#FABCONSQLCON2026

FABCON
Microsoft Fabric
COMMUNITY CONFERENCE

SQLCON
Microsoft SQL
COMMUNITY CONFERENCE

ATLANTA MARCH 16 - 20, 2026

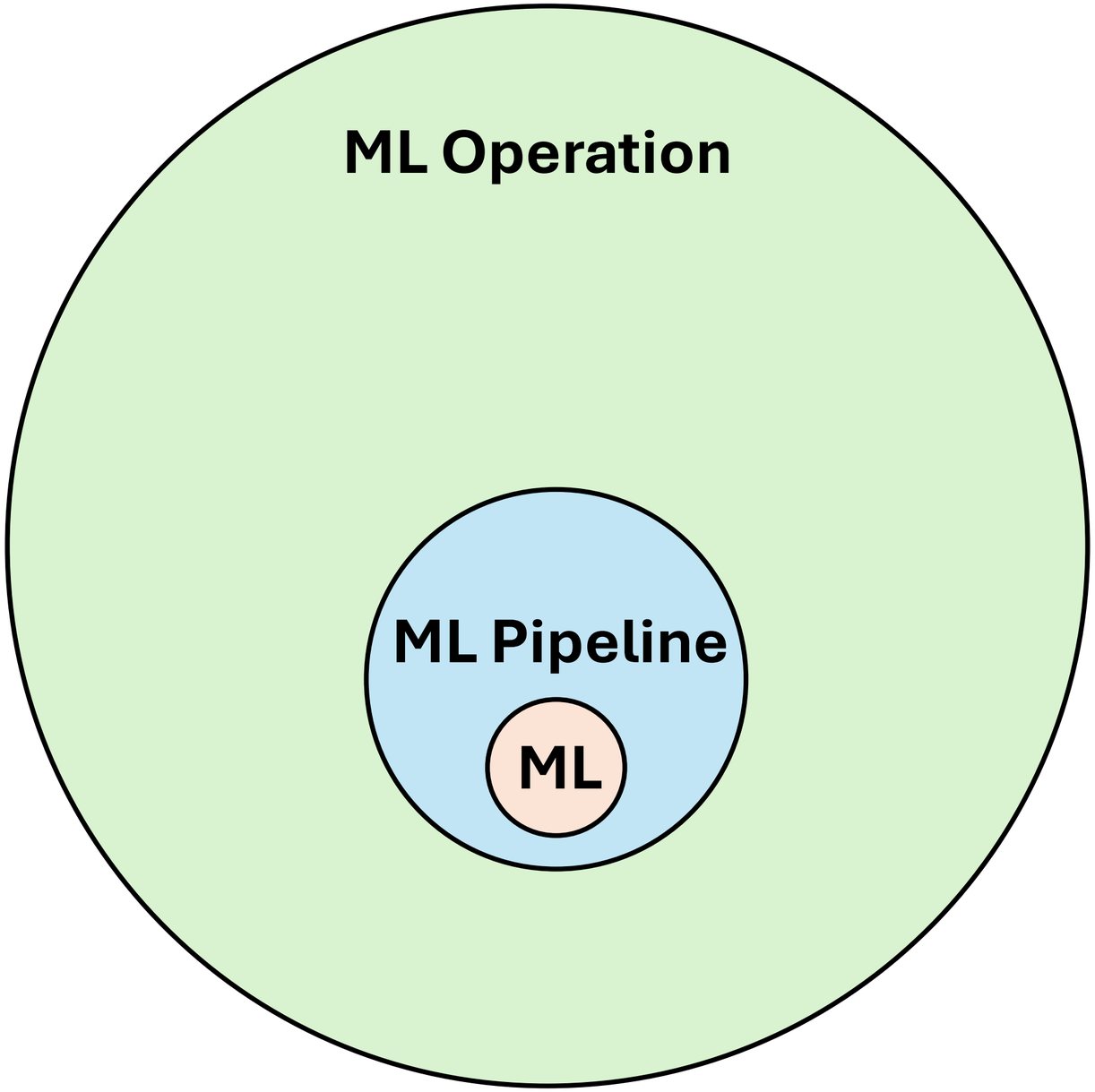
Building MLOps for Traditional ML in Microsoft Fabric

Nicholas Leonhard
Machine Learning Engineer

Connect on LinkedIn:

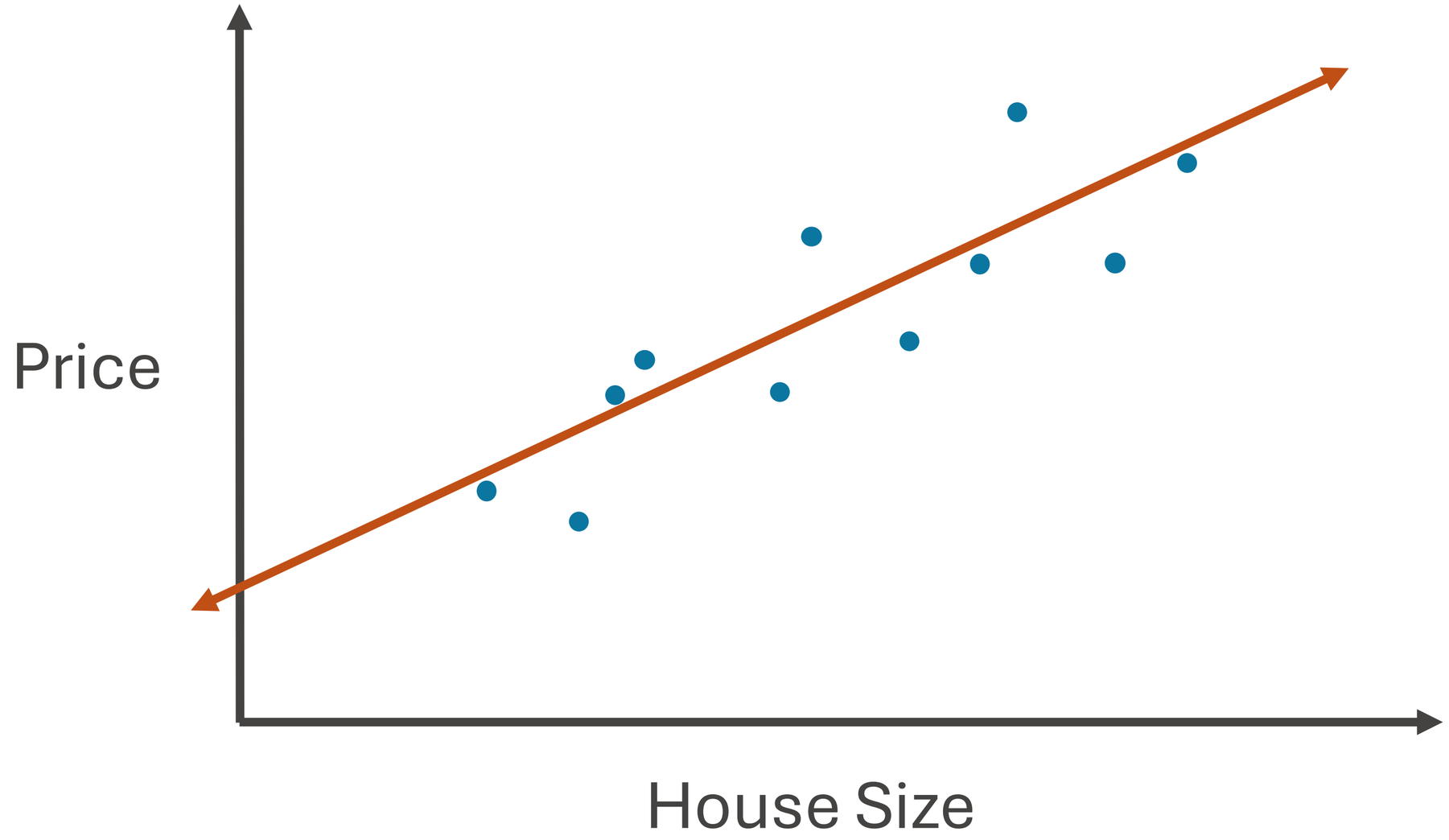
https://www.linkedin.com/in/nicholas-leonhard-21a972ab

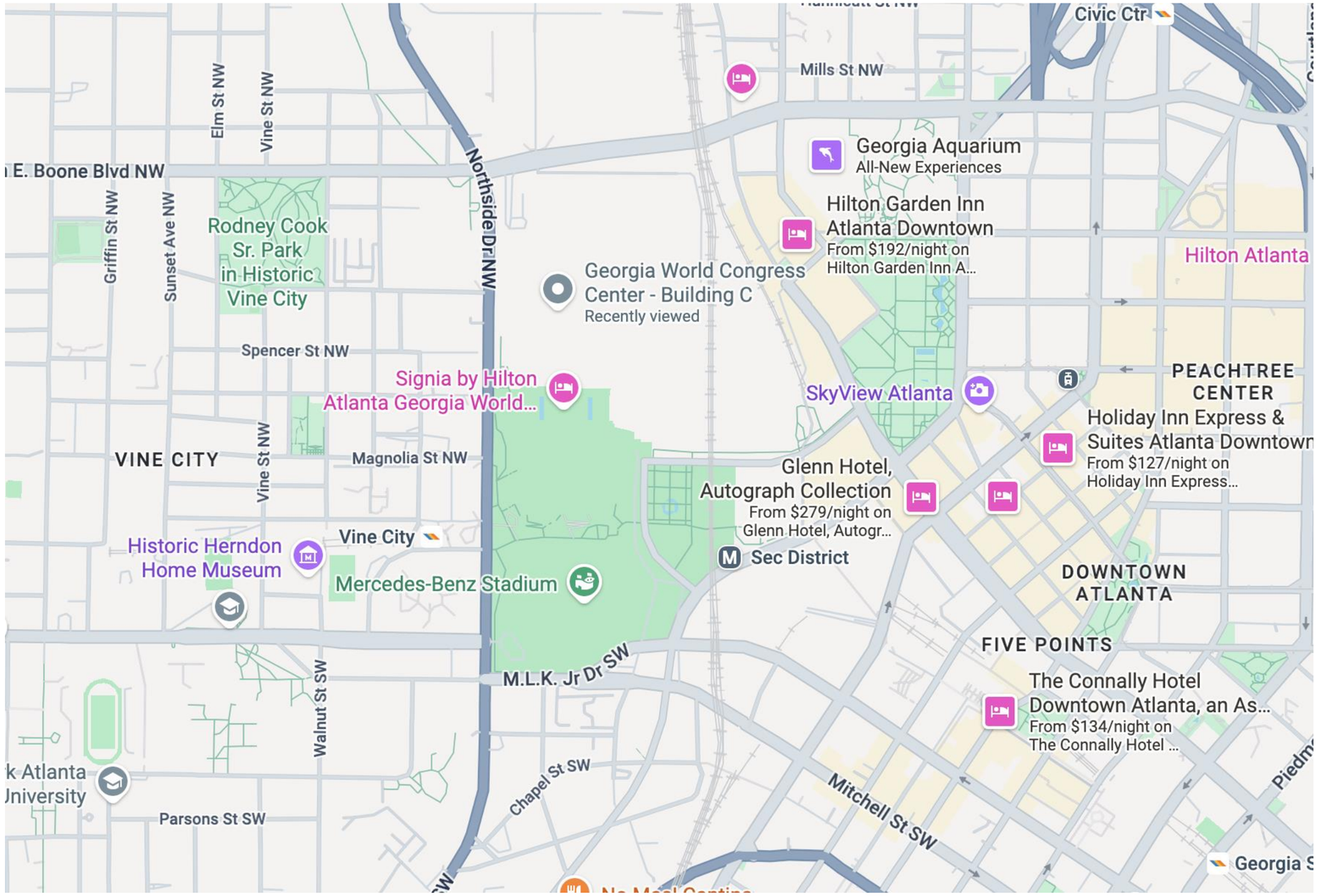
What is MLOps?

