



Week 7

Artificial Intelligence Program Infrastructure and Architecture

> Agenda // Program

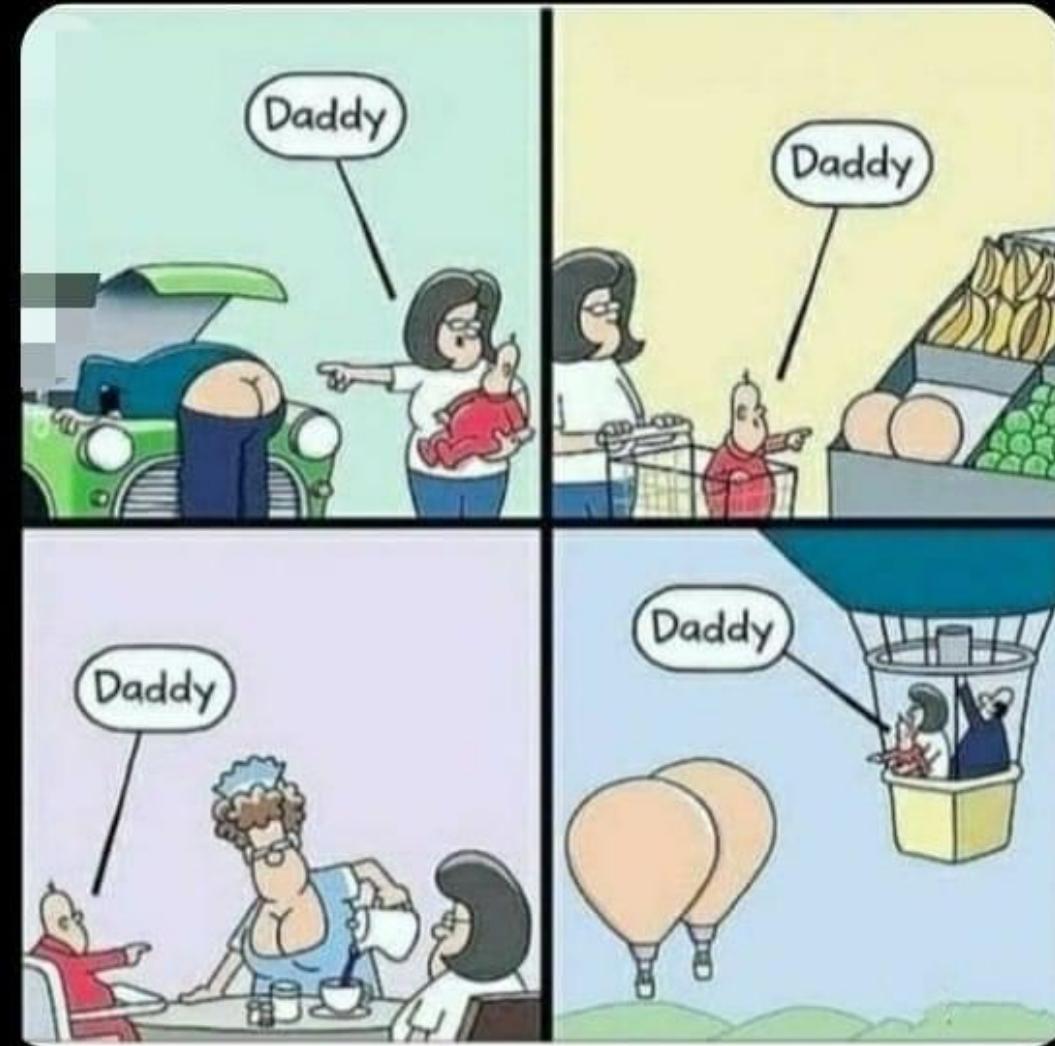
WEEK	SUBJECT	ASSIGNMENT / TO BE DELIVERED	CHALLENGES
2	Intro / AI Function / Enablers		
3	Infra and Architecture / On-prem vs. Cloud / CSPs		C 1
4	Data Pipeline / Processes / Framework / AutoML	#1 Image Classifier [5%]	
5	Data Pipeline / Processes / Framework / AutoML		C 2
6	More Data / SSIS / ADF / Data Quality	#2 Machine Learning Studio [10%]	
7	Azure services – Intro	EXAM 1 [20%]	C 3
8	READING WEEK	NO CLASSES	
9	Azure Cognitive Services 1		4.1
10	Azure Cognitive Services 2	#3 Draw your own Architecture [5%]	4.2
11	Azure Cognitive Services 3		4.3
12	Azure Cognitive Services 4	#4 Azure Pipeline / Sentiment Analysis [10%]	4.4
13	AWS Academy / Cloud Foundations		
14	AWS Academy / Machine Learning		#5 AWS Academy // Cloud Foundations [10%]
15	Enterprise Architecture	EXAM 2 [20%]	#6 AWS Academy // Machine Learning [10%]

> Agenda (7)

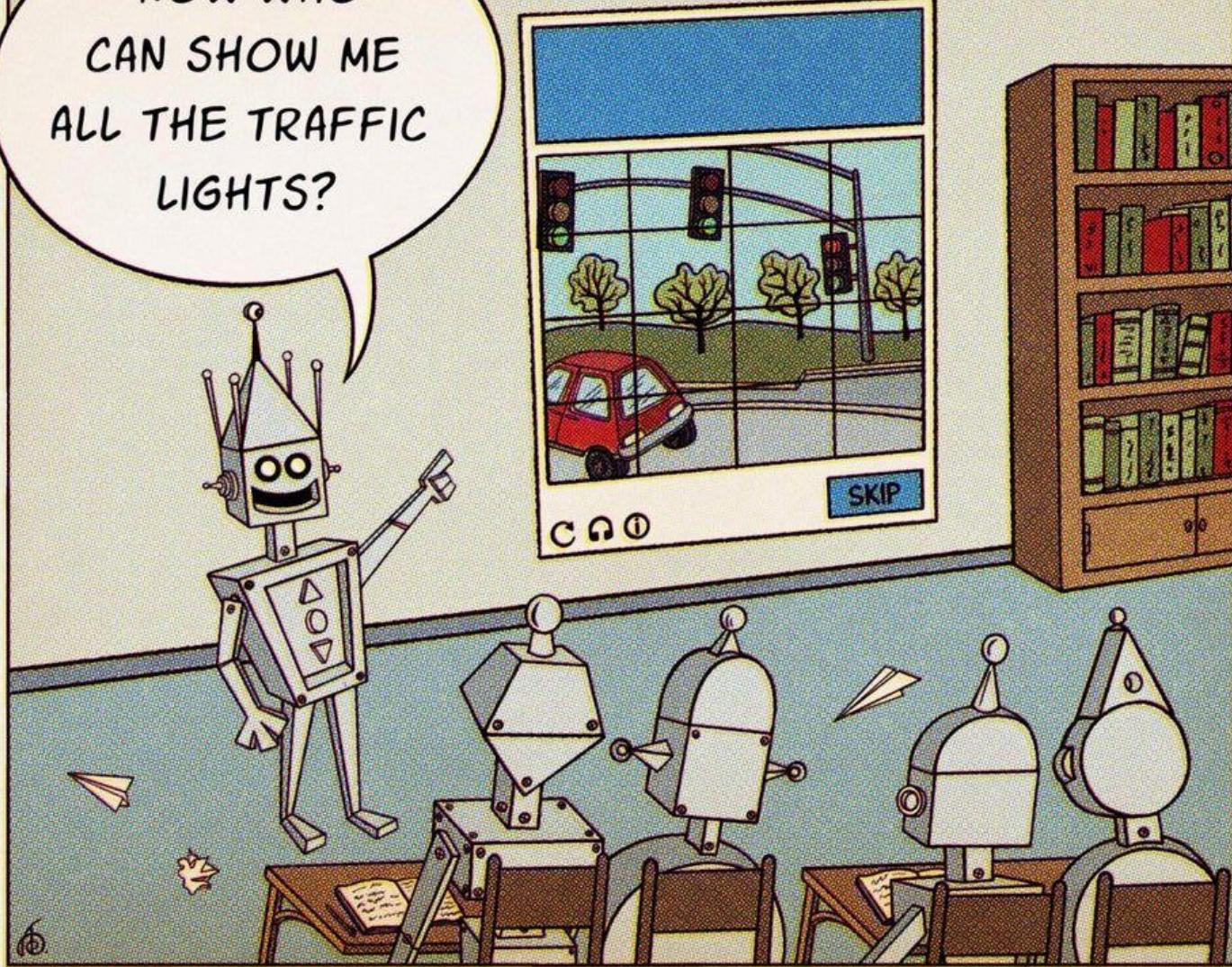
- Modern Data Platform Concepts
 - Data Lake
 - Azure Data Lake Storage Gen2
 - Azure ML
 - Azure AI
 - Event Hubs / Example
 - Stream Analytics
- Developing your own Architecture
 - Practice exercise – drawing your own architecture

Machine Learning and AI be like..

fb.me/yuva.krishna.memes



NOW WHO
CAN SHOW ME
ALL THE TRAFFIC
LIGHTS?



