

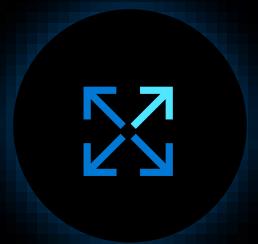
A close-up photograph of several incandescent lightbulbs hanging from a cord. One bulb in the foreground is brightly lit, casting a warm glow. Behind it, other bulbs are partially visible, some also glowing. The background is dark, making the light from the bulbs stand out.

Intelligent Data Platform

Andrea Benedetti

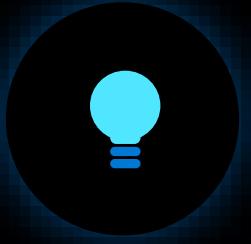
Sr. Cloud Architect | Data & AI Engineer

Your inputs guide our strategy



Scale

Limitless



A.I.

Intelligent by default



Hybrid

Operational freedom



Security

Always a step ahead

Azure

Data



Data modernization
on-premises



Data modernization
to Azure



Globally
distributed data



Cloud scale
analytics

AI



AI apps & agents



Knowledge mining



Machine learning



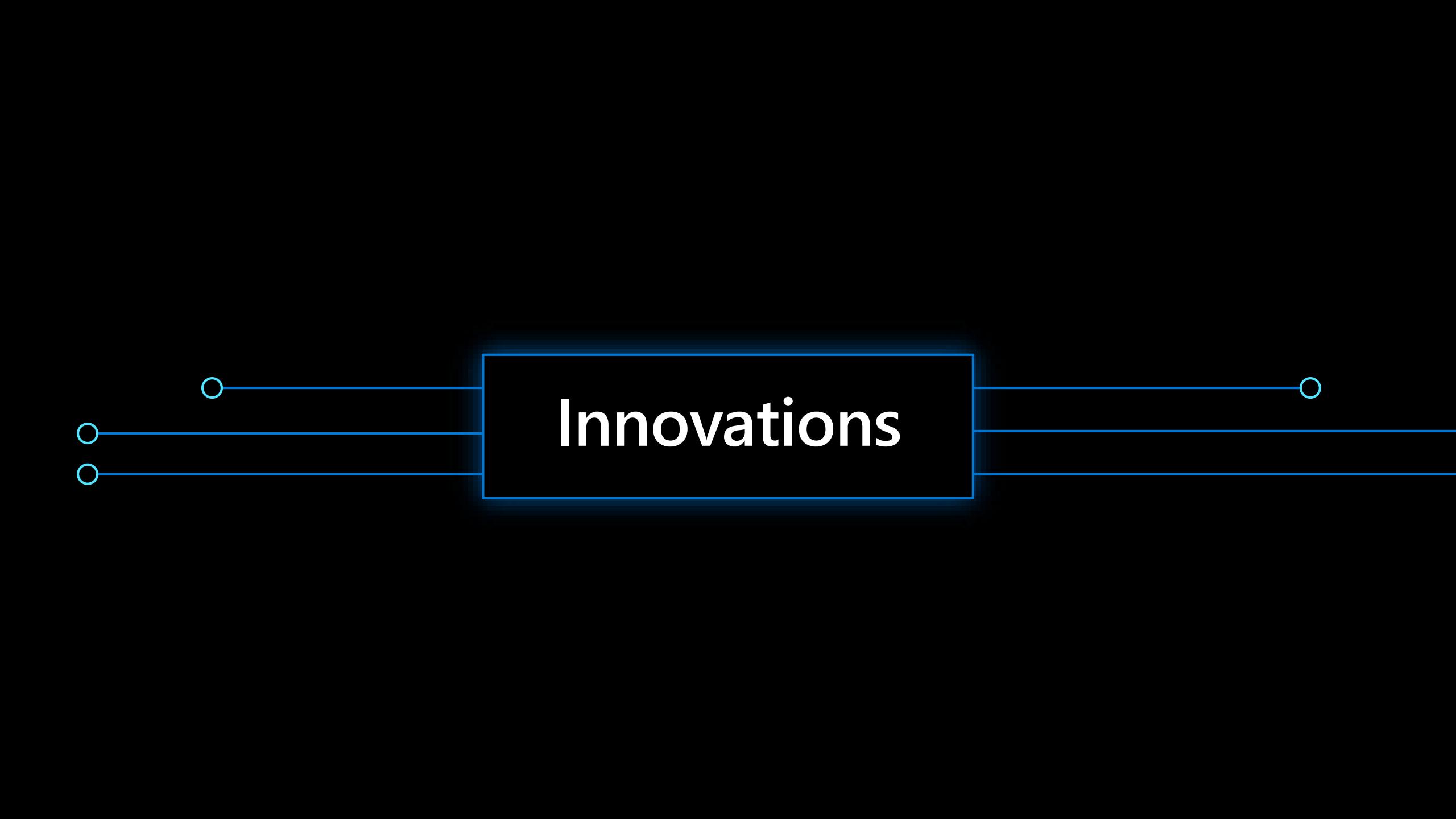
Self-service BI



On-premises BI



Apps with
intelligence



Innovations

Azure SQL Database Edge

Small footprint, optimized for Edge with built-in AI



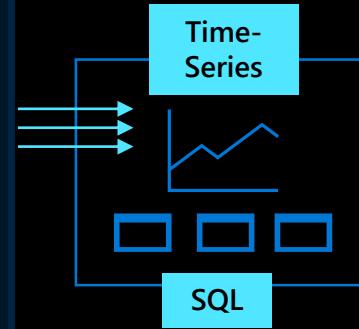
Choice of platform

Arm64 x64



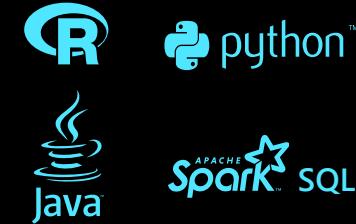
Small-footprint engine
300MB

Time-series built-in



Built-In Time-Series
Streaming and Analytics

AI built-in



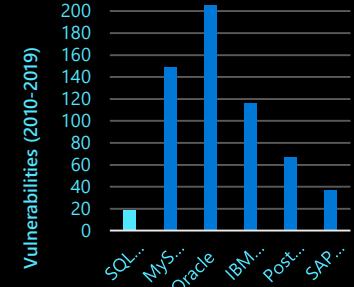
Machine learning for
low-latency analytics

Turn-key management



Single management plane
to deploy & update

Unparalleled performance and security



Most secure with industry
leading performance



Edge

Local streaming, storage, analytics all-in-one

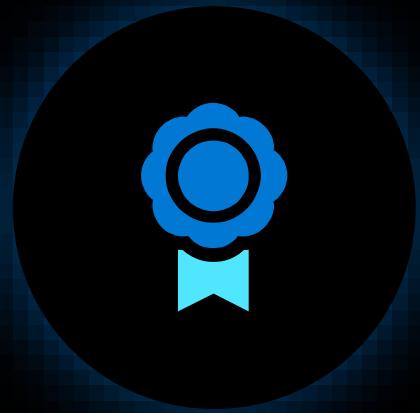
Develop once, deploy anywhere for your apps



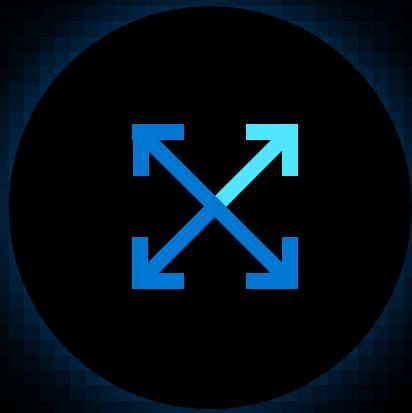
Data Center/Cloud

Azure SQL Database Hyperscale

Adapts on-demand to your workload's needs, auto-scaling up to 100TB per database.



Reliable and available



Scalable



High performance

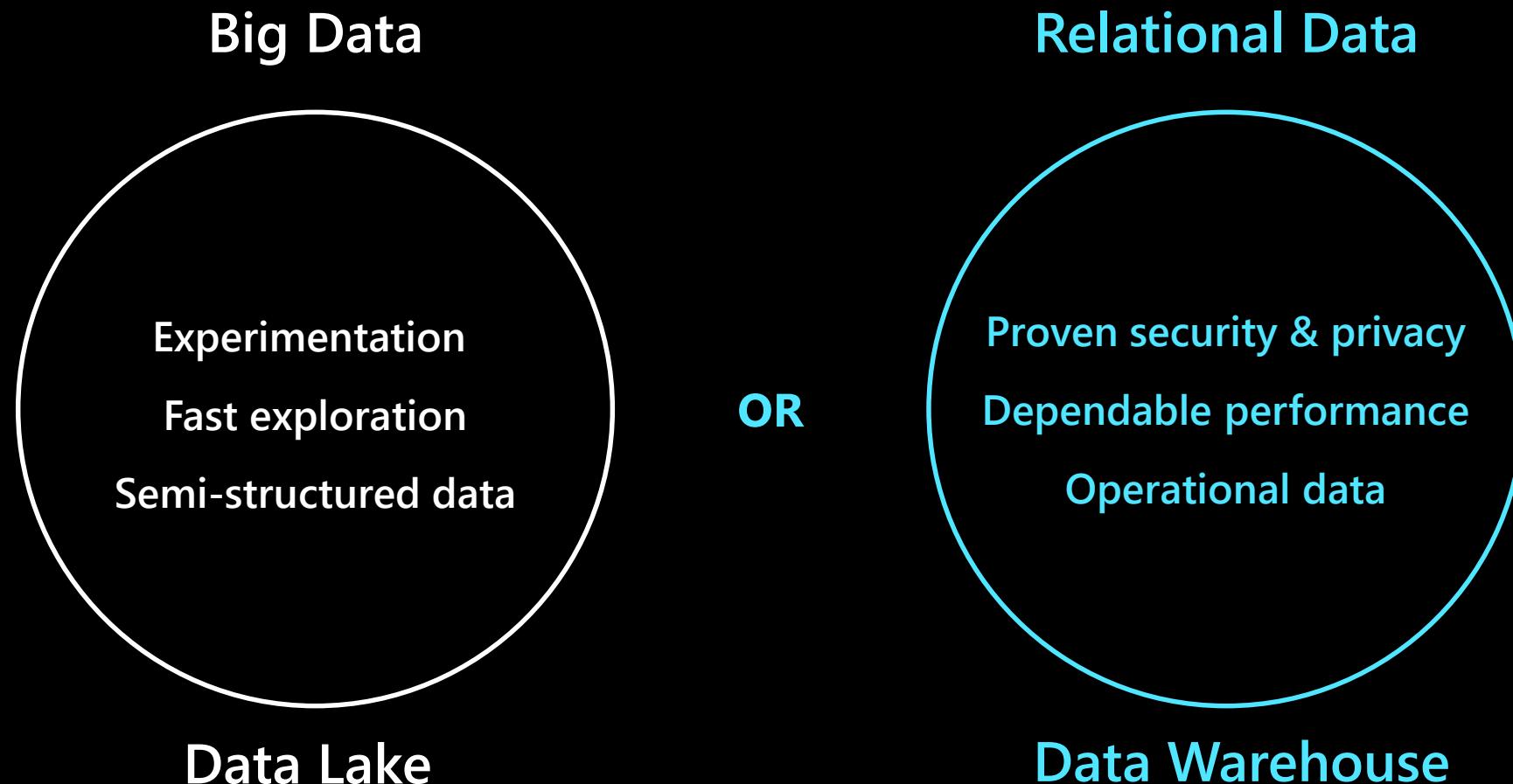
Best for VLDB workloads with highly scalable storage and read-scale requirements, optimized for OLTP and HTAP workloads.



