Taints and Tolerations in Kubernetes

Overview

Platform	Azure
Owner of this SOP:	Fully Managed POD-B
Cloud Services	AKS

Tester	Reviewer
@ Sandeep Naidu Pinnenti	

About Taints and Tolerations

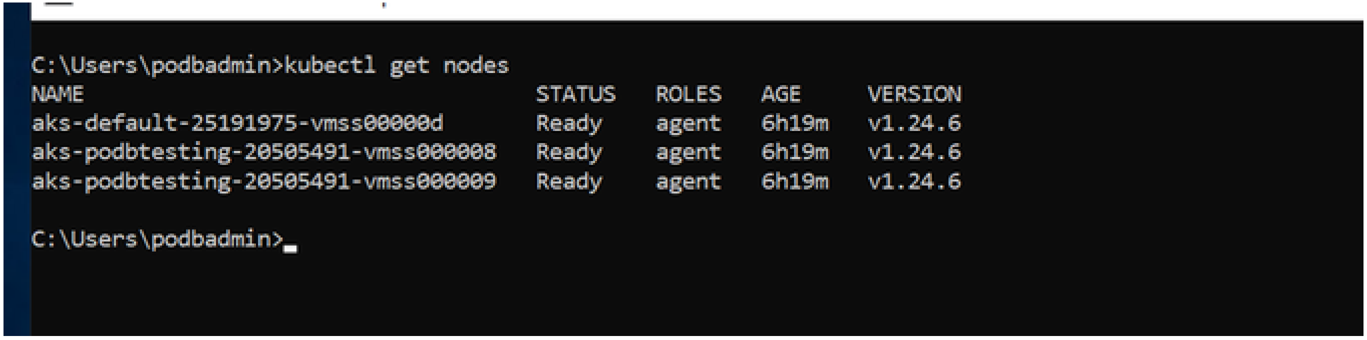
Taints allow a node to repel a set of pods. Tolerations allow the scheduler to schedule pods with matching taints. Taints and tolerations help prevent your pods from scheduling to undesirable nodes.

Taints and tolerations work together to ensure that pods are not scheduled onto inappropriate nodes. One or more taints are applied to a node; this marks that the node should not accept any pods that do not tolerate the taints.

In some cases, we do not want application pods to be placed on nodes that runs system or administration related task. In such scenarios we can taint a node so that pod that do not tolerate the taint cannot be placed in that node.

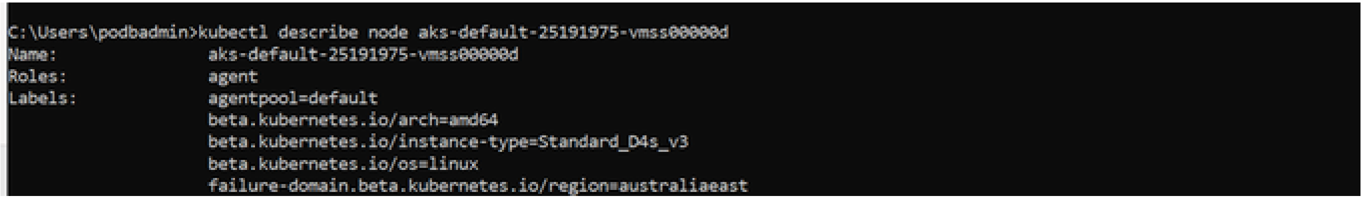
Example

List the nodes that are present in the cluster.

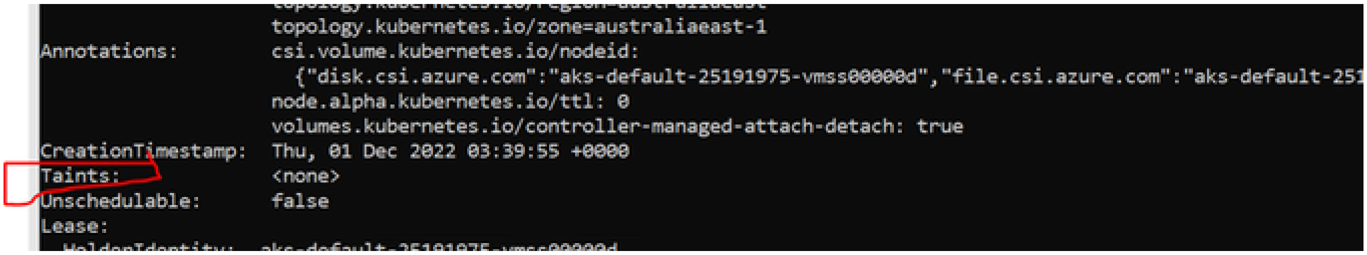


```
C:\Users\podbadmin>kubectl get nodes
NAME                                STATUS   ROLES   AGE     VERSION
aks-default-25191975-vmss00000d     Ready    agent   6h19m   v1.24.6
aks-podbtesting-20505491-vmss000008 Ready    agent   6h19m   v1.24.6
aks-podbtesting-20505491-vmss000009 Ready    agent   6h19m   v1.24.6

C:\Users\podbadmin>_
```

