ADVANCING ANALYTICS

Fabric & Databricks: Better Together?

https://craigporteous.com

@cporteous

https://github.com/cporteou

databricks PARTNER PROGRAM SOLUTIONS ARCHITECT CHAMPION

MVP

ADVANCING ANALYTICS

Fabric & Databricks: Better Together?

https://craigporteous.com

@cporteous

https://github.com/cporteou

databricks PARTNER PROGRAM SOLUTIONS ARCHITECT CHAMPION

MVP

Thank you to our Fabric February Friends!

twoday

bouvet

sopra steria

DATAMASTERMINDS

ADVANCING ANALYTICS

Evidi

Profisee
Master Data Management

Tabular Editor

KURANT

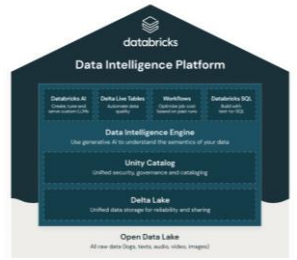
Fraktal

CluedIn

Dufrain
THE DATA COMPANY


FIGHT

Timeline of Analytics Platforms



2010 — Azure

2012 — Apache Spark

2013 — Databricks

2015 — Power BI

2018 — Data Factory

2019 — Delta Lake
- The Lakehouse is born

2019 — Azure Data Lake
- Gen 2

2020 — Azure Synapse Analytics

2023 — Data Intelligence Platform
- Databricks SQL Serverless
- Unity Catalog
- Delta Live Tables
- Databricks AI

2023 — Microsoft Fabric

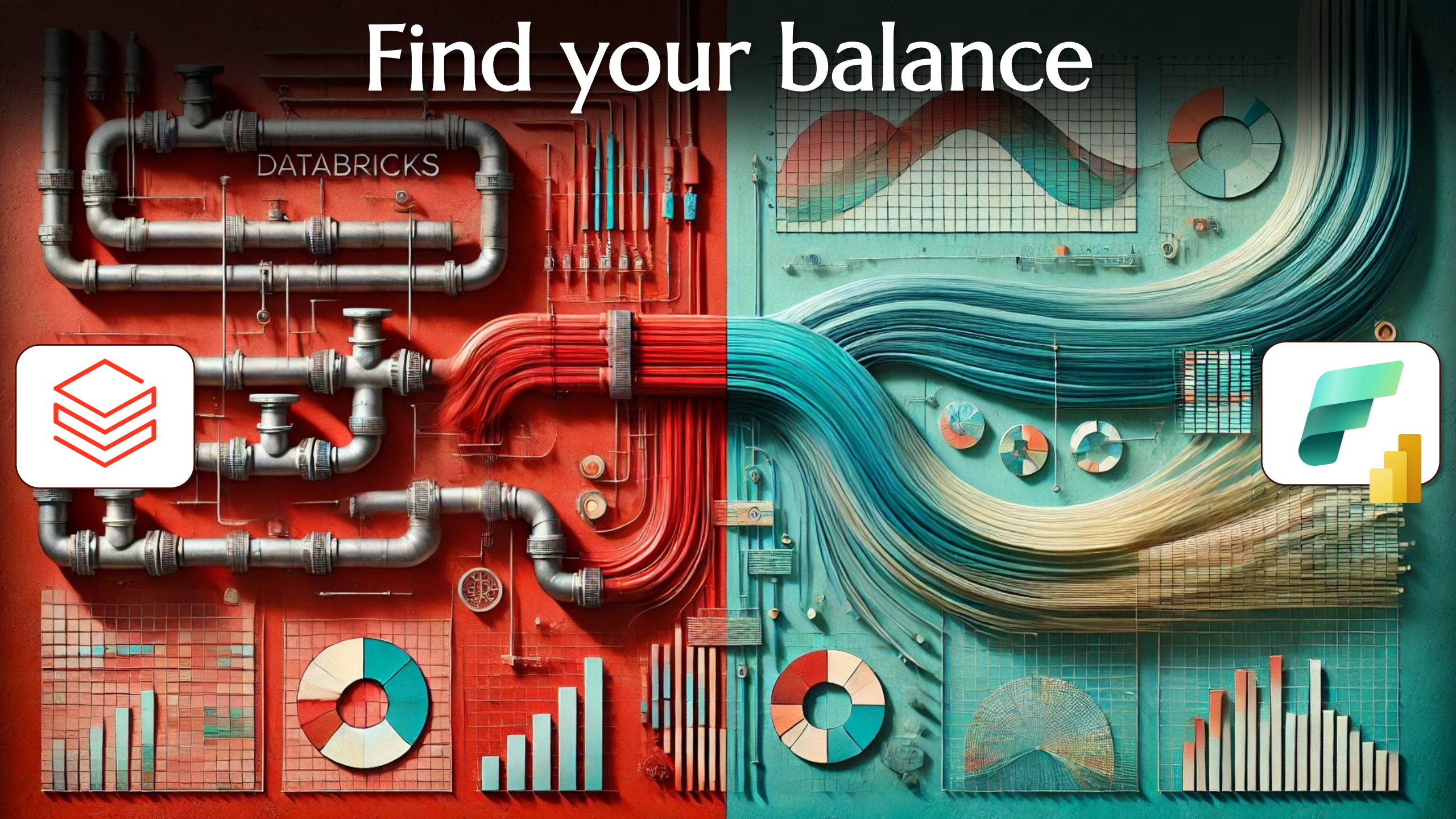
*YEAH, YEAH, WE KNOW THE TIMELINE IS NOT TO SCALE

Delta Lake

- Open source
- Based on Parquet
- ACID Transactions
- Time travel
- Schema Evolution
- Batch & Streaming support