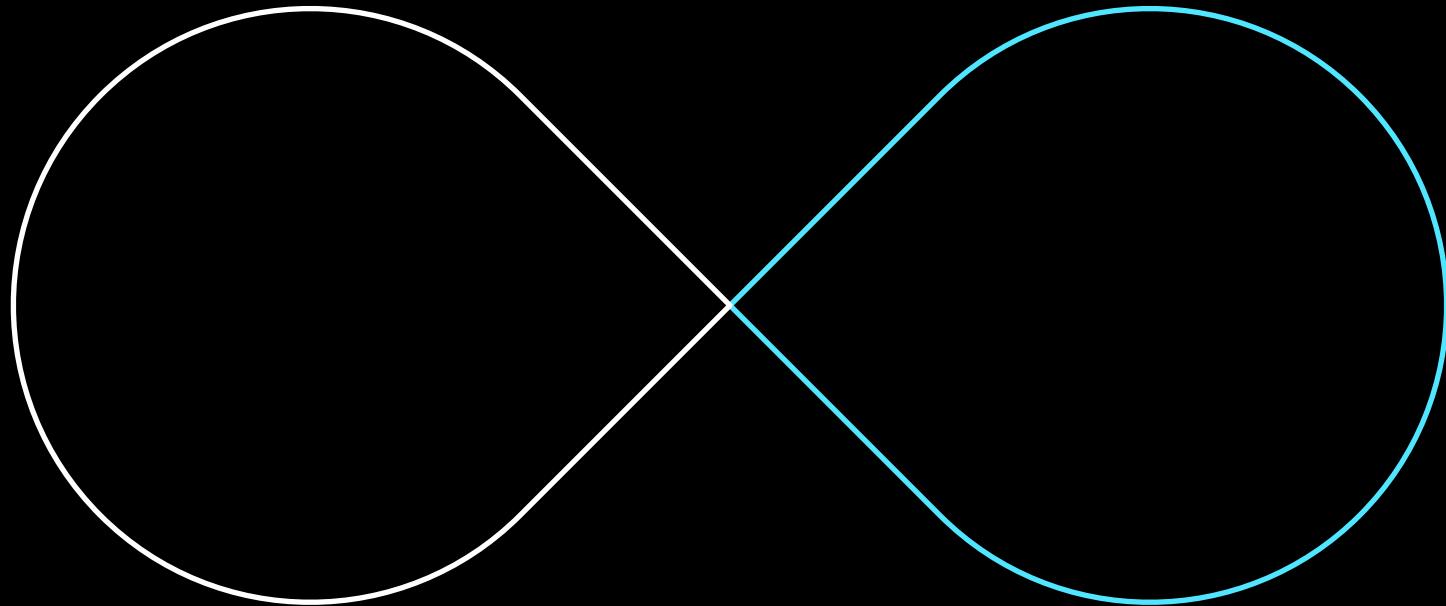
Announcing

Azure Synapse

This is a result of businesses being forced to maintain two critical, yet independent analytics systems



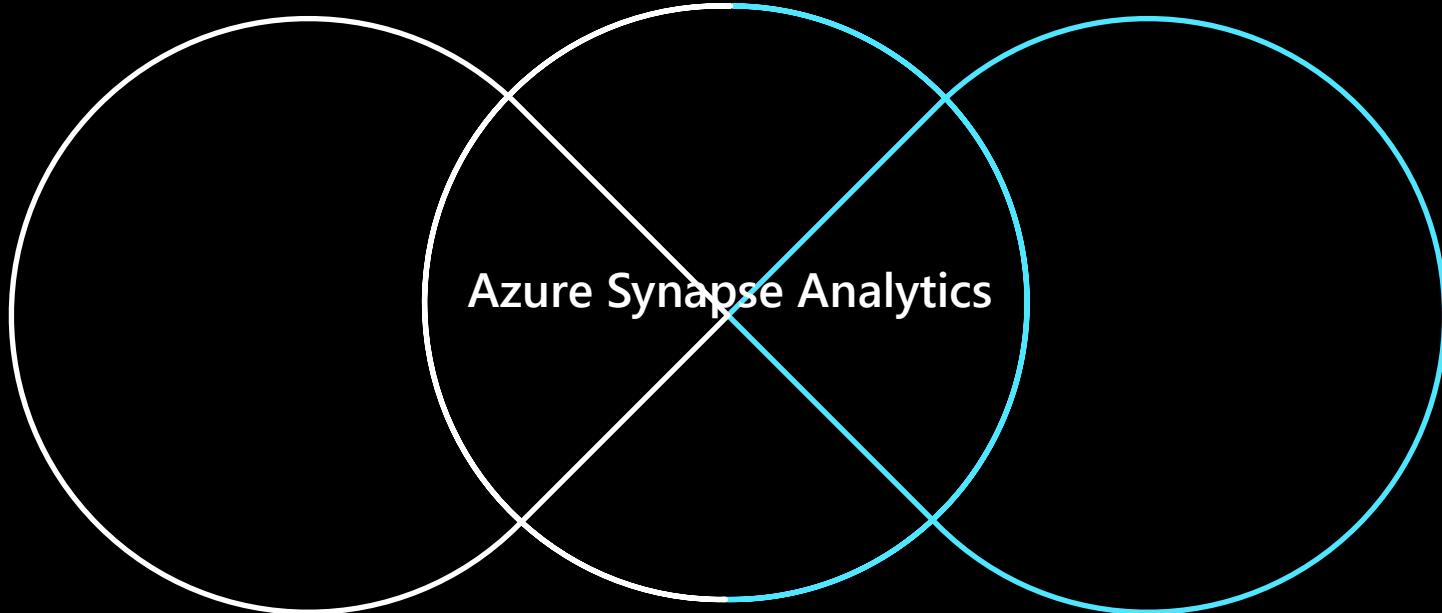
Azure brings these two worlds together, in a single service,
to provide limitless analytics



Welcome to **limitless**

Data warehousing & big data analytics—all in one service

Azure brings these two worlds together, in a single service,
to provide limitless analytics

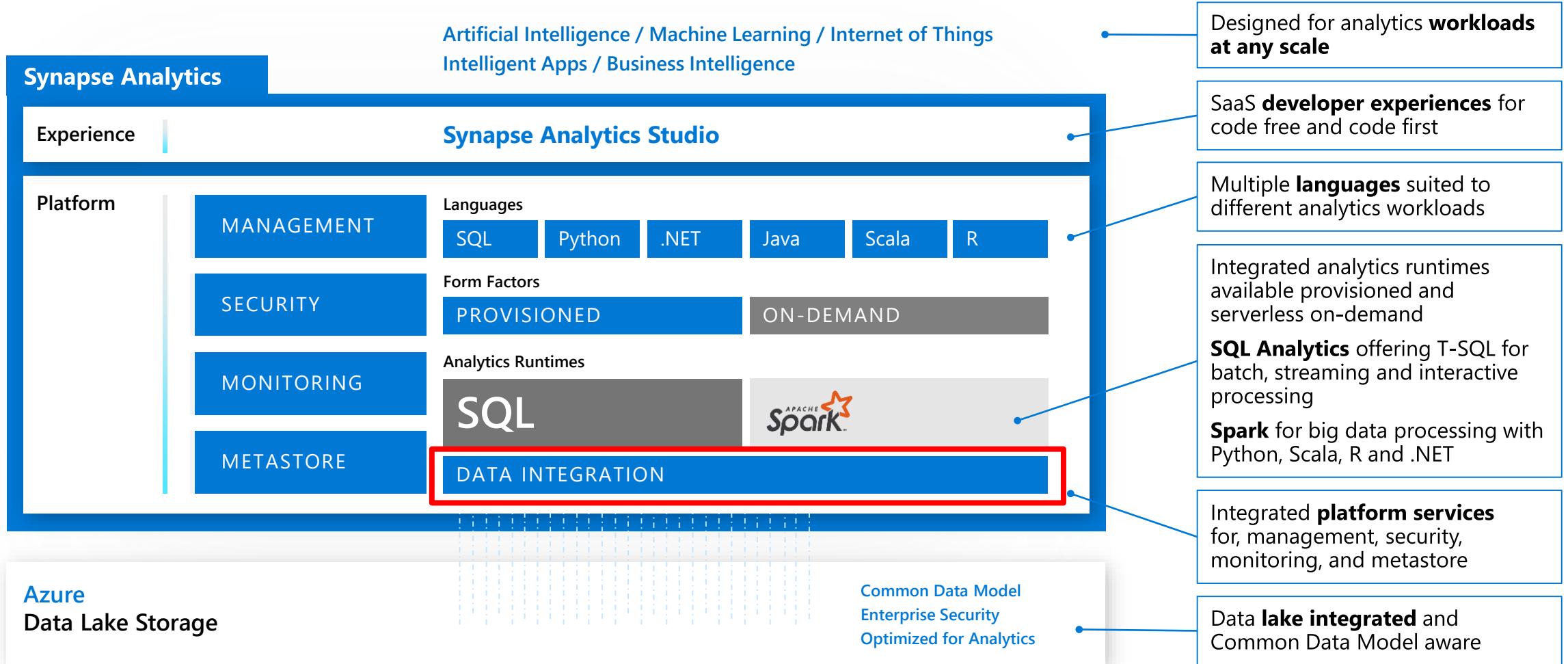


Welcome to **limitless**

Data warehousing & big data analytics—all in one service

Azure Synapse Analytics

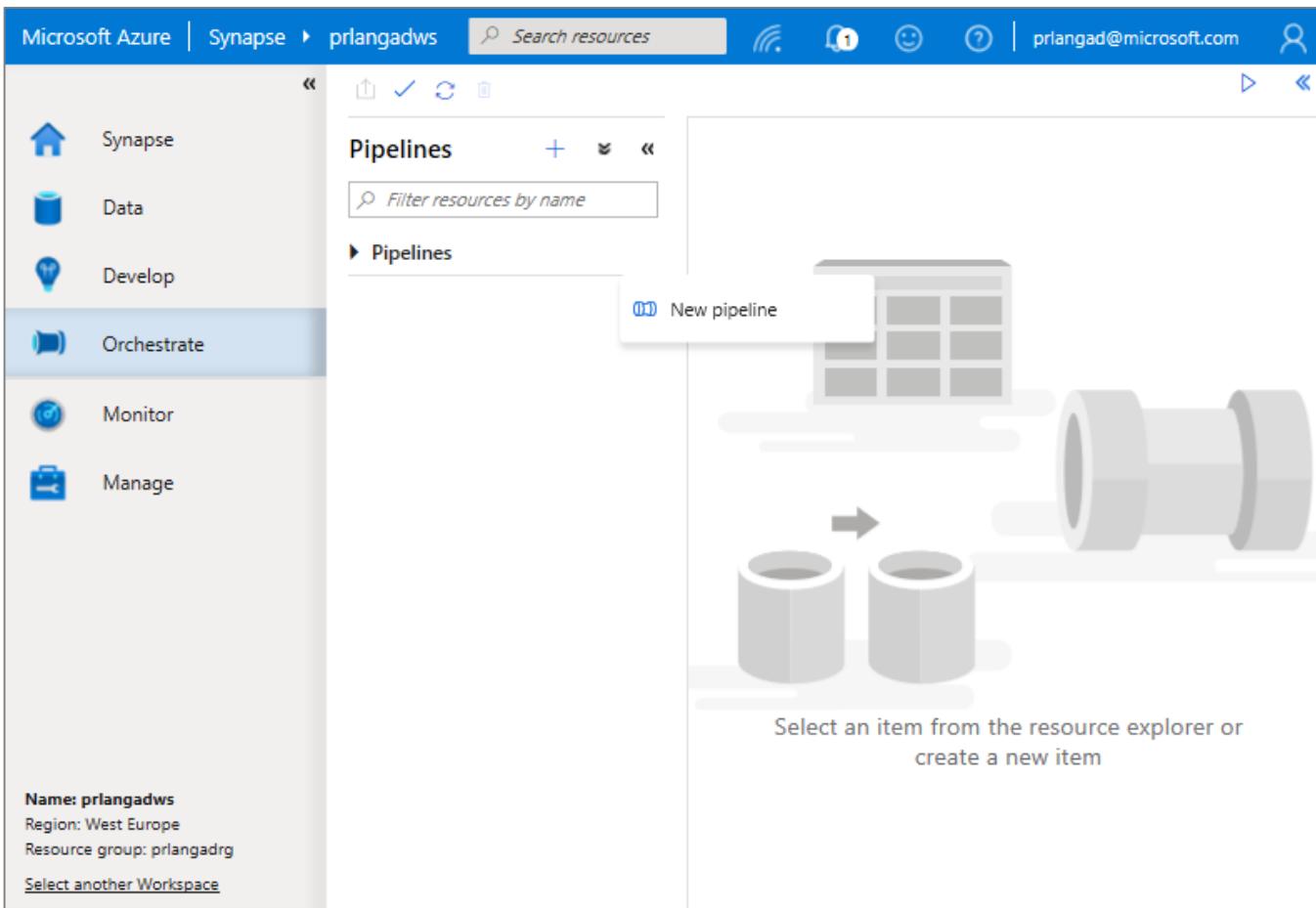
Integrated data platform for BI, AI and continuous intelligence



