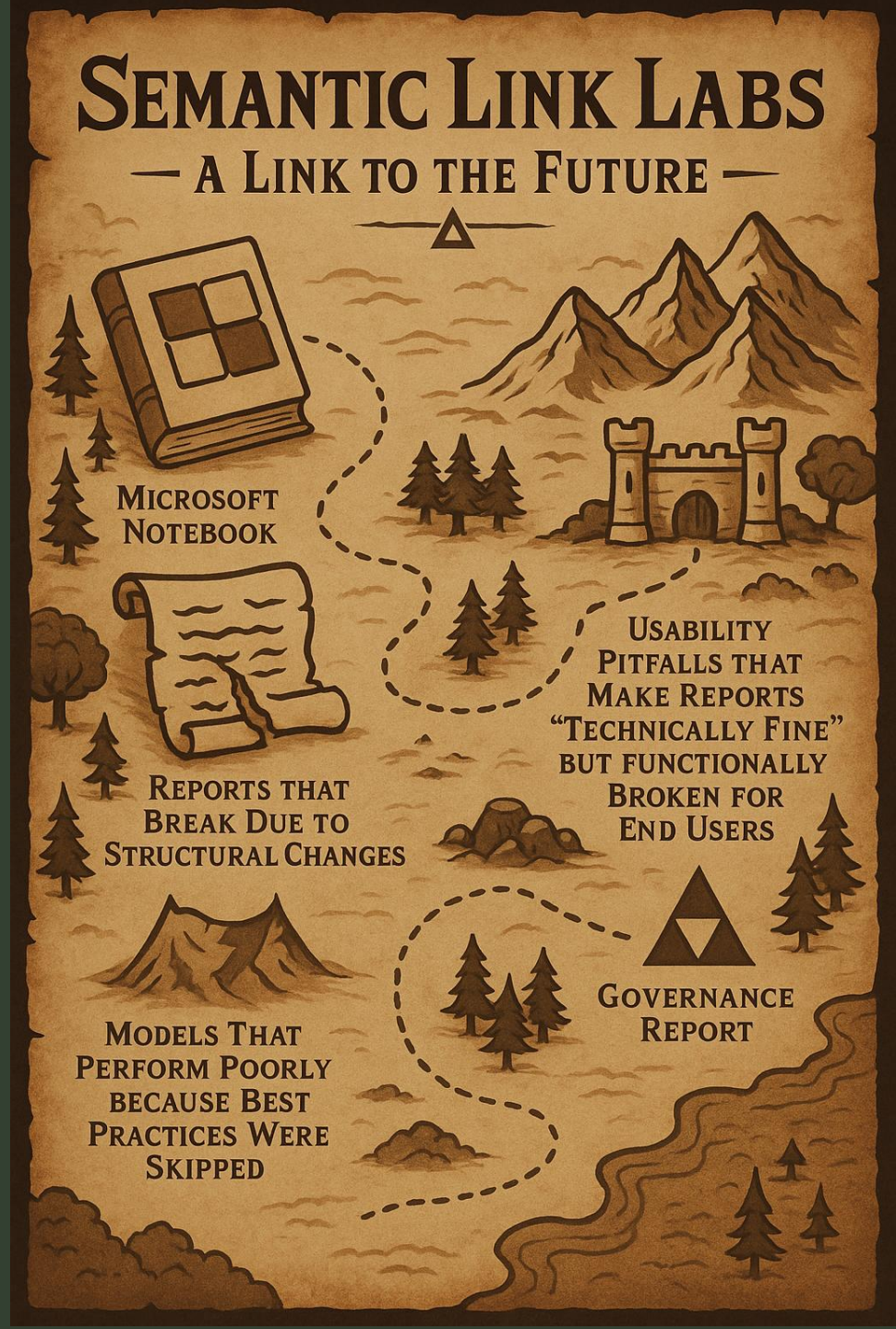
Semantic Link Labs: A Link to the Future

