From Manual to Automated: Master Metadata-Driven Design in Fabric



Erwin de Kreuk

Technology Lead Data, InSpark, Netherlands

Let’s connect

@erwindekreuk.bsky.social

linkedin.com/in/erwindekreuk

erwindekreuk.com

github.com/edkreuk

https://sessionize.com/erwin-de-kreuk/

Dutchfabricusergroup.com

Microsoft® Most Valuable Professional



THANK YOU SPONSORS!

Platinum



Gold









Bronze

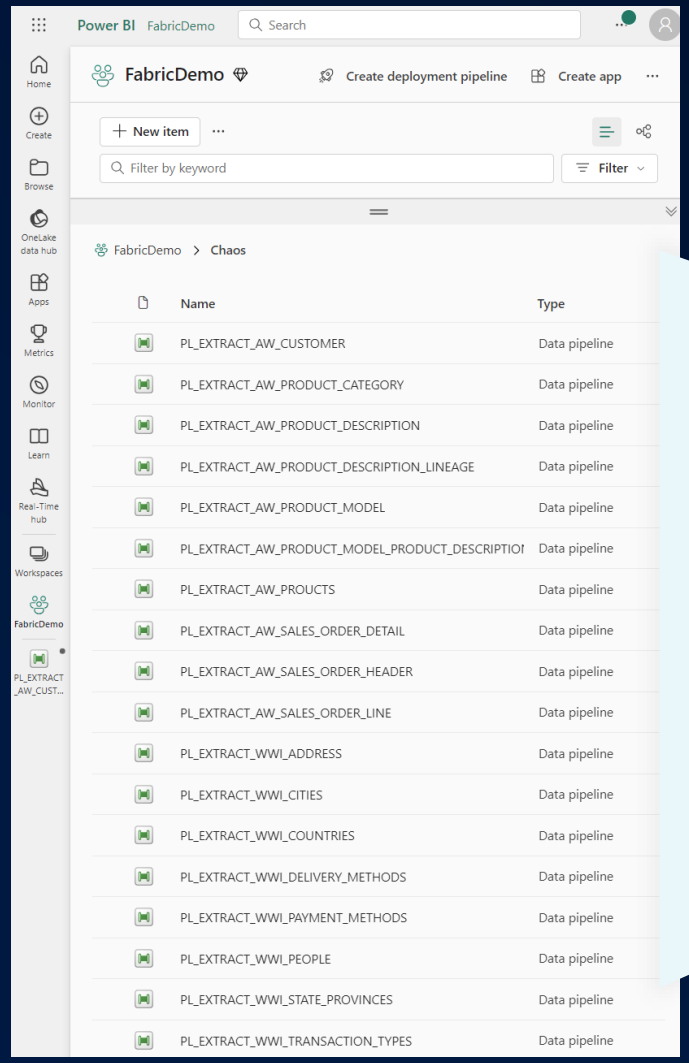




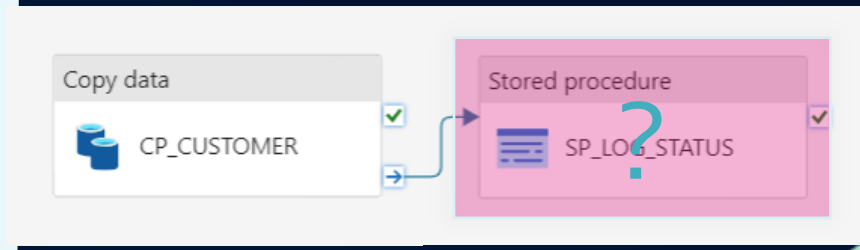


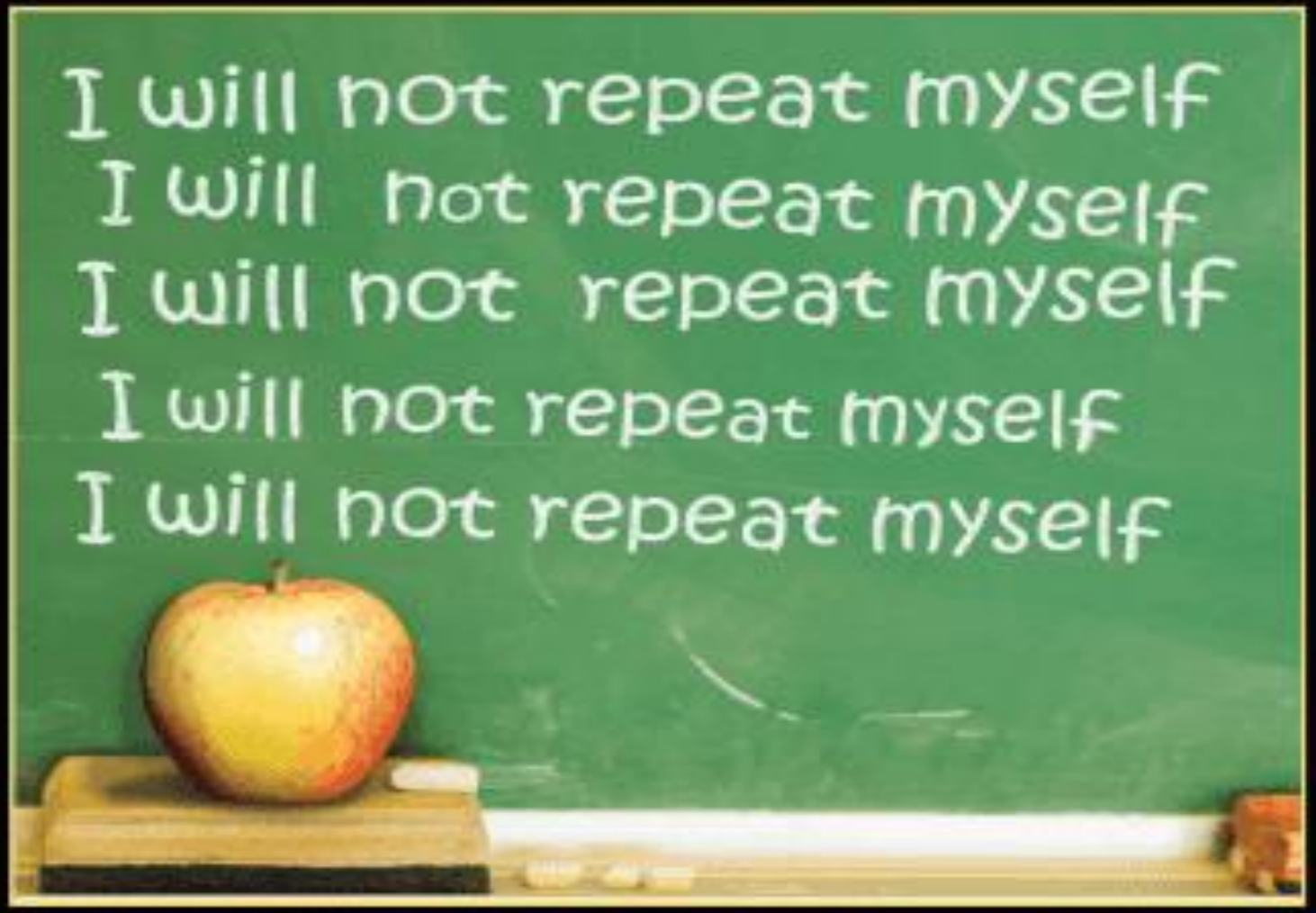


Problem statement



Content of each pipeline





DON'T REPEAT YOURSELF

Repetition is the root of all software evil.

Agenda



- Metadata Driven Framework
- Data Platform Engineering
- Data Integration
- Data Processing
- Data Monitoring
- Q & A

Out-of-the-Box Framework

- Ready-to-use.
- Rapid implementation.
- Limited customization.
- Lower development effort.
- Lower upfront costs.
- Ongoing support and updates.



Custom-Made Framework

- Tailored to specific needs.
- Full control over design and features.
- Higher development effort.
- Flexibility and extensibility.
- Higher upfront costs.



Custom-Made Framework

- Based on parameters
- Metadata => Fabric SQL Database / Json / ……
- Microsoft Fabric but also on Azure Synapse Analytics and Azure Data Factory
- Based on a Uniform Data Architecture