ADVANCING ANALYTICS

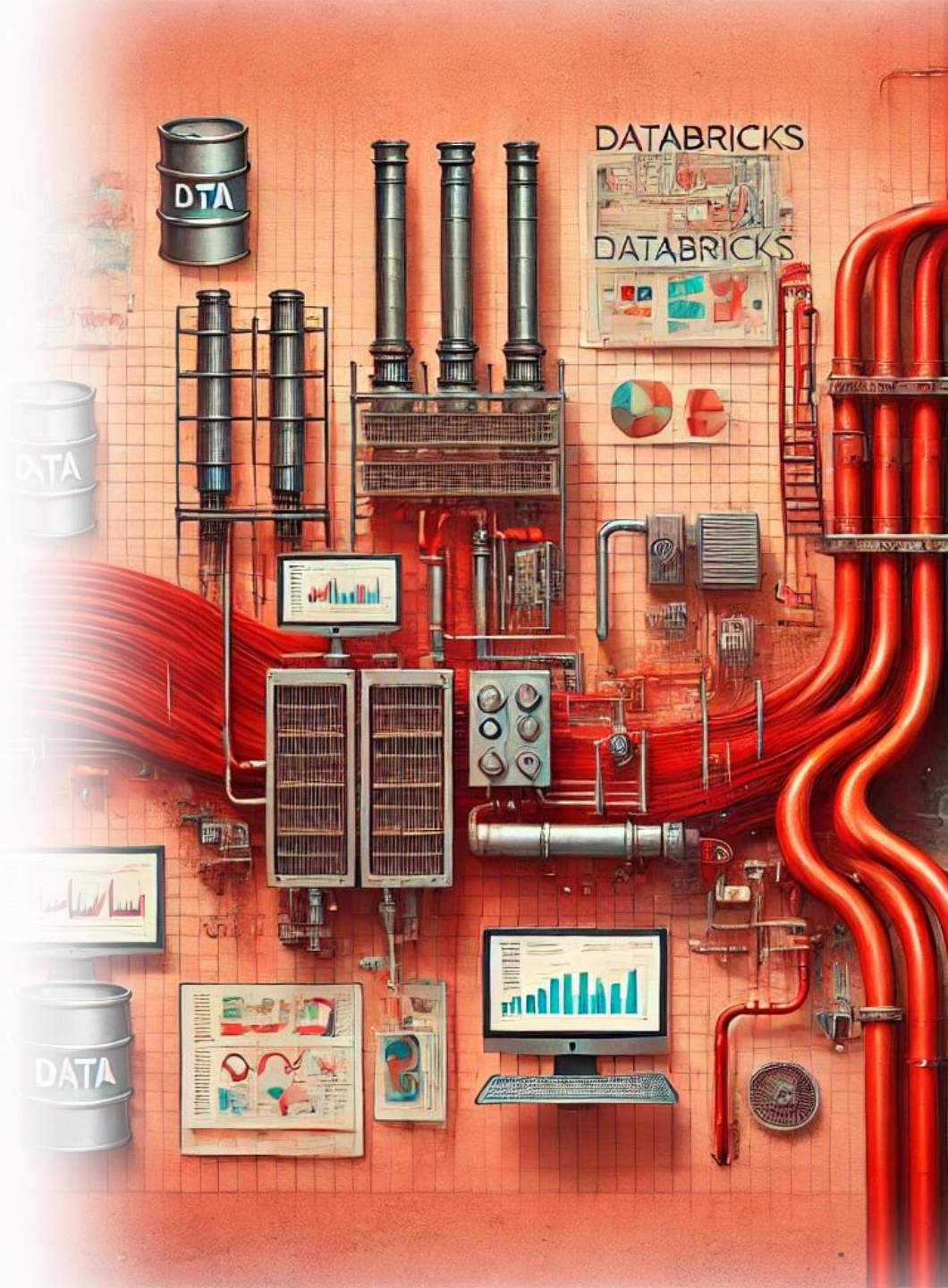
Find your balance



- Hardened, Mature Platform
- Excellent for Deep Data Engineering
- Large array of Data-Science Tools
- Baked-in Data Catalog
- Immature Analytics Experience

The classic Databricks user is code-savvy, familiar with Python/R as well as SQL. They want all of the **control** – clusters, notebooks, libraries, **ALL OF IT**

Databricks are changing, they want to lower the technical barrier



- New, Evolving Platform
- Some tools for Data Engineering
- Very little Data-Science
- Preview Data Catalog
- World-Leading Analytics Experience

The classic Fabric user is not a coder, they are analytics specialists who are branching into citizen ETL & data science as a result of their platforms newfound functionality

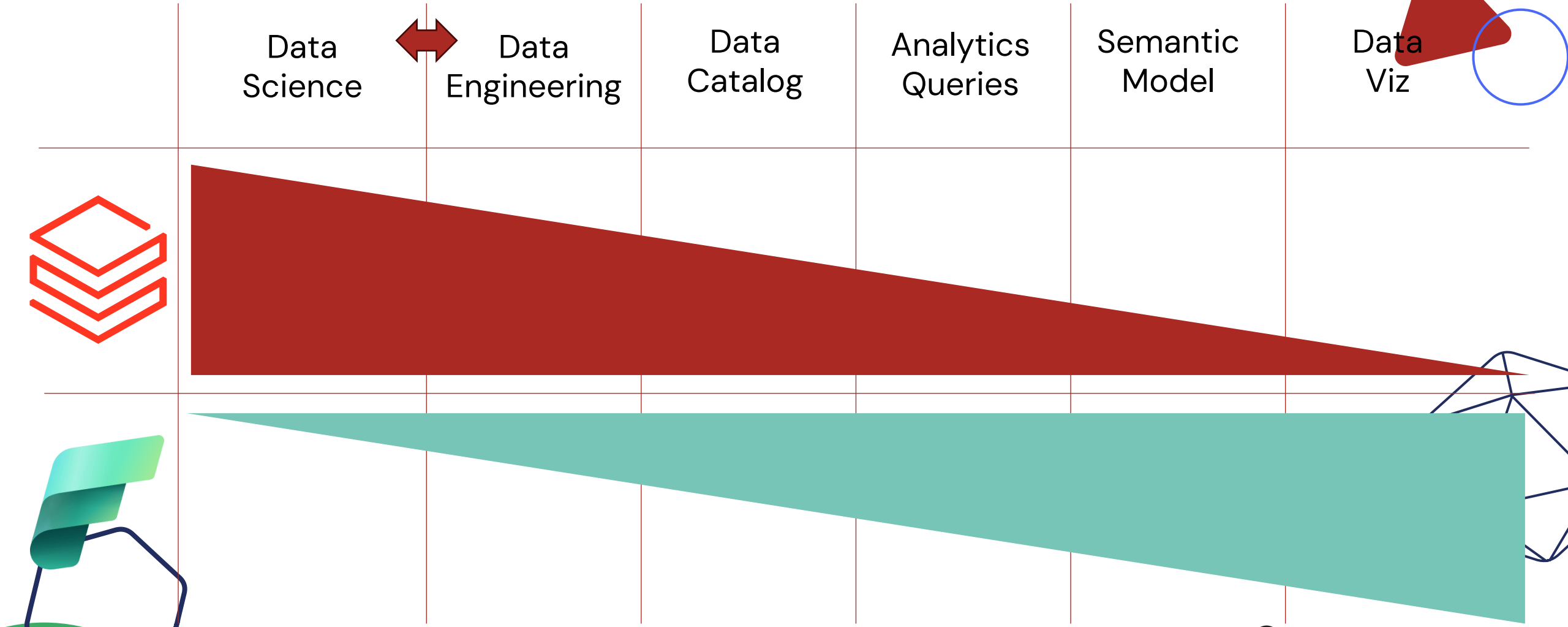
Fabric is changing, they want to attract more hardcore tech-savvy users

ADVANCING ANALYTICS

Workload Comparison

	Data Engineering	Data Science	Data Catalog	Analytics Queries	Semantic Model	Data Viz
				Photon	UC Metrics?	AI/BI
				Native Execution Engine (T-SQL)		

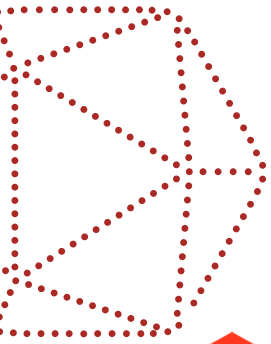
Workload Comparison



This is a very, very rough comparison – but gives you a flavour of where their respective strengths lie **at present**

ADVANCING ANALYTICS

Architectures



Serving In Fabric

Analytics in Fabric

True Hybrid

Serving via Fabric

Getting Data into Power BI

Import Mode – Data is stored in memory within Power BI, regular full/incremental syncs triggered automatically

Direct Query – Every query is pushed back to the source (Databricks SQL)

DirectLake – Rather than relying on an in-memory cached model, Power BI pulls sections of parquet files directly into memory

Note: If limitations are hit (size, complexity etc) the DirectLake query falls back to Direct Query mode

