

SRE Agent – Autonomous Incident Remediation

Learning Objective

You will gain hands-on knowledge and practical skills to design, build, and deploy intelligent AI agents tailored for Site Reliability Engineering (SRE) workflows in the insurance domain using Watsonx Orchestrate ADK and Watsonx Orchestrate. Specifically, you will be able to:

- Develop custom AI agents using Watsonx Orchestrate ADK that can autonomously monitor system health, analyze logs and metrics, and trigger predefined remediation playbooks for core insurance services such as policy management, claims processing, and customer portals.
- Orchestrate end-to-end incident management workflows using Watsonx Orchestrate, integrating multiple tools, APIs, and observability platforms to streamline root cause analysis and reduce resolution time.
- Apply these skills to real-world SRE challenges, improving system reliability, minimizing downtime, and enhancing customer satisfaction through proactive and intelligent incident response in the insurance domain.

The Problem

MetLife Insurance is striving to improve service reliability and enhance customer satisfaction by modernizing the operations behind its core insurance services—including policy management, claims processing, payments, and the customer portal.

However, the Site Reliability Engineering (SRE) team currently faces significant challenges in maintaining seamless service delivery and minimizing downtime during system incidents.

The incident management process is predominantly manual, requiring SREs to monitor disparate system dashboards, interpret performance metrics, analyze logs, and identify root causes before applying remediation. When services degrade or fail, it often leads to delays in detection and resolution, affecting both internal workflows and end-user experience.

Moreover, the lack of intelligent automation in incident response extends resolution times, introduces operational inefficiencies, and increases the risk of prolonged service disruptions. This reactive approach limits the team's ability to proactively prevent issues or scale operations effectively.

To summarize, some of the key challenges faced by SREs at MetLife Insurance include:

- Manual and fragmented processes slow down the detection, diagnosis, and resolution of service incidents.
- Reliance on human expertise to interpret logs and metrics delays response time during critical outages.

- Lack of integrated tools hinders rapid identification of root causes and effective execution of fixes.
- Extended service downtime negatively impacts customer experience and trust in digital insurance services.

An intelligent, autonomous SRE AI Agent equipped with advanced observability, automated root cause analysis, and pre-defined remediation playbooks could dramatically improve system resilience—enabling faster, more reliable, and scalable operations while reducing the burden on human engineers.

Objective

MetLife Insurance plans to implement an AI-powered SRE Agent Assistant to support their Site Reliability Engineering (SRE) team in maintaining service reliability and minimizing system downtime. The goal is to create an AI-powered agentic solution that assists SREs in executing the following tasks:

- **Check service health**
Check the health status of a service in the system.
Queries the service registry to retrieve current health metrics and determine if a service is functioning accordingly. This helps verify service availability, troubleshoot incidents, validate deployments, and assess the health of upstream/downstream dependencies.
- **Restart services**
Perform a service restart to recover from degraded or unhealthy states.
Initiates a restart of the specified service—commonly used as a remediation action during incident response when a service is unhealthy.
- **View incident history**
Retrieve historical incident data for pattern analysis and reporting.
Fetches past incidents to identify recurring issues, support root cause analysis, and improve incident response over time.
- **Find unhealthy services**
List services that are degraded or unhealthy.
Enables quick identification of problematic services to prioritize investigation and resolution.
- **List all services**
Return the current health status of all services.
Provides a unified view of overall system health, aiding in monitoring, incident management, and proactive maintenance.

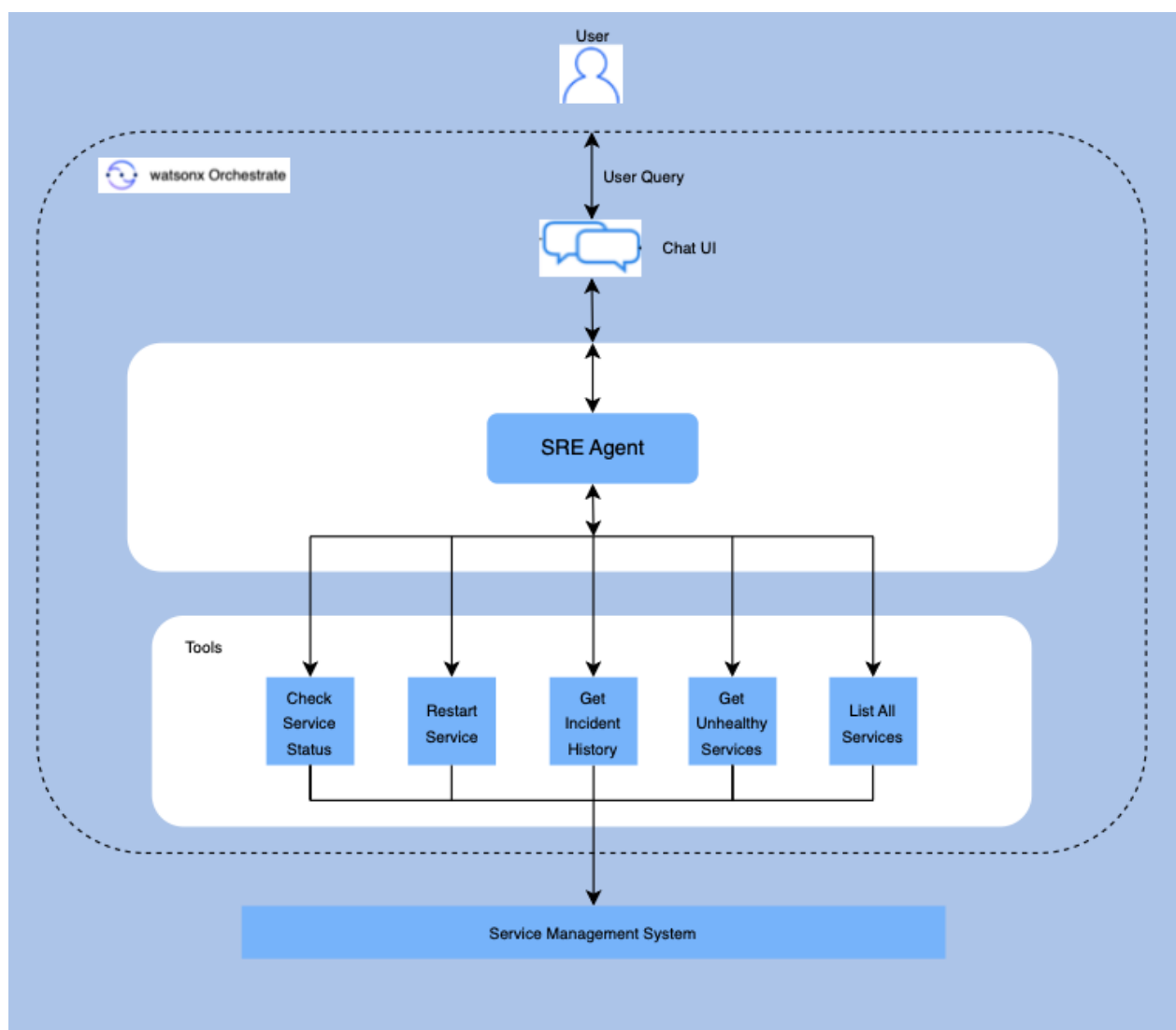
By automating these tasks, MetLife aims to reduce operational overhead, accelerate incident response, and improve system resilience—ultimately enhancing customer experience and trust in their digital insurance services at scale.

