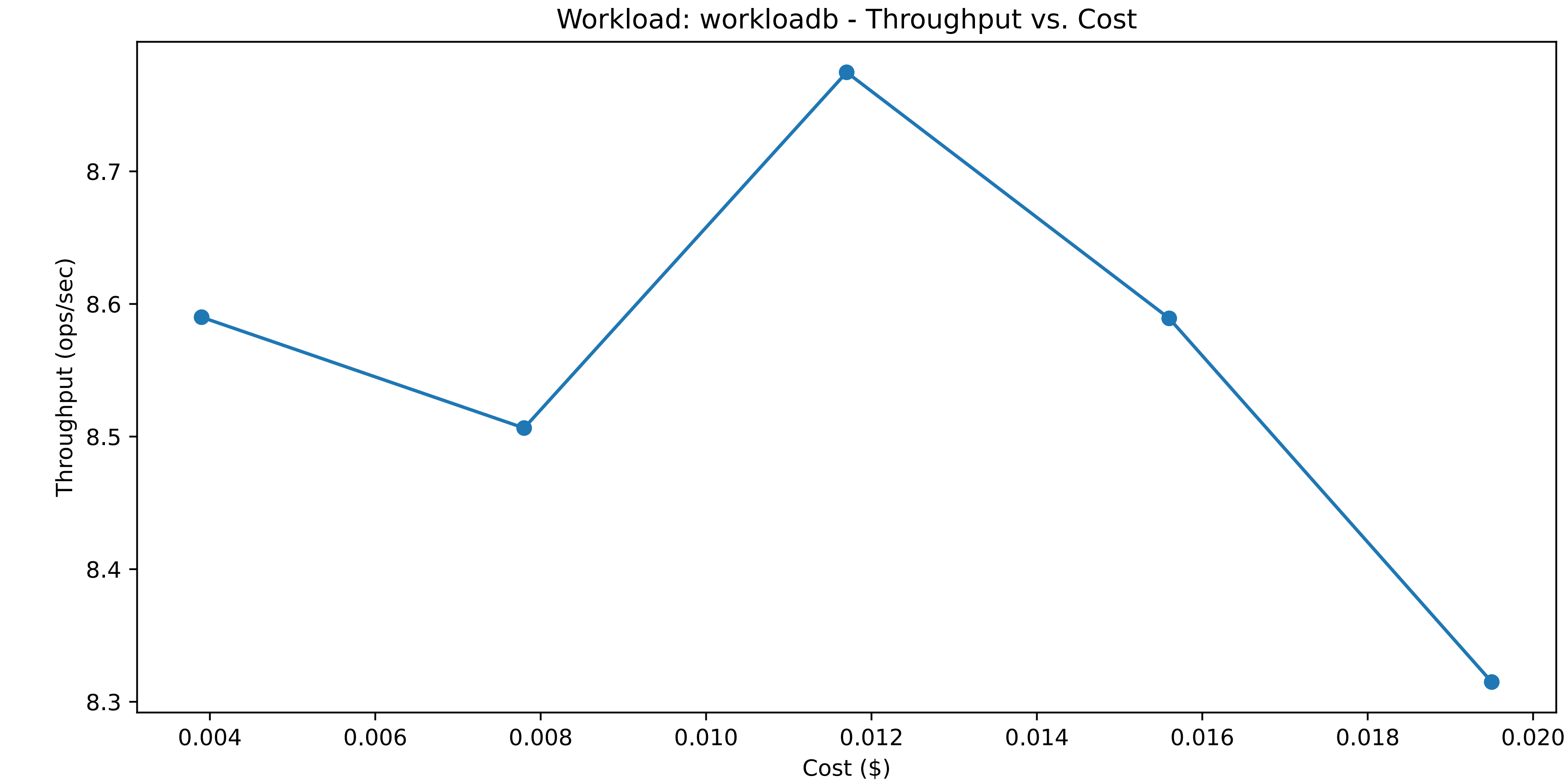
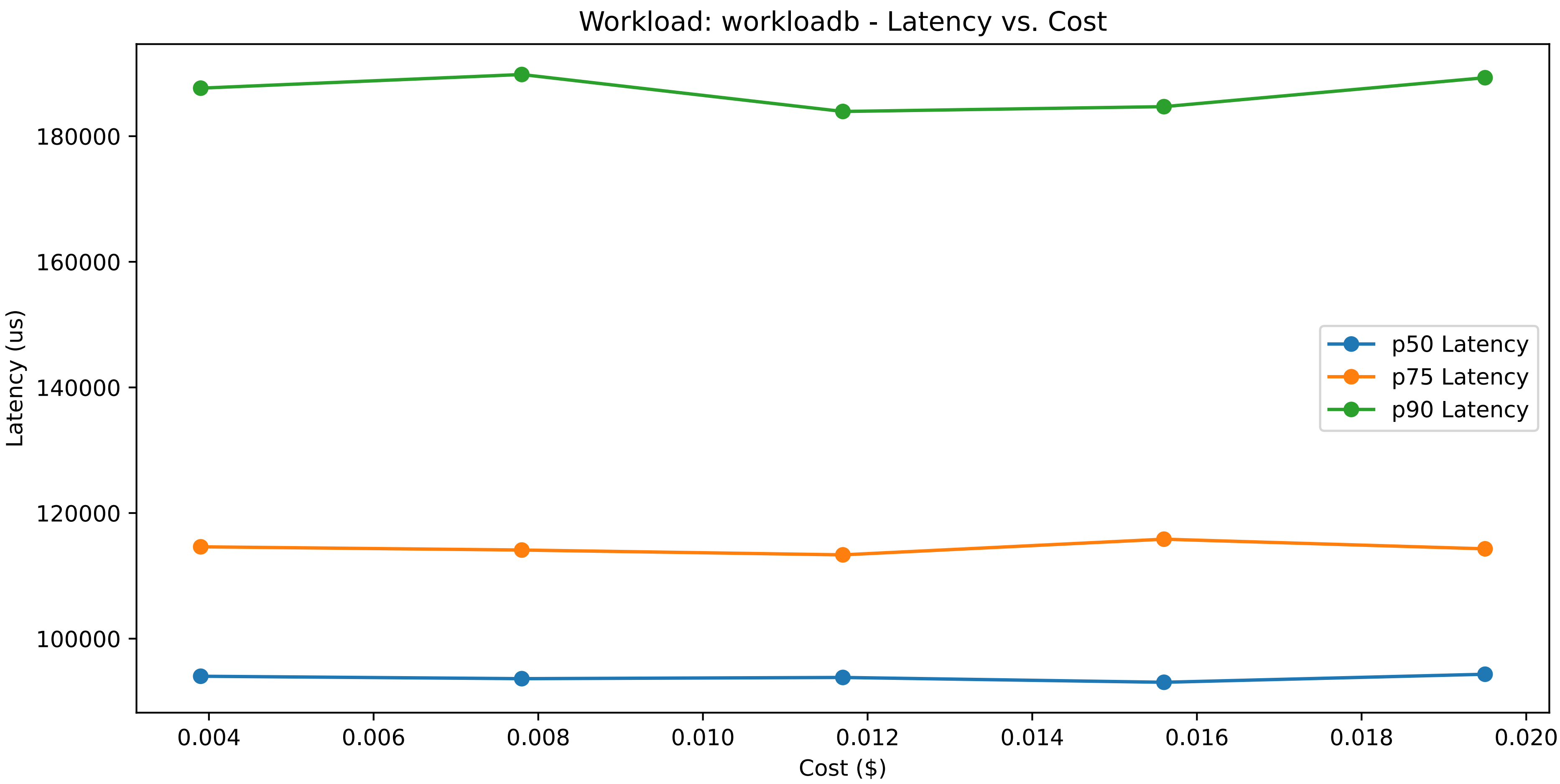
