

Azure Synapse Analytics



Azure Synapse Analytics

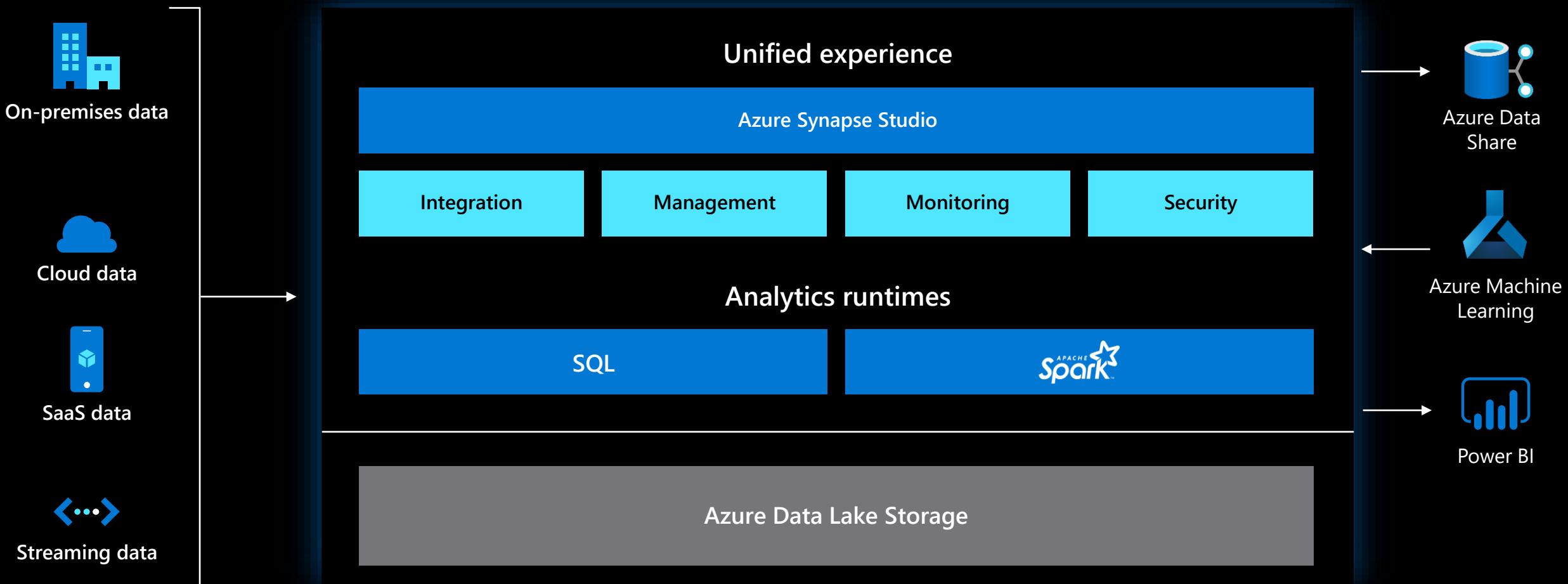
The first unified, cloud native platform for converged analytics



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.

