

# Enterprise Knowledge Mining Solution Microsoft AI Platform

LearnAI Team  
October 2018



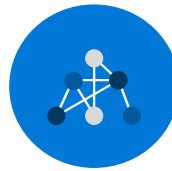
# Azure AI

## AI apps & agents



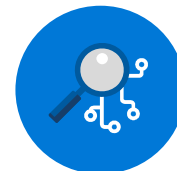
Azure Bot Service  
Azure Cognitive Services

## Machine learning



Azure Databricks  
Azure Machine Learning  
Azure AI infrastructure

## Knowledge mining



Azure Cognitive Search

AI uses **cognitive skills** to simulate human perception for:

- Vision
- Speech
- Language
- Knowledge

Customized language understanding **Text-to-speech**  
Content moderation **Spell**  
**Speech translation** check  
Custom image classification  
Speaker recognition Entity linking  
Sentiment analysis, & augmentation  
key phrase extraction **Image tagging**  
**Custom** Object detection Text translation  
**voice** **OCR handwriting** Intend  
Emotion detection analysis  
Video insights **recognition** Custom translation  
**Face** Custom speech Assisted text moderation  
**identification** Speech transcription

**Knowledge Mining** is the process of discovering actionable information from large sets of **unstructured data**, like text or images.

It uses **Artificial Intelligence** to detect hidden patterns and information. It can be used to guarantee compliance, enrich search, automate processes, among other **business processes**.





**Cognitive Search** is the Microsoft solution for Knowledge Mining.

**AI** approach for content understanding. Ingests unstructured content to create rich metadata into an **Azure Search** index, with the power of **Cognitive Services**.



# What Azure Search is

- AI-Powered cloud search service for web and mobile app development
- Enterprise Search as a Service – Azure PaaS product
- Enrich and extract insights through cognitive skills (Cognitive Search)
- Easily scale up and down
- Natural Language Processing for web search grade experience
- Connect search results to business goals with great control over search ranking
- 99.9% SLA, GDPR, Standard Azure OST (Online Service Terms)
- Creates indexed metadata about your data

# What Azure Search is

The screenshot displays the Azure Search Job Portal Demo interface. The header includes the Azure Search logo, the text 'AVAILABLE JOBS (180 jobs)', and navigation links for 'Home', 'Jobs', and 'About Azure Search'.

**Annotations and Features:**

- Spelling Mistakes:** A blue callout points to the search bar where 'analis' is entered, with a dropdown menu showing suggestions like 'DATA AND BUSINESS ANALYST' and 'Strategy & Analytics'.
- Geospatial:** A blue callout points to a map of New York City showing job locations marked with orange pins.
- Suggestions:** A blue callout points to the dropdown menu in the search bar.
- Ranking:** A blue callout points to a 'Relevance' dropdown menu.
- Paging:** A blue callout points to a pagination control showing page numbers 1 through 5.
- Hit Highlighting:** A blue callout points to the job listing for 'Budget Analyst, Family and Child Health Administration', where the title and location are highlighted in blue.
- Facets:** A blue callout points to the 'LOCATION' section on the left, which lists filters like 'Internal (92)' and 'External (88)'.

**Job Listings:**

- 2 Budget Analyst, Family and Child Health Administration**  
42-09 28th Street  
Salary: \$45,358 to \$61,754 Annual  
\*\*OPEN TO PERMANENT STAFF ANALYSTS ONLY. Under the direction of the Director of Budget for DFCH, duties of the Budget Analyst will include but not be limited to the following: [Read More](#)
- 3 PMO ANALYST - Featured Job**  
450 West 33 St, New York NY  
Salary: \$45,174 to \$62,370 Annual

# What Azure Search is

## Formats

- Microsoft Office
- PDF
- HTML
- XML, JSON
- ZIP
- EML, RTF, TXT
- JPEG, PNG, JPG
- CSV

## Data Sources

- Azure SQL
  - Text columns
  - JSON/XML columns
- SQL Server IaaS on Azure
- CosmosDB
- Azure blob storage
- Azure Table
- MySQL (PaaS and IaaS)
- Azure Files (Private Preview)

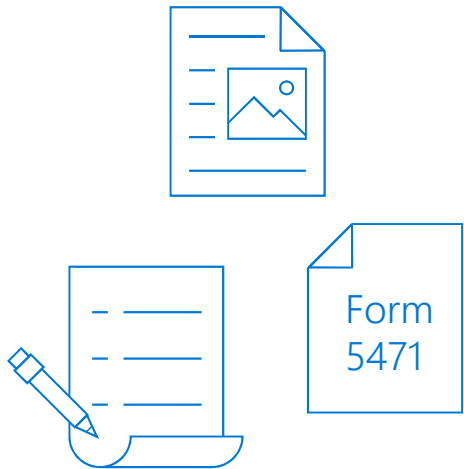


# What Cognitive Search is

- **Azure Search feature**, Announced in May 2018 (MS Build)
- It uses AI to create searchable metadata, transforming unstructured data into information
  - Data Enrichment != Data Integration
  - Original data isn't moved, changed or copied
  - Results are always loaded into an Azure Search Index
- Public Preview... **For now:**
  - 10+ regions, including South Central US, West Europe, North Europe, Brazil South, and Southeast Asia
  - Free, Azure Search regular cost only

