

# Top Online Bank Realizes 30-75 Percent Cost Reductions on Amazon EMR, Invests Savings in Further Expansion

## A Leading Online Bank With a Cost-Cutting Mandate

For nearly two decades, a large online bank has been providing its millions of customers across the US with a broad range of financial services. Underpinning these efforts are the bank's advanced credit decisioning and machine-learning models that are based on more than 150 billion cells of data.

With this significant data footprint, the bank is a major Amazon EMR customer and is running a mix of Apache Spark and Apache Hive workloads—approximately 75% Spark and 25% Hive.

The bank has a corporate initiative to maximize efficiency where possible while maintaining or exceeding defined SLAs. As part of this initiative, the bank plans to migrate all their Hive workloads to Spark for greater efficiency and eventually migrate their Spark workloads to Amazon EKS.



## Seeking Solutions to the Increasing Challenge of Cost Control

Given the bank's corporate savings mandate, its executives were receptive to a variety of cost optimization options, especially those that could be implemented quickly and without requiring development teams to build custom optimization software.

The bank was introduced to Pepperdata by their AWS Senior Account Manager, who had previously observed the power of Pepperdata in improving efficiency and reducing costs by up to 47 percent in other AWS customer environments. As part of this evaluation, the bank's large Amazon EMR environment was identified as an area of potential savings.

Transacting with Pepperdata through the AWS Marketplace for a free proof of value (POV), the bank installed Pepperdata on twenty-two of its development and production Amazon EMR on EC2 clusters. Pepperdata Capacity Optimizer was enabled on seven of the clusters for cost optimization activities. Pepperdata observability features were deployed on the remaining fifteen to provide both cluster- and application-level visibility.

## Pepperdata Capacity Optimizer Delivers Transformative Results

Once installed on the bank's Amazon EMR clusters, Pepperdata Capacity Optimizer went to work without delay. Instance hours were reduced immediately, resulting in a demonstrated savings of 30-75 percent per cluster. The following graph shows the effect of enabling Capacity Optimizer on the number of instance hours required to process the bank's workloads.

Capacity Optimizer automatically identified application waste and returned these resources to the Amazon EMR scheduler, so that the scheduler did not need to spin up more instances at additional cost. These reduced instance hours and reclaimed memory hours translated directly into lowered cloud costs. The average cost savings of \$140,000 per month across all Capacity Optimizer-enabled clusters directly helped the bank achieve its corporate savings mandate. These savings results were realized immediately and automatically, without any code changes to the company's Spark or Hive applications, and without redirecting any engineering resources away from existing activities.

After enabling Capacity Optimizer, a senior manager on the bank's FinOps team observed that the company-wide Amazon EMR spend that had previously been on a steady increase started to decrease for the first time ever in the company's history.

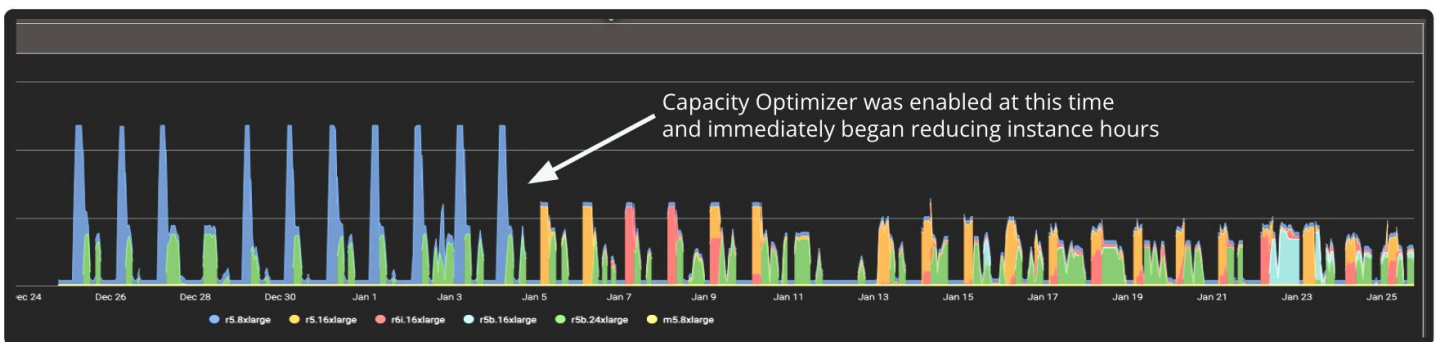


Figure 1: Capacity Optimizer immediately reduced the number of instance hours required to run the bank's workloads

