Clues Not Processing Troubleshooting Guide

Monday, 24 January 2022 11:18 AM

24 Jan 2022 uploaded to https://github.com/CluedIn-io/ImpTools/tree/master/k8s/pwsh

Purpose

To ensure clue processing is flowing through the system.

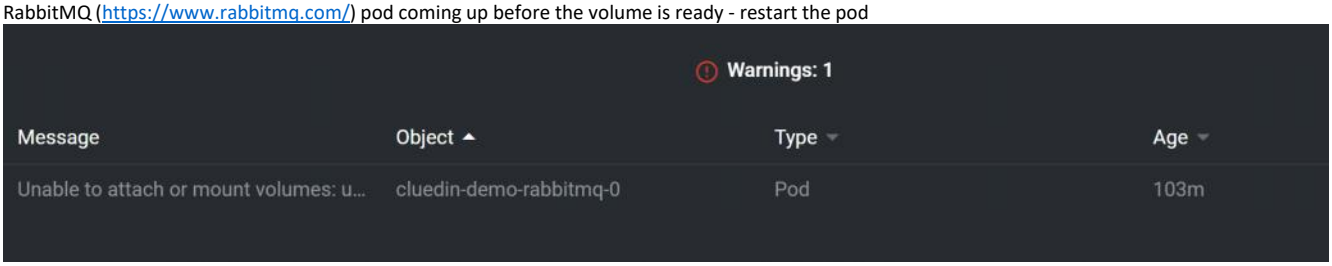
Things to Check

Methods:

- Connect to your k8s cluster with Lens (extemely helpful! - see https://k8slens.dev/)
- Use kubectl on the command line with powershell - see https://github.com/CluedIn-io/ImpTools/tree/master/k8s/pwsh for some sample scripts
- Azure Portal (https://portal.azure.com/)

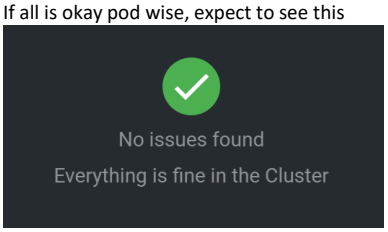
Check the Overall status

RabbitMQ (https://www.rabbitmq.com/) pod coming up before the volume is ready - restart the pod