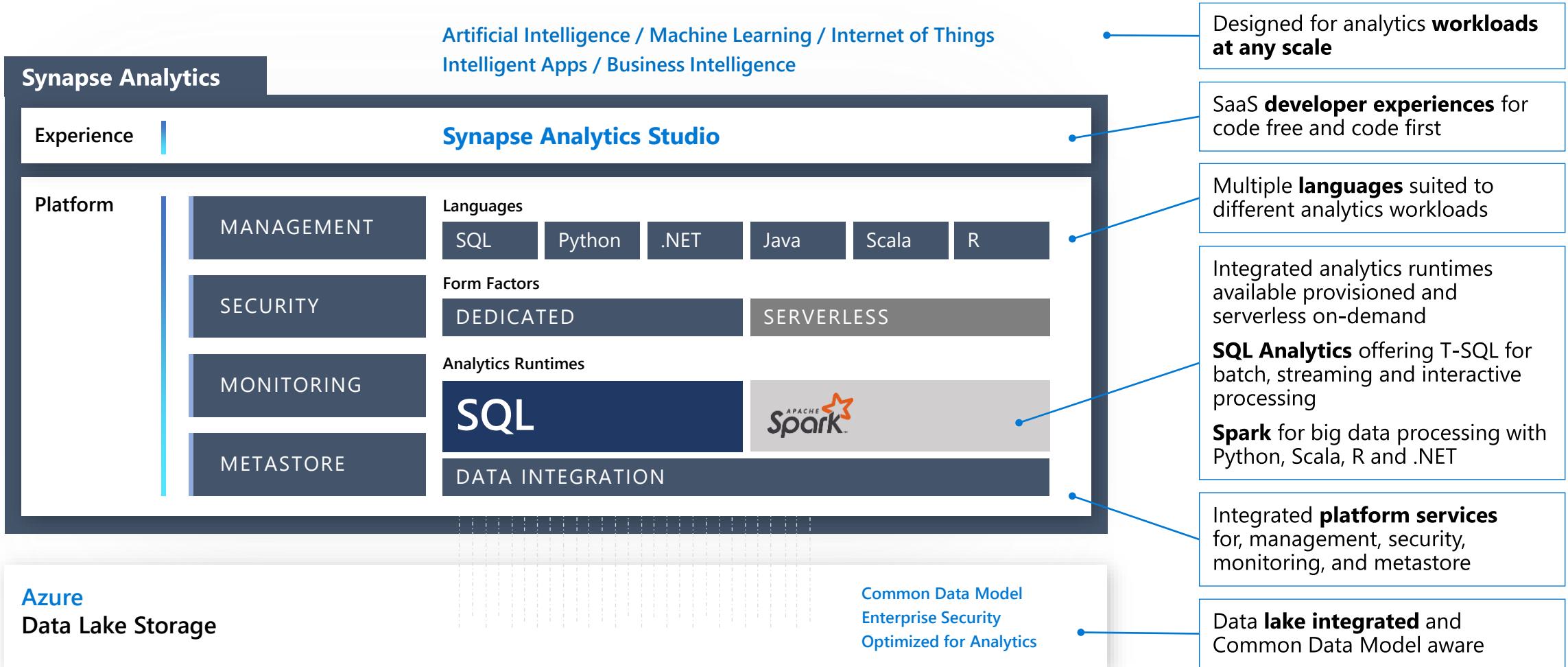
Azure Synapse

Limitless analytics service with unmatched time to insight



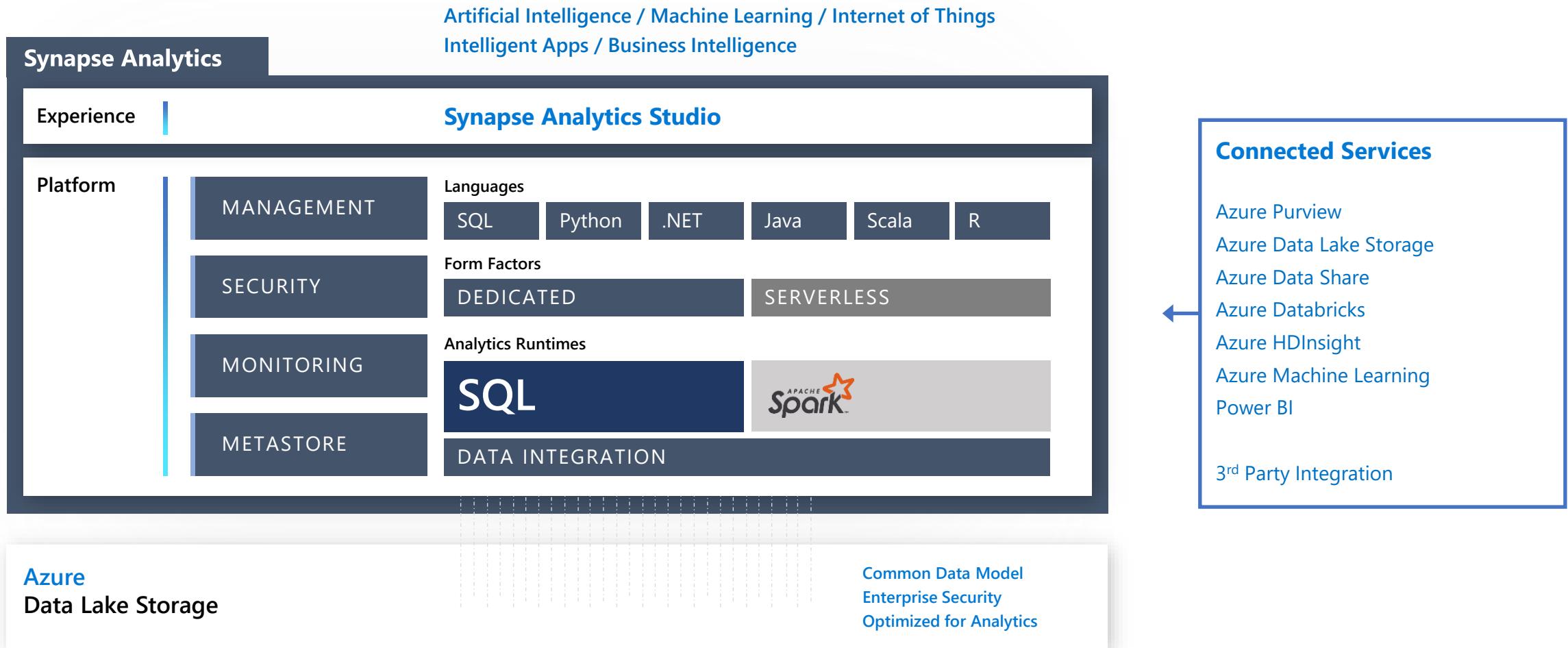
Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Azure Synapse Analytics

Built-in tools:

- Notebooks
- SQL Editor

Single pane of glass:

- Security
- Monitoring
- CI/CD
- Metastore
- Governance

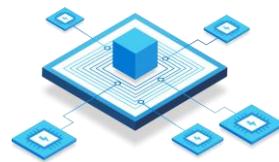


Workloads

- Data Warehousing
- Data Virtualization
- Logical Warehouses
- Data Integration
- Spark/ML
- Stream Analytics
- Business Intelligence



Synapse SQL



Synapse Spark



Synapse Pipelines



Synapse Studio

Data Warehouse

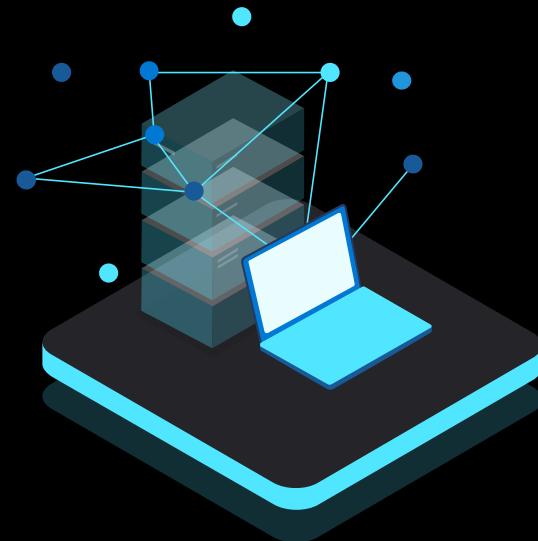
Scalable and secure SQL analytics platform

Dedicated + serverless 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

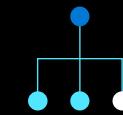
Suitable for any scale

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

The screenshot shows the Microsoft Azure Synapse Analytics Studio interface running on a laptop. The left side of the screen displays a sidebar with icons for Home, Data, Workspaces, and Linked services. The main workspace is titled "wsazuresynapseanalytics". On the right, there is a "Data" section showing a list of tables and columns from a schema named "wwi". A SQL script editor window is open, displaying the following T-SQL query:

```
1 SELECT TOP 10
2     City,
3     SUM(Quantity) AS Quantity
4 FROM
5     wwi.FactOrder f
6 INNER JOIN wwi.DimCity d ON d.CityKey = f.CityKey
7 WHERE StateProvince = 'Washington'
8 GROUP BY
9     City
10 ORDER BY
11     Quantity DESC
12
13
```

Below the script editor, there are tabs for "Results" and "Messages", with "Results" selected. A bar chart visualization is displayed, showing the quantity for various cities. The x-axis lists cities: Sekiu, Venersborg, Harbour Pointe, Lake Stevens, Koontzville, Malott, College Place, Upper Preston, Point Roberts, and Trentwood. The y-axis represents quantity, ranging from 0 to 20k. The bars show the following approximate values:

City	Quantity
Sekiu	~18k
Venersborg	~16k
Harbour Pointe	~16k
Lake Stevens	~16k
Koontzville	~15k
Malott	~14k
College Place	~13k
Upper Preston	~13k
Point Roberts	~13k
Trentwood	~12k

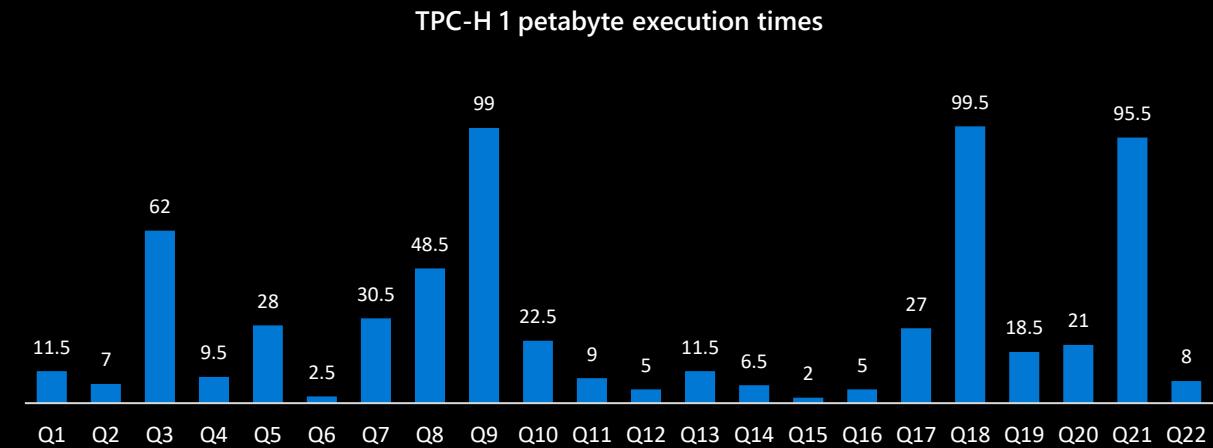
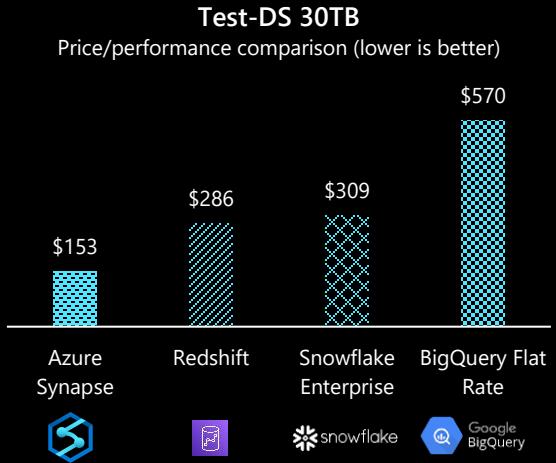
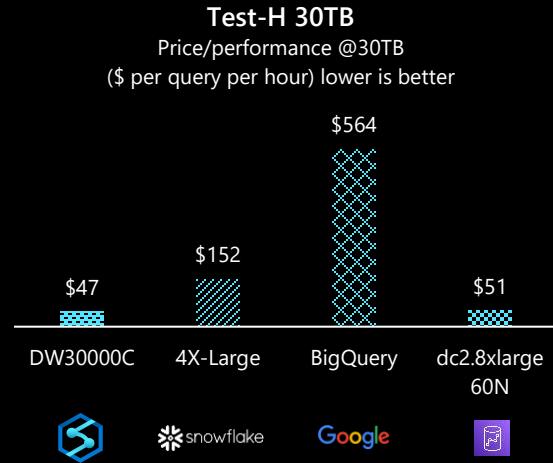
Industry Leading SQL Performance and Scalability

