



## PowerBI - Self Service Analytics

Ali Bouhaddou  
Franck Gaillard  
Frédéric Gisbert

Microsoft GPS - Cloud Solution Architect

16 Février 2023



# Agenda

## 1 From Traditional BI to Modern Analytics

PowerBI architecture pattern.

## 2 Parallel processing: Prepare the Data

Understand the dataprep paradigm and how to use them in Power BI (DataFlow pipelines & Power BI pipelines).

## 3 Machine learning: Enrich the Data

Azure Machine learning and Power BI better together.

## 4 Data Gouvernance: unlock your data

Azure Purview and Power BI better together.

Power BI Private link | MIP | MCAS

Audit PowerBI Activity

## 5 Lakehouse database, architecture considerations

Synapse analytics & Power BI (composite model, serverless data exploration )

# Thriving Community

**1.6M**

Community Members



**300+**

Independent Power BI  
User Groups Worldwide

**23k+**

Active Ideas

● Customer Ideas

**500k+**

User Votes

**95k+**

Active Users



Microsoft  
Power BI

# The only vendor with leadership in all MQ's



## Data Integration



## Analytics



## Business Intelligence

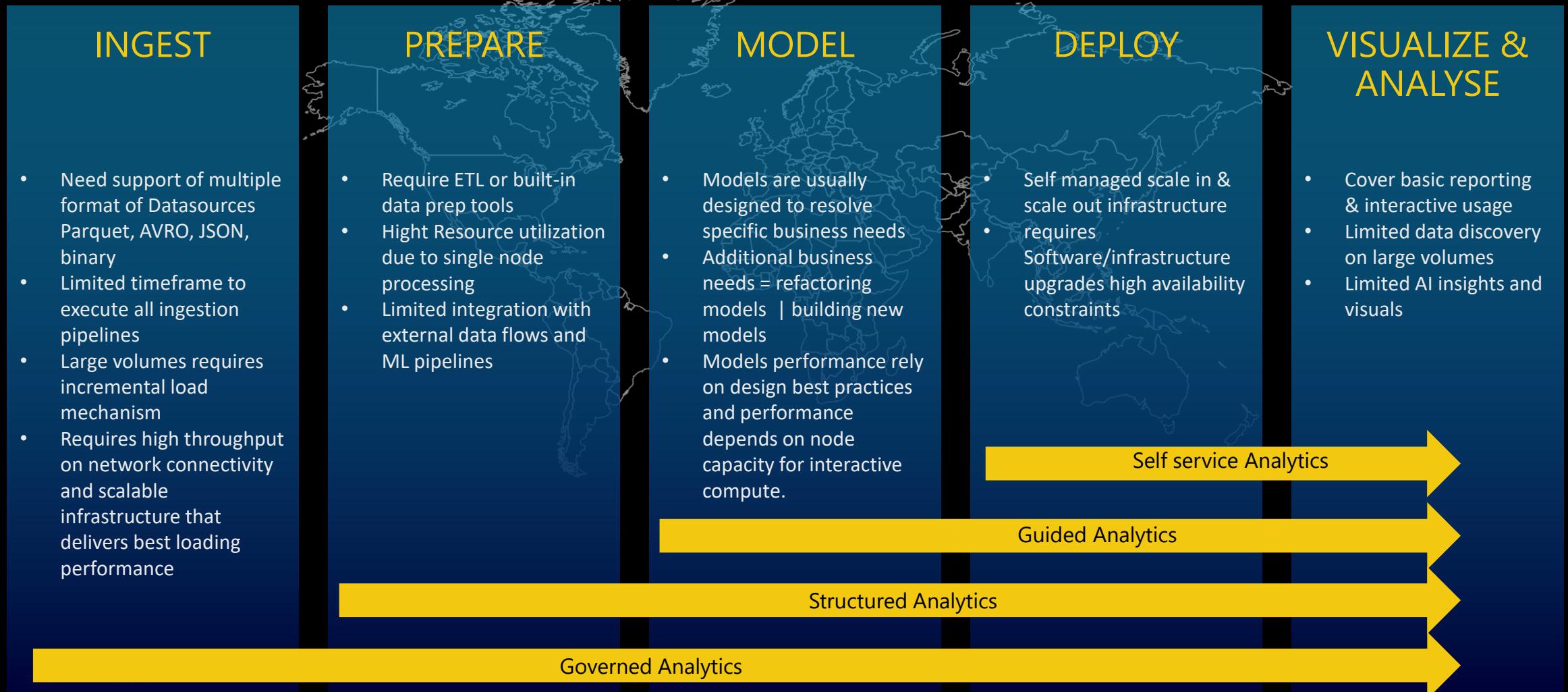


Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and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

The background of the slide is a photograph of a city skyline at dusk or dawn. The buildings are dark silhouettes against a bright, orange and yellow sky. In the foreground, there is a large amount of white noise or static.

From Traditional BI to Modern Analytics

# Growing pains in business analytics



# In Modern Analytics, Data collaboration is crucial



Frictionless collaboration

# Modern Analytics aim to **maximize the value of your data**



Powerful  
visualization and  
reporting

+



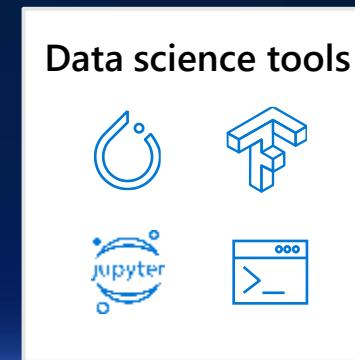
Petabyte scale  
analytics

+



Advanced  
analytics and AI

# Data silos and incompatible tooling inhibit collaboration



Incompatible tooling

Siloed data



Marketing



Sales



Product



Ops



Finance

# Unify your organization



Power BI

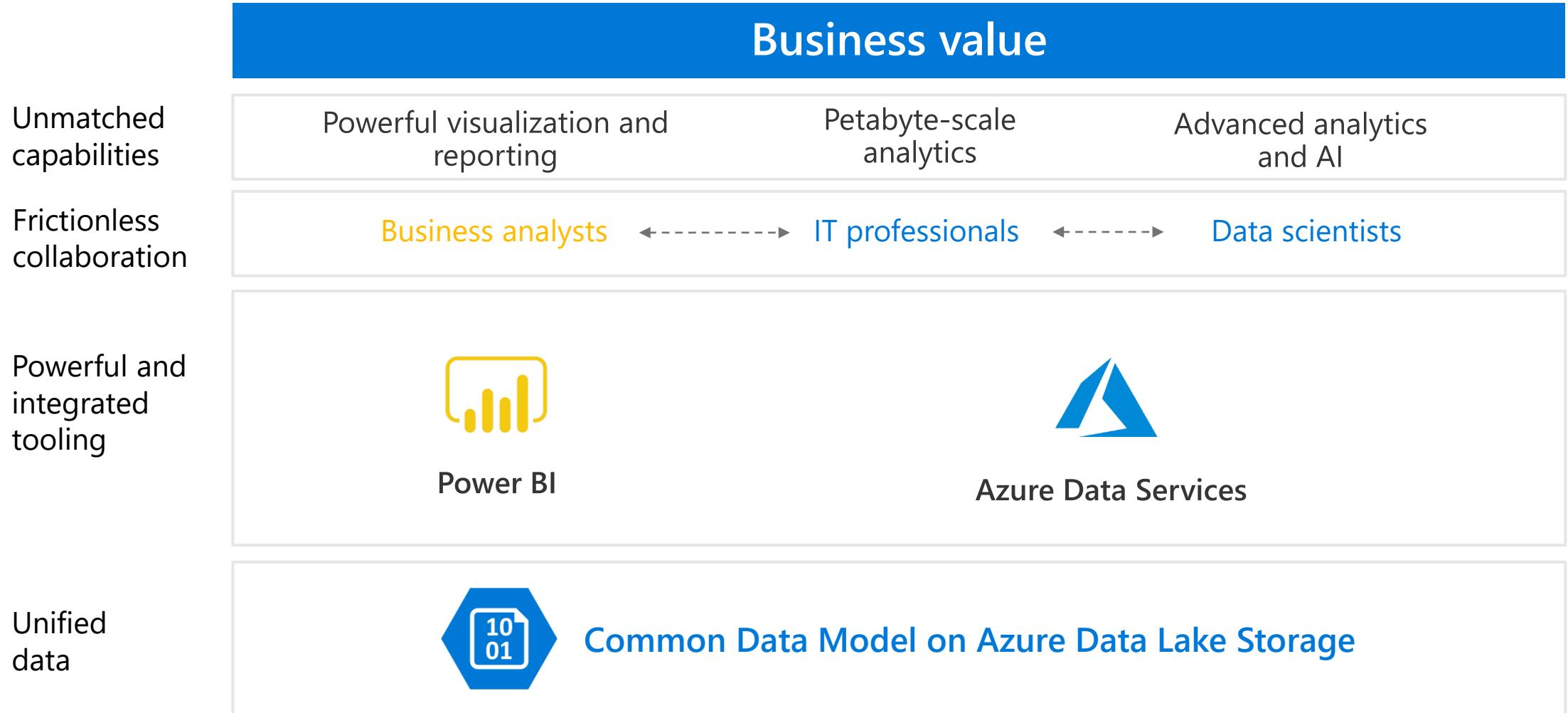


Azure Data Services



Common Data Model on Azure Data Lake Storage

# Unleash data value with Power BI and Azure Data Services



# Unleashing data value with Power BI and Azure Data Services

“

Azure gave us the **integration of back-end data analytics with front-end data consumption tools**, such as Power BI

and Microsoft Office. There are fewer moving parts to maintain and **we were able to pay for the new platform with the first-year savings.**

[Now] we go wild in terms of unleashing the value of our datasets.

- *Wilmer Peres, Information Services Director of Big Data and Analytics at Reckitt Benckiser*



**HEALTH ▾ HYGIENE ▾ HOME**

## Streaming Datasets

Realtime analytics  
IOT, Event Streaming

## Direct Query Database

Relational Datastores  
(Oracle, SAP Hana, BigQuery,  
Azure Synapse ...)

## Import Datasets

OData, parquet, Sharepoint

## Composite Models

Large datasets

## Guided analytics

Dashboards / Reports



## SelfService BI

Create your own reports online



## AI Visuals

Key influencers,  
decomposition tree, R  
visuals, Qna...



## Integrated AutoML

Train models and score  
powerbi dataflows



## Data Exploration

Query directly on Datalake  
using SQL Serverless



## DataFlows

Dedicated nodes for  
dataprep



## Embedded analytics

ISV & customers LOB Apps



## Devops integration

Devops pipelines - Powerbi  
rest api  
Deployment pipelines

## In-memory Data processing

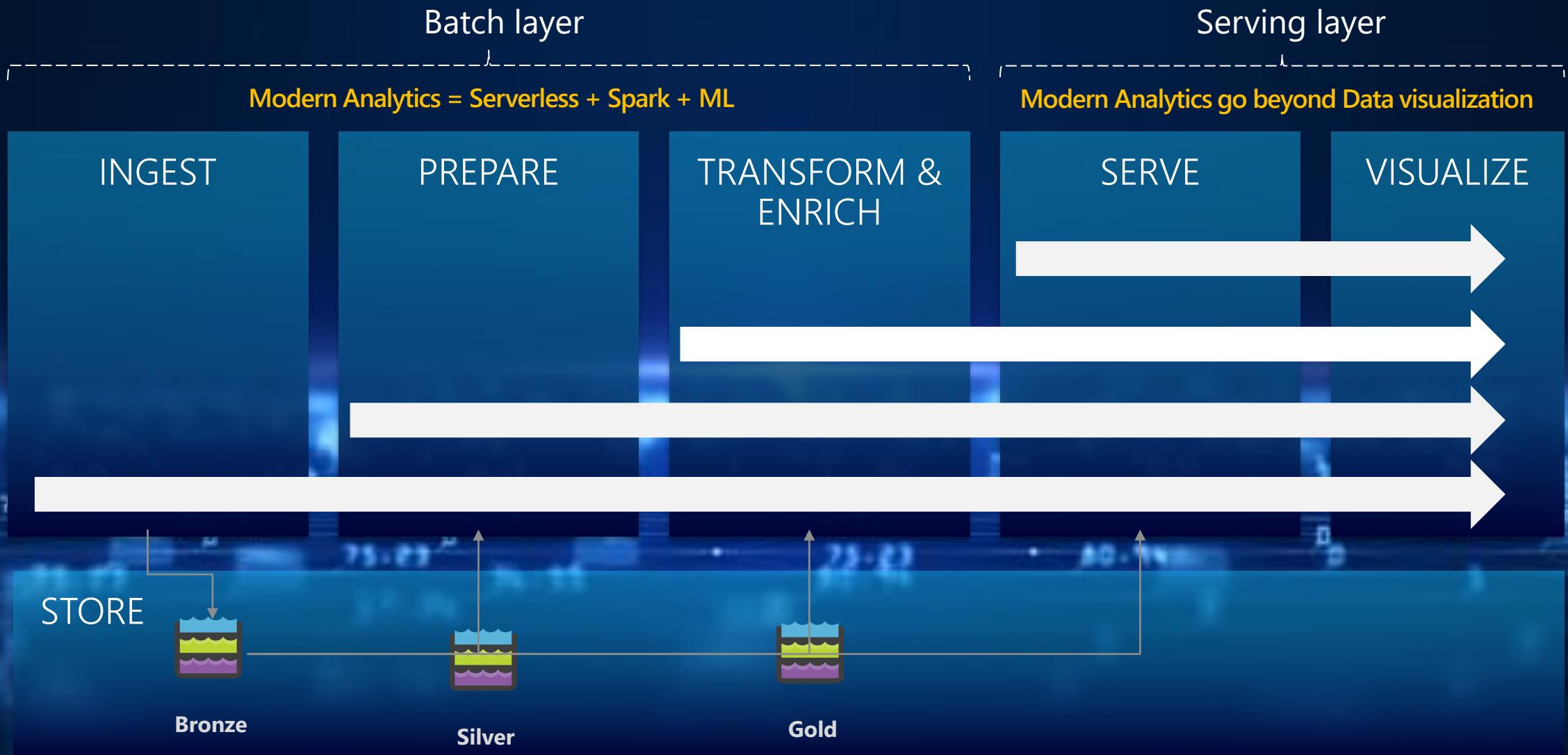
Up to 10X Data compression Column store

Integrated DataPrep/DataViz/Data Modeling

Collaboration

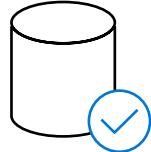
80 builtin Connectors

Native integration with Azure services



# Power BI + Azure Data Services: Better together

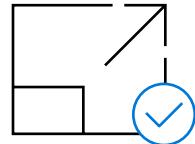
## One lake for all data



---

Game-changing data collaboration across your organization

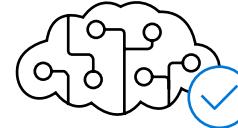
## Unlimited scale



---

Reason over petabytes of data in seconds

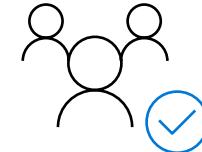
## Breakthrough insights



---

Unlock deeper insights with powerful AI models

## Easy consumption

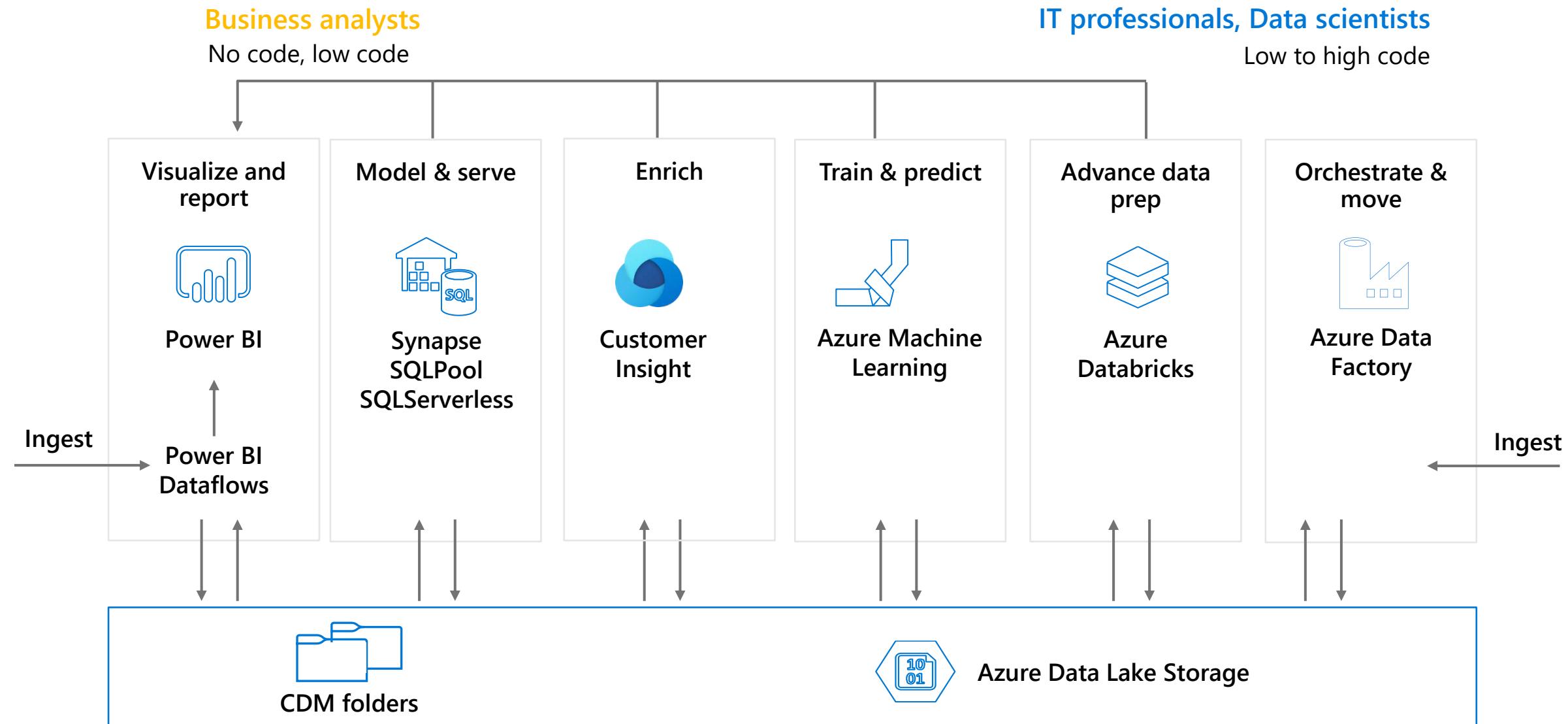


---

Put actionable insights in the hands of everyone

**Unify your organization and unlock game-changing insights for all**

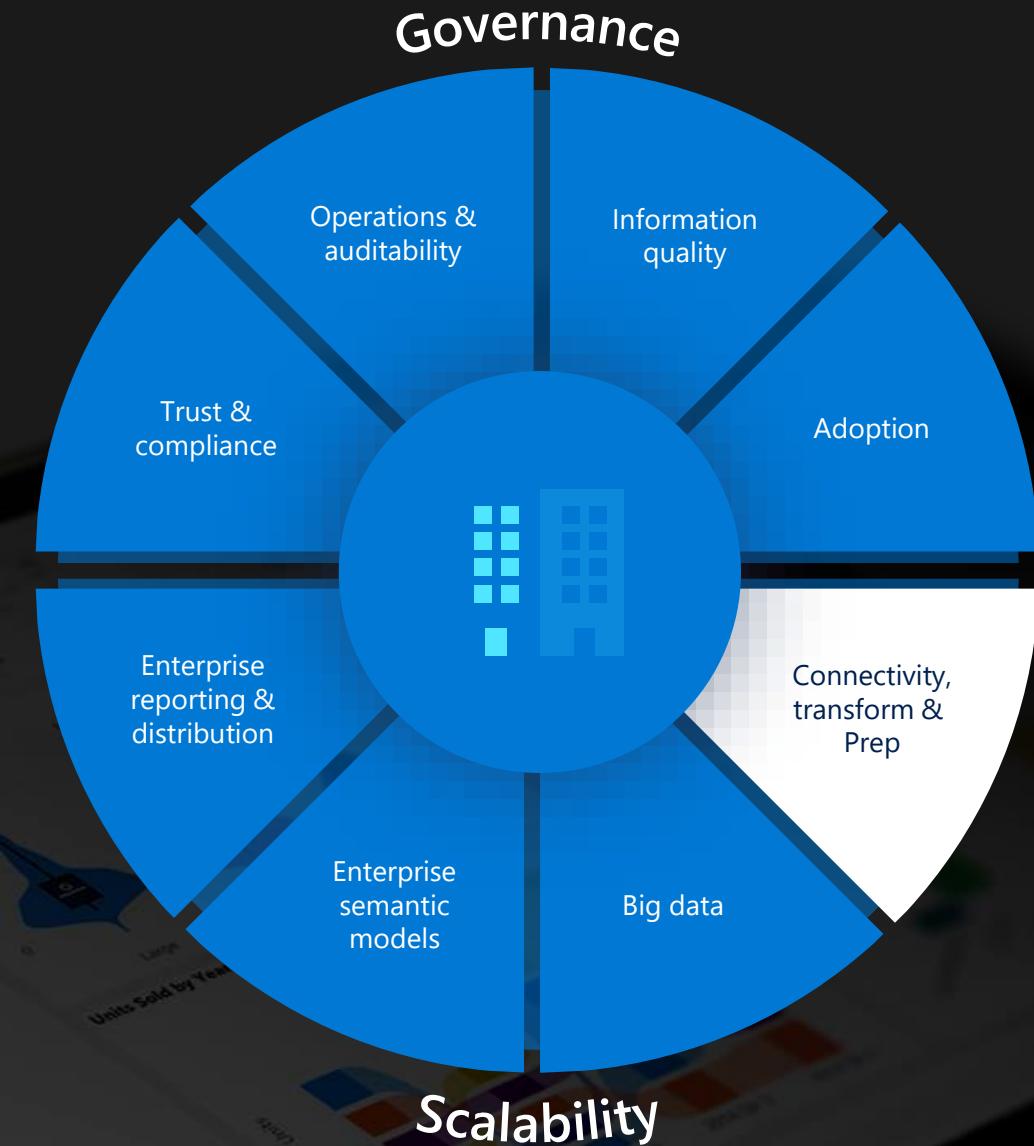
# Data Producer / Data consumer





# Empowering every organization

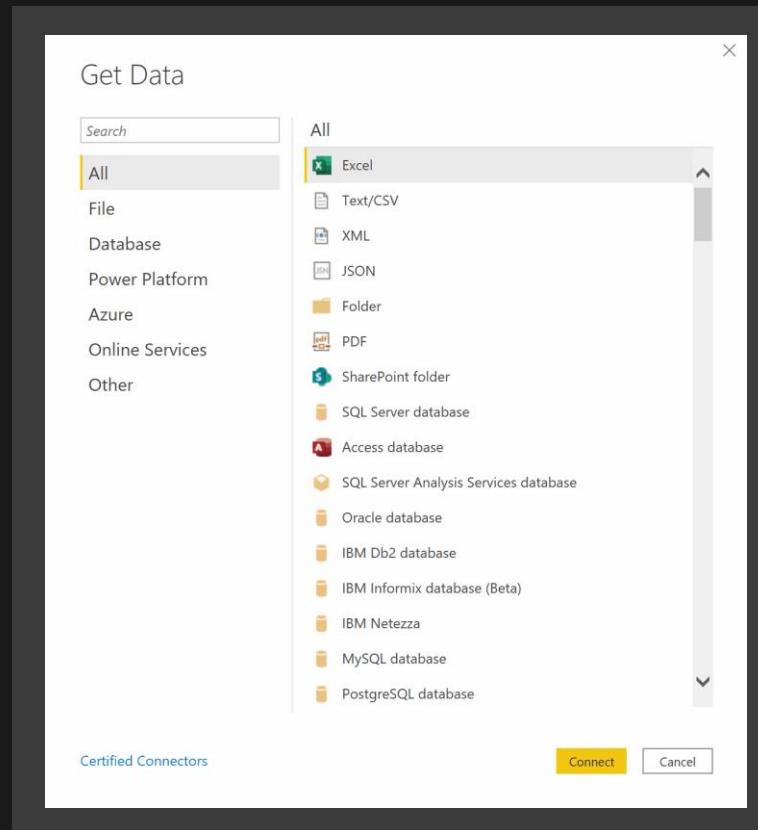
## Connectivity, transform & prep





# Power Query – richest data connectivity in the industry

## 150+ data connectors out-of-the-box



82

### Built by Microsoft

- Developed by our Product Team
- Bundled w/ Power Query
- **Examples:** SQL Server, Analysis Services, PDF Files, SAP BW & HANA, Web Page & Web API, ODBC, etc.

70

### Built by ISVs

- Built by 3rd party ISVs, using our Connectors SDK
- Certified and Shipped by Microsoft
- Bundled w/ Power Query
- **Examples:** Denodo, Databricks, etc.

1,000s

### Custom

- Built by 3rd party ISVs, SIs, partners, etc.
- Side-loaded by users/admins into Power Query & On-premises data gateway



# Self-service data prep

## Power BI Dataflows



Power Query - Edit queries

Home Transform Add column View

Get data Options Manage parameters Refresh Properties Advanced editor Choose columns Remove columns Keep rows Remove rows Sort Data type: Text Merge queries Append queries Combine files

Enter data Options Parameters Manage Manage columns Manage rows Group by Replace values

New query Options Parameters Refresh Properties Advanced editor Choose columns Remove columns Keep rows Remove rows Sort Data type: Text Merge queries Append queries Combine files

Queries

Customers

OData Navigation Choose columns Remove dupl... Filter rows

Orders

OData Navigation Choose columns Expand Remove dupl... Group by

Top customers

Merge ↗ 2

Query settings

Table.NestedJoin(Customers, {"CustomerID"}, Orders, {"CustomerID"}, "Orders", JoinKind.Inner)

	CustomerID	CompanyName	ContactName	ContactTitle	Address	City	Region	PostalCode	Country	Phone	Orders
1	RATT	Rattlesnake Canyon Grocery	Paula Wilson	Assistant Sales Representative	2817 Milton Dr.	Albuquerque	NM	87110	USA	(505) 555-5939	[Table]
2	WHITC	White Clover Markets	Karl Jablonski	Owner	305 - 14th Ave. S. Suite 3B	Seattle	WA	98128	USA	(206) 555-4112	[Table]
3	SPLIR	Split Rail Beer & Ale	Art Braunschweiger	Sales Manager	P.O. Box 555	Lander	WY	82520	USA	(307) 555-4680	[Table]
4	OLDWO	Old World Delicatessen	Rene Phillips	Sales Representative	2743 Bering St.	Anchorage	AK	99508	USA	(907) 555-7584	[Table]
5	LONEP	Lonesome Pine Restaurant	Fran Wilson	Sales Manager	89 Chiaroscuro Rd.	Portland	OR	97219	USA	(503) 555-9573	[Table]
6	THEBI	The Big Cheese	Liz Nixon	Marketing Manager	89 Jefferson Way Suite 2	Portland	OR	97201	USA	(503) 555-3612	[Table]

Create



# Data prep driven by AI

The screenshot shows the Microsoft Power Query Editor interface. The main area displays a table with four columns: CustomerKey, FirstName, LastName, and Sales. The table has 39 rows, each containing a unique customer identifier and their names along with their total sales. The 'CustomerKey' column is sorted in ascending order. The 'Sales' column shows values ranging from 5 to 167. The 'FirstName' and 'LastName' columns are grouped under the 'CustomerKey' column.

The 'Query Settings' pane on the right side of the editor provides details about the current query:

- Properties:** Name is set to "Employee Sales".
- Applied Steps:** The steps listed are: Source (Expanded InternetSales), Removed Columns, Reordered Columns, and Grouped Rows (the last step, which is currently selected).

At the bottom left, it says "4 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows". At the bottom right, it says "PREVIEW DOWNLOADED AT 8:55 AM".