Business Value

Accelerating Incident Resolution & Reinvesting Human Capital

- An SRE AI agent that autonomously troubleshoots and remediates issues can dramatically reduce Mean Time to Detect (MTTD), Mean Time to Acknowledge (MTTA), and Mean Time to Resolve (MTTR).
- By leveraging intelligent automation and contextual reasoning, the agent identifies root causes faster than manual processes, minimizing downtime and improving service reliability.
- Reducing MTTR translates directly into cost savings and improved customer experience. Every minute of reduced outage time prevents revenue loss and protects brand reputation.
- By offloading repetitive troubleshooting tasks to an autonomous agent, SRE teams reclaim valuable time.
- Instead of firefighting, engineers can focus on strategic initiatives such as capacity planning, performance optimization, and building reliability features into the architecture.

Architecture



Pre-requisites

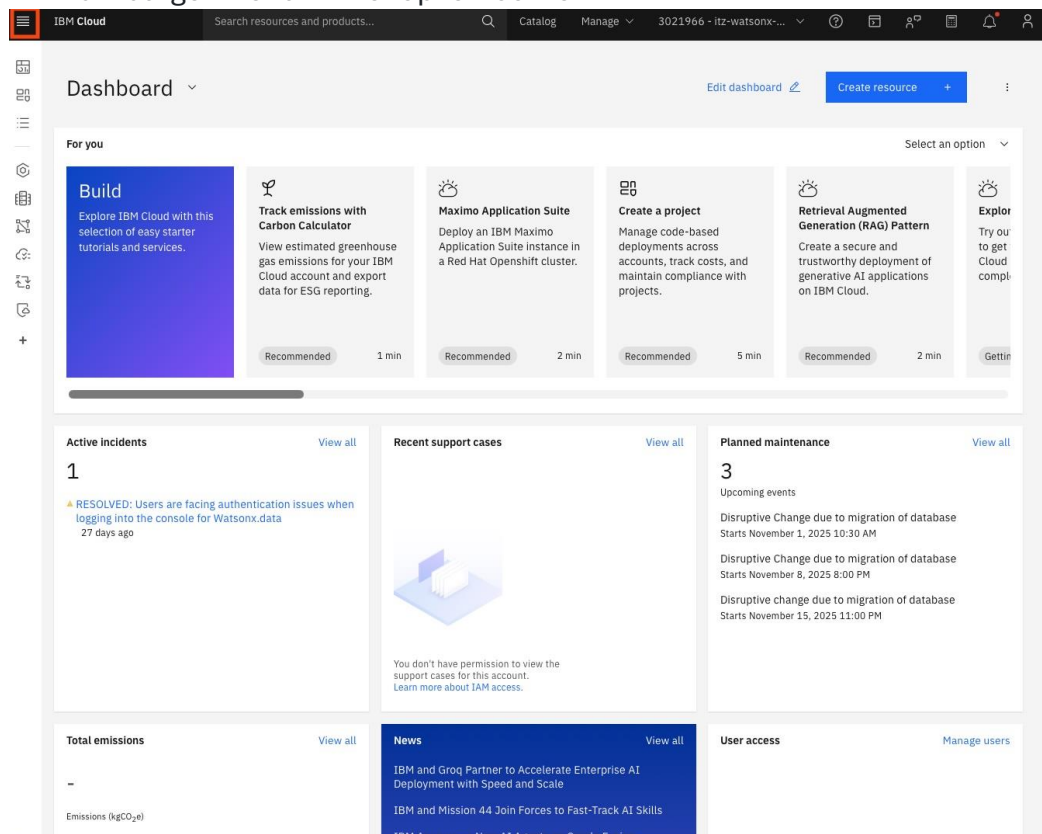
- Check with your instructor to make sure all systems are up and running before you continue.
- Validate that you have access to the right techzone environment for this lab.
- Validate that you have access to a credentials file that your instructor will share with you before starting the labs.

Accessing Your Watsonx Studio and adding the Python Notebook

- A TechZone environment has been reserved for you, and you will receive the access details via email.
- To access your Watsonx Orchestrate and Studio instance, **please check with your instructor** for the access link and the necessary steps.

