

Lab 8 - Knowledge based RAG agent

In this lab we will dive deeper in to building agent with Watsonx Orchestrate ADK. For this lap please **make sure that you have Python version 3.11 installed**. If not, please follow the [link](#) to install python 3.11 with brew.

1. Clone the following [git repo](#)

```
git clone https://github.com/edsml-kl121/watsonx-orchestrate-adk-tutorial.git
```

2. cd into the **08_knowledge_based_example** folder. Here is the structure of the folder

```
✓ 08_knowledge_based_example
  ✓ agents
    ! ibm_agent.yaml
  ✓ knowledge_base
    📄 thai_leave_policy.pdf
    ! thai_policy_knowledge_base_2.yaml
  $ import_all.sh
  $ remove_all.sh
```

3. create virtual environment using the following command. Make sure to use python 3.11 to create the environment and source to the .venv

```
python3.11 -m venv .venv
source .venv/bin/activate
```

4. install the requirement library

```
pip install ibm-watsonx-orchestrate
```

To check if the library is installed properly, you could try command **orchestrate**. The following text should pop up, indicates that Watsonx orchestrate ADK is installed

```
(.venv) → 08_knowledge_based_example git:(main) ✕ orchestrate
```

Usage: orchestrate [OPTIONS] COMMAND [ARGS]...

Options

<code>--version</code>	Show the installed version of the ADK and Developer Edition Tags
<code>--debug</code>	Enable debug mode
<code>--install-completion</code>	Install completion for the current shell.
<code>--show-completion</code>	Show completion for the current shell, to copy it or customize the installation.
<code>--help</code>	Show this message and exit.

Commands

<code>env</code>	Add, remove, or select the activate env other commands will interact with (either your local server or a production instance)
<code>agents</code>	Interact with the agents in your active env
<code>tools</code>	Interact with the tools in your active env
<code>toolkits</code>	Interact with the toolkits in your active env
<code>knowledge-bases</code>	Upload knowledge your agents can search through to your active env
<code>connections</code>	Interact with the agents in your active env
<code>server</code>	Manipulate your local Orchestrate Developer Edition server [requires entitlement]
<code>chat</code>	Launch the chat ui for your local Developer Edition server [requires entitlement]
<code>models</code>	List the available large language models (llms) that can be used in your agent definitions
<code>channels</code>	Configure channels where your agent can exist on (such as embedded webchat)
<code>evaluations</code>	Evaluate the performance of your agents in your active env
<code>copilot</code>	Access AI powered assistance to help refine your agents
<code>settings</code>	Configure the settings for your active env

5. To move on to the next step, we will need to obtain your API KEY and URL. You can find your API KEY and URL on the Watsonx Orchestrate webpage

5.1 Go to the **profile** on the top right corner > **setting**

The screenshot displays the IBM watsonx Orchestrate web interface. The top navigation bar includes 'Contact Sales', 'Buy Now', and a user profile icon. The main content area is titled 'Build agents and tools' and features three summary cards: 'Total messages' (48, +100% from the previous day), 'Failed messages' (11, +100% from the previous day), and 'Latency average' (14326 ms, +100% from the previous day). Below these cards, there are sections for 'All agents' (7) and 'All tools' (1). The 'diabetes_agent_kb' tool is highlighted, showing its description and associated tools/agents. A user profile dropdown menu is open on the right side of the screen, showing the user's name 'Tanyaton Oranrigsupak', region 'us-south', plan 'Trial', and owner 'Techzone.Adhoc01@ibm.com'. The 'Settings' option in the dropdown menu is highlighted with a red box.

5.2 Go to **API details**

5.3 Note down your “**Service instance URL**”

IBM watsonx Orchestrate

Settings

Data Retention **API details**

Generate API keys

API keys are intended to generate the JWT token that you need to authenticate your user before you make requests to the API. Use the service instance URL to request access to the environment's resources.

[Learn more about API keys](#)

[Generate API key→](#)

Service instance URL

`https://a`

5.4 Click on “**Generate API keys**”. This will redirect you to a new webpage

5.6 Click on **Create** > add the Name for you key > Click **Create**

Make sure to copy and save, or download your API KEY and URL

IBM Cloud

Search resources and products...

Catalog Manage 2784110 - itz-watsonx-...

