



Semantic Link Labs: A Link to the Future

JASON ROMANS

NASHVILLE POWER BI USER GROUP 2025

SEMANTIC LINK LABS

— A LINK TO THE FUTURE —



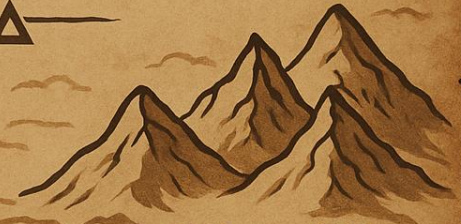
MICROSOFT
NOTEBOOK



REPORTS THAT
BREAK DUE TO
STRUCTURAL CHANGES



MODELS THAT
PERFORM POORLY
BECAUSE BEST
PRACTICES WERE
SKIPPED



USABILITY
PITFALLS THAT
MAKE REPORTS
“TECHNICALLY FINE”
BUT FUNCTIONALLY
BROKEN FOR
END USERS



GOVERNANCE
REPORT



Jason Romans

Cloud Data & Integration Developer



 Nashville, TN, USA

The DAX Shepherd



 @sql_jar

 jason-r-sql-jar

 <https://thedaxshepherd.com/>

 Began Career as a SQL Server DBA

 Transitioned to Microsoft BI Stack

 Data Engineering to Data Modeling

 Infrequent Blogger

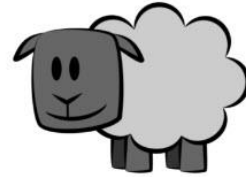
 Fan of Dimensional Models & Doctor Who

Shoulders of Giants



Slides

www.thedaxshepherd.com



The DAX Shepherd

Musings on the Microsoft BI Stack



[Home](#) [About Me](#) [Simple Talk](#) [Presentations](#) [A Speaker's Journey](#)

Presentations

Sessionize

This is my [Sessionize Profile](#) that has the conferences I have spoken at along with future events. It has a couple of my most popular sessions.

Presentation Slides

This is my [GitHub Repository](#) with the presentation slides for each event.

Recorded Sessions

Simple Talks Podcast | Episode 4 – Coffee chat with Jason Romans

About Jason Romans



I love working with the Microsoft BI Stack. I am passionate about learning.

[A Speaker's Journey](#)

Why go on this journey?

- ADMINISTRATION
- BEST PRACTICES
- DETECT ISSUES
- USE EXISTING SEMANTIC MODELS
- “HAVE YOU TRIED SEMPY?”



Our Journey



1. Intro

2. Notebooks

3. Semantic Link

4. Semantic Link Labs

5. Conclusion

Our Journey



1. Intro

2. Notebooks

3. Semantic Link

4. Semantic Link Labs

5. Conclusion

Microsoft Fabric Architecture

- Need a Fabric Capacity
- Uses what has been built in Power BI
- Data is stored in OneLake
- Choice of Compute engines

Leveraging Fabric Notebooks

- Benefit from knowledge of working with Jupyter Notebooks
 - Python
 - Pandas

Our Journey



1. Intro

2. Notebooks

3. Semantic Link

4. Semantic Link Labs

5. Conclusion

Type of Notebooks

LANGUAGE

- Python
- Scala
- Spark SQL
- R
- T-SQL

COMPUTE

- Spark
- Python (Single Node Virtual Machine)
- T-SQL Analytics



Compute & Language

 PySpark (Python)

Spark

✓ PySpark (Python)

Spark (Scala)

Spark SQL

SparkR (R)

Python

Python

T-SQL Analytics

T-SQL

Spark (Python, Scala, SQL, R)

 PySpark (Python)

Spark

✓ PySpark (Python)

Spark (Scala)

Spark SQL

SparkR (R)

