



# Turbocharge Your Data Platform with Qubole Cloud Agents

Part of Qubole's Autonomous Data Platform

Managing your data platform typically requires a lot of tedious manual work and/or writing scripts to orchestrate workloads. With Qubole Cloud Agents, your data platform self-manages, learns and is context aware. It captures and analyzes the meta-data generated by your platform usage and leverages machine learning to intelligently automate key data management tasks. Cloud agents are policy based, and provide a number of benefits:



## Qubole offers three agents, including:



## Workload Aware Auto-Scaling Agent

Big data workloads are bursty —they require varying amounts of compute to process data. A fixed-size cluster, which has to be provisioned to the peak workload usage creates a lot of idle compute and wasted spend. The Workload Aware Auto-scaling Agent augments the basic auto-scaling feature available in Qubole’s Enterprise Edition, to make it more cost efficient and able to handle a broader variety of auto-scaling scenarios.

**The Workload Aware Auto-scaling Agent can reduce compute spend by as much as 33% over basic auto-scaling solutions available in the market today.**

Qubole’s Workload Aware Auto-scaling Agent is:



### Intelligent

Constantly analyzes the workload and the infrastructure utilization to determine which resources should be scaled up/down



### Autonomous

Requires minimal input or manual management from users to trigger auto-scaling



### Performance

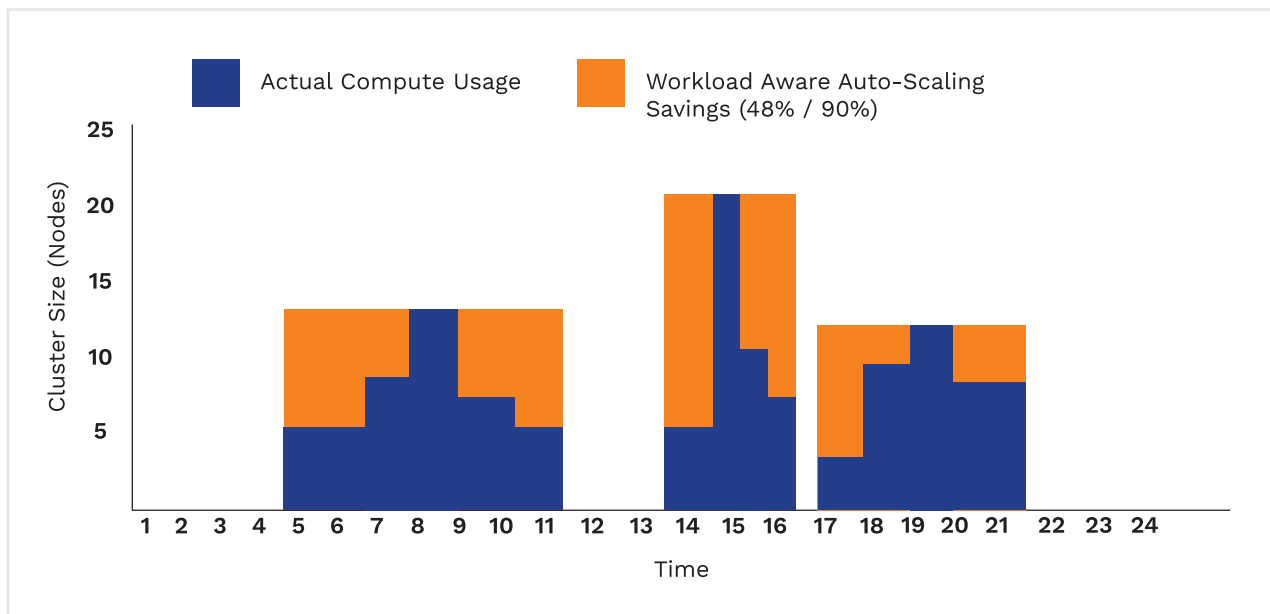
**2x faster** than static clusters



### Cost Efficient

**33% cheaper** than static clusters  
  
**Up to 90% cheaper** with heterogeneous  
  
**Up to 80% cheaper** with spot instances

## Maximizing Cost Efficiency



Workload Aware Auto-scaling offers the following capabilities:

### EBS-based

When a cluster has sufficient compute resources but requires additional storage, the agent can dynamically add storage using EBS to avoid provisioning a new compute node.

### Aggressive Downscaling

Aggressive downscaling is triggered when you reduce the maximum size of a cluster while it's running. To save costs, QDS terminates nodes that are closest to completing their tasks and closest to their billing boundary.

### Offloading

Your mappers may be idling, waiting for reducers to finish their job. Offloading conserves compute resources by saving mapper data to HDFS or object storage.

