Ventagium
Data Consulting

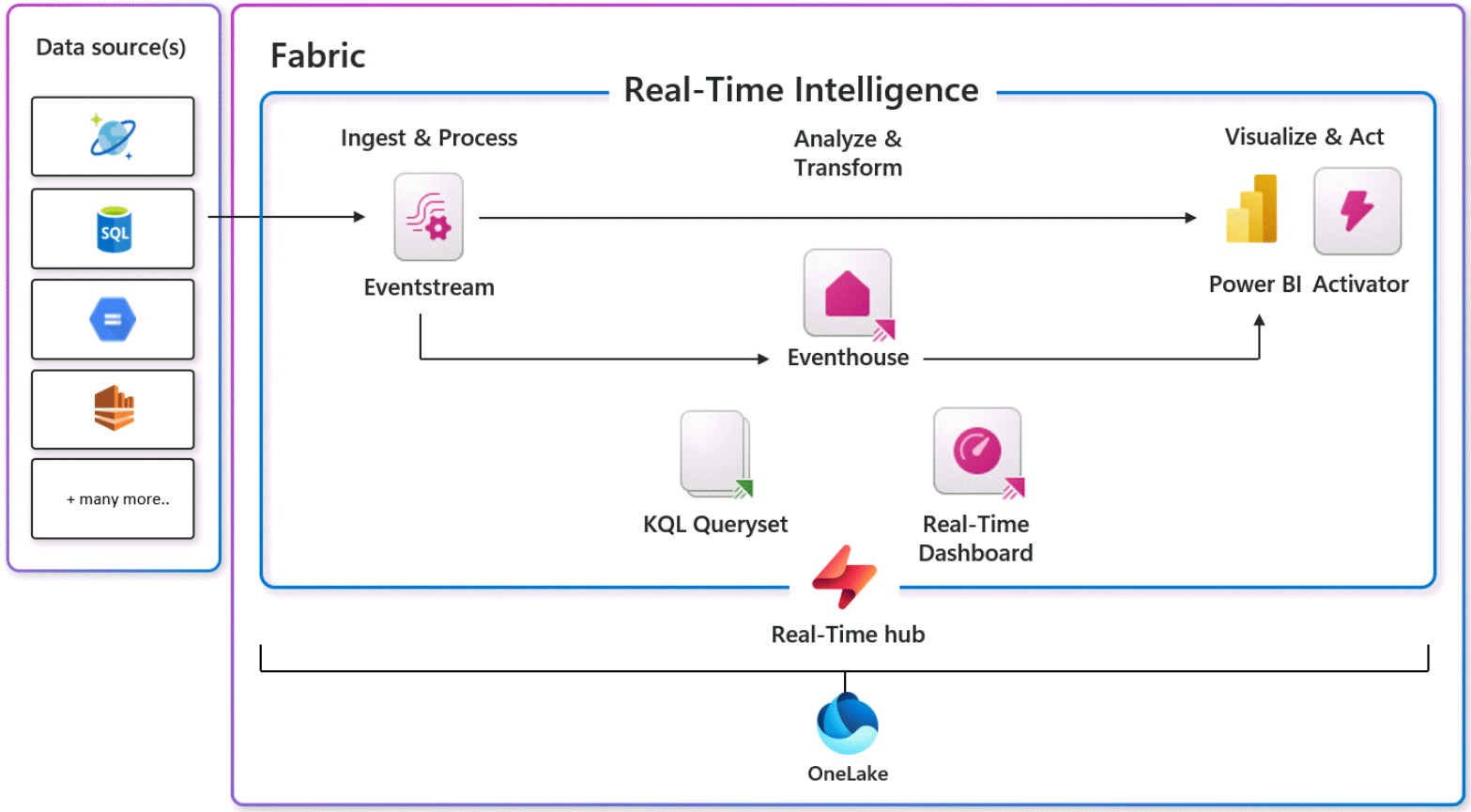
Fabric Workshop

2025-02-28 Part 1

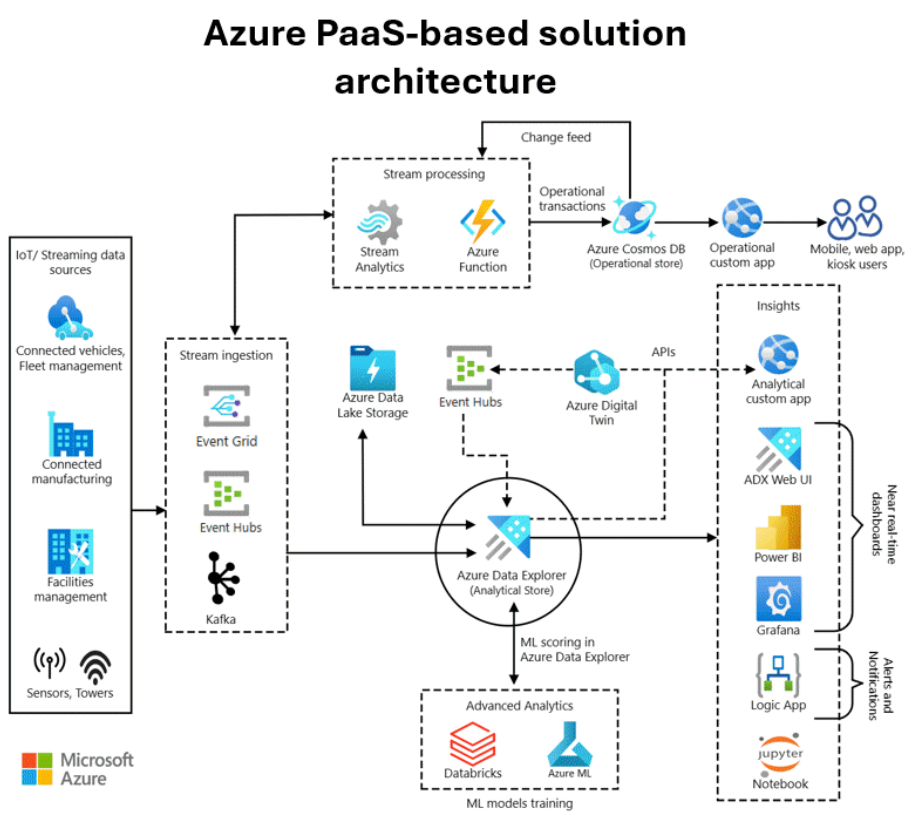
Real-Time Intelligence

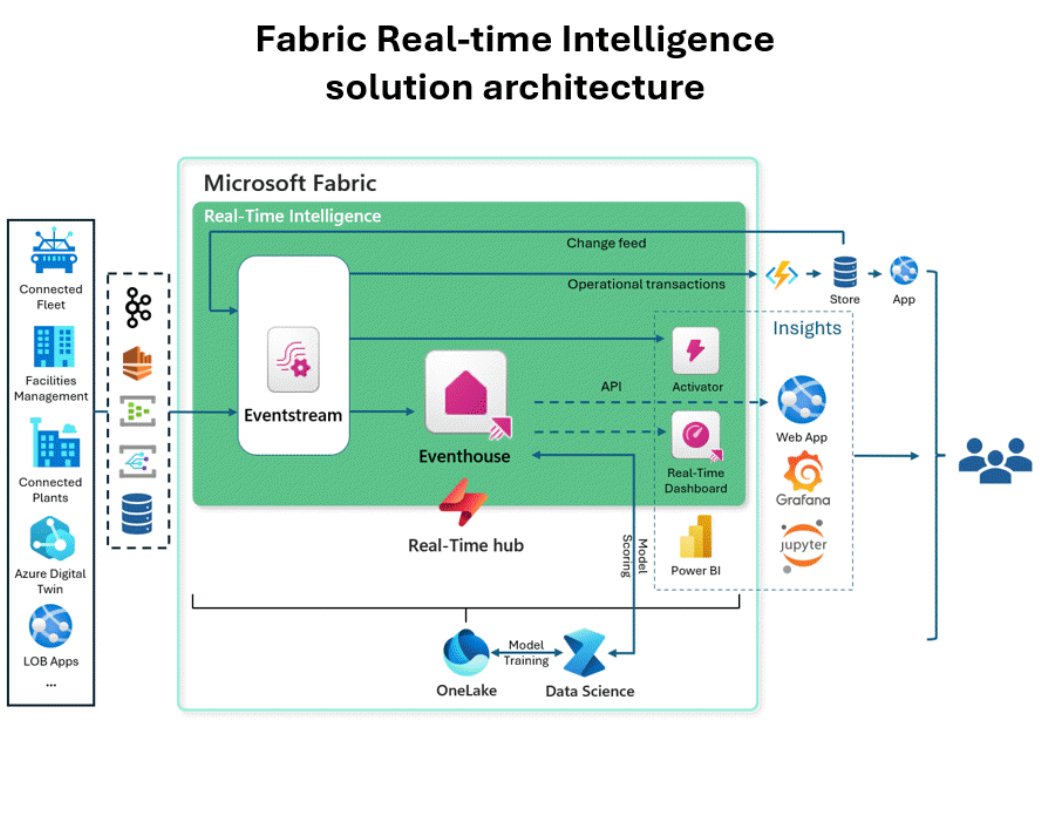
Real-Time Intelligence

- Comprehensive SaaS Offering
- Centralized Hub
- Rapid solution deployment
- Insights powered by real-time AI



Real-Time Intelligence





Real-time Intelligence: Components

- Real-time Hub
- Eventstreams
- Eventhouse
- Activator

Real-time Intelligence: Use Cases

- Manufacturing PLCs
- IoT Systems
- Fraud Detection
- Anomaly Detection

Real-time Intelligence: Use Cases

- Manufacturing PLCs
- IoT Systems
- Fraud Detection
- Anomaly Detection

Best Practices

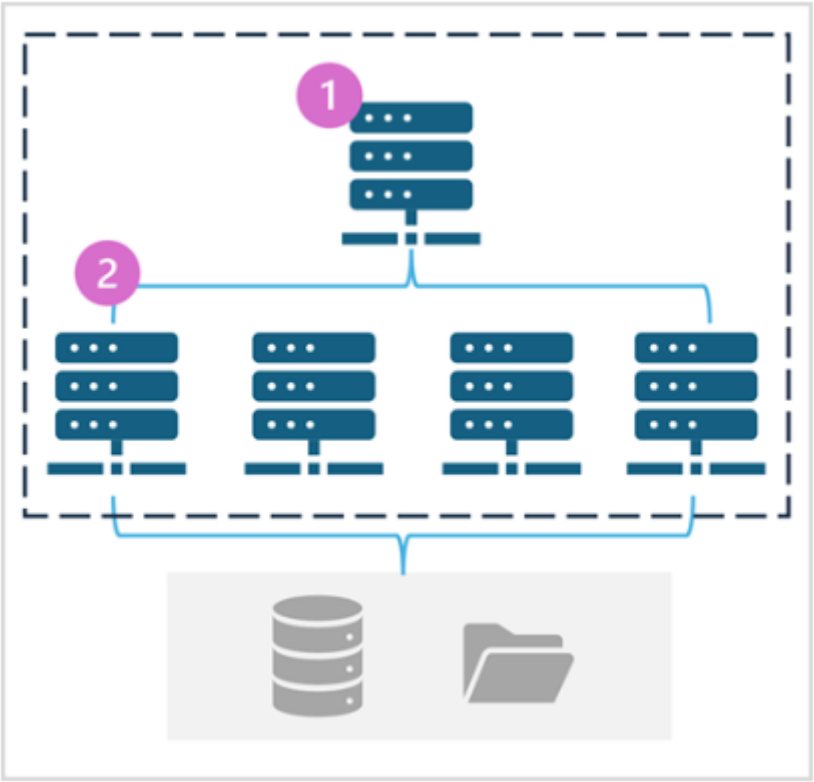
- **Be use-case driven:** not everything needs to be real-time.
- **Automate decisions:** don’t expect someone to be reacting all of the time.
- Don’t use Fabric as an Operational Platform

End-to-end Demo

Spark and Fabric Notebooks

What is Spark?

