

# Pepperdata Increases CPU Utilization by 157% When Compared to AWS Auto Scaling

## TPC-DS on Amazon EMR Benchmark Results

### Introduction

The Pepperdata 2021 Benchmark Report demonstrates the efficacy of **Pepperdata Capacity Optimizer** compared to the AWS Custom Auto Scaling Policy. The benchmarking work in this report uses TPC-DS, an industry-standard big data benchmarking workload, and measures the following:

- Instance hours
- CPU utilization
- Memory utilization
- Price/performance, as defined by the cost to run the workload divided by the time it took to run

The purpose of this report is to demonstrate the performance improvements Capacity Optimizer can deliver using TPC-DS, a documented standard workload. Pepperdata ran the benchmarks “out of the box” and did not modify or recompile them using any special libraries.

### Key Findings

This report highlights three groups of findings that demonstrate that Capacity Optimizer can automatically:

- Optimize resource utilization, as measured by an increase in CPU and memory utilization and a decrease in instance hours.
- Decrease instance duration.
- Reduce costs by up to 50%.

**On average, Capacity Optimizer decreased both overall duration by 8% and instance hours by 38%, while increasing both CPU utilization by 157% and memory utilization by 38% when compared to AWS Custom Auto Scaling for the TPC-DS workload.**

### Big Data Benchmarking in the Cloud

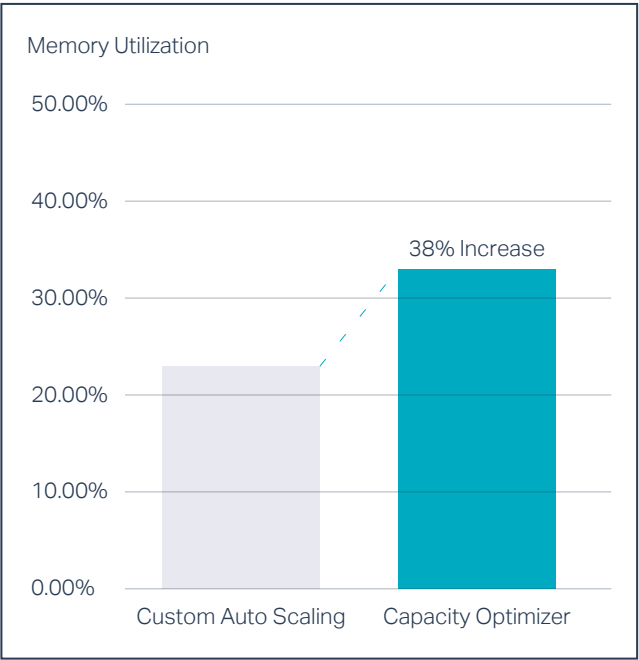
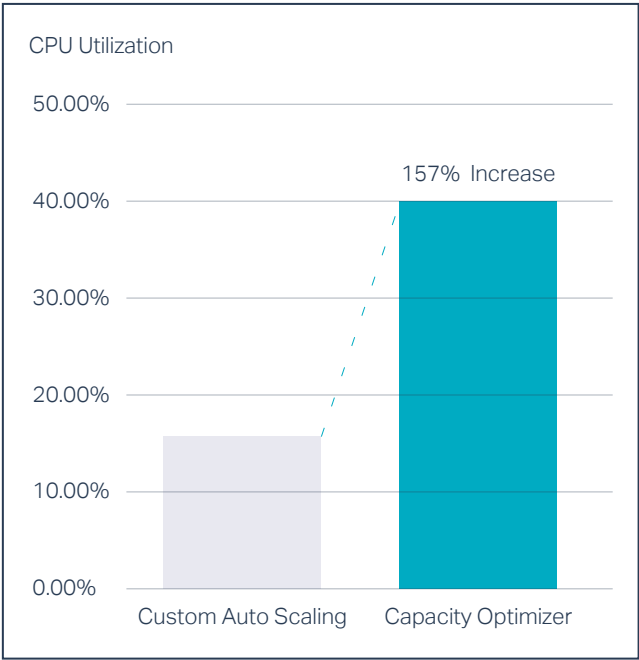
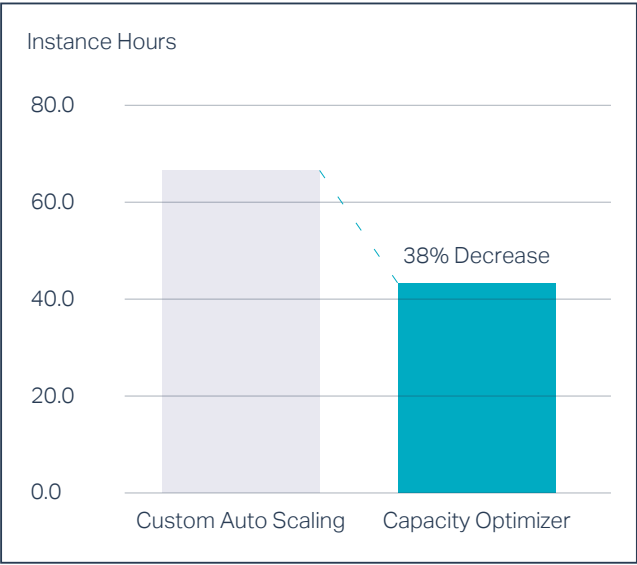
Benchmarking is the process of running a set of standard tests against some object to produce an assessment of that object’s relative performance. Imagine driving three different sports cars on the same course and measuring each car’s maximum speed, torque, and fuel consumption to compare the overall performance of the three cars.