API keys

Create, view, and work with API keys that you have access to manage. IBM Cloud API keys are associated with a user's identity and can be used to access cloud platform and classic infrastructure APIs, depending on the access that is assigned to the user. The following table displays a list of API keys created in this account. [Learn more.](#)

Looking for more options to manage API Keys? Try [IBM Cloud® Secrets Manager](#) for creating and leasing API keys dynamically and storing them securely in your own dedicated instance.

Unused or overly permissive API keys increase the risk of unauthorized access. Regularly review the [Inactive identities report](#), rotate keys, and apply only the minimum required permissions.

API keys associated with a user's identity have the same access that the user is assigned across all accounts. To update the access for an API key, assign or remove access for the user.

View
My IBM Cloud API keys

Filter by API key name or description

Status	Name	Description	Date created	Enabled
			7-24-2025 08:59 GMT	Yes

Items per page: 25 1-25 items Page 1

[Create](#)

Create IBM Cloud API key



Name

sample-api-key

Description (optional)

Enter description

Leaked action

If API key is discovered to have been leaked out in the world, what would you like the system to do?

- ☒ Disable the leaked key
- ☐ Delete the leaked key
- ☐ Nothing

Session management

Enable session management for CLI logins? ⓘ

- ☐ Yes
- ☒ No

Cancel

Create

API keys

Create, view, and work with API keys that you have access to manage. IBM Cloud API keys are associated with a user's identity and can be used to access cloud platform and classic infrastructure APIs, depending on the access that is assigned to the user. The following table displays a list of API keys created in this account. [Learn more.](#)

Looking for more options to manage API Keys? Try [IBM Cloud® Secrets Manager](#) for creating and leasing API keys dynamically and storing them securely in your own dedicated instance.

Unused or over
required perm

API keys assoc
the user.

View
My IBM Clou

Filter b

Status

miw test 24july

sample-api-key

Items per page: 25

1-25 items

Page 1

API key successfully created

Copy the API key or click download to save it. You won't be able to see this API key again, so you can't retrieve it later. The API key is no longer displayed after 280 seconds.

API key

.....

Copy

Download

Now you should already have URL and API KEY saved in your local device:

URL: <https://api.us-south.watson-orchestrate.cloud.ibm.com/xxxxxxxxxxxxxxxxxxxxxx>

API: XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

6. Add new environment to watsonx Orchestrate

6.1 Run the following command on your terminal at the **08_knowledge_based_example** folder

```
orchestrate env add -n .venv -u <YOUR_URL>

orchestrate env activate
```

put in your URL you have saved from the previous step in the first command. After running the second command, you will be asked to put in your API_KEY. Please do so using the API KEY you have saved.

Example:

```
(.venv) → 08_knowledge_based_example git:(main) ✖ orchestrate env add -n .venv -u https://api.dl.watson-orchestrate.ibm.com/instances/20250515-0617-4459-503c-8c686a2c306d

[INFO] - Environment 'new-venv' has been created
(.venv) → 08_knowledge_based_example git:(main) ✖ orchestrate env activate .venv

Please enter WXO API key:
[WARNING] - Using MCSP Auth Type. If this is incorrect please use the '--type' flag to explicitly choose one of 'ibm_iam', 'mcsp_v1', 'mcsp_v2' or 'cpd'
[INFO] - Environment '.venv' is now active
```

6.2 Add and activated the environment, you can check the list of orchestrate environment with the following command. The **.venv** environment should have the activate mark on the right

```
orchestrate env list
```

Example:

```
(.venv) → 08_knowledge_based_example git:(main) ✖ orchestrate env list

.venv      https://api.dl.watson-orchestrate.ibm.com/instances/20250515-0617-4459-503c-8c686a2c306d      (active)
local      http://localhost:4321
miew-env   https://api.eu-central-1.dl.watson-orchestrate.ibm.com/instances/20250616-0240-4655-107d-25a33ce99b20
new-venv   https://api.us-south.watson-orchestrate.cloud.ibm.com/instances/4a1f0aa7-fa04-41af-86f4-8b8fcbba91b8
```

7. We are now going to **create a RAG agent** using the following code in the folder

7.1 understanding file structure

```
✓ 08_knowledge_based_example
  ✓ agents
    ! ibm_agent.yaml
  ✓ knowledge_base
    📄 thai_leave_policy.pdf
    ! thai_policy_knowledge_base_2.yaml
  $ import_all.sh
  $ remove_all.sh
```