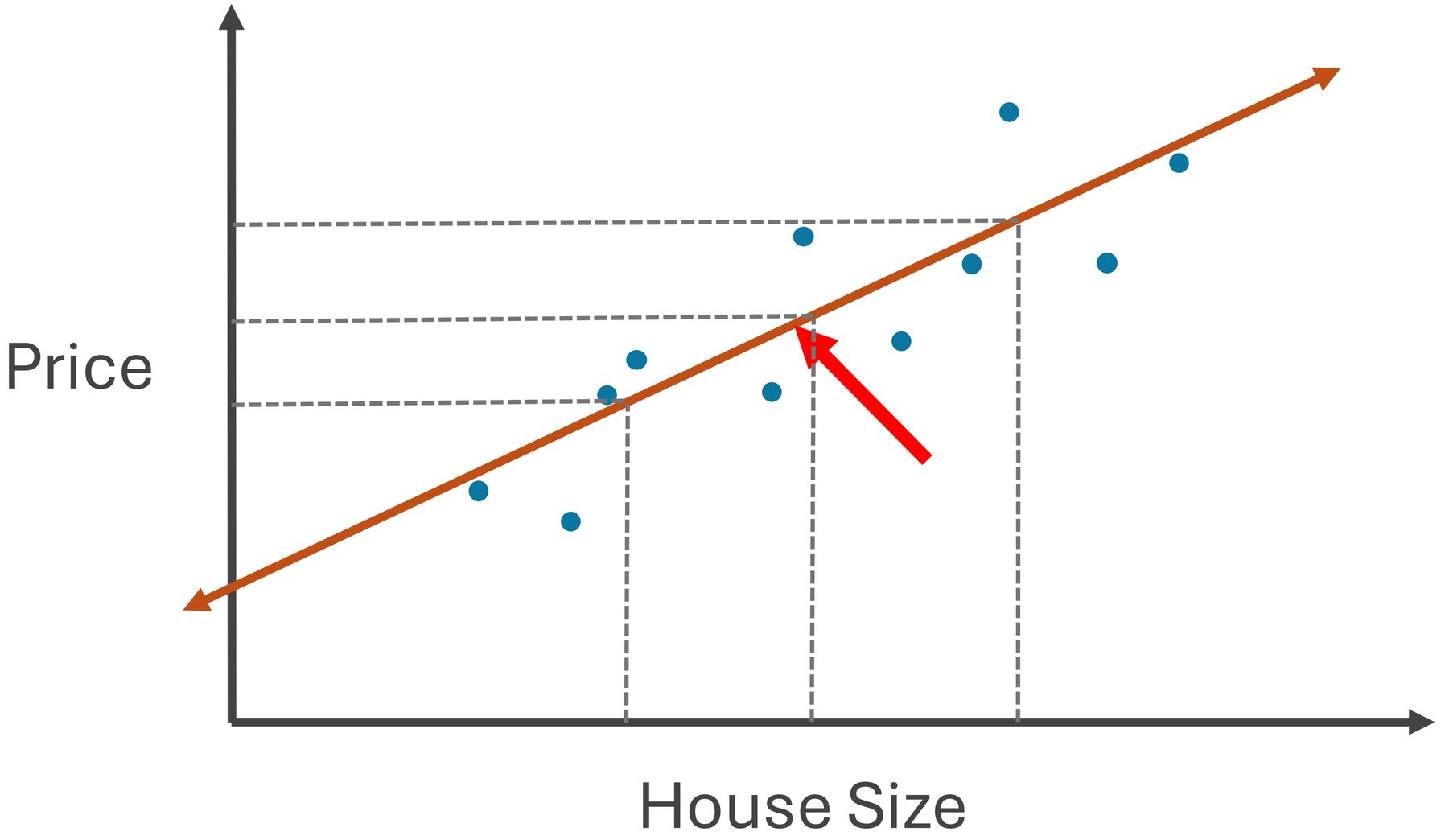

ML Operation
ML Pipeline
ML

What is Machine Learning?

- Model a relationship


Price
House Size


Civic Ctr
Mills St NW
Georgia Aquarium
All-New Experiences
E. Boone Blvd NW
Elm St NW
Vine St NW
Griffin St NW
Sunset Ave NW
Northside Dr NW
Rodney Cook Sr. Park in Historic Vine City
Hilton Garden Inn Atlanta Downtown
From $192/night on Hilton Garden Inn A...
Hilton Atlanta
Georgia World Congress Center - Building C
Recently viewed
Spencer St NW
Signia by Hilton Atlanta Georgia World...
SkyView Atlanta
PEACHTREE CENTER
Holiday Inn Express & Suites Atlanta Downtown
From $127/night on Holiday Inn Express...
VINE CITY
Magnolia St NW
Glenn Hotel, Autograph Collection
From $279/night on Glenn Hotel, Autogr...
Historic Herndon Home Museum
Vine City
Sec District
Mercedes-Benz Stadium
DOWNTOWN ATLANTA
FIVE POINTS
M.L.K. Jr Dr SW
Walnut St SW
The Connally Hotel Downtown Atlanta, an As...
From $134/night on The Connally Hotel ...
Chapel St SW
Parsons St SW
Mitchell St SW
Georgia S


Price
House Size

Statistical Methods

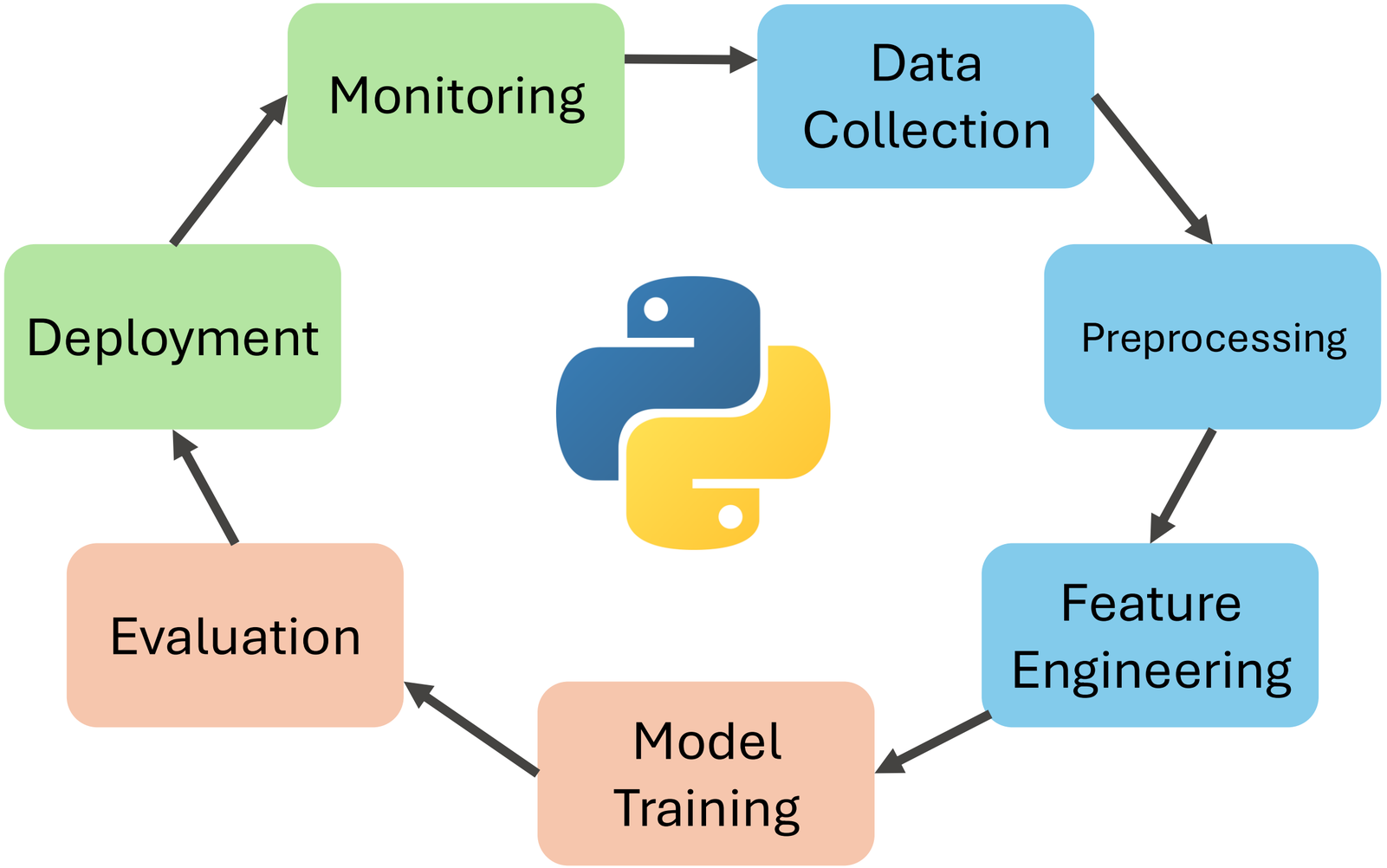
+

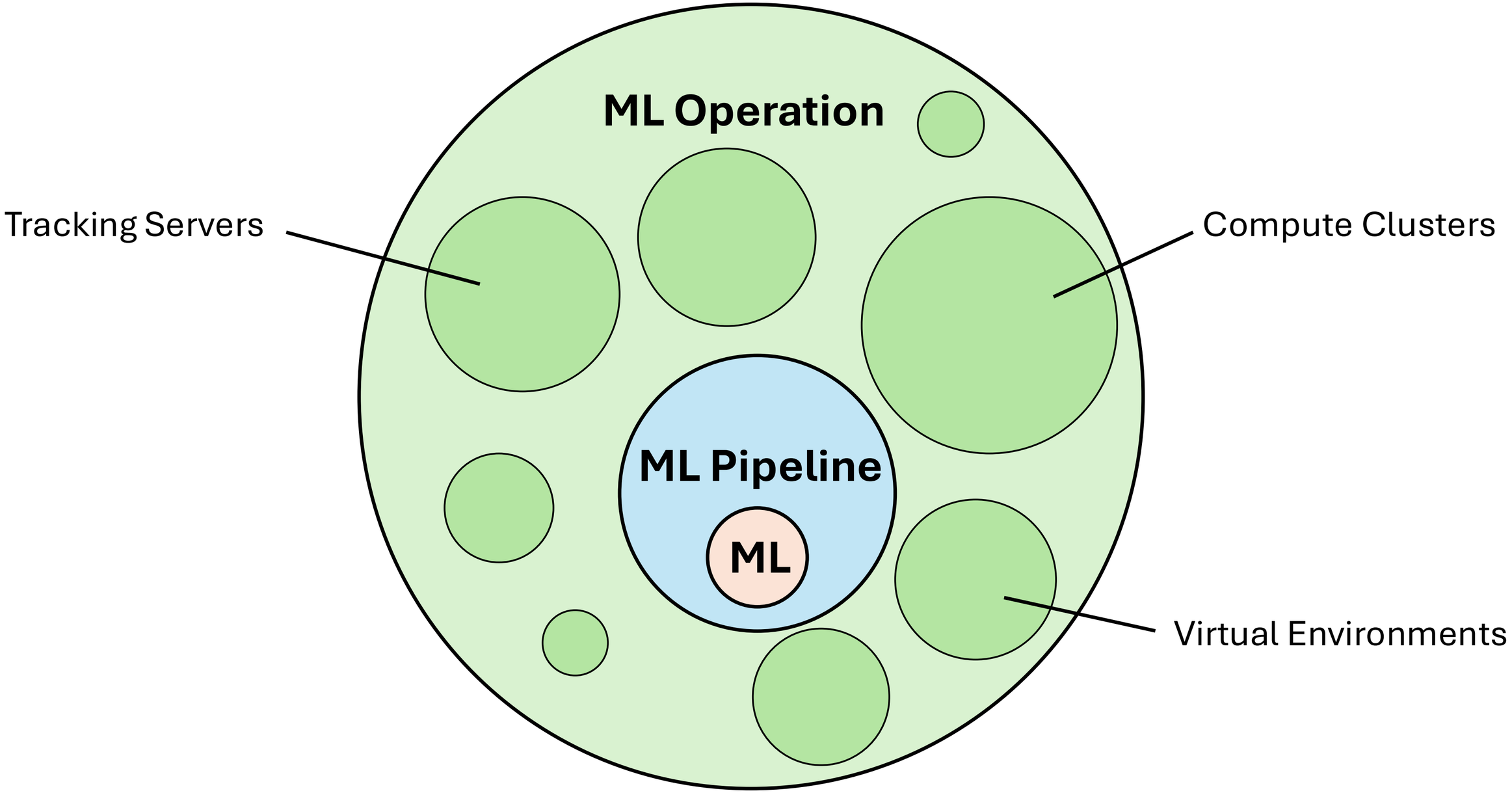
Computational Techniques

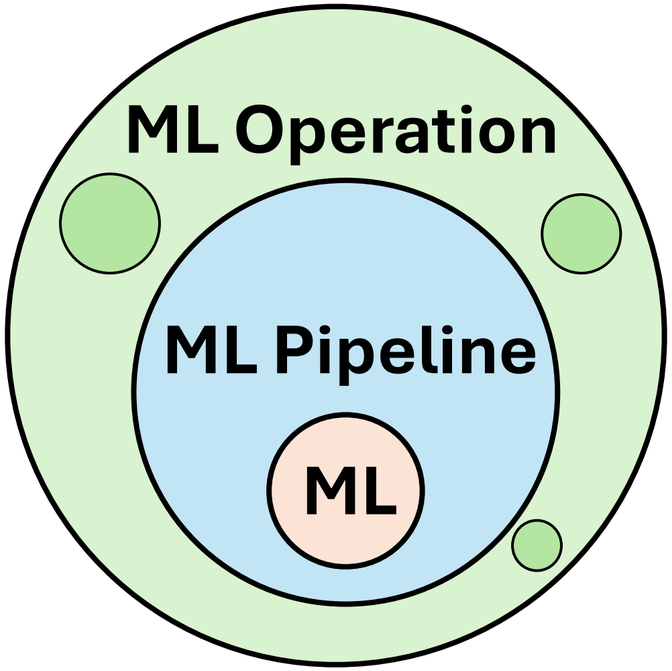
=

Machine Learning

Machine Learning ~~Lifecycle~~ Pipeline




ML Operation
Tracking Servers
Compute Clusters
ML Pipeline
ML
Virtual Environments