TPH-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>



Polaris Distributed Query Engine

Only platform to compete 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 Bockrocker, 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, Srikanth Rangarajan, Milan Ruzic, Milan Simic, Marko Sasic, Igor Stanko, Maja Stikic, Sasa Stanojkov, Vukasin Stefanovic, Milos Sukovic, Aleksandar Tomic, Dragan Tomic, Steve Toscano, Djordje Trifunovic, Veljko Vasic, Tomer Verona, Aleksandar Vujić, Nikola Vujić, 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.

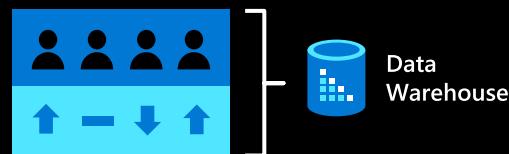
In this paper, we present the Polaris interactive relational query engine, a key component for converging warehouses and lakes in Azure Synapse [1], with a cloud-native scale-out architecture that makes novel contributions in the following areas:

- *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

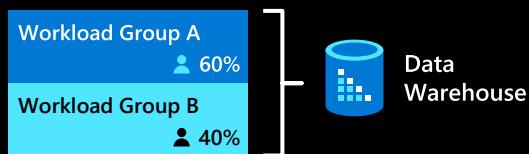
Workload Importance



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

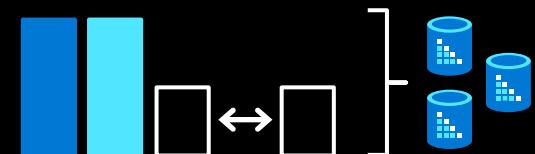


Benefits:

- 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

Multiple Cluster over shared data**



Benefits:

- Add cluster scale compute
- Full separation of workloads
- Chargeback clusters to internal business groups
- Multiple endpoints

NDA



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

Snowflake only supports elastic and multi-clusters, both which can quickly increase customer costs and limit customer control over their workloads.

** Feature in Private Preview



ABN Amro bank

**Chooses the most secure cloud DWH,
Azure Synapse Analytics, to transform its 2
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



Complete data protection

Best-in-class Security

Customer & System Managed Keys

All data encrypted by default

Up to 3x levels of data encryption at rest

Democratize data at scale with fine-grained ACL

Proactive protection

Comprehensive Compliance

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	✓
Authentication	Dynamic data masking	✓
	Column level encryption	✓
	SQL login	✓
Network Security	Azure active directory	✓
	Multi-factor authentication	✓
	Managed virtual network	✓
Threat protection	Custom virtual network	✓
	Firewall	✓
	Azure ExpressRoute	✓
Isolation	Azure Private Link	✓
	Threat detection	✓
	Auditing	✓
Isolation	Vulnerability assessment	✓
	Dedicated metadata store	✓
	Hosted in customer tenant	✓

Managed Virtual Networks

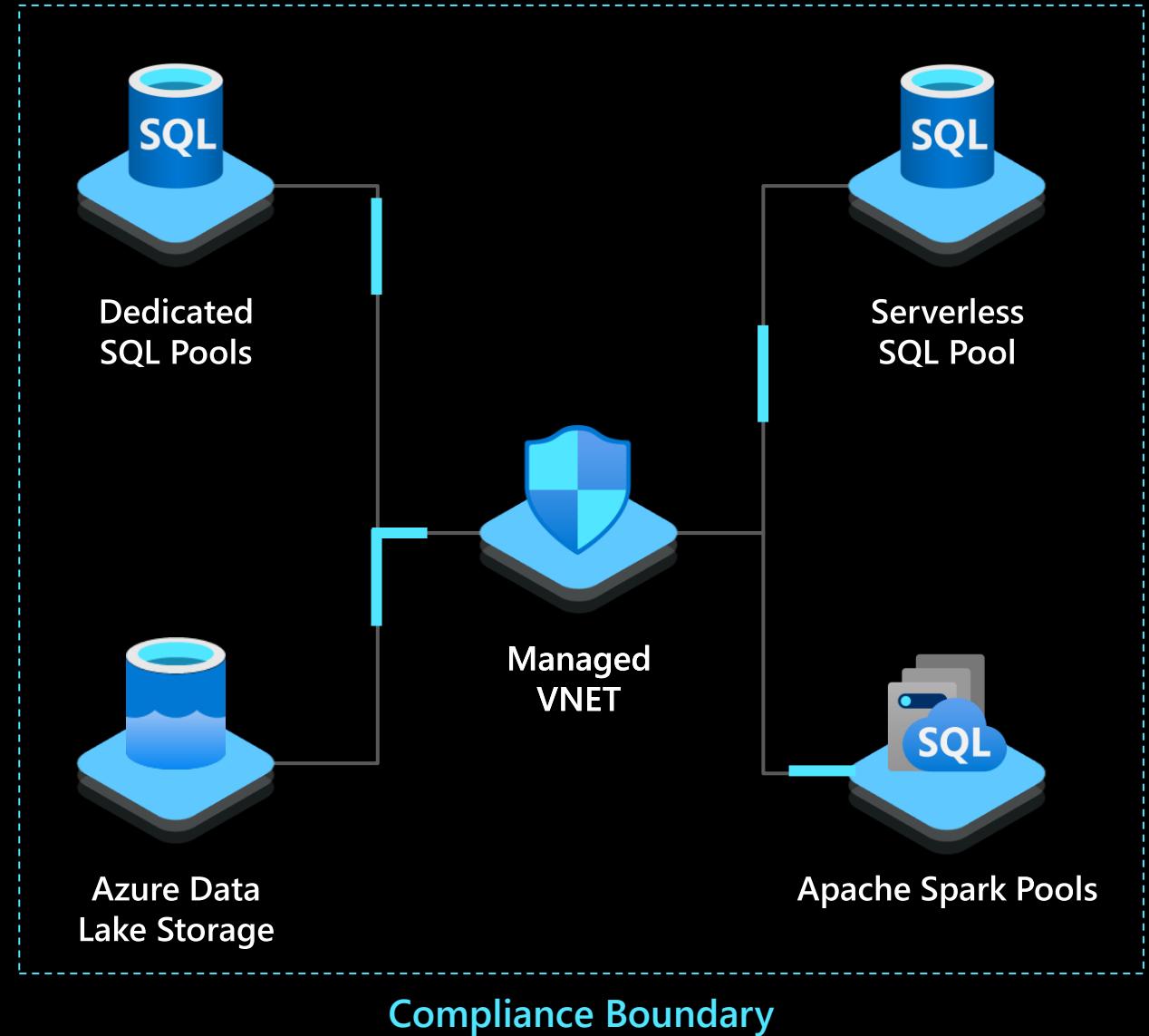
Eliminate network maintenance

One-click enables automated management of virtual networks between cluster endpoints

