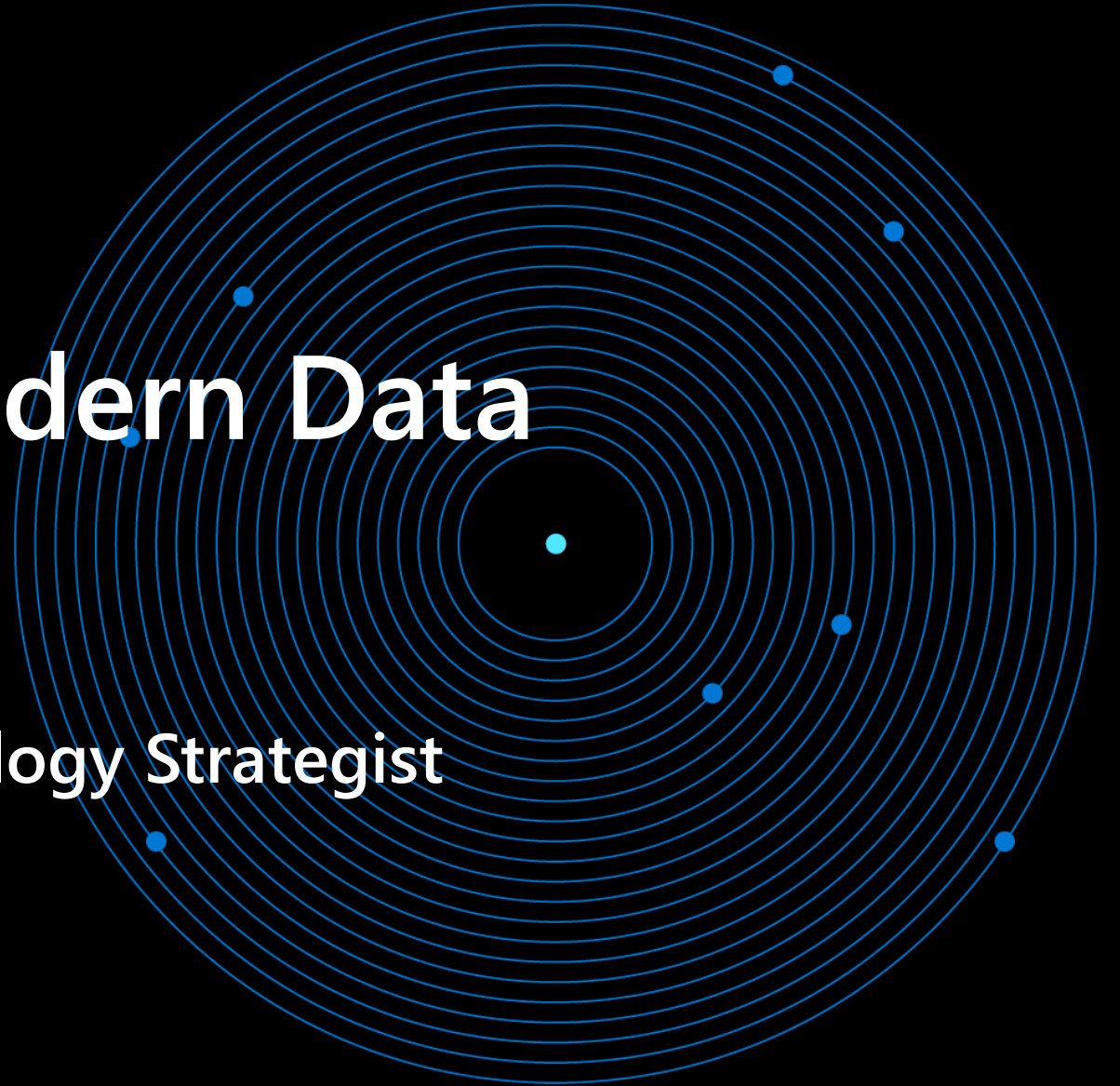




# Microsoft Azure Modern Data Platform Sharing

Methasit Tuwawit, Partner Technology Strategist



# Today's data realities



Volume



Variety



Velocity

What **data** do I have?

Is it **trustworthy**?

Can people access the **data** needed  
to make the right decisions?

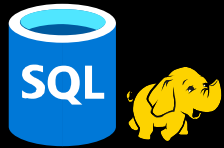
How can I **enable faster**  
**business insights**?

What's my **compliance exposure**?

# Common customer use cases

Big Data and advanced analytics

## Modern data warehouse



"Integrate all our data—including Big Data—with our data warehouse for analytics and reporting"



## Advanced analytics



"Predict next best offer and customer churn"



## Real-time analytics



"Derive insights from our devices and data streams in real-time"

# Common customer use cases

## Modern data warehouse

## Advanced analytics

## Real-time analytics

### Sources (available to migrate to Azure)

- SQL, MySQL, PostgreSQL
- SAP on Azure
- Oracle to PostgreSQL
- File storage/Backup
- DB2, AS/400 Migration



### Data Analytics Use Cases

- Sales Forecast
- Customer Segmentation
- Customer Lifetime Value
- Churn Prediction
- Recommendation Effectiveness
- Promotion
- Cross-sell/Upsell
- Enterprise Search



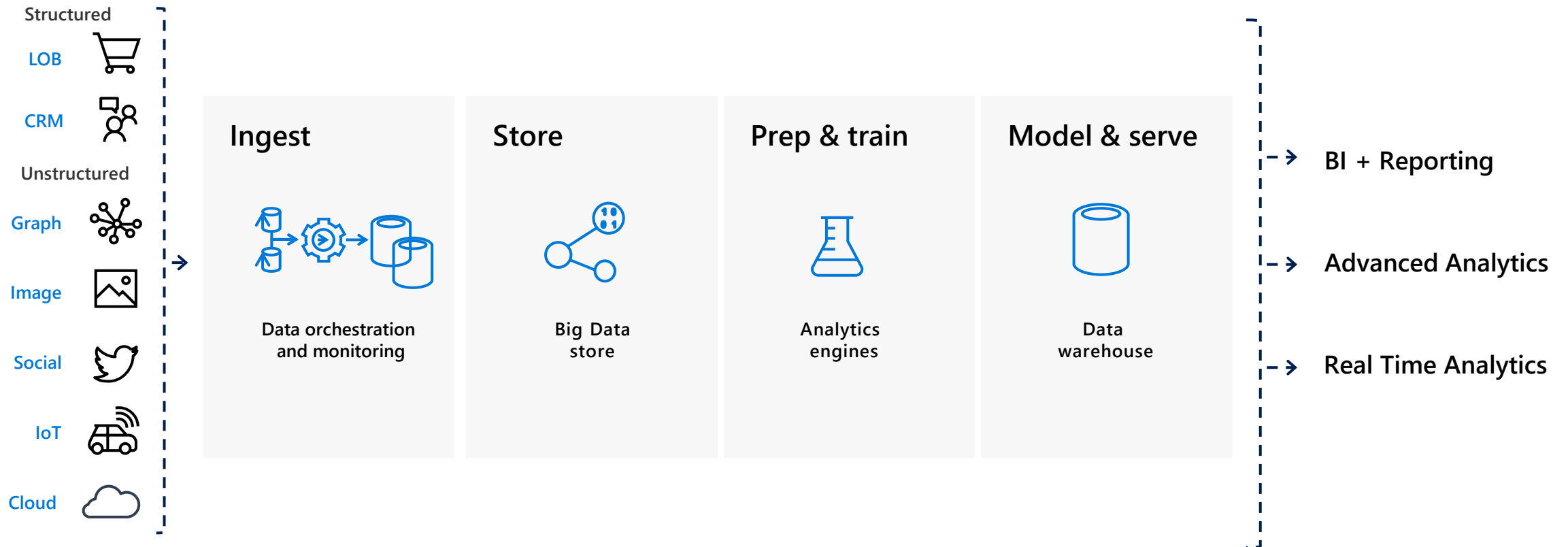
### Real-time Analytics Use Cases

- Dynamic Pricing
- Fraud Detection
- Predictive Maintenance
- Workplace Security (CCTV)
  - Face recognition
  - Security alert
- Digital Twin (simulation)

