Azure Synapse Analytics

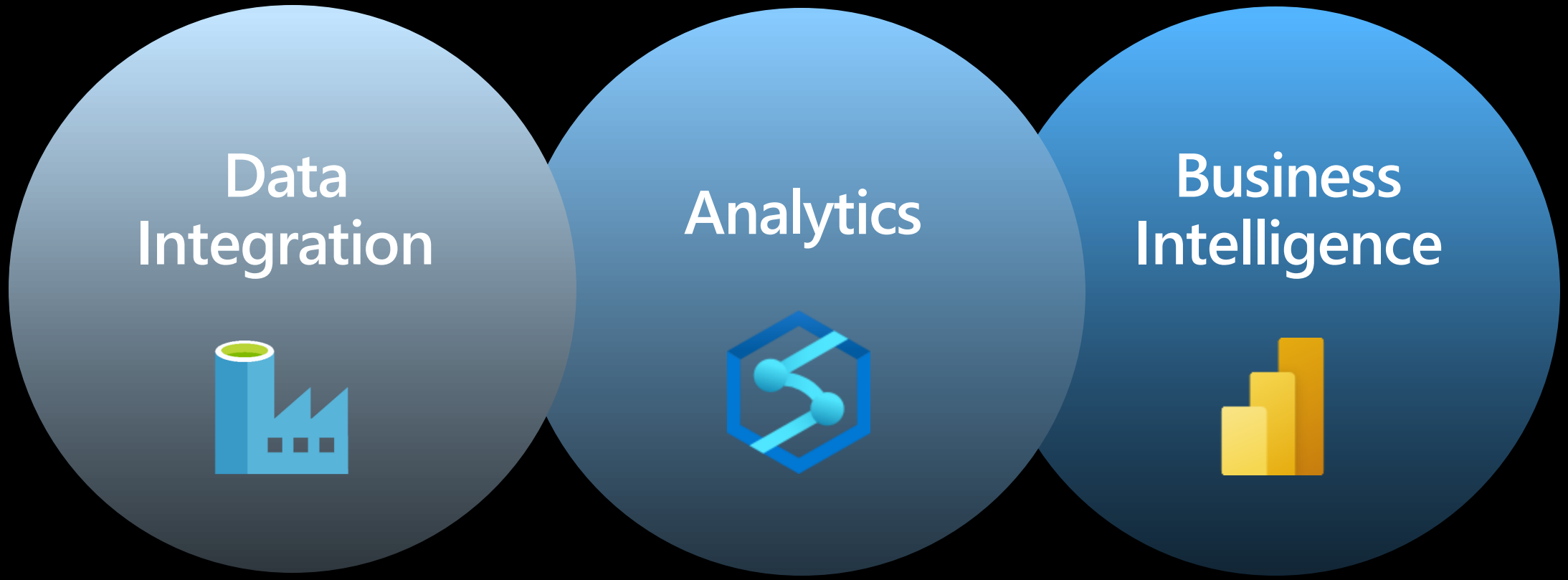
December 2021

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



The only vendor with leadership in all MQ's

Data Integration

Analytics

Business Intelligence

Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose. This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from Microsoft.

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.


Synapse + Power BI
Data Integration
Data Engineering
Data Warehouse
Data Science
Observation Analytics
Business Intelligence
Governance

Data Integration

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

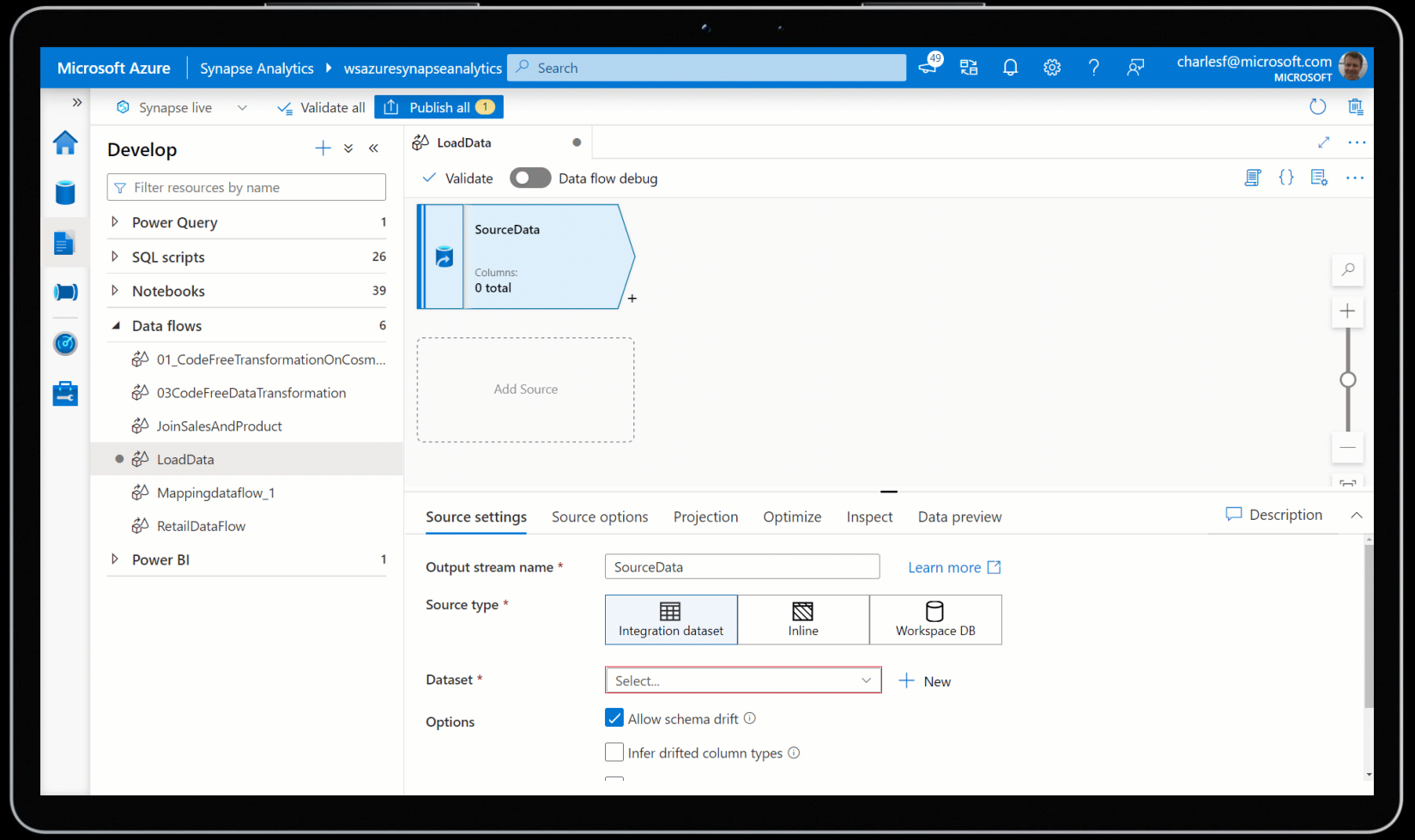
Secure hybrid connectivity

Code-free development environment

Generally Available

100+ Connectors

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

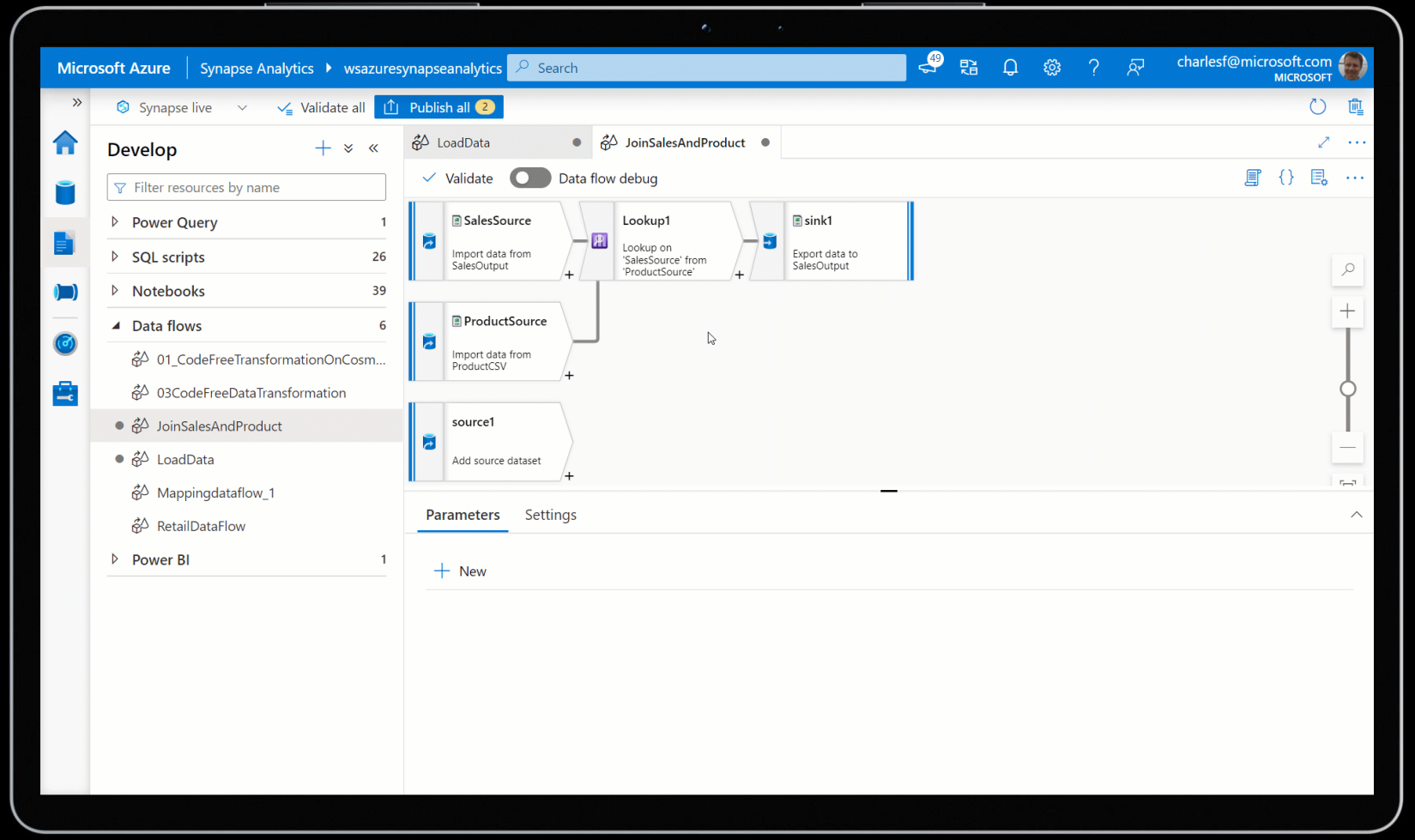


Generally Available

Code-free Data Flows

Enables developers to rapidly integrate data from a variety of sources

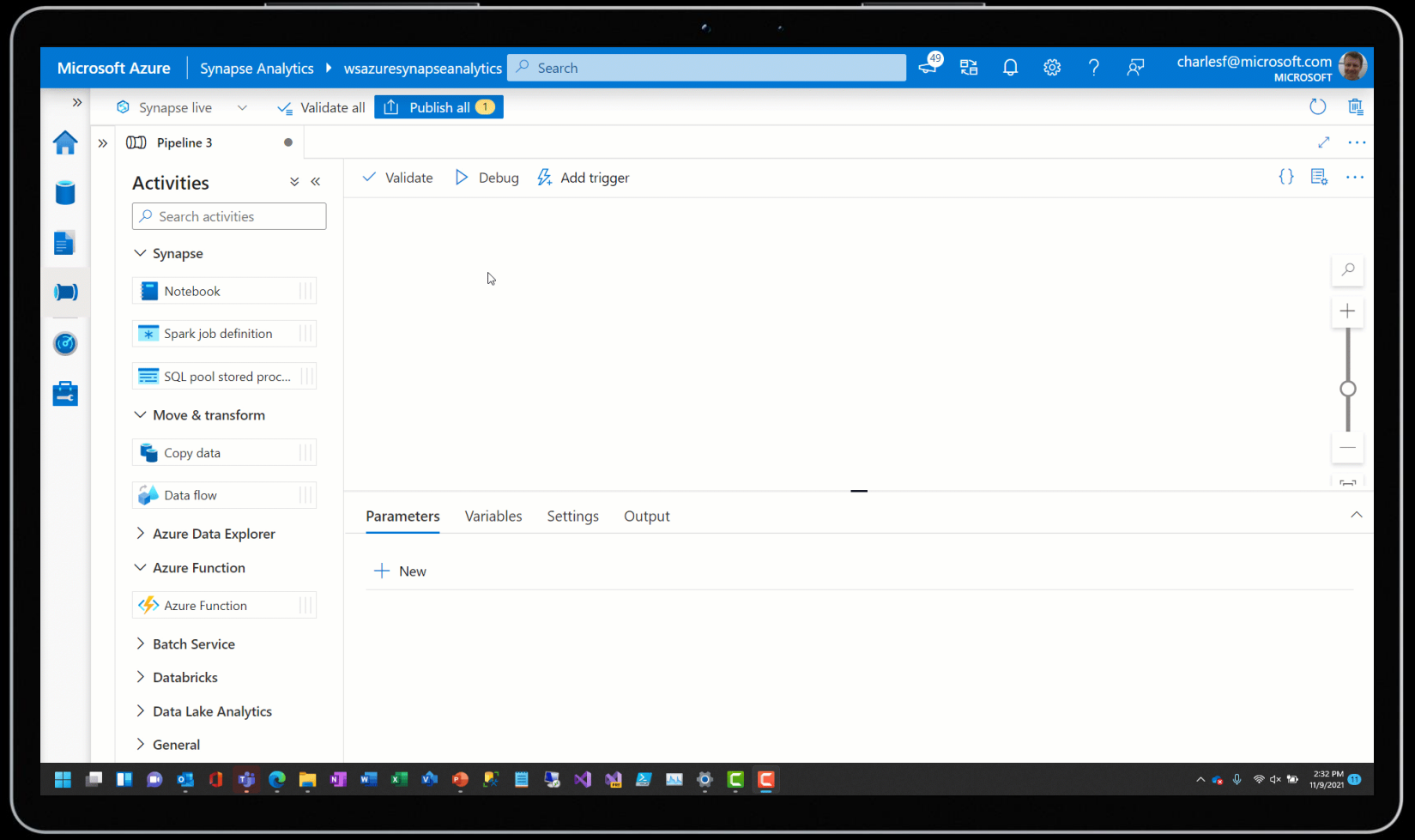
Execute on Spark for large scale processing



