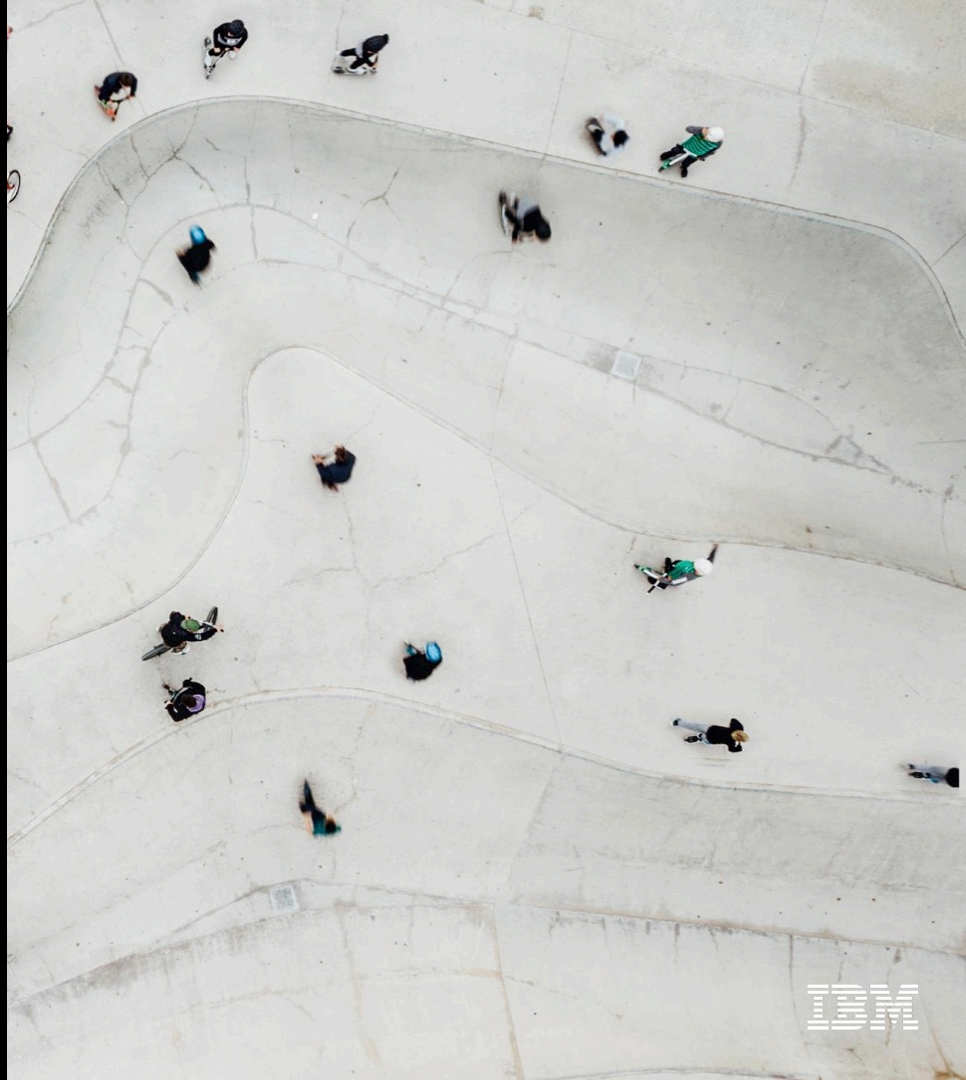
Clustered Watch Folder
—
Spectrum Scale 5.0.3



Clustered Watch Folder

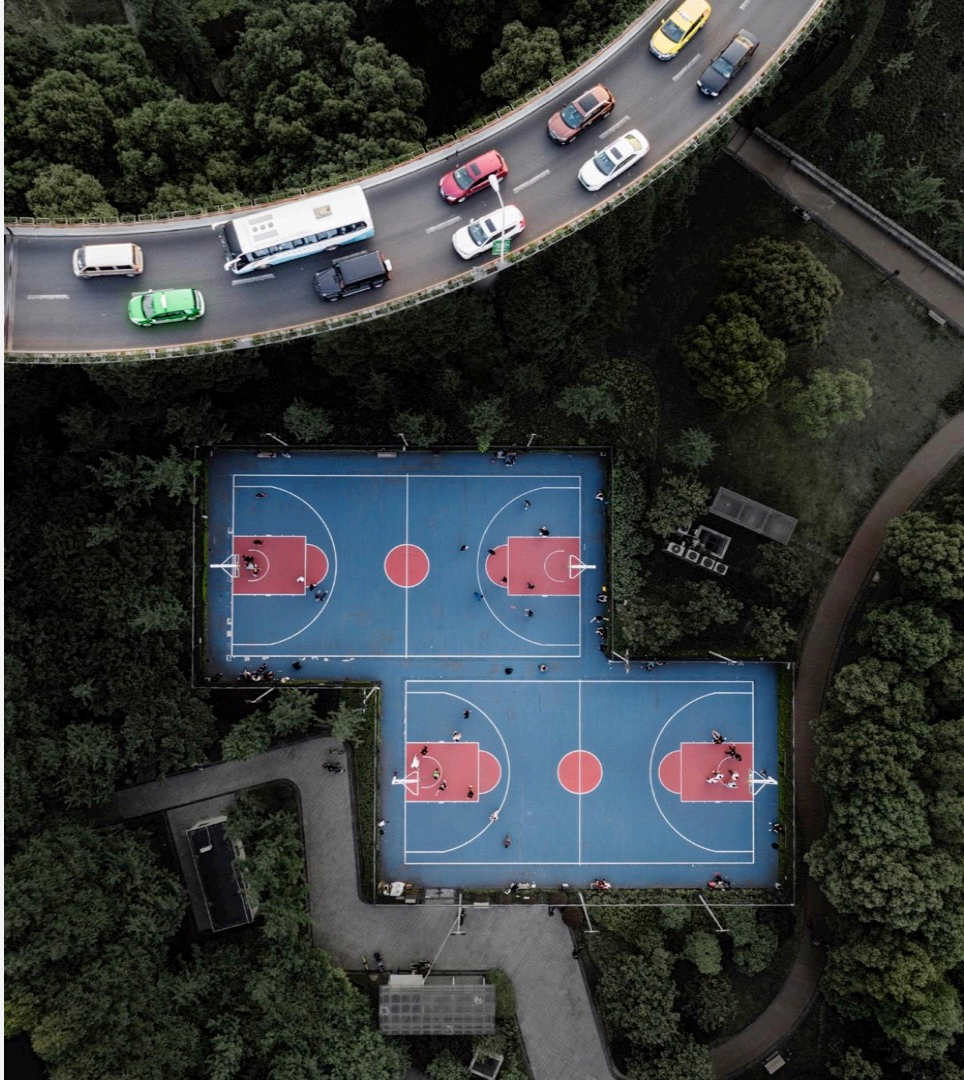
Distributed, Scalable, Multi-Cluster File System Event Notification