! ibm_agent.yaml

yaml file used to specify agent name, description, llm model used, style, and knowledge base source

! thai_policy_knowledge_base_2.yaml

yaml file used to specify agent's knowledge name, description, source document, and embedding model

◆ Thai_leave_policy.pdf

PDF document contain information for agent knowledge

You can discover more by going into each file

To create this agent on watsonx orchestrate, there are a list of command you need to execute. That list of commands is provided in the file **import_all.sh**

7.2 Run the file using the following command

```
bash import_all.sh
```

8. After running the command successfully, you should be able to see the agent on the watsonx Orchestrate UI. Click on the agent

IBM watsonx Orchestrate

Contact SalesBuy Now24 trial days left

Agent chat /

Build agents and tools

Total messages ①

10

▲ 100% from the previous day

Failed messages ①

1

▲ 100% from the previous day

Latency average ①

42864 ms

▲ 4788% from the previous day

[View all →](#)

All agents12

All tools10

thai_policy_agent

You are an helpful agent that has knowledge about Thai policies and can answer questions related to them.

ToolsAgents

— —

→

As you can see, the information specified in the yml file (eg. Name, description, knowledge) is now in the agent detail

IBM watsonx Orchestrate

Contact SalesBuy Now24 trial days left

Agent chat / Manage agents / thai_policy_agent

thai_policy_agent

Model : llama-3-405b-instr...

Profile

Knowledge

Toolset

Behavior

Channels [Preview](#)

Edit knowledge settings

Knowledge source

Upload files or connect to content repositories to add knowledge to your agent.

Change source

Files

Connected

Upload files

File name	Date uploaded
thai_leave_policy.pdf	7/29/25

Items per page: 5 1-1 of 1 items 1 of 1 page

Knowledge source description

General information about Thai Policies

Toolset

Add tools and other agents to help your agent take action.

Preview

6:30 PM

Hello, welcome to watsonx Orchestrate

Accuracy of generated answers may vary. Please double-check responses.

What can you do for me?

Learn about watsonx Orchestrate AI Agent and Gen AI capabilities

Formalize Message

Helps you craft more professional messages

Summarize Meeting Notes

Bullet key points from meeting notes

Type something...

9. The agent can now be tested with the following prompt

- พนักงานสามารถลาชั่วคราวได้กี่วันต่อปี
- การลาคลอดได้รับค่าจ้างกี่สัปดาห์
- พนักงานชายมีสิทธิลาพ่อแม่ได้กี่วัน
- สามารถนำการลาไปปีถัดไปได้สูงสุดกี่วัน
- การลาโดยไม่ได้รับค่าจ้างสามารถลาได้สูงสุดกี่เดือน

You 06:33 PM

- พนักงานสามารถลาชั่วคราวได้กี่วันต่อปี

You 06:39 PM

พนักงานหญิงลาคลอดได้กี่วัน

thai_policy_agent 06:33 PM | Show Reasoning ▾

พนักงานสามารถลาชั่วคราวได้สูงสุด 12 วัน ในหนึ่งปี ^

You 06:35 PM

- การลาคลอดได้รับค่าจ้างกี่สัปดาห์

thai_policy_agent 06:35 PM | Show Reasoning ▾

26 สัปดาห์

You 06:36 PM

- พนักงานชายมีสิทธิลาพ่อแม่ได้กี่วัน

thai_policy_agent 06:36 PM | Show Reasoning ▾

พนักงานชายมีสิทธิลาพ่อแม่ได้สูงสุด 7 วัน ^

thai_policy_agent 06:39 PM | Show Reasoning ▾

พนักงานหญิงมีสิทธิได้รับการลาคลอดพร้อมกับค่าจ้างเต็มจำนวน สำหรับระยะเวลา 26 สัปดาห์ติดต่อกัน (ไม่รวมวันหยุดนักขัตฤกษ์)

thai_leave_policy.pdf

HRMS ั้ได้รับ ั้ได้รับล่วงหน้า 15 วัน หลังจากยื่น ันคำร้องแล้ว การแจ้ง เดือนจะถูกส่งไปยังผู ่อการลาได้รับ การอนุมัติ การแจ้งเดือนจะถูกส่งไป ยังพนักงานและแผนกทรัพยากร...

[View source](#)

< 1/1 >