@ROBOTOPIAWEEKLYCOMIC

© 2019 ANGIE SIVERIA & OSKAR SCHUSTER

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

ZB

GB

Batch

Structured
data

Unstructured
data

VELOCITY

Real-time

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

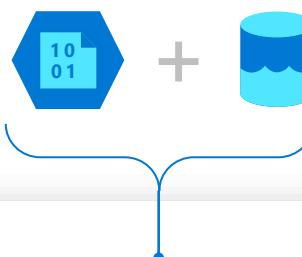


Integration ready

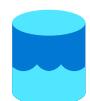
Optimized for Spark and Hadoop Analytic Engines

Tightly integrated with Azure end to end analytics solutions

Multiprotocol access

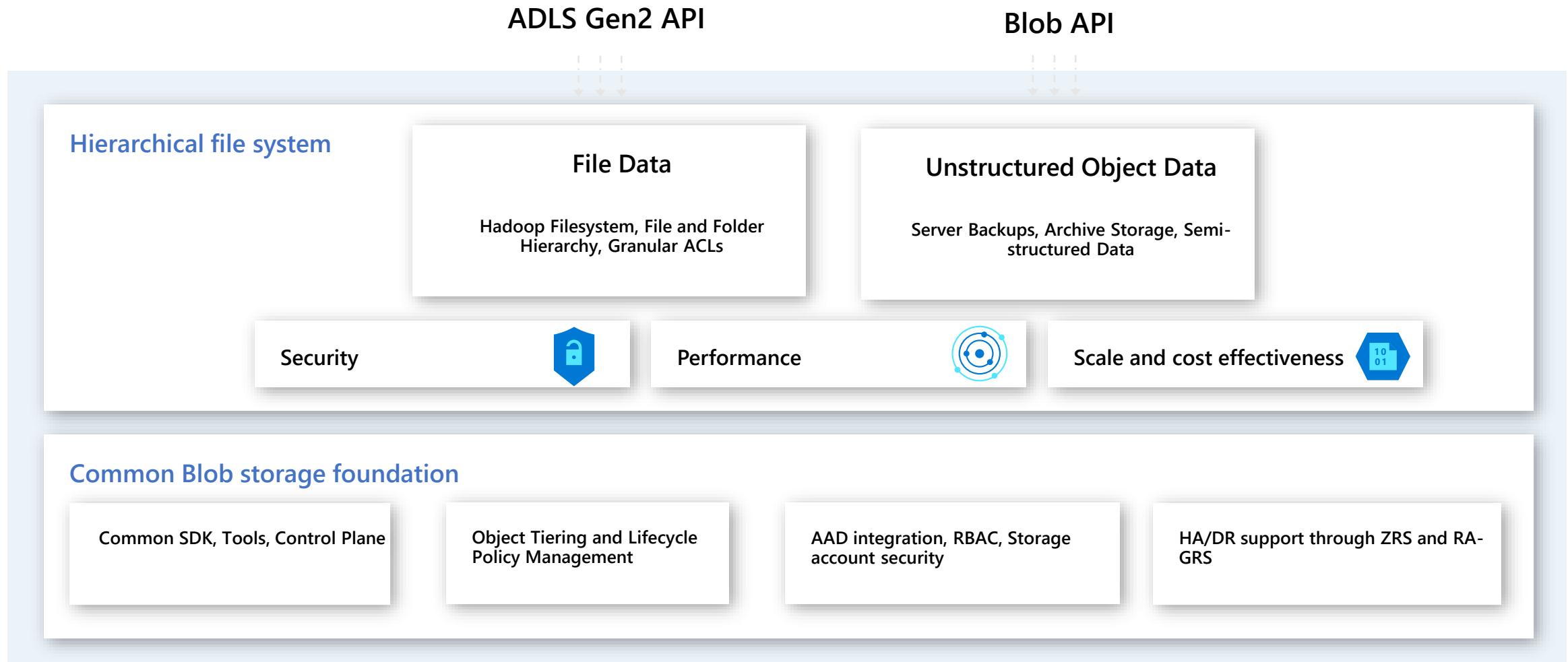


Single service



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

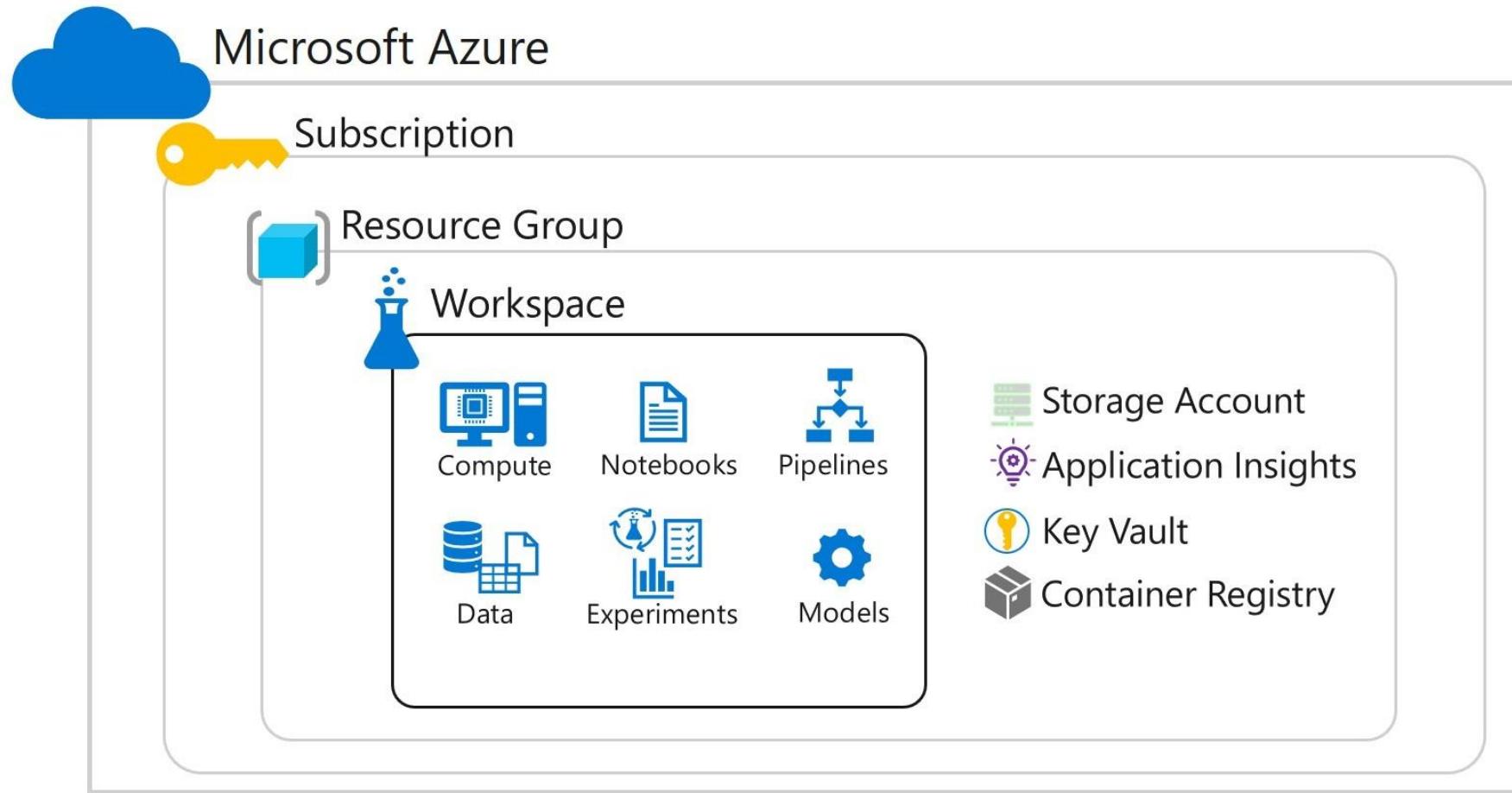
A workspace is a context for the experiments, data, compute targets, and other assets associated with a machine learning workload.

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.

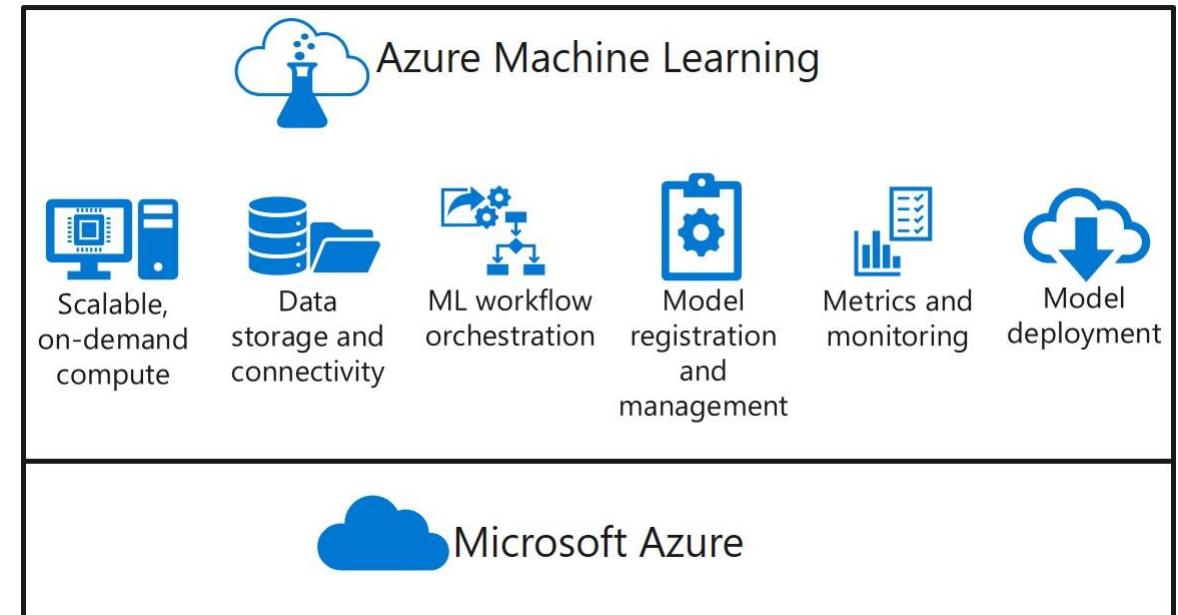
> Azure ML Workspaces



> 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

The screenshot shows the Azure Machine Learning studio interface. The left sidebar includes sections for New, Home, Author, Notebooks, Automated ML, Designer, Assets (Datasets, Experiments, Pipelines, Models, Endpoints), Compute, Datastores, and Data labeling. The main area features a 'Welcome to the studio!' section with four cards: 'Create new' (with a plus icon), 'Notebooks' (with a code icon), 'Automated ML' (with a gear icon), and 'Designer' (with a cube icon). Below this is a 'My recent resources' section with tables for 'Runs' and 'Compute'. The 'Runs' table lists several entries with columns for Run number, Experiment, Update..., and Status. The 'Compute' table lists two entries with columns for Name, Type, Provisioning status, and Creation date.

