# Pepperdata for Amazon EMR

## Reduce your Amazon EMR costs by up to 50% with zero application 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 Amazon EMR instances all in one place. Autonomously optimize your Spark jobs and reduce your Amazon EMR costs by up to 50% with no application changes.

### Optimal Capacity Utilization

Reclaim allocated but unused resources

### Augmented Autoscaling

Utilize instances fully before adding more

### Application-Aware Observability

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

> **"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."**
>
> —*Mark Kidwell, Chief Data Architect*
>   *Autodesk, Inc.*

## Optimal Capacity Utilization

Pepperdata Capacity Optimizer is a proven solution to reduce cost and optimize resources in the cloud while meeting performance SLAs. 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 Amazon EMR autoscaler to downscale as instance utilization decreases. The net effect is that scaling is autonomously optimized and waste is eliminated.

> **"Pepperdata provides great value, insight, and cost savings to the point where it pays for itself within months or sooner."**
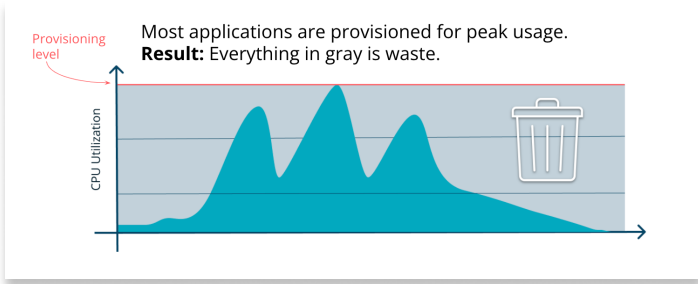>
> —*Director, Big Data*
>   *Health, Wellness and Fitness Company*

## Augmented Autoscaling

Amazon EMR autoscaling provides the elasticity required for big data workloads. Pepperdata works on top of Amazon EMR to turbocharge resources in your Amazon EMR clusters through responsive scaling based on real-time, second-by-second resource utilization. Pepperdata supports your cluster by automatically downsizing the number of instances in a time-sensitive and robust manner. By automatically responding to your cluster changes in real time, Pepperdata allows you to optimize costs related to your Amazon EMR environment up to 50% with no application changes.

## Pepperdata Responds in Real Time

Most organizations provision applications for peak usage that remains in place even when applications scale down.



Provisioning level

CPU Utilization

Most applications are provisioned for peak usage.
**Result:** Everything in gray is waste.

Pepperdata understands that workloads are dynamic; they operate at peak and also scale down below peak levels for significant amounts of time. Pepperdata Capacity Optimizer responds to fluctuating workloads in real time to eliminate the waste of overprovisioning, reduce costs, and optimize resource utilization.

## 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, and Hive 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 chargeback and showback capabilities allow you to track, measure, attribute, and charge infrastructure usage costs with ease.
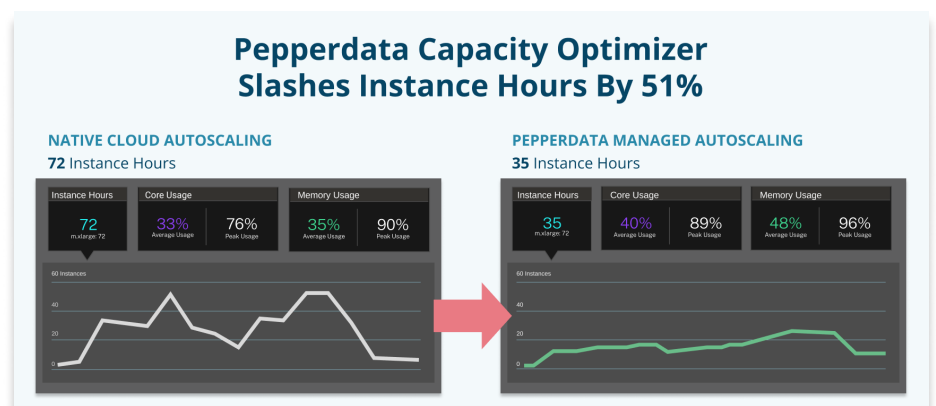
## Your Easy Button to Cost Savings

As a result of its ability to identify where more work can be done, and to find and add tasks to nodes with available resources—continuously and in real-time— Pepperdata Capacity Optimizer can reduce your instance hours by over 50% with no manual tuning and no application changes.



**Pepperdata Capacity Optimizer
Slashes Instance Hours By 51%**

NATIVE CLOUD AUTOSCALING
**72** Instance Hours

| Instance Hours | Core Usage | | Memory Usage | |
| 72 m.xlarge: 72 | 33% Average Usage | 76% Peak Usage | 35% Average Usage | 90% Peak Usage |

PEPPERDATA MANAGED AUTOSCALING
**35** Instance Hours

| Instance Hours | Core Usage | | Memory Usage | |
| 35 m.xlarge: 72 | 40% Average Usage | 89% Peak Usage | 48% Average Usage | 96% Peak Usage |

The example dashboard above is taken from the cloud cluster where Pepperdata was able to reduce instance hours from 72 per week to only 35 instance hours for the same duration.