Microsoft

GPS Data/AI Strategy FY23

Delivered by CSA Team

22/09/2022

Franck Gaillard
Cloud Solution Architect
Data AI
frgail@microsoft.com

Narjes Majdoub
Cloud Solution Architect
Data AI
nmajdoub@microsoft.com

Ali Bouhaddou
Cloud Solution Architect
Data Analytics
albouhad@microsoft.com

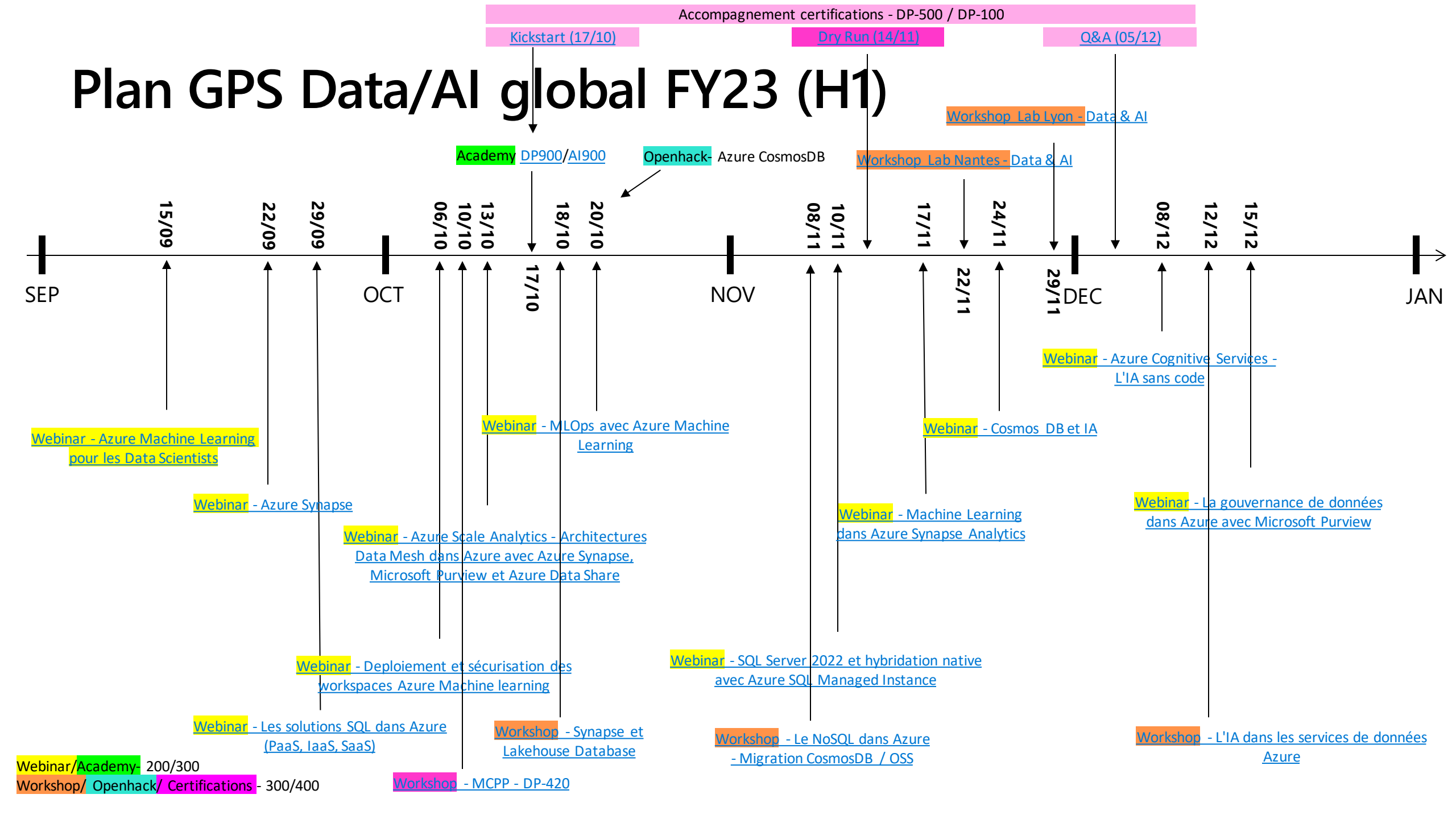
Frederic Gisbert
Cloud Solution Architect
Data Analytics
frgisber@microsoft.com

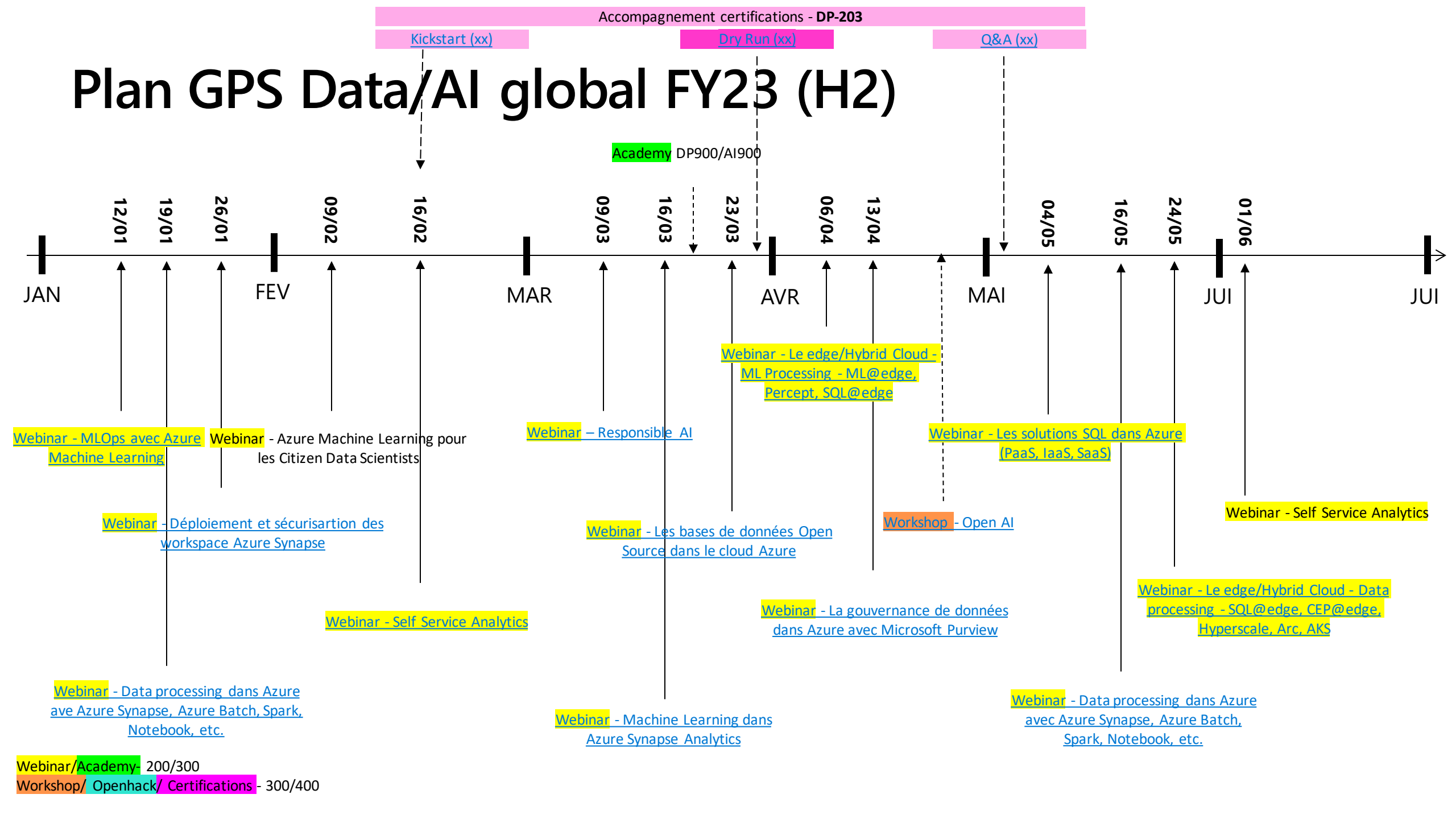
Azure Data & AI technical intensity plan

- From June 2022 to June 2023
- Focus on "Azure Data & AI" tech intensity

- Many content, from L100 Beginner to L400 Expert level
  - Academy L100
  - Webinar L200/L300
  - Workshop L300/L400
  - Certification kickstart L300/L400
  - Openhack / Microhack L400

Plan GPS Data/AI global FY23 (H1)




Accompagnement certifications - DP-203
Kickstart (xx)
Dry Run (xx)
Q&A (xx)
Plan GPS Data/AI global FY23 (H2)
Academy DP900/AI900
12/01
19/01
26/01
09/02
16/02
09/03
16/03
23/03
06/04
13/04
04/05
16/05
24/05
01/06
JAN
FEV
MAR
AVR
MAI
JUI
JUI
Webinar - MLOps avec Azure Machine Learning
Webinar - Azure Machine Learning pour les Citizen Data Scientists
Webinar – Responsible AI
Webinar - Le edge/Hybrid Cloud - ML Processing - ML@edge, Percept, SQL@edge
Webinar - Les solutions SQL dans Azure (PaaS, IaaS, SaaS)
Webinar - Déploiement et sécurisartion des workspace Azure Synapse
Webinar - Les bases de données Open Source dans le cloud Azure
Workshop - Open AI
Webinar - Self Service Analytics
Webinar - Self Service Analytics
Webinar - La gouvernance de données dans Azure avec Microsoft Purview
Webinar - Le edge/Hybrid Cloud - Data processing - SQL@edge, CEP@edge, Hyperscale, Arc, AKS
Webinar - Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.
Webinar - Machine Learning dans Azure Synapse Analytics
Webinar - Data processing dans Azure avec Azure Synapse, Azure Batch, Spark, Notebook, etc.
Webinar/Academy- 200/300
Workshop/ Openhack/ Certifications - 300/400

Liste des évènements de type Webinar 2H

Event Webinar (Les jeudis de la Data & AI) - L200/300	Date	Duration (min)	Link
Azure Machine Learning pour les Data Scientists	15/09/2022	120	https://msevents.microsoft.com/event?id=2454281594
Azure Synapse	22/09/2022	120	https://msevents.microsoft.com/event?id=857781749
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	29/09/2022	120	https://msevents.microsoft.com/event?id=502366997
Déploiement et sécurisation des workspaces Azure Machine learning	06/10/2022	120	https://msevents.microsoft.com/event?id=1505714138
Azure Scale Analytics - Architectures Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share	13/10/2022	120	https://msevents.microsoft.com/event?id=139685175
MLOps avec Azure Machine Learning	20/10/2022	120	https://msevents.microsoft.com/event?id=1245885767
SQL Server 2022 et hybridation native avec Azure SQL Managed Instance	10/11/2022	120	https://msevents.microsoft.com/event?id=145826476
Machine Learning dans Azure Synapse Analytics	17/11/2022	120	https://msevents.microsoft.com/event?id=3637723312
Azure Cosmos DB et IA	24/11/2022	120	https://msevents.microsoft.com/event?id=2646013445
Azure et les Services Cognitifs	08/12/2022	120	https://msevents.microsoft.com/event?id=3772037220
La gouvernance de données dans Azure avec Microsoft Purview	15/12/2022	120	https://msevents.microsoft.com/event?id=1499560981
MLOps avec Azure Machine Learning	12/01/2023	120	https://msevents.microsoft.com/event?id=4115194515
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	19/01/2023	120	https://msevents.microsoft.com/event?id=1537241181
Déploiement et sécurisation des workspace Azure Synapse	26/01/2023	120	https://msevents.microsoft.com/event?id=1806467748
Azure Machine Learning pour les Citizen Data Scientists	09/02/2023	120	En cours
PowerBI - Self Service Analytics	16/02/2023	120	https://msevents.microsoft.com/event?id=1401519679
L'IA responsable avec Azure machine learning	09/03/2023	120	https://msevents.microsoft.com/event?id=2072953112
Machine Learning dans Azure Synapse Analytics	16/03/2023	120	https://msevents.microsoft.com/event?id=3413014857
Les bases de données Open Source dans le cloud Azure	23/03/2023	120	https://msevents.microsoft.com/event?id=2727487131
Hybridation des services de Machine Learning Azure	06/04/2023	120	https://msevents.microsoft.com/event?id=1624914222
La gouvernance de données dans Azure avec Microsoft Purview	13/04/2023	120	https://msevents.microsoft.com/event?id=3909342839
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	04/05/2023	120	https://msevents.microsoft.com/event?id=1162207895
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	16/05/2023	120	https://msevents.microsoft.com/event?id=3517068442
Hybridation des services de données Azure	24/05/2023	120	https://msevents.microsoft.com/event?id=2996507398
Self Service Analytics	01/06/2023	120	En cours
Total	**25**		

Liste des évènements de type Workshop/Prepa Cert/Academy

Event Workshop L300/400	Date	Duration (min)	Link
Synapse et Lakehouse Database	18/10/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Le NoSQL dans Azure - Migration CosmosDB / OSS	08/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Lab Lyon - Data & AI	22/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUMlZZOURET0RSWjcyTERYRkJGTlFFUjJaUi4u
Lab Nantes - Data & AI	29/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUMlZZOURET0RSWjcyTERYRkJGTlFFUjJaUi4u
L'IA dans les services de données Azure	12/12/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Open AI	H2	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u

Event Academy, kickstart certifications, workshop certifications	Date	Duration (min)	Link
MCPP - DP-420	10/10/2022	420	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUMkJISlRKSU1RRFA0OVgzSFdTSTY0RE9WQy4u
Micro Hack CosmosDB	20/10/2022	420	H1 - Inscriptions PTA
Academy DP900	17-21/10/2022	300	https://msevents.microsoft.com/event?id=3250818161
Academy AI900	17-21/10/2022	300	https://msevents.microsoft.com/event?id=2717528090
Kickstart DP-500	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3TE9ZRS4u
Dry Run DP-500	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3TE9ZRS4u
Q&A DP-500	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3TE9ZRS4u
Kickstart DP-100	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMUFYxSkRIMi4u
Dry Run DP-100	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMUFYxSkRIMi4u
Q&A DP-100	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMUFYxSkRIMi4u
Kickstart DP-203	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u
Dry Run DP-203	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u
Q&A DP-203	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u

Total 16

Microsoft Azure Solutions Business

Azure Synapse Analytics

La modélisation de Datawarehouse n'a plus de secret pour vous ? Et si on parlait de la nouvelle approche de modélisation d'objets dans Azure Synapse.
Nous parcourrons les notions de virtualisation, de lake database et de l'orientation Spark de la plateforme.

Agenda (2h)

Service overview / updates

- Data Intégration
- Data Engineering
- Data Warehousing
- Data Science
- Observational Analytics
- Business Intelligence
- Gouvernance


Ali Bouhaddou
Cloud Solution Architect
Data Analytics


Frederic Gisbert
Cloud Solution Architect
Data Analytics


Franck Gaillard
Cloud Solution Architect
Data AI


Narjes Majdoub
Cloud Solution Architect
Data AI

Microsoft

Azure Synapse Analytics

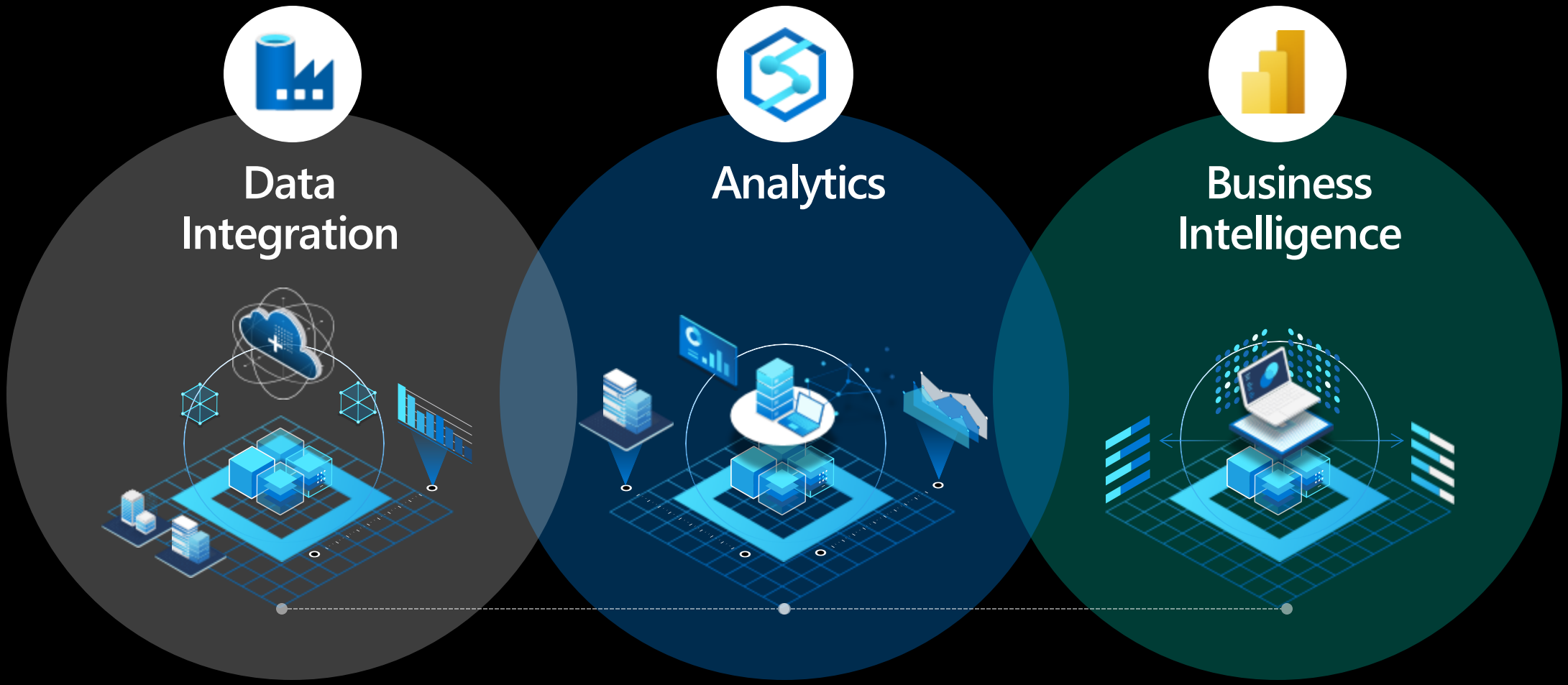
Service overview / updates March 2022

From Data to Intelligence

Drive a data culture and power a new class of data first applications

everyone | every decision | at any scale

Azure Synapse and Power BI



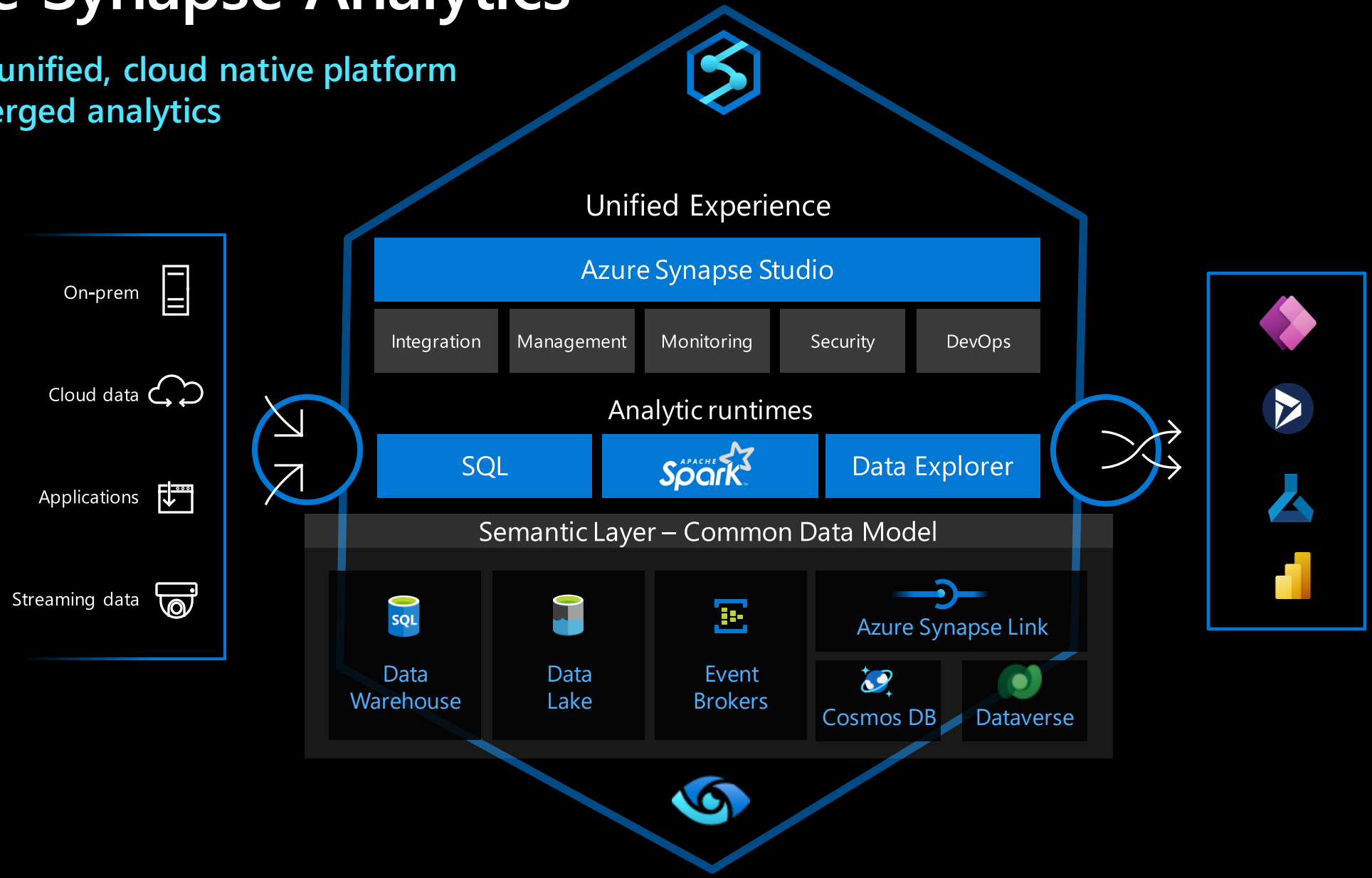
Azure Synapse Analytics

The first unified, cloud native platform for converged analytics

Azure Synapse is the only unified platform for analytics, blending big data, data warehousing, and data integration into a **single cloud native service** for end-to-end analytics at cloud scale.

Azure Synapse Analytics

The first unified, cloud native platform for converged analytics



Synapse + Power BI

- Data Integration
- Data Engineering
- Data Warehousing
- Data Science
- Observational Analytics
- Business Intelligence
- Governance

Data Integration

Over 100 connectors to ingest data from a variety of platforms

Integrate from On-Premise, PaaS, and SaaS

Batch and Real-time data integration

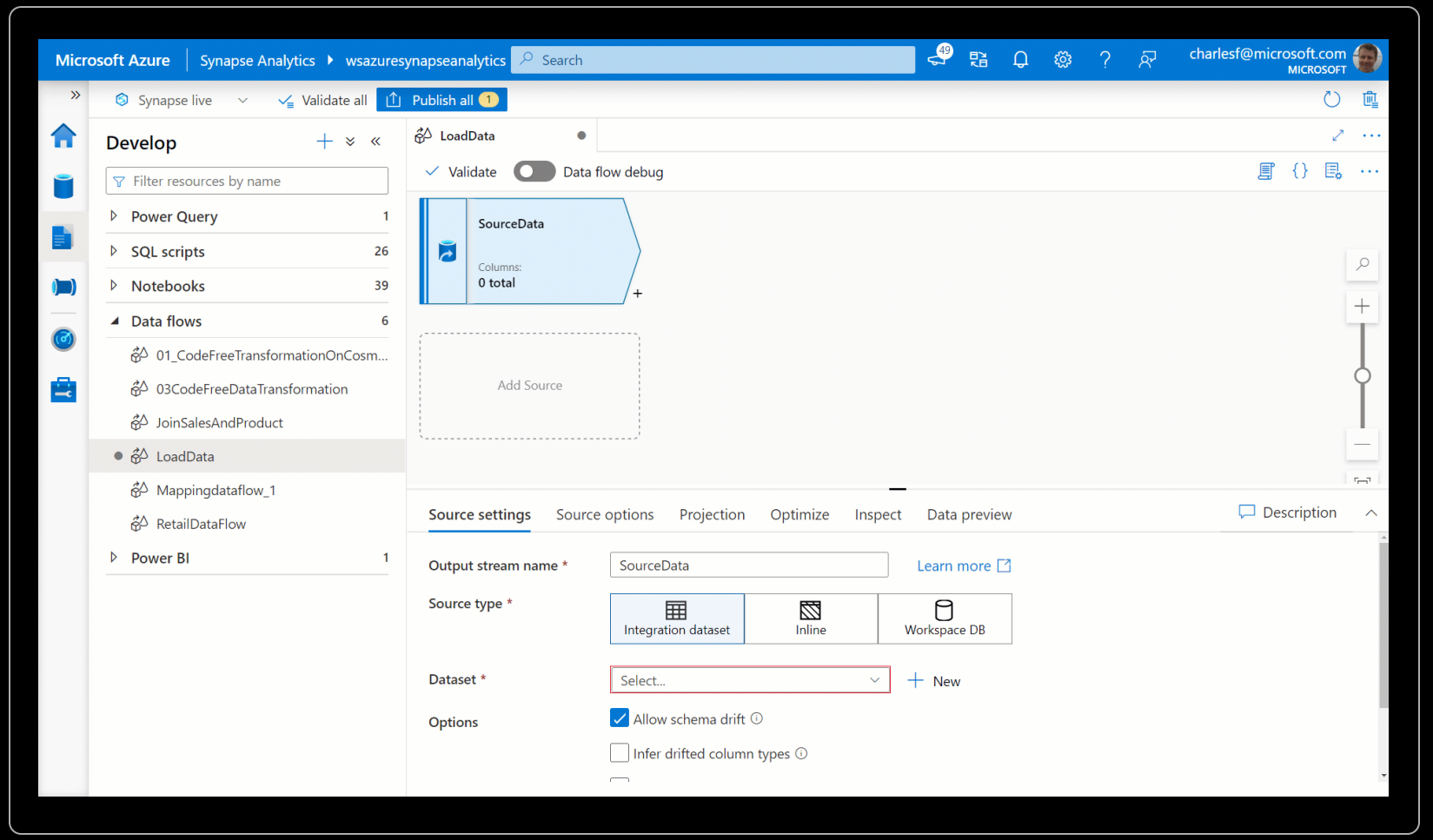
Secure hybrid connectivity

Code-free development environment

Generally Available

100+ Connectors

Connect on data sources in Azure, on-premise, other clouds, and SaaS applications

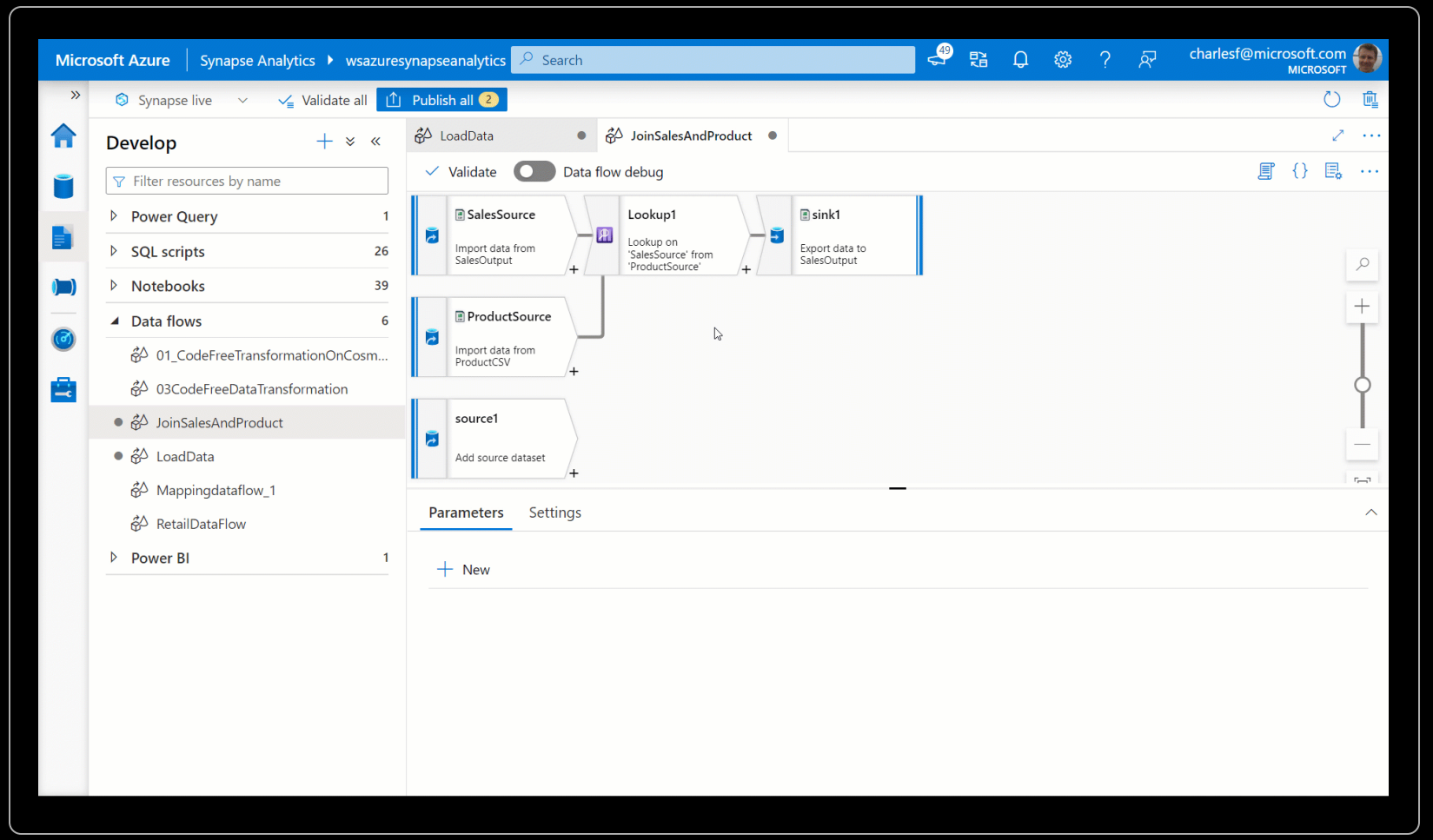


Generally Available

Code-free Data Flows

Enables developers to rapidly integrate data from a variety of sources

Execute on Spark for large scale processing



Code-free Data Flows

Handle upserts, updates, deletes on sql sinks

Add new partition methods

Add schema drift support

Add file handling (move files after read, write files to file names described in rows etc)

New inventory of functions (for e.g. Hash functions for row comparison)

Commonly used ETL patterns(Sequence generator/Lookup transformation/SCD…)