Setup and Watch File System accesses across Clusters with a Single Command

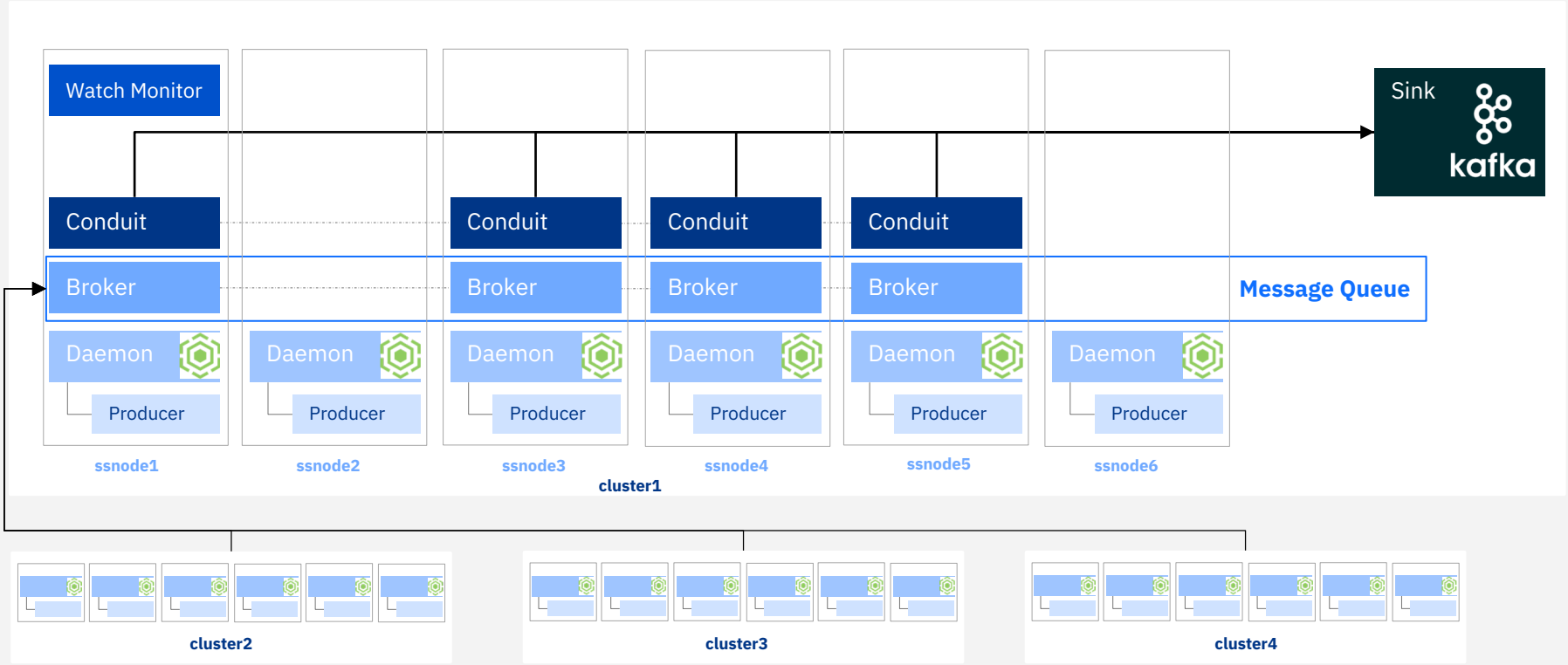
Rich Event Metadata

Take action, Respond to notifications:

- Automate Data Workflows
- Curate, Index, Discover
- Track Data Provenance and Lineage



Architecture



Spectrum Scale policies used to configure the daemons as to which accesses/files need to be watched.

Events you can Watch

- IN_OPEN *
- IN_ACCESS
- IN_MODIFY
- IN_ATTRIB *
- IN_CLOSE_WRITE *
- IN_CLOSE_NOWRITE
- IN_CREATE *
- IN_DELETE *
- IN_MOVED_FROM
- IN_MOVED_TO
- IN_DELETE_SELF
- IN_MOVE_SELF *
- IN_IGNORED **
- IN_UNMOUNT **

* IN_ISDIR also when operation concerns directory

** Cannot configure, but will be reported

http://man7.org/linux/man-pages/man7/inotify.7.html


```
{ "WF_JSON": "0.0.1",
  "wd":"1",
  "cookie":"0",
  "event": "IN_CREATE",
  "path": "/gpfs/gpfs1/DSTtest",
  "clusterName": "solar.gpfs.ibm.net",
  "nodeName": "node1.ibm.com",
  "nfsClientIp": "",
  "fsName": "gpfs1",
  "inode": "134333440",
  "fileSetID": "0",
  "linkCount": "1",
  "openFlags": "0",
  "poolName": "system",
  "fileSize": "0",
  "ownerUserId": "0",
  "ownerGroupId": "0",
  "atime": "2019-03-21_18:14:05-0500",
  "ctime": "2019-03-21_18:14:05-0500",
  "mtime": "2019-03-21_18:14:05-0500",
  "eventTime": "2019-03-21_18:14:05-0500",
  "clientUserId": "0",
  "clientGroupId": "0",
  "processId": "30432",
  "permissions": "200100644",
  "acls": null,
  "xattrs": null,
  "subEvent": "NONE"
}
```

Watch Folder JSON Fields

Field	Description
WF_JSON	Version.
wd	Watch descriptor
cookie	A unique integer that connects related events. It allows the resulting pair of IN_MOVED_FROM and IN_MOVED_TO events to be connected.
event	-
path	-
clusterName	-
nodeName	-
nfsClientIP	-
fsName	-
inode	-
filesetID	-
linkCount	-
openFlags	The open flags that are specified during the event (O_RDONLY, O_WRONLY, O_RDWR, O_CREAT, etc.) as defined in fcntl.h.