Check the taints that are present on the node. One way is to use **kubectl describe node command.**



```
C:\Users\podbadmin>kubectl describe node aks-default-25191975-vmss00000d
Name:               aks-default-25191975-vmss00000d
Roles:              agent
Labels:             agentpool=default
                    beta.kubernetes.io/arch=amd64
                    beta.kubernetes.io/instance-type=Standard_D4s_v3
                    beta.kubernetes.io/os=linux
                    failure-domain.beta.kubernetes.io/region=australiaeast
```



```
                        topology.kubernetes.io/region=australiaeast
                        topology.kubernetes.io/zone=australiaeast-1
Annotations:            csi.volume.kubernetes.io/nodeid:
                          {"disk.csi.azure.com":"aks-default-25191975-vmss00000d","file.csi.azure.com":"aks-default-251
                        node.alpha.kubernetes.io/ttl: 0
                        volumes.kubernetes.io/controller-managed-attach-detach: true
CreationTimestamp:      Thu, 01 Dec 2022 03:39:55 +0000
Taints:                 <none>
Unschedulable:          false
Lease:
  HolderIdentity:     aks-default-25191975-vmss00000d
```

Add a taint to a node using kubectl taint.

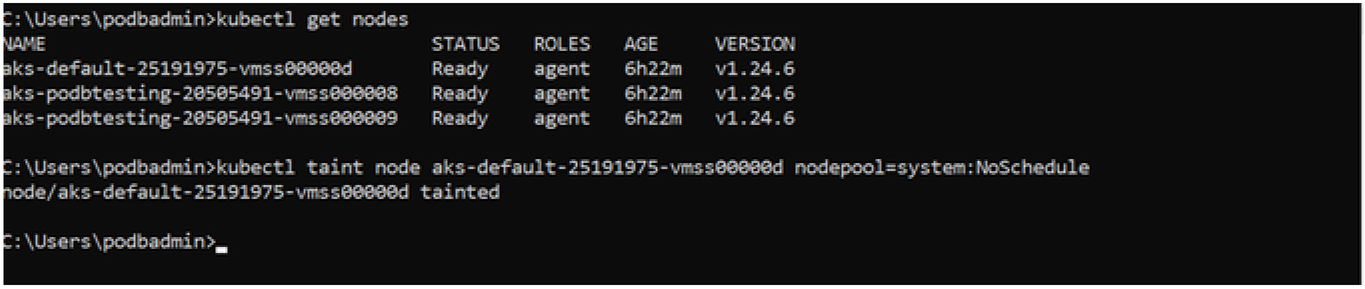
```
kubectl taint nodes node1 key1=value1:Effect
```

Below, we have placed a taint on node “**aks-default-25191975-vmss00000d**” with effect NoSchedule.

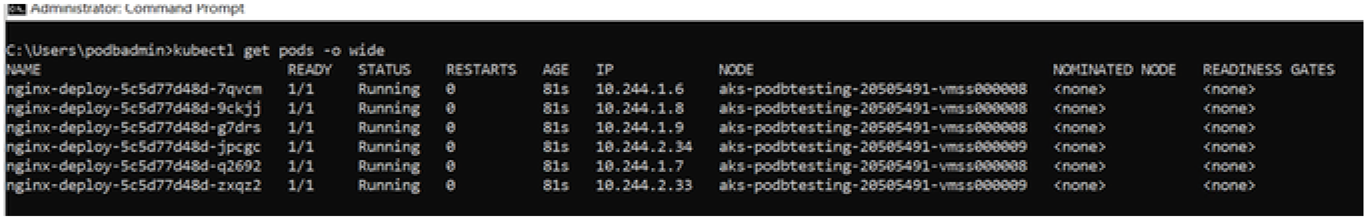
```
kubectl taint nodes node1 key1=value1:NoSchedule
```

The taint has key **nodepool**, value system, and taint effect **NoSchedule**. This means that no pod will be able to schedule onto “**aks-default-25191975-vmss00000d**” unless it has a matching toleration.