ADVANCING ANALYTICS

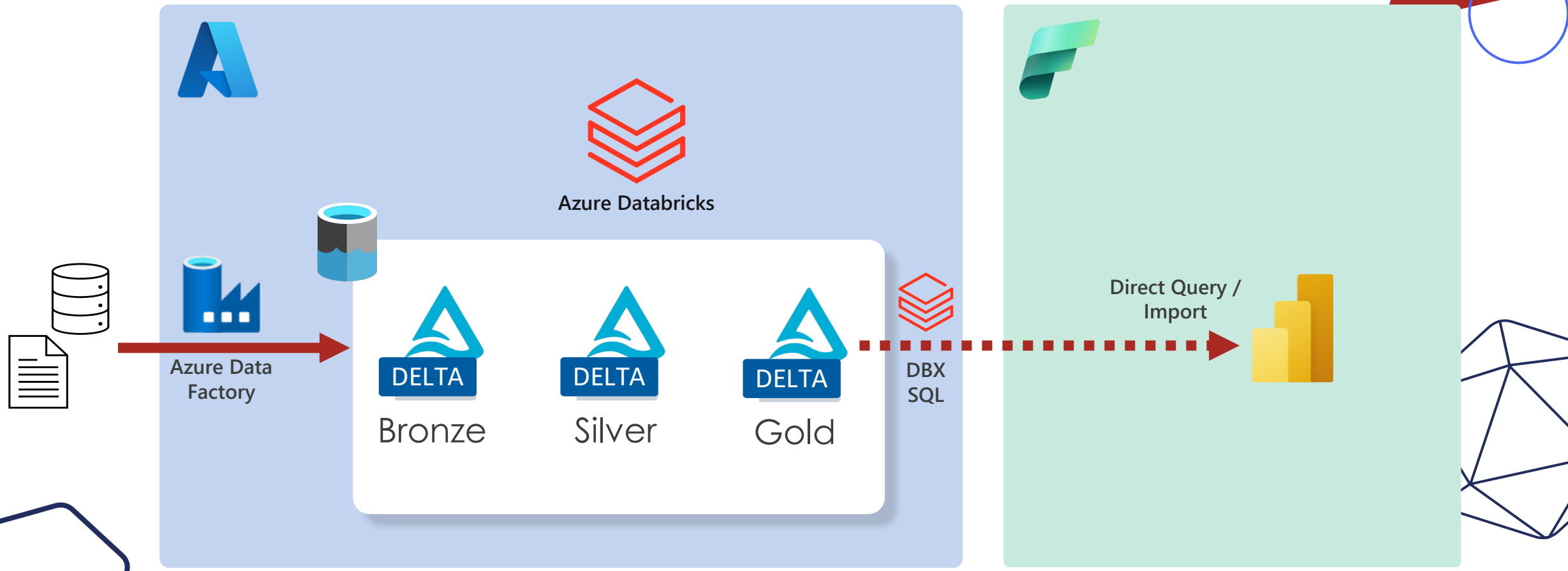
Getting Data into Power BI

DirectLake over vanilla Parquet – In order to run the query, there is a small operation to get the data in the right shape for DAX queries

DirectLake with V-Ordering – The data is already in the right shape for DAX queries and can be used immediately. This results in faster queries

