

# 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



### 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





# 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

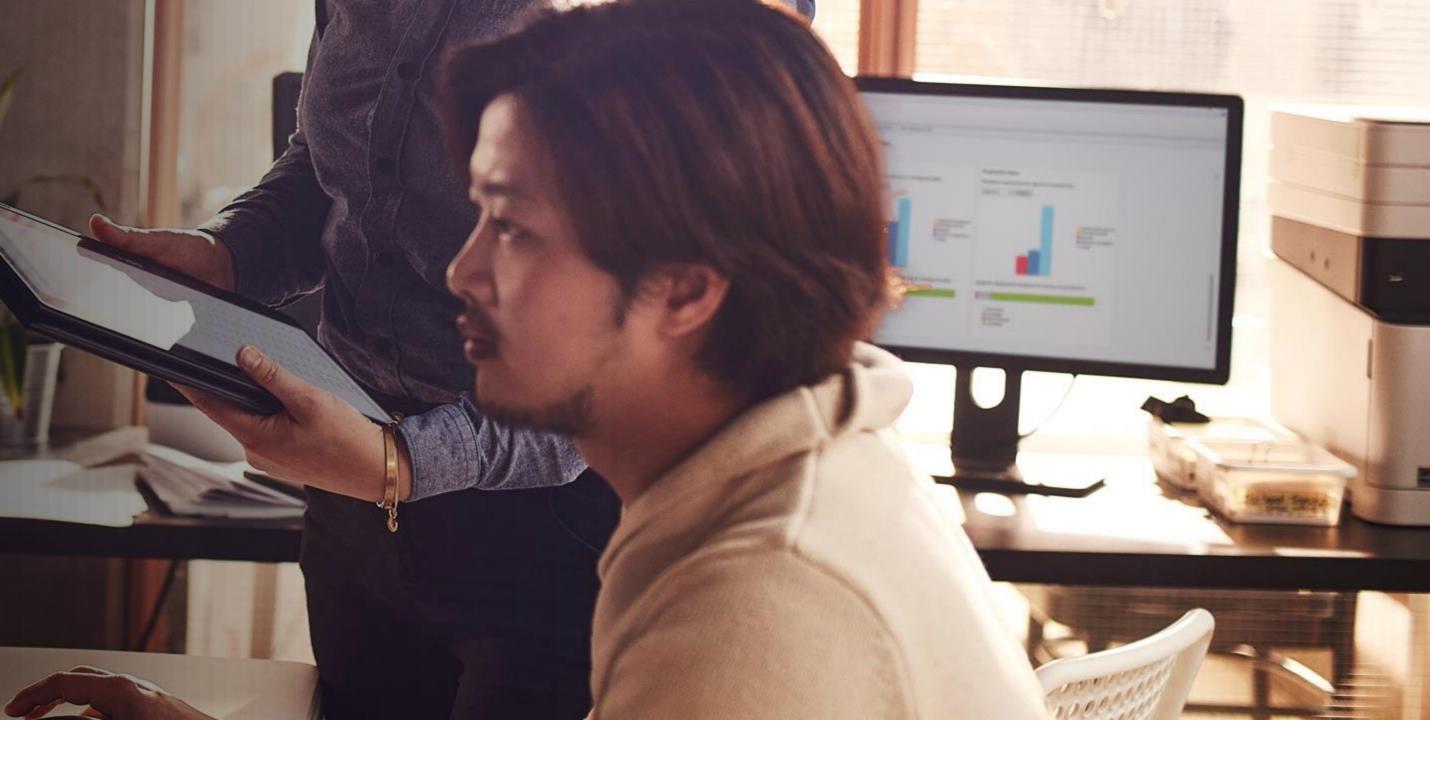
**Augmented intelligence** sees patterns we cannot

### 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

organizations



# **AI/ML creates smarter**

#### **Businesses need** scalable analytics



# 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 transforming



### **Collaboration platforms** and file-sharing are

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

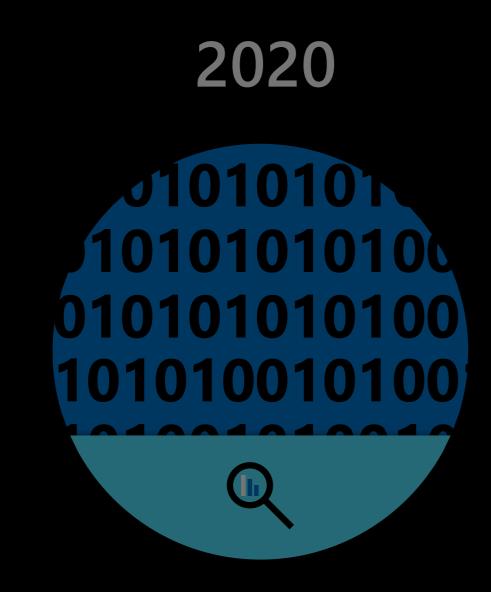


# Section 3 The paradox of analytics



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

**4**x

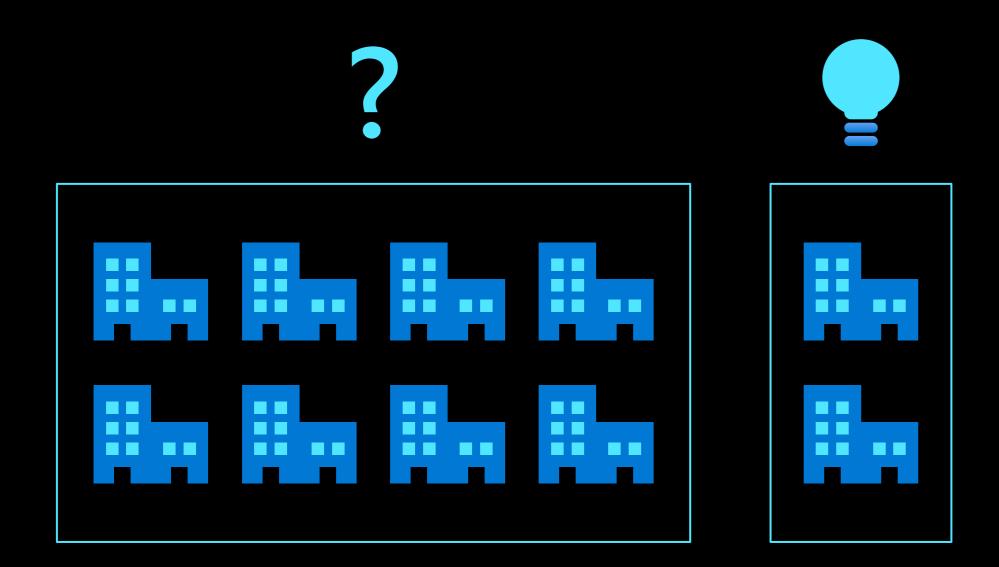


### Data collected worldwide

2025

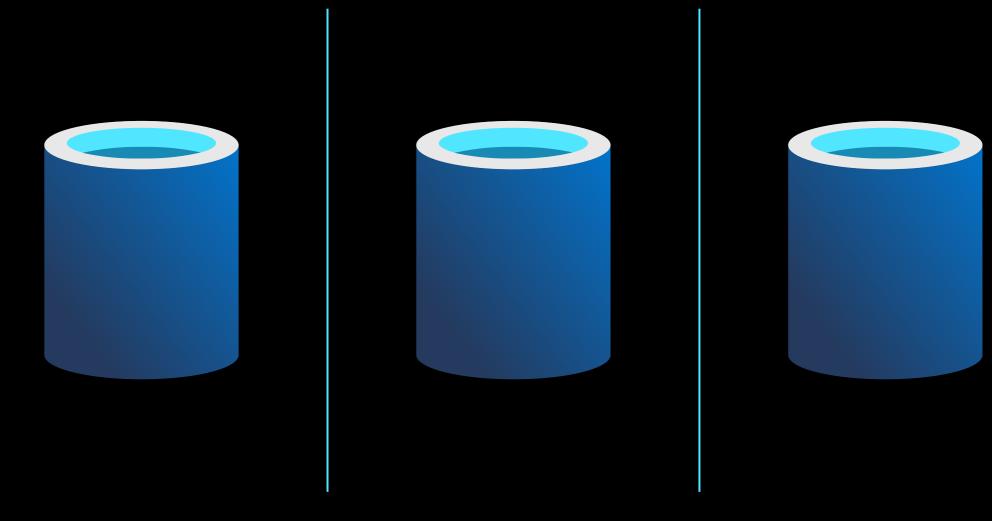
# **1752B** Data collected worldwide

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



# **BOOM** report struggling to become mature users of data\*

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



#### report data silos and data 550/ management difficulties as roadblocks\*

## More analytics solutions lead to more silos



#### Data

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



#### Technologies

- Map Reduce Open source projects Data mart Data warehouse
  - Data lake
  - Spreadsheets
    - RDMS
- Data visualization tools
  - ML models
  - Al 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

