



Week 7


# Artificial Intelligence Program

## Infrastructure and Architecture

# > Agenda // Program

Assignments [60%]

**EXAMS [40%]**

WEEK	SUBJECT	ASSIGNMENT / TO BE DELIVERED	DATES
2	Intro / AI Function / Enablers		Sep 13
3	Infra and Architecture / On-prem vs. Cloud / CSPs	C1	Sep 20
4	Data Pipeline / Processes / Framework / AutoML	#1 Image Classifier [5%]	Sep 27
5	Data Pipeline / Processes / Framework / AutoML	C2	Oct 4
6	More Data / SSIS / ADF / Data Quality	#2 Machine Learning Studio [10%]	Oct 11
7	Azure services – Intro <b>EXAM 1 [20%]</b>	C3	Oct 18
8	<b>READING WEEK</b>	<b>NO CLASSES</b>	<b>Oct 25</b>
9	Azure services – Cognitive Services 1	41	Nov 1
10	Azure services – Cognitive Services 2	#3 Draw your own Architecture [5%] 42	Nov 8
11	Azure services – Cognitive Services 3	43	Nov 15
12	Azure services – Cognitive Services 4	#4 Azure pipeline // Sentiment Analysis [20%] 44	Nov 22
13	AWS Academy – Cloud Foundations 		Nov 29
14	AWS Academy – Machine Learning	#5 AWS Academy – Cloud Foundations [10%]	Dec 6
15	Enterprise Architecture <b>EXAM 2 [20%]</b>	#6 AWS Academy – Machine Learning [10%]	Dec 13

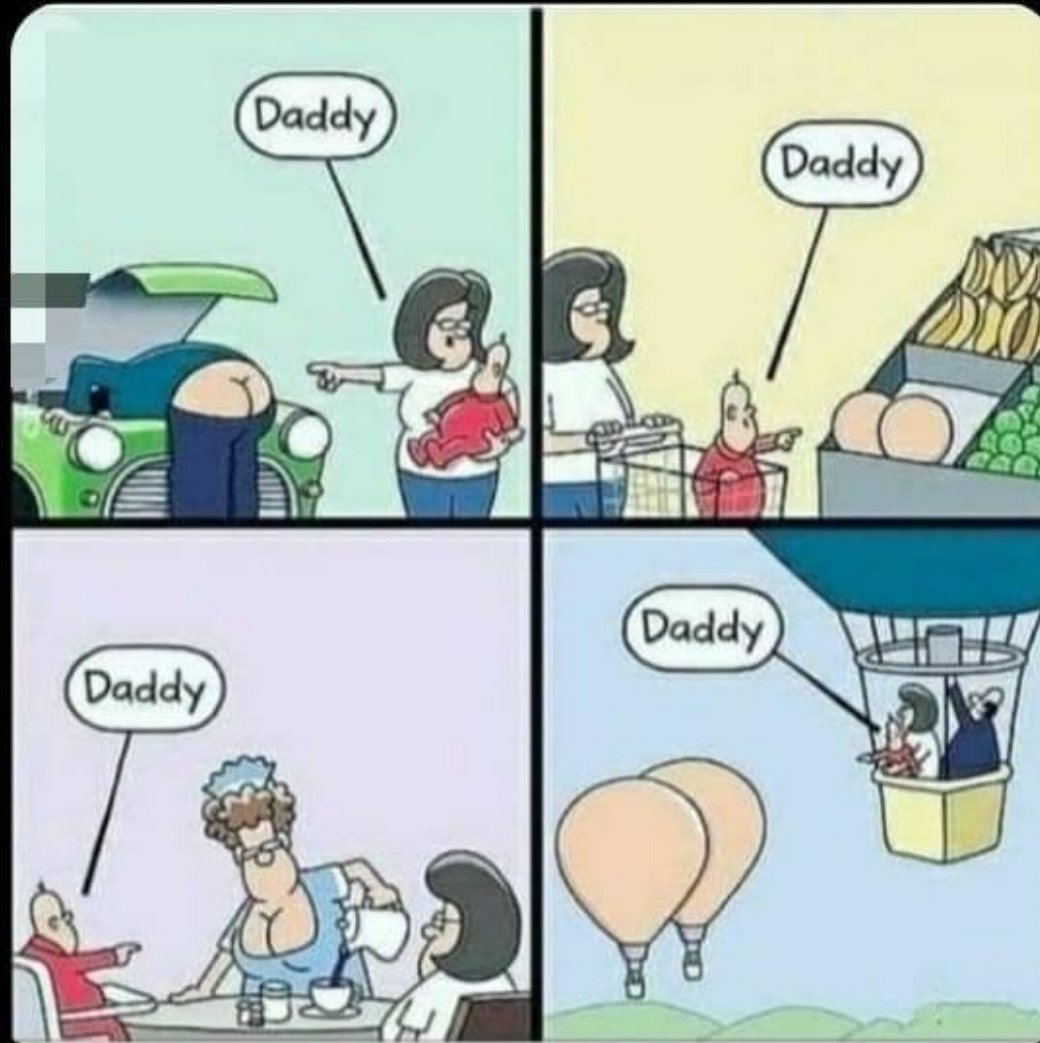
# > Agenda

---

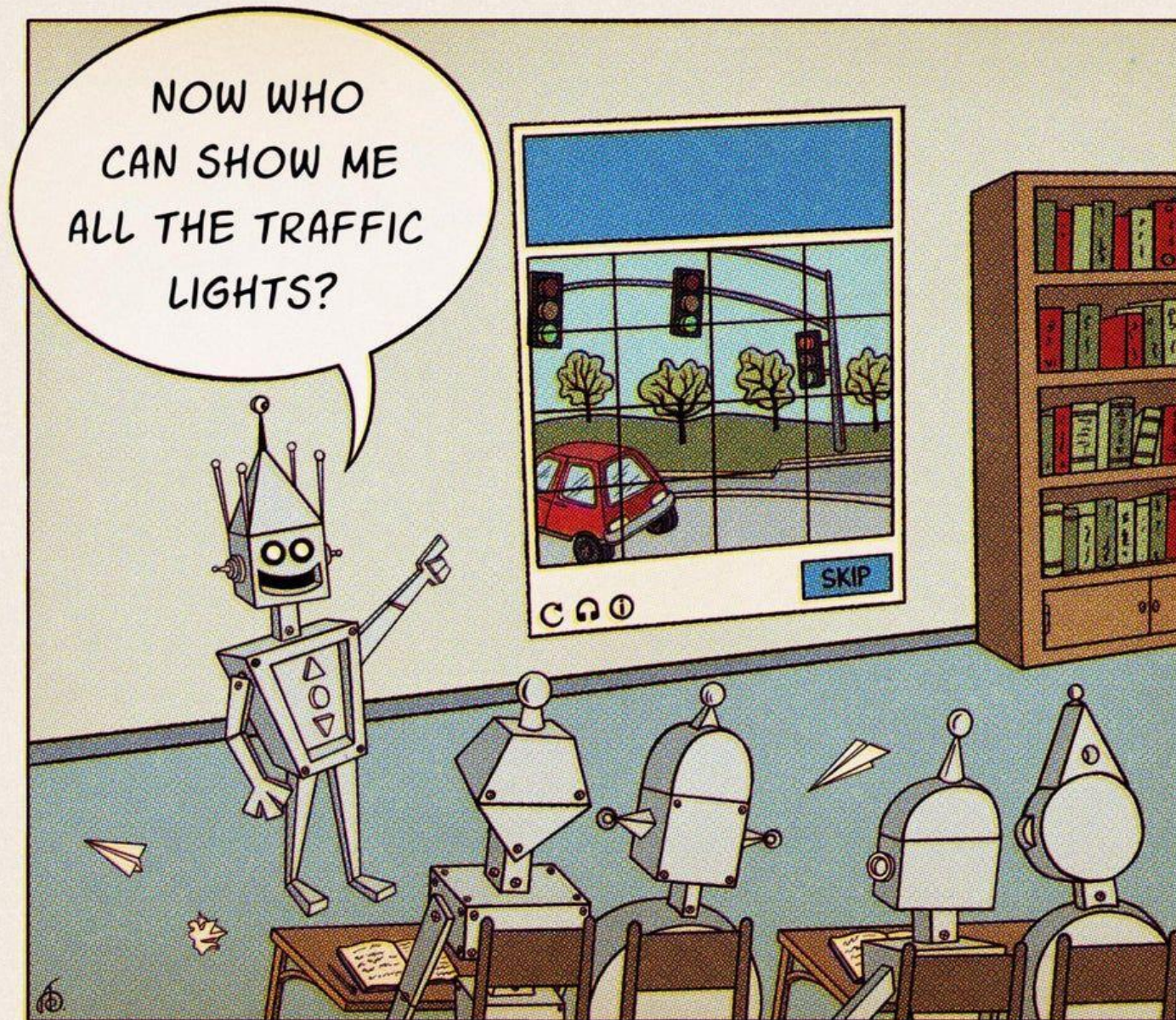
- Modern Data Platform Concepts
- Data Lake
- Azure Data Lake Storage Gen2
- Azure ML
- Azure AI
- Event Hubs / Example
- Stream Analytics
- Developing your own Architecture
- Practice exercise – drawing your own architecture

