pepperdata

Pepperdata Capacity Optimizer:

Real-Time, Automated Resource Optimization at Scale

More Pods Per Node for Kubernetes. More Executors Per Node for YARN.

The Problem of Resource Overprovisioning

Most application resources are overprovisioned. However, overprovisioning leads to underutilized nodes and wasted spend. Instance rightsizing and manual tuning or applying recommendations cannot address underutilization.

The Solution: Real-Time, Automated Resource Optimization for Data Workloads Running on Kubernetes and YARN

Pepperdata Capacity Optimizer remediates resource overprovisioning by providing the system scheduler with real-time visibility into actual CPU, GPU, and memory utilization levels **to automatically deliver more containers per node for both Kubernetes and YARN environments.**

This automatic and continuous **resource rightsizing maximizes utilization levels** for both the peaks and valleys of application runtime **without the need for manual tuning, application code changes, or applying recommendations.**

Supported Environments

- Apache Spark, Apache Airflow, Jobs, and CronJobs on Kubernetes, Amazon EKS, Google GKE
- Apache Spark on Cloudera Data Engineering
- Amazon EMR, Google DataProc
- Cloudera Data Platform (CDP)

Supported Schedulers

- Default scheduler on Amazon EMR and EKS and Google GKE
- Apache YuniKorn on Amazon EKS

Supported Autoscalers

- Amazon EMR Managed Autoscaling and Custom Autoscaling Policy on Amazon EMR
- Cluster Autoscaler and Karpenter on Amazon EKS
- Cluster Autoscaler with and without Node Auto-Provisioning (NAP) on Google GKE

Pepperdata's Four Components of Intelligent Kubernetes Resource Optimization

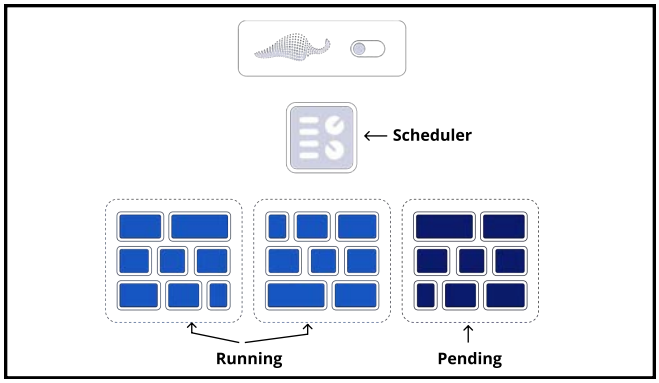
With Pepperdata Capacity Optimizer installed:

1. **The scheduler** makes more accurate and efficient resource decisions with Pepperdata-provided real-time CPU and memory usage information.
2. **Workloads on nodes** are launched based on real-time physical utilization.
3. **Pods** are launched with optimized resource requests.
4. **The autoscaler** can scale up more efficiently since nodes are packed based on actual utilization.

More Pods Per Node: The Savings Impact of Pepperdata Capacity Optimizer

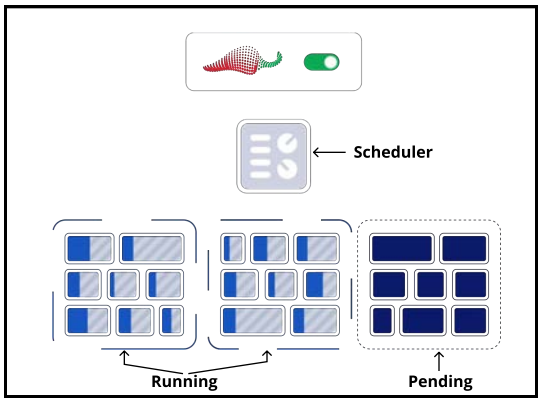
Before Pepperdata Capacity Optimizer is enabled, nodes in the cluster appear fully utilized, even when they are not. The scheduler cannot launch pending pods without resorting to autoscaling to spin up additional nodes at additional cost.