JASON ROMANS

ATLANTA BI USER GROUP 2025



Jason Romans

Cloud Data & Integration Developer

The DAX Shepherd

Microsoft® MVP Most Valuable Professional

@sql_jar

jason-r-sql-jar

https://thedaxshepherd.com/

📍 Nashville, TN, USA

🛠️ Began Career as a SQL Server DBA

➡️ Transitioned to Microsoft BI Stack

📦 Data Engineering to Data Modeling

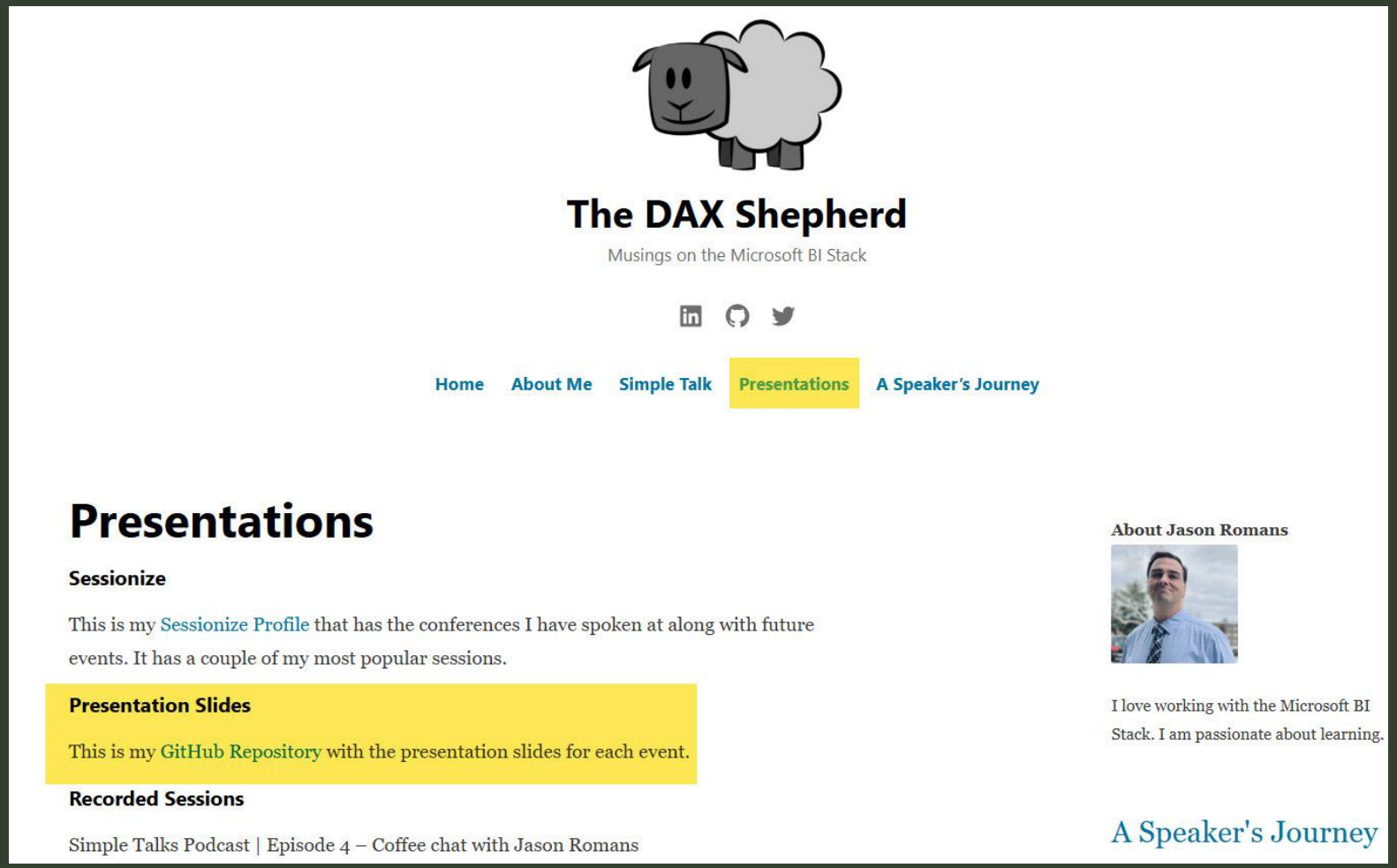
🖍️ Infrequent Blogger

🧭 Fan of Dimensional Models & Doctor Who

Shoulders of Giants

Slides

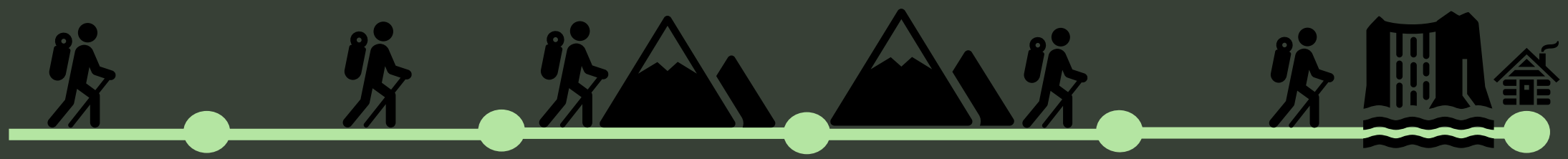
www.thedaxshepherd.com



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

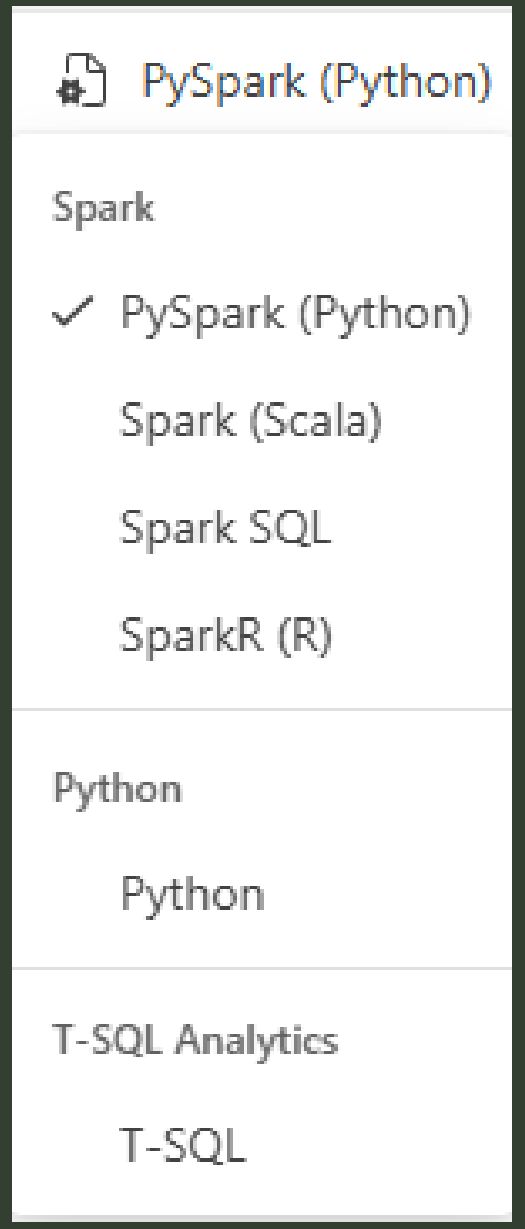
LANGUAGE

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

COMPUTE

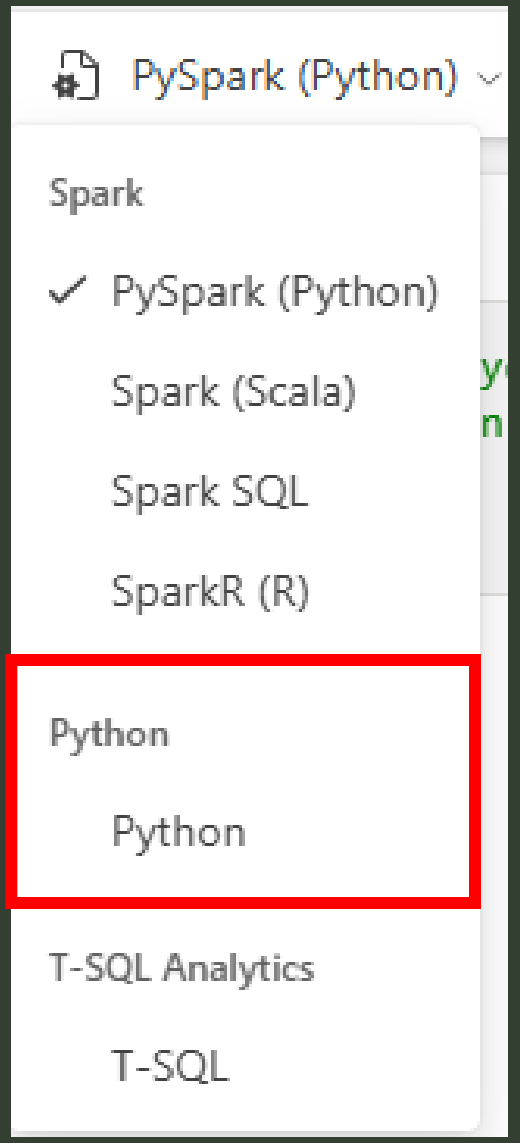
- Spark
- Python (Single Node Virtual Machine)
- T-SQL Analytics

Compute & Language

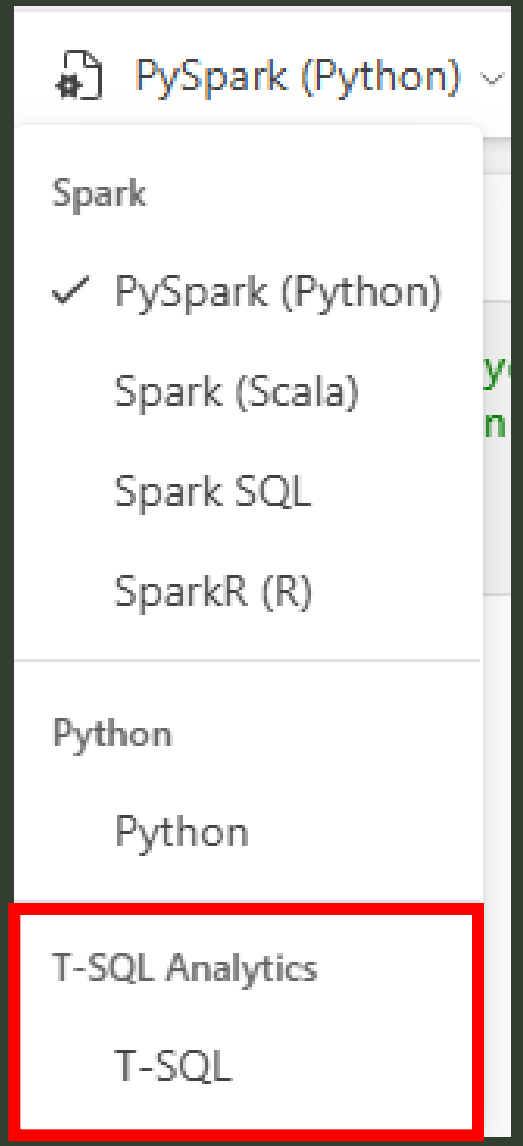


Spark (Python, Scala, SQL, R)

Python (Python)



T-SQL Analytics (T-SQL)



Choosing PySpark (Cluster) or Python Compute (Single Node)

