



Federator.ai User Guide

Federator.ai version 4.4 User Guide

ProphetStor Data Services, Inc.
830 Hillview Court, Suite 100
Milpitas, CA 95035 USA
Phone: 1.408.508.6255
Website: www.prophetstor.com

Copyright © 2020-2021 ProphetStor Data Services, Inc. All Rights Reserved.

Federator.ai® is a registered trademark of ProphetStor Data Services, Inc in the United States and other countries.

Kubernetes is a registered trademark of the Linux Foundation, and OpenShift is a trademark of Red Hat, Inc.

All other brand and product names are trademarks or registered trademarks of their respective owners.

1.26.2021

Contents

Overview	3
Terminology	3
Getting Started	5
Access the Federator.ai Portal.....	5
Setup Wizard	7
Federator.ai Administration Portal	11
Common Administration Portal Functions	11
Refresh Statistics	11
License Status	11
User Functions.....	11
Query Condition Panel.....	12
Search/Sort Information in Tables.....	12
Show/Hide Metrics in Charts.....	12
Dashboard	14
Cluster Workload Prediction	14
Cluster Charts	14
Node Chart	15
Namespace Chart	16
Application Workload Prediction.....	17
Application Charts	17
Controllers/Consumer Groups Chart	18
Insight - Cluster Health	20
Insight - Node Health	21
Planning - Workload Prediction	22
Managed Nodes Table.....	23
Managed Containers Table.....	23
Workload Prediction Table	24
CPU and Memory Charts	24
Recommendation - HPA Recommendation	25
CPU and memory Charts	25
Optimization – Kafka Consumer.....	26
Number of Replicas Chart.....	26

Production Rate and Consumption Rate Chart	26
Consumer Lag Chart	27
Consumer Queue Latency Chart.....	27
CPU and Memory Observation Charts.....	28
Cost – Multi-cloud Cost Analysis	29
CPU and Memory Utilization Charts.....	30
Recommended Cluster Configuration Tables	31
Cost - Cost Allocation	32
Current Cluster Configuration	32
Namespace Charts.....	32
Configuration - Applications.....	34
Add an Application	35
Generic:	35
Kafka:.....	36
Manage Applications.....	37
Configuration - Clusters	38
Add a Cluster	39
Manage Clusters.....	40
Configuration - System	41
Admin Password	41
Metrics Data Source	41
Licenses	42
Add a License.....	42
Manage Licenses	43
Cloud Provider Price Books.....	44
Events.....	45

Overview

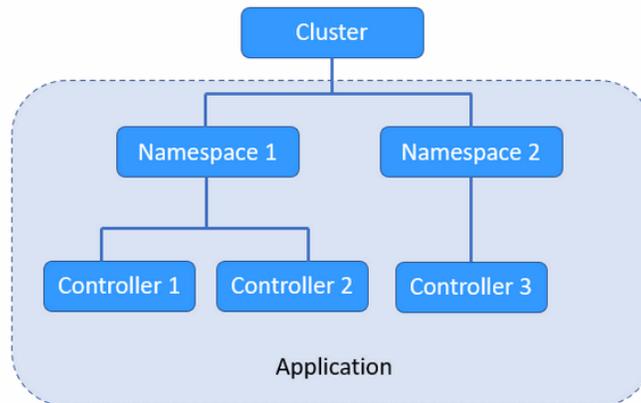
ProphetStor Federator.ai is an AI-based solution that helps enterprises manage, optimize, and auto-scale resources for any applications on Kubernetes. Using advanced machine learning algorithms to predict application workloads, Federator.ai scales the appropriate amount of resources at the right time for optimized application performance. Federator.ai offers:

- AI-based workload prediction for Kafka consumers and other applications
- Resource recommendations based on workload prediction, application, Kubernetes, and other related metrics
- Automatic scaling of application containers
- Multicloud cost analysis and recommendations based on workload predictions

If you have not installed Federator.ai yet, refer to your *Federator.ai Installation Guide* for information.

Terminology

Application – Defined by Federator.ai as a group of Kubernetes controllers that work together to serve tasks from the view of the end user. For example, an e-commerce web application consists of controllers for frontend and backend and can be considered as an application. An application is not a Kubernetes object.



Autoscaling – Horizontal Pod Autoscaling (HPA) provides the ability for the system to automatically increase or decrease containers/pods.

Container – An object that contains a software module with everything needed to run an application.

Controller – In Kubernetes, controllers are control loops that watch the state of your cluster, then make or request changes where needed. Each controller tries to move the current cluster state closer to the desired state. The types of controllers supported by Federator.ai are *Deployment* and *StatefulSet*. Additionally, Federator.ai supports *DeploymentConfig* controllers for OpenShift.

Cluster – Kubernetes cluster with one or more nodes.

Deployment – A Deployment provides declarative updates for Pods and ReplicaSets. The user describes a desired state in a deployment and the deployment controller changes the actual state to the desired state at a controlled rate.

HPA – Horizontal Pod Autoscaling – The system automatically increases or decreases the number of containers/pods based on the CPU/memory usage workload or custom workload metrics.

Namespace – Kubernetes supports multiple virtual clusters backed by the same physical cluster. These virtual clusters are called namespaces.

Node – Server-like machine, such as a virtual machine running complete systems and multiple applications. There can be master nodes and worker nodes.

Pod – A group of one or more containers with shared storage/network resources and a specification for how to run the containers. Typically, one container runs in each pod.

Replica – A copy of a pod running for an application.

StatefulSet – A Kubernetes object that manages stateful applications. Unlike a Deployment, a StatefulSet maintains a sticky identity for each of its pods that remains the same across any rescheduling.

Related topic:

[Federator.ai Administration Portal](#)

[Configure Applications](#)

[Configure Clusters](#)

[System Configuration](#)

Getting Started

After installation of Federator.ai, you must access the Federator.ai portal in order to configure your system.

Access the Federator.ai Portal

To access the Federator.ai administration portal, use the URL that is displayed at the end of the installation process.

You can also find the URL for the Federator.ai administration portal via the following methods:

Kubernetes

In a Kubernetes environment, use the `kubectl` command to find the administration portal service port number and node IP address.

```
# kubectl get svc -n federatorai |grep federatorai-dashboard-frontend-node-port
```

The output will look something like this:

```
federatorai-dashboard-frontend-node-port NodePort 10.103.181.133 <none> 9001:31012/TCP
```

Get the node's IP to access (INTERNAL-IP).

```
$kubectl get nodes -o wide
```

For example:

```
# kubectl get nodes -o wide
NAME          STATUS    ROLES    AGE   VERSION   INTERNAL-IP   EXTERNAL-IP   OS-IMAGE
KERNEL-VERSION   CONTAINER-RUNTIME
h7-130      Ready    master   35d   v1.18.5   172.31.7.130  <none>        CentOS Linux 7 (Core)
3.10.0-957.el7.x86_64  docker://19.3.13
h7-131      Ready    <none>   35d   v1.18.5   172.31.7.131  <none>        CentOS Linux 7 (Core)
3.10.0-957.el7.x86_64  docker://19.3.13
h7-132      Ready    <none>   35d   v1.18.5   172.31.7.132  <none>        CentOS Linux 7 (Core)
3.10.0-957.el7.x86_64  docker://19.3.13
h7-133      Ready    <none>   35d   v1.18.5   172.31.7.133  <none>        CentOS Linux 7 (Core)
3.10.0-957.el7.x86_64  docker://19.3.13
```

The URL will be `https://172.31.7.130:31012`.

OpenShift

In an OpenShift environment, use the `oc get route` command to find the URL.

```
# oc get route -n federatorai | grep federatorai-dashboard-frontend
```

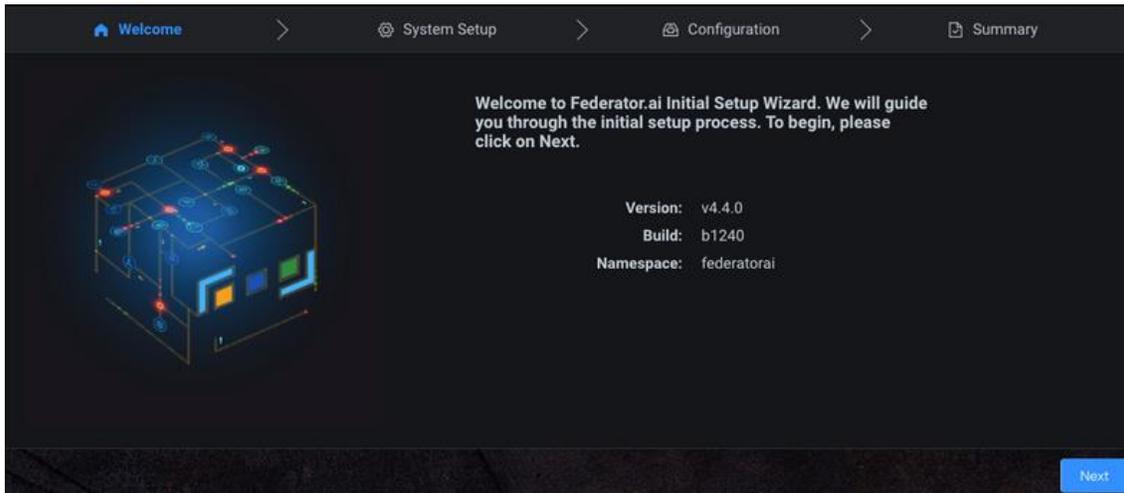
The output will look something like this:

```
federatorai-dashboard-frontend "federatorai-dashboard-frontend-federatorai.apps.ocp4.172-31-11-30.nip.io"
```

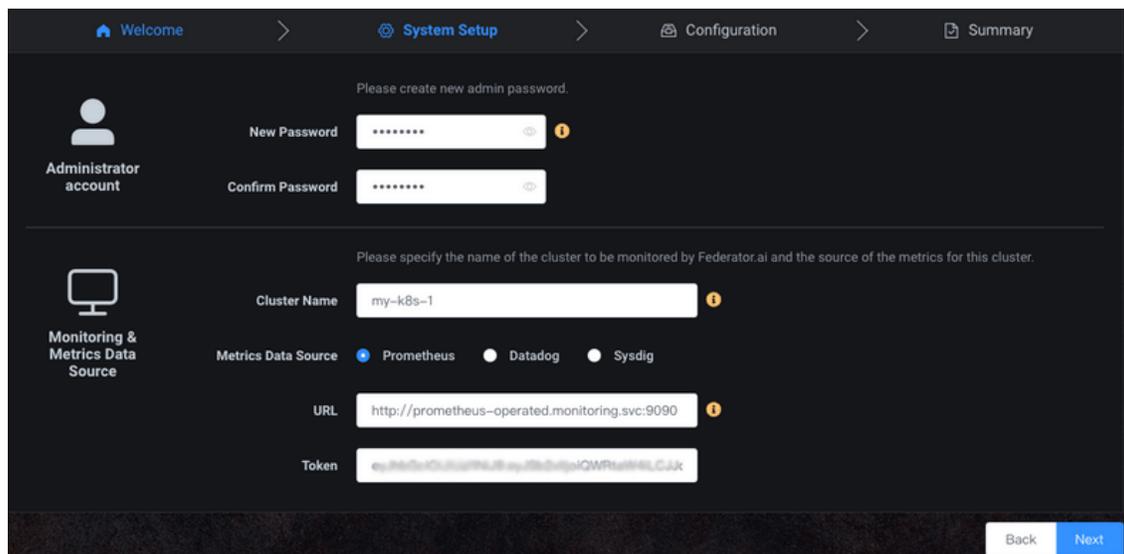
The URL will be `https://federatorai-dashboard-frontend-federatorai.apps.ocp4.172-31-11-30.nip.io`

Setup Wizard

The first time you log in after installation of Federator.ai, a configuration wizard launches that allows you to configure the applications that should be monitored by Federator.ai.



1. Click *Next* to begin configuration.
2. Set the administrator password, specify the name of a cluster to be monitored by Federator.ai, and the source of metrics for this cluster. Click *Next* when done.



You must set the administrator password in order to continue but your cluster can be configured at a later time from the *Configuration* section of the portal after completing the setup wizard.

A cluster is a Kubernetes cluster with one or more nodes. You can only configure one cluster in the setup wizard. You can add clusters from the *Configuration* section of the portal after completing the setup wizard.

For the Prometheus open-source monitoring system, the URL is required but the token is optional for authentication. If needed, you can click on the link to the Prometheus website, which is included in the popup help text.

For Datadog, the API key and application key are required for authentication. If needed, you can click on the link to the Datadog website, which is included in the popup help text.

This screenshot shows the 'Monitoring & Metrics Data Source' configuration screen. The 'Cluster Name' is 'my-k8s-1'. The 'Metrics Data Source' is set to 'Datadog'. The 'API Key' and 'Application Key' fields are filled with masked values. There are help icons next to the API and Application Key fields. At the bottom, there are 'Back' and 'Next' buttons.

For Sysdig, a URL and token are required for authentication. If needed, you can click on the link to the Sysdig website, which is included in the popup help text.

This screenshot shows the 'Monitoring & Metrics Data Source' configuration screen. The 'Cluster Name' is 'my-k8s-1'. The 'Metrics Data Source' is set to 'Sysdig'. The 'URL' field is 'https://app.sysdigcloud.com' and the 'Token' field is filled with a masked value. There are help icons next to the URL and Token fields. At the bottom, there are 'Back' and 'Next' buttons.