Step by step instructions:

1. When you launch IBM Cloud, you'll be directed to this page. Click on the hamburger menu in the top left corner:



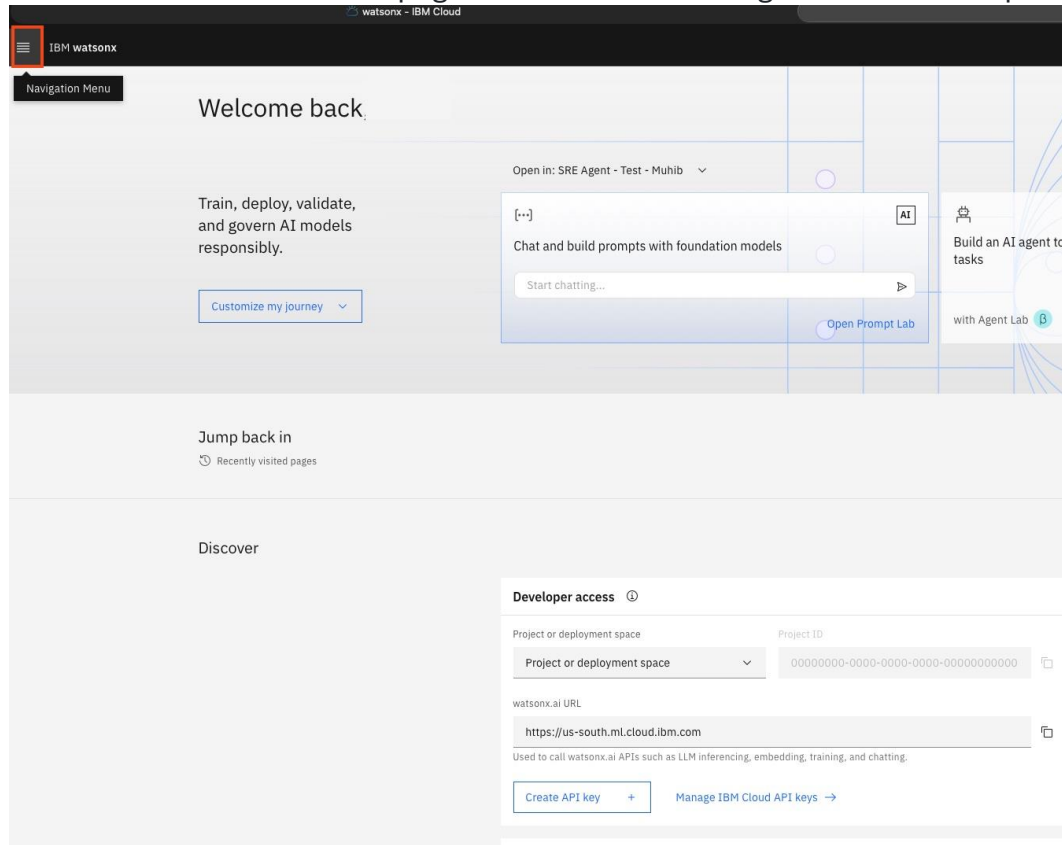
2. Click on **watsonx**:

The screenshot shows the IBM Cloud dashboard. In the left sidebar, the 'watsonx' resource card is highlighted with a red box. The main content area displays a grid of recommended resources, including 'Track emissions with Carbon Calculator', 'Maximo Application Suite', 'Create a project', 'Retrieval Augmented Generation (RAG) Pattern', and 'Explore'. Below this grid, there are sections for 'Recent support cases', 'Planned maintenance', 'News', and 'User access'.

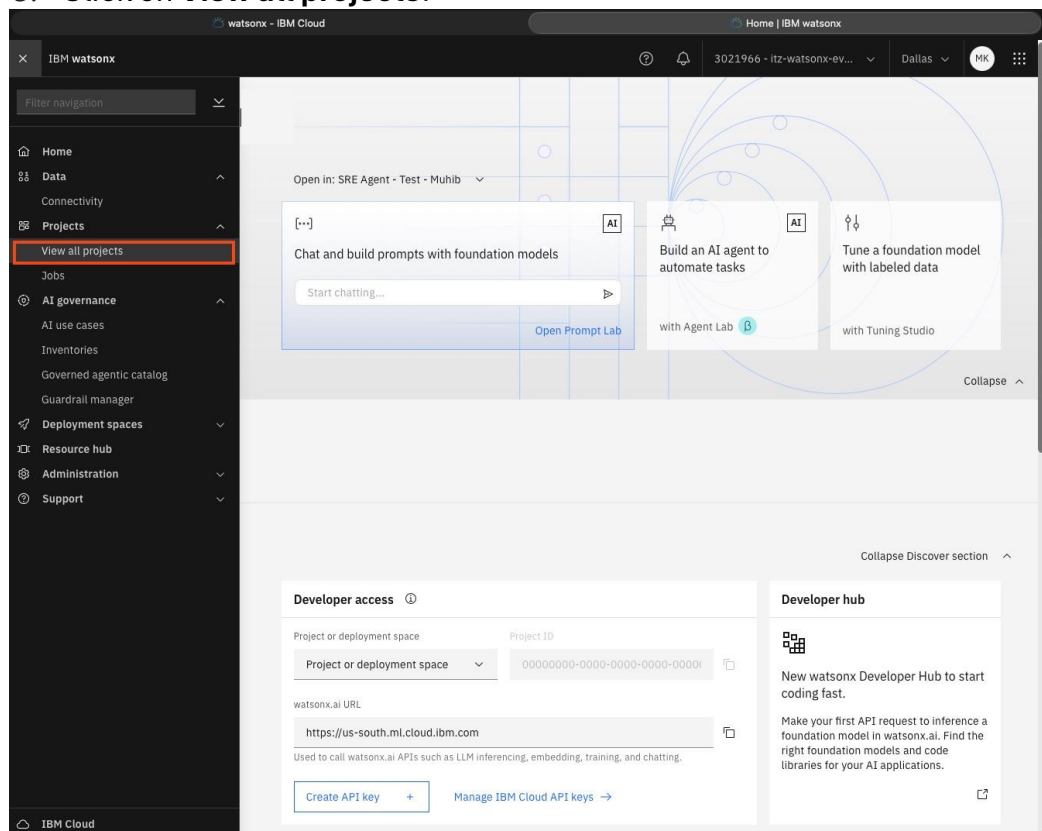
3. Click on Launch:

The screenshot shows the IBM Watsonx 'Build AI solutions with IBM watsonx' page. The 'Launch' button is highlighted with a red box. The page features a dark sidebar with navigation links for 'watsonx', 'Documentation', and various Watsonx services. The main content area includes a header section, a 'Data Fabric' section, and a 'watsonx.data' section. Below these, there are several cards for different Watsonx services, including 'watsonx.ai', 'watsonx.governance', 'watsonx.data', 'watsonx.data intelligence', and 'watsonx.data integration'.