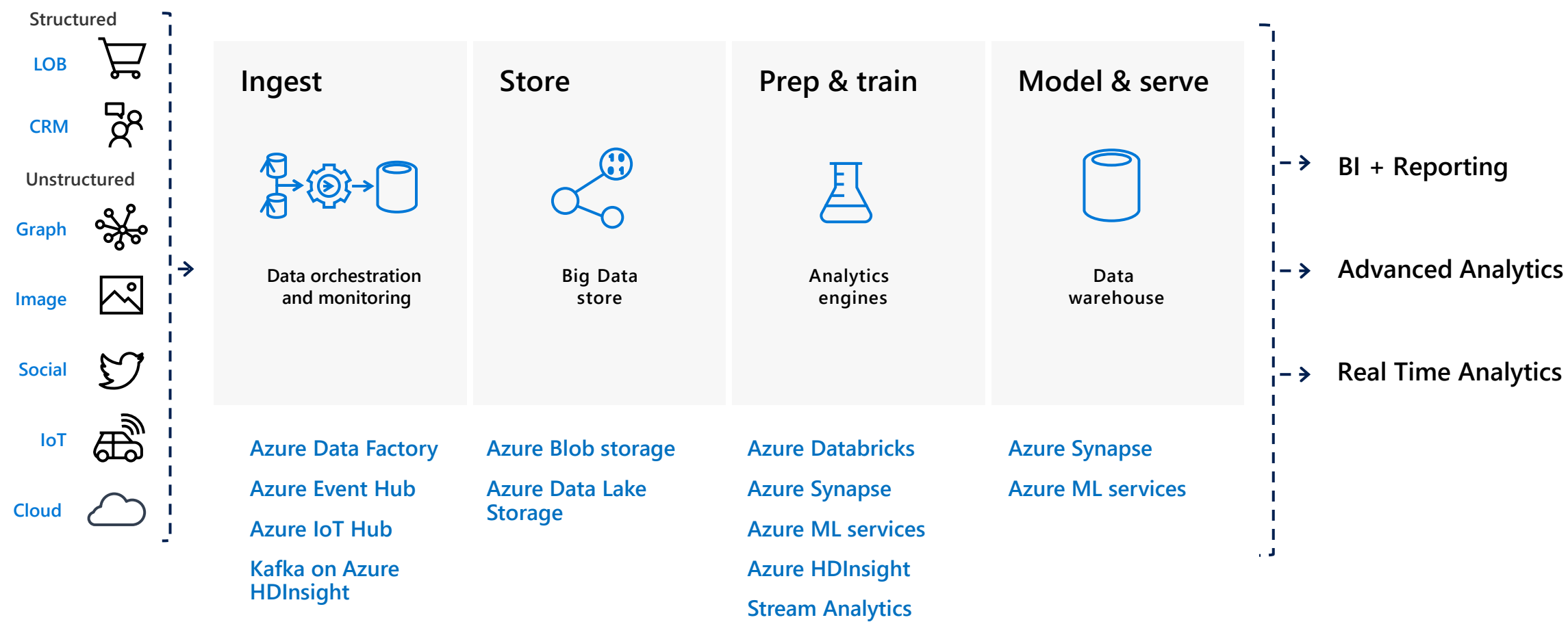
### Centralized Big Data

- Modern Data Warehouse
- PDPA
- Interactive Dashboard & Self-service BI

# How data flow?



# Data Platform



# The Azure data landscape



Azure Data Factory



Azure Import/Export service



Azure CLI



Azure SDK



Azure IoT Hub



Azure event hubs



Kafka on Azure HDInsight



Azure SQL DB



Azure Cosmos DB



Azure Blob Storage



Azure Data Lake Store



Azure Search



Azure Data Catalog



Azure Synapse Analytics  
(formerly SQL DW)



