

Table of Contents

Diagnostics Information:	
Collecting Data for Troubleshooting	2
Troubleshooting:	
Accessing MDE column data taking time.....	3
MDE Publish job appearing hang	4
Find the SQL that causing problem or failing in the MDE job.....	6
Monitoring:	
End-of-day CPD health check.....	8
Miscellaneous:	
Action.....	10

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

Data Class related issues:

Data Needs to Capture	
Diagnostic job	Gather diagnostics information from CPD UI (CCS, WKC, AE)
Application logs from swagger	<a href="https://<CPD_URL>/v2/data_profiles/api/explorer/#/Hummingbird%20tasks/getHbTaskLogs">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	<a href="https://<CPD_URL>/v2/cams/explorer/#/Assets/retrieveAssets">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	<a href="https://<CPD_URL>/v2/data_profiles/api/explorer/#/Hummingbird%20tasks/getHbTaskLogs">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	<pre>curl -i -k -H "content-type: application/json" -H "Authorization: bearer \$Bearer_TOKEN" -X GET "https://\$HOSTNAME/v3/reporting/heartbeat"</pre> <pre>curl -i -k -H "content-type: application/json" -H "Authorization: bearer \$Bearer_TOKEN" -X GET "https://\$HOSTNAME/v3/reporting/999/register"</pre> <pre>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"</pre>

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	<pre>oc adm inspect/namespace <CPD namespace></pre> <pre>oc adm inspect/namespace <IBM CPD operator namespace></pre>

OpenShift Data Fusion related issues:

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

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.

Accessing MDE column data taking time

CPD/IKC – Knowledge Sharing

24122_DR_MDE_V1_PP2

Columns (13,794)

Columns

Columns	Asset	Business terms	Data class	Data quality	Review status
CHANGEDATE	X2_DRIVERSON GPPV_24122_DisasterRecovery_C				
CHANGEDATE	X2_REPVERSION GPPV_24122_DisasterRecovery_C				
CHANGEDATE	X2_RUNVERSION GPPV_24122_DisasterRecovery_C				

CHANGEDATE

Columns from X2_RUNVERSION

Go to asset in project

Details

Description

Source

GPPV_24122_DisasterRecovery_C / SAPS...

Troubleshooting:

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

Elements Console Sources Network Performance Memory Application Security Lighthouse Recorder Performance insights

Search

Access_Point_Nam

Filter

200000000000 ms 400000000000 ms 600000000000 ms 800000000000 ms 1000000000000 ms 1200000000000 ms 1400000000000 ms 1600000000000 ms 1800000000000 ms

Name	Status	Type	Initiator	Size	Time	Waterfall
?size=24&scaleFactor=1x&showFallbackMonogram=&page...ps%3A%2F%2Fchrome.google...	200	png	chrome://res...	994 B	2 ms	
db925b85-ed34-49e3-a5f4-8cc8fa9e6719?project_id=36...225-4190-b60c-a072fed121ab&co...	500	xhr	datacatalog.v...	712 B	853 ms	
db925b85-ed34-49e3-a5f4-8cc8fa9e6719?project_id=36...225-4190-b60c-a072fed121ab&co...	304	xhr	datacatalog.v...	316 B	402 ms	
db925b85-ed34-49e3-a5f4-8cc8fa9e6719?project_id=36...225-4190-b60c-a072fed121ab&co...	500	xhr	datacatalog.v...	620 B	6.03 s	
3d87ea75f01484b48e97.svg	200	svg+xml	datacatalog.v... (disk cache)		2 ms	
tokenexpiry	200	xhr	datacatalog.v...	613 B	132 ms	

- Mainly the “Network” and “Console” panels are commonly used for investigating network activities, API call failure and error message returned 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.

Name

Headers Payload Preview Response Initiator Timing

General

Request URL: https://datacatalog.verizon.com/gov/metadata-enrichments/api/data-class-assignment/db925b85-ed34-49e3-a5f4-8cc8fa9e6719?project_id=36d882a5-f225-4190-b60c-a072fed121ab&column_name=access_amt

Request Method: GET

Status Code: 500 Internal Server Error

- 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](#) 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.