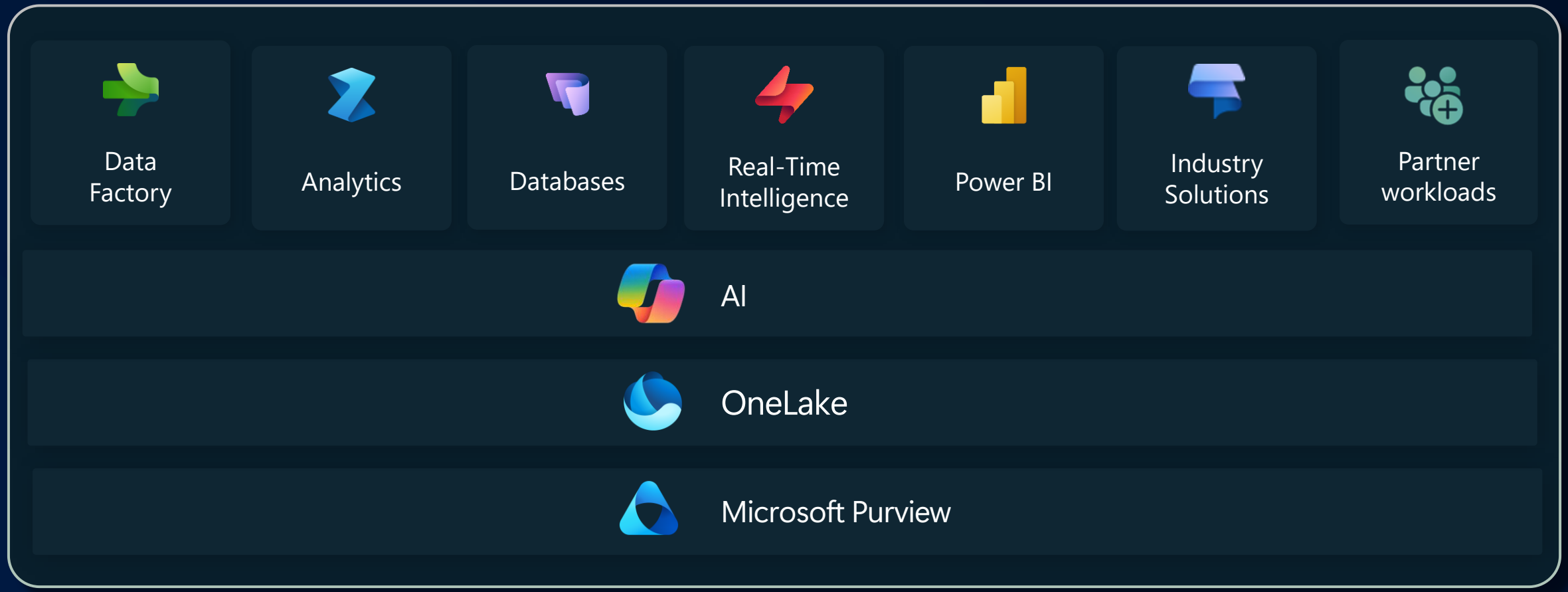
What about the Third option

- Fabric Native Solution

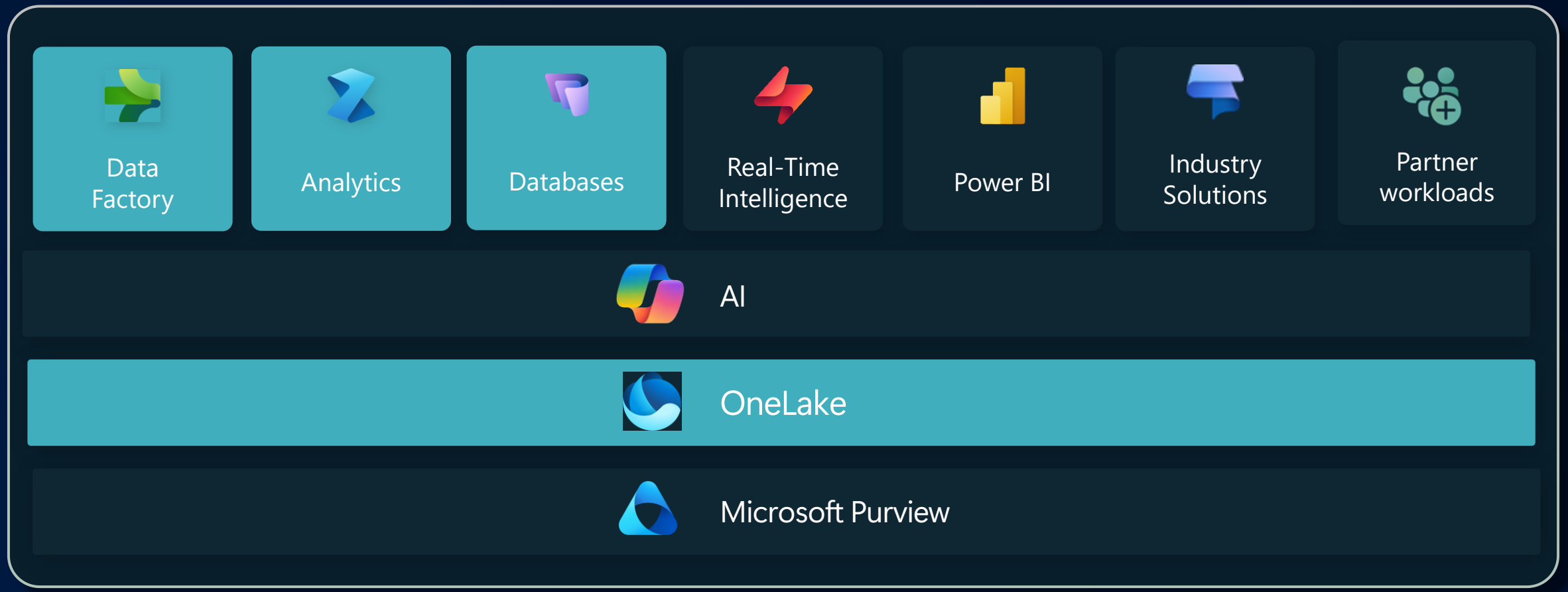


Microsoft Fabric

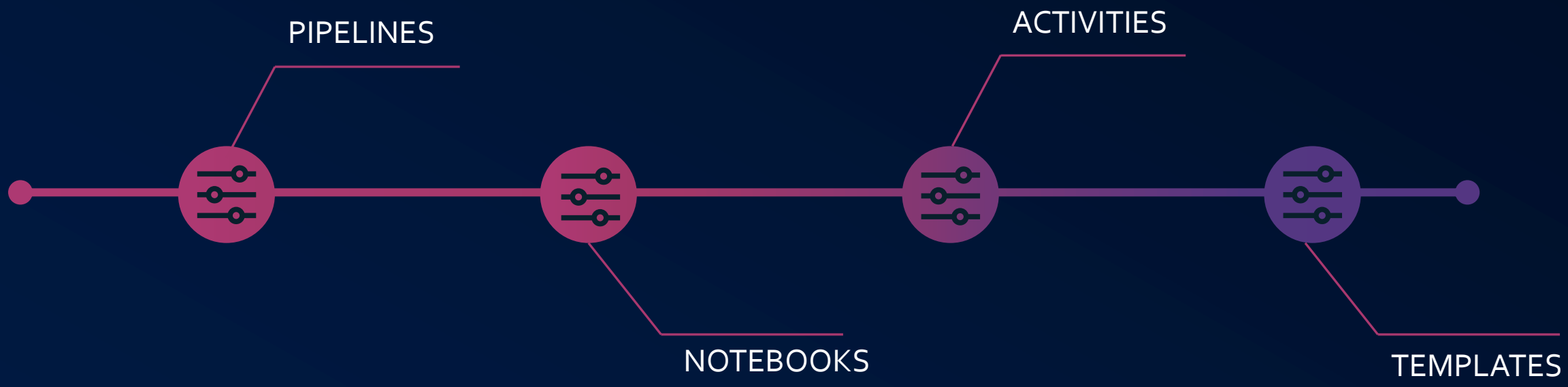
The unified data platform for AI transformation



The unified data platform for AI transformation



Parameters



Fabric Metadata Driven Framework

- Meets most of your needs
- Full control over design and features.
- Less development effort.
- Flexibility and extensibility.
- Lower upfront costs.
- 100% Fabric Native Solution



FABRIC METADATA-DRIVEN FRAMEWORK

FMD FRAMEWORK

Fabric Metadata Driven Framework

Modern data platforms demand agility, scalability, and consistency. FMD simplifies these challenges by enabling:

- **Dynamic, metadata-driven pipelines**
- **Consistent orchestration across ingestion, processing, and publishing**
- **Centralized configuration** for all data entities
- **Alignment with Microsoft Fabric Lakehouse & Medallion Architecture**
- **Reduced engineering effort** through reusable patterns
- **Faster delivery** with standardized, tested components

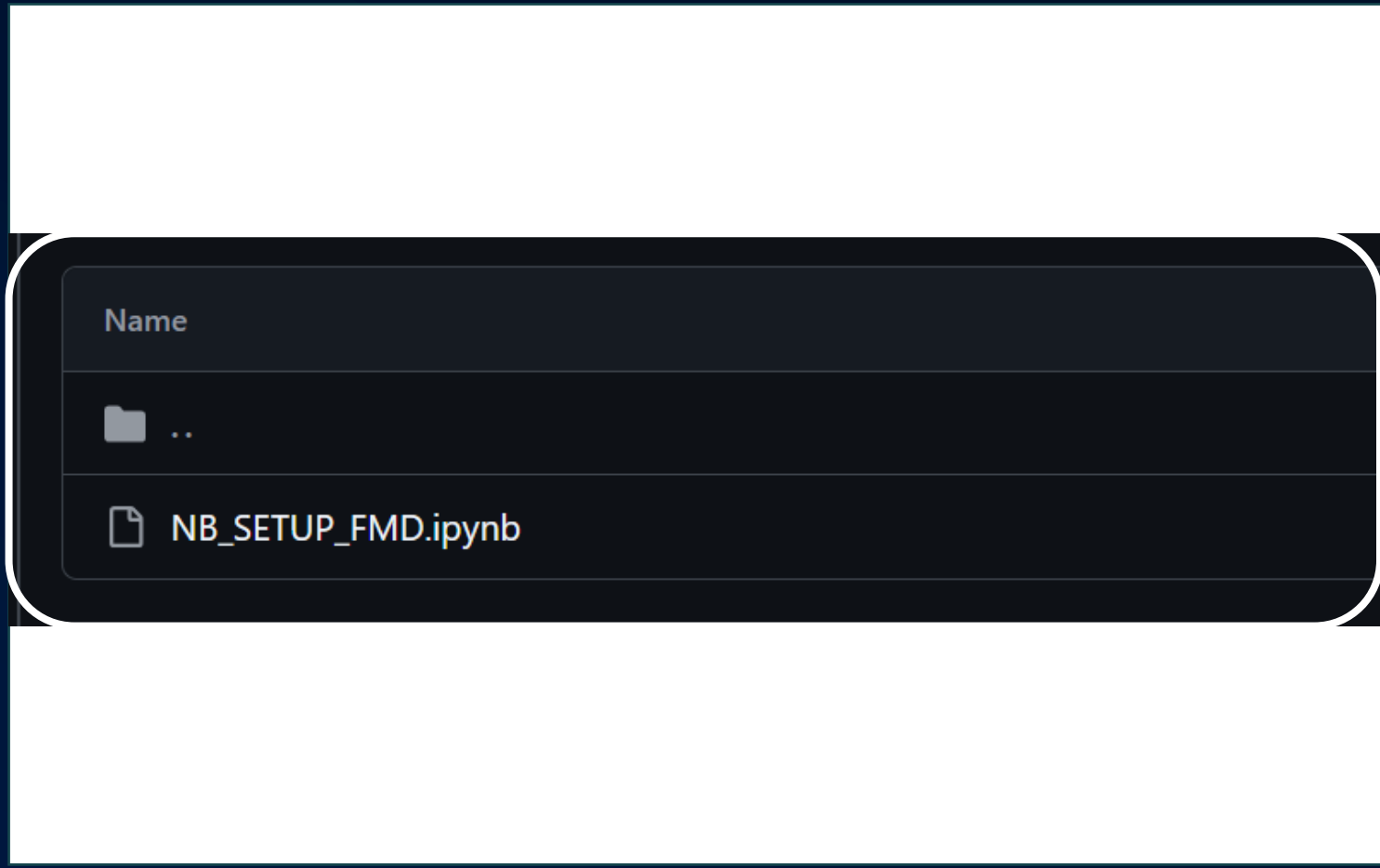


FABRIC METADATA-DRIVEN FRAMEWORK

FMD FRAMEWORK

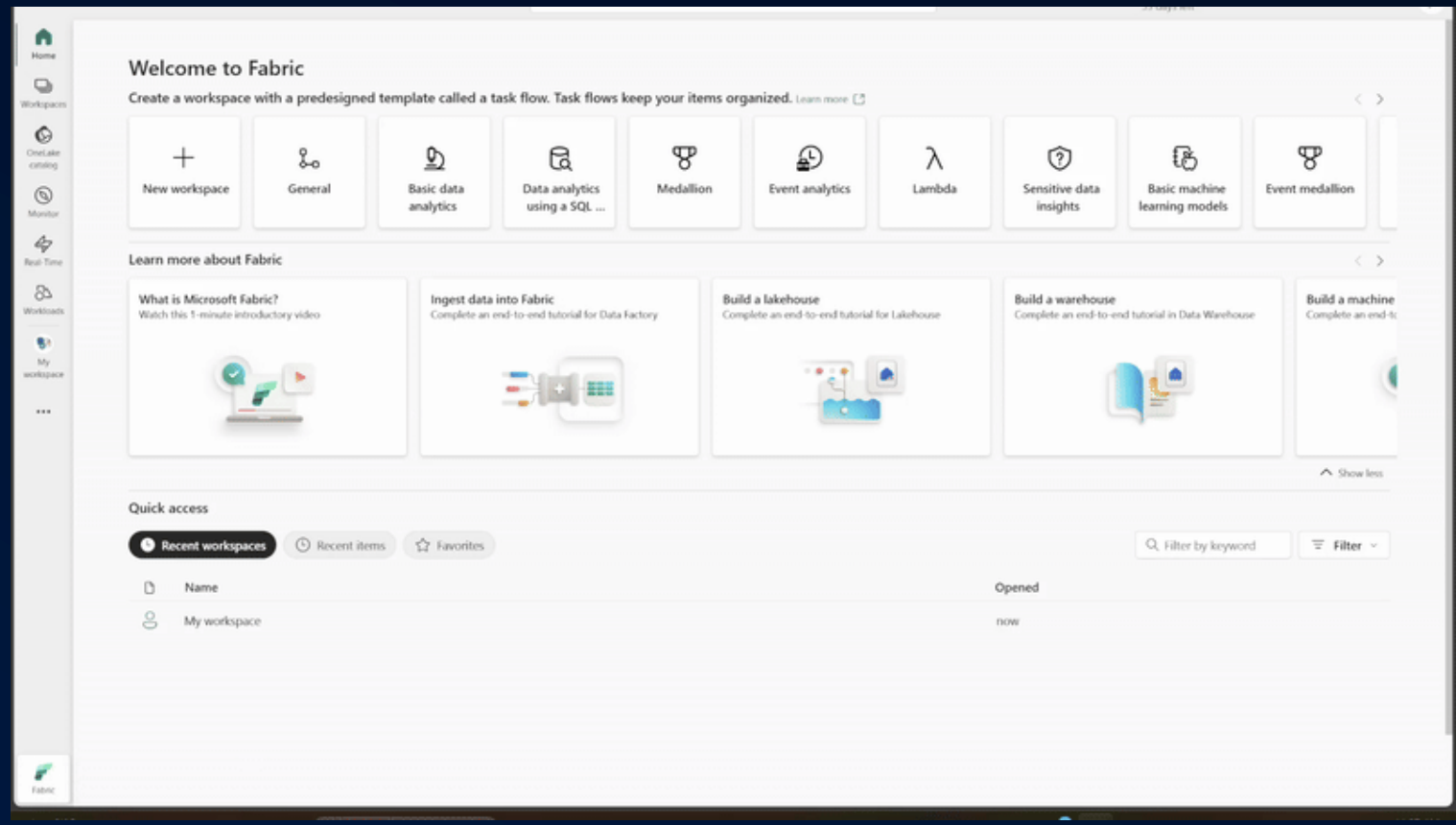
Fabric Metadata Driven Framework

- Download Notebook from Github



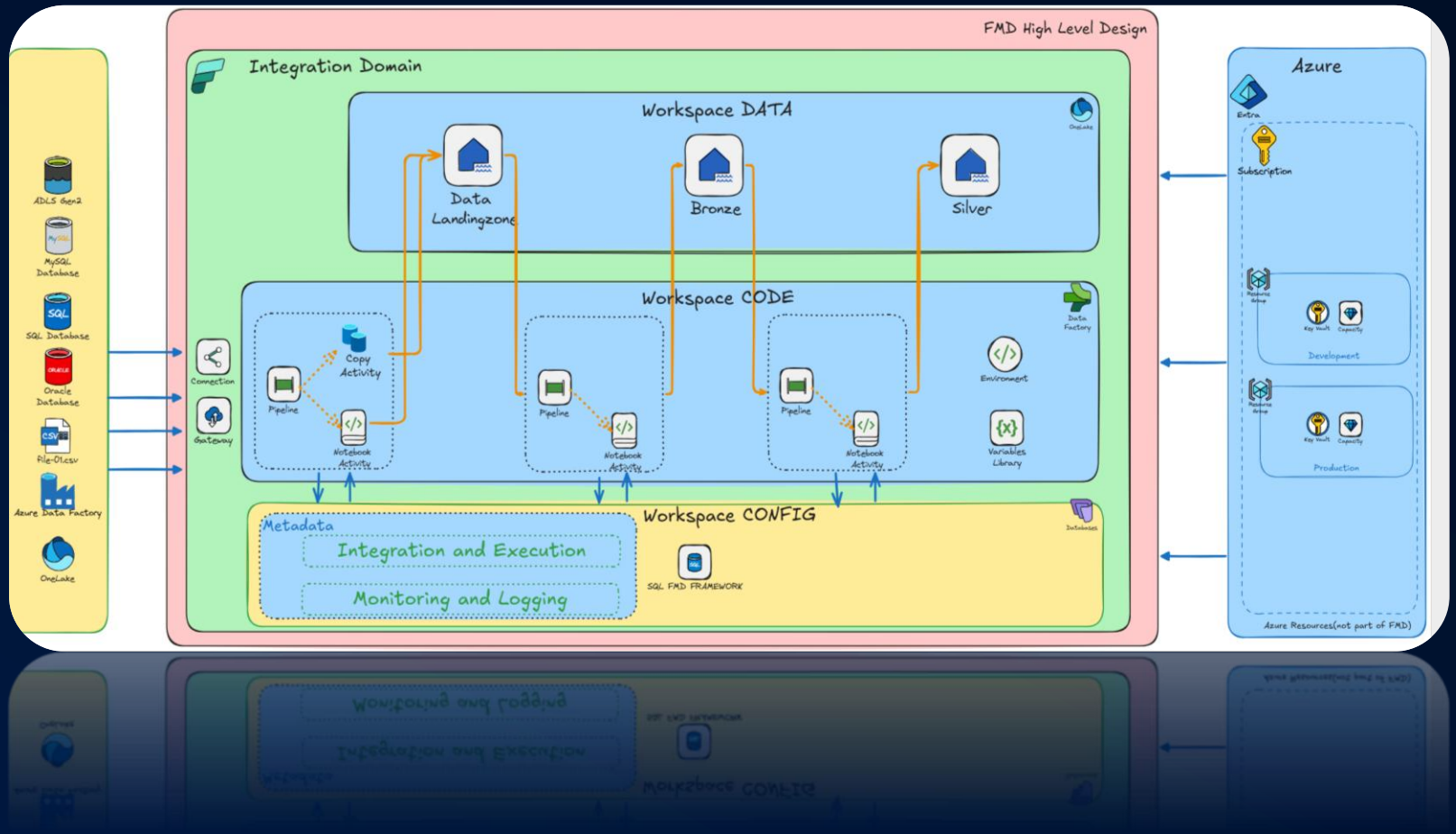
Fabric Metadata Driven Framework

- Download Notebook from Github
- Create Workspace Name is up to you:
  - **FMD_XXX_CONFIGURATION**
- Assign a capacity
- Import Notebook to this Workspace