Azure Data Lake Analytics



Azure HDInsight



Azure Databricks



Azure Stream Analytics



Azure HDInsight



Azure Databricks



Azure Analysis Services



Power BI



Azure ML



ML Server



Azure Databricks



Bot service



Cognitive services



Azure ExpressRoute



Azure Active Directory



Azure network security groups



Azure key management service



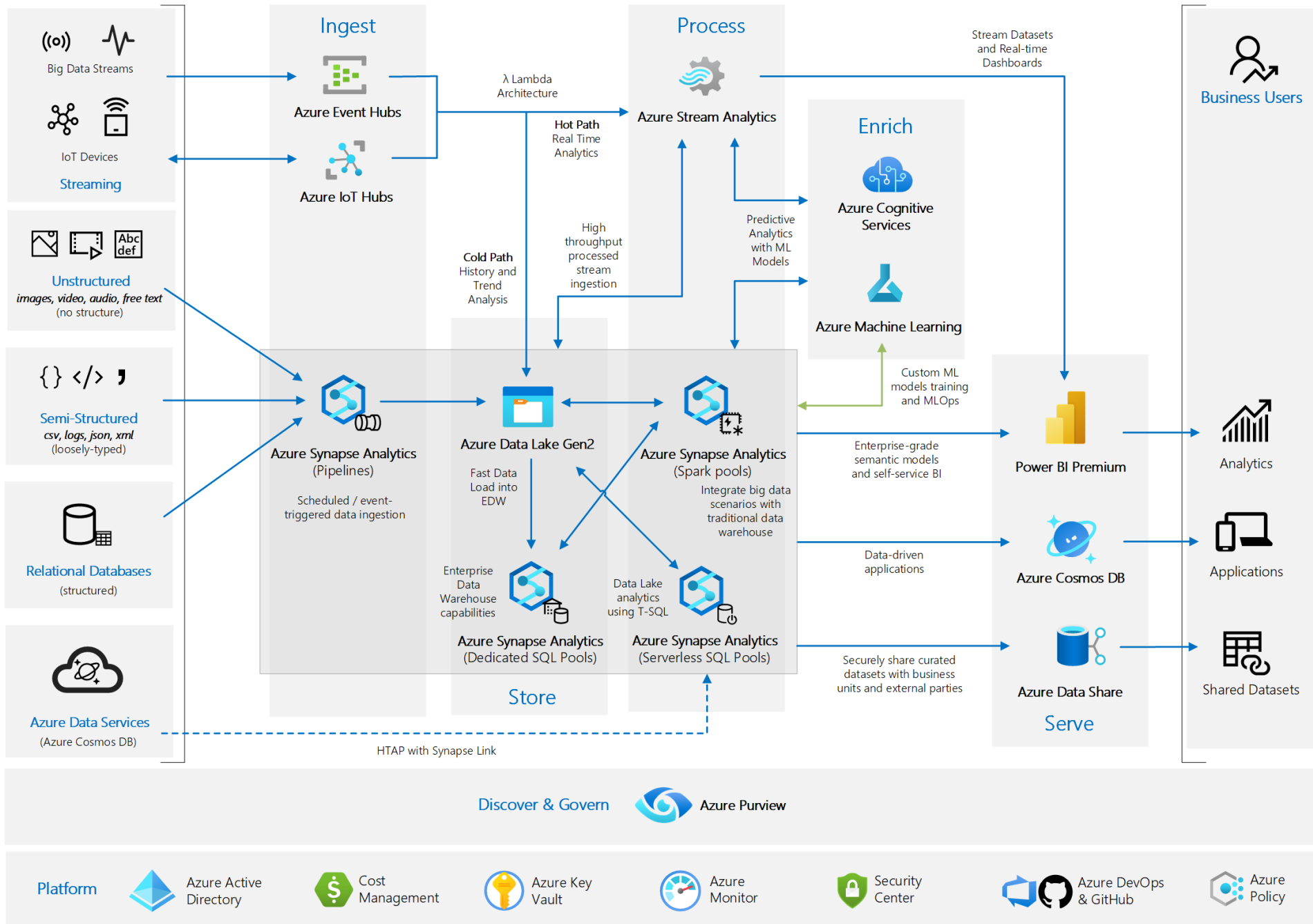
Operations Management Suite



Azure Functions

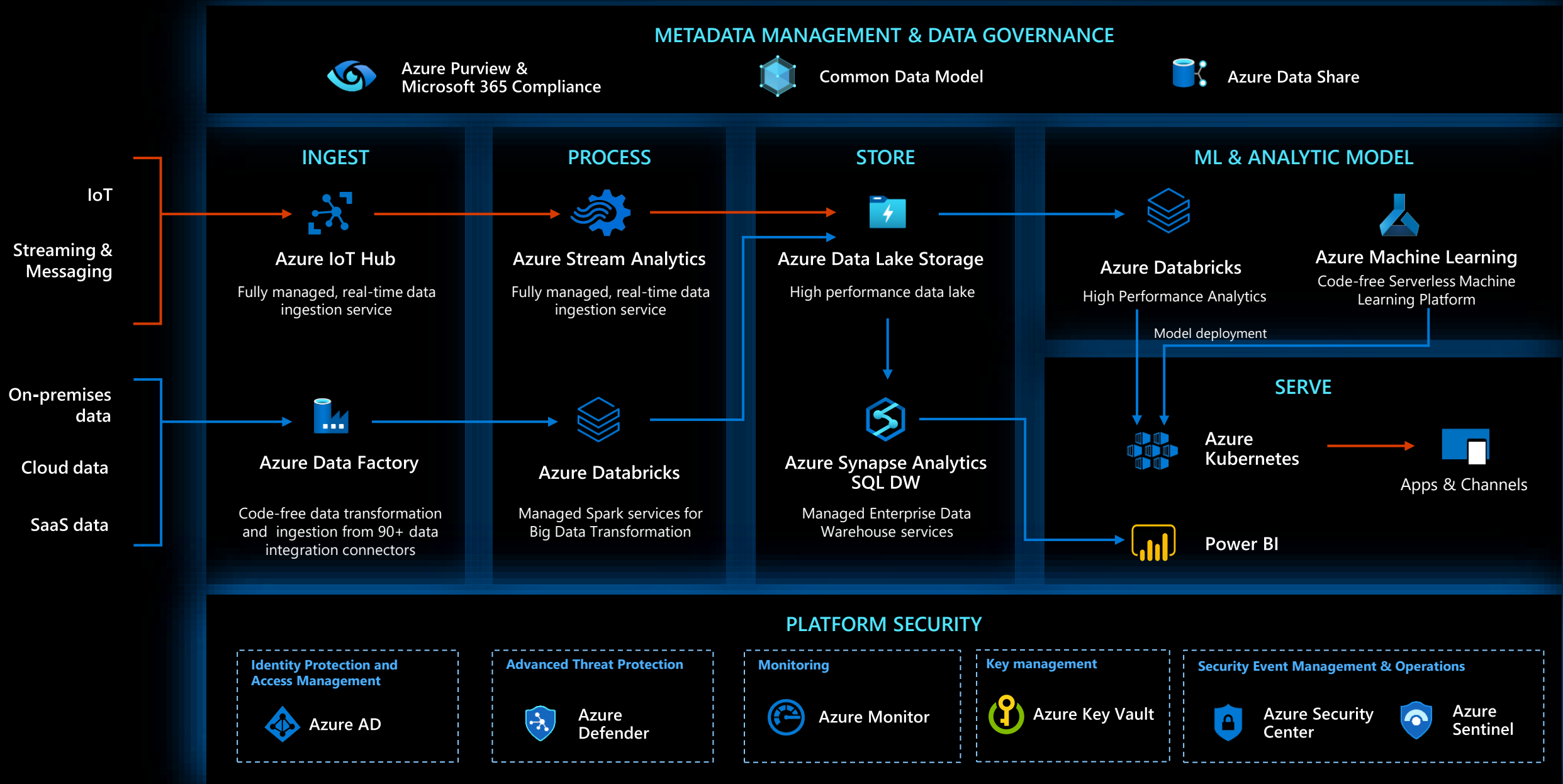


Visual Studio

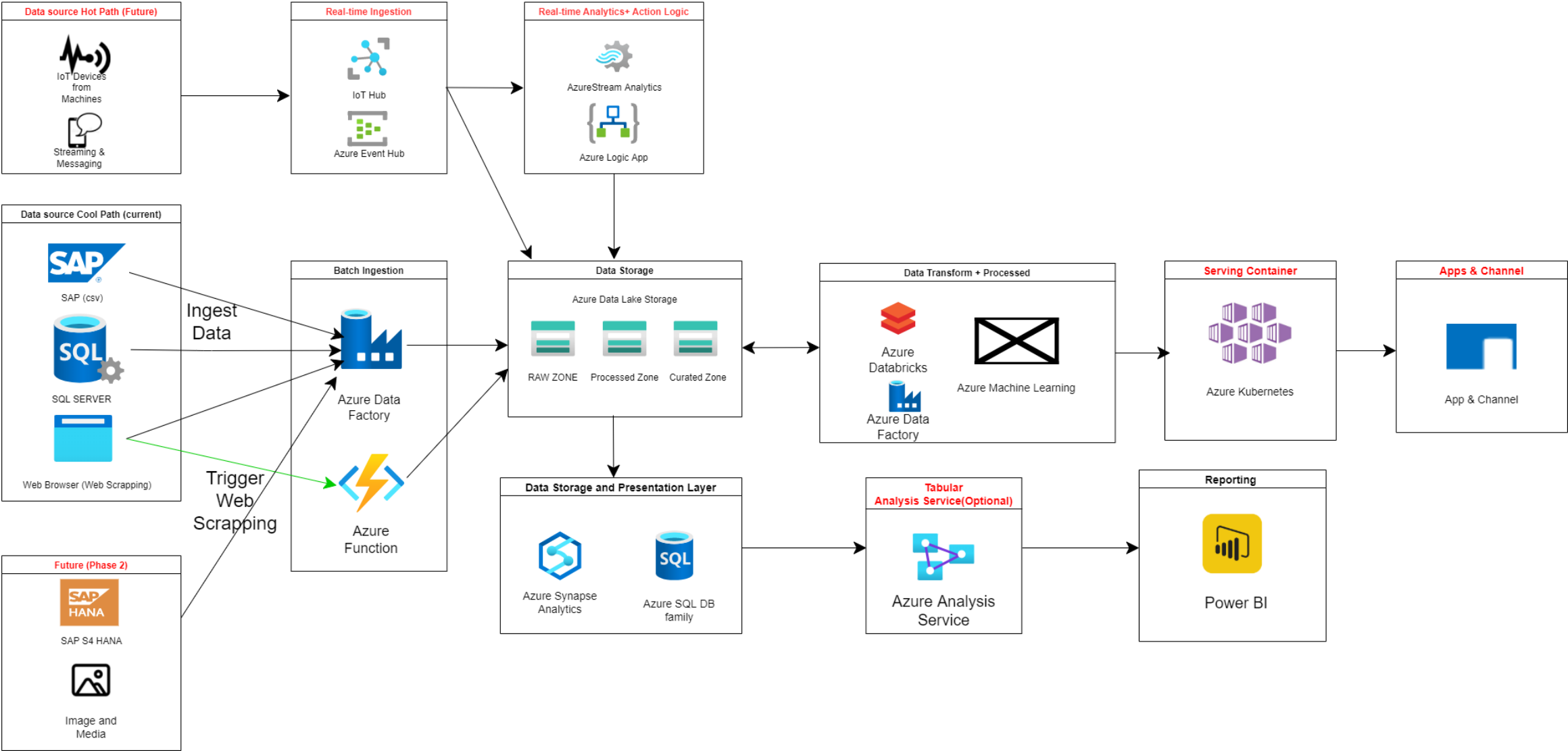




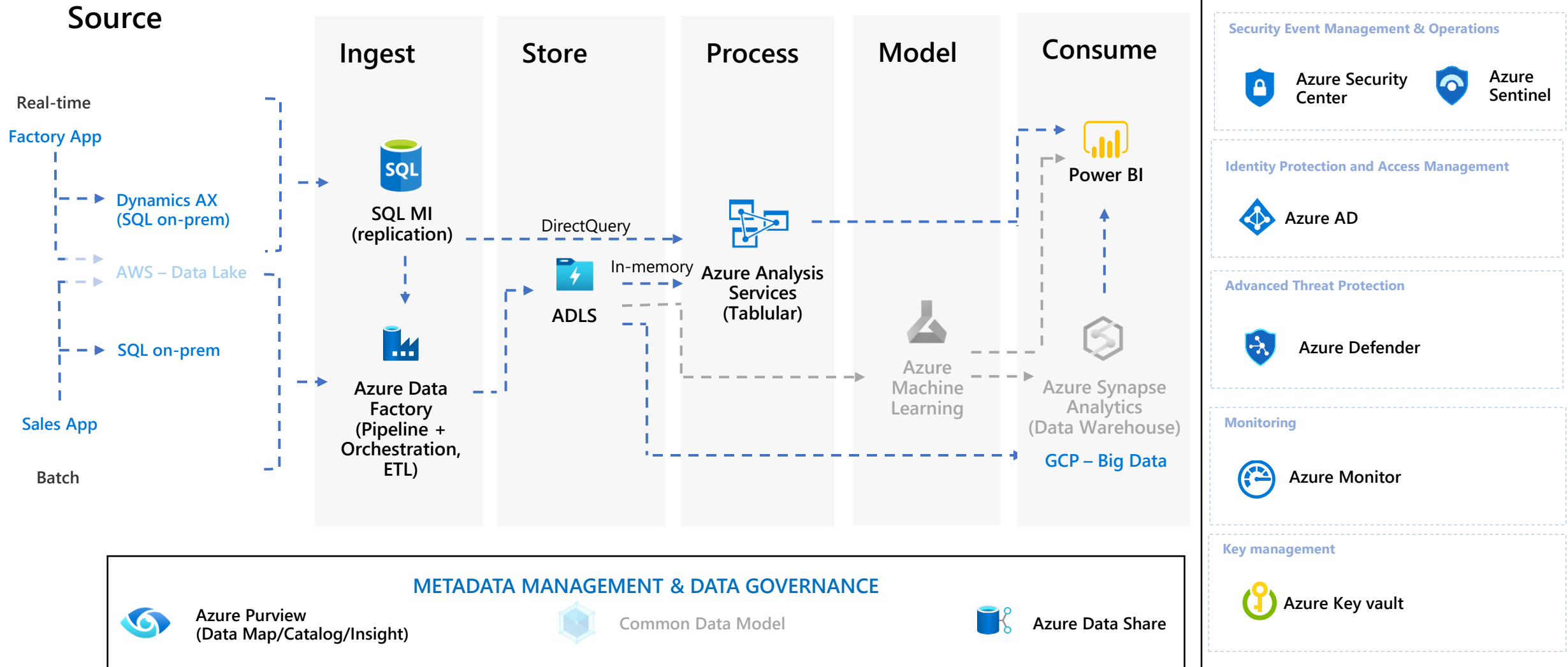
# Azure Modern Data Warehouse and Analytics Reference Architecture



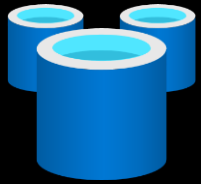
Drafted Data Architecture



# Lake House – Reference Architect



# Data + Analytics



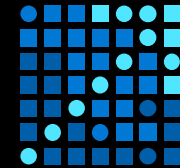
Relational



Non-relational



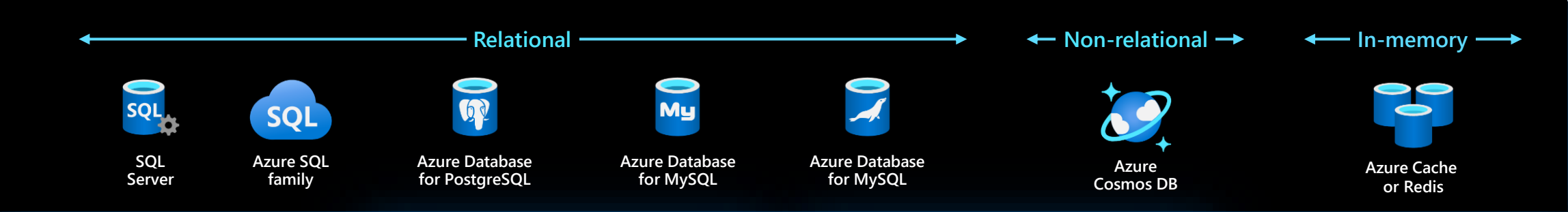
Data Warehouse



Big Data

# Data Modernization to Azure

A Comprehensive Operational Data Portfolio




**65 million**  
downloads of SQL Server

**7 million**  
databases in Azure SQL

**100 trillion**  
transactions per day on  
Azure Cosmos DB

**17 billion**  
transactions per day on one app  
in Azure Database for MySQL



Seamless database migration with Azure Migrate and Azure Database Migration Service

