Understanding the problem(s) Fabric solves

The story of Houston Electrics Inc

Houston
Electrics

Understanding the problem(s) Fabric solves

The story of Houston Electrics Inc

Houston Electrics

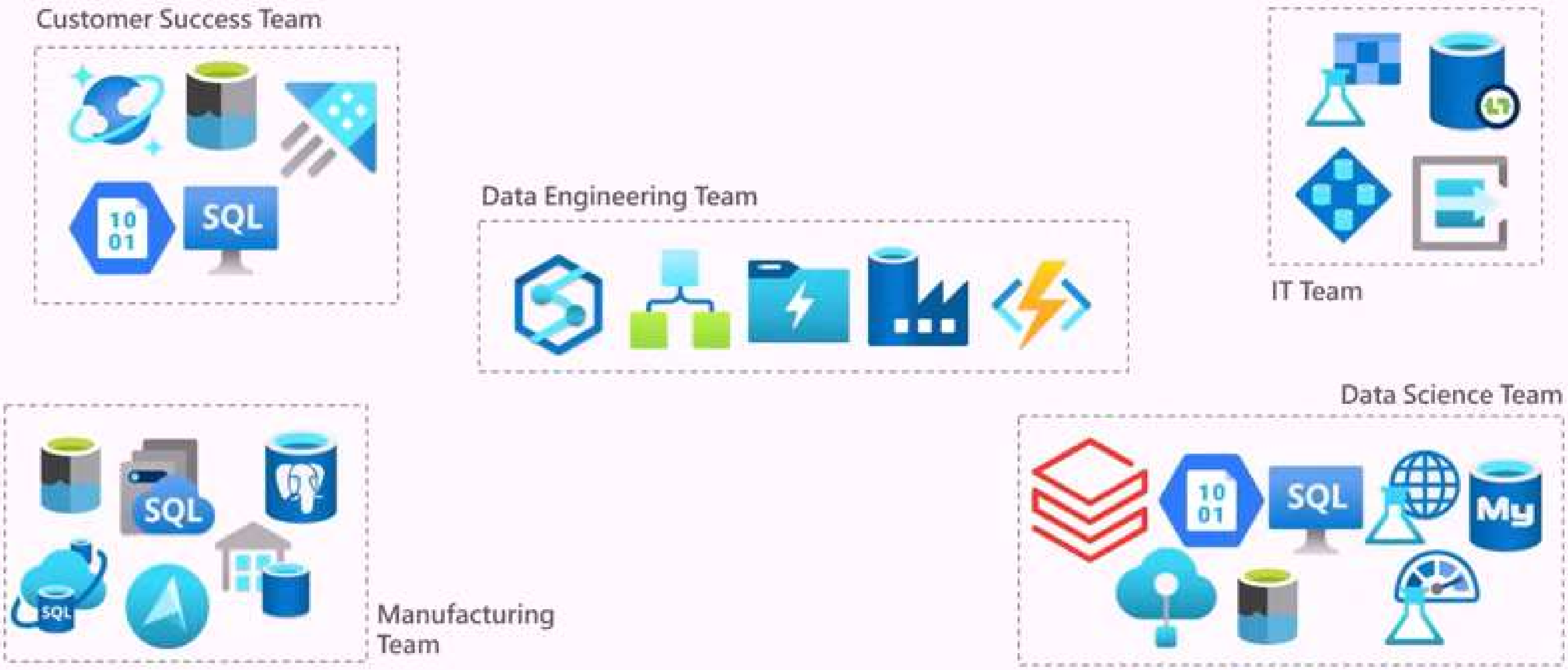
Based in the USA, ships electrical parts around the world

400+ employees

Innovative and data-driven

- 2015: Digital transformation programme to drive growth
    - Central data department with 50 employees:
    - Investment in Azure, AWS and Power BI for reporting

Existing tools @ Houston Electrics



Customer Success Team

Data Engineering Team

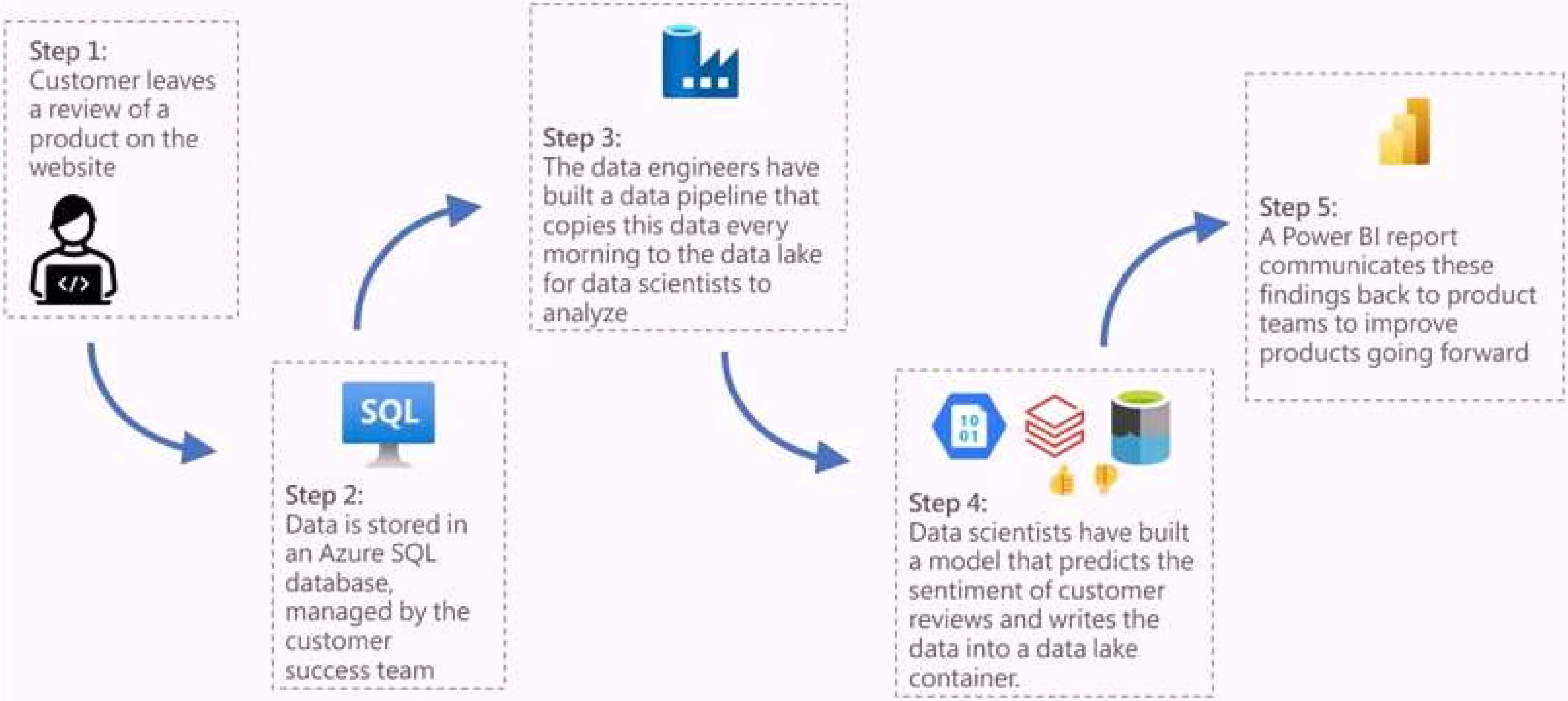
IT Team

Data Science Team

Manufacturing Team

Existing customer reviews workflow

Houston Electrics



Step 1:
Customer leaves a review of a product on the website

Step 2:
Data is stored in an Azure SQL database, managed by the customer success team

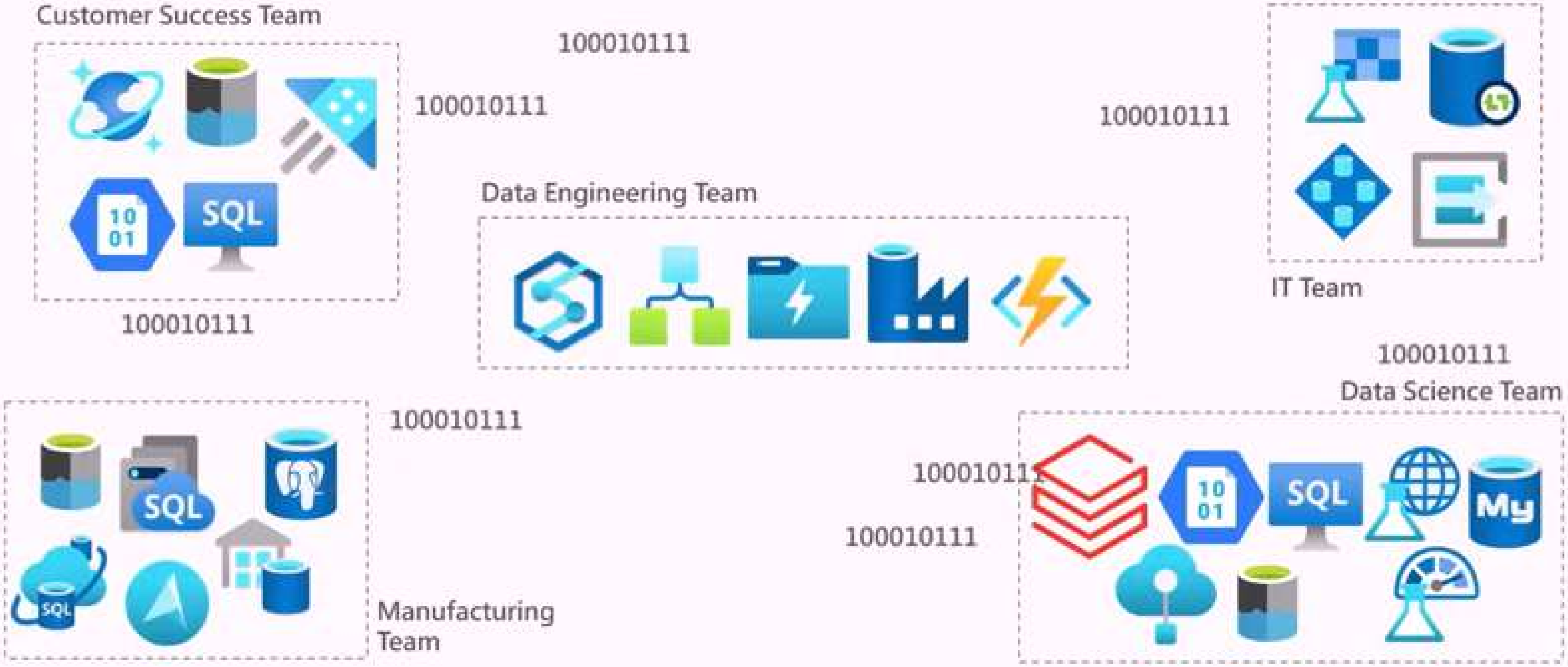
Step 3:
The data engineers have built a data pipeline that copies this data every morning to the data lake for data scientists to analyze

Step 4:
Data scientists have built a model that predicts the sentiment of customer reviews and writes the data into a data lake container.