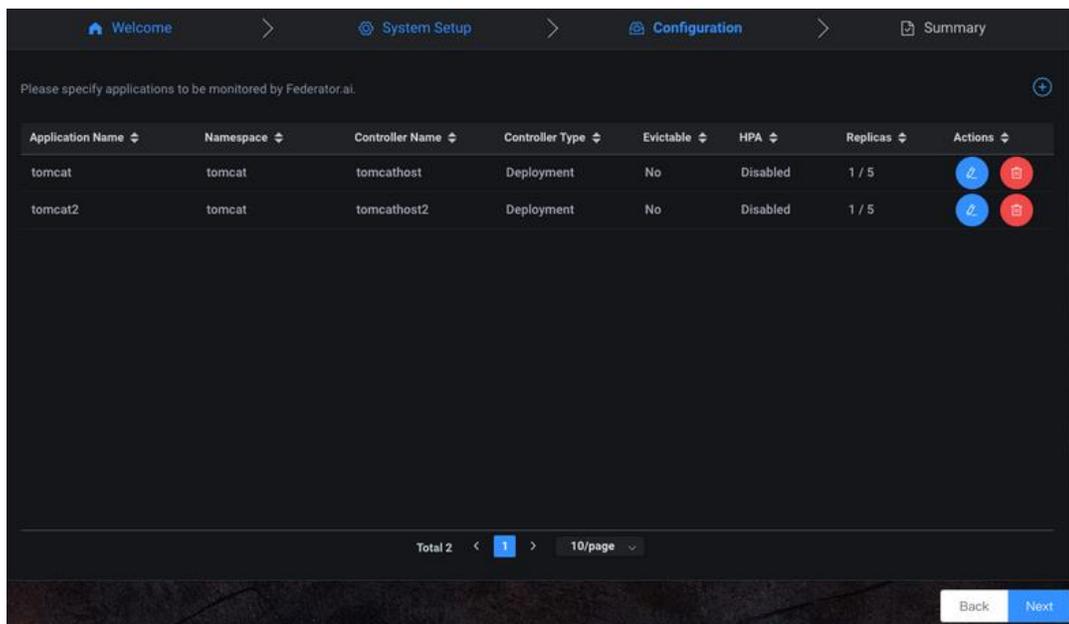
3. Enter information about the first application you want to monitor.

This screenshot shows the 'Application Info' configuration screen. It includes fields for 'Application Name', 'Namespace', and 'Controller Name'. The 'Controller Type' is set to 'Deployment'. The 'Evictable' option is set to 'No'. The 'HPA Recommendation' is set to 'Disabled'. The 'Min Replicas' is set to 1 and the 'Max Replicas' is set to 5. There are 'Skip' and 'Save' buttons at the bottom right.

You can add applications now or from the *Configuration* section of the portal after completing the setup wizard.

- *Application Name* - The name of your application to be monitored by Federator.ai. An application is a group of one or more Kubernetes controllers that work together to serve tasks from the view of the end user; an application is not a Kubernetes object.
- *Namespace* – The Kubernetes namespace where the controller is deployed.
- *Controller Name* - The name of controller to be monitored.
- *Controller Type* – Supported controller types are *Deployment*, *StatefulSet*, and *DeploymentConfig* (OpenShift only).
- *Evictable* – Indicate if the controller can be interrupted if the node is shut down. Evictable controllers are good candidates to be deployed in Spot instances.
- *HPA Recommendation* – Indicate if you want to enable Horizontal Pod Autoscaling (HPA). When enabled, CPU and memory usage is monitored, and the number of pods is automatically increased/decreased based on the CPU/memory usage workload.
- *Mix/Max Replicas* – Specify the minimum and maximum number of pods when HPA is enabled.

4. Click *Save* to view a list of applications that will be monitored by Federator.ai.



Click the + key to add an application.

Click the blue *Edit* icon to update application information or click the red *Delete* icon to remove an application.

5. Click *Next* to view the *Summary* screen.

The screenshot displays the 'Summary' screen of the Federator.ai Administration Portal. The navigation bar at the top includes 'Welcome', 'System Setup', 'Configuration', and 'Summary'. The main content is divided into three sections:

- Administrator account:** Shows a 'New Password' field with a masked input and an eye icon.
- Monitoring & Metrics Data Source:** Shows fields for 'Cluster Name' (my-k8s-1), 'Metrics Data Source' (Prometheus), 'URL' (http://prometheus-operated.monitoring.svc:9090), and 'Token'.
- Application Info:** A table listing applications with columns for Application Name, Namespace, Controller Name, Controller Type, Evictable, HPA, and Replicas.

Application Name	Namespace	Controller Name	Controller Type	Evictable	HPA	Replicas
tomcat	tomcat	tomcathost	Deployment	No	Disabled	1 / 5
tomcat2	tomcat	tomcathost2	Deployment	No	Disabled	1 / 5

At the bottom of the table, it says 'Total 2'. At the bottom right of the screen, there are 'Back' and 'Apply' buttons.

6. Click *Apply* to apply all changes or click *Back* to edit information.

Related topics:

[Terminology](#)

[Federator.ai Administration Portal](#)

[Configure Applications](#)

[Configure Clusters](#)

[System Configuration](#)

Federator.ai Administration Portal

The Federator.ai administration portal displays the overall health of each cluster, as well as application workload and resource recommendations. Information is presented in tables and charts.

Common Administration Portal Functions

The administration portal presents information in tables and charts. At the top right of each portal page, you can do the following:

- Refresh statistics
- Check license status
- Get technical support contact information
- View Federator.ai product documentation
- Display the product software version
- Log out



Refresh Statistics

By default, Federator.ai information is refreshed every five minutes. To change the interval, click the drop-down at the top right and select a 1, 5, 15, or 30-minute interval.

To force a refresh, click *Refresh Now* where the current interval is displayed.

License Status

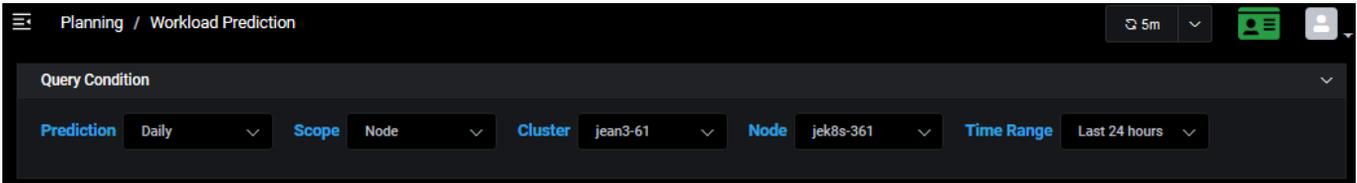
Click the *License* icon at the top right of the dashboard to see Federator.ai license information, including key code, type, CPU cores being monitored, as well as license expiration and status.

User Functions

Click the *User* icon to contact technical support, view the Federator.ai product documentation, display the product software version, or log out from the system.

Query Condition Panel

The *Query Condition* panel appears on most pages and allows you to filter the display of information. Depending upon the page, you may be able to set a time range for data and/or you may be able to select a prediction window, scope/resource type, cluster, application, namespace, controller, etc.



Search/Sort Information in Tables

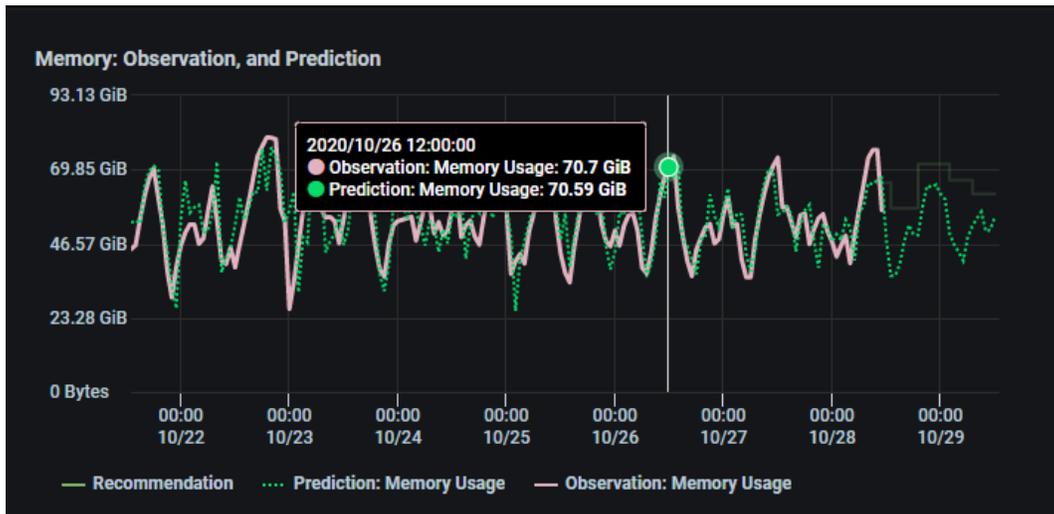
Click a column heading to sort the entire list based upon values in that column. The blue highlighted triangle or inverted triangle on the column indicates the direction of the sort. To search, type a name or value in the *Search* box; clear the search field to return to the full view of the list. As soon as you start typing, only those items that have matching text are displayed. You can also determine how many rows to show per page (5, 10, or 20).

Container ↕	Project (Names...	Application ↕	Pod ↕	Node ↕
my-nginx	nginx2	nginx2-jason-6-...	my-nginx-6f97b...	h6-182
my-nginx	nginx1	alamedascal-...	my-nginx-6c99d...	h6-182
my-nginx	nginx1	alamedascal-...	my-nginx-6c99d...	h6-182
my-nginx	nginx1	alamedascal-...	my-nginx-6c99d...	h6-182
my-nginx	nginx1	alamedascal-...	my-nginx-6c99d...	h6-182

Total 6 < 1 2 > 5/page

Show/Hide Metrics in Charts

Click anywhere on a chart to see values for a specific point in time. Highlight or click on the key at the bottom of the chart to show/hide individual metrics.



Related topics:

[Common Administration Portal Functions](#)

[Refresh Statistics](#)

[License Status](#)

[User Functions](#)

[Search/Sort Information in Tables](#)

[Show/ Hide Information in Charts](#)

[Query Condition Panel](#)

[Dashboard](#)

Dashboard

The *Dashboard* displays the number of monitored resources, as well as cluster and application workload predictions and recommendations.

Cluster Workload Prediction

Select a cluster from the drop-down list and select the timeframe (daily, weekly, or monthly) to display CPU and memory observations, predictions, and recommendations.

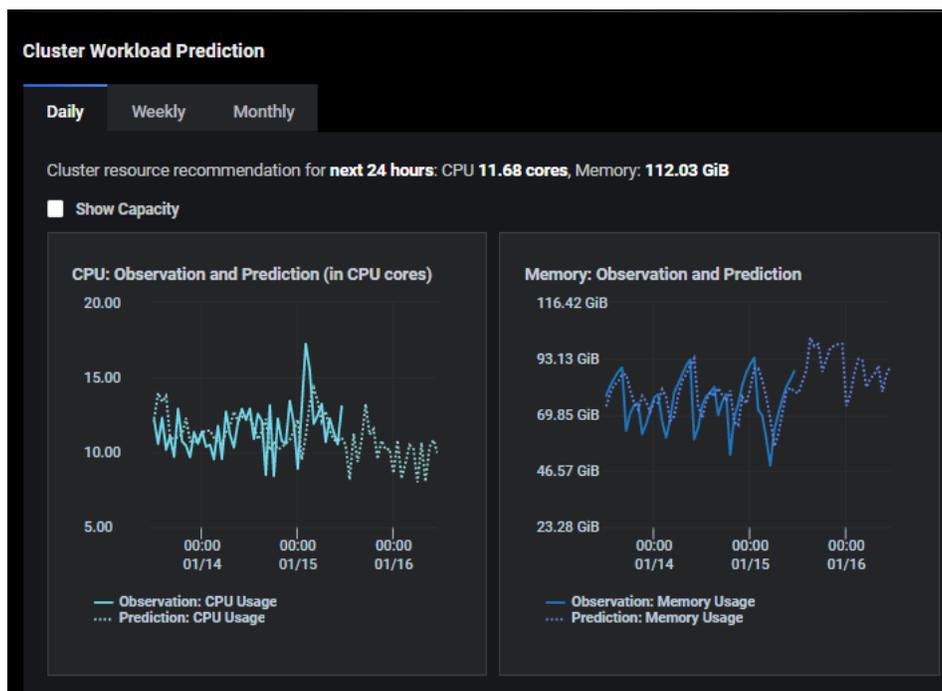
The first two charts display information for the selected cluster; the chart on the right toggles the display of information for nodes and namespaces of the selected cluster.

The text above the charts summarizes the CPU and memory recommendations for the next 24 hours (daily), 7 days (weekly), or 30 days (monthly).

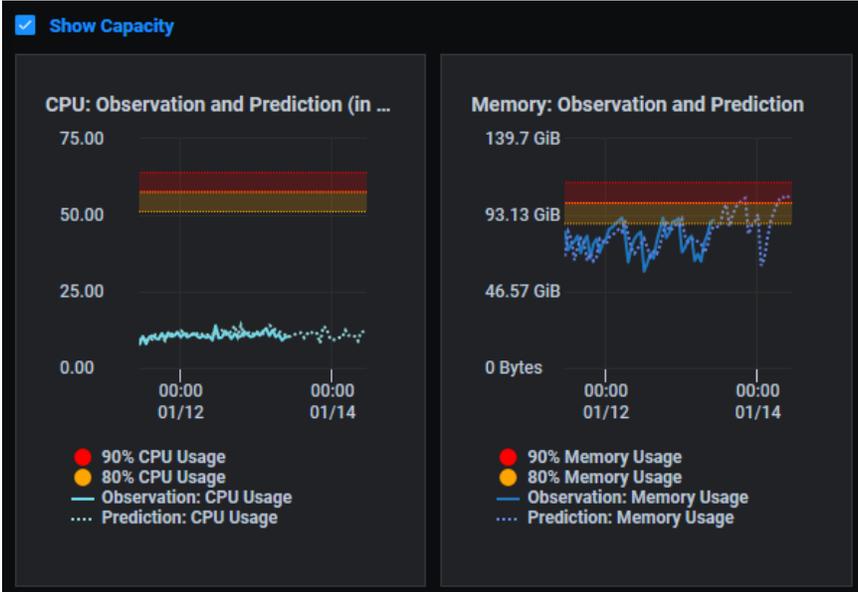
Cluster Charts

The cluster charts display CPU and memory observations and predictions for the cluster.

The solid lines represent the observed actual usage while the dotted lines show the historical and future predicted usage. Click anywhere on the charts to see values for a specific point in time. This will adjust the slider in the Node/Namespaces chart accordingly.



Select *Show Capacity* to see the maximum CPU and memory usage limits for the cluster. Orange represents 80-90% and red represents 90-100%. This is a useful way to see if the utilization of resources is approaching the overall cluster capacity.