Generally Available

Pipeline Orchestration

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

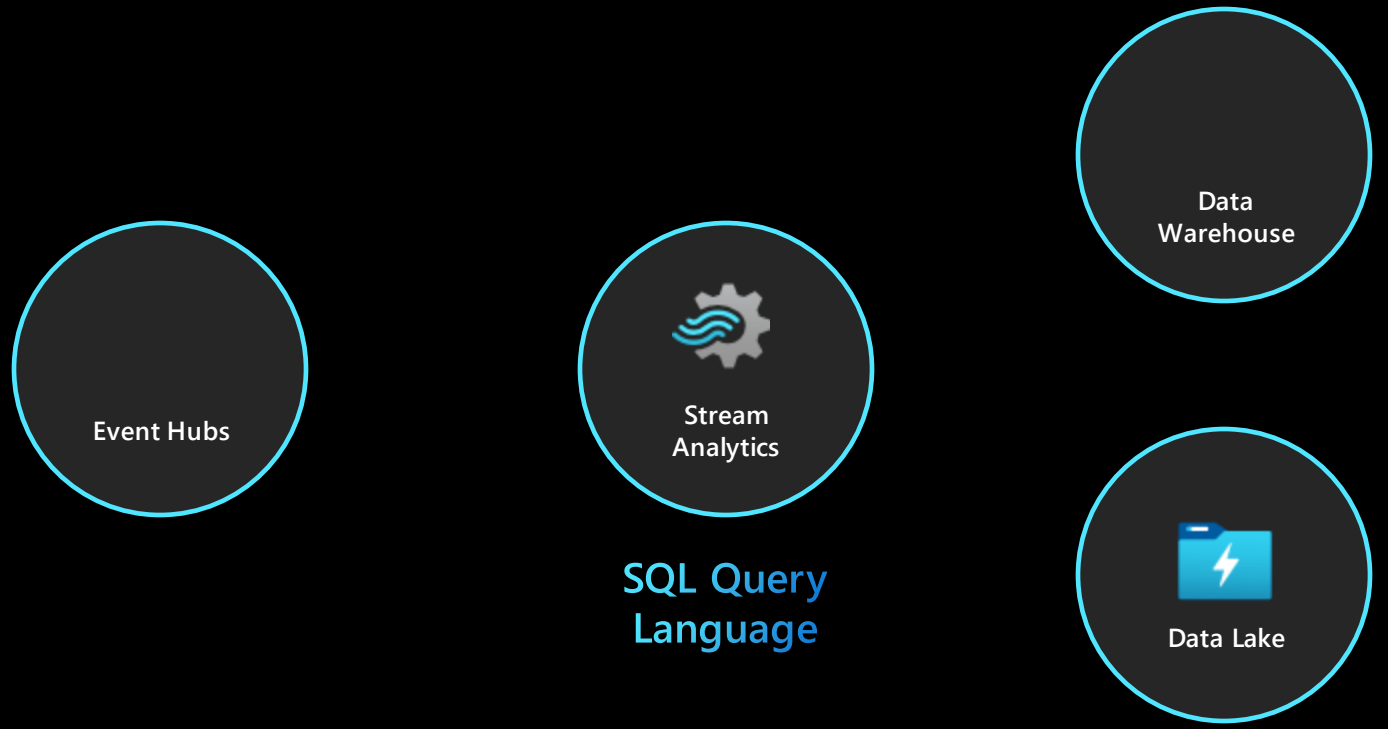


Generally Available

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



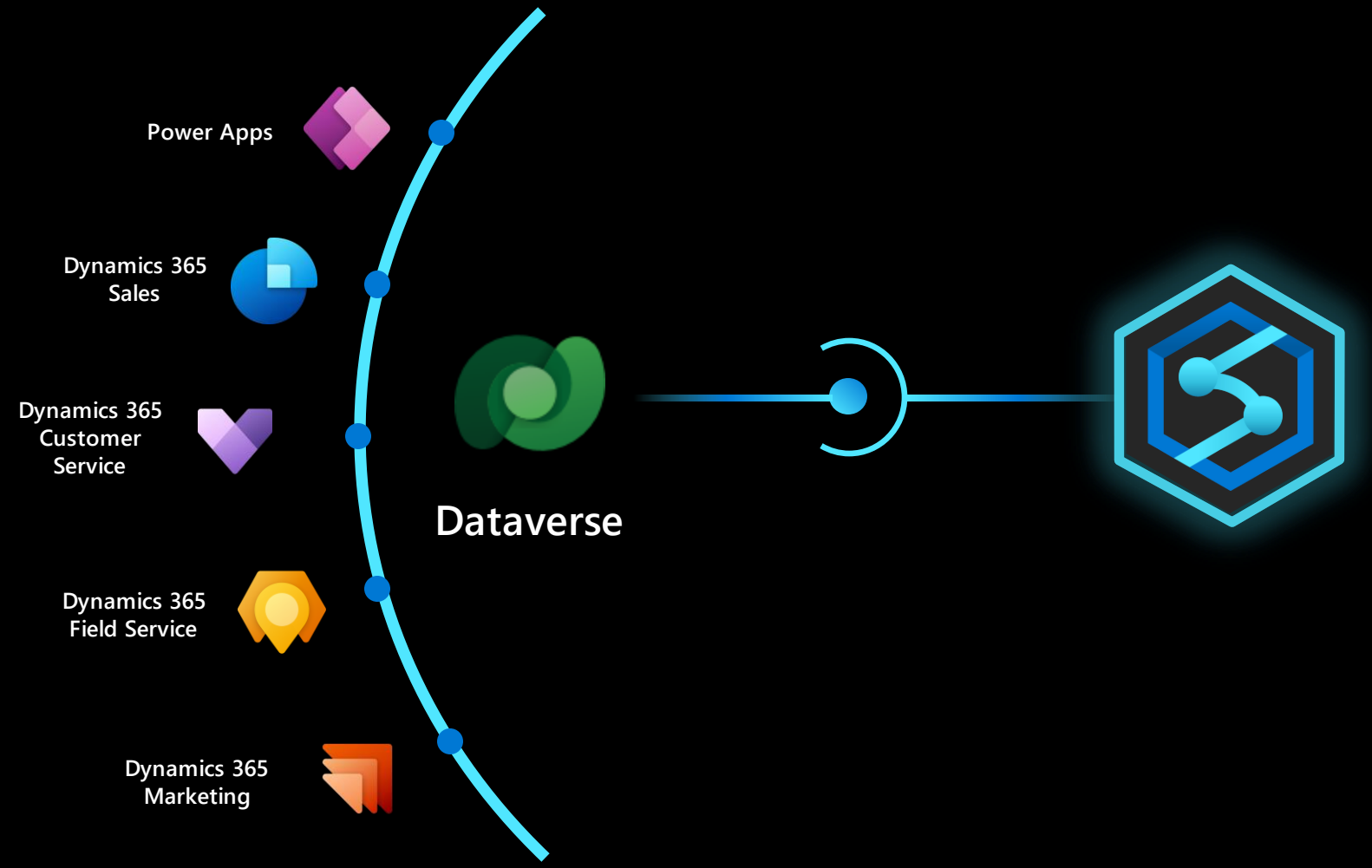
Generally Available

November 2021

Synapse Link for Dataverse

One-click integration of D365 data into Synapse for analytics

No data pipelines required



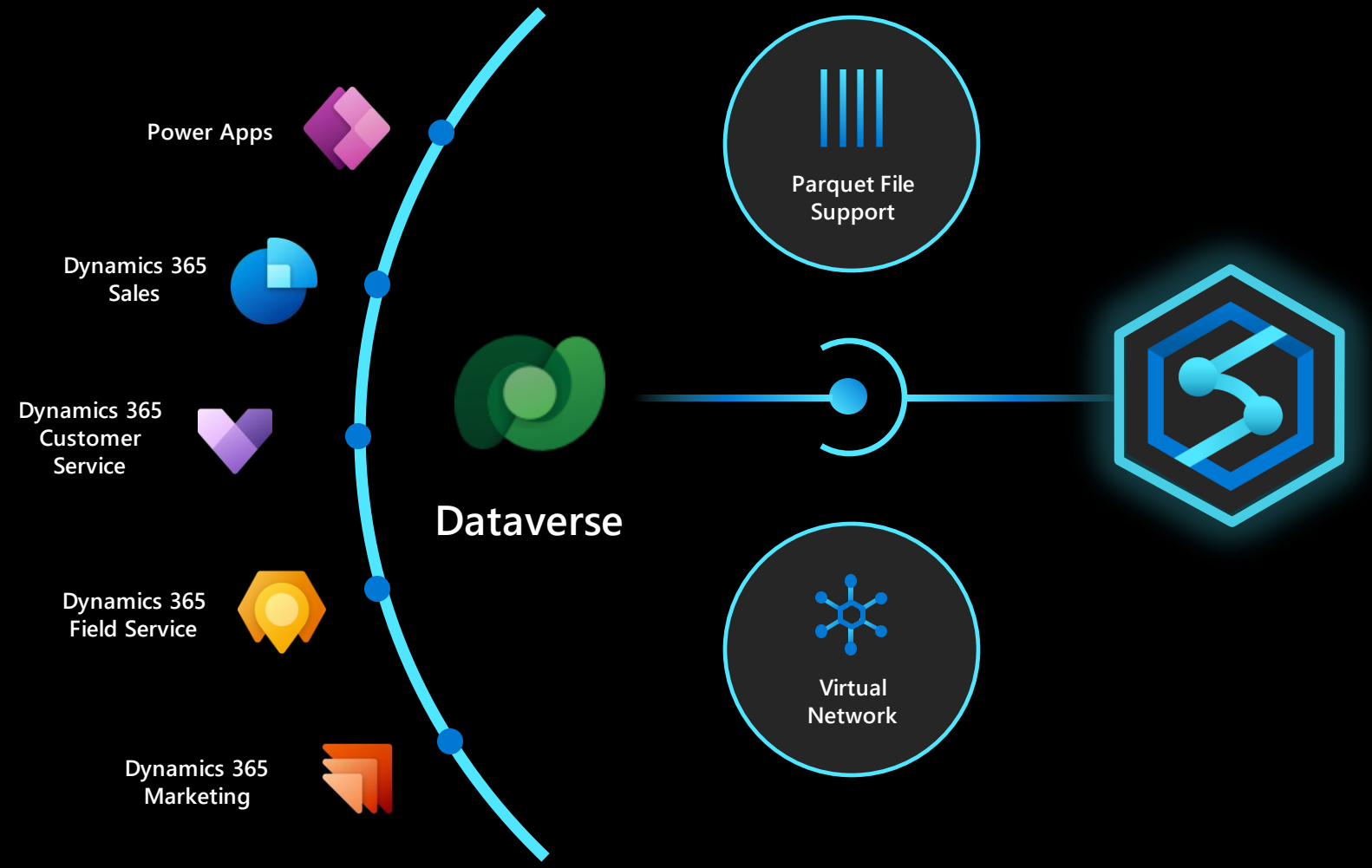
Public Preview

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)

Synapse Link for Microsoft SQL

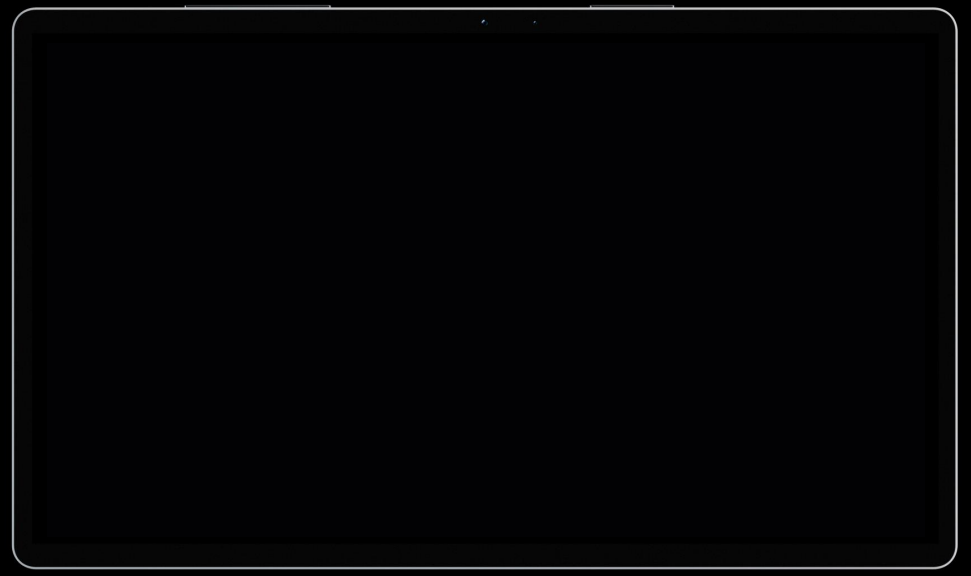
Near real-time operational analytics in Synapse

No data pipelines required

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

SQL Server 2022
Public Preview Q2 2022

Azure SQL Database
Public Preview Soon



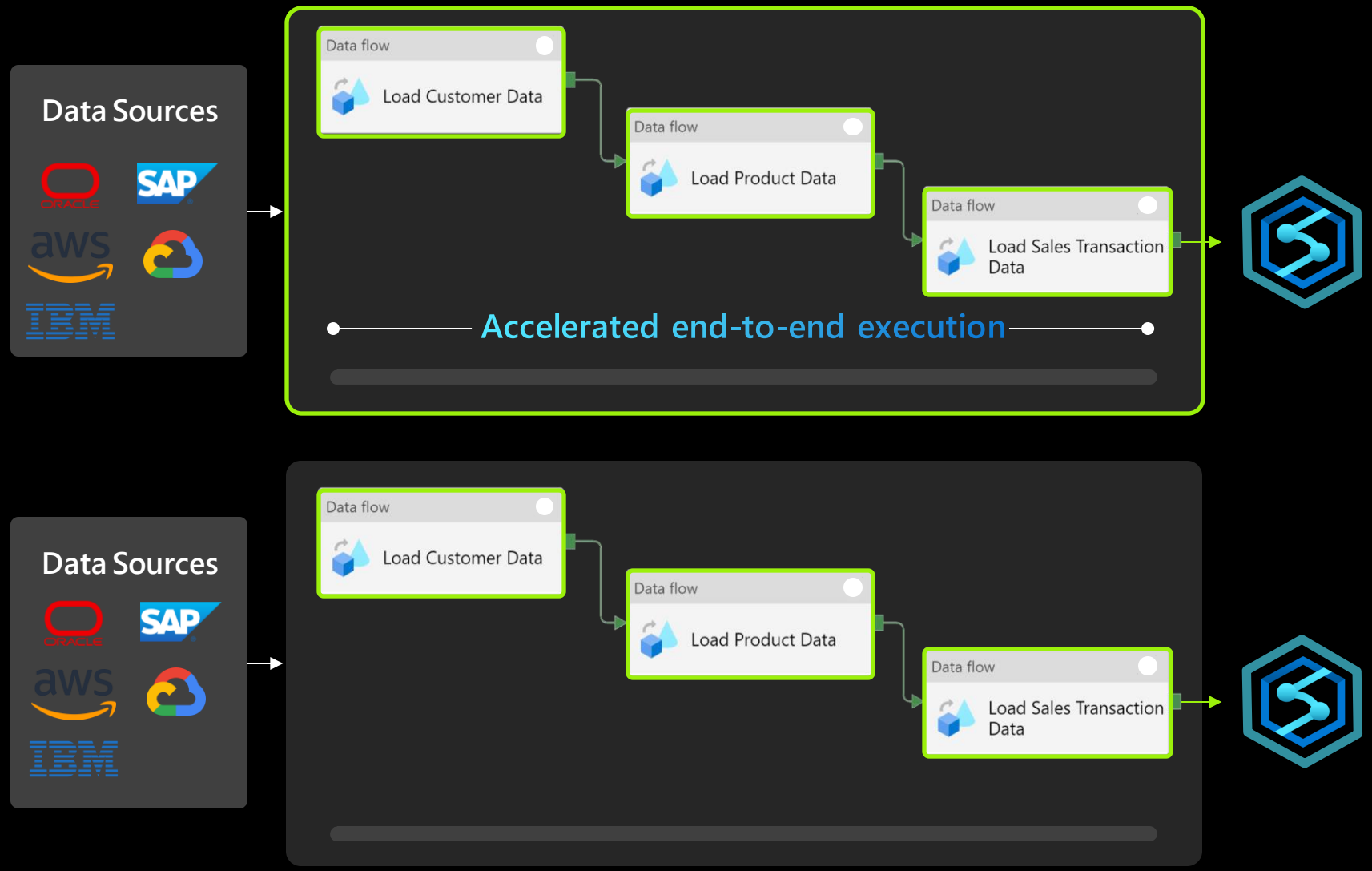
Generally Available

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 Engineering

Data Engineering

Scalable Spark engine

Industry standard languages

Delta Lake enabled

Azure DevOps integrated

Public Preview

Q1 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

RocksDB integration for managing state
Maintain streaming state for improved restartability