# Machine Learning and AI be like..

[fb.me/yuva.krishna.memes](https://fb.me/yuva.krishna.memes)







© 2019 ANGIE SIVERIA & OSKAR SCHUSTER

@ROBOTOPIAWEEKLYCOMIC

# Azure Week 2

## Microsoft



# Modern Data Platform Concepts



# The Modern Data Problem

## How to derive value from data:

What happened historically?

What is happening now?

What is going to happen?

Each dimension of data is  
**constantly expanding**

VOLUME

VELOCITY

Real-time

Batch

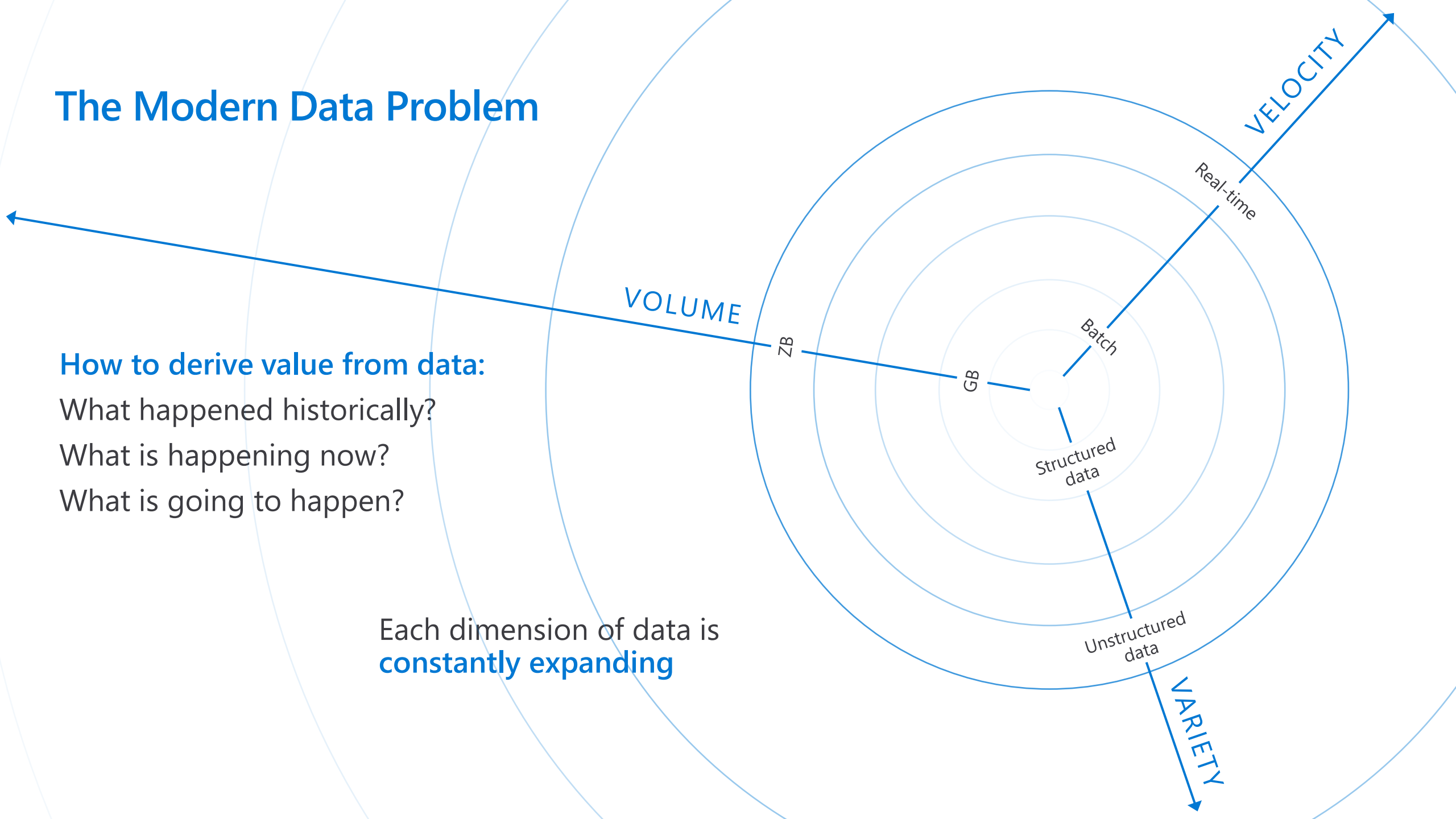
GB

ZB

Structured data

Unstructured data

VARIETY





# What is a Data Lake?

It is a central storage repository that holds data coming from many sources in a raw, granular format. It can store **structured, semi-structured, or unstructured data**, which means data ingested quickly and can be kept in a more flexible format for future use cases.



## Characteristics

- Schema-on-read (ELT)
- Collection of data, not a platform
- Perfect place for evolving data



## Benefits

- Quickly ingest high volumes of diverse data structures
- Enable advanced analytics and data exploration
- Scalability and storage cost reduction



## Best Practices

- Data Governance needed to avoid Data Swamp
- Security considerations
- Design your Data Lake
- Metadata management

# Data Warehouse or Data Lake?

Answer: both.

	Data Warehouse	Data Lake
Requirements	Relational requirements	Diverse data, scalability, low cost
Data Value	Data of recognised high value	Candidate data of potential value
Data Processing	Mostly refined calculated data	Mostly detailed source data
Business Entities	Known entities, tracked over time	Raw material for discovering entities and facts
Data Standards	Data conforms to enterprise standards	Fidelity to original format and condition
Data Integration	Data integration upfront	Data prep on demand
Transformation	Data transformed, in principle	Data repurposed later, as needs arise
Schema Definition	Schema-on-write	Schema-on-read
Metadata Management	Metadata improvement	Metadata developed on read

# Data Lake Design Considerations

## Data Lake Zones

### Transient Landing Zone

Temporary storage of data to meet regulatory and quality control requirements. Limited access. May not be required depending on requirements.

### Raw Zone

Original source of data ready for consumption. Metadata publicly available but access to data still limited.

### Trusted Zone

Standardized and enriched datasets ready for consumption to those with appropriate role-based access. Metadata available to all.

### Curated/Refined Zone

Data transformed from Trusted Zone to meet specific business requirements.

### Sandbox Zone

Playground for Data Scientists for ad hoc exploratory use cases.