## Pepperdata: Autonomous Cost Optimization Plus Industry-Leading Observability for Apache Spark

Pepperdata autonomous cost optimization delivers **30-47 percent (or more!) additional efficiency and cost savings** for data-intensive workloads such as Apache Spark on Amazon EMR with no application changes. Using patented algorithms, Pepperdata Capacity Optimizer autonomously optimizes CPU and memory in real time with no application code changes.

Capacity Optimizer automatically analyzes resource usage in real time to identify nodes where more work can be done. It then communicates this insight to the system scheduler, which then adds tasks to nodes with available resources and to spin up new nodes only when existing nodes are fully utilized. The result: **CPU and memory are autonomously and continuously optimized**, without the need for recommendations to be applied manually, safely eliminating the need for ongoing manual tuning.

**Pepperdata pays for itself**, immediately decreasing instance hours/waste, increasing Spark utilization, and freeing developers from manual tuning to focus on innovation.

Alongside autonomous cost optimization, Pepperdata also provides the option for full-stack observability and real-time insights into Apache Spark workloads. The Pepperdata observability dashboards offer industry-leading visibility into application performance, cluster health, and cost metrics that are not readily available through general-purpose monitoring or out-of-the-box performance management tools.

Pepperdata provides the deep, real-time, single-pane-of-glass views to quickly and accurately diagnose, troubleshoot, and resolve both cluster-wide and low-level application issues without having to cross-reference multiple dashboards. Executives, platform teams, and applications teams can use this knowledge to improve decision making about their Amazon EMR and Amazon EKS environments.

## Observability Complements Cost Optimization for Improved Performance

On the remaining clusters, Pepperdata's observability dashboards and data have helped the bank's operations team identify and resolve issues faster, better understand the details of their infrastructure, and measure performance against SLAs. In fact, Pepperdata's observability dashboard even helped the bank identify a critical bug in related software that they are resolving to further improve performance.

Following the successful deployment of Capacity Optimizer, one of the senior implementation leads commented, "Pepperdata promised no code changes, and we have not made a single code change. We haven't needed to work with a single developer yet to deploy Pepperdata."

Based on successes in both cost mitigation and observability, the bank plans to increase its rollout of Pepperdata Capacity Optimizer across additional clusters over time. All of the gains the bank achieved with Pepperdata required no application changes, no developer intervention, and no manual tuning.

“*Pepperdata promised no code changes, and we have not made a single code change. We haven't needed to work with a single developer yet to deploy Pepperdata.*”  
—Senior Implementation Lead, Top Online Bank

## Savings Fuel Future Expansion Plans Across Amazon EMR and Amazon EKS

Based on the successes to date, the bank expects to achieve an automatic 30 percent cost savings on Amazon EMR moving forward. Now that the bank's data-intensive workloads are being optimized both automatically and continuously to deliver the best price/performance, the bank is in a position to further expand its Amazon EMR footprint with new workloads.

The bank plans to allocate savings from the Pepperdata installation toward migrating its Spark workloads to Amazon EKS for reduced cost and improved efficiency, again leveraging Pepperdata for additional savings gains. Amazon EKS is an environment where Pepperdata has demonstrated 41.8 percent cost savings in [benchmarking environments](#) and upwards of 33 percent cost savings in [customer production environments](#). The bank's operation team also plans to use Pepperdata's observability dashboards to help inform this migration and measure the efficiency of their new Amazon EKS environment post migration.

## About Pepperdata

Pepperdata Capacity Optimizer is the only cost optimization solution for data-intensive workloads such as Apache Spark that delivers 30-47% greater cost savings continuously and in real time with no application changes or manual tuning. Our customers include top enterprises such as Citibank, Autodesk, Royal Bank of Canada, members of the Fortune 5, and mid-sized companies.

Pepperdata, Inc.  
530 Lakeside Drive, Suite 170  
Sunnyvale, CA 94085  
USA

 **Request a Demo**  
[pepperdata.com](https://pepperdata.com)

 **Send an Email**  
[eval@pepperdata.com](mailto:eval@pepperdata.com)