Secure Data Science Windows Workspace User Manual

Relevance Lab India Pvt Ltd

May 2025

Version 1.1.0

Table of Contents

1. Introduction

Relevance Lab is a specialist IT services company with re-usable technology assets in DevOps, Cloud, Automation, Digital, Service Delivery and Supply Chain Analytics that help global organizations achieve frictionless business by transforming their traditional Infrastructure, Applications and Data.

2. Installed Software Overview

This AWS Marketplace offering comes with essential development tools pre-installed to facilitate data science, software development, and system administration. Below is a list of available software:

- **Remote Access:** Powered by Amazon NICE DCV for high-performance remote desktop sessions.
- **Languages & IDEs:** Includes Python 3.x, R, Visual Studio 2022, VS Code, PyCharm, and RStudio.
- **ML & Data Frameworks:** Preinstalled with PyTorch, TensorFlow, scikit-learn, PySpark, Dask, and Vowpal Wabbit.
- **Notebook & Analytics Interfaces:** Supports Jupyter Notebook and JupyterLab for interactive development.
- **Utilities & Dev Tools:** Comes with Git, Google Chrome, AWS CLI, and 7-Zip for development and system tasks.
- **Environment Management**: Features Anaconda for Python/R package management and Docker + Docker Compose for containerized workloads.
- **Office Productivity:** LibreOffice suite (Writer, Calc, Impress) for document editing and reporting.

3. Accessing & Using Secure Desktop Applications

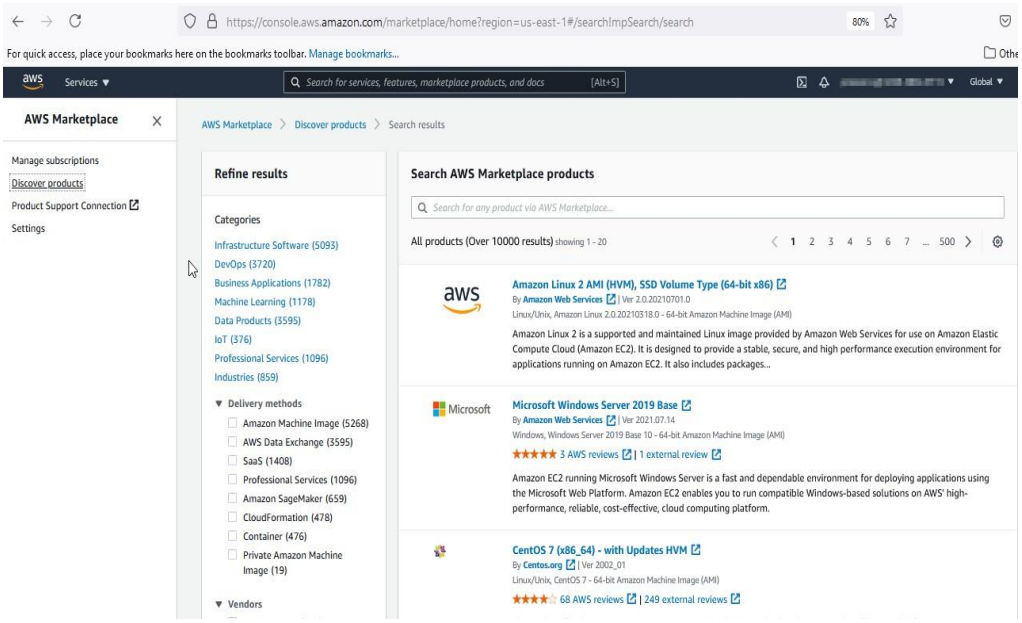
This section provides instructions on how to connect to the machine access pre-installed applications like RStudio Server, JupyterLab, and Visual Studio Code for development and data analysis

3.1 Subscribe to AWS Marketplace

You need to subscribe to **Data Science Workspace** product from AWS Marketplace.