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

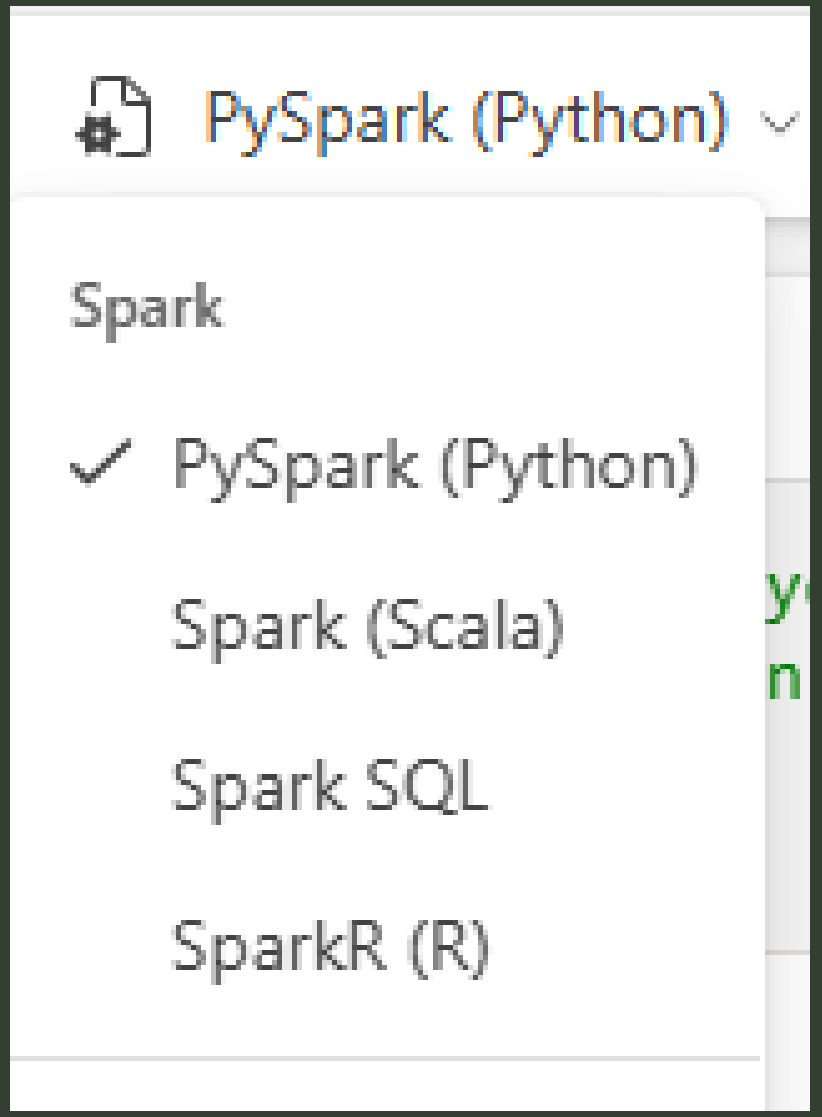
- Has libraries installed for dealing with “small-big” data
  - Less than 10 Gigabytes
  - Fits in memory
- Example Libraries installed
  - Polars
  - DuckDB

Languages for Spark

Different choices of languages

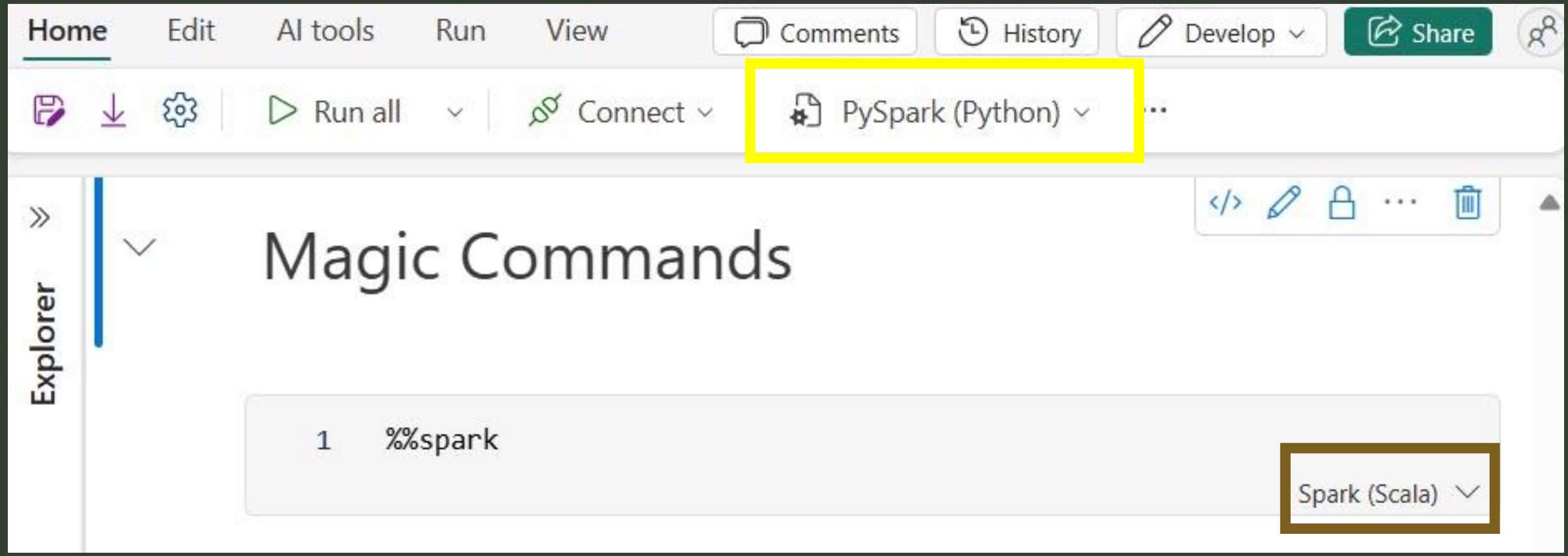
Spark was built with Scala

- 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 n **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