# What Cognitive Search is

Enrichment Pipeline



Unlock valuable  
information lying latent  
in all your content

# What Cognitive Search is

Documents


Enrichment Pipeline


Fully text-searchable  
rich index



  
Key Phrase extraction

  
Organization entity extraction

  
Face detection

  
Custom skills

  
Location entity extraction

  
Persons entity extraction

  
Celebrity recognition

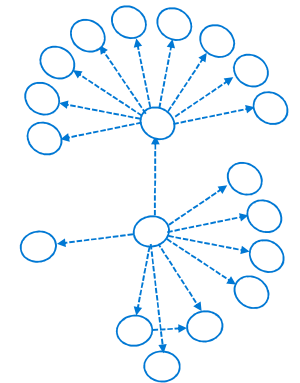
  
Landmark detection

  
Sentiment analysis

  
Language detection

  
Tag extraction

  
Printed text recognition



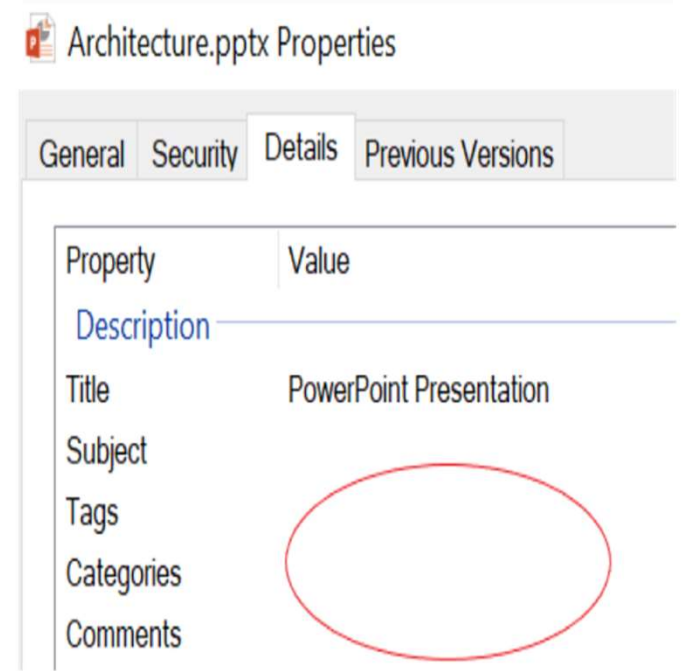
# When to use it



Documents  
Analysis



Understanding  
engineering plans



Lack of  
Metadata

# When to use It

Product Order Form & Distributor Application Kungen Water®

Enagic USA, Inc.  
Headquarters  
4115 Spencer St., Torrance, CA 90503  
Phone: (310) 542-7700 / FAX: (310) 542-1700  
Toll Free: (866) 261-9500 / [info@enagic.com](mailto:info@enagic.com)

**PRINT CLEARLY**

**Applicant Information**

Applicant Name (Last, First, Middle Initial) or Company Name: **Nguyen Van Le**

Enagic License Number: **C4574938** State: **CA** Date of Birth: **02-03-1960** Are you currently an Enagic Distributor? **No** Yes ☐ No ☒ Enagic ID#

Mailing Address (must match VISA):  
City: **San Jose** State: **CA** Zip Code: **95051**  
1234 Senter Road  
Phone Number: **(408) 765-4321**  
523-45-6789  
Fax Number: **(408) 345-6789** Email Address: **nguyentle36@gmail.com**

Shipping Address (if different from mailing address):  
City: State: Zip Code:  
Same above  
Alternate Shipping Address: City: State: Zip Code:

**Sponsor Information**

Sponsor Name: **Son Le**

Phone Number: **(408) 234-5678**

REGISTER THIS APPLICANT AS YOUR **[ 4 ] A**  
Under Sponsor ID Number: **7296044**

**ITEM ORDERED**

**SD501**

**PAYMENT METHOD**

☒ SINGLE PAYMENT ☐ ENAGIC PAYMENT: ☐ 3 months ☐ 6 months ☐ 10 months ☐ 16 months

Unit Price: **\$ 3,980.00** Tax: **348.26** Shipping: **23.00** Total: **\$ 4,351.26**

Product Retail Price: **\$ 3,980.00**

Handling: **\$** Tax: **\$** Shipping: **\$** Total Down: **\$**

Please note the first payment date must be within 45 days from purchase date. Payment date must be on the 1st or 15th of each month.

Finance Amount	Monthly Payment Amount	Withdrawal Date (Circle One)	First Payment Date
\$	\$	1st / 15th	/ /

**Payment Information - CREDIT CARD or CHECKING ACCOUNT** (void check needed for Checking Account Payment)

☒ VISA ☐ MASTER CARD ☐ AMEX ☐ DISCOVER

Credit Card Number / Checking Account Number: **2345-6789-1234-5678** Expiration Date / Checking Account Routing Number: **03/17** CVV # **376**

Card Holder Name (Please Print): **Nguyen Van Le** Card Holder Signature: *Nguyen Van Le*

Please fill out Alternate Payer Form if someone besides the applicant will be making down payment and/or monthly payment. \*\*\*

I certify that I have been furnished a copy of, and have read, understand, and agree to the provisions in Enagic USA, Inc.'s Policies and Procedures manual, which documents (with any amendments or restatements furnished by Enagic USA, Inc. after this date) are hereby incorporated by reference as if fully set forth herein and set forth the exclusive terms and conditions of my agreement with Enagic USA, Inc.

I hereby certify that the information provided on this form is complete and accurate to the best of my knowledge. I authorize ENAGIC USA, INC. to debit the amount I have indicated above from my bank account or credit card. This agreement will remain in effect until the balance is paid in full. \$20 late fee will be applied to your account every time payment is missed. By signing the line below, you are acknowledging that you have read and understood the terms and conditions. Terms and conditions are subject to change without notice. If you fail to make a monthly payment, Enagic may offset the payment amount from your commissions. FOR ALTERNATE PAYERS: By signing Alternate Payer Form, you will be jointly responsible for any and all balance owing on the account. This agreement is governed by the laws of California and proper venue will be in a court of competent jurisdiction located nearest to the Company's headquarters.