ML Operation
ML Pipeline
ML

Before Fabric

Limitations

- Not a server
- Technical overhead
- Zero redundancy

Microsoft Fabric

- All Tools, One Platform
- Reduced Complexity
- Scalability
- Microsoft Ecosystem

MLOps on Fabric

Basic Framework

x1 Lakehouse

x1 Pipeline

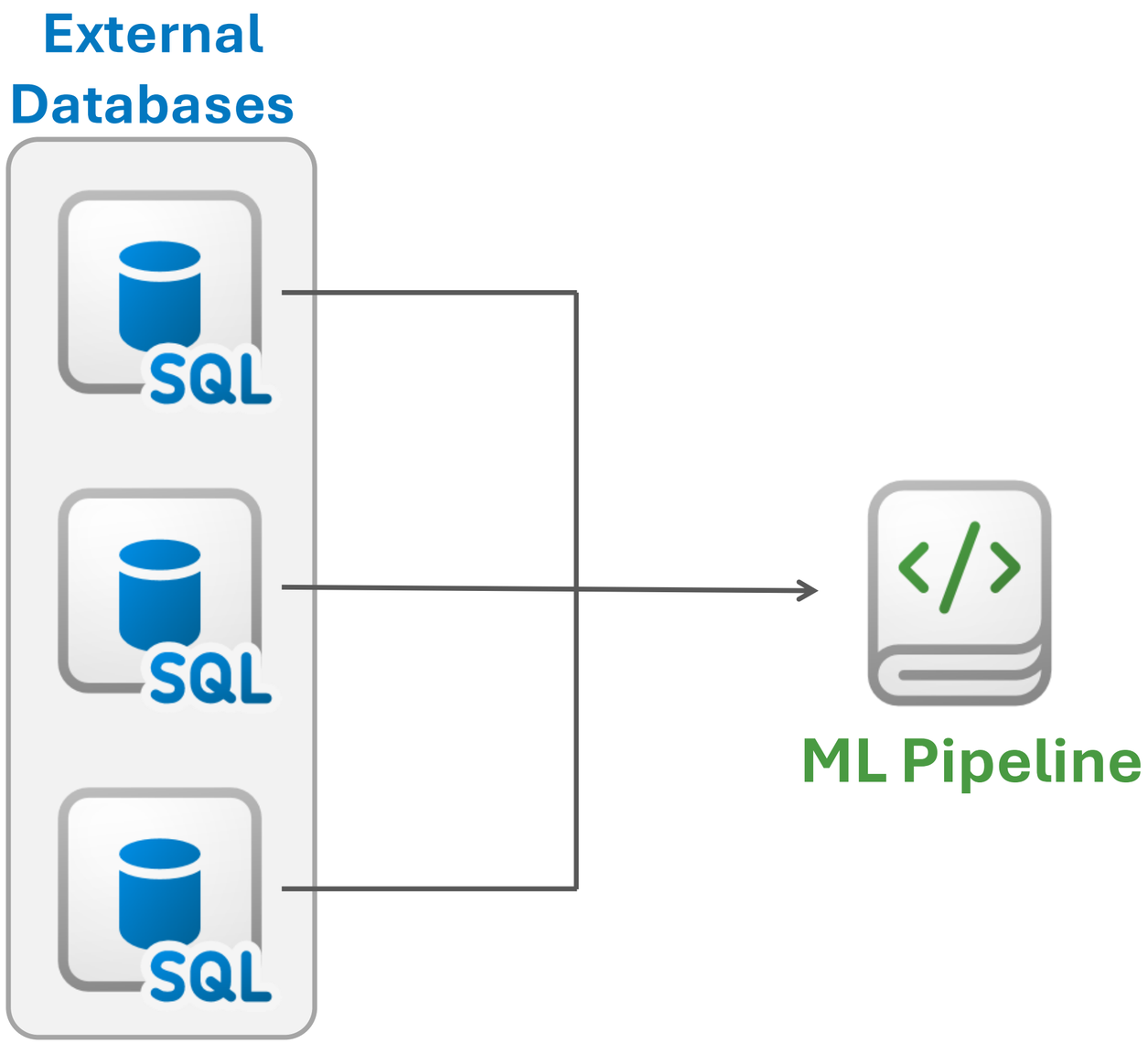
x6 Notebooks

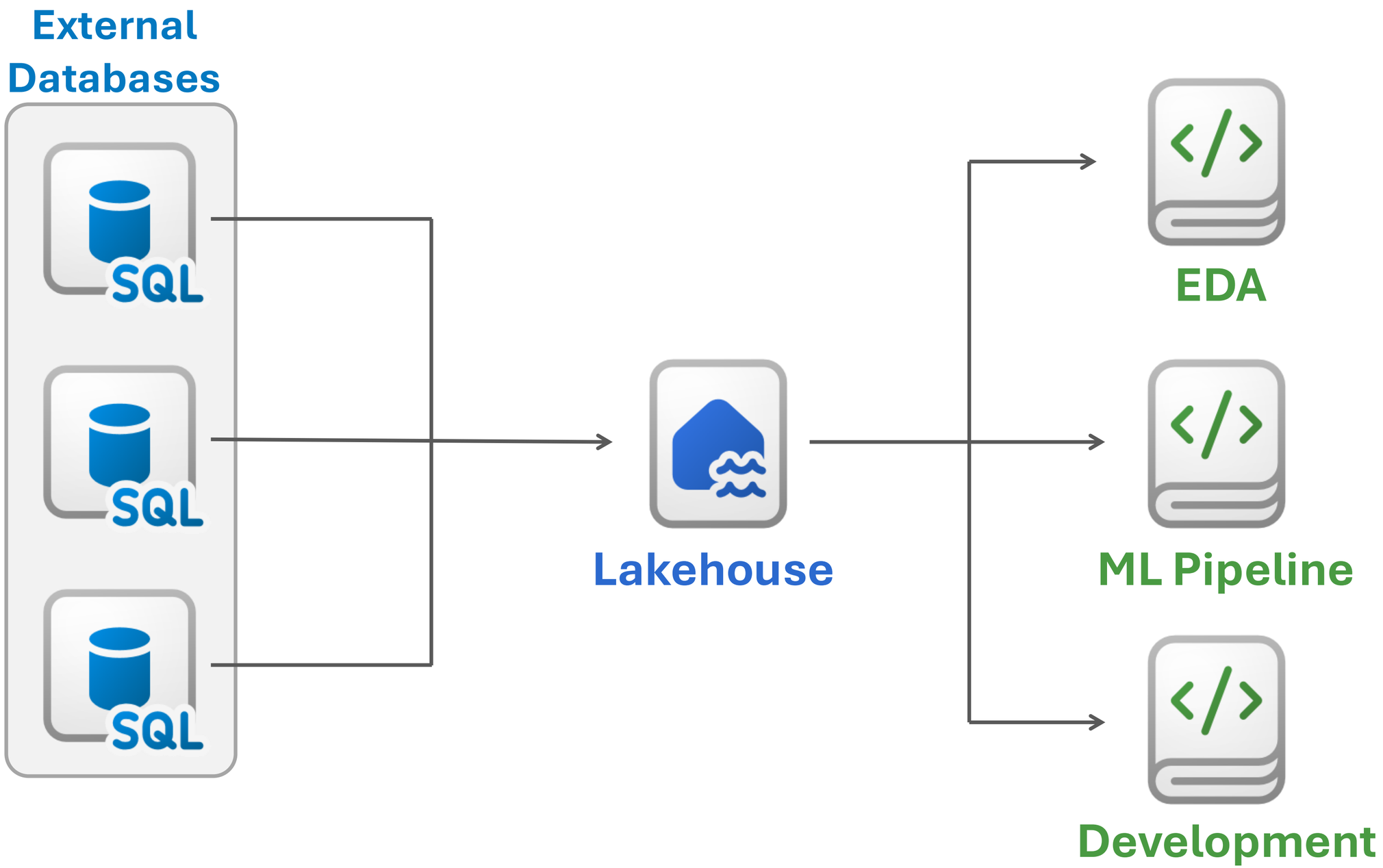
x1 Environment

x2 Experiments

Lakehouse

- Separation of Concerns (SoC)


External Databases
SQL
SQL
SQL
ML Pipeline


External Databases
SQL
SQL
SQL
Lakehouse
EDA
ML Pipeline
Development

Lakehouse

- Separation of Concerns (SoC)
- Reference Tables

11.0101 → Computer Science

CIP Crosswalk

	ABC cip_code	ABC cip_title
614	cip_15_0404	Instrumentation Technology/Technician.
615	cip_15_0405	Robotics Technology/Technician.
616	cip_15_0406	Automation Engineer Technology/Technician.
617	cip_15_0407	Mechatronics, Robotics, and Automation Engineering Te
618	cip_15_0499	Electromechanical Technologies/Technicians, Other.
619	cip_15_05	Environmental Control Technologies/Technicians.
620	cip_15_0501	Heating, Ventilation, Air Conditioning and Refrigeration
621	cip_15_0503	Energy Management and Systems Technology/Technicia
622	cip_15_0505	Solar Energy Technology/Technician.
623	cip_15_0506	Water Quality and Wastewater Treatment Management

