CloudBank Office Hours

07-JAN-2024: Get Started + NAIRR

Screencaps courtesy of ***flameshot*** (free)

Caveat: This Google slide deck is in preparation: Upgrade from a version presented 12-DEC-2024. Please treat the content as provisional.

A Note on AI Stacks for NAIRR participants

- Cloud providers are in a frenetic rush to innovate
- The only constant of cloud AI is change
- This deck is a snapshot: Be prepared to bootstrap further

Today:

- This overview (8 slides)
- Checkpointing (5 slides)
- CloudBank (4 slides)
- Fly-over of core actions: Azure, AWS, GCP (90 slides [sic])
- Cloud AI stacks (36 slides: 6 general + 3 x 10)

This slide deck is public and subject to frequent change.

We also include links to more methodical documentation.

CloudBank Research Model

- Builders: Create data science cyberinfrastructure (CI)
- Users: Do their research on the CI
- Admins: optimally manage cloud resource use
- Educators: Bring the research program into the classroom

Research team needs: Operational, technical, curriculum

CloudBank facilitates research by providing support to Builders, Users, Admins and Educators

Patterns

- CloudBank is always looking for patterns and templates
- One favorite example is the “Littlest Jupyter Hub” from UC Berkeley
    - This enables data science exploration on the “small team scale”

https://tljh.jupyter.org/

- A word on the efficacy of working with ephemeral students

Cloud use themes (1 of 5): Interaction with the cloud

3 ways to interact with a particular cloud

- A Console or Portal
- Command Line Interface (CLI)
- Application Programming Interface (API)

Cloud use themes (2 of 5): Data storage

Three types of data storage

- Block storage: Familiar “disk drive”, files
- Object storage: Cheaper per byte, unlimited, objects
  - Relative cost: 9 versus 2.3 cents per GB per month
- Database

Cloud use themes (3 of 5): Virtual Machines

Treat cloud use as an optimization problem

- A small investment of your time
- Double or triple the amount of computation you can do

Stopping VMs at night, learning machine images

Using preemptible VMs with checkpointing

Cloud use themes (4 of 5): Services

A plethora of services

- Do-it-yourself flavors are cheaper + require expertise
- Managed flavors are more expensive, less hassle

Cloud use themes (5 of 5): Laboratories

- Cloud providers offer managed “research laboratory” environments (and AI stack interface services)
  - Google Colab (Gemini etc)
  - Azure AI Studio (Azure OpenAI etc)
  - AWS SageMaker (Bedrock etc)

End Overview
Start Preemptible

Define Checkpointing

The practice of checkpointing: Storing the state of a computation, usually so that the computing task can be interrupted and then resumed “from where we left off”.

Cloud instances have appreciable (vast) unused capacity: Computers with nothing to do. These are understood to comprise a resource pool which is made available at reduced rates, often 50% to 90% discount. This is a “spot market”.

The use of spot market VMs comes with a catch: Machines can be preempted (taken away) on very short notice: 2 minutes.

This leads to an optimization problem for the Researcher: How to implement checkpointing as a hedge against preemption, with the end result that $10,000 can be used to purchase $30,000 of compute power?

Pointer to separate presentation

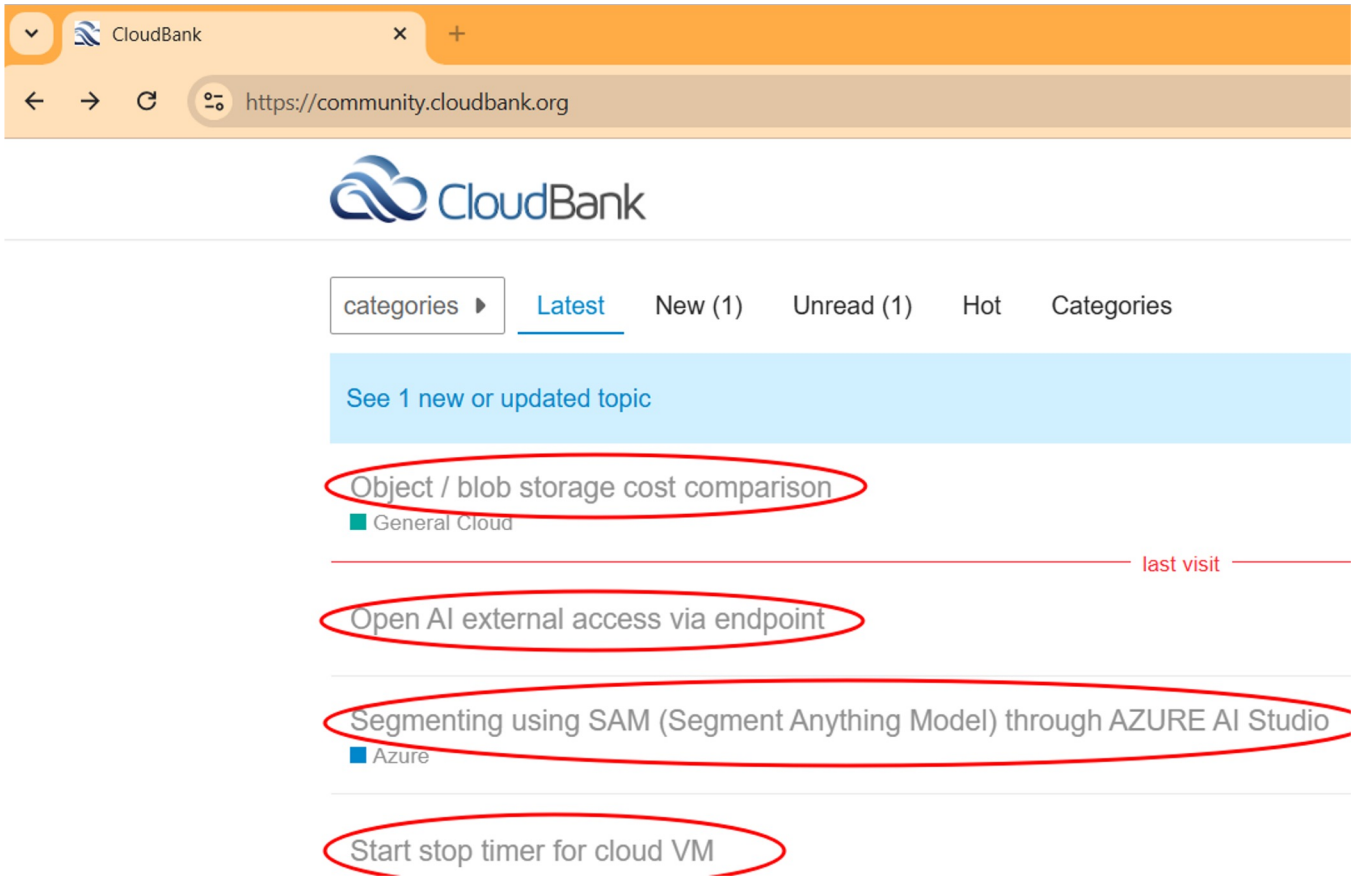
Preemptible instances and checkpointing are discussed in the Cloud Clinic deck first presented on 23-JAN-2025.

End Preemptible
Start CloudBank

CloudBank features

- CloudBank portal https://cloudbank.org >>> GCP, Azure, AWS etc
- CloudBank help help@cloudbank.org
- CloudBank community https://community.cloudbank.org Q&A
- CloudBank vendor connections… got a quota problem?
- CloudBank learning resources
  - Our videos and other content: Through the portal
  - Connections to vendor-provided content
- CloudBank cloud use machinery
- Operational tasks: Add team members, track spend

community.cloudbank.org

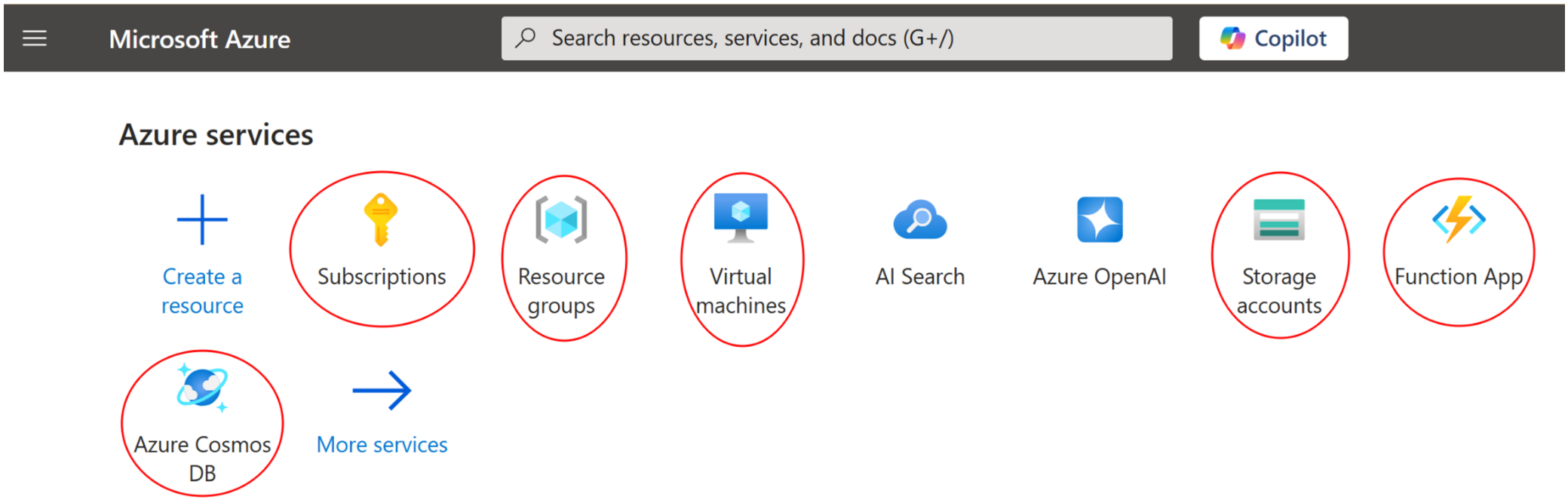


End CloudBank

Start Azure

Microsoft Azure (Azure) Fly-over

https://portal.azure.com



Azure Jargon

Subscription: Your Azure account where spend accrues

Resource Group: A logical container for related resources like VMs

Storage Account: A logical collection of data storage resources

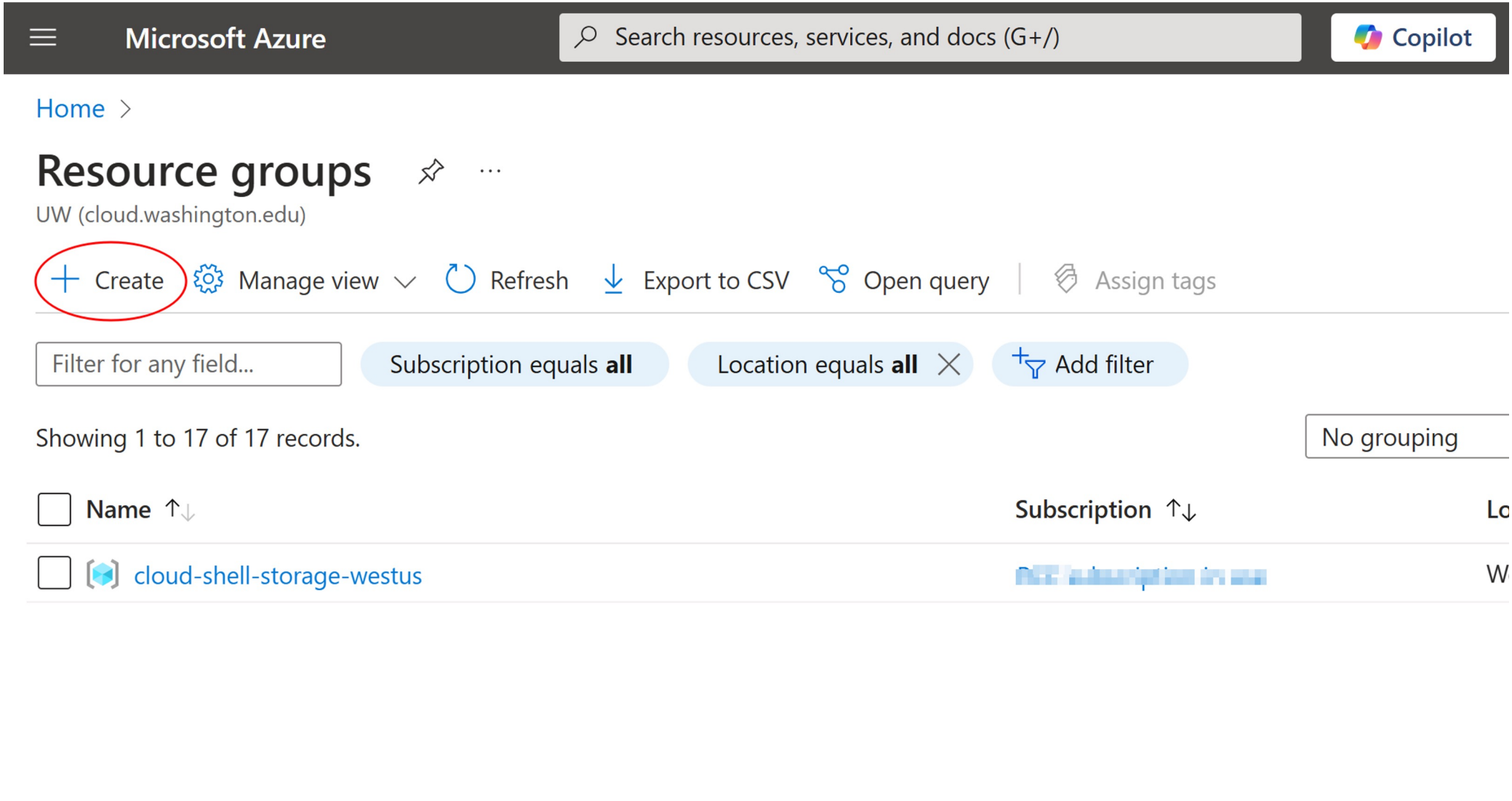
Blob: Object storage on Azure

Microsoft Azure (Azure) Fly-over

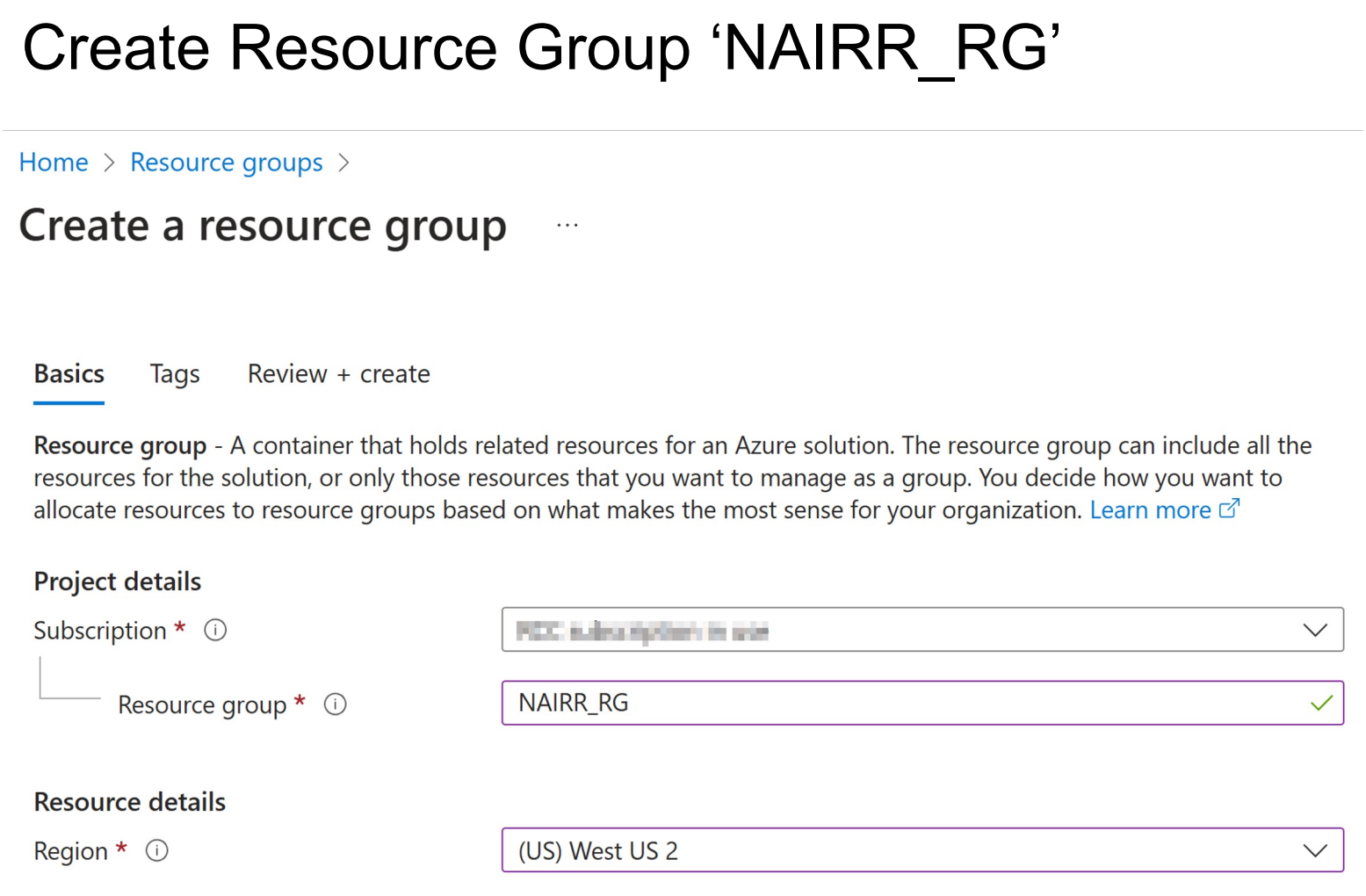
Plan

- Log in to the console
- Create a Resource Group
- Get a VM with an expanded root disk
  - Log in to the VM and install Python
- Set up object storage and upload a CSV file
- Copy the file from object to block storage
- Firewall config to starting a web server (Python, Littlest…)