Spark Streaming Enhancements
Enables streaming jobs to run for virtually infinite duration

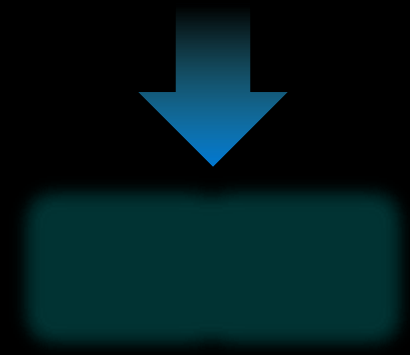
Public Preview

Q1 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

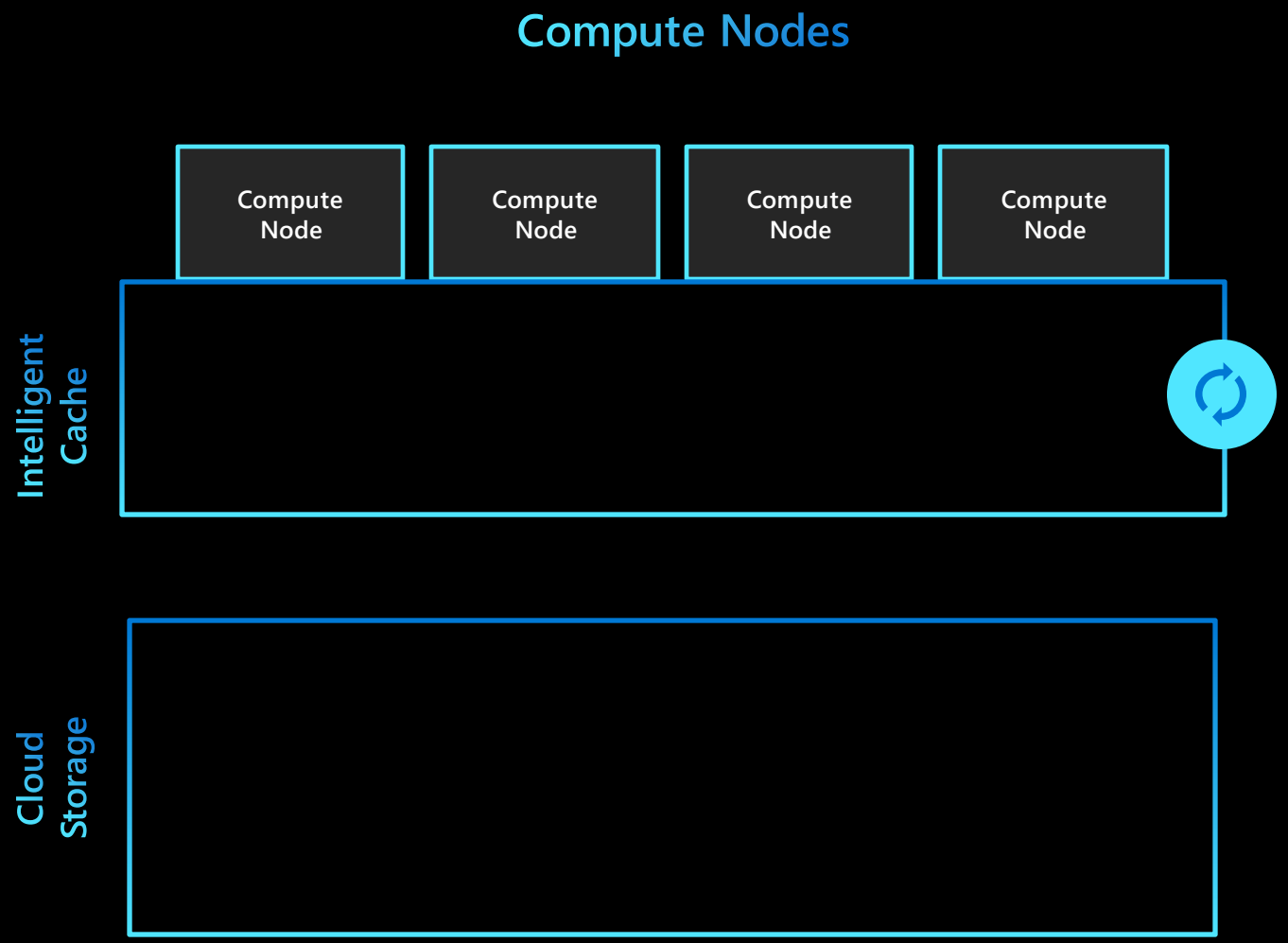
Improves filter query performance by ordering data for fast lookups on large datasets

Public Preview

Q1 2022

Spark Caching Enhancements

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



Data Warehousing

Data Warehousing

Cornerstone of enterprise analytics for decades

Industry standard SQL language

Structured and semi-structured data

Broad ecosystem of applications

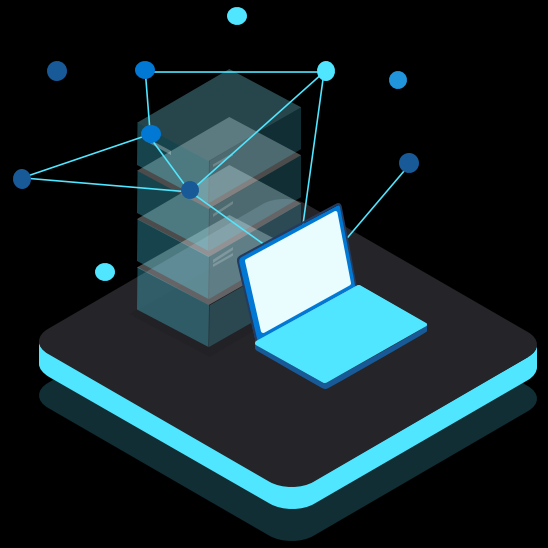
Fine-grained data security

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

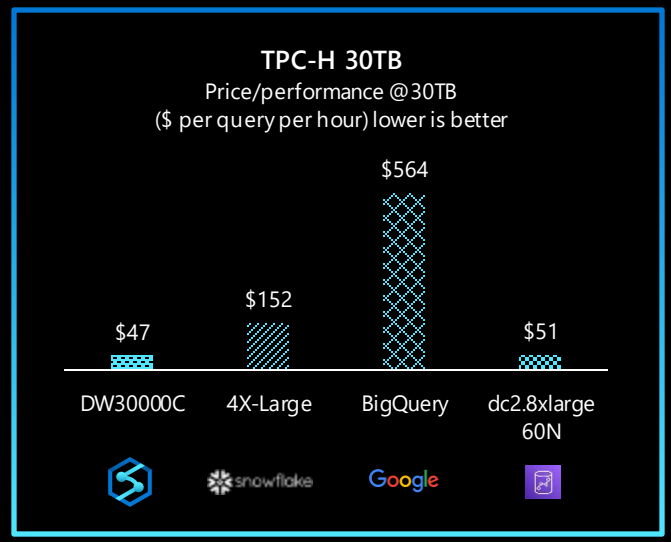
Dedicated

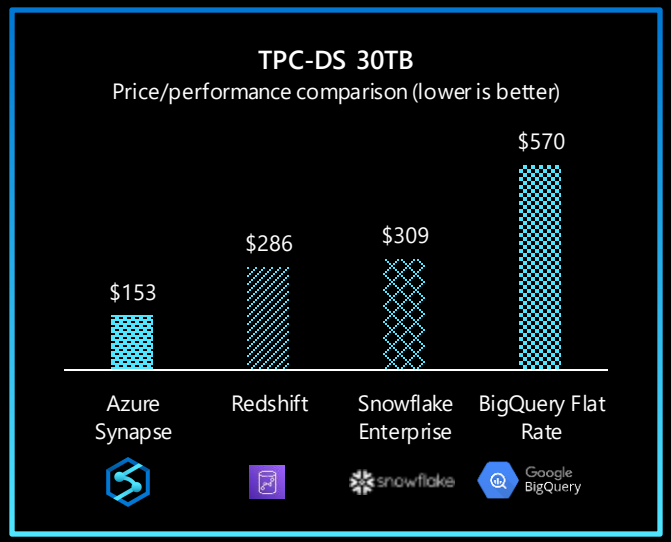
Generally Available

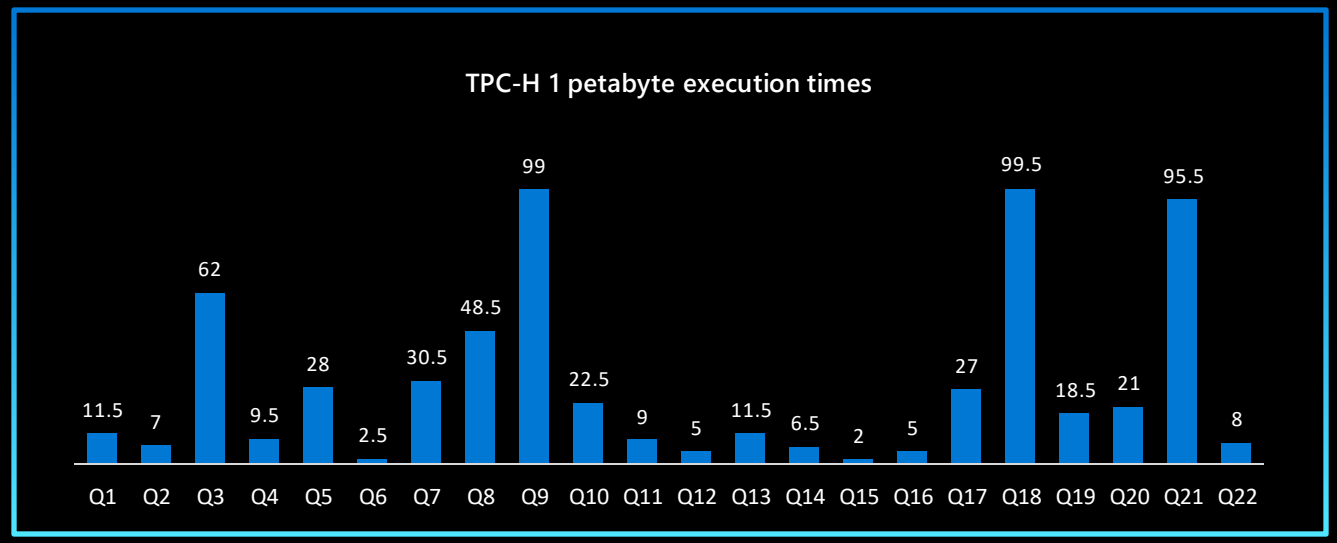
Industry Leading Performance

Price/performance leader for data warehousing

The only platform to complete TPC-H at 1PB







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

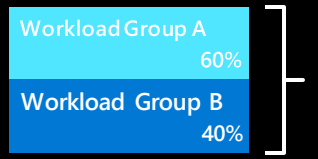
Scale-In

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

Workload Importance

Data Warehouse

Workload Isolation

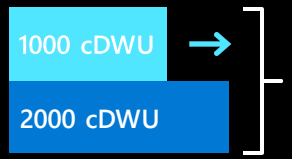


Data Warehouse

Scale-Out

- Add compute for variable workloads
- Pause compute when idle

Elastic Cluster (Scale Up)



Data Warehouse

Generally Available

Complete Data Protection

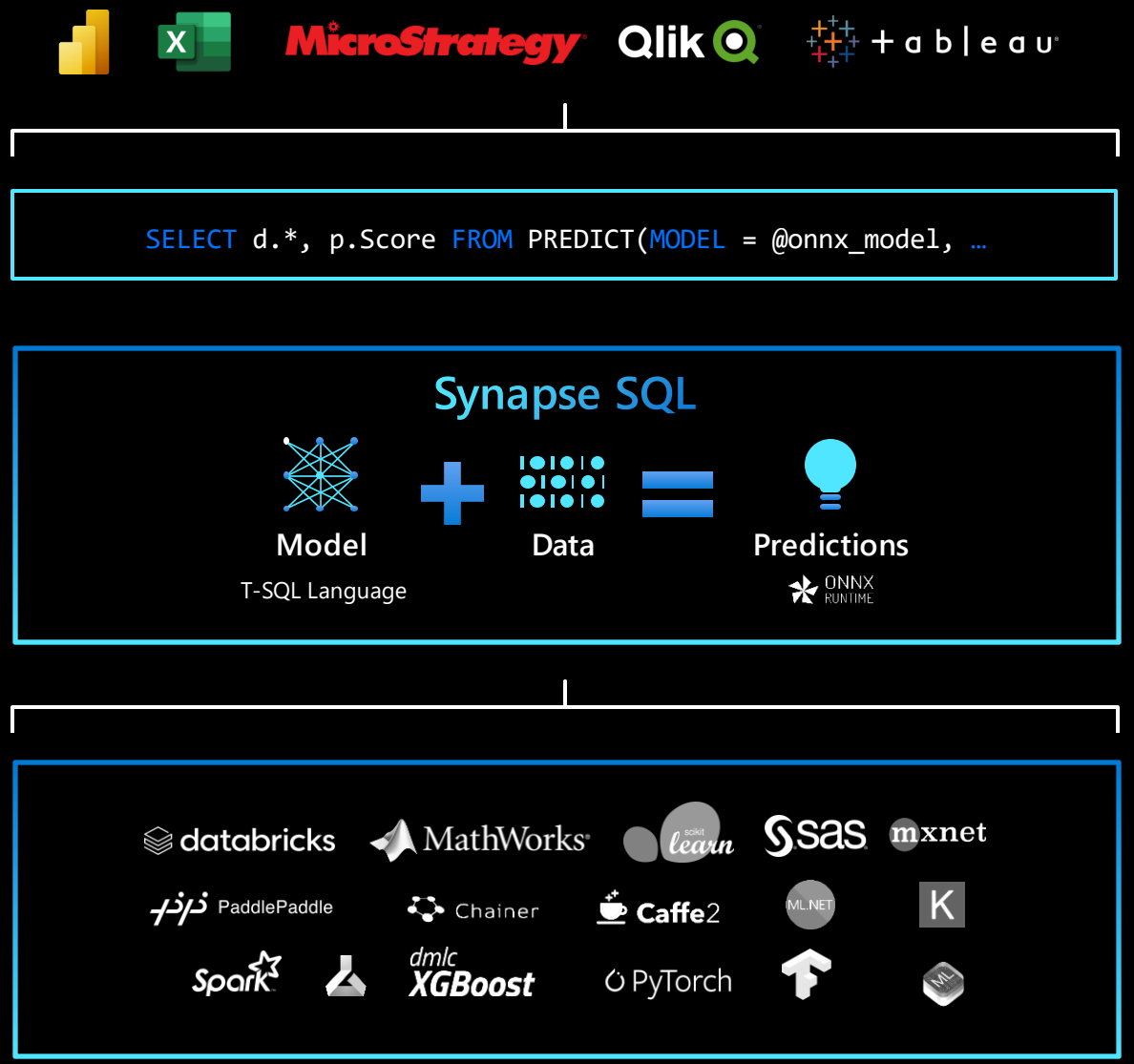
Democratize data compliantly with fine-grained access controls and multi-level encryption