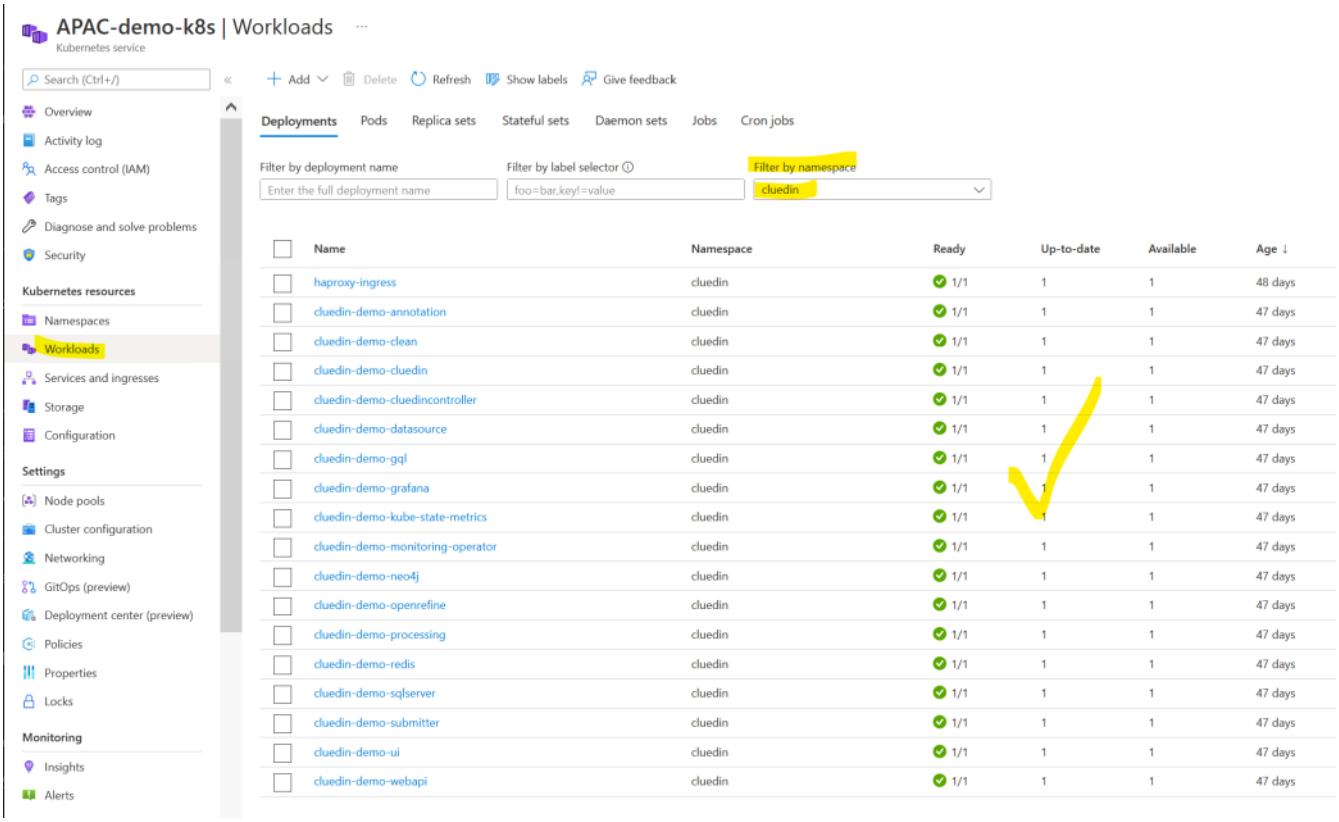


Pod is waiting for another pod to be available - this is how the start up dependancies are expressed.
In this case go sort out the gql pod first.

