Microsoft Azure Solutions Business

L'IA/ML dans les Services de Données

Agenda (2h00)

- Introduction générale (15’)
- Introduction Azure Machine Learning (15)
- AI dans Azure Synapse (15)
- Azure HDInsight (15)
- AI dans Azure SQL Database & Azure SQL Managed Instance (15)
- AI dans Azure SQL @Edge (15)
- Azure Stream Analytics (15)
- Azure Data Factory (15)

Power BI (15)


Franck Gaillard
Cloud Solution Architect
Data AI
frgail@microsoft.com


Narjes Majdoub
Cloud Solution Architect
Data AI
nmajdoub@microsoft.com


Ali Bouhaddou
Cloud Solution Architect
Data Analytics
albouhad@microsoft.com


Frederic Gisbert
Cloud Solution Architect
Data Analytics
frgisber@microsoft.com

Microsoft

GPS Data/AI Strategy FY23

Delivered by CSA Team

22/09/2022

Franck Gaillard
Cloud Solution Architect
Data AI
frgail@microsoft.com

Narjes Majdoub
Cloud Solution Architect
Data AI
nmajdoub@microsoft.com

Ali Bouhaddou
Cloud Solution Architect
Data Analytics
albouhad@microsoft.com

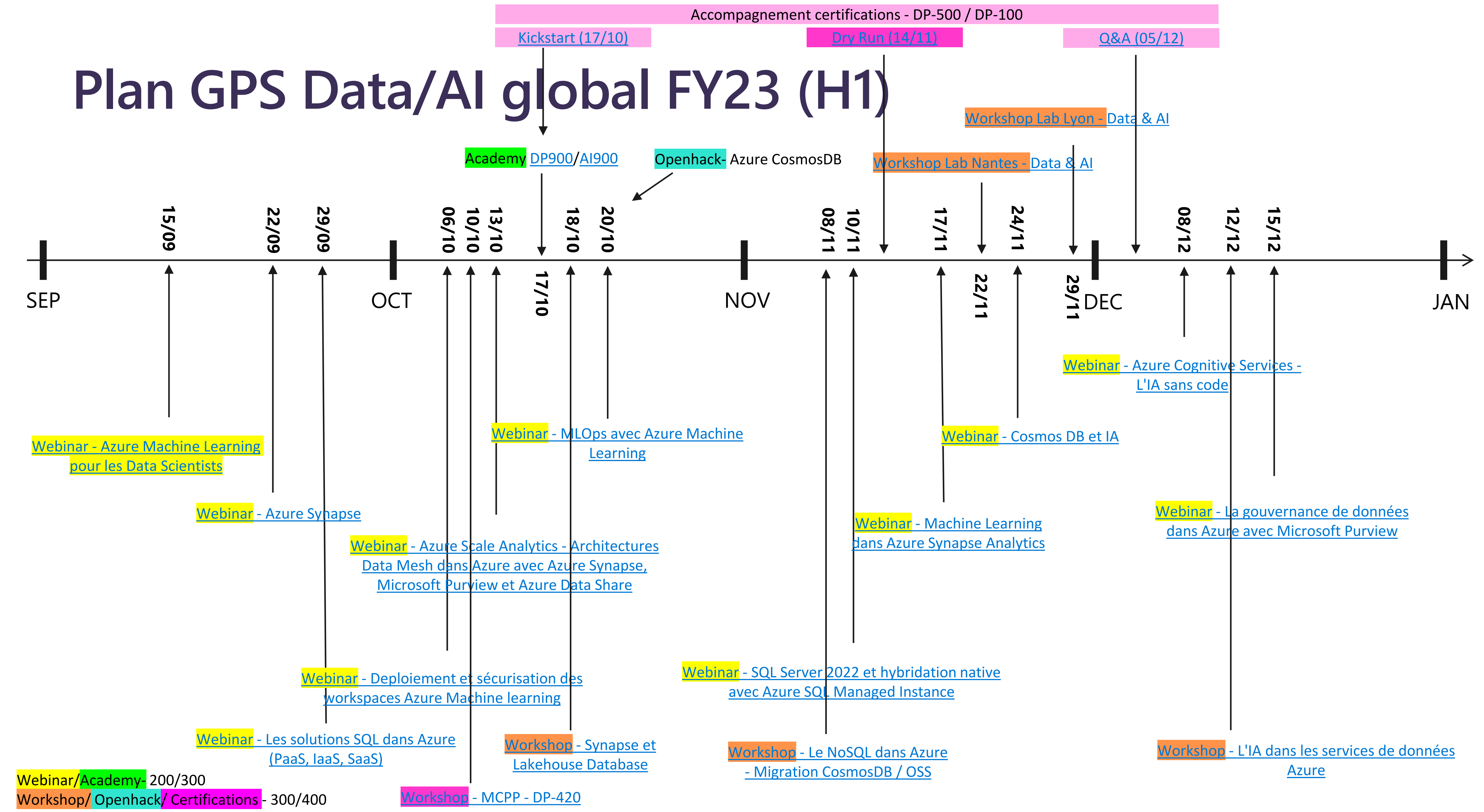
Frederic Gisbert
Cloud Solution Architect
Data Analytics
frgisber@microsoft.com

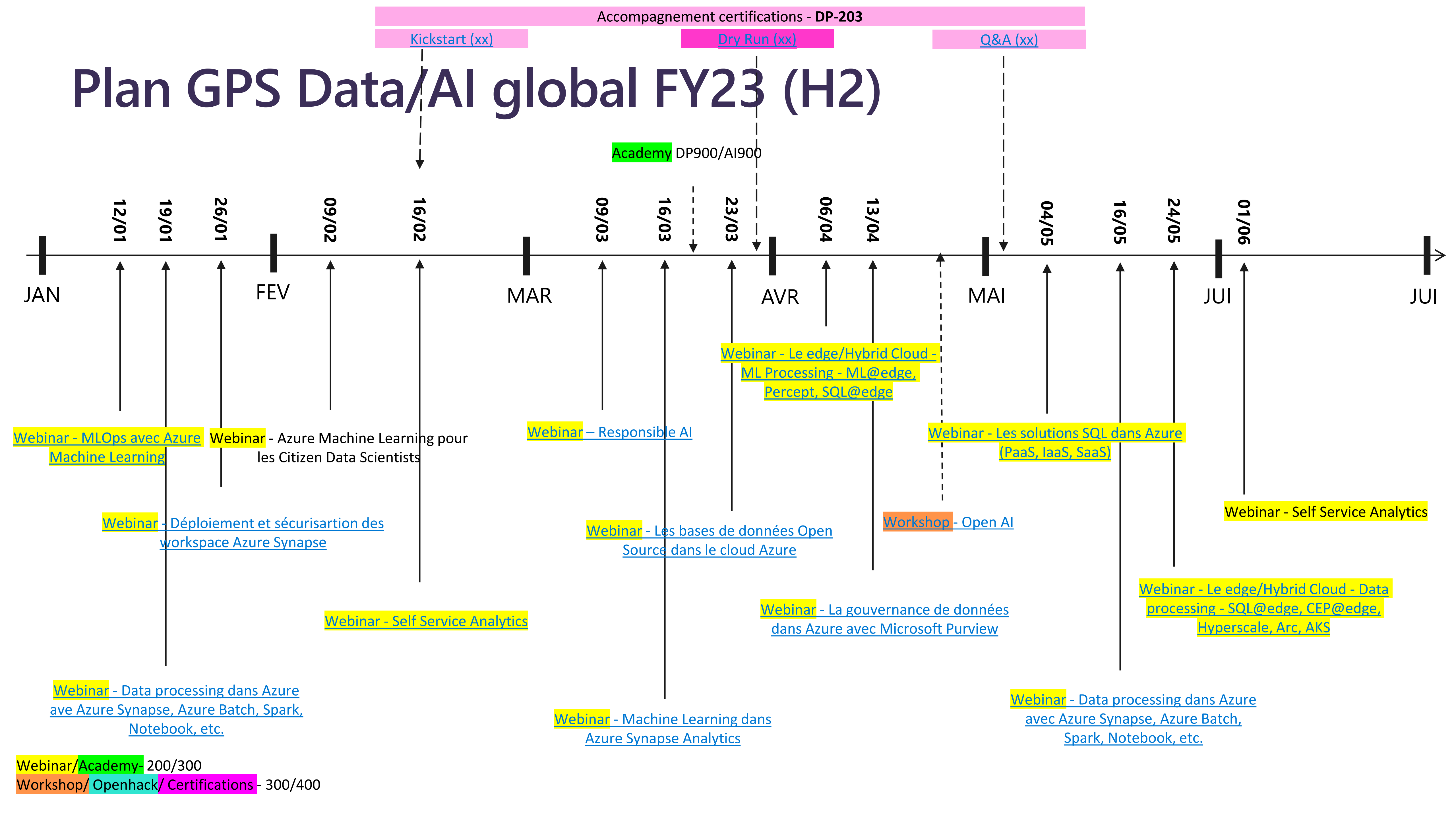
Azure Data & AI technical intensity plan

- **From June 2022 to June 2023**
- **Focus on "Azure Data & AI" tech intensity**

- **Many content, from L100 Beginner to L400 Expert level**
  - Academy L100
  - Webinar L200/L300
  - Workshop L300/L400
  - Certification kickstart L300/L400
  - Openhack / Microhack L400

Plan GPS Data/AI global FY23 (H1)




Accompagnement certifications - DP-203
Kickstart (xx)
Dry Run (xx)
Q&A (xx)
Plan GPS Data/AI global FY23 (H2)
Academy DP900/AI900
JAN
12/01
19/01
26/01
FEV
09/02
16/02
MAR
09/03
16/03
23/03
AVR
06/04
13/04
MAI
04/05
16/05
24/05
JUI
01/06
JUI
Webinar - MLOps avec Azure Machine Learning
Webinar - Azure Machine Learning pour les Citizen Data Scientists
Webinar - Déploiement et sécurisartion des workspace Azure Synapse
Webinar - Self Service Analytics
Webinar - Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.
Webinar – Responsible AI
Webinar - Les bases de données Open Source dans le cloud Azure
Webinar - Machine Learning dans Azure Synapse Analytics
Webinar - Le edge/Hybrid Cloud - ML Processing - ML@edge, Percept, SQL@edge
Webinar - La gouvernance de données dans Azure avec Microsoft Purview
Workshop - Open AI
Webinar - Les solutions SQL dans Azure (PaaS, IaaS, SaaS)
Webinar - Data processing dans Azure avec Azure Synapse, Azure Batch, Spark, Notebook, etc.
Webinar - Le edge/Hybrid Cloud - Data processing - SQL@edge, CEP@edge, Hyperscale, Arc, AKS
Webinar - Self Service Analytics
Webinar/Academy- 200/300
Workshop/ Openhack/ Certifications - 300/400

Liste des évènements de type Webinar 2H

Event Webinar (Les jeudis de la Data & AI) - L200/300	Date	Duration (min)	Link
Azure Machine Learning pour les Data Scientists	15/09/2022	120	https://msevents.microsoft.com/event?id=2454281594
Azure Synapse	22/09/2022	120	https://msevents.microsoft.com/event?id=857781749
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	29/09/2022	120	https://msevents.microsoft.com/event?id=502366997
Déploiement et sécurisation des workspaces Azure Machine learning	06/10/2022	120	https://msevents.microsoft.com/event?id=1505714138
Azure Scale Analytics - Architectures Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share	13/10/2022	120	https://msevents.microsoft.com/event?id=139685175
MLOps avec Azure Machine Learning	20/10/2022	120	https://msevents.microsoft.com/event?id=1245885767
SQL Server 2022 et hybridation native avec Azure SQL Managed Instance	10/11/2022	120	https://msevents.microsoft.com/event?id=145826476
Machine Learning dans Azure Synapse Analytics	17/11/2022	120	https://msevents.microsoft.com/event?id=3637723312
Azure Cosmos DB et IA	24/11/2022	120	https://msevents.microsoft.com/event?id=2646013445
Azure et les Services Cognitifs	08/12/2022	120	https://msevents.microsoft.com/event?id=3772037220
La gouvernance de données dans Azure avec Microsoft Purview	15/12/2022	120	https://msevents.microsoft.com/event?id=1499560981
MLOps avec Azure Machine Learning	12/01/2023	120	https://msevents.microsoft.com/event?id=4115194515
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	19/01/2023	120	https://msevents.microsoft.com/event?id=1537241181
Déploiement et sécurisation des workspace Azure Synapse	26/01/2023	120	https://msevents.microsoft.com/event?id=1806467748
Azure Machine Learning pour les Citizen Data Scientists	09/02/2023	120	En cours
PowerBI - Self Service Analytics	16/02/2023	120	https://msevents.microsoft.com/event?id=1401519679
L'IA responsable avec Azure machine learning	09/03/2023	120	https://msevents.microsoft.com/event?id=2072953112
Machine Learning dans Azure Synapse Analytics	16/03/2023	120	https://msevents.microsoft.com/event?id=3413014857
Les bases de données Open Source dans le cloud Azure	23/03/2023	120	https://msevents.microsoft.com/event?id=2727487131
Hybridation des services de Machine Learning Azure	06/04/2023	120	https://msevents.microsoft.com/event?id=1624914222
La gouvernance de données dans Azure avec Microsoft Purview	13/04/2023	120	https://msevents.microsoft.com/event?id=3909342839
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	04/05/2023	120	https://msevents.microsoft.com/event?id=1162207895
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	16/05/2023	120	https://msevents.microsoft.com/event?id=3517068442
Hybridation des services de données Azure	24/05/2023	120	https://msevents.microsoft.com/event?id=2996507398
Self Service Analytics	01/06/2023	120	En cours
Total	**25**		

Liste des évènements de type Workshop/Prepa Cert/Academy

Event Workshop L300/400	Date	Duration (min)	Link
Synapse et Lakehouse Database	18/10/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Le NoSQL dans Azure - Migration CosmosDB / OSS	08/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Lab Lyon - Data & AI	22/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUMlZZOURET0RSWjcyTERYRkJGTlFFUjJaUi4u
Lab Nantes - Data & AI	29/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUMlZZOURET0RSWjcyTERYRkJGTlFFUjJaUi4u
L'IA dans les services de données Azure	12/12/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Open AI	H2	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u

Event Academy, kickstart certifications, workshop certifications	Date	Duration (min)	Link
MCPP - DP-420	10/10/2022	420	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUMkJISlRKSU1RRFA0OVgzSFdTSTY0RE9WQy4u
Micro Hack CosmosDB	20/10/2022	420	H1 - Inscriptions PTA
Academy DP900	17-21/10/2022	300	https://msevents.microsoft.com/event?id=3250818161
Academy AI900	17-21/10/2022	300	https://msevents.microsoft.com/event?id=2717528090
Kickstart DP-500	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3TE9ZRS4u
Dry Run DP-500	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3TE9ZRS4u
Q&A DP-500	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3TE9ZRS4u
Kickstart DP-100	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMUFYxSkRIMi4u
Dry Run DP-100	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMUFYxSkRIMi4u
Q&A DP-100	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMUFYxSkRIMi4u
Kickstart DP-203	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u
Dry Run DP-203	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u
Q&A DP-203	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u
Total	**16**		


Engage customers
Empower employees
Intelligent
Data & AI
Systems & Experiences
Optimize operations
Transform products
Digital Feedback Loop

They embrace digital transformation

They rely on a modern data estate

The modern data estate

Unparalleled differentiation

- from ground to cloud

Scale
Limitless

AI
Intelligent by default

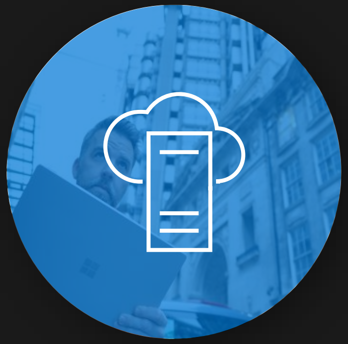
Hybrid
Operational freedom

Security
Always a step ahead

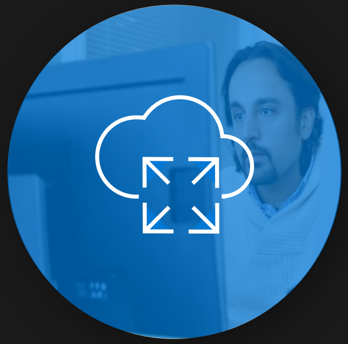
Azure Data

- Powering the Modern Data Estate

Data modernization on-premises



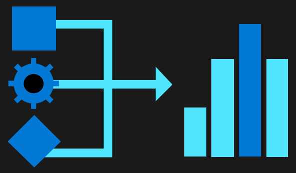
Data modernization to Azure



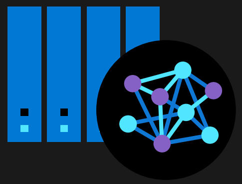
Cloud scale analytics

What if...

Simplify analytics



Use existing skills for all analytics



Apply AI where data lives

Microsoft Azure Solutions Business

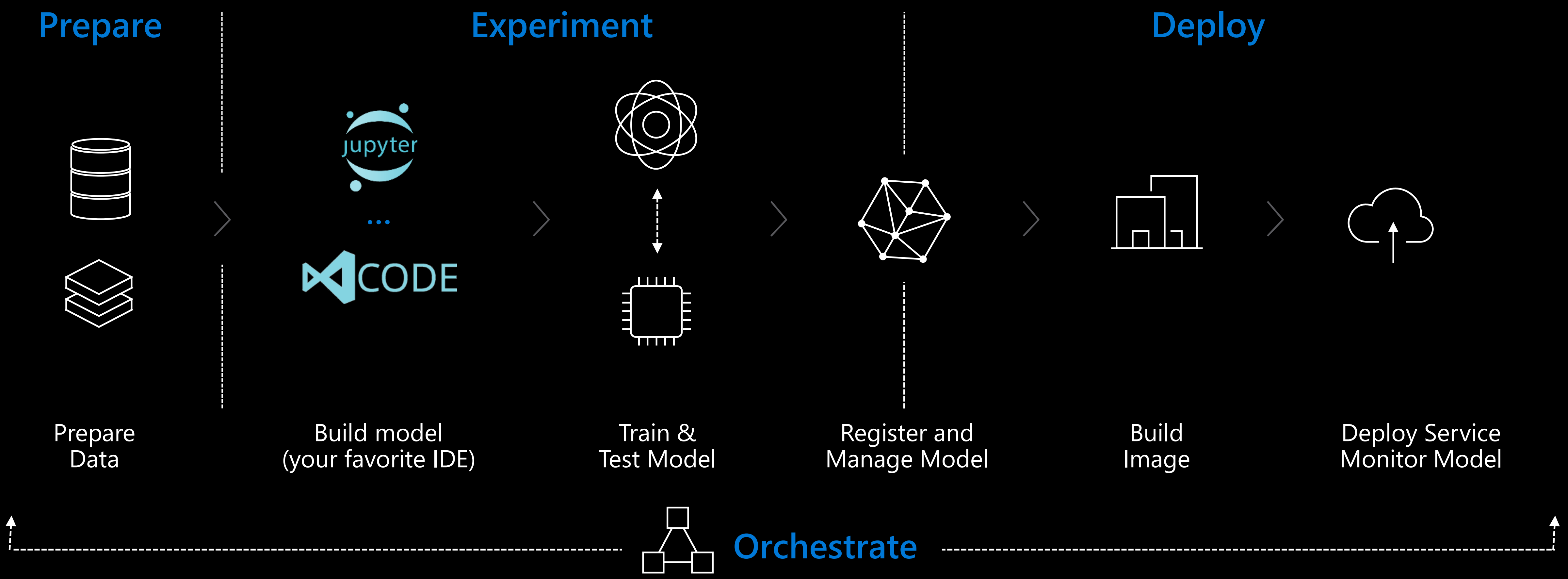
Introduction à Azure Machine Learning

Narjes Majdoub

Requirements of an advanced ML Platform

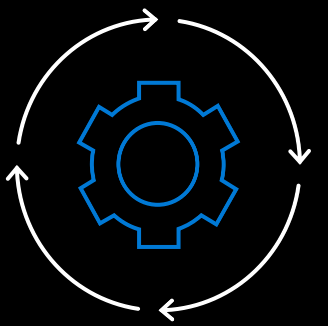
Machine Learning

Typical E2E Process



Key Trends

Data Science and ML platforms



Automation

Automated workflows for deployment and management 1000s of models

Composite AI and transfer learning techniques with models like GPT3

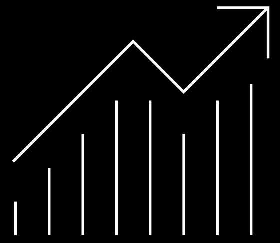
Configurable and repeatable recipes like NLP, Recommenders, many-models

Collaboration

Collaborative tools and processes as multiple roles contributing to ML practice

Robust Responsible AI approach to ensure ethical use due to multiple stakeholders

Enterprise grade data and model governance/security



Acceleration

Rise of MLOps to accelerate operationalization of models

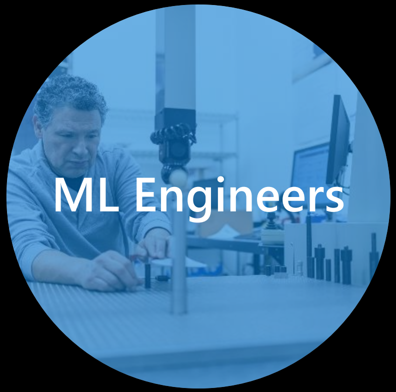
Advanced techniques like reinforcement learning, GANs and synthetic data

Azure offers a comprehensive AI/ML platform that meets—and exceeds—requirements

Intended Audience

BUILD MODELS

OPERATIONALIZE MODELS

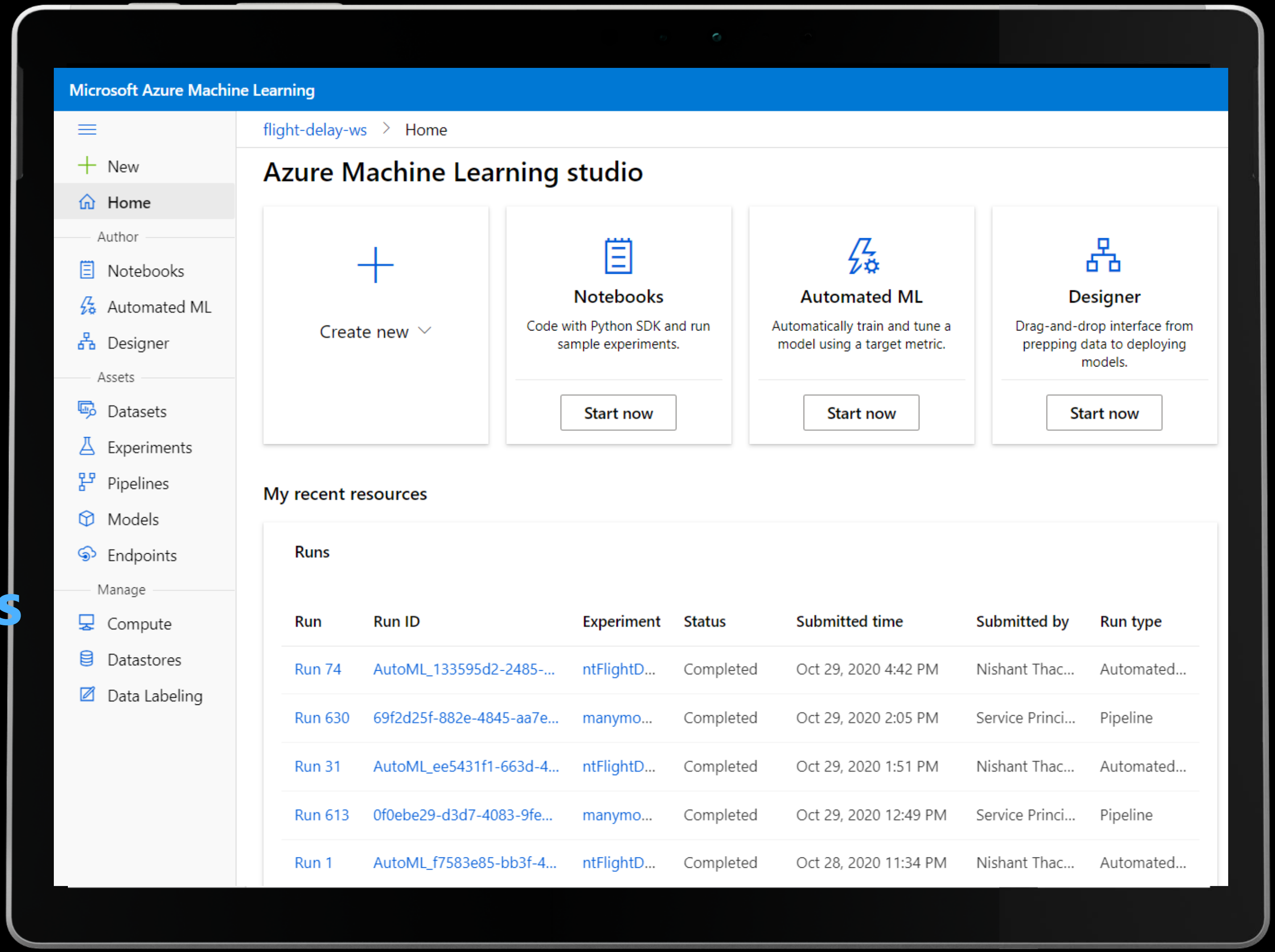


Azure Machine Learning

The one central hub for your data science team

Boosted collaboration

Integration with other Azure services



Azure Machine Learning Service

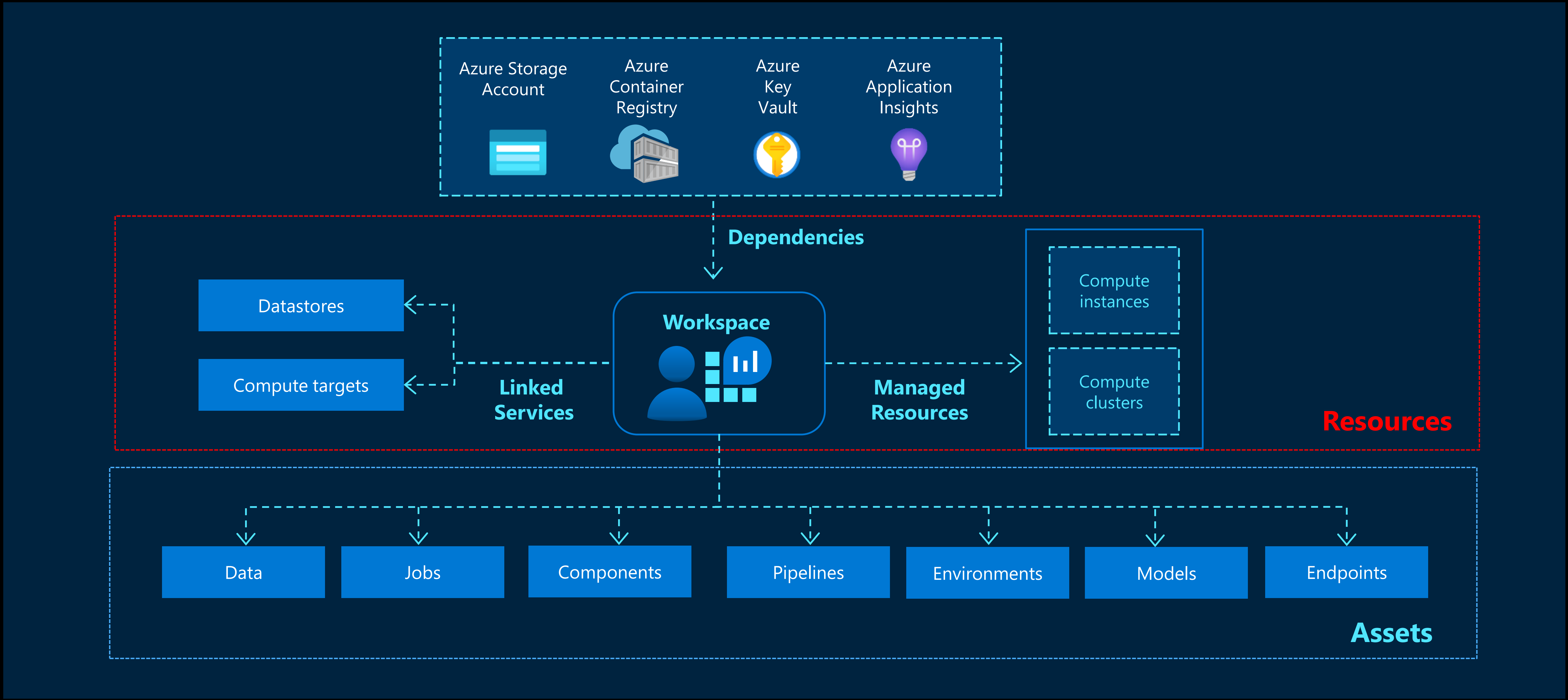
Make data scientists to be more productive
Enable your organization to manage the ML lifecycle through MLOps

Azure Cloud Services + Python SDK + Cross-Platform CLI

That enables you to:

- ✓ Prepare Data
- ✓ Build Models
- ✓ Train Models
- ✓ Manage Models
- ✓ Track Experiments
- ✓ Deploy Models

Key Elements of Azure Machine Learning



Azure Machine Learning accelerates model development

with automated machine learning

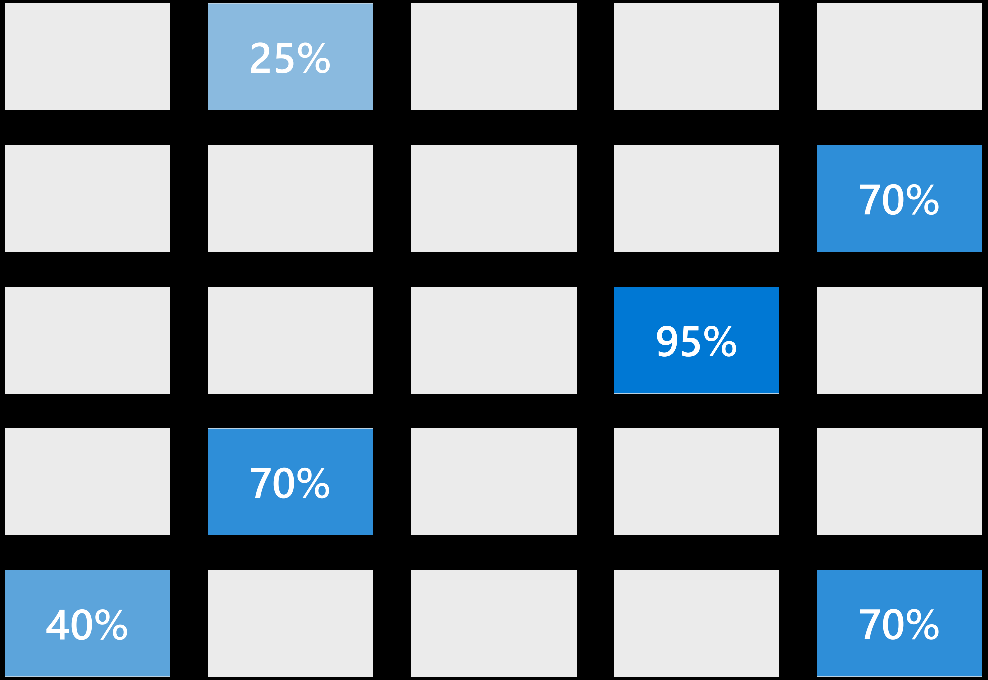
Input

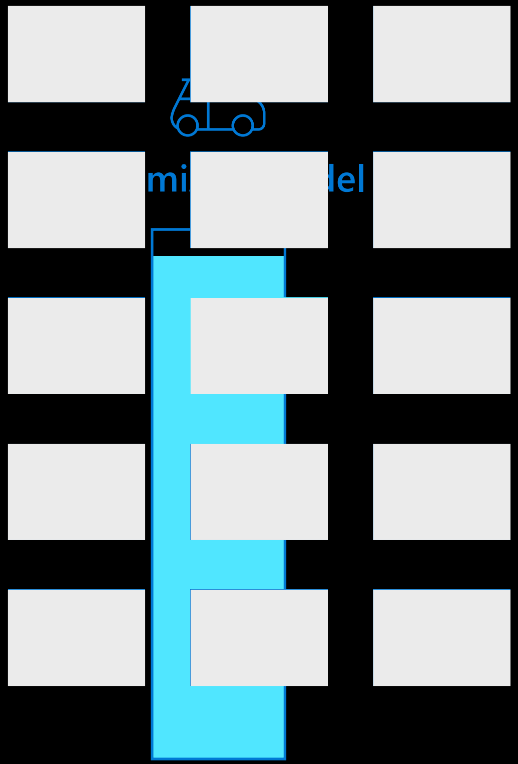
101010
010101
101010
Enter data

Define goals

Apply constraints

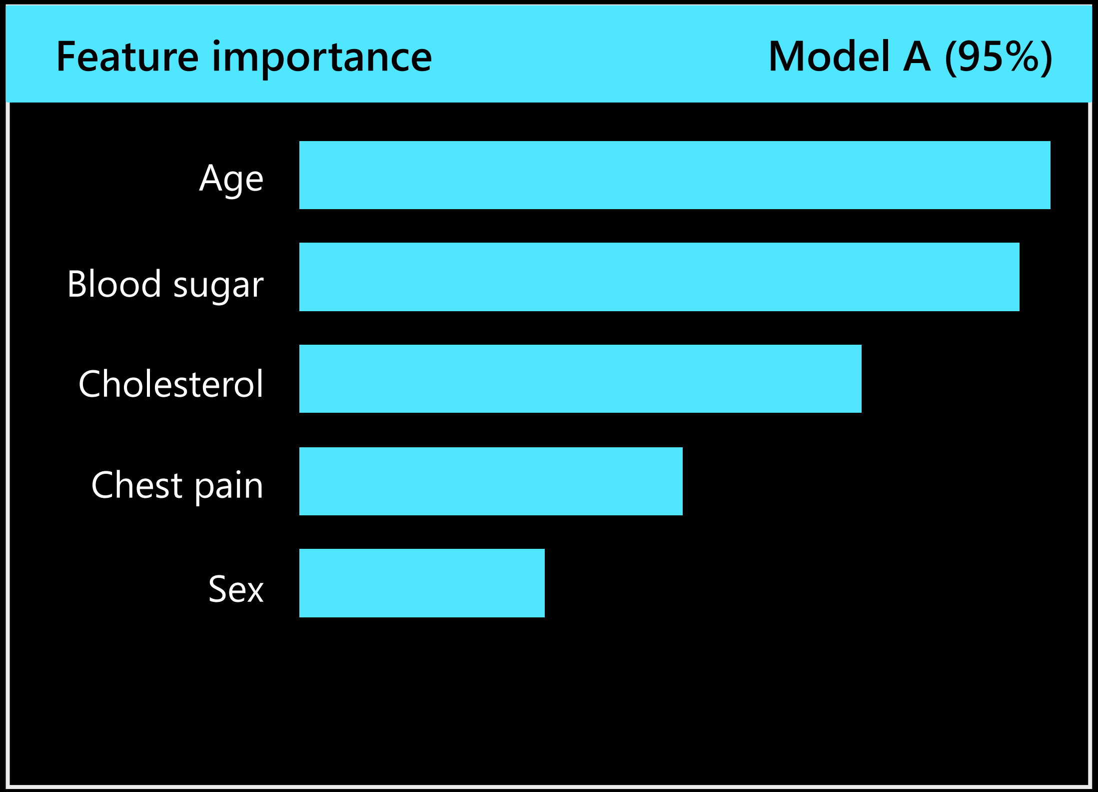
Intelligently test multiple models in parallel

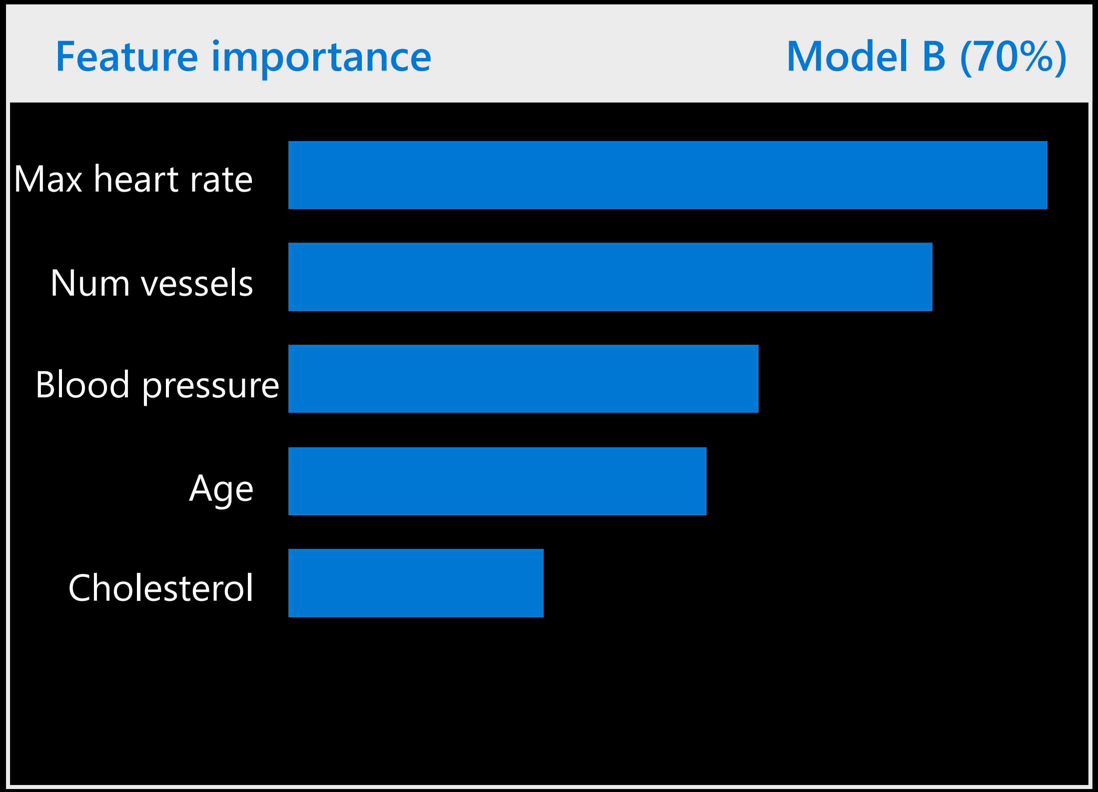




Azure Machine Learning accelerates model selection

with model explainability



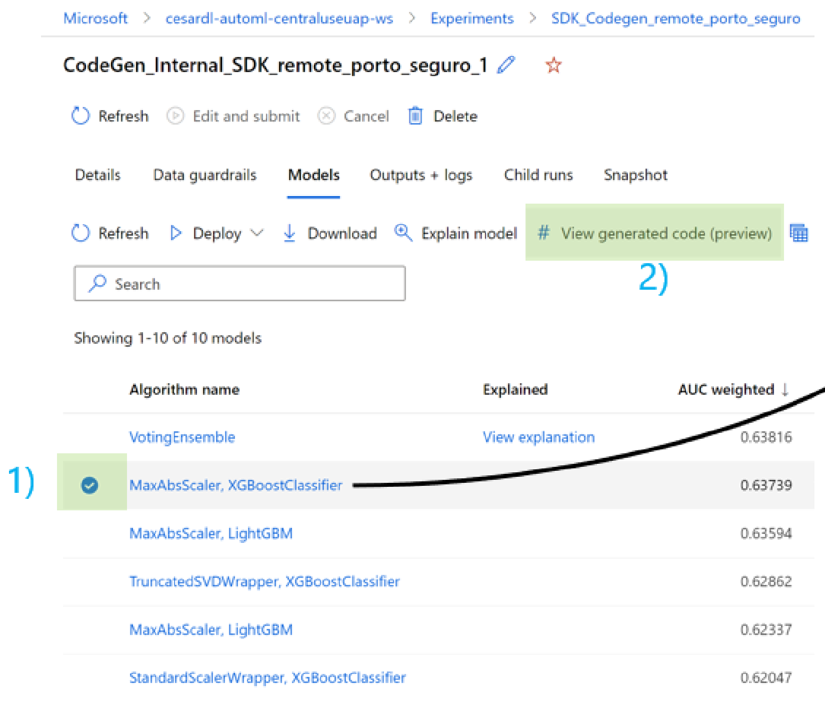


Azure Machine Learning accelerates model selection

with code generation [Public Preview]

AutoML Model's training code generation (i.e. Scikit-Learn code, Transformers and Algorithms)

AutoML UI
(Experiment's model leader board)



script.py with model's training generated code (Scikit-Learn, etc.)

```python
def generate_algorithm_config():
    from xgboost.sklearn import XGBClassifier

    algorithm = XGBClassifier(
        base_score=0.5,
        booster='gbtree',
        colsample_bylevel=1,
        colsample_bynode=1,
        colsample_bytree=1,
        gamma=0,
        learning_rate=0.1,
        max_delta_step=0,
        max_depth=3,
        min_child_weight=1,
        missing=None,
        n_estimators=100)
    return algorithm
#...
def build_model_pipeline():
    from sklearn.pipeline import Pipeline
    pipeline = Pipeline(
        steps=[
            ('featurization', generate_data_transformation_config()),
            ('preproc', generate_preprocessor_config()),
            ('model', generate_algorithm_config()),])
    return pipeline
#...

model_pipeline = build_model_pipeline()

model = model_pipeline.fit(X, y)
```

- Data preprocessing
- Algorithm selection
- Featurization
- Hyperparameters

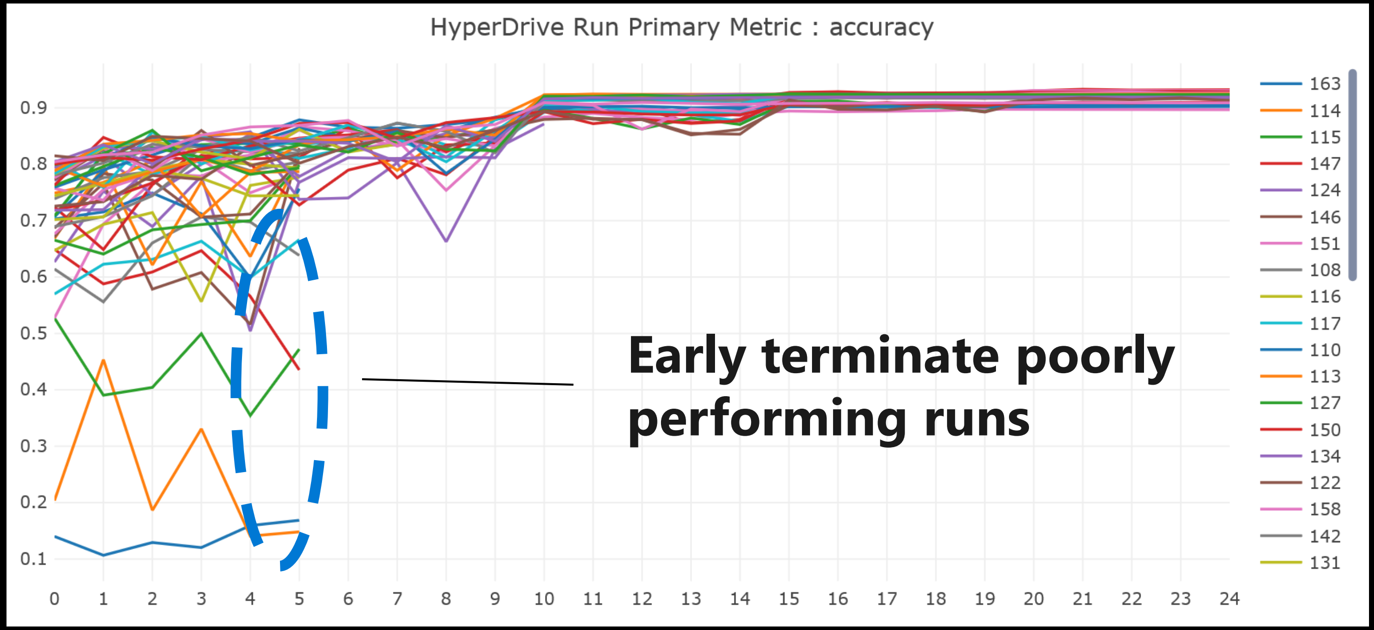
Automated Hyperparameter Tuning

Evaluate training runs for specified primary metric

Use resources to explore new configurations

Early terminate poor performing training runs. Early termination policies include:

- Bandit policy
- Median Stopping policy
- Truncation Selection policy

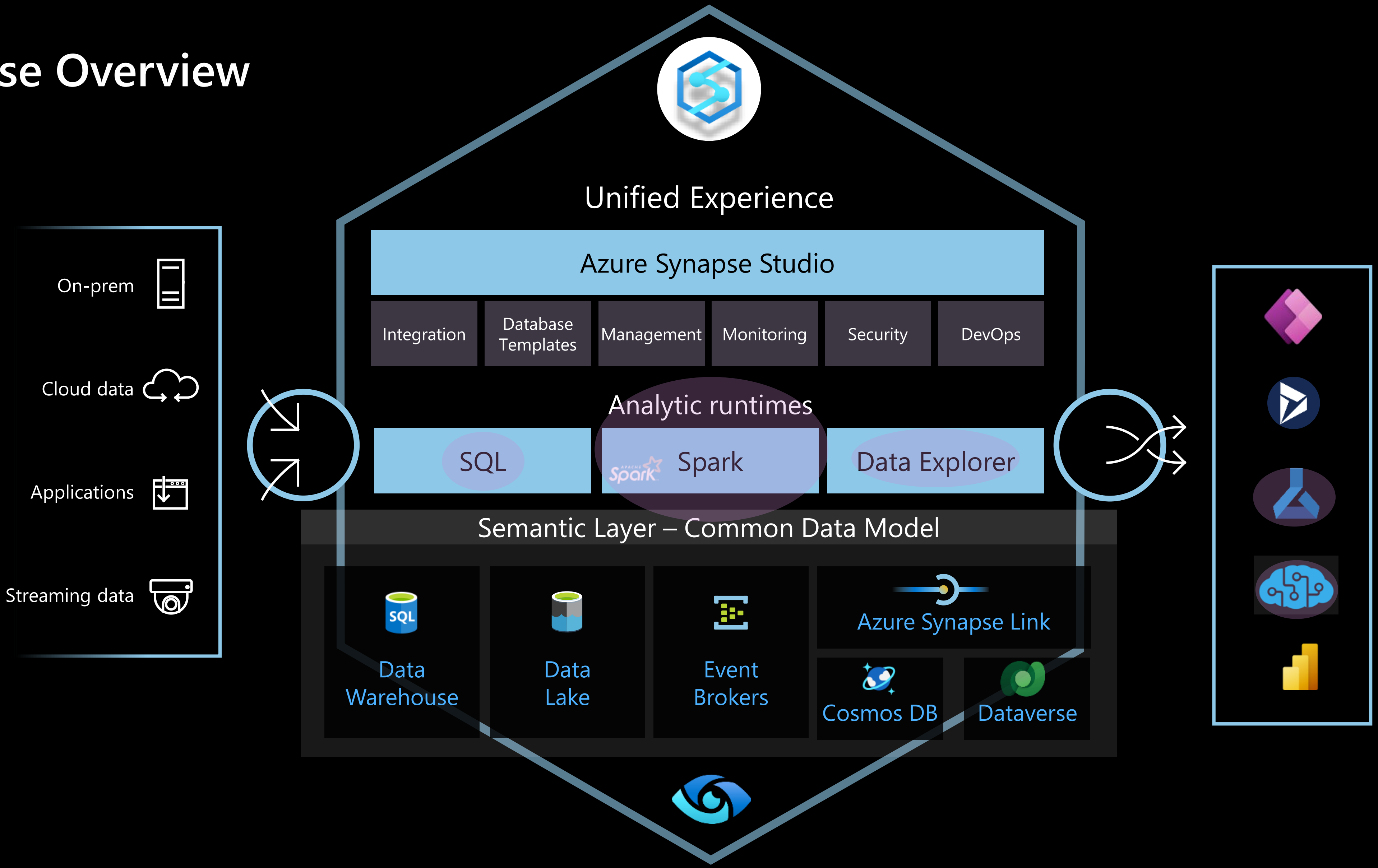


Microsoft Azure Solutions Business

AI dans Azure Synapse

Narjes Majdoub

Synapse Overview



Machine learning

Democratize predictive power

- Notebooks provide code authoring experiences
- Notebooks provides a code authoring experience for complex predictive models
- Automatic ML graphical interface provides a no-code experience for creating ML models
- Native integration with Azure Cognitive Search provides access to pre-built models

All Code
Notebook IDE
PySpark/Scala/.Net/C#, R

Low/no-Code
Classification
Regression
Time-series

Pre-built models
Anomaly Detector
Sentiment Analysis

Synapse & Azure ML: Supporting the full Data & AI lifecycle

Azure Synapse

Prepare and Transform data with code first and code free experiences

Model training on Spark with code first and code free experiences.