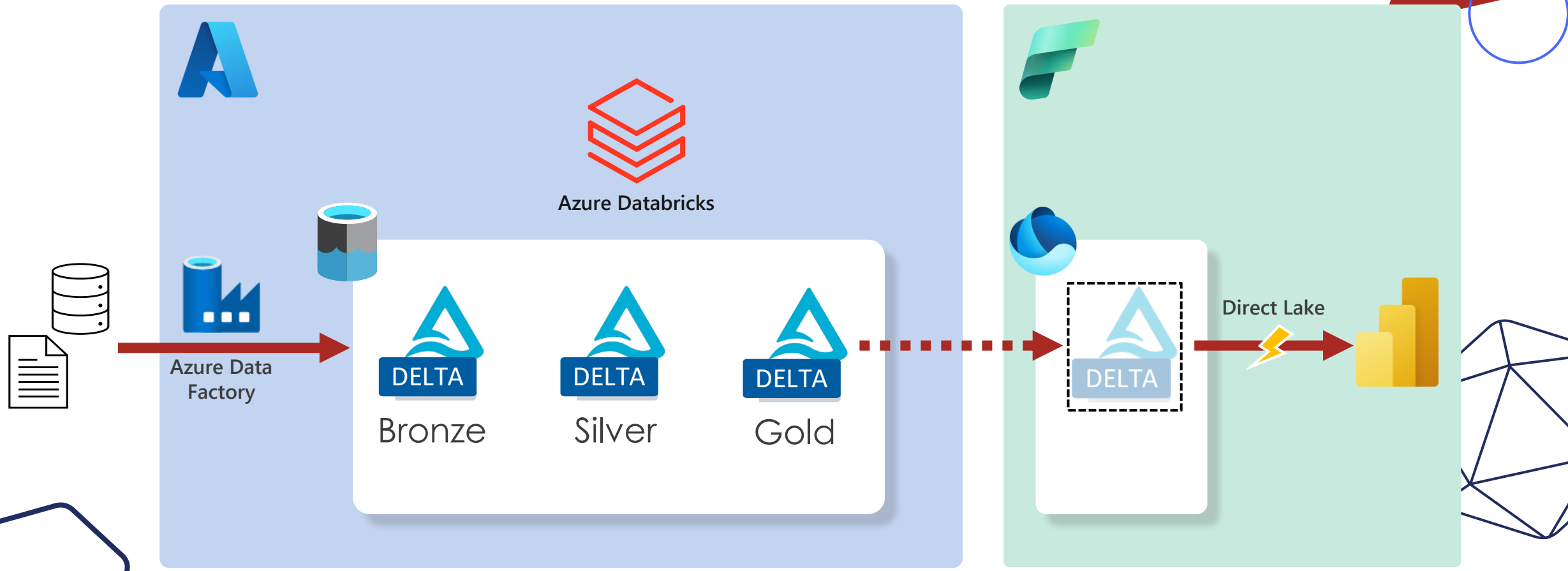
ADVANCING ANALYTICS

Serving in Fabric

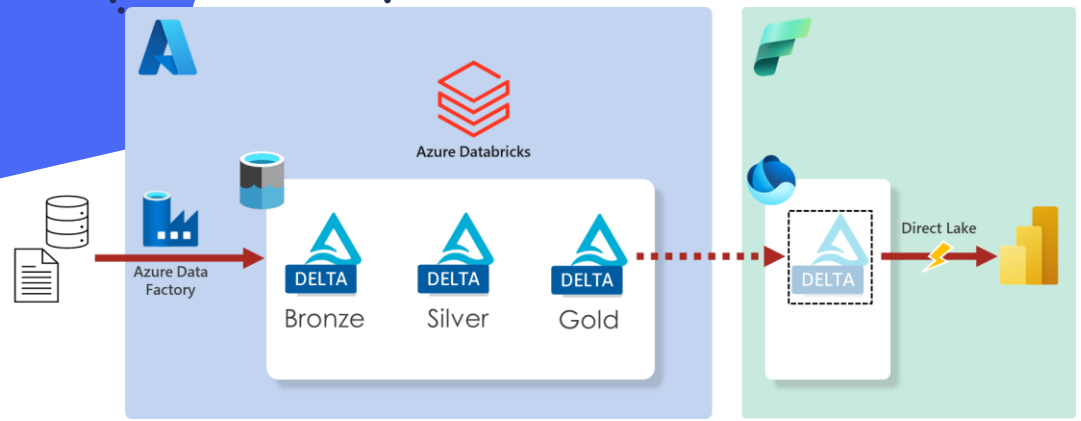


ADVANCING ANALYTICS

Serving in Fabric



ADVANCING ANALYTICS



Serving in Fabric

Pros

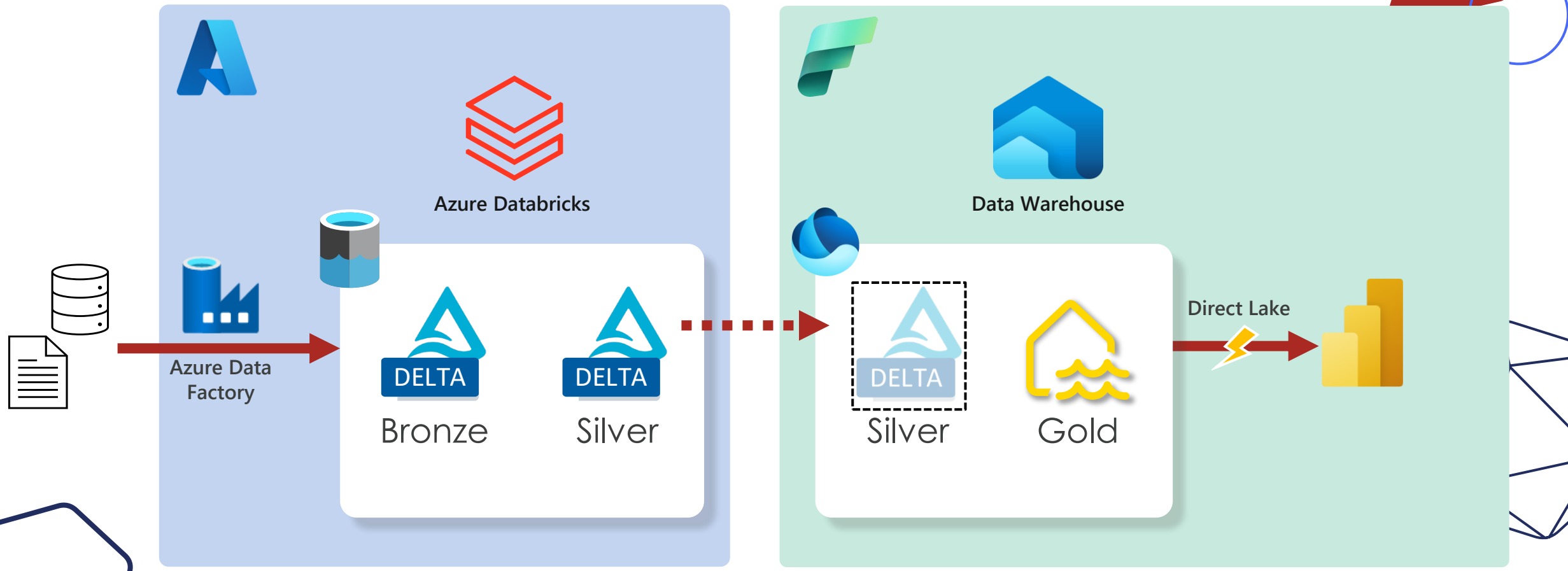
- Promoted by Microsoft and Databricks
- Retains current engineering patterns
- Direct Lake
- Single window for consumption
- Easy adoption

Cons

- Missing most Fabric features
- No V-Ordering
- Unity Catalog limitations
- Split security model
- Final mile potentially less accessible… (e.g. still quite a code heavy approach)