Field	Description
poolName	-
fileSize	-
ownerUserID	-
ownerGroupID	-
atime	-
ctime	-
mtime	-
eventTime	Time when the event was enqueued to send to message queue.
clientUserID	-
clientGroupID	-
processID	-
permissions	-
acls	-
subEvent	futures

Requirements

RPMs and packages

- GPFS™ Java™ (gpfs.java rpm/package)
- For RHEL, the librdkafka package requires the openssl-devel and cyrus-sasl-devel packages
- For Ubuntu, the librdkafka package requires the libssl-dev and libsasl2-dev packages
- librdkafka (gpfs.librdkafka rpm/package)
- Kafka (gpfs.kafka rpm/package)

OS and Hardware

- Only Linux (RHEL and Ubuntu.
- Linux Kernel > 3.10.0-123.
- Zookeeper: Minimum of three Linux quorum nodes running on approved OS and hardware
- Brokers: Minimum of three nodes to act as message queue servers (brokers) running on approved OS and hardware.

Security and Limitations

- SELinux in enforcing mode is not supported.
- Root authority is required to run **mmmsgqueue** and **mmwatch**.
- The following TCP ports must be open on all nodes in the cluster:
  - 2181, 9092, and 9093 along with the range 2888:3888

File System and Misc.

- Spectrum Scale code 5.0.3
- File system version 5.0.2
- Available in the IBM Spectrum Scale Advanced Edition or IBM Spectrum Scale Data Management Edition only.

mmwatch

mmwatch *Device* **enable** { **-F** *ConfigFilePath* |

[**--fileset** *fsetname*] [**--watch-id** *WatchID*]

[**--events** {*Event*[,*Event*...] | **ALL**}]

--event-handler *handlertype*

--sink-brokers *BrokerIP:Port*[,*BrokerIP:Port*...]

--sink-topic *Topic*

[**--sink-auth-config** *Path*]

[**--degraded**] }

mmwatch *Device* **list** [**--events**] [**--watch-id** *WatchID*] [**-Y**]
mmwatch *Device* **list --watch-id** *WatchID* **--config** [**-Y**]
mmwatch **all list** [**--events**] [**-Y**]
mmwatch **all status**
mmwatch *Device* **status** [**--watch-id** *WatchID* [**-v**]]

Sink Authentication

- SELinux in enforcing mode is not supported.
- Root authority is required to run **mmmsgqueue** and **mmwatch**.
- The following TCP ports must be open on all nodes in the cluster:
  - 2181, 9092, and 9093 along with the range 2888:3888

Enable a watch

```
[root@cluster1 ~]# mmwatch fs0 enable -F myconf.conf
[I] Beginning enablement of Clustered Watch with newly created watch ID: CLW1553290256
[I] Verifying MsgQueue nodes meet minimum local space requirements for Clustered Watch to be enabled for watch: CLW1553290256
    Depending on cluster size, this may take some time
[I] Successfully verified all configured MsgQueue nodes meet minimum local space requirements for Clustered Watch to be enabled for watch:
CLW1553290256
[I] Verified the watch type is FSYS for filesystem fs0
[I] Successfully created Clustered Watch topic on the MsgQueue for watch: CLW1553290256
[I] Successfully added Clustered Watch configuration file into CCR for watch: CLW1553290256
[I] Successfully enabled Clustered Watch consumers for watch: CLW1553290256
[I] Successfully added Clustered Watch policy rules for watch: CLW1553290256
[I] Successfully enabled Clustered Watch: CLW1553290256
```

```
# cat myconf.conf
 EVENT_HANDLER:kafkasink
 SINK_BROKERS:node1.ibm.com:9092,node2.ibm.com:9092,node3.ibm.com:9092
 SINK_TOPIC:myExternalSinkTopic
 SINK_AUTH_CONFIG:/home/secrets.txt
 DEGRADED:false
```

```
# cat /home/secrets.txt
 SINK_AUTH_TYPE:CERT
 CA_CERT_LOCATION:/opt/kafka/SSL/ca-cert
 CLIENT_PEM_CERT_LOCATION:/opt/kafka/SSL/clients/client.pem
 CLIENT_KEY_FILE_LOCATION:/opt/kafka/SSL/clients/client.key
 CLIENT_KEY_FILE_PASSWORD:DST4lyfE
```

Sink Authentication: NONE, PLAIN, SCRAM512, CERT

List Watches

```
node1:# mmwatch all list
Filesystem earthfs has no watchers.
Filesystem mercury423 is at an unsupported filesystem level for watch folder.
Filesystem venus5011 is at an unsupported filesystem level for watch folder.
Filesystem marsfs has no watchers.
Filesystem newfs has no watchers.
Filesystem jupiterfs is at an unsupported filesystem level for watch folder.
Filesystem saturnfs has 1 watcher(s):
 Watcher          WatchID          Type          Start Time                                  Path
-------------------------------------------------------------------------------------------------------------------
solar.cluster     CLW1553311967    filesystem    Tue Mar 19 15:17:24 2019                  /saturnfs
```

```
node1 # mmwatch saturnfs list --watch-id CLW1553311967
 Watcher          WatchID          Type          Start Time                                  Path
 ------------------------------------------------------------------------------------------------------------------
 solar.cluster     CLW1553311967   filesystem     Tue Mar 19 15:22:38 2019                  /saturnfs
```

```
node1:# mmwatch all status
Device          WatchID                Watch State
-------------------------------------------------------------------------------------------------------------------
saturnfs        CLW1553311967            Active
      Node Name
      solar1.ibm.com                  HEALTHY
      solar2.ibm.com                  HEALTHY
      solar3.ibm.com                  HEALTHY
      solar4.ibm.com                  HEALTHY
      solar5.ibm.com                  HEALTHY
      solar6.ibm.com                  HEALTHY
      solar7.ibm.com                  HEALTHY
```