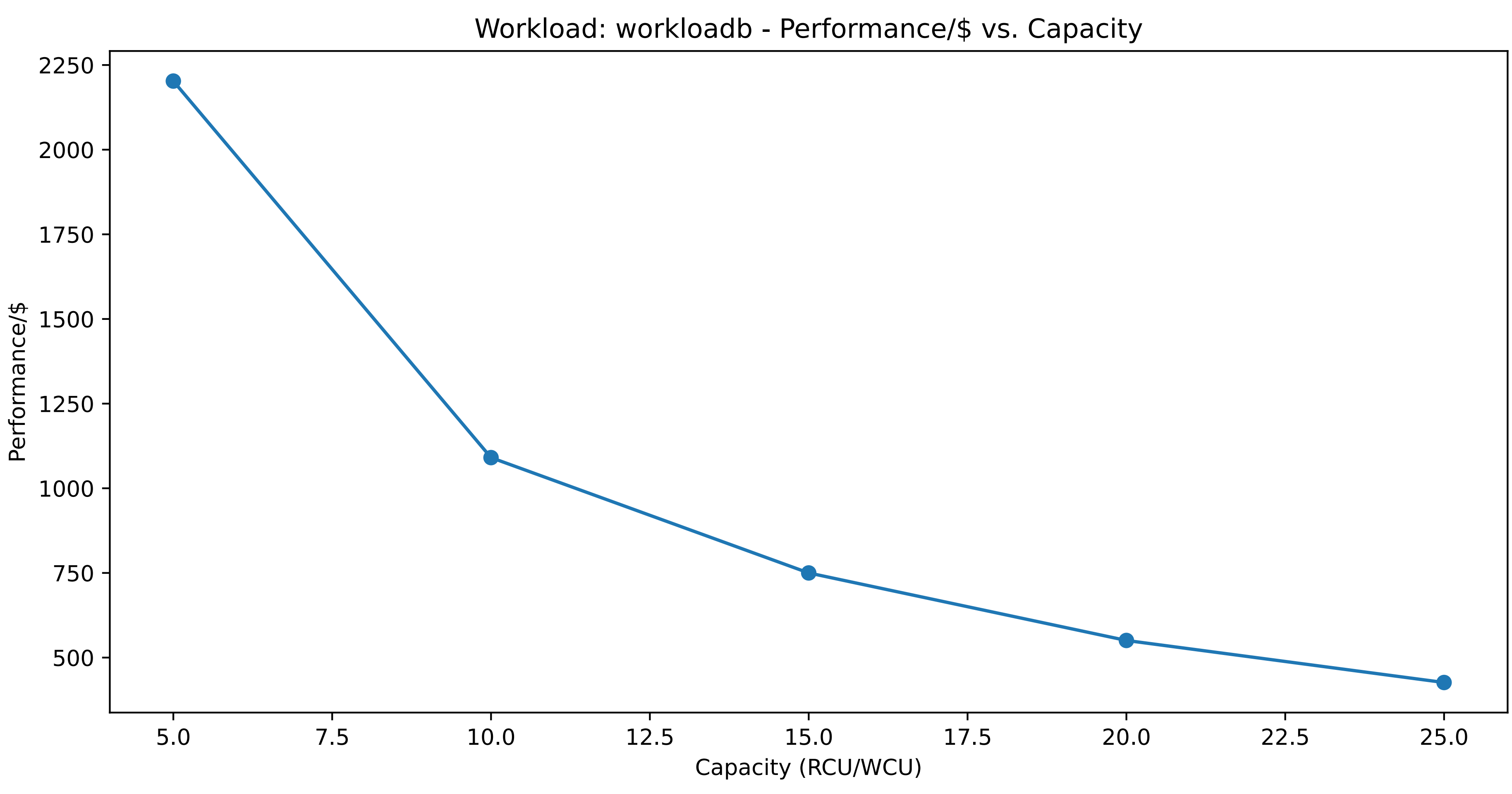