If all is okay pod wise, expect to see this



Also in the Azure Portal - check the workloads



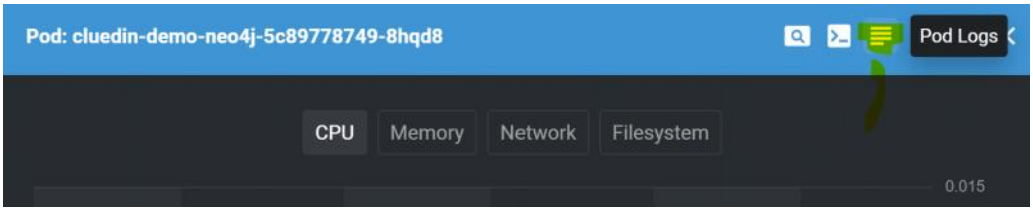
Using kubectl

https://github.com/CluedIn-io/ImpTools/blob/master/k8s/pwsh/gp.ps1 - get pods

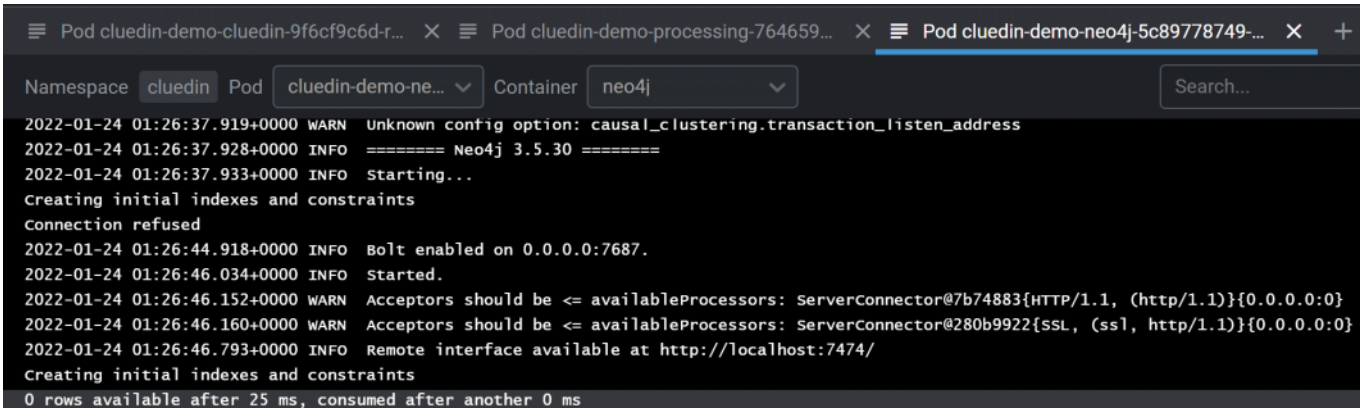
check for restarts as a sign of issues and also the STATUS column
(in this case we had some issues, but they have been resolved).

```
PS C:\rudi.harris\src\CluedIn-io\ImpTools\k8s\pwsh> cat .\gp.ps1
kubectl get pods
PS C:\rudi.harris\src\CluedIn-io\ImpTools\k8s\pwsh> .\gp.ps1
NAME                                                READY   STATUS    RESTARTS   AGE
alertmanager-cluedin-demo-monitoring-alertmanager-0   2/2     Running   0          2d21h
cluedin-demo-annotation-84764c8df9-6tnvk            1/1     Running   0          2d21h
cluedin-demo-clean-66b64c7578-b8z4w                 1/1     Running   0          2d21h
cluedin-demo-cluedin-9f6cf9c6d-rr8z5                1/1     Running   3          2d21h
cluedin-demo-cluedincontroller-65bd75f4b9-9sdzd     1/1     Running   0          2d21h
cluedin-demo-datasource-6b75cd87d8-mb977            1/1     Running   0          2d21h
cluedin-demo-gql-5bf88c9dd9-5jbvt                   1/1     Running   0          2d21h
cluedin-demo-grafana-86c89f8454-gn82s               2/2     Running   0          2d21h
cluedin-demo-kube-state-metrics-5d5bdd975-w2pn6     1/1     Running   0          2d21h
cluedin-demo-monitoring-operator-555c9bff9c-pm49b   1/1     Running   0          2d21h
cluedin-demo-neo4j-5c89778749-8hqd8                 1/1     Running   0          2d21h
cluedin-demo-openrefine-58d66f5979-g2qmr            1/1     Running   0          2d21h
cluedin-demo-processing-764659974b-r8bqx            1/1     Running   1          2d21h
cluedin-demo-prometheus-node-exporter-f5vz4         1/1     Running   0          3h30m
cluedin-demo-prometheus-node-exporter-ndcmf         1/1     Running   0          3h29m
cluedin-demo-prometheus-node-exporter-p6r7l         1/1     Running   0          3h30m
cluedin-demo-prometheus-node-exporter-sx658         1/1     Running   0          3h29m
cluedin-demo-prometheus-node-exporter-wd855         1/1     Running   0          3h29m
cluedin-demo-prometheus-node-exporter-x2fms         1/1     Running   0          3h30m
cluedin-demo-rabbitmq-0                             1/1     Running   0          8m51s
cluedin-demo-redis-78686bc979-jmkkz                 1/1     Running   0          2d21h
cluedin-demo-sqlserver-847f8c6594-2b767             1/1     Running   0          2d21h
cluedin-demo-submitter-5bbcd54774-7tqsv             1/1     Running   0          2d21h
cluedin-demo-ui-6fbfddcbf7-5gcr6                    1/1     Running   0          2d21h
cluedin-demo-webapi-5686697599-gwqcp                1/1     Running   0          2d21h
elasticsearch-master-0                              1/1     Running   0          2d21h
haproxy-ingress-5c9cb6796f-p9v6n                    1/1     Running   0          2d21h
prometheus-cluedin-demo-monitoring-prometheus-0     2/2     Running   1          2d21h
```