Fabric Metadata Driven Framework

- Open Notebook
- NB_SETUP



Data Platform Engineeri

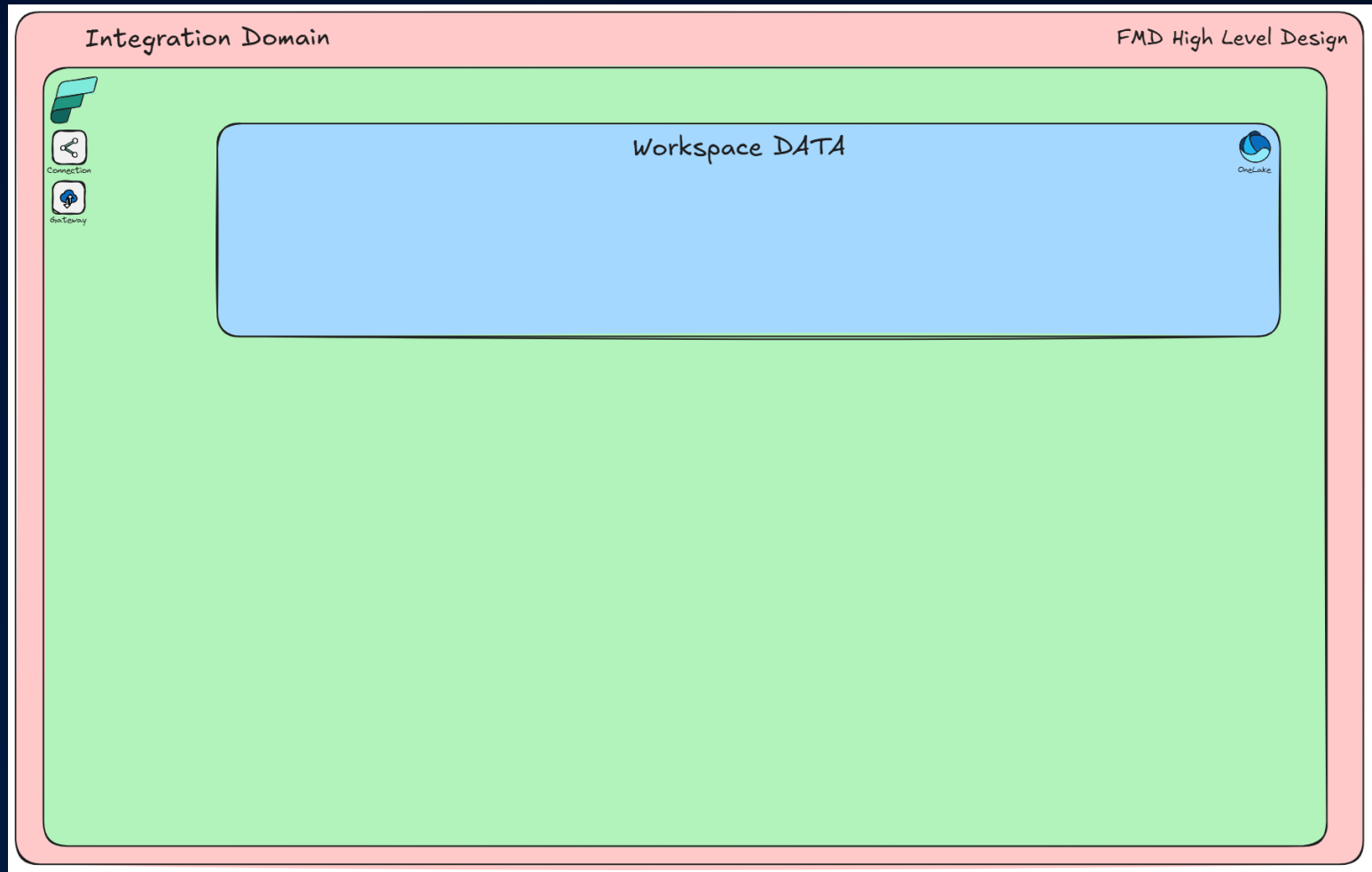
Data Platform Engineering

- Azure subscription
- Fabric Capacity (Trial)
- Key Vault optional
- Fabric Administrator role



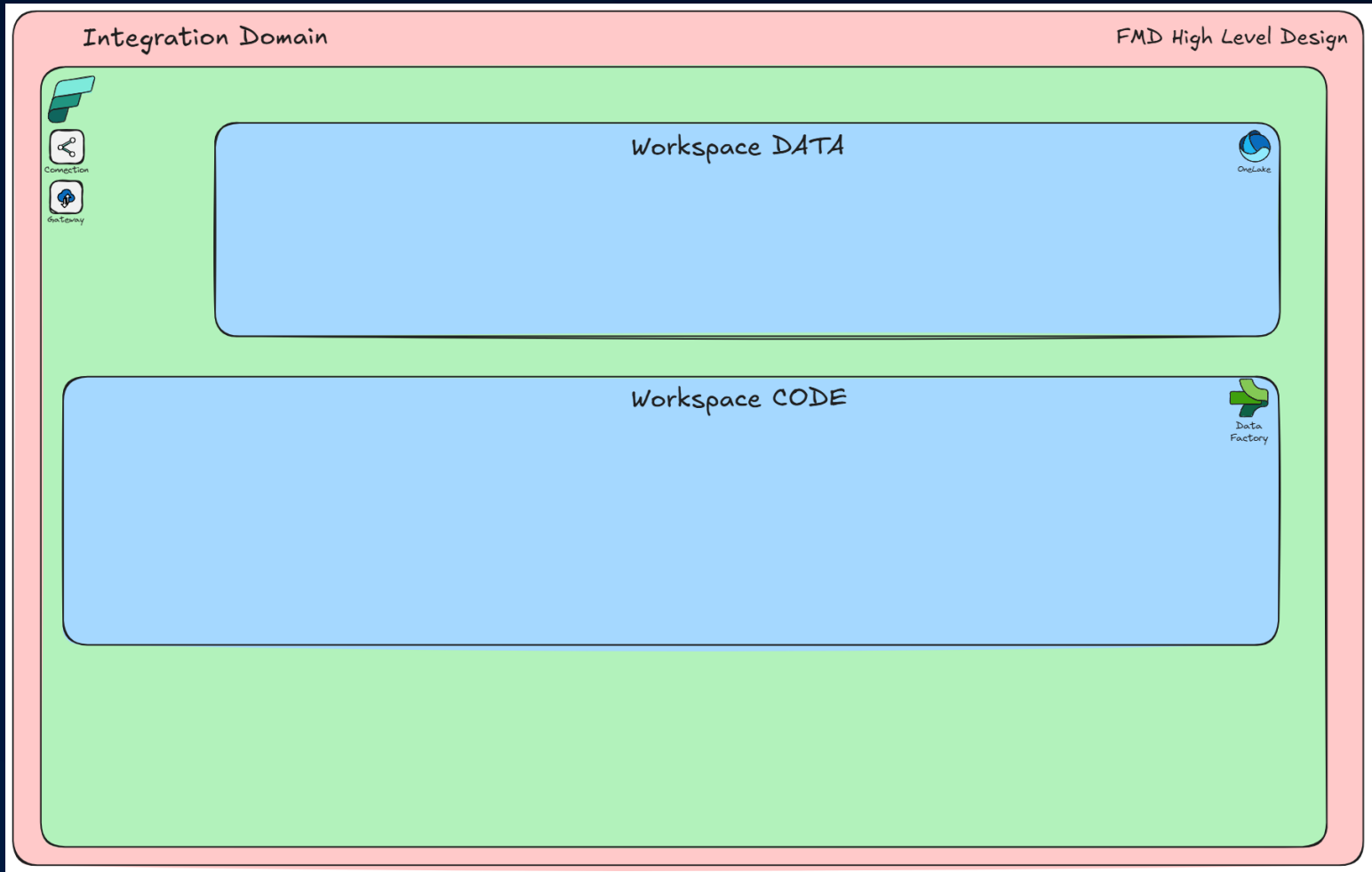
Data Platform Engineering

- Workspace DATA:
- Lakehouses



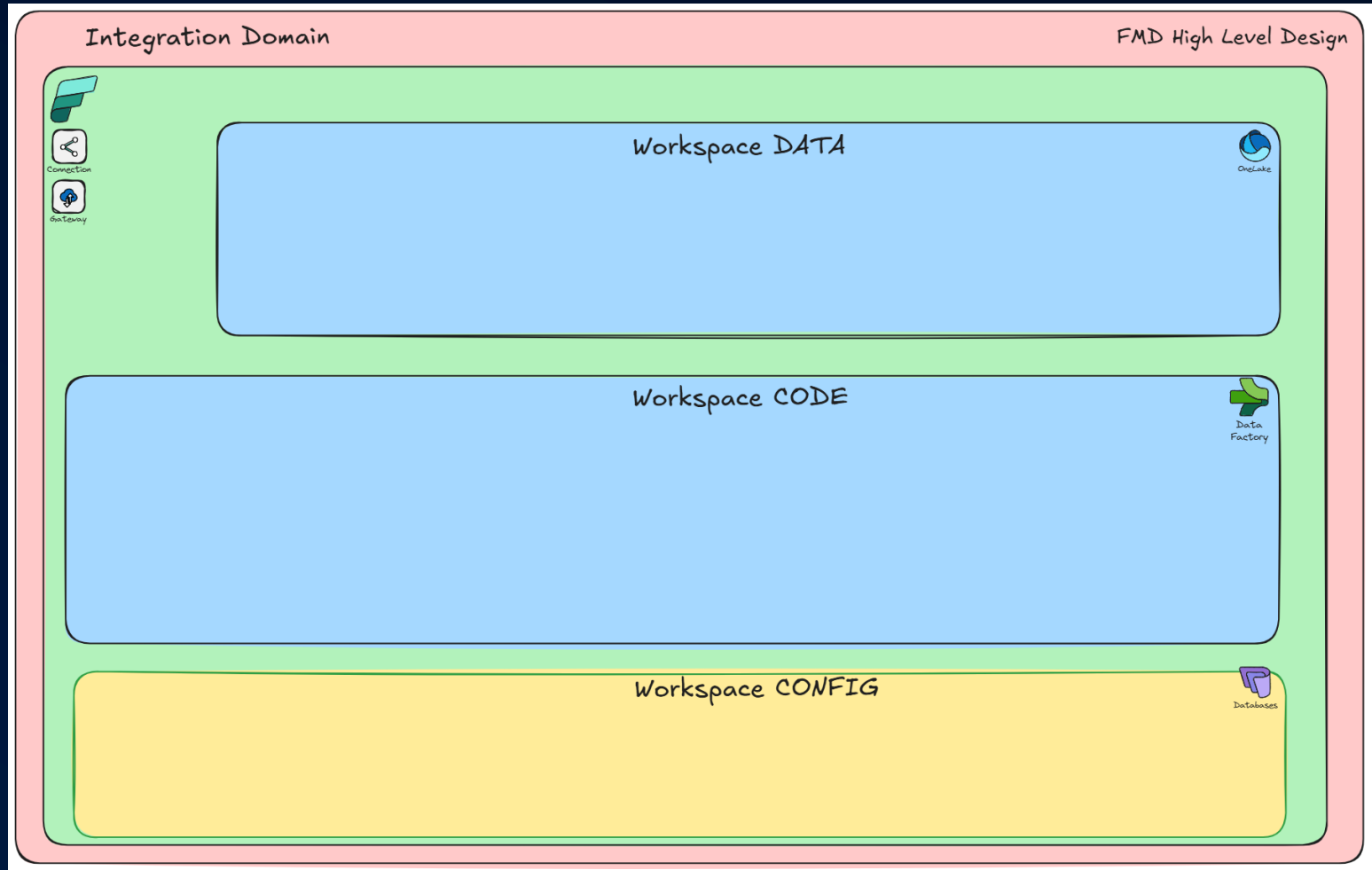
Data Platform Engineering

- Workspace CODE
- Data Pipelines
- Notebooks
- Environments
- Variable Libraries