Migrate to Direct Lake

Best Practice Analyzer

Manage Reports

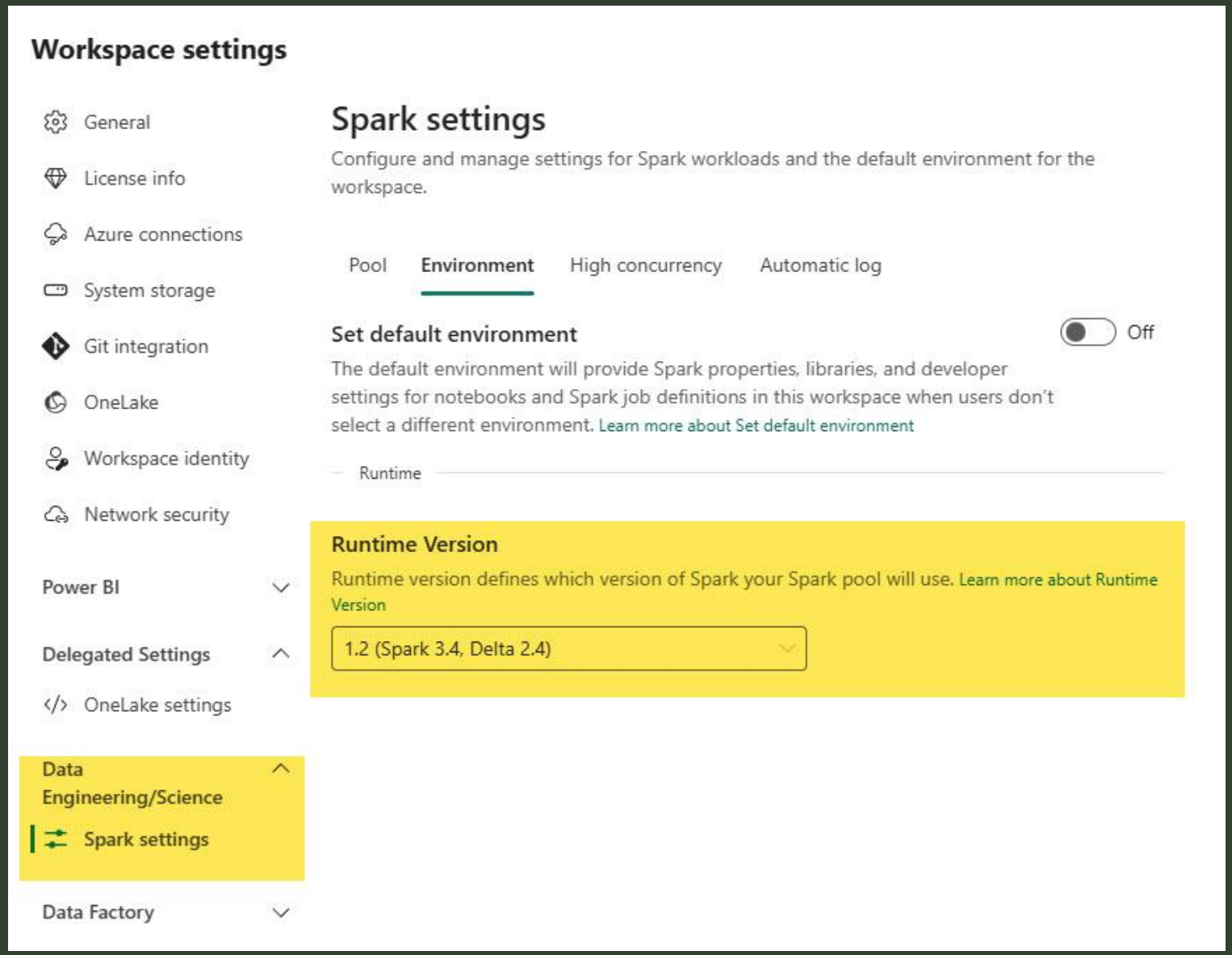
View Broken Reports

Rebind Reports

Installing

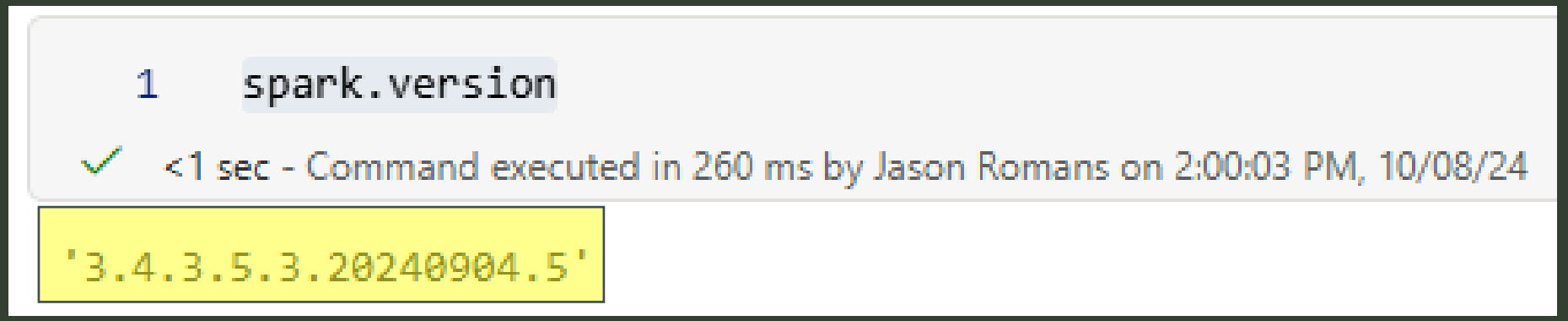
- Semantic Link
  - Spark 3.4 in default runtime
    - Update to newest version
      - %pip install –U semantic-link
  - Spark 3.3 or below – need to install
    - %pip install –U semantic-link

What Version of Spark



Spark Version Command

spark.version



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.
```

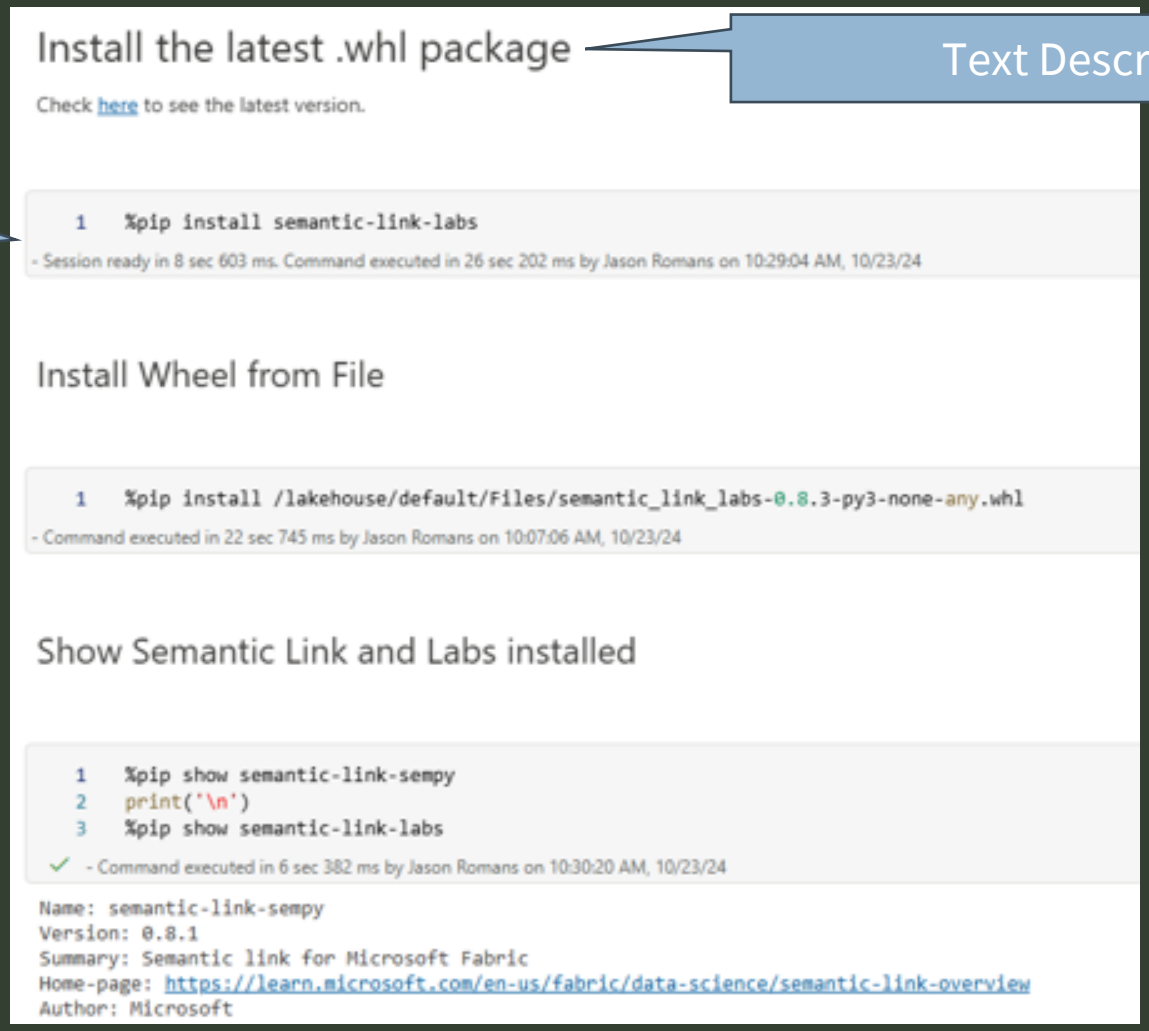
Import the Module

