CMC TECHNOLOGY & SOLUTION

Aspire to Inspire the Digital World

Harness the power of Data analytics

1. HARNESS THE POWER OF ANALYTICS
2. AZURE SNAPSE ANALYTICS
3. DATA INTEGRATION
4. DATA WAREHOUSING
5. REAL-TIME OPERATIONAL ANALYTICS
6. MACHINE LEARNING
7. POWER BI + AZURE

Microsoft

Harness the power of analytics

Section 1
All businesses are data businesses

Section 2
The cloud for modern analytics

Section 3
The paradox of analytics

Section 4
The analytics continuum

Microsoft

Section 1

All businesses are data businesses

Today, all businesses are data businesses

Data is the lifeblood of modern work

Data makes business possible

Data is crucial to businesses of all sizes

Securing data is mandatory

All data businesses need to be analytic businesses

Without analytics data is a cost center, not a resource

Businesses need to “know thyself”

Workers are empowered with intelligent insights

Augmented intelligence sees patterns we cannot

Analytic businesses need to evolve data science

Every business has opportunities to make analytics faster, easier, and more insightful

Data science empowers us to go farther

AI/ML creates smarter organizations

Businesses need scalable analytics

Microsoft

Section 2

The cloud for modern analytics

Modern businesses succeed in the cloud

The cloud is the default environment for new technology initiatives

Applications are increasingly cloud-native

Collaboration platforms and file-sharing are transforming

Hybrid businesses are replacing on-premises data management

A cloud analytics platform is an economic breakthrough

Price, performance, and agility

High infrastructure costs are avoided with cloud-based analytics platforms

Iterating designs and handling diverse data becomes more efficient

Elastic scaling makes world-class performance available to all users

A cloud analytics platform is the hub for all data models

Structured, unstructured, and streaming data integrated in a single, scalable, environment

Workloads can be stored, managed and provisioned together

No need to juggle processing schedules

Visualization and data science tools empower enterprise-wide insights

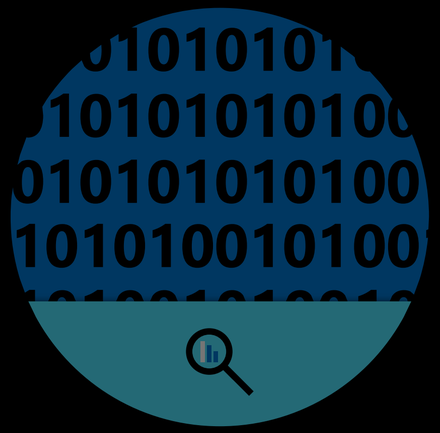
Microsoft

Section 3

The paradox of analytics

While data grows 400%, less than 30% gets analyzed

2020



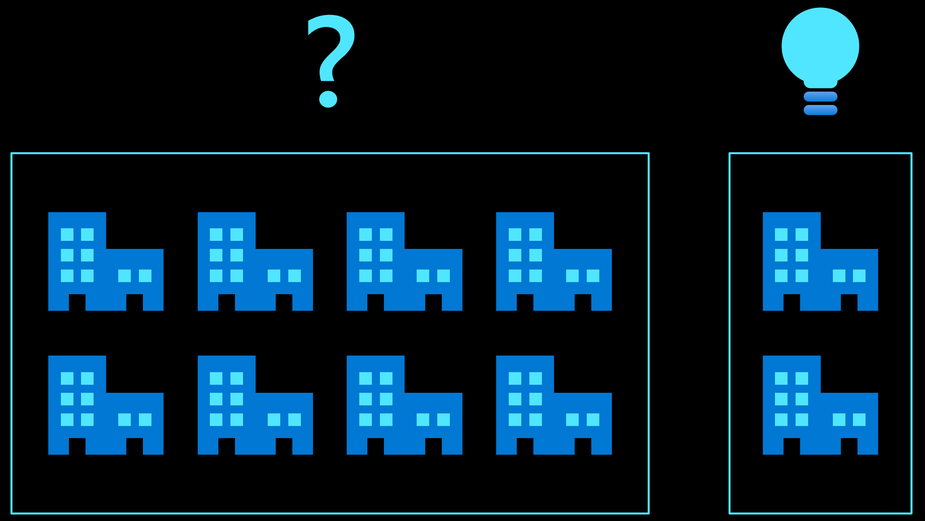
4x

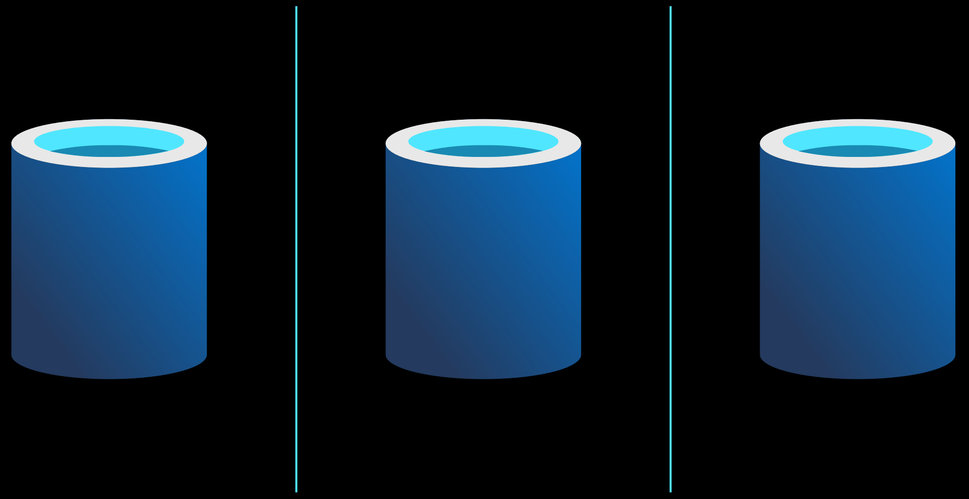
2025

44ZB Data collected worldwide

175ZB Data collected worldwide

Analytics & AI is the #1 investment for business leaders, however they struggle to maximize ROI





80% report struggling to become mature users of data*

55% report data silos and data management difficulties as roadblocks*

* Harvard Business Review (2019), Understanding why analytics strategies fall short for some, but not for others

More analytics solutions lead to more silos

Data

Structured
Unstructured
Streaming
Big data
Cloud
On-premises
IoT/edge

Technologies

Map Reduce
Open source projects
Data mart
Data warehouse
Data lake
Spreadsheets
RDMS
Data visualization tools
ML models
AI services

Skills

SQL
Python
Java
R
Industry Schemas
Data modeling
Data cleaning
Data integrations

This leads to the paradox of analytics

Each new technology creates another siloed operation

Big data

Data integration

Machine learning

Business intelligence

Data governance

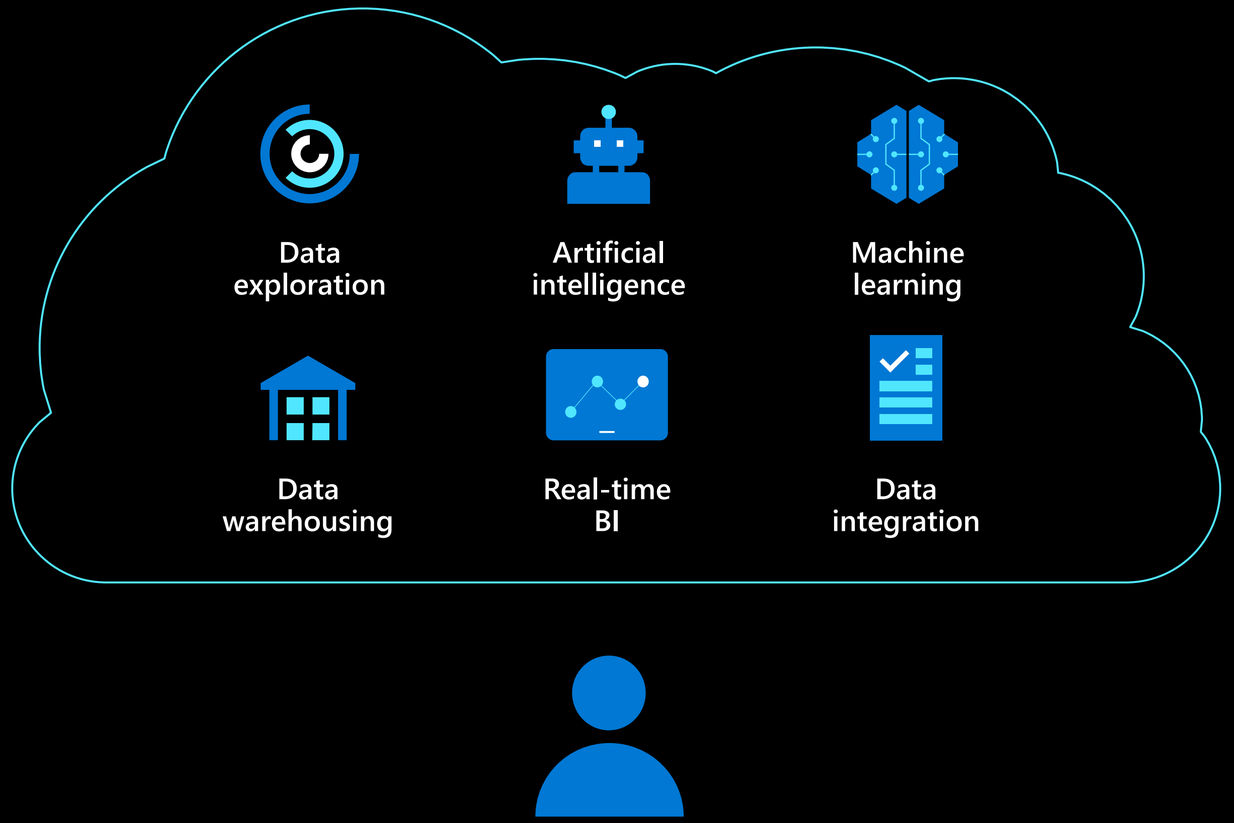
Security

Microsoft

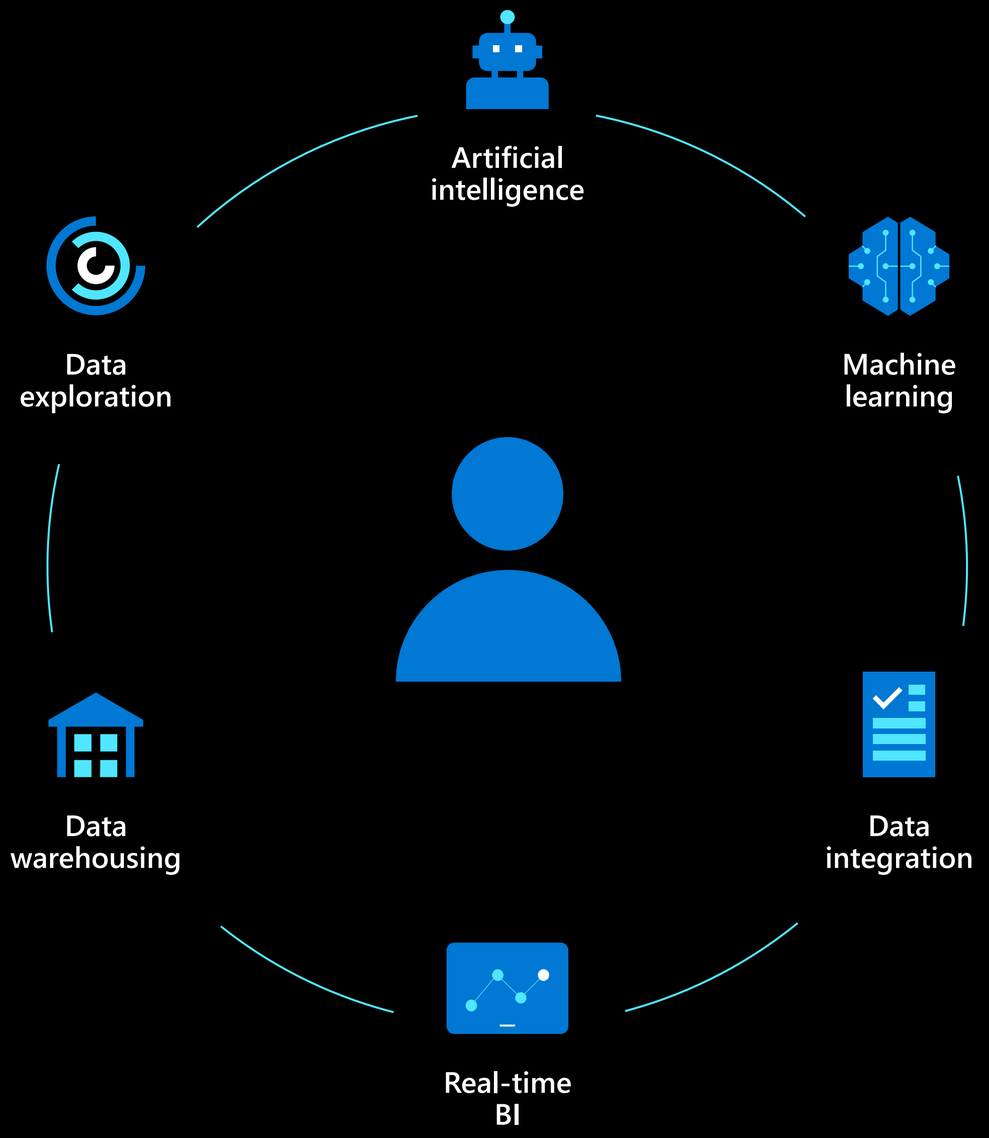
Section 4

The analytics continuum

Data professionals shouldn’t have to continually learn new skills to deliver insights



Analytics should seamlessly be part of the way users work—rather than a factor for creating more siloes