- Unified analytics engine for large-scale data processing.
- Used for batch and stream processing.
- Able to be run on single machine or multiple nodes.
  - Coordinates work across multiple processing nodes in a cluster, known in Microsoft Fabric as a **Spark pool**.

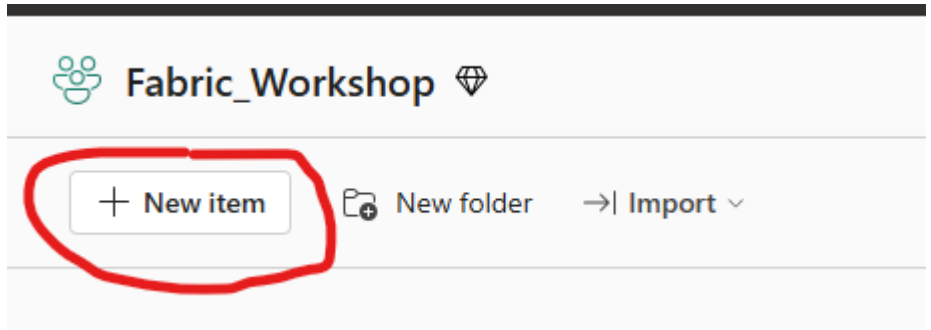


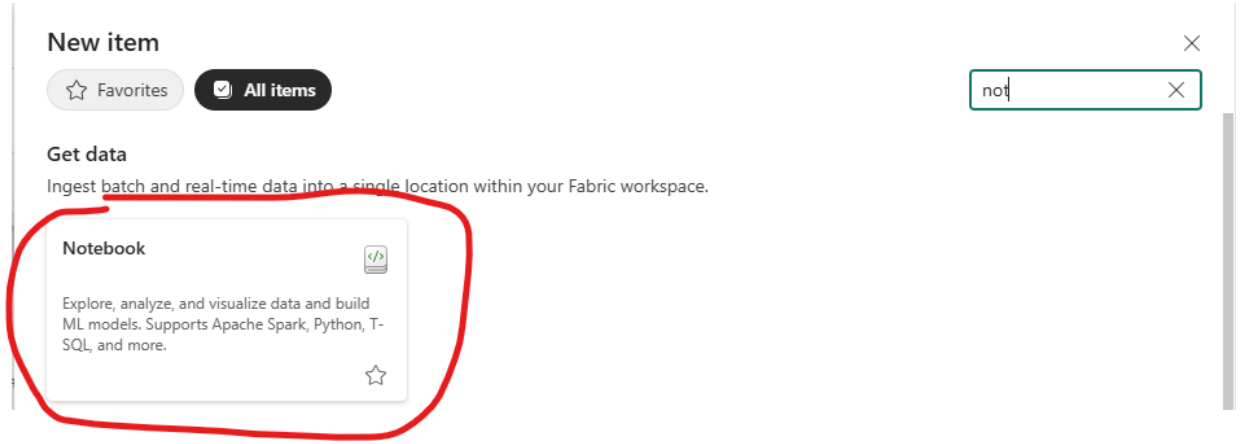
1. **Head node** coordinates via **driver**.
2. **Worker nodes** complete tasks via **executors**.

What are Fabric Notebooks?

- Primary programming experience in Microsoft Fabric.
- Multiple programming languages supported.
  - Spark – Scala, PySpark, SparkR, Spark SQL
  - Python (Preview)
  - T-SQL
- Applications
  - ETL
  - EDA
  - Machine Learning

Getting started in 2 steps…



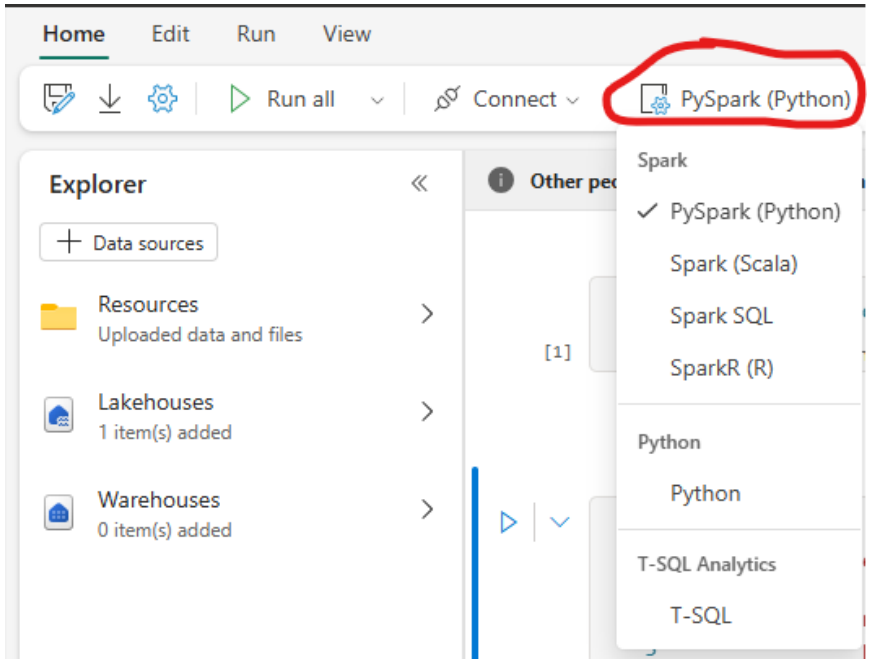


Exercise: Notebooks Demo and Tutorial

Fabric Notebooks Advanced Features

Different Programming Languages

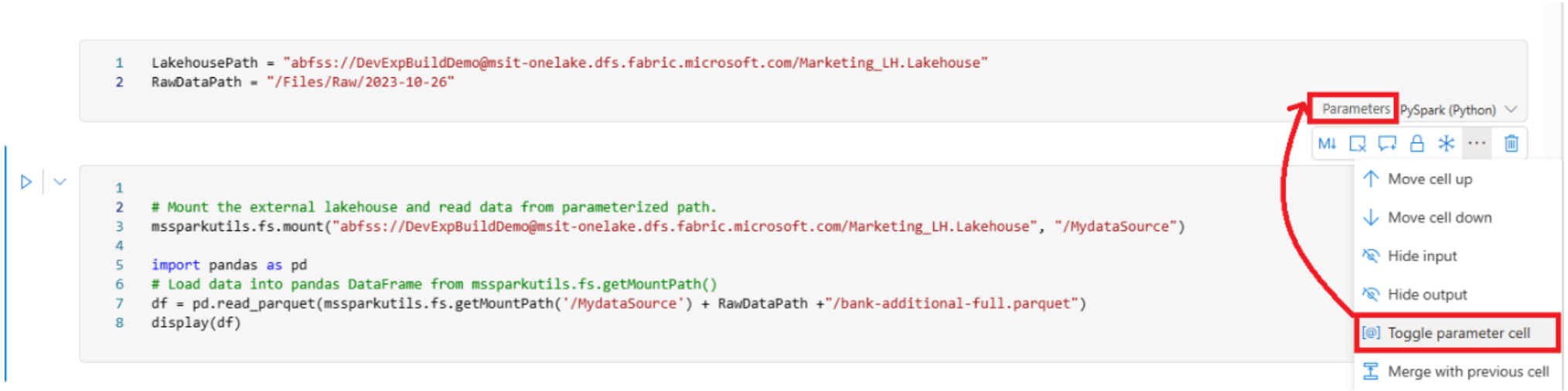
- Set for entire notebook for single cell.



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.

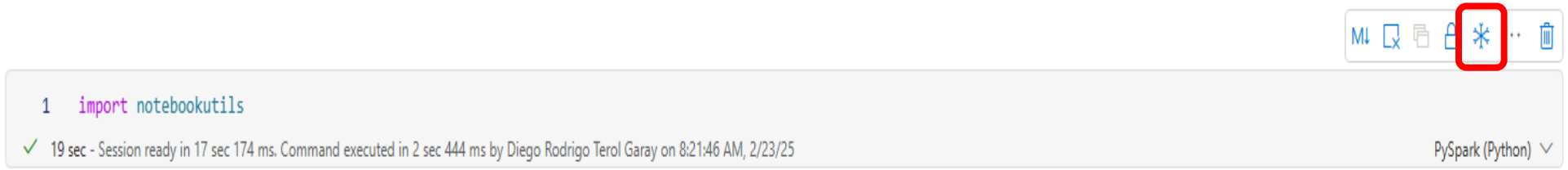
Parameter Cells

- Pass arguments via runMultiple or Pipeline.



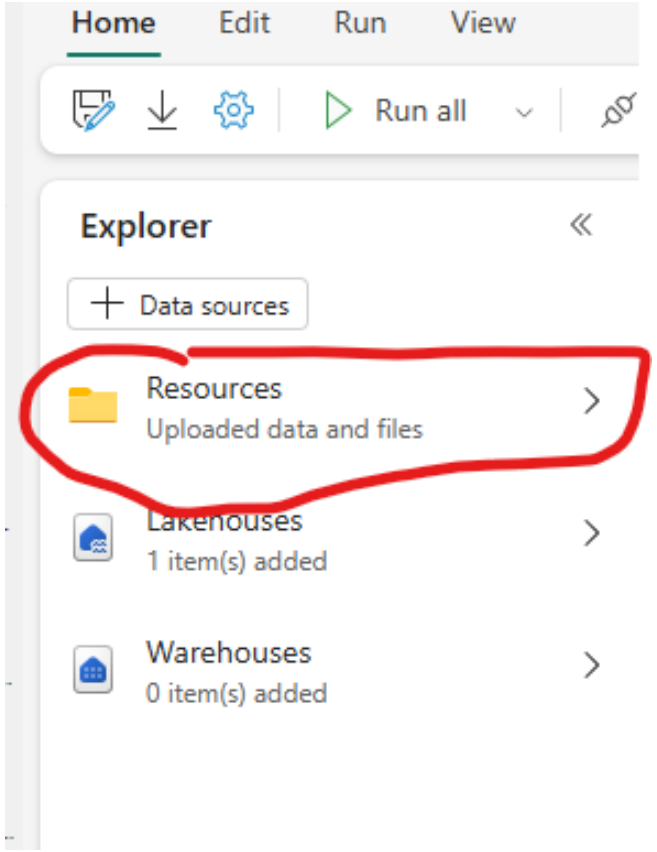
Frozen Cells

- Make cells read-only.



