

IBM watsonx Assistant for Z for Technical Sales Level 4 Lab Guide

Description	IBM watsonx Assistant for Z for Technical Sales Level 4 Lab Guide
Author(s)	Andrew R. Jones (andrewj@us.ibm.com) and Max Weiss (maxwell.g.weiss@ibm.com)
Copyright	Copyright © 2024 IBM - V 2.0

Table of Contents

Welcome

- Support
- Using the demonstration guide

IBM Technology Zone environment

- Create a reservation request
- Extend the reservation
- Join the
- Accessing the environments
- Troubleshooting

Pilot setup

Creating an assistant and configuring conversational search

- Access the
- Create your Assistant
- Configure conversational search
- More configurations
- Configure the base large language model
- Testing conversational search
- Troubleshooting

Ingesting client documents

Creating a stand-alone OpenSearch instance for document ingestion

- Install the Red Hat OpenShift command-line interface utility
- Prepare to ingest documents
 - Login to the OpenShift cluster from your local terminal
 - Create a working directory
 - Install IBM Certificate Manager on Red Hat OpenShift
 - Install the watsonx Assistant for Z Operator (for OpenSearch)
 - Deploy required secrets and the custom bring-your-own-search (BYOSearch) resource
 - Verify all the required pods are running
 - Update your assistant with the new
- Troubleshooting

Installing and using zassist to ingest client documents

- Install the zassist utility
- Ingest client documentation using
- Verify the document that is ingested is now returned as a source file for a query.

- Adjusting the search behavior
- Guidance for the pilot environment

Adding skills to the assistant

Getting started with skills and actions

- Environments
 - Watsonx Orchestrate
 - Ansible Automation Platform and Wazi as a Service

Explore Ansible Automation Platform

- Access the

Importing skills from Ansible Automation Platform

- Import skills into your assistant

Publishing and deploying your assistant

Adding other integrations

Next steps

- Earn the badge

Welcome

Welcome to the IBM watsonx Assistant for Z for Technical Sales Level 4 Lab Guide (lab guide). The lab guide is part of the IBM watsonx Assistant for Z for Technical Sales Level 4 learning plan for IBM and Business Partner Technical Sales and related badge. The learning plan is intended to teach technical sellers and Business Partners how to conduct a proof of experience (PoX) for a client.



In a fast-paced world, things change

The products and services may appear differently than what is shown in the lab guide. This can occur if the product or service is updated with a new version.

Responses generated by IBM watsonx Assistant for Z are likely to change over time. The responses you see when you run the queries in this section may differ from the screen images captured in the lab guide.

This lab guide covers the setup, configuration, and usage of watsonx Assistant for Z. This lab guide uses the [IBM watsonx Assistant for Z Velocity collection](#) and the 3 Velocity Pilot lab environments in IBM Technology Zone (ITZ).

The lab guide also enables dedicated lab environments for customized client Proof of Experiences (PoX) and demonstrations. If you are preparing for an actual pilot engagement, refer to the [Pilot Scoping Guide for watsonx Assistant for Z](#) for additional information.

The lab guide provides guidance to:

- Provisioning the lab environments
- Creating an assistant and configuring conversational search
- Configuring assistant settings
- Testing conversational search
- Deploying a dedicated instance of OpenSearch for client document ingestion (Optional)
- Importing skills for z/OS automations
- Connecting apps to assistants
- Creating assistant actions
- Creating skill flows and custom-built actions
- Importing pre-packaged z/OS skills
- Publishing and deploying your assistant

Support

Think something is down? Check the applicable status pages for any known issues like a site or service not available:

- [IBM Technology Zone](#)

For issues with provisioning the ITZ environment for this lab (for example, a failed reservation request due to insufficient quota capacity) open a ticket with ITZ support:

- Web: [IBM Technology Zone](#)
- Email: techzone.help@ibm.com

For issues related to specific steps found in the demonstration guide after the ITZ environment is provisioned, contact the authors:

- Slack:
- [#watsonx-assistant-z-technical](#) - IBM only
- [#wxo-practitioners](#) - IBM only - for questions that are related to the software as a service (SaaS) instance of watsonx Orchestrate
- Email: andrewj@us.ibm.com; maxwell.g.weiss@ibm.com

Business Partners, use the IBM Training live Chat Support service or other support methods that are found on the IBM Training portal [here](#).

Using the demonstration guide

Use these helpful tips to take full advantage of the IBM watsonx Assistant for Z for Technical Sales Level 4 Lab Guide.

🖨 Printing the demonstration guide

⚠ Printed or saved copies can be out of date

The IBM watsonx Assistant for Z for Technical Sales Level 4 Lab Guide changes regularly to match the IBM watsonx Assistant for Z offering and associated ITZ environment. Printed or saved copies of the demonstration guide can become out-of-date quickly and result in failed steps.

A ready-to-print PDF version of the IBM watsonx Assistant for Z for Technical Sales Level 4 Lab Guide is [here](#).

🖼 Viewing images

Images in the demonstration guide can be enlarged by clicking on the image. Press the **Esc** key or click the **X** to dismiss the enlarged image.

Reserve the IBM Technology Zone environment

IBM watsonx Assistant for Z for Technical Sales Level 4 Lab Guide

Welcome

Reserve the IBM Technology Zone environment

Pilot setup

Next steps

2. Click **Reserve now**.
The **Reserve now** option creates a reservation for immediate use. Optionally, schedule the reservation for a later date, like when you will be at your client's office.



3. Complete the reservation request and click **Submit**.
The first two reservations will be similar to the first image below and have fields a-e that will need to be completed.

- a. Optionally, change the **Name** field for the reservation.
- b. Select the **Education** purpose tile.
- c. Enter a **Purpose description**.
- d. Select the region nearest your physical location in the **Preferred Geography** drop-down.
- e. The **End date and time** will be set to 2 days after the current date and time.
- f. Accept the IBM Technology Zone's terms and conditions and security policies.
- g. When satisfied with the parameters, click **Submit**.



Image highlighting

In some images, the following styles of highlighting are used:

- **Solid highlight box:** This style of box highlights where to click, enter, or select an item.



- **Dash highlight box:** This style of box highlights one of two things: the path to follow to get to a specific location in the user interface, or areas to explore on your own.



Copying commands and prompts

Copying and pasting commands and prompts from this demonstration guide is easy and can eliminate typographical errors.

Click the highlighted copy icon and then use your operating system's paste function. For example, `[^ Ctrl]+[V]` or right-click and select Paste.

Prompt 1

What is the APF list in z/OS? Provide a detailed explanation.





Acronyms and terminology



IBM employees, and the tech industry in general, enjoy using acronyms. In the demonstration guide, most acronyms will appear with a dashed underline. Hover over the acronym to learn its meaning. A question mark (?) icon will first appear and after a second the tool tip with the acronym's meaning is displayed. Try it here: LPAR.

Guidance for delivering a demonstration

IBM watsonx Assistant for Z

Welcome

Reserve the IBM Technology Zone environment

Guidance for delivering a demonstration

Demonstration scenarios >

Next steps

Specific guidance for IBM watsonx Assistant for Z and the ITZ environment

Following the scripts provided in the IBM watsonx Assistant for Z for Technical Sales Level 3 Demonstration Guide will help guarantee a successful demonstration. Use these tips to help insure success with IBM watsonx Assistant for Z and the ITZ environment:

- Follow the scripts in the IBM watsonx Assistant for Z for Technical Sales Level 3 Demonstration Guide for the automations and skills to execute as expected.

Table of contents

General demonstration guidance

Specific guidance for IBM watsonx Assistant for Z and the ITZ environment



The Lab Guide table of contents

This **Demonstration Guide** uses a responsive browser-based interface to ensure a pages are usable on various devices with different screen sizes. The Demonstration Guide table of contents may be displayed as highlighted in the green dashed box in this image:

The screenshot shows a browser window for the "IBM watsonx Assistant for Z for Technical Sales Level 3 Demonstration Guide". The title bar includes the course name, a refresh icon, and a search bar. The main content area has a green dashed box highlighting the "Table of Contents" sidebar on the left. The sidebar lists several sections: "IBM watsonx Assistant for Z for Technical Sales Level 3 Demonstration Guide", "Welcome", "Guidance for delivering a demonstration", "Reserve the IBM Technology Zone environment", "Demonstration scenarios", "Introduction to scenarios", "Scenario 1: Authorized Program Facility", "Scenario 2: Certificate renewal", "Scenario 3: Db2 versioning", "Scenario 4: Initial Program Load on Z", "Additional IBM Z related prompts", "Summary", "Next steps", "Printing the guide", and "Instructions". The main content area displays the "Welcome" page, which includes an introduction to the course, information about end-user skills, and a note about Velocity Pilot information.

However, if the browser window is sized smaller, the table of contents can be accessed by clicking the main menu icon (≡):

The screenshot shows the same browser window after the main menu icon (≡) was clicked, causing the sidebar to collapse and the table of contents to disappear. The main content area now displays the "Welcome" page, which includes an introduction to the course, information about end-user skills, and a note about Velocity Pilot information.

Click the main menu icon (≡) to expand the table of contents.

Continue to the [Reserve the IBM Technology Zone environments](#) section to begin the journey to obtain the IBM watsonx Assistant for Z Technical Sales Advanced badge.

IBM Technology Zone environment

To enable sellers to both learn how to deliver client pilots of IBM Watsonx Assistant for Z, multiple environments are available in IBM Technology Zone (ITZ). The environments that are used for the Watsonx Assistant for Z Velocity lab environment can be found in the [IBM Watsonx Assistant for Z](#) collection and consist of:

- **Watsonx Assistant for Z lab – Watsonx Orchestrate:** a dedicated environment with Watsonx Orchestrate on IBM Cloud that is used for creating and configuring the assistant, configuring conversational search, importing skills, and configuring actions.
- **Ansible Automation Platform (AAP) & z/OS:** a pre-configured instance of AAP and Wazi z/OS. The environment includes preinstalled Ansible playbooks that can be imported as skills within Watsonx Orchestrate and connected to your assistant. The environment provides preinstalled templates for various use cases that are covered in a later section. Learn more about AAP [here](#). Learn more about Wazi, [here](#).
- **Single Node OpenShift with NFS storage:** used to provision a single-node Red Hat OpenShift cluster (SNO) on IBM Cloud. The cluster is used to install a dedicated instance of [OpenSearch](#) for Watsonx Assistant for Z. This environment enables ingestion of client-supplied documents.



All activities in this lab guide are required.

While all 3 environments may not be required for every client pilot, to complete the Level 4 learning plan and earn the IBM Watsonx Assistant for Z Technical Sales Advanced badge, you must provision all three ITZ environments and complete all sections in the lab guide. **Any statements in the ITZ collection regarding optional environments or tasks should be ignored.**

Follow the instructions in to create new reservation requests, extend the reservations, and access the ITZ demonstration environments. Provisioning the SNO environment in ITZ can take several hours, while the other two environments typically provisioning in under 30 minutes.

Create a reservation request

1. Click each of the links that follow to open a browser to the reservation pages of the **IBM Watsonx Assistant for Z** ITZ environments.

You may be asked to authenticate to IBM Technology Zone

The steps to authenticate to ITZ are not detailed here as they may vary between users.

[Watsonx Assistant for Z lab – Watsonx Orchestrate - reservation page](#)

[Ansible Automation Platform \(AAP\) & z/OS - reservation page](#)

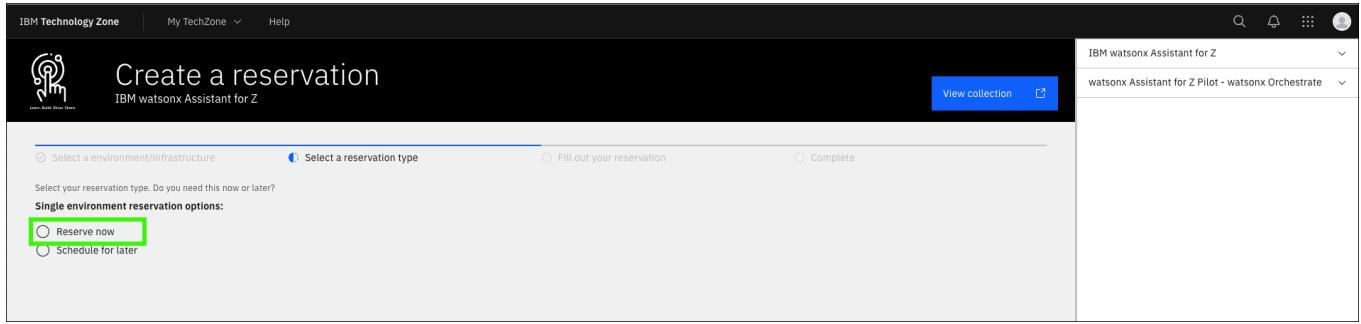
Single Node OpenShift with NFS storage - reservation page

 The following images are for 1 of the 3 environments

Be sure to follow these steps to create a reservation in ITZ for all three environments!

2. Click **Reserve now**.

The **Reserve now** option creates a reservation for immediate use. Optionally, schedule the reservation for a later date, for example, when you are at your client's office to start a pilot.



The screenshot shows the 'Create a reservation' interface. At the top, there are tabs for 'Select a environment/infrastructure', 'Select a reservation type', 'Fill out your reservation', and 'Complete'. Below these tabs, a note says 'Select your reservation type. Do you need this now or later?'. Underneath, it says 'Single environment reservation options:' and shows two radio buttons: 'Reserve now' (which is selected and highlighted with a green border) and 'Schedule for later'.

3. Complete the reservation request and click **Submit**.

The first two reservations are similar to the first image and have fields **a-e** that need to be completed.

- a.** Name: specify a name for the reservation (optional).
- b.** Purpose: select the **Education** purpose tile.



For actual client pilots...

For client pilots, set the **Purpose** field in the reservation to **Pilot** and provide an opportunity number to receive a longer reservation.

- c.** Purpose description: enter a description, for example: Level 4 education.
- d.** Preferred geography: select the region nearest to your physical location for improved performance and reduced network latency.
- e.** End date and time: the initial maximum will be set to 2 days after the current date and time. Instructions follow to extend the reservation end date.
- f.** Accept the IBM Technology Zone's terms and conditions and security policies.
- g.** When satisfied with the parameters, click **Submit**.

a Name
watsonx Assistant for Z Pilot - watsonx Orchestrate

b Purpose
Demo
Deliver a client specific demonstration based on discovery with the client and align to the identified architecture. Automatically captures a Technical Sales Activity in IBM Sales Cloud on the Opportunity code provided.

c Sales Opportunity number
Enter an opportunity number
Providing an [IBM Sales Cloud opportunity ID](#), [Contract Relationship ID](#), or a [Project Work ID](#) will allow you to extend your reservation date.

d Preferred Geography
ITZ-watsonx-2 - AMERICAS - us-south region - dat10 datacenter

e End date and time
Select a date: 11/01/2024
Select a time: 11:51 AM America/Chicago

f Notes
Enter any notes you would like to attach to this reservation

g Submit

In addition to the preceding fields, the reservation for the **Single Node OpenShift with NFS storage** has these additional fields:

- h.** OCP/Kubernetes cluster network: leave the default setting of **10.128.0.0/14**.
- i.** Enable FIPS security: leave the default setting of **No**. Learn more about the Federal Information Processing Standards (FIPS) [here](#).
- j.** Master single node flavor: select **16 vCPU x 64 GB - 300 GB ephemeral storage**.
- k.** OpenShift version: select **4.14**.
- l.** OCP/Kubernetes service network: leave the default setting of **172.30.0.0/16**.

h OCP/Kubernetes Cluster Network
10.128.0.0/14

i Enable FIPS Security
No

j Master Single Node Flavor
16 vCPU x 64 GB - 300 GB ephemeral storage

k OpenShift Version
4.14

l OCP/Kubernetes Service Network
172.30.0.0/16

f Notes
Enter any notes you would like to attach to this reservation

g I agree to IBM Technology Zone's [Terms & Conditions](#) and [End User Security Policies](#)

h Submit

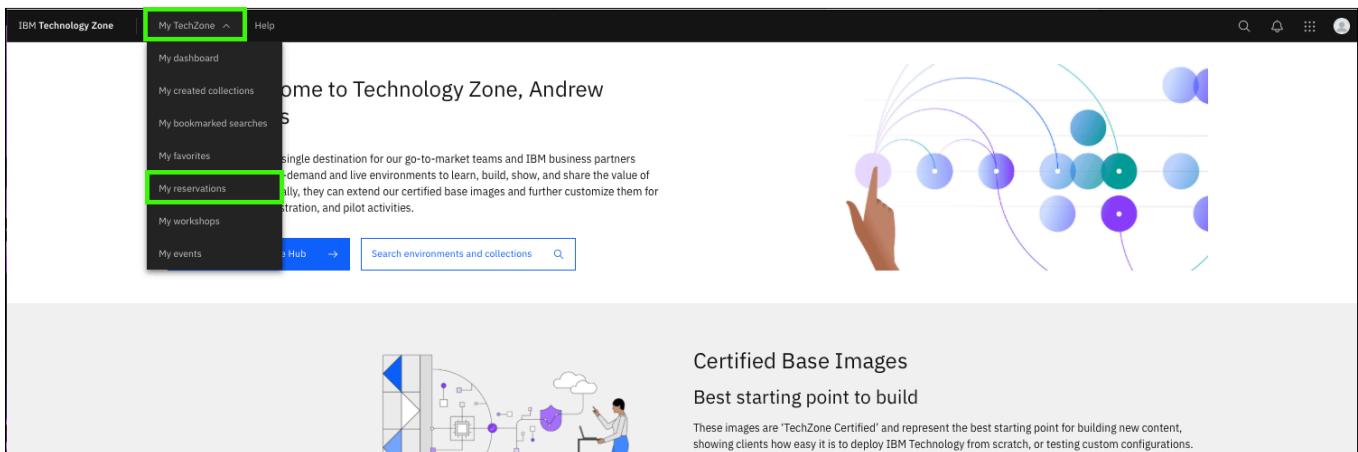
During the provisioning process, multiple emails are sent to you from ITZ as the provisioning process runs. One email states the reservation is provisioning and the other email states that the environment is **Ready**.

In rare cases, the provisioning process may fail. If you receive an email stating the reservation failed, try again by repeating Steps 1-3 for the environment that failed to provision. In addition, review the [Troubleshooting](#) section that follows. If issues continue, open an ITZ support ticket by using the methods that are mentioned in the [Support](#) section.

Extend the reservation

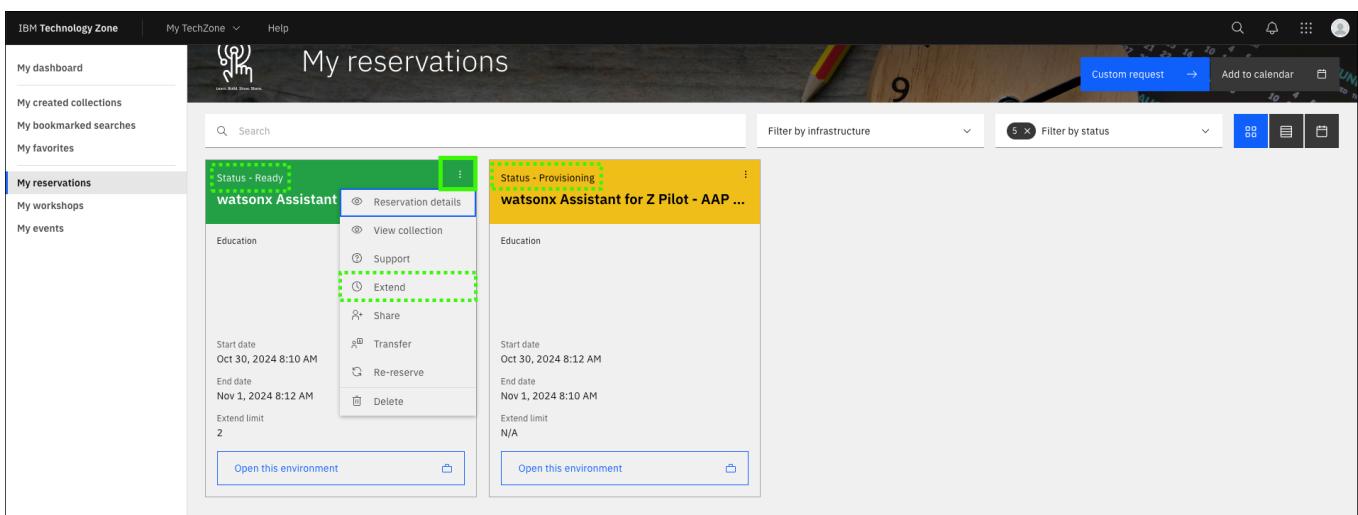
When the reservations are in the **Ready** state, you can extend each reservation to a total of 6 days.

4. In the IBM Technology Zone portal, expand **My TechZone** and select **My Reservations**.



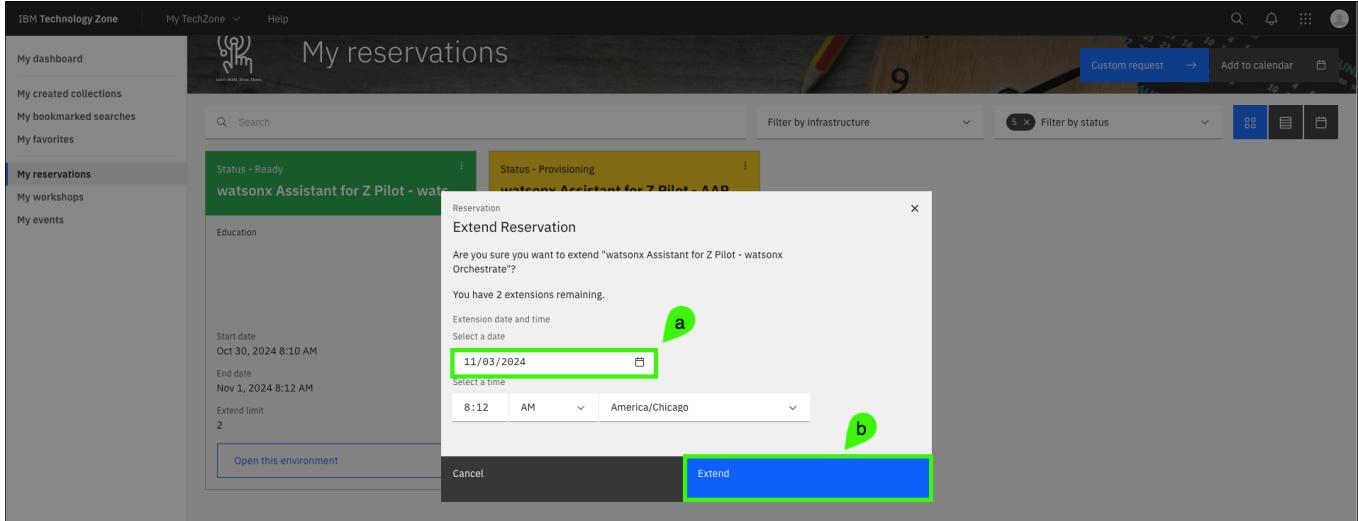
The screenshot shows the IBM Technology Zone homepage. The navigation bar at the top has 'IBM Technology Zone' on the left, 'My TechZone' with a dropdown arrow in the middle, and 'Help' on the right. A green box highlights the 'My TechZone' dropdown. Below it, a sidebar menu lists: 'My dashboard', 'My created collections', 'My bookmarked searches', 'My favorites', 'My reservations' (which is also highlighted with a green box), 'My workshops', and 'My events'. To the right of the sidebar is a main content area with a welcome message 'Welcome to Technology Zone, Andrew' and a graphic of a hand pointing at a network of nodes. Below this is a section titled 'Certified Base Images' with a sub-section 'Best starting point to build' and a note about TechZone Certified images.