Pipelines

Overview

- Scalable – per job elasticity up to 4GB/s
- Visually author or via code (Python, Spark(Scala), Spark.NET,(C#) etc.)
- Serverless, no infrastructure to manage
- 24 points of presence worldwide
- Self-hostable Integration Runtime for hybrid movement
- Supported file formats: CSV, AVRO, ORC, Parquet, JSON



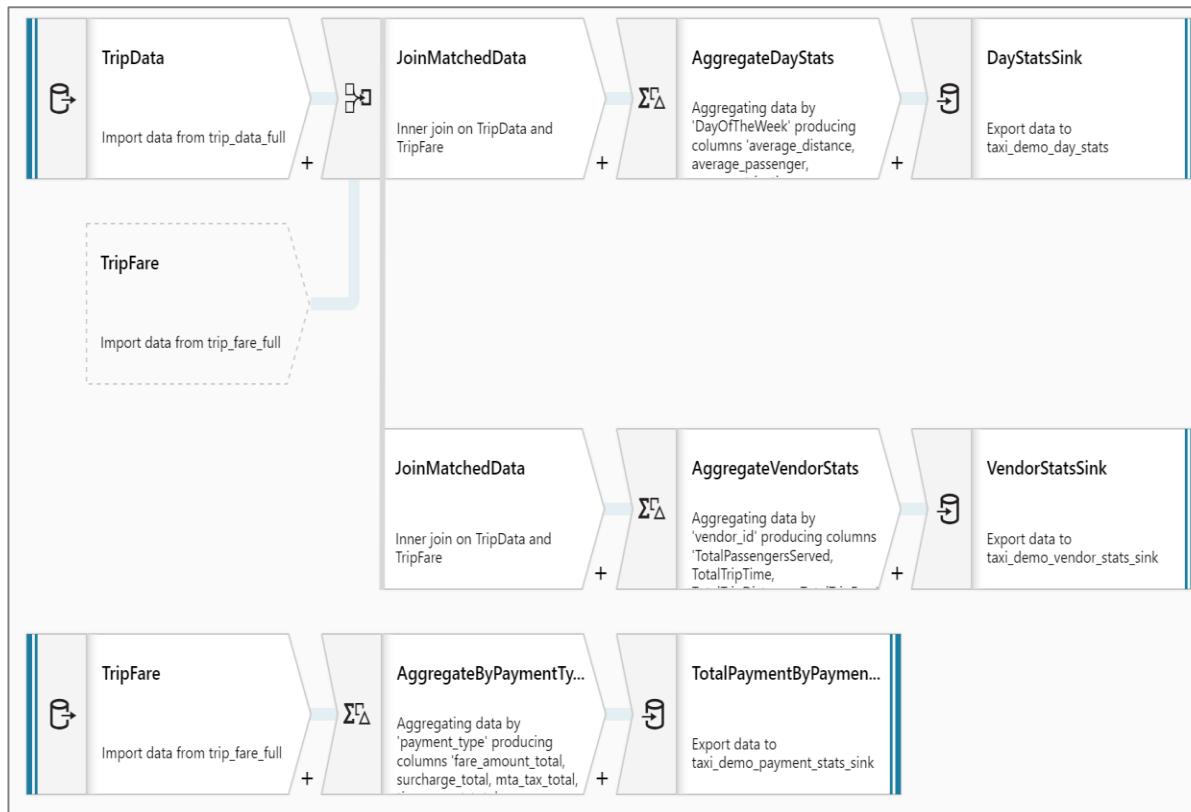
90+ Connectors out of the box

Azure (15)	Database & DW (26)		File Storage (6)	NoSQL (3)	Services and App (28)			Generic (4)	
Blob storage	Amazon Redshift		Oracle	Amazon S3	Cassandra	Amazon MWS	Oracle Service Cloud	Generic HTTP	
Cosmos DB - SQL API	DB2		Phoenix	File system	Couchbase	Common Data Service	PayPal	Generic OData	
Cosmos DB - MongoDB API	Drill		PostgreSQL	FTP	MongoDB	Concur	QuickBooks	Generic ODBC	
Data Explorer	Google BigQuery	Presto	Google Cloud Storage		Dynamics 365	Salesforce	Generic REST		
Data Lake Storage Gen1	Greenplum	SAP BW Open Hub	HDFS		Dynamics AX	Salesforce Service Cloud			
Data Lake Storage Gen2	HBase	SAP BW via MDX	SFTP		Dynamics CRM	Salesforce Marketing Cloud			
Database for MariaDB	Hive	SAP HANA	Google AdWords		SAP Cloud for Customer (C4C)				
Database for MySQL	Apache Impala	SAP table	HubSpot		SAP ECC				
Database for PostgreSQL	Informix	Spark	Jira		ServiceNow				
File Storage	MariaDB	SQL Server	Magento		Shopify				
SQL Database	Microsoft Access	Sybase	Marketo		Square				
SQL Database MI	MySQL	Teradata	Office 365		Web table				
SQL Data Warehouse	Netezza	Vertica	Oracle Eloqua		Xero				
Search index					Oracle Responsys	Zoho			
Table storage									

Prep & Transform Data

Mapping Dataflow

Code free data transformation @scale

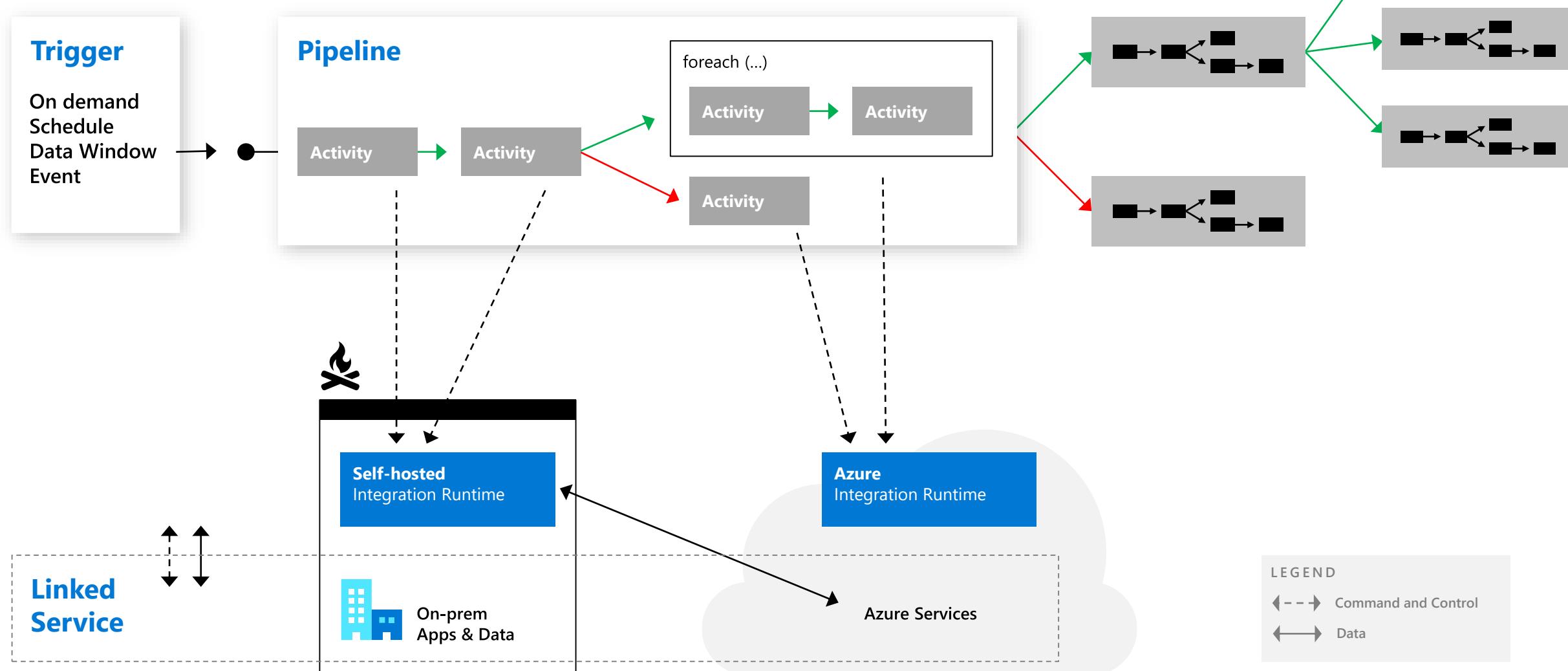


Wrangling Dataflow

Code free data preparation @scale

CustId	FirstName	LastName	City	ZIP	Email	State	BasePay
1	'Harry'	'Potter'	'Bellevue'	'98004'	'harryk@fabrikam.com'	'WA'	90000
2	'Harry'	'Potter'	'Bellevue'	'98004'	'harryk@fabrikam.com'	'WA'	90000
3	'Hermione'	'Granger'	'Wilmington'	'19801'	'hermione@fabrikam.com'	'DE'	100000
4	'Hermione'	'Granger'	'Wilmington'	'19801'	'gamalloy@fabrikam.com'	'DE'	100000
5	'Lord'	'Voldemort'	'Billings'	'59115'	'lordc@fabrikam.com'	'MT'	110000
6	'Albus'	'Dumbledore'	'Newyork'	'12345'	'albusd@fabrikam.com'	'NY'	120000
7	'Severus'	'Snape'	'Columbus'	'56789'	'severus@fabrikam.com'	'OH'	130000
8	'Draco'	'Malfoy'	'Houston'	'91019'	'dracoh@fabrikam.com'	'TX'	140000
9	'Dobby'	'Elf'	'Salt Lake C...	'11128'	'dobbyz@fabrikam.com'	'UT'	150000
10	'Ron'	'Weasley'	'Las Vegas'	'51527'	'ronag@fabrikam.com'	'NV'	160000
11	'Sirius'	'Black'	'Providence'	'61623'	'hblack@fabrikam.com'	'RI'	170000
12	'Luna'	'Lovegood'	'Kansas City'	'68692'	'lunal@fabrikam.com'	'MO'	180000
13	'Rubeus'	'Hagrid'	'Boston'	'98052'	'gammaloy@fabrikam.com'	'Malfoy'	190000
14	'Bellatrix'	'Lestrange'	'Los Angeles'	'78965'	'mlestrange@fabrikam.com'	'CA'	200000
15	'Ginny'	'Weasley'	'Redmond'	'98052'	'ginnyw@fabrikam.com'	'WA'	210000
16	'Neville'	'Longbottom'	'Bothell'	'98053'	'nevilea@fabrikam.com'	'WA'	220000
17	'Alastor'	'Moody'	'Renton'	'98054'	'albusd@fabrikam.com'	'WA'	230000
18	'Lucius'	'Malfoy'	'Bellevue'	'98004'	'luciusmalfoy@fabrikam.co...	'WA'	240000
19	'Cedric'	'Diggory'	'Seattle'	'98989'	'cedricp@fabrikam.com'	'WA'	250000
20	'Argus'	'Filch'	'Salt Lake C...	'11128'	'argusm@fabrikam.com'	'UT'	260000
21	'Lord'	'Voldemort'	'Billings'	'59115'	'lordc@fabrikam.com'	'MT'	110000
22	'Albus'	'Dumbledore'	'Newyork'	'12345'	'albusd@fabrikam.com'	'NY'	120000
23	'Severus'	'Snape'	'Columbus'	'56789'	'severus@fabrikam.com'	'OH'	130000
24	'Mason'	'Bartlet'	'Bellmore'	'98004'	'masonm@fabrikam.com'	'MA'	00000