This means that pod will be evicted from the node (if it is already running on the node) and will not be scheduled onto the node (if it is not yet running on the node).



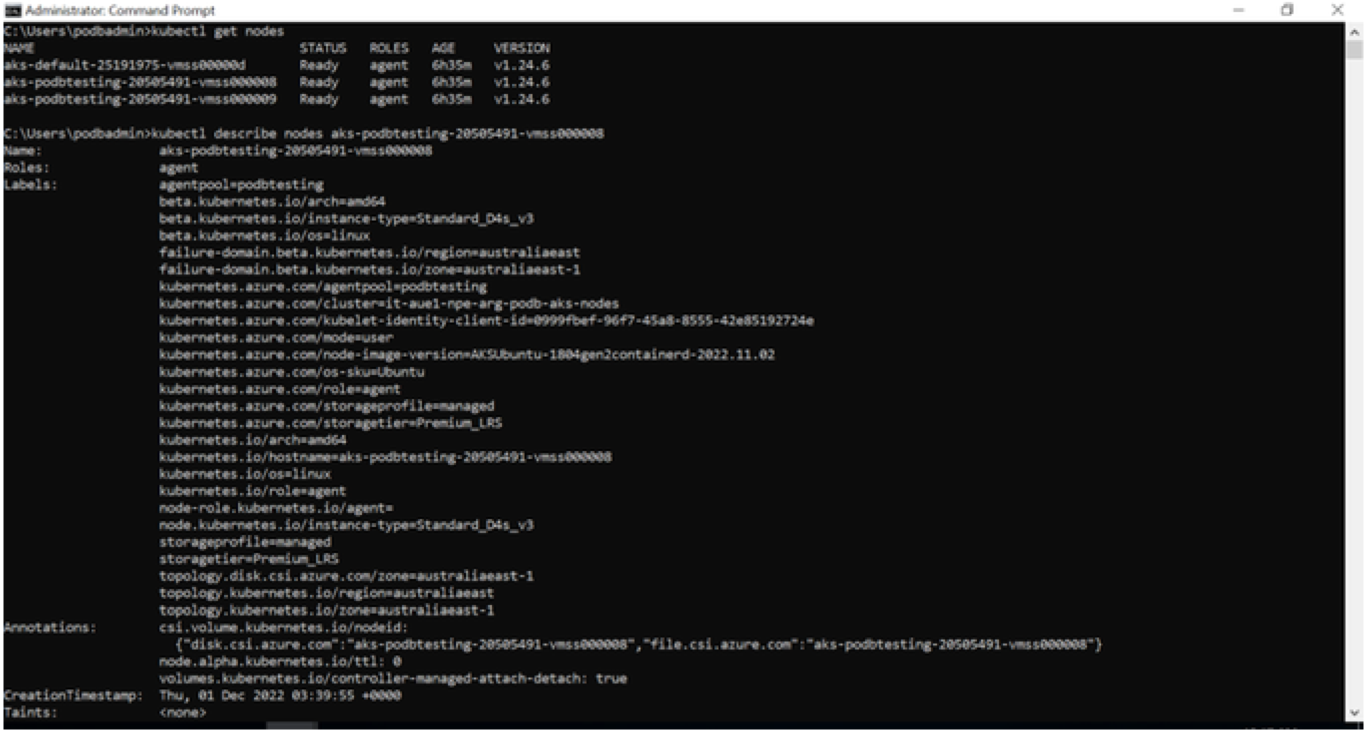
```
C:\Users\podbadmin>kubectl get nodes
NAME                                STATUS   ROLES   AGE     VERSION
aks-default-25191975-vmss00000d     Ready    agent   6h22m   v1.24.6
aks-podbtesting-20505491-vmss000008 Ready    agent   6h22m   v1.24.6
aks-podbtesting-20505491-vmss000009 Ready    agent   6h22m   v1.24.6

C:\Users\podbadmin>kubectl taint node aks-default-25191975-vmss00000d nodepool=system:NoSchedule
node/aks-default-25191975-vmss00000d tainted

C:\Users\podbadmin>_
```