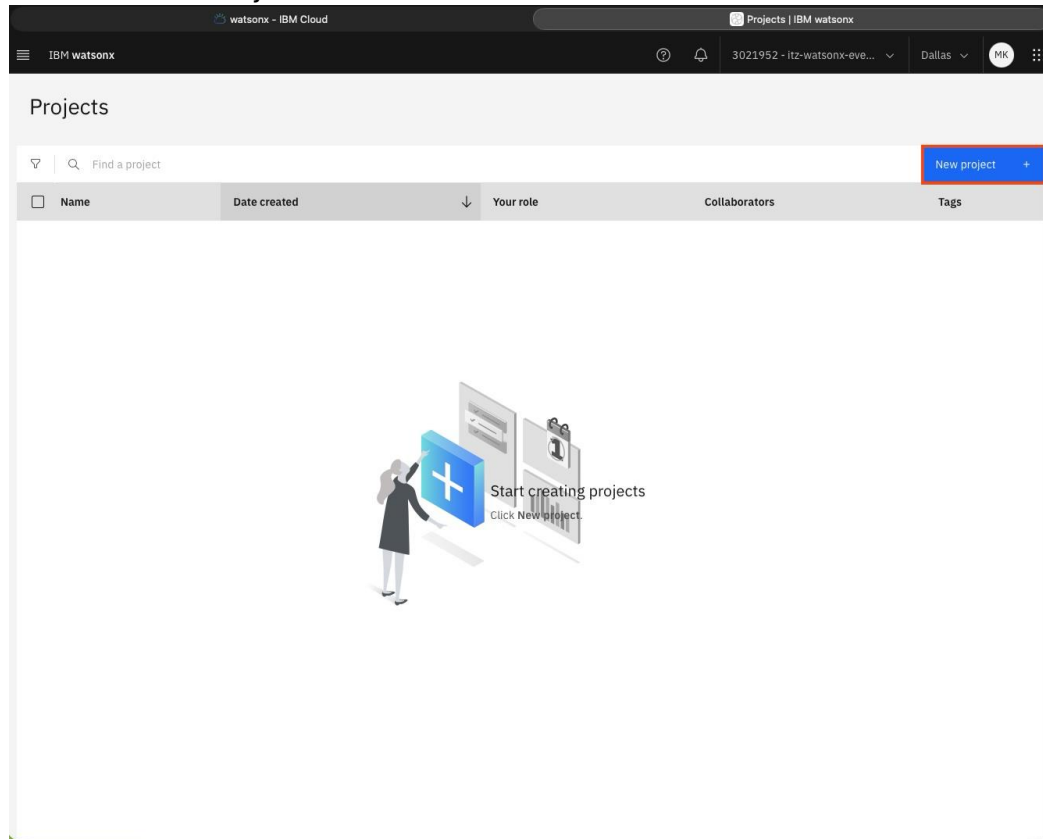
4. You'll be directed to this page. Click on the hamburger menu in the top left corner:



5. Click on **View all projects**:



6. Click New Project +



7. Create a new project and name it

A screenshot of the 'Create a project' form in IBM watsonx. The browser address bar shows 'New project | IBM watsonx'. The page header includes 'IBM watsonx' on the left and user information '3021952 - itz-watsonx-eve...', 'Dallas', and 'MK' on the right. The main heading is 'Create a project' with a subtext 'Start with a new, blank project or select from where to import an existing project.' On the left, a sidebar shows options: '+ New' (highlighted with a red box), 'Local file', and 'Sample'. The main form area is titled 'Define details' and contains the following fields: 'Name' (with 'SME Agent Notebook' entered and highlighted by a red box, and a small 'Enter a name' button below it), 'Description (optional)' (with the placeholder text 'What's the purpose of this project?'), 'Tags (optional)' (with an 'Add tags' button), 'Storage' (with the value 'cos-itx-wxo-68dec3b9e2b86cbdfa15df'), and 'Advanced settings' (with a dropdown arrow). At the bottom right, there are 'Cancel' and 'Create' buttons.