Login to your AWS account

a. In the Services, search for AWS Marketplace. Go to AWS Marketplace subscriptions
b. Navigate to AWS Marketplace
c. On the right side, click on the “Manage Subscriptions” and navigate to “Discover Products” which is on the left panel.



d. In the "Search AWS marketplace products" box, input **Secure Data Science Workspace** in the search results, select **Secure Data Science Workspace**.

e. Page will be navigated to the AWS Marketplace with detailed description of the Product.

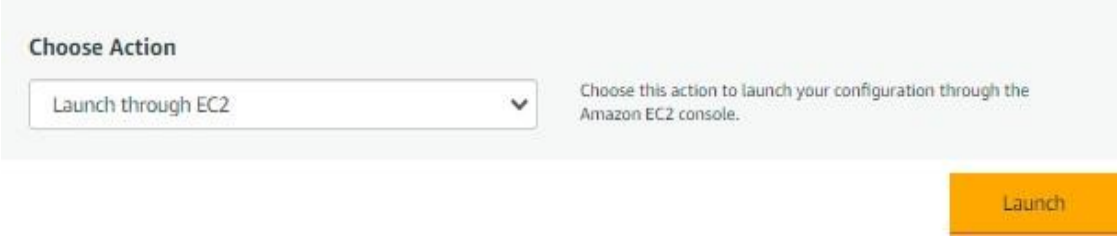
f. See the pricing and usage information and click "**Continue to Subscribe**". Kindly read the terms and conditions offered and click on
Accept Terms. Wait for a while to see the effective date of subscription.

g. Click "**Continue to Configuration**". Details such as AMI id, software version, Region and Product code are displayed in this page.

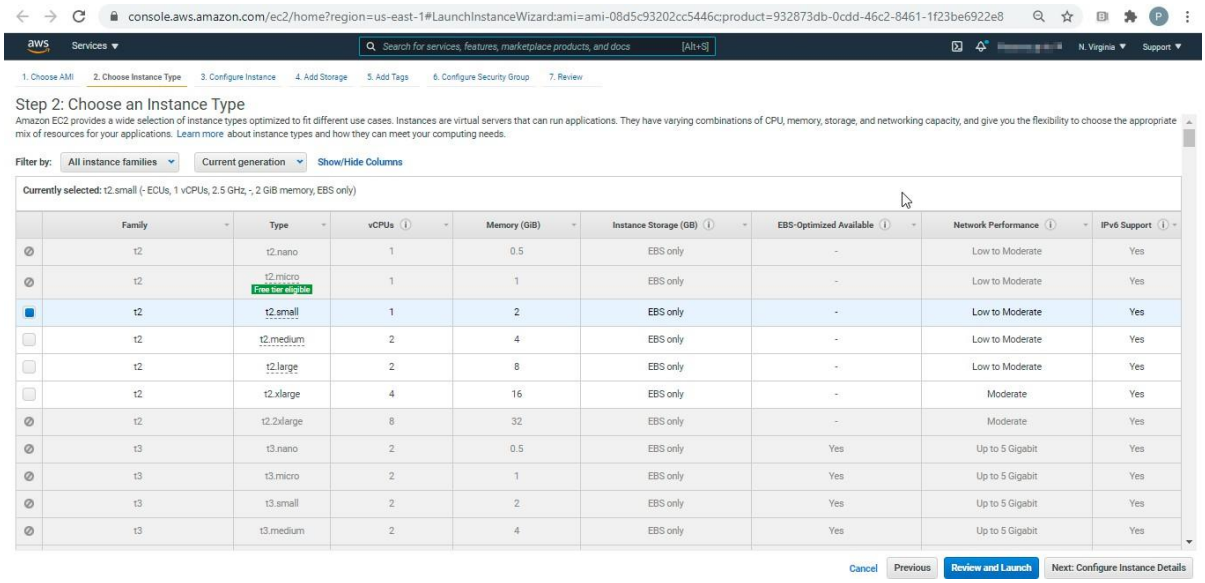
h. Click "**Continue to launch**".

i. Click "**Usage instructions**" to check a brief information about which instance type to be selected and how to launch the same.

j. Go to "**Choose Action**" dropdown and select "**Launch through EC2**".



k. Click "**Launch**", will open your AWS account, EC2 launch console to choose instance type as below.