## Data Governance Considerations

### Security and Compliance

Access Control at Folder/File level

Encryption at rest

### Metadata Management

Data Quality

Metadata Management

Lifecycle Management

# Azure Data Lake Storage Gen2



# Azure Data Lake Storage Gen2

A “no-compromises” Data Lake: Secure, performant and massively-scalable

A Data Lake that brings together the cost and scale of object storage with the performance and analytics feature set of data lake storage



## Fast

Atomic file operations  
mean jobs complete  
faster



## Manageable

Automated Lifecycle Policy  
Management

Object Level tiering



## Secure

Support for fine-grained  
ACLs, protecting data at the  
file and folder level

Multi-layered protection via  
at-rest Storage Service  
encryption & Azure Active  
Directory integration



## Scalable

No limits on data  
store size

Global footprint  
(50 regions)



## Cost effective

Object store pricing levels

File system operations  
minimize transactions  
required for job completion

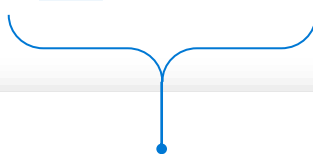


## Integration ready

Optimized for Spark and Hadoop  
Analytic Engines

Tightly integrated with Azure end  
to end analytics solutions

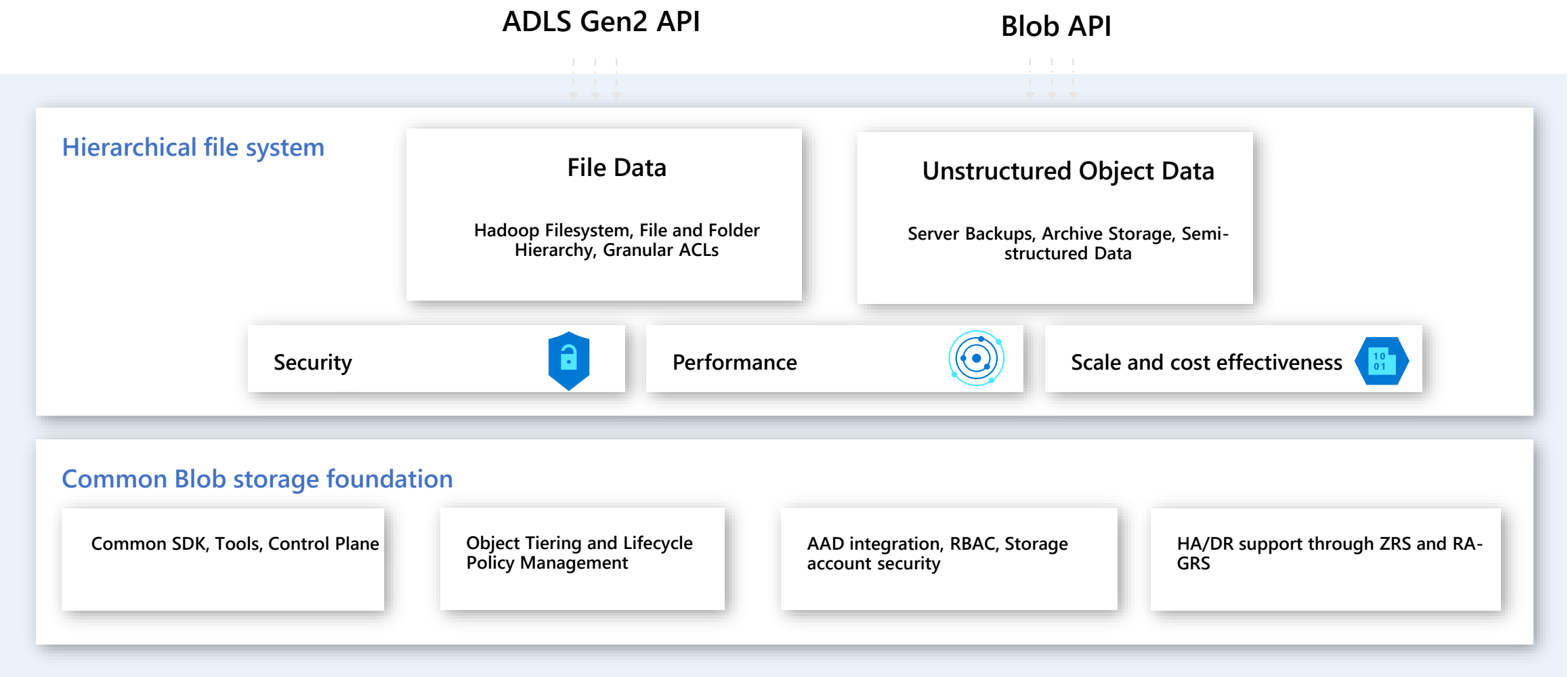
Multiprotocol access



Single service

# Azure Data Lake Storage Gen2

High performance HDFS Endpoint to Azure Blob Storage



# Modern Data Platform Concepts

# What's No-SQL?

Term coined in 2009 for a developer meetup – “Not Only SQL” -> “NoSQL”.

Databases that allow you to store and retrieve data in various structures, formats, and models other than tabular relational model.

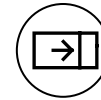
## There's a time and a place for everything

Sometimes a relational store is the right choice

Sometimes a NoSQL store is the right choice

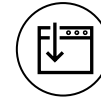
Sometimes you need more than one store for an app -> polyglot persistence

## Data Structures



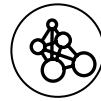
### **Key-Value Databases**

Cosmos DB, Redis Cache, Azure Table



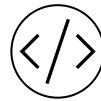
### **Column Family Stores**

Cosmos DB, Cassandra, HBase



### **Graph Databases**

Cosmos DB, Neo4j, Gremlin



### **Document Databases**

Cosmos DB, MongoDB



# Azure ML

# > Azure ML Workspaces

## Azure Machine Learning workspaces

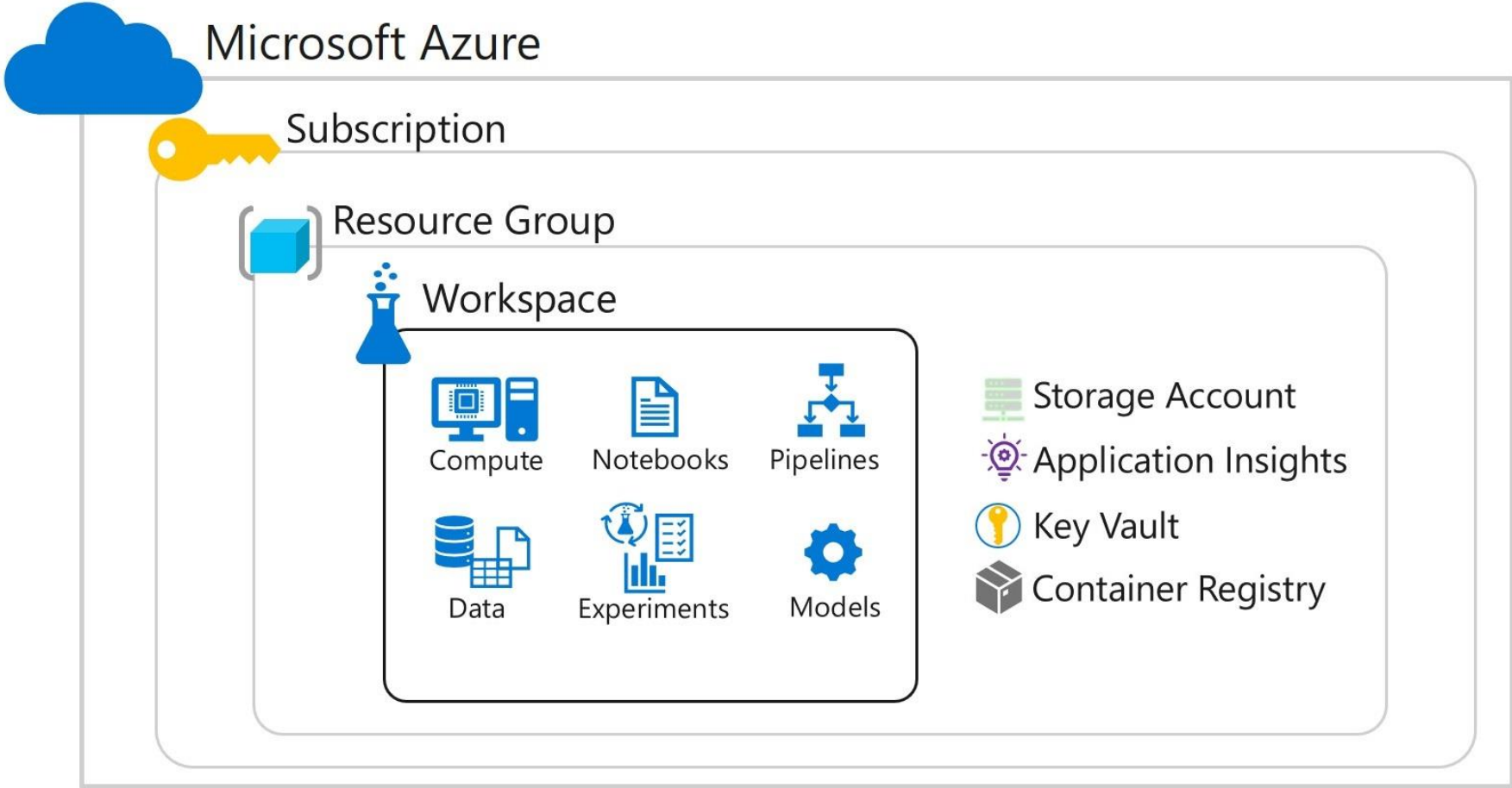
A workspace is a context for the experiments, data, compute targets, and other assets associated with a machine learning workload.