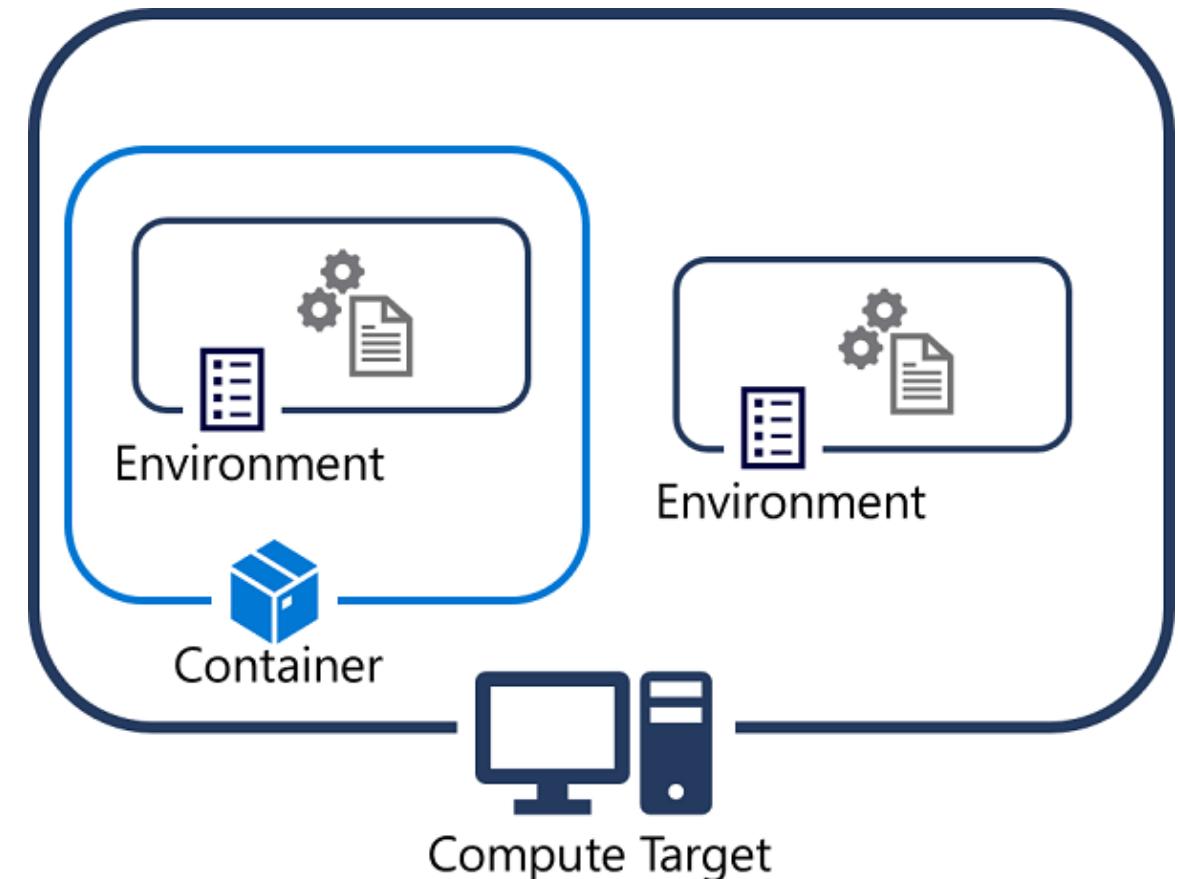
Name	Type	Provis...	Creat...
aml-cluster	Machine Learnin...	Success	Decem...
aml-instance	Compute Instanc...	Success	Decem...