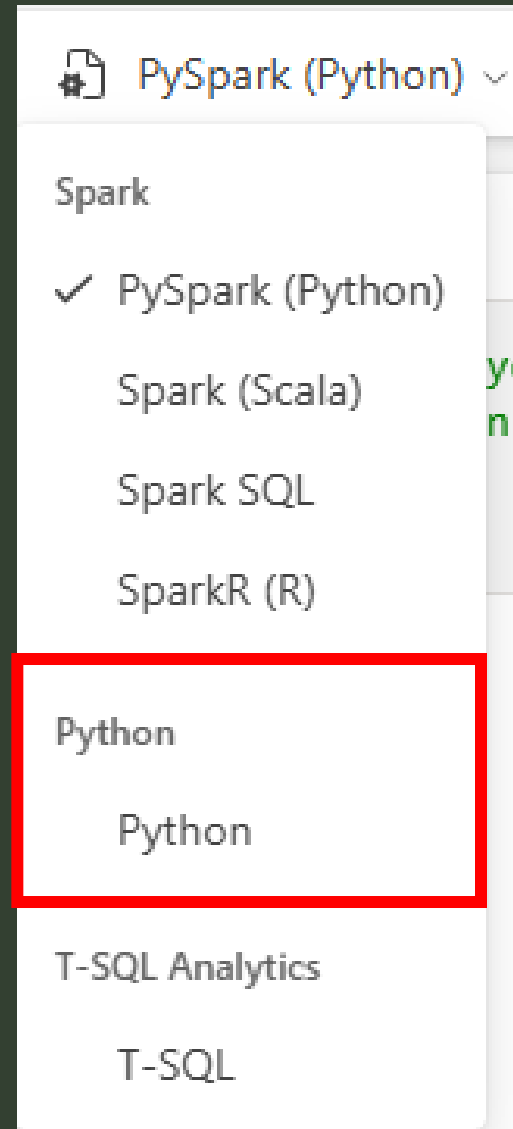
Python

Python

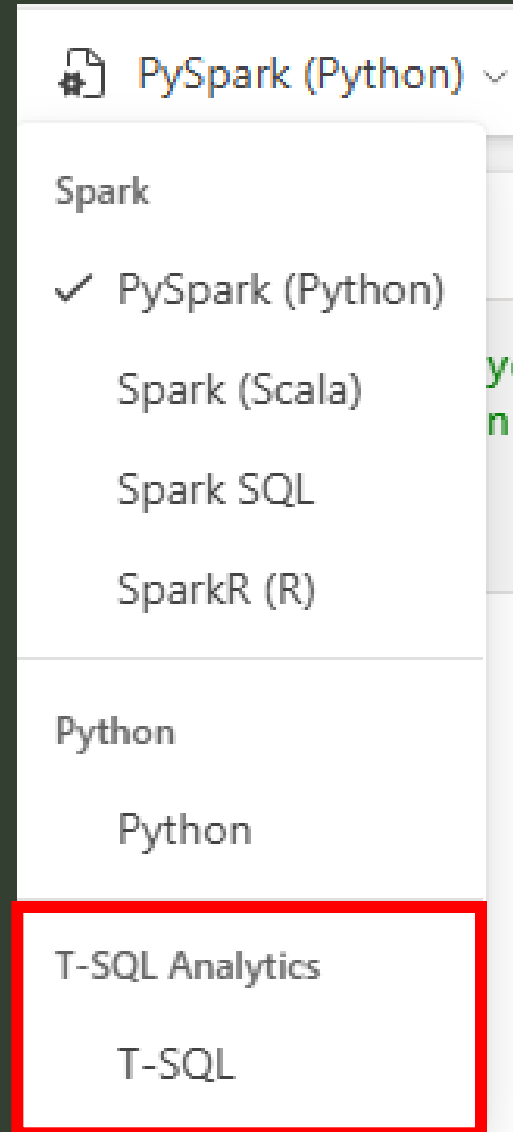
T-SQL Analytics

T-SQL

Python (Python)



T-SQL Analytics (T-SQL)



Choosing PySpark or Python Compute

Scenario	Recommended Notebook
Includes pre-installed DuckDB and Polars libraries	Python Notebooks
Small to medium data (fits in memory)	Python Notebooks (or PySpark on single-node Spark cluster)
Rapid exploration & prototyping	Python Notebooks (or PySpark on single-node Spark cluster)
Large datasets (10GB+) exceeding memory	PySpark Notebooks
Complex data workflows or ETL pipelines	PySpark Notebooks
High-concurrency or parallel execution	PySpark Notebooks
Needs Spark-native APIs (MLlib, SQL, Streaming)	PySpark Notebooks

<https://learn.microsoft.com/en-us/fabric/data-engineering/fabric-notebook-selection-guide>

Type of Compute for Notebooks

- Spark Based
 - Cluster
- Single Node Python
 - 2 vCores, 16G RAM (starter)
- T-SQL Analytics
 - Warehouse

Python Notebook - Change VM Size

Note: Longer startup times

```
%%configure
```

```
{
```

```
    "vCores": 16
```

```
}
```

- Memory is proportional to vCores
 - 2 is 2vCores with 16G
 - 4 is 4vCores with 32G
 - 8 is 8vCores with 64G
 - 16 is 16vCores with 128G
 - 32 is 32vCores with 256G
 - 64 is 64vCores with 512G
 - (32 and 64) not available in trial

Python Notebook

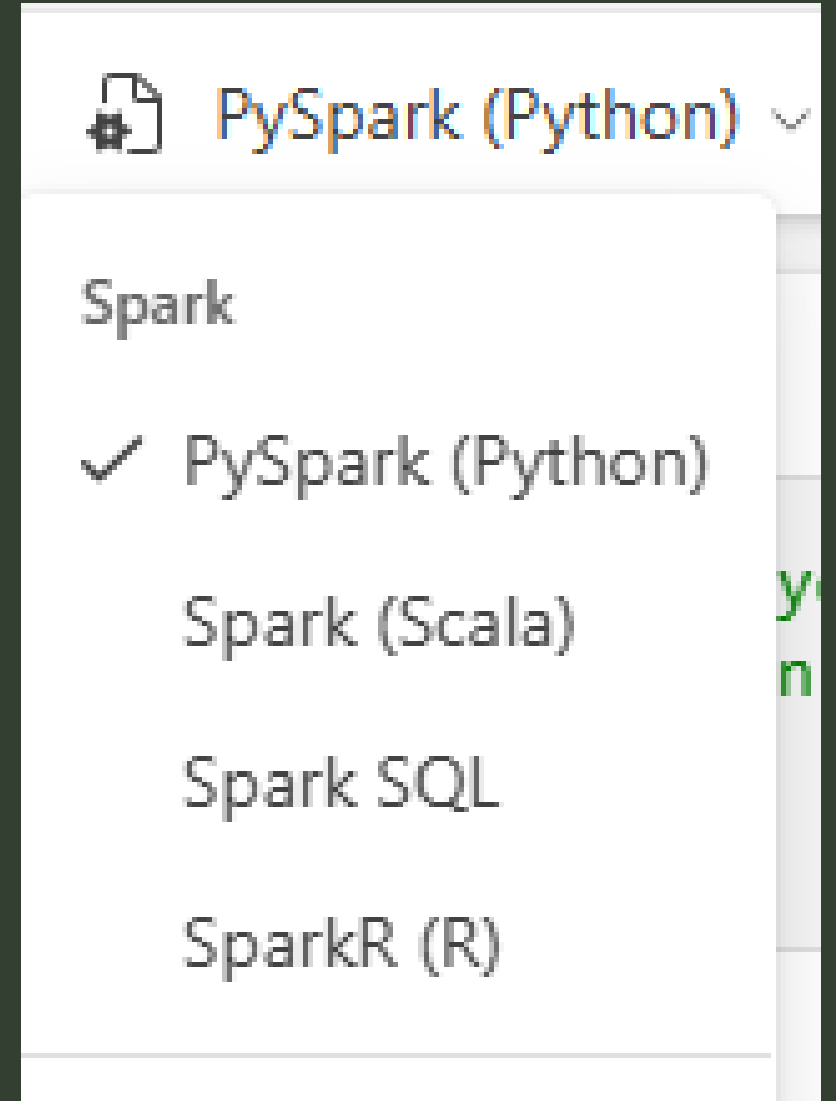
- Has libraries installed for dealing with “small-big” data
 - Less than 10 Gigabytes
 - Fits in memory
- Example Libraries installed
 - Polars
 - DuckDB