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



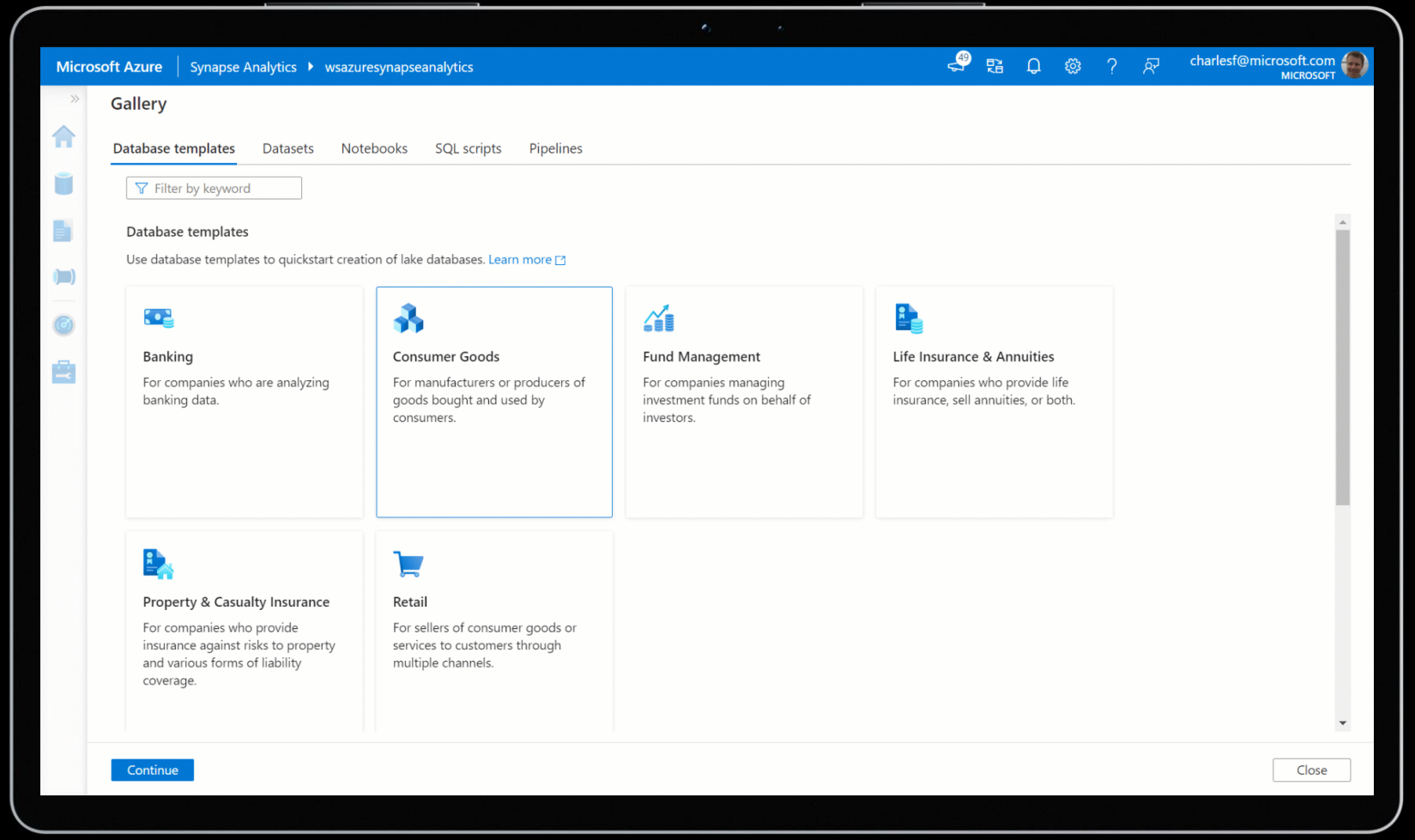
Public Preview

November 2021

Database Templates

Built-in database templates

Low code database designer

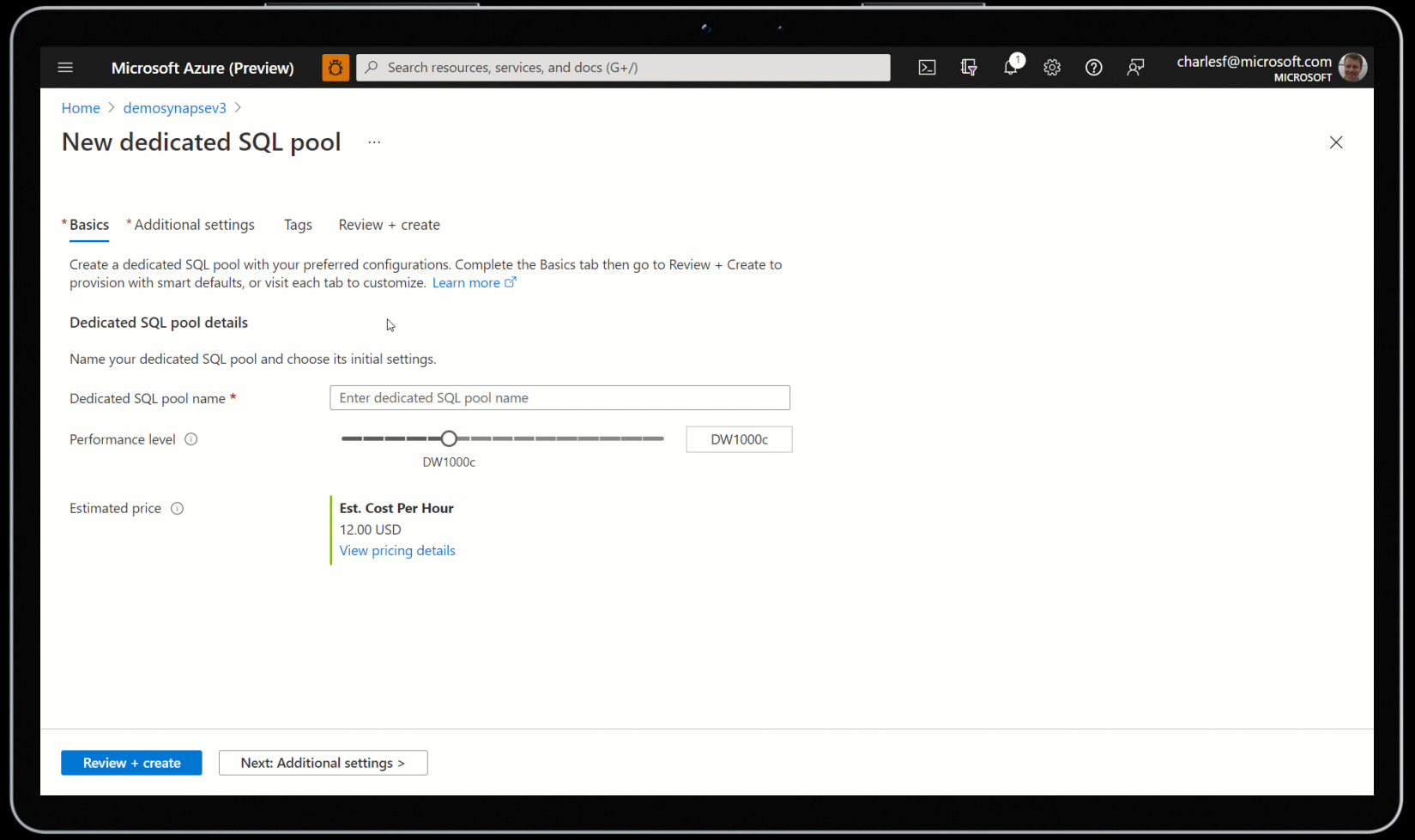


Public Preview

Q2 2022

Data Warehouse Restore SLA

Guaranteed SLA for database restore reduces disaster recovery downtime



Public Preview

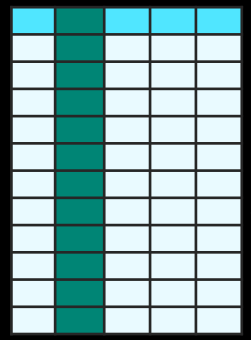
Q2 2022

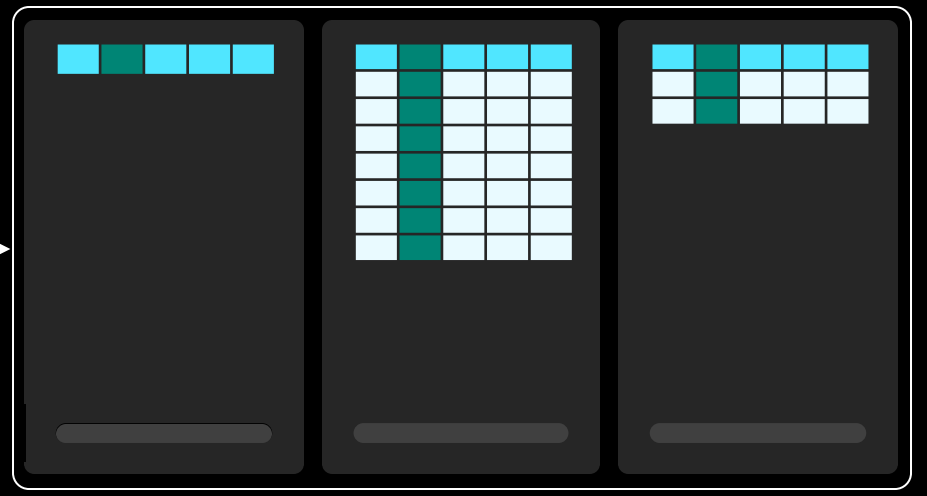
Multi-column distributed table sharding

Improved query performance and easier migrations

Uneven Distribution (Skew)

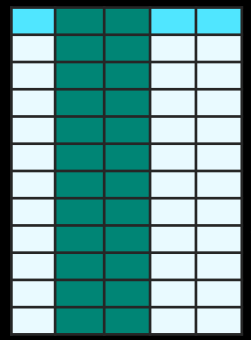
```
CREATE TABLE SalesTransactions (
WITH DISTRIBUTION =
(HASH(ProductKey))
```