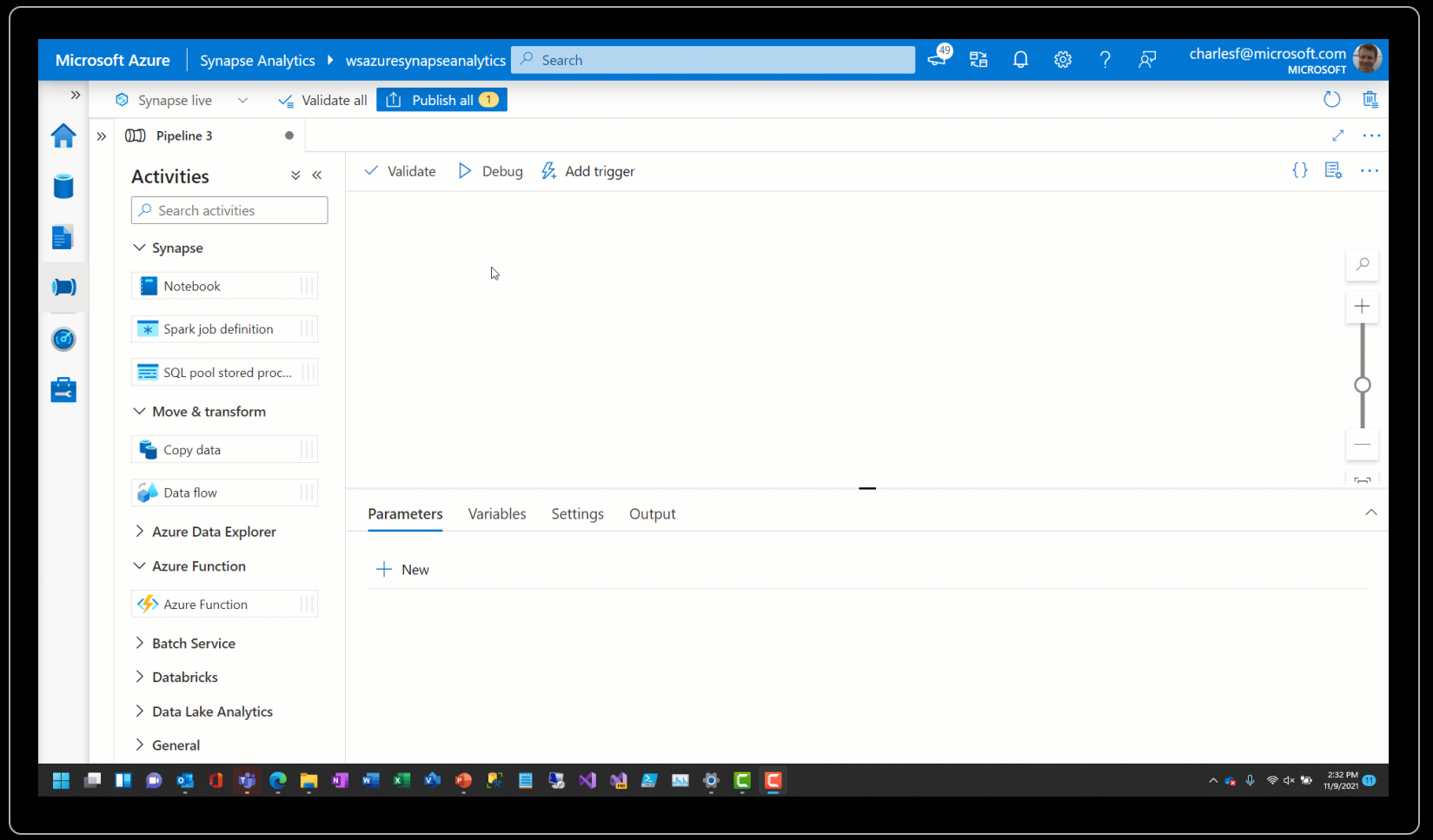
Data lineage – Capturing sink column lineage & impact analysis(invaluable if this is for enterprise deployment)

Implement commonly used ETL patterns as templates(SCD Type1, Type2, Data Vault)

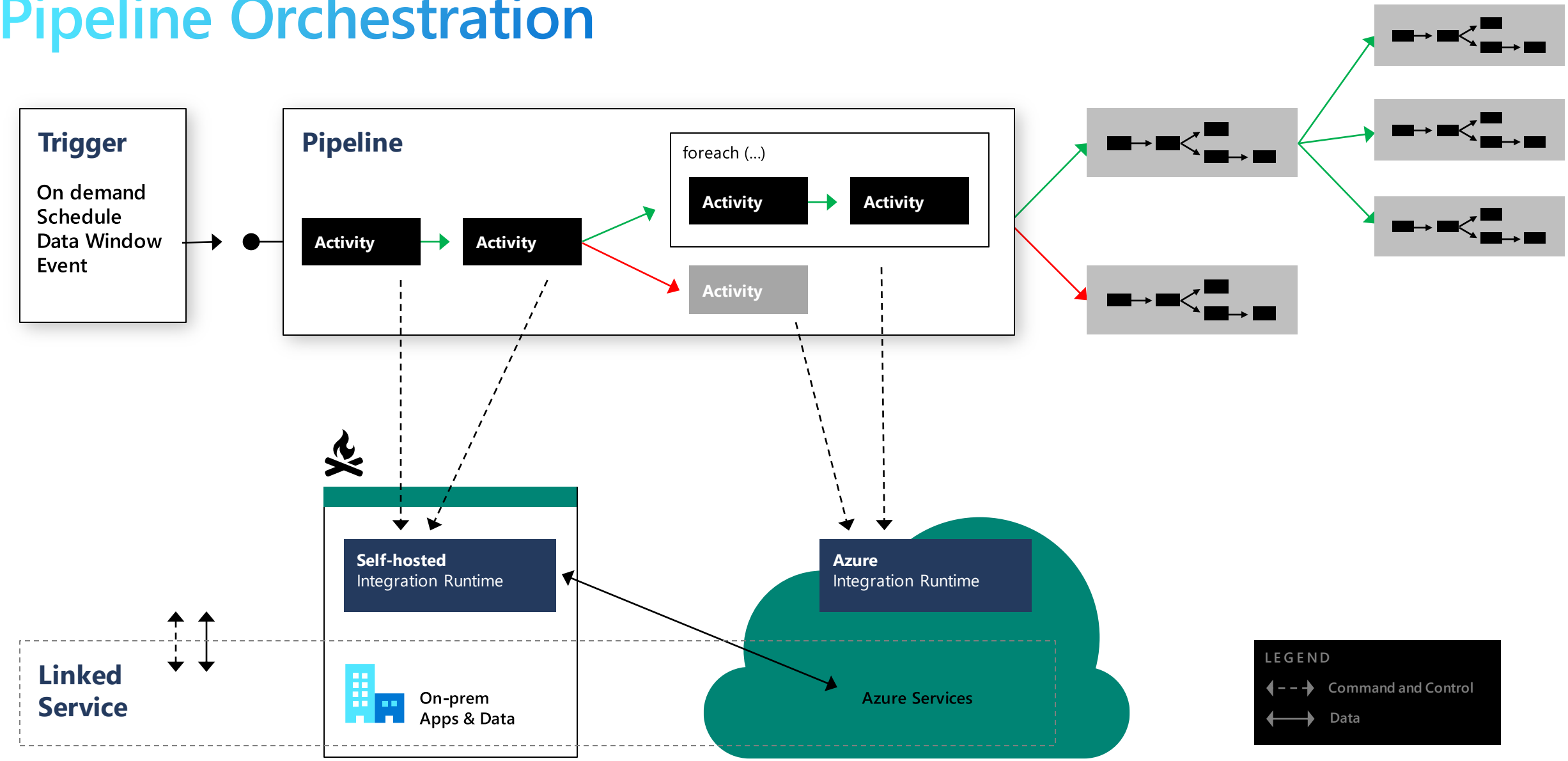
Generally Available

Pipeline Orchestration

Code-free experience for orchestrating a sequence of data integration tasks



Pipeline Orchestration

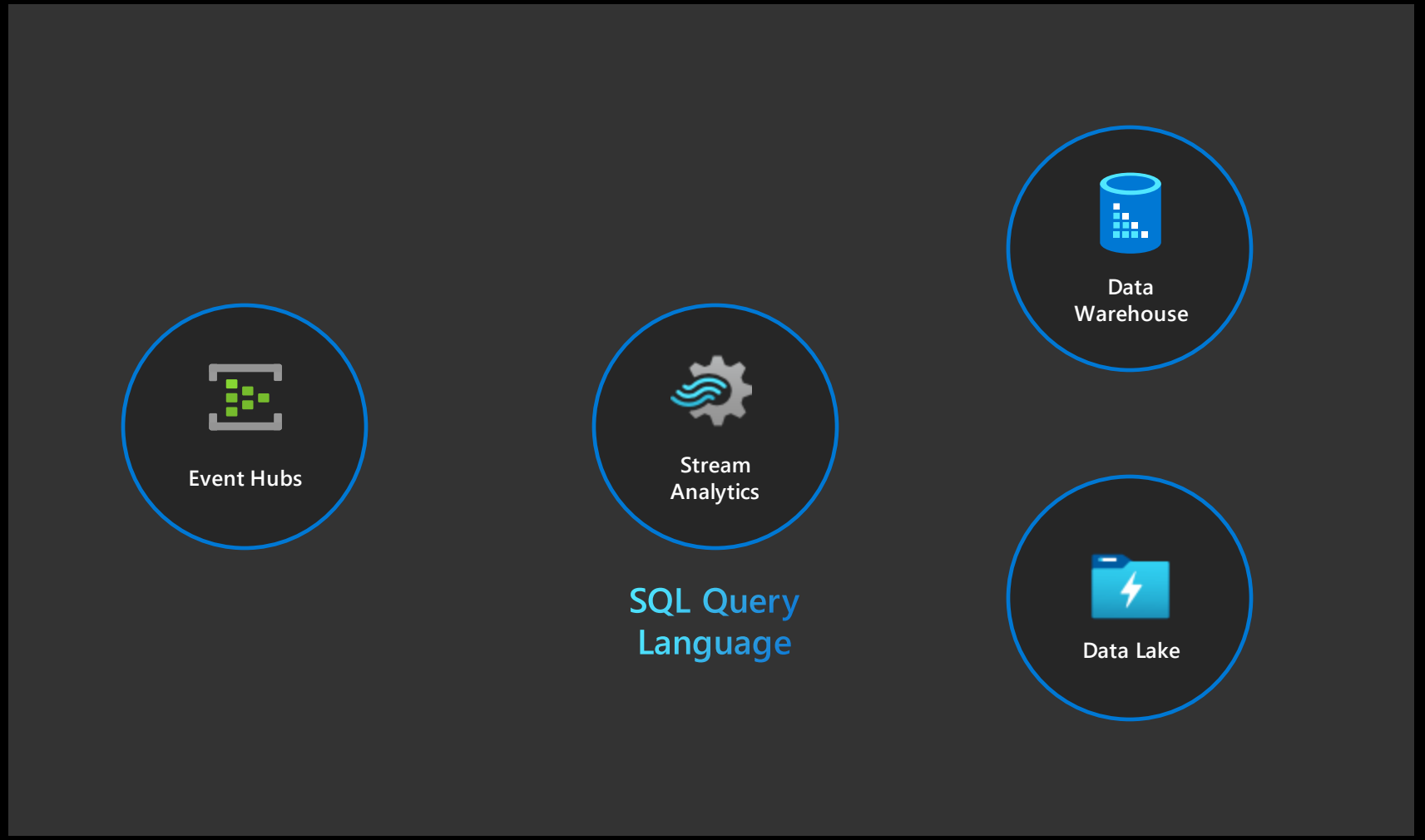


Generally Available

Real-time Streaming Data Integration

Enables IoT data streams from event brokers to load directly into the data warehouse or data lake

Analyze data in-flight with temporal T-SQL queries in Stream Analytics



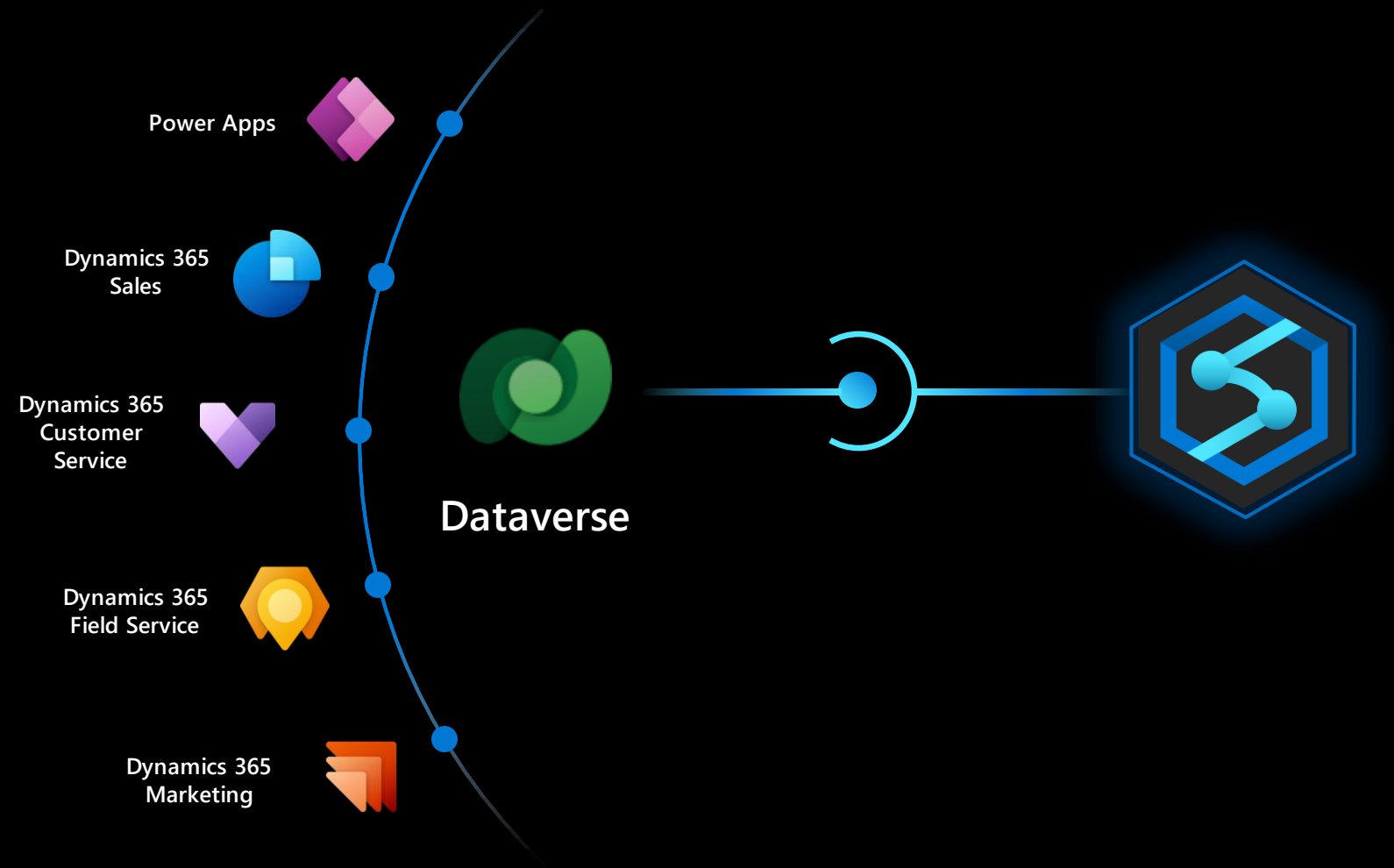
Generally Available

November 2021

Synapse Link for Dataverse

One-click integration of D365 data into Synapse for analytics

No data pipelines required



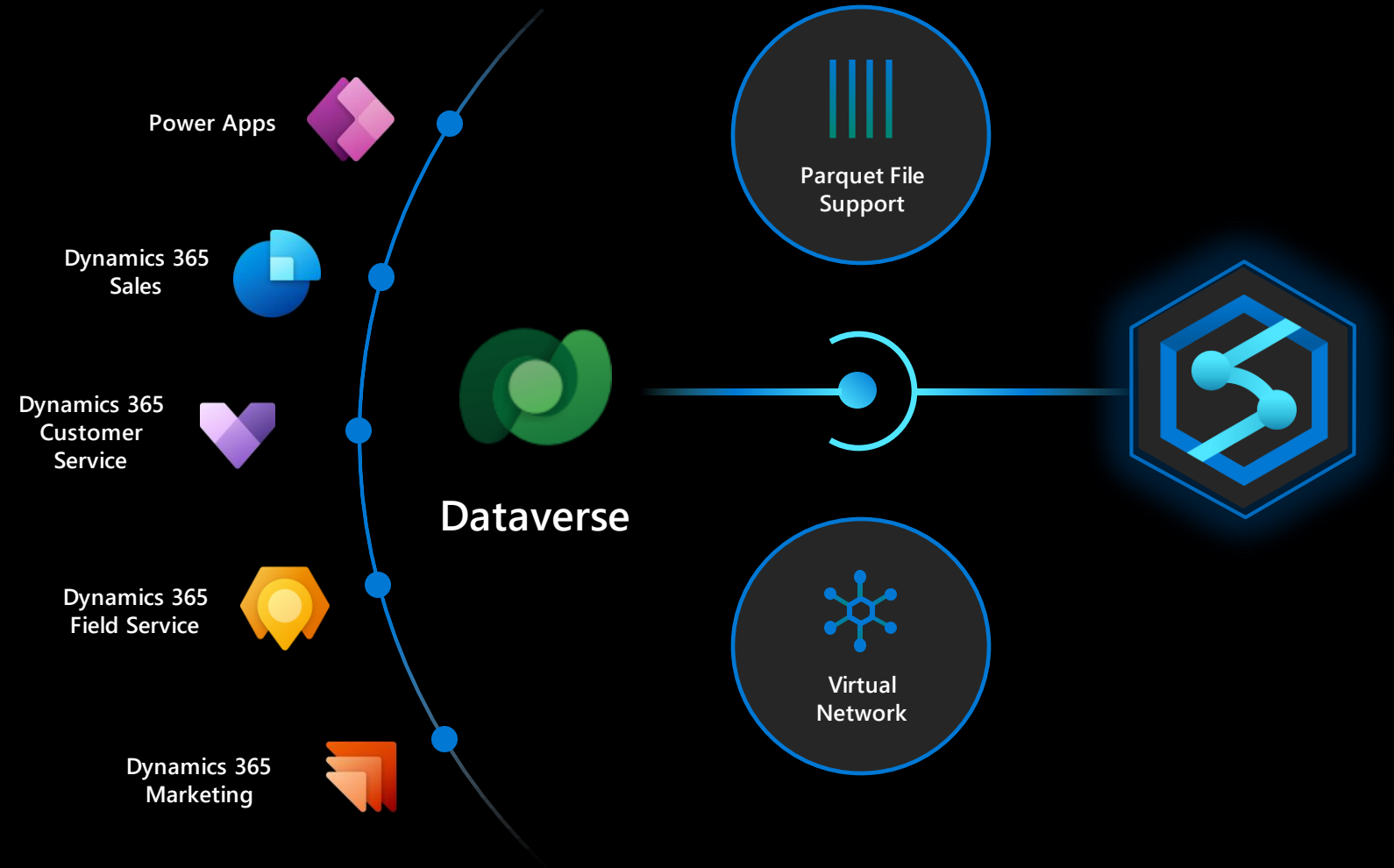
Public Preview

Q2 2022

Parquet & Virtual Network Support for Dataverse

Parquet columnar file format optimizes query performance for user queries

Enables customers to apply Virtual Network security to Dataverse connection



Public Preview

Q2 2022
(SQL Server 2022 and
Azure SQL Database)

Synapse Link for Microsoft SQL

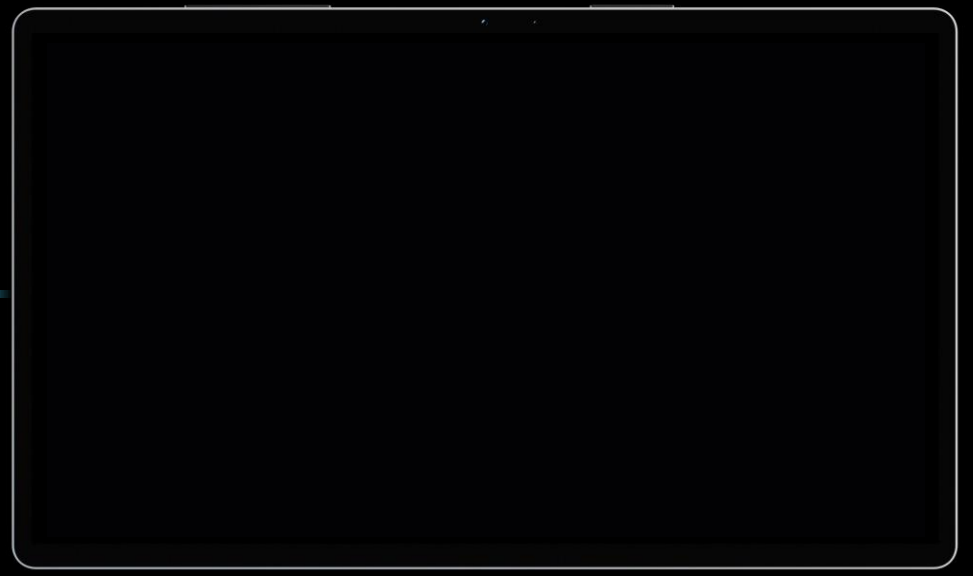
Near real-time operational analytics in Synapse

No data pipelines required

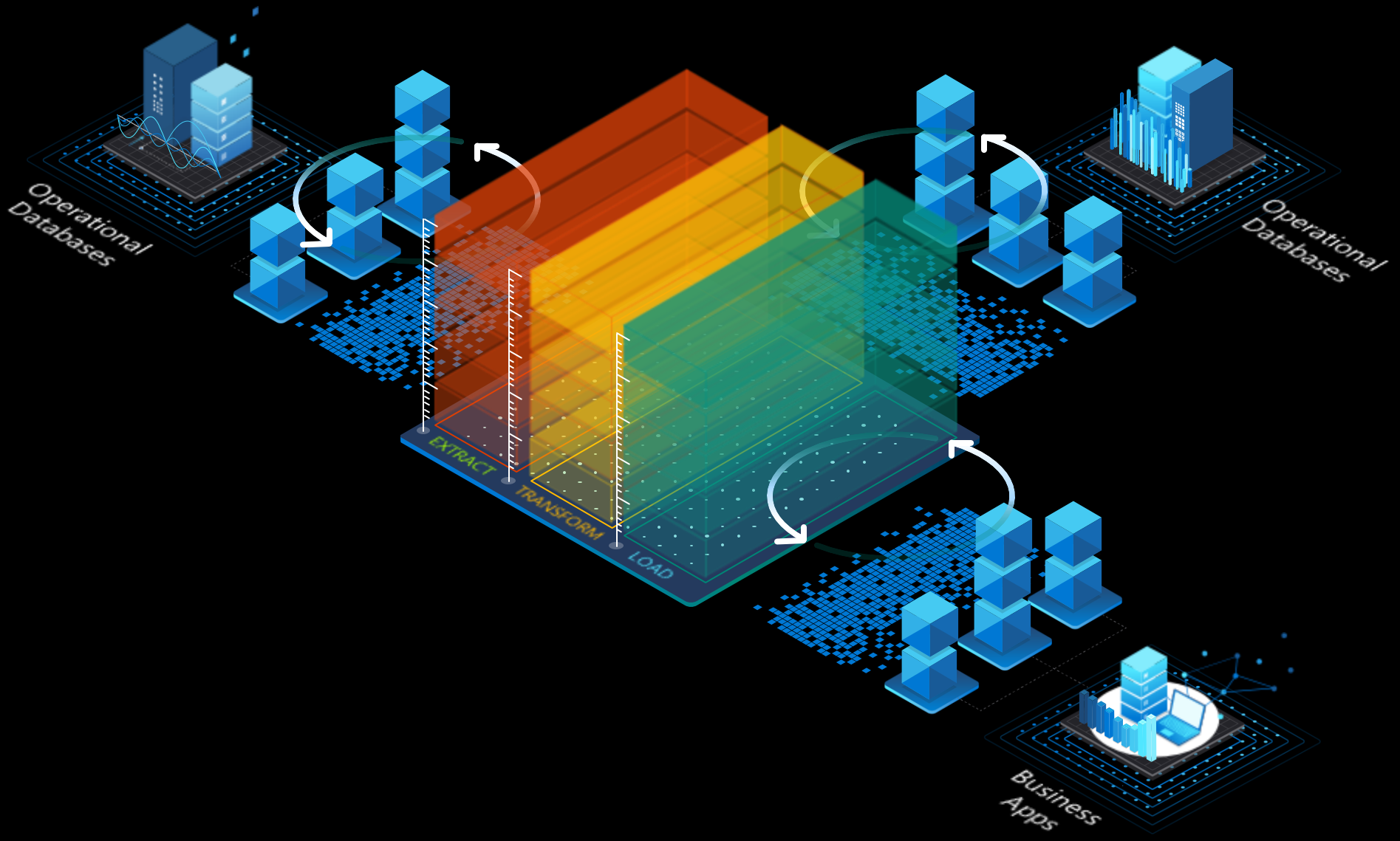
Hybrid integration for SQL Server running on-premise or in other clouds

SQL Server 2022
Public Preview Q2 2022

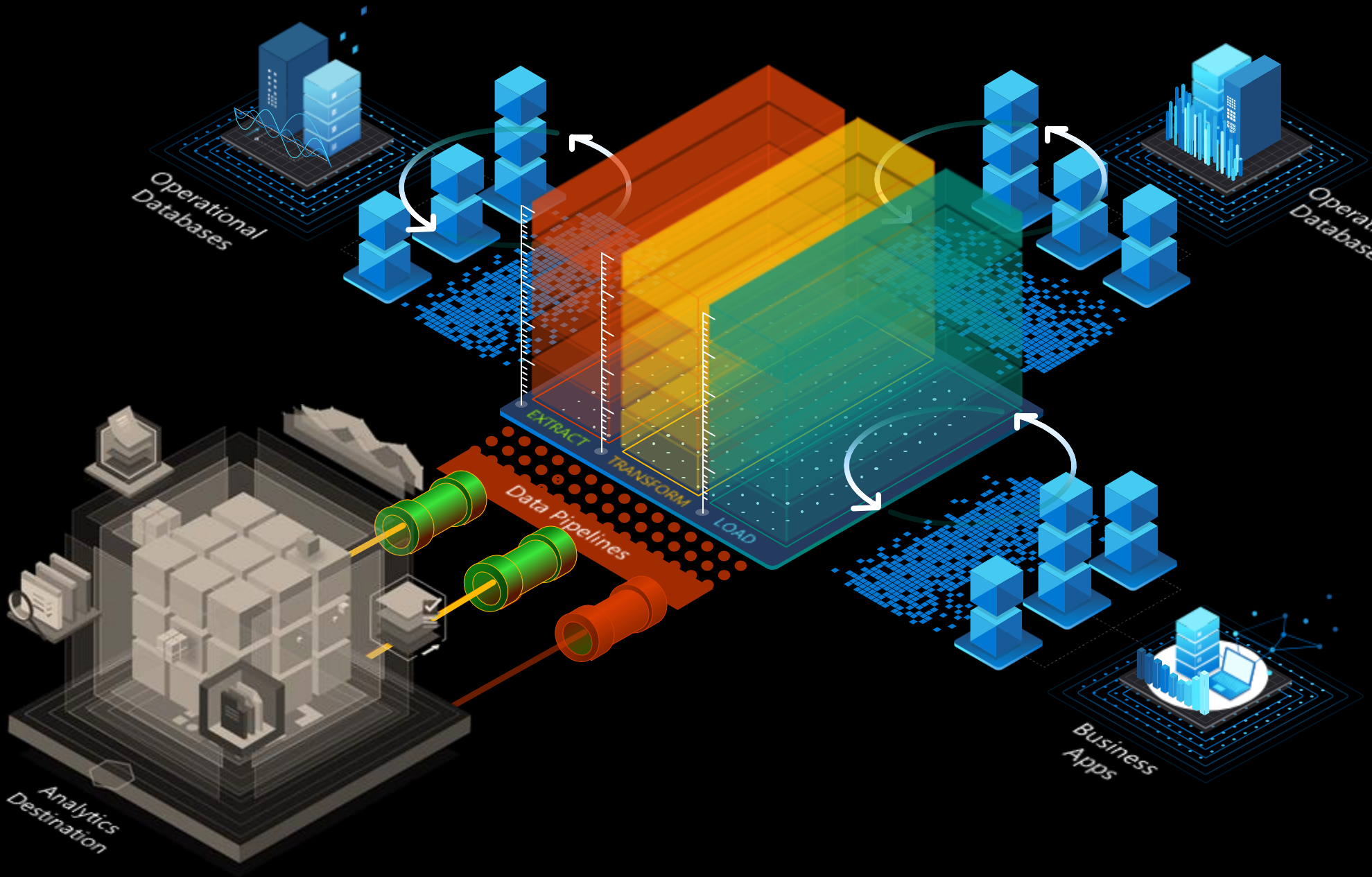
Azure SQL Database
Public Preview Q2 2022



Preparing data to be analytics-ready requires various processes



These processes result in additional issues that stall data insights



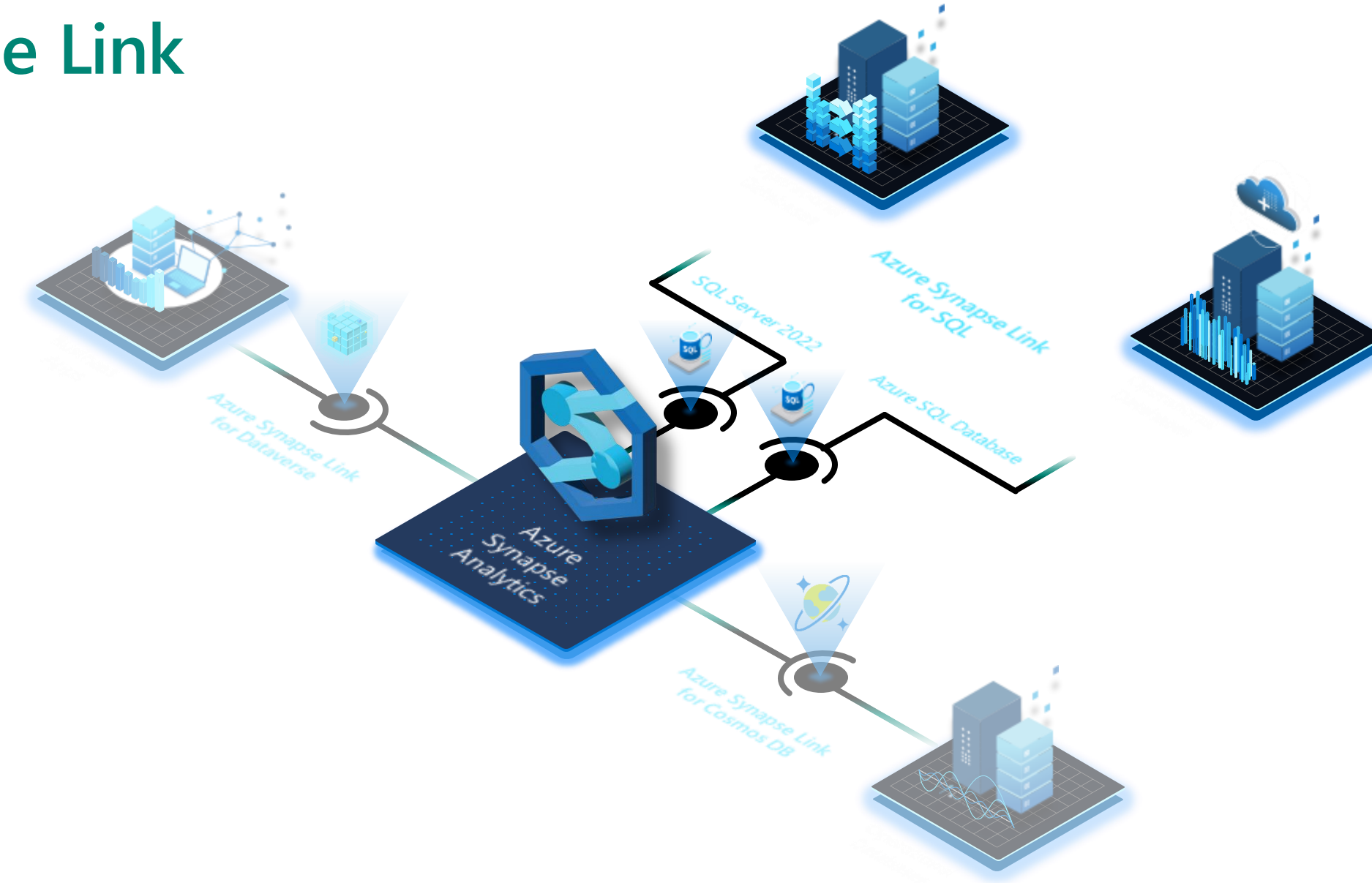
Azure Synapse Link

- No managed ETL
- Built for convergence
- Near real-time analytics

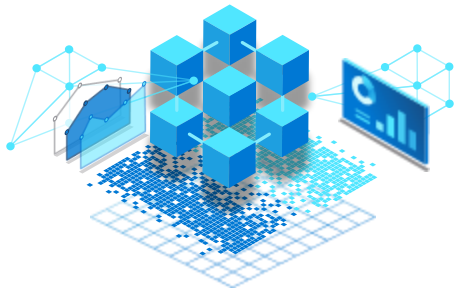
Azure Synapse Link

Go from after-the-fact analysis to near real-time insights by eliminating barriers between data and analytics. Automatically move data from both operational databases and business applications without time-consuming ETL. Deliver critical insights by easily connecting separate systems, bringing the power of analytics to every data-connected team.