Languages for Spark

Different choices of languages

Spark was built with Scala

- PySpark (Python)
- Spark (Scala)
- Spark SQL
- SparkR (R)



Magic Commands – set language by cell

Magic command	Language	Description
%%pyspark	Python	Execute a Python query against Apache Spark Context.
%%spark	Scala	Execute a Scala query against Apache Spark Context.
%%sql	SparkSQL	Execute a SparkSQL query against Apache Spark Context.
%%html	Html	Execute a HTML query against Apache Spark Context.
%%sparkr	R	Execute a R query against Apache Spark Context.

Home

Edit

AI tools

Run

View

Comments

History

Develop

Share



Run all



Connect



PySpark (Python)



Explorer



Magic Commands



1 %%spark

Spark (Scala)



Our Journey



1. Intro

2. Notebooks

3. Semantic Link

4. Semantic Link Labs

5. Conclusion

Missing Link

Semantic link is a feature that allows you to establish a connection between semantic models and Synapse Data Science in Microsoft Fabric.

-- Reference: <https://learn.microsoft.com/en-us/fabric/data-science/semantic-link-overview>

All the Things

Semantic link is a feature that allows you to establish a connection between semantic models and Synapse Data Science in Microsoft Fabric.

semantic models

Models, Reports, Lakehouse, Workspaces and more

Synapse Data Science

Fabric Notebook – Apache Spark with Python and more

A Tale of Two Links

Both are Available Only in Microsoft Fabric

- Semantic Link
 - Base
 - Driver or API
 - Included in default runtime for the current version

A Tale of Two Links

- Semantic Link Labs
 - “Expansion Pack” -- Kurt Buhler
 - Uses Semantic Link
 - `import sempy.fabric as fabric`
 - Open source – GitHub Repository
 - Under Active Development

Semantic Link Labs

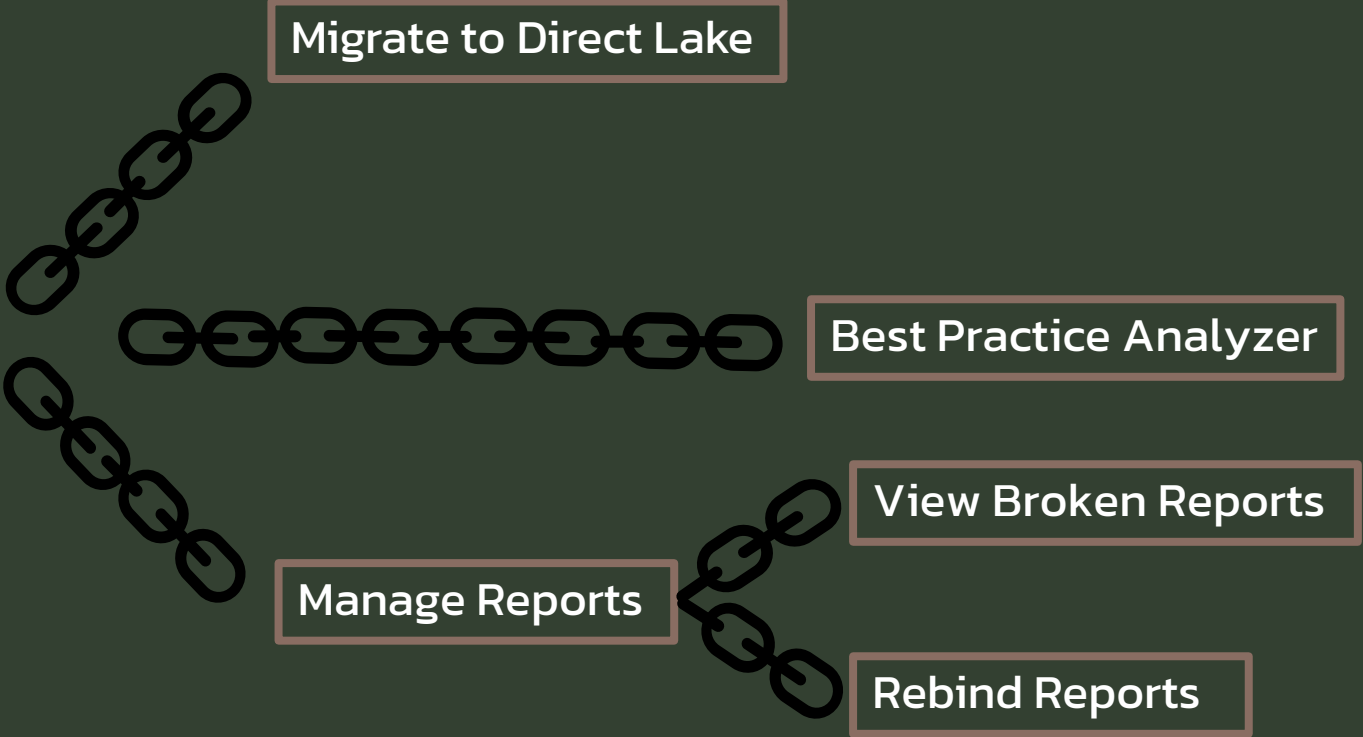
Microsoft Fabric



Semantic Link

- List Tables
- List Workspaces
- List Models
- List Reports

Semantic Link Labs



Installing

- Semantic Link
 - Spark 3.4 in default runtime
 - Update to newest version
 - %pip install -U semantic-link
- Spark 3.3 or below – need to install
 - %pip install -U semantic-link

What Version of Spark

Workspace settings

- General
- License info
- Azure connections
- System storage
- Git integration
- OneLake
- Workspace identity
- Network security
- Power BI
- Delegated Settings
- OneLake settings
- Data Engineering/Science**
- Spark settings
- Data Factory

Spark settings

Configure and manage settings for Spark workloads and the default environment for the workspace.