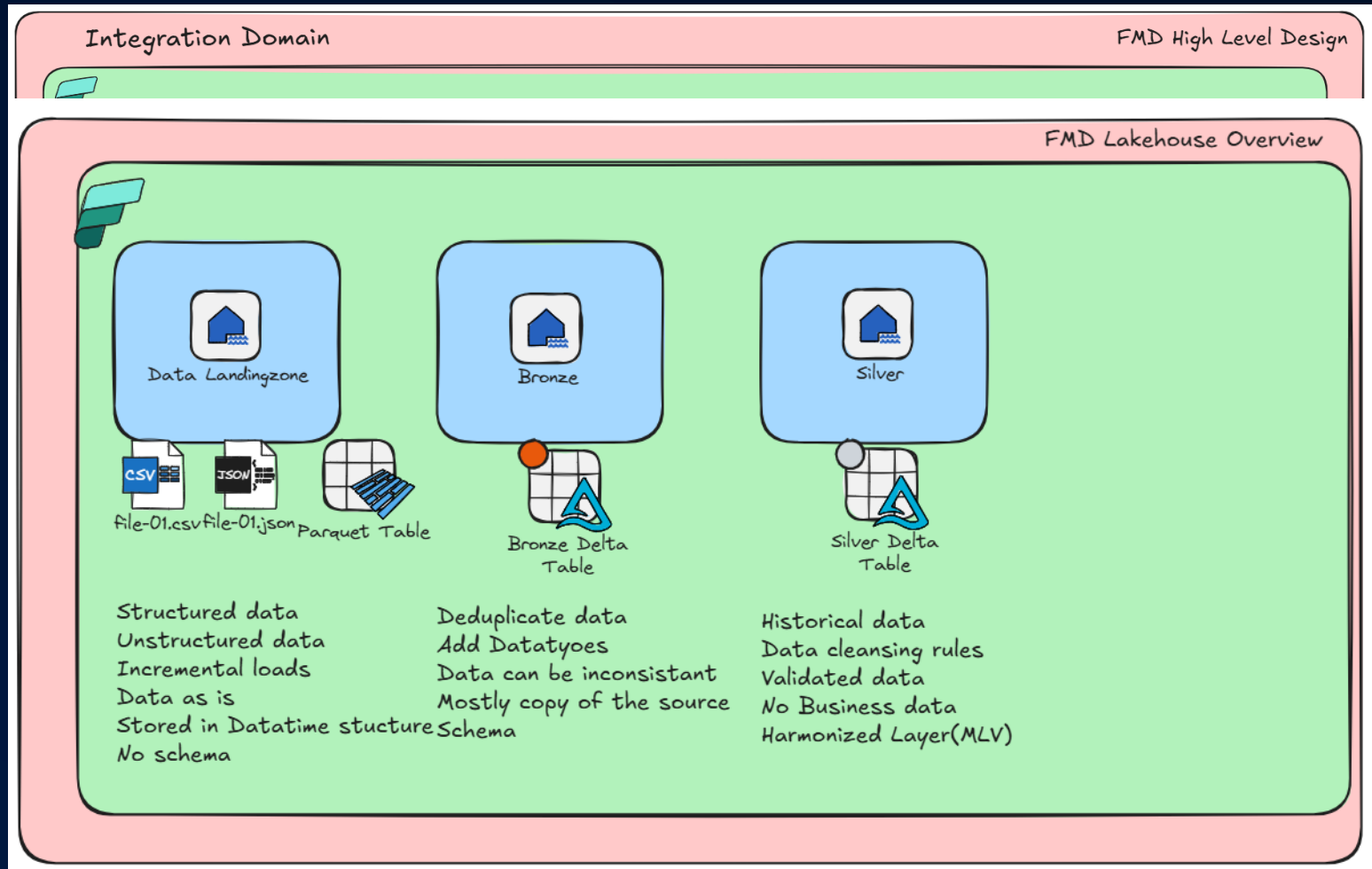
Data Platform Engineering

- Workspace CONFIG
- Fabric database
  - Metadata database



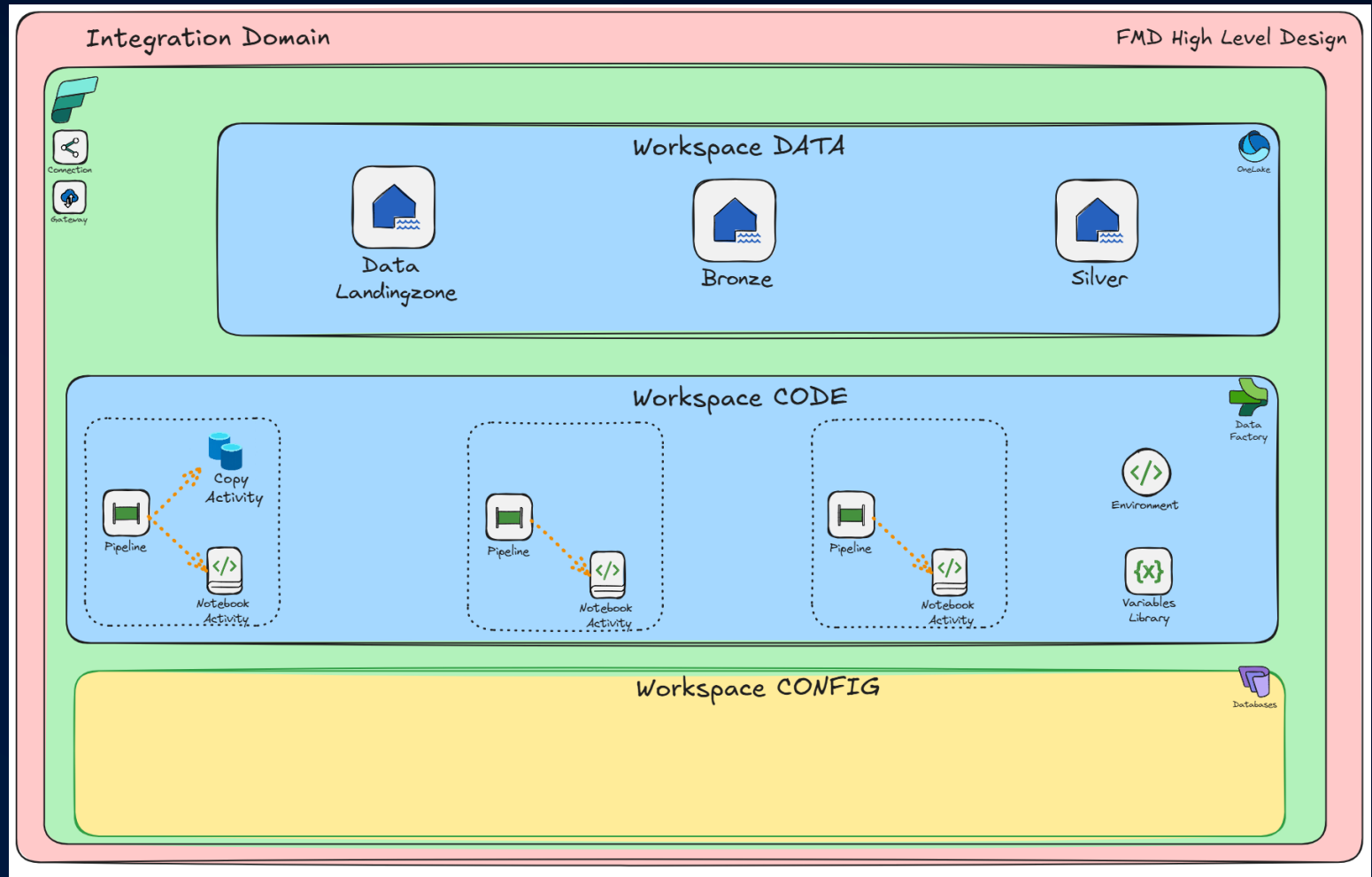
Item deployment

- Lakehouse
- Medallion Architecture
  - Data Landingzone
  - Bronze
  - Silver



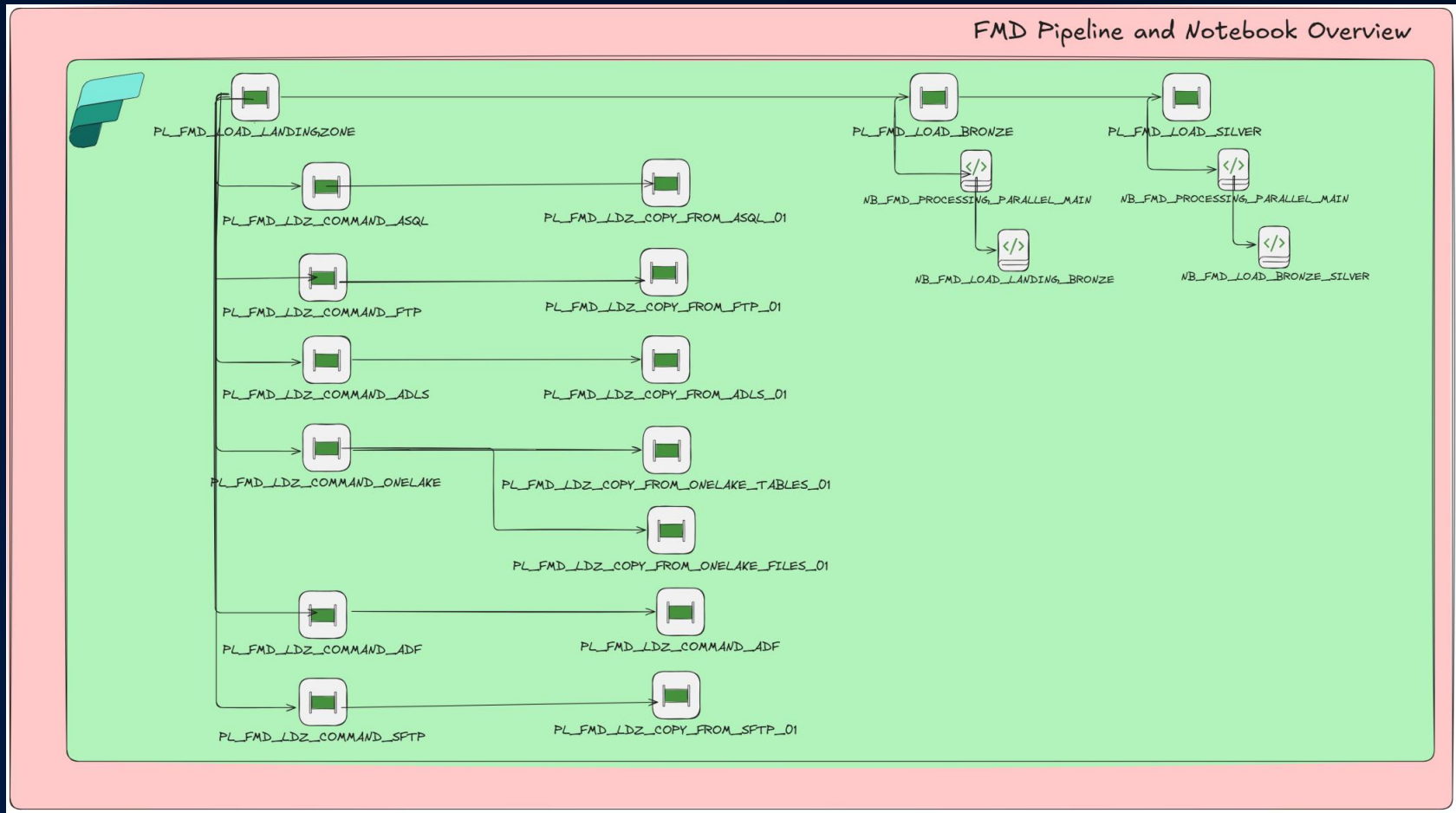
Item deployment

- Variable Library
- Environment



Item deployment

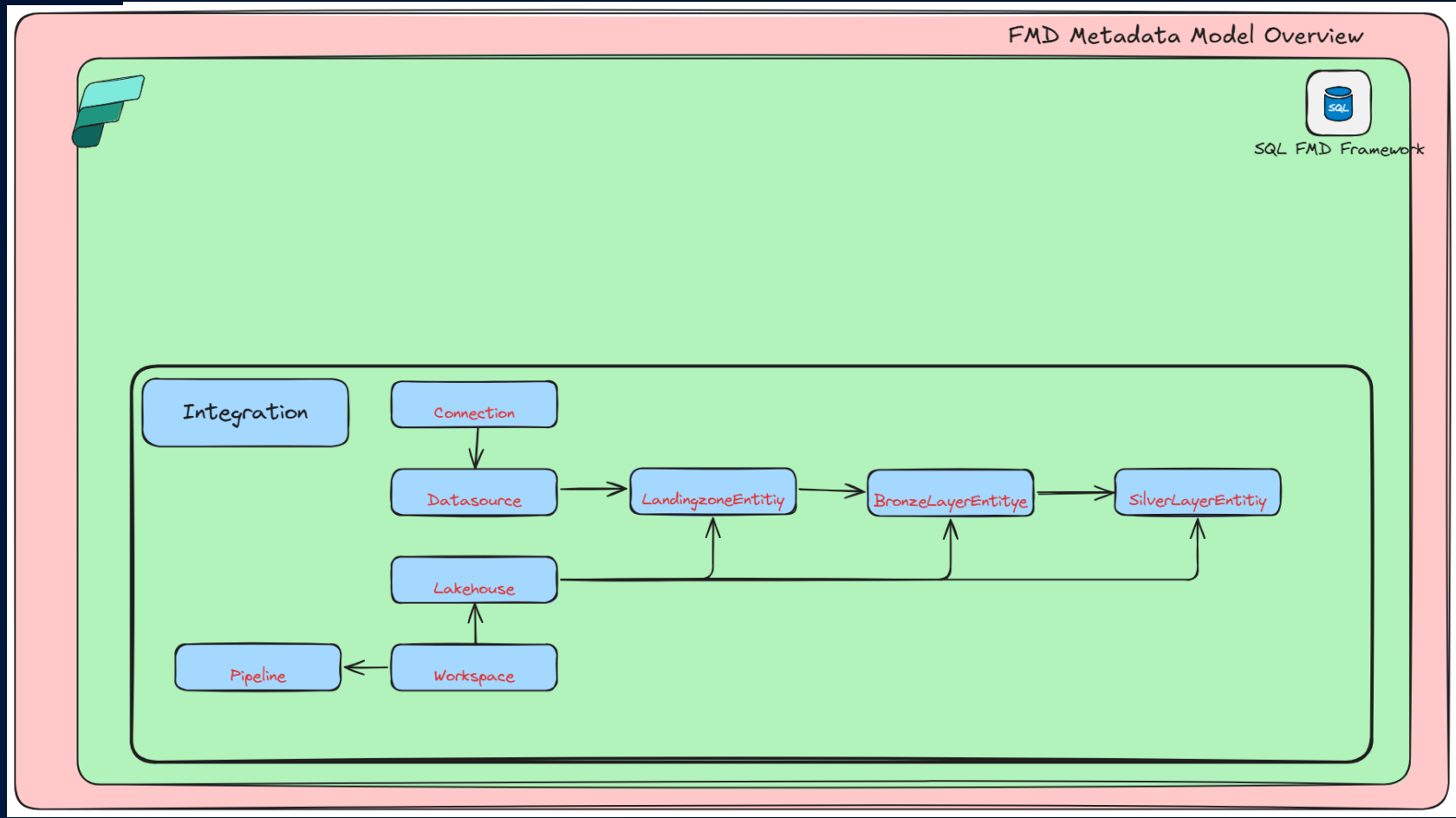
- Variable Library
- Environment
- Data Pipelines
- Notebooks