CustomerKey	FirstName	LastName	Sales
11000	Jon	Yang	57
11001	Eugene	Huang	82
11002	Ruben	Torres	18
11003	Christy	Zhu	28
11004	Elizabeth	Johnson	34
11005	Julio	Ruiz	31
11006	Janet	Alvarez	13
11007	Marco	Mehta	36
11008	Rob	Verhoff	17
11009	Shannon	Carlson	28
11010	Jacquelyn	Suarez	23
11011	Curtis	Lu	30
11012	Lauren	Walker	28
11013	Ian	Jenkins	24
11014	Sydney	Bennett	38
11015	Chloe	Young	3
11016	Wyatt	Hill	15
11017	Shannon	Wang	5
11018	Clarence	Rai	26
11019	Luke	Lai	304
11020	Jordan	King	1
11021	Destiny	Wilson	9
11022	Ethan	Zhang	6
11023	Seth	Edwards	24
11024	Russell	Xie	90
11025	Alejandro	Beck	39
11026	Harold	Sai	67
11027	Jessie	Zhao	53
11028	Jill	Jimenez	35
11029	Jimmy	Moreno	54
11030	Bethany	Yuan	40
11031	Theresa	Ramos	91
11032	Denise	Stone	167
11033	Jaime	Nath	84
11034	Ebony	Gonzalez	23
11035	Wendy	Dominguez	38
11036	Jennifer	Russell	10
11037	Chloe	Garcia	21
11038	Diana	Hernandez	25

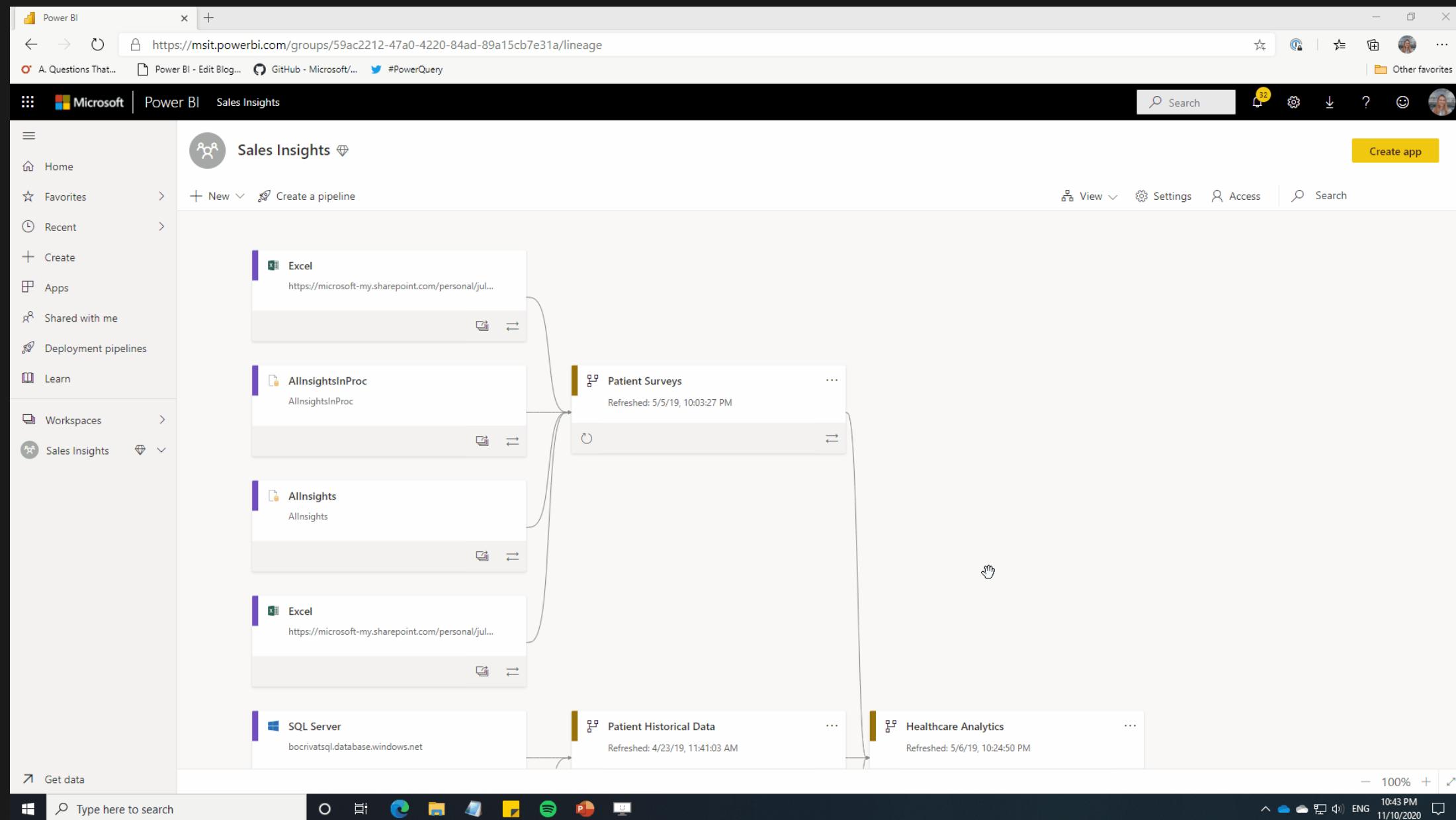


Empowering every organization

Connectivity, transform & prep

# Open & reusable data prep

Composable ETL via  
Azure Data Lake Gen 2



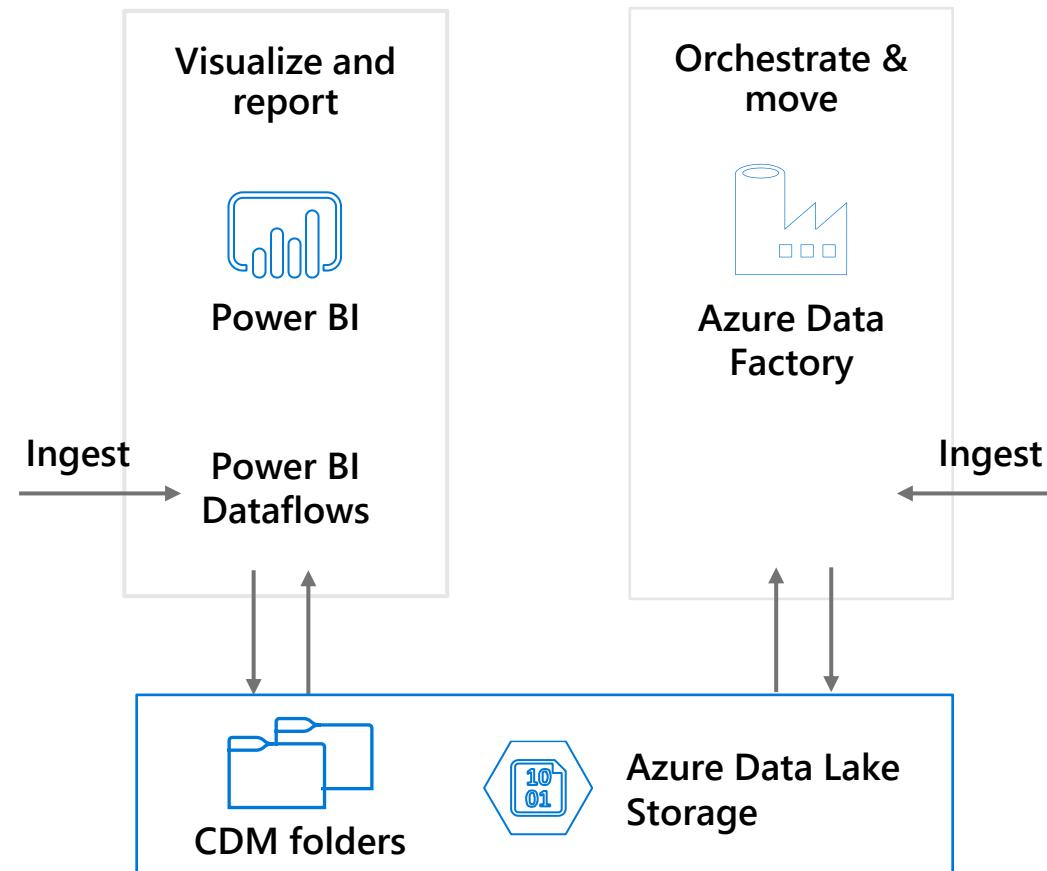
# Unify data across your organization

## Power BI + Azure Data Factory + Azure Data Lake Storage

### Power BI Dataflows

Business analysts can create reusable ETL logic to ingest any data source with point and click simplicity

Can be configured to automatically store data in CDM format on Azure Data Lake Storage



### Integration Runtime

IT professionals can run SSIS packages natively in Azure Data Factory, so you don't have to recreate years of on-premise ETL logic

**One lake for all data** enables business analysts, IT professionals, and data scientists to collaborate without having to move the data around

# Empowering 40,000 employees to work smarter



Unifying data and unlocking insights with Power BI and Azure

## Challenge

Global consumer goods company struggled with poor performance on their existing business intelligence solution.

Disconnect between data storage and consumption.

## Impact

**Unified internal and third party data stores** with Azure Data Factory and Azure Data Lake Storage.

**Easily accessible insights on both desktop and mobile** with Azure Synapse Analytics and Power BI.

**Increased cost savings on licensing and maintenance** which paid for the new platform in the first year.

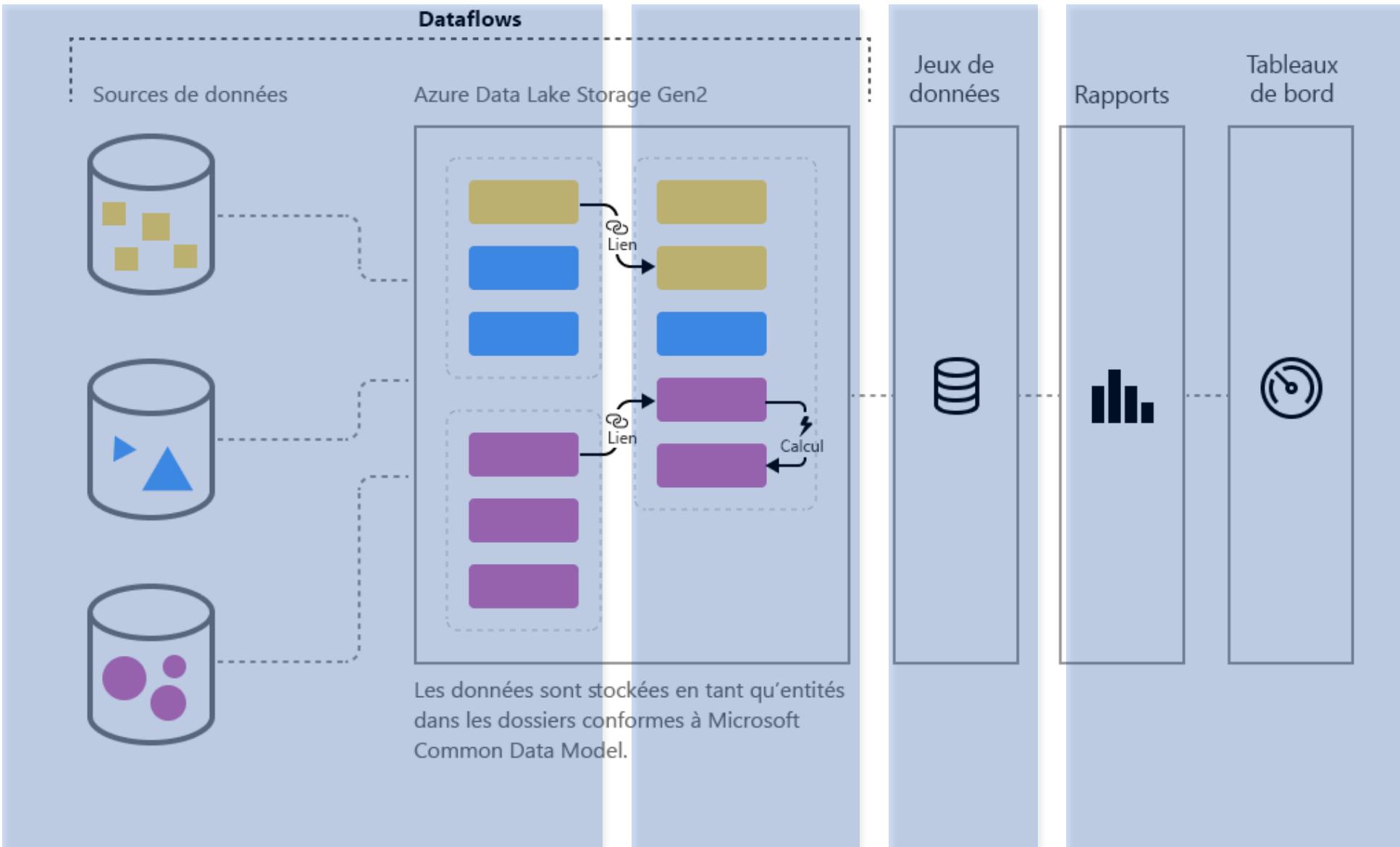




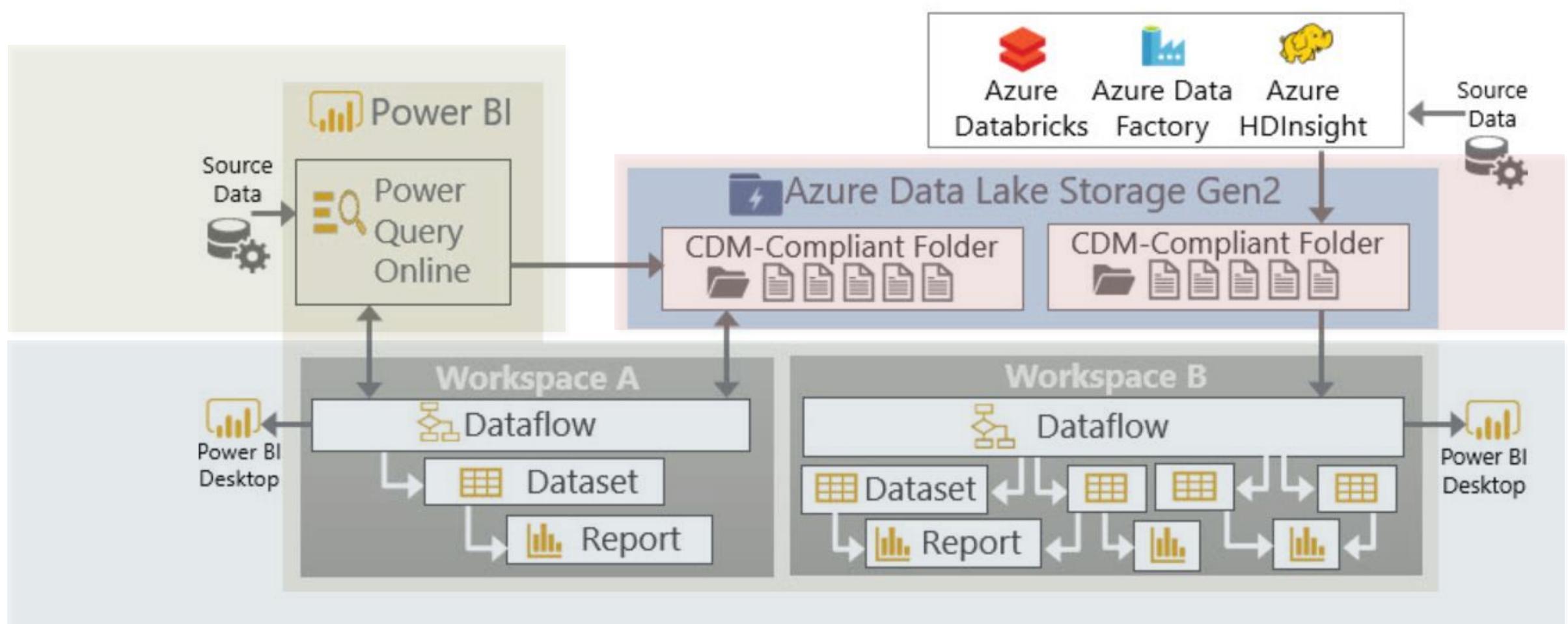
# Data Producer / Data consumer

Data Analysts

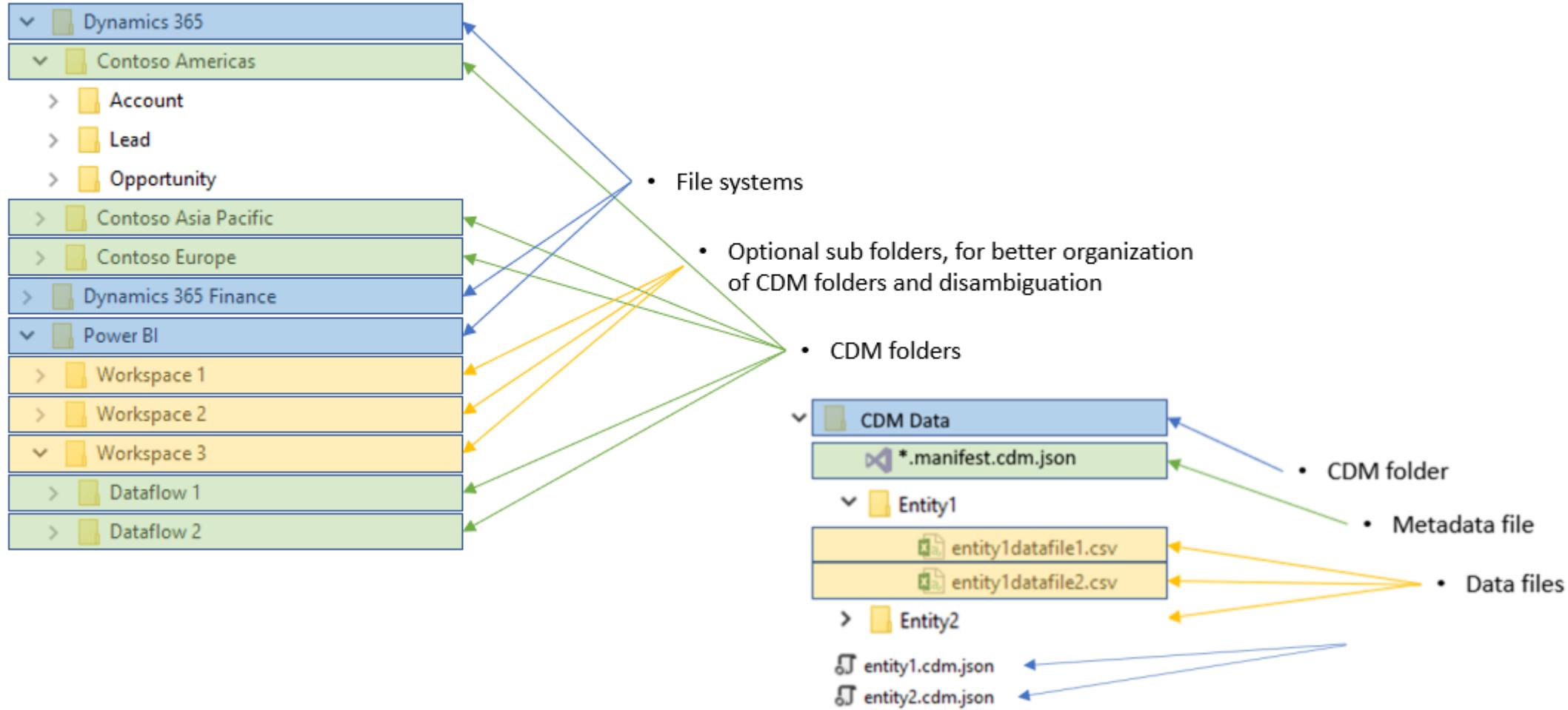
INGEST → PREPARE → SERVE → VISUALIZE



# Data producer / Data consumer – Azure Data services



# Azure Data Lake Storage Gen2 Common Data Model



# Common Data Model = Power BI + Synapse Pipelines

Create pipelines to ingest, transform and load data with 90+ inbuilt connectors.

Offers a wide range of activities that a pipeline can perform.

The screenshot shows the Azure Synapse Pipelines interface. On the left, there are three panels: "Move & transform" (Copy data, Data flow), "Machine Learning" (ML Batch Execution, ML Update Resource, ML Execute Pipeline), and "Synapse" (Notebook, Spark job definition, Stored procedure). Red arrows point from each of these panels to the corresponding activity types in the main pipeline editor. The main editor shows "Pipeline 2" with a "Stored procedure" activity followed by a "Notebook" activity. The "Settings" tab is selected, showing configuration for the stored procedure sink, including Entity reference type (Custom), Schema linked service (bronzelayer), Corpus folder (datafactory / salesEntities/salesPerfByYear), Entity (salesPerfByYear.cdm.json/salesPerfByYear), File settings (Root location, Manifest file, Partition path), Format settings (Format type Parquet), and a "Clear the folder" checkbox.

Orchestrator

Pipelines

Activities

Move & transform

Copy data

Data flow

Machine Learning

ML Batch Execution

ML Update Resource

ML Execute Pipeline

Synapse

Notebook

Spark job definition

Stored procedure

Pipeline 2

Stored procedure

sql1\_dbo\_StorePredictions

Notebook

BOOT\_Basic\_spark

Validate

Debug

Trigger (1)

Entity reference

Custom

Standard

bronzelayer

Test connection

Edit

New

datafactory

/ salesEntities/salesPerfByYear

Browse

salesPerfByYear.cdm.json/salesPerfByYear

Browse

datafactory

/ salesEntities/salesPerfByYear

Browse

Manifest file

Root location / Entity path / salesPerfByYear.cdm.json

Browse

Partition path

Root location /

Browse

Clear the folder

DelimitedText

Parquet

# Common Data Model = Power BI+ Spark

Microsoft Azure | Synapse Analytics | Search

Synapse live | Validate all | Publish all

L script 12 | SQL script 13 | Monitoring -1- see Q... | Monitoring -2- create... | Monitoring -3- create...

Cell | Run all | Publish | Attach to dev | Language | PySpark (Python) | ...

Not started

Command executed in 18s 287ms on 11-19-2020 12:43:31.546 +01:00

```
[4] taxi_df = sampled_taxi_df.select('totalAmount', 'fareAmount', 'tipAmount', 'paymentType', 'ra', 'tripDistance', 'tpepPickupDateTime', 'tpepDropoffDateTime', date_format('tpepPickupDateTime', 'hh').alias('pickupHour'), date_format('tpepPickupDateTime', 'EEEE').alias('weekdayStr'), (unix_timestamp(col('tpepDropoffDateTime')) - unix_timestamp, (when(col('tipAmount') > 0, 1).otherwise(0)).alias('tipped')))\.filter((sampled_taxi_df.passengerCount > 0) & (sampled_taxi_df.passe & (sampled_taxi_df.tipAmount >= 0) & (sampled_taxi_df.tipArou & (sampled_taxi_df.fareAmount >= 1) & (sampled_taxi_df.fareAm & (sampled_taxi_df.tipAmount < sampled_taxi_df.fareAmount)\& (sampled_taxi_df.tripDistance > 0) & (sampled_taxi_df.tripD & (sampled_taxi_df.rateCodeId <= 5) & (sampled_taxi_df.paymentType.isin("1", "2")))
```

Command executed in 2s 26ms on 11-27-2020 16:26:11.918 +01:00

```
[5] taxi_featurised_df = taxi_df.select('totalAmount', 'fareAmount', 'tipAmount', 'paymentType', 'tripDistance', 'weekdayString', 'pickupHour', when((taxi_df.pickupHour <= 6) | (taxi_df.p .when((taxi_df.pickupHour >= 7) & (taxi_df.pi .when((taxi_df.pickupHour >= 11) & (taxi_df.p .when((taxi_df.pickupHour >= 16) & (taxi_df.p .otherwise(0)).alias('trafficTimeBins')))\.filter((taxi_df.tripTimeSecs >= 30) & (taxi_df.tripTi
```

Command executed in 2s 34ms on 11-27-2020 16:26:13.984 +01:00

```
[6] display(taxi_featurised_df)
```

Synapse SparkPool

Microsoft Azure | Databricks | Portal | albouhad@microsoft.com

1-SalesEntity (Python)

Detached

Home | Workspace | Recent | Data | Clusters | Jobs | Models | Search

check bronzedatasets storage account

Cmd 5

```
1 %python
2 (sales2010.write.format("com.microsoft.cdm")
3 .option("storage", "bronzedatasets.dfs.core.windows.net")
4 .option("appId", dbutils.secrets.get(scope = "dataflow", key = "cdm-spname"))
5 .option("appKey", dbutils.secrets.get(scope = "dataflow", key = "cdm-spnsecret"))
6 .option("tenantId", dbutils.secrets.get(scope = "dataflow", key = "cdm-tenantid"))
7 .option("manifestPath", "databricks/raw/Sales2010/Sales2010.manifest.cdm.json")
8 .option("entity", "Sales2010")
9 .option("format", "csv")
10 #.option("compression", "snappy")
11 .mode("overwrite")
12 .save()
```

▶ (1) Spark Jobs

Command took 1.40 minutes -- by albouhad@microsoft.com at 10/6/2020, 2:47:27 PM on Spark2.4.5

Cmd 6

calculate store perf by year

Azure Databricks

# Common Data Model = Power BI+ AzureML

The image displays three screenshots illustrating the integration of Power BI and Azure ML:

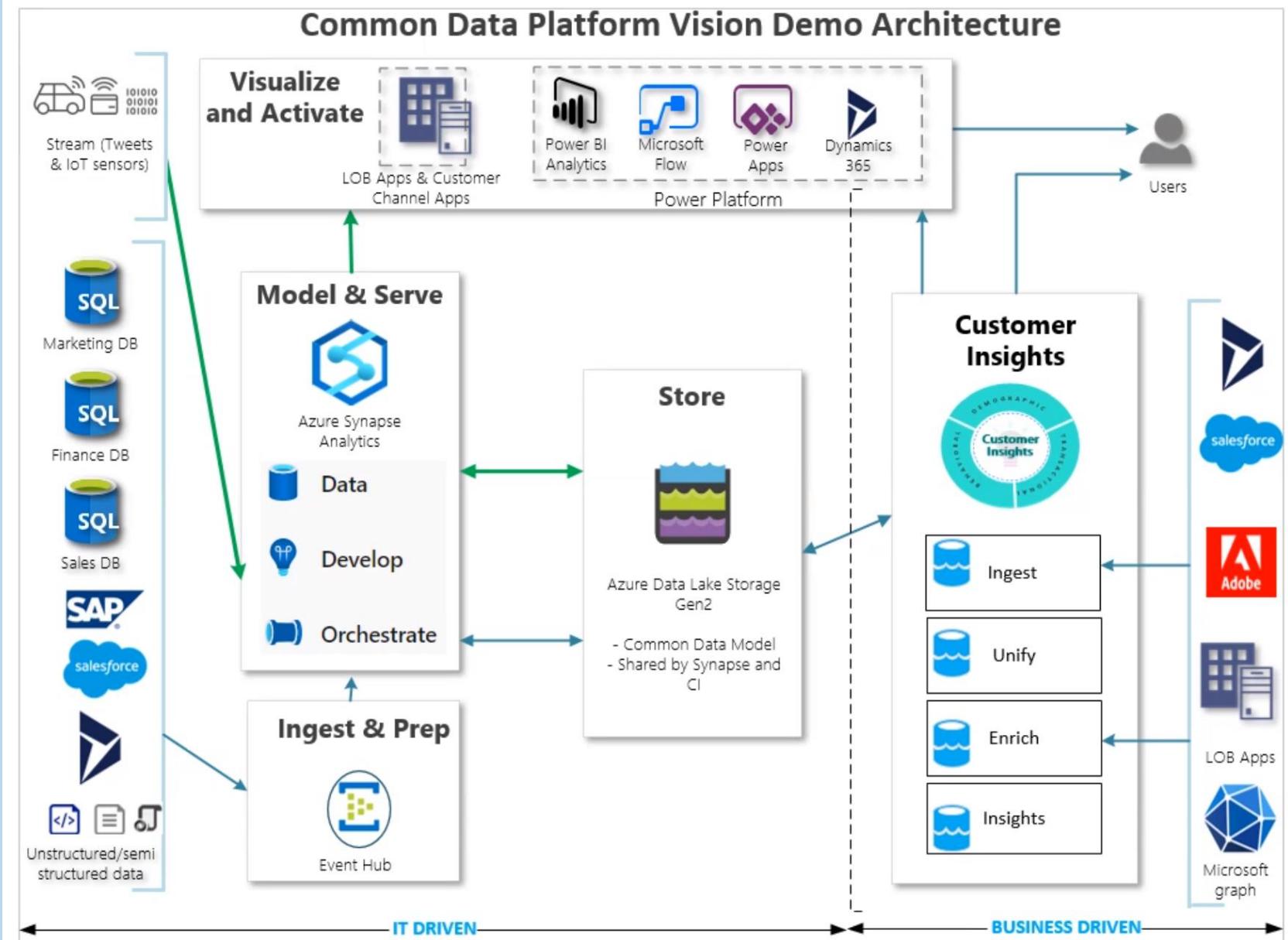
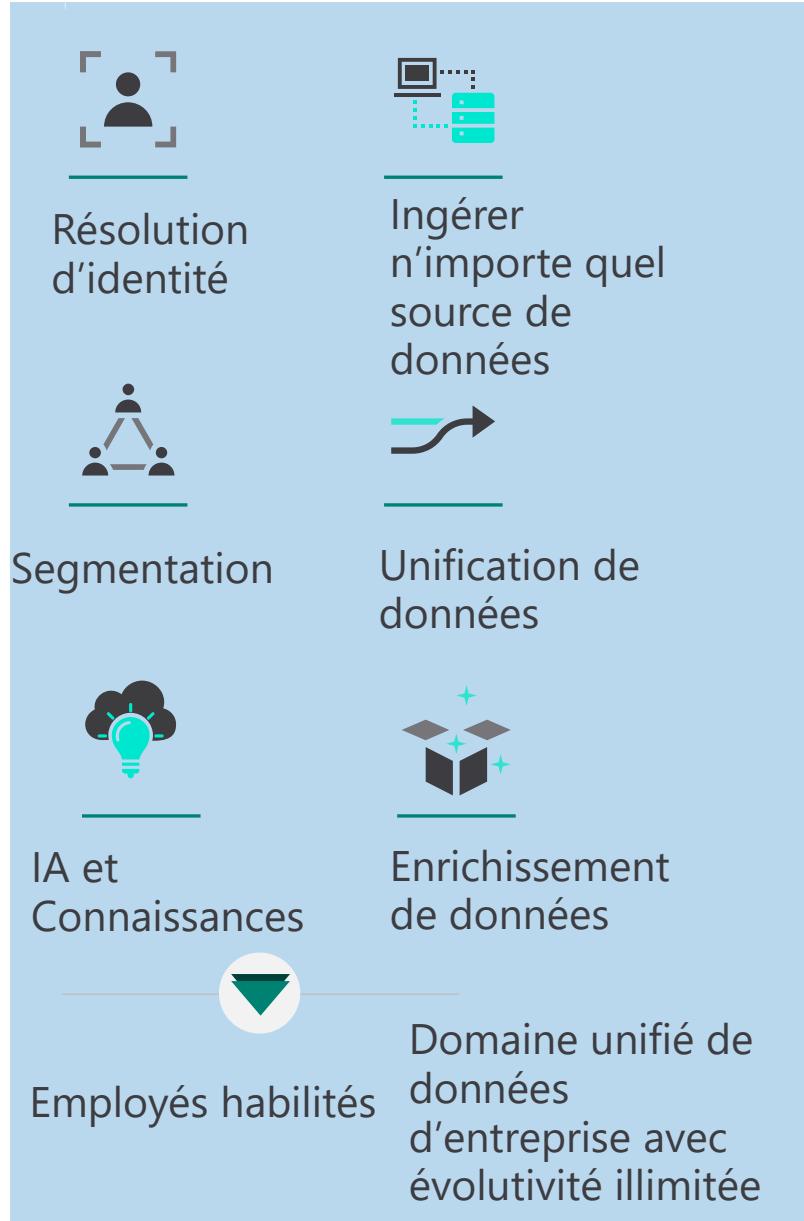
- Power BI Automl**: A screenshot of the Power BI interface showing the "Machine learning models" section. It provides a step-by-step guide: 1. Create and train your model, 2. Improve it, and 3. Apply it. Each step includes a icon and a brief description.
- Azure Machine learning studio**: A screenshot of the Azure Machine Learning studio home page. It features sections for "Create new", "Notebooks", "Automated ML", and "Designer". Below these are "Tutorials" and "Links" sections.
- Azure ML Designer**: A screenshot of the Azure ML Designer interface for a "Regression - Automobile Price Prediction" project. It shows a visual workflow for data preparation, including steps like "Select Columns in Dataset", "Clean Missing Data", "Linear Regression", "Split Data", "Train Model", "Score Model", and "Evaluate Model".

