



Pepperdata Saves Large Consumer Internet Company 33% for its Data Workloads on Amazon EKS

About The Client:

A consumer Internet brand in the Fortune 500 that provides information and services related to financial transactions.

Challenge:

The company's executive leadership team and big data platform team embarked upon an optimization initiative **to reduce the rising resource cost of its Apache Spark workloads in its massively scaled Amazon EKS environment** that was also running Karpenter and YuniKorn.

Solution:

Pepperdata [Capacity Optimizer's](#) real-time, automated resource optimization immediately reduced workload costs by providing the YuniKorn scheduler with real-time visibility into actual node utilization levels—enabling the scheduler to make more intelligent resource decisions and pack pending pods into nodes with existing capacity.

Pepperdata Capacity Optimizer also enhanced the efficiency of the Karpenter autoscaler by ensuring **new nodes were spun up only when existing nodes were fully utilized**.

Results:

In just days, Capacity Optimizer delivered an initial **33 percent reduction of instance hours (vCPU hours)** and then continued to deliver this cost reduction on an ongoing basis, equivalent to **approximately \$100,000 per month in savings**.

Concern for Growing Data Workload Costs on Amazon EKS

A consumer financial services brand in the Fortune 500 became concerned about ever-increasing costs to run its data workloads as its Amazon EKS environment continued to scale.

In an attempt to control infrastructure costs, the big data platform team embarked upon an executive-led optimization initiative to lower its monthly expenditure.

In addition to its Amazon EKS environment, the company had also pioneered the use of a series of Kubernetes-adjacent technologies, including Karpenter for advanced autoscaling and YuniKorn for advanced scheduling.

Pepperdata Proof of Value Delivered the Required Cost Savings

The company tapped [Pepperdata Capacity Optimizer](#) real-time, automated resource optimization to increase workload utilization levels and reduce costs without requiring any manual tuning or application code changes.

Pepperdata Capacity Optimizer immediately addressed the problem of overprovisioned resources *inside* this customer's Apache Spark applications—a problem that typical optimization steps such as manual application tuning, Spark Dynamic Allocation, Managed Autoscaling, and Instance Rightsizing were not able to fix. Pepperdata demonstrated its benefits through a two-week Proof of Value comprising two segments:

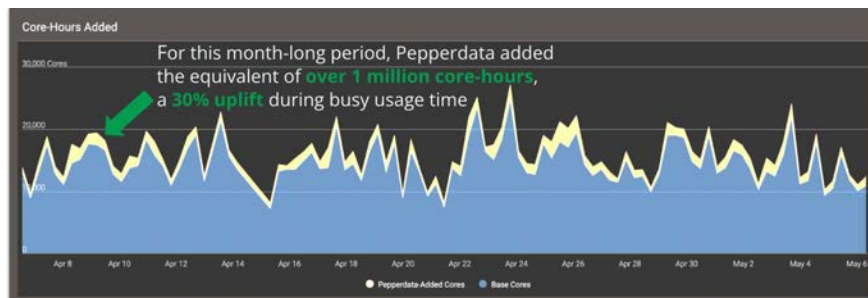
Segment 1: Capacity Optimizer was deployed to the company's development cluster to quickly validate the solution robustness and cost saving results.

Segment 2: Capacity Optimizer was deployed to the company's production cluster to double validate Pepperdata's ability to support the company's scale and robustness.

In just days, Pepperdata Capacity Optimizer increased the customer's resource utilization to help the company **run the same workloads with 33 percent less infrastructure**—delivering a **corresponding decrease in the cost** as measured by reduced vCPU hours.

This example dashboard graph from a month of recent production cluster activity showed that Pepperdata added the equivalent of **over one million core-hours** on top of the base core-hour amount, representing a 30 percent uplift in resource utilization during busy usage time.

Figure 1: Pepperdata Capacity Optimizer added the equivalent of over a million core-hours in a month, representing a 30 percent uplift during busy usage time.



On top of achieving the company executive leadership team's cost optimization goal during the two-week POV, Pepperdata

Capacity Optimizer continues to **deliver 33 percent cost savings results on an ongoing basis**—translating into **approximately \$100,000 per month in reduced costs** for this company.

Pepperdata Capacity Optimizer works automatically in real time without the need for manual tuning, applying recommendations, or changing application code, freeing up dozens of hours each week for the big data platform team to focus on higher-value projects.

