



Global Technology Firm Realizes up to 25% Cost Savings on Amazon EMR with Pepperdata

About the Client

The client is a subdivision of a large, multinational technology company that services a global customer base.

Challenge:

The subdivision's on-premises Hadoop footprint had grown over 20 percent year-over-year, causing scalability and reliability issues.

Cost reduction and operational efficiency were top priorities for the management team, while the operations team aimed to maintain high performance, reliability, and SLA adherence. The volatile nature of the subdivision's workloads, which frequently fluctuated tenfold, added further complexity.

Solution:

To help address its compounding growth challenges, the customer opted to migrate workloads to Amazon EMR. The company then enlisted Pepperdata on top of Amazon EMR to autonomously and continuously maximize its resource utilization and keep costs under control as the business expanded.

Results:

Once all the workloads had been migrated, the combination of gains from Pepperdata's contributions was able to deliver the customer up to **25 percent cost savings**.

A High-Growth Company Facing Scale Challenges

In 2022, the subdivision of a large, multinational company began migrating its workloads to Amazon EMR. Its on-prem Hadoop footprint had been growing more than 20 percent year over year and this infrastructure was facing scalability and reliability challenges.

The company was confident it would derive enormous value from the benefits of Amazon EMR, which closely parallel those of Apache Hadoop.

As the company began its migration, they were looking for additional opportunities to optimize their cloud cost, given the continually expanding nature of their workloads. The management team was also facing market pressures to reduce costs and increase operational efficiency wherever possible. At the same time, the operations team needed to maintain high levels of performance and reliability and adhere to SLAs. Finally, the uniquely volatile nature of the subdivisions workloads—which regularly expanded and contracted by a factor of ten—presented an additional challenge.



Pepperdata: The Easy Button for Cloud Cost Optimization at Scale

Pepperdata, the Cloud Cost Optimization Company, was tapped to address all these issues. Pepperdata had already been working with this customer for years, helping to optimize their on-prem Hadoop infrastructure for greater efficiency and throughput. The company had thus developed a strong level of trust in Pepperdata's ability to optimize the utilization of the company's valuable resources, whether on premises or in the cloud.

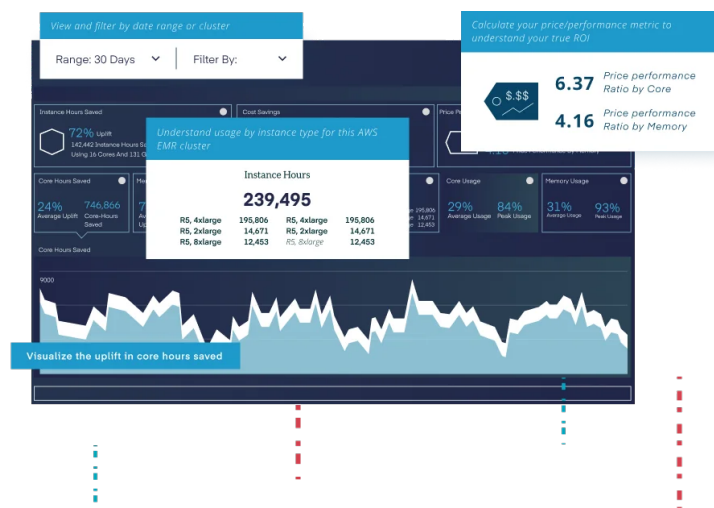
Pepperdata autonomously and continuously optimized the cost of the company's cloud deployments in two unique ways:

1. Pepperdata Increased Utilization and Reduced Waste

Capacity Optimizer improved node efficiency because it works in real time to enable the YARN scheduler to launch tasks based on actual hardware utilization rather than relying on standard allocations, which by design contain waste in the form of overhead.

By enabling this customer's scheduler to automatically set utilization levels based on actual ever-changing workloads in real time instead of arbitrary allocation levels, Pepperdata reduced the waste in their environment. This then reduced their overall cost since the company now paid only for the capacity used rather than for allocated capacity. For this company, Pepperdata allowed YARN to reclaim unused allocations to increase hardware utilization by roughly 40 percent. Pepperdata did this safely, using machine learning to intelligently and continuously maintain the cluster in a sweet spot of optimal utilization, which cannot be done manually. This Pepperdata-enhanced utilization level represented a 213 percent increased efficiency level compared to the ability of native YARN scheduling based on standard allocations alone.