Power BI Automl

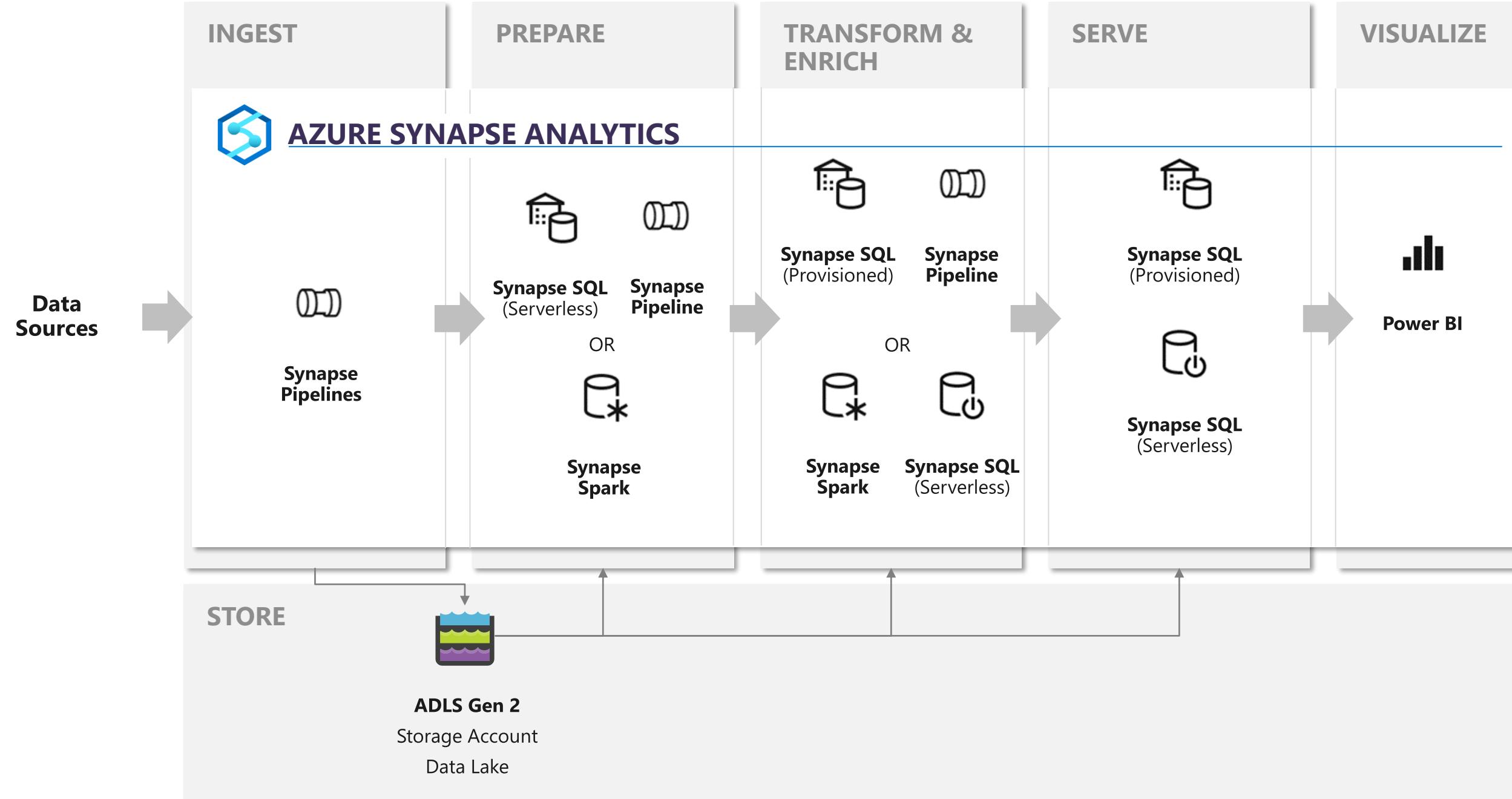
Azure Machine learning studio

Azure ML Designer

# Common Data Model = Synapse + Power BI+ Customer Insight



# Common Data Model = Power BI+ Synapse Analytics



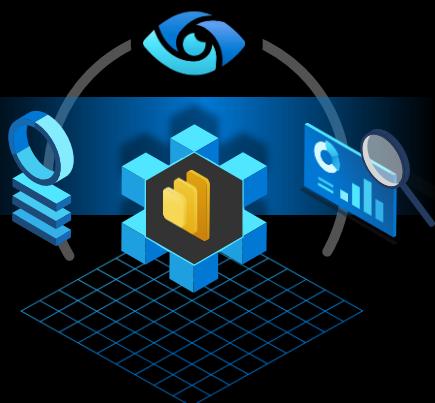
# Azure Purview and Power BI

Unified Data Governance  
to Maximize the Business  
Value of Data



# Azure Purview and Power BI: Better Together

Connect in a few clicks and automatically scan Power BI using Azure Purview to:



Search and browse  
Power BI assets in  
Azure Purview



Trace end-to-end  
lineage of Power BI  
reports



Govern sensitive data  
being used for insights  
and visualization



Open data sets from  
Azure Purview  
directly in Power BI

# Azure Purview and Power BI: Maximize the Value of your data



Make data more discoverable with catalog glossary and search



Improve trust in reporting with end-to-end Lineage



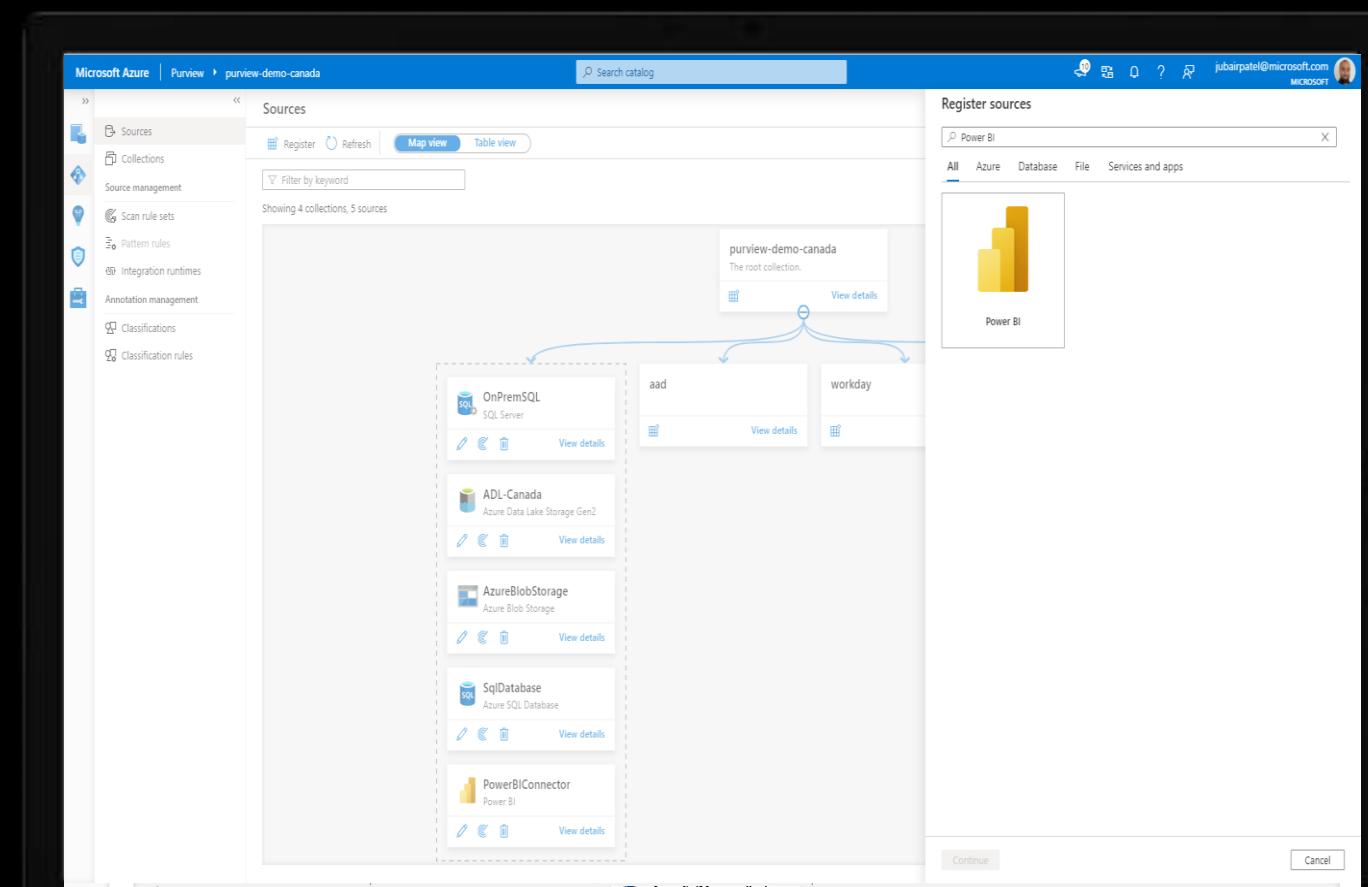
Secure sensitive data with classifications and insight reporting



Improve user experience with native integration

# Connect in a few clicks and Automatically scan data in Power BI

- Native connection with just a few clicks - no infrastructure or code setup
- Scan huge numbers of Power BI assets quickly with auto scaled performance
- Easily set up scheduled and incremental scans to keep information up-to-date



# Search and browse Power BI assets in Azure Purview

- Search across all Power BI assets using business glossary and keyword search
- Govern column, dashboard, dataset and workspace level details in Purview
- Set ownership for governance of Power BI assets

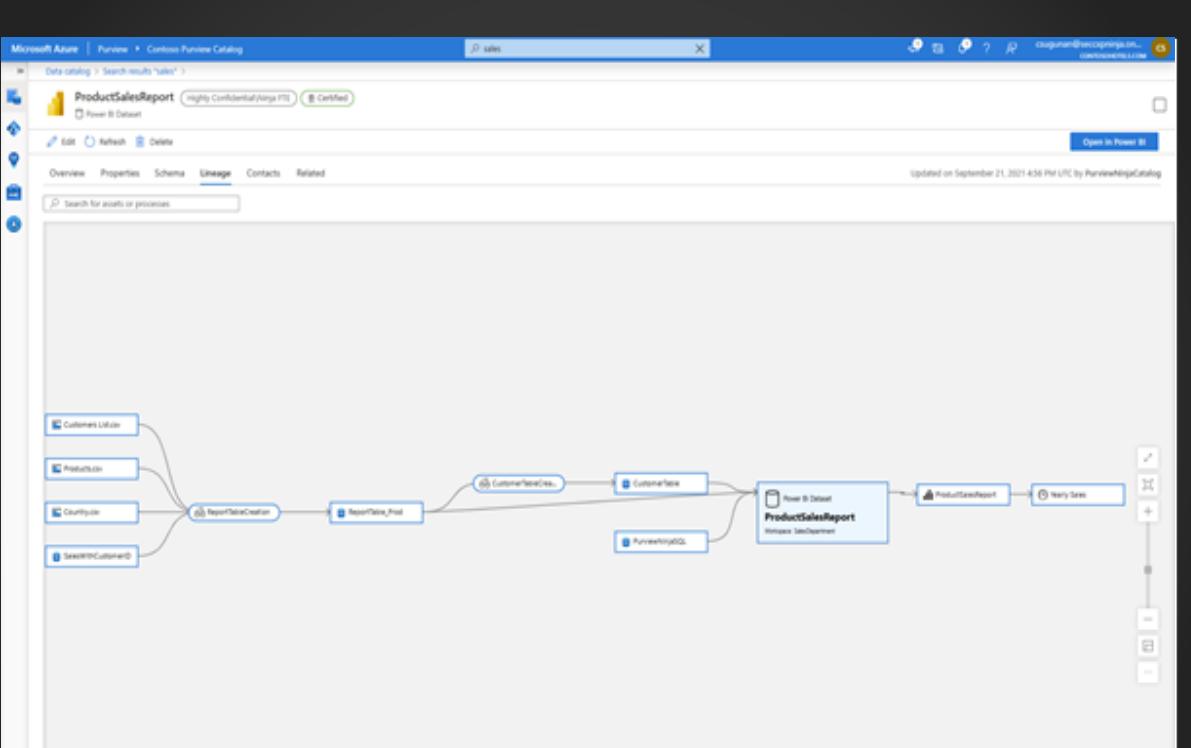
The screenshot shows the Microsoft Azure Purview Catalog interface. In the top navigation bar, 'Purview' is selected under 'Content'. The main search bar has 'Sales Department' typed into it. Below the search bar, there are several filter options: 'Status type: Power BI', 'Instance: Sales Department', 'Clear filters', 'Filter by keyword', 'Collection: Content Purview Catalog', and 'Classification' (which includes 'Country/Region', 'Framed Product', 'Imperial Unit', 'Person's Name', and 'Canada Social Insurance Number'). There is also a 'Clicked' section listing users like Franck Mercier, Will Johnson, and others. The search results are titled 'Browsing 14 out of 4 Results' and include the following items:

- Yearly Sales** [Highly Confidential, Vanya PTE] Power BI Dashboard. A thumbnail image of a bar chart.
- Sales Dashboard-SQLODB** [Power BI Report] Power BI Dataset. A thumbnail image of a bar chart.
- Sales Dashboard-SQLODB** [Highly Confidential, Vanya PTE] Power BI Report. A thumbnail image of a bar chart.
- ProductSalesReport** [Highly Confidential, Vanya PTE] Power BI Dataset. A thumbnail image of a bar chart.
- ProductSalesReport** [Highly Confidential, Vanya PTE] Power BI Report. A thumbnail image of a bar chart.
- SalesDepartment** [Power BI Workspace] Power BI Workspace. A thumbnail image of a bar chart.

At the bottom of the results, there are navigation buttons for 'Previous', 'Page 1', 'Next', and 'Last'.

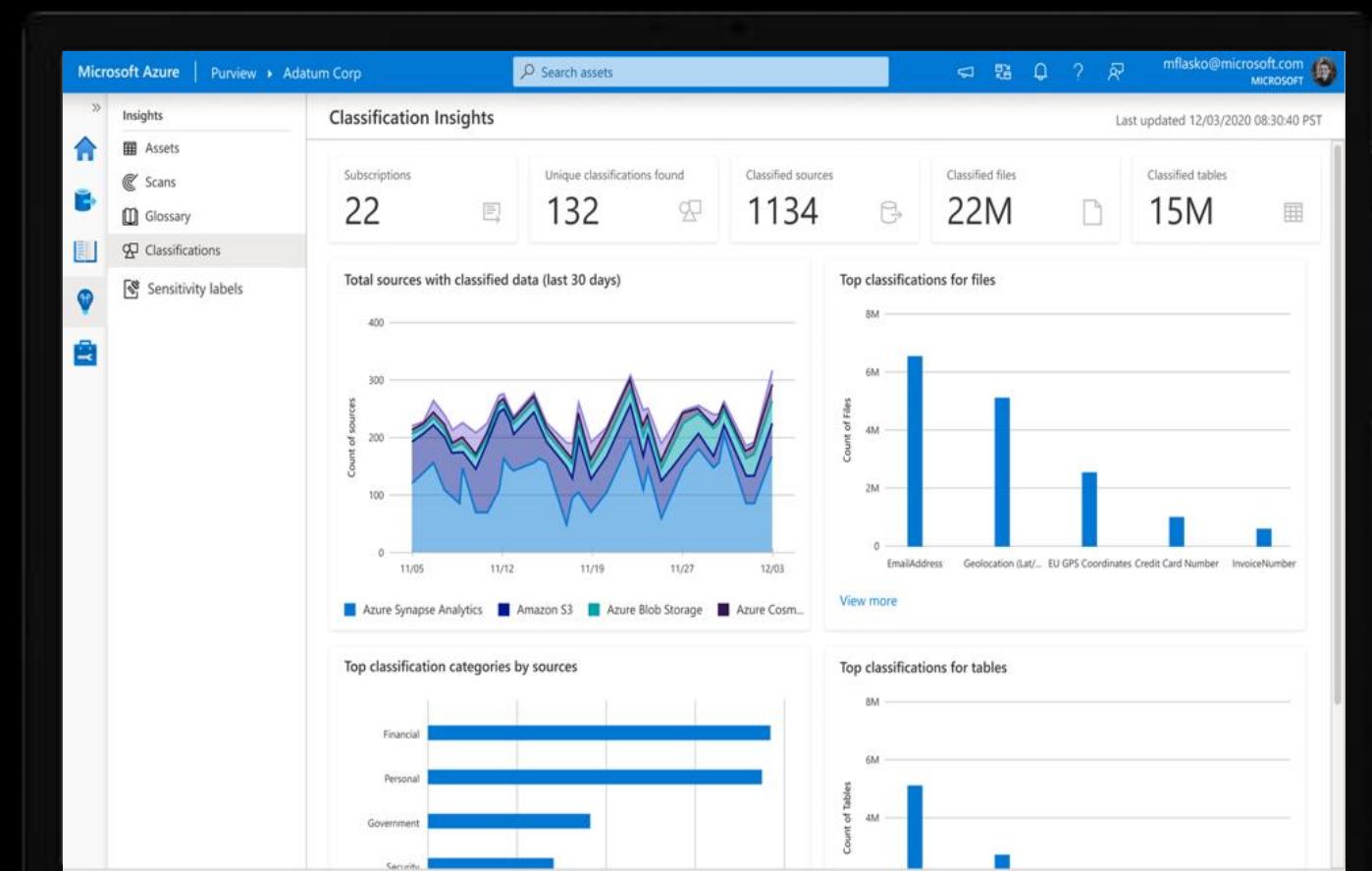
# Trace End to End lineage of Power BI Reports

- Automatically scan lineage of Power BI assets and connect with Azure data lake lineage
- Perform forward and backward-looking scenarios for impact analysis
- Lineage visualization at column level



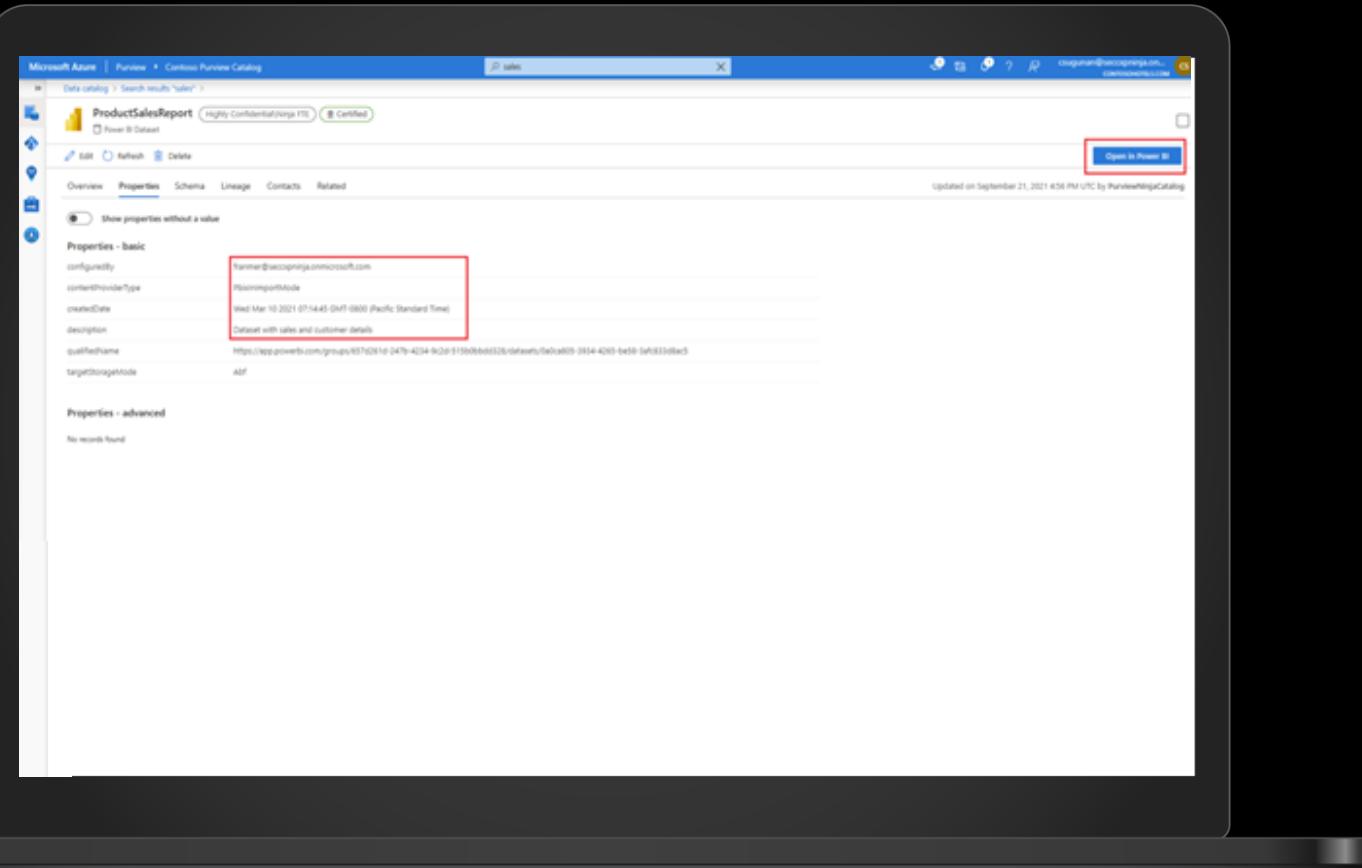
# Govern sensitive data being used for insights and visualization

- Automatically scan MIP sensitivity labels from Power BI into Purview
- Apply classifications to Power BI assets in Purview
- Apply ownership to Power BI assets
- Pre-built reports to show how sensitive data is being used

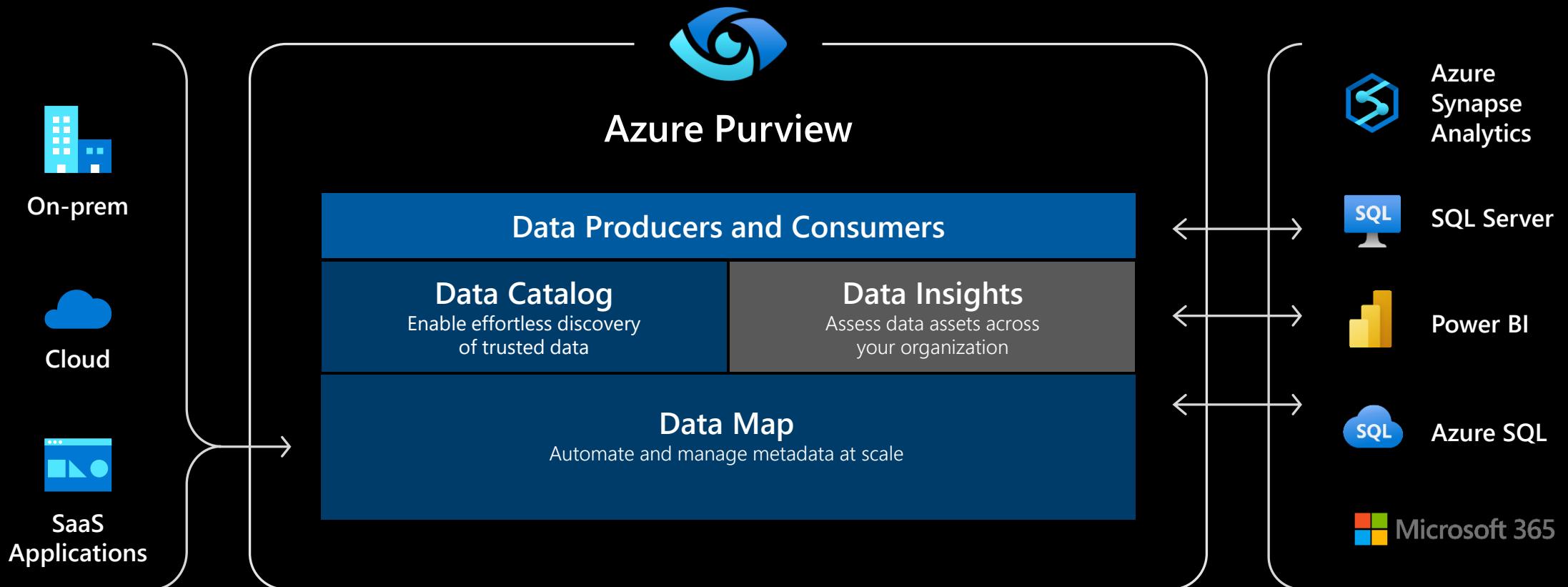


# Open Data sets from Purview directly in Power BI

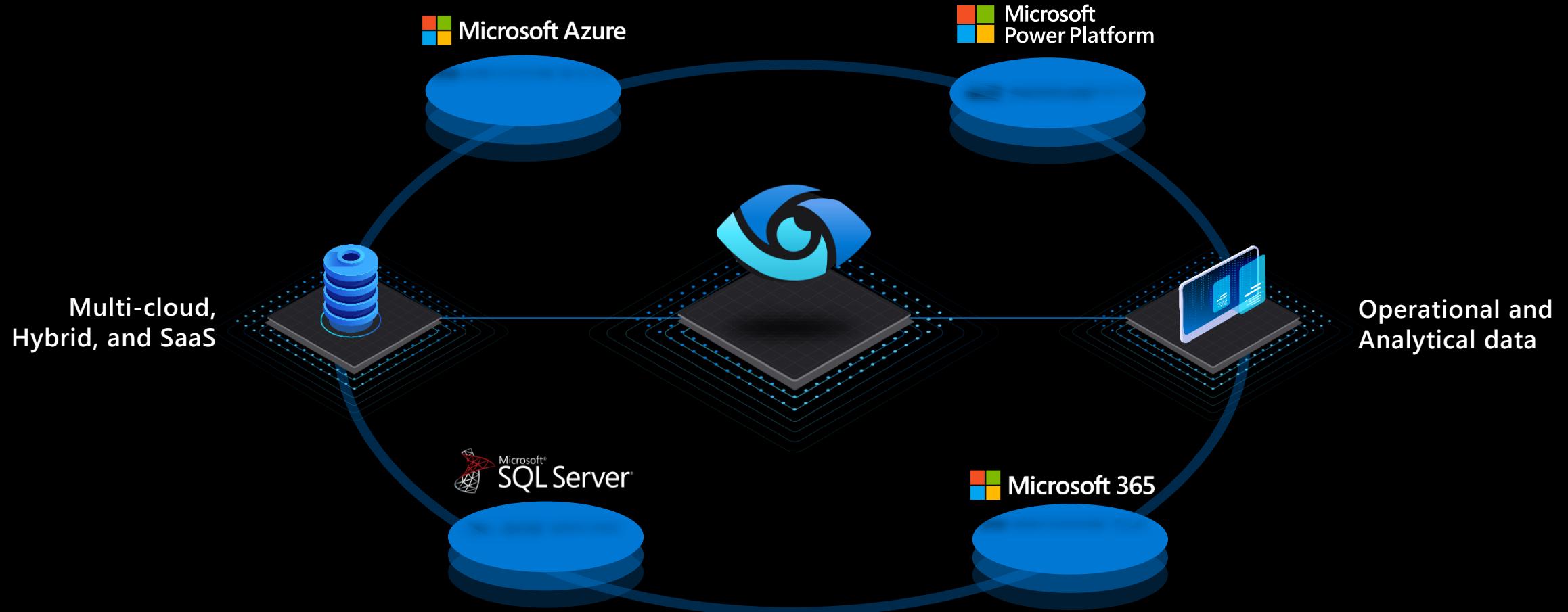
- One-click to open data sets from the Purview Data Catalog in Power BI
- Open Azure SQL and Azure Synapse datasets in Power BI securely