Participant Name: **Nguyen Van Le** Date: **03-17-15** Participant Signature: *Nguyen Van Le*

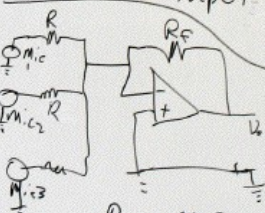
Print Sponsor Name: **Son Le** Date: **03-17-15** Sponsor Signature: *Son Le*

Change Your Water...  
Change Your Life™

Revised 10/2012

"Golden Rules" for OP AMPS.

1. No current actually flows into the - and + input terminals.  
(Very large input Resistance).
2. The feedback makes the voltage at the input terminals the same.



$$V_0 = (V_1 + V_2 + V_3) \frac{R_F}{R}$$

$$R_F = 1M\Omega$$

$$R = 10k\Omega$$

$$\frac{R_F}{R} = \text{Gain}$$



Forms  
Reading

Handwritten  
Information

Image  
Analysis

# When to use It

Product Order Form & Distributor Application Kungen Water®

Enagic USA, Inc.  
Headquarters  
4115 Spencer St., Temecula, CA 92593  
Phone: (951) 942-7700 / FAX: (951) 942-1700  
Toll Free: (866) 261-9900 / [info@enagic.com](mailto:info@enagic.com)

**PRINT CLEARLY**

**Applicant Information**  
Applicant Name (if not listed, Retail User) or Company Name: Nguyen Van Le  
City: San Jose State: CA Zip: 95128  
Address: 1234 Senter Road  
City: San Jose State: CA Zip: 95128  
Phone Number: (408) 345-6789  
Fax Number: (408) 345-6789  
Email Address: nguyenle36@gmail.com

**Sponsor Information**  
Sponsor Name: Son Le  
Phone Number: (408) 234-5678

**ITEM ORDERED**  
SD501 \$ 3,980.00  
Product Retail Price: \$ 3,980.00  
Unit Price: \$

**Payment Information - CREDIT CARD or CHECKING ACCOUNT**  
Credit Card Number / Checking Account Number: 2345-6789-1234-5678  
Cardholder Name: Nguyen Van Le

**Participant Information**  
Participant Name: Nguyen Van Le  
Signature: Nguyen Van Le Date: 03-17-15

Dataset of this training!!



Form Reading

Handwritten Information

Image Analysis



# When to use It

## How to identify Cognitive Search Opportunities

- Every company has ppts, pdfs, docx, html, images...
- Every company has contracts, forms, plans, memos, emails
- Every company needs compliance, risk detection
- Client needs a better search experience on top of business documents
- Global companies with enterprise documents in multiple languages
- Local file server is out of space
- Client already has data on Azure
- Client needs to apply AI to business documents: data != information

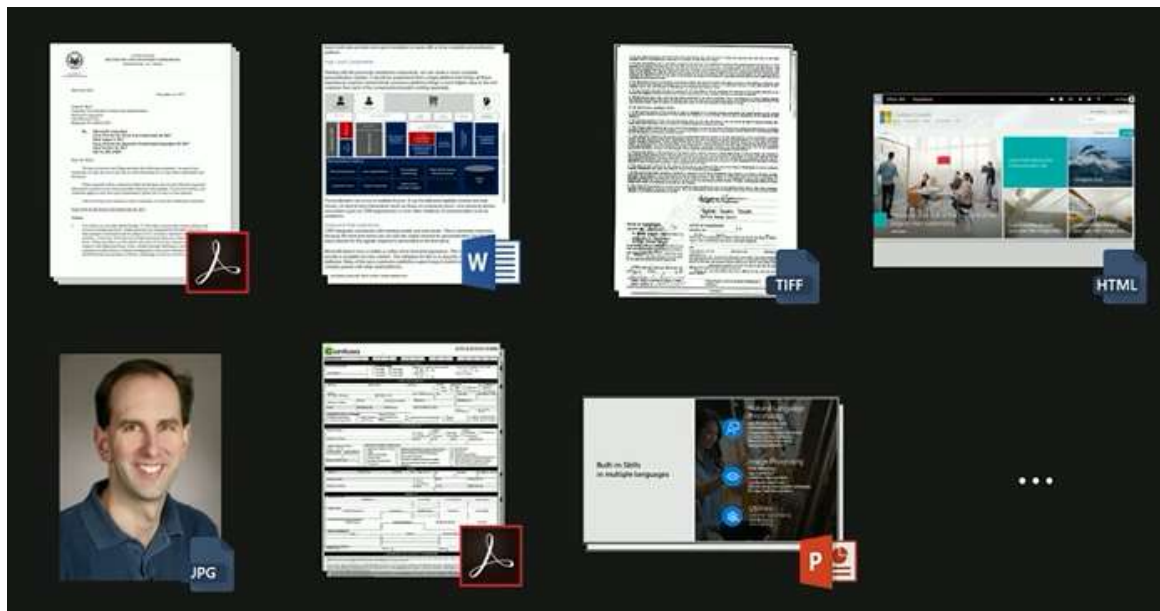
# When to use It

How to identify **CREATE** Cognitive Search Opportunities

- **Every company** has ppts, pdfs, docx, html, images...
- **Every company** has contracts, forms, plans, memos, emails
- **Every company** needs compliance, risk detection
- Client needs a better search experience on top of business documents
- Global companies with enterprise documents in multiple languages
- Local file server is out of space
- Client already has data on Azure
- Client needs to apply AI to business documents: data != information