Azure Synapse Link



Azure Synapse Link for SQL



Simplified experience

With a streamlined setup, transfer data from operational systems seamlessly

Centralized data

Consolidate data from multiple sources into a single analytics solution

Near real-time insights

Maximize impact with automated systems, capturing and updating data in near real-time

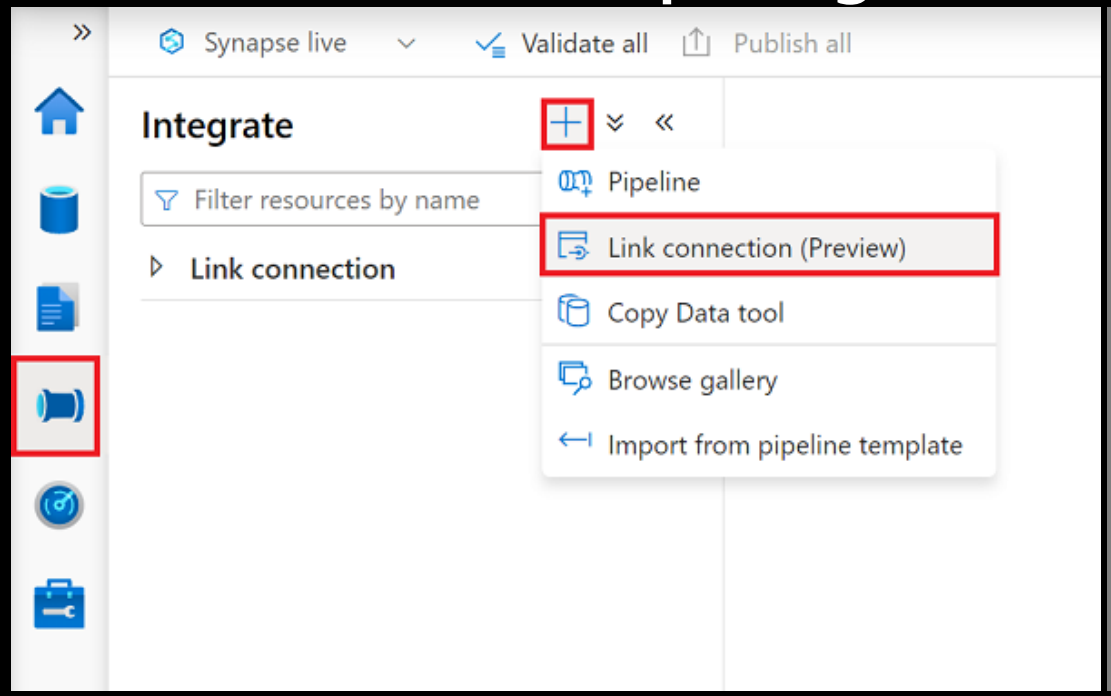
Minimized overhead

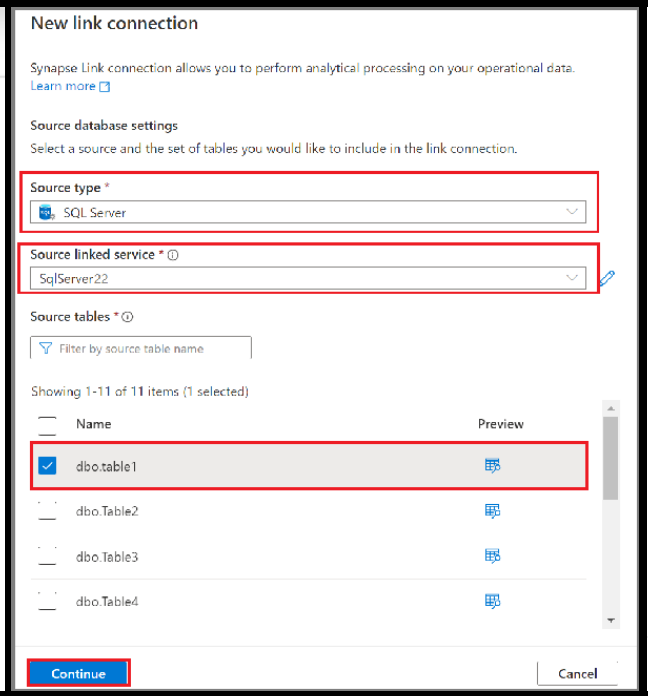
Record data changes without stretching source systems

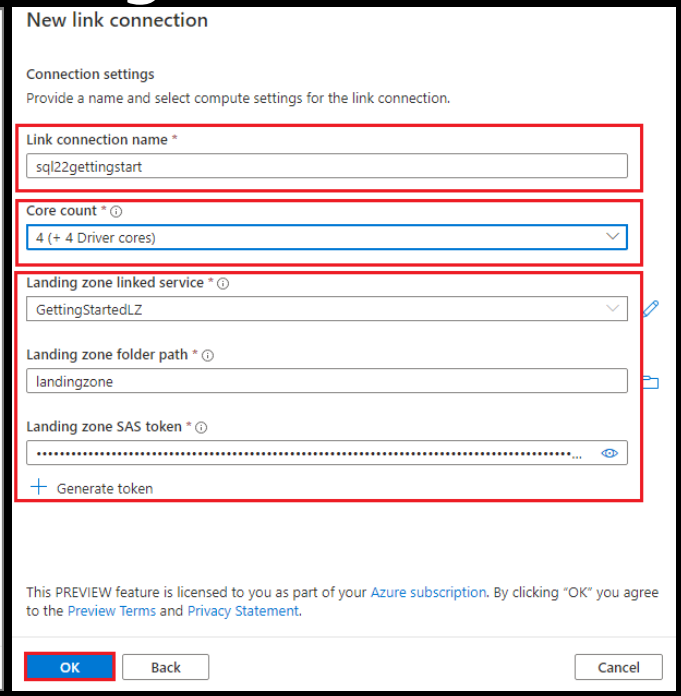
Azure Synapse Link for SQL

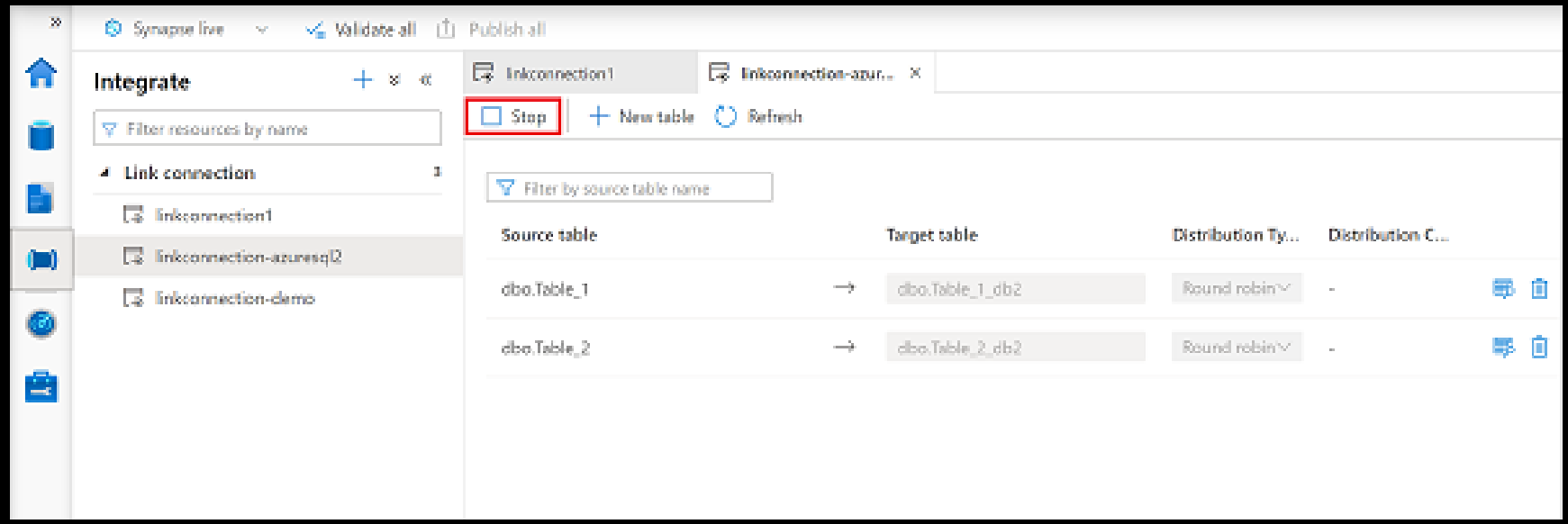
Automate data transfer and eliminate barriers with Azure Synapse Link for SQL

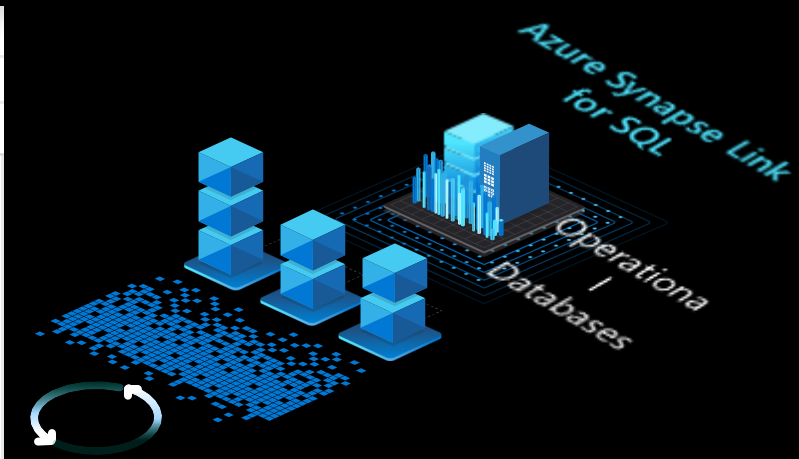
Preparing data to be near real-time insights











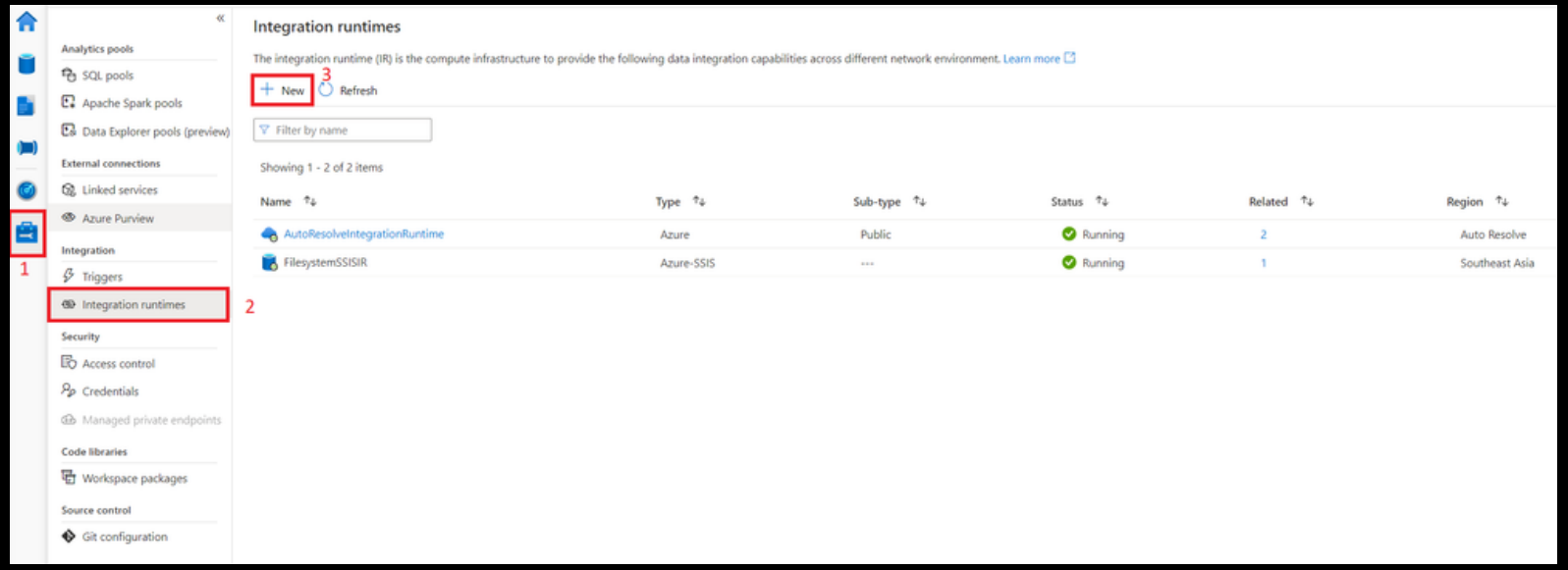
Public Preview

Q2 2022

SSIS Integration Runtime for Azure Synapse

Running packages deployed into SSIS catalog (SSISDB) hosted by Azure SQL Database server/Managed Instance (Project Deployment Model)

Running packages deployed into Azure Files (Package Deployment Model)



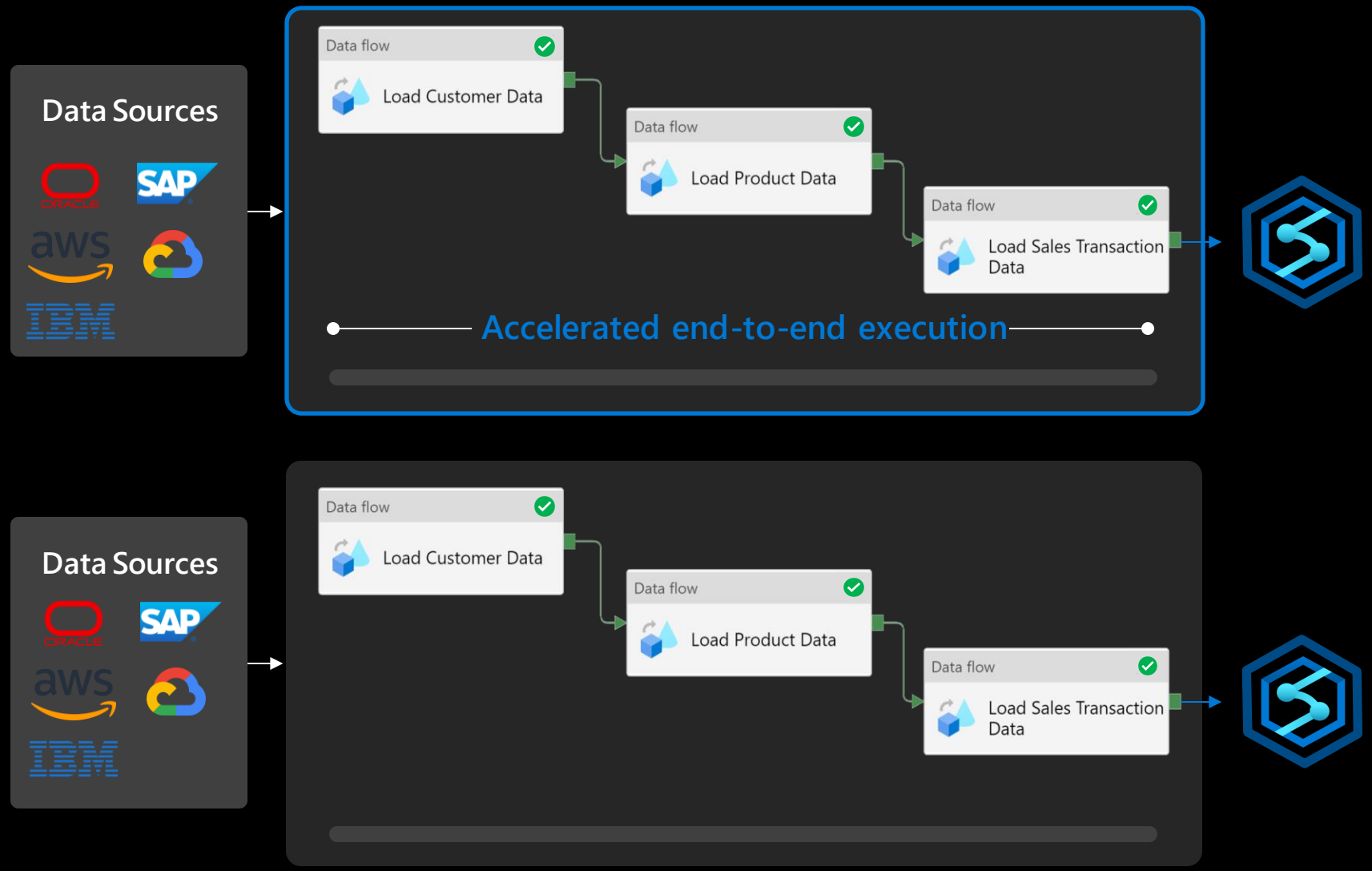
Generally Available

October 2021

Accelerated Data Pipelines

Cluster time-to-live enables near instant start of data flow pipelines for faster data integration

Data is available to the business faster to enable more timely decision making



- Data Integration
- Data Engineering
- Data Warehousing
- Data Science
- Observational Analytics
- Business Intelligence
- Governance

Data Engineering

Data Engineering

Scalable Spark engine

Industry standard languages

Delta Lake Enabled

Azure DevOps integration

Public Preview

Q2 2022

Spark 3.2

Enables developers can leverage the latest innovations in the Spark ecosystem

Pandas (Koalas) integration

A highly popular and flexible library with broad industry adoption

Adaptive Query Execution (AQE) enabled by default

Significant improvements in query performance out-of-the-box

Small Query execution improvements

Small queries run faster due to reduced initialization overhead

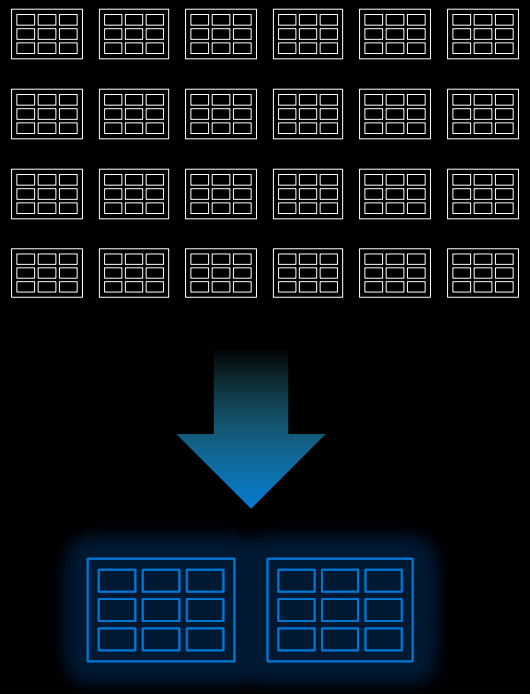
Public Preview

Q2 2022

Delta Lake Performance Enhancements

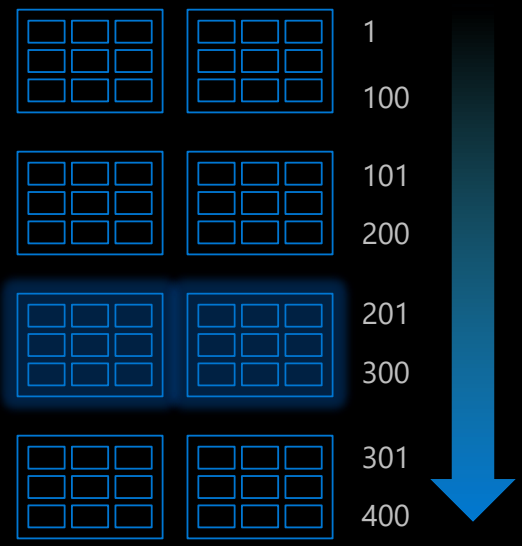
Improved performance and reduced cost with support for OPTIMIZE and Z-ORDER

OPTIMIZE



Improve query performance by coalescing small files into larger ones

Z-ORDER



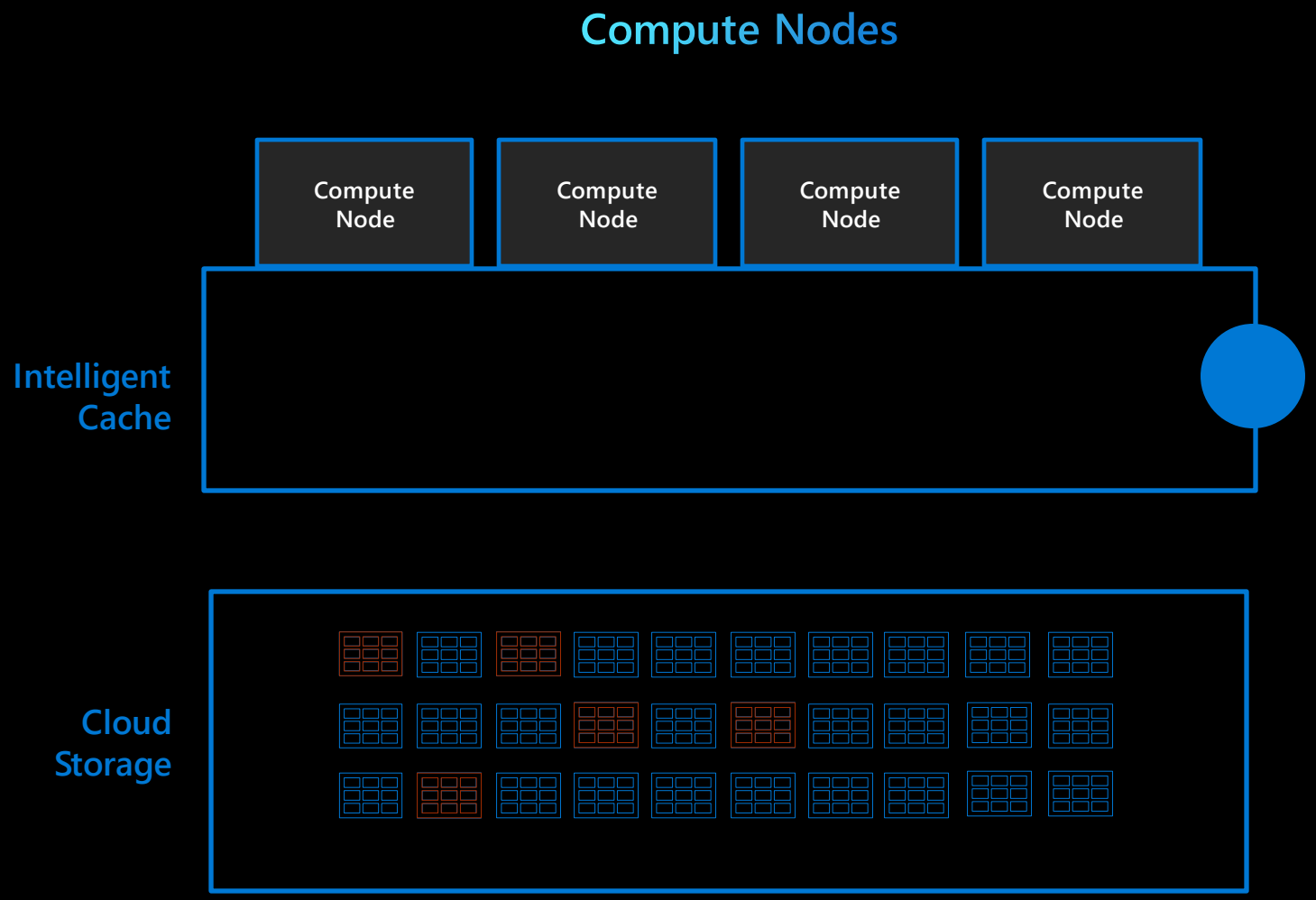
Improves filter query performance by ordering data for fast lookups on large datasets

Public Preview

Q1 2022

Spark Cashing Enhancements

Intelligent cache automatically detects changes in data to ensure data is fresh and results are accurate



- Data Integration
- Data Engineering
- Data Warehousing
- Data Science
- Observational Analytics
- Business Intelligence
- Governance

Data Warehousing

Cornerstone of enterprise analytics for decades

Industry standard SQL language

Structured and semi-structured data

Broad ecosystem of applications

Fine-grained data security

Data models tailored to business consumption

Generally Available

Dedicated & Serverless SQL

Elastic clusters with in-memory caching provide enterprise class performance combined with cloud economics

Serverless

Dedicated

Generally Available

Most Complete Workload Management

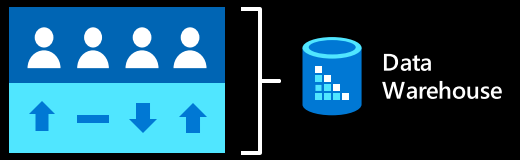
Scale-in to maximize output with the predictable cost

Scale-out to leverage cloud scale resources for spikes in demand

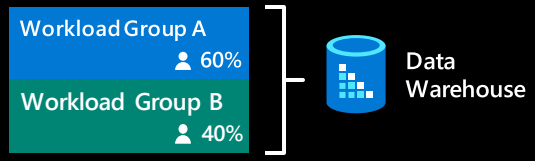
Scale-In

- Predictable cost
- Prioritize higher value work
- Prevents global contention

Workload Importance



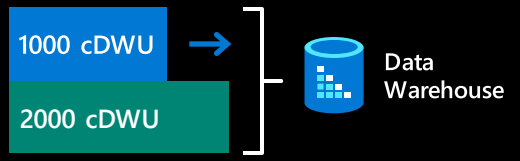
Workload Isolation



Scale-Out

- Add compute for variable workloads
- Pause compute when idle

Elastic Cluster (Scale Up)



Generally Available

Complete Data Protection

Democratize data compliantly with fine-grained access controls and multi-level encryption

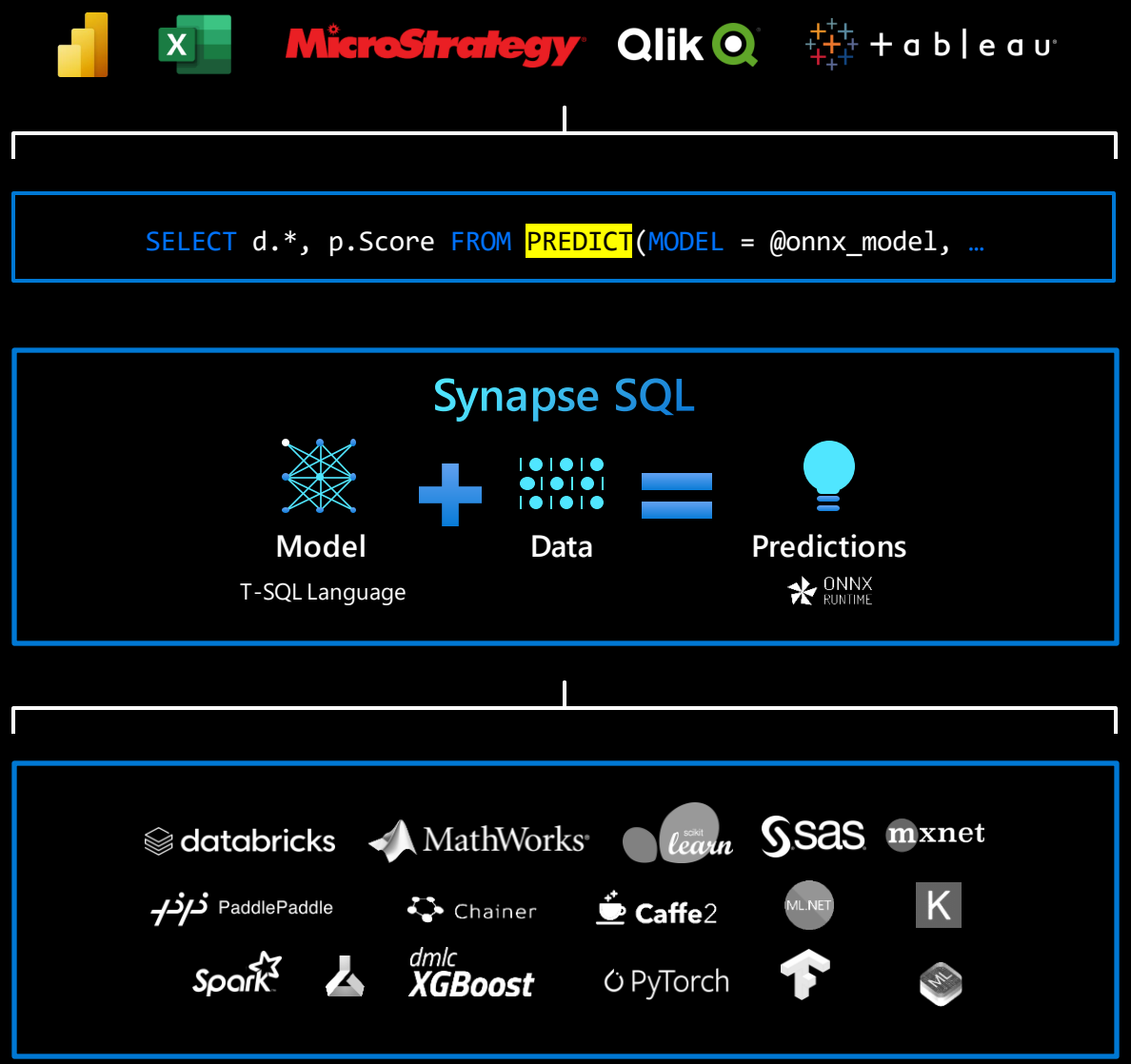
Category

Category	Feature	
Data Protection	Data in transit	✓
	Data encryption at rest	✓
	Data discovery and classification	✓
Access Control	Object level security (tables/views)	✓
	Row level security	✓
	Column level security	✓
	Dynamic data masking	✓
	Column level encryption	✓
Authentication	SQL login	✓
	Azure active directory	✓
	Multi-factor authentication	✓
Network Security	Managed virtual network	✓
	Custom virtual network	✓
	Firewall	✓
	Azure ExpressRoute	✓
	Azure Private Link	✓
Threat protection	Threat detection	✓
	Auditing	✓
	Vulnerability assessment	✓
Isolation	Dedicated metadata store	✓
	Hosted in customer tenant	✓

