[Introduction to SAP HANA Cloud](https://trials.cfapps.eu10-004.hana.ondemand.com/learning-journey/prd_hc_intro_proto)/Generative AI Hub with Vector grounding

**Hands-on Retrieval Augmented Generation (RAG) workflow**

1. Documents to be included in vector analysis are fed into the model.
2. The contents of the files are split into smaller chunks.
3. Embedding functions are used to create embeddings from the file/document chunks.
4. The embeddings are then stored as vectors in the SAP HANA Cloud Database.
5. When a query or prompt is submitted, the query itself is then embedded into vector form.
6. The query vector is compared to the values stored as vectors in SAP HANA Cloud via a similarity/semantic search.
7. The most appropriate results are forwarded, along with the original query, to a large language model such as Chat GPT.
8. The LLM uses the results of the HANA vector search to augment its own searching capabilities, and the final answer is returned to the user.

Try it out!

Jupyter Notebook Extension in Business Application Studio

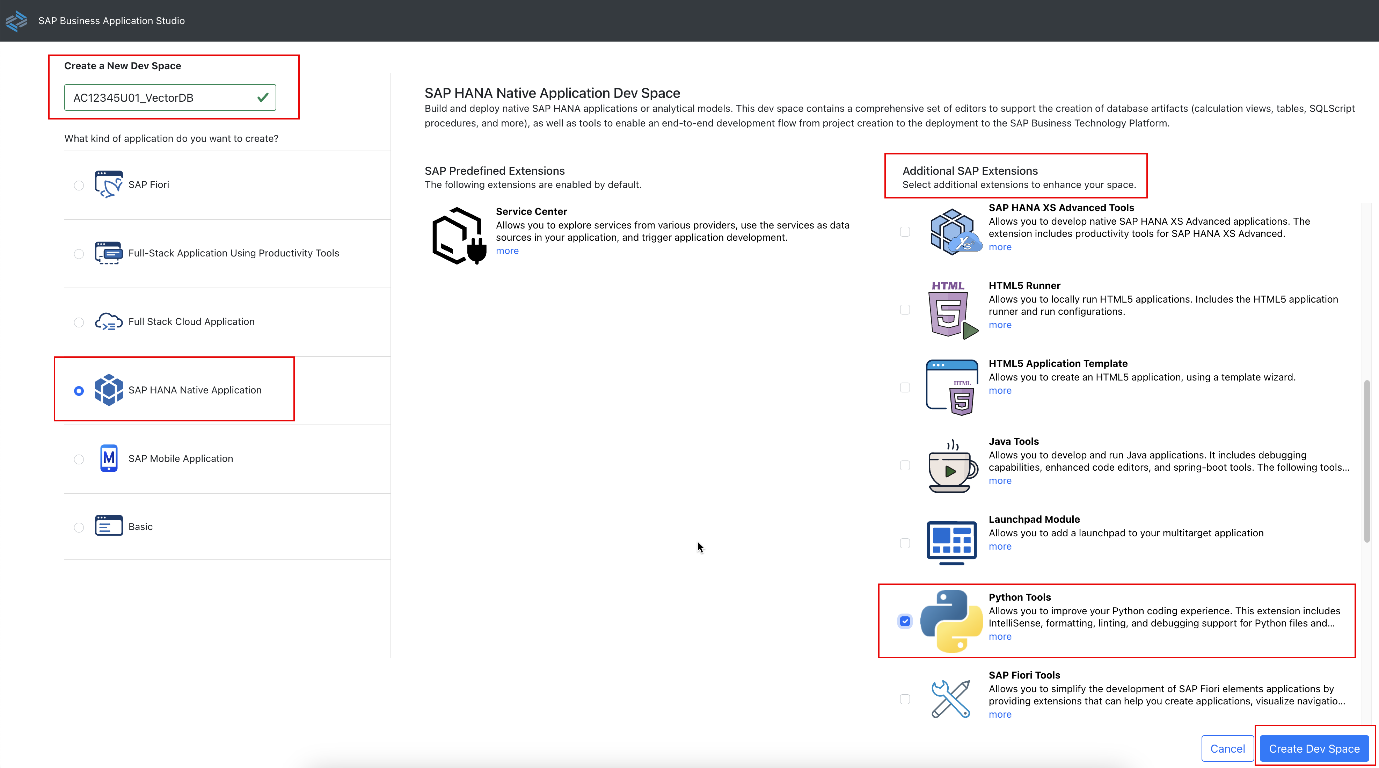
This exercise uses the Jupyter Notebooks extension in SAP Business Application Studio (BAS). This enables the use of python coding to leverage large language models in addition to the SAP HANA Cloud vector engine.

1. Select to open [Business Application Studio](https://sap-build-hana-cloud.eu10cf.applicationstudio.cloud.sap/index.html) to get started.

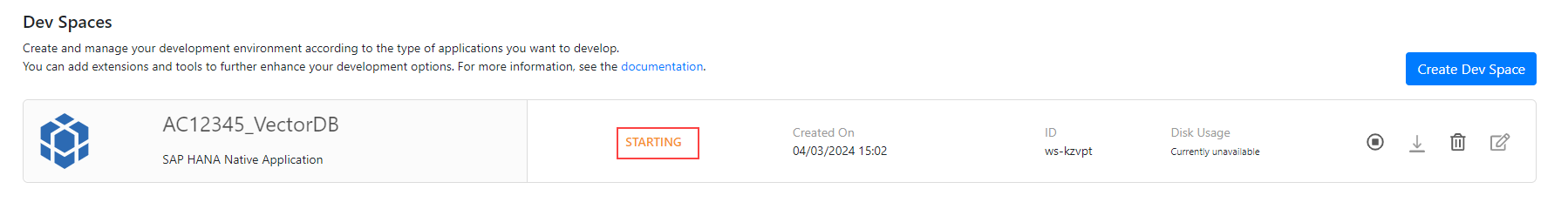
1. Log in with your user (AC207556U10) and provided credentials.

1. In the landing page for BAS, click on Create Dev Space to create a virtual environment in which to start a project.

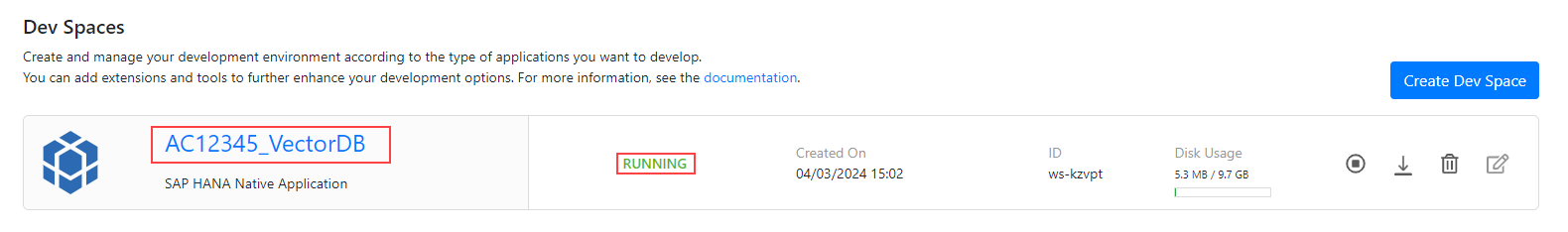
1. For the Dev Space name, use AC207556U10\_VectorDB and select SAP HANA Native Application as the application type. On the right hand side, under the list of *Additional SAP Extensions*, choose Python Tools and then select Create Dev Space to start the process. This will create a dedicated development environment with all the tools required to run the notebook.