Resource Groups



Create Resource Group ‘NAIRR_RG’



MSE544 Remark: More Thorough Documentation

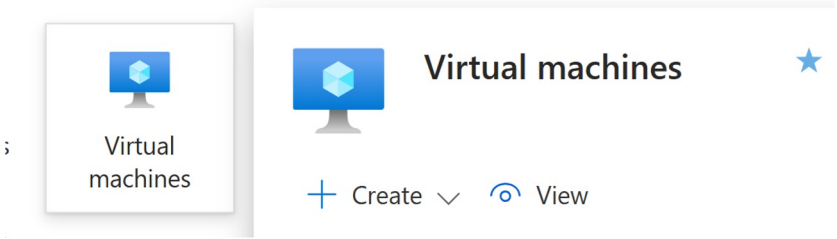
At UW: CloudBank team members teach a portion of a data science course

Focus: building 4 types of cloud compute infrastructure

See https://cloudbank-project.github.io/az-serverless-tutorial/workstation/

We proceed here in less detailed fashion…

Start a Virtual Machine (VM)



Create a virtual machine ...

Help me create a low cost VM | Help me create a VM optim

Project details

Select the subscription to manage deployed resources and costs. Use resou your resources.

Subscription * ⓘ

Resource group * ⓘ NAIRR_RG

Create new

Instance details

Virtual machine name * ⓘ vm1-jupyter

Region * ⓘ (US) West US 2

Create a virtual machine ...

Help me create a low cost VM | Help me create a VM optimized for high availability | Help me choose the right VM

Security type ⓘ
Trusted launch virtual machines
Configure security features

Image * ⓘ
Ubuntu Server 24.04 LTS - x64 Gen2
See all images | Configure VM generation

VM architecture ⓘ
Arm64
x64

Run with Azure Spot discount ⓘ

Size * ⓘ
Standard_D4s_v3 - 4 vcpus, 16 GiB memory ($147.60/month)
See all sizes

Enable Hibernation ⓘ

Administrator account

Authentication type

- (•) SSH public key
- ( ) Password

Azure now automatically generates an SSH key pair for you and allows you to store it for future use. It is a fast, simple, and secure way to connect to your virtual machine.

Username * azureuser

SSH public key source: Generate new key pair

SSH Key Type

- (•) RSA SSH Format
- ( ) Ed25519 SSH Format

Ed25519 provides a fixed security level of no more than 128 bits for 256-bit key, while RSA could offer better security with keys longer than 3072 bits.

Key pair name * vm1-jupyter_key

Inbound port rules

Select which virtual machine network ports are accessible from the public internet. You can specify more limited or granular network access on the Networking tab.

Public inbound ports *

- ( ) None
- (•) Allow selected ports

Select inbound ports * SSH (22)

Create a virtual machine ...

Help me create a low cost VM | Help me create a VM optimized for high availability | Help me choose the ri

Basics | **Disks** | Networking | Management | Monitoring | Advanced | Tags | Review + create

Azure VMs have one operating system disk and a temporary disk for short-term storage. You can attach additional data disks. The size of the VM determines the type of storage you can use and the number of data disks allowed. Learn more

VM disk encryption

Azure disk storage encryption automatically encrypts your data stored on Azure managed disks (OS and data disks) at rest by default when persisting it to the cloud.

Encryption at host ⓘ ☐

ⓘ Encryption at host is not registered for the selected subscription. Learn more

OS disk

OS disk size ⓘ 512 GiB (P20) ⌄

Create a VM: Management tab

Microsoft Entra ID login now uses SSH certificate-based authentication. You will need to use an SSH client that supports OpenSSH certificates. You can use Azure CLI or Cloud Shell from the Azure Portal. Learn more

Auto-shutdown

Enable auto-shutdown

Shutdown time: 5:30:00 PM

Time zone: (UTC-08:00) Pacific Time (US & Canada)

Notification before shutdown

Email *

Create a virtual machine ...

Validation passed

Help me create a low cost VM | Help me create a VM optimized for high availability | Help me choose

Basics | Disks | Networking | Management | Monitoring | Advanced | Tags | **Review + create**

Price

1 X Standard D4s v3
by Microsoft
Terms of use | Privacy policy

Subscription credits apply

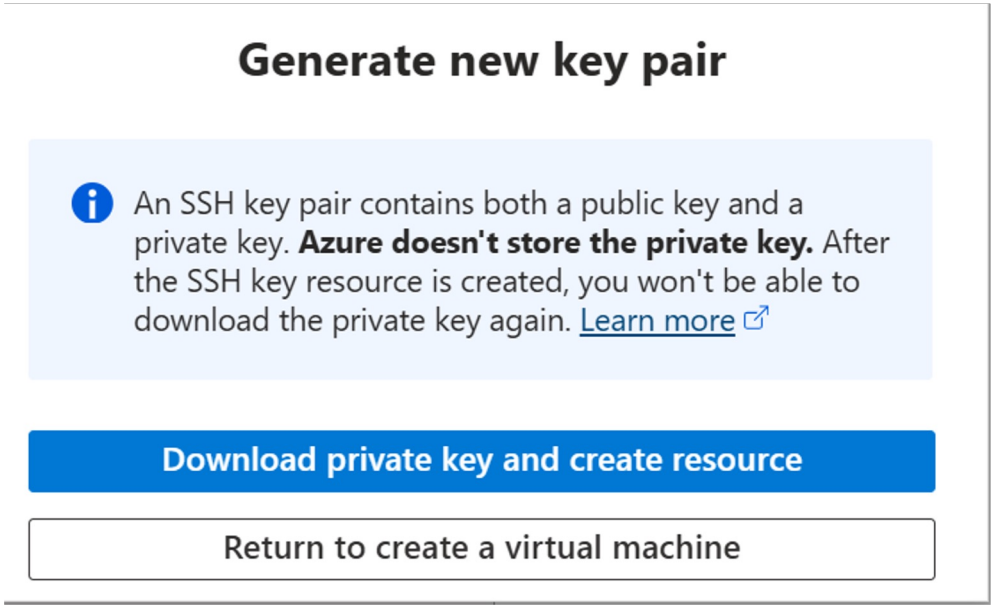
0.2022 USD/hr

Pricing for other VM sizes

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed

Download the VM access key pair



Success: We have appropriated a VM on Azure

Your deployment is complete

Deployment name: CreateVm-canonical.ubuntu-24_04-lts-server-2...
Subscription: [redacted]
Resource group: NAIRR_RG

Start time: 12/11/2024, 11:58:50 AM
Correlation ID: [redacted]

Deployment details

Next steps

Setup auto-shutdown Recommended

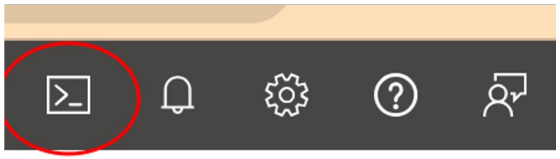
Monitor VM health, performance and network dependencies Recommended

Run a script inside the virtual machine Recommended

Go to resource | Create another VM

Operating system : Linux (ubuntu 24.04)
Size : Standard D4s v3 (4 vcpus, 16 GiB memory)
Public IP address : [redacted]
Virtual network/subnet : vm1-jupyter-vnet/default
DNS name : Not configured

VM Login 1: Portal > Cloud shell



```
Welcome to Azure Cloud Shell

Type "az" to use Azure CLI
Type "help" to learn about Cloud Shell

Your Cloud Shell session will be ephemeral so no files or system changes will persist beyond your current session.
[redacted] [ ~ ]$
```

Upload keypair file from before (**`vm1-jupyter_key.pem`**), then:

```
rob [ ~ ]$ chmod 400 vm1-jupyter_key.pem
rob [ ~ ]$ ssh -i ./vm1-jupyter_key.pem azureuser@[redacted]
Welcome to Ubuntu 24.04.1 LTS (GNU/Linux 6.8.0-1017-azure x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/pro
```

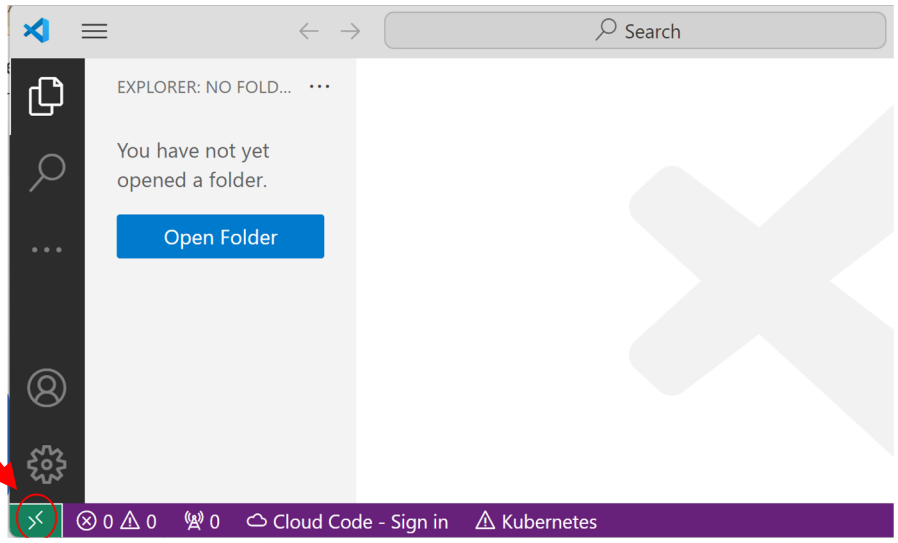
…from the VM command line: `python3` exists already…

```
azureuser@vm1-jupyter:~$ which python3
/usr/bin/python3
azureuser@vm1-jupyter:~$ python3
Python 3.12.3 (main, Sep 11 2024, 14:17:37) [GCC 13.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import this
The Zen of Python, by Tim Peters

Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.
In the face of ambiguity, refuse the temptation to guess.
There should be one-- and preferably only one --obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.
```

VM Login 2: Via VSCode IDE

- **Note: This method works for any VM / any cloud. Not Azure-specific**

- **Visual Studio Code** is a feature-rich integrated development environment (IDE)
- We use it here simply to connect to the Azure VM we just Created
- More detail: https://cloudbank-project.github.io/az-serverless-tutorial/workstation/
- Lower-left icon launches another window
- Use this to ssh to a VM