Generally Available

Democratize ML predictions with SQL

In-engine ML scoring provides interactive query response times without any data leaving the system and no additional scoring cost



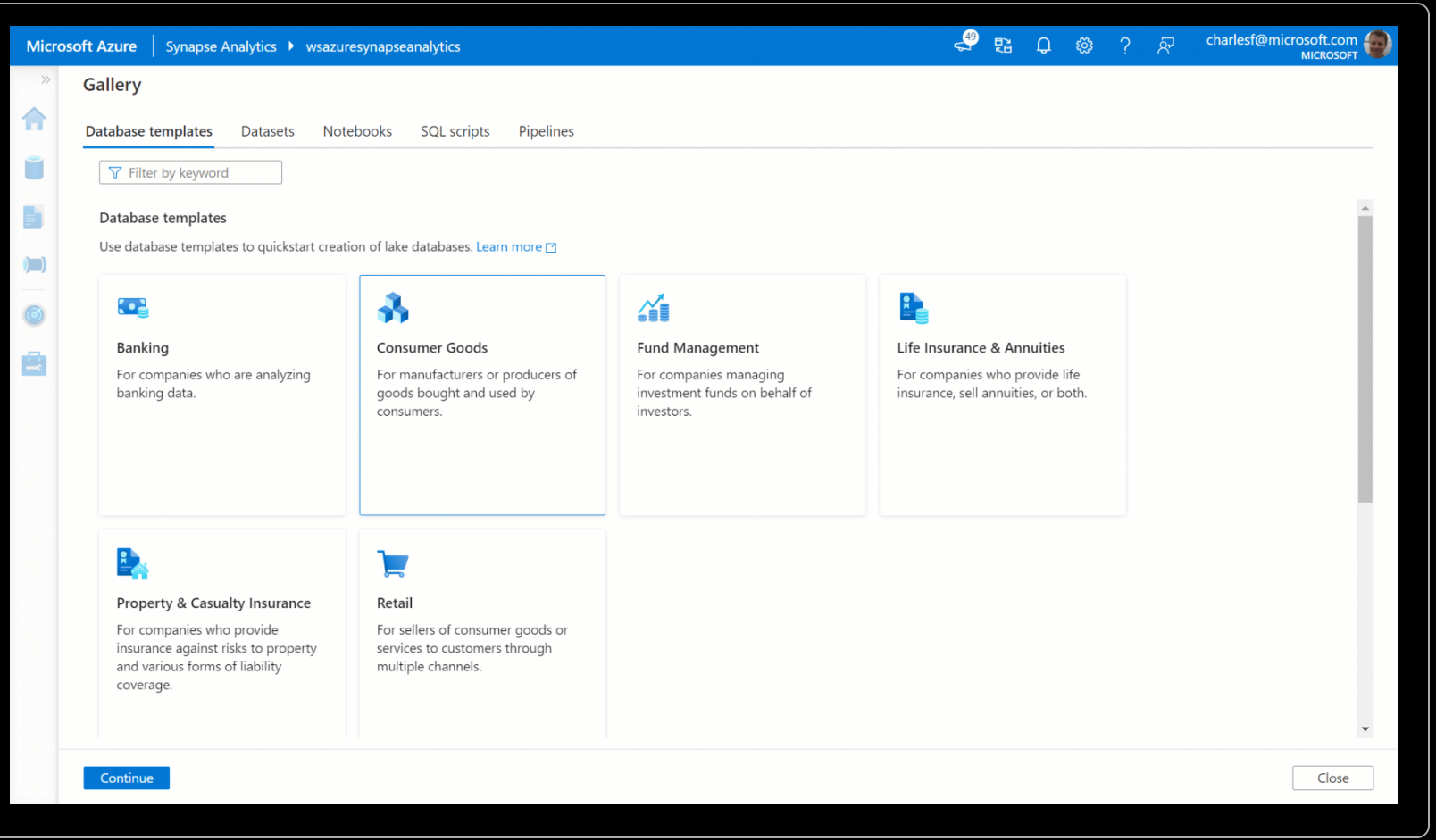
Generally Available

Q2 2022

Database template

Built-in database templates

Low code database designer



Public Preview

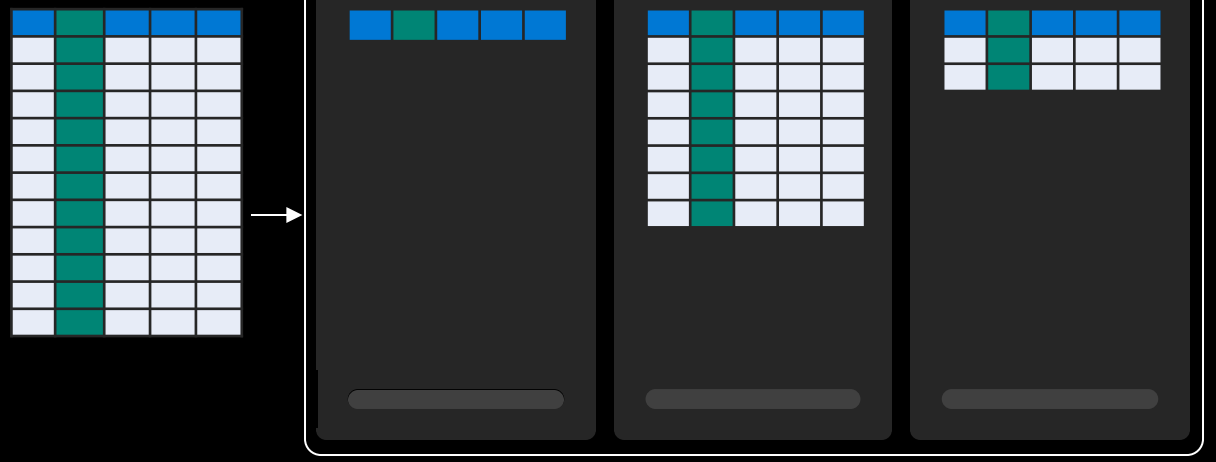
Q2 2022

Multi-column distribution table sharing

Improved query performance and easier migrations

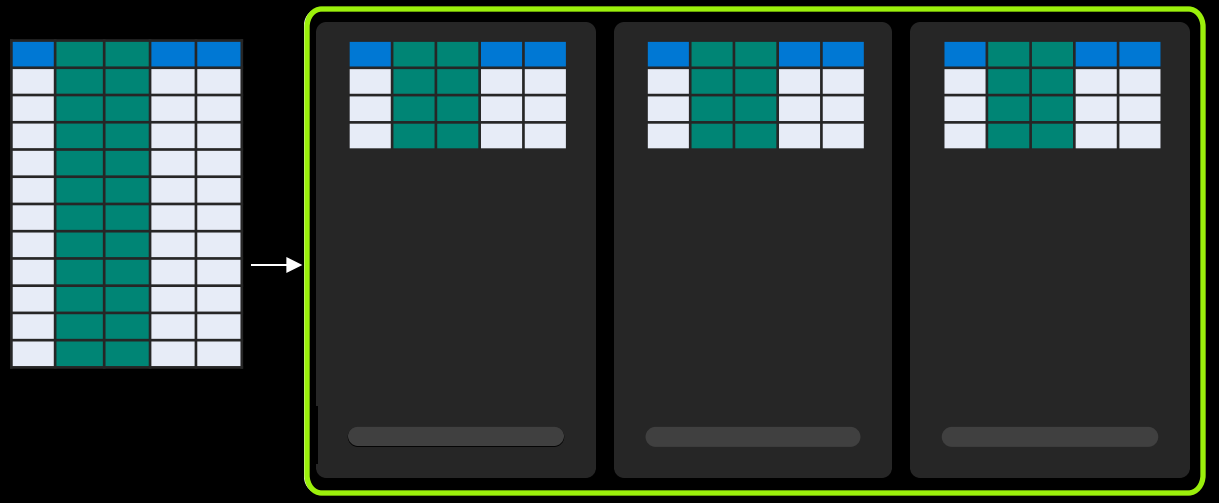
Uneven Distribution (Skew)

```
CREATE TABLE SalesTransactions (
WITH DISTRIBUTION =
(HASH(ProductKey))
```



Balanced Distribution

```
CREATE TABLE SalesTransactions (
WITH DISTRIBUTION =
(HASH(ProductKey, RegionKey))
```



Balanced Distribution: Queries execute faster

Public Preview

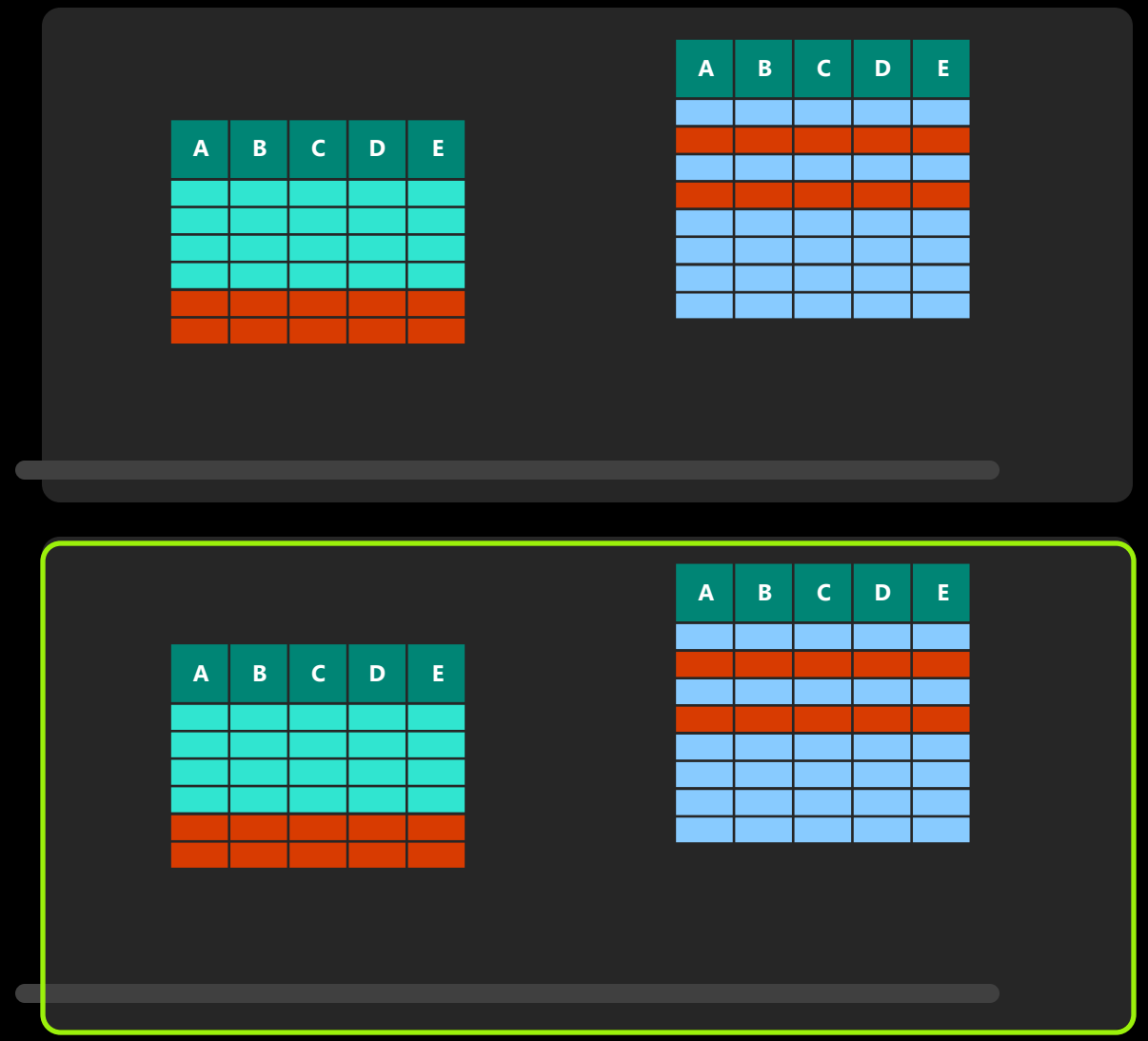
Q2 2022

MERGE SQL Statement

Improved performance and easier migration by executing INSERT, UPDATE, and DELETE functionality in a single statement

```
INSERT ...
UPDATE ...
DELETE ...
```

MERGE ...

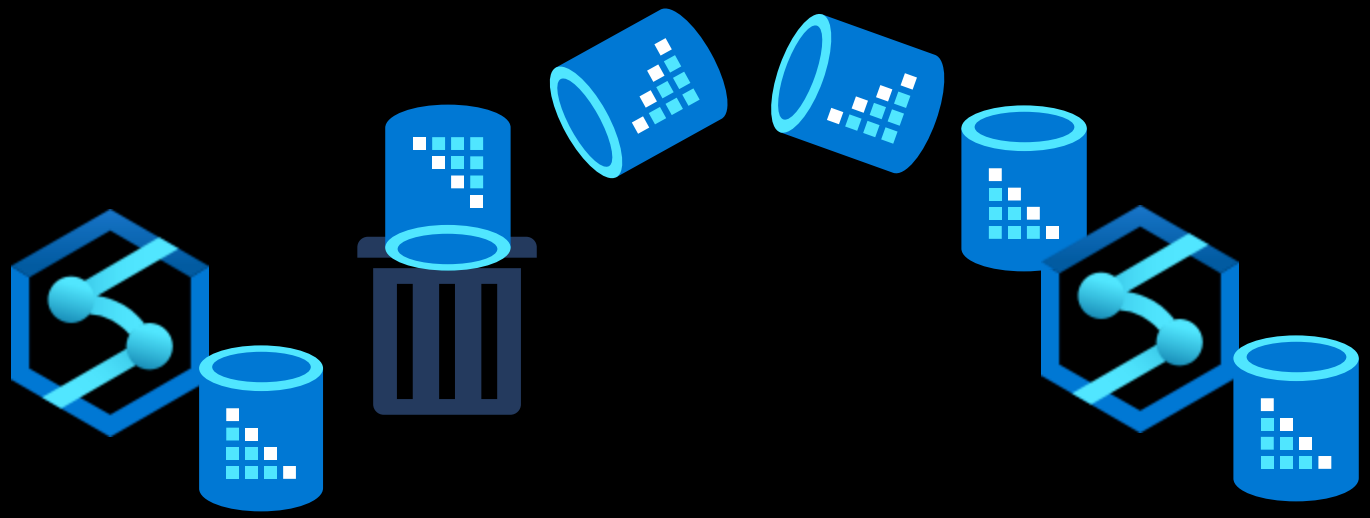


Public Preview

Q2 2022

Recover data warehouse from dropped server

No longer need support ticket to recover after an accidental server or workspace drop

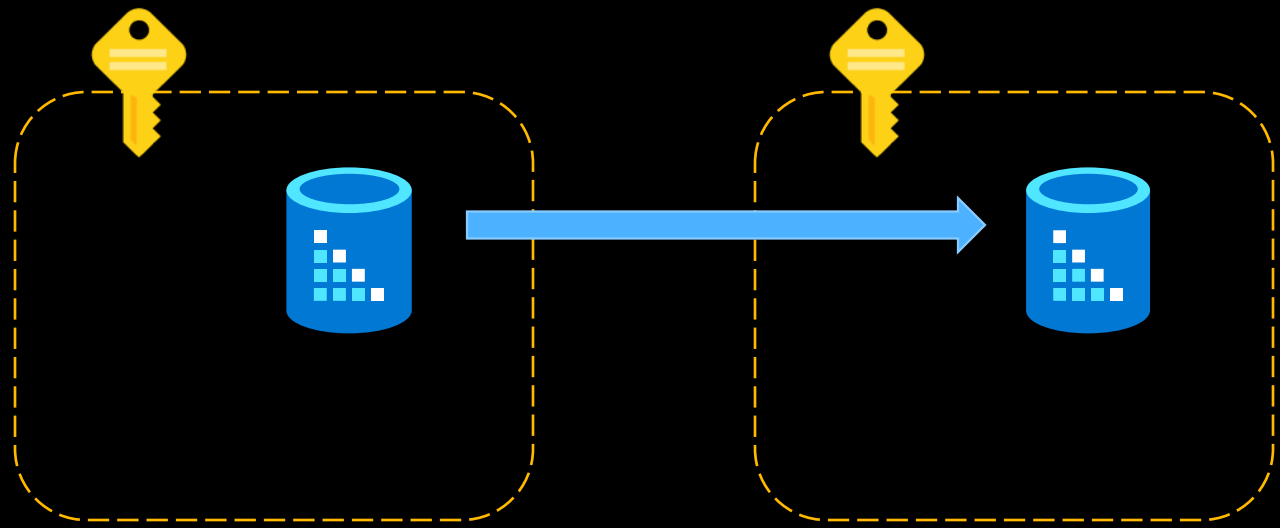


```
Restore-AzSynapseSqlPool -FromDroppedSqlPool
```

Public Preview

Q2 2022

Cross subscription restore



Restore data warehouse across subscription boundary in one command without workarounds

Unlock Dev/Test scenarios and simplify billing at the subscription level

PowerShell and API support available

Public Preview

Q2 2022

Distribution Advisor

Recommendation system that analyses the chosen queries or past run query data to provide suggestions on data distribution which can improve performance

ProductSales

AccountID	SalesAmt	...
47	$1,234.36	...
36	$2,345.47	...
14	$3,456.58	...
25	$4,567.69	...
48	$5,678.70	...
37	$6,789.81	...
...	...	...

```
CREATE TABLE ProductSales
WITH (DISTRIBUTION=HASH(AccountID))
AS...
```

- Data Integration
- Data Engineering
- Data Warehousing
- Data Science
- Observational Analytics
- Business Intelligence
- Governance

Data Science

Industry standard languages such as PySpark

Rich ecosystem of ML tools on Spark

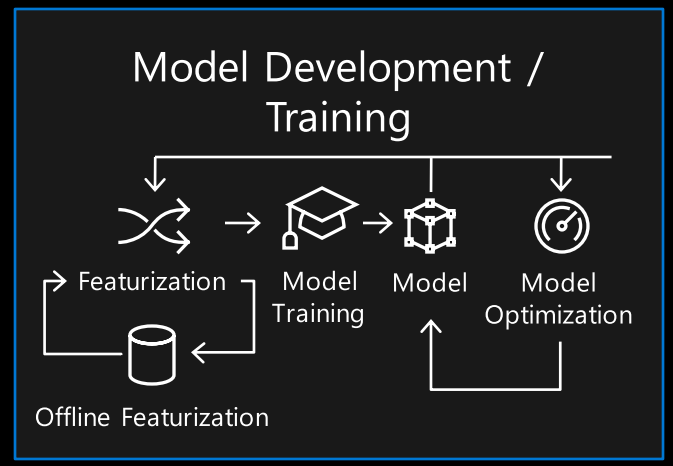
Code-first and Code-free Auto ML

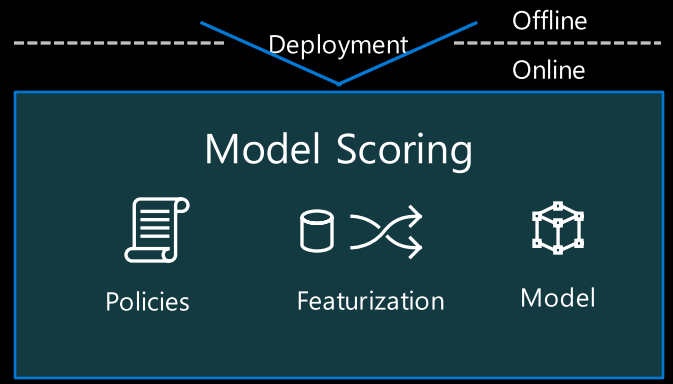
Hardware accelerated execution with GPUs

Why Synapse
Enterprise Grade-ML

Productizing AI

- Train in the cloud within the Hub
- Scoring with operational systems
- Governance everywhere (models, lineage)
- Ethical AI
- Control over deployment
- Deployment across Apps, BI, Processes
- Exchange of models (ONNX)
- Enabling Reinforcement learning

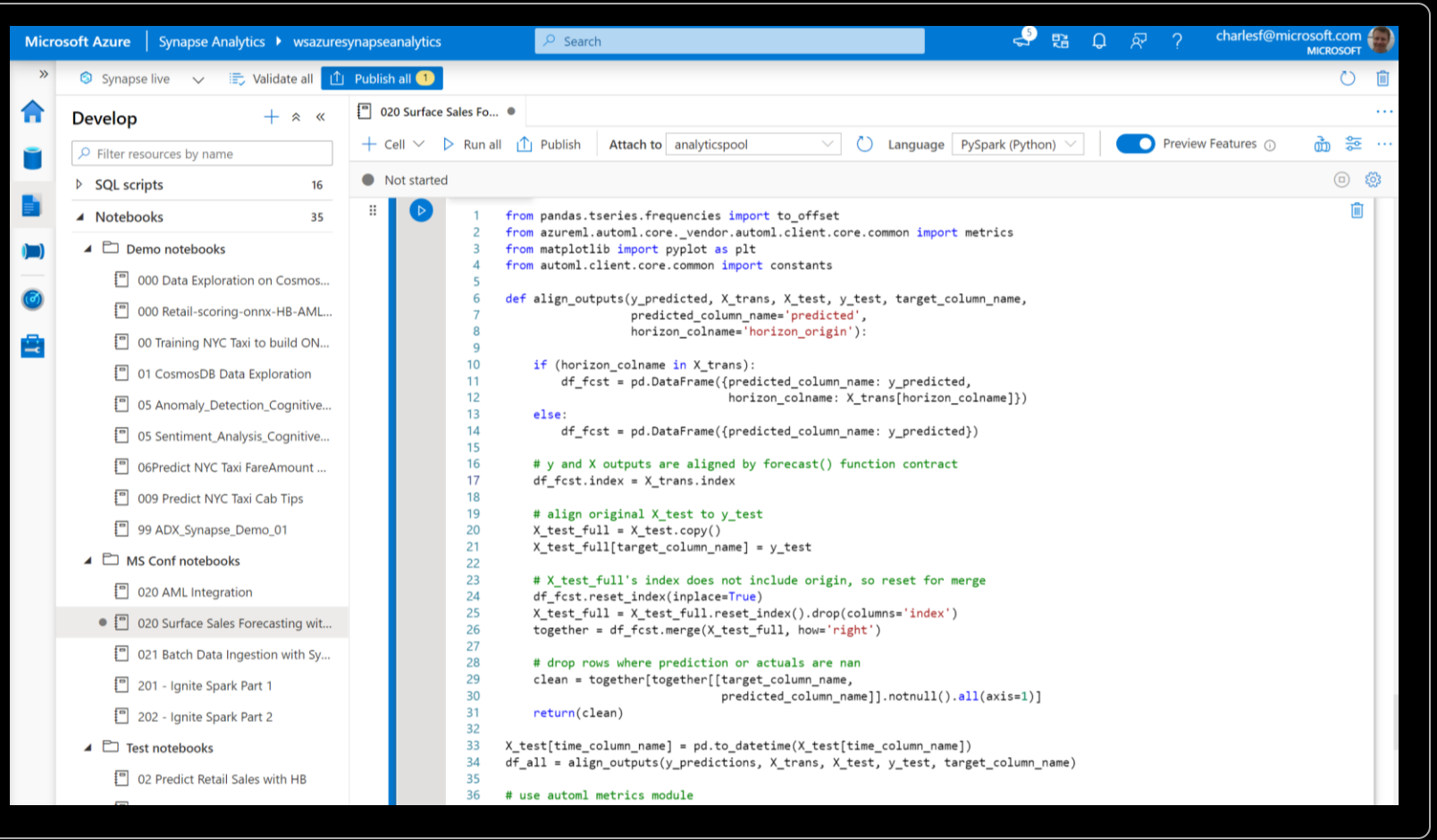




Generally Available

Notebook Development Experience

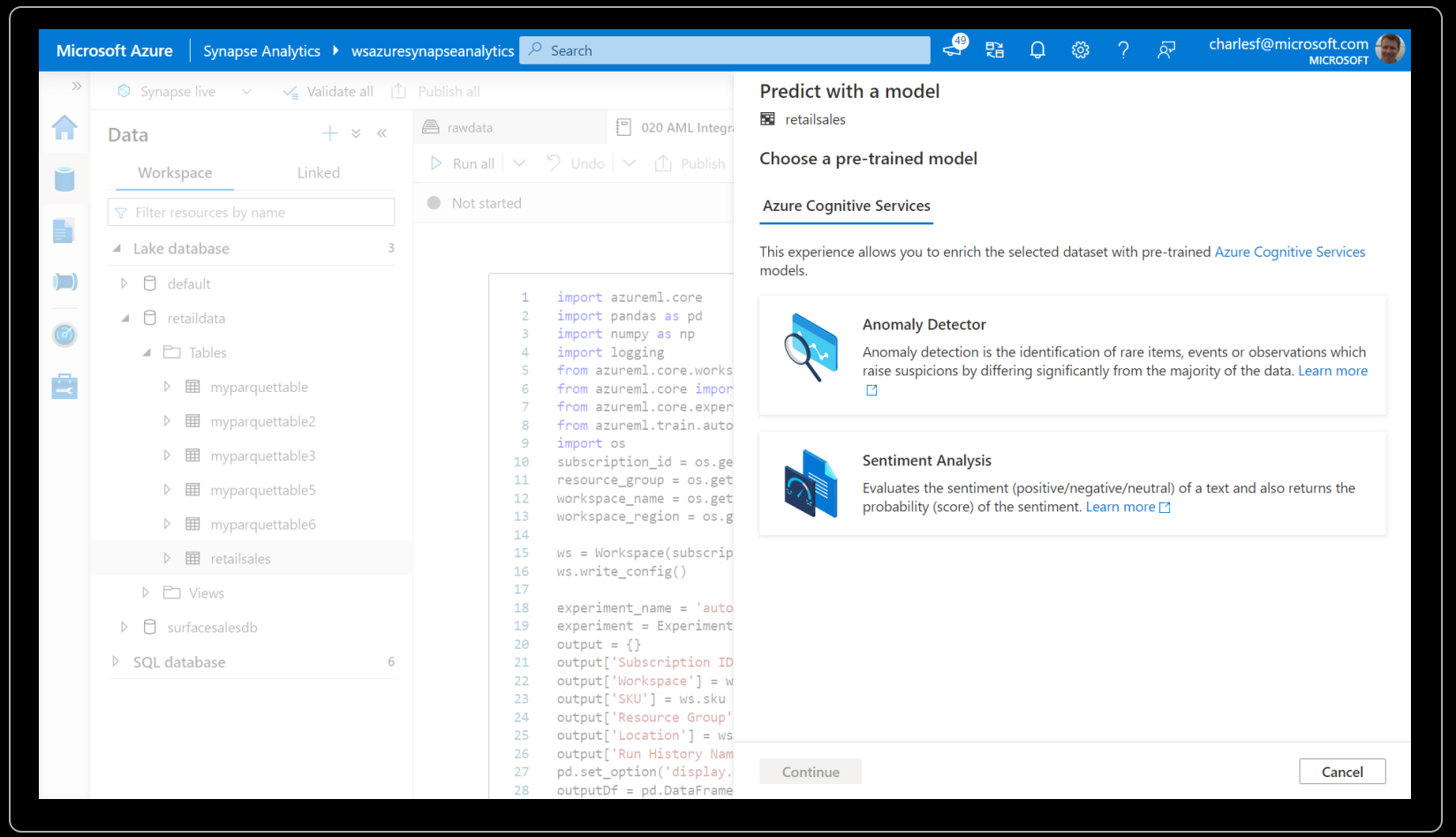
Empower data scientists will a familiar Notebook based development interface



Generally Available

Built-in Cognitive Services

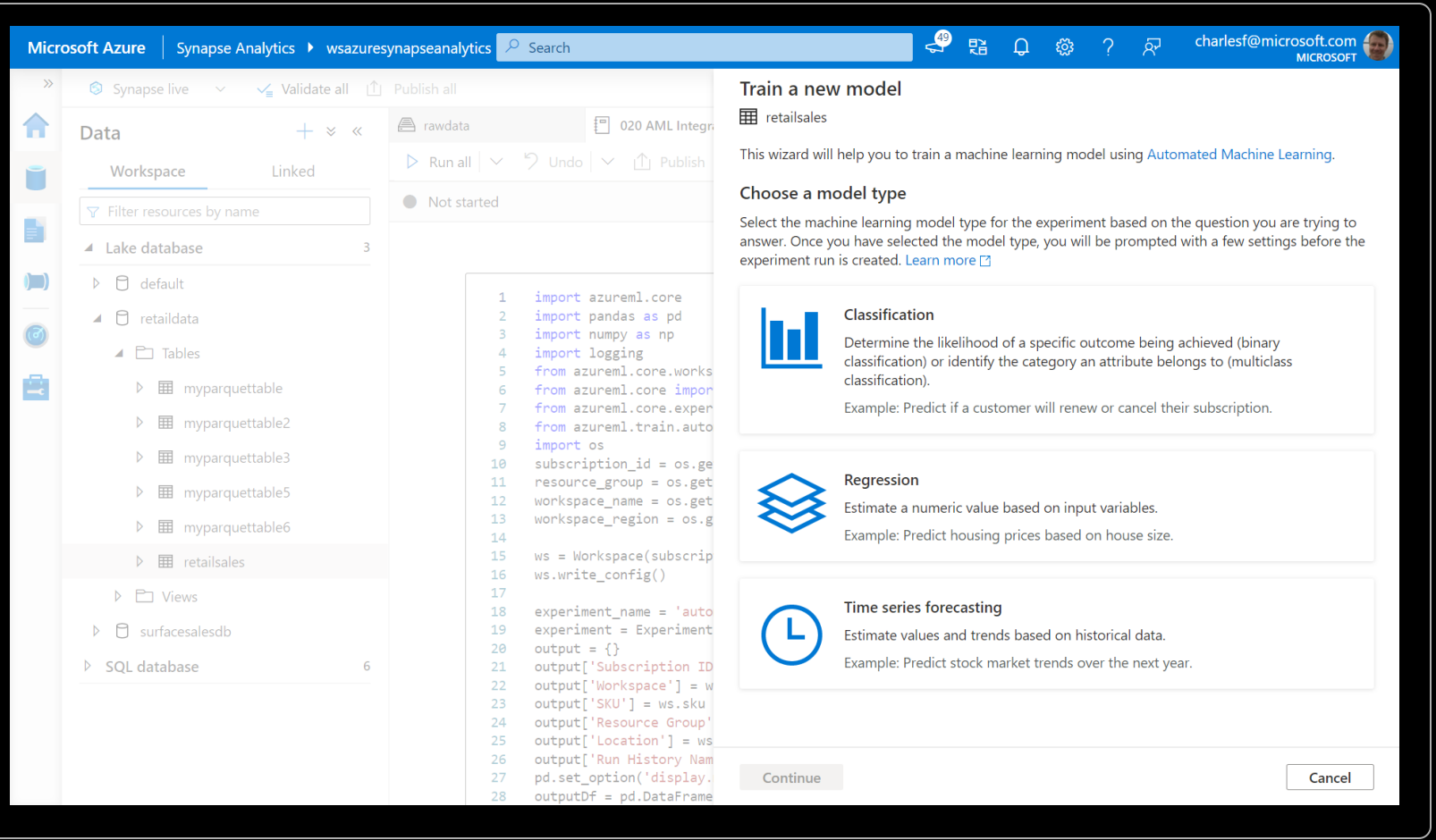
Enables simple integration of pre-built machine learning models.