Step 5:
A Power BI report communicates these findings back to product teams to improve products going forward

Existing tools @ Houston Electrics

Houston Electrics



Customer Success Team

100010111

100010111

100010111

Data Engineering Team

IT Team

100010111

100010111

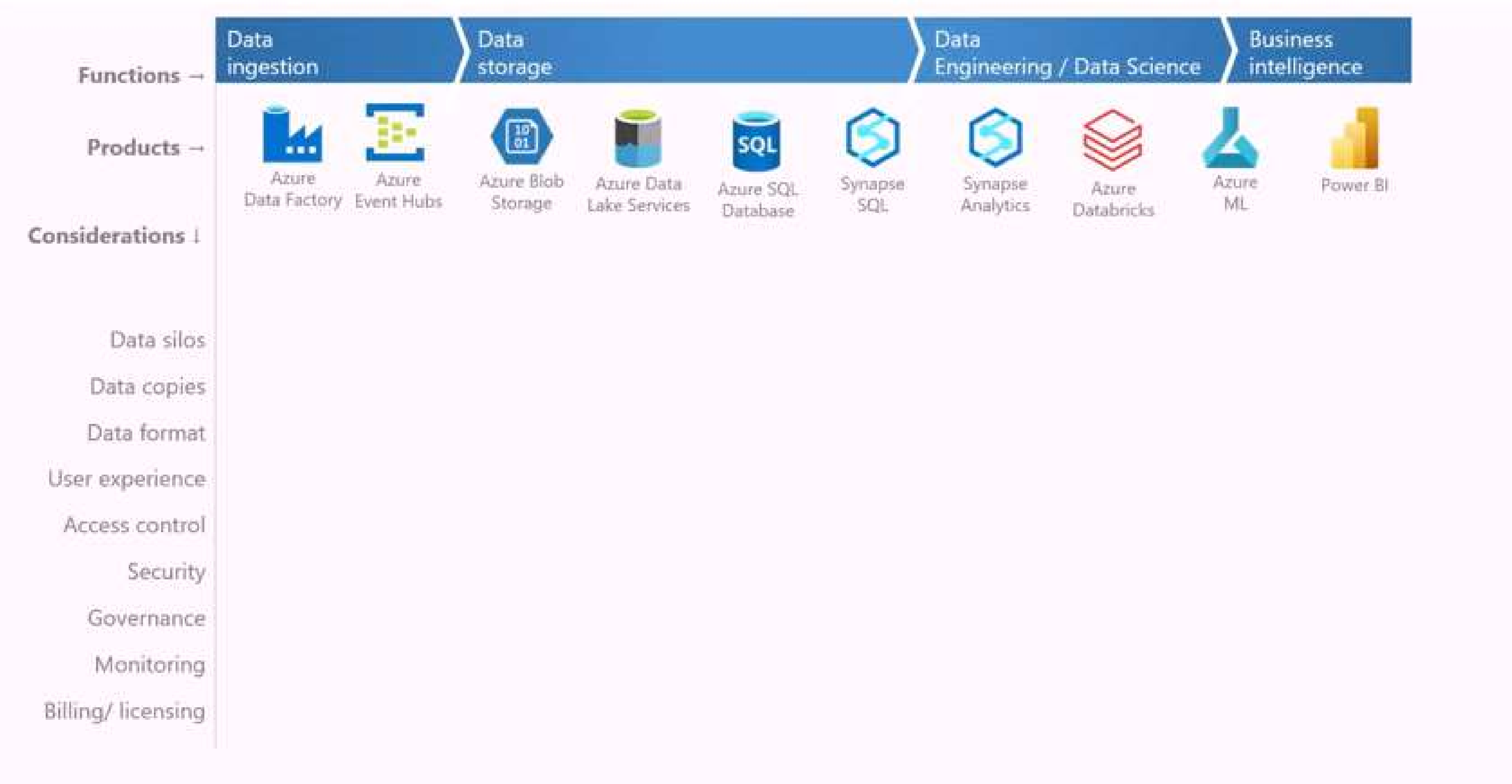
Data Science Team

100010111

100010111

100010111

Manufacturing Team



Functions →

Data ingestion

Data storage

Data Engineering / Data Science

Business intelligence

Products →

Azure Data Factory

Azure Event Hubs

Azure Blob Storage

Azure Data Lake Services

Azure SQL Database

Synapse SQL

Synapse Analytics

Azure Databricks

Azure ML

Power BI

Considerations ↓

Data silos

Data copies

Data format

User experience

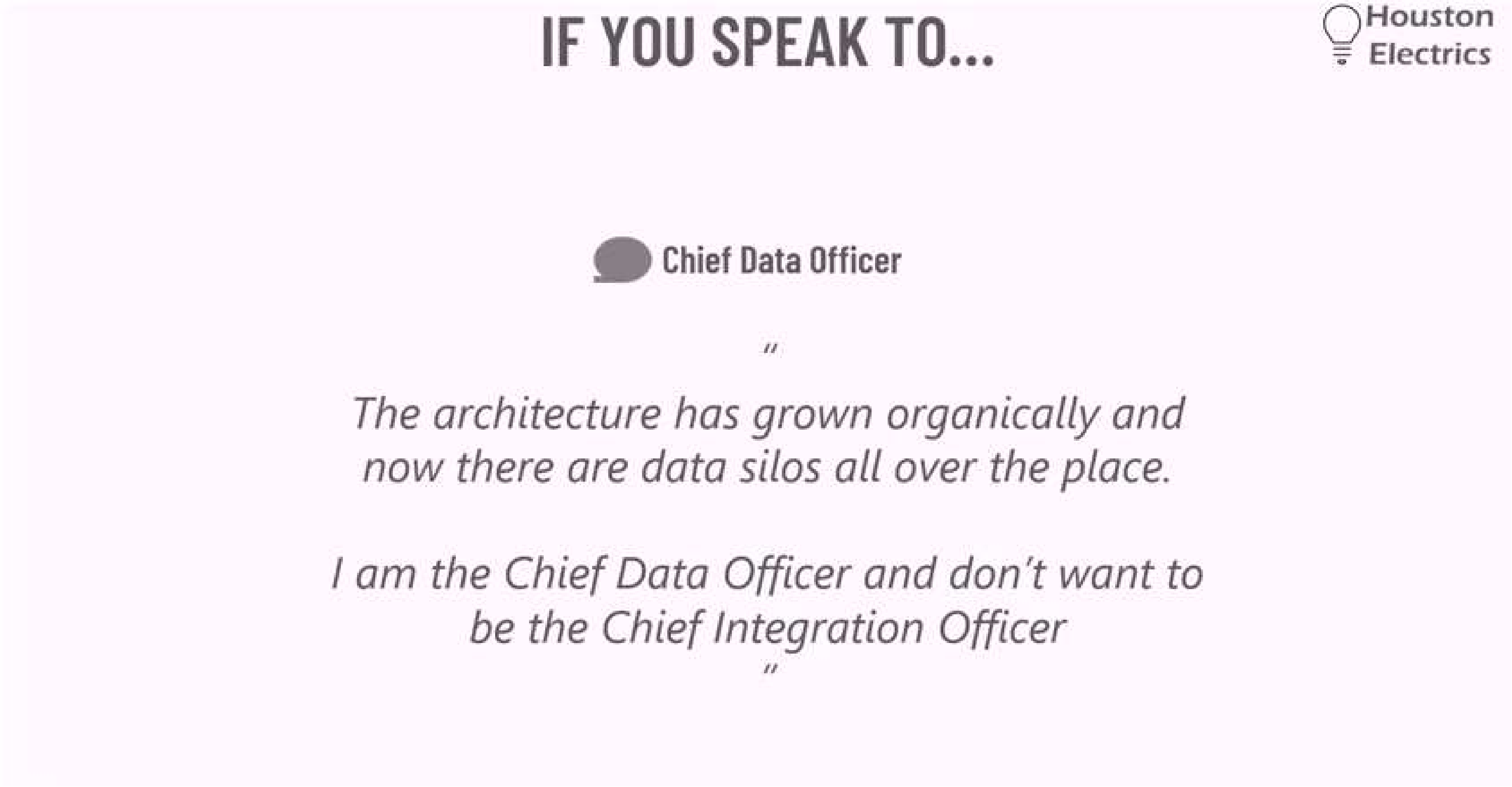
Access control

Security

Governance

Monitoring

Billing/ licensing



Houston Electrics

IF YOU SPEAK TO...

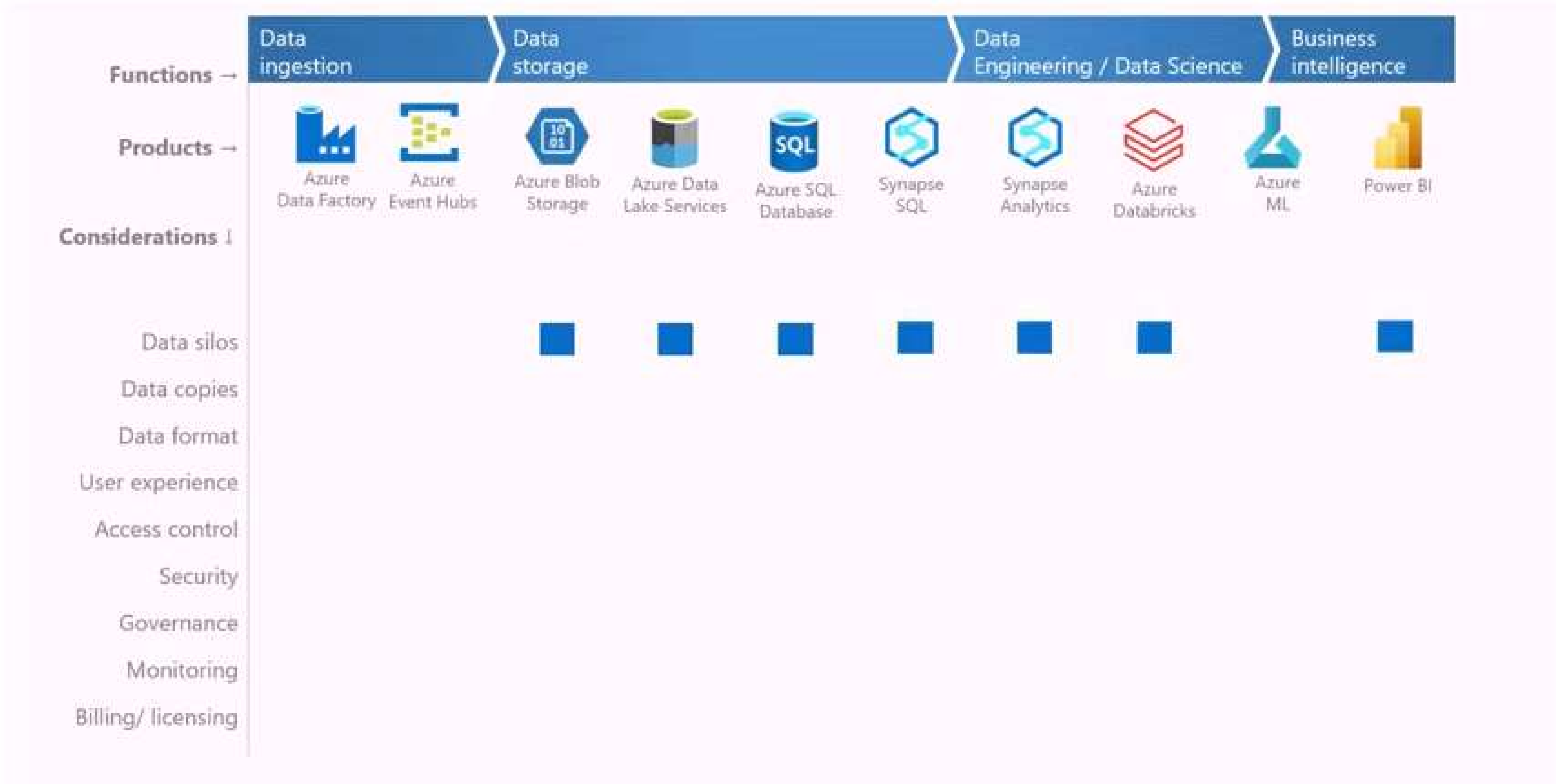
Chief Data Officer

"