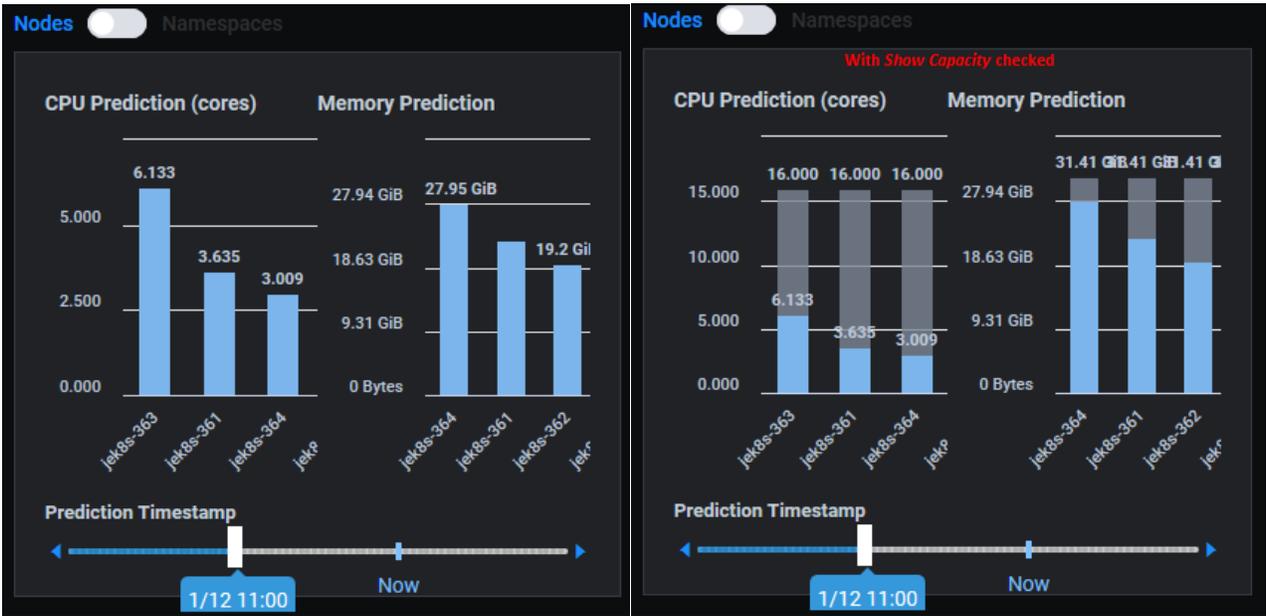


Node Chart

Toggle to *Nodes* to display CPU and memory observations and predictions for each node.

The slider always starts at the current time but allows you to select any day/hour. Slide to the left of *Now* for historical usage; slide right for future predictions.

If *Show Capacity* was selected for the cluster, the chart will include the total capacity of each node.

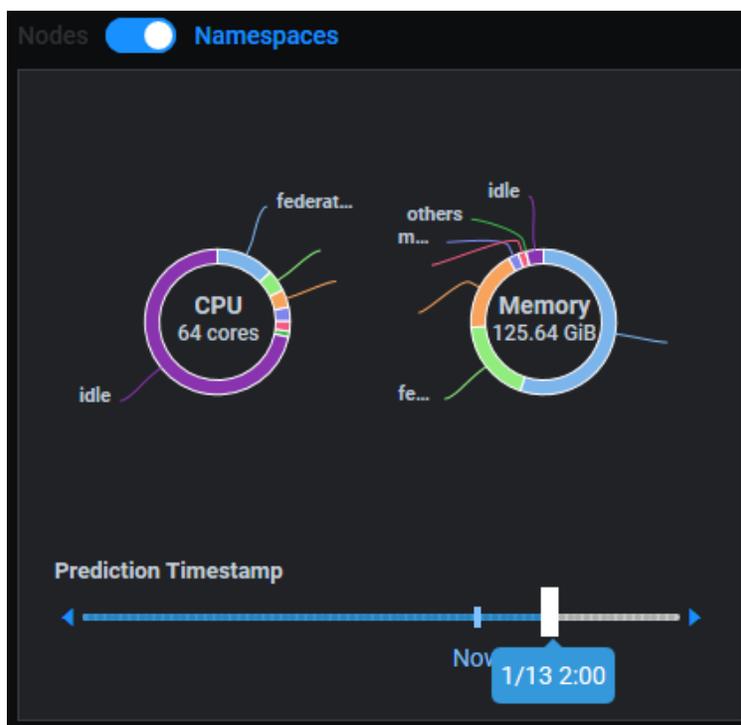


Namespace Chart

Toggle to *Namespaces* to display CPU and memory predictions for each namespace and for when the system is idle. Move your cursor over each section to show the usage or predicted usage by a specific namespace and the overall percentage used or predicted to use by the namespace or when the resource is idle.

Highlighting the amount of predicted idle (unused) resources provides a useful way to determine where your cluster is over-provisioned and can help you balance the resource allocation within the cluster.

The slider always starts at the current time but allows you to select any day/hour. Slide to the left of *Now* for historical usage; slide right for future predictions.



Application Workload Prediction

Select an application from the drop-down list and select the timeframe (daily, weekly, or monthly) to display CPU and memory observations and predictions.

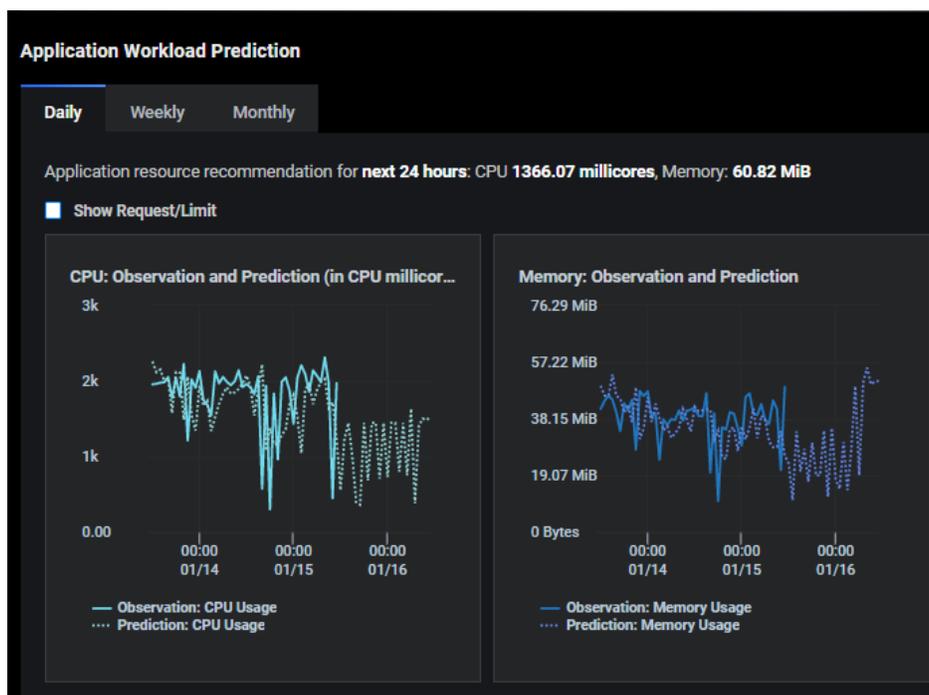
The first two charts display information for the selected application; the chart on the right displays information for controllers (generic application) or consumer groups (Kafka).

The text above the charts summarizes the CPU and memory recommendations for the next 24 hours (daily), 7 days (weekly), or 30 days (monthly).

Application Charts

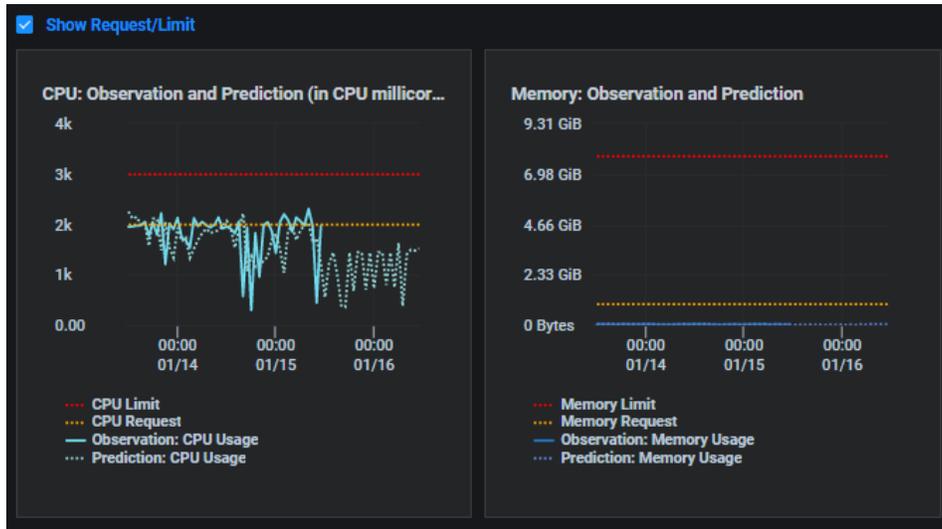
The application charts display CPU and memory observations, predictions, and recommendations for the application.

The solid lines represent the observed actual usage while the dotted lines show the historical and future predicted usage. Click anywhere on the charts to see values for a specific point in time. This will adjust the slider in the Controllers chart accordingly.



Select *Show Request/Limit* to see your application's CPU and memory usage limits in Kubernetes. Orange represents your resource request and red represents the resource limit.

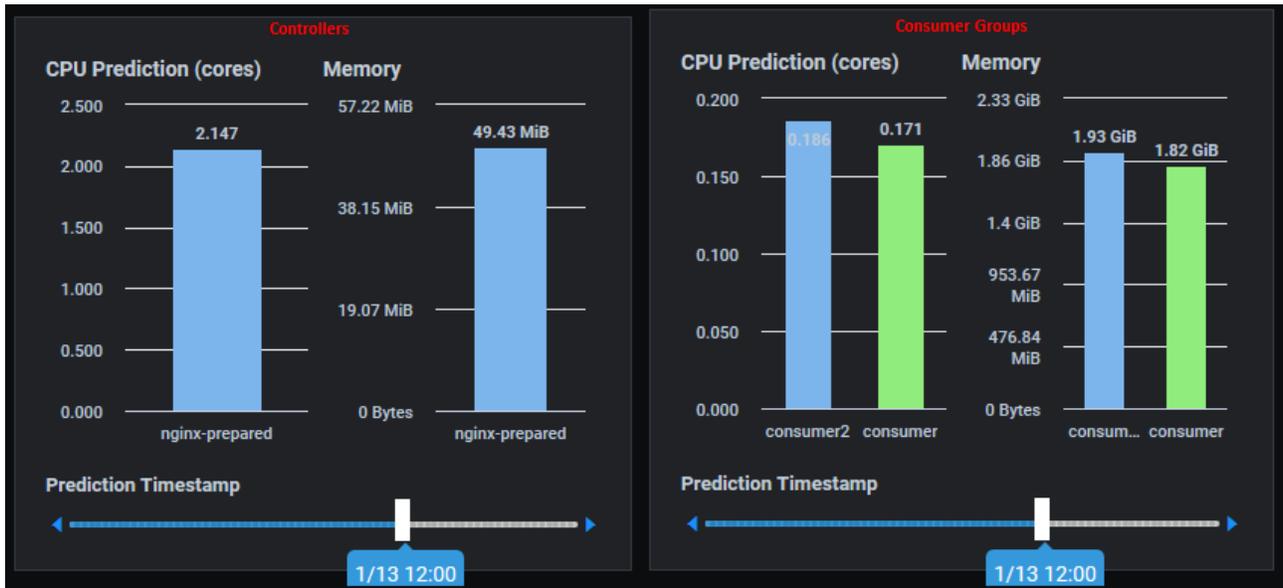
Generally, memory is a hard limit, but CPU is more *stretchable*. This is a useful way to see if you are over-provisioned or under-provisioned for your application.



Controllers/Consumer Groups Chart

The chart displays CPU and memory observations and predictions for each controller (generic application) or consumer group (Kafka) under the application.

The slider always starts at the current time but allows you to select any day/hour. Slide to the left of *Now* for historical usage; slide right for future predictions.



Related topics:

[Common Administration Portal Functions](#)

[Refresh Statistics](#)

[License Status](#)

[User Functions](#)

[Search/Sort Information in Tables](#)

[Show/ Hide Information in Charts](#)

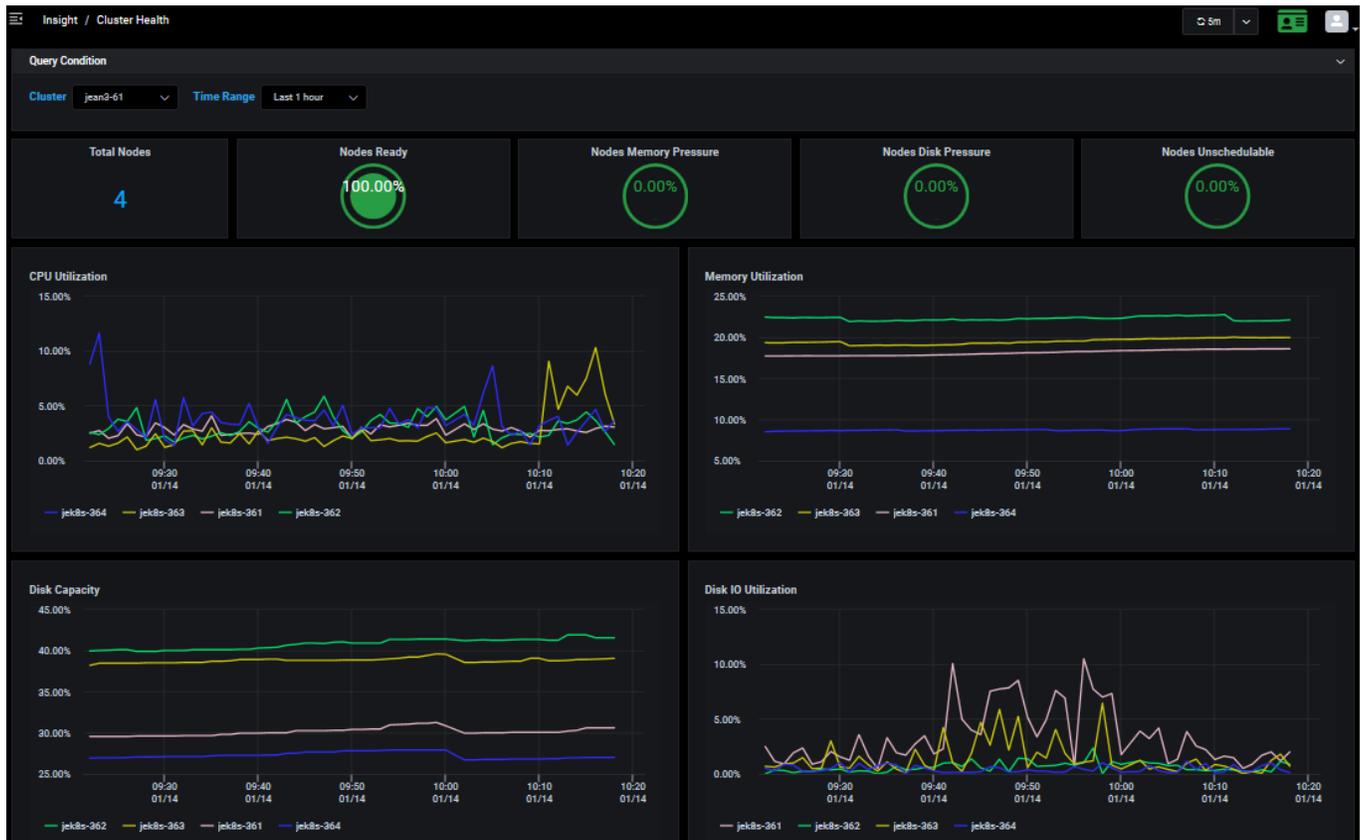
[Query Condition Panel](#)

Insight - Cluster Health

The *Cluster Health* page displays actual usage observations about the nodes in a cluster. Usage for the last 1, 2, 4, 6, or 12 hours can be displayed and can further be filtered by selecting a range of dates.