# When to use It

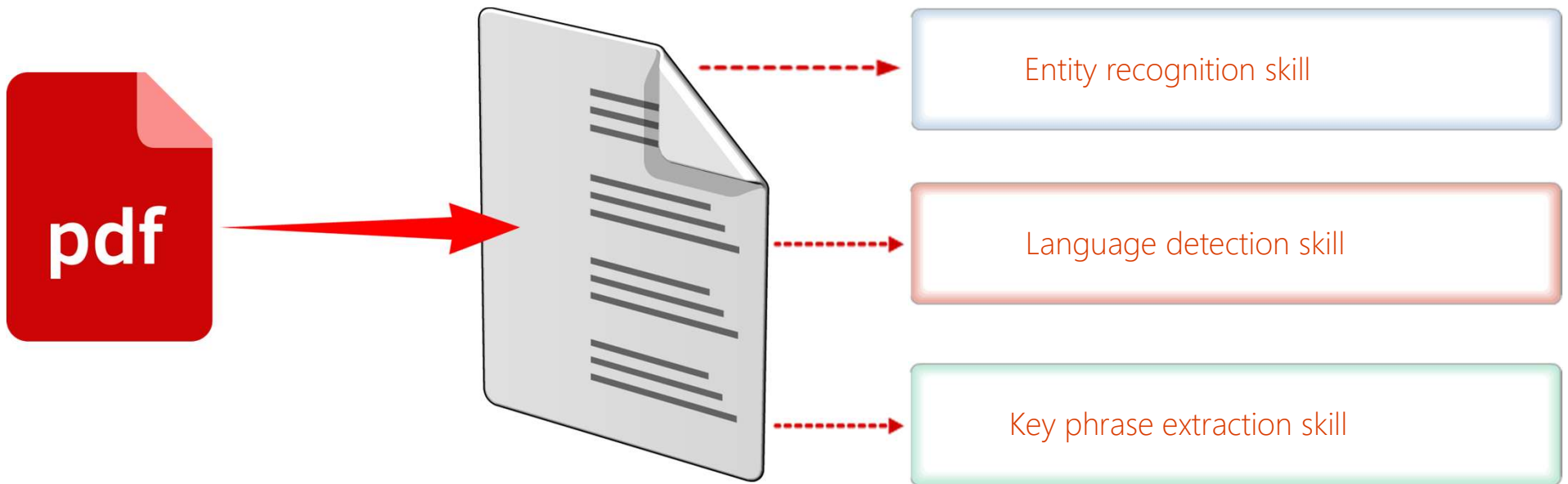
How to identify **CREATE** Cognitive Search Opportunities



80% of relevant  
business  
information is  
unstructured,  
usually text.

# How it works

Cognitive Search **enrichment pipeline**: Atomic cognitive processes, aka **Cognitive Skills**, applied for each document, creating metadata