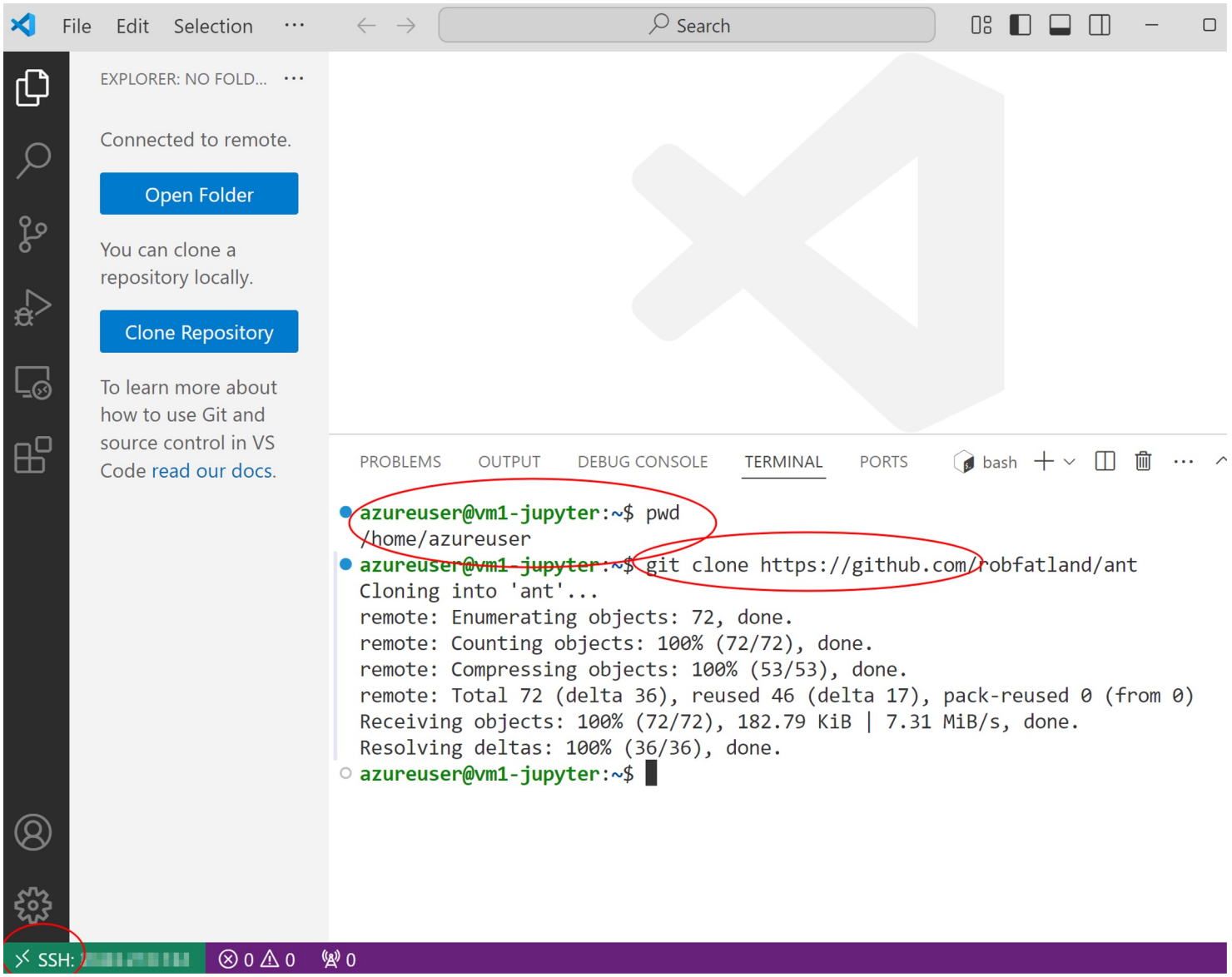


Result

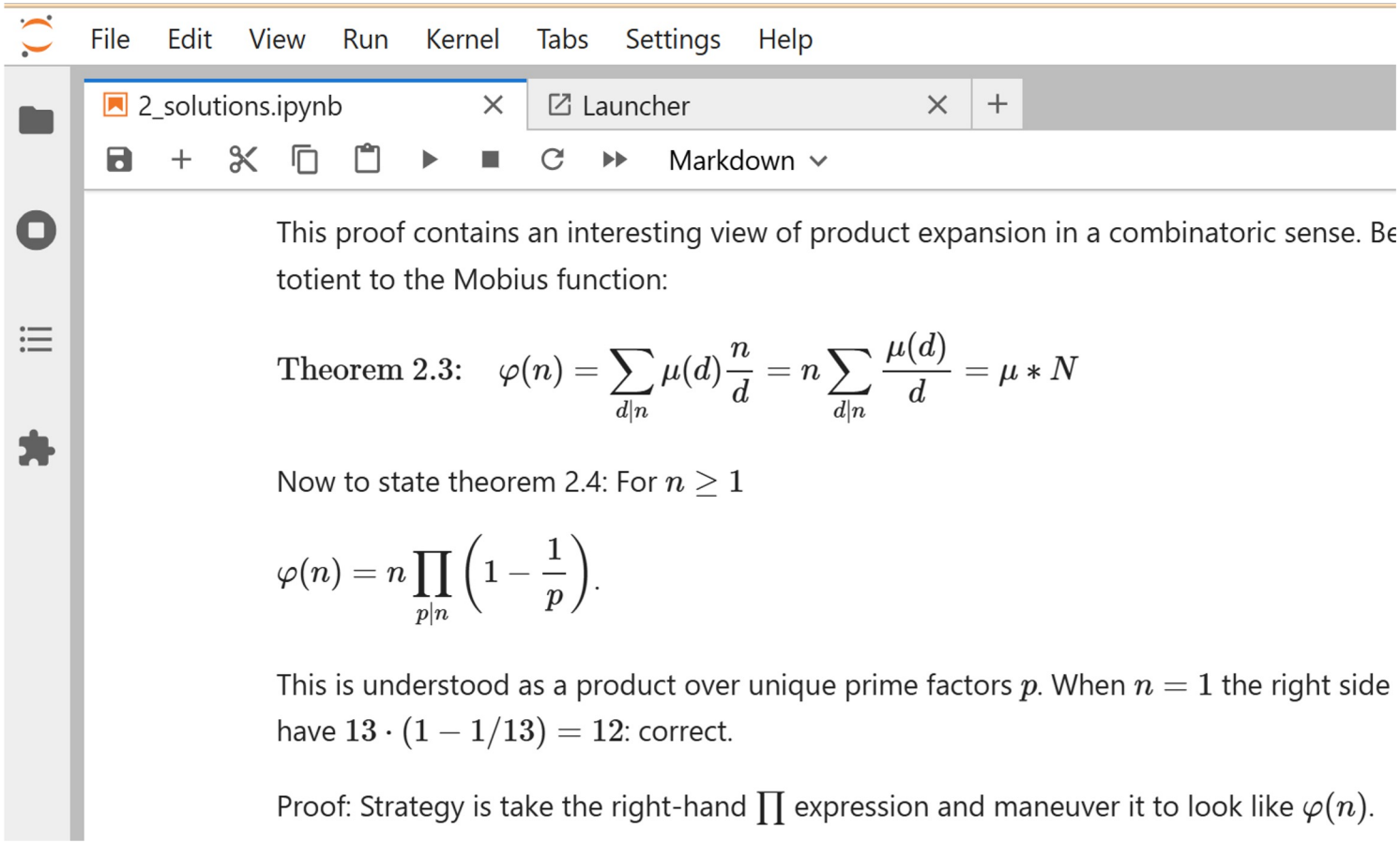
This screenshot shows a user logged in to an Azure VM as created above.

The `git clone` command creates a local clone of a GitHub repository. This particular repo contains some Jupyter notebooks.

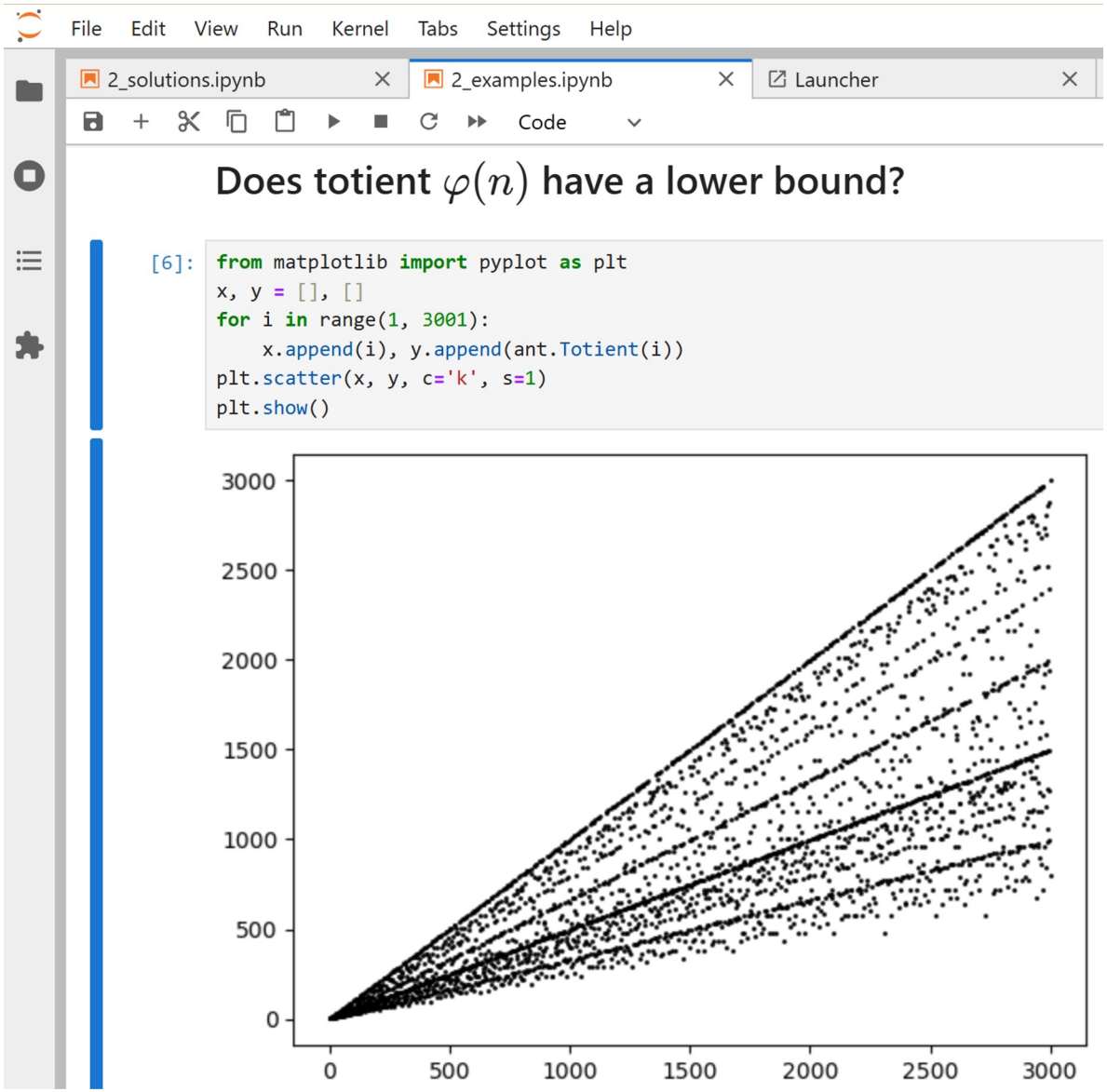
With a Jupyter notebook server installed on the PC and another installed on the VM: We can use an `ssh` connection called a tunnel to work in the VM Jupyter environment from our PC.



Jupyter notebook excerpt 1



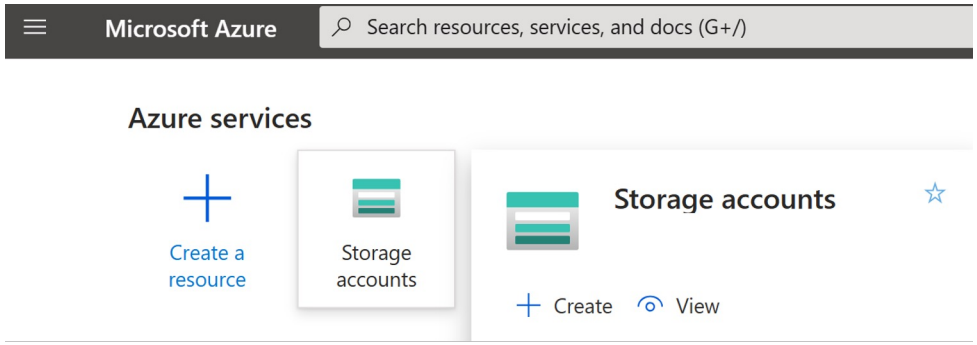
Jupyter excerpt 2



install `miniconda` on the Azure VM

Visit http://anaconda.com and follow instructions for installing miniconda on Linux

```
azureuser@vm1-jupyter:~$ mkdir -p ~/miniconda3
wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh -O ~/miniconda3/miniconda.sh
bash ~/miniconda3/miniconda.sh -b -u -p ~/miniconda3
rm ~/miniconda3/miniconda.sh
--2024-12-12 05:49:26--  https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
Resolving repo.anaconda.com (repo.anaconda.com)... 104.16.191.158, 104.16.32.241, 2606:4700::6810:20f1, ...
Connecting to repo.anaconda.com (repo.anaconda.com)|104.16.191.158|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 148337011 (141M) [application/octet-stream]
Saving to: '/home/azureuser/miniconda3/miniconda.sh'

/home/azureuser/miniconda3/minicond 100%[=====================================================================>] 141.46M   226MB/s    in 0.6s

2024-12-12 05:49:27 (226 MB/s) - '/home/azureuser/miniconda3/miniconda.sh' saved [148337011/148337011]

PREFIX=/home/azureuser/miniconda3
Unpacking payload ...

Installing base environment...

Preparing transaction: ...working... done
Executing transaction: ...working... done
installation finished.
```



Home >

Create a storage account ...

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription * [redacted]