Pool **Environment** High concurrency Automatic log

Set default environment

The default environment will provide Spark properties, libraries, and developer settings for notebooks and Spark job definitions in this workspace when users don't select a different environment. [Learn more about Set default environment](#)

Runtime

Runtime Version

Runtime version defines which version of Spark your Spark pool will use. [Learn more about Runtime Version](#)

1.2 (Spark 3.4, Delta 2.4)

Spark Version Command

spark.version

1 spark.version

✓ <1 sec - Command executed in 260 ms by Jason Romans on 2:00:03 PM, 10/08/24

'3.4.3.5.3.20240904.5'

What Version of Semantic Link

- %pip show semantic-link
- or %pip list | grep semantic-link

```
1 %pip show semantic-link
```

[23] ✓ 4 sec - Command executed in 3 sec 588 ms by Jason Romans on 11:49:10 PM, 10/19/24 PySpark (Python) ▾

```
... Name: semantic-link
Version: 0.8.1
Summary: Semantic link for Microsoft Fabric
Home-page: https://learn.microsoft.com/en-us/fabric/data-science/semantic-link-overview
Author: Microsoft
Author-email: semanticdatascience@service.microsoft.com
License: proprietary and confidential
Location: /nfs4/pyenv-dd0ba783-2069-4fdc-8c80-bc03c74db705/lib/python3.11/site-packages
Requires: semantic-link-functions-geopandas, semantic-link-functions-holidays, semantic-link-functions-meteostat, semantic-link-functions-phonenumbers, semantic-link-functions-validators, semantic-link-sempy
Required-by:
Note: you may need to restart the kernel to use updated packages.
```

Import the Module

Give it a friendly name – think Alias

```
import sempy.fabric as fabric
```

Done in a Notebook

Code - Python

Text Descriptions - Markdown

Install the latest .whl package

Check [here](#) to see the latest version.

```
1 %pip install semantic-link-labs
```

- Session ready in 8 sec 603 ms. Command executed in 26 sec 202 ms by Jason Romans on 10:29:04 AM, 10/23/24

Install Wheel from File

```
1 %pip install /lakehouse/default/Files/semantic_link_labs-0.8.3-py3-none-any.whl
```

- Command executed in 22 sec 745 ms by Jason Romans on 10:07:06 AM, 10/23/24

Show Semantic Link and Labs installed

```
1 %pip show semantic-link-sempy
2 print('\n')
3 %pip show semantic-link-labs
```

✓ - Command executed in 6 sec 382 ms by Jason Romans on 10:30:20 AM, 10/23/24

Name: semantic-link-sempy

Version: 0.8.1

Summary: Semantic link for Microsoft Fabric

Home-page: <https://learn.microsoft.com/en-us/fabric/data-science/semantic-link-overview>

Author: Microsoft

Uses Pandas Ecosystem (DataFrame)

Knowledge of working with
Pandas DataFrame helpful

DataFrame – data table



List Semantic Models

```
import sempy.fabric as fabric
fabric.list_datasets()
```

Dataset = Semantic Model

	Dataset Name	Dataset ID	Created Timestamp	Last Update
0	Contoso10K	47c34560-ef4d-46c6-825e-20cb9f11ba9d	2023-05-04 14:36:12	NaT
1	FabSLL_Lakehouse	e6d18d3a-b407-4fb3-813a-418b76388b11	2021-02-12 23:00:58	NaT
2	DataflowsStagingLakehouse	7c76f16d-2364-4c33-89bb-6960ac29cb5d	2021-02-12 23:00:58	NaT
3	DataflowsStagingWarehouse	625cb1b9-de7e-425f-90b3-727ac87268db	2021-02-12 23:00:58	NaT
4	Contoso10K_DL	261101fb-fc2d-4511-be34-1def1b4530fe	2019-09-17 05:50:29	NaT
5	Contoso10K_DLL1	40b52877-7e62-43cd-9a38-7c09b62f9048	2019-09-17 05:50:29	NaT
6	Contoso10K_ABC	cc93c0f1-e112-4adc-a009-3fb3a7a12f3b	2019-09-17 05:50:29	NaT
7	Contoso10K_SLL	e45a6ce5-5ea5-4b87-ad88-136dc4cab27	2019-09-17 05:50:29	NaT
8	Contoso10K_SL1	421040eb-ba74-4304-bb92-708c1667eb61	2019-09-17 05:50:29	NaT
9	Contoso10K_SL2	b84b790d-04a2-406d-ba17-2191186ebafa	2019-09-17 05:50:29	NaT
10	Contoso10K_SL9	5ca86b73-bef1-43d6-af09-5c12e358d391	2019-09-17 05:50:29	NaT

List Tables

```
tables = fabric.list_tables(workspace="SQLMAB", dataset = "SQLMab")  
display(tables)
```

	Name	Description	Hidden	Data Category	Type
0	Customer		False		Table
1	Sales		False		Table
2	Date		False		Table
3	Store		False		Table
4	Product		False		Table

