

# Fabric Workshop

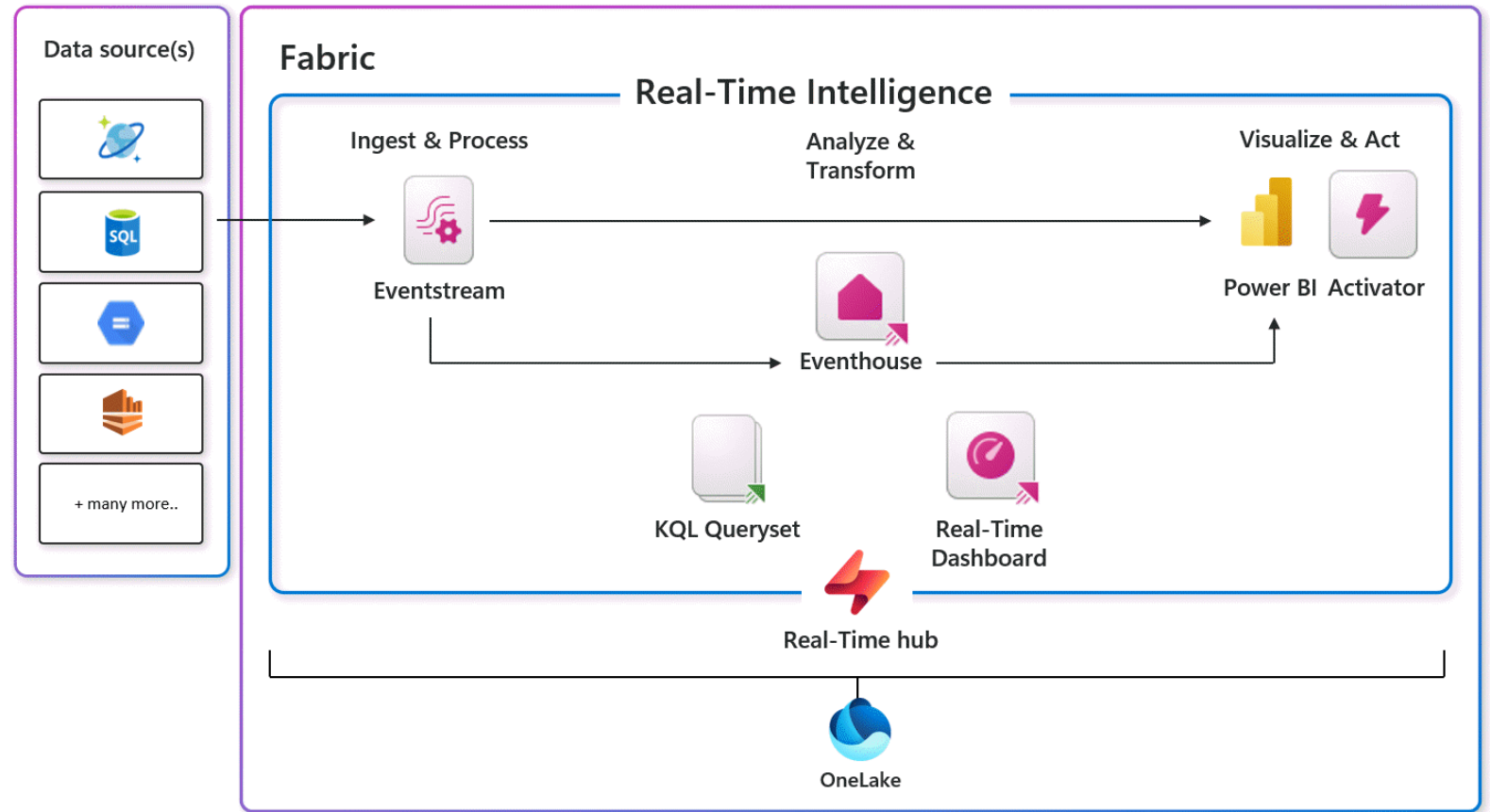
2025-02-28 Part 1

# Real-Time Intelligence

The background of the slide is split vertically. The left half is a dark charcoal gray, featuring several concentric, thick-lined circles in a slightly lighter shade of gray, creating a ripple effect. The right half is a solid, bright white.

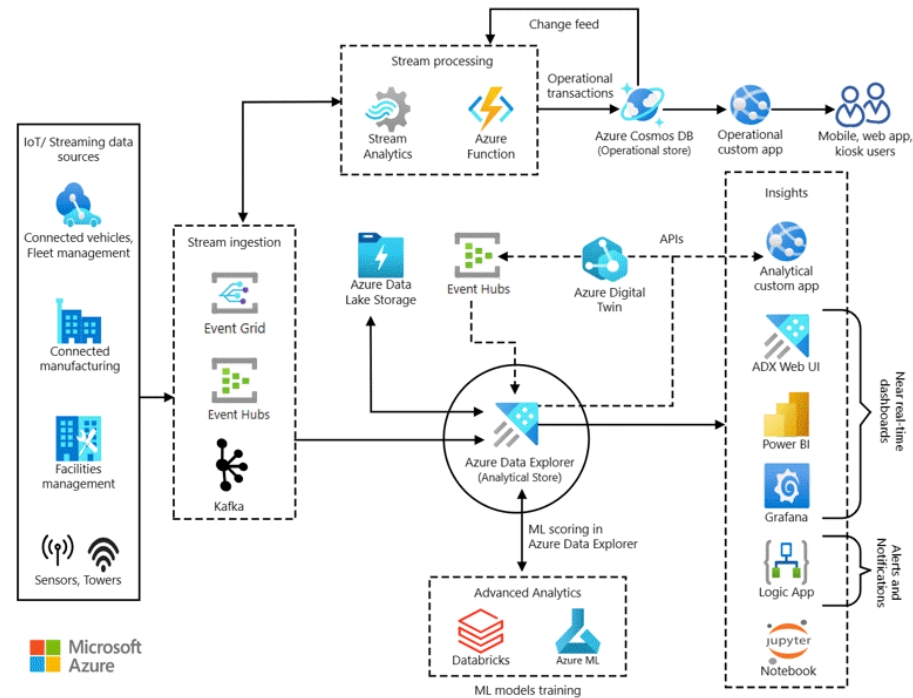
# Real-Time Intelligence

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

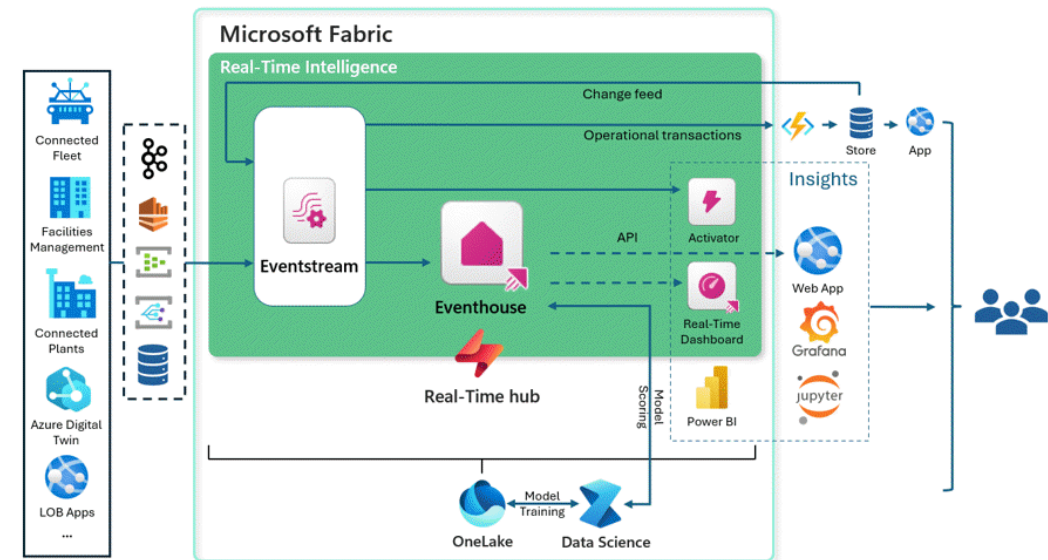


# Real-Time Intelligence

## Azure PaaS-based solution architecture



## Fabric Real-time Intelligence solution architecture



# 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



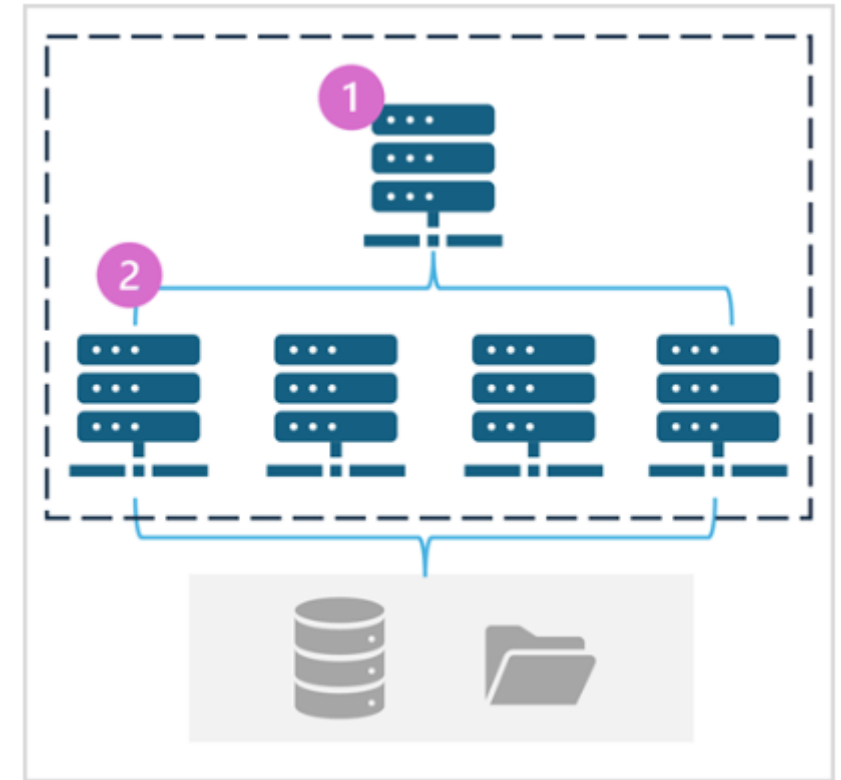
# End-to-end Demo

The background of the slide is split vertically. The left half is dark gray and features several concentric, slightly irregular circles in varying shades of gray, creating a ripple effect. The right half is a solid, bright white.