# Azure SQL

Run any workload anywhere on the industry-leading SQL Server engine

## Infrastructure-as-a-Service



Best for lift and shift and/or workloads requiring OS-level access

SQL Server on Azure Virtual Machines

## Platform-as-a-Service



Best for modernizing existing apps

Azure SQL Managed Instance



Best for supporting modern cloud apps

Azure SQL Database



Fully-managed and always up to date



Use your existing SQL experience in the cloud



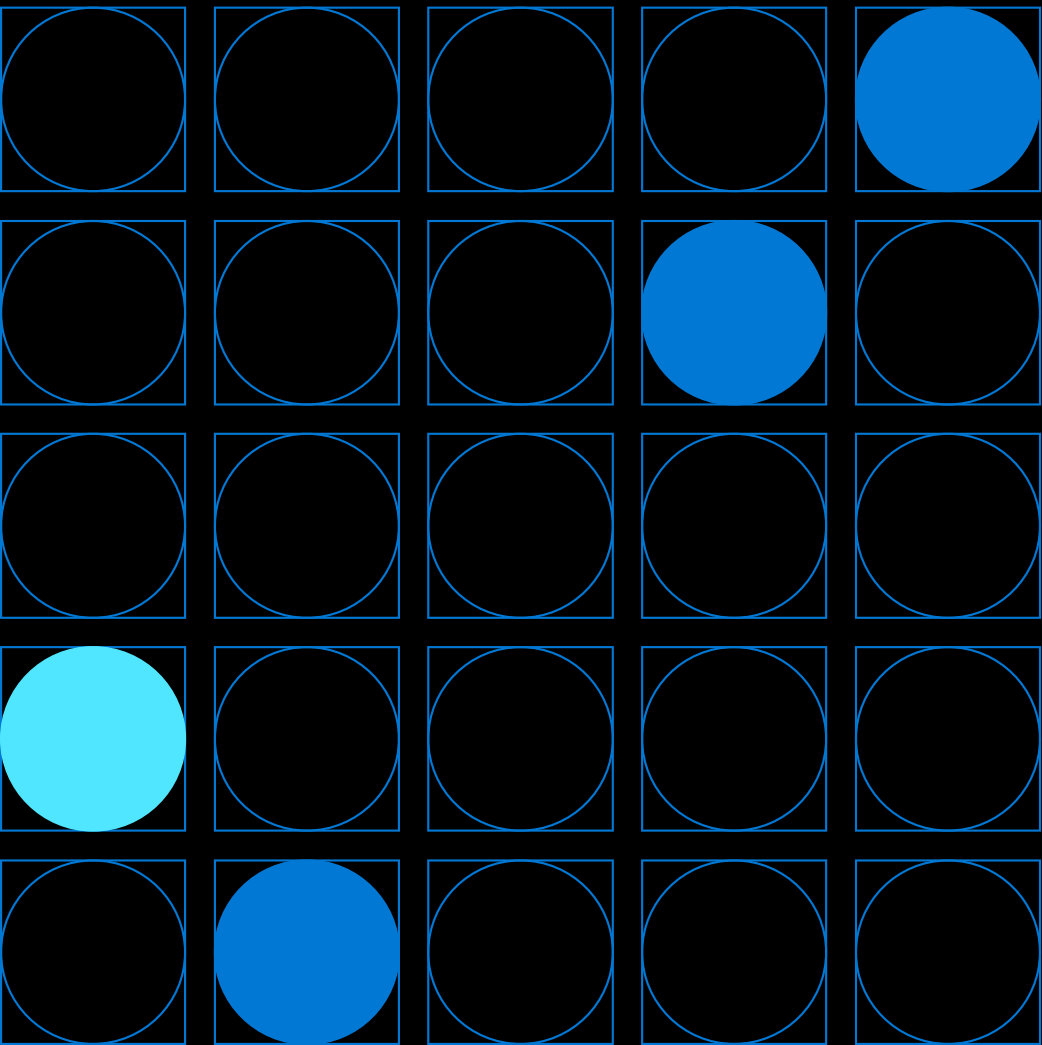
Save with the lowest total cost of ownership



Protect your data with built-in, real-time intelligent security

**The only cloud with evergreen SQL**

Software is continually updated and never requires patches



**Azure Data Factory**

# Azure Data Factory

## Data Integration Service: Serverless, Scalable, Hybrid

The screenshot displays the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' pane shows a hierarchy of Pipelines (18), Datasets (37), and Data Flows (0). The 'ImproveOnlineTrafficSales' pipeline is selected. The main canvas shows the pipeline's control flow, which includes several 'Copy Data' activities (Location\_HTTP, Customer\_Dynamics, Customer\_Salesforce, Products\_SQL) and a 'Web' activity (SendFailEmail). The pipeline is triggered by a 'File arrival trigger'. The bottom pane shows the 'General' tab for the selected pipeline, with fields for Name, Description, and Annotations.

- Control Flow with 25+ activities and growing
- File arrival trigger, Schedule roll up Trigger, Trigger Dependencies, Chaining trigger
- Activity Re-run



# Azure Data Factory

90+ including data connectors

Azure (15)	Database & DW (25)		File Storage (6)	NoSQL (3)	Services & Apps (28)		Generic (4)
Blob Storage	Amazon Redshift	Oracle	Amazon S3	Cassandra	Amazon MWS	Oracle Service Cloud	HTTP
Cosmos DB – SQL API	DB2	Phoenix	File System	Couchbase	CDS for Apps	Paypal	OData
Cosmos DB – MongoDB API	Drill	PostgreSQL	FTP	MongoDB	Concur	QuickBooks	ODBC
ADLS Gen1	Google BigQuery	Presto	Google Cloud Storage		Dynamics 365	Salesforce	REST
ADLS Gen2	Greenplum	SAP BW Open Hub	HDFS		Dynamics AX	SF Service Cloud	
Data Explorer	HBase	SAP BW MDX	SFTP		Dynamics CRM	SF Marketing Cloud	
Database for MariaDB	Hive	SAP HANA			Google AdWords	SAP C4C	
Database for MySQL	Impala	Spark			HubSpot	SAP ECC	
Database for PostgreSQL	Informix	SQL Server			Jira	ServiceNow	
File Storage	MariaDB	Sybase			Magento	Shopify	
SQL Database	Microsoft Access	Teradata			Marketo	Square	
SQL Database MI	MySQL	Vertica			Office 365	Web Table	
SQL Data Warehouse	Netezza				Oracle Eloqua	Xero	
Search Index					Oracle Responsys	Zoho	
Table Storage							

# Azure Data Factory

## Code-free data transformations in Azure Data Factory

The screenshot displays the Azure Data Factory (ADF) web interface. On the left, the 'Factory Resources' pane shows a list of pipelines, datasets, and data flows. The 'Data Flows (Preview)' section is expanded, showing a list of data flows including 'dataflow2' and 'dataflow3'. The main canvas shows a data flow named 'dataflow3' with the following activities:

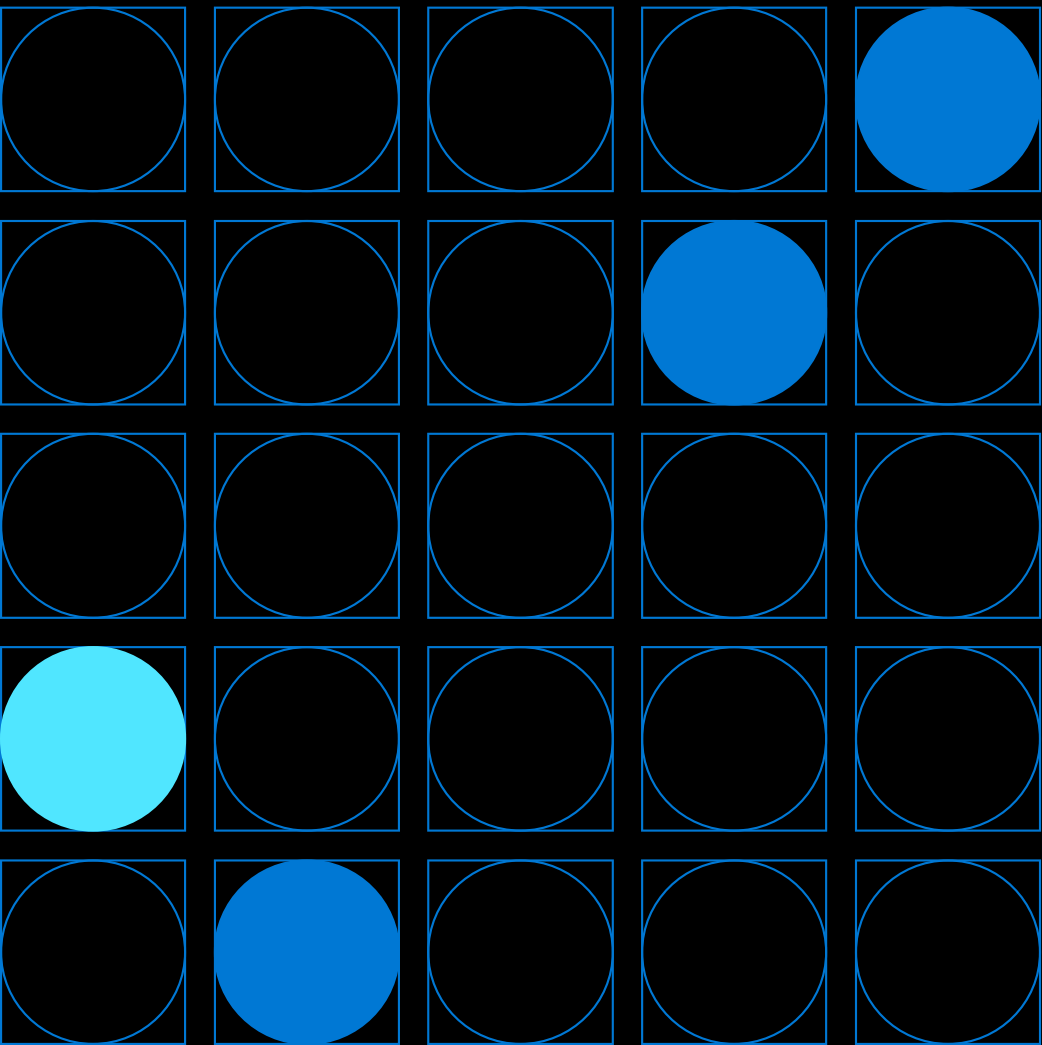
- badgesFromCsv**: Import data from src\_BadgesBlobWithHeader.
- JoinByUserId**: Inner join on badgesFromCsv and selectUsersCol.
- AggrByAge**: Aggregating data by 'Age, Name' producing columns 'rowCount'.
- AgeNotNull**: Filtering rows using expressions on columns 'Age'.
- sink1**: Export data to csv\_output3.