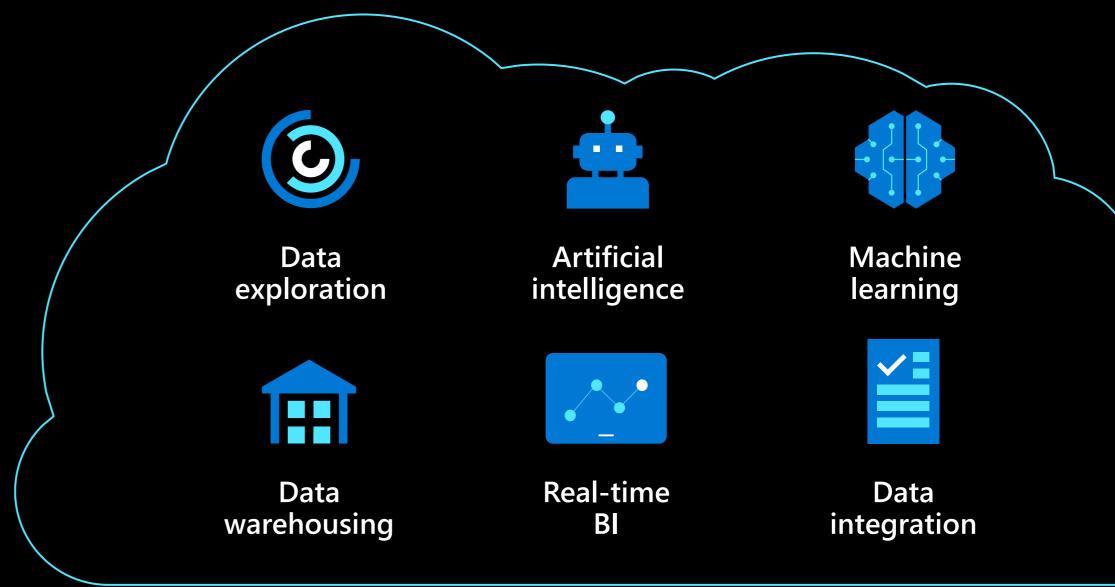


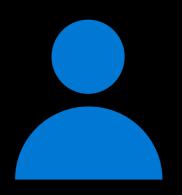


# **Section 4** The analytics continuum



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



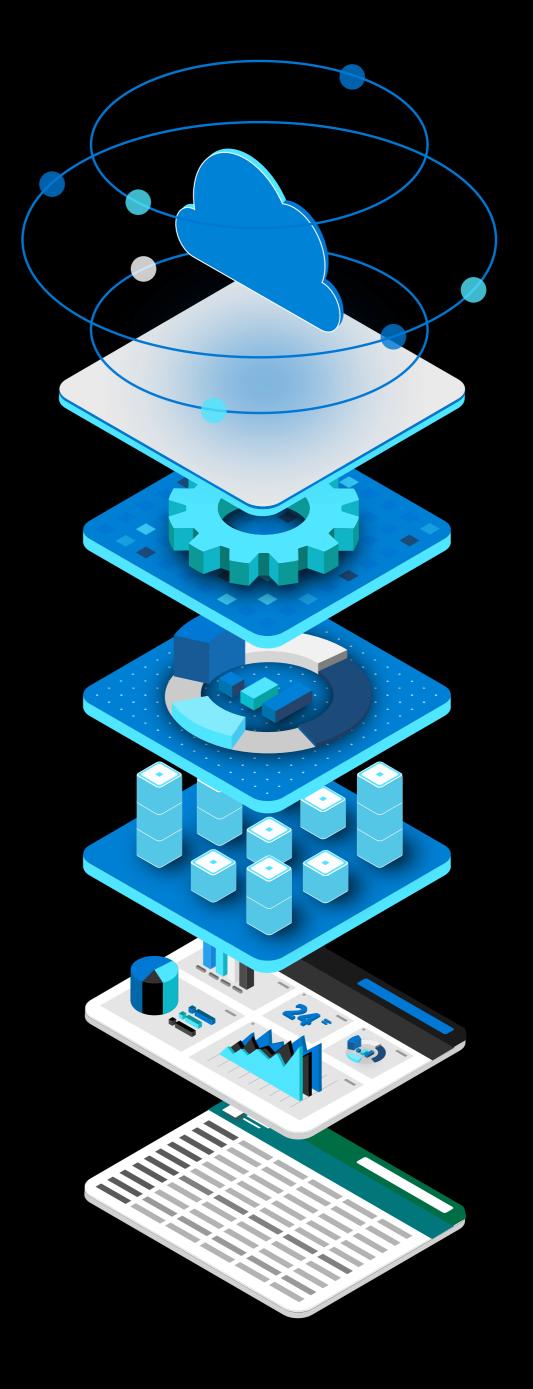




Analytics should seamlessly be part of the way users work rather than a factor for creating more siloes



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





## Azure Synapse Analytics Limitless analytics service with unmatched time to insight

••••

•

٠

•

٠

•

•

•

۲

•

•

•

•

•

۲

•

٠

•

•

۲

٠

۲

۲

٠

•

٠

۲

٠

۲

•

٠

•

•

۲

۲

•

۲

٠

۲

۲

۲

۲

۲

۲

۲

•

۲

•

۲

•

۲

۲

۲

٠

•

۲

۲

••••

•

•

•

۲

•••

•••

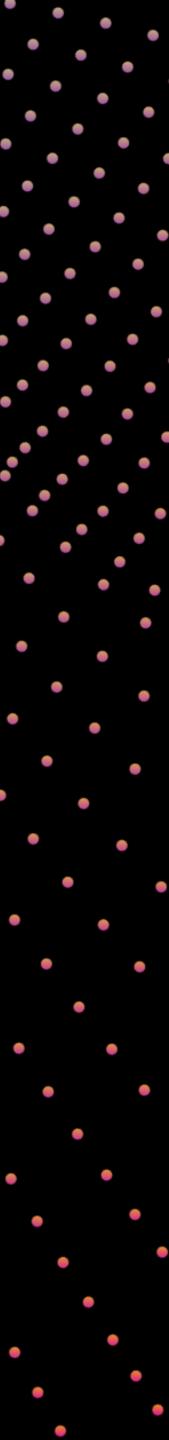
• •

• •

••••

•

•



# 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.

#### **Data integration**

#### Data warehousing

#### **Big data analytics**

# Azure Synapse Analytics

Powered by a new cloud native distributed SQL engine

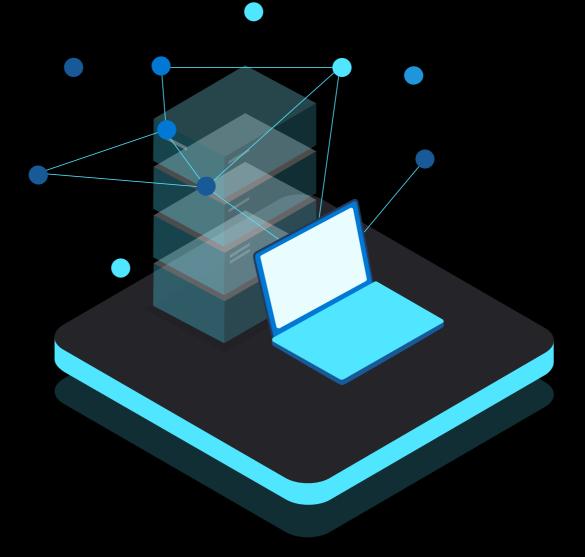


## Serverless + dedicated SQL

#### Flexible consumption models

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

Dedicated clusters optimized mission-critical data warehouse workloads









## 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**



# Code-free data wrangling

### No/low-code data transformation

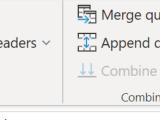
Excel-like interface is familiar to users

Transform data to desired shape completely visually

Operationalize into pipelines

										• •							
Micro	osof	t Azure	Synapse	Analytics	ws	sazures	ynapse	analytics	×	Search resource	res						Ş,
»		🗸 Validate	all ሰ I	Publish all	D 🗊	iscard a	1										
	<b>»</b>	PQSale	esPrep	×													
_		🔅 Setti	ings														
-		1 Curr	rently not a	all Power Que	ery M	functio	ns are s	supported for a	lata v	wrangling despite	being avail	able	during aut	thoring.	Learn m	nore 🖸 🛛 🛛	< compared with the second sec
		Home Transform Add column View															
		Enter data	Options	Manage parameters		Refresh	L Ad	operties vanced editor anage ~	Cł	noose Remove	Keep F rows ~			Sp colur		up 🖽 Use	a type: Text ~ first row as he lace values
		New query	Options	Parameters			Que			Manage columns	Reduce		Sort		(#2	Transfor	
-		Queries <								umnTypes(Source							
		Þ 📄 AI	DFResource	[1]	1	ab stor	reld 💌	ab productCode surface.go		12 quantity 1.2 105		0.265		sing 💌	ab price	ab week 6/15/201	CStarting 🗾 a
		U 🖽	serQuery			2		surface.go		80		3.987			269	7/27/201	
						2		surface.go		68		3.832			209	8/3/2017	
		Expand res	sources pan	ne	4	2		surface.go		28		7.966			209	8/10/201	
					5	2		surface.go		16	7	7.378	0		209	8/24/201	7 a
					6	2		surface.go		253	1	10.14	1		189	8/31/201	7 3
					7	2		surface.go		107	9	9.283	0		189	9/7/2017	/ b
					8	2		surface.go		66	8	8.803	0		189	9/14/201	7 e
					9	2		surface.go		65	8	3.794	0		179	9/21/201	7 с
					10	2		surface.go		17	7	7.455	0		269	10/12/20	)17 f4
					11	2		surface.go		337	10	).428	1		124	10/19/20	)17 5
					12	2		surface.go		19		7.56	0		159	10/26/20	017 0
					13	2		surface.go		89	g	9.101	1		159	11/2/201	7 d
					14	2		surface.go		113	g	9.341	0		129	11/9/201	7 9
					15	2		surface.go		284	10	).255	0		99	11/16/20	017 6
					16			surface.go		171		9.75	1		159	11/23/20	)17 с
					17	2		surface.do		265	10	).187	1		249	11/30/20	017 6
		Columns:	8 Rows: 99	)+	17	2		surface.ɑo		265	10	).187	1		249	11/30/20	)17

# 



#### ab id

16bd47a7-2ad6-4f0a-b8de 54cc74c2-c7da-4e12-af64a2d164b-5e44-44d7-9837 8cd9987-1d5a-4f4f-9346c0ec099-e102-4bfc-9775d22c002-b04c-4092-abcb b6e19699-d684-449e-9c98 89a5838-fb8f-413a-a23d-:3278682-16c0-4832-b76f 40190c1-b2ed-46f4-a5e0-294a983-5044-4068-b925 )a3f3796-07a3-450f-a6bcl0eab000-a45e-474f-864e-932a3bff-2f5f-4e8b-bc4d-c 558edd84-5873-4b37-a078 :1dc14fb-29a5-490a-9191 fb16536-dd0c-49cc-a7a6-

Complete

# Hybrid data integration

## Real-time operational analytics

No data integration pipelines required

Near-zero impact on operational systems

Latency <90s at 99th percentile

Micro	soft Azure Synapse Analytics 🕨 p	rlangadws2 / Search resources	(li
>>	🖞 Publish all 🕘 🛛 🗸 Validate all	C Refresh 🗈 Discard all	
<b>^</b>	Develop + × «	🕮 HolidayDataPi X 💷 Load Data to S X 💷 Pipeline 1 X 🍄 Data flow 1 X	🗐 SQL
	P Filter resources by name	🗸 Validate 🛛 Data flow debug 😐	
Ŷ	SQL scripts	HolidayData Select1 MapDrifted1	
()	<ul> <li>Notebooks</li> <li>Data flows</li> </ul>	Columns: 0 total  Columns to rename  Creates an explicit mapping for each drifted column	7
	🚰 Data flow 1		-
0	Spark job definitions	Multiple inputs/outputs	
	Power BI	Add Source 💱 New branch	
		Join	
		Source Settings Source Op Conditional Split Optimize Inspect	Data
		Output stream name * Holid: In Union I Documentatio	0.0
			- New
		Options Allo Schema modifier	
		Infe Science Column	
		Valic Valic	
		Sampling * O Enable O Disable O	

## Data warehouse

Scalable and secure analytics platform for SQL workloads

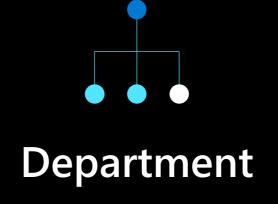
# Suitable for any scale

#### Fully-managed elastic platform

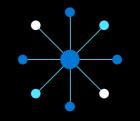
Elastic compute that can be easily optimized to different classes of workload