Resource group * NAIRR_RG

Create new

Instance details

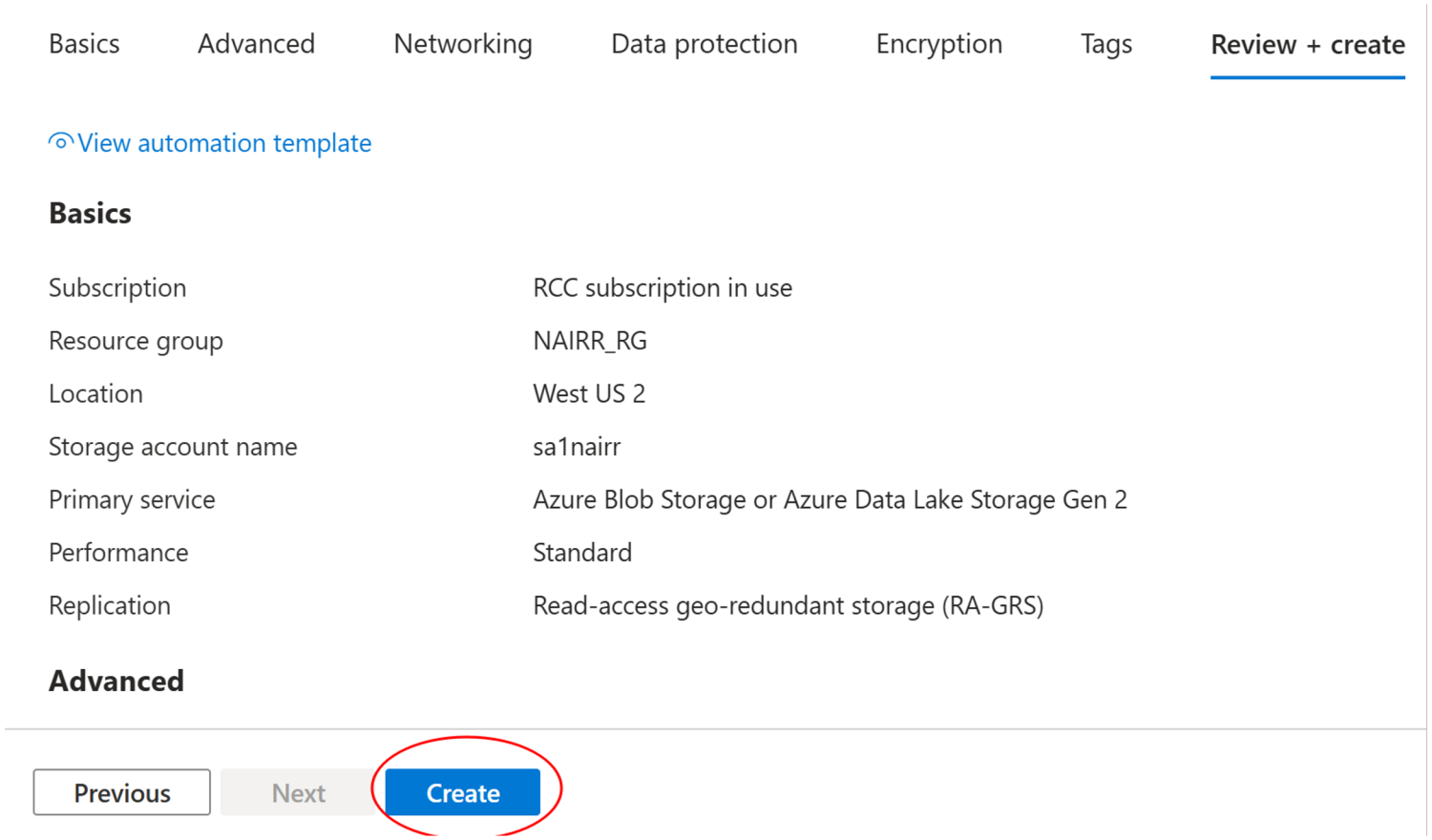
Storage account name * sa1nairr

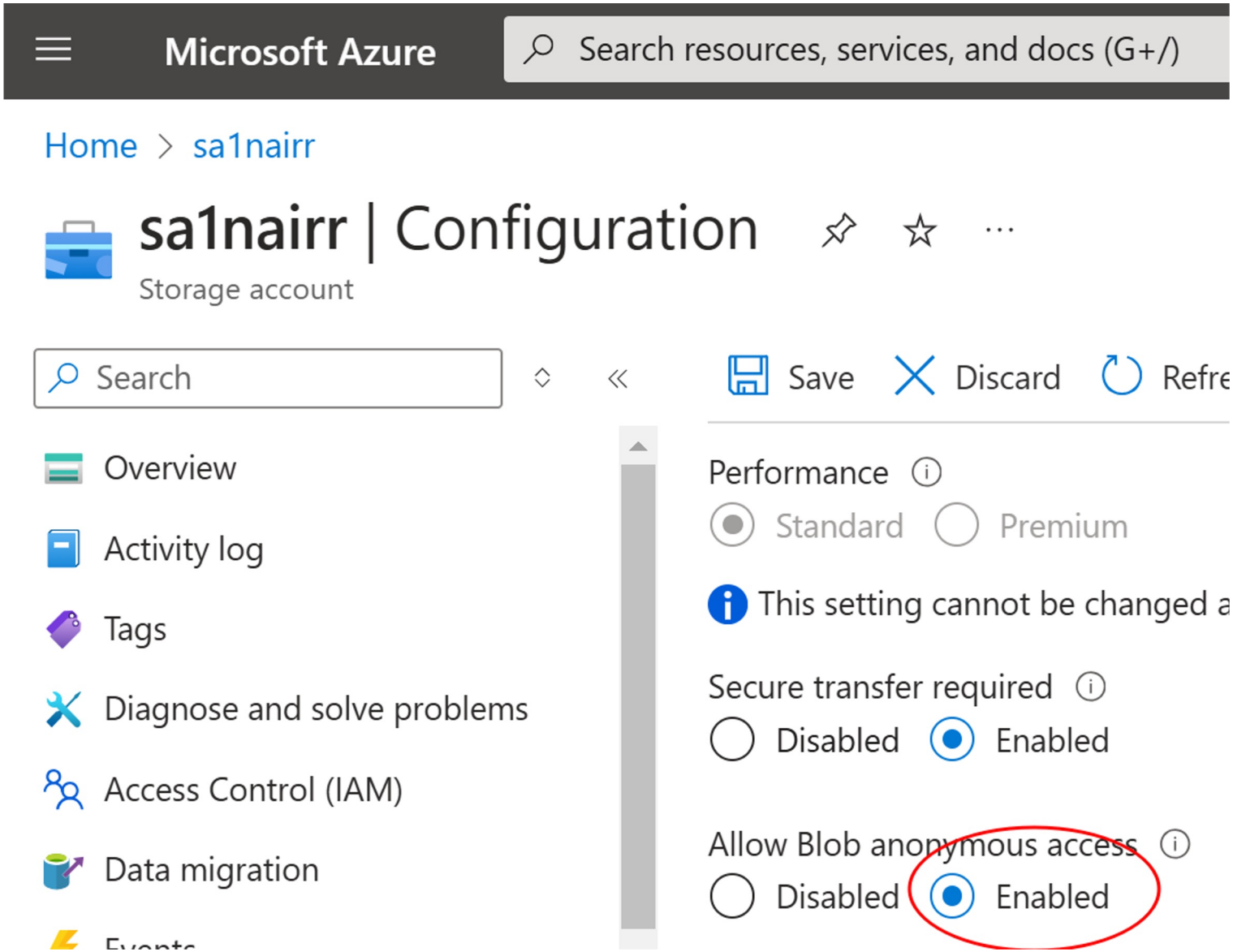
Region * (US) West US 2

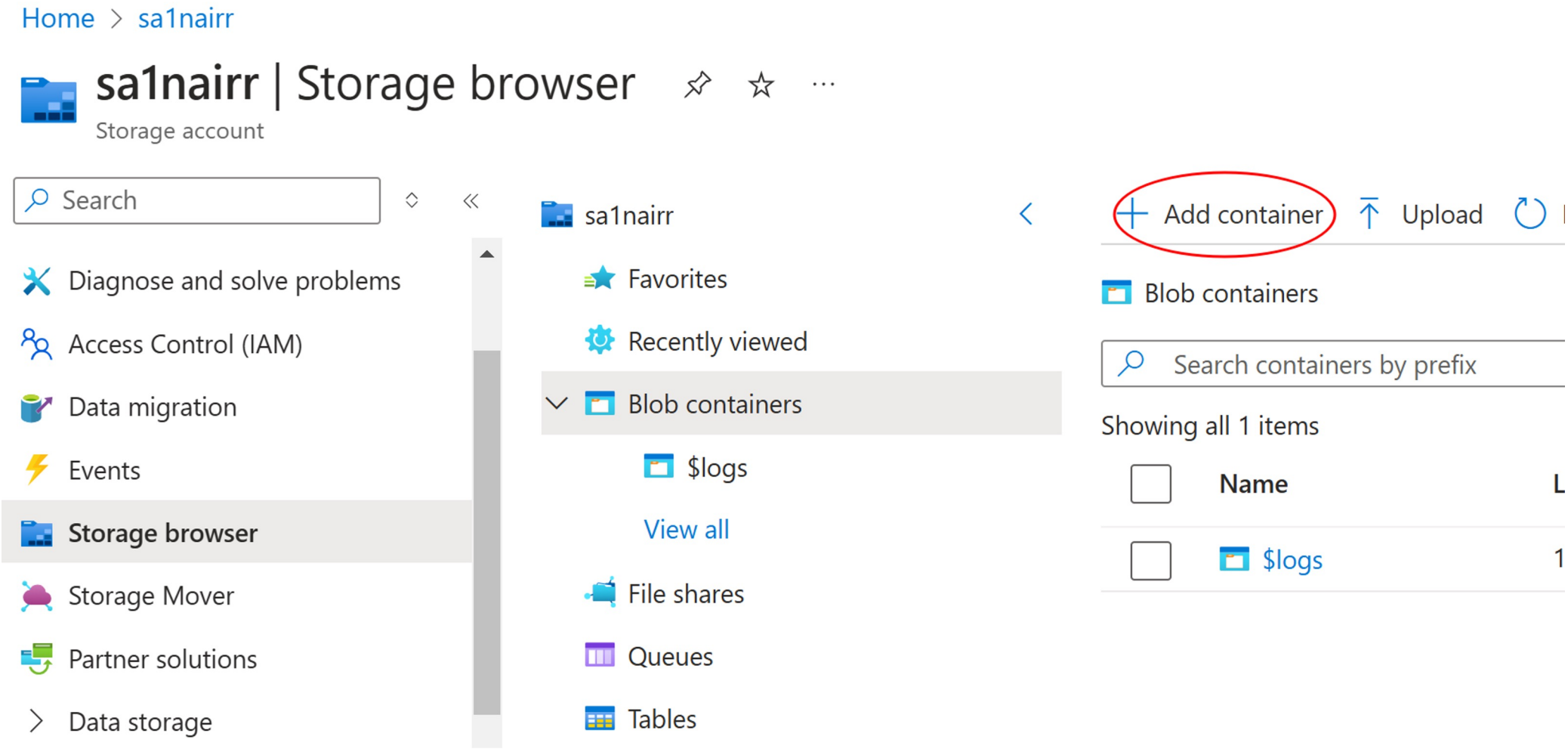
Deploy to an Azure Extended Zone

Primary service Azure Blob Storage or Azure Data Lake Storage Gen 2

Creating the storage account




Microsoft Azure
Search resources, services, and docs (G+/)
Home > sa1nairr
sa1nairr | Configuration
Storage account
Search
Save
Discard
Refre
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Performance
Standard
Premium
This setting cannot be changed a
Secure transfer required
Disabled
Enabled
Allow Blob anonymous access
Disabled
Enabled


Home > sa1nairr
sa1nairr | Storage browser
Storage account
Search
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser
Storage Mover
Partner solutions
Data storage
sa1nairr
Favorites
Recently viewed
Blob containers
$logs
View all
File shares
Queues
Tables
Add container
Upload
Blob containers
Search containers by prefix
Showing all 1 items
Name
$logs

New container

Name *

nairr-blob

Anonymous access level

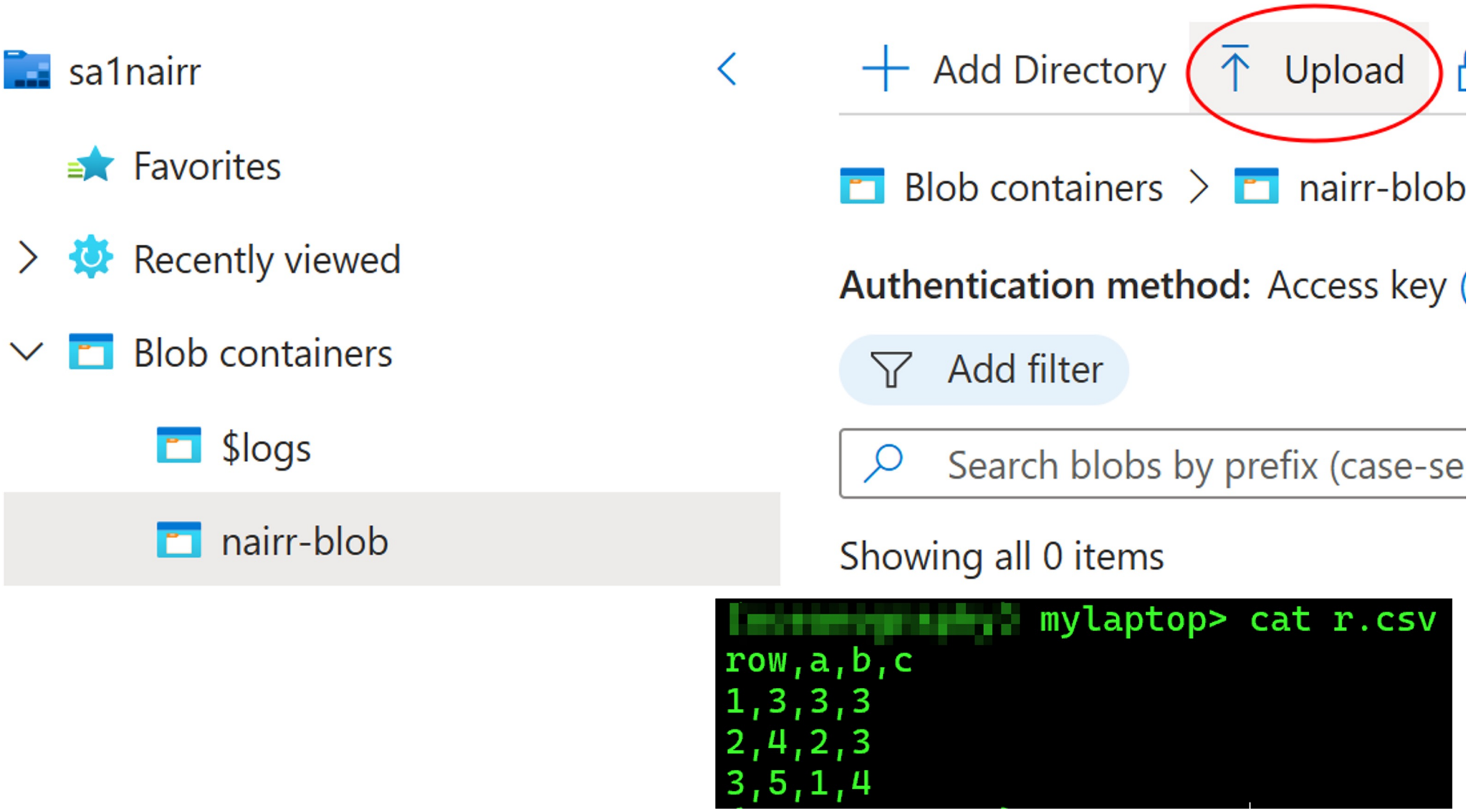
Container (anonymous read access for containers and blobs)

All container and blob data can be read by anonymous request. Clients can enumerate blobs within the container by anonymous request, but cannot enumerate containers within the storage account.

Advanced

Create

Give feedback


sa1nairr
Favorites
Recently viewed
Blob containers
$logs
nairr-blob
Add Directory
Upload
Blob containers
nairr-blob
Authentication method: Access key
Add filter
Search blobs by prefix (case-se
Showing all 0 items
mylaptop> cat r.csv
row,a,b,c
1,3,3,3
2,4,2,3
3,5,1,4

Overview

Blob: r.csv

Save | Discard | Download | Refresh | Delete | Change tier

Properties

URL	https://sa1nairr.blob.cor...
LAST MODIFIED	12/11/2024, 11:14:40 PM
CREATION TIME	12/11/2024, 11:14:40 PM
VERSION ID	-
TYPE	Block blob
SIZE	34 B
ACCESS TIER	Hot (Inferred)
ACCESS TIER LAST MODIFIED	N/A
ARCHIVE STATUS	-
REHYDRATE PRIORITY	-
SERVER ENCRYPTED	true

Reading a table directly from blob storage

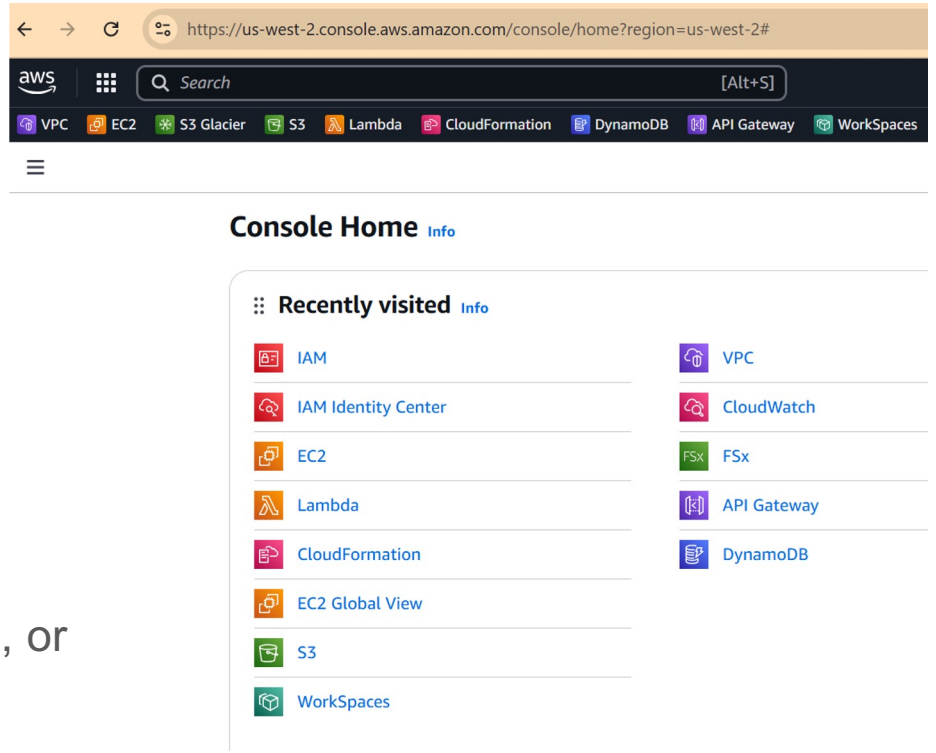
This connects the dots: First we created and configured a Virtual Machine. Second we created a storage account and therein a CSV file residing in blob (object) storage. Third (below) we used three lines of Python on the VM to read that file and confirm its contents.

```
azureuser@vm1-jupyter:~$ python
Python 3.12.7 | packaged by Anaconda, Inc. | (main, Oct  4 2024, 13:27:36) [GCC 11.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import pandas as pd
>>> d=pd.read_csv('https://sa1nairr.blob.core.windows.net/nairr-blob/r.csv')
>>> d
   row  a  b  c
0    1  3  3  3
1    2  4  2  3
2    3  5  1  4
>>>
```