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

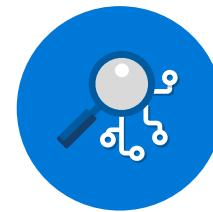
AI apps and agents



Azure Cognitive Services

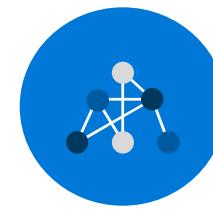
Azure Bot Service

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



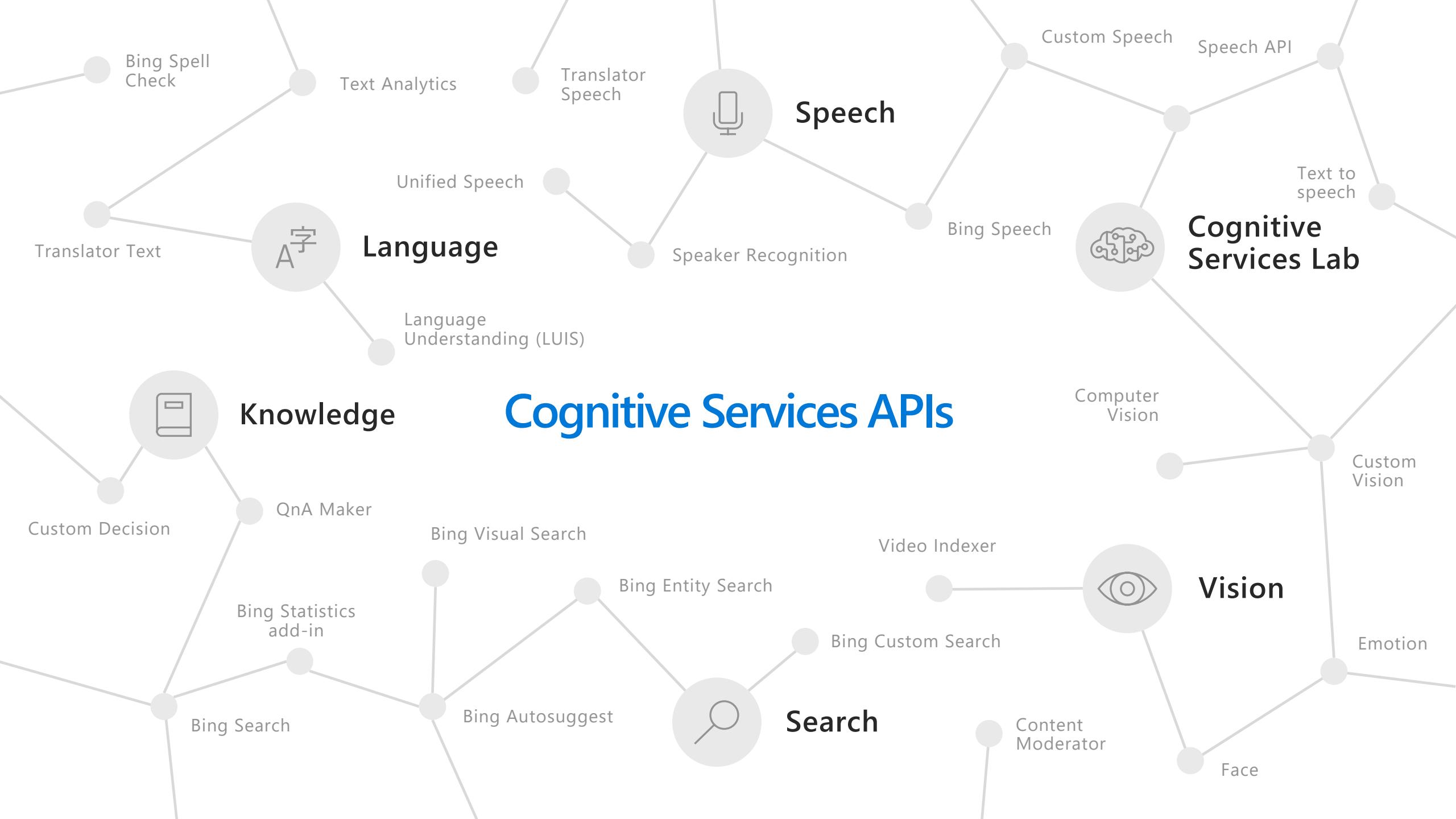
FPGA



From the Intelligent Cloud to the Intelligent Edge



Cognitive Services APIs



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

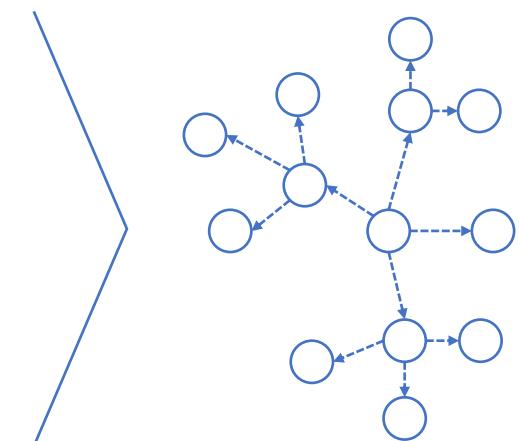


Tag extraction

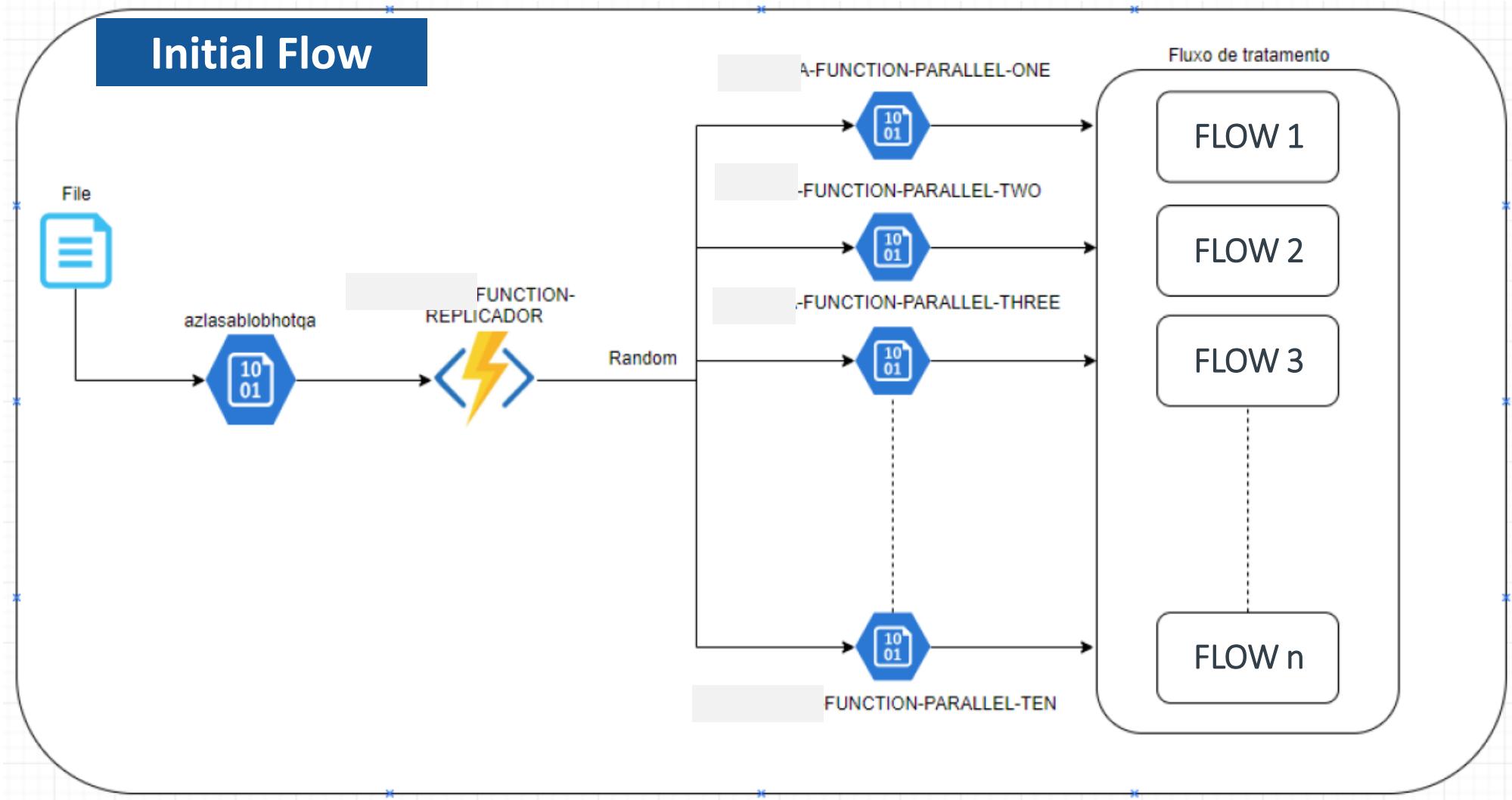


Printed text recognition

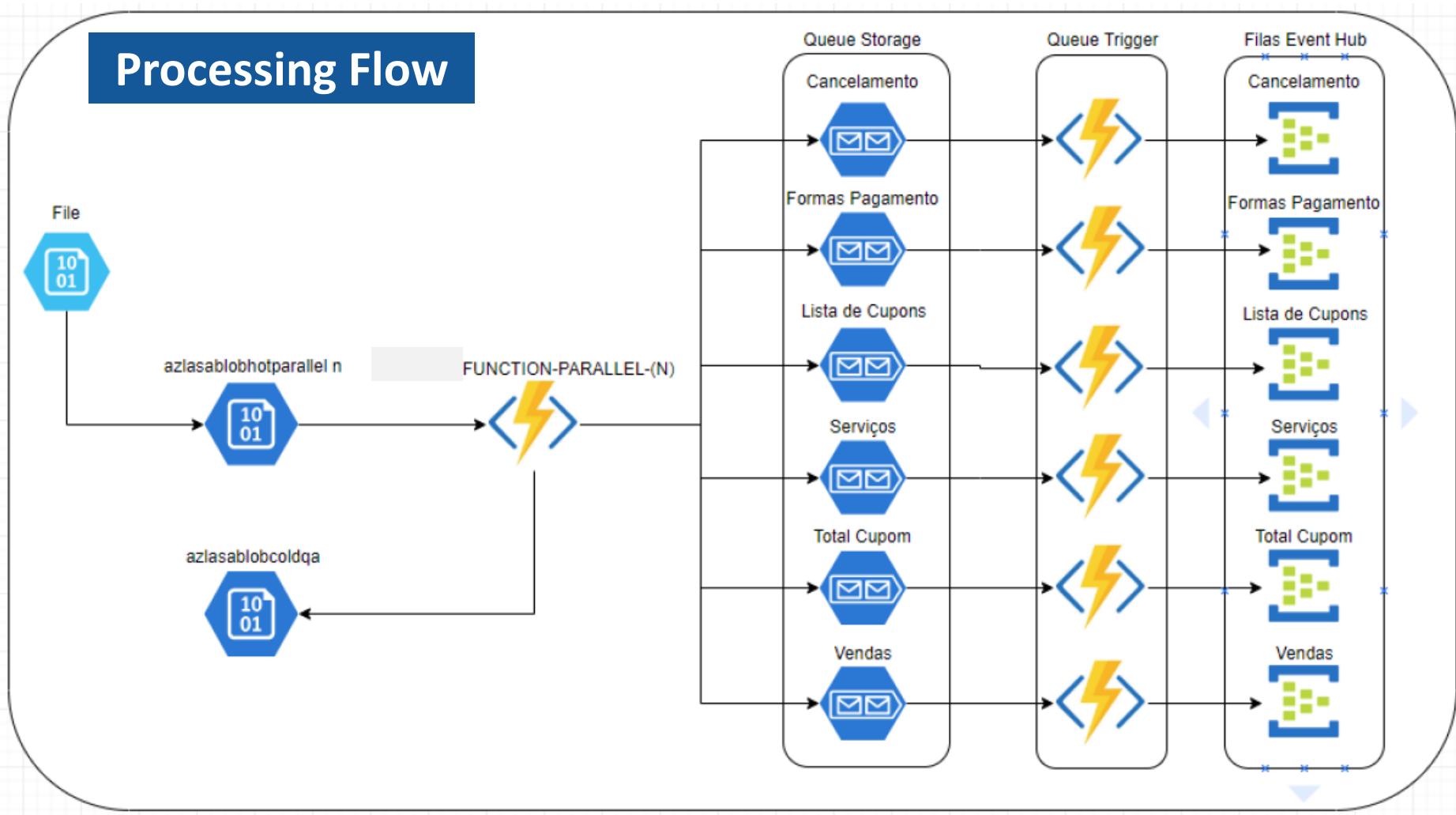
Fully text-searchable rich index



> Case // Example // Parallel processes



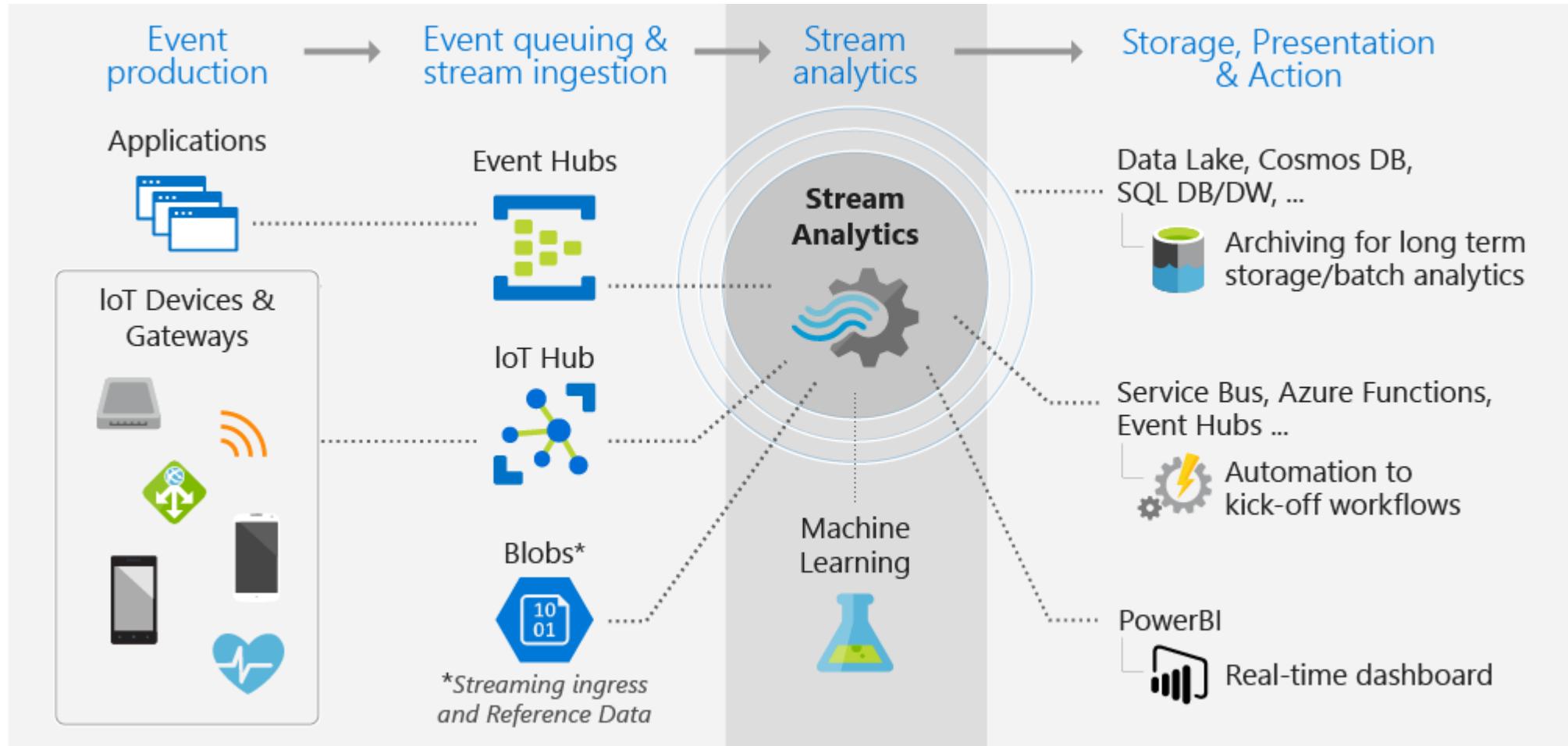
> Case // Example // Flow Detail



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](#)

