IBM Spectrum Scale Container Native(CNSA) Health Monitoring

Spectrum Scale German User Meeting 2023
Sindelfingen, Germany – March 22-23, 2023

Helene Wassmann

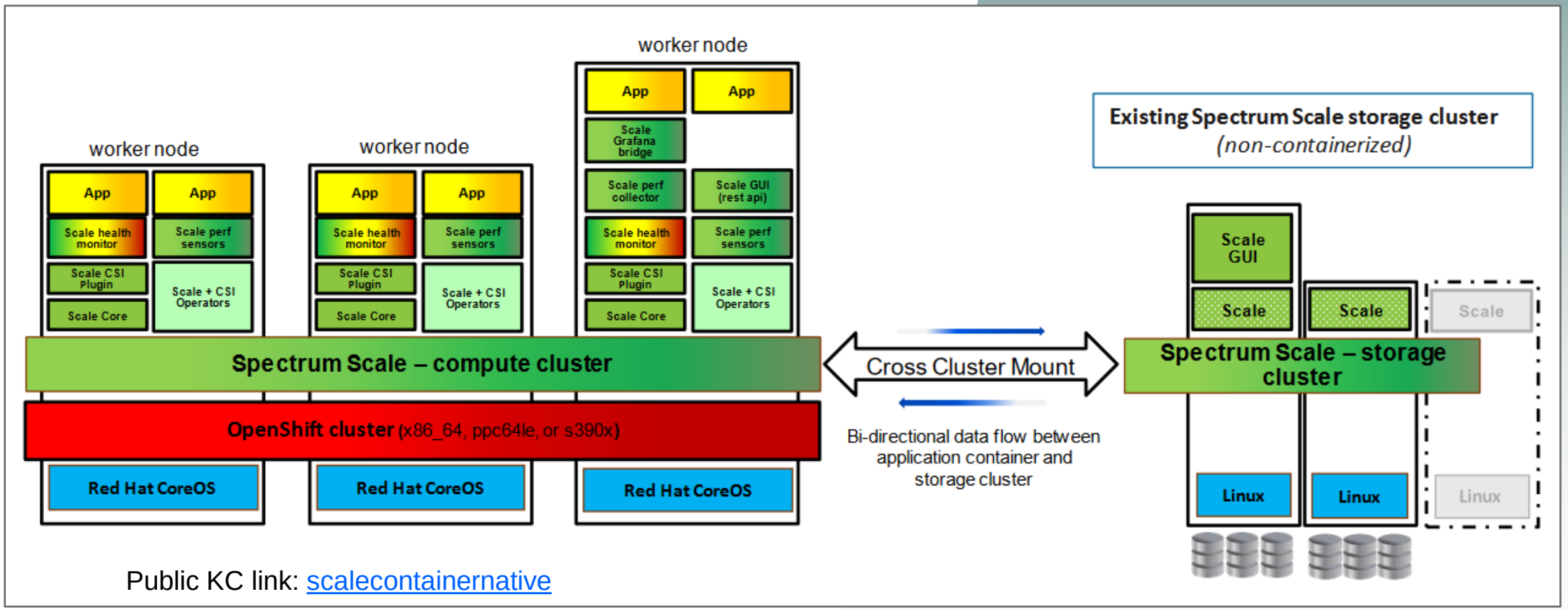
Disclaimer

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

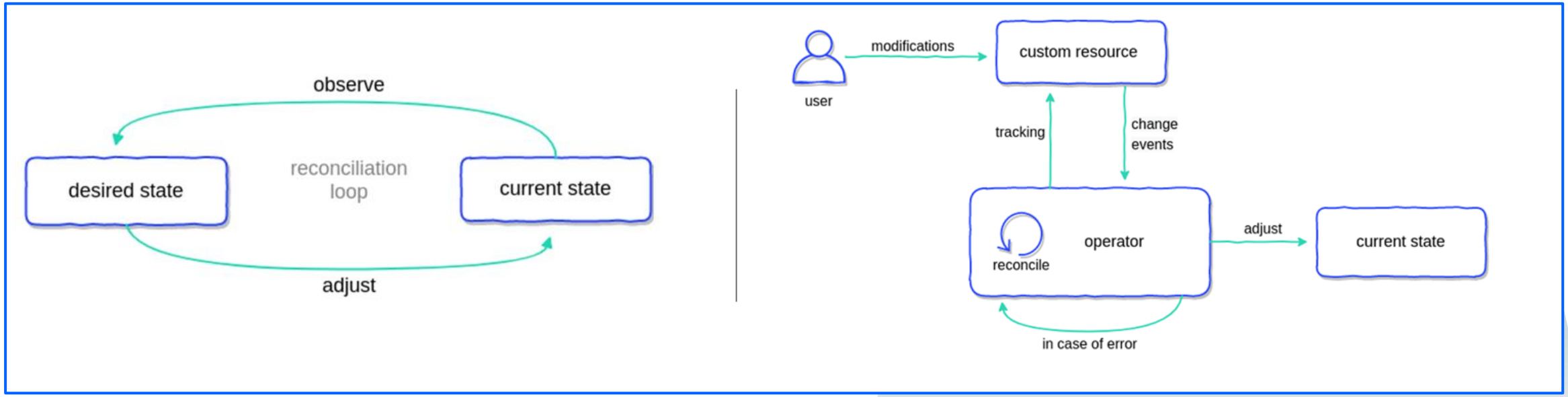
IBM Spectrum Scale container native (CNSA)

- is a containerized version of IBM Spectrum Scale
- allows the deployment of the cluster file system ***in a RedHat OpenShift cluster***



Public KC link: [scalecontainernative]