List Workspaces

```
1 fabric.list_workspaces()
```

✓ - Command executed in 806 ms by Jason Romans on 5:25:32 PM, 10/08/24

PySpark (Python) ▾

	Id	Is Read Only	Is On Dedicated Capacity	Capacity Id	Default Dataset Storage Format	Type	Name
0	bfab8dff-bdfc-4943-9996-7dc97a4e4d38	False	False	NaN	NaN	Workspace	JAXSQL2023
1	41e69008-f6c3-42c0-8c0a-739f1a7a9a0a	False	True	3b9ac229-bcbb-4aa4-8543-72b6db25e330	Small	Workspace	ONSQL2023
2	10c0ad3f-7a0b-4c0e-8b7d-f4a6170c5dde	False	True	3b9ac229-bcbb-4aa4-8543-72b6db25e330	Small	Workspace	SOFLSQL2023
3	ee7a8d30-f109-4848-9db3-be711a0f24d4	False	False	NaN	NaN	AdminInsights	Admin monitoring
4	10c2bd0a-5d3d-4e77-883c-f16af027fcd5	False	True	3b9ac229-bcbb-4aa4-8543-72b6db25e330	Small	Workspace	COLSQLSAT2023
5	365a3880-83f7-4014-bad3-a006c34e2bb1	False	True	3b9ac229-bcbb-4aa4-8543-72b6db25e330	Small	Workspace	BRSQSAT2023
6	e6e71dc2-2a8e-482a-8bc9-ff5c16ed0311	False	True	3b9ac229-bcbb-4aa4-8543-72b6db25e330	Small	Workspace	DENSQLSAT2023

List Workspaces on Dedicated Capacity

```
1 ws = fabric.list_workspaces()
2 dedicated = ws[ws["Is On Dedicated Capacity"] == True]
3 display(dedicated)
```

✓ - Command executed in 790 ms by Jason Romans on 5:37:07 PM, 10/08/24

PySpark (Python) ▾

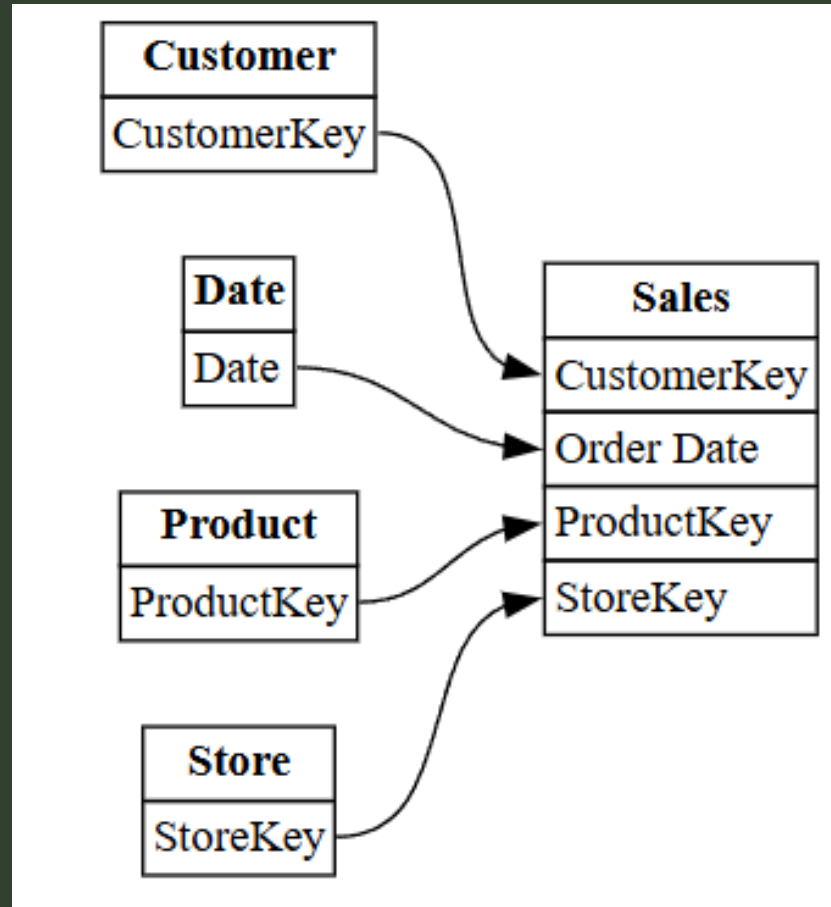
Table | Chart | Download ▾ | Showing rows 1 - 22 | Inspect | Search

	ABC Id	0/1 Is Read Only	0/1 Is On Dedicated Capacity	ABC Capacity Id	ABC Default Dataset Storage Format	ABC Type	ABC Name
1	41e69008-f...	false	true	3b9ac229-bcbb...	Small	Workspace	ONSQL2023
2	10c0ad3f-7...	false	true	3b9ac229-bcbb...	Small	Workspace	SOFLSQL2023
3	10c2bd0a-...	false	true	3b9ac229-bcbb...	Small	Workspace	COLSQLSAT2023
4	365a3880-...	false	true	3b9ac229-bcbb...	Small	Workspace	BRSQSQLSAT2023
5	e6e71dc2-...	false	true	3b9ac229-bcbb...	Small	Workspace	DENSQSQLSAT2023
6	782b40a8-...	false	true	3b9ac229-bcbb...	Small	Workspace	SQLMAB
7	62e33fa7-f...	false	true	3b9ac229-bcbb...	Small	Workspace	SQLSAT_Denver2023
8	9a5d4f1a-b...	false	true	3b9ac229-bcbb...	Small	Workspace	MNSQSQLSAT2023
9	ba1e683d-...	false	true	3b9ac229-bcbb...	Small	Workspace	ORLANDOSQSQLSAT2023
10	be75ac57-...	false	true	3b9ac229-bcbb...	Small	Workspace	SV-SQSQLSat2023
11	4d139fc8-0...	false	true	3b9ac229-bcbb...	Small	Workspace	ORWA23_SQLSAT
12	23c86de6-...	false	true	3b9ac229-bcbb...	Large	Workspace	TE Training
13	ad0afb41-...	false	true	3b9ac229-bcbb...	Small	Workspace	PremTest
14	2e6ae09d-...	false	true	3b9ac229-bcbb...	Large	Workspace	SQLBits2024
15	5f6564c6-1...	false	true	27046857-fc30-...	Small	Workspace	MetaDriven-2