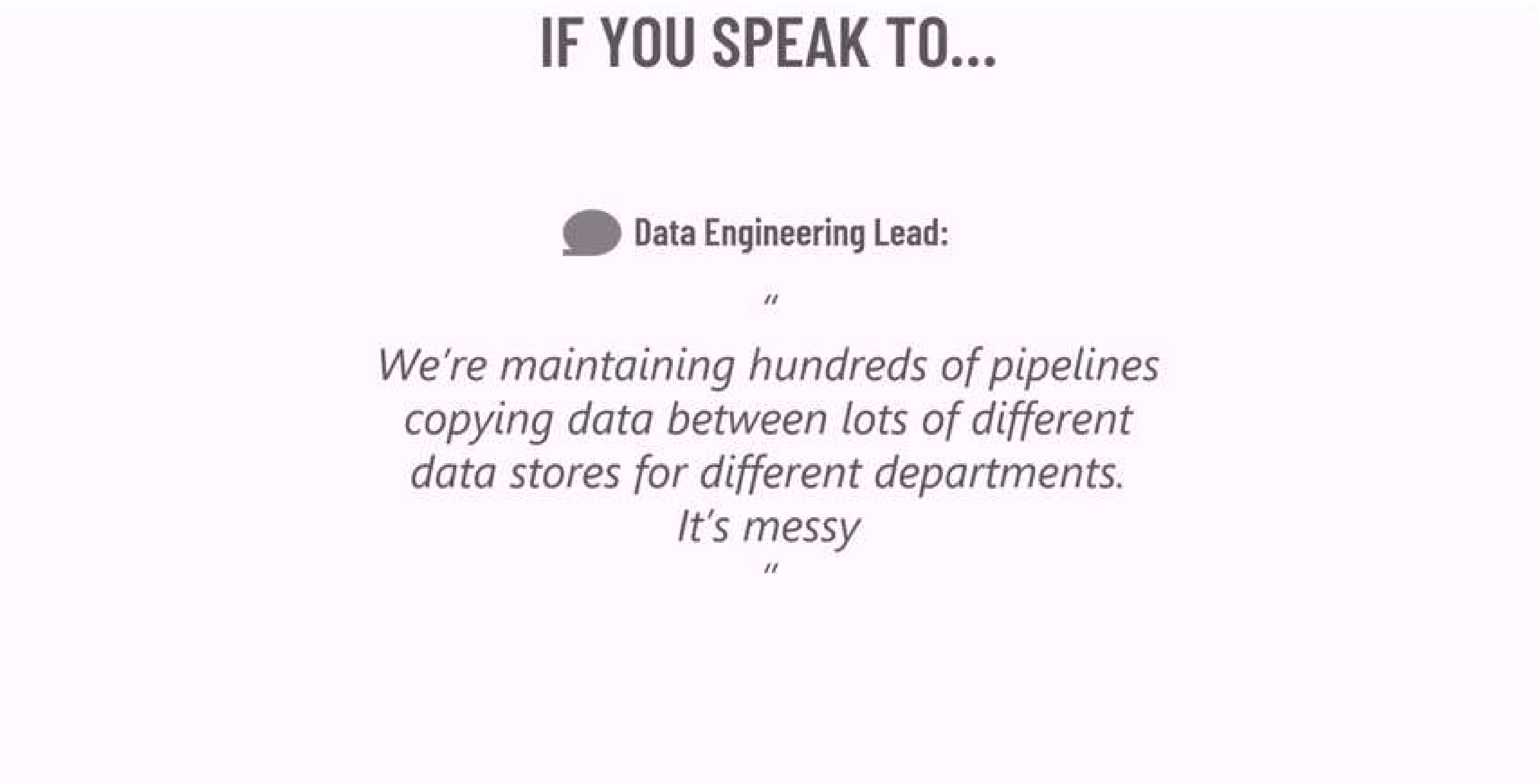
The architecture has grown organically and now there are data silos all over the place.

I am the Chief Data Officer and don't want to be the Chief Integration Officer

"



Functions →	Data ingestion		Data storage				Data Engineering / Data Science			Business intelligence
Products →	Azure Data Factory	Azure Event Hubs	Azure Blob Storage	Azure Data Lake Services	Azure SQL Database	Synapse SQL	Synapse Analytics	Azure Databricks	Azure ML	Power BI
Considerations ↓										
Data silos			■	■	■	■	■	■		■
Data copies										
Data format										
User experience										
Access control										
Security										
Governance										
Monitoring										
Billing/ licensing										



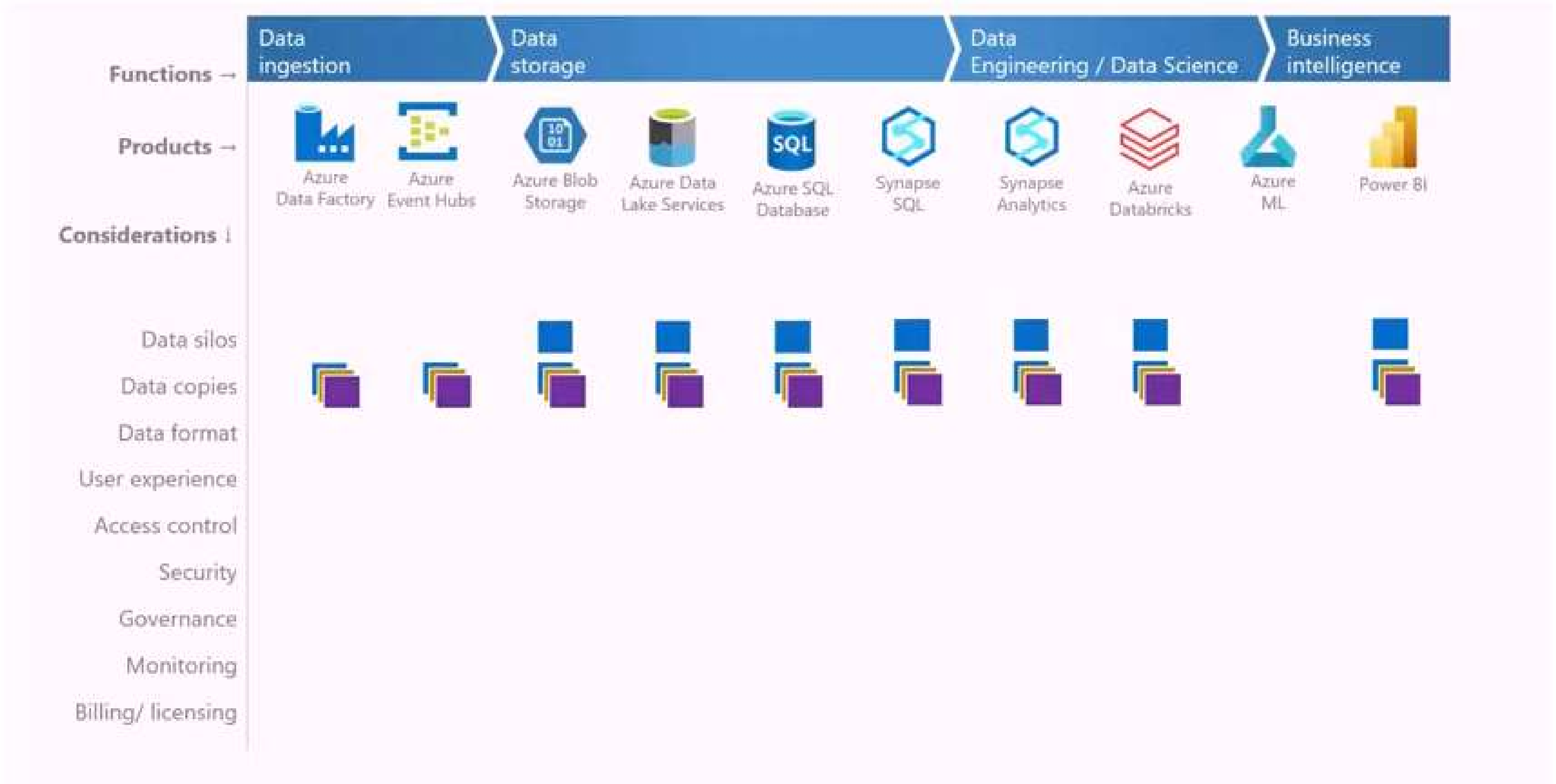
IF YOU SPEAK TO...

Data Engineering Lead:

"

We're maintaining hundreds of pipelines copying data between lots of different data stores for different departments. It's messy

"



Functions →	Data ingestion		Data storage				Data Engineering / Data Science			Business intelligence
Products →	Azure Data Factory	Azure Event Hubs	Azure Blob Storage	Azure Data Lake Services	Azure SQL Database	Synapse SQL	Synapse Analytics	Azure Databricks	Azure ML	Power BI
Considerations ↓										
Data silos			■	■	■	■	■	■		■
Data copies	■	■	■	■	■	■	■	■		■
Data format										
User experience										
Access control										
Security										
Governance										
Monitoring										
Billing/ licensing										

IF YOU SPEAK TO...

Data Scientist:

"

We have data scattered in so many formats in different places, it takes me days just to get clean datasets just to begin an analysis. Even then, I don't know if I can trust the data.

"

IF YOU SPEAK TO...