List Watch Details

```
node1: # mmwatch saturnfs list --watch-id CLW1553311967 --events
 Watcher           WatchID           Monitored Events
-------------------------------------------------------------------------------
solar.cluster     CLW1553311967      Path: /saturnfs
                                           IN_ACCESS,IN_ATTRIB,IN_CLOSE_NOWRITE,
                                           IN_CLOSE_WRITE,IN_CREATE,IN_DELETE,
                                           IN_DELETE_SELF,IN_MODIFY,IN_MOVED_FROM,
                                           IN_MOVED_TO,IN_MOVE_SELF,IN_OPEN
```

```
node1: # mmwatch saturnfs list --watch-id CLW1553311967 --config
DEVICE:saturnfs
WATCH_ID: CLW1553311967
EVENTS:ALL
EVENT_HANDLER:kafkasink
SINK_BROKERS:node1.ibm.com:9092,node2.ibm.com:9092,node3.ibm.com:9092
SINK_TOPIC: myExternalSinkTopic
SINK_AUTH_TYPE:SCRAM512
```

Logs

- `/var/adm/ras/mmmsgqueue.log`
  - Operations concerning the message queue
  - On any node containing a broker and/or zookeeper

- `/var/adm/ras/mmwatch.log`
  - Operations concerning Clustered watch
  - On a node where you run `mmwatch`

- `/var/adm/ras/mmwfclient.log`
  - Conduit logs
  - On a node that also houses the message queue broker

- `/var/adm/ras/tswatchmonitor.log`
  - Watch monitor logs
  - On a node that houses the watch monitor

- `/var/adm/ras/mmfs.log.latest`
  - Daemon log, also daemon producer logs
  - /var/log/messages (Redhat) (OR) /var/log/syslog (Ubuntu)

- **All logs collected in gpfs.snap**

Failures

- Node failure
- Daemon failure
- File system unmount
- Message queue failure
- External Sink failure
- Notification failure

Troubleshooting

Authentication failures when connecting to Kafka Sink

/var/adm/ras/mmwfclient.log:

```
2019-03-22_15:17:52: [I] WF Producer: wt: SpectrumScale_WF_C_8286659982002334447_FSYS_CLW1553311967_CLW_saturnfs t: myExternalSinkTopic a: 1
< WFProducer.configureKafkaProducer() enter. d 0 a  > b: node1.ibm.com:9092,node2.ibm.com:9092,node3.ibm.com:9092
2019-03-22_15:17:52: [I] WF Producer: wt: SpectrumScale_WF_C_8286659982002334447_FSYS_CLW1553311967_CLW_saturnfs t: myExternalSinkTopic a: 1
< Successfully created! > b: node1.ibm.com:9092,node2.ibm.com:9092,node3.ibm.com:9092
2019-03-22_15:17:52: [I] WF Producer: wt: SpectrumScale_WF_C_8286659982002334447_FSYS_CLW1553311967_CLW_saturnfs t: myExternalSinkTopic a: 1
< WFProducer:configureKafkaProducer() exit r 0 s HEALTHY. > b: node1.ibm.com:9092,node2.ibm.com:9092,node3.ibm.com:9092
[E] Producer for brokers: localhost:9092 topic: myExternalSinkTopic is unable to authenticate itself. This producer is
    shutting down.
Kafka Error: -169: Local: Authentication failure
[I] Producer state change (HEALTHY --> DOWN).
```

Troubleshooting

Brokers down on external Kafka sink:

/var/adm/ras/mmwatch.log

```
2019-03-22_22:33:29: [E] tswatchgpfsconduit: < Conduit could not process messages successfully. >
2019-03-22_22:33:29: [I] tswatchgpfsconduit: < Executing command: /usr/lpp/mmfs/bin/mmsysmonc event watchfolder watchconduit_err
gpfs1/8286659982002334447/CLW1553311967 WFProducer "WF Kafka Producer Error" >
2019-03-22_22:33:29.880-0500:1:36437:disableConduitOnNode:437: [I] Received disable conduit consumer request from MsgQueue services
monitor for node node1.ibm.com, check the /var/adm/ras/tswatchmonitor.log on the node running the msgqueueservicesmonitor for more information.
2019-03-22_22:33:39.049-0500:1:36788:conduitDownOnNode:666: [E] Disabling clustered watch because number of good consumers:
2 is less than number of required good consumers: 4
2019-03-22_22:33:39.273-0500:1:36788:conduitDownOnNode:683: [E] Disabling clustered watch for device: gpfs1 with mmwatch ID: CLW1553311967
2019-03-22_22:33:49.289-0500:1:36788:disableClusterWatch:5046: [I] Successfully disabled Clustered Watch consumers for watch: CLW1553311967
```

