Portfolio Project:

Monitoring a Data Warehouse in Microsoft Fabric

Project Overview

This project focused on monitoring and querying a data warehouse in Microsoft Fabric, which provided a relational database for large-scale analytics. The objective was to utilize dynamic management views (DMVs) and query insights to monitor activity and analyze data effectively.

Objectives:

1. **Create a Workspace:**
   - Set up a workspace with Fabric trial enabled.
2. **Create a Sample Data Warehouse:**
   - Establish a sample data warehouse for analysis.
3. **Explore Dynamic Management Views:**
   - Use DMVs to monitor connections, sessions, and requests.
4. **Query the Data Warehouse:**
   - Execute various SQL queries to analyze data.
5. **Verify Data Consistency:**
   - Check for and handle inconsistent data.
6. **Save as View:**
   - Create a view for filtered data reporting.

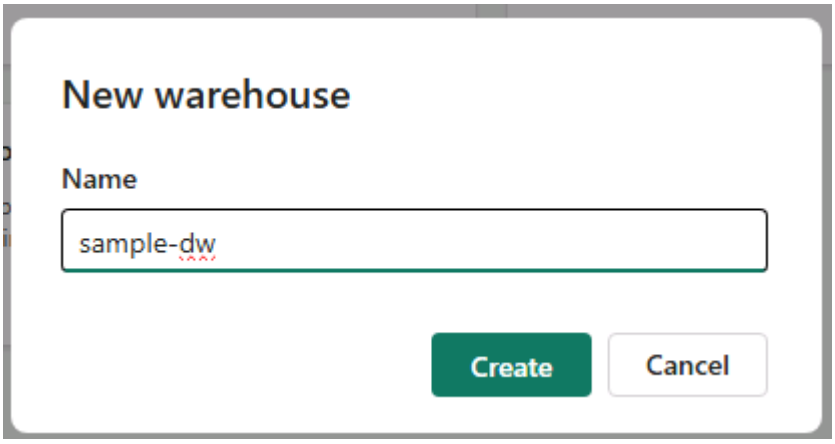
Project

Create a Workspace

- Navigated to the Microsoft Fabric home page and signed in with credentials.
- Selected Workspaces from the menu bar and created a new workspace.

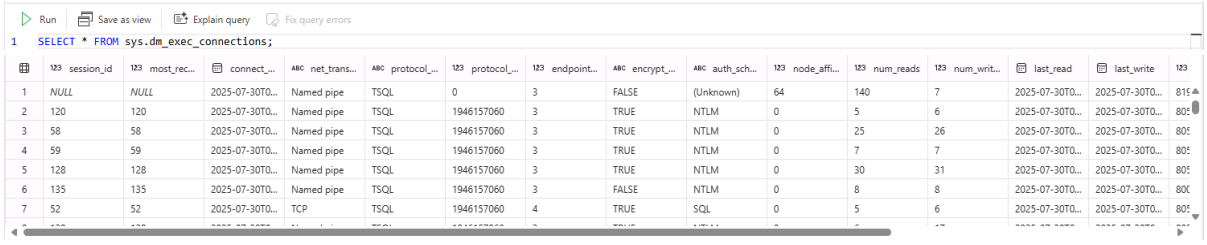
Create a Sample Data Warehouse

- Selected Create, then Sample warehouse, and created a new data warehouse named ‘sample-dw’.

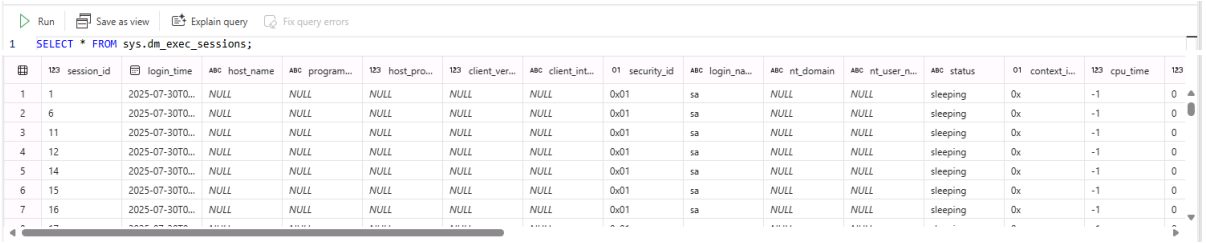


Explore Dynamic Management Views

- Query connections:



- Query sessions:



- Query requests:

	session_id	request_id	start_time	status	command	sql_handle	statemen...	statemen...	plan_han...	database...	user_id	connecti...	blocking...	wait_type	
1	1	0	2025-07-30T0...	sleeping	TASK MANAG...	NULL	NULL	NULL	NULL	1	1	NULL	0	NULL	0
2	6	0	2025-07-30T0...	sleeping	TASK MANAG...	NULL	NULL	NULL	NULL	1	1	NULL	0	NULL	0
3	11	0	2025-07-30T0...	sleeping	TASK MANAG...	NULL	NULL	NULL	NULL	1	1	NULL	0	NULL	0
4	12	0	2025-07-30T0...	sleeping	TASK MANAG...	NULL	NULL	NULL	NULL	1	1	NULL	0	NULL	0
5	14	0	2025-07-30T0...	sleeping	TASK MANAG...	NULL	NULL	NULL	NULL	1	1	NULL	0	NULL	0
6	15	0	2025-07-30T0...	sleeping	TASK MANAG...	NULL	NULL	NULL	NULL	1	1	NULL	0	NULL	0
7	16	0	2025-07-30T0...	sleeping	TASK MANAG...	NULL	NULL	NULL	NULL	1	1	NULL	0	NULL	0

- Join DMVs for running requests:

```
SELECT connections.connection_id,
sessions.session_id, sessions.login_name, sessions.login_time,
requests.command, requests.start_time, requests.total_elapsed_time
FROM sys.dm_exec_connections AS connections
INNER JOIN sys.dm_exec_sessions AS sessions
   ON connections.session_id=sessions.session_id
INNER JOIN sys.dm_exec_requests AS requests
   ON requests.session_id = sessions.session_id
WHERE requests.status = 'running'
   AND requests.database_id = DB_ID()
ORDER BY requests.total_elapsed_time DESC;
```

Messages Results

	connection_id	session_id	login_name	login_time	command	start_time	total_elapsed_time
1	F6D30735-F76B-4529-B5BB-0829...	53	User1-53405250@LODSPRODMC...	2025-07-30T07:46:39.447	SELECT	2025-07-30T07:46:39.450	9

```
SELECT connections.connection_id,
sessions.session_id, sessions.login_name, sessions.login_time,
requests.command, requests.start_time, requests.total_elapsed_time
FROM sys.dm_exec_connections AS connections
INNER JOIN sys.dm_exec_sessions AS sessions
   ON connections.session_id=sessions.session_id
INNER JOIN sys.dm_exec_requests AS requests
   ON requests.session_id = sessions.session_id
WHERE requests.status = 'running'
   AND requests.database_id = DB_ID()
ORDER BY requests.total_elapsed_time DESC;
```

Messages Results

	connection_id	session_id	login_name	login_time	command	start_time	total_elapsed_time
1	B8D7E78A-D2AC-4AA1-99BB-82...	139	User1-53405250@LODSPRODMC...	2025-07-30T07:50:20.433	SELECT	2025-07-30T07:50:20.450	28

Run Save as view Explain query Fix query errors

```
SELECT connections.connection_id,
sessions.session_id, sessions.login_name, sessions.login_time,
requests.command, requests.start_time, requests.total_elapsed_time
FROM sys.dm_exec_connections AS connections
INNER JOIN sys.dm_exec_sessions AS sessions
   ON connections.session_id=sessions.session_id
INNER JOIN sys.dm_exec_requests AS requests
   ON requests.session_id = sessions.session_id
WHERE requests.status = 'running'
   AND requests.database_id = DB_ID()
ORDER BY requests.total_elapsed_time DESC;
```

Messages Results

	connection_id	session_id	login_name	login_time	command	start_time	total_elapsed_time
1	14832F5C-7FD7-495D-AB9E-65D...	55	User1-53405250@LODSPRODMC...	2025-07-30T07:47:47.223	SELECT	2025-07-30T07:47:47.227	17036
2	0E2094A8-5114-4A08-8B54-F416...	179	User1-53405250@LODSPRODMC...	2025-07-30T07:48:04.243	SELECT	2025-07-30T07:48:04.247	53