5. Click the overflow icon (≡) on the reservation tile and select **Extend**.



The screenshot shows the 'My reservations' page. The left sidebar is identical to the previous screenshot, with 'My reservations' selected. The main area displays two reservation tiles. The first tile, for 'watsonx Assistant', is green and labeled 'Status - Ready'. Its context menu (accessed via an overflow icon) includes options like 'Reservation details', 'View collection', 'Support', 'Extend' (which is highlighted with a green box), 'Share', 'Transfer', 'Re-reserve', and 'Delete'. The second tile, for 'watsonx Assistant for Z Pilot - AAP ...', is orange and labeled 'Status - Provisioning'. It shows a start date of Oct 30, 2024 8:10 AM and an end date of Nov 1, 2024 8:10 AM. The 'Extend' option is also present in its context menu. Both tiles have an 'Open this environment' button at the bottom.

6. Click the **Select a date** option, (a) specify the date to extend to, and then (b) click **Extend**.

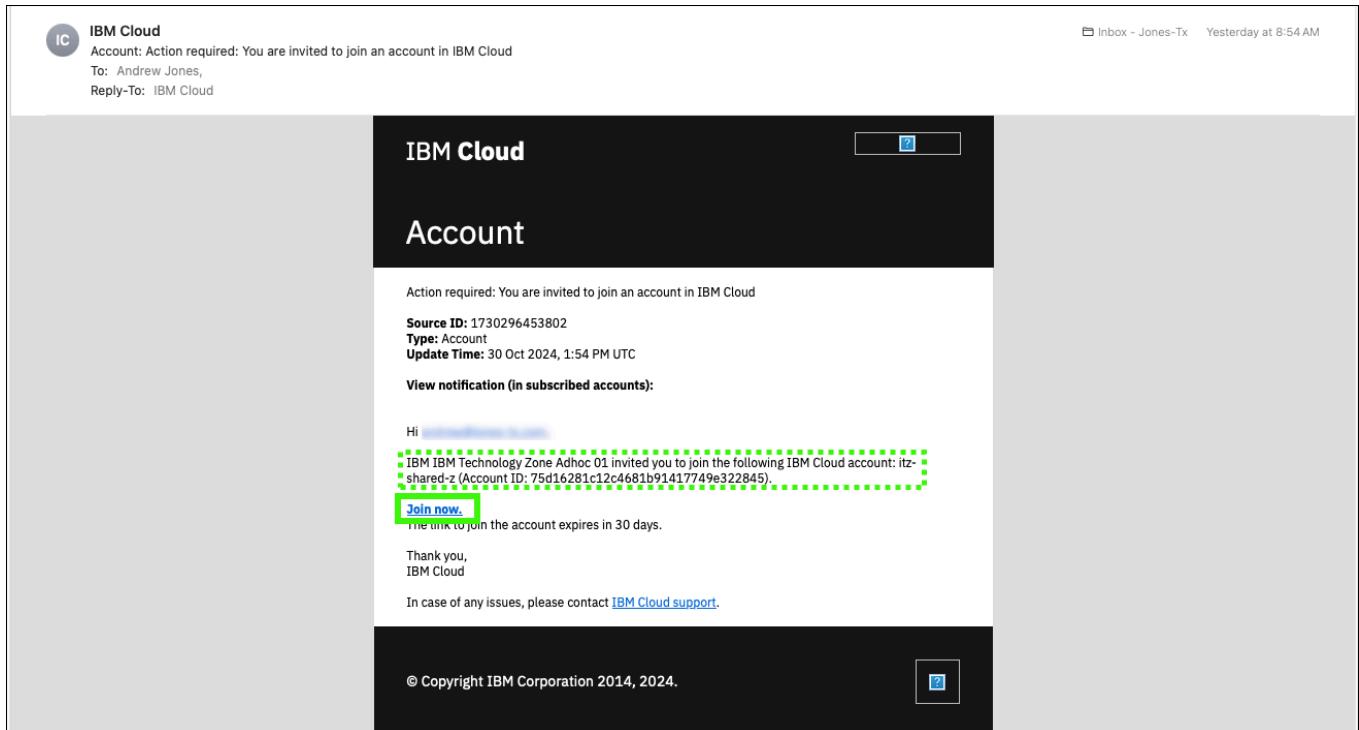


If you anticipate needing more time, repeat Steps 5 and 6 to extend the reservation to the maximum of 6 days. Repeat these steps for the other two reservations.

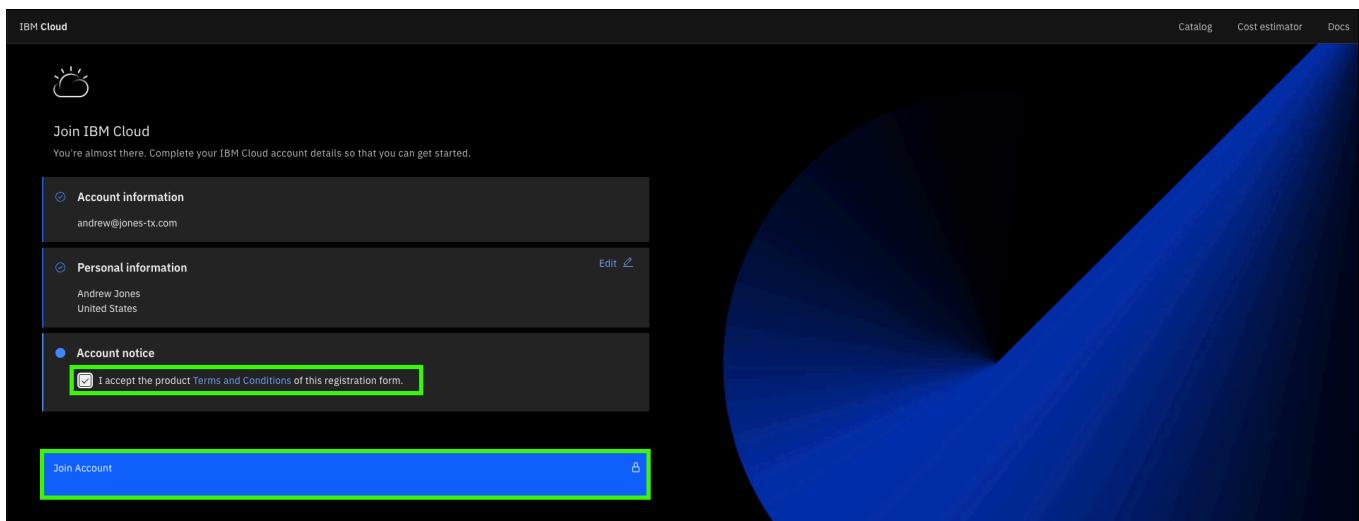
Join the ITZ IBM Cloud accounts

Both the **watsonx Assistant for Z lab – watsonx Orchestrate** and the **Ansible Automation Platform (AAP) & z/OS** environments add you to an IBM Cloud account while your reservation is active. During the provisioning process of these **ITZ** environments, you receive two emails from IBM Cloud. To access the environment, you must first accept the invitations to join both of the IBM Cloud accounts.

7. Open the emails from **IBM Cloud** and click the **Join now** links.



8. In the **Join IBM Cloud** browser windows that open, select the **I accept the product Terms and Conditions of this registration form**, and then click **Join Account****.



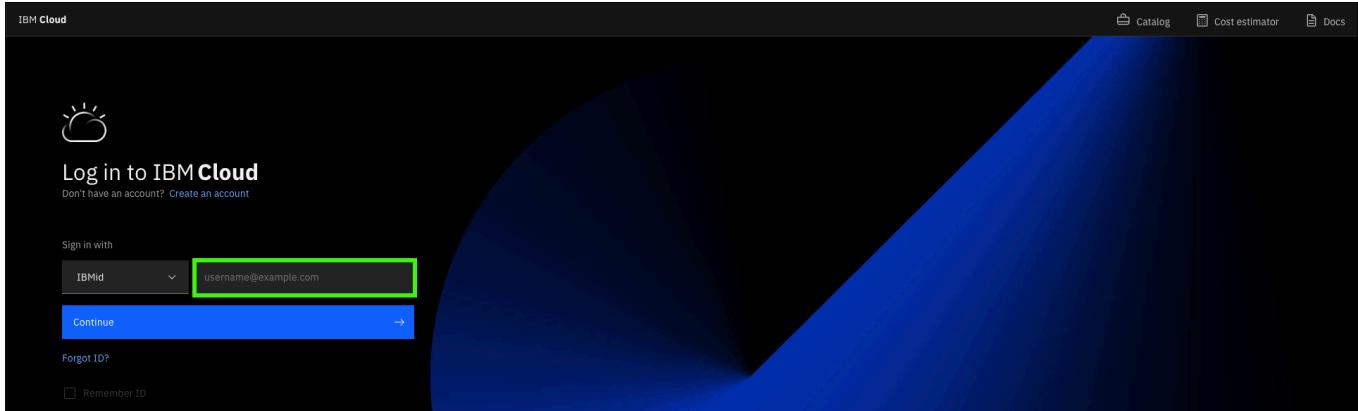
Repeat steps 7 and 8 for the second invitation.

After joining both accounts, verify that both accounts appear in your available account list in the IBM Cloud portal.

- Click the following link to open a browser to the IBM Cloud portal.

[IBM Cloud portal](#)

- Follow the directions to complete the authentication to IBM Cloud using the same email address you used to login to ITZ. The login steps very depending on any two-factor authentication methods enabled.



- Click the **account** menu and verify access to the two IBM Cloud accounts listed in your ITZ reservations.



These accounts may change within ITZ.

Over time, the accounts may change for the environments. The accounts names should align with the accounts named in the invitation emails you received.

 Does your IBM Cloud portal view look different?

If your IBM Cloud portal looks different from the images above, it could be because the IBM Cloud portal has done through a design change, or your browser window is set to smaller size. Instead of the current selected account appearing in the top menu, you may see this **change account** icon:  Click this icon to view the list of accounts you can access.



Accessing the environments

Each reservation provides access to its respective environment. Details for accessing each environment are provided in the **Pilot setup** sections that follow in the lab guide.

After all three reservations are in the **Ready** state and you accepted the invitations to the IBM Cloud accounts, proceed to the next section to complete the pilot setup.

Troubleshooting

 If your reservation for the Single Node OpenShift environment fails...

If your reservation for the Single Node OpenShift environment fails, try selecting one of the **eu-gb region** options as the **Preferred Geography**.

Pilot setup

Creating an assistant and configuring conversational search

To create your watsonx Assistant for Z, you use [watsonx Orchestrate](#) to create the assistant and configure conversational search. You can configure your assistant to use conversational search by using a hosted [OpenSearch](#) instance. The pre-configured instance in IBM Technology Zone (ITZ) has over 220 knowledge sources and supports Retrieval Augmented Generation (RAG). The large language model (LLM) providing the conversational AI augments this knowledge based on IBM Z documentation. All of these elements create IBM Z context-aware responses to queries with its content-grounded knowledge.

Access the ITZ IBM Cloud account for the watsonx Assistant for Z Pilot environment

1. In the IBM Technology Zone portal, expand **My TechZone** and select **My Reservations**, or click the following link.

ITZ My reservations

2. Click the **watsonx Assistant for Z Pilot - watsonx Orchestrate** tile.

3. Record the ITZ IBM Cloud account name associated with the reservation.

Purpose

- Purpose
- Education
- Opportunity Product(s)
- Customer(s)

Environment

Reservation ID	Type
672a091a8f85062f891e081	IBM Cloud
Request method	Transaction ID
watsonx-orchestrate	115897c9-58a1-4f17-af9c-b16dc2a97590
ITZ-WATSONX-036	Geo
Region	americas
us-south	Datacenter
Customer data	dal10
false	Environment
Idle runtime limit	watsonx-orchestrate-wusdf
10800	Timeout action

Reservation Details

IBM Cloud Login
<https://cloud.ibm.com/resources>

4. Click the IBM Cloud Login link.

Purpose

- Purpose
- Education
- Opportunity Product(s)
- Customer(s)

Environment

Reservation ID	Type
672a091a8f85062f891e081	IBM Cloud
Request method	Transaction ID
watsonx-orchestrate	115897c9-58a1-4f17-af9c-b16dc2a97590
Cloud Account	Geo
ITZ-WATSONX-036	americas
Region	Datacenter
us-south	dal10
Customer data	Environment
false	watsonx-orchestrate-wusdf
Idle runtime limit	Timeout action
10800	

Reservation Details

IBM Cloud Login
<https://cloud.ibm.com/resources>



Steps to authenticate to IBM Cloud are not illustrated here.

You may need to authenticate to IBM Cloud after clicking the link. These steps are not shown here as they may vary by individual.

5. Verify that the current IBM Cloud account is the same as the account name recorded in step 3, if not the same, switch to the proper account.

Note: the formatting of the name can appear differently than what is shown in the ITZ reservation.

If the proper account is not listed, click the account drop down and select the proper account. Note, if your browser window is narrow, the account drop down can be depicted with the switch account icon ().

Create your Assistant

6. Click the **resources** icon ().

7. Expand the **AI / Machine Learning** section and click the **watsonx Orchestrate** instance listed (the instance name is different than shown in the following image).

Resource list

Name	Group	Location	Product	Status	Tags
itwxo-2700039nft-erspw	watsonx-orchestrate-erspw	Dallas	watsonx Orchestrate	Active	-

8. Click Launch watsonx Orchestrate.

itwxo-2700039nft-erspw Active Add tags

Manage Service credentials

Start by launching the tool

Launch watsonx Orchestrate

Getting started tutorial

Credentials

API key: [Download](#) [Show credentials](#)

URL: <https://api.us-south.assistant-builder.watson.cloud.ibm.com/instances/a7675d8e-e885-489f>

Plan Essentials Plan

9. Click the AI assistant builder tile to start creating a new assistant.

Welcome, Andrew Jones!

Take productivity to the next level.

Try Skills in Chat →

AI assistant builder

Boost productivity and customer care by creating conversational experiences.

Skill studio

Build the skills your team needs to get their work done more quickly.

Automate how your business works

Standard plan | Skill studio

Learn more →

10. Enter a name and optional description for your assistant and click Next.

Welcome to AI assistant builder

Create your first assistant

Let's get your assistant up and running. Name your assistant, add a description, and choose a language. In following steps we'll gather more information, show you basic customizations, and give you a preview of what your assistant will look like.

Assistant name: Zeeves

Description (optional): Zeeves, the Jeeves for everything IBM Z.

Assistant language: English (US)

Next

11. Complete the **Personalize your assistant** form and click **Next**.

Explore the personalization options. In creating an assistant for a client pilot, consider specifying attributes that align with the client's business.

- Select **Web**.
- Select the industry of your choice.
- Select the role of your choice.
- Select the need of your choice.

Welcome to AI assistant builder

Personalize your assistant

Tell us where your assistant will live
Where do you plan on deploying your assistant?

a. Web

Tell us about yourself
Which industry do you work in?

b. Software

c. Developer

d. I want to automate common tasks in a natural way

This is what your customers will experience

watsonx Assistant

Do you have the Speed Demons in stock?

The Speed Demons are in stock at our Downtown and Northgate locations, which are both within 5 miles of you.

What size and color do you need?

I'm looking for a size 9 in white

Great news! The Speed Demons are available in white in a size 9.

You can purchase them for curbside pickup or we can ship them to you. Which would you prefer?

I'll pick them up! Ship them to me!

Type something...

Next

12. Complete the **Customize your chat UI** form and click **Next**.

Explore the customization options. When creating an assistant for a client pilot, consider specifying attributes that align with the client (for example, colors and logos).

Welcome to AI assistant builder

Create Personalize Customize Preview

Customize your chat UI

Update the style to match your brand and your website. You can change these settings later. A developer can also add more advanced styling changes with code. [Learn more](#)

Assistant's name as known by customers: **Zeeves**

Intended purpose: Standard: For virtual agents and customer support experiences. Carbon for AI: For use in internal IBM products.

Choose a theme: Light Dark

Primary color: #FFFFFF Secondary color: #30303D Chat header: User message bubble:

Accent color: #035AE9 Significant and interactive objects:

Size: The size of the web chat on this page will not change by updating these fields.

Width: 380px Height: 640px

IBM Watermark: Shade IBM Watermark: On: Off:

Streaming: Enable Streaming: Off:

Zeeves

Hi! I'm a virtual assistant.
How can I help you today?

Example: Find nearby location
Example: Check account balance
Example: See how I can help

Type something...

Powered by IBM watsonx

13. Preview your assistant and then click **Create**.

Welcome to AI assistant builder

Create Personalize Customize Preview

Preview your assistant

See what your assistant will look like as a chatbot on your website.

Certain settings do not work on this page and are disabled.

Sample website

Hi! I'm a virtual assistant.
How can I help you today?

Example: Find nearby location
Example: Check account balance
Example: See how I can help

Type something...

Powered by IBM watsonx

The assistant is now created.

The screenshot shows the IBM Watsonx Assistant AI assistant builder interface. At the top, there's a navigation bar with 'IBM Watsonx Orchestrate', 'AI assistant builder', 'Zeeves', and other options like 'What's new' and 'Documentation'. Below the navigation is a sidebar with sections like 'Assistant Builder Home', 'Zeeves' (last updated on October 31, 2024), and 'Enhance your assistant' which includes 'Build actions', 'Customize your greeting', 'Create a fallback plan', 'Set up search', 'Preview & debug', 'Customize web chat', 'Set up a channel', 'Set up live agent', and 'Publ assis'. A large central area displays the 'Assistant architecture' diagram. This diagram shows a flow from 'Default behavior' through 'Actions' (which includes 'Search', 'AI-guided', 'Skill-based', and 'Custom-built') and 'Fallback' to 'Live agent'. The 'Actions' section also lists 'General purpose' and 'Conversational search' with their respective configurations.

Configure conversational search

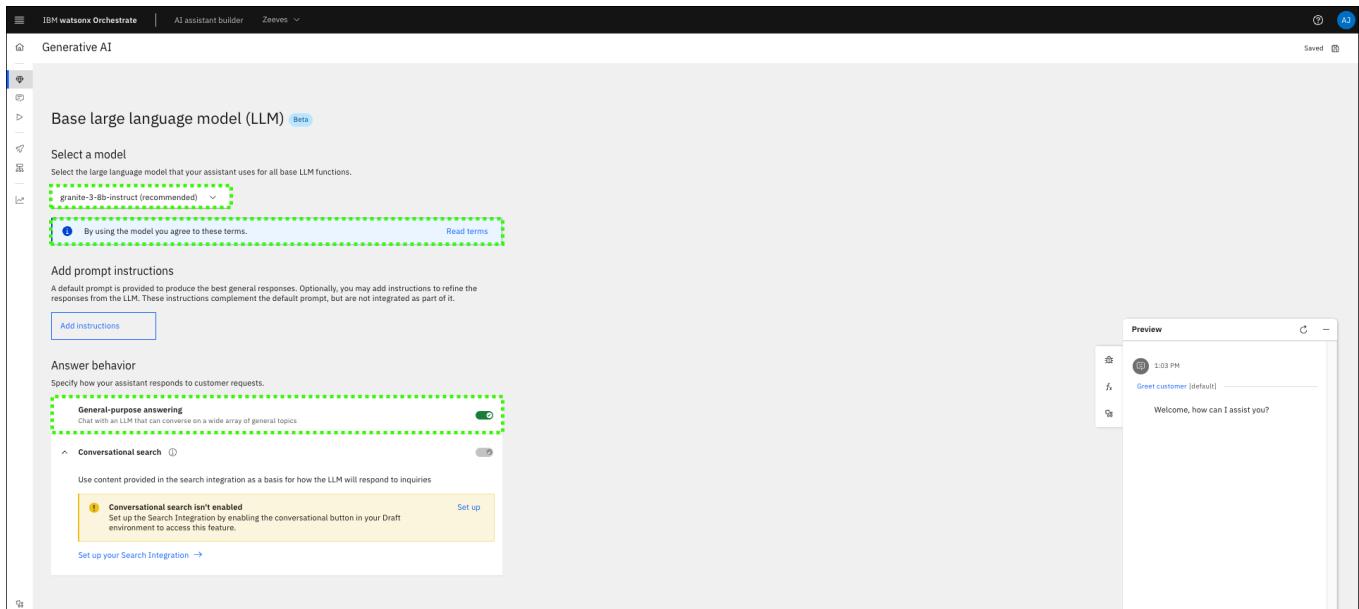
The next step will be to configure **conversational search** for your assistant that uses a hosted instance of OpenSearch.

- Click **Generative AI** menu item (💡) in the left navigation.

This screenshot is similar to the previous one but with a green highlight around the 'Generative AI' menu item in the left navigation bar. The rest of the interface, including the 'Assistant architecture' diagram, remains the same.

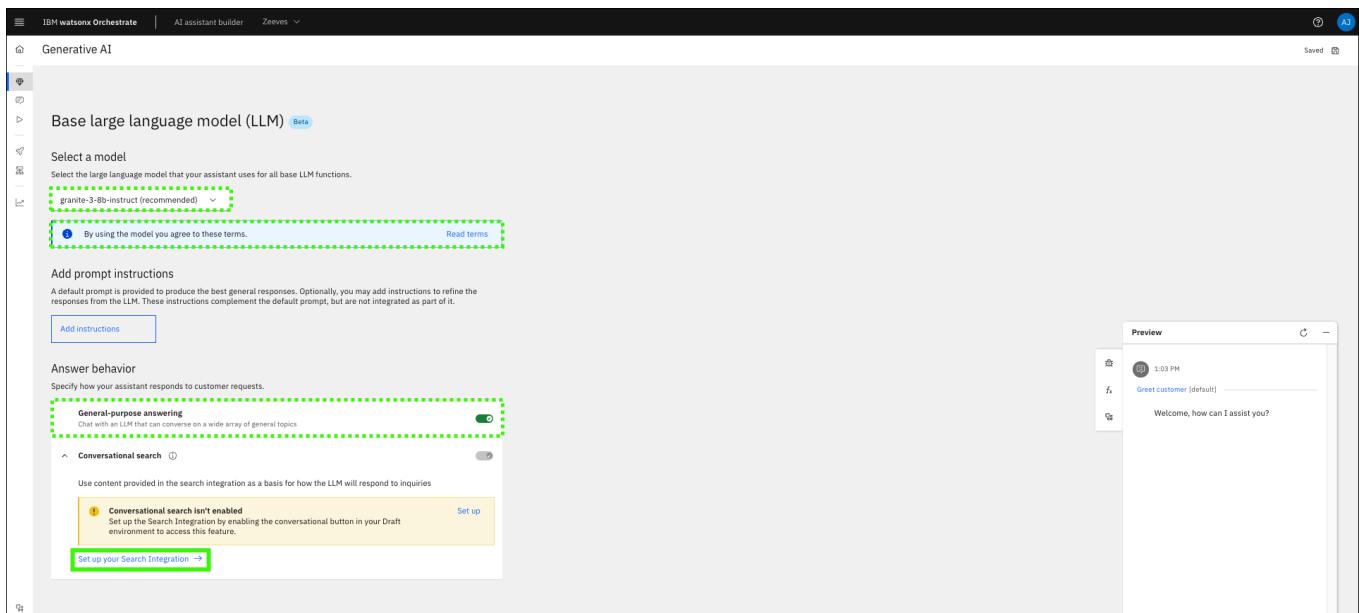
- Review the base large language model (LLM) settings.

