

Deep Insight into the Performance and Cost of Your Large-Scale Data Analytics Clusters and Applications

Gain full-stack observability alongside Pepperdata's industry-leading Real-Time Cost Optimization

Real-Time Cost Optimization: Controlling Large-Scale Data Analytics Expenses

With cost and resource discipline becoming an increasing business imperative—and even a competitive differentiator—leading companies are turning to Pepperdata to maximize their large-scale data analytics investments and rein in expenses.

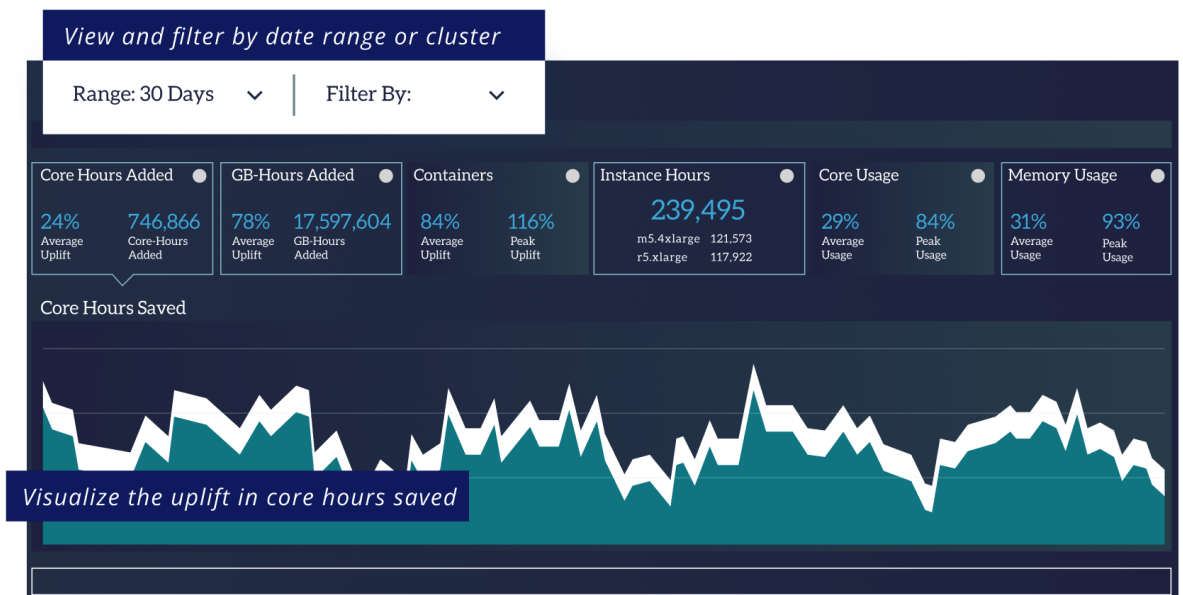
For clients ranging from global giants to tech startups, Pepperdata Capacity Optimizer's ability to continuously and autonomously reduce instance hours and costs and improve utilization in **real time** through its **Continuous Intelligent Tuning** has saved millions of dollars. By informing the scheduler of exact resource usage and ensuring that every instance is fully utilized before a new instance is added, Pepperdata delivers **30 to 47 percent immediate cost savings** with **no manual application tuning** and offers a guaranteed ROI between 100 and 662 percent.

Optimization with Enhanced Observability for Apache Spark

Alongside our Real-Time Cost Optimization, Pepperdata also provides the option for **full-stack observability** and **real-time insights** across all of your Apache Spark workloads on Amazon EMR and Amazon EKS.

For customers interested in additional visibility into their Apache Spark cluster and application activity, Pepperdata's observability dashboard offers deep insights not readily available through general-purpose monitoring or out-of-the-box performance management tools. These observability features are available to customers of Capacity Optimizer at no extra cost. The combination of Pepperdata Real-Time Cost Optimization plus additional observability empowers you to both optimize and understand the performance and cost of your large-scale data analytics clusters and applications.

Figure 1: Pepperdata quantifies the dollar cost savings that would otherwise go to waste without Pepperdata Capacity Optimizer installed. For example, this cloud-based customer achieved a 30% uplift in their core and memory usage, translating into cluster cost savings of almost \$50,000.



Apache Spark Observability: Gain Control of Your Large-Scale Data Environment

Apache Spark observability is an integral aspect of managing your large-scale data analytics environment. Unlike monitoring, which only tells you *when* something went wrong, observability tells you *why*. It is a term from control theory that helps determine how well a system is functioning. You can't control the output of a system if you can't observe what is happening inside. **More observable details mean greater control.**

As compared to general-purpose monitoring or out-of-the-box performance management tools, Pepperdata provides everything you need for holistic, enterprise-wide platform observability—and therefore control—at both the cluster and the application level.

Pepperdata works by instrumenting every node of your cloud cluster with a micro-footprint PepAgent. Developed and hardened over the last decade in some of the world's most complex compute environments, PepAgent continuously collects and correlates hundreds of real-time operational metrics, including host-level CPU, RAM, disk I/O, and network metrics, as well as job, task, queue, workflow, and user info. PepAgent is the secret sauce that powers the Pepperdata dashboard. From the single pane of glass of the Pepperdata dashboard, you can quickly diagnose, troubleshoot, and resolve both cluster-wide and low-level application issues.