Workload: workloada - Throughput vs. Cost
Throughput (ops/sec)
Cost ($)
8.5
8.6
8.7
8.8
8.9
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020

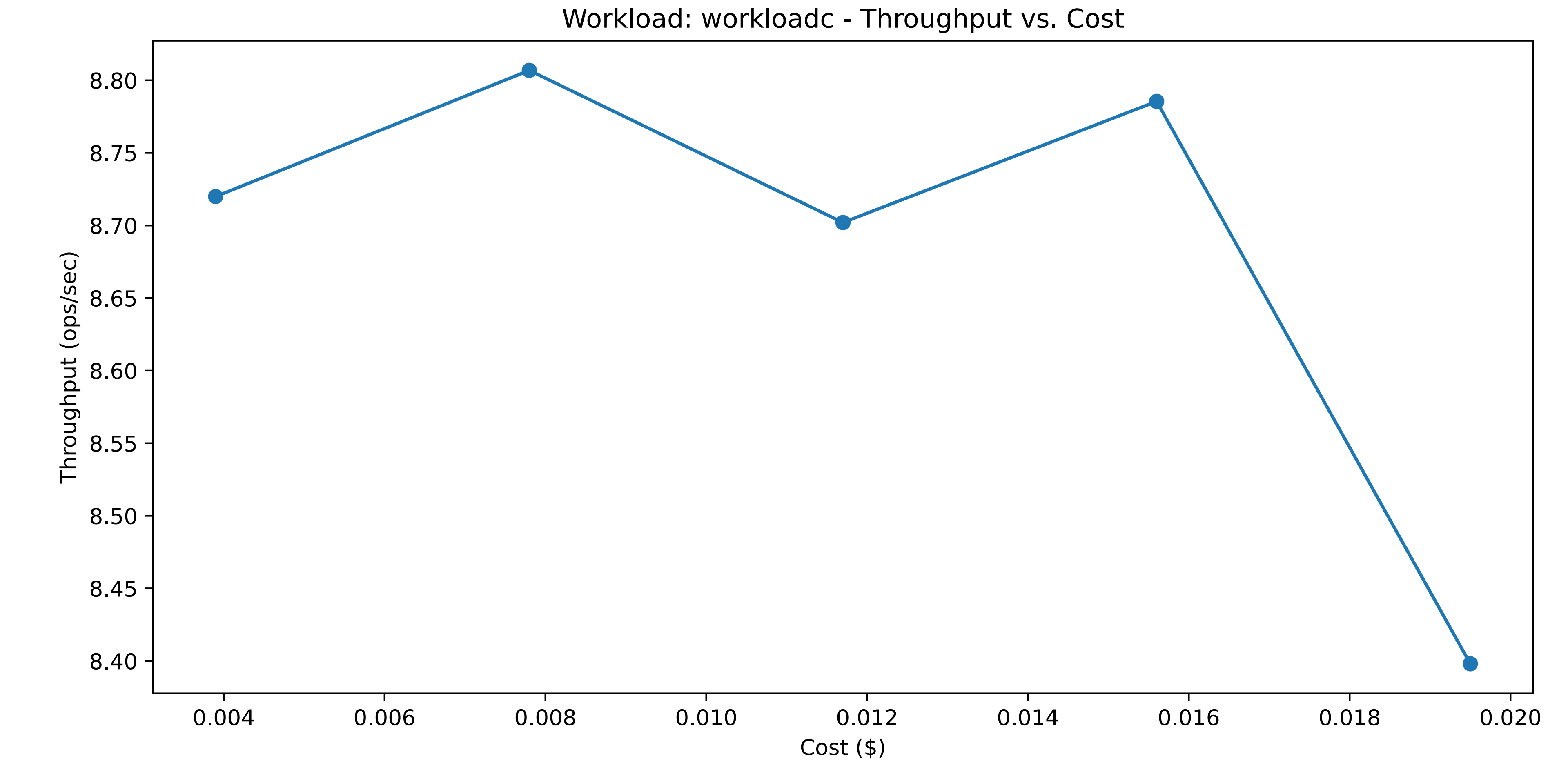

Workload: workloada - Latency vs. Cost
Latency (us)
Cost ($)
p50 Latency
p75 Latency
p90 Latency
100000
120000
140000
160000
180000
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020


Workload: workloada - Performance/$ vs. Capacity
Performance/$
Capacity (RCU/WCU)
500
750
1000
1250
1500
1750
2000
2250
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0