Data Integration

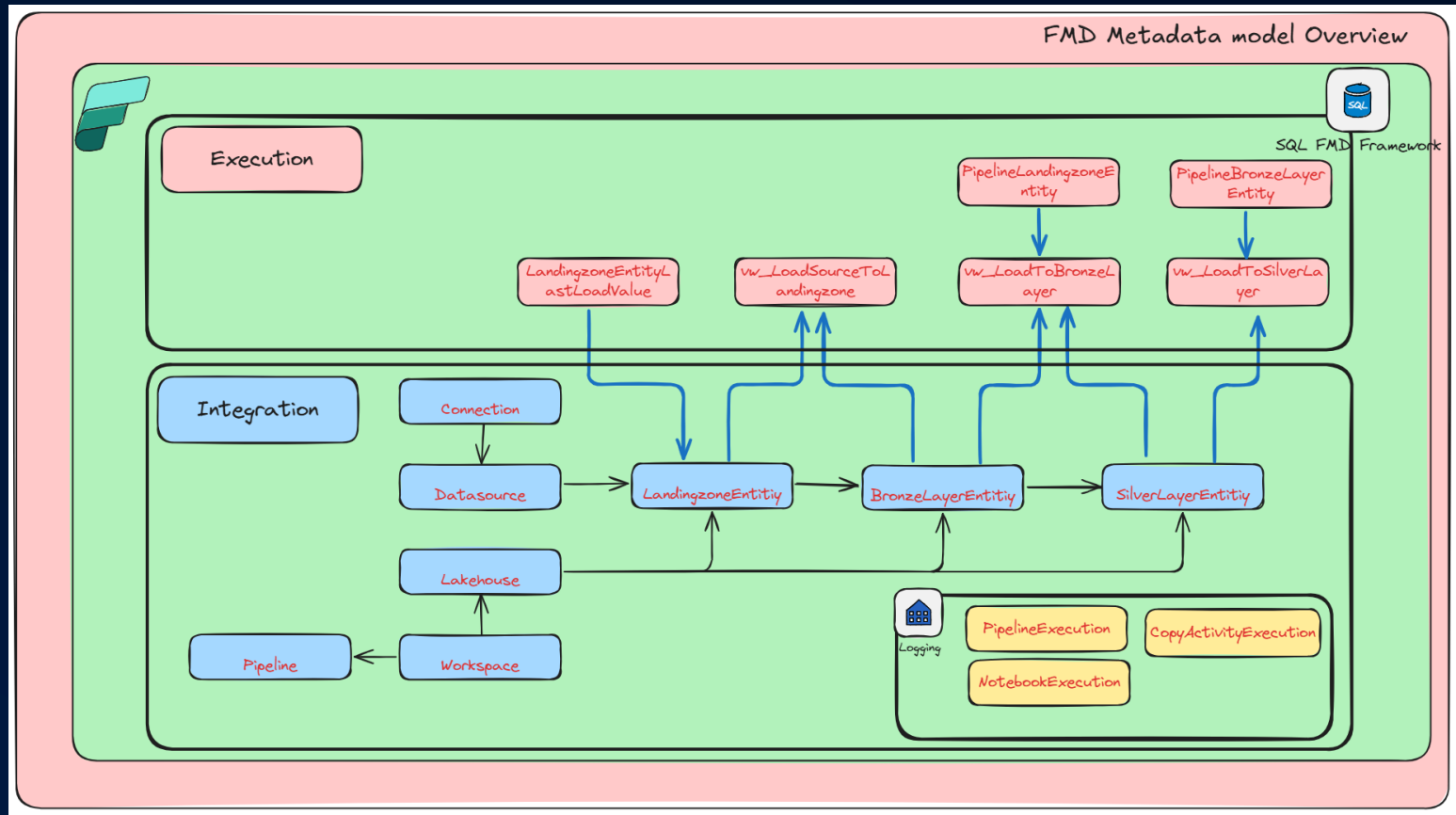
Data Integration

- Fabric database for metadata
- Integration
  - Tables
  - Views
  - Stored Procedures



Data Integration

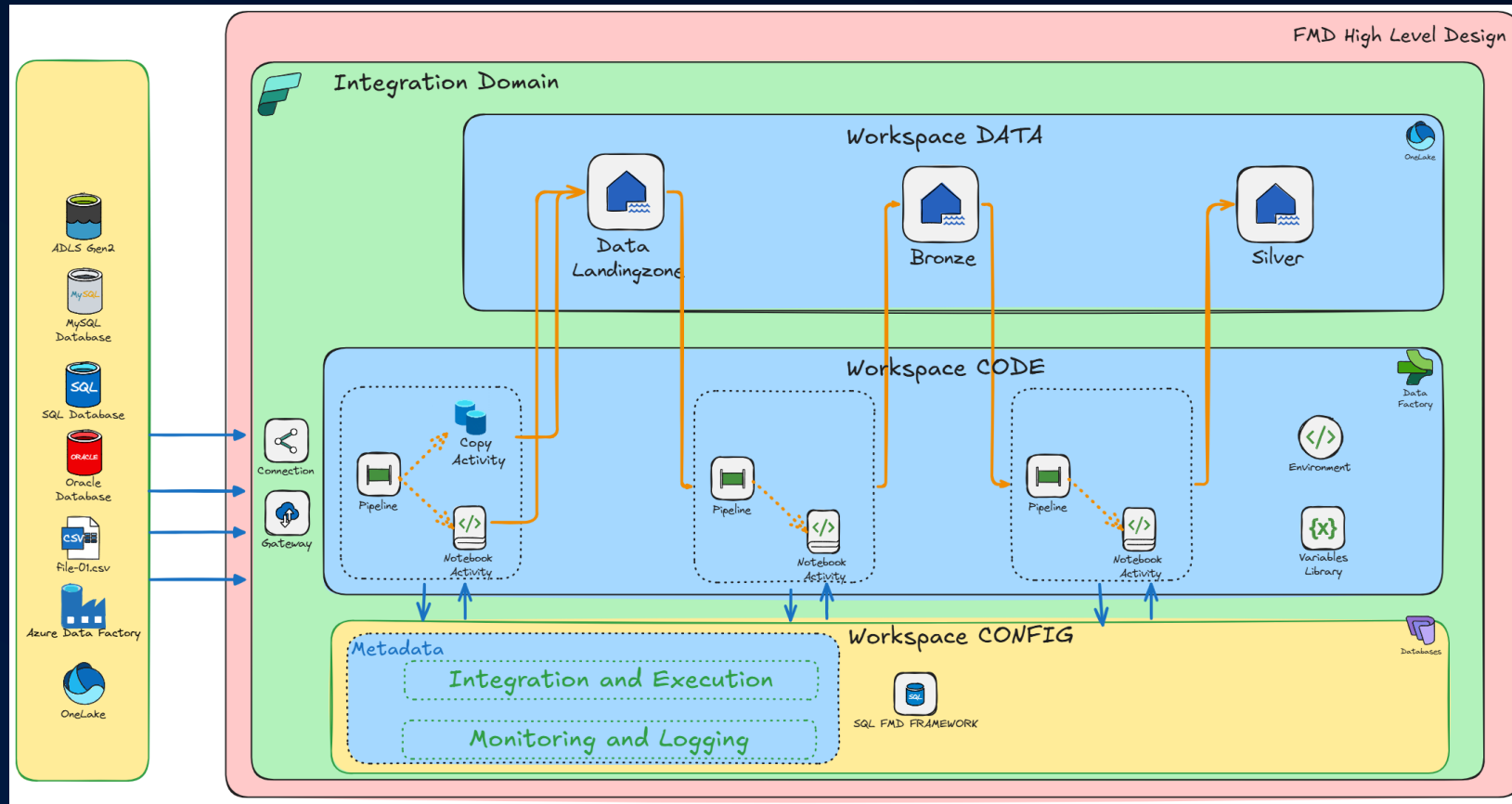
- Execution
  - Tables
  - Views
  - Stored Procedures
- Logging
  - Tables
  - Views
  - Stored Procedures



Data Sources

- SQL Server
- Data Lake Gen 2
- Onelake Tables
- Onelake Files
- FTP / SFTP
- Oracle
- Custom Notebooks(API)
- ADF (Connect to ADF pipeline)

All other connections can easily be added
Every connection needs a new Data Pipeline



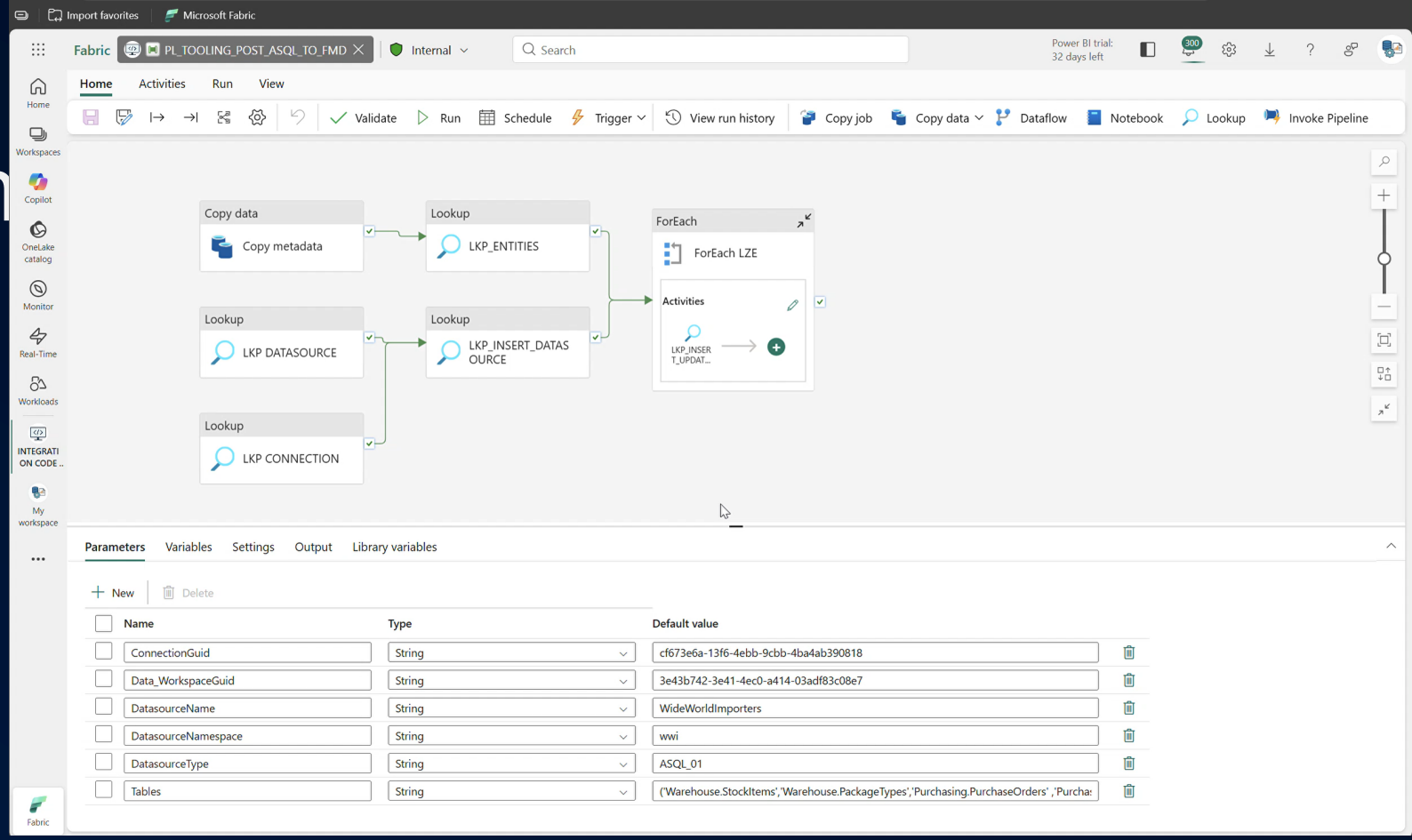
How to get data easily into Fabric Metadata Driven data model

