

1. Create a Kubernetes Container

Create a brand new Azure Kubernetes Services (AKS) cluster using the portal, with the following characteristics:

- Dev/Test (\$) cluster
- B2ms node size
- 3 nodes with manual scaling

1. From the Azure Home page, click on **create a resource**.
2. Find **Kubernetes Service** from the list.
3. Choose the correct subscription and resource group.
4. Choose the **Dev/Test (\$)** cluster preset configuration.
5. Give the cluster a name.
6. Choose a region close to you.

The screenshot shows the 'Create Kubernetes cluster' form in the Microsoft Azure portal. The form is titled 'Create Kubernetes cluster' and includes a search bar at the top. The form is divided into several sections: 'Subscription', 'Resource group', 'Cluster details', 'Kubernetes cluster name', 'Region', 'Availability zones', and 'Kubernetes version'. Each section has a dropdown menu for selection. The 'Subscription' dropdown is set to 'Pay-As-You-Go'. The 'Resource group' dropdown is set to '(New) azsjdmykube'. The 'Cluster details' section is expanded, showing 'Cluster preset configuration' set to 'Dev/Test (\$)'. Below this, there is a note: 'To quickly customize your Kubernetes cluster, choose one of the preset configurations above. You can modify these configurations at any time. [Learn more and compare presets](#)'. The 'Kubernetes cluster name' dropdown is set to 'azsjdmykube' with a green checkmark. The 'Region' dropdown is set to '(Europe) West Europe'. The 'Availability zones' dropdown is set to 'None'. The 'Kubernetes version' dropdown is set to '1.23.8 (default)'.

Microsoft Azure Search resources, services, and docs (G+/)

Home > Create a resource >

Create Kubernetes cluster

Subscription * ⓘ Pay-As-You-Go

Resource group * ⓘ (New) azsjdmykube [Create new](#)

Cluster details

Cluster preset configuration Dev/Test (\$)
 To quickly customize your Kubernetes cluster, choose one of the preset configurations above. You can modify these configurations at any time. [Learn more and compare presets](#)

Kubernetes cluster name * ⓘ azsjdmykube ✓

Region * ⓘ (Europe) West Europe

Availability zones ⓘ None

Kubernetes version * ⓘ 1.23.8 (default)

7. Choose a node size of **B2ms**. You will be creating **3 nodes with manual scaling**.


API server availability ⓘ

☐ 99.9%

Optimize for availability. 99.95% is available when at least one availability zone is selected.

☒ 99.5%

Optimize for cost.



99.5% API server availability is recommended for dev/test configuration.


Primary node pool

The number and size of nodes in the primary node pool in your cluster. For production workloads, at least 3 nodes are recommended for resiliency. For development or test workloads, only one node is required. If you would like to add additional node pools or to see additional configuration options for this node pool, go to the 'Node pools' tab above. You will be able to add additional node pools after creating your cluster. [Learn more about node pools in Azure Kubernetes Service](#)

Node size * ⓘ

Standard B2ms

2 vcpus, 8 GiB memory




Standard B4ms is recommended for dev/test configuration.

[Change size](#)

Scale method * ⓘ

☒ Manual

☐ Autoscale



Autoscaling is recommended for dev/test configuration.

Node count * ⓘ

3

Review + create

< Previous

Next : Node pools >

8. Keep the default for the rest on this page. Click **next**.
9. Keep the default for the node pools page. Click **next**.
10. Keep the default for the access page. Click **next**.
11. Keep the default for the networking page. Click **next**.

12. For integrations, choose your **Azure Container Registry** or create a **new** one. Click **Review and Create**.

The screenshot shows the 'Create Kubernetes cluster' wizard in the Microsoft Azure portal. The 'Integrations' tab is selected, showing options to connect the AKS cluster with additional services. The 'Container registry' dropdown is set to 'TestKubernetesApp20220505184946'. The 'Azure Monitor' section is expanded, showing options to enable Container Insights for more comprehensive data. The 'Container monitoring' and 'Azure Policy' sections are also visible, both with 'Enabled' radio buttons selected. At the bottom, there are buttons for '< Previous', 'Review + create', and 'Next: Advanced >'.