Check logs



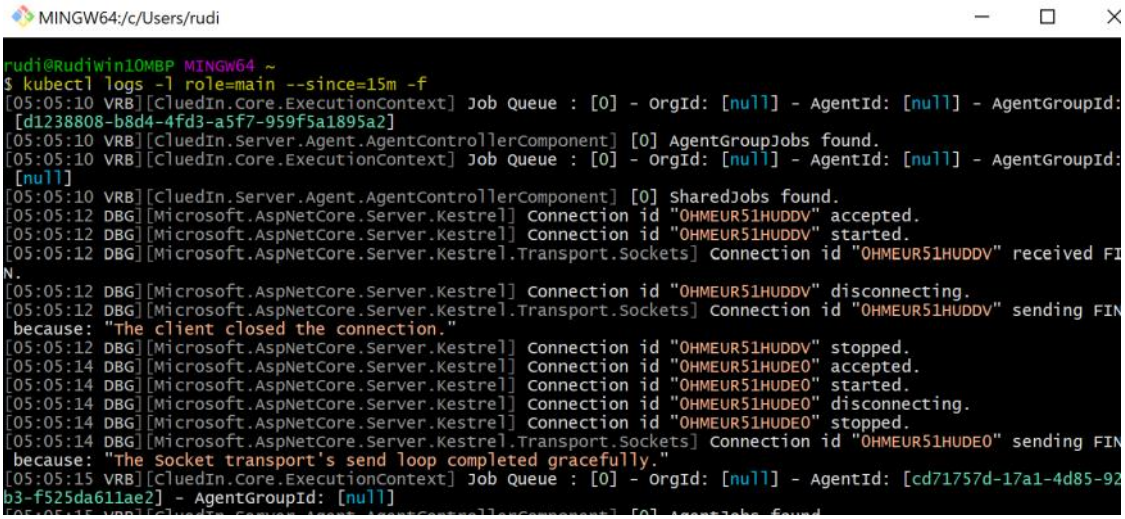
Main server pod, processing pod, neo4j pod - these are the best 3 to start with
Look for out of memory issues in neo4j
Look for C# error stacks and errors in the others



Alternatively, use the kubectl command under git bash (so the coloured logs are shown correctly - https://gitforwindows.org/)

`-l` can be used to reference the pod using a label - or use the pod name from the get pods command

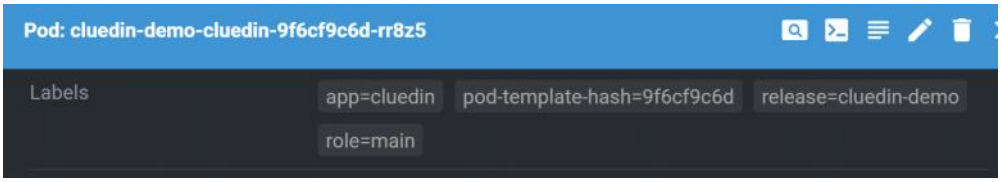
The below example is show the logs for the server (i.e. `role=main`) starting from 15 minutes ago and `-f` for follow the log as new entries come in



Here is a few examples of labels:

```
rudi@RudiWin10MBP MINGW64 /c/rudi.harris/src/CluedIn-io/ImpTools/k8s/pwsh (master)
$ grep "\-l" restart*.ps1
restart_ingress_ui.ps1:kubectl delete pod -l app.kubernetes.io/name=haproxy-ingress
restart_ingress_ui.ps1:kubectl delete pod -l app=ui
restart_main_processing_crawling.ps1:kubectl delete pod -l role=main
restart_main_processing_crawling.ps1:kubectl delete pod -l role=processing
restart_main_processing_crawling.ps1:kubectl delete pod -l role=crawling
restart_rmq.ps1:kubectl delete pod -l app.kubernetes.io/name=rabbitmq
restart_submitter.ps1:kubectl delete pod -l app=submitter
```