All features available in a single tier

Infinite cost effective PAYG storage



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

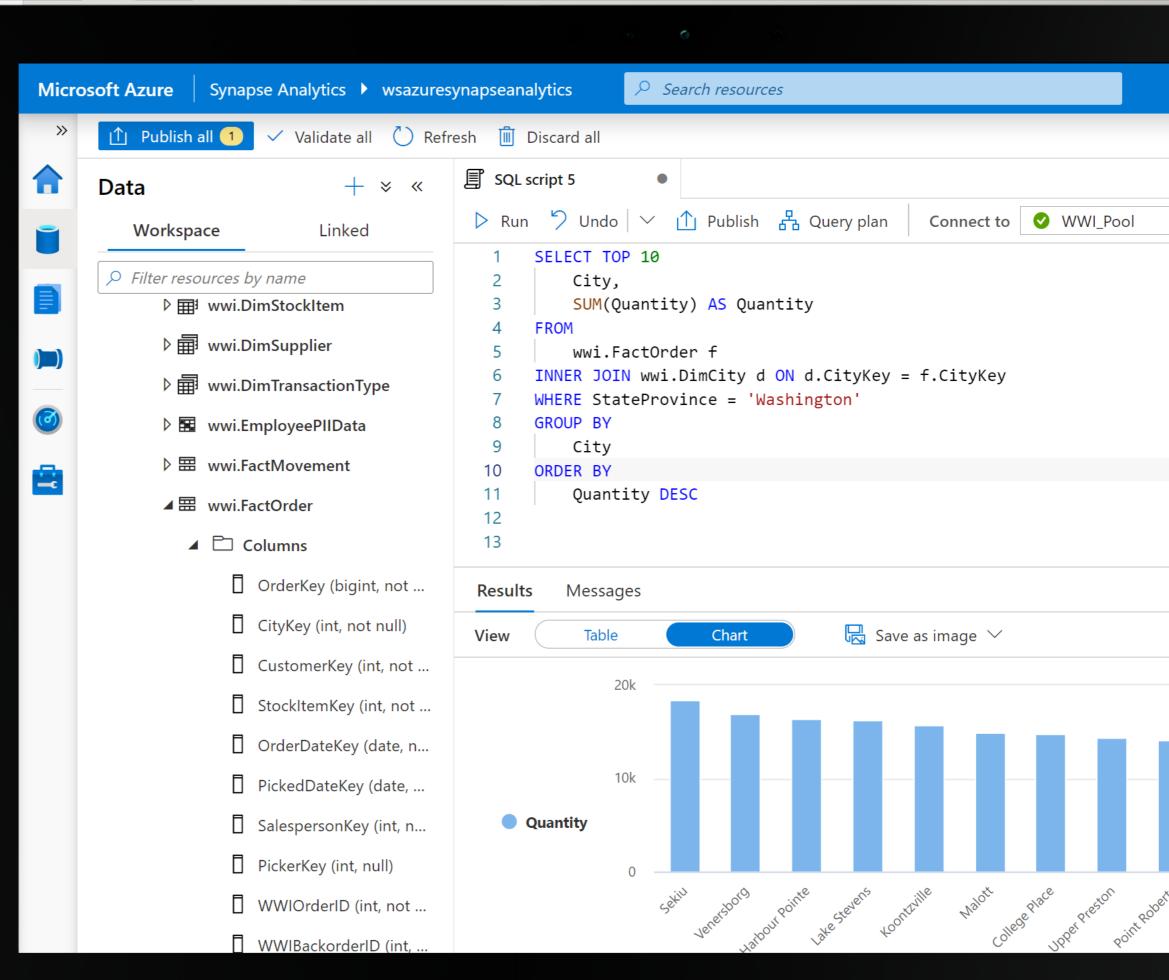
#### **SQL Editor**

Automatic code completion (Intellisense)

Script collaboration within the Workspace

**Built-in visualizations** 

Easily switch between clusters



3	B	Q
~	/	Use data

			lef
			Leg
			Leg
5		à	Leg
	Trentwi	20	
			Cat

# Industry leading SQL performance and scalability

#### **TPC-H and TPC-DS Leader**

Price/performance leadership relative to other cloud data warehouses

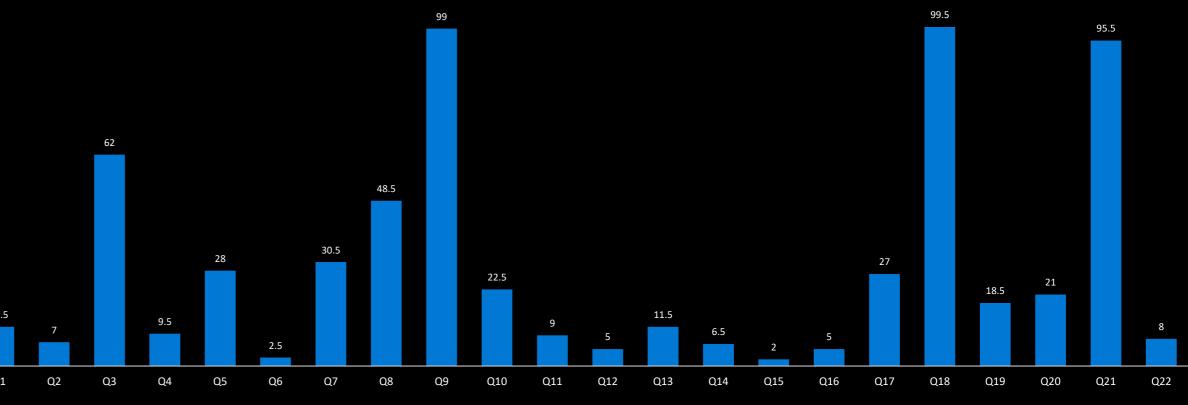
"Polaris" is the only query engine to successfully complete TPC-H at 1PB scale

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 rearchitect 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 businessintelligence (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

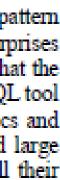
phase of interactive analysis and reporting. While this pattern bridges the lake and warehouse paradigms and allows enterprises to benefit from their complementary strengths, we believe that the two approaches are converging, and that the full relational SQL tool chain (spanning data movement, catalogs, 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:

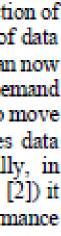
 Cell data abstraction: Polaris builds on the abstraction of a data "cell" to run efficiently on a diverse collection of data formats and storage systems. The full SQL tool chain can now be brought to bear over files in the lake with on-demand interactive performance at scale, eliminating the need to 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.

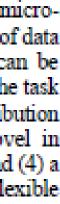
 Fine-grained scale-out: The highly-available microservice architecture is based on (1) a careful packaging of data and query processing into units called "tasks" that can be readily moved across compute nodes and re-started at the task level; (2) widely-partitioned data with a flexible distribution model; (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.

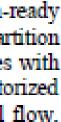
 Combining scale-up and scale-out: Production-ready scale-up SQL systems offer excellent intra-partition parallelism and have been tuned for interactive queries with deep enhancements to query optimization and vectorized processing of columnar data partitions, careful control flow,



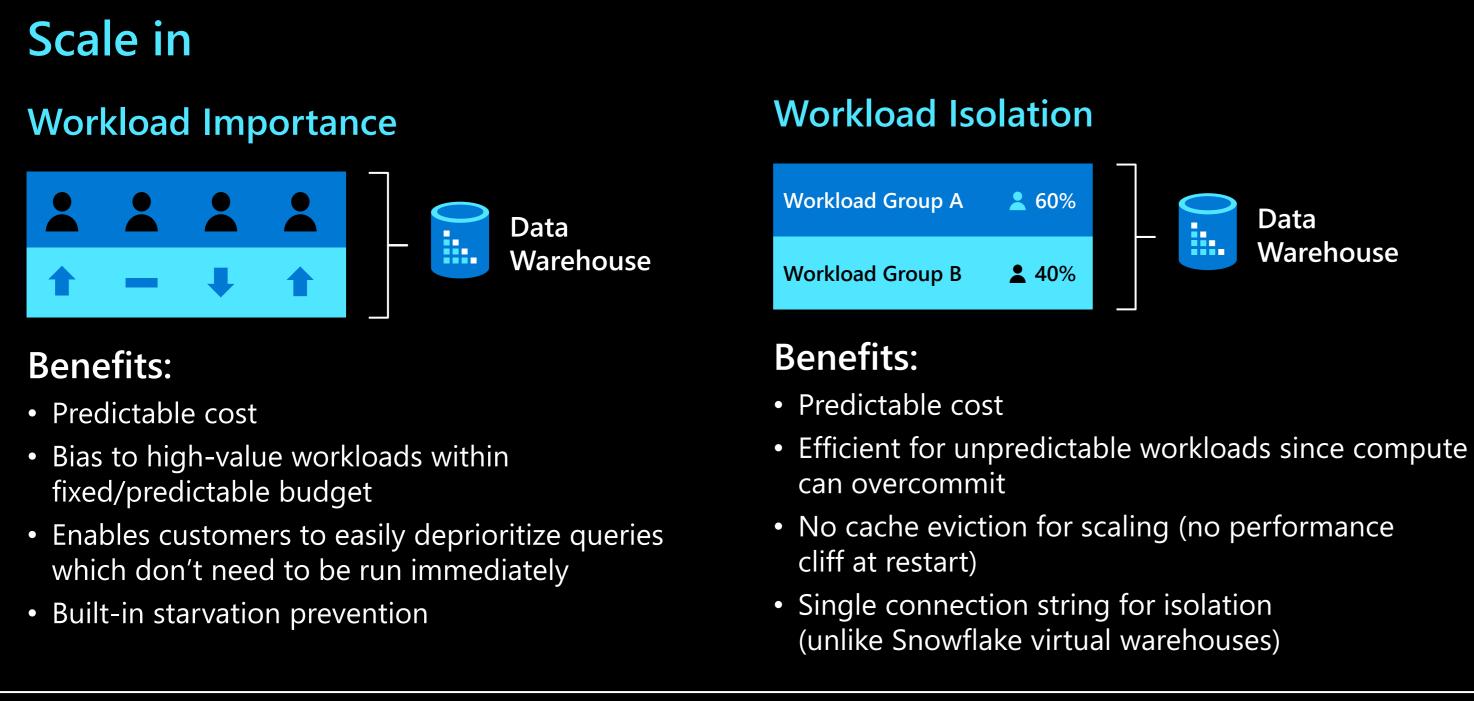










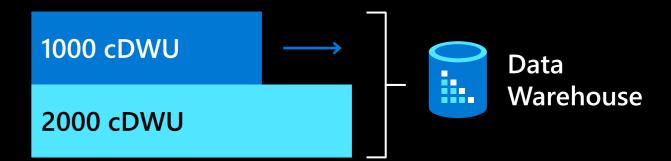


Azure Synapse supports a more diverse set of workload management tools through workload importance, intra-cluster isolation, and elastic clusters.