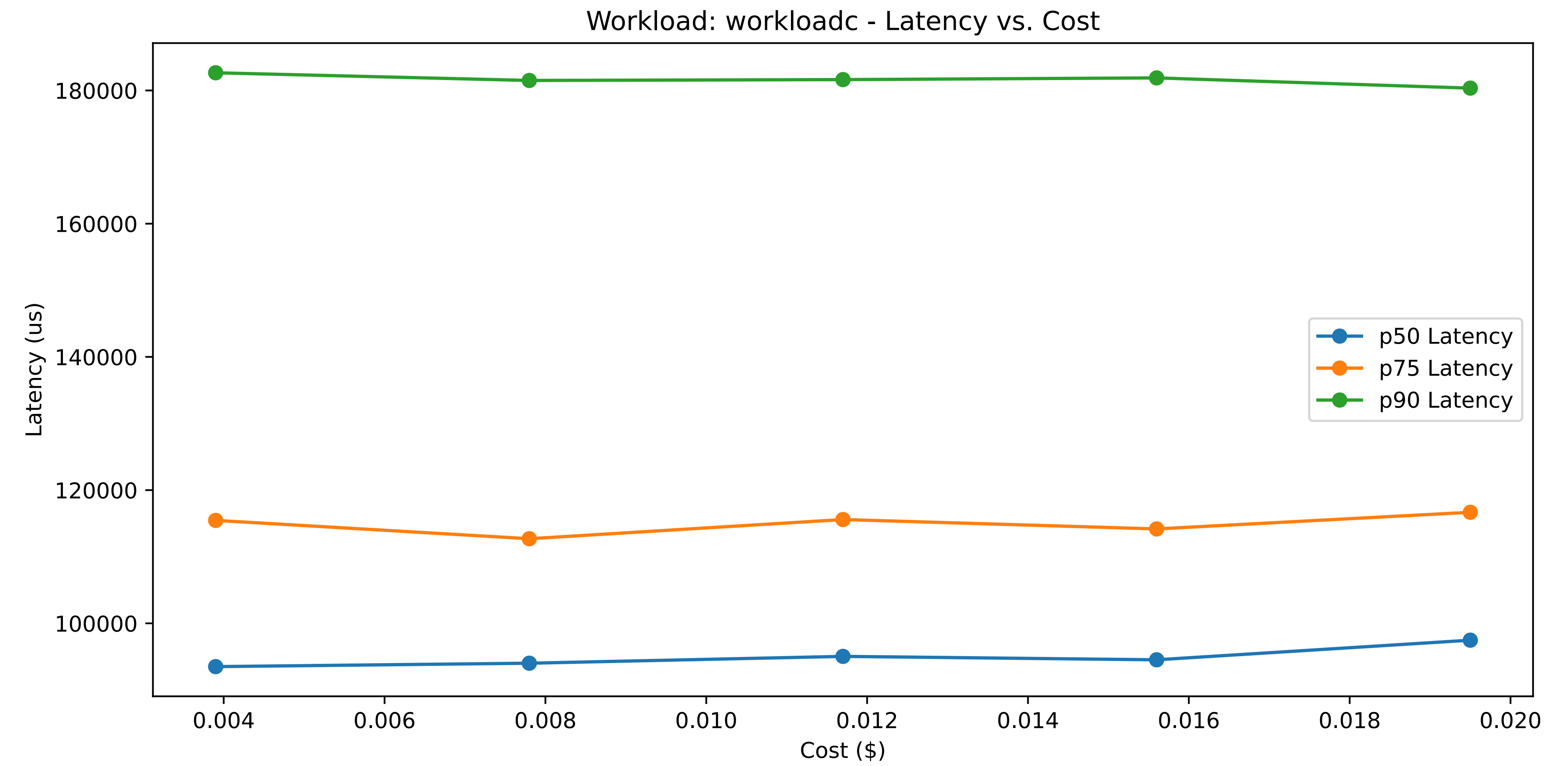

Workload: workloadb - Throughput vs. Cost
Throughput (ops/sec)
Cost ($)
8.3
8.4
8.5
8.6
8.7
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020


Workload: workloadb - Latency vs. Cost
Latency (us)
Cost ($)
p50 Latency
p75 Latency
p90 Latency
100000
120000
140000
160000
180000
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020

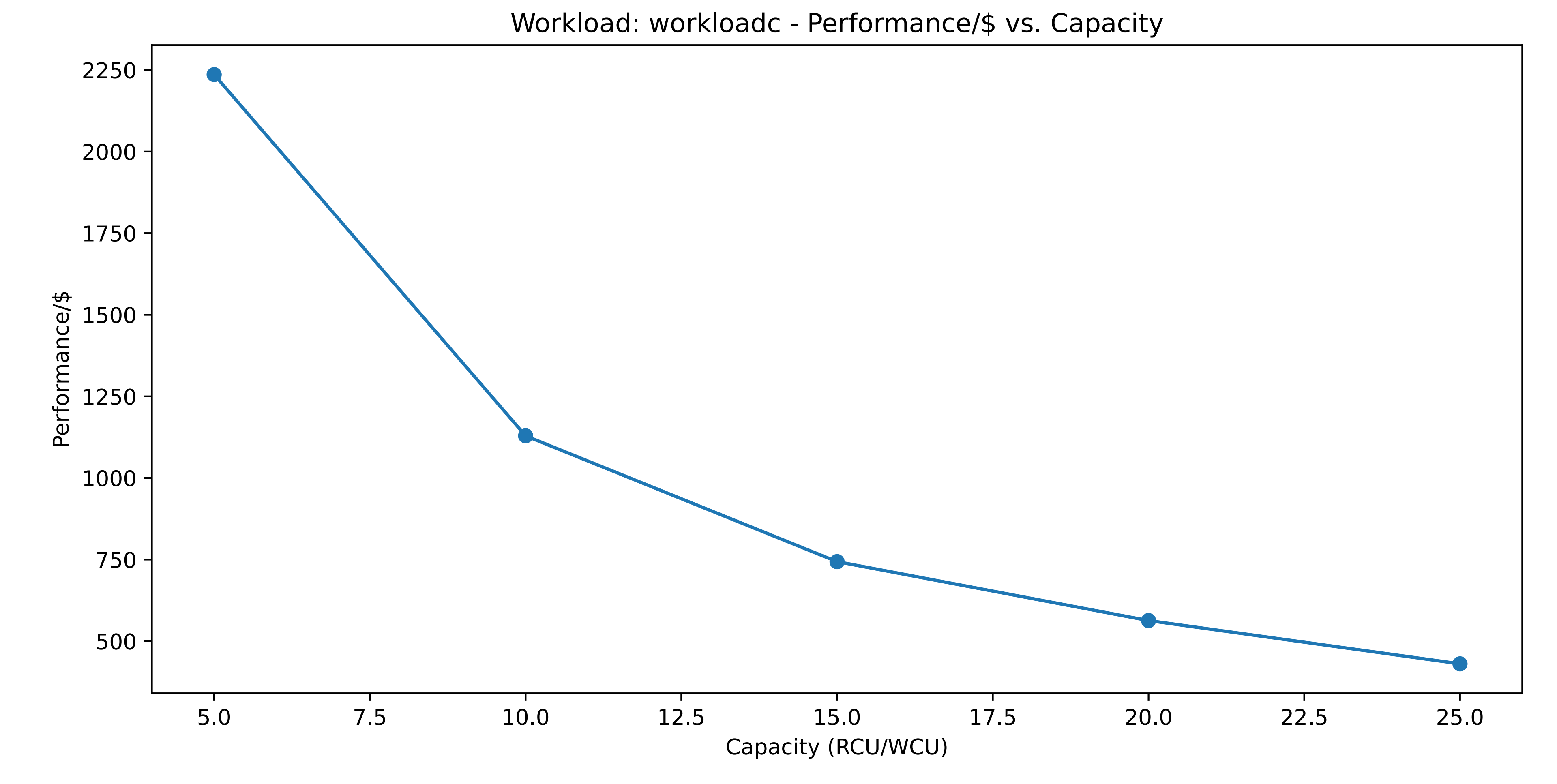

Workload: workloadb - Performance/$ vs. Capacity
Performance/$
Capacity (RCU/WCU)
500
750
1000
1250
1500
1750
2000
2250
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0