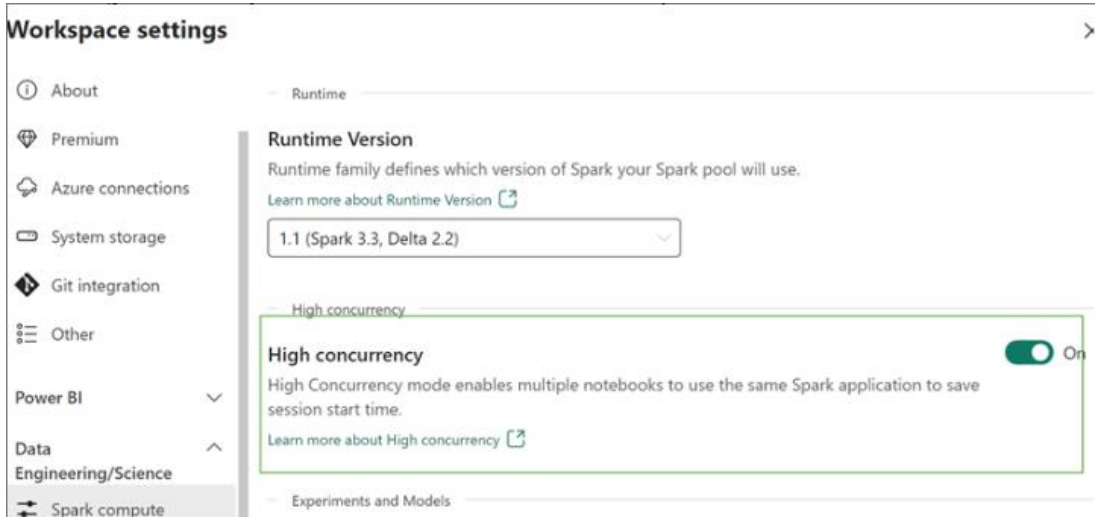
Notebook Resources

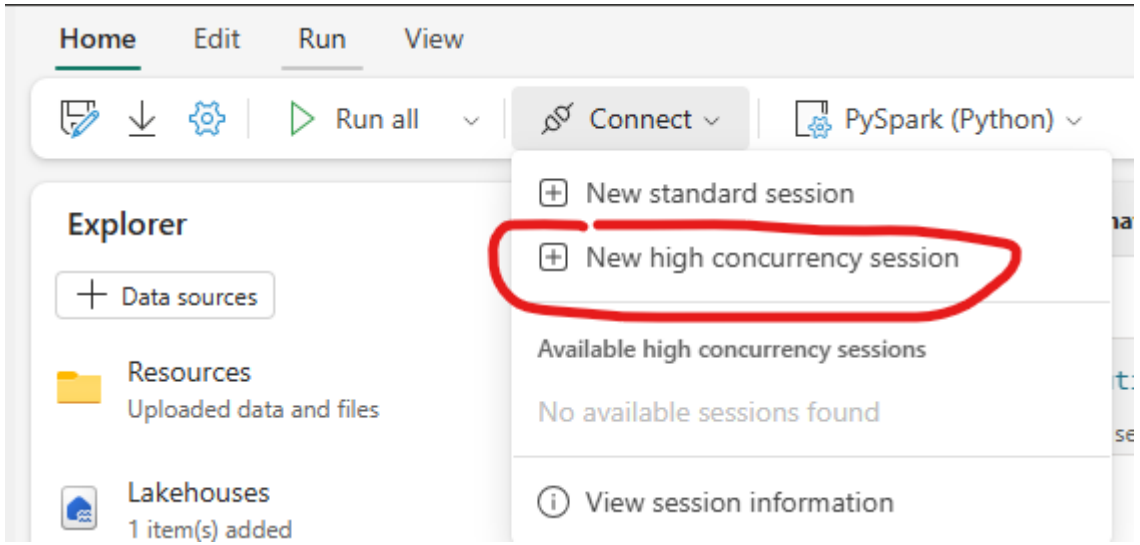
- File system space to store small-sized files (e.g., code modules, semantic models, images…).



High Concurrency

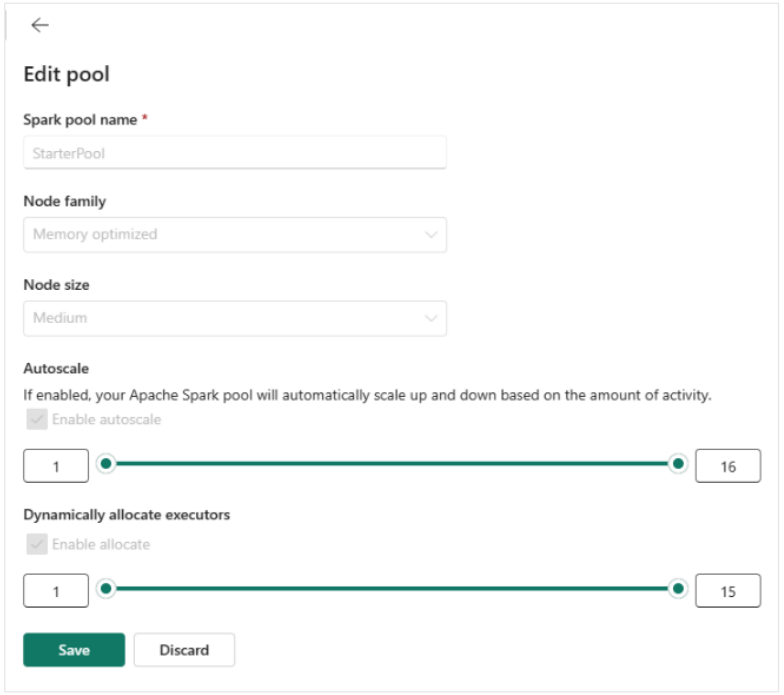
- Share Spark Sessions across multiple Notebooks.





Configure Spark Starter Pools

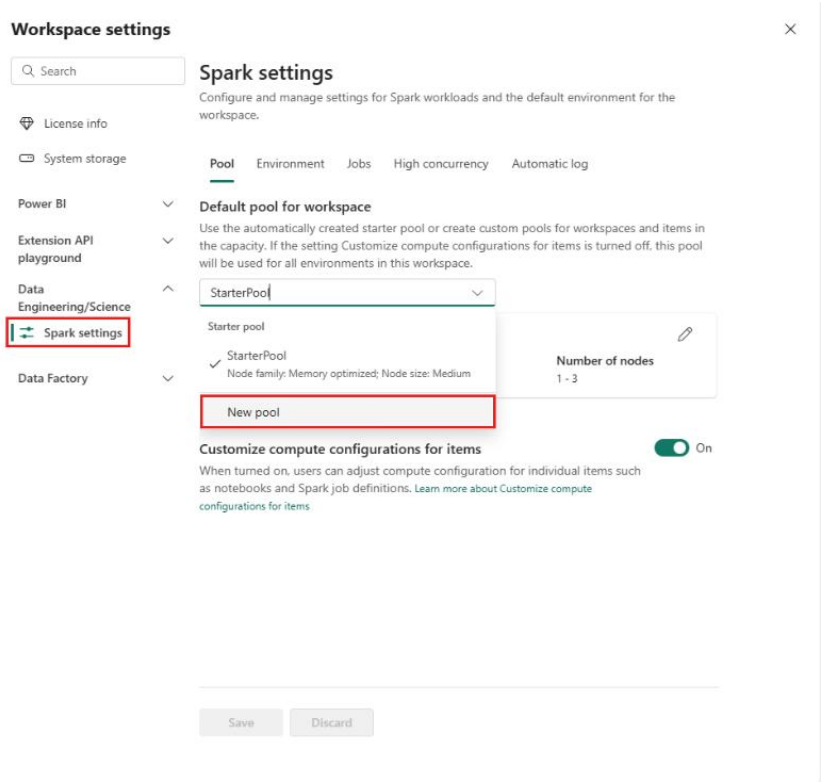
- Configure Starter Pool based on Capacity size ans required resources.

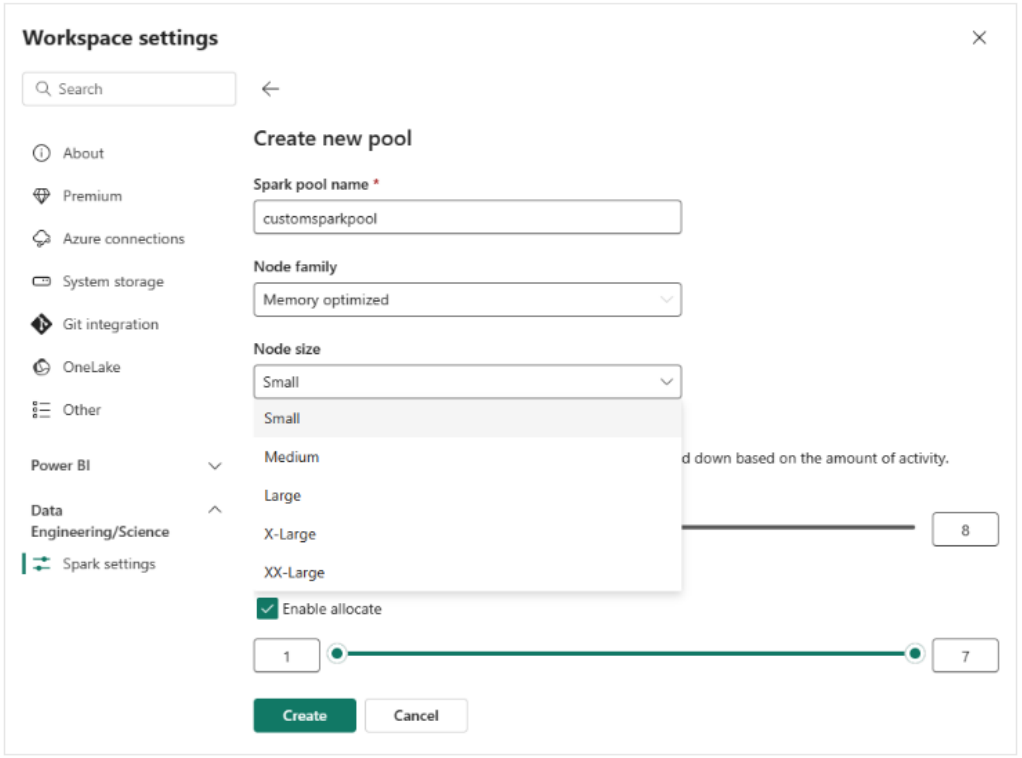


SKU name	Capacity units	Spark VCores	Node size	Default max nodes	Max number of nodes
F2	2	4	Medium	1	1
F4	4	8	Medium	1	1
F8	8	16	Medium	2	2
F16	16	32	Medium	3	4
F32	32	64	Medium	8	8
F64	64	128	Medium	10	16
(Trial Capacity)	64	128	Medium	10	16
F128	128	256	Medium	10	32
F256	256	512	Medium	10	64
F512	512	1024	Medium	10	128
F1024	1024	2048	Medium	10	200
F2048	2048	4096	Medium	10	200

Work with Custom Spark Pools

- Create and configure based on required workloads.





Managing Libraries

- Available via Inline installation and Environment setup.

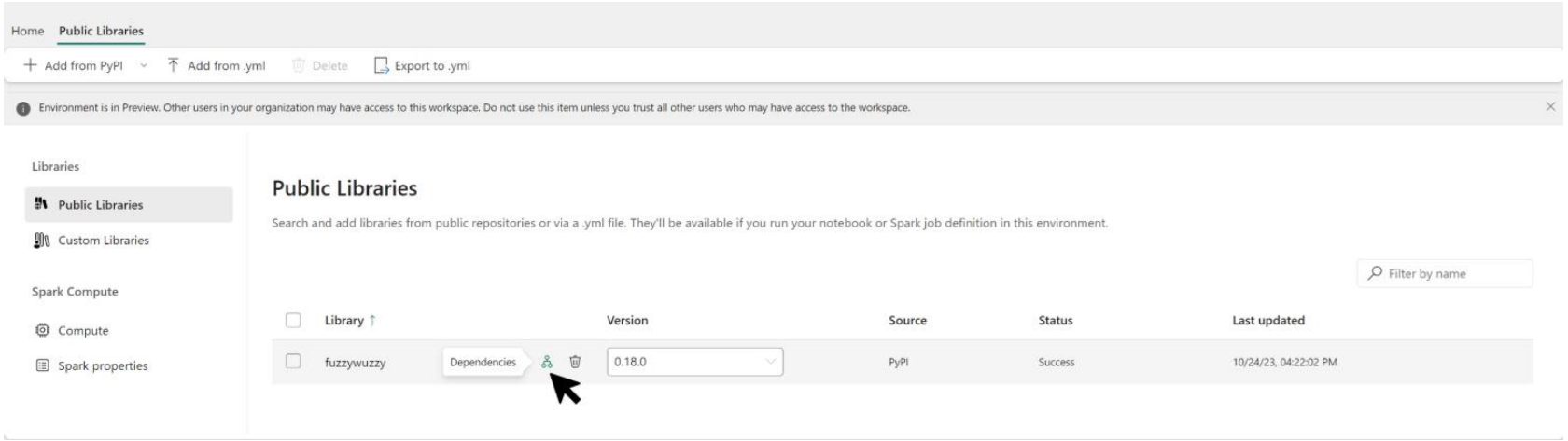
Python Copy

```
%conda install altair          # install latest version through conda command
%conda install vega_datasets   # install latest version through conda command
```

Library type	Environment library management	Inline installation
Python Public (PyPI & Conda)	Supported	Supported
Python Custom (.whl)	Supported	Supported
R Public (CRAN)	Not supported	Supported
R custom (.tar.gz)	Supported as custom library	Supported
Jar	Supported as custom library	Supported

Environments

- Configure Spark compute.
- Manage public and custom Libraries.
- Manage resources.