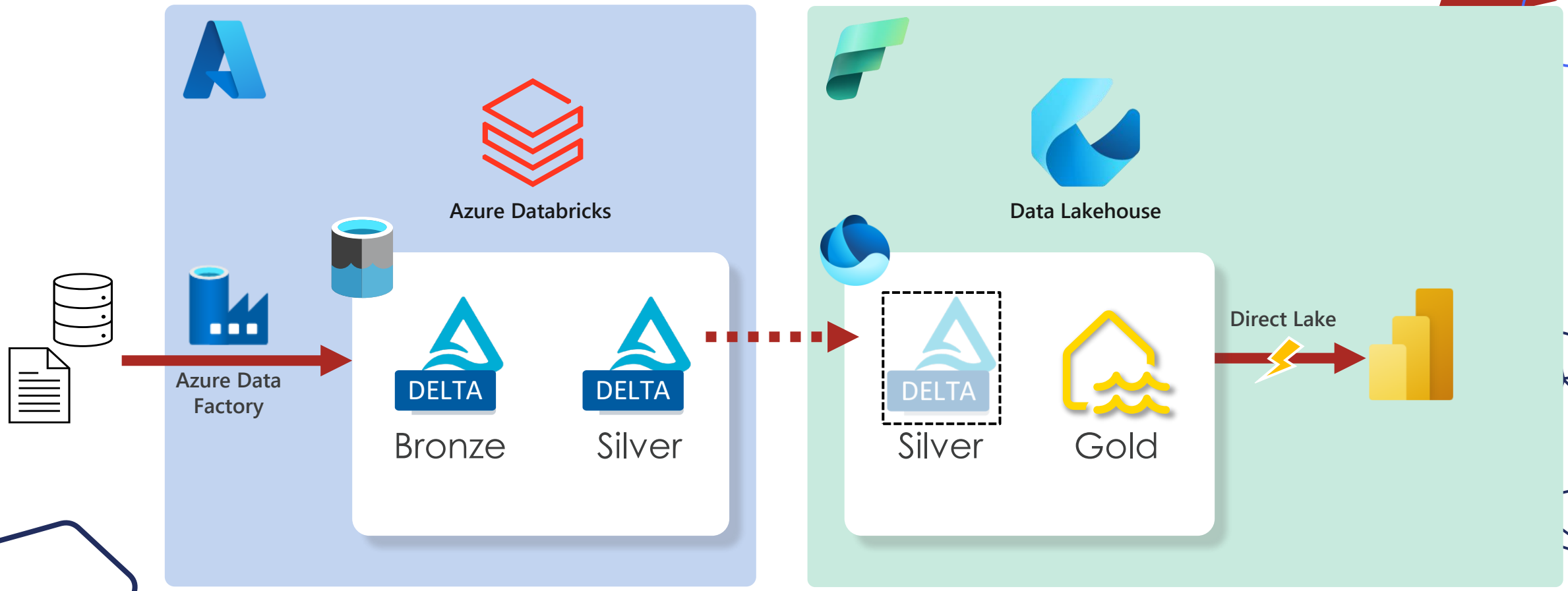
Analytics (Gold) in Fabric

Analytics in Fabric

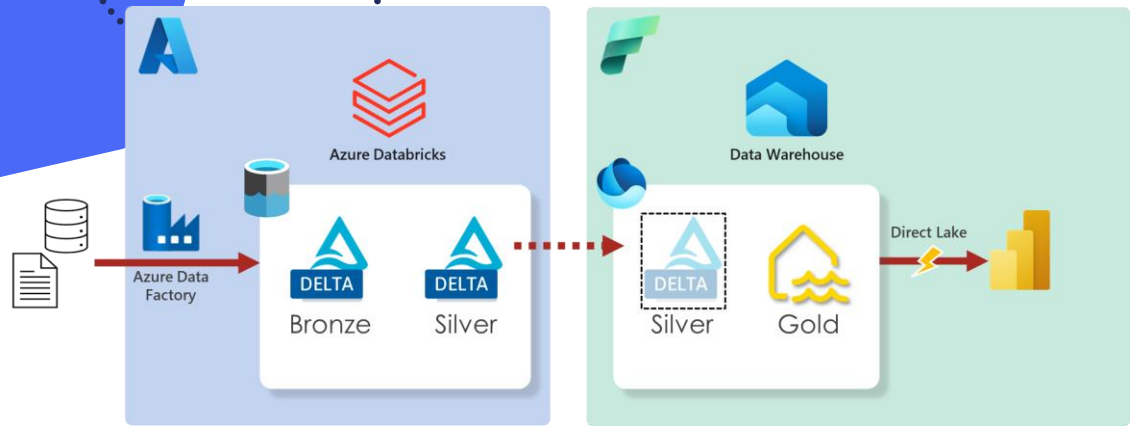


ADVANCING ANALYTICS

Analytics in Fabric



ADVANCING ANALYTICS



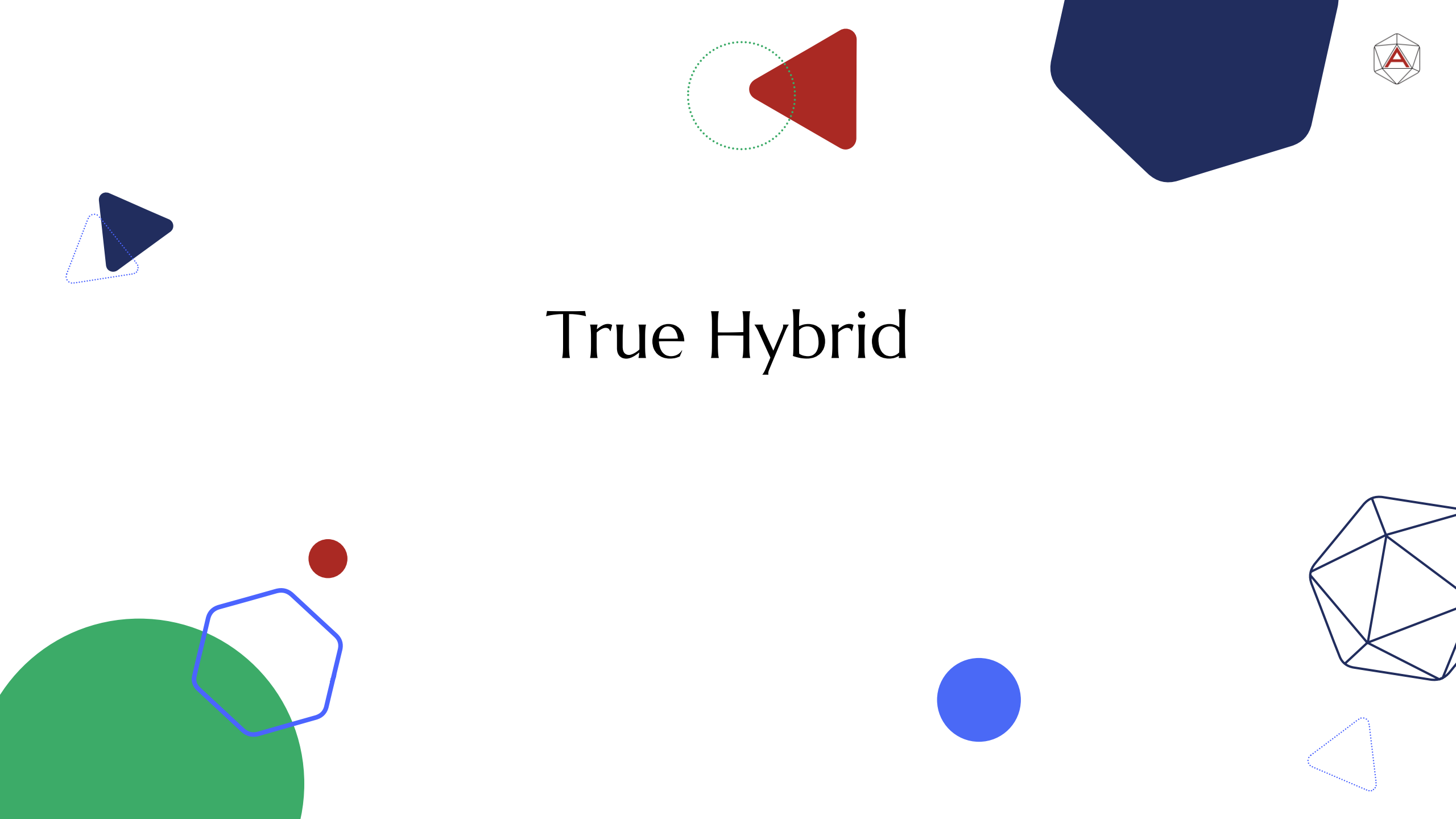
Analytics in Fabric

Pros

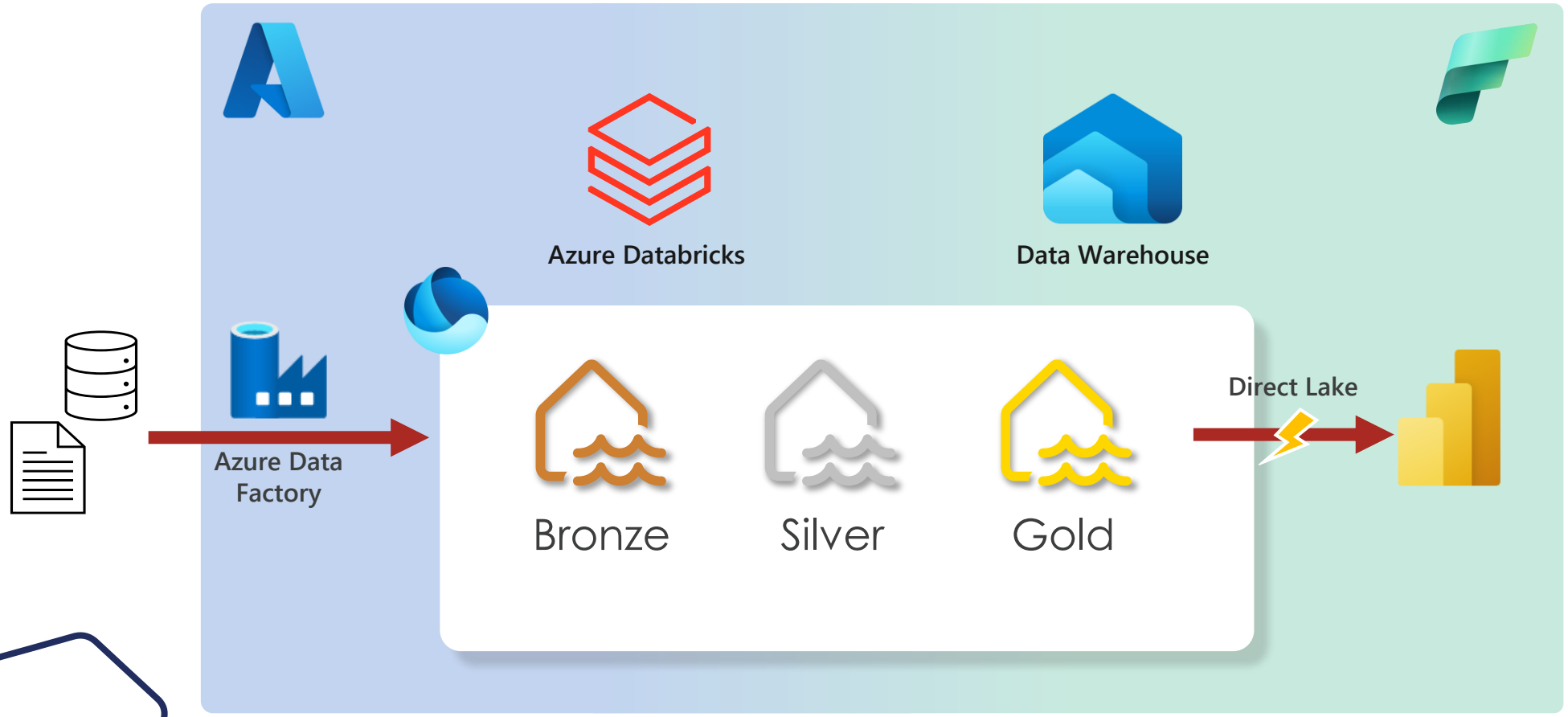
- Retains current ingestion patterns
- Direct Lake
- Final mile processing – V-ordering by default
- Wider range of processing options (T-sql, dataflows etc)
- OneLake adoption

Cons

- Split security model
- Split billing/cost profile (compute in fabric)
- Orchestration is… tricky
- Shortcuts in Unity Catalog (GUIDs)
- No single unified view of data
- Where do Data Scientists & Deep Analysts work?