# 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**.



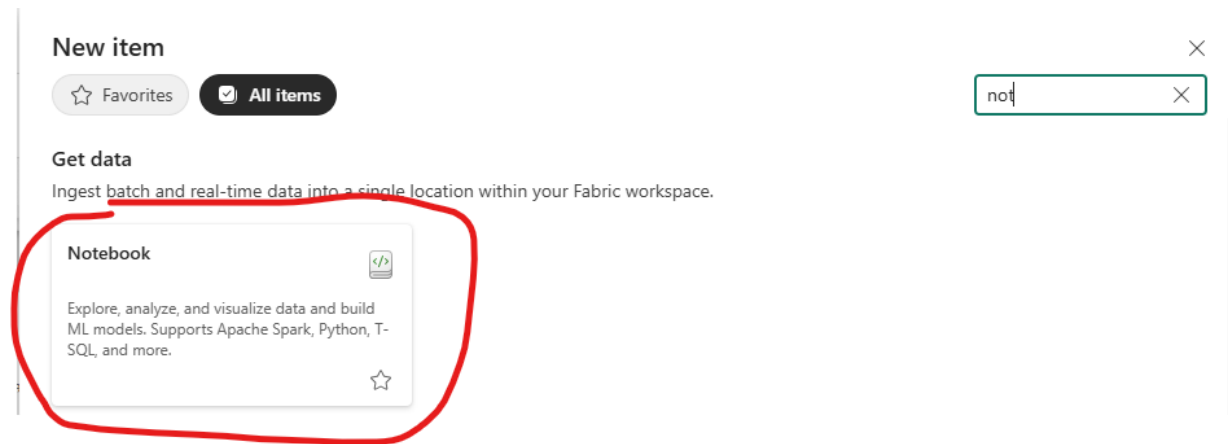
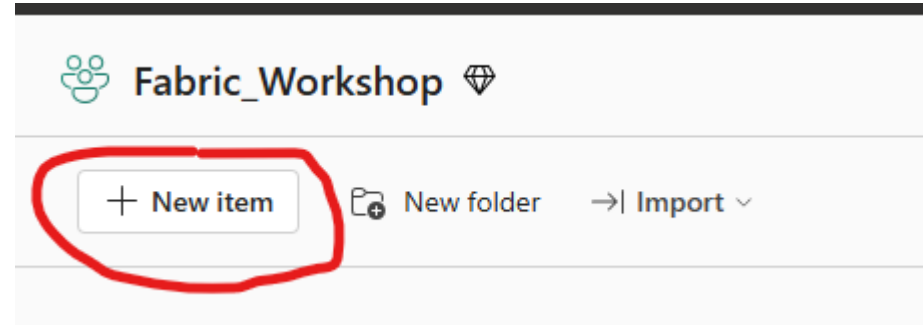
1. **Head node** coordinates via **driver**.
2. **Worker nodes** complete tasks via **executors**.



# 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

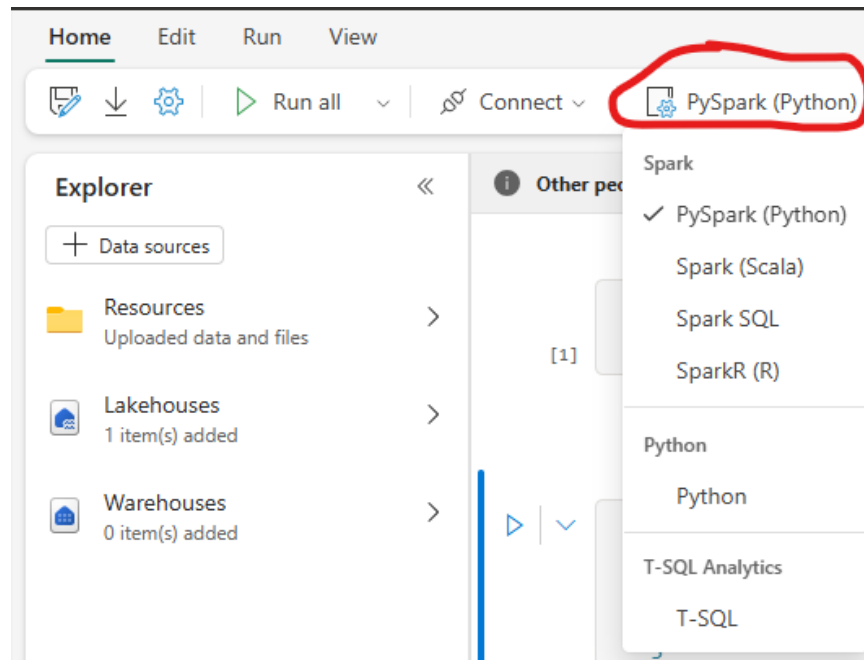




# Fabric Notebooks Advanced Features

# Different Programming Languages

- Set for entire notebook for single cell.

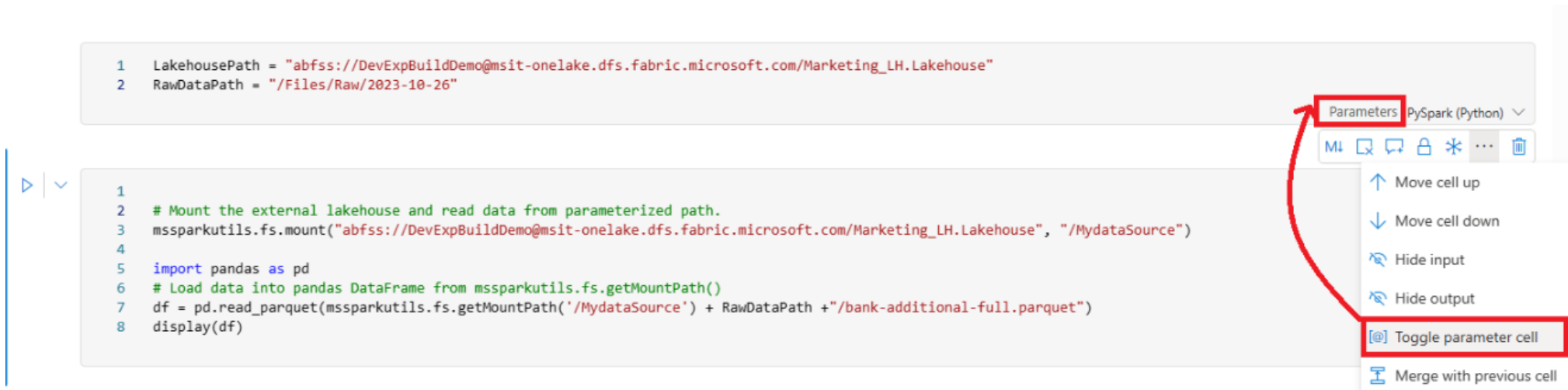


| Magic command | Language | Description   |
|---------------|----------|---|
| %%pyspark     | Python   | Execute a <b>Python</b> query against Apache Spark Context.   |
| %%spark       | Scala    | Execute a <b>Scala</b> query against Apache Spark Context.    |
| %%sql         | SparkSQL | Execute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



# Parameter Cells

- Pass arguments via runMultiple or Pipeline.