Data Scientist:

"

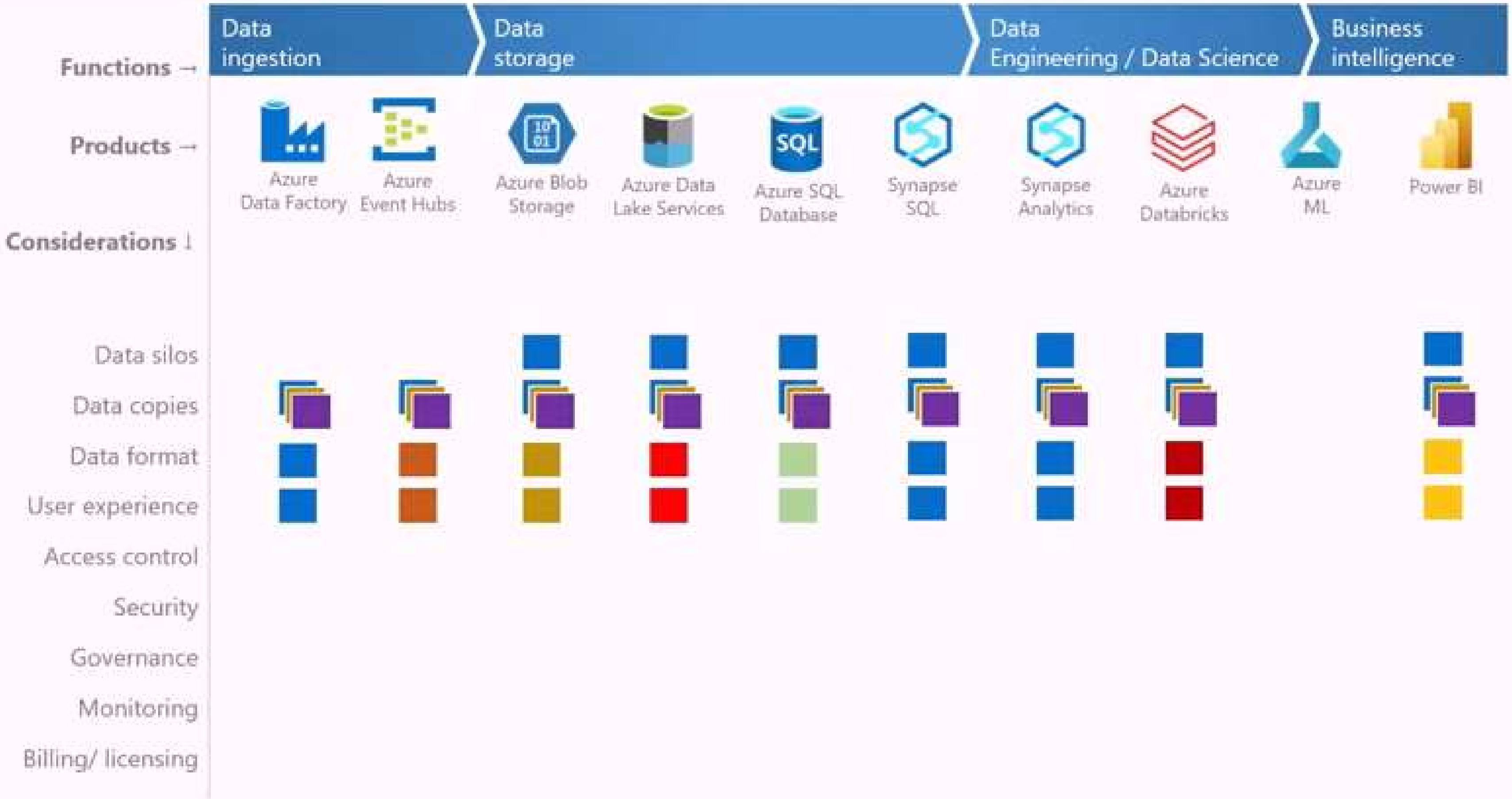
We have data scattered in so many formats in different places, it takes me days just to get clean datasets just to begin an analysis. Even then, I don't know if I can trust the data.

"

"

I had to learn the intricacies of many different data technologies, and each one is different. We had a new starter. It took 3 months to upskill them in all the different platforms we use.

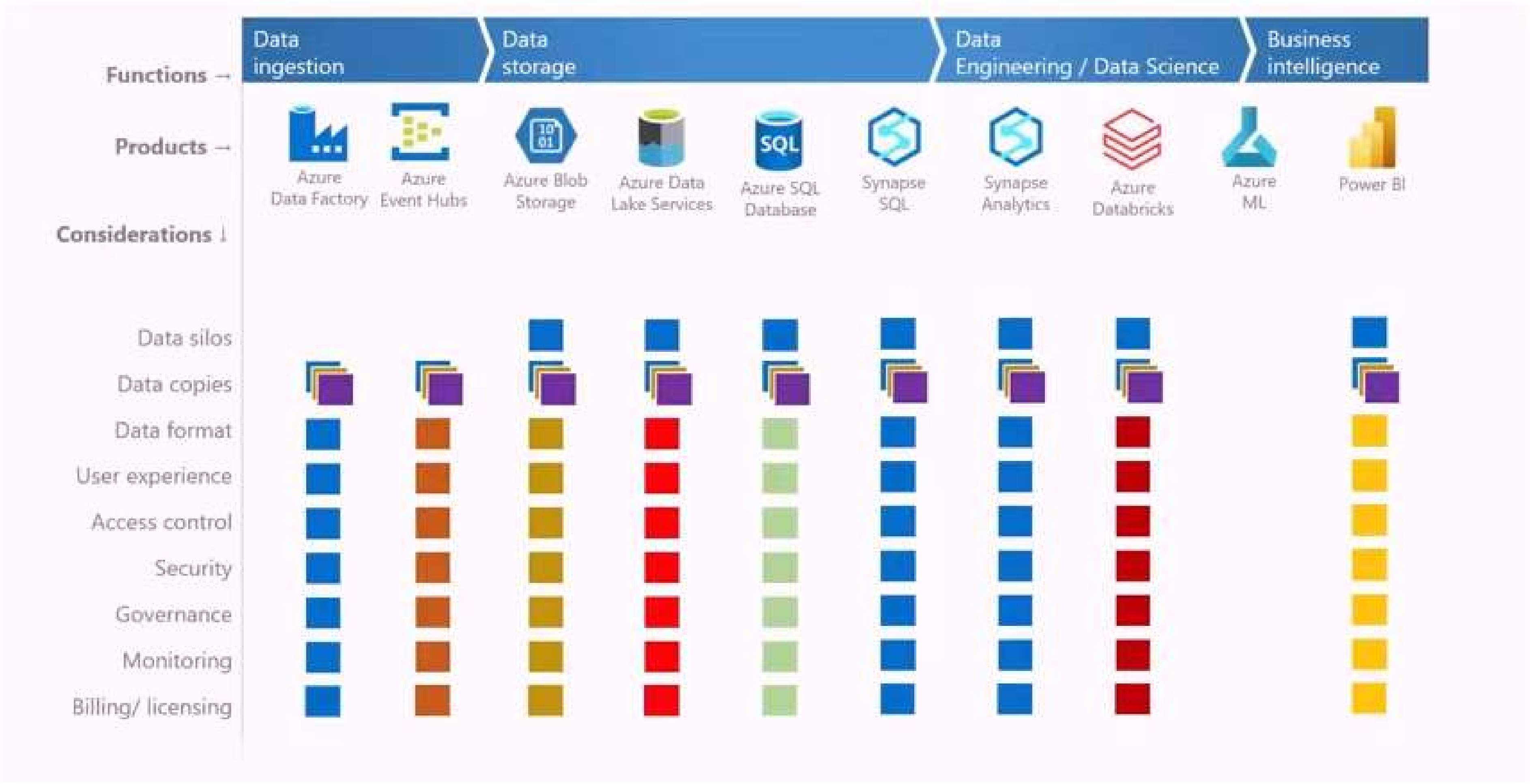
"



Functions →	Data ingestion		Data storage				Data Engineering / Data Science			Business intelligence
Products →	Azure Data Factory	Azure Event Hubs	Azure Blob Storage	Azure Data Lake Services	Azure SQL Database	Synapse SQL	Synapse Analytics	Azure Databricks	Azure ML	Power BI
Considerations ↓										
Data silos			🟦	🟦	🟦	🟦	🟦	🟦		🟦
Data copies	🟪	🟪	🟪	🟪	🟪	🟪	🟪	🟪		🟪
Data format	🟦	🟧	🟨	🟥	🟩	🟦	🟦	🟥		🟨
User experience	🟦	🟧	🟨	🟥	🟩	🟦	🟦	🟥		🟨
Access control										
Security										
Governance										
Monitoring										
Billing/ licensing										

IT Director

"We're using too many systems, all have different security profiles and requirements to keep data at-rest and in-transit secure. It's a nightmare"



Functions →	Data ingestion		Data storage				Data Engineering / Data Science			Business intelligence
Products →	Azure Data Factory	Azure Event Hubs	Azure Blob Storage	Azure Data Lake Services	Azure SQL Database	Synapse SQL	Synapse Analytics	Azure Databricks	Azure ML	Power BI
Considerations ↓										
Data silos			🟦	🟦	🟦	🟦	🟦	🟦		🟦
Data copies	🟪	🟪	🟪	🟪	🟪	🟪	🟪	🟪		🟪
Data format	🟦	🟧	🟫	🟥	🟩	🟦	🟦	🟥		🟨
User experience	🟦	🟧	🟫	🟥	🟩	🟦	🟦	🟥		🟨
Access control	🟦	🟧	🟫	🟥	🟩	🟦	🟦	🟥		🟨
Security	🟦	🟧	🟫	🟥	🟩	🟦	🟦	🟥		🟨
Governance	🟦	🟧	🟫	🟥	🟩	🟦	🟦	🟥		🟨
Monitoring	🟦	🟧	🟫	🟥	🟩	🟦	🟦	🟥		🟨
Billing/ licensing	🟦	🟧	🟫	🟥	🟩	🟦	🟦	🟥		🟨

IF YOU SPEAK TO...

IT Director

"

I dread getting our Azure bill every month. So unpredictable, and sometimes scary. Each data product has their own pricing structure so it's difficult to predict how much we will be charged month-to-month.

"

IF YOU SPEAK TO...