Balanced Distribution

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



Balanced Distribution: Queries execute faster

Public Preview

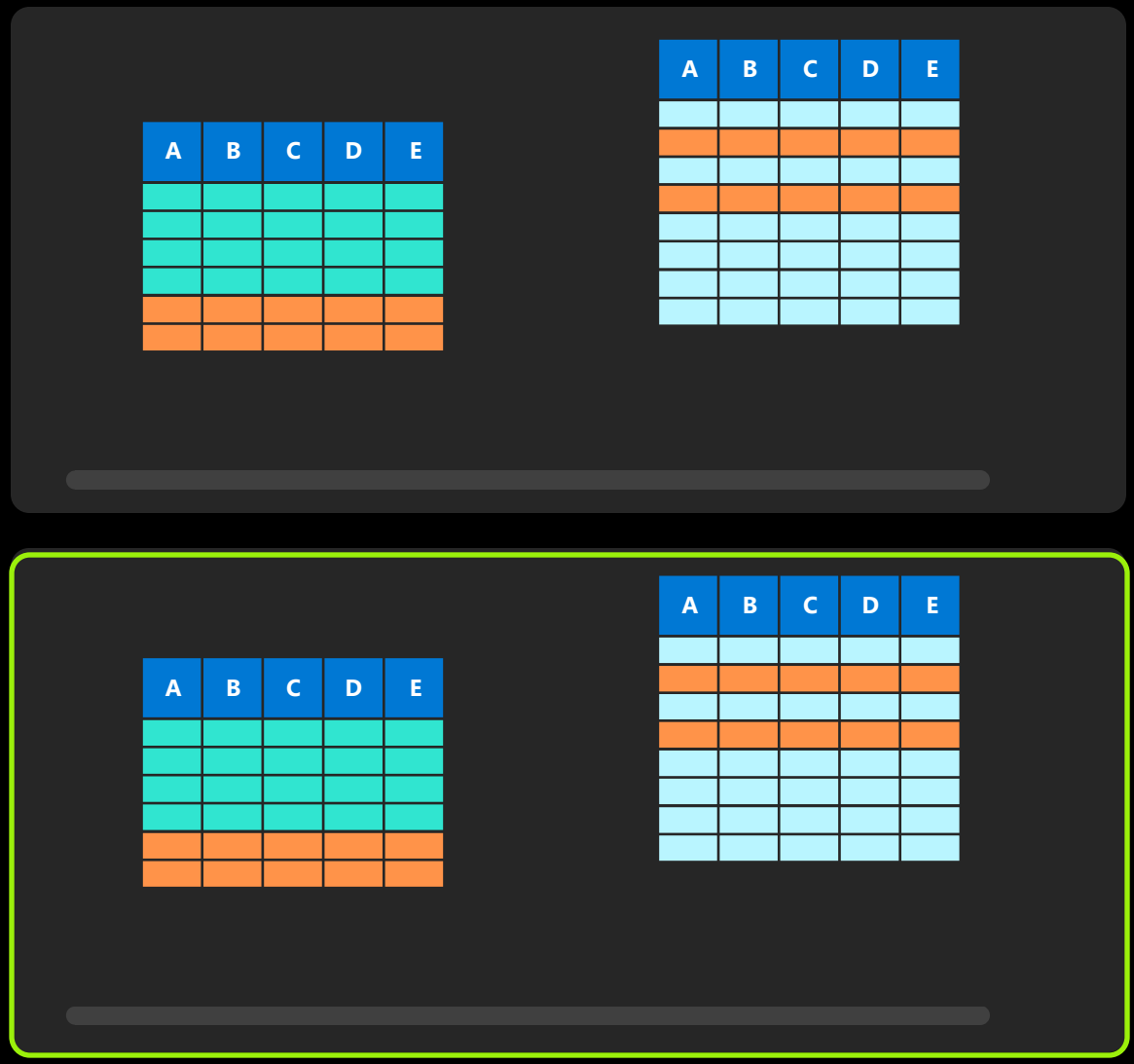
Q2 2022

MERGE SQL Statement

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

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

```
MERGE ...
```



Generally Available

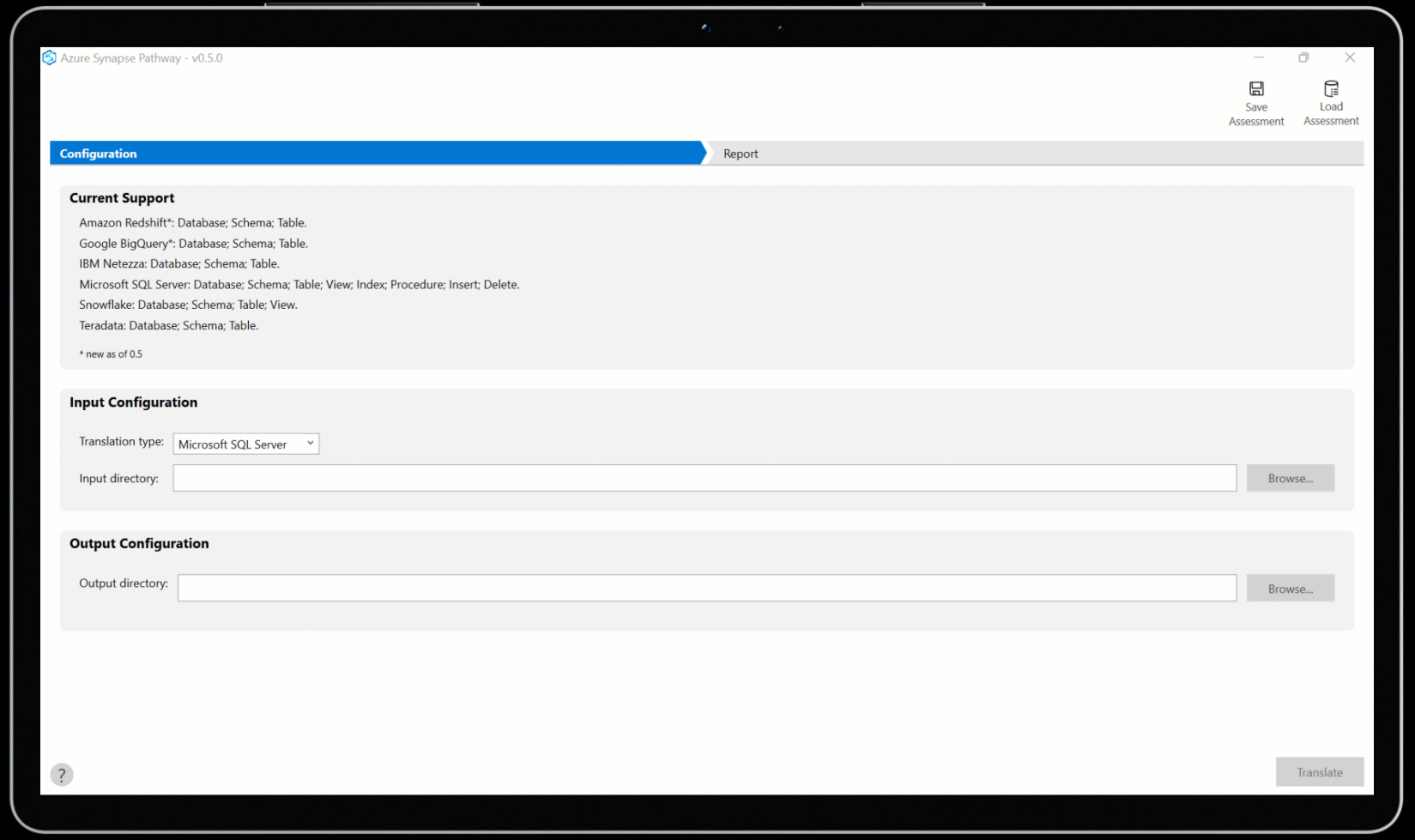
Updated Frequently

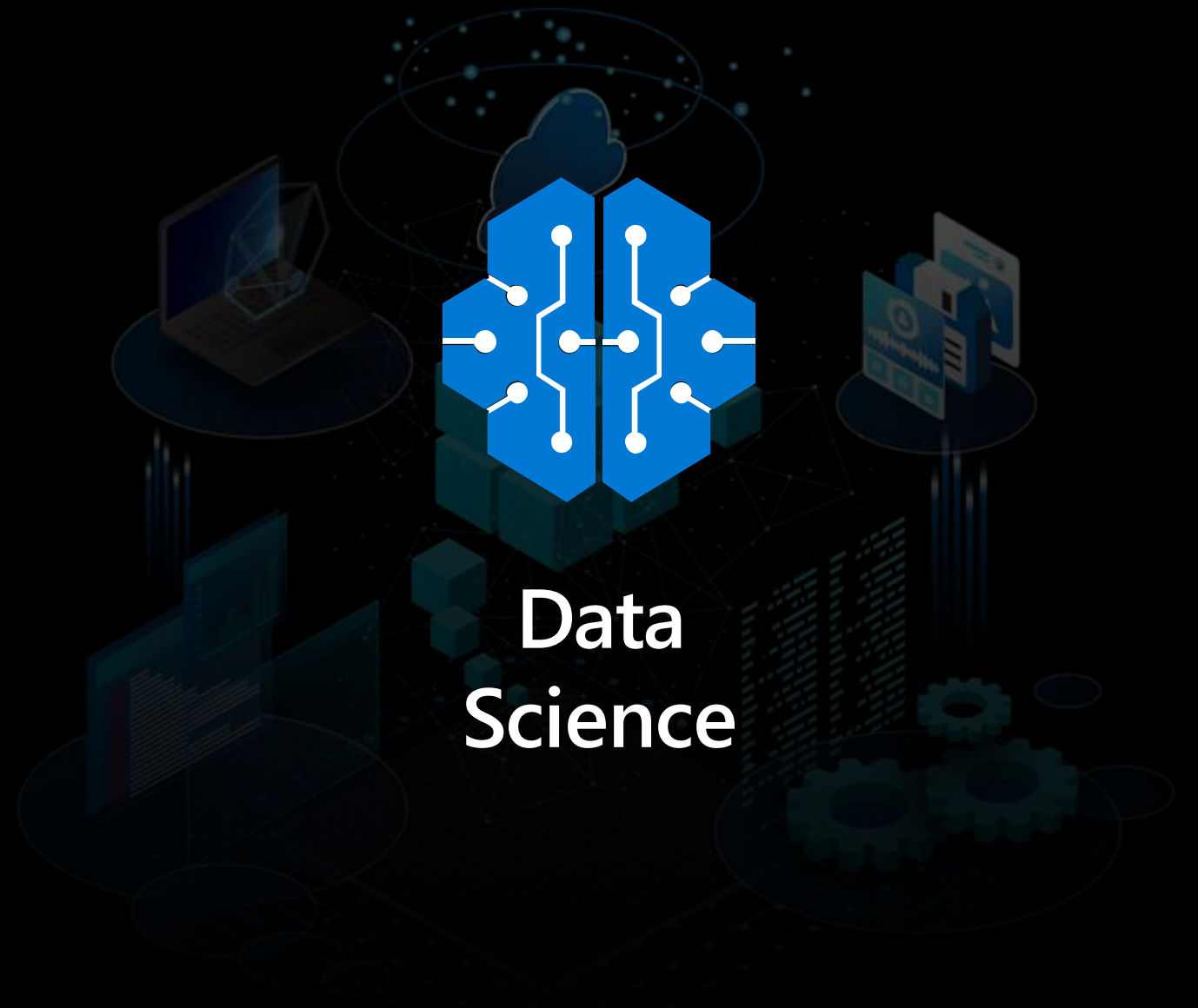
Free SQL Migration Utility

Automate SQL conversion from existing systems

Supported Sources

- Teradata
- Netezza
- SQL Server
- Redshift
- Snowflake
- BigQuery




Data
Science

Data Science

Industry standard languages such as PySpark

Code-first and Code-free Auto ML

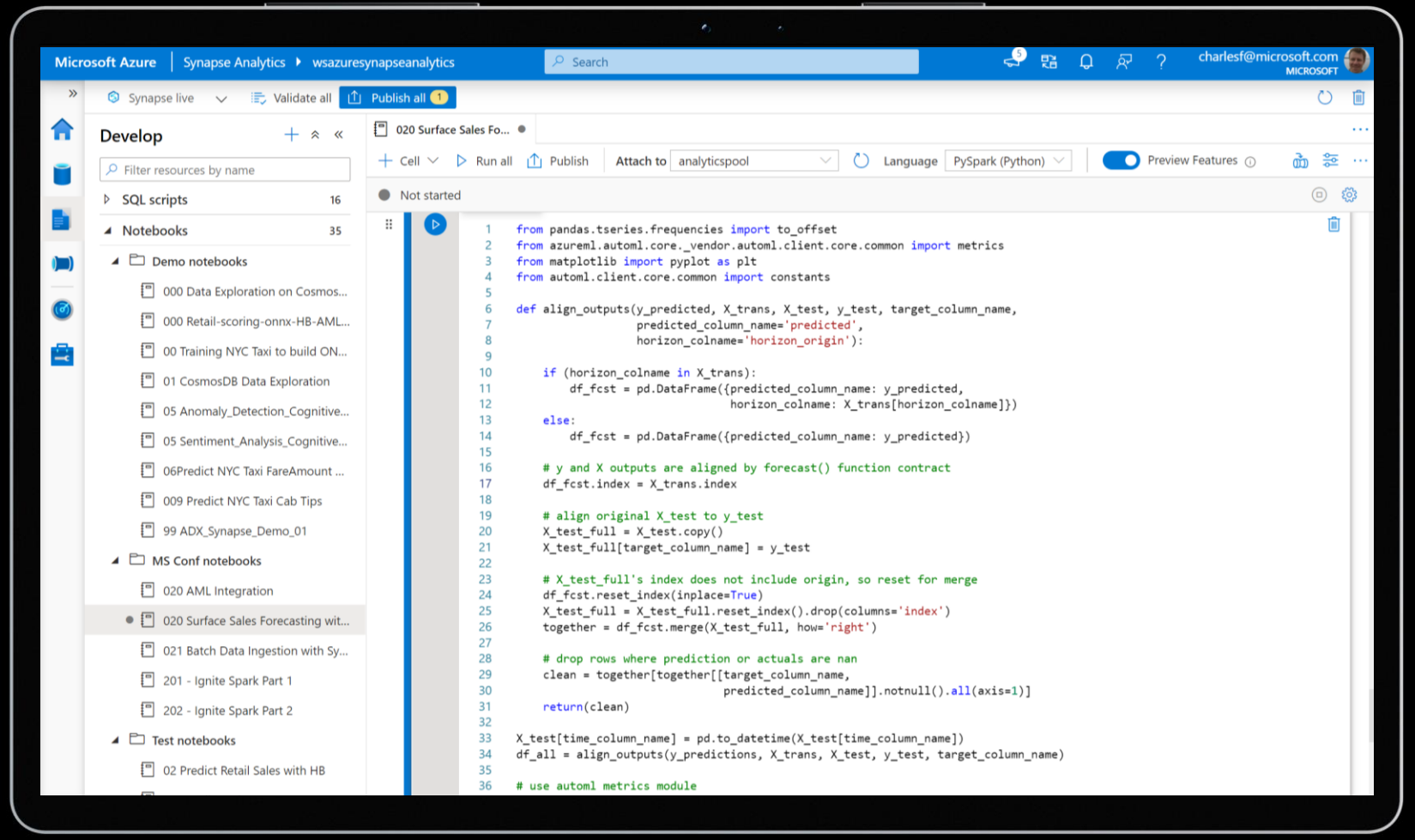
Integrated ML Model Repository

Hardware accelerated GPU model training

Generally Available

Notebook Development Experience

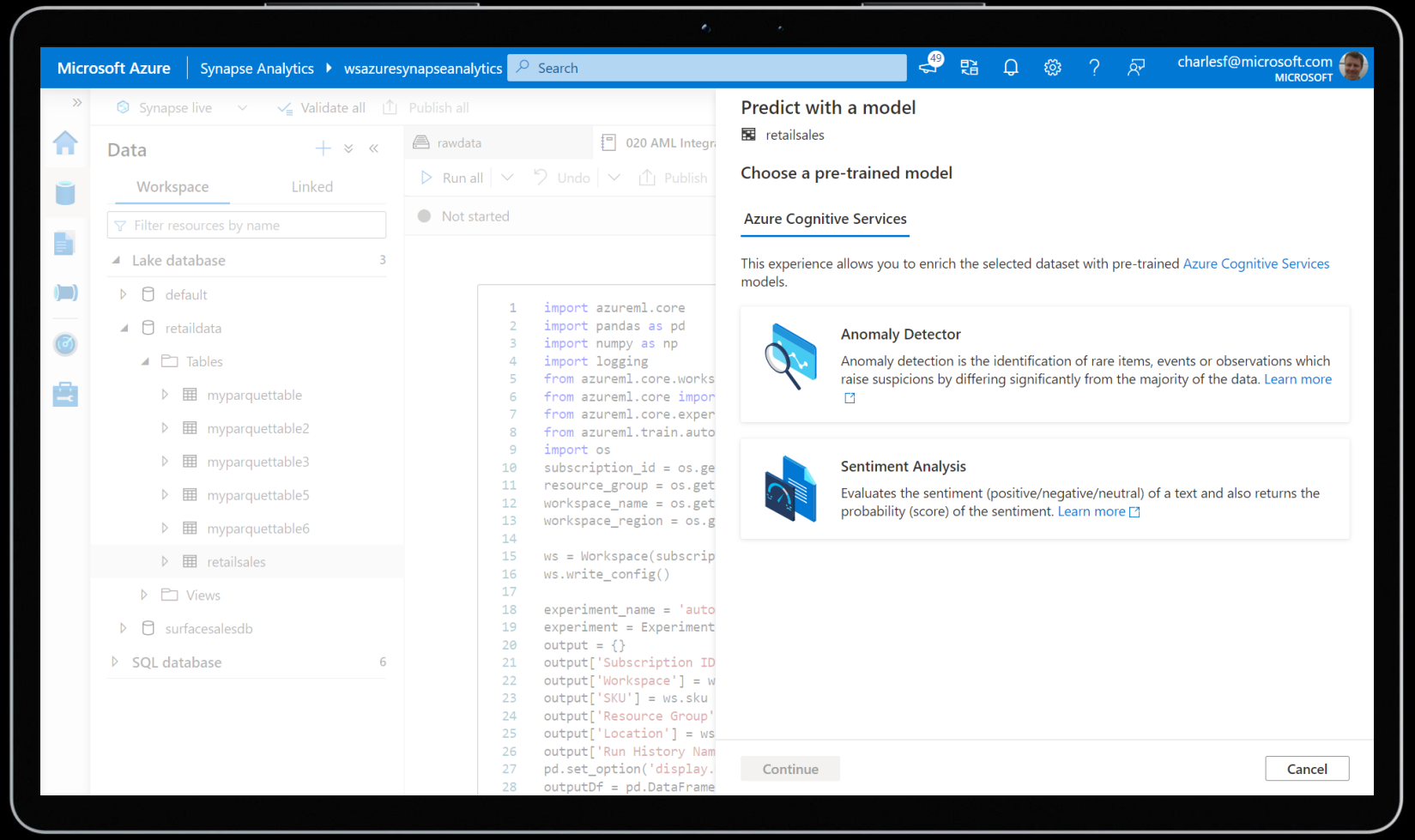
Empower data scientists will a familiar Notebook based development interface