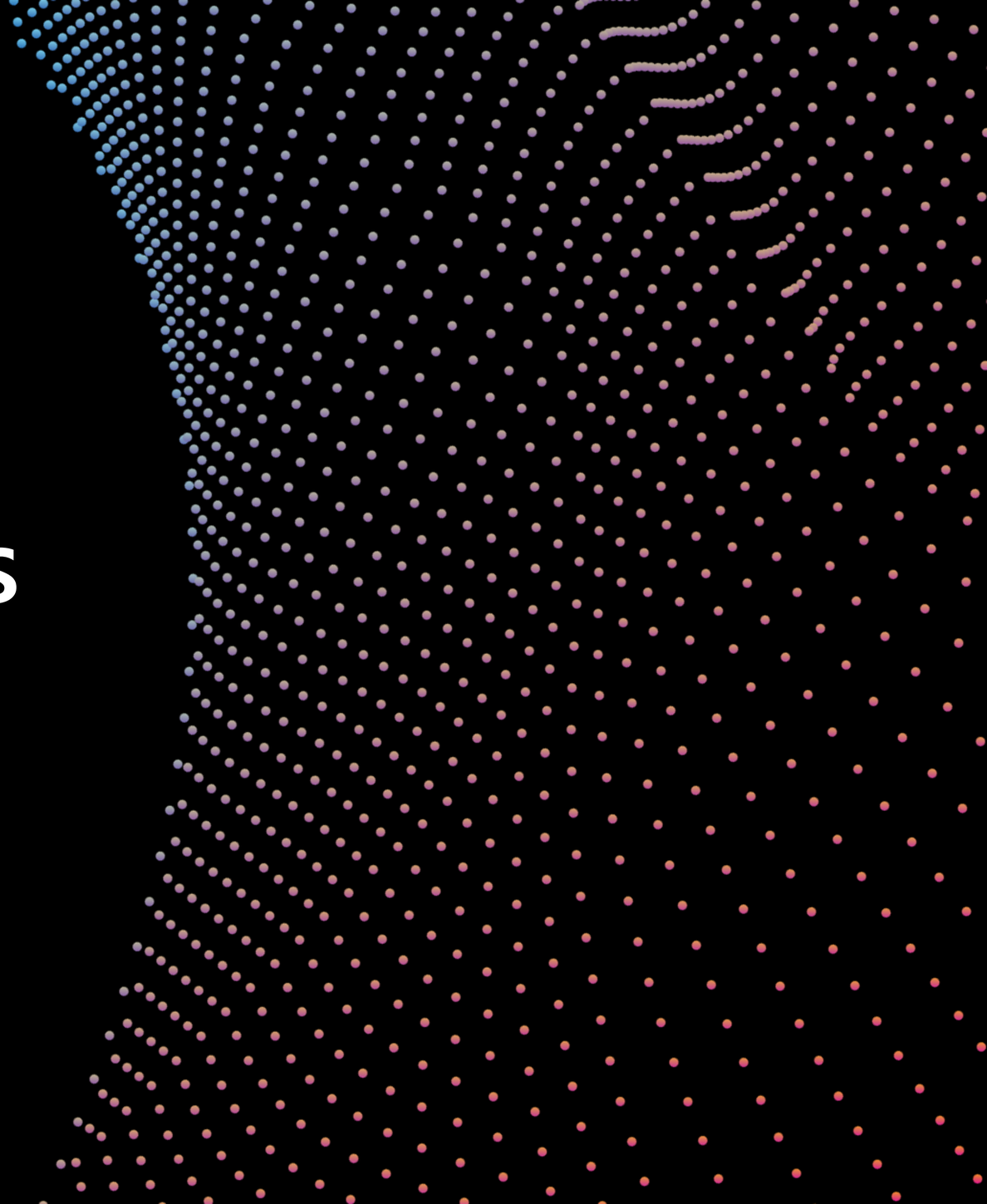


Azure Synapse delivers the full continuum of analytics in a single service

Microsoft

Azure Synapse Analytics

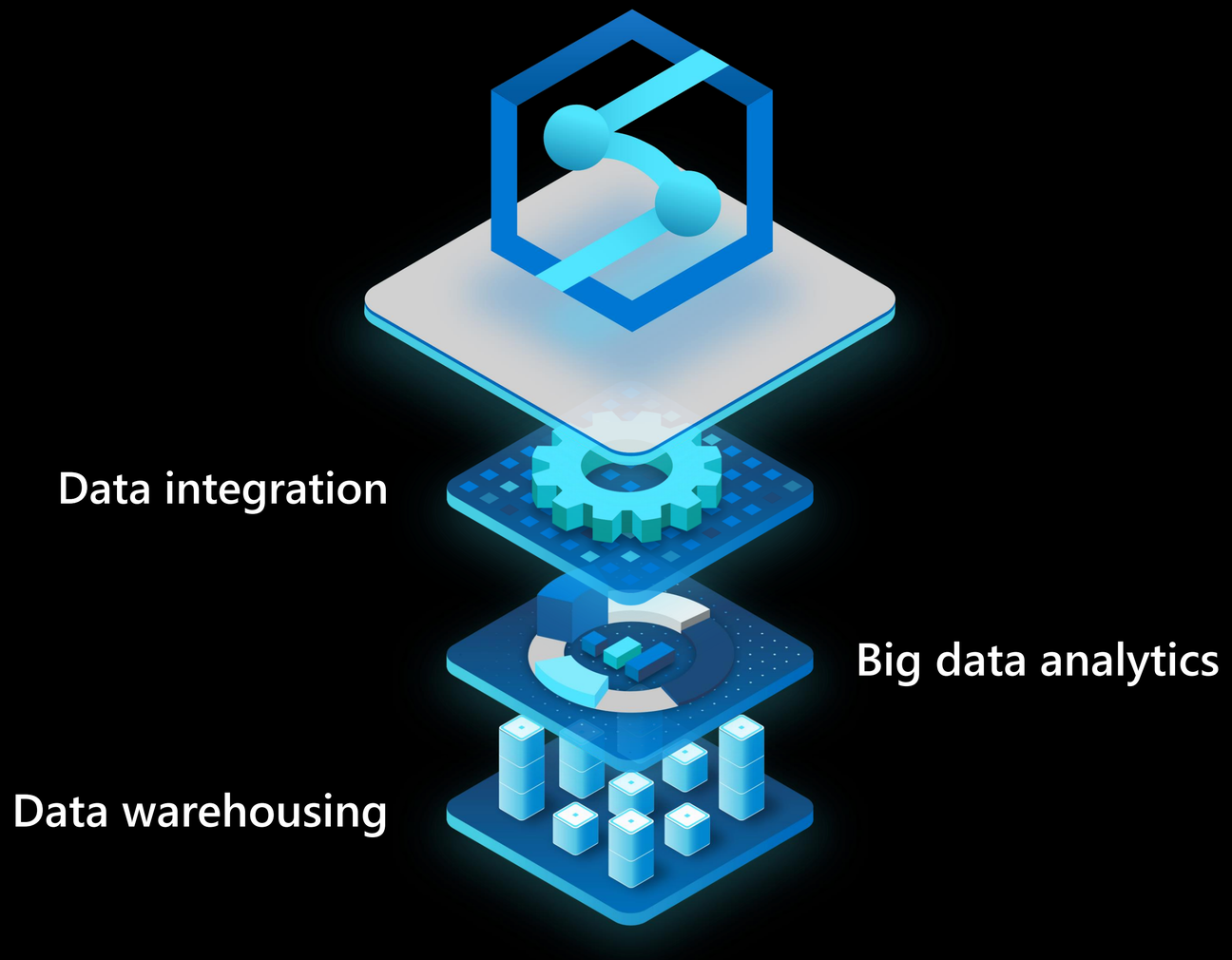
Limitless analytics service with unmatched time to insight



Azure Synapse Analytics

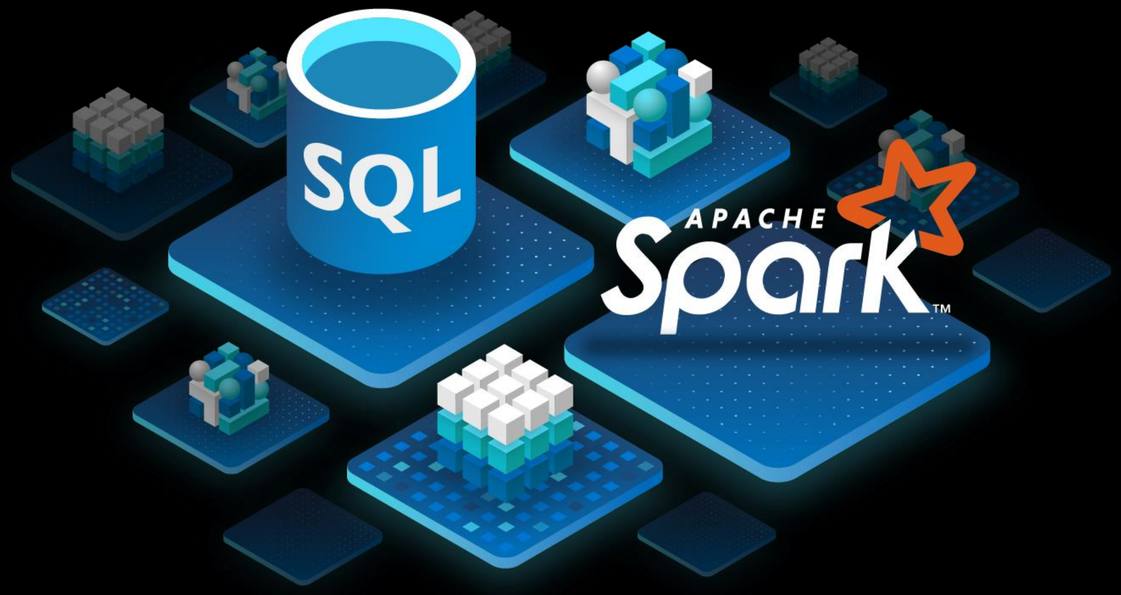
The first unified, cloud native platform for converged analytics

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

Powered by a new cloud native distributed SQL engine

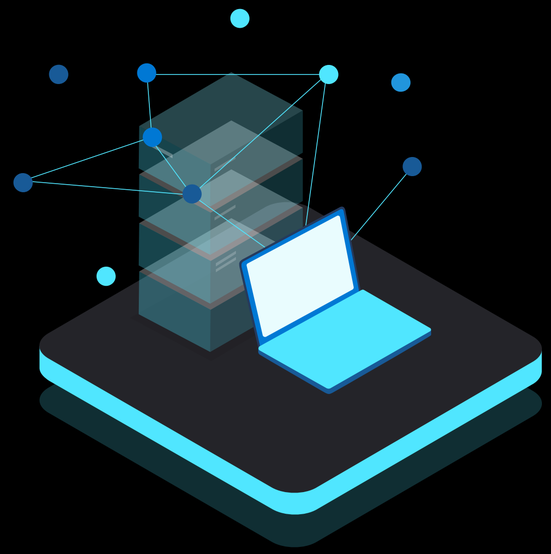


Serverless + dedicated SQL

Flexible consumption models

Serverless pay-per-query ideal for ad-hoc data lake exploration and transformation

Dedicated clusters optimized mission-critical data warehouse workloads



Serverless

Dedicated

Microsoft

Let’s dive deeper

- Comprehensive security and compliance
- Streamlined data integration
- Flexible data warehousing
- Real-time operational analytics
- Integrated machine learning
- Power BI + Azure Synapse

Data integration

Code-free hybrid data integration

Hybrid data integration

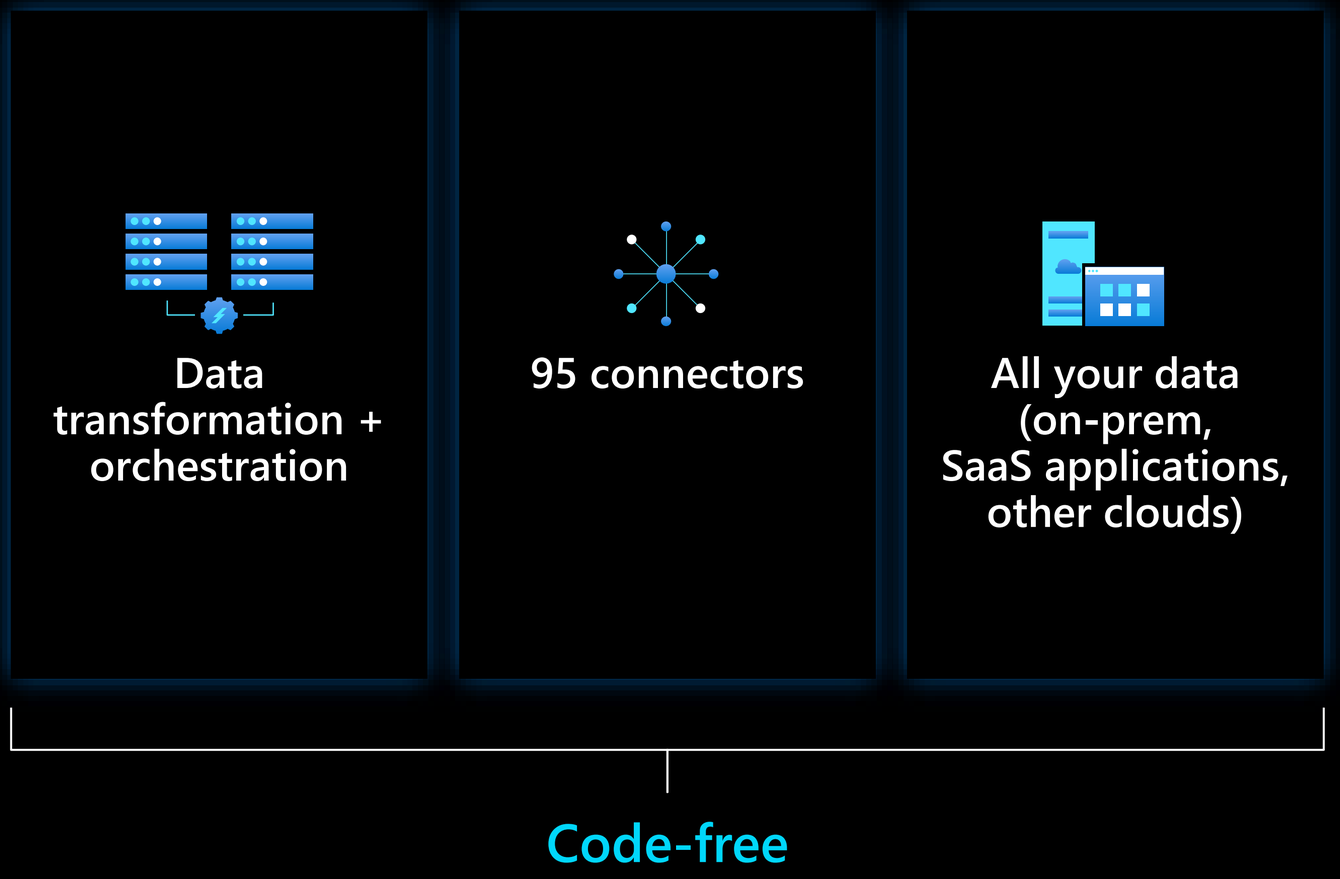
Cloud native ETL/ELT

95+ connectors available

Secure connectivity to on-premise data sources, other clouds, and SaaS applications