True Hybrid

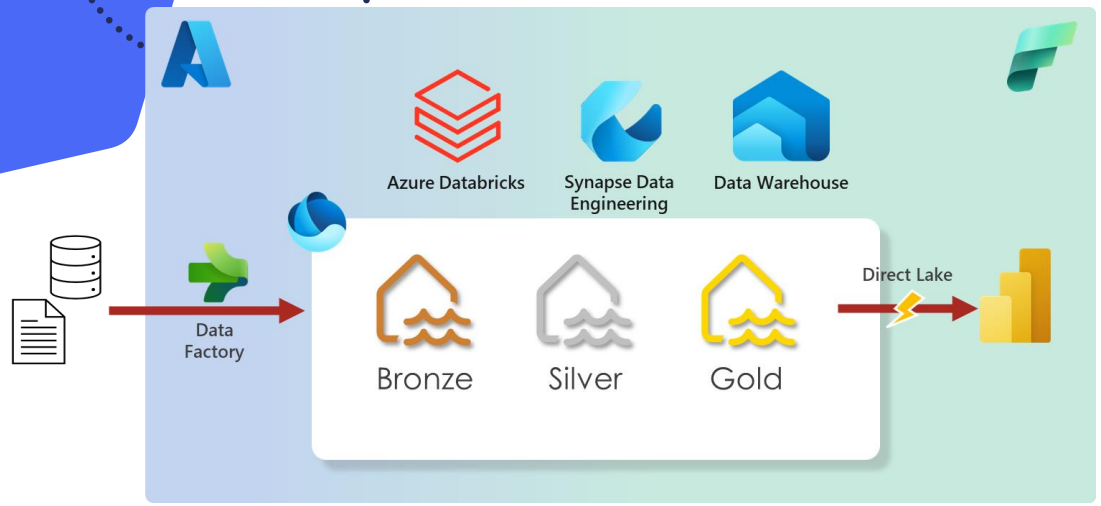
“Best tool for the job”



ADVANCING ANALYTICS

Best tool for the job

ADVANCING ANALYTICS



Best tool for the Job

Pros

- Retains current engineering patterns
- Direct Lake
- Full OneLake adoption – “single copy”
- Best tool for the job
- Single security model

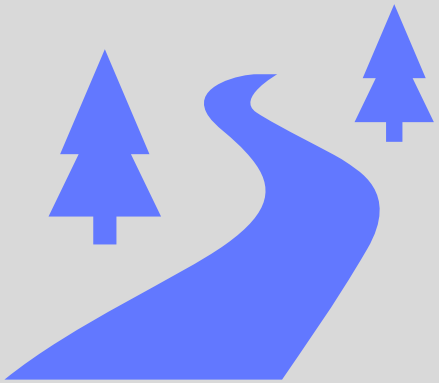
Cons

- Not an officially supported pattern
- Limited Unity Catalog support
- Split/complex billing/cost profile

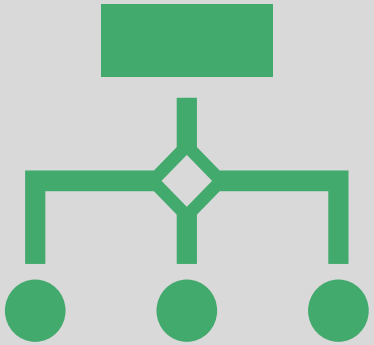
What else affects the decision?