### Workload management

#### Scale out

#### Elastic Cluster (Scale Up)



#### **Benefits:**

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







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

#### <sup>©</sup> 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

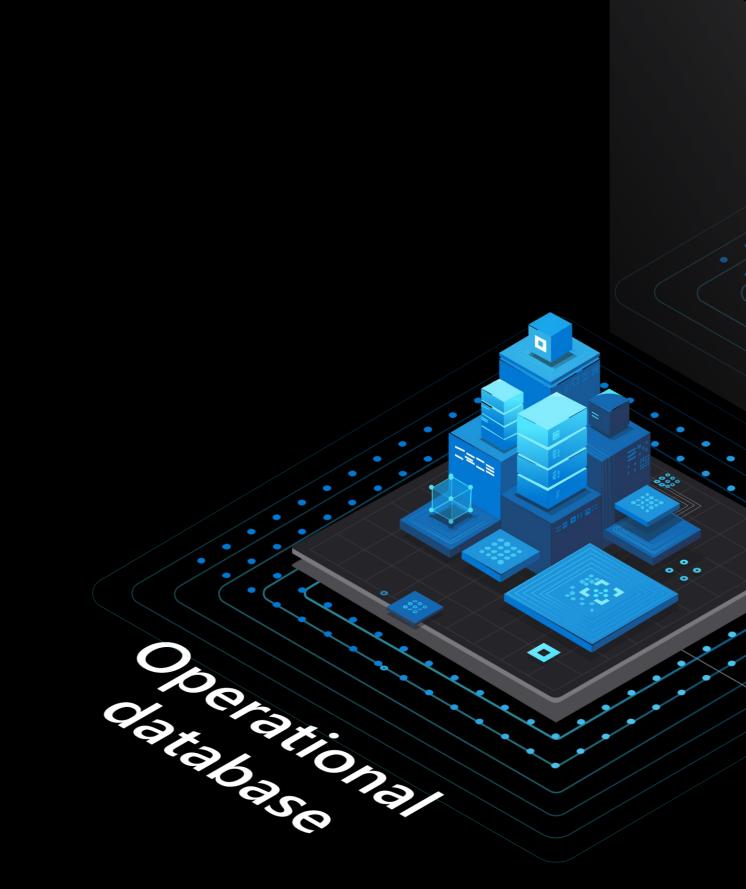
- Dramatically cost reduction eliminating the Teradata contract along with 70% savings using Azure Synapse Analytics reserved capacity and reducing operational complexity.
- Increased agility, By decoupling consumers from data providers, ABN AMRO will be able to bring new services and use cases much 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

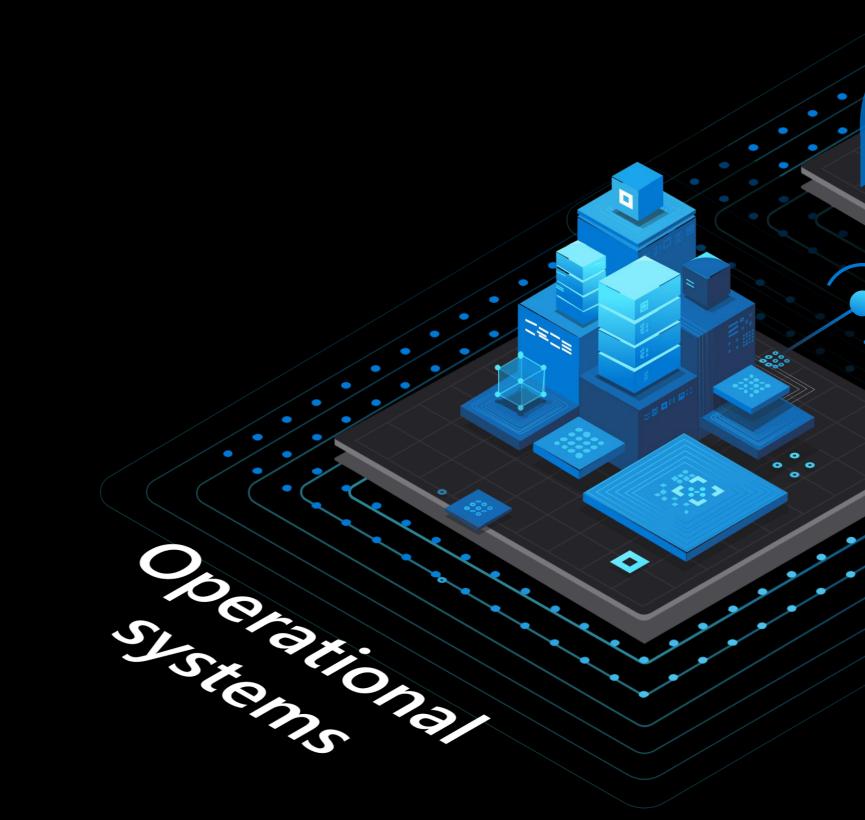
Eliminate latency and accelerate decision making

### Integrating operational data with analytics systems





### Integrating operational data with analytics systems



### Azure Synapse Link

Staller.