Selected Cell 14 of 21 cells

Relationships

```
from sempy.relationships import plot_relationship_metadata as prm
prm(fabric.list_relationships(workspace="SQLMAB", dataset = "SQLMab"))
```





Semantic Link Demo



Our Journey



1. Intro

2. Notebooks

3. Semantic Link

4. Semantic Link Labs

5. Conclusion

Install the Package

```
# Install Semantic Link Labs
```

```
%pip install semantic-link-labs
```

```
# Import module with a shorter name
```

```
# Easier to Type
```

```
import sempy_labs as labs
```

Helper Notebooks

- MIGRATION TO DIRECT LAKE



Helper Notebooks

- BEST PRACTICE ANALYZER REPORT
- CAPACITY MIGRATION
- DELTA ANALYZER
- MODEL OPTIMIZATION
- QUERY SCALE OUT
- REPORT ANALYSIS



Helper Notebooks Part 2

- SQL
- SEMANTIC MODEL MANAGEMENT
- SEMANTIC MODEL REFRESH
- SERVICE PRINCIPAL
- TABULAR OBJECT MODEL



Obstacles

- May Require Spark
 - This function may be executed in either a PySpark or pure Python notebook. If executing in a pure Python notebook, the dataframe must be a pandas dataframe.
- May not fully work with Lakehouse Schemas

Update Dependencies

Installing Semantic Link Labs will update
Semantic Link



Expands Semantic Link

- Does not replace or overwrite
- Have both available
- Same as if imported DuckDB or other packages

In the Environment:
Install the Expansion Pack (DLC)


Create Environment (Workspace)

Synapse Data Science

BPA

Q Search

New


Current workspace:  BPA

Items will be saved to this workspace.

Data Activator

Detect patterns and conditions in your Power BI reports and streaming data, and then take actions such as alert users or kick-off workflows.

Reflex (preview)




Monitor datasets, queries, and event streams for patterns to trigger actions and alerts.

Data Engineering


Create a lakehouse and operationalize your workflow to build, transform, and share your data estate.

Lakehouse




Store big data for cleaning, querying, reporting, and sharing.

Notebook




Explore data and build machine learning solutions with Apache Spark applications.

Environment




Set up shared libraries, Spark compute settings, and resources for notebooks and Spark job definitions.

Spark Job Definition



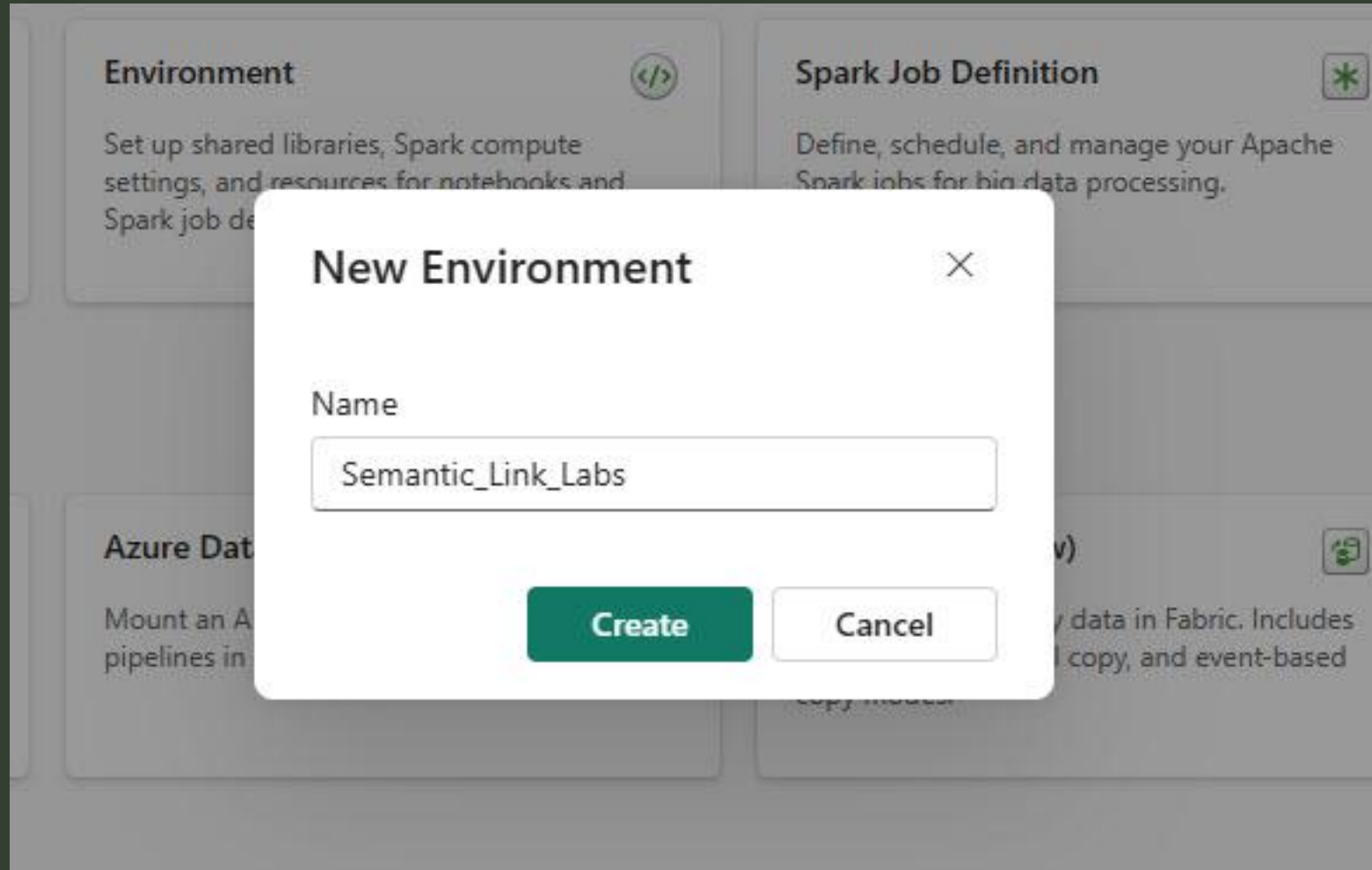
Define, schedule, and manage your Apache Spark jobs for big data processing.