Track models and experiments in AzureML

Model scoring in Synapse SQL and Spark. Leverage AzureML for model endpoints and MLOps.

Orchestrate data prep, model training and other steps with Synapse and AzureML pipelines

Serve predictions to Power BI

Data discovery → **Data access** → **Data prep** → **Model** → **Deploy** → **Orchestration & Monitor** → **Business Insights**

[Data wrangling with Apache Spark pools (preview) - Azure Machine Learning | Microsoft Learn]

[MachineLearningNotebooks/how-to-use-azureml/azure-synapse at master · Azure/MachineLearningNotebooks (github.com)]

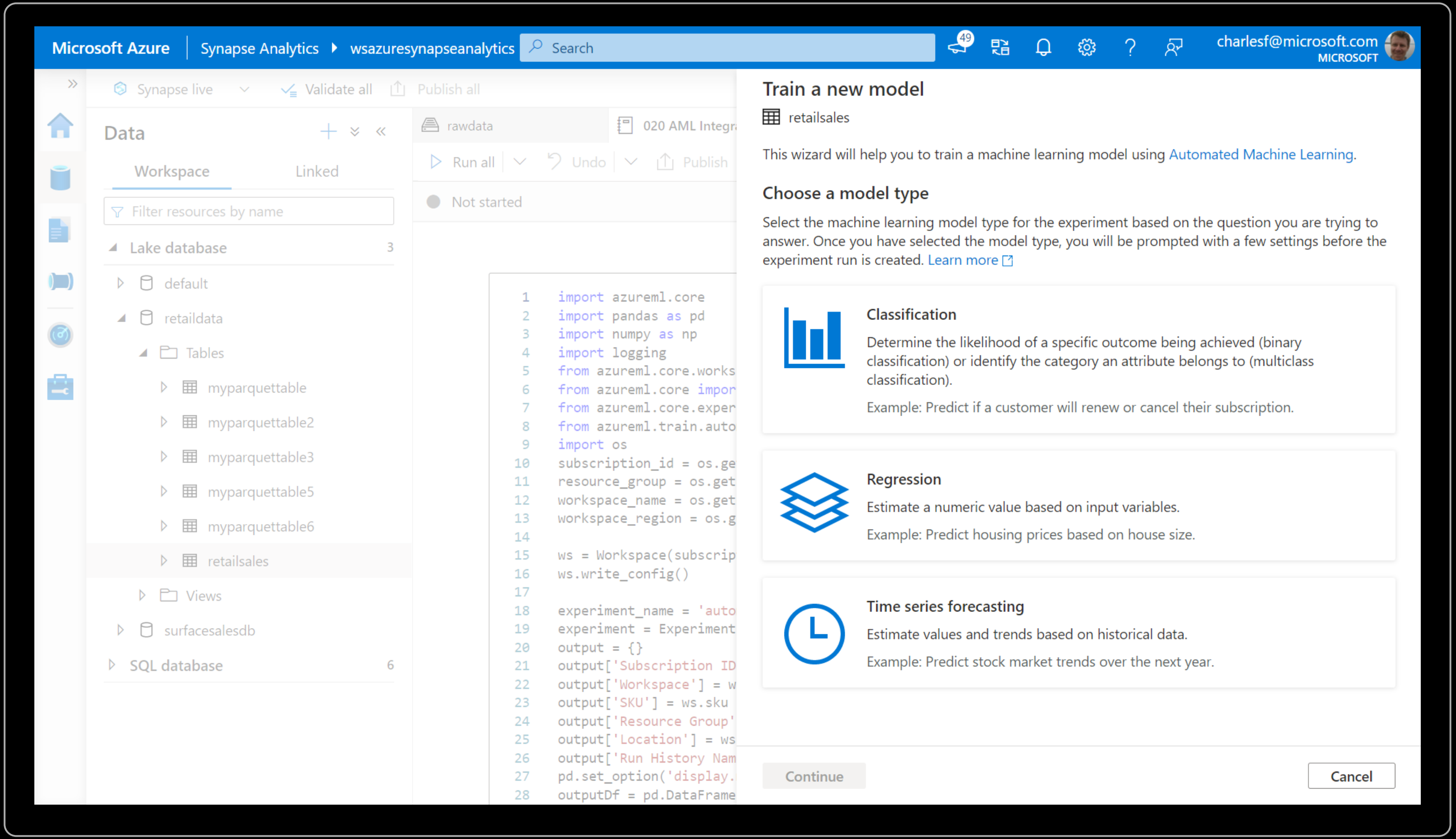
Synapse

Data Science

GA

Automated Machine Learning

Code first + No-code training for ML models empowers everyone with data science

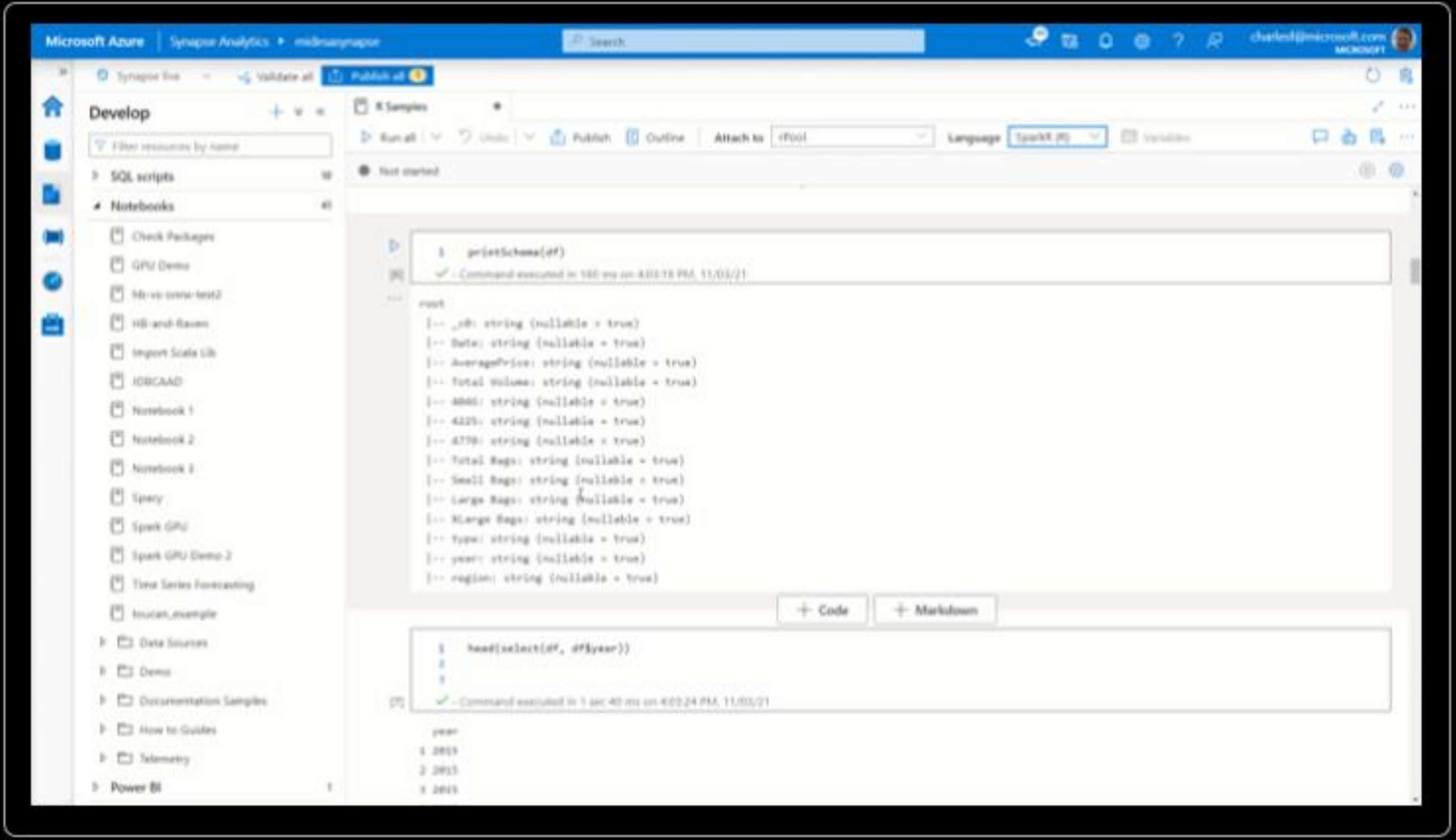


Private Preview

Q2 2022

R Language Support

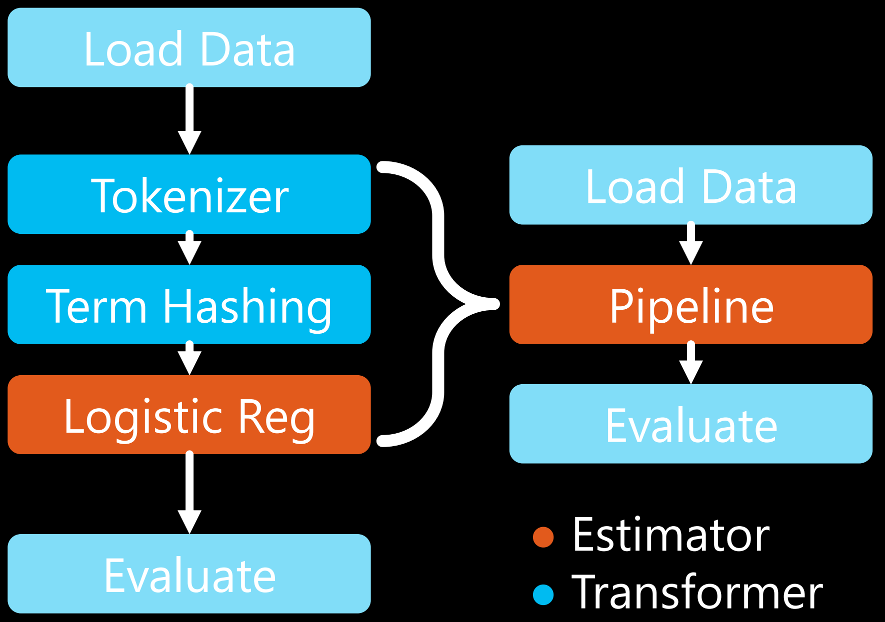
Enables data scientists to apply the industry standard R language to developing ML models



Spark ML

```
data = spark.read.csv("hdfs://…")
train, test = data.randomSplit([.5,.5])
model = LogisticRegression().fit(train)
predictions = model.transform(test)
```

- High level library for distributed machine learning
- Inspired by scikit-learn
- All models have a uniform interface
  - Compose models into pipelines
  - Save, load, and transport models



APACHE Spark ML

Public Preview

GPU Accelerated Workloads

Accelerates data transformation and reduces ML model training time by dramatically increasing throughput vs. traditional CPU

CPU

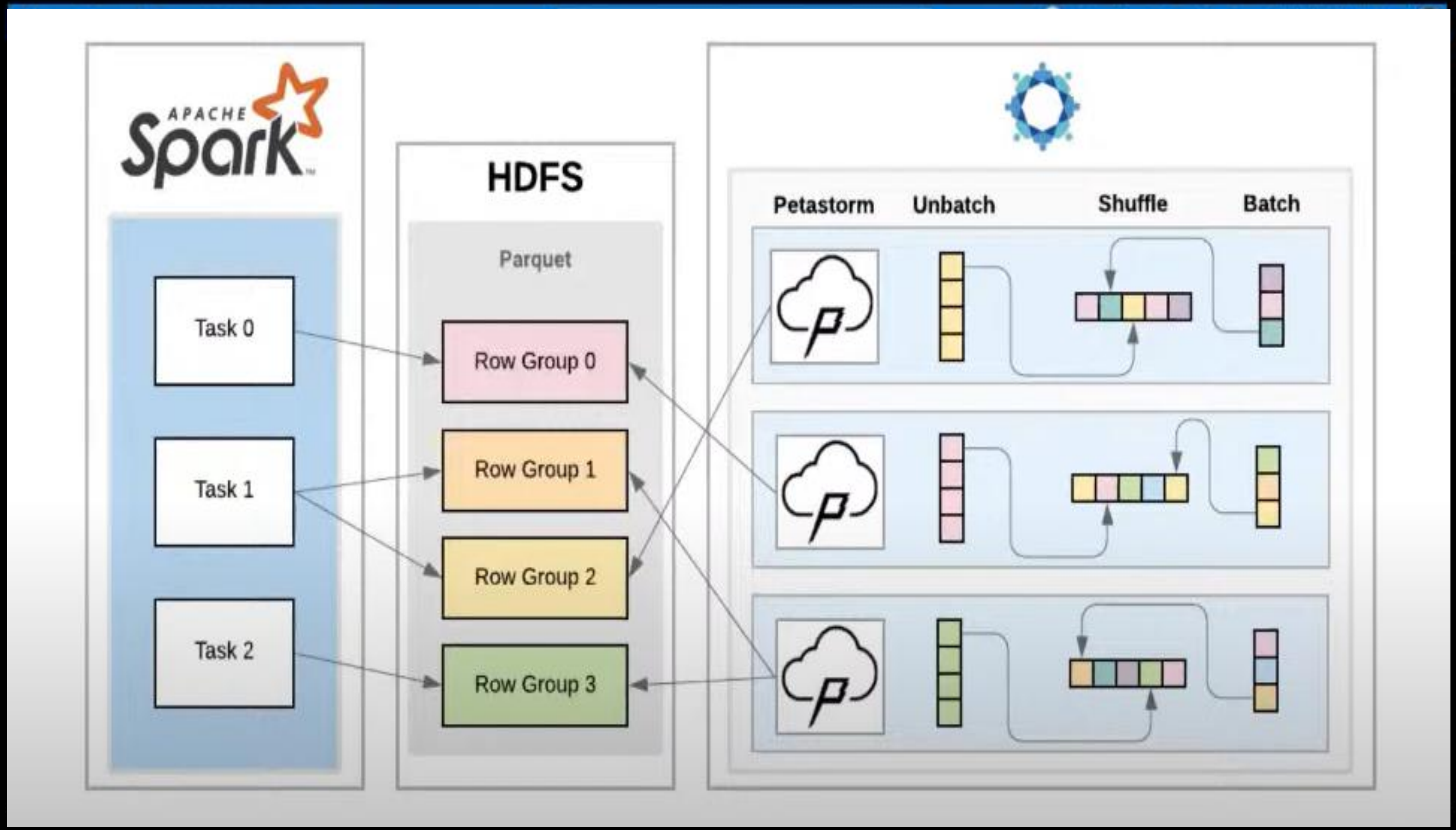
100
MB/SECOND

GPU

1,200
MB/SECOND

Public Preview

Distributed Deep Neural Network Training with Horovod and Petastorm



https://jameskle.com/writes/spark-ai-2020

Generally Available

https://aka.ms/spark

SynapseML

A machine learning library that's

Simple

Quickly create, train, and use distributed machine learning tools in only a few lines of code.

Scalable

Scale ML workloads to hundreds of machines on your Apache Spark cluster.

Multilingual

Use SynapseML from any Spark compatible language including Python, Scala, R, Java, .NET and C#.

Open

SynapseML is Open Source and can be installed and used on any Spark 3 infrastructure including your local machine, Databricks, Synapse Analytics, and others.

Synapse ML = Apache Spark ML +

Gradient Boosting

Cognitive Services

Deep learning

Reinforcement learning

Responsible AI

Language Modeling

Search engine creation

Content retrieval

Anomaly detection

Cybersecurity

Explainable Models

Image processing

Synapse ML = Apache Spark ML +

LightGBM

Cognitive Services

ONNX

Vowpal Wabbit

Responsible AI

PREDICT Keyword

Azure Search

Conditional KNNs

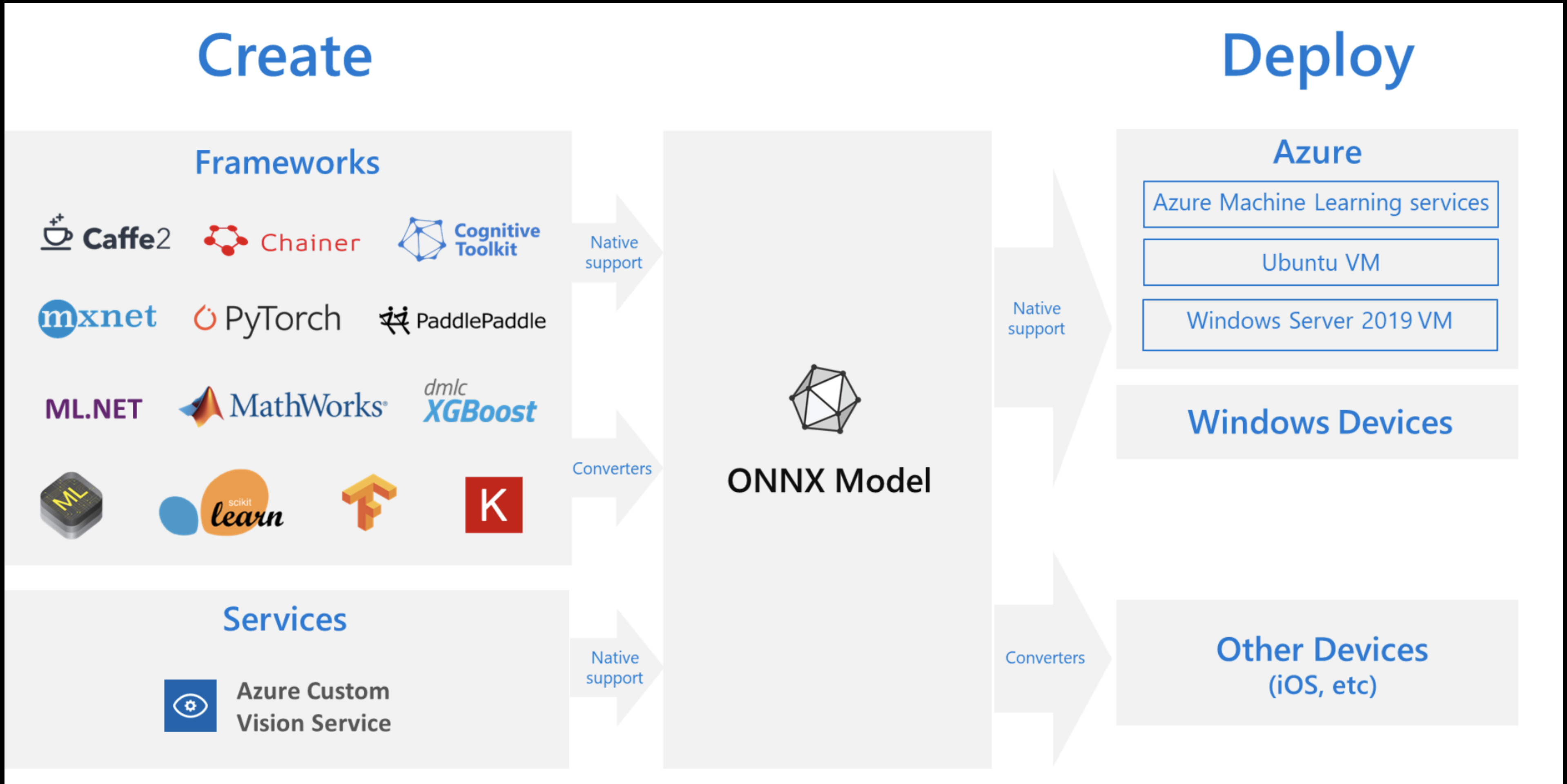
Isolation Forests

CyberML

Explainable Boosting Machines

OpenCV

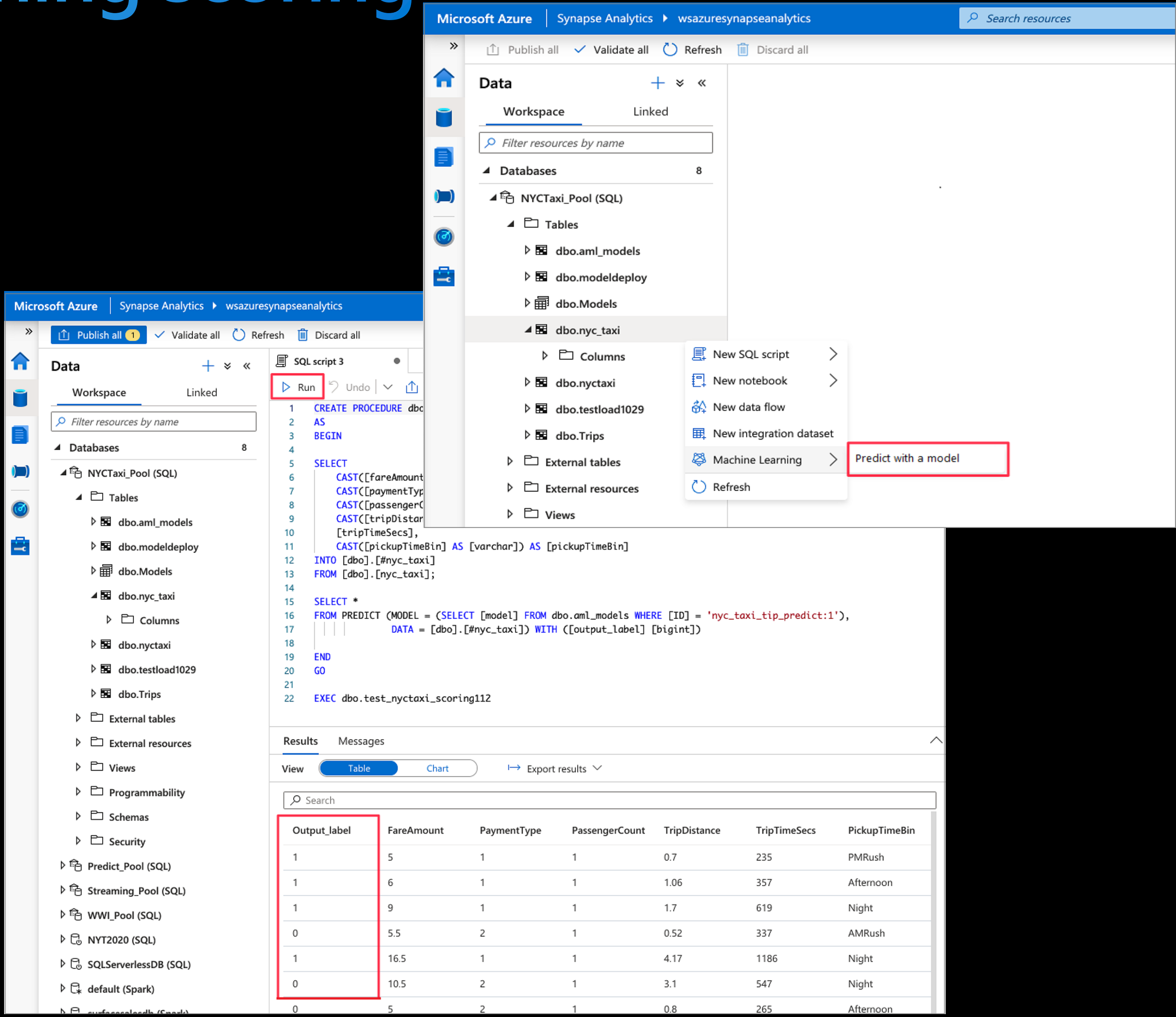
Scoring - what is ONNX?



ONNX flow diagram showing training, converters, and deployment

Code-free machine learning scoring

- No-code **references** to machine learning models
- Democratize ML to everyone since no data science domain knowledge required
- Easily embed in SQL stored procedures for transformation of Views for reporting



Model Prediction with Spark

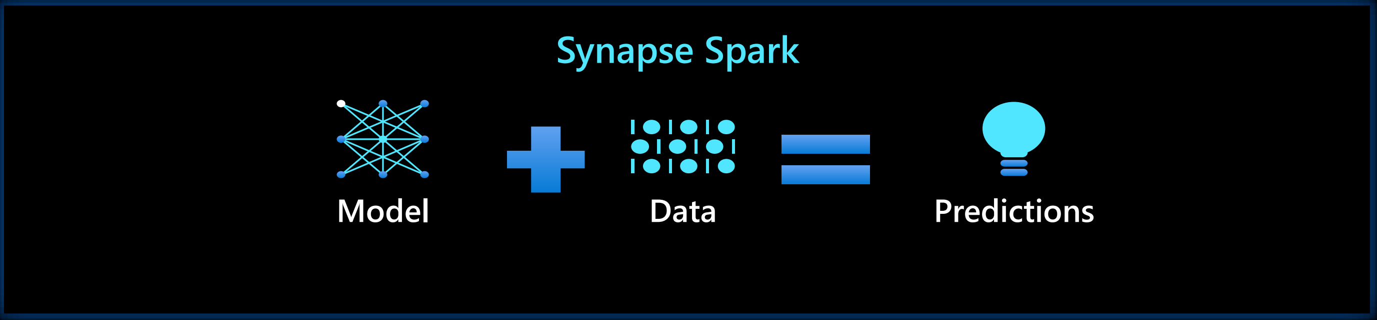
Simplifying model scoring at scale on Spark

- Drastically simplifies handover of models from producer to consumer
- Machine Learning models from Azure ML, ADLS Gen2, REST
- “In-engine” for performance and scalability
- No data leaves the platform for scoring
- No additional cost for scoring

Scope public preview: Sklearn, Pytorch, TensorFlow, Onnx and PyFunc model flavors

Scope GA: All mlflow model flavors

```
Prediction  = spark.sql("""
                            SELECT
                                PREDICT('{MODEL_ID}', *) AS preds
                            FROM data
                        """)
```



PyTorch
scikit learn
TensorFlow
ONNX
mlflow

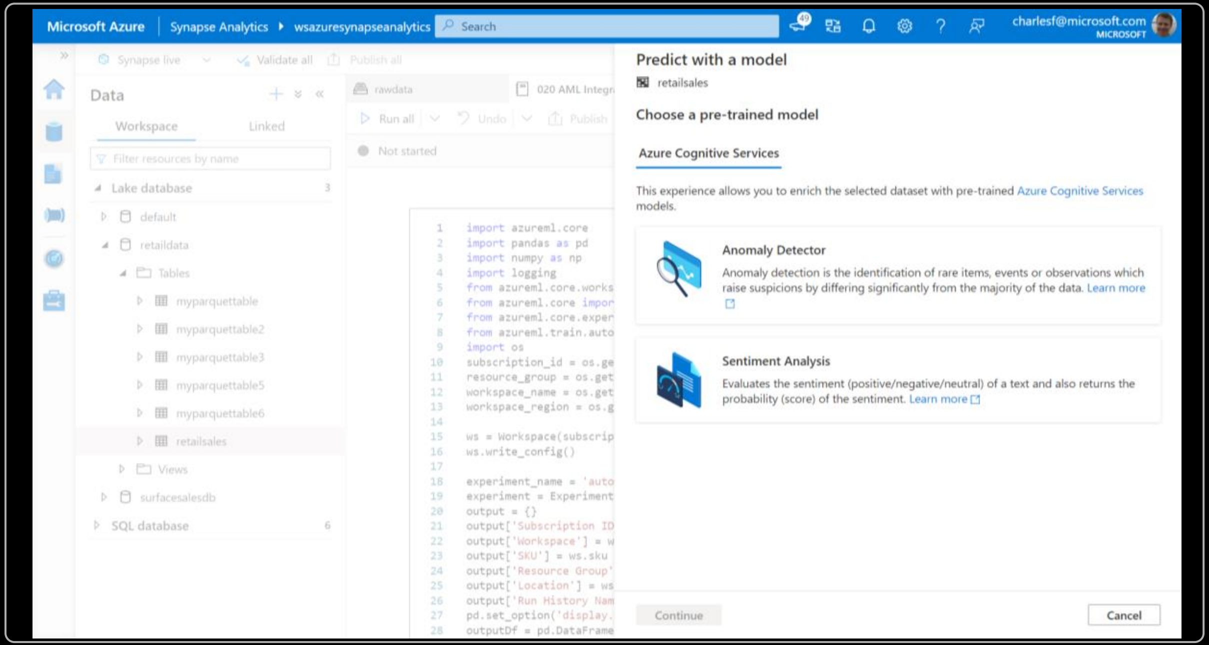
Public Preview

Generally Available

Built-in Cognitive Services

Enables simple integration of pre-built machine learning models.

- Guided UI experience for data enrichment
- SynpaseML
- Pyspark – code first



[Cognitive Services in Azure Synapse Analytics - Azure Synapse Analytics | Microsoft Learn]()

Microsoft Azure Solutions Business

Azure HDInsight

Franck Gaillard

Amplify the power of analytics in the cloud with Azure HDInsight

Azure HDInsight is a managed, full-spectrum cloud distribution of Hadoop components.

Effortlessly process massive amounts of data and get all the benefits of the broad open-source ecosystem with the global scale of Azure.

Build your projects in an open-source ecosystem

Stay up to date with the newest releases of open-source frameworks, including Kafka, HBase, and Hive LLAP.

HDInsight supports the latest open-source projects from the Apache Hadoop and Spark ecosystems.

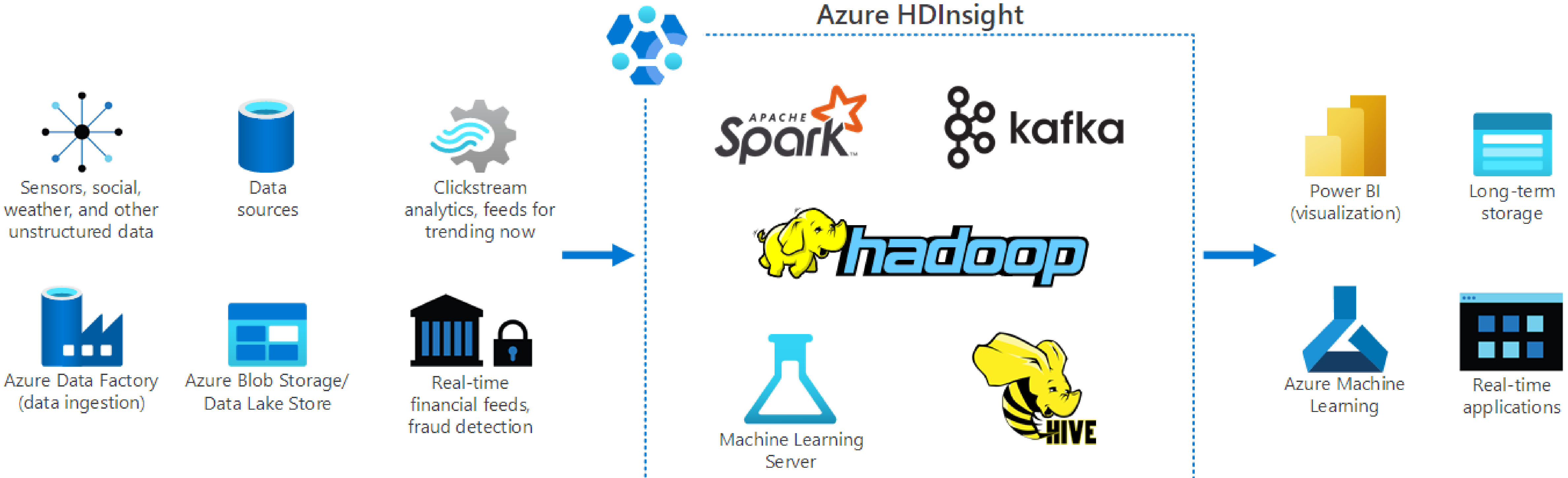
Open-source projects and clusters are easy to spin up quickly without the need to install hardware or manage infrastructure

Big data clusters reduce costs through autoscaling and pricing tiers that allow you to pay for only what you use

Enterprise-grade security and industry-leading compliance with more than 30 certifications helps protect your data

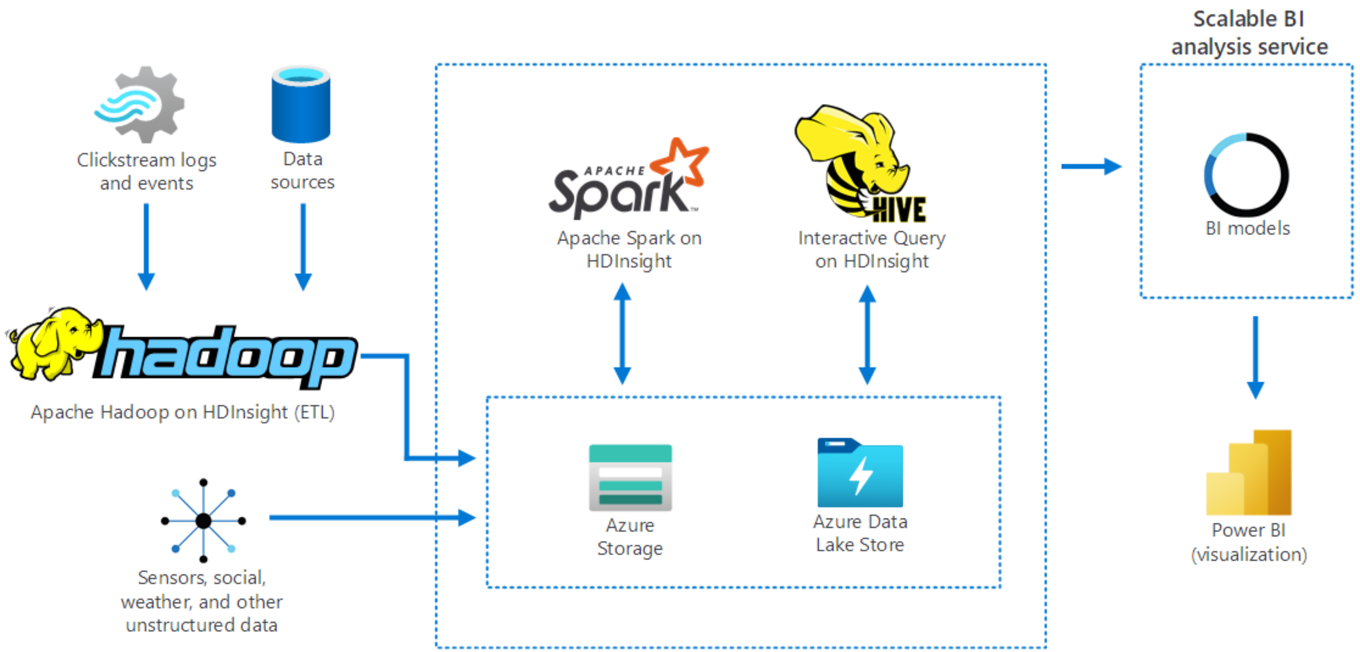
Optimized components for open-source technologies such as Hadoop and Spark keep you up to date

Overview of HDInsight Scenarii



HDInsight processes data from many locations. It then makes it available in long-term storage for real-time apps and additional analysis.

Data Warehousing



The following diagram depicts how Apache Hadoop on HDInsight gathers and stores data from several sources.

- Apache Spark and Apache Hive prepare and analyze the data.
- Finally, the data is modeled for use with Power BI for data visualization.

A data warehouse provides an organization with somewhere to store big data while waiting to analyze it.

Data warehousing enables you to:

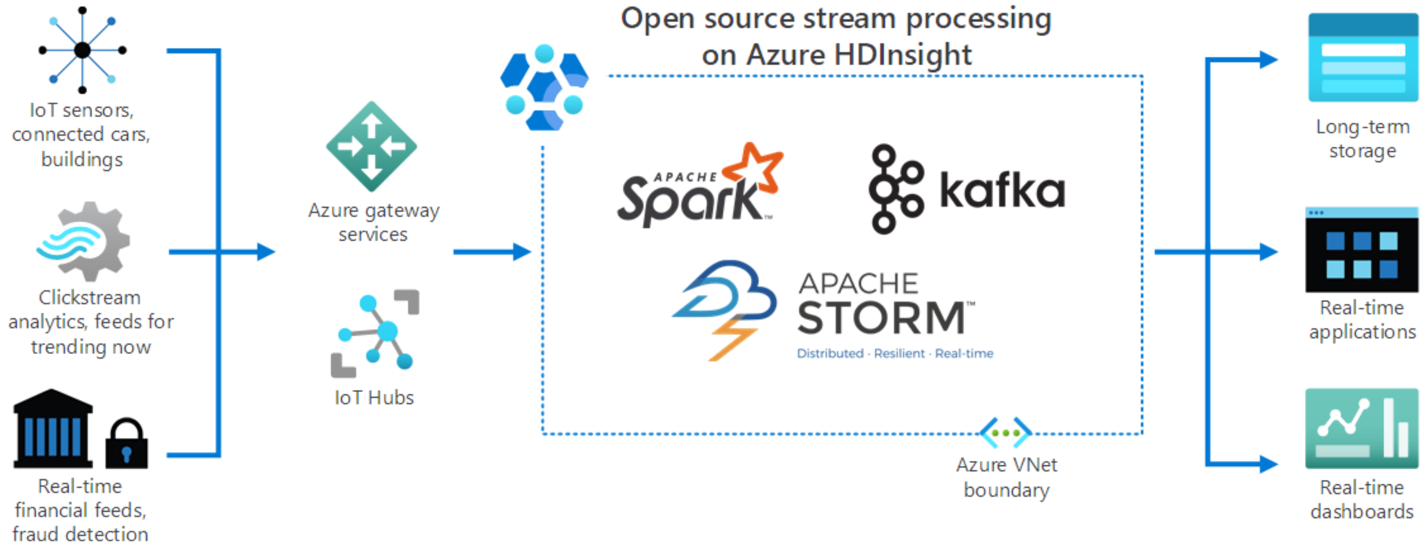
- Store your data.
- Prepare your data for analysis.
- Provide the prepared data in a structured format. You then can query the data by using analytical tools.