3.2 Prerequisites

Before you create an instance, you need to have the following resources available in your AWS account.

- VPC in which to launch your instance. This can be the default VPC, or you can create a custom VPC
- Subnets within the VPC
- Route Tables
- Internet Gateway
- Security Groups

Once you have the above pre-requisites in place you can create an EC2 instance from the AMI. Setting up the prerequisites.

3.2.1 Create VPC.

- Login to AWS
- Switch to N. Virginia (or any region of your choice)
- Search for VPC under Services. Click on VPC. This will take you to the VPCs console.
- Click on Your VPCs in the left panel. You will see all your VPCs listed.
- Click on Action and click on “Create default VPC”.
- Click on Create default VPC. VPC is created. Click on the close.
- In the “Your VPCs” table, select the VPC you created, Input VPC name and save it.

Ref: Documentation from AWS on creating VPC

A default VPC is created along with public subnets in each availability zone. Select one of those zones. This is where you will be launching an AMI. You can give a name to the subnet for easier reference. Click “Subnets” in the left panel. Select a subnet associated with the above VPC. Input the Subnet name and save.

3.2.2 Create Security Groups.

Let’s create a Security group now. In the Services search for EC2 and click on EC2 in the results. You should see the EC2 console.

- Click “Security groups” in the left side panel.
- Click on the default security group which is linked to the VPC Id of the VPC created.
- Click the “Action” tab followed by view details.
- Click the “Edit inbound Ports” button from inbound rules descriptions.

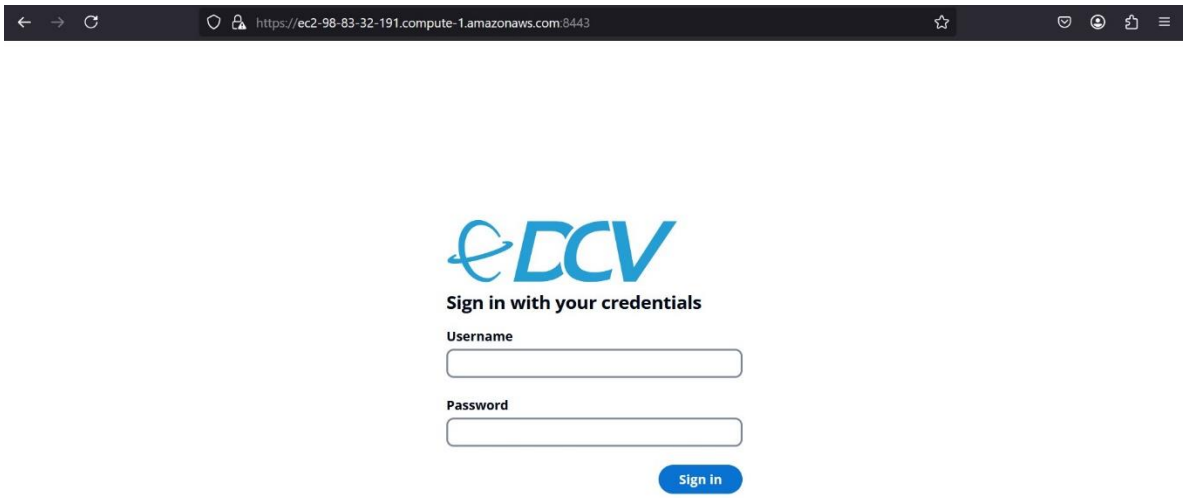
- Add the following Inbound ports to make the application work.
- Add a rule to allow HTTP traffic with source Anywhere (0.0.0.0/0)
- Add a rule to allow HTTPS traffic with source Anywhere (0.0.0.0/0)
- Add a rule to allow SSH traffic with source Anywhere (0.0.0.0/0)
- Click “Save rules”. Ensure you have added 3 inbound rules (80, 8443, 443) and verify.
- Allow all traffic and all IP addresses for Outbound rules.

