



Pepperdata for Amazon EMR

Reduce your Amazon EMR costs by up to 50% with zero code changes

Enterprises grappling with controlling cloud costs in these uncertain economic times rely on Pepperdata to inform and optimize their cloud spend.

Pepperdata for Amazon EMR provides automated tuning, full-stack observability, and real-time insights across all of your Spark workloads and EMR instances all in one place. Autonomously optimize your Spark jobs and reduce your Amazon EMR costs by up to 50% with no code changes.

Optimal Capacity Utilization

Reclaimed allocated but unused resources

Optimal Autoscaling

Utilizes instances fully before adding more

Optimal Instance Selection

Choose optimal instances for your workloads

Application-Aware Observability

Troubleshoot, alert, and charge back at a granular level, under a single pane of glass

“Pepperdata provides great value insight and cost savings to the point where it pays for itself within months or sooner.”

—Director, Big Data | Health, Wellness and Fitness Company

Optimal Capacity Utilization

Pepperdata Capacity Optimizer is a radical new way to reduce cost and resources in the cloud. 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. Capacity Optimizer ensures that all nodes are fully utilized before additional nodes are created and directs the EMR autoscaler to downscale as instance utilization decreases. The net effect is that scaling is autonomously optimized and waste is eliminated.

“Pepperdata allowed us to significantly increase capacity for our Amazon EMR workloads and reduce our costs by over 50%. We can focus on our business, while they optimize for costs and performance.”

—Chief Data Architect, Data Platforms and Insights | Global Design Software Company

Optimal Autoscaling

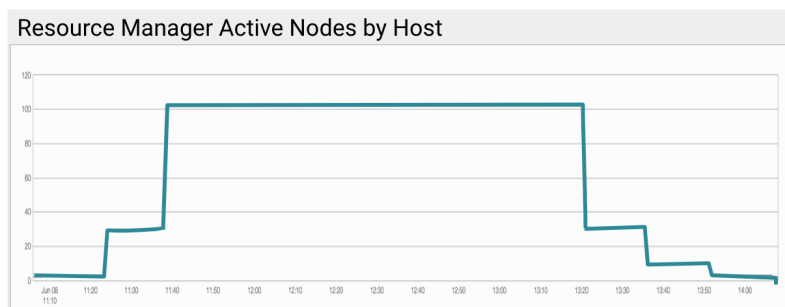
While Amazon EMR autoscaling provides the elasticity required for big data workloads, that flexibility can also lead to increased costs. Pepperdata autonomously remediates any inefficiencies in your Amazon EMR clusters by reclaiming resources and scaling based on actual and not requested resource utilization. Pepperdata also supports your cluster in automatically downsizing the number of autoscaling instances in a time-sensitive and robust manner. By automatically reducing the number of autoscaling instances, Pepperdata allows you to achieve up to 35% cost savings over the default EMR autoscaler configuration with no code changes.

Optimal Instance Selection

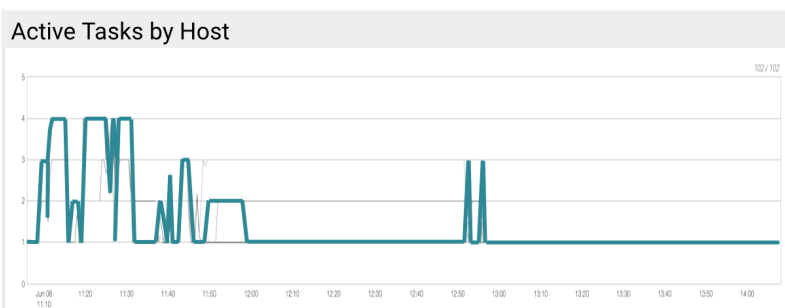
Overprovisioning is all too common in the cloud. Nearly one-third of businesses exceed their cloud budgets by as much as 40%, in part due to poorly sized instances. Pepperdata remediates overprovisioning by analyzing the historical utilization of the CPU and memory of your workloads, applying a machine learning algorithm that sorts through the hundreds of EC2 instance types, and identifying the instance types that are the best fit for your workload's unique characteristics. Instance rightsizing with Pepperdata allows you to achieve up to a 40% cost reduction while preserving the same level of performance.

Pepperdata enhances cloud-native autoscalers

Cloud-native autoscalers tend to operate in large increments:



Workloads are dynamic and require much smaller increments:



Application-Aware Observability

Pepperdata provides full-stack observability of your infrastructure and resources at a granular level, enabling you to quickly diagnose performance issues anywhere in your environment and make resource decisions based on business priorities. Pepperdata makes it easy to observe and manage mixed workloads from Spark, MapReduce, Hive, and Presto/Trino under a single pane of glass. Your developers can receive prescriptive recommendations to troubleshoot applications that need attention. They can also automatically identify bottlenecks and alert on duration, failure conditions, and resource usage. In addition, Pepperdata's chargeback and showback capabilities allow you to track, measure, attribute, and charge infrastructure usage costs.

"It's hard to understand how we were getting by without Pepperdata."

—Chief Data Architect | DPI

Ready to get started with a 15-day, no-risk trial of [Pepperdata](#) to reduce inefficiencies in your Amazon EMR clusters?

Visit pepperdata.com/big-data or reach out to us at sales@pepperdata.com. Pepperdata installs in most clusters in about 30 minutes, providing you a rapid path to cost savings.



Pepperdata and the Pepperdata logo are trademarks or registered trademarks of Pepperdata Inc. All other trademarks are the property of their respective owners. Pepperdata reserves the right to change this document without notice. To ensure you have the latest version of this document, visit www.pepperdata.com.

About Pepperdata

Pepperdata products help customers transform the performance of their big data cloud and Kubernetes workloads. Unlike solutions from infrastructure monitoring and APM vendors that provide only summary dashboards, **Pepperdata** automatically scales system resources while providing a detailed and correlated understanding of each application using hundreds of real-time application and infrastructure metrics. This helps IT maintain business continuity, ensuring that applications and workloads meet SLAs, and track resource spend for clear accountability. Companies like Expedia and Royal Bank of Canada depend on Pepperdata to deliver big data success. For more information, visit www.pepperdata.com.