Public Preview

Automated Machine Learning

No-code training for ML models empowers everyone with data science

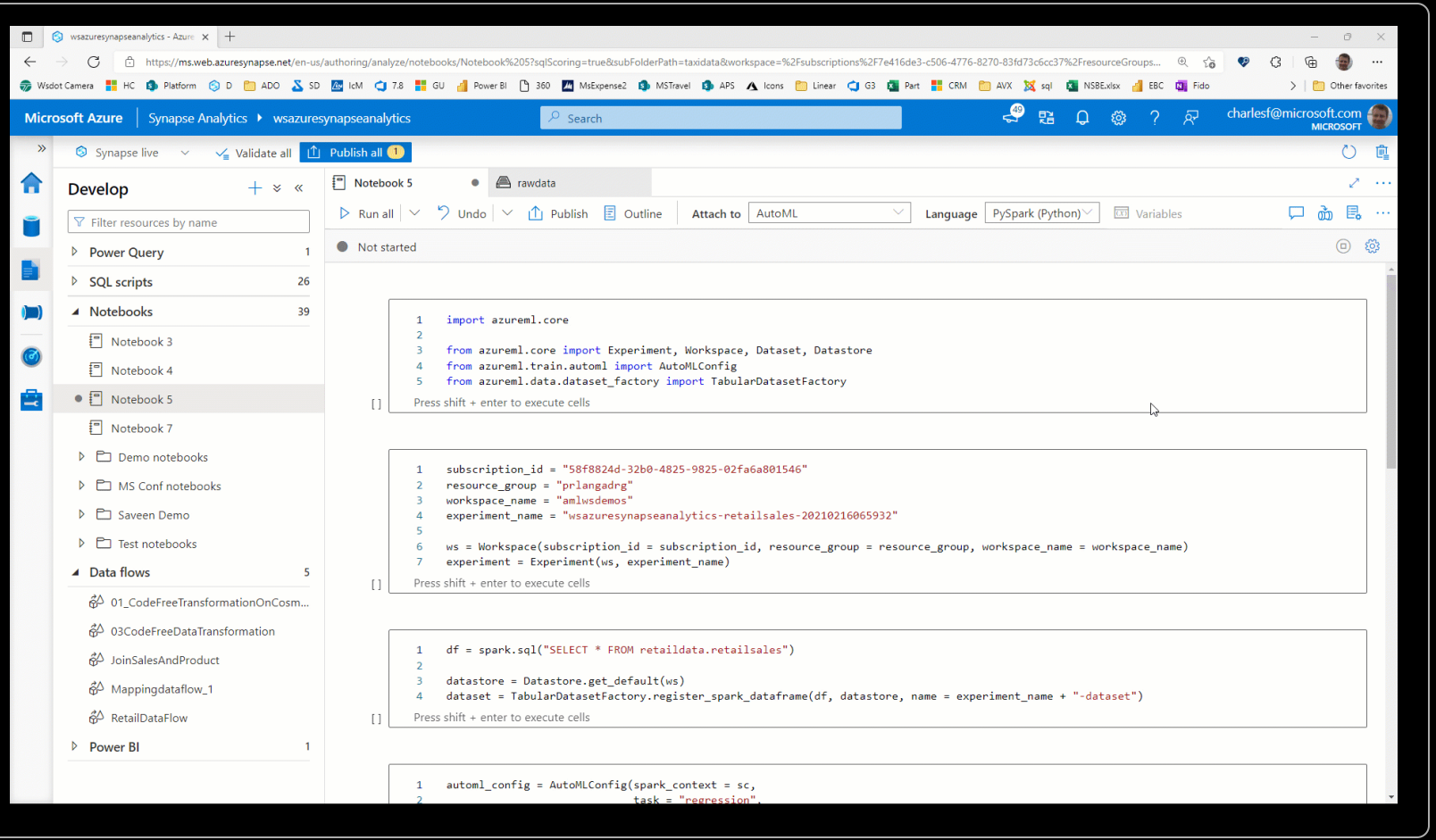


Generally Available

Industry Standard Open Ecosystem

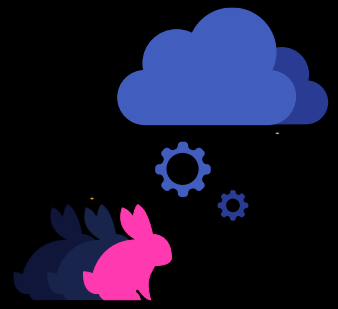
Open file formats enable easy integration with other data services

Industry standard languages make it easy for developers to get started



Generally Available

Synapse ML Library GA



Open-Source Spark Library

Distributed ML algorithms for Apache Spark, e.g.:

- Linear models & reinforcement learning: Vowpal Wabbit
- Gradient boosted trees: LightGBM
- Novelty detection: Isolation Forests

Cognitive Services Integration for Spark

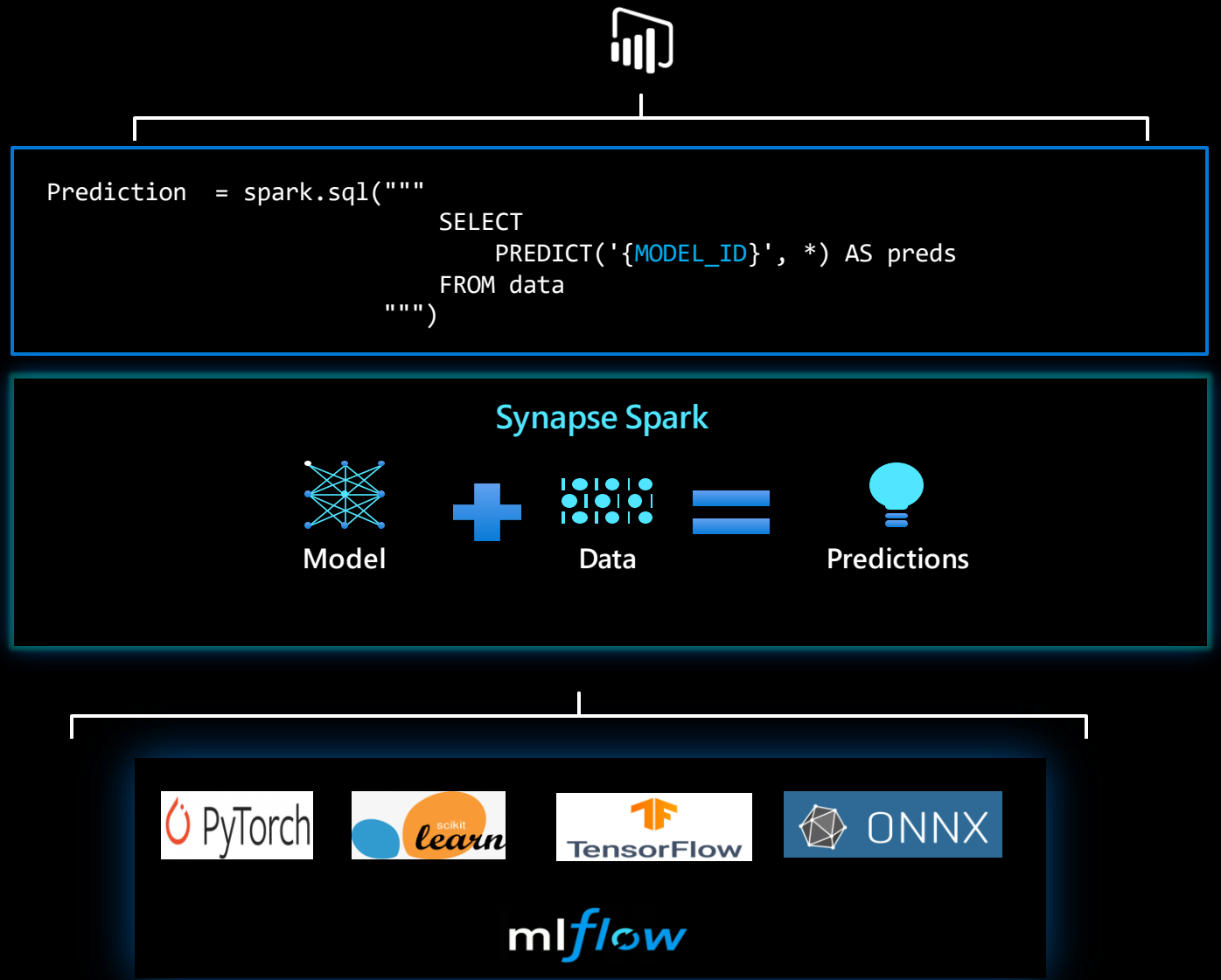
Language Binding Generators: Python, R, Java (, .NET)

https://aka.ms/spark

Public Preview

Simplify model scoring at scale on Spark

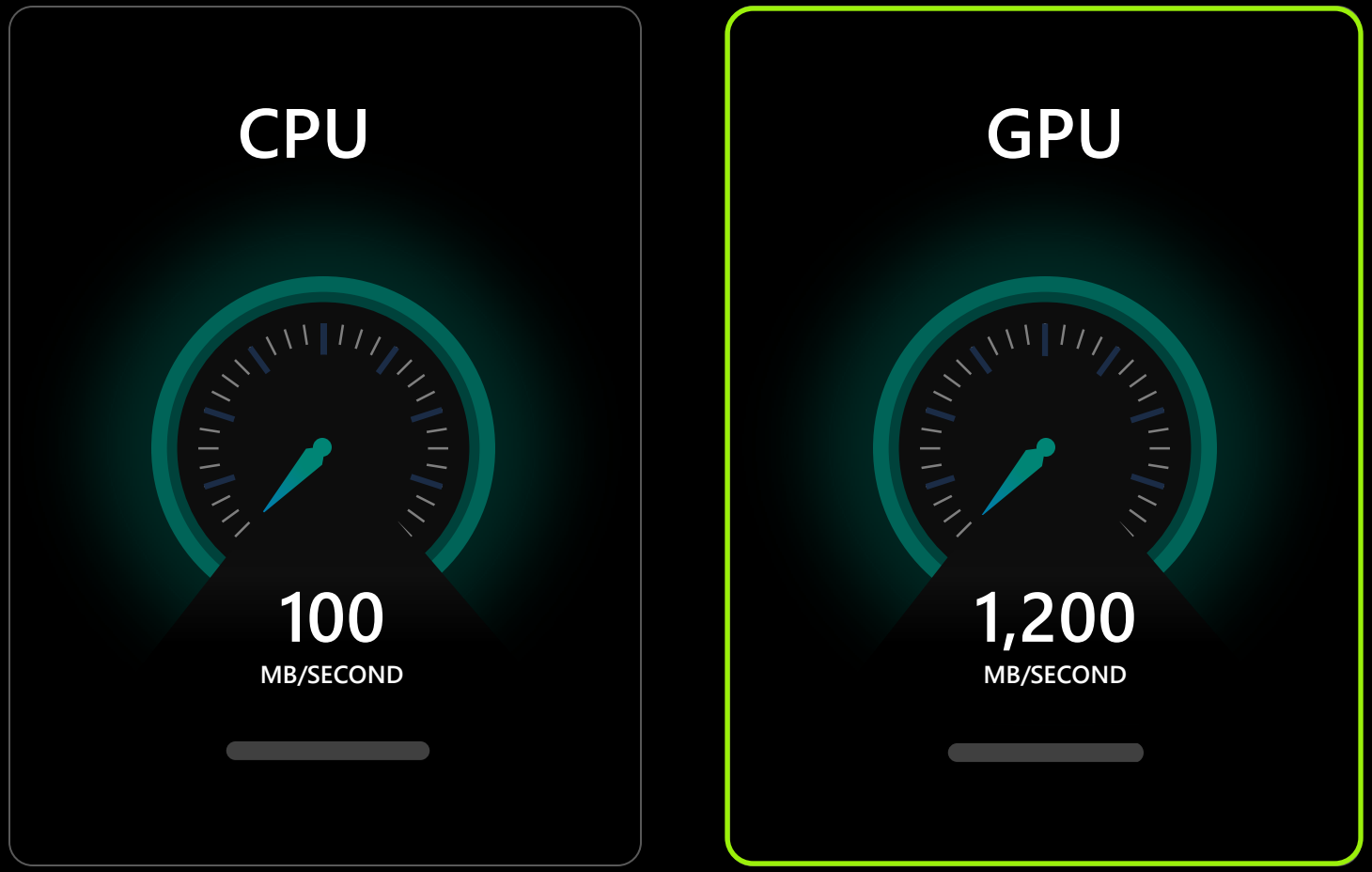
Drastically simplifies handover of models from producer to consumer for “in-engine” scoring. Use machine Learning models from Azure ML in Synapse without moving any data



Public Preview

GPU Accelerated Workloads

Accelerates data transformation and reduces ML model training time by dramatically increasing throughput vs. traditional CPU

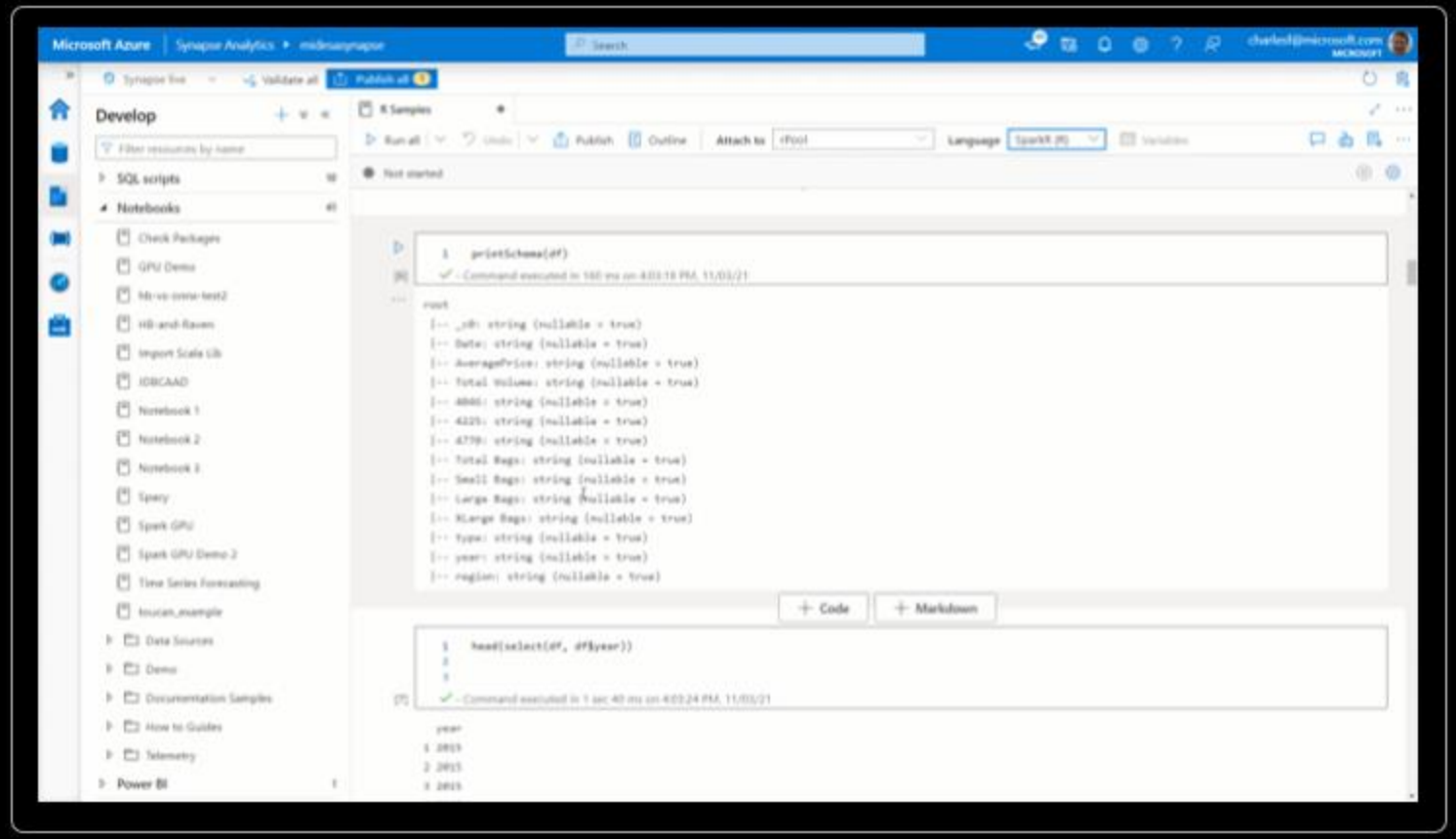


Private Preview

Q2 2022

R Language Support

Enables data scientists to apply the industry standard R language to developing ML models



- Data Integration
- Data Engineering
- Data Warehousing
- Data Science
- Observational Analytics
- Business Intelligence
- Governance

Observational Analytics

What is it?

Semi-structured: text, json, time series

Machine generated or machine recorded human interactions

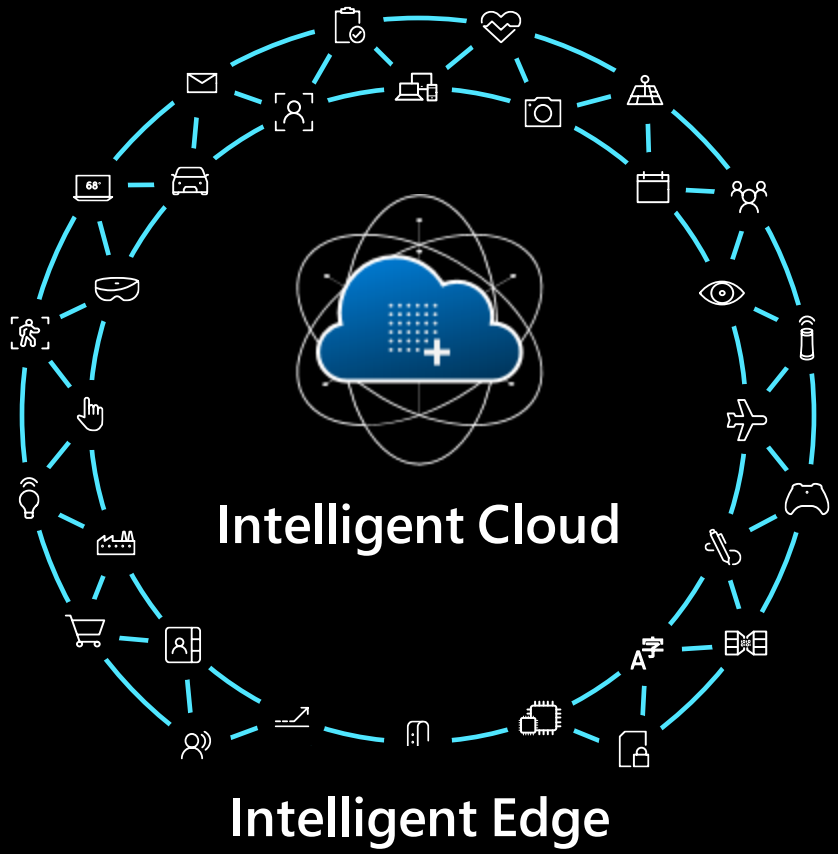
Mass volume

High velocity

Few large fact tables/streams

Observational Data

50 BN
connected devices
by 2030



175 ZB
total amount of
data by 2025

Looking for unpredictable phenomena

Constantly changing schema

Near real time visibility required

Analytics systems costs are often prohibitive

Frequently changing business questions

Why is it challenging to analyze?

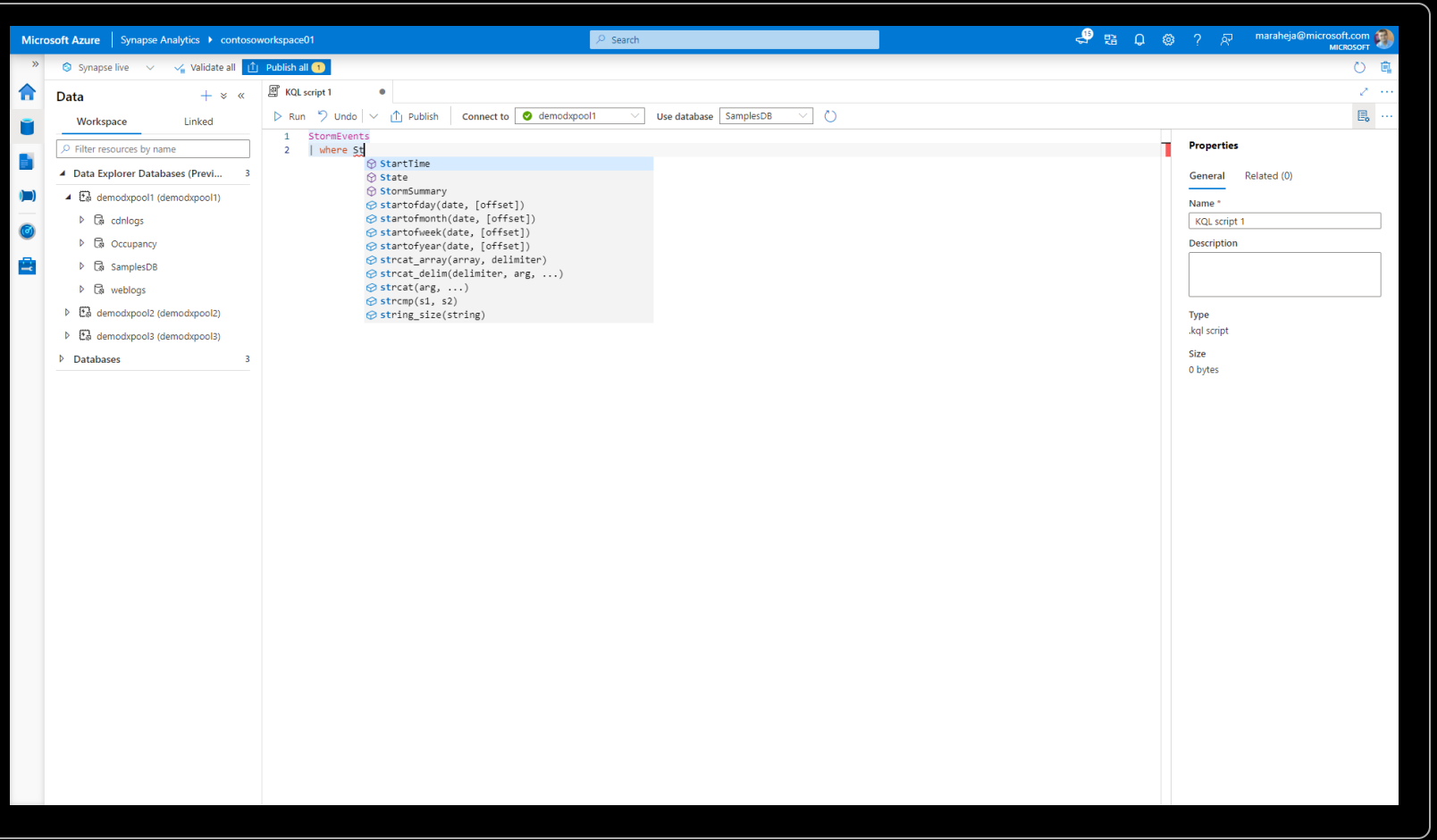
Near real-time analytics on Observational data at petabyte scale

Public Preview

November 2021

Synapse Data Explorer Engine

Industry leading free-text and semi-structured data indexing for sub second observational analytics

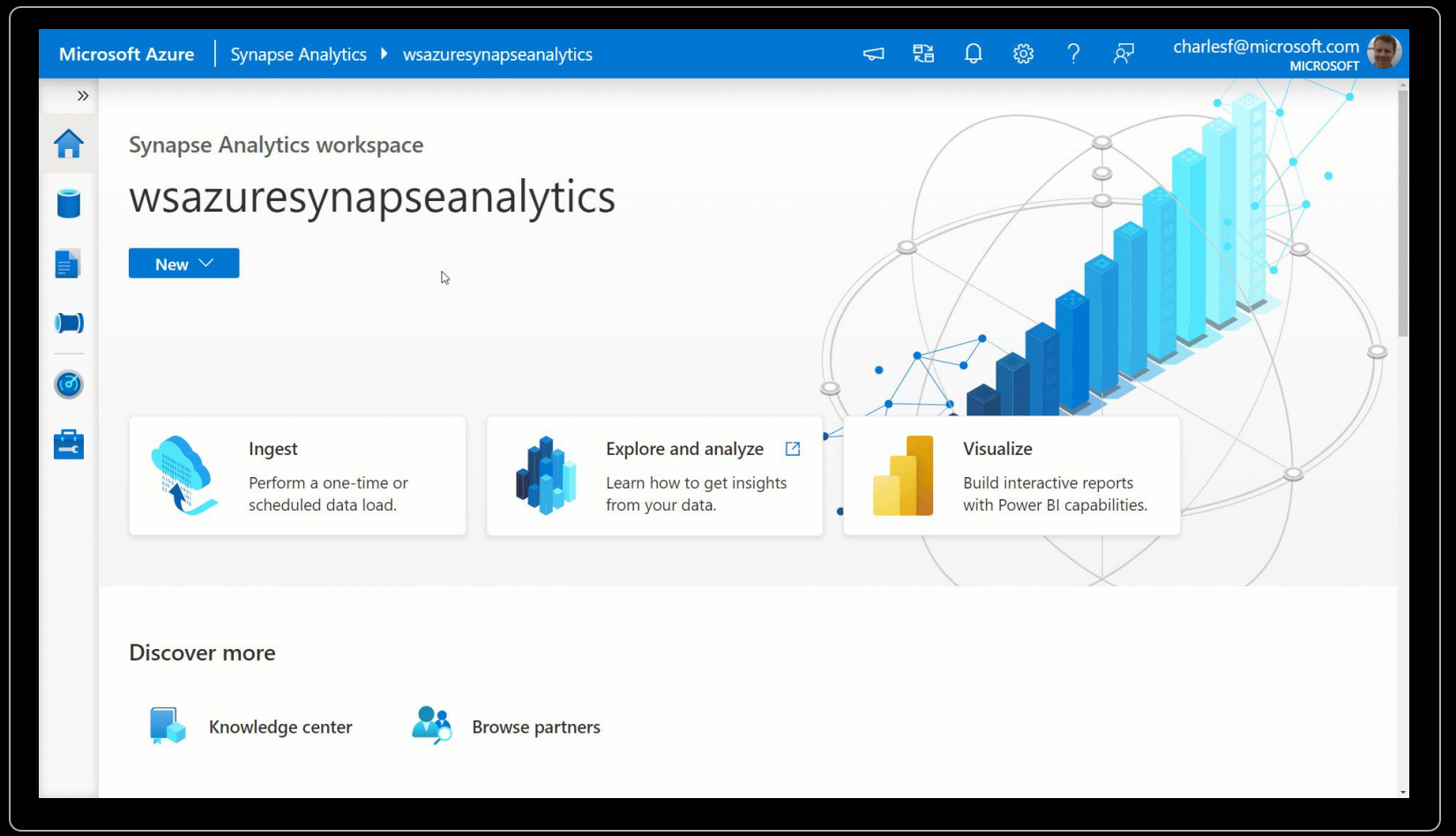


Public Preview

Q4 2021

Auto-Pause and Private Link for Data Explorer Clusters

Reduce costs by enabling clusters to automatically pause based on pre-defined timeout



Public Preview

Q4 2021

100,000 databases in a cluster

Enable developers to build large scale multi-tenant solutions with cluster compute reuse across workloads

Modélisation des objets logiques et "Lake Database"

Approche moderne de données - Lakehouse

- Nouvelle approche d'objets métiers
- Exploiter les bénéfices des solutions « Cloud Native » comme serverless, autoscale, on-demand….
- Notion forte de croisement entre le Datawarehouse, le Datalake et les modèles d'analyse
- **La Datalake est vu comme une variable d'entrée et non une source de donnée**