1. It will take a few minutes for the Dev Space to start up. The status can be seen beside the name.



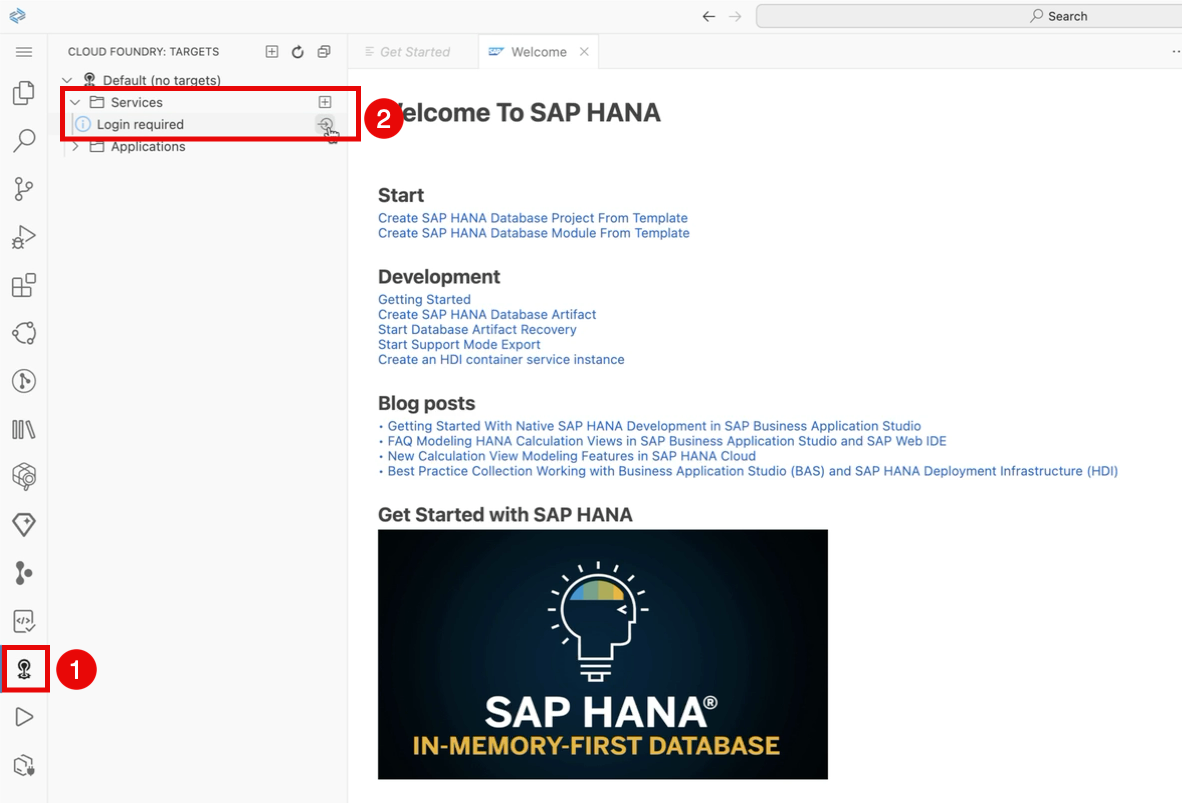
1. Once the status changes to Running, select the name (AC207556U10\_VectorDB) to open it.



Create Cloud Foundry SAP AI Core service key

1. The *Welcome to SAP HANA* page will appear. Follow the below sequence to log in to Cloud Foundry:

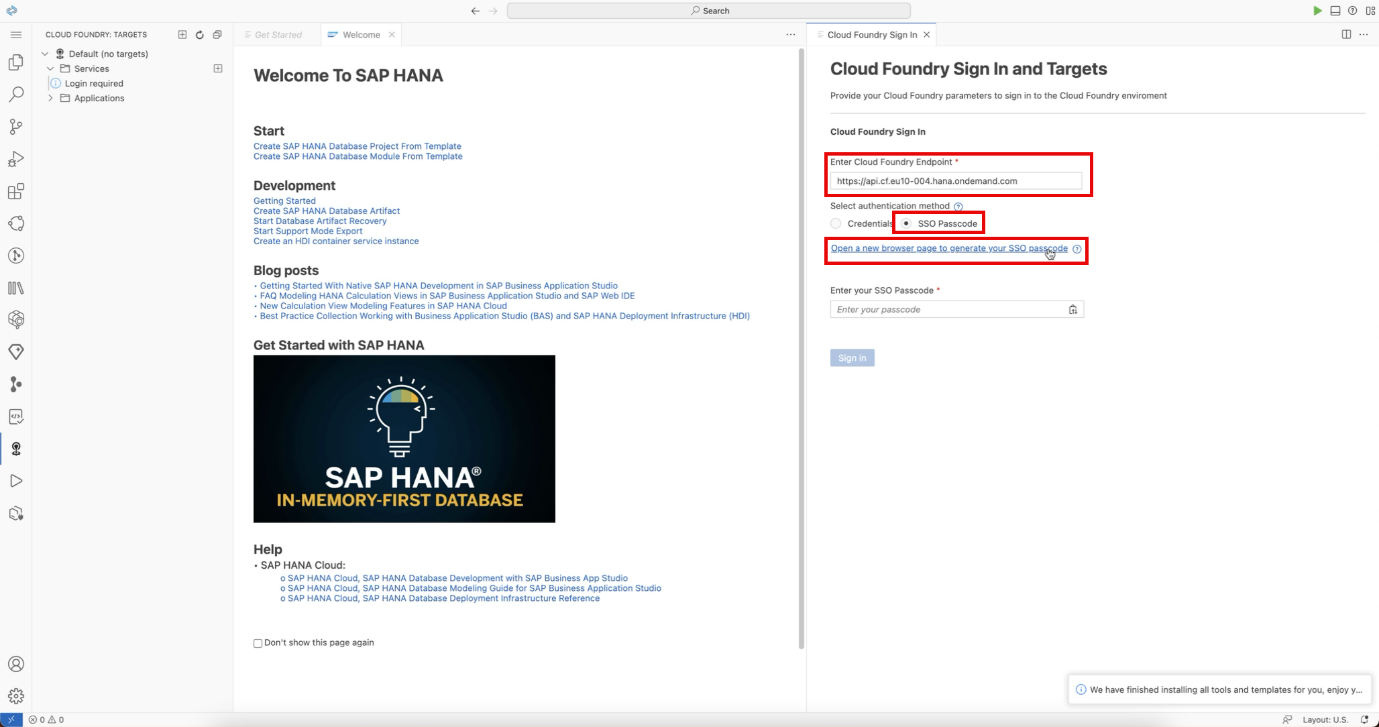
* Select the Cloud Foundry icon
* Expand Services folder and select the login icon.



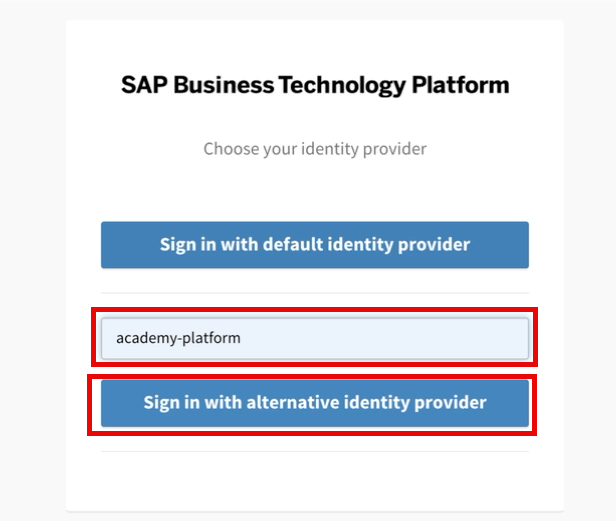
1. The Cloud Foundry Sign In panel will appear. Overwrite the Cloud Foundry Endpoint with the following value, and then choose SSO as the authentication method:

| Login Option | Selection |
| --- | --- |
| Cloud Foundry Endpoint | https://api.cf.eu10-004.hana.ondemand.com |
| Authentication method | SSO Passcode |

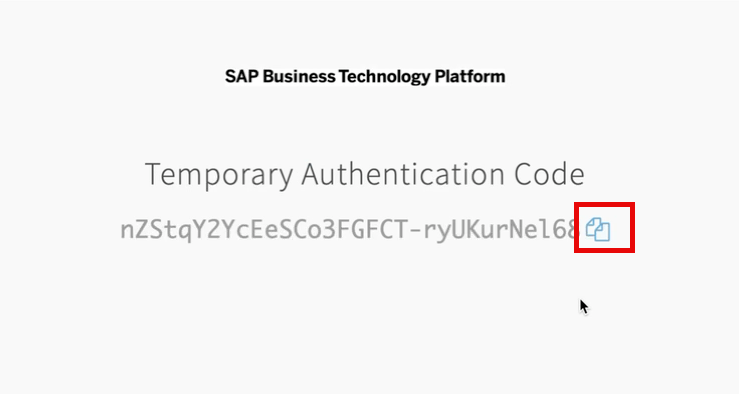
1. Select the Open a new browser page to generate your SSO passcode.