Components in this scenario include:

- Apache Spark is a parallel processing framework. It supports in-memory processing, which helps boost the performance of big-data analytic applications.
- Apache Hive in HDInsight is a data warehouse system for Apache Hadoop. Hive enables data summarization, querying, and analysis.

You can use these components to perform queries at petabyte scales on structured and unstructured data, in any format.

Internet of Things



Azure gateway services and IoT hubs direct data from various sources to these frameworks.

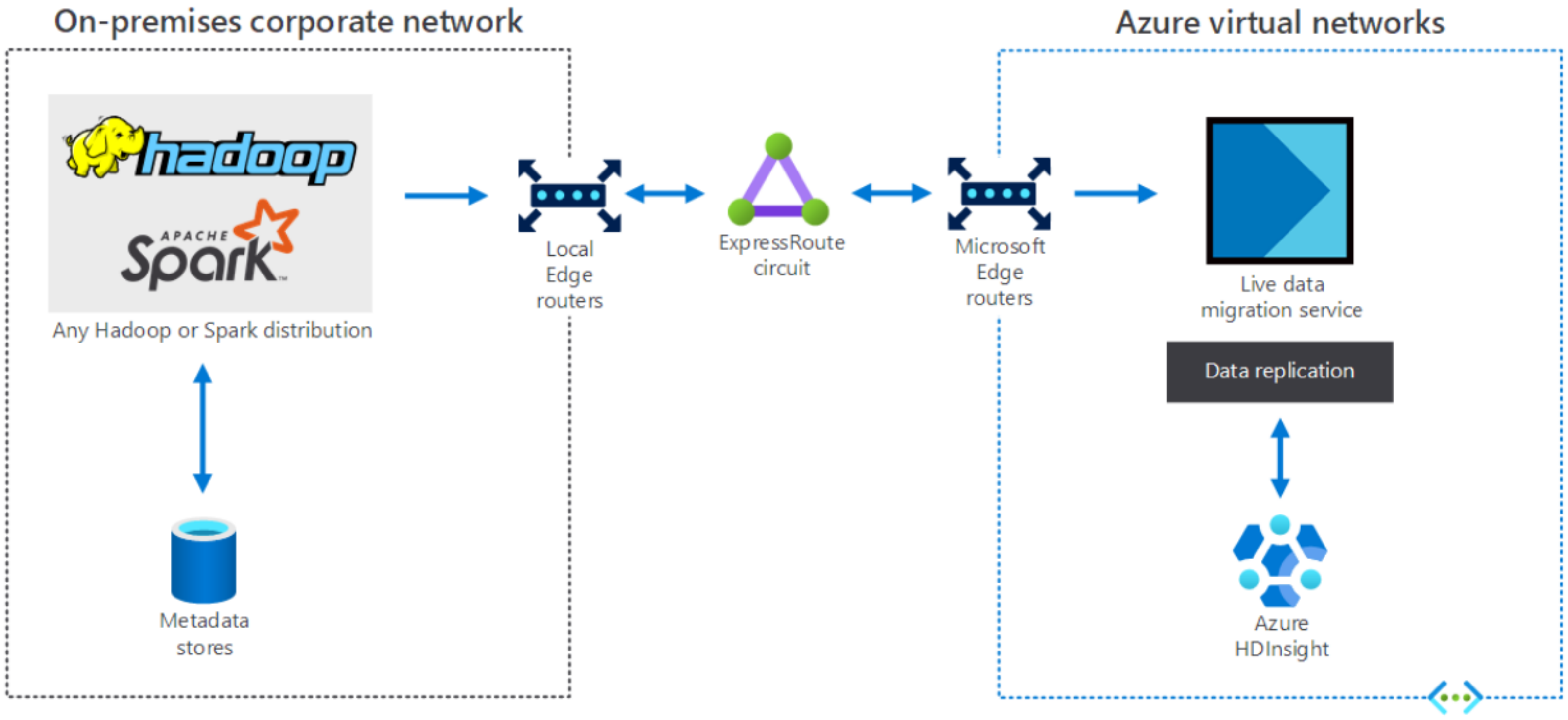
The frameworks then process the data, and it passes to:

- Long-term storage
- Real-time apps
- Real-time dashboards

As the following diagram depicts, HDInsight processes streaming data received in real time from different devices and sensors. In this example, several open-source frameworks provide stream processing, including:

- Apache Spark
- Apache Kafka
- Apache Storm

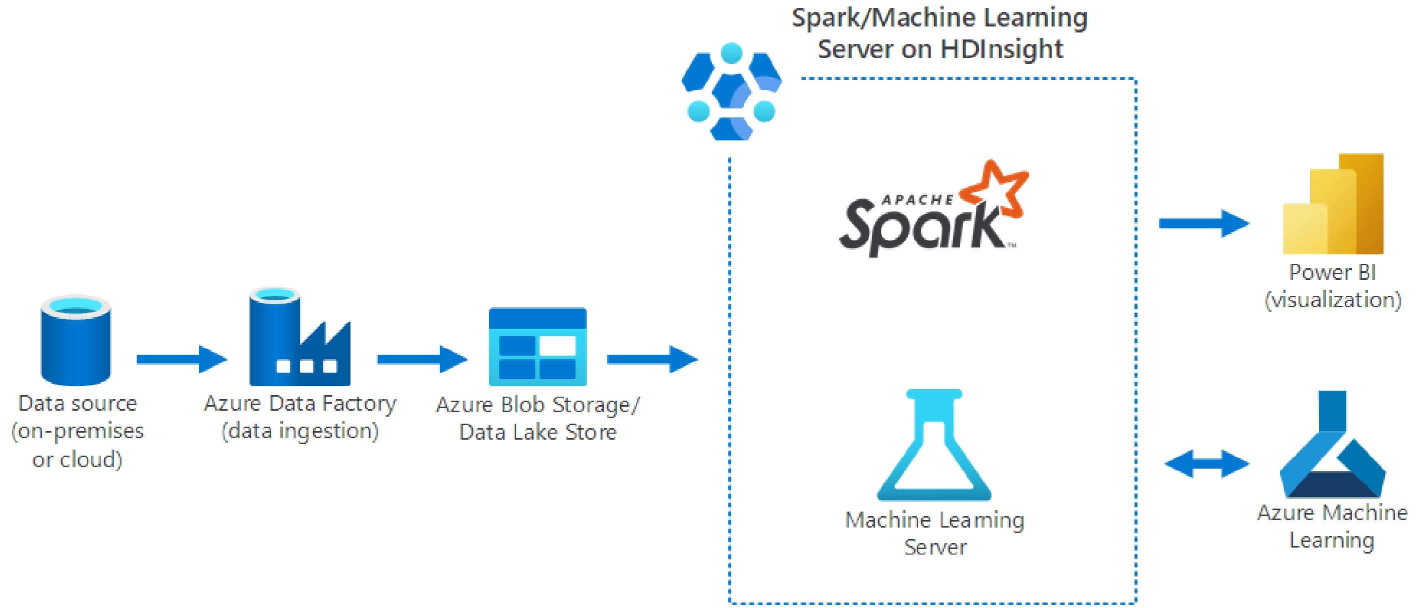
Hybrid



Organizations that have an on-premises, big-data infrastructure can use HDInsight to extend into Azure.
This provides you with the benefits of the Azure cloud's advanced analytics capabilities.
The following diagram depicts the hybrid scenario, in which:

- The on-premises big-data infrastructure consists of metadata stores and a Hadoop or Spark distribution on local VMs.
- An Azure ExpressRoute circuit connects the on-premises corporate network environment to Azure virtual networks.
- A live data migrator for Azure replicates the data received from on-premises to HDInsight.

Data Science



You can use HDInsight to complete common data science tasks such as:

- Data ingestion
- Feature engineering
- Modeling
- Model evaluation

The following diagram depicts a data-science scenario, in which:

1. Data is collected from an on-premises data source by using Azure Data Factory.
2. The ingested data is then stored in Azure storage (either Azure Blob Storage or a Data Lake Store).
3. Azure Spark on HDInsight processes and prepares the data for Azure Machine Learning.
4. Data is also visualized by using Power BI.

Spark ML and MLlib

- HDInsight Spark is an Azure-hosted offering of Apache Spark, a unified, open source, parallel data processing framework supporting in-memory processing to boost big data analytics.
- The Spark processing engine is built for speed, ease of use, and sophisticated analytics.
- Spark's in-memory distributed computation capabilities make it a good choice for the iterative algorithms used in machine learning and graph computations.
- There are two scalable machine learning libraries that bring algorithmic modeling capabilities to this distributed environment: MLlib and SparkML.
- SparkML is a newer package that provides a higher-level API built on top of DataFrames for constructing ML pipelines.
- SparkML doesn't yet support all of the features of MLlib, but is replacing MLlib as Spark's standard machine learning library.

SynapseML

- **SynapseML** (previously known as MMLSpark), is an open-source library that simplifies the creation of massively scalable machine learning (ML) pipelines. SynapseML provides simple, composable, and distributed APIs for a wide variety of different machine learning tasks such as text analytics, vision, anomaly detection, and many others.
- This library is designed to make data scientists more productive on Spark, increase the rate of experimentation, and leverage cutting-edge machine learning techniques, including deep learning, on very large datasets.

Vowpal Wabbit on Spark	The Cognitive Services for Big Data	LightGBM on Spark	Spark Serving
Fast, Sparse, and Effective Text Analytics	Leverage the Microsoft Cognitive Services at Unprecedented Scales in your existing SparkML pipelines	Train Gradient Boosted Machines with LightGBM	Serve any Spark Computation as a Web Service with Sub-Millisecond Latency

HTTP on Spark	ONNX on Spark	Responsible AI	Spark Binding Autogeneration
An Integration Between Spark and the HTTP Protocol, enabling Distributed Microservice Orchestration	Distributed and Hardware Accelerated Model Inference on Spark	Understand Opaque-box Models and Measure Dataset Biases	Automatically Generate Spark bindings for PySpark and SparklyR

[GitHub - microsoft/SynapseML: Simple and Distributed Machine Learning]

Benefits of Azure HDInsight

- Open-source. Enables you to create optimized clusters for various open-source frameworks.
  - ✓ Hadoop
  - ✓ Apache Spark
  - ✓ Apache Hive
  - ✓ Apache Kafka
- Reliable. Provides an end-to-end SLA for all production workloads.
- Scalable. Enables you to scale workloads to respond to demand changes.
- Secure. Enables you to protect your enterprise data assets through integration with:
  - ✓ Azure Virtual Network
  - ✓ Azure encryption technologies
  - ✓ Azure Active Directory
- Compliant. Meets popular industry and government compliance standards.
- Monitored. Integrates with Azure Monitor logs to provide a single interface. Monitor all clusters by using the single interface.

Microsoft Azure Solutions Business

AI dans Azure SQL Database et Azure SQL Managed Instance

Ali Bouhaddou

Azure SQL Database deployment option

Azure SQL Database

Single

Database-scoped deployment option with predictable workload performance

Best for apps that require resource guarantee at database level

Elastic Pool

Shared resource model optimized for greater efficiency of multi-tenant applications

Best for SaaS apps with multiple databases that can share resources at database level, achieving better cost efficiency

Managed Instance

Instance-scoped deployment option with high compatibility with SQL Server and full PaaS benefits

Best for modernization at scale with low friction and effort

Service Tiers

General Purpose

Business Critical

Hyperscale

Serverless

Managed Instance key capabilities

Azure SQL Database

Single

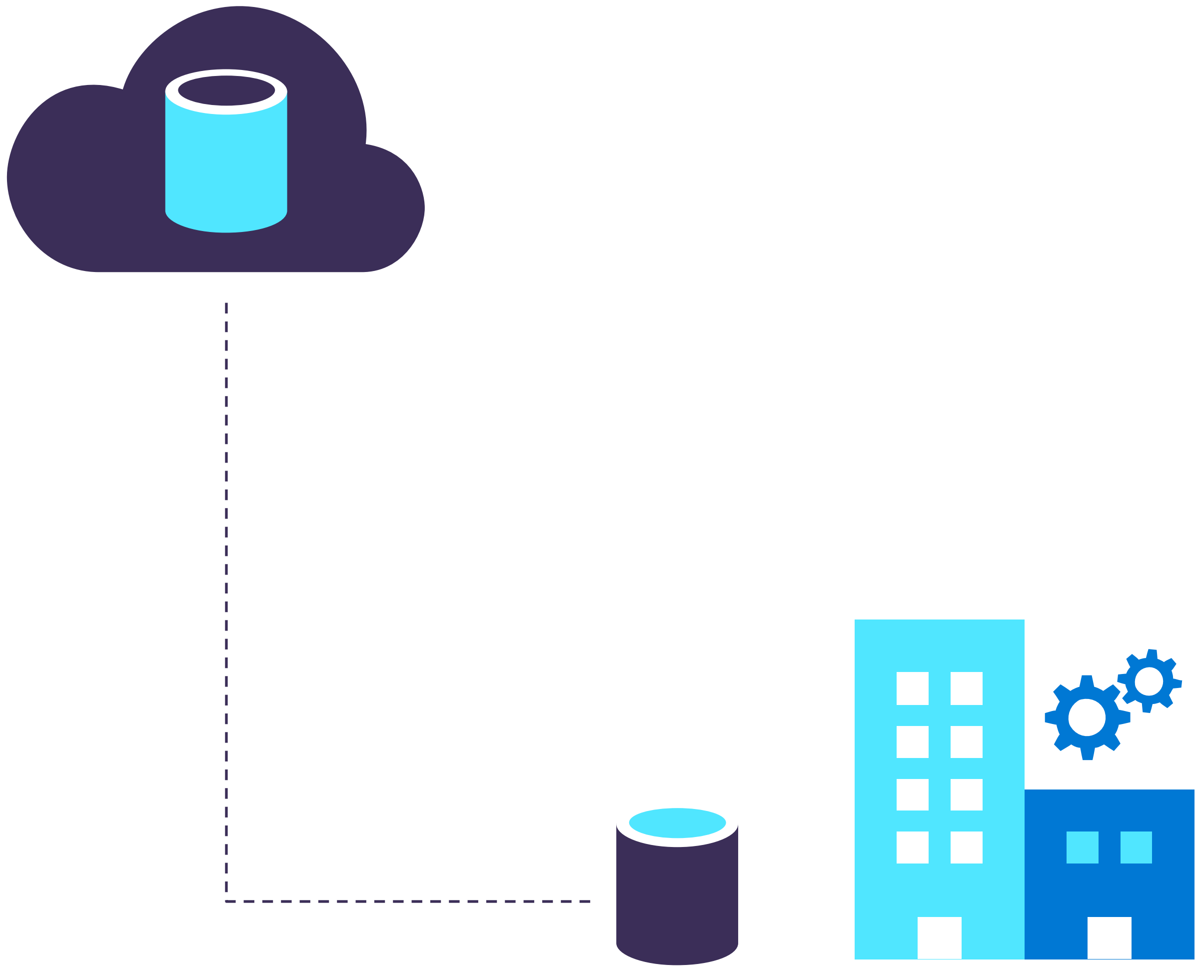
Managed Instance

Elastic Pool

Fully-managed service	SQL Server compatibility	Full isolation and security	New pricing options
• Built on the same infrastructure as SQL Database • Provides the same benefits (PaaS)	• Fully-fledged SQL instance with nearly 100% compat with on-premise	• Contained within your VNet • Private IP addresses • Express Route / VPN connectivity	• Transparent • Frictionless • Competitive

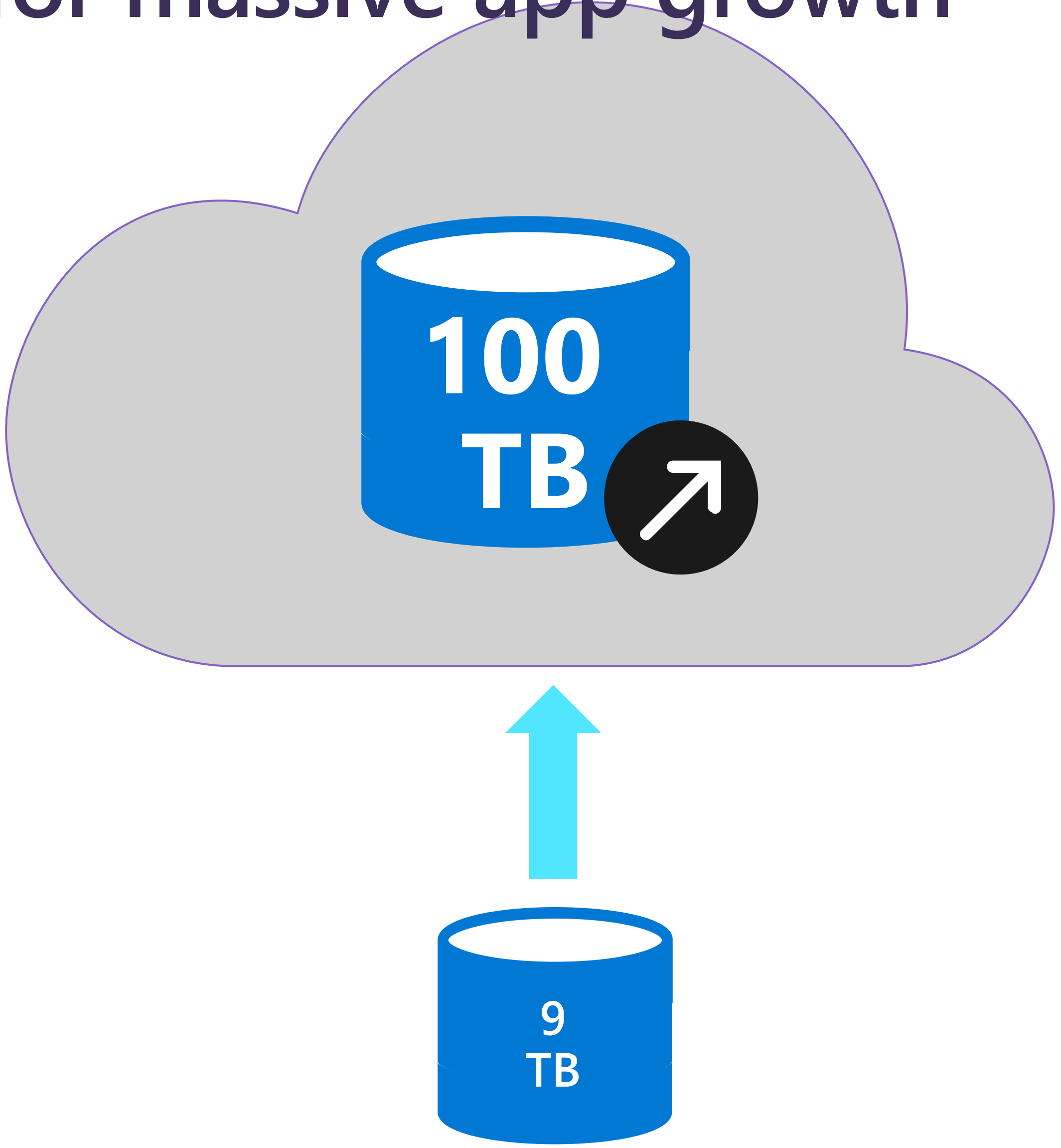
Broader SQL Server support for improved compatibility on Azure

- Online index rebuild capability for clustered and non-clustered indexes for greater availability
- Build highly optimized schemas to improve query processing with table partitioning support
- Access Common Language Runtime (CLR) and define CLR types, aggregates, functions, and procedures written in C#
- In-Memory Columnstore index for data marts
- Support for additional Dynamic Management Views (DMVs) for deeper insight into application health



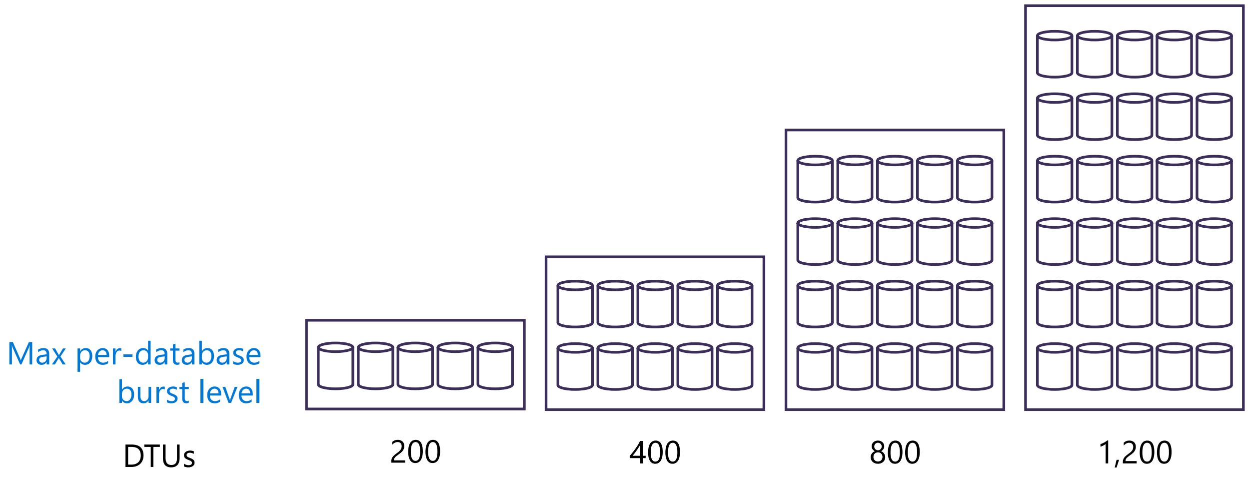
Hyperscale is the foundation for massive app growth

- **Hyperscale is a new, highly scalable service tier that adapts on-demand to your workload's needs, auto-scaling up to 100TB per database.**
  - Storage dynamically adapts to your workloads' needs, auto-scaling up to 100TB.
  - Provision one or more additional compute nodes that can serve your read-only workload and use them as a hot-standby, in case of failover.
  - Perform operations in constant time, regardless of the size of the data operation.
  - Compute and storage resources scale rapidly and independently without sacrificing performance.



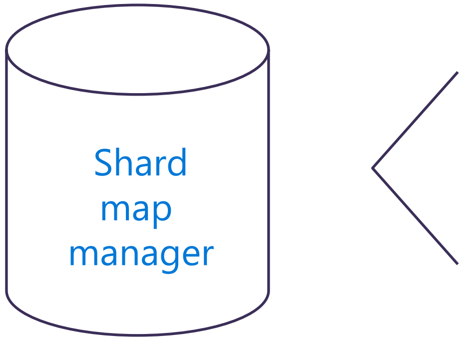
Elastic database model

- Elastic databases in elastic database pools
- Pooled resources are used by many databases
- Standard elastic database pools provide 50-3000 database throughput units (DTUs) for up to 500 databases
- Max eDTUs per database can be set if available based on utilization by other database in the pool
- Create/configure pools using portal, Azure PowerShell,
  REST APIs
- Move databases in/out using portal, Azure PowerShell, REST APIs, and T-SQL
- Databases remain online throughout
- Monitoring and alerting available on both pools and databases



Elastic database client library overview

- Two types of shard maps
  - Range: contiguous values
  - List: explicit values
- Four types of sharding keys
  - INT, BIGINT, GUID, VARBINARY



[shardmaps_global]

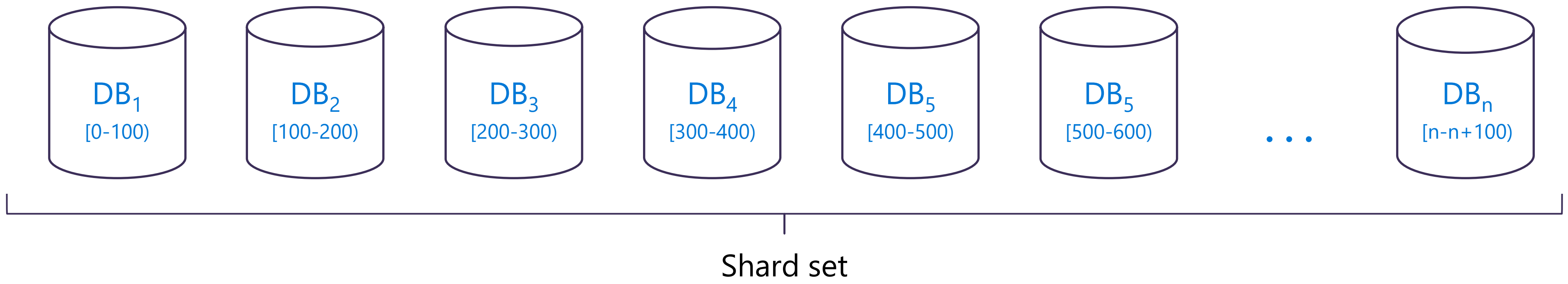
smid	name
1	RangeShardMap

[shards_global]

sid	smid	Datasource	Databasename
1	1	serverName	DB2
2	1	serverName	DB2

[shard_mappings_global]

mid	smid	min	max	Sid
1	1	0	100	1
2	1	100	200	2



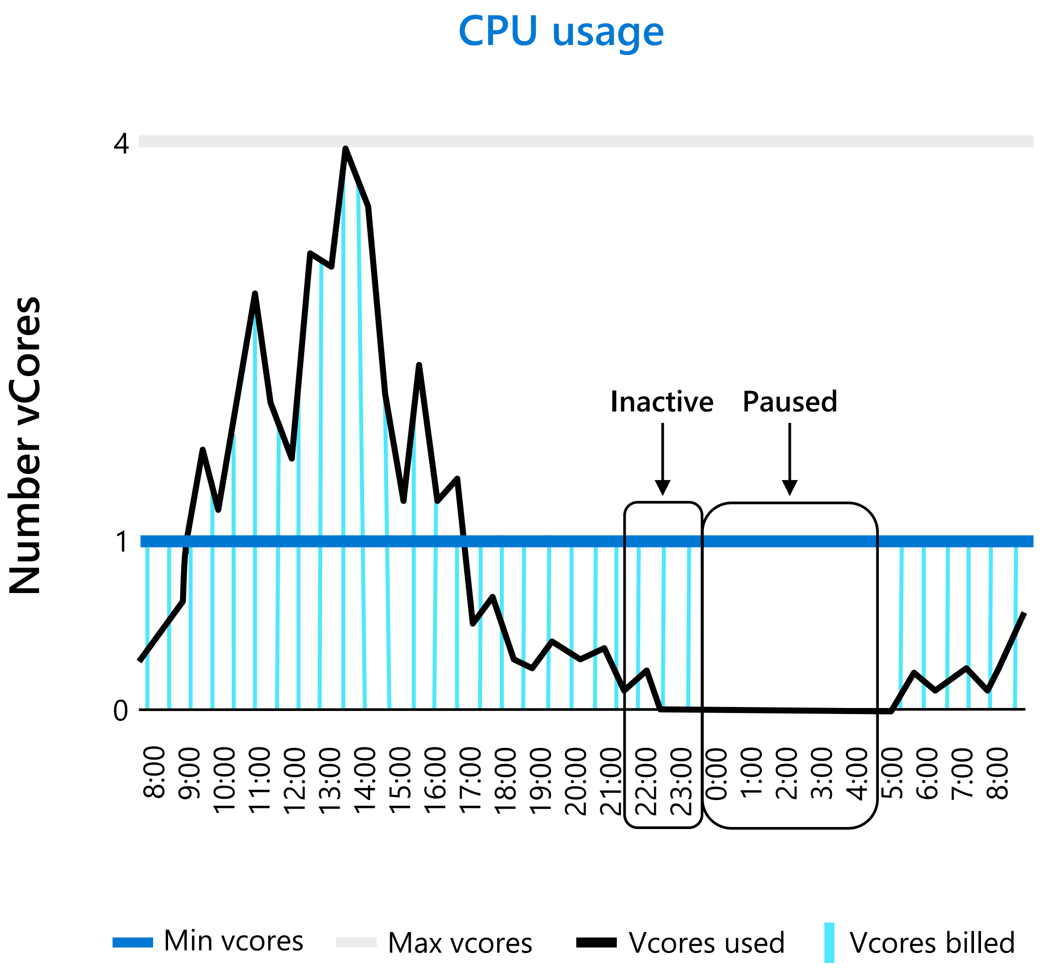
Optimize price to performance with per-second billing

Compute resources scale dynamically up or down based on workload requirements

Configure minimum and maximum vCores to define the range of available compute capacity

Use auto-pause delay to define the time period the dataset must be inactive before pausing

Pay for compute based on the vCores and memory used per second, with lowest billing based on configured vCore minimum



Machine Learning in Azure Databases

The Team Data Science Process

Business Understanding

- Define Objectives
- Identify Data Sources

Data Acquisition and Understanding

- Ingest Data
- Explore Data
- Update Data

Modeling

- Feature Selection
- Create and Train Model

Deployment

- Operationalize

Customer Acceptance

- Testing and Validation
- Handoff
- Re-train and re-score

http://aka.ms/tdsp

The Azure Platform for Analytics and AI

Information Management
- Purview
- Data Factory
- Event Hubs

Big Data
- Azure Storage
- Data Lake
- Azure Synapse
- Azure Cosmos DB

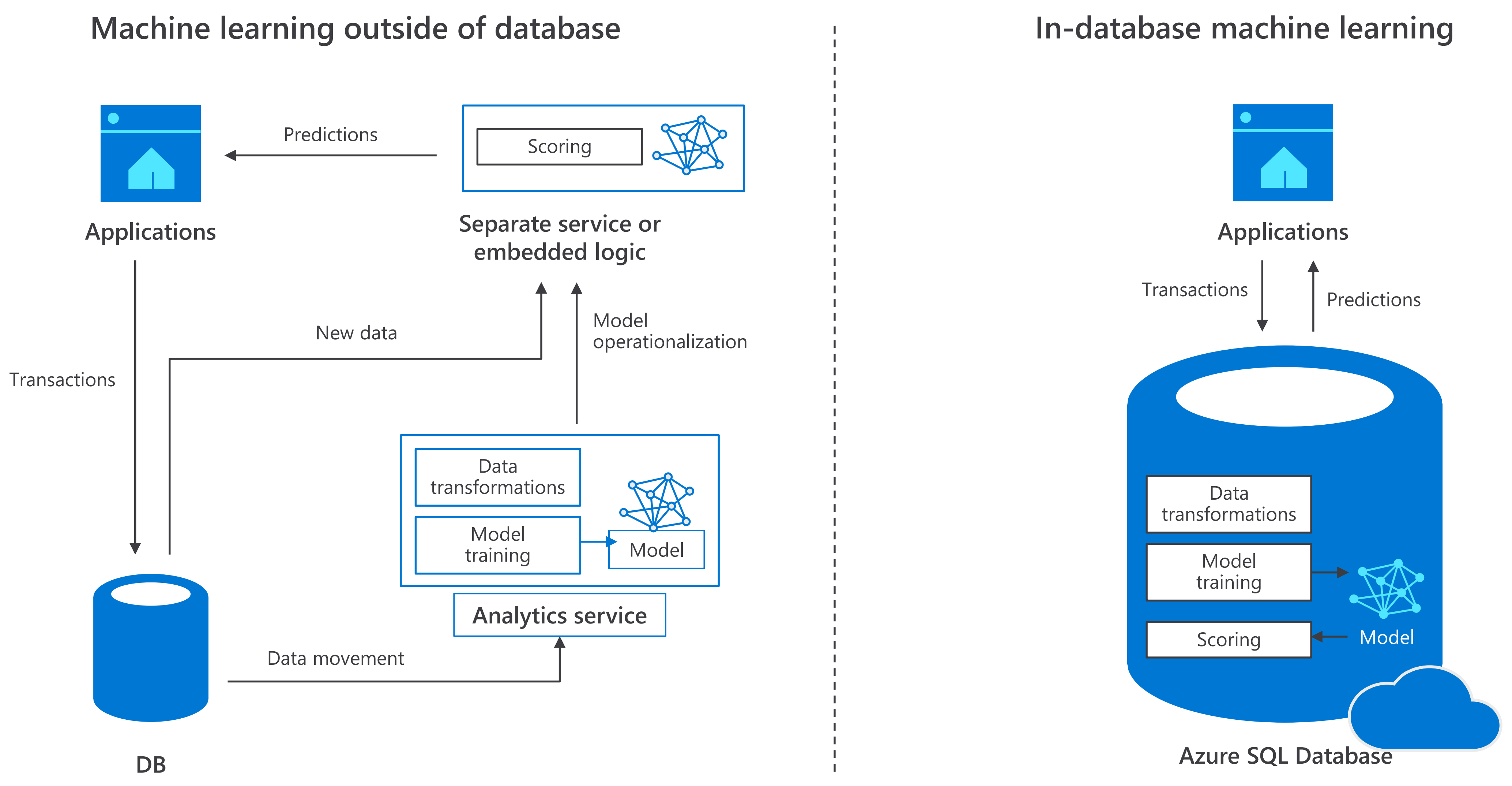
Intelligence and Advanced Analytics
- Cortana, Bot Service, Cognitive Framework
- **Azure Machine Learning Services**
- HDInsight
- Data Lake
- Stream Analytics
- Analysis Services (oups :))

Visualization
- Power BI
- R/Python

Solutions
- Templates and Gallery

http://aka.ms/cisc

Machine Learning Services in Azure SQL Database



Machine Learning Services in Azure SQL Database

Capabilities

Extensible in-database analytics, exposed through T-SQL

With R, Python

No data movement, resulting in faster time to insights

Real-time analytics on transactional data with native PREDICT

Integration with existing application workflows

Unified governance across analytics and storage

Running R script in Azure SQL Database:

```
/* Input table schema */
create table Iris_Data (name varchar(100), length int, width int);

/* Model table schema */
create table my_iris_model (model varbinary(max));

declare @iris_model varbinary(max) = (select model from my_iris_model);

exec sp_execute_external_script
  @language = 'R'
, @script = '
    IrisPredict <- function(data, model){
    library(e1071)
    predicted_species <- predict(model, data)
    return(predicted_species)
    }
    IrisPredict(input_data_1, model);
  '
, @parallel = default
, @input_data_1 = N'select * from Iris_Data'
, @params = N'@model varbinary(max)'
, @model = @iris_model
with result sets ((name varchar(100), length int, width int
, species varchar(30)));
```

Values highlighted in yellow are SQL queries embedded in the original R script

Values highlighted in aqua are R variables that bind to SQL variables by name

Installation, Overview and Setup

- https://github.com/microsoft/sqlworkshops-sqlmlsvc
- https://docs.microsoft.com/en-us/sql/machine-learning/tutorials/demo-data-nyctaxi-in-sql?view=sql-server-ver15

In [1]:

```
-- Examine the current settings on your SQL Server Instance
sp_configure
```

Commands completed successfully.

Total execution time: 00:00:00.161

Out[1]:

name	minimum	maximum	config_value	run_value
allow polybase export	0	1	0	0
allow updates	0	1	0	0
backup checksum default	0	1	0	0
backup compression default	0	1	0	0
clr enabled	0	1	0	0
column encryption enclave type	0	2	0	0
contained database authentication	0	1	0	0
cross db ownership chaining	0	1	0	0
default language	0	9999	0	0
external scripts enabled	0	1	0	0

Installation, Overview and Setup

In [2]:

```
-- Enable SQL Server Machine Learning Services
EXEC sp_configure  'external scripts enabled', 1
RECONFIGURE WITH OVERRIDE

-- Next, restart the Instance before you run the next cell
```

Configuration option 'external scripts enabled' changed from 0 to 1. Run the RECONFIGURE statement to install.

Total execution time: 00:00:00.021

In [1]:

```
-- Check to see if the service is running (run_value should be 1)
EXECUTE sp_configure  'external scripts enabled'
```

Commands completed successfully.

Total execution time: 00:00:00.029

Out[1]:

name	minimum	maximum	config_value	run_value
external scripts enabled	0	1	1	1

Installation, Overview and Setup

In [2]:

```
-- Check R, and then Python

EXEC sp_execute_external_script  @language =N'R',
@script=N'
OutputDataSet <- InputDataSet;
',
@input_data_1 =N'SELECT 1 AS R_Is_Functional'
WITH RESULT SETS (([R_Is_Functional] int not null));
GO

EXEC sp_execute_external_script  @language =N'Python',
@script=N'
OutputDataSet = InputDataSet;
',
@input_data_1 =N'SELECT 1 AS Python_Is_Functional'
WITH RESULT SETS (([Python_Is_Functional] int not null));
GO
```

Out[2]:

R_Is_Functional
1

(1 row affected)

(1 row affected)

Total execution time: 00:00:24.955

Out[2]:

Python_Is_Functional
1

Explore and Visualize the data

- **Create a plot as varbinary data**
- A stored procedure returns a serialized Python figure object as a stream of varbinary data.

- You cannot view the binary data directly, but you can use Python code on the client to deserialize and view the figures, and then save the image file on a client computer.

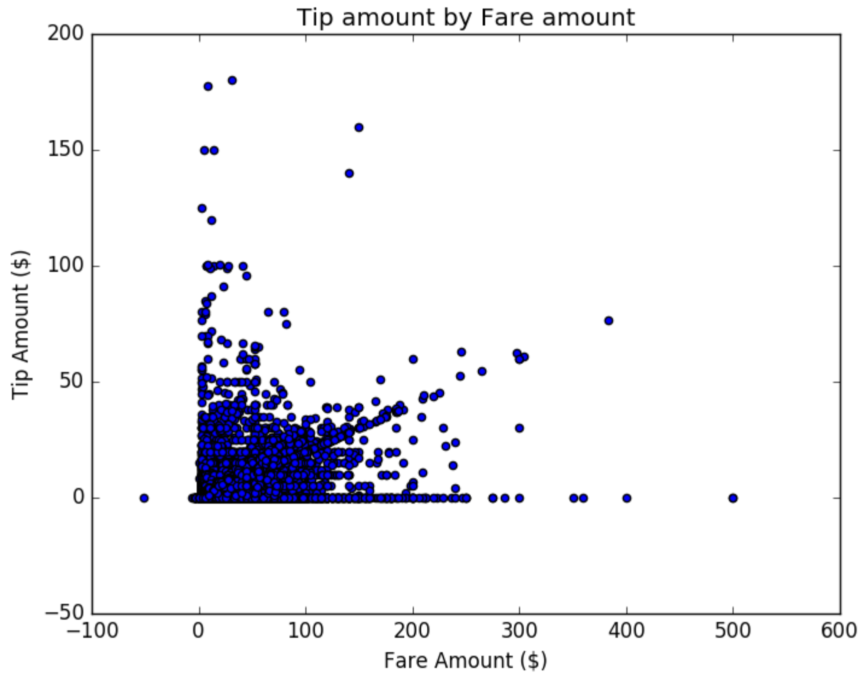
```
In [1]: USE NYCTaxi;
GO

DROP PROCEDURE IF EXISTS PyPlotMatplotlib;
GO

CREATE PROCEDURE [dbo].[PyPlotMatplotlib]
AS
BEGIN
    SET NOCOUNT ON;
    DECLARE @query nvarchar(max) =
    N'SELECT cast(tipped as int) as tipped, tip_amount, fare_amount FROM [dbo].[nyctaxi_sample]'
    EXECUTE sp_execute_external_script
    @language = N'Python',
    @script = N'
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt
import pandas as pd
import pickle

fig_handle = plt.figure()
plt.hist(InputDataSet.tipped)
plt.xlabel("Tipped")
plt.ylabel("Counts")
plt.title("Histogram, Tipped")
plot0 = pd.DataFrame(data =[pickle.dumps(fig_handle)], columns =["plot"])
plt.clf()

plt.hist(InputDataSet.tip_amount)
plt.xlabel("Tip amount ($)")
plt.ylabel("Counts")
plt.title("Histogram, Tip amount")
plot1 = pd.DataFrame(data =[pickle.dumps(fig_handle)], columns =["plot"])
plt.clf()

plt.hist(InputDataSet.fare_amount)
plt.xlabel("Fare amount ($)")
plt.ylabel("Counts")
plt.title("Histogram, Fare amount")
plot2 = pd.DataFrame(data =[pickle.dumps(fig_handle)], columns =["plot"])
plt.clf()
```

Explore and Visualize the data

- **Use Python from an External Client to Create Graphs**
- The Developer or Data Scientist can now use the stored procedure to create graphical outputs.

```
%matplotlib notebook
import pyodbc
import pickle
import os
cnxn = pyodbc.connect('DRIVER=SQL Server;SERVER={SERVER_NAME};DATABASE={DB_NAME};Trusted_Connection=True;')
cursor = cnxn.cursor()
cursor.execute("EXECUTE [dbo].[PyPlotMatplotlib]")
tables = cursor.fetchall()
for i in range(0, len(tables)):
    fig = pickle.loads(tables[i][0])
    fig.savefig(str(i)+'.png')
print("The plots are saved in directory: ",os.getcwd())
```

- From a Python tool of your choice, you can execute a script like this one, replacing the connection information for your SQL Server Instance



Train and save model

- **Train and save a Python model using T-SQL**
- These Python libraries are already installed with SQL Server Machine Learning Services
- The following code creates the stored procedure *PyTrainScikit*.

In [3]:
```
DROP PROCEDURE IF EXISTS PyTrainScikit;
GO

CREATE PROCEDURE [dbo].[PyTrainScikit] (@trained_model varbinary(max) OUTPUT)
AS
BEGIN
EXEC sp_execute_external_script
  @language = N'Python',
  @script = N'
import numpy
import pickle
from sklearn.linear_model import LogisticRegression

##Create SciKit-Learn logistic regression model
X = InputDataSet[["passenger_count", "trip_distance", "trip_time_in_secs", "direct_distance"]]
y = numpy.ravel(InputDataSet[["tipped"]])

SKLalgo = LogisticRegression()
logitObj = SKLalgo.fit(X, y)

##Serialize model
trained_model = pickle.dumps(logitObj)
',
@input_data_1 = N'
select tipped, fare_amount, passenger_count, trip_time_in_secs, trip_distance,
dbo.fnCalculateDistance(pickup_latitude, pickup_longitude,  dropoff_latitude, dropoff_longitude) as direct_distance
from nyctaxi_sample_training
',
@input_data_1_name = N'InputDataSet',
@params = N'@trained_model varbinary(max) OUTPUT',
@trained_model = @trained_model OUTPUT;
;
END;
GO
```

Train and save model

- Insert the trained model into table

In [4]:
```sql
-- Insert the trained model into table nyc_taxi_models:

DECLARE @model VARBINARY(MAX);
EXEC PyTrainScikit @model OUTPUT;
INSERT INTO nyc_taxi_models (name, model) VALUES('SciKit_model', @model);
GO
```

STDERR message(s) from external script: C:\Program Files\Microsoft SQL Server\MSSQL15.MSSQLSERVER\PYTHON_SERVICES\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning. FutureWarning)

(1 row affected)

Total execution time: 00:00:35.857

In [5]:
```sql
-- Show the model - note that the code above could have specified more columns to serve as a model version:

SELECT * FROM nyc_taxi_models;
GO
```

(1 row affected)

Total execution time: 00:00:00.016

Out[5]:

model
0x800363736B6C6561726E2E6C696E6561725F6D6F64656C2E6C6F6769737469630A4C6F67697374696352656772657373696F6E0A7100298171017D710228580700

Run Batch and predictions using T-SQL

- **Batch Scoring**

- This procedure illustrates the basic syntax for wrapping a Python prediction call in a stored procedure

- The name of the exact model to use is provided as input parameter to the stored procedure

In [1]:

```
DROP PROCEDURE IF EXISTS PredictTipSciKitPy;
GO

CREATE PROCEDURE [dbo].[PredictTipSciKitPy] (@model varchar(50), @inquery nvarchar(max))
AS
BEGIN
DECLARE @lmodel2 varbinary(max) = (select model from nyc_taxi_models where name = @model);
EXEC sp_execute_external_script
  @language = N'Python',
  @script = N'
import pickle;
import numpy;
from sklearn import metrics

mod = pickle.loads(lmodel2)
X = InputDataSet[["passenger_count", "trip_distance", "trip_time_in_secs", "direct_distance"]]
y = numpy.ravel(InputDataSet[["tipped"]])

probArray = mod.predict_proba(X)
probList = []
for i in range(len(probArray)):
  probList.append((probArray[i])[1])

probArray = numpy.asarray(probList)
fpr, tpr, thresholds = metrics.roc_curve(y, probArray)
aucResult = metrics.auc(fpr, tpr)
print ("AUC on testing data is: " + str(aucResult))

OutputDataSet = pandas.DataFrame(data = probList, columns = ["predictions"])
',
  @input_data_1 = @inquery,
  @input_data_1_name = N'InputDataSet',
  @params = N'@lmodel2 varbinary(max)',
  @lmodel2 = @lmodel2
WITH RESULT SETS ((Score float));
END
GO
```

Commands completed successfully.