For the selected cluster, you will see the total number of nodes, the percentage of nodes that are ready, the percentage of nodes that are under memory or disk pressure, as well as the percentage of nodes that are not schedulable. Memory pressure and disk pressure are defined by Kubernetes. Refer to <https://kubernetes.io/docs/tasks/administer-cluster/out-of-resource/#node-conditions> for more information.

CPU utilization, memory utilization, disk capacity, and disk IO utilization charts are displayed for each node.



Related topics:

[Search/Sort Information in Tables](#)

[Terminology](#)

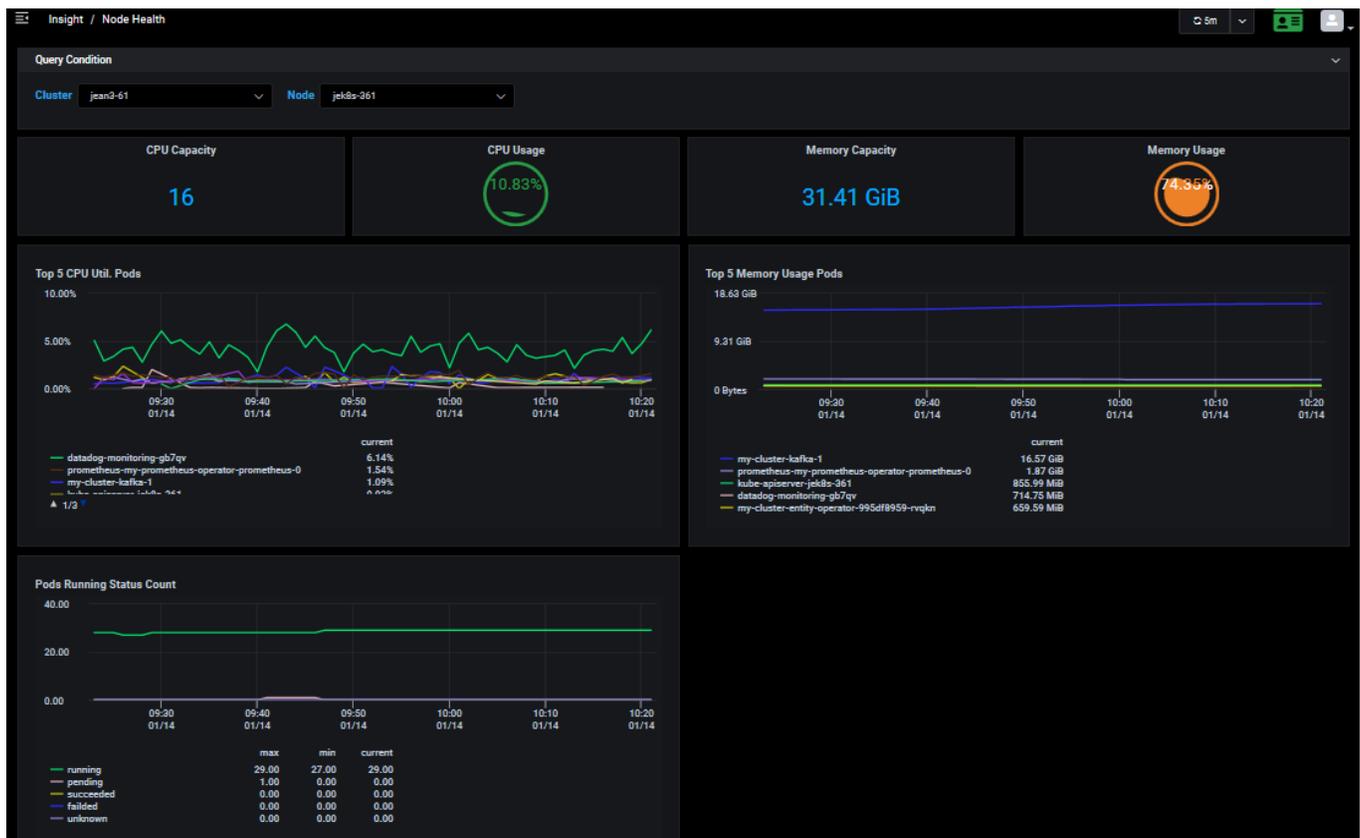
Insight - Node Health

The *Node Health* page displays actual usage observations about each node in a cluster. Select which cluster and node to display.

For the selected node, you will see the total CPU capacity and usage as well as memory capacity and usage.

The *Top 5* charts below display the CPU utilization of the top five pods on each node and the memory usage of the top five pods.

The *Pods Running Status Count* chart displays the number of pods running, the minimum and maximum number of pods that can run, along with the status of each pod.



Related topics:

[Search/Sort Information in Tables](#)

[Show/Hide Information in Charts](#)

[Terminology](#)

Planning - Workload Prediction

The *Workload Prediction* page displays actual CPU and memory usage observations, predicted usage, and recommendations for selected clusters, nodes, namespaces, applications, and controllers.

Federator.ai provides workload prediction and recommendations for resources at multiple levels of resources in a Kubernetes cluster.

For each monitored cluster, Federator.ai monitors resource usage for the entire cluster, every node, and user-selected namespaces in the cluster and provides predictions and recommendation for each.

In addition, Federator.ai monitors user-defined applications and their controllers, providing predictions and recommendations at the controller level and at the application level.

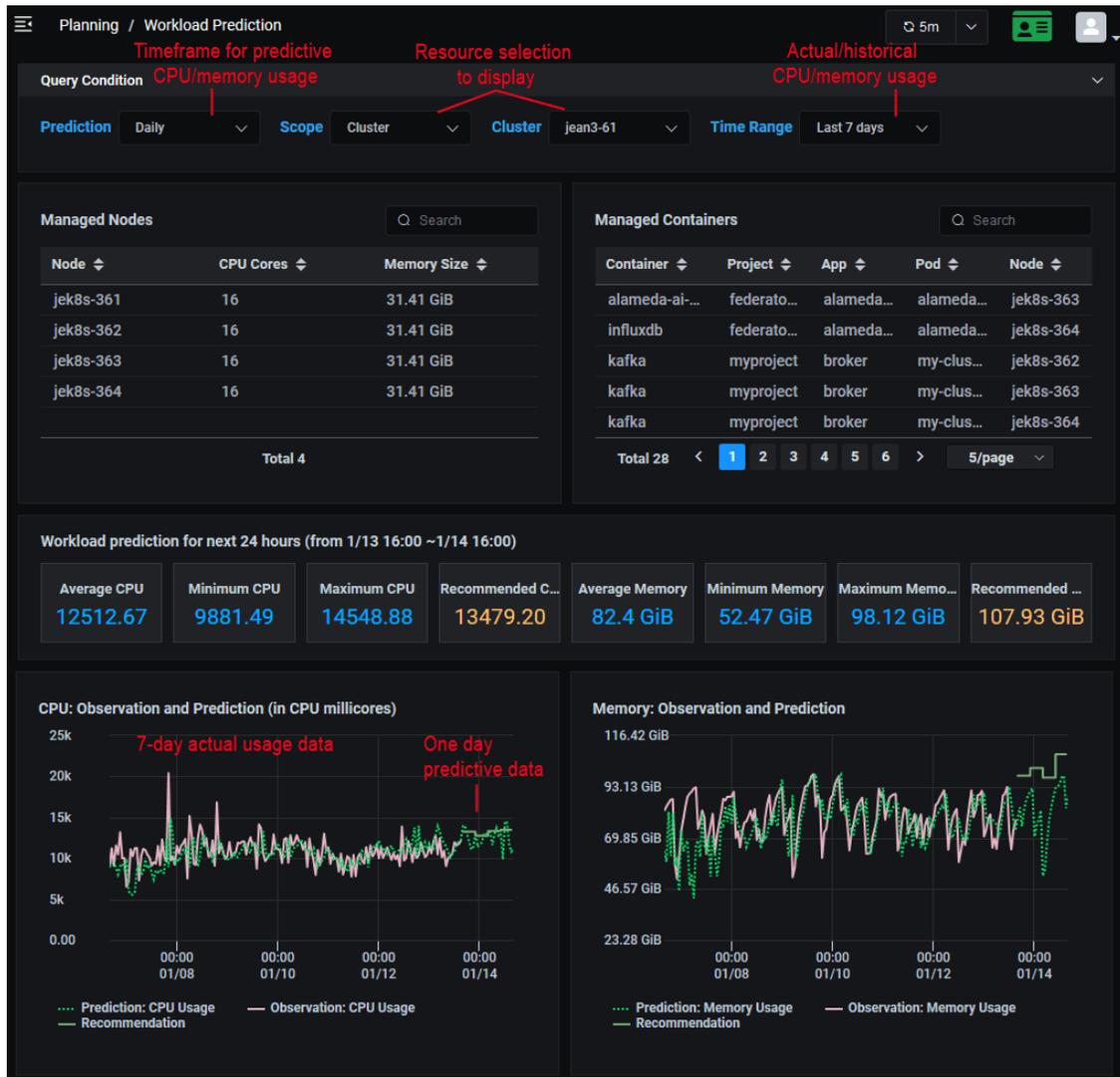
With the recommendations, you can decide if a controller, namespace, node, or cluster is over-provisioned (wasting resources), or if it is under-provisioned and will not sustain an increased workload.

In the *Query Condition* panel, select the level of information you want to display, cluster, node, namespace, application, or controller, as well as the timeframe. When you select a namespace, the namespace status will be displayed.

The *Workload Prediction* charts display daily, weekly, and monthly predictive CPU and memory data:

- Daily – Predicts CPU and memory usage every hour for the next 24 hours.
- Weekly – Predicts CPU and memory usage every 6 hours for the next 7 days.
- Monthly – Predicts CPU and memory usage every day for the next 30 days.

Actual CPU and memory usage can be displayed for a historical perspective. Use the *Time Range* field to determine how far back to display.



Managed Nodes Table

This table shows the list of nodes in the selected cluster with the number of CPU cores and the size of memory for each node.

Managed Containers Table

Based on the selected scope (cluster, node, namespace, application, or controller), this table lists all of the containers for the selected scope. Each container is listed along with its project (namespace), application, Kubernetes pod name, and the node where this container runs.

Workload Prediction Table

This table displays average/minimum/maximum CPU and memory usage and recommendations for the upcoming time selected - 24 hours (daily), 7 days (weekly), 30 days (monthly).

If you are displaying information for a namespace and the status is anything but *Monitoring*, this section will provide more information. For example, you will see a message for newly added namespaces that do not have enough information for predictions.

CPU and Memory Charts

These charts display CPU and memory observations and predictions for all resources specified in the *Query Condition* panel.

- The solid line represents the observed actual usage.
- The dotted green line represents the past and future predicted usage.
- The solid green line represents the recommended usage, which can help you from over-provisioning resources.
- If the selected scope is *Node*, the red line represents the node's total CPU and memory. A big difference between total resources and actual and predicted usage can indicate that you are over-provisioned. A small difference between total resources and actual and predicted usage can indicate that you might be under-provisioned.

Related topics:

[Terminology](#)

[Search/Sort Information in Tables](#)

[Show/Hide Information in Charts](#)

[Cluster Configuration](#)

Recommendation - HPA Recommendation

The *HPA Recommendation* page displays usage and recommendation information about replicas for selected controllers. These controllers must be enabled with autoscaling during configuration. Refer to the Configuration section for information about how to enable autoscaling for a controller. When autoscaling is enabled, CPU and memory usage is monitored, and the number of pods is increased/decreased based on the workload. An auto-scaled pod is called a *replica*.

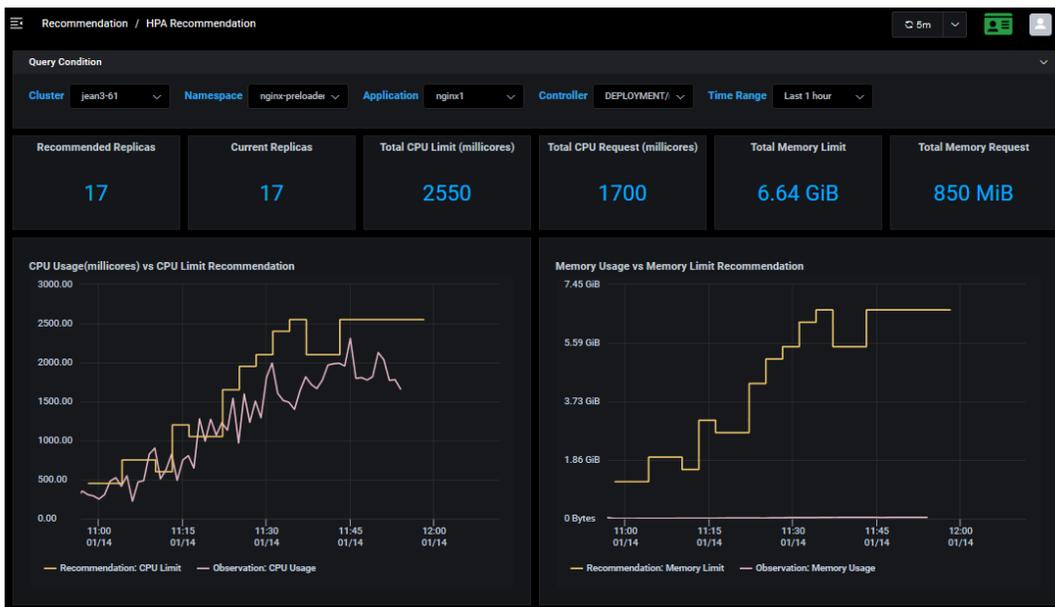
Use the correct namespace, application, and controller name to see the history of autoscaling for a specific controller. The historical number of replicas and CPU/memory usage for the last 1, 2, 4, 6, or 12 hours can be displayed and can further be filtered by selecting a range of dates.