3.3 Access the Desktop GUI

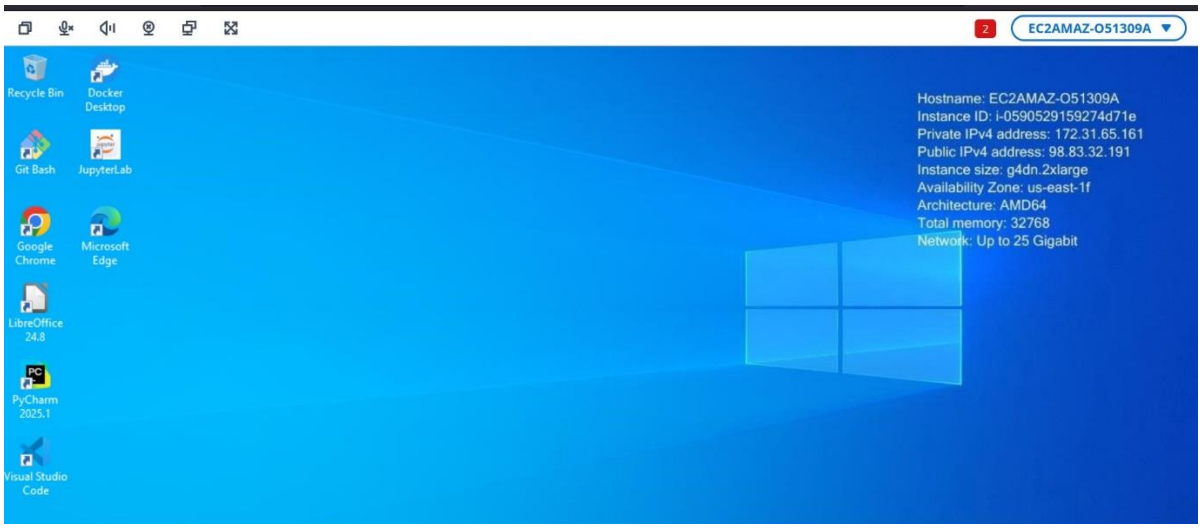
After the instance is up and running, retrieve the Public DNS Address from the EC2 console. Then, open a web browser and enter https://<Public DNS Name>:8443, which will take you to the DCV authentication page. Use the default credentials & Click Login to proceed.

Username: Administrator
Password: Please follow the instructions to get the password shared (https://repost.aws/knowledge-center/retrieve-windows-admin-password)

RDP access is disabled by default to enhance the security of remote connections to the machine



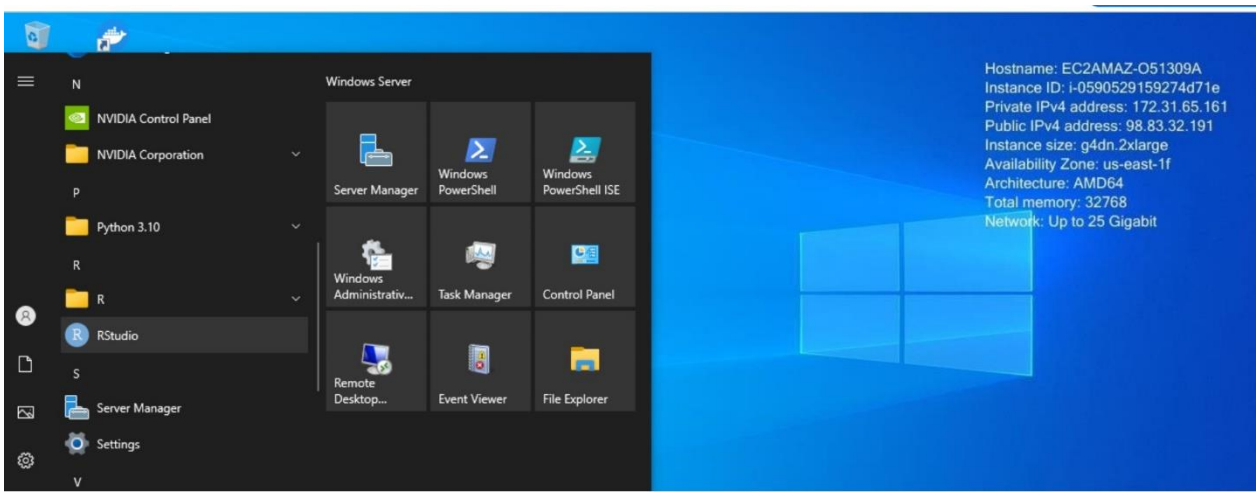
After logging in, you will be taken to the Desktop environment via Amazon DCV as below



