|  |  |
| --- | --- |
|  | **THE MIZUHO SECURITIES**  **ASIA LIMITED** |

**Mizuho Securities Asia**

**Data Analytics Platform**

**Operation guide**

|  |  |
| --- | --- |
|  | Logo  Description automatically generated |
| Prepared by | Eastech Systems Limited |
| Document Version | v0.1 |
| Date | Dec 2, 2024 |

Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Author** | **Version** | **Description** |
| 2nd Dec, 2024 | Jasper Cheung | v0.1 | Document created. |
| 5th Dec, 2024 | Ellis Chiang | v0.2 | Formatting and additional content. |
|  |  |  |  |
|  |  |  |  |

Reviewers

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Position** | **Version** | **Date** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Distributions

|  |  |  |
| --- | --- | --- |
| **Department** | **Name** | **Position** |
| Information Technology Department | Gordon Lo | Vice President |
|  |  |  |
|  |  |  |
|  |  |  |

**Table of Contents**

[1 Introduction 6](#_Toc184844308)

[1.1 Purpose of Document 6](#_Toc184844309)

[1.2 Architecture of the Data Analytics Platform 6](#_Toc184844310)

[2 RunMyJob Scheduler (RMJ) 7](#_Toc184844311)

[2.1 Trigger a RMJ Job 7](#_Toc184844312)

[2.2 Monitor RMJ Job Status 9](#_Toc184844313)

[2.3 Resubmit a RMJ Job 9](#_Toc184844314)

[2.4 Access the Log Message of a RMJ Job 9](#_Toc184844315)

[2.5 Schedule and Alert 10](#_Toc184844316)

[2.5.1 To Enable/Disable a Schedule 10](#_Toc184844317)

[2.5.2 To Enable/Disable an Alert 11](#_Toc184844318)

[3 Data Factory 12](#_Toc184844319)

[3.1 Trigger An ETL Job 12](#_Toc184844320)

[3.2 Monitor ETL Job Status 14](#_Toc184844321)

[4 Synapse Dedicated SQL Pool 17](#_Toc184844322)

[4.1 Resume and Pause Dedicated SQL Pool 17](#_Toc184844323)

[4.2 Check the Log of Stored Procedure Executed 17](#_Toc184844324)

[4.3 Check the Parameters of the Latest Run of a Report 19](#_Toc184844325)

[5 SQL Server Reporting Services (SSRS) 21](#_Toc184844326)

[5.1 Trigger a Report Subscription 21](#_Toc184844327)

[5.2 Monitor Subscription Status 23](#_Toc184844328)

[5.3 Restart Reporting Services 25](#_Toc184844329)

[5.4 Check Report Server Log 26](#_Toc184844330)

[6 Power BI Services 28](#_Toc184844331)

[6.1 Refresh a Semantic Model (Dataset) 28](#_Toc184844332)

[6.2 Update the Content of a Power BI App 28](#_Toc184844333)

[7 Troubleshooting 31](#_Toc184844334)

[7.1 Drill Down from RMJ Alert to Data Factory Monitor with Run ID 31](#_Toc184844335)

[7.2 Common Issue 34](#_Toc184844336)

[7.2.1 Repository Database is Inactive/Paused 34](#_Toc184844337)

[7.2.2 Failed to Connect to Data Sources 35](#_Toc184844338)

[7.2.3 Missing Input Files in Shared Folder 37](#_Toc184844339)

[7.2.4 Report Subscription Execution Failure 39](#_Toc184844340)

[7.2.5 Power BI Semantic Model Refresh Failure 41](#_Toc184844341)

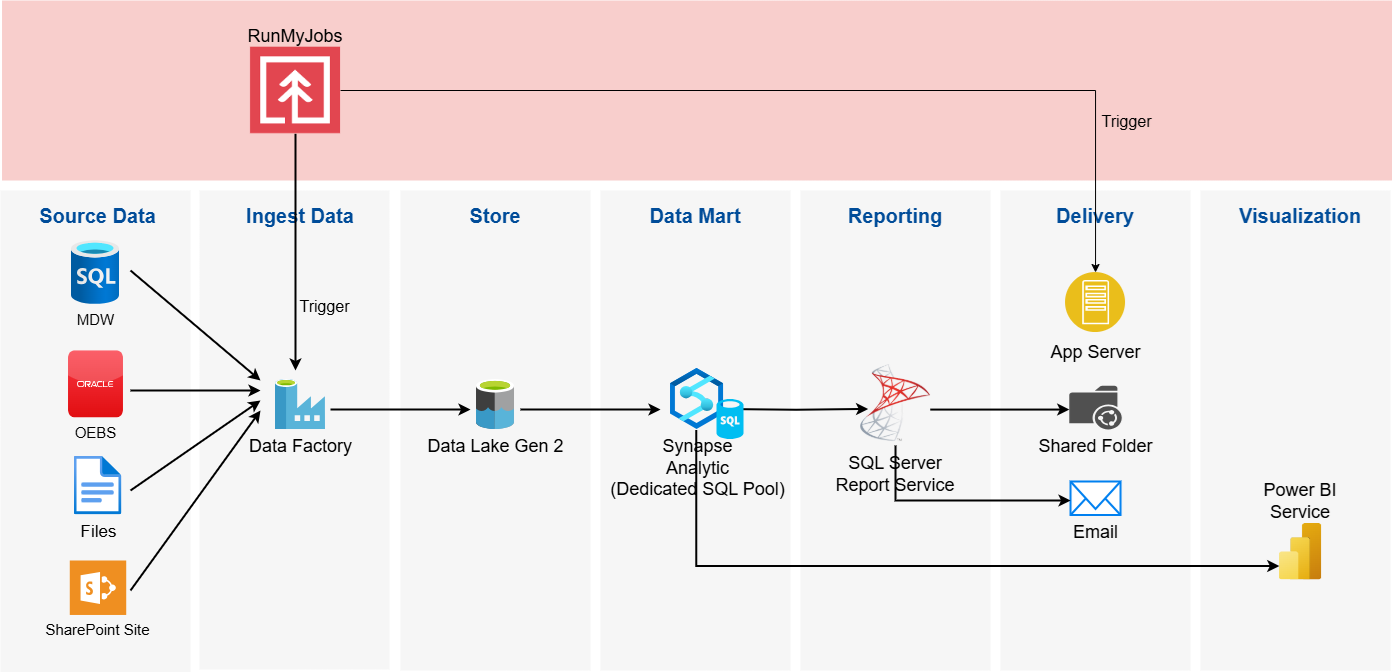
# Introduction

## Purpose of Document

This operation guide is designed to provide a comprehensive understanding of the operation related tasks on the key components in the Azure Data Analytics platform. These components include the RunMyJob (RMJ) scheduler, Data Factory, Synapse Analytics Workspace, SQL Server Reporting Services (SSRS), and Power BI service. It provides detailed instructions on essential tasks such as triggering and monitoring data pipelines, starting and stopping services, responding to alerts and so on.

## Architecture of the Data Analytics Platform

The diagram below provides an overview of the architecture of the Data Analytics Platform.



The architecture consists of the following Azure components:

* Data Factory: Orchestrates and automates data integration and ETL processes.
* Data Lake Gen 2: Scalable, secure, and cost-effective data lake storage solution.
* Synapse Analytics: Unified analytics platform for big data and real-time analytics.
* SQL Server Reporting Services (SSRS): Reporting platform for create, deploy and manage reports.
* Power BI Services: Cloud based business analytics service for data visualization.

In brief, the Data Factory pipeline is designed to extract data from source databases and input files located in a shared folder. It transfers the necessary data into the Azure Data Analytics Platform, where it is stored in the data lake. This data is then mounted to the SQL Pool in Synapse Analytics as an external table, creating a result data set for reporting. SQL Server Reporting Services (SSRS) generates reports using these result datasets. The generated reports can either be saved in the shared folder or sent via email, depending on the subscription settings in SSRS. Additionally, Power BI services will be used to visualize the data in the SQL Pool through a dashboard.

# RunMyJob Scheduler (RMJ)

RMJ serves as the central entry point for managing and automating the ETL process and report generation within the Data Platform. Through RMJ, users can efficiently schedule, monitor, and manage the lifecycle of the report process.

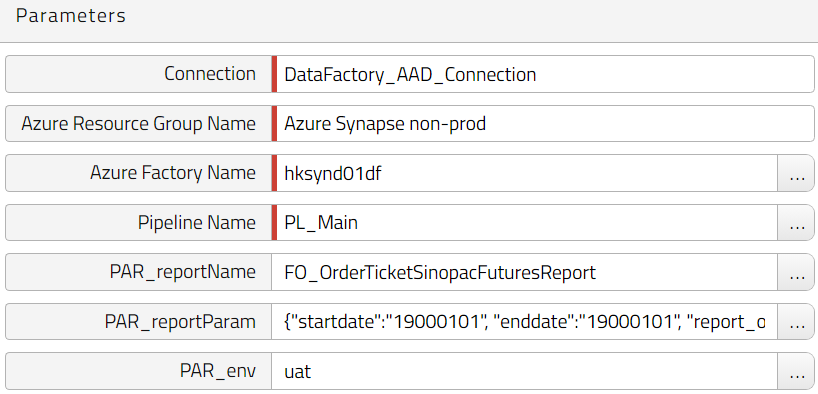
## Trigger a RMJ Job

1. In RMJ portal, navigate to **Process Definitions** / **Chain Definition** tab.
2. Search by Keywords on the search bar to filter the corresponding process definition / chain definition.
3. To Trigger a process definition, right click on the row, and click on “**Submit**”.

A screenshot of a computer

Description automatically generated

1. A Submit Process window will then show up, check the parameter values before submission. Click on “**Submit**”.

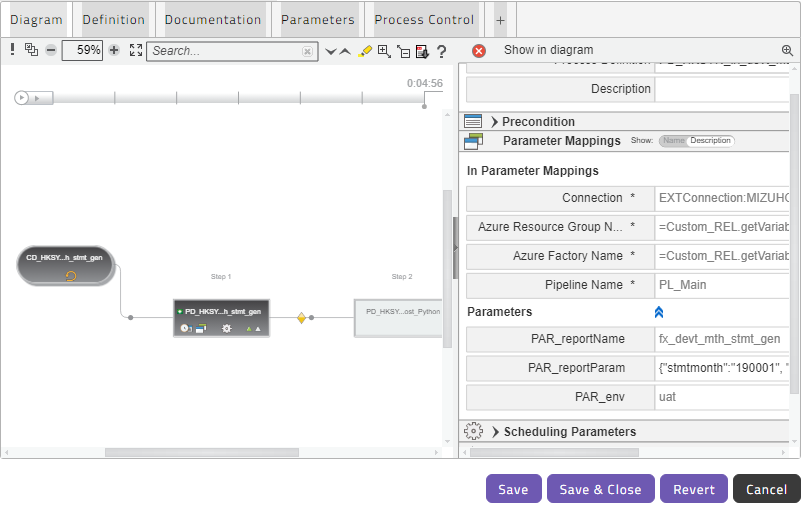


1. To Trigger a chain definition, right click on the row, and click on “**Edit**”.

A screenshot of a computer

Description automatically generated

1. An Edit Chain Definition window will then show up. In the diagram, click on the process that triggers a data factory pipeline. Click on the blue arrow on the right pane to toggle the text fields form. Update the report parameter values if required. Click on “**Save and Close**”.



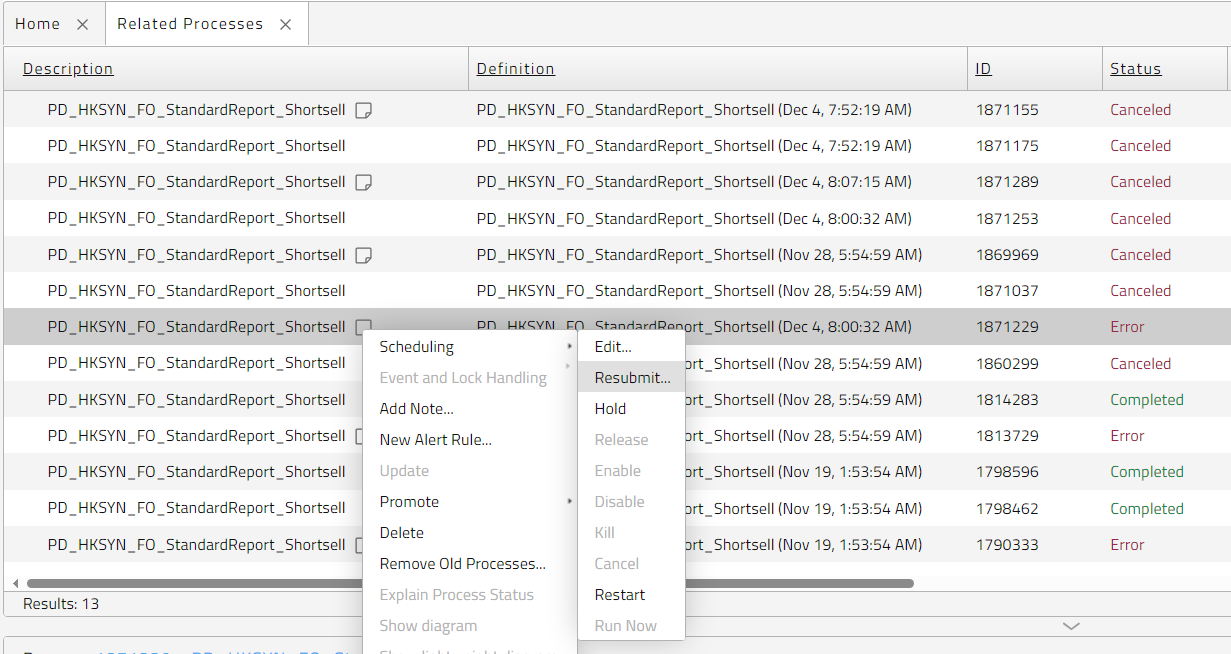
1. Right click on the chain definition row again, click on “**Submit**”.
2. A Submit Process window will then show up, click on “**Submit**”.