The screenshot shows a Jupyter Notebook interface with two code cells. The first cell contains two lines of Python code defining paths. The second cell contains a multi-line Python script that mounts a lakehouse, imports pandas, and reads a parquet file. A context menu is open for the second cell, with a red arrow pointing to the 'Parameters' tab and another red box highlighting the '@ Toggle parameter cell' option.

```
1 LakehousePath = "abfss://DevExpBuildDemo@msit-onelake.dfs.fabric.microsoft.com/Marketing_LH.Lakehouse"
2 RawDataPath = "/Files/Raw/2023-10-26"
```

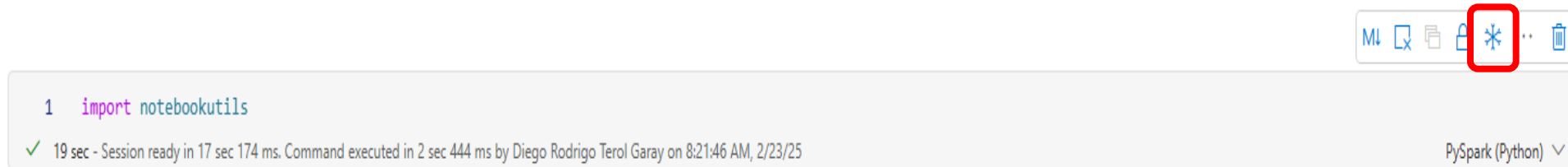
```
1
2 # Mount the external lakehouse and read data from parameterized path.
3 mssparkutils.fs.mount("abfss://DevExpBuildDemo@msit-onelake.dfs.fabric.microsoft.com/Marketing_LH.Lakehouse", "/MydataSource")
4
5 import pandas as pd
6 # Load data into pandas DataFrame from mssparkutils.fs.getMountPath()
7 df = pd.read_parquet(mssparkutils.fs.getMountPath('/MydataSource') + RawDataPath + "/bank-additional-full.parquet")
8 display(df)
```

Parameters PySpark (Python) ▾

- Move cell up
- Move cell down
- Hide input
- Hide output
- @ Toggle parameter cell**
- Merge with previous cell

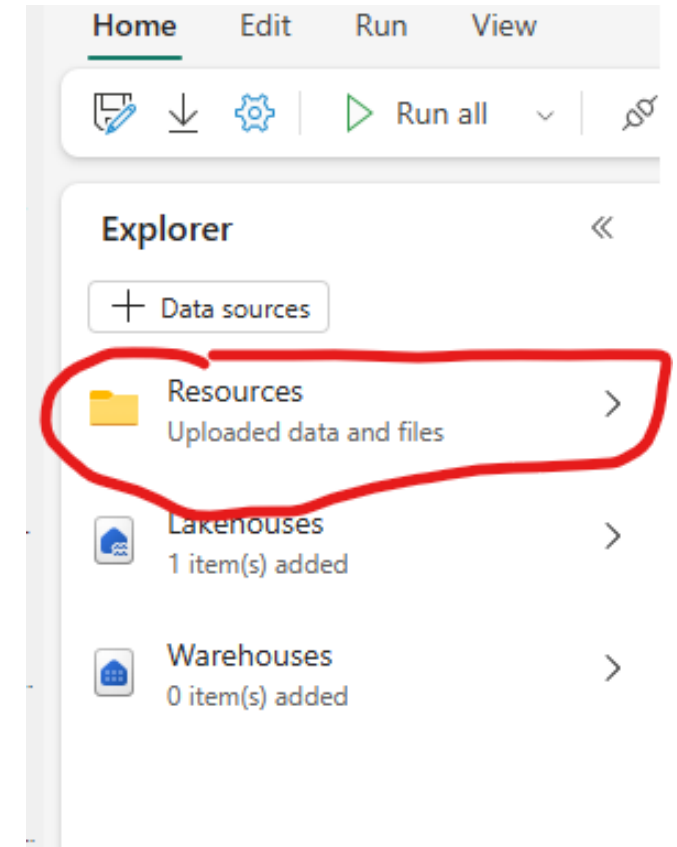
# Frozen Cells

- Make cells read-only.



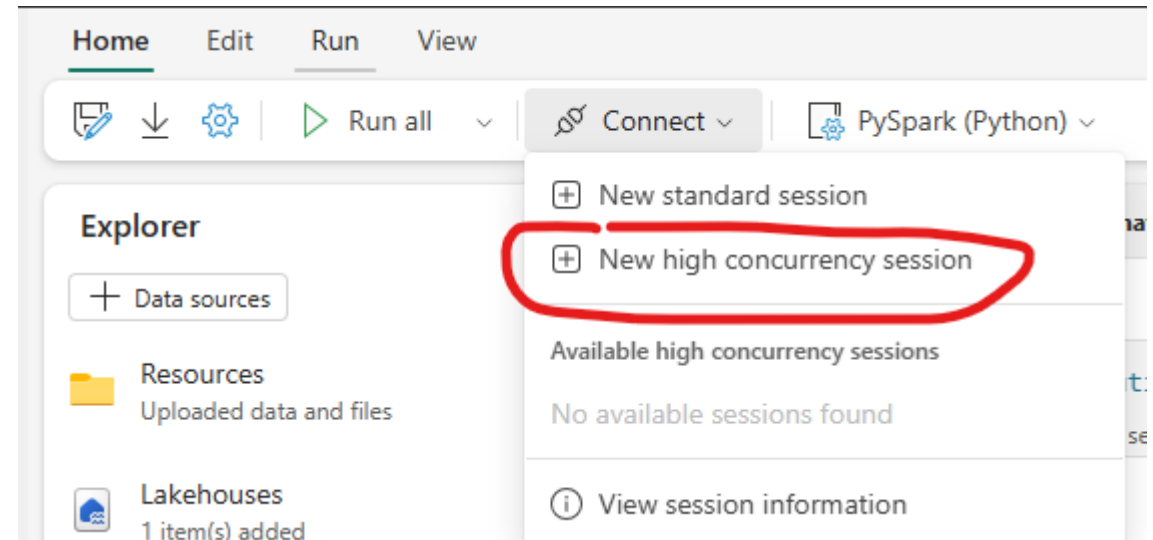
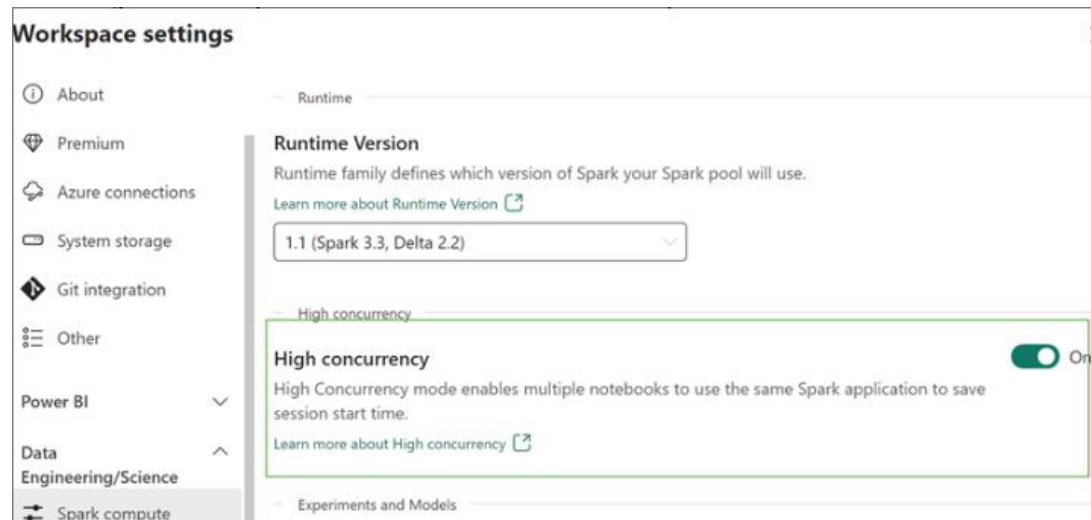
# Notebook Resources

- 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 and required resources.

**Edit pool**

Spark pool name \*

StarterPool

Node family

Memory optimized

Node size

Medium

Autoscale

If enabled, your Apache Spark pool will automatically scale up and down based on the amount of activity.