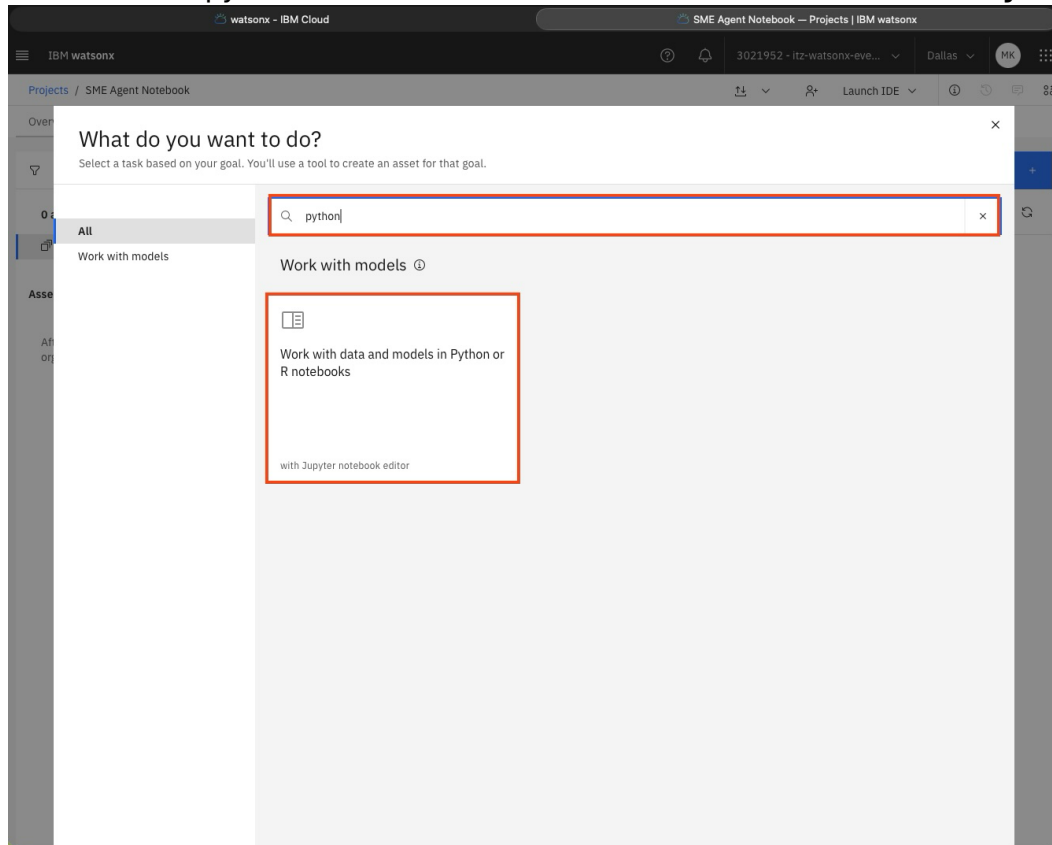
8. Click **Assets**:

The screenshot shows the IBM watsonx SRE Agent Notebook interface. The top navigation bar includes the IBM watsonx logo, a search bar, and a user profile. The main navigation bar has tabs for Overview, Assets (highlighted with a red box and an arrow), Jobs, and Manage. Below the navigation bar, there's a 'Start working' section with four cards: 'Add users as collaborators', 'Add data to work with', 'Chat and build prompts with foundation models', and 'Tune a foundation model with labeled data'. Below this, there's a 'View all' link and a 'Collapse' button. The main content area is divided into three columns: 'Jump back in' (showing assets created with tools), 'Resource usage' (showing metrics for CUH, Tokens, Hosting hours, and Pages), and 'Your documentation' (with a 'New!' badge and a link to 'Open Documentation editor'). The 'Project history' section shows a recent project creation.

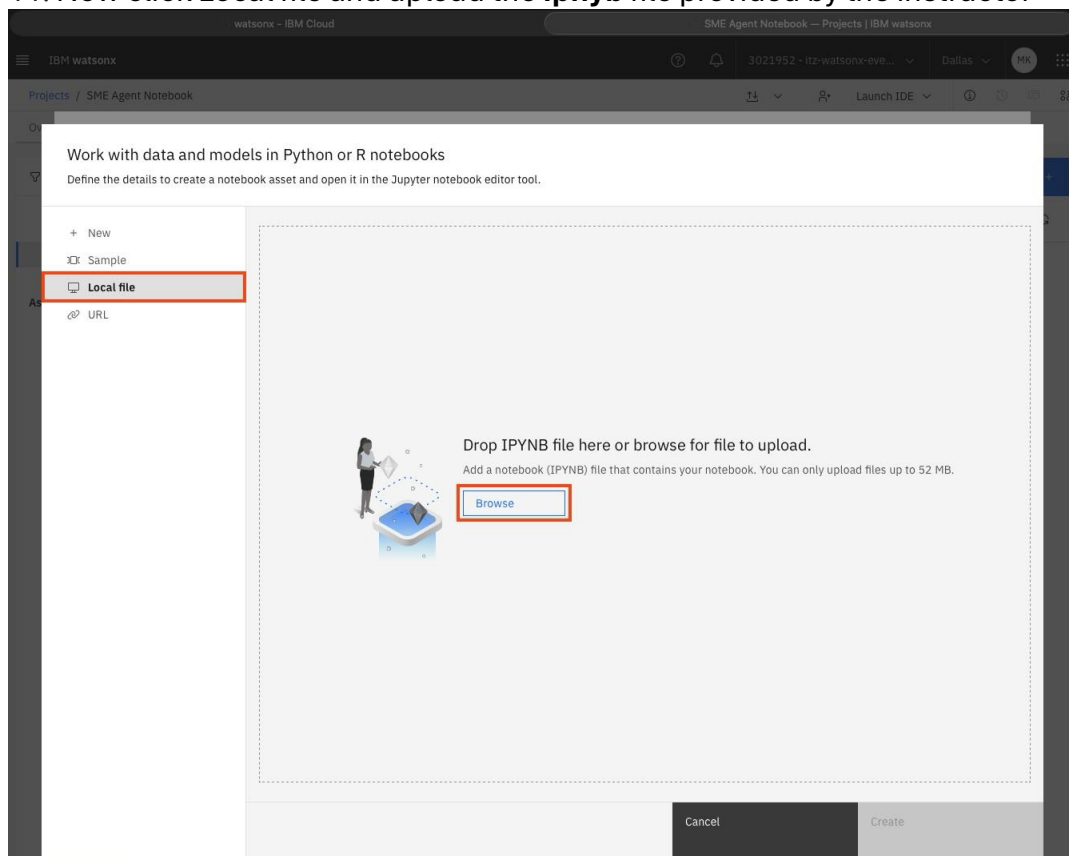
9. Click on **New Asset +**:

The screenshot shows the IBM watsonx SRE Agent Notebook interface with the 'Assets' tab selected. The top navigation bar is the same as in the previous screenshot. The main navigation bar has tabs for Overview, Assets (selected), Jobs, and Manage. Below the navigation bar, there's a 'Find assets' search bar and a 'New asset +' button (highlighted with a red box). The main content area is divided into two columns: 'All assets' (showing 0 assets) and 'Asset types' (showing a list of asset types). The 'All assets' section has a 'Start working' message: 'To get started with assets, click **New asset** to create one in a tool, or **Import assets** to add existing ones.'

