Pepperdata Capacity Optimizer for Apache Airflow

Realize up to 40% Cost Savings for Apache Airflow with Dynamic Resource Optimization

The Challenge of Resource Utilization with Apache Airflow

Apache Airflow is a powerful open-source platform for orchestrating and scheduling data workflows. It provides a framework that enables users to manage, run, and monitor their workflows.

A significant challenge with Airflow clusters is that they often suffer from overprovisioning of CPU and memory, leading to low utilization and wasted spending. This occurs because the KubernetesExecutor launches a separate pod for each task, and task resource needs vary. DevOps teams, finding it cumbersome to set individual task resources, often configure a single, large global pod size. This results in hundreds of overprovisioned tasks, causing significant wasted spend, unused capacity, and low resource utilization—especially when multiplied over hundreds of workloads launching simultaneously.

The Solution: Dynamic Resource Optimization with Pepperdata Capacity Optimizer

Pepperdata Capacity Optimizer for Apache Airflow enables more work on fewer nodes by

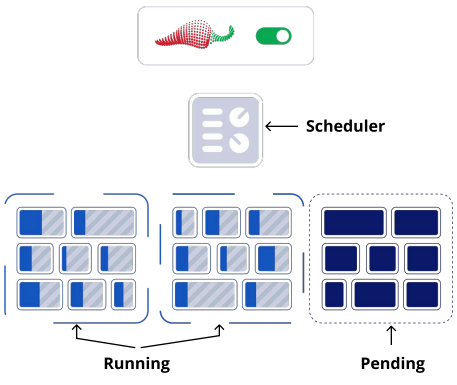
1. Reducing resource requests so more pods can be packed on existing nodes
2. Providing the scheduler with visibility into utilized resources so that it can schedule more pods on existing nodes
3. Packing all existing nodes first before enabling the autoscaler to launch new nodes

The result is reduced infrastructure cost in the cloud, increased resource utilization, and more workloads running on the same cluster.

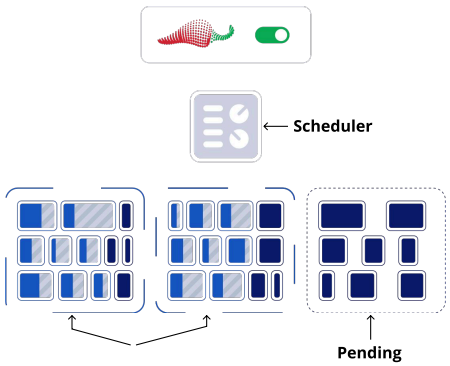
With Capacity Optimizer Enabled

Before Capacity Optimizer is enabled, the scheduler sees all instances as full. The moment Capacity Optimizer is enabled, more pods automatically run per node since the scheduler now perceives available capacity and can launch pending pods based on actual utilization—resulting in nodes running at their greatest utilization and lowest cost.

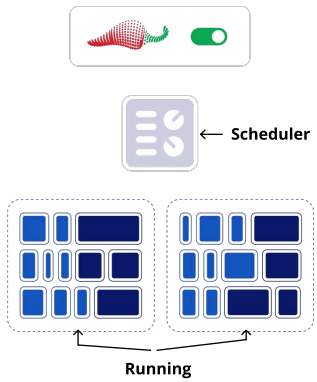
1. **The scheduler** immediately becomes aware of existing but unused node capacity.



2. **Pending pods** are now launched to fill actual available resources.



3. **Resource utilization is increased**, costs are reduced, and throughput performance is improved.



Capacity Optimizer Benefits for Apache Airflow

Cost-Effective Airflow Operations

Eliminate the cost and waste resulting from task pods allocating more resources than they ever use.

Improved Throughput

Reduce pod resource requests automatically so that all existing nodes are continuously optimally packed at optimal capacity.

Significantly More Efficient Autoscaling

Ensure new nodes are provisioned only when existing nodes are fully utilized.

No Need For Manual Tuning, Recommendations, or Application Code Changes

Free developers from optimization tasks so they can focus on revenue-generating projects and innovation.

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.

aws partner network
Select Technology Partner

AICPA SOC
aicpa.org/soc
Formerly SAS 70 Reports

100% ROI GUARANTEE

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

Start a Free Trial
www.pepperdata.com

Send an Email
info@pepperdata.com