Synapse resources only ever interop with private endpoints

No management of subnets or IP Ranges

Prevents data exfiltration



Integrated Data Governance

More than just data security

Native integration with Azure Purview

Automatically discover and classify data assets

End-to-end data lineage





Grab

Reducing time to insights with Azure Synapse Analytics

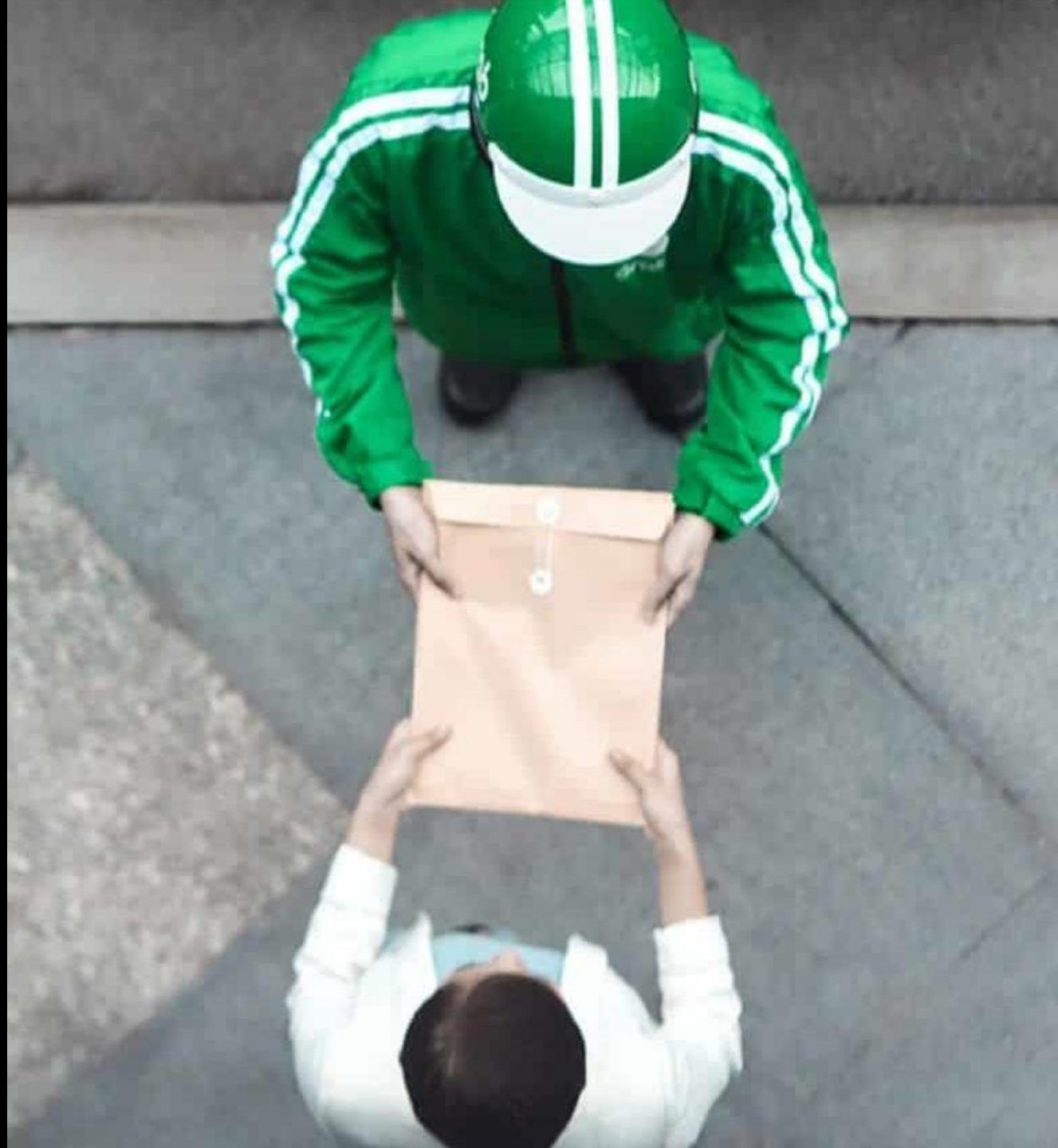
Singaporean multinational ride-hailing, food delivery and digital payments services company

CHALLENGE

- Required ability to dynamically experiment and build very light weight POV to determine viability of projects to unlock new business opportunities
- Dependency on a solution that could then scale to address the large customer base across 8 countries
- Critical data security and sovereignty requirements to function in those markets

SOLUTION

- Azure Data Lake Storage, Azure Synapse Analytics, Azure Databricks
- Grab was able to achieve the required data governance through Synapse with a data warehouse that provides the central source of truth for reporting and analysis to their data scientists, analysts and data communities.