## Monitor RMJ Job Status

1. In RMJ portal, navigate to **Process Definitions** / **Chain Definition** tab.
2. Search by Keywords on the search bar to filter the corresponding process definition / chain definition.
3. To Monitor the job status of historical runs. Right click on the row and click on “**Monitor related processes**”.
4. The Related Processes tab will show up, the list of previous runs will be displayed from the most recent run to the oldest run.

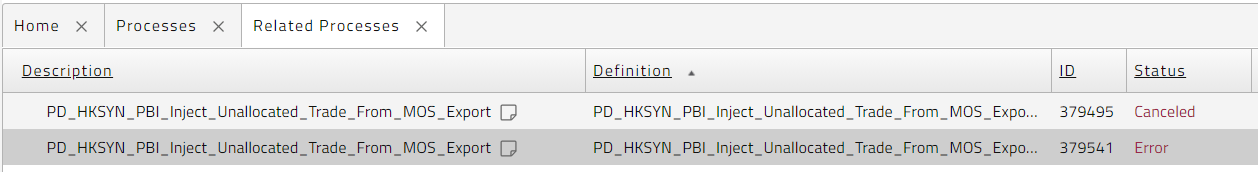
## Resubmit a RMJ Job

1. In case a process or chain encountered error and is required a job resubmission, in the RMJ portal, navigate to the failed job in the processes monitoring tab or monitor related processes tab.
2. Right click on the failed job entry and select “**Resubmit**” from the menu.

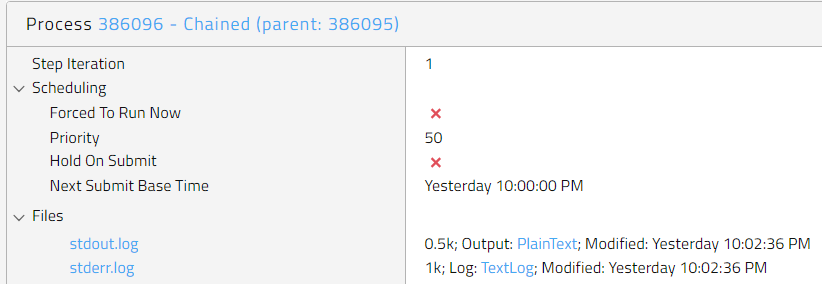


## Access the Log Message of a RMJ Job

1. In RMJ portal, navigate to the process or chain from the processes monitoring tab or monitor related processes tab.
2. Highlight the job entry that need to access the job log.

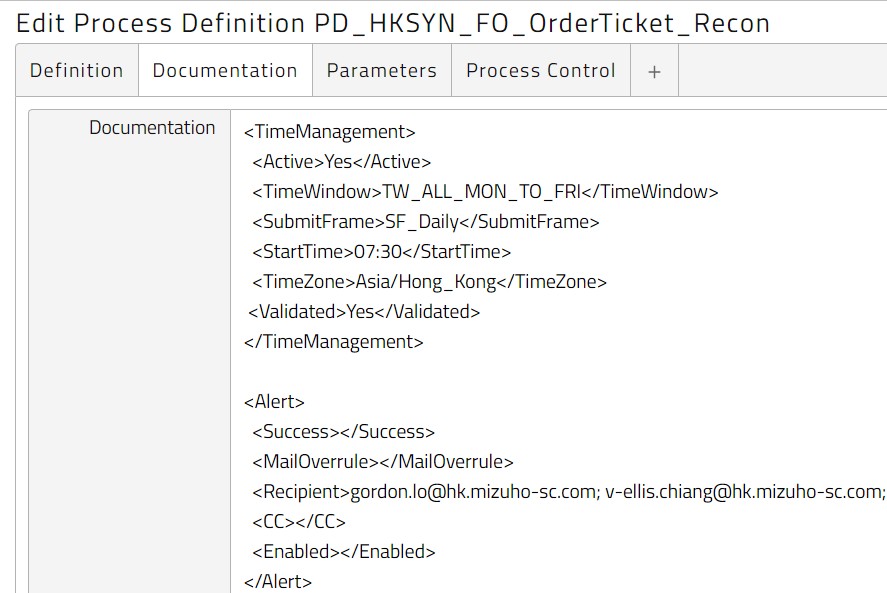


1. In the process details pane, scroll down to “**Files**” and there will be a “**stdout.log**” and/or a “**stderr.log**” (depends on setup and remote system type) attached with the RMJ process. The log message from the remote system can be located from these log files.



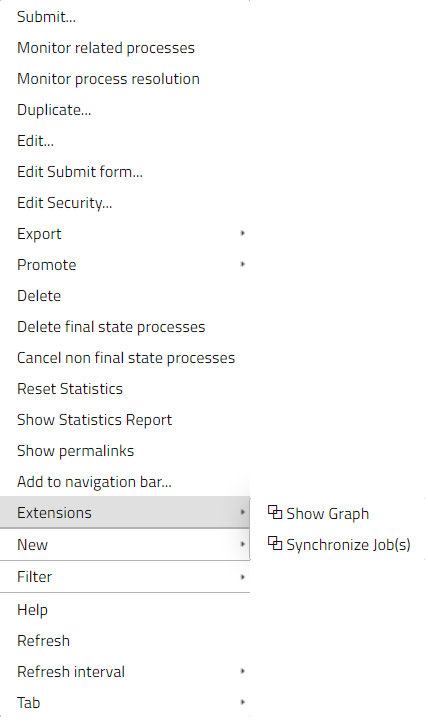
## Schedule and Alert

The schedule and alert information are configured in the “**Documentation**” tab on the “**Edit Process Definition**” or “**Edit Chain Definition**” window.



### To Enable/Disable a Schedule

1. To enable or disable a schedule of a process or chain, it can be achieved by updating the value of the **“<Active>**” tag in the “**<TimeManagement>**” section of the “**Documentation**”. Using value “**Yes**” for enable and “**No**” for disable. Click “**Save & Close**” to save the change.
2. Right click on the process or chain and select “**Synchronize Job(s)**” from the menu to enable or disable a schedule.



### To Enable/Disable an Alert

1. To enable or disable an alert of a process or chain, it can be achieved by updating the value of the **“<Enabled>**” tag in the “**<Alert>**” section of the “**Documentation**”. Using value “**Yes**” for enable and “**No**” for disable. By default, it is enabled for a blank value. Click “**Save & Close**” to save the change.
2. The alert does not require to synchronize job after the setting has been updated. The change will take effect in the next job submission.

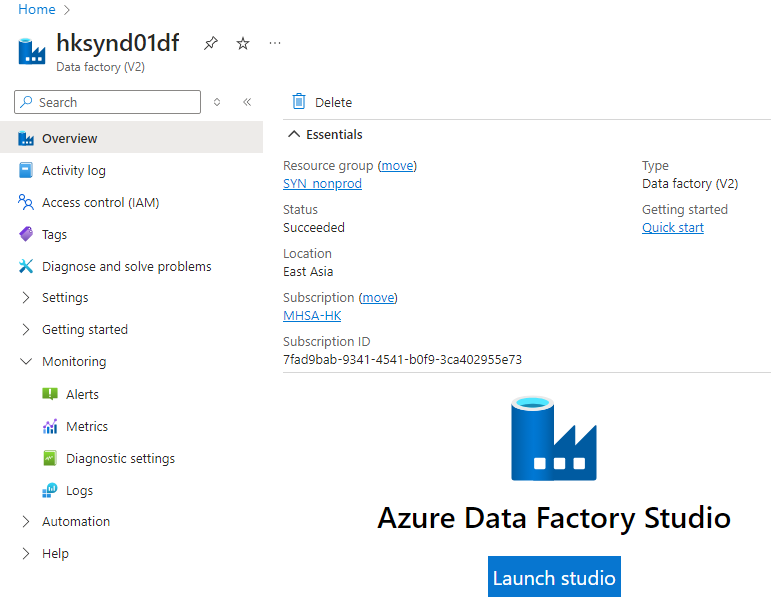
# Data Factory

## Trigger An ETL Job

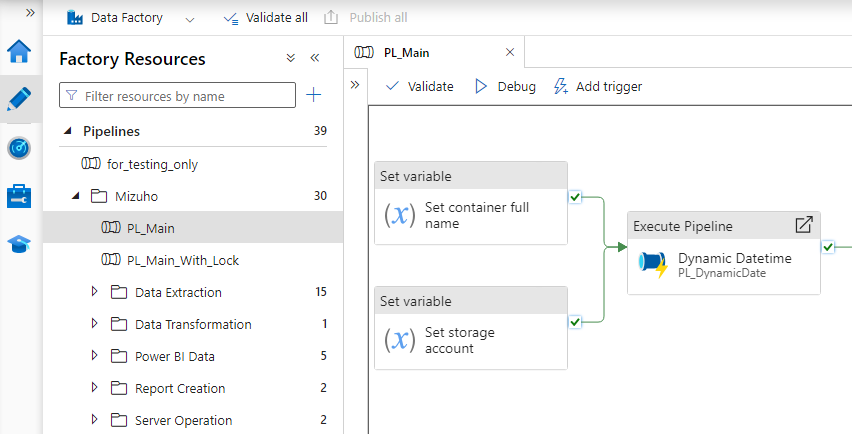
1. In Azure Portal, navigate to **Azure Data Factory**.
2. Access the target Azure Data Factory by clicking the name from the list.

DEV: hksynd01-df  
PROD: hksynp01-df

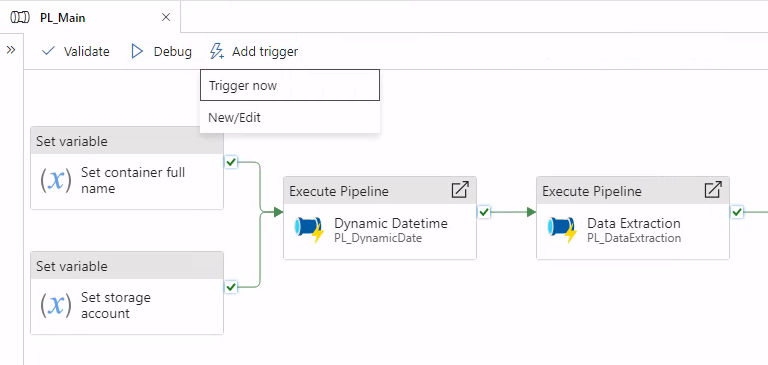
1. Click on “**Launch studio**” and Azure Data Factory Studio will be started in a new browser tab.



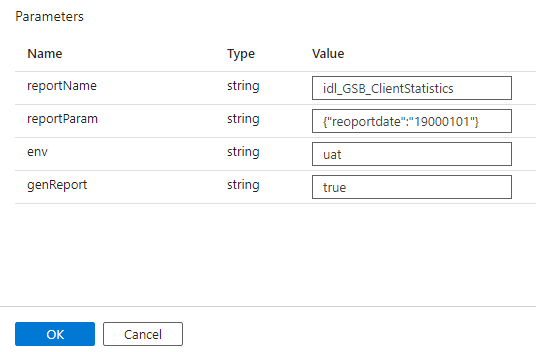
1. To trigger a ETL job for a report, go to “Author” in the side menu. Expand “**Pipelines**” and click on “**PL\_Main**” under the “**Mizuho**” folder.



1. On the main canvas, click on “**Add Trigger**”, select “**Trigger Now**”.



1. A parameter panel will be shown on the right, fill in the parameter values. Click on “**OK**”.



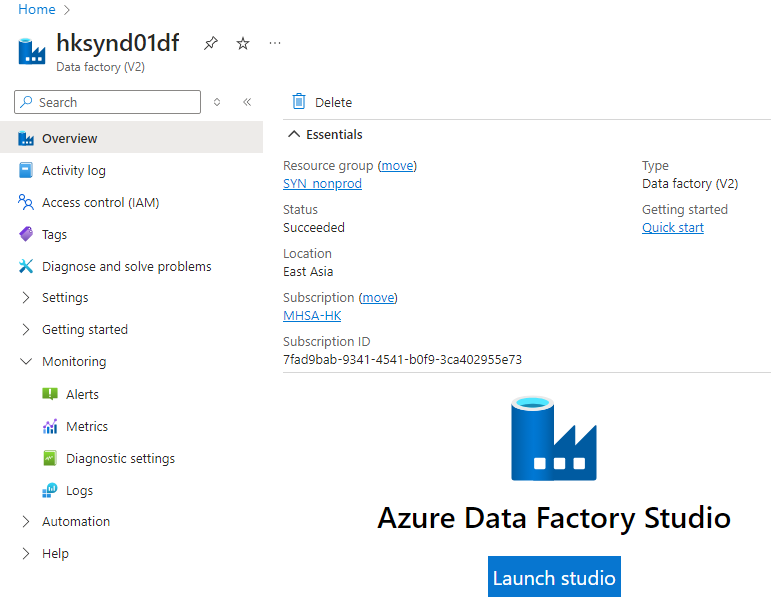
1. The pipeline will be triggered. The output status can be found on the “Output” tab at the bottom.