```
# Give it a friendly name – think Alias
import sempy.fabric as fabric
```

Done in a Notebook

Code - Python

Text Descriptions - Markdown



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-ad88-136dc4cabc27	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	BRSQLSAT2023
6	e6e71dc2-2a8e-482a-8bc9-ff5c16ed0311	False	True	3b9ac229-bcbb-4aa4-8543-72b6db25e330	Small	Workspace	DENSQLSAT2023

List Workspaces on Dedicated Capacity

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

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

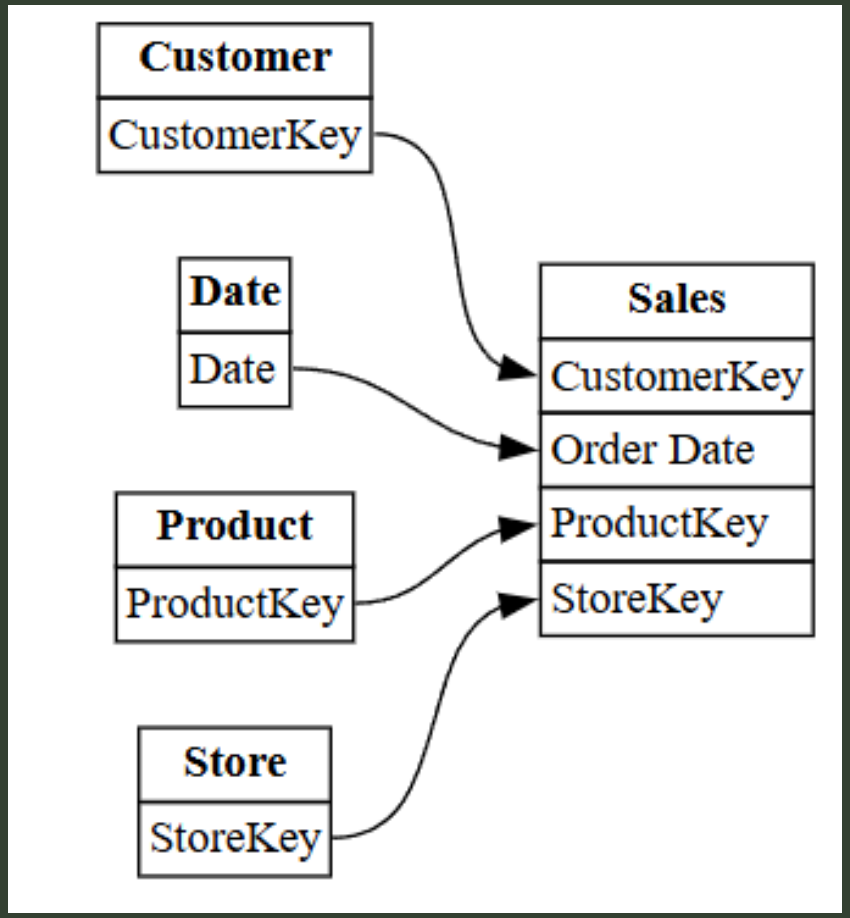
Table | Chart | Download | Showing rows 1 - 22 | Inspect | Search

	Id	Is Read Only	Is On Dedicated Capacity	Capacity Id	Default Dataset Storage Format	Type	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	BRSQLSAT2023
