Table of Contents

Collecting Data for Troubleshooting

CPD platform related problems:

	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI

Connectivity related issues:

	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI (CCS)
Application logs from pod	oc cp <wdp-connect-connection pod>/logs ./<wdp-connect-connection pod>-logs oc cp <wdp-connect-connector pod>/logs ./<wdp-connect-connector pod>-logs

CPD UI related issues:

	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI
HAR files	Chrome: More Tools > Developer Tools > Network > Export HAR

Data Class related issues:

	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI (CCS, WKC, AE)
Application logs from swagger	https://<CPD_URL>/v2/data_profiles/api/explorer/#/Hummingbird%20tasks/getHbTaskLogs

Global Search related issues:

	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI (CCS, WKC)
Pod logs	catalog-api
Application logs from swagger	https://<CPD_URL>/v2/cams/explorer/#/Assets/retrieveAssets

Metadata Enrichment related issues:

	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI (CCS, WKC, AE)
Application logs from pod	oc cp <wdp-profiling pod>/logs ./<wdp-profiling pod>-logs oc cp <spark-hb-control-plane pod>/logs ./< spark-hb-control-plane pod>-logs
Job log from UI	Job log for MDE job & Screenshot of the job start time
Application logs from swagger	https://<CPD_URL>/v2/data_profiles/api/explorer/#/Hummingbird%20tasks/getHbTaskLogs

Metadata Enrichment related issues:

	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI (CCS, WKC)

WKC Reporting related issues:

	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI (CCS, WKC)
Reporting status from API	curl -i -k -H "content-type: application/json" -H "Authorization: bearer $Bearer_TOKEN" -X GET "https://$HOSTNAME/v3/reporting/heartbeat" curl -i -k -H "content-type: application/json" -H "Authorization: bearer $Bearer_TOKEN" -X GET "https://$HOSTNAME/v3/reporting/999/register" curl -i -k -H "content-type: application/json" -H "Authorization: bearer $Bearer_TOKEN" -X GET "https://$HOSTNAME/v3/reporting/bistatus?tenant_id=999&table_name=all"

OpenShift related issues:

	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI (CCS)
Must-gather from CLI	cc adm must-gather
Collect data for projects from CLI	oc adm inspect/namespace <CPD namespace> oc adm inspect/namespace <IBM CPD operator namespace>

OpenShift Data Fusion related issues:

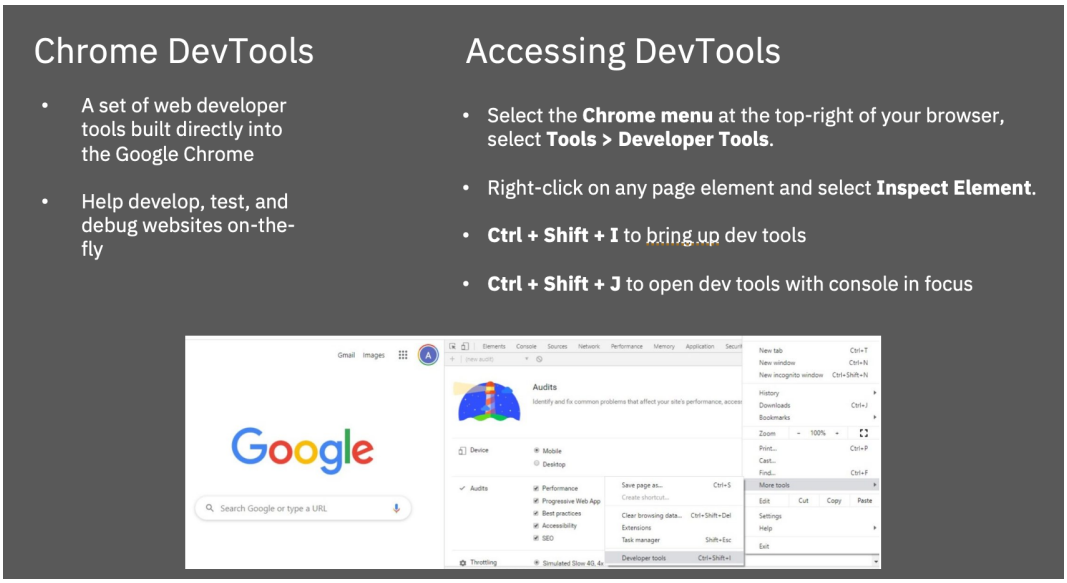
	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI (CCS)
ODF must-gather from CLI	oc adm must-gather --image=registry.redhat.io/odf4/ocs-must-gather-rhel8:v4.12 --dest-dir=/tmp/odf-must-gather or oc adm must-gather --image=docker-virtual.oneartifactoryci.verizon.com/odf4/ocs-must-gather-rhel8:v4.12

IBM Storage Fusion related issues:

	Data Needs to Capture
Diagnostic job	Gather diagnostics information from CPD UI (CCS)
Collect logs from Fusion console	Fusion Web Console > “Help icon” > Support logs > Collect logs.