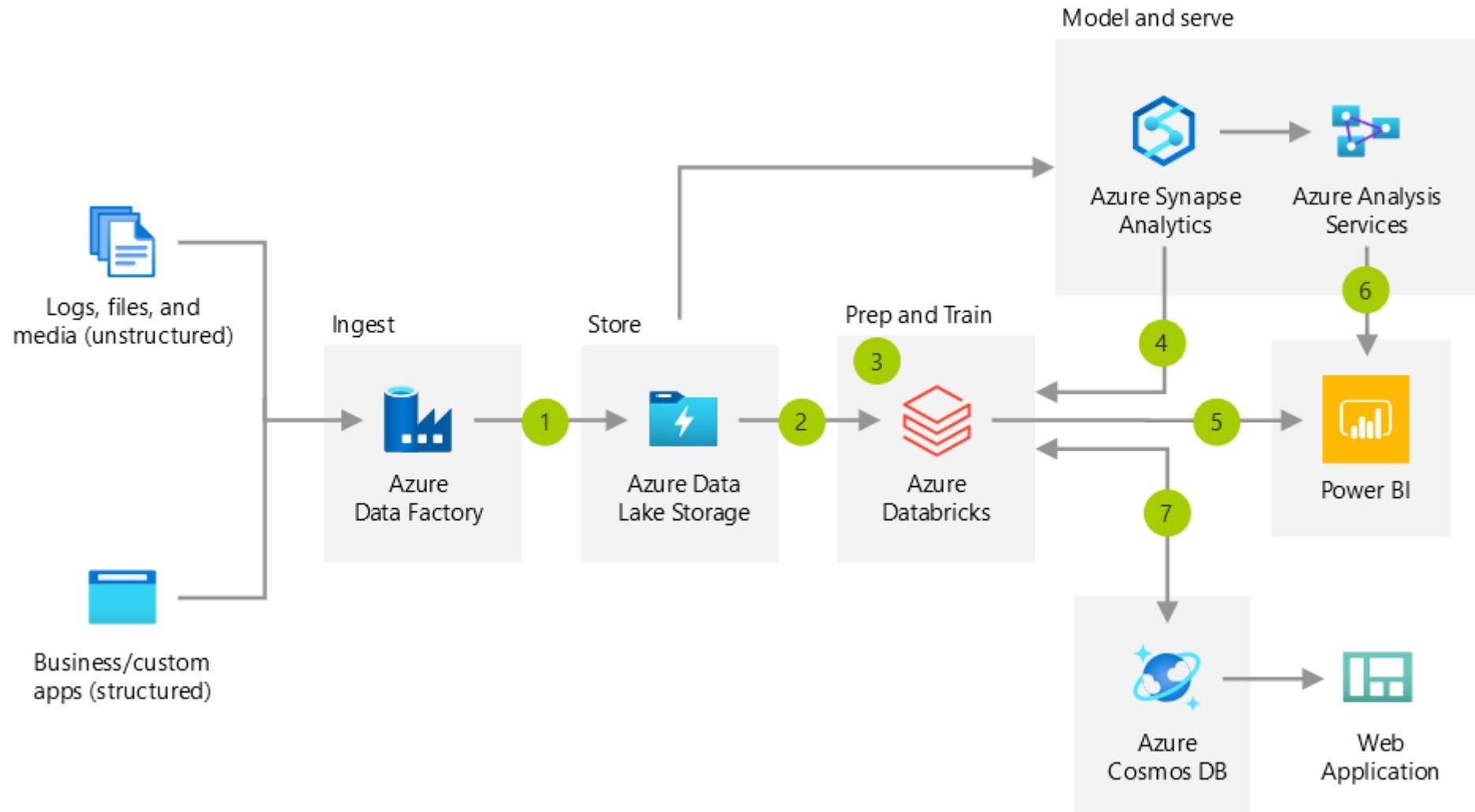
Microsoft permits the use of these icons in architectural diagrams, training materials, or documentation. You may copy, distribute, and display the icons only for the permitted use unless granted explicit permission by Microsoft. Microsoft reserves all other rights.

I agree to the above terms

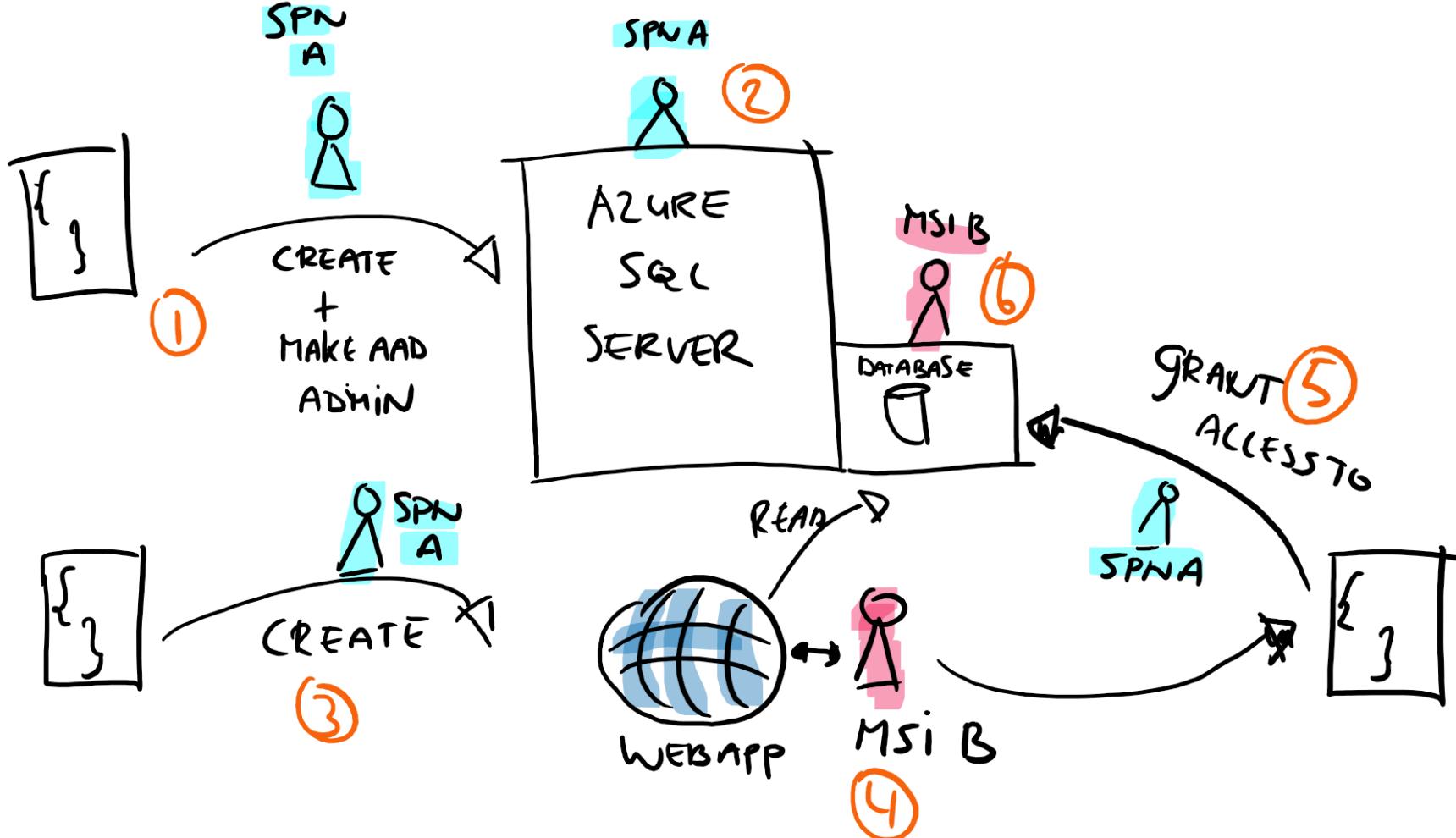
[Download SVG icons](#)



> 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: Please confirm on the blackboard

> #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?

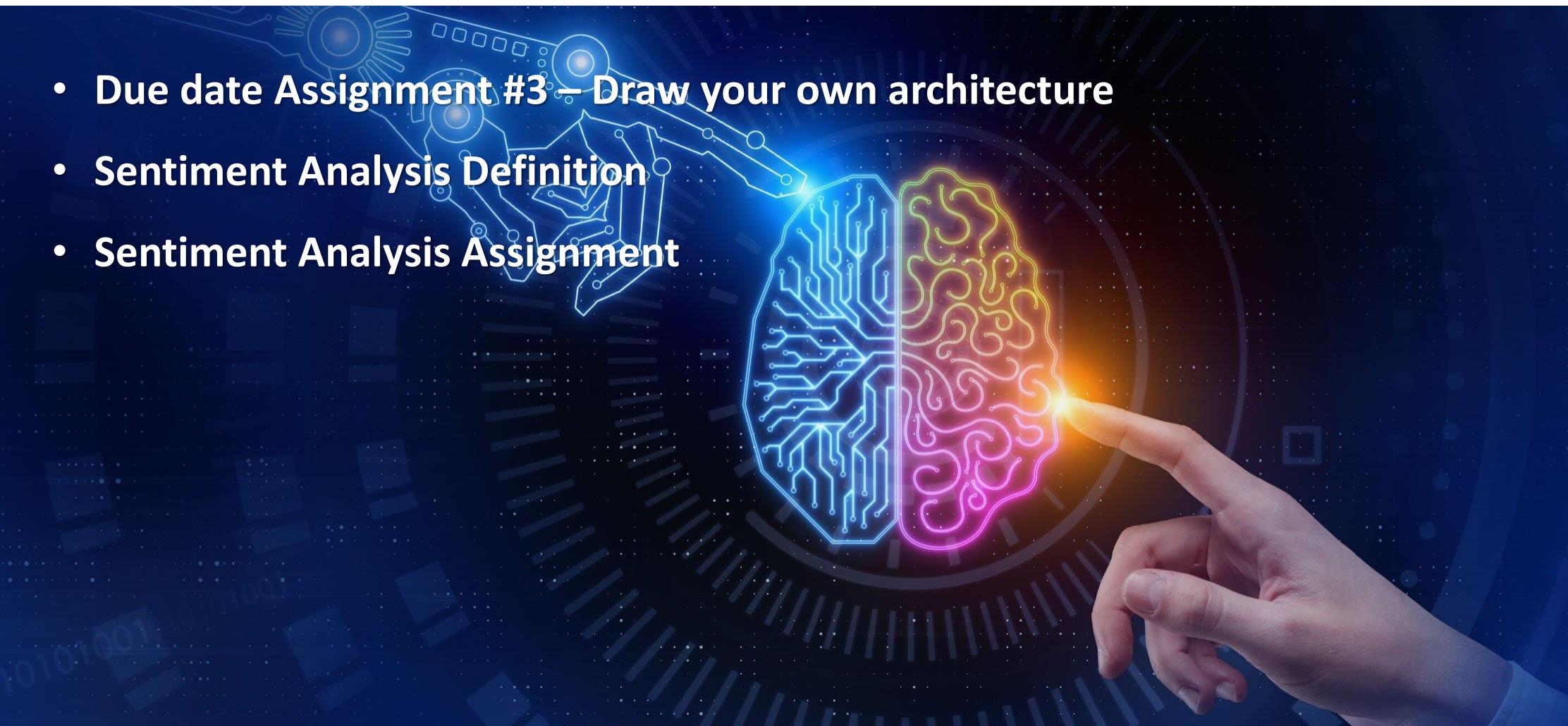


Georgian

END OF DAY 8

> Agenda (9)