Manage Notebooks with APIs

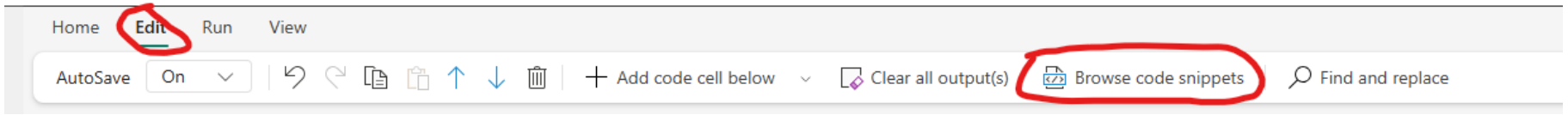
- Manage CRUD operations and Notebook Run instances.

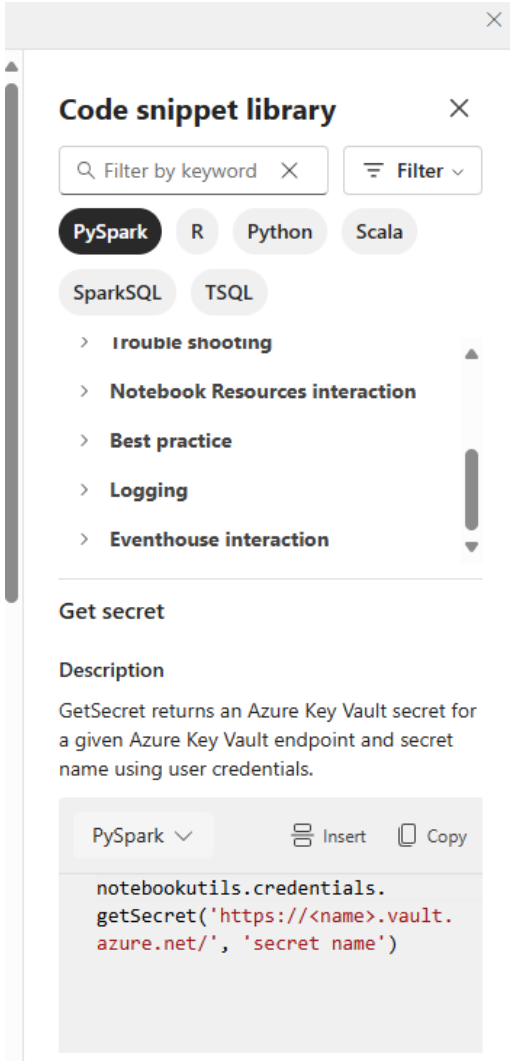
Action	Description
Create item	Creates a notebook inside a workspace.
Update item	Updates the metadata of a notebook.
Update item definition	Updates the content of a notebook.
Delete item	Deletes a notebook.
Get item	Gets the metadata of a notebook.
Get item definition	Gets the content of a notebook.
List item	List all items in a workspace.

Action	Description
Run on demand Item Job	Run notebook with parameterization.
Cancel Item Job Instance	Cancel notebook job run.
Get Item Job Instance	Get notebook run status.

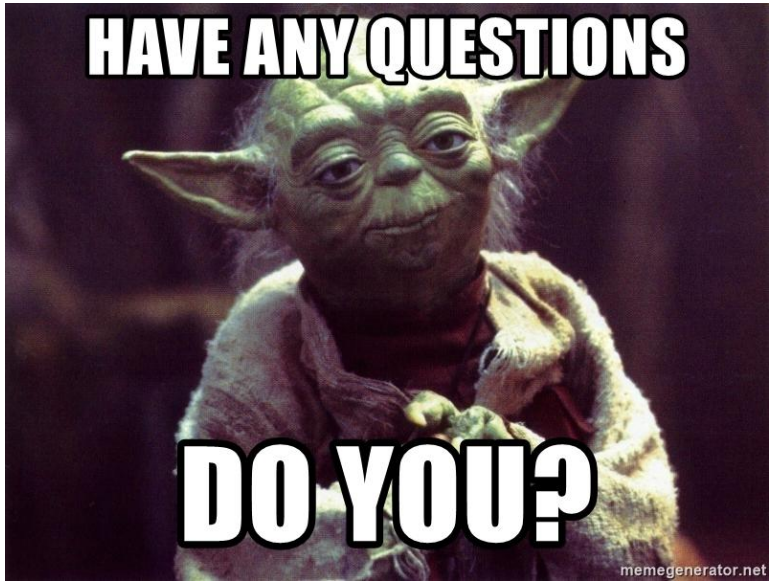
Code Snippets

- Get commonly used code patterns.





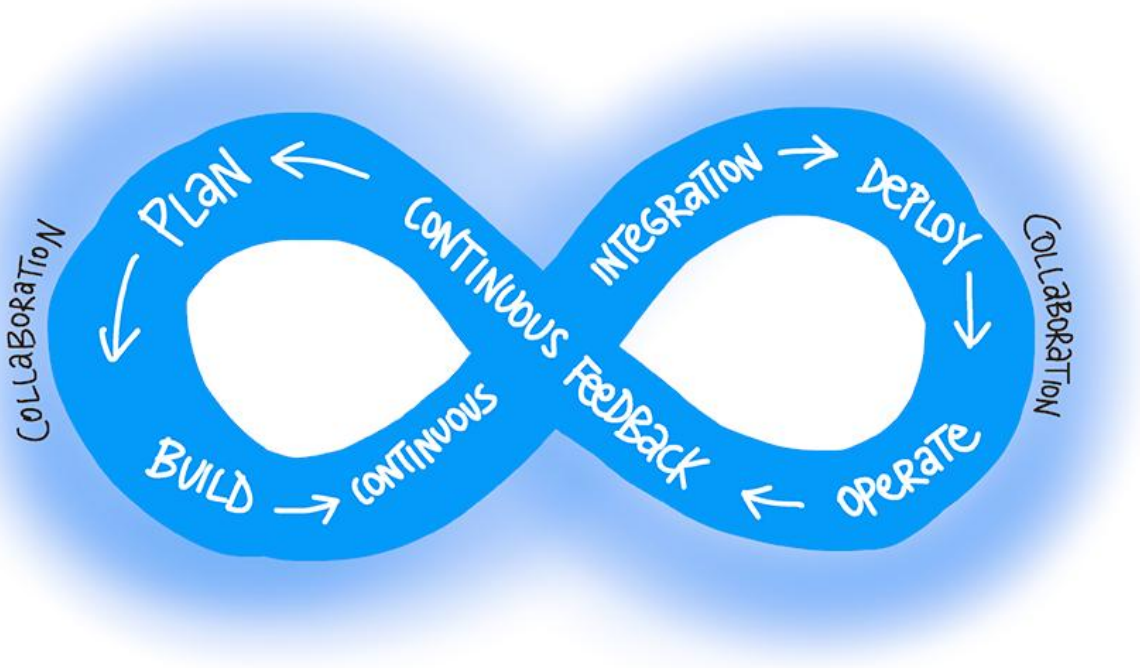
Questions?



CI with Azure DevOps and Git Integration

What is DevOps?

- [What is DevOps? - Training | Microsoft Learn]



DevOps benefits for Fabric

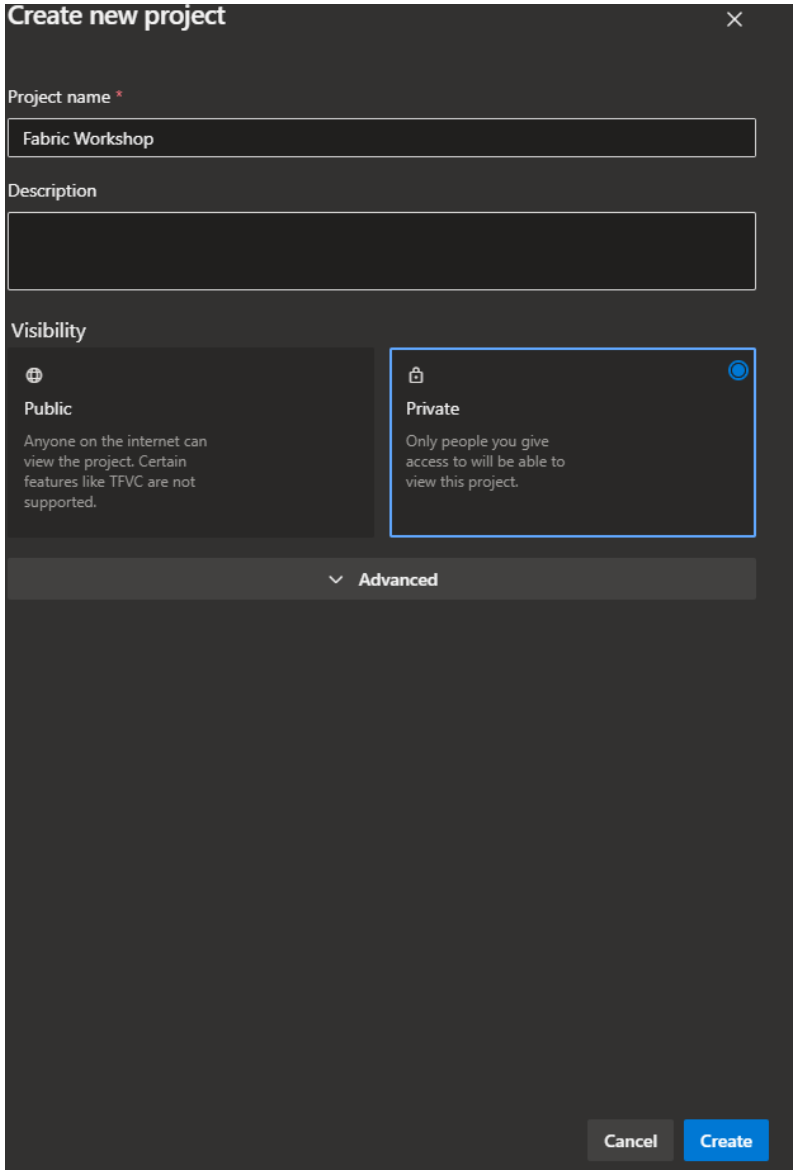
Enhance collaboration (CI)

Leverage your development (CD)