Notice the other LLM models available. For most pilots, the **granite-3-8b-instruct** model is appropriate.



16. Click Set up your Search Integration.

By default, conversational search is not enabled when an assistant is created. Conversational search takes priority over general-purpose answering if both are enabled. Learn more about conversational search in watsonx [here](#).



17. Click Custom service.

18. Complete the **Custom service form and then click **Next**.**

- Select **By providing credentials**.
- Enter the following value in the **URL** field (use the copy icon to avoid typographical errors).

https://wxa4z-opensearch-wrapper-wxa4z-demo.wxo4z-opc-opensearch-clus-47e063e6a3ad1f71bf2e58f91c3b4c2e-0000.us-south.containers.appdomain.cloud/v1/query

- Select **None** in the **Choose an authentication type** drop-down list.

19. Enable **conversational search and then click **Save**.**

20. Update the conversational search **custom service settings based on your requirements.**

Note: the **Settings** page is divided into two sections in the following images to enhance the visibility of the screen captures. Learn more about these settings [here](#).

The following settings are proven to work well. You can experiment with these settings to see how they affect queries for your client's pilot.

a. Enable Conversational search.

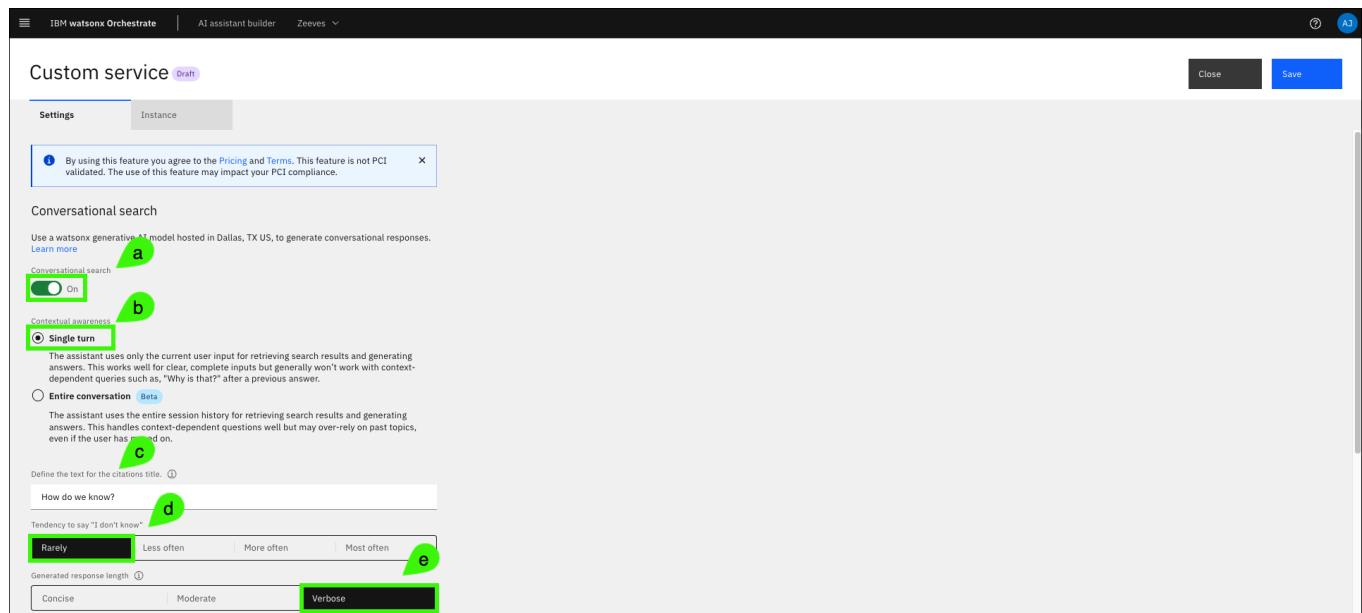
b. Select Single turn. Enabling multi-turn conversation (by selecting Entire conversation) is not yet supported for the solution on-premises. Be mindful in using this option and help ensure that the client understands what is supported in the solution.

c. Specify the text appears to expand the list of citations in the assistant (except web chat client).

d. Select Rarely for the tendency to say "I don't know" setting.

e. Select Verbose for the generated response length. This setting affects the average response length.

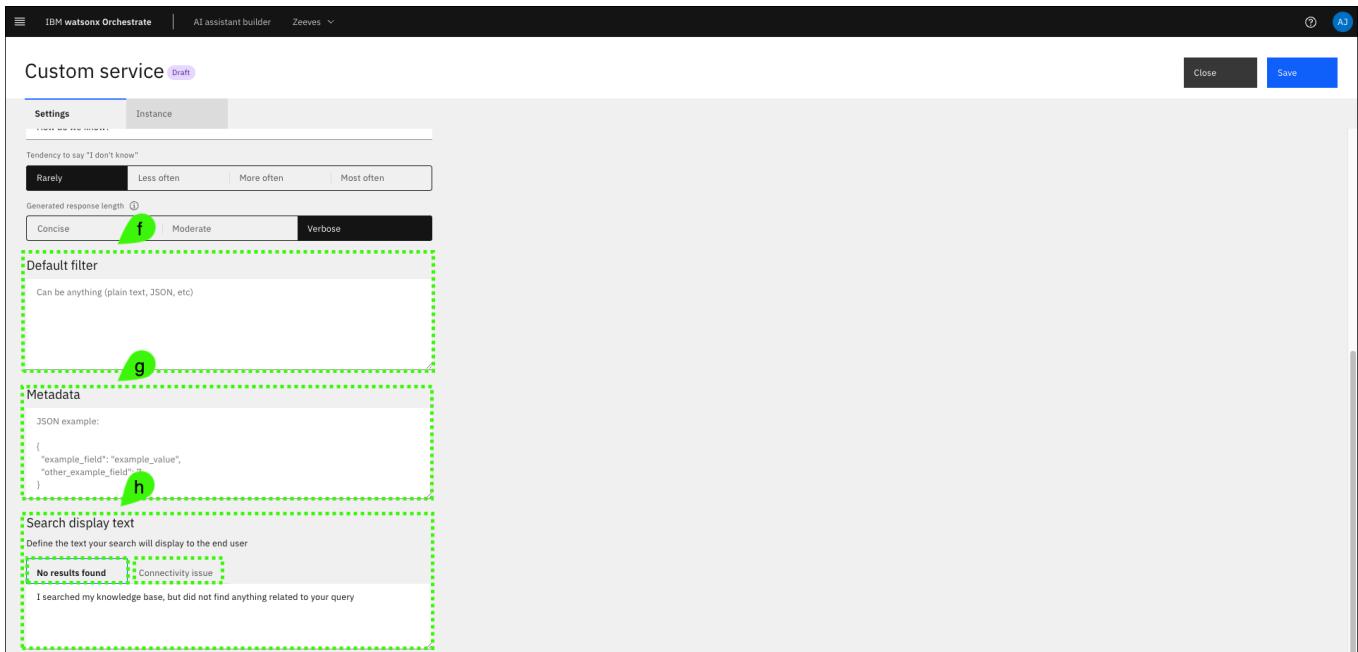
Depending on user input, variations from the selected length can occur.



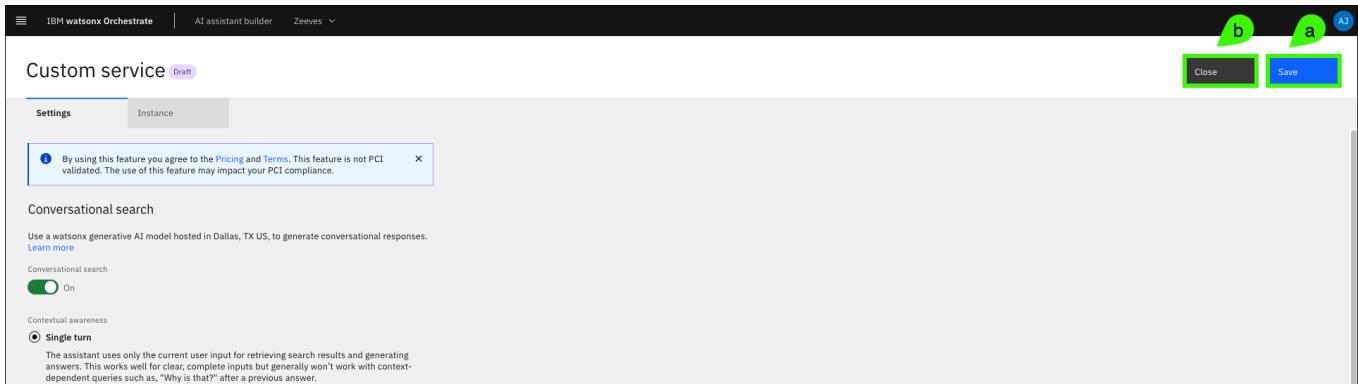
f. Leave the Default filter field empty.

g. The Metadata field provides a way to adjust your assistant's behavior during conversational search for your OpenSearch instance. This option is explored in detail in the [Installing and using zassist to ingest client documents](#). Leave the field empty for now.

h. The Search display text options specify the default text displayed when no results are found or when connectivity issues to the backend search service occur. You can keep the defaults or customize.



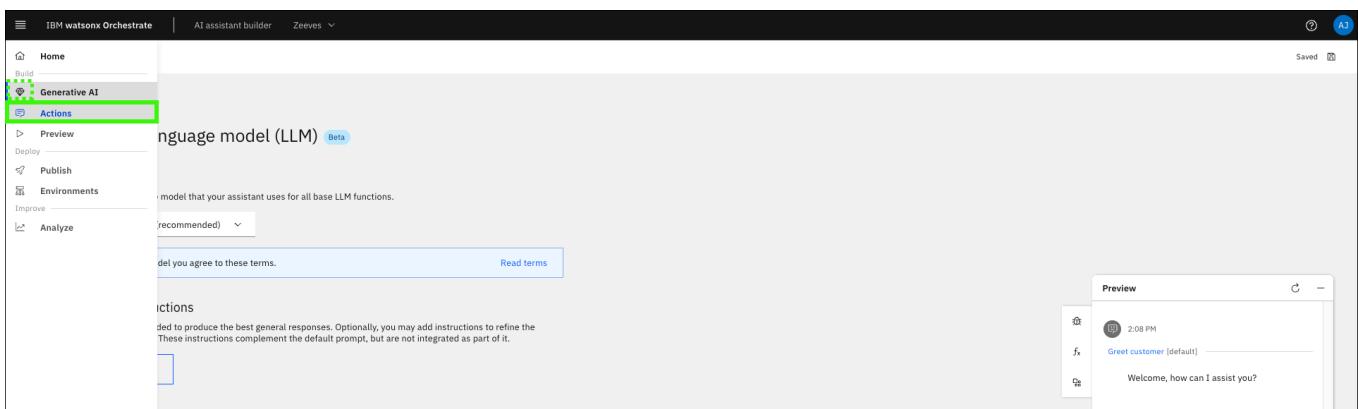
21. Click (a) **Save** and then click (b) **Close**.



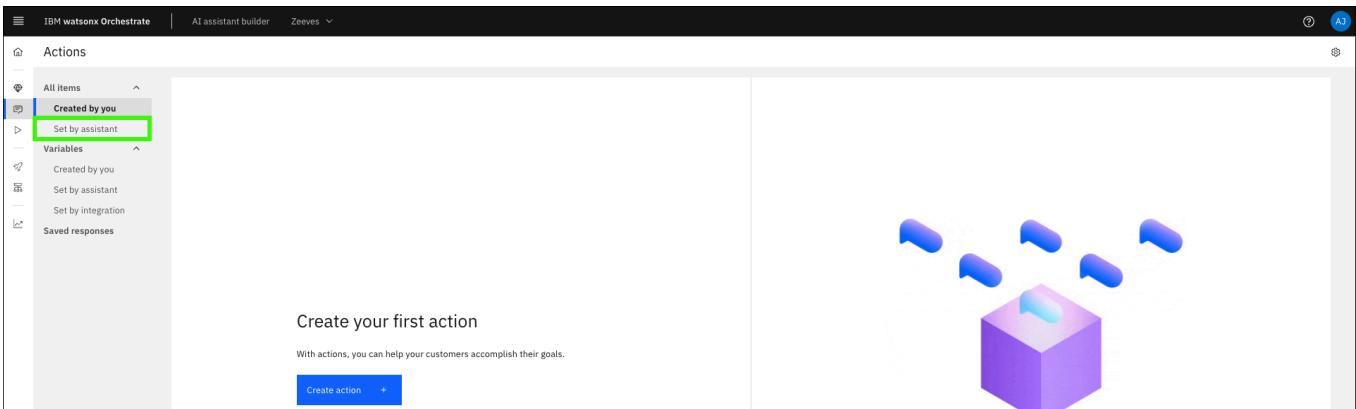
More configurations

After you save and close the **Conversational search** configuration page, a few more configurations are needed to get the best experience from your conversational chat. Details on these settings are available [here](#).

22. Hover over the **Generative AI** icon (💡) in the left navigation and click **Actions**.



23. Click **Set by assistant** under the **All items** menu.



24. Click no matches.

25. Click Step 1 under Conversation steps.

26. Select (a) without conditions in the Is taken drop-down menu and then click (b) Clear conditions.

Note, the **Is taken** value does not change from **with conditions** after selecting **without conditions**.

27. Delete the default text in the Assistant says entry field.

The screenshot shows the AI assistant builder interface with the following configuration:

- Customer starts with:** Example: Can I have a sandwich?
- Conversation steps:**
 - Step 1: Is taken without conditions. Assistant says: For example: Please select from the following options. Define customer response.
 - And then: End the action.

28. Select **Search for the answer** in the **And then** drop-down menu.

The screenshot shows the AI assistant builder interface with the following configuration:

- Customer starts with:** Example: Can I have a sandwich?
- Conversation steps:**
 - Step 1: Is taken without conditions. Assistant says: For example: Please select from the following options. Define customer response.
 - And then: End the action. The dropdown menu is open, showing the following options:
 - Continue to next step
 - Re-ask previous step(s)
 - Go to a subaction
 - Use an extension
 - Search for the answer** (highlighted with a green box)
 - Connect to agent
 - End the action

29. Click **Edit settings**.

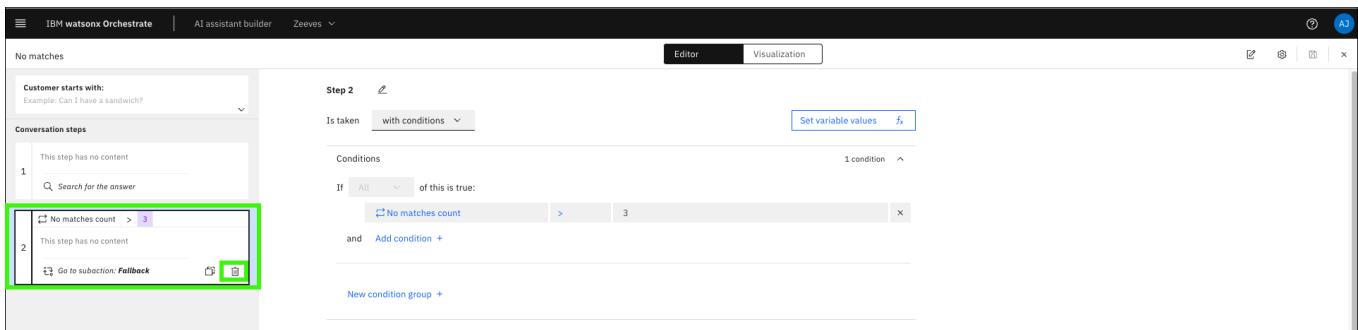
The screenshot shows the AI assistant builder interface with the following configuration:

- Customer starts with:** Example: Can I have a sandwich?
- Conversation steps:**
 - Step 1: Is taken without conditions. Assistant says: For example: Please select from the following options. Define customer response.
 - And then: **Q. Search for the answer**. The dropdown menu is open, showing the following options:
 - Custom query: None (Customer's message will be the query)
 - Custom filter: None (Optional)
 - Edit settings** (highlighted with a green box)

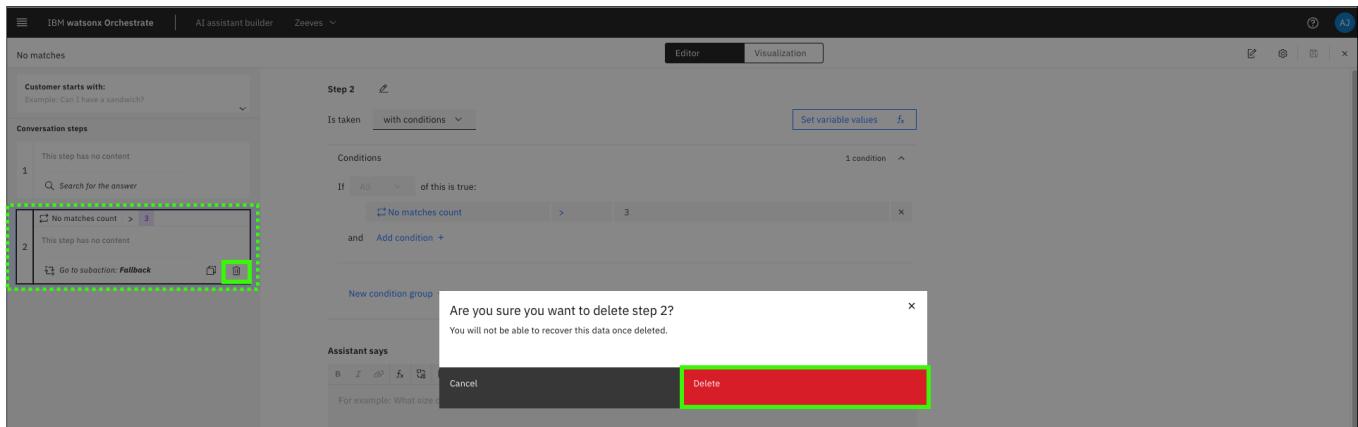
30. Select **End the actions after returning results** and then click **Apply**.



31. Select Step 2 (No matches count) under **Conversation steps** and click the delete icon (>Delete).



32. Click **Delete** in the confirmation dialog to delete step 2.



33. Click the x to close the **Editor** window.



34. Click Fallback in the Actions table.

Name	Last edited	Examples Count	Status
Greet customer	3 hours ago	1	Green
Trigger word detected	3 hours ago	1	Green
No matches	a minute ago	1	Green
Fallback	3 hours ago	1	Green

35. Delete all of the Conversation steps.

Note: the following image is edited. Only 5 steps are shown, but all 6 need to be deleted. You need to select each step individually, click the delete icon (trash), and confirm the deletion.

36. Verify that all Conversation steps are deleted and then click the x to close the Editor window.

37. Click the global settings (⚙️).

Name	Last edited	Examples Count	Status
Greet customer	4 days ago	1	Green
Trigger word detected	4 days ago	1	Green
No matches	4 days ago	1	Green
Fallback	4 days ago	1	Green

38. Click **No matches** under the **Conversation routing** tab.

The screenshot shows the 'Global settings' page with the 'Conversation routing' tab selected. In the 'No matches' section, there is a note about handling user input when no relevant options are found. A toggle switch labeled 'Enable disambiguation' is set to 'On'. Below it, a text box contains the placeholder 'Did you mean:'. A dropdown menu lists 'None of the above', 'Something else', and 'Connect to support'. To the right, there is an example of a 'Bank Bot' interaction where the user asks 'open an account' and the bot responds with 'Did you mean:' followed by a list of suggestions: 'Open a new savings account', 'Open a new checking account', 'I want to apply for a mortgage loan', and 'None of the above'.

39. Move the slider to **More often** (or select **More often** in the drop-down).

The setting helps ensure that actions are triggered less often unless the user's query specifically matches the action's input.

The screenshot shows the 'Global settings' page with the 'Conversation routing' tab selected. The 'No matches' section includes a note about triggering the 'No matches' action for unrecognized input. A slider is shown with a green box around its track, and the dropdown menu next to it has 'More often' selected.

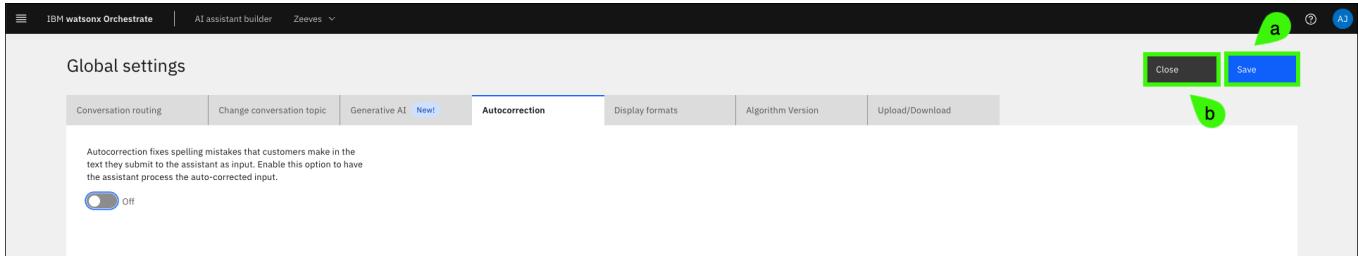
40. Click **Autocorrection**.

The screenshot shows the 'Global settings' page with the 'Autocorrection' tab selected. The 'Autocorrection' section contains a note about fixing spelling mistakes in customer input. A toggle switch labeled 'Off' is shown with a green box around it.

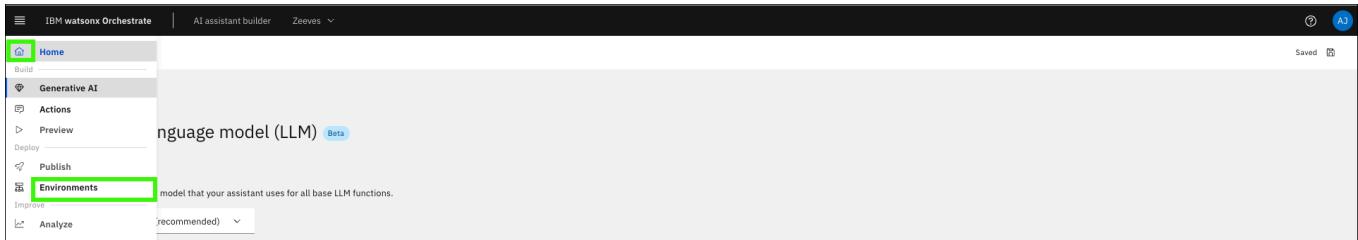
41. Click the **autocorrection** toggle to turn the feature **off**.

