

# 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)
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 & configuring conversational search

- Access the
- Create your Assistant
- Configure conversational search
- Additional configuration
- Configure the base Large Language Model (
- Testing conversational search
- Troubleshooting

## Ingesting client documents

### Creating a standalone 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 machine
  - Create a working directory
  - Install IBM Certificate Manager on 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
  - Install the
- Troubleshooting

### Installing and using zassist to ingest client documents

- Install the zassist utility
- Ingest client documentation using

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

Adding skills to the assistant

### Getting started with skills and actions

- Environments
  - watsonx Orchestrate
  - Ansible Automation Platform (

### Explore Ansible Automation Platform (

- Access the

### Importing skills from Ansible Automation Platform

- Import skills into your assistant

## 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 leverages 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](mailto: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 related to the software as a service (SaaS) instance of watsonx Orchestrate
- Email: [andrewj@us.ibm.com](mailto:andrewj@us.ibm.com); [maxwell.g.weiss@ibm.com](mailto:maxwell.g.weiss@ibm.com)

Business Partners should 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



#### Search

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,  $\text{[ } \wedge \text{Ctrl} \text{]} + \text{[V]}$  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 Demonstration 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:

**IBM watsonx Assistant for Z for Technical Sales Level 3 Demonstration Guide**

Search

Welcome

Welcome to the **IBM watsonx Assistant for Z for Technical Sales Level 3** course and demonstration guide! The goal of this course is to provide IBM and Business Partner technical sellers the knowledge and tools to demonstrate **IBM watsonx Assistant for Z**.

This course is not intended to serve as an in-depth study of IBM watsonx Assistant for Z, watsonx, or IBM Z; rather the focus is on learning how to deliver a client demonstration of IBM watsonx Assistant for Z. More specifically, the focus is on the *end-user* creating queries and by using trusted automations through skills that are suggested by IBM watsonx Assistant for Z.

Other aspects of IBM watsonx Assistant for Z including creating and configuring assistants, importing and creating new skills, and ingestion of client documentation are **not** covered in this demonstration guide. To learn more and get hands-on with these additional capabilities, check out the Velocity Pilot information in IBM Technology Zone (ITZ) [here](#).

Obtaining the IBM watsonx Assistant for Z Technical Sales Intermediate badge

IBM watsonx Assistant for Z for Technical Sales Level 3 Demonstration Guide

Search

Welcome

Welcome to the **IBM watsonx Assistant for Z for Technical Sales Level 3** course and demonstration guide! The goal of this course is to provide IBM and Business Partner technical sellers the knowledge and tools to demonstrate **IBM watsonx Assistant for Z**.

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

IBM watsonx Assistant for Z for Technical Sales Level 3 Demonstration...

Search

Welcome

Welcome to the **IBM watsonx Assistant for Z for Technical Sales Level 3** course and demonstration guide! The goal of this course is to provide IBM and Business Partner technical sellers the knowledge and tools to demonstrate **IBM watsonx Assistant for Z**.

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 obtaining the IBM watsonx Assistant for Z Technical Sales Advanced badge.

# IBM Technology Zone environment

To enable sellers to both learn how to perform client pilots of IBM Watsonx Assistant for Z, multiple environments have been created in IBM Technology Zone (ITZ). The environments leveraged 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 tenant of Watsonx Orchestrate on IBM Cloud and is leveraged for creating and configuring the assistant, configuring conversational search, importing skills, and configuring assistant actions.
- **Ansible Automation Platform (AAP) & z/OS** – a pattern used to deploy a pre-configured instance of AAP and Wazi z/OS with pre-loaded Ansible playbooks that can be imported as skills within Watsonx Orchestrate and connected to your assistant. Provides pre-loaded templates for various use cases which will be 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 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 learning.

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 this section 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](#)

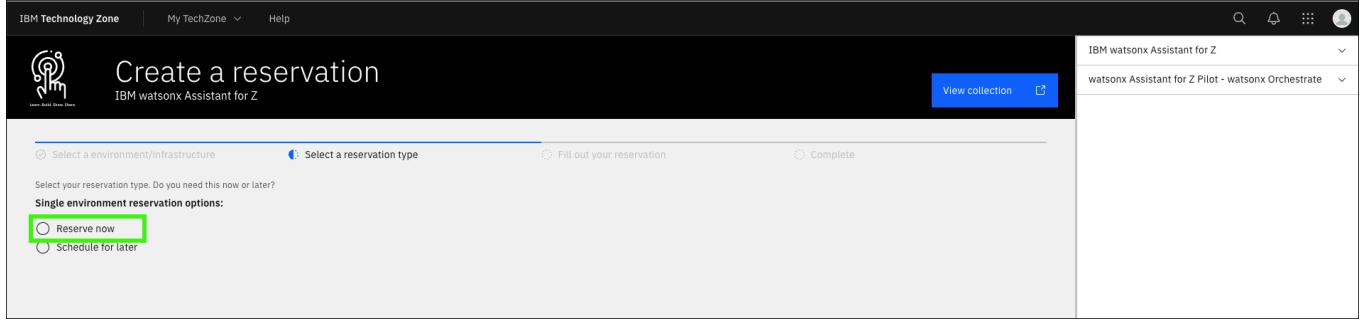
## Single Node OpenShift with NFS storage - reservation page

 Images below 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 will be at your client's office to start a pilot.



The screenshot shows the 'Create a reservation' interface. At the top, there are tabs for 'IBM Technology Zone', 'My TechZone', and 'Help'. On the right, there are navigation icons for search, notifications, and more. The main title is 'Create a reservation' with a subtitle 'IBM watsonx Assistant for Z'. Below the title, there are four steps: 'Select a environment/infrastructure', 'Select a reservation type', 'Fill out your reservation', and 'Complete'. Under 'Select a reservation type', it says 'Select your reservation type. Do you need this now or later?'. It lists 'Single environment reservation options': '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 will be similar to the first image below and have fields **a-e** that will 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 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: Education (Gaining experience with specific technology, product, or solutions)

**c.** Sales Opportunity number: Enter an opportunity number (e.g., [IBM Sales Cloud opportunity ID](#), [Project ID](#), or [Project Work ID](#))

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

**e.** End date and time: 11/01/2024 at 11:51 AM America/Chicago

**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](#)  Submit

In addition to the above fields, the reservation for the **Single Node OpenShift with NFS storage** will have 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

**m.** Notes: Enter any notes you would like to attach to this reservation

**n.** I agree to IBM Technology Zone's [Terms & Conditions](#) and [End User Security Policies](#)  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 below. If issues continue, open an ITZ support ticket using the methods mentioned in the [Support](#) section of this guide.

## Extend the reservation

When the reservations are in the **Ready** state, you can extend each reservation to a total of 6 days. Remember, IBM sellers need the environment to record their Stand and Deliver and Business Partners need an environment to answer quiz questions. Plan your time accordingly.

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

Welcome to Technology Zone, Andrew

Certified Base Images

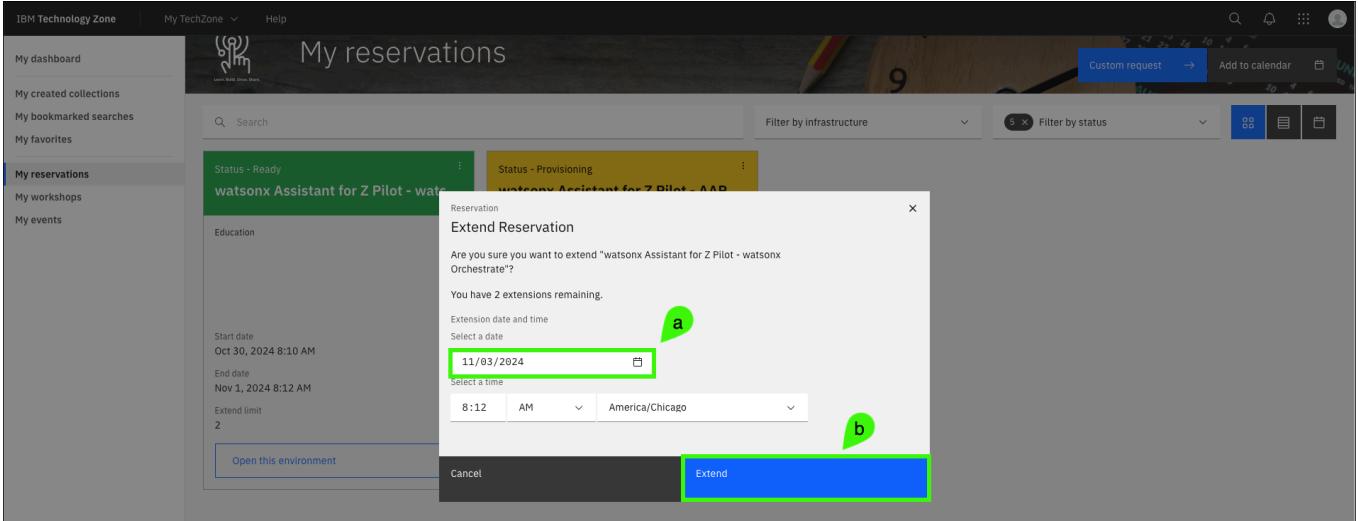
Best starting point to build

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.

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

Reservation Status	Reservation Name	Start Date	End Date	Extend Limit
Status - Ready	watsonx Assistant	Oct 30, 2024 8:10 AM	Nov 1, 2024 8:12 AM	2
Status - Provisioning	watsonx Assistant for Z Pilot - AAP ...	Oct 30, 2024 8:12 AM	Nov 1, 2024 8:10 AM	N/A

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

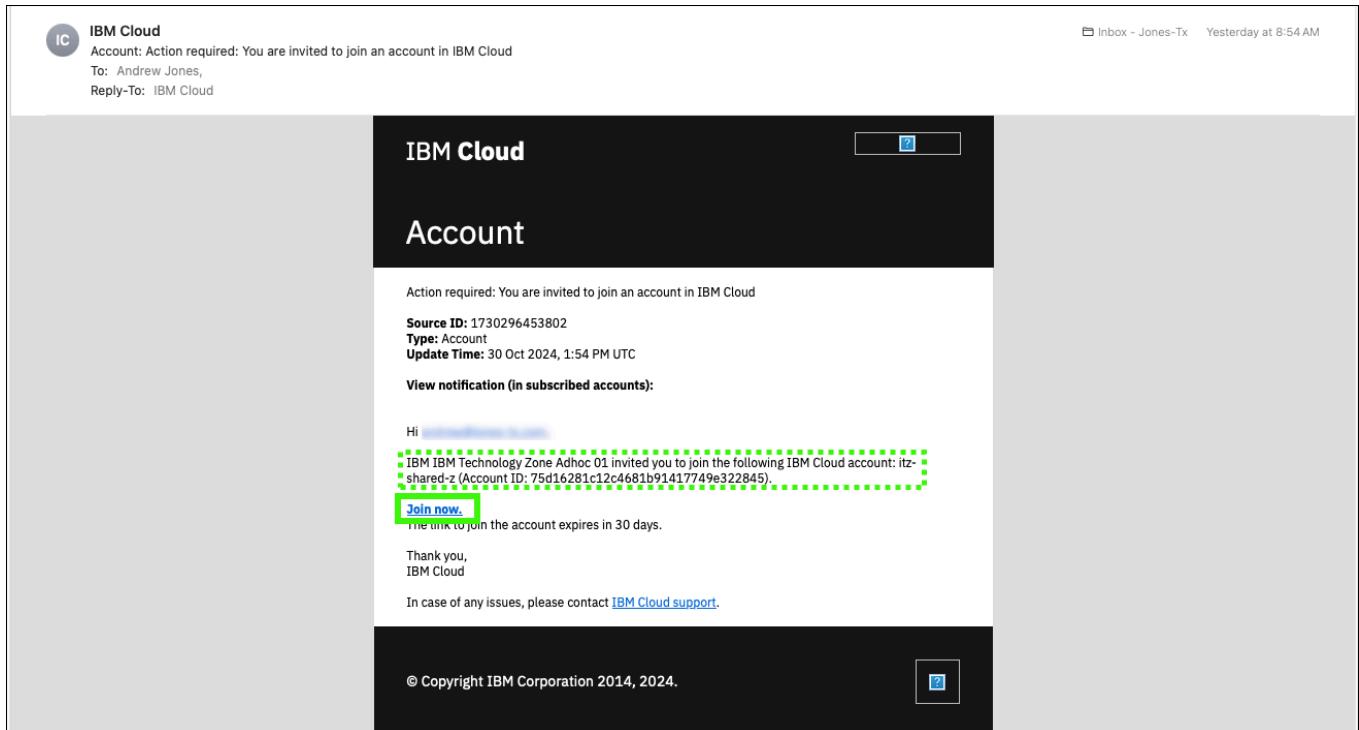


If you anticipate needing more time, repeat Steps 5 and 6 to extend the reservation to the maximum of 6 days. Repeat these steps for the other two reservations.

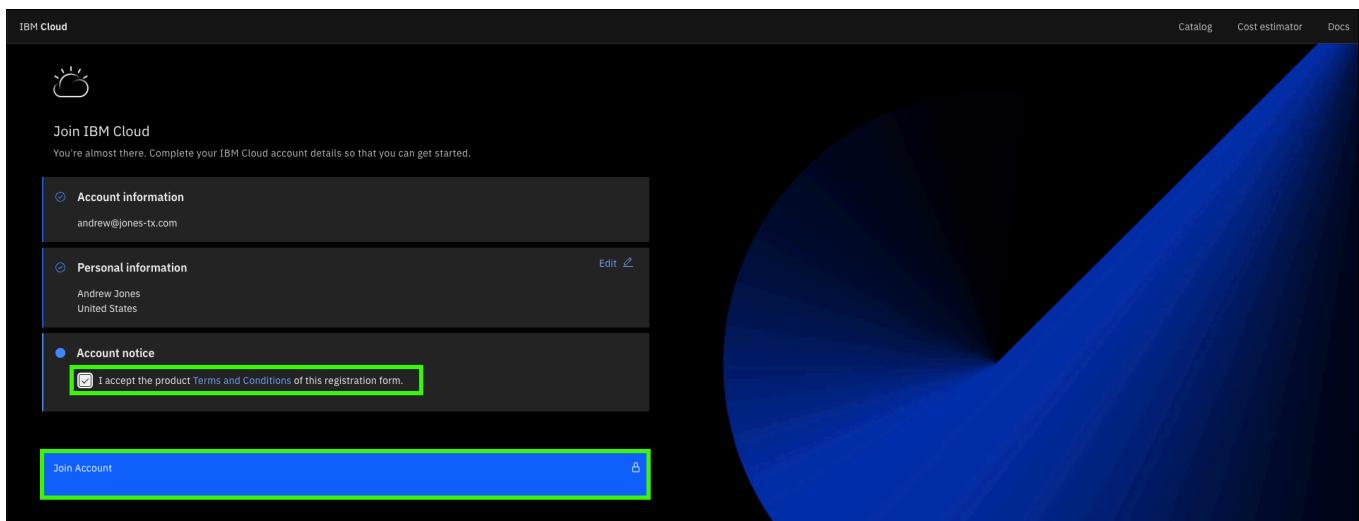
## Join the ITZ IBM Cloud accounts

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 should receive two emails from IBM Cloud. In order 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\*\***.