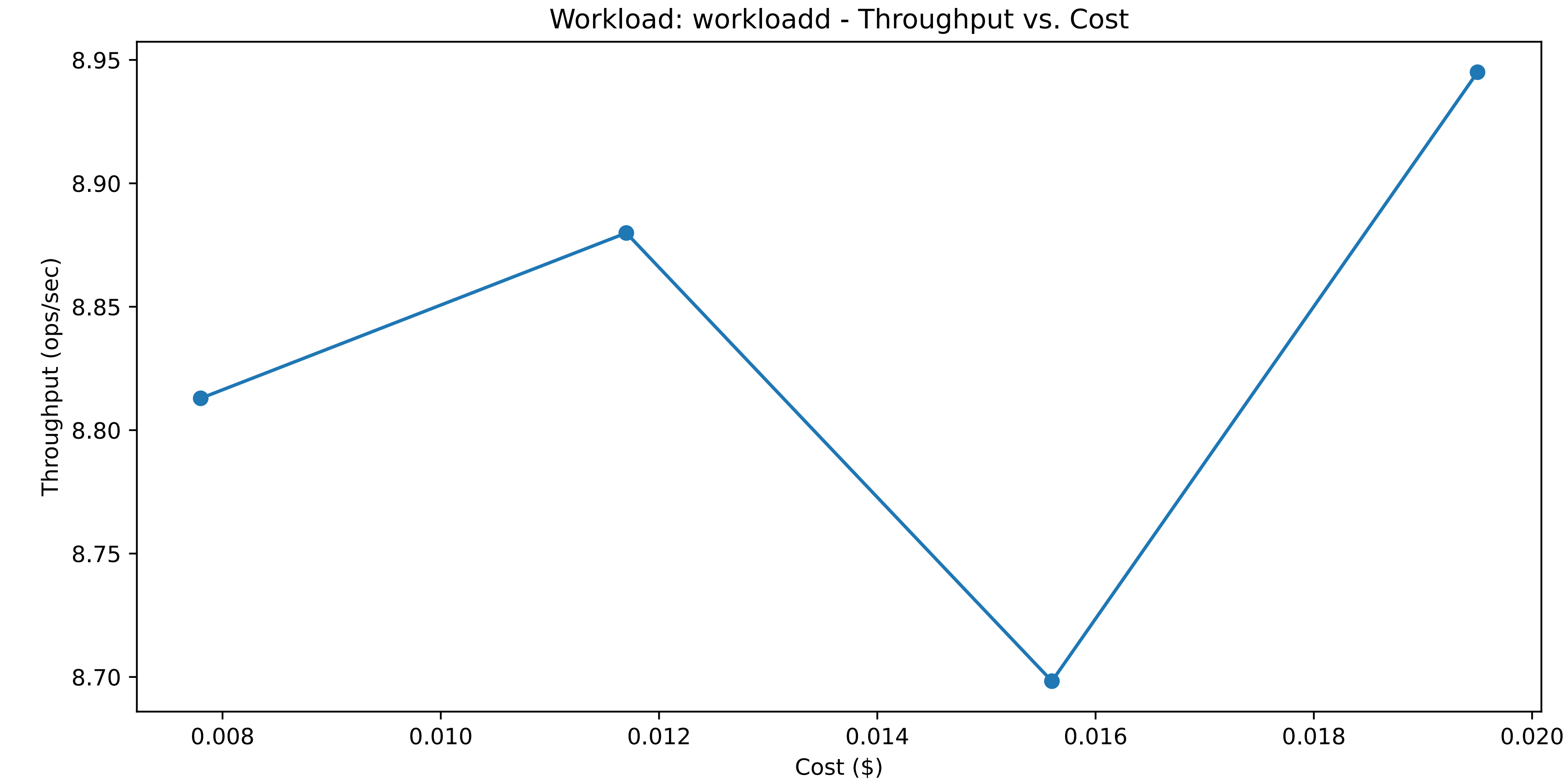
Workload: workloadc - Throughput vs. Cost
Throughput (ops/sec)
8.80
8.75
8.70
8.65
8.60
8.55
8.50
8.45
8.40
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020
Cost ($)