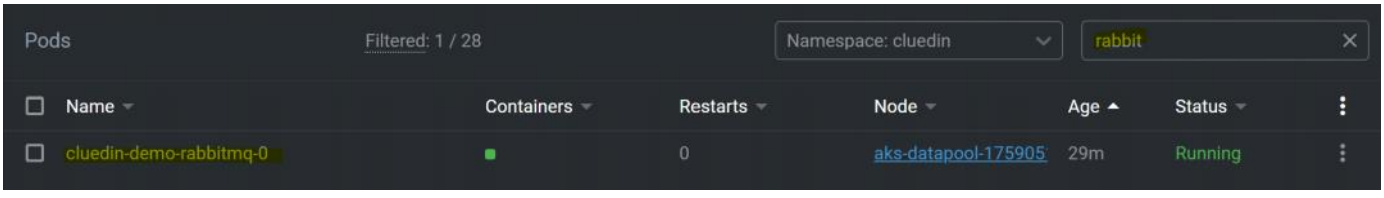
These can be found under labels in lens



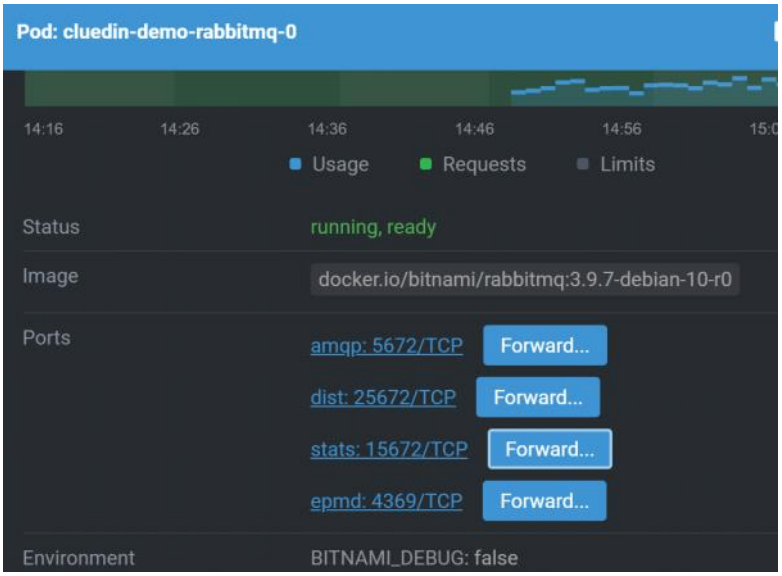
or use kubectl describe pod <pod name>
and search for the labels section:

```
Labels:       app=cluedin
              pod-template-hash=9f6cf9c6d
              release=cluedin-demo
              role=main
```