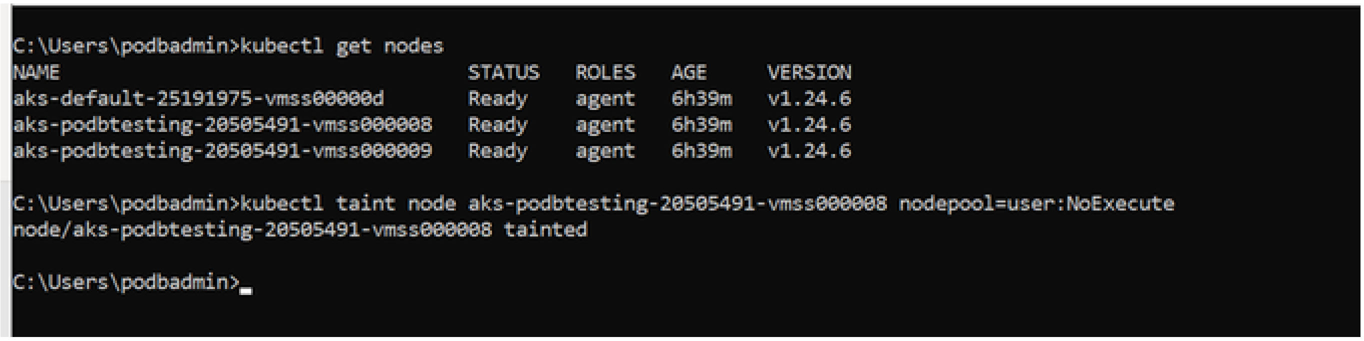
```
C:\Users\podbadmin>kubectl get pods -o wide
NAME                              READY   STATUS    RESTARTS   AGE   IP            NODE                                  NOMINATED NODE   READINESS GATES
nginx-deploy-5c5d77d48d-7qvcm     1/1     Running   0          81s   10.244.1.6    aks-podbtesting-20505491-vmss000008   <none>           <none>
nginx-deploy-5c5d77d48d-9ckjj     1/1     Running   0          81s   10.244.1.8    aks-podbtesting-20505491-vmss000008   <none>           <none>
nginx-deploy-5c5d77d48d-g7drs     1/1     Running   0          81s   10.244.1.9    aks-podbtesting-20505491-vmss000008   <none>           <none>
nginx-deploy-5c5d77d48d-jpcgc     1/1     Running   0          81s   10.244.2.34   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-5c5d77d48d-q2692     1/1     Running   0          81s   10.244.1.7    aks-podbtesting-20505491-vmss000008   <none>           <none>
nginx-deploy-5c5d77d48d-zxqz2     1/1     Running   0          81s   10.244.2.33   aks-podbtesting-20505491-vmss000009   <none>           <none>
```

The above example used effect of NoSchedule. Alternatively, you can use effect of **NoExecute**. This means that pod will be evicted from the node (if it is already running on the node) and will not be scheduled onto the node (if it is not yet running on the node).

```
kubectl taint nodes node1 key1=value1:NoExecute
```