Commands completed successfully.

Total execution time: 00:00:00.011

Run Batch and predictions using T-SQL

- **Run batch scoring using a SELECT query**
- The stored procedures *PredictTipSciKitPy* and *PredictTipRxPy* require two input parameters:
- The query that retrieves the data for scoring
- The name of a trained model

In [3]:
```
DECLARE @query_string nvarchar(max) -- Specify input query
  SET @query_string='
  select tipped, fare_amount, passenger_count, trip_time_in_secs, trip_distance,
  dbo.fnCalculateDistance(pickup_latitude, pickup_longitude,  dropoff_latitude, dropoff_longitude) as direct_distance
  from nyctaxi_sample_testing'
EXEC [dbo].[PredictTipSciKitPy] 'SciKit_model', @query_string;
GO
```

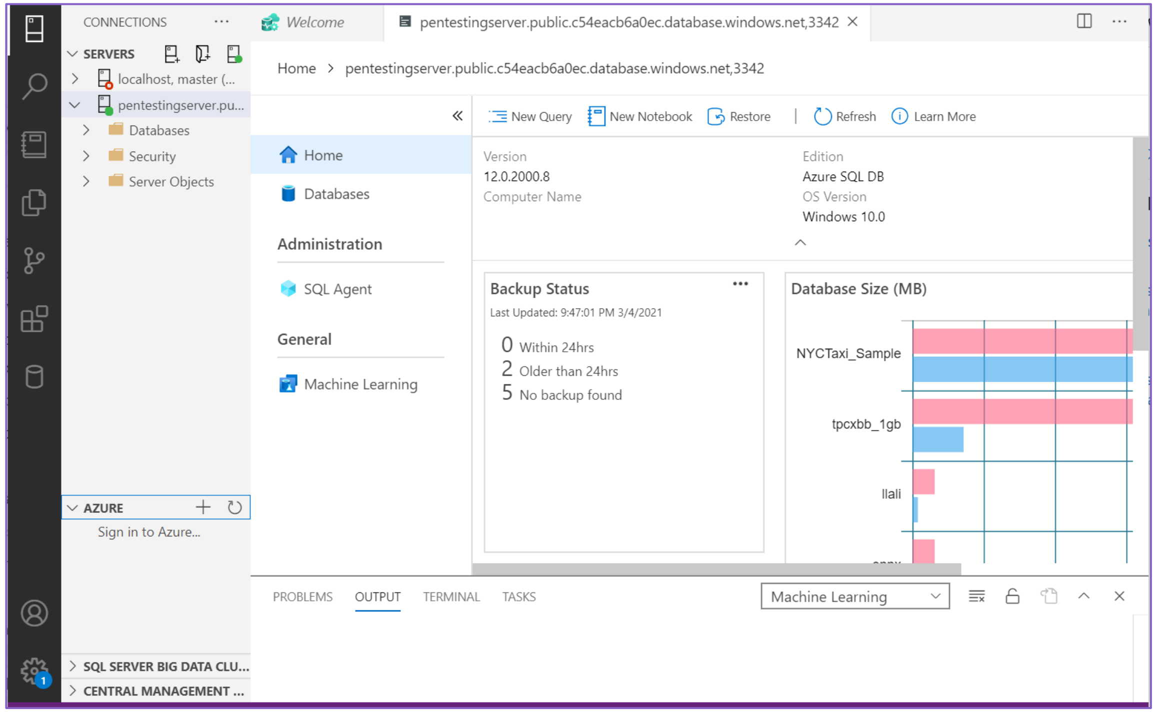
Total execution time: 00:00:18.430

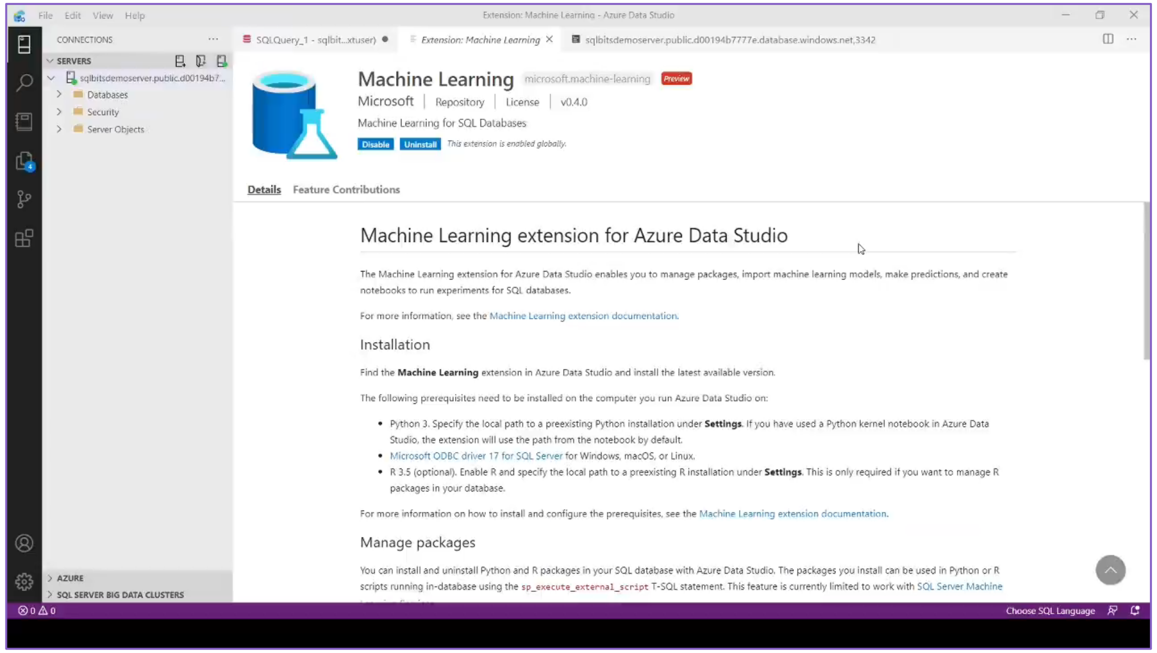
Out[3]:

Score
0.508259735691279

Machine Learning on Azure SQL Managed Instance GA – Getting Started

- Managed Instance subscription
- [New support request - Microsoft Azure]()
- Machine Learning Extension on Azure Data Studio
- [Machine Learning Services in Azure SQL Managed Instance - Azure SQL Managed Instance | Microsoft Docs]()





Microsoft Azure Solutions Business

AI dans Azure SQL @Edge

Frédéric Gisbert

Introduction

Data and compute pushing to the edge

By 2025, 75% of enterprise-generated data will be created and processed at the edge, up from less than 20% in 2018

80B
Connected IoT devices by 2025

180ZB
Amount of data from IoT devices by 2025

↑29%
CAGR of IoT data, 2018 - 2025

Impacting a variety of industries

Manufacturing	Health Care	Aerospace
Defense	Automotive	Financial
Energy	Retail	Smart cities

Sources: "Edge Computing Solutions for Industrial IoT", Gartner, July 2018. "Top Strategic IoT Trends and Technologies Through 2023", Gartner, September 2018

Handling data and compute at the edge presents new challenges

Keeping latency low and working with bandwidth constraints

Managing large quantities and various types of data

Enforcing security and maintaining compliance with regulations

Ensuring business continuity when connectivity isn't an option

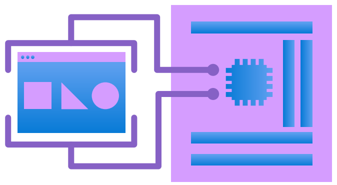
Wrangling complex architectures and combining numerous solutions

Accelerate your edge analytics by enabling a flexible hybrid strategy

Keeping latency low and working with bandwidth constraints	Process data at the edge and send only the data you need, optimizing reaction time
Managing large quantities and various types of data	Apply in-database machine learning to time series and streaming data
Enforcing security and maintaining compliance with regulations	Deploy and update from Azure for consistent security and turnkey management. Use responsible ML to ensure data privacy.
Ensuring business continuity when connectivity isn’t an option	Train models in the cloud and run at the edge in connected, disconnected, and semi-connected scenarios.
Wrangling complex architectures and combining numerous solutions	Rapidly build and deploy machine learning models regardless of skill level with automated ML.

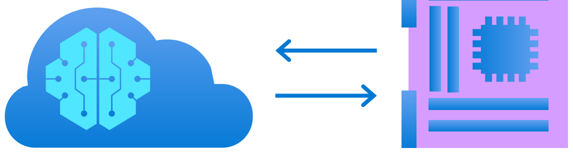
Connected **Semi-connected** **Disconnected**

Run real-time analytics at the edge with edge to cloud natively integrated machine learning



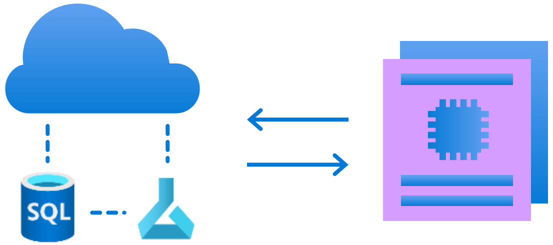
Run your applications at the edge and close to the data

Deploy the **Azure SQL Edge** container directly to IoT devices for real-time insights at the edge where data is created and collected. Stream, store, and analyze data natively whether online or off, sending only the data you need to on-prem data centers or to the cloud for further processing or storage.



Create accurate models quickly and easily

Make your edge devices even smarter. Upload datasets or subsets to the cloud to build, train and retrain your models with **Azure Machine Learning**. Prepare data quickly by automating iterative labeling tasks. Use automated ML to make it easy. Understand models with interpretability and fairness, protect data with differential privacy and confidential computing, and control the ML lifecycle with responsible ML.



Integrate with other Azure services and open-source frameworks.

Azure Machine Learning and **Azure SQL Edge** natively integrate with other Azure services enabling accelerated productivity. **Azure Machine Learning** offers best-in-class support for open-source frameworks and languages including MLflow, Kubeflow, TensorFlow, Python, and R, allowing you to use your language of choice for training before deploying to the edge with ONNX.

Better together: Azure Machine Learning and Azure SQL Edge

Time series, data streaming, and AI

Native data movement to Azure

Unparalleled performance and security

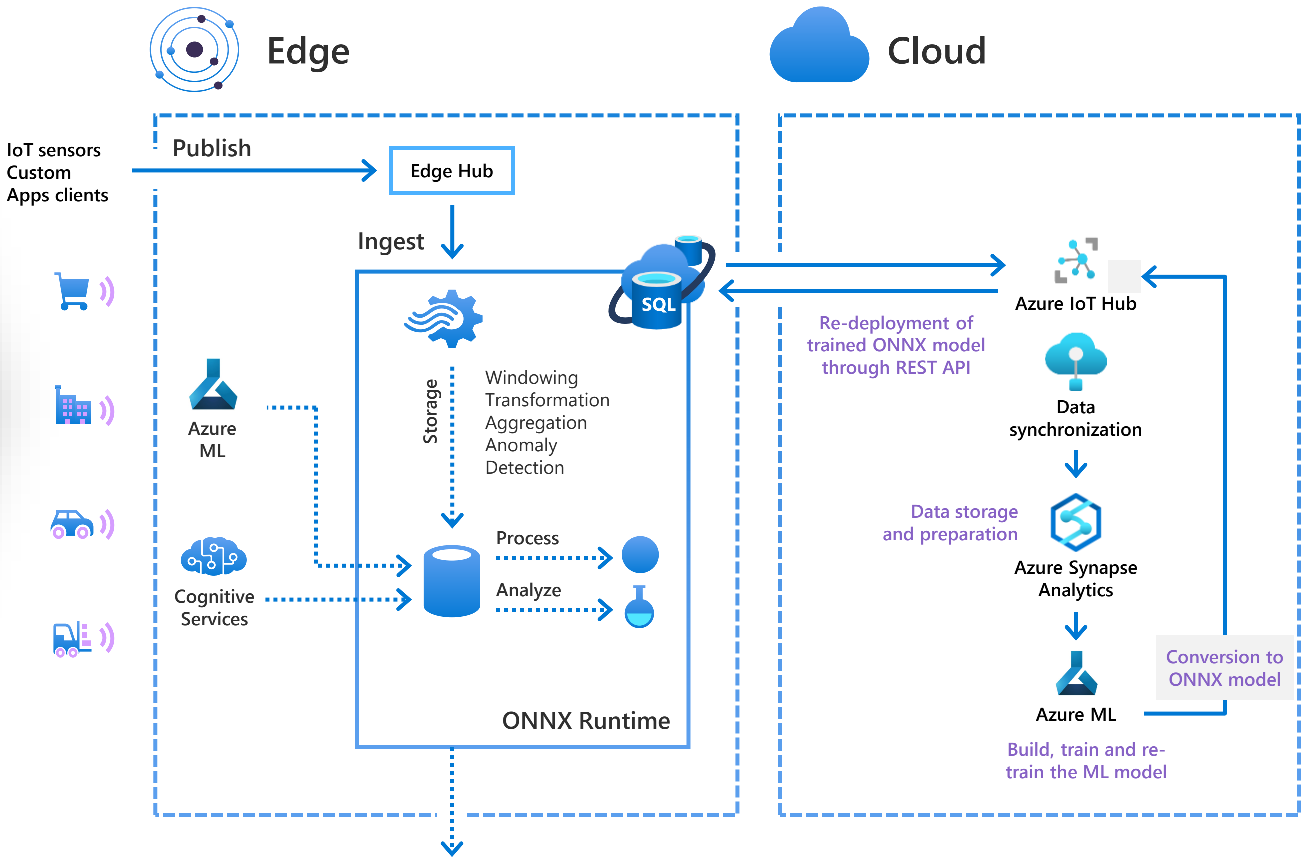
Microsoft SQL code base

Automated ML

State-of-the-art responsible ML

Open and interoperable framework

Integration with Azure services



Native integration means seamless re-training of machine learning modules running on the edge

Industry edge cases

Healthcare

Track assets to maximize up-time of critical equipment

Analyze imagery data for anomaly detection in patients

Estimate duration of stay with predictive analysis

Manufacturing

Detect and remove defects faster

Identify maintenance needs before they cause downtime

Improve employee safety

Predict human/machine failure on plant floor

Enhance workflow, asset tracking, and inventory processes

Retail

Reduce revenues loss from unstocked shelves

Optimize inventory management

Increase customer satisfaction

Improve employee productivity

Enable new revenue streams

Energy

Improve efficiency of oil and gas exploration

Detect well or pipeline anomalies before they impede work

Improve site safety

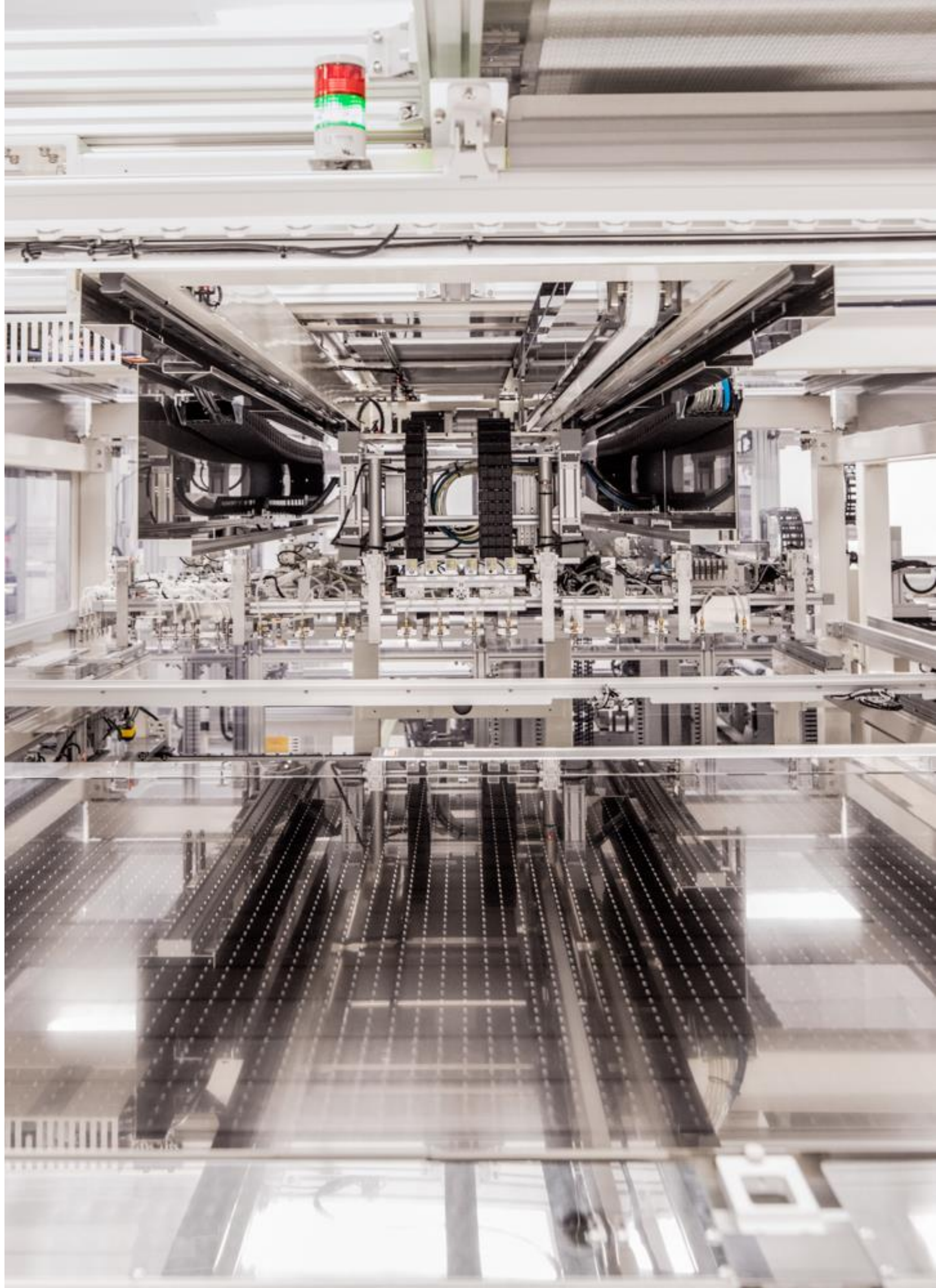
Government/ Public Sector

Enable real-time insights during field operations

Improve safety and security in public space

Track health and location of all assets

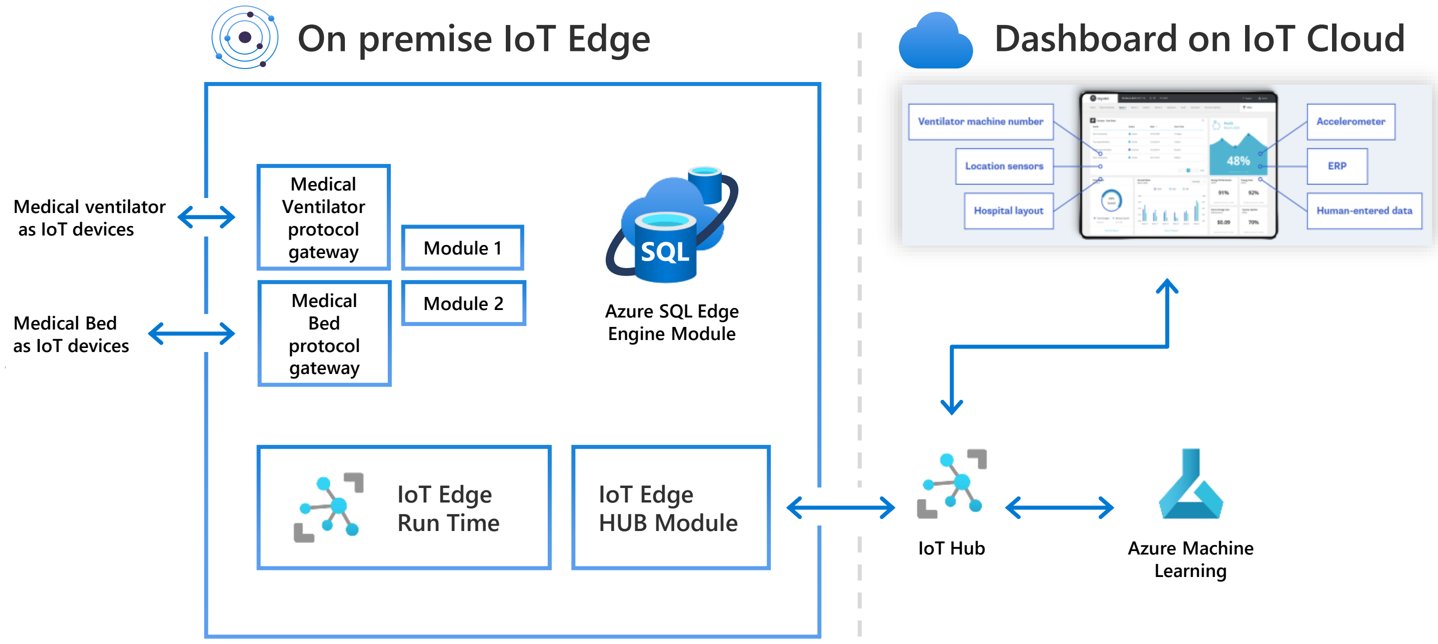
Analyze drone and satellite data



Healthcare use case

Healthcare

- Gain greater insights by facilitating data movement whilst ensuring patient data privacy and security
- Analyze patient data at the edge to overcome latency in life-threatening situations and provide real-time alerts
- Increase asset utilization by tracking and monitoring vital hospital equipment with machine learning applied while in connected or disconnected states
- Continually improve care by retraining models in the cloud and re-deploying on the edge



Only pay for what you use with edge to cloud integrated machine learning

Simply order Azure SQL Edge from the Azure portal in a pay-as-you-go model, paid monthly, or yearly via your Azure subscription.

Train models in the cloud with Azure Machine Learning and only pay for the Azure resources used to train your models. Once trained and deployed back on Azure SQL Edge, there are no ongoing costs associated with containerized models created in Azure Machine Learning.

	Per month
Azure SQL Edge	$10 per IoT device[1]
Azure Machine Learning	Consumed Azure resources Only (No Azure Machine Learning fee for training/inference)

1. See pricing page at aka.ms/sqledge for additional offers and details.
2. See Azure.com for offers and details.

Real-time insights powered by the cloud and run at the edge

Ready for faster insights at the edge? Learn more about Azure SQL Edge and Azure Machine Learning

Follow us at @AzureSQL and @Azure

Azure SQL Edge

Learn about at aka.ms/sqledge

Try out at aka.ms/sqledge-azuremarketplace

Customer stories at aka.ms/sqledge-customers

Documentation at aka.ms/sqledge-docs

See Azure SQL Edge pricing

Azure Machine Learning

Learn about at aka.ms/azureml

Try out at aka.ms/mltrial

Customer stories at customers.microsoft.com

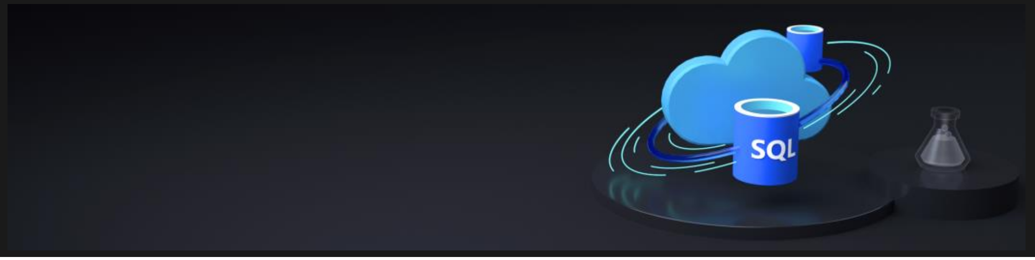
Documentation at docs.microsoft.com

See Azure Machine Learning pricing

- No Upfront Costs
- No Termination fee
- Pay only for what you use

Azure SQL Edge

Edge optimized IoT Database with built-in streaming, storage and AI



Data streaming built-in

Easy-to-use, low-latency, real-time analytics with Azure Stream Analytics

Time-series built-in

Stream, store, and analyze data using time-windowing, aggregation, & filtering

Native data movement to Azure

Consistent app development & management from cloud to data center to edge.

ML & analytics built-in

Detect anomalies and apply business logic using the built-in ML capabilities.

#1 DB in performance & security

Flexible high availability & disaster recovery. Industry-leading data protection & security tools.

Azure SQL

DEVELOP ONCE

Consistent application development and management experience

SQL Server

DEPLOY ANYWHERE

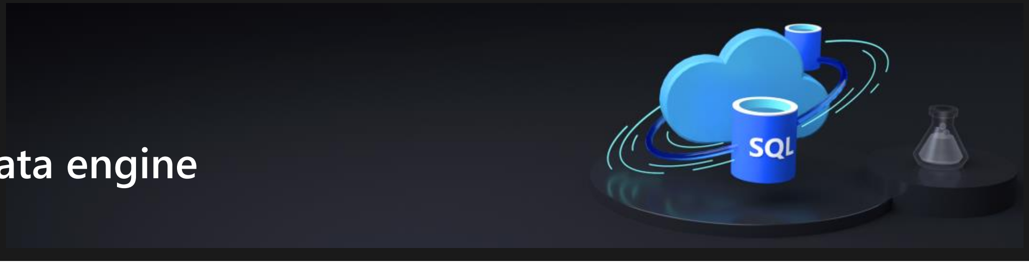
Simplify your architecture from ground to cloud to edge

Azure SQL Edge

Product Page | Case Study: ZEISS | Demo | Documentation

Azure SQL Edge

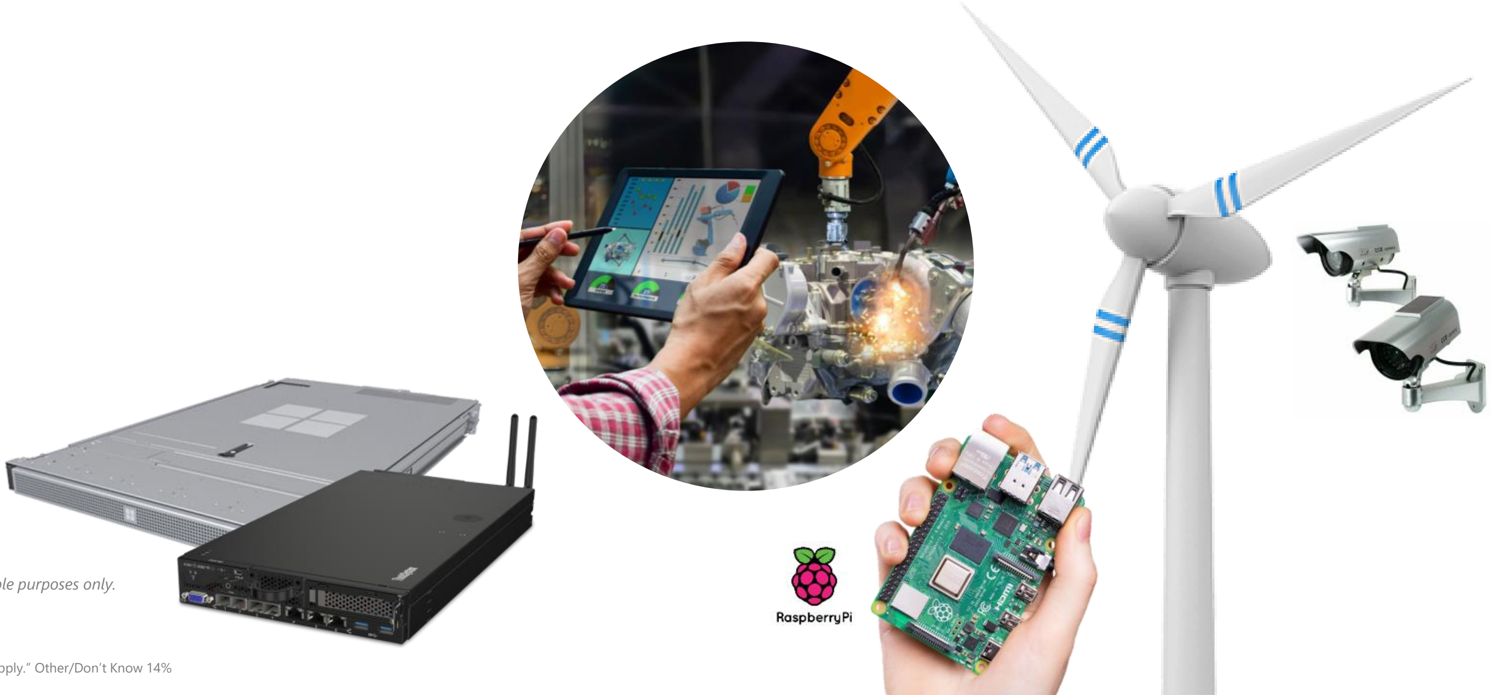
Meeting the demands of IoT edge data & compute - with the world's #1 data engine



Intelligent Cloud	Data Centers	Edge Servers 18%*	Edge Gateways 36%*	IoT Devices 32%*	Intelligent Edge
Azure SQL ⟷	SQL Server ⟷	**Upper Boundary** 16 core (3p HW) No limit (1p HW)		**Lower Boundary** 1 core 1 GB memory	Azure Sphere

Develop once, deploy anywhere.

- Extends the #1 data engine to IoT edge & ARM devices
- Built, optimized, & priced specifically for IoT
- No upskilling or retraining needed for SQL customers
- Connected, disconnected, & semi-connected scenarios



All images for example purposes only.

SQL Database Edge Early Adopter Program Survey, 1/9/20, "What device types will you use for the Azure SQL DB Edge solution. Please select all device types that apply." Other/Don't Know 14%

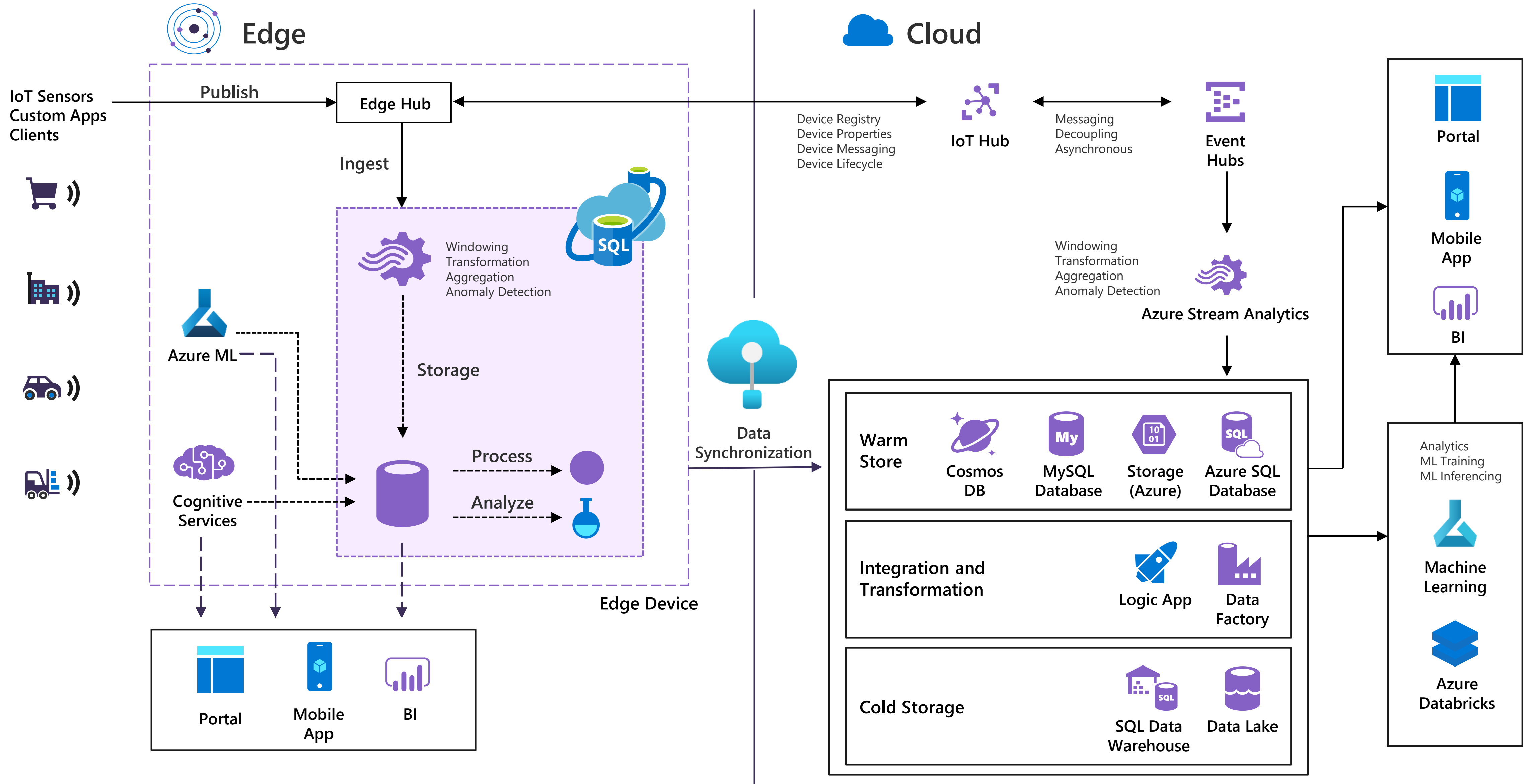
Azure SQL Edge

Feature Breakdown

Applied from SQL Server 2019 on Linux	Security	Row-Level Security
		Always Encrypted
		Transparent Database Encryption (TDE)
		Data Classification and auditing built-in
	Performance	Intelligent Database: Intelligent Query Processing features
		UTF-8 Support: Support GB18030 Char set
		Column Store Indexes and Compression
		Table and Index Partitioning
		Data Compression
		Accelerated Database Recovery
	Change Feed	Change Data Capture
		Change Tracking
NEW IoT edge features	Data Streaming	Data Streaming from IoT Edge and Kafka
		Stream Processing with Reference Data
	Time Series	Date_Bucket()
		Date_Bucket_GapFill()
		GapFill()
		Data Retention and Purging
	Machine Learning	Data Inferencing with ONNX and T-SQL Predict() Function

Azure SQL Edge

Reference Architecture – Connected Scenario



Implémentation

ML Implementation

- Création d'un workspace Azure ML
- Installation des prérequis
- Définition du compute

Resource properties

Status
Running

Virtual machine size
Standard_DS12_v2 (4 cores, 28 GB RAM, 56 GB disk)

Processing unit
CPU - Memory optimized

Applications
JupyterLab Jupyter VS Code RStudio Terminal

Created on
6/2/2021, 6:02:51 PM

SSH access
Disabled

Private IP address
10.0.0.4

Virtual network/subnet
--

Public IP address
20.74.64.170

In [1]:
```
!pip install scikit-learn==0.22.1 joblib==0.14.1
```

```
Collecting scikit-learn==0.22.1
  Downloading scikit_learn-0.22.1-cp36-cp36m-manylinux1_x86_64.whl
     |████████████████████████████████| 7.0 MB 6.1 MB/s eta 0:00:0
Requirement already satisfied: joblib==0.14.1 in /anaconda/envs/a
Requirement already satisfied: scipy>=0.17.0 in /anaconda/envs/azu
0.22.1) (1.5.2)
Requirement already satisfied: numpy>=1.11.0 in /anaconda/envs/azu
0.22.1) (1.18.5)
ERROR: raiwidgets 0.2.2 has requirement lightgbm>=3.1.1, but you'
ERROR: azureml-automl-dnn-nlp 1.28.0 has requirement scikit-learn
mpatible.
ERROR: autokeras 1.0.13 has requirement tensorflow>=2.3.0, but you
Installing collected packages: scikit-learn
  Attempting uninstall: scikit-learn
    Found existing installation: scikit-learn 0.22.2.post1
    Uninstalling scikit-learn-0.22.2.post1:
      Successfully uninstalled scikit-learn-0.22.2.post1
Successfully installed scikit-learn-0.22.1
```

In [37]:
```
!pip install azureml-contrib-interpret
```

```
Installing collected packages: interpret-community, cryptography,
ret
  Attempting uninstall: interpret-community
    Found existing installation: interpret-community 0.17.2
    Uninstalling interpret-community-0.17.2:
      Successfully uninstalled interpret-community-0.17.2
  Attempting uninstall: cryptography
    Found existing installation: cryptography 3.4.7
    Uninstalling cryptography-3.4.7:
      Successfully uninstalled cryptography-3.4.7
  Attempting uninstall: azureml-core
    Found existing installation: azureml-core 1.28.0
    Uninstalling azureml-core-1.28.0:
      Successfully uninstalled azureml-core-1.28.0
  Attempting uninstall: azureml-interpret
    Found existing installation: azureml-interpret 1.28.0
    Uninstalling azureml-interpret-1.28.0:
      Successfully uninstalled azureml-interpret-1.28.0
Successfully installed azureml-contrib-interpret-1.23.0 azureml-c
erpret-community-0.16.0
```

Experiment creation

In our script, there are three distinct sections:

1. Setting up the scikit-learn logistic regression model pipeline (including encoding our features).
2. Analyzing and logging the results of the model training.
3. Running the model explainer to understand the key model drivers.

In [17]:

```
%%writefile $project_folder/train.py

from azureml.core import Run

from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler

from utils import *

# Fetch current run
run = Run.get_context()

# Fetch dataset from the run by name
dataset = run.input_datasets['training']

# Convert dataset to Pandas data frame
X_train, X_test, y_train, y_test = split_dataset(dataset)

# Setup scikit-learn pipeline
numeric_transformer = Pipeline(steps=[('scaler', StandardScaler())])
preprocessor = ColumnTransformer(transformers=[('num', numeric_transformer, list(X_train.columns.values))])

clf = Pipeline(steps=[('preprocessor', preprocessor),
                      ('classifier', LogisticRegression())])

model = clf.fit(X_train, y_train)

# Analyze model performance
analyze_model(clf, X_test, y_test)

# Save model
model_id = save_model(clf)
```

```
Overwriting ./scripts/train.py
```

Workspace experiment creation

The Experiment constructor allows to create an experiment instance. The constructor takes in the current workspace, which is fetched by calling `Workspace.from_config()` and an experiment name.

For more information on **Experiment**, please visit: [Microsoft Experiment Documentation]()

In [18]:

```
from azureml.core.experiment import Experiment

# Get an instance of the Workspace from the config file
ws = Workspace.from_config()

experiment_name = 'wake-detection-experiment'

# Create Experiment
experiment = Experiment(ws, experiment_name)
```

Create Automated ML Compute cluster

Firstly, check for the existence of the cluster. If it already exists, we are able to reuse it. Checking for the existence of the cluster can be performed by calling the constructor `ComputeTarget()` with the current workspace and name of the cluster.

In case the cluster does not exist, the next step will be to provide a configuration for the new AML cluster by calling the function `AmlCompute.provisioning_configuration()`. It takes as parameters the VM size and the max number of nodes that the cluster can scale up to. After the configuration has executed, `ComputeTarget.create()` should be called with the previously configuration object and the workspace object.

For more information on **ComputeTarget**, please visit: [Microsoft get_data Documentation]

For more information on **AmlCompute**, please visit: [Microsoft get_data Documentation]

Note: Please wait for the execution of the cell to finish before moving forward.