Kubernetes Controller/Operator Concept

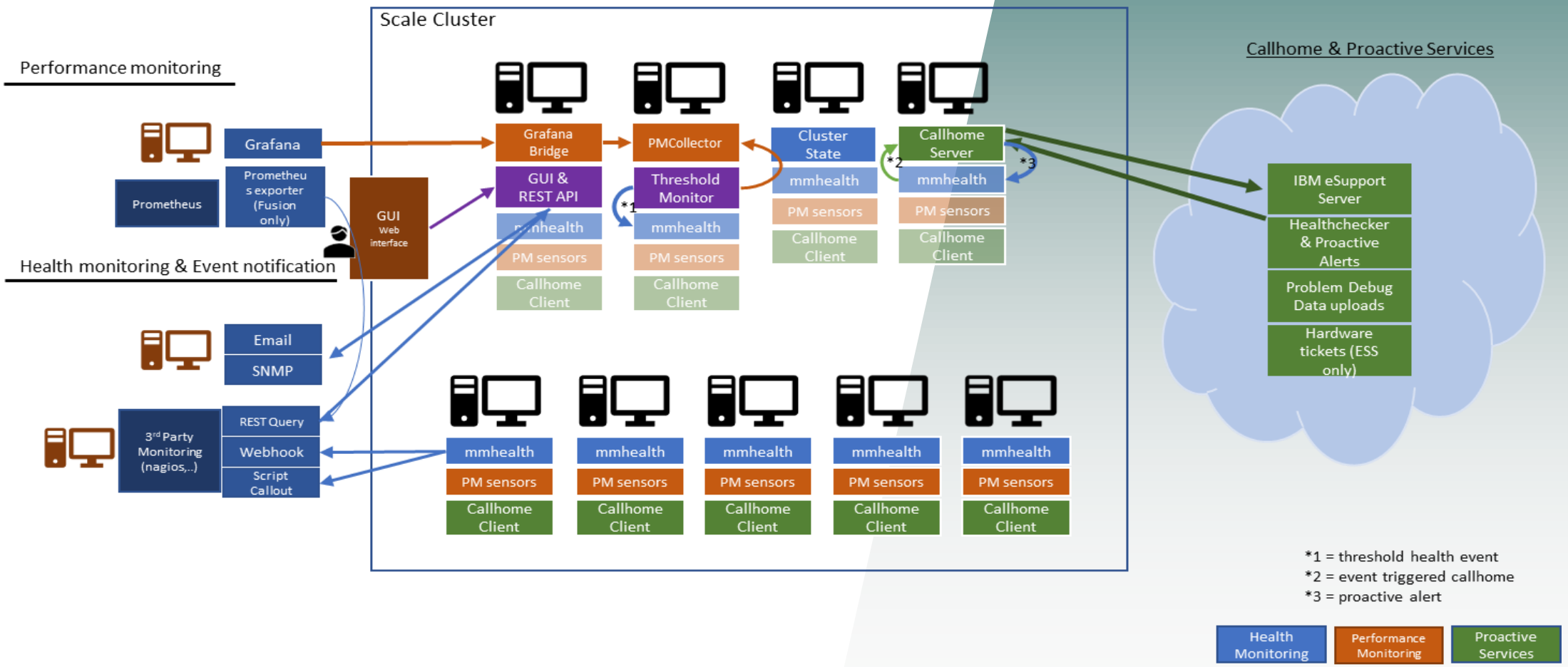


- A Kubernetes Controller is a special software that, in the loop, responds to changes and performs adaptation actions in the cluster
- A controller tracks at least one Kubernetes resource type
- Kubernetes has a set of built-in controllers that run inside the kube-controller-manager

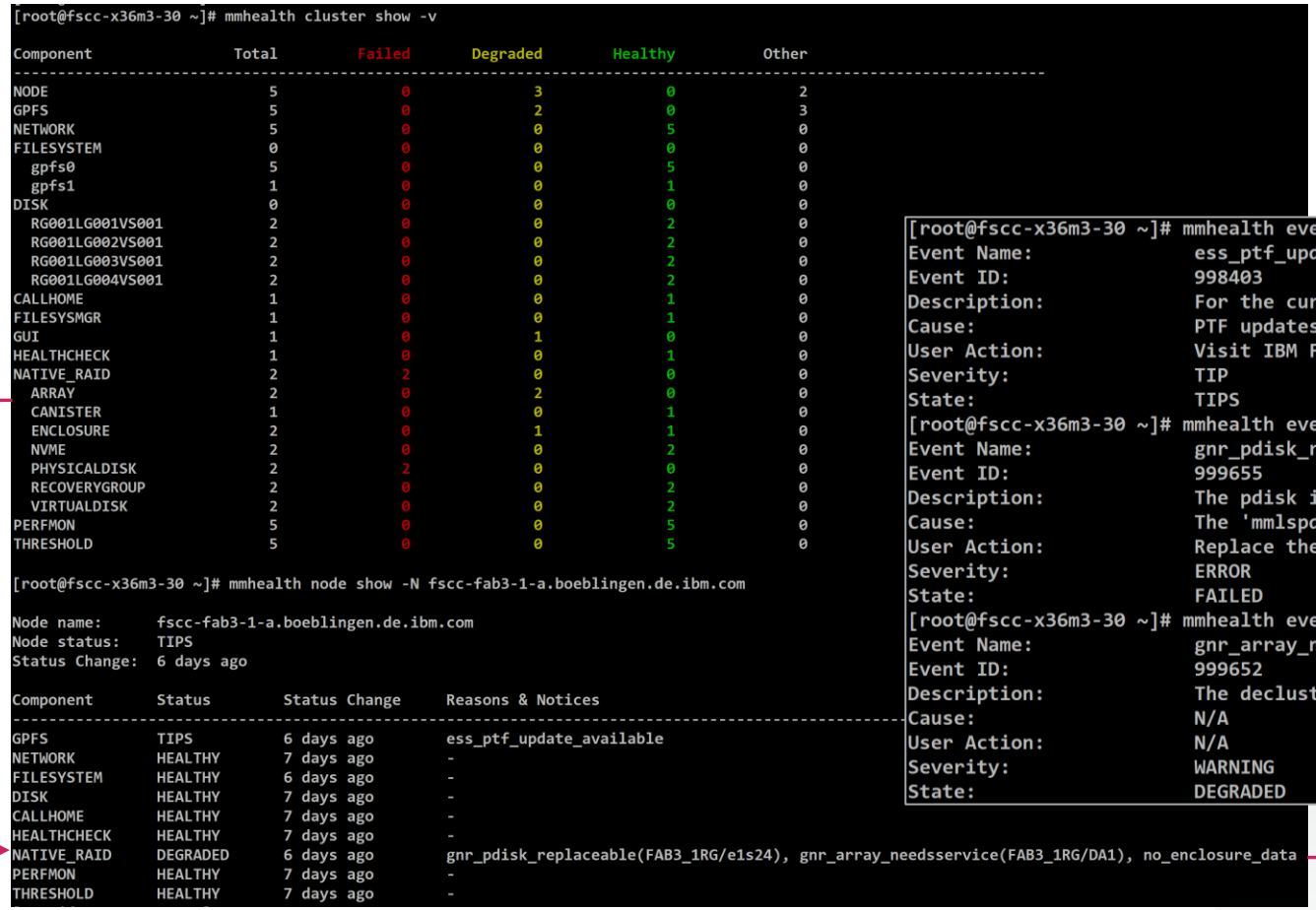
"Operators are software extensions to Kubernetes that make use of custom resources to manage applications and their components. Operators follow Kubernetes principles, notably the control loop"
Kubernetes Documentation

