pepperdata

CASE STUDY

Zeotap Increases Big Data Application Performance 23.78% with Pepperdata

With Pepperdata, Zeotap achieved and maintained optimal usage for all Apache Spark jobs on their platform. The software enabled them to tune big data applications, gain powerful insights, and implement data-driven recommendations, resulting in a 23.78% increase in tasks performed and over $31,000 savings in just over three months.

"We use Google Dataproc clusters for our data engineering pipelines. Pepperdata lets us track critical metrics for those pipelines." —Aditya "Ady" Chandra, *Director of IT*

Zeotap supplies its clients with a powerful Customer Intelligence Platform, allowing companies to leverage first-party data so they can understand their customers, anticipate consumer behaviors, attract new customers, and grow their customer base. As they gained more clients, Zeotap needed a better way to gain visibility and build visualizations to ensure optimal usage of all Spark jobs. Linking data across multiple clusters and workflows was among its pressing big data problems. Pepperdata delivered a comprehensive big data application performance suite tailored to meet Zeotap's unique requirements, from monitoring and tracking Spark jobs to tuning applications for continuous optimal performance. With Pepperdata, Zeotap has achieved superior visibility and performance, especially as their Spark workloads increase. Powerful metrics enable them to monitor jobs and optimize compute utilization, effectively driving down costs while fulfilling the SLAs promised to their clients.

Superior Visibility and Data Visualization = Better Big Data Application Performance and Massive Savings

Zeotap is one of the market leaders in the customer intelligence and analytics industry. Zeotap provides companies with powerful customer intelligence solutions, enabling them to gain a better and more comprehensive understanding of their customers through data analytics and modeling, data unification, identity resolution, and more. With Zeotap, businesses can build on a nucleus of first-party data. This allows them to predict

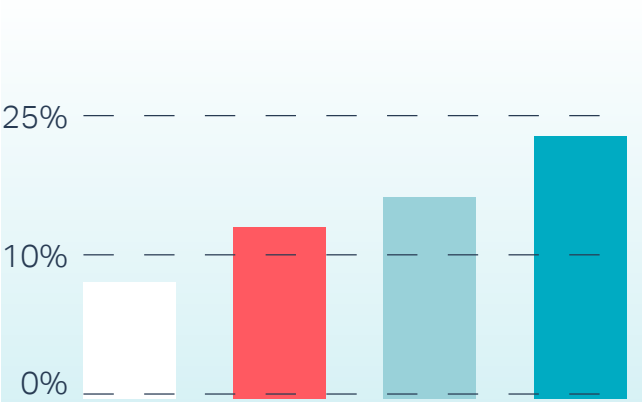
consumer behaviors, create and deliver more meaningful experiences, attract new customers, and grow their consumer base.

Pepperdata provides Zeotap with a robust and powerful big data application optimization solution. The solution gives Zeotap the visibility and observability they need into their applications and workloads. Through the visibility Pepperdata provides into Zeotap's big data infrastructure performance, Zeotap further improved the performance of multiple jobs—including scheduled and ad-hoc jobs.

Pepperdata Capacity Optimizer makes it possible for Zeotap to optimize their resources and run more tasks. This resulted in a cost savings of $31,414.76 in just over three months. With Pepperdata, Zeotap implemented powerful, data-based optimization configurations to tune applications and minimize costly jobs.

Working with some of the world's largest brands—including Red Bull, Mercedes Benz, Pepe Jeans, and FMCG—Zeotap's fully optimized big data applications and Apache Spark environments not only benefited the company but also impacted their clients' outcomes.

By optimizing their resources, Zeotap is able to run 23.7% more tasks. This resulted in a cost savings of over $31,000 in just over three months.



The Situation

Zeotap leverages Apache Spark to process incoming data from their partners. Spark also helps Zeotap deliver analytics for their clients, testing sets of ID for matches and exporting based on linkage. Spark gives Zeotap a fast and powerful engine to process large amounts of data. However, Spark creates a lot of big data challenges when running unoptimized. Spark queries take longer to process, and jobs consume more resources than needed. This results in poor performance and overspending.

Overspending, due to unresolved big data challenges, is a real concern. In our recent **big data survey,** we found that the majority of businesses that migrated to the cloud and embraced new digital technologies are concerned with the "optimization of their cloud resources." 64% of respondents are worried about "cost management and containment," and one in 12 enterprises are spending 40% more than their initial big data and cloud allocation.

Aside from overspending, visibility is another big data problem experienced by enterprises. Zeotap was no exception. Prior to using Pepperdata, Zeotap used SparkLens to gain some degree of visibility and build visualizations. However, Zeotap struggled to link data across multiple clusters and workflows as well as monitor multiple jobs that go through multiple Google Dataproc clusters.

Using Google Dataproc clusters to manage data engineering pipelines, Zeotap wanted better visibility into these applications to ensure optimal resource usage for all Spark jobs on the platform. The company also wanted effortless tracking across multiple workflows, Oozie and Airflow included.