Code-first and low/no code design interfaces

Schedule and Event based triggering



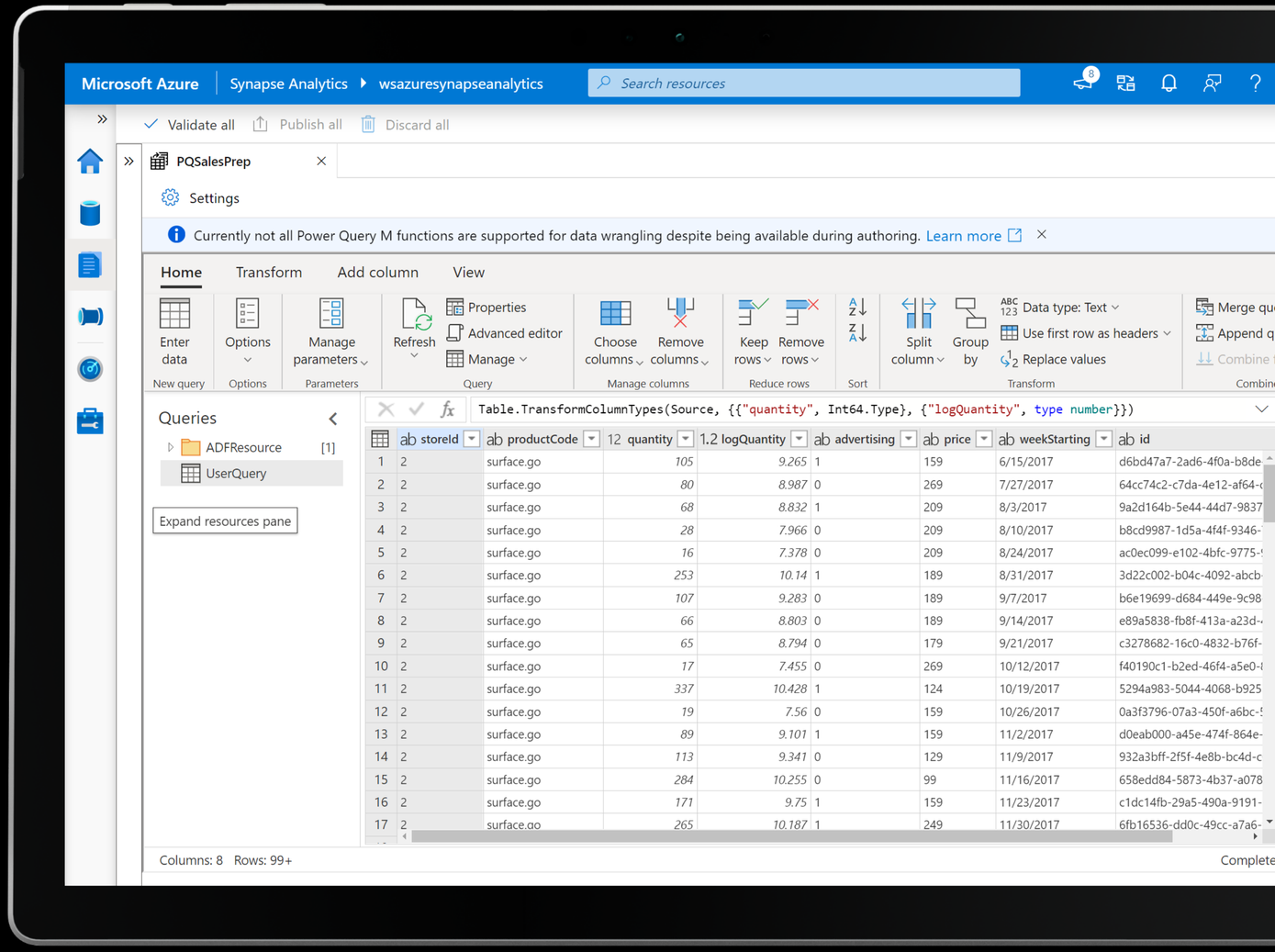
Code-free data wrangling

No/low-code data transformation

Excel-like interface is familiar to users

Transform data to desired shape completely visually

Operationalize into pipelines



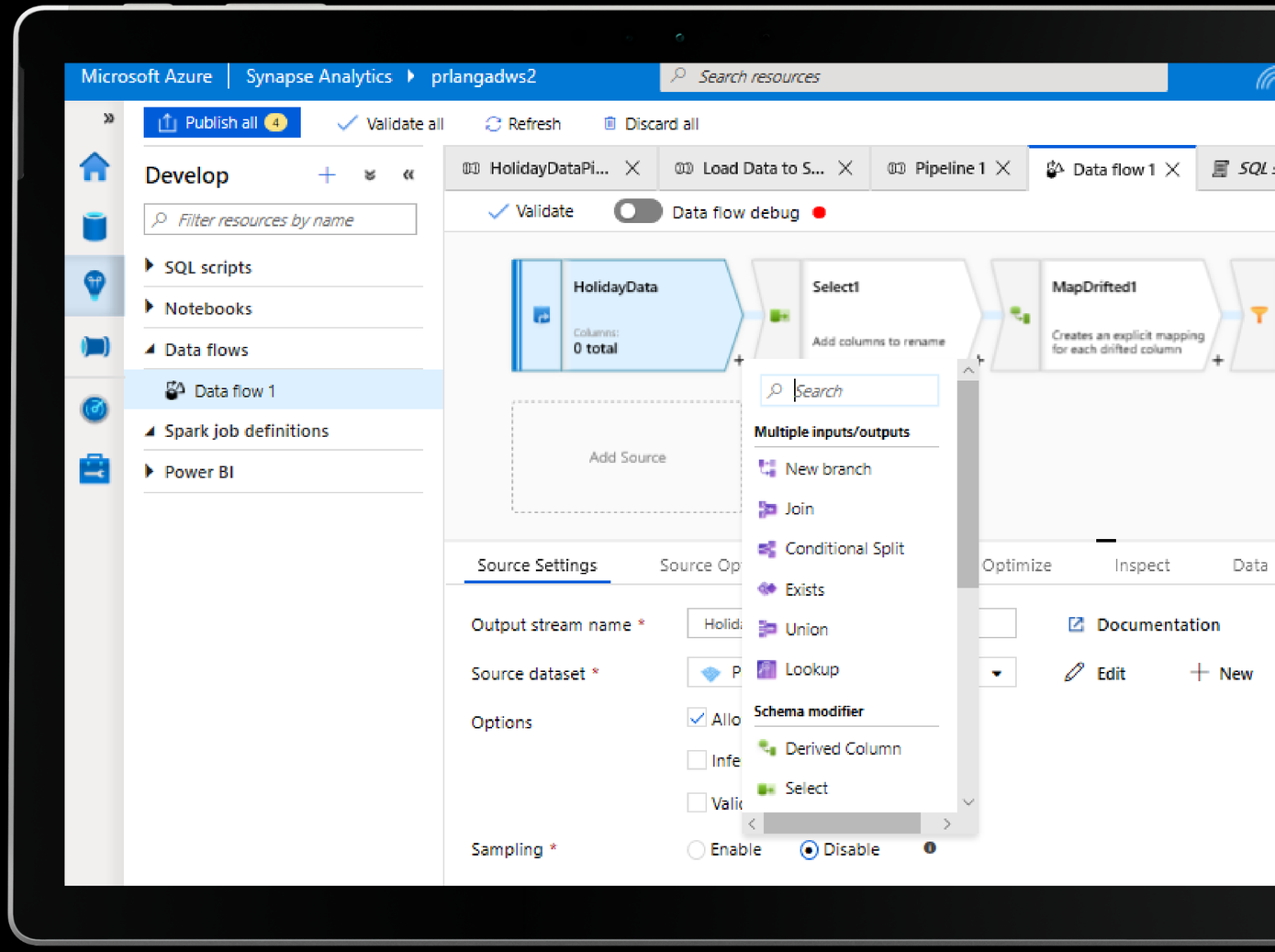
Hybrid data integration

Real-time operational analytics

No data integration pipelines required

Near-zero impact on operational systems

Latency <90s at 99th percentile



Data warehouse

Scalable and secure analytics platform for SQL workloads

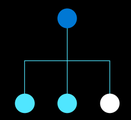
Suitable for any scale

Fully-managed elastic platform

Elastic compute that can be easily optimized to different classes of workload

All features available in a single tier

Infinite cost effective PAYG storage



Department

Gigabytes of data

10’s of users

Weekday usage

Organization

Terabytes of data

100’s of users

24/7 usage

Enterprise

Petabytes of data

1000’s of users

Tier-0 Availability

Synapse Studio

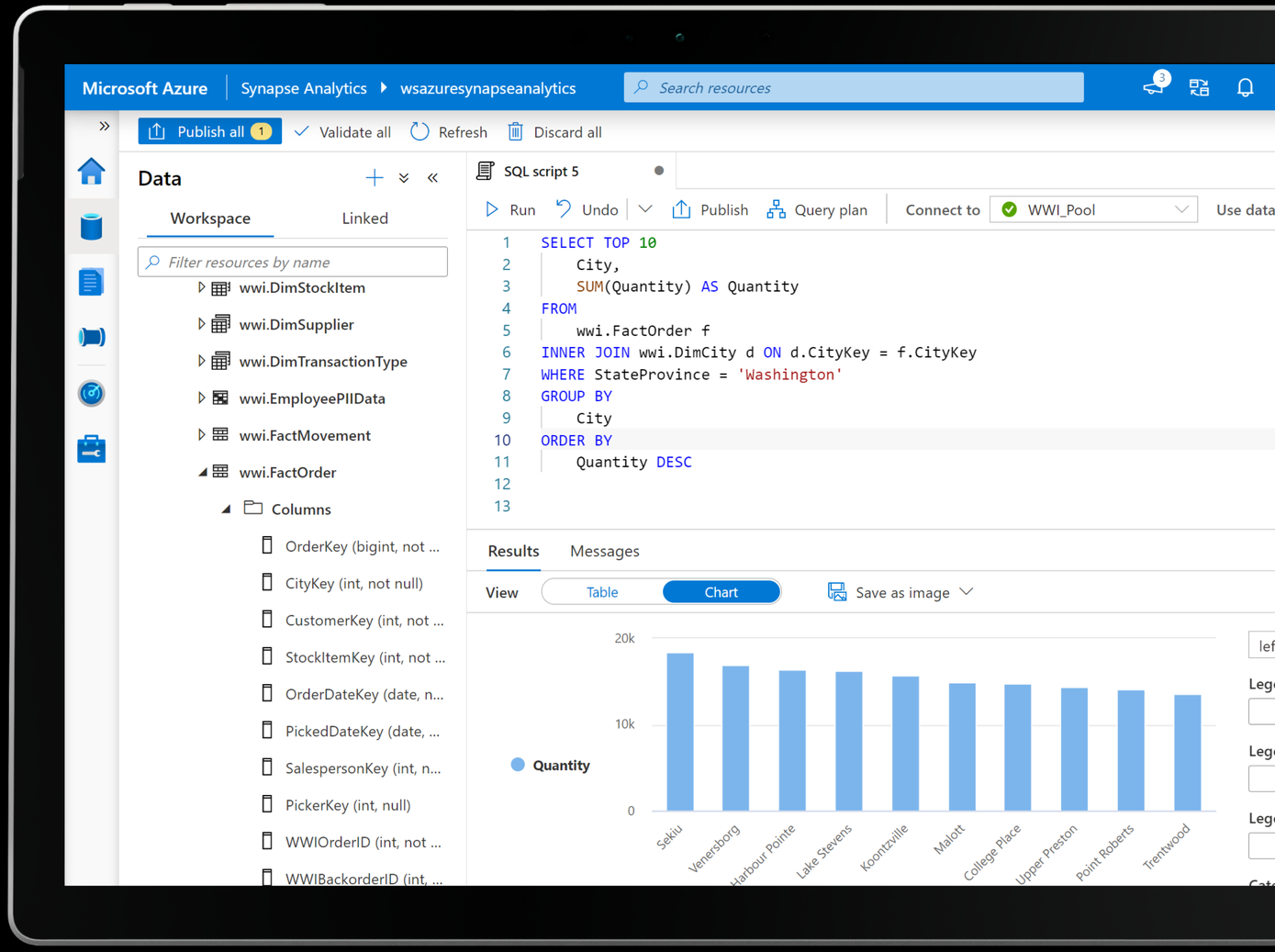
SQL Editor

Automatic code completion (Intellisense)

Script collaboration within the Workspace

Built-in visualizations

Easily switch between clusters



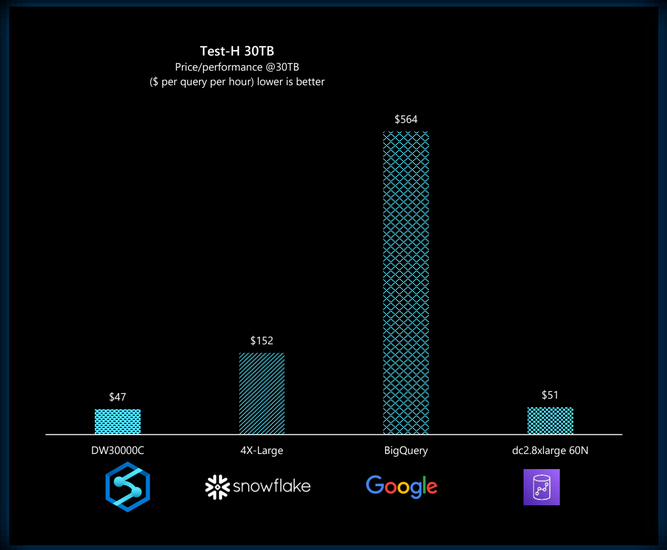
Industry leading SQL performance and scalability

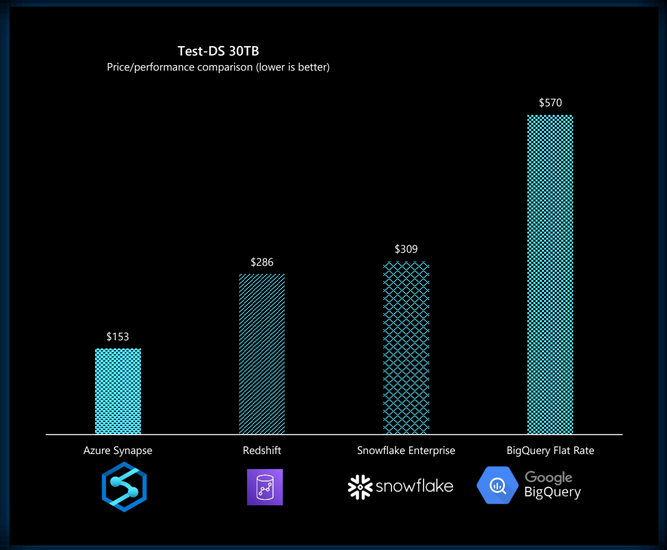
TPC-H and TPC-DS Leader

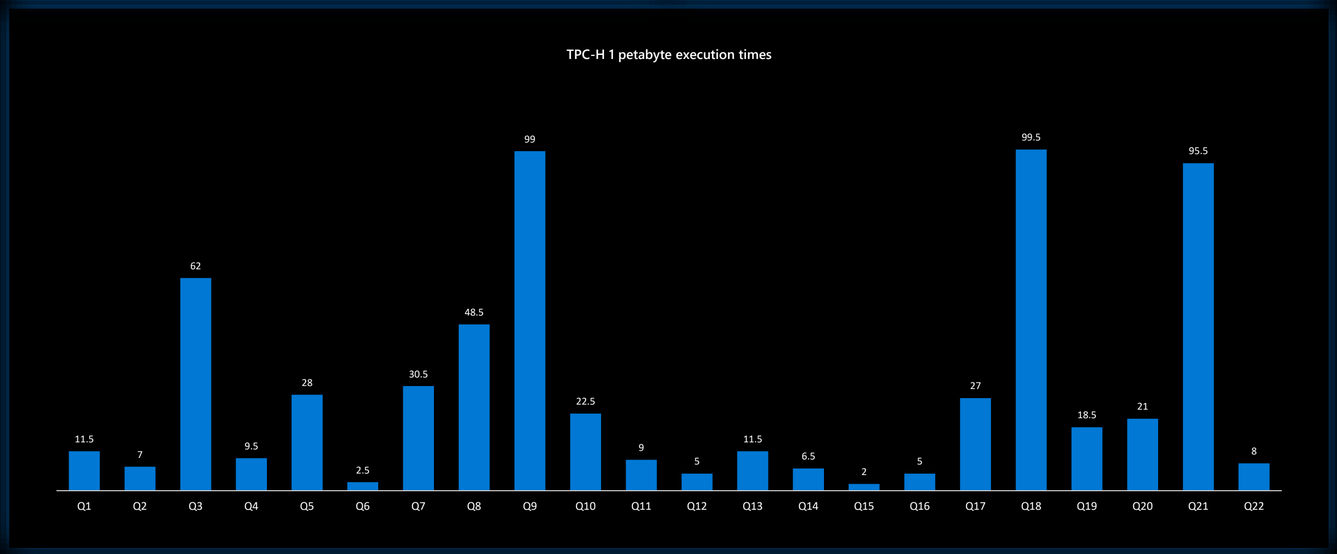
Price/performance leadership relative to other cloud data warehouses