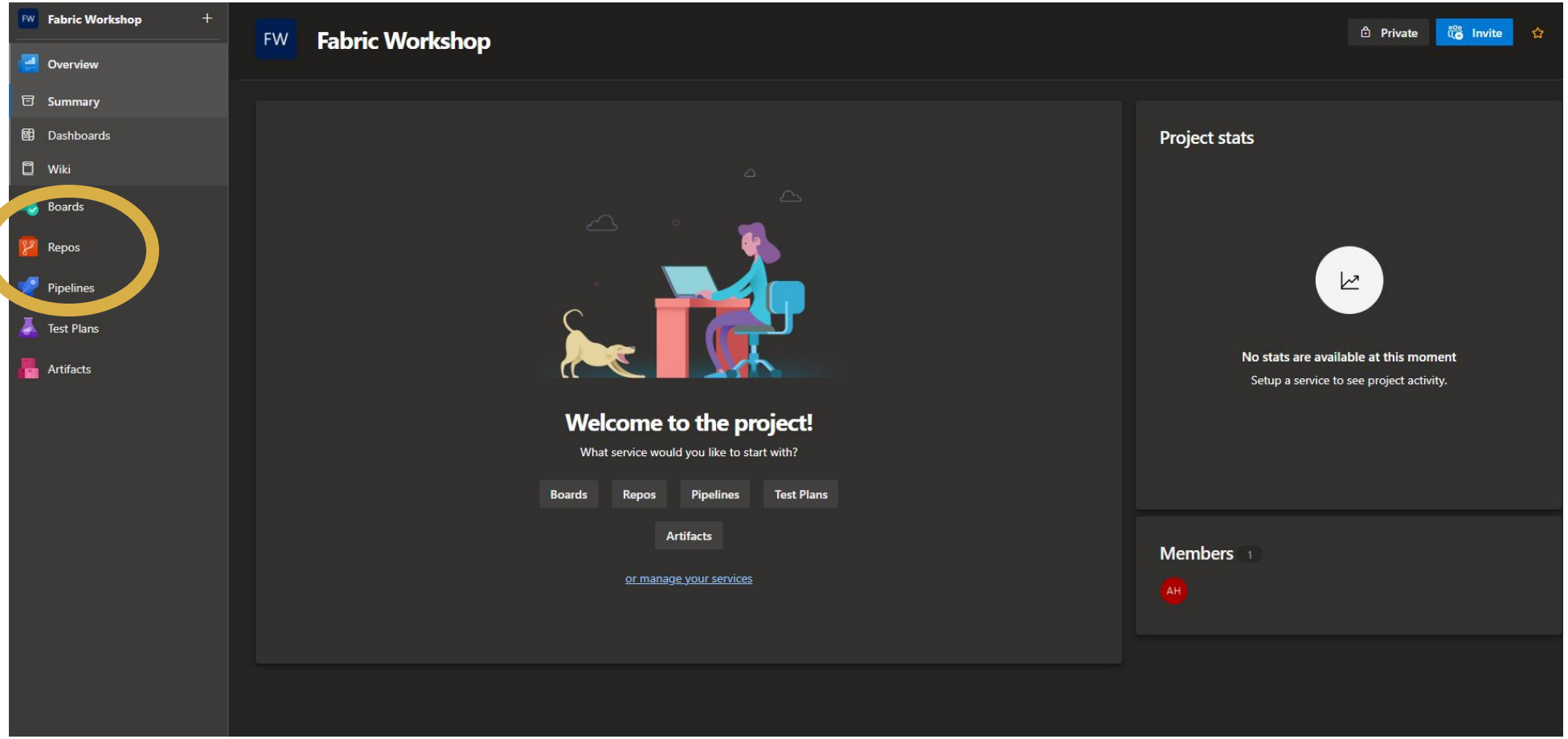
Version control

Create a DevOps Project

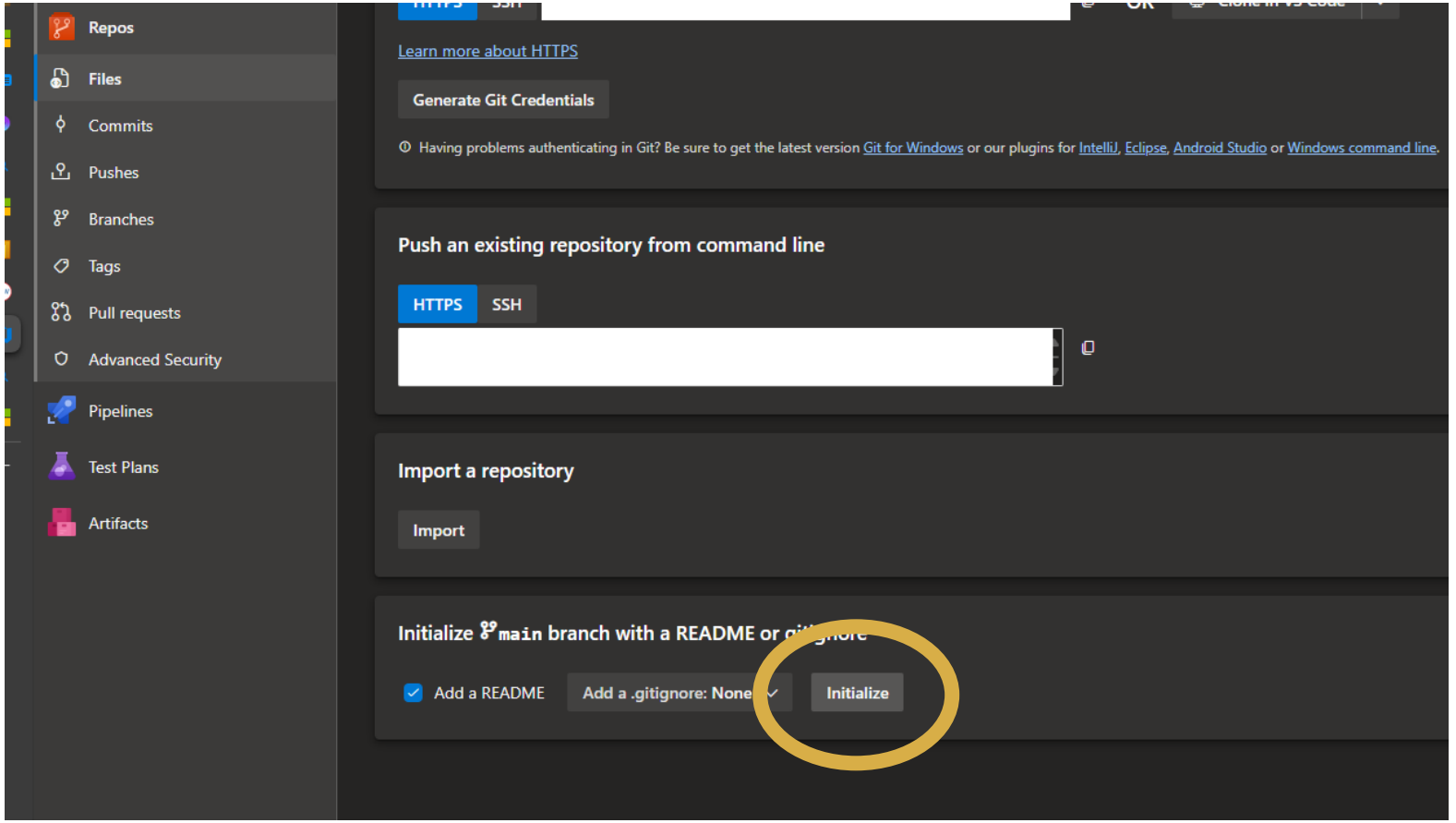
Create a DevOps Project



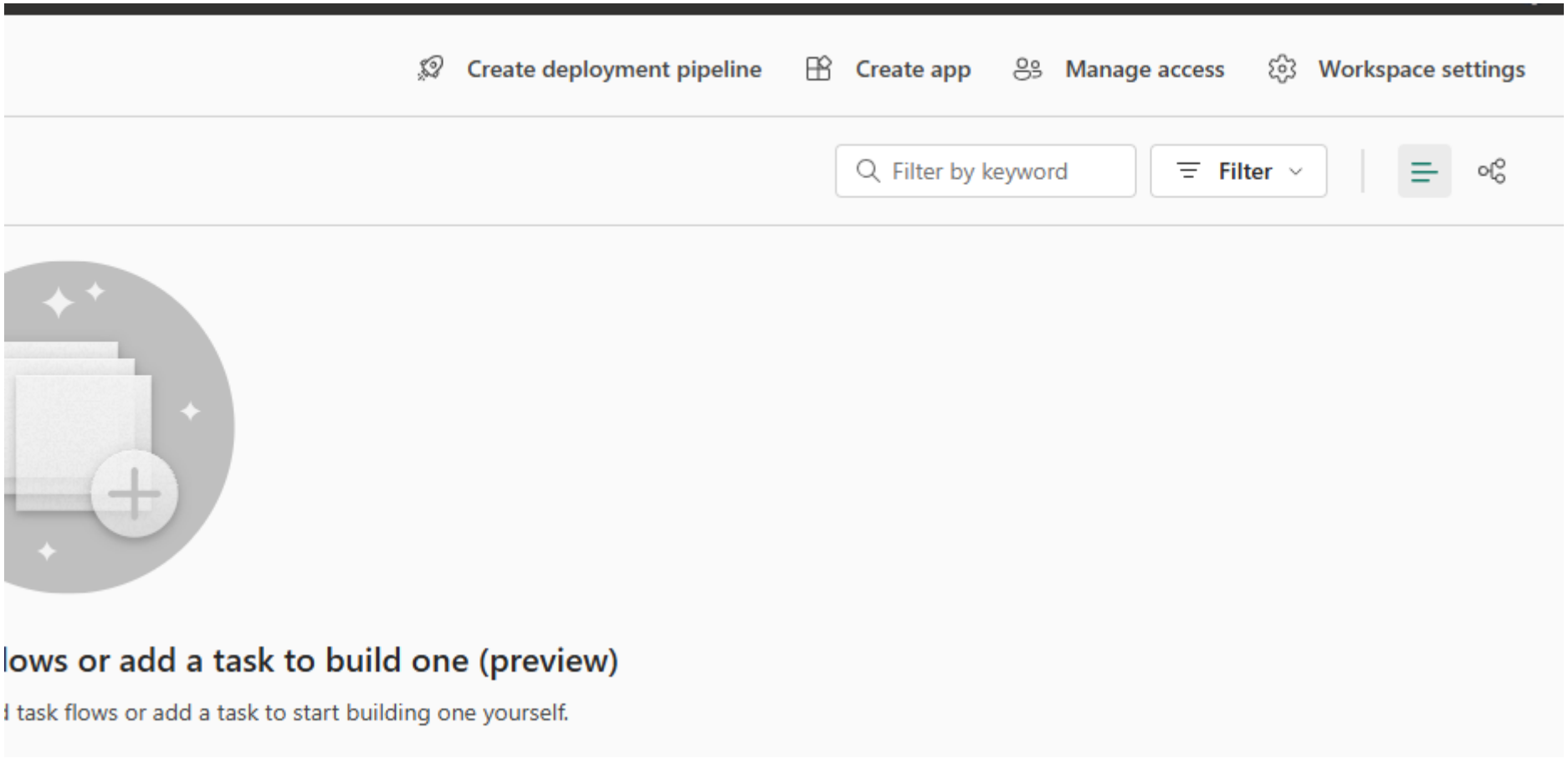
Initialize a Repo



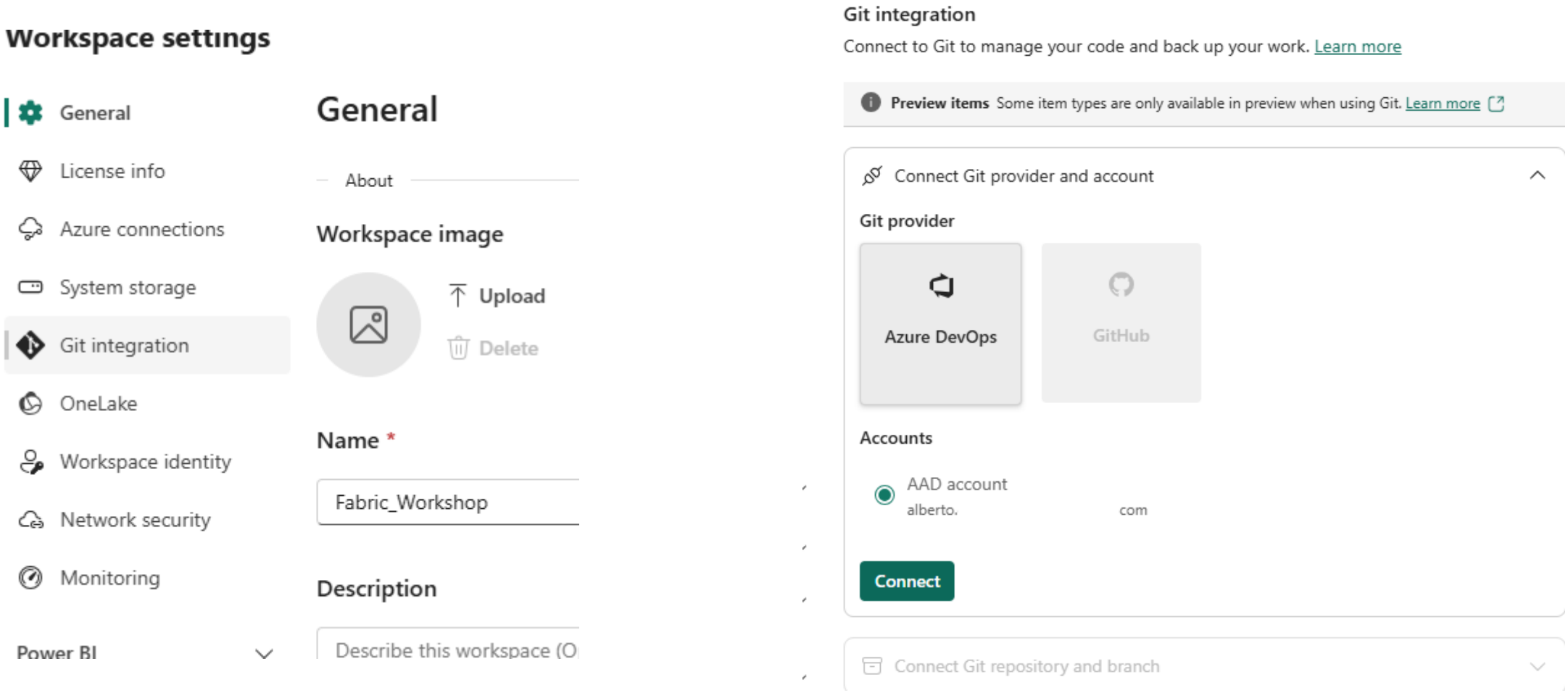
Initialize a Repo



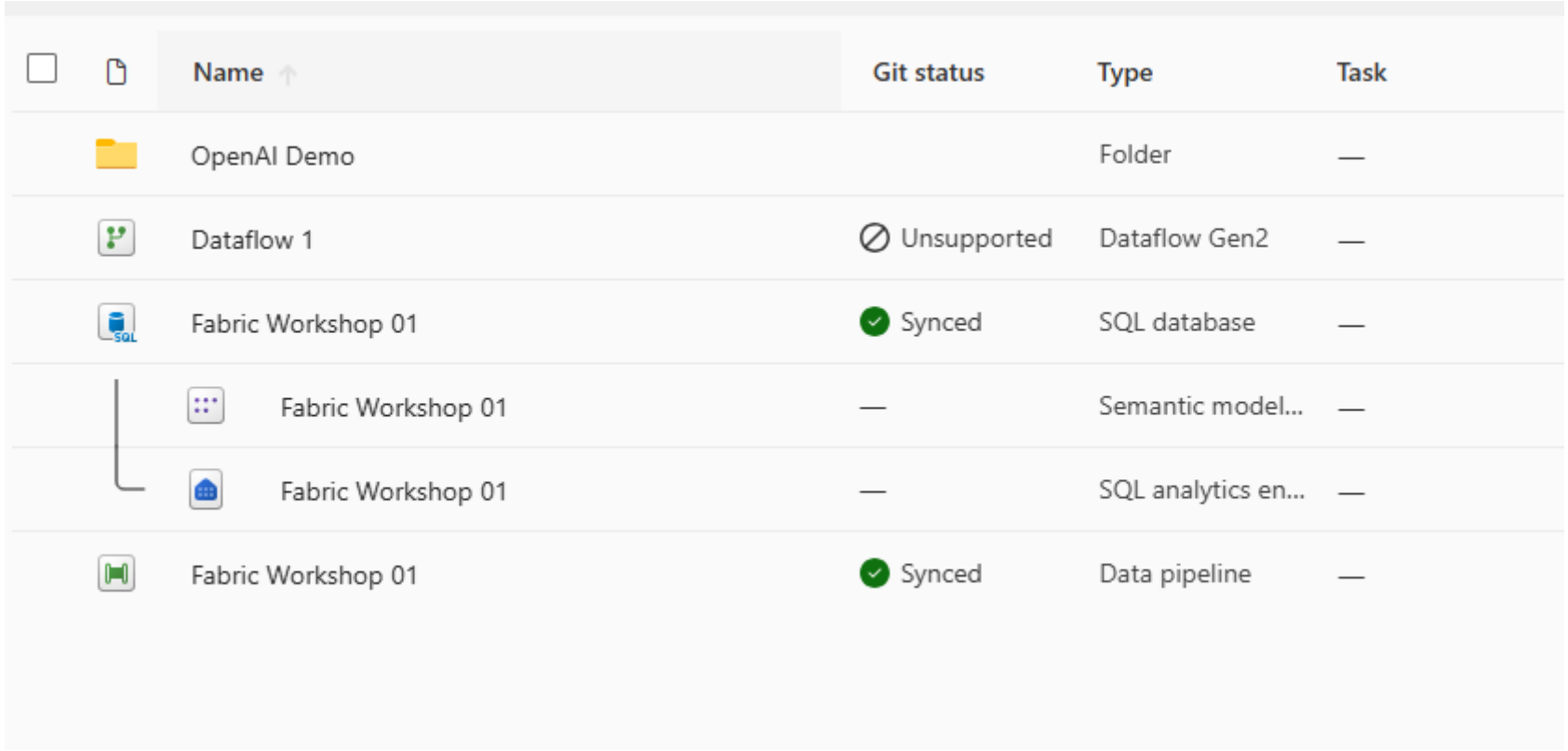
Enable Git Integration



Enable Git Integration



