

# Pepperdata Capacity Optimizer for Long-Running Microservices on Kubernetes

Achieve Immediate Utilization Improvements with Dynamic Kubernetes Resource Optimization

## The Challenge of Resource Utilization in Long-Running Microservices

Setting resource requests (CPU and memory) for microservices directly impacts performance, stability, and costs. DevOps often add large buffers to ensure their workloads run to completion, leading to overallocated resources and wasteful spending. Kubernetes resource requests and limits can be managed either manually or automatically.

**Manual configuration** requires constant monitoring and adjustments, which are time-consuming and quickly become outdated, leading to wasted resources or performance issues. As resource requirements constantly fluctuate, responding to those changes is difficult, if not impossible, to match manually—especially at scale.

**Automatic configuration** removes this overhead by continuously adjusting requests and limits in real time based on actual usage. For dynamic or growing workloads, automation is the more scalable and sustainable choice.

## The Solution: Dynamic Resource Optimization

Pepperdata Capacity Optimizer aligns the resource allocations of service pods with *actual* resource usage to provide more pods per node, increased utilization, and lower costs to operate your Kubernetes environment.

Capacity Optimizer monitors individual node utilization and ensures that a node is never under resource pressure due to resource adjustments. Capacity Optimizer then generates pod rightsizing recommendations based on observed services running in a cluster. Workload owners can choose to have the recommended optimizations applied automatically, manually, or skipped altogether. Optimization can be done at the service level, namespace level, or cluster level, and the application of recommendations does not disrupt the service or pods in any way.

Capacity Optimizer also enhances cloud autoscaling efficiency by ensuring that new nodes are provisioned only when the existing nodes are fully utilized.

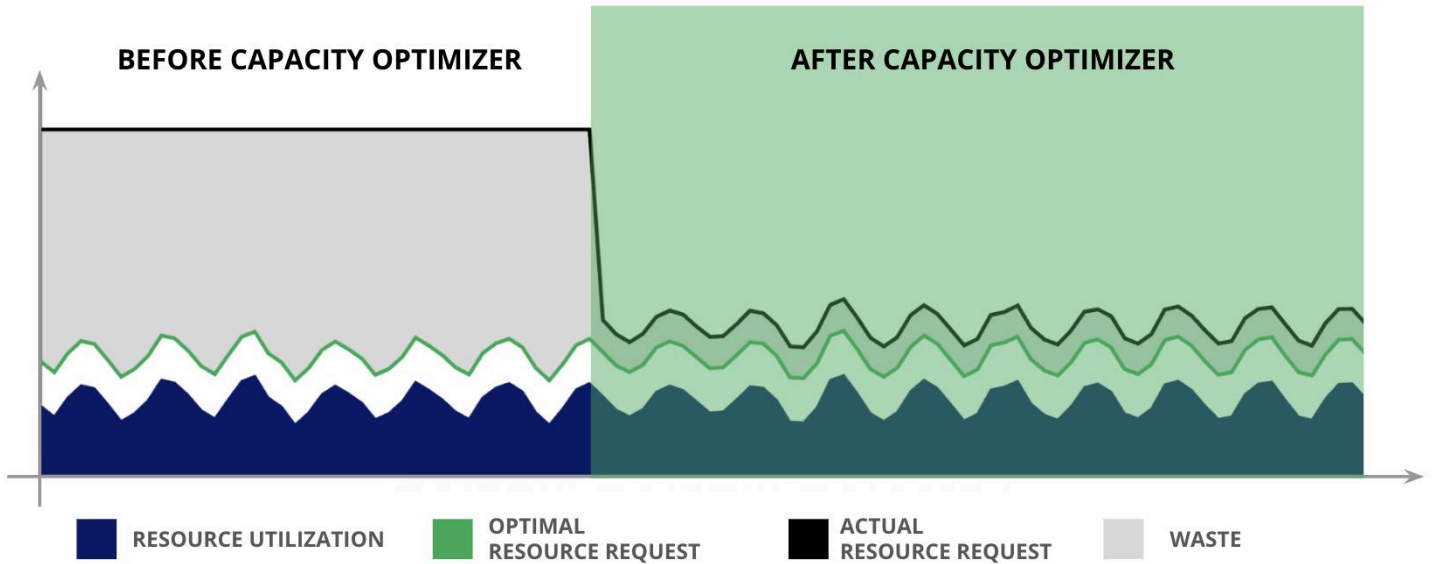
## How Capacity Optimizer Enhances the Autoscaler

Capacity Optimizer functions like a Vertical Pod Autoscaler (VPA), but unlike VPA, it operates in real time and is fully compatible with the Horizontal Pod Autoscaler (HPA). Note: To avoid conflicts, any existing VPA should either be disabled or set to recommendation-only mode.

- As workload traffic grows, HPA scales out by increasing the number of pods.
- With Capacity Optimizer actively reducing pod resource requests based on actual usage, more pods can be scheduled onto existing nodes—maximizing node utilization.
- If additional pods are still pending, the cluster autoscaler provisions new nodes, but scaling decisions are now based on true utilization rather than inflated, static resource requests.

## After Enabling Capacity Optimizer

Pepperdata Capacity Optimizer aligns the CPU and memory resource allocations of service pods with *actual* resource usage to provide more pods per node, increased utilization, and lower costs to operate your Kubernetes environment.



### Capacity Optimizer Benefits for Long-Running Microservices



#### Real-Time Reclamation of Resource Waste

Align resource requests with actual hardware usage to reduce waste and lower cloud bills, without impacting response time.



#### Increased Utilization Without Disruption

Pack more pods on existing physical nodes by working seamlessly with HPA and without any pod restarts.



#### Enhanced Autoscaling Efficiency

Launch new nodes only when existing nodes are fully utilized.

### Supported Technologies

- Amazon EKS, GKE, Microsoft Azure
- Static clusters or autoscaled clusters
- Cluster Autoscaler, Karpenter, GKE NAP
- Horizontal Pod Autoscaler (HPA)
- Kubernetes Event Driven Autoscaler (KEDA)
- GitOps tools that control the service lifecycle
- Supported Kubernetes controllers: Deployments, ReplicaSet, StatefulSets, DaemonSets, Jobs and CronJobs, Custom Controllers

### About Pepperdata

Deployed on over 30,000+ clusters, Capacity Optimizer optimizes resources in some of the largest and most complex environments in the world, providing more pods per node in Kubernetes environments. Since 2012 Pepperdata has helped companies ranging from startups and mid-sized ISVs to top enterprises such as Citibank, Autodesk, Magnite, Royal Bank of Canada, and members of the Fortune 10 save over \$250 million. For more information, visit [www.pepperdata.com](http://www.pepperdata.com).



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



**Start a Free Trial**  
[www.pepperdata.com](http://www.pepperdata.com)



**Send an Email**  
[info@pepperdata.com](mailto:info@pepperdata.com)