When the system was configured, the requested (minimum) CPU/memory and the limit (maximum) CPU/memory were set. This page shows the number of recommended and current replicas along with the total CPU/memory limit/request being used by all current replicas.

CPU and memory Charts

These charts display CPU and memory usage and recommendations.

- The solid pink/white line represents the observed actual usage.
- The solid yellow line represents the limits after autoscaling.



Related topics:

[Common Administration Portal Functions](#)

[Terminology](#)

[Configuration](#)

[Search/Sort Information in Tables](#)

[Show/Hide Information in Charts](#)

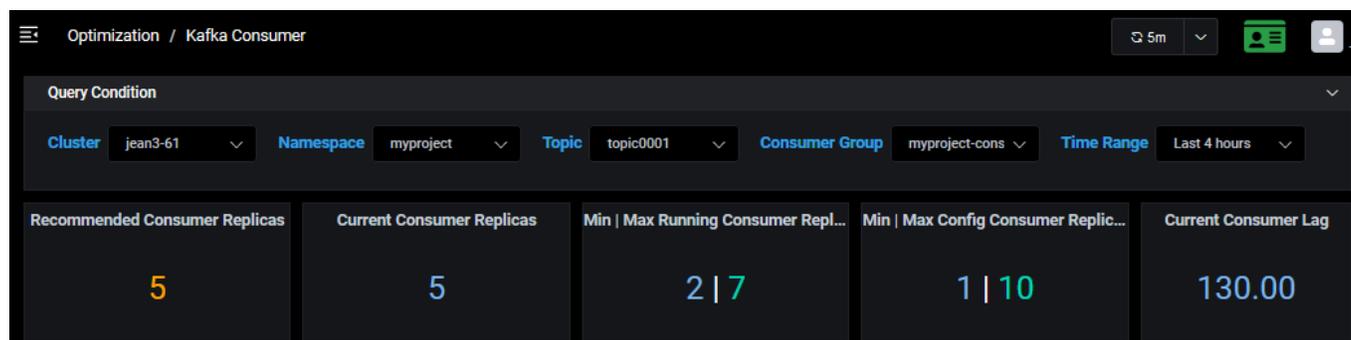
[Configure Applications](#)

Optimization – Kafka Consumer

The traditional method of autoscaling, based on Kafka consumer CPU/memory usage, is not enough to achieve desired performance goals. Kafka production and consumption of messages (including message arrival rate and consumer lag) offers a better indicator of workload and performance, and is monitored by Federator.ai. Using message production rate predictions, Federator.ai auto-scales the number of Kafka consumer pods to fit the workload and optimize performance.

If you are using Kafka software and you have configured Federator.ai to monitor and auto-scale Kafka consumers, the *Kafka* page displays predictions for the message production rate and scales Kafka consumer replicas to satisfy the workload. You can configure multiple Kafka topics and consumer groups. Federator.ai will predict and auto-scale consumers for each individual topic/consumer group. Refer to the [Add an Application](#) section for information about how to configure Kafka consumer monitoring and auto-scaling.

In the *Query Condition* panel, select the correct cluster, namespace, topic, and consumer group for the Kafka consumer being monitored. The number of replicas and Kafka message production/consumption rate and the prediction of message production rate for the last 1, 2, 4, 6, or 12 hours can be displayed and can further be filtered by selecting a range of dates.



The number of recommended and current consumer replicas is displayed along with the minimum and maximum number of running and configured replicas and the current consumer lag.

Number of Replicas Chart

The *Number of Replicas* chart displays the Kafka consumer replicas for the specified time range as a result of recommendations from Federator.ai.

Production Rate and Consumption Rate Chart

The *Production Rate and Consumption Rate* chart displays the actual observed message production rate (solid green line) and consumption/processed rate (solid blue line) for the specified time range and the historical and future predicted rate (dotted green line).

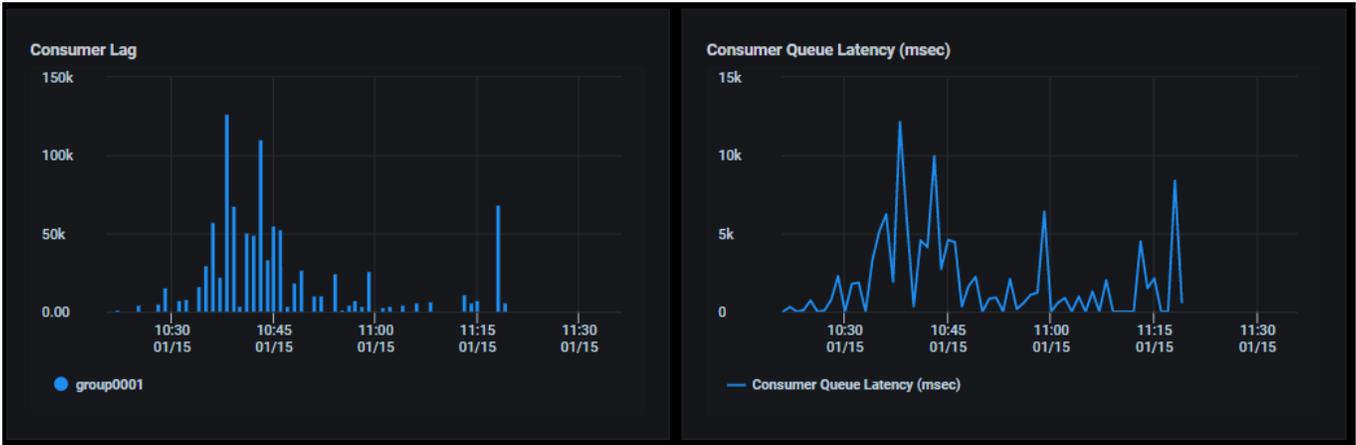


Consumer Lag Chart

The *Consumer Lag* chart displays the number of messages in the Kafka brokers that are yet to be processed by the Kafka consumers for the selected topic and consumer group and time range. This represents messages in the queue waiting to be processed.

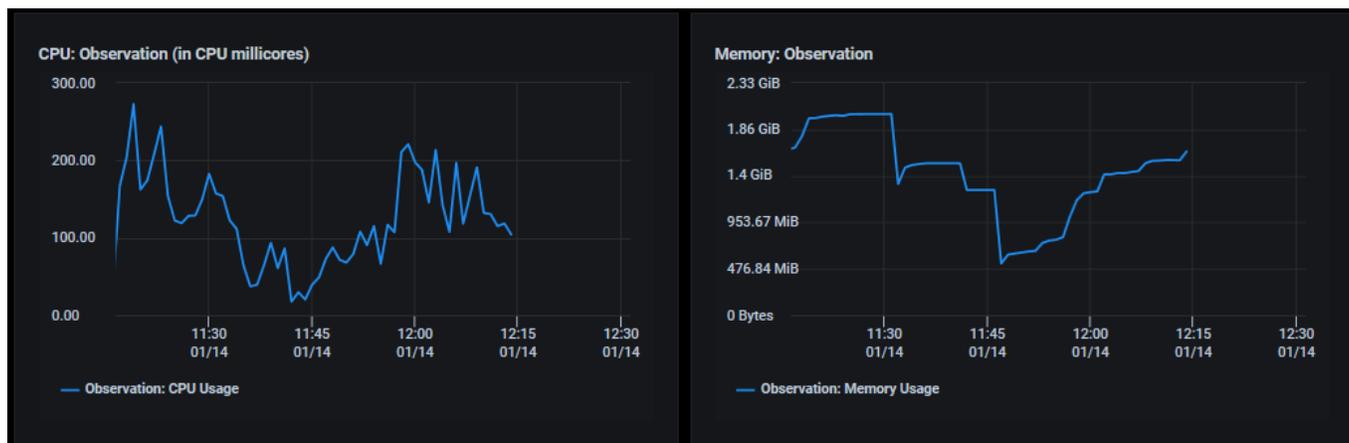
Consumer Queue Latency Chart

The *Consumer Queue Latency* chart displays how long it takes for each message to be processed for the selected time range.



CPU and Memory Observation Charts

The *CPU Observation* and *Memory Observation* charts display observed actual usage for the selected time range. This reflects the resources used as a result of autoscaling consumer pods.



Related topics:

[Common Administration Portal Functions](#)

[Terminology](#)

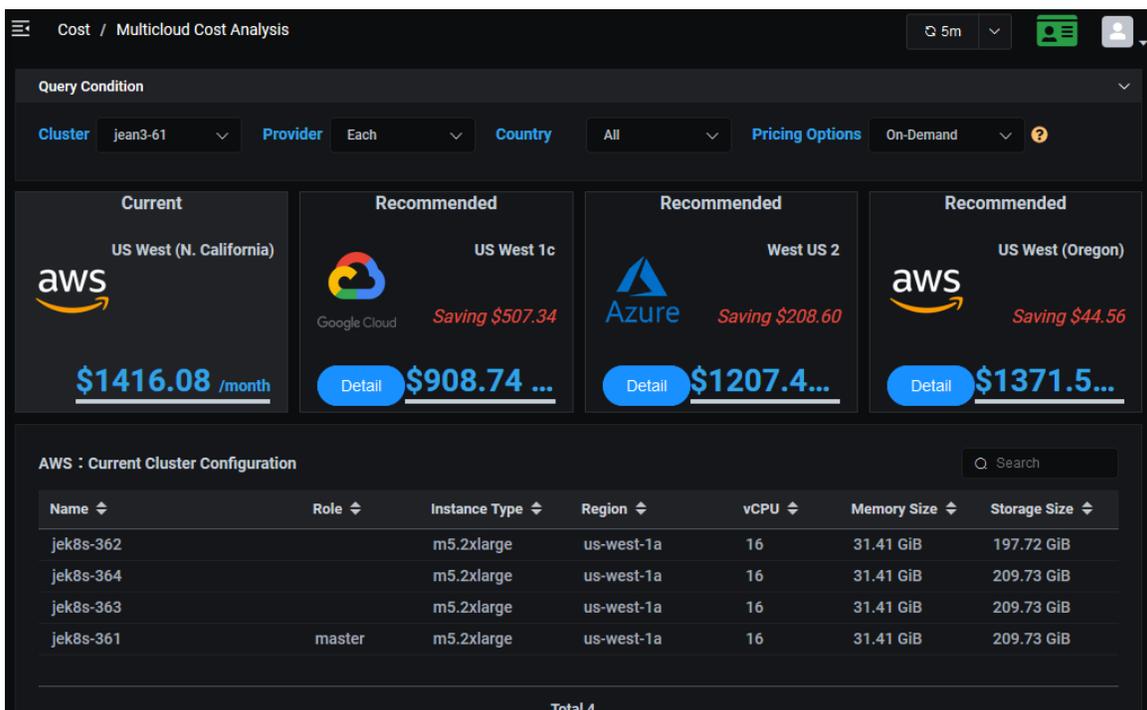
[Show/Hide Information in Charts](#)

[Setup Wizard](#)

[Configure Applications](#)

Cost – Multicloud Cost Analysis

The *Multicloud Cost Analysis* page displays the amount you are spending with your current cloud provider (if applicable) and outlines your estimated cost for doing business with a variety of cloud providers based on fees charged by each provider and recommended changes to your current resource configuration. Federator.ai monitors and analyzes the overall CPU/memory usage of a cluster and predicts the usage for the next 30 days. Based on the past usage and the forecast, Federator.ai recommends the right instance types of cluster nodes for the workload with the lowest cost from each cloud provider.



You can filter the data that appears on the whole page by selecting:

- Cluster
- Provider – Display data for the:
 - Three lowest costs among all providers.
 - Lowest cost from each provider.
 - Three lowest costs from a selected provider.
- Country – All or a specific country.
- Pricing option - The following pricing options are available:
 - Reserved - Stationary workloads will be served with Reserved instances and temporary, increased workloads will be served with On-Demand instances.
 - On-Demand - All workloads will be served with On-Demand instances.
 - Spot - Evictable workloads will be served with Spot instances while the other workloads will be served with On-Demand instances.
 - Spot + Reserved - Evictable workloads will be served with Spot instances while the other workloads will be served with Reserved and On-Demand instances.

The *Current* box displays what you are currently paying to your cloud provider with your existing resource configuration. The current configuration for each cluster node, including role (master, worker), instance types being used, region, number of CPUs, memory size, and storage size is displayed in the table below.

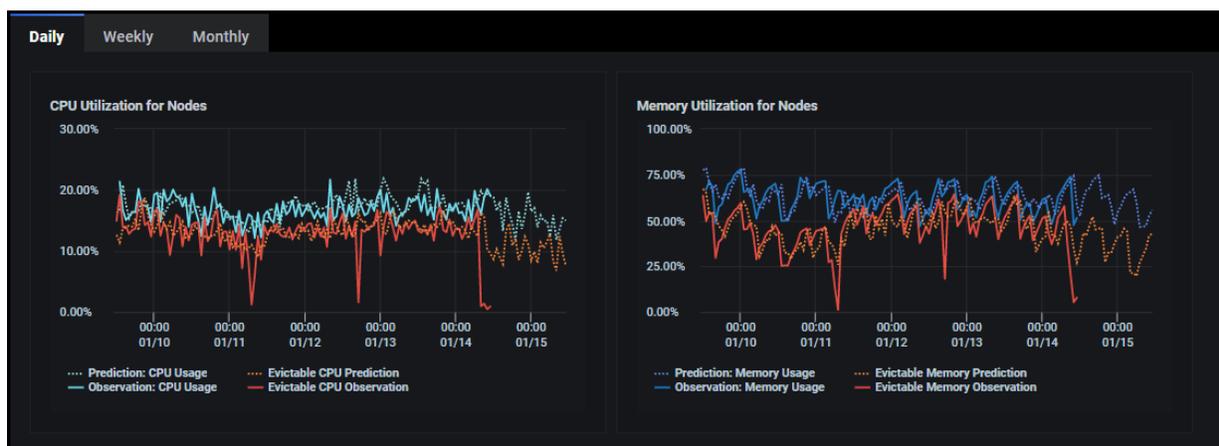
The *Recommended* boxes display potential savings with a different provider or different region for your current provider. Click the *Detail* button to see how these savings are calculated. The savings typically come from lower provider fees, as well as from reducing idle resources and using more cost-effective instance types with the provider. The information continually refreshes itself as new data becomes available.