5	e6e71dc2-...	false	true	3b9ac229-bcbb...	Small	Workspace	DENSQLSAT2023
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	MNSQLSAT2023
9	ba1e683d-...	false	true	3b9ac229-bcbb...	Small	Workspace	ORLANDOSQLSAT2023
10	be75ac57-...	false	true	3b9ac229-bcbb...	Small	Workspace	SV-SQLSat2023
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 ce

Relationships

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

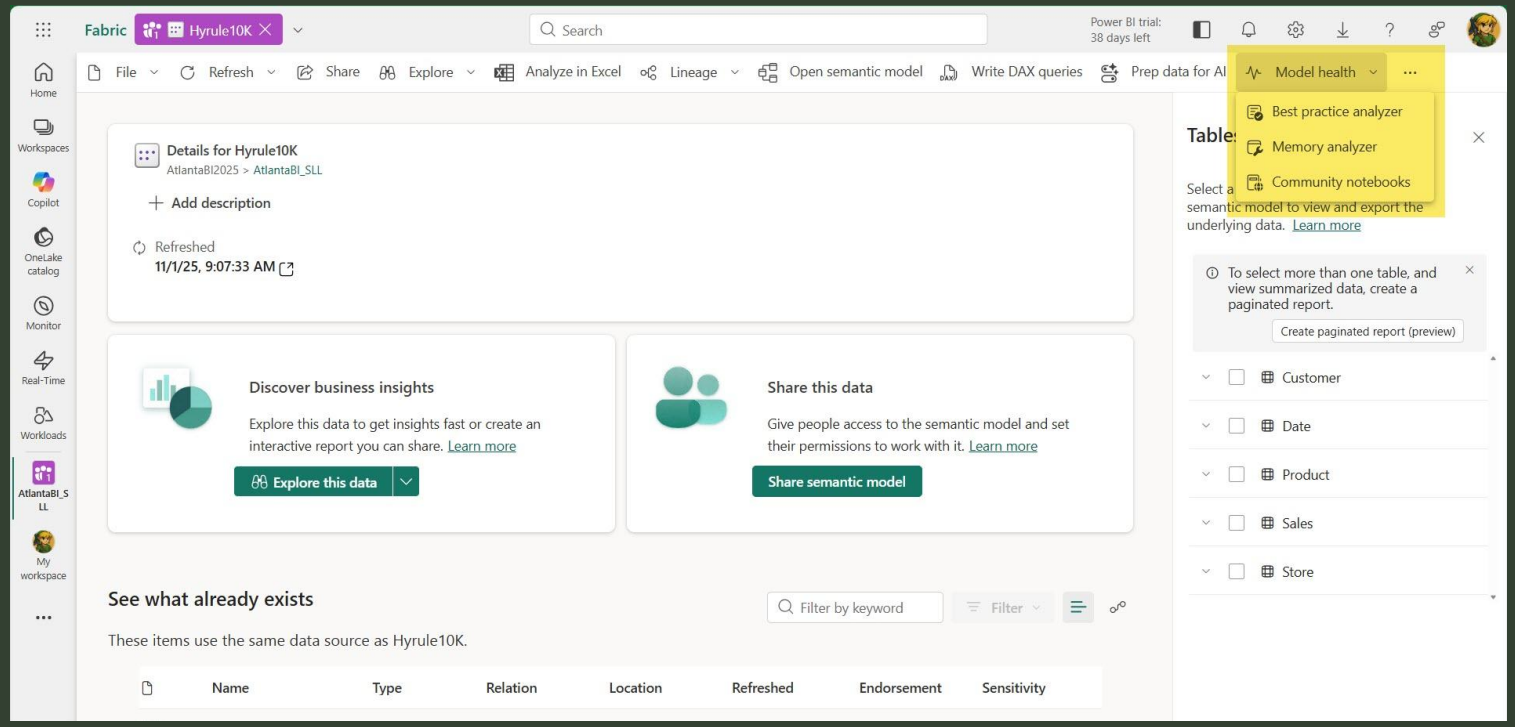


Model Health

Best Practice Analyzer

Memory Analyzer

Community Notebooks



Semantic Link Demo

Our Journey

1. Intro
2. Notebooks
3. Semantic Link
4. **Semantic Link Labs**
5. Conclusion

Version

0.4.2 in June 18, 2004

Now up to 0.12.5

Created

Migrate from Import to Direct Lake

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

Helper Notebooks Part 2