13. Review this page, and click **Create**. **NOTE: Deploying this will take a while. Wait for it to complete.**

The screenshot shows the 'Overview' page for the AKS cluster 'microsoft.aks-20220831185404'. The page indicates that the 'Deployment is in progress'. It provides details about the deployment, including the name, subscription, resource group, start time, and correlation ID. A table at the bottom shows the deployment details.

Resource	Type	Status	Operation details
azsjdnewkubecuster	Microsoft.ContainerServic...	Created	Operation details

Resources

Documentation

- [Create AKS Cluster](#)

2. Deploy the Container Image to the AKS Cluster

Deploy the image from ACR into the AKS cluster that you just created. You'll need to create a YAML file so that AKS knows how to deploy the image within the cluster.

1. Open the **Azure Cloud Shell**, and choose the **Bash/CLI** command prompt. Azure Cloud Shell is represented by the ">" symbol at the top of the Azure Portal. Choose **Bash**.



2. Get the credentials for AKS using the following command "**az aks get-credentials --resource-group <GROUP> --name <CLUSTER>**" where the **<GROUP>** is the name of your resource group and **<CLUSTER>** is the name of your AKS cluster.
3. Type "**kubectl get nodes**" to see the running nodes.

```
Microsoft Azure | Search resources, services, and docs (G+/) | student_7pmnzhdul2u...
Azure services

Bash
Requesting a Cloud Shell.Succeeded.
Connecting terminal...

Welcome to Azure Cloud Shell

Type "az" to use Azure CLI
Type "help" to learn about Cloud Shell

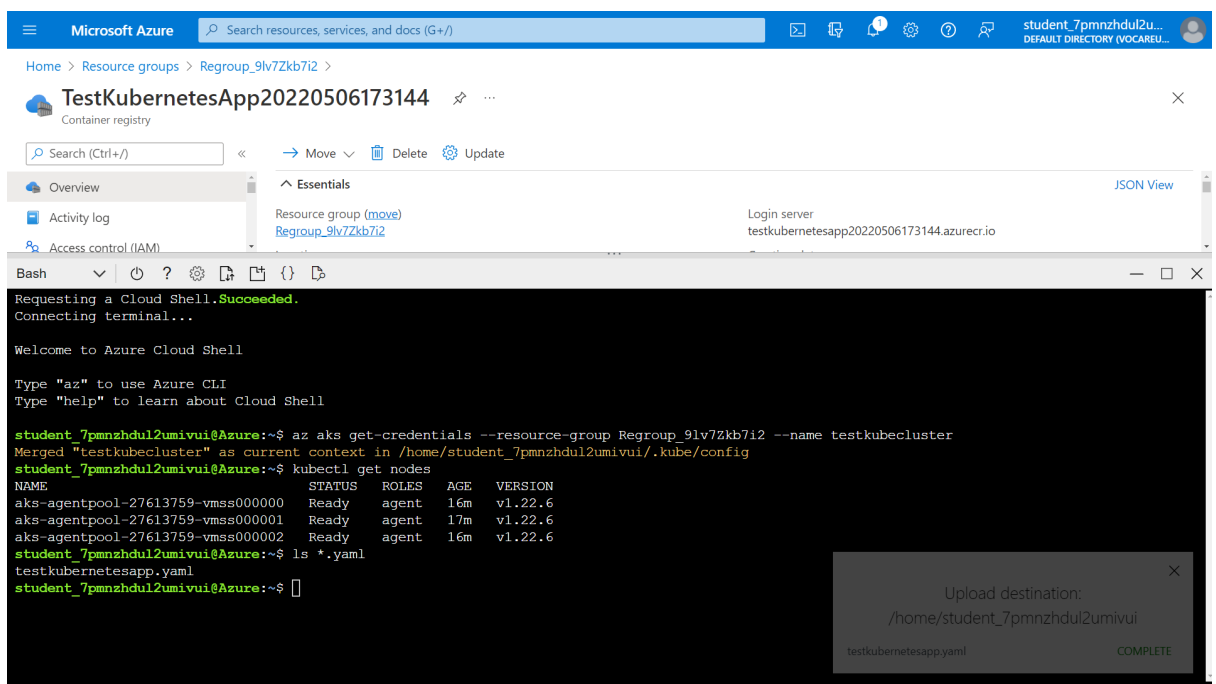
student_7pmnzhdul2umivui@Azure:~$ az aks get-credentials --resource-group Regroup_9lv7Zkb7i2 --name testkubecuster
Merged "testkubecuster" as current context in /home/student_7pmnzhdul2umivui/.kube/config
student_7pmnzhdul2umivui@Azure:~$ kubectl get nodes
NAME                                STATUS    ROLES    AGE    VERSION
aks-agentpool-27613759-vmss000000   Ready    agent    16m    v1.22.6
aks-agentpool-27613759-vmss000001   Ready    agent    17m    v1.22.6
aks-agentpool-27613759-vmss000002   Ready    agent    16m    v1.22.6
student_7pmnzhdul2umivui@Azure:~$
```