# Unified Data Governance with Azure Purview



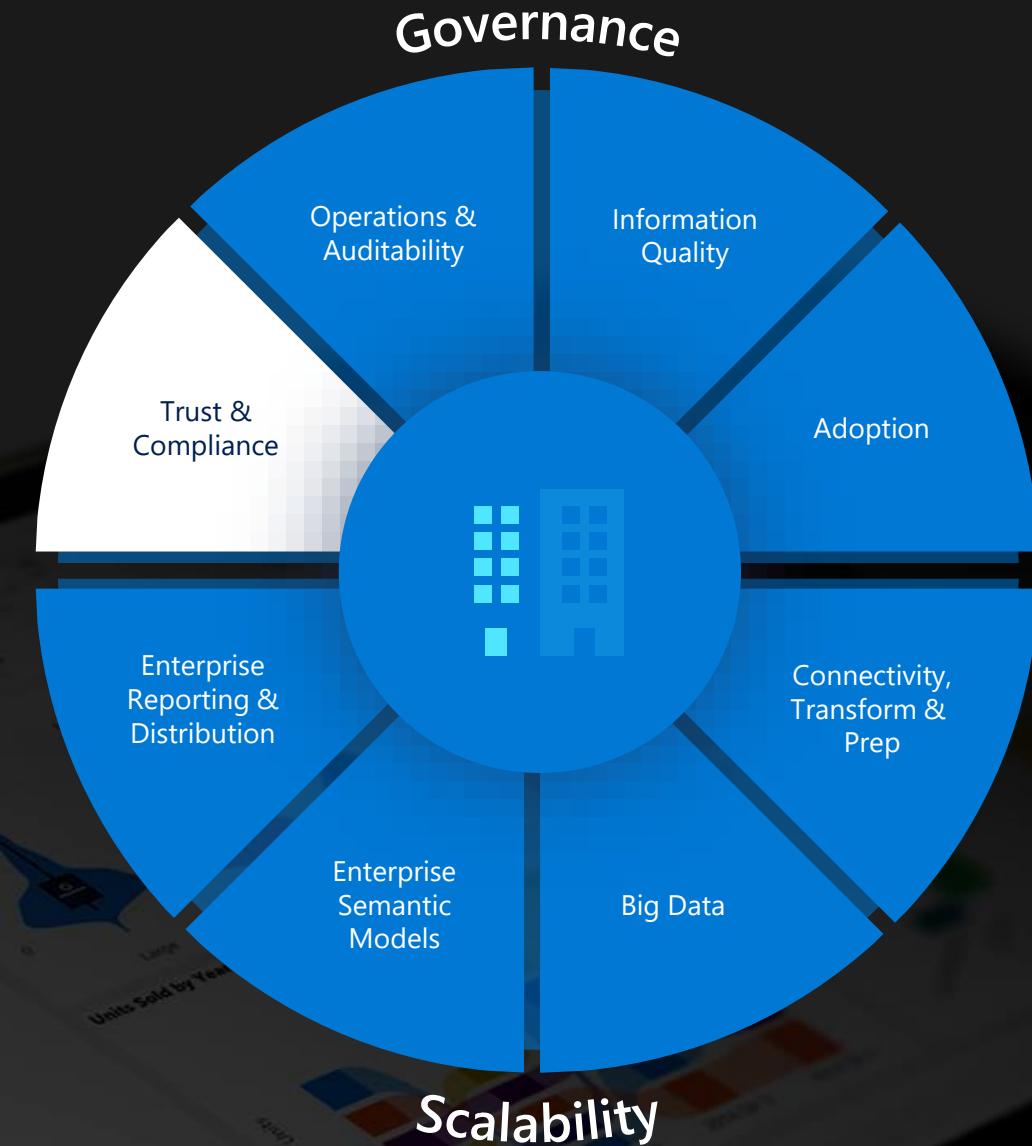
# Make the most of your Microsoft investments





# Empowering every organization

## Trust & compliance





# Power BI security strategy

## Cloud & remote work

### Data residency

Large global footprint,  
offering storage in 54  
data centers WW

Multi-geo capacities

### Certifications

119 local and  
international  
compliance standards

34 local and  
international industry  
certifications

### Network security

Service tags  
Private links

VNet

TLS 1.2 enforcement

### Data encryption

BYOK encryption via  
Azure Key Vault  
  
Double encryption  
at rest

### Data loss prevention

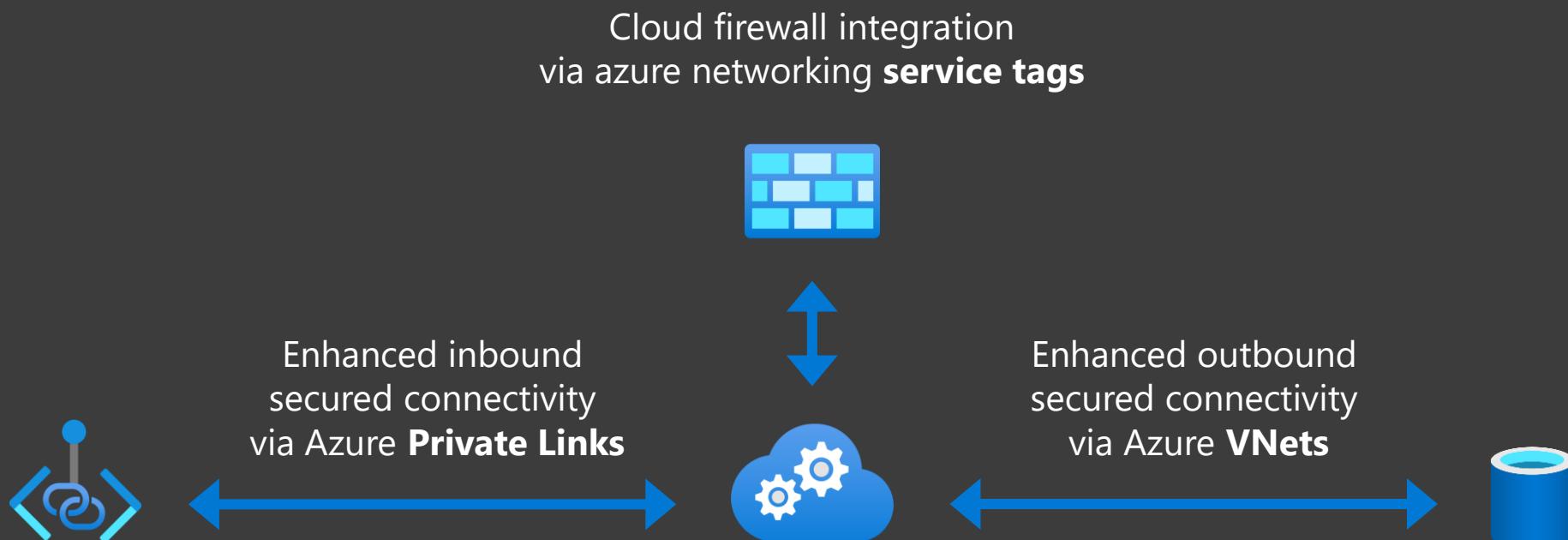
Classify and protect  
data with MIP  
  
Monitor and protect  
data with MCAS

Integrated security policy  
management, audit and  
compliance

Power BI tenant admin  
governance - permissions,  
settings, metrics

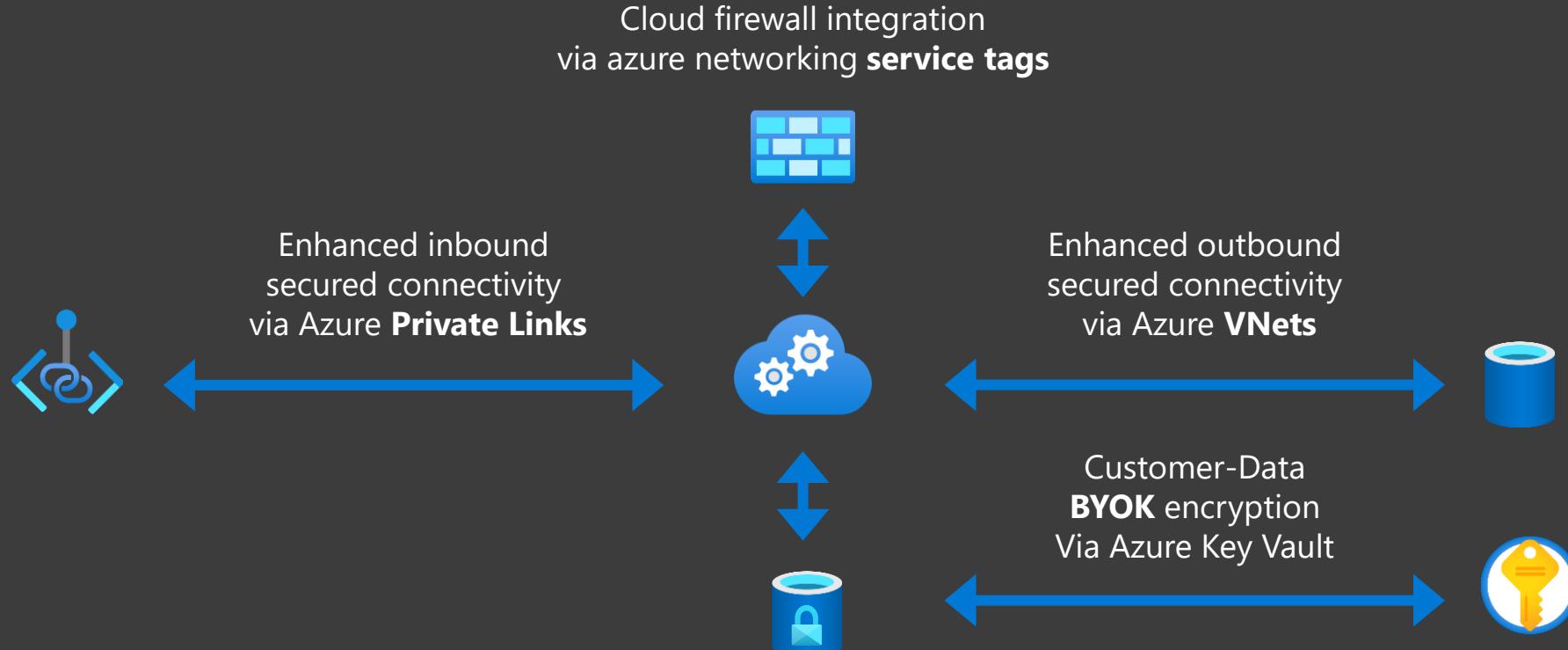


# Network security





# Encryption





Power BI ContosoFinanceDepartment

ContosoFinanceDepartment

Create app

+ New Create a pipeline

View Filters Settings Access Search

All Content Datasets + dataflows

Name	Type	Owner	Refreshed	Next refresh	Endorsement	Sensitivity	Include in app
Business Units & Regions	Report	ContosoFinanceDepa...	4/21/21, 10:38:10 AM	—	<input checked="" type="checkbox"/> Promoted	—	<input checked="" type="checkbox"/> Yes
Company Metrics	Dataset	ContosoFinanceDepa...	4/21/21, 10:38:10 AM	N/A	<input checked="" type="checkbox"/> Promoted	Confidential\Internal...	<input type="checkbox"/>
Executive Summary	Dashboard	ContosoFinanceDepa...	—	—	—	Highly Confidential\In...	<input checked="" type="checkbox"/> Yes
Finance	Report	ContosoFinanceDepa...	4/21/21, 10:38:10 AM	—	<input checked="" type="checkbox"/> Promoted	Highly Confidential\In...	<input checked="" type="checkbox"/> Yes
HR Metrics	Report	ContosoFinanceDepa...	4/21/21, 10:38:10 AM	—	<input checked="" type="checkbox"/> Promoted	Highly Confidential\In...	<input checked="" type="checkbox"/> Yes



Screenshot of Microsoft Excel showing the Data tab ribbon selected. The 'Get Data' button is highlighted in the 'From File' section of the ribbon.

The ribbon tabs include: AutoSave (Off), Home, Insert, Draw, Page Layout, Formulas, Data (selected), Review, View, Help.

The 'Data' tab ribbon groups include:

- Queries & Connections**: Recent Sources, Existing Connections, Refresh All, Properties, Edit Links.
- Data Types**: Stocks, Geography.
- Sort & Filter**: Sort (A-Z, Z-A), Filter, Clear, Reapply, Advanced.
- Data Tools**: Text to Columns, What-if Analysis, Forecast Sheet, Group, Ungroup, Subtotal, Outline, Automation.
- Forecast**: What-if Analysis, Forecast Sheet.
- Outline**: Group, Ungroup, Subtotal.
- Automation**: Flow.

The main Excel interface shows a blank worksheet titled 'Sheet1'. The columns are labeled C through T, and rows 12 through 26 are visible in the left margin. The status bar at the bottom indicates 'Ready' and 'Accessibility: Good to go'.



Microsoft Azure | Purview >

Search assets

Browse assets > Azure Dedicated SQL Pool (formerly SQL DW) > 02\_company\_> ip\_address

ip\_address Highly Confidential/Internal Only

Azure Dedicated SQL Pool Column

Edit Refresh

Overview Lineage Contacts Related

Description

No description for this asset.

Last updated

02/18/2021 09:27:54 UTC by [redacted]

Classifications (2)

IP Address Personal IP Address

Properties

userTypeId	167
columnEncryptionKeyDatabaseName	
replicatedTo	
replicatedFrom	
qualifiedName	mssql://.database.windows.net/02_company_ /dbo/PII_Customers#ip_address
precision	0
length	20
encryptionType	0
columnEncryptionKeyId	0
description	
scale	0
is XmlDocument	false
is Masked	false
encryptionTypeDesc	
xmlCollectionId	0

Hierarchy

Glossary terms

No glossary terms for this asset.



Power BI Sales HQ US Retail Sales Report | Highly Confidential\Int... ▾

Search ...

File Export Share Chat in Teams Comment Subscribe Edit ...

☰

Contoso

Revenue: \$23.94M (Goal: \$20.00M)

Gross Margin: 5.27% (Goal: 3.5%)

Inventories: \$2.27M (Goal: \$2.15 M)

Total Store Count: 125

Region: United States

Region Manager:

- Select all
- Isaiah Langer
- Jeff Hay
- Joni Sherman
- Molly Clark
- Renee Lo

State: All

Top Retailers Operated by 3rd Party:

Retailer	Revenue
City Power & Light	2.02M
Proseware, Inc.	1.52M
Wide World Importers	0.58M
The Phone Company	0.51M

Top Retailers Leased and Operated by 3rd Party:

Retailer	Revenue
Proseware, Inc.	1.821.2K
Wide World Importers	831.8K
The Phone Company	217.6K
City Power & Light	67.0K
Wingtip Toys	60.0K
Coho Winery	15.0K

Revenue by State:

Revenue by Region Manager:

New Store Opportunity:

Stage	Value
Qualify	\$30M
Propose	\$67M
Develop	\$40M
Close	\$0M

Target Consumer Size by Market:

Market	Size
Entertainment Retail	15K
Consumer Services	12K
Food Services	2K
Automotive	2K

Store Status by Region Manager:

Store Status:

Legend:

- Increase
- Decrease
- Total



# Empowering every organization

## Operations & auditability





# Capacity utilization

## Utilization metrics



Microsoft Power BI | Utilization Report Demo | Data updated 2/10/21 ▾

Search Bookmarks View

Reset to default

Home Favorites Recent Create Datasets Apps Shared with me Deployment pipelines Learn Workspaces 2 - Demo Gen2

Premium capacity usage  
P1-002C84C3-2698-497E-AC63-DB7E93880073  
<https://pbipeu.kusto.windows.net>

Date  
Last 4 Weeks  
1/14/2021 - 2/10/2021

The visualizations below show the daily peak usage of your Premium capacity as a percentage of the capacity's cores, as well as the workspaces and operations that contributed to the usage. Click on a day to see how the usage was spread across the day.

Daily peak usage

Premium Cores Auto-scale Cores Total 100%

Core Count

Feb 02 Feb 03 Feb 04 Feb 05 Feb 06 Feb 07 Feb 08 Feb 09 Feb 10

Peak usage - 15m intervals

Ctrl + click on any data point to see which workspaces and artifacts were in use during that time.

Total by Workspace ID and Operation

Operation (Blank) Dataflow Refresh Dataset On-Demand Refresh Dataset Scheduled Refresh Query Render XMLA Read Operation

Workspace ID

B9D9838F-8422-4E13-B3B5-3306EB4... 39.07%  
BE94FF18-4389-4FF2-A675-9400F349... 12.99%  
82D8D21E-1DC8-407F-8570-81986E0... 11.94%  
1984D987-6B28-4D3B-81A5-FE6B4D...  
324AD25C-9E43-4908-B16B-3EBF4A8...  
DE24AB1A-5D1E-4F19-86B8-7E40800...  
0dc497a6-b138-4363-aee7-462639c3...  
B8DA9835-48EB-4D49-88C4-675F801...

% of Total

Workspaces Artifacts Identity Interactive usage (s) Background usage (s) Total usage (s)

Workspace ID: All Operation: All

Total Time Interactive Background

Click a column to the left to see details



# Performance management



Microsoft Power BI BYOLA Template Apps - AS Engine DEV Power BI Log Analytics for Analysis Services Engine(Option1) | General

Search Reset to default Workspace(User) View

Pages Workspace summary

Engine activities Dataset refreshes Query statistics User activities Error summary Help

Date: 14-May-2021 - 21-May-2021 Premium capacity: All Workspace: All Dataset: All Report id: All User: All

Filters

Top users with high CPU usage

User	CPU Time (ms)
NT AUTHORITY\SYSTEM	~28K
@microsoft.com	~12K
@microsoft.com	~5K

Which users/reports have the most impact on load?

Top 10 datasets by query executions

Dataset	Users	Query Executions	Duration (ms) 50th Percentile	Duration (ms) 90th Percentile
Auto-aggs-user-testing	1	2	0	5
P&G Demo v7	3	113	0	3,100

How many data refreshes failed /succeeded over time?

Dataset refresh success vs failures

Status: Succeeded

Refreshes

Date

15-May-21

Operations by segment

Operations

Segment

[100, 300ms] [10s, 30s] [1s, 3s] [300ms, 1s] [30ms, 100ms] [3s, 10s] <30ms >30s

Operations by date and segment

Segment

Operations

Date

May 14 May 15 May 16 May 17 May 18 May 19 May 20 May 21



# Tracing & troubleshooting



Power BI Analysis Services Engine

Workspace summary Engine activities Dataset refreshes **Query statistics** User activities Error summary Help

Date: 21-Jun-2021 Premium capacity: All Dataset: All Report id: All User: All

Filters

How many queries are executed for given time by duration segments?

**Queries by date and segment**

Segment ● [100, 300ms] ● [10s, 30s] ● [1s, 3s] ● [300ms, 1s] ● [30ms, 100ms] ● [3s, 10s] ● <30ms ● >30s

Operations

Jun 21 Date

How many queries are using aggregation feature?

**Queries by aggregation usage**

IsAggregationUsed

- No
- Yes

Yes 63 (28.25%)

No 160 (71.75%)

What are the details by CPU variability?

**Query duration variability**

Capacity | Workspace | Dataset | Report | Query

	Count	Duration (ms)	Std Dev %
IB-88A1869C5295	3	11%	

AutoAggs

AutoAggs Demo v4

DEFINE VAR \_\_H0FilterTable = TREATAS((DATE(2...

What are the top N queries by duration / CPU variability?

**Top 5 queries by duration P50**

Capacity | Workspace | Dataset | Report | Query

	Count	Duration (ms) 50th Percentile	Duration (ms) 90th Percentile	Max Duration (ms)
70DF2531-C...	1	49.604	49.604	49.604

AutoAggs

AutoAggs Demo v5

DEFINE VAR \_\_DSOFilterTable = FILTER(...

Select N : 5



Office 365 | Security & Compliance

Home > Audit log search

## Audit log search

Need to find out if a user deleted a document or if an admin reset someone's password? Search the Office 365 audit log to find out what the users and admins in your organization have been doing. You'll be able to find activity related to email, groups, documents, permissions, directory services, and much more. [Learn more about searching the audit log](#)

**Search** [Clear](#)

**Results** 150 results found (More items available, scroll down to see more.)

[Hide filtering](#) [Export results](#)

Date	IP address	User	Activity	Item	Detail
2018-07-23 23:59:21	40.118.175.16:31760	Unknown	Accessed other entity type	Get Organization	"AdminCenter" @ "https://..."
2018-07-23 23:59:21	40.118.175.16:31760	Unknown	Accessed other entity type	Get Organization	"AdminCenter" @ "https://..."
2018-07-23 23:59:15	40.118.175.16:23425	Unknown	Accessed other entity type	Get Organization	"AdminCenter" @ "https://..."
2018-07-23 23:59:15	40.118.175.16:23425	Unknown	Accessed other entity type	Get Organization	"AdminCenter" @ "https://..."
2018-07-23 23:58:44	71.197.217.152:14473	BenW@BAPpartners.onmic...	Accessed other entity type	Configure Organization	"AdminCenter" @ "https://..."
2018-07-23 23:58:42	104.42.186.44:1249	Unknown	Accessed other entity type	Post UpdateInstanceSetting	"AdminApi" @ "https://ad..."
2018-07-23 23:58:42	104.42.186.44:1249	Unknown	Accessed other entity type	Post RetrieveInstancesOper...	"AdminApi" @ "https://ad..."
2018-07-23 23:58:41	104.42.186.44:1249	Unknown	Accessed other entity type	Post RetrieveInstancesOper...	"AdminApi" @ "https://ad..."
2018-07-23 23:56:08	71.197.217.152:14473	BenW@BAPpartners.onmic...	Accessed other entity type	Edit Organization	"AdminCenter" @ "https://..."
2018-07-23 23:53:54	40.83.145.50:27916	Unknown	Accessed other entity type	Post RetrieveInstancesOper...	"AdminApi" @ "https://bap..."

[Search](#) [+ New alert policy](#)



# Adoption tracking



Microsoft Power BI BYOLA Template Apps - AS Engine DEV

Usage Metrics Report | Data updated 6/16/21

Pages < File Export ... New usage report on Reset to default Usage by distribution View

## Usage Metrics (Multiple reports selected)

Report opens: 196

Report page views: 657

Unique viewers: 9

Report open trend: -42.3%

Report open requests:

Daily report opens Weekly report opens

Report viewers:

Daily report viewers Weekly report viewers

Report usage based on data from 5/18/2021 to 6/15/2021

Dataset last refreshed: 6/16/2021 8:16:15 PM (UTC)

5/16/2021 6/16/2021

Distribution Platform

Workspace

100%

Users Pages

User	Report opens	Total page views
user@contoso.microsoft.com	1	2
user@contoso.microsoft.com	58	164
user@contoso.microsoft.com	8	15
user@contoso.microsoft.com	66	202
user@contoso.microsoft.com	1	19
user@contoso.microsoft.com	11	94
user@contoso.microsoft.com	17	48
user@contoso.microsoft.com	27	91

Unnamed users are masked by the Power BI tenant administrator

Report Id  
Multiple reports selected



# Empowering every individual

## Augmented analytics

PowerPoint  
for data

Insights  
on the move

Augmented  
analytics





# Microsoft's AI Innovation

**5,000+**

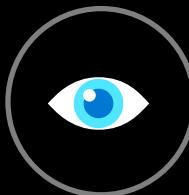
Researchers & Engineers

**4,000+**

WW Patents

**22,000+**

Papers



**2016**

Object  
recognition  
human parity



**2017**

Speech  
recognition  
human parity



**2018**

Reading  
comprehension  
human parity



**2018**

Machine  
translation  
human parity



**2018**

Speech synthesis  
near-human  
parity



**2019**

General  
Language  
Understanding  
human parity



**2020**

Document  
summary at  
human parity

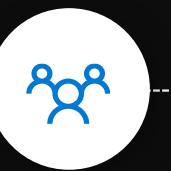
# The Most Complete AI Capabilities in a BI Product



## Data Scientists

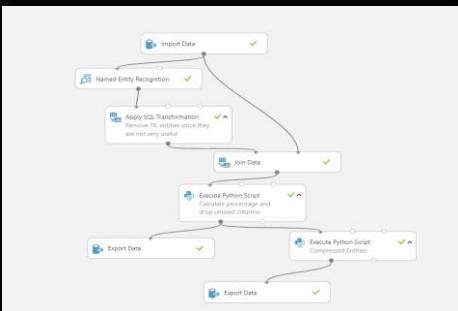


## Analysts

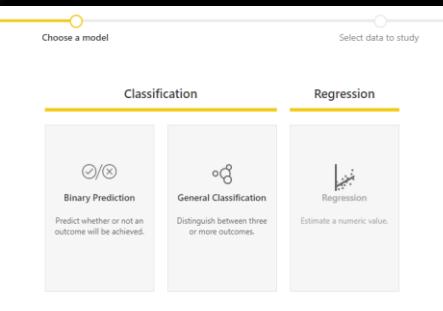


## End users

## Extend with Azure ML



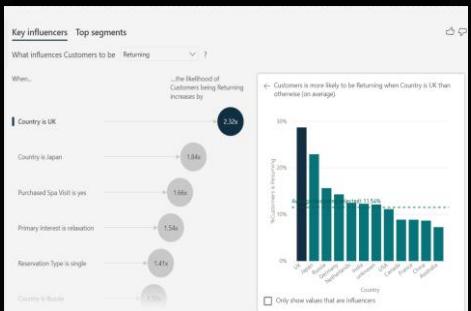
## Create ML models



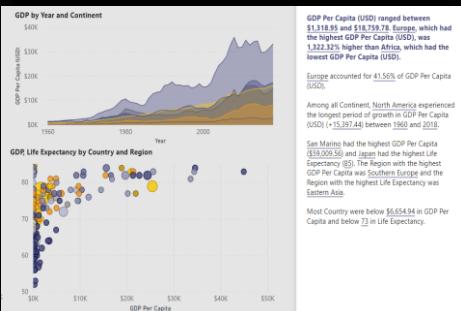
## Sentiment Analysis



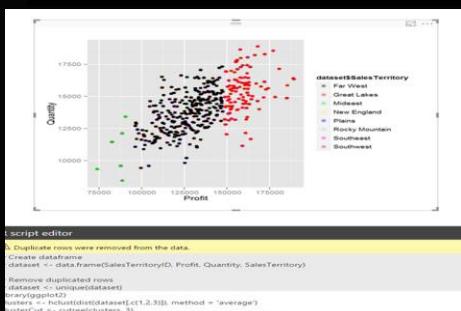
## Key Driver Analysis



## Smart Narratives



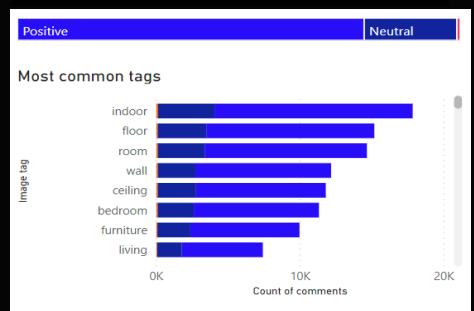
# Python & R Integration



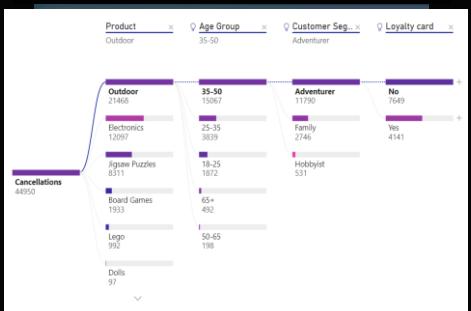
## Explore Predictions



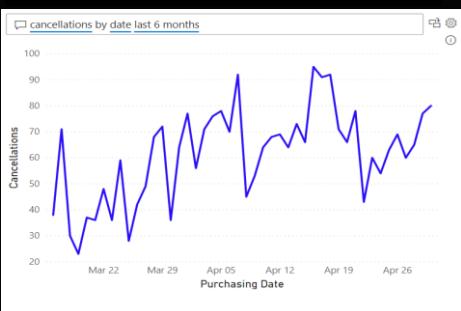
# Key Phrase Extraction



## Root Cause Analysis



Q&A





Online Sales - Power BI Desktop

Search

Justyna Lucznik

File Home Insert Modeling View Help

Get data Refresh New visual More visuals New measure Sensitivity (preview) Publish

### Product Analysis

Sales across time

Revenue

Sales by Product

Cancellations and Returns

Visualizations

Fields

Add data fields here

Drill through

Cross-report

Off

Keep all filters

On

Add drill-through fields here

Transactions Products Final Product Anomalies Sales across time Products Explore Cancellations Ask Questions +

Page 5 of 7

Detailed description: The screenshot shows a Power BI desktop workspace titled 'Online Sales - Power BI Desktop'. The ribbon at the top has tabs for File, Home, Insert, Modeling, View, and Help. The Home tab is selected. Below the ribbon are standard file operations like Get data, Refresh, and Publish. The main area contains four visualizations: 1) 'Sales across time' is a stacked area chart showing sales over time from Jan 2018 to Jan 2020 for five product categories: Board Games (red), Electronics (blue), Jigsaw Puzzles (purple), Lego (light blue), and Outdoor (dark blue). 2) 'Revenue' is a line chart showing revenue over time from Sep 2019 to Jan 2020, with values ranging from \$2K to \$6K. 3) 'Sales by Product' is a donut chart showing the percentage distribution of sales by product category: Outdoor (54.67%), Electronics (17.37%), Jigsaw Puzzles (12.98%), Board Games (11.96%), and Lego (0%). 4) 'Cancellations and Returns' is a bar chart comparing the number of cancellations and returns for five products: Outdoor, Electronics, Jigsaw Puzzles, Board Games, and Lego. The visualization pane on the right lists various visualization types and their properties.