Real-time Operational Analytics

Eliminate latency and accelerate decision making

Cloud Native HTAP with Synapse Link

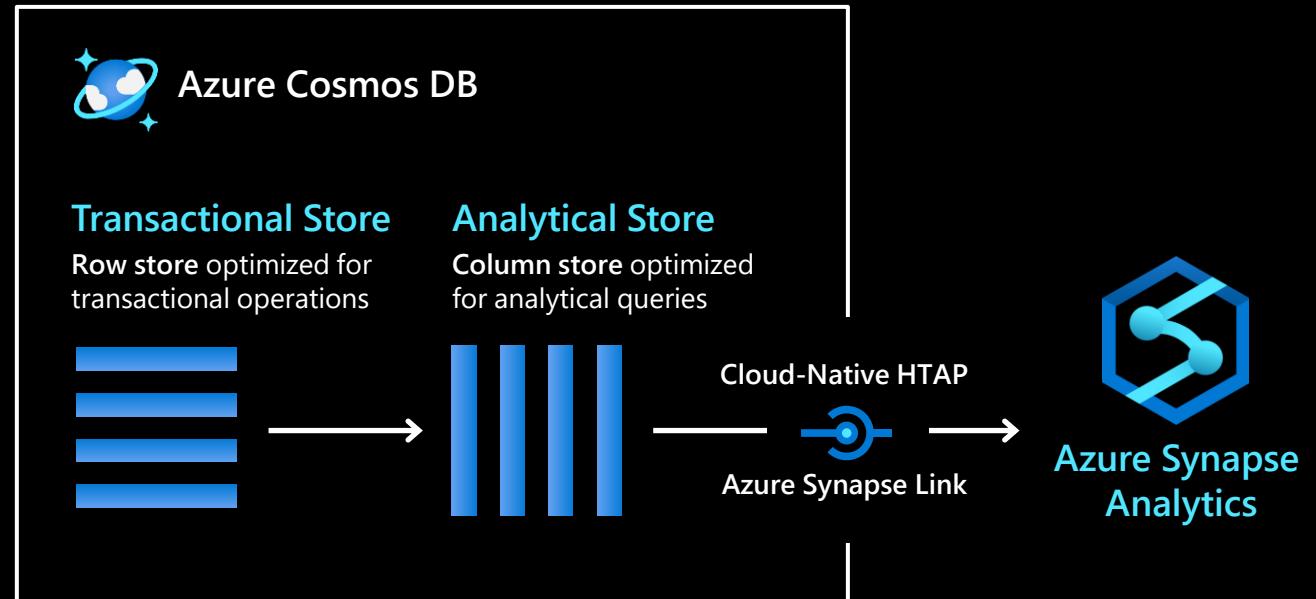
Real-time operational analytics

One-click enablement in Azure Portal

No data integration pipelines required

Near-zero impact on operational systems

Latency <90s at 99th percentile



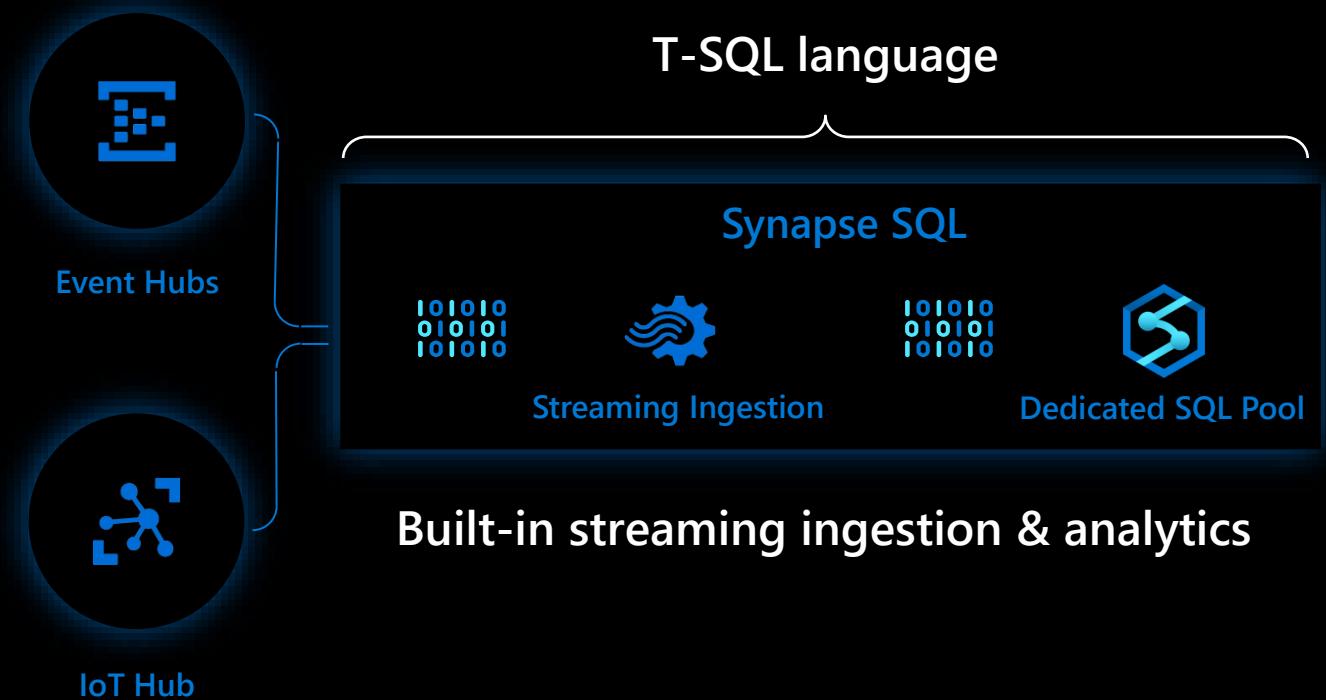
Streaming Event Ingestion

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 DataLake and [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 data science to all