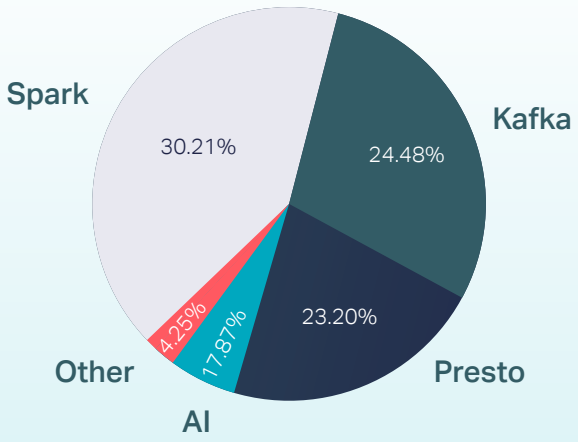
Lack of Visibility Impeded Spark Performance

The limited visibility into their big data applications and workloads, coupled with Zeotap's rapidly growing customer base, meant that more workloads were entered into Spark without an effective means to track, manage, and optimize them.

To ensure their Spark applications perform optimally, developers need to tune the applications constantly. Spark applications consist of many different types of stages. One particular configuration that's optimal for one stage might be inappropriate for another stage.

One in every three companies spend 20% to 40% more than their initial big data/cloud budget. The complexity of Spark is a big reason for this.

Without the right level of visibility, it's extremely difficult to tune Spark jobs and achieve optimal application performance without a significant rise in cost. Failure to optimize Spark applications and big data infrastructure could easily lead to over-spending.



Pepperdata: Optimizing Spark Apps and More

Zeotap turned to Pepperdata to help them with their visibility requirements. They needed to closely monitor their Spark applications and ensure their workloads are running optimally.

Pepperdata Platform Spotlight helped Zeotap correlate applications and workflows. This allowed them to truly measure the cost and efficiency of each application and workflow.

Pepperdata Application Spotlight provided Zeotap with real-time configurations for big data applications based on performance data and other KPIs.

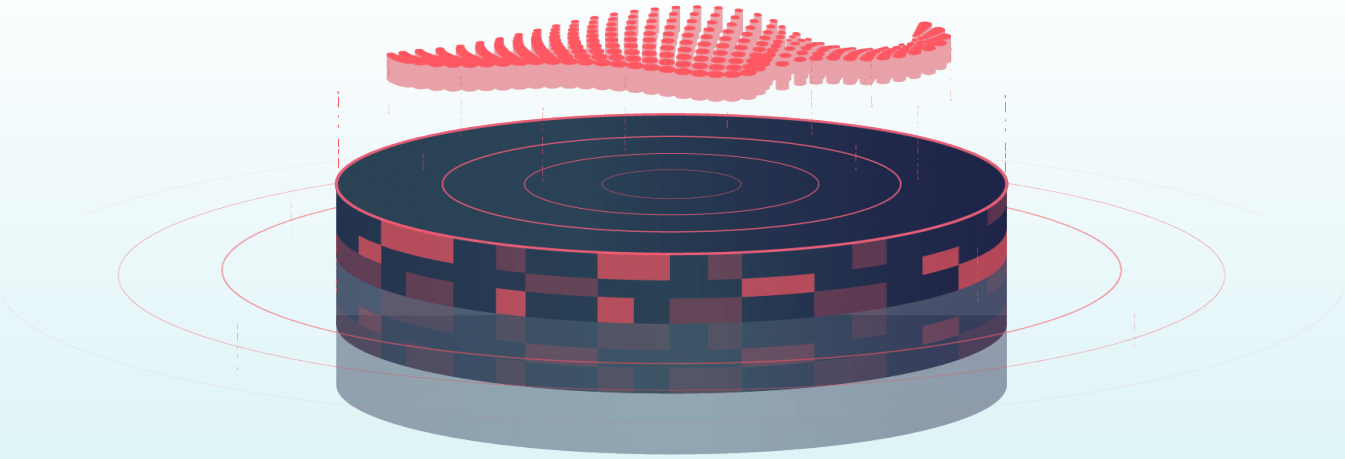
Capacity Optimizer ensured that Zeotap's compute resources and capacity were adequate and optimized, guaranteeing SLA-level performance whenever usage increases. Pepperdata autoscaling significantly reduced resource wastage, contributing to massive resource recovery, and ultimately, a substantial reduction in costs.

The Pepperdata Impact

Shortly after deploying the Pepperdata solution, Zeotap experienced immediate positive results. The costs for all of their Spark engineering efforts decreased and became more manageable, even as Spark workloads increased.

In the future, Zeotap is looking to fully automate tuning processes based on Pepperata recommendations and automatically integrate them into new big data applications.

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.



About Pepperdata

Pepperdata provides analytics performance management that guarantees SLAs and reliability for your infrastructure and apps.

pepperdata

3945 Freedom Circle, Suite 920
Santa Clara, CA 95054

Start a Free Trial
www.pepperdata.com/trial

Send an Email
eval@pepperdata.com