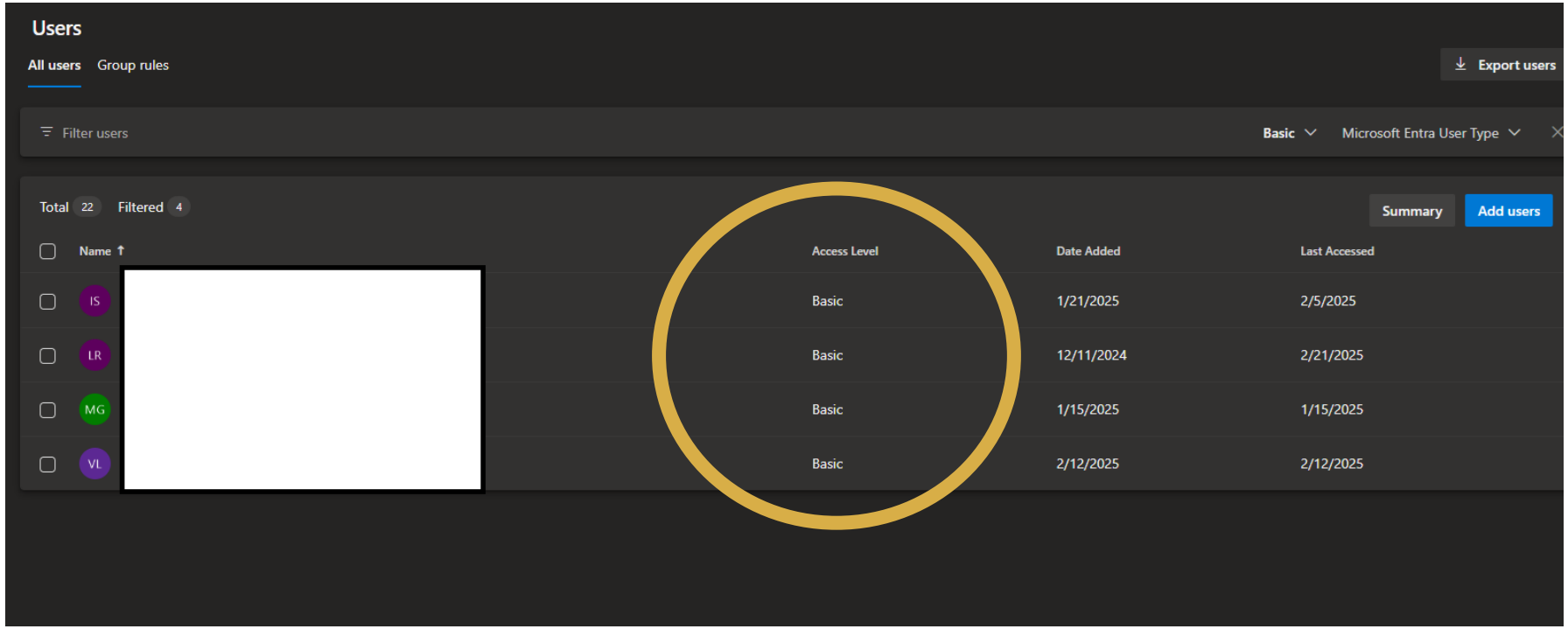
Enable Git Integration



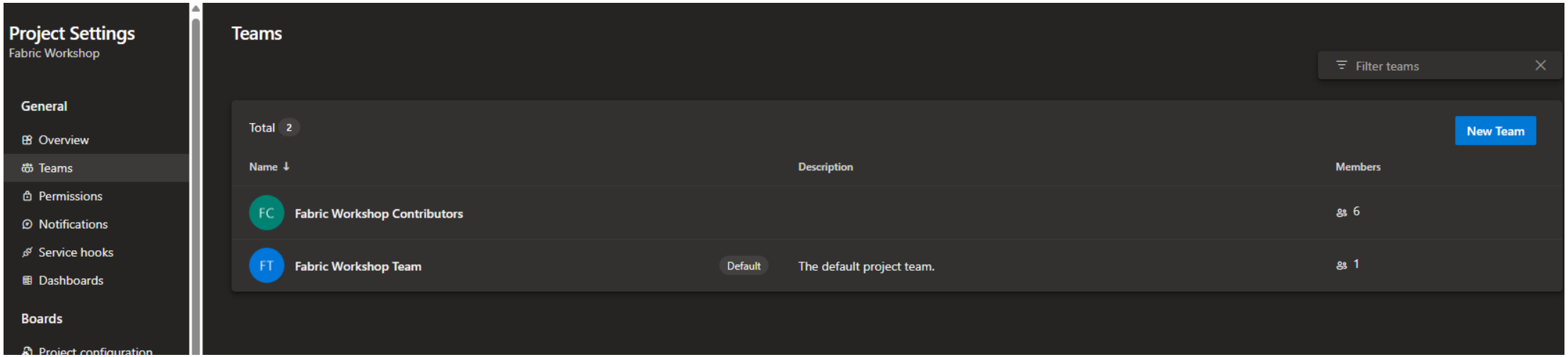
Name	Git status	Type	Task
OpenAI Demo		Folder	—
Dataflow 1	Unsupported	Dataflow Gen2	—
Fabric Workshop 01	Synced	SQL database	—
Fabric Workshop 01	—	Semantic model...	—
Fabric Workshop 01	—	SQL analytics en...	—
Fabric Workshop 01	Synced	Data pipeline	—