Dashboard user:

"

I'm not sure if I can trust the data that is being presented to me – doesn't always reflect reality

"

HOW FABRIC IS DIFFERENT

Functions →

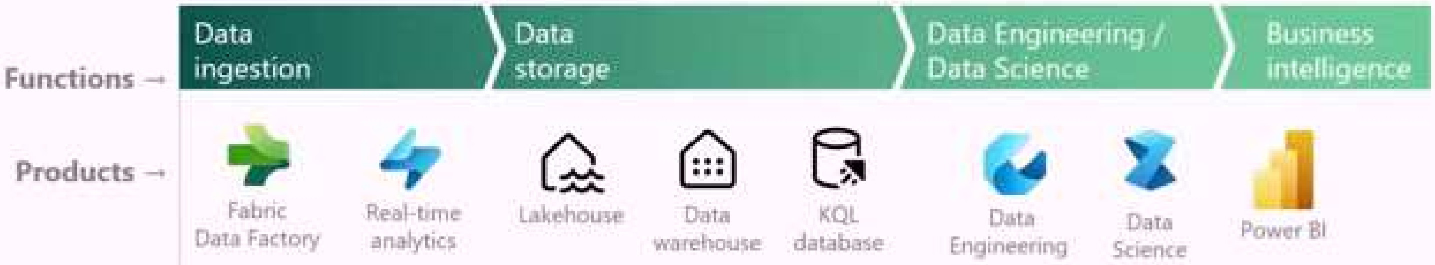
Data ingestion

Data storage

Data Engineering / Data Science

Business intelligence

HOW FABRIC IS DIFFERENT



Functions →

Data ingestion

Data storage

Data Engineering / Data Science

Business intelligence

Products →

Fabric Data Factory

Real-time analytics

Lakehouse

Data warehouse

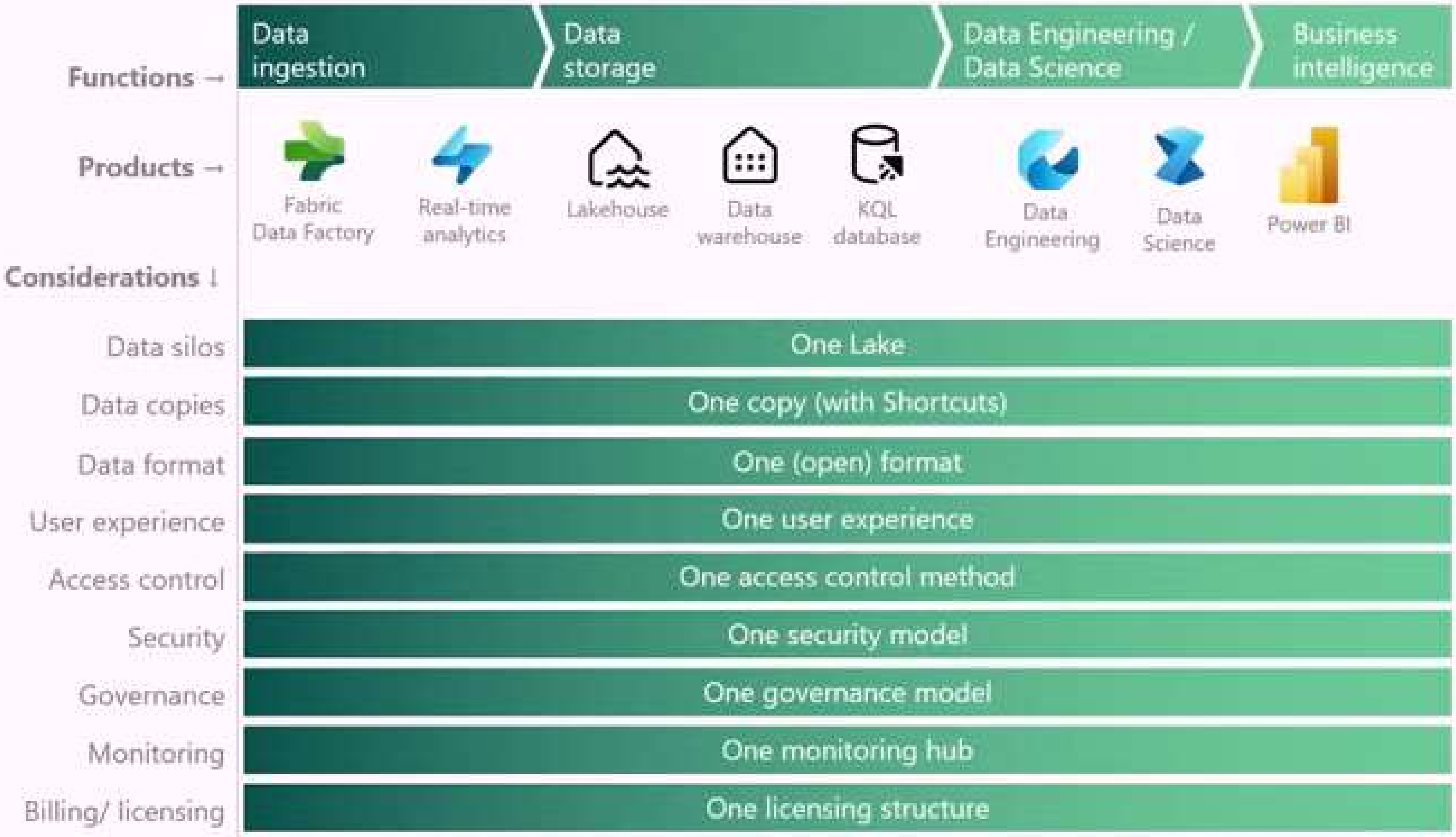
KQL database

Data Engineering

Data Science

Power BI

HOW FABRIC IS DIFFERENT



Functions →

Data ingestion

Data storage

Data Engineering / Data Science

Business intelligence

Products →

Fabric Data Factory

Real-time analytics

Lakehouse

Data warehouse

KQL database

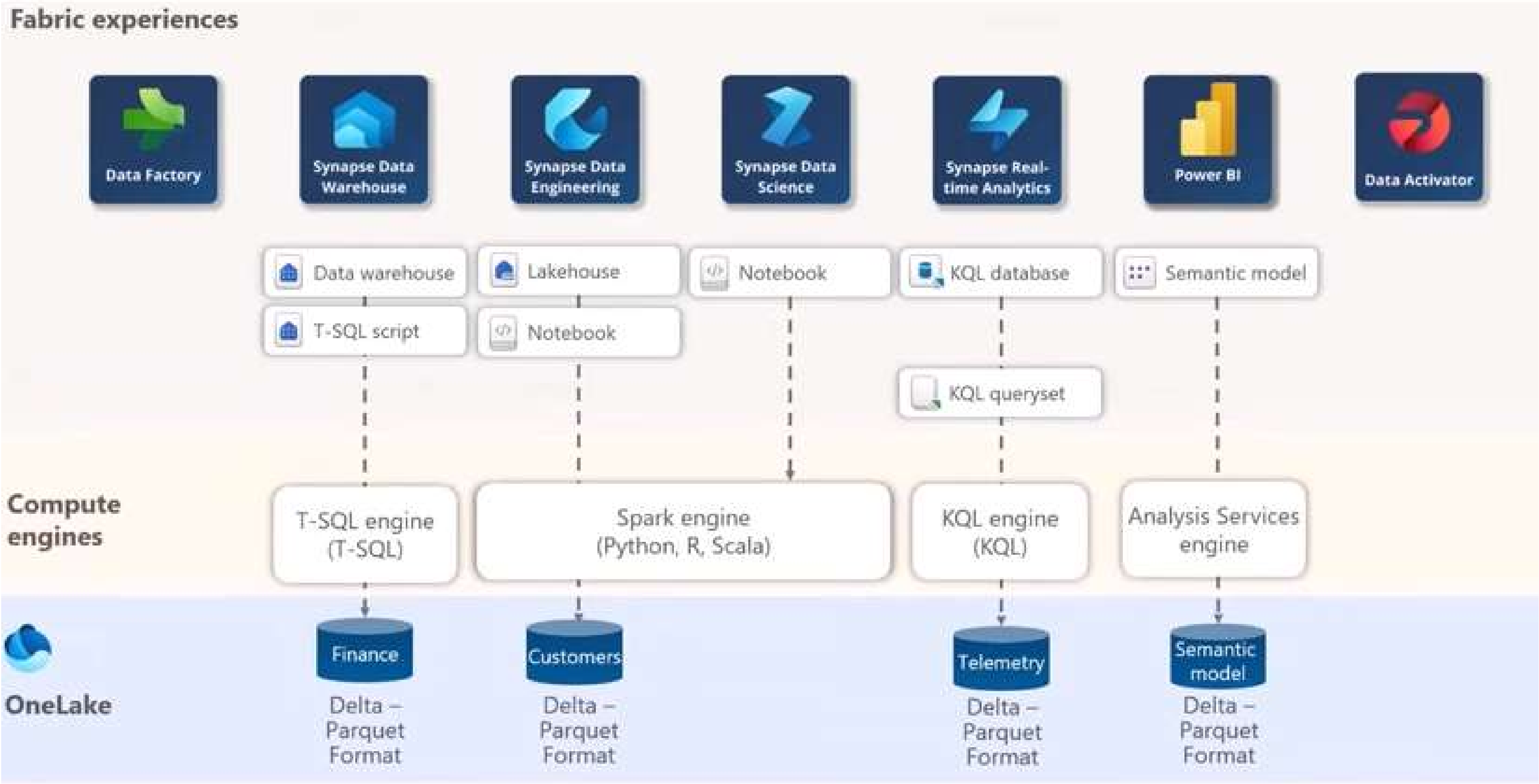
Data Engineering

Data Science

Power BI

Considerations ↓

Data silos	One Lake
Data copies	One copy (with Shortcuts)
Data format	One (open) format
User experience	One user experience
Access control	One access control method
Security	One security model
Governance	One governance model
Monitoring	One monitoring hub
Billing/ licensing	One licensing structure



Fabric experiences

Data Factory

Synapse Data Warehouse

Synapse Data Engineering

Synapse Data Science

Synapse Real-time Analytics

Power BI

Data Activator

Data warehouse

T-SQL script

Lakehouse

Notebook

Notebook

KQL database

KQL queryset

Semantic model

Compute engines

T-SQL engine (T-SQL)

Spark engine (Python, R, Scala)

KQL engine (KQL)