Microsoft | Power BI | My workspace Supply Chain Demo | Data updated 10/12/20 ▾

File Export Share Chat in Teams Comment Subscribe Edit ...

Reset to default Bookmarks View

## Backorder Analysis

Ask a question about your data

Try one of these to get started

what is the % on time orders by demand type	what is the product availability by demand type	top plants by backorder \$
top plants by product availability	what is the % on backorder by region	what is the % on time orders by month
what is the backorder \$ by brand	what is the % on backorder by brand	what is the % on time orders by brand
sort backorder percentages by shipment type		Show fewer suggestions

% on backorder by Demand Type

Volatile (High Price)	Stable (Low Price)	Volatile (Lo...	Stable (Hig...
Intermittent	Cyclical	Growing	
Seasonal	No Segment	Declining	

Backorder \$ by Month

Backorder %

Month



Q&A setup

### Getting started

- Getting started
- Field synonyms
- Review questions
- Teach Q&A
- Manage terms
- Suggest questions

**Field synonyms**  
Add terms people might use as synonyms for the fields and tables in your data.

**Field synonyms**

**Teach Q&A**  
Teach Q&A to understand questions and terms people might use.

**Teach Q&A**

**Review questions**  
Review questions people have asked and fix misunderstandings.

**Review questions**

**Suggest questions**  
Help people explore your data by adding suggested questions.

**Suggest questions**

Help Q&A understand people better

The screenshot shows a 'Teach Q&A' interface with a search bar and a bar chart. The chart has several bars, with one bar for 'Teach Q&A' being circled in black.

[Learn more about Q&A](#)

This feature is in preview. [Learn more](#)



Microsoft Power BI My workspace Supply Chain Demo | Data updated 10/12/20

## Supply Chain Analytics

### Number of Products by Demand Type

Demand Type	Sub-Segment	Count
Stable (Low Price)	Growing	100
	Seasonal	80
Intermittent	Volatile (High Price)	20
	No Segment	10
Stable (High ...)	Declining	10
	Volatile (Low Price)	5

### Backorder \$ by Region

Region	Backorder \$
Midwest	\$45K
West	\$35K
Northeast	\$20K
Southwest	\$15K
Southeast	\$5K

### Key influencers Top segments

What influences Product to be on backorder

- When...
  - Demand Type is Volatile (High Price) → 2.32x
  - Demand Type is Volatile (Low Price) → 2.14x
  - Manufactured Goods is <70% → 1.73x
  - Forecast Bias is Accurate (5% to -5%) → 1.50x
  - Forecast Accuracy is Below 50% → 1.44x
  - Demand Type is No Segment → 1.35x
  - Demand Type is Growing → 1.27x

← Product is more likely to be on backorder when Demand Type is Volatile (High Price) than otherwise (on average).

Demand Type	%Product is on backorder
Volatile (High Price)	~28%
Volatile (Low Price)	~23%
No Segment	~16%
Growing	~14%
Seasonal	~13%
Declining	~12%
Cyclical	~11%
Intermittent	~11%
Stable (High Price)	~10%
Stable (Low Price)	~8%

Average (excluding selected): 11.34%

Only show values that are influencers



Microsoft Power BI My workspace Supply Chain Demo | Data updated 10/12/20 ▾

Search (2) ... ? ! ? ... User profile

File Export Share Chat in Teams Comment Subscribe Edit ...

Reset to default Bookmarks View Filters

### Root Cause Analysis

#### Average of Backorder % by Month

The chart displays the average backorder percentage for each month. The y-axis is labeled 'Month' and lists the months from September at the top to June at the bottom. The x-axis is labeled 'Backorder %' with tick marks at 0% and 5%. Each month has a dark blue horizontal bar. A tooltip for March shows a value of 5.07%.

Month	Backorder %
September	~6.5%
October	~5.5%
November	~4.5%
December	~4.5%
January	~6.0%
February	~5.5%
March	5.07%
April	~4.5%
May	~4.5%
June	~5.0%

High Risk Low Risk



Online Sales - Power BI Desktop Justyna Lucznik (MSIT)

File Home Insert Modeling View Help Format Data / Drill

Get data Excel Power BI datasets SQL Server Enter data Recent sources Transform data Refresh New visual Text box More visuals New measure Quick measure Publish

Product Analysis

Sales across time

Revenue

Drill through

Cross-report

Keep all filters

Add drill-through fields here

Transactions Products Products Final Sales across time Explore Cancellations Ask Questions Product Anomalies +



https://msit.powerbi.com/groups/c5135f90-6a81-4bf5-82b2-07ac37e20e93/dataflows/d06c809d-c317-4814-8df2-e181d1be5fe

Power BI Hotel Analytics > Hotel Reviews

Edit queries

Power Query

Get data Refresh Options Manage columns Transform table Reduce rows Add column AI insights Map to standard Combine tables

AI insights [2] Hotel Reviews

= Table.AddColumn(#"Invoked CognitiveServices.ScoreSentiment", "CognitiveServices.ExtractKeyPhrases", each CognitiveServices.ExtractKeyPhrases([reviews\_text], "en"))

	A <sup>B</sup> categories	A <sup>B</sup> city	A <sup>B</sup> country	1.2 latitude	1.2 longitude	A <sup>B</sup> name	A <sup>B</sup> province	reviews_date	reviews_dateAdded	A <sup>B</sup> reviews_text
1	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	5/5/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	We had A/C issues at 3:30 ..
2	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	6/1/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	A/C was broken. Hotel was ..
3	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	10/7/2015, 5:00:00 PM	2/25/2017, 12:32:57 PM	We had a one night stay at ..
4	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	6/23/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	Elevator was broken.
5	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	9/16/2015, 5:00:00 PM	2/25/2017, 12:32:57 PM	Unprepared for the unwea..
6	Hotels	Kapaa	US	22.043	-159.338	Hotel 5	HI	5/31/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	I expected that the Jacuzzi ..
7	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	9/26/2015, 5:00:00 PM	2/25/2017, 12:32:57 PM	For the price that I paid for ..
8	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	6/12/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	At Night A/C very loud, als..
9	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	6/13/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	The A/C in my room broke.
10	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	7/29/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	Great beach park off the la..
11	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	6/4/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	Our room was on the bott..
12	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	2/15/2016, 4:00:00 PM	2/25/2017, 12:32:57 PM	We spent 2 weeks in this h..
13	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	7/1/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	Terrible view from my \$300.
14	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	9/2/2015, 5:00:00 PM	2/25/2017, 12:32:57 PM	Older property but it is sup..
15	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	8/18/2015, 5:00:00 PM	2/25/2017, 12:32:57 PM	We stayed here for over a ..
16	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	12/10/2015, 4:00:00 PM	2/25/2017, 12:32:57 PM	When we had booked this ..
17	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	7/29/2016, 5:00:00 PM	2/25/2017, 12:32:57 PM	Loved the beach and servic..
18	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	2/15/2016, 4:00:00 PM	2/25/2017, 12:32:57 PM	I hesitate to share negative.
19	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	7/29/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	Beautiful renovation. The h..
20	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	3/26/2016, 5:00:00 PM	2/25/2017, 12:32:57 PM	Positives: Location! It is on ..
21	Hotels	Kapaa	US	22.043	-159.338	Hotel 5	HI	7/30/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	This hotel is on the beach ..
22	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	10/4/2015, 5:00:00 PM	2/25/2017, 12:32:57 PM	Clean room, old style, 196..
23	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	12/12/2015, 4:00:00 PM	2/25/2017, 12:32:57 PM	The accommodation is bas..
24	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	3/1/2016, 4:00:00 PM	2/25/2017, 12:32:57 PM	The entrance to the hotel i..
25	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	9/16/2015, 5:00:00 PM	2/25/2017, 12:32:57 PM	Rooms were nice, basic bu..
26	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	10/25/2015, 5:00:00 PM	2/25/2017, 12:32:57 PM	I booked this hotel for mid..
27	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	5/5/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	Loved the view from my ro..
28	Hotels	Princeville	US	22.226	-159.481	Hotel 2	HI	4/19/2016, 5:00:00 PM	3/31/2017, 11:32:19 AM	A/C unit was disgusting an..
29	Vacation Rentals,Resorts &..	Honolulu	US	21.282	-157.831	Hotel 4	HI	2/28/2016, 4:00:00 PM	2/25/2017, 12:32:57 PM	The staff were really friend..
30										

Name: Hotel Reviews  
Entity type: Custom  
Applied steps: Source, Navigation, Navigation 1, Invoked CognitiveSer..., Invoked CognitiveSer...

1 warning Done



Microsoft Power BI 2 - Demo Gen2

Online Shoppers Intent

Search

Tables Machine learning models

TABLE NAME	TABLE TYPE	ACTIONS
Online Visitors	Custom	
Purchase Intent Prediction v1 Training Data	Custom	
Purchase Intent Prediction v1 Testing Data	Custom	
Online Visitors enriched Purchase Intent Prediction v1	Custom	
Online Visitors enriched Purchase Intent Prediction v1 explanations	Custom	



Python - Power BI Desktop

File Home Insert Modeling View Help External Tools

Paste Cut Get data v Excel Power BI SQL Enter data Dataverse Recent sources v Transform Refresh data v New visual Text box More visuals v New measure Quick measure Sensitivity (preview) v Sensitivity Publish Clipboard Data Queries Insert Calculations Share

Filters Visualizations Fields

Search

Build visuals with your data

Select or drag fields from the Fields pane onto the report canvas.

Add data fields here

Add data fields here

Values

Add data fields here

Drill through

Cross-report

Off —

Keep all filters

On —

Add drill-through fields here

Page 2 +

Page 1 of 1



Power BI Supply Chain - Power BI https://msit.powerbi.com/groups/0f8d48b2-6b16-4d50-87c0-0f2304577cc3/dataflows/c1562d0b-6af1-486c-bfb5-9a23e2f69a29

A: Questions That... Power BI - Edit Blog... GitHub - Microsoft... #PowerQuery ALP Project

Microsoft Power BI Supply Chain Final View Sales Orders Search 32 ?

### Backorder risk model training report

This report summarizes the model performance and training details and enables you find an optimal threshold for defining your business outcome.

Apply model Edit model

Home Favorites Recent Create Apps Shared with me Deployment pipelines Learn Workspaces Supply Chain Fin... ▾

How the model was evaluated

The model predicted On backorder probabilities for a test set of 9044 records and compared the predicted outcomes (based on the selected threshold) to the historical outcomes.

Model performance

The Area under the curve (AUC) observed on the test set is : **85%**

Different features have varying influence on the predicted outcome. Click below for details.

See top predictors

Predicted high risk Predicted low risk

	Predicted high risk	Predicted low risk
Actual high risk	1.77K	2.26K
Actual low risk	316.00	3.48K

84% Precision of records predicted as high risk are likely to actually be high risk

46% Recall of records that are actually high risk are likely to be predicted as high risk

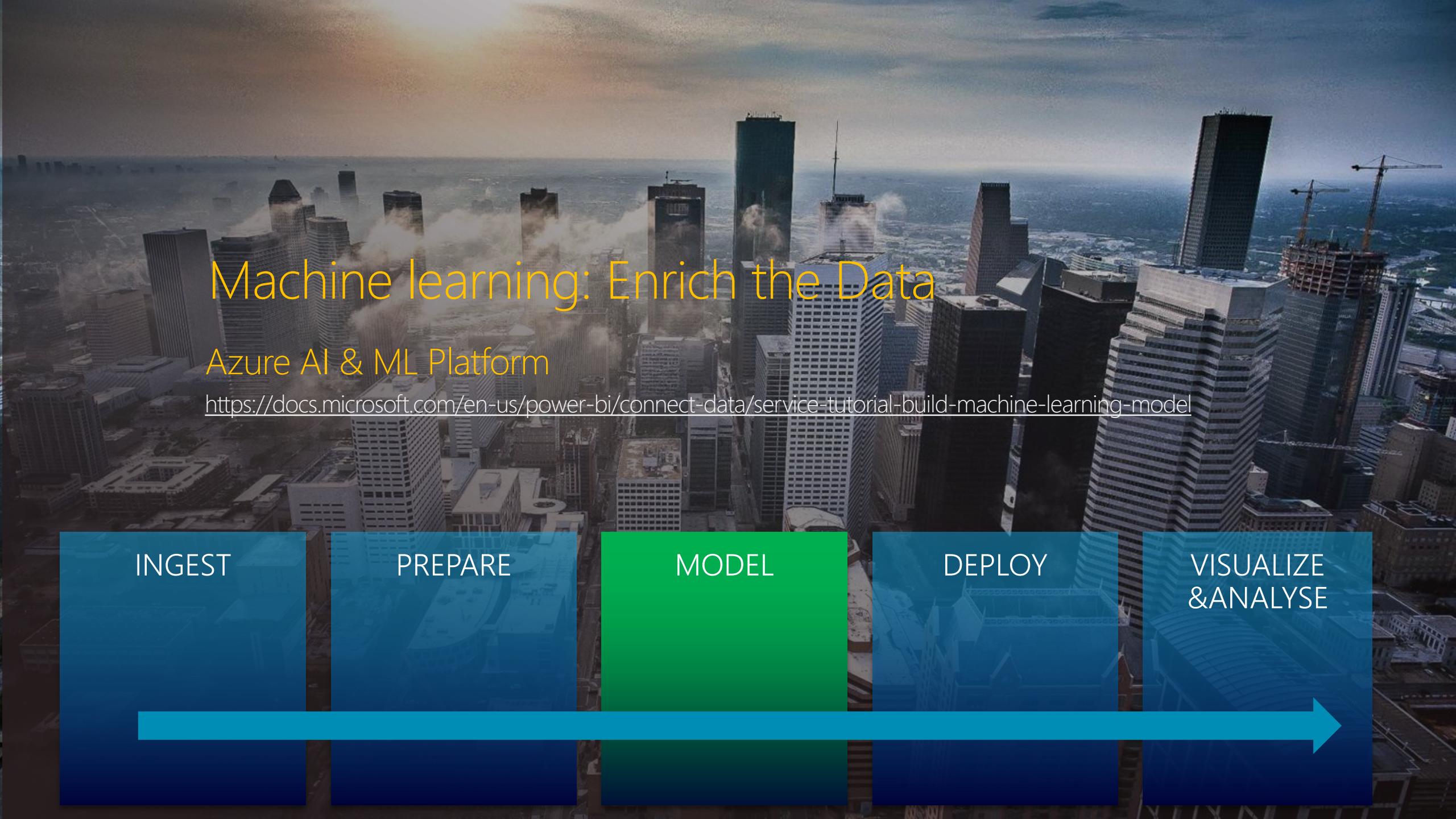
Probability Threshold 0.00 0.78 Increase Recall Increase Precision

### Cost-Benefit Analysis

Get data Model Performance Accuracy Report Training Details

Type here to search

4K 7:56 PM ENG 11/17/2020



# Machine learning: Enrich the Data

Azure AI & ML Platform

<https://docs.microsoft.com/en-us/power-bi/connect-data/service-tutorial-build-machine-learning-model>

INGEST

PREPARE

MODEL

DEPLOY

VISUALIZE  
& ANALYSE

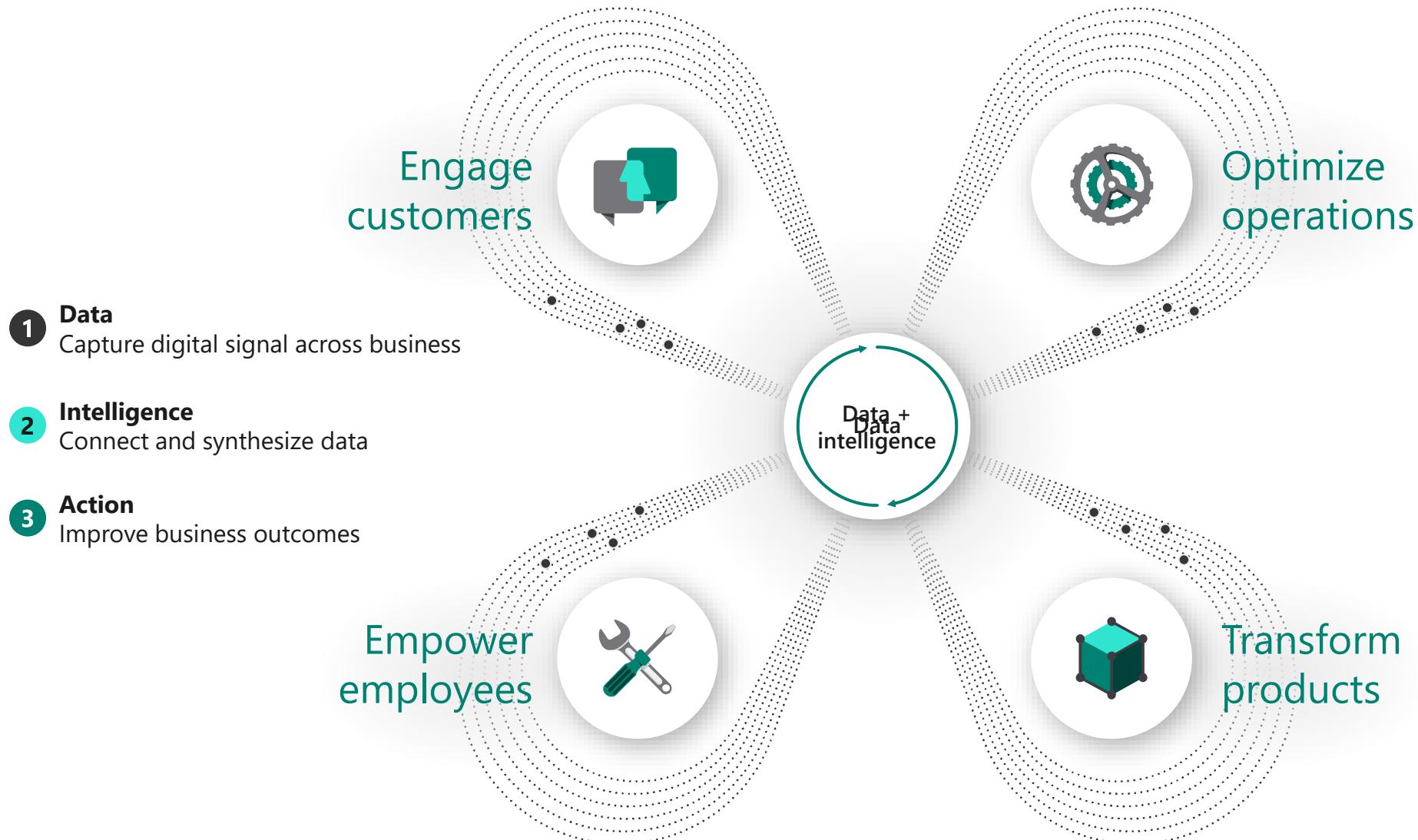
# What is Artificial Intelligence ?

*Software that imitates human capabilities*

- Make decision based on data and past experience
- Recognizing abnormal events
- Interpreting visual input
- Understanding written and spoken language
- Engaging in dialog and conversations



# DIGITAL FEEDBACK LOOP



# Pervasive Artificial Intelligence for BI

**Help end users  
understand their data**



Easily navigate data and  
gain insights

**Assist analysts to  
prepare their data**



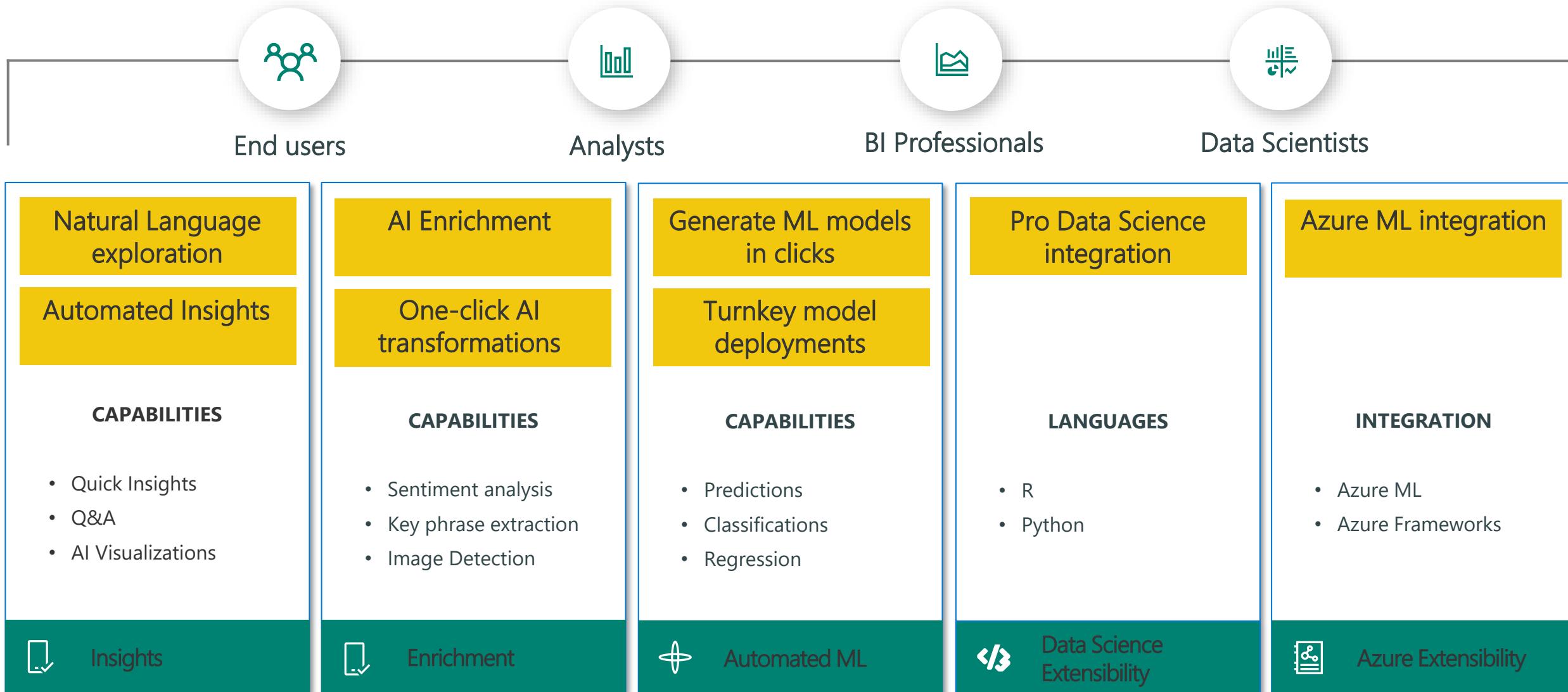
Enrich reports and dashboards  
with low- or no-code AI  
capabilities

**Enable collaboration  
between data scientists  
and analysts**



Seamless and contextual access  
to custom built ML models from  
across the organization

# Current AI investments in Power BI



# AI in Power bi



End users



Analysts



BI Professionals



Data Scientists

Natural Language exploration

Explore influencers

Capabilities:

**Quick Insights**

**Q&A**



Automated  
Insights

Out of box AI in PQ

One-click AI transformations

Capabilities:

**Sentiment analysis**

**Key phrase extraction**

**OCR, Text Translations**



Enrichment

Transforms & Visualizations

Custom Python models

Languages:

**R**

**Python**



Scripting

Generate ML models in clicks

Turnkey model deployments

Capabilities:

**Predictions,  
Classifications,**

**Forecasting, Clustering,**

**Recommendations**



Automated ML

PQ integration for Azure ML

Other Azure hosted models

Integration:

**Azure ML**

**Azure Frameworks**

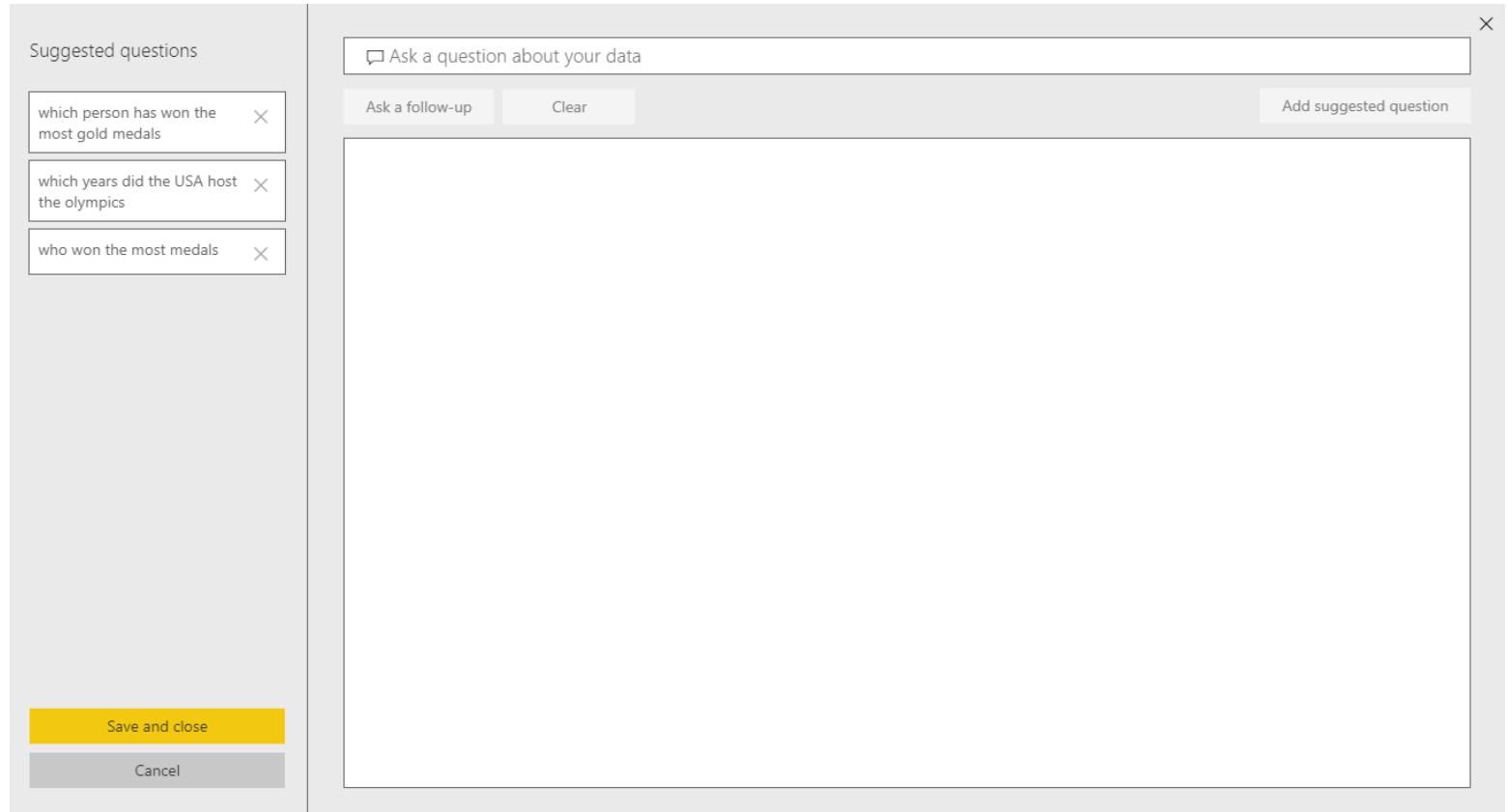


Azure ML

Easily  
navigate  
data and  
gain insights

[Q&A for Power BI business users -  
Power BI | Microsoft Docs](#)

Q&A: lets users navigate their data by asking natural language questions

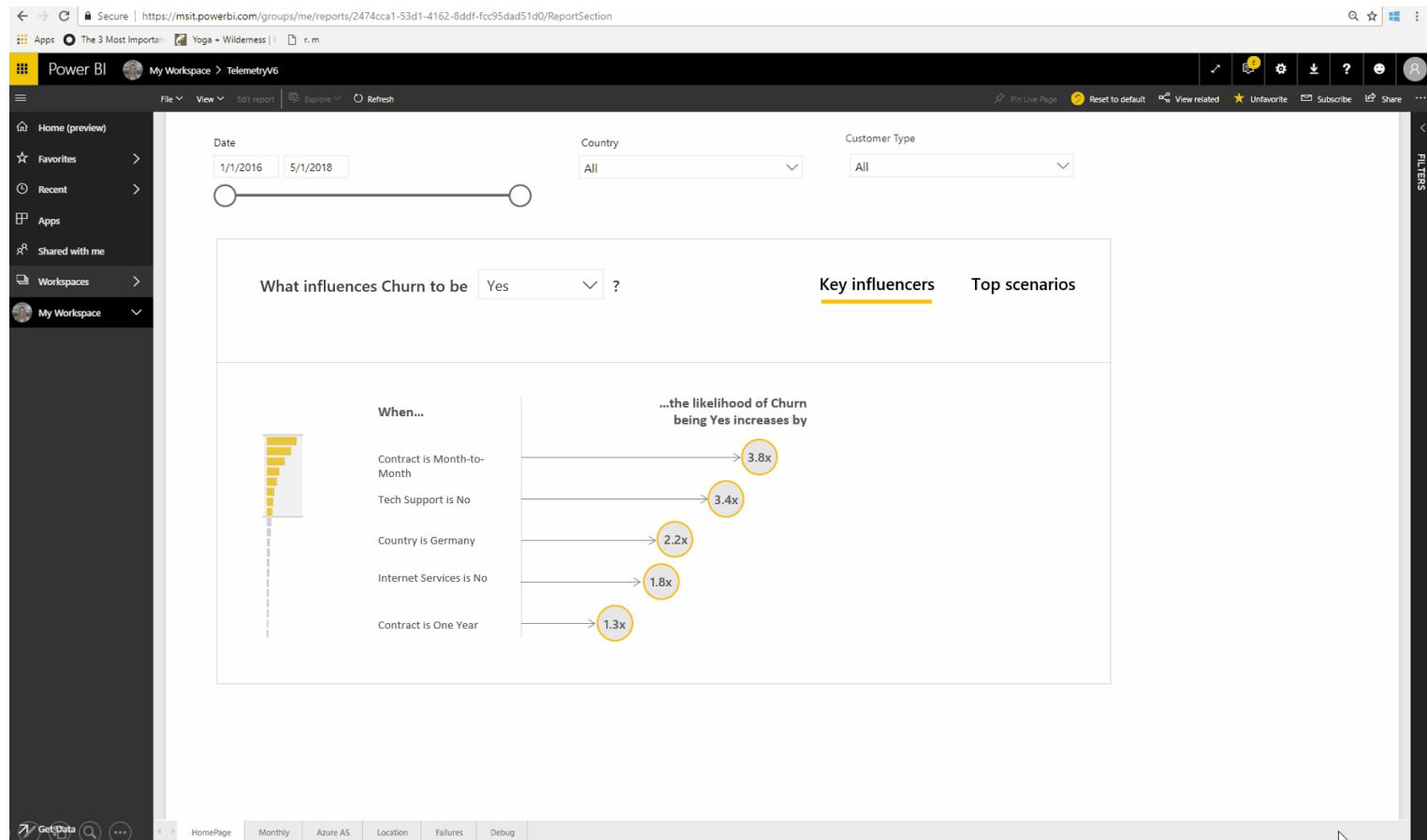


Use Q&A to explore your data using natural language capabilities and receive answers in the form of charts and graphs.

