Pepperdata for Amazon EKS

Reduce your cost of running Spark and Amazon EMR on Amazon EKS—autonomously, continuously and in real time—with zero application code changes.

Benchmarking on Amazon EKS

Pepperdata selected TPC-DS, the decision support benchmark from the Transaction Processing Performance Council, as the standardized workload model for its benchmarking using a GitHub repository that closely adheres to the TPC-DS model. Pepperdata ran this benchmark "out of the box" to create an unofficial audited benchmark as defined by TPC.

Using a 1 TB dataset on 500 nodes running 10 parallel applications (with 275 executors per application) to simulate a multi-tenant environment, Pepperdata found that Capacity Optimizer:

- **Reduced cost:** Reduced instance hours consumed by 41.8 percent
- **Increased throughput:** Reduced total workload run time by 45.5 percent

This benchmark showed the value of Pepperdata for batch workloads running on Amazon EKS at scale. Read more here.

Supported Technologies

- Amazon EMR for EKS
- Apache Spark on EKS

Driving Savings for Global Enterprises

AUTODESK   citi

EXTOLE   IQVIA™

Magnite   RBC

Pepperdata for Amazon EKS

Kubernetes has become the de facto standard for managing applications and services in the cloud. Many organizations choose to run their Kubernetes workloads on Amazon Elastic Kubernetes Service (Amazon EKS) to gain the additional deployment and scaling benefits offered by a fully managed service, including streamlined container management and simplified integration with tooling and plug-ins from the Kubernetes open-source community.

Significant Cost Reductions and Performance Enhancements at Scale

According to a recent survey of enterprise users, Apache Spark is one of the primary workloads deployed on Kubernetes. This same survey identified "significant or unexpected spend" as the top challenge to Kubernetes adoption. Pepperdata Capacity Optimizer, a cost-optimization software solution battle-tested in some of the world's largest and most complex computing environments, delivers automatic cost control for Kubernetes—including Spark on Kubernetes—without the need for manual application tuning. Capacity Optimizer helps improve the performance and cost of Spark workloads running on Amazon EKS by automatically packing additional pending pods onto underutilized nodes, increasing node utilization and reducing the need for additional nodes. According to a recent benchmarking study, Capacity Optimizer reduced the cost of running a standard Spark workload at scale on 500 Amazon EKS nodes by 41.8 percent, as measured by reduced instance hour consumption. Capacity Optimizer also enabled the Spark workload to run 45.5 percent faster, as measured by decreased run time.
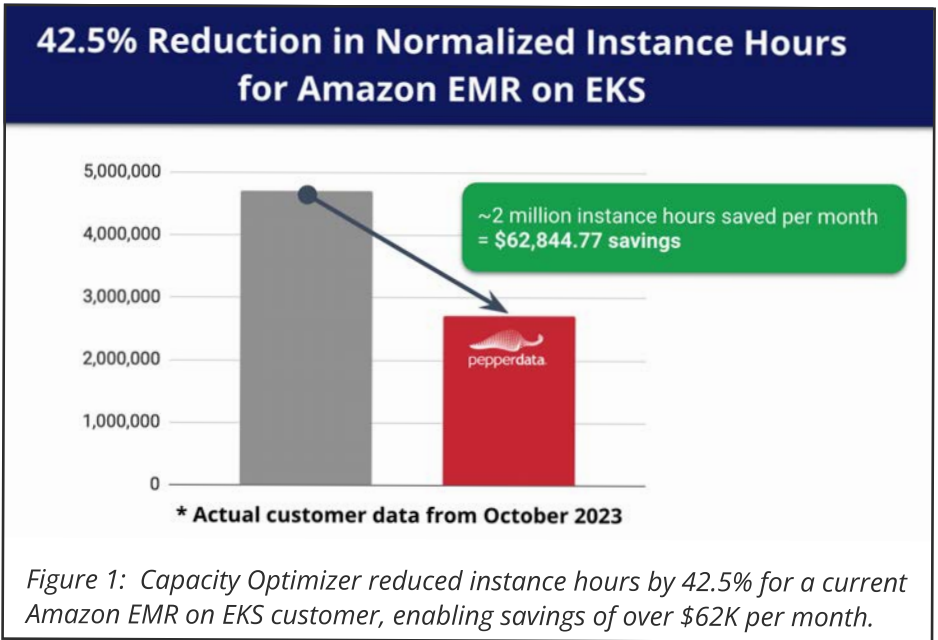
Real-Time Capacity Optimization

Pepperdata Capacity Optimizer offers a patented **Continuous Intelligent Tuning that autonomously reduces costs and optimizes resource utilization in real time** for Spark on Amazon EKS and also for Amazon EMR on EKS, eliminating the hassle of manually configuring pods or responding to ongoing recommendations.

For batch workloads such as Spark, Capacity Optimizer ensures that you pay only for the resources you actually use, rather than paying for higher levels of allocated resources that go unused and wasted. Capacity Optimizer improves node efficiency by enabling the YARN scheduler to launch tasks based on hardware utilization rather than relying on allocations, which by design contain waste in the form of overhead.

Capacity Optimizer further enables the autoscaler to scale only based on actual usage instead of allocations, which ensures that existing nodes in the cluster are fully utilized before the autoscaler adds additional nodes. And it does all of this in real time in response to dynamically changing workloads and datasets. As a result, Capacity Optimizer always maintains nodes in a sweet spot of optimal utilization, which cannot be done manually. In this way, **your applications consume only the resources they actually use**, which translates into significant cost savings.

For Amazon EMR workloads running on Amazon EKS, just as with Spark workloads, Capacity Optimizer delivers additional performance improvements and significant cost reductions. Initial data from a current Pepperdata customer shows a 42.5 percent reduction in instance hours when Pepperdata Capacity Optimizer with Autoscaling Optimization is enabled. Autoscaling Optimization intelligently augments the native autoscaler to ensure all pods are fully utilized before additional pods are launched, eliminating waste and reducing costs.

Figure 1: Capacity Optimizer reduced instance hours by 42.5% for a current Amazon EMR on EKS customer, enabling savings of over $62K per month.

Advanced Visibility and Optimization Flexibility

In addition to providing immediate autonomous improvements in cost and performance, Pepperdata dashboards provide advanced visibility into Spark applications and the clusters in which they are running, including Spark metrics, containers, clusters, pods, nodes, and workflows, in addition to the total realized savings and potential savings available at the cluster level. This advanced visibility translates into greater understanding of your workloads and the ability to adjust optimization levels to your desired price/performance SLAs.

Pepperdata installs in under 30 minutes in most enterprise environments. We guarantee a 100% ROI, with a typical ROI between 100% and 660%.

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, Autodesk, Royal Bank of Canada, and those in the Fortune 5. For more information, visit 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