4. Download the sample YAML file called "testkubernetesapp.yaml".
5. Using a text editor, modify the YAML file to connect with your Azure Container Registry, replacing the **<youracr>** and **<yourappname>** with the appropriate values. Be sure and save the changes.

```
C:\Users\scott\Downloads\testkubernetesapp - Copy.yaml - Sublime Text (LICENSE UPGRADE REQUIRED)
File Edit Selection Find View Goto Tools Project Preferences Help

testkubernetesapp - Copy.yaml
1 apiVersion: apps/v1
2 kind: Deployment
3 metadata:
4   labels:
5     app: testkubernetesapp <- YOUR APP NAME
6   name: testkubernetesapp <- YOUR APP NAME
7 spec:
8   replicas: 1
9   selector:
10    matchLabels:
11      app: testkubernetesapp <- YOUR APP NAME
12   template:
13     metadata:
14       labels:
15         app: testkubernetesapp <- YOUR APP NAME
16     spec:
17       containers:
18         - image: testkubernetesapp20220505184946.azurecr.io/testkubernetesapp:latest <- YOUR ACR URL INCLUDING NAME AND LABEL
19           name: testkubernetesapp <- YOUR APP NAME
20           ports:
21             - containerPort: 80
22 ---
23 apiVersion: v1
24 kind: Service
25 metadata:
26   name: testkubernetesapp <- YOUR APP NAME
27 spec:
28   type: LoadBalancer
29   ports:
30     - port: 80
31   selector:
```