Easily  
navigate  
data and  
gain insights

[Insights - Power BI | Microsoft Docs](#)

# Advanced Insights: surface hidden patterns



This *Insights* feature helps you easily explore and find insights such as anomalies and trends in your data as you interact and consume your reports.

# Enrich reports and dashboards with low- or no- code AI capabilities

[Tutorial: Use Cognitive Services in Power BI \(Preview\) - Power BI | Microsoft Docs](#)

## Cognitive Services: Enrich data with prebuilt ML models

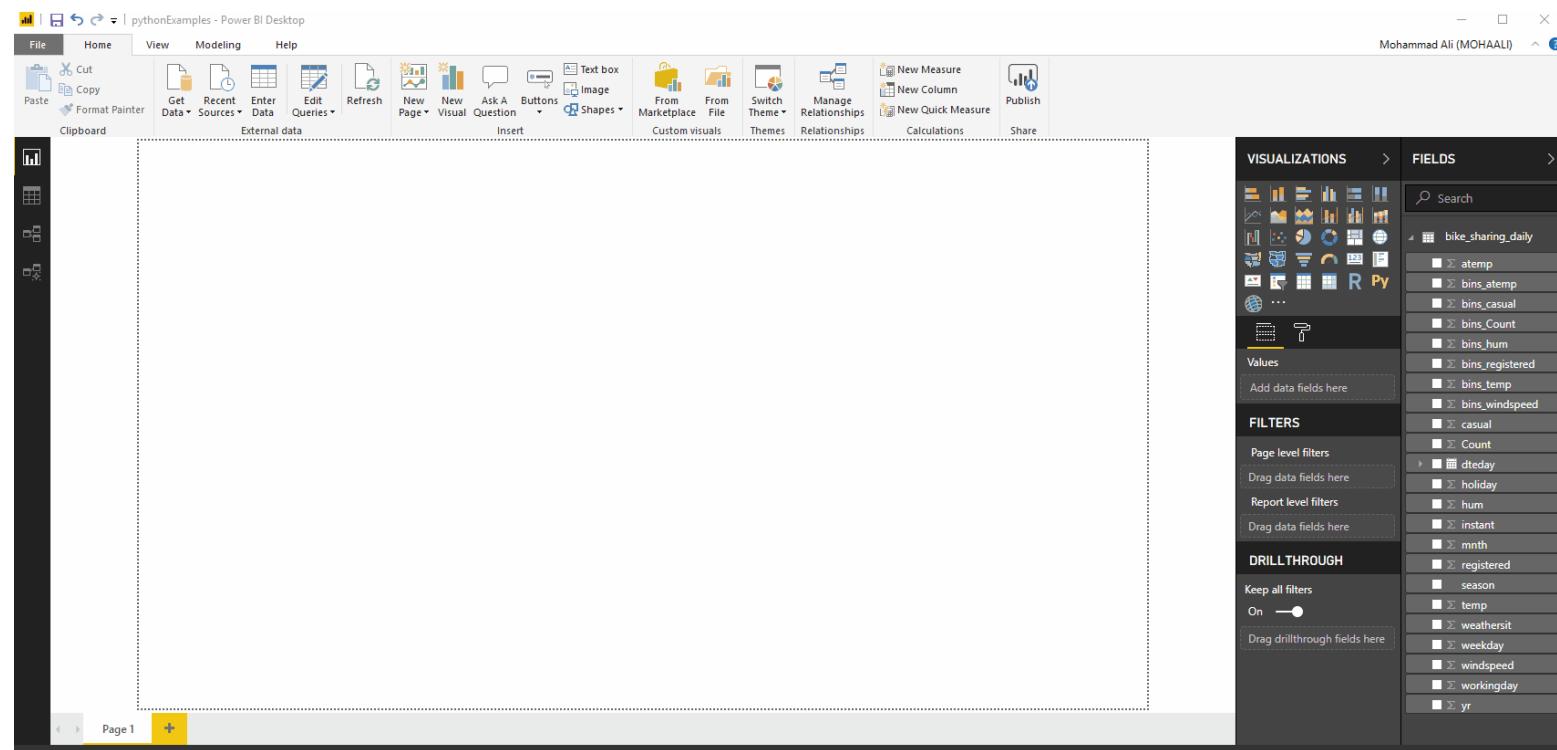
The screenshot shows the Power BI 'Edit queries' interface. The left sidebar lists various items under 'AI Insights Test': Home (preview), Favorites, Recent, Apps, Shared with me, Workspaces, and AI Insights Test. Under 'DATAFLOWS', 'HotelReviews' is selected. The main area displays a table titled 'Edit queries' with the following columns: reviews.date, reviews.dateAdded, reviews.text, reviews.title, and Image. The table contains 24 rows of hotel review data. The Power Query ribbon at the top includes tabs for Get data, Refresh, Options, Manage columns, Transform table, Reduce rows, Add column, AI insights, Map to standard, and Combine tables. On the right, a pane shows the 'Name' as HotelReviews, 'Entity type' as Custom, and 'Applied steps' which include Source (https://ignitedemoai.blob...), Imported CSV, Promoted Headers, Changed Type, and Invoked CognitiveServ... (highlighted in yellow). A status bar at the bottom shows 1 warning.

Seamless and contextual access to custom built ML models from across the organization

[Run Python Scripts in Power BI Desktop - Power BI | Microsoft Docs](#)

[Run R scripts in Power BI Desktop - Power BI | Microsoft Docs](#)

# R and Python Integration



Seamless and contextual access to custom built ML models from across the organization

[Tutorial: Create the predictive model by using automated ML \(part 1 of 2\) - Azure Machine Learning | Microsoft Docs](#)

# Azure ML Integration

The screenshot shows the Power BI 'Edit queries' interface. On the left, the navigation pane lists 'Home (preview)', 'Favorites', 'Recent', 'Apps', 'Shared with me', 'Workspaces', and 'AI Insights Test'. Under 'AI Insights Test', 'HotelReviews' is selected. The main area displays a table titled 'Edit queries' with columns: reviews.text, reviews.title, Image, ExtractKeyPhrases, and AzureML.HotelImageClassifier. The table contains 24 rows of hotel review data. The 'AzureML.HotelImageClassifier' column shows the results of the ML model. The bottom right corner of the interface shows a warning icon with '1 warning' and a 'Done' button.

reviews.text	reviews.title	Image	ExtractKeyPhrases	AzureML.HotelImageClassifier
1 We had A/C issues at 3:30 ...	Please Keep Away Until Co...	https://iytimg.com/vi/-3sD... (null)	[Record]	Not recommended
2 AC was broken. Hotel was...	Please Keep Away Until Co...	https://iytimg.com/vi/xcEB... (null)	[Record]	Not recommended
3 We had a one night stay at...	worst ever !	https://iytimg.com/vi/xcEB... (null)	[Record]	Not recommended
4 Elevator was broken.	Gutes Preis- Leistungsverhi...	https://media-cdn.tripadv...	[Record]	Recommended
5 Unprepared for the unwe... 6 expected that the Jacuzzi ...	Very much a budget place Did not live up to the Hilt...	https://media-cdn.tripadv...	[Record]	Not recommended
7 For the price that I paid for...	Rooms	https://media-cdn.tripadv...	[Record]	Not recommended
8 At Night A/C very loud, als...	AC in room Too loud!	https://media-cdn.tripadv...	[Record]	Not recommended
9 The A/C in my room broke...	Don't waste your money	https://media-cdn.tripadv...	[Record]	Not recommended
10 Great beach park off the la...	Nice surprise.	https://s3-media1.fl.yelpcd...	[Record]	Recommended
11 Our room was on the bott...	Great staff, excellent getaw...	https://s3-media1.fl.yelpcd...	[Record]	Recommended
12 We spent 2 weeks in this h...	IN SEVERE NEED OF UPDA...	https://s3-media1.fl.yelpcd...	[Record]	Not recommended
13 terrible view from my \$300...	Beautiful renovations locat...	https://s3-media1.fl.yelpcd...	[Record]	Recommended
14 Older property but it is su...	Combines great price with ...	https://s3-media1.fl.yelpcd...	[Record]	Recommended
15 We stayed here for over a ...	Affordable, Clean, Friendly ...	https://s3-media1.fl.yelpcd...	[Record]	Recommended
16 When we had booked this ...	Average	https://s3-media1.fl.yelpcd...	[Record]	Not recommended
17 Loved the beach and service	Well the location is nice	https://s3-media1.fl.yelpcd...	[Record]	Recommended
18 hesitate to share negative...	Advertisement Sham	https://s3-media1.fl.yelpcd...	[Record]	Not recommended
19 Beautiful renovation. The h...	Beautiful renovations locat...	https://s3-media2.fl.yelpcd...	[Record]	Recommended
20 Positives: Location! It is on ...	Location!	https://s3-media2.fl.yelpcd...	[Record]	Recommended
21 This hotel is on the beach ...	Great Location	https://s3-media2.fl.yelpcd...	[Record]	Recommended
22 Clean room, old style, 196...	Location	https://s3-media2.fl.yelpcd...	[Record]	Recommended
23 The accommodation is bas...	Polynesian Plaza, Honolulu	https://s3-media2.fl.yelpcd...	[Record]	Not recommended

# What is Azure Machine Learning service?

Set of Azure Cloud  
Services



Python  
SDK

---

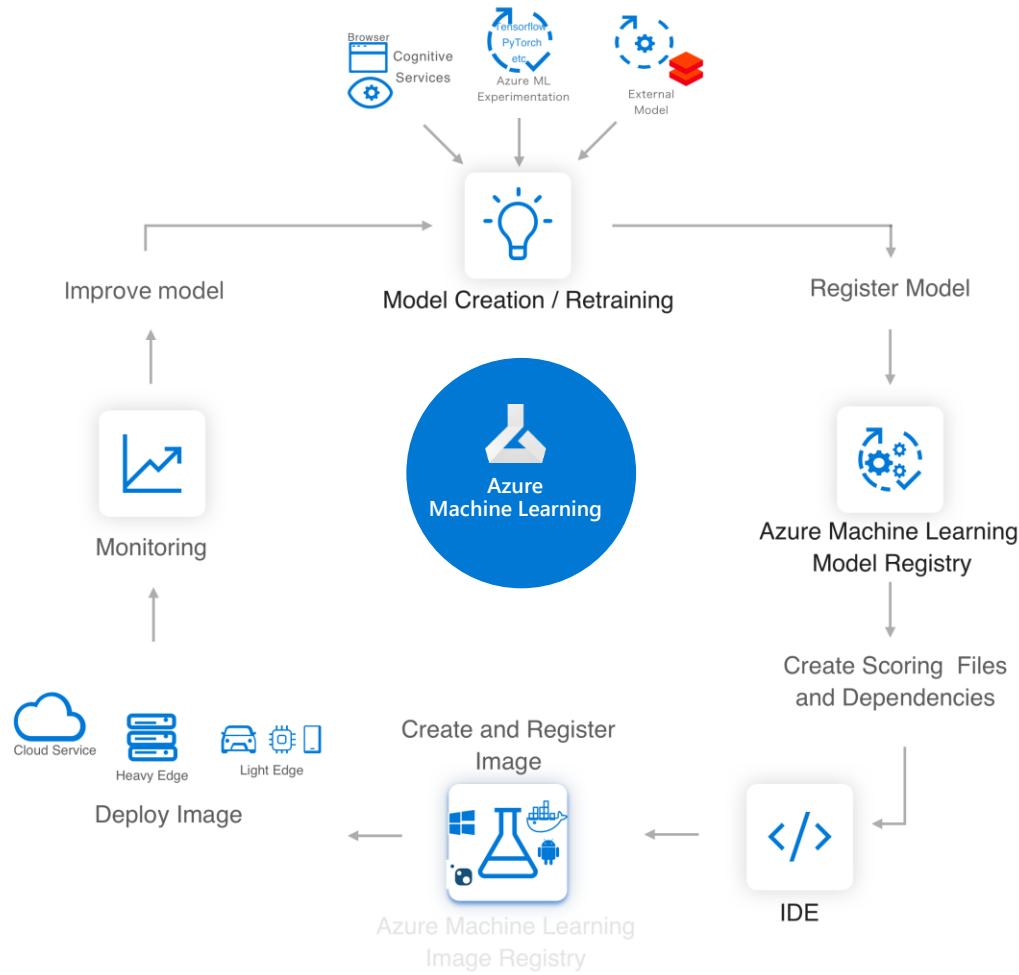
That enables  
you to:

- ✓ Prepare Data
- ✓ Build Models
- ✓ Train Models

- ✓ Manage Models
- ✓ Track Experiments
- ✓ Deploy Models

# Azure ML service

Let's you easily implement this AI/ML Lifecycle

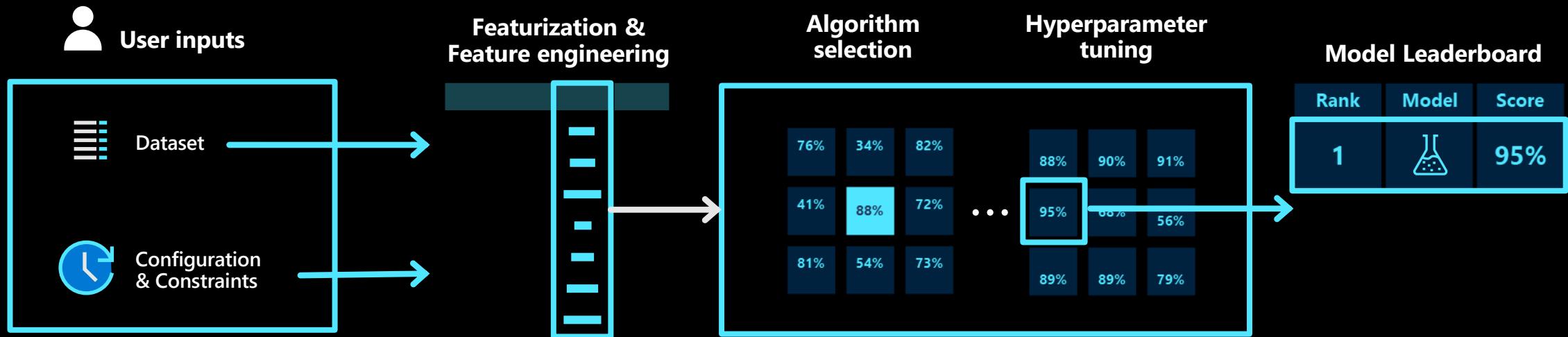


## Workflow Steps

- ✓ Develop machine learning training scripts in Python.
- ✓ Create and configure a compute target.
- ✓ Submit the scripts to the configured compute target to run in that environment. During training, the compute target stores run records to a datastore. There the records are saved to an experiment.
- ✓ Query the experiment for logged metrics from the current and past runs. If the metrics do not indicate a desired outcome, loop back to step 1 and iterate on your scripts.
- ✓ Once a satisfactory run is found, register the persisted model in the model registry.
- ✓ Develop a scoring script.
- ✓ Create an Image and register it in the image registry.
- ✓ Deploy the image as a web service in Azure.

# What is Automated Machine Learning?

Automated machine learning (automated ML) automates feature engineering, algorithm and hyperparameter selection to find the 'best model' for your data.



# Model Creation Is Typically Time-Consuming (1 of 3)

Which features?	Which algorithm?	Which parameters?	
Driver Address	Gradient Boosted	Parameter 1	
Driver Age	Nearest Neighbors	Parameter 2	
Max Mileage Year	SVM	Min Samples Split	
Engine power	Bayesian Regression	Min Samples Leaf	
Driver Phone number	LGBM	Others	
...	...		

30%  
Model

# Model Creation Is Typically Time-Consuming (2 of 3)

Which features?

Max Mileage Year
Driver Age
Engine Power
Car Color
Driver Address
...

Which algorithm?

Gradient Boosted
Nearest Neighbors
SVM
Bayesian Regression
LGBM
...

Which parameters?

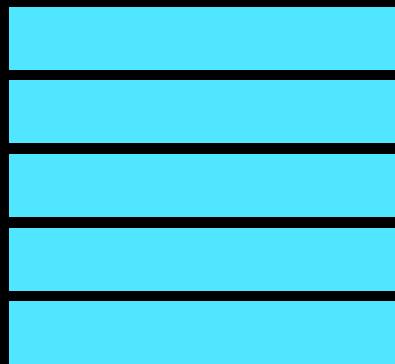
Neighbors
Weights
Min Samples Split
Min Samples Leaf
Others

30% Model

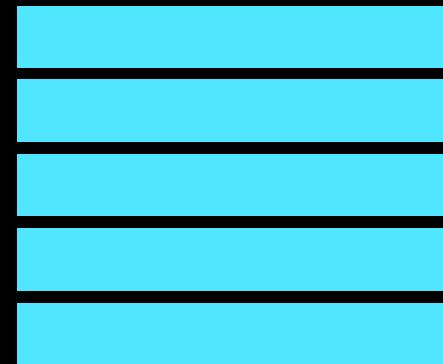
Iterate

# Model Creation Is Typically Time-Consuming (3 of 3)

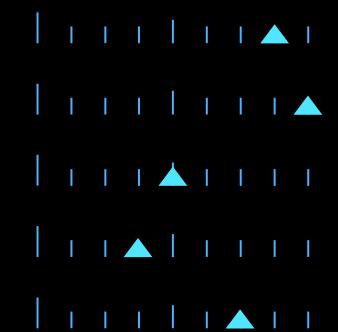
Which features?



Which algorithm?



Which parameters?



30%

15%

Iterate

# Automated ML

## How it works

ML Tasks supported:

Classification Regression and Time-Series Forecasting

+ DNN-based text featurization and algos (BERT).

Based on Microsoft Research

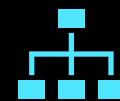
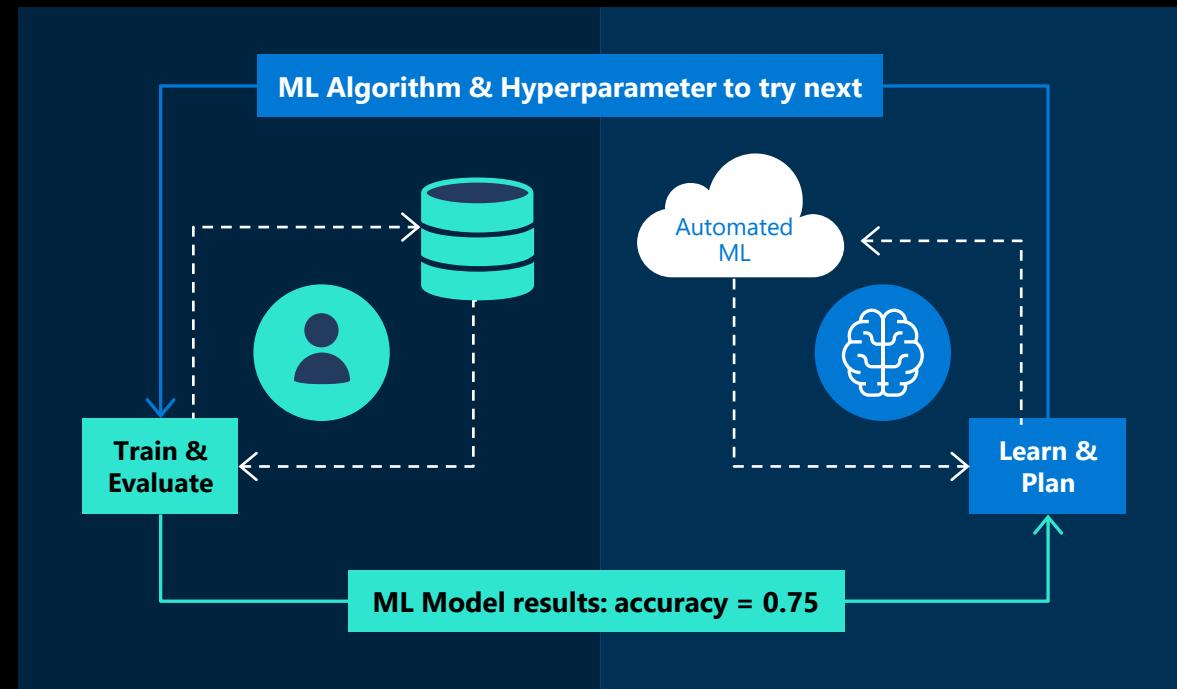
Brain trained with millions of experiments

Personalized recommendation approach:

Collaborative filtering and Bayesian optimization

Privacy preserving:

No need to "see" the data



Classification



Regression



Time Series Forecasting

# Azure Automated ML – Sample Output

AutoML\_ab755820-4bfd-4e8a-8b4b-9e0a2446b1c2:

Status: Completed

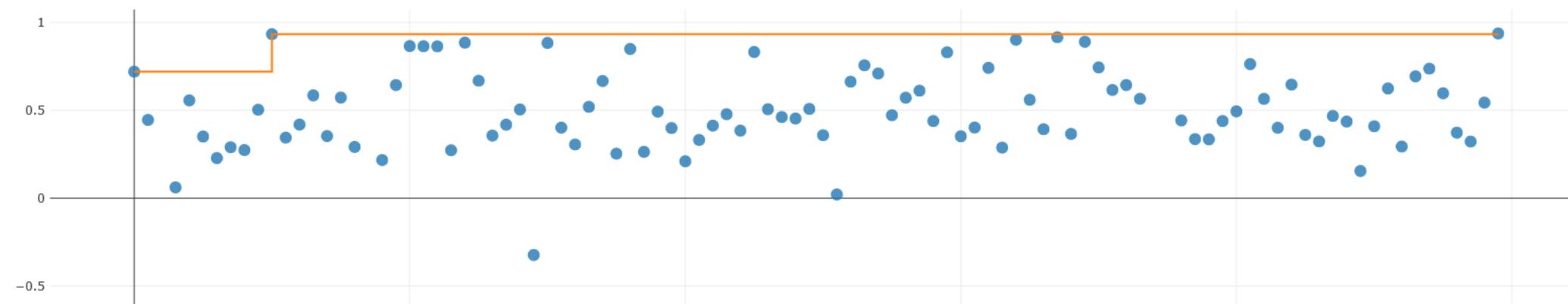


Iteration	Pipeline	Iteration metric	Best metric	Status	Duration	Started	Run Id
99	Ensemble	0.93702349	0.93702349	Completed	0:02:18	Dec 4, 2018 12:18 AM	<a href="#">View</a>
10	MaxAbsScaler, LightGBM	0.93289307	0.93289307	Completed	0:01:22	Dec 3, 2018 7:49 PM	<a href="#">View</a>
67	SparseNormalizer, LightGBM	0.9154763	0.93289307	Completed	0:01:31	Dec 3, 2018 10:19 PM	<a href="#">View</a>
64	MaxAbsScaler, LightGBM	0.90148724	0.93289307	Completed	0:01:24	Dec 3, 2018 10:09 PM	<a href="#">View</a>
69	MaxAbsScaler, LightGBM	0.88975241	0.93289307	Completed	0:00:55	Dec 3, 2018 10:22 PM	<a href="#">View</a>

Pages: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... Next Last [5 ▾] per page

r2\_score ▾

AutoML Run with metric : r2\_score



# Demo

The background of the slide is a black and white aerial photograph of a dense urban skyline, likely Houston, Texas. The city is filled with numerous skyscrapers of varying heights, some with visible dust or steam rising from their tops, suggesting construction or industrial activity. In the foreground, there's a large, dark, semi-transparent rectangular area containing the main text.

