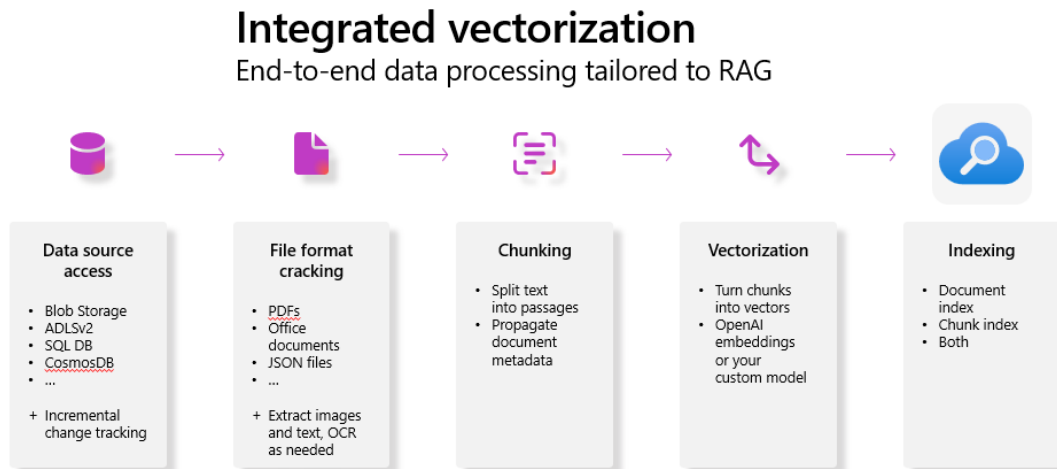# Azure AI Search - Create a Hybrid Search Solution (including vector search)

In this lab, you will create a hybrid search solution using Azure AI Search integrated vectorization solution. You can do this task using Azure AI Studio or Azure Portal. We are going to do it using Azure AI Studio.

Integrated vectorization
End-to-end data processing tailored to RAG

Data source access	File format cracking	Chunking	Vectorization	Indexing
• Blob Storage • ADLSv2 • SQL DB • CosmosDB • … + Incremental change tracking	• PDFs • Office documents • JSON files • … + Extract images and text, OCR as needed	• Split text into passages • Propagate document metadata	• Turn chunks into vectors • OpenAI embeddings or your custom model	• Document index • Chunk index • Both

We will use sample data (uploaded to Azure storage account) in this step. We will also use Azure OpenAI embedding model to turn document chunks into vectors.

Therefore, this step will require Azure AI Studio to connect to your Azure Storage account (to get pdf documents) and Azure OpenAI Service (for embedding model)

1. Prerequisites:

    a. Azure OpenAI service with embedding model deployed.
    b. Azure storage account
    c. Azure AI Studio with AI Hub created.

2. Upload sample data (PDF documents) to Azure storage container (using Azure portal).

    a. <u>Create a container within the Azure storage account</u> you created before. To do this
        i. Go to the Azure storage account (in Azure portal)
        ii. Click on Containers
        iii. Click on the "+ Container" to create a new container.
        iv. Give the container a name (Ex: ragdatacontainer) and click Create.
    b. Upload sample data within this container. To do this:
        i. Get all PDF documents from the GitHub repository. Sample data available on GitHub at https://github.com/amulchapla/GenAI-LLMapps-lifecycle-workshop/tree/main/AzureAISearch-VectorSearch
        ii. Go to the Container you created and click Upload.
        iii. VERY IMPORTANT: Enter "Upload to folder" name as "hrbenefitsdata" (as shown below)
        iv. Select all PDF documents to upload and click "Upload" button.

Upload blob   ✕

6 file(s) selected: Benefit_Options.pdf, employee_handbook.pdf, Nort...

Drag and drop files here or Browse for files

☐ Overwrite if files already exist

⌄ Advanced

Blob type ⓘ

| Block blob | ⌄ |

☑ Upload .vhd files as page blobs (recommended)

Block size ⓘ

| 4 MiB | ⌄ |

Access tier ⓘ

| Hot (Inferred) | ⌄ |

Upload to folder

| hrbenefitsdata |

Blob index tags ⓘ

3. Add your data (in Azure AI Studio)

In this step, we will bring our own data so you can chat with your own data using the GPT models.