Troubleshooting with Browser Developer Tools

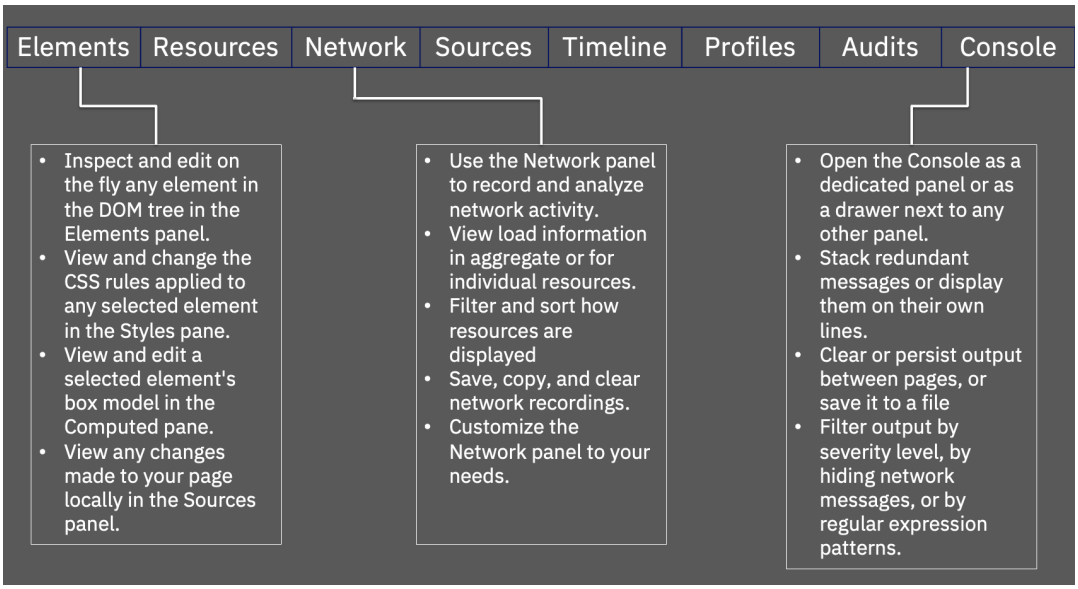
Debug with Chrome DevTools.



Mainly any API related problem can be investigate using the browser developer tool.

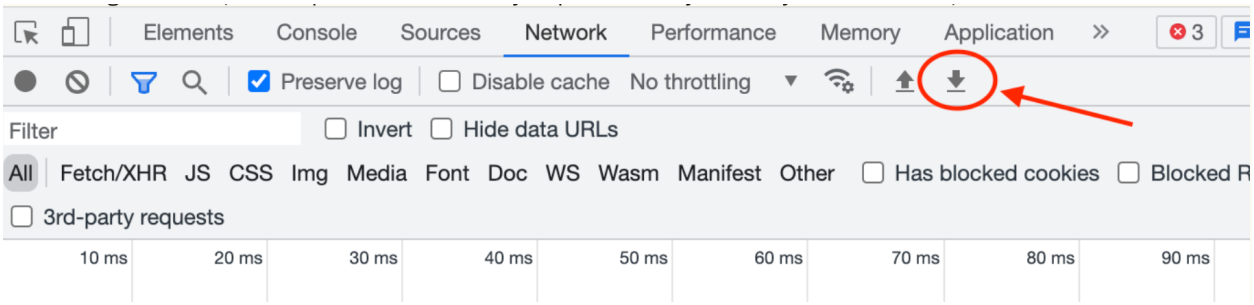
Each browser has their own way to use their developer tool. In this example we use Chrome as a browser.

A set of web developer tools built directly into the Google Chrome, that you can use for test, and debug issues on-the-fly.



There are different dev tools options to investigate different kind of problems.