The screenshot shows the 'Global settings' page with the 'Autocorrection' tab selected. The 'Autocorrection' section contains a note about fixing spelling mistakes. A toggle switch labeled 'Off' is shown with a green box around it.

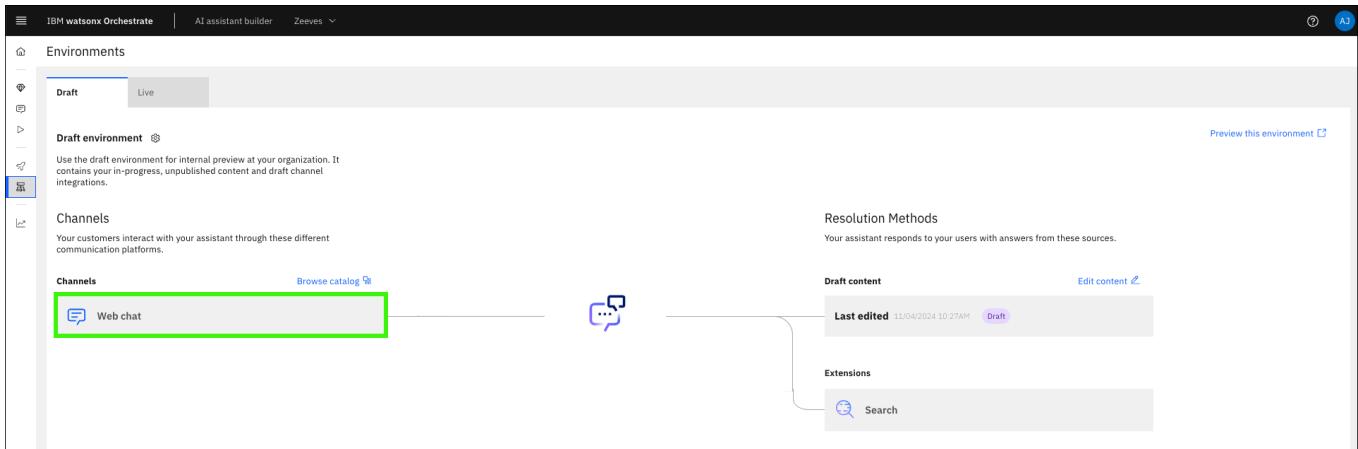
42. Click (a) **Save** and then (b) **Close**.



43. Hover over the **home** (🏠) and click **Environments**.



44. Click **Web chat**.



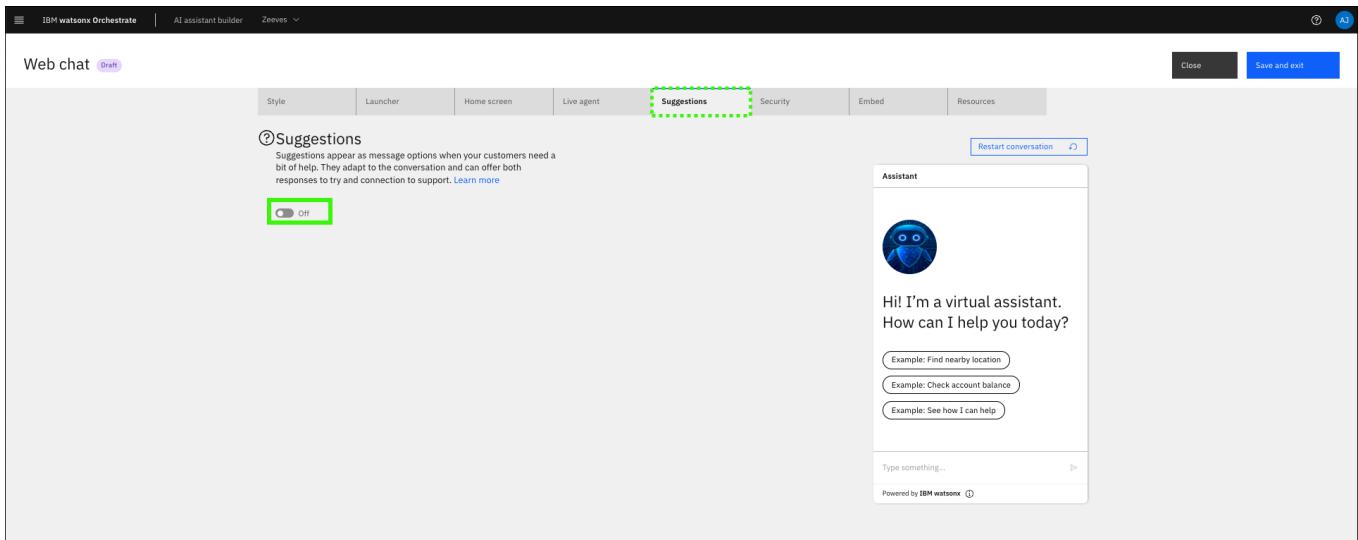
45. On the **Style** tab, click the **Streaming** toggle to enable streaming.

The screenshot shows the 'Style' tab of the AI assistant builder. On the left, there's a configuration panel for a 'Web chat' with various styling options like color themes, fonts, and sizes. On the right, a preview window shows a blue owl-like assistant with the message: 'Hi! I'm a virtual assistant. How can I help you today?'. Below the preview is a text input field and a 'Powered by IBM watsonx' logo.

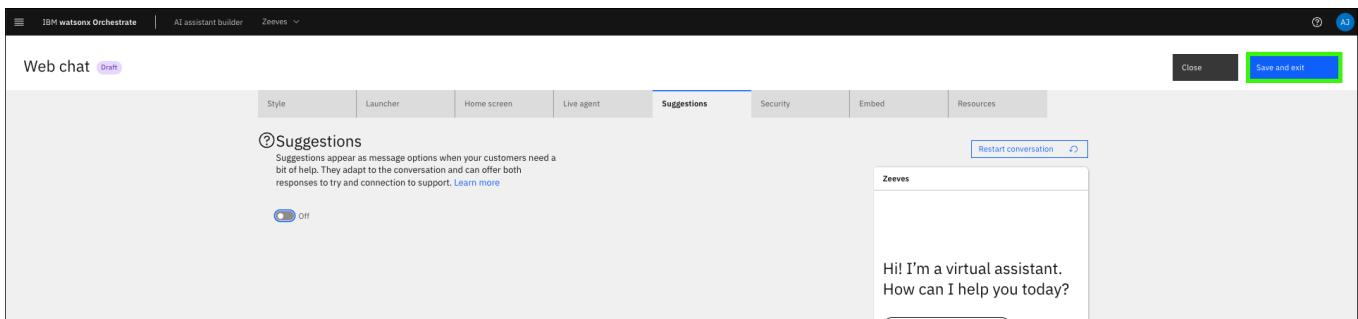
46. Click Suggestions.

This screenshot is identical to the previous one, but the 'Suggestions' toggle in the configuration panel is now turned on (green). In the preview window, the assistant's message has changed to 'Zeeves'.

47. Click the Suggestions toggle to turn this feature off.



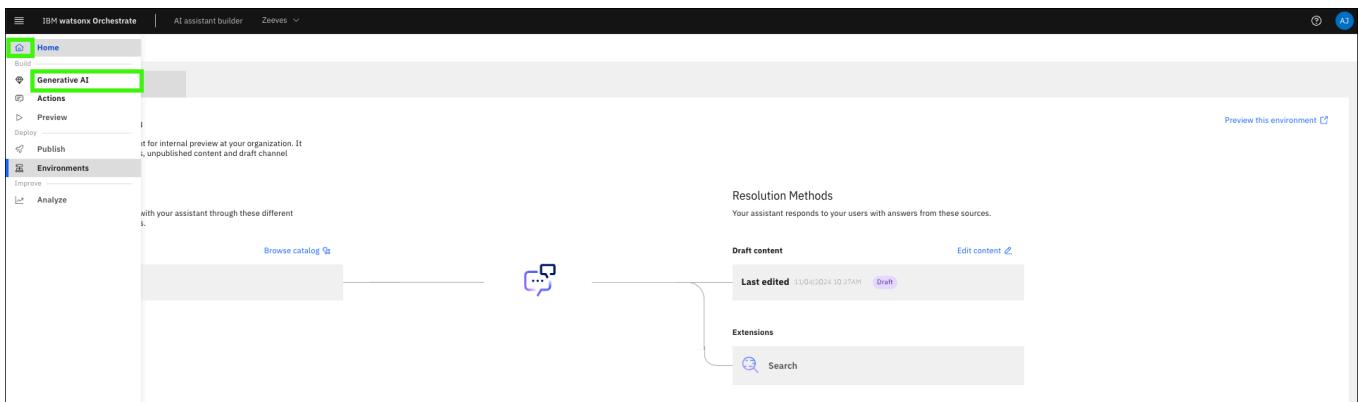
48. Click Save and exit.



Configure the base large language model

After the preceding steps are completed, there are enhancements that can be made to configure how the large language model (LLM) responds to your queries. Including adding prompt instructions and configuring the LLM's answer behavior. These options can be summarized [here](#).

1. Hover over the home () and click Generative AI.



2. Click Add instructions.

3. Enter a prompt instruction.

This option instructs the LLM in your assistant to give refined responses by adding prompt instructions. The instructions help the LLM guide the conversations with clarity and specifically to achieve the end-goal of an action.

Enter the prompt instructions in the field. The maximum number of characters you can enter in the prompt instruction field is 1,000.

The following is an example prompt instruction that works well. Experiment with different prompt instructions.

You are a subject matter expert on mainframe systems. Please respond to all prompts with truth and accuracy. Keep all answers short and concise, unless requested to provide details.

Note: When the instructions are typed in, they are automatically saved and the LLM is immediately trained on them.

4. Toggle General-purpose answering to off.

The ability exists to configure the answering behavior of your assistant to provide responses that are based on the preinstalled content or general content.

On the **Generative AI** page (under **Prompt Instructions**), you see the **Answer behavior** section. After you configure **Conversational search**, you see that it is enabled (toggled on) with the search integration added.

If you enable both General-purpose answering as well as Conversational search, the Conversational search answering takes precedence over General-purpose answering.

For purposes of retrieving Z-specific answers and responses, it is recommended that you turn off General-purpose answering and leave only Conversational search turned on.

The screenshot shows the 'Generative AI' configuration page in the IBM Watsonx Assistant builder. In the 'Answer behavior' section, the 'General-purpose answering' toggle switch is off (indicated by a greyed-out appearance), while the 'Conversational search' toggle switch is on (indicated by a green outline). To the right, a 'Preview' window displays a simulated interaction where the assistant greets a customer and asks for assistance. The overall interface includes sections for selecting a model, adding prompt instructions, and viewing the status of the LLM.

Testing conversational search

Now, you can begin issuing queries to test the assistant's responses.

It is important to keep in mind that many of the settings configured earlier can be iteratively modified based on your assessment of the quality of responses. The settings can be reviewed and changed at any time. For example, adding extra prompt instructions, changing the verbosity of the responses, and modifying the indexes used for OpenSearch.

53. Hover over the **home** (🏠) and click **Preview**.

The screenshot shows the 'Home' page of the IBM Watsonx Assistant builder. The 'Preview' button in the main content area is highlighted with a green box. The left sidebar shows navigation links for Home, Generative AI, Actions, and other features like Deploy, Publish, Environments, Improve, and Analyze.

54. Experiment with different prompts and validate the answers are reasonable and related to IBM Z.

Other prompts and responses follow. Note: the responses that you receive can vary from the ones shown.

Prompt:

What is the APF list in z/OS? Provide a detailed explanation?

Example output:

The screenshot shows the IBM Watsonx Orchestrate interface. On the left, there's a sidebar with icons for preview, AI assistant builder, and Zeeves. The main area is titled "Sample website". On the right, there's a "Zeeves" AI chat window. The chat history shows a user asking "What is the APF list in z/OS? Provide a detailed explanation?" and the AI responding with a detailed explanation of the APF (Authorized Program Facility) list in z/OS.

User: What is the APF list in z/OS? Provide a detailed explanation?

Zeeves:

The APF (Authorized Program Facility) list in z/OS is a security control mechanism which lists the programs and data sets that are authorized to run in privileged mode. It is used to ensure system security and integrity by restricting access to sensitive system resources. The APF list is located in the system's parmlib (parameter library) and can be managed using various system commands and macros.

The APF list contains a list of libraries and data sets that are authorized to run in privileged mode. These libraries and data sets are typically system-critical programs and utilities that require elevated privileges to function correctly. By authorizing these

Prompt:

Why is Db2 different than other database systems?

Example output:

The screenshot shows the IBM Watsonx Orchestrate interface. On the left, there's a sidebar with icons for preview, AI assistant builder, and Zeeves. The main area is titled "Sample website". On the right, there's a "Zeeves" AI chat window. The chat history shows a user asking "Why is Db2 different than other databases?" and the AI responding with a comparison between Db2 and other database systems.

User: Why is Db2 different than other databases?

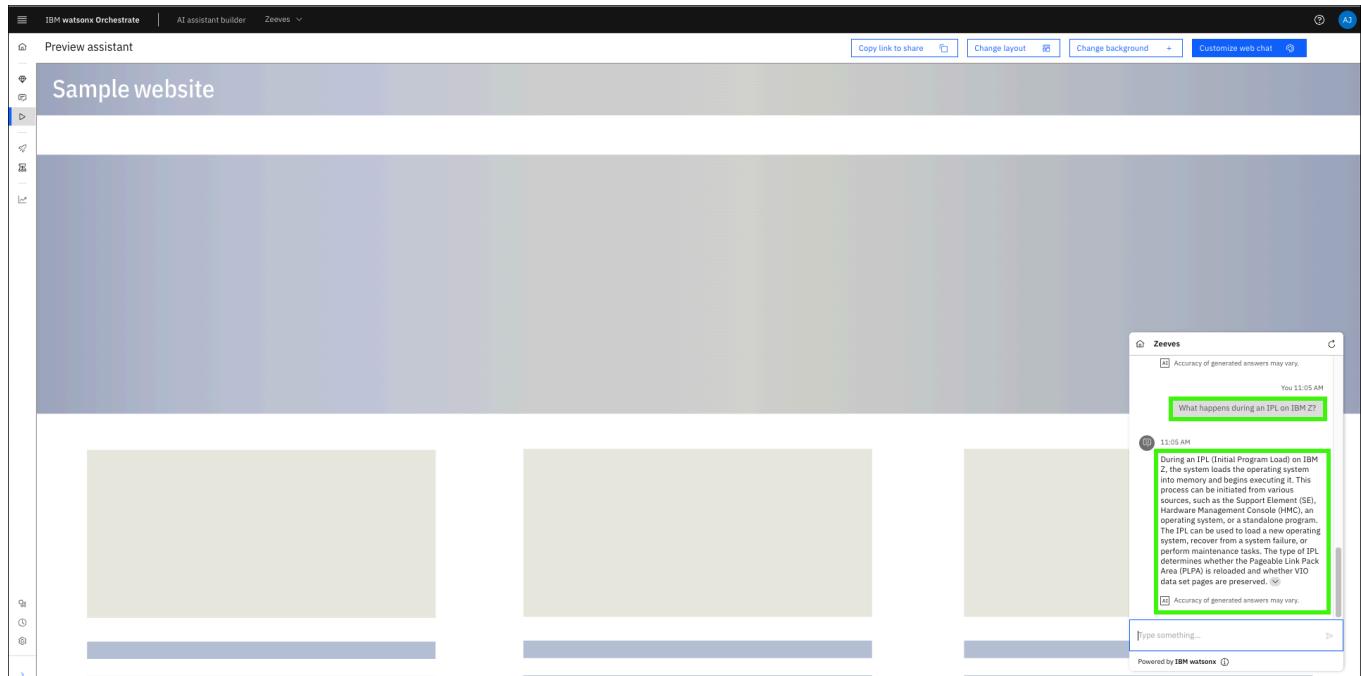
Zeeves:

Db2 is a relational database management system (DBMS) that runs under the z/OS operating system, which is specific to mainframe computers. Other database systems include Oracle, MySQL, PostgreSQL, and Microsoft SQL Server. Additionally, Db2 provides a higher level of data management and security features compared to other database systems, such as managing data stored in databases and controlling access to the data.

Prompt:

What happens during an IPL on IBM Z?

Example output:



You now have a working assistant that uses IBM watsonx Assistant for Z. Take time to explore different prompt instructions and settings. If you encounter any issues, the Troubleshooting section that follows can help resolve them.

Continue to the [Creating a stand-alone OpenSearch instance for document ingestion](#) learn how to configure a dedicated OpenSearch instance for ingesting client-specific documentation into the RAG model.

Troubleshooting

The following are issues that you may encounter. If the provided resolutions do not work, contact support by using the methods that are mentioned in the [Support](#) section.

✖ Assistant responds to all prompts with, "I might have information related to your query to share, but am unable to connect to my knowledge base at the moment" ✖

This Assistant is unable to connect to the custom service URL specified. This could be a network issue, the service may be down, the service may be restarting, or the service is no longer running at that URL.

Before reaching out to [Support](#), try the following:

- Wait a few minutes and try again. It may be the service was in the process of restarting.
- If you printed this demonstration guide or saved a copy, verify you are using the most current version of the [lab guide](#) and the correct service URL (<https://wxa4z-opensearch-wrapper-wxa4z-demo.wxo4z-opc-opensearch-clus-47e063e6a3ad1f71bf2e58f91c3b4c2e-0000.us-south.containers.appdomain.cloud/v1/query>). The URL may have changed since you saved or printed the lab guide.

Ingesting client documents

Creating a stand-alone OpenSearch instance for document ingestion

Next, learn to enable a client to ingest their own documentation into the Retrieval Augmented Generation (RAG) used by IBM watsonx Assistant for Z. The first step is to deploy a dedicated [OpenSearch](#) instance. The dedicated search instance will be referred to as bring-your-own-search (BYOS).

Earlier, you provisioned three IBM Technology Zone (ITZ) environments. One of which was a single-node Red Hat OpenShift (SNO) cluster. If you have not reserved this environment, or it is not in the **Ready** state, return to the [IBM Technology Zone environment](#) section to complete the reservation.

Install the Red Hat OpenShift command-line interface utility

The Red Hat OpenShift command-line interface (CLI) utility, which is known as **oc**, must be installed on your local workstation. If you already installed the **oc** utility, you can proceed to [log in to the SNO cluster](#).

1. Click the following link to open a browser window to your ITZ reservations.

[ITZ My reservations](#)

2. Click the **Single Node OpenShift** tile.

The screenshot shows the 'My reservations' section of the IBM Technology Zone interface. On the left, a sidebar lists options like 'My dashboard', 'My created collections', 'My bookmarked searches', 'My favorites', and 'My reservations'. The 'My reservations' option is selected and highlighted with a blue background. The main area displays three environment tiles:

- Single Node OpenShift (VMware on IBM Cloud)**: Status - Ready. This tile is highlighted with a green border around its title and the 'Open this environment' button below it. It shows details: Start date Oct 31, 2024 7:02 AM, End date Nov 6, 2024 6:49 AM, Extend limit 0.
- watsonx Assistant for Z Pilot - AAP & z/OS**: Status - Ready. It shows details: Start date Oct 30, 2024 8:12 AM, End date Nov 5, 2024 7:53 AM, Extend limit 0.
- watsonx Assistant for Z Pilot - watsonx O...**: Status - Ready. It shows details: Start date Oct 30, 2024 8:10 AM, End date Nov 5, 2024 7:12 AM, Extend limit 0.

At the top right of the main area, there are filters for 'Filter by infrastructure' and 'Filter by status', along with icons for 'Custom request', 'Add to calendar', and view modes ('List', 'Grid').

3. Scroll down and record the **Cluster Admin Username** and **Cluster Admin Password**.

The screenshot shows the IBM Technology Zone interface. At the top, there are two cluster nodes listed:

- 672371d38376796fb96a6c4d-master-1: Red Hat Enterprise Linux 8 (64-bit), IP 192.168.252.11, Running. A blue "Console" button is visible.
- 672371d38376796fb96a6c4d-nfs: Red Hat Enterprise Linux 8 (64-bit), IP 192.168.252.17, Running. A blue "Console" button is visible.

Below the nodes is a section titled "Reservation Details". It includes the following information:

- API URL: <https://api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:6443>
- Bastion Password: [REDACTED]
- Bastion RDP Address: api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:43389
- Bastion SSH connection: ssh itzuser@api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com -p 40222
- Bastion Username: itzuser
- Cluster Admin Username: kubeadmin
- Cluster Admin Password: [REDACTED]
- OCP Console: <https://console.openshift-console.apps.672371d38376796fb96a6c4d.ocp.techzone.ibm.com> (This link is highlighted with a green box)
- OCP Version: 4.14
- vCenter: itzeu-vc.eu.cloud.techzone.ibm.com

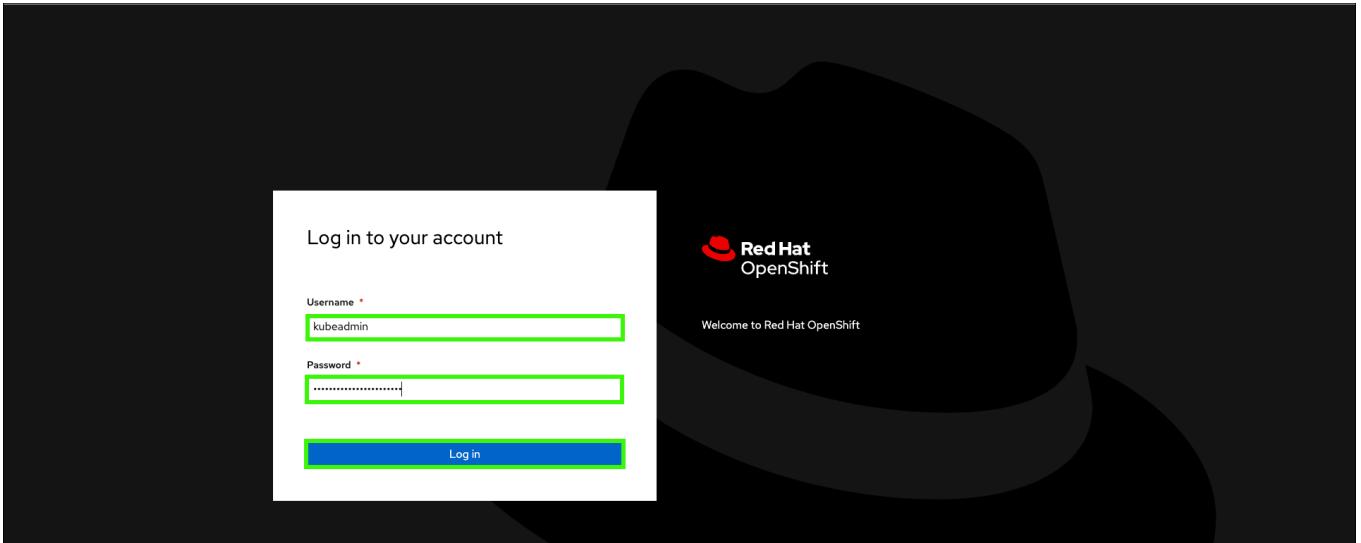
At the bottom right of the "Reservation Details" section is a blue "Download kubeconfig" button with a downward arrow icon.

