

Pepperdata Boosts 3D Design Software Enterprise's Amazon EMR Performance, Cuts Costs

One of our biggest clients, a 3D design software company, now benefits from more efficient workloads, reduced operational costs, and faster troubleshooting to nip performance issues in the bud. Discover how Pepperdata's Capacity Optimizer and Spotlight solutions help them do just that.

About The Client

The client is a global leader in design and manufacturing software with primary markets in engineering, architecture, construction, manufacturing, media, and entertainment industries.

The company constantly analyzes large volumes of data using Apache Spark on Amazon EMR. This allows them to derive critical insights, improve existing products, and create new solutions while ensuring that products and services perform well to meet critical SLAs.

"Pepperdata allowed us to significantly increase capacity for our Amazon EMR workloads and reduce our EC2 costs by over 50%. We can focus on our business, while they optimize for costs and performance."

—Chief Data Architect, Data Platforms and Insights

The Situation: Apache Spark on Amazon EMR Tuning Complications

The client used Apache Spark on Amazon EMR to process and analyze large sets of big data and turn them into insights. While this approach proved effective, performance became a significant issue when Spark was left unoptimized.

The company's Chief Data Architect added:

"Spark is notoriously hard to tune correctly. People don't have time to go into every job. As a result, our entire platform just wasn't as efficient as it could have been."

Challenge:

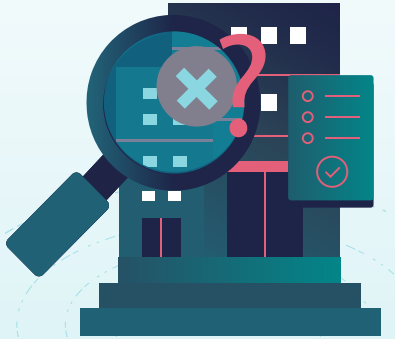
The software company found that scaling Amazon EMR resources to handle workloads resulted in runaway costs. Its goal was to reduce costs by 50% by increasing capacity and rightsizing compute for the company's Apache Spark on Amazon EMR applications.

Solution:

The software company used Pepperdata's Spotlight and Capacity Optimizer solutions for increased visibility and autonomous optimization.

Results:

With Pepperdata, our client significantly increased its capacity for Amazon EMR workloads, optimized processes for better business results, and successfully reduced its Amazon EC2 costs by over 50%.



The software development company used Amazon EMR autoscaling to dynamically scale Spark for better performance and lower costs. While autoscaling improved its efficiency, the benefits needed to be more significant to help it achieve performance goals or reduce cost. The company's data team tried to increase capacity by adjusting maximum instance size and autoscaling policies to ensure smooth performance, but costs continued to add up.

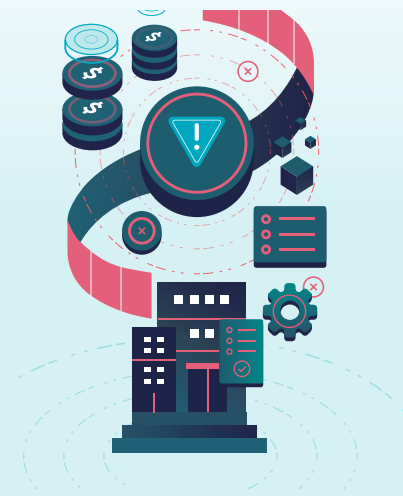
As their workloads grew, so did the team's Spark issues. The increased compute consumption was quickly eating through its budget. Each Amazon EMR cluster was consuming two or three times the planned capacity. In 2020, the company saw its data processing needs increase 10x over the previous year. If this trend of doubling capacity and overprovisioning resources continued, our client would have been overwhelmed by runaway costs, low latencies, and increased downtime.

Resolution: Visibility and Autonomous Optimization in One Package

Turning to Pepperdata, the 3D software company found a comprehensive solution that gave it visibility into its Spark applications, significantly improved Spark performance, and effectively reduced its compute consumption. Our client set a goal to reduce costs on Amazon EMR by 50%. With Pepperdata, they were well on their way to reaching that goal.

As the enterprise's Chief Data Architect put it:

"We didn't have an automated way to identify potential problems or make our systems more efficient. We needed observability and insights."



[Pepperdata Capacity Optimizer](#) solved the problem of inefficient cloud autoscalers by enabling the scheduler or cluster manager to schedule workloads based on actual resource utilization instead of resource allocation—cutting the organization's EC2 instance cost by 50%.

Pepperdata Capacity Optimizer not only maximizes the utilization of each of the existing instances: It manages the autoscaling behavior of the cloud platforms and ensures that the new instances are added only when existing instances are fully utilized in an autoscaling environment.

Capacity Optimizer allowed the company to implement a more sophisticated approach to Spark performance tuning. The solution automatically optimizes the resources in its clusters and recaptures compute waste, resulting in 15% reduction of instance hours. With more resources available, our client can run more applications without adding additional hardware and personnel to tune them.

[Pepperdata Application Spotlight](#) helps the company developers determine the amount of resources needed for specific applications and workloads, ensuring adequate resources are available to avoid performance issues and lags. It provides the client with tuning recommendations that they can quickly apply to ensure all applications are running smoothly and efficiently—while eliminating applications and processes deemed wasteful by the system.

[Pepperdata Platform Spotlight](#) gives the company data team a complete picture of their big data architecture and its processes, which is instrumental to all their Spark performance tuning efforts. With everything visible and accessible in one place, their big data team can spot issues and determine if the issues are application-based or a symptom of cluster performance.

With a hand in a multitude of industries—engineering, architecture, construction, manufacturing, media, and entertainment—the client needs its big data workloads to capture as many critical insights as possible in order to consistently provide new products and solutions for its customers.

After implementing Pepperdata, our client significantly increased capacity for Amazon EMR workloads, optimized processes for better business results, and reduced Amazon EC2 costs by over 50%. Now, their big data team can focus entirely on business priorities and innovation rather than cost control and manual infrastructure tuning.



[Try a Demo](#)

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.

©2022 Pepperdata Inc. All rights reserved. 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.



3945 Freedom Circle,
Suite 920 Santa Clara, CA
95054



Start Free Trial
www.pepperdata.com/trial



Send an Email
eval@pepperdata.com