## Workspaces for Machine Learning Assets

A workspace defines the boundary for a set of related machine learning assets. You can use workspaces to group machine learning assets based on projects, deployment environments (for example, test and production), teams, or some other organizing principle. The assets in a workspace include:

- Compute targets for development, training, and deployment.
- Data for experimentation and model training.
- Notebooks containing shared code and documentation.
- Experiments, including run history with logged metrics and outputs.
- Pipelines that define orchestrated multi-step processes.
- Models that you have trained.

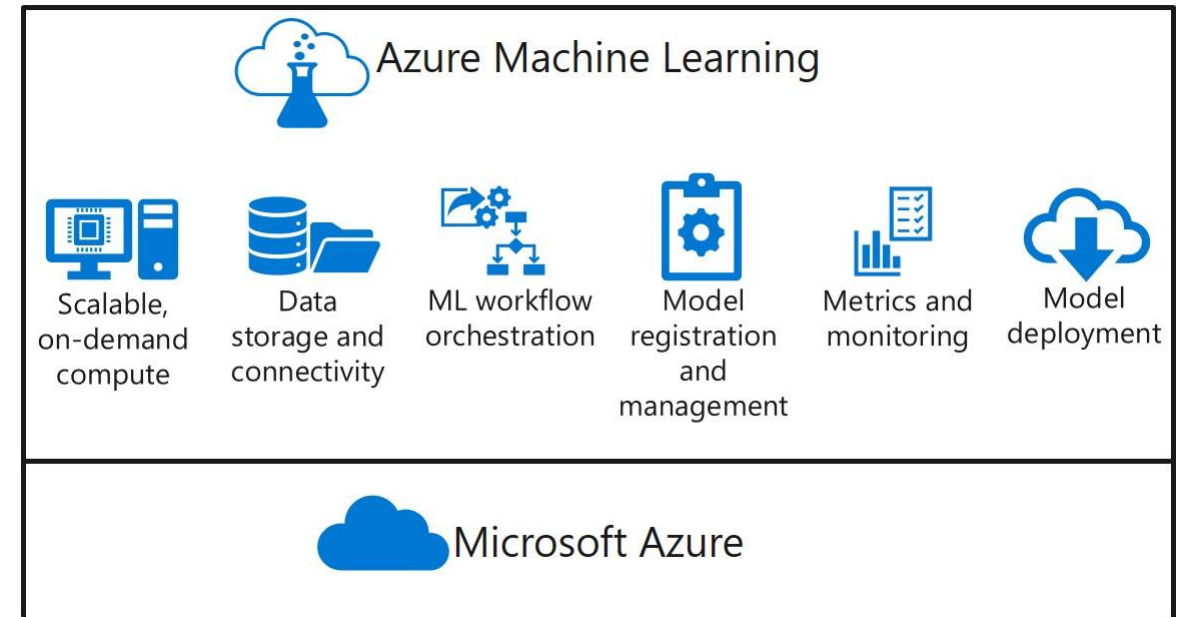
# > Azure ML Workspaces



# > Azure ML Studio

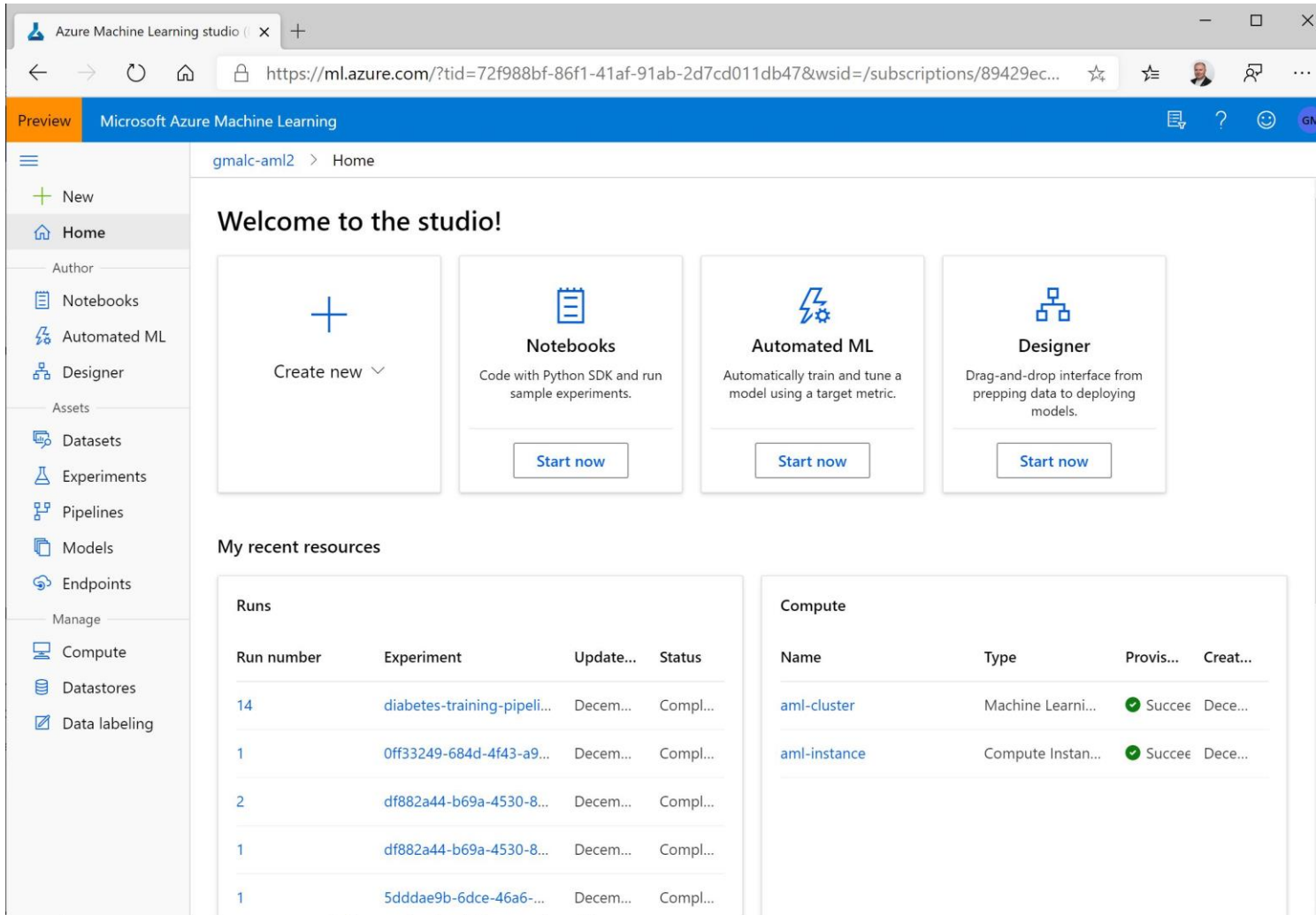
Built on the Microsoft Azure cloud platform, Azure Machine Learning enables you to manage:

- Scalable on-demand compute for machine learning workloads.
- Data storage and connectivity to ingest data from a wide range sources.
- Machine learning workflow orchestration to automate model training, deployment, and management processes.
- Model registration and management, so you can track multiple versions of models and the data on which they were trained.
- Metrics and monitoring for training experiments, datasets, and published services.
- Model deployment for real-time and batch inferencing.