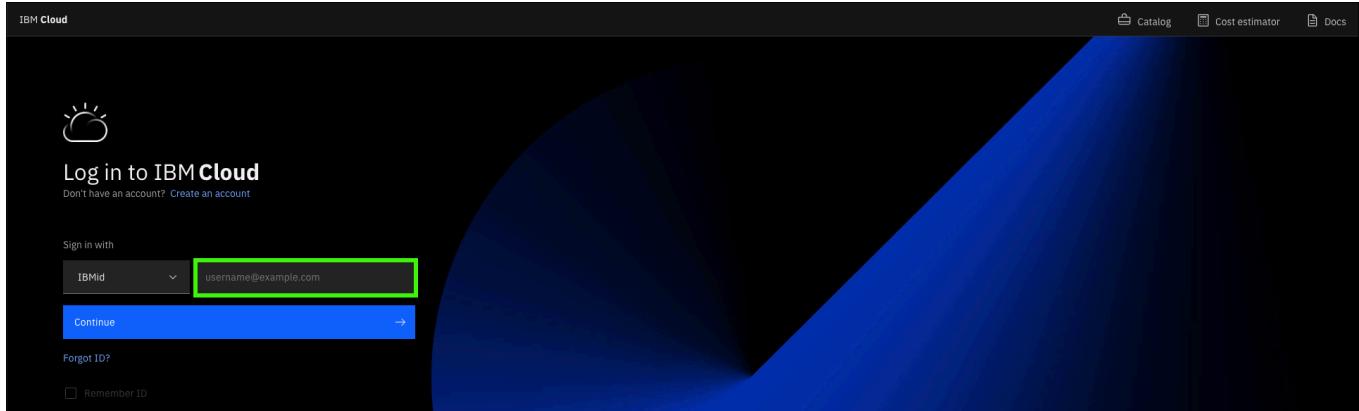
**Repeat steps 7 and 8 for the second invitation.**

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

9. Click the link below to open a browser to the IBM Cloud Portal.

### [IBM Cloud portal](#)

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



11. 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.

Once all three reservations are in the **Ready** state and you have 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 & configuring conversational search

This section will cover how to begin using [watsonx Orchestrate](#) to create a new assistant for watsonx Assistant for Z and configure conversational search. You will be able to configure your assistant to use conversational search using a hosted [OpenSearch](#) instance. The pre-configured instance in IBM Technology Zone (ITZ) has over 220 knowledge sources and supports the Retrieval Augmented Generation (RAG) in which the Large Language Model (LLM) providing the conversational AI is augmented by this knowledge based on IBM Z documentation. All of which helps 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** at the top and select **My Reservations**, or click the link below.

## ITZ My reservations

Welcome to Technology Zone, Andrew

Certified Base Images

Best starting point to build

These images are 'TechZone Certified' and represent the best starting point to build 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 - watsonx Orchestrate** tile.

Status	Title	Start Date	End Date	Action
Status - Ready	watsonx Assistant for Z Pilot - AAP & z/OS	Nov 5, 2024 6:05 AM	Nov 11, 2024 6:59 AM	<a href="#">Open this environment</a>
Status - Ready	watsonx Assistant for Z Pilot - watsonx O...	Nov 5, 2024 6:03 AM	Nov 11, 2024 6:08 AM	<a href="#">Open this environment</a>
Status - Ready	Single Node OpenShift (VMware on IBM C...	Nov 5, 2024 12:58 PM	Nov 7, 2024 5:39 PM	<a href="#">Open this environment</a>

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

**Purpose**

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

**Environment**

- Reservation ID: 672a09a1a8f85062f891e081
- Type: IBM Cloud
- Request method: Watsonx-orchestrate
- Cloud Account: ITZ-WATSONX-036 (highlighted with a green dashed box)
- Region: us-south
- Customer data: false
- Idle runtime limit: 10800
- Transaction ID: 115897c9-58a1-4f17-af9c-b16dc2a97590
- Geo: americas
- Datacenter: dal10
- Environment: watsonx-orchestrate-wusdf
- 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: 672a09a1a8f85062f891e081
- Type: IBM Cloud
- Request method: Watsonx-orchestrate
- Cloud Account: ITZ-WATSONX-036 (highlighted with a green dashed box)
- Region: us-south
- Customer data: false
- Idle runtime limit: 10800
- Transaction ID: 115897c9-58a1-4f17-af9c-b16dc2a97590
- Geo: americas
- Datacenter: dal10
- Environment: watsonx-orchestrate-wusdf
- Timeout action

**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 the current IBM Cloud account is the same as the account name/number recorded in step 3, if not the same, switch to the proper account.

Note: the formatting of the name may appear differently than what was 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 may 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 will be different than shown in the image below).

Resource list

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

## 8. Click Launch watsonx Orchestrate.

itzwxo-2700039nft-erspw Active Add tags

Manage Service credentials

Start by launching the tool

**Launch watsonx Orchestrate**

Credentials

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

URL: <https://api.us-south.assistant-builder.watson.cloud.ibm.com/instances/a7675d8e-e885-4891>

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

Your assistant name will be kept internally and not visible to your customers

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

40/128

Assistant language  
English (US)

This is the language your assistant will speak.

**Next**

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

Explore the personalization options. When 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  
You may add multiple channels from your dashboard.

Where do you plan on deploying your assistant?  
**a** Web

Tell us about yourself  
This information will be used to personalize your onboarding experience.

Which industry do you work in?  
**b** Software

What is your role on the team building the assistant?  
**c** Developer

Which statement describes your needs best?  
**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... ▶

**Back** **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  
Enable IBM Watermark  
 On

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.

[Copy link to share](#) [Change layout](#) [Customize web chat](#)

Sample website

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

At this point, the assistant has been created.

The screenshot shows the IBM Watsonx Orchestrate interface for the 'Zeeves' assistant. 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 section titled 'Enhance your assistant' with a list of recommendations: '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 note says 'Further improve and customize your assistant with these recommendations.' To the right of these is a detailed 'Assistant architecture' diagram. This diagram shows a central node 'Default behavior' branching into 'General purpose' (granite-3-8b-instruct) and 'Conversational search' (No content). It also branches to 'Actions' (View →), which further branches into 'AI-guided' (0 actions), 'Skill-based' (0 actions), and 'Custom-built' (0 actions). Another branch from 'Default behavior' goes to 'Fallback' (Add →), which then leads to 'Live agent' (No agent). There's also a direct connection from 'Zeeves' to 'Actions'. The diagram is labeled '1/2' at the bottom right.

## Configure conversational search

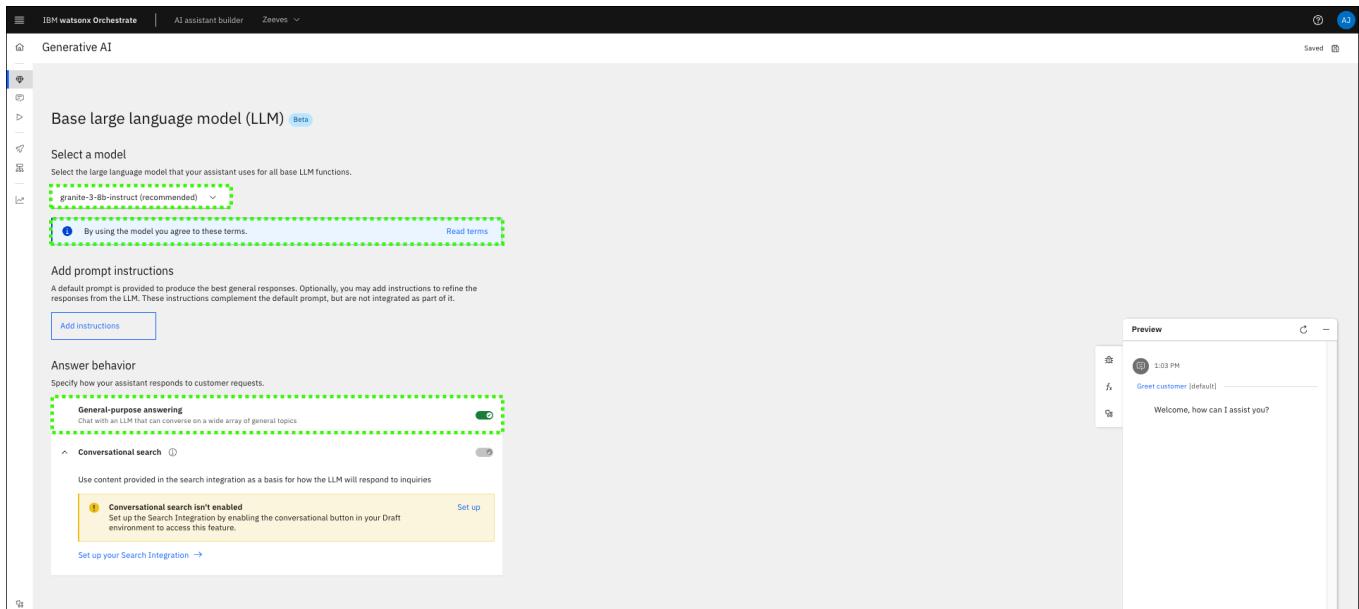
The next step will be to configure **conversational search** for your assistant using a hosted instance of OpenSearch.

14. 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.

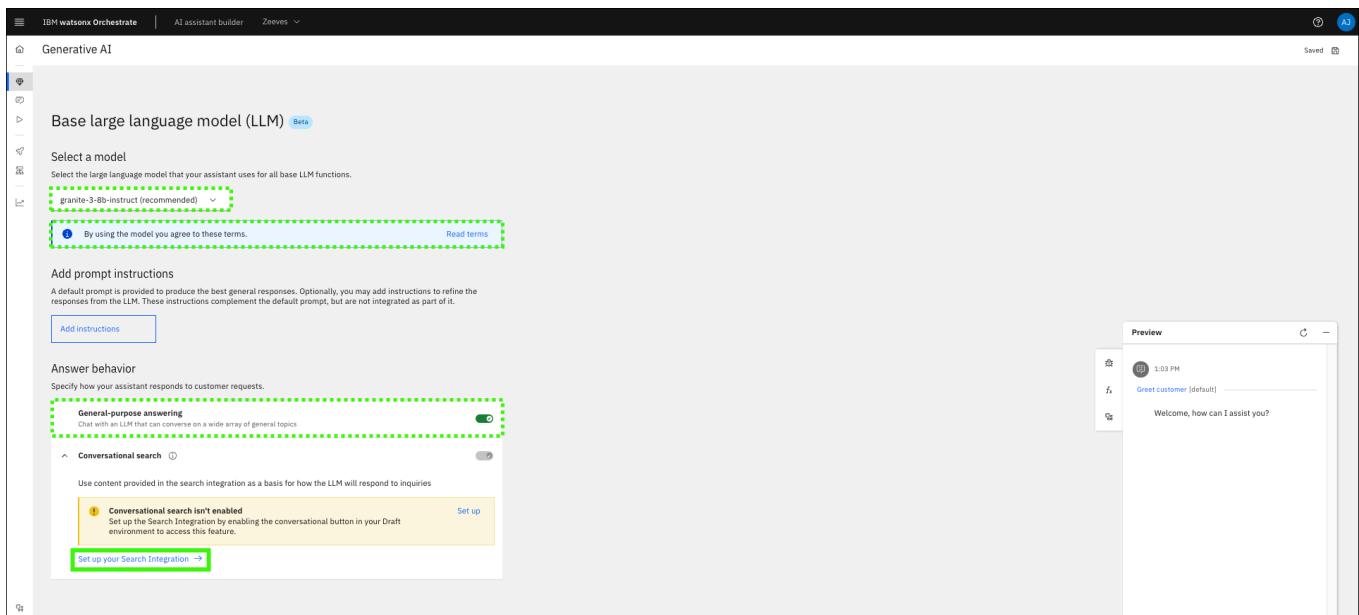
15. Review the base large language model (LLM) settings.

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 will take 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 upon your requirements.**

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

The following settings have 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.**

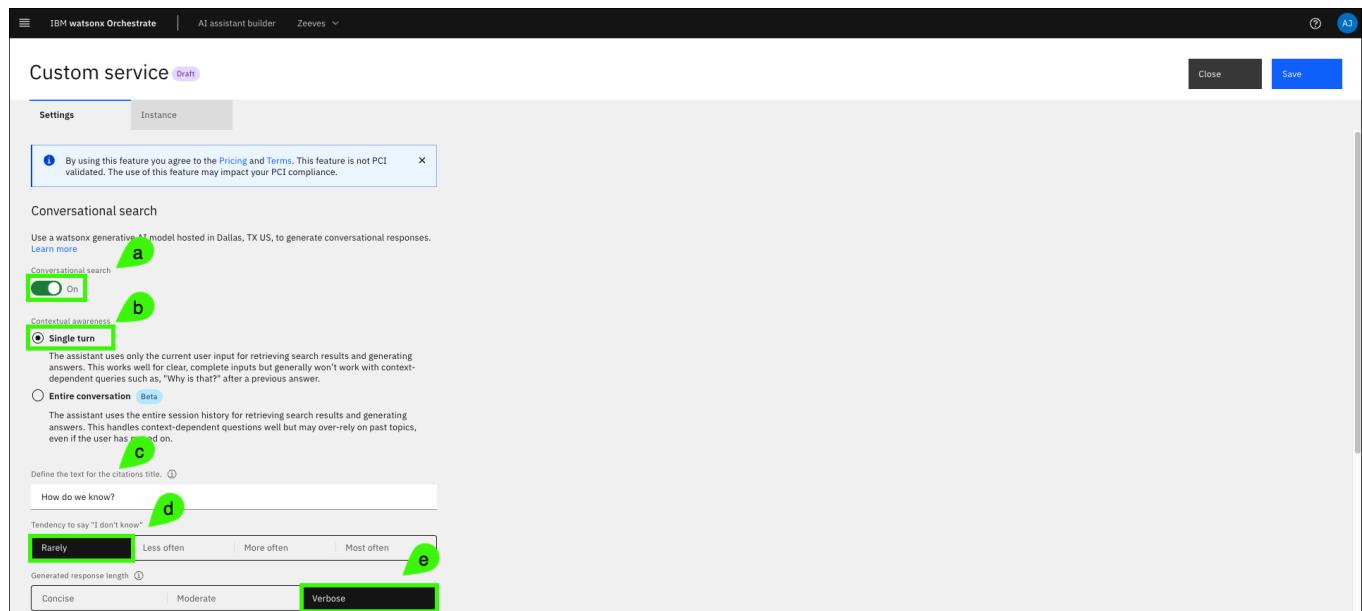
**b. Select Single turn.** Enabling multi-turn conversation (by selecting Entire conversation) is not yet supported for the solution on-prem. Please be mindful in using this option and ensure the client understands what is currently supported today by the solution.