4. Click the **OCP Console** link.

Note, OCP stands for OpenShift Container Platform.

This screenshot is identical to the one above, showing the IBM Technology Zone interface with two cluster nodes and their reservation details. The "OCP Console" link is again highlighted with a green box.

5. Enter the **Cluster Admin Username** and **Cluster Admin Password** values from step 3 and click **Log in**.



6. Click the help (?) and then click **Command Line Tools**.

7. Click the link under **oc - OpenShift Command Line Interface (CLI)** for the operating system of your local machine.

Clicking the preceding link automatically downloads either a **.zip** or **.tar** file specific to your operating system. Unzip or untar the file and place the **oc** binary in a directory that is in your default PATH. The PATH will vary by operating system and individual settings, detailed instructions are not provided.

After properly installed, verify the installation by running the **oc** command on your local workstation.

```
oc --help
```

Sample output:

```
andrewjones@Andrews-MBP ~ % oc --help
OpenShift Client

This client helps you develop, build, deploy, and run your applications on any
OpenShift or Kubernetes cluster. It also includes the administrative
commands for managing a cluster under the 'adm' subcommand.

Basic Commands:
  login           Log in to a server
  new-project     Request a new project
  new-app          Create a new application
  status           Show an overview of the current project
  project          Switch to another project
  projects         Display existing projects
  explain          Get documentation for a resource

Build and Deploy Commands:
  rollout          Manage a Kubernetes deployment or OpenShift deployment
  config           Revert part of an application back to a previous deployment
```

Prepare to ingest documents

Before ingesting documents, a few setup steps must be taken.

Login to the OpenShift cluster from your local terminal

Note: if you just installed the **oc** utility, you can skip the next 5 steps.

1. Click the following link to open a browser window to your ITZ reservations.

[ITZ My reservations](#)

2. Click the **Single Node OpenShift** tile.

3. Scroll to the bottom of the reservation page and record the **Cluster Admin Username** and **Cluster Admin Password**.

Reservation Details

API URL
<https://api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:6443>

Bastion Password [REDACTED]

Bastion RDP Address
api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:43389

Bastion SSH connection
ssh itzuser@api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com -p 40222

Bastion Username
itzuser

Cluster Admin Username
kubeadmin

Cluster Admin Password [REDACTED]

OCP Console
<https://console.openshift-console.apps.672371d38376796fb96a6c4d.ocp.techzone.ibm.com>

OCP Version
4.14

vCenter
itzeu-vc.eu.cloud.techzone.ibm.com

Download kubeconfig [\[button\]](#)

4. Click the **OCP Console** link.

Reservation Details

API URL
<https://api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:6443>

Bastion Password [REDACTED]

Bastion RDP Address
api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:43389

Bastion SSH connection
ssh itzuser@api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com -p 40222

Bastion Username
itzuser

Cluster Admin Username
kubeadmin

Cluster Admin Password [REDACTED]

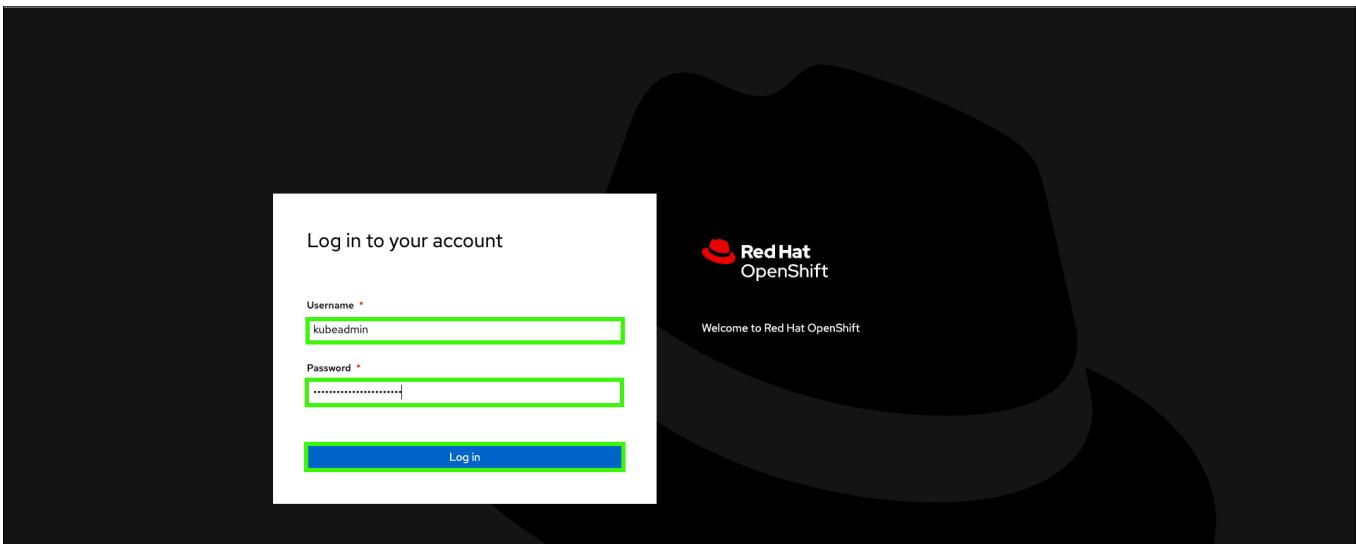
OCP Console
<https://console.openshift-console.apps.672371d38376796fb96a6c4d.ocp.techzone.ibm.com>

OCP Version
4.14

vCenter
itzeu-vc.eu.cloud.techzone.ibm.com

Download kubeconfig [\[button\]](#)

5. Enter the **Cluster Admin Username** and **Cluster Admin Password** values from step 3 and click **Log in**.



6. Click the **kube:admin** profile drop-down and click **Copy login command**.

You are logged in as a temporary administrative user. Update the cluster OAuth configuration to allow others to log in.

- Copy login command**
- User Preferences
- Log out

7. Click **Display Token**.



8. Select and copy the **Log in with this token** string.

The steps to select and copy the value may differ by operating system, but for most, you can double-click the value and then right click and select **Copy**.

Your API token is
sha256~zuWR0KDnkYniIY0m8g8iKoUXPdFFFmou-o4s5FsrDNA

Log in with this token

```
oc login --token=sha256~zuWR0KDnkYniIY0m8g8iKoUXPdFFFmou-o4s5FsrDNA --server=https://api.672b79320c7a71b728e523b4.ocp.techzone.ir
```

Use this token directly against the API

```
curl -H "Authorization: Bearer sha256~zuWR0KDnkYniIY0m8g8iKoUXPdFFFmou-o4s5FsrDNA" "https://api.672b79320c7a71b728e523b4.ocp.techzone.ir/v1/users/~"
```

Request another token

Logout

9. Open a command prompt or terminal window on your local workstation.

10. Paste the login command and press **enter**.

```
andrewjones@Andrews-MBP ~ % oc login --token=sha256~mJ4L8K6cUMyNyk2Z69KMm3vbP1sWc8SW0eeOdVqtA94 --server=https://api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:6443
Logged into "https://api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:6443" as "kube:admin" using the token provided.
You have access to 70 projects, the list has been suppressed. You can list all projects with 'oc projects'
Using project "default".
andrewjones@Andrews-MBP ~ %
```

Create a working directory

11. Create a directory that will be used to store the configuration files that will be created in the next steps.



Instructions vary by your local workstation's operating system

The directions that follow may vary depending on your operating system. The examples provided are based upon MacOS.

```
mkdir watsonxAssistant
```

12. Change to the new directory.

```
cd watsonxAssistant
```

```
watsonxAssistant ~ % oc login --token=sha256~mJ4L8K6cUMyNyk2Z69KMm3vbP1sWc8SW0eeOdVqtA94 --server=https://api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:6443
Logged into "https://api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:6443" as "kube:admin" using the token provided.
You have access to 70 projects, the list has been suppressed. You can list all projects with 'oc projects'
Using project "default".
andrewjones@Andrews-MBP ~ % mkdir watsonxAssistant
andrewjones@Andrews-MBP ~ % cd watsonxAssistant
andrewjones@Andrews-MBP watsonxAssistant %
```

Install IBM Certificate Manager on Red Hat OpenShift

13. Create a file that is named **catalogCertManager.yaml** using a text editor and paste the following text into the file.



Formatting of the yaml file is critical!

The content of the YAML file must be formatted exactly as shown. Use the copy icon to prevent typographical errors.

```

apiVersion: operators.coreos.com/v1alpha1
kind: CatalogSource
metadata:
  name: ibm-cert-manager-catalog
  namespace: openshift-marketplace
spec:
  displayName: ibm-cert-manager-4.2.7
  grpcPodConfig:
    securityContextConfig: restricted
  image: icr.io/cpopen/ibm-cert-manager-operator-
catalog@sha256:4dcf4ace4b5f166f83b31063f7e6404dbf78d8e98a9d4fcf52fedf576a55ca6c
  publisher: IBM
  sourceType: grpc
  updateStrategy:
    registryPoll:
      interval: 30m0s

```

14. Install the IBM Certificate Manager operator in the Red Hat OpenShift cluster.

```
oc apply -f catalogCertManager.yaml
```

The preceding command returns a message that states the **ibm-cert-manager-catalog** was created.

15. In the OpenShift web console, click **Operators** and select **OperatorHub**.

The screenshot shows the Red Hat OpenShift web console interface. The left sidebar has a 'Operators' dropdown with 'OperatorHub' selected, highlighted with a green box. The main content area displays information about the 'oc' command-line tool, including download links for various platforms. The top navigation bar shows the cluster name 'Red Hat OpenShift on IBM Techzone' and the user 'kube:admin'.

16. Click the **Project** to pull-down menu and click the **Show default projects** toggle.

The screenshot shows the Red Hat OpenShift web console with the 'OperatorHub' page open. A green box highlights the 'Project: All Projects' dropdown menu. Another green box highlights the 'Show default projects' toggle switch, which is turned on. The main content area lists several projects: 'All Projects', 'default', 'kube-node-lease', and 'kube-public'. At the bottom, there are tabs for 'Community', 'Marketplace', and 'Community' again. The top navigation bar shows the cluster name 'Red Hat OpenShift on IBM Techzone' and the user 'kube:admin'.

17. Scroll down and select **openshift-marketplace**.

You are logged in as a temporary administrative user. Update the cluster OAuth configuration to allow others to log in.

Project: All Projects

operators

operators

operatorHub

Installed Operators

Workloads

Networking

Storage

Builds

Observe

Compute

User Management

Administration

624 items

Community

[DEPRECATED] CrowdStrike Operator provided by CrowdStrike

[DEPRECATED] Use the CrowdStrike Falcon Operator from the certified channel instead

Community

ABot

ibm-cert-manager-4.2.7

Marketplace

Community

Hazelcast Platform Operator provided by Hazelcast, Inc.

[DEPRECATED] Use the certified Hazelcast Operator instead

Certified

Accuknox Operator provided by Accuknox Inc.

Red Hat

Advanced Cluster Management for Kubernetes

Create Project

18. Enter IBM Cert Manager in the search field and then click the IBM Cert Manager tile.

You are logged in as a temporary administrative user. Update the cluster OAuth configuration to allow others to log in.

Project: openshift-marketplace

All Items

Q IBM Cert Manager

1 items

ibm-cert-manager-4.2.7

IBM Cert Manager provided by IBM

Operator for managing deployment of cert-manager service.

19. Click Install.

You are logged in as a temporary administrative user. Update the cluster OAuth configuration to allow others to log in.

Project: openshift-marketplace

All Items

Q IBM Cert Manager

ibm-cert-manager-4.2.7

IBM Cert Manager provided by IBM

Operator for managing deployment of cert-manager service.

IBM Cert Manager

4.2.7 provided by IBM

Install

Channel: v4.2

Version: 4.2.7

Capability level:

- Basic Install
- Seamless Upgrades
- Full Lifecycle
- Deep Insights
- Auto Pilot

Source: ibm-cert-manager-4.2.7

20. Keep the default settings and click Install.

You are logged in as a temporary administrative user. Update the [cluster OAuth configuration](#) to allow others to log in.

Install Operator

Install your Operator by subscribing to one of the update channels to keep the Operator up to date. The strategy determines either manual or automatic updates.

Update channel *

v4.2

Version *

4.2.7

Installation mode *

All namespaces on the cluster (default)
Operator will be available in all Namespaces.

A specific namespace on the cluster
This mode is not supported by this Operator

Installed Namespace *

Operator recommended Namespace: ibm-cert-manager

Select a Namespace

Namespace creation
Namespace ibm-cert-manager does not exist and will be created.

Update approval *

Automatic

Manual

Provided APIs

- IBM Cert Manager** provided by IBM
- CR CertificateRequest** Not available
- C Certificate** A Certificate resource should be created to ensure an up to date and signed x509 certificate is stored in the Kubernetes Secret resource named in `spec.secretName`. Documentation For additional details regarding install parameters check...
- CMC Cert Manager Config** CertManagerConfig is the Schema for the certmanagerconfigs API. Documentation For additional details regarding install parameters check: <https://ibm.biz/icpf39install>. License By installing this product you accept the license terms...
- Challenge** Not available
- ClusterIssuer** Not available
- Issuer** An issuer represents a certificate issuing authority which can be referenced as

Install **Cancel**

The installation process takes a few minutes. Do not continue until you see the message: **Installed operator: ready for use.**

You are logged in as a temporary administrative user. Update the [cluster OAuth configuration](#) to allow others to log in.

Installed operator: ready for use

View Operator [View installed Operators in Namespace ibm-cert-manager](#)

Install the watsonx Assistant for Z Operator (for OpenSearch)

- In your command prompt or terminal window, create a new namespace called **wxa4z-byos** in the Red Hat OpenShift cluster.

```
oc create namespace wxa4z-byos
```

- Create or obtain your IBM Container Software **production entitlement key**.

A **production entitlement key** is required to pull the container images that get deployed by the operator.

To create this key, follow the [instructions here](#) to create or retrieve your existing entitlement key.

After locating your existing key or creating a new key, continue to the next step.

23. Copy the **production entitlement key**.

The screenshot shows the IBM Container Software and Cloud Pak Access Management interface. On the left, there's a sidebar with 'Entitlement keys' selected. The main area is titled 'Entitlement keys (1)'. It says 'Access your container software' and provides a brief description of what an entitlement key is. Below that, it lists 'Active entitlement keys' with a note about having up to 5 keys. A specific key is shown with its issue date as 'October 17, 2022'. At the bottom right of this card, there are 'Copy', 'Edit', and 'Delete' buttons, with 'Copy' being highlighted by a green box.

24. Using the following command in your command prompt or terminal window, set an environment variable with your **production entitlement key** and then create a pull secret for the Container Registry.

Substitute your **production entitlement key** for the <entitlement key> string.

```
export IBM_CS_ENT_KEY=<entitlement key>
```

```
oc -n wxa4z-byos create secret docker-registry icr-pull-secret --docker-server=cp.icr.io --  
docker-username=cp --docker-password=$IBM_CS_ENT_KEY
```

The screenshot shows a terminal window with a command-line interface. The user has run the command 'oc -n wxa4z-byos create secret docker-registry icr-pull-secret --docker-server=cp.icr.io --docker-username=cp --docker-password=\$IBM_CS_ENT_KEY'. The entire command line is highlighted with a green box.

25. Create a file called **catalogSource.yaml** with the following content.

Formatting of the yaml file is critical!

The content of the YAML file must be formatted exactly as shown. Use the copy icon to prevent typographical errors.

```

apiVersion: operators.coreos.com/v1alpha1
kind: CatalogSource
metadata:
  name: ibm-wxa4z-operator-catalog
  namespace: wxa4z-byos
spec:
  displayName: "IBM watsonx Assistant for Z Operator Catalog"
  image: icr.io/cpopen/ibm-wxa4z-
catalog:2.0.2@sha256:d457aa24af07c23d7b36727a9cff95995ca1c4ed49fb980ef4f386fd09a29b4
  publisher: 'IBM'
  sourceType: grpc
  secrets:
    - icr-pull-secret

```

26. Create your document catalog in the Red Hat OpenShift operator.

```
oc apply -f catalogSource.yaml
```

```

andrewjones@Andrews-MBP watsonxAssistant % oc create namespace wxa4z-byos
namespace/wxa4z-byos created
andrewjones@Andrews-MBP watsonxAssistant % export IBM_CS_ENT_KEY=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9eyJpc3MiOiJJQk0gTWFya2V0cGxhY2UiLCJpXXQiOjE2NjYwMTk1ODAsImp0aSI6IjNKOWUyMzzjZTAzMDQzMzVhNTJhYTkzMWNmOTcyMDR1In0.4M3XRD4XzkHMS0kFNJ4uKvcWZ6SnEA0Z03eL_11A2XY
andrewjones@Andrews-MBP watsonxAssistant % oc -n wxa4z-byos create secret docker-registry icr-pull-secret --docker-server=cp.icr.io --docker-username=cp --docker-password=$IBM_CS_ENT_KEY
secret/icr-pull-secret created
andrewjones@Andrews-MBP watsonxAssistant % vi catalogSource.yaml
andrewjones@Andrews-MBP watsonxAssistant % vi catalogSource.yaml
andrewjones@Andrews-MBP watsonxAssistant % oc apply -f catalogSource.yaml
catalogsource.operator.coreos.com/ibm-wxa4z-operator-catalog created:
andrewjones@Andrews-MBP watsonxAssistant %

```

27. In the Red Hat OpenShift web console, click **OperatorHub** and select the **wxa4z-byos** project.

The screenshot shows the Red Hat OpenShift web console interface. The left sidebar is collapsed. The main navigation bar has 'Administrator' selected. The 'OperatorHub' menu item is highlighted with a green box. The search bar at the top right contains the text 'Project: openshift-marketplace'. Below the search bar, the 'Operators' section is expanded, showing a list of installed operators like 'openshift-node', 'openshift-nutanix-infra', etc. On the right, there is a large grid of operator cards. One card for 'wxa4z-byos' is highlighted with a green box. The card details are as follows:

- Community**
- Operator**: wxa4z-byos
- Description**: partners, curated by Red Hat. You can purchase commercial software through Red Hat Marketplace. You can install Operators on your clusters to provide optional add-ons and shared libraries will appear in the Developer Catalog providing a self-service experience.
- Provider**: Red Hat
- Version**: 2.0.2
- Category**: Infra
- Created**: 2023-07-10
- Last Updated**: 2023-07-10
- Image**: icr.io/cpopen/ibm-wxa4z-
- Source Type**: grpc
- Secrets**: icr-pull-secret
- Annotations**: `operator.coreos.com/wxa4z-operator-catalog: "true"`
- Labels**: `app.kubernetes.io/instance: ibm-wxa4z-operator-catalog, app.kubernetes.io/name: ibm-wxa4z-operator-catalog, app.kubernetes.io/version: 2.0.2, chart: ibm-wxa4z-operator-catalog-0.1.0, component: ibm-wxa4z-operator-catalog, controller-revision-hashing: true, helm: v3, name: ibm-wxa4z-operator-catalog, release: ibm-wxa4z-operator-catalog, version: v1alpha1`

28. Enter **ibm watsonx** in the search field and the click the **IBM watsonx Assistant for Z Operator Catalog** tile.

Note, it may take a few minutes for the operator to appear. Try again if it is not initially found.

You are logged in as a temporary administrative user. Update the cluster OAuth configuration to allow others to log in.

Project: wxa4z-byos

OperatorHub

Discover Operators from the Kubernetes community and Red Hat partners, curated by Red Hat. You can purchase commercial software through Red Hat Marketplace. You can install Operators on your clusters to provide optional add-ons and shared services to your developers. After installation, the Operator capabilities will appear in the Developer Catalog providing a self-service experience.

All Items All Items

Search: ibm watsonx

1 items

IBM watsonx Assistant for Z Operator Catalog
IBM watsonx Assistant for Z provided by IBM
IBM watsonx Assistant for Z Operator

29. Click Install.

You are logged in as a temporary administrator.

Project: wxa4z-byos

OperatorHub

Discover Operators from the Kubernetes community and Red Hat partners, curated by Red Hat. You can purchase commercial software through Red Hat Marketplace. You can install Operators on your clusters to provide optional add-ons and shared services to your developers. After installation, the Operator capabilities will appear in the Developer Catalog providing a self-service experience.

All Items All Items

Search: ibm watsonx

IBM watsonx Assistant for Z
2.0.1 provided by IBM

Install

Channel: stable

Version: 2.0.1

Capability level:

- Basic Install
- Seamless Upgrades
- Full Lifecycle
- Deep Insights
- Auto Pilot

Source: IBM watsonx Assistant for Z Operator Catalog

Provider: IBM

Repository: N/A

Container Image: N/A

Created at: Oct 4, 2024, 1:31 PM

30. Select A specific namespace on the cluster under Installation mode and wxa4z-byos for the Installed Namespace, then click Install.

You are logged in as a temporary administrative user. Update the cluster OAuth configuration to allow others to log in.

Install Operator

Install your Operator by subscribing to one of the update channels to keep the Operator up to date. The strategy determines either manual or automatic updates.

Update channel *: stable

Version *: 2.0.1

Installation mode *:

- All namespaces on the cluster (default)
Operator will be available in all Namespaces
- A specific namespace on the cluster
Operator will be available in a single Namespace only.

Installed Namespace *: `wxa4z-byos`

Update approval *:

- Automatic
- Manual

Install **Cancel**

The installation process takes a few minutes. Do not continue until you see the message: **Installed operator: ready for use**.

You are logged in as a temporary administrative user. Update the cluster OAuth configuration to allow others to log in.

Installed Operators

ibm-wxa4z-operator:v2.0.1 provided by IBM

ibm-wxa4z-operator:v2.0.1 provided by IBM

Installed operator: ready for use

View Operator **View installed Operators in Namespace `wxa4z-byos`**

- In your command prompt or terminal window, run the following commands to add the Container Registry credential to the operator's service account.

```
oc project wxa4z-byos
```

```
oc patch serviceaccount ibm-wxa4z-operator-controller-manager --type merge -p
'{"imagePullSecrets": [{"name": "icr-pull-secret"}]}'
```

```
watsonxAssistant -- cecuser@p1365-pvm1:~ -zsh - 135x37
andrewjones@Andrews-MBP watsonxAssistant % oc create namespace wxa4z-byos
namespace/wxa4z-byos created
andrewjones@Andrews-MBP watsonxAssistant % export IBM_CS_ENT_KEY=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJpc3MiOiJJQk0gTWFya2V0cGxhY2UiLCJpYXQiOjE2NjYwMTk1ODAsImp0aS16IjNkOWUyMzjZTAzMDQzMzVhNTJhYTkzMWNmOTcyMDR1In0.4M3XRD4XzkHMSOkFNJ4uKVcwZ6SnEAZO3eL_11A2xY
andrewjones@Andrews-MBP watsonxAssistant % oc -n wxa4z-byos create secret docker-registry icr-pull-secret --docker-server=cp.icr.io --docker-username=cp --docker-password=$IBM_CS_ENT_KEY
secret/icr-pull-secret created
andrewjones@Andrews-MBP watsonxAssistant % vi catalogSource.yaml
andrewjones@Andrews-MBP watsonxAssistant % vi catalogSource.yaml
andrewjones@Andrews-MBP watsonxAssistant % oc apply -f catalogSource.yaml
catalogsource.operator.coreos.com/ibm-wxa4z-operator-catalog created
andrewjones@Andrews-MBP watsonxAssistant % oc project wxa4z-byos
Now using project "wxa4z-byos" on server "https://api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:6443".
andrewjones@Andrews-MBP watsonxAssistant % oc patch serviceaccount ibm-wxa4z-operator-controller-manager --type merge -p '{"imagePullSecrets": [{"name": "icr-pull-secret"}]}'
serviceaccount/ibm-wxa4z-operator-controller-manager patched
andrewjones@Andrews-MBP watsonxAssistant %
```

32. In the Red Hat OpenShift web console, click **Workloads** and select the **Pods**.

33. Verify the two pods that start with **ibm-wxa4z-operator** have a Status of **Running** and that all pods are **Ready**.