# > Azure ML Studio

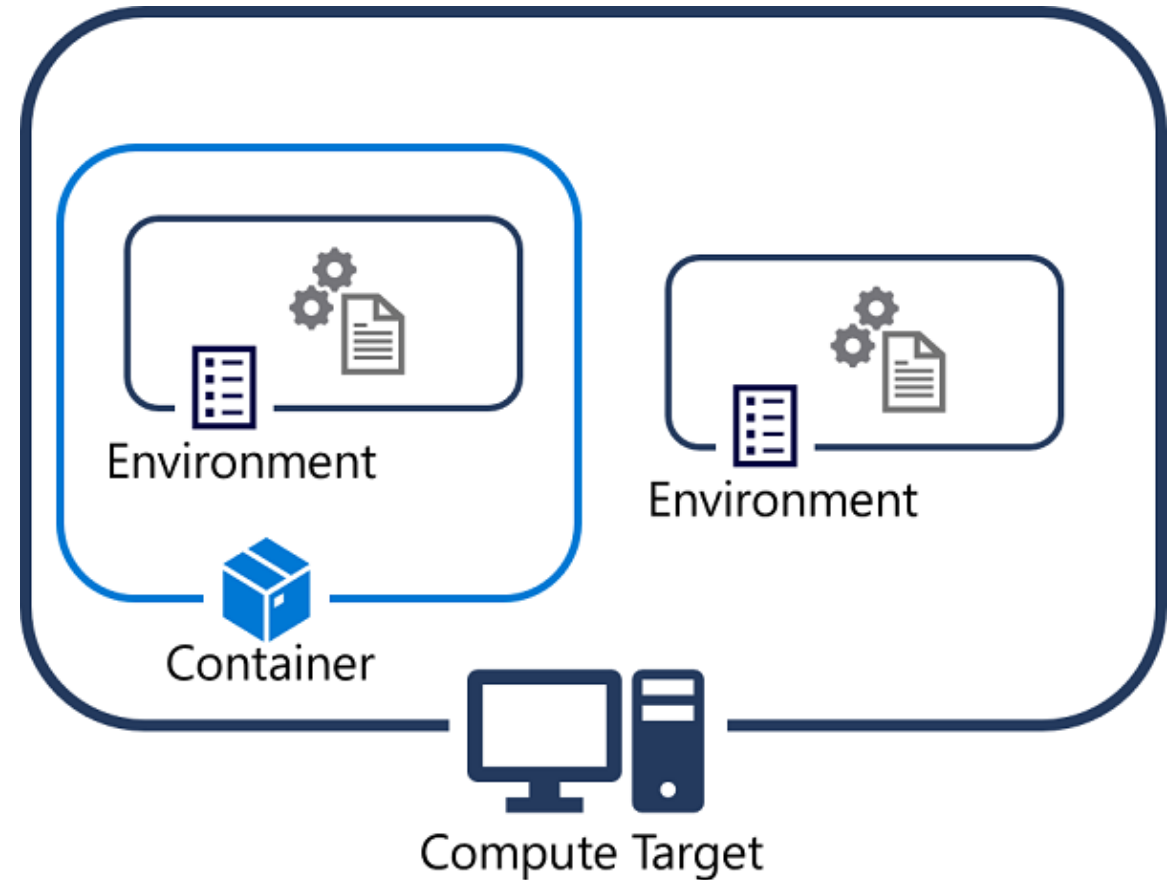


**Azure Machine Learning studio**  
You can manage the assets in your Azure Machine Learning workspace in the Azure portal, but as this is a general interface for managing all kinds of resources in Azure, data scientists and other users involved in machine learning operations may prefer to use a more focused, dedicated interface.

# > Azure ML Studio

Python code runs in the context of a *virtual environment* that defines the version of the Python runtime to be used as well as the installed packages available to the code. In most Python installations, packages are installed and managed in environments using **Conda** or **pip**.

To improve portability, we usually create environments in docker containers that are in turn be hosted in compute targets, such as your development computer, virtual machines, or clusters in the cloud.



# Azure AI

# Azure AI

## Solution Areas

### AI apps and agents



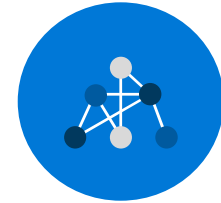
Azure Cognitive Services  
Azure Bot Service

### Knowledge mining



Azure Search

### Machine learning



Azure Databricks  
Azure Machine Learning  
Azure AI Infrastructure

---

Productive   Built for enterprises   Trusted



# Machine Learning on Azure

## Domain specific pretrained models

To simplify solution development



Vision



Speech



Language



Search

## Familiar Data Science tools

To simplify model development



Visual Studio Code



Azure Notebooks



Jupyter



Command line

## Popular frameworks

To build advanced deep learning solutions



PyTorch



TensorFlow



Scikit-Learn



ONNX

## Productive services

To empower data science and development teams



Azure  
Databricks



Azure Machine  
Learning



Machine  
Learning VMs

## Powerful infrastructure

To accelerate deep learning



CPU



GPU

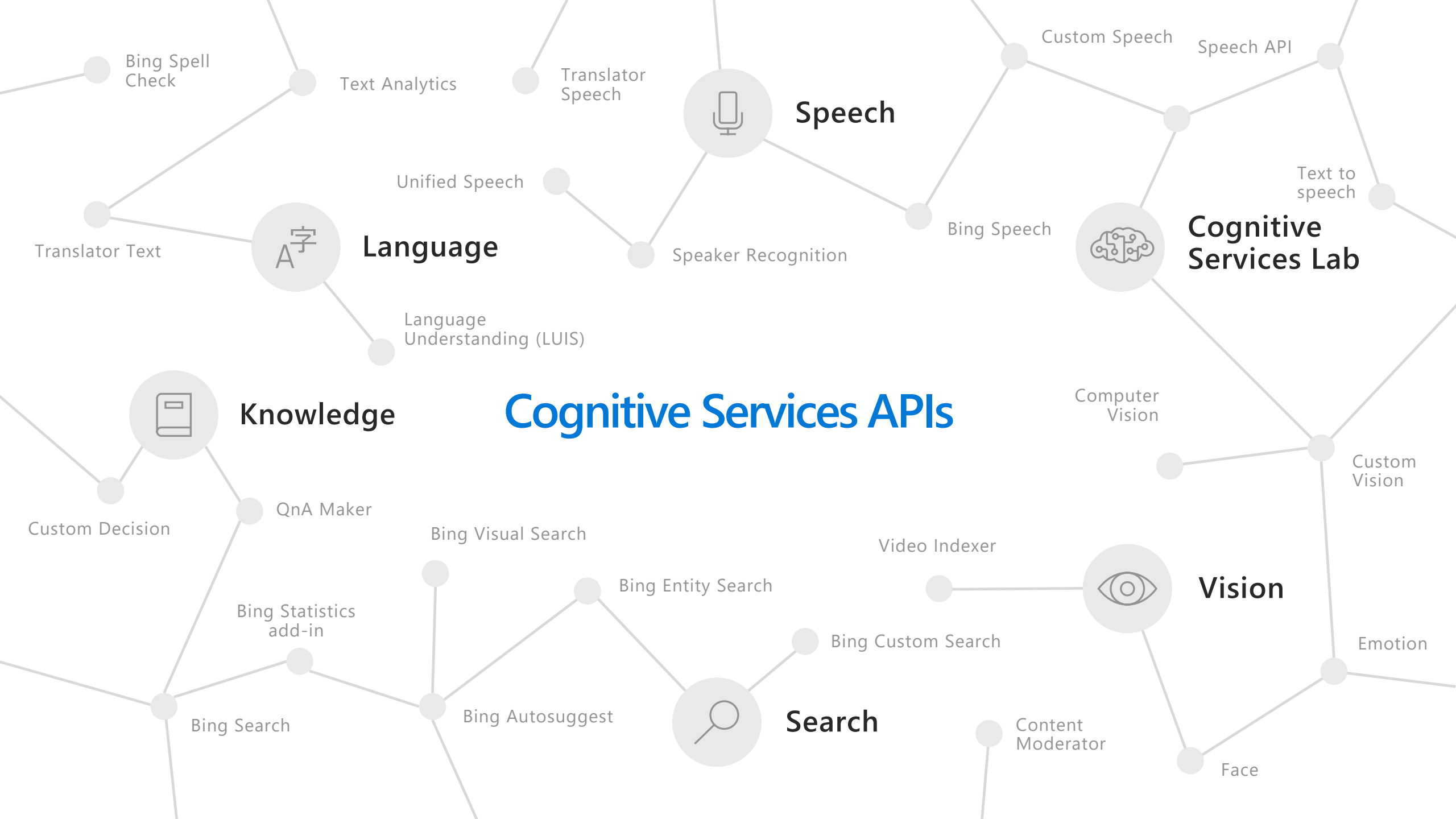


FPGA



From the Intelligent Cloud to the Intelligent Edge





# Cognitive Services capabilities

Infuse your apps, websites, and bots with human-like intelligence



## Vision

- Object, scene, and activity detection
- Face recognition and identification
- Celebrity and landmark recognition
- Emotion recognition
- Text and handwriting recognition (OCR)
- Customizable image recognition
- Video metadata, audio, and keyframe extraction and analysis
- Explicit or offensive content moderation