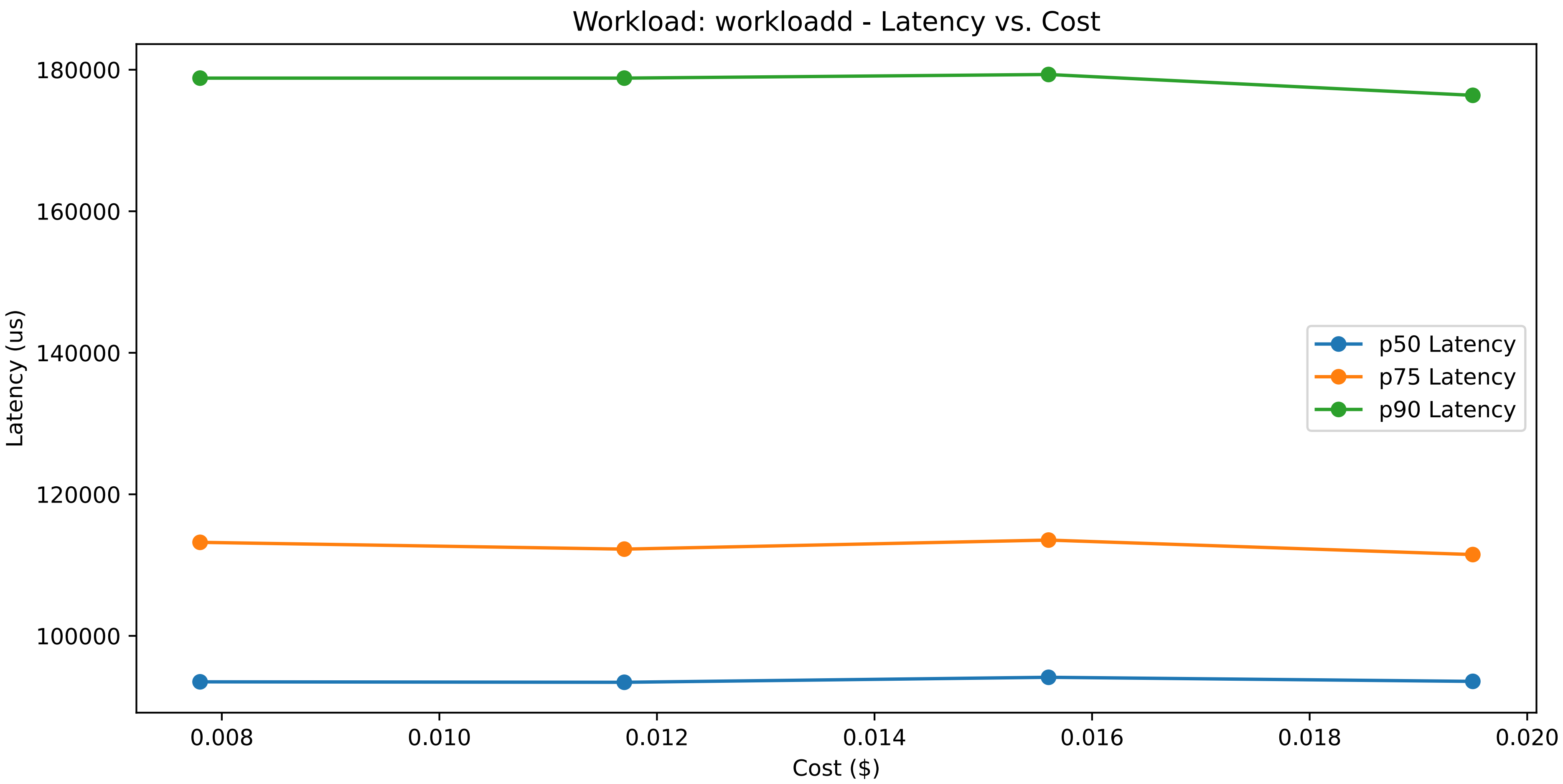


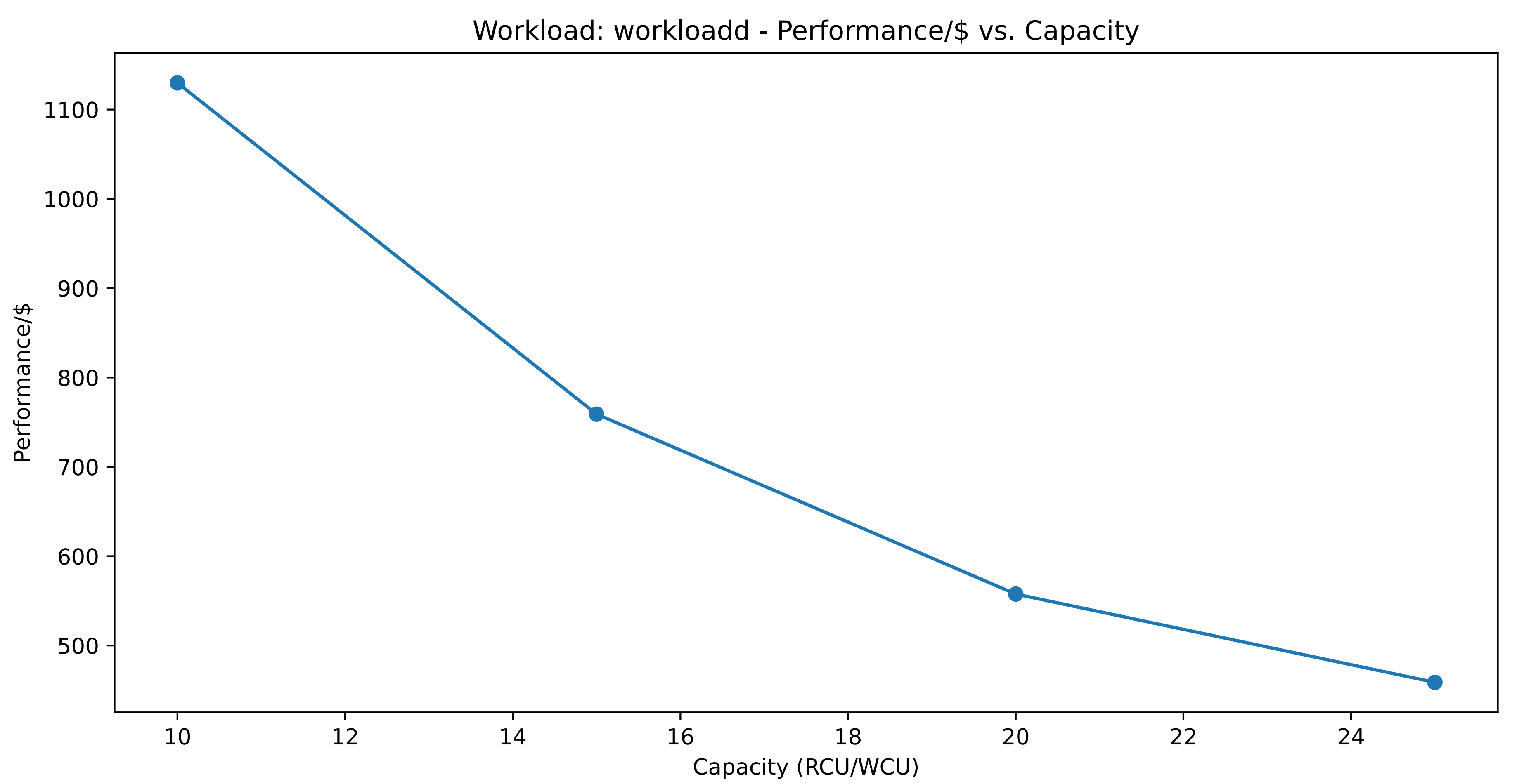
Workload: workloadc - Latency vs. Cost
Latency (us)
180000
160000
140000
120000
100000
p50 Latency
p75 Latency
p90 Latency
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020
Cost ($)