  a. Go to Azure AI Studio. You can login to AI Studio at https://ai.azure.com/ OR go to your Azure OpenAI service overview page and click on "Build in Azure AI Studio" button.
  b. Once in Azure AI Studio, click on the Project.
  c. Click on Playground (on the left hand side)
  d. Go to "Add your data" tab
  e. Click on "Add your data" and click on "Add connection" button. Here you are going to connect AI Studio to your Azure Storage account.
  f. Enter information related to your storage account (see screenshot below). You will need to enter Account Key for your Azure storage account which you can get from Azure Portal.

g. Select Azure AI Search service

h. Configure Azure OpenAI service connection for Embedding model. Embedding model is used to create embedding of your documents in this step.

i. Enter index name and other settings (you can leave them default)

j. Review information and click Create.
k. It will take a few minutes to complete the process (shown below). It will update the status as shown below.
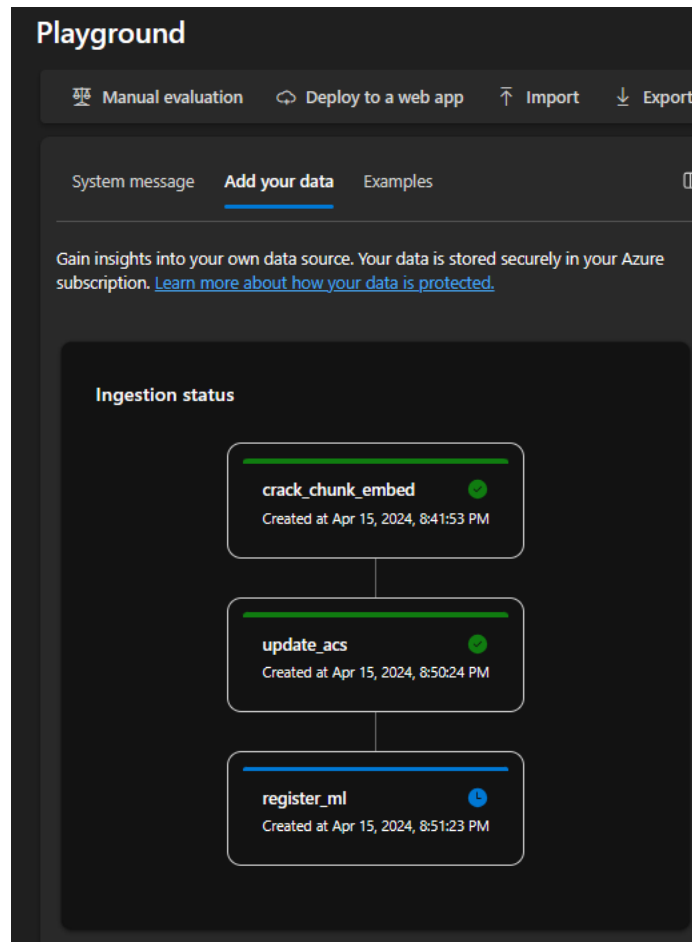
l. You can now chat with your data using the GPT model.
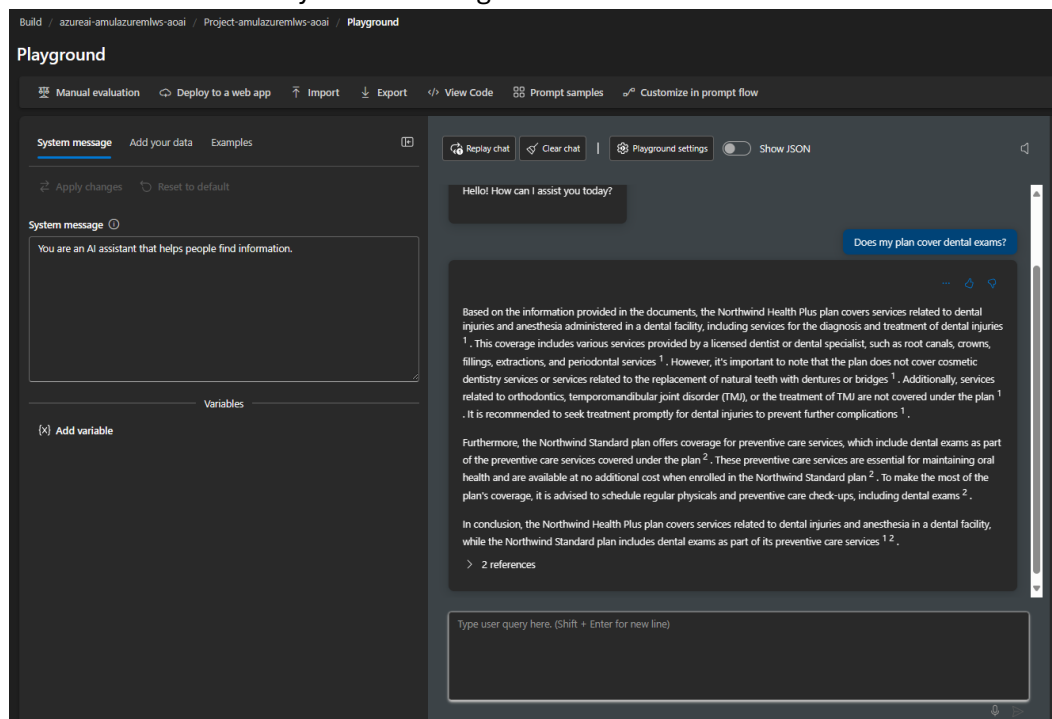
4. [OPTIONAL] Create Azure AI Search service and create vector store