**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.

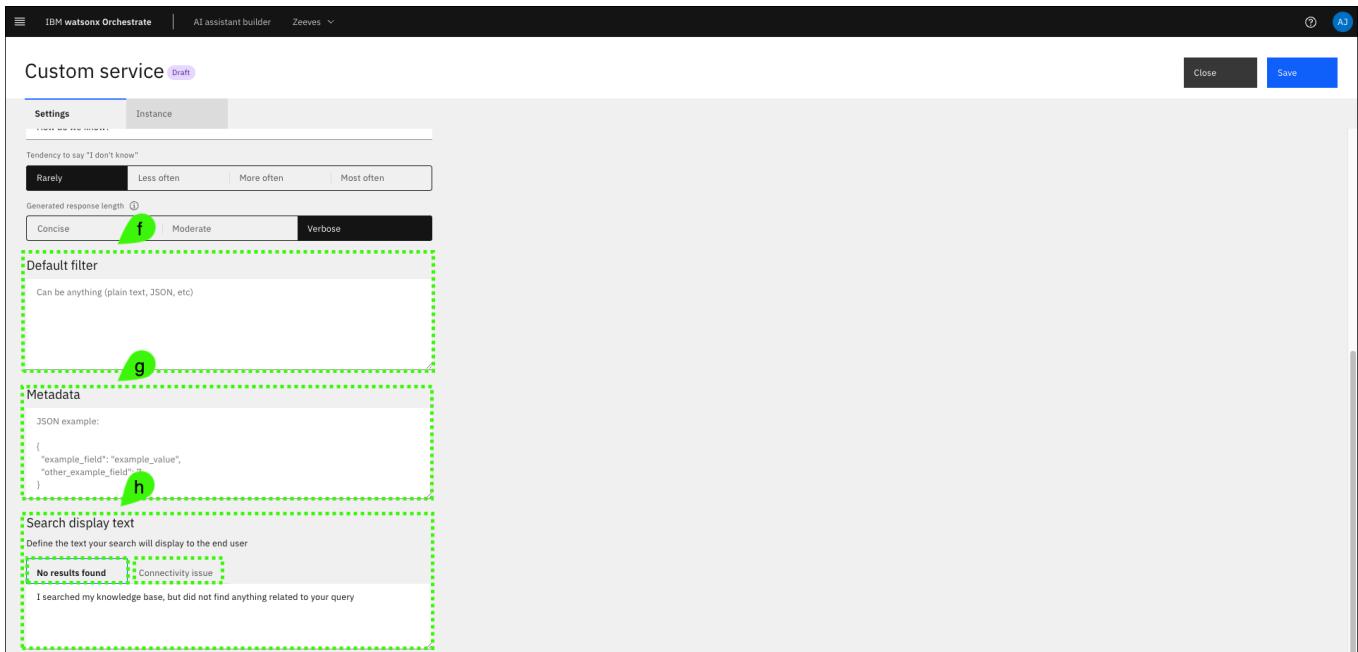
Depending on user input, variations from the selected length may occur.



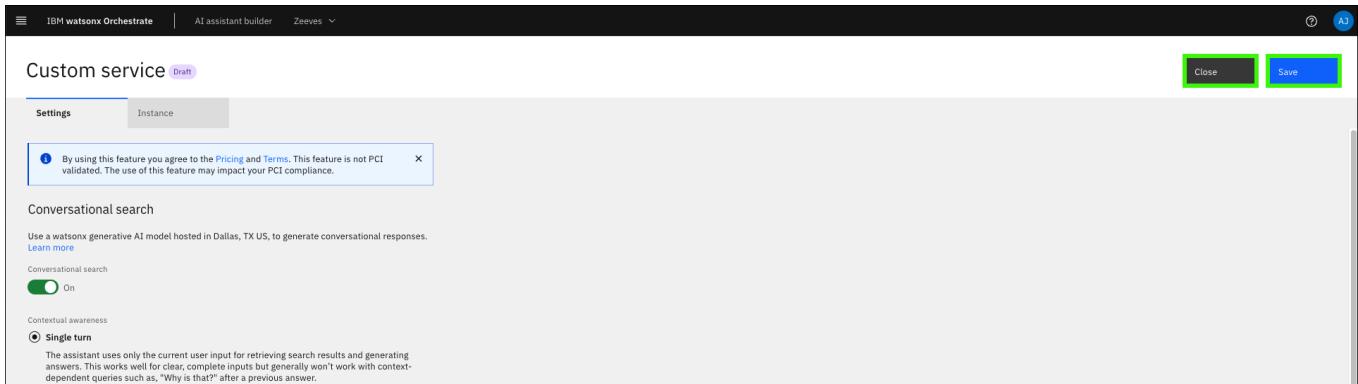
**f. Leave the Default filter field empty.**

**g. The Metadata field provides a way to adjust your assistant's behavior during conversational search for your OpenSearch instance. This option will be 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 there are connectivity issues to the backend search service. You can keep the defaults or customize.**



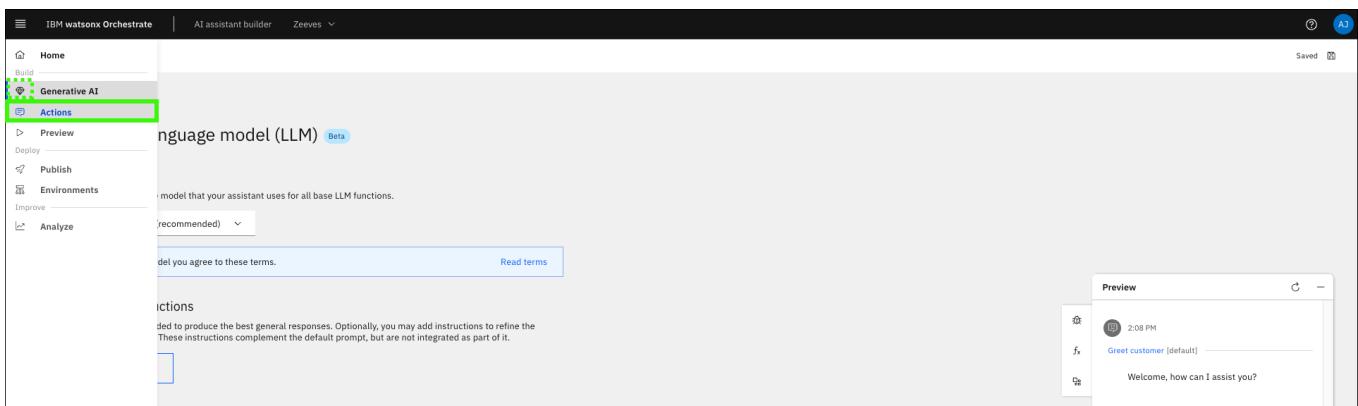
21. Click **Save** and then click **Close**.



## Additional configuration

Once you have saved and closed the **Conversational search** configuration page, there are a few more configurations 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 left navigation and click **Actions**.



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

Create your first action  
With actions, you can help your customers accomplish their goals.  
Create action +

24. Click **no matches**.

Name	Last edited	Examples Count	Status
Greet customer	3 hours ago	0	Green
Trigger word detected	3 hours ago	0	Green
No matches	3 hours ago	0	Green
Fallback	3 hours ago	0	Green

25. Click **Step 1** under **Conversation steps**.

Action starts  
When no action can be matched to the customer's message.  
Use the assistant's default retry message or customize your own.

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**.

a  
Is taken with conditions

b  
Clear conditions?

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

The screenshot shows the AI assistant builder interface. On the left, there's a sidebar with 'Conversation steps' and two items: '1 This step has no content' and '2 This step has no content'. Item 1 has a green dashed border around it. In the main area, under 'Step 1', 'Is taken' is set to 'without conditions'. Under 'Assistant says', there is a rich text editor with placeholder text: 'For example: Please select from the following options:'. Below it, there's a dropdown menu with 'Define customer response'.

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

The screenshot shows the AI assistant builder interface with the 'And then' dropdown menu open. The 'Search for the answer' option is highlighted with a green box. Other options in the menu include 'Continue to next step', 'Re-ask previous step(s)', 'Go to a subaction', 'Use an extension', 'Connect to agent', and 'End the action'.

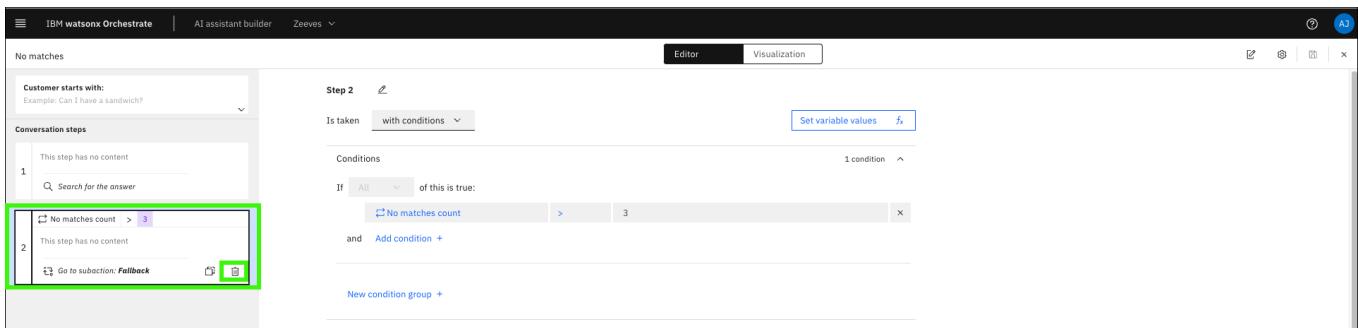
## 29. Click **Edit settings**.

The screenshot shows the AI assistant builder interface with the 'And then' dropdown menu open. The 'Search for the answer' option is highlighted with a green box. In the 'Custom query' field, 'None (Customer's message will be the query)' is selected. In the 'Custom filter' field, 'None (Optional)' is selected. The 'Edit settings' button is highlighted with a green box.

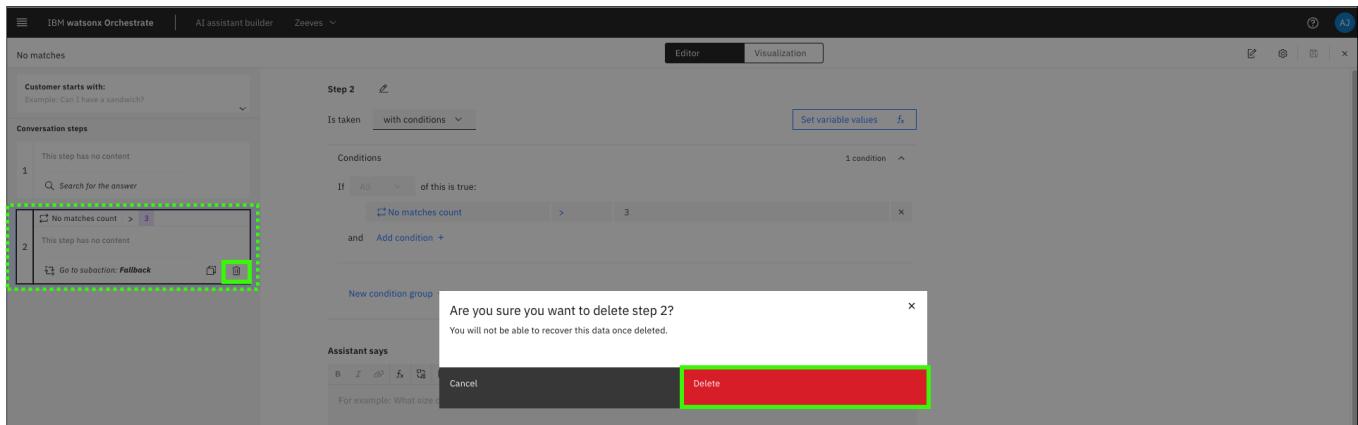
## 30. Select **End the actions after returning results** and then click **Apply**.



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
<b>Fallback</b>	3 hours ago	1	Green

### 35. Delete all of the Conversation steps.

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

### 36. Verify 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
<b>Fallback</b>	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 a user asks 'open an account' and the bot responds with 'Did you mean:' followed by a list of options: '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).

This will 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 is also highlighted with a green box. The option 'More often' is selected.

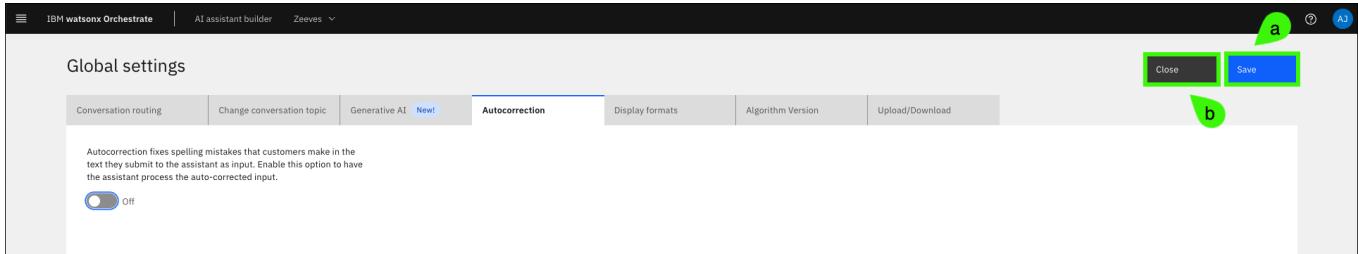
### 40. Click Autocorrection.

The screenshot shows the 'Global settings' page with the 'Autocorrection' tab selected. The 'Autocorrection' section has a note about fixing spelling mistakes. A toggle switch is shown with a green box around it, indicating it is turned off.

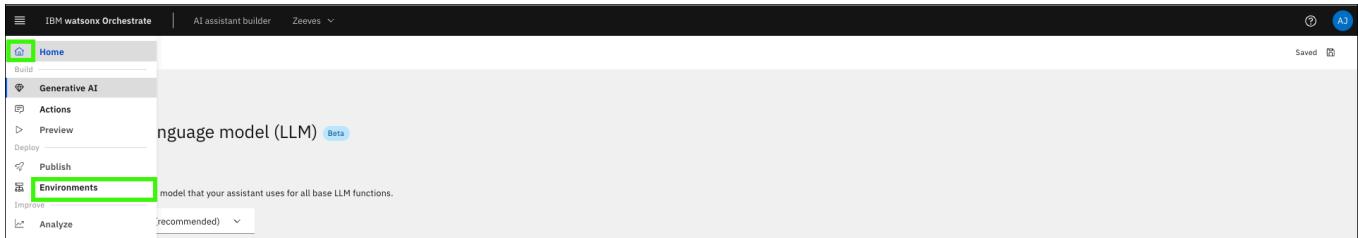
### 41. Click the autocorrection toggle to turn the feature off.