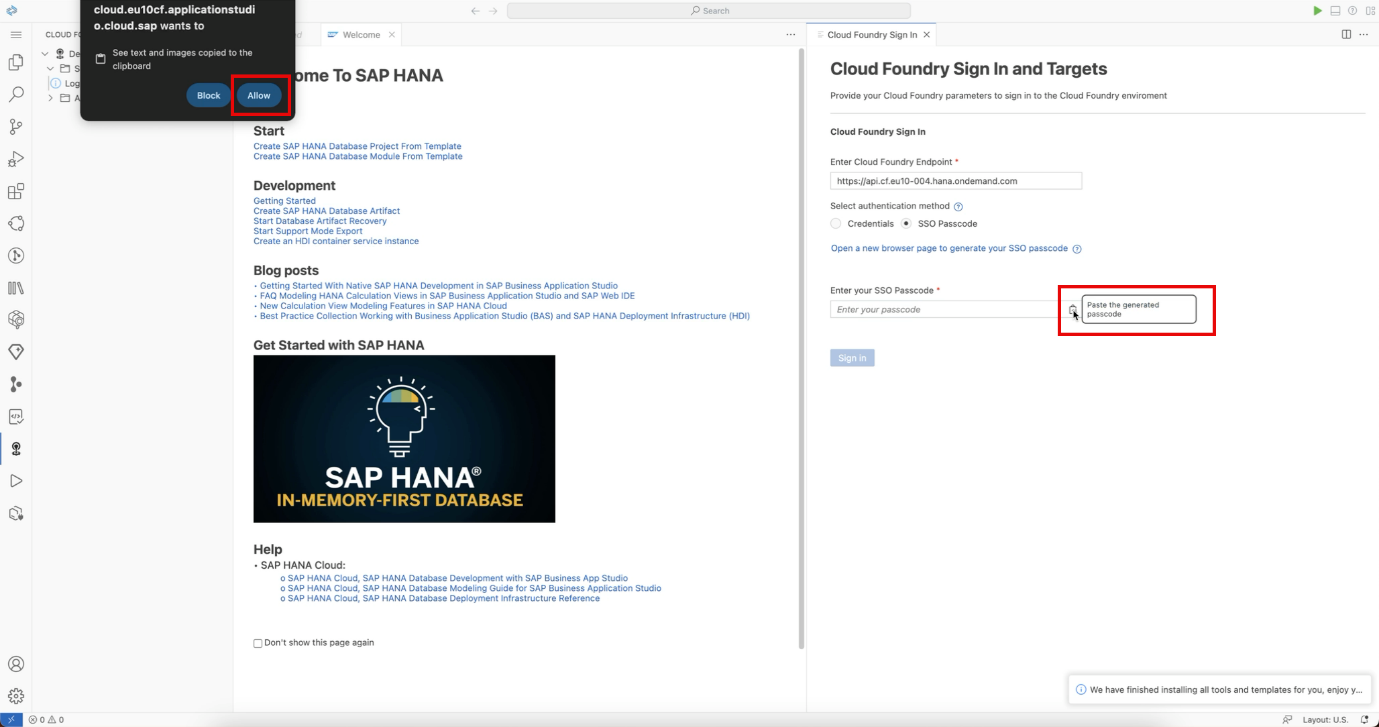
1. Enter academy-platform as the identity provider and then select the Sign in with alternative provider option.



1. Copy the generated passcode to the clipboard.

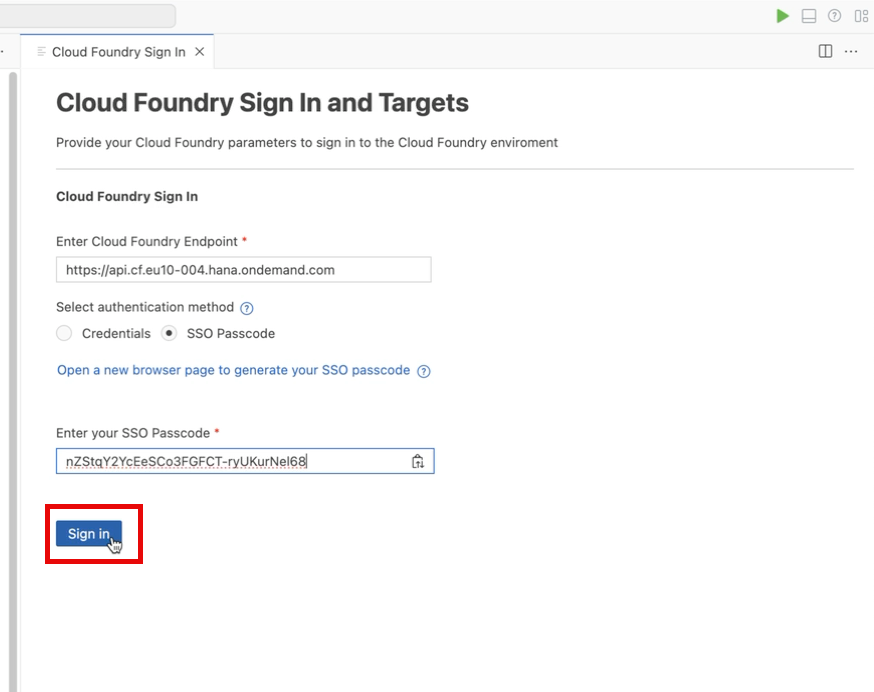


1. Paste the copied passcode in the Enter your SSO passcode field.



Note: If prompted for clipboard permissions by your browser, select Allow.

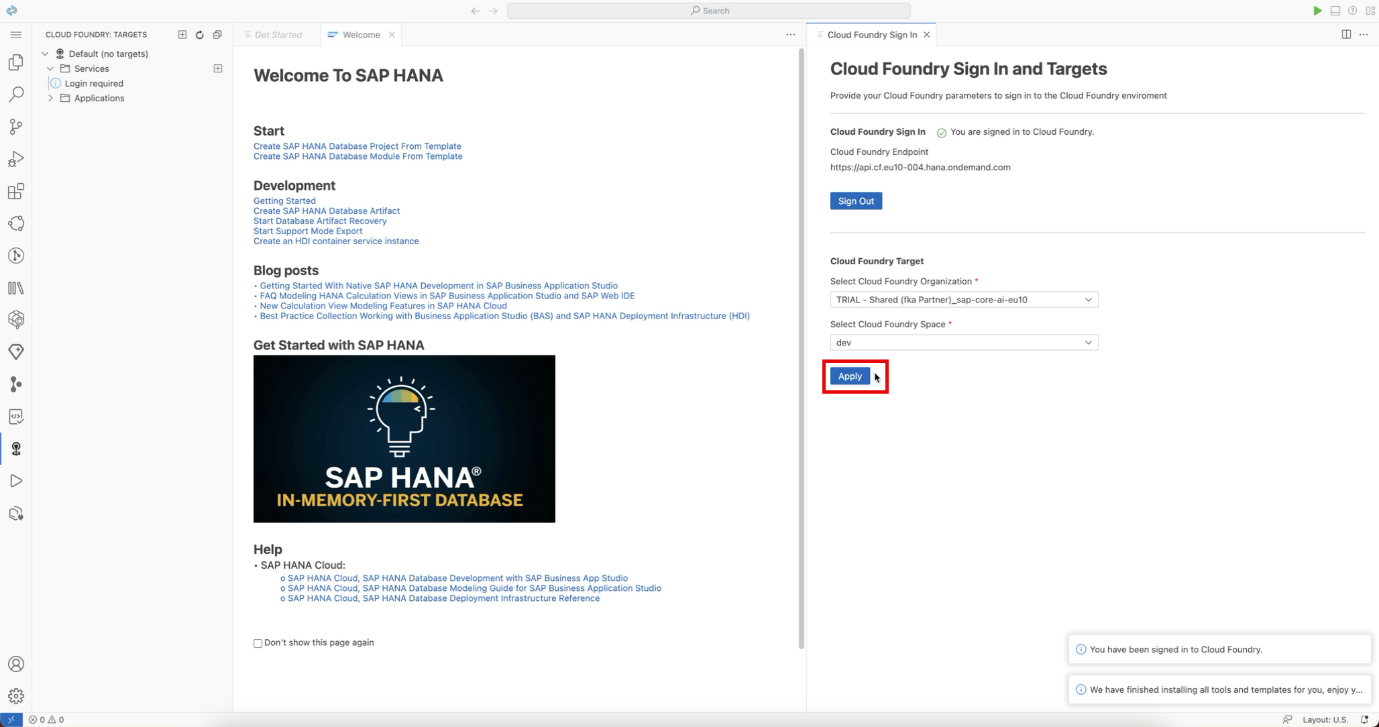
1. Sign into the Cloud Foundry Target by clicking the Sign In button.