"Polaris" is the only query engine to successfully complete TPC-H at 1PB scale

https://aka.ms/synapse-dqp







* "GigaOm Analytics Field Test-H Benchmark Report" January 2019; "GigaOm Analytics Field Test-DS Benchmark Report" April 2019

Polaris distributed query engine

Only platform to complete TPC-H benchmark at 1 Petabyte

Massive concurrency

Global workload graph

Workload aware query scheduling

https://aka.ms/synapse-dqp

POLARIS: The Distributed SQL Engine in Azure Synapse

Josep Aguilar-Saborit, Raghu Ramakrishnan, Krish Srinivasan

Kevin Bocksrocker, Ioannis Alagiannis, Mahadevan Sankara, Moe Shafiei

Jose Blakeley, Girish Dasarathy, Sumeet Dash, Lazar Davidovic, Maja Damjanic, Slobodan Djunic, Nemanja Djurkic, Charles Feddersen, Cesar Galindo-Legaria, Alan Halverson, Milana Kovacevic, Nikola Kicovic, Goran Lukic, Djordje Maksimovic, Ana Manic, Nikola Markovic, Bosko Mihic, Ugljesa Milic, Marko Milojevic, Tapas Nayak, Milan Potocnik, Milos Radic, Bozidar Radivojevic, Srikumar Rangarajan, Milan Ruzic, Milan Simic, Marko Sosic, Igor Stanko, Maja Stikic, Sasa Stanojkov, Vukasin Stefanovic, Milos Sukovic, Aleksandar Tomic , Dragan Tomic, Steve Toscano, Djordje Trifunovic, Veljko Vasic, Tomer Verona, Aleksandar Vujic, Nikola Vujic, Marko Vukovic, Marko Zivanovic

Microsoft Corp

ABSTRACT

In this paper, we describe the Polaris distributed SQL query engine in Azure Synapse. It is the result of a multi-year project to re-architect the query processing framework in the SQL DW parallel data warehouse service, and addresses two main goals: (i) converge data warehousing and big data workloads, and (ii) separate compute and state for cloud-native execution.

From a customer perspective, these goals translate into many useful features, including the ability to resize live workloads, deliver predictable performance at scale, and to efficiently handle both relational and unstructured data. Achieving these goals required many innovations, including a novel "cell" data abstraction, and flexible, fine-grained, task monitoring and scheduling capable of handling partial query restarts and PB-scale execution. Most importantly, while we develop a completely new scale-out framework, it is fully compatible with T-SQL and leverages decades of investment in the SQL Server single-node runtime and query optimizer. The scalability of the system is highlighted by a 1PB scale run of all 22 TPC-H queries; to our knowledge, this is the first reported run with scale larger than 100TB.

PVLDB Reference Format:
Josep Aguilar-Saborit, Raghu Ramakrishnan et al.
VLDB Conferences. *PVLDB*, 13(12): 3204 – 3216, 2020.
DOI: https://doi.org/10.14778/3415478.3415545

1. INTRODUCTION

Relational data warehousing has long been the enterprise approach to data analytics, in conjunction with multi-dimensional business-intelligence (BI) tools such as Power BI and Tableau. The recent explosion in the number and diversity of data sources, together with the interest in machine learning, real-time analytics and other advanced capabilities, has made it necessary to extend traditional relational DBMS based warehouses. In contrast to the traditional approach of carefully curating data to conform to standard enterprise schemas and semantics, *data lakes* focus on rapidly ingesting data from many sources and give users flexible analytic

phase of interactive analysis and reporting. While this pattern bridges the lake and warehouse paradigms and allows enterprises to benefit from their complementary strengths, we believe that the two approaches are converging, and that the full relational SQL tool chain (spanning data movement, catalogs, business analytics and reporting) must be supported directly over the diverse and large datasets stored in a lake; users will not want to migrate all their investments in existing tool chains.

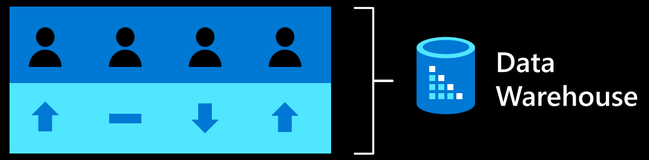
In this paper, we present the Polaris interactive relational query engine, a key component for converging warehouses and lakes in Azure Synapse [1], with a cloud-native scale-out architecture that makes novel contributions in the following areas:

- *Cell data abstraction:* Polaris builds on the abstraction of a data "cell" to run efficiently on a diverse collection of data formats and storage systems. The full SQL tool chain can now be brought to bear over files in the lake with on-demand interactive performance at scale, eliminating the need to move files into a warehouse. This reduces costs, simplifies data governance, and reduces time to insight. Additionally, in conjunction with a re-designed storage manager (Fido [2]) it supports the full range of query and transactional performance needed for Tier 1 warehousing workloads.
- *Fine-grained scale-out:* The highly-available micro-service architecture is based on (1) a careful packaging of data and query processing into units called "tasks" that can be readily moved across compute nodes and re-started at the task level; (2) widely-partitioned data with a flexible distribution model; (3) a task-level "workflow-DAG" that is novel in spanning multiple queries, in contrast to [3, 4, 5, 6]; and (4) a framework for fine-grained monitoring and flexible scheduling of tasks.
- *Combining scale-up and scale-out:* Production-ready scale-up SQL systems offer excellent intra-partition parallelism and have been tuned for interactive queries with deep enhancements to query optimization and vectorized processing of columnar data partitions, careful control flow,

Workload management

Scale in

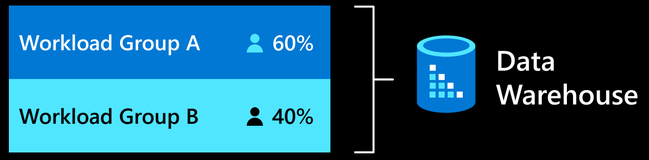
Workload Importance



Benefits:

- Predictable cost
- Bias to high-value workloads within fixed/predictable budget
- Enables customers to easily deprioritize queries which don't need to be run immediately
- Built-in starvation prevention

Workload Isolation

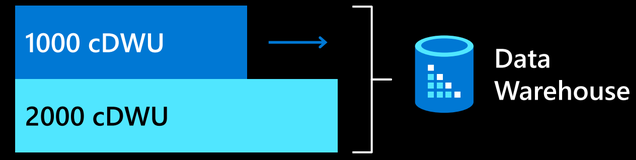


Benefits:

- Predictable cost
- Efficient for unpredictable workloads since compute can overcommit
- No cache eviction for scaling (no performance cliff at restart)
- Single connection string for isolation (unlike Snowflake virtual warehouses)

Scale out

Elastic Cluster (Scale Up)



Benefits:

- Incremental add compute
- Increase large query performance
- Single cache for heterogeneous workloads
- Single endpoint

Azure Synapse supports a more diverse set of workload management tools through workload importance, intra-cluster isolation, and elastic clusters.

ABN·AMRO

Chooses the most secure cloud DW, Azure Synapse Analytics, to transform two business critical Teradata systems

Challenge

- ABN Amro had two aging, very expensive to maintain Teradata systems housing business critical information and processes of customer, marketing, and financial reporting data
- The bank was looking for a secure and integrated data platform able to decouple consumers from data providers

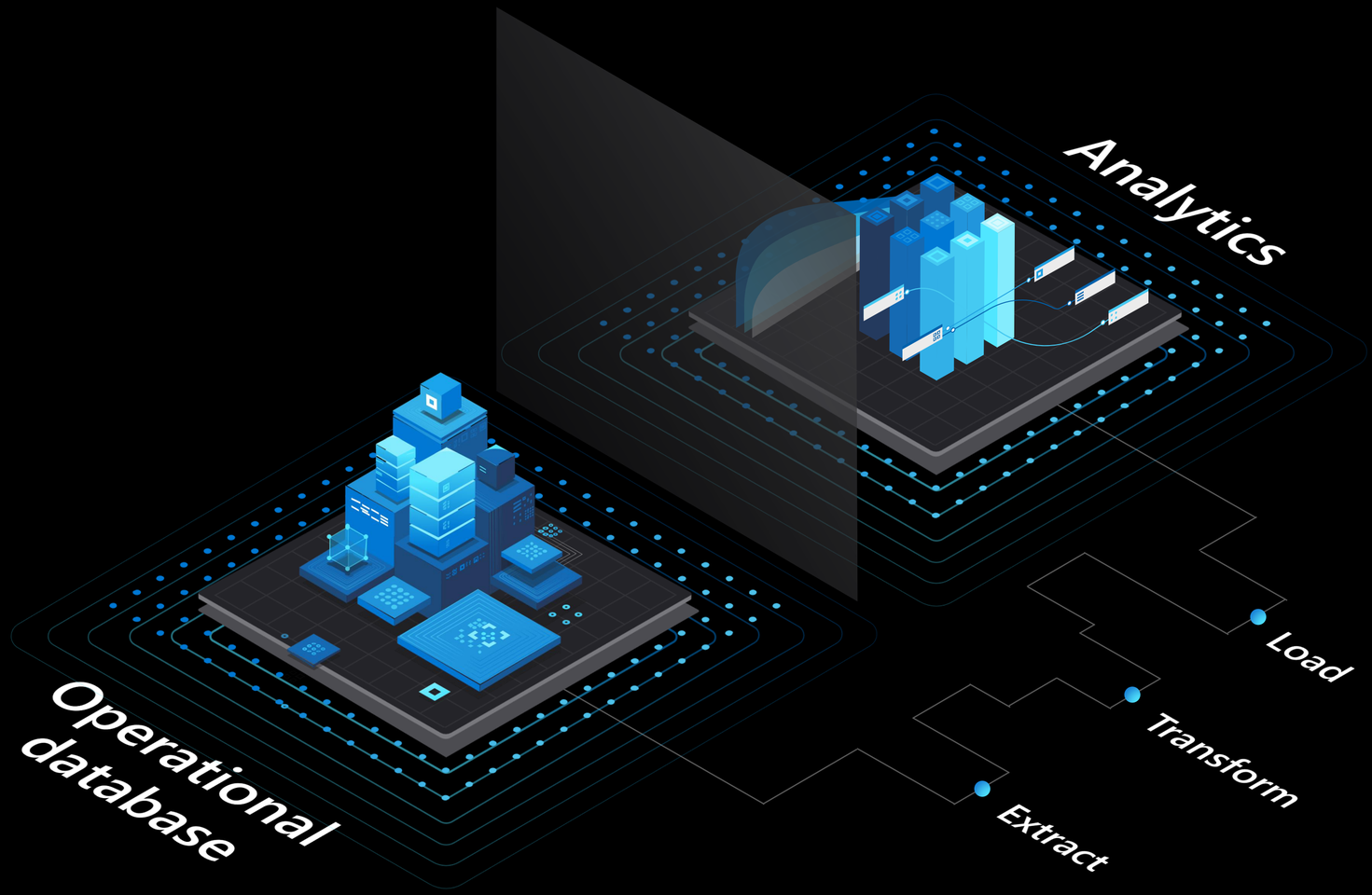
Solution

- **Dramatically cost reduction** eliminating the Teradata contract along with 70% savings using Azure Synapse Analytics reserved capacity and reducing operational complexity.
- **Increased agility**, By decoupling consumers from data providers, ABN AMRO will be able to bring new services and use cases much more quickly
- **Strong security**, Azure Synapse was considered the most secure cloud DWH, other cloud options were too risky for a bank to implement
- **Better data insights**, The new platform creates an integrated data platform, giving the company an end-to-end view of customer, marketing, and financial data empowering users to make better business decisions