# How it works

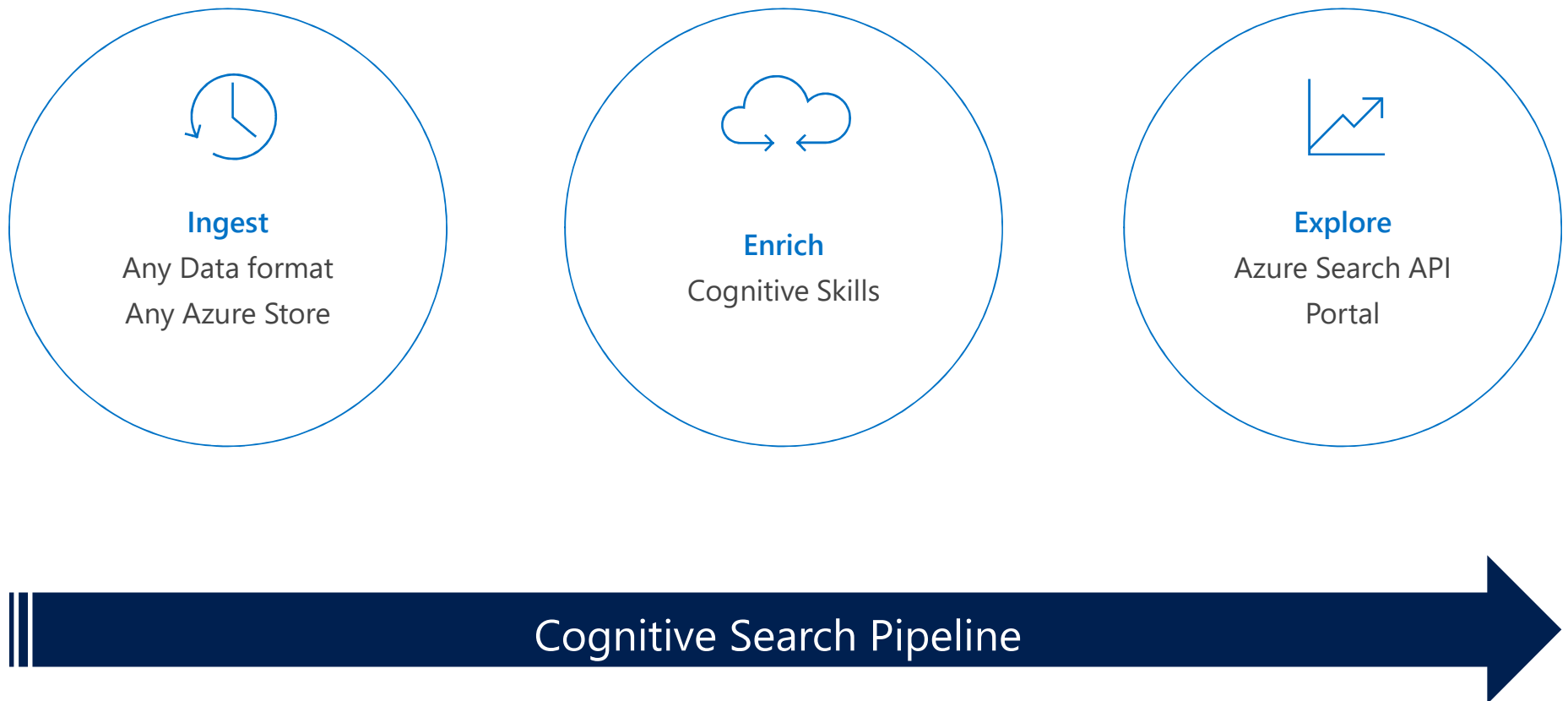
## Predefined Cognitive Skills

- Key Phrases
- Language Detection
- Text Merger
- Entity Recognition: Names, Locations, Organizations
- Sentiment Analysis
- Text Splitter
- Image Analysis: categories, tags, description, faces, type, colors, adult content
- OCR
- Shaper: Complex types (matrix) – Private Preview

## Custom Skills

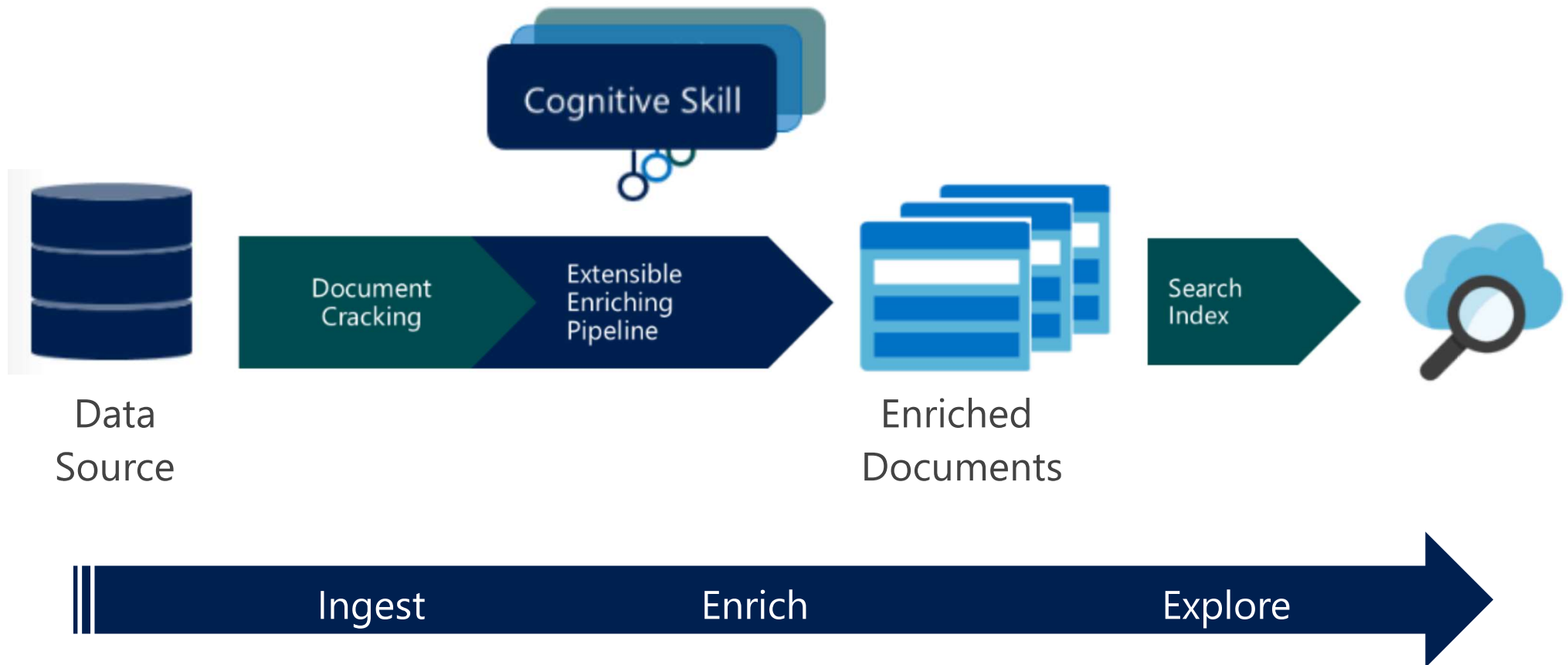
- Any REST API
  - Public
  - Client Specific
  - Microsoft Cognitive Services
- Create your own API
  - Industry specific
  - Hosted on Azure
    - Azure Functions
    - Web APP