The screenshot shows the 'Global settings' page with the 'Autocorrection' tab selected. The 'Autocorrection' section has a note about fixing spelling mistakes. A toggle switch is shown with a green box around it, indicating it is turned off.

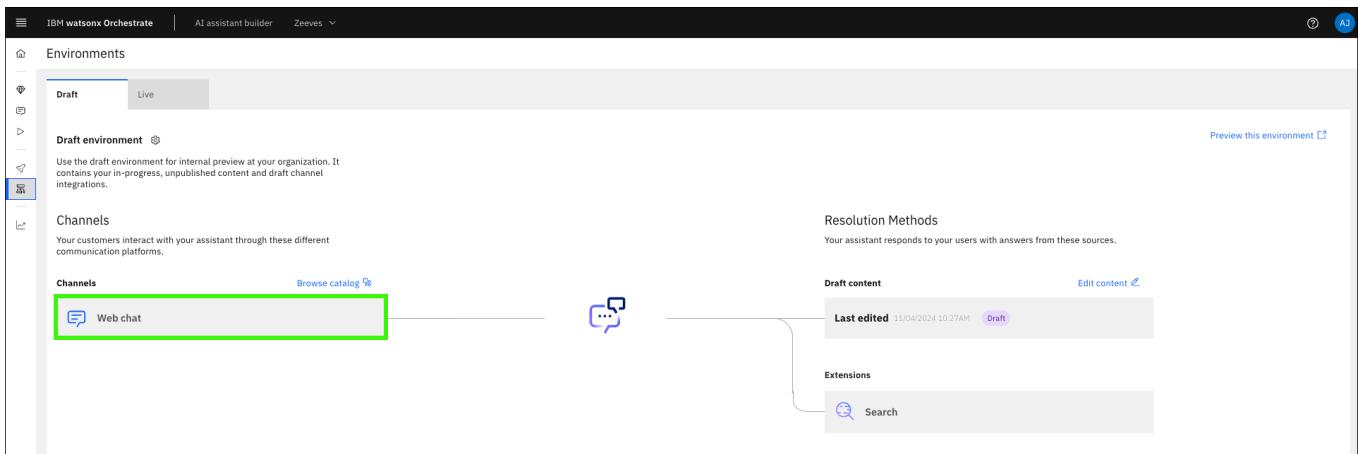
42. Click (a) **Save** and then (b) **Close**.



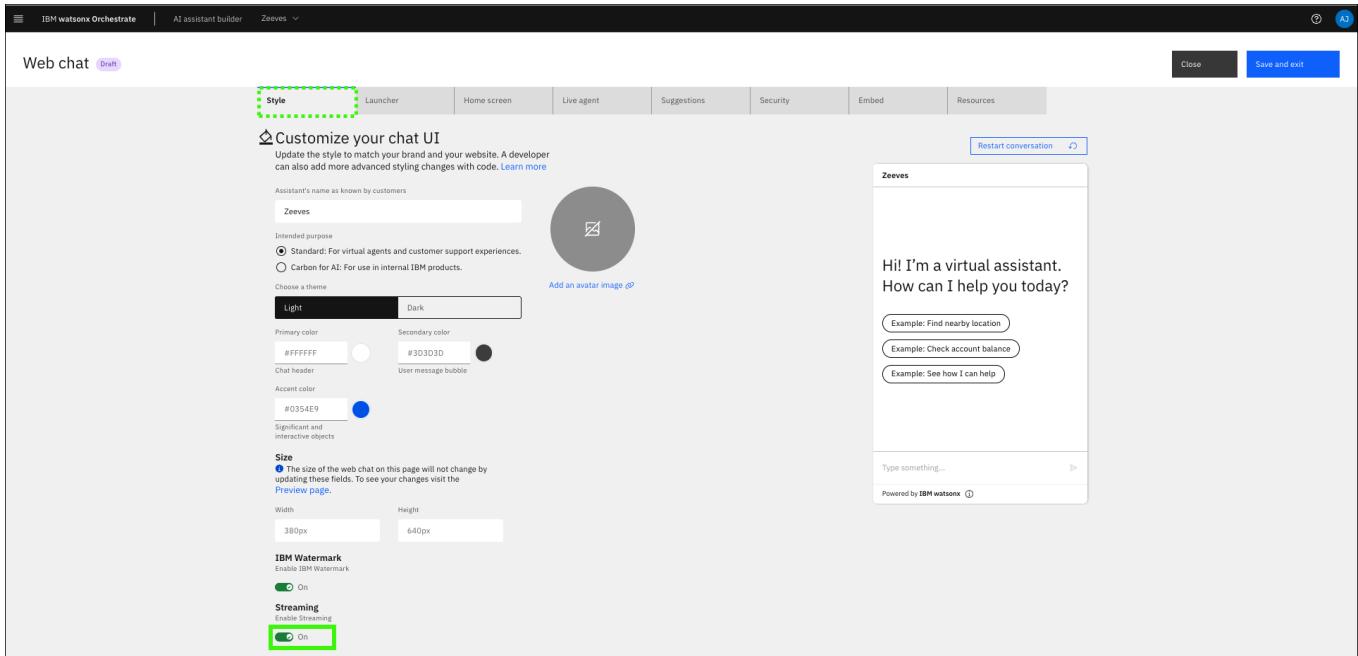
43. Hover over the **home** (🏠) and click **Environments**.



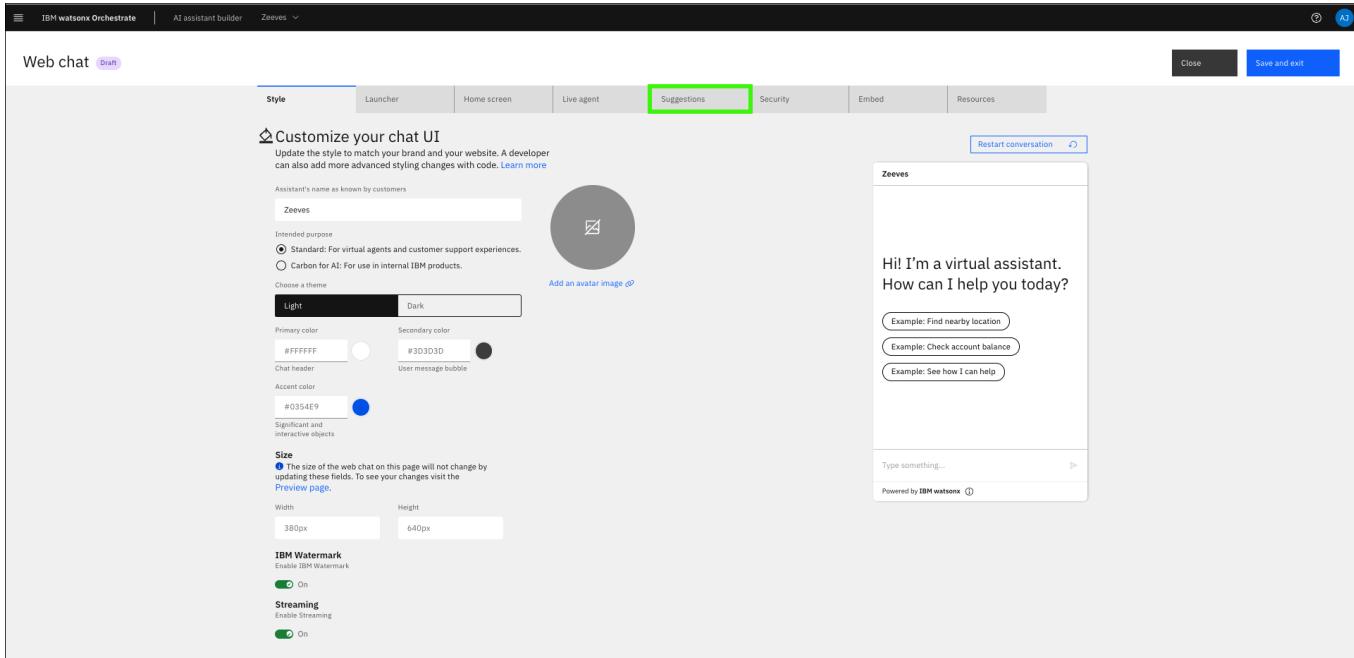
44. Click **Web chat**.



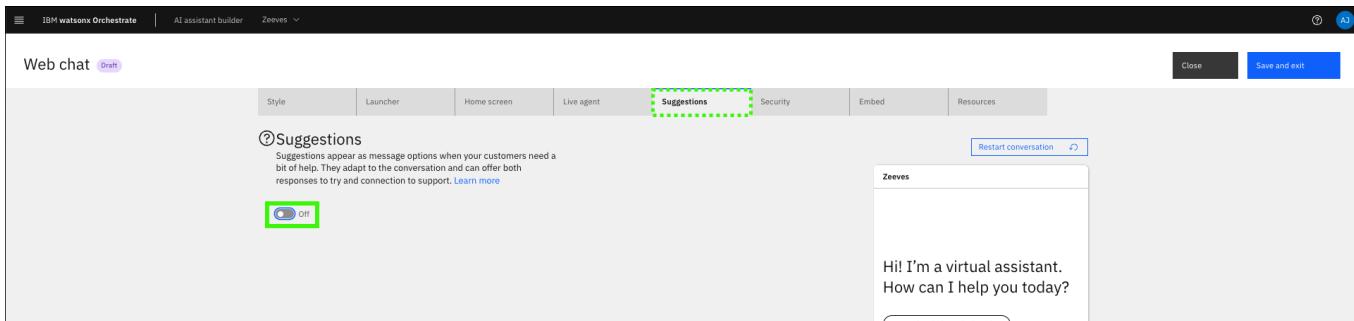
45. On the **Style** tab, click the **Streaming** toggle to enable streaming.



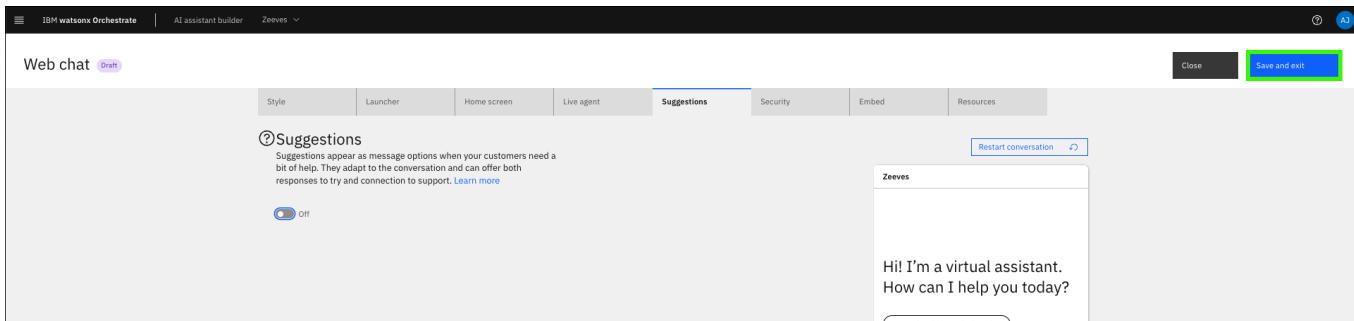
#### 46. Click **Suggestions**.



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



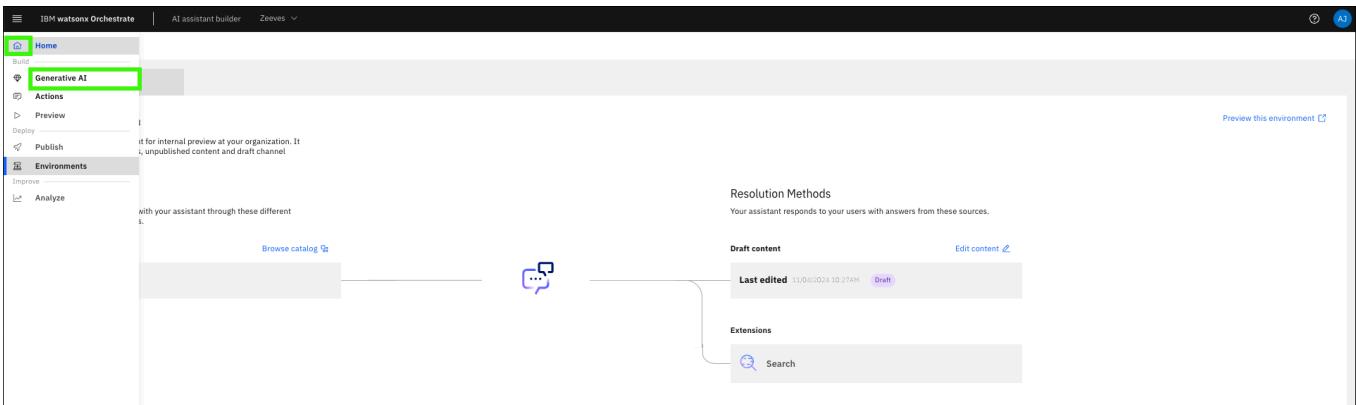
#### 48. Click **Save and exit**.



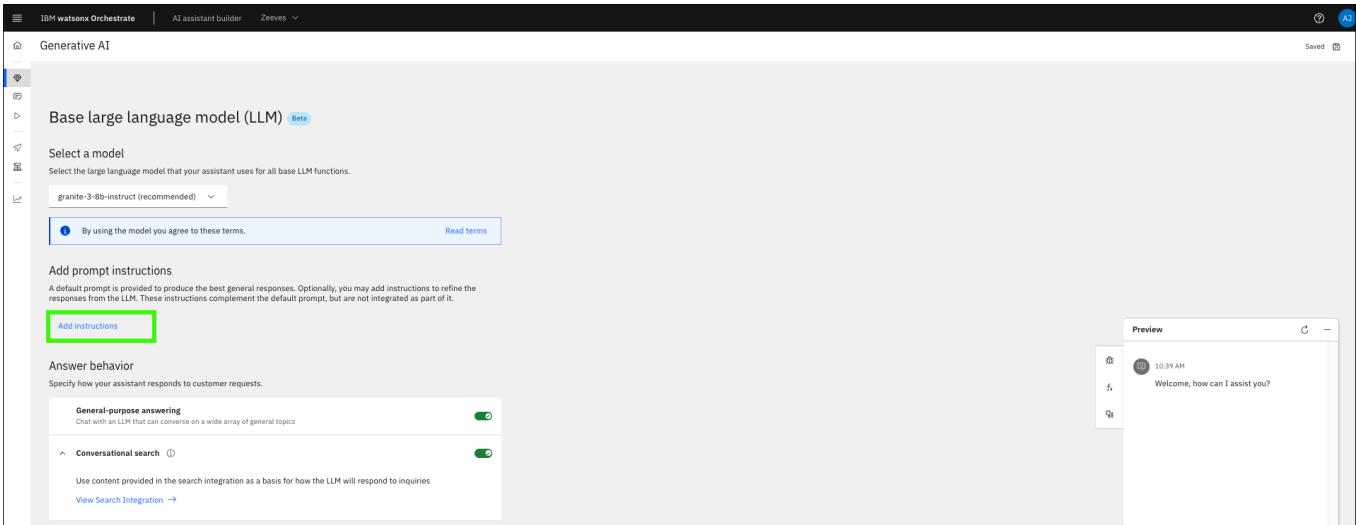
## Configure the base Large Language Model (LLM)

Once the above steps have been completed, there are some final enhancements that can be made to configure how the LLM will respond to your queries. This includes Adding prompt instructions as well as configuring the LLM's answer behavior. These options can be summarized [here](#).