☒ Enable autoscale

1

16

Dynamically allocate executors

☒ Enable allocate

1

15

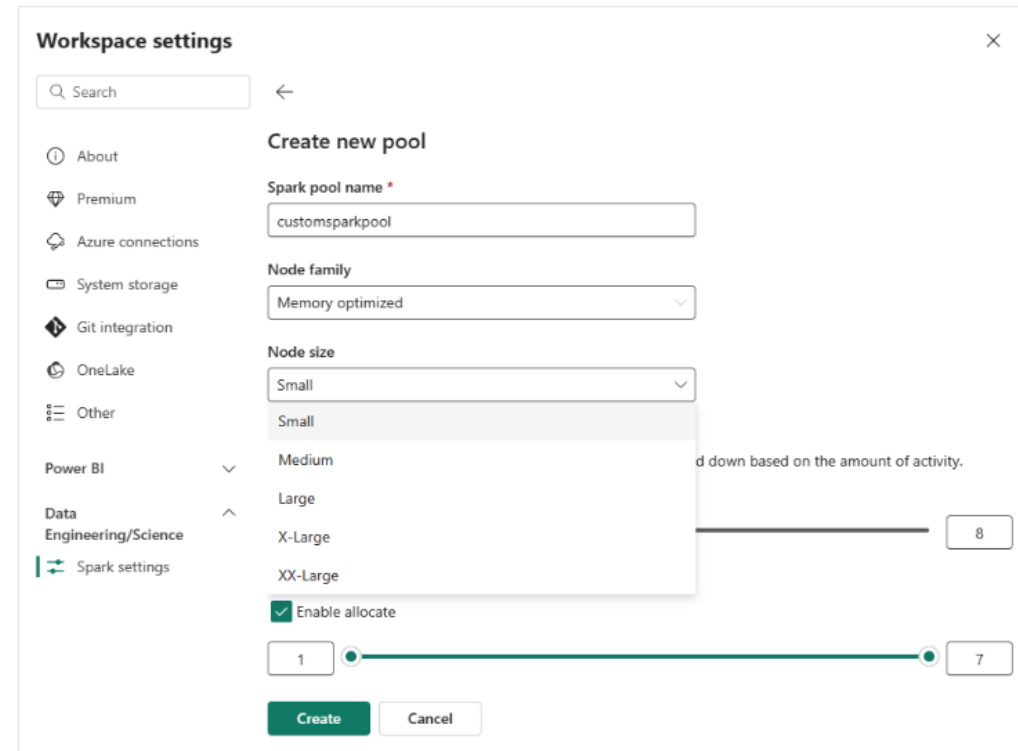
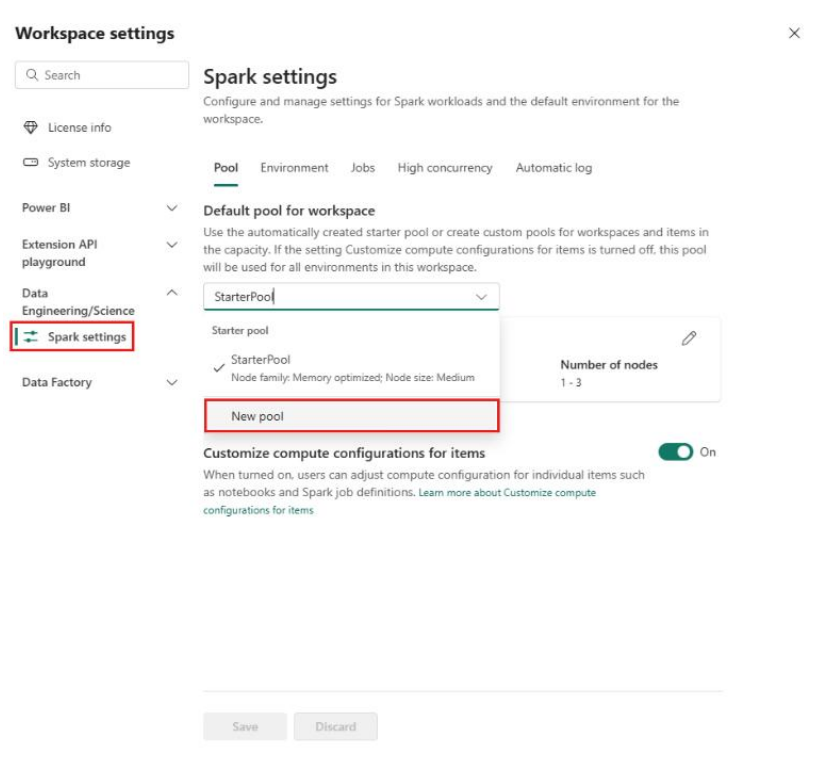
Save

Discard

| 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.

The screenshot shows the 'Public Libraries' tab in the Databricks workspace. The interface includes a top navigation bar with 'Home' and 'Public Libraries' tabs. Below the navigation bar, there are buttons for '+ Add from PyPI', '+ Add from .yaml', 'Delete', and 'Export to .yaml'. A warning message states: 'Environment is in Preview. Other users in your organization may have access to this workspace. Do not use this item unless you trust all other users who may have access to the workspace.' The left sidebar contains a 'Libraries' section with 'Public Libraries' and 'Custom Libraries' options, and a 'Spark Compute' section with 'Compute' and 'Spark properties' options. The main content area is titled 'Public Libraries' and contains a search bar 'Filter by name'. Below the search bar is a table with columns: 'Library', 'Version', 'Source', 'Status', and 'Last updated'. The table has one row for the library 'fuzzywuzzy' with version '0.18.0', source 'PyPI', status 'Success', and last updated '10/24/23, 04:22:02 PM'. A 'Dependencies' button is located next to the library name, and a mouse cursor is pointing at it.

Home Public Libraries

+ Add from PyPI + Add from .yaml Delete Export to .yaml

Environment is in Preview. Other users in your organization may have access to this workspace. Do not use this item unless you trust all other users who may have access to the workspace.

Libraries

- Public Libraries
- Custom Libraries

Spark Compute

- Compute
- Spark properties

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.

Filter by name

| Library    | Version | Source | Status  | Last updated          |
|------------|---------|--------|---------|-----------------------|
| fuzzywuzzy | 0.18.0  | PyPI   | Success | 10/24/23, 04:22:02 PM |



# Manage Notebooks with APIs

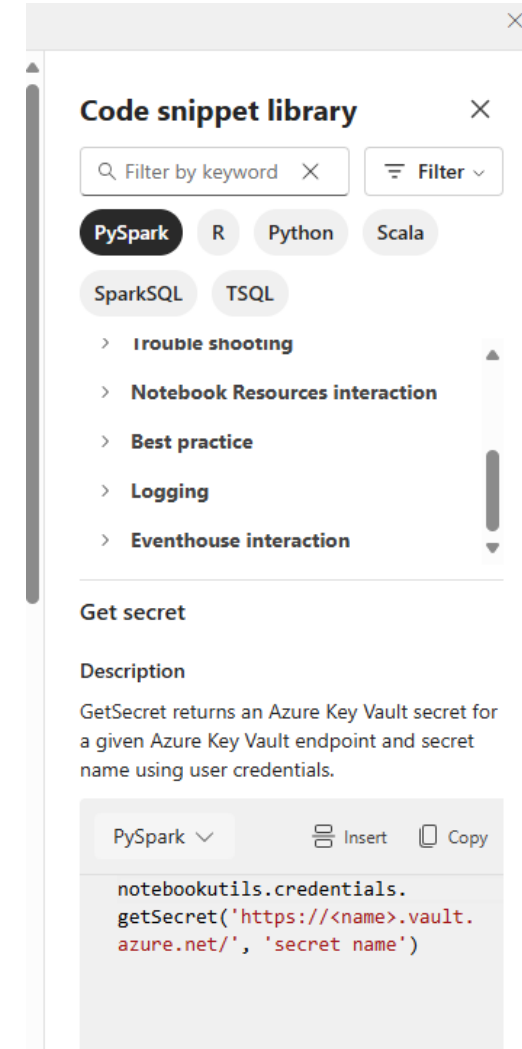
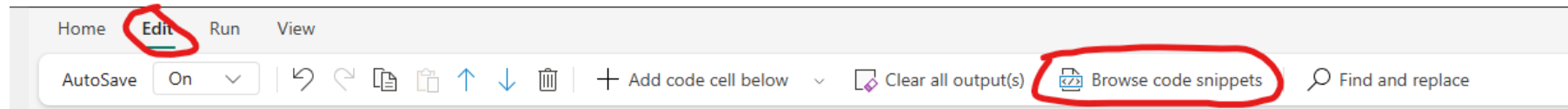
- Manage CRUD operations and Notebook Run instances.

| 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.



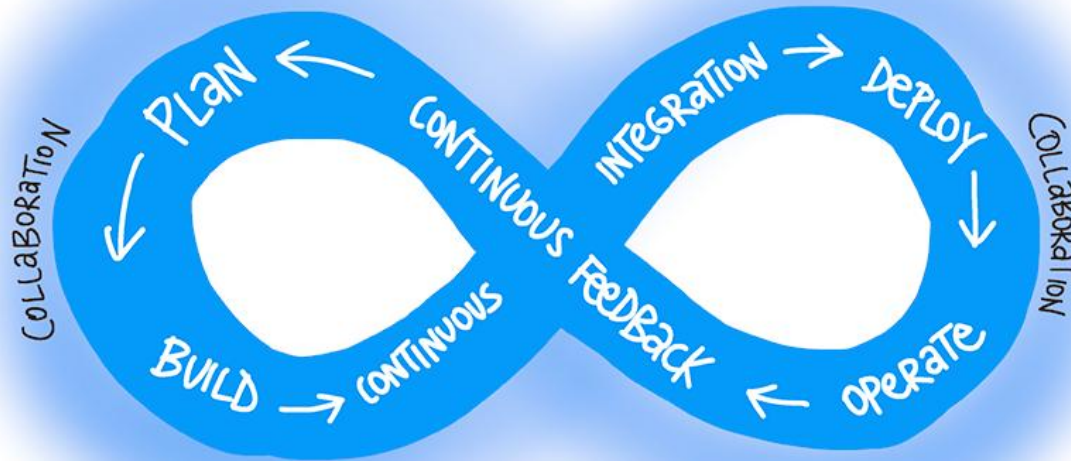
# Questions?



# CI with Azure DevOps and Git Integration

# What is DevOps?

- [What is DevOps? - Training | Microsoft Learn](#)



# DevOps benefits for Fabric



Enhance collaboration (CI)

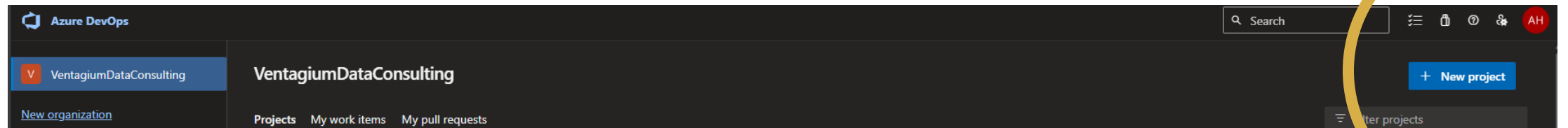


Leverage your development (CD)



Version control

# Create a DevOps Project



# Create a DevOps Project

Create new project

Project name \*

Fabric Workshop

Description

Visibility

Public

Anyone on the internet can view the project. Certain features like TFVC are not supported.

Private

Only people you give access to will be able to view this project.

