In OpenShift 4 the "kube-apiserver" pods are in "CrashLoopBackOff", and there are thousands of "installer-N-retry" pods in failed state in the "openshift-kube-apiserver" namespace

SOLUTION VERIFIED - Updated October 13 2022 at 12:14 PM - English

Environment

- Red Hat OpenShift Container Platform (RHOCP)
  - 4.10 or higher.

Issue

- The `openshift-kube-apiserver` namespace has thousands of pods `installer-REVISION-retry-NODE` in failed state.
- The failed reason for most of these failed pods is `OOMKilled` .

Resolution

Clean the `etcd` database by removing the old, and outdated, `install-retry` entries with related events, following the next steps:

1. Connect to an `etcd` pod running in a healthy node:

```
$ oc -n openshift-etcd exec -ti etcd-<HEALTHY_NODE> -c etcdctl -- /bin/bash
```

- If the command `etcdctl` is working well, then execute:

```
$ etcdctl del /kubernetes.io/pods/openshift-kube-apiserver/installer- --prefix
$ etcdctl del /kubernetes.io/events/openshift-kube-apiserver/installer- --prefix
```

- If the command `etcdctl` is trying to connect to an unhealthy node, then execute:

```
$ unset ETCDCTL_ENDPOINTS
$ etcdctl --endpoints=localhost:2379 del /kubernetes.io/pods/openshift-kube-apiserver/installer- --prefix
$ etcdctl --endpoints=localhost:2379 del /kubernetes.io/events/openshift-kube-apiserver/installer- --prefix
```

2. Force an update of the `kube-apiserver` revision by finding and deleting the most recent revision `configmap` :

```
$ oc -n openshift-kube-apiserver get configmap | grep kube-apiserver-pod
$ oc -n openshift-kube-apiserver delete configmap kube-apiserver-pod-<LATEST_REVISION>
```

Wait a few minutes for the cluster to consolidate and estabilize. You may delete the crashed `kube-apiserver` pod to speed the process.

Root Cause

- It is uncertain how this state was reached in the first place. In the current state, it was possible to ascertain that:
  - the `etcd` database was filled with all the records of the previous attempts,
  - the `installer` pod loads the current data into memory at run time,
  - the `installer` pod has a limit of 200M:

```
$ oc -n openshift-kube-apiserver get pod installer-474-retry-NNNN-NODE -o yaml
[...]
    resources:
      limits:
        cpu: 150m
        memory: 200M
      requests:
        cpu: 150m
        memory: 200M
```

- The high amount of data together with the 200M limit was causing the `OOMKilled` events. Because, the `installer` pod was not able to perform the update of the `kube-apiserver` configuration, which includes the certificates rotation, this service failed with an outdated configuration.

Diagnostic Steps

- There are no memory constrains in the control plane nodes.
- The `etcd` status is OK.

- The `kube-apiserver` operator is degraded:

```
$ oc get co
NAME                                       VERSION   AVAILABLE   PROGRESSING
DEGRADED   SINCE
etcd                                       4.10.16   True        False
False      7h7m
kube-apiserver                             4.10.16   True        True
True       8h11m
```

- The `kube-apiserver` operator reports the error:

```
$ oc get co kube-apiserver -o yaml
status:
  conditions:
  - lastTransitionTime: "2022-10-10T05:09:38Z"
    message: "GuardControllerDegraded: Missing operand on node
NODE\nNodeInstallerDegraded:
      1 nodes are failing on revision 474:\n
    [...]
    reason: GuardController_SyncError::NodeInstaller_InstallerPodFailed
    status: "True"
    type: Degraded
  - lastTransitionTime: "2022-10-03T04:58:48Z"
    message: 'NodeInstallerProgressing: 2 nodes are at revision 0; 1 nodes are
at
      revision 468; 0 nodes have achieved new revision 474'
    reason: NodeInstaller
    status: "True"
    type: Progressing
```

- There are thousands of attempts to change the `kube-apiserver` control plane node configuration. Attempts from the current revision, and from previous revisions as well:

```
$ oc -n openshift-kube-apiserver get pod
NAME                            READY   STATUS      RESTARTS   AGE
[...]
installer-471-NODE              0/1     Failed      0          4d
installer-471-retry-NNNN-NODE   0/1     Failed      0          4d    (Repeated
more than 1200 times)
[...]
installer-472-NODE              0/1     Failed      0          2d
installer-472-retry-NNNN-NODE   0/1     Failed      0          2d    (Repeated
more than 3100 times)
[...]
installer-473-NODE              0/1     Failed      0          9h30m
installer-473-retry-NNNN-NODE   0/1     Failed      0          9h30m (Repeated 5
times)
[...]
installer-474-NODE              0/1     Failed      0          9h
installer-474-retry-NNNN-NODE   0/1     Failed      0          9h    (Repeated
158 times)
[...]
kube-apiserver-guard-NODE       1/1     Running     0          7d
kube-apiserver-NODE             1/5     CrashLoopBackOff nnnn  9h
revision-pruner-474-NODE        0/1     Completed   0          9h
```

- There may be references of `no such file or directory` pointing to non-existente locations such as `/etc/kubernetes/static-pod-resources/secrets/` or `/etc/kubernetes/static-pod-certs/` :

```
E1013 10:51:45.091310      16 runtime.go:77] Observed a panic: open
/etc/kubernetes/static-pod-resources/secrets/etcd-client/tls.crt: no such file
or directory
```

This is to be expected, because these are the internal mount points of the `kube-apiserver` pod:

```
$ oc -n openshift-kube-apiserver describe pod kube-apiserver-NODE
[...]
    Mounts:
      /etc/kubernetes/static-pod-certs from cert-dir (rw)
      /etc/kubernetes/static-pod-resources from resource-dir (rw)
[...]
Volumes:
  resource-dir:
    Type:          HostPath (bare host directory volume)
    Path:          /etc/kubernetes/static-pod-resources/kube-apiserver-pod-
REVISION
    HostPathType:
  cert-dir:
    Type:          HostPath (bare host directory volume)
    Path:          /etc/kubernetes/static-pod-resources/kube-apiserver-certs
    HostPathType:
```

Product(s) Red Hat OpenShift Container Platform **Component** kubernetes

Category Troubleshoot **Tags** kubernetes troubleshooting

This solution is part of Red Hat's fast-track publication program, providing a huge library of solutions that Red Hat engineers have created while supporting our customers. To give you the knowledge you need the instant it becomes available, these articles may be presented in a raw and unedited form.

Was this helpful?

YES NO

People who viewed this solution also viewed

kube-apiserver pod does not come up

Solution - Apr 8, 2024

Why the revision and installer pods for kube-apiserver operator rolls out automatically?

Solution - Sep 30, 2022

How to recover static static kube-apiserver Pod on SNO

Solution - Aug 14, 2023

Get notified when this content is updated FOLLOW

Comments

Add comment

Formatting Help

Send notifications to content followers

Submit

Quick Links

Help

Site Info

Related Sites

About

Red Hat Subscription Value

About Red Hat

Red Hat Jobs