r is a Kubernetes component. Although the AKS cluster uses a virtual machine
es, don't manually enable or edit settings for scale set autoscale in the Azure
zure CLI. Let the Kubernetes cluster autoscaler manage the required scale
formation, see **Can I modify the AKS resources in the node resource group?**

creates an AKS cluster with a single node pool backed by a virtual machine scale
cluster autoscaler on the node pool for the cluster and sets a minimum of *1* and

Copy Try It

```
resource group
-name myResourceGroup --location eastus

AKS cluster and enable the cluster autoscaler

p myResourceGroup \
ster \
\
irtualMachineScaleSets \
-sku standard \
r-autoscaler \
```

o create the cluster and configure the cluster autoscaler settings.

existing AKS cluster to enable the cluster

command to enable and configure the cluster autoscaler on the node pool for the
e `--enable-cluster-autoscaler` parameter, and specify a node `--min-count`

r is a Kubernetes component. Although the AKS cluster uses a virtual machine
es, don't manually enable or edit settings for scale set autoscale in the Azure
zure CLI. Let the Kubernetes cluster autoscaler manage the required scale
formation, see **Can I modify the AKS resources in the node resource group?**

updates an existing AKS cluster to enable the cluster autoscaler on the node pool
a minimum of *1* and maximum of *3* nodes:

Copy Try It

```
p myResourceGroup \
```

```
ster \
r-autoscaler \
```

o update the cluster and configure the cluster autoscaler settings.

cluster autoscaler settings

node pools in your AKS cluster, skip to the **autoscale with multiple agent pools**
multiple agent pools require use of the `az aks nodepool` command set to
ecific properties instead of `az aks`.

create an AKS cluster or update an existing node pool, the cluster autoscaler
vas set to *1*, and the maximum node count was set to *3*. As your application
may need to adjust the cluster autoscaler node count.

unt, use the az aks update command.

Copy Try It

```
p myResourceGroup \
ster \
r-autoscaler \
```

dates cluster autoscaler on the single node pool in *myAKSCluster* to a minimum of
odes.

r makes scaling decisions based on the minimum and maximum counts set on
it does not enforce them after updating the min or max counts. For example,
of 5 when the current node count is 3 will not immediately scale the pool up to 5.
t on the node pool has a value higher than the current number of nodes, the
ings will be respected when there are enough unschedulable pods present that
additional nodes and trigger an autoscaler event. After the scale event, the new
ected.

ce of your applications and services, and adjust the cluster autoscaler node counts
performance.

utoscaler profile

more granular details of the cluster autoscaler by changing the default values in
aler profile. For example, a scale down event happens after nodes are under-

s. If you had workloads that ran every 15 minutes, you may want to change the ale down under utilized nodes after 15 or 20 minutes. When you enable the cluster ofile is used unless you specify different settings. The cluster autoscaler profile has hat you can update:

iption	**Default value**
often cluster is reevaluated for scale up or down	10 seconds
ong after scale up that scale down evaluation resumes	10 minutes
ong after node deletion that scale down evaluation resumes	scan-interval
ong after scale down failure that scale down evaluation resumes	3 minutes
ong a node should be unneeded before it is eligible for scale down	10 minutes
ong an unready node should be unneeded before it is eligible for scale down	20 minutes
utilization level, defined as sum of requested resources divided by capacity, which a node can be considered for scale down	0.5
num number of seconds the cluster autoscaler waits for pod termination when to scale down a node	600 seconds
ts similar node pools and balances the number of nodes between them	false
of node pool expander to be used in scale up. Possible values: `most-pods`, m, `least-waste`, `priority`	random
cluster autoscaler will never delete nodes with pods with local storage, for le, EmptyDir or HostPath	true
cluster autoscaler will never delete nodes with pods from kube-system (except emonSet or mirror pods)	true

ıum number of empty nodes that can be deleted at the same time	10 nodes

iption	**Default value**
enarios like burst/batch scale where you don't want CA to act before the netes scheduler could schedule all the pods, you can tell CA to ignore eduled pods before they're a certain age.	0 seconds
ıum percentage of unready nodes in the cluster. After this percentage is ded, CA halts operations	45%
ıum time the autoscaler waits for a node to be provisioned	15 minutes
ıer of allowed unready nodes, irrespective of max-total-unready-percentage	3 nodes

r profile affects all node pools that use the cluster autoscaler. You can't set an r node pool.

r profile requires version *2.11.1* or greater of the Azure CLI. If you need to install ll Azure CLI.

r autoscaler profile on an existing AKS cluster

command with the *cluster-autoscaler-profile* parameter to set the cluster ur cluster. The following example configures the scan interval setting as 30s in the

Copy Try It

```
p myResourceGroup \
ster \
caler-profile scan-interval=30s
```

luster autoscaler on node pools in the cluster, those clusters will also use the le. For example:

Copy Try It

```
pdate \
p myResourceGroup \
myAKSCluster \
```

```
ot \
r-autoscaler \
```

uster autoscaler profile, any existing node pools with the cluster autoscaler
ng the profile immediately.

r autoscaler profile when creating an AKS cluster

uster-autoscaler-profile parameter when you create your cluster. For example:

Copy Try It

```
p myResourceGroup \
ster \
\
r-autoscaler \

caler-profile scan-interval=30s
```

reates an AKS cluster and defines the scan interval as 30 seconds for the cluster-
The command also enables the cluster autoscaler on the initial node pool, sets the
o 1 and the maximum node count to 3.

autoscaler profile to default values

command to reset the cluster autoscaler profile on your cluster.