Mainly the “Network” and “Console” panels are commonly used for investigating network activities, API call failure and error message retuned by a URL.



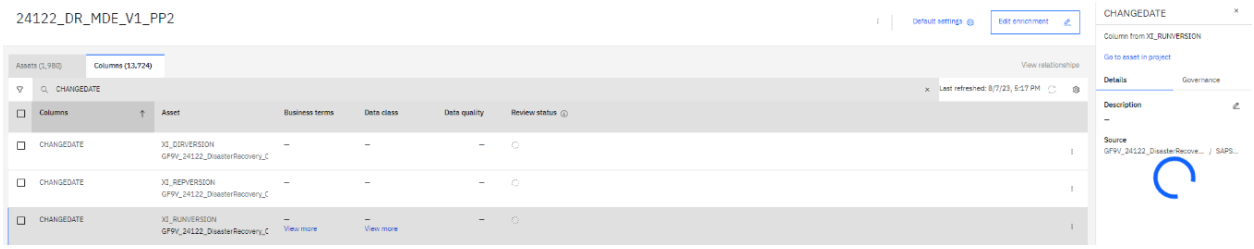
Using HAR Files to troubleshoot web pages that are failing to fully load.

- Track web browser requests
- Include response headers
- The body content
- Page load time.

A HAR (HTTP ARchive) file is a way to preserve and archive the interactions and exchanges between a web browser and a site. Among a lot of performance data and other factors that are generally only of interest to web developers, they also contain things like HTTP status codes for all of the transactions that go on behind the scenes whenever a given web page is loaded.

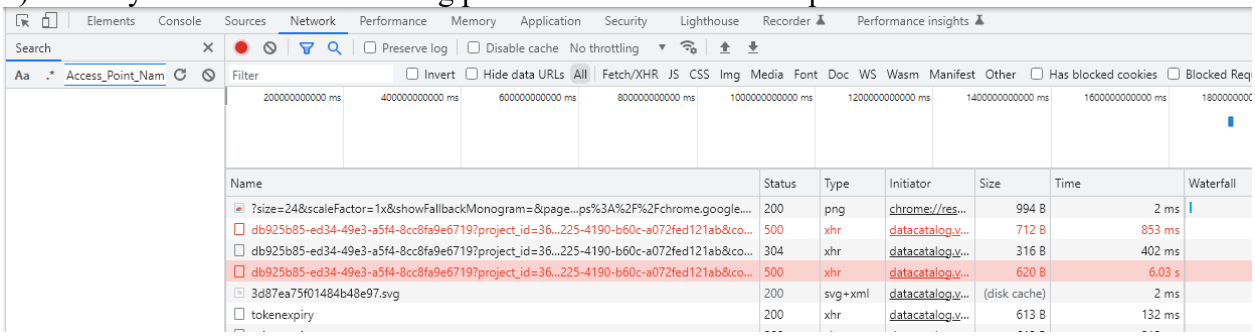
A HAR file can include sensitive details such as passwords, payment information, and private keys. There are open tools available to sanitize HAR file.

Accessing MDE column data taking time

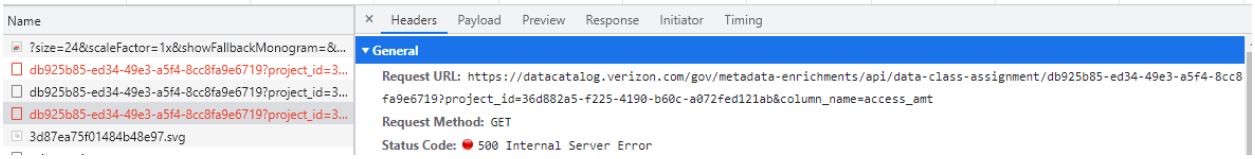


Troubleshooting:

1) For any CPD UI issue a starting point is the browser developer tool.



- Mainly the "Network" and "Console" panels are commonly used for investigating network activities, API call failure and error message retuned by a URL.
- Focused on the HTTP return/status code and elapsed time to find the problematic API. Compare elapsed time with previous successful run.



- The URL in the "Network > Headers" tab can help to identify the CPD service or component having problem.