## Monitor ETL Job Status

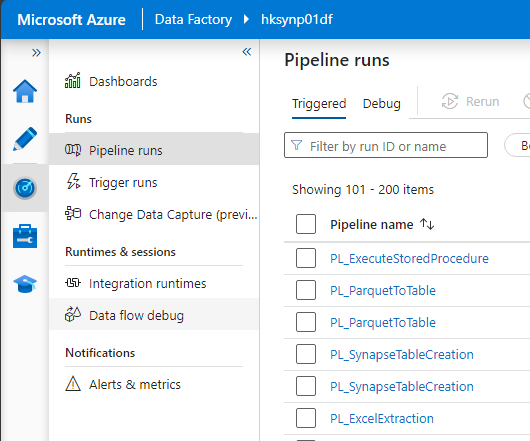
1. In Azure Portal, navigate to **Azure Data Factory**.
2. Access the target Azure Data Factory by clicking the name from the list.

DEV: hksynd01-df  
PROD: hksynp01-df

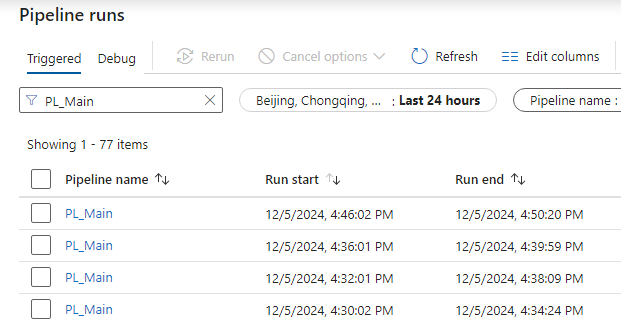
1. Click on “**Launch studio**” and Azure Data Factory Studio will be started in a new browser tab.



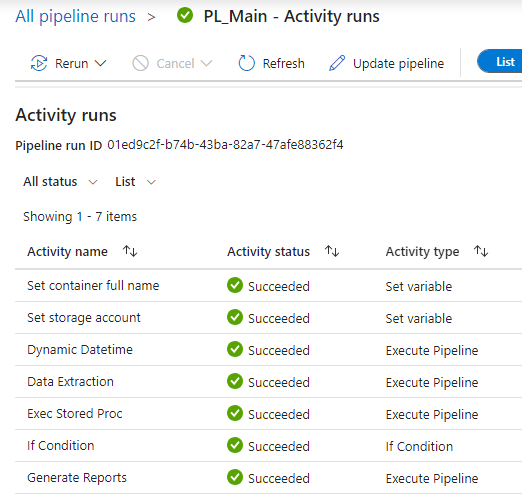
1. To monitor the status of the pipeline execution, go to “**Monitor**” in the side menu. Select “**Pipeline runs**” from the Table of content pane and click on “**Triggered**”.



1. To further narrow down the number of pipelines being shown in the Monitor’s Pipeline runs pane, enter a keyword of a pipeline name in the search box.

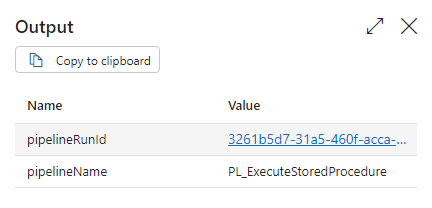


1. Click on the pipeline name to show the status of each activity being executed in the pipeline.



1. To further drill into any sub-pipelines from the root pipeline, highlight the sub-pipeline and then click on the output arrow button. A pop-up box will be shown and click on the link of the pipeline ID.





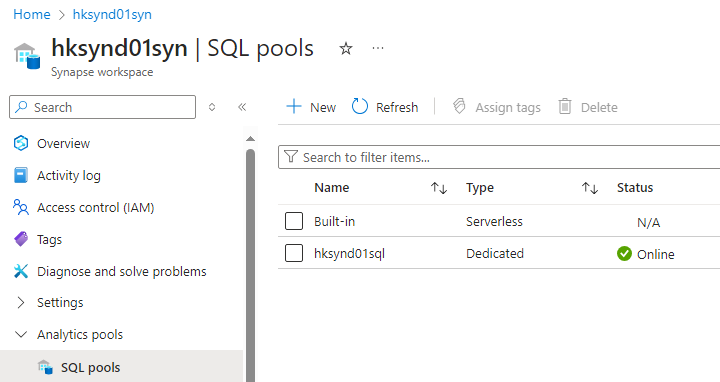
1. It will then drill down to the desired sub-pipeline and show all the activities and the corresponding status. Repeat above steps to drill down to another sub-pipelines if needed.
2. To go back to the pipelines(s) in the upper level, use the navigation bar on the top to go back to the previous level by directly clicking on the name of the pipeline in the navigation bar.



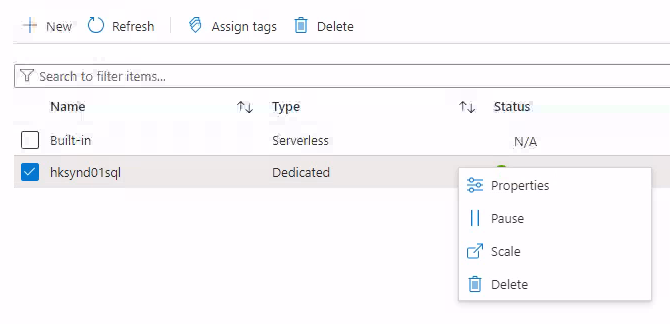
# Synapse Dedicated SQL Pool

## Resume and Pause Dedicated SQL Pool

1. In the Azure portal, navigate to **Azure Synapse Analytic**.
2. Expand Analytics pools on the left side menu, click on “**SQL pools**”.

****

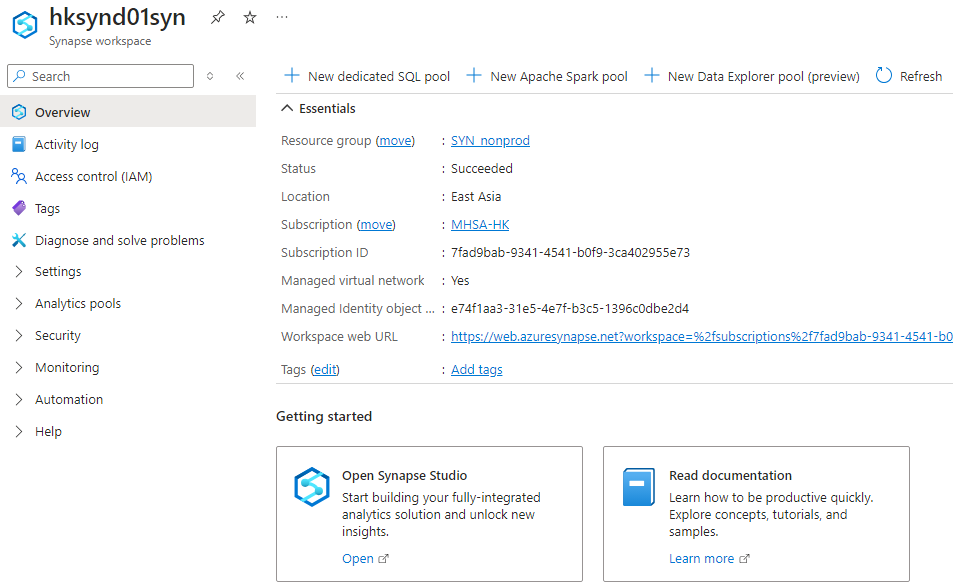
1. To pause a dedicated SQL pool, right click on the row and then click “**Pause**”.



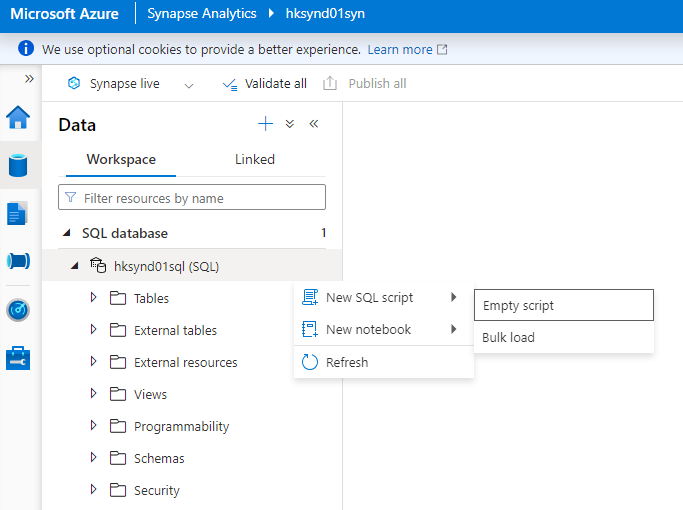
1. To resume a dedicated SQL pool, right click on the row and then click on “**Resume”**.

## Check the Log of Stored Procedure Executed

1. In the Azure portal, navigate to **Azure Synapse Analytics**.
2. Access the target Synapse workspace by clicking the name from the list.
3. Click on “**Open**” and Azure Synapse Studio will be started in a new browser tab.



1. Navigate to “**Data**” in the left menu, right click on the Database, and then click on “**Empty script**”.

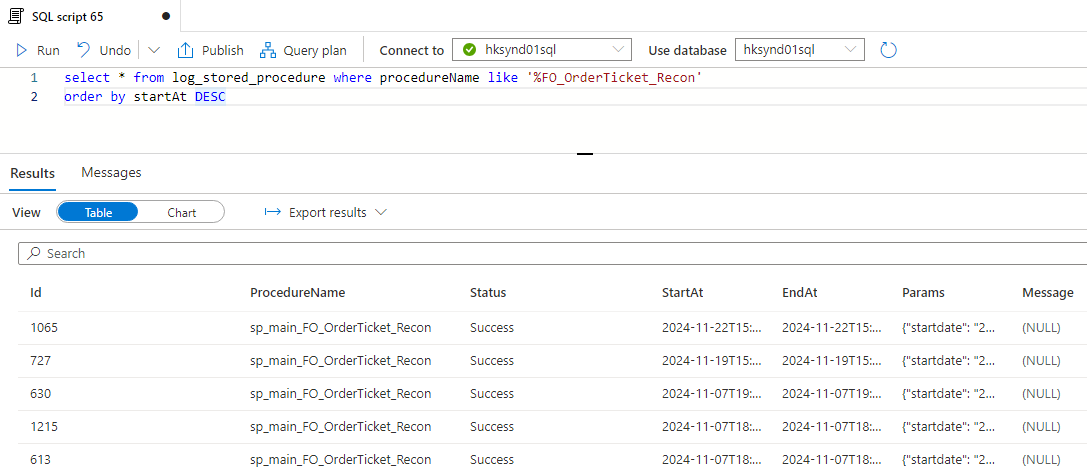


1. A new SQL script will pop-up, SQL statement can be executed.
2. To check the execution log of a stored procedure, query from the table “**log\_stored\_procedure**”.

SELECT \* FROM log\_stored\_procedure

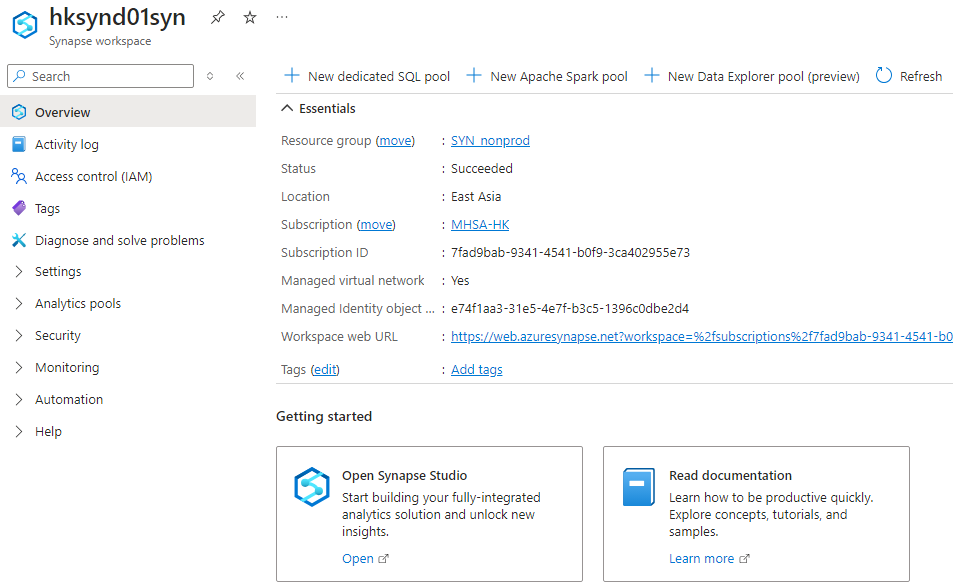
WHERE ProcudureName = '<prodcedure\_name>'

ORDER BY StartAt DESC;

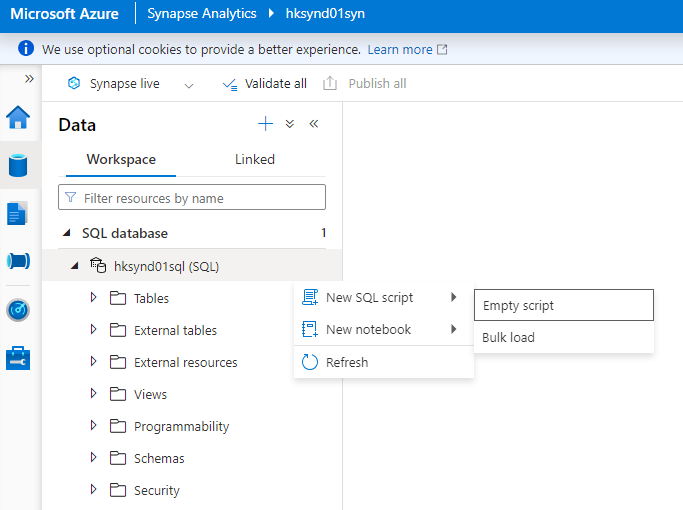


## Check the Parameters of the Latest Run of a Report

1. In the Azure portal, navigate to **Azure Synapse Analytics**.
2. Access the target Synapse workspace by clicking the name from the list.
3. Click on “**Open**” and Azure Synapse Studio will be started in a new browser tab.