In addition, by communicating with the scheduler about actual available resources, Pepperdata Capacity Optimizer enabled an average memory uplift of 43 percent during busy times, leading to 46 percent more running containers than without Pepperdata Capacity Optimizer.



2. Pepperdata Enabled Dynamic Autoscaling

Pepperdata Capacity Optimizer's intelligent automation ensured that existing nodes in the cluster were fully utilized by communicating to the YARN scheduler that more resources were actually available than the scheduler was aware of. This enabled the scheduler to automatically add more pending jobs to the existing nodes, increasing throughput without adding cost. Even when two very large new jobs were added to the cluster, no new nodes had to be added since Capacity Optimizer had enabled the scheduler to understand that more space was actually available.

Capacity Optimizer's ability to increase job throughput and performance by automatically reclaiming underutilized resources from existing nodes saved the customer significant resources by avoiding unnecessary scale up operations and its associated costs. By working autonomously and in real time around the clock to tune the autoscaler, Capacity Optimizer works at a level that cannot be matched by even the most diligent engineer.

The Net Result: Up to 25 Percent Cost Savings

Pepperdata Capacity Optimizer was activated over a two-week period. This rollout cadence was chosen to be fairly conservative so as not to disrupt the scale of the customer’s production workloads. Along the way, the return on investment was calculated in a two-step fashion. The first set of gains was achieved by implementing Capacity Optimizer alone. A second set of gains was achieved by enabling Pepperdata’s Autoscaling Optimization. At the end of two weeks, once all the workloads had been migrated, and once both optimization steps had been fully implemented, the customer achieved close to **25 percent cost savings from Pepperdata on top of their Amazon EMR cost savings.**



Future Directions: Onward and Upward

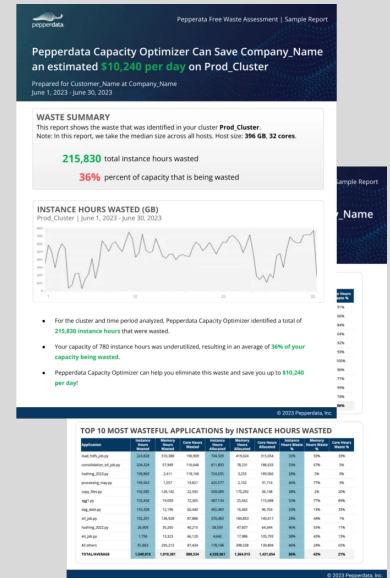
Based on the successes achieved in this two-week deployment, the customer is currently planning on migrating additional on-prem workloads to Amazon EMR. They will continue to deploy Pepperdata on their Amazon EMR workloads to achieve the ultimate in cloud cost optimization.

Would You Like to Achieve Similar Results?

Whether or not you’re a global enterprise like this customer, Pepperdata can help you achieve similar results on your Amazon EMR cluster. To learn more about Pepperdata’s capabilities for Amazon EMR, download the Pepperdata for Amazon EMR datasheet.

You can also sign up for a free two-day Customized Waste Assessment that will provide a detailed report of total estimated instance hours, top 10 most wasteful queues, and estimated savings from running Pepperdata Capacity Optimizer in your environment.

For more information, or to get started, contact Pepperdata at sales@pepperdata.com.



About Pepperdata

Pepperdata is the only cost optimization solution that delivers up to 47% greater cost savings—continuously and in real-time—on Amazon EMR and EKS with no application changes or manual tuning. Our customers include the largest, most complex, and highly-scaled clusters in the world, at top enterprises such as Citibank, T-Mobile, Autodesk, Securonix, Royal Bank of Canada, and those in the Fortune 5. For more information, visit pepperdata.com.



530 Lakeside Drive
Suite 170
Sunnyvale, CA 94085



Start a Free PoV

www.pepperdata.com/free-cost-optimization-proof-of-value



Send an Email

eval@pepperdata.com