Analysis Services engine

OneLake

Finance

Delta – Parquet Format

Customers

Delta – Parquet Format

Telemetry

Delta – Parquet Format

Semantic model

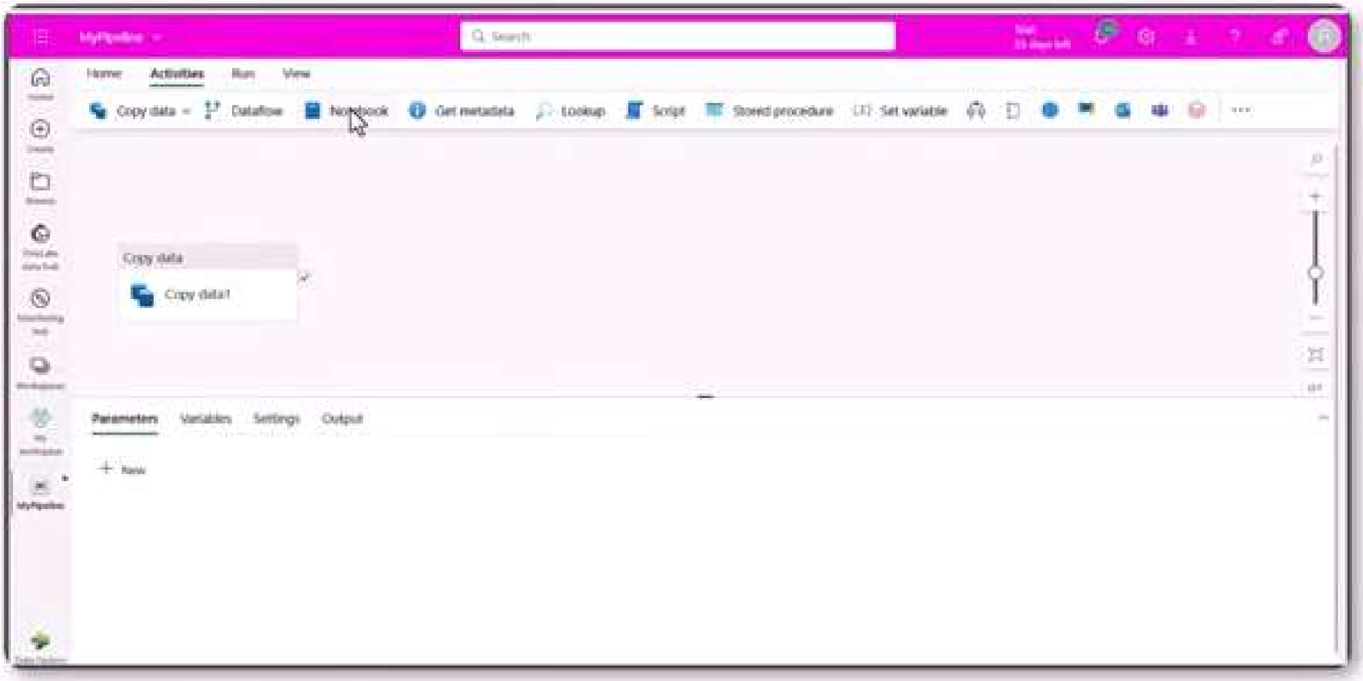
Delta – Parquet Format

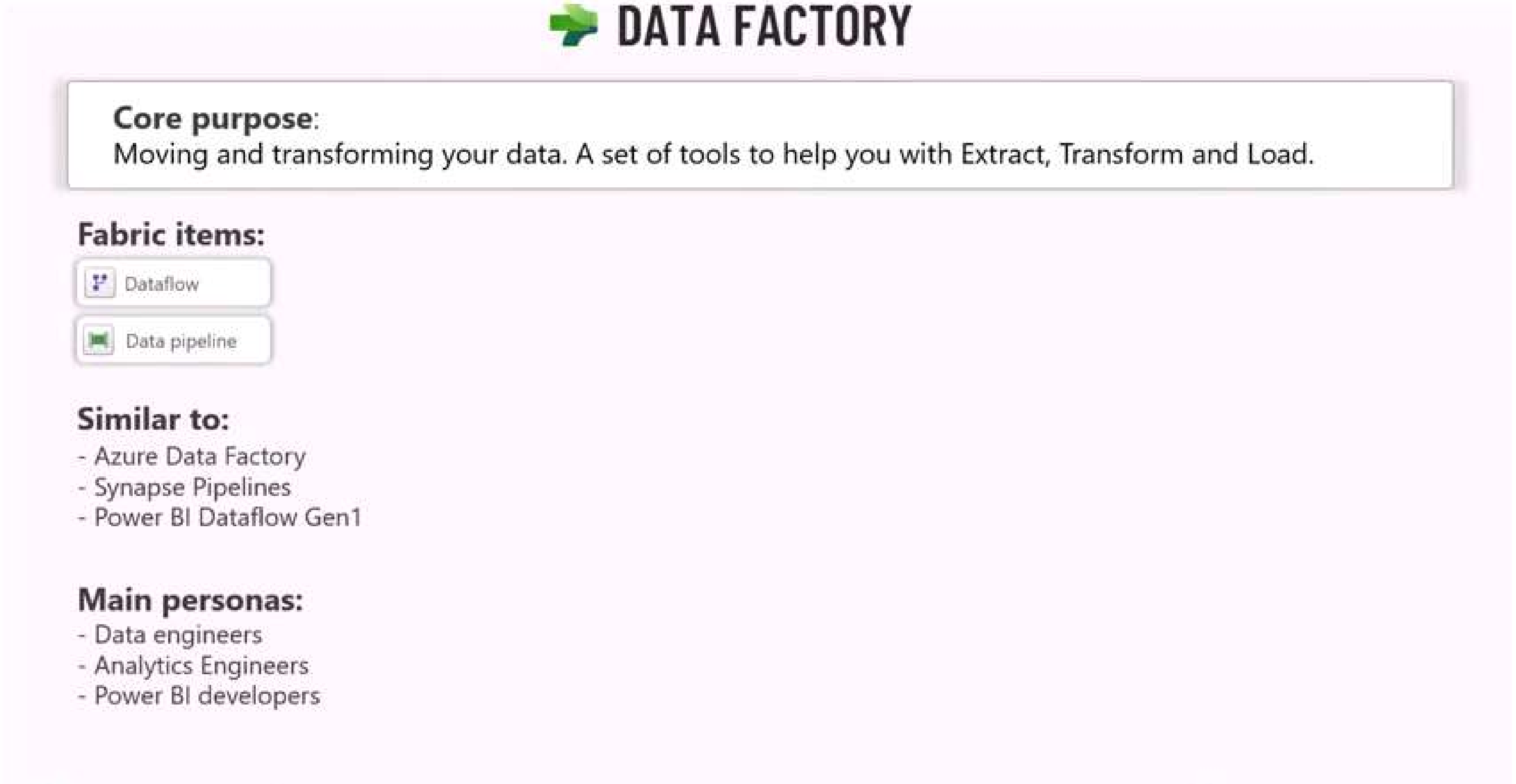
DATA FACTORY

Core purpose:
Moving and transforming your data. A set of tools to help you with Extract, Transform and Load.

Fabric items:

- Dataflow
- Data pipeline





DATA FACTORY

Core purpose:
Moving and transforming your data. A set of tools to help you with Extract, Transform and Load.

Fabric items:

- Dataflow
- Data pipeline

Similar to:

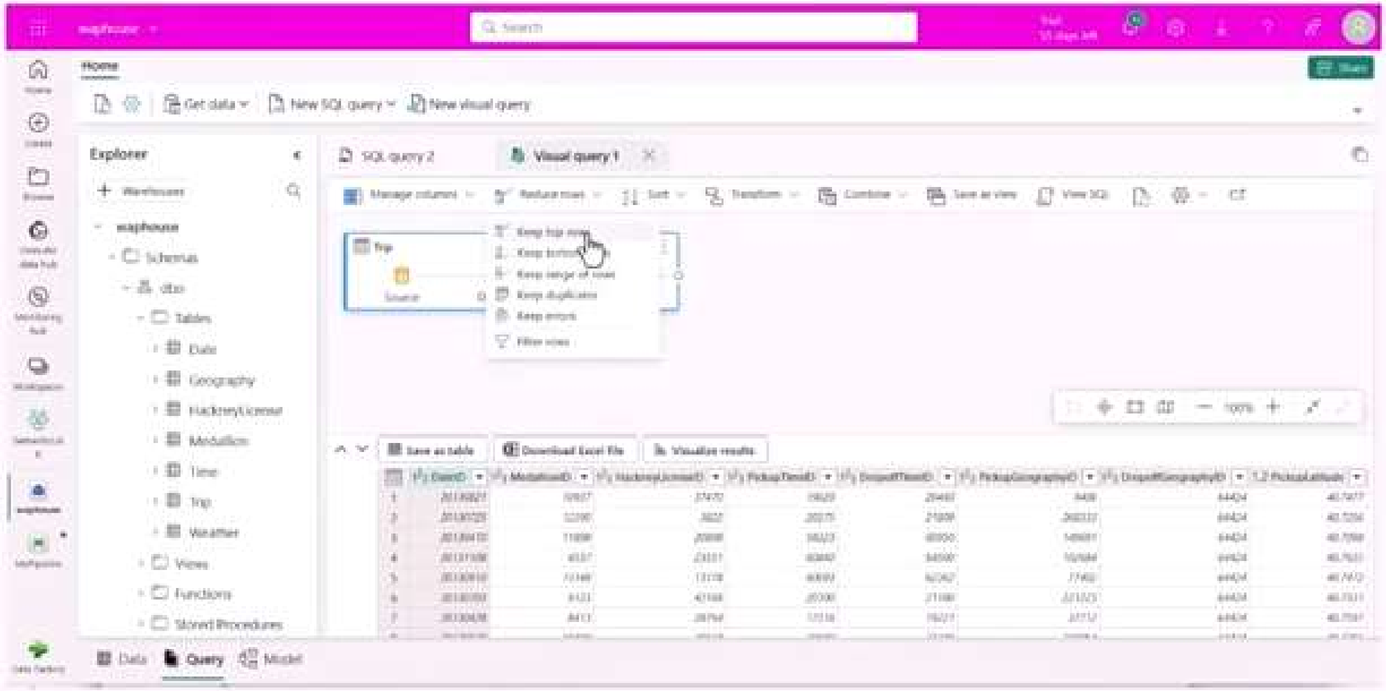
- Azure Data Factory
- Synapse Pipelines
- Power BI Dataflow Gen1

Main personas:

- Data engineers
- Analytics Engineers
- Power BI developers

SYNAPSE DATA WAREHOUSE

Core purpose:
provides a familiar transactional data warehouse solution with tables, schemas, views, stored procedures etc. Query-able using Structured Query Language (T-SQL)



	DateID	MedallionID	HackneyLicenseID	PickupTimeID	DropoffTimeID	PickupGeographyID	DropoffGeographyID	PickupLatitude
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	64424	[illegible]
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	64424	[illegible]
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	64424	[illegible]
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	64424	[illegible]
5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	64424	[illegible]
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	64424	[illegible]
7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	64424	[illegible]

SYNAPSE DATA WAREHOUSE

Core purpose:
provides a familiar transactional data warehouse solution with tables, schemas, views, stored procedures etc. Query-able using Structured Query Language (T-SQL)

Fabric items:

Data warehouse

Similar to:

- SQL Server/ Azure SQL
- Synapse SQL Serverless/Dedicated
- Snowflake

Main personas:

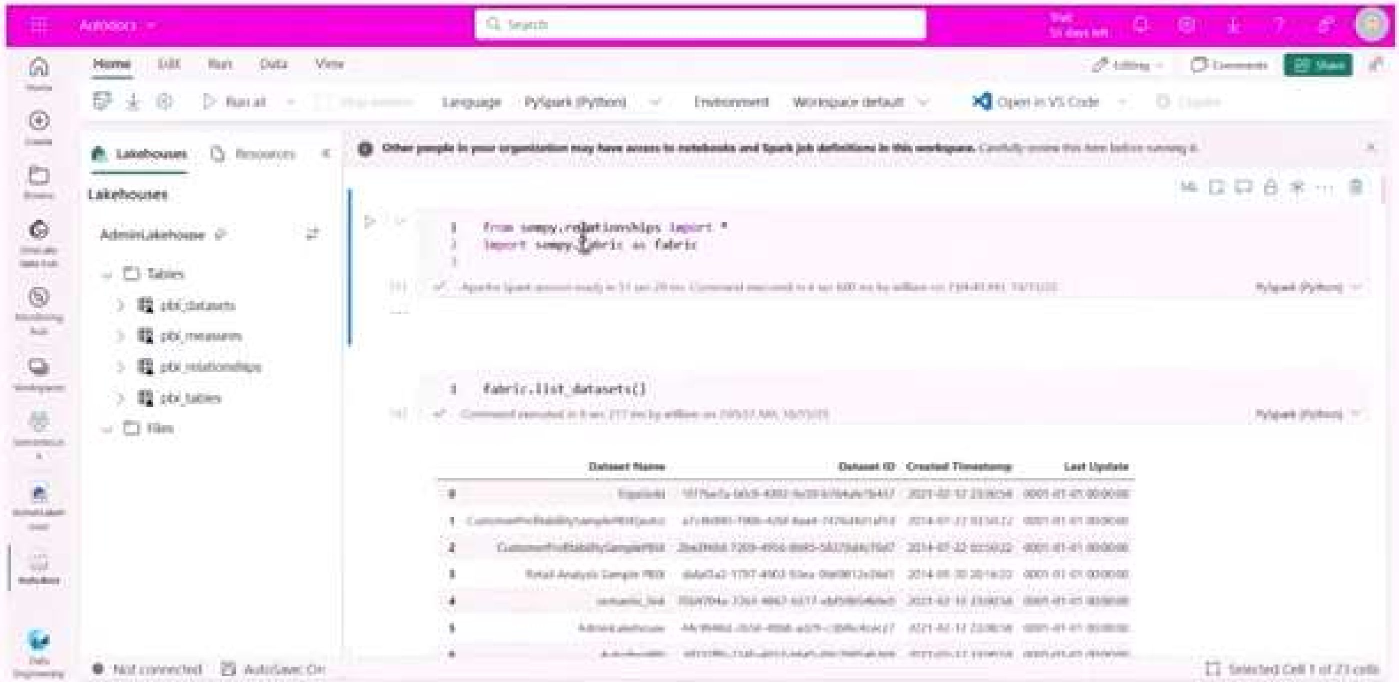
- Database administrators
- Data engineers
- Data analysts

SYNAPSE DATA ENGINEERING

Core purpose:
Enables users to design, build, and maintain **infrastructures and systems** that enable their organizations to collect, store, process, and analyze large volumes of data .

Fabric items:

- Lakehouse
- Notebook

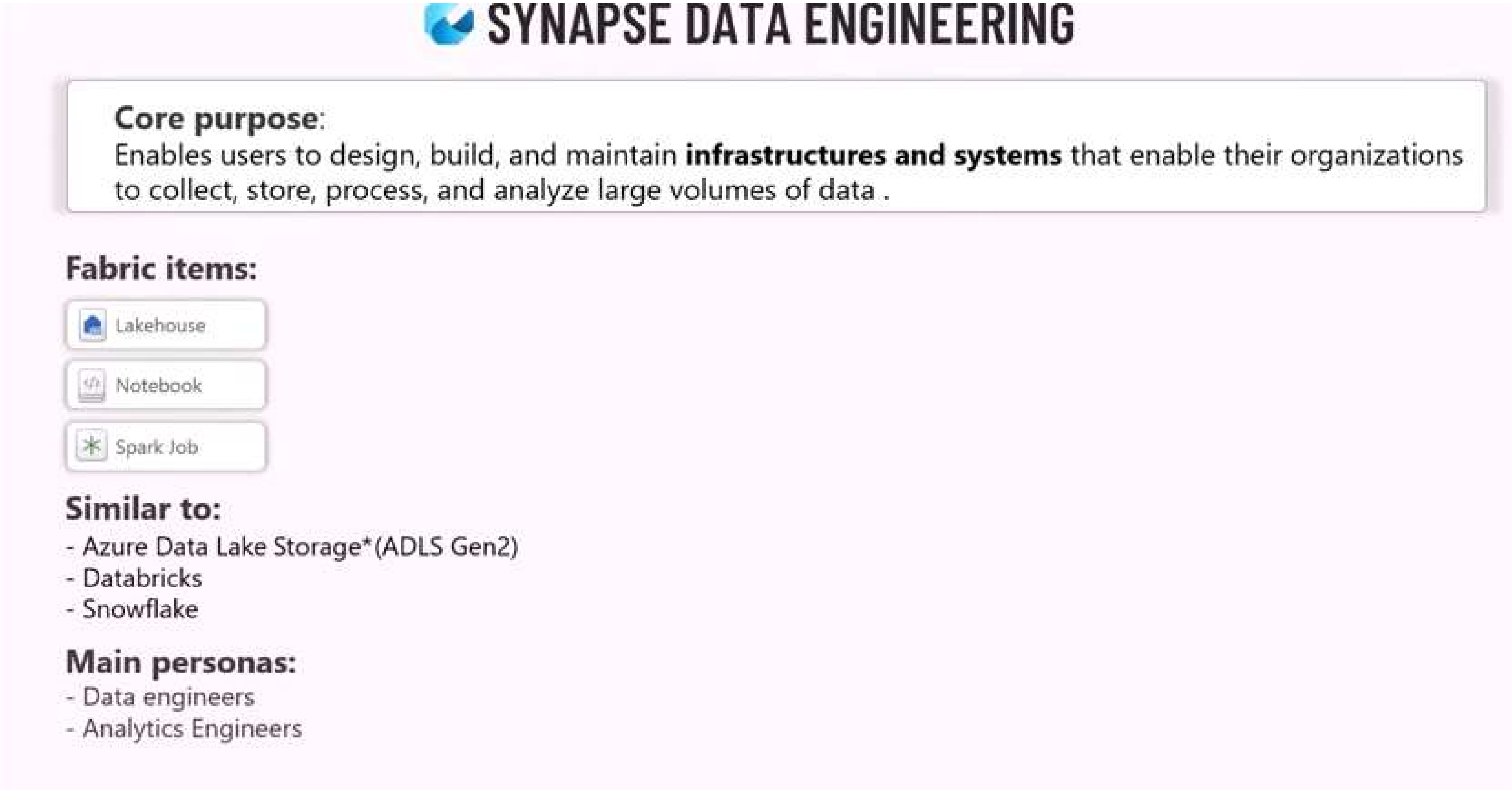


Other people in your organization may have access to notebooks and Spark job definitions in this workspace. Carefully review this item before running it.

```
from sempy.relationships import *
import sempy.fabric as fabric
```

```
fabric.list_datasets()
```

	Dataset Name	Dataset ID	Created Timestamp	Last Update
0	[illegible]	[illegible]	[illegible]	0001-01-01 00:00:00
1	CustomerProfitabilitySamplePBIX(auto)	[illegible]	[illegible]	0001-01-01 00:00:00
2	CustomerProfitabilitySamplePBIX	[illegible]	[illegible]	0001-01-01 00:00:00
3	Retail Analysis Sample PBIX	[illegible]	[illegible]	0001-01-01 00:00:00
4	semantic_link	[illegible]	[illegible]	0001-01-01 00:00:00
5	AdminLakehouse	[illegible]	[illegible]	0001-01-01 00:00:00
6	[illegible]	[illegible]	[illegible]	[illegible]



SYNAPSE DATA ENGINEERING

Core purpose:
Enables users to design, build, and maintain **infrastructures and systems** that enable their organizations to collect, store, process, and analyze large volumes of data .

Fabric items:

Lakehouse

Notebook

Spark Job

Similar to:

- Azure Data Lake Storage*(ADLS Gen2)
- Databricks
- Snowflake

Main personas:

- Data engineers
- Analytics Engineers

SYNAPSE DATA SCIENCE

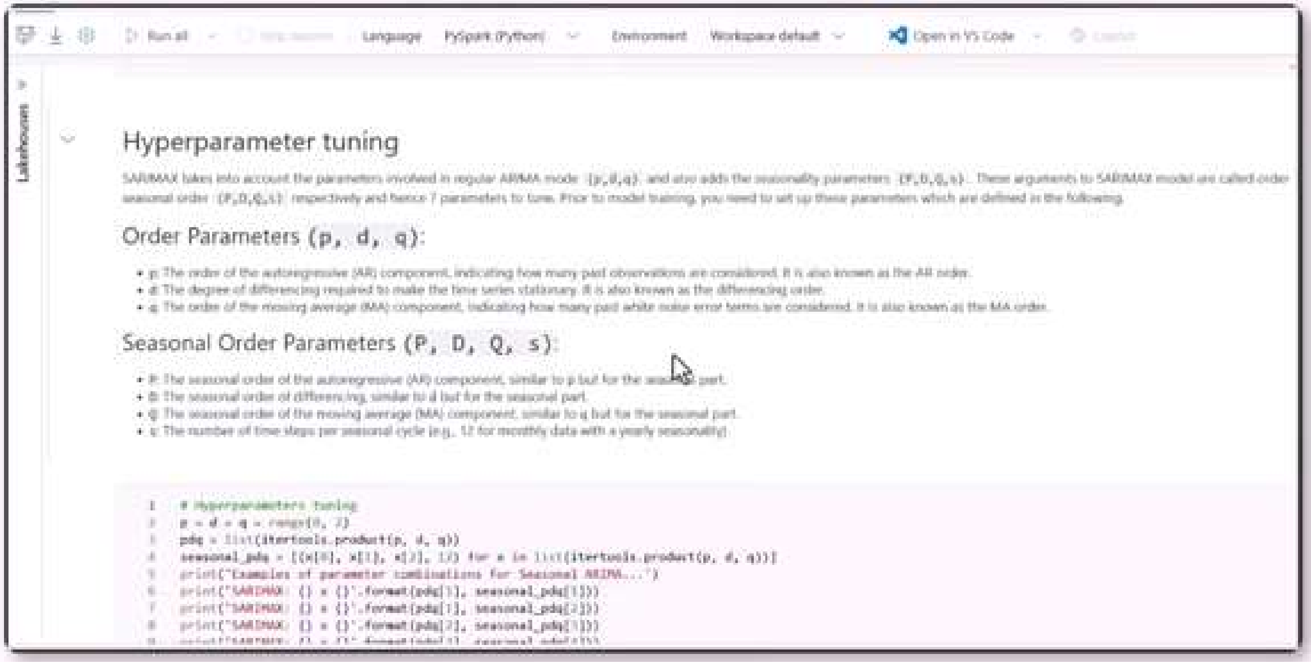
Core purpose:
Supports the entire data science workflow in an organisation, from data exploration, preparation and cleansing to experimentation, modeling, model scoring and serving of predictive insights to BI reports.