Classic Scale Monitoring Overview

IBM Storage



Health Reporting with MMHEALTH



```
[root@fscc-x36m3-30 ~]# mmhealth cluster show -v

Component               Total         Failed       Degraded        Healthy          Other
-------------------------------------------------------------------------------------------------------------
NODE                        5              0              3              0              2
GPFS                        5              0              2              0              3
NETWORK                     5              0              0              5              0
FILESYSTEM                  0              0              0              0              0
  gpfs0                     5              0              0              5              0
  gpfs1                     1              0              0              1              0
DISK                        0              0              0              0              0
  RG001LG001VS001           2              0              0              2              0
  RG001LG002VS001           2              0              0              2              0
  RG001LG003VS001           2              0              0              2              0
  RG001LG004VS001           2              0              0              2              0
CALLHOME                    1              0              0              1              0
FILESYSMGR                  1              0              0              1              0
GUI                         1              0              1              0              0
HEALTHCHECK                 1              0              0              1              0
NATIVE_RAID                 2              2              0              0              0
  ARRAY                     2              0              2              0              0
  CANISTER                  1              0              0              1              0
  ENCLOSURE                 2              0              1              1              0
  NVME                      2              0              0              2              0
  PHYSICALDISK              2              2              0              0              0
  RECOVERYGROUP             2              0              0              2              0
  VIRTUALDISK               2              0              0              2              0
PERFMON                     5              0              0              5              0
THRESHOLD                   5              0              0              5              0

[root@fscc-x36m3-30 ~]# mmhealth node show -N fscc-fab3-1-a.boeblingen.de.ibm.com

Node name:      fscc-fab3-1-a.boeblingen.de.ibm.com
Node status:    TIPS
Status Change:  6 days ago

Component       Status        Status Change     Reasons & Notices
-------------------------------------------------------------------------------------------------------------
GPFS            TIPS          6 days ago        ess_ptf_update_available
NETWORK         HEALTHY       7 days ago        -
FILESYSTEM      HEALTHY       6 days ago        -
DISK            HEALTHY       7 days ago        -
CALLHOME        HEALTHY       7 days ago        -
HEALTHCHECK     HEALTHY       7 days ago        -
NATIVE_RAID     DEGRADED      6 days ago        gnr_pdisk_replaceable(FAB3_1RG/e1s24), gnr_array_needsservice(FAB3_1RG/DA1), no_enclosure_data
PERFMON         HEALTHY       7 days ago        -
THRESHOLD       HEALTHY       7 days ago        -
```

```
[root@fscc-x36m3-30 ~]# mmhealth event show ess_ptf_update_available
Event Name:              ess_ptf_update_available
Event ID:                998403
Description:             For the currently installed IBM ESS packages a PTF update is available.
Cause:                   PTF updates are available.
User Action:             Visit IBM Fix Central to download and install the updates.
Severity:                TIP
State:                   TIPS
[root@fscc-x36m3-30 ~]# mmhealth event show gnr_pdisk_replaceable
Event Name:              gnr_pdisk_replaceable
Event ID:                999655
Description:             The pdisk is ready to be replaced, which means that all the data is drained out of the disk.
Cause:                   The 'mmlspdisk' command shows replaceable user condition for the disk.
User Action:             Replace the pdisk.
Severity:                ERROR
State:                   FAILED
[root@fscc-x36m3-30 ~]# mmhealth event show gnr_array_needsservice
Event Name:              gnr_array_needsservice
Event ID:                999652
Description:             The declustered array state needs service.
Cause:                   N/A
User Action:             N/A
Severity:                WARNING
State:                   DEGRADED
```

https://www.ibm.com/docs/en/spectrum-scale/5.1.7?topic=monitoring-system-health-by-using-mmhealth-command
https://www.ibm.com/docs/en/spectrum-scale/5.1.7?topic=events-monitoring-mmhealth