- Due date Assignment #3 – Draw your own architecture
- Sentiment Analysis Definition
- Sentiment Analysis Assignment



Azure Week 3

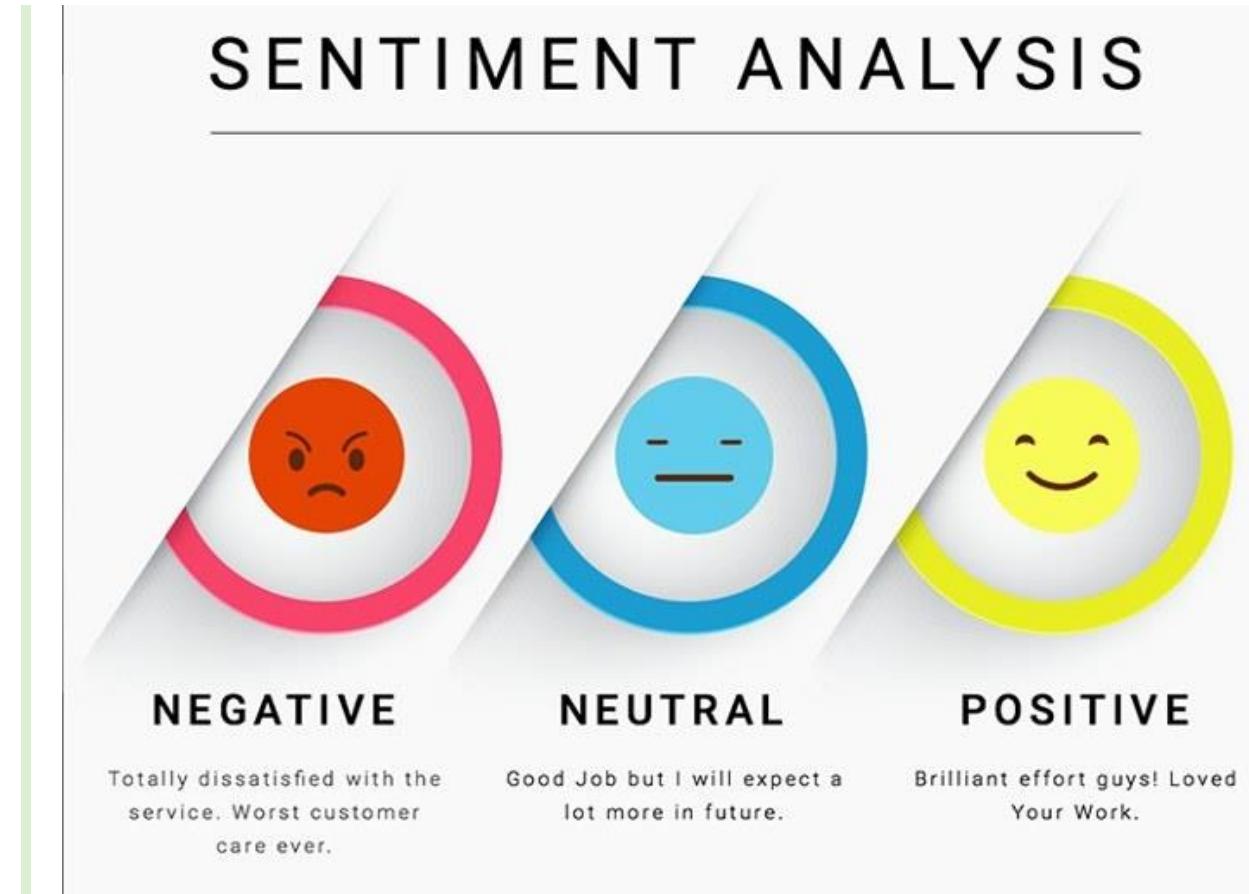
Microsoft



> Sentiment Analysis | Practice

Sentiment analysis (also known as **opinion mining** or **emotion AI**) refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.



> Sentiment Analysis | Practice

Sentiment analysis is the process of detecting positive or negative sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers.

The most popular applications of sentiment analysis in real life:

- Social media monitoring
- Customer support
- Customer feedback
- Brand monitoring and reputation management
- Voice of customer (VoC)
- Voice of employee
- Product analysis
- Market research and competitive research

> #4 Sentiment Analysis Assignment

INSTRUCTIONS:

Record a video with 10-15 minutes explaining how to create a **Sentiment Analysis solution** and your challenges.

- Use your Azure instance (student)
- Select your data sources
- Perform all the steps to acquire data
- Select your preferred lib (python or other)
- Use ADF to ingest data to your preferred database
- Generate visuals using your preferred visualization tool (PowerBI, Tableau, etc.)
- Validate the results
- Evaluate the results

In your video explain how you performed each step and interpret the result. You can use the example below as a reference.

Be prepared to present your video / solution in our next class.

EVALUATION:

Mark: 10%

Ensure that you recorded yourself using the tool

Ensure that you showed all the performed steps

Ensure that you analyse the results

Explain what kind of ML you are using in this exercise and why

Will be considered:

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

Due date: Please confirm on the blackboard

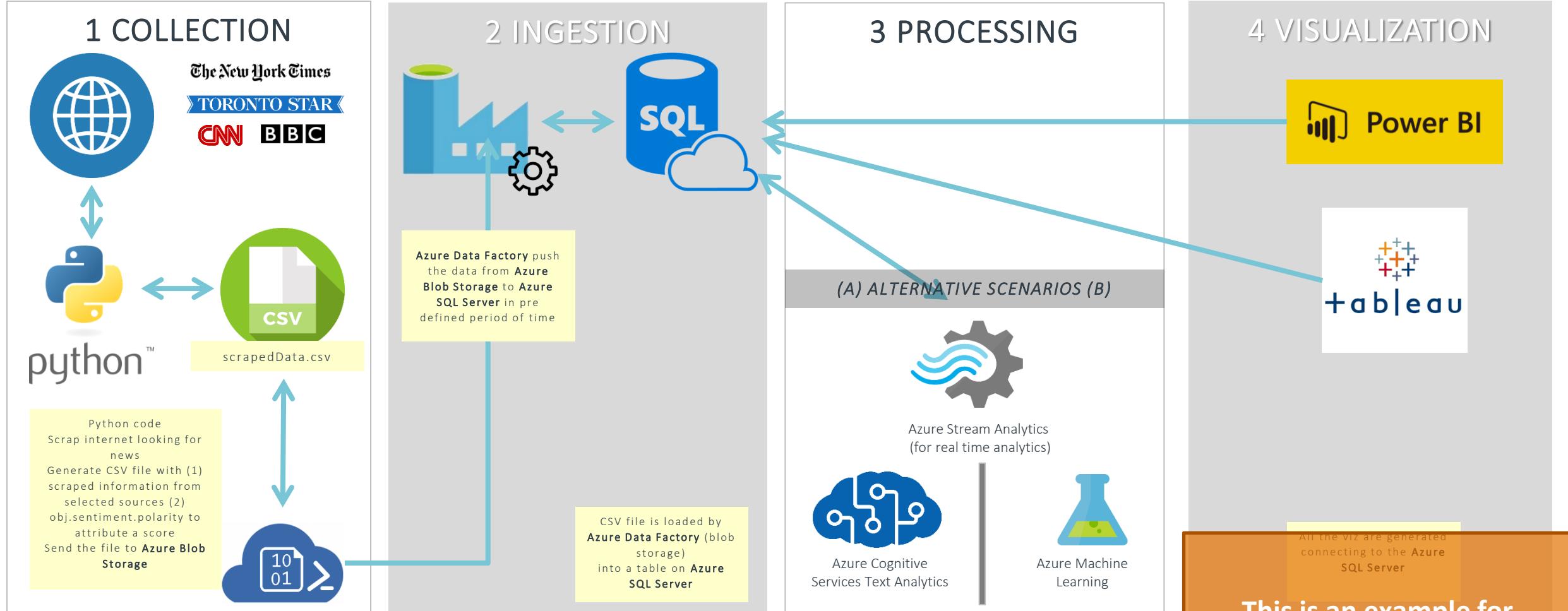
Practice #4

Sentiment Analysis



> Sentiment Analysis | Practice

PROJECT ARCHITECTURE

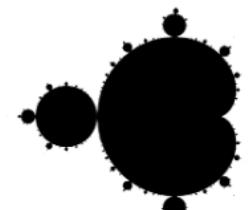


> Sentiment Analysis | Practice

TEXT BLOB LIBRARY

SENTIMENT ANALYSIS

```
from textblob import TextBlob
sentiment =
obj.sentiment.polarity
Features
• Noun phrase extraction
• Part-of-speech tagging
• Sentiment analysis
• Classification (Naive Bayes, Decision Tree)
• Language translation and detection powered by
Google Translate
• Tokenization (splitting text into words and
sentences)
• Word and phrase frequencies
• Parsing
• n-grams
• Word inflection (pluralization and singularization)
and lemmatization
• Spelling correction
• Add new models or languages through extensions
• WordNet integration
$ pip install -U textblob
$ python -m textblob.download_corpora
```



TextBlob

Star 6,859

TextBlob is a Python (2 and 3) library for processing textual data. It provides a consistent API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, and more.