Generally Available

Built-in Cognitive Services

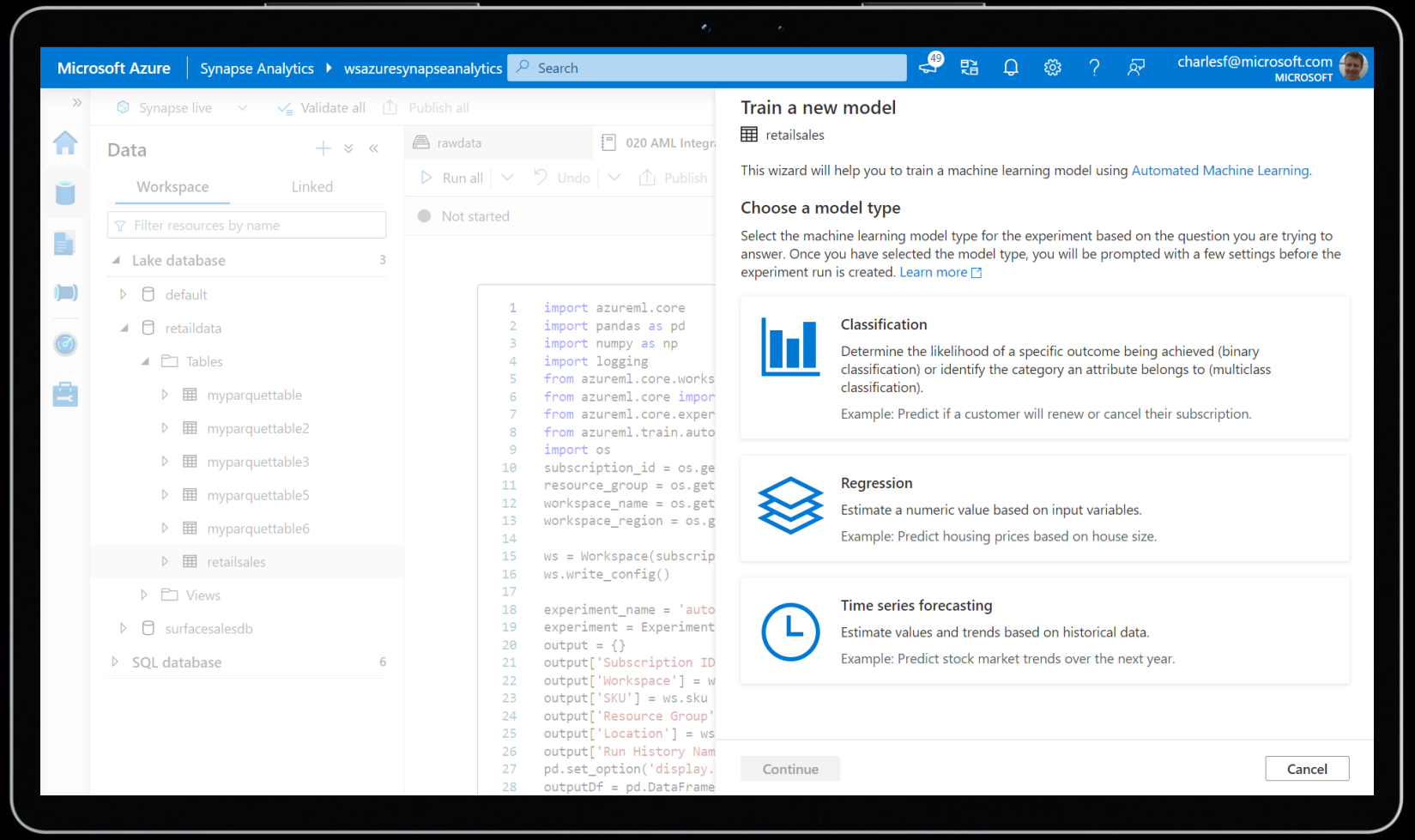
Enables simple integration of pre-built machine learning models



Generally Available

Automatic Machine Learning

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

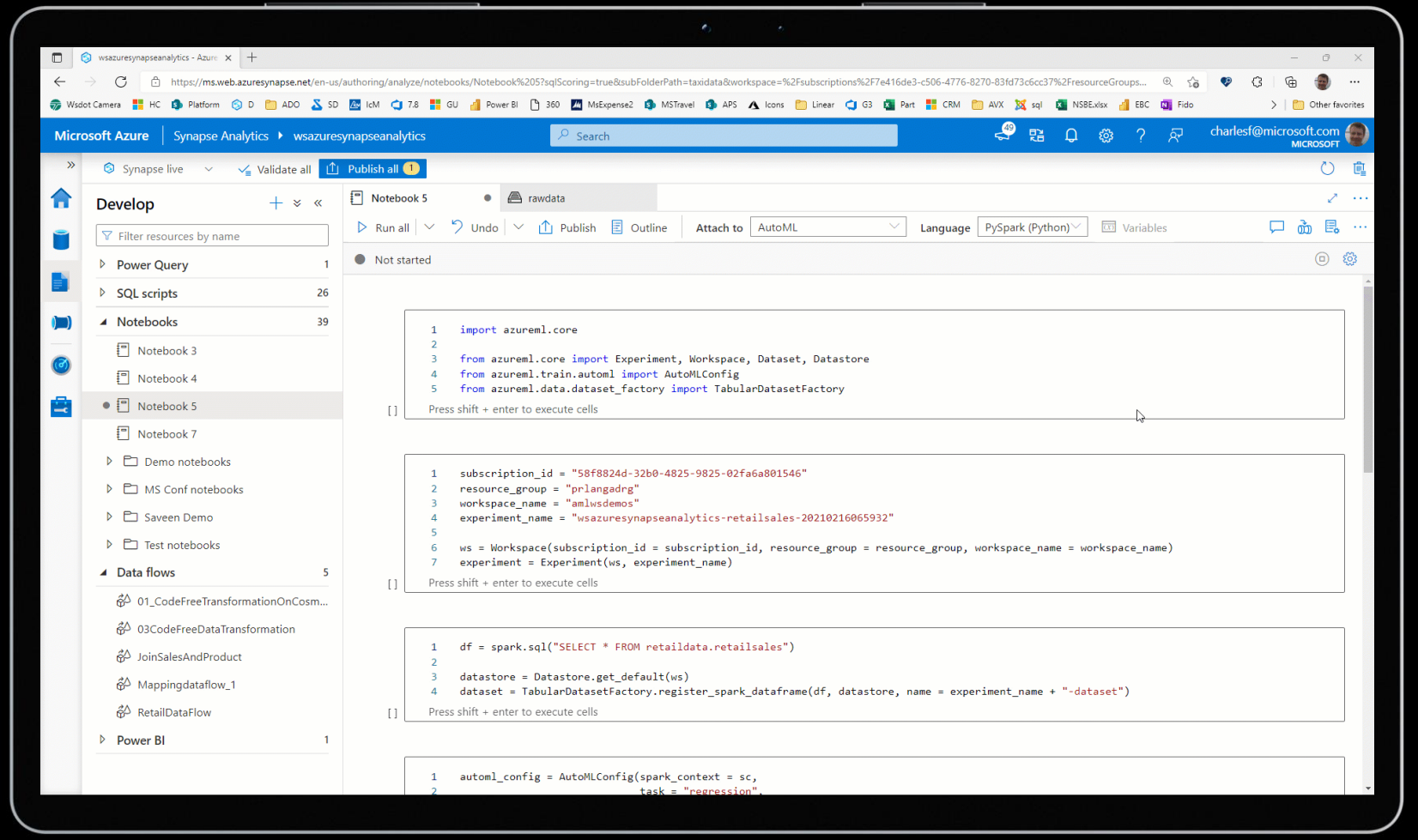


Generally Available

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

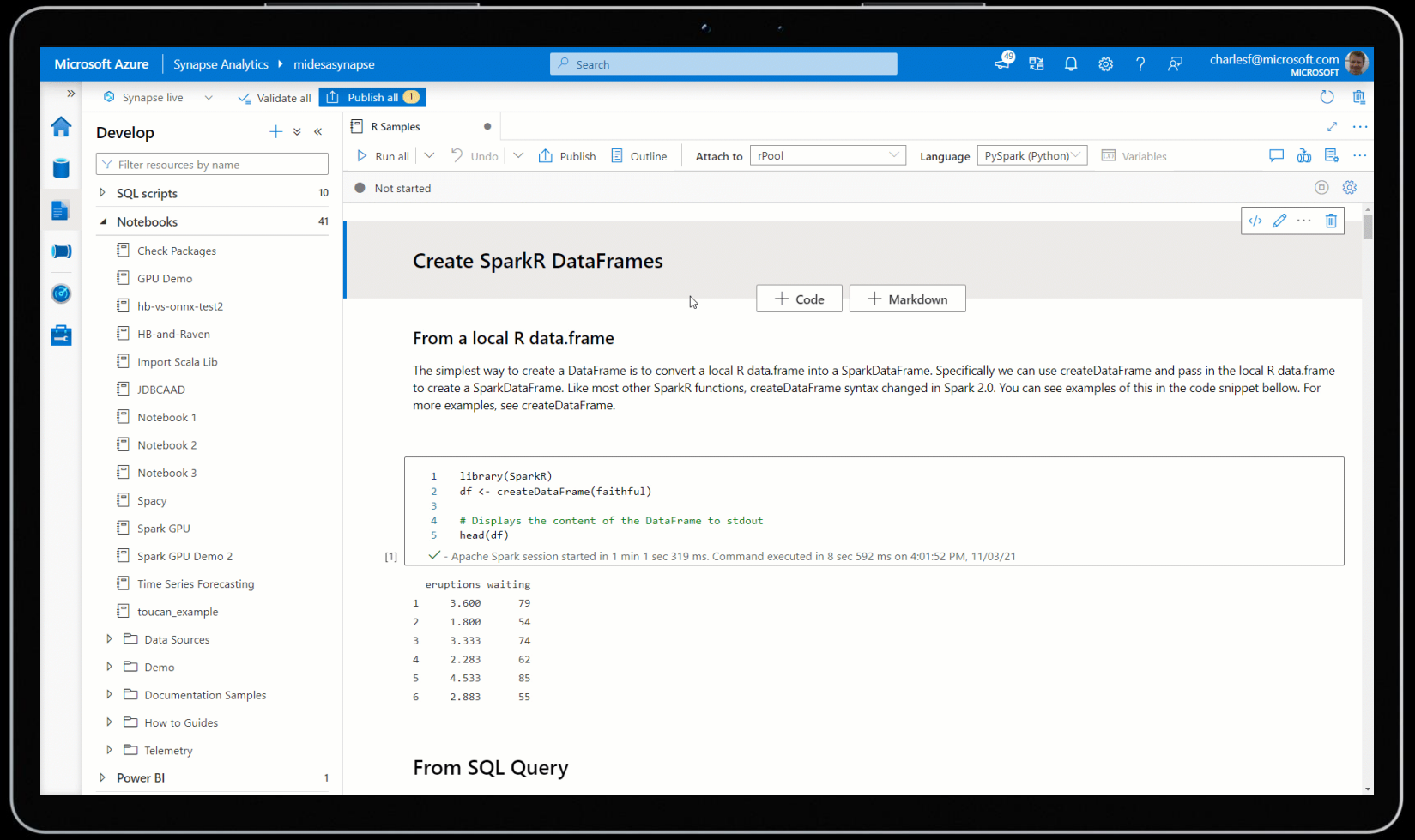


Public Preview

Q2 2022

R Language Support

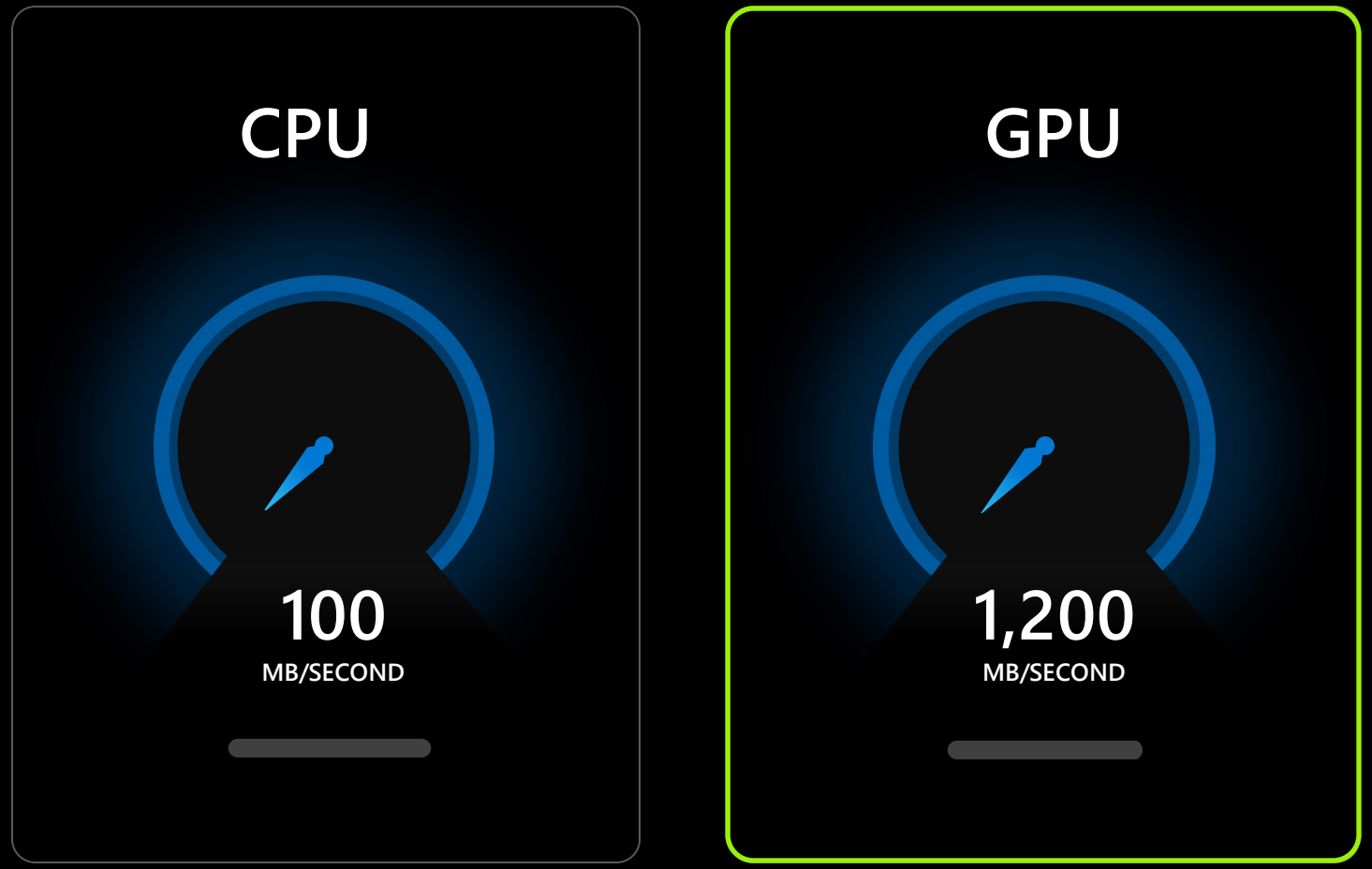
Enables data scientists to apply the industry standard R language to developing ML models



Public Preview
November 2021

GPU Accelerated Workloads

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

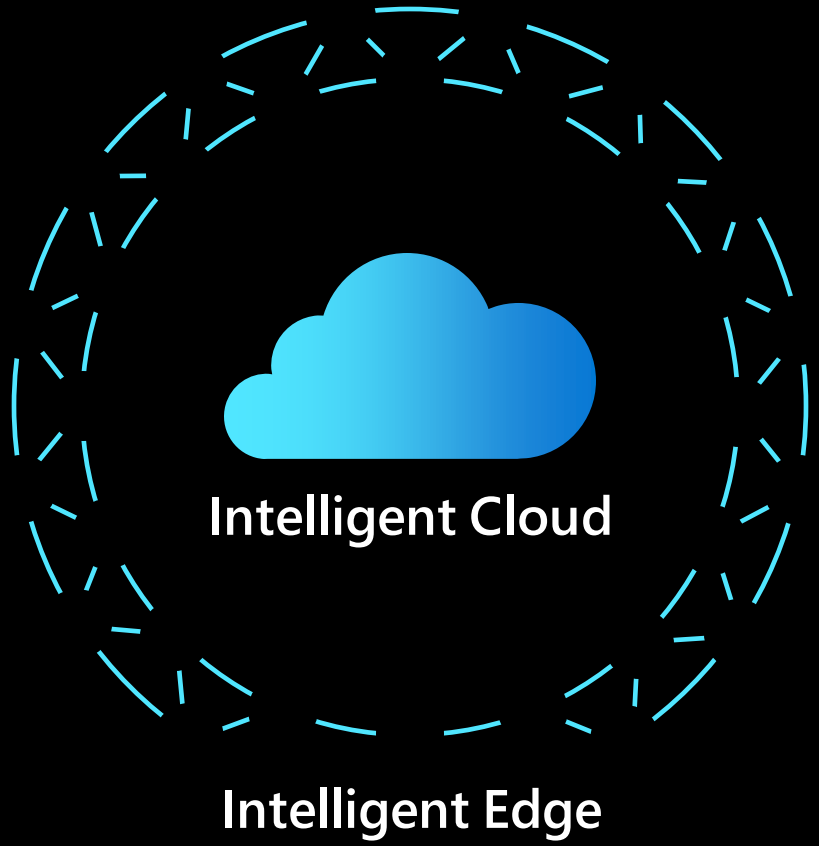


Observational Analytics

Observational Data

The fastest growing data segment

50 BN
connected devices
by 2030



175 ZB
total amount of
data by 2025

Observational Data

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

Why is it challenging to analyze?