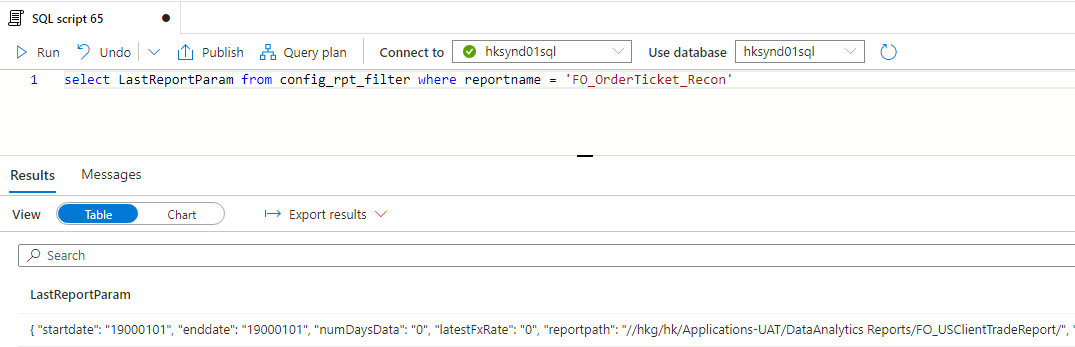
1. Navigate to “**Data**” in the left menu, right click on the Database, and then click on “**Empty script**”.



1. A new SQL script will pop-up, SQL statement can be executed.
2. To check the parameter used in the latest run, query “**LastReportParam**” from the table “**config\_rpt\_filter**”.

SELECT LastReportParam FROM config\_rpt\_filter

WHERE ReportName = '<report\_name>';



# SQL Server Reporting Services (SSRS)

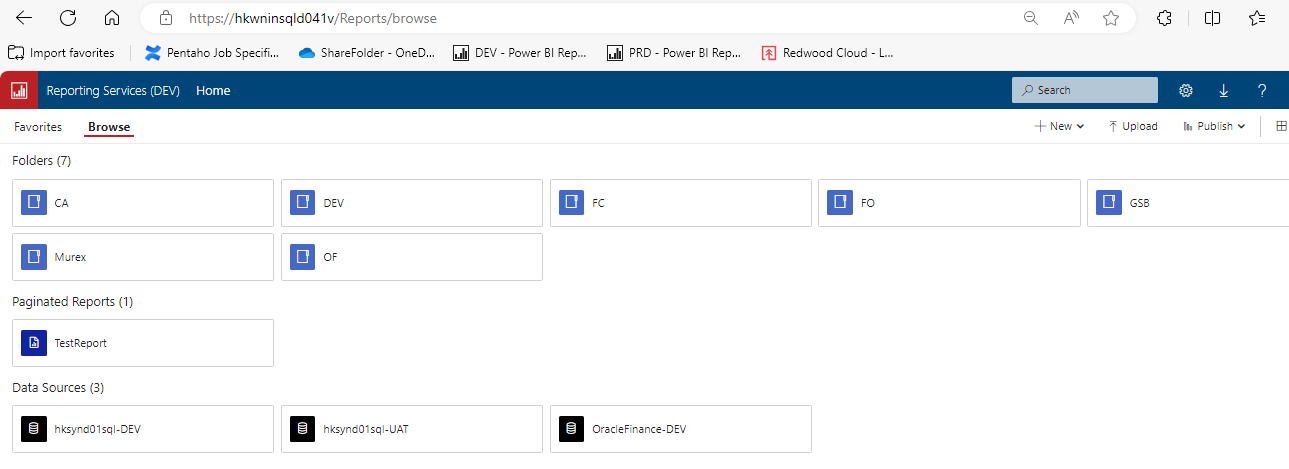
## Trigger a Report Subscription

1. Access the Reporting Services portal with a web browser.

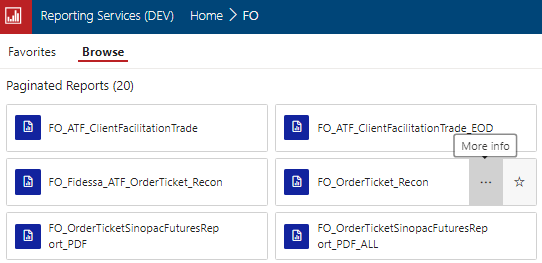
DEV: [Home - SQL Server 2019 Reporting Services](https://hkwninsqld041v/Reports/browse?_fr=1)

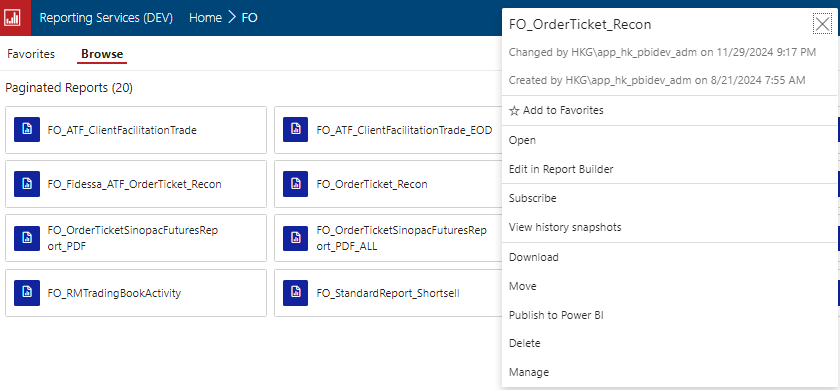
PROD: [Home - SQL Server 2019 Reporting Services](https://hkwninsqlp034v/Reports/browse?_fr=1)

1. In the Reporting Services Home, select the corresponding folder of a report.

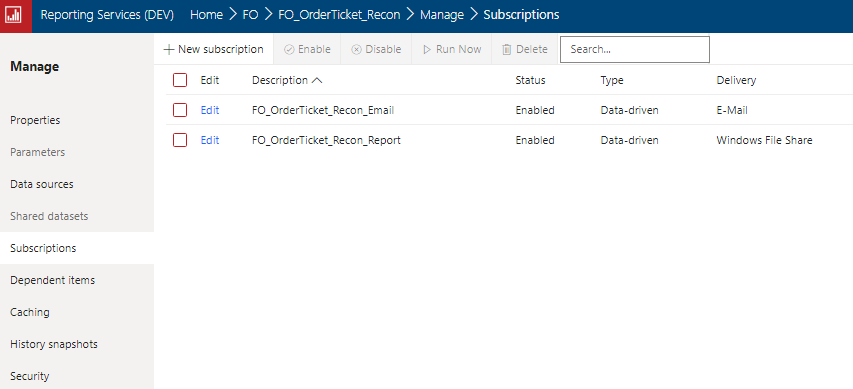


1. Hover on a report, click on the “**ellipse**” icon. From the menu, select “**Manage**”.

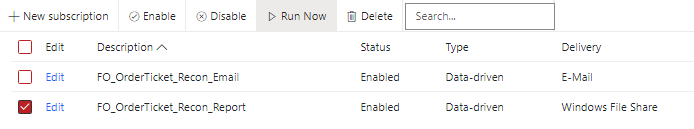




1. Click on “S**ubscriptions**” on the left menu, the subscription table will be displayed.



1. To trigger a report subscription, check the checkbox and then click on “**Run Now**”.



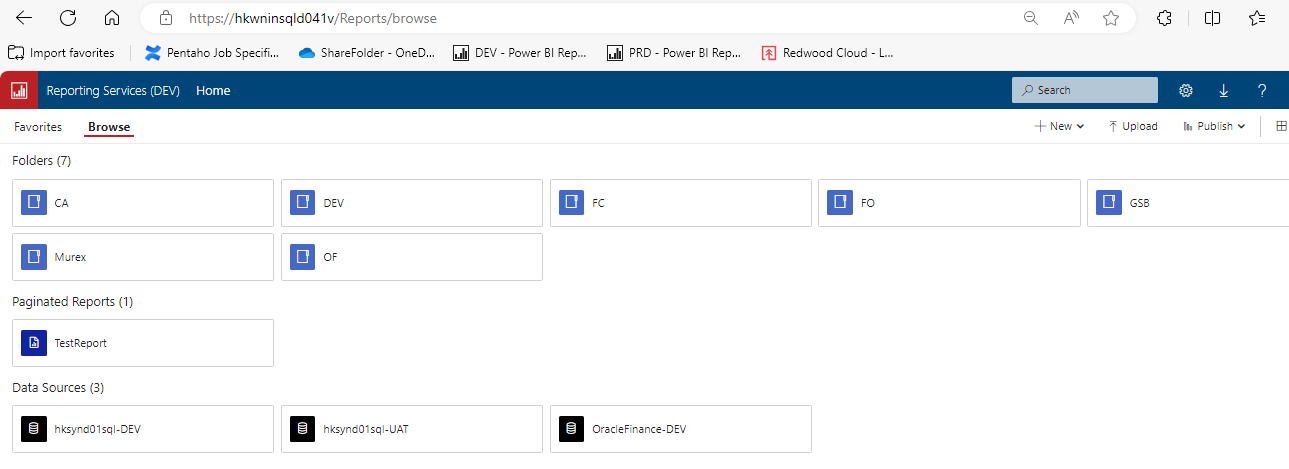
## Monitor Subscription Status

1. Access the Reporting Services portal with a web browser.

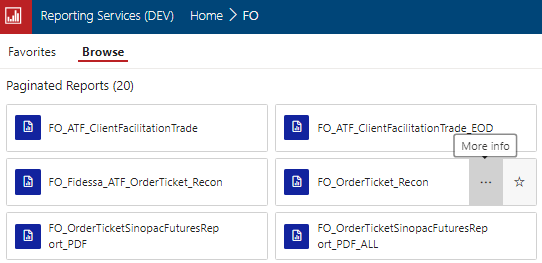
DEV: [Home - SQL Server 2019 Reporting Services](https://hkwninsqld041v/Reports/browse?_fr=1)

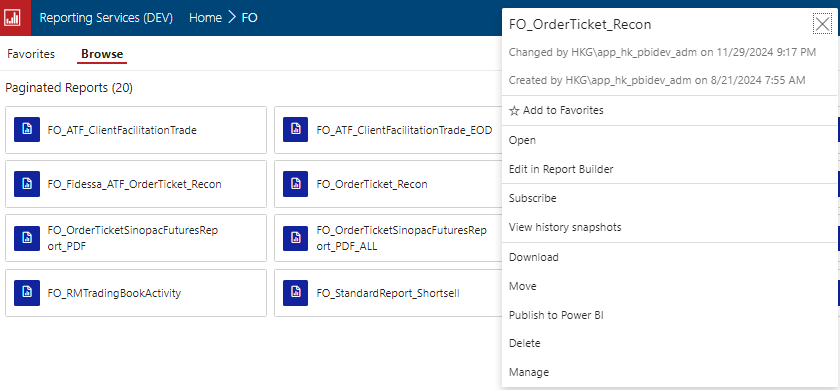
PROD: [Home - SQL Server 2019 Reporting Services](https://hkwninsqlp034v/Reports/browse?_fr=1)

1. In the Reporting Services Home, select the corresponding folder of a report.

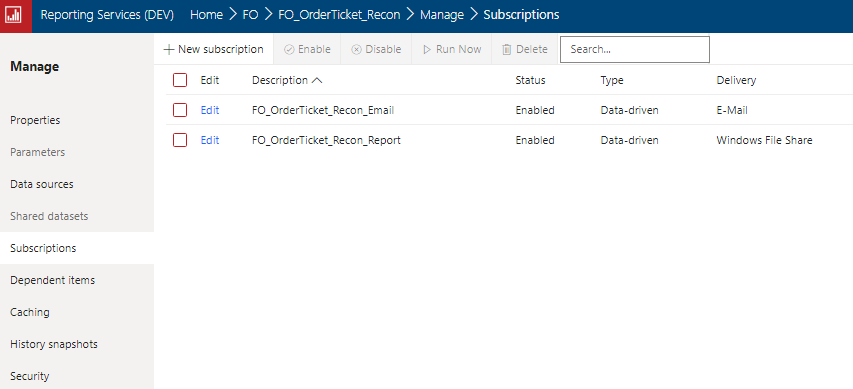
fff

1. Hover on a report, click on the “**ellipse**” icon. From the menu, select “**Manage**”.

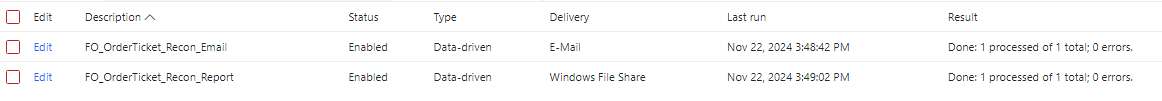




1. Click on “**Subscriptions**” on the left menu, the subscription table will be displayed.



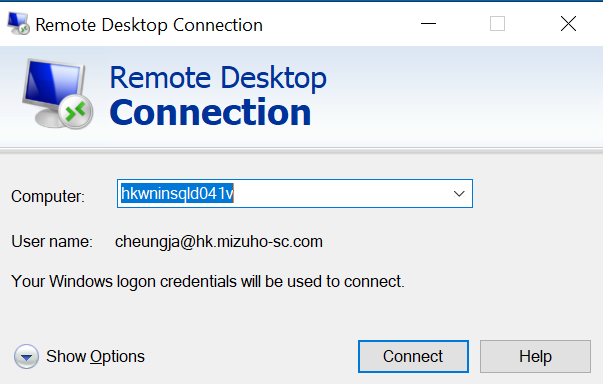
1. The last run time and the result message of a subscription will be displayed in the table.