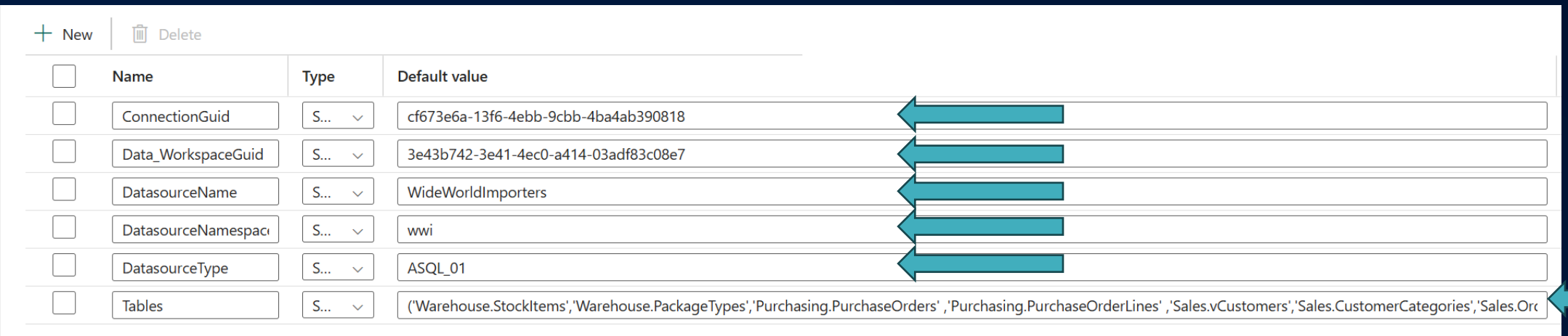
- Metadata driven ingestion Pipeline
- Manual



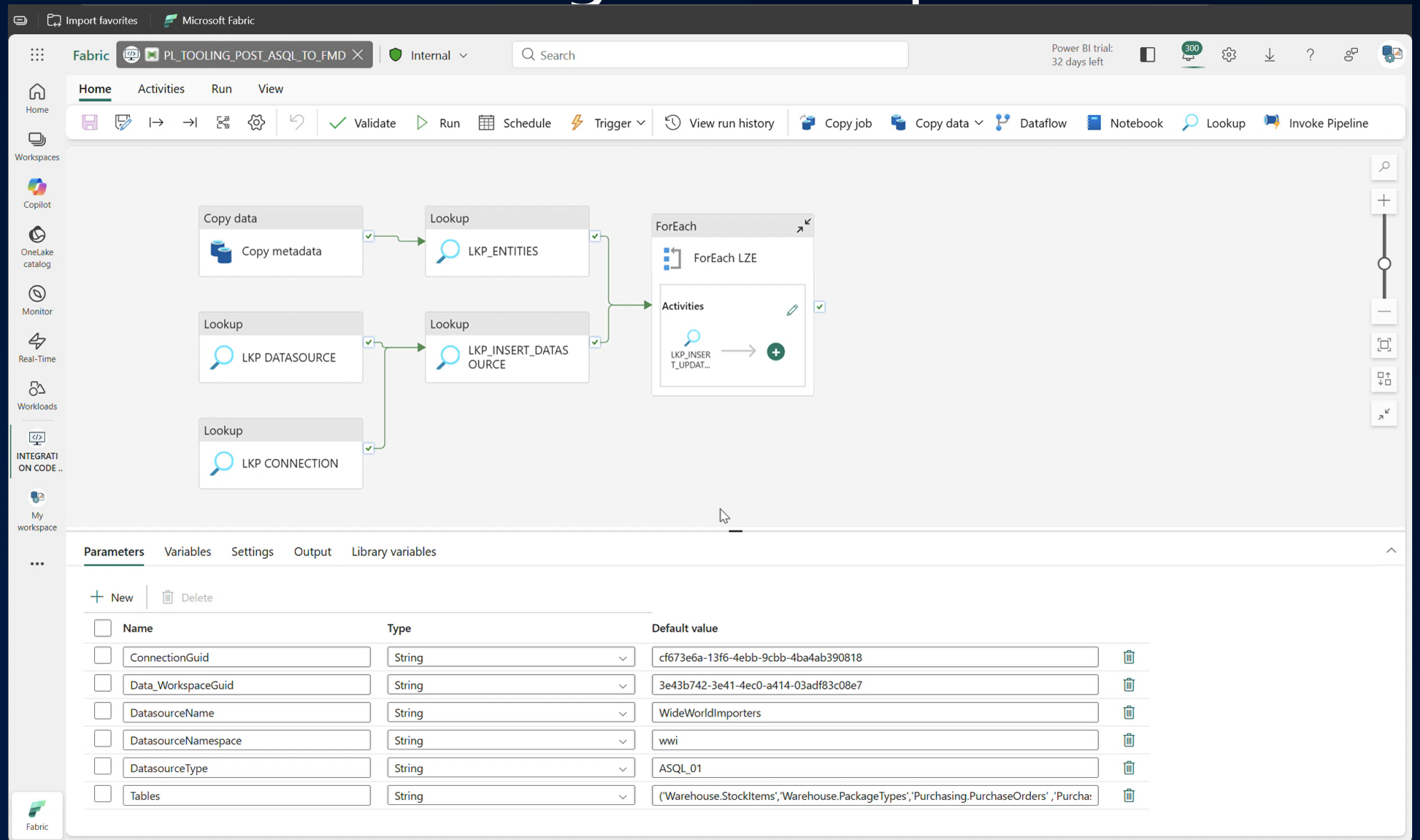
Metadata driven

- ConnectionGuid
- DatasourceName
- Namespace
- DatasourceType
- Tables(schema+table





Metadata driven ingestion Pipeline



How to add a new connection and entity

```
set @ConnectionGuid='11a8e5fe-fbca-4822-9ba4-9162cf56e6dd'
set @Name='CON_FMD_CONTOSO'
set @Type='ASQL_01'

EXECUTE @RC = [integration].[sp_UpsertConnection]
   @ConnectionGuid
  ,@Name
  ,@Type
```

Add Connection

```
set @ConnectionId='Id of connection you want to connect to'
set @DataSourceId=0
set @Name='Databasename'
set @Namespace-'loaction in Landingzone'
set @Type='SQL'
set @Description=''


EXECUTE @RC = [integration].[sp_UpsertDataSource]
   @ConnectionId
  ,@DataSourceId
  ,@Name
  ,@Namespace
  ,@Type
  ,@Description
```

Add Datasource

```
SET @DataSourceId=''
SET @WorkspaceGuid='3e43b742-3e41-4ec0-a414-03adf83c08e7'
SET @SourceSchema='dbo'
SET @SourceName='Employees'
SET @SourceCustomSelect=''
SET @FileName='dbo_Employees'
SET @FilePath='FMD'
SET @FileType='parquet'
SET @IsIncremental=''
SET @IsIncrementalColumn=''
SET @CustomNotebookName=''
SET @PrimaryKeys='EmployeeID'

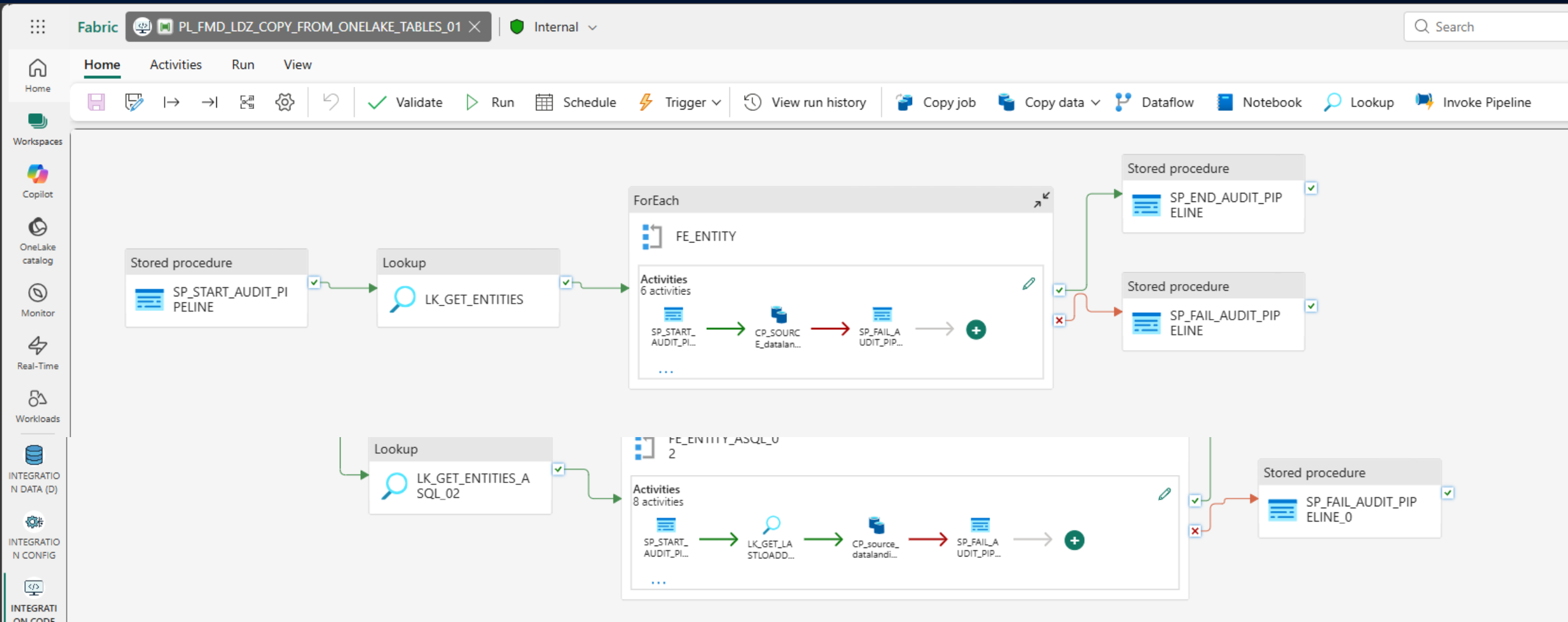
EXECUTE @RC = [integration].[sp_UpsertLandingzoneBronzeSilver]
   @DataSourceId
  ,@WorkspaceGuid
  ,@SourceSchema
  ,@SourceName
  ,@SourceCustomSelect
  ,@FileName
  ,@FilePath
  ,@FileType
  ,@IsIncremental
  ,@IsIncrementalColumn
  ,@CustomNotebookName
  ,@PrimaryKeys
```

Add Entity

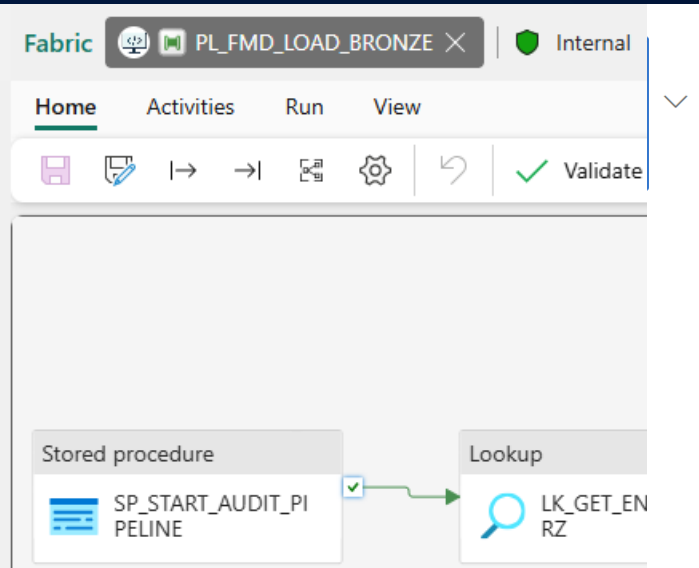
Demo

Data Processing

Data Processing (source to Data Landingzone)



Data Processing (Data LDZ to Bronze/Silver)



Parameters

```
1  from json import loads, dumps
2  import uuid
3  from datetime import datetime
4  NotebookExecutionId = str(uuid.uuid4())
5
```

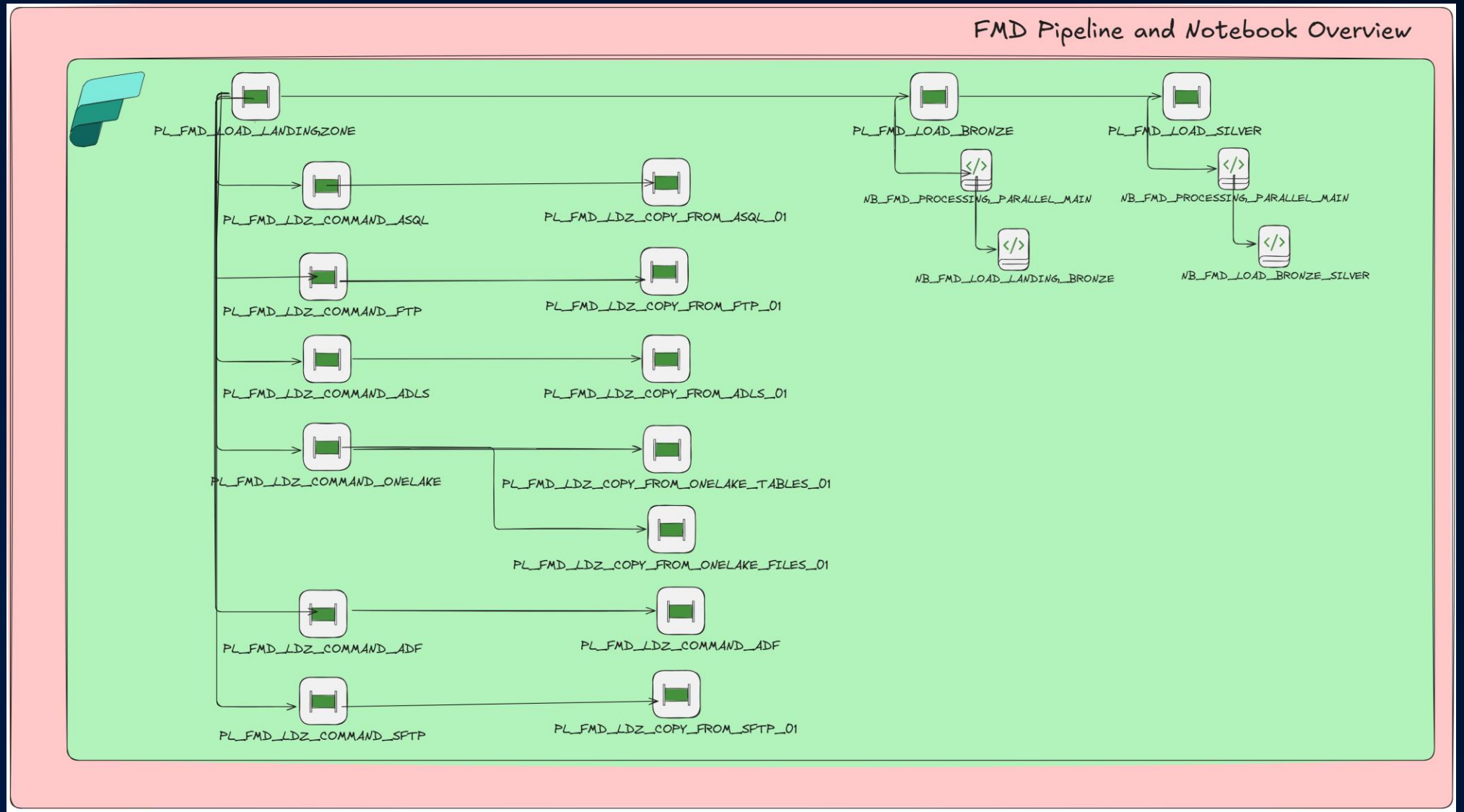
PySpark (Python)

```
1  variable_settings=notebookutils.variableLibrary.getLibrary("VAR_FMD")
```