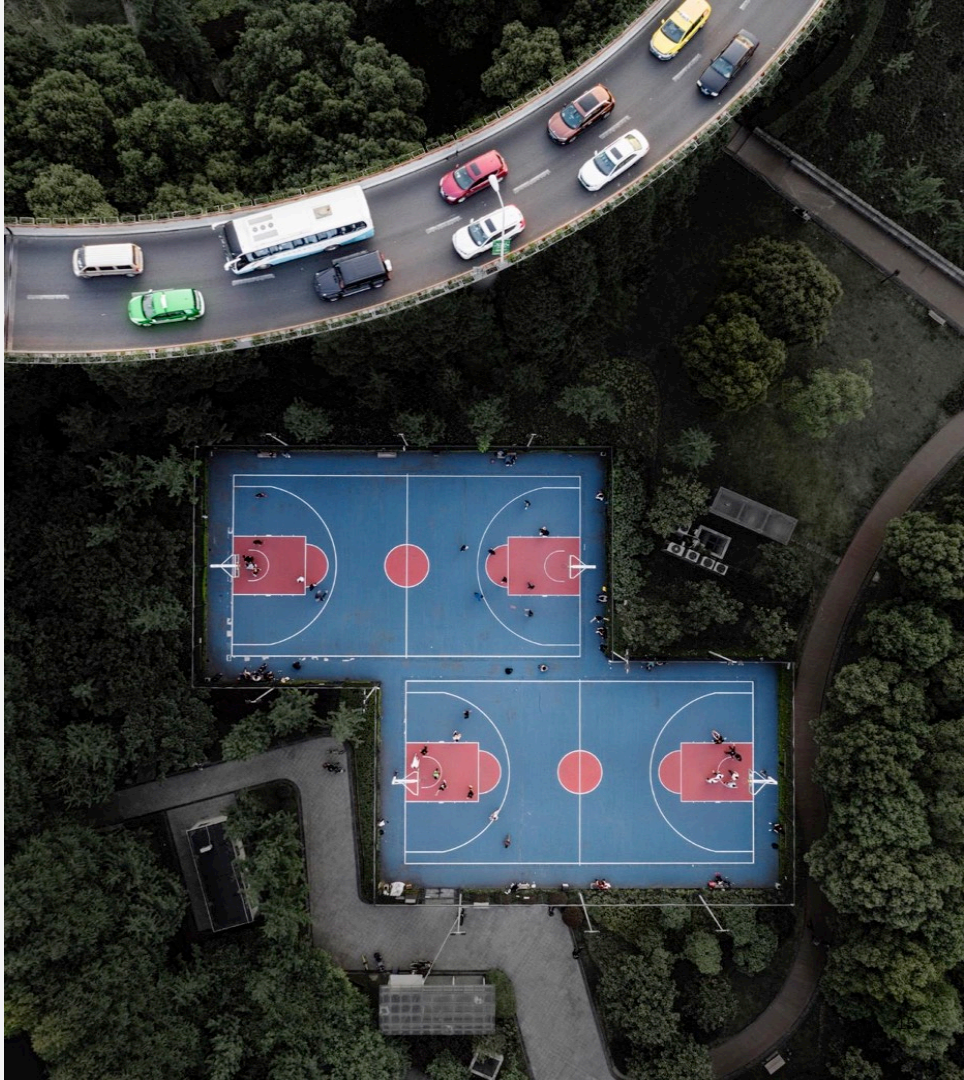
/var/adm/ras/mmwfclient.log

```
2019-03-22_22:33:29: [W] WF Producer: wt: SpectrumScale_WF_C_8286659982002334447_FSYS_CLW1553311967_CLW_saturnfs t: myExternalSinkTopic
a: 3 < Kafka Producer state change (HEALTHY --> ERROR). Kafka Error: -187: Local: All broker connections are down:
Reason: 1/1 brokers are down. > b: node1.ibm.com:9092,node2.ibm.com:9092,node3.ibm.com:9092
2019-03-22_22:33:30: [I] WF Producer: wt: SpectrumScale_WF_C_8286659982002334447_FSYS_CLW1553311967_CLW_gpfs1 t: myExternalSinkTopic, a: 3, < WFProducer
shut down. State is now: DOWN> b: node1.ibm.com:9092,node2.ibm.com:9092,node3.ibm.com:9092
```

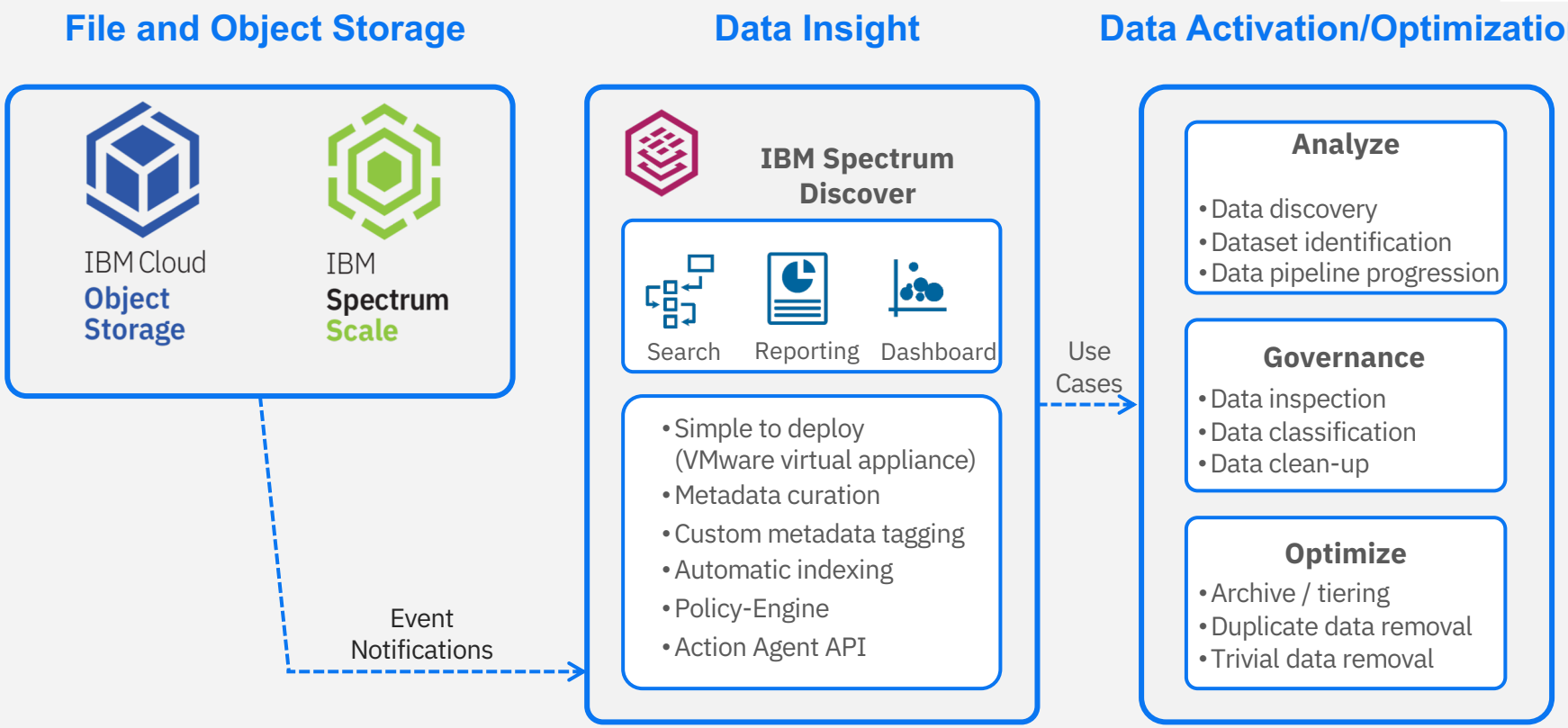
```
node1:# mmwatch all status
Device          WatchID                 Watch State
-------------------------------------------------------------------------------------------------------------------------------------------
saturnfs        CLW1553311967           Auto Disabled
      Node Name
      solar1.ibm.com                  DOWN
      solar2.ibm.com                  DOWN
      solar3.ibm.com                  DOWN
      solar4.ibm.com                  DOWN
      solar5.ibm.com                  DOWN
      solar6.ibm.com                  DOWN
      solar7.ibm.com                  DOWN
```