3.4 Using RStudio Server

To access RStudio Server and begin analysing data with R, follow these steps:

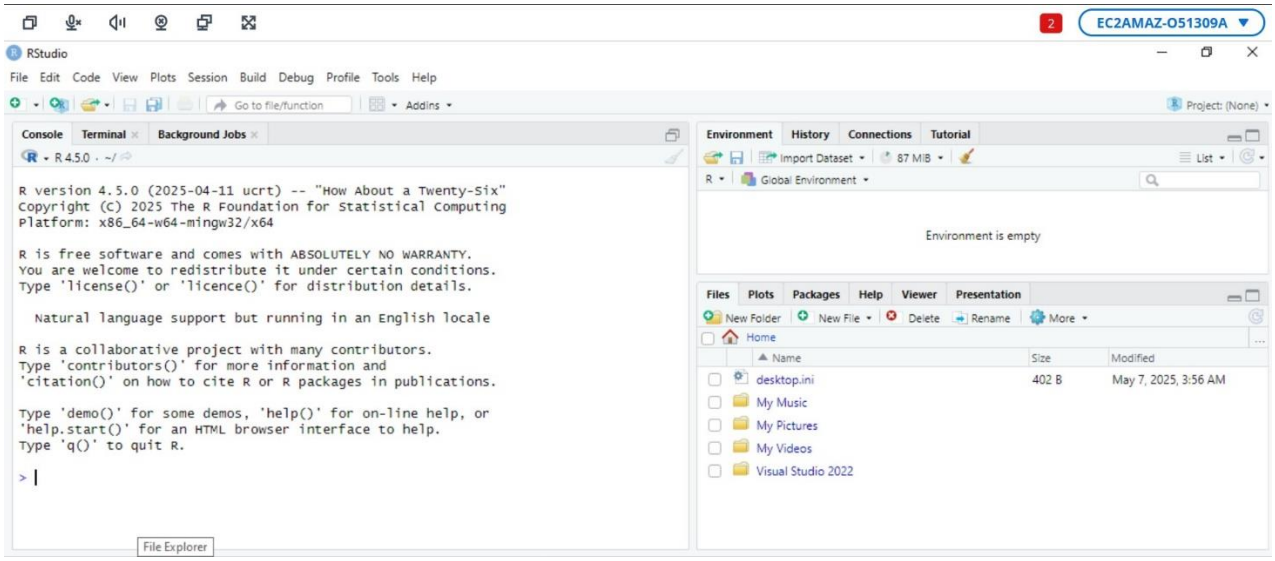
1. Open your preferred web browser (such as Google Chrome or Mozilla Firefox). From the desktop environment, click on the Start menu, then select RStudio. This will launch the RStudio application in your browser and take you directly to the RStudio home page.



No login credentials are required. For the best experience, we recommend using either Chrome or Firefox. Once opened, you should see the RStudio interface as shown below.