PySpark (Python)

```
1  Path = ""
2  useRootDefaultLakehouse= True
3  PipelineRunGuid = ""
4  PipelineGuid = ""
5  TriggerGuid = ""
6  TriggerTime = ""
```

Data Processing



Data Cleansing · edkreuk/FMD_FRAMEWORK Wiki

Data Processing Cleansing rules (BRZ/SLV)

- NB_FMD_DQ_CLEANSING
  - Built in Rules
    - normalize_text
    - split
    - fill_nulls
    - parse_datetime

```
df_norm = normalize_text(
    df,
    columns=["title"],
    args={"case": "lower", "empty_as_null": False}
)
```

```
df_norm = normalize_text(
    df,
    columns=["name"],
    args={"case": "title", "collapse_spaces": False}
)
```

```
DECLARE @RC int
DECLARE @SilverLayerEntityId int
DECLARE @CleansingRules nvarchar(max)

-- TODO: Set parameter values here.

EXECUTE @RC = [integration].[sp_UpsertSilverCleansingRule]
   24
  ,[{"function": "normalize_text","columns": "CustomerName","parameters": {"case": "upper"}}]
GO
```

ssages | Results | {} | Search

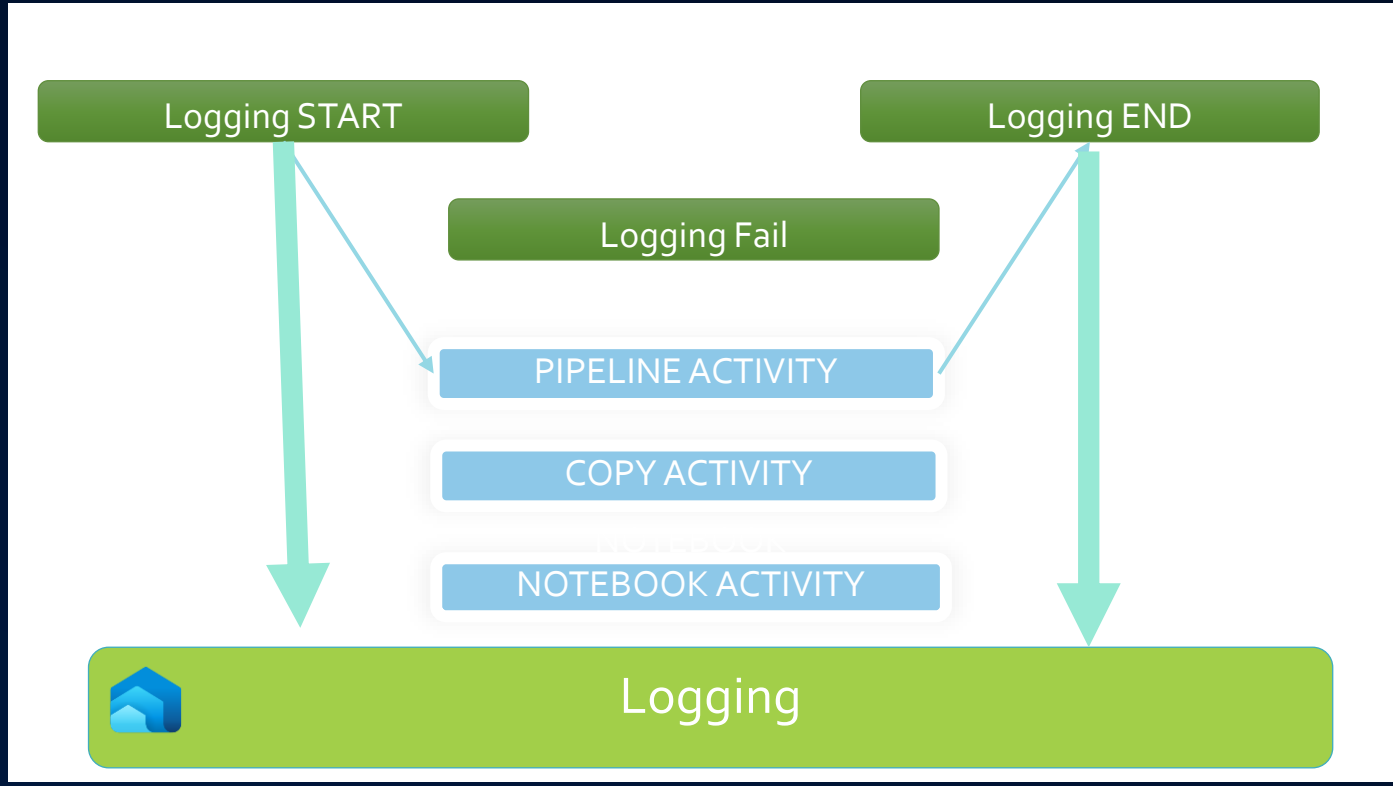
123 SilverLayerEntityId	ABC CleansingRules
24	{"function": "normalize_text","columns": "CustomerName","parameters": {"case": "upper"}}

Demo

Data Monitoring

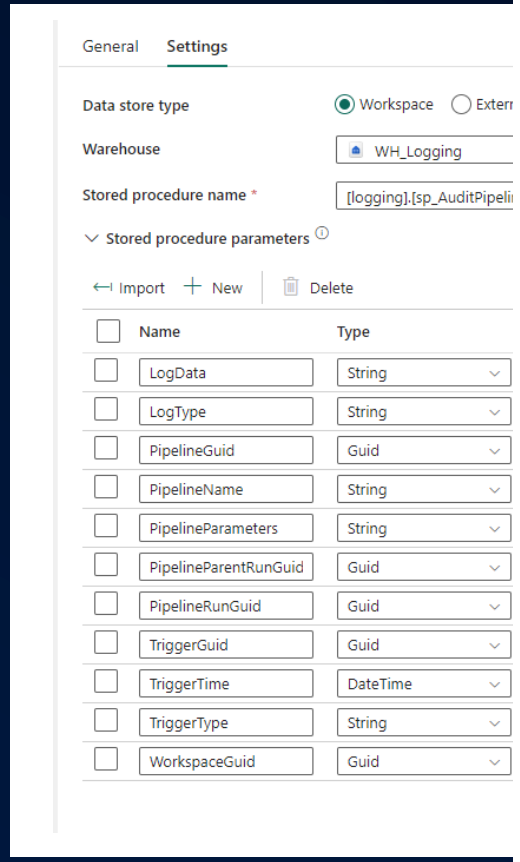
Data Monitoring

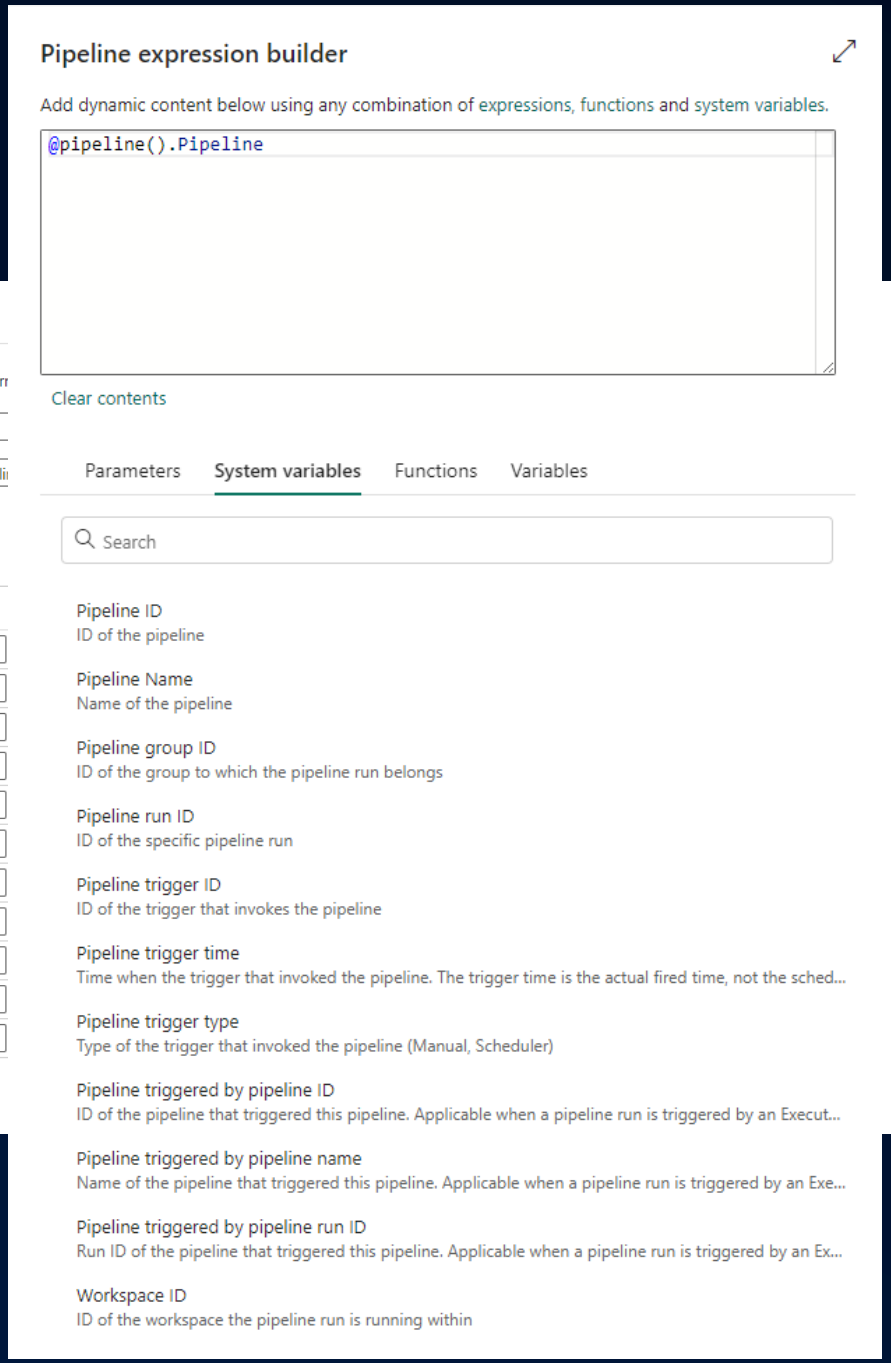
- Log Start and End Time of records
- Log Extracted Records
- Log Execution Failure



Data Monitoring

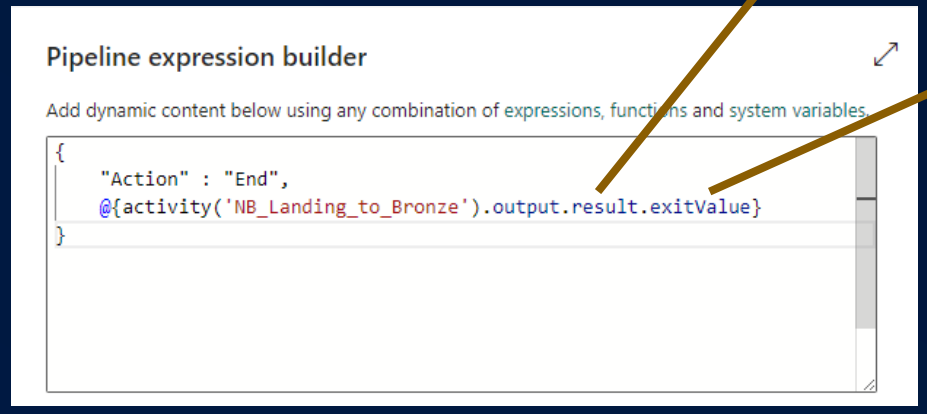
- Add Information about pipelines
- Adding System Variables





Data Monitoring

- Add Information about pipelines
- Adding System Variables
- Add Information about Notebooks




```
"status": "Succeeded",
"result": {
    "runId": "10883ea1-cac8-4d1b-92d3-03f8ca451054",
    "runStatus": "Succeeded",
    "sessionId": "188bc363-7872-453e-9369-8a18c115f6e8",
    "sparkPool": "1acc1c9e-b8d2-4f1d-ba36-5ba8e5088067",
    "error": null,
    "lastCheckedOn": "2025-03-17T18:29:37.44Z",
    "metadata": {
        "isForPipeline": null,
        "runStartTime": "2025-03-17T18:25:13.2077314Z",
        "runEndTime": "2025-03-17T18:29:37.4521264Z"
    },
    "highConcurrencyModeStatus": null,
    "exitValue": "[{'TableName': 'Sales_Orders_0_0', 'exitVal': \"{'CopyOutput':
},
"message": "Notebook execution is in Succeeded state, runId: 10883ea1-cac8-4d1b-9
"SparkMonitoringURL": "https://app.powerbi.com/workloads/de-ds/sparkmonitor/fcd42
"executionDuration": 266
```