This report covers our initial work with TPC-DS, the Decision Support framework from the Transaction Processing Performance Council. TPC-DS is a sophisticated, industry-standard big data analytics benchmark developed over decades and is a de facto standard for SQL including Hadoop. Our work is not an official audited benchmark as defined by TPC.

The TPC-DS workload consists of three distinct disciplines: Database Load, Query Run, and Data Maintenance. The query run is executed twice, once before and once after the data maintenance step. Each query run executes the same 99 query templates with different variables in permuted order, thereby simulating a workload of multiple concurrent users accessing the system<sup>1</sup>.

## Resource Utilization Results

On average, Capacity Optimizer decreased both overall duration by 8% and instance hours by 38%, while increasing both CPU utilization by 157% and memory utilization by 38% when compared to AWS Custom Auto Scaling for the TPC-DS workload:

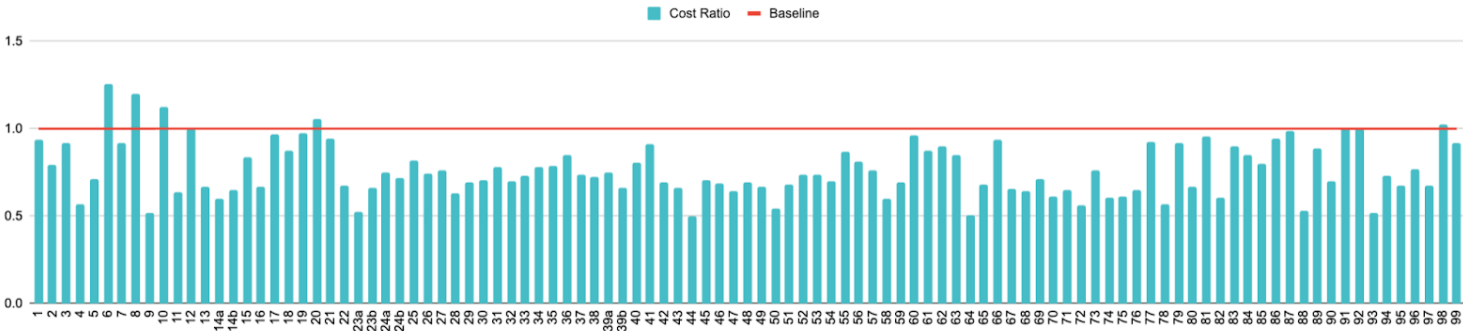


<sup>1</sup> [Why You Should Run TPC-DS: A Workload Analysis](#)

# Cost Ratio

Pepperdata compared the cost of the two options for each of the 103 queries. In the following chart, the baseline cost line represents the average cost of the queries with TPC-DS data using the AWS Custom Auto Scaling. Any queries where Capacity Optimizer provided a savings is represented by bars which run below the baseline. More than 90% of the queries which occurred while running Capacity Optimizer resulted in savings versus running the AWS Custom Auto Scaling policy alone.

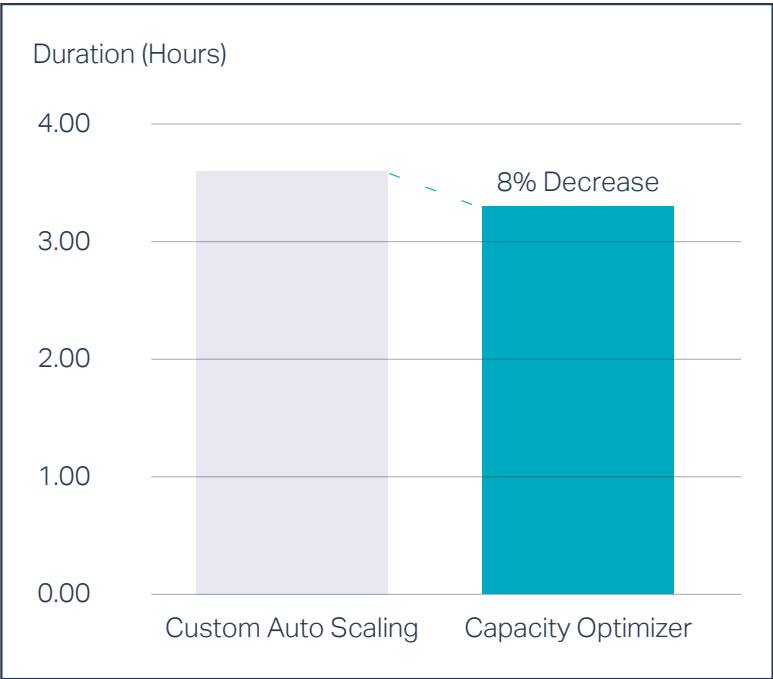
Cost Ratio by Query Before and After Capacity Optimizer