Real-time operational analytics

Eliminate latency and accelerate decision making

Integrating operational data with analytics systems



Integrating operational data with analytics systems

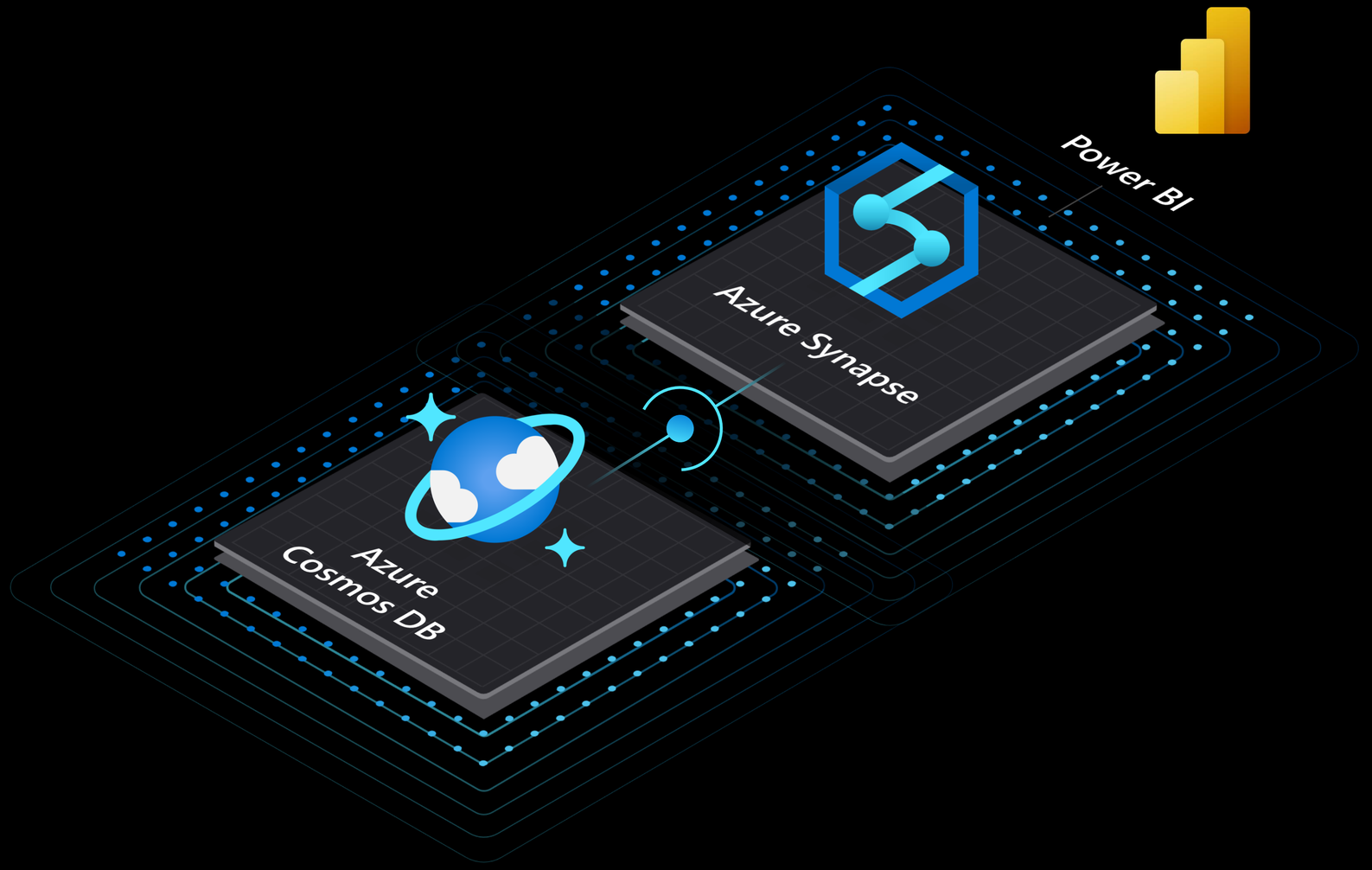
Azure Synapse Link

Real-time data analytics

No ETL required

No performance impact on transactions

Integrating operational data with analytics systems



Azure Synapse Link

Enable real-time data analytics

Cloud native HTAP with Synapse Link

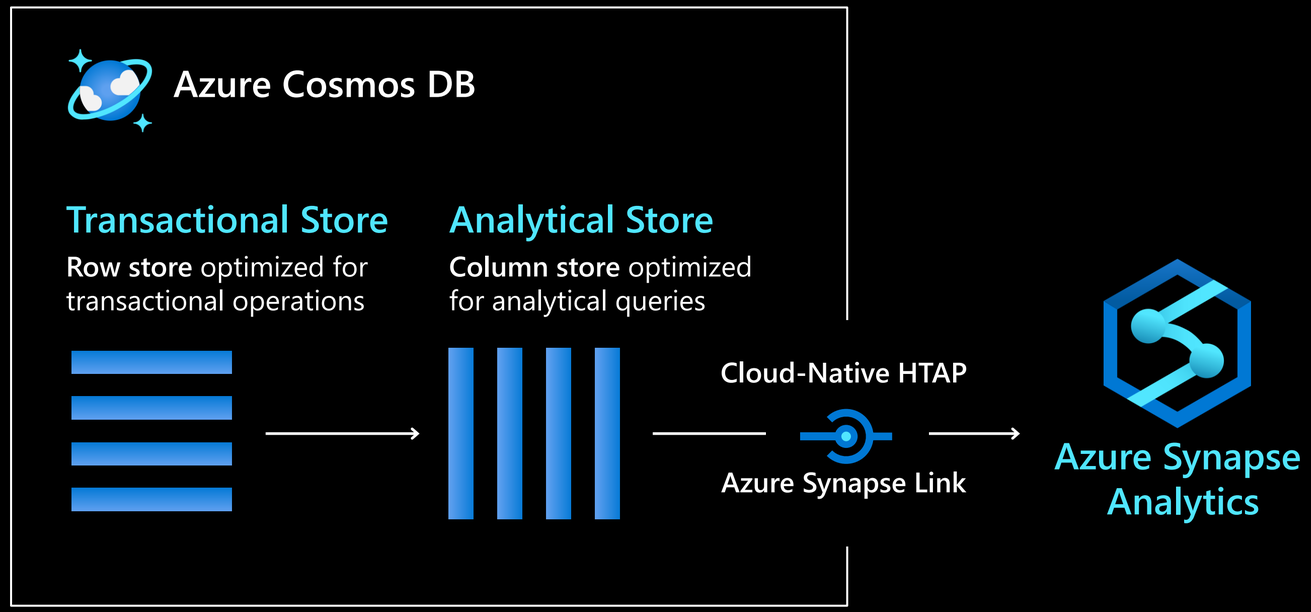
Real-time operational analytics

One-click enablement in Azure Portal

No data integration pipelines required

Near-zero impact on operational systems

Latency <90s at 99th percentile



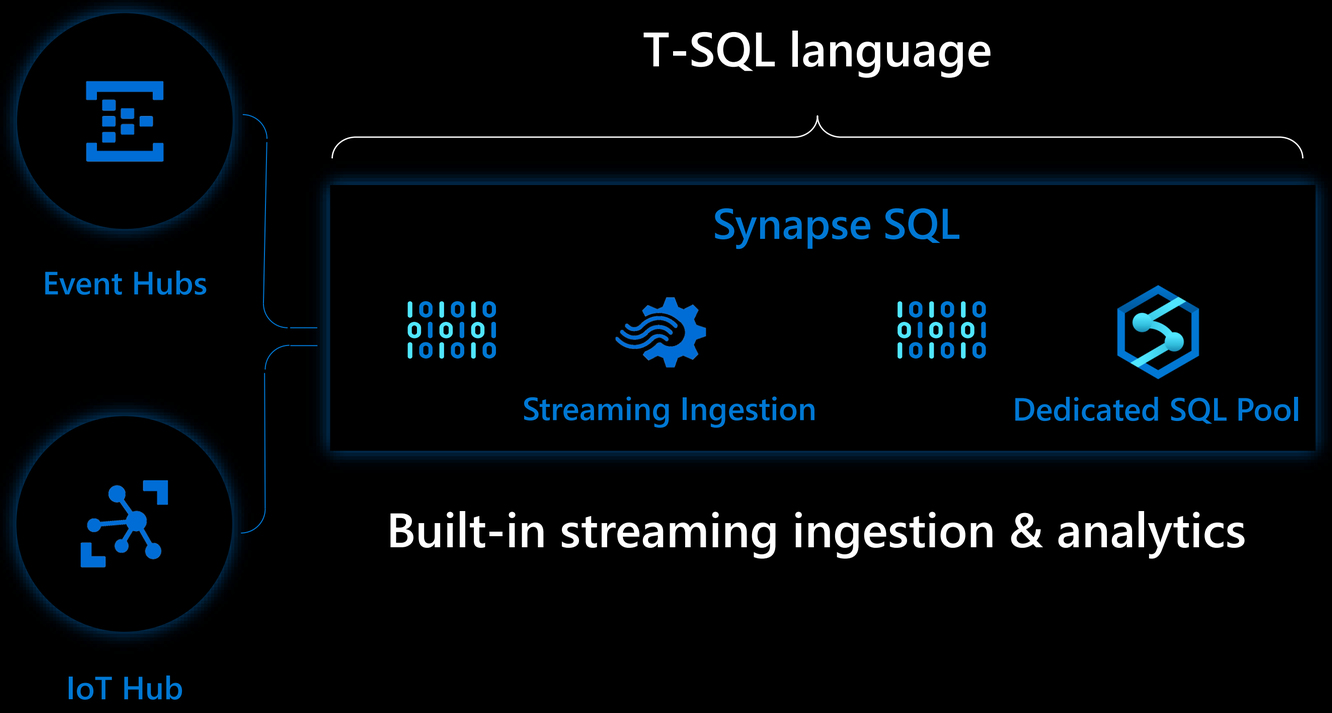
Streaming event ingestion

IoT ingestion without aggregation

Ingress up to 720 gigabytes/hour of raw events

Analyze data in-flight using SQL language Azure Stream Analytics

Join streaming data with other data assets in the data warehouse and data lake



TECHNOALPIN®

New ways for optimizing Snow production and operational costs with Azure Synapse

The global market leader in snow-making technology

Challenge

- Provide more reliable plants adopting predictive assistance and maintenance with increasing amount of data
- Real time access to all data – an important limitation of the old infrastructure
- Make snowmaking more qualitative and resource-efficient, reducing operational costs, saving valuable resources like water and energy

Solution

- An advance analytics platform enabling TechnoAlpin to monitor over 30.000 snow producers and thousands of components of the snow-making plants of nearly 400 clients world-wide with Azure Synapse Analytics
- Service Improvement with faster reaction on emerging issues
- Offering their customers new ways for optimizing their snow production and the operational costs, thanks to the unification of different data sources such as weather forecasts or snow depth measurements
- TechnoAlpin is now developing training offers that are individually tailored to client's needs and the specific characteristic of their plant

"Azure data services have been seamlessly integrated into existing infrastructure, which was especially helpful with respect to authentication and user access management."

Machine learning

Empower everyone with predictive insights

Machine learning

Democratize predictive power

Synapse makes predictive analytics accessible to all

Notebooks provides a code authoring experience for complex predictive models

Automatic ML graphical interface provides a no-code experience for creating ML models

Native integration with Azure Cognitive Search provides access to pre-built models

Notebook IDE code authoring

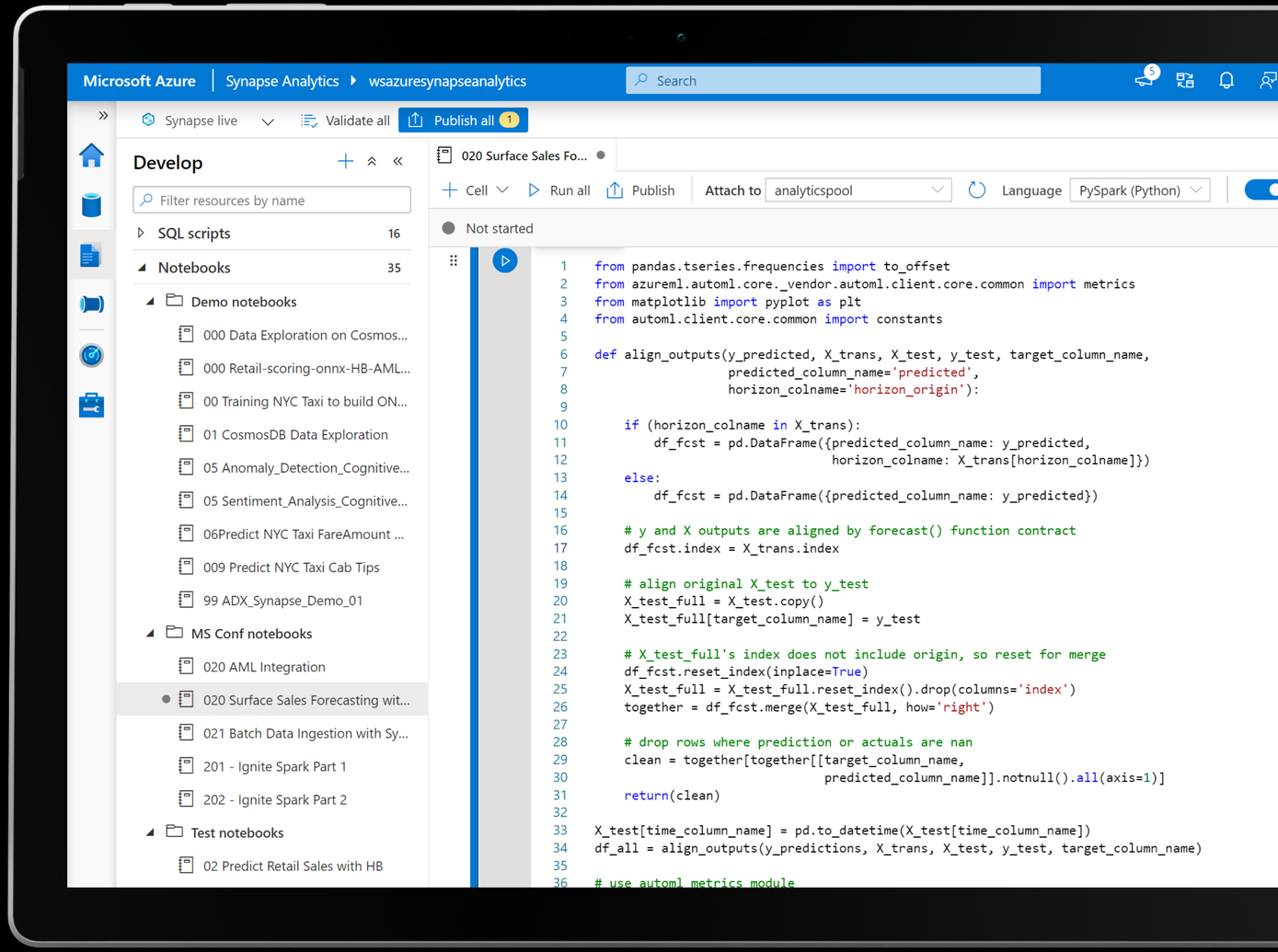
Code-first ML model development

PySpark, Scala, and C# languages supported

Automatic code completion (Intellisense)

Author multiple languages in a single notebook

Analyze data from the data warehouse, data lake, and real-time operational data from one place

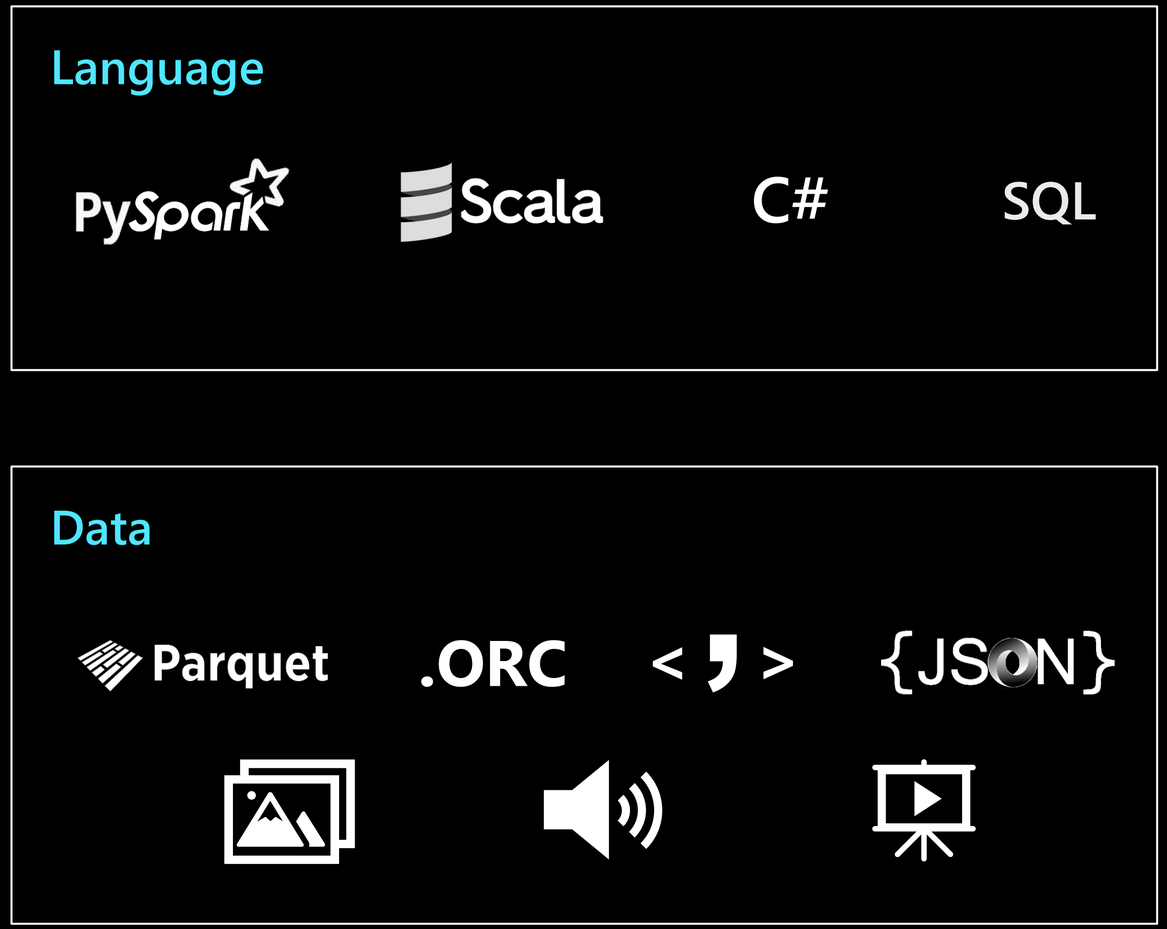


Open ecosystem with support for industry standards

Data + Languages

Languages such as SQL, PySpark, Scala and C# in support of data science and data warehouse workloads