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



All Code

Notebook IDE
PySpark/Scala



Low/No-Code

Classification
Regression
Time-series



Pre-built models

Anomaly Detector
Sentiment Analysis



Notebook IDE Code Authoring

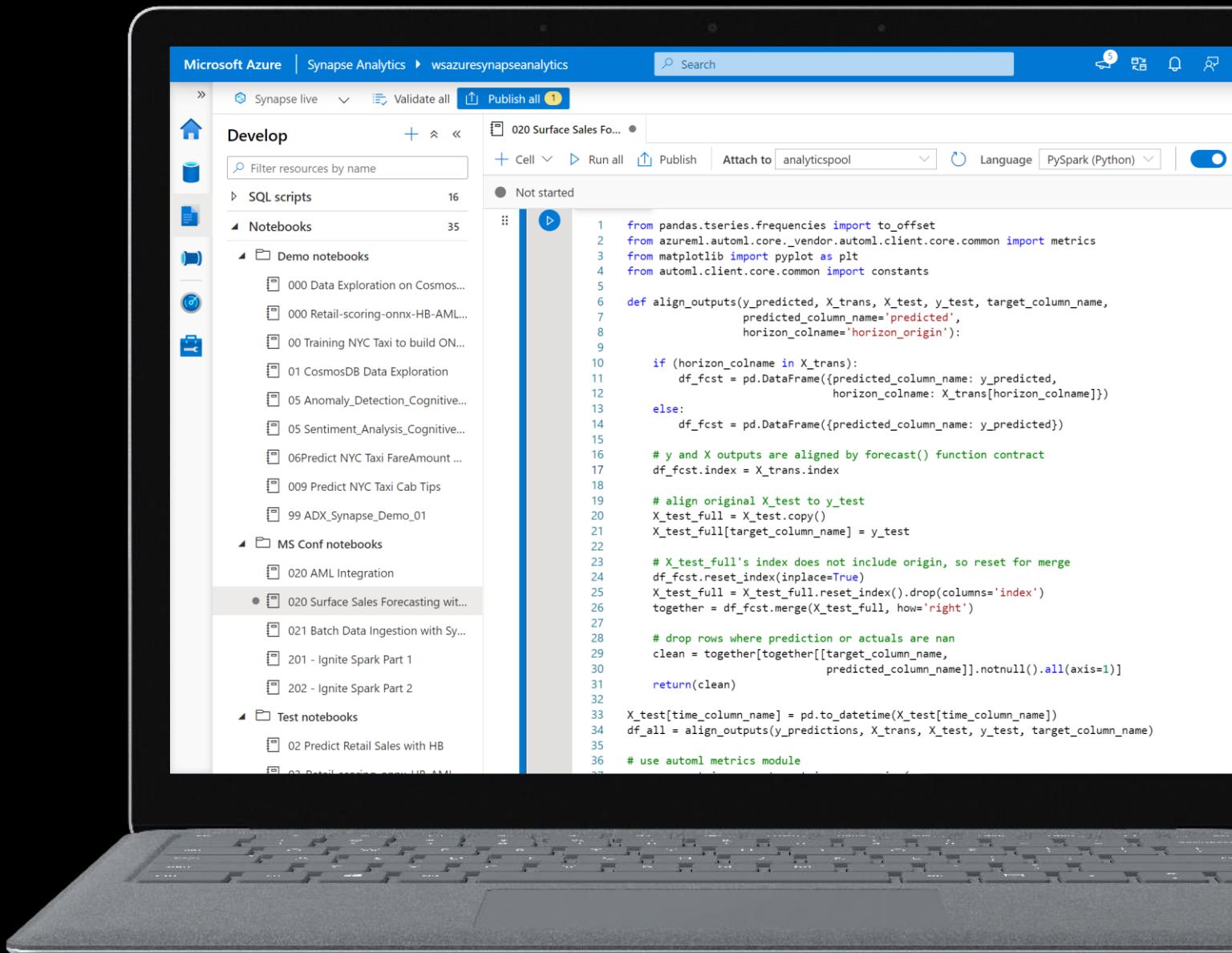
Code-first ML model development

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



The screenshot shows the Microsoft Azure Synapse Analytics Notebook IDE interface. The top navigation bar includes 'Microsoft Azure', 'Synapse Analytics', and 'wsazuresynapseanalytics'. The left sidebar has sections for 'Develop' (SQL scripts, Notebooks), 'Data', 'Machine Learning', and 'Analytics'. The main area shows a notebook titled '020 Surface Sales Fo...'. The code editor displays a Python script for surface sales forecasting:

```
from pandas.tseries.frequencies import to_offset
from azureml.core._vendor.azureml.client.core.common import metrics
from matplotlib import pyplot as plt
from automl.core.common import constants

def align_outputs(y_predicted, X_trans, X_test, y_test, target_column_name,
                  predicted_column_name='predicted',
                  horizon_colname='horizon_origin'):

    if (horizon_colname in X_trans):
        df_fcst = pd.DataFrame({predicted_column_name: y_predicted,
                               horizon_colname: X_trans[horizon_colname]})

    else:
        df_fcst = pd.DataFrame({predicted_column_name: y_predicted})

    # y and X outputs are aligned by forecast() function contract
    df_fcst.index = X_trans.index

    # align original X_test to y_test
    X_test_full = X_test.copy()
    X_test_full[target_column_name] = y_test

    # X_test_full's index does not include origin, so reset for merge
    df_fcst.reset_index(inplace=True)
    X_test_full = X_test_full.reset_index().drop(columns='index')
    together = df_fcst.merge(X_test_full, how='right')

    # drop rows where prediction or actuals are nan
    clean = together[[target_column_name,
                      predicted_column_name]].notnull().all(axis=1)

    return(clean)

X_test[time_column_name] = pd.to_datetime(X_test[time_column_name])
df_all = align_outputs(y_predictions, X_trans, X_test, y_test, target_column_name)

# use automl metrics module
```

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 an unlimited set of file formats including Parquet, ORC and JSON as well as audio, image, and video formats

Language



C#

SQL

Data



.ORC

< , >

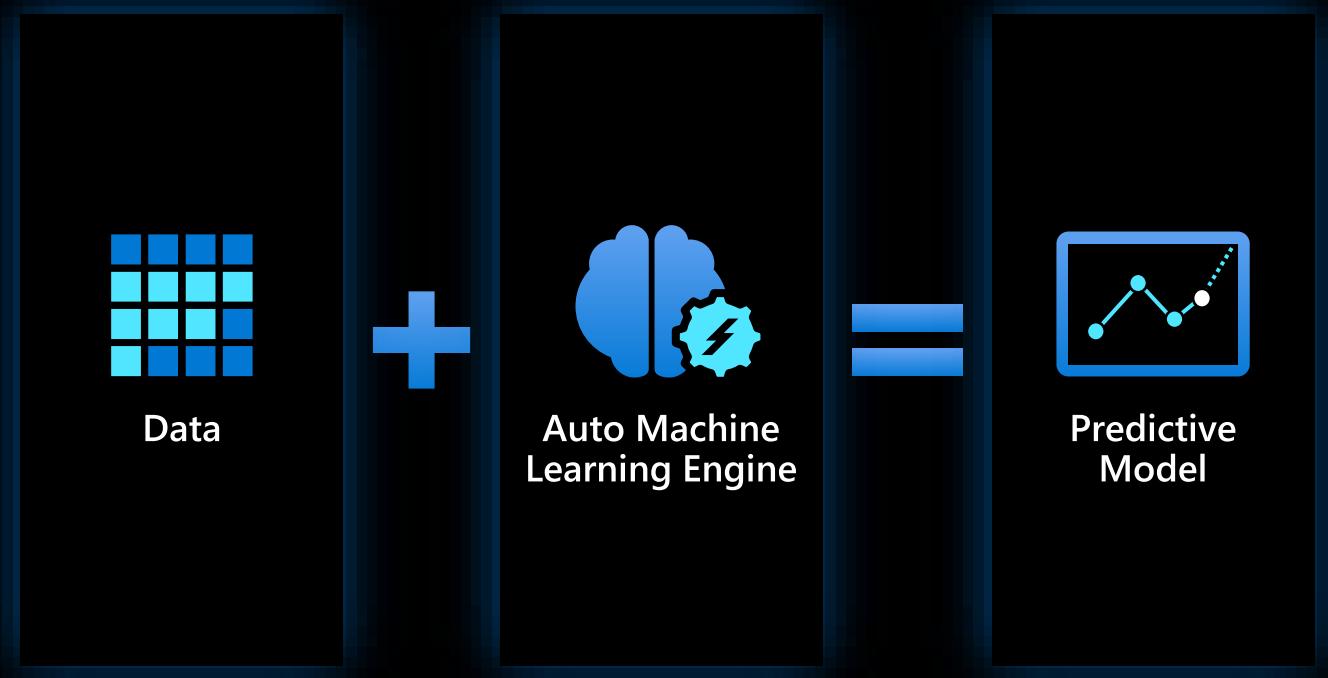
{JSON}



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

The screenshot shows a Microsoft Azure Synapse Analytics workspace titled 'wsazuresynapseanalytics'. On the left, the 'Data' section lists various databases and tables, including 'newpoll' (SQL), 'NYCTaxi_Pool' (SQL), 'Predict_Pool' (SQL), 'Streaming_Pool' (SQL), 'WWI_Pool' (SQL), 'NYT2020' (SQL), 'SQLServerlessDB' (SQL), 'default' (Spark), 'retaildata' (Spark), 'retailsales' (Table), and 'surfacesalesdb' (Spark). The main area displays a Python script for machine learning:

```
from pandas import *
from azureml import *
from matplotlib import *
from automl import *
def align_(df):
    if (horizon == 'short-term'):
        df = df[['X', 'y']]
    else:
        df = df[['X', 'y']]
    # align X_train, X_test, X_test together
    # drop clean :
    return df
X_test[time] = df_all['time'].max()
# use autor
```

To the right, there's a sidebar titled 'Choose a model type' with three options: 'Classification' (with a bar chart icon), 'Regression' (with a stack of bars icon), and 'Time series forecasting' (with a circular arrow icon). Each option includes a brief description and an example.