6. Upload the YAML file to Azure Cloud Shell using the upload function.
7. Type **"ls *.yaml"** to verify the upload worked.



```
Microsoft Azure Search resources, services, and docs (G+/) student_7pmnzhdul2u... DEFAULT DIRECTORY (VOCAREU...)

Home > Resource groups > Regroup_9lv7Zkb7i2 >

TestKubernetesApp20220506173144 Container registry

Search (Ctrl+/) Move Delete Update

Overview Activity log Access control (IAM) Essentials JSON View

Resource group (move) Regroup_9lv7Zkb7i2

Login server testkubernetesapp20220506173144.azurecr.io

Bash ? ? ? ? ? ? ? ? ? ?

Requesting a Cloud Shell.Succeeded.
Connecting terminal...

Welcome to Azure Cloud Shell

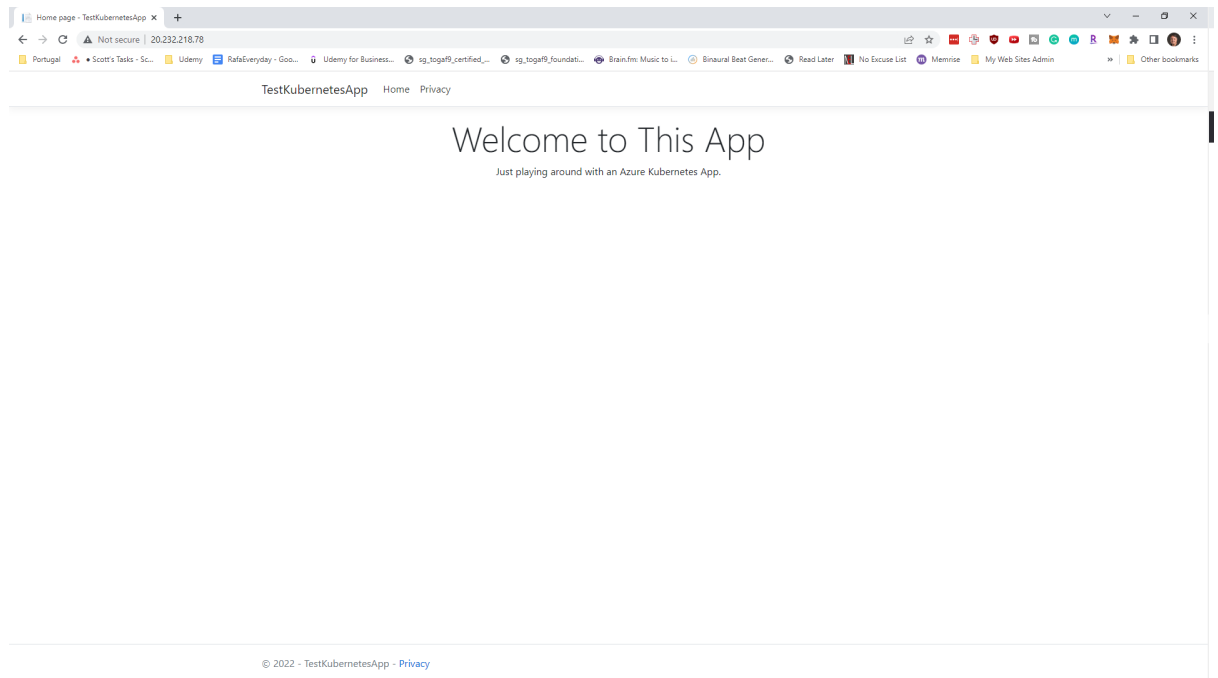
Type "az" to use Azure CLI
Type "help" to learn about Cloud Shell

student_7pmnzhdul2umivui@Azure:~$ az aks get-credentials --resource-group Regroup_9lv7Zkb7i2 --name testkubecoluster
Merged "testkubecoluster" as current context in /home/student_7pmnzhdul2umivui/.kube/config
student_7pmnzhdul2umivui@Azure:~$ kubectl get nodes
NAME STATUS ROLES AGE VERSION
aks-agentpool-27613759-vmss000000 Ready agent 16m v1.22.6
aks-agentpool-27613759-vmss000001 Ready agent 17m v1.22.6
aks-agentpool-27613759-vmss000002 Ready agent 16m v1.22.6
student_7pmnzhdul2umivui@Azure:~$ ls *.yaml
testkubernetesapp.yaml
student_7pmnzhdul2umivui@Azure:~$
```

Upload destination:
/home/student_7pmnzhdul2umivui
testkubernetesapp.yaml COMPLETE

8. Deploy the YAML file to Kubernetes using the command **"kubectl apply -f <yourfile>.yaml"** where <yourfile> is your YAML file name.

13. See that your application is successfully running in Azure Kubernetes Service.



Resources

Assets

- [testkubernetesapp.yaml](#)

Documentation

- [Deploy to AKS](#)

3. Scaling

You need to scale horizontally and vertically. Horizontal scaling means that the response to increased load is to deploy more Pods. This is different from vertical scaling, which for Kubernetes would mean assigning more resources (for example memory or CPU) to the Pods that are already running for the workload. Start with vertical scaling.

1. Open the Azure Cloud Shell, and choose the Bash/CLI command prompt. Azure Cloud Shell is represented by the ">_" symbol at the top of the Azure Portal. Choose Bash.