API for GraphQL™ (preview)



Create an API for GraphQL to easily connect your applications to Fabric data sources.

Create Environment




The image shows a 'New Environment' dialog box overlaid on a blurred background of a software interface. The dialog box is white with a dark title bar and a close button (X) in the top right corner. It contains a text input field with the name 'Semantic_Link_Labs' and two buttons at the bottom: 'Create' (green) and 'Cancel' (white).

New Environment ✕


Name

Semantic_Link_Labs


Create Cancel

Environment 

Set up shared libraries, Spark compute settings, and resources for notebooks and Spark job de

Spark Job Definition 

Define, schedule, and manage your Apache Spark jobs for big data processing.


Azure Data 


Mount an A pipelines in


y) data in Fabric. Includes copy, and event-based

Create Environment


Libraries


 Built-in Libraries

 Public libraries


 Custom libraries

Spark compute

 Compute


 Spark properties

Storage

 Resources

Public libraries

Search and add libraries from public repositories or via a .yaml file. They'll be available if you run your notebook or Spark job definition in this environment. [Learn more](#)




There's nothing here yet


Add libraries from public repositories or via a .yaml file.


Add from PyPI

Create Environment


Libraries


 Built-in Libraries

 Public libraries 1


 Custom libraries

Spark compute

 Compute 1

 Spark properties

Storage

 Resources

Public libraries

Search and add libraries from public repositories or via a .yml file. They'll be available if you run your notebook or Spark job definition in this environment. [Learn more](#)

<input type="checkbox"/>	Library ↑	Version	Source	Status	Last updated
<input type="checkbox"/>	<div>semantic-link-labs</div>	<div>0.8.3</div>	PyPI	New	New

Specific Versions

- Allows you to develop with specific versions
- Move to newer version when ready

Publish (Important)

Libraries

Built-in Libraries

Public libraries

1

Custom libraries

Spark compute

Compute

1

Spark properties

Storage

Resources

You have unpublished changes. To apply these changes to notebooks and Spark job definition run in this environment, select Publish. To save your changes without updating the environment, select Save.