In this set of runs, the five queries where Capacity Optimizer did not perform as well as the AWS Custom Auto Scaling Policy tend to be the simpler and less complicated queries. These are TPC-DS Query 6, TPC-DS Query 8, TPC-DS Query 10, TPC-DS Query 20, and TPC-DS Query 98. Pepperdata excelled in the most complicated queries that reflect the most demanding real-world environments.

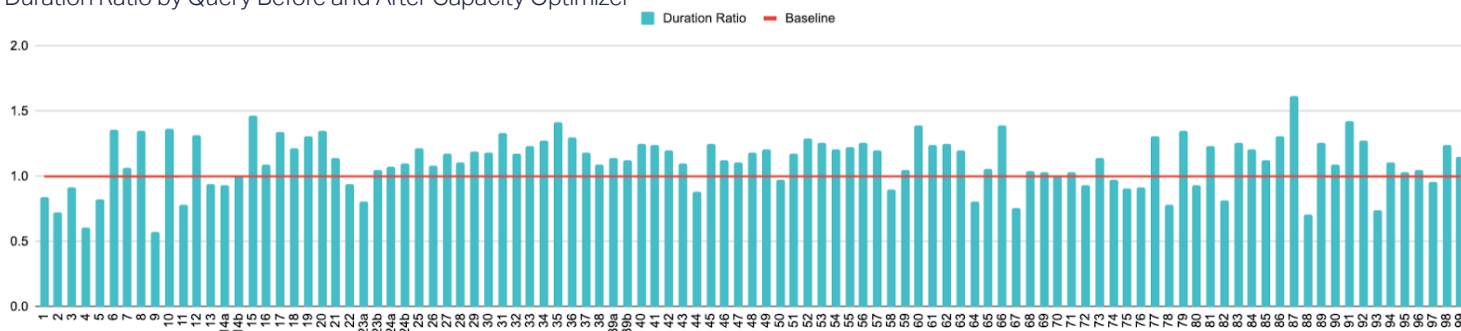
# Overall Duration

Although overall duration was not a primary metric of interest in evaluating the effectiveness of Capacity Optimizer, we did observe an approximate 8% decrease in the overall runtime of the entire suite of 104 queries, with the overall duration decreasing from 3.62 hours to 3.33 hours.



At the individual query level, the duration difference ranged from an increase of 38% with Capacity Optimizer to a decrease of 74% with Capacity Optimizer, as shown in the following graph. As in the previous graph, any queries where Pepperdata using Capacity Optimizer decreased the duration of the query is represented by bars which run below the baseline.

Duration Ratio by Query Before and After Capacity Optimizer



## Conclusion

Using the industry-standard benchmark dataset TPC-DS, Pepperdata was able to demonstrate that Capacity Optimizer provides an uplift over the Amazon AWS Custom Auto Scaling Policy in both CPU and memory utilization while decreasing instance hours and overall time to run the entire suite of 103 queries.

As more companies migrate big data workloads to the cloud, these findings have important implications for cost and resource management. A recent [survey](#) conducted by Pepperdata identified budget control as a primary challenge of organizations moving to the cloud. Without a lot of effort, Capacity Optimizer can help big data workloads in the cloud run more efficiently, resulting in substantial cost and resource savings. This helps make the cloud an even more attractive and viable option for big data workloads.

For information on the methodology and configurations used and referenced in this report, please contact us at [info@pepperdata.com](mailto:info@pepperdata.com).

## About Pepperdata

Pepperdata provides analytics performance management that guarantees SLAs and reliability for your infrastructure and apps.

©2021 Pepperdata Inc. All rights reserved. Pepperdata and the Pepperdata logo are trademarks or registered trademarks of Pepperdata Inc. All other trademarks are the property of their respective owners. Pepperdata reserves the right to change this document without notice. To ensure you have the latest version of this document, visit [www.pepperdata.com](http://www.pepperdata.com).



3945 Freedom Circle, Suite 920  
Santa Clara, CA 95054

**Start a Free Trial**

[www.pepperdata.com/trial](http://www.pepperdata.com/trial)

**Send an Email**

[eval@pepperdata.com](mailto:eval@pepperdata.com)