Copy Try It

```
p myResourceGroup \
ster \
caler-profile ""
```

cluster autoscaler

o use the cluster autoscaler, you can disable it using the az aks update command, `le-cluster-autoscaler` parameter. Nodes aren't removed when the cluster

Copy Try It

```
p myResourceGroup \
ster \
er-autoscaler
```

your cluster after disabling the cluster autoscaler by using the az aks scale
ıe horizontal pod autoscaler, that feature continues to run with the cluster
t pods may end up unable to be scheduled if all node resources are in use.

disabled cluster autoscaler

ɛ the cluster autoscaler on an existing cluster, you can re-enable it using the az aks
ɔecifying the `--enable-cluster-autoscaler`, `--min-count`, and `--max-count`

ster autoscaler logs and status

ɟ autoscaler events, logs and status can be retrieved from the autoscaler add-on.

ɐr autoscaler on your behalf and runs it in the managed control plane. You can
ɔde to see the logs and operations from CA.

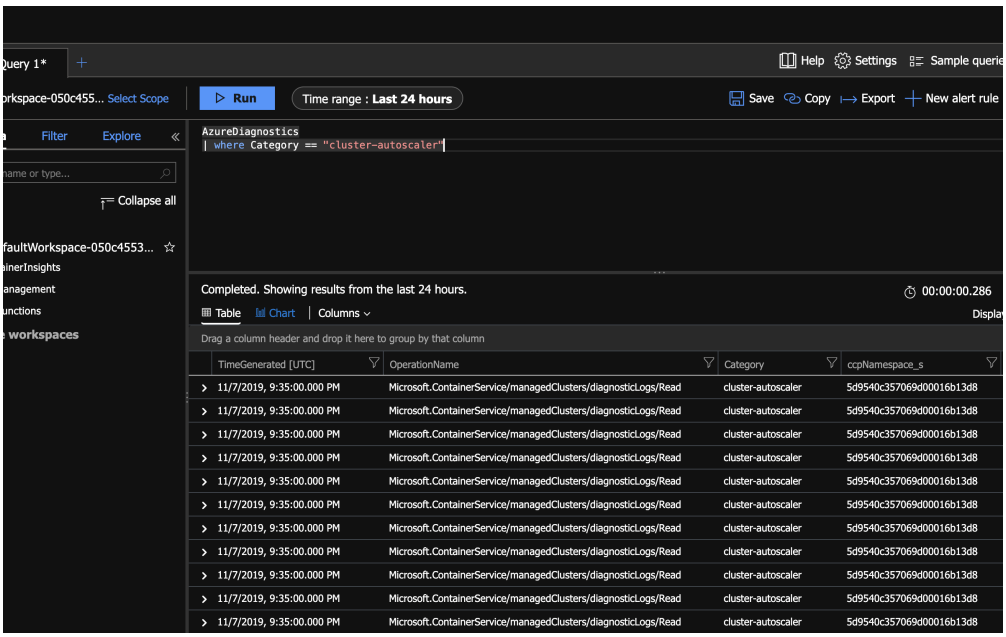
pushed from the cluster autoscaler into Log Analytics, follow these steps.

esource logs to push cluster-autoscaler logs to Log Analytics. Instructions are
ure you check the box for `cluster-autoscaler` when selecting options for "Logs".
section on your cluster via the Azure portal.
g example query into Log Analytics:

```
Copy

== "cluster-autoscaler"
```

ıilar to the following example as long as there are logs to retrieve.



vill also write out health status to a `configmap` named `cluster-autoscaler-`
se logs, execute the following `kubectl` command. A health status will be reported
ıfigured with the cluster autoscaler.

Copy

```
gmap -n kube-system cluster-autoscaler-status -o yaml
```

hat is logged from the autoscaler, read the FAQ on the Kubernetes/autoscaler

ter autoscaler with multiple node pools

an be used together with multiple node pools enabled. Follow that document to ltiple node pools and add additional node pools to an existing cluster. When using you enable the cluster autoscaler on each individual node pool in the cluster and aling rules to each.

ssumes you followed the initial instructions earlier in this document and you want ode pool's max-count from *3* to *5*. Use the az aks nodepool update command to e pool's settings.

Copy Try It

```
pdate \
p myResourceGroup \
myAKSCluster \
1 \
r-autoscaler \
```

an be disabled with az aks nodepool update and passing the `--disable-` parameter.

Copy Try It

```
pdate \
p myResourceGroup \
myAKSCluster \
1 \
er-autoscaler
```

the cluster autoscaler on an existing cluster, you can re-enable it using the az aks mmand, specifying the `--enable-cluster-autoscaler`, `--min-count`, and `--` .

n using the cluster autoscaler with nodepools that span multiple zones and features related to zones such as volume topological scheduling, the o have one nodepool per zone and enable the `--balance-similar-node-` autoscaler profile. This will ensure that the autoscaler will scale up succesfully sizes of the nodepools balanced.

how to automatically scale the number of AKS nodes. You can also use the
ler to automatically adjust the number of pods that run your application. For steps
pod autoscaler, see Scale applications in AKS.

pful?

ed content

ernetes Service (AKS) cluster - Azure Kubernetes Service
number of nodes in an Azure Kubernetes Service (AKS) cluster.

ool to an Azure Kubernetes Service (AKS) cluster - Azure Kubernetes Service
ot node pool to an Azure Kubernetes Service (AKS) cluster.

pools in Azure Kubernetes Service (AKS) - Azure Kubernetes Service
nd manage multiple node pools for a cluster in Azure Kubernetes Service (AKS)

in Azure Kubernetes Services (AKS) - Azure Kubernetes Service
Azure Kubernetes Service (AKS), including volumes, persistent volumes, storage classes,

Show more