#### 49. Hover over the **home** (🏠) and click **Generative AI**.



## 50. Click Add instructions.



## 51. 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.

Below is an example prompt instruction that has worked well. Feel free to explore using your own 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:** Once the instructions are typed in, they are automatically saved and the LLM is immediately trained on them.

## 52. Toggle General-purpose answering to off.

There is also the ability to configure the answering behavior of your assistant to provide responses that are based on the preloaded content or general content.

At the bottom of the Generative AI page (under Prompt Instructions), you will see the Answer behavior section. Once you've configured Conversational search, you will now see that it is enabled (toggled on) with the search integration you added previously.

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

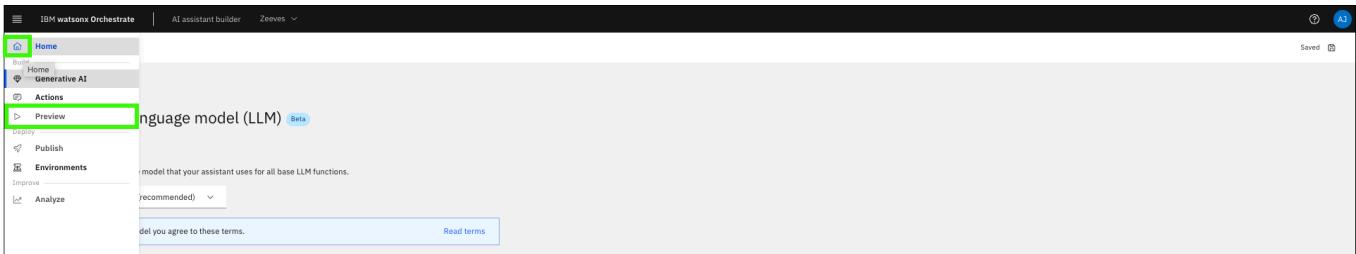
**For purposes of retrieving Z specific answers and responses, it is recommended that you turn off General-purpose answering and only leave Conversational search turned on.**

## Testing conversational search

Now that all of your assistant settings have been made and you have enabled conversational search, you can begin issuing queries to test the responses.

It is important to keep in mind that many of the settings configured above can be iteratively modified based on your assessment of the quality of responses. These can be revisited and changed at any point (for example, adding additional prompt instructions, changing verbosity of the responses, and modifying the indexes used for OpenSearch).

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



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

Here are a few prompts and responses. Note: the responses you receive may vary from those shown below.

**Prompt:**

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

**Example output:**

**Prompt:**

Why is Db2 different than other database systems?

**Example output:**

The screenshot shows the IBM Watsonx Assistant for Z interface. On the left, there's a sidebar with icons for preview, AI assistant builder, and Zeeves. The main area has a title 'Sample website' and a large, mostly blank content area. At the top right, there are buttons for 'Copy link to share', 'Change layout', 'Change background', and 'Customize web chat'. A small A3 icon is also present. On the far right, a chat window titled 'Zeeves' is open. It shows a message from 'You' at 11:03 AM asking 'Why is DB2 different than other database systems?'. Below it, a response from 'Zeeves' at 11:03 AM explains that DB2 is a relational database management system (RDBMS) that runs under the z/OS operating system, which is specific to mainframe computers. Other database systems run on various operating systems like Windows, Linux, or Unix. Additionally, DB2 provides a higher level of data management and security features compared to other database systems, as it manages data stored in databases and controls access to the data.

### Prompt:

What happens during an IPL on IBM Z?

### Example output:

The screenshot shows the same interface as the previous one. The main area is a 'Sample website' placeholder. The chat window on the right shows a message from 'You' at 11:05 AM asking 'What happens during an IPL on IBM Z?'. Below it, a response from 'Zeeves' at 11:05 AM explains that during an IPL (Initial Program Load) on IBM Z, the system reloads the operating system into memory and begins executing it. This process can be initiated from various sources, such as the Support Element (SE), Hardware Management Console (HMC), an Open Systems Port, or a console terminal. The IPL can be used to load a new operating system, recover from a system failure, or perform maintenance tasks. The type of IPL determines whether the Primary Link Pack Area (PLPA) is reloaded and whether VTO data set pages are preserved.

You should now have a working assistant using IBM Watsonx Assistant for Z. Feel free to explore with different prompt instructions and settings. If you encounter any issues, the Troubleshooting section that follows may help resolve them.

Continue to the [Creating a standalone 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 you may encounter. If the provided resolutions do not work, contact support using the methods mentioned in the [Support](#) section of this guide.

 **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 standalone OpenSearch instance for document ingestion

This section will cover how to enable a client to ingest their own documentation into the Retrieval Augmented Generation (RAG) used by IBM watsonx Assistant for Z. This requires deploying a dedicated [OpenSearch](#) instance.

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 already 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, known as **oc**, must be installed on your local machine. If you already have the **oc** utility installed, you can proceed to [Log into the SNO cluster](#).

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

[ITZ My reservations](#)

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

Environment Tile	Status	Category	Start Date	End Date	Extend Limit	Action
Single Node OpenShift (VMware on IBM Cloud)	Ready	Education	Oct 31, 2024 7:02 AM	Nov 6, 2024 6:49 AM	0	<a href="#">Open this environment</a>
watsonx Assistant for Z Pilot - AAP & z/OS	Ready	Education	Oct 30, 2024 8:12 AM	Nov 5, 2024 7:53 AM	0	<a href="#">Open this environment</a>
watsonx Assistant for Z Pilot - watsonx O...	Ready	Education	Oct 30, 2024 8:10 AM	Nov 5, 2024 7:12 AM	0	<a href="#">Open this environment</a>

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 [Download](#)

#### 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]

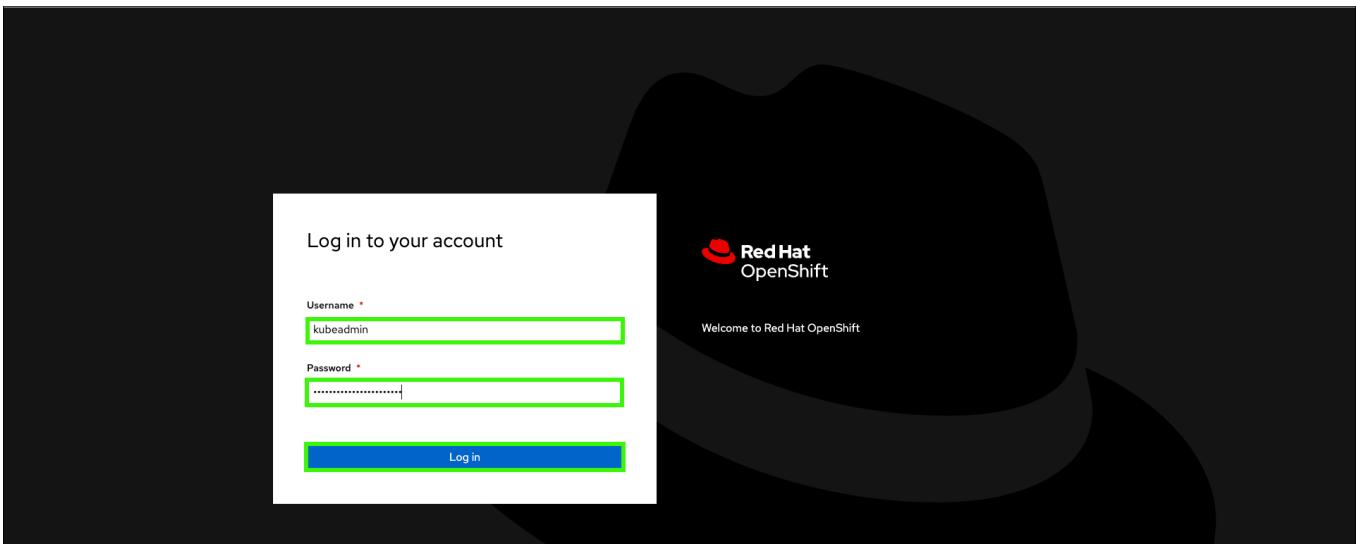
OCP Console  
<https://console.openshift-console.apps.672371d38376796fb96a6c4d.ocp.techzone.ibm.com>

OCP Version  
4.14

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

Download kubeconfig [Download](#)

#### 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 above link will automatically download either a **zip** or **tar** file specific to your operating system. You should unzip or un-tar the file and place the **oc** binary in a directory that is in your default PATH. Since this will vary by operating system and individual settings, details for doing this are not provided.

Once properly installed, verify the installation by running the **oc** command in a terminal on your local machine.

```
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

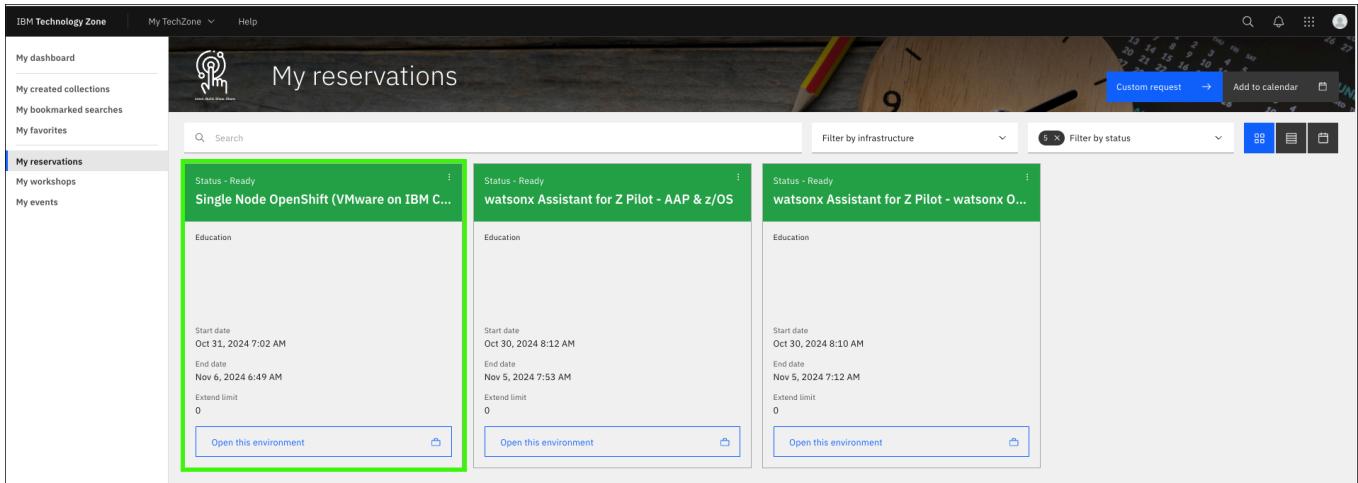
Before ingesting documents, a few setup steps must be taken.

### Login to the OpenShift cluster from your local machine

Note: if you just installed the **oc** utility, you should be able to skip the next 5 steps. 1. Click the link below to open a browser window to your ITZ reservations.

```
<a href="https://techzone.ibm.com/my/reservations" target="_blank">ITZ My reservations</a>
```

#### 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](#)

#### 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]

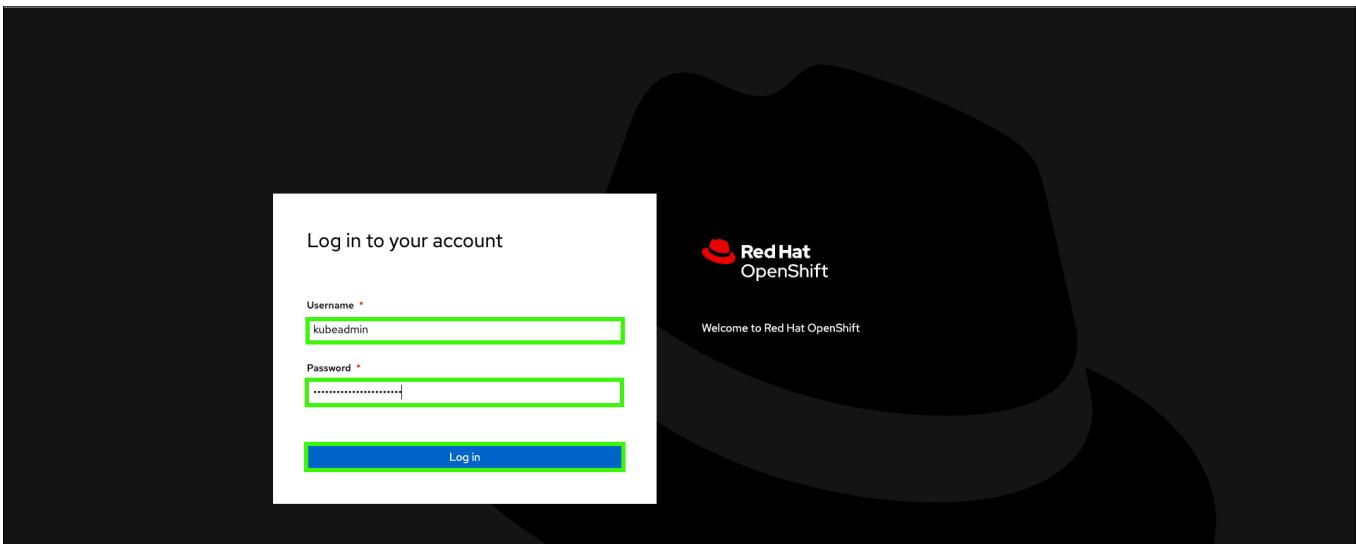
OCP Console  
<https://console.openshift-console.apps.672371d38376796fb96a6c4d.ocp.techzone.ibm.com>

OCP Version  
4.14

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

[Download kubeconfig](#)

#### 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 should be able to 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 machine.

10. Paste the login command line in the terminal window 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 may vary by your local machines 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 OpenShift

13. Create a file 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 OpenShift cluster.

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

The above command should return a message stating the **ibm-cert-manager-catalog** was created.

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

The screenshot shows the 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 user is logged in as 'kube:admin'.

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

The screenshot shows the OperatorHub page in the OpenShift web console. The 'Project: All Projects' dropdown is highlighted with a green box. To its right, a toggle switch labeled 'Show default projects' is also highlighted with a green box. The main content area lists several projects under 'All Projects'. The top navigation bar shows the user is logged in as '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

openshift-kube-controller-manager-operator  
openshift-kube-scheduler  
openshift-kube-scheduler-operator  
openshift-kube-storage-version-migrator  
openshift-kube-storage-version-migrator-operator  
openshift-machine-api  
openshift-machine-config-operator  
**openshift-marketplace**  
openshift-monitoring  
openshift-multus  
openshift-network-diagnostics  
openshift-network-node-identity  
openshift-network-operator  
openshift-node  
openshift-nutanix-infra  
openshift-oauth-apiserver

Community      Marketplace

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

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

[DEPRECATED] Hazelcast Platform Operator  
[DEPRECATED] Use the certified Hazelcast Operator instead

ABot  
Abot Operator-v3.0.0  
provided by Rehaca Technologies

Accuknox Operator  
provided by Accuknox Inc.

Red Hat  
Advanced Cluster Management for Kubernetes

624 items

## 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      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.

1 items

## 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      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 install process will take 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 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.

Once you locate your existing key or create 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 a maximum of 5 keys. A specific key is shown with its issue date as 'October 17, 2022'. To the right of the key details are 'Copy' and 'Delete' buttons, with 'Copy' highlighted by a green box.

### 24. Using the command below 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 terminal window shows the following commands being run:

```
andrewjones@Andrews-MBP watsonxAssistant % oc create namespace wxa4z-byos  
namespace/wxa4z-byos created  
andrewjones@Andrews-MBP watsonxAssistant % export IBM_CS_ENT_KEY=  
andrewjones@Andrews-MBP watsonxAssistant % oc -n wxa4z-byos create secret docker-registry icr-pull-secret --docker-server=cp.icr.io --  
secret/icr-pull-secret created  
andrewjones@Andrews-MBP watsonxAssistant %
```

The 'secret/icr-pull-secret created' 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 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 OpenShift web console, click **OperatorHub** and select the **wxa4z-byos** project.

The screenshot shows the Red Hat OpenShift web console interface. The left sidebar has a navigation menu with items like Home, Operators, Workloads, Networking, Storage, Builds, Observe, Compute, User Management, and Administration. The 'Operators' item is currently selected and highlighted with a green border. The main content area displays a list of operators under the 'operatorhub' project. A search bar at the top of the list has 'Project: openshift-marketplace' selected. The search results show several operators, including 'ibm-wxa4z-operator-catalog' which is highlighted with a green border. Other operators listed include 'openshift-node', 'openshift-nutanix-infra', 'openshift-oauth-apiserver', 'openshift-openstack-infra', 'openshift-operator-lifecycle-manager', 'openshift-operators', 'openshift-ovirt-infra', 'openshift-ovn-kubernetes', 'openshift-route-controller-manager', 'openshift-service-ca', 'openshift-service-ca-operator', 'openshift-user-workload-monitoring', 'openshift-vsphere-infra', 'techzone-infra', and 'wxa4z-byos'. The 'wxa4z-byos' operator is also highlighted with a green border. The bottom of the list shows a total of 624 items.

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

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.

**OperatorHub > Operator Installation**

### 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 install process will take 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.

**OperatorHub**

**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 OpenShift web console, click **Workloads** and select the **Pods**.

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

**IBM watsonx Assistant for Z**  
ibm-wxa4z-operator/v2.0.1 provided by IBM

### 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-7c7898d7d4-87htm	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=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9eyJpc3MiOiJJQk0gTWFya2V0cGxhY2UiCJpYXQiOjE2NjYwMTk1ODAsImp0aSI6ijNkOWUyMzZjTAzM0gMzVhNTJhYTkzMWNmOTcyMDR1In0.4M3XRD4XzkHMS0kFNJ4uKVcWZ6SnEA0Z03eL_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.operators.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 just 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 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 will begin and could 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 all pods are either have the status of **Running** or **Completed**.

Name	Status	Ready	Restarts	Owner	Memory	CPU	Created
da80909aab1563ff0c15b7c6d	Completed	0/1	0	da80909aab1563ff0c15b7c6da81 belefabaa605065bb6c6b490e87 f067f45	-	-	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-719b8d6c58-9bzth	Running	1/1	0	wxa4z-client-ingestion-719b8d6c58	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

Once finished deploying and all the pods have a status of "Ready" or "Running", you must retrieve your BYOS endpoint URL.

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

Name	Status	Ready	Restarts	Owner	Memory	CPU	Created
da80909aab1563ff0c15b7c6d	Completed	0/1	0	da80909aab1563ff0c15b7c6da81 belefabaa605065bb6c6b490e87 f067f45	-	-	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-719b8d6c58-9bzth	Running	1/1	0	wxa4z-client-ingestion-719b8d6c58	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 previous step.

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

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

Note, the above URL will not work for you. You must use the value of your specific OpenSearch instance 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 do this is illustrated in the animated gif below. You can revisit this step in the prior section [here](#).

The screenshot shows the 'Routes' section of the Red Hat OpenShift interface. The left sidebar shows 'Operators' and 'Workloads' sections. Under 'Networking', 'Routes' is selected. The main area displays a table of routes with columns for Name, Status, Location, and Service. Two routes are listed: 'wxa4z-client-ingestion' (Status: Accepted, Location: https://wxa4z-client-ingestion-wxa4z-byos.apps.67b79320c7a7b728e523b4.ocp.techzone.ibm.com) and 'wxa4z-opensearch-wrapper' (Status: Accepted, Location: https://wxa4z-opensearch-wrapper-wxa4z-byos.apps.67b79320c7a7b728e523b4.ocp.techzone.ibm.com). The URL for 'wxa4z-client-ingestion' is highlighted with a yellow box and a cursor icon.

## Install the `zassist` command

## Troubleshooting

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

### ✖ 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 located [here](#).

## Install the zassist utility

The **zassist** utility is an executable program that automates the ingestion of client documentation into the RAG for Watson 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 below to download and install **zassist**.

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

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

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

2. Unzip the **zassist.zip** file.

3. Locate the appropriate executable for your local machine'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
<b>zassist</b>	<b>Sep 5, 2024 at 11:15 AM</b>	--	<b>Folder</b>
linux	Sep 18, 2024 at 8:15 AM	--	Folder
mac	Sep 18, 2024 at 8:15 AM	--	Folder
<b>zassist</b>	<b>Sep 18, 2024 at 8:17 AM</b>	<b>10.2 MB</b>	<b>Unix Executable File</b>
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** executable to a directory in your command prompt or terminal window's PATH, or copy it to a directory and add that directory PATH to your PATH environment variable.

Additional information for performing the above tasks can be found [here](#).

## 5. Verify 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 executable ready, 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, rather the video below illustrates all of the steps to ingest a single document. This video has no audio.

## 404 - Not found

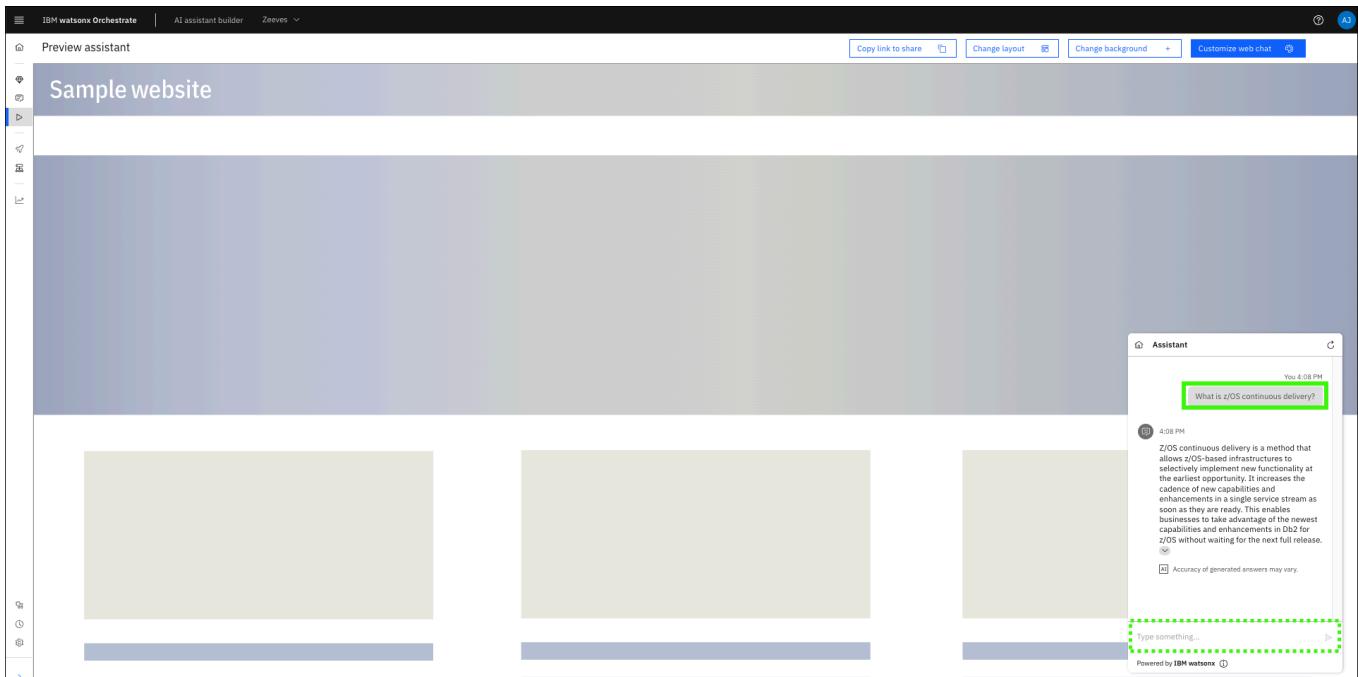
The document ingested in the video is a compressed PDF version of the [IBM z/OS Continuous Delivery Red Piece](#). You can download a copy of this document [here](#).

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

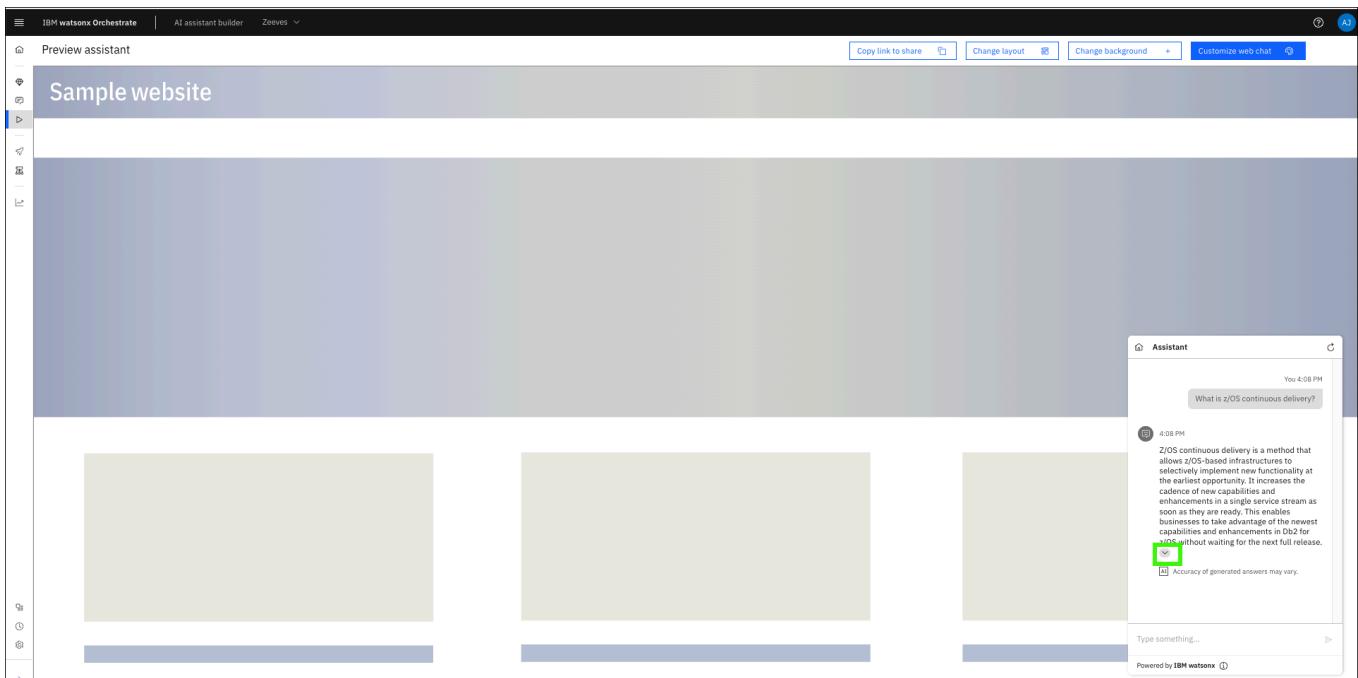
Use the watsonx Orchestrate AI assistant builder to verify your document ingestion.

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.

The screenshot shows the IBM Watsonx AI assistant builder interface. On the left, there's a sidebar with icons for file operations like preview, edit, and share. The main area is titled "Sample website" and displays a blurred preview of a web page. To the right, there's a "Assistant" panel with a summary of a document. The summary includes a title, a detailed description of the document's content, and a "View source" button. At the bottom of the assistant panel, there's a note about accuracy and a "Powered by IBM watsonx" footer.

#### 10. Click the ingested document reference.

This screenshot is similar to the previous one, but the "View source" button in the AI-generated document summary is highlighted with a green box. This indicates that the user has interacted with the document reference, fulfilling the task requirement.

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

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

```

```

## Adjusting your assistant's search behavior  
Do you recall the **Metadata** field when you were configuring your assistant?

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 ingested for conversational search, you can set the metadata field for your assistant to use those documents in it's content-grounded search. If you leave the metadata field empty, then it defaults to settings found to perform well experimentally. This replaces having to paste a complicated search string. By default (without any string in the Metadata field), it will search all the default IBM provided documentation and all ingested customer documentation using the following:

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

Replacing the wildcard string with an explicit list of indices allows personalization. This is where you could 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 which the OpenSearch instance has IBM documentation for. You can find those indices and products [here](#).

You have the ability to input a subset of indices into the "Metadata" field in cases where you only want your assistant gathering 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 could 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 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 assigned to “ibm\_indices” and “customer\_docs” is the weight assigned to “customer\_indices”.

Once you’ve configured all the above settings for Conversational Search on that page, you can click “Save” in the top-right of the page.

For more information on customizing the metadata field for conversational search, please 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_indices":"*_ibm_docs_slate",  
"customer_indices":"customer_*"  
}
```

Now redo steps 6 through 8 above. Notice the ingested Red Piece document is now the first sited reference!