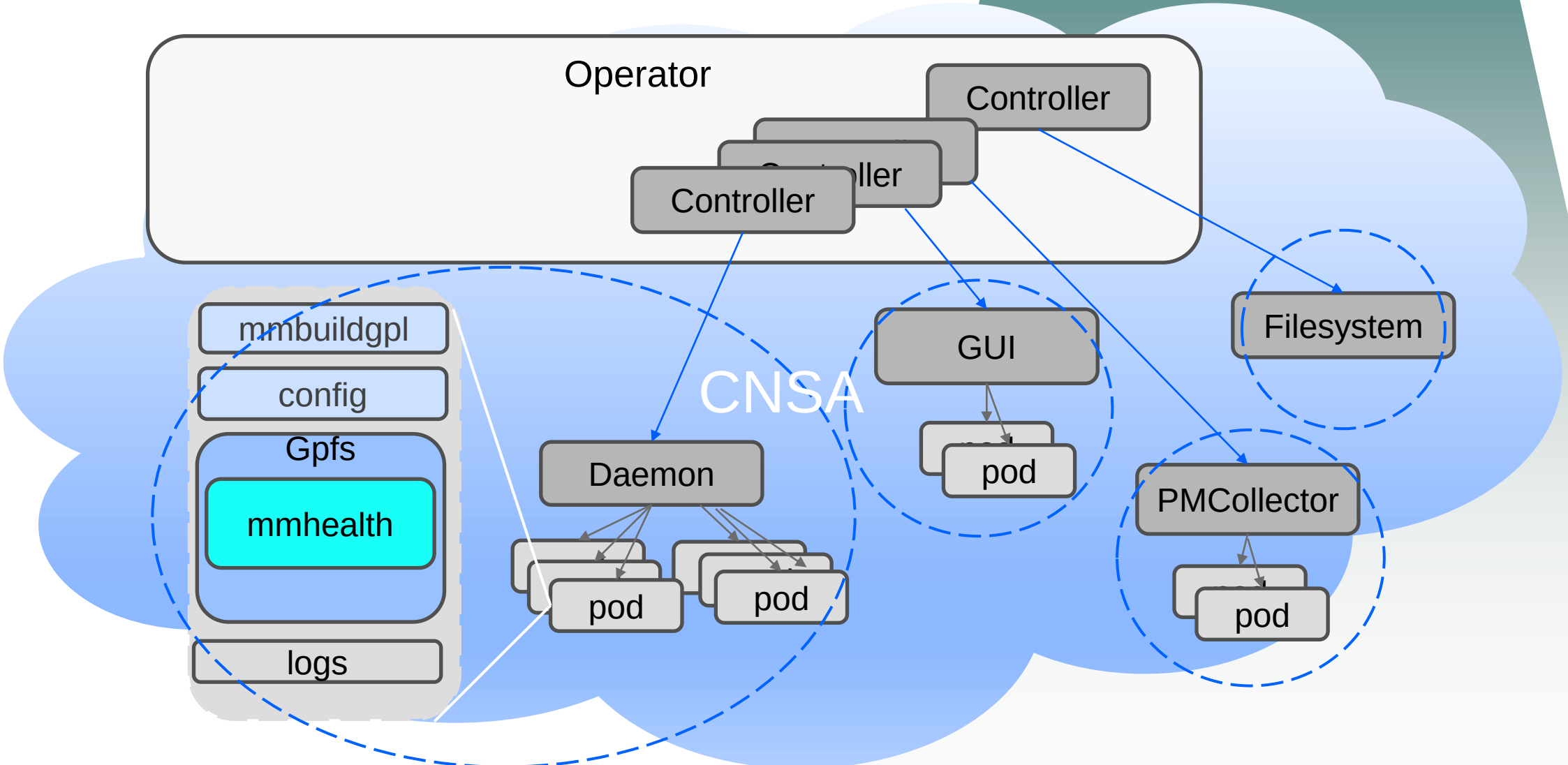
MMHEALTH - CNSA objects mapping

MMHEALTH Component	CNSA Custom Resource (CR)
Filesystem	**Filesystem**
GPFS	**Daemon**
Perfmon	**PMCollector**
GUI	**GUI**
Native Raid, Network, Disk	Tbd. <future release>