A detailed view of the 'selectUsersCol' activity is shown, indicating it renames usersFromDatabase to selectUsersCol with columns 'Id, Age, CreationDate, DisplayName, DownVotes'.

Below the main canvas, the 'General' tab of the 'dataflow3' activity is visible, showing the name 'dataflow3' and a description field.

### Available Mapping Activities

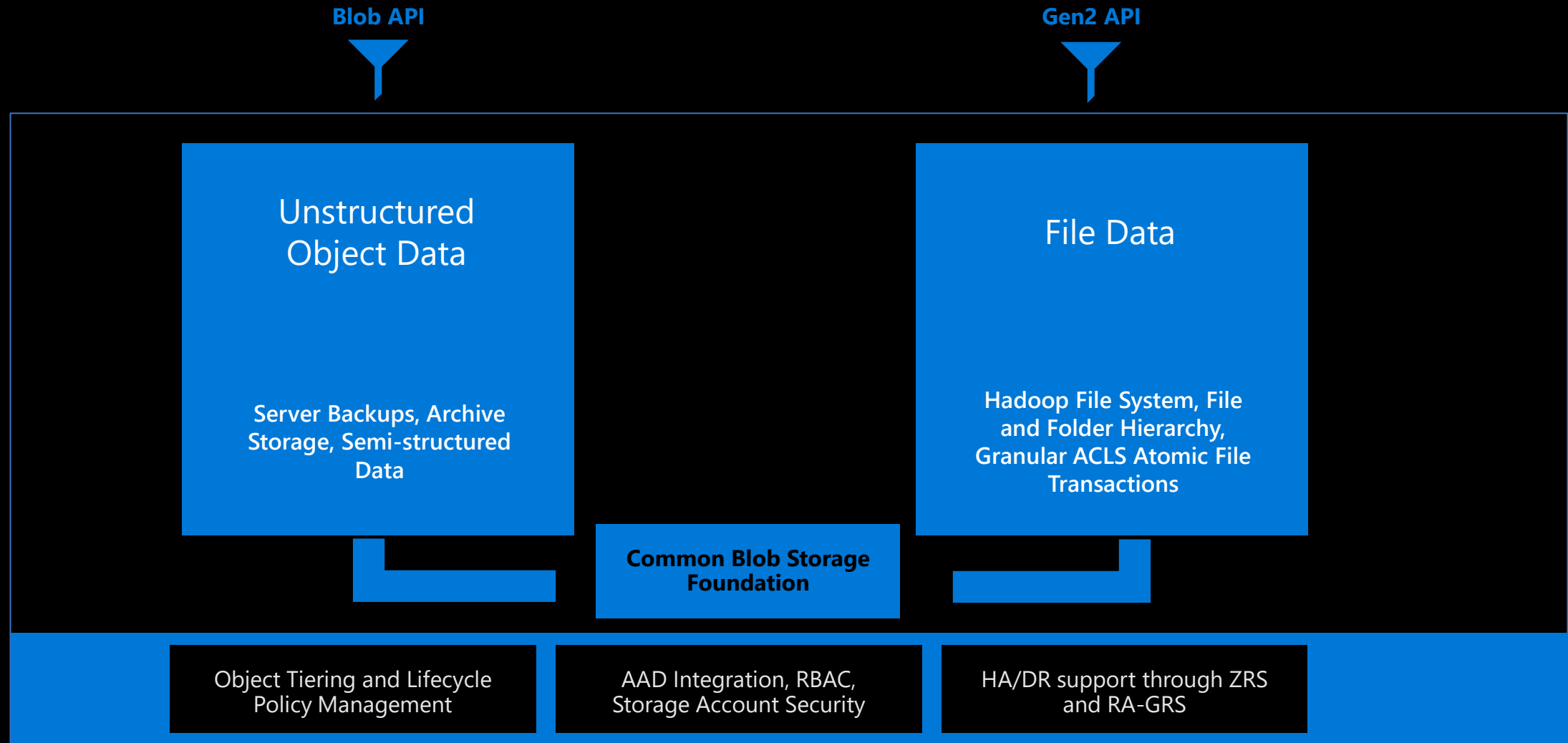
- Source
- New Branch
- Join
- Conditional Split
- Union
- Lookup
- Derived Column
- Surrogate Key
- Pivot
- Unpivot
- Window
- Exists
- Select
- Filter
- Sort
- Alter Row
- Sink



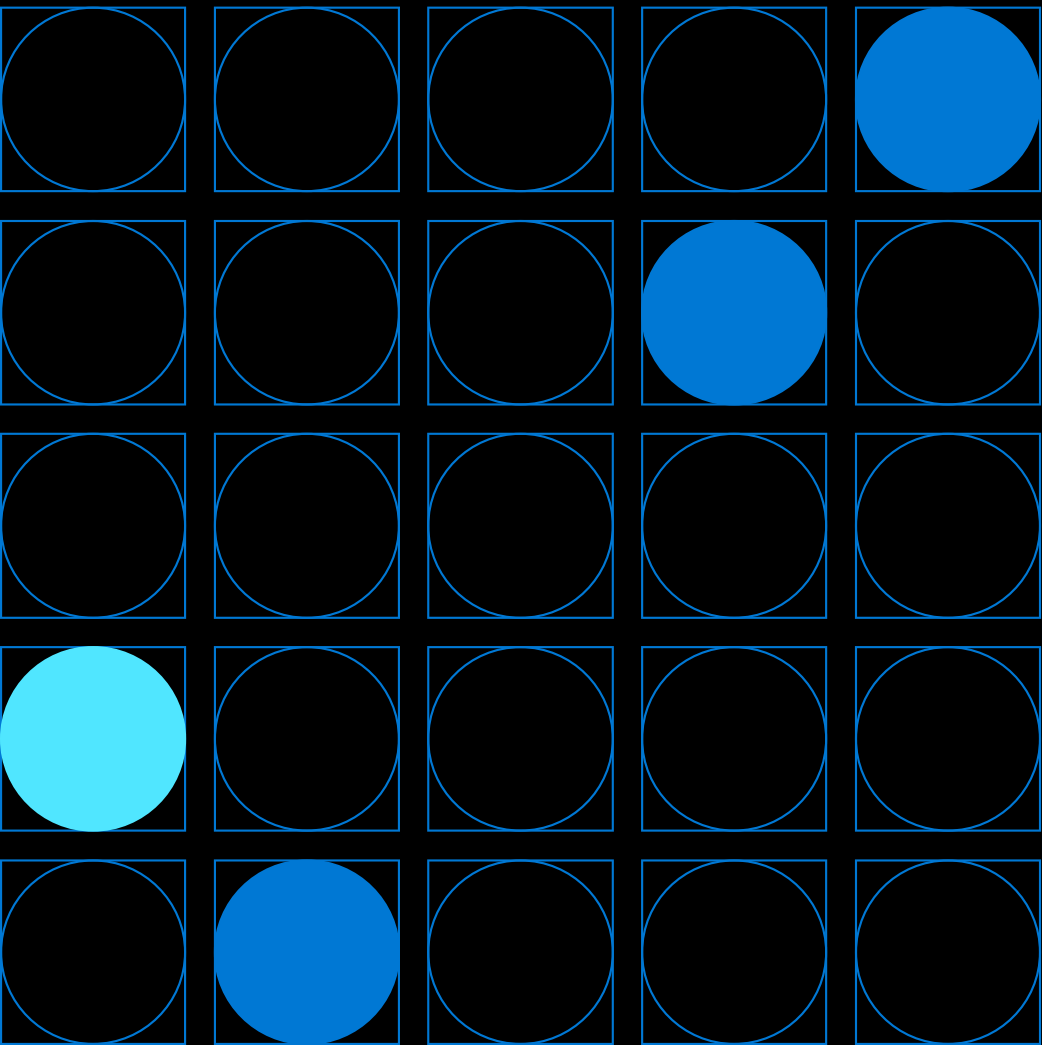
**Azure Data Lake Storage**

# Azure Data Lake Storage Gen2

ADLS Gen2 adds a high performance HDFS Endpoint to Azure Blob Storage and inherits the rich feature set of Azure Blob Storage \*



*\* Full Blob and HNS Interoperability available in post-GA*



**Azure Synapse Analytics**

# Azure Synapse Analytics

Limitless MPP data warehouse with unmatched time to insights



## Synapse SQL

Query and analyze data with T-SQL using both provisioned and serverless models



## Apache Spark for Synapse

Apache Spark in Synapse for quick creation of notebooks with your choice of language



## Synapse Pipelines

Build end-to-end workflows for your data movement and data processing scenarios



## Synapse Studio

Execute all data tasks with a simple UI and unified environment



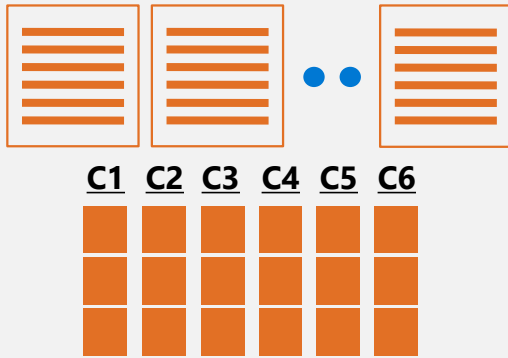
## Common Data Models

Comprehensive data models in CDM format providing information blueprint for customers to describe their data estate for analytics