2. Type **kubectl scale --replicas=10 deployment/azure-vote-front** to increase the number of pods to 10.
3. Type **kubectl get pods -o wide** and check to see if you now have 10 pods running:

```
jordi_koenderink@Azure:~$ kubectl get pods -o wide
```

NAME	READY	STATUS	RESTARTS	AGE	IP	NODE	NOMINATED NODE	READINESS GATES
azure-vote-back-59cb7dc555-tt2r2	1/1	Running	0	13m	10.244.1.2	aks-agentpool-21567423-vmss000000	<none>	<none>
azure-vote-front-5f4d7db9c8-26wpv	1/1	Running	0	15s	10.244.0.11	aks-agentpool-21567423-vmss000002	<none>	<none>
azure-vote-front-5f4d7db9c8-28f9h	1/1	Running	0	15s	10.244.0.13	aks-agentpool-21567423-vmss000002	<none>	<none>
azure-vote-front-5f4d7db9c8-577nw	1/1	Running	0	13m	10.244.2.2	aks-agentpool-21567423-vmss000001	<none>	<none>
azure-vote-front-5f4d7db9c8-98c4m	1/1	Running	0	15s	10.244.2.4	aks-agentpool-21567423-vmss000001	<none>	<none>
azure-vote-front-5f4d7db9c8-99f5b	1/1	Running	0	4m13s	10.244.1.3	aks-agentpool-21567423-vmss000000	<none>	<none>
azure-vote-front-5f4d7db9c8-bb7rw	1/1	Running	0	15s	10.244.0.14	aks-agentpool-21567423-vmss000002	<none>	<none>
azure-vote-front-5f4d7db9c8-dqkb4	1/1	Running	0	15s	10.244.1.6	aks-agentpool-21567423-vmss000000	<none>	<none>
azure-vote-front-5f4d7db9c8-dv8tk	1/1	Terminating	0	15s	10.244.1.5	aks-agentpool-21567423-vmss000000	<none>	<none>
azure-vote-front-5f4d7db9c8-h7d86	1/1	Terminating	0	15s	10.244.1.4	aks-agentpool-21567423-vmss000000	<none>	<none>
azure-vote-front-5f4d7db9c8-kgdb8	1/1	Terminating	0	15s	10.244.2.5	aks-agentpool-21567423-vmss000001	<none>	<none>
azure-vote-front-5f4d7db9c8-lm24b	1/1	Terminating	0	15s	10.244.2.3	aks-agentpool-21567423-vmss000001	<none>	<none>
azure-vote-front-5f4d7db9c8-w5st2	1/1	Running	0	15s	10.244.2.6	aks-agentpool-21567423-vmss000001	<none>	<none>
azure-vote-front-5f4d7db9c8-wcq92	1/1	Terminating	0	15s	10.244.1.7	aks-agentpool-21567423-vmss000000	<none>	<none>
azure-vote-front-5f4d7db9c8-xjt7m	1/1	Running	0	15s	10.244.0.12	aks-agentpool-21567423-vmss000002	<none>	<none>
azure-vote-front-5f4d7db9c8-zn9nw	1/1	Running	0	4m13s	10.244.0.10	aks-agentpool-21567423-vmss000002	<none>	<none>

4. Perform the same commands but now scale down to 1 pod by typing **kubectl scale --replicas=1 deployment/azure-vote-front**
5. Double check how many pods you now have running by using the command **kubectl get pods -o wide**.