In [19]:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException

# Create AML CPU Compute Cluster
try:
    compute_target = ComputeTarget(workspace=ws, name='cpucluster')
    print('Found existing compute target.')
except ComputeTargetException:
    print('Creating a new compute target...')
    compute_config = AmlCompute.provisioning_configuration(vm_size='Standard_DS12_v2',
                                                           max_nodes=4)

    compute_target = ComputeTarget.create(ws, 'cpucluster', compute_config)
    compute_target.wait_for_completion(show_output=True)
```

```
Creating a new compute target...
Creating.........
SucceededProvisioning operation finished, operation "Succeeded"
Succeeded
AmlCompute wait for completion finished

Minimum number of nodes requested have been provisioned
```

We'll use remote compute for this job. We need to install a couple of extra libraries, including those required for model interpretability.

The `experiment.submit()` function is called to send the experiment for execution. The only parameter received by this function is the `Estimator` object.

Submit and monitor experiment

In [20]:

```
from azureml.train.sklearn import SKLearn

estimator = SKLearn(source_directory=project_folder,
                    compute_target=compute_target,
                    entry_script='train.py',
                    inputs=[tabular.as_named_input('training')],
                    pip_packages=['azureml-dataprep[fuse,pandas]','joblib==0.14.1','azureml-interpret','azureml-contrib-inter

run = experiment.submit(estimator)
run
```

```
'SKLearn' estimator is deprecated. Please use 'ScriptRunConfig' from 'azureml.core.script_run_config' with your own defined
environment or the AzureML-Tutorial curated environment.
'enabled' is deprecated. Please use the azureml.core.runconfig.DockerConfiguration object with the 'use_docker' param instea
d.
You have specified to install packages in your run. Note that you have overridden Azure ML's installation of the following p
ackages: ['joblib', 'scikit-learn']. We cannot guarantee image build will succeed.
WARNING:root:If 'script' has been provided here and a script file name has been specified in 'run_config', 'script' provided
in ScriptRunConfig initialization will take precedence.
```

Out[20]:

Experiment	Id	Type	Status	Details Page	Docs Page
wake-detection-experiment	wake-detection-experiment_1622652133_96e7e592	azureml.scriptrun	Preparing	Link to Azure Machine Learning studio	Link to Documentation

Monitor Experiment

The creation of an object of type `Run` will enable us to observe the experiment's progress and results. The object is created by calling the constructor `Run()`. It takes, as arguments, the experiment and the identifier of the run to fetch. After the object has been instantiated, the `RunDetails()` function will retrieve the progress, metrics, and tasks for the specified run. They will be displayed by calling the function `show()` over the mentioned object.

Note: Please wait for the execution of the cell to finish before moving forward. (Status should be **Completed**)

In [21]:

```
from azureml.core import Run
from azureml.widgets import RunDetails

run = Run(experiment, run.id)
RunDetails(run).show()
```

```
_UserRunWidget(widget_settings={'childWidgetDisplay': 'popup', 'send_telemetry': False, 'log_level': 'INFO', '…
```

Encode, download, convert and save

Encode dataset and download trained model

First step is to encode our training data to take the shape expected by the Onnx converter. Next, download the model obtained from the best run. In order to download the model, the function `download_model()` should be called. This will take care of downloading the model obtained from the best run.

In [49]:
```
from utils import *
from scripts.utils import *

# Convert dataset to Pandas data frame
X_train, X_test, y_train, y_test = split_dataset(tabular)
model = download_model(run)
```

Convert model to Onnx format

Export the Sklearn model to Onnx format by using `skl2onnx`. This step will output an Onnx model that we will be able to publish to the Azure SQL Edge Database Instance to use along with our `PREDICT` statement.

In [50]:
```
import skl2onnx
import onnxmltools

# Convert the scikit model to onnx format
onnx_model = skl2onnx.convert_sklearn(model, 'Wind Turbine Dataset', convert_dataframe_schema(X_train))
# Save the onnx model locally
onnx_model_path = 'windturbinewake.model.onnx'
onnxmltools.utils.save_model(onnx_model, onnx_model_path)
```

```
WARNING:skl2onnx:The maximum opset needed by this model is only 1.
WARNING:skl2onnx:The maximum opset needed by this model is only 11.
```

Save model to Azure Blob Storage

Let's save our Onnx model to the default workspace Datastore.

In [51]:
```
datastore.upload_files(files=[onnx_model_path],
                       overwrite=True,
                       show_progress=True)
```

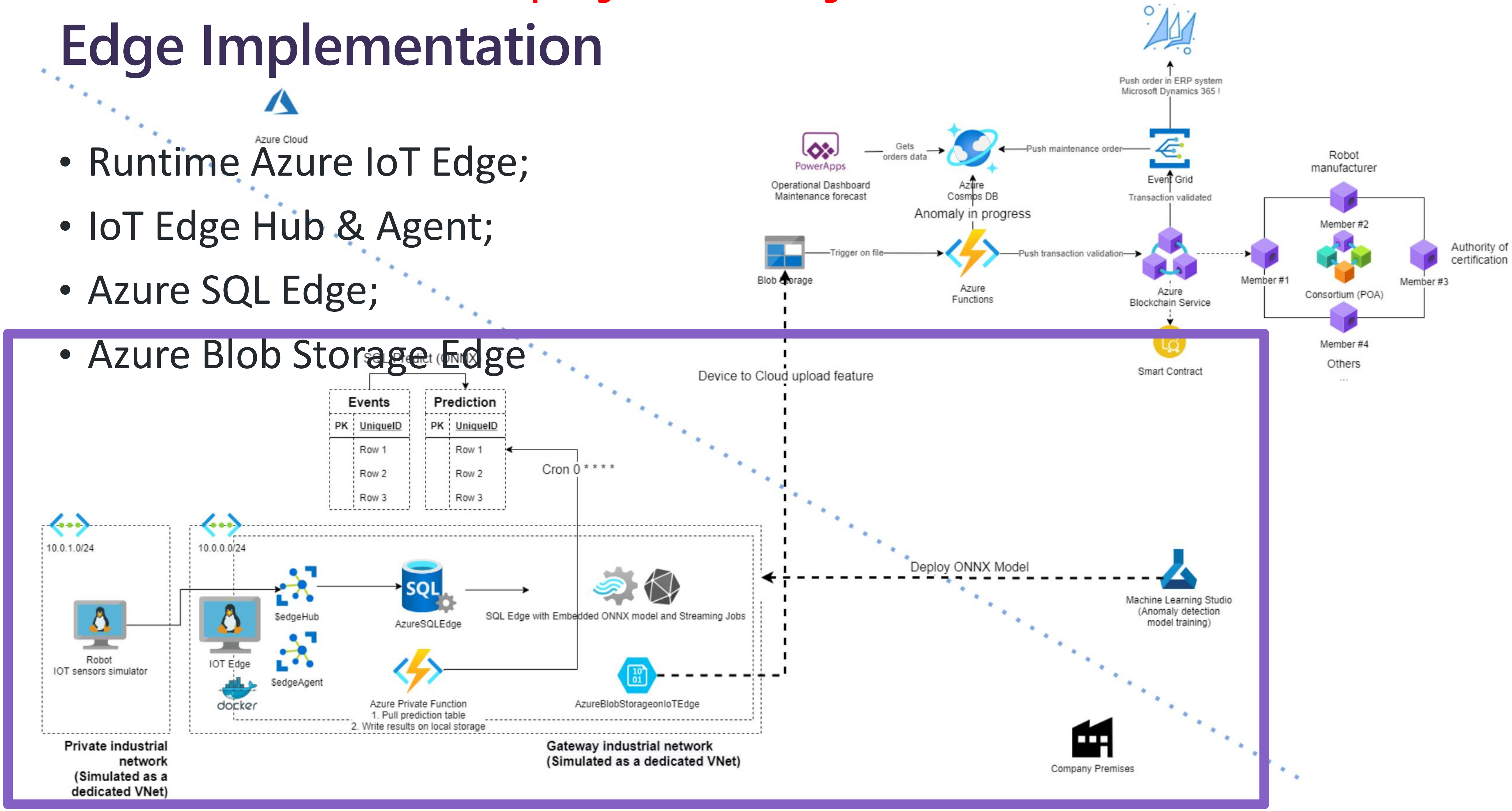
```
Uploading an estimated of 1 files
Uploading windturbinewake.model.onnx
Uploaded windturbinewake.model.onnx, 1 files out of an estimated total of 1
Uploaded 1 files
```

Out[51]:
```
$AZUREML_DATAREFERENCE_workspaceblobstore
```

https://github.com/fredgis/AIRobot

Edge Implementation

- Runtime Azure IoT Edge;
- IoT Edge Hub & Agent;
- Azure SQL Edge;
- Azure Blob Storage Edge



SQL Edge module

```
--CREATE [dbo].[Events] TABLE
CREATE TABLE [dbo].[Events](
        [ID] [bigint] IDENTITY(1,1) NOT NULL,
        [Timestamp] [bigint] NOT NULL,
        [DrillingTemperature] [decimal](9, 5) NULL,
        [DrillBitFriction] [decimal](9, 5) NULL,
        [DrillingSpeed] [decimal](9, 5) NULL,
        [LiquidCoolingTemperature] [decimal](9, 5) NULL,
 CONSTRAINT [PK_Events] PRIMARY KEY CLUSTERED
(
        [ID] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_L
) ON [PRIMARY]
GO

-- CREATE [dbo].[Models] TABLE
CREATE TABLE Models (
        [ID] [int] IDENTITY(1,1) NOT NULL,
        [Data] [varbinary](MAX) NULL,
        [Description] varchar(1000))
GO
```

SQL Edge - Streaming Job - Input

```sql
--Create an external file format of the type JSON.
CREATE EXTERNAL FILE FORMAT InputFileFormat
WITH
(
   format_type = JSON
)
GO

--Create an external data source for Azure IoT Edge hub
CREATE EXTERNAL DATA SOURCE EdgeHubInput
WITH
(
    LOCATION = 'edgehub://'
)
GO

--Create the external stream object for Azure IoT Edge hub.
CREATE EXTERNAL STREAM RobotSensors
WITH
(
    DATA_SOURCE = EdgeHubInput,
    FILE_FORMAT = InputFileFormat,
          LOCATION = N'RobotSensors',
    INPUT_OPTIONS = N'',
    OUTPUT_OPTIONS = N''
);
GO
```

SQL Edge - Streaming Job - Output

```
--Create the external stream object for local SQL Edge database.
CREATE DATABASE SCOPED CREDENTIAL SQLCredential
WITH IDENTITY = 'edgejob', SECRET = [redacted]
GO

CREATE EXTERNAL DATA SOURCE LocalSQLOutput
WITH
(
    LOCATION = 'sqlserver://tcp:.,1433',
    CREDENTIAL = SQLCredential
)
GO

CREATE EXTERNAL STREAM EventsTableOutput
WITH
(
    DATA_SOURCE = LocalSQLOutput,
    LOCATION = N'airobotedgedb.dbo.Events',
    INPUT_OPTIONS = N'',
    OUTPUT_OPTIONS = N''
);
GO
```

SQL Edge - Streaming Job - Start

Une fois l'input et l'ouput définit, le STREAM job peut être créé et démarré.

```
--Create the streaming job and start it.
EXEC sys.sp_create_streaming_job @name=N'StreamingJob1',
        @statement= N'SELECT [Timestamp],
                                                 [drillingTemperature] AS [DrillingTemperature],
                                                 [drillBitFriction] AS [DrillBitFriction],
                                                 [drillingSpeed] AS [DrillingSpeed],
                                                 [liquidCoolingTemperature] AS [LiquidCoolingTemperature]
                                  INTO [EventsTableOutput]
                                  FROM [RobotSensors]'

exec sys.sp_start_streaming_job @name=N'StreamingJob1'
GO
```

> **Notes:** L'ensemble du script SQL est disponible dans le répertoire Src/SQLEdge de ce repo.

SQL Edge - ONNX

Déploiement du modèle ONNX dans SQLEdge

Si ce n'est pas déjà je cas, se connecter à la VM simulant la gateway IoT Edge via le service `Azure Bastion` ou autre méthode de votre choix

Prendre la main sur le conteneur `SQL Edge` :

```
sudo docker exec -it AzureSQLEdge bash
```

Télécharger les deux modèles ONNX présents dans ce repo via la méthode de votre choix. Ici, par exemple, nous utilisons l'utilitaire `wget` .

> **Notes:** Le téléchargement d'éléments depuis Internet peut nécessiter l'ouverture de flux adéquats dans le Network Security Group `nsg-vnet-airobot-edge` .

```
cd /var/opt/mssql
wget https://raw.githubusercontent.com/fredgis/AIRobot/main/Src/Models/model_final.onnx
wget https://raw.githubusercontent.com/fredgis/AIRobot/main/Src/Models/pipeline_std.onnx
```

Une fois téléchargés, il ne reste plus qu'à les insérer dans la table `dbo.Models` .

Se connecter à l'instance SQL Edge via `sqlcmd` :

```
/opt/mssql-tools/bin/sqlcmd -S localhost -U sa -P <SQL_PASSWORD>
```

Insérer les modèles:

```
INSERT INTO dbo.Models ([Description], [Data]) SELECT N'model_final.onnx', * FROM OPENROWSET(BULK N'/var/opt/m

INSERT INTO dbo.Models ([Data]) SELECT N'pipeline_std.onnx', * FROM OPENROWSET(BULK N'/var/opt/mssql/pipeline_
```

Preview and What's next ?

Azure Machine Learning as an IOT Edge Module

Tutorial: Deploy Azure Machine Learning as an IoT Edge module

07/29/2020 • 7 minutes to read • +21

Applies to: IoT Edge 1.1 IoT Edge 1.2

Use Azure Notebooks to develop a machine learning module and deploy it to a device running Azure IoT Edge with Linux containers. You can use IoT Edge modules to deploy code that implements your business logic directly to your IoT Edge devices. This tutorial walks you through deploying an Azure Machine Learning module that predicts when a device fails based on simulated machine temperature data. For more information about Azure Machine Learning on IoT Edge, see Azure Machine Learning documentation.

> **Note**
>
> Azure Machine Learning modules on Azure IoT Edge are in public preview.

The Azure Machine Learning module that you create in this tutorial reads the environmental data generated by your device and labels the messages as anomalous or not.

In this tutorial, you learn how to:

- ✓ Create an Azure Machine Learning module.
- ✓ Push a module container to an Azure container registry.
- ✓ Deploy an Azure Machine Learning module to your IoT Edge device.
- ✓ View generated data.

If you don't have an Azure subscription, create a free account before you begin.

https://docs.microsoft.com/en-us/azure/iot-edge/tutorial-deploy-machine-learning?view=iotedge-2020-11

Microsoft Azure Solutions Business

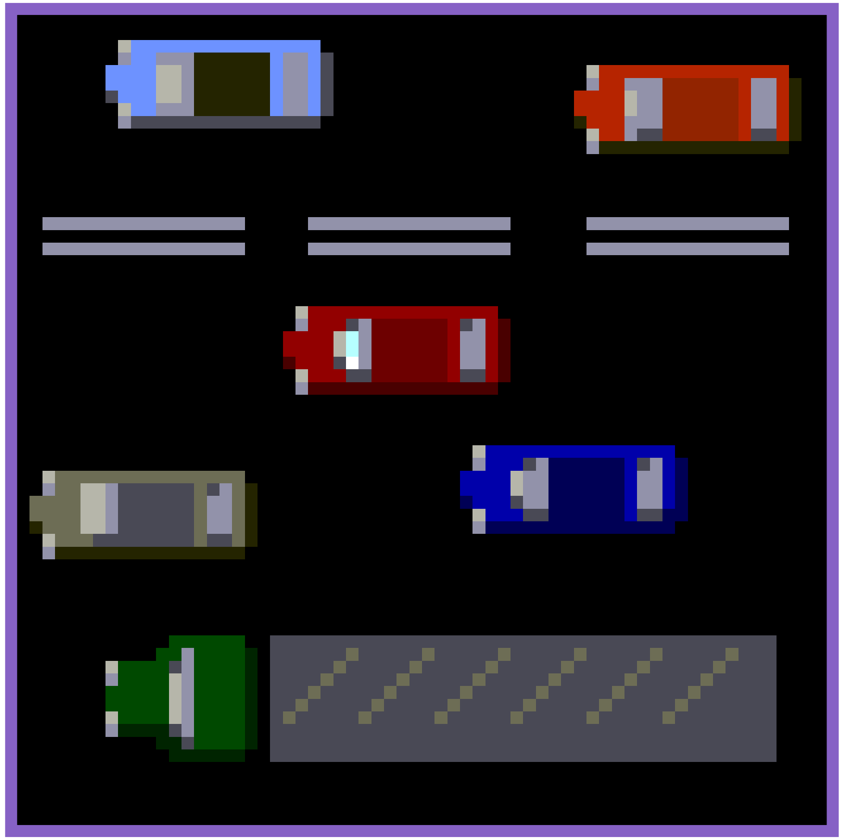
Azure Stream Analytics

Franck Gaillard

What is streaming?

Data at rest

Data in motion



What is Azure Stream Analytics?

Azure Stream Analytics is a real-time analytics and complex event-processing engine that is designed to analyze and process high volumes of fast streaming data from multiple sources simultaneously.

Patterns and relationships can be identified in information extracted from several input sources including devices, sensors, clickstreams, social media feeds, and applications.

These patterns can be used to trigger actions and initiate workflows such as creating alerts, feeding information to a reporting tool, or storing transformed data for later use.

Also, Stream Analytics is available on Azure IoT Edge runtime, enabling to process data on IoT devices.

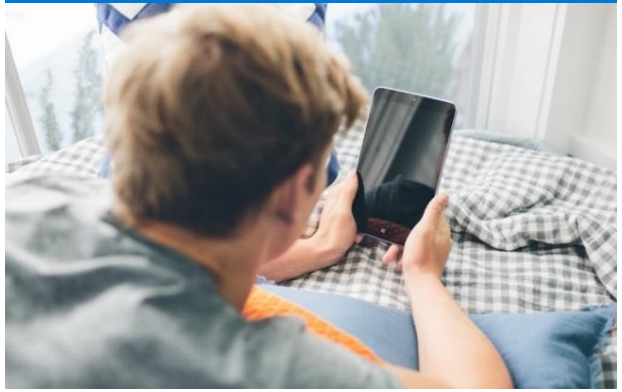
The need for real-time processing

Real-time
fraud detection

Connected car

Click-stream analysis

Remote device
monitoring



Smart grid

CRM alerting sales

Predictive
maintenance

Real-time financial
sales tracking

How does Stream Analytics work?

An Azure Stream Analytics job consists of an input, query, and an output.

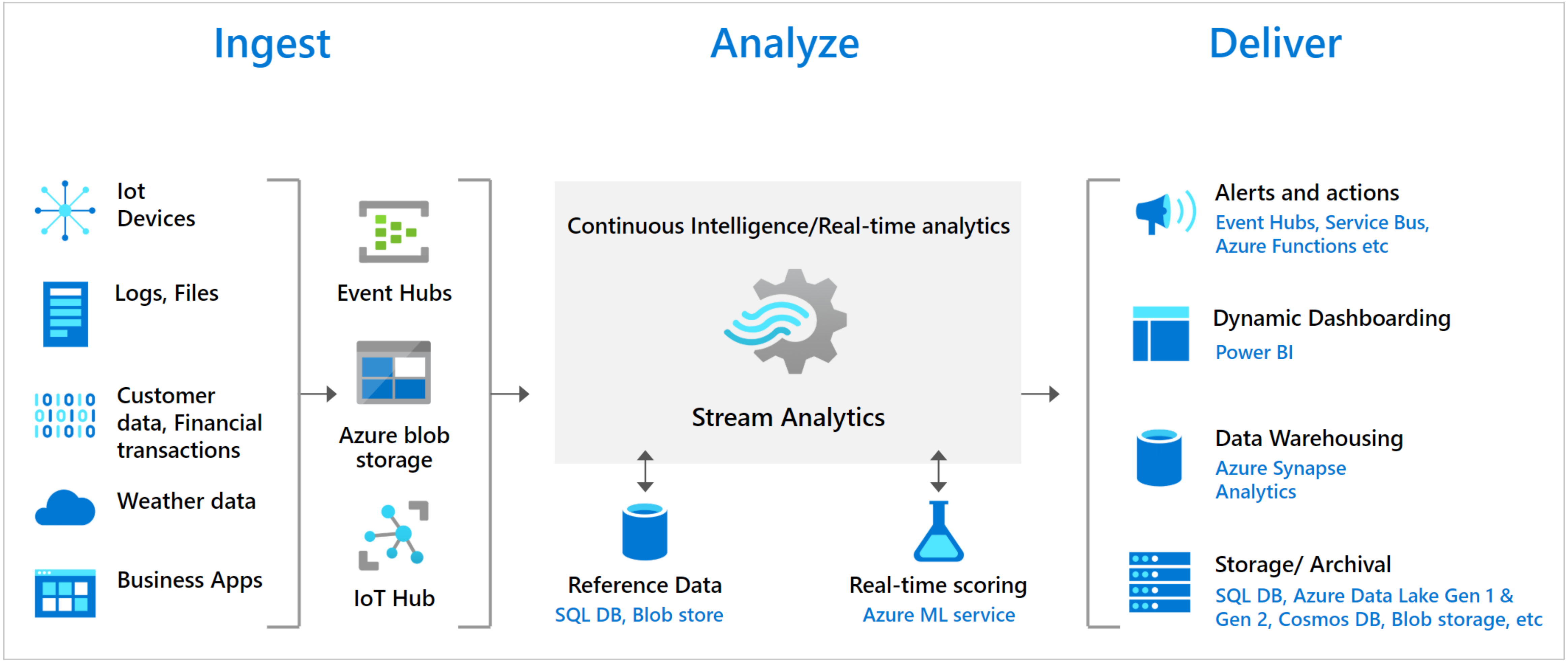
Stream Analytics ingests data from Azure Event Hubs (including Azure Event Hubs from Apache Kafka), Azure IoT Hub, or Azure Blob Storage.

The query, which is based on SQL query language, can be used to easily filter, sort, aggregate, and join streaming data over a period of time.

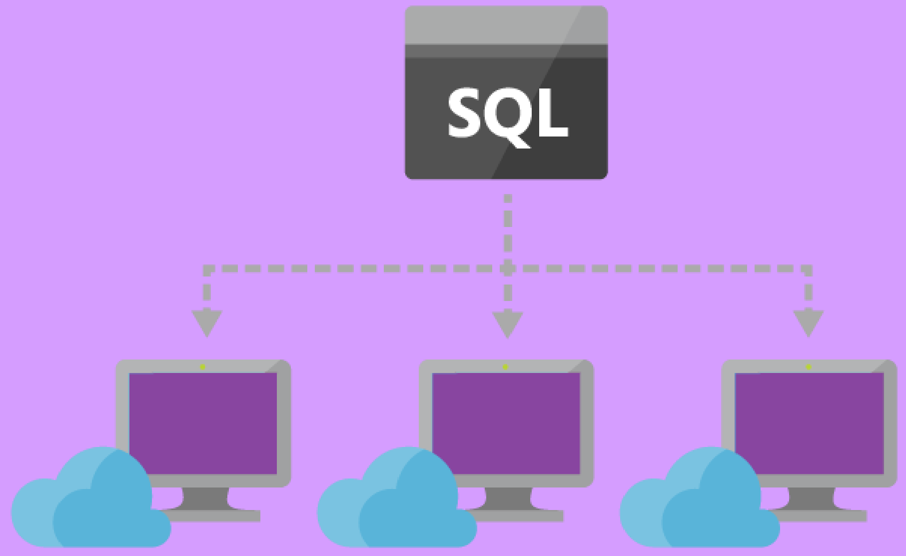
Each job has one or several outputs for the transformed data, and you can control what happens in response to the information you've analyzed. For example, you can:

- Send data to services such as Azure Functions, Service Bus Topics or Queues to trigger communications or custom workflows downstream.
- Send data to a Power BI dashboard for real-time dashboarding.
- Store data in other Azure storage services (for example, Azure Data Lake, Azure Synapse Analytics, etc.) to train a machine learning model based on historical data or perform batch analytics.

Stream Analytics Architecture



Rapid development

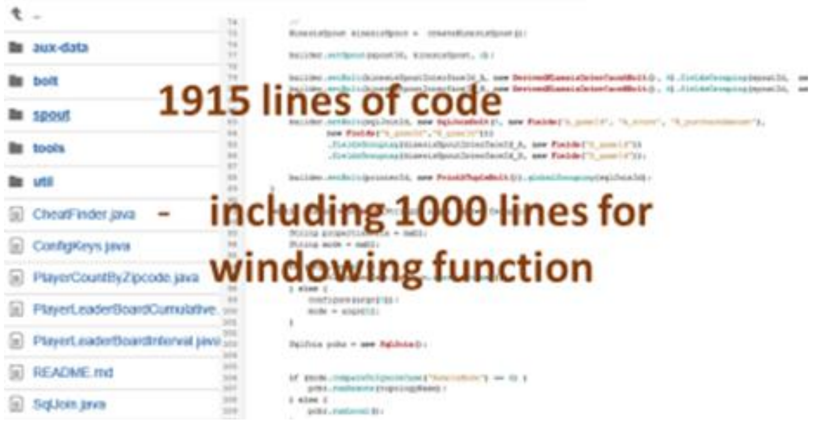


Only SQL queries needed

Developers uses declarative SQL commands

Some functions take several lines of code versus thousands from other solutions

Thousand lines of code in other solutions, such as Apache Storm

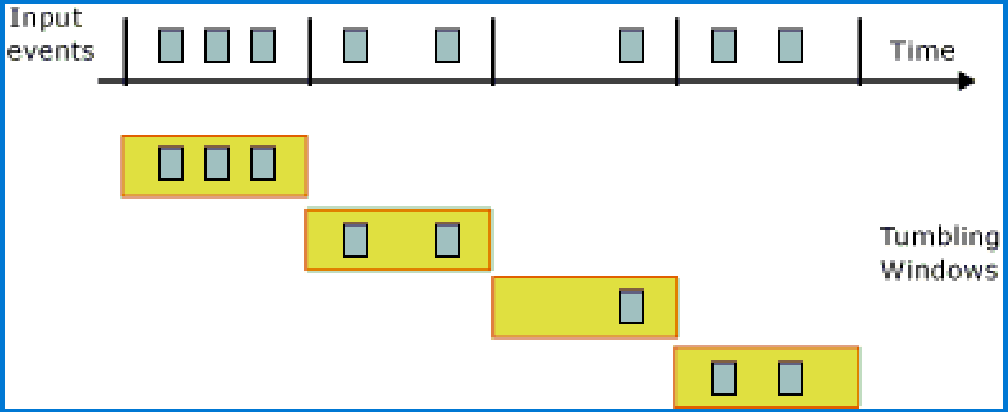


Versus

3 lines of code in Stream Analytics

```
SELECT Avg(Purchase), Score
FROM GameDataStream
GROUP BY TumblingWindow(5, Minute), Score
```

Built in temporal semantics



Implement temporal functions

Tumbling Windows

Hopping Windows

Sliding Windows

Manage out-of-order events

With configuration instead of code

Manage actions on late events

Using policy settings instead of code

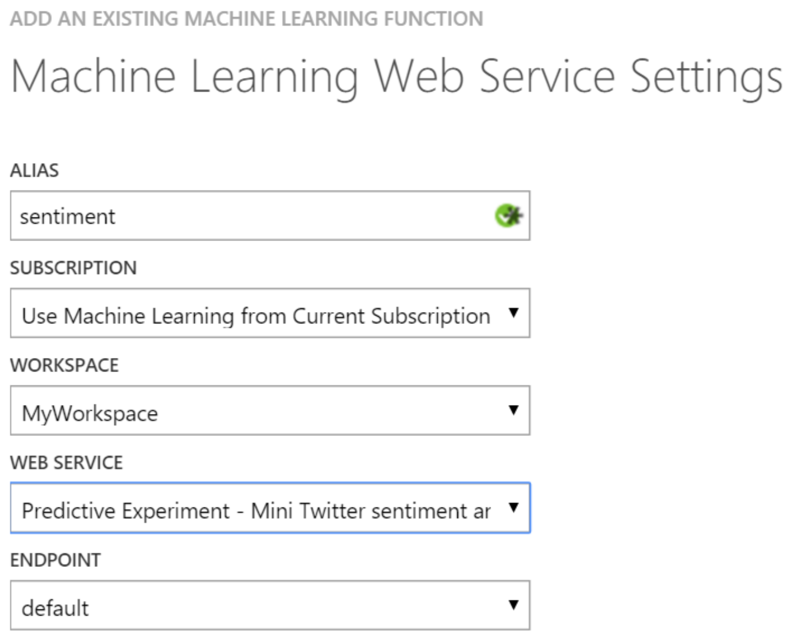
Functions
Use Azure ML web services

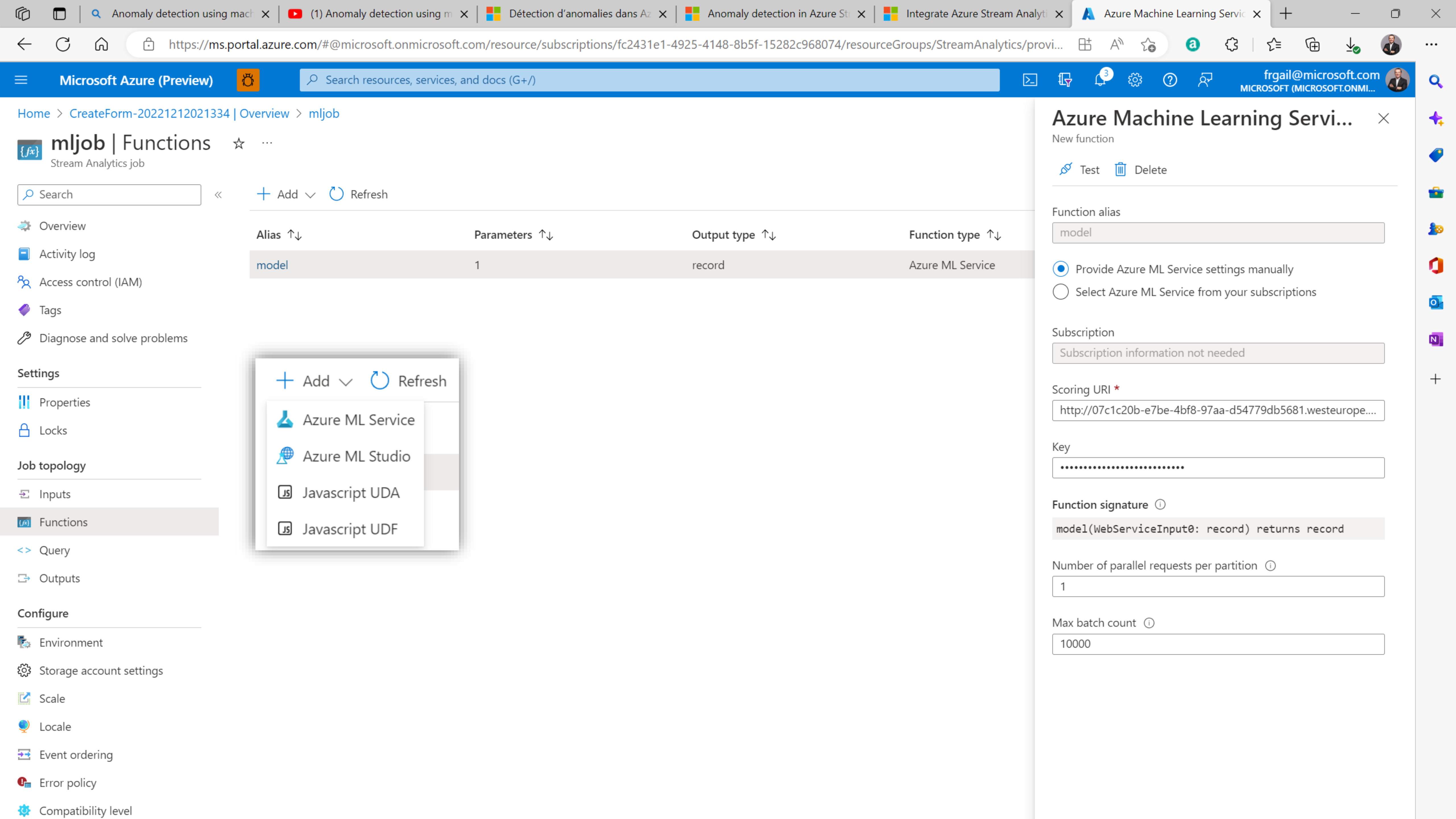
Implement
Azure ML web services in Real-Time

Includes Real-time R/Python models

Example Query

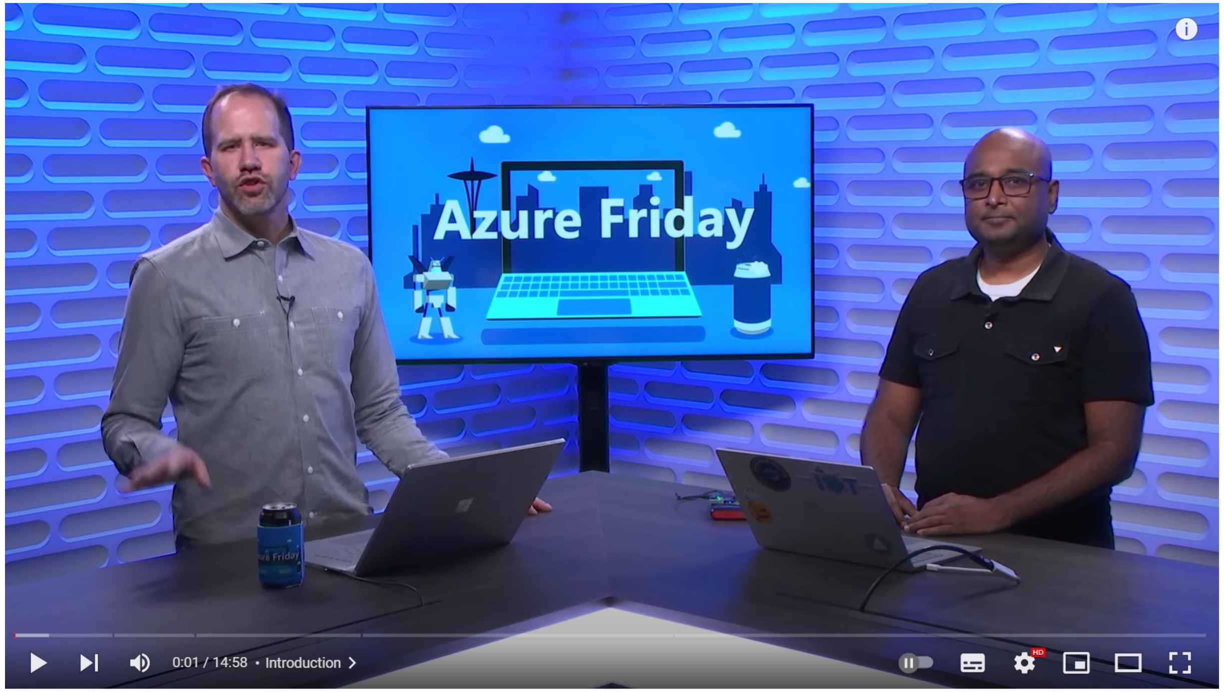
```
WITH subquery AS(SELECT text, sentiment(text) AS result FROM myinput)

SELECT text, result.[Score]
INTO myoutput
FROM subquery
```




Anomaly detection using mach
(1) Anomaly detection using m
Détection d'anomalies dans Az
Anomaly detection in Azure St
Integrate Azure Stream Analyti
Azure Machine Learning Servic
https://ms.portal.azure.com/#@microsoft.onmicrosoft.com/resource/subscriptions/fc2431e1-4925-4148-8b5f-15282c968074/resourceGroups/StreamAnalytics/provi...
Microsoft Azure (Preview)
Search resources, services, and docs (G+/)
frgail@microsoft.com
MICROSOFT (MICROSOFT.ONMI...
Home > CreateForm-20221212021334 | Overview > mljob
mljob | Functions
Stream Analytics job
Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Settings
Properties
Locks
Job topology
Inputs
Functions
Query
Outputs
Configure
Environment
Storage account settings
Scale
Locale
Event ordering
Error policy
Compatibility level
Add
Refresh
Alias
Parameters
Output type
Function type
model
1
record
Azure ML Service
Add
Refresh
Azure ML Service
Azure ML Studio
Javascript UDA
Javascript UDF
Azure Machine Learning Servi...
New function
Test
Delete
Function alias
model
Provide Azure ML Service settings manually
Select Azure ML Service from your subscriptions
Subscription
Subscription information not needed
Scoring URI *
http://07c1c20b-e7be-4bf8-97aa-d54779db5681.westeurope....
Key
Function signature
model(WebServiceInput0: record) returns record
Number of parallel requests per partition
1
Max batch count
10000

Anomaly Detection in Azure Stream Analytics

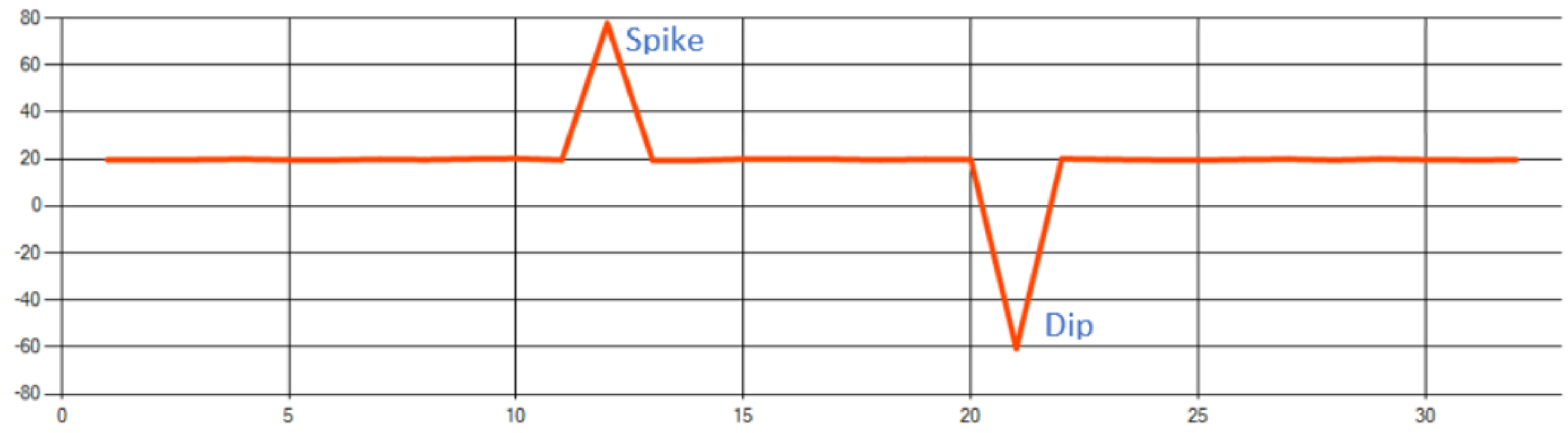
Perform anomaly detection directly in your Stream Analytics job



#machinelearning #azure

Anomaly detection using machine learning in Azure Stream Analytics | Azure Friday

Spike and dip



Examples of Spike and Dip anomalies

Temporary anomalies in a time series event stream are known as spikes and dips. Spikes and dips can be monitored using the Machine Learning based operator, [AnomalyDetection_SpikeAndDip].

Spike and dip

SQL

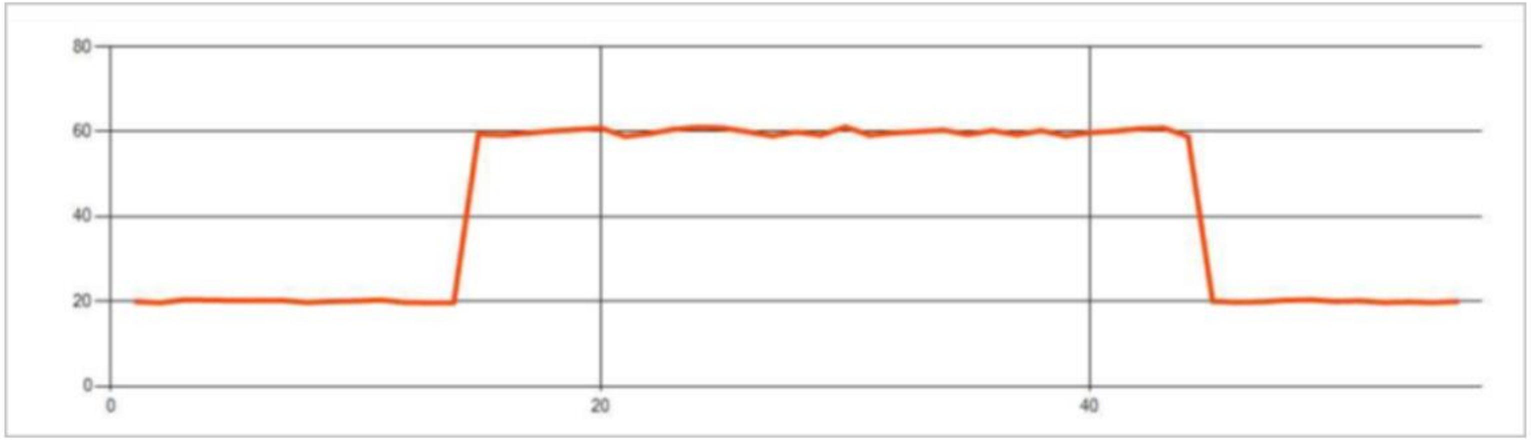
Copy

```sql
WITH AnomalyDetectionStep AS
(
    SELECT
        EVENTENQUEUEDUTCTIME AS time,
        CAST(temperature AS float) AS temp,
        AnomalyDetection_SpikeAndDip(CAST(temperature AS float), 95, 120, 'spikesanddips')
            OVER(LIMIT DURATION(second, 120)) AS SpikeAndDipScores
    FROM input
)
SELECT
    time,
    temp,
    CAST(GetRecordPropertyValue(SpikeAndDipScores, 'Score') AS float) AS
    SpikeAndDipScore,
    CAST(GetRecordPropertyValue(SpikeAndDipScores, 'IsAnomaly') AS bigint) AS
    IsSpikeAndDipAnomaly
INTO output
FROM AnomalyDetectionStep
```

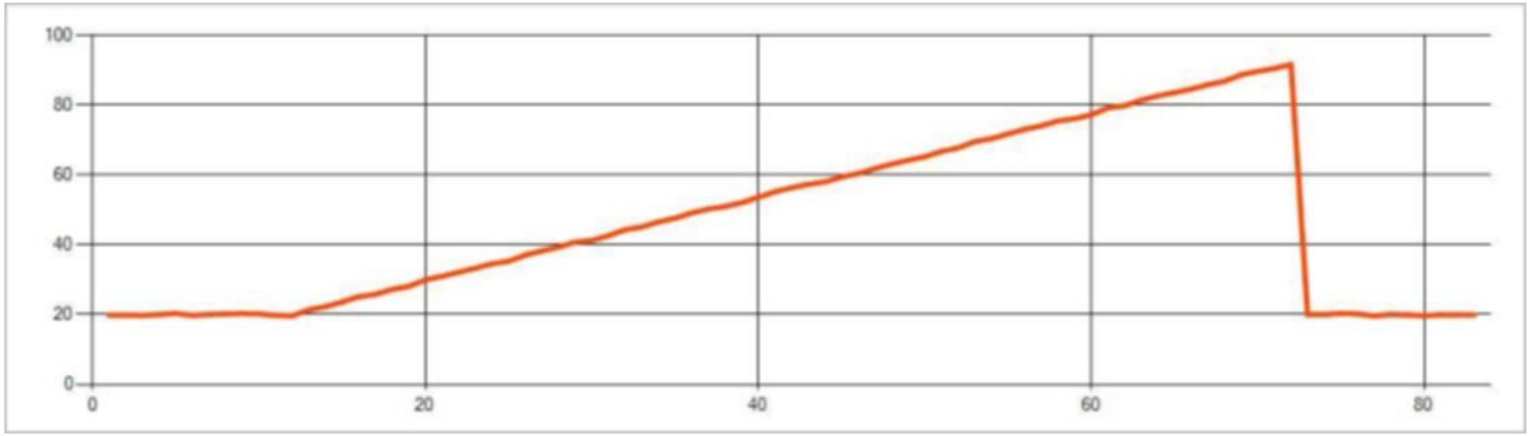
The following example query assumes a uniform input rate of one event per second in a 2-minute sliding window with a history of 120 events. The final SELECT statement extracts and outputs the score and anomaly status with a confidence level of 95%.

Change Point

The following image is an example of a level change:



The following image is an example of a trend change:



Change Point

SQL

Copy