Name	Status	Ready	Restarts	Owner	Memory	CPU	Created
ld126367b1ca53dcf2b0c93acd733e33875cd4b6c382dd5eb412032ac2h6	Completed	0/1	0	ld126367b1ca53dcf2b0c93acd733e33875cd4b6c382dd5eb412032ac42b38	-	-	Nov 4, 2024, 4:44 PM
ibm-wxa4z-operator-catalog-n9m5m	Running	1/1	0	ibm-wxa4z-operator-catalog	19.0 MiB	0.003 cores	Nov 4, 2024, 4:32 PM
ibm-wxa4z-operator-controller-manager-7c7898d7d4-87hmt	Running	2/2	0	ibm-wxa4z-operator-controller-manager-7c7898d7d4	31.0 MiB	0.001 cores	Nov 4, 2024, 4:44 PM

34. Run the following command to set the administrative policy for the workspace.

```
oc -n wxa4z-byos adm policy add-scc-to-user privileged -z byos
```

```
watsonxAssistant — cecuser@p1365-pvm1:~ -- zsh — 135x37
andrewjones@Andrews-MBP watsonxAssistant % oc create namespace wxa4z-byos
namespace/wxa4z-byos created
andrewjones@Andrews-MBP watsonxAssistant % export IBM_CS_ENT_KEY=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJpc3MiOiJJQk0gTWFya2V0cGxhY2UiCJpYXQiOjE2NjYwMTk1ODAsImp0aSI6ijNkOWUyMzZjTAzM0QzMzvhNTJhYTkzMWNmOTcyMDR1In0.4M3XRD4XzkHMS0kFNJ4uKVcWZ6SnEA0Z03eL_11A2xY
andrewjones@Andrews-MBP watsonxAssistant % oc -n wxa4z-byos create secret docker-registry icr-pull-secret --docker-server=cp.icr.io --docker-username=cp --docker-password=$IBM_CS_ENT_KEY
secret/icr-pull-secret created
andrewjones@Andrews-MBP watsonxAssistant % vi catalogSource.yaml
andrewjones@Andrews-MBP watsonxAssistant % vi catalogSource.yaml
andrewjones@Andrews-MBP watsonxAssistant % oc apply -f catalogSource.yaml
catalogsource.operator.coreos.com/ibm-wxa4z-operator-catalog created
andrewjones@Andrews-MBP watsonxAssistant % oc project wxa4z-byos
Now using project "wxa4z-byos" on server "https://api.672371d38376796fb96a6c4d.ocp.techzone.ibm.com:6443".
andrewjones@Andrews-MBP watsonxAssistant % oc patch serviceaccount ibm-wxa4z-operator-controller-manager --type merge -p '{"imagePullSecrets": [{"name": "icr-pull-secret"}]}'
serviceaccount/ibm-wxa4z-operator-controller-manager patched
andrewjones@Andrews-MBP watsonxAssistant % oc -n wxa4z-byos adm policy add-scc-to-user privileged -z byos
clusterrole.rbac.authorization.k8s.io/system:openshift:scc:privileged added: "byos"
andrewjones@Andrews-MBP watsonxAssistant %
```

Deploy required secrets and the custom bring-your-own-search (BYOSearch) resource

35. Create a file called **os-secret.yaml** with the following content.

Substitute a secure password of your choosing for the string <OPENSEARCH_PASSWORD>.

```
apiVersion: v1
stringData:
  password: <OPENSEARCH_PASSWORD>
kind: Secret
metadata:
  name: opensearch-creds
  namespace: wxa4z-byos
  type: Opaque
```

36. Create the secret by running the following command.

```
oc apply -f os-secret.yaml
```

37. Create a file called **client-ingestion-secret.yaml** with the following content.

Substitute a secure authentication key of your choosing for the string <CLIENT_INGESTION_AUTHKEY>. The authentication key can be a random password.

```
apiVersion: v1
stringData:
  authkey: <CLIENT_INGESTION_AUTHKEY>
kind: Secret
metadata:
  name: client-ingestion-authkey
  namespace: wxa4z-byos
  type: Opaque
```

38. Create the secret by running the following command.

```
oc apply -f client-ingestion-secret.yaml
```

39. Obtain and record your cluster domain that is used for routes by running the following command.

```
oc -n openshift-ingress-operator get ingresscontroller default -o jsonpath=".status.domain"
```

**Remove the trailing % from the cluster domain name.**

The value returned for the cluster domain may include a % character at the end. Do not include the % in the next step!

Note, the output of the command will be a string similar to:

apps.672b79320c7a71b728e523b4.ocp.techzone.ibm.com

40. Create a file called **byos.yaml** with the following content.

Substitute the domain name recorded in the previous step for the string <YOUR_CLUSTER_DOMAIN>. Do not include the % at the end of the domain name.

```

apiVersion: wxa4z.watsonx.ibm.com/v1
kind: BYOSearch
metadata:
  name: byosearch
  namespace: wxa4z-byos
spec:
  imagePullSecrets:
    - name : icr-pull-secret
  namespace: wxa4z-byos
  clusterName: wxa4z-byos-cluster
  clusterDomain: <YOUR_CLUSTER_DOMAIN>

  opensearch:
    secretName: opensearch-creds

  persistence:
    enabled: true
    storageClass: "managed-nfs-storage"
    accessModes:
      - ReadWriteOnce
    size: 24Gi

  wrapper:
    createRoute: true
    resources:
      requests:
        cpu: 4
        memory: "500Mi"
      limits:
        cpu: 4
        memory: "1Gi"

  clientIngestion:
    secretName: client-ingestion-authkey

    resources:
      limits:
        cpu: "500m"
        memory: 2Gi
        nvidia.com/gpu: "0"
      requests:
        cpu: "500m"
        memory: 1Gi
        nvidia.com/gpu: "0"
    pvc:
      storageClass: "managed-nfs-storage"
      enabled: true
      size: 24Gi

```

41. Run the following command to deploy BYOS on your cluster.

```
oc apply -f byos.yaml
```

The BYOS deployment begins and might take up to 20 minutes or more for the images to download and the deployment to complete. You can check the status by looking at the Pods view in the OCP console.

Verify all the required pods are running

42. Verify that all pods have the status of **Running** or **Completed**.

Name	Status	Ready	Restarts	Owner	Memory	CPU	Created
da80909aab1563ff0c15b7c6d	Completed	0/1	0	da80909aab1563ff0c15b7c6da81 a8bfefabaa605065bb6c6b490e87 0e87f0k96l	-	-	Nov 6, 2024, 9:26 AM
ibm-wxa4z-operator-catalog-cfvw	Running	1/1	0	ibm-wxa4z-operator-catalog	25.0 MiB	0.004 cores	Nov 6, 2024, 9:25 AM
ibm-wxa4z-operator-controller-manager-556fcf98bb-4bhmk	Running	2/2	0	ibm-wxa4z-operator-controller-manager-556fcf98bb	142.8 MiB	0.023 cores	Nov 6, 2024, 9:26 AM
wxa4z-byos-cluster-0	Running	1/1	0	wxa4z-byos-cluster	1,095.8 MiB	0.017 cores	Nov 6, 2024, 10:22 AM
wxa4z-byos-cluster-1	Running	1/1	0	wxa4z-byos-cluster	1,169.1 MiB	0.016 cores	Nov 6, 2024, 10:22 AM
wxa4z-byos-cluster-2	Running	1/1	0	wxa4z-byos-cluster	1,113.3 MiB	0.015 cores	Nov 6, 2024, 10:22 AM
wxa4z-client-ingestion-7f9b8d6c58-9bzth	Running	1/1	0	wxa4z-client-ingestion-7f9b8d6c58	501.8 MiB	0.071 cores	Nov 6, 2024, 12:18 PM
wxa4z-opensearch-wrapper-5cb879f5fb-qw7qt	Running	1/1	0	wxa4z-opensearch-wrapper-5cb879f5fb	547.1 MiB	0.031 cores	Nov 6, 2024, 10:22 AM
wxa4z-snapshot-setup-job-nsqtz	Completed	0/1	0	wxa4z-snapshot-setup-job	-	-	Nov 6, 2024, 10:22 AM

When deployment completes and all the pods have a status of "Ready", "Running", or "Completed", the next step is to retrieve your BYOS endpoint URL.

43. In the OCP console, click **Networking** and then **Routes**.

Name	Status	Ready	Restarts	Owner	Memory	CPU	Created
da80909aab1563ff0c15b7c6d	Completed	0/1	0	da80909aab1563ff0c15b7c6da81 a8bfefabaa605065bb6c6b490e87 0e87f0k96l	-	-	Nov 6, 2024, 9:26 AM
ibm-wxa4z-operator-catalog-cfvw	Running	1/1	0	ibm-wxa4z-operator-catalog	25.0 MiB	0.004 cores	Nov 6, 2024, 9:25 AM
ibm-wxa4z-operator-controller-manager-556fcf98bb-4bhmk	Running	2/2	0	ibm-wxa4z-operator-controller-manager-556fcf98bb	141.2 MiB	0.022 cores	Nov 6, 2024, 9:26 AM
wxa4z-byos-cluster-0	Running	1/1	0	wxa4z-byos-cluster	1,095.8 MiB	0.016 cores	Nov 6, 2024, 10:22 AM
wxa4z-byos-cluster-1	Running	1/1	0	wxa4z-byos-cluster	1,169.1 MiB	0.016 cores	Nov 6, 2024, 10:22 AM
wxa4z-byos-cluster-2	Running	1/1	0	wxa4z-byos-cluster	1,113.3 MiB	0.015 cores	Nov 6, 2024, 10:22 AM
wxa4z-client-ingestion-7f9b8d6c58-9bzth	Running	1/1	0	wxa4z-client-ingestion-7f9b8d6c58	534.6 MiB	0.087 cores	Nov 6, 2024, 12:18 PM
wxa4z-opensearch-wrapper-5cb879f5fb-qw7qt	Running	1/1	0	wxa4z-opensearch-wrapper-5cb879f5fb	547.1 MiB	0.030 cores	Nov 6, 2024, 10:22 AM
wxa4z-snapshot-setup-job-nsqtz	Completed	0/1	0	wxa4z-snapshot-setup-job	-	-	Nov 6, 2024, 10:22 AM

44. Copy and record the location for the **wxa4z-opensearch-wrapper** route.

Name	Status	Location	Service
wxa4z-client-ingestion	Accepted	https://wxa4z-client-ingestion-wxa4z-byos.apps.672b79320c7a71b728e523b4.ocp.techzone.ibm.com	wxa4z-client-ingestion
wxa4z-opensearch-wrapper	Accepted	https://wxa4z-opensearch-wrapper-wxa4z-byos.apps.672b79320c7a71b728e523b4.ocp.techzone.ibm.com	wxa4z-opensearch-wrapper

Update your assistant with the new BYOS search instance

45. Configure your assistant with the route recorded in the previous step.

First, append the string **/v1/query** to complete the URL endpoint. The URL should look similar to:

https://wxa4z-opensearch-wrapper-wxa4z-byos.apps.672b79320c7a71b728e523b4.ocp.techzone.ibm.com/v1/query

Note, the above URL will not work for you. Use the value of your specific OpenSearch instance that is recorded in the previous step.

Next, you need to return to your assistant in the watsonx Orchestrate AI assistant builder and update the custom search integration URL. The steps to update the URL are illustrated in the animated gif that follows. You can see the steps [here](#).

The screenshot shows the 'Routes' section of the Red Hat OpenShift interface. The left sidebar shows 'Workloads' and 'Networking' sections, with 'Routes' currently selected under 'Networking'. The main area displays two routes: 'wxa4z-client-ingestion' and 'wxa4z-opensearch-wrapper'. The 'wxa4z-client-ingestion' route is highlighted with a yellow box and a cursor icon, indicating it is the active item. The URL for this route is visible in the status bar at the bottom of the browser window.

Troubleshooting

The following are issues that you may encounter. If the provided resolutions do not work, contact support by using the methods that are mentioned in the [Support](#) section.

✖ Pods have a status of ErrImagePull or ImagePullBackoff

If the pods starting with **ibm-wxa4z-operator** have a status of “ErrImagePull” or “ImagePullBackoff”, you can delete the pod and it will automatically restart and pull the image successfully. Wait until the pod is re-created successfully.

✖ The wxa4z-client-ingestion pod does not start

Did you include the % character in the **clusterDomain** name when creating the **byos.yaml**? To resolve, edit the **byos.yaml** file and run the following command again. The current pod will be terminated and a new one started. This will take about 20 minutes to start.

```
oc apply -f byos.yaml
```

Installing and using zassist to ingest client documents

With bring-your-own-search (BYOS) installed and configured in your assistant, you can now prepare for document ingestion. Currently, only PDF, HTML, and DOCX file formats are supported for ingestion.

To prepare for document ingestion, you can also reference the setup instructions that are located [here](#).

Install the zassist utility

The **zassist** utility is an executable program that automates the ingestion of client documentation into the RAG for watsonx Assistant for Z. The utility is available to clients through [IBM Passport Advantage](#).

A version of zassist is available for download for IBMers and Business Partners for conducting pilots. Follow the steps that follow to download and install **zassist**.

1. Click the following link and download the **zassist.zip** file.

<https://ibm.box.com/s/j3nt5iw4fqd5w2jgcqwxnjlsu8bpvl77>

Name	Modified	Size
zassist	Sep 5 at 9:15 AM	--

2. Extract the **zassist.zip** file.

3. Locate the appropriate file for your local workstation's operating system.

Name	Date Modified	Size	Kind
byos.yaml	Today at 12:17 PM	956 bytes	YAML
ca.crt	Today at 11:36 AM	Zero bytes	certificate
catalogCertManager.yaml	Today at 9:15 AM	476 bytes	YAML
catalogSource.yaml	Today at 9:25 AM	397 bytes	YAML
client-ingestion-secret.yaml	Today at 10:05 AM	155 bytes	YAML
os-secret.yaml	Today at 10:04 AM	149 bytes	YAML
zassist	Sep 5, 2024 at 11:15 AM	--	Folder
linux	Sep 18, 2024 at 8:15 AM	--	Folder
mac	Sep 18, 2024 at 8:15 AM	--	Folder
zassist	Sep 18, 2024 at 8:17 AM	10.2 MB	Unix Executable File
mac-arm	Sep 18, 2024 at 8:15 AM	--	Folder
windows	Sep 18, 2024 at 8:15 AM	--	Folder
zassist.zip	Today at 1:47 PM	23 MB	ZIP archive

4. Either copy the appropriate **zassist** file to a directory in your PATH, or copy it to a directory and add that directory to your PATH environment variable.

Additional information for running the preceding tasks can be found [here](#).

5. Verify that the **zassist** utility is working.



```
andrewjones@Andrews-MacBook-Pro watsonxAssistant % PATH=$PATH:.
andrewjones@Andrews-MacBook-Pro watsonxAssistant % zassist
zassist: error: expected one of "version", "init", "login", "ingest", "load", ...
andrewjones@Andrews-MacBook-Pro watsonxAssistant %
```

Ingest client documentation using **zassist**

With the **zassist** command installed, you are now able to begin ingesting data.

Step-by-step guidance for ingesting documents using zassist are provided in the IBM watsonx Assistant for Z documentation [here](#).

These steps are not repeated in this lab guide. The following video illustrates the steps to ingest a single document. This video has no audio.

 **Don't see the video in the PDF version of the lab guide?**



If you are viewing the PDF of the lab guide, you can access the video [here](#).

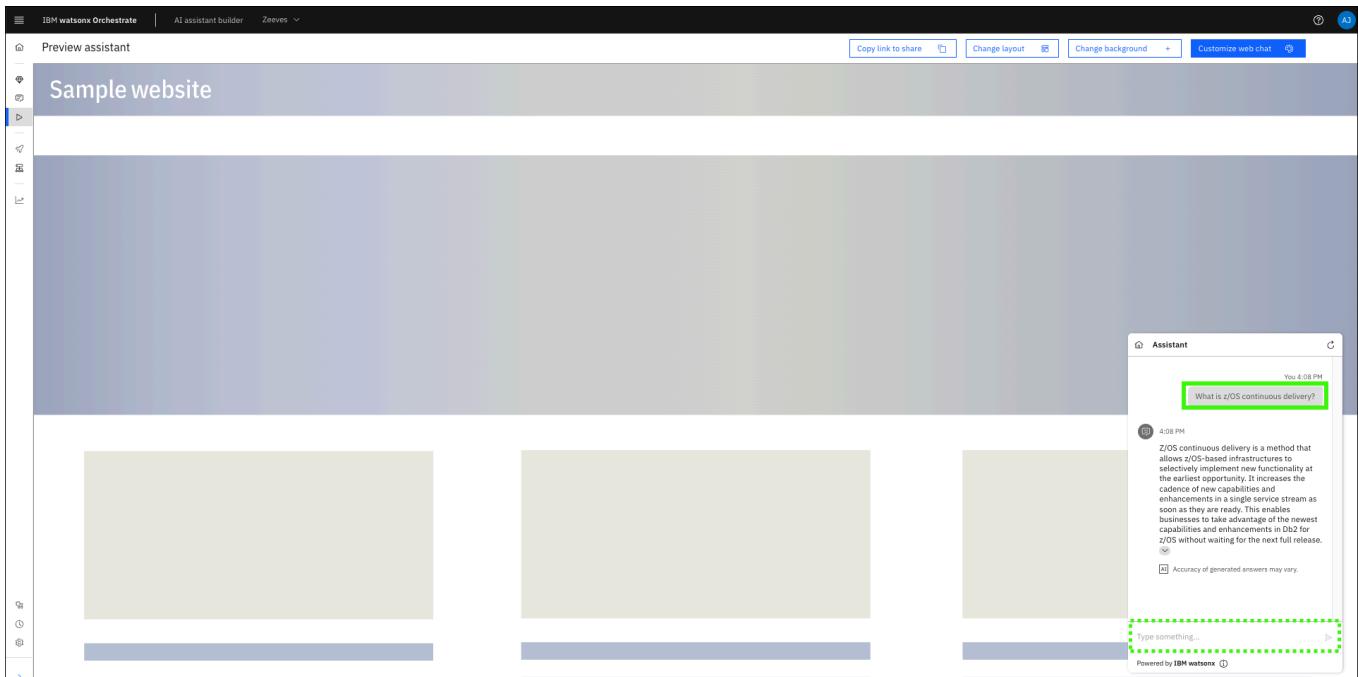
The document that is ingested in the video is a compressed PDF of the **IBM z/OS Continuous Delivery Red Piece**. You can download a copy of this document [here](#).

Verify the document that is ingested is now returned as a source file for a query.

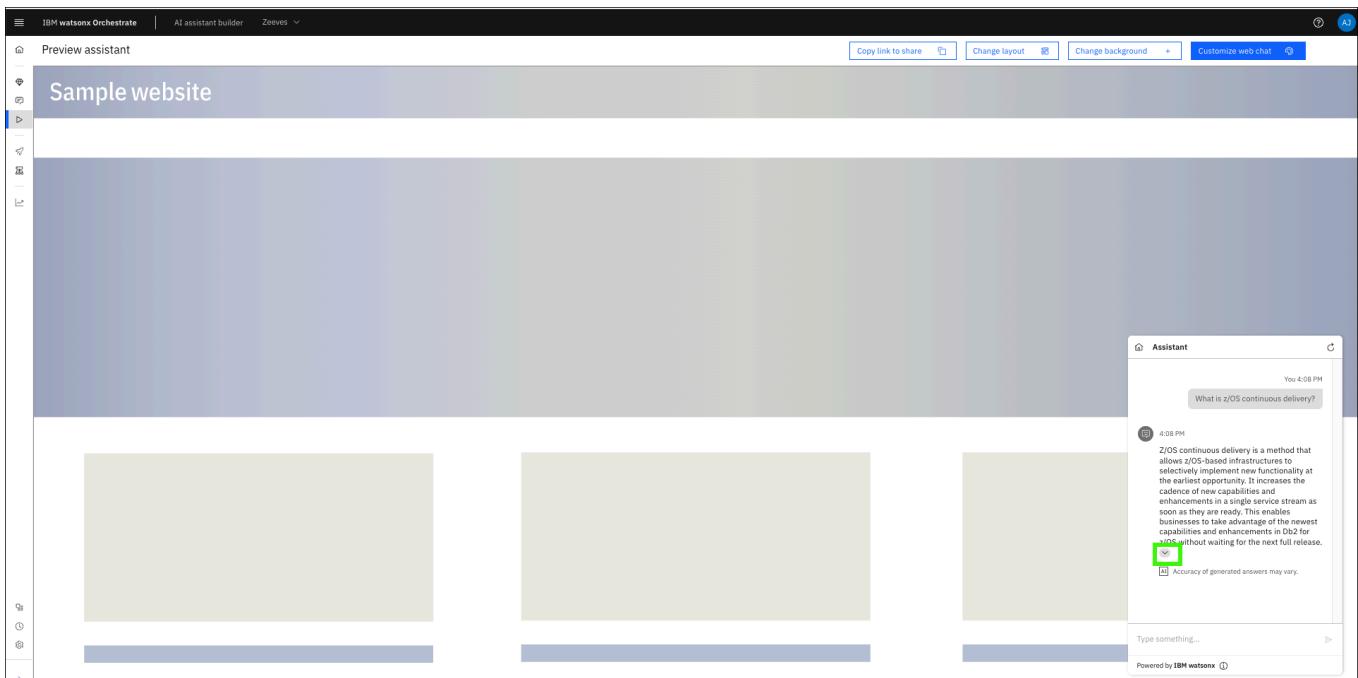
Use the watsonx Orchestrate AI assistant builder to verify your document ingestion.