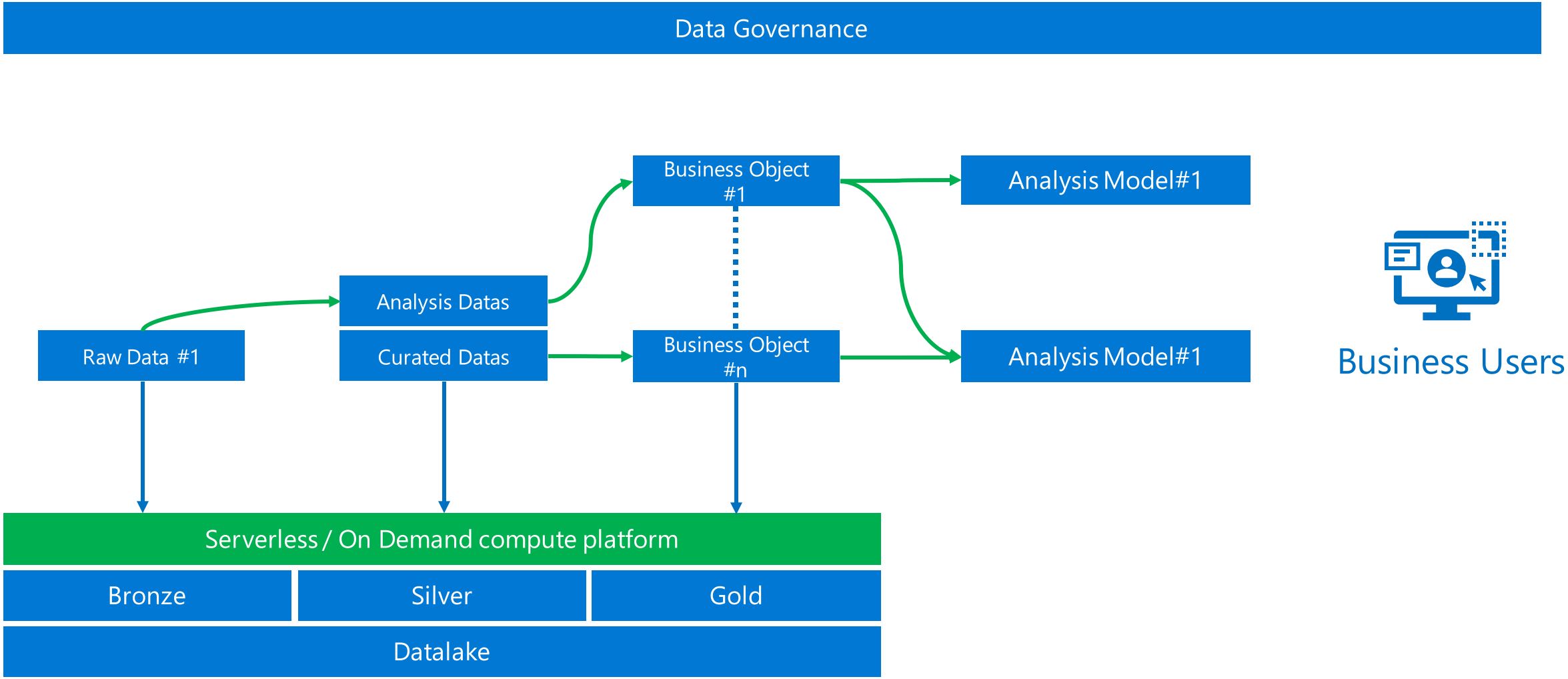
La finalité de l'approche

Exposer des objets métiers dynamiques
Exposer des objets métiers logiques
La matérialisation des objets n'est plus nécessaire
S'affranchir des notions de scalabilité et performance
S'affranchir de la consommation ou non de l'objet exposé

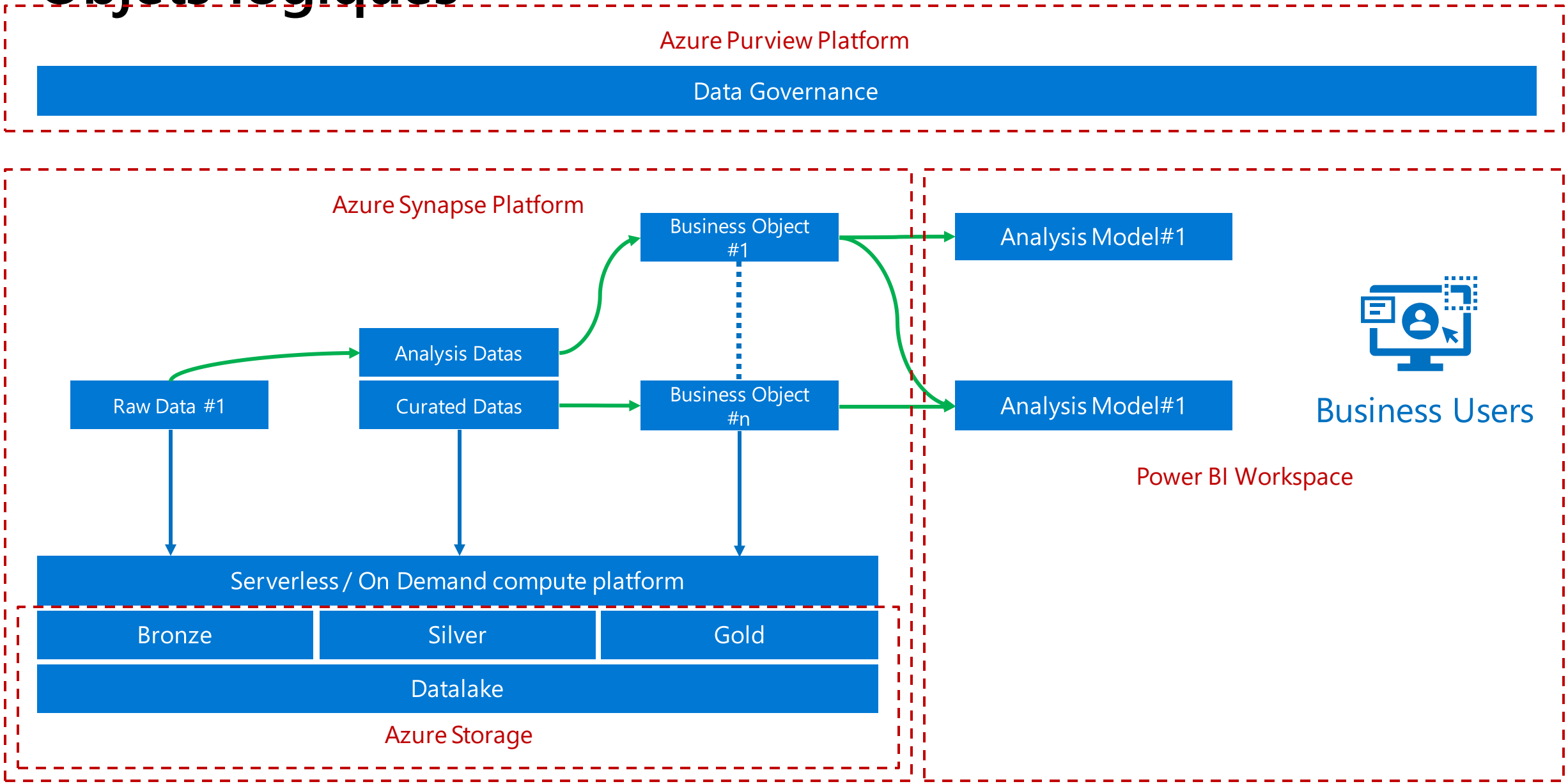
Explications et retour d'expérience

- Besoin d'expliquer finement l'approche analytique moderne
- Besoin d'expliquer les nouvelles avancées du cloud
  - Détachement de la scalabilité des services, plus de "capacity planning"
  - Détachement de la notion de performance
  - Consommation "on demand" de la BI moderne
  - Les limites d'une telle approche
  - **Réflexion sur "où" est calculé l'indicateur, modèles hybrides/composite**
  - Couts variables pour les métiers (dépendant de l'utilisation), besoin de changer l’approche budgétaire.

Objets logiques



Objets logiques

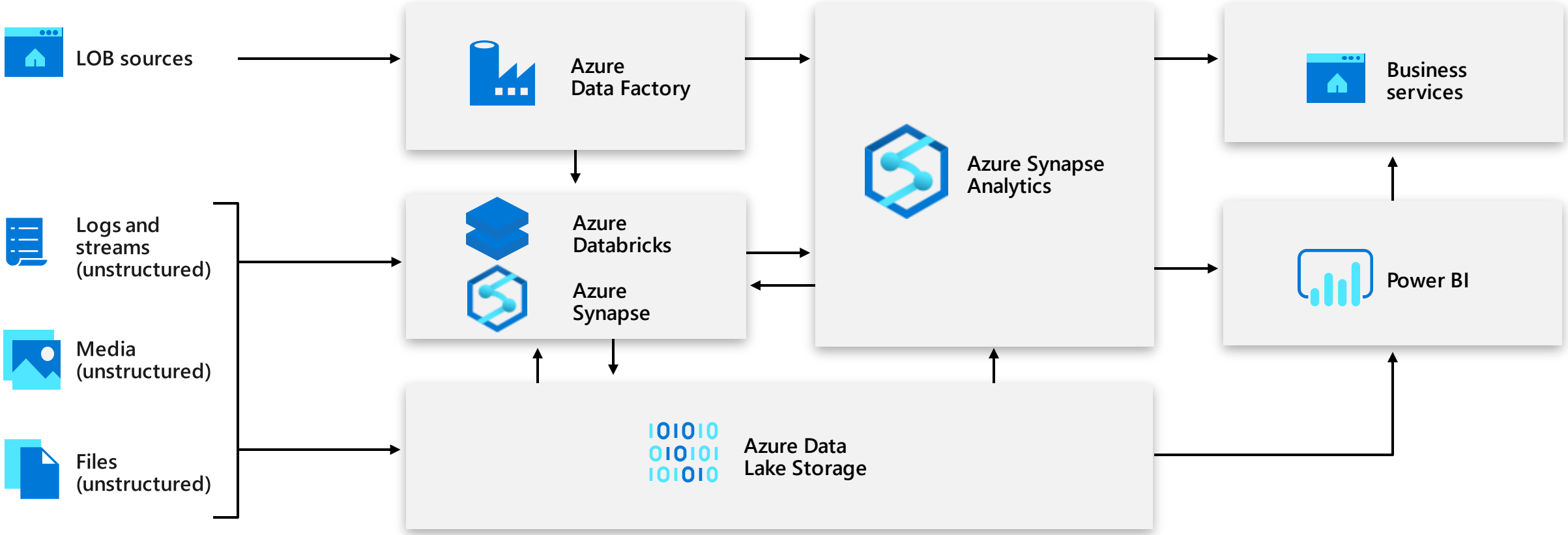


Les zones de l'approche "Datalake(house)"

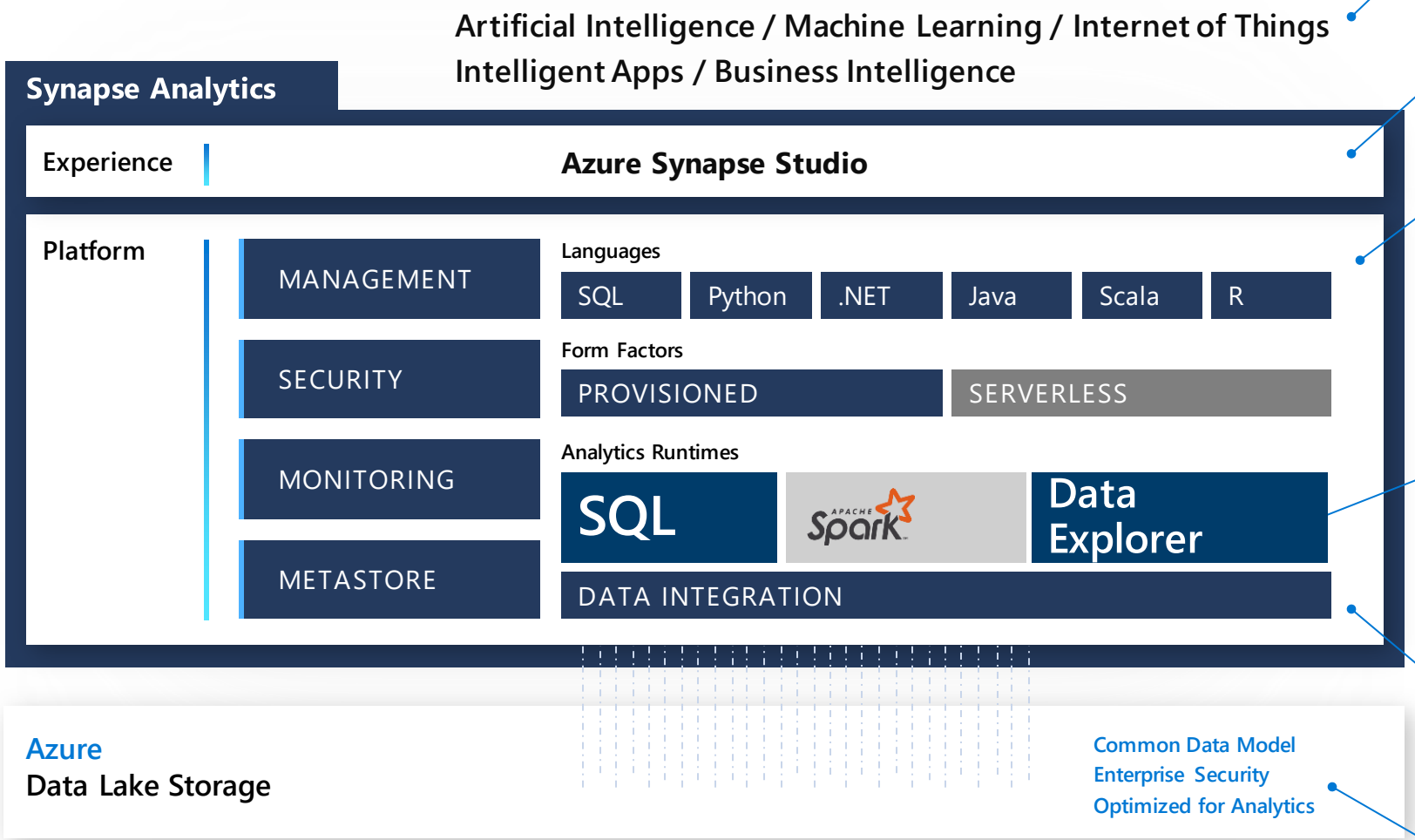
- Le "Datalake" se décompose en trois zones distinctes
- Ces zones sont réparties sur trois comptes de stockage Azure

- Zone Bronze
  - Cette zone correspond au dépôt des données brutes issues des systèmes amonts
  - Cette zone n'a pas de gouvernance spécifique
- Zone Silver
  - Cette zone est une zone de données partiellement nettoyées et analysables
  - L'accès est réservé aux "Data Engineer et Data Scientist" dans des approches de recherche et analyse
  - La gouvernance est minimaliste mais existe
- Zone Gold
  - Cette zone est la zone d'exposition des informations
  - Toutes données présentent dans celle-ci est une donnée vérifiée et exploitable dans l'entreprise (notion de trust)
  - Ces données sont exposées dans un outil de gouvernance d'accès dans l'entreprise
  - Gouvernance forte de la structure du datalake
  - Tout objet consommé dans l'entreprise provient de cette zone

L'approche moderne d'exposition de données



Moteurs de la plateforme Azure Synapse



Designed for analytics **workloads at any scale**

SaaS **developer experiences** for code free and code first

Multiple **languages** suited to different analytics workloads

Integrated analytics runtimes available provisioned and serverless on-demand

SQL Analytics offering T-SQL for batch, streaming and interactive processing

Spark for big data processing with Python, Scala, R and .NET

Data Explorer for time series processing

Integrated **platform services** for, management, security, monitoring, and metastore

Data **lake integrated** and Common Data Model aware

Performance

Elastic Architecture

Columnar Storage

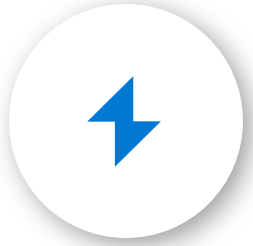
Columnar Ordering

Table Partitioning

Secondary Indexes

Hash Distribution

Materialized Views



Resultset Cache

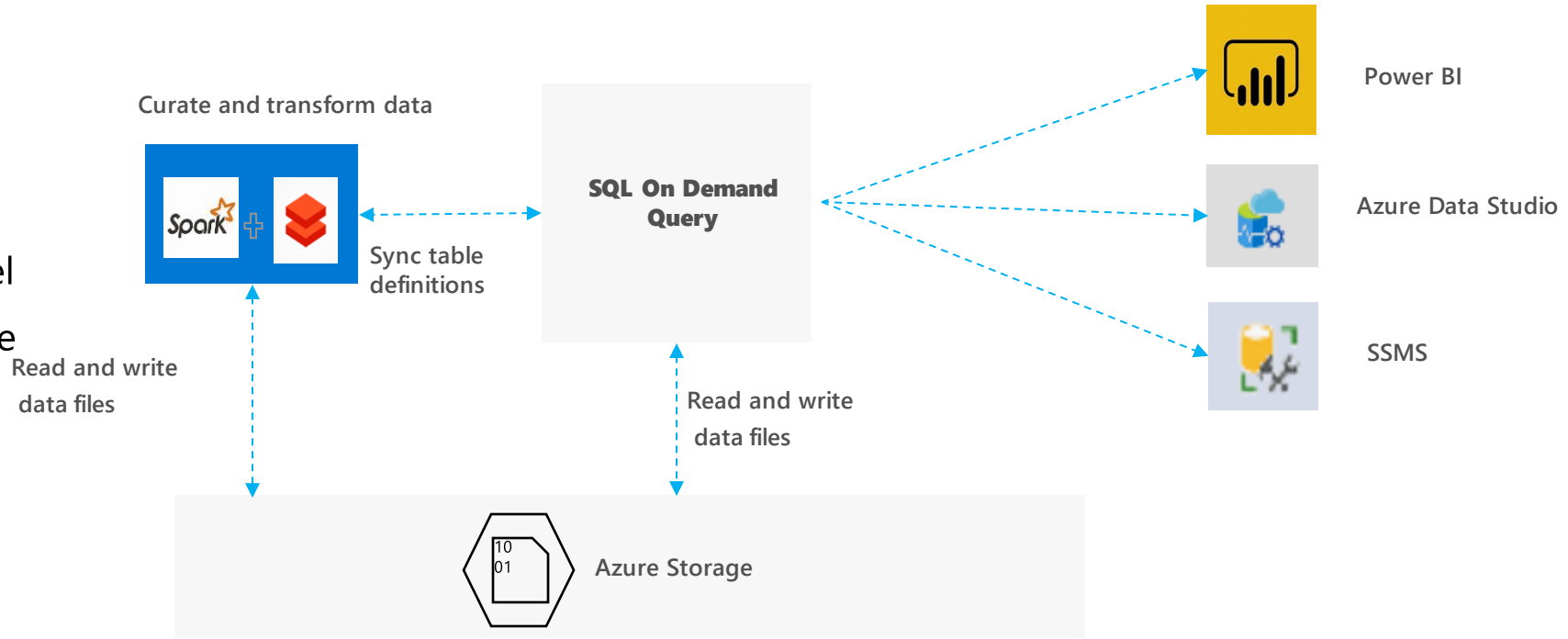
Transform with Synapse SQL Serverless / Pricing

Overview

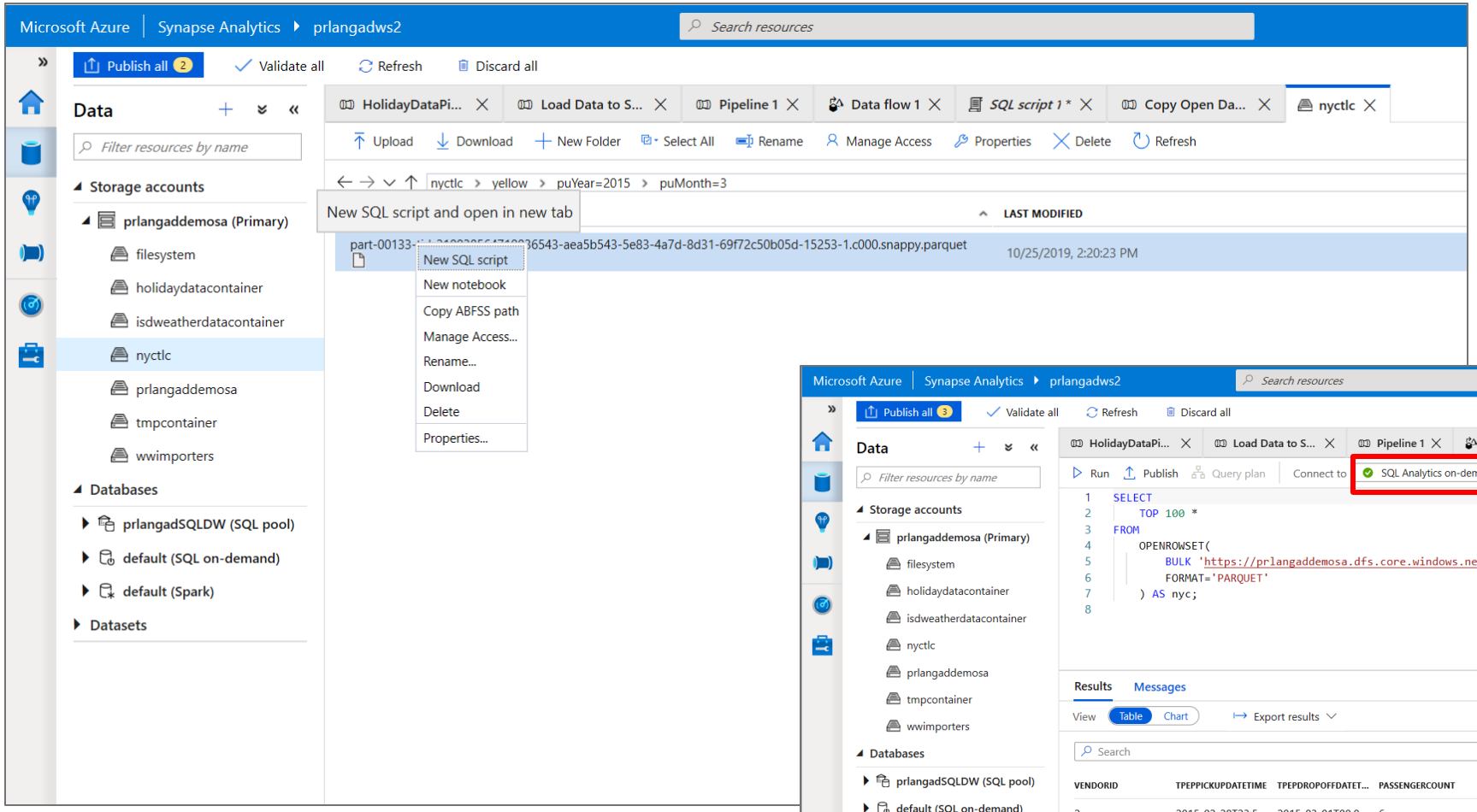
An interactive query service that provides T-SQL queries over high scale data in Azure Storage.

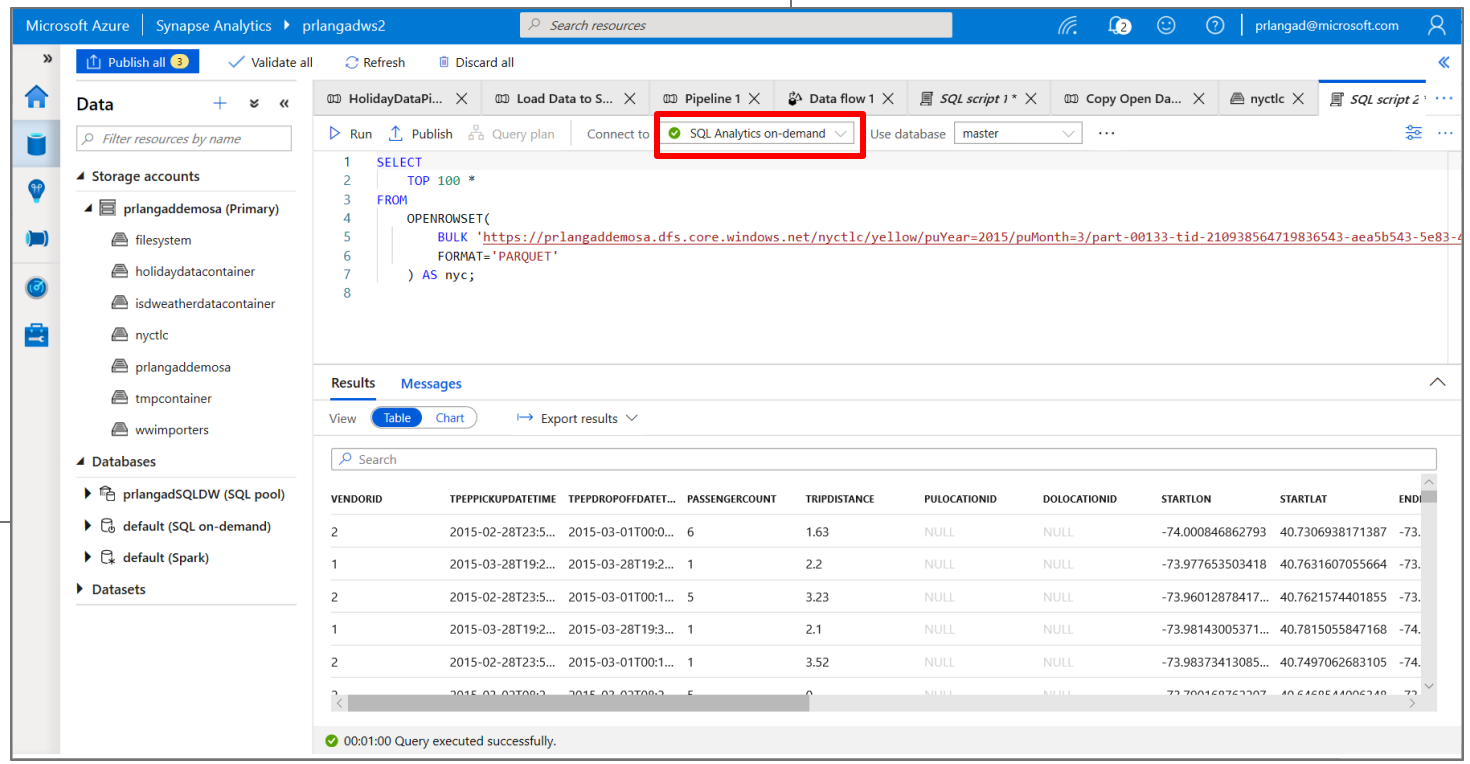
Benefits

- Pay-per-query with serverless model
- Query data in-place on the data lake with T-SQL (no ETL)
- Supports various file formats (Parquet, CSV, JSON)
- Integrates with Databricks, HDInsight, PowerBI, and the shared Synapse metastore



Synapse SQL Serverless – Querying on storage





Synapse SQL Serverless – Querying CSV File

Overview

Uses OPENROWSET function to access data

Benefits

Ability to read CSV File with

- no header row, Windows style new line
- no header row, Unix-style new line
- header row, Unix-style new line
- header row, Unix-style new line, quoted
- header row, Unix-style new line, escape
- header row, Unix-style new line, tab-delimited
- without specifying all columns

```
SELECT *
FROM OPENROWSET(
    BULK 'https://XXX.blob.core.windows.net/csv/population/population.csv',
    FORMAT = 'CSV',
    FIELDTERMINATOR =',',
    ROWTERMINATOR = '\n'
  )
WITH (
  [country_code] VARCHAR (5) COLLATE Latin1_General_BIN2,
  [country_name] VARCHAR (100) COLLATE Latin1_General_BIN2,
  [year] smallint,
  [population] bigint
) AS [r]
WHERE
  country_name = 'Luxembourg'
  AND year = 2017
```

	country_code	country_name	year	population
1	LU	Luxembourg	2017	594130

Synapse SQL Serverless – Querying specific files

Overview

filename – Provides file name that originates row result

filepath – Provides full path when no parameter is passed or part of path when parameter is passed that originates result

Benefits

Provides source name/path of file/folder for row result set

```
SELECT
    r.filepath() AS filepath
    ,r.filepath(1) AS [year]
    ,r.filepath(2) AS [month]
    ,COUNT_BIG(*) AS [rows]
FROM OPENROWSET(
        BULK 'https://XXX.blob.core.windows.net/csv/taxi/yellow_tripdata_*-*.csv',
        FORMAT = 'CSV',
        FIRSTROW = 2   )
WITH (
    vendor_id INT,
    pickup_datetime DATETIME2,
    dropoff_datetime DATETIME2,
    passenger_count SMALLINT,
    trip_distance FLOAT,
   <... columns>
) AS [r]

WHERE r.filepath(1) IN ('2017')
    AND r.filepath(2) IN ('10', '11', '12')

GROUP BY    r.filepath() ,r.filepath(1) ,r.filepath(2)
ORDER BY    filepath
```

Example of filename function

```
SELECT
    r.filename() AS [filename]
    ,COUNT_BIG(*) AS [rows]
FROM OPENROWSET(
        BULK 'https://XXX.blob.core.windows.net/csv/taxi/yellow_tripdata_2017-1*.csv',
        FORMAT = 'CSV',
        FIRSTROW = 2
    )
    WITH (
        vendor_id INT,
        pickup_datetime DATETIME2,
        dropoff_datetime DATETIME2,
        passenger_count SMALLINT,
        trip_distance FLOAT,
        <...columns>
    ) AS [r]

GROUP BY r.filename()

ORDER BY [filename]
```

	filename	rows
1	yellow_tripdata_2017-10.csv	9768815
2	yellow_tripdata_2017-11.csv	9284803
3	yellow_tripdata_2017-12.csv	9508276

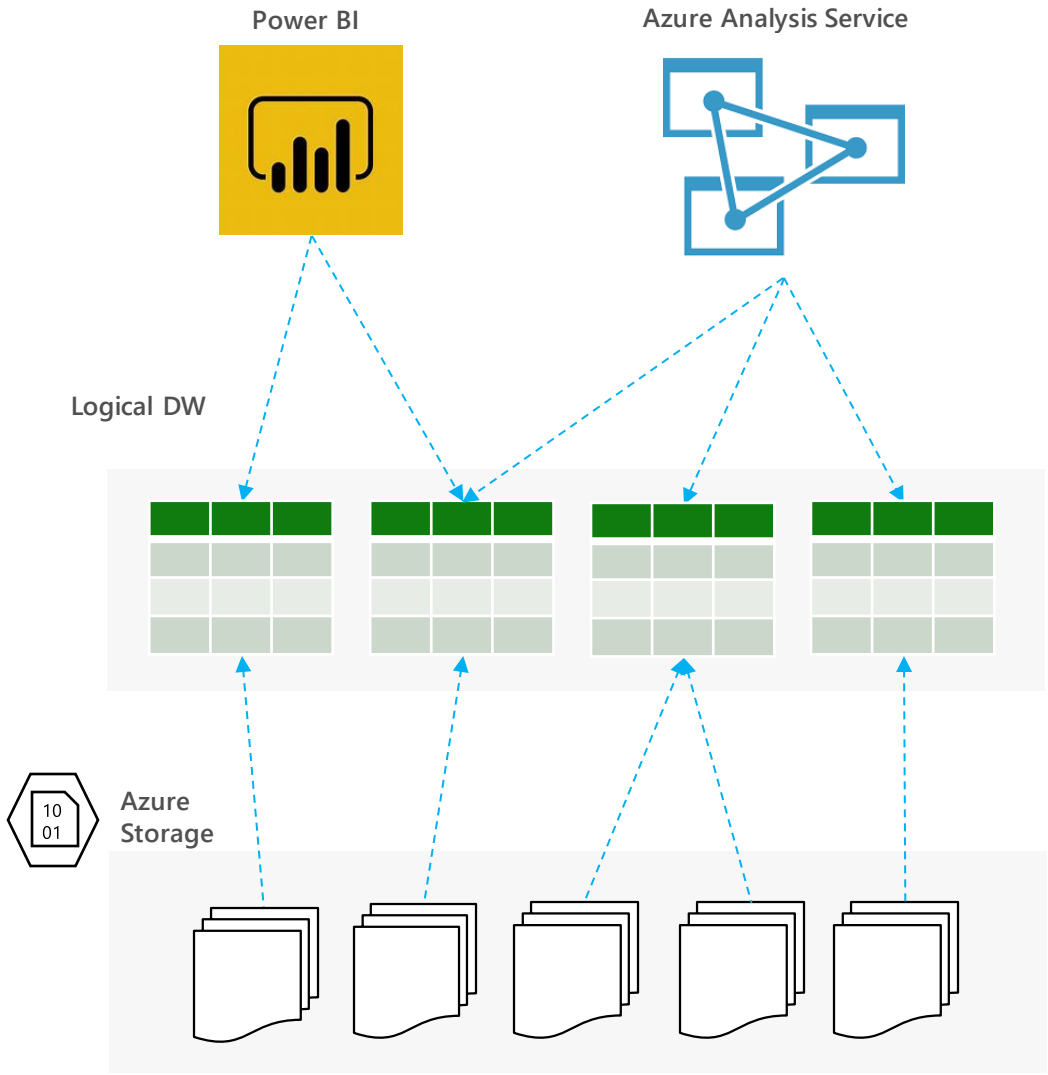
Synapse serverless SQL pool as a logical data warehouse

Overview

Logical relational layer on top of physical files in Azure Storage.