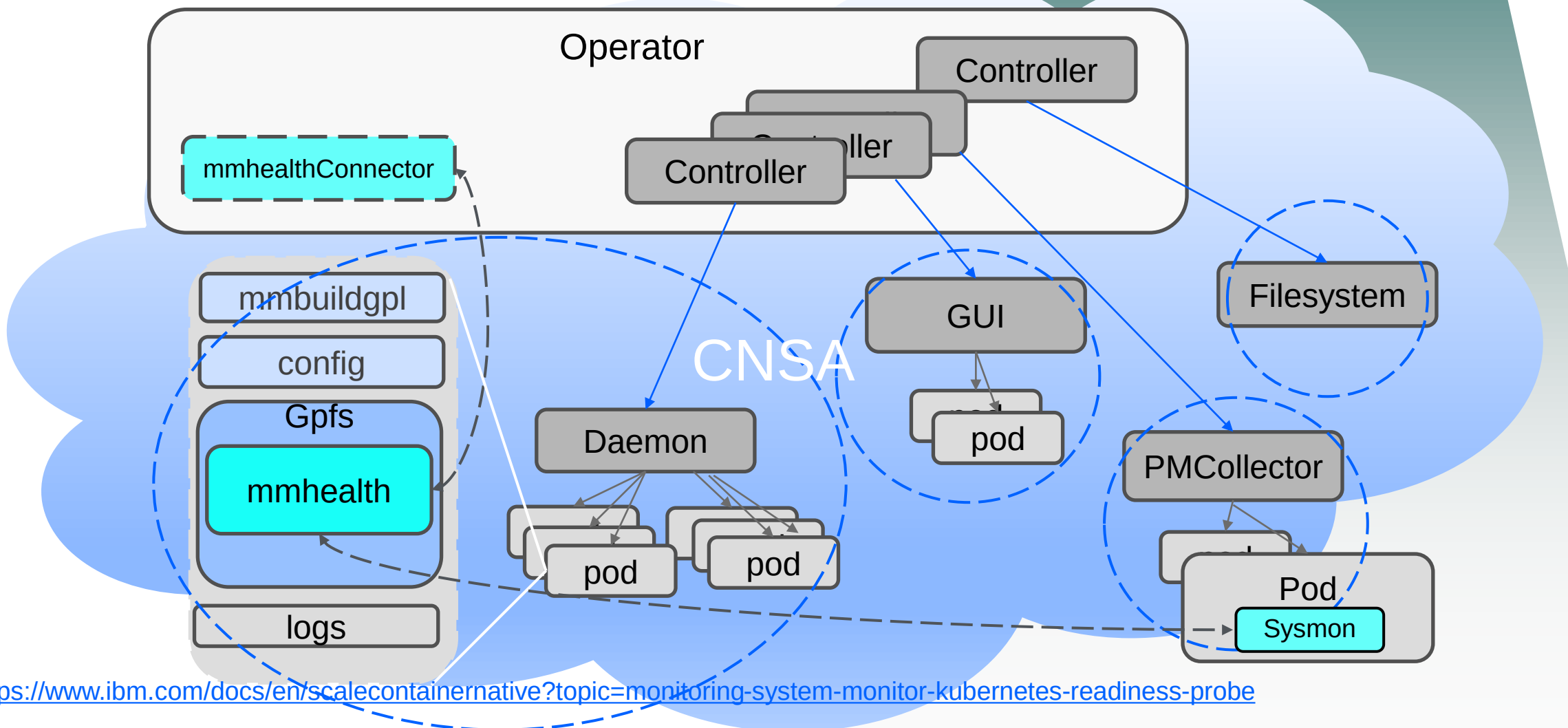
afm, array, auth, auth_obj, block, callhome, canister/server, ces, cesip, cesnetwork, cloudgateway, disk, enclosure, encryption, fileauditlog, filesysmgr, filesystem, gds, gpfs, gui, hadoopconnector, hdfs_datanode, hdfs_namenode, healthcheck, localcache, msgqueue, native_raid, network, nfs, nfspod, node, noobaa, nvme, nvmeof, object, perfmon, physicaldisk, powerhw, recoverygroup, scalemgmt, serverraid, smb, stretchcluster, threshold, virtualdisk, watchfolder

Callhome, CloudCSIDisk, Cluster, CompressionJob, Daemon, DiskJob, DNSConfig, DNS, EncryptionConfig, Filesystem, GrafanaBridge, Gui, Job, LocalDisk, Pmcollector, RecoveryGroup, RemoteCluster, RestripeFSJob, StretchClusterInitNodes, StretchCluster, StretchClusterTiebreaker, UpgradeApproval