How Pepperdata Capacity Optimizer Immediately Reduced Costs by 33 Percent for Apache Spark on Amazon EKS

Without Pepperdata Capacity Optimizer installed, the system scheduler was only able to respond to resource requests provisioned for peak usage levels since it was unaware that actual node resources were underutilized—even when the workload ran well under peak levels.

This gap between resource request levels and actual usage levels led to underutilized nodes that drove up cost. And even with the company adopting Karpenter—which helps rightsize the correct instance types—new nodes continued to be launched despite there being underutilized capacity in existing nodes.

Installed in about an hour, Capacity Optimizer automatically increased utilization of CPU and memory for the company's data workloads by providing the system scheduler with real-time visibility into actual node utilization levels.

Pepperdata Capacity Optimizer helped the company run the same workloads on 33% less infrastructure—translating into approximately \$100K/month in reduced cost.

This visibility enabled more intelligent resource allocation decisions by the scheduler which could then launch more pods on nodes with available capacity.

Pepperdata Capacity Optimizer also enhanced the efficiency of Karpenter by ensuring new nodes were only spun up when existing nodes became fully utilized. The company could now run more applications on less infrastructure while freeing its platform team from manual config tuning.

Pepperdata Contributes to YuniKorn to Drive Further Efficiencies with Optimized Scheduling

For more efficient scheduling in their large-scale Amazon EKS cluster, the company uses YuniKorn. YuniKorn is an open-source scheduler that dynamically adjusts resource allocations based on both developer-configured priorities and workload demands to optimize resource utilization and performance for heterogeneous workloads on large scale, multi-tenant, cloud-native environments.

YuniKorn integrates with Kubernetes to either complement or replace the default Kubernetes scheduler.

Despite its incredible power and flexibility, YuniKorn is still early in its evolution. Pepperdata worked closely with this customer to ensure flawless operation of Capacity Optimizer alongside the YuniKorn scheduler.

As part of this work, the Pepperdata engineering team discovered a stability issue and spent the time to create and contribute an improvement to the YuniKorn codebase, working with the vibrant and active open-source community to improve the overall robustness of YuniKorn.

Above and Beyond: Pepperdata Achieves a Tenfold Footprint Decrease for its Bundled Prometheus

Prometheus is a powerful and versatile yet remarkably simple open-source application for event monitoring and alerting, capable of reliably handling the billions of metrics generated in dynamic cloud environments. Prometheus has become a de facto standard for monitoring Kubernetes environments, including the large-scale environment operated by this customer.

In this customer's environment, Pepperdata made use of Prometheus as a bundled component to provide key optimization metrics to Pepperdata Capacity Optimizer.

Pepperdata developed a specialized configuration for Prometheus **that reduced the memory footprint by nearly tenfold**. Despite this customer's very large-scale environment, Pepperdata Capacity Optimizer with Prometheus as a bundled component exhibited a very small memory footprint, which improved stability and reliability.

Outsized Gains in a Massive, Innovative Kubernetes Environment

Pepperdata's work delivered significant utilization improvements and cost savings for this customer's data workloads through real-time, automated resource optimization in a highly scaled and innovative Amazon EKS environment that was also running Karpenter and YuniKorn.

Pepperdata Capacity Optimizer reduced infrastructure costs by 33 percent, resulting in approximately \$100K/month of savings without any need for manual tuning, applying recommendations, or changing application code.

The big data platform team is now freed to focus on revenue-generating projects while the company can reinvest its savings into continued growth and expansion.



About Pepperdata

Pepperdata resource optimization for data workloads on Kubernetes and Amazon EMR automatically increases utilization levels by up to 80% to reduce overprovisioning waste, delivering an average 30% cost savings—automatically, continuously, and in real time—with no application code changes, recommendations, or manual tuning.

Deployed on over 30,000+ clusters, Pepperdata Capacity Optimizer optimizes Kubernetes resources in some of the largest and most complex environments in the world. Since 2012, Pepperdata has helped companies ranging from startups and mid-sized ISVs to top enterprises such as Citibank, Autodesk, Magnite, Royal Bank of Canada, and members of the Fortune 10 save over \$250 million. For more information, visit www.pepperdata.com.

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



Start a Free PoV

www.pepperdata.com



Send an Email

eval@pepperdata.com