End Azure
Start AWS

Welcome to AWS! Here is the console view*

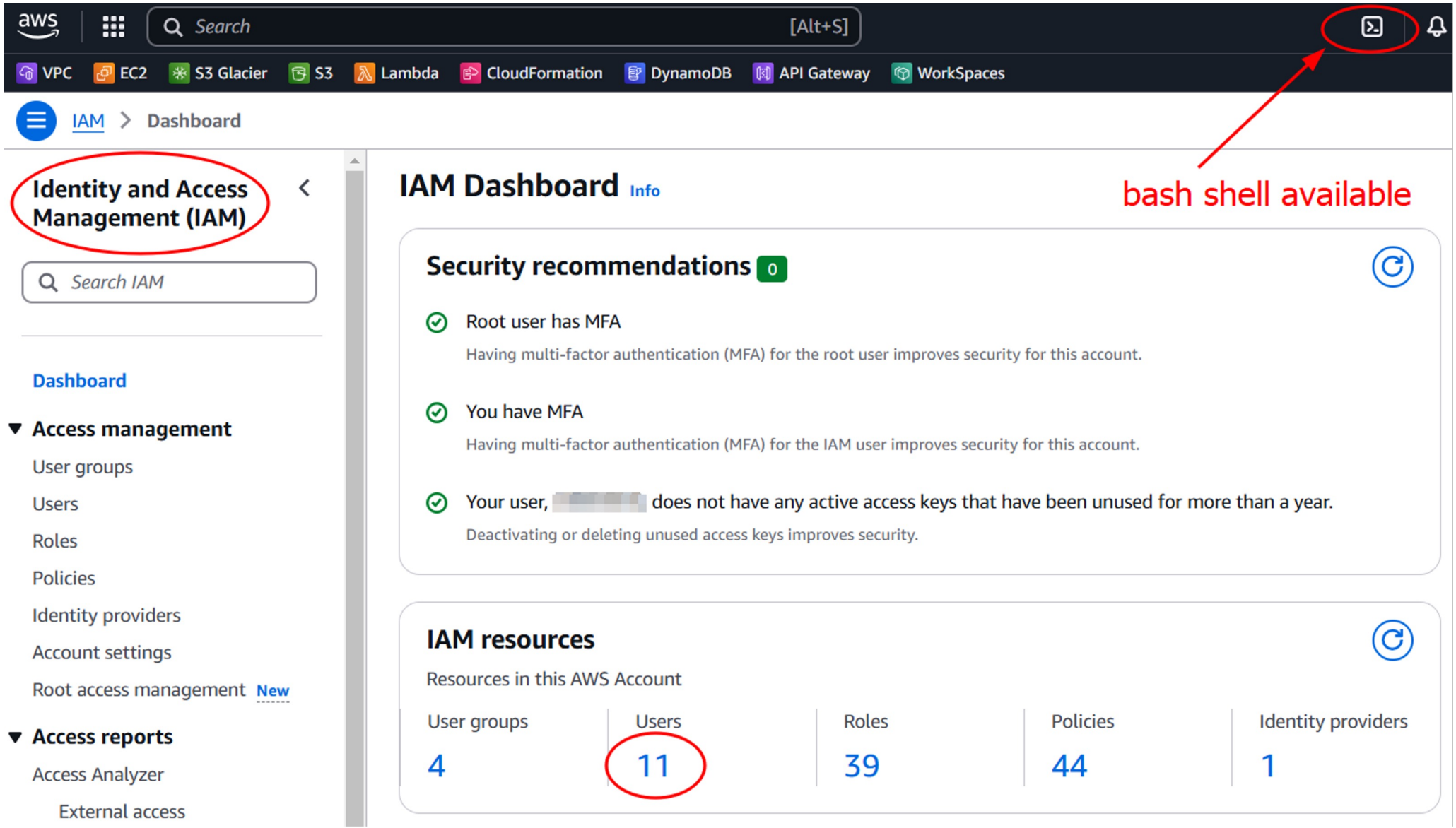
- Services have dedicated landing pages
- These are shown here (below the search bar) as bookmarks
- As with all cloud platforms: Some time is required…
  - …to learn the service vocabulary
  - …to focus on what you need to use



* Talk to the AWS cloud via console, command line, or with a programmatic library such as `boto3`

AWS get-started terms

- VPC: ‘Virtual Private Cloud’ (logical collection of related resources)
- EC2: ‘Elastic Cloud Compute’ is AWS’ term for a virtual machine
  - ‘E’ frequently means ‘Elastic/Scalable’
- S3: ‘Simple Storage Service’: Object storage (buckets / blobs)
- AWS Account: Administrative entity for AWS cloud use (12 digit number)
  - Roughly equivalent to an Azure *Subscription* / GCP *Project*
- Spot Market: Preemptible resource pool
  - Cheap Virtual Machines but some probability of being kicked off on short notice (minute)
- EBS: Elastic Block Store: A detachable disk volume
- EFS: Elastic File System: A shareable disk volume equivalent to NFS

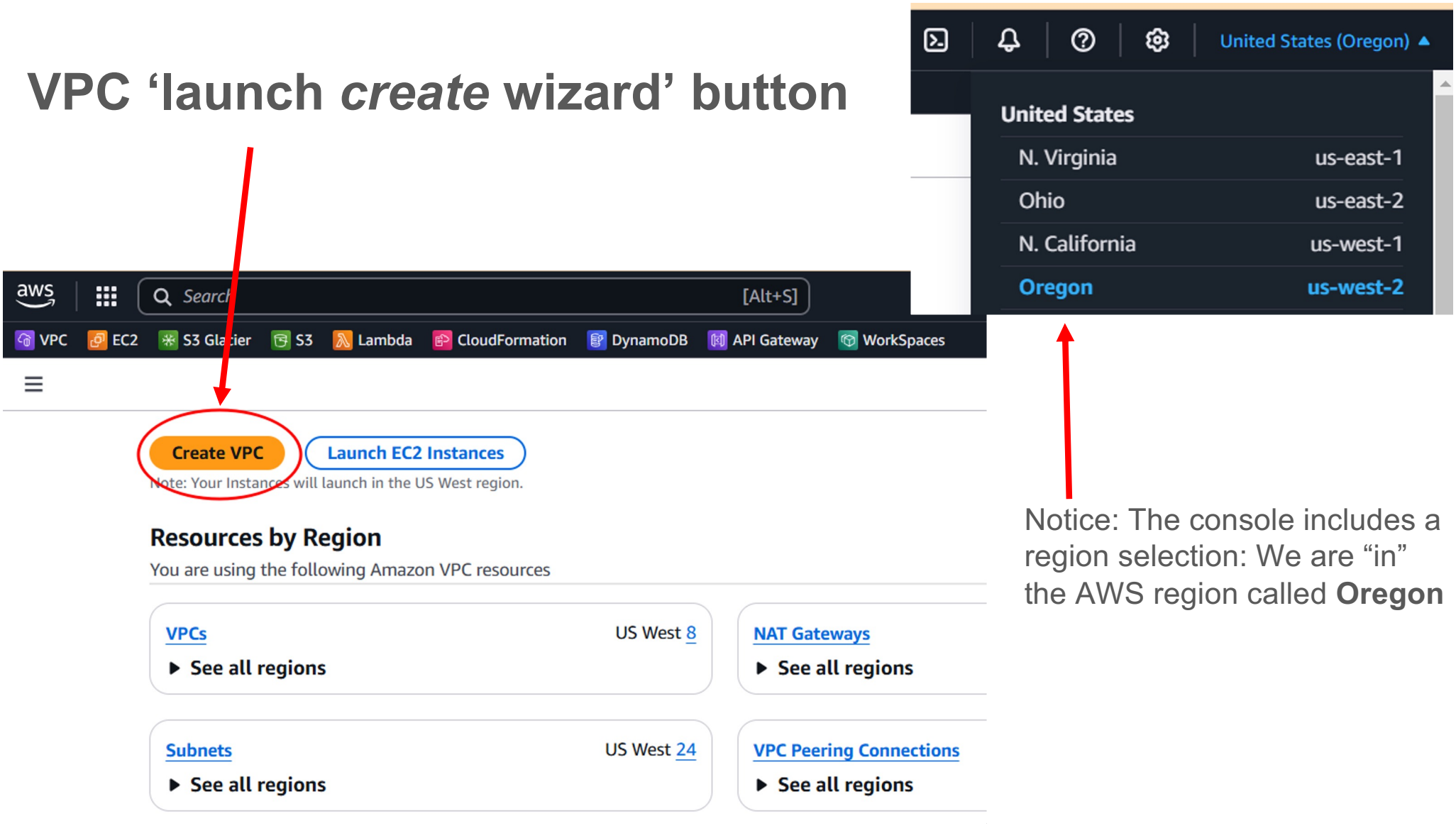

aws
Search
[Alt+S]
VPC
EC2
S3 Glacier
S3
Lambda
CloudFormation
DynamoDB
API Gateway
WorkSpaces
IAM
Dashboard
Identity and Access Management (IAM)
Search IAM
Dashboard
Access management
User groups
Users
Roles
Policies
Identity providers
Account settings
Root access management
New
Access reports
Access Analyzer
External access
IAM Dashboard
Info
bash shell available
Security recommendations
0
Root user has MFA
Having multi-factor authentication (MFA) for the root user improves security for this account.
You have MFA
Having multi-factor authentication (MFA) for the IAM user improves security for this account.
Your user, does not have any active access keys that have been unused for more than a year.
Deactivating or deleting unused access keys improves security.
IAM resources
Resources in this AWS Account
User groups
4
Users
11
Roles
39
Policies
44
Identity providers
1

Amazon Web Services Fly-over

Plan

- Log in to the console (done on previous slide!)
- Create a Virtual Private Cloud or VPC
- Get a VM from the Spot market with an expanded root disk
  - Log in to the VM, install Python
- Set up S3 object storage and upload a CSV file
- Copy this file from object to block storage, load it into a Python program

VPC ‘launch *create* wizard’ button



Notice: The console includes a region selection: We are “in” the AWS region called **Oregon**

Create VPC Info

A VPC is an isolated portion of the AWS Cloud populated by AWS objects, such as

VPC settings

Resources to create Info

Create only the VPC resource or the VPC and other networking resources.

- ○ VPC only
- ◉ VPC and more

Name tag auto-generation Info

Enter a value for the Name tag. This value will be used to auto-generate Name tags for all resources in the VPC.

- ☑ Auto-generate

CloudBankExampleAWS

IPv4 CIDR block Info

Determine the starting IP and the size of your VPC using CIDR notation.

10.0.0.0/16 65,536 IPs

CIDR block size must be between /16 and /28.

IPv6 CIDR block Info

- ◉ No IPv6 CIDR block
- ○ Amazon-provided IPv6 CIDR block

Tenancy Info

Default ▼

Create VPC → View VPC

vpc-08[redacted] / CloudBankExampleAWS-vpc

Details Info

VPC ID
vpc-08[redacted]

State
Available

Block Public Access
Off

DNS resolution
Enabled

Tenancy
default

DHCP option set
dopt-[redacted] / Oregon DHCP options

Main network ACL
acl-0d[redacted]

Default VPC
No

IPv4 CIDR
10.0.0.0/16

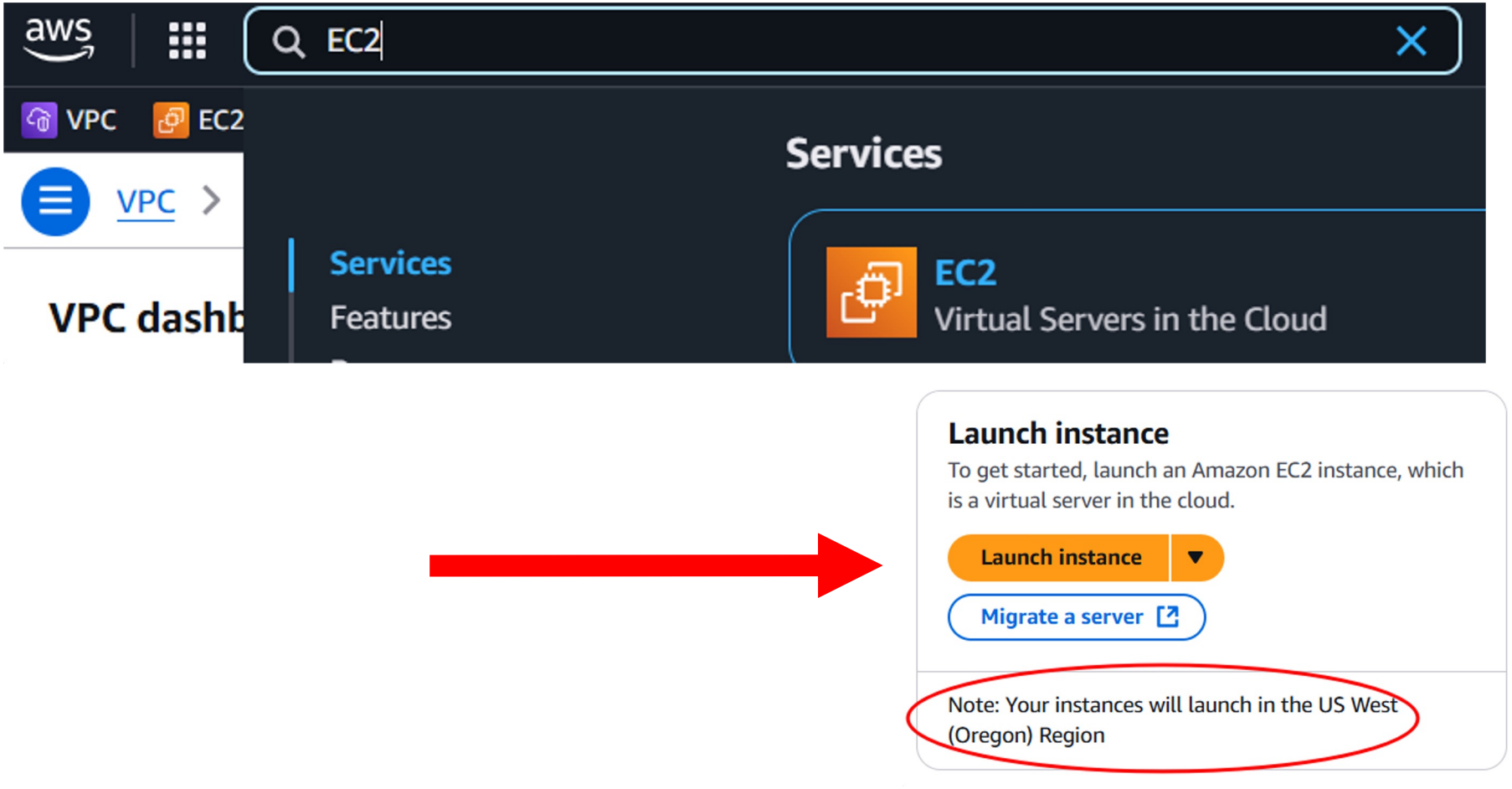
IPv6 CIDR (Network border group)
–

Network Address Usage metrics
Disabled

Route 53 Resolver DNS Firewall rule groups
–

Resource map | CIDRs | Flow logs | Tags | Integrations

Find the EC2 (Virtual Machine) service and click to go there



Launch an instance Info

Amazon EC2 allows you to create virtual machines, or instances, that run on the

Name and tags Info

Name

CloudBankExampleAWS_VM

▼ Application and OS Images (Amazon Machine Imag

An AMI is a template that contains the software configuration (operating syst
instance. Search or Browse for AMIs if you don't see what you are looking for

Search our full catalog including 1000s of application and OS images

Recents | My AMIs | Quick Start

Amazon Linux	macOS	Ubuntu	Windows	Red Hat
aws	Mac	ubuntu	Microsoft	Red Hat

Amazon Machine Image (AMI)

Ubuntu Server 24.04 LTS (HVM), SSD Volume Type
ami-05d38da78ce859165 (64-bit (x86)) / ami-0d4eea77bb23270f4 (64-bit (Arm))
Virtualization: hvm ENA enabled: true Root device type: ebs

Description

Ubuntu Server 24.04 LTS (HVM),EBS General Purpose (SSD) Volume Type. Support available from Canonical (http://www.ubuntu.com/cloud/services).

Canonical, Ubuntu, 24.04, amd64 noble image

Architecture	AMI ID	Username	
64-bit (x86)	ami-05d38da78ce859165	ubuntu	Verified provider

▼ Instance type Info | Get advice

Instance type

c5.xlarge
Family: c5 4 vCPU 8 GiB Memory Current generation: true
On-Demand SUSE base pricing: 0.226 USD per Hour On-Demand Linux base pricing: 0.17 USD per Hour
On-Demand Windows base pricing: 0.354 USD per Hour
On-Demand Ubuntu Pro base pricing: 0.177 USD per Hour On-Demand RHEL base pricing: 0.228 USD per Hour

Key pair (login) Info

You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.

Key pair name - *required*

Select

Create new key pair

Network settings Info

VPC - *required* | Info

vpc-08[redacted] (CloudBankExampleAWS-vpc)
10.0.0.0/16

Subnet | Info

subnet-05[redacted]
CloudBankExampleAWS-subnet-public1-us-west-2a
VPC: vpc-08[redacted] Owner: [redacted]
Availability Zone: us-west-2a Zone type: Availability Zone
IP addresses available: 4091 CIDR: 10.0.0.0/20)

Create new subnet

Auto-assign public IP | Info

Enable

Additional charges apply when outside of free tier allowance

Firewall (security groups) | Info

A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.