6. Hover over the home () icon and click **Preview**.
7. Enter the following prompt in your assistant.

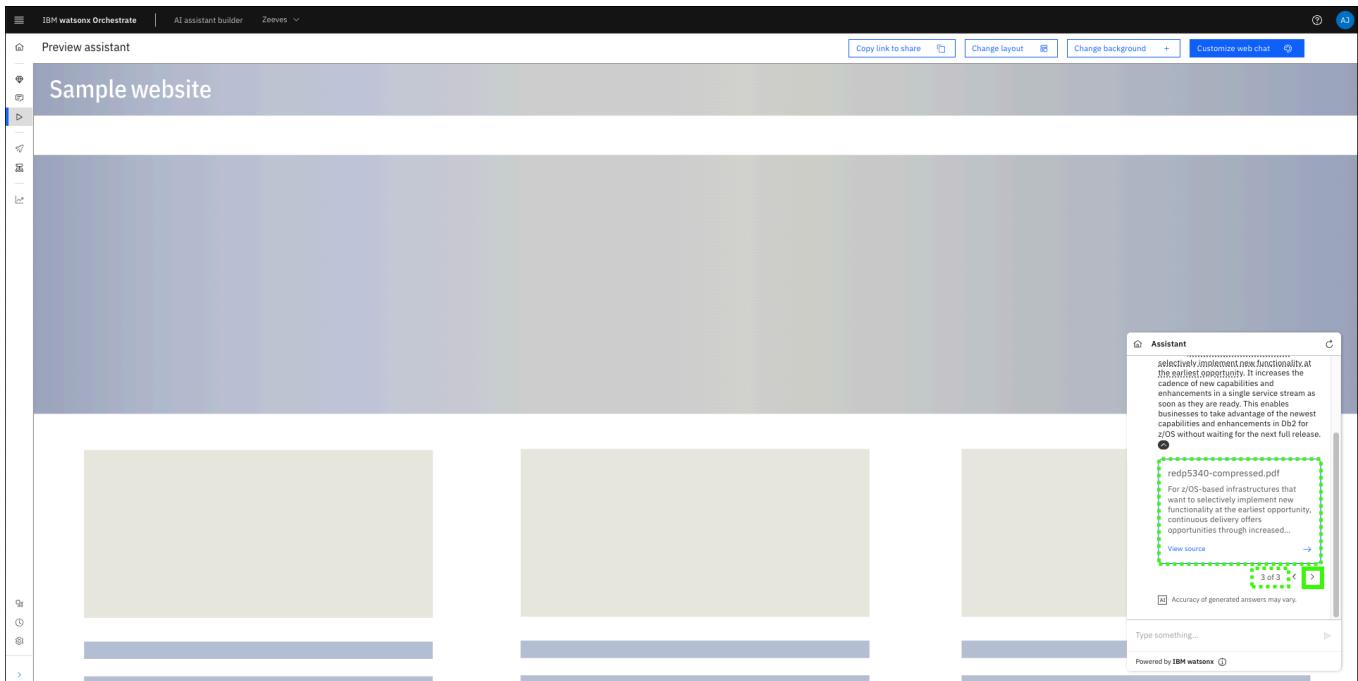
What is z/OS continuous delivery?



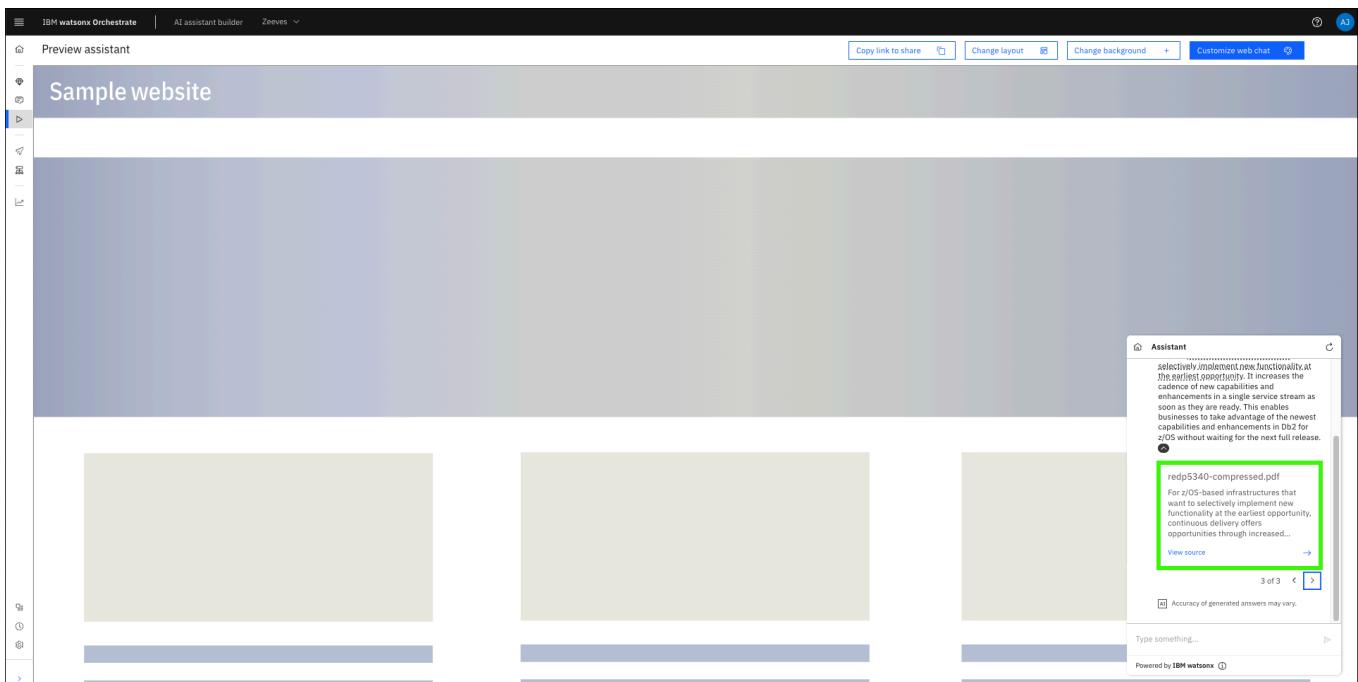
8. Expand the sources section by clicking the (▼).



9. Click through the list of resources and find the reference to the Red Piece document you ingested.

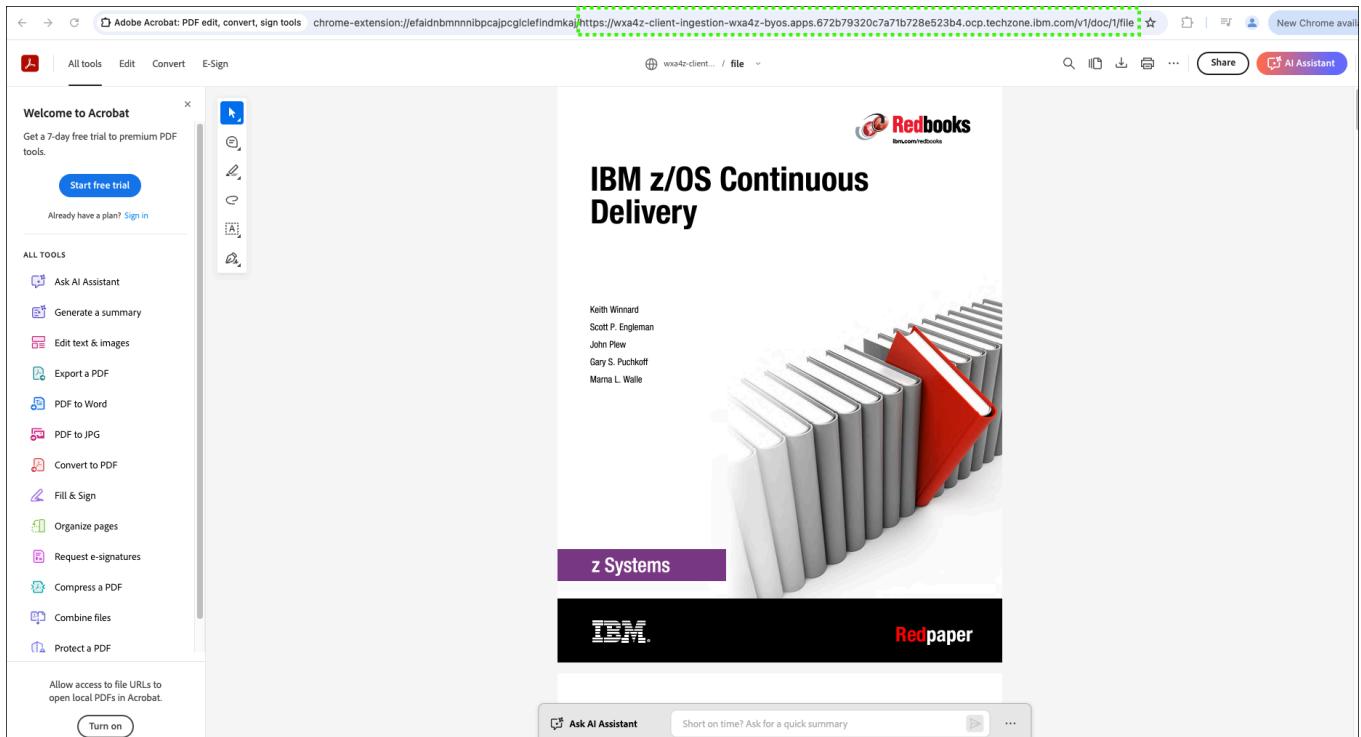


10. Click the ingested document reference.



11. Accept the security risk to view the source document.

The steps to accept the security risk for the document are not shown. The risk occurs because the certificate for the connection to the SNO instance is not secure. Notice that the URL contains the path to your SNO instance route.



Adjusting the search behavior

Do you recall the **Metadata** field when you configured your assistant?

Custom service Draft

Settings Instance Close Save

even if the user has moved on.

Define the text for the citations title. ⓘ

How do we know?

Tendency to say "I don't know":

Rarely	Less often	More often	Most often
--------	------------	------------	------------

Generated response length:

Concise	Moderate	Verbose
---------	----------	---------

Default filter

Can be anything (plain text, JSON, etc)

Metadata

```
JSON example:
{
  "example_field": "example_value",
  "other_example_field": 7
}
```

Search display text

Define the text your search will display to the end user

No results found Connectivity issue
I searched my knowledge base, but did not find anything related to your query

The Metadata field provides a way to adjust your assistant's behavior during conversational search for your OpenSearch instance. Now that you have your own docs that are ingested for conversational search, you can set the metadata field for your assistant to use those documents in its content-grounded search. If you leave the metadata field empty, then it defaults to settings found to perform well. This replaces having to paste a complicated search string. By default (without any string in the Metadata field), it searches all the default IBM provided documentation and all ingested customer documentation using the following value:

```
{"ibm_indices": "*_ibm_docs_slate",
"customer_indices": "customer_*"}
```

Replacing the wildcard string with an explicit list of indices allows personalization. The metadata setting is where you can input specific indices (pointing to the underlying documentation) that you want your assistant to use for the content-grounded search. Out of the box there are over 220 products and topics that the OpenSearch instance has IBM Documentation for. You can find those indices and products [here](#).

You can input a subset of indices into the “Metadata” field in cases where you only want your assistant to gather context for specific IBM products or topics. The specific indices can be listed out in this format:

```
{"ibm_indices": "<comma separated index values>", "customer_indices": "customer_*"}
```

For example, if you only want your assistant to reference documentation for “Db2 Analytics Accelerator for z/OS” and no ingested client documentation, you can enter the following into the metadata field:

```
{"ibm_indices": "ss4lq8_ibm_docs_slate"}
```

If you have a mix of IBM Documentation and client documentation ingested, then there's an optional search string that you can use to set the “weights” used for each.

For example:

```
{"doc_weight": {
"product_docs": 0.5,
"customer_docs": 0.5},
"ibm_indices": "*_ibm_docs_slate",
"customer_indices": "customer_*"
}
```

In this case, “product_docs” is the weight that is assigned to “ibm_indices” and “customer_docs” is the weight that is assigned to “customer_indices”.

After you have configured all the settings for Conversational Search on the page, click “Save” in the upper-right of the page.

For more information on customizing the metadata field for conversational search, refer to this supplemental video found [here](#).

You are encouraged to experiment with the metadata field!

Try setting the metadata field to the following, which weights ingested docs higher than the product docs:

```
{"doc_weight":  
{"product_docs":0.2,  
"customer_docs":0.8},  
"ibm_indicies":"*_ibm_docs_slate",  
"customer_indicies":"customer_**"  
}
```

Now, repeat steps 6 through 8 (make sure you start a fresh preview of the assistant). Notice that the ingested Red Piece document is now the first sited reference!

Guidance for the pilot environment

If you or your client have other documents to ingest, you can do so by repeating the steps using zassist. The Velocity Pilot ITZ environment is limited in compute and storage capacity. The following limits should be adhered to:

- Greater than 15,000 documents is not supported.
- Individual documents should be less than 10 megabytes (MB).
- Larger documents greater than 50 MB should work if the document contains mostly images.
- Loading documents can take a long time, especially with > 100 MB of text.
- It is recommended to run large loads late at night.
- When loading, ensure your workstations does not sleep during the process.
- If you receive a **batch time error**, set the batch size to a lower number for that command. For example:

```
zassist ingest . -s 50
```

After ingesting all the your additional documents, proceed to the next section to learn about adding skills to your assistant.

Adding skills to the assistant

Getting started with skills and actions

Watsonx Assistant for Z can import skills to automate a range of IBM Z related tasks through assistant interactions. Each skill is a pre-defined automation that runs tasks to accomplish some unit or units of work. For example, you can use a skill to view z/OS IPL information, or work with z/OS datasets.

IBM Watsonx Assistant for Z provides an extension within Watsonx Orchestrate to build new skills from Ansible Automation platform or z/OS Management Facility (z/OSMF). The Z Skills Accelerator extension connects Ansible and z/OS application programming interfaces (APIs) and can import automation as Ansible Playbooks, JCL, or REXX as skills. Learn more importing and building skills [here](#).

Environments

Watsonx Orchestrate

The Z Skills Accelerator extension is already configured in your Watsonx Orchestrate IBM Technology Zone (ITZ) environment. You can use this component to import new skills.

Ansible Automation Platform and Wazi as a Service

To import skills for automations, you use Ansible Automation Platform (AAP) and Wazi as a Service (Wazi aaS) to serve as the z/OS back-end. Learn more about AAP [here](#). Learn more about Wazi, [here](#).

The two resources are provisioned together in the ITZ environment that you reserved earlier. This environment enables the ability to manage and automate z/OS tasks and subsystems with various preinstalled ansible playbooks. It includes a z/OS back-end (Wazi as a Service) with all needed prerequisites.

The playbooks provided cover various use cases for automating z/OS management. Ansible's capabilities for automating various Z-specific tasks are not limited to the use cases that are preinstalled in the AAP instance. The preinstalled playbooks are tasks from the 'IBM z/OS core collection'. Using this environment accelerates the ability to showcase the value of Watsonx Assistant for Z, and to get started with simple automations that can be expanded.

The ITZ environment gives you access to AAP, which is preconfigured to target the accompanying z/OS Wazi system, along with web-based access to AAP to experiment with different playbook templates. These templates are imported into Watsonx Orchestrate as skills and connected to your assistant.

For more information on the AAP and Wazi z/OS environments, refer to this document.

The playbook templates that are preinstalled in AAP cover various use cases, which you can explore, including:

- z/OS Certificate Management (create, delete, list, and renew certificates)
- dataset management (create, delete, fetch datasets)
- Submit JCL
- Run Operator commands

- Run TSO commands
- And more

Each of the sections that follow build upon each other. Complete each to successfully enhance your assistant by starting with [Explore Ansible Automation Platform](#).

Explore Ansible Automation Platform

After you reserved the Ansible Automation Platform (AAP) and Wazi z/OS environment in IBM Technology Zone (ITZ) and the reservation is in the **Ready** state, follow these steps to explore AAP.

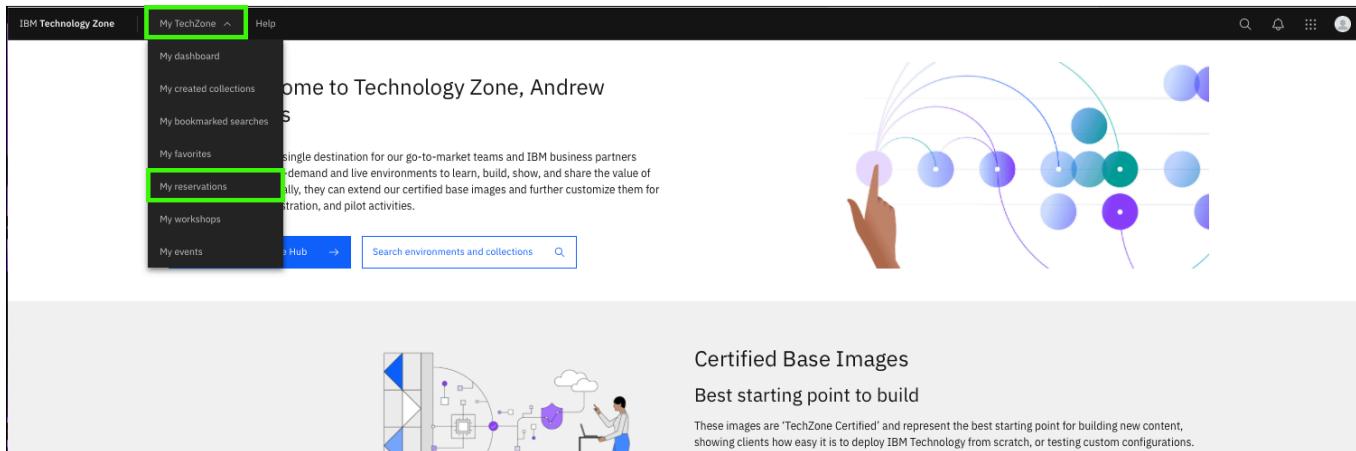
Access the AAP and Wazi as a Service environment

Be sure to record the information as instructed

Several of the steps below instruct you to record values from your ITZ reservation. Be sure to do this as they will not only be used in this section, but also in later sections of the lab guide.

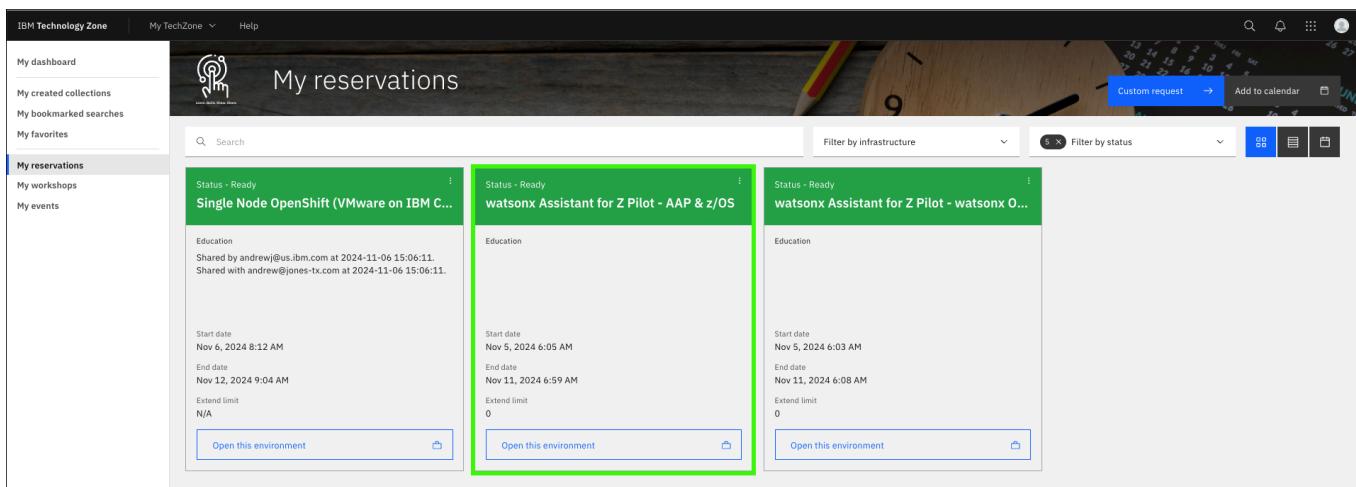
1. In the IBM Technology Zone portal, expand **My TechZone** and select **My Reservations**, or click the following link.

[ITZ My reservations](#)



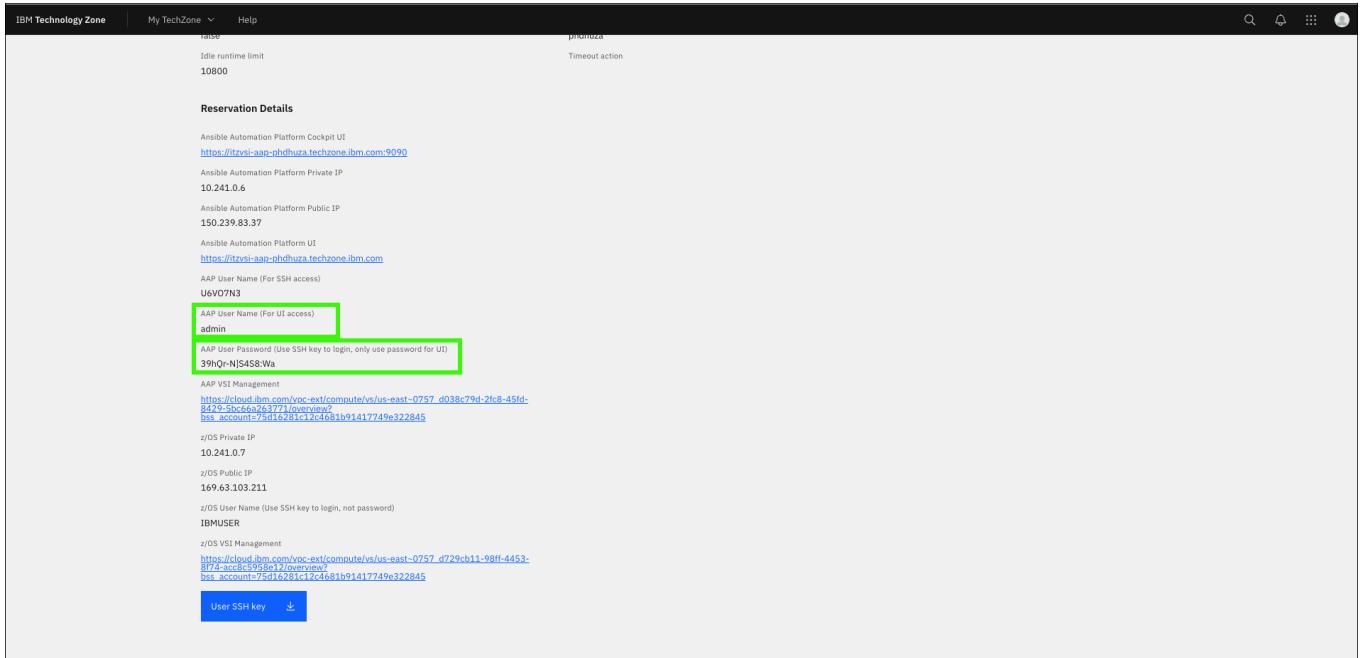
Certified Base Images
Best starting point to build
These images are "TechZone Certified" and represent the best starting point for building new content, showing clients how easy it is to deploy IBM Technology from scratch, or testing custom configurations.