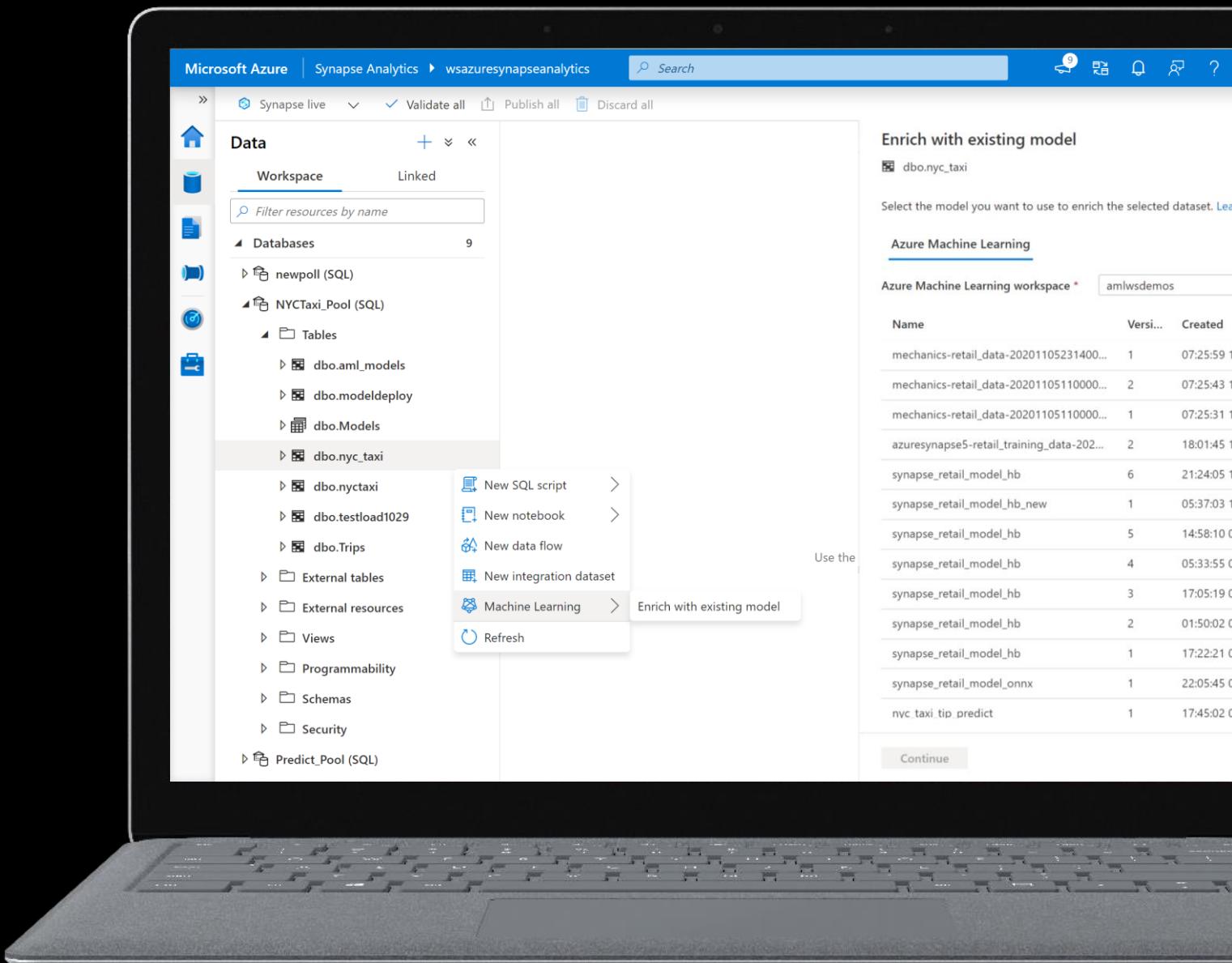
Code-free Machine Learning Scoring

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



+ a b | e a u

Qlik

MicroStrategy

SISIENSE

```
SELECT d.* , p.Score FROM PREDICT(MODEL = @onnx_model, ...)
```

Synapse SQL



Model



Predictions

databricks

MathWorks

learn

Sas

mxnet

PaddlePaddle

Chainer

Caffe2

ML.NET

K

Spark

XGBoost

PyTorch

F

M



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, Azure Databricks, Azure Data Lake Gen2, Azure Data Factory, Logic Apps, Azure Active Directory
- 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."



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



95 connectors



All your data
(on-prem,
SaaS applications,
other clouds)

Code-free

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

The screenshot shows the Microsoft Azure Synapse Analytics Power Query Editor interface. The top navigation bar includes 'Microsoft Azure', 'Synapse Analytics', 'wsazuresynapseanalytics', and a 'Search resources' bar. Below the navigation is a toolbar with various icons for data operations like 'Validate all', 'Publish all', and 'Discard all'. The main workspace is titled 'PQSalesPrep' under the 'Settings' tab. A message states: 'Currently not all Power Query M functions are supported for data wrangling despite being available during authoring. Learn more'. The interface features a ribbon with 'Home', 'Transform', 'Add column', and 'View' tabs. The 'Transform' tab is selected, showing a set of tools for data manipulation. On the left, there's a 'Queries' pane listing 'ADFResource' and 'UserQuery'. The main area displays a table with 17 rows and 8 columns, showing data from 'surface.go' with columns including 'ab storeId', 'ab productCode', '12 quantity', '1.2 logQuantity', 'ab advertising', 'ab price', 'ab weekStarting', and 'ab id'. The table has a header row and several data rows. At the bottom of the table, it says 'Columns: 8 Rows: 99+'. The bottom of the screen shows a portion of a laptop keyboard.