```sql
WITH AnomalyDetectionStep AS
(
    SELECT
        EVENTENQUEUEDUTCTIME AS time,
        CAST(temperature AS float) AS temp,
        AnomalyDetection_ChangePoint(CAST(temperature AS float), 80, 1200)
        OVER(LIMIT DURATION(minute, 20)) AS ChangePointScores
    FROM input
)
SELECT
    time,
    temp,
    CAST(GetRecordPropertyValue(ChangePointScores, 'Score') AS float) AS
    ChangePointScore,
    CAST(GetRecordPropertyValue(ChangePointScores, 'IsAnomaly') AS bigint) AS
    IsChangePointAnomaly
INTO output
FROM AnomalyDetectionStep
```

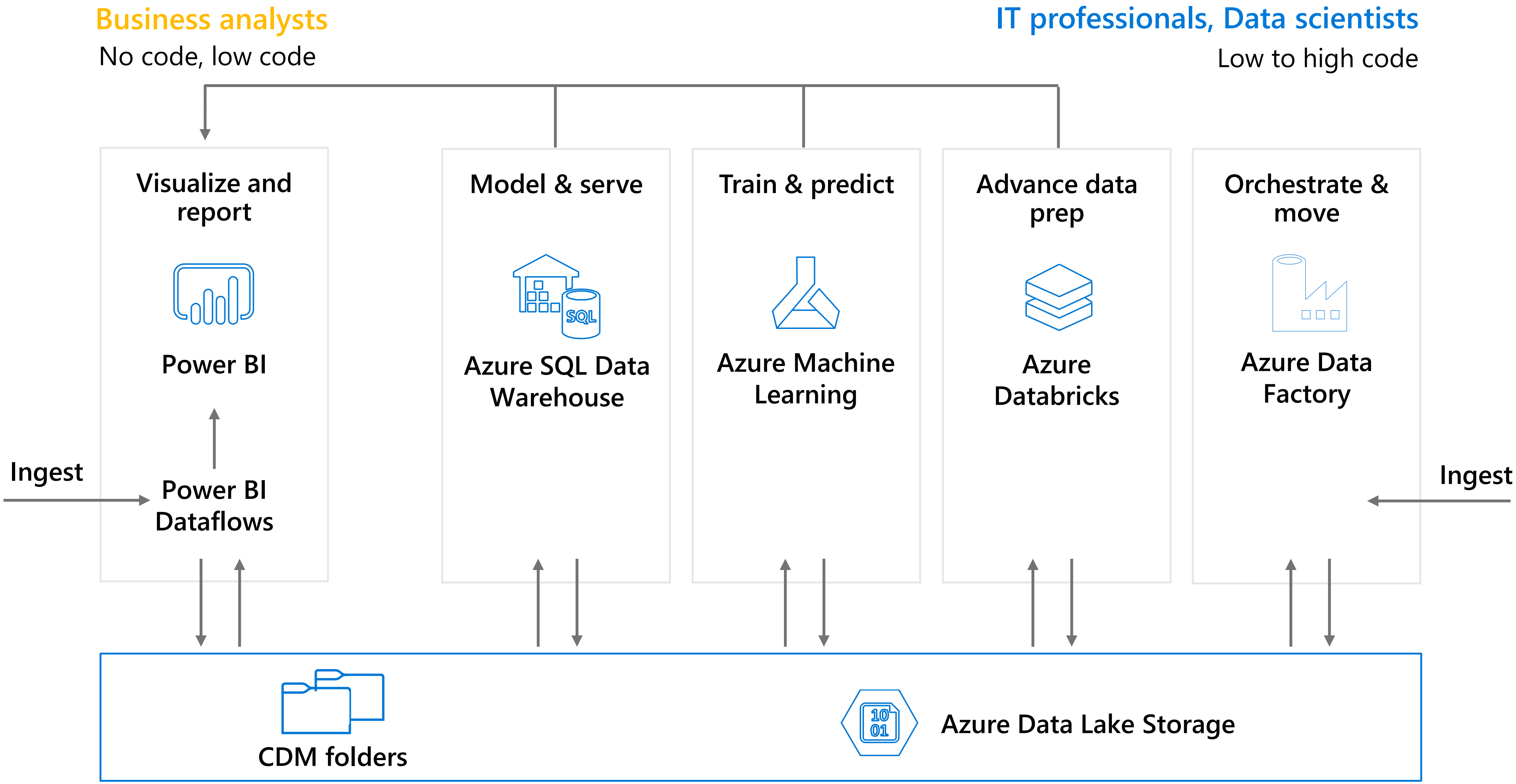
The following example query assumes a uniform input rate of one event per second in a 20-minute sliding window with a history size of 1200 events. The final SELECT statement extracts and outputs the score and anomaly status with a confidence level of 80%.

Microsoft Azure Solutions Business

Azure Data Factory

Ali Bouhaddou

Data Producer / Data consumer

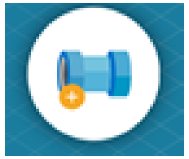


Code-Free ETL as a Service

INGEST

- Multi-cloud and on-prem hybrid copy data
- 90+ native connectors
- Serverless and auto-scale
- Use wizard for quick copy jobs

CONTROL FLOW



- Design code-free data pipelines
- Generate pipelines via SDK
- Utilize workflow constructs: loops, branches, conditional execution, variables, parameters, ...

DATA FLOW

- Code-free data transformations that execute in Spark
- Scale-out with Azure Integration Runtimes
- Generate data flows via SDK
- Designers for data engineers and data analysts

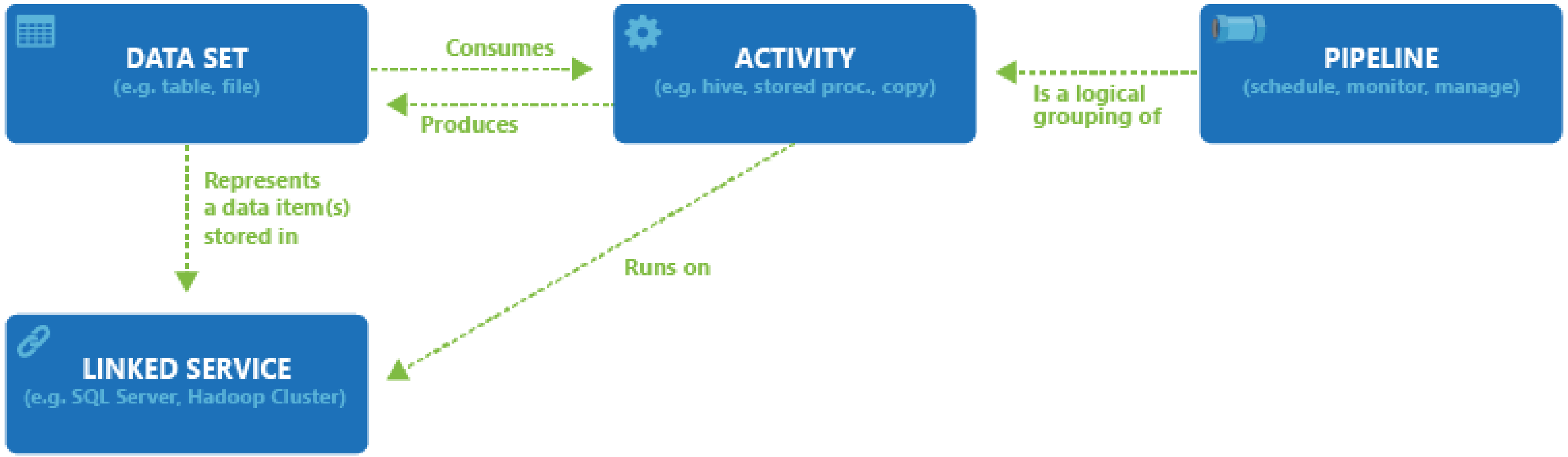
SCHEDULE

- Build and maintain operational schedules for your data pipelines
- Wall clock, event-based, tumbling windows, chained

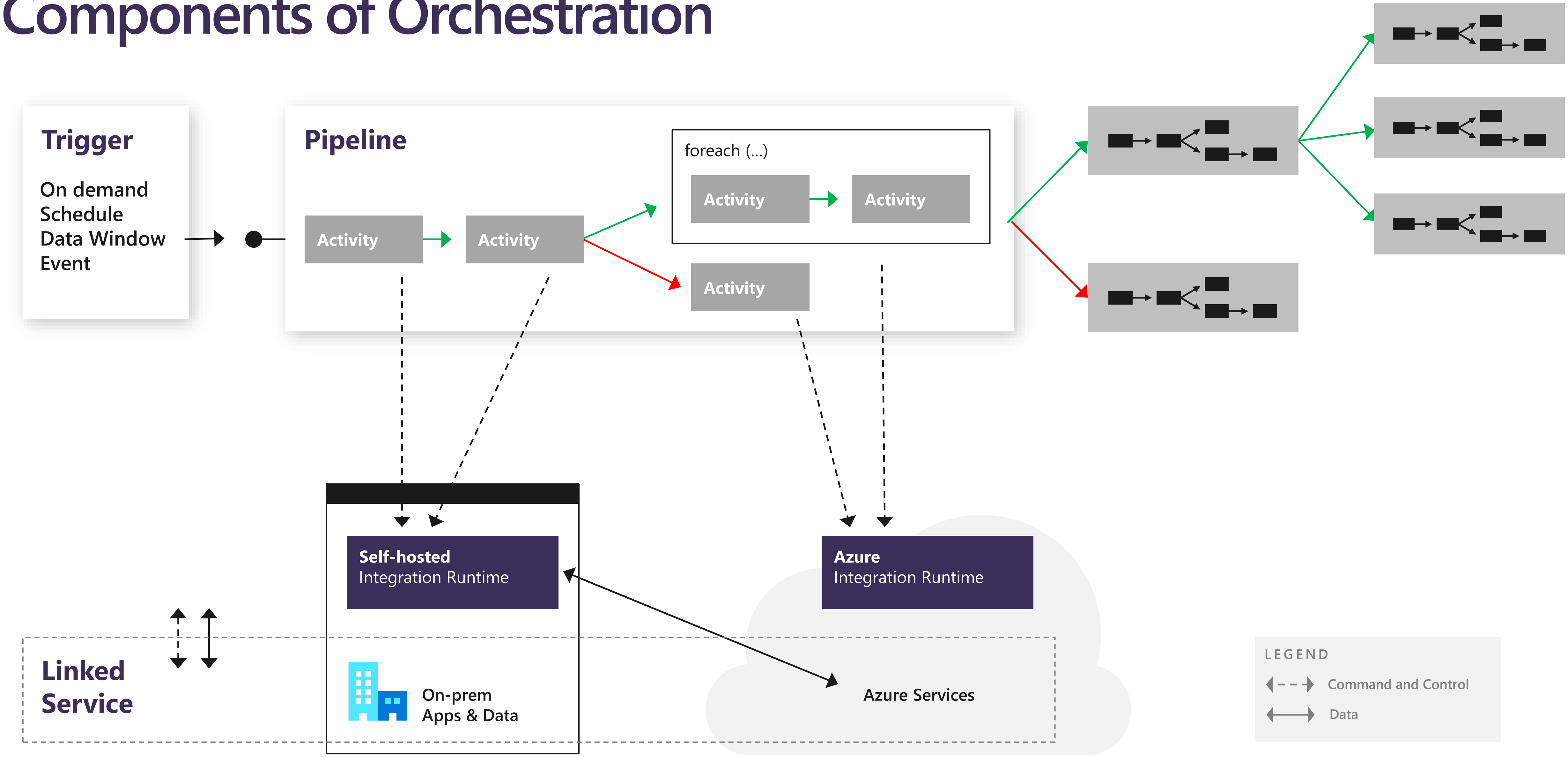
MONITOR

- View active executions and pipeline history
- Detail activity and data flow executions
- Establish alerts and notifications

Control Flow - Orchestration with Pipelines



Components of Orchestration

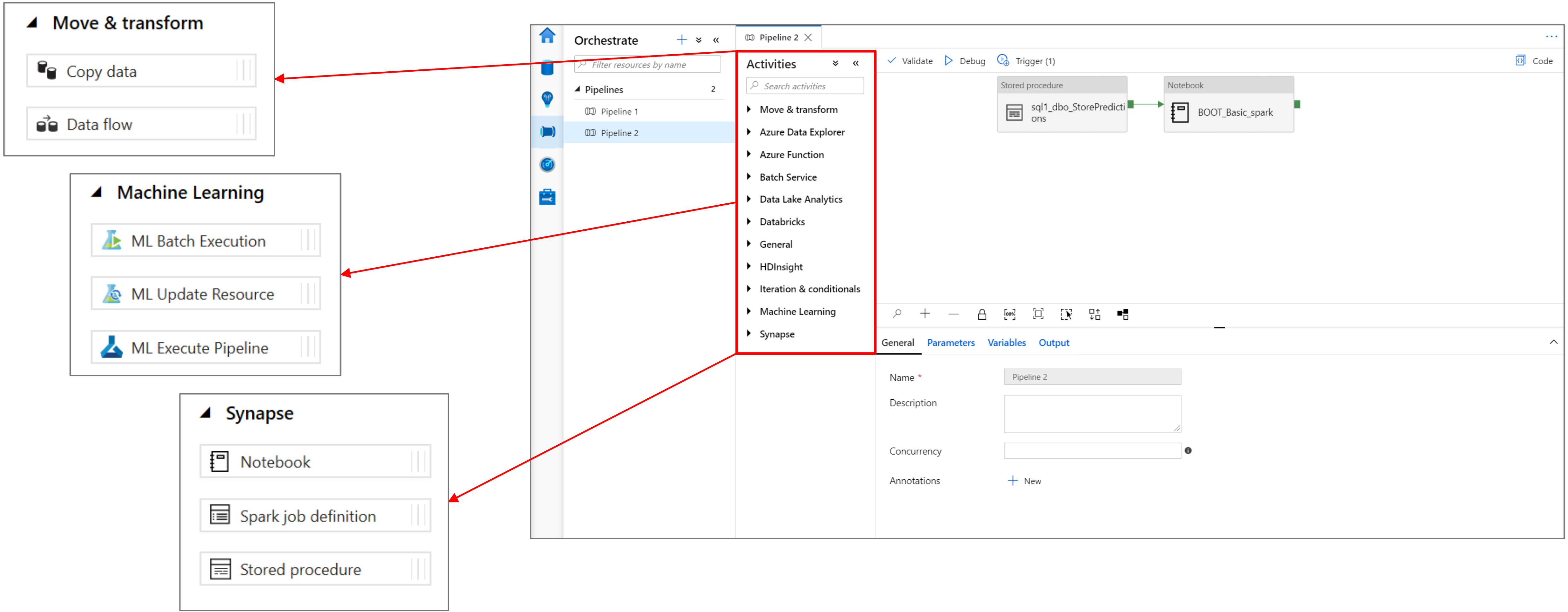


Synapse Pipelines shares codebase with Azure Data Factory

Pipelines

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

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



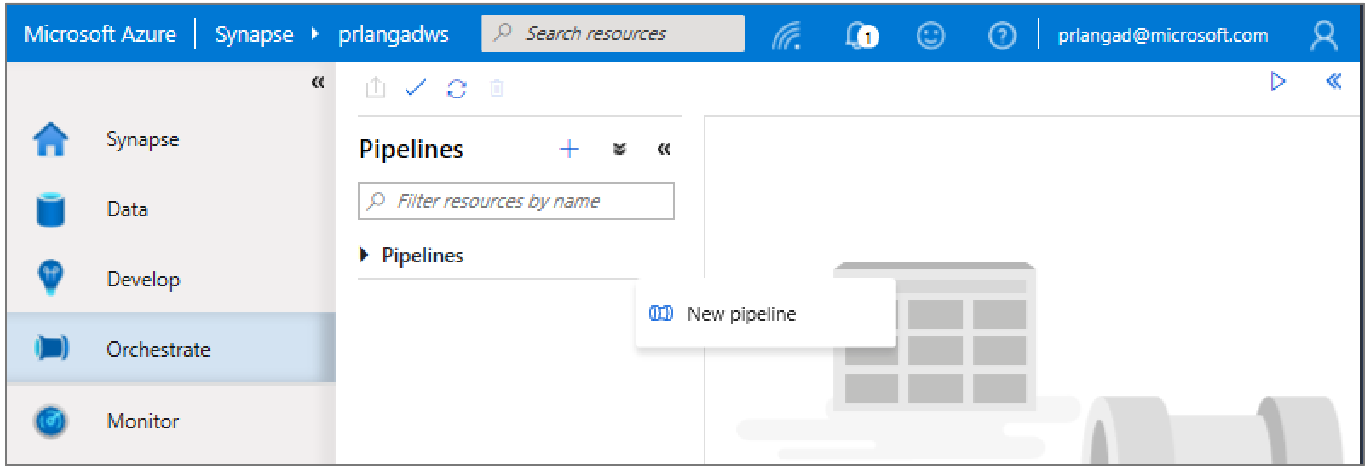
Pipelines

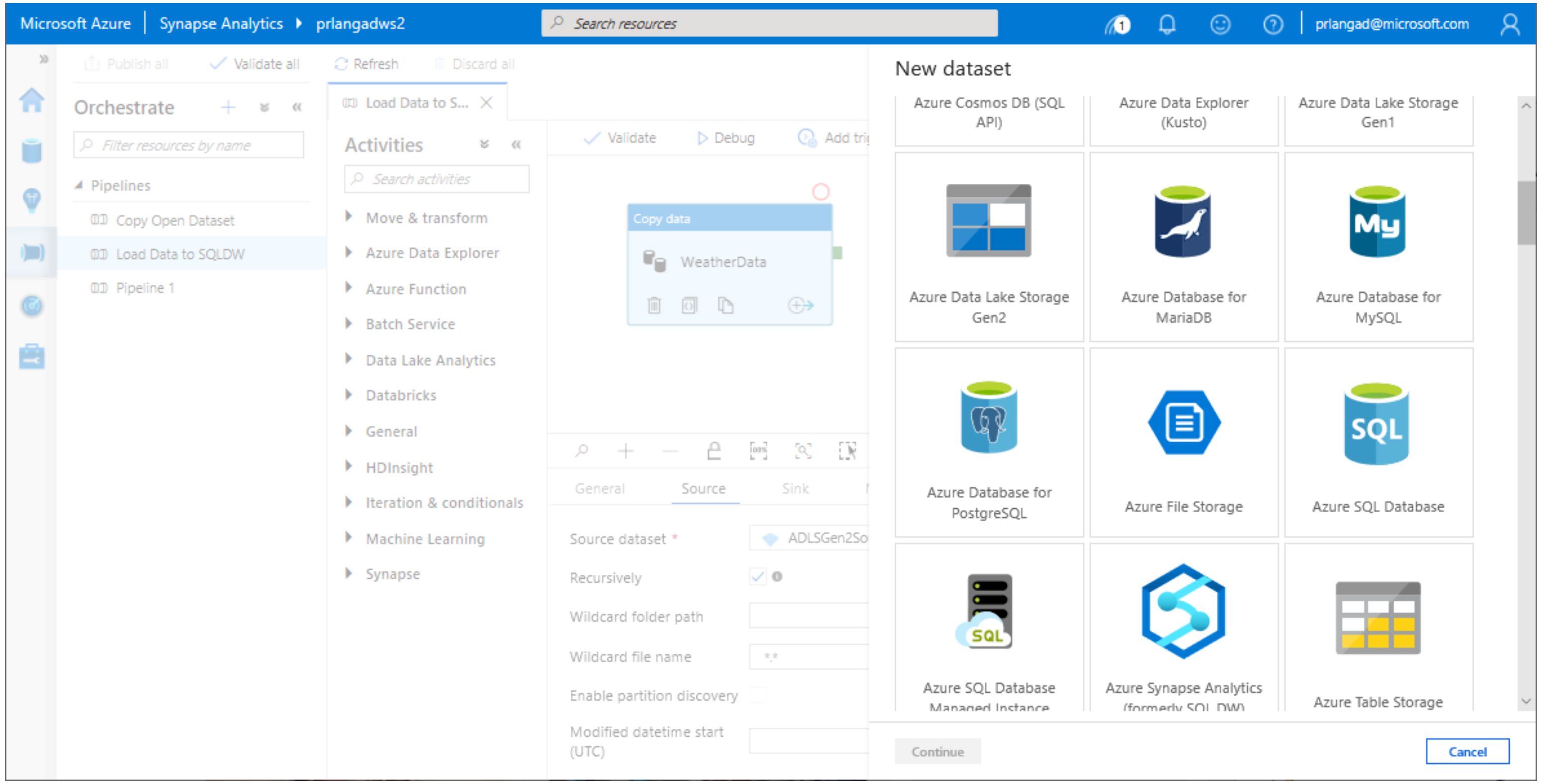
Overview

- Provide ability to load data from storage account to desired linked service.
- Load data by manual execution of pipeline or by orchestration.

Benefits

- Supports common loading patterns.
- Fully parallel loading into data lake or SQL tables.
- Graphical development experience.





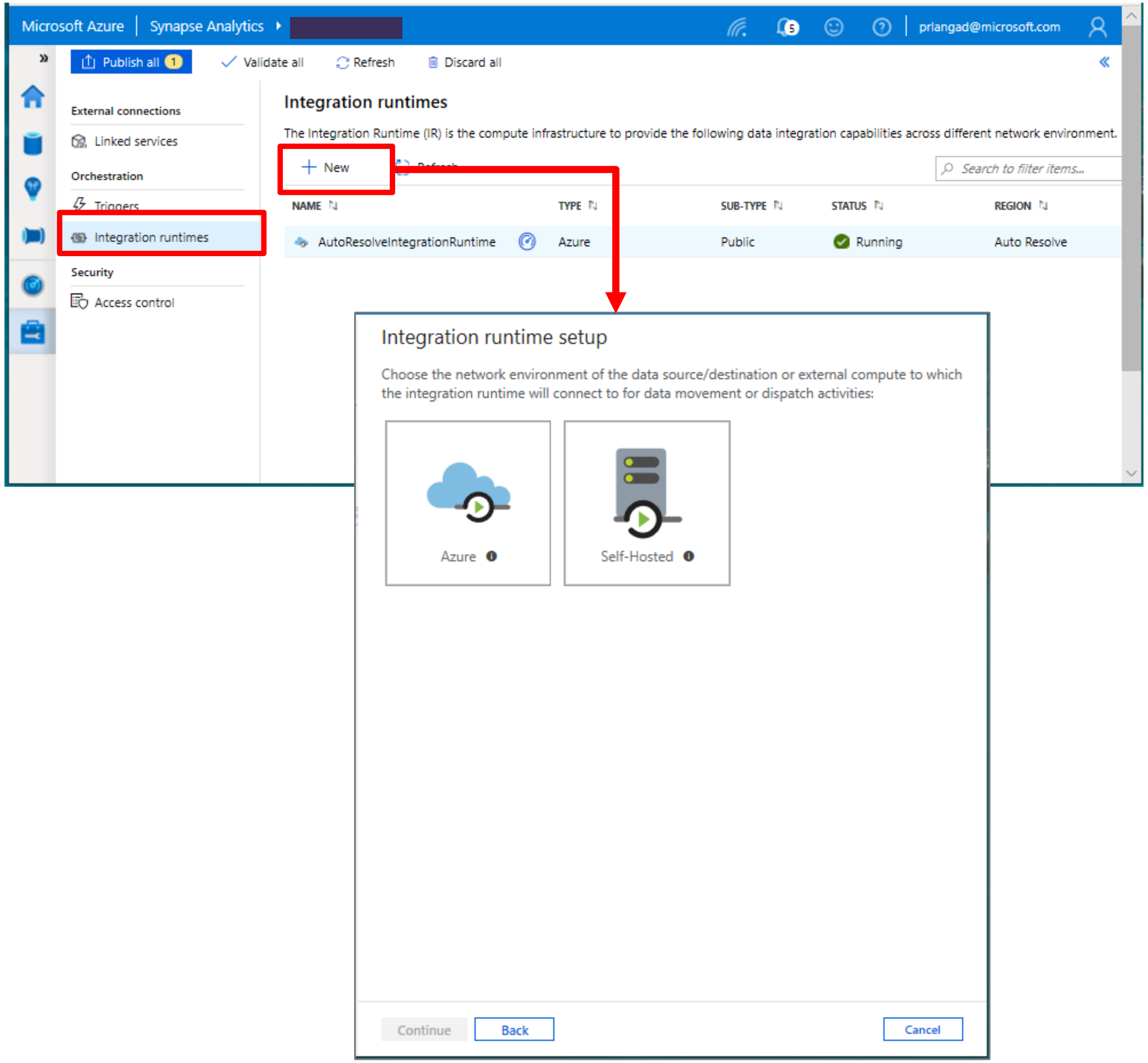
Integration runtimes

Overview

Integration runtimes are the compute infrastructure used by Pipelines to provide the data integration capabilities across different network environments. An integration runtime provides the bridge between the activity and linked services.

Benefits

- Offers Azure Integration Runtime or Self-Hosted Integration Runtime
- Azure Integration Runtime – provides fully managed, serverless compute in Azure
- Self-Hosted Integration Runtime – use compute resources in on-premises machine or a VM inside private network



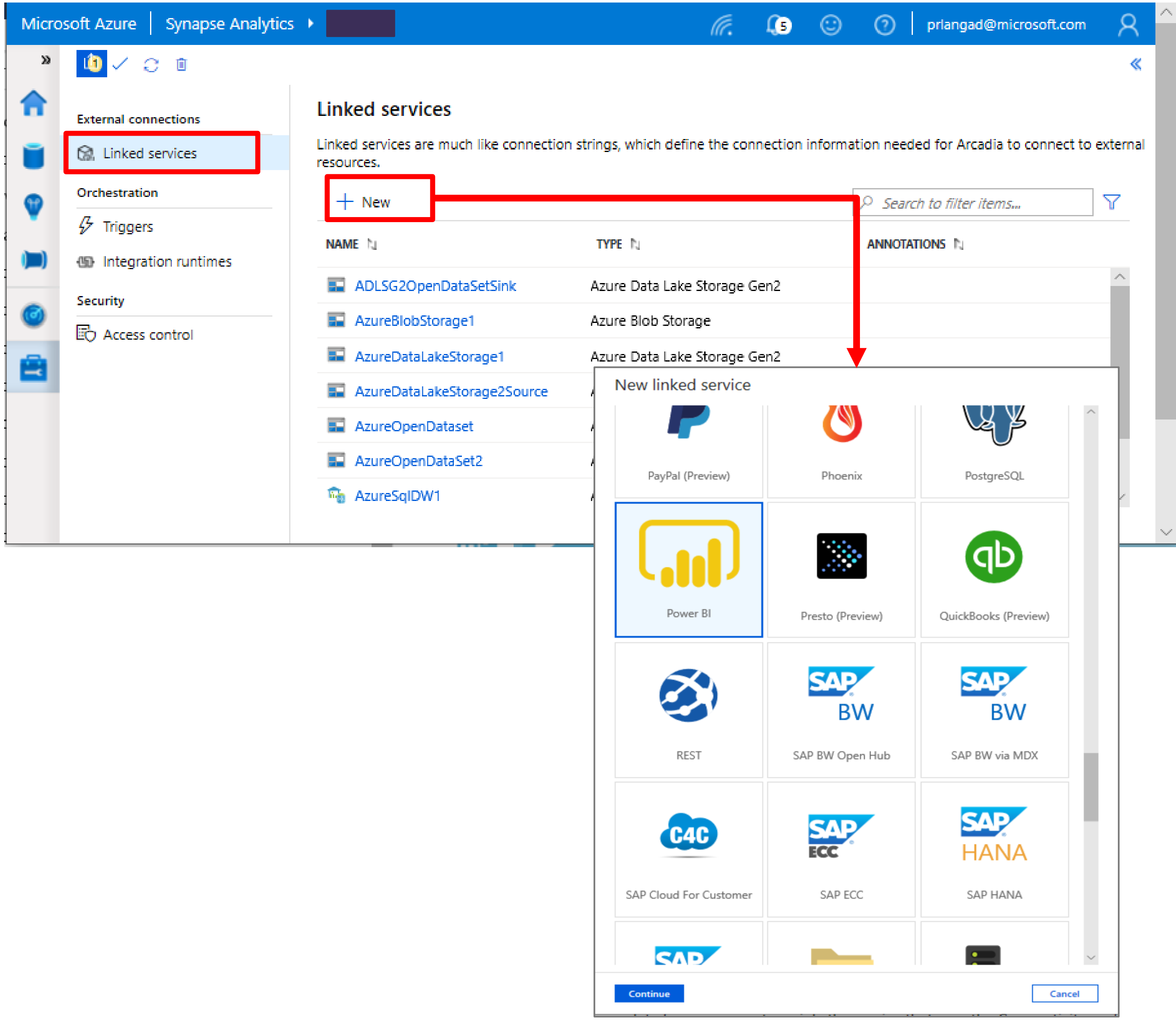
Linked services

Overview

Linked services define the connection information needed to connect to external resources.

Benefits

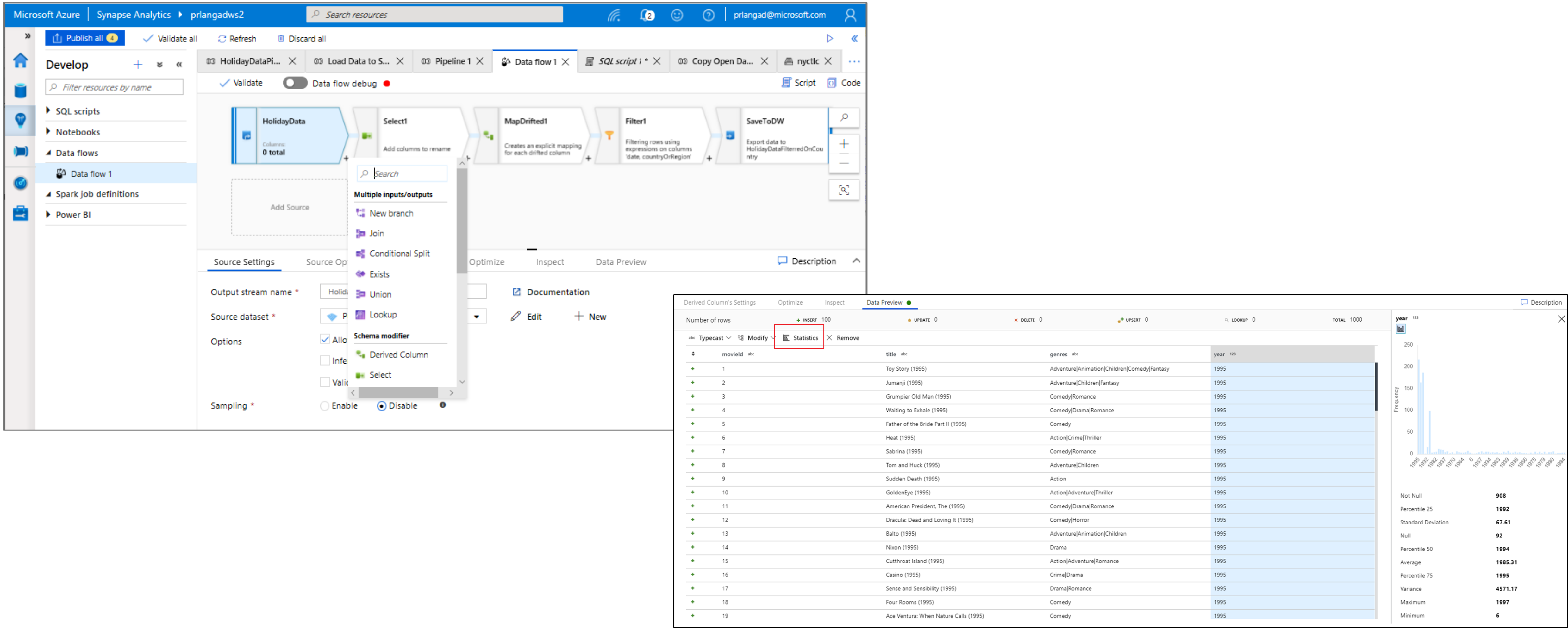
- Offers pre-build 90+ connectors
- Easy cross platform data migration
- Represents data store *or* compute resources



Data Flows

Mapping Data flows are pipeline activities providing a visual way of specifying how to transform data.

Provides a code-free experience.



Dataflow Capabilities

- Handle upserts, updates, deletes on sql sinks
- Add new partition methods
- Add schema drift support
- Add file handling (move files after read, write files to file names described in rows etc)
- New inventory of functions (for e.g Hash functions for row comparison)
- Commonly used ETL patterns(Sequence generator/Lookup transformation/SCD…)
- Data lineage – Capturing sink column lineage & impact analysis(invaluable if this is for enterprise deployment)
- Implement commonly used ETL patterns as templates(SCD Type1, Type2, Data Vault)

Triggers

Overview

Triggers represent a unit of processing that determines when a pipeline execution needs to be kicked off.

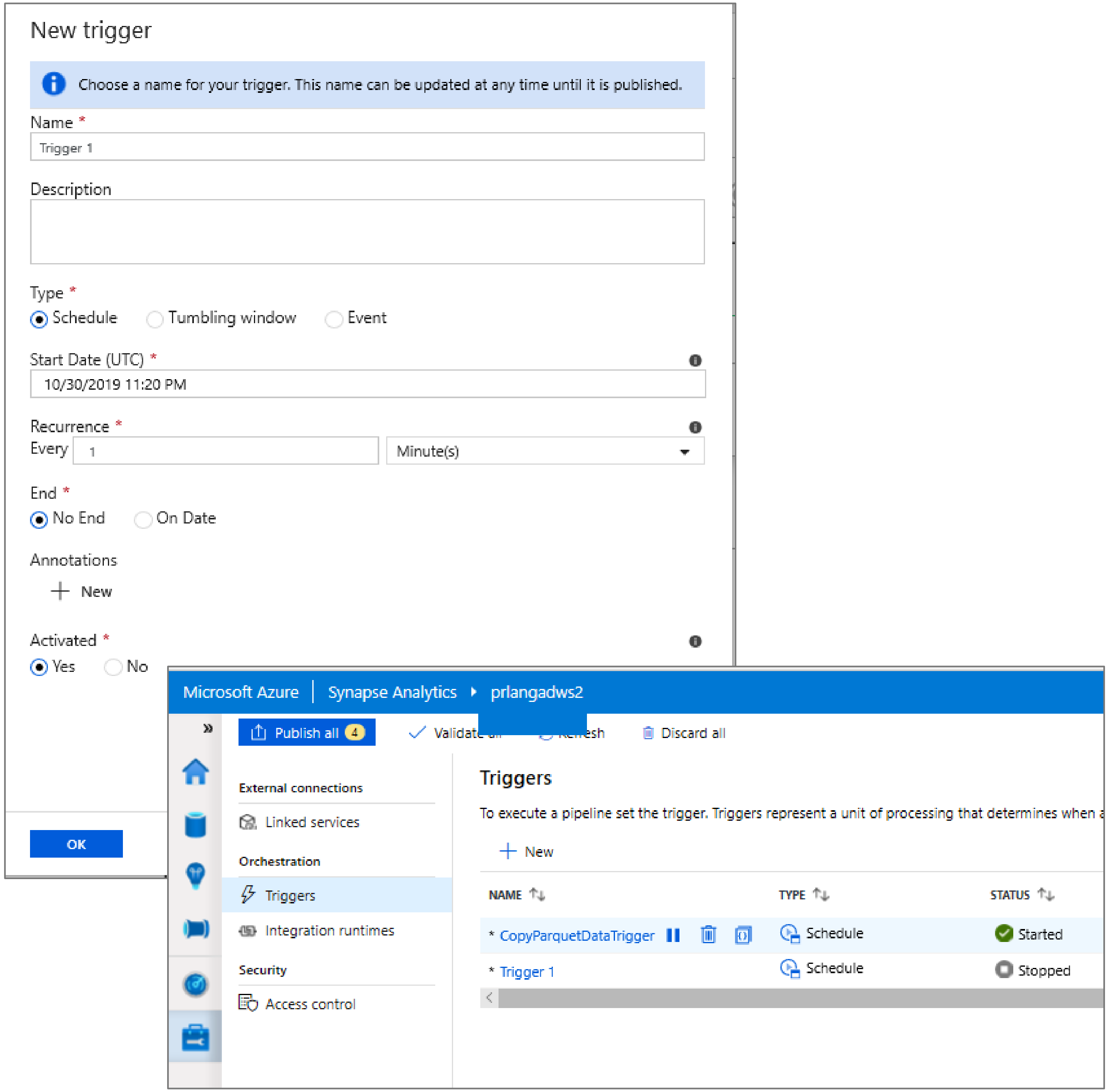
Data Integration offers 3 trigger types as –

1. Schedule – gets fired at a schedule with information of start date, recurrence, end date

2. Event – gets fired on specified event

3. Tumbling window – gets fired at a periodic time interval from a specified start date, while retaining state

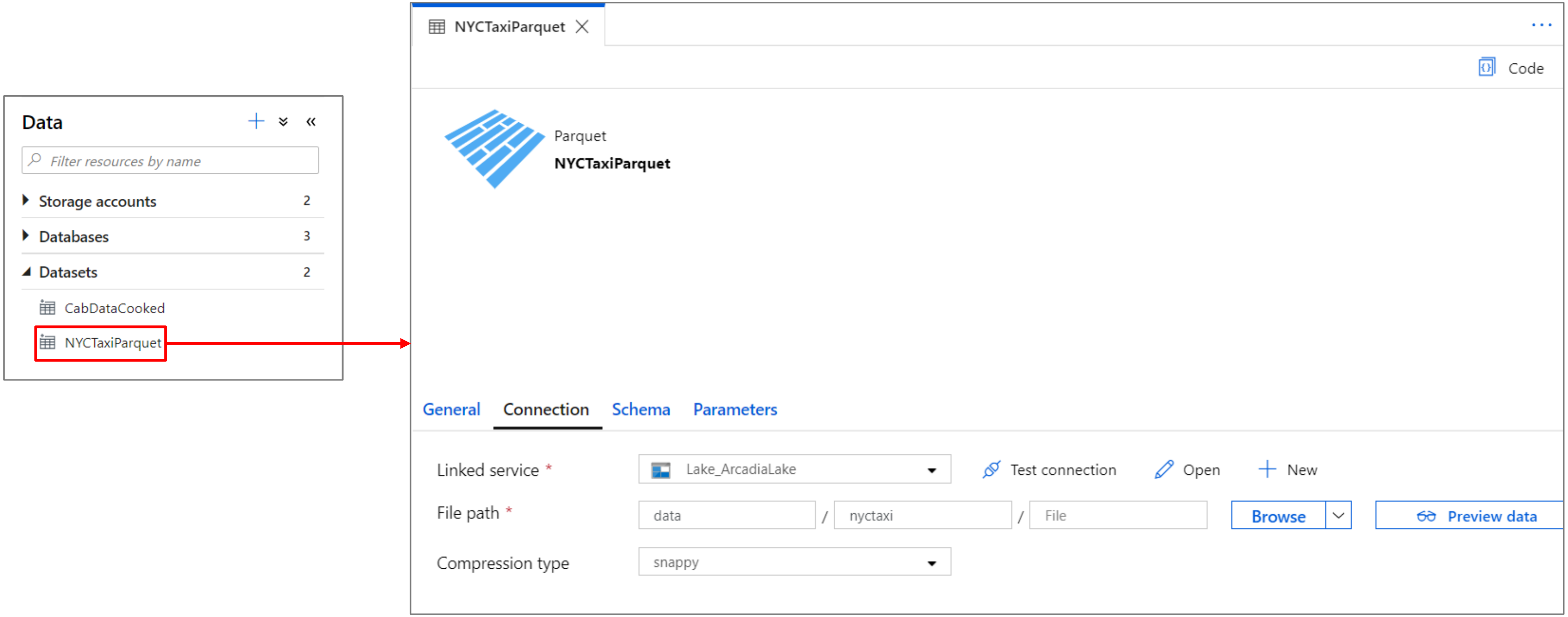
It also provides ability to monitor pipeline runs and control trigger execution.



Datasets

Orchestration datasets describe data that is persisted.