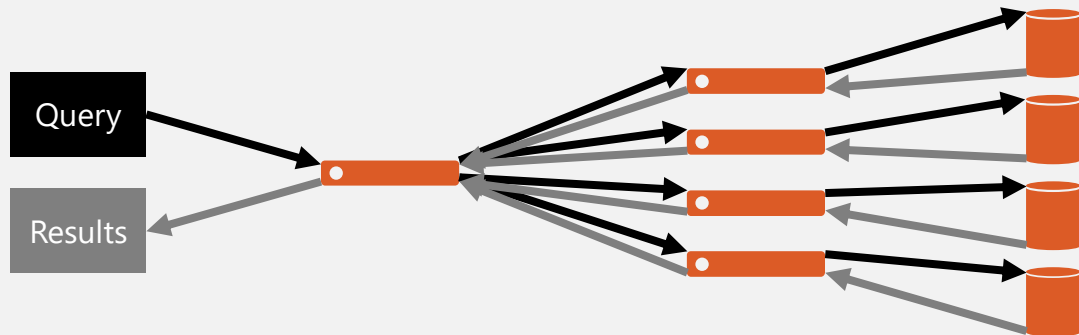
# Azure Synapse: Massive Parallel Processor

MPP and in-memory columnstore for next-generation performance

## Columnstore index representation



## Parallel query execution



Updateable clustered columnstore vs. table with customary indexing

Up to **100x**  
faster queries

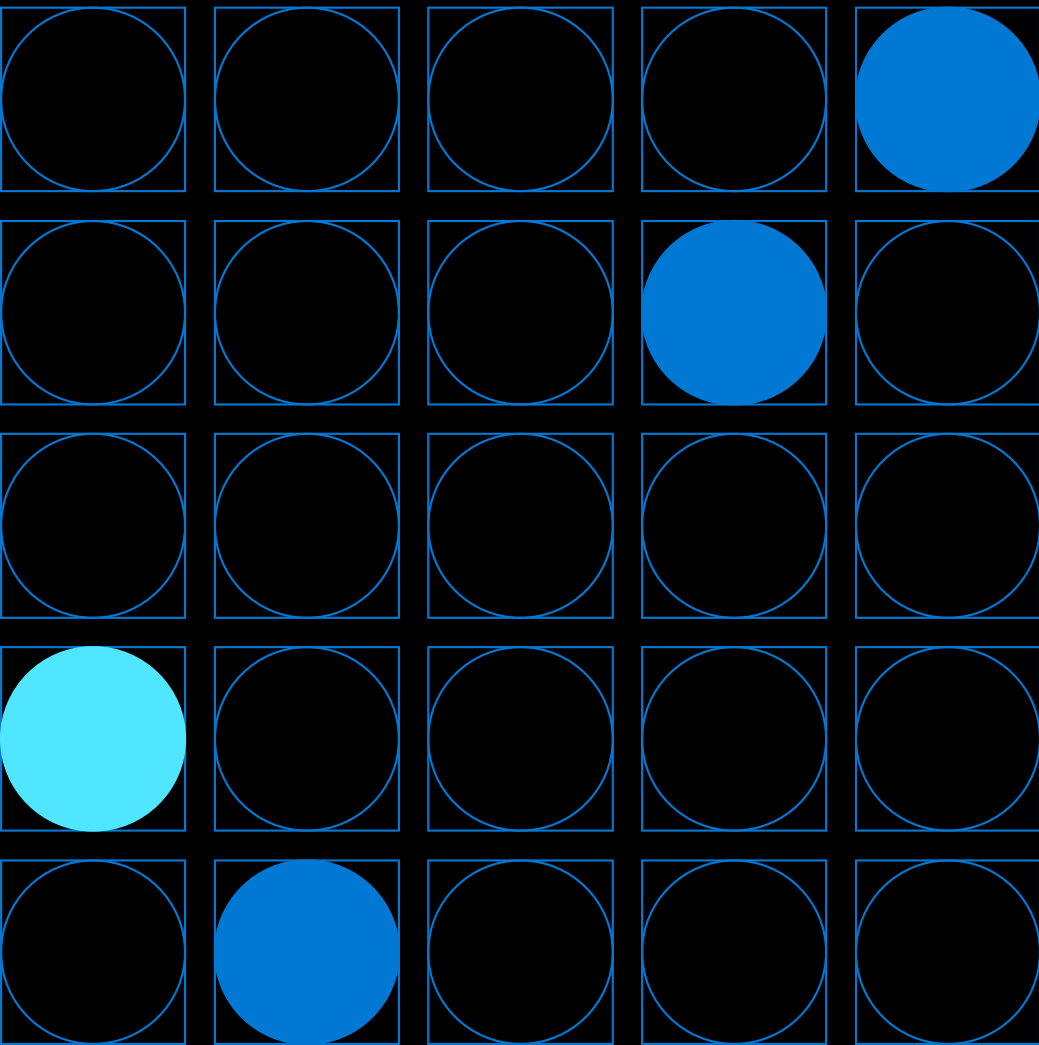
Up to **15x**  
more compression



Data storage in columnar format for massive compression

Data loading into or out of memory for next-generation performance, with up to 60% improvement in data loading speed

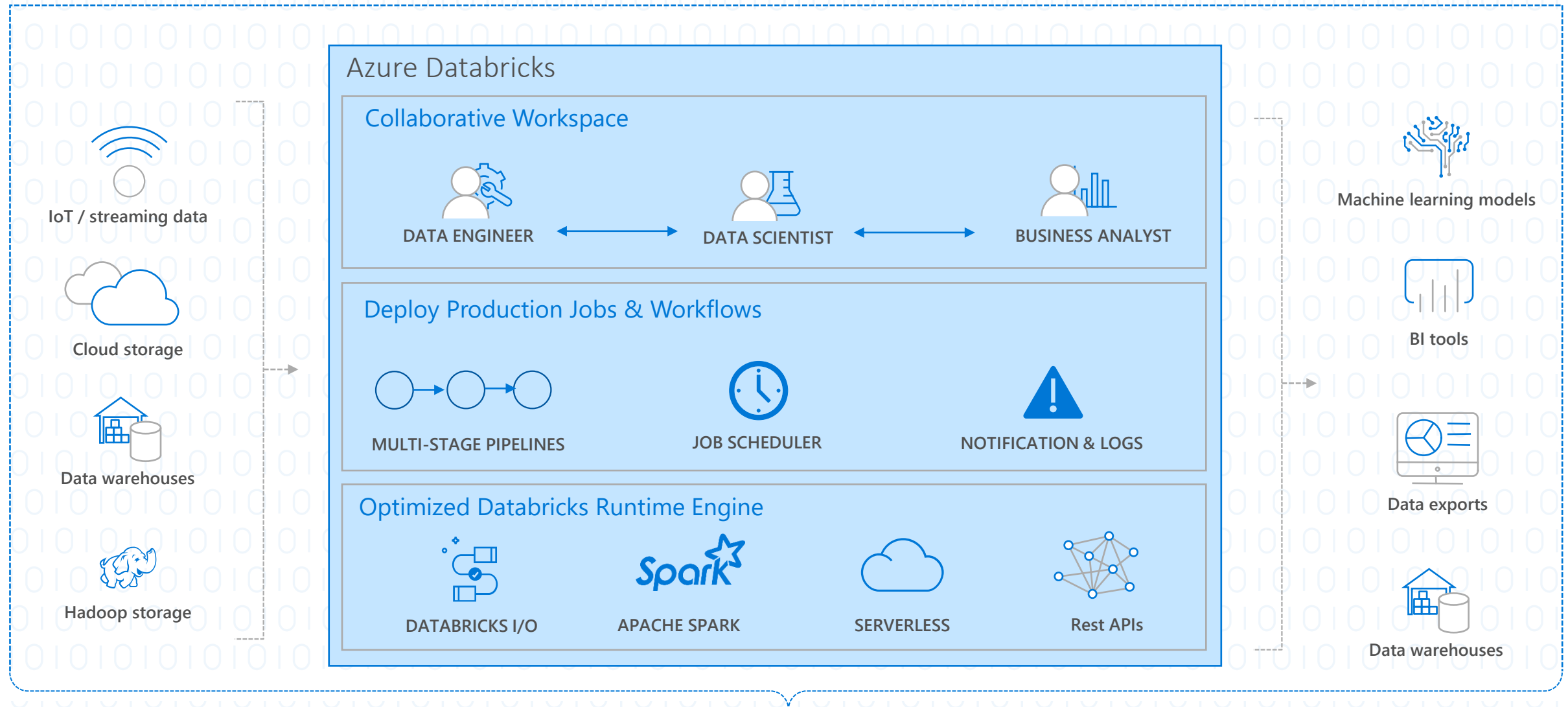
Updateable and clustered for real-time trickle loading