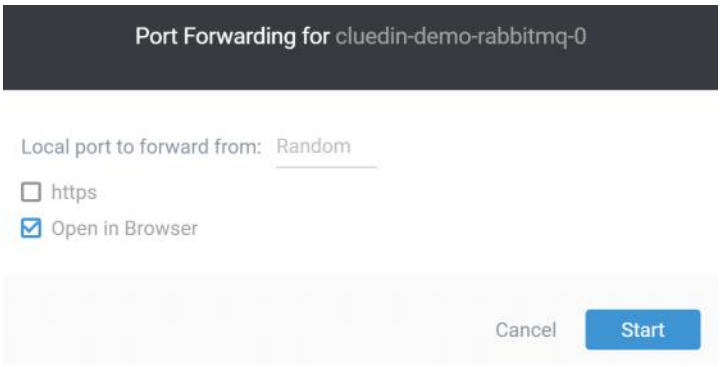
Check the queues in RabbitMQ



Select then scroll to the Ports section



Click Forward for the stats port - accept defaults then click Start



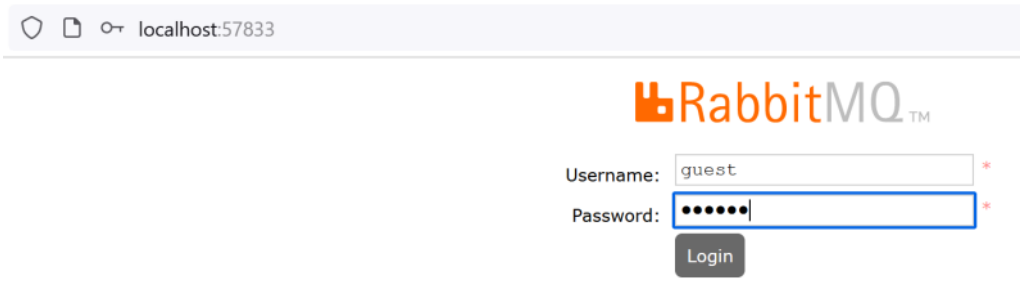
Or do the port forwarding using kubectl

https://github.com/CluedIn-io/ImpTools/blob/master/k8s/pwsh/pf_rmq.ps1

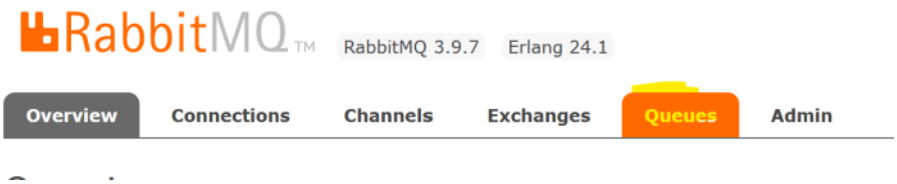
```
PS C:\rudi.harris\src\CluedIn-io\ImpTools\k8s\pwsh> .\pf_rmq.ps1
label is "app.kubernetes.io/name=rabbitmq", looking up name...
name is "pod/cluedin-demo-rabbitmq-0"
Forwarding from 127.0.0.1:15672 -> 15672
Forwarding from [::1]:15672 -> 15672
```

And open your browser to http://localhost:15672/

Login guest/guest



Select queues



Check all the process command queues with the Regex filter for `.*Processing.Process.*Command`

Overview				Messages			Message rates		
Name	Type	Features	State	Ready	Unacked	Total	incoming	deliver / get	ack
CluedIn.Core.Messages.Processing.ProcessBigClueCommand, CluedIn.Core_CluedIn	classic	D lazy-queue	idle	0	0	0			
CluedIn.Core.Messages.Processing.ProcessEdgesCommand, CluedIn.Core_CluedIn	classic	D lazy-queue	idle	0	0	0			
CluedIn.Core.Messages.Processing.ProcessLowPriorityClueCommand, CluedIn.Core_CluedIn	classic	D lazy-queue	idle	0	0	0			
CluedIn.Core.Messages.Processing.ProcessPrioritizedClueCommand, CluedIn.Core_CluedIn	classic	D lazy-queue	idle	0	0	0			
CluedIn.Core.Messages.Processing.ProcessVersionHistoryCommand, CluedIn.Core_CluedIn	classic	D lazy-queue	idle	0	0	0			
CluedIn.Core.Messages.Processing.ProcessWebHookClueCommand, CluedIn.Core_CluedIn	classic	D lazy-queue	idle	0	0	0			

And also for the clue queue
Regex filter `^clue$`

Page 1 of 1 - Filter: ^clue$ Regex ?

Overview				Messages			Message rates			+/-
Name	Type	Features	State	Ready	Unacked	Total	incoming	deliver / get	ack	
clue	classic	D lazy-queue	idle	0	0	0				

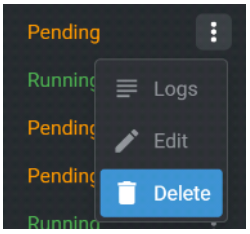
If Ready is all zeros then the submitter has not submitted any clues

If Ready is non-zero then we expect that number to go down over time - if it is not going down then it's time to **restart the processing pod** - see Things to Try below for the method