Use Cases

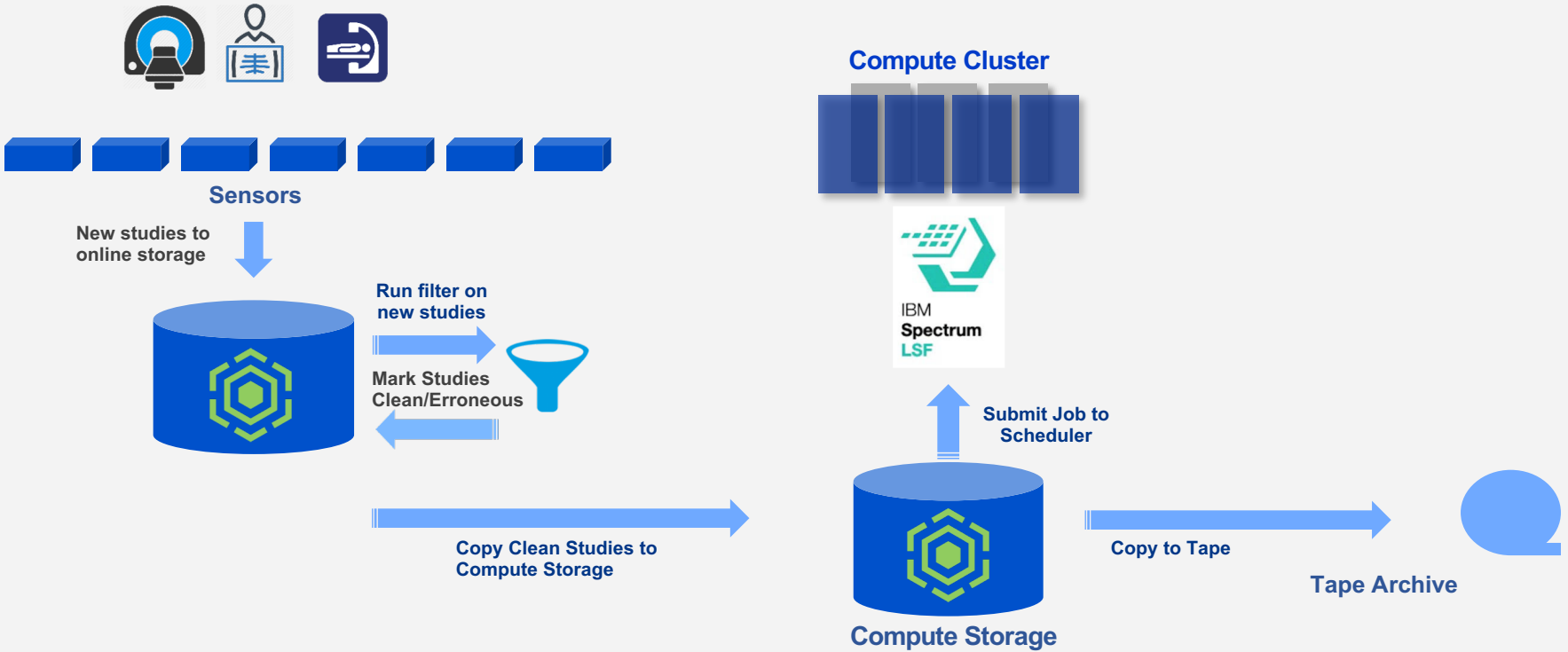
- File System Catalog for curating data – Spectrum Discover
- Workflow Automation
- File Sharing Portal



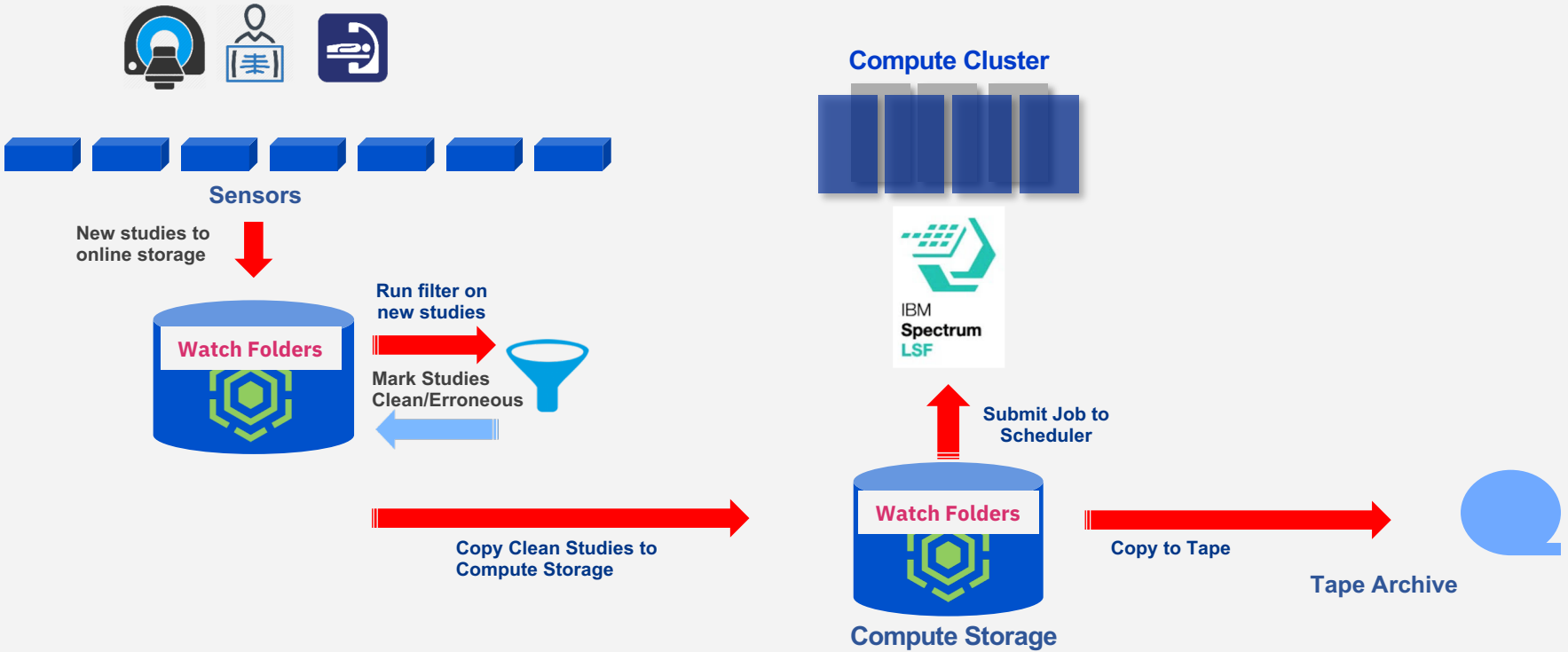
IBM Spectrum Discover



Workflow Automation (POC)