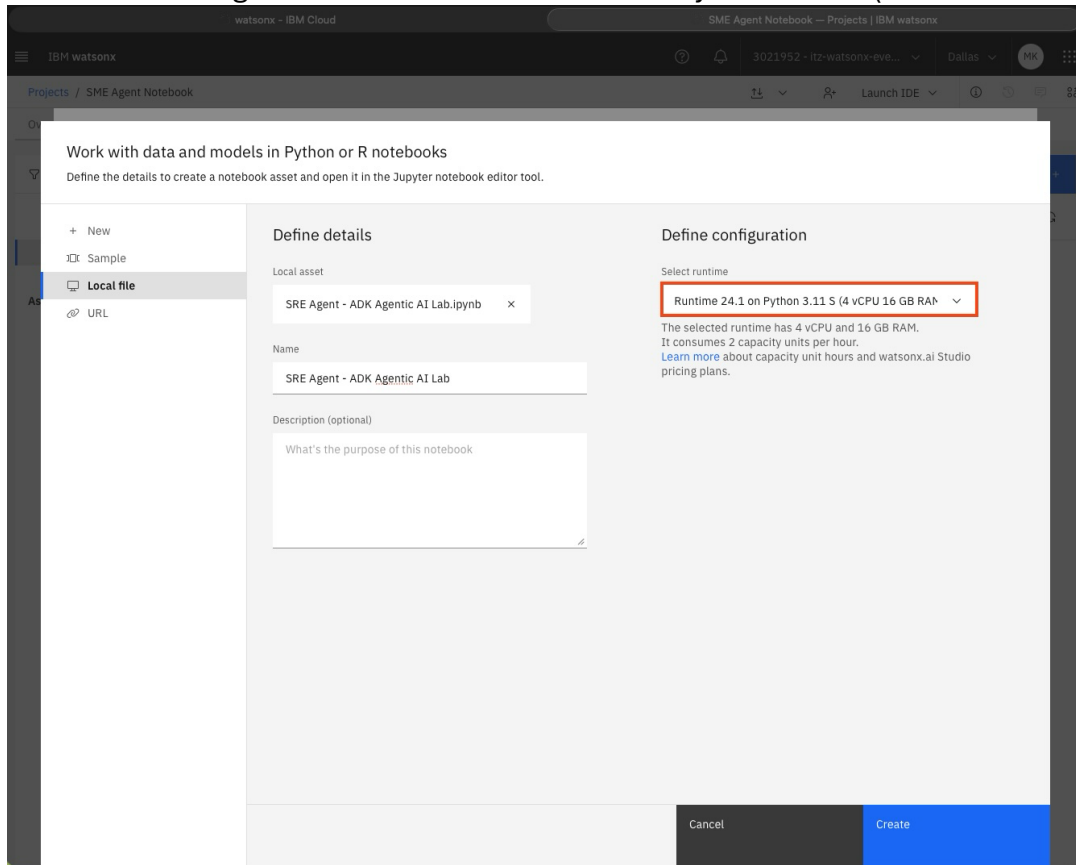
10. Search for python and then click on **Work with data and models in Python or R notebooks**:



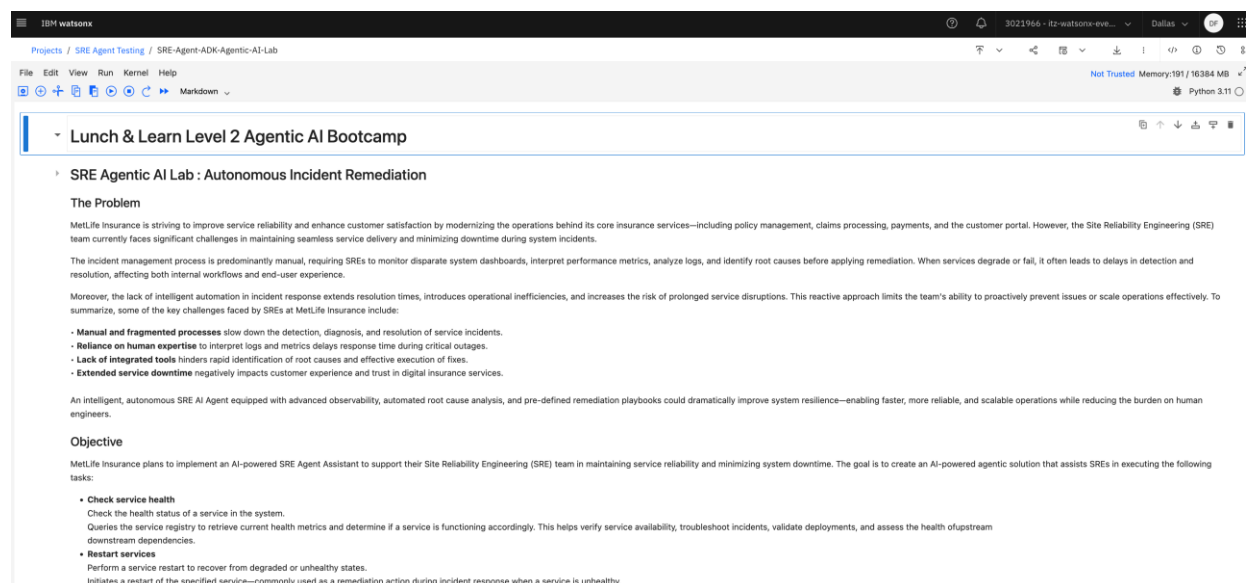
11. Now click Local file and upload the **ipnyb** file provided by the instructor