**Real-time data analytics** No ETL required No performance impact on transactions

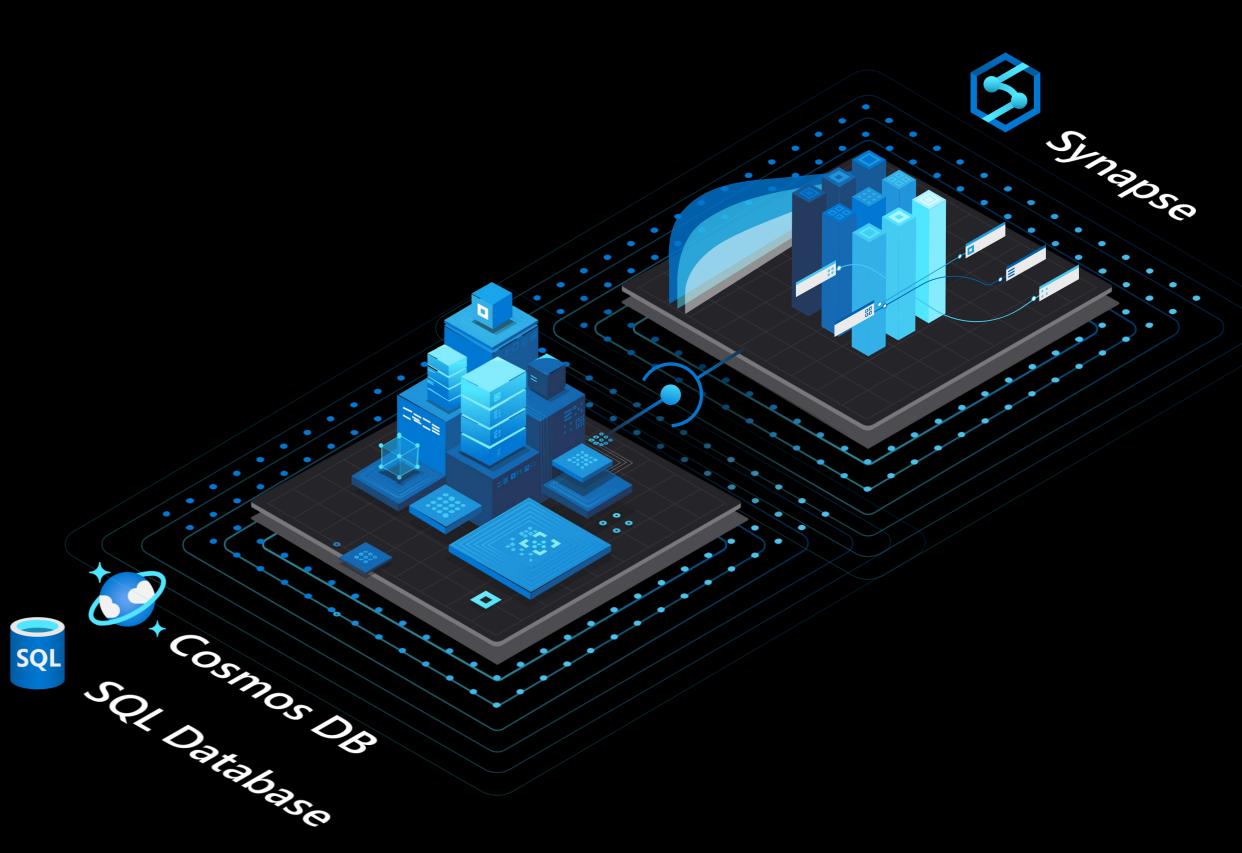


### Integrating operational data with analytics systems



### Azure Synapse Link

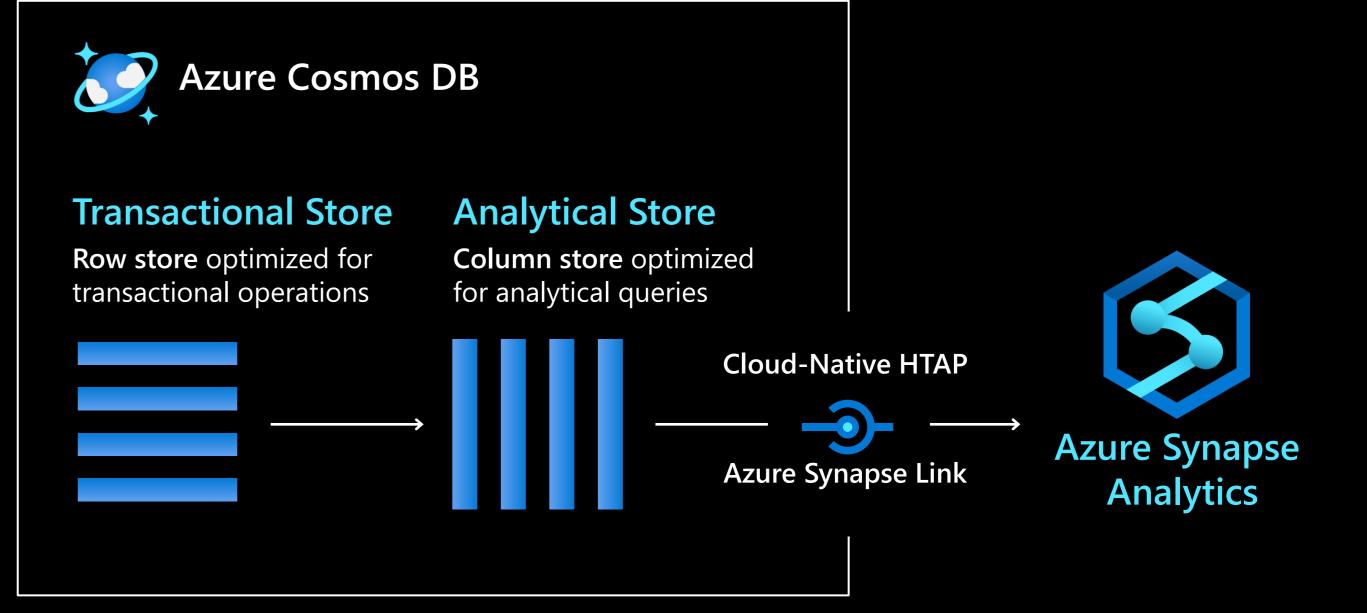
Enable real-time data analytics



## 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 99<sup>th</sup> percentile



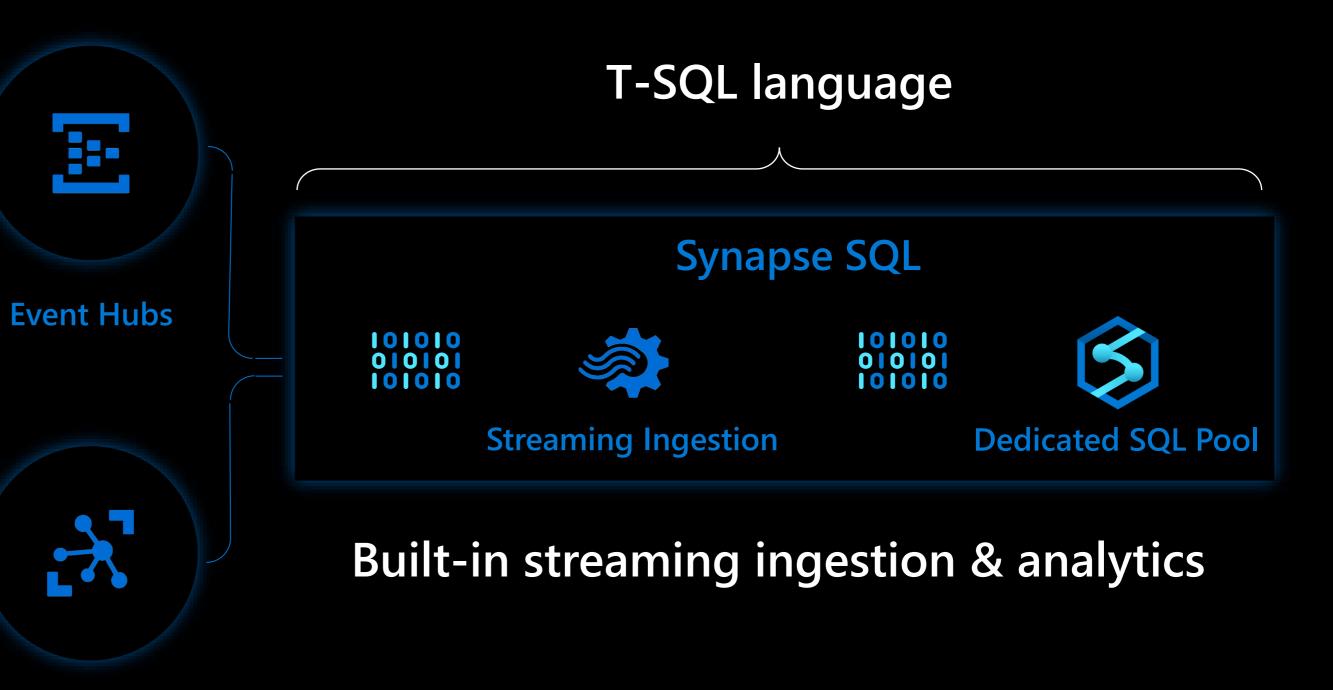
### Streaming event ingestion

#### **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



IoT Hub

### TECHNO-LPIN<sup>®</sup>

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

The global market leader in snow-making technology

#### <sup>o</sup>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

Micro	osoft Azure Synapse Analytics > wsazures	ynapseanalytics 🔎 Search 🖉 🛱 🗘 🖗						
»	🛇 Synapse live 🗸 📄 Validate all 🔟	Publish all 1						
	Develop + * «	□ 020 Surface Sales Fo ●						
	$\wp$ Filter resources by name	<ul> <li>+ Cell ∨ ▷ Run all ↑ Publish Attach to analyticspool ∨ ℃ Language PySpark (Python) ∨</li> <li>Not started</li> </ul>						
	SQL scripts 16							
	Notebooks 35	<pre>     from pandas.tseries.frequencies import to_offset     from azureml.automl.corevendor.automl.client.core.common import metrics </pre>						
))	Demo notebooks	3 from matplotlib import pyplot as plt						
	🗐 000 Data Exploration on Cosmos	4 from automl.client.core.common import constants 5						
	000 Retail-scoring-onnx-HB-AML	<pre>6 def align_outputs(y_predicted, X_trans, X_test, y_test, target_column_name, 7 predicted_column_name='predicted',</pre>						
	00 Training NYC Taxi to build ON	<pre>8 horizon_colname='horizon_origin'): 9</pre>						
	01 CosmosDB Data Exploration	<pre>10 if (horizon_colname in X_trans): 11 df_fcst = pd.DataFrame({predicted_column_name: y_predicted,</pre>						
	05 Anomaly_Detection_Cognitive	12horizon_colname: X_trans[horizon_colname]})13else:						
	05 Sentiment_Analysis_Cognitive	<pre>14 df_fcst = pd.DataFrame({predicted_column_name: y_predicted}) 15</pre>						
	06Predict NYC Taxi FareAmount	<pre>16 # y and X outputs are aligned by forecast() function contract 17 df_fcst.index = X_trans.index</pre>						
	009 Predict NYC Taxi Cab Tips	18 19 # align original X_test to y_test						
	99 ADX_Synapse_Demo_01	<pre>20 X_test_full = X_test.copy()</pre>						
	MS Conf notebooks	<pre>21 X_test_full[target_column_name] = y_test 22</pre>						
	020 AML Integration	<pre>23 # X_test_full's index does not include origin, so reset for merge 24 df_fcst.reset_index(inplace=True)</pre>						
	• 🗐 020 Surface Sales Forecasting wit	<pre>25 X_test_full = X_test_full.reset_index().drop(columns='index') 26 together = df_fcst.merge(X_test_full, how='right')</pre>						
	021 Batch Data Ingestion with Sy	27 28 # drop rows where prediction or actuals are nan						
	201 - Ignite Spark Part 1	<pre>29 clean = together[together[[target_column_name, 30 predicted_column_name]].notnull().all(axis=1)]</pre>						
	202 - Ignite Spark Part 2	31 return(clean) 32						
	Test notebooks	<pre>33 X_test[time_column_name] = pd.to_datetime(X_test[time_column_name])</pre>						
	02 Predict Retail Sales with HB	<pre>34 df_all = align_outputs(y_predictions, X_trans, X_test, y_test, target_column_name) 35</pre>						
		36 # 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 and unlimited set of file formats including Parquet, ORC and Json as well as audio, image, and video formats

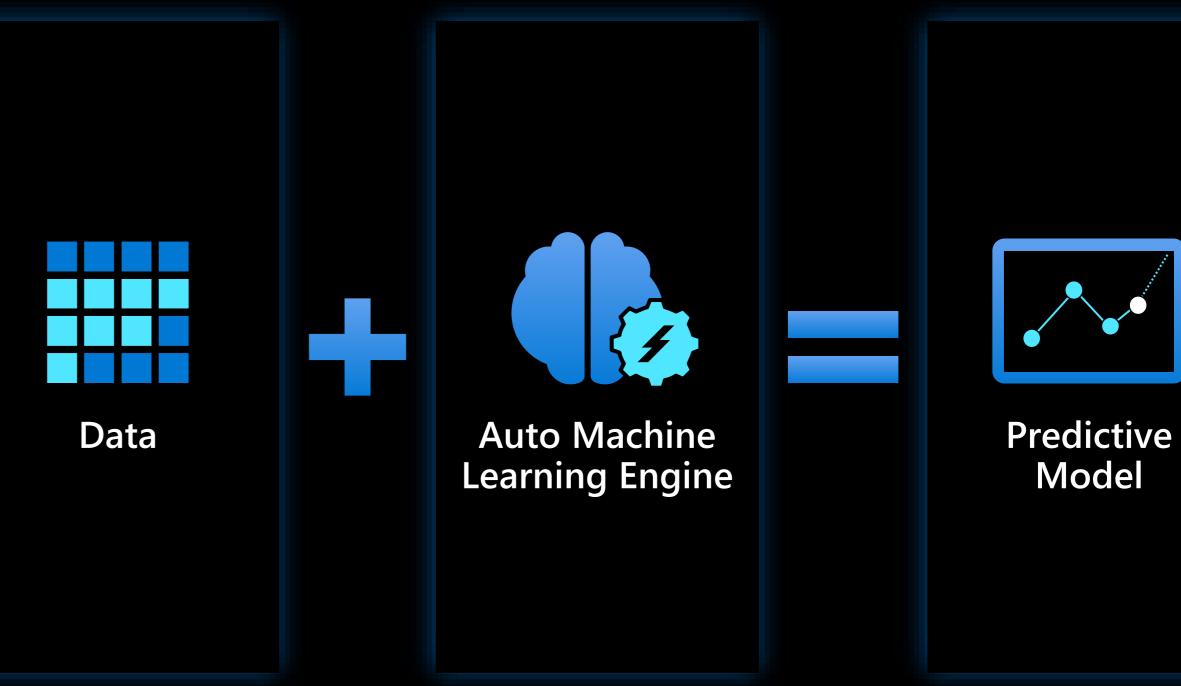




### Automatic machine learning

#### 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

Micro	soft Azure Synapse Analytics > wsazure	rch			\$	5 B	Û	ন্দ				
>>>	🛇 Synapse live 🗸 📑 Validate all 🚺	Publis	h all 1			Enrich with	new mode					
	Data + × «	02	20 Surface S	ales Fo.	•	retailsales						
-	Workspace Linked	+ 0	+ Cell ∨ ▷ Run all 1 Publist Choose a model type									
	Filter resources by name	Not started					del type for the experimer	nt based on the que	estion you	ı are try	/ing to	
	Databases 10			1	from panda:		ou have selected is created. <mark>Lear</mark> r	l the model type, you will I n more 🖸	be prompted with a	a few sett	ings bef	fore th
)==)	▷ 🔂 newpoll (SQL)			3	from azurer from matple							
_	ト 合 NYCTaxi_Pool (SQL)			4 5	from autom:	1.1.1.1	Classification					
	▷ 🔂 Predict_Pool (SQL)			6 7	<pre>def align_(</pre>		classification) c	likelihood of a specific out or identify the category an				
	▷ 🔂 Streaming_Pool (SQL)			8 9			classification).	ct if a customor will repow	or cancel their cub	ccription		
	▷ 🛱 WWI_Pool (SQL)			10 11	if (hou df_		Example. Preur	ct if a customer will renew	or cancer their sub	scription.		
	▷ 🔂 NYT2020 (SQL)			12 13	else:		Democia					
	SQLServerlessDB (SQL)			14 15	df_		Regression	neric value based on input	variables.			
	▷ 🕞 default (Spark)			16 17	# y and	$\Leftrightarrow$		ct housing prices based or				
	🔺 🕞 retaildata (Spark)			18	df_fcst							
	▲ 🗁 Tables			19 20	# alig X_test		Time series fo	precasting				
	▷ ⊞ retailsales			21 22	X_test_	(L)		s and trends based on hist	orical data.			
	Surfacesalesdb (Spark)			23 24	# X_te: df_fcsi	$\cup$	Example: Predi	ct stock market trends ove	er the next year.			
	, C* contraction (chant)			25	X_test_							
				26 27	togethe							
				28 29	# drop clean :							
				30	CIEdii -							
				31 32	return							
				33	X_test[time_							
				34 35	df_all = ai	Continue	Back				Cá	ancel
				36	# use autor							

