Contents

[Steps to create a New KnowledgeBase (Text Based) (in DEV) 2](#_Toc178257070)

[Example - creating milvus collection for new BU 2](#_Toc178257071)

[Steps to create a New Image KnowledgeBase (in DEV) 4](#_Toc178257072)

[Example - Creating Image milvus collection for new BU 4](#_Toc178257073)

[Steps to create a New Text and Image KnowledgeBase (in PROD) – Vector DB in PROD. 5](#_Toc178257074)

[Example – After creating both knowledgebases (milvus collections) in dev we need to create again in PROD 5](#_Toc178257075)

[Steps to create an ADD document from Text and Image KnowledgeBase – 6](#_Toc178257076)

[Steps to create DELETE document from Text and Image KnowledgeBase (DEV & PROD) – 8](#_Toc178257077)

# Steps to create a New KnowledgeBase (Text Based) (in DEV)

### Example - creating milvus collection for new BU

**Prerequisites Process -**

1. Create the folder with given BU Name in the KNOWLEDGEBASE folder in **S3 bucket** and upload all pdf docs inside BU Name folder.
2. Create BU\_NAME\_Metadata.json file and upload it in KnowledgeBase\_Metadata folder

A screenshot of a computer

Description automatically generated

1. BU\_NAME\_Metadata.json contains author and title information for each pdf in the knowledge base. Below is the sample -

A computer screen with text

Description automatically generated

1. Open WinSCP, Go to the Genai/knowledge\_base\_temp folder.

A screenshot of a computer

Description automatically generated

1. Open **ParseKnowledgeBase.py** file
2. Change the Name of KnowledgeBase (BU Name) for KnowledgeBaseName variable in the code

A black text on a white background

Description automatically generated

1. Save it, then open **milvus\_standalone.py** file
2. Change the Name of KnowledgeBase (BU Name) for KnowledgeBaseName variable in the code, Same as above
3. Save it and open **check\_collection\_info.py**
4. Change the Name of Collection(BU Name) for colletion variable in the code, Same as above and save it.

**Note –** Keep KnolwedgeBaseName variable value same as BU Name for example for Tepezza BU we keep collection name and KnowledgeBaseName as “TEPEZZA”

**Running Codes (python scripts) inside VM ->**

1. Open DEV VM terminal via putty/ec2 instance.
2. Go to Genai folder –



1. Activate genai\_env by using following command –

source genai\_env/env/bin/activate

1. Open knowledge\_base\_temp folder and run **ParseKnowledgeBase.py** script using command – python ParseKnowledgeBase.py
2. This will parse documents using textract and create chunks of these documents and will save them in s3 bucket
3. After above step is complete and run **milvus\_standalone.py** script using command – python milvus\_standalone.py.
4. This would read all pkl files and create collection in milvus.
5. After step 17 is done, wait for 5-10 minutes and then run **check\_collection\_info.py** script. Given below is a sample output, it should show the information about created collection.

A black screen with white text

Description automatically generated

# Steps to create a New Image KnowledgeBase (in DEV)

### Example - Creating Image milvus collection for new BU

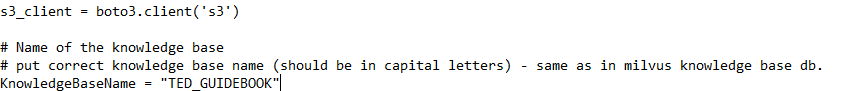
**Prerequisites -**

1. Ensure that text knowledgebase is already created for the BU.
2. For Image BU it would utilize data in KNOWLEDGEBASE and KnowledgeBase\_Extracts folder (which will be automatically created while creating KnowledgeBase.
3. Open WinSCP, Go to the Genai/Image\_KB\_Temp folder.

A screenshot of a computer

Description automatically generated

1. Open **Parse\_Images.py** file
2. Change the Name of KnowledgeBase (BU Name) for KnowledgeBaseName variable in the code



1. Save it, then open **Create\_Milvus\_DB.py** file
2. Change the Name of KnowledgeBase (BU Name) for KnowledgeBaseName variable in the code, Same as above
3. Save it and open Check\_Collection\_Info.py
4. Change the Name of Collection(BU Name) for colletion variable in the code, Same as above and save it.