- (•) Create security group
- ( ) Select existing security group

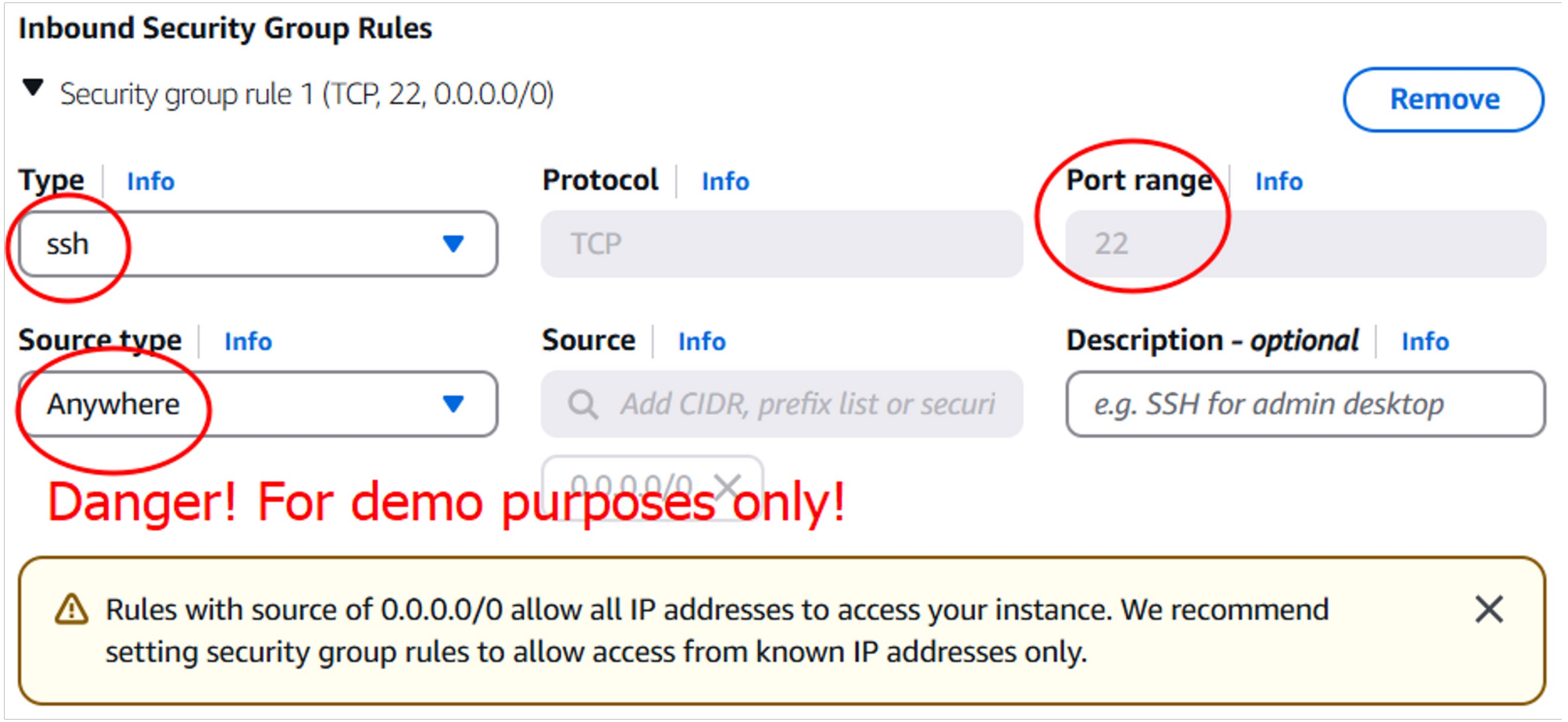
Security group name - *required*

CloudBankExampleAWS_Security_Group

This security group will be added to all network interfaces. The name can't be edited after the security group is created. Max length is 255 characters. Valid characters: a-z, A-Z, 0-9, spaces, and ._-:/()#,@[]+=&;{}!$*

Description - *required* | Info

created 2025-01-06T00:49:45.049Z


Inbound Security Group Rules
Security group rule 1 (TCP, 22, 0.0.0.0/0)
Remove
Type Info
ssh
Protocol Info
TCP
Port range Info
22
Source type Info
Anywhere
Source Info
Add CIDR, prefix list or securi
0.0.0.0/0
Description - optional Info
e.g. SSH for admin desktop
Danger! For demo purposes only!
Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

Configure storage Info

1x 32 GiB gp3 Root volume 3000 IOPS (Not encrypted)

Advanced details Info

Purchasing option Info

- None
- Capacity Blocks
  Launch instances for your active capacity blocks
- Spot instances
  Request Spot Instances at the Spot price, capped at the On-Demand price

Launch instance

Preview code

Console-to-Code

Console-to-Code > Preview code

Preview code

Preview code generates CLI commands for your current in-progress action.

Close preview | Start recording

EC2 / Launch Instances

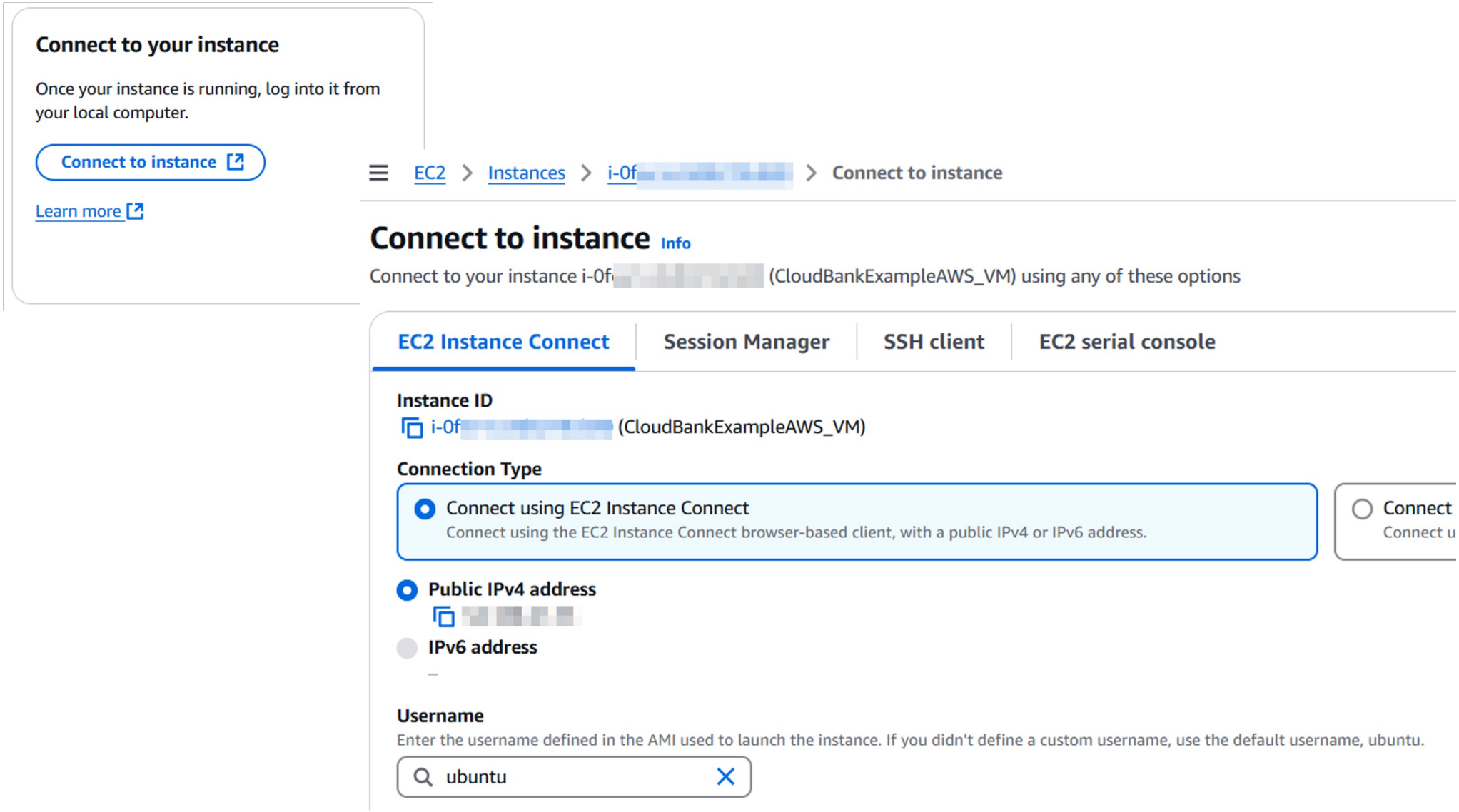
Last updated Sun Jan 05 2025 17:23:53 GMT-0800 (Pacific Standard Time)

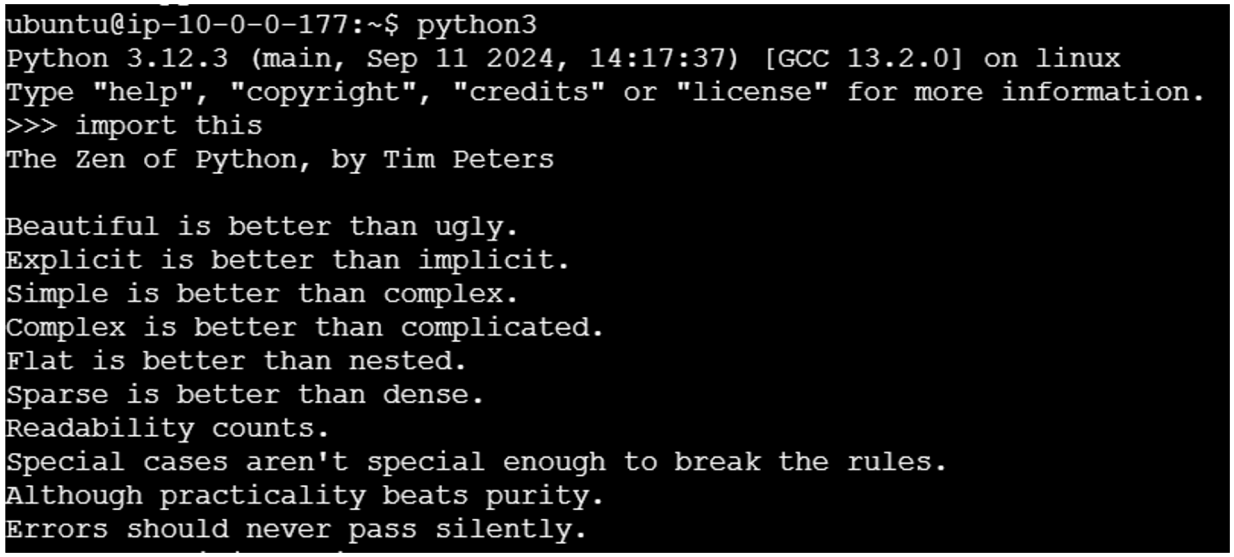
CreateSecurityGroup

```
aws ec2 create-security-group --group-name "CloudBankExampleAWS_Security_Group" --description "created 2025-01-06T00:49:45.049Z" --vpc-id "vpc-08[redacted]'
```

AuthorizeSecurityGroupIngress

```
aws ec2 authorize-security-group-ingress --group-id "sg-preview-1" --ip-permissions '{"IpProtocol":"tcp","FromPort":22,"ToPort":22,"IpRanges":[{"CidrIp":"0.0.0.0/0"}]}'
```

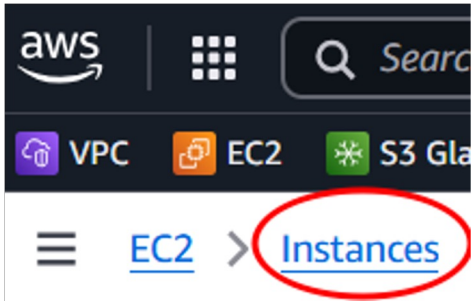

Connect to your instance
Once your instance is running, log into it from your local computer.
Connect to instance
Learn more
EC2 > Instances > i-0f > Connect to instance
Connect to instance Info
Connect to your instance i-0f (CloudBankExampleAWS_VM) using any of these options
EC2 Instance Connect
Session Manager
SSH client
EC2 serial console
Instance ID
i-0f (CloudBankExampleAWS_VM)
Connection Type
Connect using EC2 Instance Connect
Connect using the EC2 Instance Connect browser-based client, with a public IPv4 or IPv6 address.
Connect
Public IPv4 address
IPv6 address
Username
Enter the username defined in the AMI used to launch the instance. If you didn't define a custom username, use the default username, ubuntu.
ubuntu

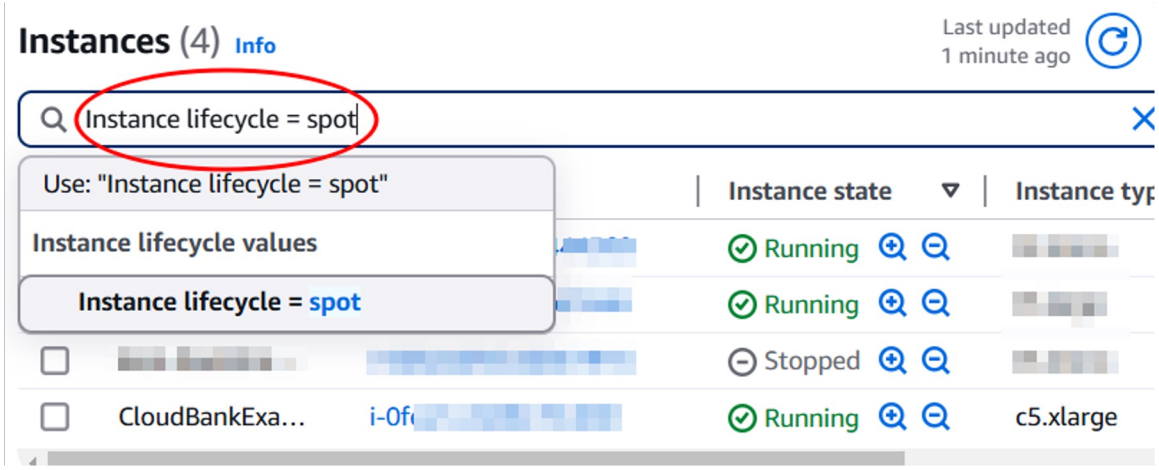


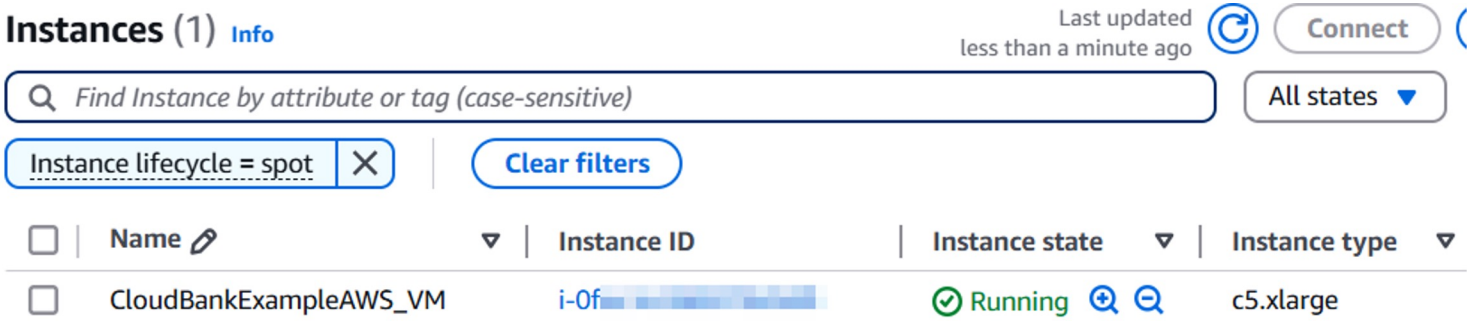
Here we are logged in and running Python on our AWS test EC2 instance (VM)

- On-demand would run us about $0.18 per hour
- We stipulated: Configure a larger root disk of 32GB (> 8GB default)
  - This will add $0.003 per hour
- We stipulated: Use a VM from the preemptible Spot Market
  - This should drop our price to less than $0.09 per hour