	database_name	median_total_elapse...	last_run_total_elaps...	last_run_start_time	last_dist_statement_id	last_run_session_id	number_of_runs	query_hash	last_run_command
1	sample-dw	32603	18464	2025-07-30 07:47:47.220...	2A6E90EF-23CE-4346-98...	55	3	0x29EBCE2B988E151C	SELECT * FROM Trip
2	sample-dw	842	842	2025-07-30 07:41:35.463...	8AD1E5B2-5A32-4E0A-A...	56	1	0x43E4F3D439F30C75	COPY INTO [dbo].[Time]...
3	sample-dw	1046	1046	2025-07-30 07:41:34.400...	603C2B6B-8207-4719-9...	56	1	0x4A7FDB0D04BF3F63	COPY INTO [dbo].[Medal...
4	sample-dw	1158	1158	2025-07-30 07:41:33.223...	AD8663A8-0BE9-47D6-B...	56	1	0x57338B4B7612E76B	COPY INTO [dbo].[Hackn...
5	sample-dw	15742	15742	2025-07-30 07:41:37.743...	335A6882-5AA6-4FE9-9...	56	1	0x5D65C730D405D935	COPY INTO [dbo].[Trip]#...
6	sample-dw	2043	2043	2025-07-30 07:41:31.160...	F74D4568-F350-43E7-A...	56	1	0x82F02BDAE2A3287F	COPY INTO [dbo].[Geogr...
7	sample-dw	1414	1414	2025-07-30 07:41:36.323...	67014084-1B46-4E4E-A2...	56	1	0x85657FCCFD685B55	COPY INTO [dbo].[Weath...
8	sample-dw	7274	7274	2025-07-30 07:41:23.893...	47478F35-1E85-45BB-A6...	56	1	0x96E9E7D49E344340	COPY INTO [dbo].[Date]...

Explore Query Insights

- Query executed requests:

```
SELECT * FROM queryinsights.exec_requests_history;
```

- Query frequently run queries:

```
SELECT * FROM queryinsights.frequently_run_queries;
```

	database_n...	number_of...	min_run_to...	max_run_t...	avg_total_e...	number_of...	number_of...	number_of...	last_run_to...	last_run_st...	last_dist_st...	query_hash	last_run_co...
1	sample-dw	3	18464	47532	32866	3	0	0	18464	2025-07-30 07:...	2A6E90EF-23CE...	0x29EBCE2B98...	SELECT * FROM...
2	sample-dw	1	842	842	842	1	0	0	842	2025-07-30 07:...	8AD1E5B2-5A3...	0x43E4F3D439F...	COPY INTO [db...
3	sample-dw	1	1046	1046	1046	1	0	0	1046	2025-07-30 07:...	603C2B6B-820...	0x4A7FDB0D04...	COPY INTO [db...
4	sample-dw	1	1158	1158	1158	1	0	0	1158	2025-07-30 07:...	AD8663A8-0BE...	0x57338B4B76...	COPY INTO [db...
5	sample-dw	1	15742	15742	15742	1	0	0	15742	2025-07-30 07:...	335A6882-5AA...	0x5D65C730D4...	COPY INTO [db...
6	sample-dw	1	2043	2043	2043	1	0	0	2043	2025-07-30 07:...	F74D4568-F350...	0x82F02BDAE2...	COPY INTO [db...
7	sample-dw	1	1414	1414	1414	1	0	0	1414	2025-07-30 07:...	67014084-1B46...	0x85657FCCFD...	COPY INTO [db...
8	sample-dw	1	7274	7274	7274	1	0	0	7274	2025-07-30 07:...	47478F35-1E85...	0x96E9E7D49E...	COPY INTO [db...

- Query long-running queries:

```
SELECT * FROM queryinsights.long_running_queries;
```

	database_name	median_total_elapse...	last_run_total_elaps...	last_run_start_time	last_dist_statement_id	last_run_session_id	number_of_runs	query_hash	last_run_command
1	sample-dw	32603	18464	2025-07-30 07:47:47.220...	2A6E90EF-23CE-4346-98...	55	3	0x29EBCE2B988E151C	SELECT * FROM Trip
2	sample-dw	842	842	2025-07-30 07:41:35.463...	8AD1E5B2-5A32-4E0A-A...	56	1	0x43E4F3D439F30C75	COPY INTO [dbo].[Time]...
3	sample-dw	1046	1046	2025-07-30 07:41:34.400...	603C2B6B-8207-4719-9...	56	1	0x4A7FDB0D04BF3F63	COPY INTO [dbo].[Medal...
4	sample-dw	1158	1158	2025-07-30 07:41:33.223...	AD8663A8-0BE9-47D6-B...	56	1	0x57338B4B7612E76B	COPY INTO [dbo].[Hackn...
5	sample-dw	15742	15742	2025-07-30 07:41:37.743...	335A6882-5AA6-4FE9-9...	56	1	0x5D65C730D405D935	COPY INTO [dbo].[Trip]...
6	sample-dw	2043	2043	2025-07-30 07:41:31.160...	F74D4568-F350-43E7-A...	56	1	0x82F02BDAE2A3287F	COPY INTO [dbo].[Geogr...
7	sample-dw	1414	1414	2025-07-30 07:41:36.323...	67014084-1B46-4E4E-A2...	56	1	0x85657FCCFD685B55	COPY INTO [dbo].[Weath...
8	sample-dw	7274	7274	2025-07-30 07:41:23.893...	47478F35-1E85-45BB-A6...	56	1	0x96E9E7D49E344340	COPY INTO [dbo].[Date]...