While it is difficult to move between cloud providers, this information can help you reduce costs by changing instance types, reducing idle resources, and selecting a different region from your current provider.

Even if you are not currently using a cloud provider, this information can show you what your options are if you move to the cloud.

CPU and Memory Utilization Charts

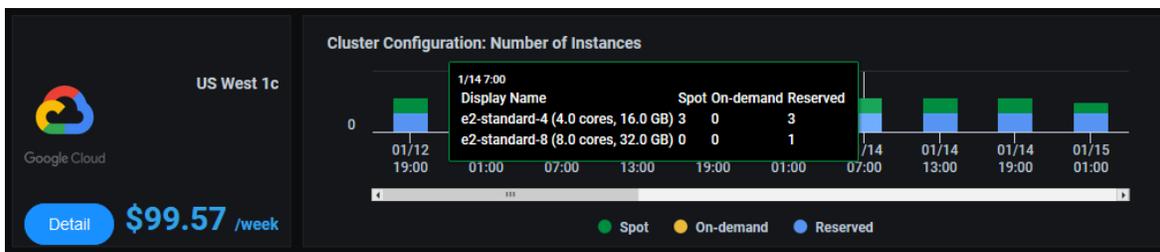
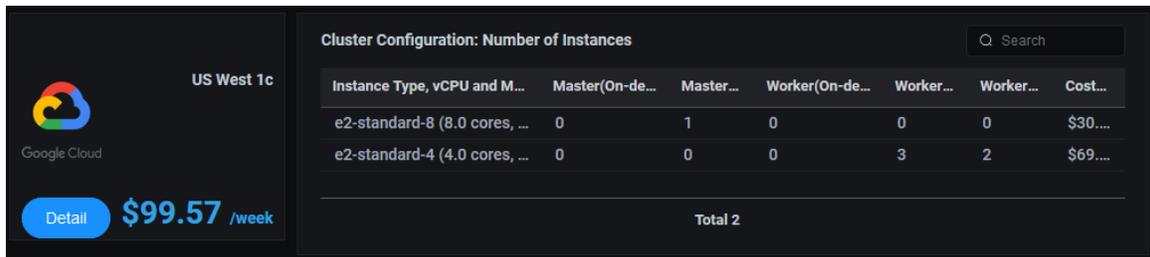
The *CPU Utilization* and *Memory Utilization* charts display utilization information for all nodes in the selected cluster based on your daily, weekly, or monthly workload.



The solid lines represent the observed actual usage while the dotted lines show the historical and future predicted usage. The predicted usage is used in the analysis for cluster configuration recommendations.

Recommended Cluster Configuration Tables

The recommended configuration with each provider is displayed based on your daily, weekly, or monthly workload prediction (e.g., if *Daily* is selected for the *CPU Utilization* and *Memory Utilization* charts, the recommendation and saving are calculated based on the workload prediction for the next 24 hours). This includes the recommended instance type, CPU, and memory as well as the number of master and worker servers with each pricing model.



Information can be displayed in a table format or as a *Time Series* chart, which shows the recommended instances (number and type) for future workloads at specific times and can help you determine how many resources to reserve.

For daily predictions, data is displayed in the charts for each hour. For weekly predictions, data is displayed for every six hours. For monthly predictions, data is displayed for each day.

The recommendations for each data point include the number of on-demand, reserved, or spot instances, depending on the pricing option selected.

Place your cursor over one of the bars and you can see the recommended number of instances. In the above example, 3 spot and 3 reserved are recommended for the first instance type, while 1 reserved is recommended for the second instance type.

Related topics:

[Common Administration Portal Functions](#)

[Terminology](#)

[Search/Sort Information in Tables](#)

[Show/Hide Information in Charts](#)

[Cloud Provider Price Books](#)

Cost - Cost Allocation

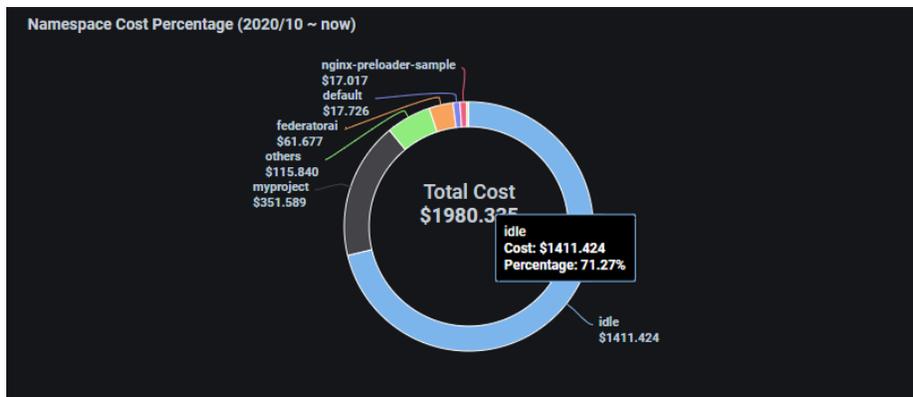
The *Cost Allocation* page provides an in-depth cost analysis for individual namespaces of a cluster hosted by the current cloud service provider. With the information displayed on this page, you can compare the cost of each namespace and determine the estimated cost of each namespace based on the workload prediction. You can also see the percentage of cost of the system spent while the system is idle. A high percentage of cost while the system is idle indicates that the cluster is over-provisioned.

Current Cluster Configuration

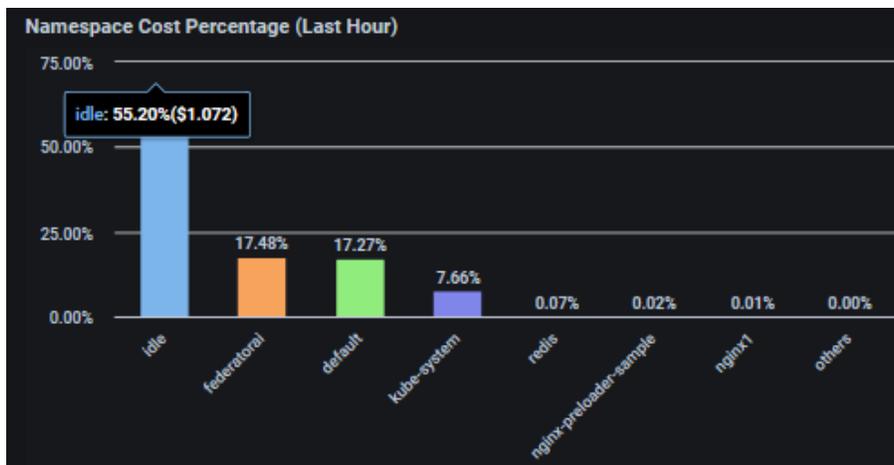
The existing configuration with your provider is displayed. This includes the role (master, worker), instance types being used, region, number of CPUs, memory size, and storage space.

Namespace Charts

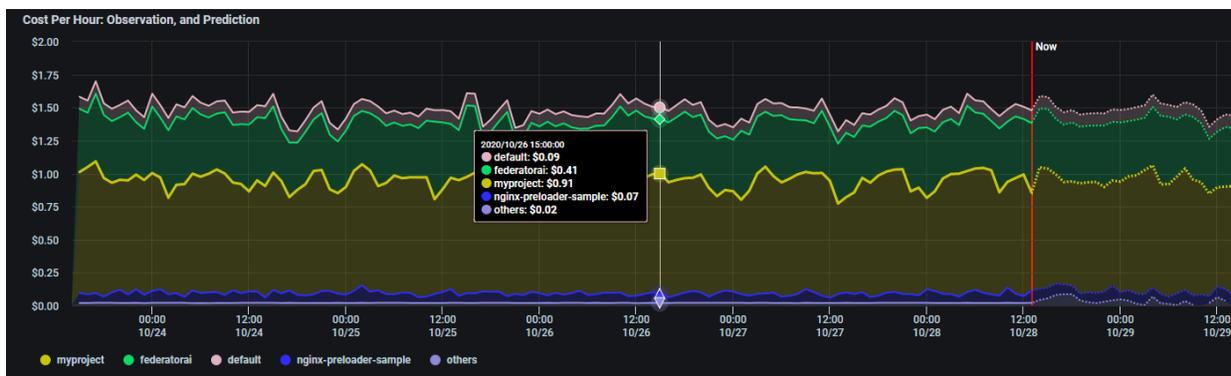
The *Namespace Cost Percentage (month)* chart displays this month's actual percentage of cost to-date for each namespace and when the system is idle. The *others* tag represents the unmonitored namespaces in a cluster. Move your cursor over each section to show the percentage of total cost for a namespace or when the system is idle.



The *Namespace Cost Percentage (Last Hour)* chart displays the actual cost for each namespace and when system was idle during the past hour. The *others* tag represents the unmonitored namespaces in a cluster. Move your cursor over each bar to show the cost of a namespace or when system is idle.



The *Cost Per Hour* chart displays hourly costs for each namespace during the past five days and the predicted hourly cost for the next 24 hours. The solid lines represent the observed actual cost while the dotted line shows the future predicted cost based on the namespace resource usage prediction. The *others* tag represents the unmonitored namespaces in a cluster. Move your cursor to a specific date/hour to see the detailed cost of each namespace.



The *Cumulative Cost Projection* chart displays a cumulative cost from the beginning of the month and a cost projection for this month for each namespace. The solid lines represent the observed actual cost while the dotted lines show the future predicted cost based on the namespace resource usage prediction. The *others* tag represents the unmonitored namespaces in a cluster. Click on the calendar icon to choose the cost analysis for a specific month.



Related topics:

[Common Administration Portal Functions](#)

[Terminology](#)

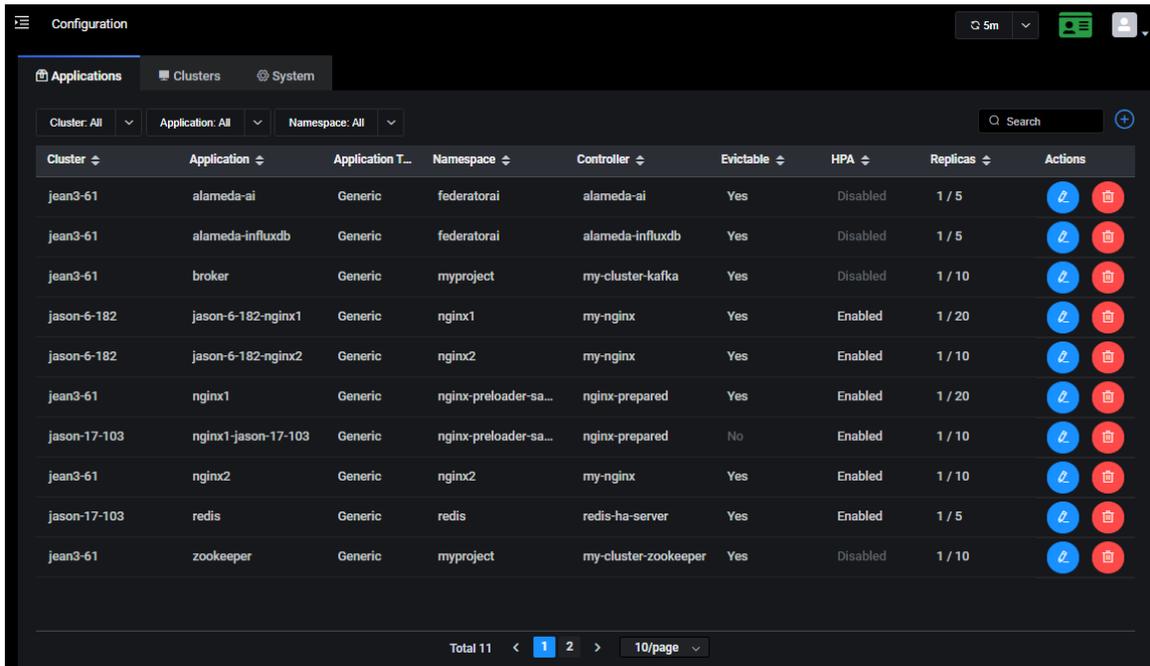
[Search/Sort Information in Tables](#)

[Show/Hide Information in Charts](#)

[Cloud Provider Price Books](#)

Configuration - Applications

The *Applications* page displays all applications being monitored by Federator.ai and allows you to manage applications.

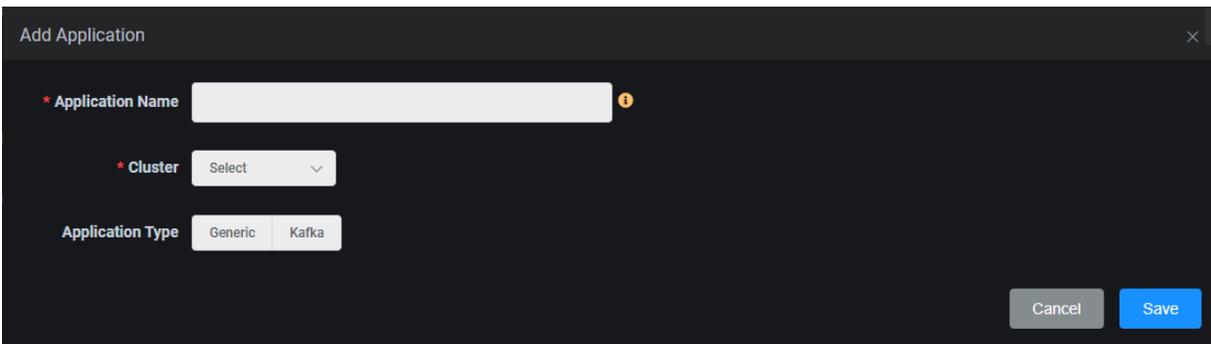


You can perform the following functions:

Icon	Function
	Add an application.
	Edit an application.
	Remove an application.