# 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

						• •		
Micro	osoft Azure	Synapse Analytics 🕨 w	sazures	synapseanalytics		𝒫 Search		
>>	🗐 Synapse	e live 🗸 🗸 Validate	all ⊥≜	Publish all 📗	Discar	rd all		
	Data	+ *	«					Enrich with existing model
-	Worksp	ace Linked						🖬 dbo.nyc_taxi
	<i>P</i> Filter res	ources by name						Select the model you want to use to enrich
	A Databas	es	9					Azure Machine Learning
)==)	⊳ 🔂 new	poll (SQL)						Azure Machine Learning workspace *
	▲ 🔓 NYC	Taxi_Pool (SQL)						
		Tables						Name
		dbo.aml_models						mechanics-retail_data-20201105231400
		dbo.modeldeploy						mechanics-retail_data-20201105110000 mechanics-retail_data-20201105110000
	▷≣	dbo.Models						
	▷ 🖻	dbo.nyc_taxi						azuresynapse5-retail_training_data-202
		dbo.nyctaxi	📑 Ne	ew SQL script	>			synapse_retail_model_hb
		dbo.testload1029	📮 Ne	ew notebook	>			synapse_retail_model_hb_new
		dbo.Trips	😽 Ne	ew data flow			Use the	synapse_retail_model_hb
	▷ 🗅	External tables	🖽 Ne	ew integration dat	taset		ose the	synapse_retail_model_hb
	▷□	External resources	😂 M	achine Learning	>	Enrich with existing model		synapse_retail_model_hb
	▷ 🗅	Views	🖒 Re	fresh				synapse_retail_model_hb
	▷ 🗅	Programmability						synapse_retail_model_hb
	▷ 🗅	Schemas						synapse_retail_model_onnx
	▷□	Security						nyc_taxi_tip_predict
		ict_Pool (SQL)						Continue

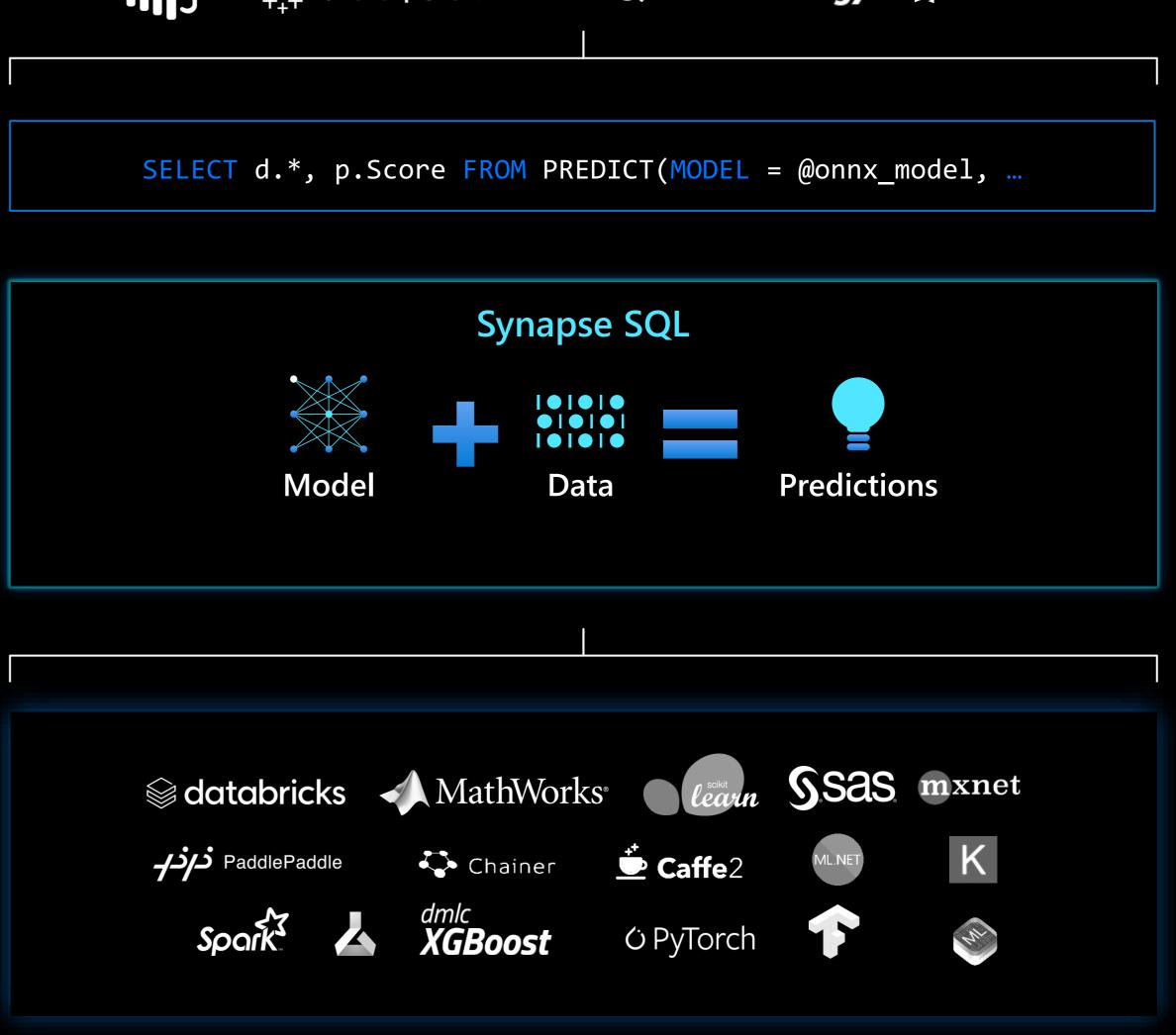
	Û	র্ম প	?
n the	e selected	d dataset.	Lea
am	lwsdemo	DS	
	Versi	Create	ed
	1	07:25:	59 1
	2	07:25:	43 1
	1	07:25:	31 1
	2	18:01:	45 1
	6	21:24:	05 1
	1	05:37:	03 1
	5	14:58:	10 0
	4	05:33:	55 0
	3	17:05:	19 (
	2	01:50:	02 (
	1	17:22:	21 (
	1	22:05:	45 (
	1	17:45:	02 0