# BI and Data Warehousing comes together

INGEST

PREPARE

MODEL

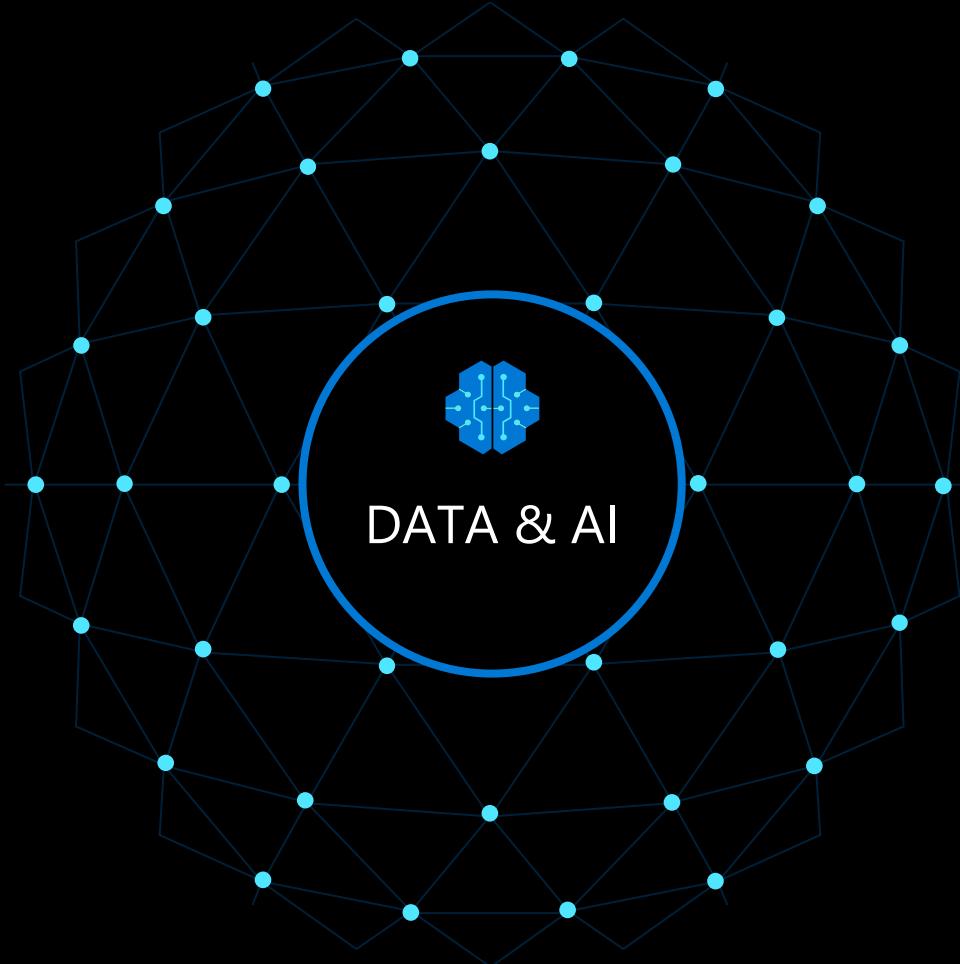
DEPLOY

VISUALIZE  
& ANALYSE

# The Data Driven Enterprise

Engage customers 

Transform products 

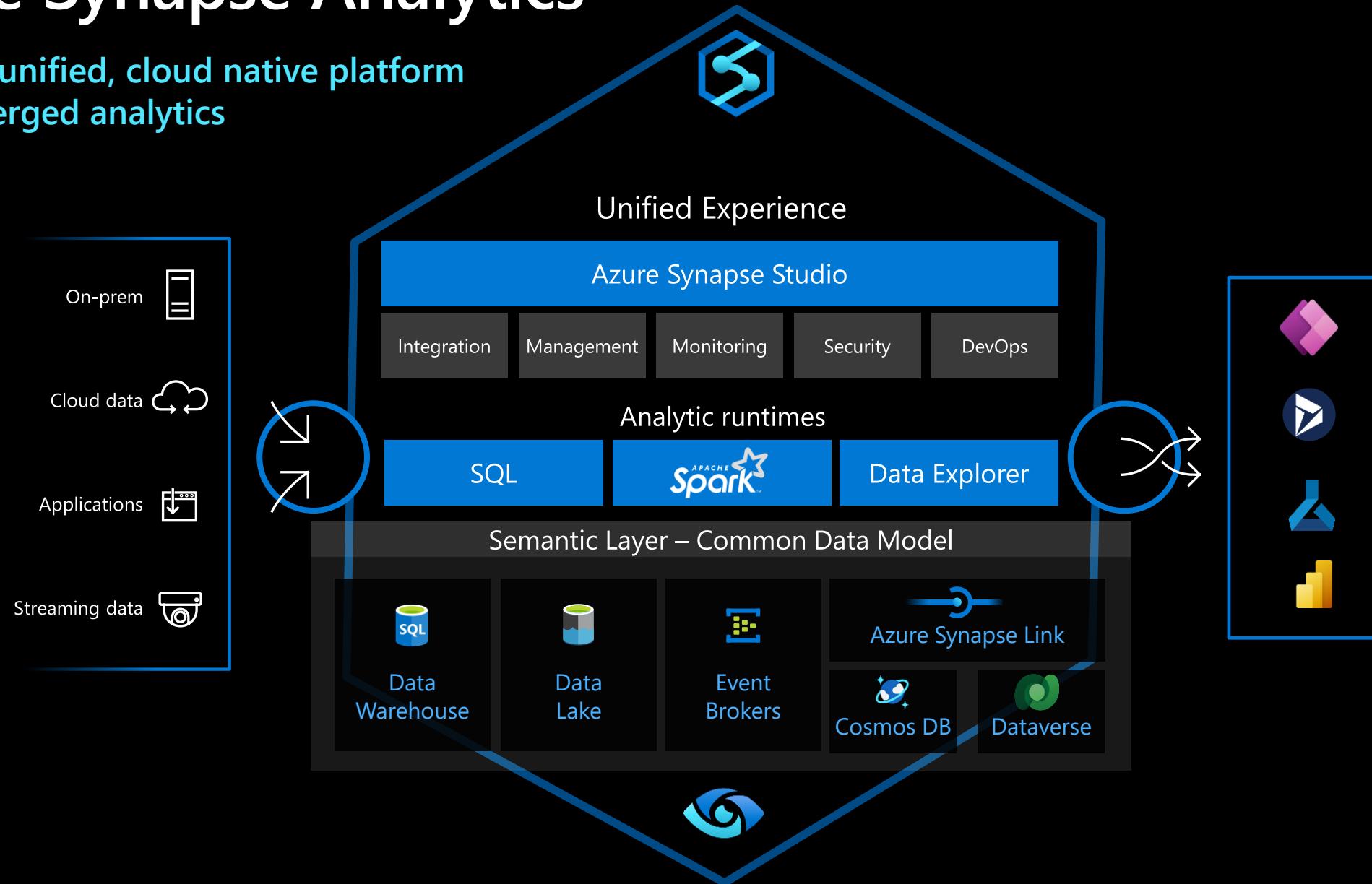


 Optimize operations

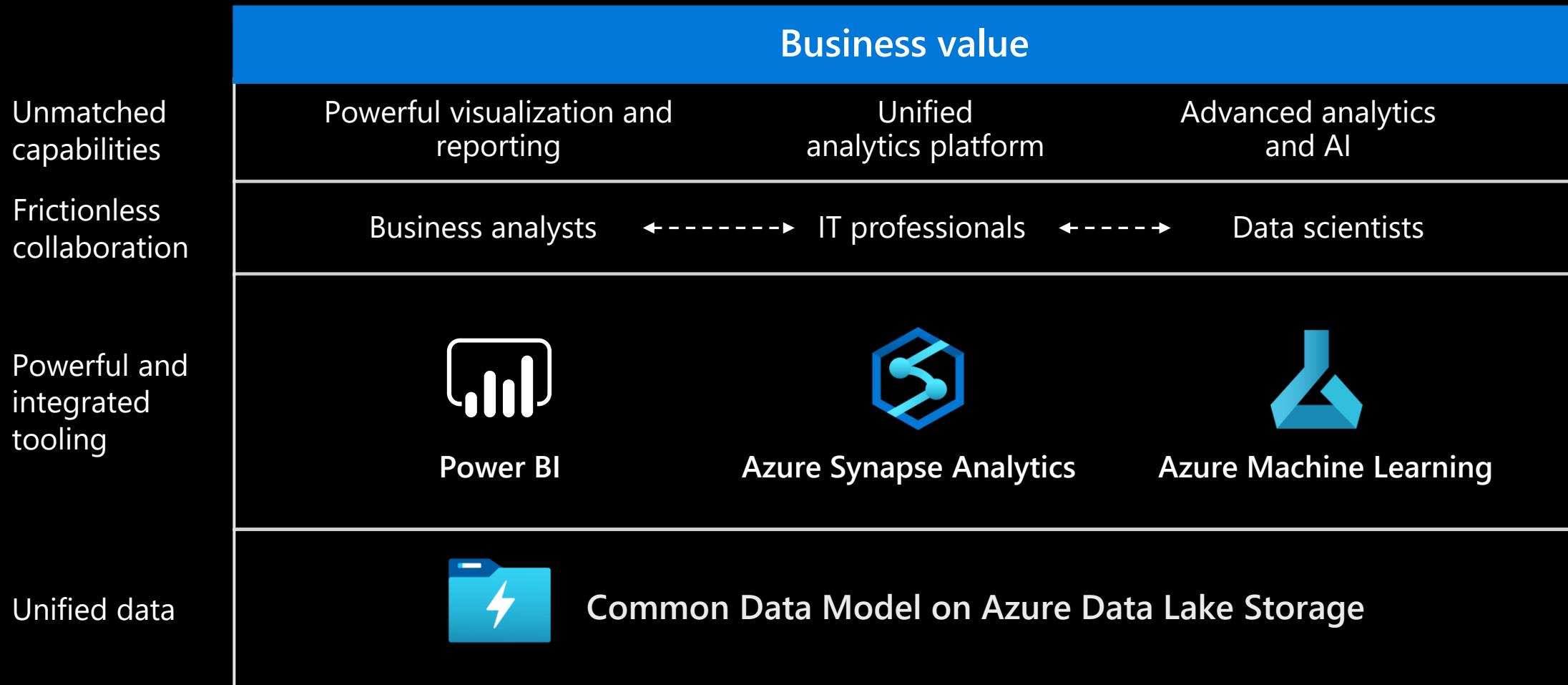
 Empower people

# Azure Synapse Analytics

The first **unified, cloud native platform**  
for converged analytics



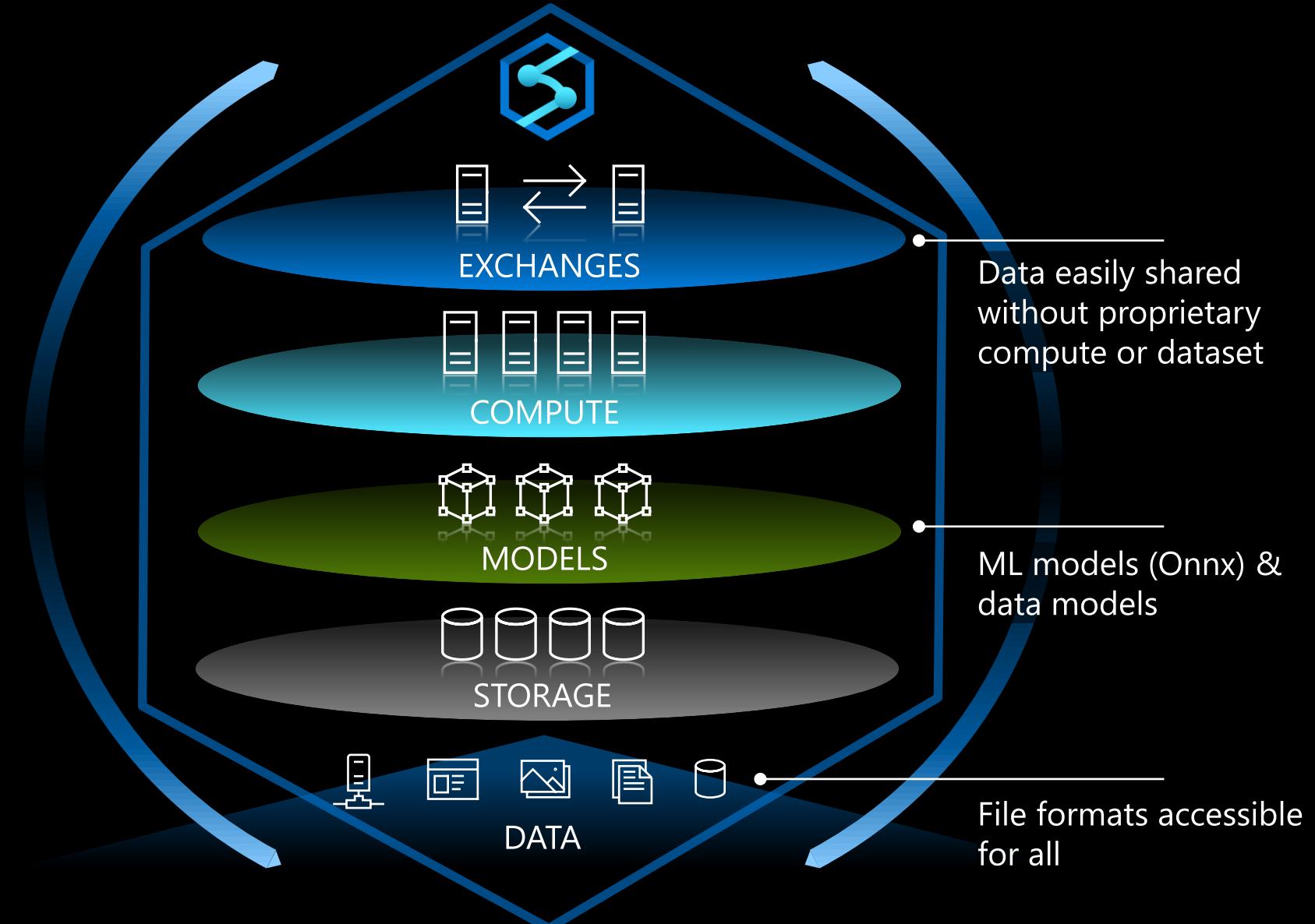
# Accelerate business value with a powerful analytics platform



# Why Synapse Open Hub

## Hub & Spoke Model

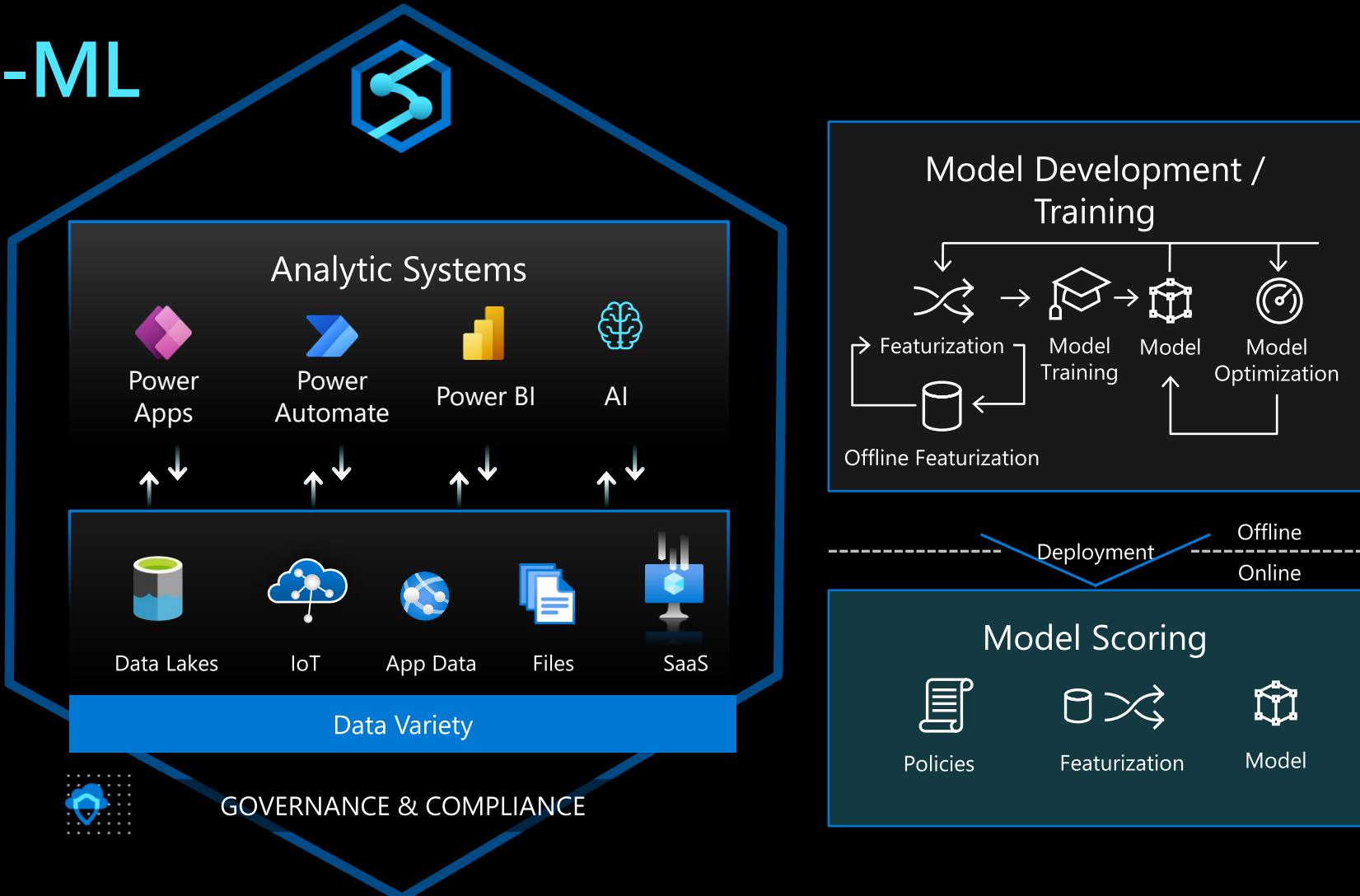
- Integrated with hundreds of ISVs
- Compute choices using both proprietary and open-source tech
- Open lake format
- Open file formats



# Why Synapse Enterprise Grade-ML

## Productizing AI

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



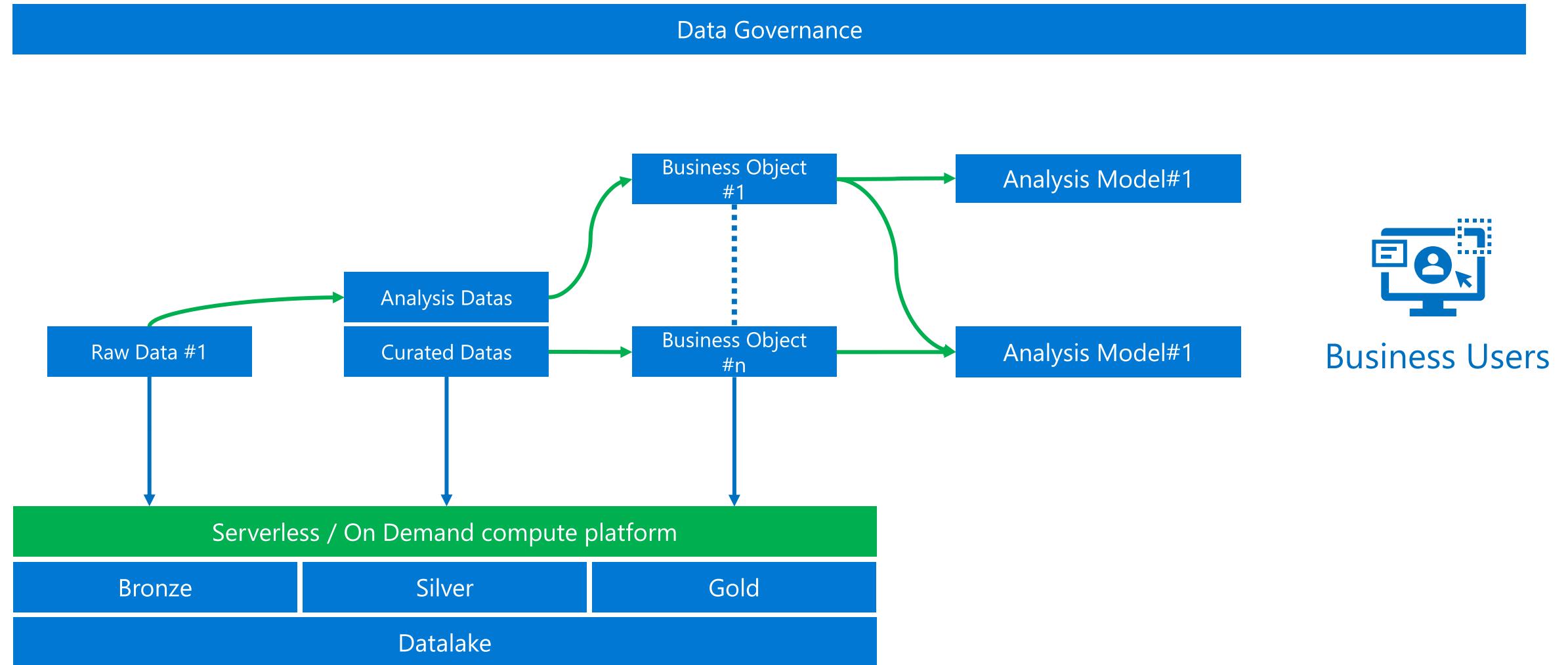
# Approche moderne de données - Lakehouse database

- Nouvelle approche d'objets métiers
- Exploiter les bénéfices des solutions « Cloud Native » comme serverless, autoscale, on-demand....
- Notion forte de croisement entre le Datawarehouse, le Datalake et les modèles d'analyse
- La Datalake est vu comme une variable d'entrée et non une source de donnée
- La finalité de l'approche
  - Exposer des objets métiers dynamiques
  - Exposer des objets métiers logiques
  - La matérialisation des objets n'est plus nécessaire
  - S'affranchir des notions de scalabilité et performance
  - S'affranchir de la consommation ou non de l'objet exposé

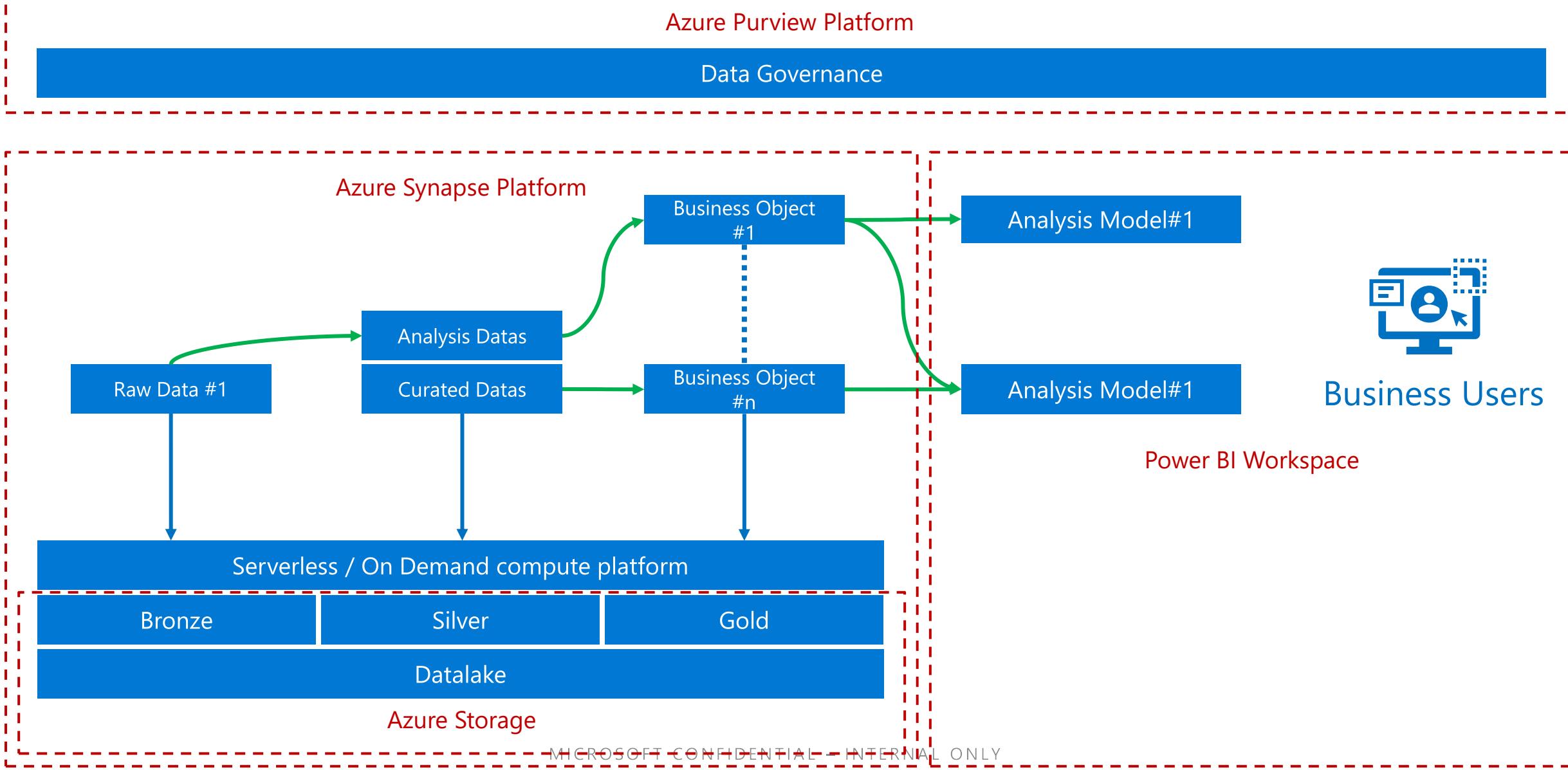
# Explications et retour d'expérience

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

# Objets logiques



# Objets logiques



# Hypothèses et approche

- Besoin d'extraction des sources de données
- Structuration du/des Datalake en zones
- Moteur de la plateforme Synapse
  - <https://aka.ms/synapse-dqp>
  - Moteur Synapse dédié pour interne les besoins internes
  - Moteur Synapse Serverless pour les accès depuis le portail Web avec BI embarquée
- Consommation Web via PBI Embedded ou JS
- Etude en amont
  - Nombre de systèmes sources
  - Structuration du datalake centralisé et gouvernance de la zone Gold
  - Fréquence de mise à jour des données et sur quel mode (full/delta)

# Extraction des informations

- Extraction depuis les systèmes sources à l'aide de Azure Synapse Pipeline
- Structuration des données dans des zones de datalake
- Implémentation des "Integration Runtime" pour des captures sur les systèmes "on premise"

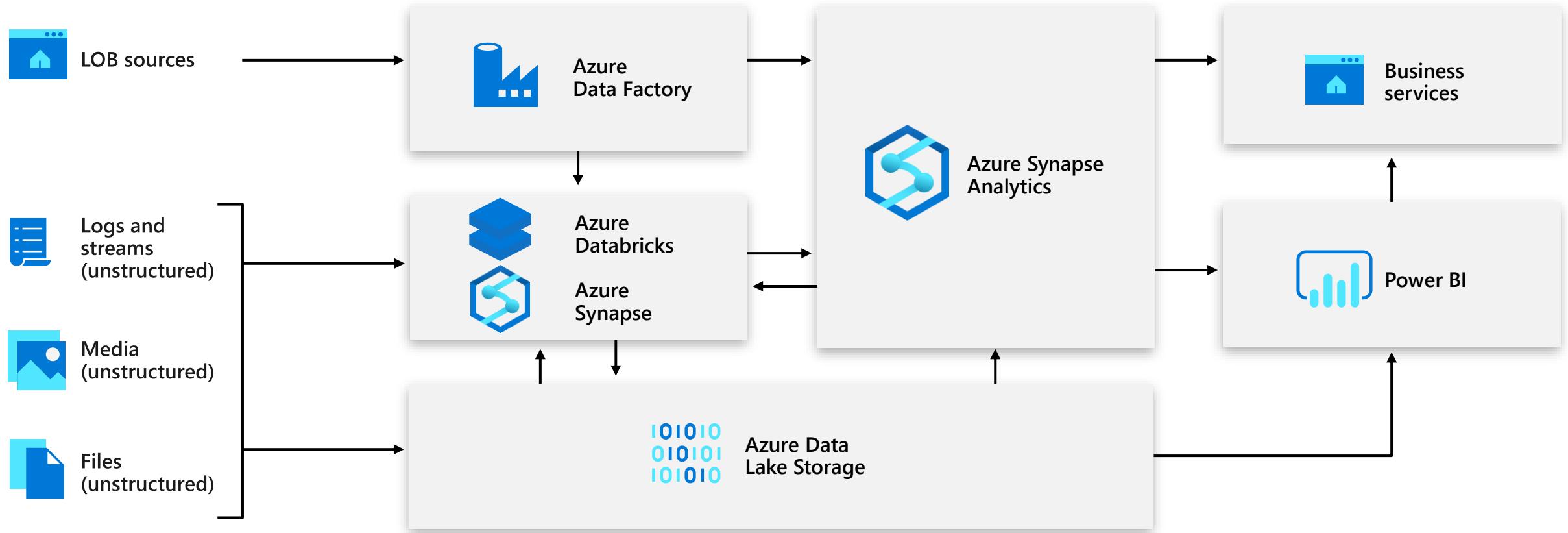
# Structuration du datalake

- Les données sont mises à disposition dans un concept de "(Data) Lakehouse"
- Une définition par Databricks :  
<https://databricks.com/fr/blog/2020/01/30/what-is-a-data-lakehouse.html>
- Réflexion en amont sur les cas d'usages des différentes données afin de déterminer
  - Leur volumétrie
  - Leur fréquence d'utilisation
  - Leur granularité

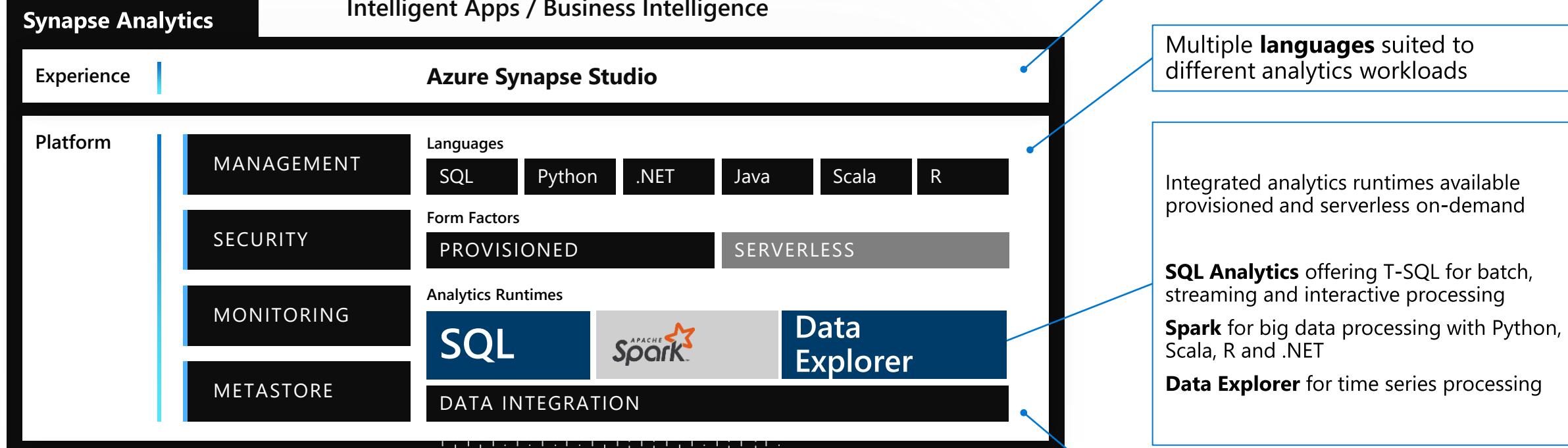
# Les zones de l'approche "Datalake(house)"