## Speech

- Speech transcription (speech-to-text)
- Custom speech models for unique vocabularies or complex environment
- Text-to-speech
- Custom Voice
- Real-time speech translation
- Customizable speech transcription and translation
- Speaker identification and verification



## Language

- Language detection
- Named entity recognition
- Key phrase extraction
- Text sentiment analysis
- Multilingual and contextual spell checking
- Explicit or offensive text content moderation
- PII detection for text moderation
- Text translation
- Customizable text translation
- Contextual language understanding



## Knowledge

- Q&A extraction from unstructured text
- Knowledge base creation from collections of Q&As
- Semantic matching for knowledge bases
- Customizable content personalization learning



## Search

- Ad-free web, news, image, and video search results
- Trends for video, news
- Image identification, classification and knowledge extraction
- Identification of similar images and products
- Named entity recognition and classification
- Knowledge acquisition for named entities
- Search query autosuggest
- Ad-free custom search engine creation

# Knowledge mining with Azure Search

Documents



Key Phrase extraction



Organization entity extraction



Face detection



Custom skills

Cognitive skills



Location entity extraction



Persons entity extraction



Celebrity recognition



Landmark detection



Sentiment analysis



Language detection

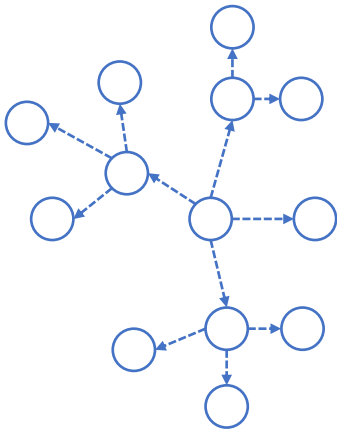


Tag extraction

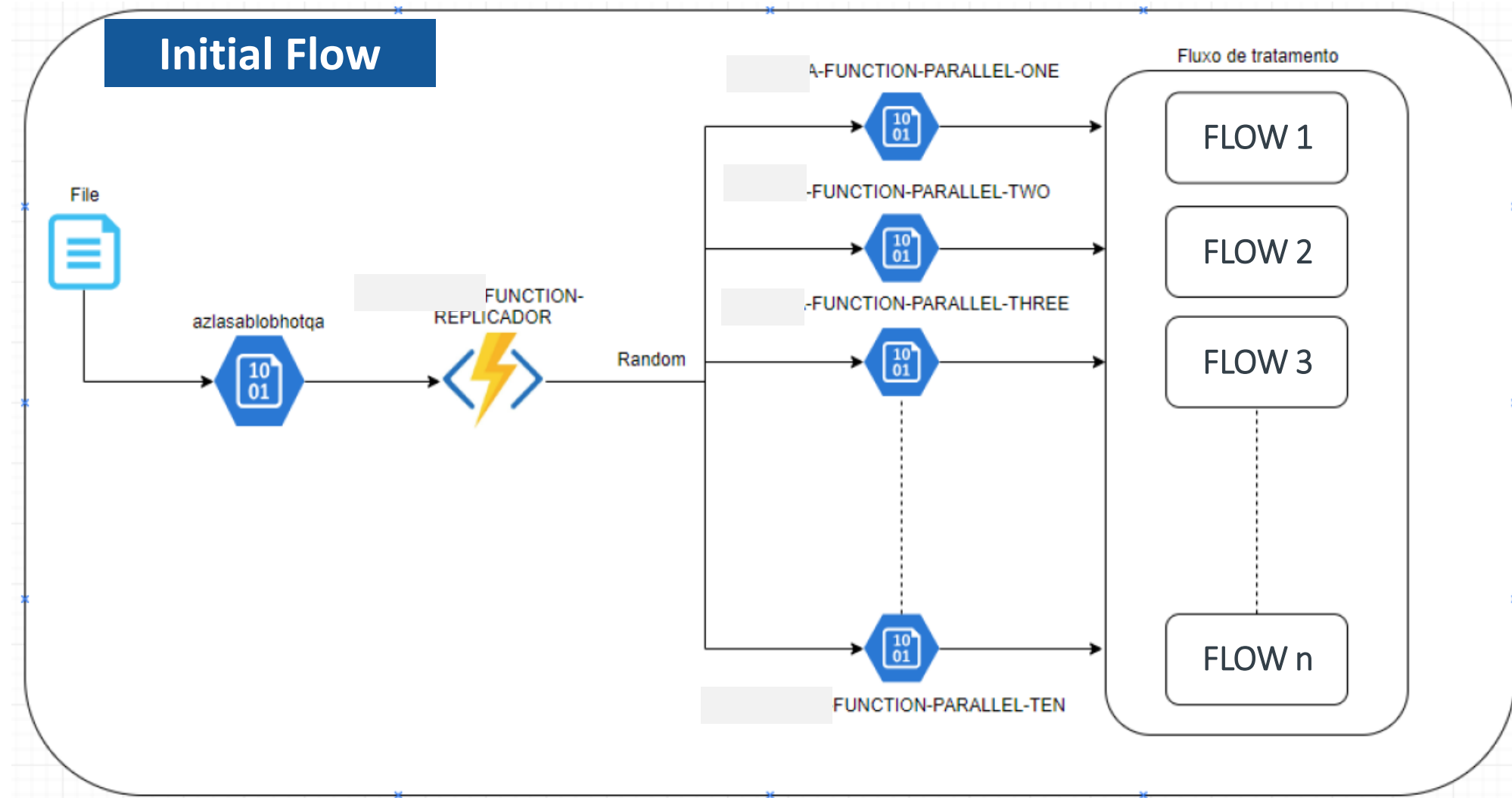


Printed text recognition

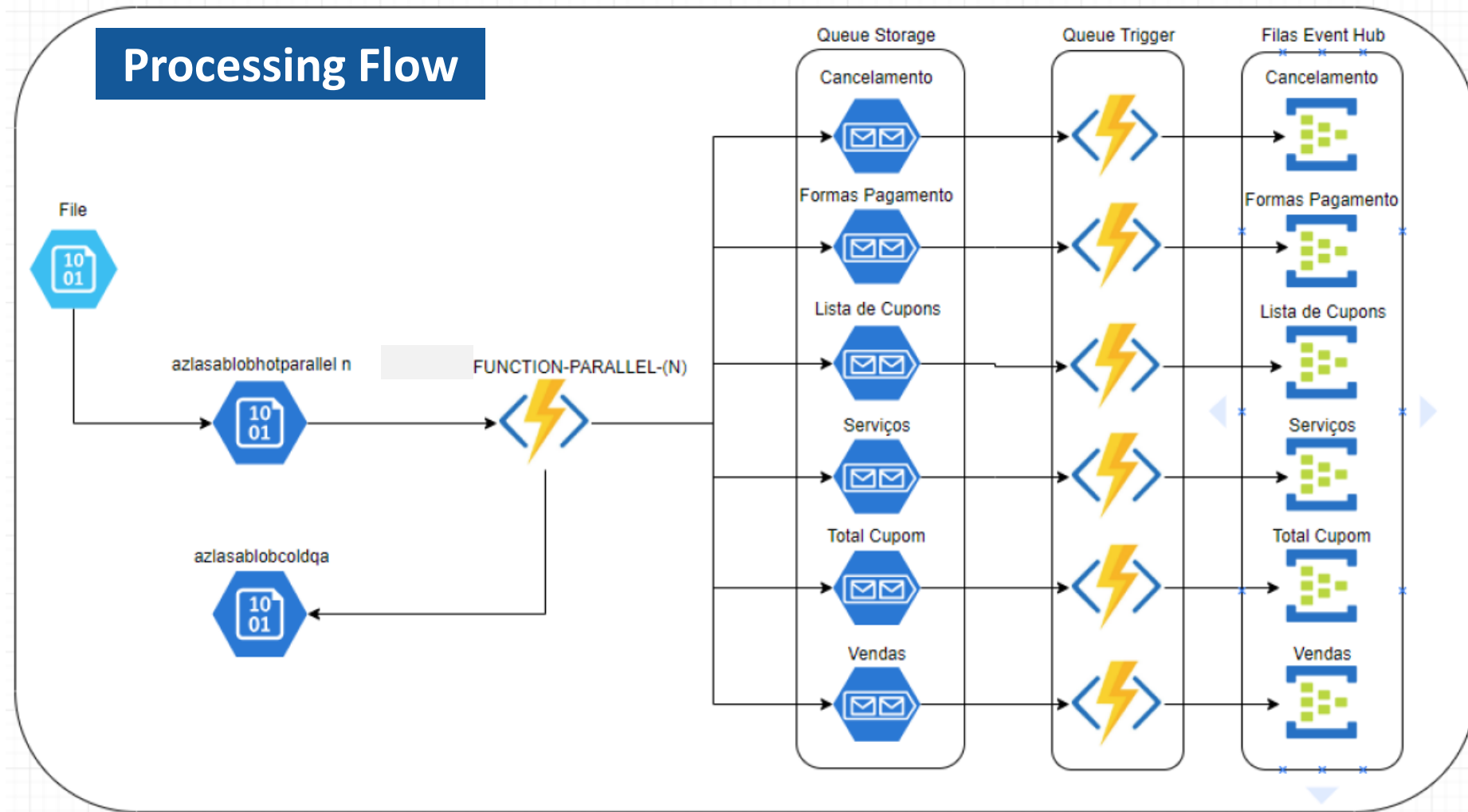
Fully text-searchable  
rich index



# > Case // Example // Parallel processes



# > Case // Example // Flow Detail

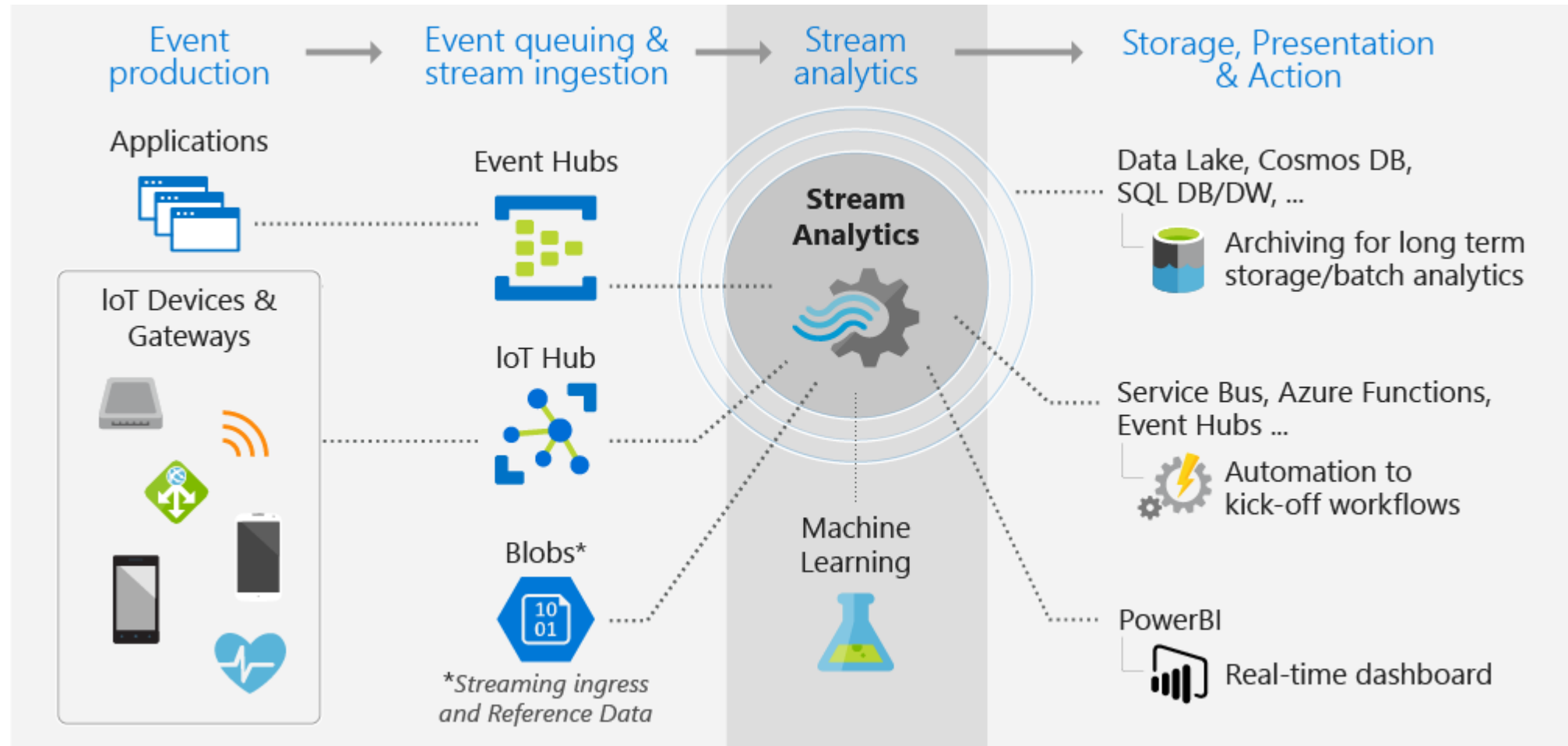


# Stream Analytics



# Stream Analytics

Event-processing engine that allows you to examine high volumes of data streaming from devices



# > Draw your own architecture...

Microsoft Azure Cloud and AI Symbol / Icon Set - SVG

– Pointer

[Azure Icons - Azure Architecture Center | Microsoft](#)

[Docs](#)

## Terms

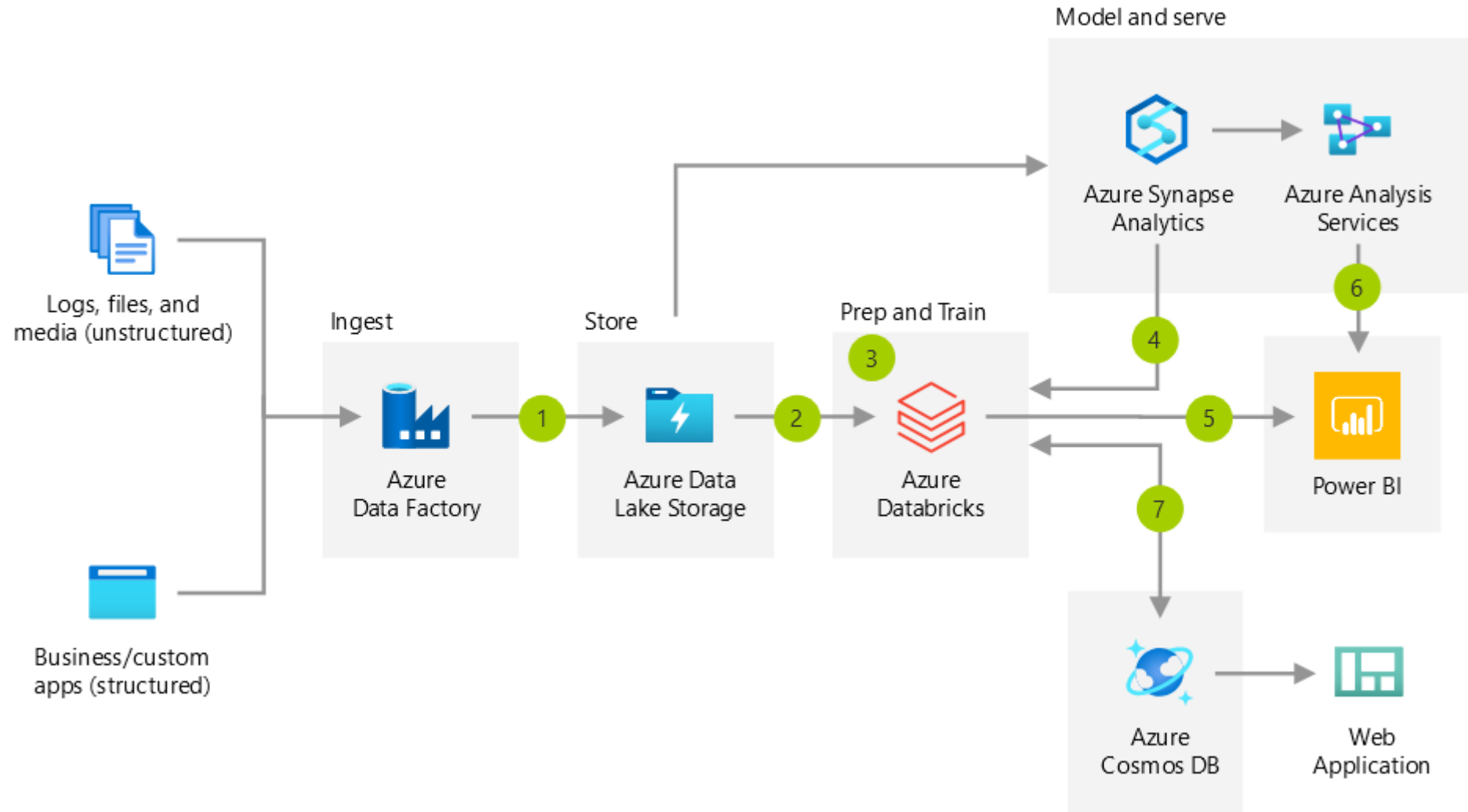
Microsoft permits the use of these icons in architectural diagrams, training materials, or documentation. You may copy, distribute, and display the icons only for the permitted use unless granted explicit permission by Microsoft. Microsoft reserves all other rights.