- REPORT ANALYSIS
- 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

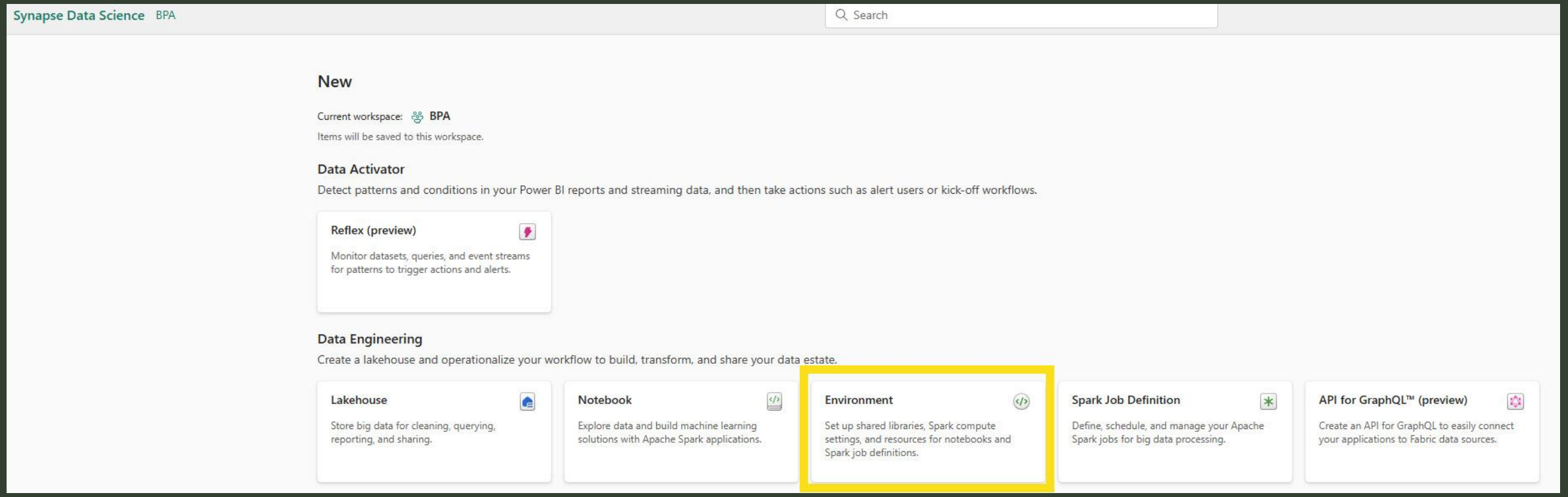
Installing Semantic Link Labs will update Semantic Link

Expands Semantic Link

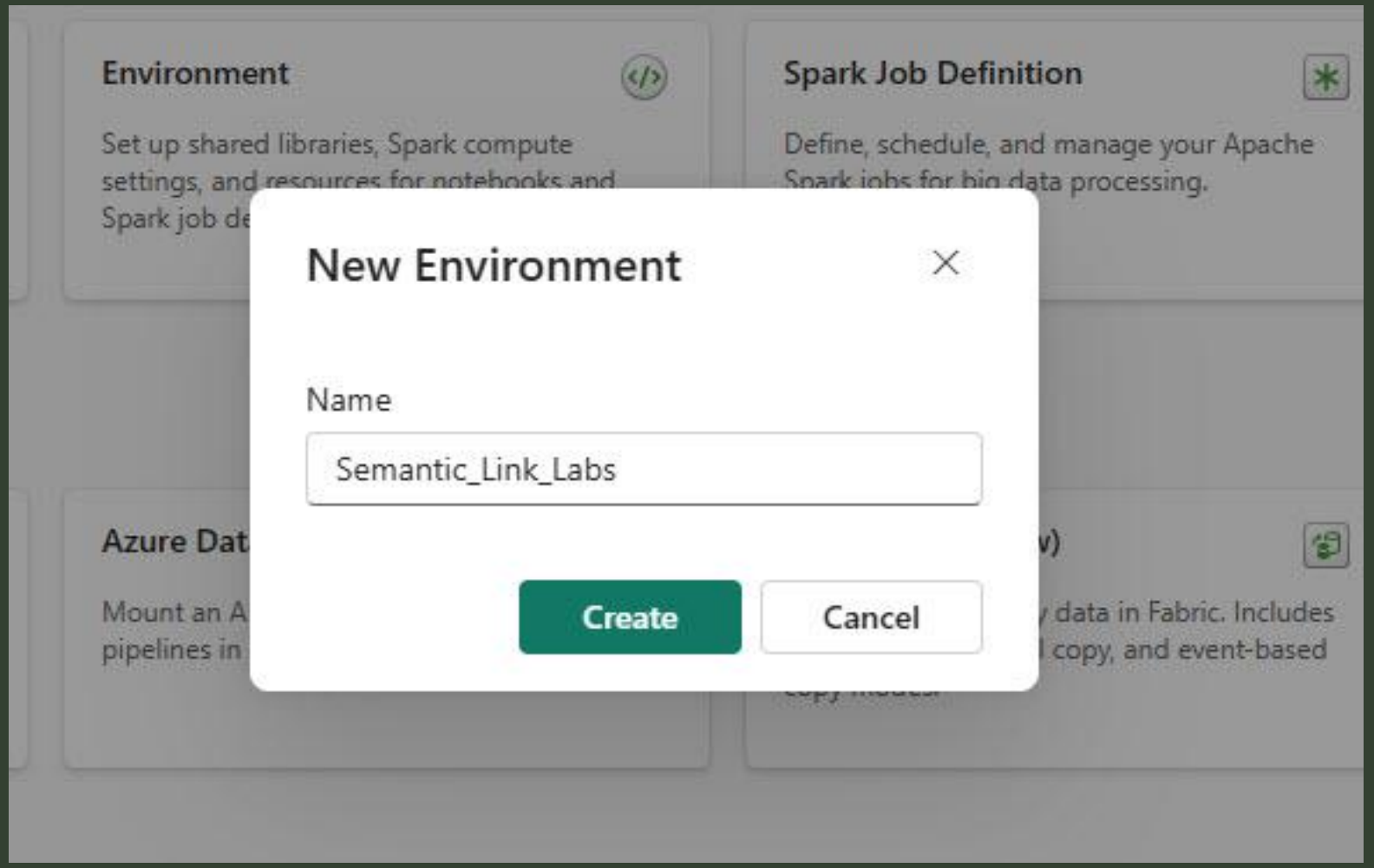
- Does not replace or overwrite
- Have both available
- Same as if imported DuckDB or other packages

In the Environment:
Install the Expansion Pack (DLC)

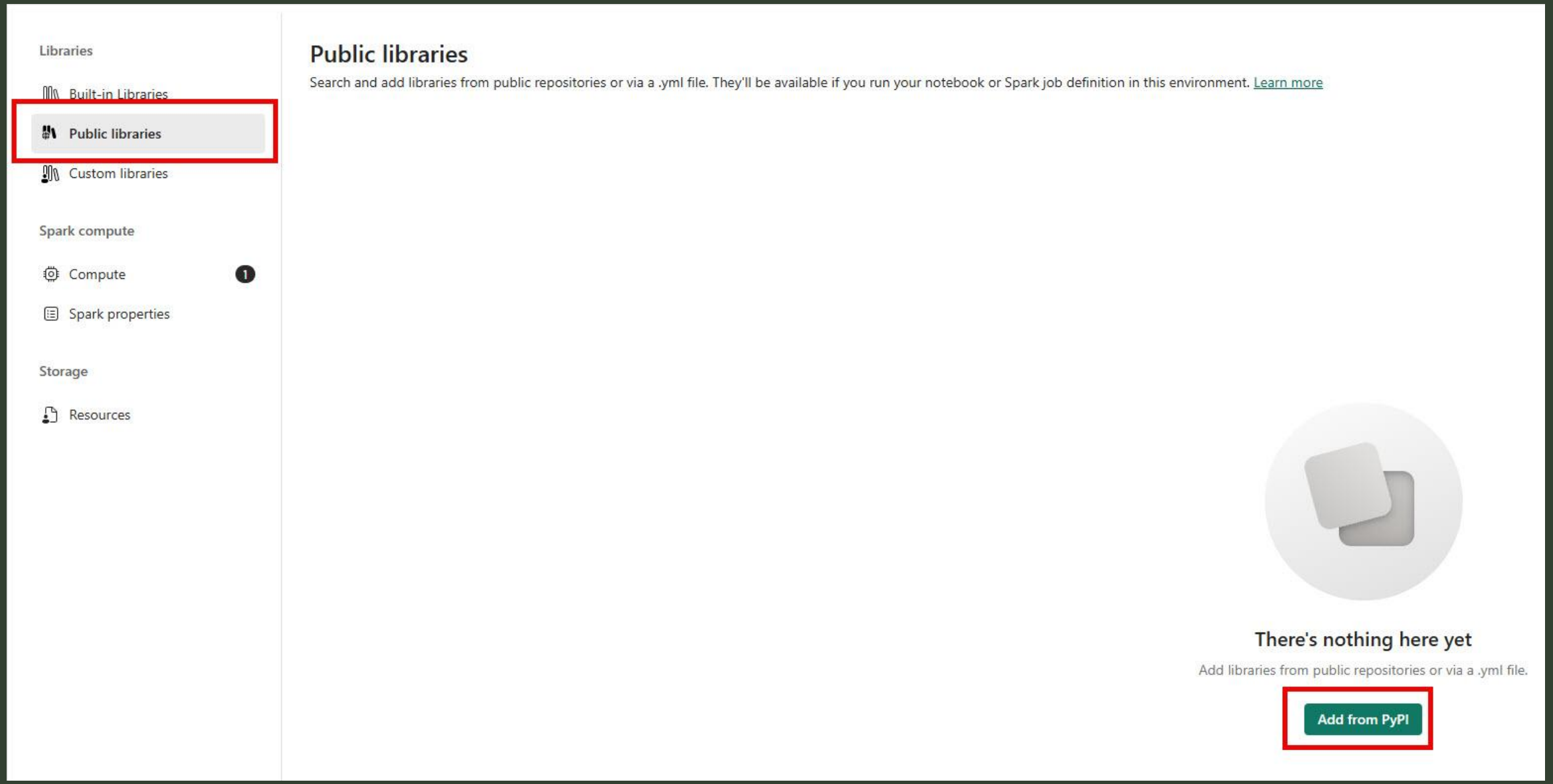
Create Environment (Workspace)



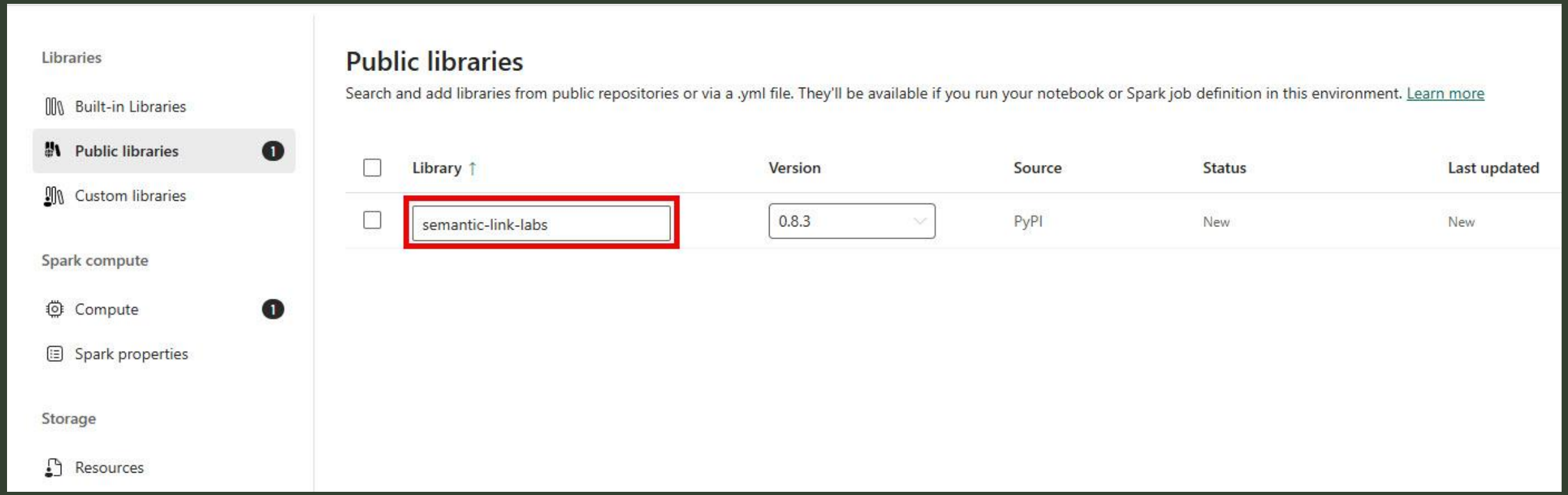
Create Environment



Create Environment



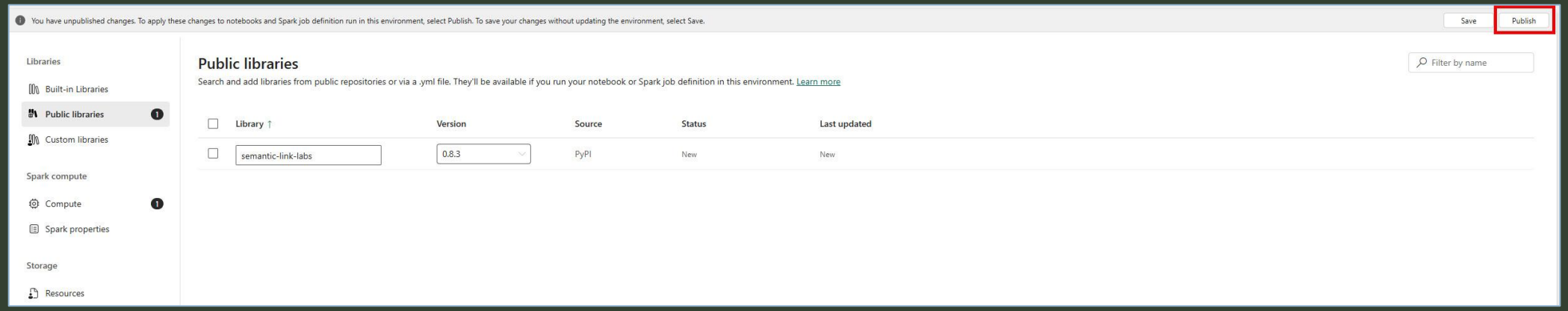
Create Environment



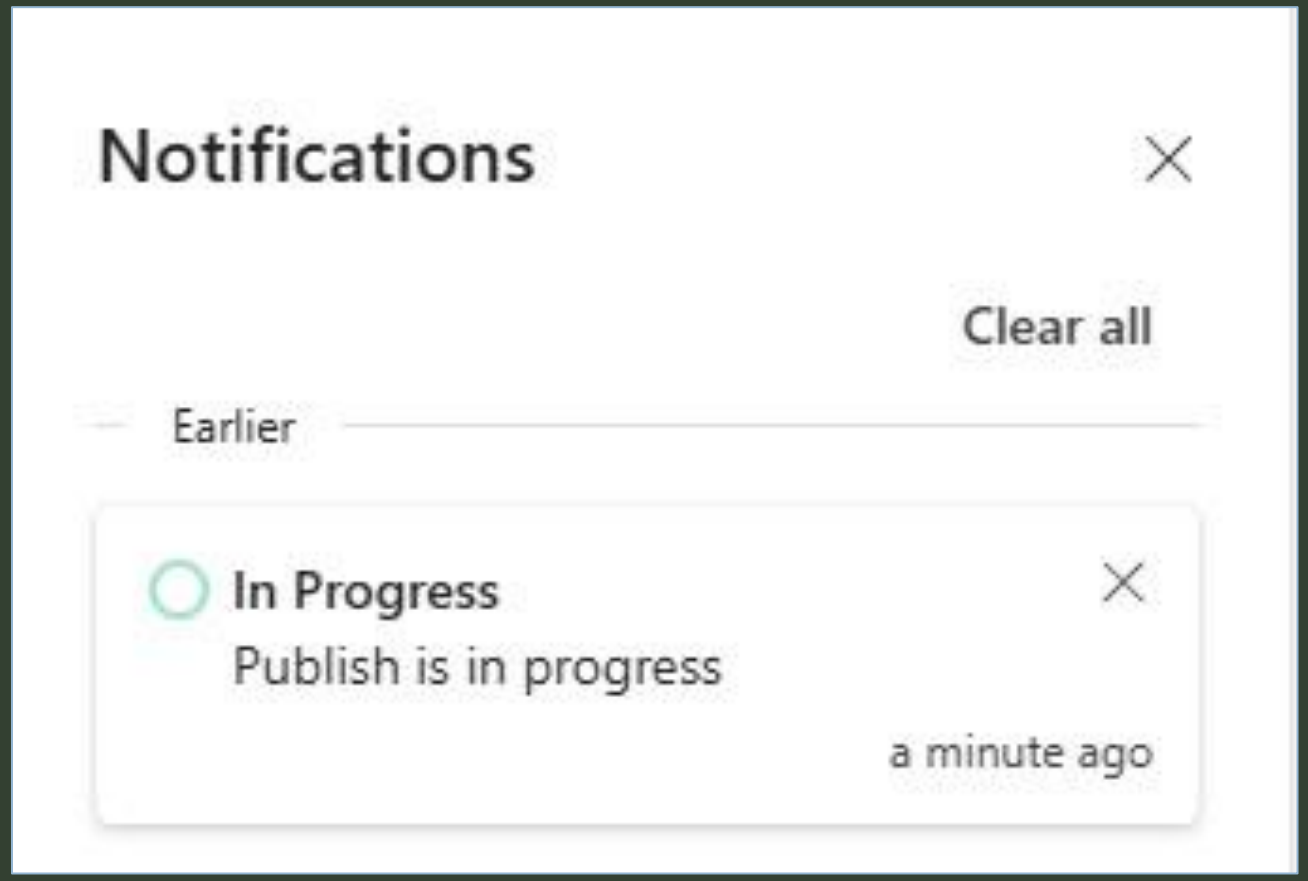
Specific Versions

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

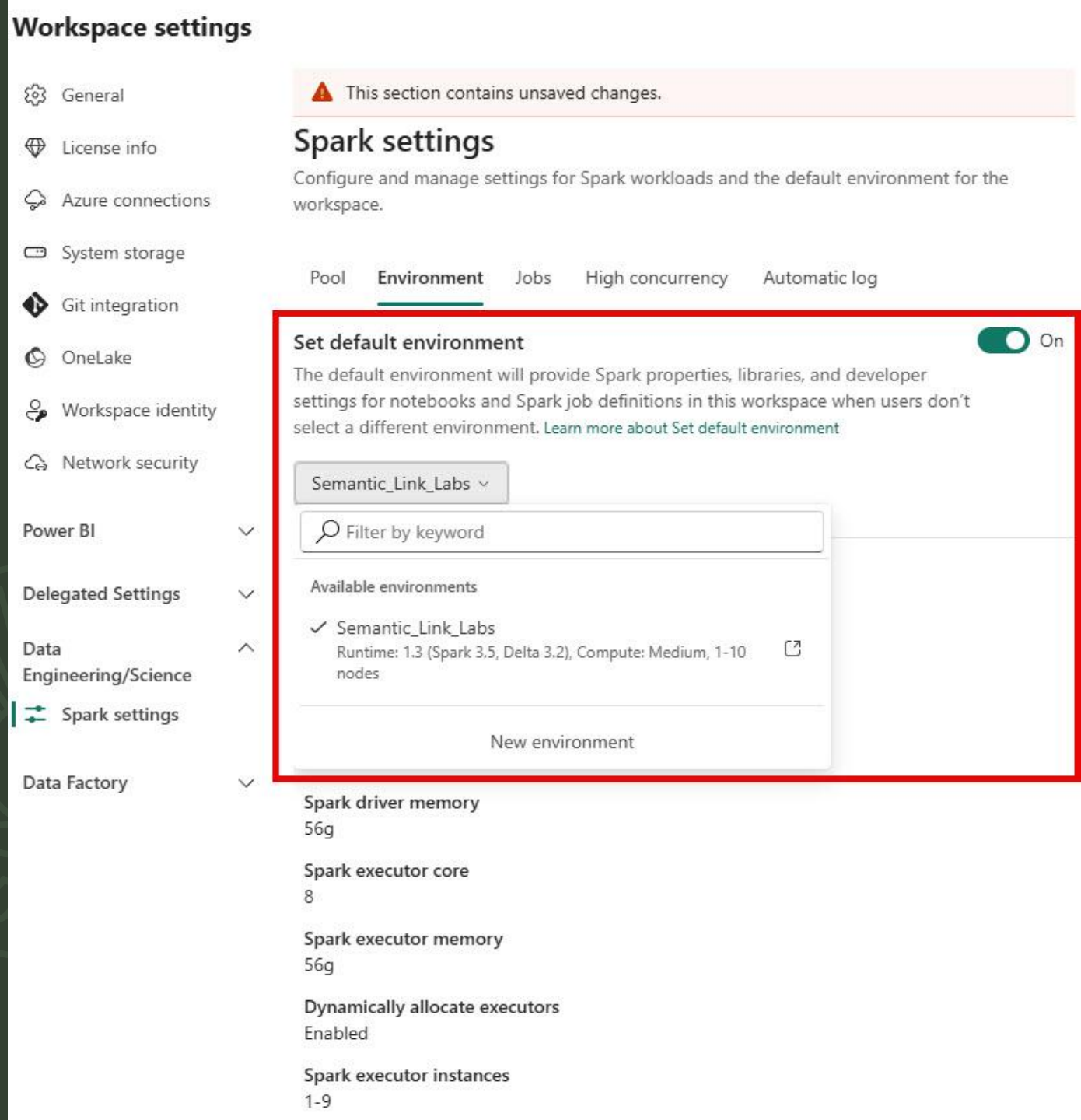
Publish (Important)