CNSA Health Status reporting flow



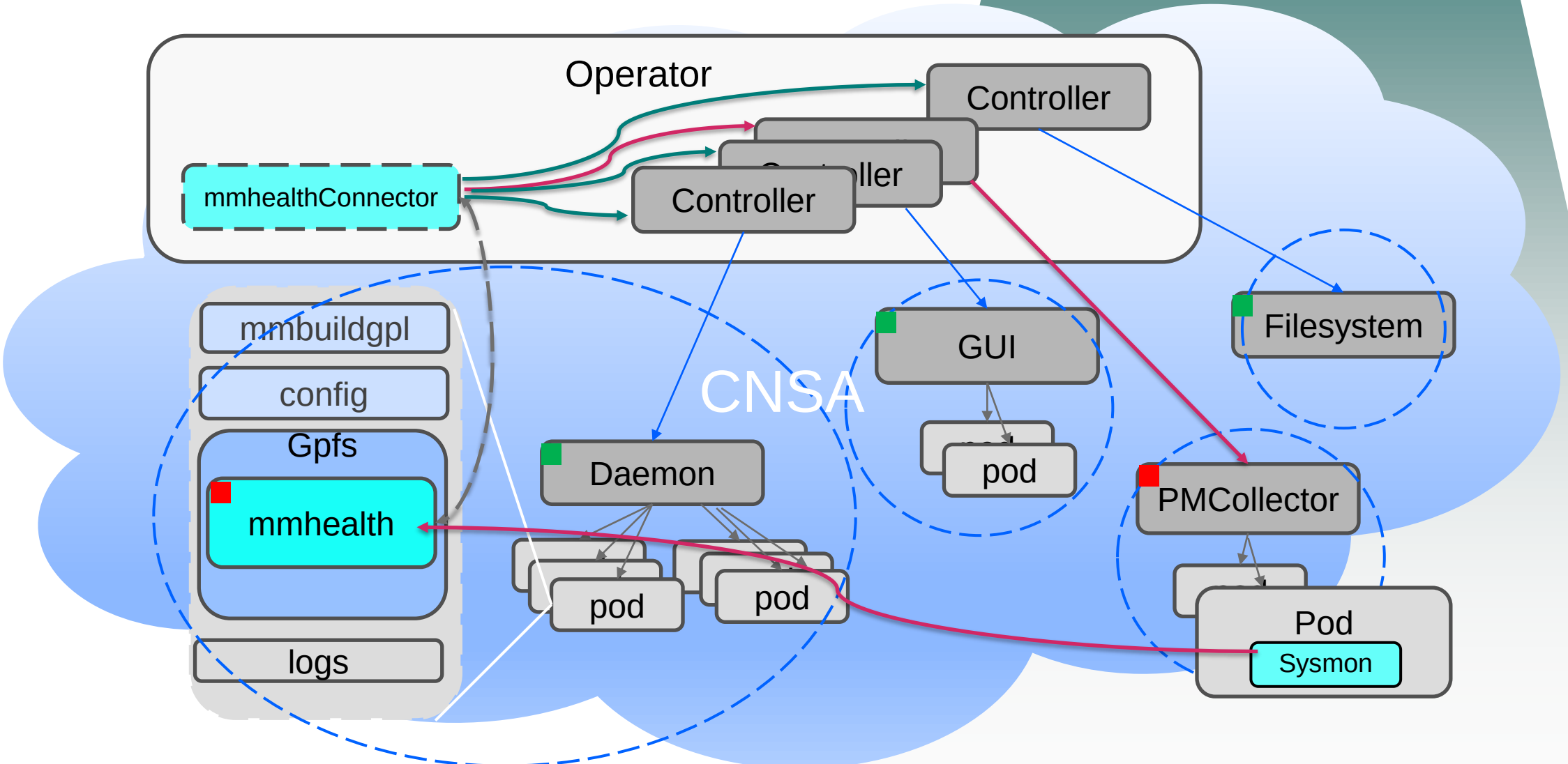
CNSA Health Status reporting flow



https://www.ibm.com/docs/en/scalecontainernative?topic=monitoring-system-monitor-kubernetes-readiness-probe

CNSA Health Status reporting flow

IBM Storage



Health Status reprsentation in CNSA

Kubernetes API conventions

Every Kubernetes object includes two nested object :

- Spec - a complete description of the **desired state**
- Status - the **current state** of the object in the system

Status Condition

- contains **details** for one aspect of the current state of the resource

Events

- provide **additional information** for debugging and tracing, akin to log messages

CNSA implementation

„Healthy“ Condition added to the status of a most important CNSA resources: Daemon, Filesystem, Pmcollector, GUI

MMHEALTH status is the source of the Healthy condition

MMHEALTH events passed through and published to a resource

Healthy Condition

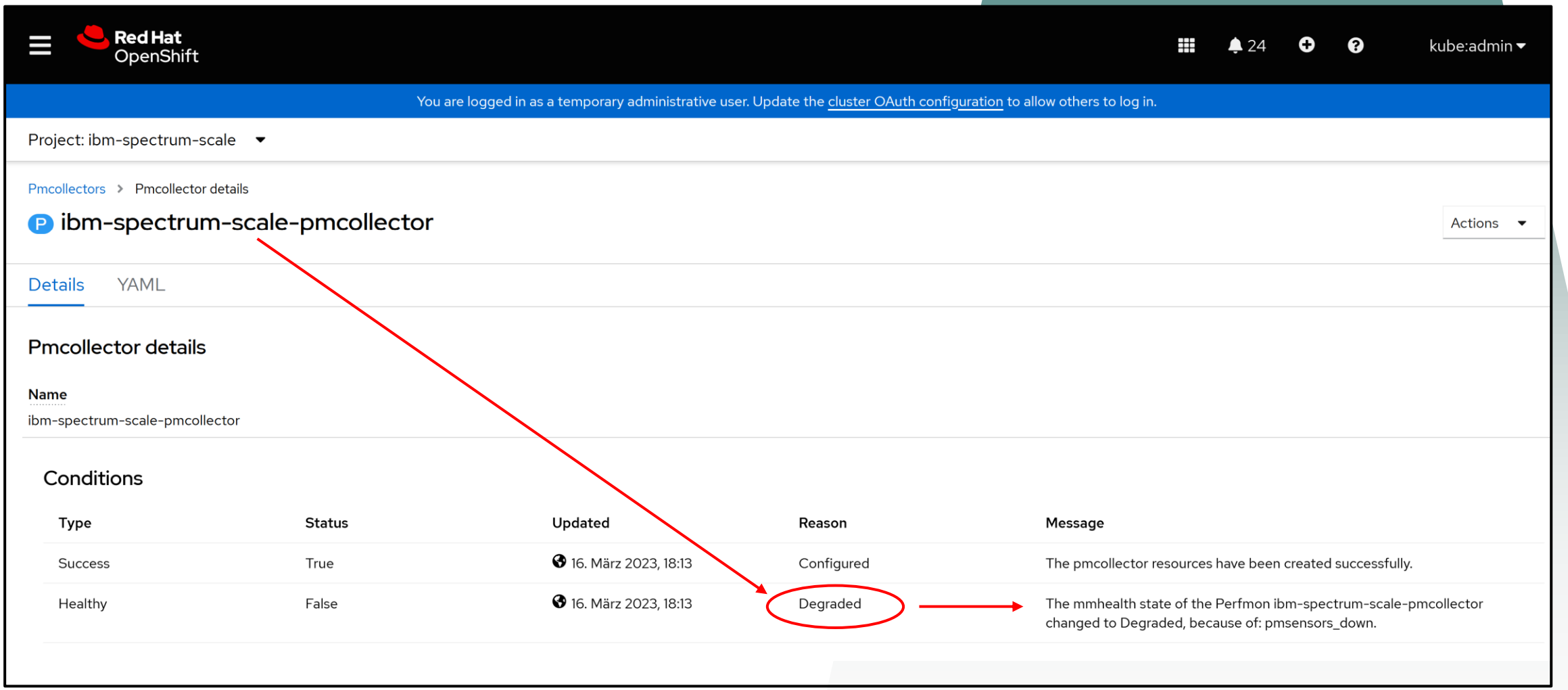
- Type - *Healthy*
- Status - *True*, *False*, or *Unknown*
- Reason - Health status obtained by mmhealth
- Message - short message including 3 most important happenings causing the last status transition