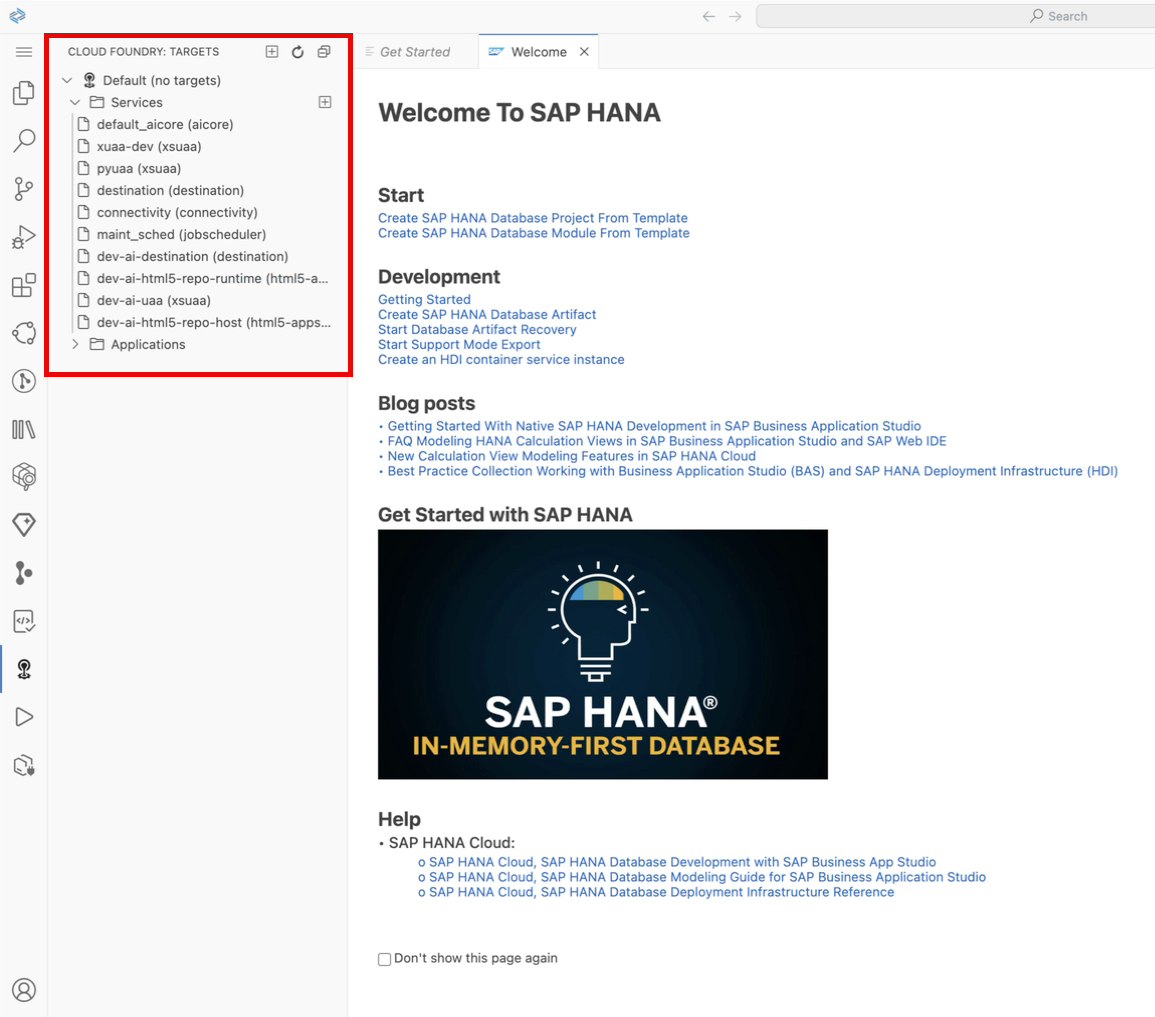
1. Proceed by selecting the desired Cloud Foundry Target and Space as seen below.

| Cloud Foundry Target | Selection |
| --- | --- |
| Cloud Foundry Organization | TRIAL - Shared\_sap-core-ai-eu10 |
| Space | dev |

1. Select Apply



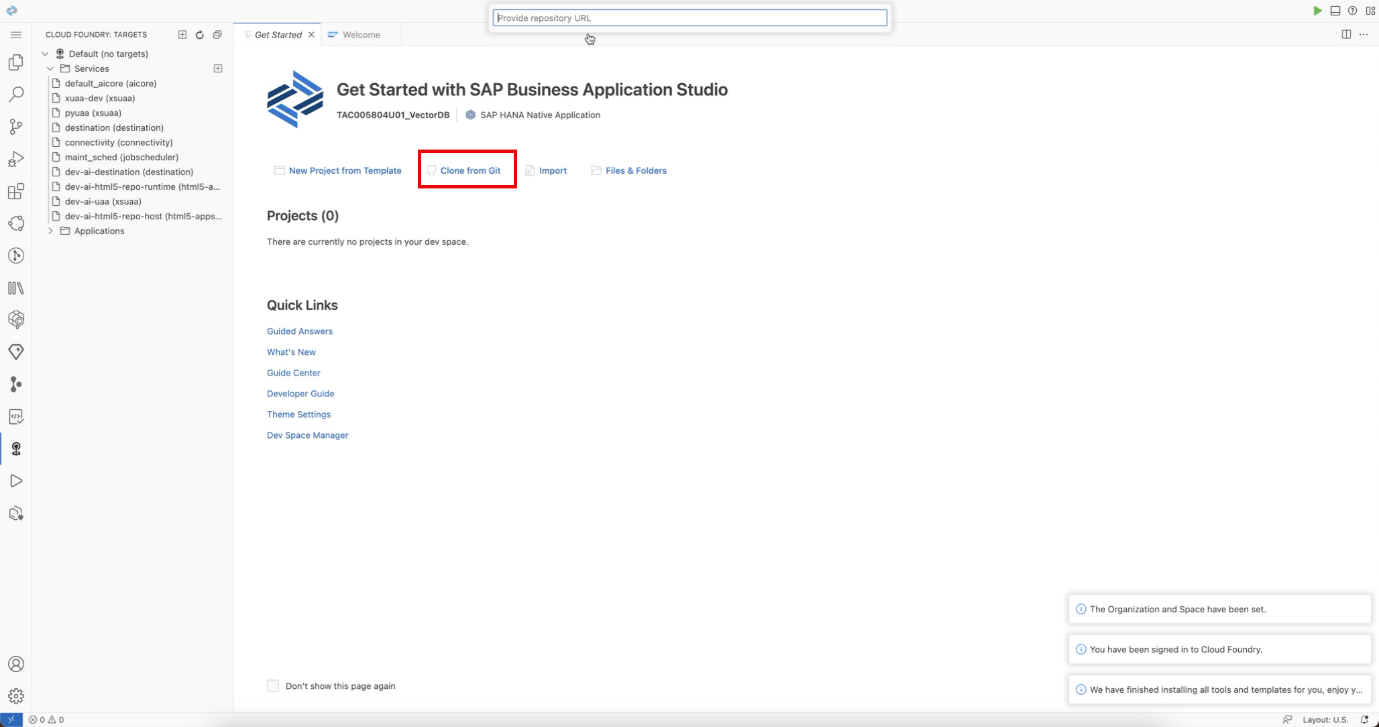
Observe all the services associated with the target:



Create SAP GenAI Core service key

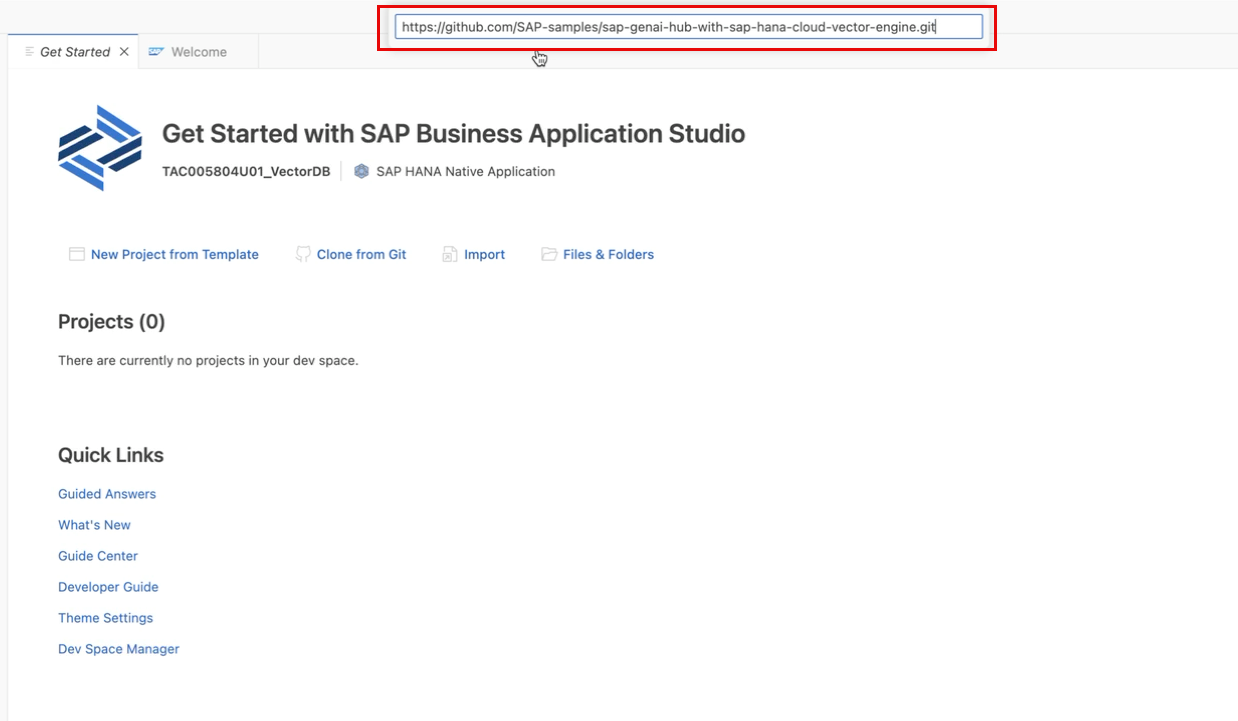
The following steps are to setup the required notebook.

1. On the Get Started page in Business Application Studio, select the option Clone from Git to import a repository from Github.