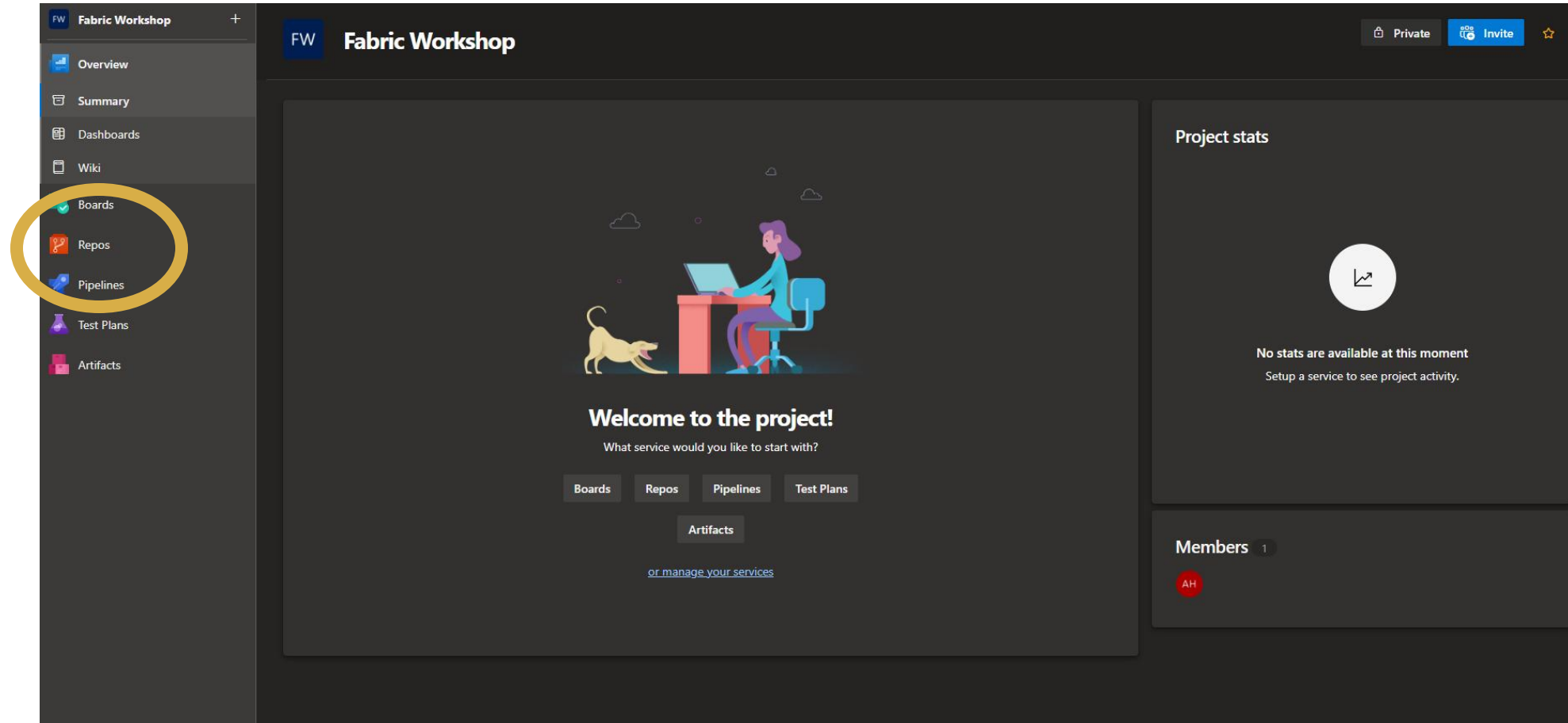
Advanced

Cancel

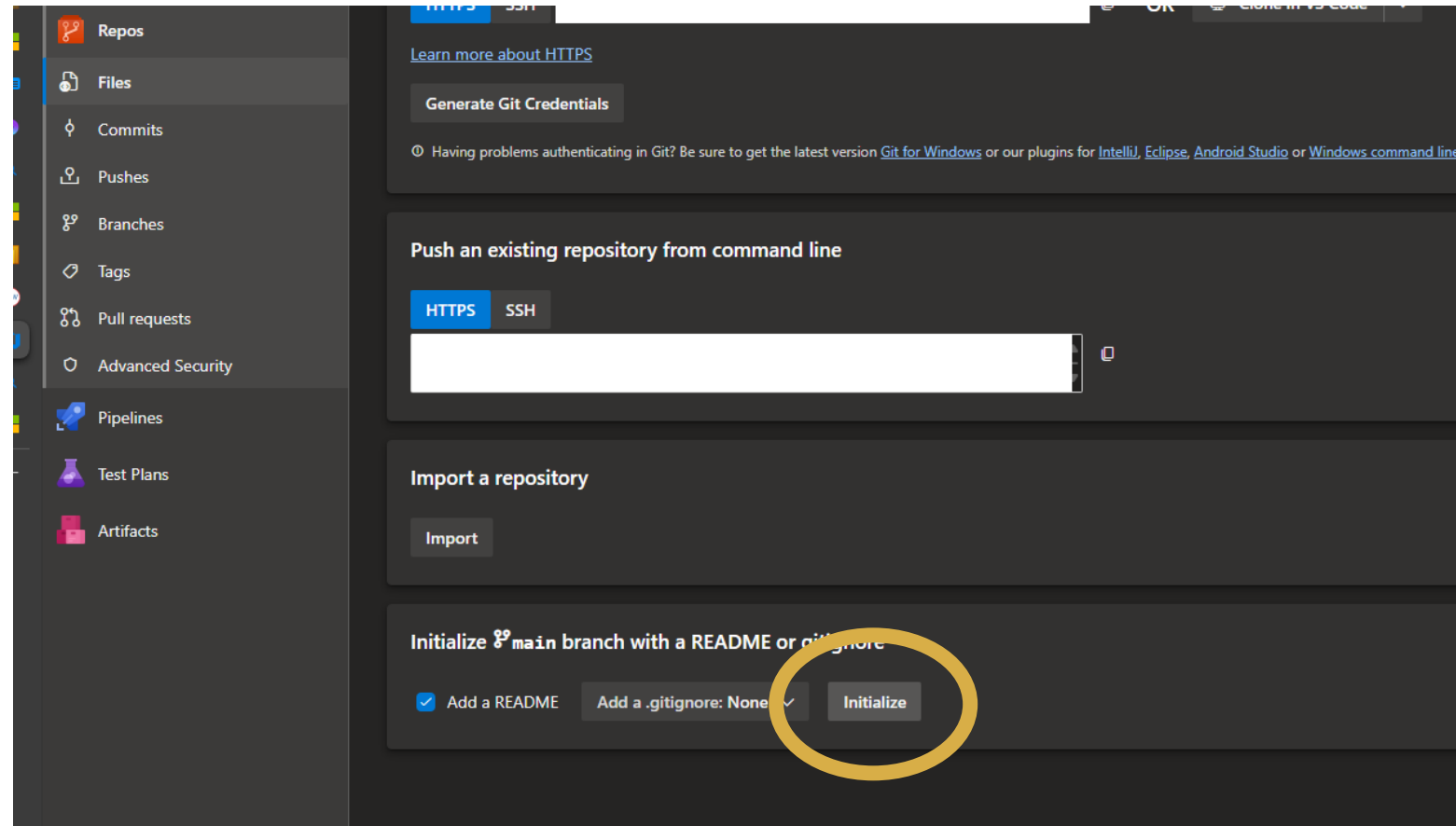
Create



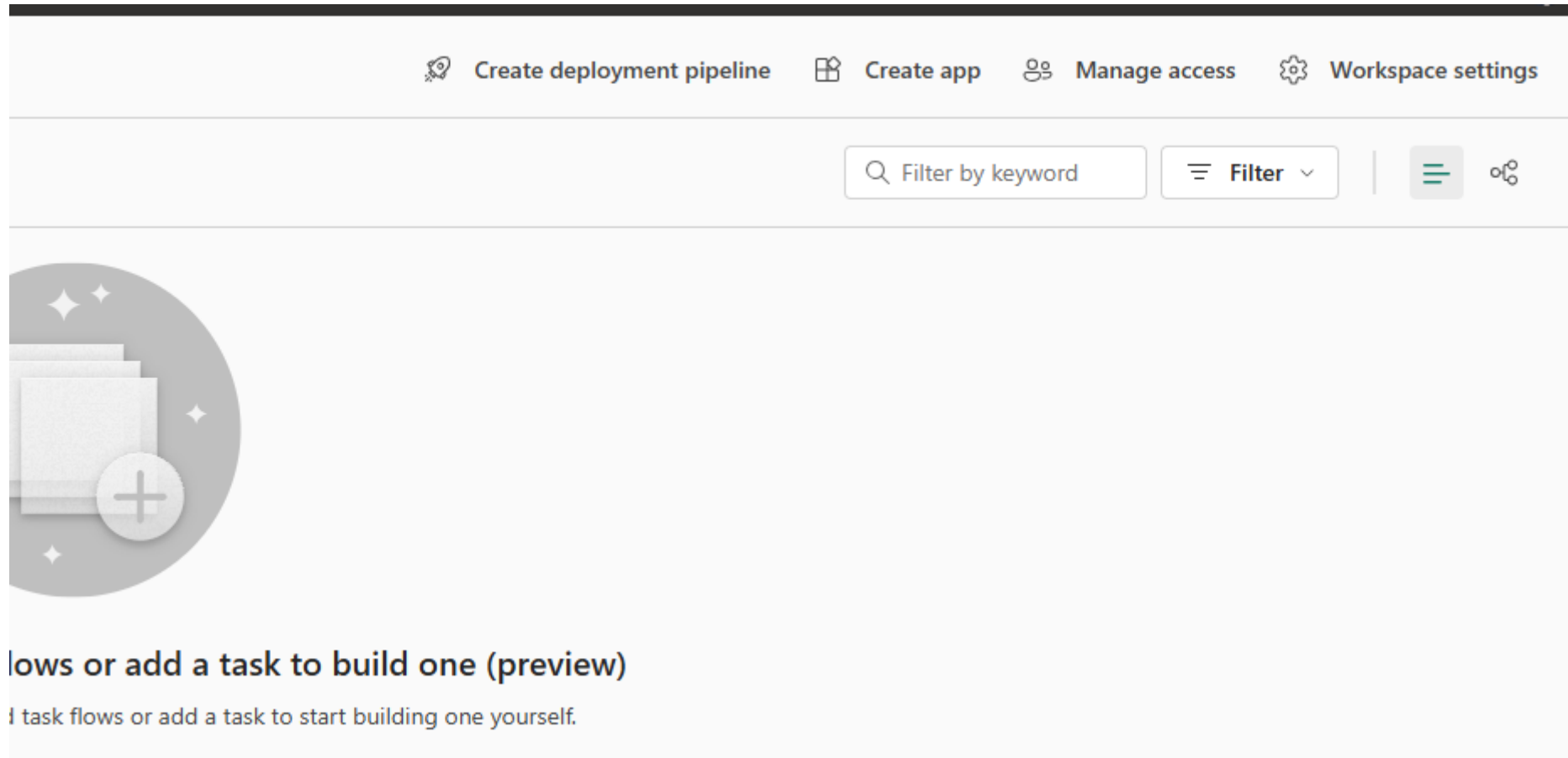
# Initialize a Repo



# Initialize a Repo



# Enable Git Integration



# Enable Git Integration

## Workspace settings

### General

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

Power BI

## General

About

### Workspace image



### Name \*

Fabric\_Workshop

### Description

Describe this workspace (O

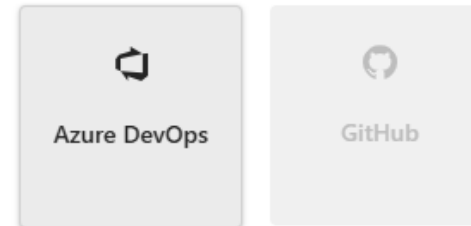
## Git integration

Connect to Git to manage your code and back up your work. [Learn more](#)

**Preview items** Some item types are only available in preview when using Git. [Learn more](#)

### Connect Git provider and account

#### Git provider
















#### Accounts

☒ AAD account  
alberto. com

Connect

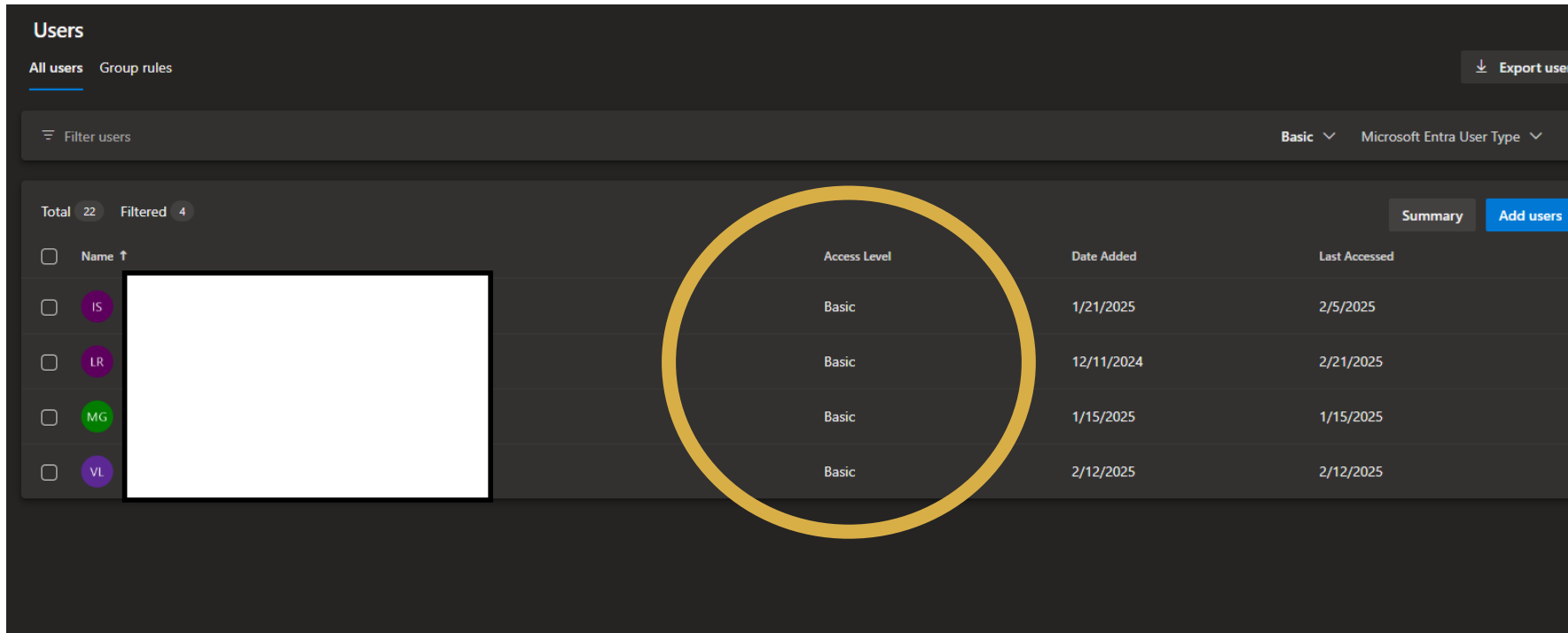
Connect Git repository and branch

# 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... | —    |
|   |                     |  Synced      | Data pipeline       | —    |

# Create a Contributors Team

(Users require Basic access at least)



The screenshot displays the 'Users' management interface. At the top, there are tabs for 'All users' and 'Group rules', with 'All users' selected. An 'Export users' button is located in the top right corner. Below the tabs is a 'Filter users' section with a search bar. On the right side of the filter section, there are dropdown menus for 'Basic' and 'Microsoft Entra User Type'. Below the filter section, there is a summary bar showing 'Total 22' and 'Filtered 4'. To the left of the user list is a sidebar with a 'Name' column and a list of user initials (IS, LR, MG, VL) each with a checkbox. The main table lists users with columns for 'Access Level', 'Date Added', and 'Last Accessed'. A yellow circle highlights the 'Access Level' column, which shows 'Basic' for all four users. At the bottom right of the table, there are 'Summary' and 'Add users' buttons.

| Name | Access Level | Date Added | Last Accessed |
|------|--------------|------------|---------------|
| IS   | Basic        | 1/21/2025  | 2/5/2025      |
| LR   | Basic        | 12/11/2024 | 2/21/2025     |
| MG   | Basic        | 1/15/2025  | 1/15/2025     |
| VL   | Basic        | 2/12/2025  | 2/12/2025     |

# Create Contributors Team

Project Settings

Fabric Workshop

General

Overview

Teams

Permissions

Notifications

Service hooks

Dashboards

Boards

Project configuration

Teams

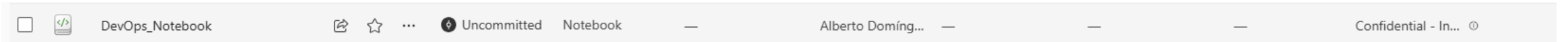
Total 2

New Team

| Name ↓                                     | Description                                  | Members |
|--|--|---------|
| <div>FC</div> Fabric Workshop Contributors |  | 6       |
| <div>FT</div> Fabric Workshop Team         | <div>Default</div> The default project team. | 1       |

# Commit and Sync your Items

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





# Commit and Sync your Items

[Create deployment pipeline](#) [Create app](#) [Manage access](#) [Workspace settings](#)

Source control 2

Filter by keyword

Filter

| Next refresh | Endorsement |
|--------------|-------------|
| —            | —           |
| N/A          | —           |
| —            | —           |
| —            | —           |
| N/A          | —           |
| —            | —           |
| —            | —           |