### 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.

#### <sup>8</sup>O 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.

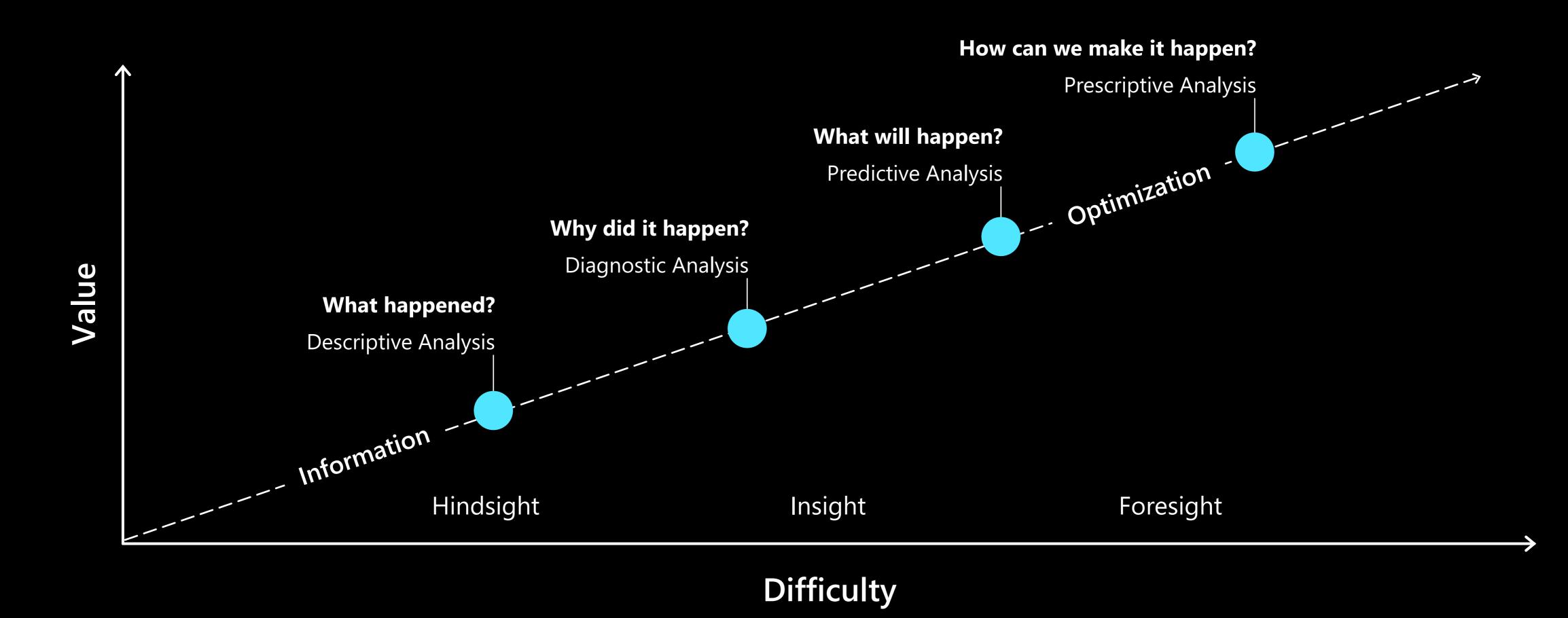




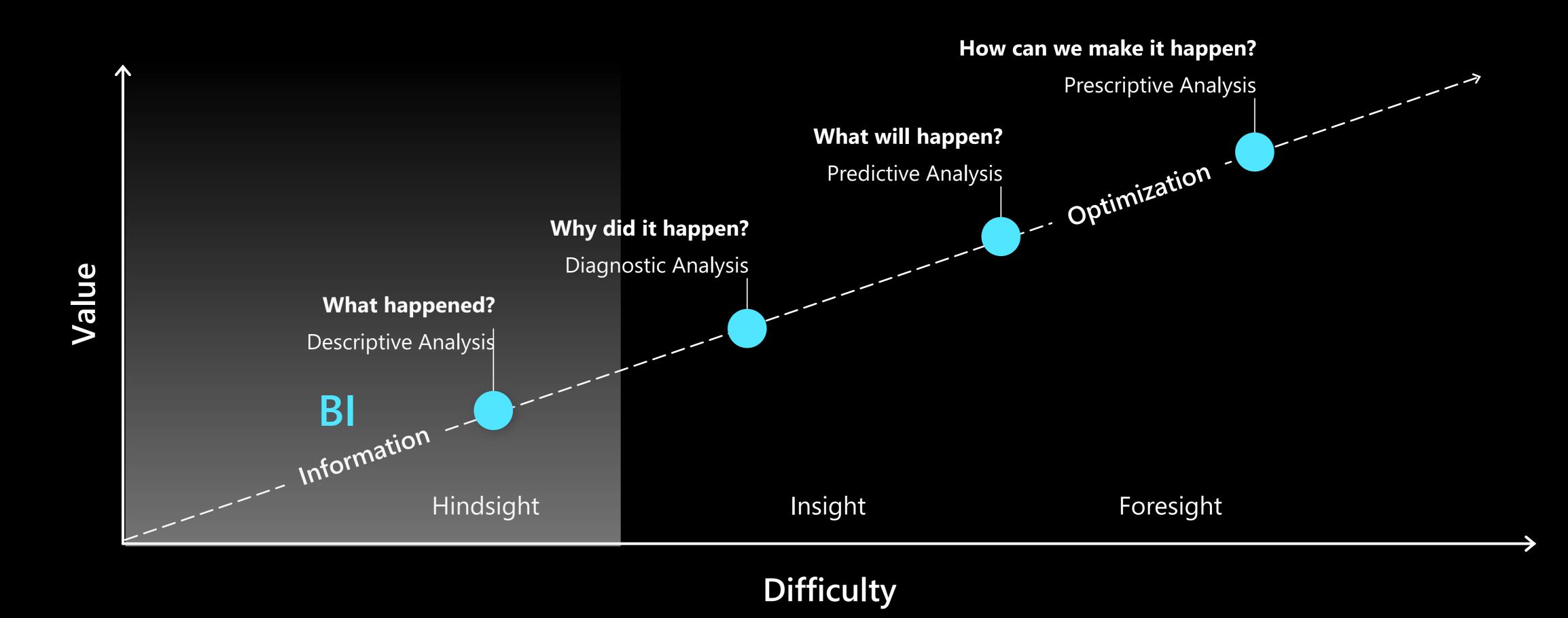
### Power BI + Azure Synapse

An unmatched combination

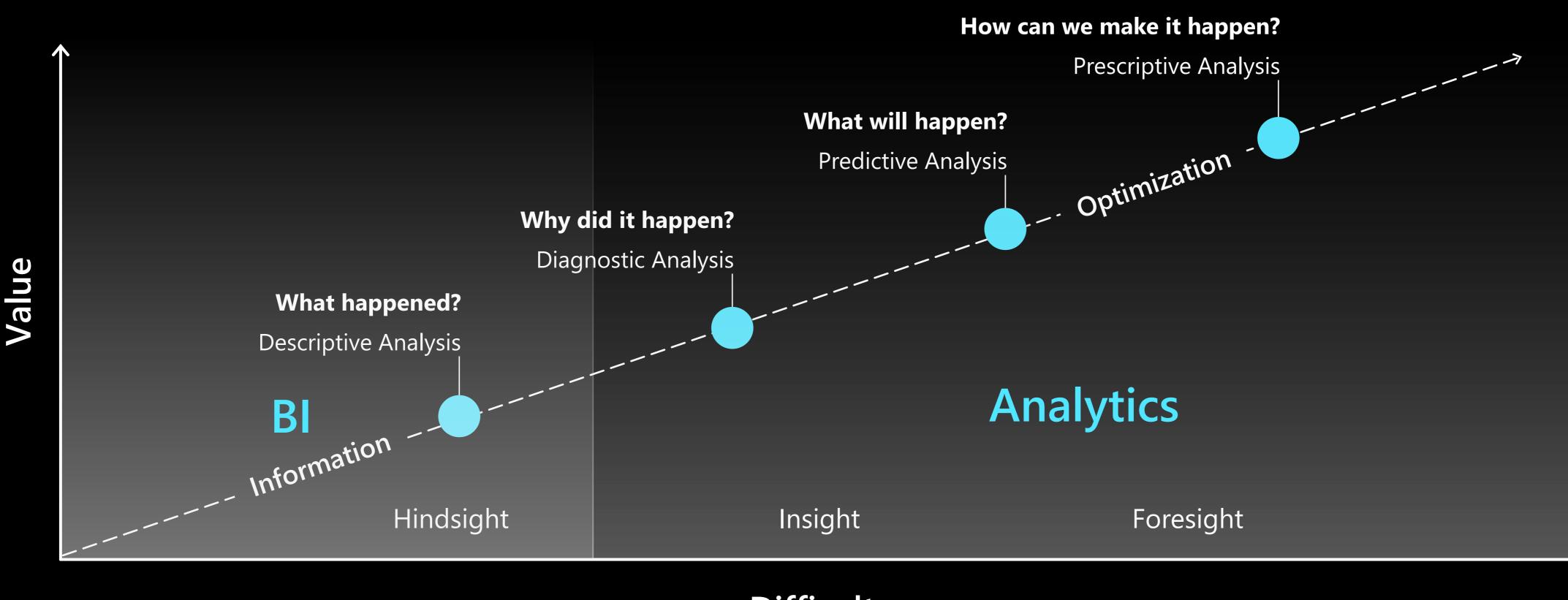
### Where do you find yourself on the curve?



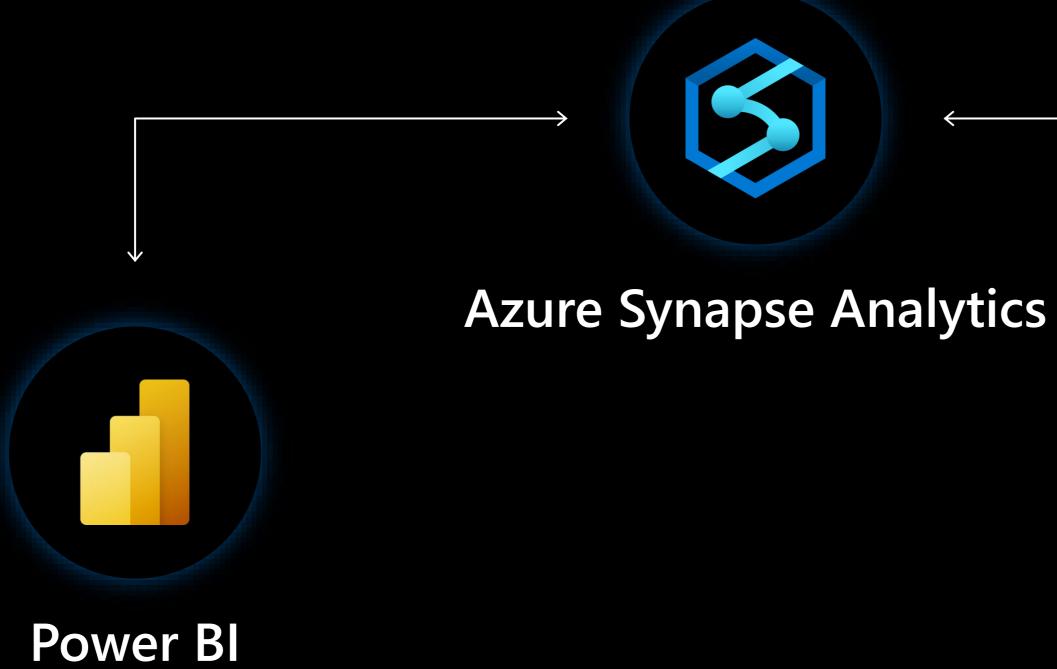
### Where do you find yourself on the curve?



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



Difficulty



Unified experience to enrich data Automated ML for rapid development **Seamless collaboration** 



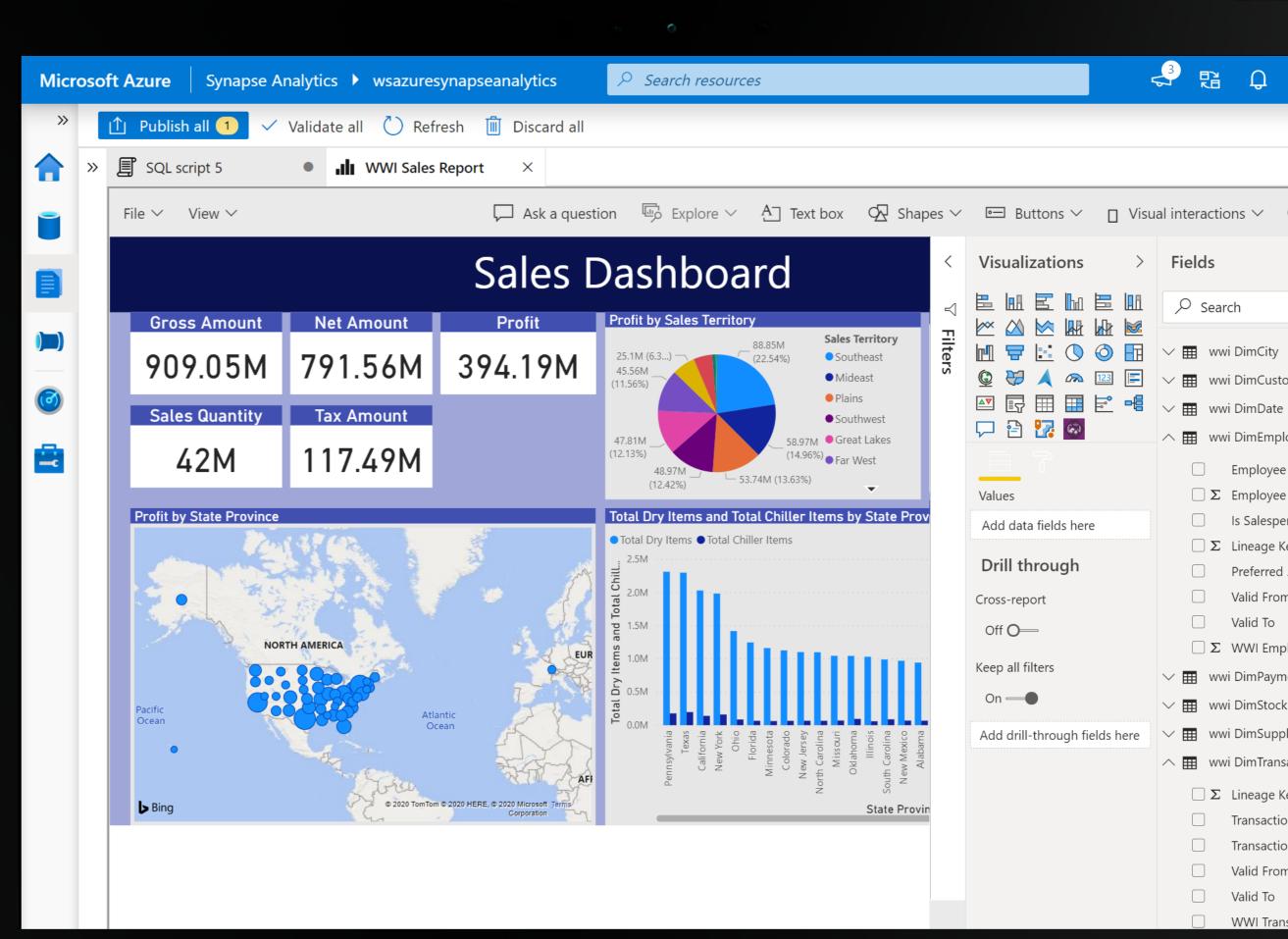
#### **Azure Machine Learning**

### Power Bl integration

### **Build dashboard in Synapse Studio**

Code-free experience for development rich visualizations

One-click publishing to for secure consumption across the enterprise



u	ual interactions 🗸 💍						
	Fields						
		Q s	Sear	ch			
	$\sim$	Ⅲ	ww	i DimCity			
	$\sim$		ww	i DimCust	tome		
	$\sim$	===	ww	i DimDate	e		
	$\wedge$		ww	i DimEmp	oloye		
				Employe	е		
			Σ	Employe	e		
				ls Salesp	er		
			Σ	Lineage	Key		
				Preferred	:		
				Valid Fro	m		
				Valid To			
			Σ	WWI Em	pl		
	$\sim$		ww	i DimPayr	nen.		
	$\sim$		ww	i DimStoc	klte		
	$\sim$		ww	i DimSup	plier		
	$\wedge$		ww	i DimTran	sact		
			Σ	Lineage	Key		
				Transacti	0		
				Transacti	0		
				Valid Fro	m		
				Valid To			
				WWI Tra	ns		