Save

Publish

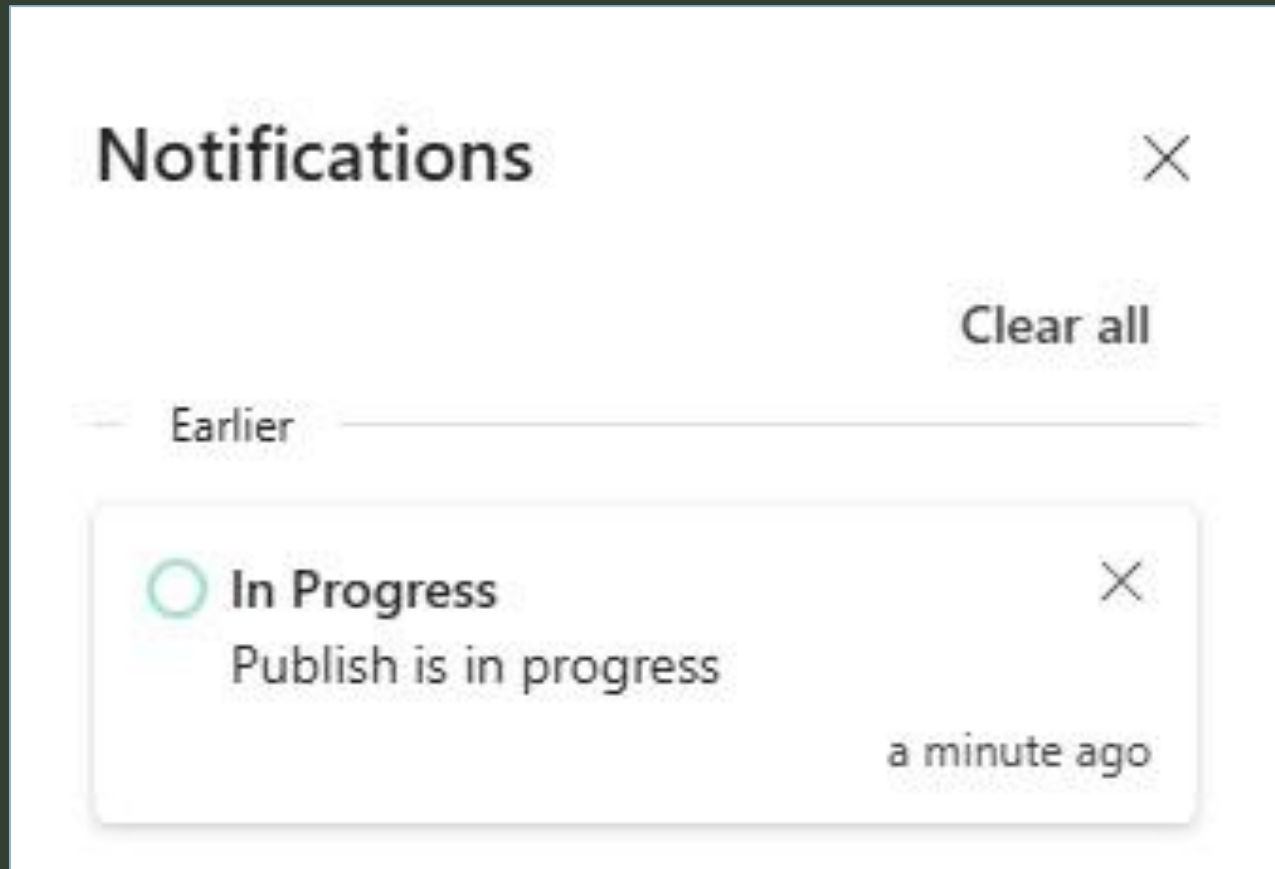
Public libraries

Search and add libraries from public repositories or via a .yaml file. They'll be available if you run your notebook or Spark job definition in this environment. [Learn more](#)

Filter by name

<input type="checkbox"/>	Library ↑	Version	Source	Status	Last updated
<input type="checkbox"/>	<input type="text" value="semantic-link-labs"/>	<input type="text" value="0.8.3"/>	PyPI	New	New

Environment takes time to publish



Workspace Settings - Environment

Workspace settings

- General
- License info
- Azure connections
- System storage
- Git integration
- OneLake
- Workspace identity
- Network security

Power BI

Delegated Settings

Data Engineering/Science

Spark settings

Data Factory

This section contains unsaved changes.

Spark settings

Configure and manage settings for Spark workloads and the default environment for the workspace.

Pool **Environment** Jobs High concurrency Automatic log

Set default environment

On

The default environment will provide Spark properties, libraries, and developer settings for notebooks and Spark job definitions in this workspace when users don't select a different environment. [Learn more about Set default environment](#)

Semantic_Link_Labs

Filter by keyword

Available environments

✓ Semantic_Link_Labs

Runtime: 1.3 (Spark 3.5, Delta 3.2), Compute: Medium, 1-10 nodes



New environment

Spark driver memory

56g

Spark executor core

8

Spark executor memory

56g

Dynamically allocate executors

Enabled

Spark executor instances

1-9

Notebook Setting - Environment

Best Practice Analyzer Report | Saved ▾

⚠ Your free Fabric trial is ending. This item will be deleted if you do not upgrade to a paid capacity. [Learn more](#) ↗

Home Edit Run View

📄 ⬇ ⚙ ▶ Run all ▾ 🔗 Connect ▾ 📄 PySpark (Python) ▾ Environment Workspace default ▾ 📊 Data Wrangler ▾

🔍 Filter by keyword

Workspace default

- ✓ Workspace Settings
Runtime: 1.3 (Spark 3.5, Delta 3.2), Compute: Medium, 1-10 nodes

Available environments

- Semantic_Link_Labs
Runtime: 1.3 (Spark 3.5, Delta 3.2), Compute: Medium, 1-10 nodes

New environment

Explorer

- + Data sources
- Resources
Uploaded data and files
- Lakehouses
1 item(s) added
- Warehouses
0 item(s) added

Install the latest version of the library

Check [here](#) to see the latest version

```
1 %pip install
```

- Session ready in 10 sec 930

Install Wheel for

Executing Notebook using a Pipeline

- Either
 1. Use Environment with Semantic Link Labs
 - Microsoft recommended way
 2. Python inline installation
 - Enable %pip install for pipeline, add "_inlineInstallationEnabled" as bool parameter equals True in the notebook activity parameters.

Reference:

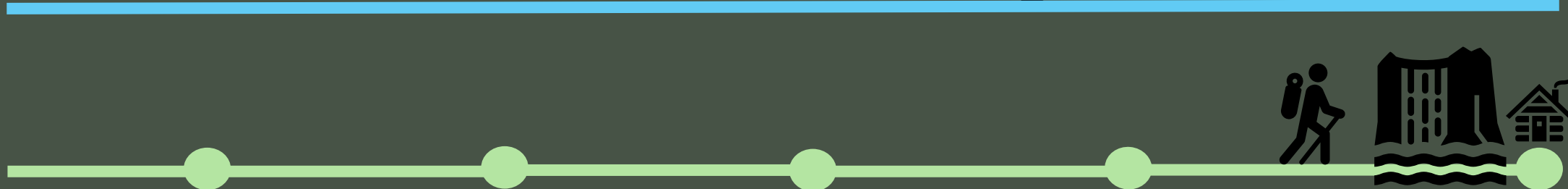
- <https://learn.microsoft.com/en-us/fabric/data-engineering/library-management>



Semantic Link Labs Demo



Our Journey



1. Intro

2. Notebooks

3. Semantic Link

4. Semantic Link Labs

5. Conclusion

Not if but when

- Recommended solution at some point
- Continues to add functionality
- Evolving to adapt to Fabric changes
 - Lakehouse Schemas

Resources

Semantic Link

- <https://learn.microsoft.com/en-us/fabric/data-science/semantic-link-overview>

Semantic Link labs

- <https://github.com/microsoft/semantic-link-labs>

Semantic Link Labs – Read the Docs

- https://semantic-link-labs.readthedocs.io/en/stable/sempy_labs.html

Thank you

JASON ROMANS

THEDAXSHEPHERD@GMAIL.COM

WWW.THEDAXSHEPHERD.COM



The DAX Shepherd