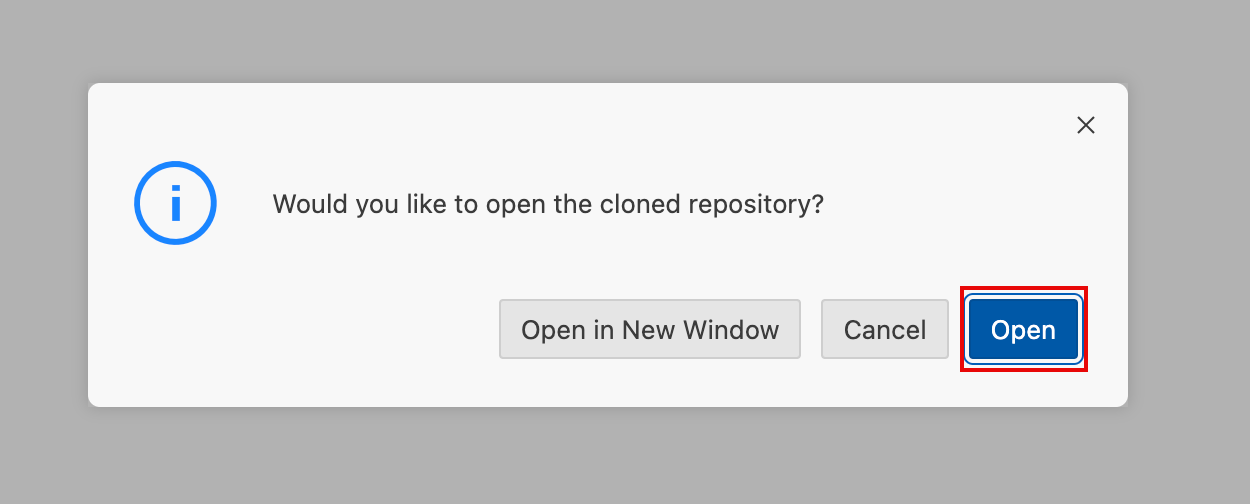


1. A prompt will appear on top asking for a URL to the repository. Copy and paste in the following URL and then press Enter.

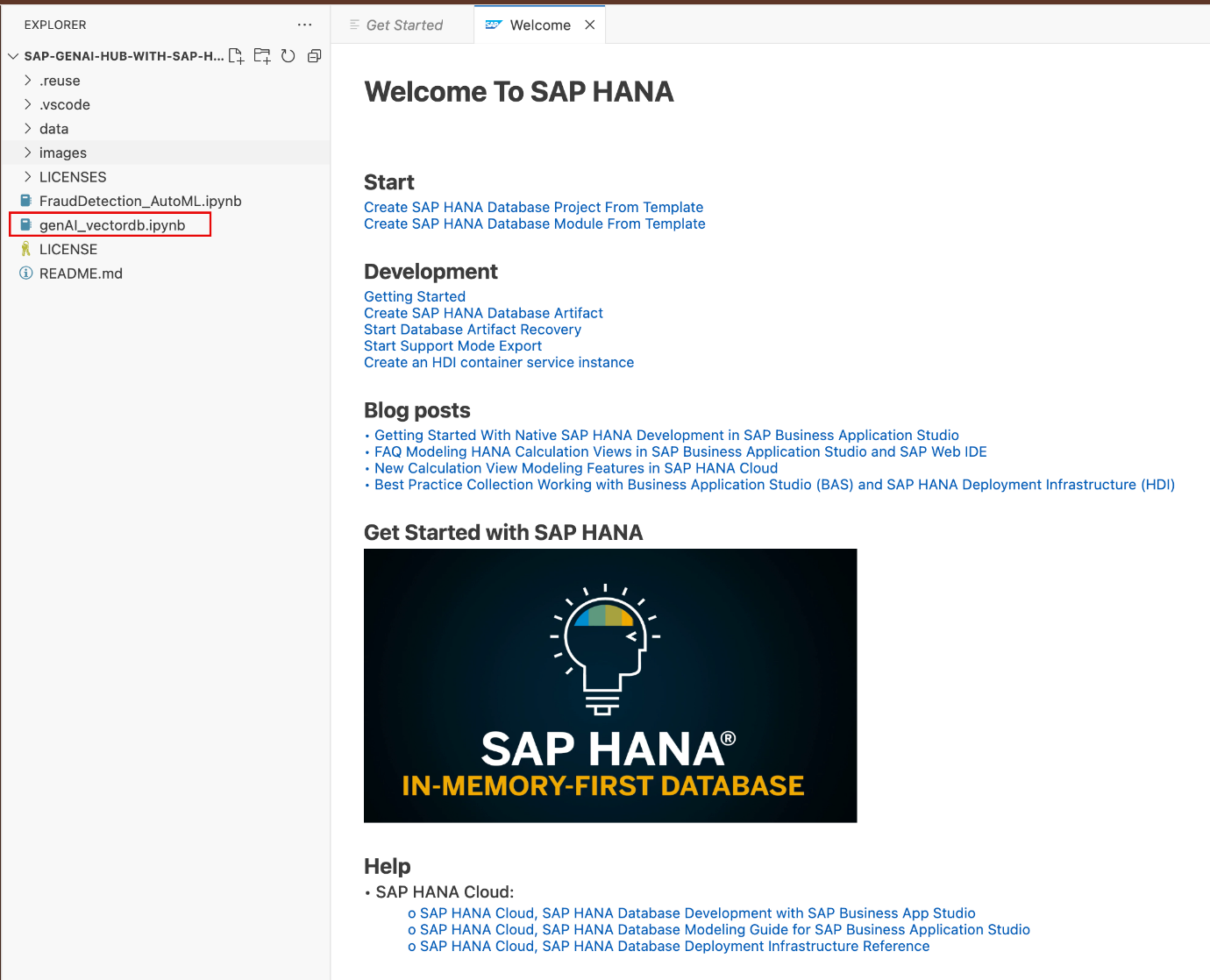
https://github.com/SAP-samples/sap-genai-hub-with-sap-hana-cloud-vector-engine.git



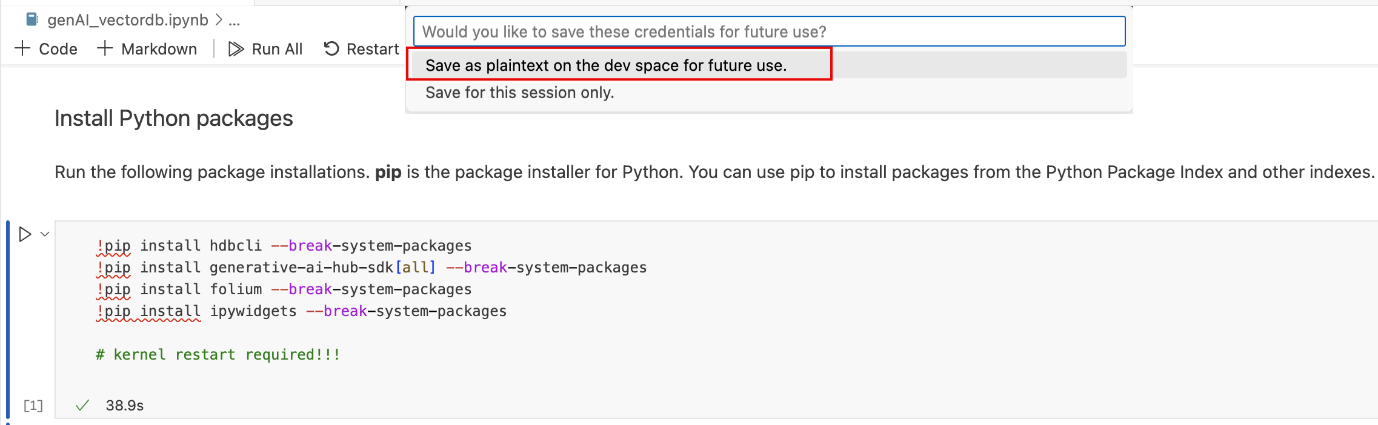
1. The following message will appear. Select Open to see the imported files in the Explorer window.