# How it works

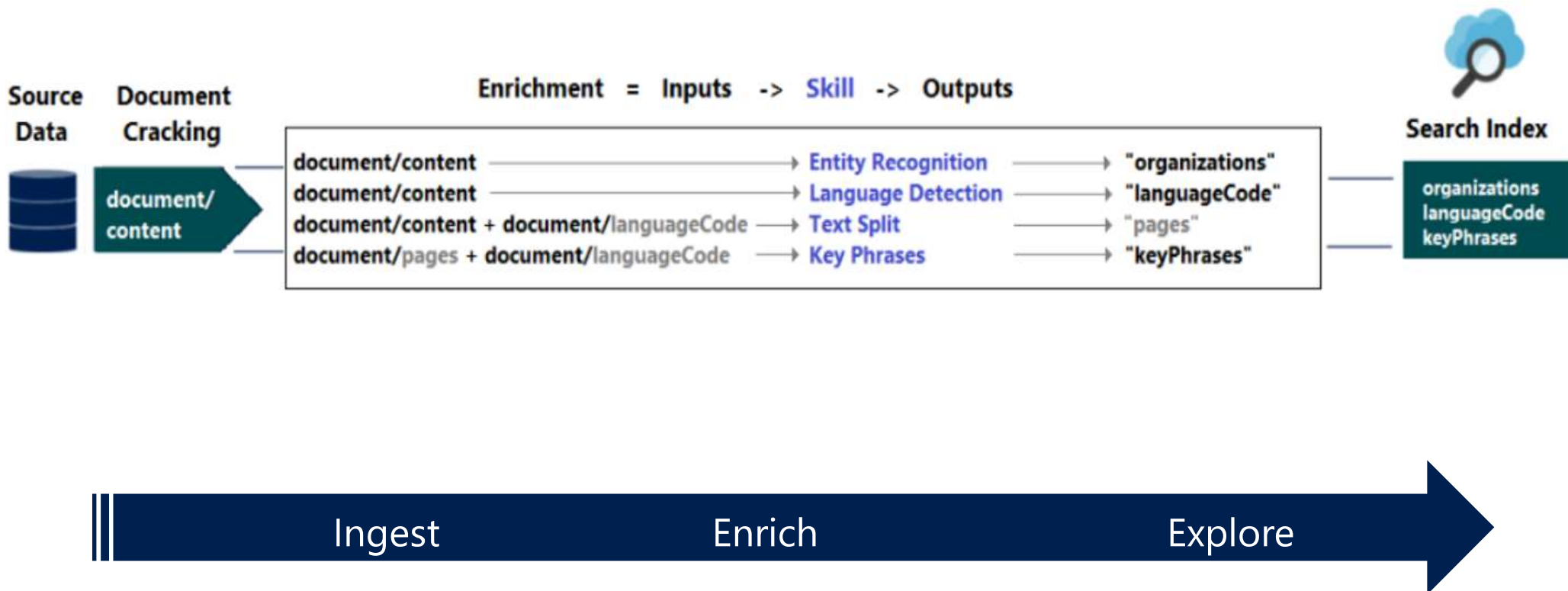




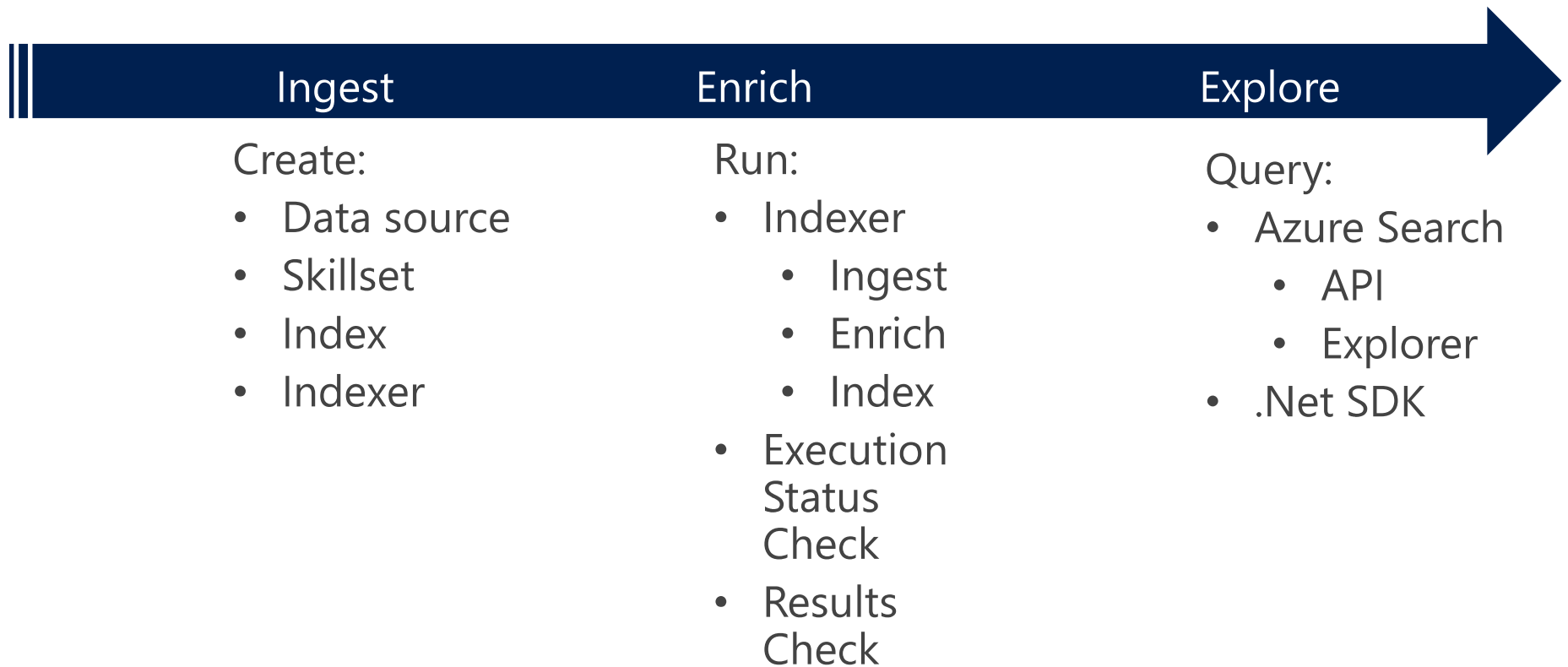
# How it works – Cognitive Search Pipeline