Request URL: /gov/metadata-enrichments/api/data-class-assignment/

- Service: gov/WKC/IKC
- Component: metadata-enrichments/data-class-assignment

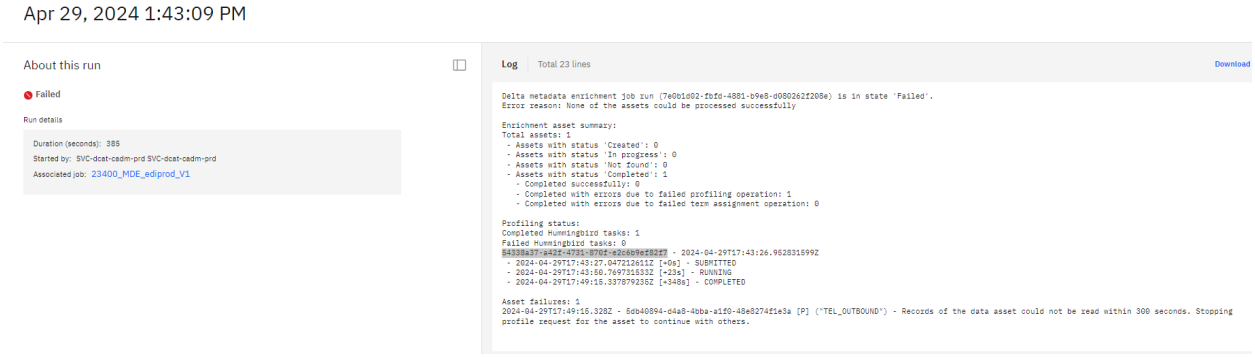
2) Check pod reference table[A6] to find pods associated to the Service/Component.
   - Are all pods healthy? [A1]
   - Any recent warning/fail event associated to the pod? [A2]
   - Has pod restarted recently? [A3]
   - Any error in the pod log during the problem? [A4]
   - What is memory, CPU consumption of pod? [A5]

MDE Publish job appearing hang

1. Check the job log for any error.
2. Monitor the job log for any progress in statistical values.
3. Check pod reference table[A6] to find pods associated to publishing
    a. Are all pods healthy? [A1]
    b. Any recent warning/fail event associated to the pod? [A2]
    c. Has pod restarted recently? [A3]
    d. Any error in the pod log during the problem? [A4]
    e. What is memory, CPU consumption of pod? [A5]

Find the SQL that causing problem or failing in the MDE job

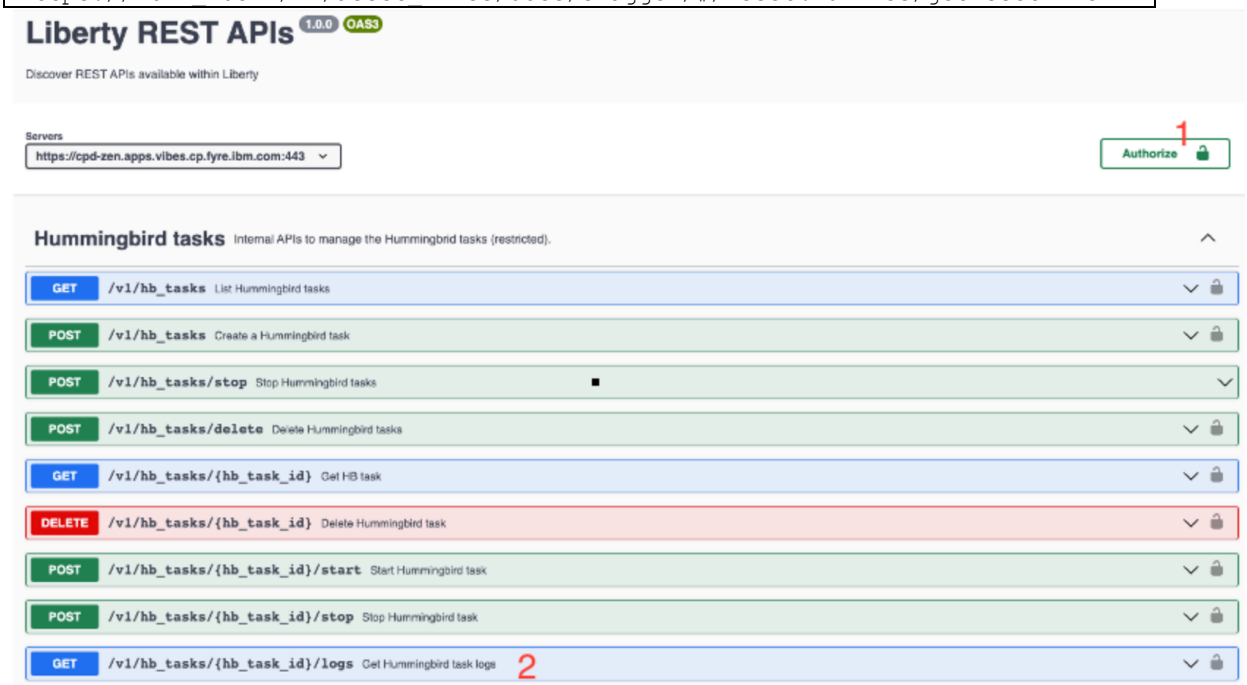
For example, an MDE job log reported profiling on the table “TLE_OUTBOUND” taking longer than 300 seconds and it skipped. You like to know the SQL used in this MDE job.