Orchestration @ Scale



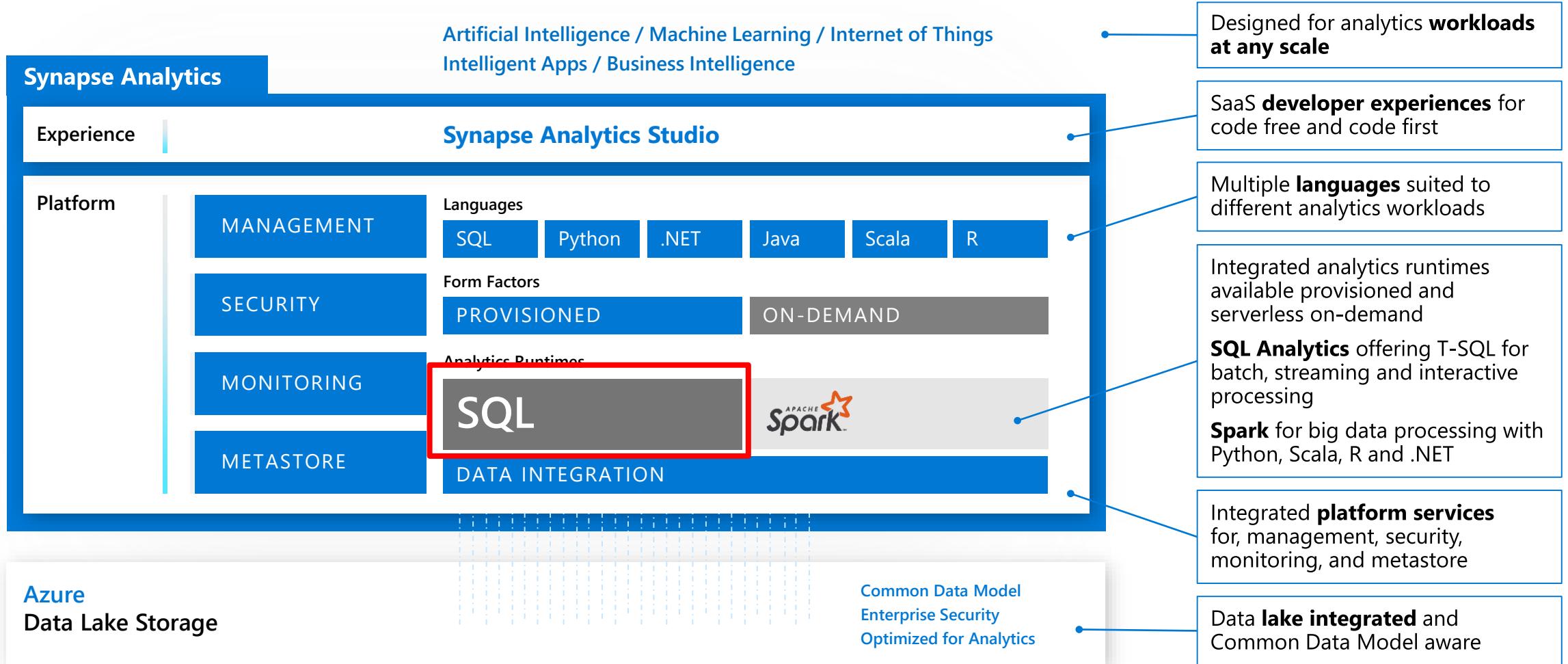


Azure Synapse Analytics

SQL Analytics

Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Azure Advisor recommendations

Suboptimal Table Distribution

Reduce data movement by replicating tables

Data Skew

Choose new hash-distribution key

Slowest distribution limits performance

Cache Misses

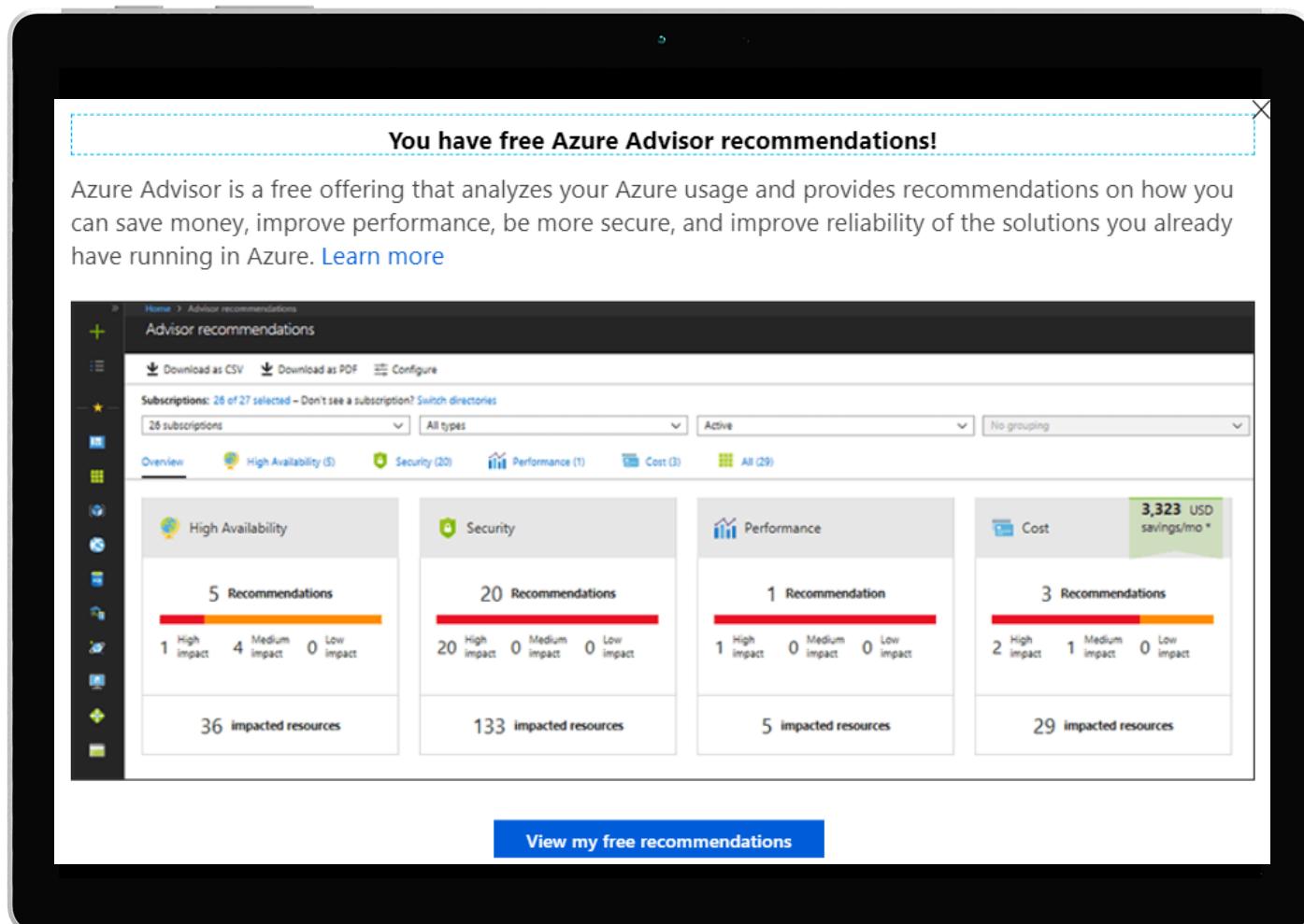
Provision additional capacity

Tempdb Contention

Scale or update user resource class

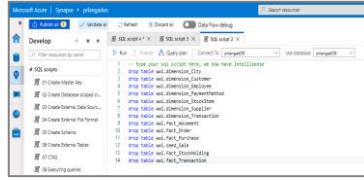
Suboptimal Plan Selection

Create or update table statistics



Developer Tools

Azure Synapse Analytics

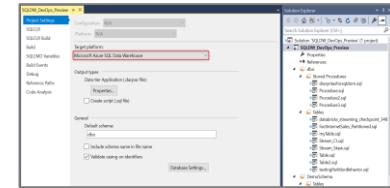


Azure Cloud Service

Offers end-to-end lifecycle for analytics

Connects to multiple services

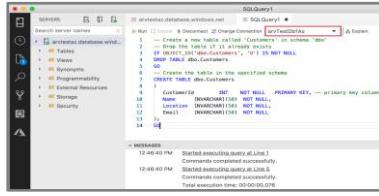
Visual Studio - SSDT database projects



Runs on Windows

Create, maintain database code, compile, code refactoring

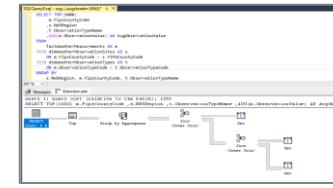
Azure Data Studio



Runs on Windows, Linux, macOS

Light weight editor, (queries and extensions)

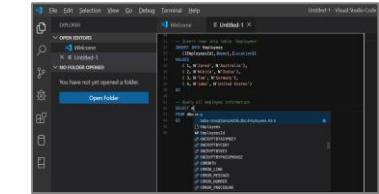
SQL Server Management Studio



Runs on Windows

Offers GUI support to query, design and manage

Visual Studio Code



Runs on Windows, Linux, macOS

Offers development experience with light-weight code editor

Continuous integration and delivery (CI/CD)