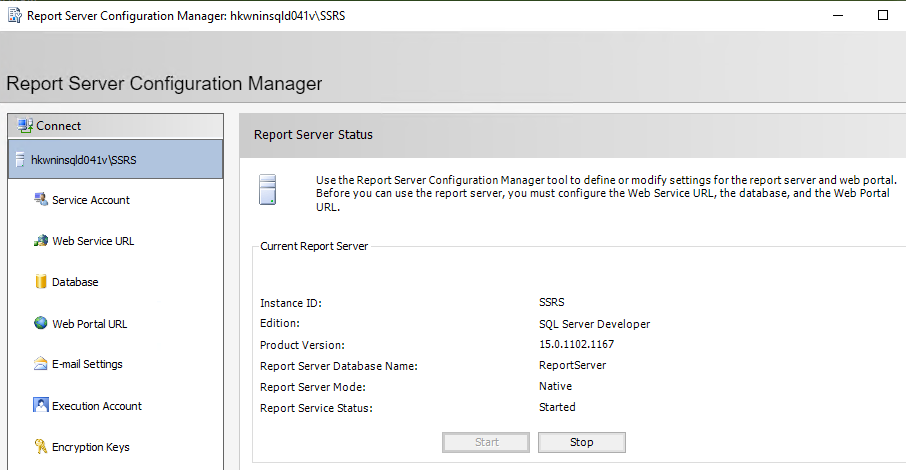
## Restart Reporting Services

1. Use remote desktop connection to connect to the Report Server.

**DEV**: hkwninsqld041v  
**PROD**: hkwninsqlp034v



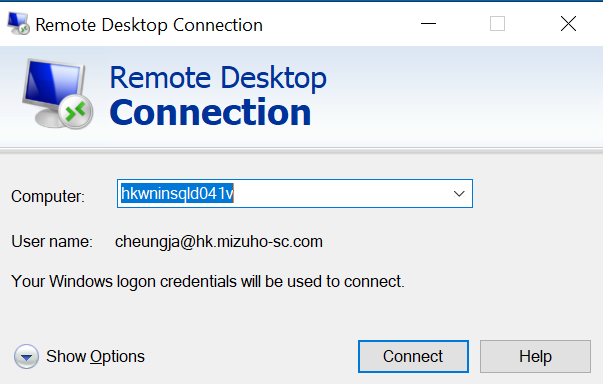
1. Open the desktop APP “**Report Server Configuration Manager**”.
2. To restart the reporting services, click on “**Stop**” and then resume by clicking on “**Start**”.



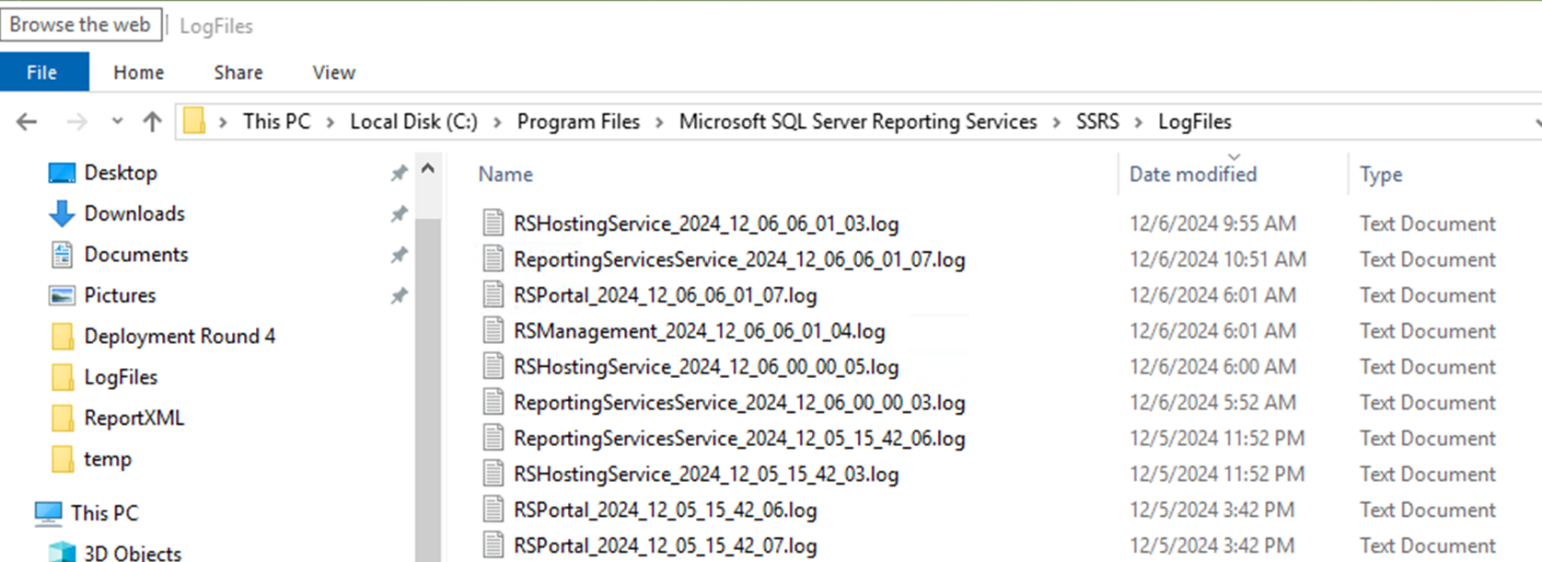
## Check Report Server Log

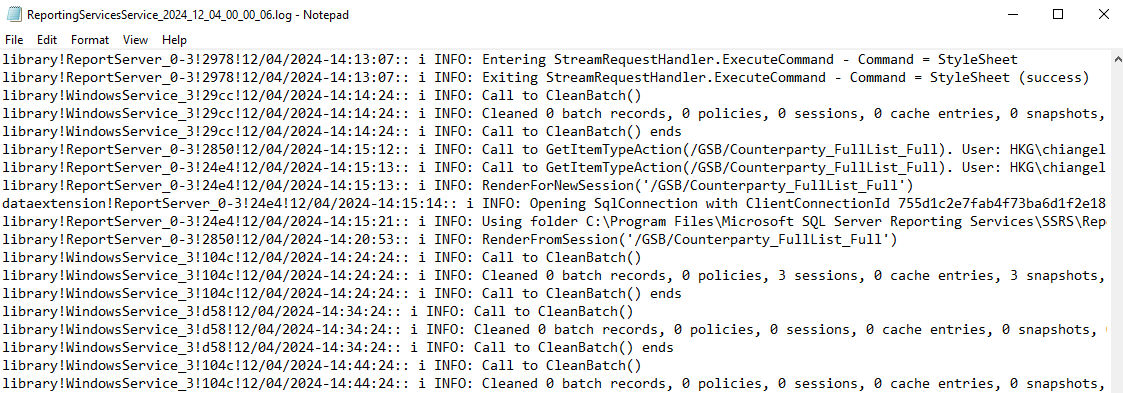
1. Use remote desktop connection to connect to the Report Server.

**DEV**: hkwninsqld041v  
**PROD**: hkwninsqlp034v



1. With File Explorer, go to the folder “**C:\Program Files\Microsoft SQL Server Reporting Services\SSRS\LogFiles**”.
2. The detail logs are written to files with prefix “**ReportingServicesService\_**” and suffix in date time format “**yyyy\_MM\_dd\_HH\_mm\_ss**”.





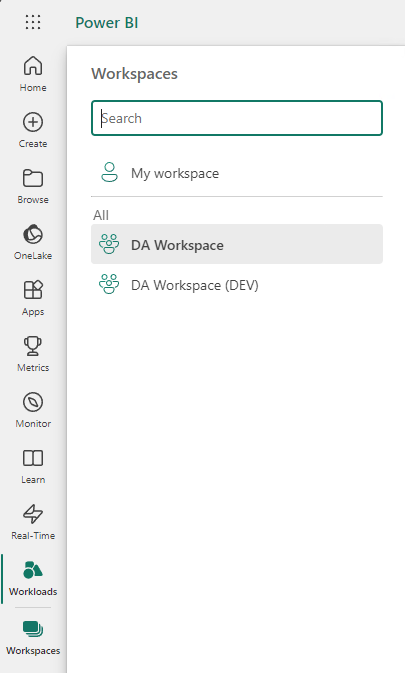
# Power BI Services

## Refresh a Semantic Model (Dataset)

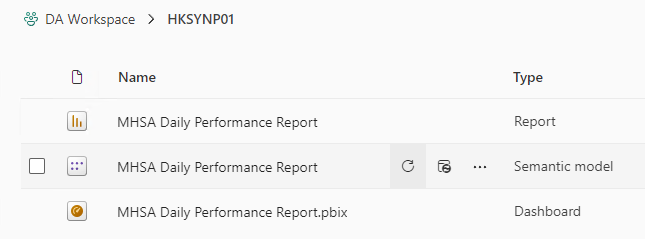
1. Access the Power BI services portal, <https://app.powerbi.com/>, with a web browser.
2. By using the navigation bar on the left-hand side, go to the Workspace where the semantic model located.

DEV: DA Workspace (DEV)

PROD: DA Workspace



1. Hover on the semantic model and click the “**Refresh now**“ button to trigger a manual refresh.

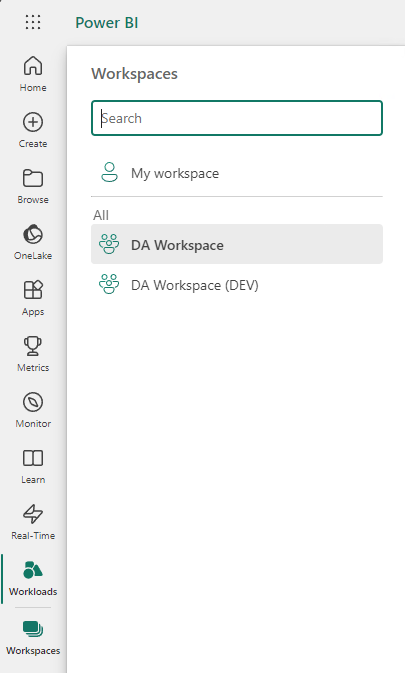


## Update the Content of a Power BI App

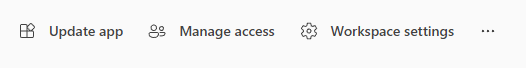
1. Access the Power BI services portal, <https://app.powerbi.com/>, with a web browser.
2. By using the navigation bar on the left-hand side, go to the Workspace where the Power BI App being created and shared.