Add an Application

1. On the *Configuration* page, click the + (Add Application) icon.



The screenshot shows a dark-themed dialog box titled "Add Application" with a close button (X) in the top right corner. It contains three main input fields: a text field for "* Application Name" with an information icon (i) to its right; a dropdown menu for "* Cluster" with "Select" as the current selection; and a radio button group for "Application Type" with "Generic" selected and "Kafka" as an alternative. At the bottom right, there are "Cancel" and "Save" buttons.

Application Name - Name of the application you want to manage. The name must be a maximum of 253 characters, contain only lowercase alphanumeric characters, "-", or ".", and start and end with an alphanumeric character. Note, an application is not a native Kubernetes object. You will define the controllers that are part of your application later.

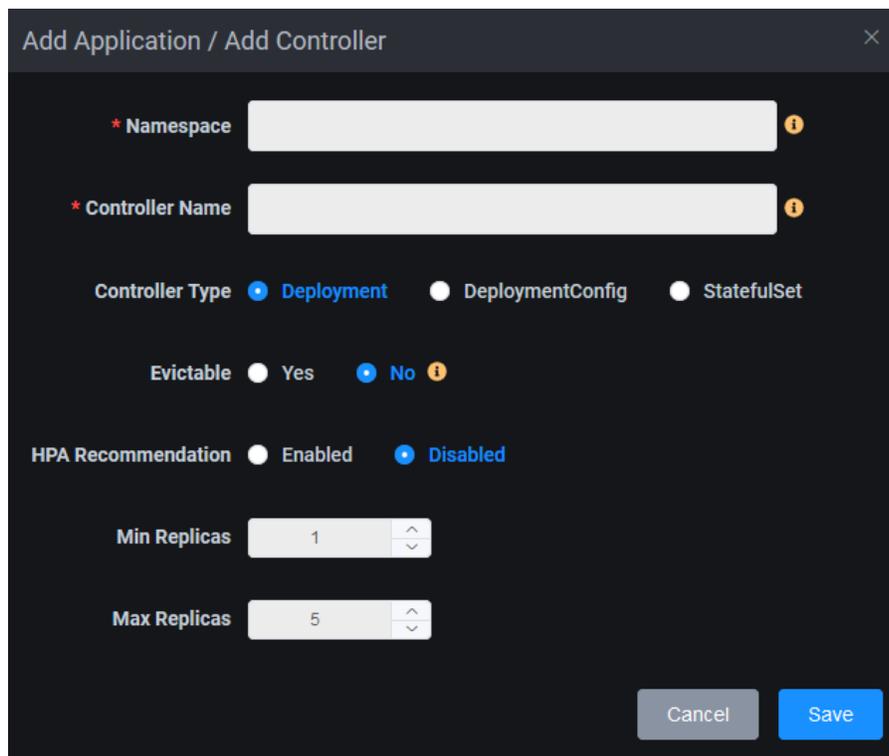
Cluster – Select an existing cluster where this application resides.

Application Type - For a *Generic* application, CPU, memory, and usage are monitored. For scaling Kafka consumers, choose *Kafka* as the application type

2. Click the + icon to add controllers (Generic) or consumer groups (Kafka).

The options will vary, depending upon the selected *Application Type*.

Generic:



The screenshot shows a dark-themed dialog box titled "Add Application / Add Controller" with a close button (X) in the top right corner. It contains several input fields and radio button groups: a text field for "* Namespace" with an information icon (i); a text field for "* Controller Name" with an information icon (i); a radio button group for "Controller Type" with "Deployment" selected, and "DeploymentConfig" and "StatefulSet" as alternatives; a radio button group for "Evictable" with "No" selected and "Yes" as an alternative; a radio button group for "HPA Recommendation" with "Disabled" selected and "Enabled" as an alternative; a numeric input field for "Min Replicas" with the value "1"; and a numeric input field for "Max Replicas" with the value "5". At the bottom right, there are "Cancel" and "Save" buttons.

- *Namespace* – The Kubernetes namespace where the controller is deployed.
- *Controller Name* - The name of controller to be monitored.
- *Controller Type* – Supported controller types are *Deployment*, *StatefulSet*, and *DeploymentConfig* (OpenShift only).
- *Evictable* – Indicate if the controller can be interrupted if the node is shut down. Evictable controllers are good candidates to be deployed in Spot instances.
- *HPA Recommendation* – Indicate if you want to enable Horizontal Pod Autoscaling (HPA). When enabled, CPU and memory usage is monitored, and the number of pods is automatically increased/decreased based on the CPU/memory usage workload.
- *Min/Max Replicas* – Specify the minimum and maximum number of pods when HPA is enabled.

Kafka:

The screenshot shows a dark-themed dialog box titled "Add Application / Add Consumer Group". It contains the following fields and controls:

- * Kafka Broker Namespace: Text input field.
- * Consumer Group Namespace: Text input field.
- * Consumer Group Name: Text input field with an information icon.
- * Kafka Consumer ID: Text input field with an information icon.
- Controller Type: Radio buttons for Deployment (selected), DeploymentConfig, and StatefulSet.
- Evictable: Radio buttons for Yes and No (selected), with an information icon.
- * Topic Name: Text input field.
- Min Replicas: Spin box with value 1.
- Max Replicas: Spin box with value 5.
- Buttons: Cancel and Save.

- *Kafka Broker Namespace* – The namespace for the Kafka broker that receives and stores messages from producers and allows consumers to fetch the messages.
- *Consumer Group Namespace* – The namespace of the consumer group used for checking the offset on topics and processing messages.
- *Consumer Group Name* – The name of the Kafka consumer group used for checking the offset on topics and processing messages.

- *Kafka Consumer ID* – The unique ID of the Kafka consumer group for the set of consumers within the same consumer group.
- *Controller Type* – The controller type of the consumer group. Supported controller types are *Deployment*, *StatefulSet*, and *DeploymentConfig* (OpenShift only).
- *Evictable* – Indicate if the controller can be interrupted if the node is shut down. Evictable controllers are good candidates to be deployed in Spot instances.
- *Topic Name* – Indicate the topic (where messages are stored and published) that will be processed by consumers.
- *Min/Max Replicas* – Specify the minimum and maximum number of consumers to be auto-scaled.

3. Click *Save*.

4. Continue adding controllers/consumer groups and click *Save* when done.

Manage Applications

You can do the following from the *Applications* page:

- Add, edit, or remove a controller/consumer group. To do this, click the *Edit Application* icon and click the appropriate icon:

Icon	Function
	Add a controller/consumer group.
	Edit a controller/consumer group.
	Remove a controller/consumer group.

- Remove an application. To do this, click the *Remove Application* icon and confirm the removal.

Related topics:

[Terminology](#)

[Search/Sort Information in Tables](#)

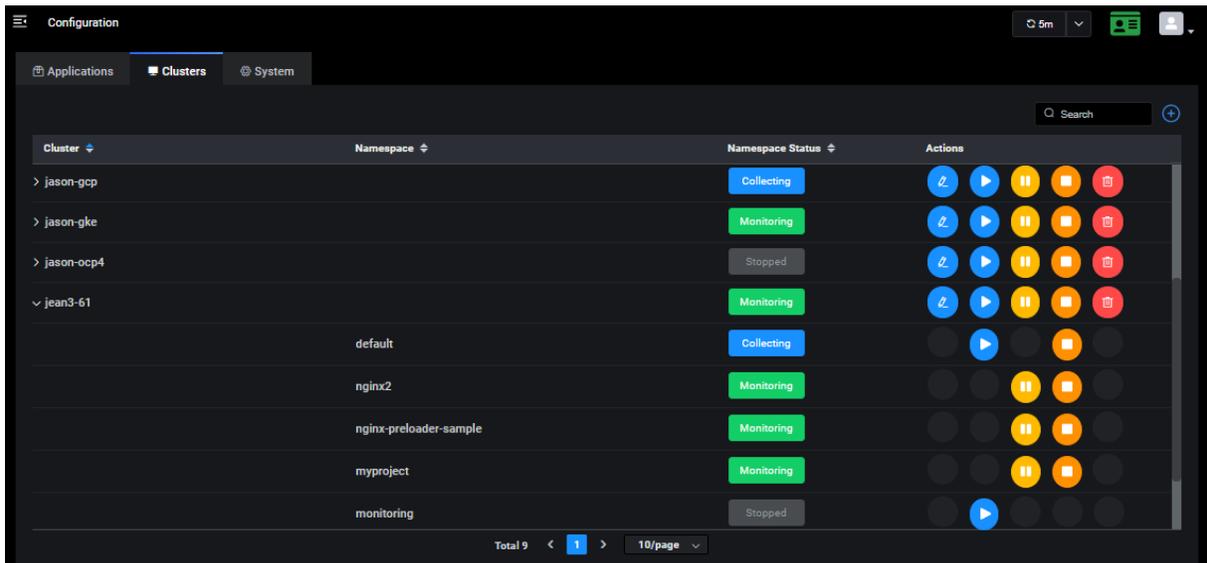
[Configure Clusters](#)

Configuration - Clusters

The *Clusters* page displays all clusters being monitored by Federator.ai. Expand a cluster to see the namespaces for a cluster.

You can also see namespace status for the cluster and individual namespaces. The status will be *Monitoring* when the system is collecting metrics and providing workload predictions. The status will be *Collecting* when all the namespaces in the cluster are in collecting state. When a namespace is in collecting state, the system will still collect metrics, but prediction tasks are paused. A namespace will also collect data if it was added into a cluster that does not automatically monitor for new namespaces.

A status of *Stopped* means no there is no collecting of metrics and no predictions. A cluster is *Stopped* if all of its namespaces are stopped.

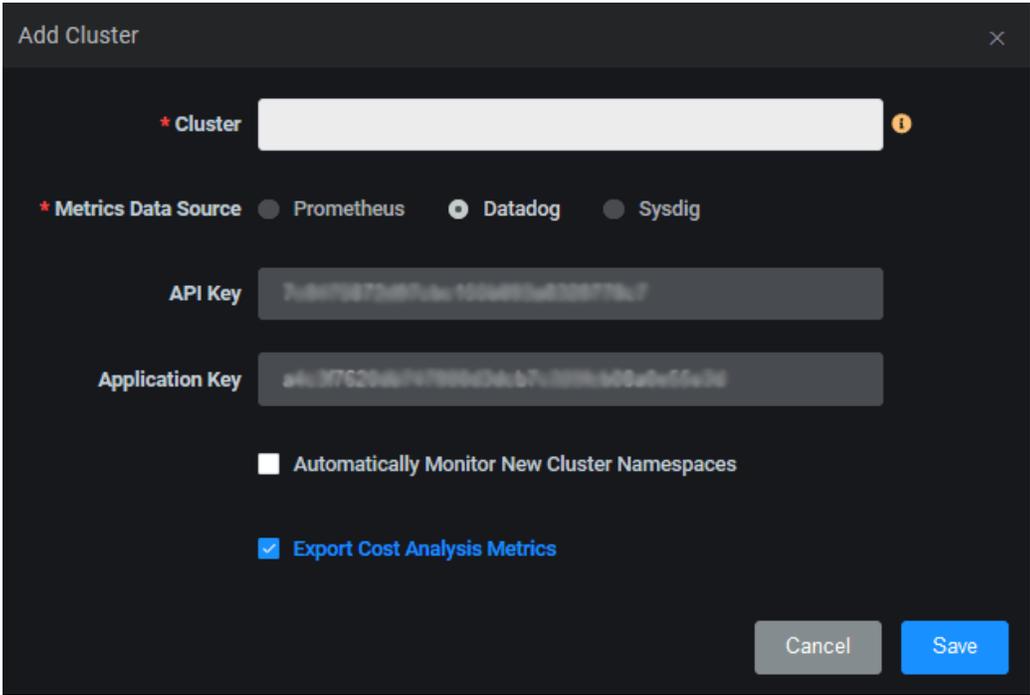


Functions are different for clusters and namespaces. You can perform the following functions:

Icon	Function
	Add a cluster.
	Edit a cluster.
	Start monitoring and prediction for all namespaces or a specific namespace.
	Pause monitoring and prediction for all namespaces or a specific namespace.
	Stop collecting metrics and making predictions for all namespaces or a specific namespace.
	Remove a cluster.

Add a Cluster

1. On the *Configuration* page, click the + (Add Cluster) icon.



Cluster – Specify a Kubernetes cluster with one or more nodes.

Metrics Data Source – Specify the source of metrics for this cluster. If you created a cluster during the setup wizard, this source will be selected for you and cannot be changed.

API Key/Application Key – For Datadog, the API key and application key are required for authentication. They can only be set/changed from the *Metrics Data Source* tab under *Configuration* → *System*.

URL/Token – For Sysdig, a URL and token are required for authentication. They can only be set/changed from the *Metrics Data Source* tab under *Configuration* → *System*. For the Prometheus open-source monitoring system, the URL is required but the token is optional.

Automatically Monitor New Cluster Namespaces – Specify if you want to automatically monitor new namespaces in this cluster. If unchecked (default), you must manually specify which namespaces to monitor.

Export Cost Analysis Metrics – Specify if you want cost analysis metrics to be automatically exported to Datadog, so that the information can appear in their user interface.

2. Click *Save*.

Manage Clusters

You can do the following from the *Clusters* page:

- Edit the settings for monitoring new namespaces and exporting cost analysis metrics. To do this, click the *Edit Cluster* icon.
- Start monitoring and prediction for all namespaces in the cluster or a specific namespace. To do this, click the *Start Monitoring and Predictions* icon.
- Pause monitoring and prediction for all namespaces or a specific namespace. To do this, click the *Pause Monitoring and Predictions* icon for a cluster or for a namespace.
- Stop collecting metrics and making predictions for all namespaces or a specific namespace. To do this, click the *Stop Collecting Metrics and Predictions* icon for a cluster or for a namespace.
- Remove a cluster that does not have any applications configured. To do this, click the *Remove Cluster* icon.

Related topic:

[Terminology](#)

[Search/Sort Information in Tables](#)

[Configure Applications](#)

Configuration - System

The *System* page has tabs that allow you to:

- Change the admin password.
- Change the authentication needed to access metrics from the data source.
- Manage the system license.
- Set the policy/update cloud provider price books.

Admin Password

The *Admin Password* page allows you to change the admin password. To do this, select *Configuration, System, Admin Password* tab.

You must know the current password and *New Password* must match *Confirm Password*.