Query the Data Warehouse

- Opened a new SQL query and run the following code to get total trips and revenue by month:

```
SELECT
D.MonthName,
COUNT(*) AS TotalTrips,
SUM(T.TotalAmount) AS TotalRevenue
FROM dbo.Trip AS T
JOIN dbo.[Date] AS D
    ON T.[DateID]=D.[DateID]
GROUP BY D.MonthName;
```

Messages Results Search

	MonthName	TotalTrips	TotalRevenue
1	September	231384	3562548
2	December	233767	3540532
3	November	236513	3547993
4	June	238432	3627186
5	May	249340	3794240
6	January	237842	3401334
7	April	246153	3679261
8	July	228327	3417355

- Ran another query to get average trip duration and distance by day of the week:

```sql
SELECT
D.DayName,
AVG(T.TripDurationSeconds) AS AvgDuration,
AVG(T.TripDistanceMiles) AS AvgDistance
FROM dbo.Trip AS T
JOIN dbo.[Date] AS D
    ON T.[DateID]=D.[DateID]
GROUP BY D.DayName;
```

Messages Results Search

	DayName	AvgDuration	AvgDistance
1	Tuesday	763	2.81239639559921
2	Thursday	796	19.3689186988126
3	Friday	790	2.84746039283699
4	Sunday	1147	3.21572794977795
5	Saturday	835	33.0886460668962
6	Monday	742	6.82623660465939
7	Wednesday	780	2.8184765031265

- Queried the top 10 most popular pickup and dropoff locations:

```sql
SELECT TOP 10
    G.City,
    COUNT(*) AS TotalTrips
FROM dbo.Trip AS T
JOIN dbo.Geography AS G
    ON T.DropoffGeographyID=G.GeographyID
GROUP BY G.City
ORDER BY TotalTrips DESC;
```

Messages Results Search

	City	TotalTrips
1	Manhattan	1523186
2	New York	595638
3	Brooklyn	153569
4	Prince	92711
5	Flushing	80533
6	Queens	58181
7	Planetarium	52907
8	Brooklyn Heights	47909

Verify Data Consistency

- Checked for trips with unusually long duration:

```sql
-- Check for trips with unusually long duration
SELECT COUNT(*) FROM dbo.Trip WHERE TripDurationSeconds > 86400; -- 24 hours
```

	untitled1
1	49

- Checked for trips with negative trip duration:

```sql
-- Check for trips with negative trip duration
SELECT COUNT(*) FROM dbo.Trip WHERE TripDurationSeconds < 0;
```

	untitled1
1	4

- Removed trips with negative trip duration:

```sql
-- Remove trips with negative trip duration
DELETE FROM dbo.Trip WHERE TripDurationSeconds < 0;
```

Results:

- ✓ A workspace was successfully created in Microsoft Fabric.
- ✓ A sample data warehouse named sample-dw was established and populated with sample data for analysis.
- ✓ Dynamic Management Views (DMVs) were utilized to monitor:
  - Active connections to the data warehouse.
  - Current authenticated sessions.
  - Ongoing requests and their statuses.
- ✓ SQL queries were executed to analyze data, including:
  - Total trips and revenue by month.
  - Average trip duration and distance by day of the week.
  - Identification of the top 10 most popular pickup and drop-off locations.
- ✓ Data consistency checks were performed, revealing:
  - Trips with unusually long durations and negative trip durations were identified and handled appropriately.
- ✓ A view named vw_JanTrip was created to filter and report on January trip data.

Conclusion

This project provided a comprehensive introduction to monitoring and querying a data warehouse in Microsoft Fabric. Key insights were gained into the use of dynamic management views for real-time monitoring of database activity and the execution of SQL queries for data analysis. Overall, this exercise demonstrated the capabilities of Microsoft Fabric in managing and analyzing large-scale data efficiently.

Resources

GitHub: https://github.com/ThatoMTNG/Microsoft-Fabric-Analytics-Engineer-DP-600-

Mentions

Project Author: Thato Metsing (https://www.linkedin.com/in/thatometsing/)

Project Mentor: Maureen Direro (https://www.linkedin.com/in/maureen-direro-46a6b220/)