Useful Links

[TextBlob @ PyPI](#)
[TextBlob @ GitHub](#)
[Issue Tracker](#)

Stay Informed

Follow @sloria

TextBlob: Simplified Text Processing

Release v0.15.2. [\(Changelog\)](#)

TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

```
from textblob import TextBlob
text = ...
The titular threat of The Blob has always struck me as the ultimate movie
monster: an insatiably hungry, amoeba-like mass able to penetrate
virtually any safeguard, capable of--as a doomed doctor chillingly
describes it--"assimilating flesh on contact.
Snide comparisons to gelatin be damned, it's a concept with the most
devastating of potential consequences, not unlike the grey goo scenario
proposed by technological theorists fearful of
artificial intelligence run rampant.
...
blob = TextBlob(text)
blob.tags           # [('The', 'DT'), ('titular', 'JJ'),
# ('threat', 'NN'), ('of', 'IN'), ...]
blob.noun_phrases # WordList(['titular threat', 'blob',
#                                'ultimate movie monster',
#                                'amoeba-like mass', ...])
for sentence in blob.sentences:
    print(sentence.sentiment.polarity)
```

This is an example for
academic purposes

> Sentiment Analysis | Practice

NLTK 3.5 documentation

NEXT | MODULES | INDEX

Natural Language Toolkit

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to [over 50 corpora and lexical resources](#) such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active [discussion forum](#).

Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, plus comprehensive API documentation, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source.



Search projects



Help Sponsor Log in Register

gensim 3.8.3

pip install gensim



Python framework for fast Vector Space Modelling

Navigation

Project description

build passing wheel yes

Gensim is a Python library for *topic modelling*, *document indexing* and *similarity retrieval* with large corpora. Target audience is the *natural language processing* (NLP) and *information retrieval* (IR) community.

Features

- All algorithms are **memory-independent** w.r.t. the corpus size (can process input larger than RAM, streamed, out-of-core),
- Intuitive interfaces
 - easy to plug in your own input corpus/stream (simple streaming API)

spaCy Out now: spaCy v3.0

USAGE MODELS API UNIVERSE



19,512



Search docs

Industrial-Strength Natural Language Processing

IN PYTHON

Done

Blazing fast

Awesome ecosystem

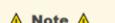
StanfordNLP

[Github repo](#) [Quick links](#)

Latest version



Released: May 3, 2020



All development, issues, ongoing maintenance, and support have been moved to our [new GitHub repository](#) as the toolkit is being renamed as Stanza since version 1.0.0. Please visit our [new website](#) for more information. You can still download stanfordnlp via pip, but newer versions of this package will be made available as stanza. This site is kept for archival purposes.

StanfordNLP

[PyPI](#) v0.2.0 [python](#) 3.6 | 3.7

Overview Usage Pipeline Resources

StanfordNLP 0.2.0 - Python NLP Library for Many Human Languages

Table of Contents

- About
- Get Started
- License
- Citing StanfordNLP in papers
- Links



All development, issues, ongoing maintenance, and support have been moved to our [new GitHub repository](#) as the toolkit is being renamed as Stanza since version 1.0.0. Please visit our [new website](#) for more information. You can still download stanfordnlp via pip, but newer versions of this package will be made available as stanza. This site is kept for archival purposes.

This is an example for academic purposes

> Sentiment Analysis | Practice

PYTHON (TEXT BLOB LIB) CODE RUNNING...

```
Command Prompt - python sentimental_analysis.py
6 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/QiHRxemDpA8/index.html
7 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/JJjL1CmYXo4/index.html
8 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/IIZ-KmkgRdM/index.html
9 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/0qZf_v-EnOQ/index.html
10 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/7s0xgqltfzw/spicous-device-in-car.html
er-biden-nr-vpx.cnn
11 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/b-CZarEugen0/iwith-complicated-results.html
12 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/xgu0Pc6inAw/i13 articles from Toronto Star @ https://www.thestar.com/news/world/us/2020/03/06/washington-dc-gets-its-first-coronavirus-pop-up-shop.html
13 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/-OfGKZ1SHLw/i14 articles from Toronto Star @ https://www.thestar.com/news/world/europe/2020/03/06/asylum-seekers-coronavirus-collide-with-greece.html
14 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/UtuitD5jEH0/i15 articles from Toronto Star @ https://www.thestar.com/news/world/middleeast/2020/03/05/turkey-vows-justice-for-migrants-starr-dnt-lead-vpx.cnn
15 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/vT2T03cBPsk/i16 articles from Toronto Star @ https://www.thestar.com/news/world/eu/ministers-tackle-risks-of-shortages.html
16 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/C2jMr8LCPDc/i17 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/4e8NUM59Cv4/i18 articles from Toronto Star @ https://www.thestar.com/news/world/asia/2020/03/06/indian-depositors-scramble-for-money-ta-nr-vpx.cnn
17 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/7_IlWCVh4sg/i19 articles from Toronto Star @ https://www.thestar.com/news/world/asia/2020/03/06/as-virus-cases-near-100000-fear-of-death-and-it-may-affect-their-lifespan-and-warranty-bombardier-says.html
18 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/_ks27Yfkq70/i20 articles from Toronto Star @ https://www.thestar.com/news/world/middleeast/2020/03/06/ap-explains-militant-fighters-in-final-battle-in-syria.html
19 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/iB9NYmjbh0I/i21 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/4L6BczFsAfw/i22 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/qh30HcwmHVA/i23 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/7Rkc6-28_Dk/i24 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/V_QreGWLLSk/i25 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/EfiJQCfs-i0/i26 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/Nr0FH4kfR_4/i27 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/WRJHX_wI8qU/i28 articles from CNN @ http://rss.cnn.com/~r/rss/cnn_topstories/~3/enwiid8aZq0/ekg-ebof-vpx.cnn
Uploading to Blob storage as blob -> scrapedData.csv
List blobs in the container
Blob name: mydata.csv
Blob name: mydata.json
Blob name: scrapedData.csv
File SUCESSFULLY transferred to blob storage C:\sentimental_analysis\scrapedData.csv
C:\sentimental_analysis>
```

This is an example for
academic purposes

> Sentiment Analysis | Practice

AZURE DATA FACTORY - DATA INGESTION PROCESS

The screenshot illustrates the Azure Data Factory interface for a 'Copy Data' pipeline named 'rl-datafactory-csv'. The pipeline is configured to copy data from an 'Azure Blob Storage' source (webscraping, Region: Brazil South) to an 'Azure SQL Database' destination (1 table(s), Region: East US). The 'Copy Run Time Region' is set to 'East US'. The deployment process has completed successfully, as indicated by the 'Deployment complete' message and the validation log:

- Validation runtime environment: Validation passed
- Registering Connections: ✓
- Creating Datasets: ✓
- Creating Pipelines: ✓

A link is provided to 'Click here to monitor copy pipeline'.

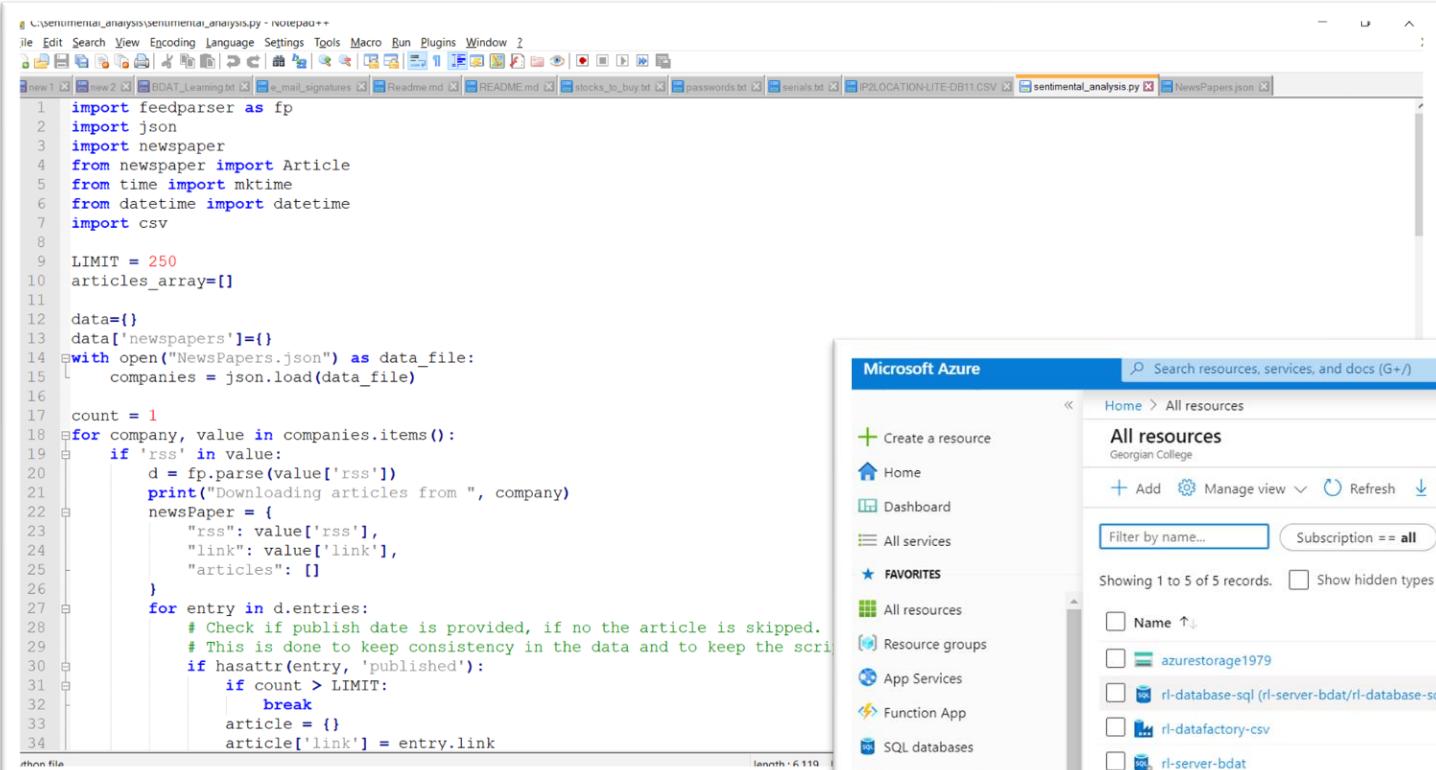
The main interface shows the pipeline structure with three activities: InputDataset-sfj (AZURE BLOB STORE...), Activity-0-scrapedData_csv->[dbo]_[t1...], and OutputDataset-sfj (FREQ: DAY INTVL: 1 AZURE SQL DATABASE...).