Once a dataset is defined, it can be used in pipelines and sources of data or as sinks of data.



Data Movement

Scalable

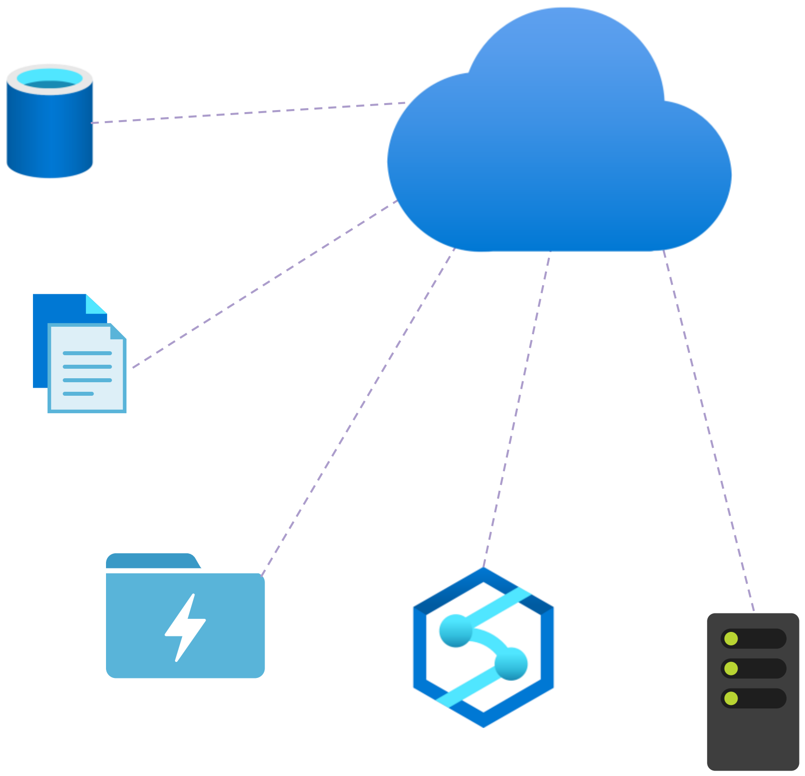
- per job elasticity
- Up to 4 GB/s

Simple

- Visually author or via code (Python, .Net, etc.)
- Serverless, no infrastructure to manage

Access all your data

- 90+ connectors provided and growing (cloud, on premises, SaaS)
- Data Movement as a Service: 25 points of presence worldwide
- Self-hostable Integration Runtime for hybrid movement

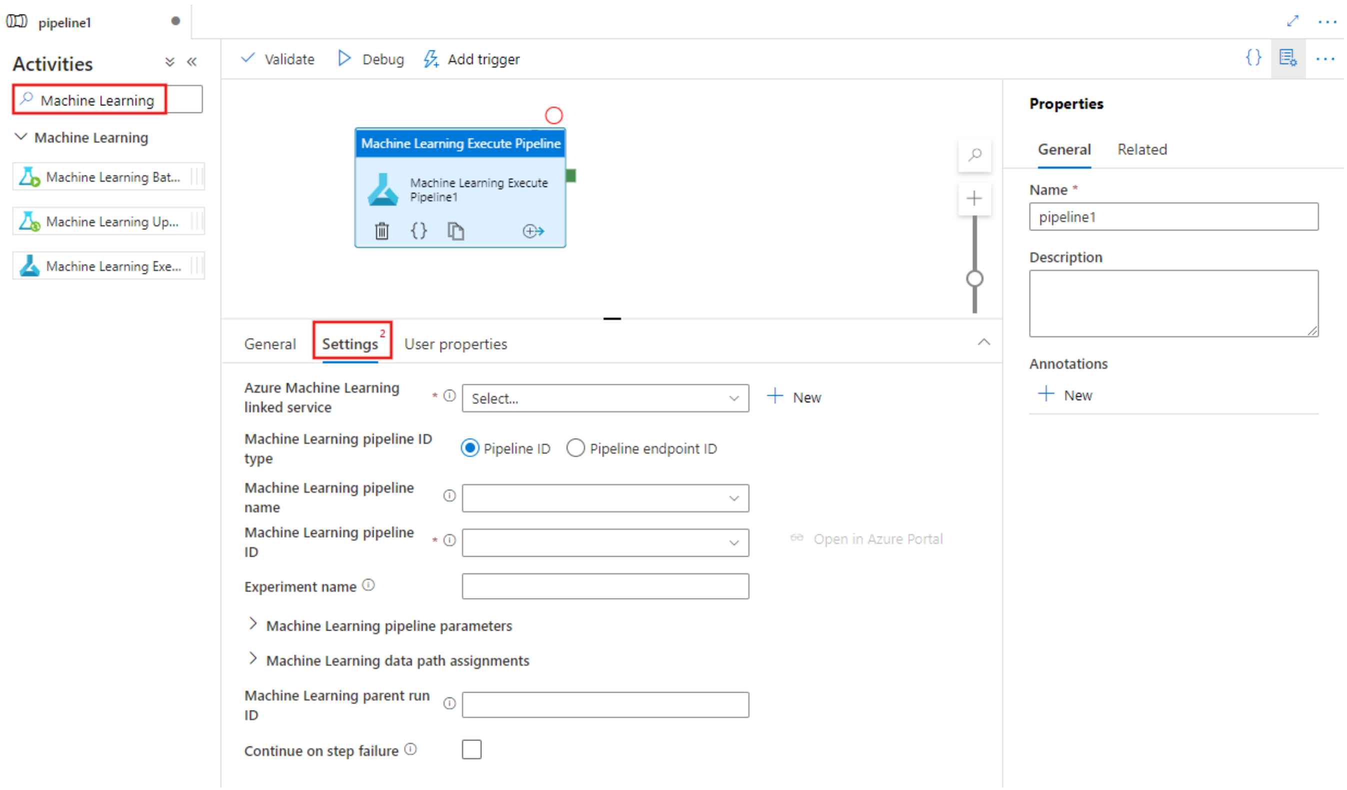


90+ Connectors out of the box

Azure (15)	Database & DW (26)		File Storage (6)	File Formats(6)	NoSQL (3)	Services and App (28)		Generic (4)
Blob storage	Amazon Redshift	Oracle	Amazon S3	AVRO	Cassandra	Amazon MWS	Oracle Service Cloud	Generic HTTP
Cosmos DB - SQL API	DB2	Phoenix	File system	Binary	Couchbase	CDS for Apps	PayPal	Generic OData
Cosmos DB - MongoDB API	Drill	PostgreSQL	FTP	Delimited Text	MongoDB	Concur	QuickBooks	Generic ODBC
Data Explorer	Google BigQuery	Presto	Google Cloud Storage	JSON		Dynamics 365	Salesforce	Generic REST
Data Lake Storage Gen1	Greenplum	SAP BW Open Hub	HDFS	ORC		Dynamics AX	SF Service Cloud	
Data Lake Storage Gen2	HBase	SAP BW via MDX	SFTP	Parquet		Dynamics CRM	SF Marketing Cloud	
Database for MariaDB	Hive	SAP HANA				Google AdWords	SAP 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								

Batch Execution ML Pipelines

Property	Description	Allowed values	Required
name	Name of the activity in the pipeline	String	Yes
type	Type of activity is 'AzureMLExecutePipeline'	String	Yes
linkedServiceName	Linked Service to Azure Machine Learning	Linked service reference	Yes
mlPipelineId	ID of the published Azure Machine Learning pipeline	String (or expression with resultType of string)	Yes
experimentName	Run history experiment name of the Machine Learning pipeline run	String (or expression with resultType of string)	No
mlPipelineParameters	Key, Value pairs to be passed to the published Azure Machine Learning pipeline endpoint. Keys must match the names of pipeline parameters defined in the published Machine Learning pipeline	Object with key value pairs (or Expression with resultType object)	No
mlParentRunId	The parent Azure Machine Learning pipeline run ID	String (or expression with resultType of string)	No
dataPathAssignments	Dictionary used for changing datapaths in Azure Machine learning. Enables the switching of datapaths	Object with key value pairs	No
continueOnStepFailure	Whether to continue execution of other steps in the Machine Learning pipeline run if a step fails	boolean	No

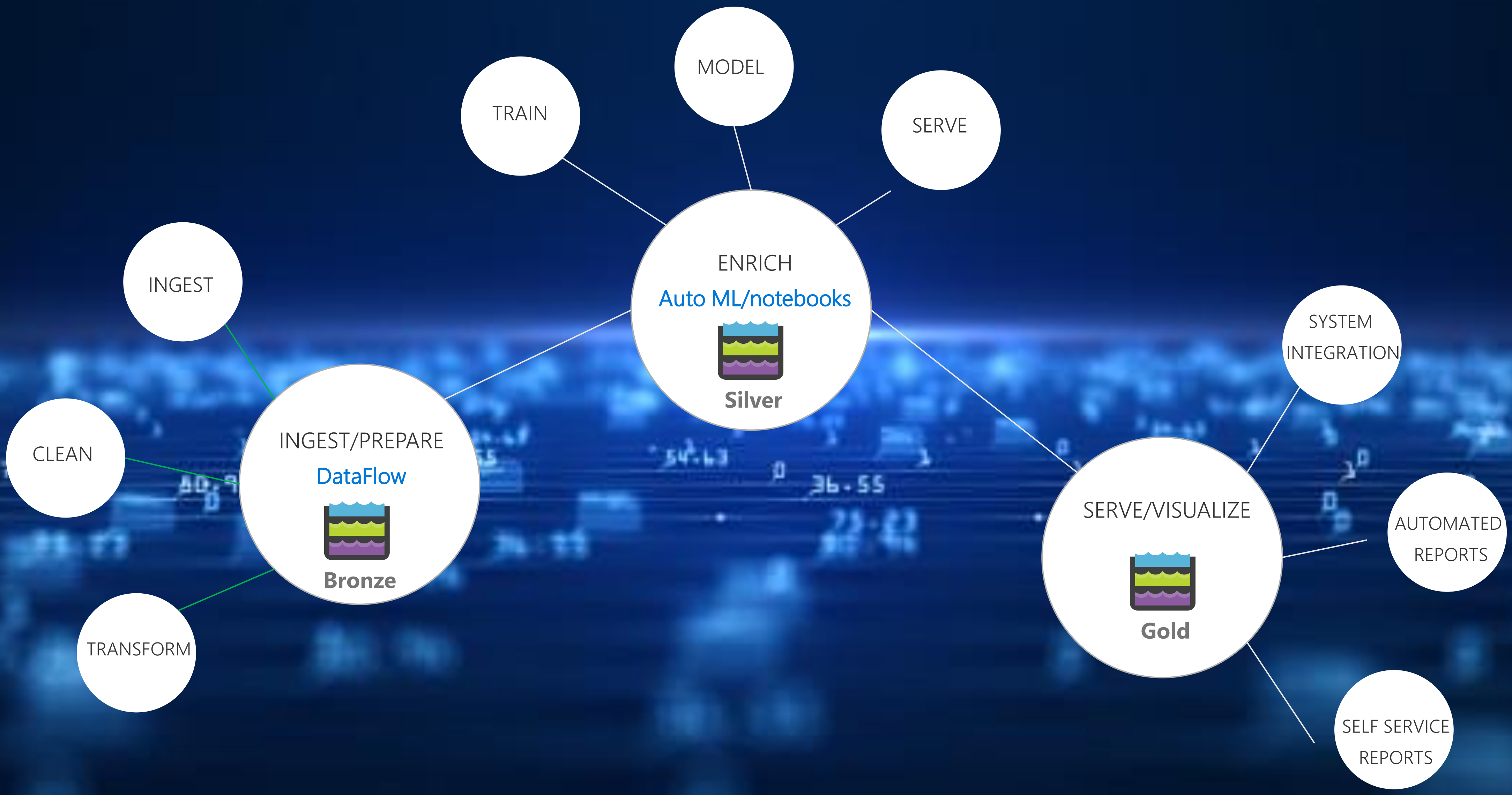


Microsoft Azure Solutions Business

Power BI

Ali Bouhaddou

Modern BI workload



Modern BI < > reporting

Modern BI = Serverless + Spark + ML

	Data Engineers	Data Analysts	Data Scientists
INGEST	Synapse Pipelines – Copy activities	PowerBI online DataFlows	notebooks
TRANSFORM & ENRICH	Mapping DataFlows (Synapse / Data Factory)	PowerBI online DataFlows Customer Insight Synapse Wrangling DataFlows	
SERVE	Azure SQLDB/Synapse	PowerBI Dataset	

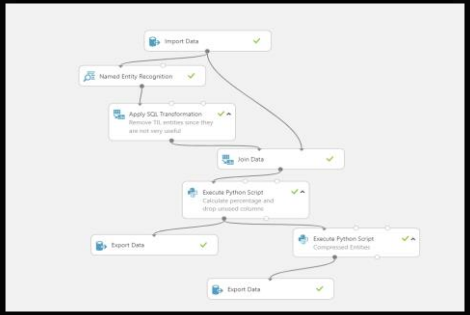
The Most Complete AI Capabilities in a BI Product

Data Scientists

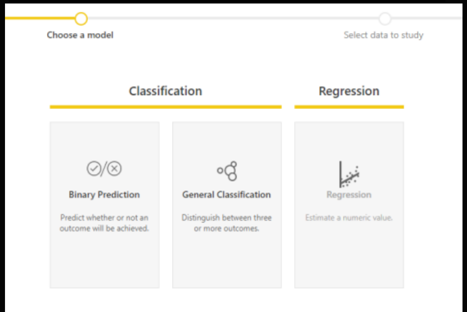
Analysts

End users

Extend with Azure ML

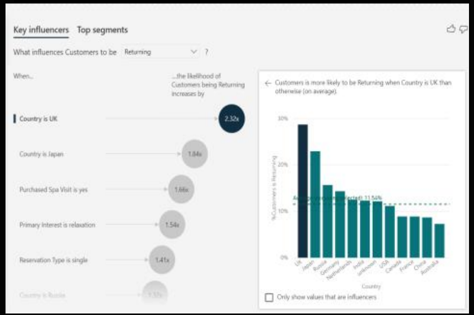


Create ML models

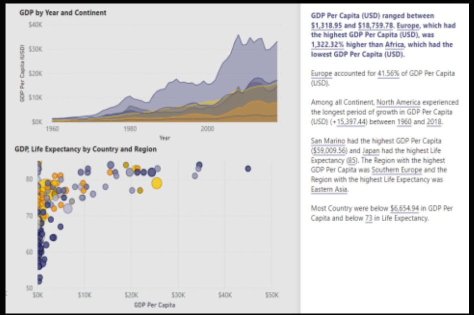


Sentiment Analysis

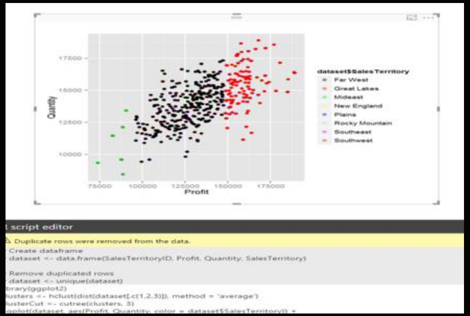
Key Driver Analysis



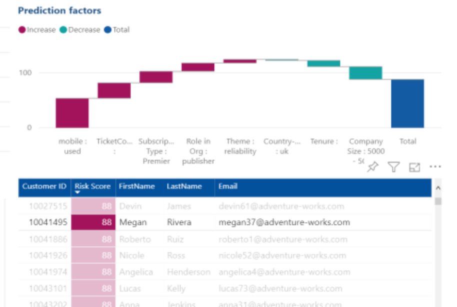
Smart Narratives



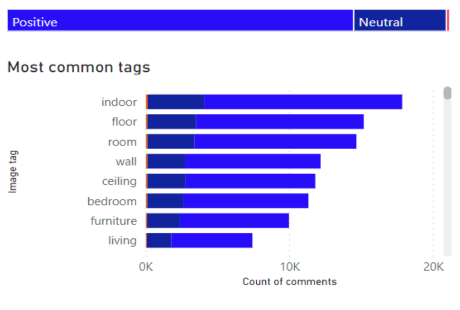
Python & R Integration



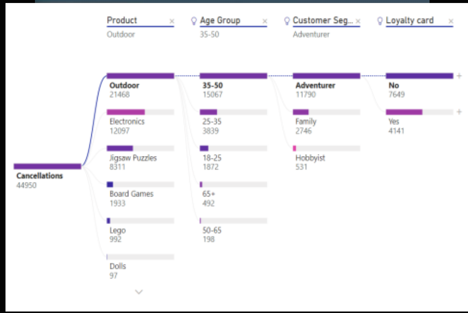
Explore Predictions



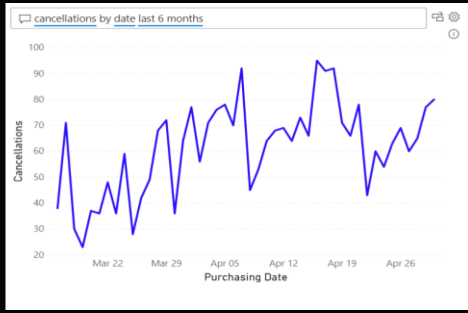
Key Phrase Extraction



Root Cause Analysis



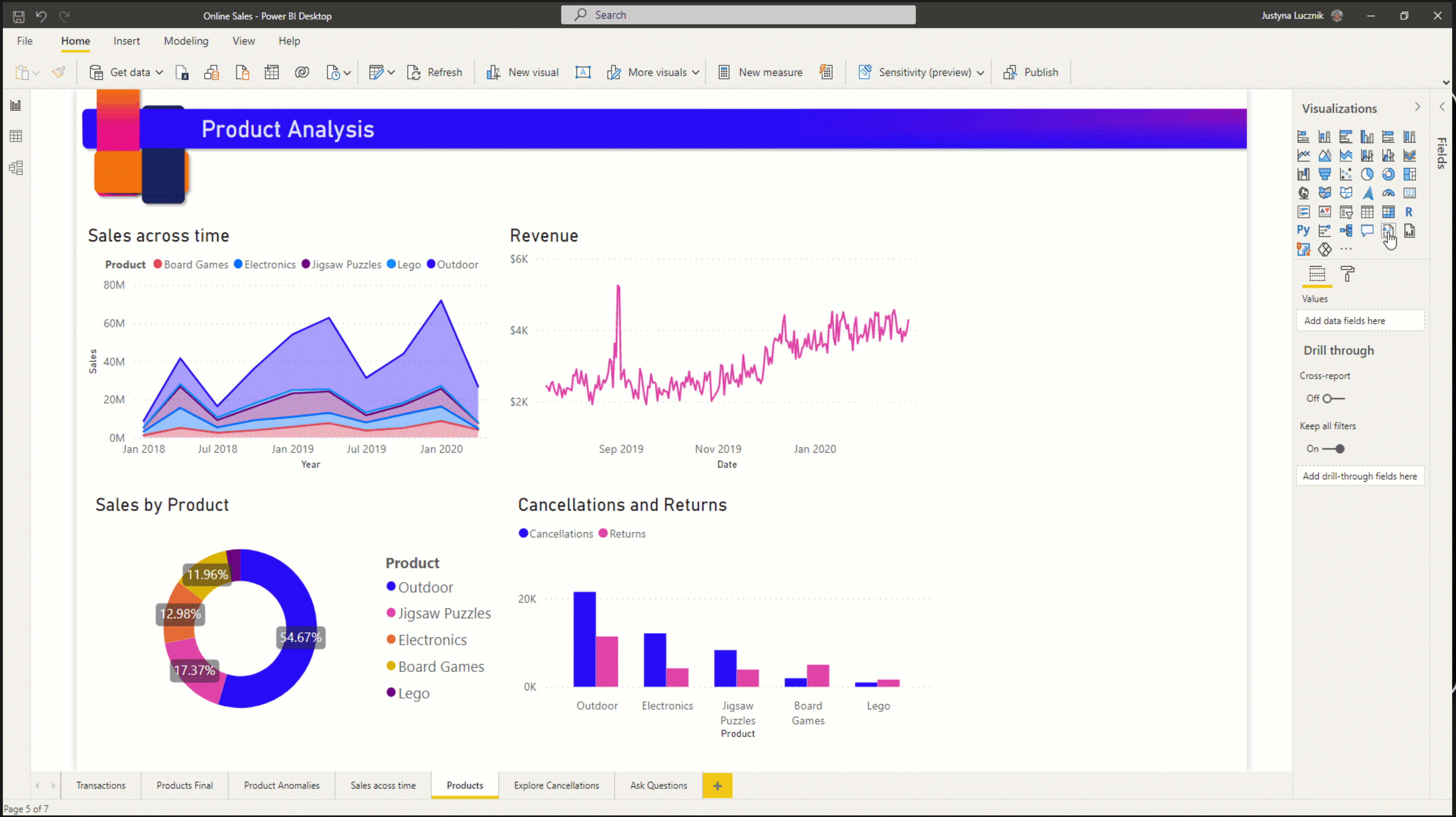
Q&A



Empowering every individual
Augmented analytics

Information workers

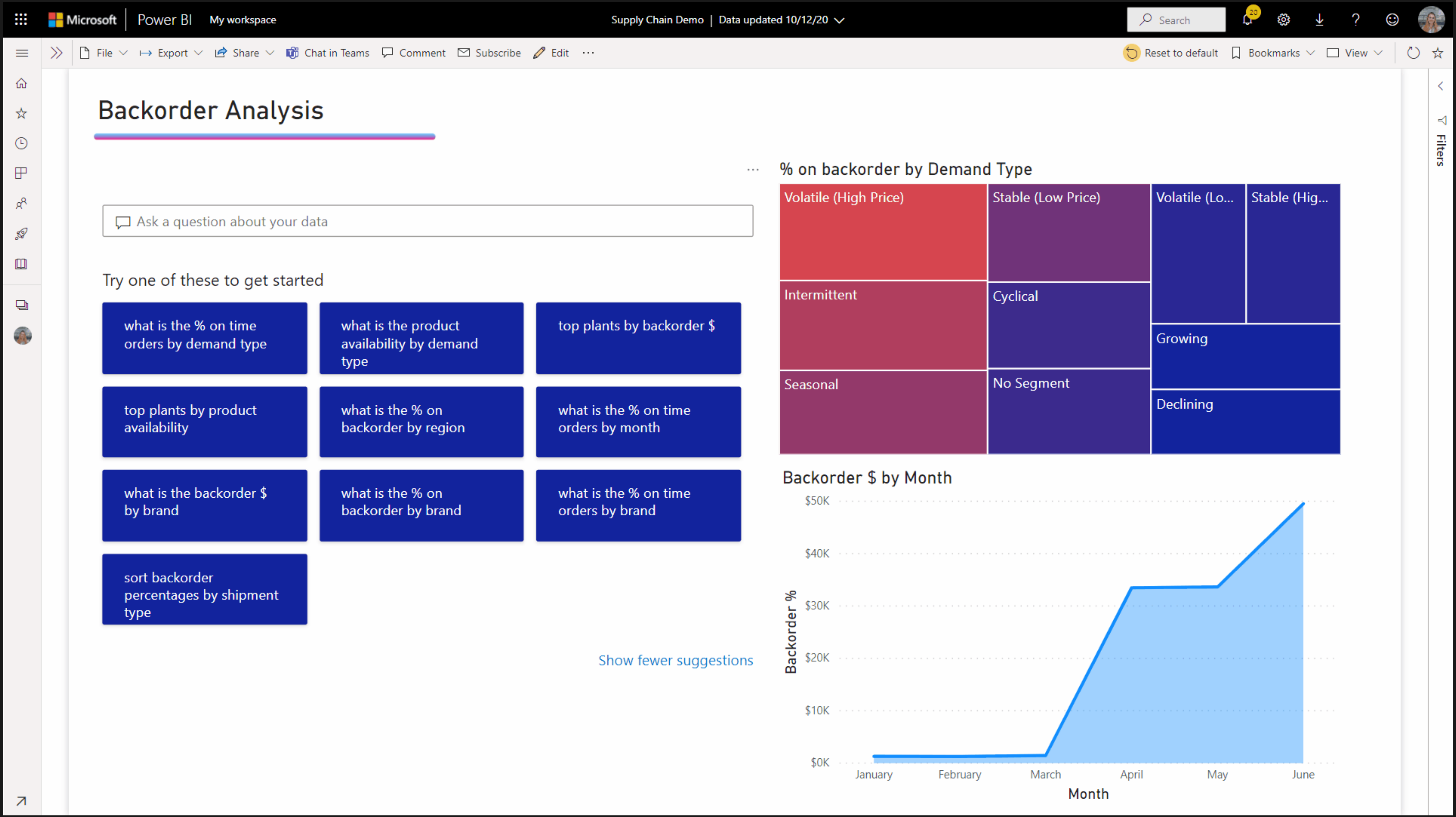
Smart narratives



Empowering every individual
Augmented analytics

Information workers

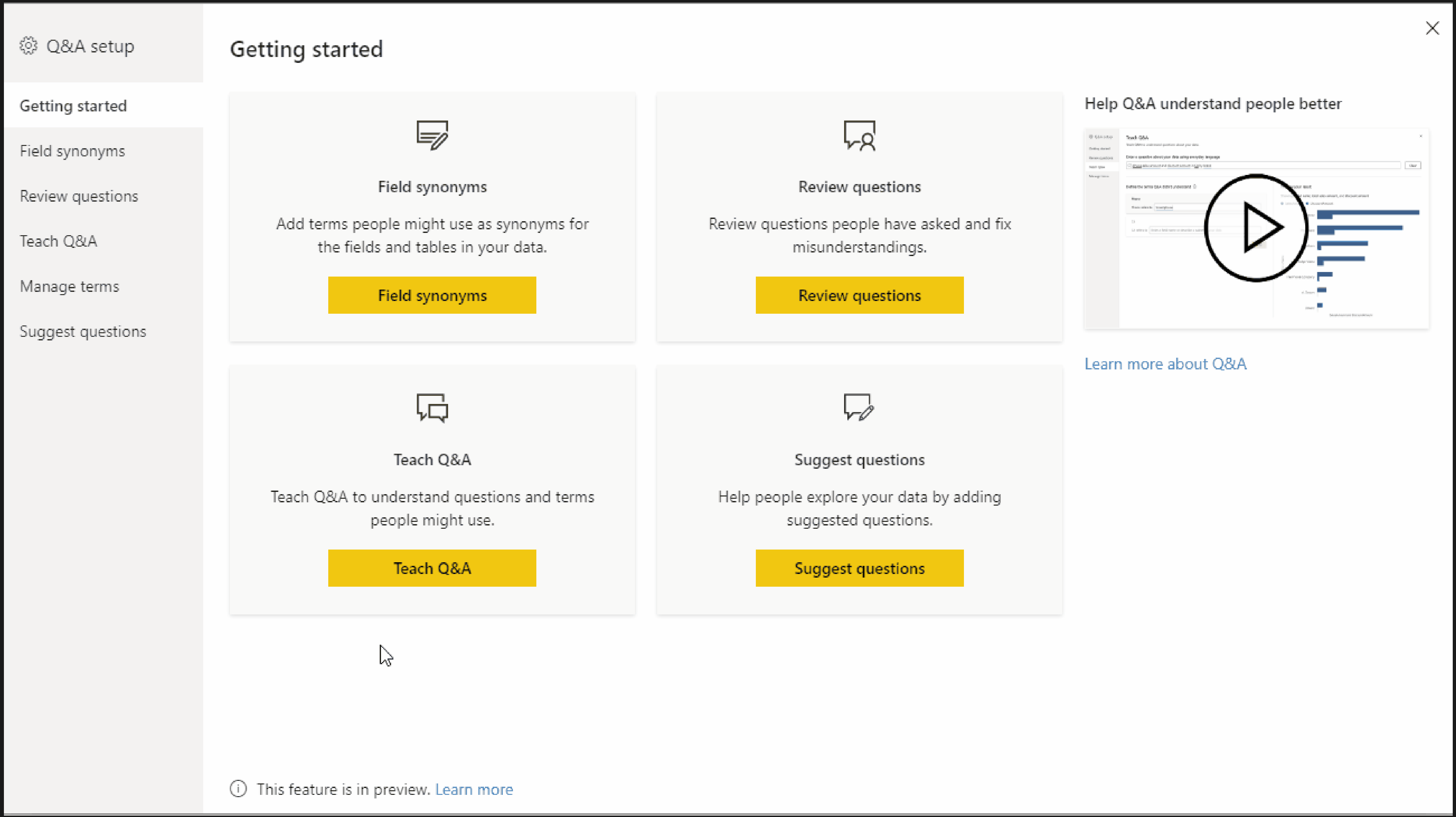
Q & A



Empowering every individual
Augmented analytics

Information workers

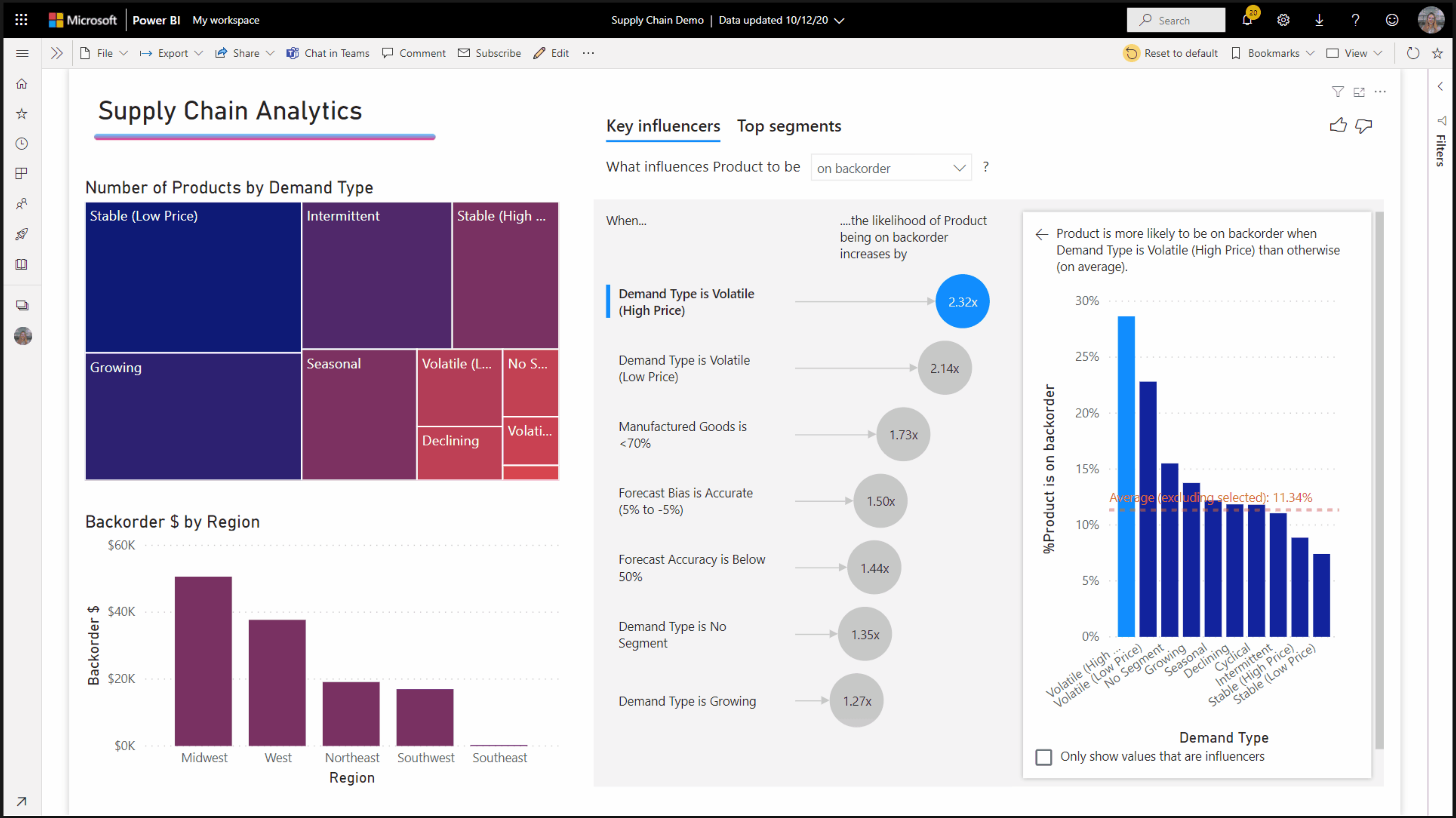
Q & A Tools



Empowering every individual
Augmented analytics

Business Analysts

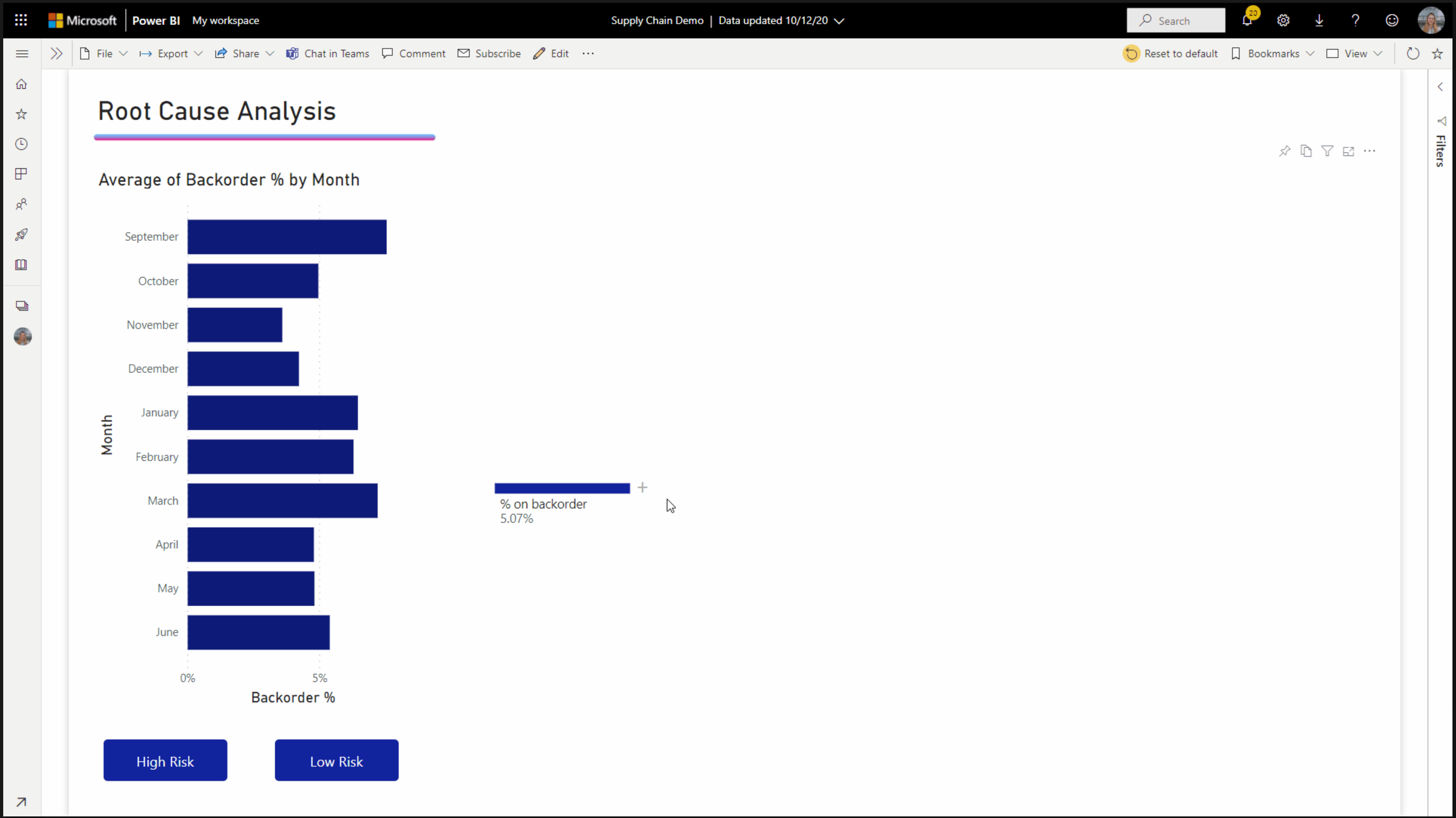
Key Influencers



Empowering every individual
Augmented analytics

Business Analysts

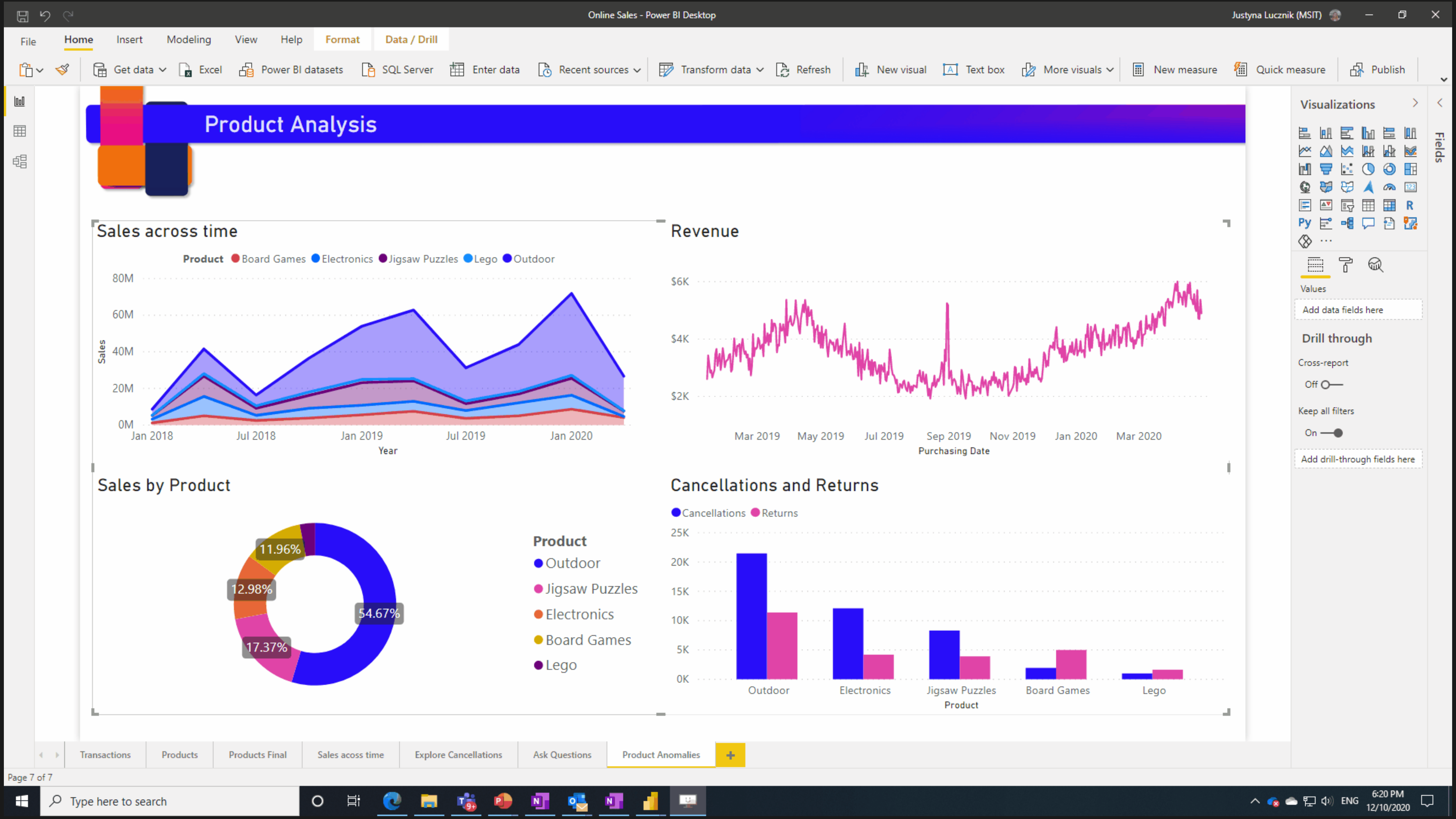
Decomposition Tree



Empowering every individual
Augmented analytics

Business Analysts

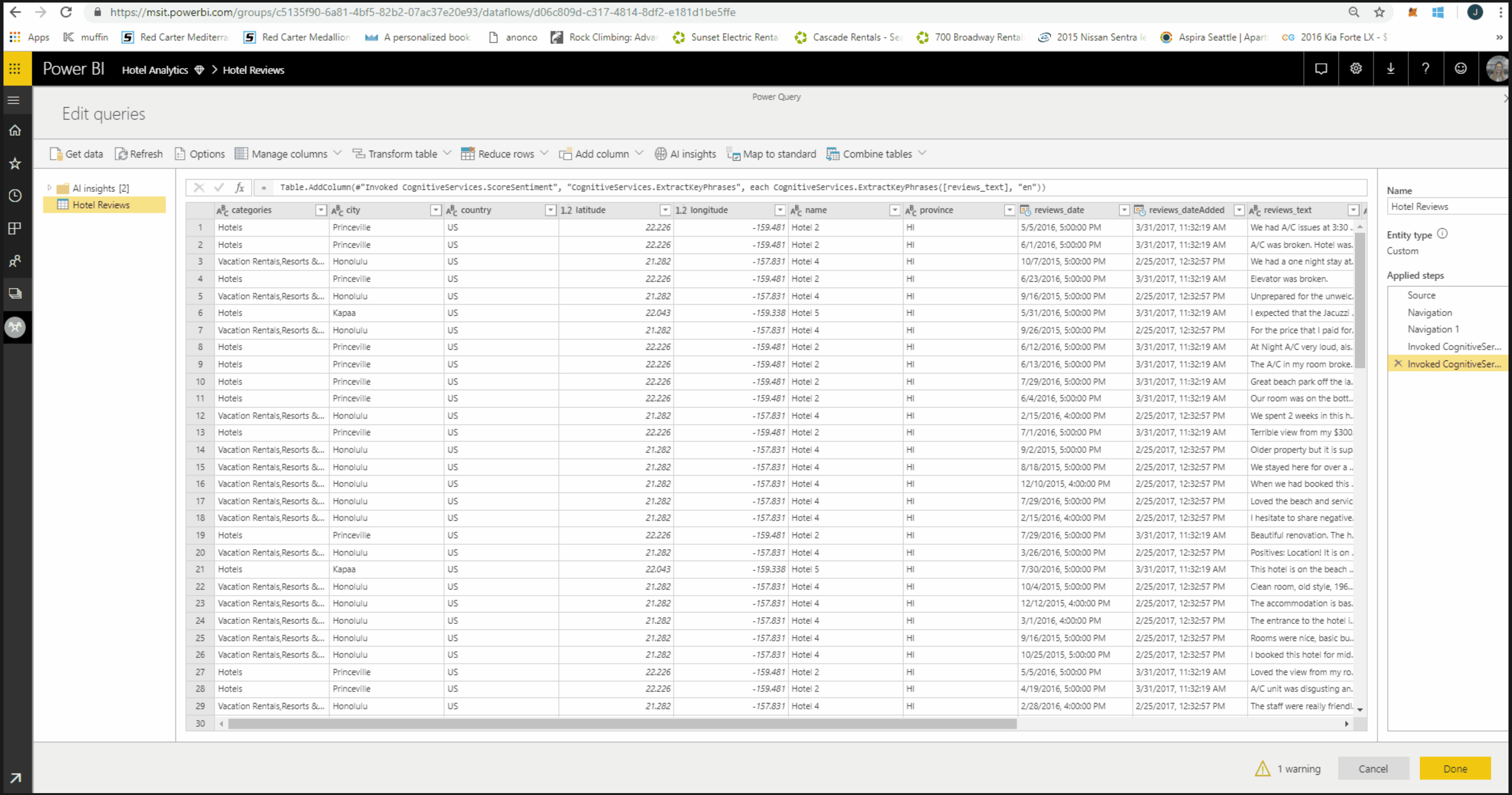
Anomaly Detection



Empowering every individual
Augmented analytics

Data Scientist

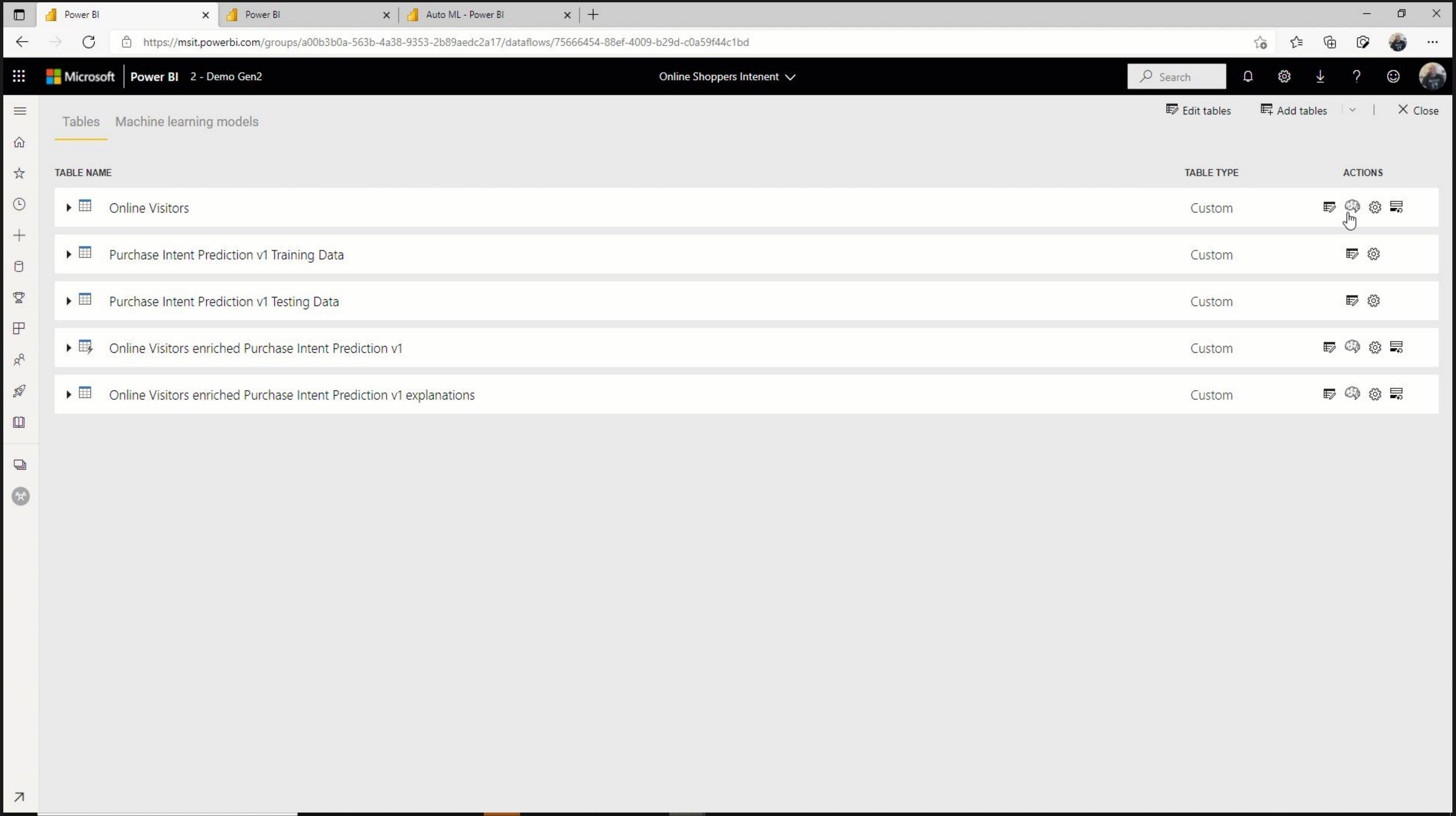
Extend with Azure ML



	categories	city	country	latitude	longitude	name	province	reviews_date	reviews_dateAdded	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 unwelc.
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 friendl.