Health Status reprsentation in CNSA


```
[root@ocp1-helper ~]#
[root@ocp1-helper ~]# oc get Pmcollector ibm-spectrum-scale-pmcollector -o json |jq '.status.conditions[]'
{
  "lastTransitionTime": "2023-03-16T17:13:31Z",
  "message": "The pmcollector resources have been created successfully.",
  "reason": "Configured",
  "status": "True",
  "type": "Success"
}
{
  "lastTransitionTime": "2023-03-16T17:13:32Z",
  "message": "The mmhealth state of the Perfmon ibm-spectrum-scale-pmcollector changed to Degraded, because of: pmsensors_down.",
  "reason": "Degraded",
  "status": "False",
  "type": "Healthy"
}
[root@ocp1-helper ~]# oc get event --namespace ibm-spectrum-scale --field-selector involvedObject.name=ibm-spectrum-scale-pmcollector,involvedObje
ct.kind=Pmcollector
LAST SEEN   TYPE      REASON           OBJECT                                  MESSAGE
5m17s       Warning   pmsensors_down   pmcollector/ibm-spectrum-scale-pmcollector   The pmsensors service should be 0, but is 1. (worker0.daemon.i
bm-spectrum-scale.stg.ocp1.vmlocal.)
[root@ocp1-helper ~]#
[root@ocp1-helper ~]#
[root@ocp1-helper ~]#
```

Health Status reprsentation in CNSA

IBM Storage



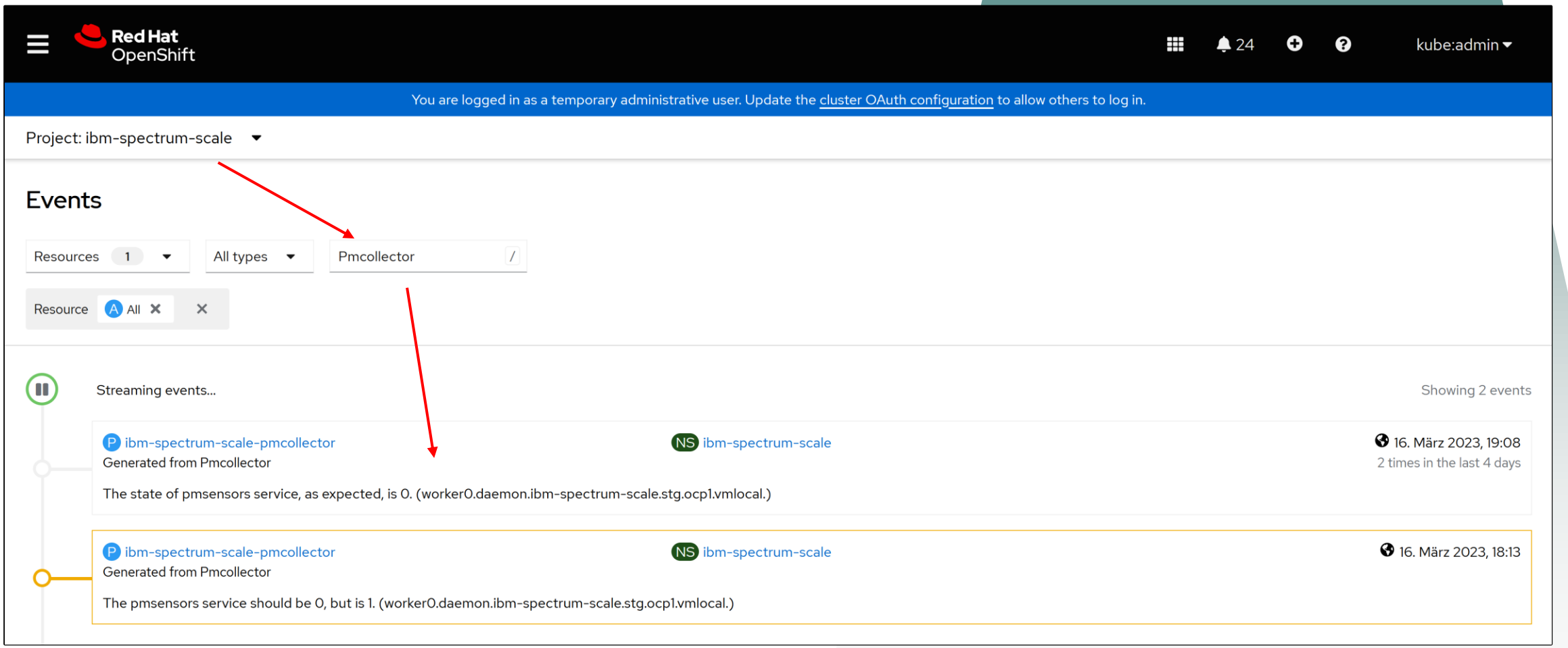
Health Status reprsentation in CNSA


```
[root@ocp1-helper ~]# oc exec worker0 -c gpfs -n ibm-spectrum-scale -- /opt/IBM/zimon/sbin/pmsensors -C /etc/scale-pmsensors-configuration/ZIMonSensors.cfg -R /var/run/perfmon
SensorFactory: CPU registered
SensorFactory: Memory registered
SensorFactory: Network registered
SensorFactory: TopProc registered
SensorFactory: Netstat registered
SensorFactory: Diskstat registered
```

