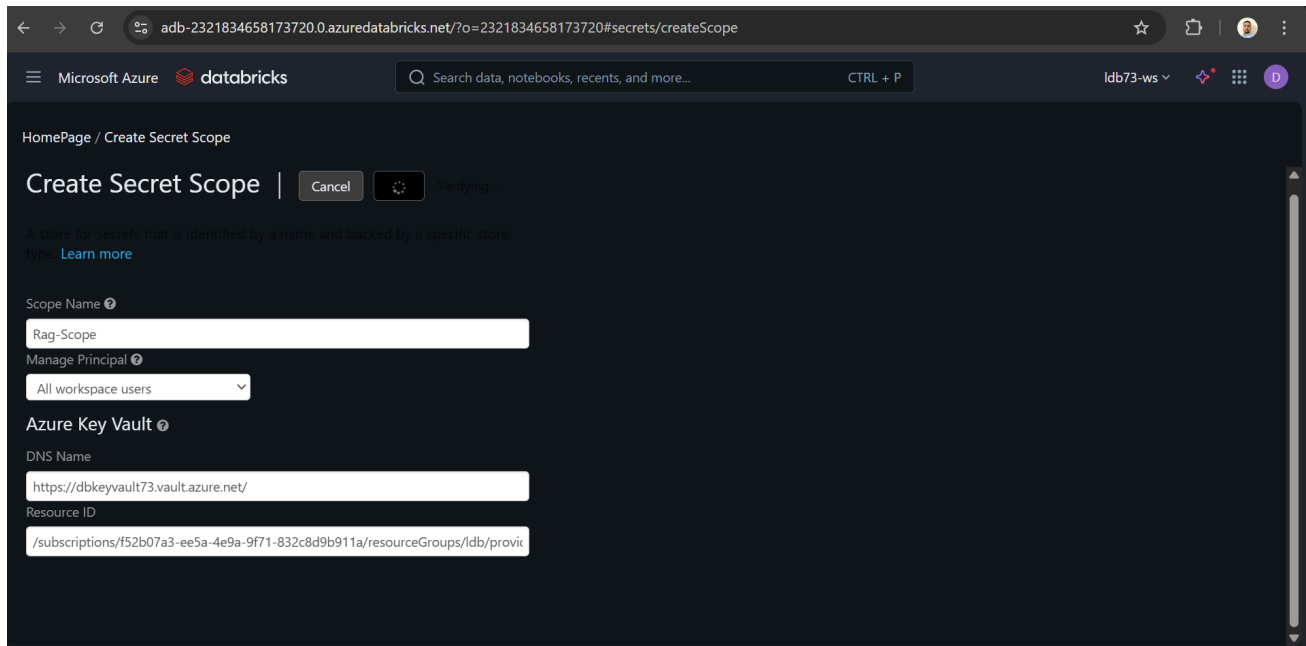


P4 - Raw

After running 1st and 2nd notebook you do the steps below:

1. login to databricks -> settings -> developer -> access token -> copy it
2. go to azvault-> new secret -> create it

<https://adb-2321834658173720.0.azuredatabricks.net/#secrets/createScope>



3. Assign the Azure Key Vault user role to the **OID** (error occurs when we try to access the secret in notebook 3 -2nd cell). copy the **OID** from the ERROR.

4. Creating the Vector DB on Azure databricks

Databricks → Compute → Vector Search → Create your endpoint

endpoint_name = doc_vector_endpoint

index → select the table which contains the data which is docs_text

llm.rag.docs_idx → refer to the screenshots below to understand better.

Microsoft Azure databricks

Search data, notebooks, recents, and more... CTRL + P

ldb73-ws

Compute

All-purpose compute Job compute SQL warehouses **Vector Search** Pools Policies Apps Lakebase Postgres

Filter vector search endpoints by name Owned by me Empty endpoints Type

Status	Name	Type	Indexes	Creator
Provisioning	doc_vector_endpoint	Standard	0	admin@nikhil0073outlook.onmicrosoft.com

Just Provide the endpointname and craete it.

Microsoft Azure databricks

Search data, notebooks, recents, and more... CTRL + P

ldb73-ws

Catalog

Serverless Starter Wareho... Serverless S

Type to search...

My organization

- ldb73_ws
- system
- llm2
 - default
 - information_schema
 - rag
 - Tables(2)
 - docs_text**
 - docs_track
 - Volumes (1)
 - Models (1)
 - old-stuff
 - Delta Shares Received
 - samples
 - Legacy
 - hive_metastore

Catalog Explorer > llm2 > rag > docs_text

Open in a dashboard Share Create

Overview Sample Data Details Permissions Policies History Lineage Insights Quality

Description

AI generate Add

Filter columns...

Column	Type	Comment	Tags	Column maskir
id	bigint			
text	string			

About this table

Owner Databricks Admin

Type Managed

Open in a dashboard Share Create

Query Notebook Dashboard Genie space **Vector search index** Synced table Online table Metric view

Create vector search index

Index structure

Name docs_idx

Primary key id

Columns to sync text

Primary key and embedding column will automatically be synced to the index, even if they are not selected.

Embeddings

Embedding source Compute embeddings Use existing embedding column

Embedding source column text

Embedding model databricks-gte-large-en

Cancel Create

Create vector search index

Embedding source Compute embeddings Use existing embedding column

Embedding source column text

Embedding model databricks-gte-large-en

Sync computed embeddings

Compute resources

Vector search endpoint doc_vector_endpoint

Sync mode Triggered Continuous

Advanced settings

Cancel Create

Keep this exactly docs_idx

change the embedding model in the code also

The screenshot shows the Databricks workspace interface. In the left sidebar, the 'Catalog' view is open, and 'docs_idx' is highlighted under the 'Tables (3)' section. The main notebook area displays two code cells. The first cell, titled '2: Set needed parameters', contains the following Python code:

```
host = "https://" + spark.conf.get("spark.databricks.workspaceUrl")
os.environ["DATABRICKS_TOKEN"] = dbutils.secrets.get(scope="P4-Scope", key="RAG06")
index_name = "llm2_rag_docs_idx"
host = "https://" + spark.conf.get("spark.databricks.workspaceUrl")
VECTOR_SEARCH_ENDPOINT_NAME = "doc_vector_endpoint"
```

The second cell, titled '4: Build Retriever', contains the following Python code:

```
from databricks.vector_search.client import VectorSearchClient
from langchain_community.vectorstores import DatabricksVectorSearch
from langchain_community.embeddings import DatabricksEmbeddings

embedding_model = DatabricksEmbeddings(endpoint="databricks-gte-large-en")

def get_retriever(persist_dir: str = None):
    os.environ["DATABRICKS_HOST"] = host
    # Get the vector search index
```

Go to the Notebook 3 and run all the cells one-by-one.