30										

Data Scientist

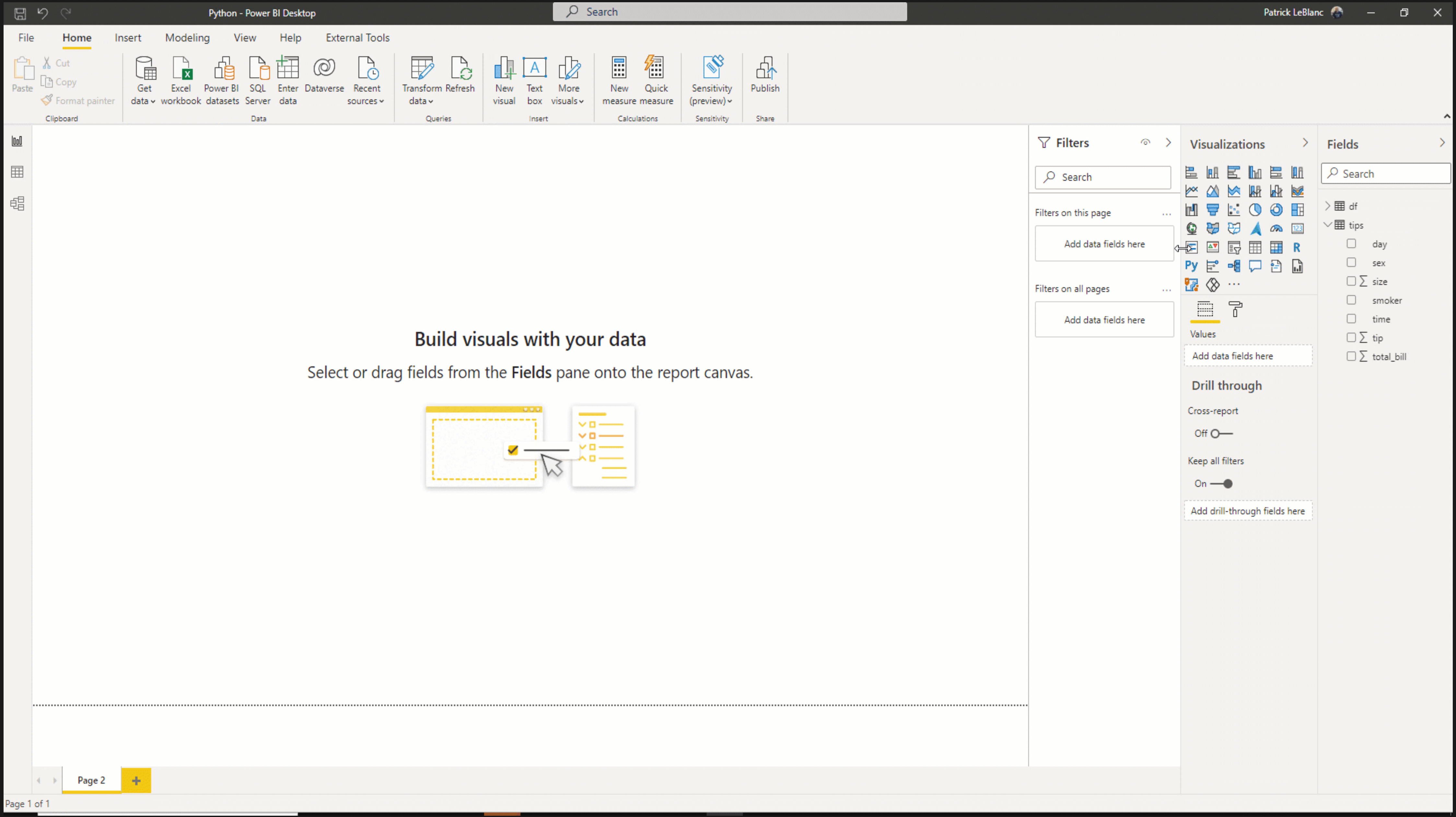
Automated ML



Empowering every individual
Augmented analytics

Data Scientist

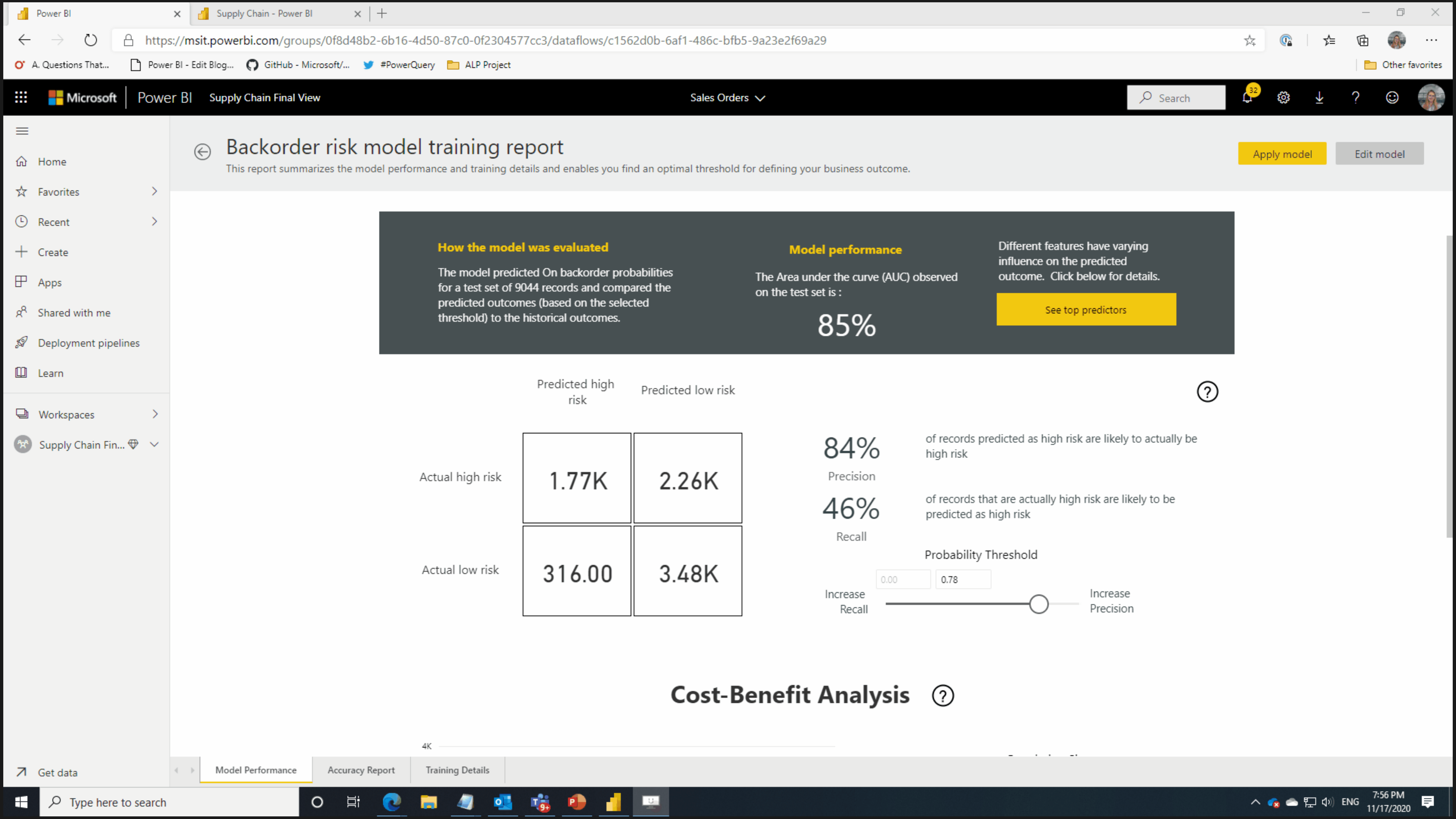
Python & R integration

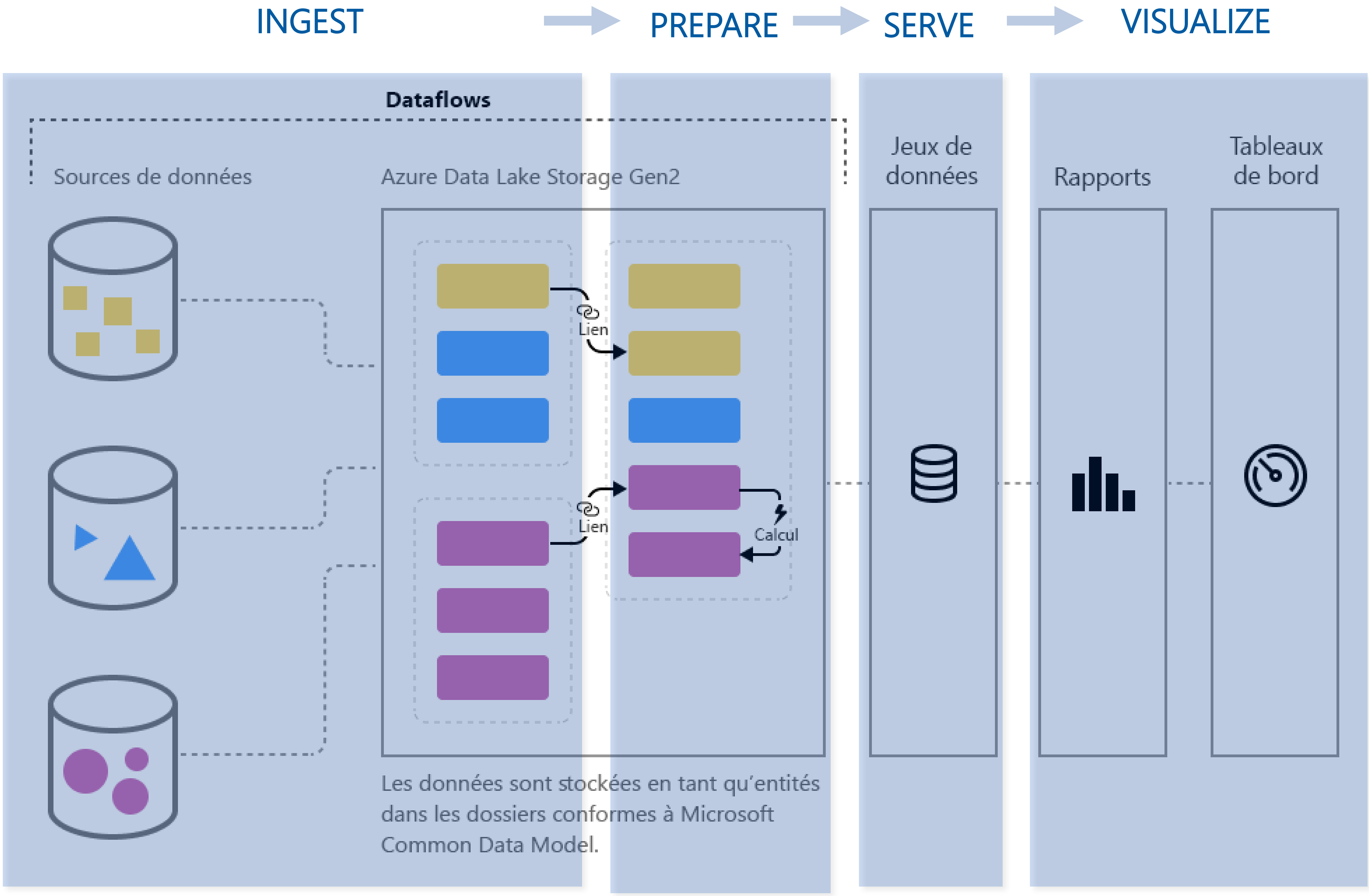


Empowering every individual
Augmented analytics

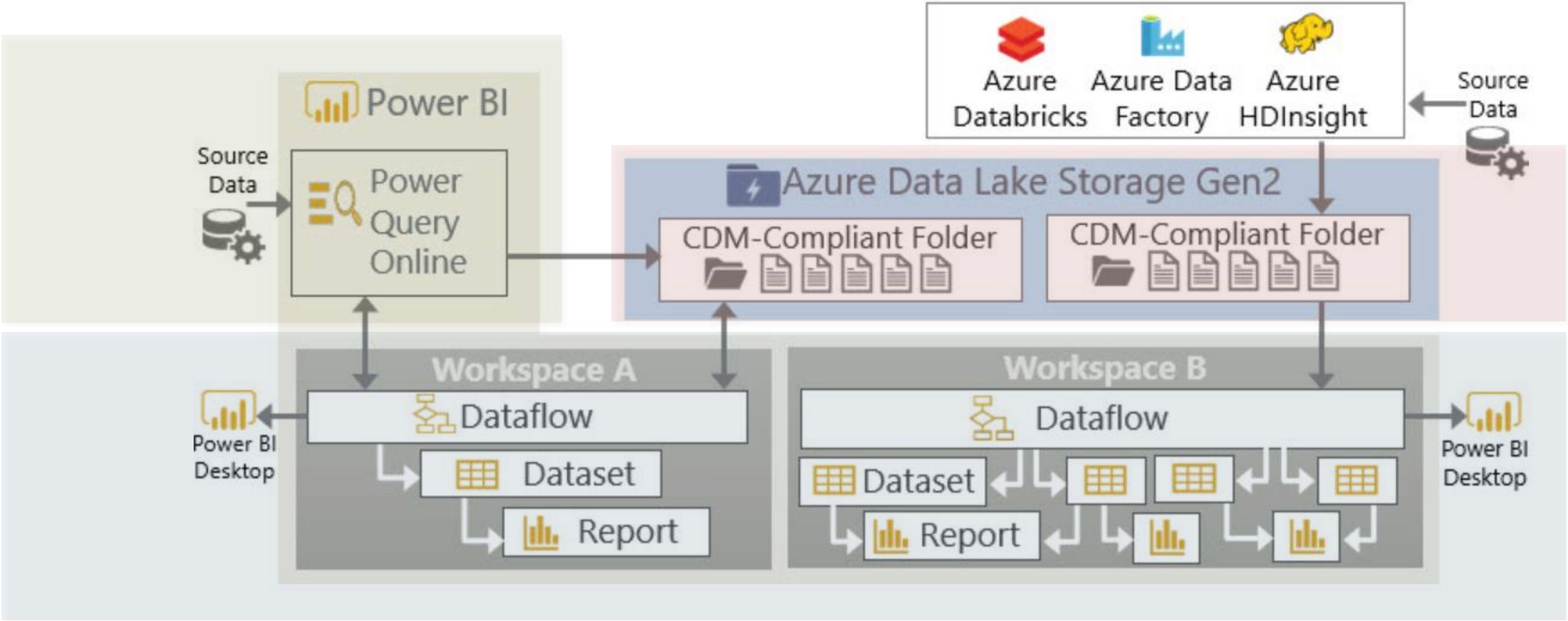
Data Scientist

Explore predictions




INGEST
PREPARE
SERVE
VISUALIZE
Dataflows
Sources de données
Azure Data Lake Storage Gen2
Jeux de données
Rapports
Tableaux de bord
Lien
Lien
Calcul
Les données sont stockées en tant qu'entités dans les dossiers conformes à Microsoft Common Data Model.

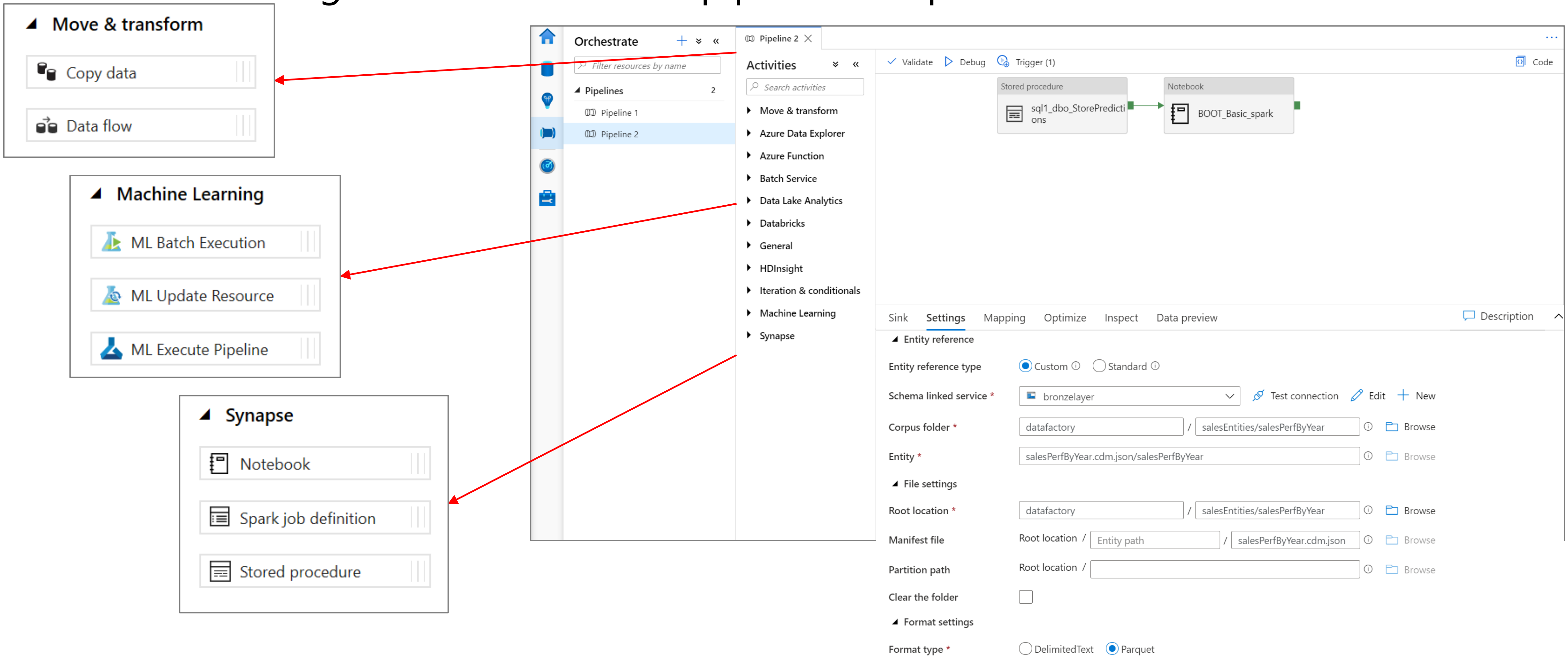
Data producer / Data consumer



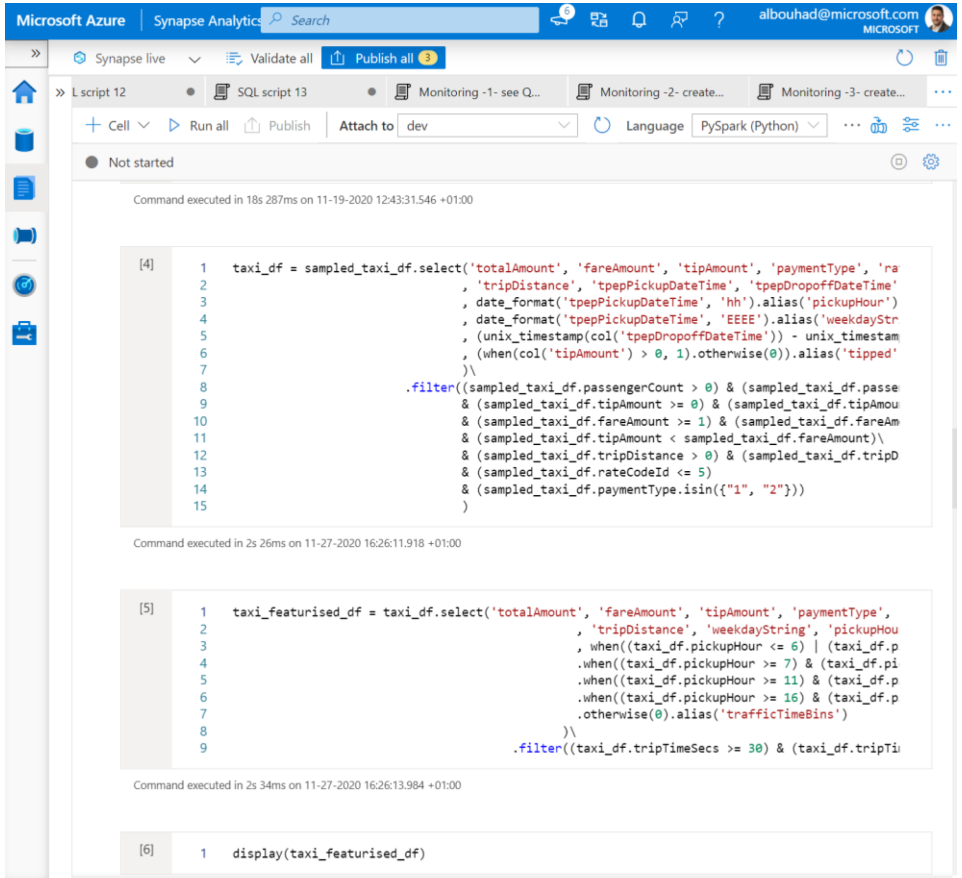
Common Data Model = Power BI + Synapse Pipelines

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

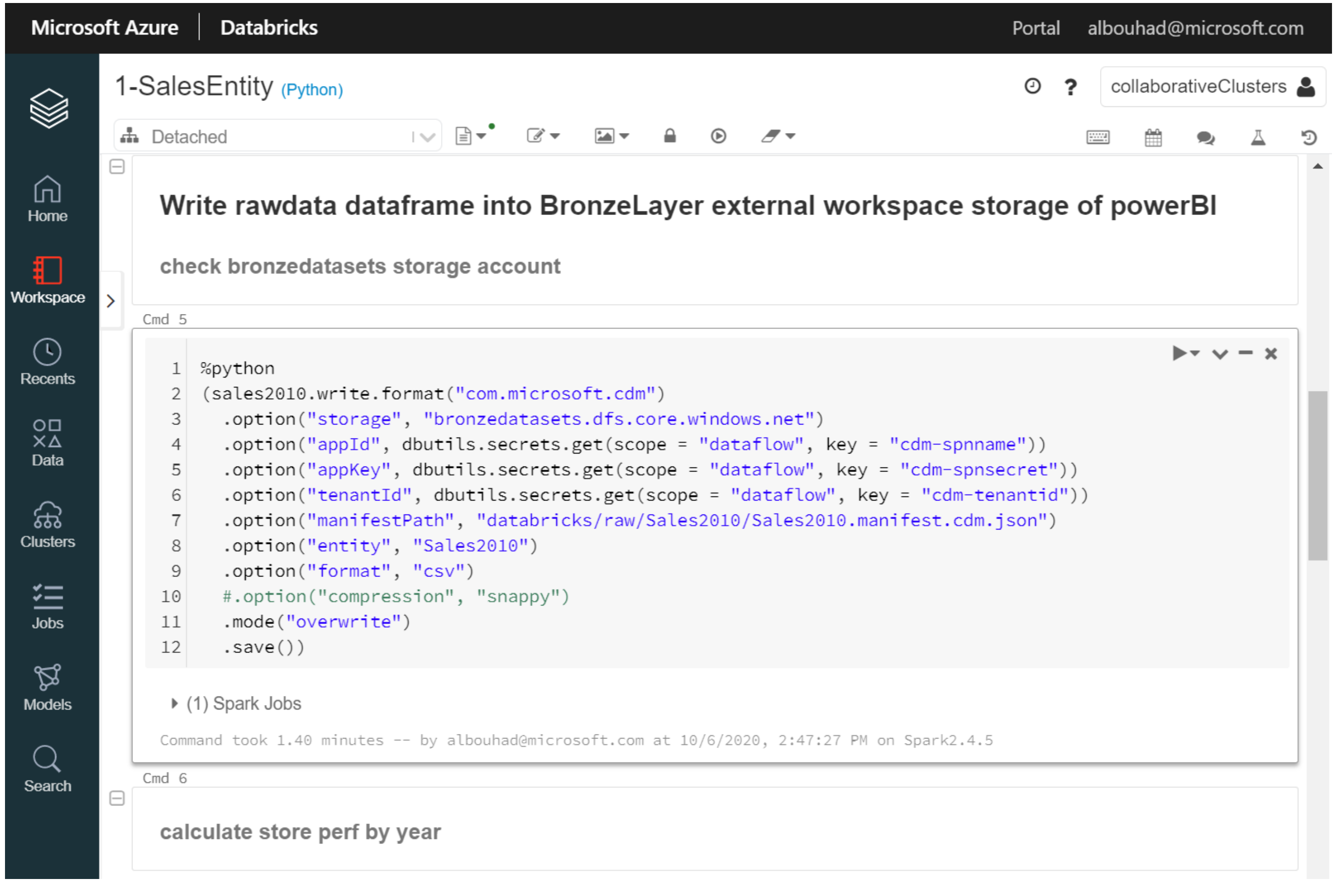
Offers a wide range of activities that a pipeline can perform.



Common Data Model = Power BI+ Spark

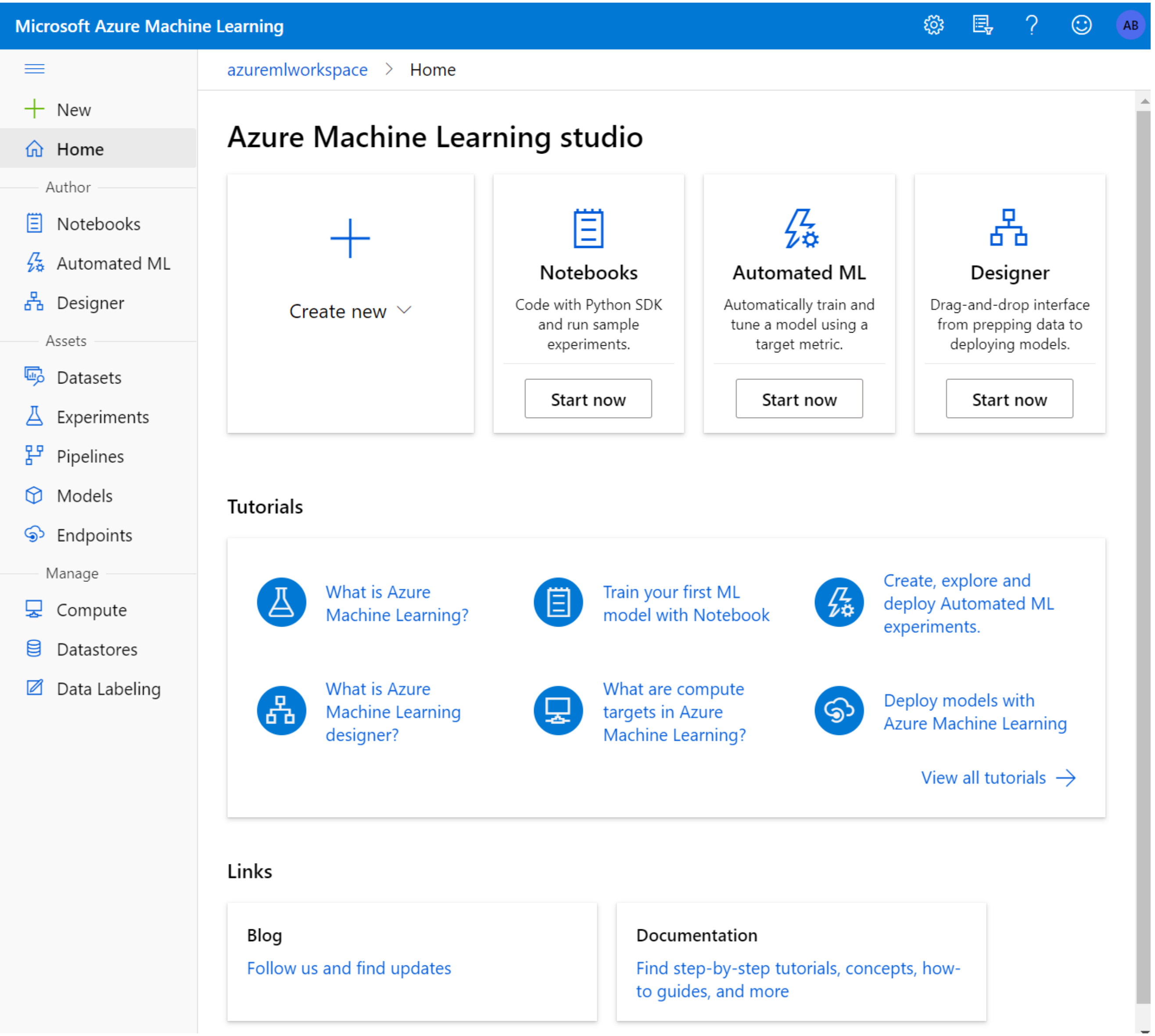


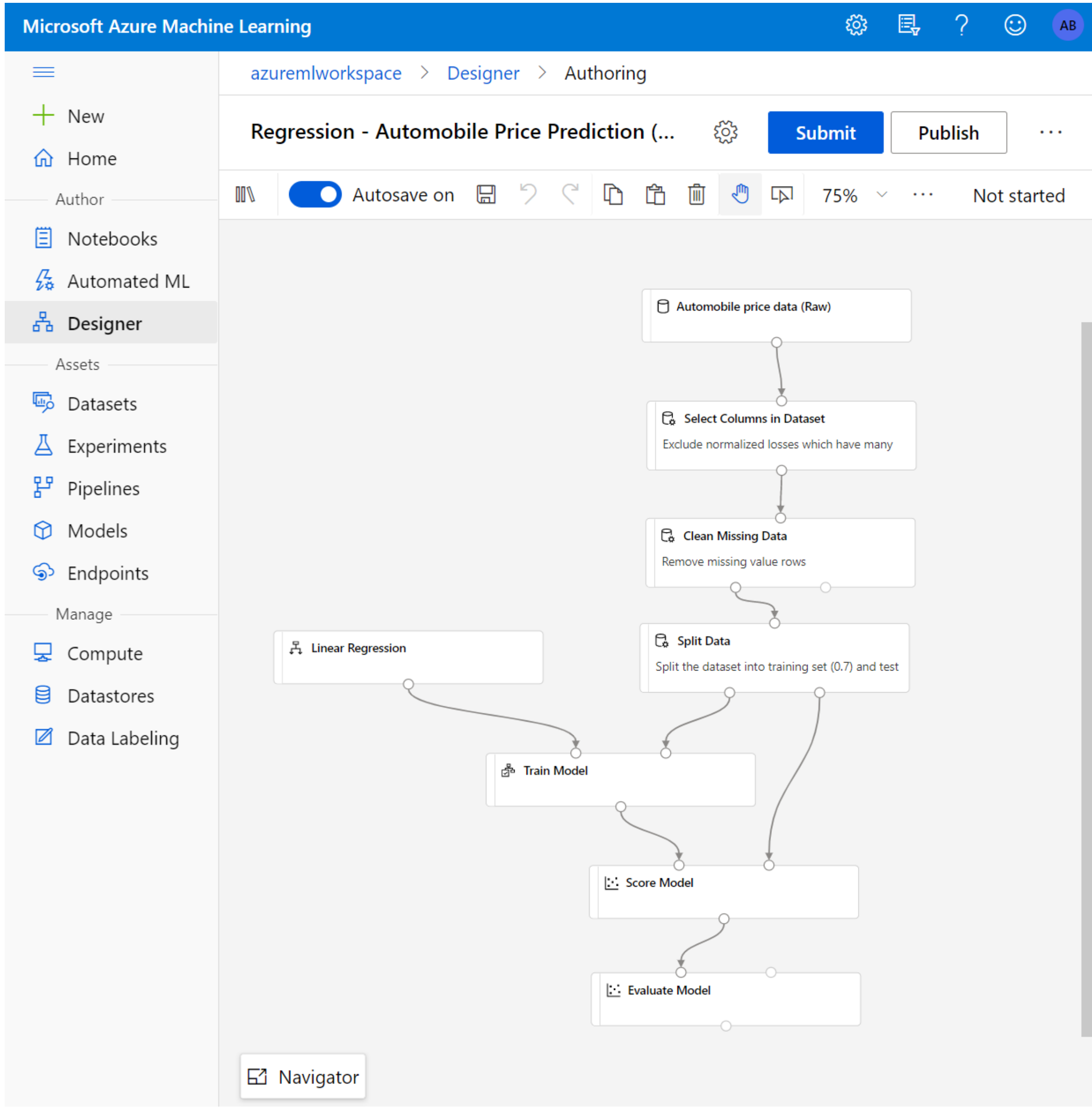
Synapse SparkPool



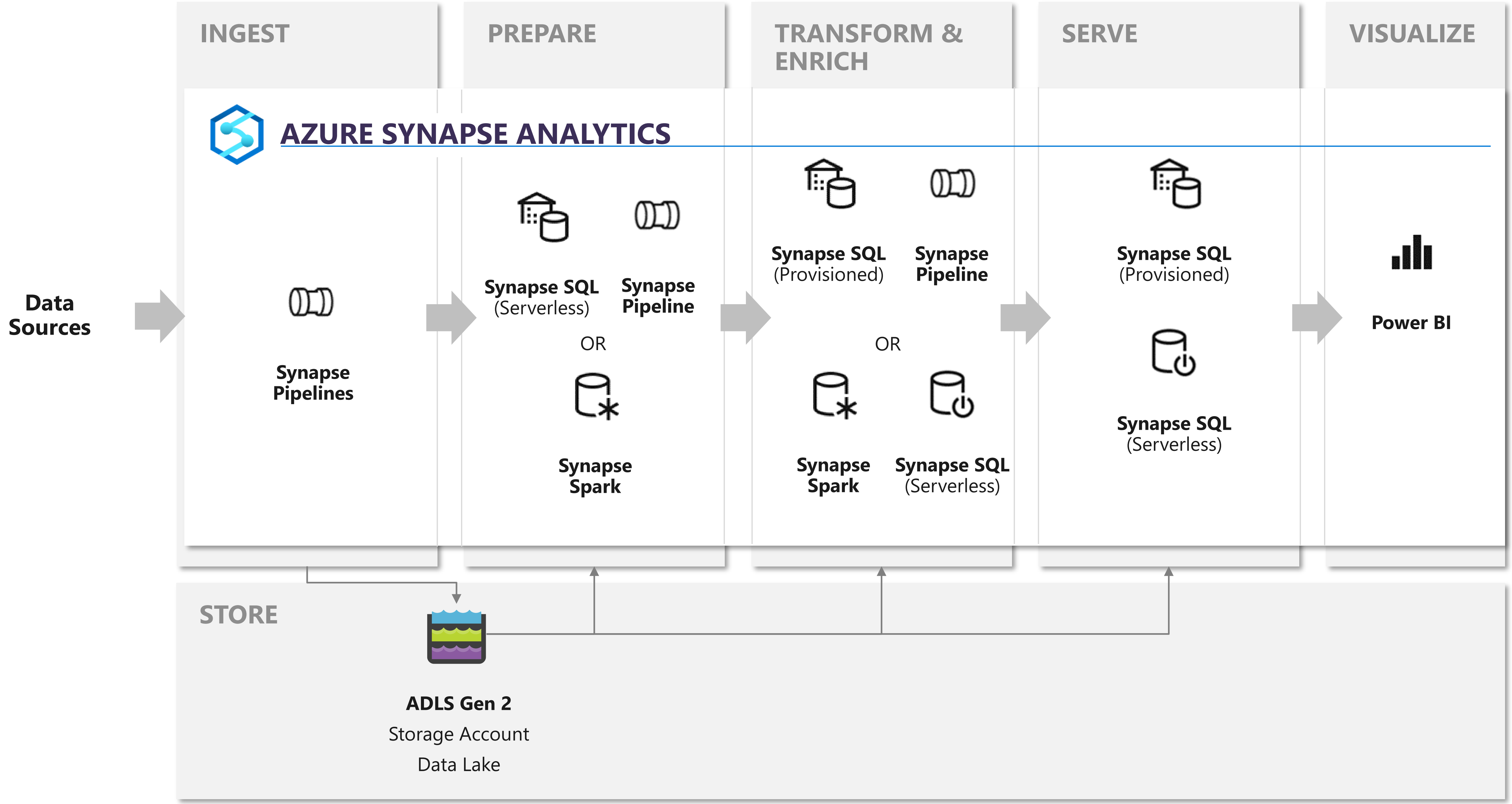
Azure Databricks

Common Data Model = Power BI+ AzureML





Common Data Model = Power BI+ Synapse Analytics



Power BI Builtin integration

SHARE WITH [Online/Offline]

- /
- Powerbi PRO
- Powerbi Embedded
- Powerbi Premium

VISUALIZE [Interactive]

- /
- Powerbi PRO
- Powerbi Embedded A1-A6
- Powerbi Premium

SERVE [Interactive]

Live Query

- Powerbi Premium P1
- Powerbi Embedded A1-A6
- Azure SQLDB

Direct Query

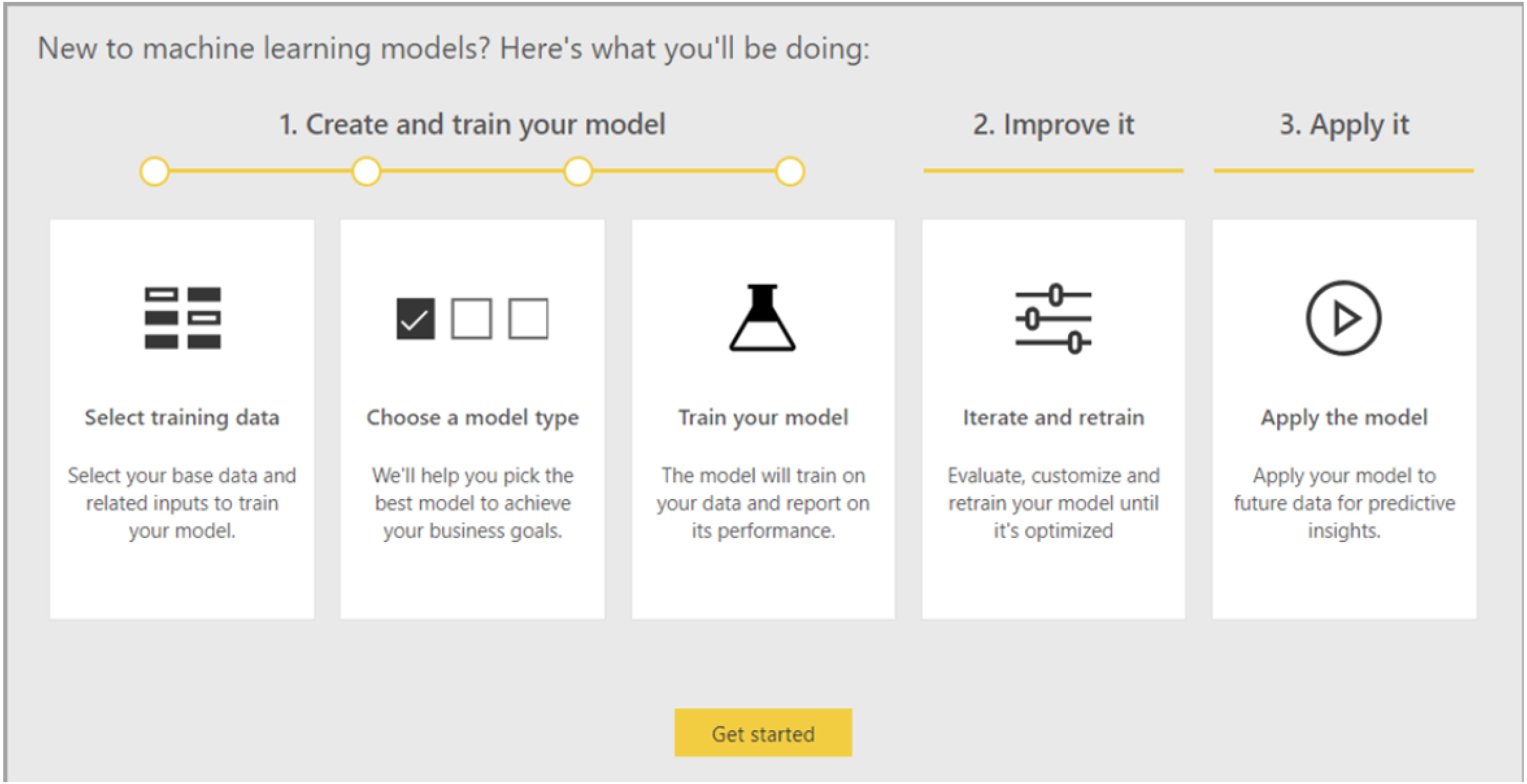
- Synapse SQLPool
- Synapse SQL Serverless

PREPARE [batch]

- Powerbi Embedded A3 [*integrated DF]
- Powerbi Embedded A4-A6 [*integrated DF, [*integrated ML]
- Synapse Mapping/wrangling Dataflows
- Synapse SQL Serverless
- Synapse notebooks / Databricks

INGEST [batch]

- Powerbi Embedded A3
- Synape Pipelines
- Synapse notebooks / Databricks


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.
Get started

Thank you!

Microsoft

Enregistrez vous dès maintenant au prochain Webinars Data AI

Event Webinar (Les jeudis de la Data & AI) - L200/300	Date	Duration (min)	Link
Azure Synapse	22/09/2022	120	https://msevents.microsoft.com/event?id=857781749
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	29/09/2022	120	https://msevents.microsoft.com/event?id=502366997
Déploiement et sécurisation des workspaces Azure Machine learning	06/10/2022	120	https://msevents.microsoft.com/event?id=1505714138
Azure Scale Analytics - Architectures Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share	13/10/2022	120	https://msevents.microsoft.com/event?id=139685175
MLOps avec Azure Machine Learning	20/10/2022	120	https://msevents.microsoft.com/event?id=1245885767
SQL Server 2022 et hybridation native avec Azure SQL Managed Instance	10/11/2022	120	https://msevents.microsoft.com/event?id=145826476
Machine Learning dans Azure Synapse Analytics	17/11/2022	120	https://msevents.microsoft.com/event?id=3637723312
Azure Cosmos DB et IA	24/11/2022	120	https://msevents.microsoft.com/event?id=2646013445
Azure et les Services Cognitifs	08/12/2022	120	https://msevents.microsoft.com/event?id=3772037220
La gouvernance de données dans Azure avec Microsoft Purview	15/12/2022	120	https://msevents.microsoft.com/event?id=1499560981
MLOps avec Azure Machine Learning	12/01/2023	120	https://msevents.microsoft.com/event?id=4115194515
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	19/01/2023	120	https://msevents.microsoft.com/event?id=1537241181
Déploiement et sécurisation des workspace Azure Synapse	26/01/2023	120	https://msevents.microsoft.com/event?id=1806467748
Azure Machine Learning pour les Citizen Data Scientists	09/02/2023	120	En cours
PowerBI - Self Service Analytics	16/02/2023	120	https://msevents.microsoft.com/event?id=1401519679
L'IA responsable avec Azure machine learning	09/03/2023	120	https://msevents.microsoft.com/event?id=2072953112
Machine Learning dans Azure Synapse Analytics	16/03/2023	120	https://msevents.microsoft.com/event?id=3413014857
Les bases de données Open Source dans le cloud Azure	23/03/2023	120	https://msevents.microsoft.com/event?id=2727487131
Hybridation des services de Machine Learning Azure	06/04/2023	120	https://msevents.microsoft.com/event?id=1624914222
La gouvernance de données dans Azure avec Microsoft Purview	13/04/2023	120	https://msevents.microsoft.com/event?id=3909342839
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	04/05/2023	120	https://msevents.microsoft.com/event?id=1162207895
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	16/05/2023	120	https://msevents.microsoft.com/event?id=3517068442
Hybridation des services de données Azure	24/05/2023	120	https://msevents.microsoft.com/event?id=2996507398
Self Service Analytics	01/06/2023	120	En cours