Metrics Data Source

The *Metrics Data Source* page allows you to set the authentication needed to access metrics from the data source (e.g., Prometheus, Datadog, Sysdig). You will need to do this if you did not specify a data source during initial configuration. You will also need to change existing information if your authentication changes. To set/change authentication, select *Configuration, System, Metrics Data Source* tab.

For Datadog, the *API Key* and *Application Key* are required for authentication.

For Sysdig, a *URL* and *Token* are required for authentication.

For the Prometheus open-source monitoring system, the *URL* is required but the *Token* is optional.

Licenses

The *Licenses* page displays your current system license, including license type, status, and expiration date. It also shows what is included in your license. To access this page, select *Configuration, System, Licenses* tab.

You can perform the following functions:

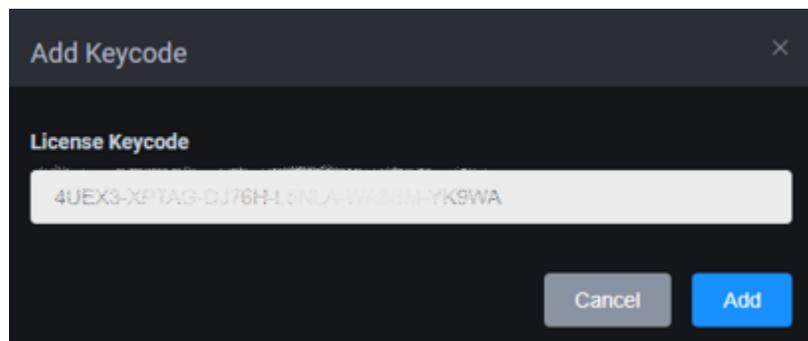
Icon	Function
	Add a license.
	Activate a license.
	Show license keycode.
	Remove a license.

Add a License

There is one license per system; you must delete the existing license before adding a new one.

Follow the steps below to add and register a license:

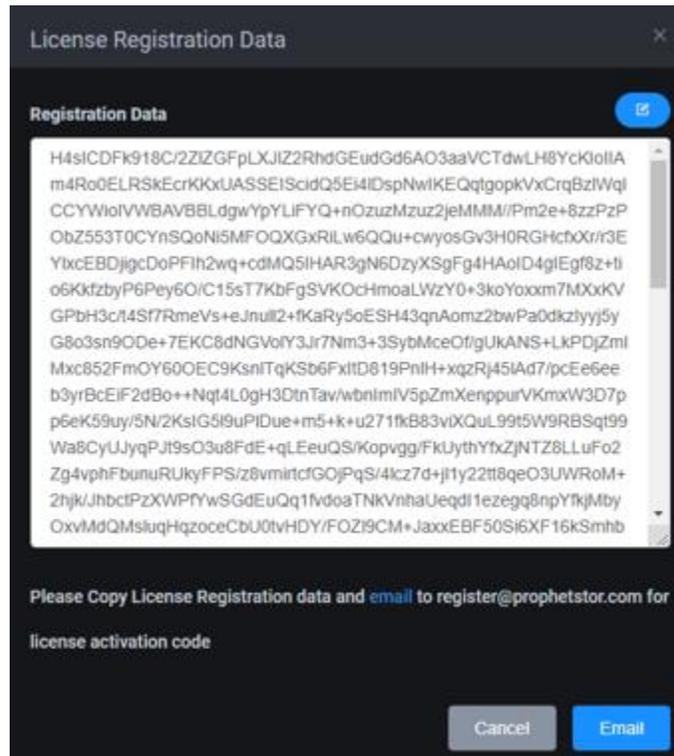
1. Click the *Add License* icon and enter the keycode.



2. Click *Add*.

A trial license is valid immediately after it is added. A regular license needs to be registered in order to be activated.

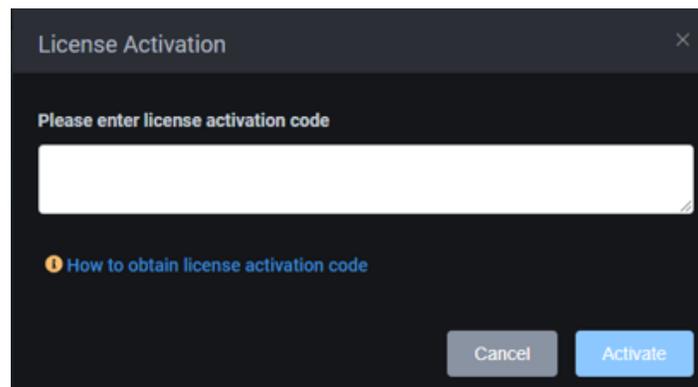
3. Email the registration data to register@prophetstor.com for a license activation code.



Click the *Email* button to launch your email program. Click the *Copy* icon to copy the registration text and paste it into your email before sending.

The license status will now say *Pending Activation*. It must be activated within 30 days.

- Once you have received the license activation code, click the *Activate License* icon and paste the activation code.



- Click *Activate*.

Manage Licenses

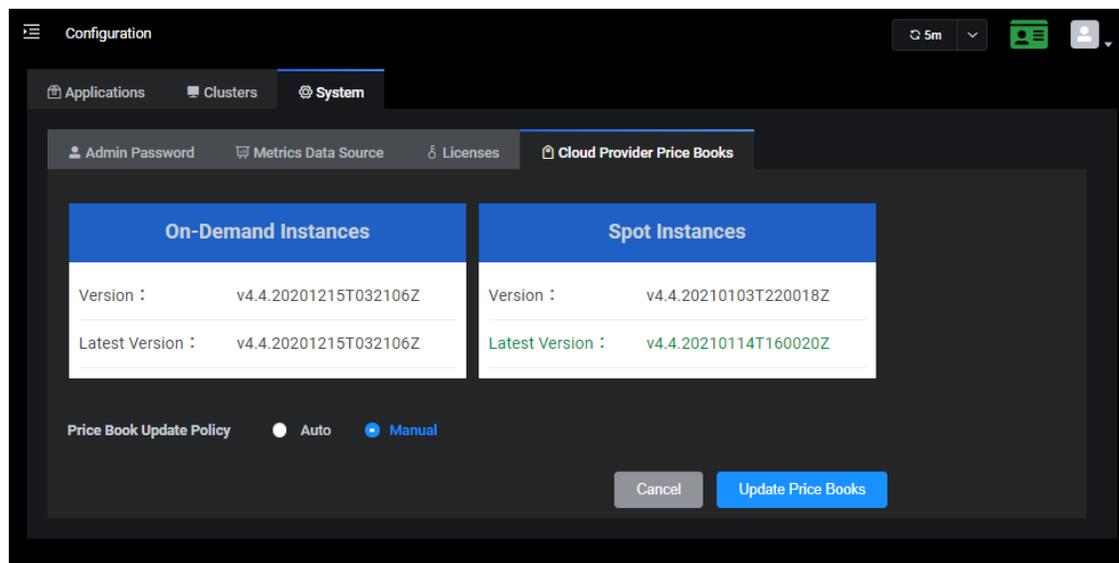
You can do the following from the *Licenses* page:

- View the keycode for your license. To do this, click the *Show License Keycode* icon.
- Remove the license. To do this, click the *Remove License* icon and confirm the removal.

Cloud Provider Price Books

The *Cloud Provider Price Books* page displays the versions of the on-demand and spot instance price books that have been collected from the following cloud providers: Amazon Web Services (AWS), Google Cloud, and Microsoft Azure.

To access this page, select *Configuration, System, Cloud Provider Price Books* tab.



The version numbers represent the ProphetStor version of the compiled price books.

The *Price Book Update Policy* allows you to specify how you want the price books updated. *Auto* regularly checks availability and automatically downloads the latest version, which requires Internet access.

If *Manual* is selected and the *Latest Version* is different from *Version*, Federator.ai is aware of a newer price book but it has not been applied. Click the *Update Price Books* button to update.

Note that the browser needs Internet access in order to display the *Latest Version*. If your computer does not have access, the data must be pushed to system.

Events

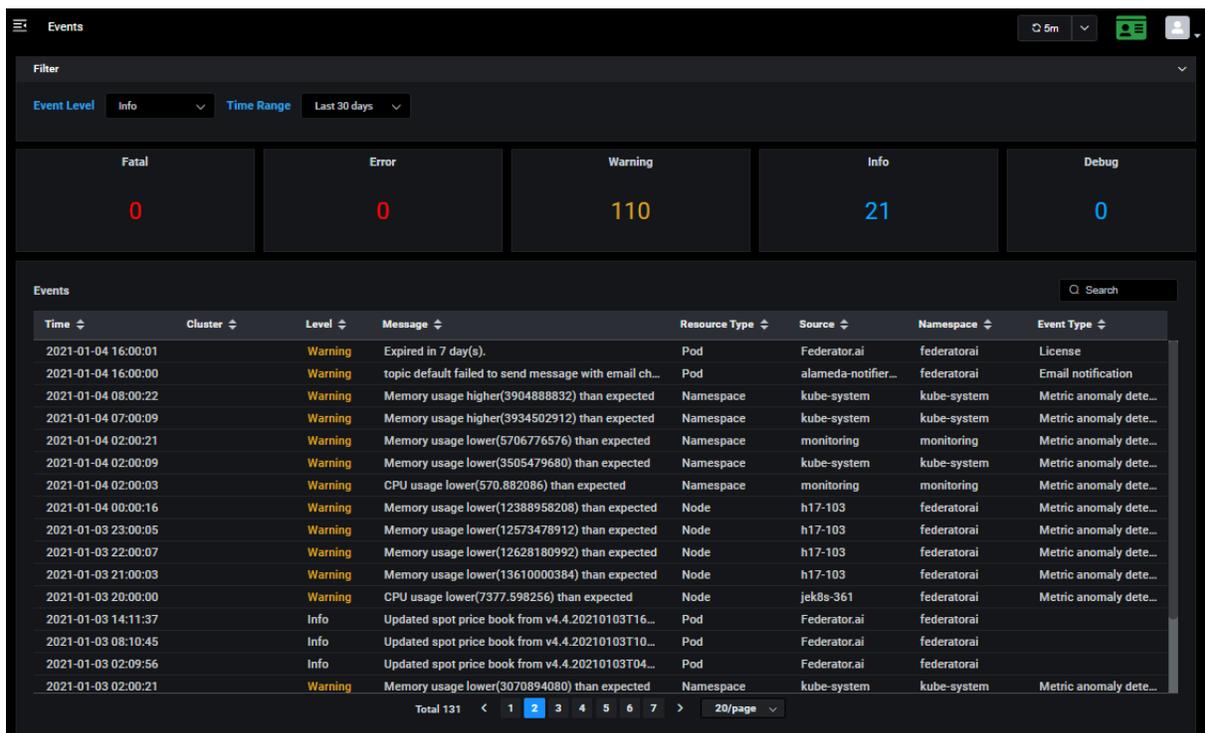
The *Events* page displays all system events that have occurred.

There are five levels of events:

- Fatal - Issues that may stop the system from operating properly.
- Error - Indicates that a failure has occurred.
- Warning - Indicates that something occurred that may require maintenance or corrective action; however, the system is still operational.
- Info - Day-to-day activities, which require no action.
- Debug - Detailed activities used for troubleshooting.

You can select the event level, which includes all levels above, and the time frame to display. For example, if you select the warning level, all warnings, errors, and fatal events will be displayed for the specified time period.

To specify a custom range, select *Custom* under *Time Range* and then specify a date range.



The screenshot shows the 'Events' page in a dark theme. At the top, there's a 'Filter' section with 'Event Level' set to 'Info' and 'Time Range' set to 'Last 30 days'. Below this, a summary bar shows counts for Fatal (0), Error (0), Warning (110), Info (21), and Debug (0). The main area is a table of events with columns for Time, Cluster, Level, Message, Resource Type, Source, Namespace, and Event Type. The table lists various events, including warnings about memory usage and CPU usage, and info events about updated spot price books. At the bottom, it shows 'Total 131' and a pagination control for 20 items per page.

Time	Cluster	Level	Message	Resource Type	Source	Namespace	Event Type
2021-01-04 16:00:01		Warning	Expired in 7 day(s).	Pod	Federator.ai	federatorai	License
2021-01-04 16:00:00		Warning	topic default failed to send message with email ch...	Pod	alameda-notifier...	federatorai	Email notification
2021-01-04 08:00:22		Warning	Memory usage higher(3904888832) than expected	Namespace	kube-system	kube-system	Metric anomaly dete...
2021-01-04 07:00:09		Warning	Memory usage higher(3934502912) than expected	Namespace	kube-system	kube-system	Metric anomaly dete...
2021-01-04 02:00:21		Warning	Memory usage lower(5706776576) than expected	Namespace	monitoring	monitoring	Metric anomaly dete...
2021-01-04 02:00:09		Warning	Memory usage lower(3505479680) than expected	Namespace	kube-system	kube-system	Metric anomaly dete...
2021-01-04 02:00:03		Warning	CPU usage lower(570.882086) than expected	Namespace	monitoring	monitoring	Metric anomaly dete...
2021-01-04 00:00:16		Warning	Memory usage lower(12388958208) than expected	Node	h17-103	federatorai	Metric anomaly dete...
2021-01-03 23:00:05		Warning	Memory usage lower(12573478912) than expected	Node	h17-103	federatorai	Metric anomaly dete...
2021-01-03 22:00:07		Warning	Memory usage lower(12628180992) than expected	Node	h17-103	federatorai	Metric anomaly dete...
2021-01-03 21:00:03		Warning	Memory usage lower(13610000384) than expected	Node	h17-103	federatorai	Metric anomaly dete...
2021-01-03 20:00:00		Warning	CPU usage lower(7377.598256) than expected	Node	jek8s-361	federatorai	Metric anomaly dete...
2021-01-03 14:11:37		Info	Updated spot price book from v4.4.20210103T16...	Pod	Federator.ai	federatorai	
2021-01-03 08:10:45		Info	Updated spot price book from v4.4.20210103T10...	Pod	Federator.ai	federatorai	
2021-01-03 02:09:56		Info	Updated spot price book from v4.4.20210103T04...	Pod	Federator.ai	federatorai	
2021-01-03 02:00:21		Warning	Memory usage lower(3070894080) than expected	Namespace	kube-system	kube-system	Metric anomaly dete...

At the bottom of the page, the total number messages being displayed is shown, along with navigation to other pages. You can also determine how many events to show per page.

Related topics:

[Terminology](#)

[Search/Sort Information in Tables](#)