**Azure Databricks**



# A Z U R E   D A T A B R I C K S

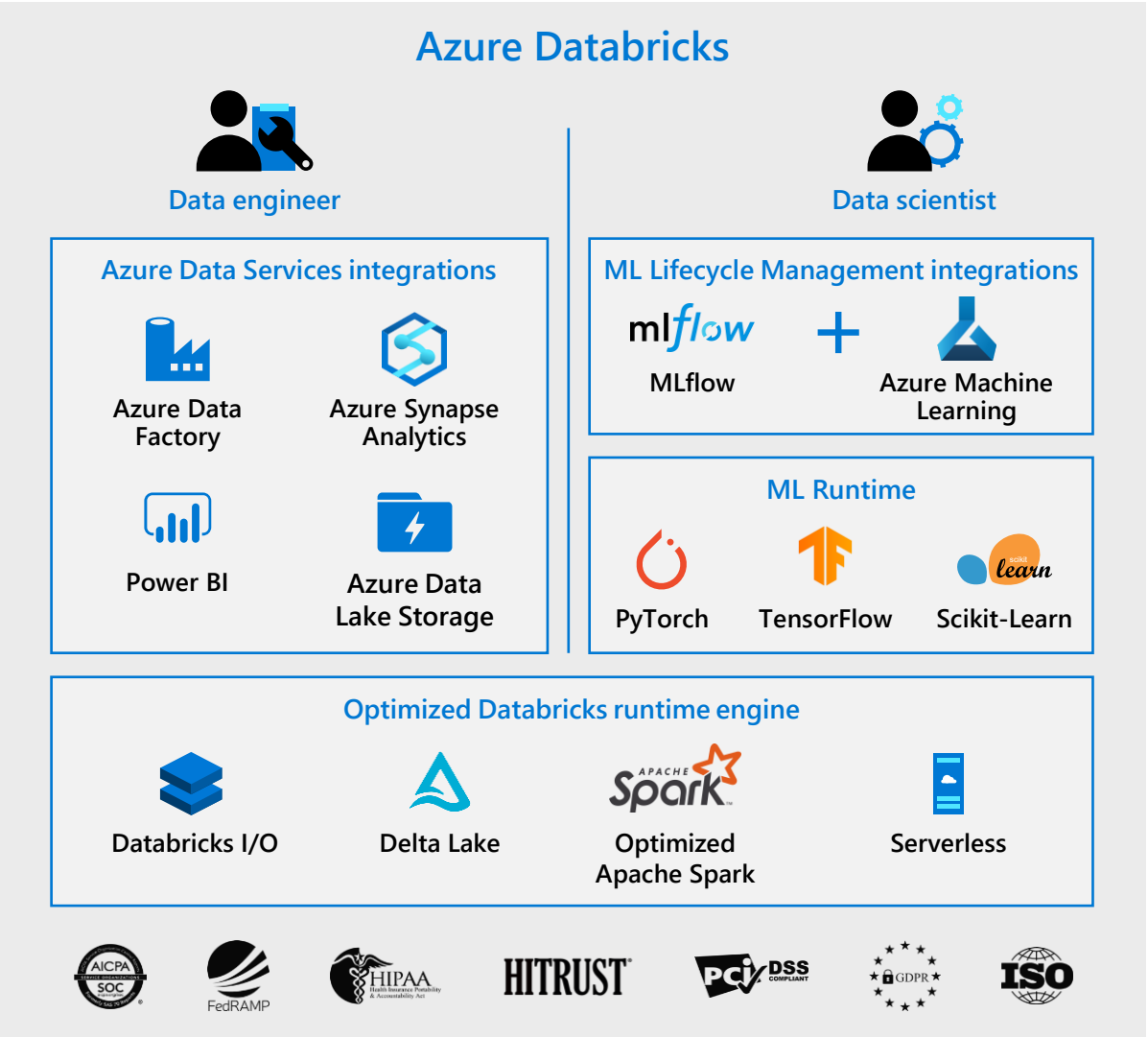


Enhance Productivity

Build on secure & trusted cloud

Scale without limits

# Databricks accelerates data-driven innovation



Collaborative workspace for data teams across the full lifecycle

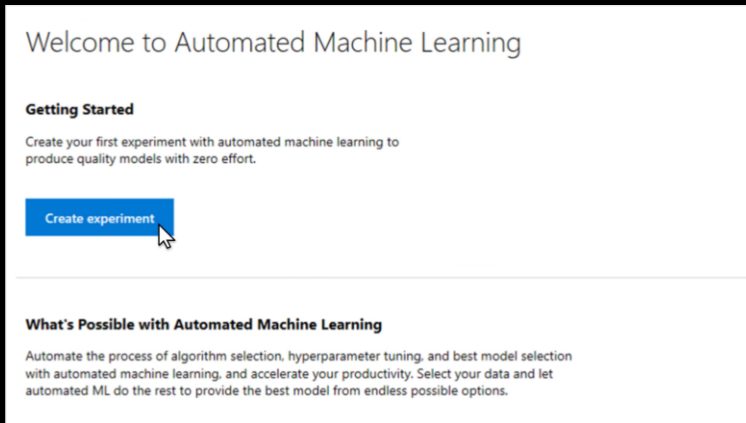
Native integration with AML and Azure Data Services

Scalable, reliable, and fast data - built on your existing data lake powered by most optimized Spark Engine

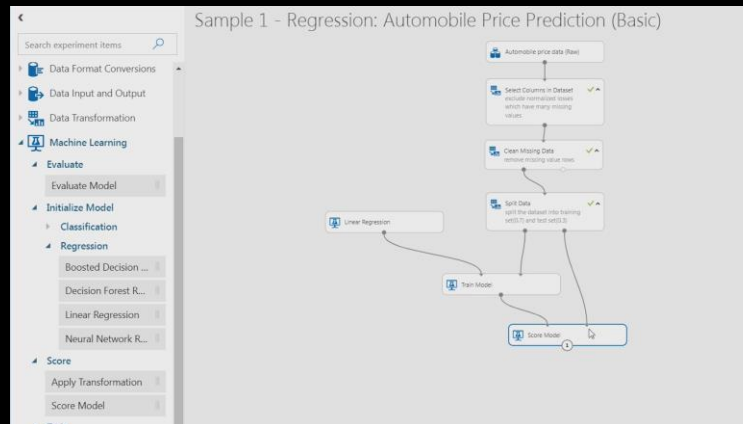
One fully-integrated security model for production infrastructure

# Simplify machine learning for any skill level

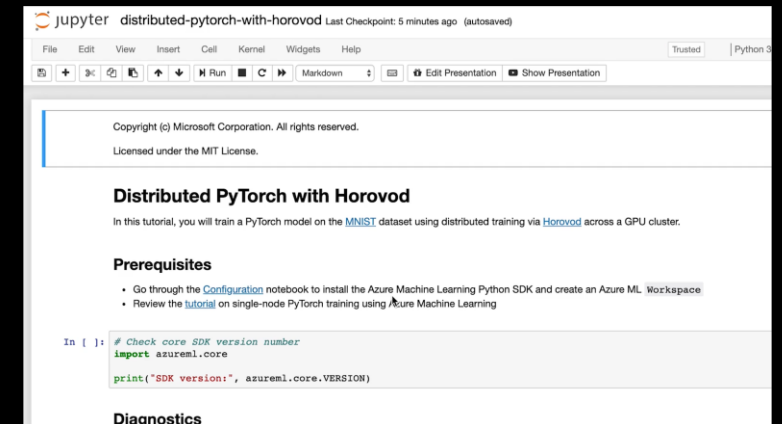
## Azure Machine Learning service



Automated  
machine learning UI



Visual interface



Machine learning notebooks

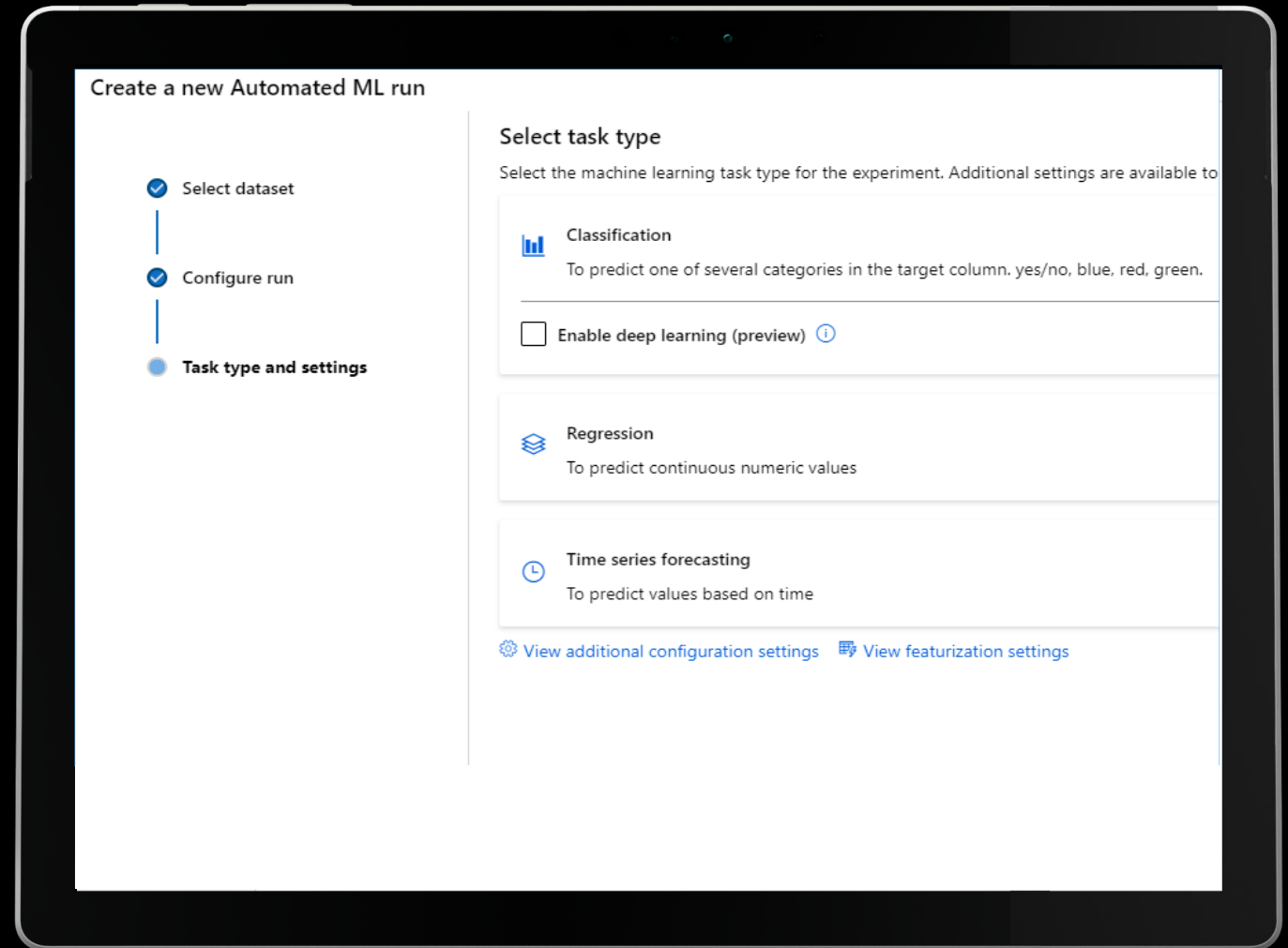
# Automated Machine Learning

Automatically build and deploy predictive models using the no-code UI or through a code-first notebooks experience.