### Accelerate business value with a powerful analytics platform

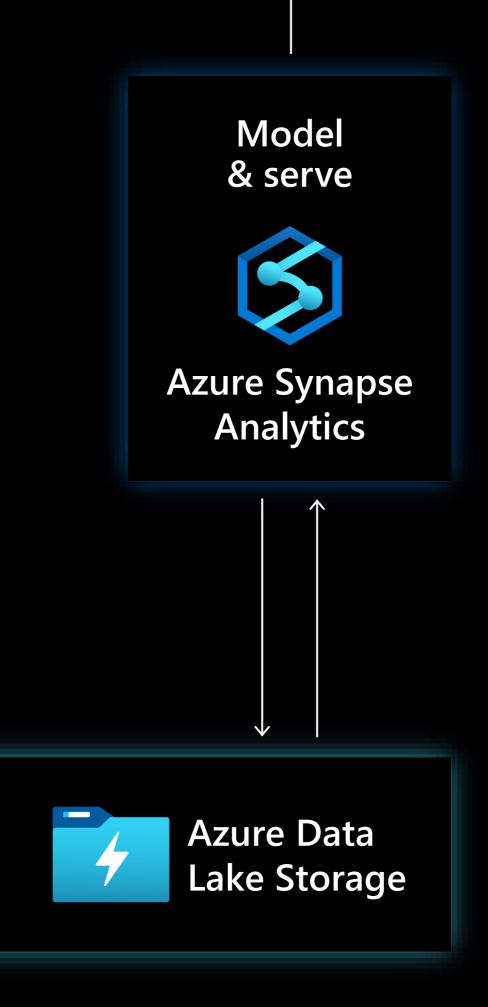


#### **Respond instantly**

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 industryleading 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



Increased customer satisfaction

27%
Faster time to insights

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





Lower total cost of ownership

## P&G

#### Understanding insights in the consumer goods market with Azure Synapse Analytics American multinational consumer goods corporation

#### Challenge စ္ပဲ

- 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.



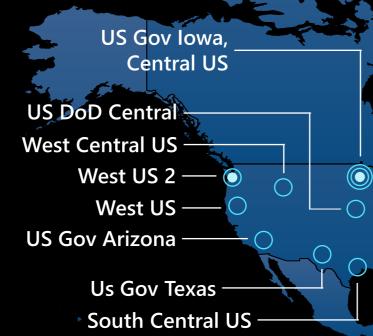
#### Solution

- 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

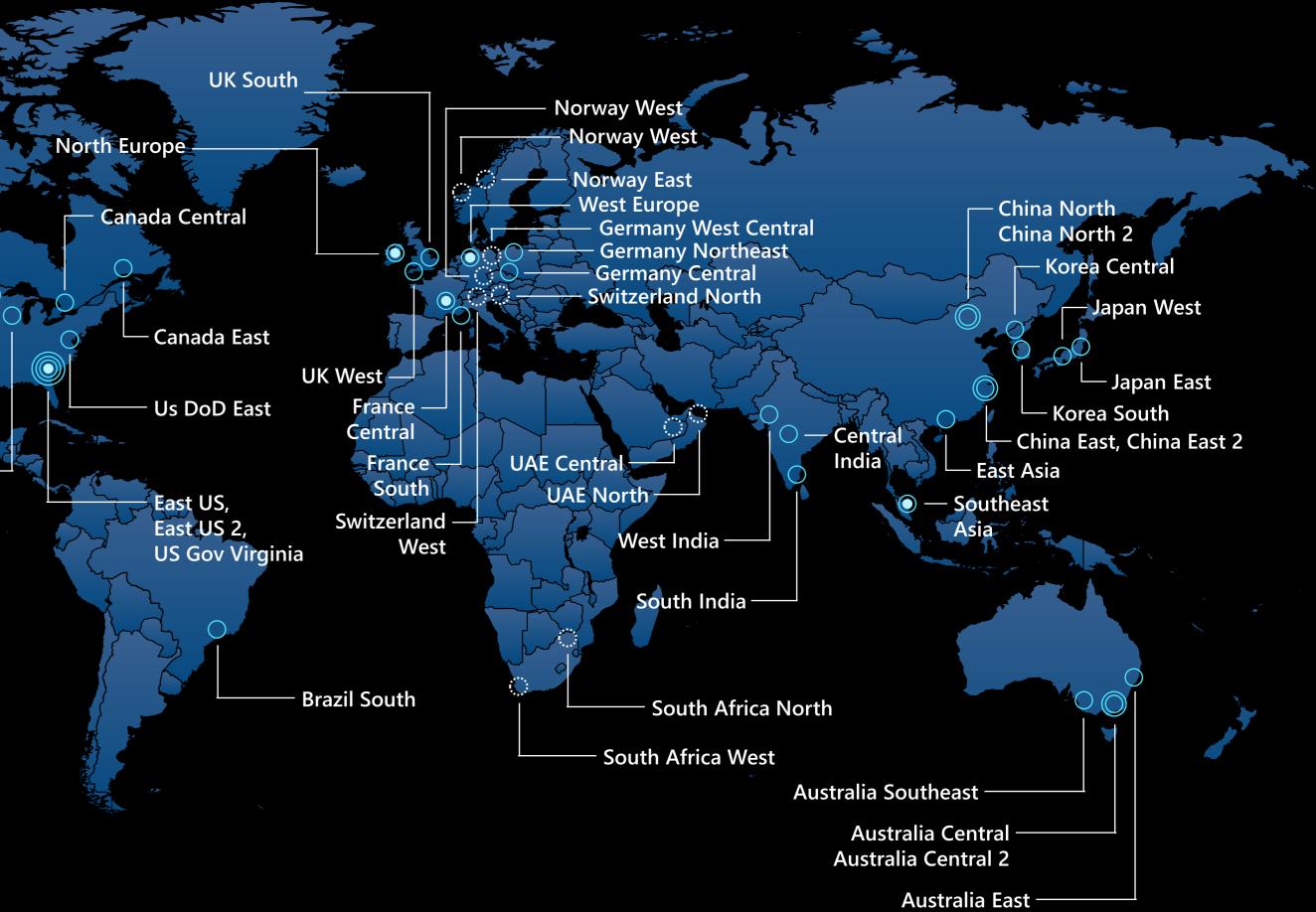
Australia Southeast	Korea Central
Australia East	North Central US
Brazil South	North Europe
Canada Central	South Africa North
Canada East	South Central US
Central India	Southeast Asia
Central US	Switzerland North
East Asia	UK West
East US	UK South
East US 2	West Central US
France Central	West Europe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



North Central US



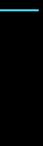
Japan West



#### ATSCALE 🔆 loö Partners TAR Data integration alter $\nabla p$ ivitio **"**Fivetran () **HV** FI preci HEDDA.IO talend TIMEXTE Data management **C** aginity $\bowtie$ A the data governance company SqIDBM ........tamr

#### **Business intelligence**

oirs	st 🖗 dren	nio <b>()</b> Dunda:	S COGNOS	Inf%rmation Builders	JReport
Sker	MicroStrat	egy MOD			+ <sup>++</sup> ++ ++++++ <b>a b   e a</b>
GIT	Thoughts	Spot <b>Vellowf</b>	n 🏶 Kylige	nce° TIBC⊘° Spotfire	e°
ryx	ð attunity		datacoral.	denodo	DIMODELO solutions
<b>V</b> R	incorta.	Informatica <sup>®</sup>	Inf%rmation Builders		A MATILLION
isely	RŻVERY	🔆 Segment	skyvia	snapLogic	StreamSets
ENDER	${oldsymbol{\mathbb{Q}}}$ trifacta	WhereScape®	Xplenty	/	
nt			Mach	nine learnir	ng and Al
lation	Collibra			lataiku Qu	bole°
	► MicroERE		DO		1athWorks <sup>®</sup>
	teleran		S.S.	as	



U

## Knowledge Center

#### Accelerate time to solution

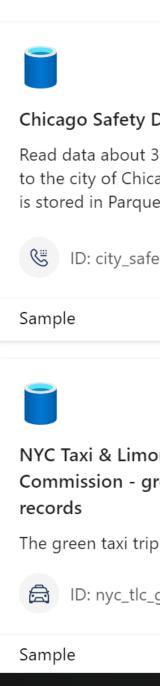
Azure Open Data sets

Pre-built samples to accelerate development

- SQL Scripts
- Notebooks
- Data Pipelines

		• <b>0</b>						
Microsoft	Microsoft Azure Synapse Analytics V wsazuresynapseanalytics							
» S	» Sample center							
<b>^</b>	Datasets Notebooks SQL scripts Pipelines							
	Datasets     Notebooks     SQL scripts     Pipelines       V     Filter by keyword     Tags : All							
	Bing COVID-19 Data	Boston Safety Data	COVID Tracking Project					
<b>(</b>	Bing COVID-19 data includes confirmed, fatal, and recovered cases from all regions, updated da	Read data about 311 calls reported to the city of Boston. This dataset is stored in Parquet format and is up	The COVID Tracking Project dataset provides the latest numbers on tests, confirmed cases, hospitalizat					
	ID: bing-covid-19-data	Statesty_boston	ID: covid-tracking					
	Sample	Sample	Sample					
	European Centre for Disease Prevention and Control (ECDC) Covid-19 Cases	NOAA Integrated Surface Data (ISD) NOAA Integrated Surface Data (ISD)	NYC Taxi & Limousine Commission - For-Hire Vehicle (FHV) trip records					
	The latest available public data on	provides Worldwide hourly weath	The For-Hire Vehicle trip records i					
	ID: ecdc-covid-19-cases	ざ ID: isd	D: nyc_tlc_fhv					
	Sample	Sample	Sample					

3 5 7







## THANK YOU!