Lakehouse

- Separation of Concerns (SoC)
- Reference Tables
- Store Predictions

current_predictions

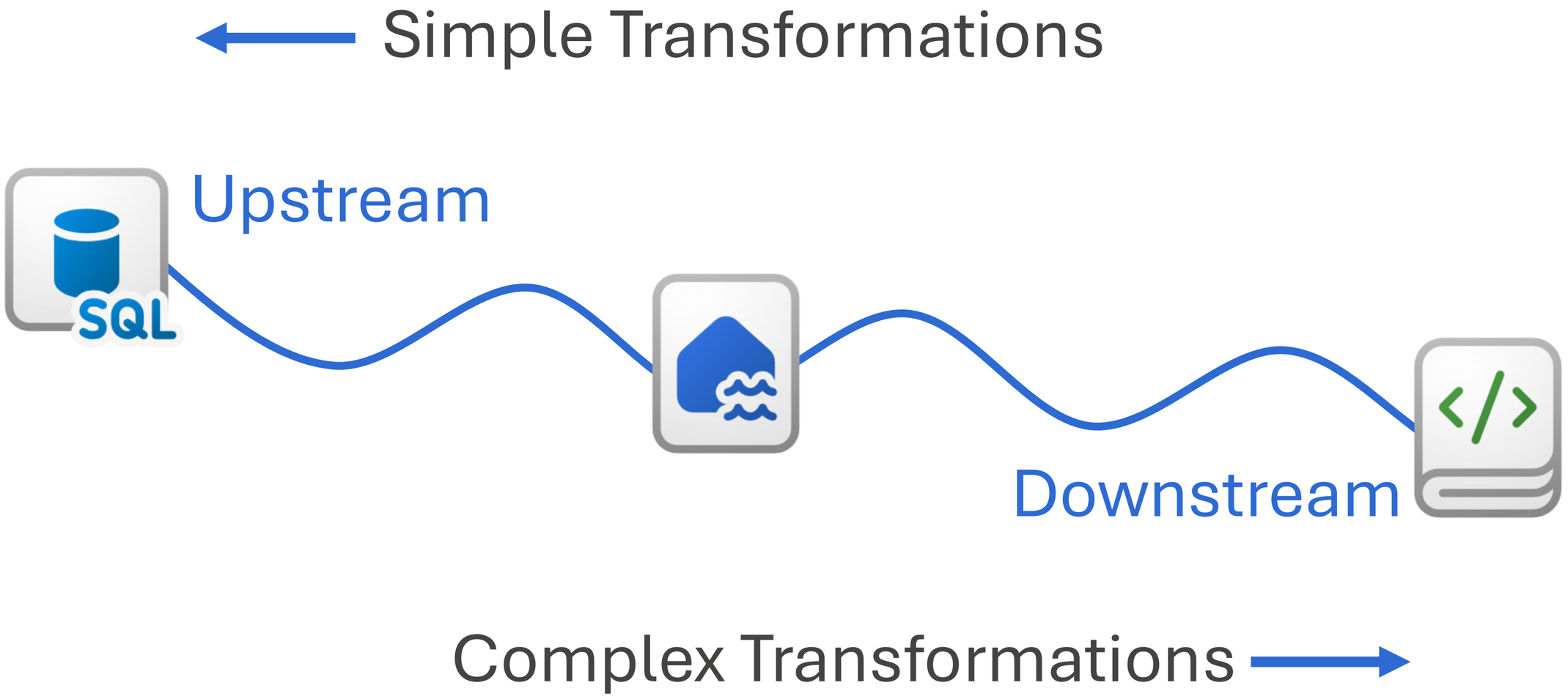
prediction_history

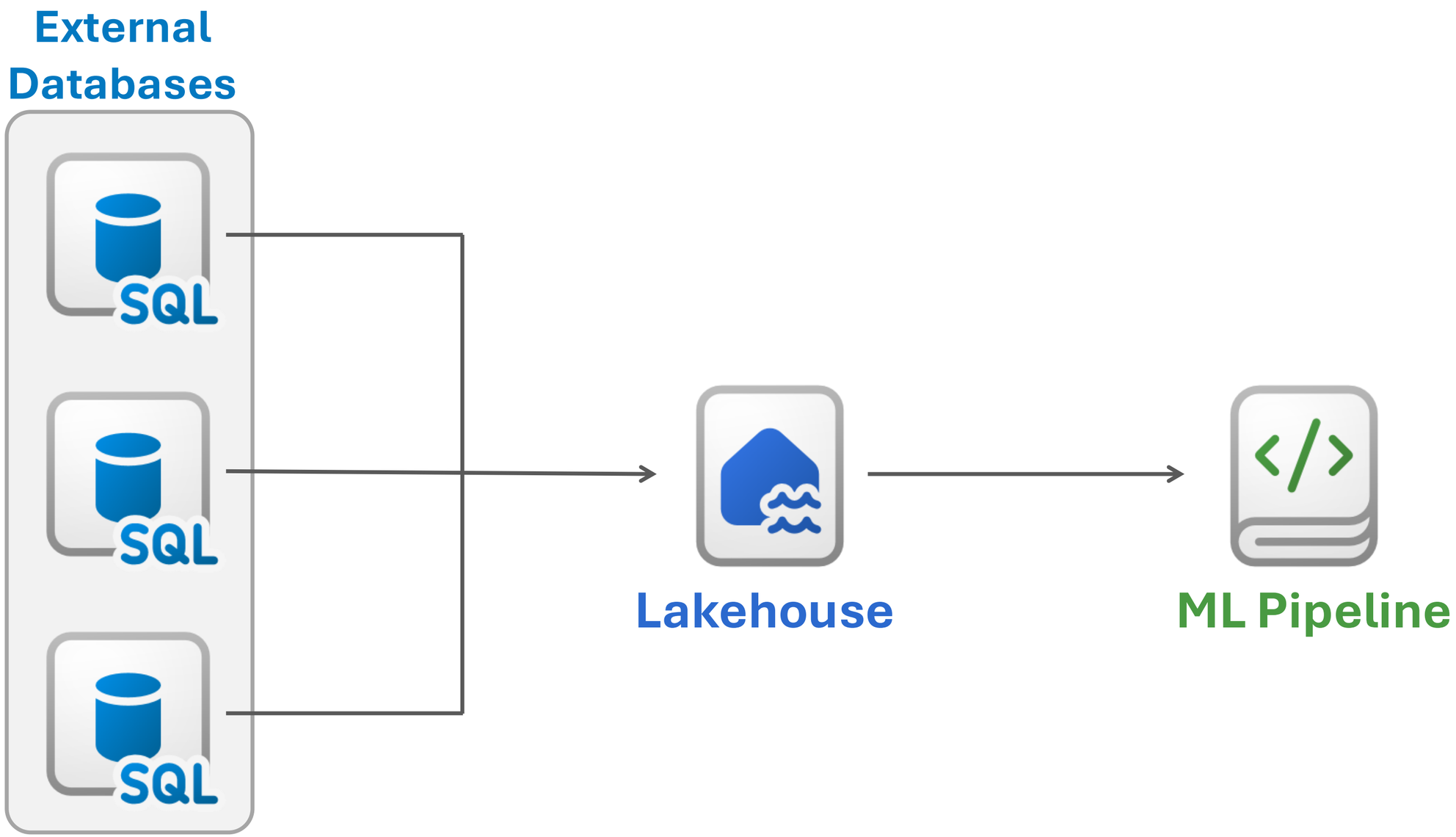
Lakehouse

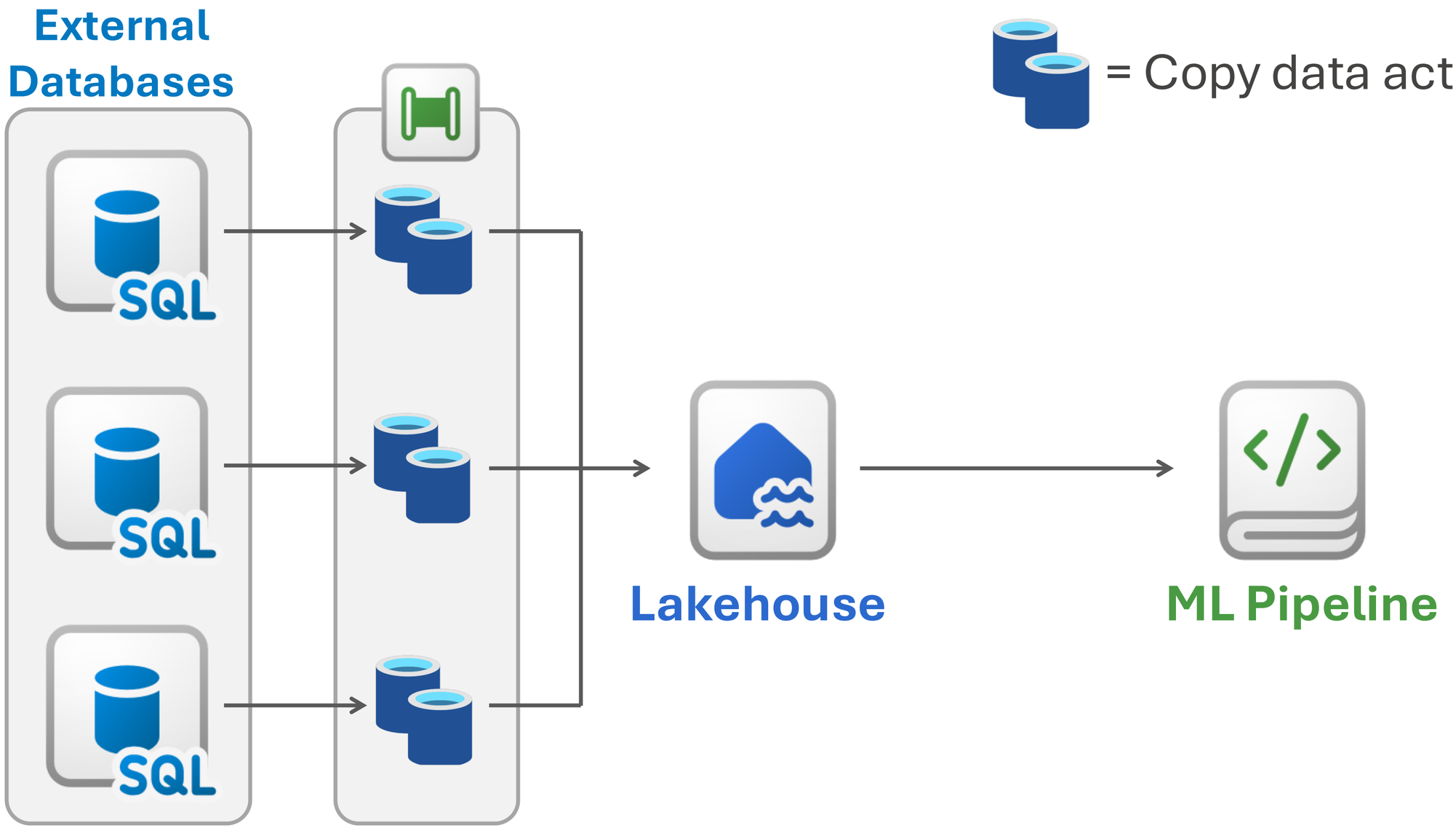
- Separation of Concerns (SoC)
- Reference Tables
- Store Predictions
- Unstructured Data
  - Configuration files
  - CSV files

Data for ML

- Subset of clean data
- Numerous transformations
  - Feature selection
  - Type casting
  - Encoding
  - Feature scaling
  - Imputation


Simple Transformations
Upstream
SQL
Downstream
Complex Transformations

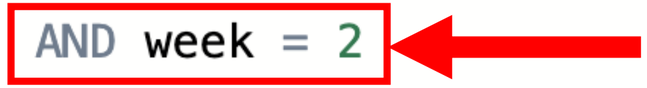

External Databases
SQL
SQL
SQL
Lakehouse
ML Pipeline


External Databases
= Copy data activity
SQL
SQL
SQL
Lakehouse
ML Pipeline

Copy data activity

- Extract External Data
- Custom SQL Query
- String Interpolation

```sql
SELECT
    hs_gpa,
    ap_credits,
    act_math_score,
    cip_code,
    midterm_gpa,
    CASE WHEN first_gen = 'Y' THEN 1 ELSE 0 END AS is_first_gen
FROM student_enrollment
WHERE
    academic_level = 'freshman'
    AND semester = 'fall'
    AND week = 2
    AND academic_load = 'F'
```



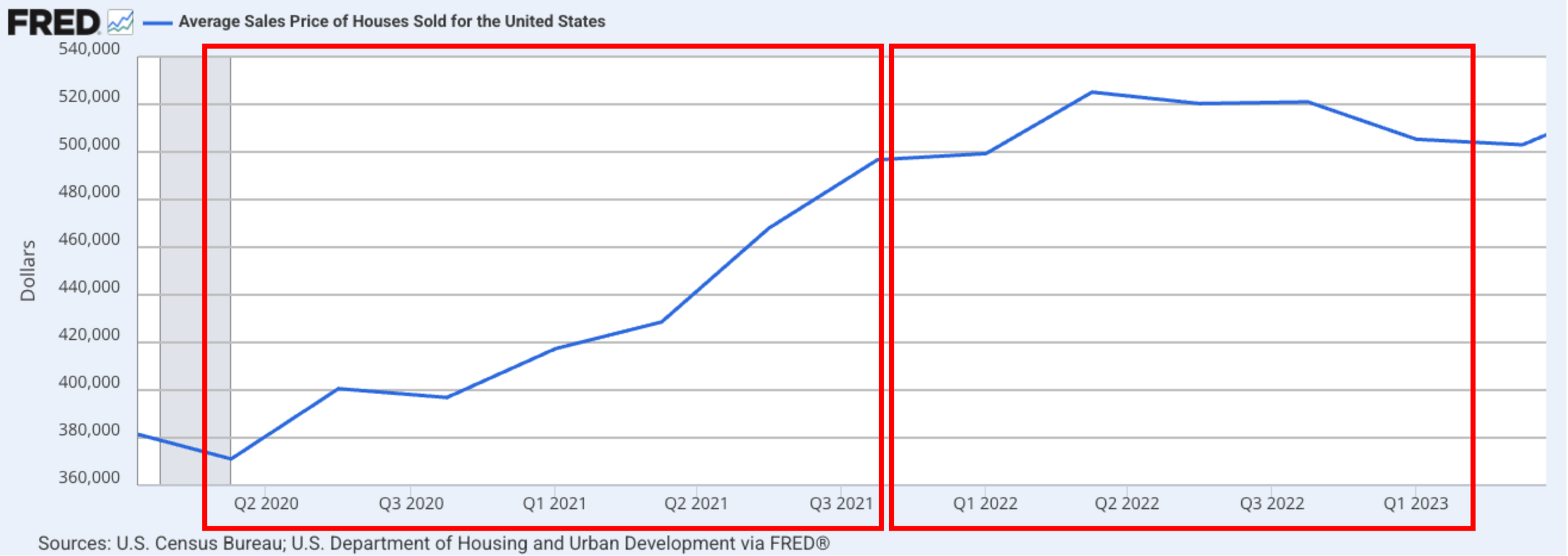
```sql
SELECT
    hs_gpa,
    ap_credits,
    act_math_score,
    cip_code,
    midterm_gpa,
    CASE WHEN first_gen = 'Y' THEN 1 ELSE 0 END AS is_first_gen
FROM student_enrollment
WHERE
    academic_level = 'freshman'
    AND semester = 'fall'
    AND week = '@{variables('week')}'
    AND academic_load = 'F'
```

Copy data activity

- Extract External Data
- Custom SQL Query
- String Interpolation
- Explicit Column Mapping

Inferred Schema		Target Schema	
STUDENT_ID	String ✓	STUDENT_ID	String ✓
ACT_MATH	String ✗	ACT_MATH	Int32 ✗

Zillow®
Zillow Offers


FRED
Average Sales Price of Houses Sold for the United States
Dollars
540,000
520,000
500,000
480,000
460,000
440,000
420,000
400,000
380,000
360,000
Q2 2020
Q3 2020
Q1 2021
Q2 2021
Q3 2021
Q1 2022
Q2 2022
Q3 2022
Q1 2023
Sources: U.S. Census Bureau; U.S. Department of Housing and Urban Development via FRED®

Data Validation

- Many checks per column
- Difficult to maintain

Column Registry

- YAML file
- Column metadata

```yaml
- group name: standarized_test_scores
  columns:
    - name: act_math
      display_name: ACT Math Score
      role: feature
      dtype: numeric
      validation:
        range:
          min: 1
          max: 36
        non_null: false
      drift:
        monitor: true
        method: ks
        warn_threshold: 0.1
        alert_threshold: 0.2
      enabled: true
```