**Note –** Keep KnolwedgeBaseName variable value same as BU Name for example for Tepezza BU we keep KnowledgeBaseName as “TEPEZZA” and Collection name would be made as “TEPEZZA\_IMAGES)

**Running Codes (python scripts) inside VM ->**

1. Open DEV VM terminal via putty/ec2 instance.
2. Go to Genai folder –



1. Activate genai\_env by using following command –

source genai\_env/env/bin/activate

1. Open Image\_KB\_Temp folder and run **Parse\_Images.py** script using command – python Parse\_Image.py.
2. This will parse documents using GPT 4o, crop images, create chunks of these images and will save them in s3 bucket.
3. Run **Create\_Milvus\_Image\_DB.py** script using command – python Create\_Milvus\_Image\_DB.py.
4. This would read all pkl files and create collection in milvus.
5. After step 16 is done, wait for 5-10 minutes and then run **Check\_Collection\_Info.py** script. Given below is a sample output, it should show the information about created collection.

# Steps to create a New Text and Image KnowledgeBase (in PROD) – Vector DB in PROD.

### Example – After creating both knowledgebases (milvus collections) in dev we need to create again in PROD

**Prerequisites -**

1. Copy the folders created or made while creating vector db in DEV. Below are the list of folders that should be copied or updated with new BU data in PROD –
   1. KNOWLEDGEBASE (pdf files)
   2. KnowledgeBase\_Metadata
   3. KnowledgeBase\_Extracts
   4. KnowledgeBase\_Images
   5. KnowledgeBase\_Images\_Archive
2. Open WinSCP, Go to the genai/knowledge\_base\_temp folder.
3. Change KnowledgeBaseName and collection variable values to BU\_NAME in **milvus\_standalone.py** and **check\_collection\_info.py**.
4. Go to the Genai/Image\_KB\_Temp folder.
5. Change KnowledgeBaseName and collection variable values to BU\_NAME and BU\_NAME\_IMAGES respectively in **Create\_Milvus\_Image\_DB.py** and **Check\_Collection\_Info.py**.

**Running Codes (python scripts) inside VM ->**

1. Open PROD VM terminal via putty/ec2 instance.
2. Go to genai folder –



1. Activate env by using following command –

source env/bin/activate

1. Open knowledge\_base\_temp folder and run **milvus\_standalone.py** script using command – python milvus\_standalone.py.
2. This would read all pkl files and create collection in milvus.
3. After step 10 is done, wait for 5-10 minutes and then run **check\_collection\_info.py** script to confirm if collection is created successfully.
4. Open Image\_KB\_Temp folder and run **Create\_Milvus\_Image\_DB.py** script using command – python Create\_Milvus\_Image\_DB.py.
5. This would read all pkl files and create collection in milvus.
6. After step 14 is done, wait for 5-10 minutes and then run **Check\_Collection\_Info.py** script.

# Steps to create an ADD document from Text and Image KnowledgeBase –

**Prerequisites Process -**

1. Add new Document in the KNOLWELDGEBASE/BU\_NAME folder.
2. Add document metadata in BU\_NAME\_Metadata.json file present in KnowledgeBase\_Metadata folder
3. Open WinSCP, Go to **Genai/knowledge\_base\_temp** folder
   1. Change KnowledgeBaseName in **ParseKnolwedgeBase.py, milvus\_standalone.py**
   2. Change collection name in **check\_collection\_info.py and remove\_collection.py**
   3. Add the document name (example Test\_Doc.pdf) in the list in **ParseKnolwedgeBase.py**

**A white background with black text

Description automatically generated**

1. Similarly open Genai/Image\_KB\_Temp folder
   1. Change KnowledgeBaseName in **Parse\_Imags.py, Create\_Milvus\_Image\_DB.py**
   2. Change collection name in **Check\_Collection\_Info.py and Remove\_Collection.py**
   3. Add the document name (example Test\_Doc.pdf) in the list in **Parse\_Images.py**