Note: it can take some time to scale down, hence the Terminating status:

```
jordi_koenderink@Azure:~$ kubectl get pods -o wide
```

NAME	READY	STATUS	RESTARTS	AGE	IP	NODE	NOMINATED NODE	READINESS GATES
azure-vote-back-59cb7dc555-tt2r2	1/1	Running	0	15m	10.244.1.2	aks-agentpool-21567423-vmss000000	<none>	<none>
azure-vote-front-5f4d7db9c8-26wpv	1/1	Terminating	0	2m29s	10.244.0.11	aks-agentpool-21567423-vmss000002	<none>	<none>
azure-vote-front-5f4d7db9c8-28f9h	1/1	Terminating	0	2m29s	10.244.0.13	aks-agentpool-21567423-vmss000002	<none>	<none>
azure-vote-front-5f4d7db9c8-577nw	1/1	Terminating	0	15m	10.244.2.2	aks-agentpool-21567423-vmss000001	<none>	<none>
azure-vote-front-5f4d7db9c8-98c4m	1/1	Terminating	0	2m29s	10.244.2.4	aks-agentpool-21567423-vmss000001	<none>	<none>
azure-vote-front-5f4d7db9c8-99f5b	1/1	Running	0	6m27s	10.244.1.3	aks-agentpool-21567423-vmss000000	<none>	<none>
azure-vote-front-5f4d7db9c8-bb7rw	1/1	Terminating	0	2m29s	10.244.0.14	aks-agentpool-21567423-vmss000002	<none>	<none>
azure-vote-front-5f4d7db9c8-dqkb4	1/1	Terminating	0	2m29s	10.244.1.6	aks-agentpool-21567423-vmss000000	<none>	<none>
azure-vote-front-5f4d7db9c8-w5st2	1/1	Terminating	0	2m29s	10.244.2.6	aks-agentpool-21567423-vmss000001	<none>	<none>
azure-vote-front-5f4d7db9c8-xjt7m	1/1	Terminating	0	2m29s	10.244.0.12	aks-agentpool-21567423-vmss000002	<none>	<none>
azure-vote-front-5f4d7db9c8-zn9nw	1/1	Terminating	0	6m27s	10.244.0.10	aks-agentpool-21567423-vmss000002	<none>	<none>

6. Now move on with horizontal scaling. Create the HorizontalPodAutoscaler by typing the command **kubectl autoscale deployment azure-vote-back --cpu-percent=50 --min=1 --max=10**
7. Check the status by using the command **kubectl get hpa**. The current CPU consumption is 2%:

```
jordi_koenderink@Azure:~$ kubectl get hpa
```

NAME	REFERENCE	TARGETS	MINPODS	MAXPODS	REPLICAS	AGE
azure-vote-back	Deployment/azure-vote-back	2%/50%	1	10	1	43s

8. Type **kubectl get service** and write down the external IP.
9. Increase the load to test the autoscaler by typing **kubectl run -i --tty load-generator --rm --image=busybox:1.28 --restart=Never -- /bin/sh -c "while sleep 0.01; do wget -q -O- 20.23.34.186; done"**. Replace the IP address with the IP address from step 8.

Note: type **CTRL+C** to end the loop.

10. Now, run **kubectl get hpa azure-vote-back --watch**
11. After a couple of minutes you should see a higher CPU load, and more replicas:

```
Bash
jordi_koenderink@Azure:~$ kubectl get hpa azure-vote-front --watch
```

NAME	REFERENCE	TARGETS	MINPODS	MAXPODS	REPLICAS	AGE
azure-vote-front	Deployment/azure-vote-front	66%/50%	1	10	3	4m50s

12. Type **CTRL+C** to end the Watch.
13. Type **kubectl get deployment azure-vote-front** to see the replica count matching the figure from the HorizontalPodAutoscaler in another view:

```
Bash
jordi_koenderink@Azure:~$ kubectl get deployment azure-vote-front
```

NAME	READY	UP-TO-DATE	AVAILABLE	AGE
azure-vote-front	3/3	3	3	66m

14. After a couple of minutes the number of replicas have scaled down:

```
jordi_koenderink@Azure:~$ kubectl get hpa azure-vote-front --watch
```

NAME	REFERENCE	TARGETS	MINPODS	MAXPODS	REPLICAS	AGE
azure-vote-front	Deployment/azure-vote-front	1%/50%	1	10	3	8m56s
azure-vote-front	Deployment/azure-vote-front	1%/50%	1	10	3	10m
azure-vote-front	Deployment/azure-vote-front	1%/50%	1	10	1	10m