← Structural Validation

← Distributional Validation

Basic Framework

x1 Lakehouse

x1 Pipeline

x6 Notebooks

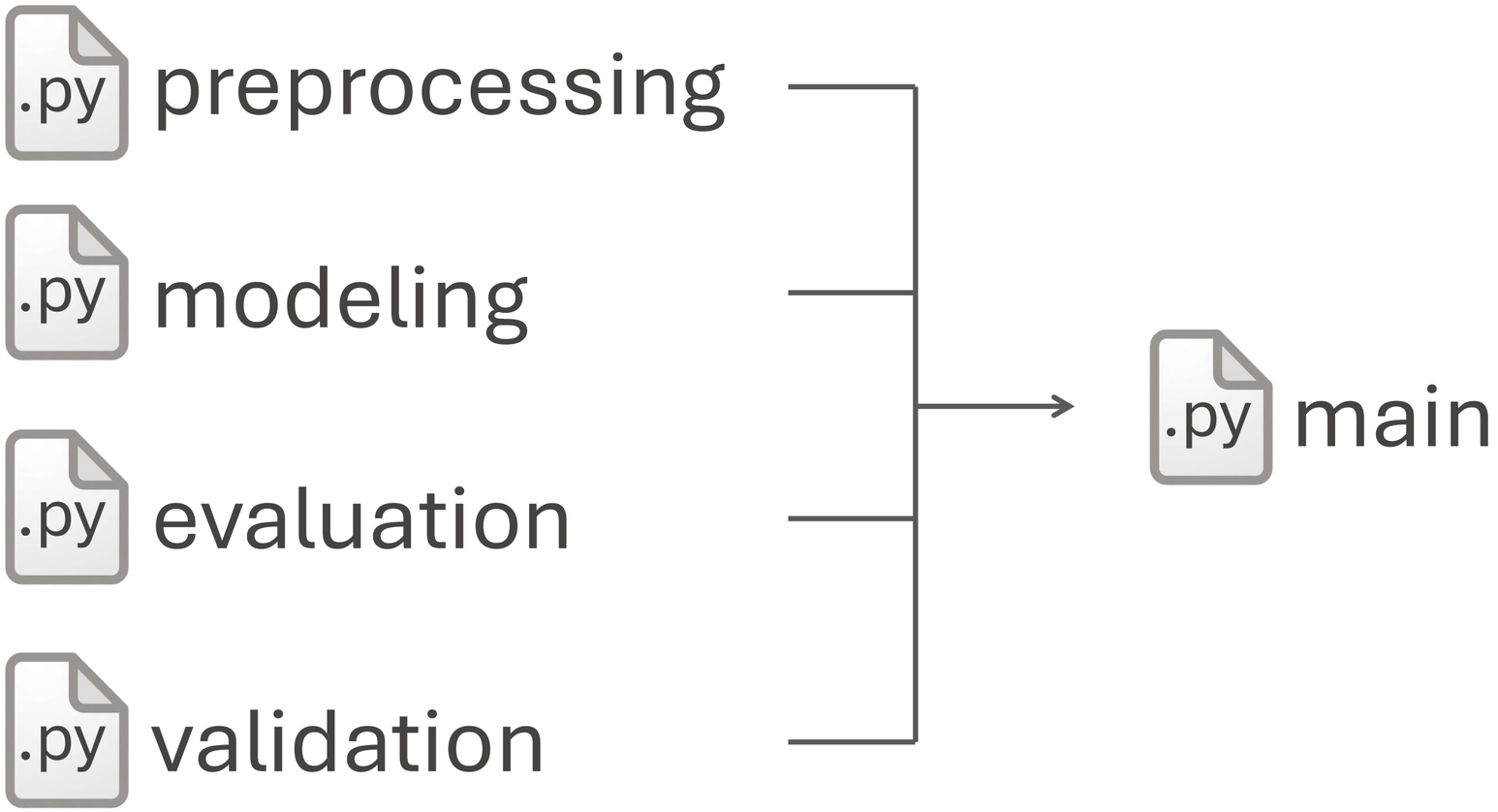
x2 Experiments

x1 Environment

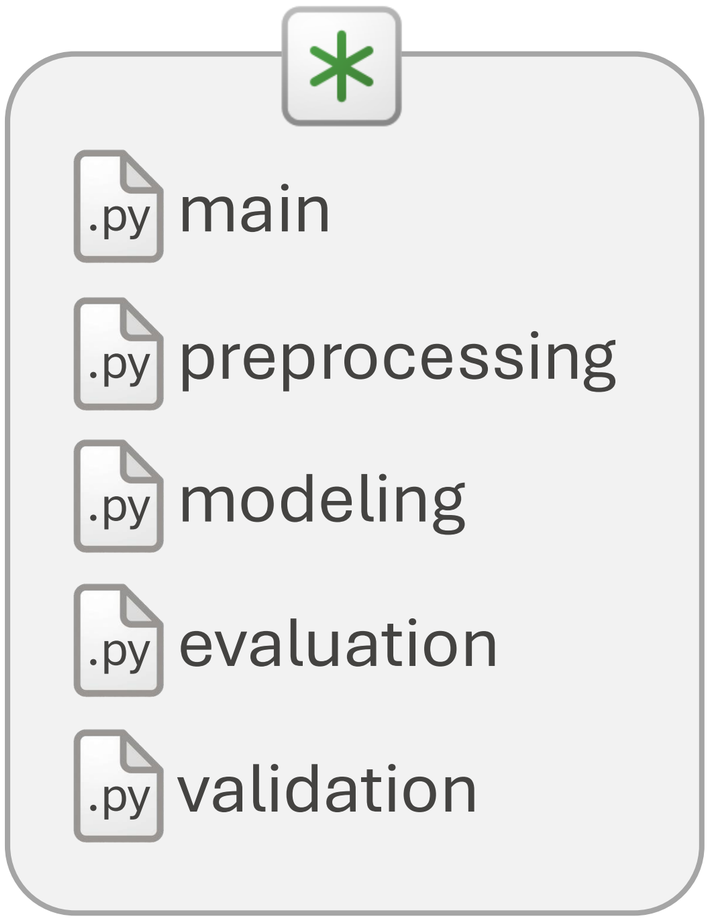
ML Pipeline

- Software
- Python
- Package

ml_pipeline


.py preprocessing
.py modeling
.py evaluation
.py validation
.py main

Spark Job Definition



Notebook

Hybrid

Run | Language: PySpark (Python) | Environment: Workspace default

Main definition file *

Upload or enter the url of a .py file. File size must be less than 250MB.

○ Upload a local file ◉ ADLS-Gen2 path

e.g. "abfss://.../path/to/shakespeare.txt abfss://.../path/to/result"

.py main

Reference file

Upload or enter the urls of one or multiple files. File size must be less than 250MB.

○ Upload local files ◉ ADLS-Gen2 path

Each line should be a storage URI. e.g.
"abfss://.../path/to/shakespeare.txt abfss://.../path/to/result"

.py preprocessing

.py modeling

.py evaluation

.py validation

Command line arguments

e.g. "abfss://.../path/to/shakespeare.txt abfss://.../path/to/resu

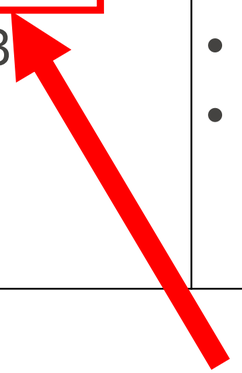
Lakehouse Reference *

+ Add

All lakehouses (0)

Environment (Default)

Runtime	Libraries	Compute
• OS Mariner 2.0 • Apache Spark 3.5 • Delta Lake 3.2 • Python 3.11	• scikit-learn 1.2.2 • xgboost 2.0.3 • shap 0.42.1	• Node size: medium • # of Nodes: 1-10 • Autoscale: enabled

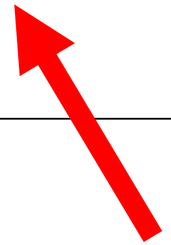


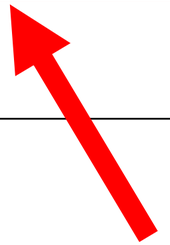
Machine Learning Libraries

- Single-node libraries
  - Scikit-Learn
  - XGBoost
- Multi-node libraries
  - Synapse ML
  - MLlib

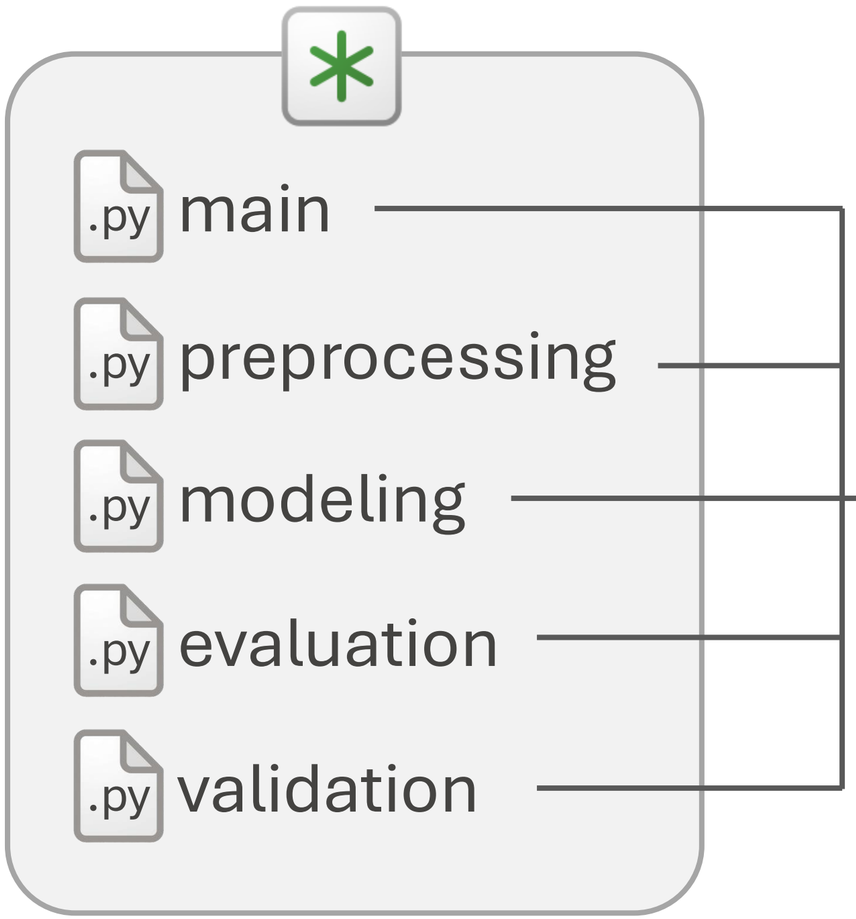
Environment (Custom)

Runtime	Libraries	Compute
• OS Mariner 2.0 • Apache Spark 3.5 • Delta Lake 3.2 • Python 3.11	• scikit-learn 1.6.0 • xgboost 3.0.0 • shap 0.48.0	• Node size: large • # of Nodes: 1 • Autoscale: disabled



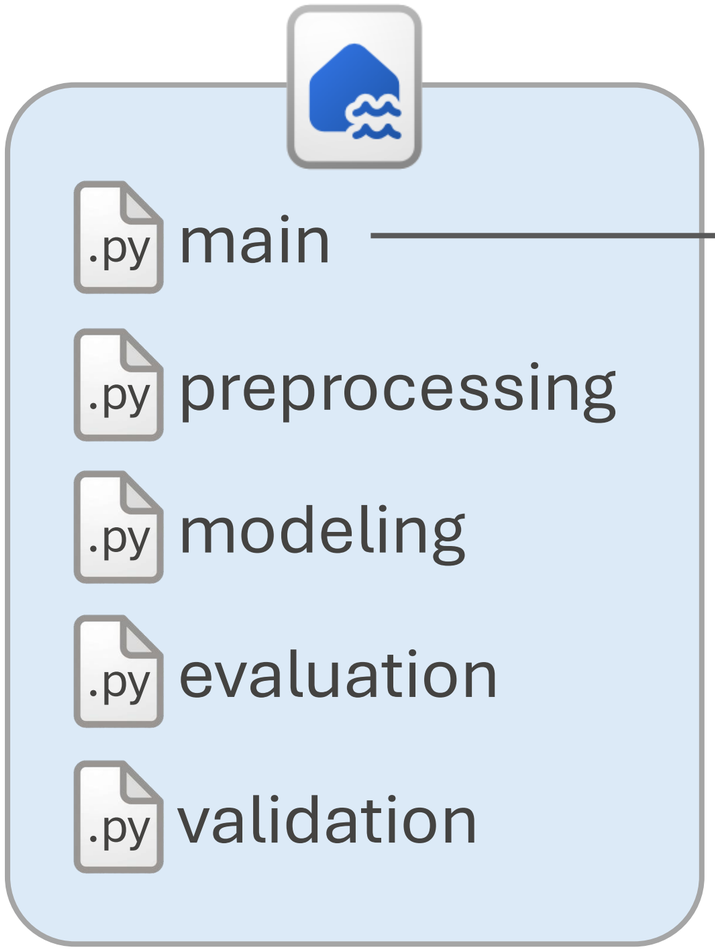


Spark Job Definition



Environment (Default)