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

# L'approche moderne d'exposition de données



# Moteurs de la plateforme Azure Synapse



Azure  
Data Lake Storage

Common Data Model  
Enterprise Security  
Optimized for Analytics

Designed for analytics **workloads at any scale**

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

Multiple **languages** suited to different analytics workloads

Integrated analytics runtimes available provisioned and serverless on-demand

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

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

**Data Explorer** for time series processing

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

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

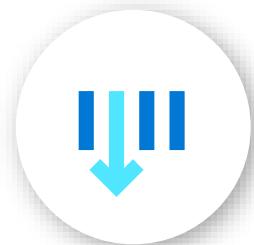
# Performance



Elastic Architecture



Columnar Storage



Columnar Ordering

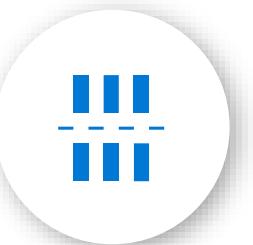


Table Partitioning



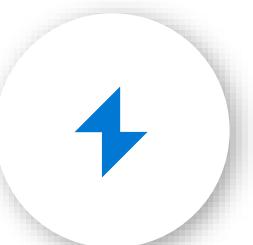
Secondary Indexes



Hash Distribution



Materialized Views



Resultset Cache

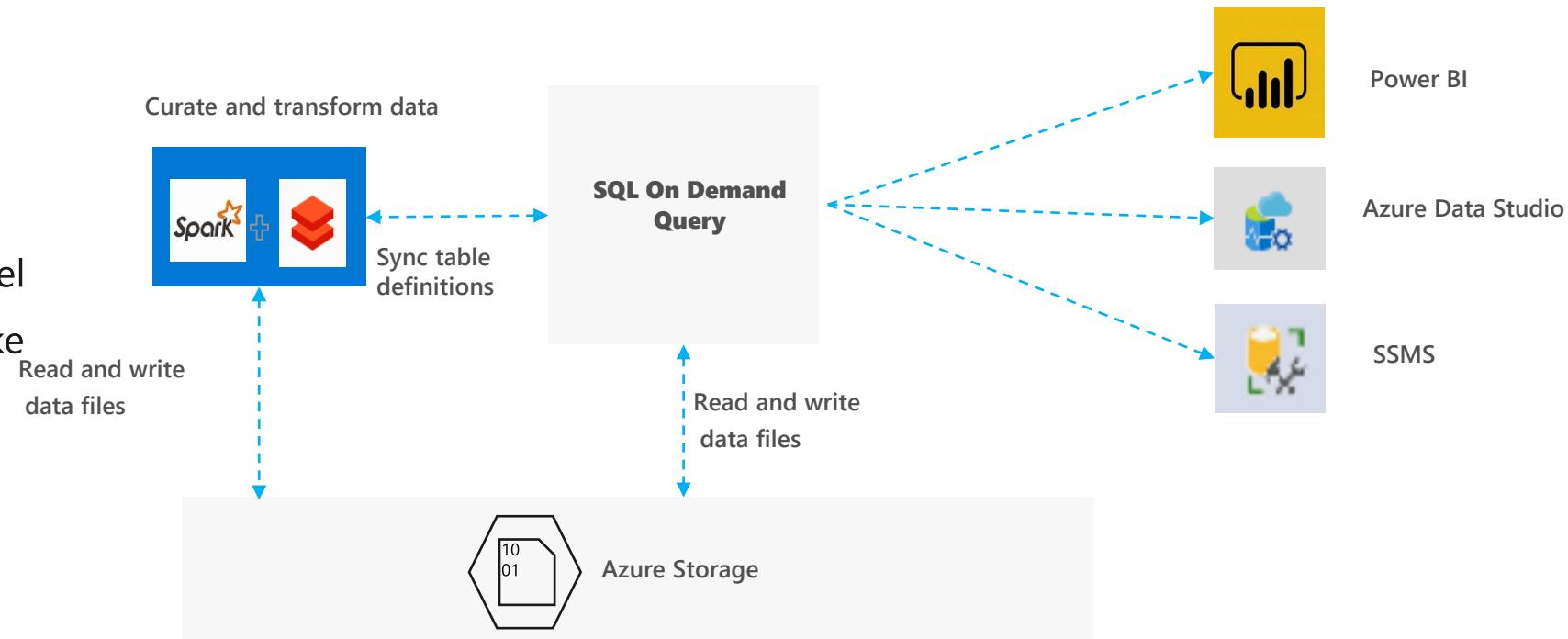
# Transform with Synapse SQL Serverless / Pricing

## Overview

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

## Benefits

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



# Synapse SQL Serverless – Querying on storage

The screenshot displays two side-by-side views of the Microsoft Azure Synapse Analytics portal.

**Left View:** Shows the storage exploration interface under the 'Data' tab. A context menu is open over a file named 'part-00133-19836543-aea5b543-5e83-4a7d-8d31-69f72c50b05d-15253-1.c000.snappy.parquet'. The menu options include 'New SQL script' (which is highlighted), 'New notebook', 'Copy ABFS path', 'Manage Access...', 'Rename...', 'Download', 'Delete', and 'Properties...'. The left sidebar shows 'Storage accounts' (including 'nyctlc' which is selected), 'Databases' (including 'prlangaddemosa'), and 'Datasets'.

**Right View:** Shows the query editor interface. The 'SQL Analytics on-demand' dropdown is highlighted with a red box. The query window contains the following T-SQL code:

```
1 SELECT
2     TOP 100 *
3 FROM
4     OPENROWSET(
5         BULK 'https://prlangaddemosa.dfs.core.windows.net/nyctlc/yellow/puYear=2015/puMonth=3/part-00133-tid-210938564719836543-aea5b543-5e83-4a7d-8d31-69f72c50b05d-15253-1.c000.snappy.parquet'
6         FORMAT='PARQUET'
7     ) AS nyc;
```

The results pane shows a table with the following columns: VENDORID, TPEPICKUPDATETIME, TEPDROPPOFFDATETIME, PASSENGERCOUNT, TRIPDISTANCE, PULOCATIONID, DLOCATIONID, STARTLON, STARTLAT, and ENDLON. The results are as follows:

VENDORID	TPEPICKUPDATETIME	TEPDROPOFFDATETIME	PASSENGERCOUNT	TRIPDISTANCE	PULOCATIONID	DLOCATIONID	STARTLON	STARTLAT	ENDLON
2	2015-02-28T23:5...	2015-03-01T00:0...	6	1.63	NULL	NULL	-74.000846862793	40.7306938171387	-73.
1	2015-03-28T19:2...	2015-03-28T19:2...	1	2.2	NULL	NULL	-73.977653503418	40.7631607055664	-73.
2	2015-02-28T23:5...	2015-03-01T00:1...	5	3.23	NULL	NULL	-73.96012878417...	40.7621574401855	-73.
1	2015-03-28T19:2...	2015-03-28T19:3...	1	2.1	NULL	NULL	-73.98143005371...	40.7815055847168	-74.
2	2015-02-28T23:5...	2015-03-01T00:1...	1	3.52	NULL	NULL	-73.98373413085...	40.7497062683105	-74.

At the bottom of the results pane, a message indicates: **00:01:00 Query executed successfully.**

# Synapse SQL Serverless – Querying CSV File

## Overview

Uses OPENROWSET function to access data

## Benefits

Ability to read CSV File with

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

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

	country_code	country_name	year	population
1	LU	Luxembourg	2017	594130

# Synapse SQL Serverless – Querying specific files

## Overview

**filename** – Provides file name that originates row result

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

## Benefits

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

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

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

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

## Example of filename function

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

GROUP BY r.filename()

ORDER BY [filename]

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

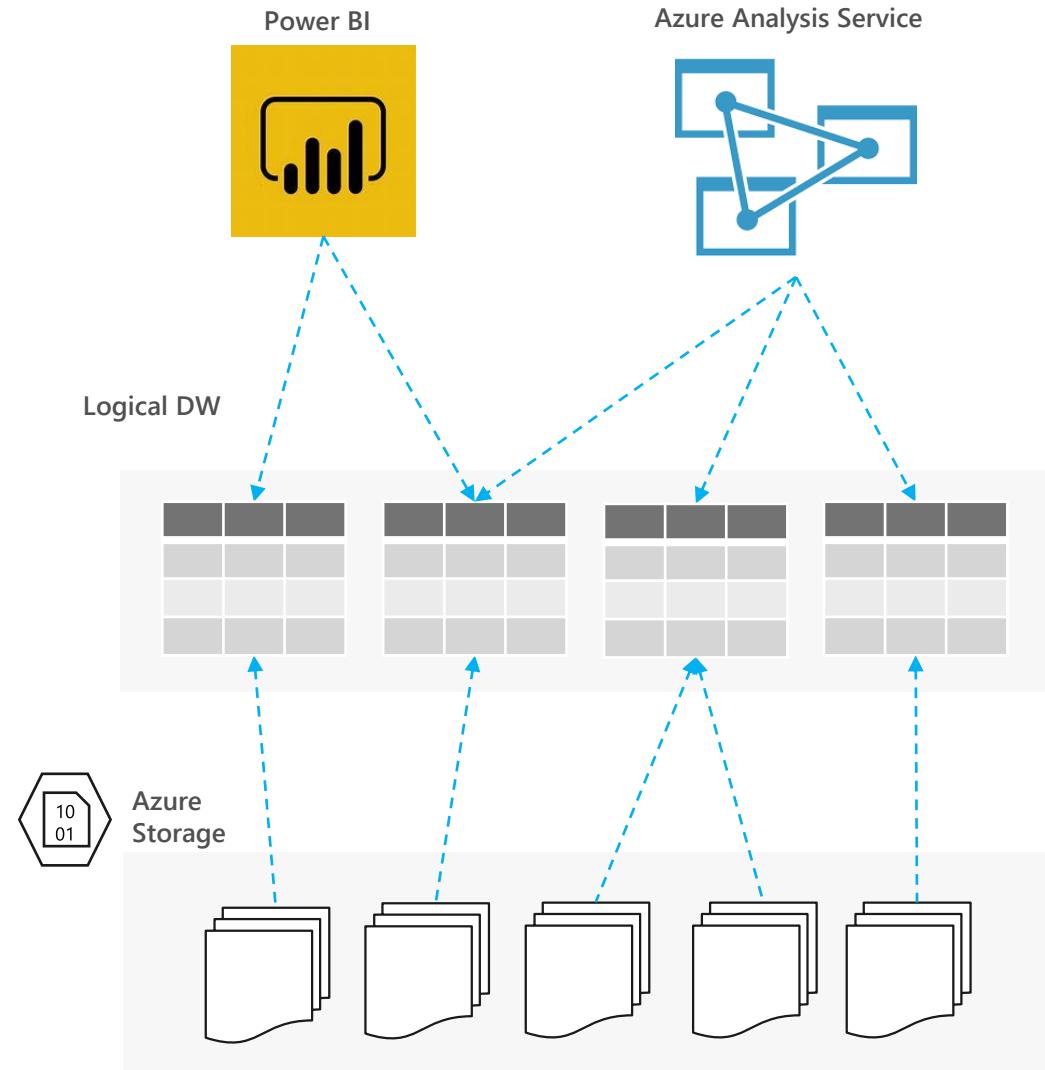
# Synapse serverless SQL pool as a logical data warehouse

## Overview

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

## Benefits

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



# Logical Data Warehouse views

## Overview

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

## Benefits

Create SQL views on externally stored data

Access files using the view from various tools and language

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

Create PowerBI reports on the views created on external data

```
USE [mydbname]
GO

DROP VIEW IF EXISTS populationView
GO

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

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

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

# Creating views

The screenshot displays three main windows from the Microsoft Azure Synapse Analytics portal:

- Top Left Window:** Shows the "Develop" tab with an open SQL script. The script creates a view named "yellow\_2017" that selects all columns from an open dataset "opendatafile" located at [https://internalsandboxwe.dfs.core.windows.net/opendatafile/nyctlc/yellow/puYear=2017/\\*/\\*](https://internalsandboxwe.dfs.core.windows.net/opendatafile/nyctlc/yellow/puYear=2017/*/*). The "Run" button is highlighted.
- Bottom Left Window:** Shows the "Results" tab with the same SQL script. The "Chart" button is highlighted. A chart is displayed below, showing the distribution of passenger counts. The Y-axis ranges from 0 to 100M, and the X-axis shows passenger counts from 0 to 10. The "cnt" series (green line with squares) peaks at 1 passenger with approximately 80M counts, while the "passengerCount" series (black line with dots) has values ranging from 0 to 5.
- Right Window:** Shows the "Develop" tab with the same SQL script. The "Table" button is highlighted. The results show the count of trips by passenger count for the year 2017, ordered by passenger count. The data is as follows:

(NO COLUMN NAME)	PASSENGERCOUNT	CNT
2017	0	166086
2017	1	81034075
2017	2	16545571
2017	3	4748869
2017	4	2257813
2017	5	5407319

A message at the bottom right indicates the query was executed successfully in 0:01:00.

# Logical Data Warehouse - tables

## Overview

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

## Benefits

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

Join and transform multiple tables in the same query.

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

Manage column statistics in external tables.

Manage access rights per table.

Create PowerBI reports on the views created on external data

```
USE [mydbname]
```

```
GO
```

```
DROP TABLE IF EXISTS dbo.Population
```

```
GO
```

```
CREATE EXTERNAL TABLE dbo.Population (
```

```
country_code VARCHAR (5) COLLATE Latin1_General_BIN2,  
country_name VARCHAR (100) COLLATE Latin1_General_BIN2,  
year smallint,  
population bigint
```

```
)
```

```
WITH(
```

```
LOCATION = '/csv/population/population-* .csv',
```

```
DATA_SOURCE = MyAzureStorage,
```

```
FILE_FORMAT = MyAzureCSVFormat
```

```
)
```

```
CREATE STATISTICS stat_country_name
```

```
ON dbo.Population(country_name);
```

```
SELECT
```

```
country_name, population
```

```
FROM population
```

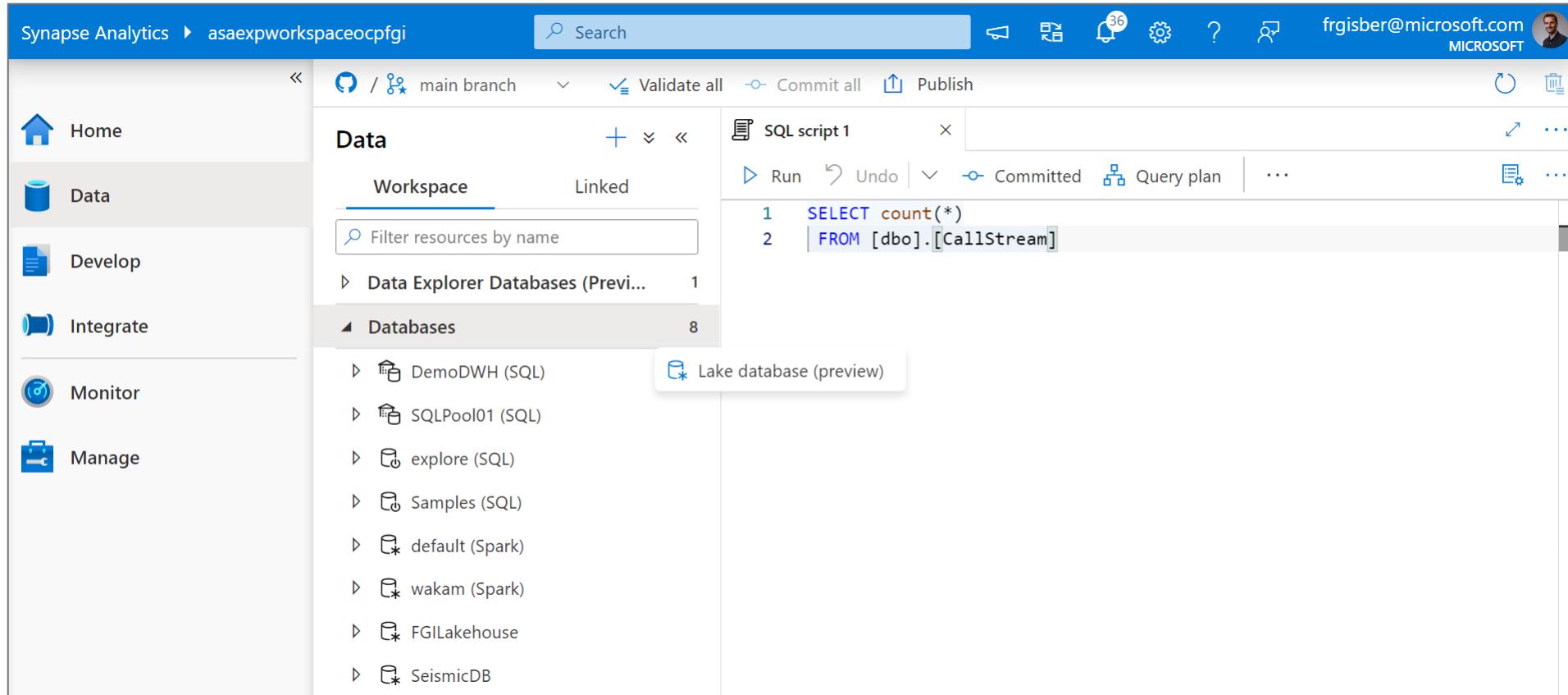
```
WHERE year = 2019
```

```
ORDER BY population DESC
```

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

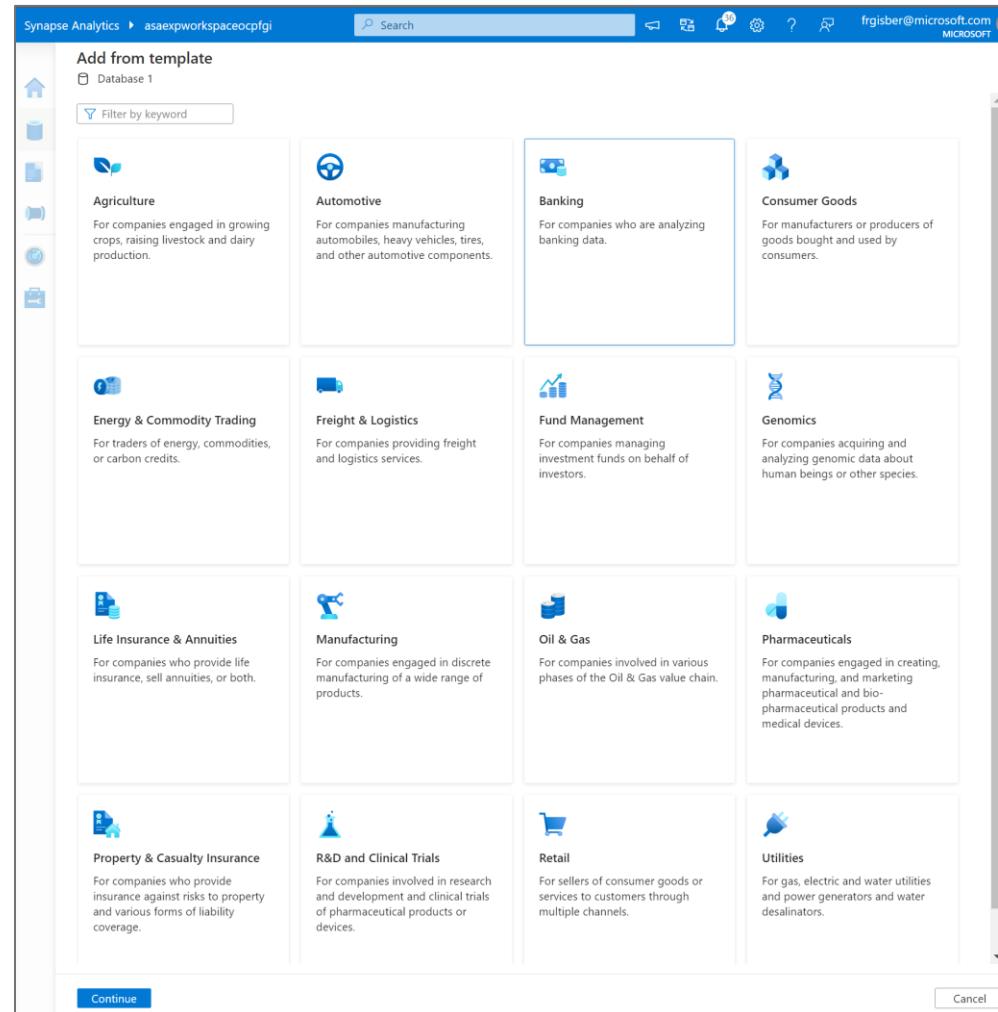
# Lakehouse Database Engine

- New Lakehouse Database capabilities based on Synapse Spark



# Data Lakehouse database

- Based on entities templates or custom design



Microsoft Azure | Synapse Analytics > asaexpworkspaceocpfgi

main branch | Validate all | Commit all | Publish | Search

**Data** | SQL script 1 | Database 1 | SeismicDB

Table | Committed

**Tables**

- Filter resources by name
- Data Explorer Databases (Prev...)** 1
  - kustopool (kustopool)
- Databases** 9
  - Database 1
  - DemoDWH (SQL)
  - SQLPool01 (SQL)
  - explore (SQL)
  - Samples (SQL)
  - default (Spark)
  - wakam (Spark)
  - FGILakehouse
- SeismicDB**
  - Tables**
    - SeismicChannelType
    - SeismicDataAcquisitionEvent
    - SeismicShot
    - ShotChannel
    - ShotFile

**SeismicShot**

- SeismicShotId PK
- SeismicShotTimestamp
- SeismicShotEnergySourceA...
- SeismicShotLocationId
- GeographicAreaId
- GeographicAreaPolygonVer...
- SeismicShotSourceArrayTy...
- SeismicShotRecordingPolar...
- SeismicShotSeismicEnergyT...
- SeismicShotSeismicSensorT...

**SeismicData AcquisitionEvent**

- SeismicDataAcquisitionEve... PK
- SeismicDataAcquisitionEve...
- SeismicDataAcquisitionEve...
- SeismicDataAcquisitionEve...
- SeismicDataAcquisitionEve...
- SurveyMethodTypeId
- NumberOfChannels

**ShotFile**

- SeismicDataFileId PK
- SeismicShotId PK,FK
- ShotFileNote

**ShotChannel**

- SeismicShotId PK,FK
- ShotChannelId PK
- SeismicDataFileId PK,FK
- PeriodStartTimestamp PK
- PeriodEndTimestamp
- SeismicChannelTypeId FK
- ShotChannelNote

**General Columns Relationships**

Filter by keyword | + Column | Clone | Delete

Name	Keys	Description	Nullability	Data type	Format / Length
SeismicDataFileId	PK	The unique identifier of a seismic data file.	Null	integer	1024
SeismicShotId	PK, FK	The unique identifier of a seismic shot.	Null	integer	1024
ShotFileNote	PK	A note, comment or additional information regarding the shot file.	Null	string	1024

# Spécificités des moteurs SQL

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

# Spécificités du moteurs Spark

- Le moteur Spark est disponible dans le service Azure Synapse
- Ce moteur possède deux "form factor"
  - Cluster Spark à disposition dans un format managé, "auto scalable" et activé à la demande
  - Objets Spark disponibles en mode "serverless"
- Possibilité de matérialiser les résultats des travaux de recherches, analyses, etc. dans des tables Spark
- Les tables Spark sont exposées en mode "Serverless"
- L'accès est réalisé via le point d'accès SQL Serverless sans que le moteur Spark soit démarré

# Power BI Embedded pour l'exposition et REST API

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

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

## Getting started

[Set up your Power BI embedding environment](#)  
[Power BI JavaScript API wiki](#)  
[Power BI embedding documentation](#)

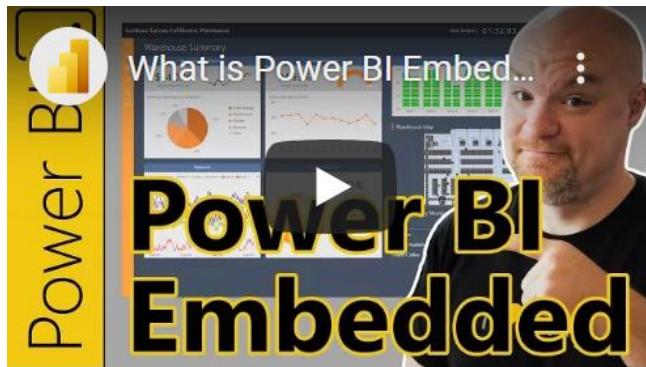
## Useful links

[Power BI Embedded on Azure](#)  
[Power BI community](#)  
[Power BI Ideas - APIs and embedding](#)

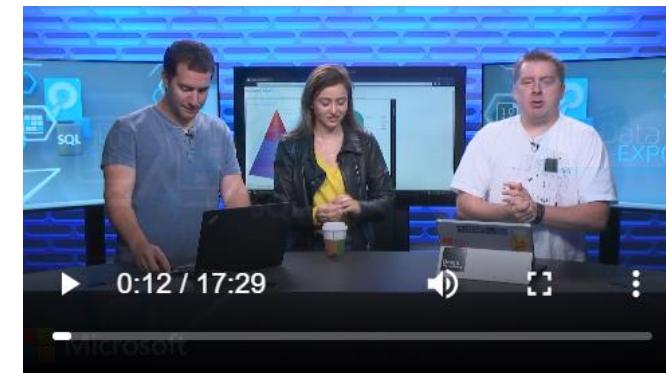
## Support

[Power BI Embedded FAQ](#)  
[Power BI Embedded troubleshooting](#)  
[Power BI Support](#)

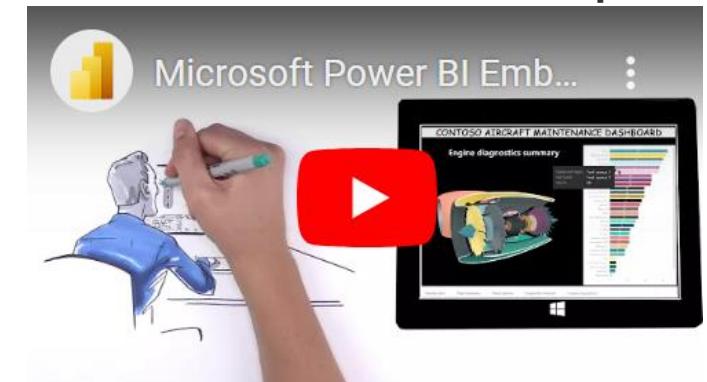
## What is Power BI Embedded



## Microsoft Power BI Embedded update



## Microsoft Power BI Embedded update



<https://playground.powerbi.com/>

# Shared datasets - paginated reports

offline/online distribution through

The screenshot shows a paginated report interface. At the top, there's a navigation bar with File, View, Export, Parameters, Print, Comments, Subscribe, and Share buttons. A dropdown menu says "Select to View Performance" with "Mickey" selected. A yellow "View report" button is also present. On the left, a sidebar has icons for Home, Favorites, Recent, and various report types. The main content area has a yellow header "Garage sale results report for Mickey". Below it, there's a section for "Badge Photo" (empty), "Age: 6", "Birthplace: Iowa", "Year Hired: 2016", and "Manager Comments: Bit of a troublemaker. Loves Paginated Reports". Under "Items Sold", there's a table:

Category	9/1/2018	9/8/2018	Total
Electronics	\$32.00	\$75.00	<b>\$107.00</b>
Furniture		\$400.00	<b>\$400.00</b>
Handbag		\$6.00	<b>\$6.00</b>
Household Items	\$21.00	\$15.00	<b>\$36.00</b>
Kitchen Items		\$17.00	<b>\$17.00</b>
Toys	\$63.00	\$51.00	<b>\$114.00</b>
<b>Total</b>	<b>\$116.00</b>	<b>\$564.00</b>	<b>\$680.00</b>

At the bottom, a footer says "© 2018 Paginated Report Bear".

## Subscribe to emails

SAMPLE PAGINATED REPORT

This is a configuration page for email subscriptions. It includes a yellow header "Add new subscription" and a yellow "sample paginated report" section. It features a "Run Now" button with a "On" toggle switch. A dropdown menu for "Format" is open, showing options like PDF (.pdf) (which is selected), XML (.xml), Comma Separated Values (.csv), PDF (.pdf), Accessible PDF (.pdf), Microsoft Excel (.xlsx), Microsoft Word (.docx), and Microsoft PowerPoint (.pptx). There are fields for "Include an optional message...", "Frequency" (set to "Monthly"), "Every month on day(s)" (set to "3"), and a "Scheduled Time" field with dropdowns for hour, minute, AM/PM, and time zone "(UTC+01:00) Brussels, Copenhagen".

# Demo