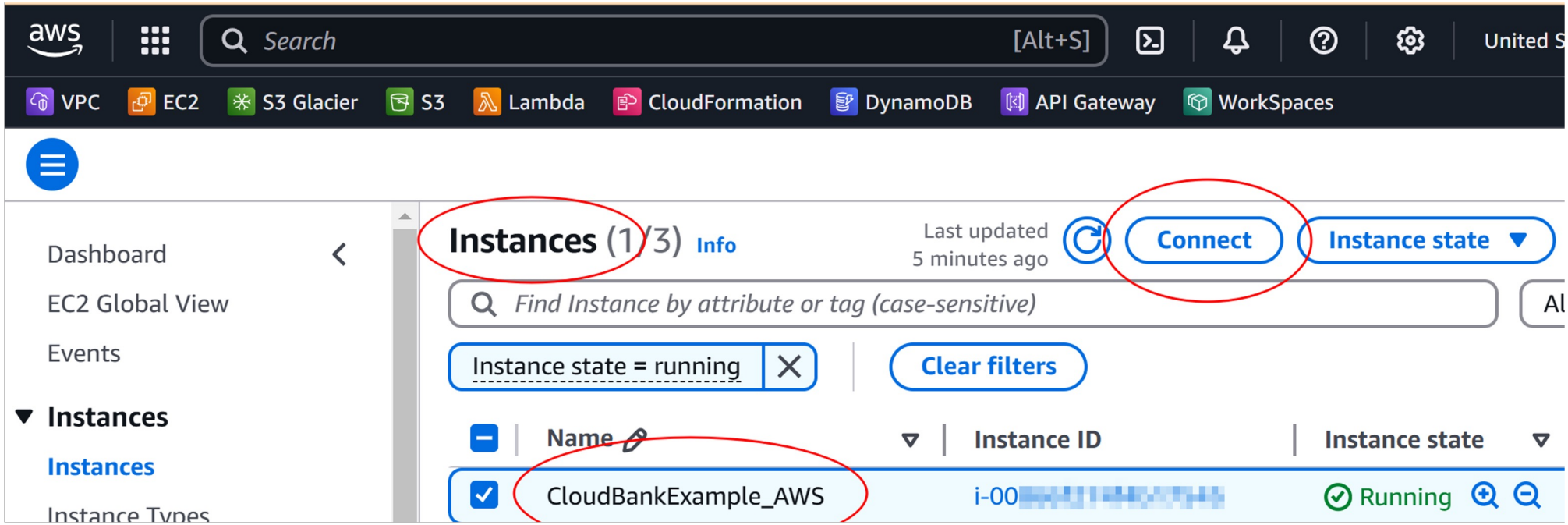
Verify this EC2 instance is in fact from the AWS Spot Market…



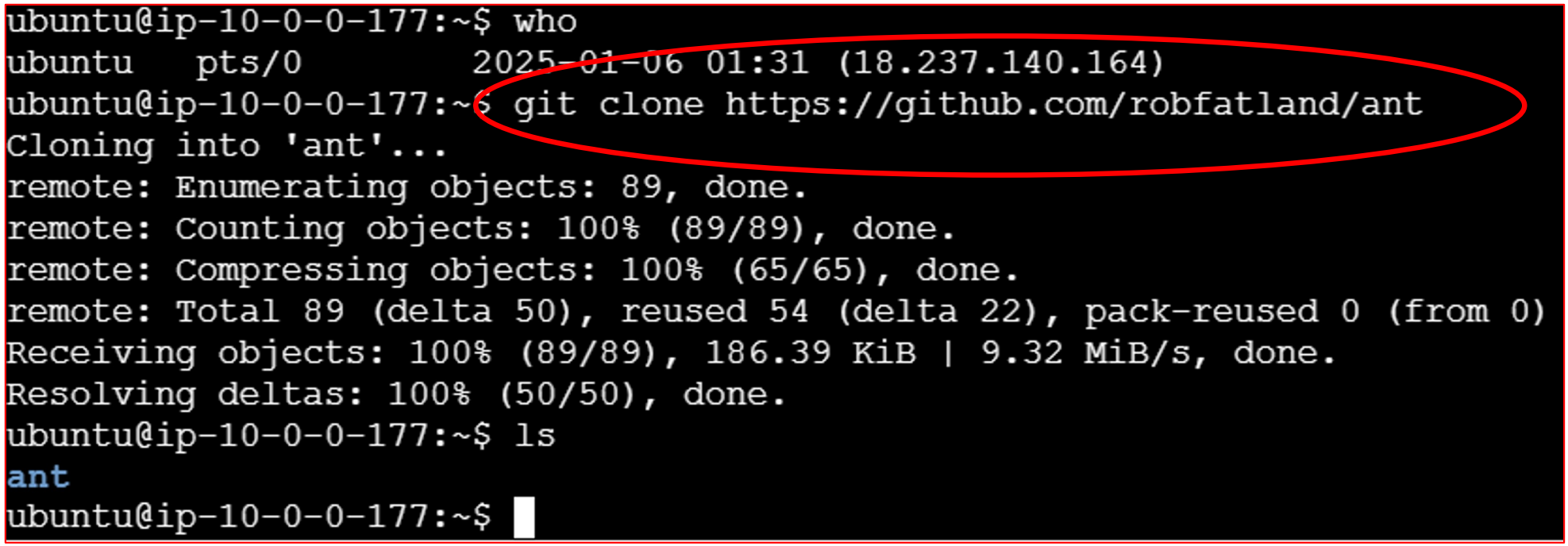




Connect to VM ‘in console’

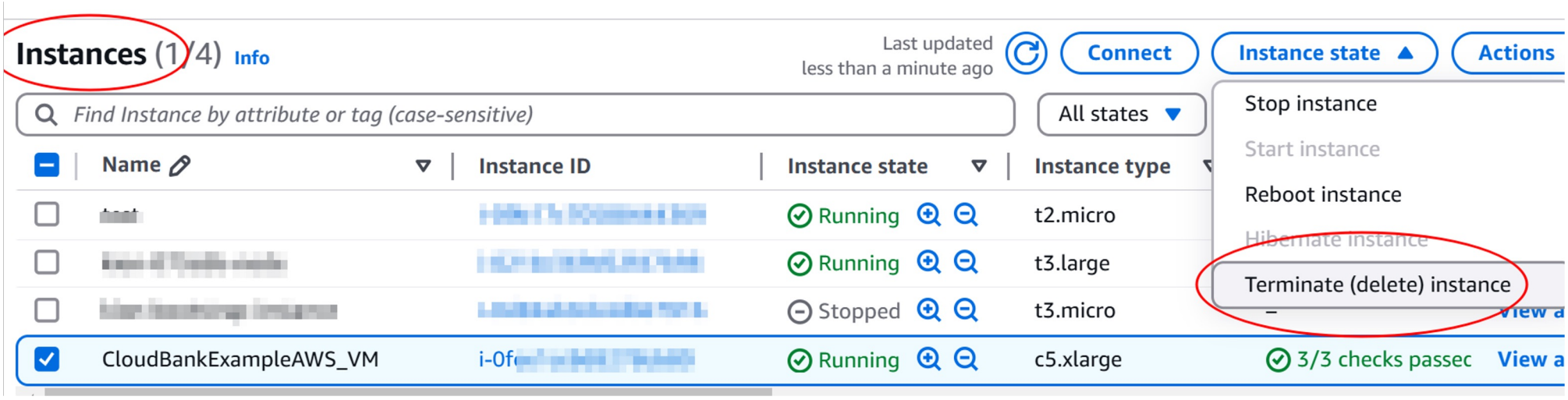


Load in (clone) a Github repo



```
ubuntu@ip-10-0-0-177:~$ pwd
/home/ubuntu
ubuntu@ip-10-0-0-177:~$ mkdir -p ~/miniconda3
ubuntu@ip-10-0-0-177:~$ wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh -O
--2025-01-06 02:40:14--  https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
Resolving repo.anaconda.com (repo.anaconda.com)... 104.16.191.158, 104.16.32.241, 2606:4700::6810:20f
Connecting to repo.anaconda.com (repo.anaconda.com)|104.16.191.158|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 147784736 (141M) [application/octet-stream]
Saving to: '/home/ubuntu/miniconda3/miniconda.sh'

/home/ubuntu/miniconda3/miniconda.sh      100%[=================================================

2025-01-06 02:40:15 (160 MB/s) - '/home/ubuntu/miniconda3/miniconda.sh' saved [147784736/147784736]

ubuntu@ip-10-0-0-177:~$ bash ~/miniconda3/miniconda.sh -b -u -p ~/miniconda3
PREFIX=/home/ubuntu/miniconda3
Unpacking payload ...

Installing base environment...

Preparing transaction: ...working... done
Executing transaction: ...working... done
installation finished.
ubuntu@ip-10-0-0-177:~$ rm ~/miniconda3/miniconda.sh
ubuntu@ip-10-0-0-177:~$
```

Sometimes One Must Terminate a VM



From Allocated Spot Market VM to Research Environment

- ***WARNING***: This example: loose connection security :(

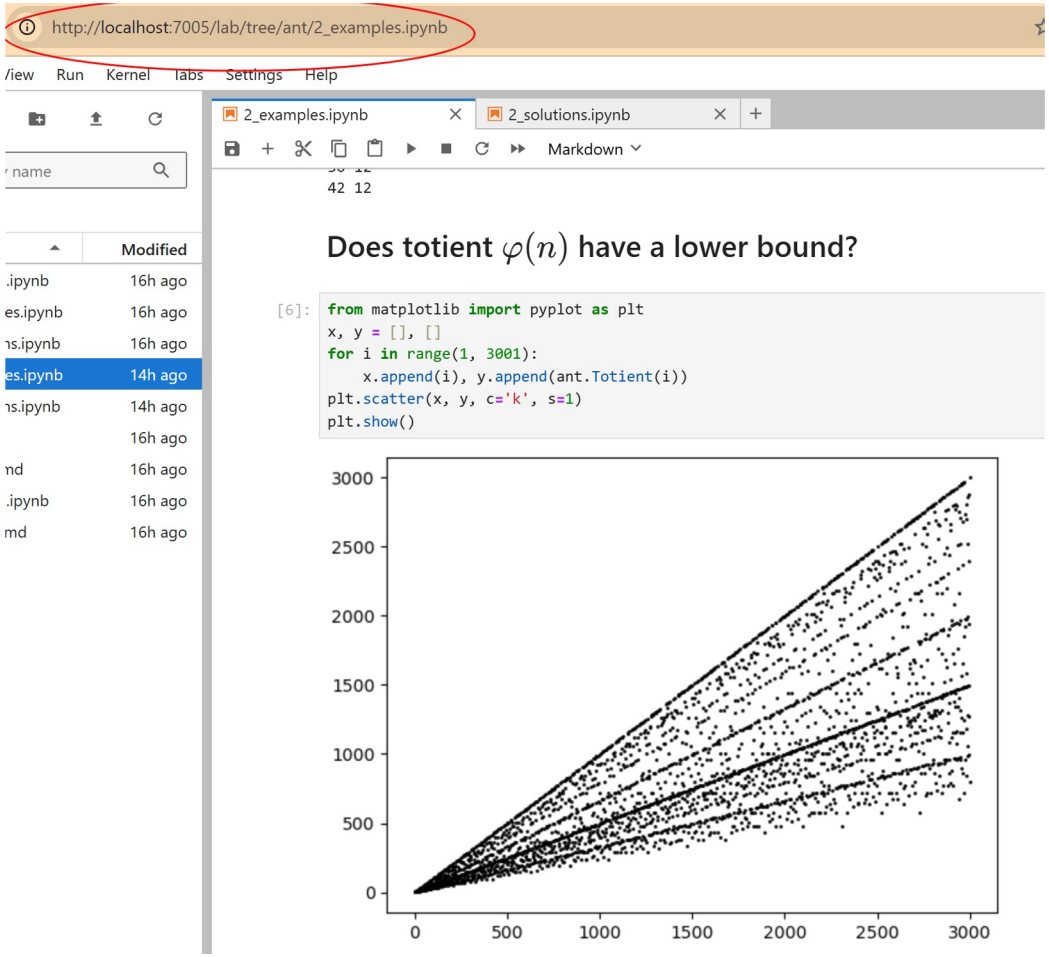
- Login to the VM (VSCode remote window, ssh, Connect from console)
- Install (say) `miniconda`
- Clone a GitHub repo: Project-specific code base, notebooks etc
- Create a `conda` environment, install useful packages like `pandas`
- Start a Jupyter lab server with two important characteristics
  - No browser
  - Port 8889

- On my local computer
  - Use ssh to map my port 7005 to VM port 8889
  - Browser address bar: localhost:7005

Sidebar on the nexus project

- Work in progress (Github > robfatland > nexus > bash)
- Objective: Mediate the gap between research and cyberinfrastructure

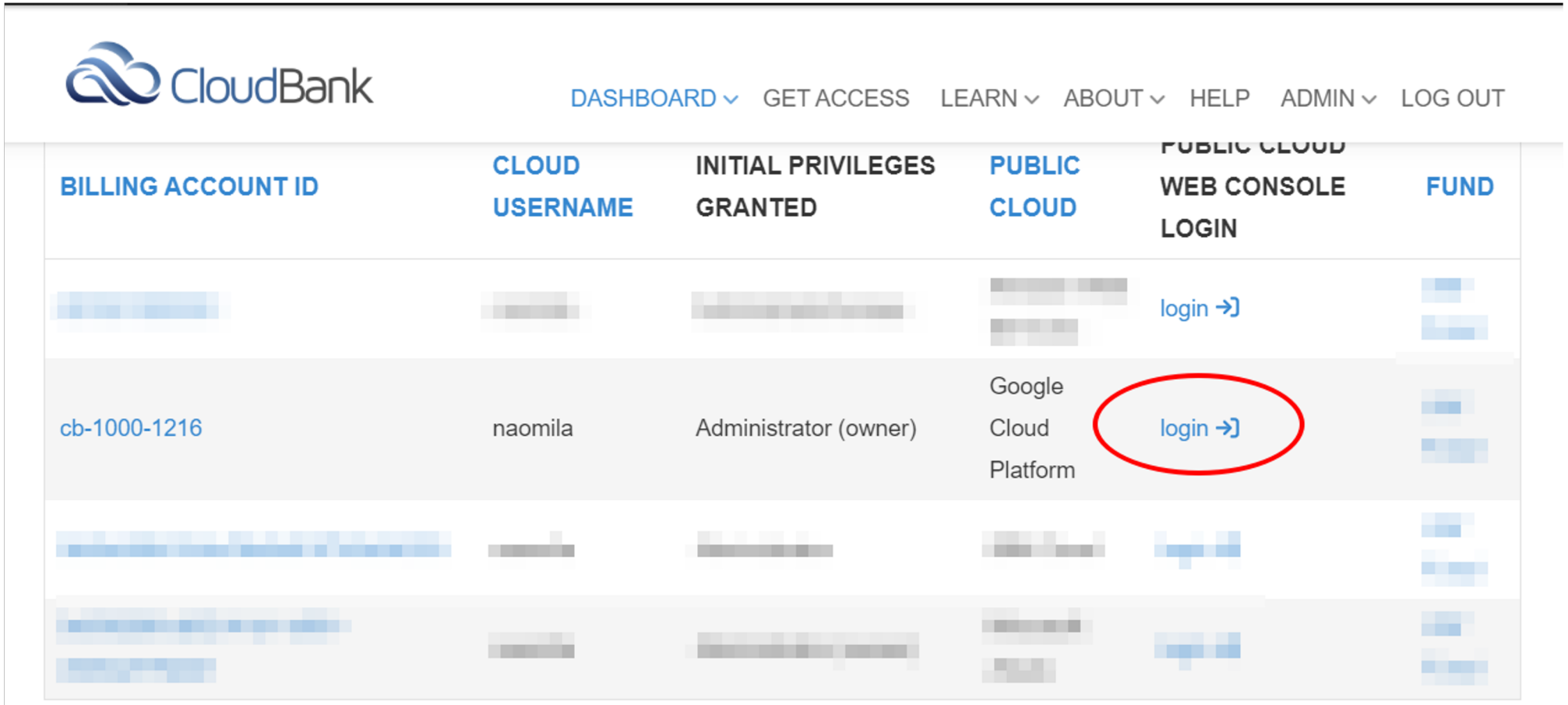
Jupyter lab running on the AWS Spot Market