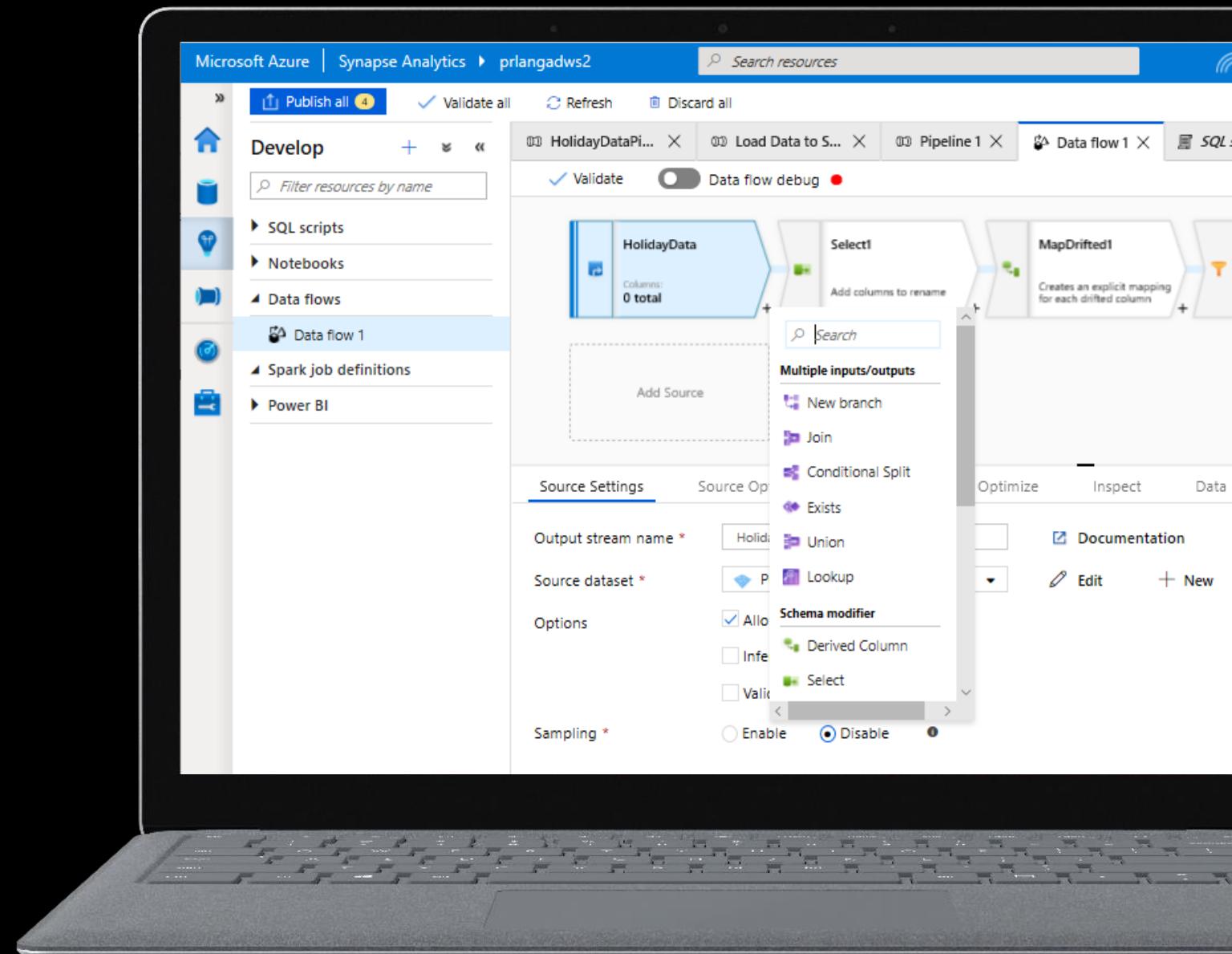
Hybrid Data Integration

Real-time operational analytics

No data integration pipelines required

Near-zero impact on operational systems

Latency <90s at 99th percentile





RHI MAGNESITA

RHI Magnesita

Azure Synapse Analytics empowering AI & IoT

The global leader in refractories with the largest number of locations around the world

CHALLENGE

- As an outcome of the merger of the former separate companies RHI and Magnesita, the ERP data from two different SAP systems needed to be consolidated in one single point of truth.
- On-prem DWH reached its limits in size and ability to support cloud based digital use cases focusing on ingestion of streaming data from the edge and machine learning capabilities.

SOLUTION

- Consolidation in [Azure Synapse Analytics](#) using Spark and SQL engines build the foundation for a cloud native data platform leveraged by next-gen applications and transforming ideas into solutions / services to generate business benefit
- 5-10x better performance for business by utilizing scaling and better ease of use
- Enabling multiple cloud native digital business use cases by providing a single point of truth for data
- Data driven customer facing business applications generate direct benefit for RHI Magnesita's customers in the steel and industrial segments
- RHI Magnesita is now proactively alerting their customers when vessels need to be changed or if measures need to be taken to prolong its lifetime



Knowledge Center

Accelerate time to solution

Azure Open Data sets

Pre-built samples to accelerate development

- SQL Scripts
- Notebooks
- Data Pipelines

The screenshot shows a Microsoft Azure Synapse Analytics interface titled "Sample center". The top navigation bar includes "Microsoft Azure", "Synapse Analytics", and a workspace name "wsazuresynapseanalytics". Below the navigation is a breadcrumb trail "Datasets > Sample center". The main content area has tabs for "Datasets", "Notebooks", "SQL scripts", and "Pipelines", with "Datasets" selected. A search bar "Filter by keyword" and a tag filter "Tags : All" are present. The page displays a grid of sample datasets:

Dataset	Description	ID	Action
Bing COVID-19 Data	Bing COVID-19 data includes confirmed, fatal, and recovered cases from all regions, updated daily.	ID: bing-covid-19-data	Sample
Boston Safety Data	Read data about 311 calls reported to the city of Boston. This dataset is stored in Parquet format and is updated daily.	ID: city_safety_boston	Sample
COVID Tracking Project	The COVID Tracking Project dataset provides the latest numbers on tests, confirmed cases, hospitalizations, and deaths.	ID: covid-tracking	Sample
Chicago Safety Data	Read data about 311 calls reported to the city of Chicago. This dataset is stored in Parquet format and is updated daily.	ID: city_safe...	Sample
European Centre for Disease Prevention and Control (ECDC) Covid-19 Cases	The latest available public data on COVID-19 cases from ECDC.	ID: ecdc-covid-19-cases	Sample
NOAA Integrated Surface Data (ISD)	NOAA Integrated Surface Data (ISD) provides Worldwide hourly weather observations.	ID: isd	Sample
NYC Taxi & Limousine Commission - For-Hire Vehicle (FHV) trip records	The For-Hire Vehicle trip records from the NYC TLC.	ID: nyc_tlc_fhv	Sample
NYC Taxi & Limousine Commission - green taxi trip records	The green taxi trip records from the NYC TLC.	ID: nyc_tlc_g...	Sample

Power BI Authoring

Build dashboard in Synapse Studio

Code-free experience for development rich visualizations

One-click publishing to for secure consumption across the enterprise

The screenshot shows a Microsoft Azure Synapse Analytics workspace titled "wsazuresynapseanalytics". The main area displays a "WWI Sales Report" dashboard titled "Sales Dashboard". The dashboard includes several key figures: Gross Amount (909.05M), Net Amount (791.56M), Profit (394.19M), Sales Quantity (42M), and Tax Amount (117.49M). Below these, there is a pie chart titled "Profit by Sales Territory" showing distribution across Southeast, Midwest, Plains, Southwest, Great Lakes, Far West, and Northeast regions. A map titled "Profit by State Province" shows data points across North America and Europe. To the right, a bar chart titled "Total Dry Items and Total Chiller Items by State Prov" compares dry items (blue bars) and chiller items (black bars) across various US states. The interface features a left sidebar with navigation icons and a top bar with "Microsoft Azure", "Synapse Analytics", and "wsazuresynapseanalytics" along with "Search resources" and notification icons. The right side of the screen shows a "Visualizations" pane with a grid of icons for different chart types and a "Fields" pane listing various data dimensions and measures.



Proctor & Gamble

Understanding insights in the consumer goods market with 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
- 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.



Globally Available

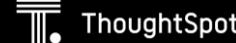
Australia Southeast	Korea Central
Australia East	North Central US
Brazil South	North Europe
Canada Central	South Africa North
Canada East	South Central US
Central India	Southeast Asia
Central US	Switzerland North
East Asia	UK West
East US	UK South
East US 2	West Central US
France Central	West Europe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



Partners

Business intelligence

AT SCALE



Data integration

ab initio

alteryx

attunity

cdata

datacoral.

denodo

DIMODELO solutions

Fivetran

HVR

incorta

Informatica

Information Builders

flytron

MATILLION



precisely

RIVERY

Segment



snapLogic

StreamSets

talend

TIMEXTENDER

TRIFACTA

WhereScape®

Xplenty

Data management

aginity

Alation

Collibra

DATOMETRY

erwin

InfoLibrarian CORPORATION

MicroERD

REDPOINT GLOBAL



tamr

telera.n

Machine learning and AI

dataiku

Qubole

DOMO

MathWorks

Sas

Questions