DEV: DA Workspace (DEV)

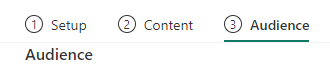
PROD: DA Workspace

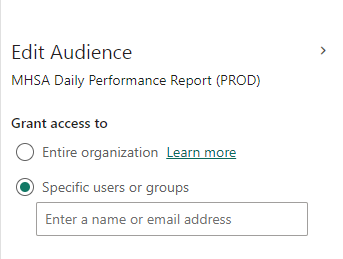


1. Click on “Update app” on the upper right corner to update the Power BI app.

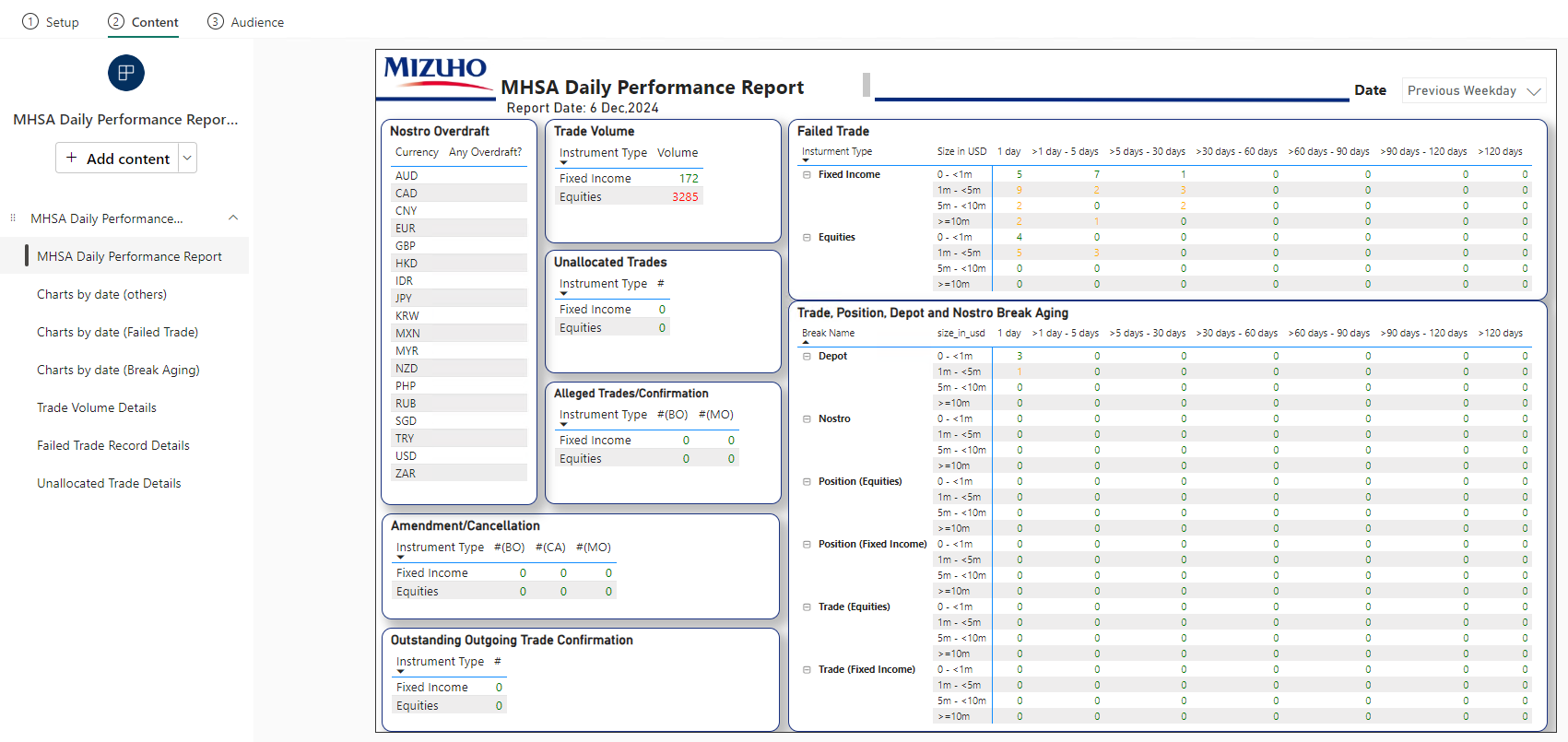


1. If there are new audience need to grant the access to the Power BI app or a user found he/she does not have access to the Power BI App in their Power BI services portal, then go to the “**Audience**” tab and grant the access to the user by using the “**Edit Audience**” pane on the right-hand side of the page.





1. If the dashboard has been updated, then go to the “Content” tab and check the content of the dashboard. When a dashboard has been updated, say for instance added additional visual or change on the filter condition, a manual update on the Power BI App is required.



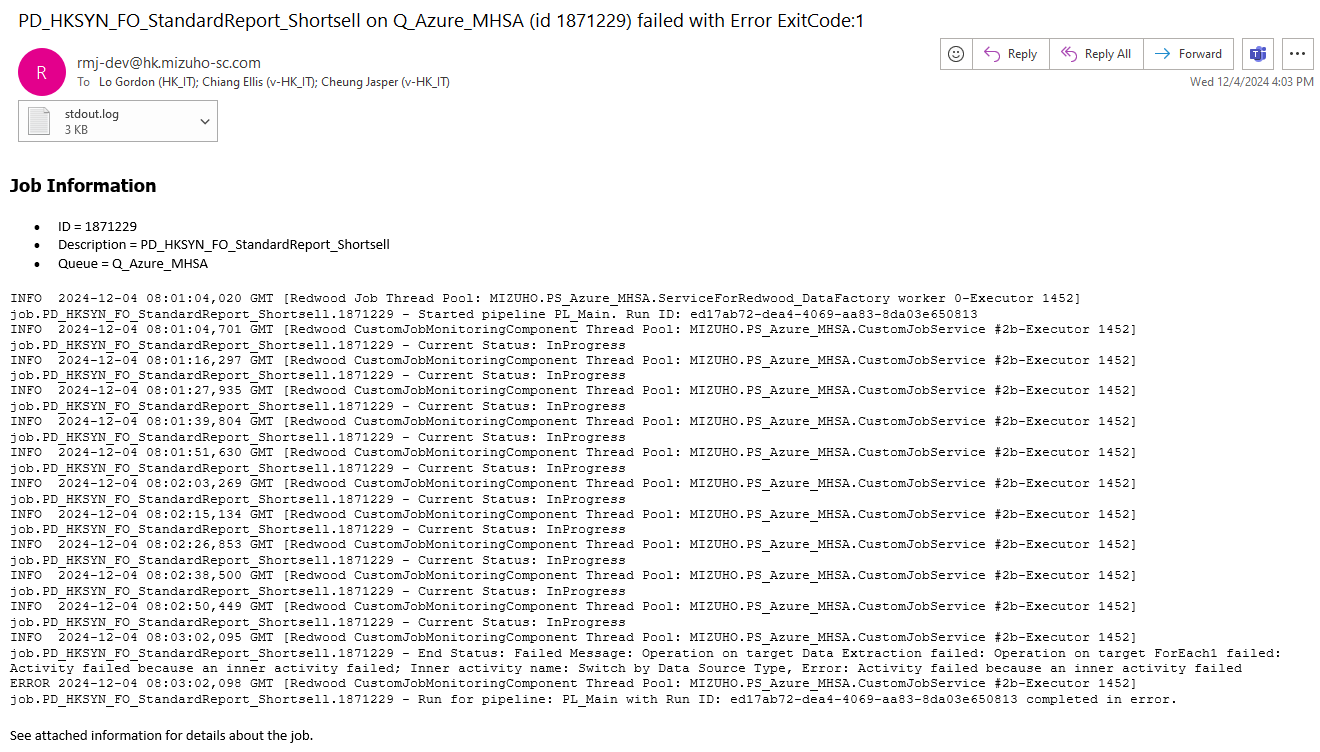
1. After completing the update on the either content or audience, click on the “Update App” button on the lower right corner to apply the update on the Power BI App.



# Troubleshooting

## Drill Down from RMJ Alert to Data Factory Monitor with Run ID

1. When a RMJ job fails, an email alert is triggered. This email includes a Run ID, which matches the Data Factory pipeline Run ID.



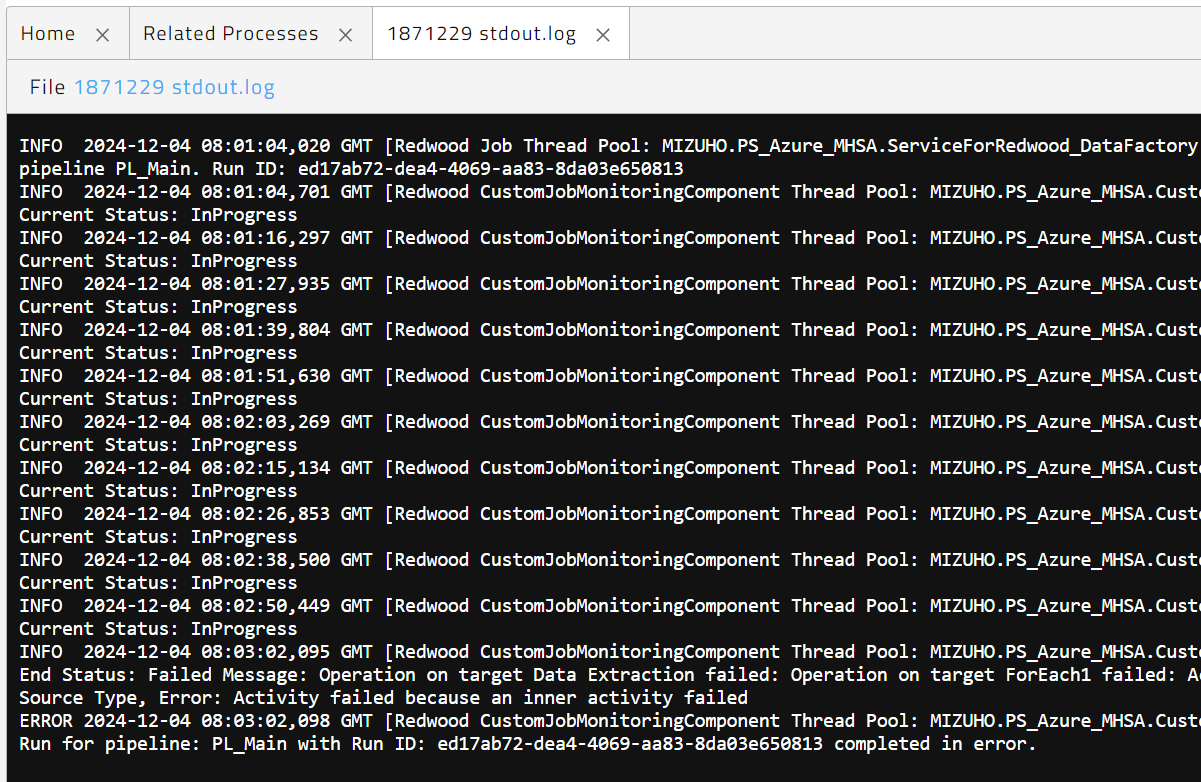
1. An alternative way to retrieve the Run ID of the pipeline is through the RMJ console. In the RMJ console, navigate to the process (job) which encountered an error during execution.



1. In the process details pane, scroll down to “**Files**” and click on “**stdout.log**” to open the log.



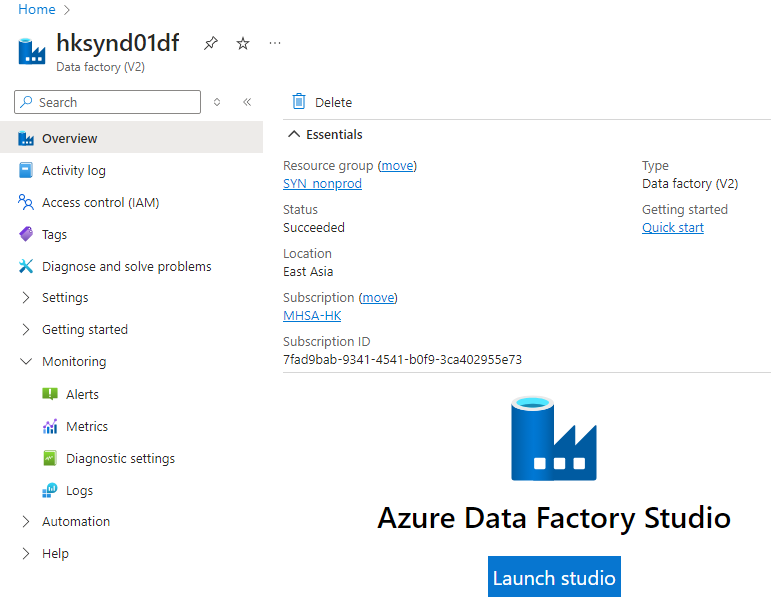
1. The pipeline Run ID can then be located from the last entry of the log.



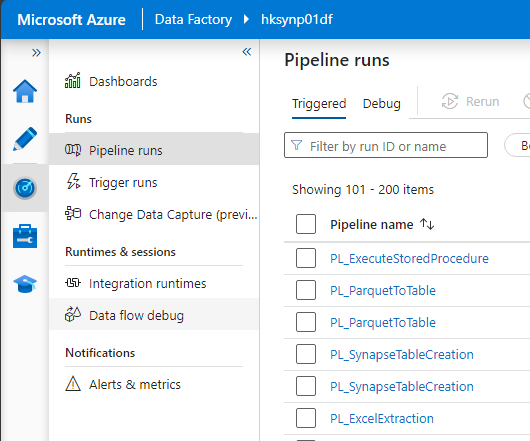
1. In Azure Portal, navigate to **Azure Data Factory**.
2. Access the target Azure Data Factory by clicking the name from the list.

**DEV**: hksynd01-df  
**PROD**: hksynp01-df

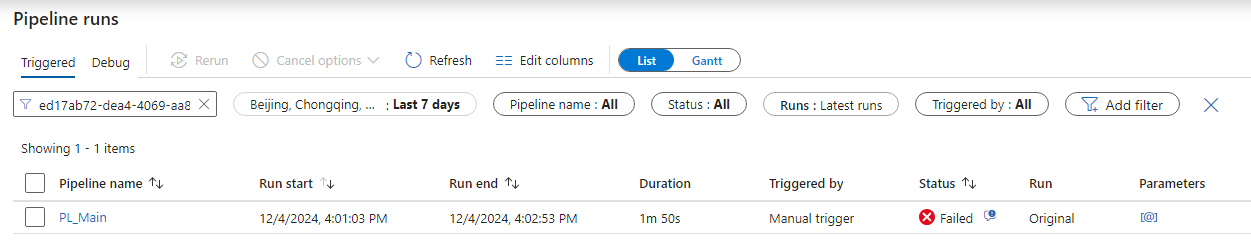
1. Click on “**Launch studio**” and Azure Data Factory Studio will be started in a new browser tab.