Data Monitoring

PipelineExecution

Data preview - PipelineExecution

Showing 1000 rows

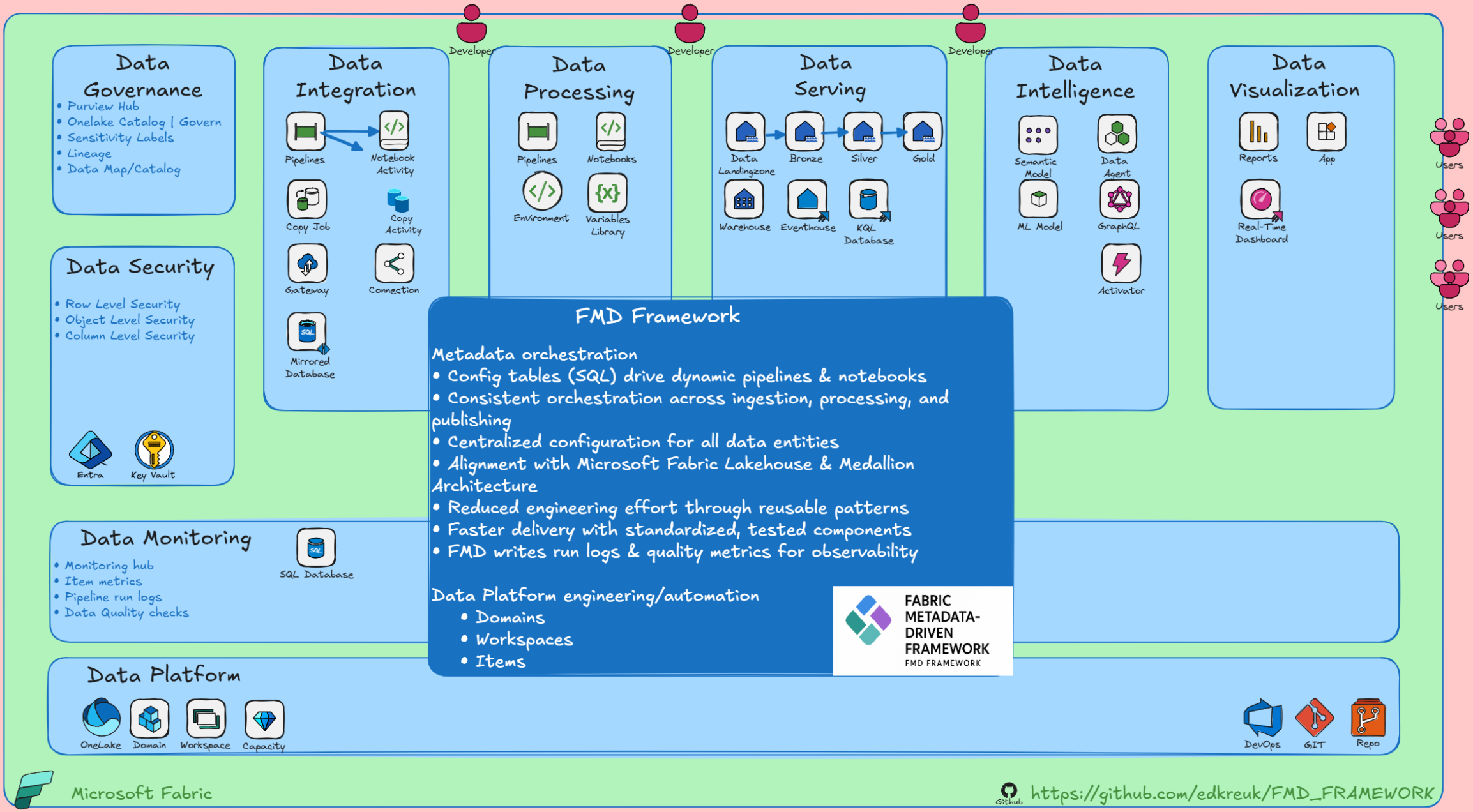
Search

	Workspace...	PipelineRun...	PipelinePare...	PipelineGuid	PipelineName	PipelinePara...	TriggerType	TriggerGuid	TriggerTime	LogType	LogDateTime	LogData
1	1ACC1C9E-B8D2...	C19BCD8F-6F46...	*NULL*	67B39D07-7CE2...	PL_FMD_LOAD_...	*NULL*	Manual	AA34BEA8-E884...	2025-03-17 18:1...	EndPipeline	2025-03-17 18:3...	{ "Action" : "End" }
2	1ACC1C9E-B8D2...	1C939008-FF26-...	*NULL*	DC3E6A24-2155...	PL_FMD_LOAD_S...	*NULL*	Manual	6C182932-0447-...	2025-03-17 18:3...	EndPipeline	2025-03-17 18:3...	{↵ "Action" : "En...
3	1ACC1C9E-B8D2...	1C939008-FF26-...	*NULL*	DC3E6A24-2155...	PL_FMD_LOAD_S...	*NULL*	Manual	6C182932-0447-...	2025-03-17 18:3...	StartPipeline	2025-03-17 18:3...	{ "Action" : "Start...
4	1ACC1C9E-B8D2...	2BB79DC5-CB19...	*NULL*	69523C39-B819-...	PL_FMD_LOAD_B...	*NULL*	Manual	AA45F0C5-24D3...	2025-03-17 18:2...	StartPipelineActi...	2025-03-17 18:2...	{ "Action" : "Start...
5	1ACC1C9E-B8D2...	C19BCD8F-6F46...	*NULL*	67B39D07-7CE2...	PL_FMD_LOAD_...	*NULL*	Manual	AA34BEA8-E884...	2025-03-17 18:1...	EndPipeline	2025-03-17 18:2...	{↵ "Action" : "Err...
6	1ACC1C9E-B8D2...	2BB79DC5-CB19...	*NULL*	69523C39-B819-...	PL_FMD_LOAD_B...	*NULL*	Manual	AA45F0C5-24D3...	2025-03-17 18:2...	EndPipelineActiv...	2025-03-17 18:3...	{↵ "Action" : "En...
7	1ACC1C9E-B8D2...	CF5D046E-7D58...	*NULL*	C0737F04-B4E9-...	PL_FMD_LDZ_CO...	*NULL*	Manual	F26FF0CD-71A9-...	2025-03-17 18:1...	EndPipeline	2025-03-17 18:2...	{↵ "Action" : "En...
8	1ACC1C9E-B8D2...	D238165C-FCDF...	*NULL*	B8236D27-F8C2...	PL_FMD_LDZ_CO...	*NULL*	Manual	800DB4CE-CA74...	2025-03-17 18:1...	EndPipeline	2025-03-17 18:2...	{↵ "Action" : "En...
9	1ACC1C9E-B8D2...	D238165C-FCDF...	*NULL*	B8236D27-F8C2...	PL_FMD_LDZ_CO...	Invoices	Manual	800DB4CE-CA74...	2025-03-17 18:1...	EndCopyActivity	2025-03-17 18:2...	{↵ "Action" : "En...
10	1ACC1C9E-B8D2...	D238165C-FCDF...	*NULL*	B8236D27-F8C2...	PL_FMD_LDZ_CO...	InvoiceLines	Manual	800DB4CE-CA74...	2025-03-17 18:1...	EndCopyActivity	2025-03-17 18:2...	{↵ "Action" : "En...
11	1ACC1C9E-B8D2...	D238165C-FCDF...	*NULL*	B8236D27-F8C2...	PL_FMD_LDZ_CO...	OrderLines	Manual	800DB4CE-CA74...	2025-03-17 18:1...	EndCopyActivity	2025-03-17 18:2...	{↵ "Action" : "En...
12	1ACC1C9E-B8D2...	D238165C-FCDF...	*NULL*	B8236D27-F8C2...	PL_FMD_LDZ_CO...	Orders	Manual	800DB4CE-CA74...	2025-03-17 18:1...	EndCopyActivity	2025-03-17 18:2...	{↵ "Action" : "En...
13	1ACC1C9E-B8D2...	D238165C-FCDF...	*NULL*	B8236D27-F8C2...	PL_FMD_LDZ_CO...	BuyingGroups	Manual	800DB4CE-CA74...	2025-03-17 18:1...	EndCopyActivity	2025-03-17 18:2...	{↵ "Action" : "En...
14	1ACC1C9E-B8D2...	D238165C-FCDF...	*NULL*	B8236D27-F8C2...	PL_FMD_LDZ_CO...	PurchaseOrderLi...	Manual	800DB4CE-CA74...	2025-03-17 18:1...	EndCopyActivity	2025-03-17 18:2...	{↵ "Action" : "En...

Conclusion

	Out of the Box Framework	Custom Made Framework	Fabric Metadata-driven Framework
Ready to use	✓	✗	✓
Rapid Implementation	✓	✗	✓
Lower development effort	✓	✗	✓ ✗
Lower upfront cost	✓	✗	✓
Support and Updated	✓	✗	✓ ✗
Tailored to specific needs.	✗	✓	✗
Full control over design and features.	✗	✓	✓
Easy to extend.	✗	✓	✓

Fabric Metadata-Driven Framework



Key Learnings and Best Practices

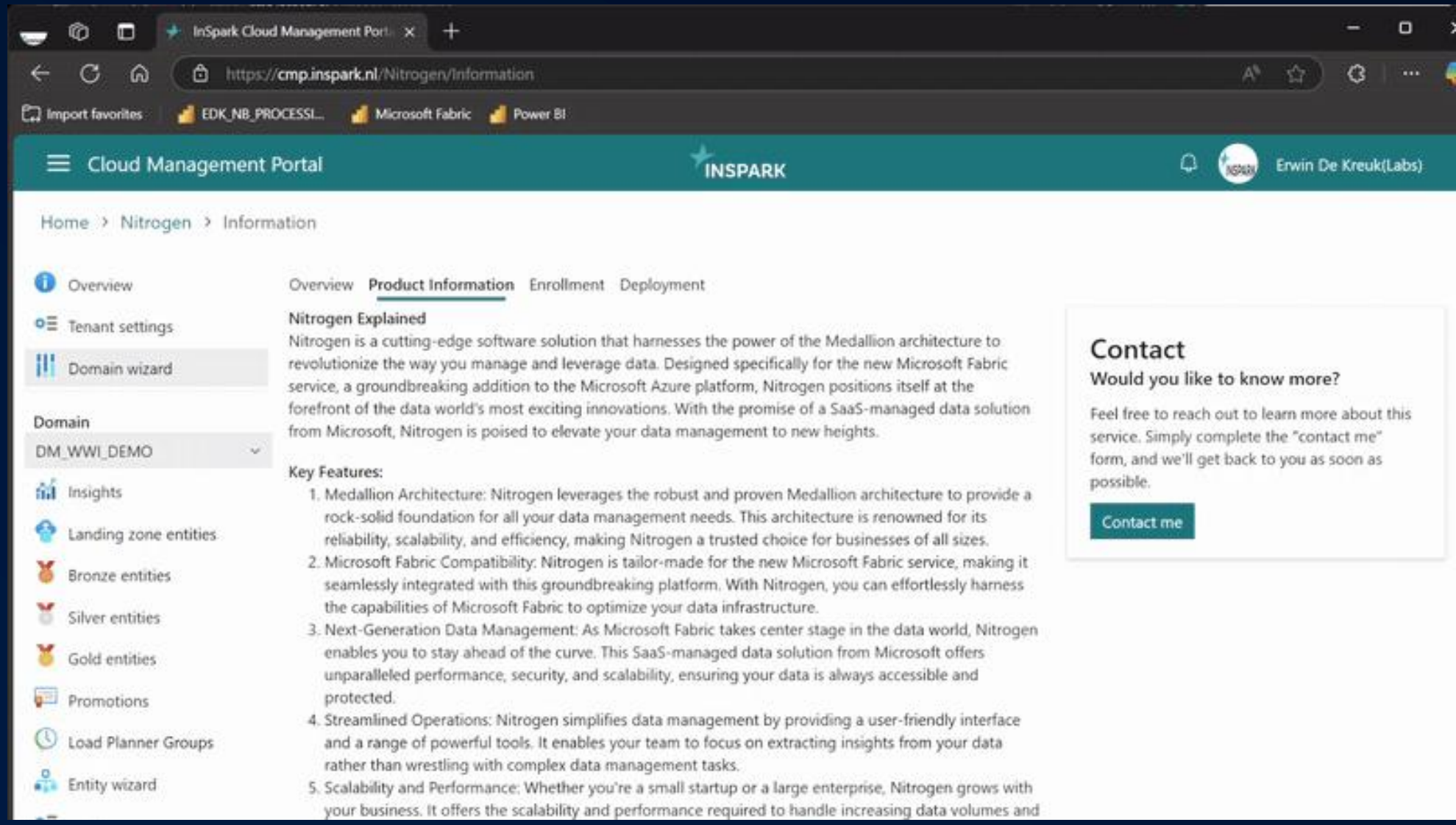


Metadata -driven approach is a best practice for managing data pipelines

Medallion Lakehouse architecture is a proven framework for implementing Metadata -driven approach

Start small and gradually expand the Metadata -driven approach across the organization

Nitrogen Control Center



Feedback



From Manual to Automated: Master Metadata-Driven Design in Fabric

LINKS

From Manual to Automated: Master Metadata-Driven Design in Fabric

FMD Framework

Center

Let’s connect LinkedIn