Source control

Current branch: main

Changes 2

Updates

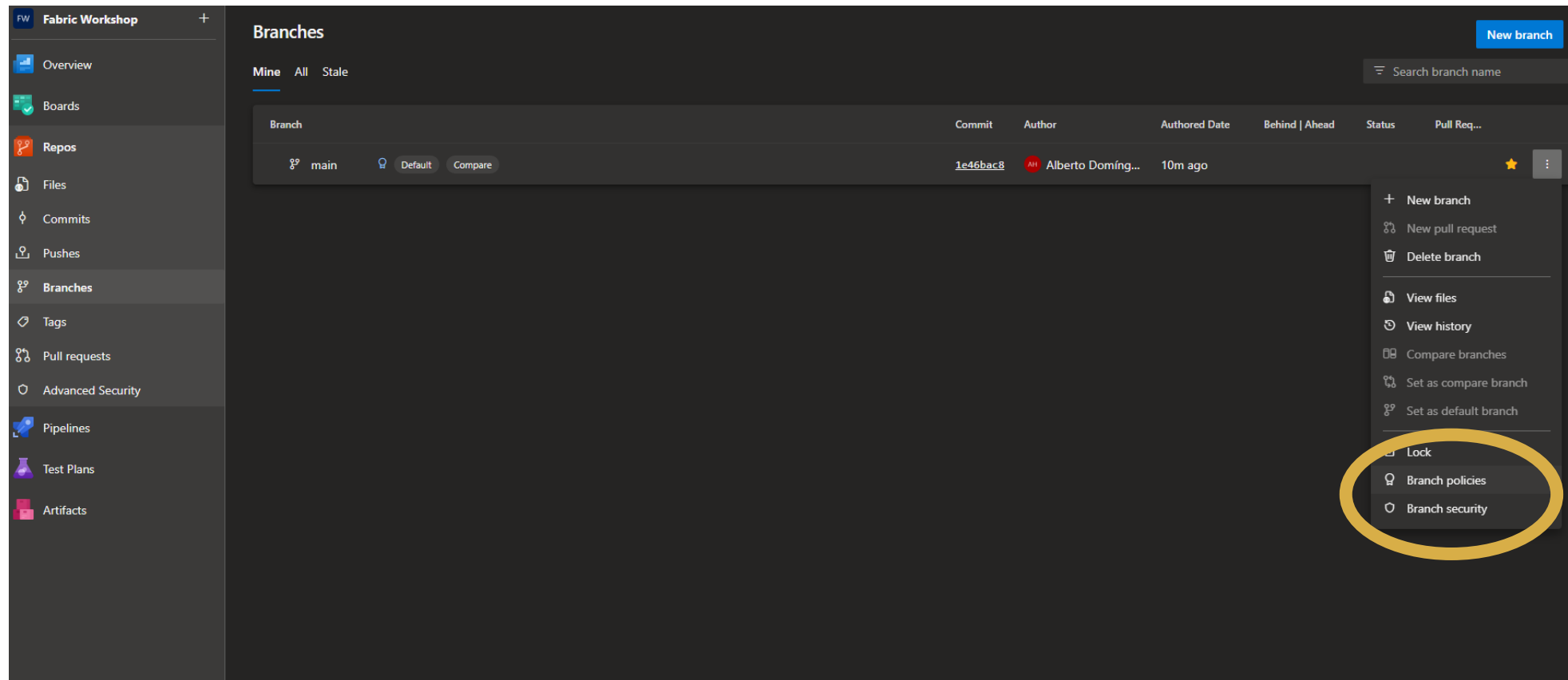
feat: idle change to track it on devOps

|                                     | Item                | Status |
|-------------------------------------|---------------------|--------|
| <input type="checkbox"/>            | OpenAI_API_Notebook | +      |
| <input checked="" type="checkbox"/> | DevOps_Notebook     | +      |

Commit

Undo

# Create Branch policies



# Branch Policies

Fabric\_Workshop

Create deployment pipeline>Create app

+ New item

New folder

→ Import

Source control 2

Enter by key

|  | Name               | Git status    | Type              | Task | Owner             | Refreshed             | Next refresh | Endorsement | Se |
|--|--------------------|---------------|-------------------|------|-------------------|-----------------------|--------------|-------------|----|
|  | OpenAI Demo        |               | Folder            | —    | —                 | —                     | —            | —           | —  |
|  | Dataflow 1         | ⚠ Unsupported | Dataflow Gen2     | —    | Roberto Hernán... | —                     | N/A          | —           | —  |
|  | DevOps_Notebook    | 📍 Uncommitted | Notebook          | —    | Alberto Domíng... | —                     | —            | —           | Co |
|  | Fabric Workshop 01 | ✅ Synced      | SQL database      | —    | Roberto Hernán... | —                     | —            | —           | Co |
|  | Fabric Workshop 01 | —             | Semantic model... | —    | Fabric_Workshop   | 2/18/2025, 10:45:1... | N/A          | —           | Co |
|  | Fabric Workshop 01 | —             | SQL application   | —    | Roberto Hernán... | —                     | —            | —           | Co |

Notifications

Clear all

Earlier

✖ Unable to complete commit request

View details below to learn more.

View details

6 minutes ago

✅ Your selected changes were committed.

14 minutes ago

ℹ Session stopped successfully.

19 minutes ago

# Branch Policies

The screenshot displays a GitHub Pull Request (PR) titled "PR Example - DevOps Tutorial". The PR is in the "Active" state, with the number "11469" and the author "Alberto Domínguez Horner" (AH). The PR description is "Alberto Domínguez Horner proposes to merge development into main".

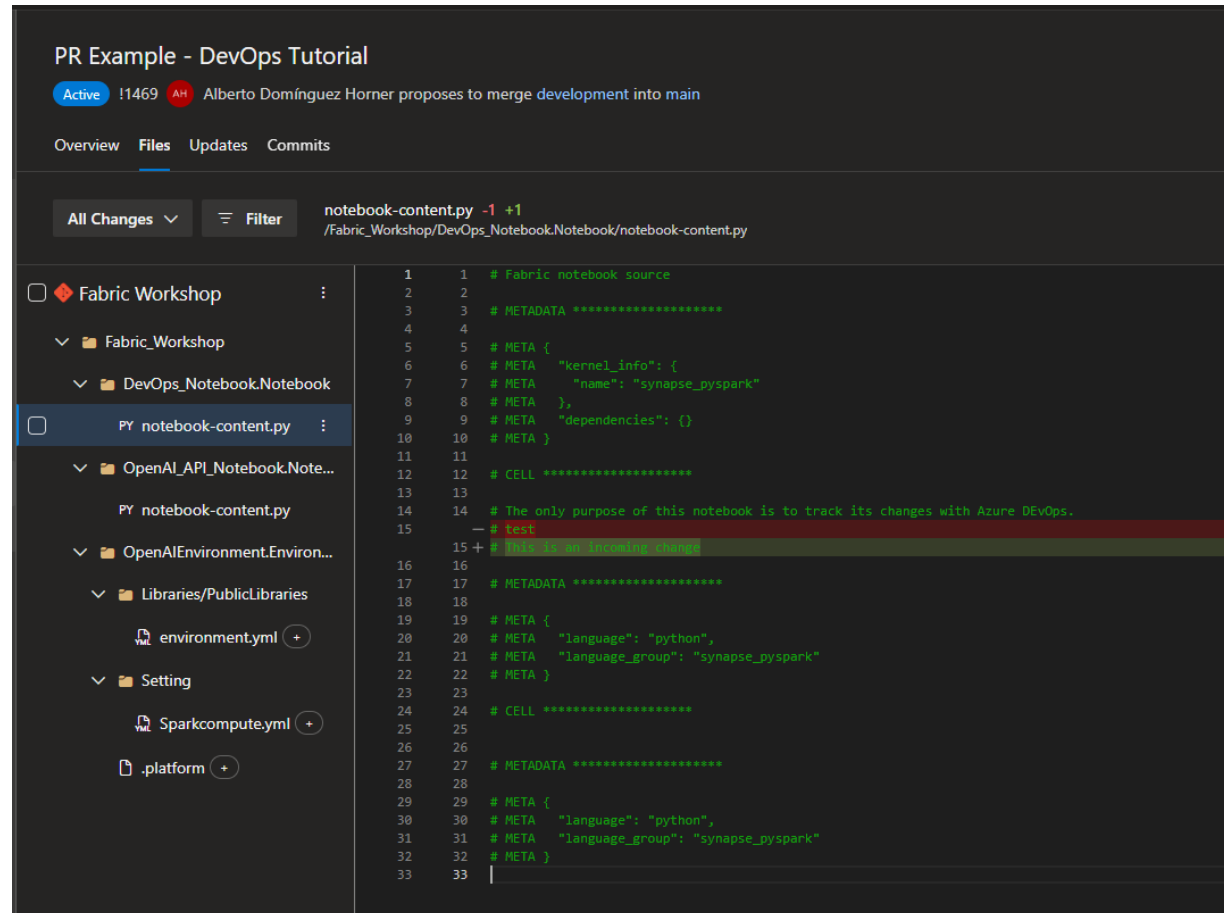
The "Overview" tab is selected, showing the following information:

- Reviewers:** Alberto Domínguez Horner must approve.
- Checks:** No merge conflicts (Last checked: Just now).
- Description:** A text area for the PR description.
- Comments:** A section for adding comments, with a "Show everything (3)" dropdown.
- Commit History:** A list of commits, including "Alberto Domínguez Horner pushed 1 commit" and "test: incoming changes".
- Branch Policies:** A section showing the branch policy for the "development" branch, which requires approval from Alberto Domínguez Horner.

The "Reviewers" section on the right shows the required reviewers: Alberto Domínguez Horner (No review yet). The "Optional" section shows "No optional reviewers". The "Tags" section shows "No tags". The "Work items" section shows "No work items".

A yellow oval highlights the "Branch Policies" section, which shows the policy for the "development" branch. The policy is titled "Branch policy for development" and lists the required reviewers: Alberto Domínguez Horner. The policy is currently in the "Active" state.