The screenshot shows the IBM Watsonx Assistant interface. On the left, there's a preview of a "Sample website" with three large, blurred rectangular sections. On the right, the "Assistant" panel is open, displaying a message from "Zeeves" at 4:30 PM. The message discusses "z/OS continuous delivery" and includes a link to "redp5340-compressed.pdf". The assistant panel also has a "View source" button and navigation controls (1 of 3, < >).

## Adding skills to the assistant

# Getting started with skills and actions

Watsonx Assistant for Z provides the ability to import skills to automate a range of IBM Z related tasks through assistant interactions. Each skill is a pre-defined automation that performs a different task to accomplish some 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 that allows you to build new skills from Ansible Automation platform or z/OS Management Facility (z/OSMF). This extension, Z Skills Accelerator, allows you to connect to Ansible and z/OS application programming interfaces (APIs) and enables you to import automation as Ansible Playbooks, JCL, or REXX as skills. Learn more importing and building skills [here](#).

## Environments

### watsonx Orchestrate

The extension for Watsonx Assistant for Z that allows for importing skills from Ansible Automation platform (AAP) or z/OSMF is called the 'z/OS Skills Accelerator' which is already configured in your Watsonx Orchestrate IBM Technology Zone (ITZ) environment. You will be using this component when it comes to importing new skills.

### Ansible Automation Platform (AAP) and Wazi as a Service (aaS)

To import skills for automations, you will be using Ansible Automation Platform (AAP) and 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 you reserved earlier. This environment will enable the ability to manage and automate z/OS tasks and subsystems with a variety of pre-loaded ansible playbooks ready for immediate use, as well as z/OS installed with all prerequisites needed (no additional setup required).

The playbooks provided cover a variety of simple 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 pre-loaded in the AAP instance. These are mostly tasks from the 'IBM z/OS core collection'. Leveraging this environment will accelerate the ability to showcase the value of Watsonx Assistant for Z when it comes to automation, 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 will be imported into Watsonx Orchestrate as skills and connected to your assistant.

For more information on the AAP & Wazi z/OS environments, refer to this [document](#).

The playbook templates pre-loaded into AAP cover various use cases which you will be able to explore, including:

- z/OS Certificate Management (create, delete, list and renew certificates)
- dataset management (create, delete, fetch datasets)

- Submit JCL
- Execute Operator commands
- Execute TSO commands
- And more

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

# Explore Ansible Automation Platform (AAP)

Once you have provisioned the AAP & Wazi z/OS environments in ITZ and the reservation is in the **Ready** state, follow these steps to explore AAP.

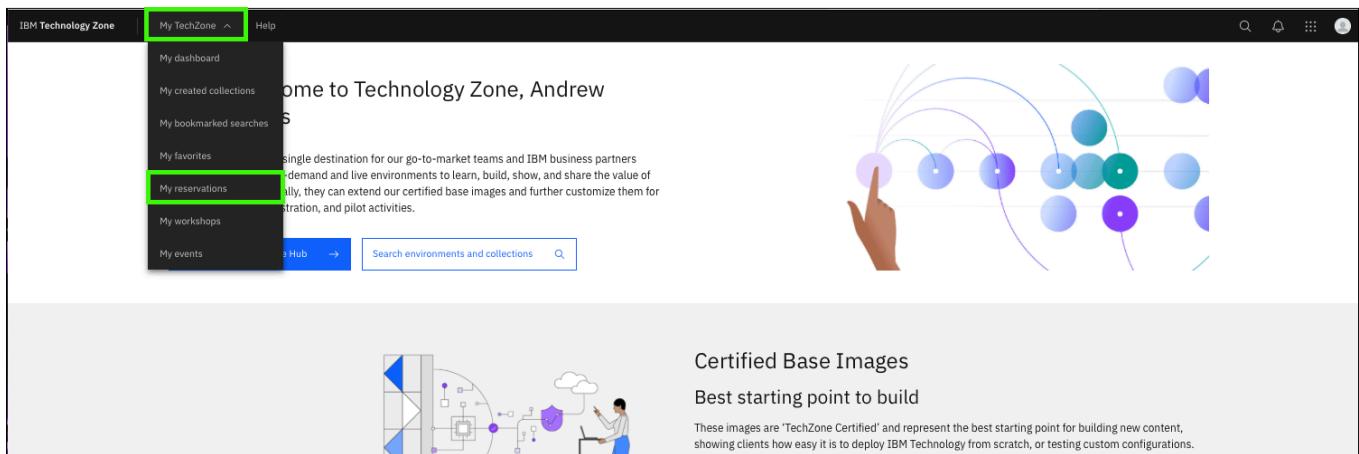
Access the AAP and Wazi as a Service environment

## Be sure to record the information as instructed

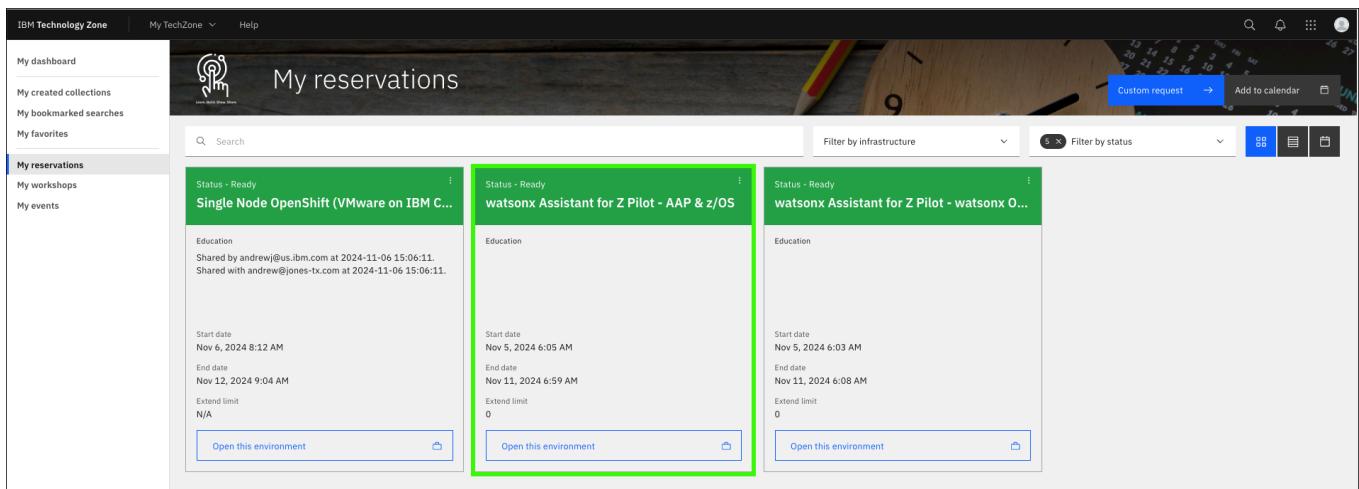
Several of the steps below instruct you to record values from your ITZ reservation. Be sure to do this as they will not only be used in this section, but also in later sections of the lab guide.

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

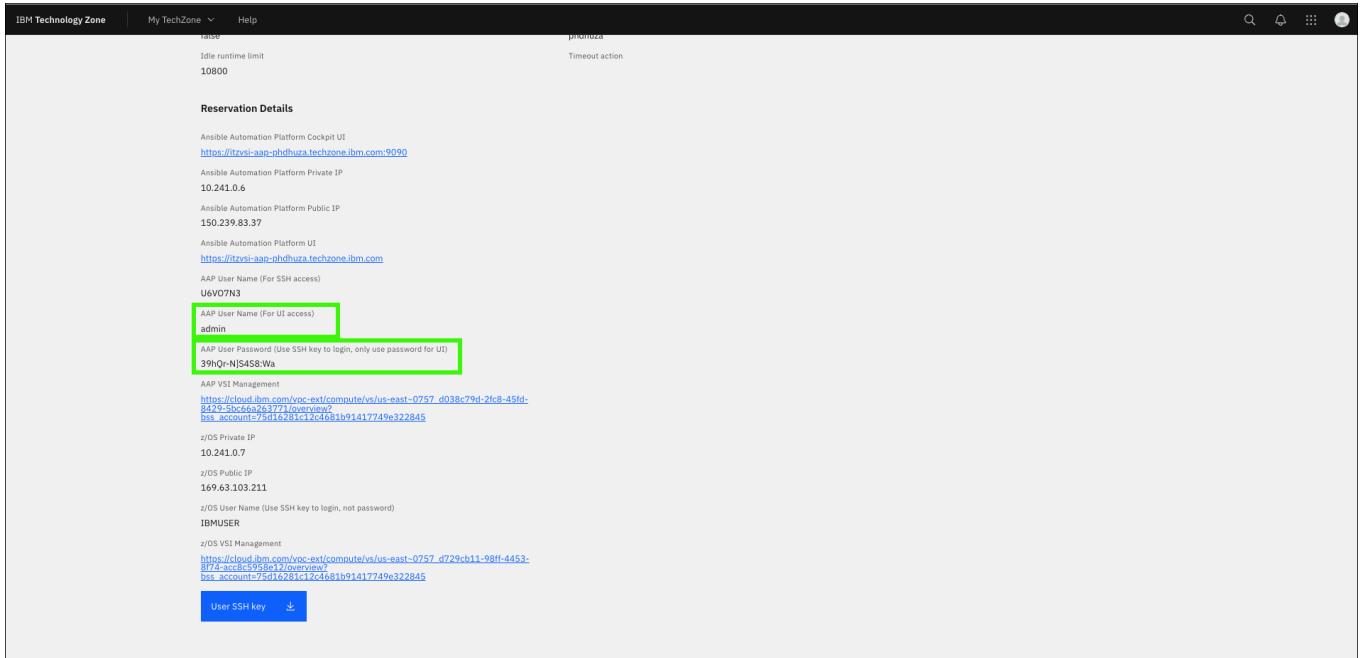
### ITZ My reservations



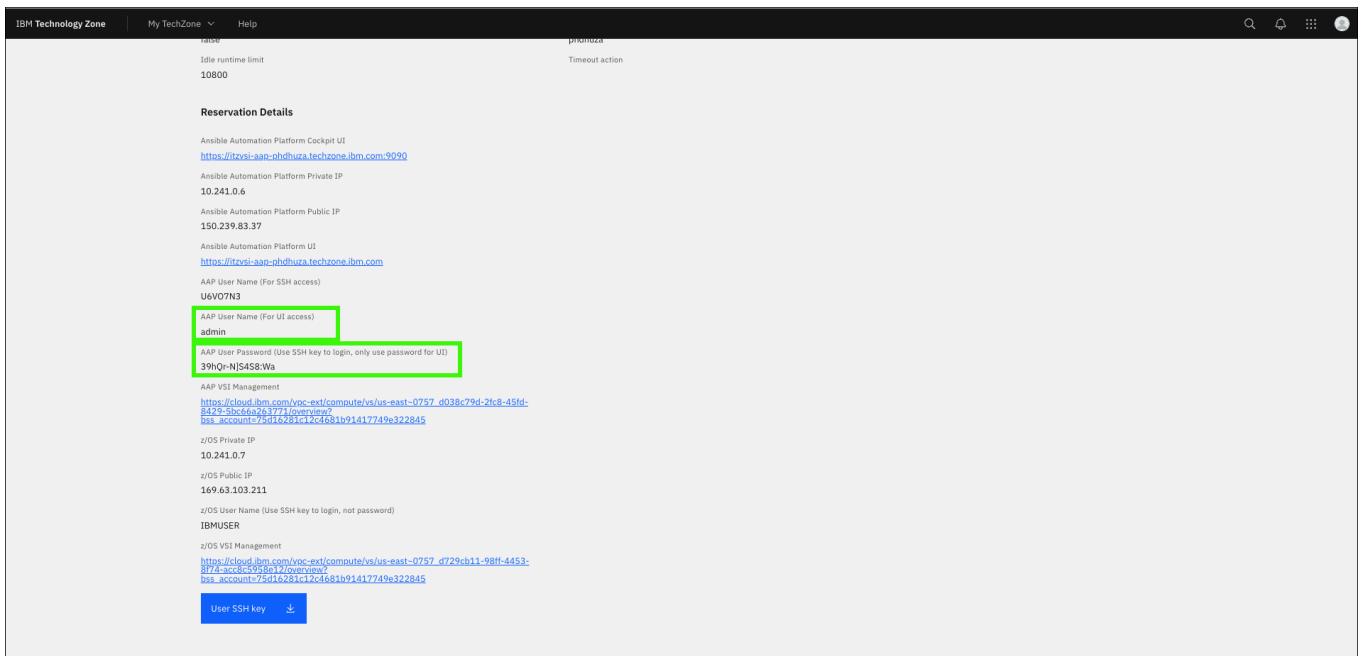
2. Click the **watsonx Assistant for Z Pilot - AAP & z/OS** tile.



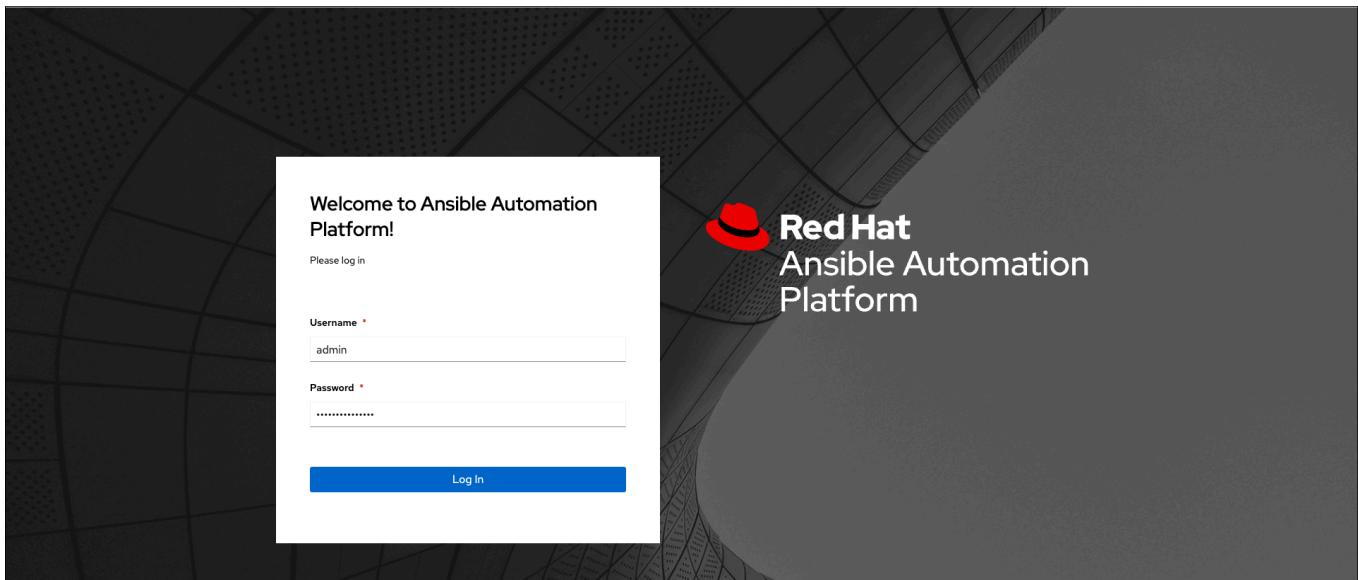
3. Locate and record the **AAP User Name (For UI access)** and **AAP User Password** fields.