Increase productivity with easy data exploration and profiling and with intelligent feature engineering.

Easily create accurate models customized to your data and refined by a wide array of algorithms and hyperparameters.

Build responsible AI solutions with model interpretability and fine-tune your models to improve accuracy.



# Accelerate model development

with automated machine learning

Input

101010  
010101  
101010

Enter data

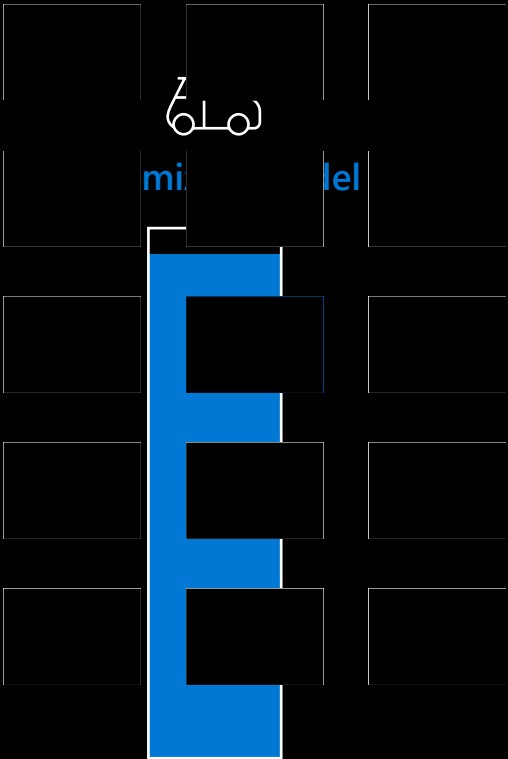
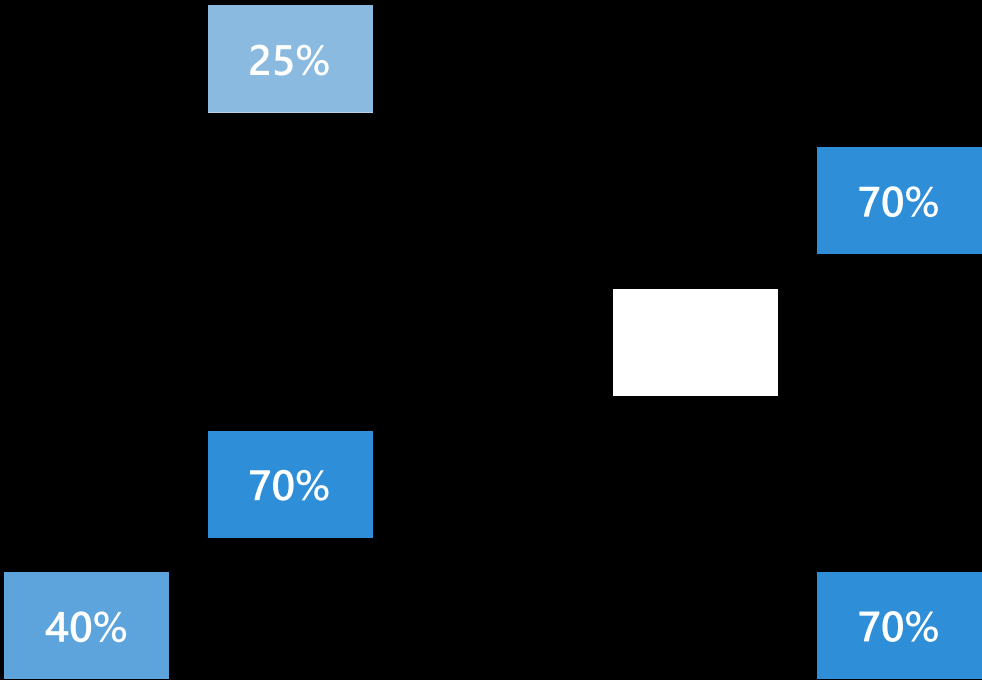


Define goals



Apply constraints

Intelligently test multiple models in parallel



Refresh Cancel Delete

Details   Data guardrails   **Models**   Outputs + logs   Child runs   Snapshot

Deploy   Download   Explain model   Refresh   Edit columns   Reset view

Search

Submitted time All filters Clear all

Showing 1-10 of 10 models

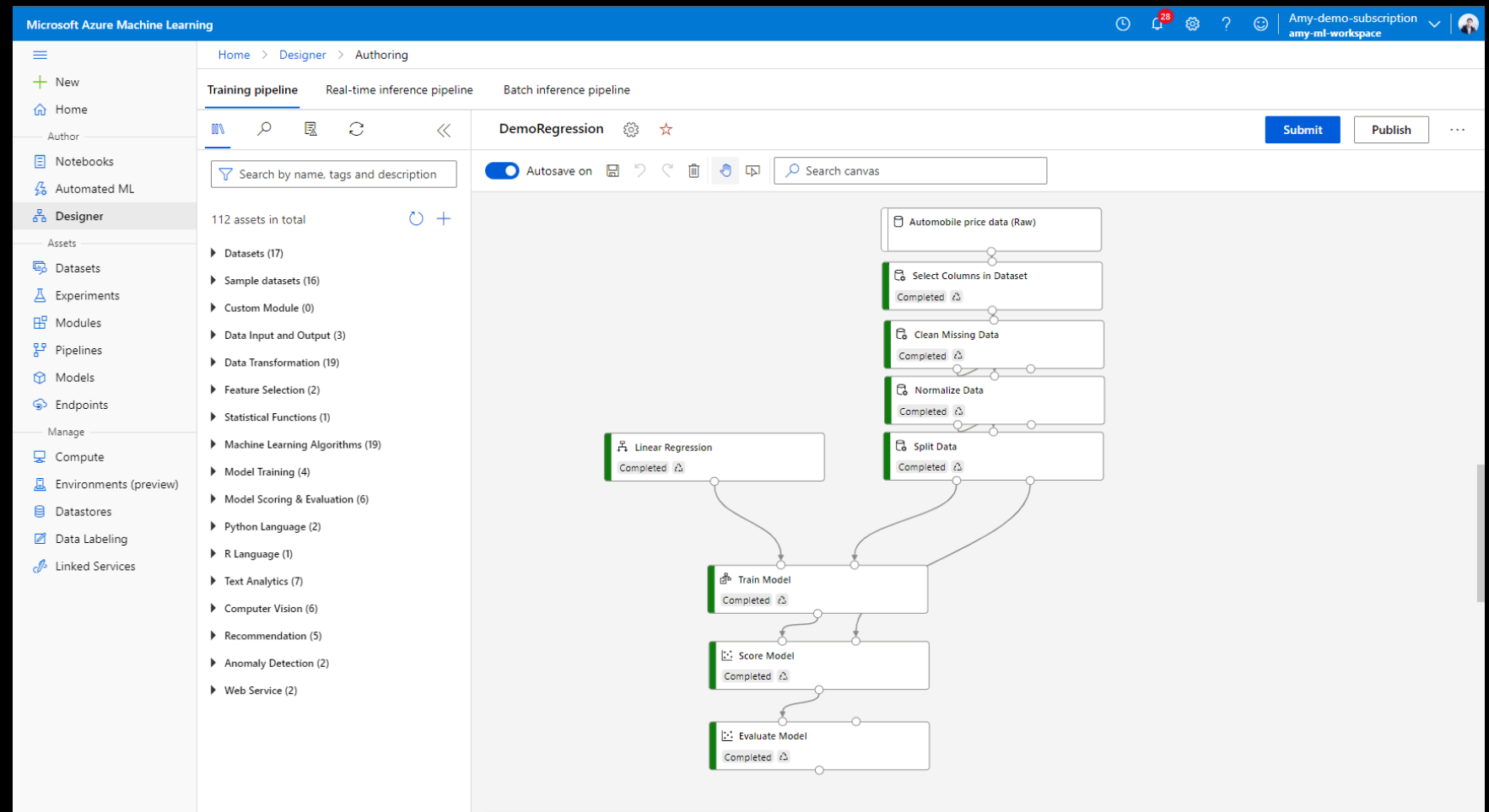
Page size: 25

Algorithm name	Explained	Norm macro recall ↓	Sampling	Submitted time	Duration	Hyperparameters
VotingEnsemble	<a href="#">View explanation</a>	0.23878	100.00 %	Jul 3, 2021 4:09 PM	38s	algorithm : [
MaxAbsScaler, ExtremeRandomTrees		0.21612	100.00 %	Jul 3, 2021 4:03 PM	36s	bootstrap : t
MaxAbsScaler, RandomForest		0.20684	100.00 %	Jul 3, 2021 4:01 PM	35s	bootstrap : t
StackEnsemble		0.17269	100.00 %	Jul 3, 2021 4:09 PM	41s	algorithm : [
MaxAbsScaler, LightGBM		0.09651	100.00 %	Jul 3, 2021 4:01 PM	34s	min_data_in
MaxAbsScaler, XGBoostClassifier		0.08788	100.00 %	Jul 3, 2021 4:01 PM	35s	tree_method
MaxAbsScaler, RandomForest		0.00943	100.00 %	Jul 3, 2021 4:02 PM	5m 27s	bootstrap
MaxAbsScaler, ExtremeRandomTrees		0.00000	100.00 %	Jul 3, 2021 4:06 PM	37s	bootstrap
MaxAbsScaler, ExtremeRandomTrees		0.00000	100.00 %	Jul 3, 2021 4:05 PM	33s	bootstrap : t
MaxAbsScaler, RandomForest		0.00000	100.00 %	Jul 3, 2021 4:04 PM	34s	bootstrap : t

# Azure Machine Learning

## ML for all skills

- Code-free Drag & Drop Designer
- Automated Machine Learning Model
- Supports code based
- Integrated with other development tools and frameworks





# Microsoft Azure Machine Learning Algorithm Cheat Sheet

This cheat sheet helps you choose the best machine learning algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the goal you want to achieve with your data.

