Autodesk Reduces Amazon EMR Costs by 50% and Boosts Performance with Pepperdata

AUTODESK

About The Client

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

Challenge:

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

Solution:

Autodesk used Pepperdata Capacity Optimizer for autonomous and continuous cloud cost optimization in real time, and granular visibility.

Results:

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

The Situation: Apache Spark on Amazon EMR Tuning Complications

Autodesk 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 for the company became a significant issue when Spark was left unoptimized.

> *"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."*
>
> **—Mark Kidwell, Chief Data Architect, Autodesk Data Platforms and Services**

Amazon EMR autoscaling was able to dynamically scale Spark for better performance and lower costs.

While autoscaling improved Autodesk's efficiency, even higher rates of utilization were needed to help reduce costs. The company's data team tried to increase capacity by adjusting maximum instance size and autoscaling policies to ensure smooth performance, but manually tuning was difficult and costs continued to add up.

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

## Resolution: Autonomous Optimization and Observability in One Package

Turning to Pepperdata, Autodesk found a comprehensive solution that autonomously reduced its compute consumption, maximized resource utilization of its applications, and provided visibility into its Spark applications—all continuously and in real time. The company set a goal to reduce costs on Amazon EMR by 50 percent. With Pepperdata's cost optimization capabilities, they were well on their way to reaching that goal.

> *"We didn't have an automated way to identify potential problems or make our systems more efficient. We needed observability and insights."*
>
> **—Mark Kidwell, Chief Data Architect, Autodesk Data Platforms and Solutions**

The problem of skyrocketing cloud costs was solved by Pepperdata Capacity Optimizer enabling the scheduler or cluster manager to schedule workloads based on actual resource utilization instead of resource allocation—cutting the organization's Amazon EC2 instance cost by 50 percent.

Pepperdata Capacity Optimizer not only maximized the utilization of each of the existing instances: it enhanced the autoscaling behavior of the cloud platforms and ensured that the new instances were added only when existing instances were fully utilized.
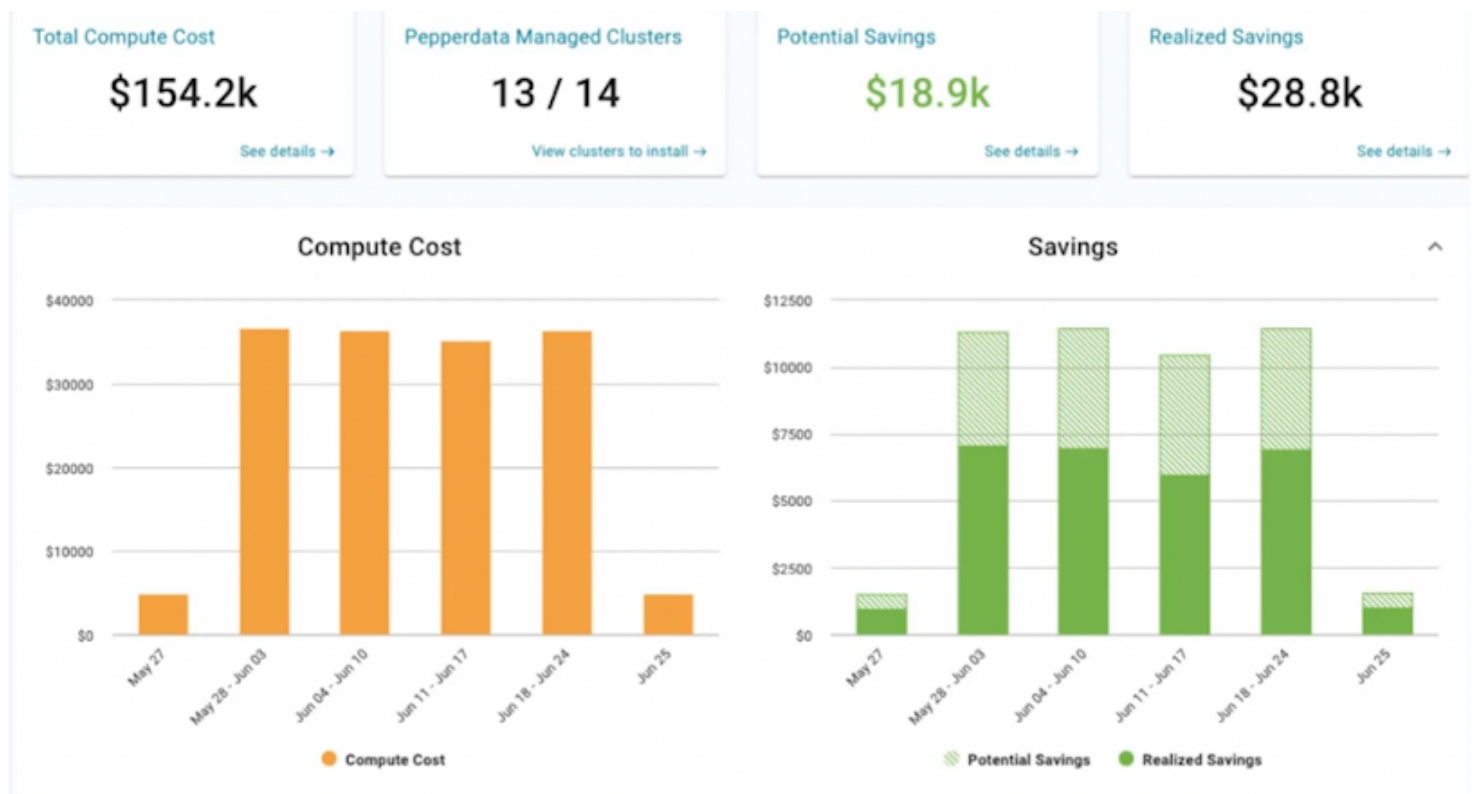
Capacity Optimizer also allowed Autodesk to implement a more sophisticated approach to Spark application tuning. The solution automatically optimized the resources in its clusters and recaptured compute waste, resulting in a 15 percent reduction of instance hours. With more resources available, the company could run more applications without adding additional hardware and personnel to tune them.

Pepperdata Capacity Optimizer further helped company developers determine the amount of resources needed for specific applications and workloads, ensuring that adequate resources were available to avoid performance issues and lags. It provided Autodesk with thousands of application-level metrics in one aggregated view —while eliminating manual application tuning with no code changes required.

Capacity Optimizer gave the Autodesk data team a complete picture of their big data architecture and its processes, which were instrumental to all their Spark application tuning efforts. With everything visible and accessible in one place, their big data team could spot issues and determine if the issues were application-based or a symptom of cluster performance.

After implementing Pepperdata, the client significantly increased capacity for its Amazon EMR workloads, optimized processes for better business results, and reduced Amazon EC2 costs by over 50 percent. With an automated solution to optimize their workloads, their big data team became free to focus entirely on business priorities and innovation rather than cost control and manual infrastructure tuning.

Autoscaling Optimizer Cuts Instances Hours By 51%

BEFORE = Native Cloud Autoscaling
72 Instances Hours

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

AFTER = 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

"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."

—Mark Kidwell, Chief Data Architect, Autodesk Data Platforms and Services

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, T-Mobile, Autodesk, Securonix, Royal Bank of Canada, and those in the Fortune 5. For more information, visit pepperdata.com.

pepperdata.
530 Lakeside Drive
Suite 170
Sunnyvale, CA 94085

Start a free PoV
www.pepperdata.com/
free-cost-optimization-
proof-of-value

Send an Email
eval@pepperdata.
com