Runtime	Libraries	Compute
• OS Mariner 2.0 • Apache Spark 3.5 • Delta Lake 3.2 • Python 3.11	• scikit-learn 1.2.2 • xgboost 2.0.3 • shap 0.42.1	• Node size: medium • # of Nodes: 1-10 • Autoscale: enabled



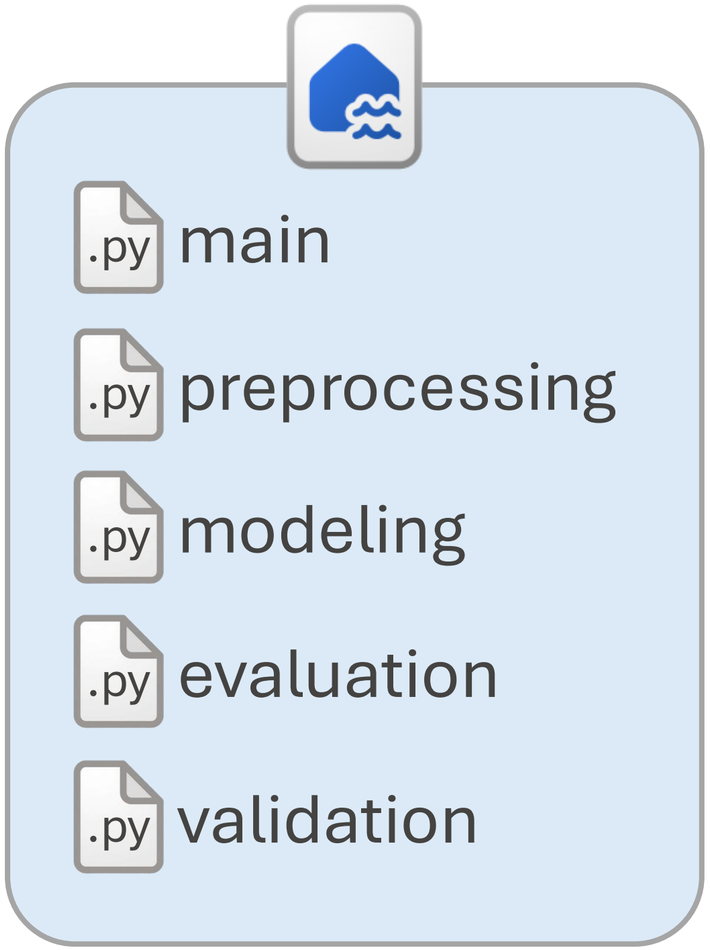

```python
import sys
import os

# Add directory to your Python path
sys.path.append("/lakehouse/default/Files/")
```

PySpark (Python)

```
%load file:/lakehouse/default/Files/main.py
```

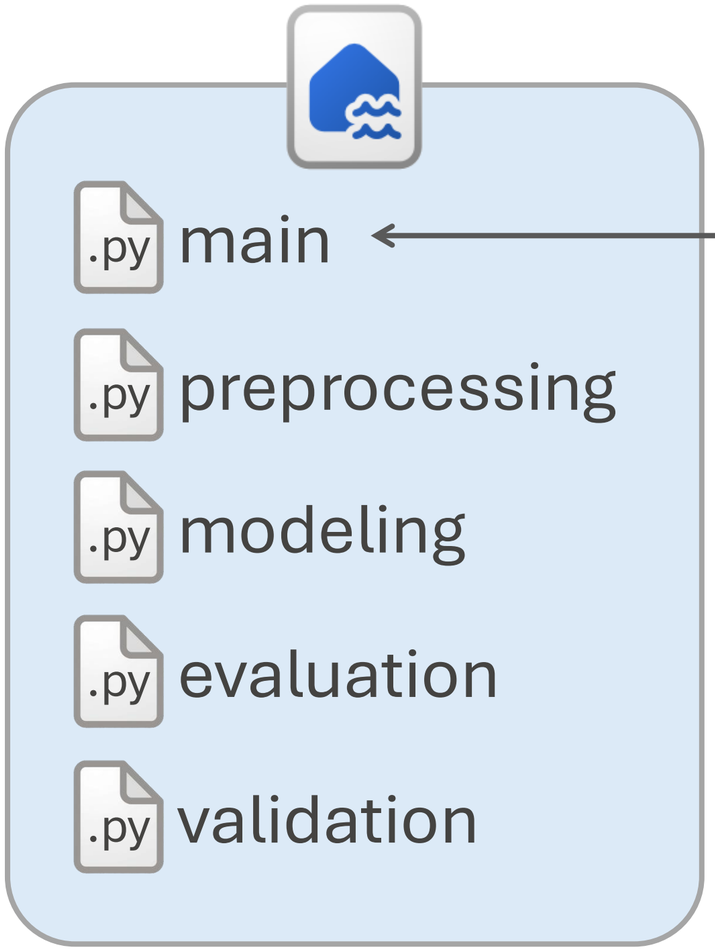
PySpark (Python)




```
1   import sys
2   import os
3
4   # Add directory to your Python path
5   sys.path.append("/lakehouse/default/Files/")
```

PySpark (Python)

```
1   # %load file:/lakehouse/default/Files/main.py
2   import modeling
3   def main():
4       spark = SparkSession.builder.getOrCreate()
```



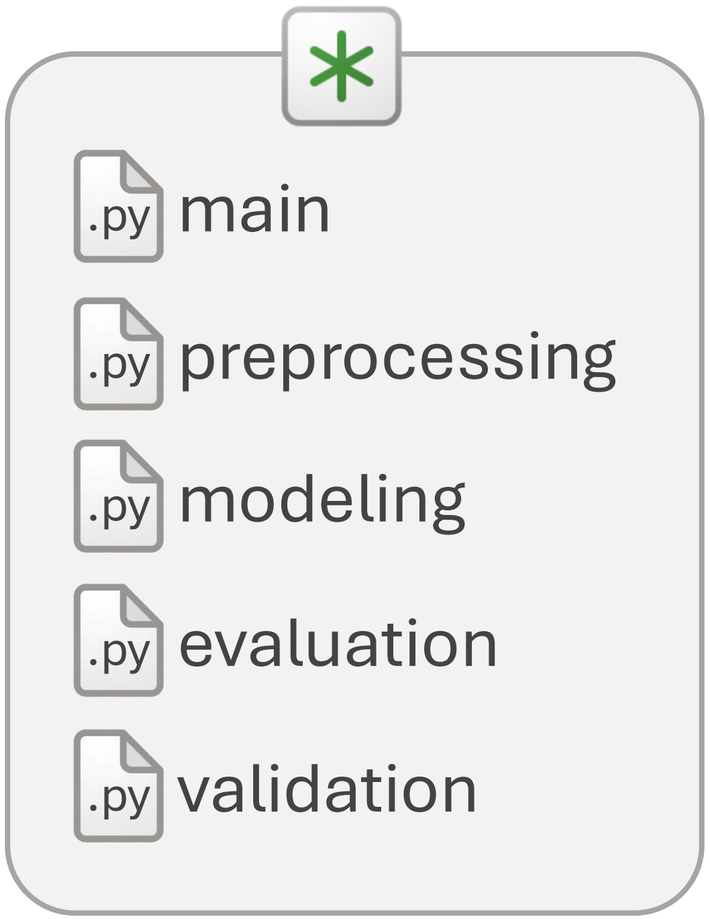

```
1  import sys
2  import os
3
4  # Add directory to your Python path
5  sys.path.append("/lakehouse/default/Files/")
```

PySpark (Python)

```
1  %%writefile file:/lakehouse/default/Files/main.py
2  import modeling
3  def main():
4      spark = SparkSession.builder.getOrCreate()
5      print('Hello, Fabcon!')
```

This will create a .crc file in your lakehouse

Spark Job Definition



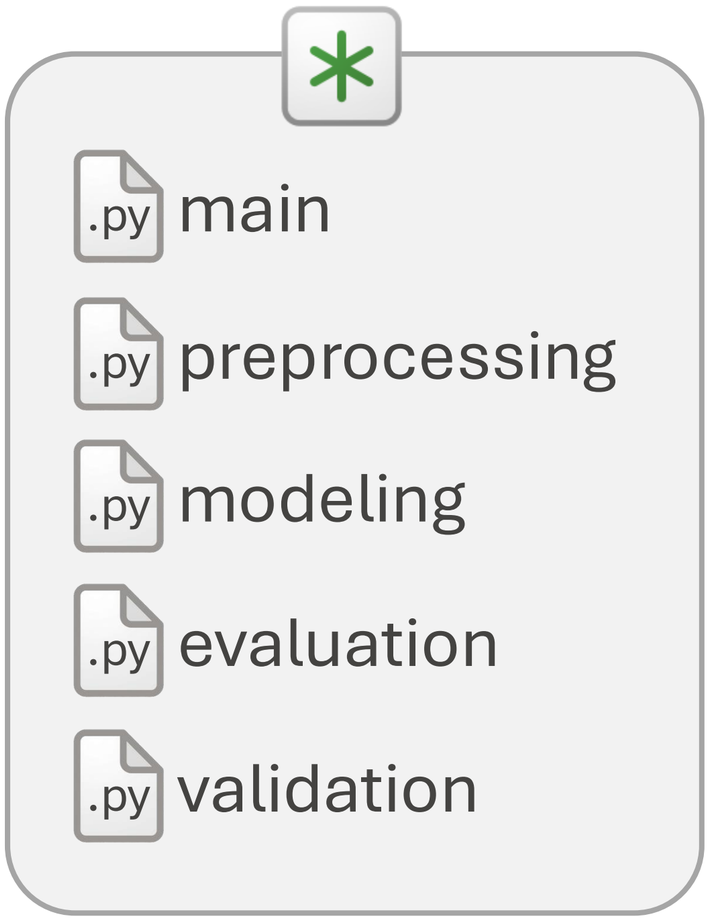
Notebook

Hybrid

Notebook Approach

- Notebooks in production
- “Primary coding item”
- Convenient
- Use %run to import modules into main

Spark Job Definition



Notebook

Hybrid

Hybrid Approach

- Best of both worlds
- Leverage agentic coding tools
  - Claude Code
  - GitHub Copilot
- Perform better on .py files

Experiment Tracking

- Train/test multiple different models
- Analyze each “run”
- Compare different runs

XGB Run

```
with mlflow.start_run():
    model.fit()
    mlflow.log_metrics()
    mlflow.log_params()
    mlflow.set_tags()
    mlflow.log_model()
```

Experiment

Run name	Start time	Duration	Accuracy
RF Run	3/19/26	3:13	0.76
XGB Run	3/18/26	4:01	0.78

XGB Run

```
with mlflow.start_run():
    model.fit()
    mlflow.log_metrics()
    mlflow.log_params()
    mlflow.set_tags()
    mlflow.log_model()
```

RF Run

```
with mlflow.start_run():
    model.fit()
    mlflow.log_metrics()
    mlflow.log_params()
    mlflow.set_tags()
    mlflow.log_model()
```

Experiment

Run name	Start time	Duration	Accuracy
XGB Run	3/20/26	4:25	0.81
RF Run	3/19/26	3:13	0.76
XGB Run	3/18/26	4:01	0.78

XGB Run

```
with mlflow.start_run():
    model.fit()
    mlflow.log_metrics()
    mlflow.log_params()
    mlflow.set_tags()
    mlflow.log_model()
```

RF Run

```
with mlflow.start_run():
    model.fit()
    mlflow.log_metrics()
    mlflow.log_params()
    mlflow.set_tags()
    mlflow.log_model()
```

Experiment

Run name	Start time	Duration	Accuracy
RF Run	3/20/26	3:11	0.79
XGB Run	3/20/26	4:25	0.81
RF Run	3/19/26	3:13	0.76
XGB Run	3/18/26	4:01	0.78

XGB Run

```
with mlflow.start_run():
    model.fit()
    mlflow.log_metrics()
    mlflow.log_params()
    mlflow.set_tags()
    mlflow.log_model()
```

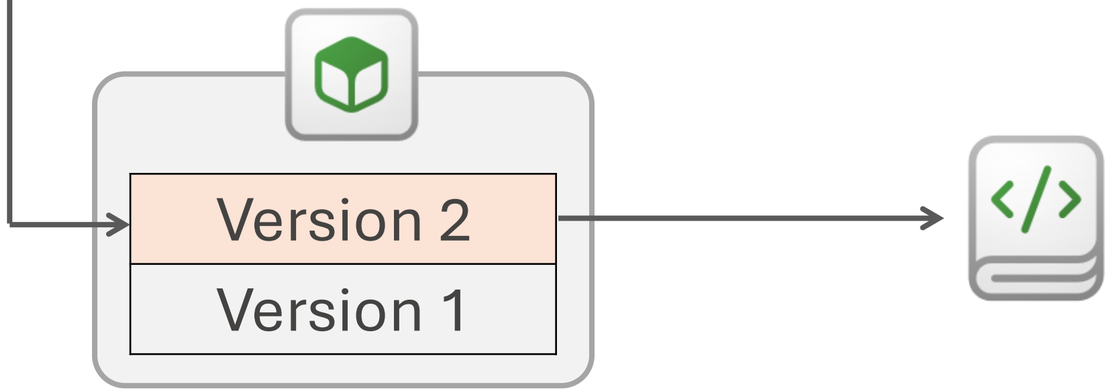
RF Run

```
with mlflow.start_run():
    model.fit()
    mlflow.log_metrics()
    mlflow.log_params()
    mlflow.set_tags()
    mlflow.log_model()
```