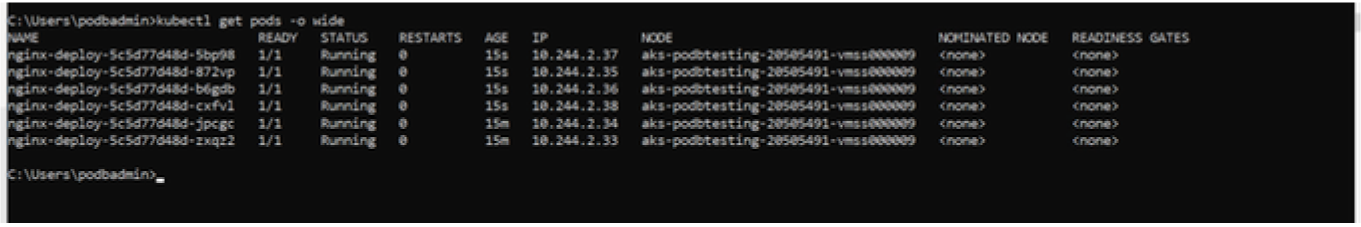


```
C:\Users\podbadmin>kubectl get nodes
NAME                                STATUS   ROLES   AGE     VERSION
aks-default-25191975-vmss00000d     Ready    agent   6h35m   v1.24.6
aks-podbtesting-20505491-vmss000008   Ready    agent   6h35m   v1.24.6
aks-podbtesting-20505491-vmss000009   Ready    agent   6h35m   v1.24.6

C:\Users\podbadmin>kubectl describe nodes aks-podbtesting-20505491-vmss000008
Name:               aks-podbtesting-20505491-vmss000008
Roles:              agent
Labels:             agentpool=podbtesting
                    beta.kubernetes.io/arch=amd64
                    beta.kubernetes.io/instance-type=Standard_D4s_v3
                    beta.kubernetes.io/os=linux
                    failure-domain.beta.kubernetes.io/region=australiaeast
                    failure-domain.beta.kubernetes.io/zone=australiaeast-1
                    kubernetes.azure.com/agentpool=podbtesting
                    kubernetes.azure.com/cluster=it-aue1-npe-arg-podb-aks-nodes
                    kubernetes.azure.com/kubelet-identity-client-id=0999fbef-96f7-45a8-8555-42e85192724e
                    kubernetes.azure.com/mode=user
                    kubernetes.azure.com/node-image-version=AKSUbuntu-1804gen2containerd-2022.11.02
                    kubernetes.azure.com/os-sku=Ubuntu
                    kubernetes.azure.com/role=agent
                    kubernetes.azure.com/storageprofile=managed
                    kubernetes.azure.com/storagetier=Premium_LRS
                    kubernetes.io/arch=amd64
                    kubernetes.io/hostname=aks-podbtesting-20505491-vmss000008
                    kubernetes.io/os=linux
                    kubernetes.io/role=agent
                    node-role.kubernetes.io/agent=
                    node.kubernetes.io/instance-type=Standard_D4s_v3
                    storageprofile=managed
                    storagetier=Premium_LRS
                    topology.disk.csi.azure.com/zone=australiaeast-1
                    topology.kubernetes.io/region=australiaeast
                    topology.kubernetes.io/zone=australiaeast-1
Annotations:        csi.volume.kubernetes.io/nodeid:
                      {"disk.csi.azure.com":"aks-podbtesting-20505491-vmss000008","file.csi.azure.com":"aks-podbtesting-20505491-vmss000008"}
                    node.alpha.kubernetes.io/ttl: 0
                    volumes.kubernetes.io/controller-managed-attach-detach: true
CreationTimestamp:  Thu, 01 Dec 2022 03:39:55 +0000
Taints:             <none>
```



```
C:\Users\podbadmin>kubectl get nodes
NAME                                  STATUS   ROLES   AGE     VERSION
aks-default-25191975-vmss00000d       Ready    agent   6h39m   v1.24.6
aks-podbtesting-20505491-vmss000008   Ready    agent   6h39m   v1.24.6
aks-podbtesting-20505491-vmss000009   Ready    agent   6h39m   v1.24.6

C:\Users\podbadmin>kubectl taint node aks-podbtesting-20505491-vmss000008 nodepool=user:NoExecute
node/aks-podbtesting-20505491-vmss000008 tainted

C:\Users\podbadmin>
```