Activity Windows section: Pipeline, Activity, Window Start, Window End, Status. Status: There are currently no activity windows to display.

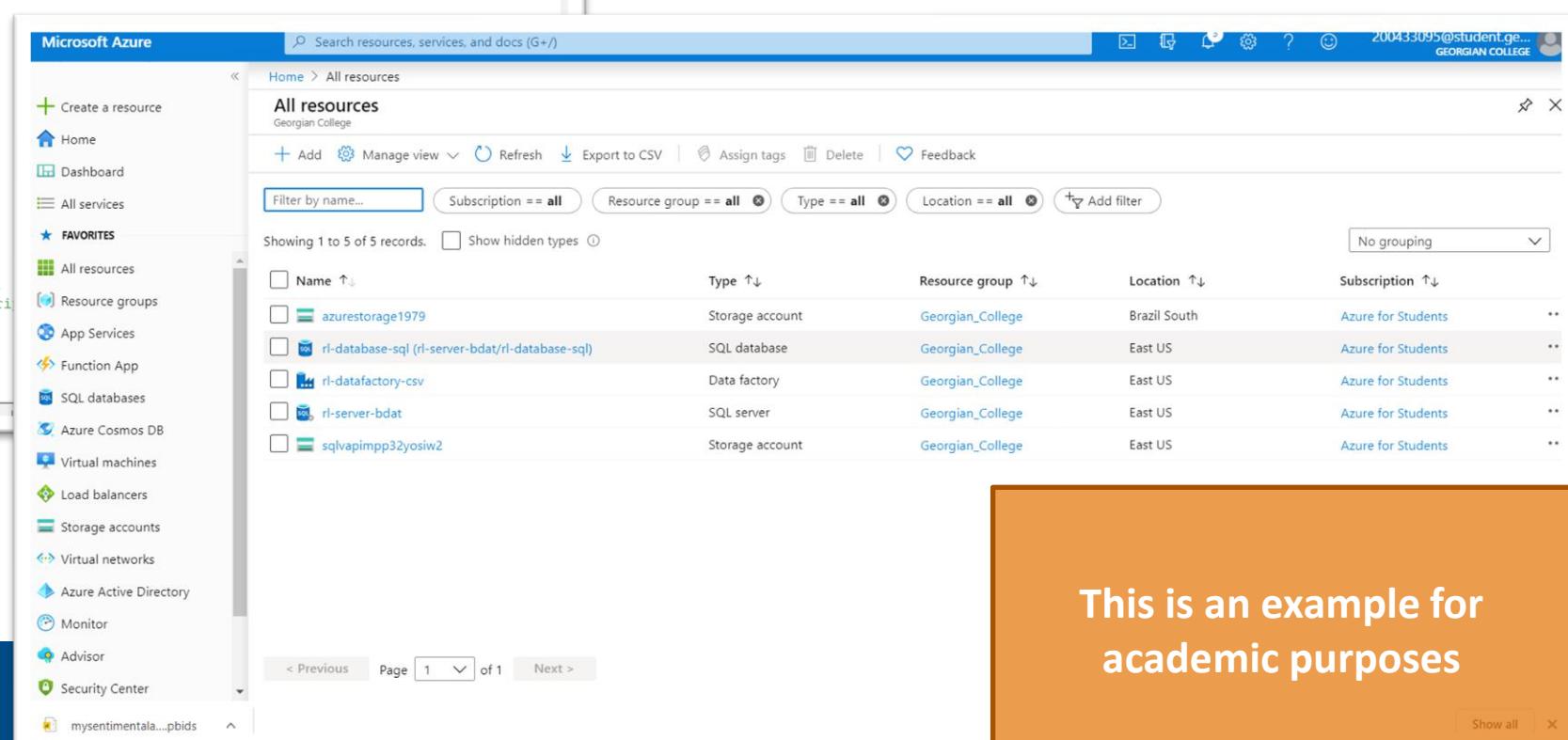
This is an example for academic purposes

> Sentiment Analysis | Practice

PYTHON CODE AND AZURE RESOURCES



```
sentimental_analysis.py ~ ivotepeau++  
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window 2  
new1 new2 BDAT_Learning.txt e_email_signatures Readme.md README.md stocks_to_buy.txt passwords.txt serials.txt IPLOCATION-LITE-DB11.CSV sentimental_analysis.py NewsPapers.json  
1 import feedparser as fp  
2 import json  
3 import newspaper  
4 from newspaper import Article  
5 from time import mktime  
6 from datetime import datetime  
7 import csv  
8  
9 LIMIT = 250  
10 articles_array=[]  
11  
12 data={}  
13 data['newspapers']= {}  
14 with open("NewsPapers.json") as data_file:  
15     companies = json.load(data_file)  
16  
17 count = 1  
18 for company, value in companies.items():  
19     if 'rss' in value:  
20         d = fp.parse(value['rss'])  
21         print("Downloading articles from ", company)  
22         newsPaper = {  
23             "rss": value['rss'],  
24             "link": value['link'],  
25             "articles": []  
26         }  
27         for entry in d.entries:  
28             # Check if publish date is provided, if no the article is skipped.  
             # This is done to keep consistency in the data and to keep the script  
             # running  
29             if hasattr(entry, 'published'):   
30                 if count > LIMIT:  
31                     break  
32                 article = {}  
33                 article['link'] = entry.link  
34  
about file  
length: 6119
```



Microsoft Azure Search resources, services, and docs (G+) 20043095@student.ge... GEORGIAN COLLEGE

All resources Georgian College

+ Add Manage view Refresh Export to CSV Assign tags Delete Feedback

Filter by name... Subscription == all Resource group == all Type == all Location == all + Add filter

Showing 1 to 5 of 5 records. Show hidden types

<input type="checkbox"/> Name ↑	Type ↑	Resource group ↑	Location ↑	Subscription ↑
azurestorage1979	Storage account	Georgian_College	Brazil South	Azure for Students
rl-database-sql (rl-server-bdat/rl-database-sq)	SQL database	Georgian_College	East US	Azure for Students
rl-datafactory-csv	Data factory	Georgian_College	East US	Azure for Students
rl-server-bdat	SQL server	Georgian_College	East US	Azure for Students
sqlvapimpp3yosiw2	Storage account	Georgian_College	East US	Azure for Students

< Previous Page 1 of 1 Next >

mysentimental...pbds

This is an example for academic purposes

> Sentiment Analysis | Practice

SAMPLE DASHBOARD - POWER BI

CNN Sentiment Analysis FINAL ASSIGNMENT

Sentiment Breakdown

Sentiment	Count	Percentage
Positive	91	43.7%
Negative	105	50.0%
Neutral	12	5.77%

Coronavirus Word Cloud

Article by Score

Title	Year	Text
4-year-old only child sings 'All By Myself' in quarantine	2020	Zachary Frenette likes working as an Uber driver in Phoenix. He is a top-rated driver who often chats with his customers on their trips.
A 7-year-old boy was shot in the face and killed in a drive-by shooting	2020	During the outbreak of the coronavirus last month, business began to slow. Then, a possible exposure to the virus prompted Mr. Frenette, 29, to quarantine himself. Off the road, he had heard that the ride-hailing app was one of several companies that announced policies to offer
A fire at a Florida airport destroyed more than 3,500 rental cars	2020	He had heard that the ride-hailing app was one of several companies that announced policies to offer
A Historic Unemployment Crisis	2020	He had heard that the ride-hailing app was one of several companies that announced policies to offer
A New Covid-19 Crisis: Domestic Abuse Rises Worldwide	2020	He had heard that the ride-hailing app was one of several companies that announced policies to offer
A Quarantined Uber Driverâ€™s Quest for Paid Sick Leave	2020	He had heard that the ride-hailing app was one of several companies that announced policies to offer
A slow-moving storm will dump feet of snow and flooding rain across California Monday	2020	He had heard that the ride-hailing app was one of several companies that announced policies to offer

> Sentiment Analysis | Practice

SAMPLE DASHBOARD - POWER BI

CNN Sentiment Analysis FINAL ASSIGNMENT

Sentiment Breakdown

Count of Score 39

Positive 19 (9.1%)

Negative 18 (8.6%)

Neutral 2 (0.96%)

Picture from article

Article by Score

Title	Year	Text
Boris Johnson is under observation in hospital for coronavirus but 'remains in charge'	2020	Western Europe may have reached an important turning point in the coronavirus epidemic: while the total number of patients continues to climb, the rate of new infections is no longer rising.
Bronx Zoo Tiger Is Sick With the Coronavirus	2020	The shift seems clearest in the two hardest-hit countries, Italy and Spain, though incomplete and inconsistent data make it hard to be sure.
CNN's Fareed Zakaria: The US has abandoned this crucial role during the coronavirus	2020	Italy's daily tally of confirmed new infections peaked on March 21, at more than 6,500, but for the past week the number has not gone above 5,000. In Spain, the number seems to have reached a
Coronavirus concerns delay treatment for cancer patient	2020	
Coronavirus death rate: What are the chances of dying?	2020	
Coronavirus in F1: Williams & Racing Point drivers take pay cut as staff put on leave	2020	
Coronavirus information: What should I do?	2020	

This is an example for academic purposes

> 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 8