Fabric items:

- Notebook
- Experiments
- ML Models

Similar to:

- Azure Machine Learning
- Synapse notebooks
- Databricks notebooks



Hyperparameter tuning

SARIMAX takes into account the parameters involved in regular ARIMA mode (p,d,q) and also adds the seasonality parameters (P,D,Q,s). These arguments to SARIMAX model are called order seasonal order (P,D,Q,s) respectively and hence 7 parameters to tune. Prior to model training, you need to set up these parameters which are defined in the following.

Order Parameters (p, d, q):

- p: The order of the autoregressive (AR) component, indicating how many past observations are considered. It is also known as the AR order.
- d: The degree of differencing required to make the time series stationary. It is also known as the differencing order.
- q: The order of the moving average (MA) component, indicating how many past white noise error terms are considered. It is also known as the MA order.

Seasonal Order Parameters (P, D, Q, s):

- P: The seasonal order of the autoregressive (AR) component, similar to p but for the seasonal part.
- D: The seasonal order of differencing, similar to d but for the seasonal part.
- Q: The seasonal order of the moving average (MA) component, similar to q but for the seasonal part.
- s: The number of time steps per seasonal cycle (e.g., 12 for monthly data with a yearly seasonality).

```python
# Hyperparameters tuning
p = d = q = range(0, 2)
pdq = list(itertools.product(p, d, q))
seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]
print('Examples of parameter combinations for Seasonal ARIMA...')
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[1]))
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[2]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[3]))
```

SYNAPSE REAL-TIME ANALYTICS

Core purpose:
Provides a set of tools to ingest, manage and analyze real-time event data

Fabric items:

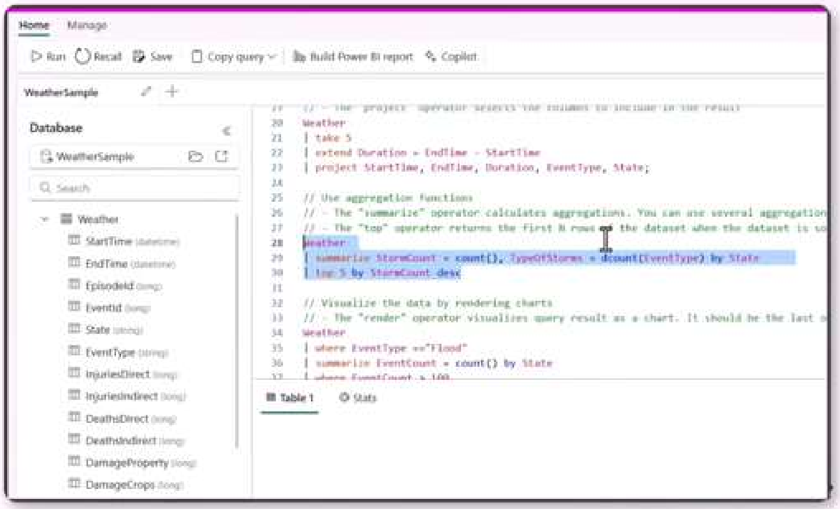
- KQL database
- Eventstream
- KQL queryset

Similar to:

- Azure Data Explorer

Main personas:

- Data engineers
- Analytics Engineers



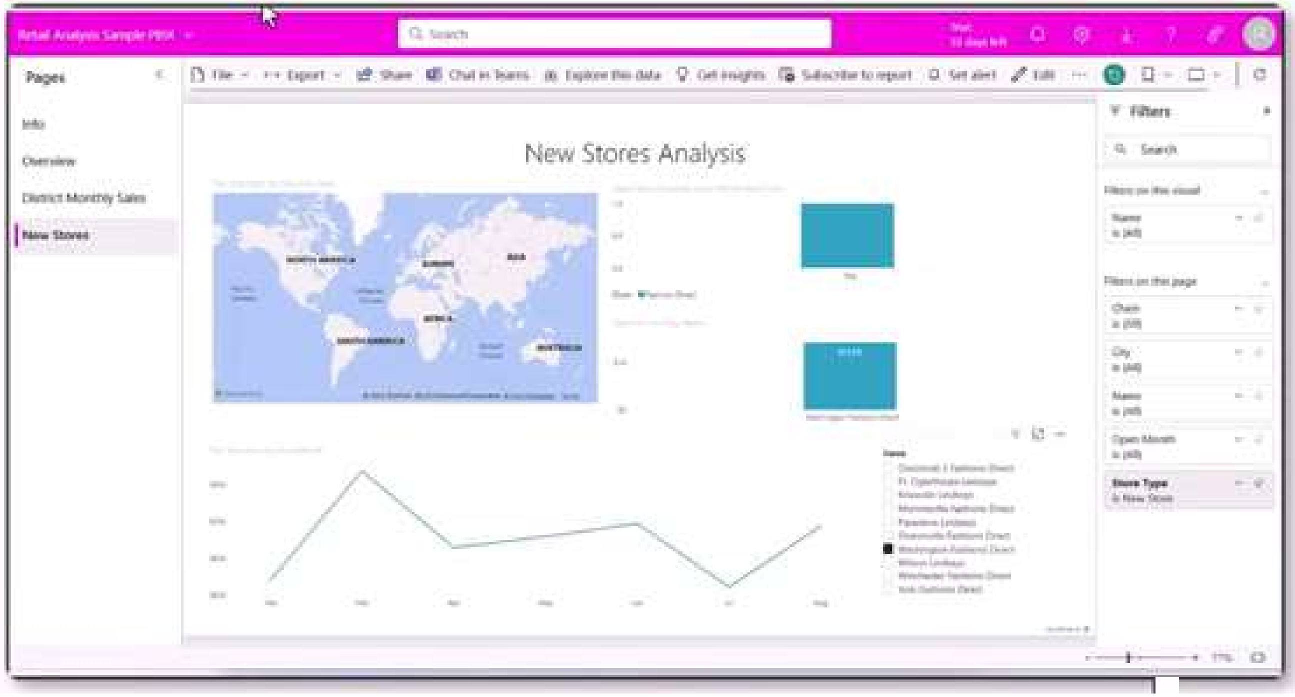
```
// - The "project" operator selects the columns to include in the result
Weather
| take 5
| extend Duration = EndTime - StartTime
| project StartTime, EndTime, Duration, EventType, State;

// Use aggregation functions
// - The "summarize" operator calculates aggregations. You can use several aggregation
// - The "top" operator returns the first N rows of the dataset when the dataset is so
Weather
| summarize StormCount = count(), TypeOfStorms = dcount(EventType) by State
| top 5 by StormCount desc

// Visualize the data by rendering charts
// - The "render" operator visualizes query result as a chart. It should be the last o
Weather
| where EventType =="Flood"
| summarize EventCount = count() by State
| where EventCount > 100
```

POWER BI

Core purpose:
Power BI is Microsoft's business intelligence solution that allows you to create reports to present visual insights to business users.

Fabric items:

- Report
- Semantic model



Similar to:

- Tableau
- Looker

Main personas:

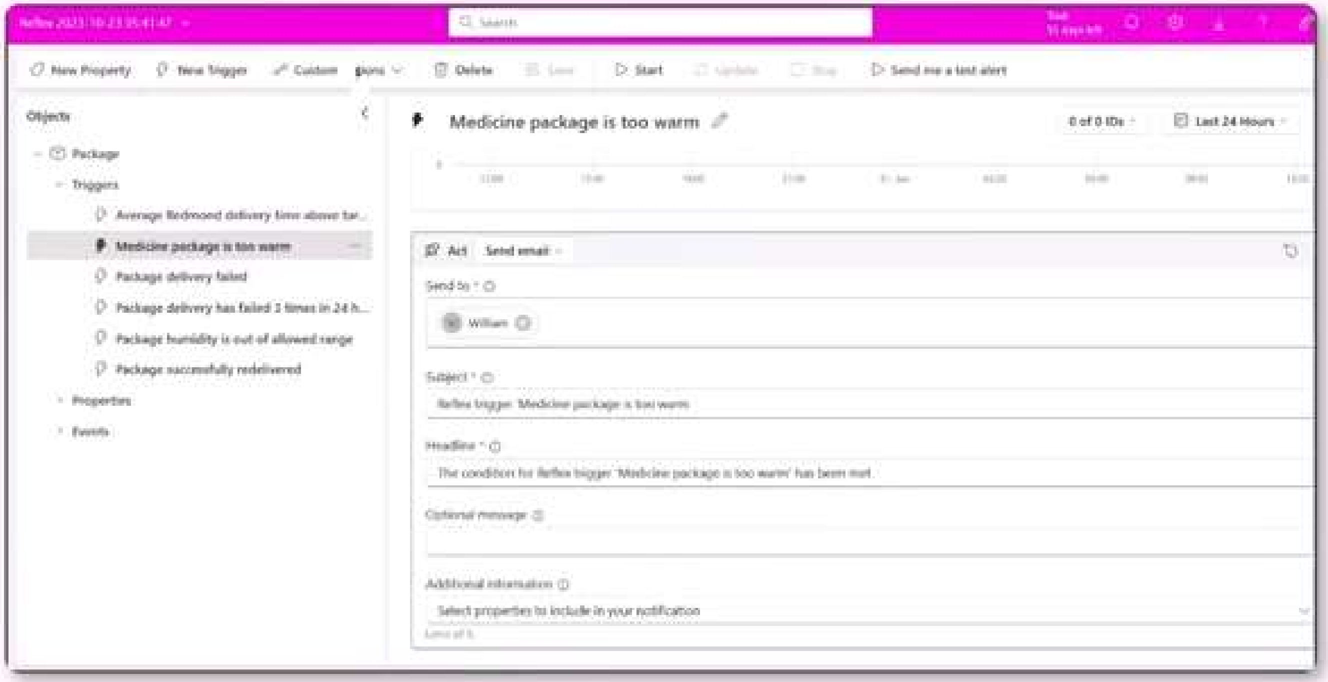
- Business users
- Power BI developers
- Data analysts

DATA ACTIVATOR

Core purpose:
Automatically taking actions (like running a Power Automate routine) when **patterns** or **conditions** are detected in changing data, such as data in Power BI reports and Eventstreams

Fabric items:

Reflex



Search
New Property
New Trigger
Custom
Delete
Save
Start
Update
Stop
Send me a test alert
Objects
Package
Triggers
Average Redmond delivery time above tar...
Medicine package is too warm
Package delivery failed
Package delivery has failed 3 times in 24 h...
Package humidity is out of allowed range
Package successfully redelivered
Properties
Events
0 of 0 IDs
Last 24 Hours
Act
Send email

Medicine package is too warm

Send to *: William

Subject *: Reflex trigger: Medicine package is too warm

Headline *: The condition for Reflex trigger 'Medicine package is too warm' has been met

Optional message

Additional information: Select properties to include in your notification

Similar to:

- Power Automate
- Azure Functions

Main personas:

- Business users
- Power BI developers
- Data analysts