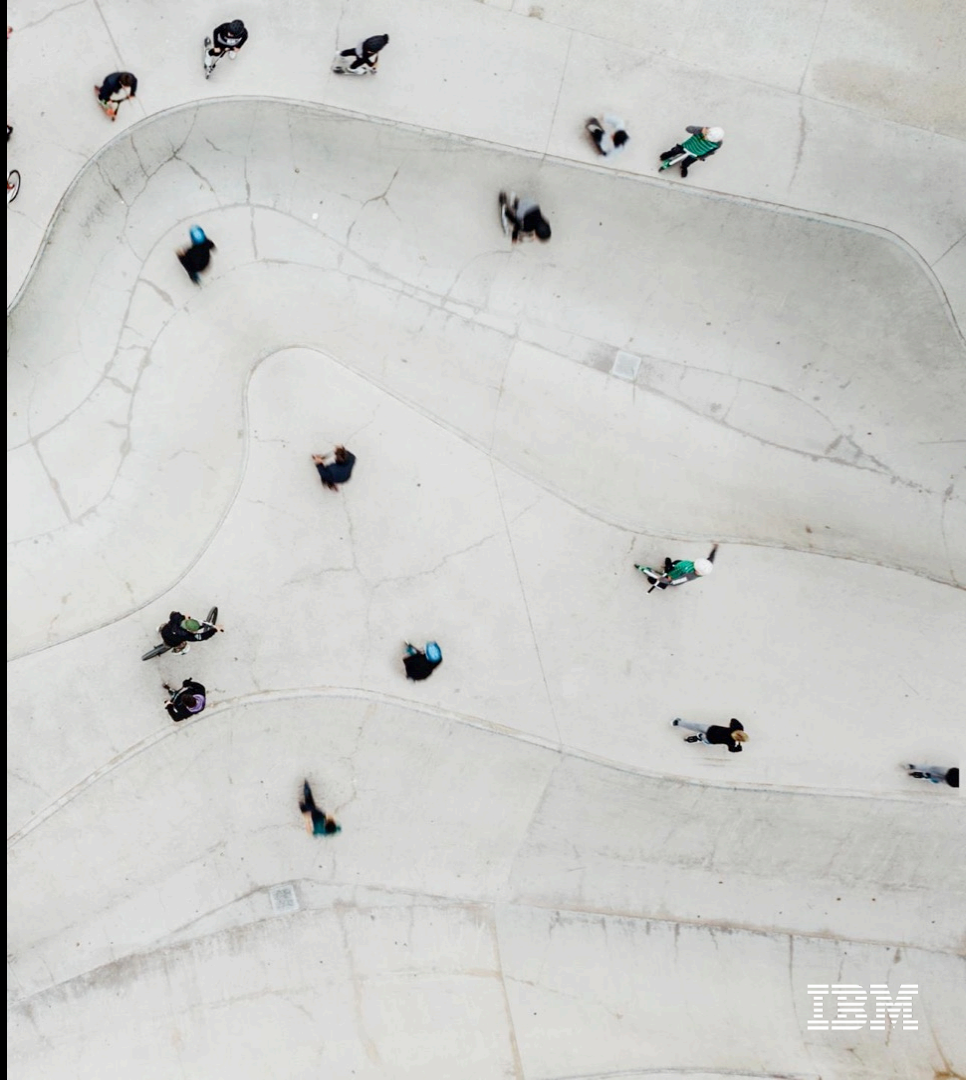
Workflow Automation



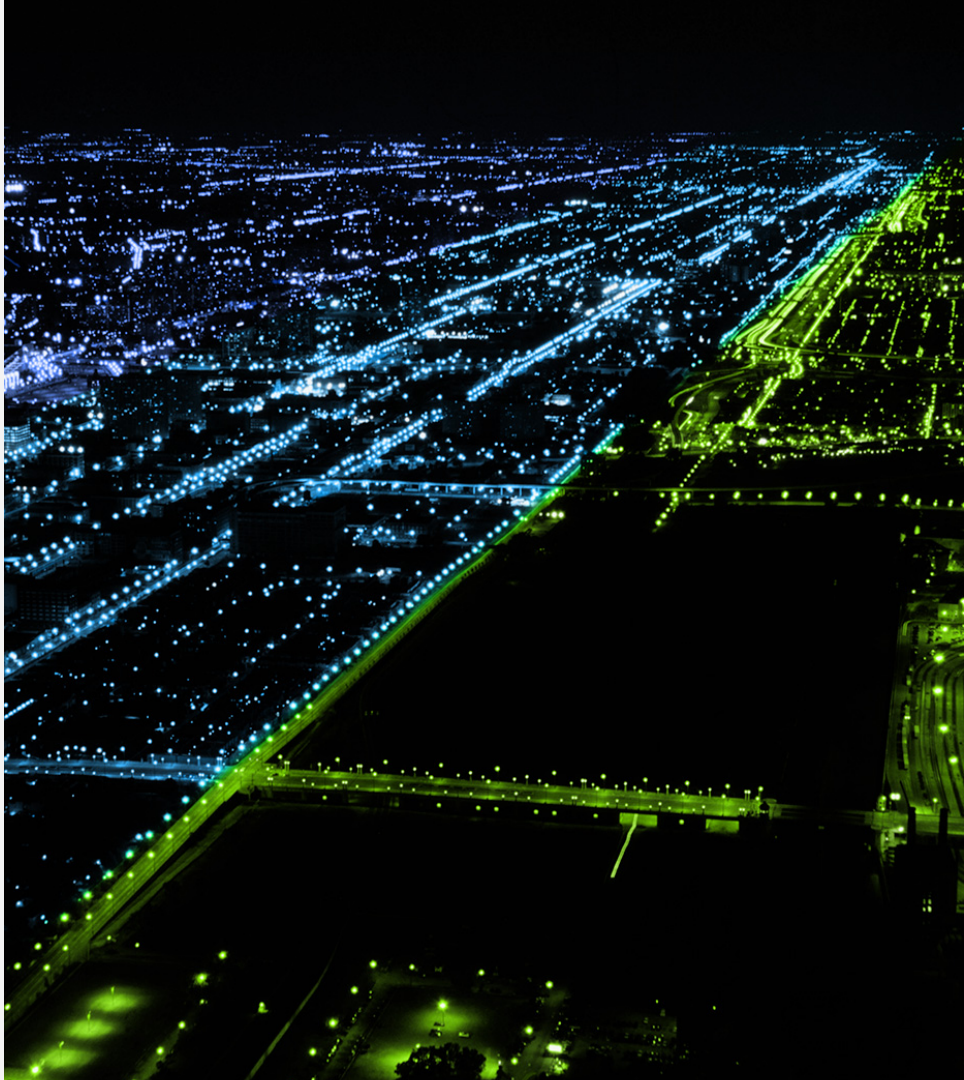
File Sharing Portal (POC)