A black text on a white background

Description automatically generated

**Running Codes (python scripts) inside VM ->**

1. Open DEV VM terminal via putty/ec2 instance.
2. Go to Genai folder –



1. Activate genai\_env by using following command –

source genai\_env/env/bin/activate

1. In **knowledge\_base\_temp** run –
   1. ParseKnowledgeBase.py (will create pkl file in s3)
   2. remove\_collection.py (wait 2 minutes after success)
   3. milvus\_standalone.py (creates db again)
   4. check\_collection\_info.py (run this after waiting 5 minutes)
2. In **Image\_KB\_Temp** run –
   1. Parse\_Images.py (will create pkl file in s3)
   2. Remove\_Milvus\_Collection.py (wait 2 minutes after success)
   3. Create\_Milvus\_Image\_DB.py (creates db again)
   4. Check\_Collection\_Info.py (run this after waiting 5 minutes)

**To ADD Document in PROD Databases ->**

1. We can repeat same steps as above from 1-9 or do following steps
2. Copy Paste the newly created folders from DEV to PROD –
   1. Upload PDF in KNOWLEDGEBASE folder
   2. Update the metadata file
   3. Copy the new folder created in KnowledgeBase\_Extracts/BU\_NAME/NEW\_PDF to same location in PROD
   4. Do same for KnowledgeBase\_Images.
3. Open WinSCP, Go to **genai/knowledge\_base\_temp** folder
   1. Change KnowledgeBaseName in **milvus\_standalone.py**
   2. Change collection name in **check\_collection\_info.py and remove\_collection.py**
4. Similarly open Genai/Image\_KB\_Temp folder
   1. Change KnowledgeBaseName in **Create\_Milvus\_Image\_DB.py**
   2. Change collection name in **Check\_Collection\_Info.py and Remove\_Collection.py**
5. Open PROD VM terminal via putty/ec2 instance.
6. Go to genai folder, activate env
7. In **knowledge\_base\_temp** run –
   1. remove\_collection.py (wait 2 minutes after success)
   2. milvus\_standalone.py (creates db again)
   3. check\_collection\_info.py (run this after waiting 5 minutes)
8. In **Image\_KB\_Temp** run –
   1. Remove\_Milvus\_Collection.py (wait 2 minutes after success)
   2. Create\_Milvus\_Image\_DB.py (creates db again)
   3. Check\_Collection\_Info.py (run this after waiting 5 minutes)

# Steps to create DELETE document from Text and Image KnowledgeBase (DEV & PROD) –

**Prerequisites Process -**

1. Delete the Document from below list of folders –
   1. KNOLWELDGEBASE/BU\_NAME folder.
   2. KnowledgeBase\_Extracts/BU\_NAME
   3. KnowledgeBase\_Images/BU\_NAME
   4. KnowledgeBase\_Images\_Archive/BU\_NAME
2. Delete document metadata in BU\_NAME\_Metadata.json file present in KnowledgeBase\_Metadata folder
3. Open WinSCP, Go to **genai/knowledge\_base\_temp** folder
   1. Change KnowledgeBaseName in **milvus\_standalone.py**
   2. Change collection name in **check\_collection\_info.py and remove\_collection.py**
4. Similarly open Genai/Image\_KB\_Temp folder
   1. Change KnowledgeBaseName in **Create\_Milvus\_Image\_DB.py**
   2. Change collection name in **Check\_Collection\_Info.py and Remove\_Collection.py**

**Running Codes (python scripts) inside VM ->**

1. Open DEV/PROD VM terminal via putty/ec2 instance.
2. Go to genai folder, activate env
3. In **knowledge\_base\_temp** run –
   1. remove\_collection.py (wait 2 minutes after success)
   2. milvus\_standalone.py (creates db again)
   3. check\_collection\_info.py (run this after waiting 5 minutes)
4. In **Image\_KB\_Temp** run –
   1. Remove\_Milvus\_Collection.py (wait 2 minutes after success)
   2. Create\_Milvus\_Image\_DB.py (creates db again)
   3. Check\_Collection\_Info.py (run this after waiting 5 minutes)