Experiment

Run name	Start time	Duration	Accuracy
RF Run	3/20/26	3:11	0.79
XGB Run	3/20/26	4:25	0.81
RF Run	3/19/26	3:13	0.76
XGB Run	3/18/26	4:01	0.78

Model



Basic Framework

x1 Lakehouse

x1 Pipeline

x6 Notebooks

x1 Environment

x2 Experiments

External Database

Custom Environment

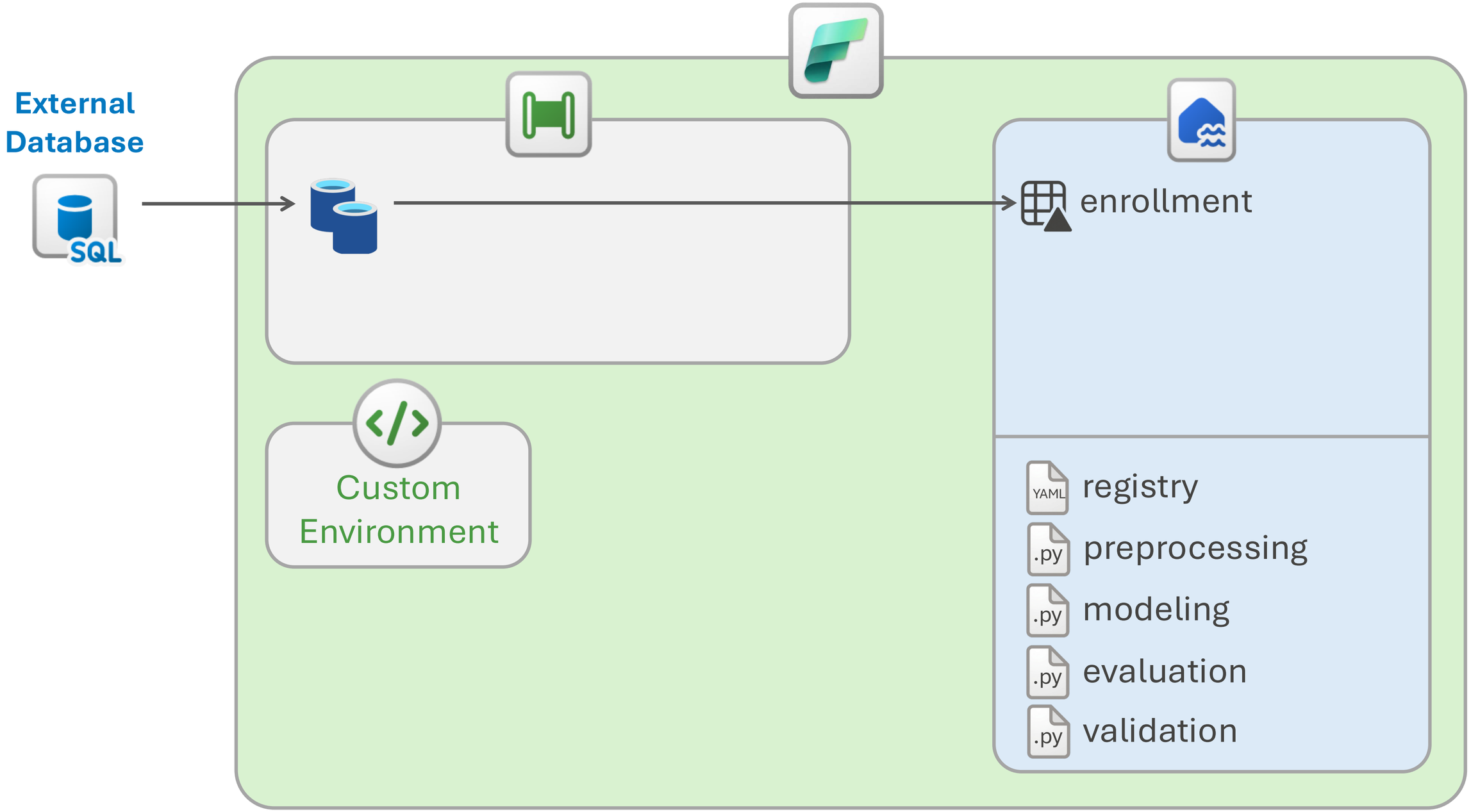
registry

preprocessing

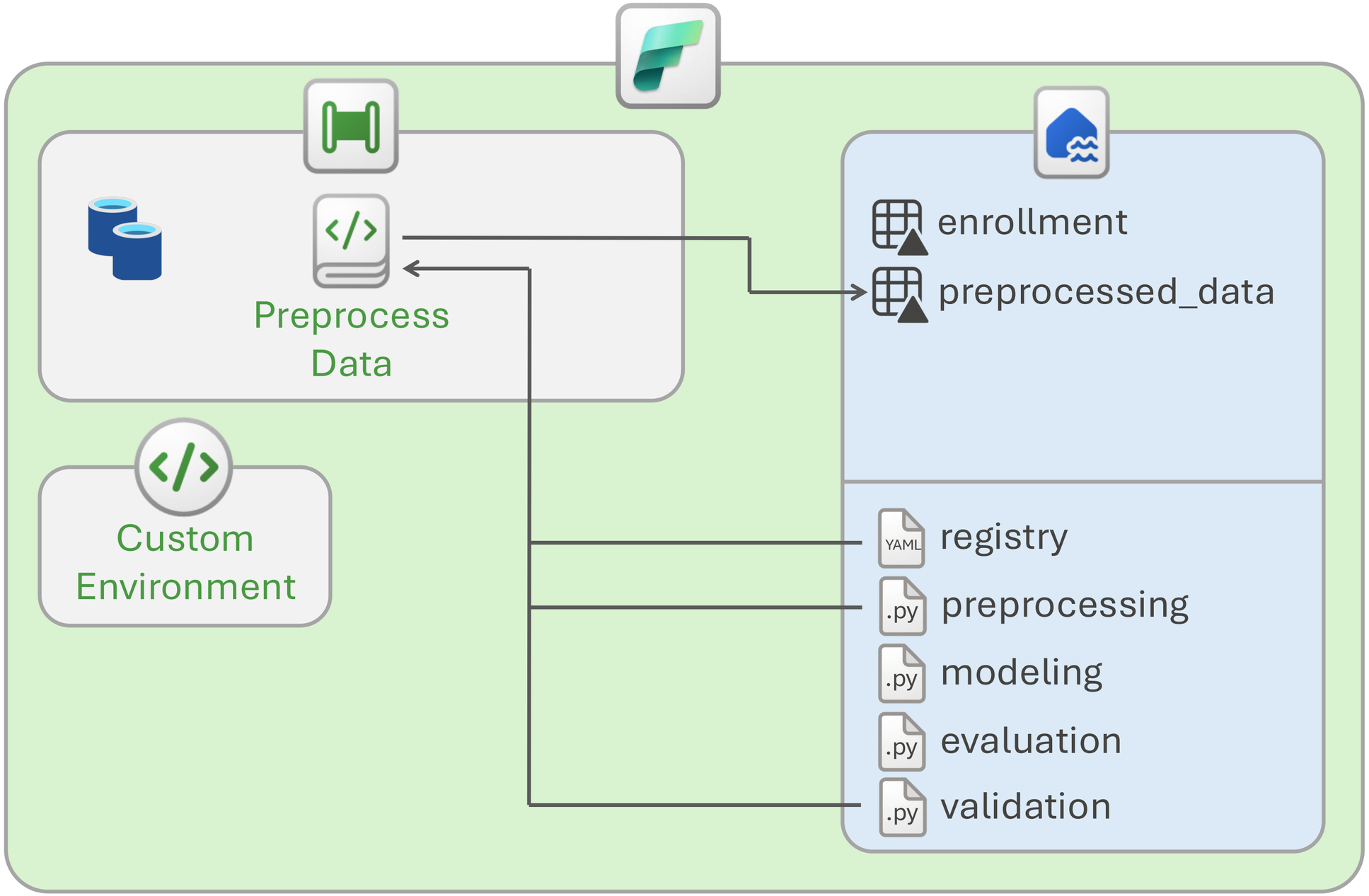
modeling

evaluation

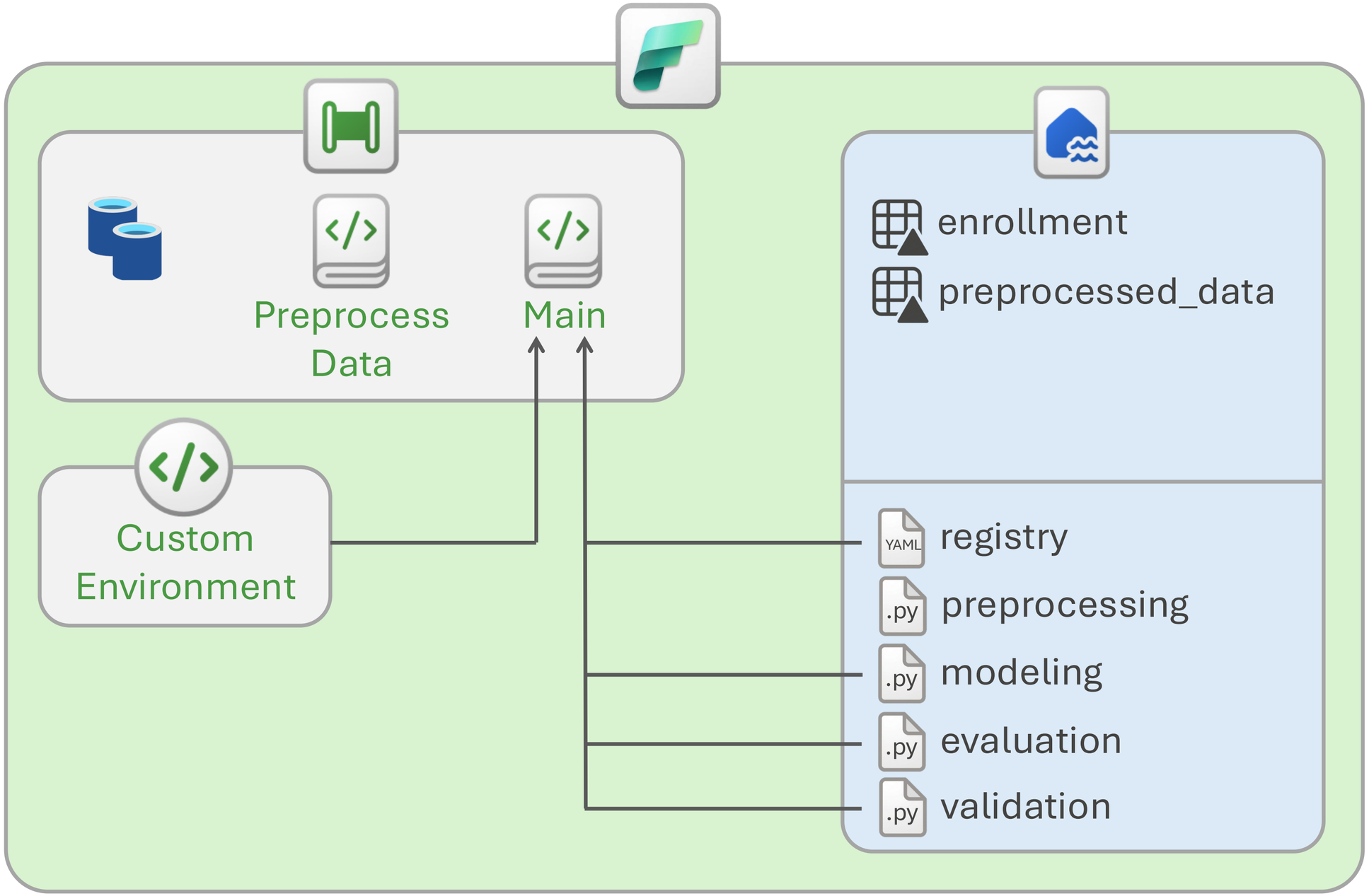
validation


External Database
SQL
enrollment
Custom Environment
YAML
registry
.py
preprocessing
.py
modeling
.py
evaluation
.py
validation