12. For the configuration select Runtime 24.1 on Python 3.11 S (4 vCPU 16 GB RAM)



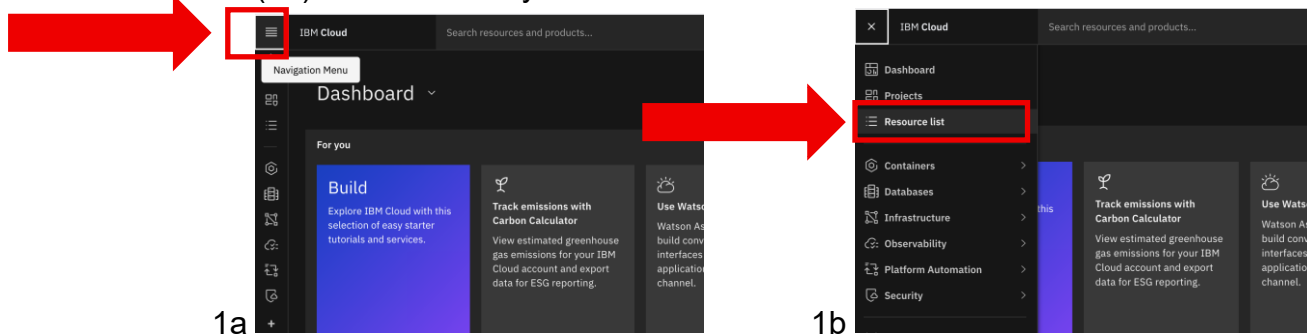
13. You should end up with this page with the notebook loaded and follow the instructions within the notebook.



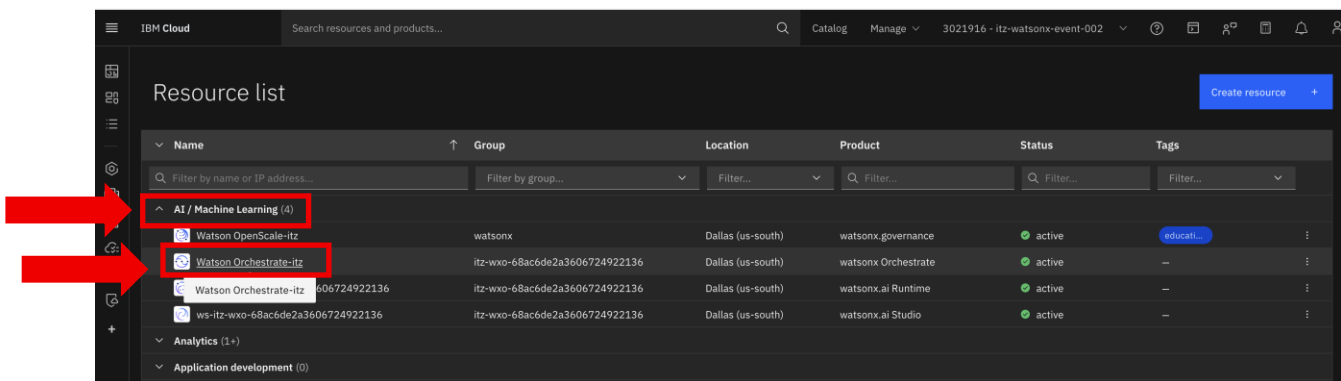
Accessing Your Watsonx Orchestrate Instance :

During the lab you will use the watsonx Orchestrate ADK to import agents, tools, knowledge base to the Orchestrate environment. Once you have imported these assets you can access the Orchestrate instance using the instructions below

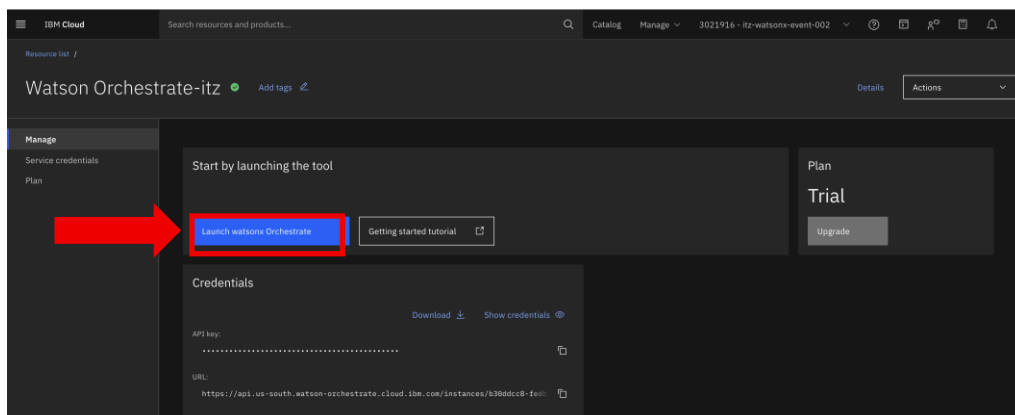
- 1.) Navigate to the [IBM Cloud](#) Home Page. To view IBM Resources select the “Navigation Menu” button on the top left (1a), and then select the “Resource List” button (1b). This will take you back to the Resource List section.