Looking for unpredictable phenomena

Constantly changing schema

Near real time visibility required

Analytics systems costs are often prohibitive

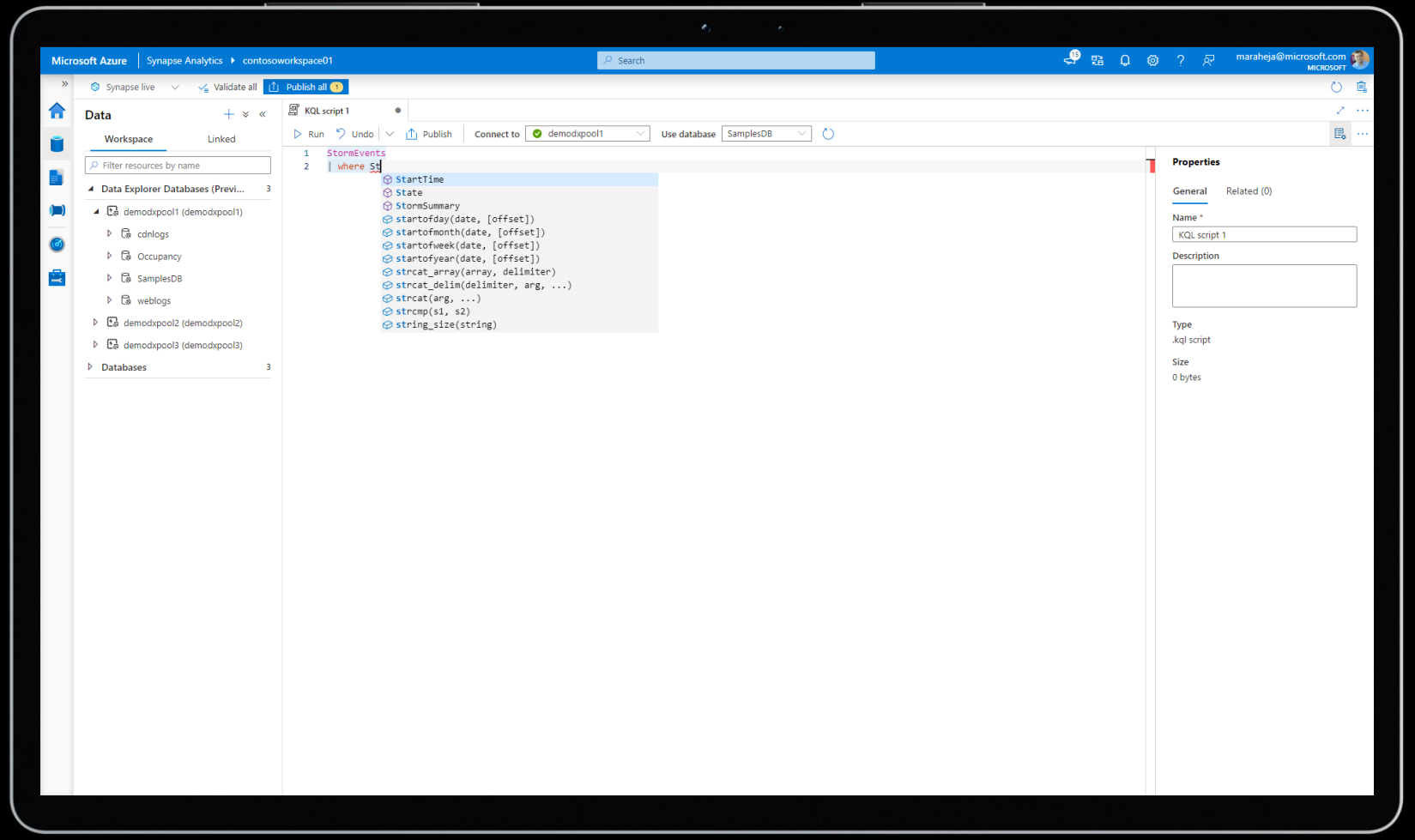
Frequently changing business questions

Near real-time analytics on Observational data at petabyte scale

Public Preview

Synapse Data Explorer Engine

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

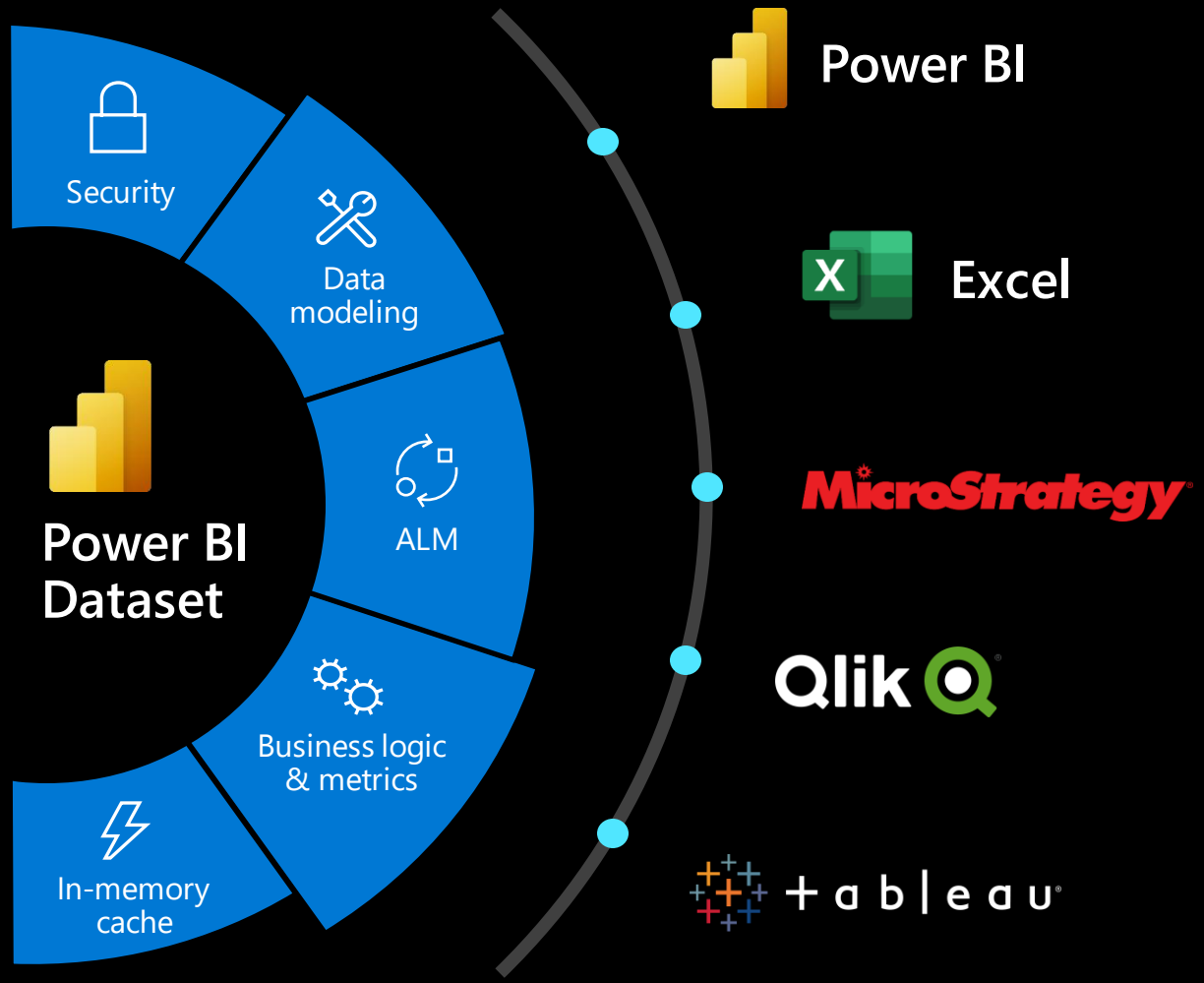


Business Intelligence

Generally Available

Worlds leading OLAP engine

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



Generally Available

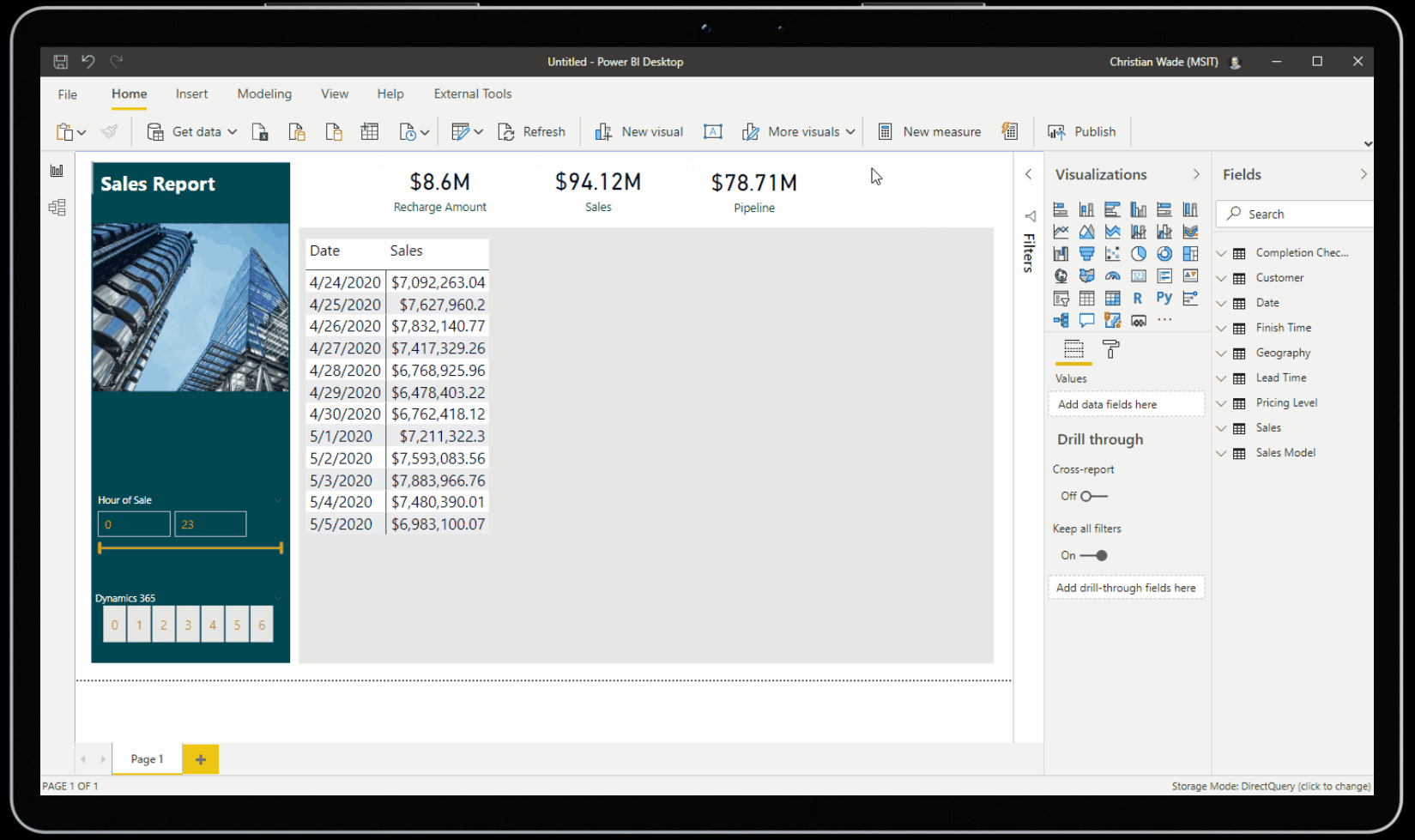
Combine enterprise and local datasets

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

Public Preview

Composite models

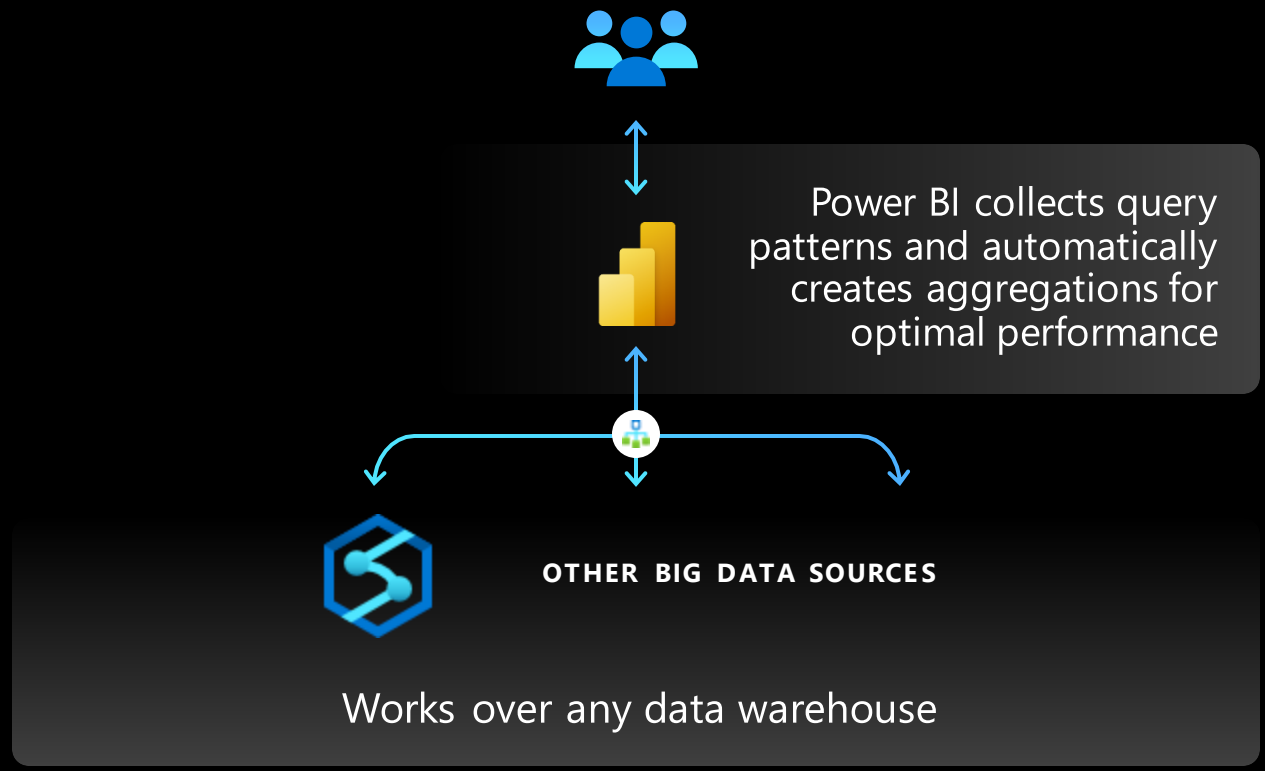
Seamlessly combine and extend self-service BI with corporate BI models