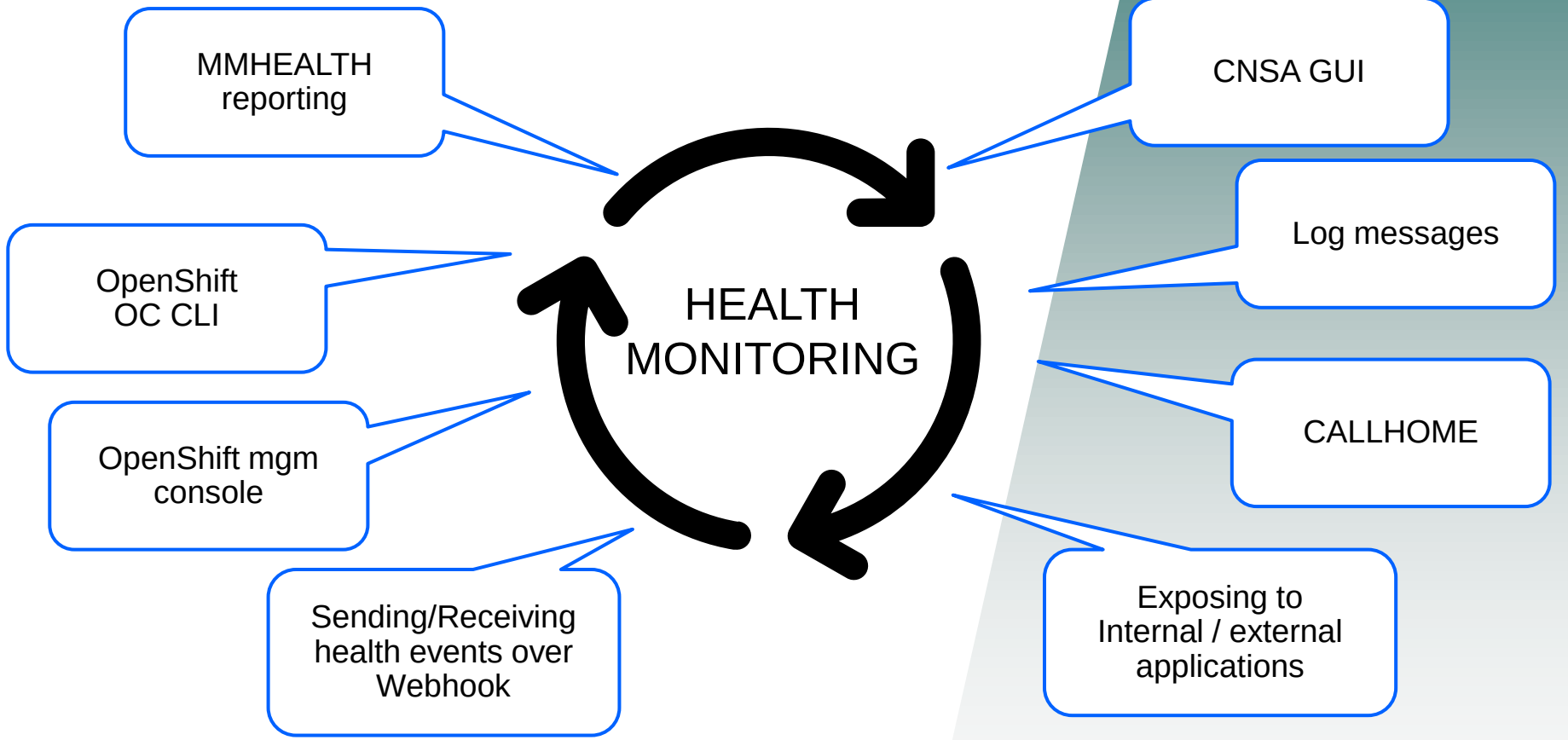
```
Mar-16 18:07:42  [Info   ]: Successfully read configuration from file /etc/scale-pmsensors-configuration/ZIMonSensors.cfg
[root@ocp1-helper ~]#
[root@ocp1-helper ~]# oc get Pmcollector ibm-spectrum-scale-pmcollector -o json |jq '.status.conditions[]'
{
  "lastTransitionTime": "2023-03-16T18:08:06Z",
  "message": "The pmcollector resources have been created successfully.",
  "reason": "Configured",
  "status": "True",
  "type": "Success"
}
{
  "lastTransitionTime": "2023-03-16T18:08:07Z",
  "message": "The mmhealth state of the Perfmon ibm-spectrum-scale-pmcollector changed to Healthy.",
  "reason": "Healthy",
  "status": "True",
  "type": "Healthy"
}
[root@ocp1-helper ~]# oc get event --namespace ibm-spectrum-scale --field-selector involvedObject.name=ibm-spectrum-scale-pmcollector,involvedObject.kind=Pmcollector
LAST SEEN   TYPE      REASON           OBJECT                                      MESSAGE
77s         Normal    pmsensors_up     pmcollector/ibm-spectrum-scale-pmcollector   The state of pmsensors service, as expected, is 0. (worker0.daemon.ibm-spectrum-scale.stg.ocp1.vmlocal.)
55m         Warning   pmsensors_down   pmcollector/ibm-spectrum-scale-pmcollector   The pmsensors service should be 0, but is 1. (worker0.daemon.ibm-spectrum-scale.stg.ocp1.vmlocal.)
[root@ocp1-helper ~]#
[root@ocp1-helper ~]#
```

Health Status reprsentation in CNSA

IBM Storage



Consistent Scale Status Reporting



https://www.ibm.com/docs/en/spectrum-scale/5.1.7?topic=monitoring
https://www.ibm.com/docs/en/scalecontainernative?topic=517-monitoring

Thank you for using
IBM Spectrum Scale!