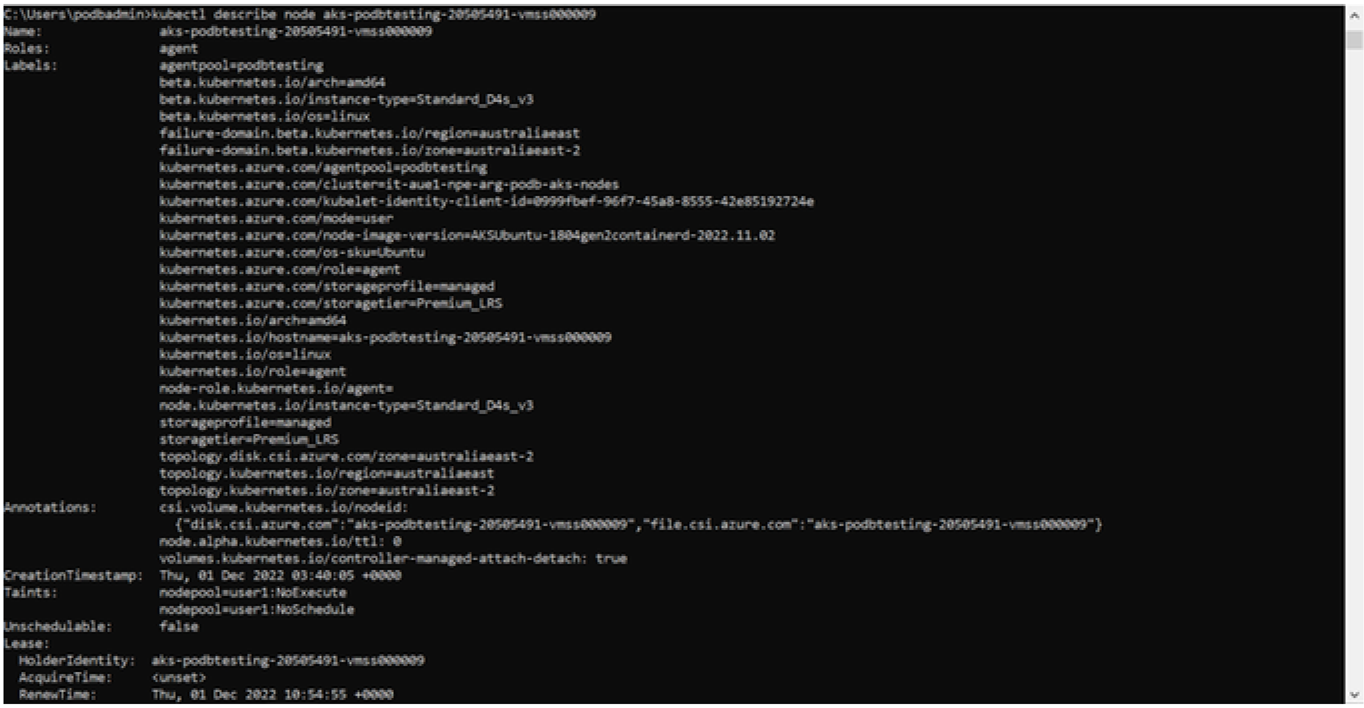
Once the taint is applied to the node, all the pods that are present in the node are evicted and are scheduled in other node.



```
C:\Users\podbadmin>kubectl get pods -o wide
NAME                             READY   STATUS    RESTARTS   AGE   IP            NODE                                  NOMINATED NODE   READINESS GATES
nginx-deploy-5c5d77d48d-5bp98    1/1     Running   0          15s   10.244.2.37   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-5c5d77d48d-872vp    1/1     Running   0          15s   10.244.2.35   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-5c5d77d48d-b6gdb    1/1     Running   0          15s   10.244.2.36   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-5c5d77d48d-cxfvl    1/1     Running   0          15s   10.244.2.38   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-5c5d77d48d-jpcgc    1/1     Running   0          15m   10.244.2.34   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-5c5d77d48d-zxqz2    1/1     Running   0          15m   10.244.2.33   aks-podbtesting-20505491-vmss000009   <none>           <none>

C:\Users\podbadmin>
```

