PodSupervisor
Are your pods sleeping?

Glenn West

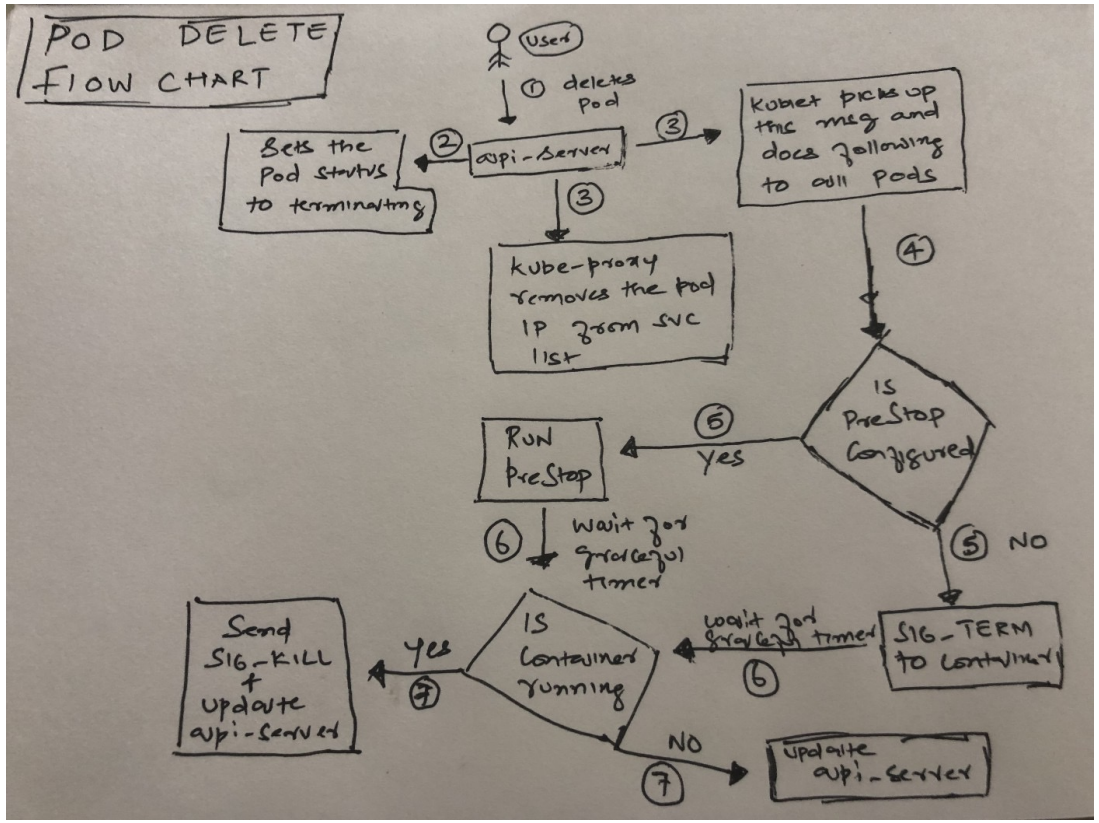
gwest@redhat.com

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 un-scheduable should also be avoided for better uptime.

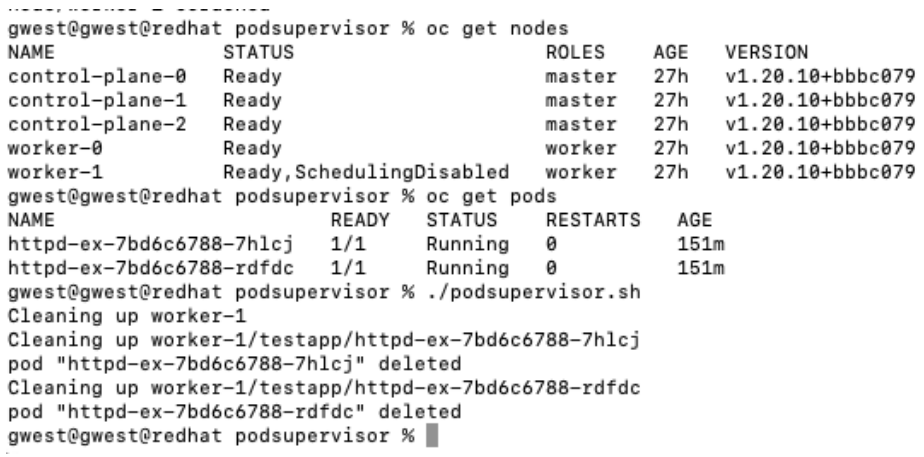
Solution

- 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



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