Benefits

- Abstract physical storage and file formats using well understandable relational concepts such as tables and views.
- Direct connector to Azure storage for large ecosystem of BI tools
- BI tools that use SQL can work with files on storage
  - Analytic tools use external tables that represent proxy to actual files.
  - No need for custom connectors in BI tools.
- Provides complex data processing (joining and aggregation) on top of raw files.
- Apply enterprise-ready security model and access control using battle-tested SQL Server permission model on top of Azure storage files



Logical Data Warehouse views

Overview

serverless SQL pool logical data warehouse views are created on external files placed in customer Azure storage

Benefits

Create SQL views on externally stored data

Access files using the view from various tools and language

Leverage rich T-SQL language to process and analyze data in external files exposed via views

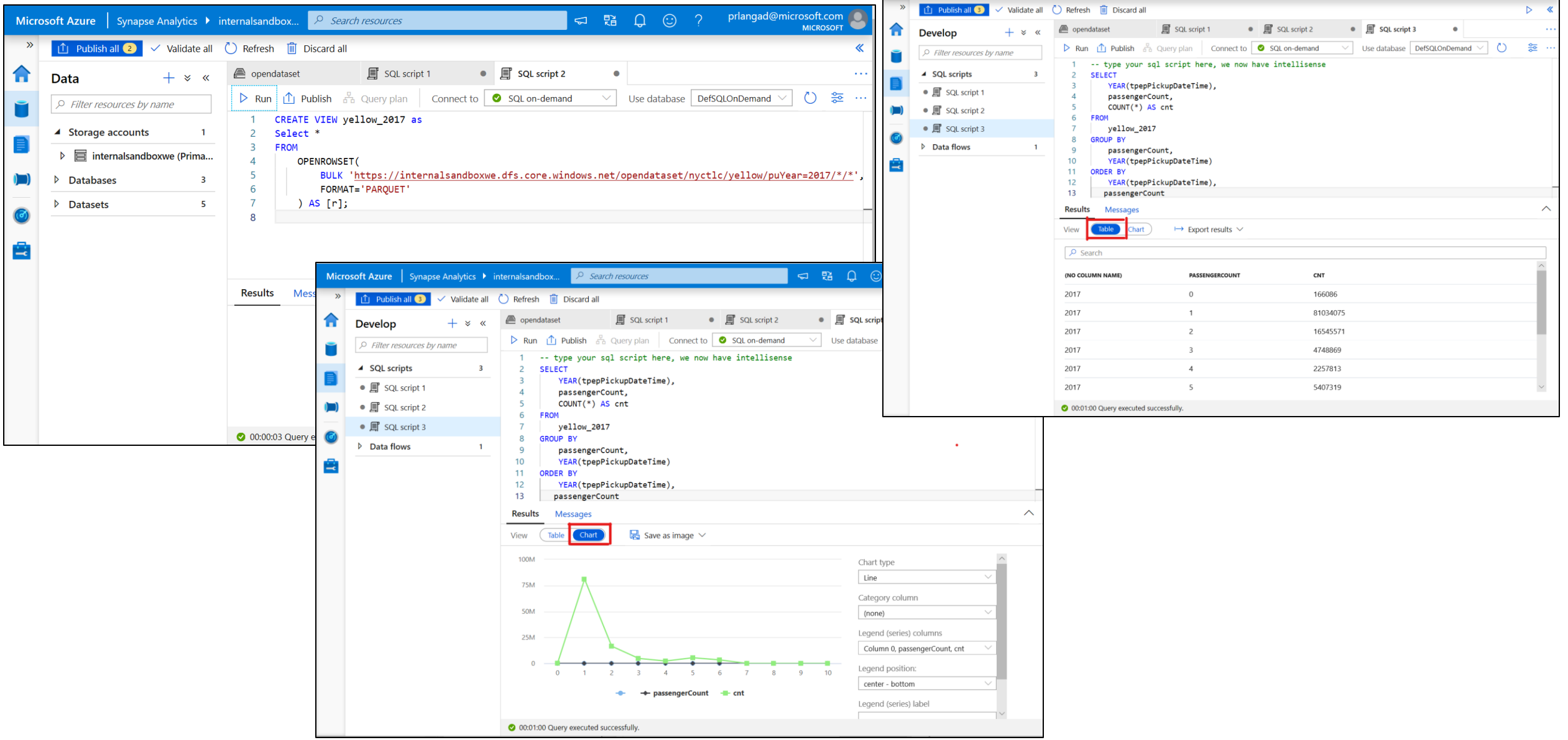
Create PowerBI reports on the views created on external data

```
USE [mydbname]
GO

DROP VIEW IF EXISTS  populationView
GO

CREATE VIEW populationView AS
SELECT *
FROM OPENROWSET(
      BULK 'https://XYZ.blob.core.windows.net/csv/population/*.csv',
      FORMAT = 'CSV',
      FIELDTERMINATOR =',',
      ROWTERMINATOR = '\n'
   )
WITH (
   [country_code] VARCHAR (5) ,
   [country_name]  VARCHAR (100),
   [year] smallint,
   [population] bigint
) AS [r]
```

```
SELECT
   country_name, population
FROM populationView
WHERE
   [year] = 2019
ORDER BY
   [population] DESC
```

	country_name	population
1	China	1389618778
2	India	1311559204
3	United States	331883986
4	Indonesia	264935824
5	Pakistan	210797836
6	Brazil	210301591
7	Nigeria	208679114
8	Bangladesh	161062905
9	Russia	141944641
10	Mexico	127318112

Creating views



Logical Data Warehouse - tables

Overview

Create external tables that reference external files in your serverless SQL pool logical data warehouse

Benefits

Create external tables that reference set of files on Azure storage.

Join and transform multiple tables in the same query.

Enables you to analyze external files with the same experience that you have in classic databases.

Manage column statistics in external tables.

Manage access rights per table.

Create PowerBI reports on the views created on external data

```
USE [mydbname]
GO

DROP TABLE IF EXISTS dbo.Population
GO
CREATE EXTERNAL TABLE dbo.Population (
   country_code VARCHAR (5) COLLATE Latin1_General_BIN2,
   country_name VARCHAR (100) COLLATE Latin1_General_BIN2,
   year smallint,
   population bigint
)
WITH(
      LOCATION = '/csv/population/population-*.csv',
      DATA_SOURCE = MyAzureStorage,
      FILE_FORMAT = MyAzureCSVFormat
)
```

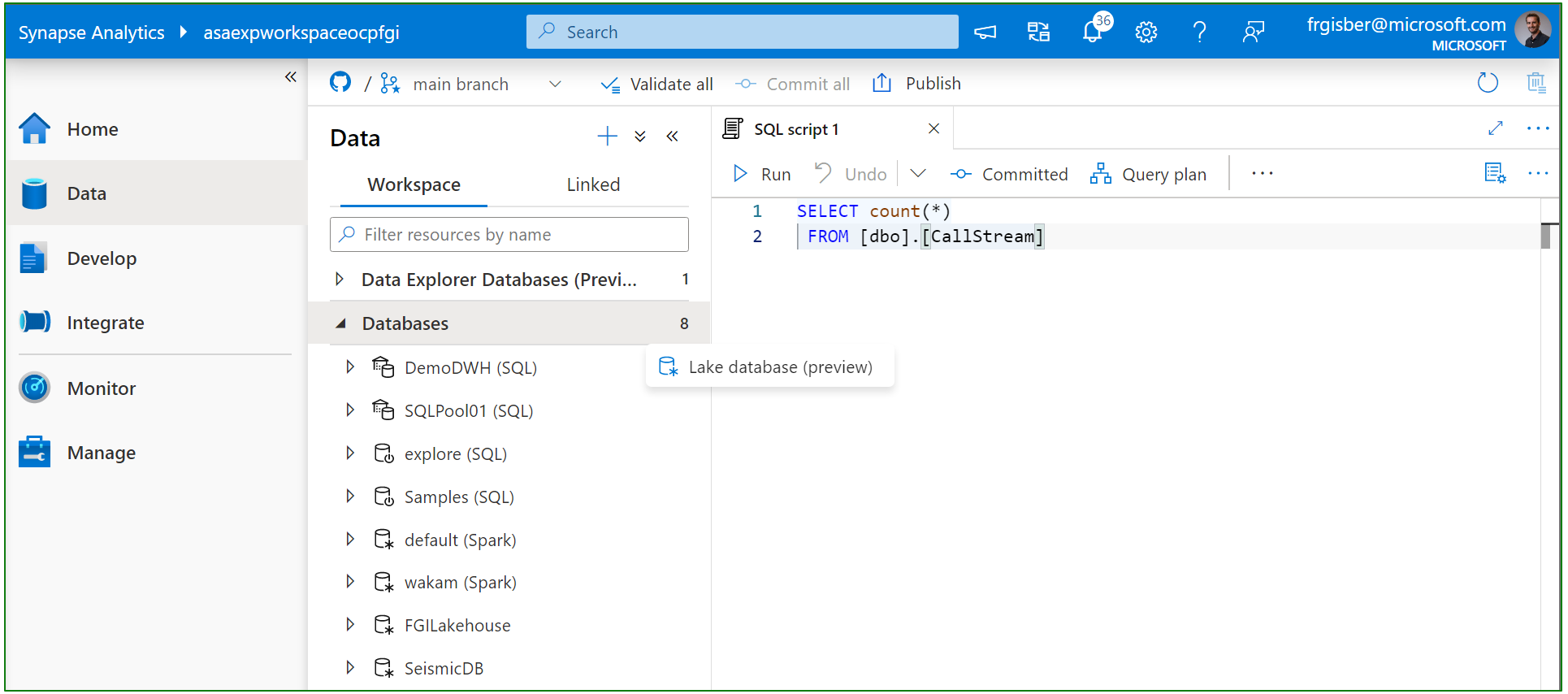
```
CREATE STATISTICS stat_country_name
ON dbo.Population(country_name);
```

```
SELECT
   country_name, population
FROM population
WHERE year = 2019
ORDER BY population DESC
```

	country_name	population
1	China	1389618778
2	India	1311559204
3	United States	331883986
4	Indonesia	264935824
5	Pakistan	210797836
6	Brazil	210301591
7	Nigeria	208679114
8	Bangladesh	161062905
9	Russia	141944641
10	Mexico	127318112

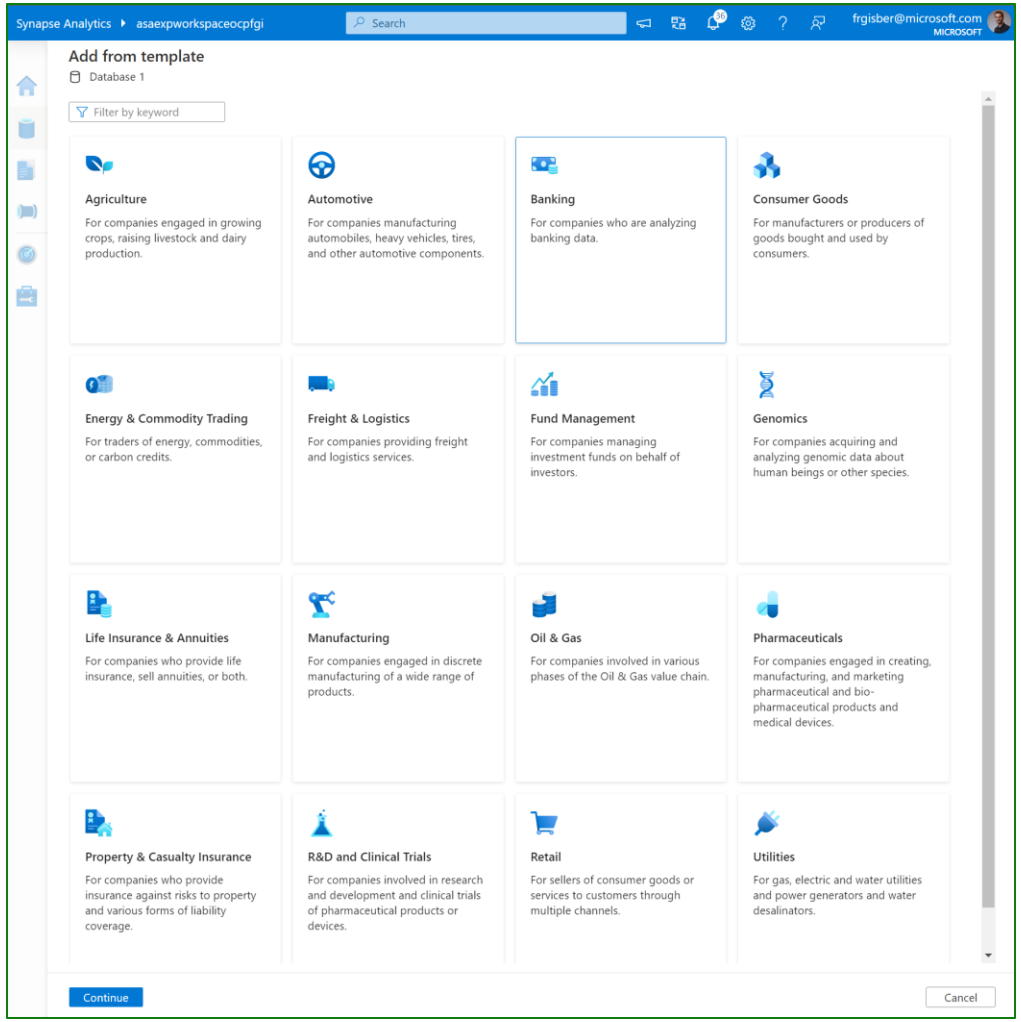
Lakehouse Database Engine

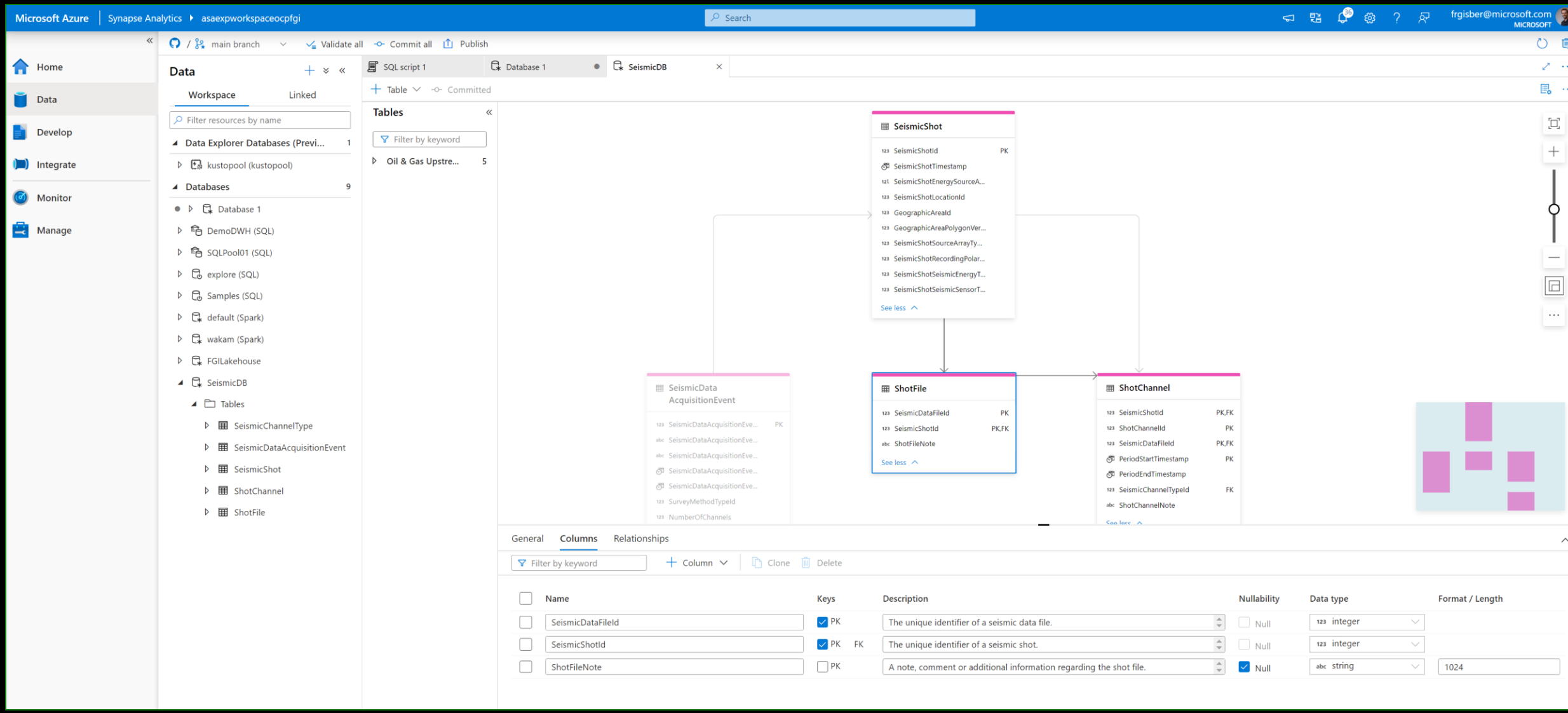
- New Lakehouse Database capabilities based on Synapse Spark



Data Lakehouse database

- Based on entities templates or custom design




Microsoft Azure
Synapse Analytics
asaexpworkspaceocpfgi
Search
frgisber@microsoft.com
MICROSOFT
Home
Data
Develop
Integrate
Monitor
Manage
main branch
Validate all
Commit all
Publish
Data
Workspace
Linked
Filter resources by name
Data Explorer Databases (Previ... 1
kustopool (kustopool)
Databases 9
Database 1
DemoDWH (SQL)
SQLPool01 (SQL)
explore (SQL)
Samples (SQL)
default (Spark)
wakam (Spark)
FGILakehouse
SeismicDB
Tables
SeismicChannelType
SeismicDataAcquisitionEvent
SeismicShot
ShotChannel
ShotFile
SQL script 1
Database 1
SeismicDB
Table
Committed
Tables
Filter by keyword
Oil & Gas Upstre... 5
SeismicShot
SeismicShotId PK
SeismicShotTimestamp
SeismicShotEnergySourceA...
SeismicShotLocationId
GeographicAreaId
GeographicAreaPolygonVer...
SeismicShotSourceArrayTy...
SeismicShotRecordingPolar...
SeismicShotSeismicEnergyT...
SeismicShotSeismicSensorT...
See less
SeismicData AcquisitionEvent
SurveyMethodTypeId
NumberOfChannels
ShotFile
SeismicDataFileId PK
SeismicShotId PK,FK
ShotFileNote
See less
ShotChannel
SeismicShotId PK,FK
ShotChannelId PK
SeismicDataFileId PK,FK
PeriodStartTimestamp PK
PeriodEndTimestamp
SeismicChannelTypeId FK
ShotChannelNote
General
Columns
Relationships
Filter by keyword
Column
Clone
Delete
Name
Keys
Description
Nullability
Data type
Format / Length
SeismicDataFileId
PK
The unique identifier of a seismic data file.
Null
integer
SeismicShotId
PK FK
The unique identifier of a seismic shot.
Null
integer
ShotFileNote
PK
A note, comment or additional information regarding the shot file.
Null
string
1024

Spécificités des moteurs SQL

- Moteur SQL dédié
  - Puissance dédiée scalable horizontalement
  - Moteur MPP (Massivement parallèle)
  - 60 nœuds de stockages et n nœuds de calcul
  - Réponse linéaire des requêtes
  - Exposition d'objets
    - Logiques sous la forme de tables externes
    - Logiques sous la forme de vues (pouvant être matérialisées, donc objets physiques)
    - Physiques sous la forme de tables ou vues avec données embarquées

- Moteur SQL Serverless
  - Puissance "on demand"
  - Moteur non MPP
  - Scalabilité automatique et non linéaire (mais prédictible)
  - Exposition d'objets **logiques**

Spécificités du moteurs Spark

- Le moteur Spark est disponible dans le service Azure Synapse
- Ce moteur possède deux "form factor"
  - Cluster Spark à disposition dans un format managé, "auto scalable" et activé à la demande
  - Objets Spark disponibles en mode "serverless"

- Possibilité de matérialiser les résultats des travaux de recherches, analyses, etc. dans des tables Spark
- Les tables Spark sont exposées en mode "Serverless"

- L'accès est réalisé via le point d'accès SQL Serverless **sans que le moteur Spark soit démarré**

Power BI Embedded pour l'exposition et REST API

PowerBI REST APIs https://docs.microsoft.com/en-us/rest/api/power-bi

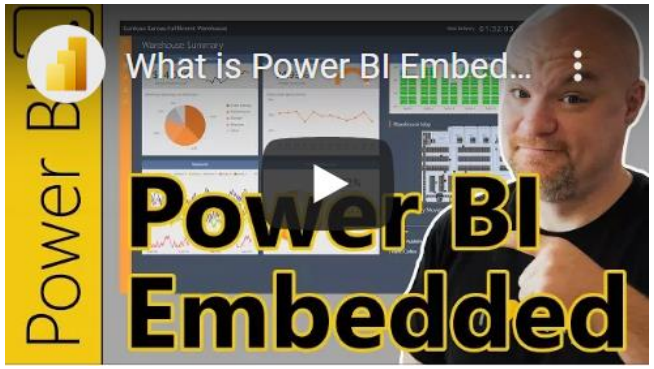
PowerBI Playground https://microsoft.github.io/PowerBI-JavaScript/demo/v2-demo/index.html#

Getting started
Set up your Power BI embedding environment
Power BI JavaScript API wiki
Power BI embedding documentation

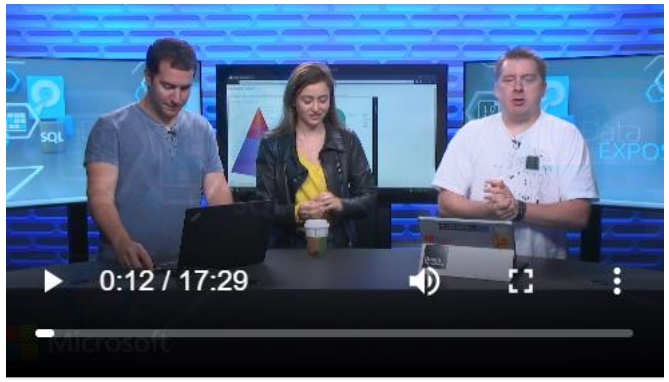
Useful links
Power BI Embedded on Azure
Power BI community
Power BI Ideas - APIs and embedding

Support
Power BI Embedded FAQ
Power BI Embedded troubleshooting
Power BI Support

What is Power BI Embedded



Microsoft Power BI Embedded update



Microsoft Power BI Embedded update

https://playground.powerbi.com/

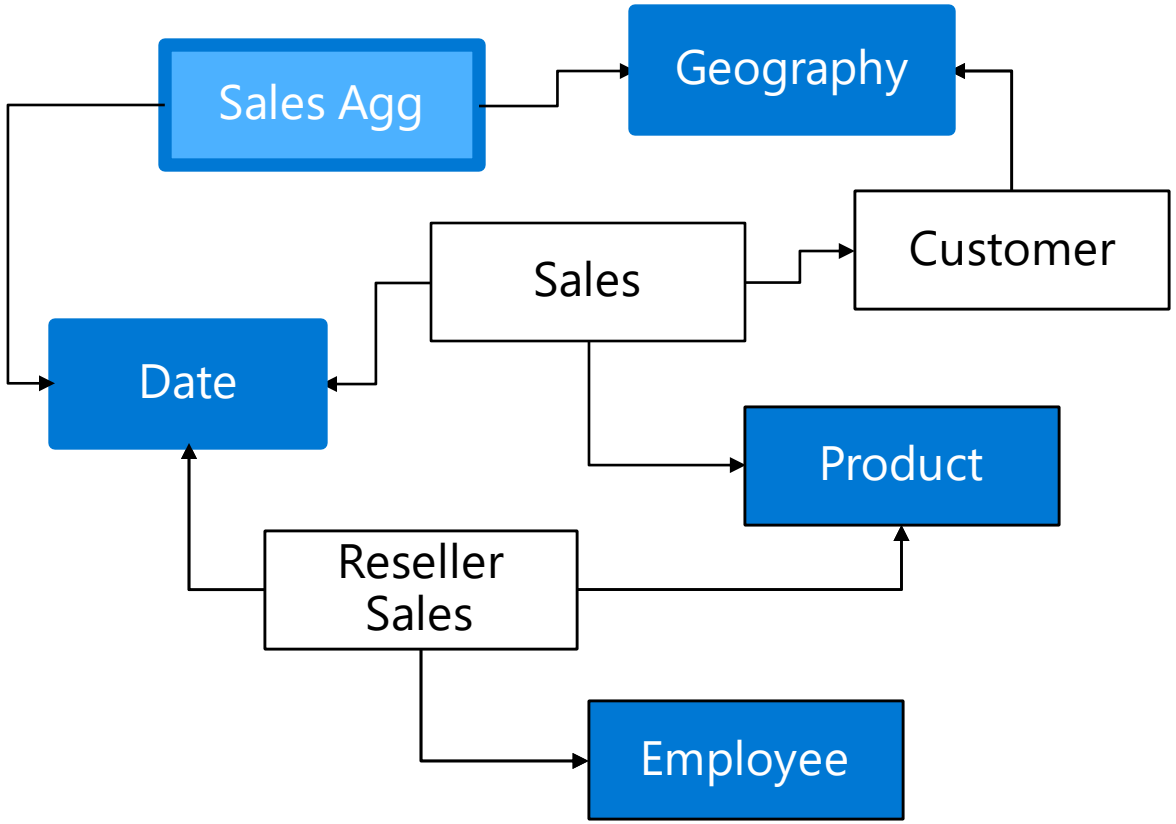
Rapports Dynamiques et Paginés

- Deux approches de rapports consommant les objets

- Rapport / Dashboard dynamiques PowerBI
  - Rapports très graphiques
  - Modèle de donnée embarqué ou délégué
  - Accès aux objets logiques dans une approche hybride (modèle composite)
  - Très adapté aux exploitations de résultats calculés, agrégés, etc.
  - Requêtes générées par l'outil
  - Utilisation de Power BI Desktop

- Rapport Paginés
  - Rapports graphiques, paramétrables
  - Adaptés aux approches de rapports sur les données de détails (listing, recherche d'information très fines)
  - Contrôle totale de la requête
  - Utilisation de Power BI Report Designer

Import, Direct Query et Agrégats




```
SummarizeColumns(
        Date[Year],
        Geography[City],
        "Sales", Sum(Sales[Amount])
)
```

DirectQuery

```
SELECT [Year],
       [Name],
       SUM([Amount]) AS [Amount]
FROM   [Sales]
INNER JOIN [Date] ON ...
INNER JOIN [Customer] ON ...
GROUP BY [Year],
         [Name]
```

- Data Integration
- Data Engineering
- Data Warehousing
- Data Science
- Observational Analytics
- Business Intelligence
- Governance