Public Preview

Automatic Aggregations

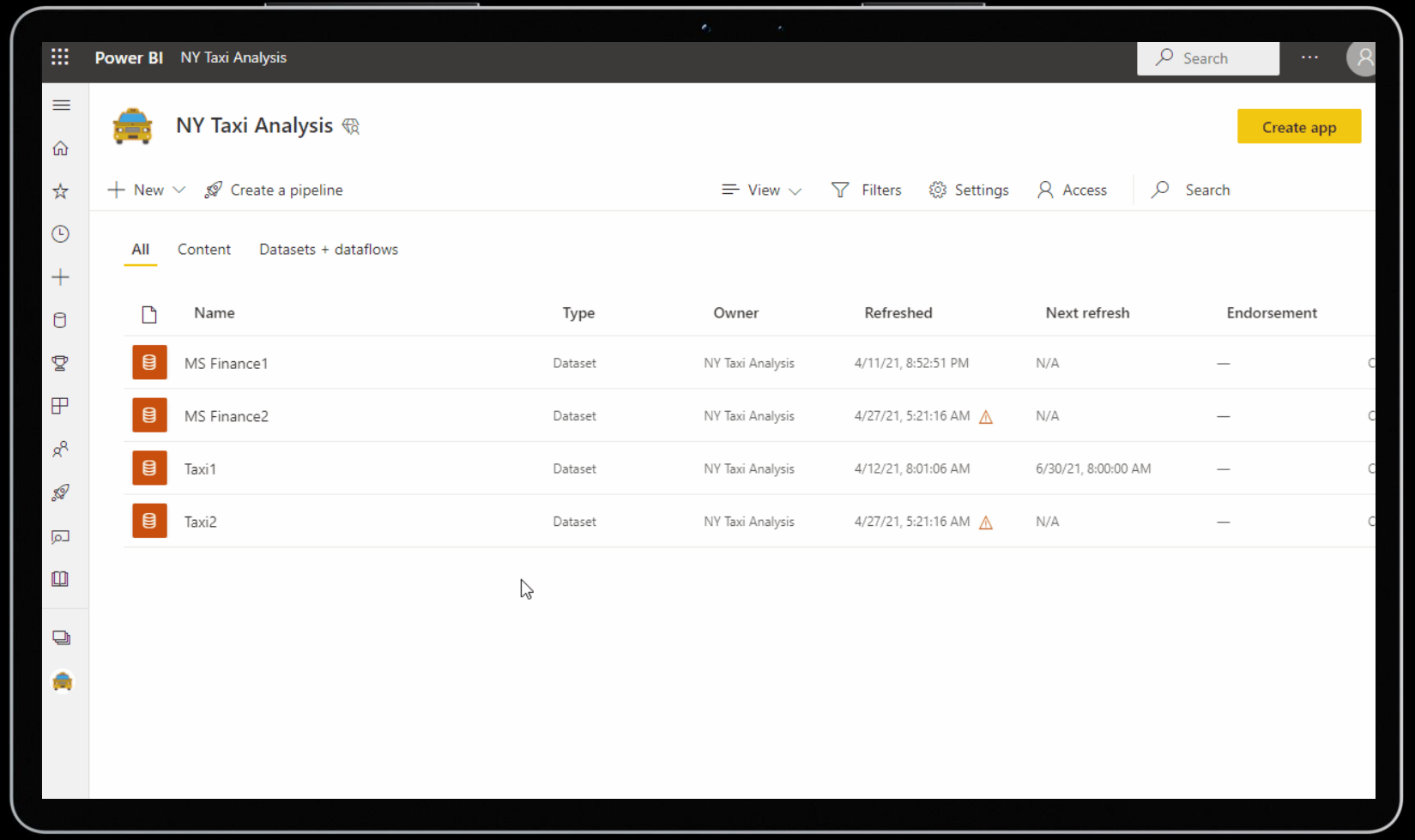
Automatically learns about customer usage patterns and create aggregates to optimize performance and reduce cost



Public Preview

Automatic aggregations

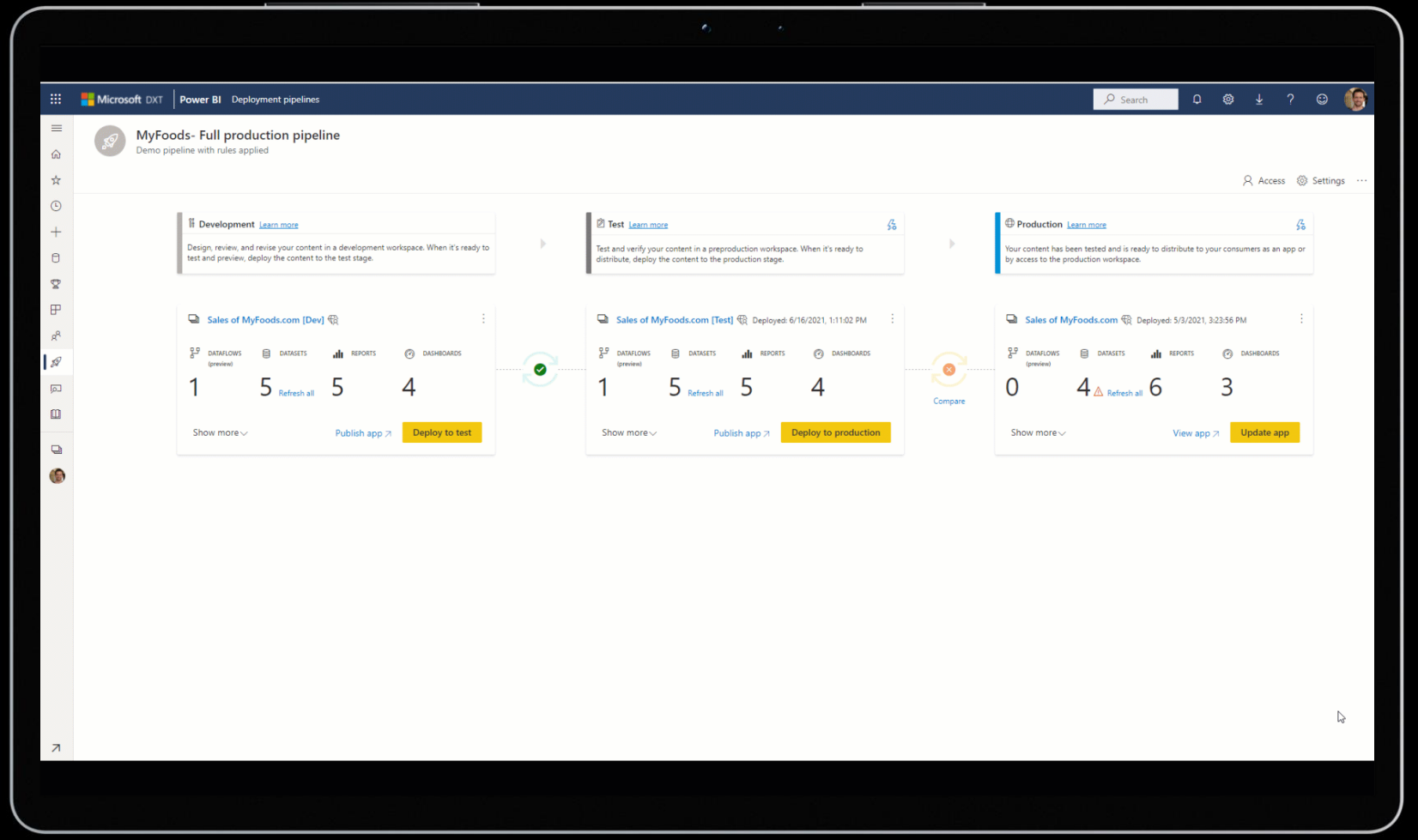
AI driven self-optimizing performance improvement



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

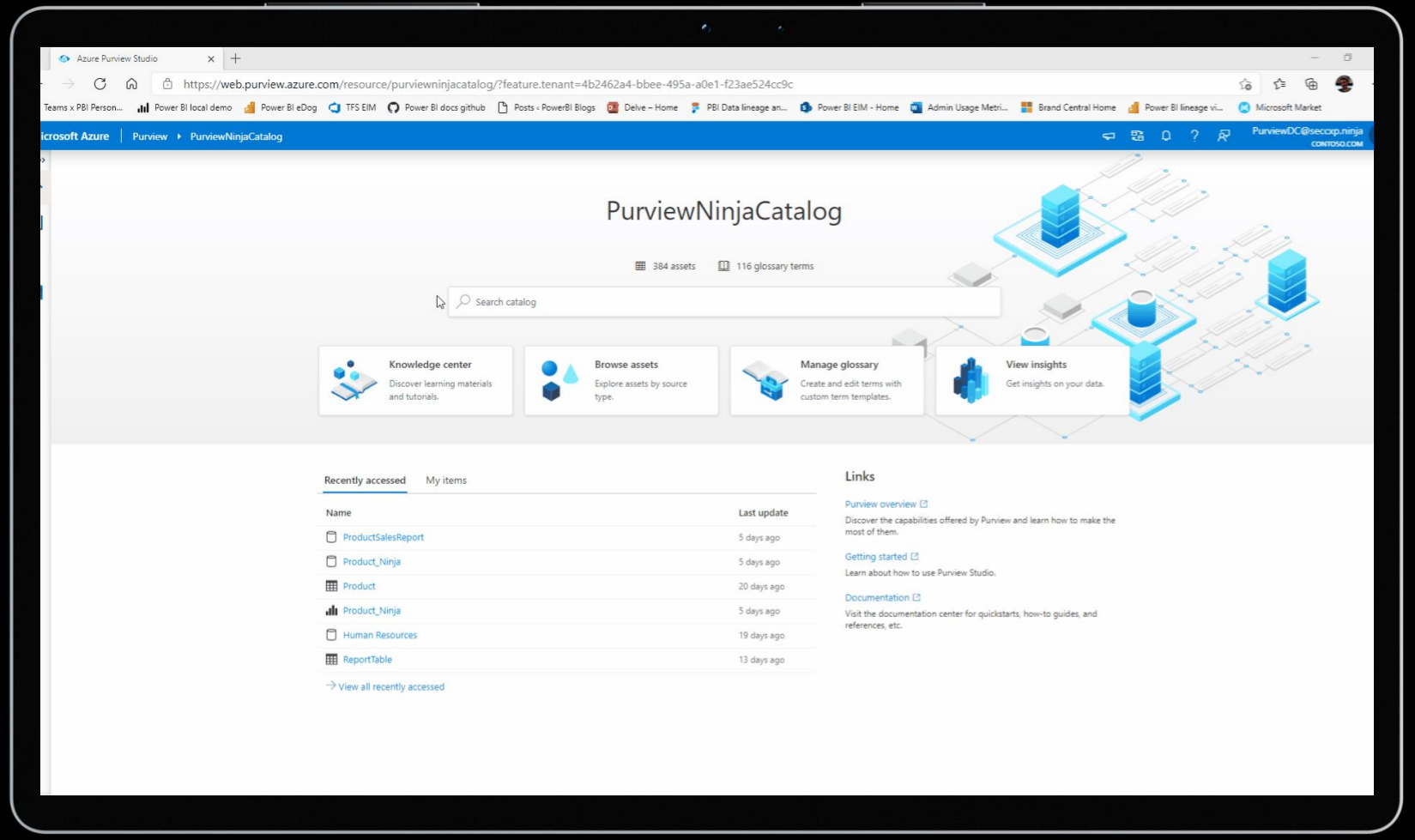
Over 50%

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

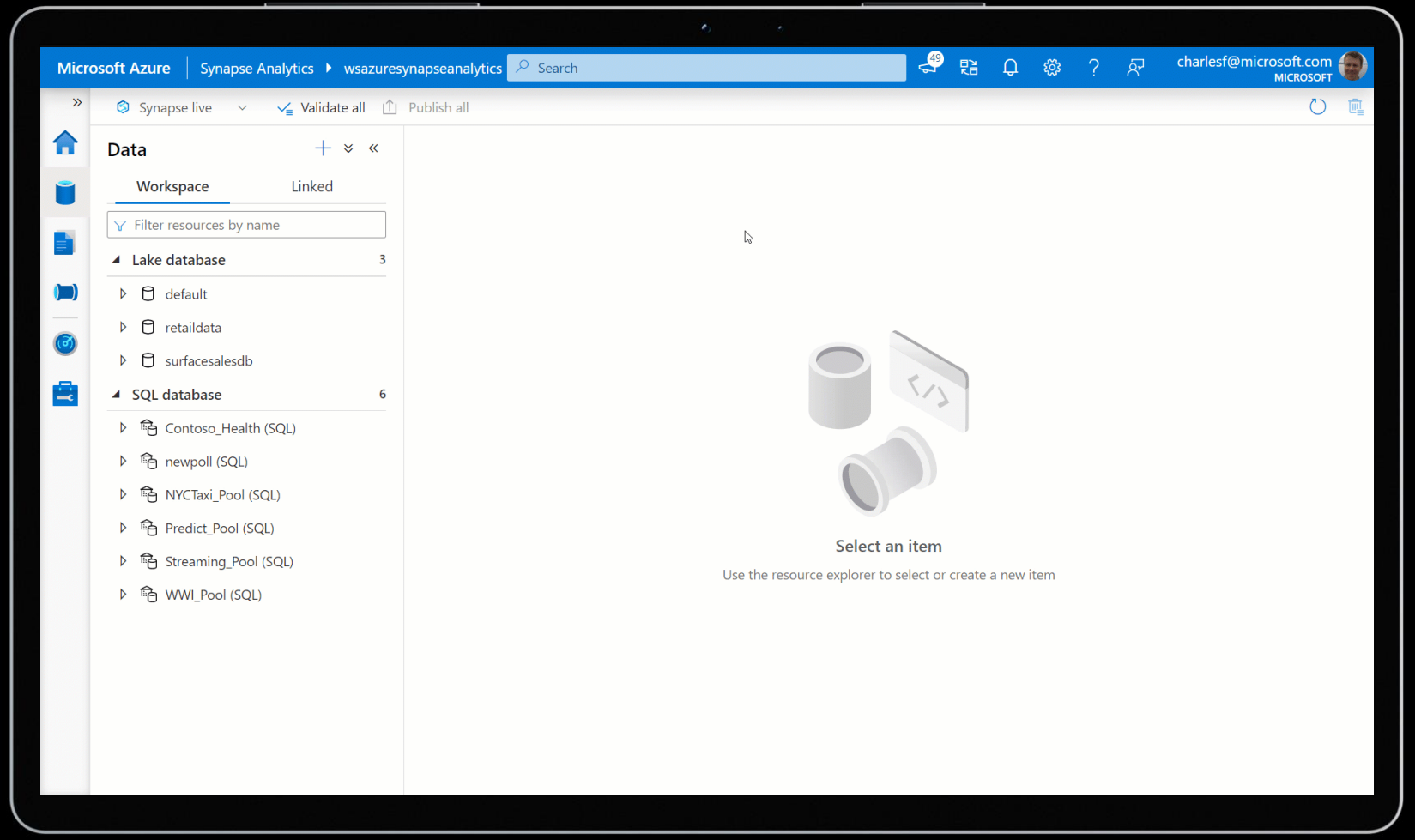


Governance

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 Warehouse
Data Science
Observation Analytics
Business Intelligence
Governance