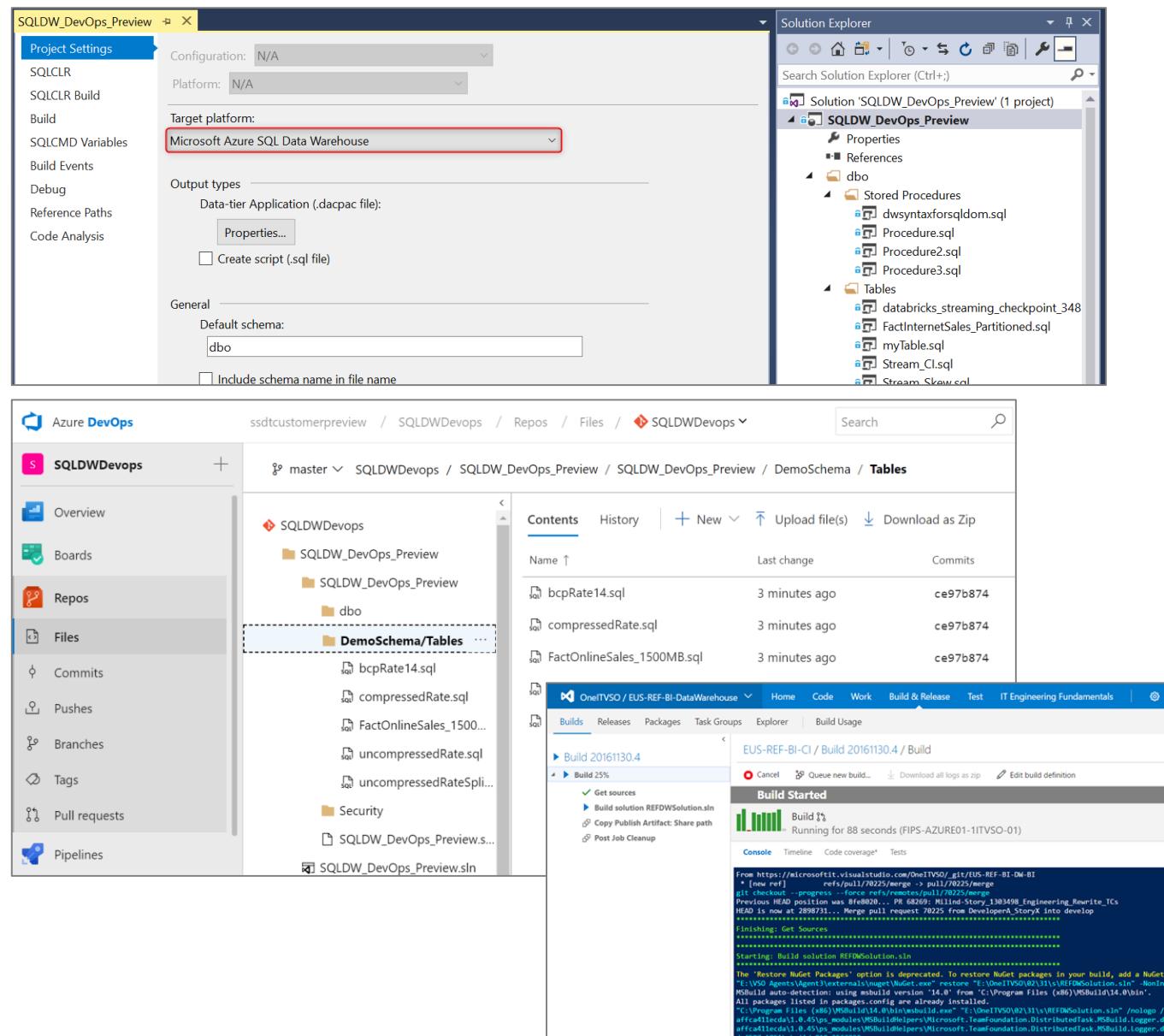
Overview

Database project support in SQL Server Data Tools (SSDT) allows teams of developers to collaborate over a version-controlled data warehouse, and track, deploy and test schema changes.

Benefits

Database project support includes first-class integration with Azure DevOps. This adds support for:

- Azure Pipelines** to run CI/CD workflows for any platform (Linux, macOS, and Windows)
- Azure Repos** to store project files in source control
- Azure Test Plans** to run automated check-in tests to verify schema updates and modifications
- Growing ecosystem of third-party integrations that can be used to complement existing workflows (Timetracker, Microsoft Teams, Slack, Jenkins, etc.)



SQL On-Demand
SQL On Demand

SQL On-Demand

Overview

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

Benefits

Serverless

No infrastructure

Pay only for query execution

No ETL

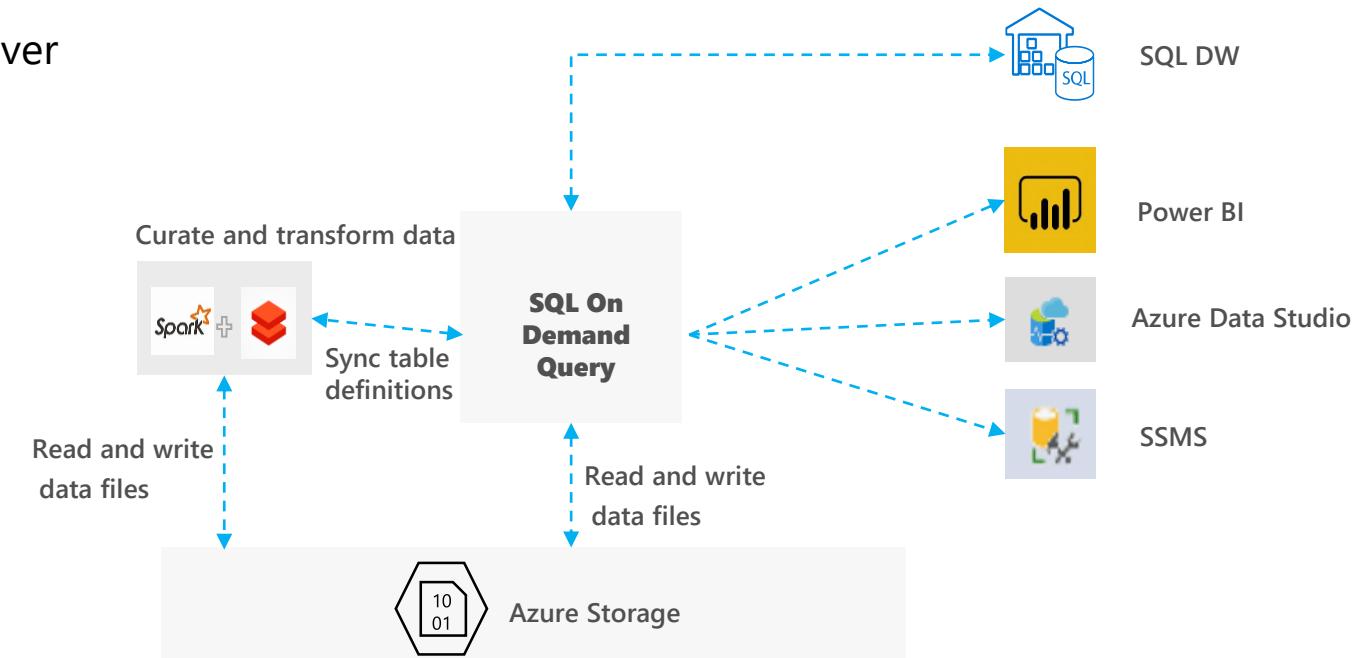
Offers security

Data integration with Databricks, HDInsight

T-SQL syntax to query data

Supports data in various formats (Parquet, CSV, JSON)

Support for BI ecosystem

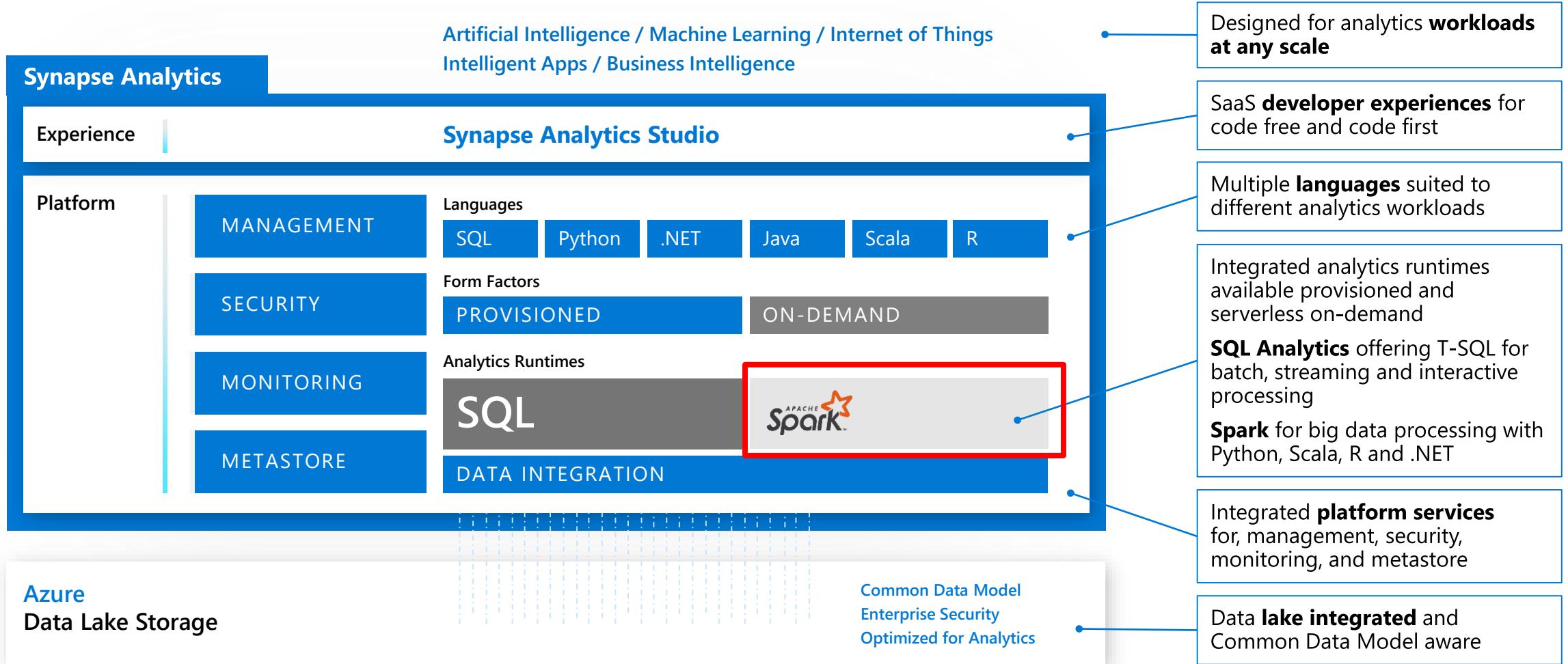




Azure Synapse Analytics Spark

Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Library Management - Python

Overview

Customers can add new python libraries at Spark pool level

Benefits

Input requirements.txt in simple pip freeze format

Add new libraries to your cluster

Update versions of existing libraries on your cluster

Libraries will get installed for your Spark pool during cluster creation

Ability to specify different requirements file for different pools within the same workspace

Constraints

The library version must exist on PyPI repository

Version downgrade of an existing library not allowed

In the Portal

Specify the new requirements while creating Spark Pool in Additional Settings blade

Microsoft Azure (Preview) Restore default configuration Report a bug Search resources, services, and data

Home > nushuklasynapsewestus2 > Create Apache Spark pool

Create Apache Spark pool

Enter required settings for this Apache Spark pool, including setting auto-pause and picking versions.

Auto-pause * Enabled Disabled

Number of minutes idle *

Component versions

Select the Apache Spark version for your Apache Spark pool.

Component	Version
Apache Spark *	2.4
Python	3.6.1
Scala	2.11.12
Java	1.8.0_222
.NET Core	3.0
.NET for Apache Spark	0.6.0
Delta Lake	0.4.0

Packages

Upload environment configuration file ("PIP freeze" output).