Team Shape

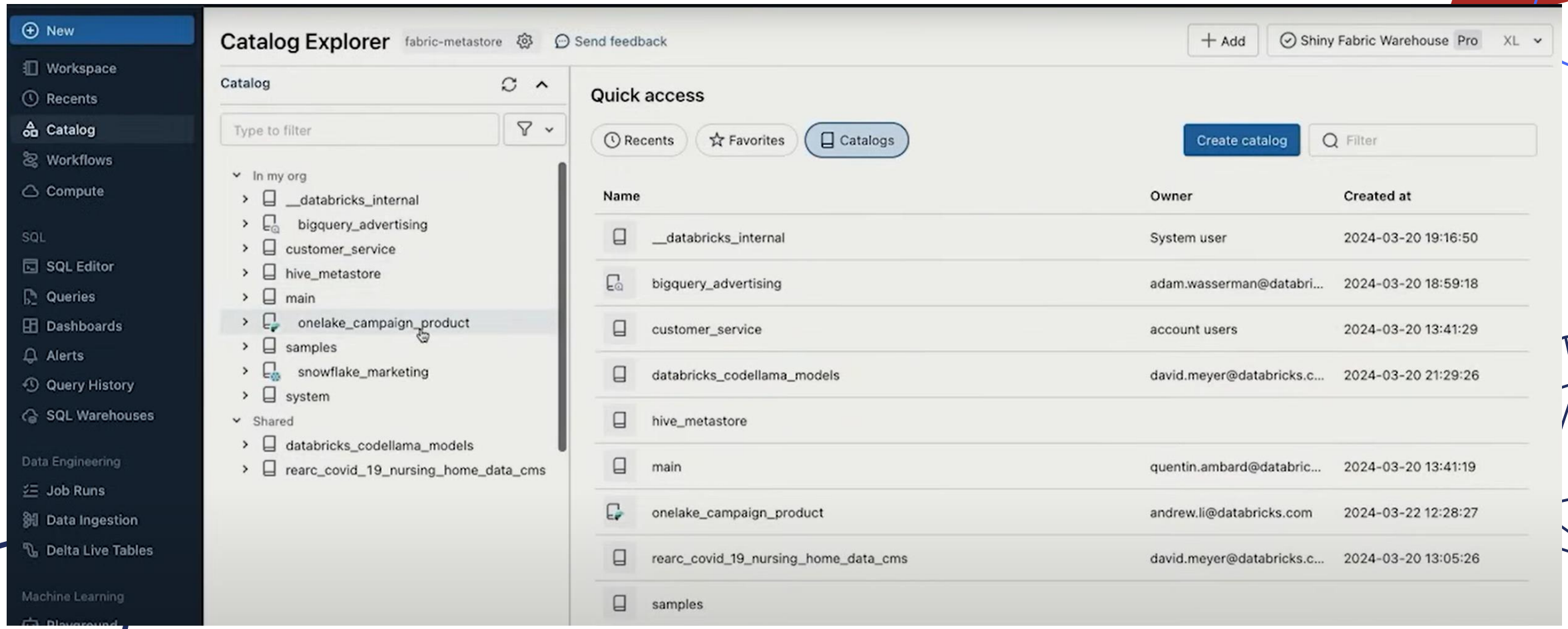
Stage in Journey



Operating Model

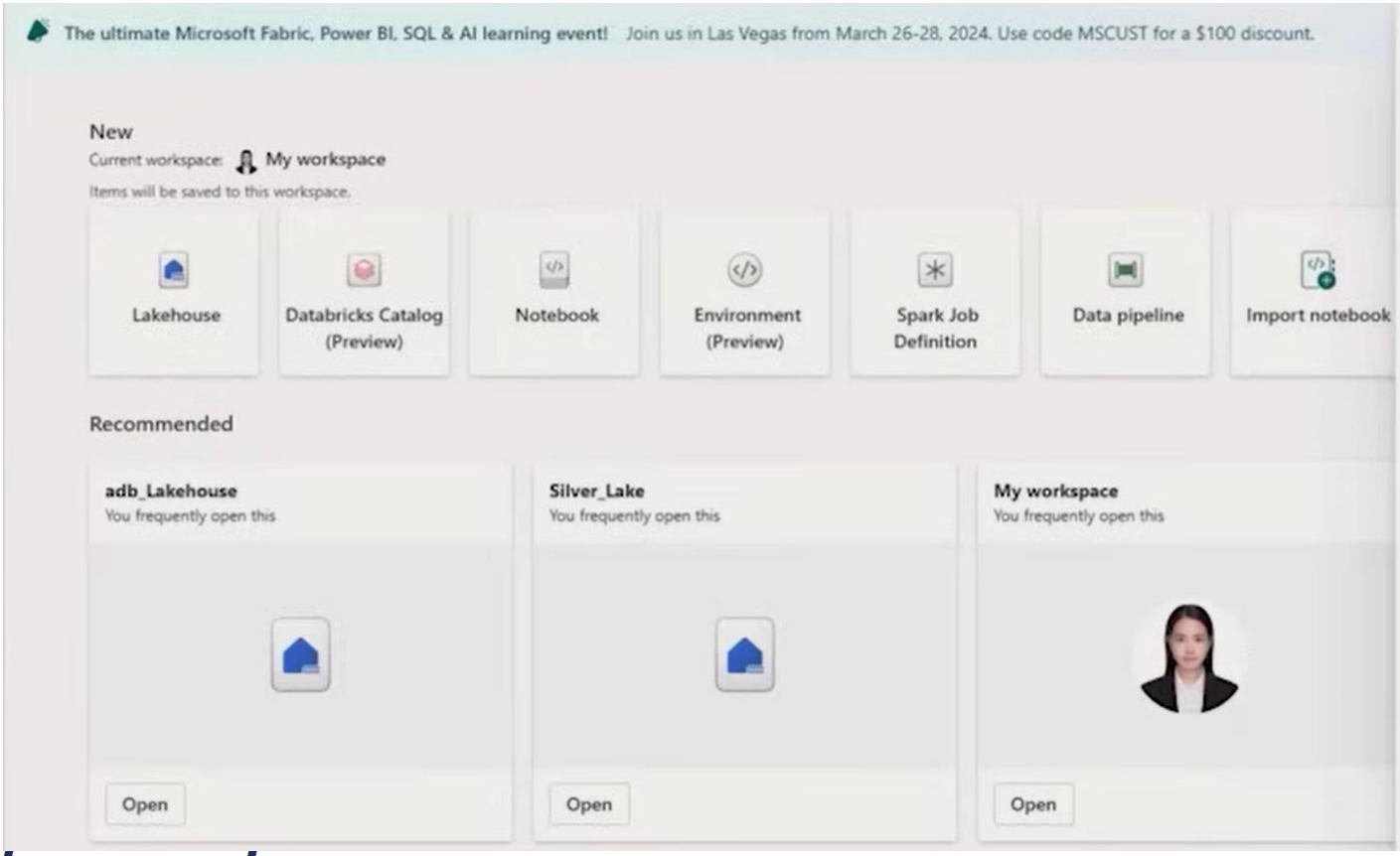


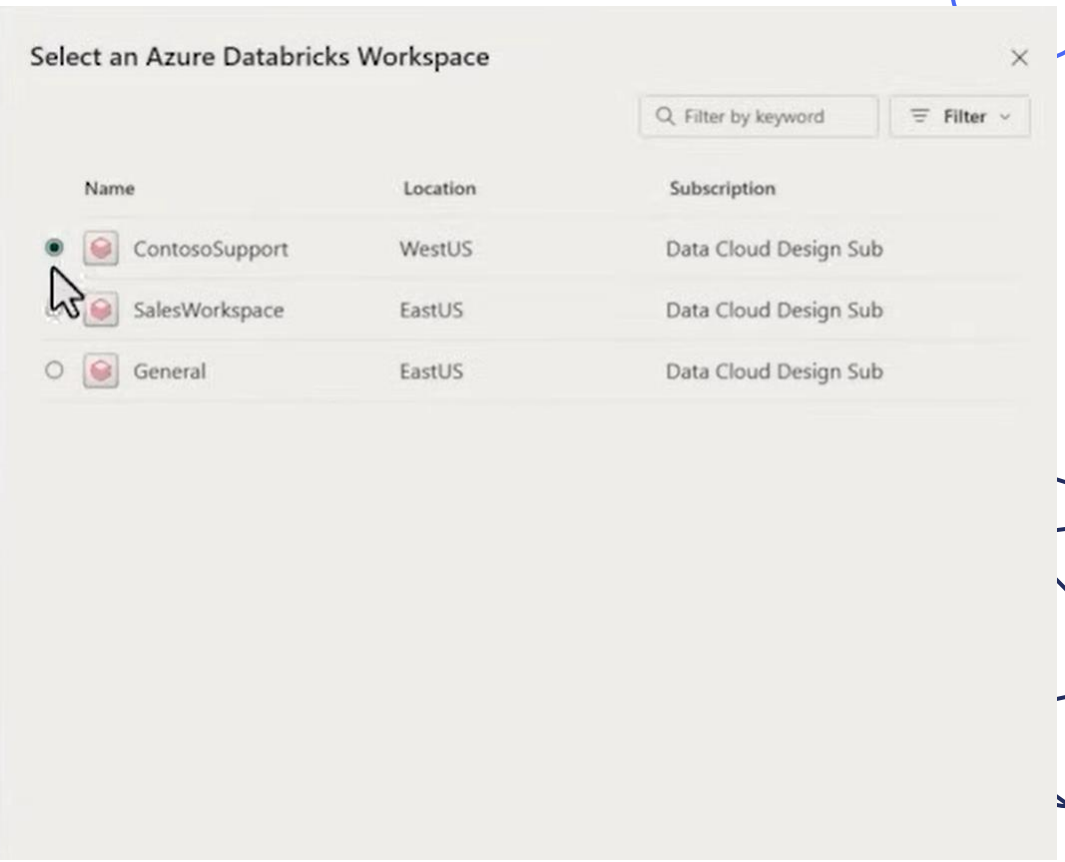
Always evolving



ADVANCING ANALYTICS

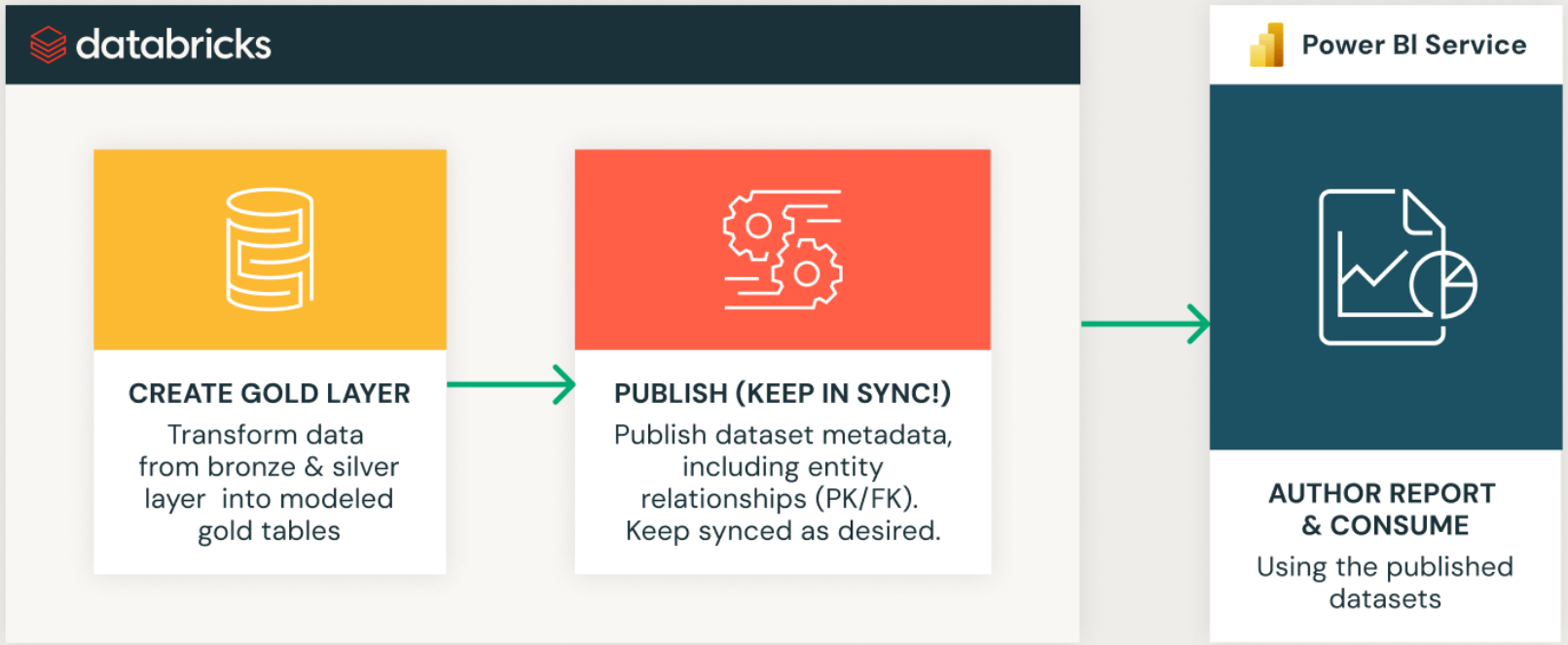
Always evolving





ADVANCING ANALYTICS

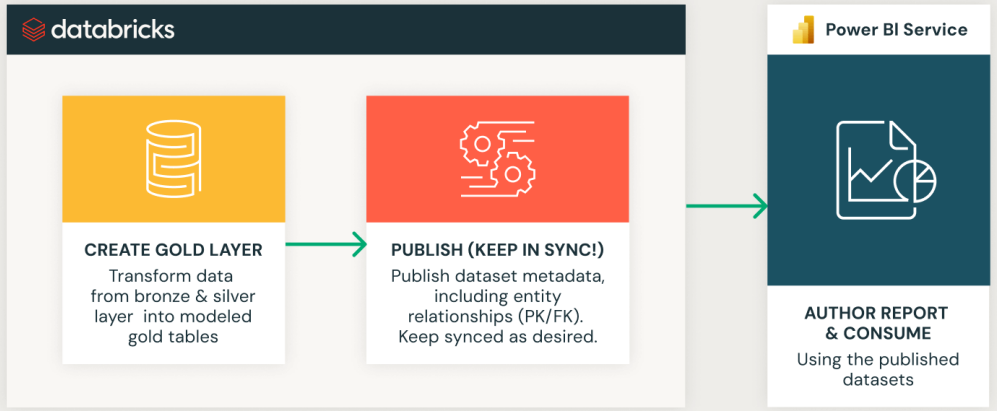
Publish to Power BI from Unity Cat



"This eliminates the need to switch contexts between Databricks and Power BI Desktop, significantly simplifying the process of making your data available for visualization and analysis."

ADVANCING ANALYTICS

Publish to Power BI from Unity Cat



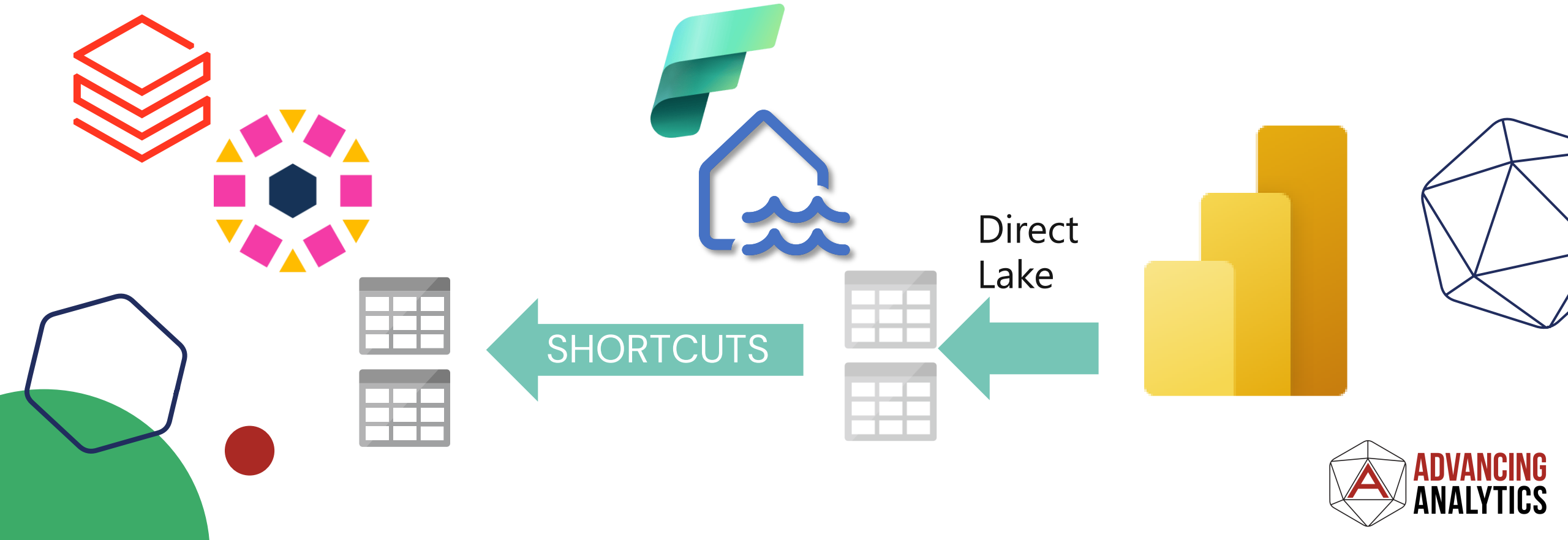
- Publish whole schemas as a data model
- Choose Import or DirectQuery mode
- Config Security as required

This option is going to use Databricks SQL to serve the data – so it's more of a shortcut to getting to the two previous approaches

ADVANCING ANALYTICS

Unity Catalog Mirroring in Fabric (Preview)

"When you use Fabric to read data that is registered in Unity Catalog, there is no data movement or data replication. Only the Azure Databricks catalog structure is mirrored to Fabric and the underlying catalog data is accessed through shortcuts."



ADVANCING ANALYTICS

Unity Catalog Mirroring in Fabric (Preview)

Network Limitations

- Azure Databricks workspaces shouldn't be behind a private endpoint.
- ADLS Gen 2 is utilized by your Azure Databricks workspace must also be accessible to Fabric.
- Storage accounts containing unity catalog data can't be behind a firewall.
- Azure Databricks IP Access lists aren't supported.

Table Limitations

The following table types are not supported:

- Tables with RLS/CLM policies
- Lakehouse federated tables
- Delta sharing tables
- Streaming tables
- Views, Materialized views

ADVANCING ANALYTICS

Which architecture is best?

It depends!

Understand your requirements

Understand your team and organisation's strengths

Understand how each platform can best support these

Thank You

CRAIG PORTEOUS

Assoc. Head of Data Engineering

https://craigporteous.com

@cporteous

https://github.com/cporteou

ADVANCING ANALYTICS

Share **your thoughts** and help our speakers!

fabfeb.app/feedback

#FabricFebruary

ADVANCING ANALYTICS