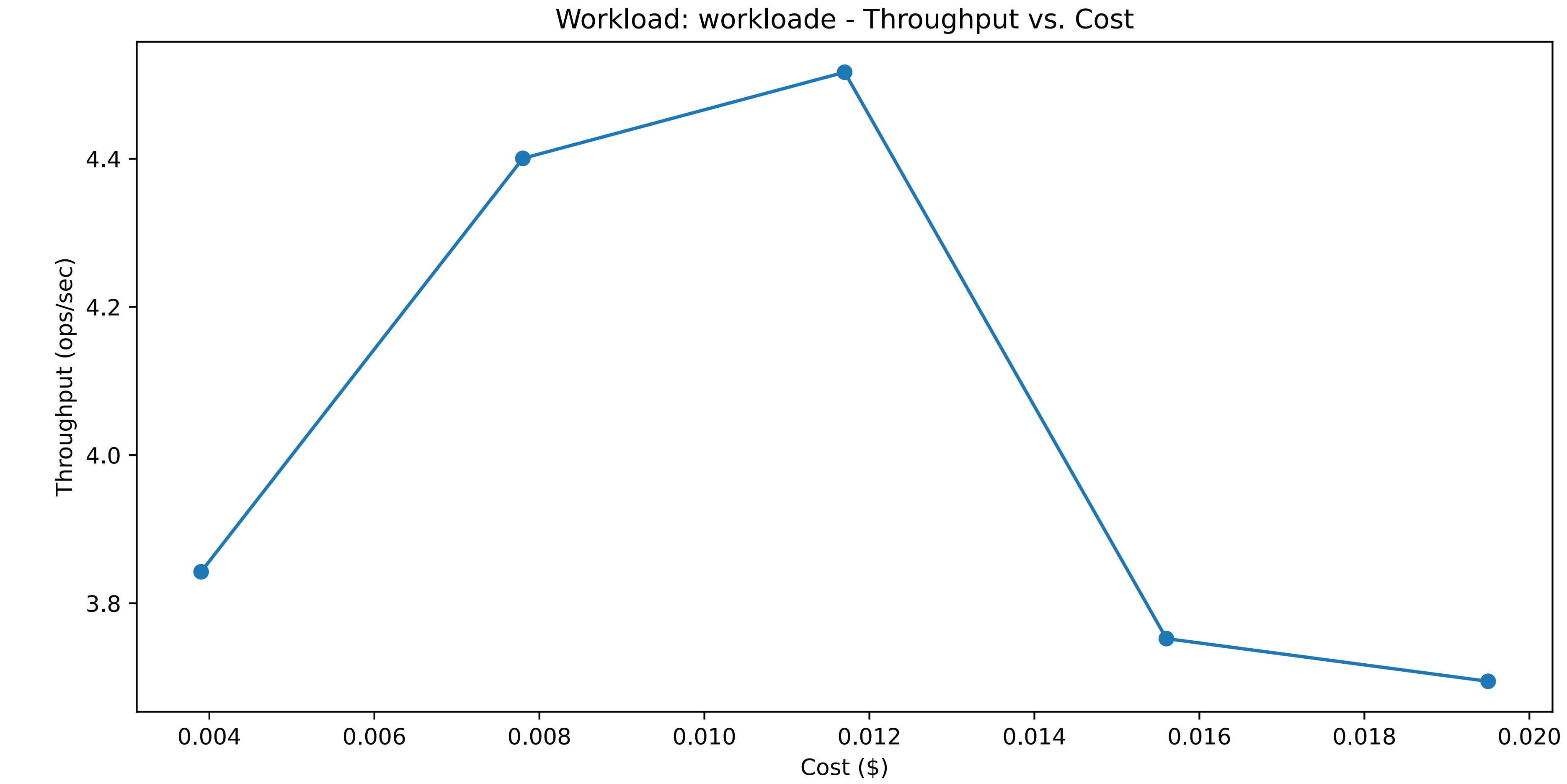


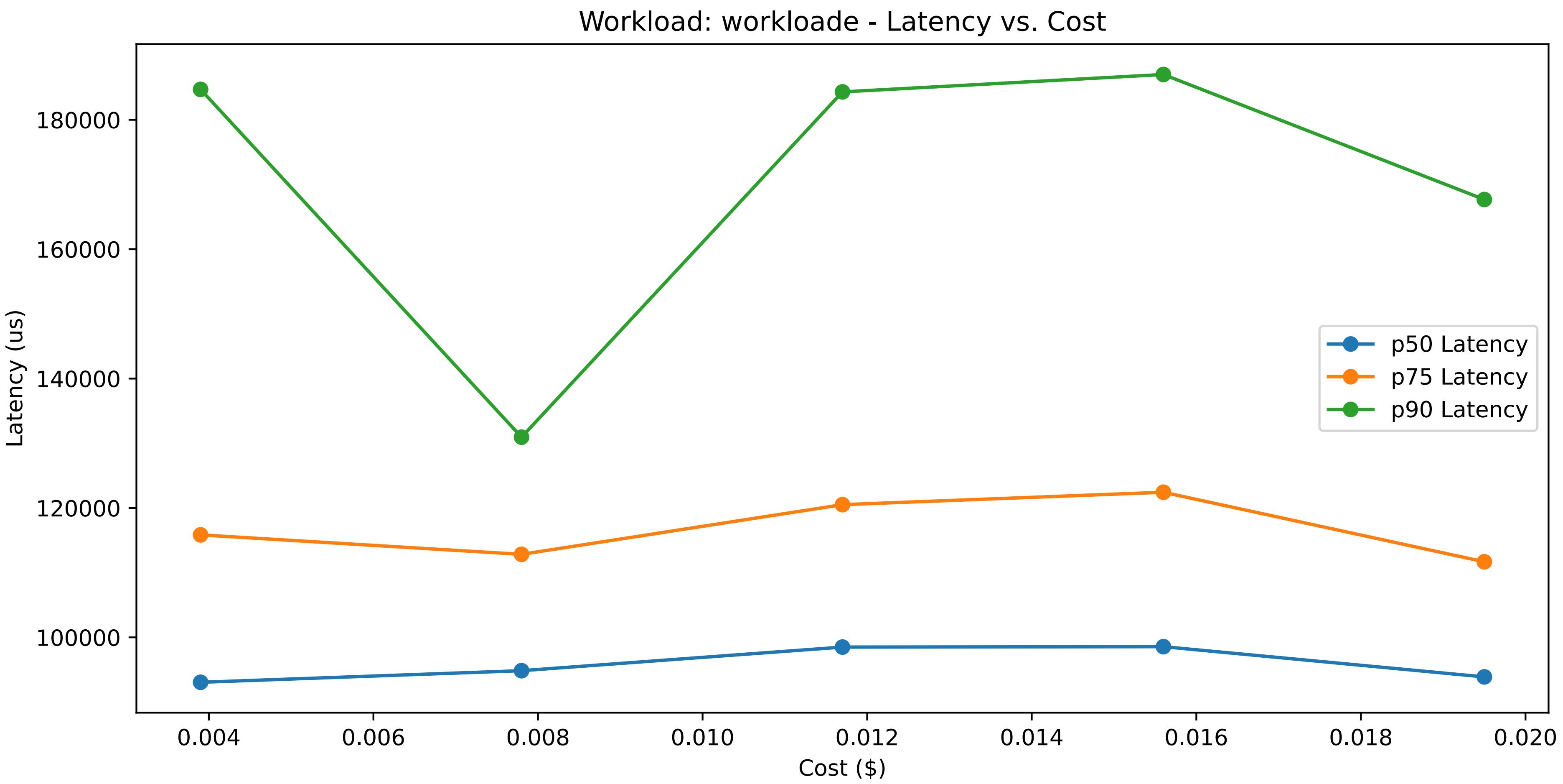
Workload: workloadc - Performance/$ vs. Capacity
Performance/$
2250
2000
1750
1500
1250
1000
750
500
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Capacity (RCU/WCU)