Without Pepperdata Capacity Optimizer, Cluster Nodes Appear Full

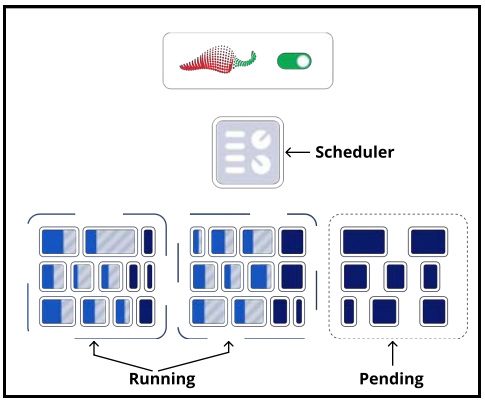


With Pepperdata Capacity Optimizer Enabled

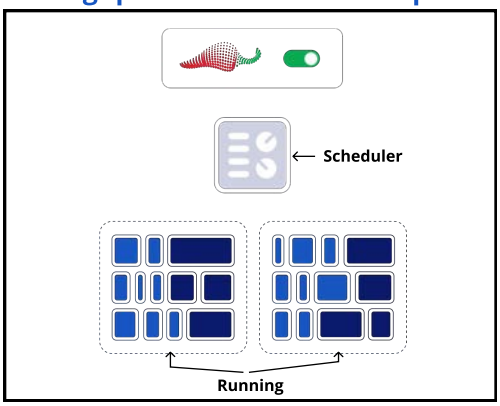
The Scheduler Immediately Becomes Aware of Existing But Unused Node Capacity



Pending Pods Are Now Launched To Fill Actual Available Resources



Costs Are Reduced, Resource Utilization Is Increased, and Throughput Performance Is Improved



The moment Capacity Optimizer is enabled, it shares information on **actual hardware utilization**, signaling to the scheduler where more node capacity is available. **Pending pods can then be launched on existing nodes with extra capacity**—nodes that previously appeared fully utilized to the scheduler.

With Pepperdata Capacity Optimizer, more pods can run per node since the scheduler now perceives available capacity and can launch pending pods based on actual utilization. With Capacity Optimizer, nodes run at greatest utilization and lowest cost automatically.

A Pepperdata TPC-DS benchmark for an Apache Spark workload running on Amazon EKS found a 41.8 percent reduction in total instance hours consumed, showcasing a dramatic improvement in resource utilization that ultimately translates to reduced cost.

Pepperdata Capacity Optimizer Enhances Your Autoscaler's Efficiency

Capacity Optimizer also enhances the efficiency of your cloud autoscaler by ensuring new nodes are provisioned only when existing nodes are fully utilized. This optimizes node scaling without altering the autoscaler's downscaling behavior. In practice, Pepperdata's customers have noted up to a **71 percent decrease in instance hours when using autoscaling** once Capacity Optimizer was enabled.

If there are any pods in the pending state, Pepperdata Capacity Optimizer enables the autoscaler (e.g. Karpenter) to add new nodes **only when all existing nodes are fully utilized**. Otherwise, there is no need for the autoscaler to add new nodes since all nodes are running at maximum utilization.

YARN/Amazon EMR Environments

Similar to how it works in Kubernetes, Pepperdata Capacity Optimizer provides the actual usage information to the YARN/Amazon EMR scheduler so that new containers are launched based on actual usage rather than on allocated resource levels. Capacity Optimizer also guides the Amazon EMR autoscaler to add new nodes only when existing ones are fully utilized.

In this manner, Capacity Optimizer improves utilization, reduces instance hours, and lowers cloud costs in YARN, Amazon EMR, and Google Dataproc environments.

Pepperdata Capacity Optimizer Benefits

Improve Resource Utilization
By creating more containers per node, Capacity Optimizer increases utilization by up to 80 percent—eliminating waste and maximizing efficiency.

Reduce Cost Across All Your Clusters
By optimizing resources automatically in real time, Capacity Optimizer cuts cloud spend by 30 percent on average—with no developer effort.

Free Engineers with No Manual Tuning, No Recommendations, No Application Code Changes
Capacity Optimizer works continuously and automatically in the background with no tedious recommendations to implement and no need for developers to modify code or tune configs.

Seamless Installation
Capacity Optimizer is deployed in under an hour using a simple bootstrap script (for Amazon EMR) or a Helm chart (for Amazon EKS).

About Pepperdata

Pepperdata Capacity Optimizer provides automated, continuous, and real-time Kubernetes and YARN resource optimization. It boosts utilization by up to 80%, enhances performance, and cuts costs without requiring application code changes.

Customers typically reduce infrastructure by 30% on average, freeing developers from manual tuning to focus on revenue-generating innovation. Our customers include top enterprises such as Citibank, Autodesk, Royal Bank of Canada, members of the Fortune 10, and mid-sized companies.

Pepperdata, Inc.
530 Lakeside Drive
Suite 170
Sunnyvale, CA 94085

Start a Free Trial
www.pepperdata.com

Send an Email
eval@pepperdata.com