a. Login to Azure Portal
b. Create a new Azure AI Search service (if one doesn't exist already)
c. Go to Azure AI Search service and click "Import and vectorize data". Follow the data import wizard.

      i. **Set up your data connection**: Connect to an external Azure Blob storage containing PDFs and other unstructured data files. You can leave Blob folder empty.

      ii. **Vectorize your data with Azure OpenAI**: Connect to an Azure OpenAI service and select an embedding model for vector generation.
You need Azure OpenAI service with embedding model deployed for this step. If you don't have this then complete that task first.

      iii. **Update the search index name (optional):** The wizard generates an index, indexer, data source, and skillset on your search service. You can view and independently manage these objects once they exist, but the names and many other properties are fixed for the lifetime of the object. To customize the name, change the object name prefix.

Import and vectorize data ...

gptkb-thxdj6b6uiqtk

✓ Connect to your data	The wizard generates an index, indexer, data source, and skillset on your search service. You can view and independently manage these objects once they exist, but the names and many other properties are fixed for the lifetime of the object. To customize the name, change the object name prefix.
	Objects name prefix
✓ Vectorize and enrich data	vector-hrbenefits
● Review and create	**Review**
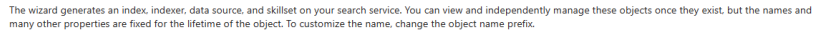
Attached Azure OpenAI service	aoaiamuleastus2
Deployment model	text-embedding-ada-002
Extracting text from images	Disabled
Semantic ranker	Enabled
Indexer run schedule	Once

iv. **Wait for search index setup to complete.**

gptkb-thxdj6b6uiqtk | Indexes ☆ ···
Search service

+ Add index ∨   ↻ Refresh   🗑 Delete

🔍 Search

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

vector

Name	Document count	Vector index size	Total storage size
vector-hrbenefits	410	2.43 MB	12.14 MB

d. **Test Azure AI Search**: Go to the Azure AI Search service and click on Indexes and select the search index you created above.

    i. Enter a search term related to HR benefits and make sure you get search results.

vector-hrbenefits ···

💾 Save   ✕ Discard   ↻ Refresh   ✂ Create Demo App   {} Edit JSON   🗑 Delete

Documents ⓘ	Total storage ⓘ	Vector index size ⓘ
410	12.14 MB	2.43 MB

Search explorer   Fields   CORS   Scoring profiles   Semantic configurations   Vector profiles

⚙ Query options   ◎ View ∨

🔍 what is performance review   ✕   **Search**

Results

```
 1  {
 2    "@odata.context": "https://gptkb-thxdj6b6uiqtk.search.windows.net/indexes('vector-hrbenefits')/$metadata#do
 3    "@search.answers": [
 4      {
 5        "key": "1227e000194b_aHR0cHM6Ly9zdGF6dXJlYWlhbhXU4NjExODk2NDI5ODDAuYmxvYi5jb3J1LndpbmRvd3MubumV0L3NvdXJjZ\
 6        "text": "Performance reviews are a two-way dialogue between managers and employees. We   encourage all
 7        "highlights": "Performance reviews are<em> a two-way dialogue between managers and employees.</em> We
 8        "score": 0.99560546875
 9      }
10    ],
11    "@search.nextPageParameters": {
12      "search": "what is performance review",
13      "queryType": "semantic",
14      "semanticConfiguration": "vector-hrbenefits-semantic-configuration",
15      "captions": "extractive",
16      "answers": "extractive|count-3",
17      "queryLanguage": "en-US",
18      "skip": 50,
19      "vectorQueries": [
20        {
21          "kind": "text",
22          "k": 5,
23          "oversampling": null,
24          "fields": "vector",
25          "vector": [],
```