Clustered Watch Folder

Deepavali Bhagwat
Lead Architect



Backup



Monitoring Watches using mmhealth

```
# mmhealth node show
Node name:      protocolnode1.ibm.com
Node status:    HEALTHY
Status Change:  1 day ago
Component       Status        Status Change     Reasons
------------------------------------------------------------------
GPFS            HEALTHY       1 day ago         -
NETWORK         HEALTHY       1 day ago         -
FILESYSTEM      HEALTHY       5 hours ago       -
CES             HEALTHY       1 day ago         -
GUI             HEALTHY       1 day ago         -
MSGQUEUE        HEALTHY       1 day ago         -
PERFMON         HEALTHY       1 day ago         -
THRESHOLD       HEALTHY       1 day ago         -
WATCHFOLDER     HEALTHY       1 day ago         -
```

```
# mmhealth cluster show
Component            Total        Failed       Degraded       Healthy          Other
-------------------------------------------------------------------------------------------------
NODE                 4            0            2              0                2
GPFS                 4            0            1              0                3
NETWORK              4            0            0              4                0
FILESYSTEM           3            0            0              3                0
DISK                 6            0            0              6                0
CES                  2            0            1              1                0
CESIP                1            0            0              1                0
GUI                  1            0            1              0                0
MSGQUEUE             4            0            1              3                0
PERFMON              4            0            0              4                0
THRESHOLD            4            0            0              4                0
WATCHFOLDER          4            0            0              4                0
```

```
# mmhealth node show watchfolder -v
Node name:      protocolnode1.ibm.com
Component                                   Status        Status Change            Reasons
-------------------------------------------------------------------------------------------------------------
WATCHFOLDER                                 HEALTHY       2019-03-19 16:50:57      -
gpfs0/13222185860284578504/CLW1553028625    HEALTHY       2019-03-19 16:50:57      -
Event                Parameter                      Severity   Active Since           Event Message
-------------------------------------------------------------------------------------------------------------
watchc_service_ok    gpfs0/13222185860284578504/    INFO       2019-03-19 16:50:57    Watchfolder consumer service
                     CLW1553028625                                                    for file system gpfs0
                                                                                      is running
watchconduit_ok      gpfs0/13222185860284578504/    INFO       2019-03-19 16:50:57    GPFS Watch Conduit for watch
                     CLW1553028625                                                    id gpfs0/
                                                                                      13222185860284578504/
                                                                                      CLW1553028625 is running
```

Difference between WF and Clustered Watch

Watch folder API

- Flexible API, requires C program
- Ability to customize ingestion of events and decide what to do with them
- Runs on a single node
- Stops when program stops or node reboots, etc.
- Output very similar to output of standard Linux inotify
- No replication of watch data within the message queue
- Does not check for required local disk space, but limits maximum overall watches and watches per filesystem
- Allows directory, inode space and fileset based watches

Clustered Watch

- mm command, no programming required
- Provides clustered output to an external Kafka cluster sink
- Runs on multiple nodes
- Remains resident over node reboots and intermittent failures
- Output is structured in highly parsable JSON format (similar to file audit logging)
- Maintains 3 copies of each message within the message queue
- Specifically checks for required local disk space on broker nodes to host each clustered watch
- Allows device, inode space and fileset based watches

Install and Setup with the Spectrum Scale installation toolkit

```
# /usr/lpp/mmfs/5.0.3.0/installer/spectrumscale watchfolder list
[ INFO  ] Watch folder is Disabled.

# /usr/lpp/mmfs/5.0.3.0/installer/spectrumscale watchfolder enable
[ INFO  ] Enabling watch folder in the cluster configuration file.
[ INFO  ] Tip :If watch folder is enabled for the cluster, the message queue will be enabled. If you have at least 3 protocol
nodes defined, you do not need to use the '-b' flag to designate a node as a broker node. Otherwise, you need to specify at least
3 nodes to have the Kafka packages installed on and act as broker nodes for the message queue. Use the following command to
designate a node as a broker node: ./spectrumscale node add node_name -b.

# /usr/lpp/mmfs/5.0.3.0/installer/spectrumscale node add protocolnode1.ibm.com -b
[ INFO  ] Setting protocolnode1.ibm.com as a Kafka broker node.
[ INFO  ] Configuration updated.

# /usr/lpp/mmfs/5.0.3.0/installer/spectrumscale node add protocolnode2.ibm.com -b
[ INFO  ] Setting protocolnode2.ibm.com as a Kafka broker node.
[ INFO  ] Configuration updated.

# /usr/lpp/mmfs/5.0.3.0/installer/spectrumscale watchfolder list
[ INFO  ] Watch folder is Enabled.
[ INFO  ] User has defined Node ['protocolnode1.ibm.com', 'protocolnode2.ibm.com', 'protocolnode3.ibm.com',
'protocolnode4.ibm.com', 'protocolnode5.ibm.com', 'protocolnode6.ibm.com', 'protocolnode7.ibm.com', 'protocolnode8.ibm.com'] as
broker node in the current configuration.
```

Clustered Watch Folder
—
Spectrum Scale 5.0.3