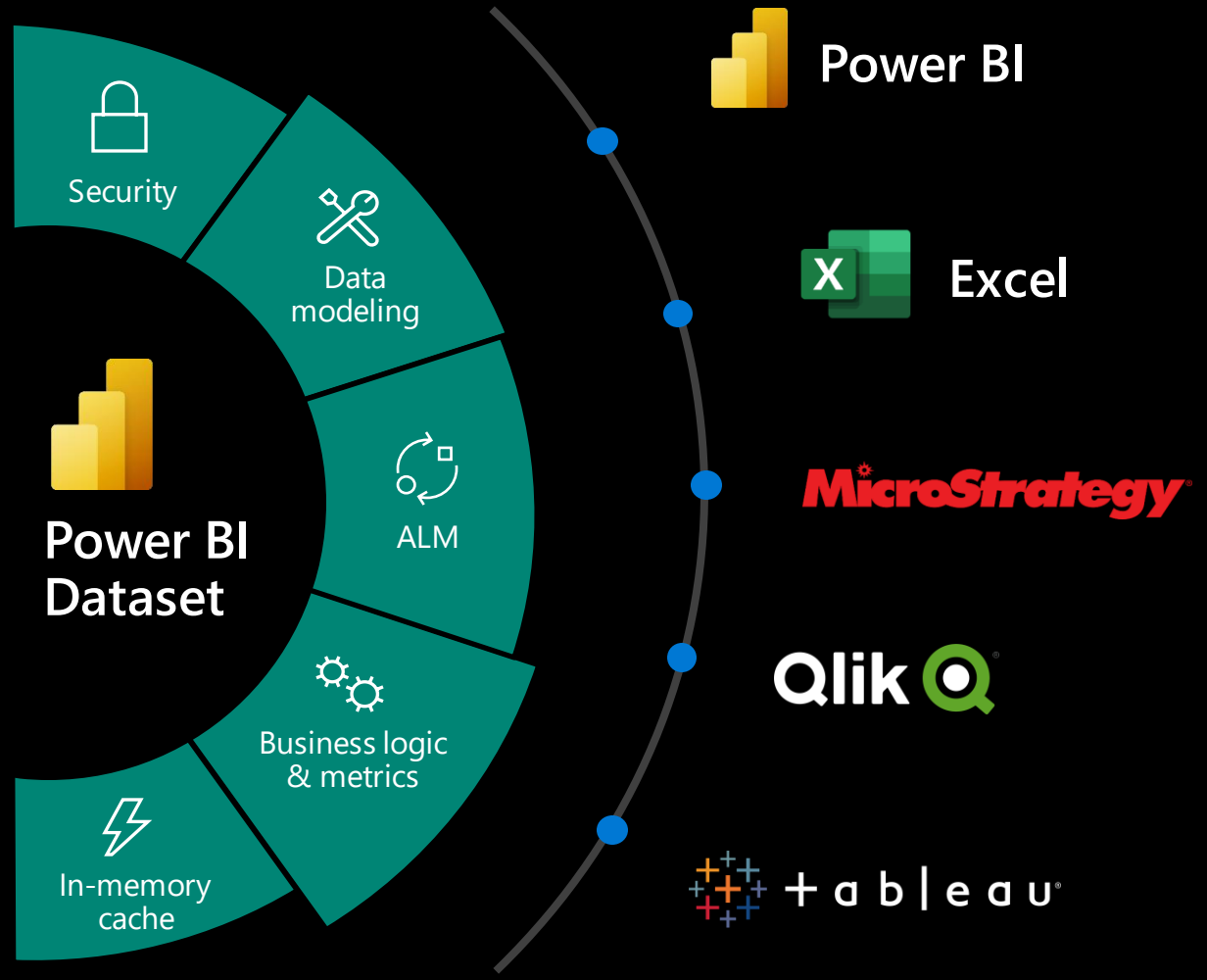
Business Intelligence

Business Intelligence

Generally Available

World's leading OLAP engine

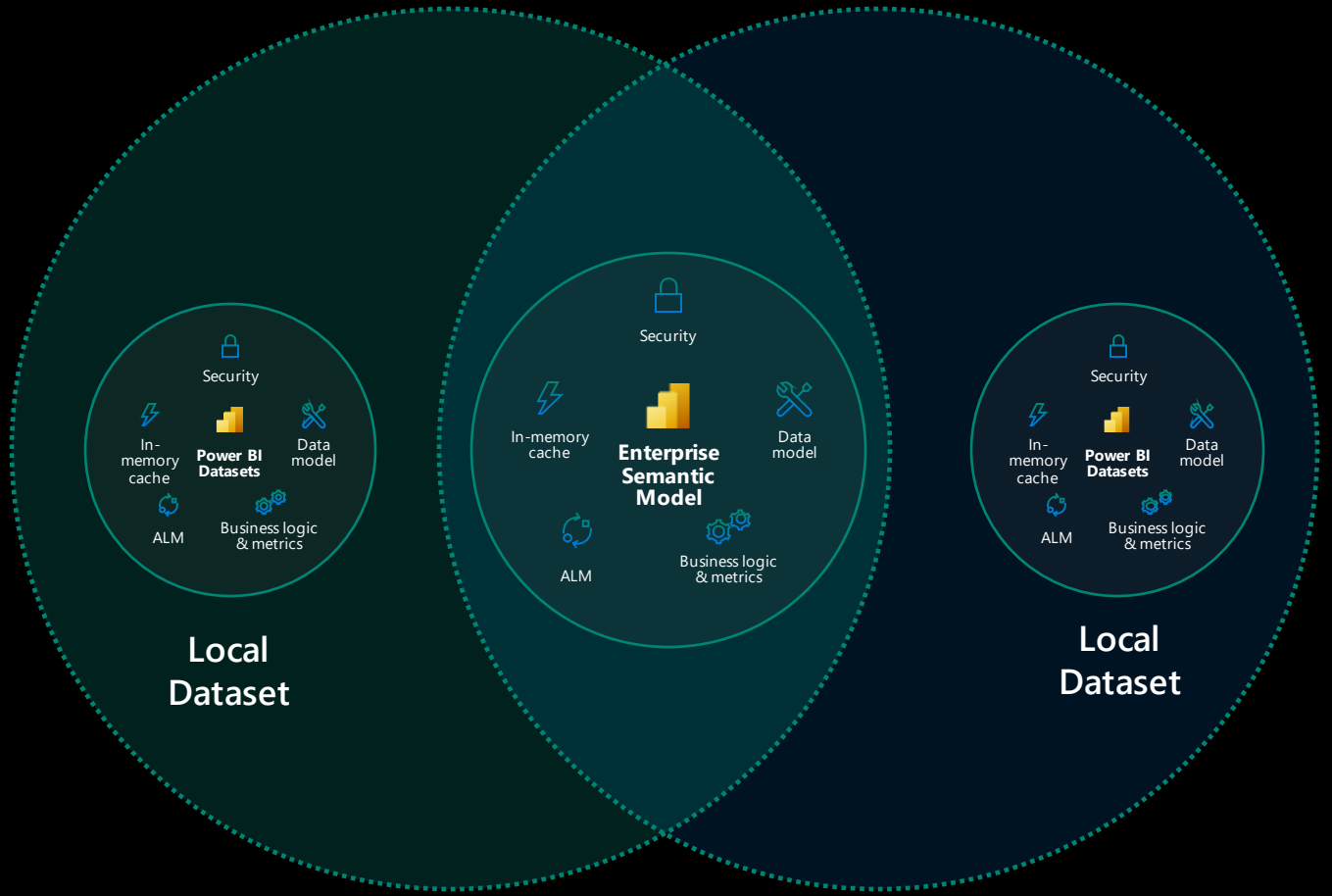
Blazing fast performance with connectivity for a variety of data visualization applications



Generally Available

Combine enterprise and local datasets

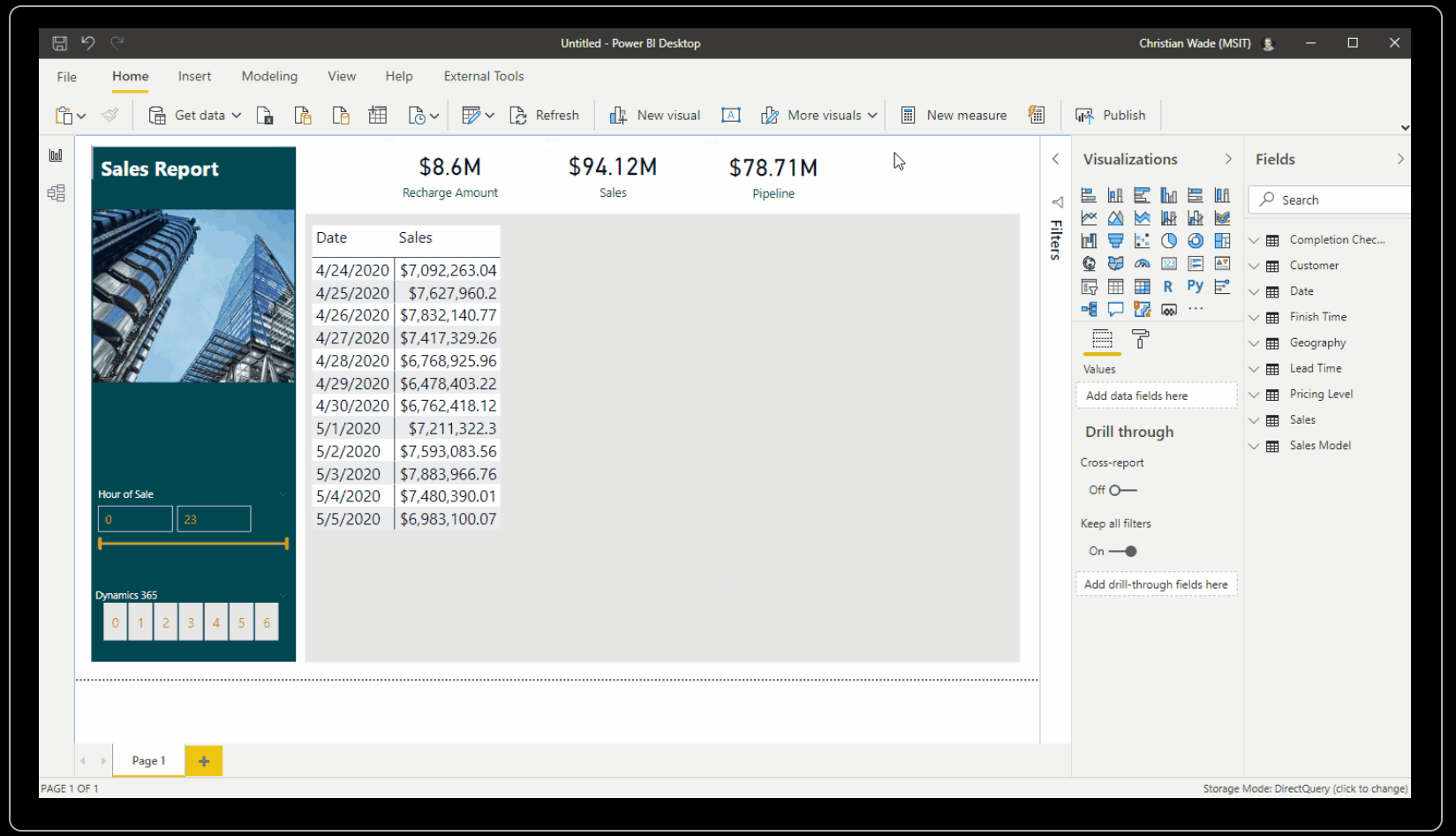
Seamless evolution from self-service BI to an enterprise semantic model for company wide adoption



Public Preview

Composite models

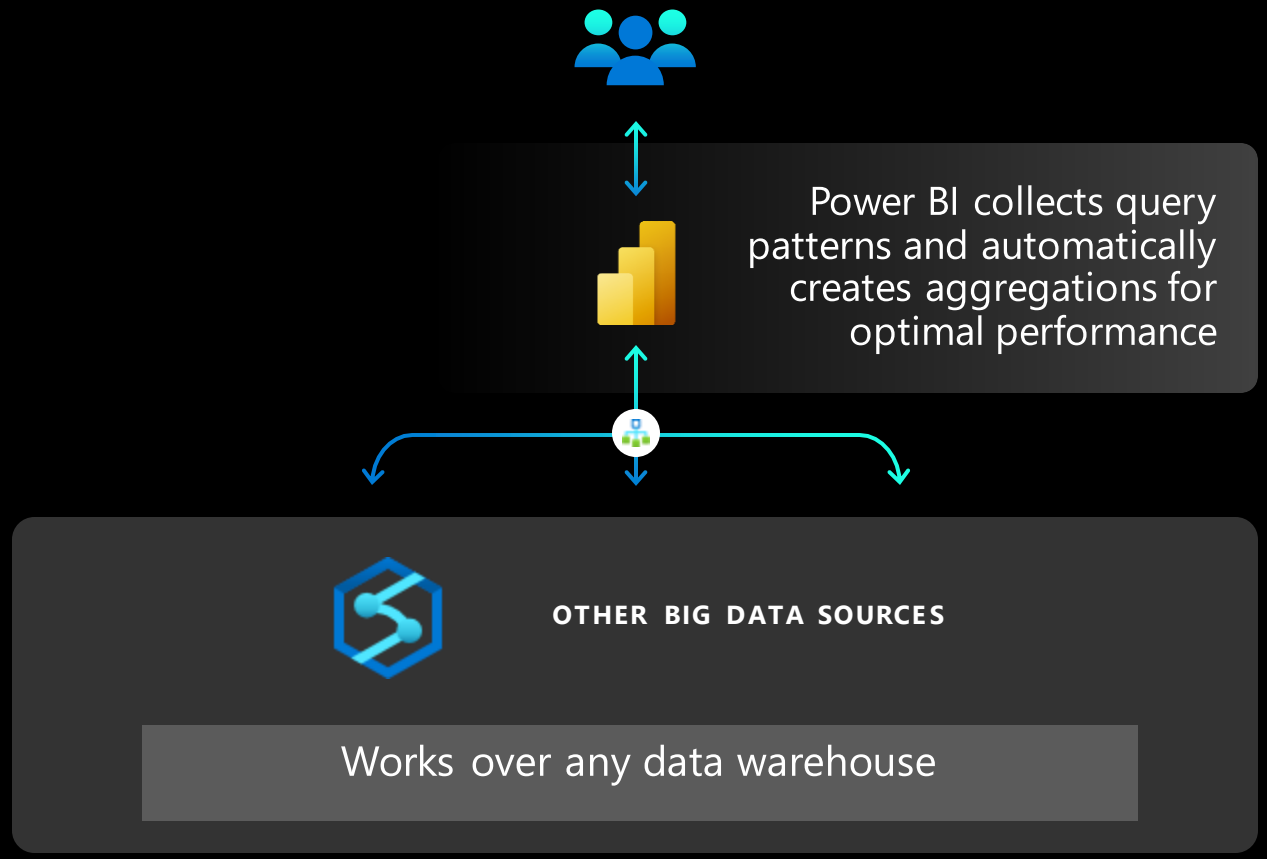
Seamlessly combine and extend self-service BI with corporate BI models



Public Preview

Automatic Aggregations

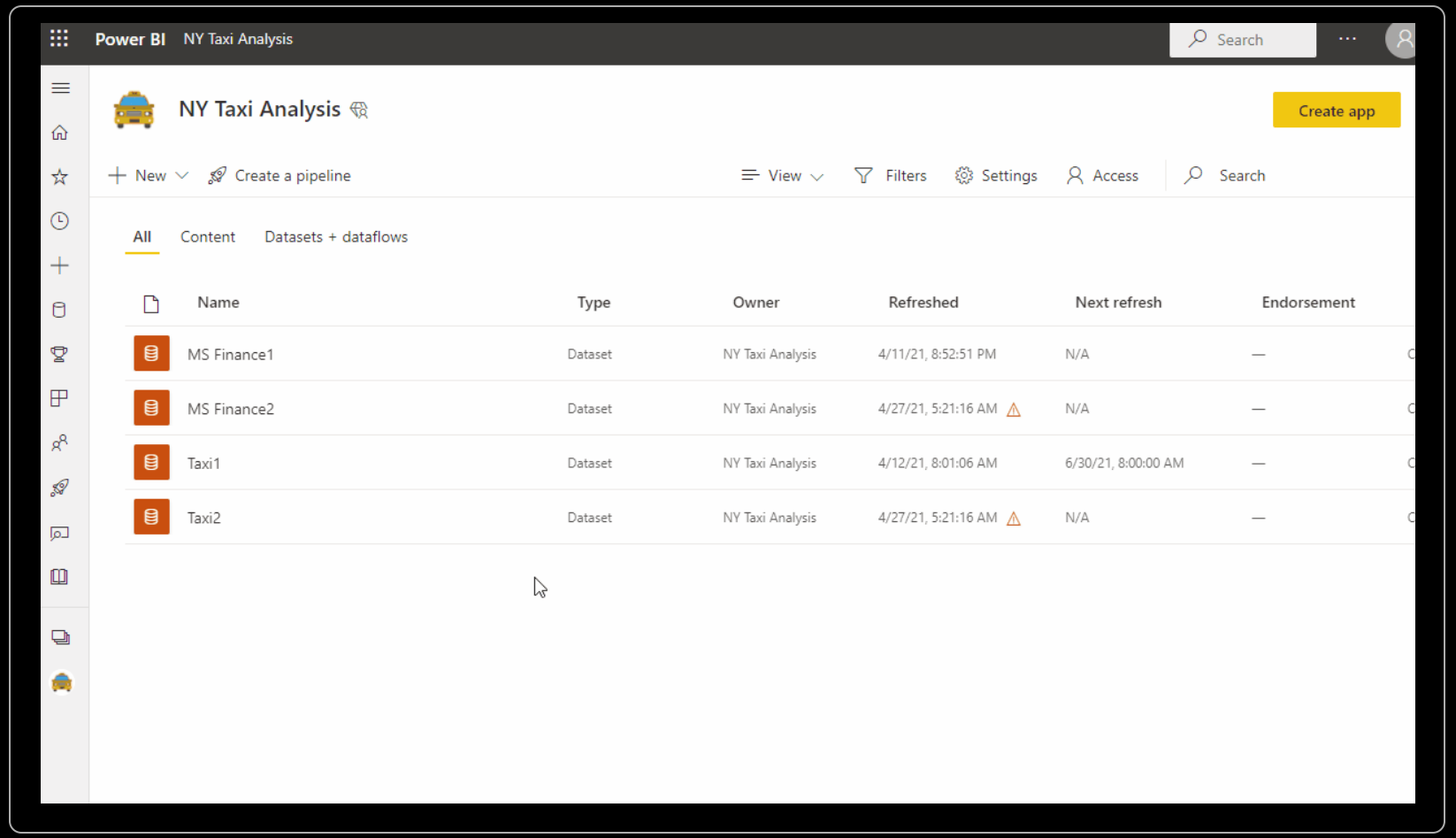
Automatically learns about customer usage patterns and create aggregates to optimize performance and reduce cost



Public Preview

Automatic Aggregations

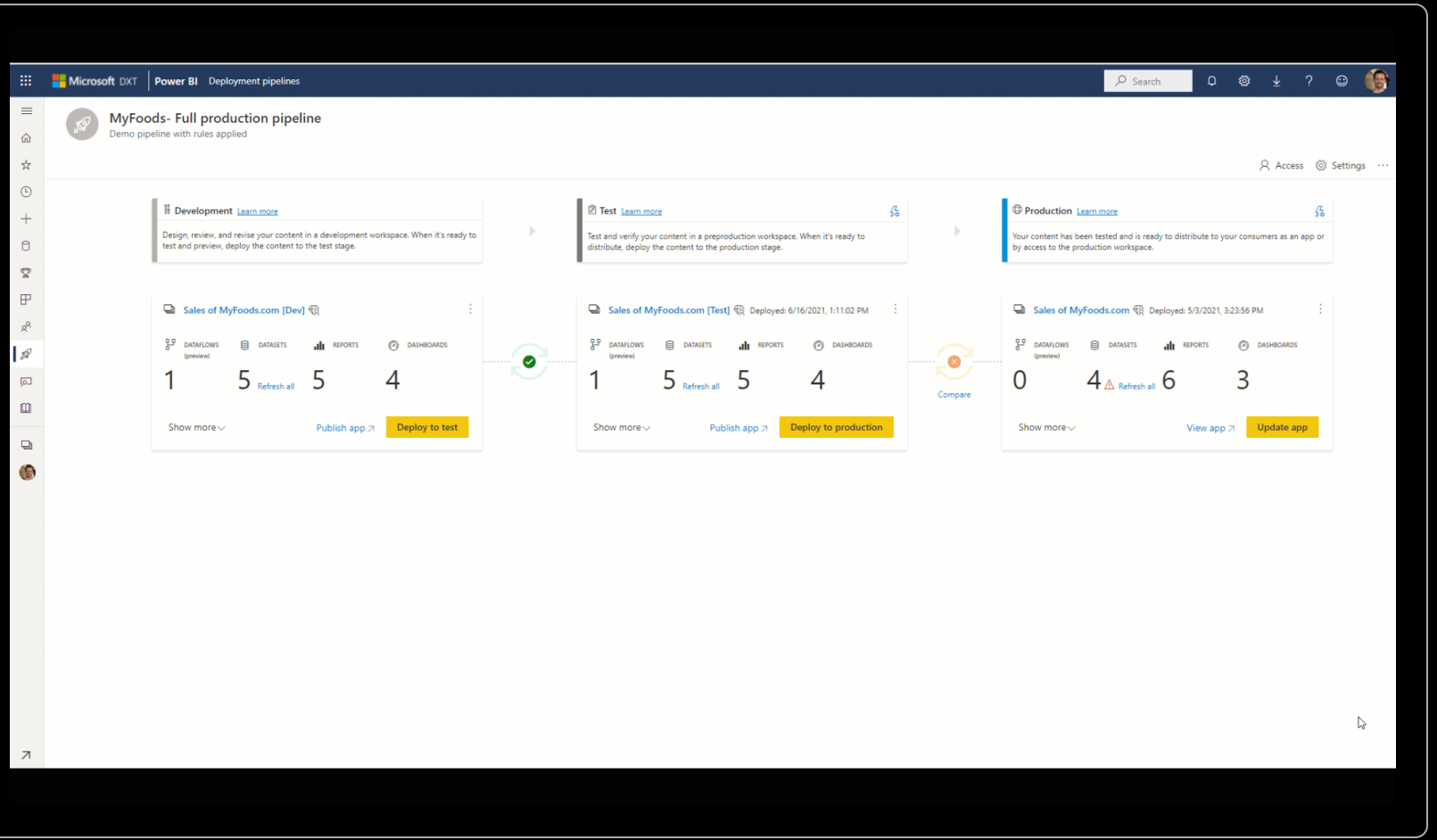
AI driven self-optimizing performance improvement



Generally Available

Controlled change management

Power BI deployment pipelines enable efficient and reusable release processes



Generally Available

Premium Gen2

Unmatched large-scale analytics with simple low-overhead administration

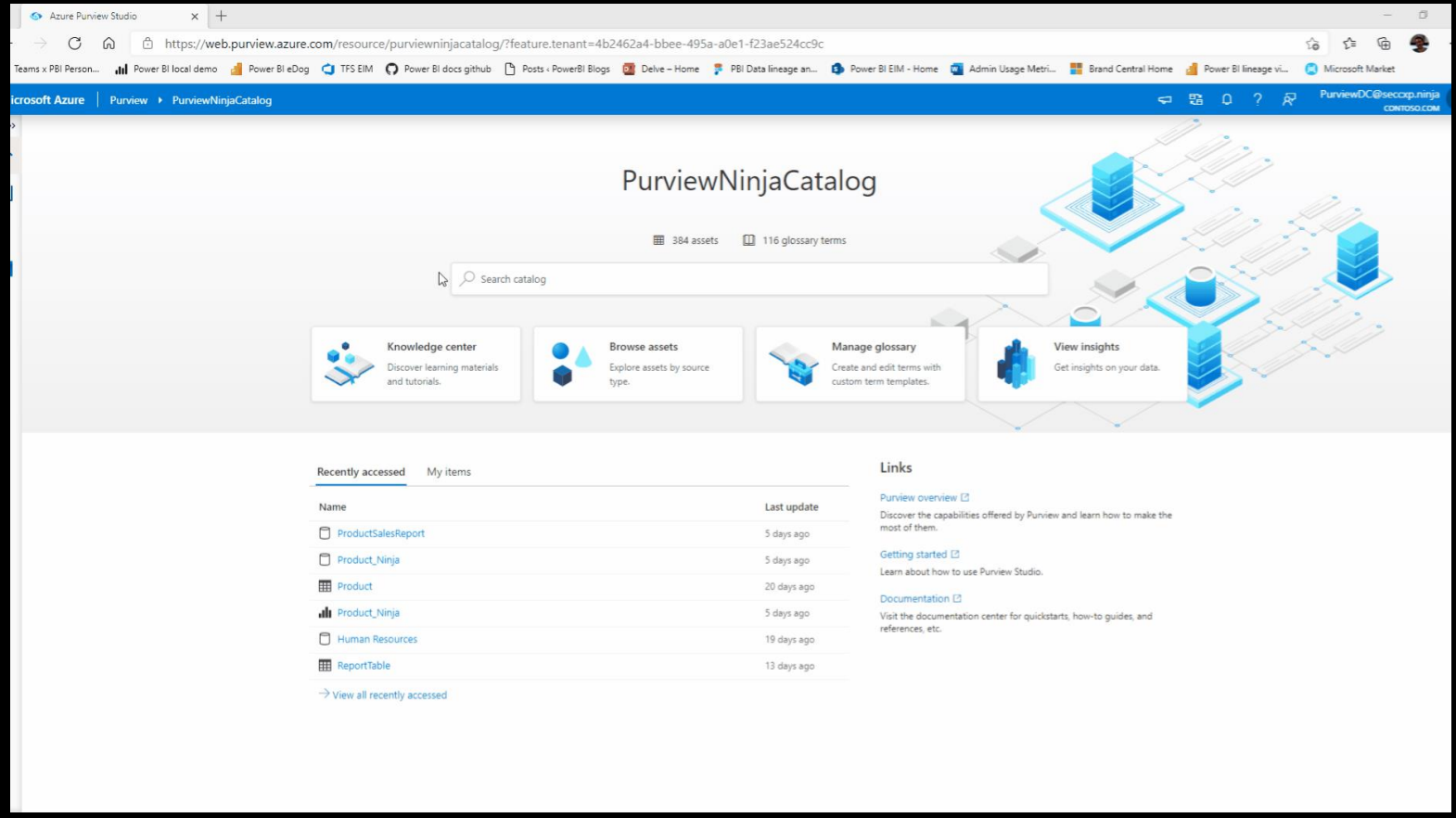
Over 50%

Premium Customer Nodes running on Gen 2 two weeks after release

Generally Available

Power BI + Azure Purview

Enhanced governance and cataloging capabilities integrated with Power BI



- Data Integration
- Data Engineering
- Data Warehousing
- Data Science
- Observational Analytics
- Business Intelligence
- Governance

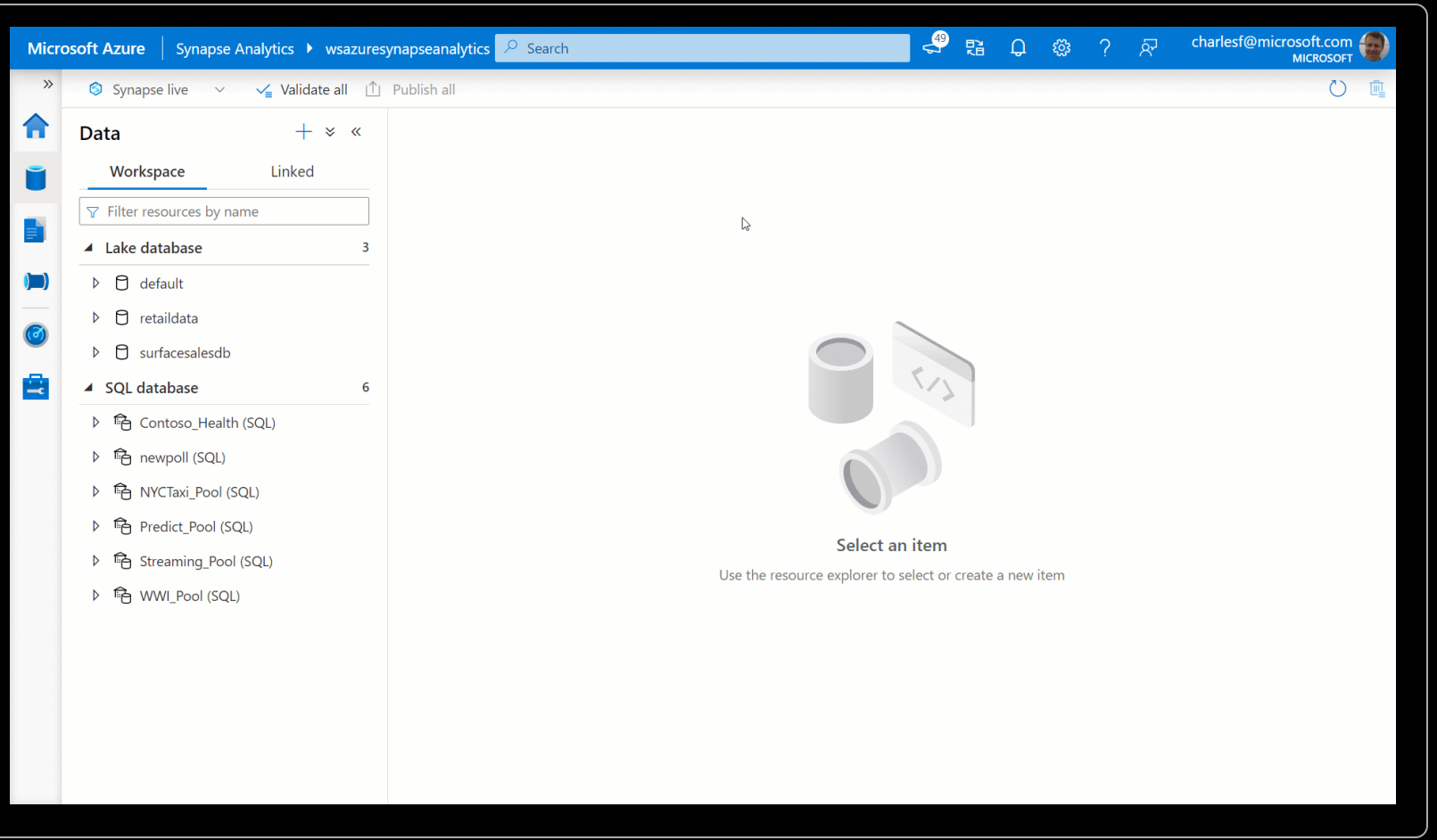
Governance


Governance

Generally Available

Integrated Catalog search in Synapse

Enables developers to search for data assets across the entire data estate to analyze with Synapse



Synapse + Power BI

- Data Integration
- Data Engineering
- Data Warehousing
- Data Science
- Observational Analytics
- Business Intelligence
- Governance

Microsoft

Partners make more possible

Microsoft Enregistrez vous dès maintenant au prochain Webinars Data AI

Event Webinar (Les jeudis de la Data & AI) - L200/300	Date	Duration (min)	Link
Azure Machine Learning pour les Data Scientists	15/09/2022	120	https://msevents.microsoft.com/event?id=2454281594
Azure Synapse	22/09/2022	120	https://msevents.microsoft.com/event?id=857781749
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	29/09/2022	120	https://msevents.microsoft.com/event?id=502366997
Déploiement et sécurisation des workspaces Azure Machine learning	06/10/2022	120	https://msevents.microsoft.com/event?id=1505714138
Azure Scale Analytics - Architectures Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share	13/10/2022	120	https://msevents.microsoft.com/event?id=139685175
MLOps avec Azure Machine Learning	20/10/2022	120	https://msevents.microsoft.com/event?id=1245885767
SQL Server 2022 et hybridation native avec Azure SQL Managed Instance	10/11/2022	120	https://msevents.microsoft.com/event?id=145826476
Machine Learning dans Azure Synapse Analytics	17/11/2022	120	https://msevents.microsoft.com/event?id=3637723312
Azure Cosmos DB et IA	24/11/2022	120	https://msevents.microsoft.com/event?id=2646013445
Azure et les Services Cognitifs	08/12/2022	120	https://msevents.microsoft.com/event?id=3772037220
La gouvernance de données dans Azure avec Microsoft Purview	15/12/2022	120	https://msevents.microsoft.com/event?id=1499560981
MLOps avec Azure Machine Learning	12/01/2023	120	https://msevents.microsoft.com/event?id=4115194515
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	19/01/2023	120	https://msevents.microsoft.com/event?id=1537241181
Déploiement et sécurisation des workspace Azure Synapse	26/01/2023	120	https://msevents.microsoft.com/event?id=1806467748
Azure Machine Learning pour les Citizen Data Scientists	09/02/2023	120	En cours
PowerBI - Self Service Analytics	16/02/2023	120	https://msevents.microsoft.com/event?id=1401519679
L'IA responsable avec Azure machine learning	09/03/2023	120	https://msevents.microsoft.com/event?id=2072953112
Machine Learning dans Azure Synapse Analytics	16/03/2023	120	https://msevents.microsoft.com/event?id=3413014857
Les bases de données Open Source dans le cloud Azure	23/03/2023	120	https://msevents.microsoft.com/event?id=2727487131
Hybridation des services de Machine Learning Azure	06/04/2023	120	https://msevents.microsoft.com/event?id=1624914222
La gouvernance de données dans Azure avec Microsoft Purview	13/04/2023	120	https://msevents.microsoft.com/event?id=3909342839
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	04/05/2023	120	https://msevents.microsoft.com/event?id=1162207895
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	16/05/2023	120	https://msevents.microsoft.com/event?id=3517068442
Hybridation des services de données Azure	24/05/2023	120	https://msevents.microsoft.com/event?id=2996507398
Self Service Analytics	01/06/2023	120	En cours
Total	**25**		