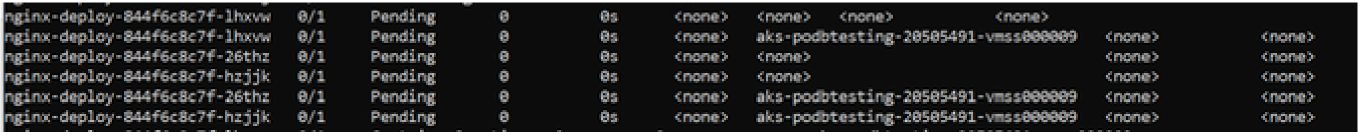
Tolerations

Tolerations allow the scheduler to schedule pods with matching taints.



```
C:\Users\podbadmin>kubectl describe node aks-podbtesting-20505491-vmss000009
Name:               aks-podbtesting-20505491-vmss000009
Roles:              agent
Labels:             agentpool=podbtesting
                    beta.kubernetes.io/arch=amd64
                    beta.kubernetes.io/instance-type=Standard_D4s_v3
                    beta.kubernetes.io/os=linux
                    failure-domain.beta.kubernetes.io/region=australiaeast
                    failure-domain.beta.kubernetes.io/zone=australiaeast-2
                    kubernetes.azure.com/agentpool=podbtesting
                    kubernetes.azure.com/cluster=it-aue1-npe-arg-podb-aks-nodes
                    kubernetes.azure.com/kubelet-identity-client-id=0999fbef-96f7-45a8-8555-42e85192724e
                    kubernetes.azure.com/mode=user
                    kubernetes.azure.com/node-image-version=AKSUbuntu-1804gen2containerd-2022.11.02
                    kubernetes.azure.com/os-sku=Ubuntu
                    kubernetes.azure.com/role=agent
                    kubernetes.azure.com/storageprofile=managed
                    kubernetes.azure.com/storagetier=Premium_LRS
                    kubernetes.io/arch=amd64
                    kubernetes.io/hostname=aks-podbtesting-20505491-vmss000009
                    kubernetes.io/os=linux
                    kubernetes.io/role=agent
                    node-role.kubernetes.io/agent=
                    node.kubernetes.io/instance-type=Standard_D4s_v3
                    storageprofile=managed
                    storagetier=Premium_LRS
                    topology.disk.csi.azure.com/zone=australiaeast-2
                    topology.kubernetes.io/region=australiaeast
                    topology.kubernetes.io/zone=australiaeast-2
Annotations:        csi.volume.kubernetes.io/nodeid:
                      {"disk.csi.azure.com":"aks-podbtesting-20505491-vmss000009","file.csi.azure.com":"aks-podbtesting-20505491-vmss000009"}
                    node.alpha.kubernetes.io/ttl: 0
                    volumes.kubernetes.io/controller-managed-attach-detach: true
CreationTimestamp:  Thu, 01 Dec 2022 03:40:05 +0000
Taints:             nodepool=user1:NoExecute
                    nodepool=user1:NoSchedule
Unschedulable:      false
Lease:
  HolderIdentity:  aks-podbtesting-20505491-vmss000009
  AcquireTime:     <unset>
  RenewTime:       Thu, 01 Dec 2022 10:54:55 +0000
```

Pods are in pending state as the toleration is not defined in POD YAML spec.

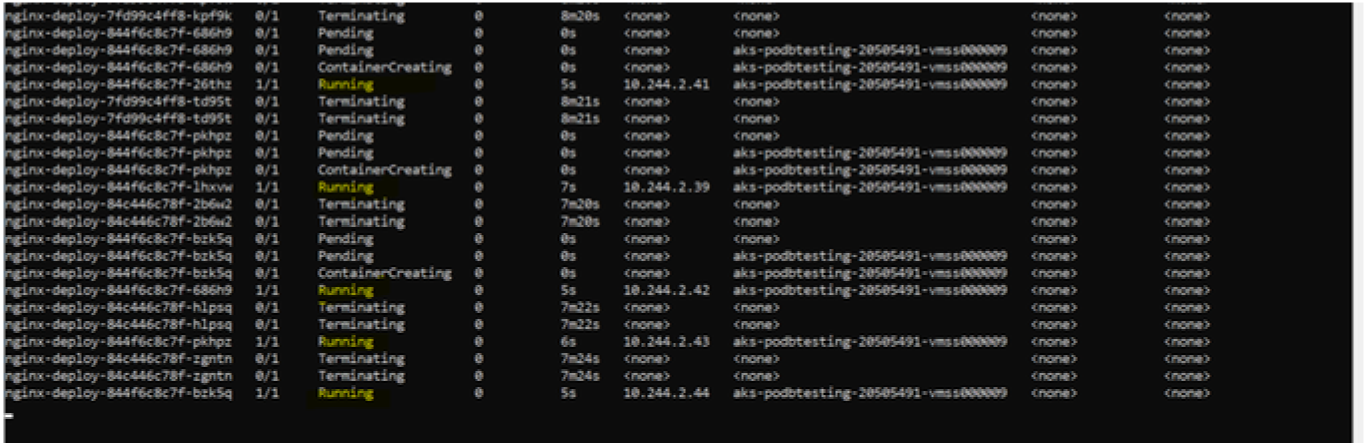


```
nginx-deploy-844f6c8c7f-lhxvw   0/1     Pending     0          0s      <none>    <none>   <none>             <none>
nginx-deploy-844f6c8c7f-lhxvw   0/1     Pending     0          0s      <none>    aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-26thz   0/1     Pending     0          0s      <none>    <none>                                <none>           <none>
nginx-deploy-844f6c8c7f-hzjjk   0/1     Pending     0          0s      <none>    <none>                                <none>           <none>
nginx-deploy-844f6c8c7f-26thz   0/1     Pending     0          0s      <none>    aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-hzjjk   0/1     Pending     0          0s      <none>    aks-podbtesting-20505491-vmss000009   <none>           <none>
```