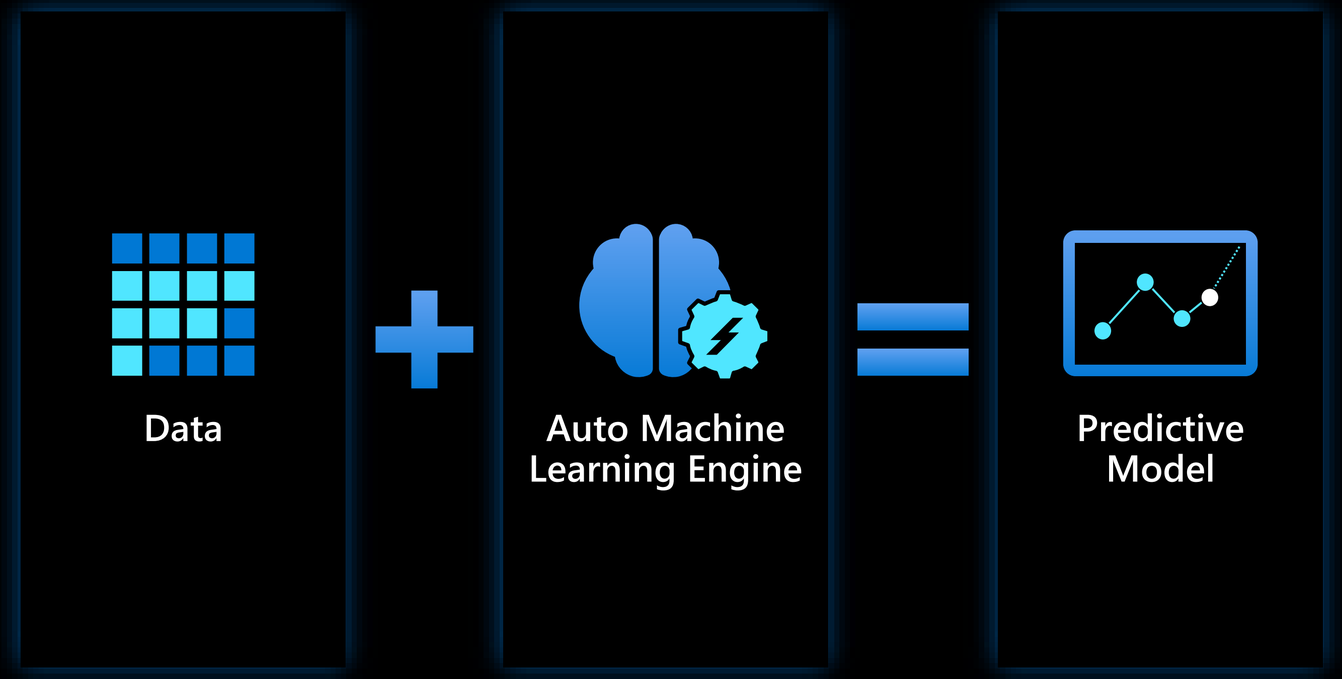
The data lake supports and unlimited set of file formats including Parquet, ORC and Json as well as audio, image, and video formats



