Pepperdata for Amazon EKS

Reduce the cost of running Spark and Amazon EMR on Amazon EKS by up to 35% with zero code changes and no manual intervention.

Driving Productivity and Business Growth for the World's Largest Brands

citi

Cigna

Expedia

indeed
one search. all jobs.

IQVIA

RBC

securonix

Kubernetes has become the standard for managing modern-day instances and services in the cloud. Pepperdata's 2023 State of Kubernetes Report found that cloud-native Kubernetes deployments are maturing, growing to six to ten clusters per organization.

Kubernetes orchestrates computing, networking, and storage infrastructure and is often relied upon to streamline the migration of on-prem workloads to the cloud. However, despite its power and flexibility, running Kubernetes at scale can pose some challenges—most notably, controlling costs without constant manual tweaking.

Capacity Optimization

Developers typically configure their pods to request more resources than needed. While this practice may be manageable in development environments, it can create needless waste when jobs are moved in production. Pepperdata Capacity Optimizer is a radical new way to reduce cost and resources for big data workloads on EKS. By packing additional pending pods onto underutilized nodes, Capacity Optimizer increases the node utilization and reduces the need for additional nodes, which translates directly to reduced costs. By working autonomously and continuously in the background, Capacity Optimizer uses machine learning to eliminate the hassle of manual tuning, freeing your developers to focus on production and innovation.

Optimal Autoscaling

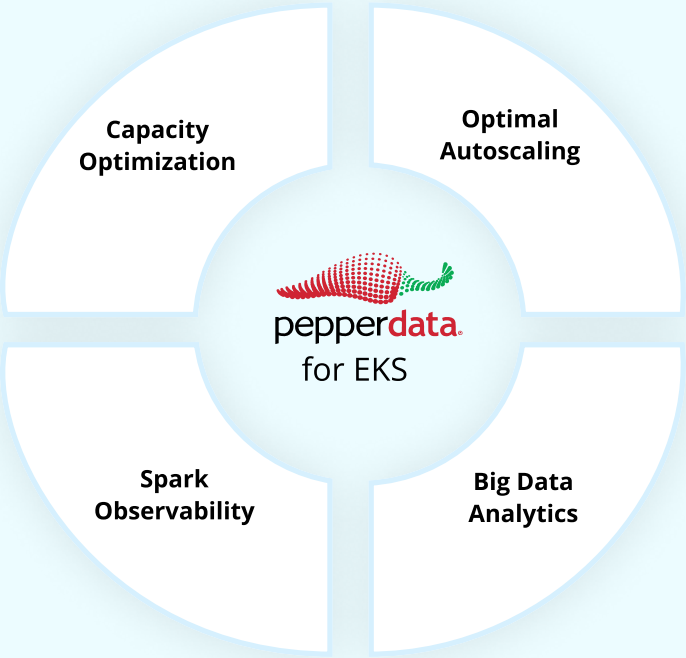
While EKS autoscaling provides the elasticity required for big data workloads, that flexibility can also lead to increased costs. Pepperdata autonomously remediates any inefficiencies in your Amazon EKS clusters by reclaiming resources and scaling based on actual and not requested resources. By packing more pods onto existing nodes, Pepperdata automatically reduces the number of nodes needed, allowing you to achieve up to 35% cost savings over the default EKS autoscaler configuration with no code changes.

pepperdata

Optimize Cloud Costs and Operate Kubernetes Autonomously

Kubernetes management doesn't need to be manual. Pepperdata's Capacity Optimizer autonomously optimizes your cluster resources, recapturing wasted capacity so you can run more applications, and get the most out of your infrastructure investment.

Pepperdata helps you improve price/performance, increase throughput, meet SLAs, and more. Pepperdata does all this while significantly recovering resource waste and reducing overprovisioning.



Spark Observability

Pepperdata provides comprehensive visibility into the health and performance of Spark applications on Kubernetes in real time. Managers, developers, and operations teams can monitor cluster resource usage and optimize the performance of their clusters and applications under a single pane of glass. Your developers can receive prescriptive recommendations to troubleshoot Spark applications that need attention. They can also automatically identify bottlenecks and receive alerts on duration, failure conditions, and resource usage. In addition, Pepperdata's chargeback or showback capabilities allow you to track, measure, attribute, and charge infrastructure usage costs.

For Spark on Kubernetes, Pepperdata provides:

- Autonomous optimization of resources and workloads
- Application and infrastructure observability
- A self-service dashboard for granular visibility into the Spark job details as well as performance tuning recommendations
- Detailed usage attribution for IT chargeback

Benchmarking EKS with the TPC-DS Query Set

TPC-DS is the Decision Support framework from the Transaction Processing Performance Council. TPC-DS is a sophisticated, industry-standard big data analytics benchmark. (Pepperdata's work is not an official audited benchmark as defined by TPC.)

Using a 1 TB dataset on 30 nodes with 225 executors and 103 jobs, Pepperdata achieved the following results for Capacity Optimizer on EKS:

- **Improved Performance:** Decreased query duration by 30%
- **Improved Utilization:** Decreased idle CPU by 13%
- **Increased Throughput:** Increased workload capacity by 35%
- **Increased Available Resource Utilization:** Decreased available memory by 27%

Supported Technologies

- **Amazon EKS**
- **Amazon EMR for EKS**
- **Apache Spark**

Sign up for the free self interactive demo to see how automatic optimization and full-stack observability can improve big data application performance across your entire big data stack.

Request Free Demo

www.pepperdata.com/trial | sales@pepperdata.com