1. Go to “**Monitor**” in the side menu, select “**Pipeline runs**” from the Table of content pane and click on the “**Triggered**” tab.



1. To drill down the pipeline that encountered error, copy & paste the Run ID in the search box.

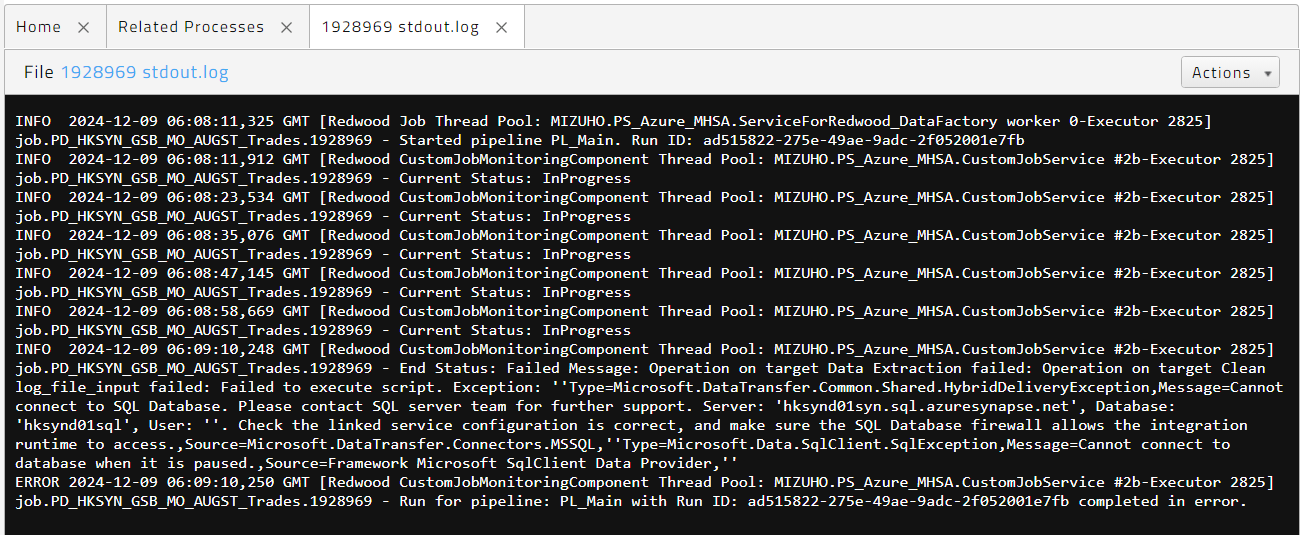


1. Follow the steps in section 3.2. to drill down the error of a sub-pipeline.

## Common Issue

### Repository Database is Inactive/Paused

The report creation pipeline in the Data Factory has been configured to use the Synapse Dedicated SQL Pool as the report job repository. The pipeline cannot execute without accessing to the report job repository. From the log message (refer to section 2.4 for access to the job log), it stated that the Synapse Dedicated SQL Pool, hksynd01syn.sql.azuresynpase.net, is out of reach and caused the pipeline failure.

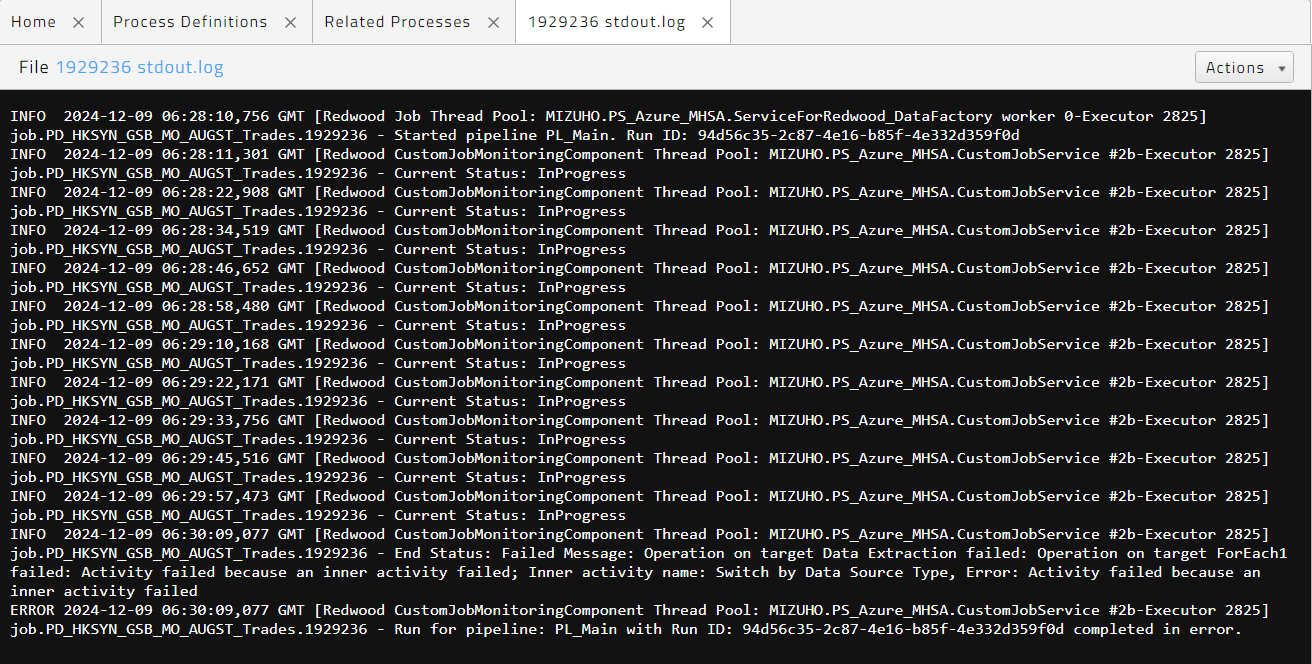


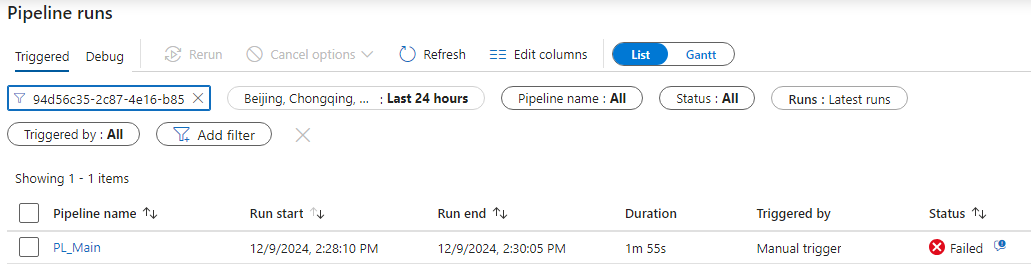
Possible cause of error:

* The Synapse Dedicated SQL Pool was paused.
* The database connection configuration is incorrect.
* Service outage of the Synapse Dedicated SQL Pool.

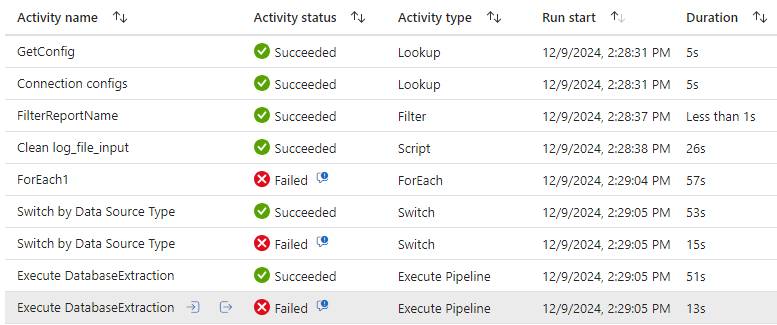
### Failed to Connect to Data Sources

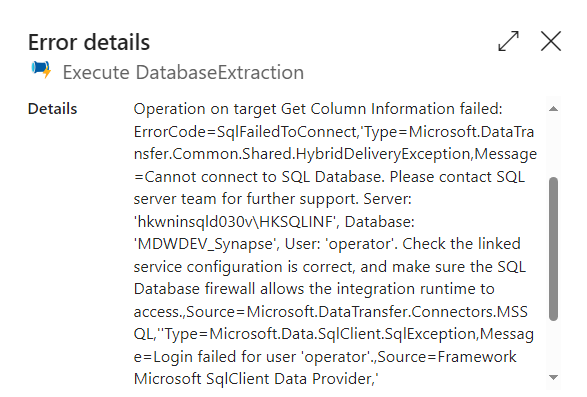
Since the report pipeline needs to copy data from source database resided in on-premises database system to the Azure Synapse Dedicated SQL Pool for data processing and create reports, failure to connect to the source database will cause the pipeline to fail. From the log message (refer to section 2.4 for access to the job log), the message is not clear enough to identify the cause of the issue as the error is from the sub-pipeline. We will need to use the Run ID from the log and drill down to the sub-pipeline for the root error message as described in section 3.2.





After drill down to the sub-pipeline, the error was encountered during database extraction. Click on the information button at the Failed status and it will show the error message returned when the pipeline trying to connect to the source database. In the example below, the cause of the error was login failure for the specified user account.



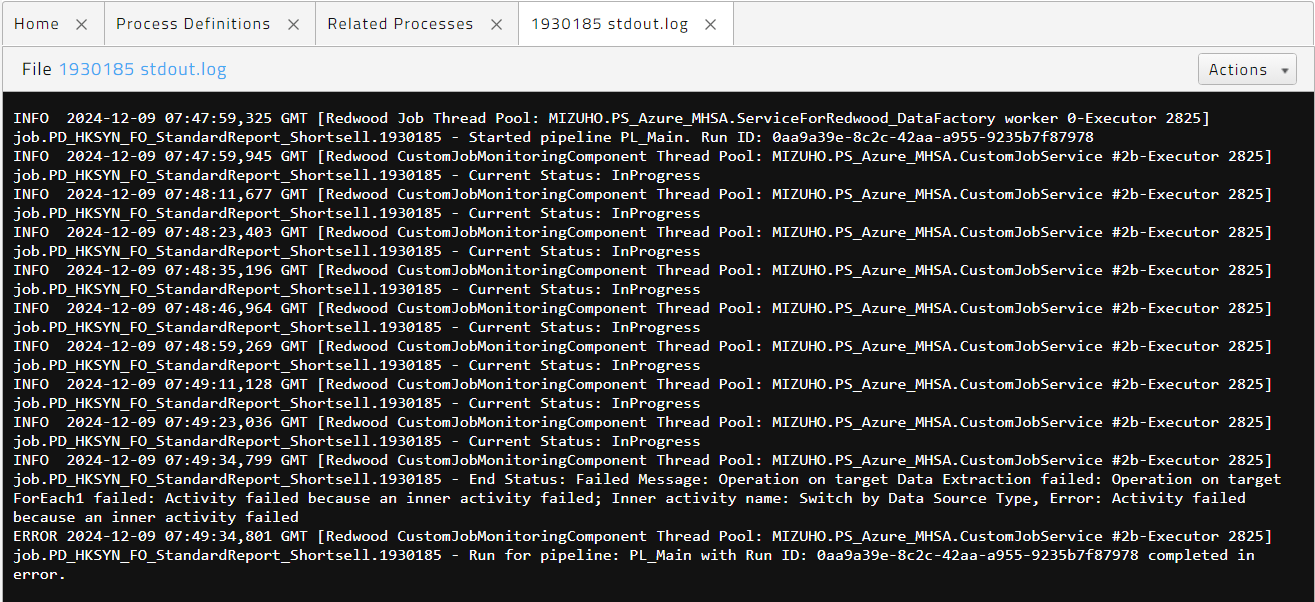


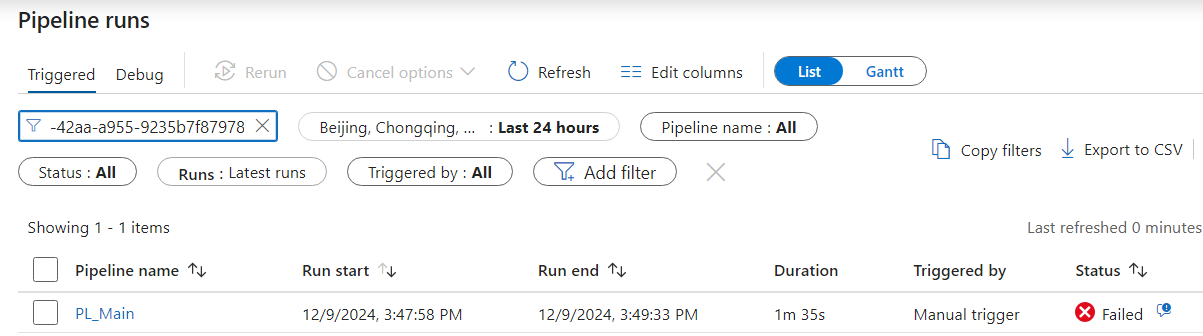
Possible cause of error:

* Login credential is incorrect.
* The source database is out of services.
* The database connection configuration is incorrect.