# Continuous Integration by Pull Requests



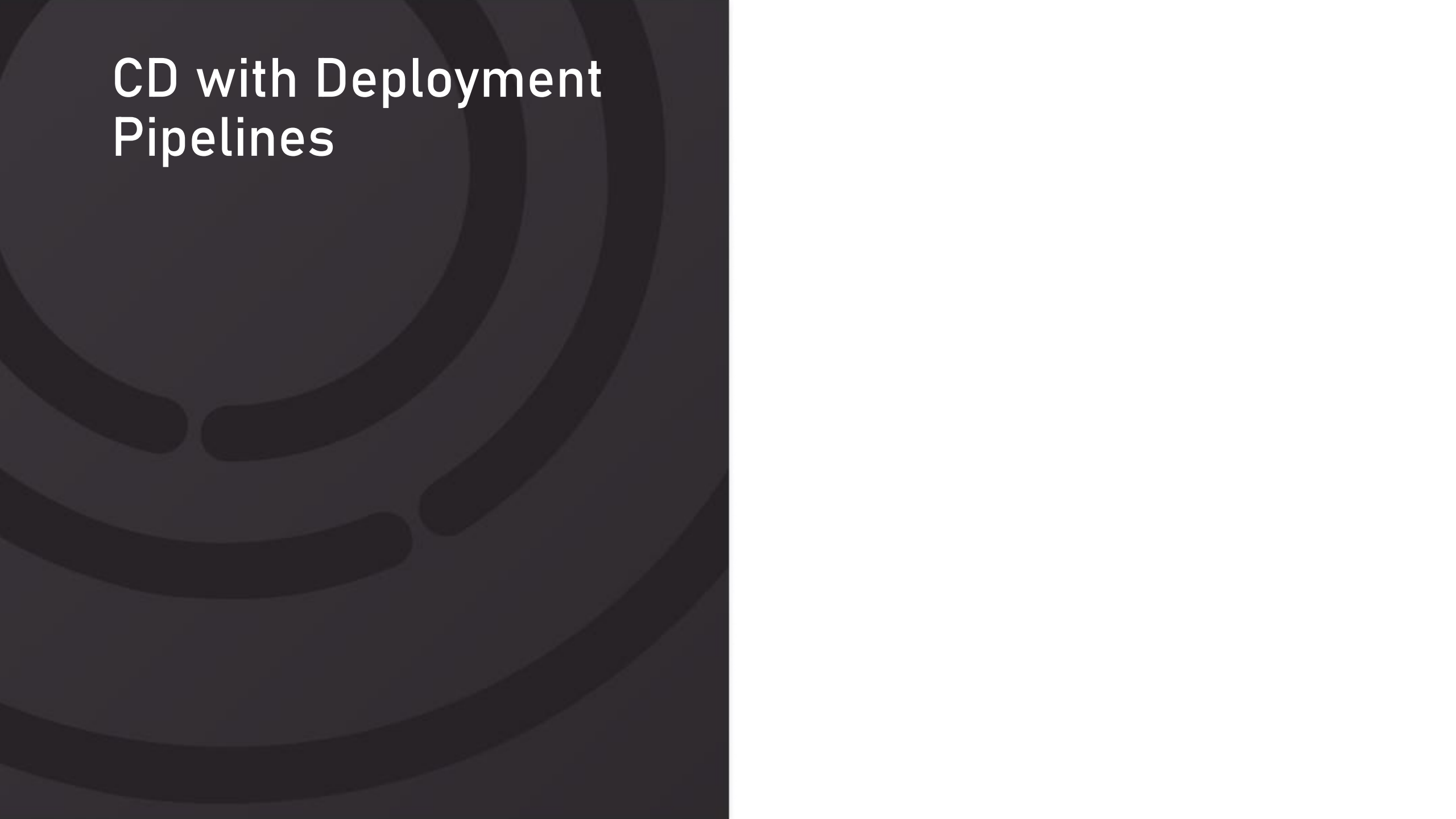
The screenshot displays a GitHub Pull Request (PR) titled "PR Example - DevOps Tutorial". The PR is active and proposed by Alberto Domínguez Horner, aiming to merge the 'development' branch into 'main'. The interface shows the 'Files' tab, highlighting a change in the file 'notebook-content.py' within the 'Fabric\_Workshop/DevOps\_Notebook.Notebook' directory. The diff shows a single line added (line 15):

```
15 + # This is an interesting change
```

The file's content is displayed in a dark-themed editor. It is a Jupyter Notebook file with the following structure:

- Line 1: # Fabric notebook source
- Line 2: (empty)
- Line 3: # METADATA \*\*\*\*\*
- Line 4: (empty)
- Line 5: # META {
- Line 6: # META "kernel\_info": {
- Line 7: # META "name": "synapse\_pyspark"
- Line 8: # META },
- Line 9: # META "dependencies": { }
- Line 10: # META }
- Line 11: (empty)
- Line 12: # CELL \*\*\*\*\*
- Line 13: (empty)
- Line 14: # The only purpose of this notebook is to track its changes with Azure DEVOps.
- Line 15: # This is an interesting change (added)
- Line 16: (empty)
- Line 17: # METADATA \*\*\*\*\*
- Line 18: (empty)
- Line 19: # META {
- Line 20: # META "language": "python",
- Line 21: # META "language\_group": "synapse\_pyspark"
- Line 22: # META }
- Line 23: (empty)
- Line 24: # CELL \*\*\*\*\*
- Line 25: (empty)
- Line 26: (empty)
- Line 27: # METADATA \*\*\*\*\*
- Line 28: (empty)
- Line 29: # META {
- Line 30: # META "language": "python",
- Line 31: # META "language\_group": "synapse\_pyspark"
- Line 32: # META }
- Line 33: (empty)

# CD with Deployment Pipelines

The background of the slide is split vertically. The left half is dark gray and features several concentric, thick-lined circles of varying shades, creating a tunnel-like or ripple effect. The right half is a solid, bright white.

# AI Skills

The background of the slide is split vertically. The left half is dark grey with several concentric, slightly broken circles in a lighter shade of grey. The right half is a solid, light grey.

# 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

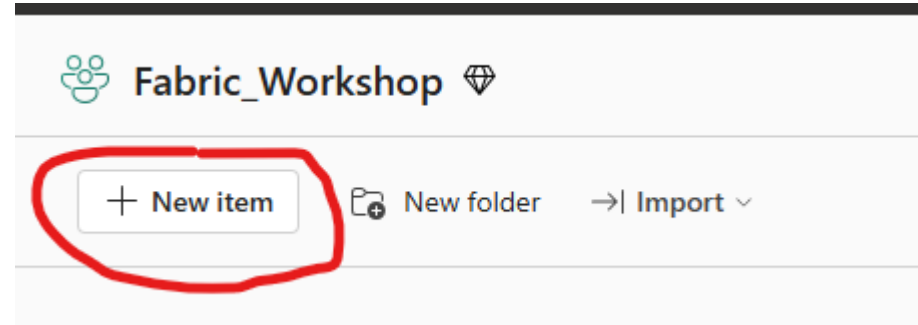


# Prerequisites.

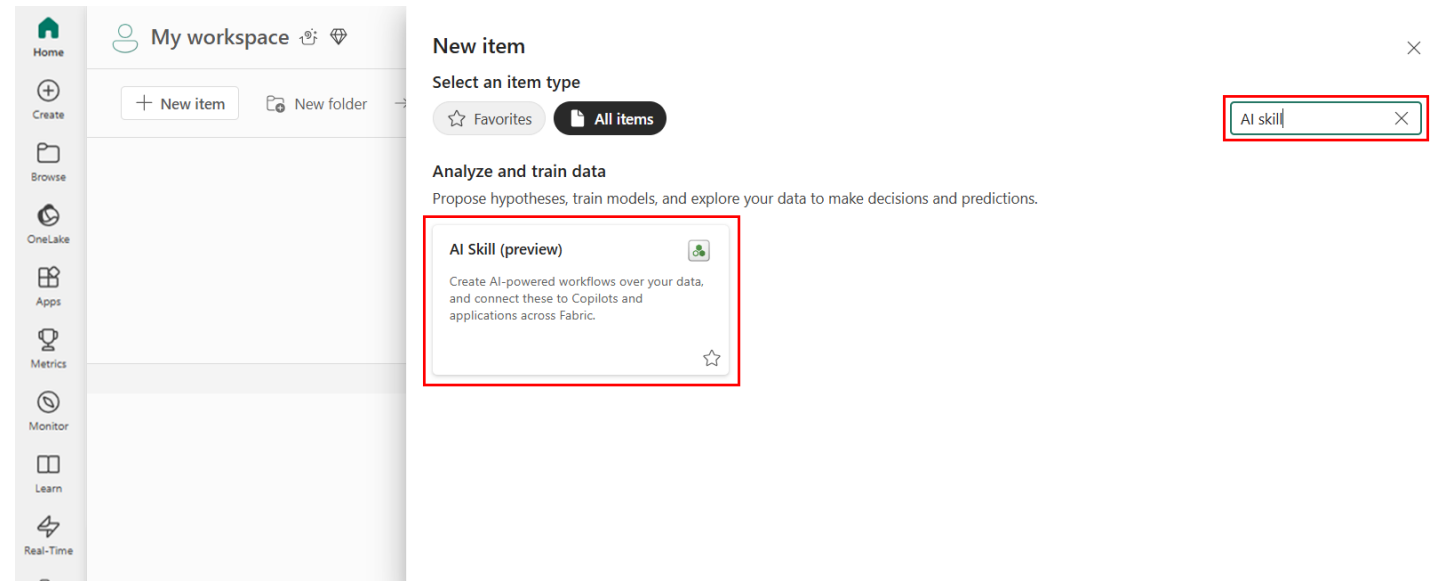
- A paid F64 or higher Fabric capacity resource.
- AI skill tenant switch is enabled.
- Copilot tenant switch is enabled.
- Cross-geo sharing for AI is enabled, if relevant.
- A warehouse or lakehouse with data.

# 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 performance 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.