Automatic machine learning

All you need is data

Fully automated feature exploration



Automatic machine learning

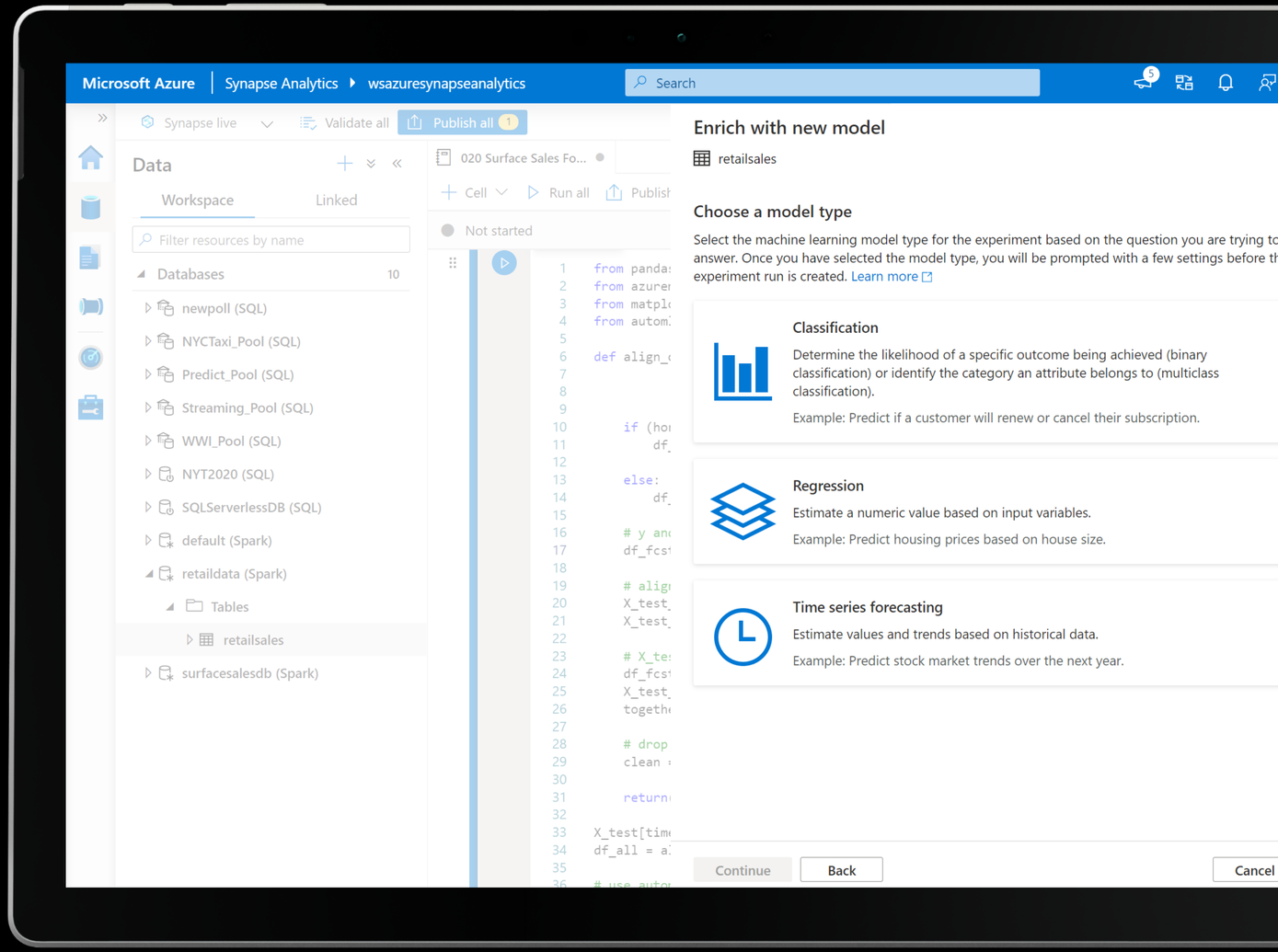
Code-free in Synapse Studio

No-code creation on Machine Learning models

Democratize ML to everyone since no data science domain knowledge required

Support for ensemble models

Supports classification, regression, and time-series forecasting



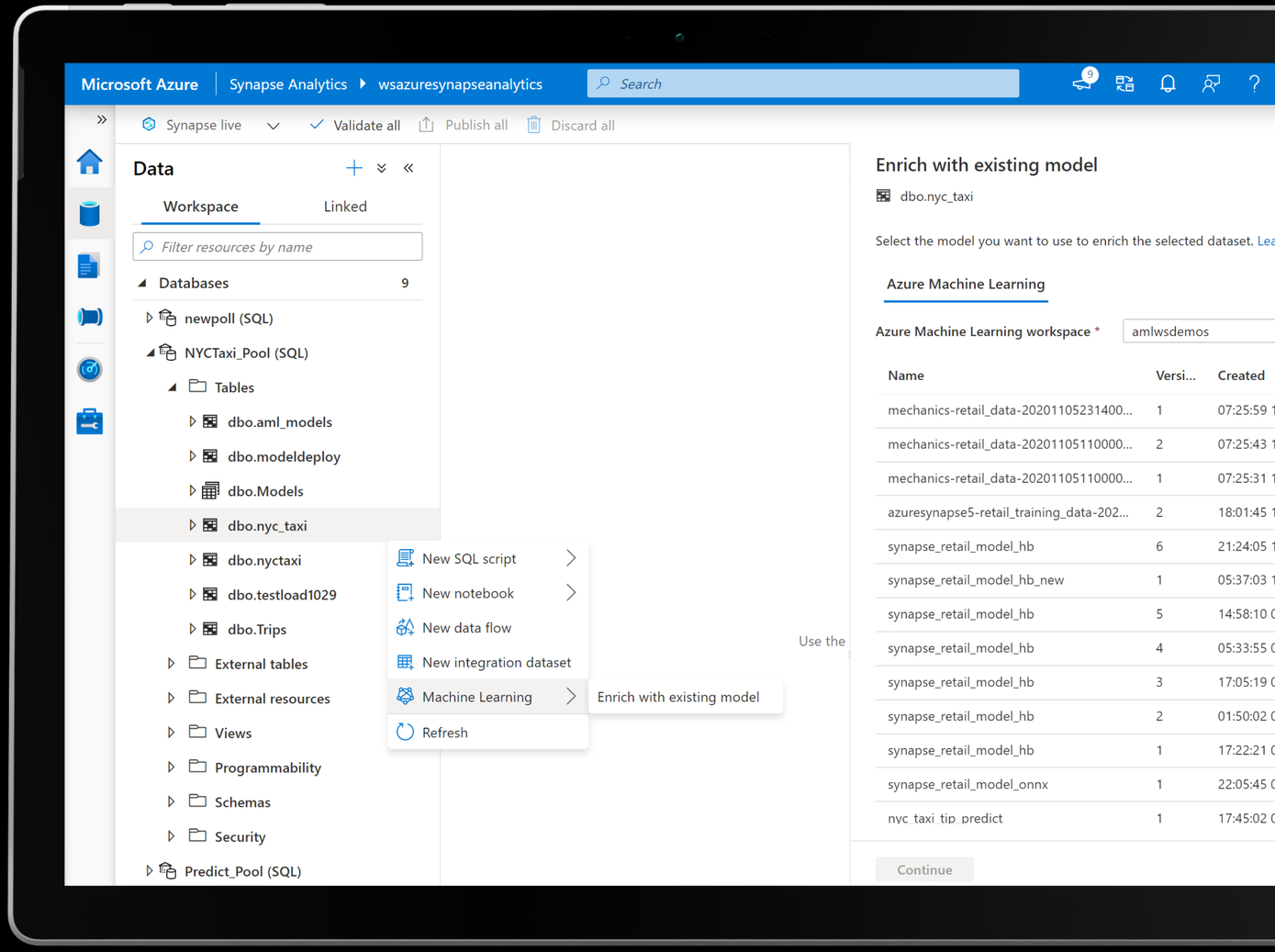
Code-free machine learning scoring

Code-free in Synapse Studio

No-code references to machine learning models

Democratize ML to everyone since no data science domain knowledge required

Easily embed in SQL stored procedures for transformation of Views for reporting



Democratize predictions to all

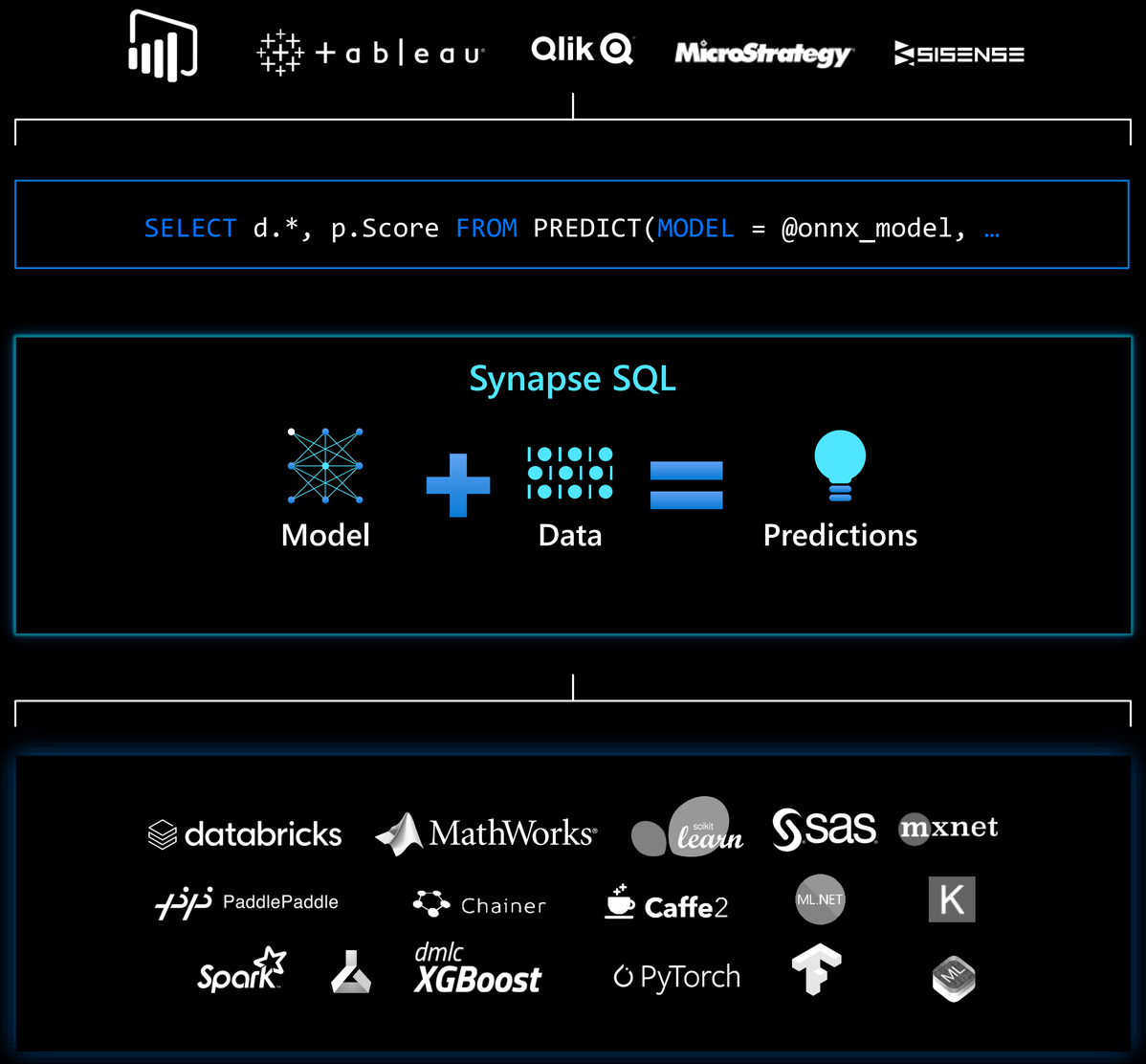
In-engine ML scoring

Machine learning models executed using SQL

“In-engine” for performance and scalability

No data leaves the platform for scoring

No additional cost for scoring



AmerisourceBergen®

Gains a unified analytics platform at scale with Azure Synapse Analytics

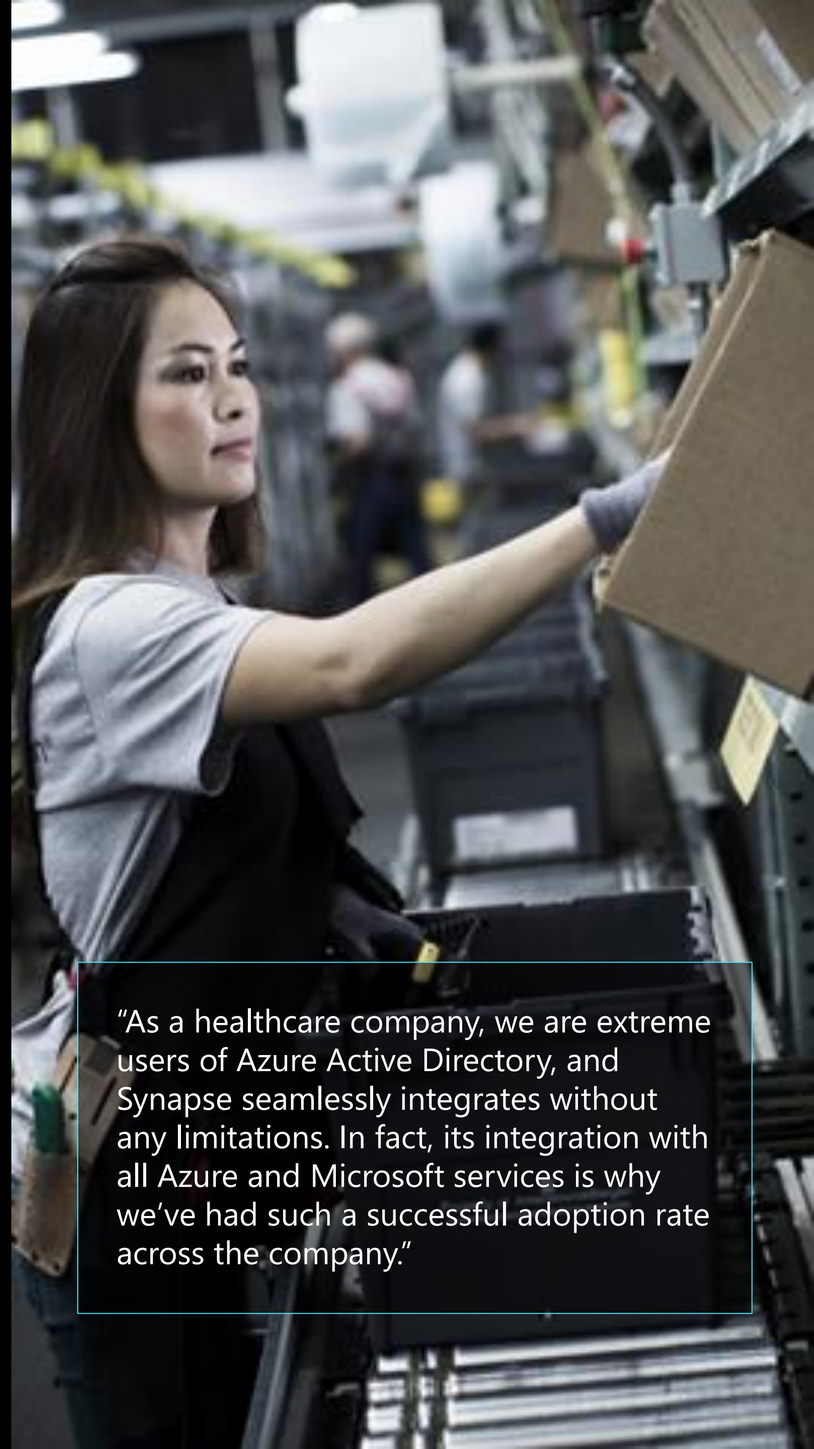
AmerisourceBergen is a leading pharmaceutical sourcing and distribution company, handling around 20 percent of the pharmaceuticals distributed throughout the US.

Challenge

- Need a centralized Enterprise Analytics Platform where business users could access data to apply their own analytics and get insights without help from data engineers.
- Struggle to produce the timely analytics at scale. Business groups began building their own analytics models on-premises, leading to reporting and performance delays.

Solution

- Azure Synapse Analytics provides them a versatile and self-served platform that accelerates time to insight, boosts analytics performance, and optimizes the operation costs.

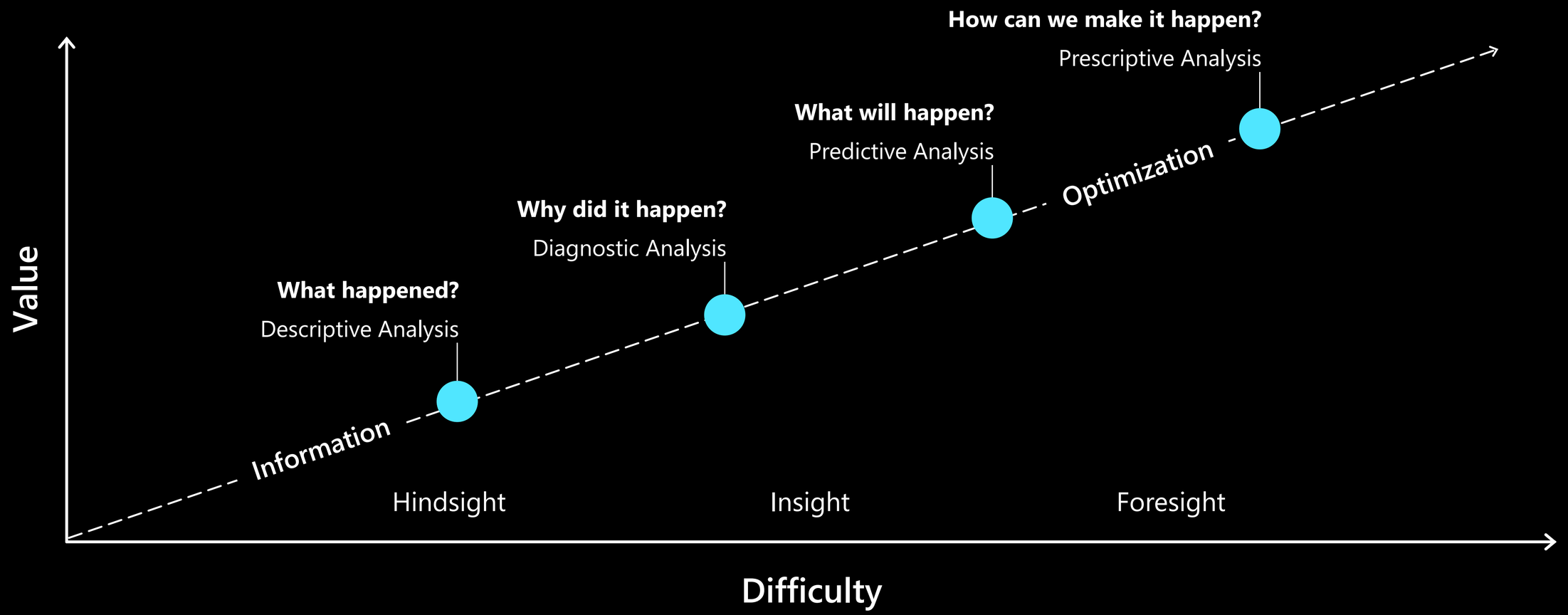


"As a healthcare company, we are extreme users of Azure Active Directory, and Synapse seamlessly integrates without any limitations. In fact, its integration with all Azure and Microsoft services is why we've had such a successful adoption rate across the company."

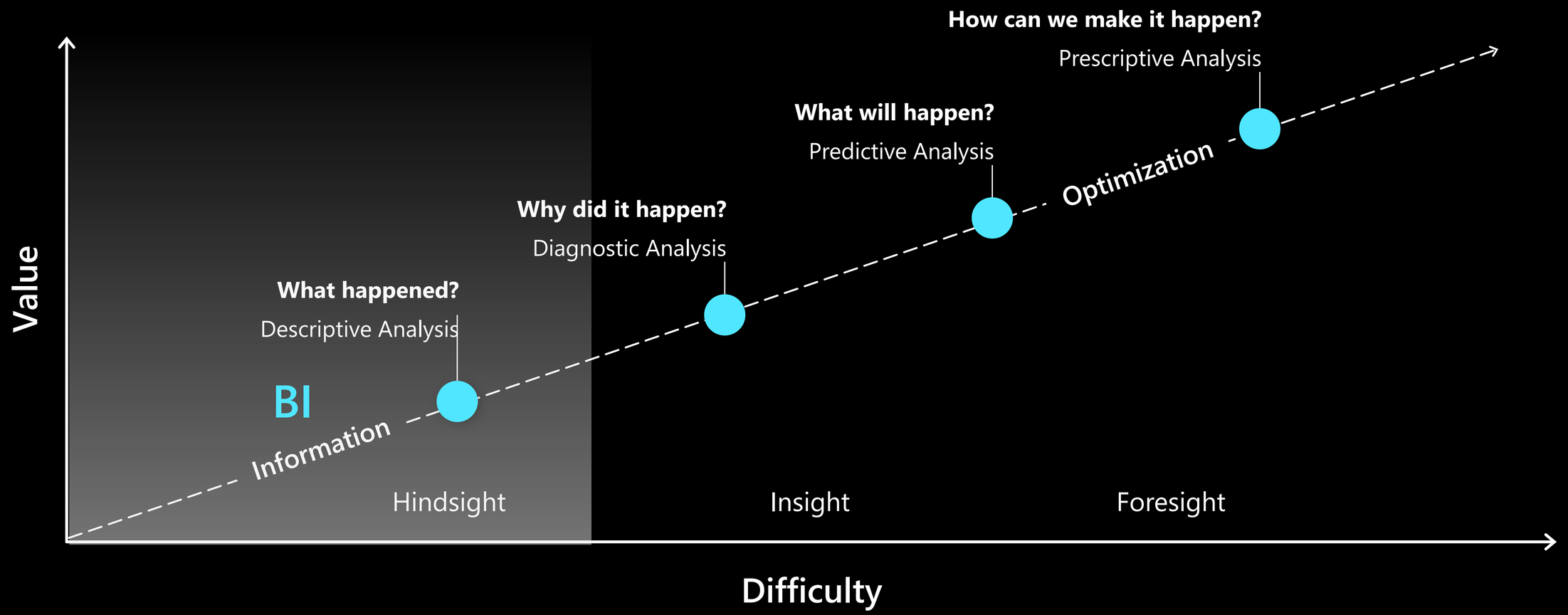
Power BI + Azure Synapse

An unmatched combination

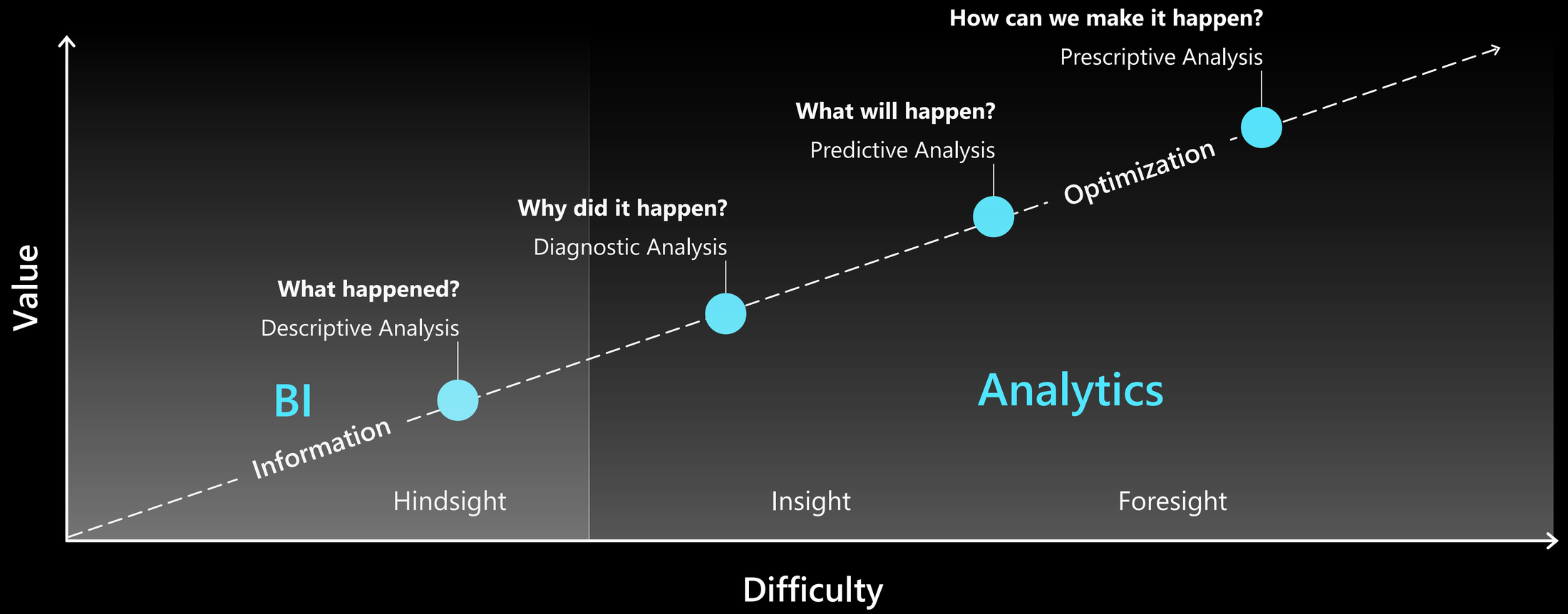
Where do you find yourself on the curve?