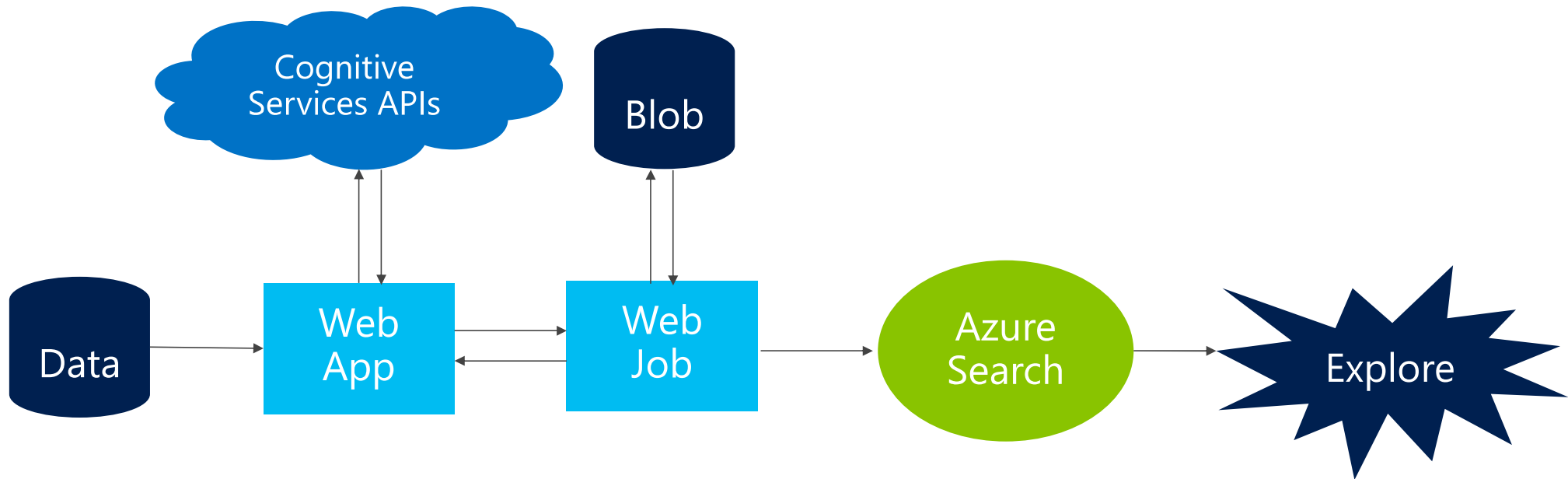
# How it works – Cognitive Search Pipeline



# How it works – Cognitive Skills



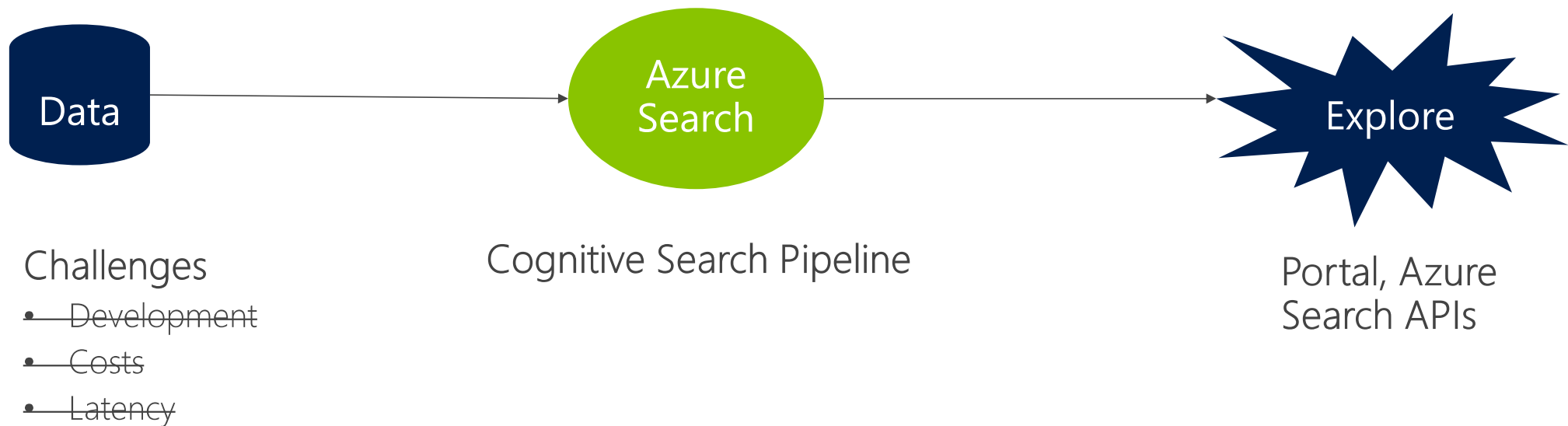
# How it works – Old High Level Architecture



## Challenges

- Development
- Costs
- Latency

# How it works – New High Level Architecture



# Good Practices

Supplement the indexes with:


- Natural Language Processing
- Custom Analyzers (50+ Languages)
- Fuzzy Search
- Filtered Search
- Scoring Profiles
- Autocomplete (new – June 2018)
- Synonymous (new – June 2018)

The screenshot displays the Azure Search Job Portal Demo interface. The header includes the Azure Search logo, 'AVAILABLE JOBS (180 jobs)', and navigation links for 'Home', 'Jobs', and 'About Azure Search'. The main content area is divided into a search sidebar and a results section. The sidebar features a search bar with the text 'analys' and a dropdown menu showing suggestions like 'DATA AND BUSINESS ANALYST Strategy & Analytics'. Below this, a list of job titles is shown, including 'Procurement Analyst (6)', 'Change Order Analyst (4)', and 'Claiming Analyst, Bureau of Budget and Revenue (4)'. The results section on the right shows a map of New York City with red pins indicating job locations. Below the map, a list of jobs is displayed, including 'Budget Analyst, Family and Child Health Administration' and 'PMO ANALYST - Featured Job'. The interface is annotated with several blue callout boxes: 'Spelling Mistakes' points to the search bar; 'Geospatial' points to the map; 'Suggestions' points to the dropdown menu; 'Ranking' points to the 'Relevance' dropdown; 'Paging' points to the pagination controls; 'Facets' points to the 'LOCATION' section; and 'Hit Highlighting' points to the job description text.