2. Click the **watsonx Assistant for Z Pilot - AAP & z/OS** tile.

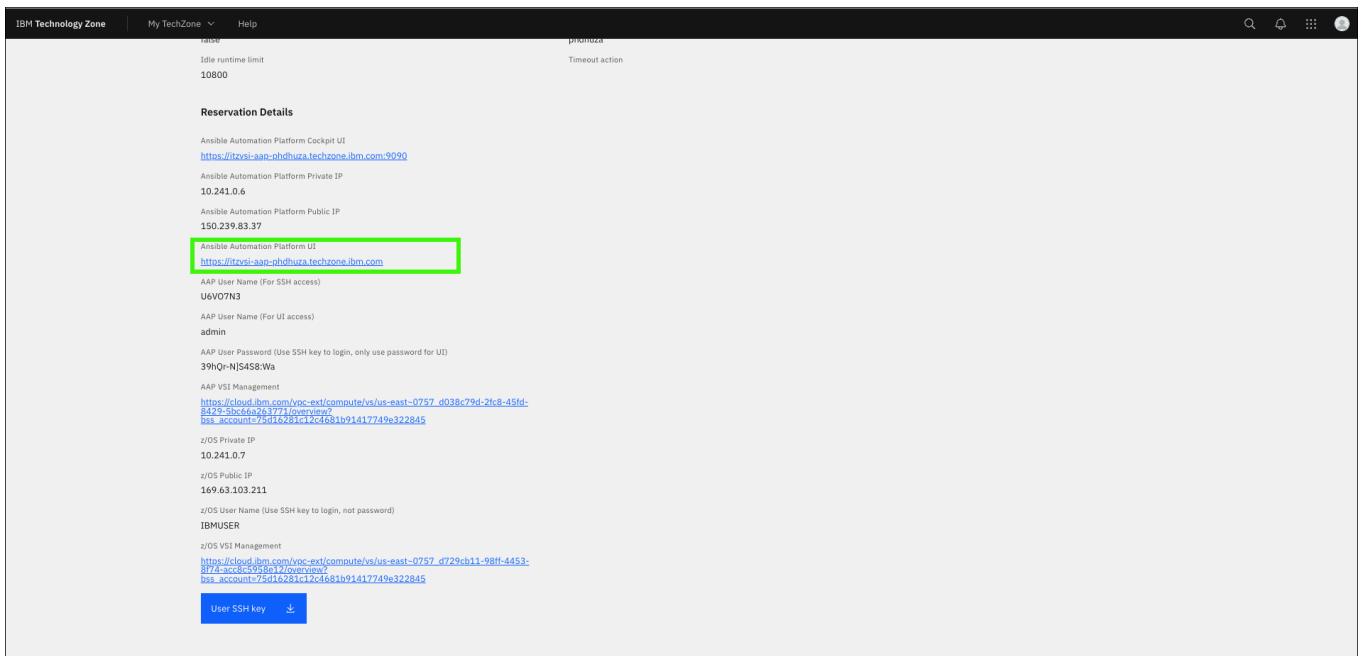


Reservation Type	Status	Start Date	End Date	Extend Limit	Action
Single Node OpenShift (VMware on IBM Cloud)	Ready	Nov 6, 2024 8:12 AM	Nov 12, 2024 9:04 AM	N/A	Open this environment
watsonx Assistant for Z Pilot - AAP & z/OS	Ready	Nov 5, 2024 6:05 AM	Nov 11, 2024 6:59 AM	0	Open this environment

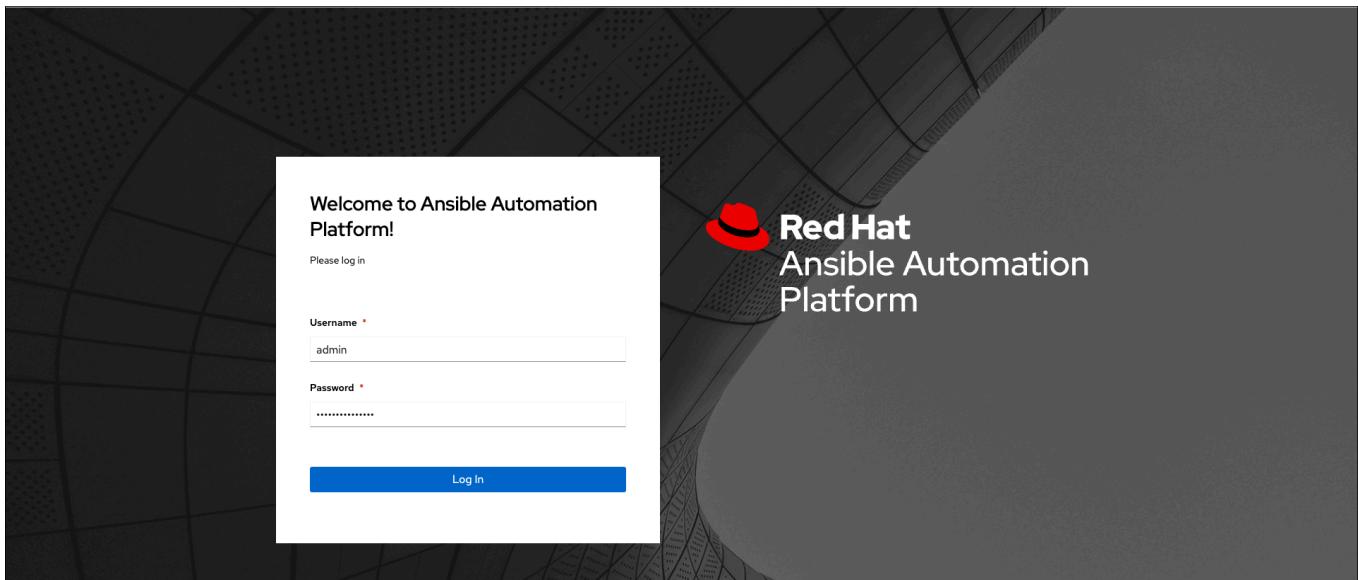
3. Locate and record the **AAP User Name (For UI access)** and **AAP User Password** fields.



4. Record and then click the Ansible Automation Platform UI link.



5. Enter the **Username** and **Password** that is recorded in step 3 and click **Log In**.

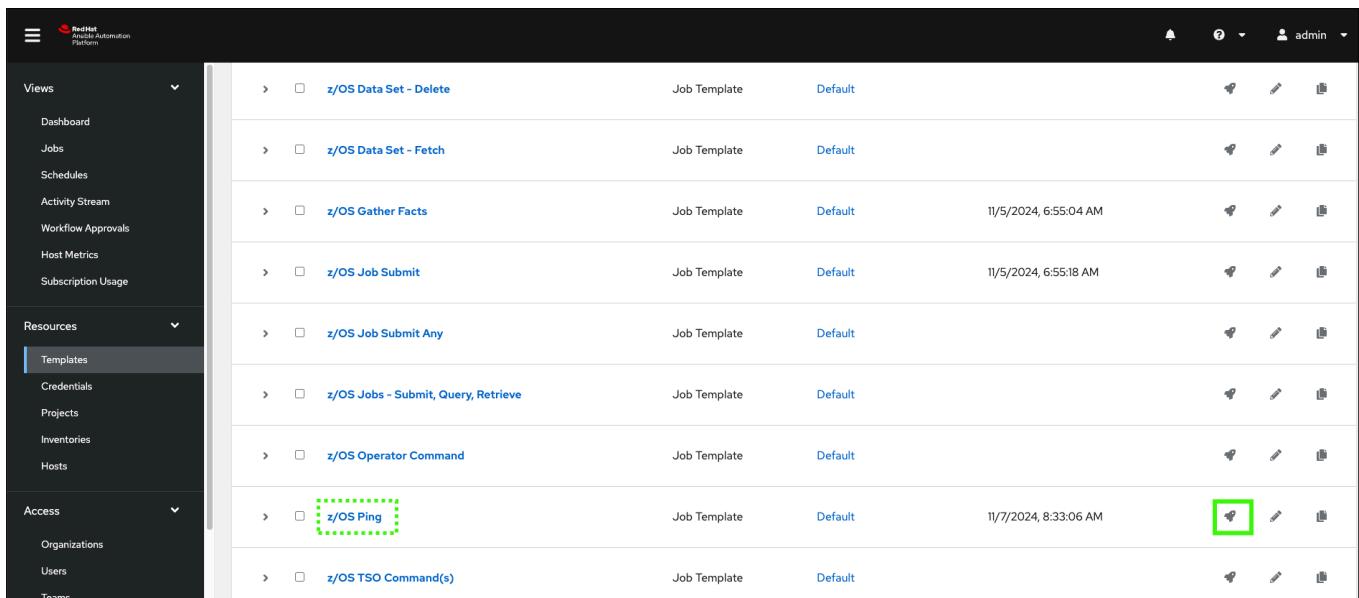


6. Click **Templates** under the **Resources** section.

i **The AAP instance is preconfigured to the Wazi aaS instance**

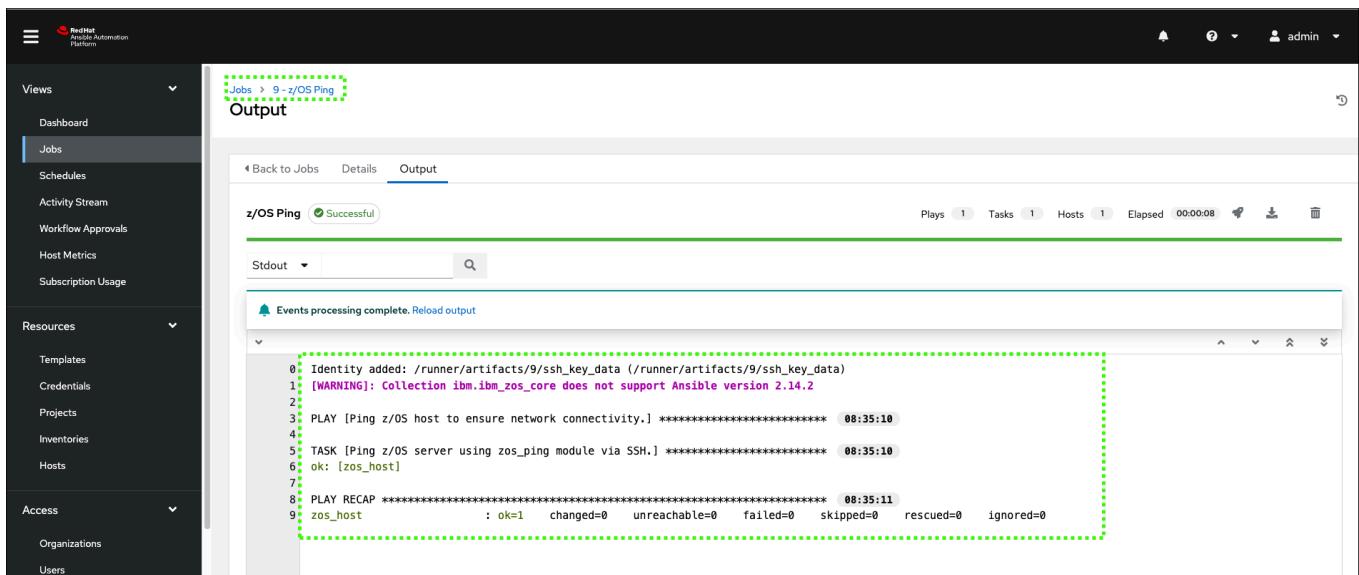
Note that because the AAP instance and the back-end z/OS system are preconfigured, no changes are needed to execute the templates and they will target your provisioned z/OS system by default.

7. Locate the **z/OS Ping** template and click the rocket () icon to start the template.



Views	z/OS Data Set - Delete	Job Template	Default			
Dashboard	z/OS Data Set - Fetch	Job Template	Default			
Schedules	z/OS Gather Facts	Job Template	Default	1/5/2024, 6:55:04 AM		
Activity Stream	z/OS Job Submit	Job Template	Default	1/5/2024, 6:55:18 AM		
Workflow Approvals	z/OS Job Submit Any	Job Template	Default			
Host Metrics	z/OS Jobs - Submit, Query, Retrieve	Job Template	Default			
Subscription Usage	z/OS Operator Command	Job Template	Default			
Resources	z/OS Ping	Job Template	Default	1/7/2024, 8:33:06 AM		
Templates	z/OS TSO Command(s)	Job Template	Default			
Credentials						
Projects						
Inventories						
Hosts						
Access						
Organizations						
Users						
Teams						

8. Observe the z/OS Ping job run.



Jobs > 9 - z/OS Ping

Output

Events processing complete. Reload output

```

0 Identity added: /runner/artifacts/9/ssh_key_data (/runner/artifacts/9/ssh_key_data)
1 [WARNING]: Collection ibm.ibm_zos_core does not support Ansible version 2.14.2
2
3 PLAY [Ping z/OS host to ensure network connectivity.] **** 08:35:10
4
5 TASK [Ping z/OS server using zos_ping module via SSH.] **** 08:35:10
6 ok: [zos_host]
7
8 PLAY RECAP **** 08:35:11
9 zos_host : ok=1    changed=0    unreachable=0    failed=0    skipped=0    rescued=0    ignored=0

```

Take time to explore the other templates that are ready to use. Learn more about the automation capabilities [here](#).

Importing skills from Ansible Automation Platform

Now that you understand Ansible Automation Platform (AAP) and the preinstalled automations available, you can import them as skills into your watsonx Orchestrate instance, which is used for assistant guided actions.

Import skills into your assistant

The next steps assume that you have an active browser window to the watsonx Orchestrate ITZ cloud account. If you do not, refer to the initial steps in [Creating an assistant and configuring conversational search](#).

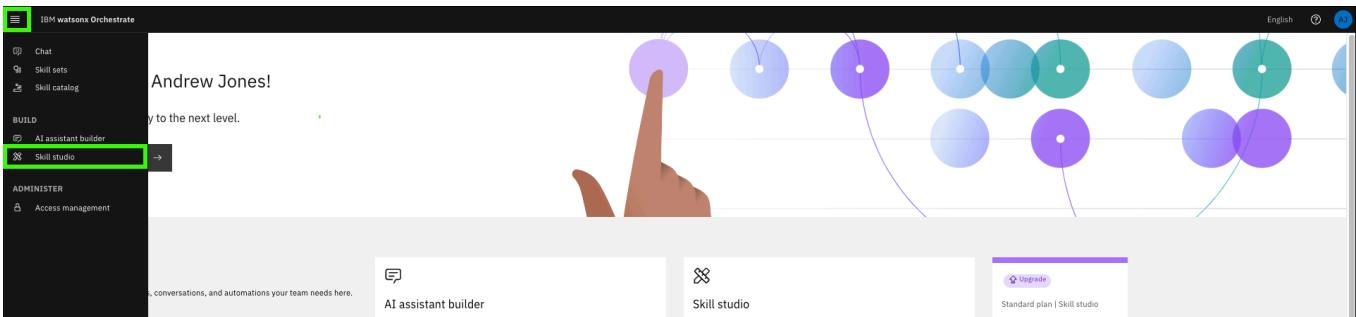
1. Click the **resources** icon (☰).

The screenshot shows the IBM Cloud dashboard with a green box highlighting the 'Build' card in the 'For you' section. The 'Build' card has a blue background and text: 'Explore IBM Cloud with this selection of easy starter tutorials and services.' Below it are other service cards: 'Use Watson Assistant', 'IBM Cloud Essential Security and Observability Services', 'Use Cloud Object Storage', 'Maximo Application Suite', 'Build with Watson', 'Use Watson Studio', and 'Get Started with Studio'. Each card includes a small icon, a title, a brief description, and a status badge (Popular or Recommended).

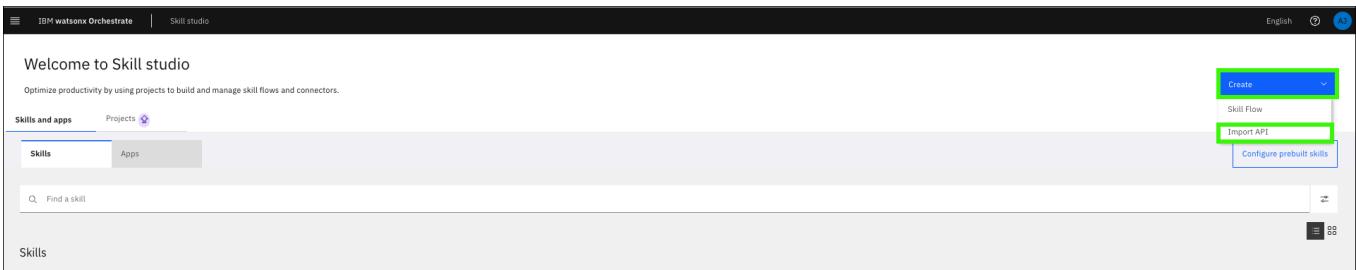
2. Click **Launch watsonx Orchestrate**.

The screenshot shows the 'Resource list' page for the watsonx instance 'itzwxo-2700039nft-wusdf'. The 'Manage' tab is selected. On the left, there's a sidebar with 'Service credentials'. The main area has a 'Start by launching the tool' section with a 'Launch watsonx Orchestrate' button (highlighted with a green box) and a 'Getting started tutorial' link. Below that is a 'Credentials' section showing an API key and a URL. To the right, there's a 'Plan' section titled 'Essentials Plan'.

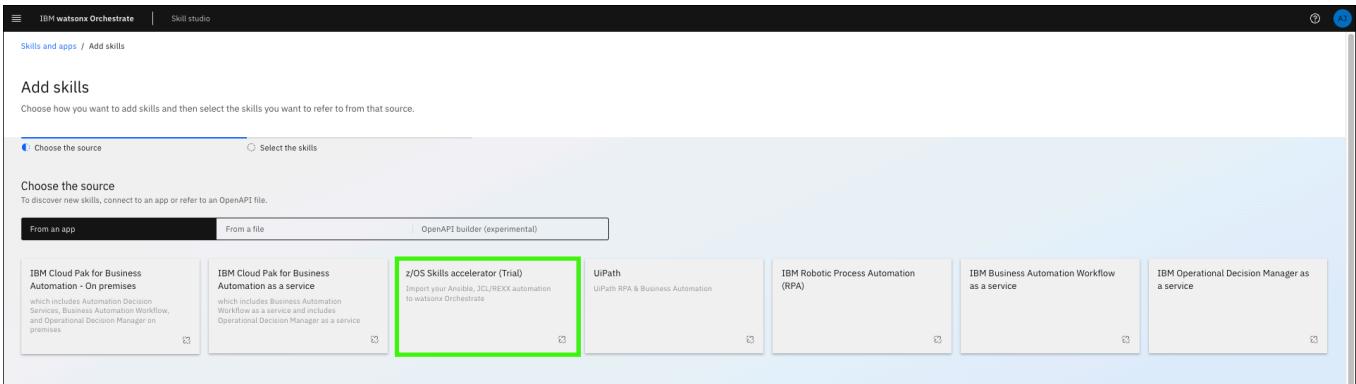
3. Click the main menu (☰) and then click **Skill studio**.



4. Expand Create and click Import API.



5. Click the z/OS Skills accelerator (Trial) tile.



6. Enter the following values in the z/OS Skills accelerator form and then click Connect.

Use the **URL**, **User Name**, and **Password** values recorded in the [Explore Ansible Automation Platform](#) section earlier.

a: Connection Type: **ansible**

b: Application Name: <enter a meaningful name for the skills that you will import>

c: Connection URL: <enter the URL for your AAP UI>

d: User Name: <enter the AAP User Name (for UI access)>

e: Password: <enter the AAP User Password>

f: Search Pattern: *

Add skills

Choose how you want to add skills and then select the skills you want to refer to from that source.

Choose the source Select the skills

Choose the source

To discover new skills, connect to an app or refer to an OpenAPI file.

From an app From a file OpenAPI builder (experimental)

IBM Cloud Pak for Business Automation - On premises which includes Automation Decision Services, Business Automation Workflow, and Operational Decision Manager on-premises	IBM Cloud Pak for Business Automation as a service which includes Business Automation Workflow as a service and includes Operational Decision Manager as a service	z/OS Skills accelerator (Trial) Import your Ansible, JCL/REXX automation	UiPath UiPath RPA & Business Automation	IBM Robotic Process Automation (RPA)	IBM Business Automation Workflow as a service
--	---	--	--	--------------------------------------	---

Connection Type: ansible

Either 'ansible' or 'zos'

Application Name: z skills

Short application name that uniquely identifies the connection url. Use the same name you previously used if you are importing additional skills from the same connection. Must contain only alphanumeric characters, spaces, and underscores.

Connection URL: https://itervsap-phdhubz.techzone.ibm.com

For example https://www.google.com

User name: admin

Password: password

Search Patterns: z

Necessary for 'zosim' connections. Pattern cannot be '*' or '**'.

Cancel Connect

TBD

TBD

TBD

TBD

TBD

Publishing and deploying your assistant

To this point, acting as an Assistant Builder, you have built out the assistant, configured conversational search, and added skills and automations. And while doing so, you have been testing your assistant using the **preview** capability of AI Assistant Builder. The **preview** capability is a closed environment for experimenting with prompts.

Adding other integrations

Other integrations can be added to your assistant to allow you to publish your assistant to the channels where your end-users are likely to interact with the assistant.

To do so, channel integrations must be added. By default, a web chat integration is created, allowing an assistant to be embedded in a website. Other channel integrations are available in the Integrations catalog inside the AI assistant builder interface. Learn more about adding integrations [here](#).

Within the AI assistant builder, click on the Integrations page:

Scroll down to **Channels** to view the integration deployment options for your assistant.

Clicking on any of the available channels will provide guidance on how to configure the integration.

Next steps

This IBM watsonx Assistant for Z for Technical Sales Level 4 demonstrated several of the IBM watsonx Assistant for Z capabilities. However, there is still more to learn about IBM watsonx Assistant for Z. Be sure to check out the [offering landing page](#) and the [product documentation](#).

Also, the focus of the IBM watsonx Assistant for Z for Technical Sales Level 4 Lab Guide was the *end-user* use of IBM watsonx Assistant for Z. To learn more about deploying, configuring, and customizing IBM watsonx Assistant for Z explore the other capabilities that are included in the [ITZ collection](#) [here](#). Included are guidance for running client pilots and the ability to manage an IBM watsonx Assistant for Z deployment. In the Velocity Pilot environment, you can explore all the capabilities of the offering like creating an assistant, importing documents, and adding automation skills.

Earn the badge

Finally, remember to complete the IBM watsonx Assistant for Z for Technical Sales Level 4 learning plan to earn the IBM watsonx Assistant for Z Technical Sales Advanced badge.

IBM technical sellers must complete a Stand and Deliver exercise to validate their ability to demonstrate IBM watsonx Assistant for Z to potential clients. Refer to the evaluation checklist in the [YourLearning learning plan](#) for the Stand and Deliver.

Business Partners must pass the IBM watsonx Assistant for Z for Technical Sales Level 4 quiz that is located in the [IBM Training learning plan](#). The Business Partner quiz verifies use of the demonstration environment and click-through demonstrations, and not one's general knowledge of the offering.