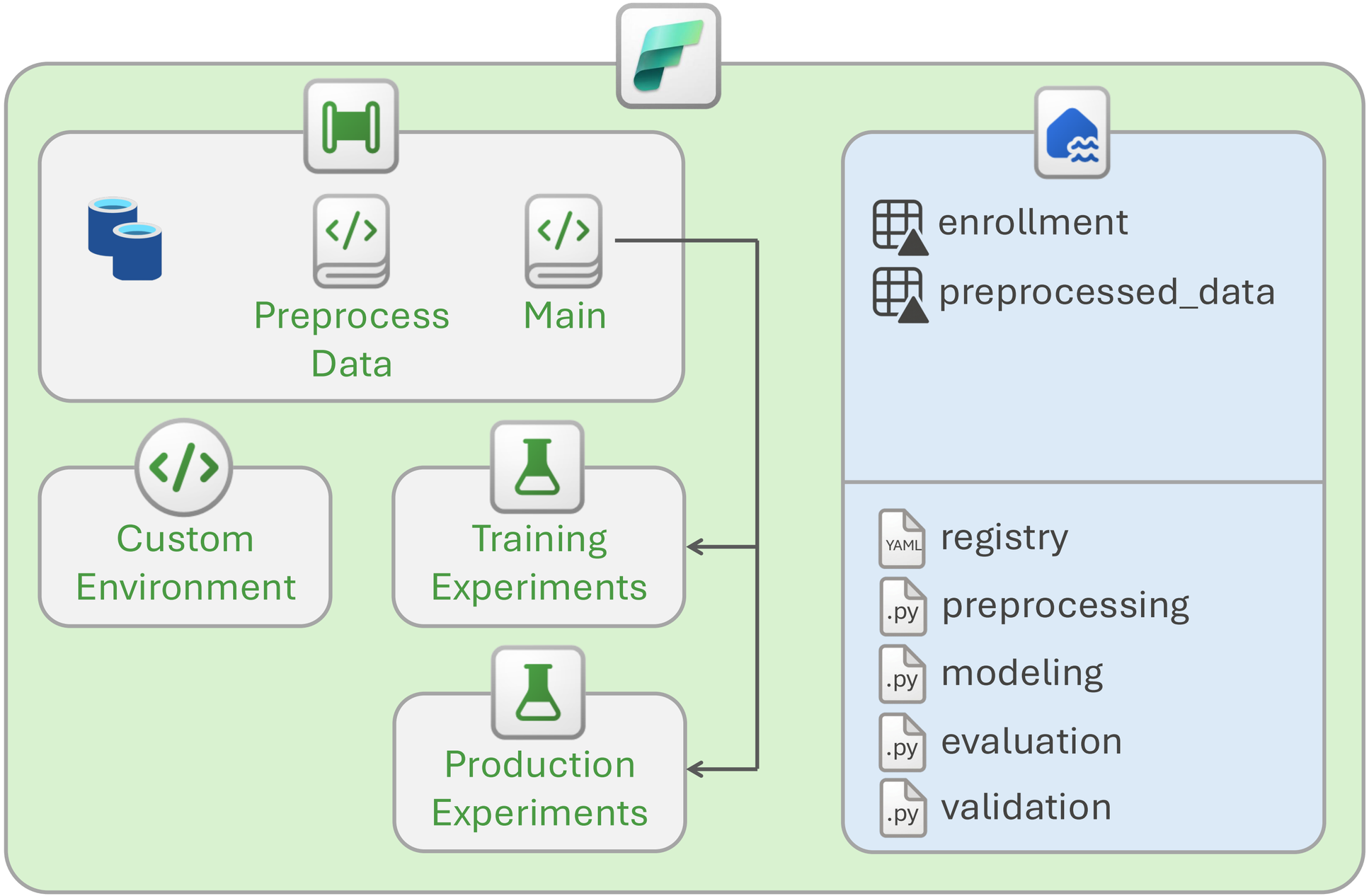
External Database



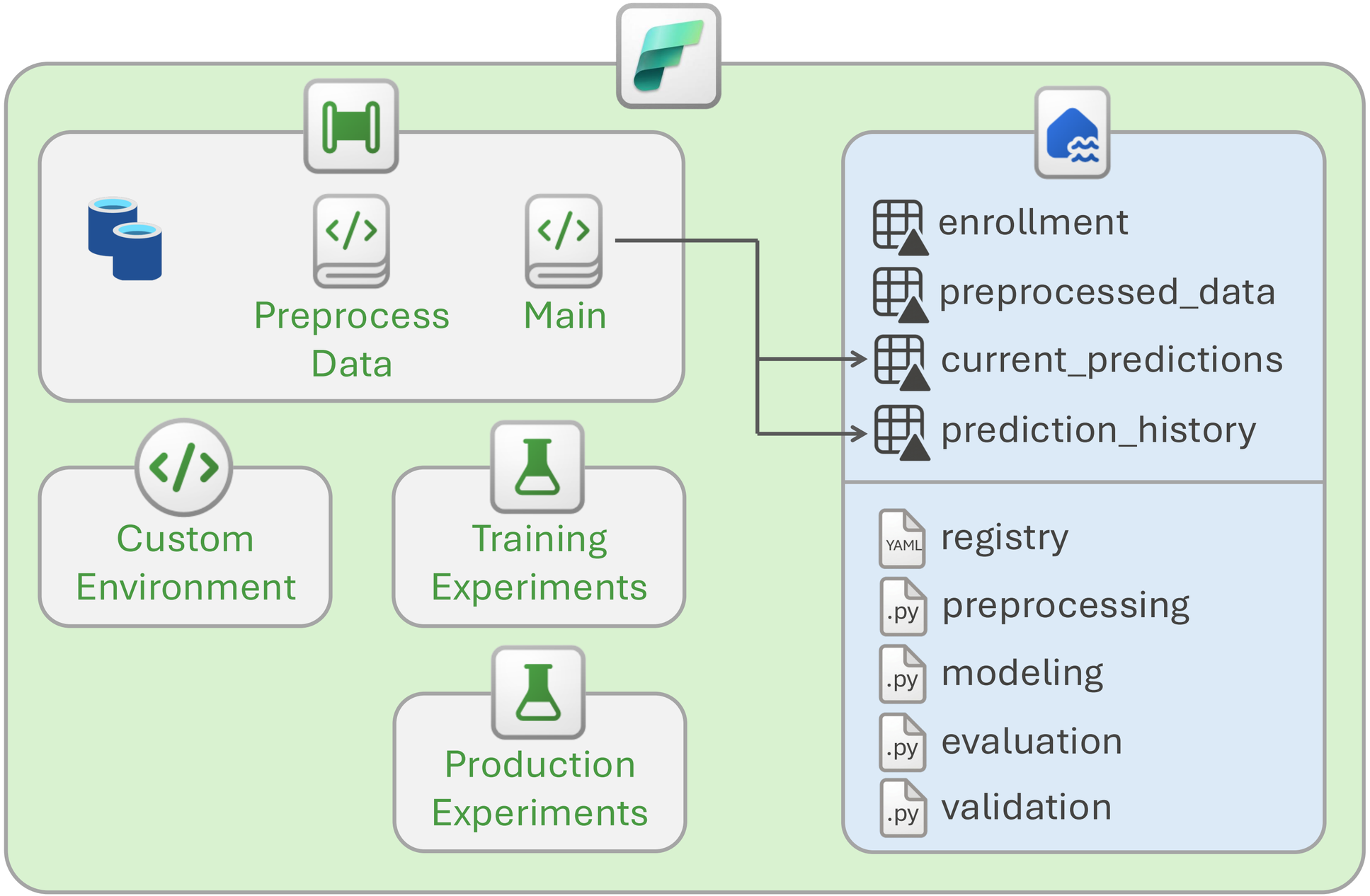
External Database



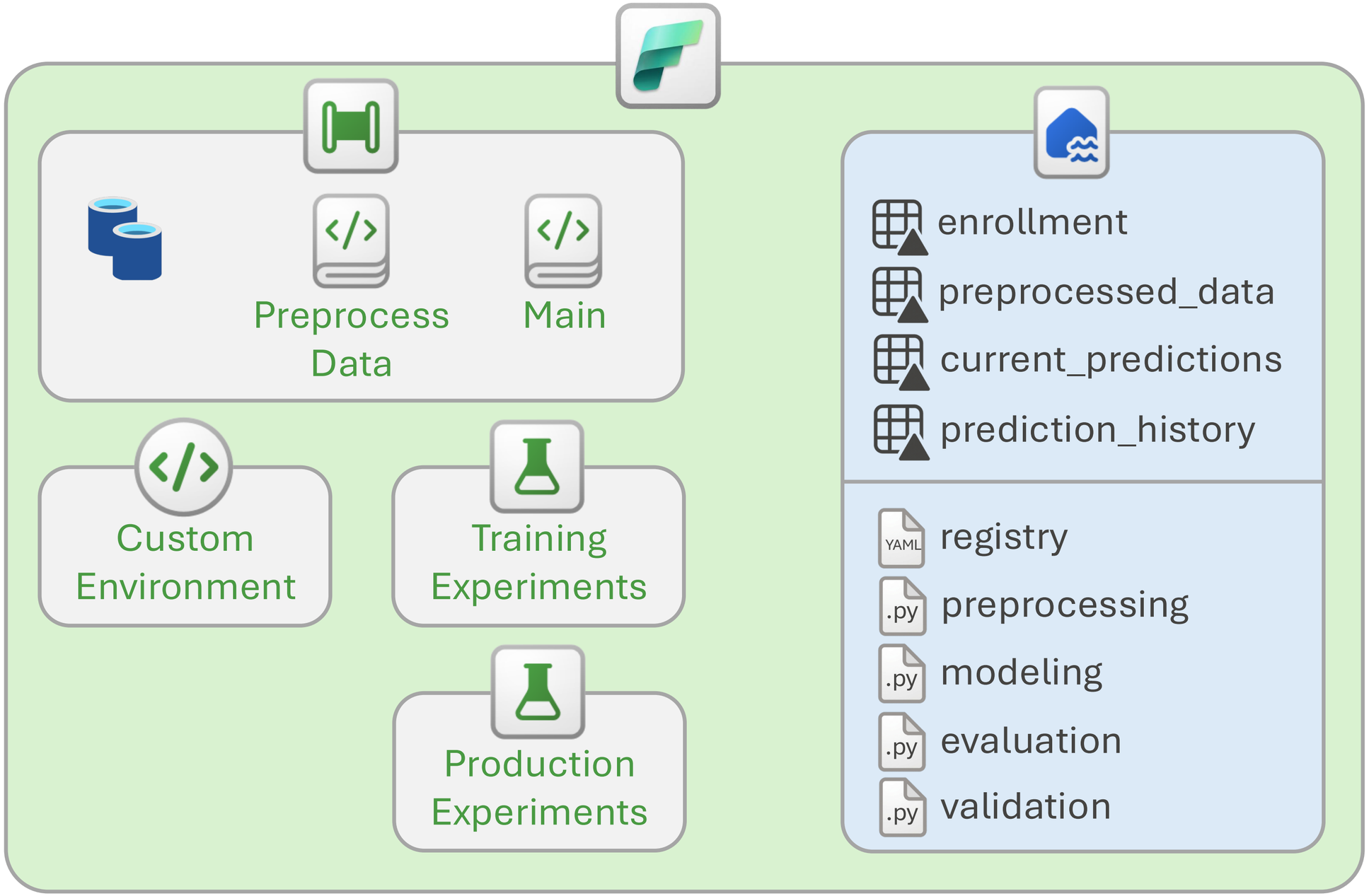
External Database

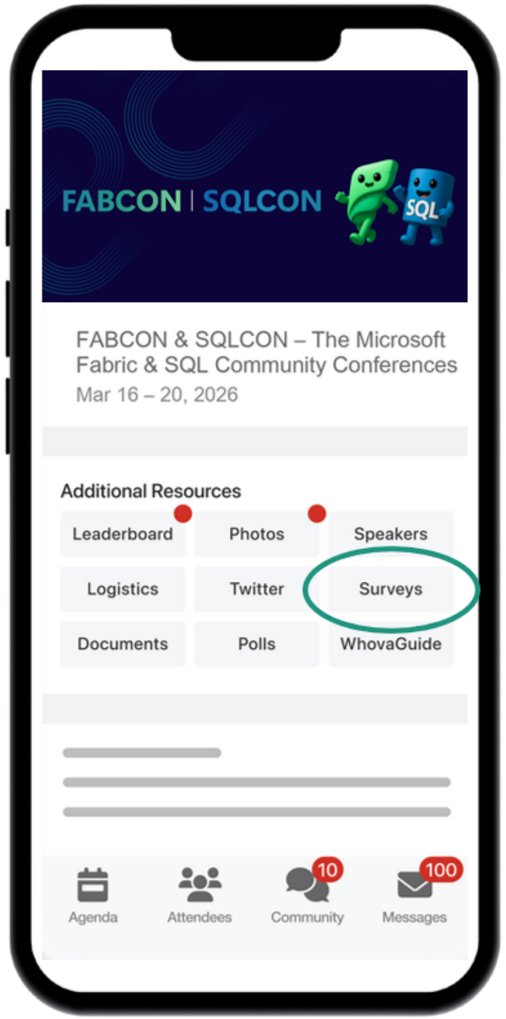


External Database



External Database



How was the session?

Complete Session Surveys in Whova for your chance to WIN PRIZES!