2. Start Using RStudio

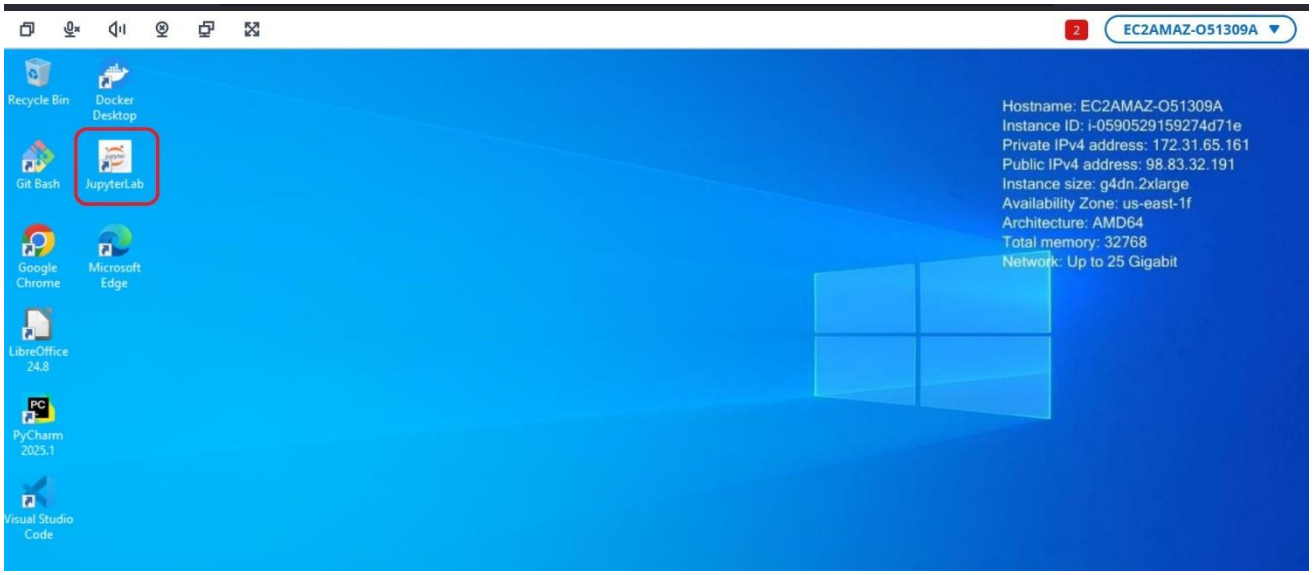
Once you've successfully logged in, you'll be taken to the RStudio home interface. From here, you can start creating and editing R scripts, perform data analysis, and take advantage of RStudio's built-in tools for data visualization, package installation, and environment management. The interface you'll see upon logging in is illustrated below.



3.5 Opening JupyterLab

To access JupyterLab for interactive data analysis and notebook-based development, follow these steps:

1. Open your preferred web browser (such as Google Chrome or Mozilla Firefox). From the desktop, double-click the JupyterLab shortcut icon.