File upload Upload

Review + create < Previous Next: Tags >

Languages

Overview

Supports multiple languages to develop notebook

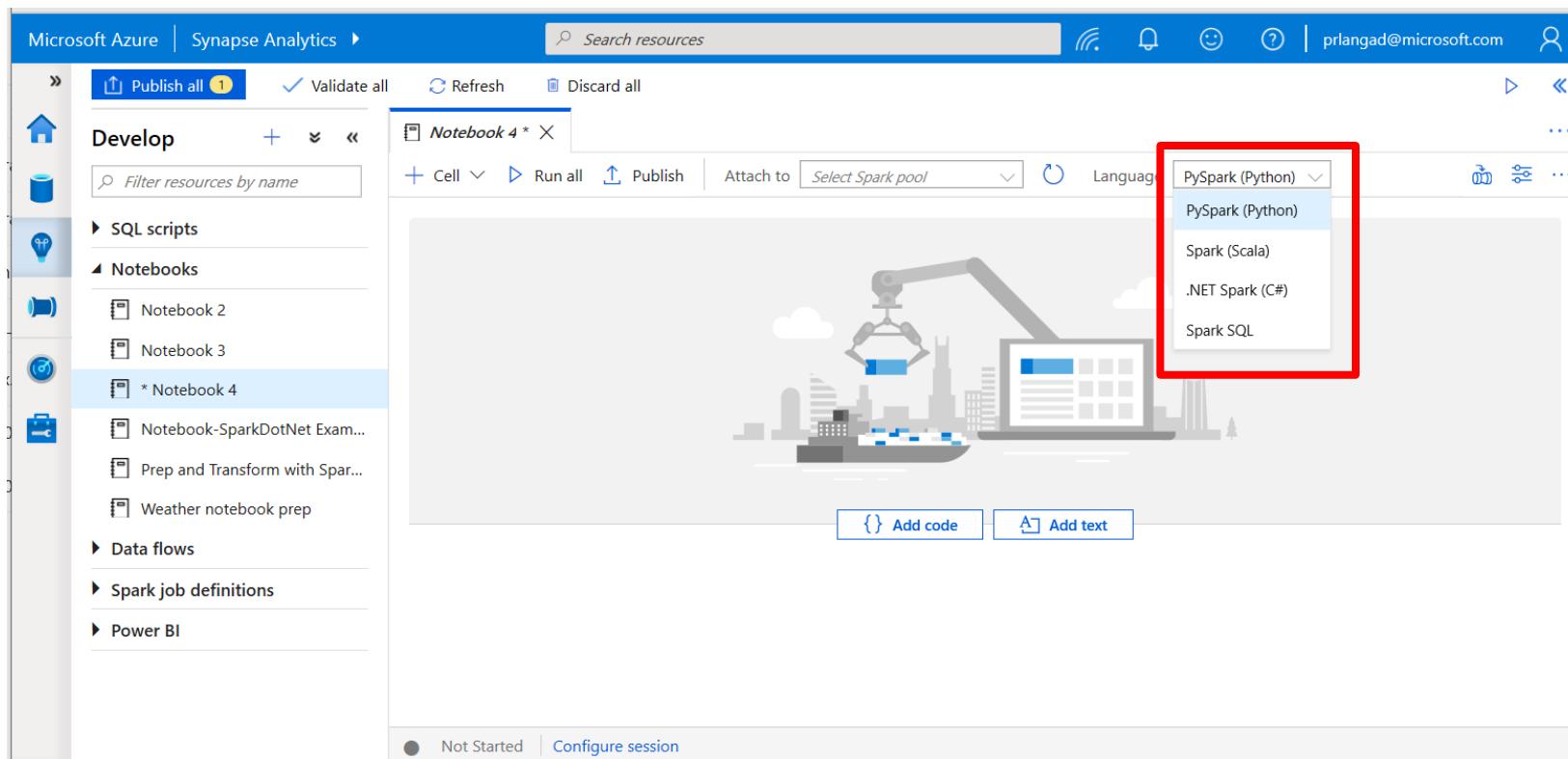
- PySpark (Python)
- Spark (Scala)
- .NET Spark (C#)
- Spark SQL

Benefits

Allows to write multiple languages in one notebook

%%<Name of language>

Offers use of temporary tables across languages



INDUSTRY LEADING SECURITY

Category	Feature	SQL Data Warehouse	Amazon Redshift	Snowflake	Google Big Query
Data Protection	Data In Transit	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>	No
	Data encryption at rest (Service & User Managed Keys)	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>
	Data In Use (Always Encrypted)	No	No	No	No
	Data Discovery and Classification	Yes (March 2019)	No	No	No
Access Control	Native Row Level Security	<u>Yes</u>	No	No	No
	Table and View Security (GRANT / DENY)	Yes	Yes	Yes	Yes
	Column Level Security	<u>Yes</u>	No	No	No
Authentication	SQL Authentication	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>	No
	Native Azure Active Directory	Yes	No	Yes	No
	Integrated Security	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>
	Multi-Factor Authentication	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>
Network Security	Virtual Network (VNET)	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>
	SQL Firewall (server)	<u>Yes</u>	<u>Yes</u>	<u>Yes</u>	No
Threat Protection	SQL Threat Detection	<u>Yes</u>	<u>Yes</u>	No	No
	SQL Auditing	<u>Yes</u>	<u>Yes</u>	No	<u>Yes</u>
	Vulnerability Assessment	<u>Yes</u>	<u>Yes</u>	No	No