☒ I agree to the above terms

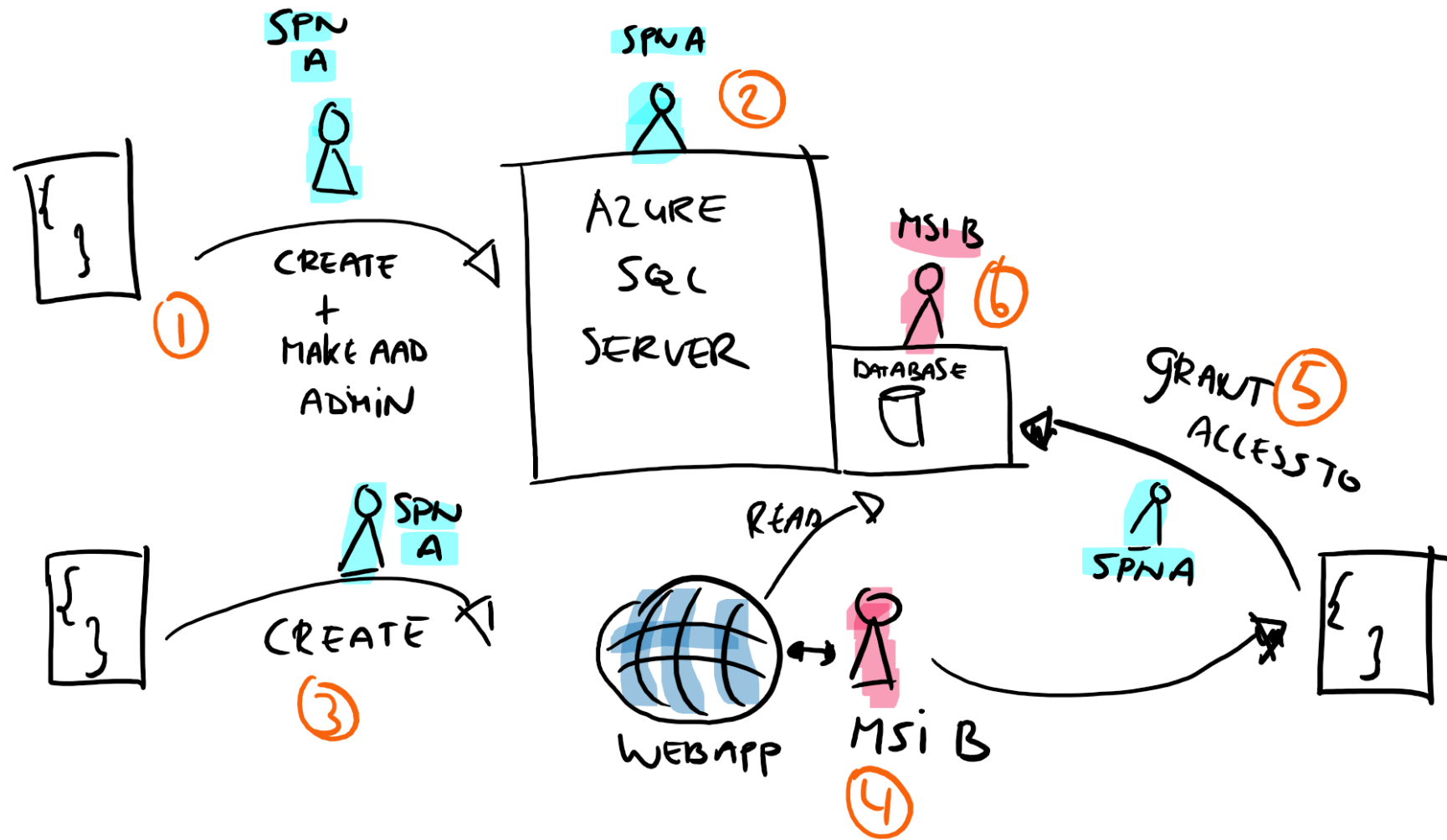
[Download SVG icons](#)



# > Why should you use professional icons?



# > Why should you use professional icons?



# Assignment #3

**Microsoft**



# > #3 Draw your own Architecture

## INSTRUCTIONS:

Think about a possible architecture / business needs based on the case detailed in the next slide.

Consider all the data sources and steps to ingest data and to show the visualizations.

- Use PowerPoint or other drawing tool (draw.io)
- <https://app.diagrams.net/> | <https://www.diagrams.net/>
- Use the standard icons for each service available
- Explain why do you select the resource / Clarify your expectations

After you finish your Architecture , please explain why you selected each of the different services and make sure that you draw the arrows showing the data flow.

## EVALUATION:

**Mark: 5 points**

**Delivery: PPT OR Video 1-4 minutes (explanations)**

Ensure that you recorded yourself explaining your data flow

Ensure that you showed all the performed steps

Data Sources / Data ingestion / Store / Process / Serve

Ensure to explain each resource and why you choose them

Will be considered:

Your results, level of detail and clarity to explain and video quality.

**Due date: Week 9 class**

SCREENCAST  MATIC

# > #3 Draw your own Architecture

## SCOPE

The STK company is a brand-new start-up responsible to deliver 95% of all the products sold by Amazon in Canada. This company is using SAP, ORACLE, and Microsoft Dynamics 365 CRM, as the main data sources. The company also has some data stored in a blob storage service on Azure (CSV files and unstructured data).

The main idea is to move ALL the data to a cloud instance (Azure). They need a unique place to store all the data and to help them to explore the data, generating data analysis, and prep the structure for future AI projects.

## SOME QUESTIONS

Your architecture should answer questions like:

- How to ingest the data from the different data sources?
- Where to store the data?
- What are the tools to perform data analysis?
- What are the resources you are planning to use for future AI projects?
- Where to process and train your data?
- What are the tools / resources to perform AI models?
- Where are you planning to generate the management data visualization? Dashboards?



# References



# > References

---

- Big Data Analytics Program, 2019/2020 – Georgian College, Barrie, Ontario
- Microsoft, Azure data platform, <https://docs.microsoft.com/en-us/azure/architecture/example-scenario/dataplate2e/data-platform-end-to-end>
- Microsoft, Data warehousing and analytics, <https://docs.microsoft.com/en-us/azure/architecture/example-scenario/data/data-warehouse>
- Microsoft, Advanced Analytics Architecture, <https://docs.microsoft.com/en-us/azure/architecture/solution-ideas/articles/advanced-analytics-on-big-data>
- Microsoft, Azure Synapse Analytics - dedicated SQL pool Videos, <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-videos>
- Microsoft, Success by Design Implementation Guide, First Edition, 2021
- Monkey Learn, Sentiment Analysis, <https://monkeylearn.com/sentiment-analysis/>
- Cloud Geeks, Jerry Hargrove, website, <https://www.lucidchart.com/blog/what-are-cloud-regions>
- Microsoft, Authentication, Microsoft Docs, <https://docs.microsoft.com/en-us/learn/modules/recognize-dynamics-365-security/4-authentication>
- Microsoft, Dataverse, Microsoft Docs, <https://docs.microsoft.com/en-us/learn/modules/connect-analyze-dynamics-365-data/3-benefits-dataverse>
- Microsoft, Azure Data Platform End-to-End, Implement a Modern Data Platform Architecture, Official Material



# Georgian

END OF DAY 7