

PodSupervisor

Are your pods sleeping?

Glenn West
gwest@redhat.com

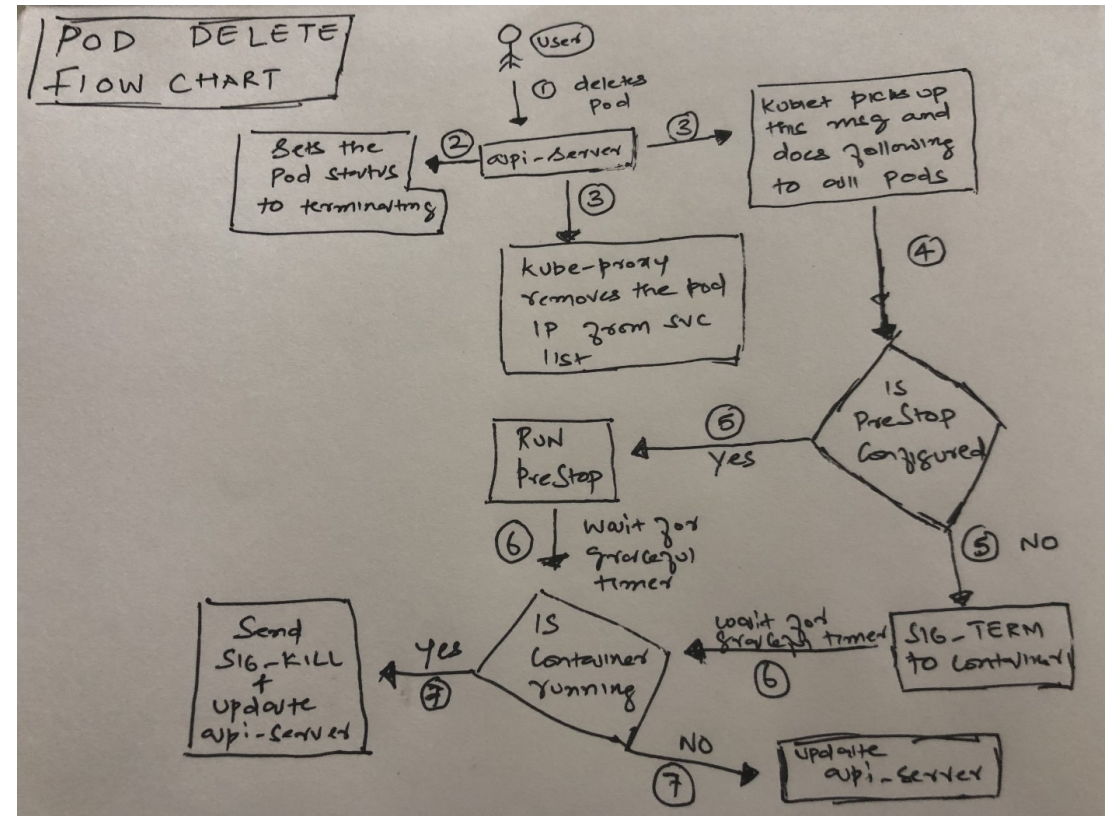
Issue

- In a highly redundant OpenShift environment – Fail Fast is important to uptime of applications.
- OpenShift “kills” pods on not-ready node when node reboots. In addition, Nodes marked unscheduable should also be avoided for better uptime.



Solution

- A new monitor script that watches all nodes for not-ready or un-schedeable are cleaned up of all non-system pods
- Non-system pods do not have “openshift-” in there namespace/project name.



Example

- Node is marked un-scheduable
- Script is run
- Two pods are deleted

```
gwest@gwest@redhat podsupervisor % oc get nodes
NAME                                STATUS    ROLES    AGE   VERSION
control-plane-0                    Ready     master   27h   v1.20.10+bbbc079
control-plane-1                    Ready     master   27h   v1.20.10+bbbc079
control-plane-2                    Ready     master   27h   v1.20.10+bbbc079
worker-0                           Ready     worker   27h   v1.20.10+bbbc079
worker-1                           Ready,SchedulingDisabled worker   27h   v1.20.10+bbbc079
gwest@gwest@redhat podsupervisor % oc get pods
NAME                                READY    STATUS    RESTARTS   AGE
httpd-ex-7bd6c6788-7hlcj           1/1      Running   0           151m
httpd-ex-7bd6c6788-rdfdc           1/1      Running   0           151m
gwest@gwest@redhat podsupervisor % ./podsupervisor.sh
Cleaning up worker-1
Cleaning up worker-1/testapp/httpd-ex-7bd6c6788-7hlcj
pod "httpd-ex-7bd6c6788-7hlcj" deleted
Cleaning up worker-1/testapp/httpd-ex-7bd6c6788-rdfdc
pod "httpd-ex-7bd6c6788-rdfdc" deleted
gwest@gwest@redhat podsupervisor % █
```

Result

- Pods have moved to worker-0, as the worker-1 is un-scheduable.
- No system/openshift pods were touched or harmed.

```
[gwest@gwest@redhat podsupervisor % oc get pods -o wide
NAME                                READY   STATUS    RESTARTS   AGE   IP            NODE       NOMINATED NODE   READINESS GATES
httpd-ex-7bd6c6788-7pvkt            1/1    Running   0          80s   10.131.0.32   worker-0   <none>           <none>
httpd-ex-7bd6c6788-lt269            1/1    Running   0          77s   10.131.0.33   worker-0   <none>           <none>
```

Things to add/Enhancements

- AvoidList for things like operators or other infra components
- State Transition Support – Wait till node is ready once before processing
- Event/Notification to be compatible with prometheus

Source

- Upstream source:
- <https://github.com/glennswest/podsupervisor>