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



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

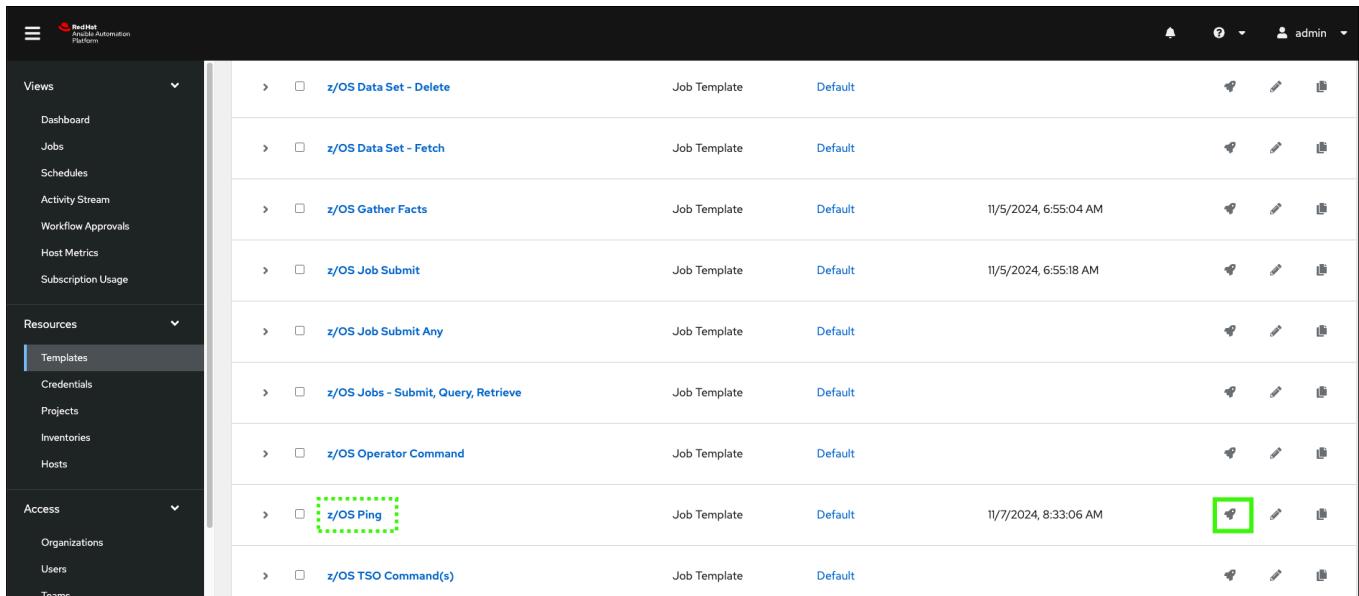


6. Click **Templates** under the **Resources** section.

**i** The AAP instances is preconfigured to the Wazi aaaS instance

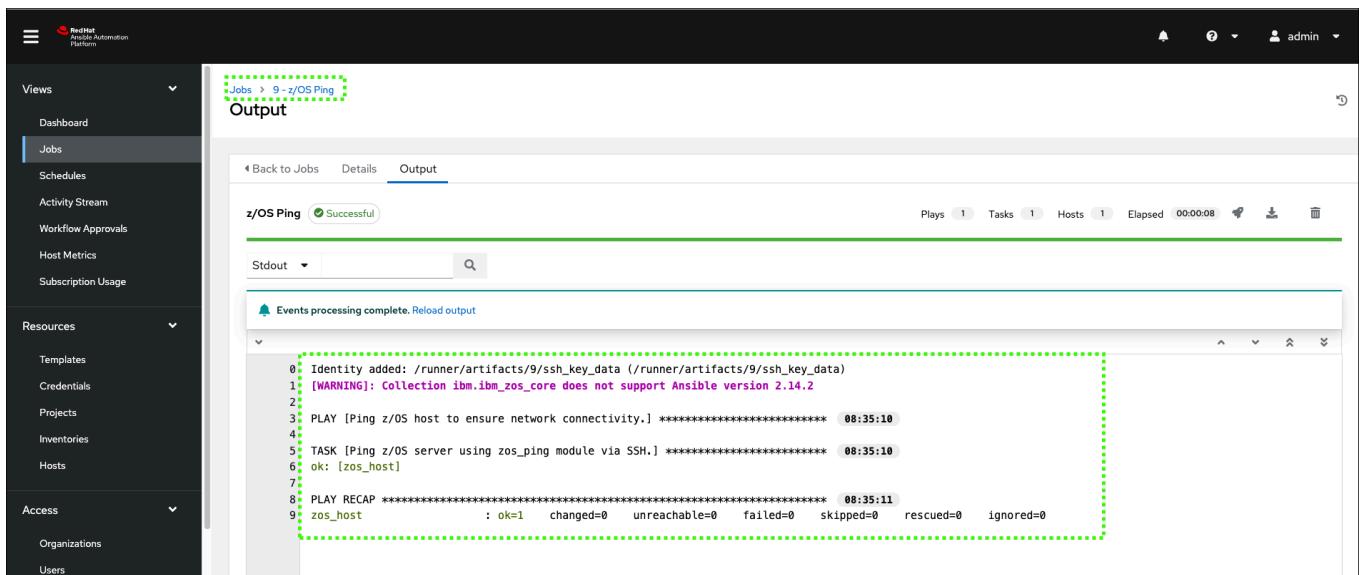
Note that because the AAP instance and the back-end z/OS system are preconfigured, no changes are needed to execute the templates and they will target your provisioned z/OS system by default.

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



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

## 8. Observe the z/OS Ping job run.



The screenshot shows the 'Output' tab for the 'z/OS Ping' job. The terminal window displays the following log 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
-----
```

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

# Importing skills from Ansible Automation Platform

Once you have a better understanding of AAP and the pre-loaded automations available, you can begin to import them as skills into your watsonx Orchestrate instance which will ultimately be used for assistant guided actions.

## Import skills into your assistant

The next steps assume 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 (☰).

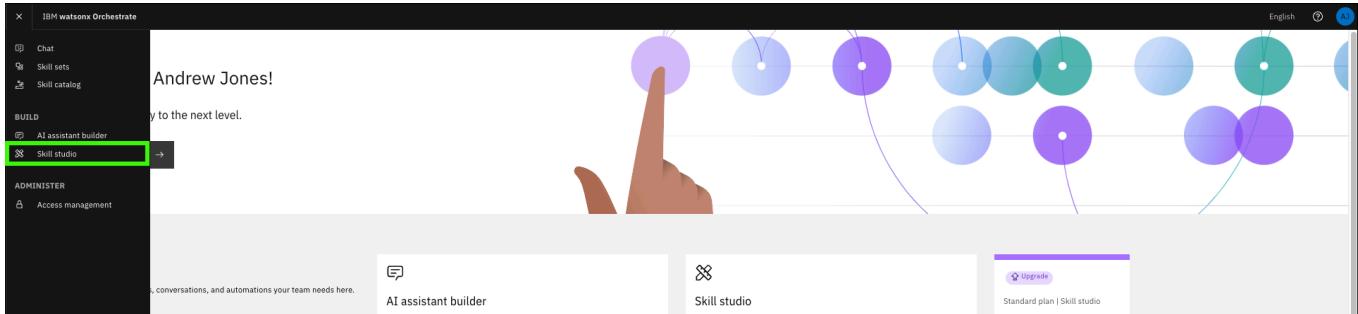
The screenshot shows the IBM Cloud dashboard with a sidebar on the left containing icons for various services. The main area displays a 'For you' section with several cards:

- Build**: Explore IBM Cloud with this selection of easy starter tutorials and services. (Popular)
- Use Watson Assistant**: Watson Assistant lets you build conversational interfaces into any application, device, or channel. (Popular)
- IBM Cloud Essential Security and Observability Services**: Deploy core security and other supporting services to get set up to manage the security compliance of the resources in your account. (Recommended)
- Use Cloud Object Storage**: Store limitless unstructured data with Cloud Object Storage and access it via self serve portal or S3 API. (Recommended)
- Maximo Application Suite**: Deploy an IBM Maximo Application Suite instance in a Red Hat OpenShift cluster. (Recommended)
- Build with Watson**: Chatbots, insights, recognizers, and more. Explore the AI platform for business. (Popular)
- Use Watson Studio**: Watson Studio provides a suite of tools and a collaborative environment for data scientists, developers and domain experts. (Popular)
- Get Started with Watson Studio**: Get started with u and Cloud Object 15 minutes. (Popular)

2. Click **Launch watsonx Orchestrate**.

The screenshot shows the watsonx Orchestrate Resource list page. On the left, there's a sidebar with 'Manage' selected. The main area has sections for 'Start by launching the tool' (with 'Launch watsonx Orchestrate' highlighted) and 'Plan' (showing 'Essentials Plan'). Below these are sections for 'Credentials' (with an API key and URL) and 'Getting started tutorial'.

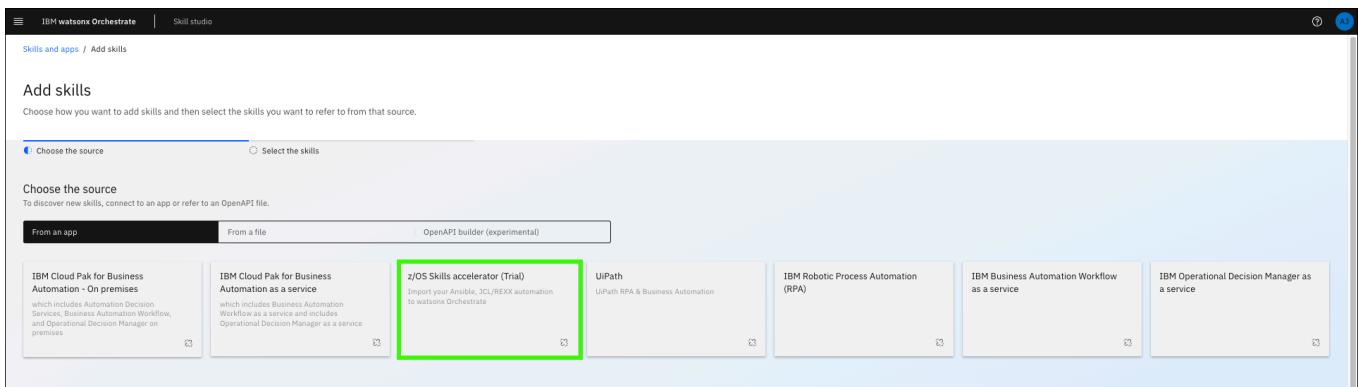
3. Click the main menu (☰) and then click **Skill studio**.



#### 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**

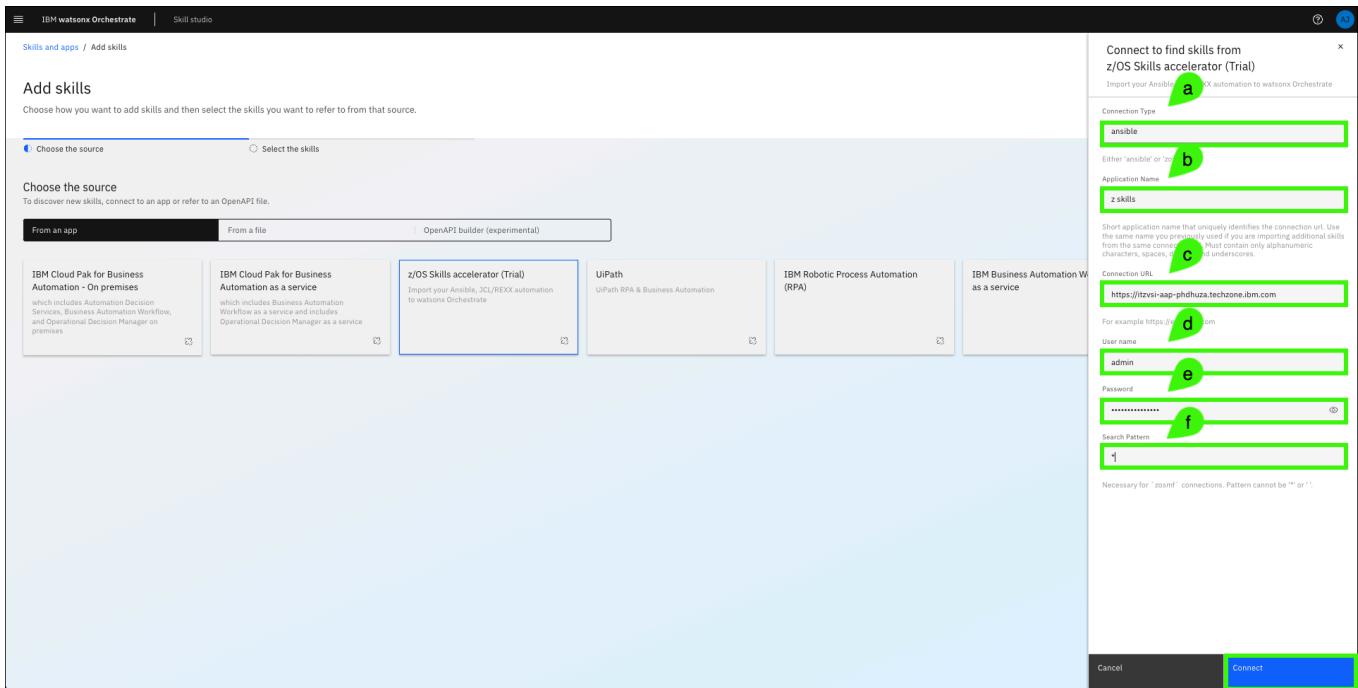
**b:** Application Name: <enter a meaningful name for the skills you will import>

**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: \*



TBD

TBD

TBD

TBD

TBD

TBD

TBD

# 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.