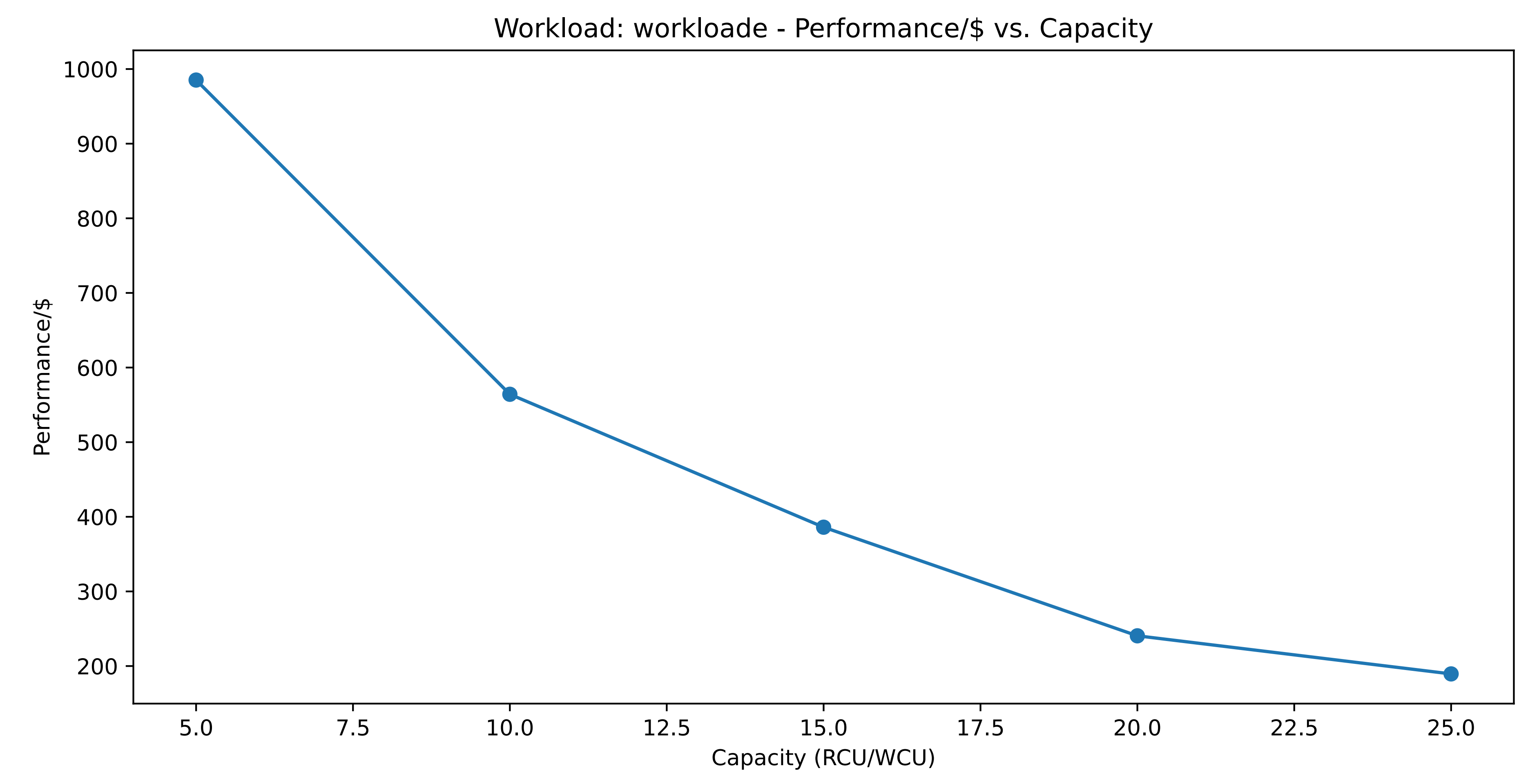

Workload: workloadd - Throughput vs. Cost
Throughput (ops/sec)
Cost ($)
8.95
8.90
8.85
8.80
8.75
8.70
0.008
0.010
0.012
0.014
0.016
0.018
0.020


Workload: workloadd - Latency vs. Cost
Latency (us)
Cost ($)
p50 Latency
p75 Latency
p90 Latency
180000
160000
140000
120000
100000
0.008
0.010
0.012
0.014
0.016
0.018
0.020


Workload: workloadd - Performance/$ vs. Capacity
Performance/$
Capacity (RCU/WCU)
1100
1000
900
800
700
600
500
10
12
14
16
18
20
22
24


Workload: workloade - Throughput vs. Cost
Throughput (ops/sec)
Cost ($)
3.8
4.0
4.2
4.4
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020


Workload: workloade - Latency vs. Cost
Latency (us)
Cost ($)
p50 Latency
p75 Latency
p90 Latency
100000
120000
140000
160000
180000
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020


Workload: workloade - Performance/$ vs. Capacity
Performance/$
Capacity (RCU/WCU)
200
300
400
500
600
700
800
900
1000
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0

Jagannath Prabhakaran

Analysis Report

The findings underscore the importance of tailoring capacity settings to the unique access patterns and demands of each workload to achieve the best balance between cost and performance. Periodically conducting such analyses is crucial to maintaining optimal configurations as application demands evolve.

Based on the graphs generated and the optimal point calculations, workloads a,b,c, and e had an optimal capacity of 5 RCU/WCU. However, for workload d the optimal point was 10 RCU/WCU, which suggests that in general lower capacities may be more efficient for Performance/$, but there are instances where a higher capacity is beneficial. For Latency, there was not too much discernible correlation between the difference in latency percentiles, but this is likely attributed to my network variability. Overall this underscores the importance of tailoring capacity to specific access patterns and demands to efficiently optimize performance and cost.