1. Find out the Hummingbird task ID from the MDE job log where the asset enrichment attempted. In this case it is “5433a37-a42f-4731-870f-e2c6b9ef82f7”. If there are multiple Hummingbird tasks, you need to consider the asset enriched time and the Hummingbird tasks start time to find the correct one.

2. Go to Hummingbird tasks explorer (Swagger) from web browser and capture the HB task log. You need to authenticate yourself on the Swagger. Secondly, make sure you have access to the respective project where MDE executed.

```
https://<CPD_HOST>/v2/asset_files/docs/swagger/#/Asset%20Files/getAssetFile
```



3. Download the HB task log and search the table to find out the SQL. For example:

```
    % grep -i TEL_OUTBOUND hb_task_log.txt
```

```
INFO: CDICO0004I: Interaction properties: {query_timeout=300,
row_limit=10001, schema_name=EDIMON, table_name=TEL_OUTBOUND}.

INFO: CDICO2019I: The specified table name is: TEL_OUTBOUND

INFO: CDICO2020I: The connector will run the statement: SELECT
"FILE_NAME", "HEADER", "CONTENT" FROM "EDIMON"."TEL_OUTBOUND" FETCH FIRST
10001 ROWS ONLY
```

End-of-day CPD health check

Process of identifying potential issue:

1) What are you going to run over night?
    a. MDI/MDE/Publishing jobs

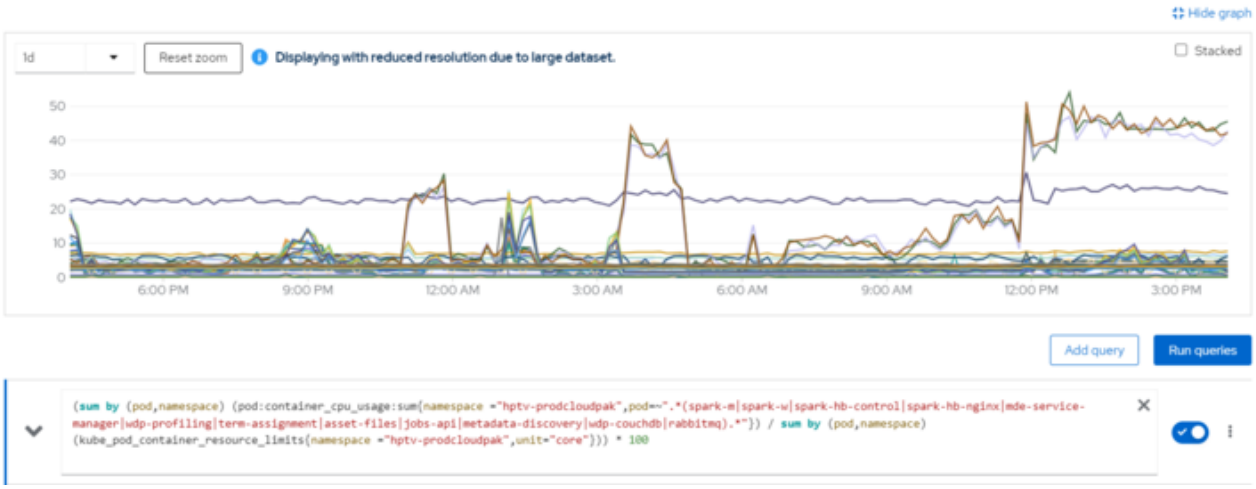
2) What are pods associated to these operations?
    a. Check pod reference table[A6] to find pods associated to the Service/Component.