The Azure data landscape



Azure Data Factory



Azure Import/Export service



Azure CLI



Azure SDK



Azure IoT Hub



Azure Event Hubs



Kafka on Azure HDInsight



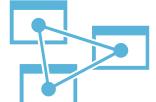
Azure SQL DB



Azure Cosmos DB



Azure Synapse Analytics



Azure Analysis Services



Power BI



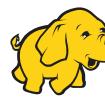
Azure Blob Storage



Azure Data Lake Store



Azure Data Lake Analytics



Azure HDInsight



Azure Databricks



Azure ML



ML Server



Azure Databricks



Azure Search



Azure Data Catalog



Azure Stream Analytics



Azure HDInsight



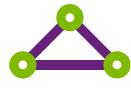
Azure Databricks



Bot service



Cognitive Services



Azure ExpressRoute



Azure Active Directory



Azure network security groups



Azure key management service



Operations Management Suite



Azure Functions



Visual Studio

Power BI



Transform data to insights



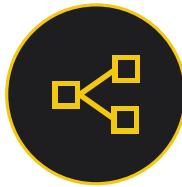
Get your data



Analyze



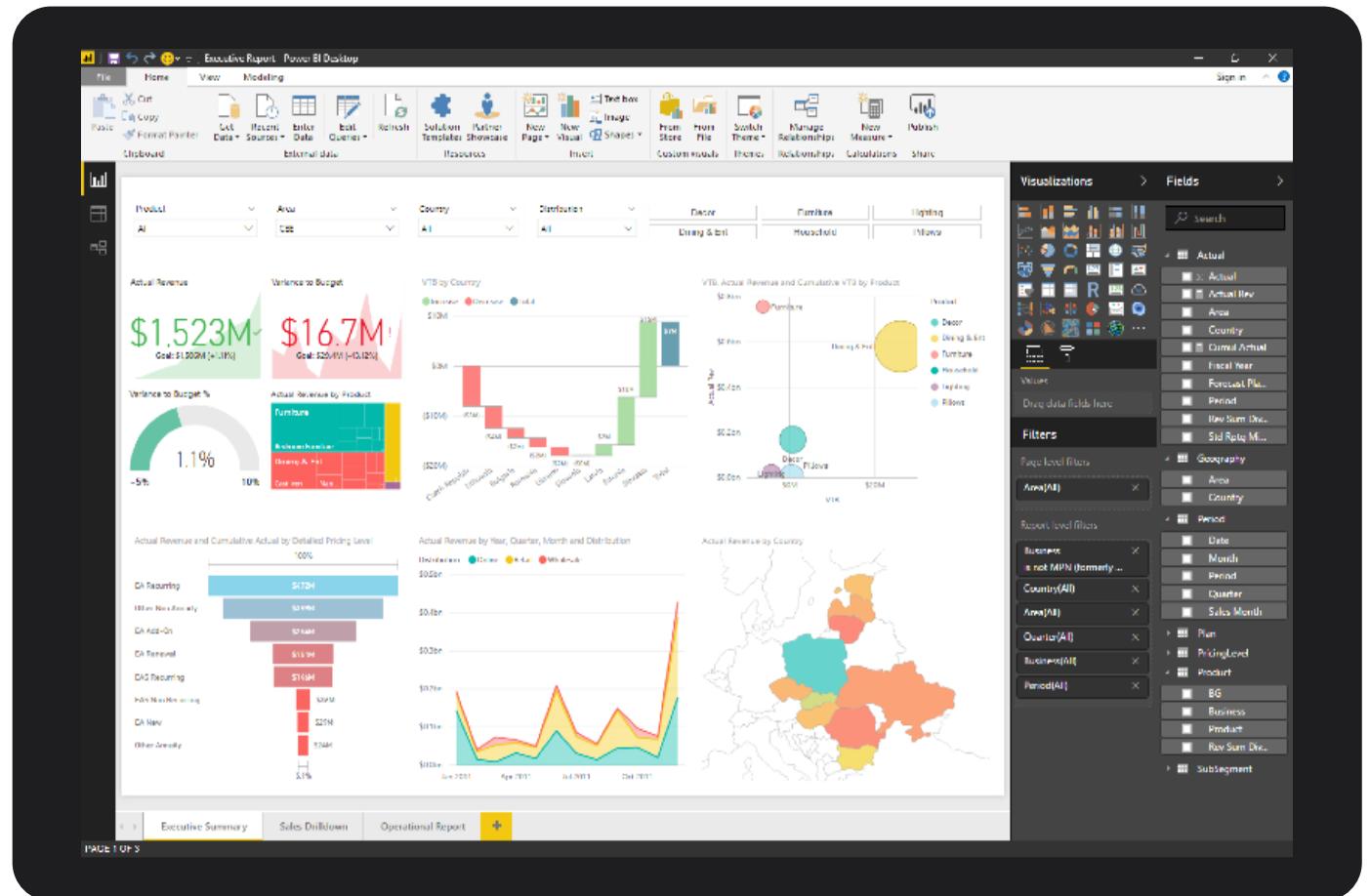
Visualize



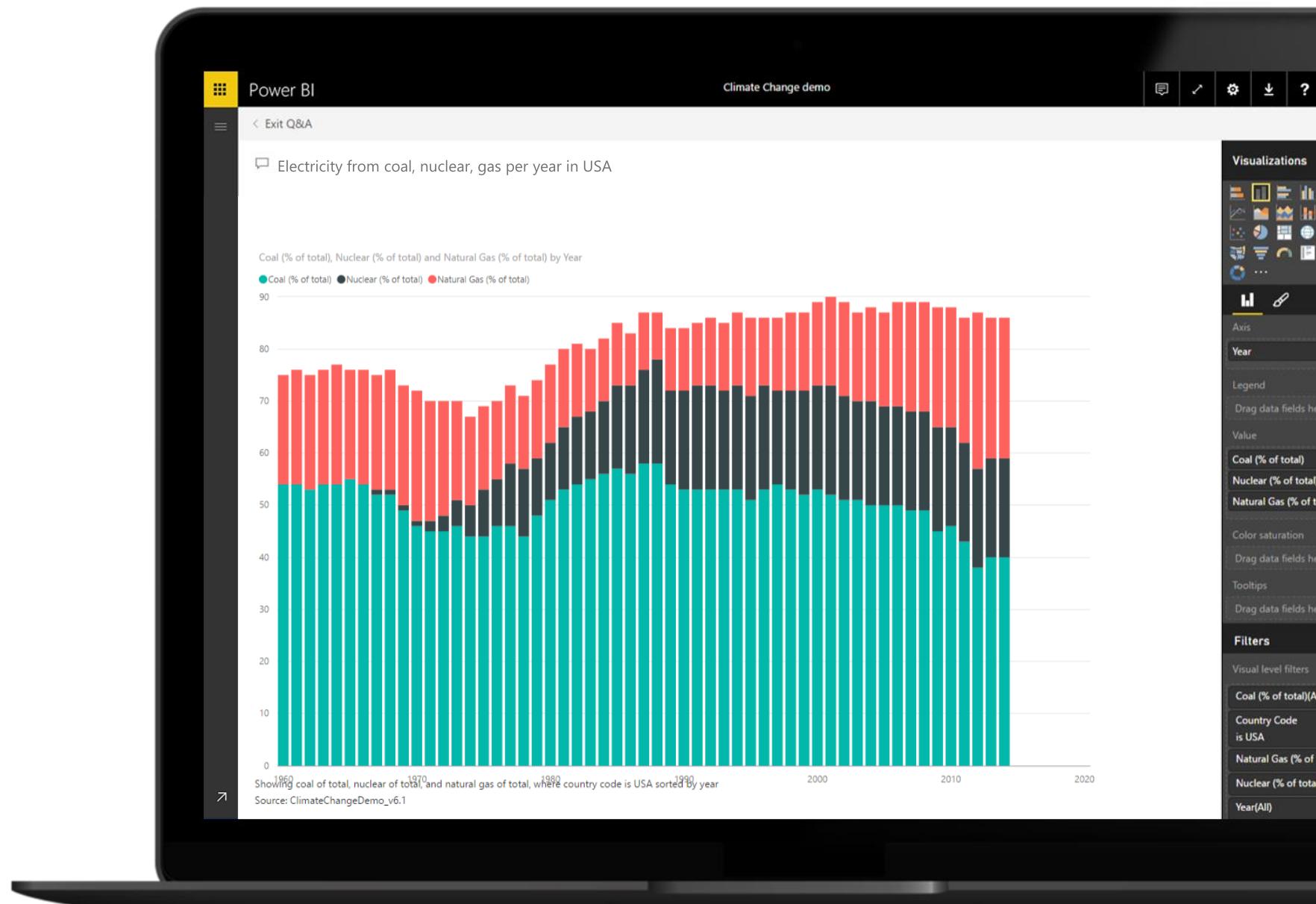
Publish



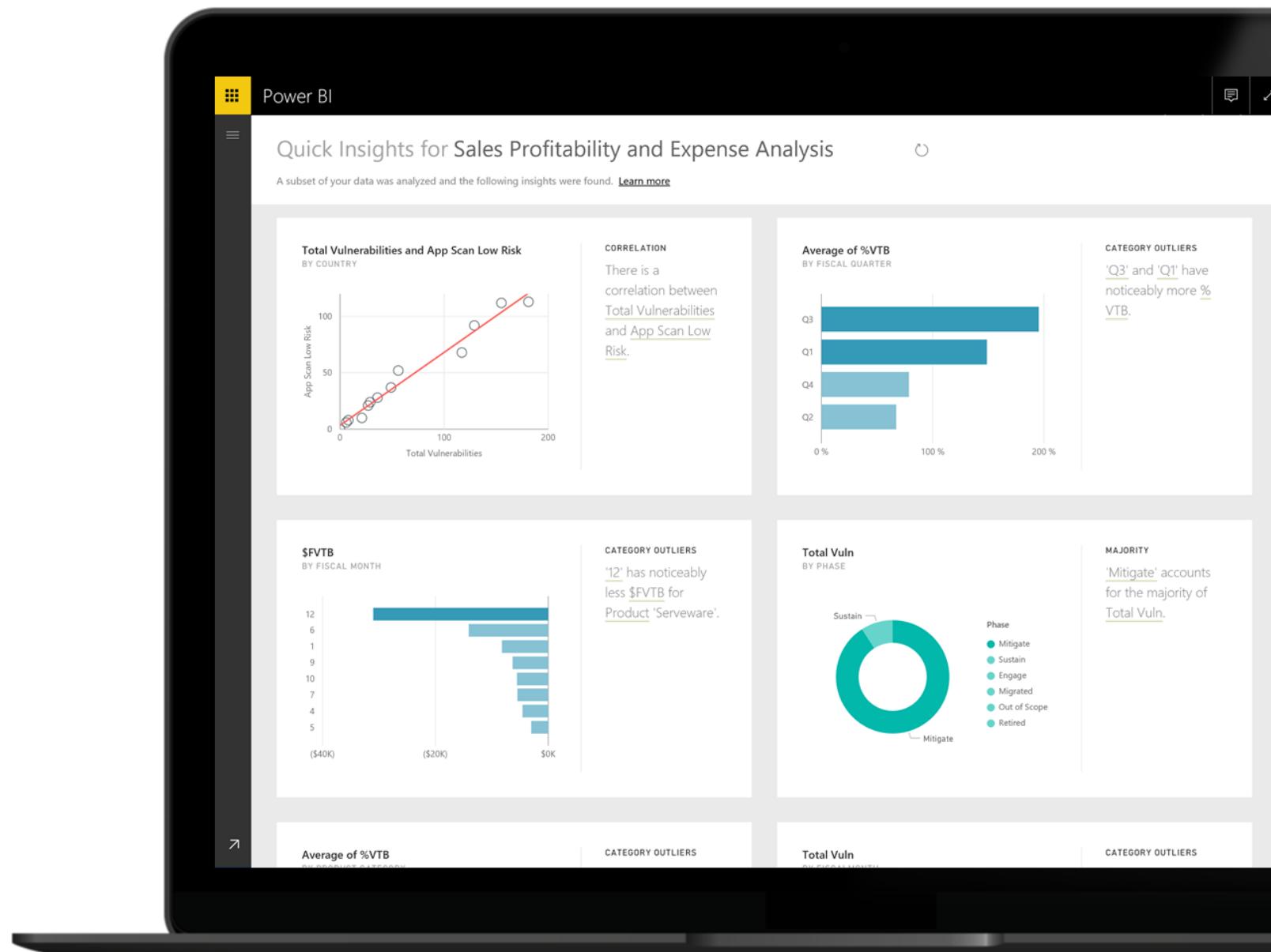
Collaborate



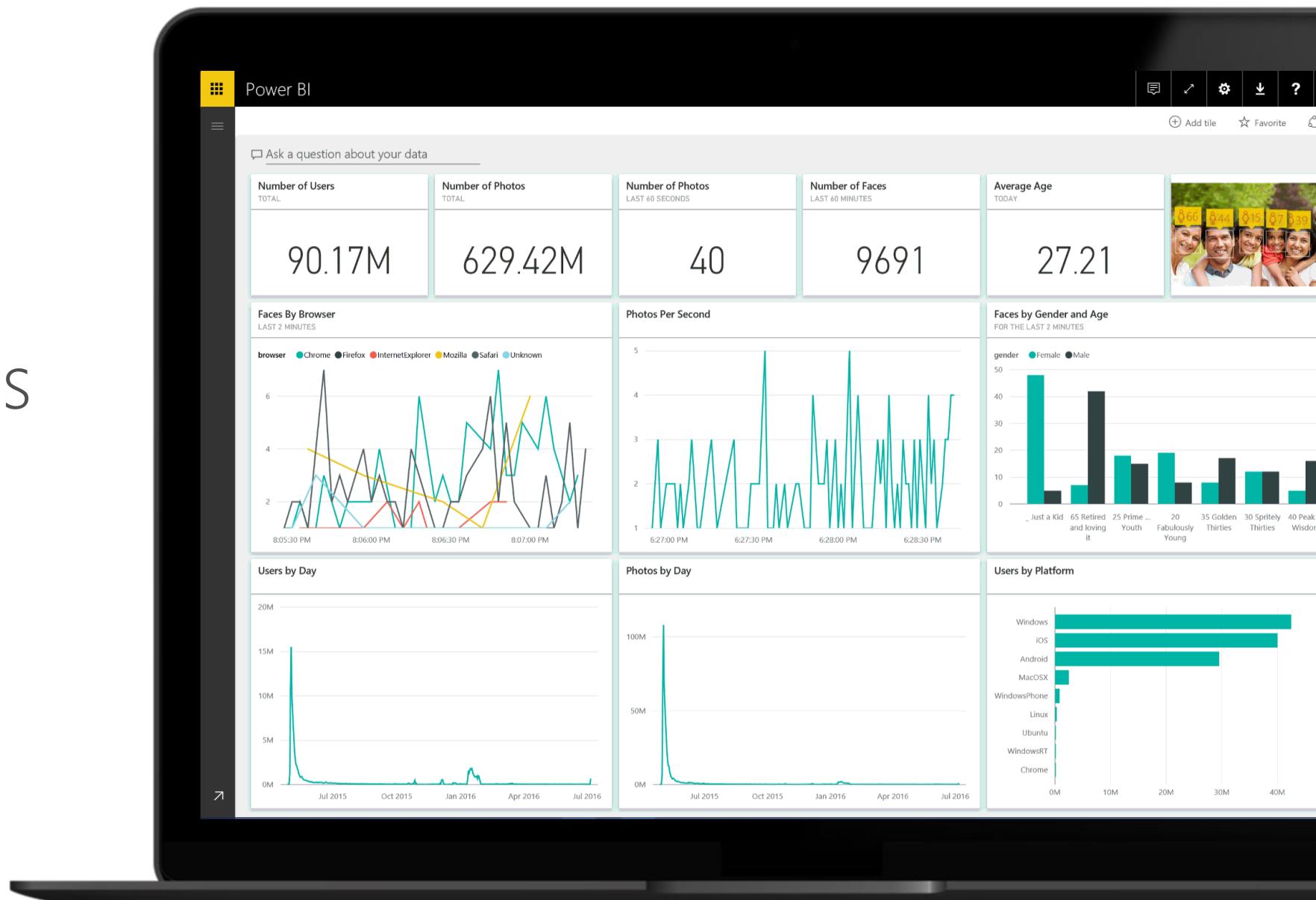
Ask questions of your data



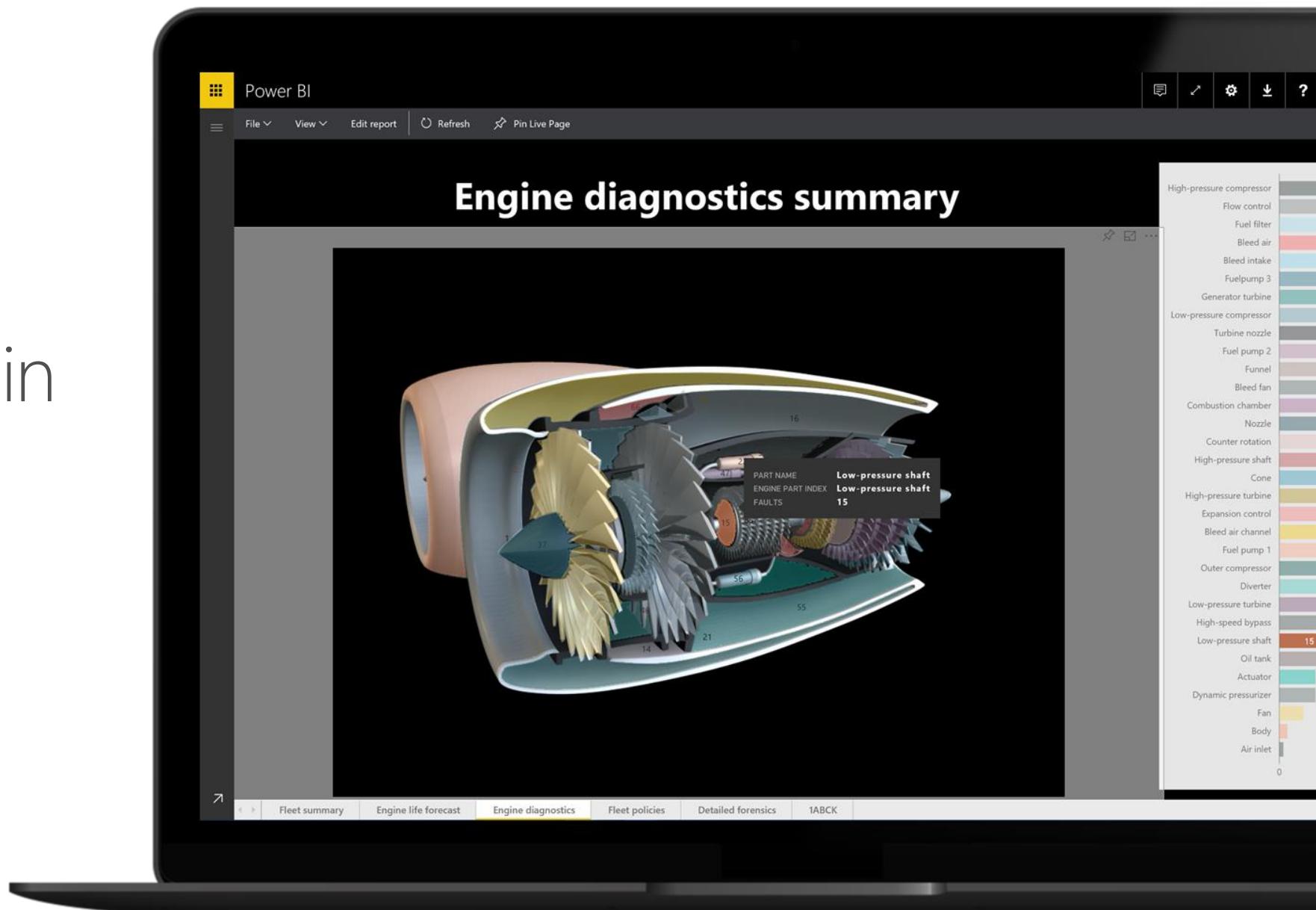
Automatically discover insights from your data

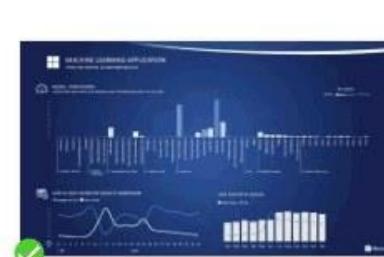
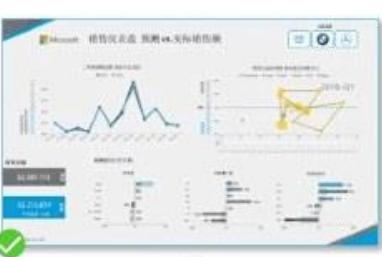
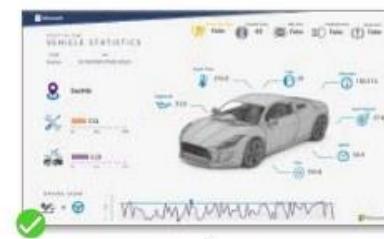


Run your business in real time with live dashboards



Visualize insights in
the context of
your business

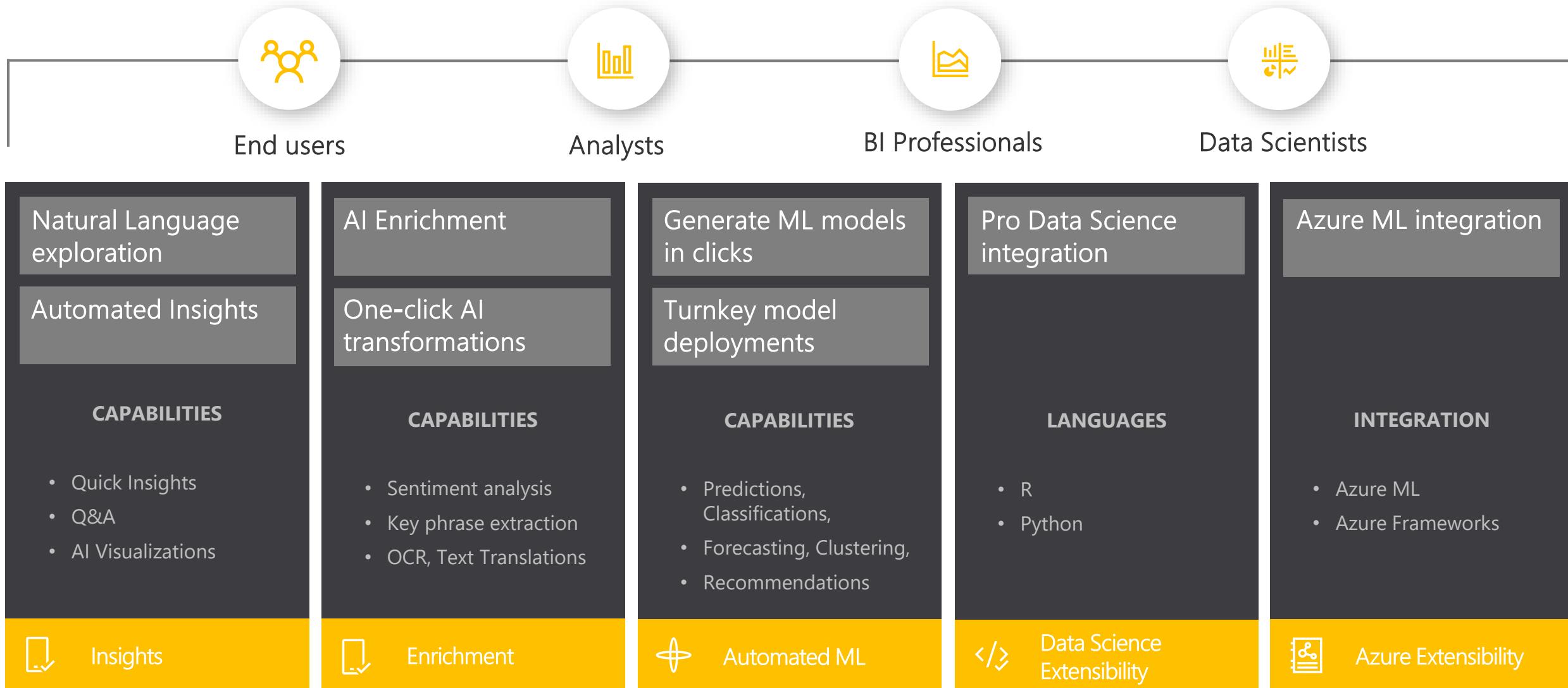




Power BI innovation areas



AI in Power BI



Creating AutoML models in Power BI

- The data science behind the creation of ML models is automated by Power BI
- Today we support ML models for Binary Predictions, Classifications and Regressions. Timeseries forecasting will also be available shortly.

New to machine learning models? Here's what you'll be doing:

1. Create and train your model

2. Improve it

3. Apply it

Select training data
Select your base data and related inputs to train your model.

Choose a model type
We'll help you pick the best model to achieve your business goals.

Train your model
The model will train on your data and report on its performance.

Iterate and retrain
Evaluate, customize and retrain your model until it's optimized

Apply the model
Apply your model to future data for predictive insights.

Purchase Intent Prediction model accuracy preview

This report summarizes the accuracy of the prediction model and enables you to find an optimal threshold for defining your business outcome.

How the model was evaluated

The model predicted revenue predictions for a total set of 1643 records and compared the predicted outcomes (based on the selected threshold) to the historical outcomes.