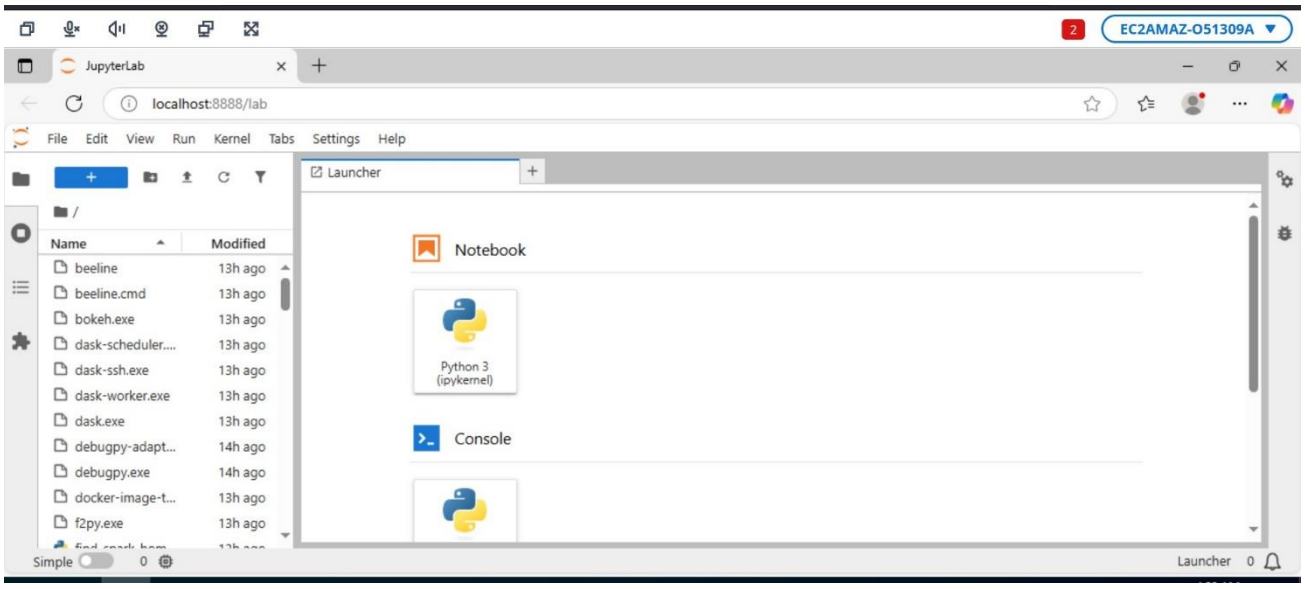


2. Access JupyterLab

A web browser will open automatically, taking you directly to the JupyterLab interface without requiring login credentials.

3. Start Using JupyterLab

Once inside JupyterLab, you can create and manage Python and R notebooks, write scripts, and perform interactive data analysis.



3.6 Terminate your instance

Once you have determined that the instance is no longer required, you can proceed with terminating it to stop any further charges. Follow these steps to terminate the instance properly:

1. Access AWS Marketplace
   - Navigate to the AWS Marketplace and click on Your Software to view your purchased and active instances.
2. Locate the Instance
   - On the Your Software page, find the instance you want to terminate.
   - Next to the instance, click on Manage in AWS Console, which will redirect you to the AWS Management

Console.

3. Terminate the Instance in EC2 Console
   - In the AWS Management Console, open the EC2 console to view all running instances.
   - Locate the instance you wish to terminate, right-click on it, and select Terminate from the options.
4. Confirm Termination
   - A prompt will appear asking for confirmation. Click on Yes, terminate to proceed.
   - The instance will now begin the termination process and will be permanently removed once completed.

By following these steps, you can safely decommission your instance when it is no longer needed, ensuring that you are not billed for any unused resources.

4. Summary

This document serves as a detailed guide for accessing and working within the Secure Desktop environment. The environment is pre-configured with a rich set of tools and applications tailored for data science and software development workflows. Included by default are popular platforms such as RStudio Server, JupyterLab, Visual Studio Code, Docker, Anaconda, LibreOffice, and Git.

Additionally, the system comes preloaded with a wide range of Python libraries and frameworks, including PyTorch, TensorFlow, scikit-learn, PySpark, Dask, and Vowpal Wabbit, enabling users to dive straight into machine learning, data processing, and analytical tasks with minimal setup

4.1 Security Best Practices

- **Regularly update** installed software packages to ensure security patches are applied.
- **Restrict access** to the instance using AWS **security groups and IAM roles**.
- **Monitor resource usage** and optimize instance types based on workload requirements.

4.2 Next Steps

- **Expand functionality** by installing additional R and Python packages using CRAN or Conda.
- **Integrate with AWS services** like **S3 for data storage**, **RDS for databases**, and **IAM for secure authentication**.