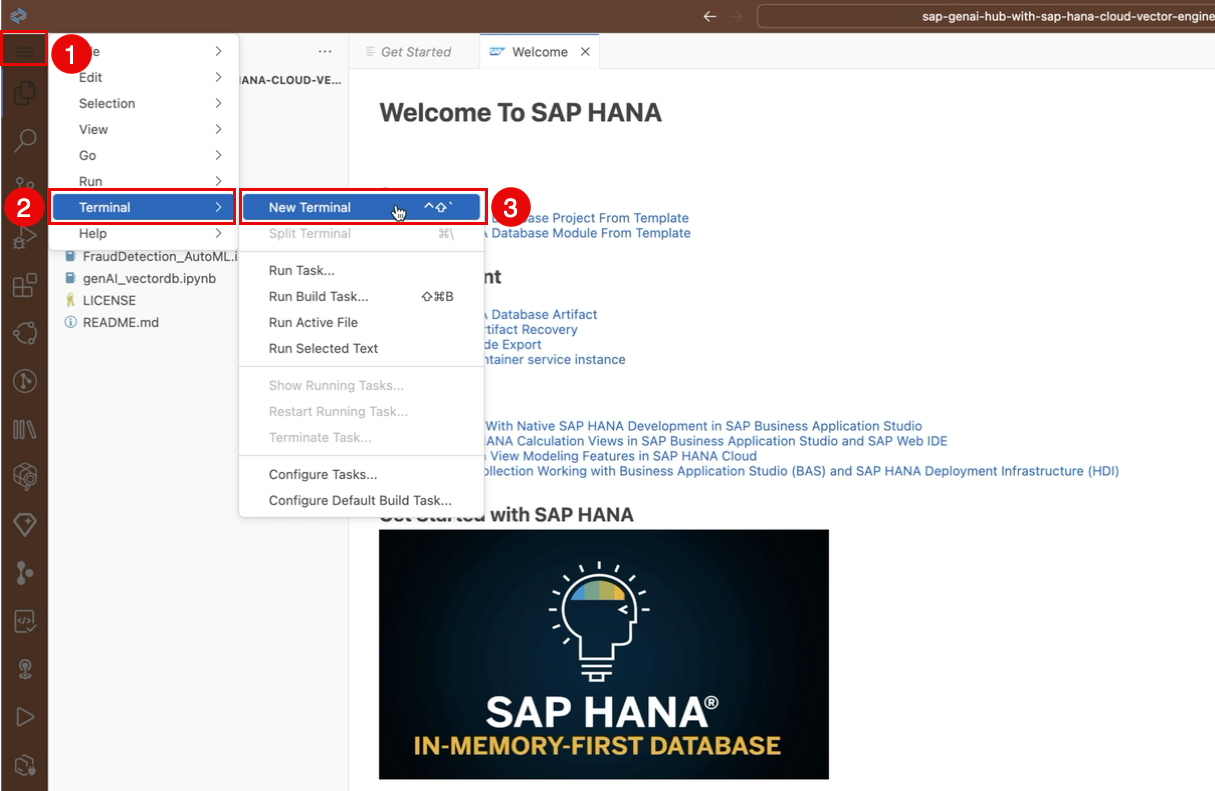
1. The files should now be visible in the Explorer pane on the left hand side. Select the Jupyter Notebook genAI\_vectordb.ipynb to open it.



Note: If the following message appears, select *Save as plain text on the dev space for future use*.



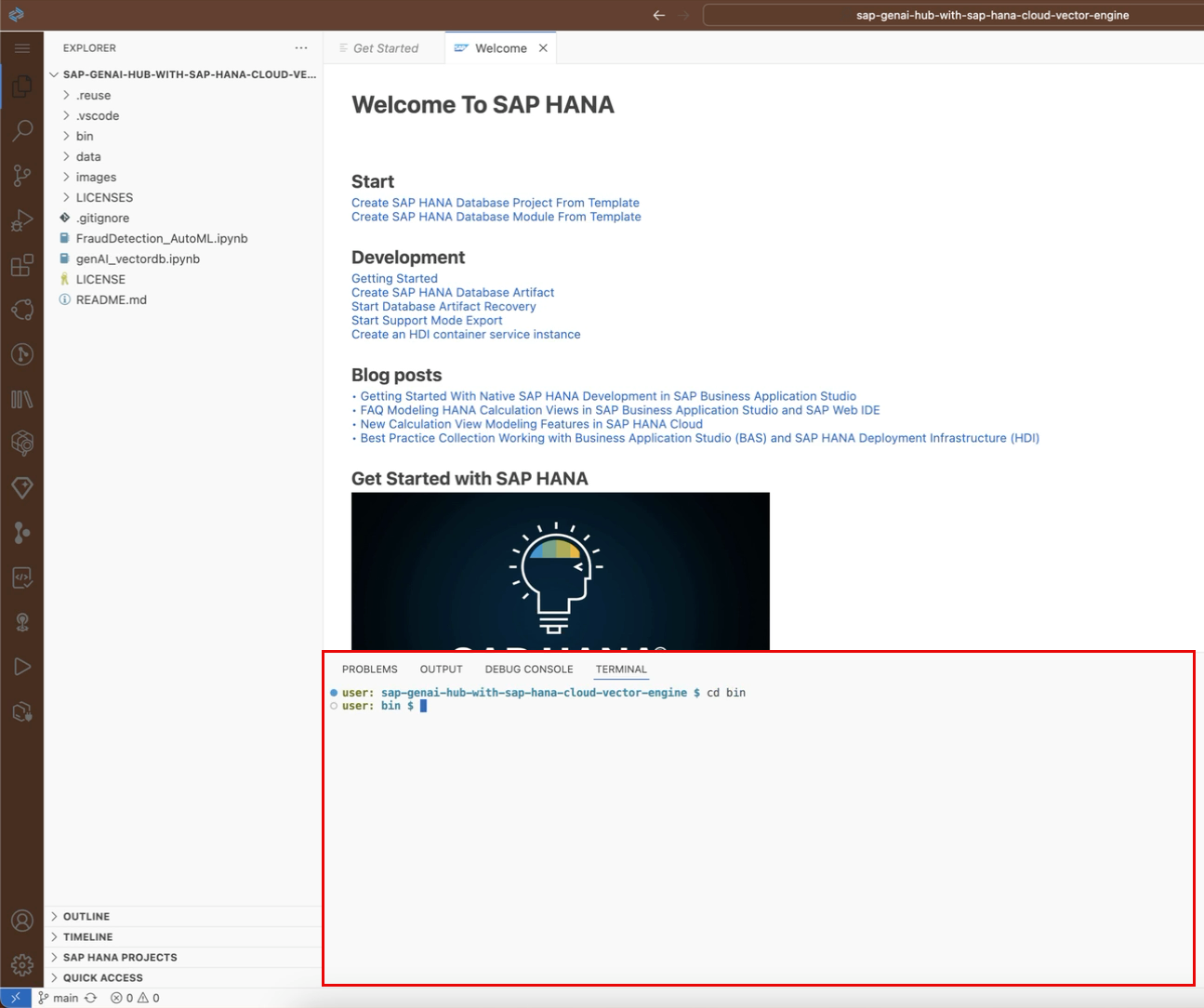
1. Open a new terminal in BAS by following the numbered sequence of steps.



1. Change to the bin folder where the cf\_create\_bas\_sso.py script is located.

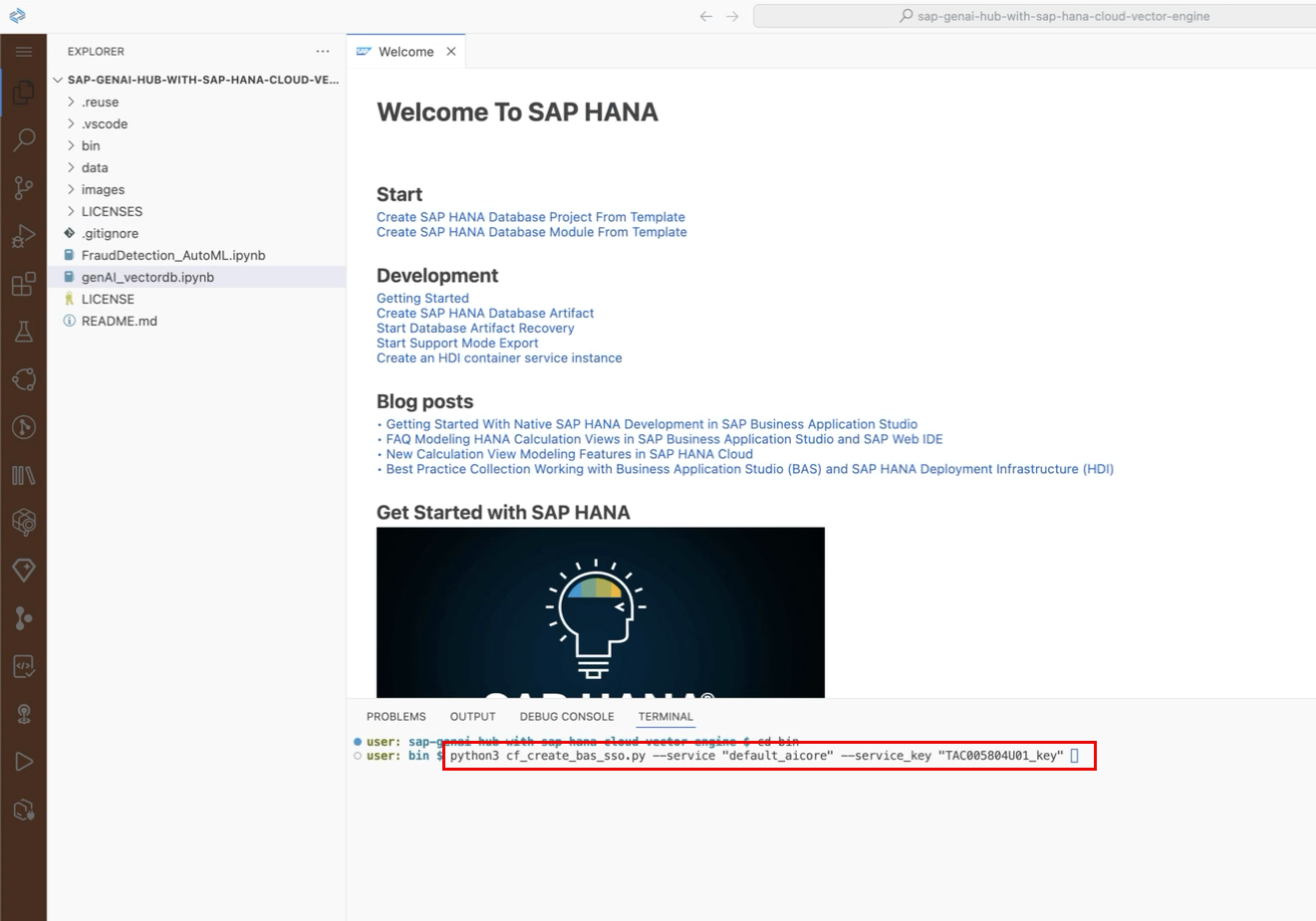
To change folder use the following shell command as seen below:

cd bin

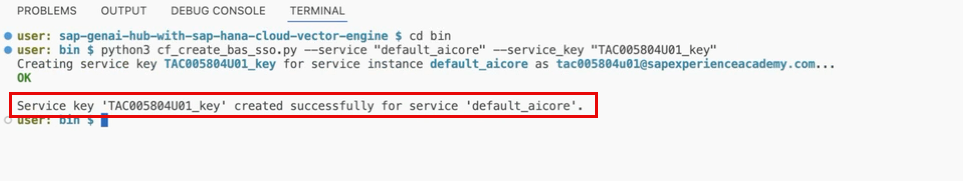


1. Create the SAP AI Core service key by executing the following command from the bin folder.

python3 cf\_create\_bas\_sso.py --service "default\_aicore" --service\_key AC207556U10\_key

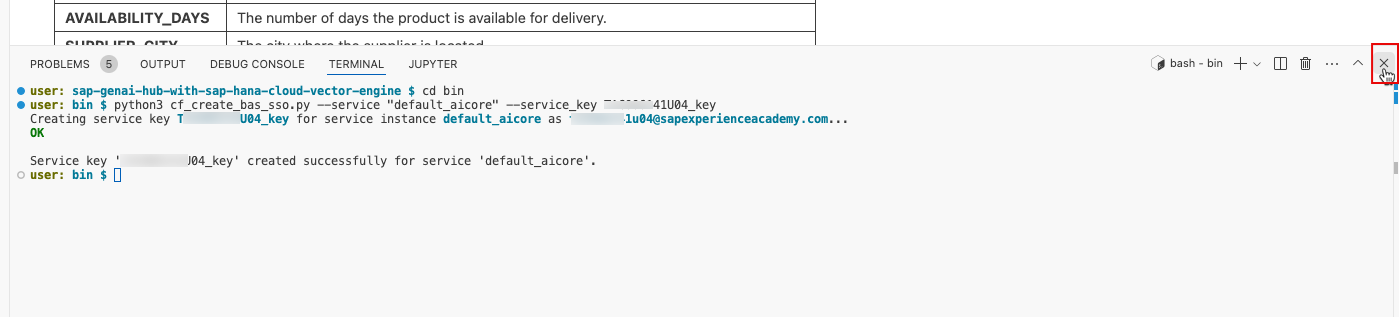


The script will indicate if the creation was successful or not.



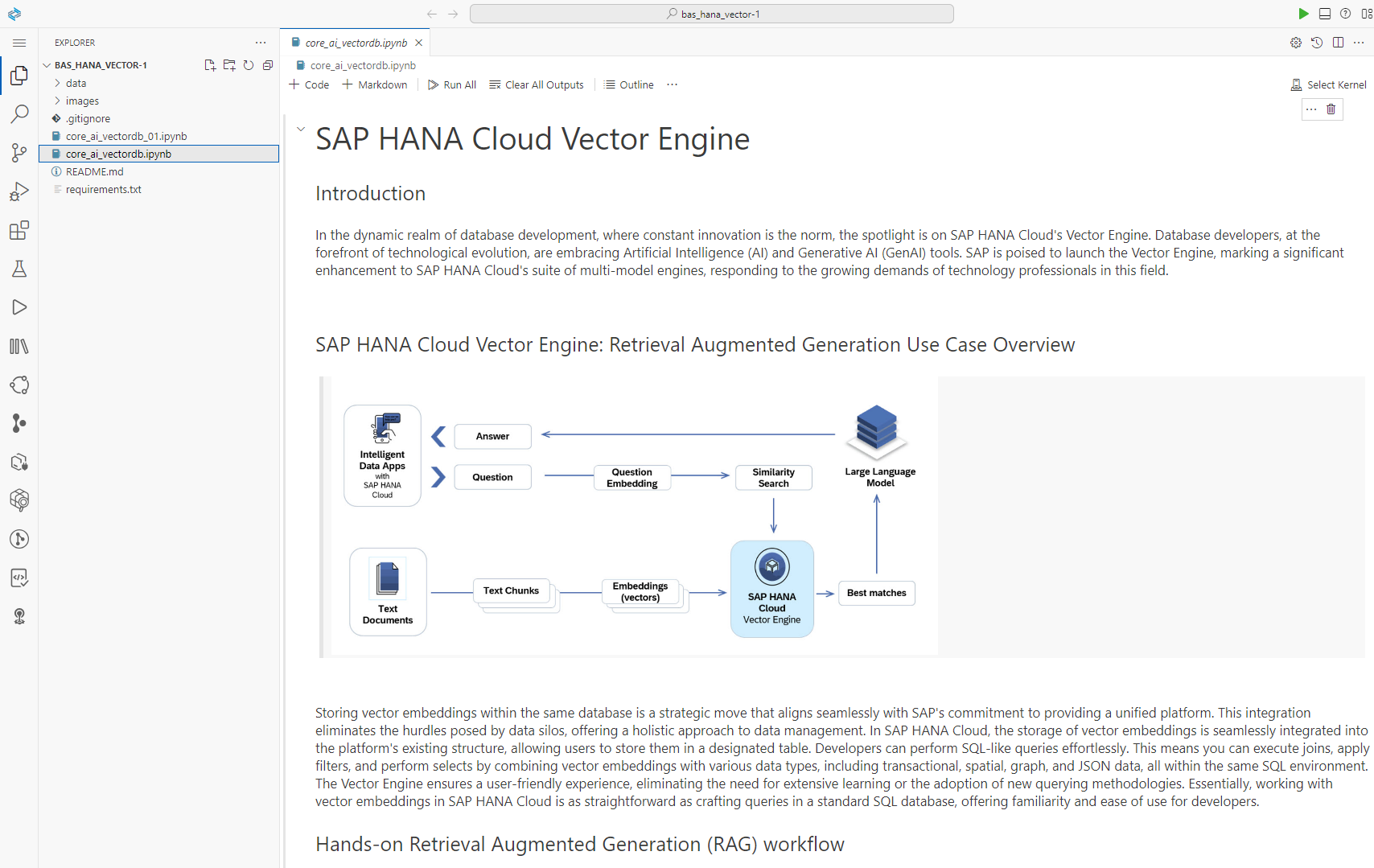
Note: This service key will be deleted once the workshop registration expires.

1. Once the key has been successfully created, the terminal window can be closed. Select the ‘X’ button to close it and return to the jupyter notebook.



Configure Jupyter and Python Extensions

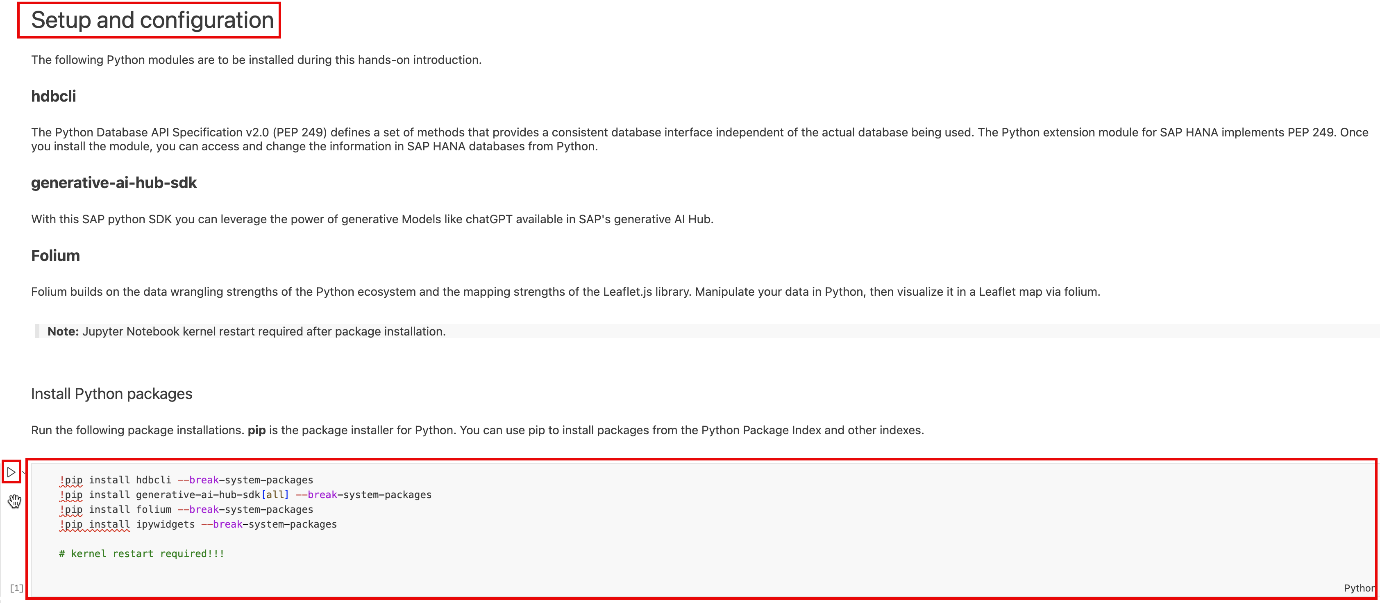
1. The notebook, with executable code cells, will be displayed on the screen.