Final model type: Pre-fitted Soft Voting Classifier

5503	Input row count
3038	Sampled row count
2465	Training row count
1578	Validation row count

Different features have a different influence on the predicted value. Click below for details.

See key influences

Outcome distribution

68% of records predicted as Purchase are likely to actually be Purchase.

70% of records that are actually Purchase are likely to be predicted as Purchase.

Precision: 68%

Recall: 70%

Probability threshold: 0.75

Precision and recall

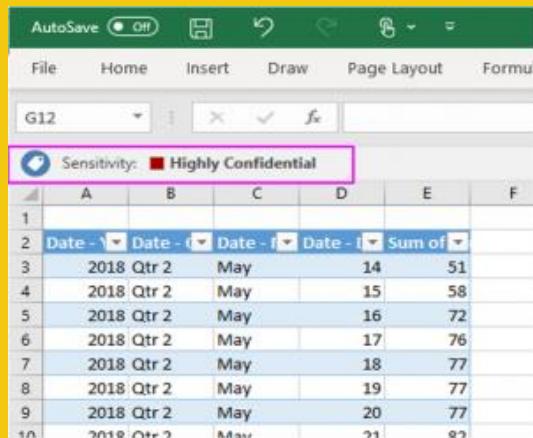
If applied to 10,000 records with the same distribution of Purchase and NonPurchase events, respectively, False Positives are observations where a NonPurchase was predicted as a Purchase. Conversely, False Negatives are observations where a Purchase was predicted as a NonPurchase. Note, the probability threshold used to classify a prediction as a Purchase or a NonPurchase affects these values.

Decide how to apply this model to your scenario

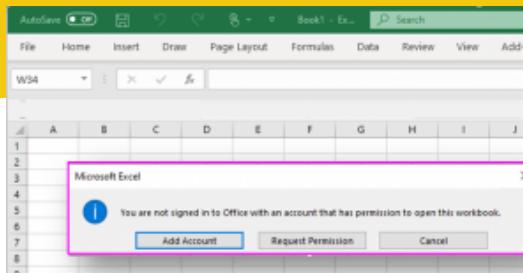
The model looks every record with a probability of Purchase ranging from 0 to 1. You can decide the threshold above which an observation is predicted as a Purchase. Increasing the threshold will result in fewer false positives (higher precision). Decreasing it will result in fewer false negatives (higher recall).

Power BI Data Protection

- Based on the same MS Information Protection built into Office
- The first BI product to protect your data even after it is exported out



With that policy in place, any Power BI data exported to Excel, PowerPoint or PDF file will carry that sensitivity label with it, and only authorized, authenticated users will be able to access the exported data



A screenshot of the Power BI web interface. The URL is https://dxt.powerbi.com/groups/80e26d0a-e85e-499b-ad17-897e0b9052c4/list/reports. The page title is 'Contoso Finance Department'. At the top, there are buttons for '+ Create', 'Settings', 'Access', 'List view', and 'Publish app'. Below this is a search bar with 'Search content...'. There are tabs for 'Dashboards', 'Reports', 'Workbooks', 'Datasets', and 'Dataflows', with 'Reports' selected. A table lists two reports: '2019 Q2 Revenue - Internal' and 'Sales 2019 - External'. Each report has a star icon, a preview thumbnail, and a row of icons for actions like edit, delete, and copy. The table includes columns for 'NAME', 'ACTIONS', 'OWNER', 'SENSITIVITY', and 'INCLUDED IN APP'. The 'SENSITIVITY' column shows 'General' with a gear icon, and the 'INCLUDED IN APP' column shows a toggle switch that is turned on.

Next Step

-

Q&A