Create a Contributors Team

(Users require Basic access at least)

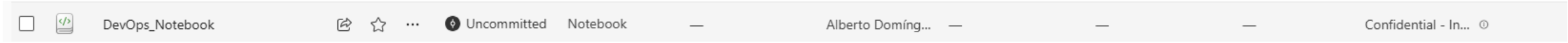


Create Contributors Team

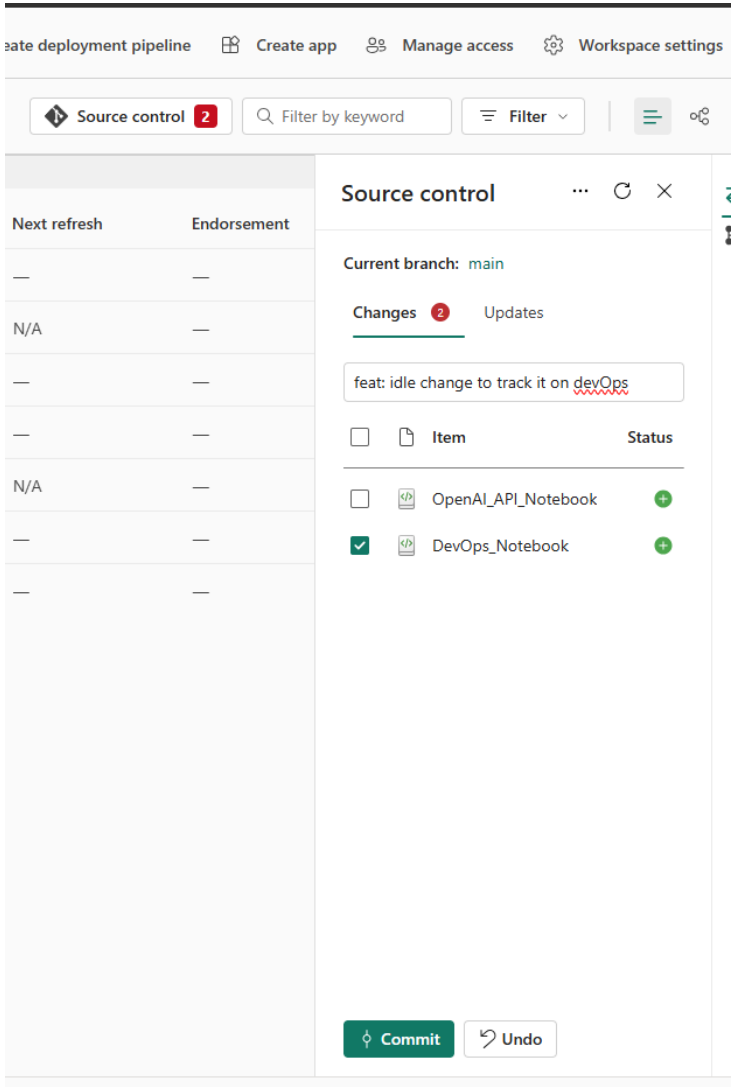


Commit and Sync your Items

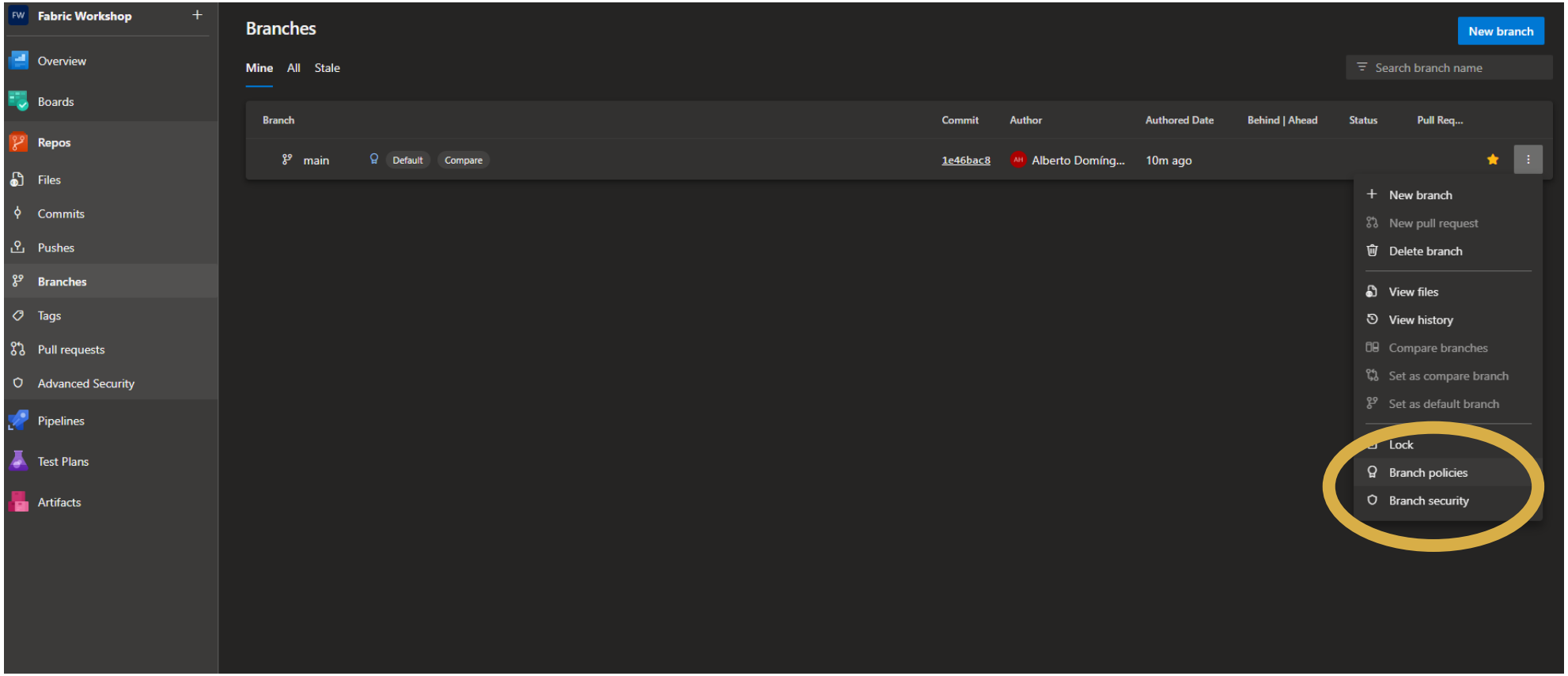
When someone edits an item, it will appear as uncommited:



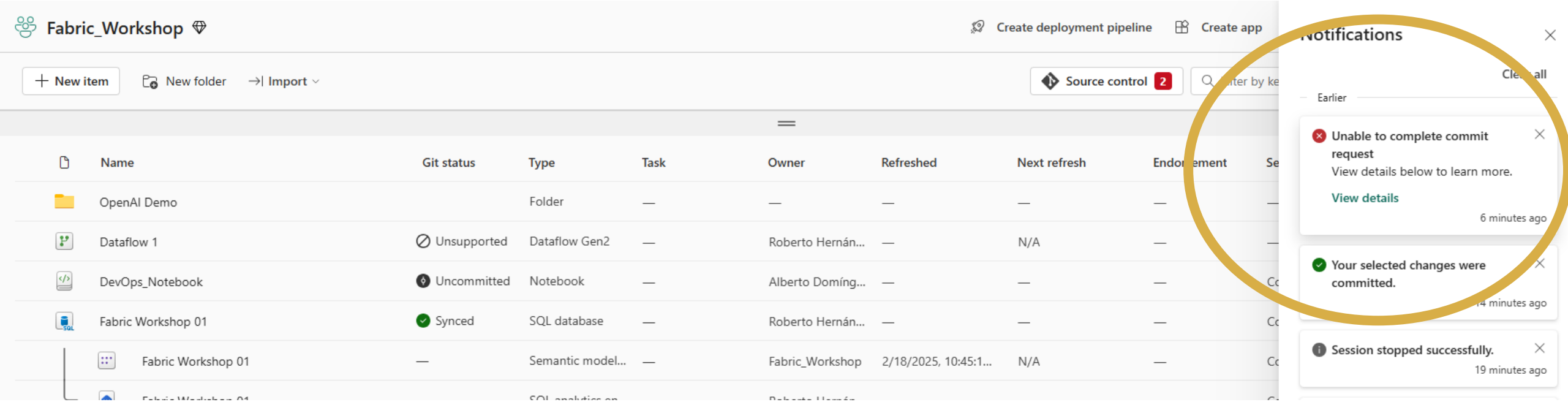
Commit and Sync your Items



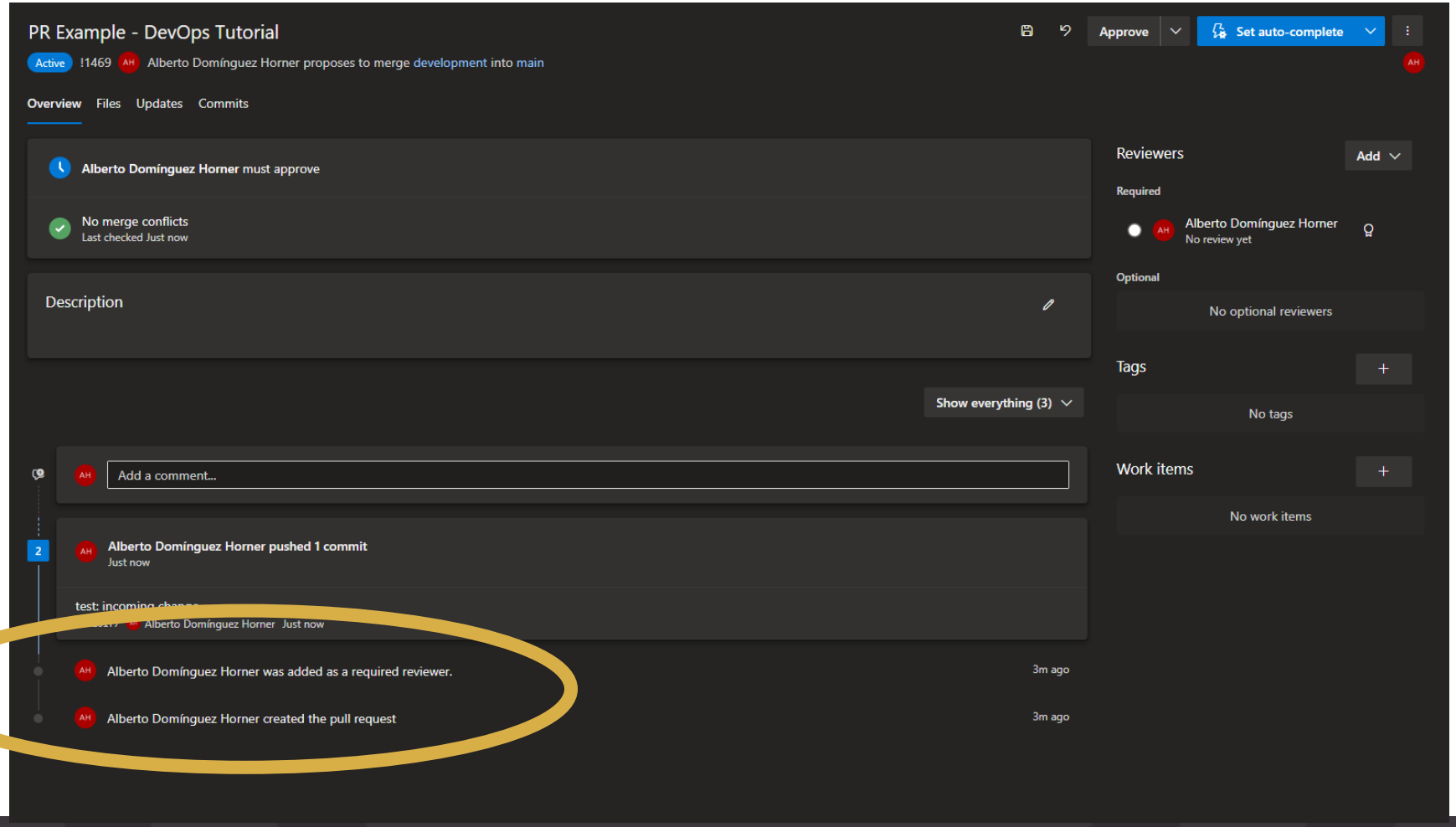
Create Branch policies



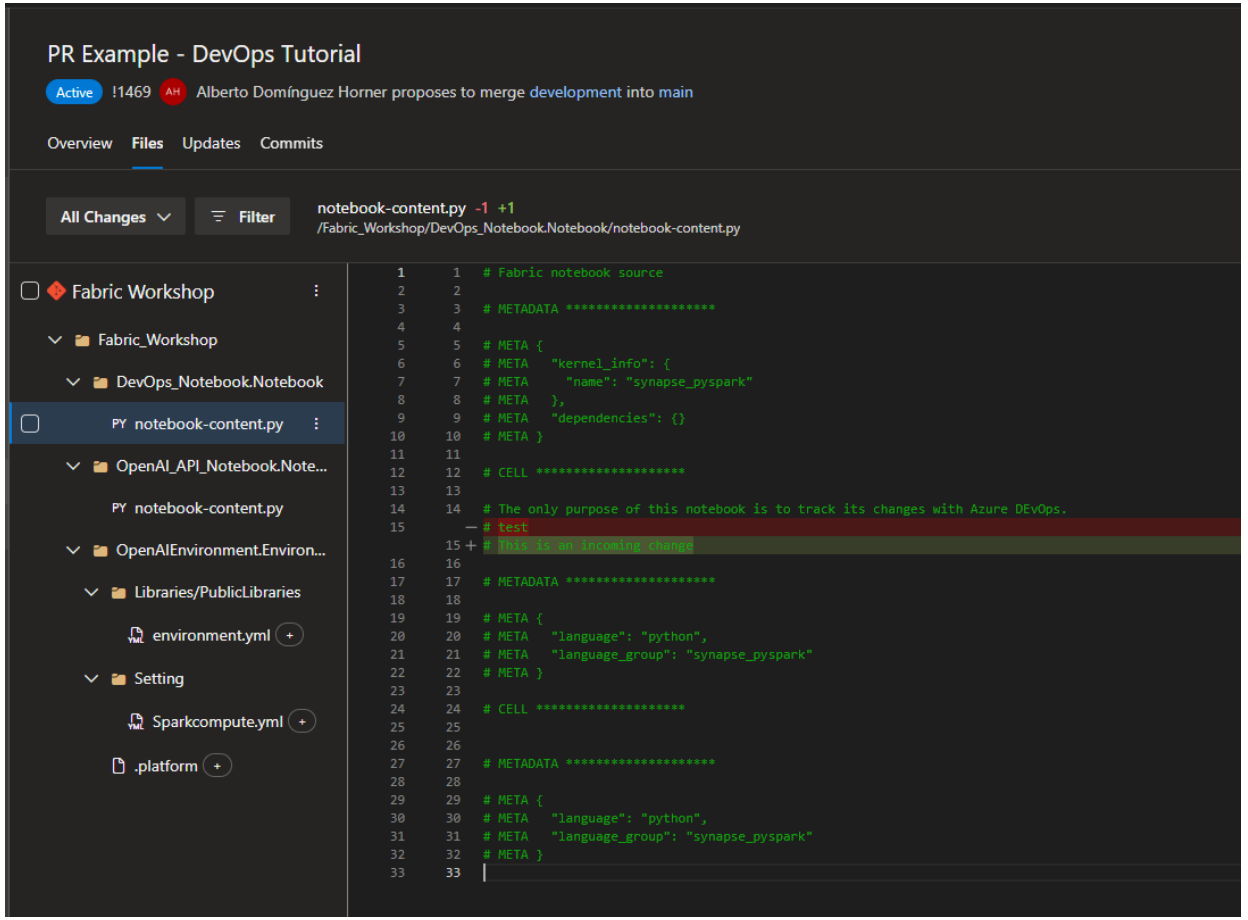
Branch Policies



Branch Policies



Continuous Integration by Pull Requests



CD with Deployment Pipelines

AI Skills

What are AI Skills?

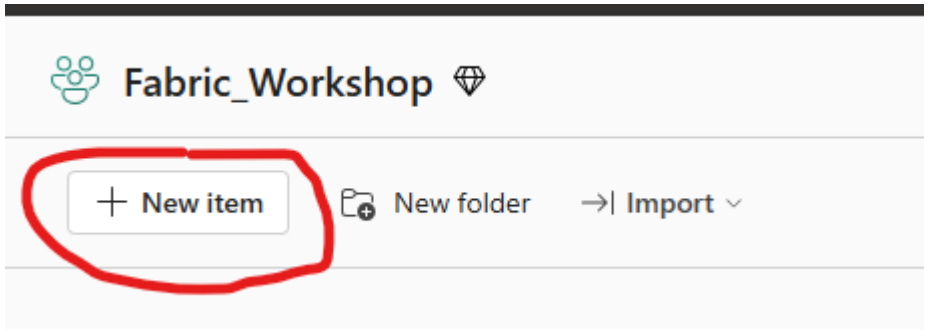
- Generative AI system designed to generate queries that answer questions about your data.
- Trained via data from a Warehouse or Lakehouse.
- Additionally trained and tweaked via User-defined queries and instructions.
- Shareable and accessible via numerous options:
  - Programmatically
  - Microsoft Teams
  - Microsoft Fabric

Prerequisites.

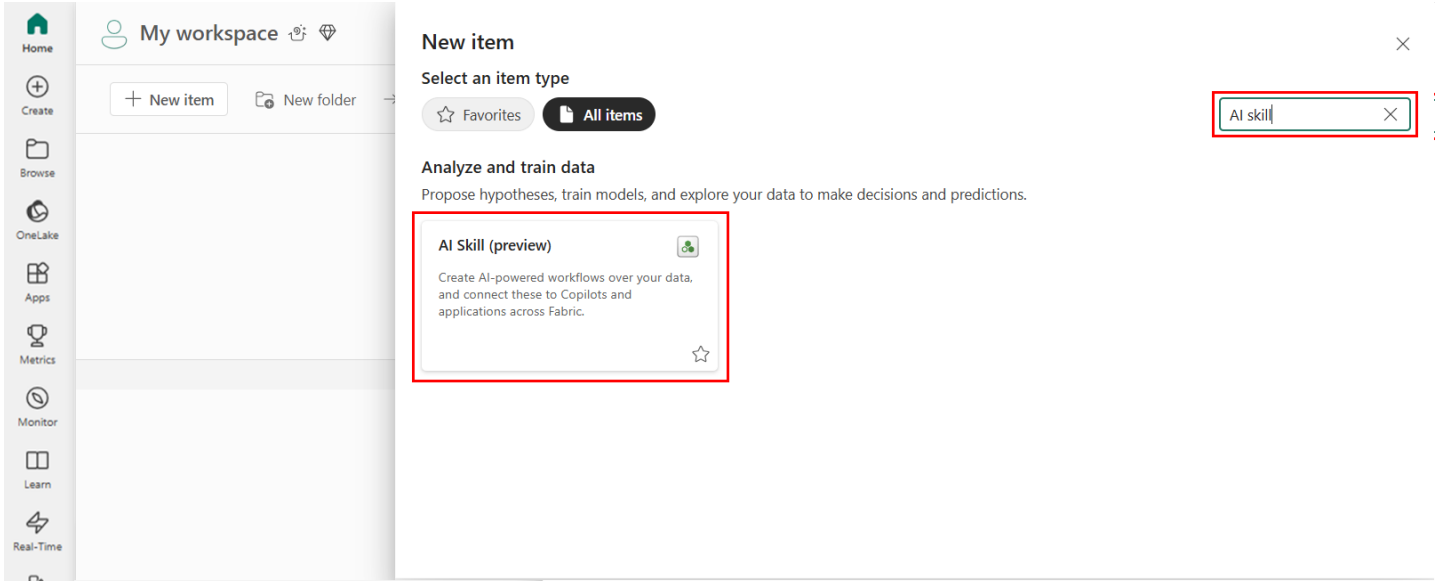
- A paid F64 or higher Fabric capacity resource.
- AI skill tenant switch is enabled.
- Copilot tenant switch is enabled.
- Cross-geo sharing for AI is enabled, if relevant.
- A warehouse or lakehouse with data.

Getting started in 2 steps…

1. Create Fabric Item “AI Skills”



2. Give it a name



Long list of limitations…

Limitations

The AI skill is currently in public preview and has limitations. Updates will improve the AI skill over time.

- Generative AI doesn't interpret the results of an executed T-SQL query. It only generates that query.
- The AI skill might return incorrect answers. You should test the AI skill with your colleagues to verify that it answers questions as expected. If it makes mistakes, provide it with more examples and instructions.
- Only T-SQL queries on warehouses and lakehouses are supported.
- The AI skill only generates T-SQL "read" queries. It doesn't generate T-SQL queries that create, update, or delete data.
- The AI skill can only access data that you provide. It only uses the data resource configurations that you provide.
- The AI skill has data access permissions that match the permissions granted to the AI skill questioner. This is true when the AI skill is published to other locations, for example, Copilot for Microsoft 365 or Microsoft Copilot Studio.
- You can't use the AI skill to access unstructured data resources. These resources include .pdf, .docx, or .txt files, for example.
- At this time, you can only select a single warehouse or a single lakehouse.
- The AI skill doesn't support a conversational interface. Every question must be fully self-contained. It doesn't remember earlier questions.
- It blocks non-English language questions or instructions.
- You can't connect the AI skill to Fabric copilots, Microsoft Teams, or other experiences outside of Fabric.
- You can't change the LLM that the AI skill uses.
- The AI skill loses accuracy if you use nondescriptive column names.
- The AI skill loses accuracy if you use large schemas with dozens of tables.
- The AI skill is in a preview status. It has a limited scope and it might have bugs. Because of these considerations, we recommend that you avoid its use in production systems. Also avoid its use for critical decisions.
- Nondescriptive data resource column and table names have a significant, negative impact on generated T-SQL query quality. We recommend the use of descriptive names.
- Use of too many columns and tables might lower AI skill performance.
- The AI skill is currently designed to handle simple queries. Complex queries that require many joins or sophisticated logic tend to have lower reliability.

(Narrowed down) Limitations

- Only T-SQL queries provided on a single Warehouse/Lakehouse data.
- Read-only queries.
- Can’t access unstructured data.
- No conversation interface supported (each question/query is self-contained).
- Loss of perfomance with wider tables and with non-descriptive columns.
- The AI skill loses accuracy if you use large schemas with dozens of tables.
- Designed for simpler queries.
- Still in Preview.