3) Find status of pods using command line or CPD monitor console.
    a. Are all pods healthy? [A1]
    b. Any pods restarted recently within last 24 hours? If yes, investigate why in restart and take corrective action. [A3]

```
# oc get pods | egrep 'spark-m|spark-w|spark-hb-control|spark-hb-nginx|mde-service-manager|wdp-
profiling|term-assignment|asset-files|jobs-api|metadata-discovery|wdp-couchdb|rabbitmq'

asset-files-api-7b74f7d4cf-82hzb                                  1/1     Running     1 (4d4h ago)     4d23h
asset-files-api-7b74f7d4cf-b4xjj                                  1/1     Running     2 (3d10h ago)    4d23h
asset-files-api-7b74f7d4cf-dgbmz                                  1/1     Running     0                4d23h
asset-files-api-7b74f7d4cf-hxtpq                                  1/1     Running     2 (38h ago)      4d23h
asset-files-api-7b74f7d4cf-lr2t8                                  1/1     Running     1 (4d6h ago)     4d23h
asset-files-api-7b74f7d4cf-z5nqc                                  1/1     Running     2 (27h ago)      4d23h
jobs-api-6484667d79-5xjv8                                         1/1     Running     0                5d23h
metadata-discovery-68bfc986dd-gd8zr                               1/1     Running     0                5d19h
metadata-discovery-68bfc986dd-rxt7d                               1/1     Running     0                5d19h
metadata-discovery-68bfc986dd-w67fn                               1/1     Running     0                5d19h
rabbitmq-ha-0                                                     1/1     Running     0                5d3h
rabbitmq-ha-1                                                     1/1     Running     0                5d3h
rabbitmq-ha-2                                                     1/1     Running     0                5d3h
rabbitmq-ha-3                                                     1/1     Running     0                5d3h
spark-hb-control-plane-59759fcdb6-89678                           2/2     Running     0                8d
spark-hb-nginx-85f7c6995d-6hjbj                                   1/1     Running     0                5d18h
wdp-couchdb-0                                                     2/2     Running     0                5d21h
wdp-couchdb-1                                                     2/2     Running     0                5d21h
wdp-couchdb-2                                                     2/2     Running     0                5d21h
wdp-profiling-745bd5897b-4vhs4                                    1/1     Running     0                4d23h
wdp-profiling-745bd5897b-bj262                                    1/1     Running     0                4d23h
```

    c. What is memory, CPU consumption of pod over last 24 hours? Measure the pods resource usage from OpenShift console using metrics query. Current resource usage below 40% will be ideal.


Hide graph
1d
Reset zoom
Displaying with reduced resolution due to large dataset.
Stacked
50
40
30
20
10
0
6:00 PM
9:00 PM
12:00 AM
3:00 AM
6:00 AM
9:00 AM
12:00 PM
3:00 PM
Add query
Run queries
(sum by (pod,namespace) (pod:container_cpu_usage:sum{namespace ="hptv-prodcloudpak",pod=~".*(spark-m|spark-w|spark-hb-control|spark-hb-nginx|mde-service-manager|wdp-profiling|term-assignment|asset-files|jobs-api|metadata-discovery|wdp-couchdb|rabbitmq).*"}) / sum by (pod,namespace)
(kube_pod_container_resource_limits{namespace ="hptv-prodcloudpak",unit="core"})) * 100

Action

A1. Are all pods healthy? If not:
- Investigate pod issues
- Try restart the pod

A2. Any recent warning/fail event associated to the pod? If yes:
- Get an understanding of the event and take corrective action.

A3. Has pod restarted recently? If yes:
- Need to investigate if pod restarted multiple times.
- Check pod description to find the reason of restart.
- Capture the previous pod log. Look for any error message at the end of log to figure out the cause of restart.
- Search exiting Jira for similar error message.
- If need help open support ticket with IBM along with all necessary diagnostics data.

A4. Any error in the pod log during the problem? If yes:
- Search exiting Jira for similar error message.
- If need help open support ticket with IBM along with all necessary diagnostics data.

A5. What is memory, CPU consumption of pod?
- Check the pod description for current resource configured.
- Check the pod description for restart due to lack of resource.
- Using OpenShift console monitor pod resource usage for a period.
- Check with IBM support if pod restarted due to lack of resource.

A6. Pod reference table
- https://github.com/sanjitc/Cloud-Pak-for-Data/blob/main/wkc/troubleshooting/Pods_Making_CPD.md