Things to Try

Restarting pods

RabbitMQ pod

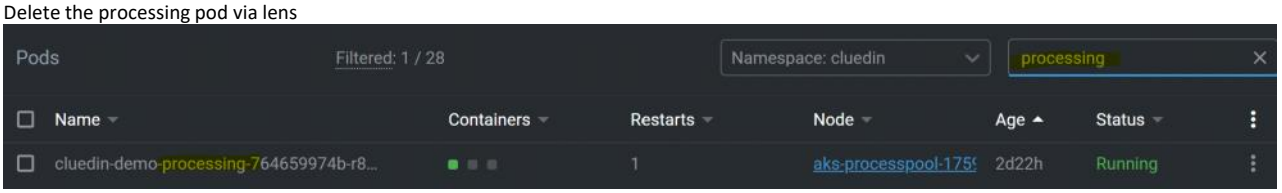


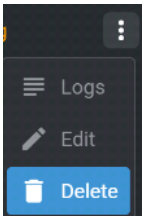
Or using kubectl and Powershell
https://github.com/CluedIn-io/ImpTools/blob/master/k8s/pwsh/restart_rmq.ps1

```
kubectl delete pod -l app.kubernetes.io/name=rabbitmq
```

Processing pod (and the other main pods)

Delete the processing pod via lens





Or using kubectl and Powershell
see the processing role line from https://github.com/CluedIn-io/ImpTools/blob/master/k8s/pwsh/restart_main_processing_crawling.ps1

```
kubectl delete pod -l role=processing
```

Restart submitter pod
https://github.com/CluedIn-io/ImpTools/blob/master/k8s/pwsh/restart_submitter.ps1

```
kubectl delete pod -l app=submitter
```

Likewise, restart neo4j or gql

Restart by group

Restart the main pod (server), the processing pod and the crawling pod
https://github.com/CluedIn-io/ImpTools/blob/master/k8s/pwsh/restart_main_processing_crawling.ps1

```
kubectl delete pod -l role=main
kubectl delete pod -l role=processing
kubectl delete pod -l role=crawling
```

Restart the Ingress and UI
https://github.com/CluedIn-io/ImpTools/blob/master/k8s/pwsh/restart_ingress_ui.ps1

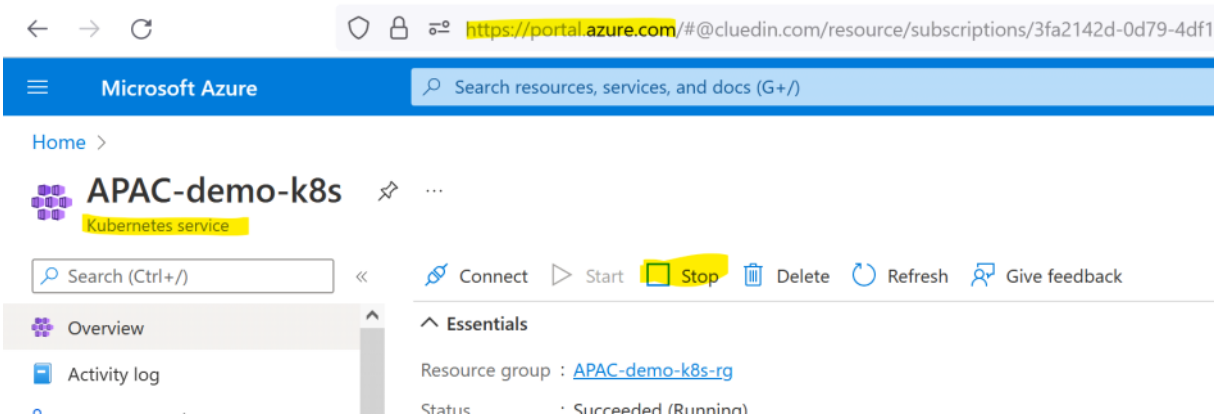
If All Else Fails

https://github.com/CluedIn-io/ImpTools/blob/master/k8s/pwsh/restart_main_processing_crawling.ps1

```
kubectl delete pod -l role=main
kubectl delete pod -l role=processing
kubectl delete pod -l role=crawling
```

Likewise, restart neo4j or gql

Hit the Stop button on the whole cluster



Wait until you are notified that everything has stopped.

Wait some more…

Then when you think it's ready, hit the Start button and start from the top of this advice to troubleshoot the pods as they start up - this takes 15 to 20 minutes typically.