Resolve System-Wide Issues with High-Level Cluster Observability

Pepperdata's observability at the cluster level helps you ensure that your infrastructure is optimally sized for your unique workloads. Observability at the cluster level also enables you to understand your cluster environment as well as rapidly diagnose, troubleshoot, and resolve system-wide issues by providing answers to such questions as:

- What are all the Spark applications running on my cluster at any given time?
- Which applications are running on Kubernetes?
- What resources are my applications consuming?
- Is my cluster under- or overprovisioned?
- When and where are bottlenecks occurring in my cluster?



Figure 2: Pepperdata's dashboard delivers easy-to-digest high-level stats about overall cluster performance and resources, in terms of CPU, memory, and node health.

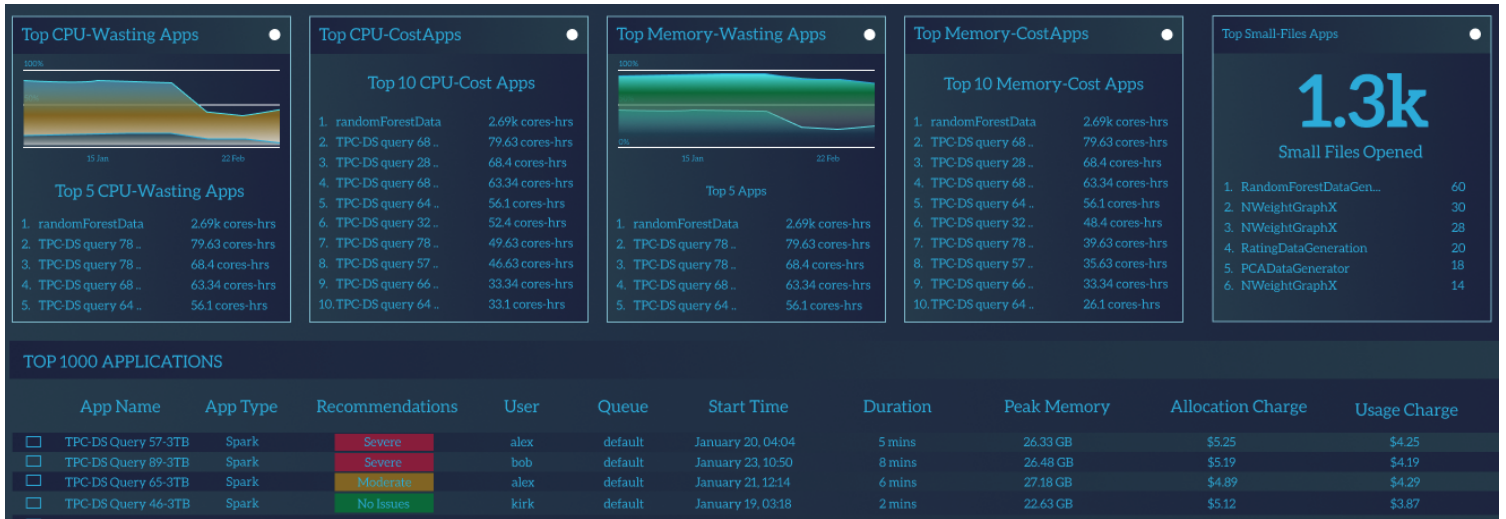


Figure 3: Pepperdata's dashboard provides a summary of all the applications that are currently running or have run in the past—even on now-deleted ephemeral clusters. Simple rollups highlight the top apps that may need attention in terms of excessive waste and cost. From here, it's easy to drill down into any application to understand its unique performance characteristics and its impact on the cluster.

Gain Deeper Insights with Granular Application-Level Observability

Pepperdata's observability and recommendations at the application level provide your developers the deep insight they need into the performance of tens of thousands (or even more!) Spark applications running concurrently. Application-level observability both improves application efficiency and speeds issues to resolution by delivering answers to such questions as:

- What are all the applications that ran in my environment in the last month (including those that ran on now-deleted ephemeral clusters)?
- What is the average/max/min CPU and memory utilization for each of those applications?
- Why is Application X consuming all the available memory and CPU?
- For an application that ran multiple times, which runs were fastest/slowest and why?

Key Takeaways

Pepperdata's industry-leading Real-Time Cost Optimization via Capacity Optimizer, combined with its observability dashboard, empower you to:

- Automatically decrease instance hours
- Continuously reduce your Amazon EMR and Amazon EKS costs in real time
- Safely and securely optimize your Apache Spark and other large-scale data analytics applications
- Quickly grasp the status of your large-scale data analytics environment, both at the cluster level and the granular application level
- Gain real-time insight to quickly diagnose, troubleshoot, and resolve both cluster-wide issues and low-level application failures

It's easy to get started with Pepperdata Capacity Optimizer and our observability dashboard via a [free Savings Demo](#). Pepperdata typically installs in most enterprise clusters in under 60 minutes and starts delivering efficiencies and cost savings as soon as it starts running. **Pepperdata also guarantees a minimum of 100 percent ROI, with a typical ROI between 100 and 660 percent.**

We also offer [free Proofs of Value](#), which allow organizations to explore the full value of Pepperdata in their environment. Leading companies such as Citibank, Autodesk, Royal Bank of Canada, IQVIA, and those in the Fortune 5 depend on Pepperdata to optimize and visualize the performance of their large-scale data analytics investments. To learn more or get started, visit pepperdata.com or contact us at sales@pepperdata.com.