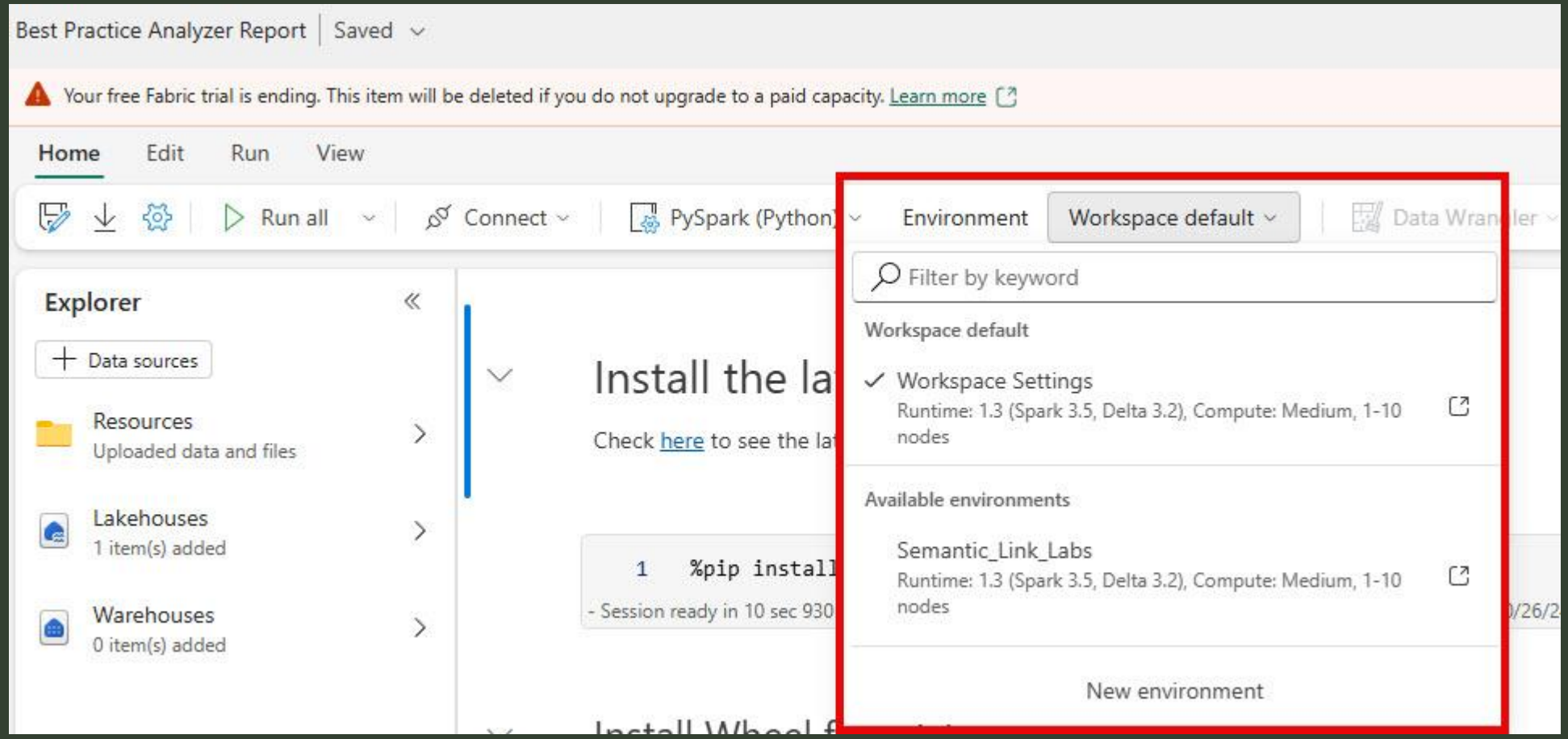
Environment takes time to publish



Workspace Settings - Environment



Notebook Setting – Environment



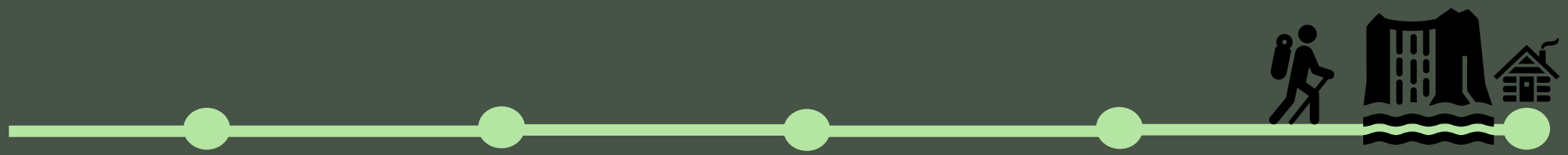
Executing Notebook using a Pipeline

- Either
  1. Use Environment with Semantic Link Labs
     - Microsoft recommended way
     - Only for Spark
  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

Resources

Tabular Editor Training (Notebooks, Semantic Link, and semantic-link-labs)

- https://tabulareditor.com/learn

- Sandeep Pawar
  - https://fabric.guru/fabric-semantic-link-and-use-cases
- Zelda Dataset (Zelda: Breath of the Wild Hyrule Compendium)
  - https://www.kaggle.com/datasets/elisamork/zelda-breath-of-the-wild-hyrule-compendium

Thank you

JASON ROMANS

THEDAXSHEPHERD@GMAIL.COM

WWW.THEDAXSHEPHERD.COM

The DAX Shepherd