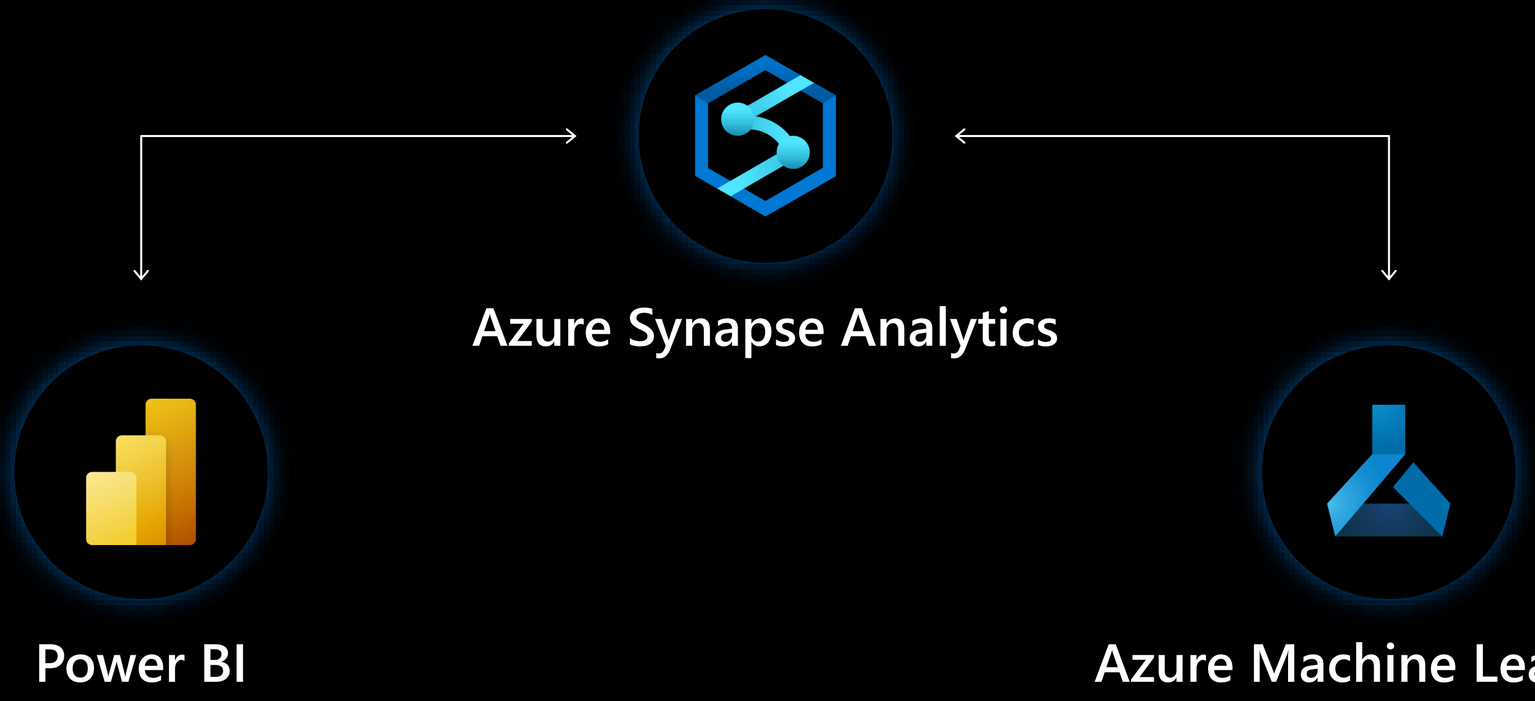


Where do you find yourself on the curve?



BI + Analytics unlock the door to AI, machine learning, and real-time insights





Unified experience to enrich data

Automated ML for rapid development

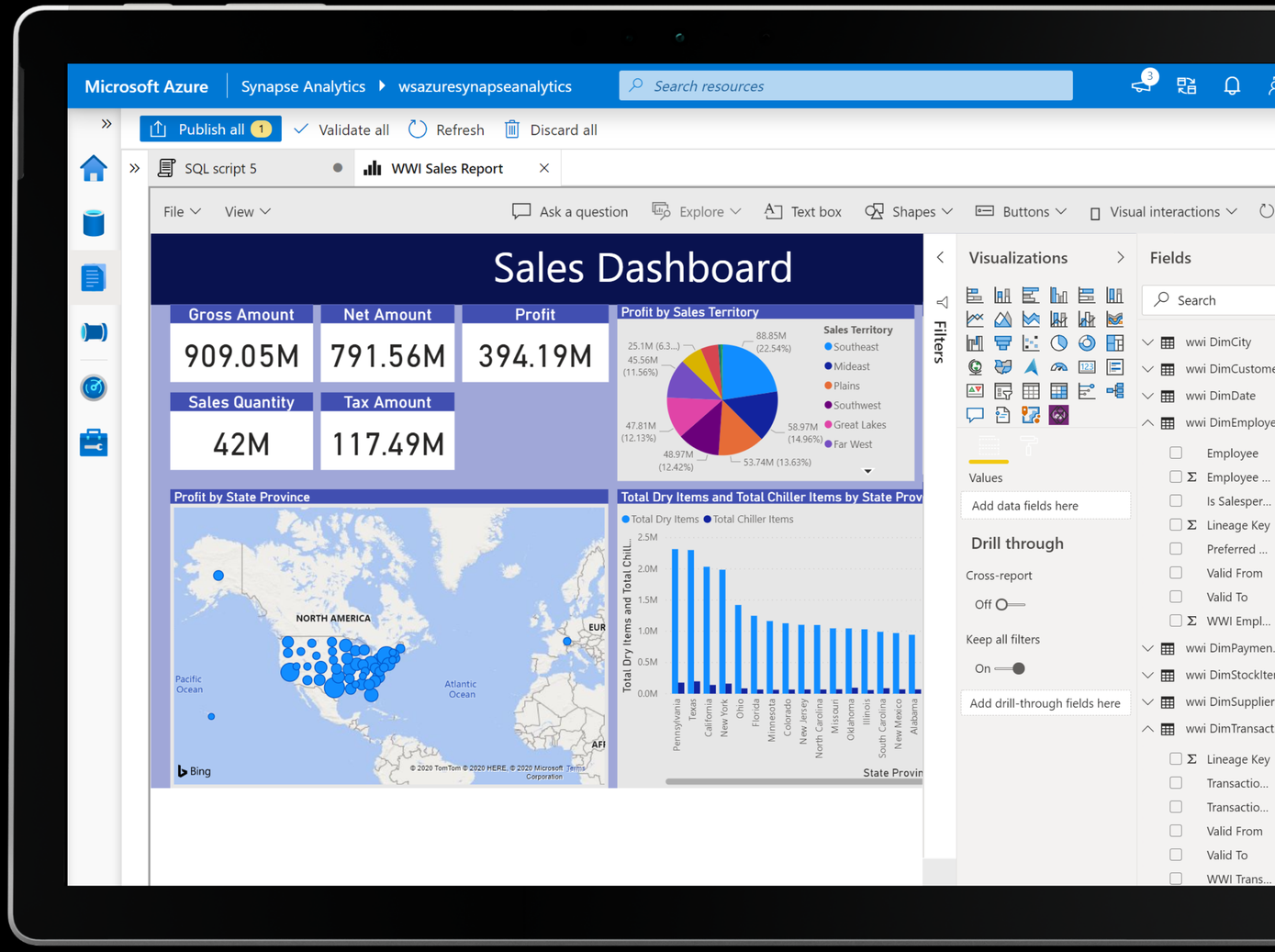
Seamless collaboration

Power BI integration

Build dashboard in Synapse Studio

Code-free experience for development rich visualizations

One-click publishing to for secure consumption across the enterprise



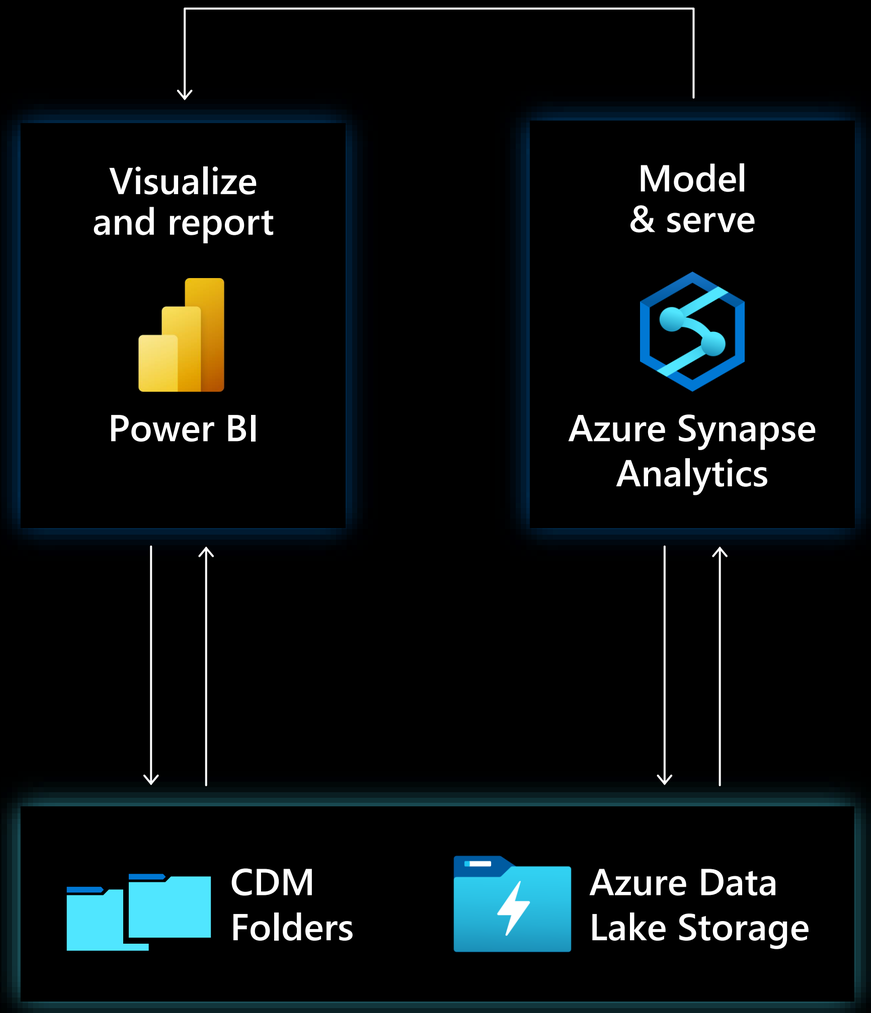
Accelerate business value with a powerful analytics platform

Respond instantly

Enable instant response times with Power BI Aggregations on massive datasets when querying at the aggregated level

Get granular with your data

Queries at the granular level are sent to Azure Synapse Analytics with DirectQuery leveraging its industry-leading performance



Save money with industry-leading performance

Azure Synapse Analytics is up to 14x faster and 94% cheaper than other cloud providers

View reports with a single pane of glass

Skip the configuration when connecting to Power BI with integrated Power BI-authoring directly in the Azure Synapse Studio

Customers using Azure Synapse & Power BI today are transforming their business with purpose

271% Average ROI

▲ **60%**
Increased customer satisfaction

▲ **27%**
Faster time to insights

▲ **26%**
Lower total cost of ownership

** Forrester, October 2019, "The Total Economic Impact of Microsoft Azure Analytics with Power BI"*

P&G

Understanding insights in the consumer goods market with Azure Synapse Analytics

American multinational consumer goods corporation

Challenge

- Looking to incorporate geographical dispersed data types for a single view for their business users
- As they are depending on data to inform their decisions at a more granular and at real time level, they could not afford for those data signals to have quality issues.

Solution

- Azure Synapse Analytics, Power BI, Azure Data Lake Storage
- In consumer goods, Procter & Gamble is now able to generate insights from over half a billion queries performed on petabytes of data and hundreds of data signals.
- Account managers, brand managers, supply chain planners and analysts can make faster and more accurate decisions, and tens of thousands of employees at the company can access the data to gain insights that serve consumers and build the business.

Azure Synapse regional availability

Australia Southeast	Korea Central
Australia East	North Central US
Brazil South	North Europe
Canada Central	South Africa North
Canada East	South Central US
Central India	Southeast Asia
Central US	Switzerland North
East Asia	UK West
East US	UK South
East US 2	West Central US
France Central	West Europe
Germany West Central	West US
Japan East	West US 2
Japan West	

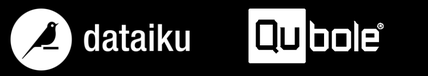
Partners

Business intelligence

Data integration

Data management

Machine learning and AI



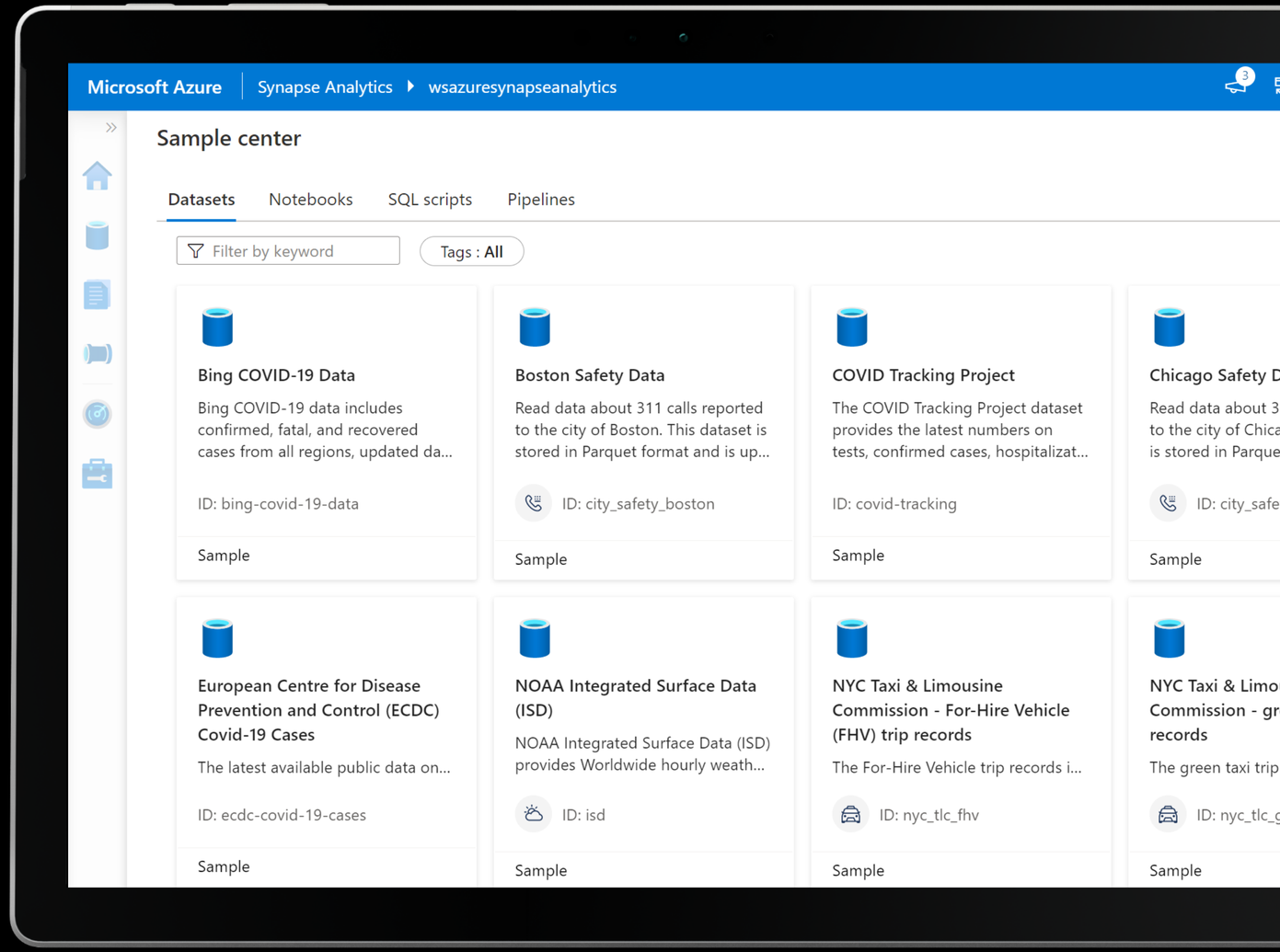
Knowledge Center

Accelerate time to solution

Azure Open Data sets

Pre-built samples to accelerate development

- SQL Scripts
- Notebooks
- Data Pipelines



CMC TECHNOLOGY & SOLUTION

Aspire to Inspire the Digital World

THANK YOU!