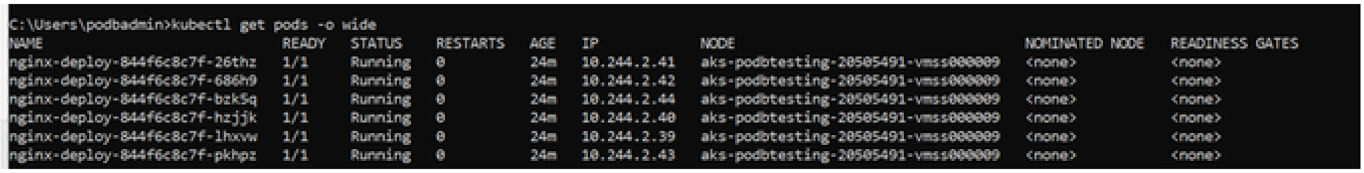
Add tolerations to the Pod YAML file that will tolerate the taint.

```
- key: "nodepool"
  operator: "Equal"
  value: "user1"
  effect: "NoSchedule"
- key: "nodepool"
  operator: "Equal"
  value: "user1"
  effect: "NoExecute"
```

Once the tolerations are added to the Pod YAML spec, we can see the pods are in running state.



```
nginx-deploy-7fd99c4ff8-kpf9k   0/1     Terminating         0          8m20s   <none>        <none>                                <none>           <none>
nginx-deploy-844f6c8c7f-686h9   0/1     Pending             0          0s      <none>        <none>                                <none>           <none>
nginx-deploy-844f6c8c7f-686h9   0/1     Pending             0          0s      <none>        aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-686h9   0/1     ContainerCreating   0          0s      <none>        aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-26thz   1/1     Running             0          5s      10.244.2.41   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-7fd99c4ff8-td95t   0/1     Terminating         0          8m21s   <none>        <none>                                <none>           <none>
nginx-deploy-7fd99c4ff8-td95t   0/1     Terminating         0          8m21s   <none>        <none>                                <none>           <none>
nginx-deploy-844f6c8c7f-pkhpz   0/1     Pending             0          0s      <none>        <none>                                <none>           <none>
nginx-deploy-844f6c8c7f-pkhpz   0/1     Pending             0          0s      <none>        aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-pkhpz   0/1     ContainerCreating   0          0s      <none>        aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-lhxvw   1/1     Running             0          7s      10.244.2.39   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-84c446c78f-2b6w2   0/1     Terminating         0          7m20s   <none>        <none>                                <none>           <none>
nginx-deploy-84c446c78f-2b6w2   0/1     Terminating         0          7m20s   <none>        <none>                                <none>           <none>
nginx-deploy-844f6c8c7f-bzk5q   0/1     Pending             0          0s      <none>        <none>                                <none>           <none>
nginx-deploy-844f6c8c7f-bzk5q   0/1     Pending             0          0s      <none>        aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-bzk5q   0/1     ContainerCreating   0          0s      <none>        aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-686h9   1/1     Running             0          5s      10.244.2.42   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-84c446c78f-hlpsq   0/1     Terminating         0          7m22s   <none>        <none>                                <none>           <none>
nginx-deploy-84c446c78f-hlpsq   0/1     Terminating         0          7m22s   <none>        <none>                                <none>           <none>
nginx-deploy-844f6c8c7f-pkhpz   1/1     Running             0          6s      10.244.2.43   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-84c446c78f-zgntn   0/1     Terminating         0          7m24s   <none>        <none>                                <none>           <none>
nginx-deploy-84c446c78f-zgntn   0/1     Terminating         0          7m24s   <none>        <none>                                <none>           <none>
nginx-deploy-844f6c8c7f-bzk5q   1/1     Running             0          5s      10.244.2.44   aks-podbtesting-20505491-vmss000009   <none>           <none>
```



```
C:\Users\podbadmin>kubectl get pods -o wide
NAME                            READY   STATUS    RESTARTS   AGE   IP            NODE                                     NOMINATED NODE   READINESS GATES
nginx-deploy-844f6c8c7f-26thz   1/1     Running   0          24m   10.244.2.41   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-686h9   1/1     Running   0          24m   10.244.2.42   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-bzk5q   1/1     Running   0          24m   10.244.2.44   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-hzjjk   1/1     Running   0          24m   10.244.2.40   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-lhxvw   1/1     Running   0          24m   10.244.2.39   aks-podbtesting-20505491-vmss000009   <none>           <none>
nginx-deploy-844f6c8c7f-pkhpz   1/1     Running   0          24m   10.244.2.43   aks-podbtesting-20505491-vmss000009   <none>           <none>
```

Related articles

- https://kubernetes.io/docs/concepts/scheduling-eviction/taint-and-toleration/