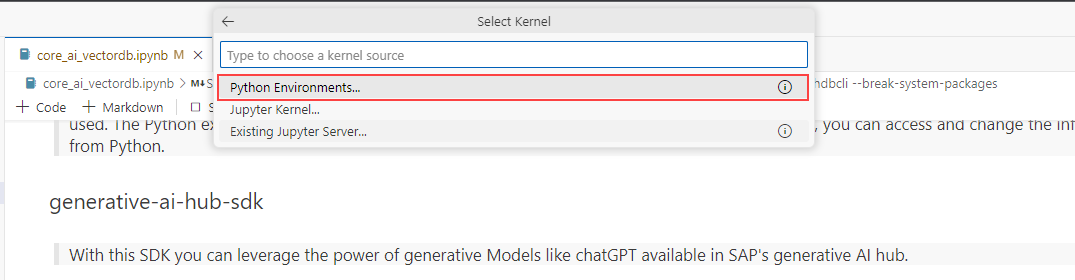
In the next section the python environment will be configured so that the code cells can be executed from within the Jupyter notebook.

1. Run the first code cell to install the required python modules.

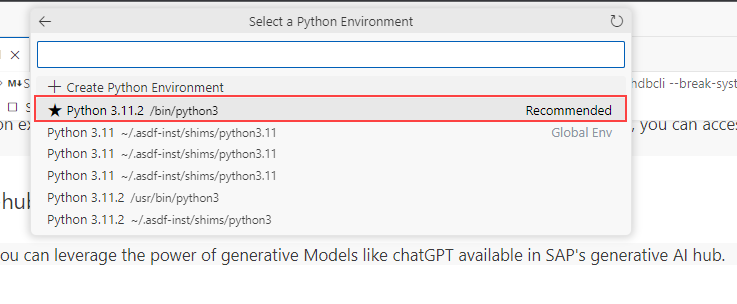
Note: To execute a code cell, click on the play icon beside the cell. It is also possible to execute it by clicking into the code cell and pressing *Shift+Enter*.



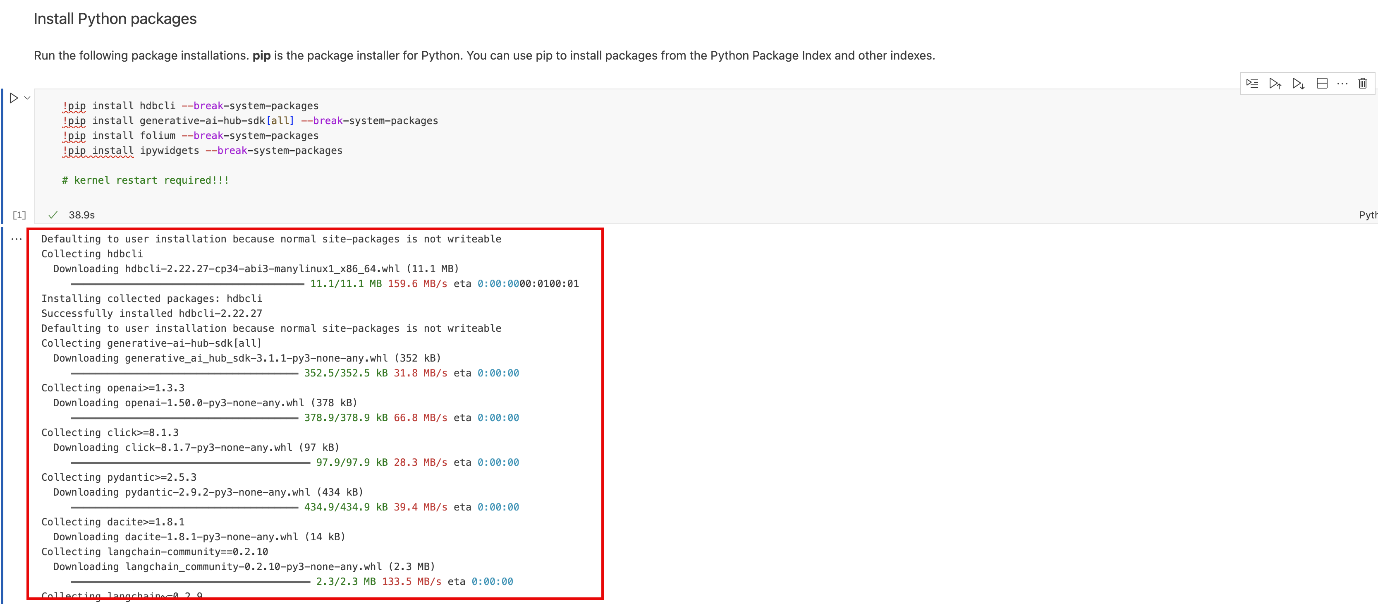
1. When the first cell is executed, a prompt asking to choose a kernel source will appear. Select Python Environments.



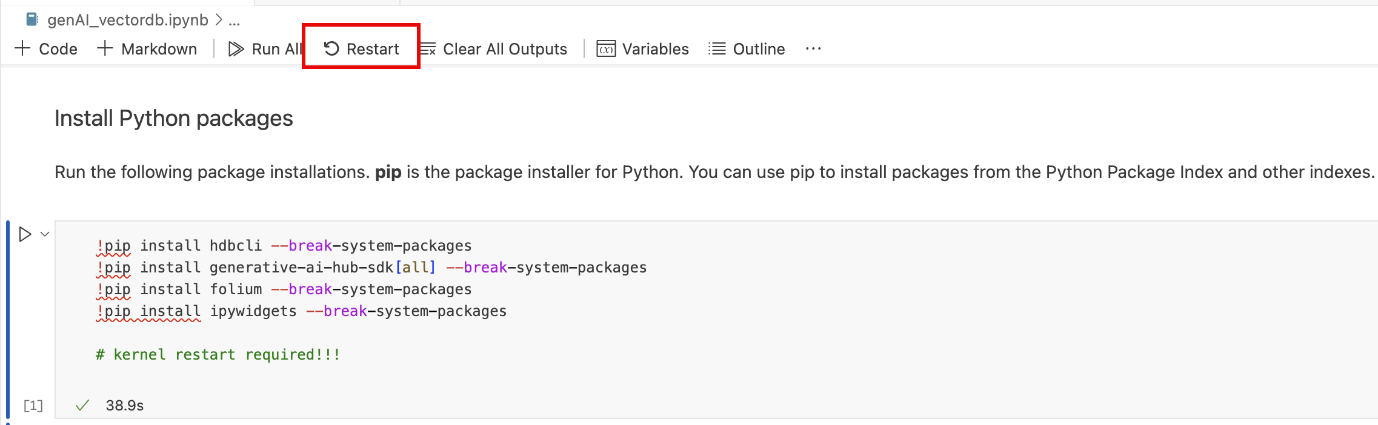
1. On the next screen, select the recommended Python kernel (3.11.2).



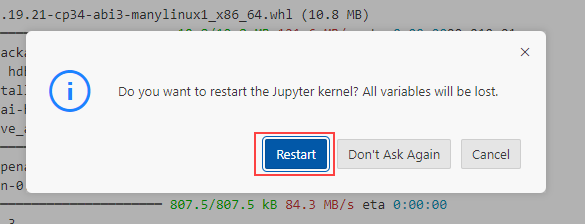
1. The required libraries will now start installing, and some messages about port numbers will also be visible in the lower right corner. This could take a few minutes to complete also.



1. Once the libraries have been successfully installed, the kernel must be restarted. Do this by clicking on the Restart option in the menu bar at the top.



1. On the following pop-up message box, click on Restart again to complete the process.



Note: It will take a few seconds for the kernel to restart and once done, the environment is ready for use!

1. Please follow the instructions in the Jupyter Notebook in BAS to continue with the rest of the lesson.

Note: For the section on connecting to the SAP HANA Cloud instance with the vector engine, please use the following as the HANA hostname:  
13b7c15d-848f-40b5-9259-c9c36ab85f56.hna1.prod-eu10.hanacloud.ondemand.com