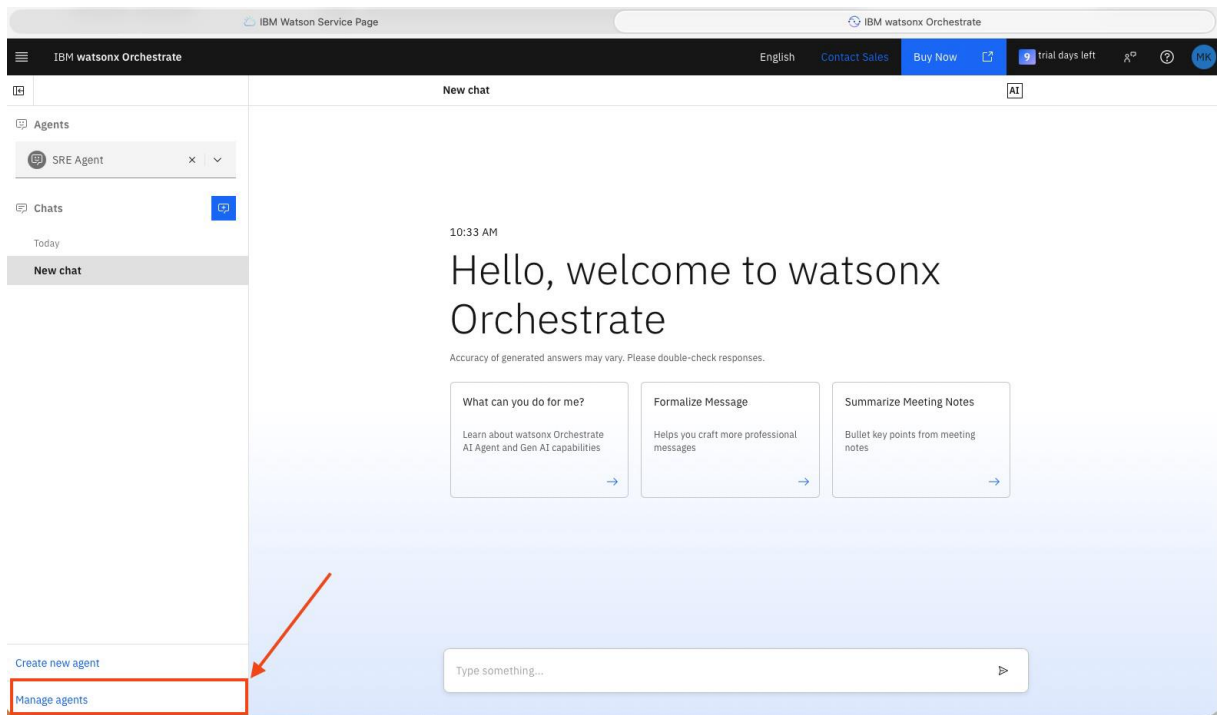
- 2.) Select the “AI/Machine Learning” Drop Down on the left. And then click the Resource “Watson Orchestrate-itz”



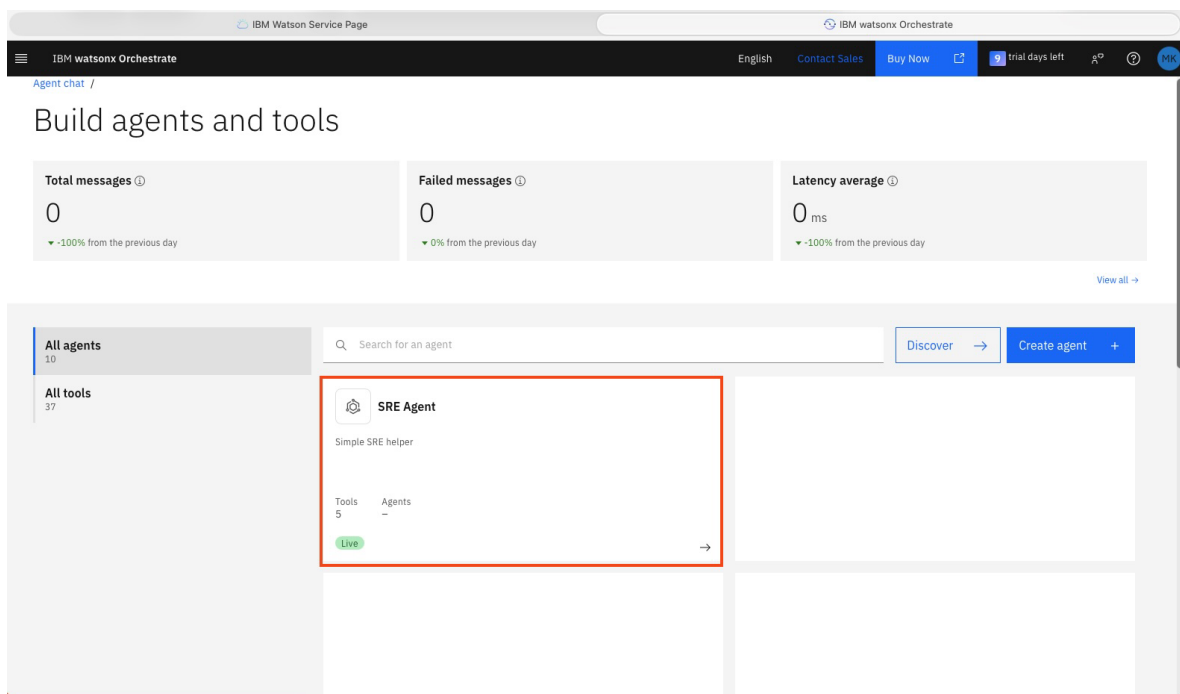
- 3.) Select “Launch watsonx Orchestrate”



4.) You have now logged into watsonx Orchestrator. Click Manage Agents on the bottom



5.) You will navigated to the list of the agents created. Click on the agent you created.



6.) Now, you can start testing the agent you have created!

The screenshot displays the IBM watsonx Orchestrate web interface. The top navigation bar includes the IBM Watson Service Page link, the product name 'IBM watsonx Orchestrate', and links for 'English', 'Contact Sales', 'Buy Now', and 'trial days left'. The main header shows the agent name 'SRE Agent' and the model 'llama-3-2-90b-visi...'. The left sidebar contains tabs for 'Profile', 'Knowledge', 'Toolset', 'Behavior', and 'Channels' (which is highlighted with a 'Preview' label). The 'Profile' tab is active, showing a 'Description*' field with the text 'Simple SRE helper' and a 'Welcome message' field with the text 'Hello, welcome to watsonx Orchestrate'. A 'Quick start prompts' section is also visible. The right panel, titled 'Preview', shows a live chat interface with a timestamp of '10:38 AM' and a large welcome message: 'Hello, welcome to watsonx Orchestrate'. Below this, there are three suggested prompts: 'What can you do for me?', 'Formalize Message', and 'Summarize Meeting Notes'. A red box highlights the input field at the bottom of the chat interface, which contains the placeholder text 'Type something...'.