Apr 29, 2024 1:43:09 PM

About this run

Failed

Run details

Duration (seconds): 385

Started by: SVC-dcat-cadm-prd SVC-dcat-cadm-prd

Associated job: 23400_MDE_ediprod_V1

Log Total 23 lines

Download

```
Delta metadata enrichment job run (7e0b102-f5fe-4801-b9e8-d09262f208e) is in state 'Failed'.
Error reason: None of the assets could be processed successfully

Enrichment asset summary:
Total assets: 1
- Assets with status 'Created': 0
- Assets with status 'In progress': 0
- Assets with status 'Not found': 0
- Assets with status 'Completed': 1
- Completed successfully: 0
- Completed with errors due to failed profiling operation: 1
- Completed with errors due to failed term assignment operation: 0

Profiling status:
Completed Hummingbird tasks: 1
Failed Hummingbird tasks: 0
5433a37-a42f-4731-870f-e2c6b9ef82f7 - 2024-04-29T17:43:26.962816992
2024-04-29T17:43:27.0472116112 [-08] - SUBMITTED
2024-04-29T17:43:16.7697215132 [-23s] - RUNNING
2024-04-29T17:49:15.337879236Z [-348s] - COMPLETED

Asset failures: 1
2024-04-29T17:49:15.3362 - 50b46894-d4a8-40ba-a180-49ed274f1e3a [PJ ("TLE_OUTBOUND") - Records of the data asset could not be read within 300 seconds. Stopping profile request for the asset to continue with others.
```

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

Liberty REST APIs 1.0.0 OAS3

Discover REST APIs available within Liberty

Servers

<https://cpd-zen.apps.vibes.cp.fyre.ibm.com:443>

Authorize

Hummingbird tasks Internal APIs to manage the Hummingbird tasks (restricted).

GET	/v1/hb_tasks	List Hummingbird tasks
POST	/v1/hb_tasks	Create a Hummingbird task
POST	/v1/hb_tasks/stop	Stop Hummingbird tasks
POST	/v1/hb_tasks/delete	Delete Hummingbird tasks
GET	/v1/hb_tasks/{hb_task_id}	Get HB task
DELETE	/v1/hb_tasks/{hb_task_id}	Delete Hummingbird task
POST	/v1/hb_tasks/{hb_task_id}/start	Start Hummingbird task
POST	/v1/hb_tasks/{hb_task_id}/stop	Stop Hummingbird task
GET	/v1/hb_tasks/{hb_task_id}/logs	Get Hummingbird task logs

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]

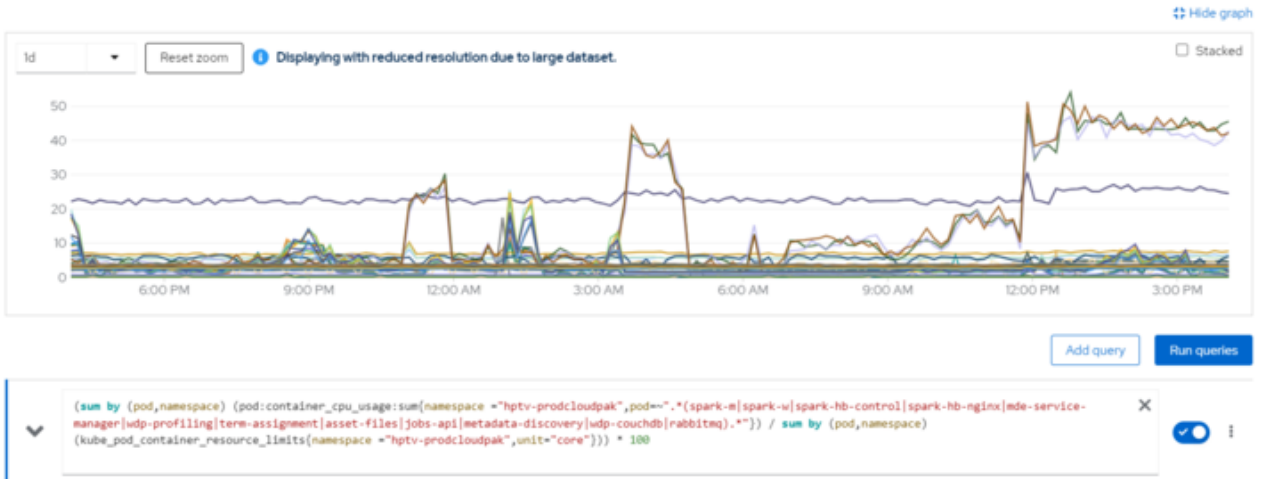
```
# oc get pods | egrep 'spark-m|spark-w|spark-hb-control|spark-hb-nginx|mde-service-manager|wdp-
profiling|term-assignment|asset-files|jobs-api|metadata-discovery|wdp-couchdb|rabbitmq'
```

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-dgvmz	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-59759fcd6-89678	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.



CPD/IKC – Knowledge Sharing



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