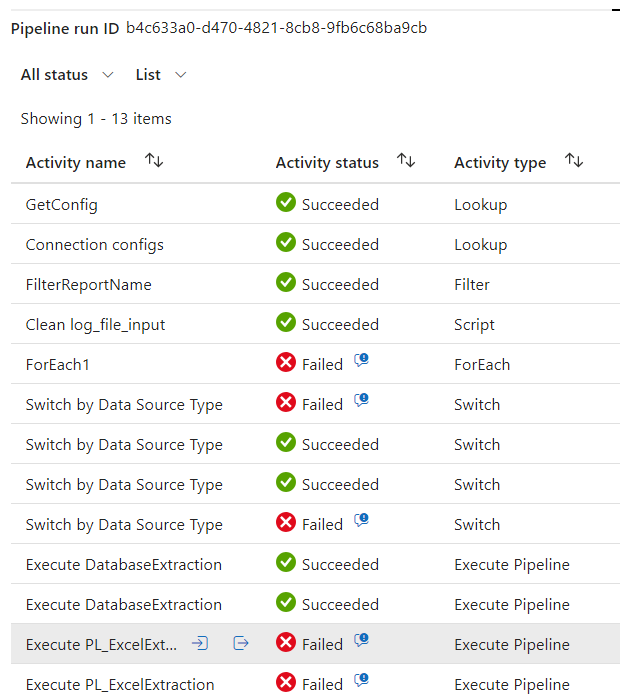
### Missing Input Files in Shared Folder

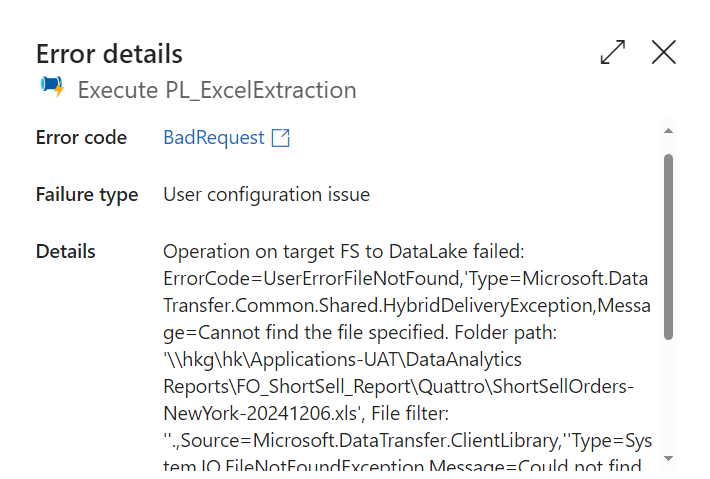
Some of the reports rely on input files generated from other system and placed in the shared folder. Failure to read the input files specified for the report job will cause the pipeline to fail. From the log message (refer to section 2.4 for access to the job log), the message is not clear enough to identify the cause of the issue as the error is from the sub-pipeline. We will need to use the Run ID from the log and drill down to the sub-pipeline for the root error message as described in section 3.2.





After drill down to the sub-pipeline, the error was encountered during Excel file extraction. Click on the information button at the Failed status and it will show the error message returned when the pipeline trying to access the specified input file from the shared folder. In the example below, the cause of the error was the file not found in the specified path.



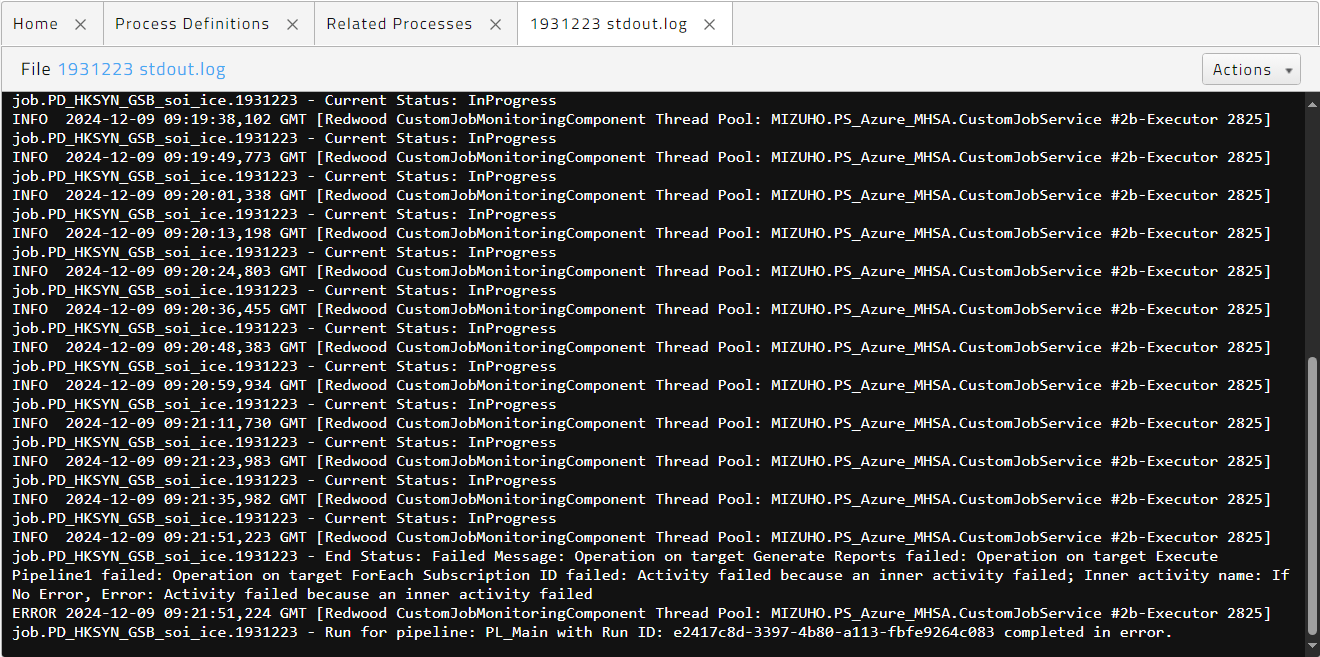


Possible cause of error:

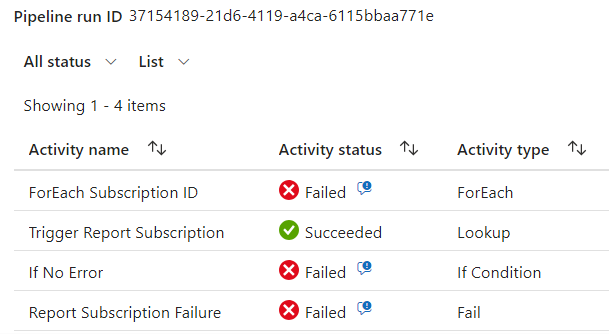
* The input file does not exist.
* Access denied to the input file path.
* The credential for accessing the shared folder is not valid.

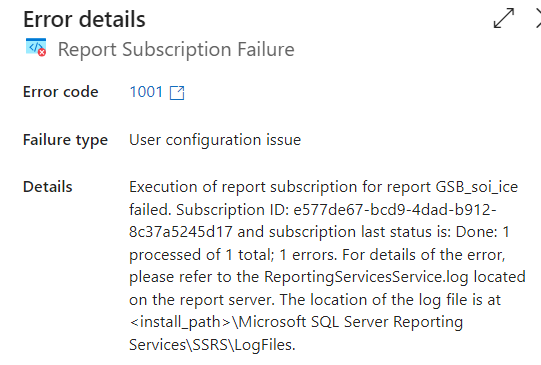
### Report Subscription Execution Failure

The pipeline will fail when the report generation failed in the SQL Server Reporting Services (SSRS). From the log message (refer to section 2.4 for access to the job log), the message is not clear enough to identify the cause of the issue as the error is from the sub-pipeline. We will need to use the Run ID from the log and drill down to the sub-pipeline for the root error message as described in section 3.2.

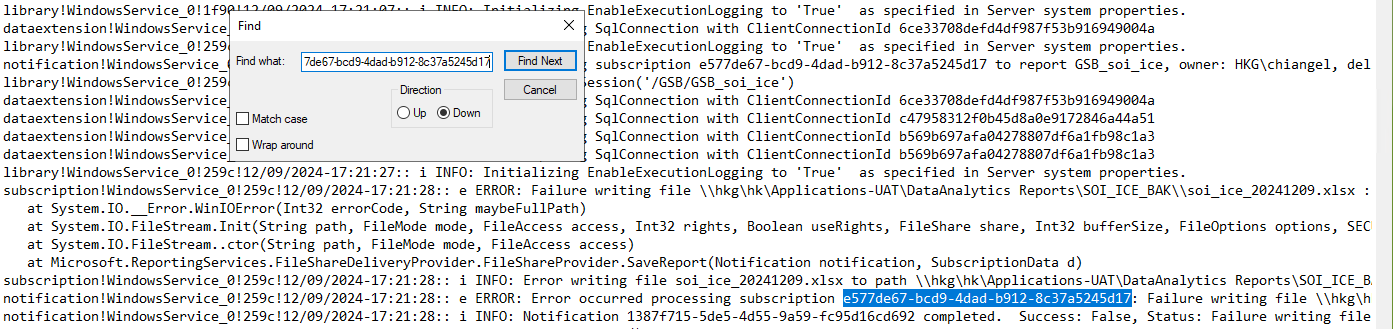


After drill down to the sub-pipeline, the error was encountered during executing the report subscription. Click on the information button at the Failed status and it will show the error message returned when the pipeline trying to generate a report with the SSRS. For retrieving the actual error message from SSRS, please refer to section 5.4 for the procedures to access to the reporting services log file.





In the error message from the Data Factory information box, a corresponding subscription ID of the report will be noted in the error details. One can use this subscription ID to search for related entries in the report server reporting services log. In the example below, the issue was caused by the SSRS failed to write the output file to the target output path.

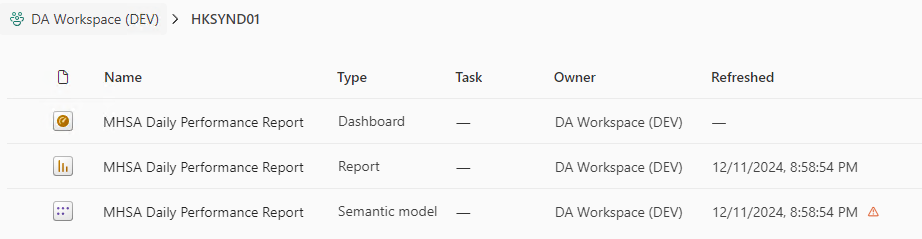


Possible cause of error:

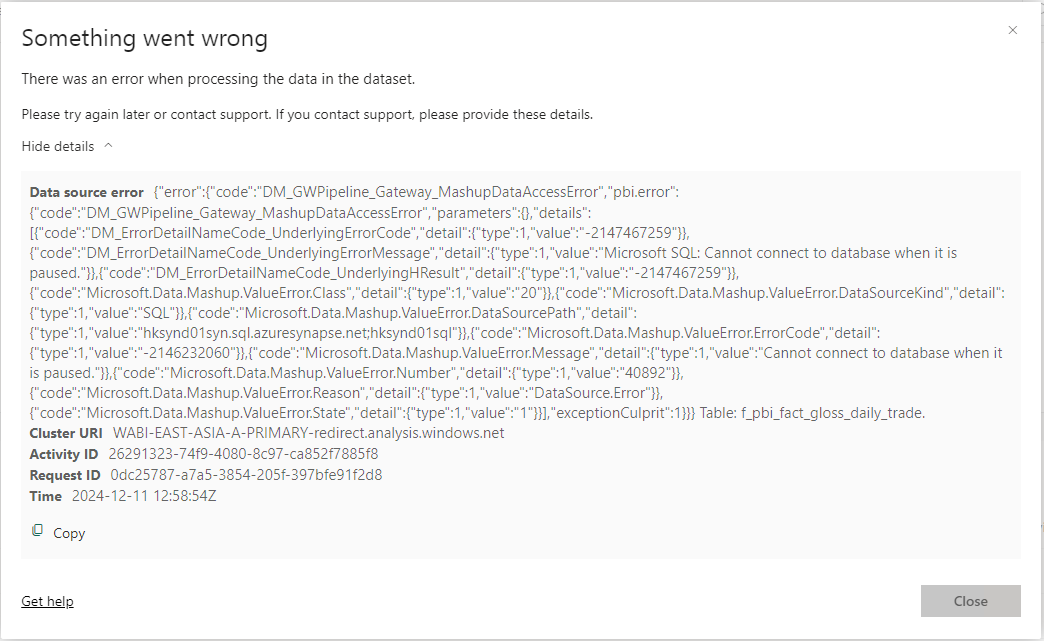
* The target folder does not exist or no permission to access to the target folder.
* The report template cannot connect to the source data.
* The report template is not valid or corrupted.
* A non-reporting services error caused by the operating system.

### Power BI Semantic Model Refresh Failure

When encountered an error during refreshing the Power BI semantic model, an exclamation mark will be shown at the back of the refresh timestamp.



Click on the exclamation mark and the details of the error will be displayed in a message box. In the example below the issue was caused by Power BI failed to connect to the source database to retrieve and process the data.



Possible cause of error:

* The target database is inactive and not providing service.
* The login credential to the database is incorrect.
* The data gateway cannot be connected.