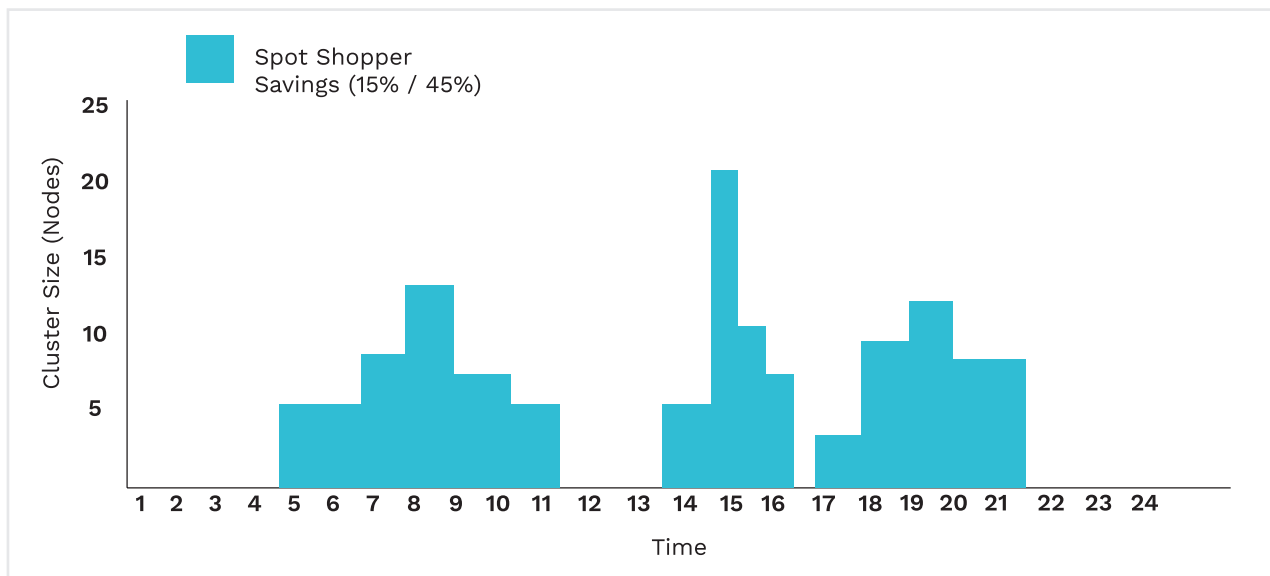
### Container Packing

New resource allocation strategy that makes more nodes available for downscaling in an elastic computing environment. This prevents hot spots in a cluster, while trying to honor data locality preferences.

## Spot Shopper Agent

The Spot Shopper Agent 'shops' for the best combination of price and performance, based on the policies you set. It achieves this by shopping across different instance types, by dynamically rebalancing Spot and On-demand nodes and by considering different Availability Zones.

**The Spot Shopper Agent can reduce compute spend by as much as 50% over solutions that exclusively rely on on-demand instance types.**



Spot Shopper offers the following capabilities:

### **Heterogeneous Clusters**

With Heterogeneous Clusters, slave nodes comprising the cluster may be of different instance types. Heterogeneity in Spot nodes is highly beneficial because Spot prices can change rapidly, and Spot Shopper Agent can make the lowest-cost purchasing decision in real time.

### **Availability Zone Selection**

Unless you specify a particular AZ when you configure the cluster, Qubole can automatically select the AZ with the lowest Spot prices for the region and instance type you've specified.

### **Spot Rebalancing**

Fluctuations in the market may mean that QDS cannot always obtain as many Spot instances as you have specified for a cluster. In these circumstances, the Spot Shopper Agent will automatically rebalance the cluster at a later time, when prices drop by swapping out on-demand nodes for Spot nodes. This ensures that your cluster is continuously optimized for the lowest prices possible.

### **Placement Policy**

The Placement Policy option enables QDS to make a best effort to store one replica of each HDFS block on a stable node. This prevents job failures that could occur if all replicas were lost as a result of AWS reclaiming many Spot instances at once.

## **Data Caching Agent**

The Data Caching Agent automates the movement of data for performance optimization.

### **Caching from Object Store**

The Data Caching Agent automatically determines the right set of data to cache in the cluster so that interactive, ad-hoc queries run faster and don't need to retrieve data for each query.

### **Caching of Index Metadata**

Qubole's Data Caching Agent makes optimal use of ORC, Parquet and Avro data formats by minimizing the amount of data that's read when selecting only specific columns.

“ Qubole’s container packing has helped us reduce our EC2 costs by enabling a lazy upscale and an aggressive downscale without compromising on performance



**ENDY LAMBEY**

Data Governance, Operations & Architecture

traveloka 

## Turbocharge Your Data Platform Today

Contact us to learn about our subscription based pricing. Our Cloud Agents product is available to Qubole Enterprise Edition customers only.

Learn more at [www.qubole.com](http://www.qubole.com)

### About Qubole

Qubole is passionate about making data-driven insights easily accessible to anyone. Qubole customers currently process nearly an exabyte of data every month, making us the leading cloud-agnostic big-data-as-a-service provider. Customers have chosen Qubole because we created the industry’s first autonomous data platform. This cloud-based data platform self-manages, self-optimizes and learns to improve automatically and as a result delivers unbeatable agility, flexibility, and TCO. Qubole customers focus on their data, not their data platform. Qubole investors include CRV, Lightspeed Venture Partners, Norwest Venture Partners and IVP. For more information visit [www.qubole.com](http://www.qubole.com)