# Good Practices

- Results are **always** loaded into an Azure Search Index. But you can do extra Analytics: Create a Custom Skill to save the enriched metadata:
  - CosmosDB
  - Azure SQL json column
  - Blob Storage
  - Others



Roadmap: Automatic  
export of the enriched  
metadata

# Good Practices

- Clear Business requirements
- SQL Server on-prem is supported but not recommended
  - Performance
  - SQL Server open to the internet
  - Use ADF to copy data to Azure
- Co-locate the Azure Search Service with your data: Costs, Latency

# Good Practices

- Leverage Azure Free Account and Azure Search Free Tier
  - 50 MB
  - 5 Cognitive Skills, 3 Indexes
  - Keep it up and running – demo in minutes with clients own data
- Security Trimming to filter content based on user identity
- Multiple Indexes for the same dataset may be required
  - Ranking, languages, filters...

# Good Practices

- Incremental Updates for new files (portal or command line)
  - On Demand, File Watch, Schedule
  - Execution Overlap: The schedule essentially tells the indexer to proceed where it left off at the next hour
- Image Skills Performance
  - Deep Learning behind the sciences
  - While in public preview, performance will be better if you break documents with multiple images on multiple documents. When GA this will be done automatically

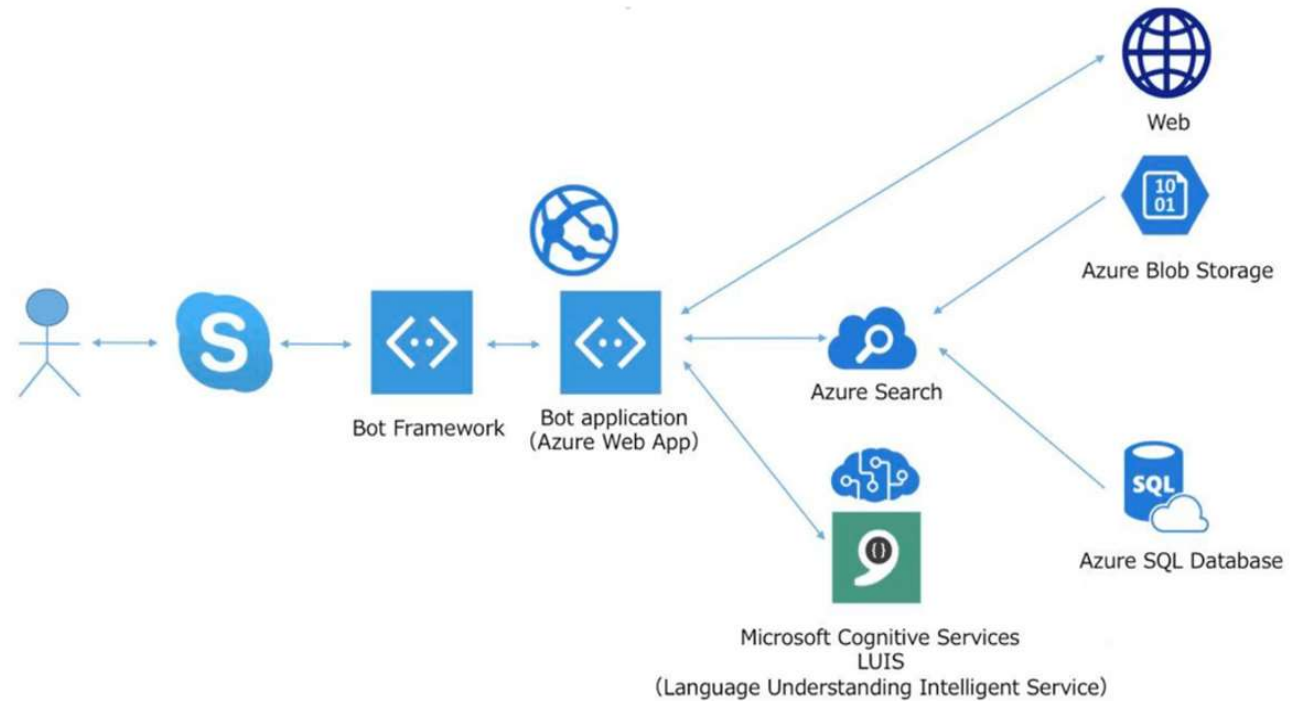
# Good Practices

- Troubleshooting
  - Start with a small dataset
  - Make Sure your credentials are correct
  - See what works even if there are some failures
    - Check Status
    - Error Parameters
  - Enriched documents under the hood: saving the raw transformations into the index
  - Free and basic tears have limits for document sizes
  - Parallel Indexing
  - Maximum runtime

# Good Practices

## End to End Solutions

- .NET SDK
- REST API
- BOT + LUIS





# Demos – Cognitive Search

Enrichment Pipeline

<https://text-analytics-demo-dev.azurewebsites.net/>

JFK Files

<https://jfk-demo.azurewebsites.net/>

Business Documents Demo - Wolters Kluwer

<https://wolterskluwereap.azurewebsites.net/>

Oil & Gas Demo - Exxon Mobile

<http://seismicsearch.azurewebsites.net/>

Healthcare - CTakes

<http://webmedsearch.azurewebsites.net/>

## Additional Resources

<http://aka.ms/LearnAI-trainings>

<http://aka.ms/LearnAI-csw>

<http://aka.ms/LearnAI-kmb>

<http://aka.ms/LearnAI-links>

# LearnAI – Knowledge Mining Bootcamp