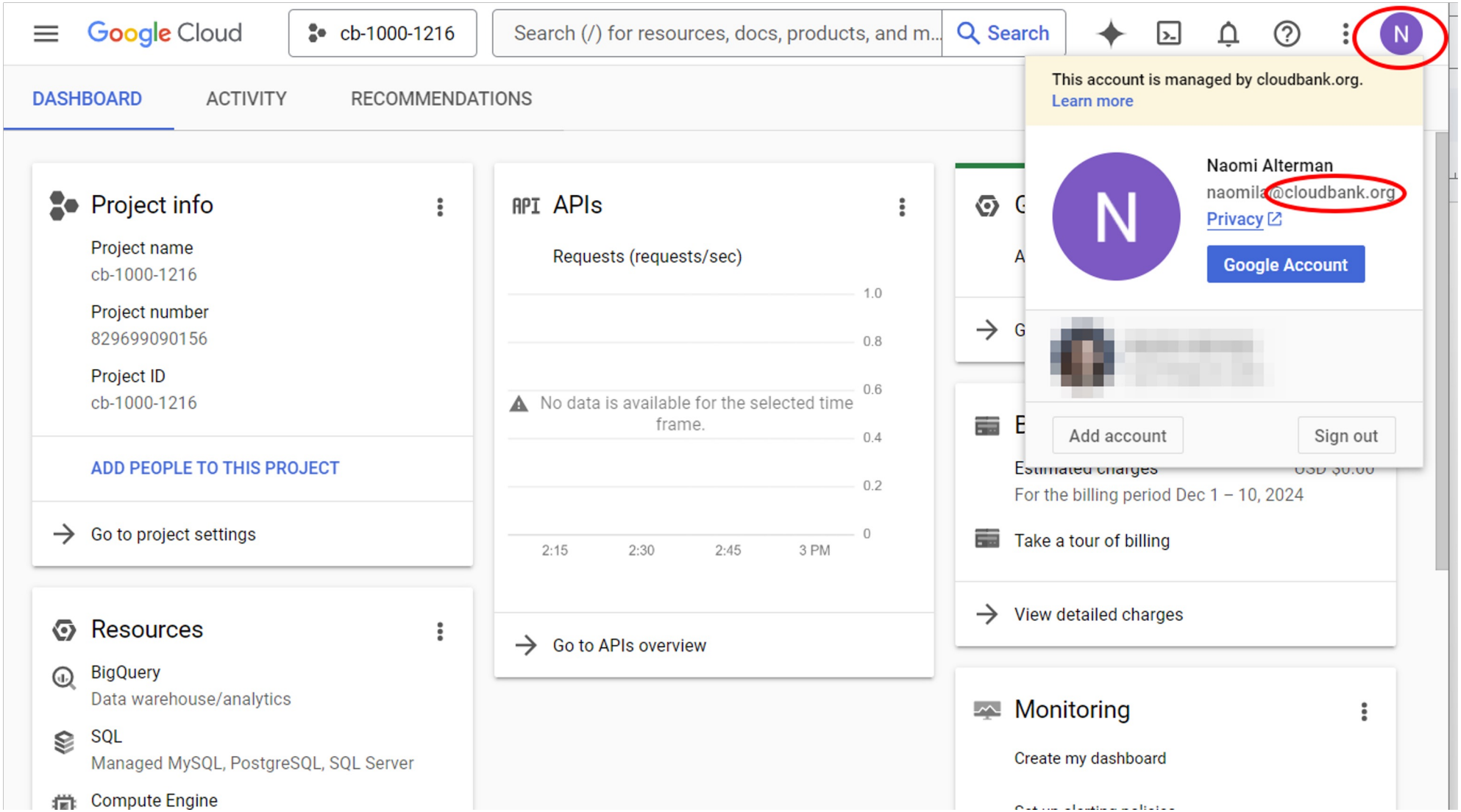
End AWS
Start GCP

Google Cloud Platform (GCP) Fly-over

Looking ahead: Firewall interface



* No specific limit specified on this billing account, value is inherited from fund.


Google Cloud
cb-1000-1216
Search (/) for resources, docs, products, and m...
Search
N
DASHBOARD
ACTIVITY
RECOMMENDATIONS
This account is managed by cloudbank.org.
Learn more
Naomi Alterman
naomila@cloudbank.org
Privacy
Google Account
Add account
Sign out
Project info
Project name
cb-1000-1216
Project number
829699090156
Project ID
cb-1000-1216
ADD PEOPLE TO THIS PROJECT
Go to project settings
APIs
Requests (requests/sec)
1.0
0.8
0.6
0.4
0.2
0
No data is available for the selected time frame.
2:15
2:30
2:45
3 PM
Go to APIs overview
For the billing period Dec 1 – 10, 2024
Take a tour of billing
View detailed charges
Resources
BigQuery
Data warehouse/analytics
SQL
Managed MySQL, PostgreSQL, SQL Server
Compute Engine
Monitoring
Create my dashboard

CloudBank

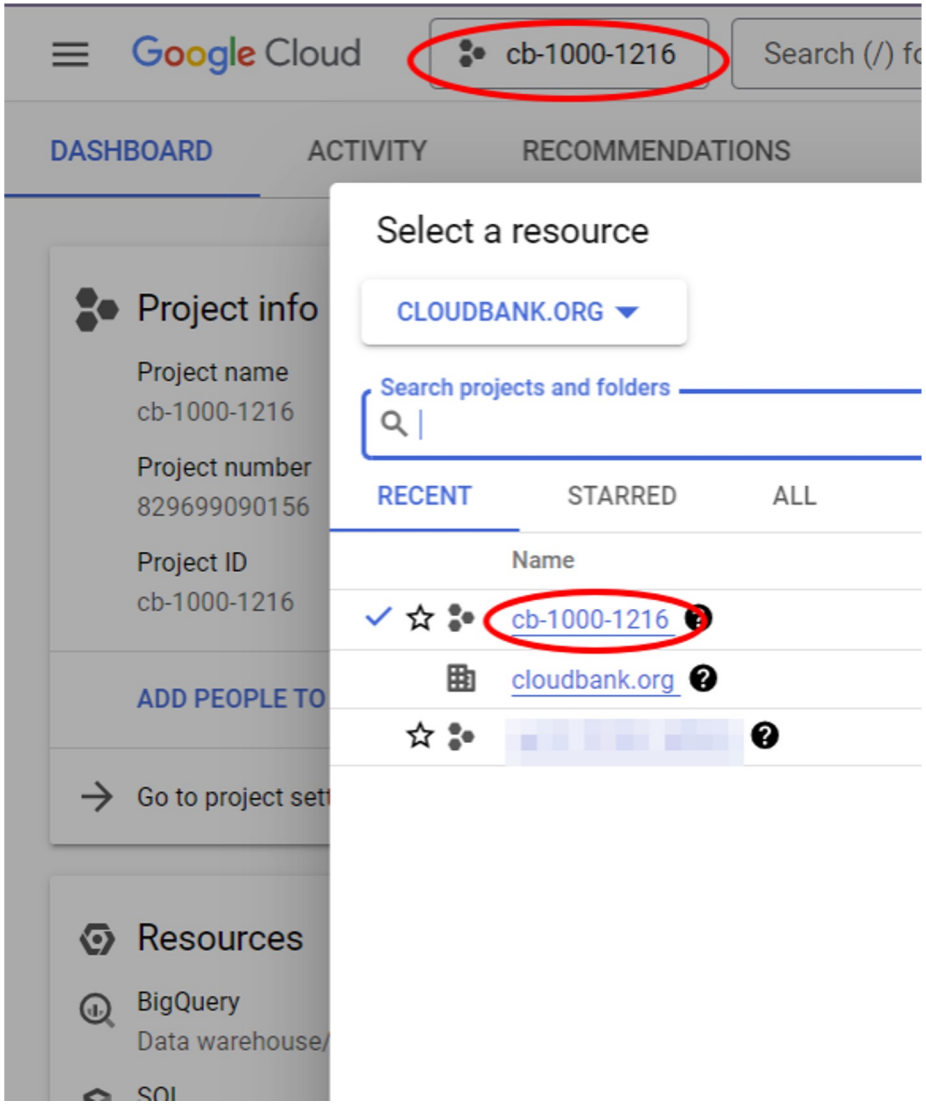
DASHBOARD

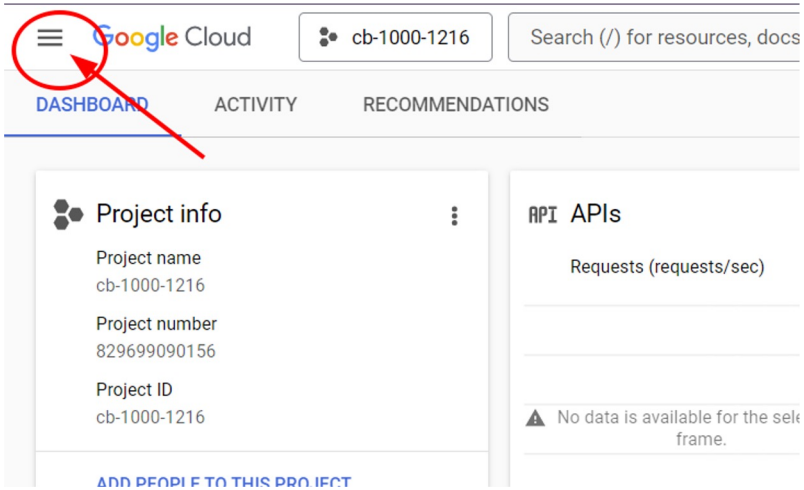
HOME / DASHBOARD / ACCESS CLOUDBANK BILLING ACCOUNTS

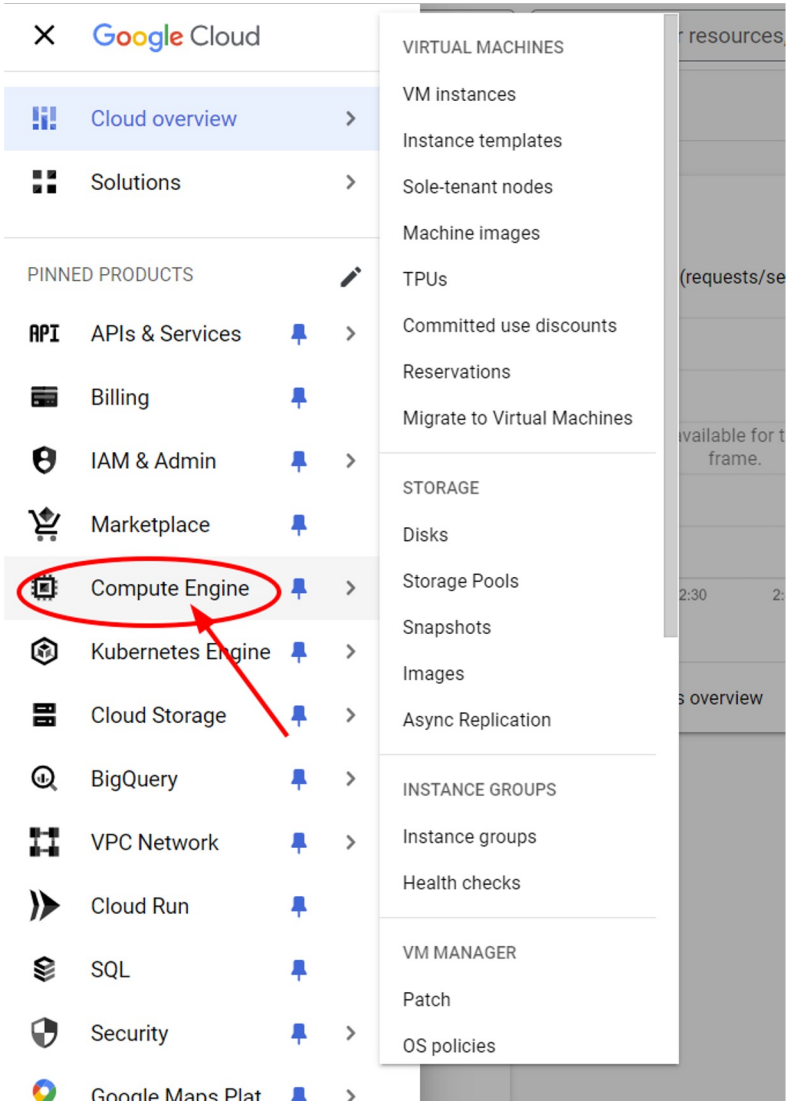
You have access to the following billing accounts. A billing acc
in a shared billing account, please note that we can only track
manager for assistance if you need to track usage at a finer gr

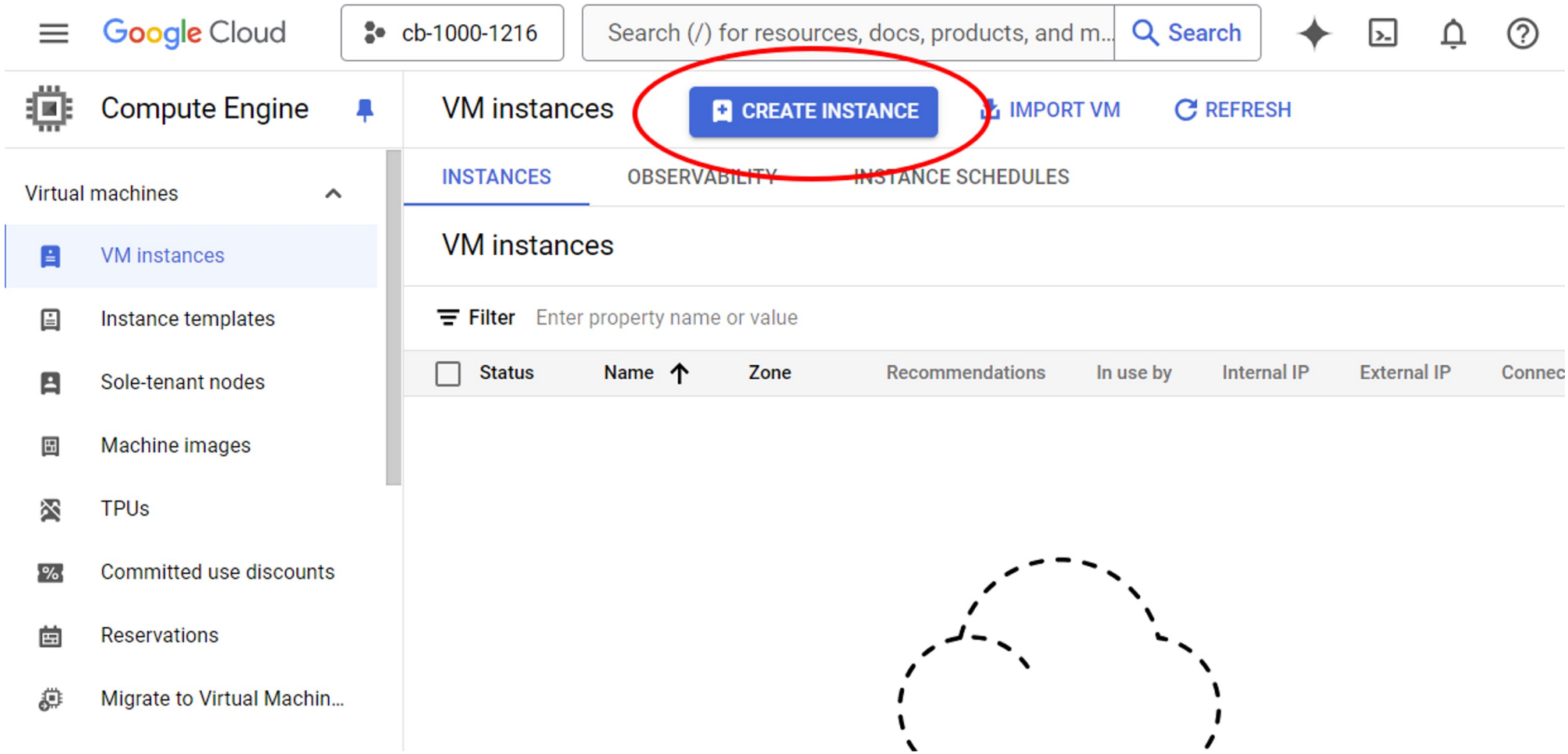
BILLING ACCOUNT ID	CLOUD USERNAME	INITIA GRAI
[redacted]	naomila	Admir
cb-1000-1216	naomila	Admir
[redacted]	naomila	Admir
[redacted]	naomila	Admir

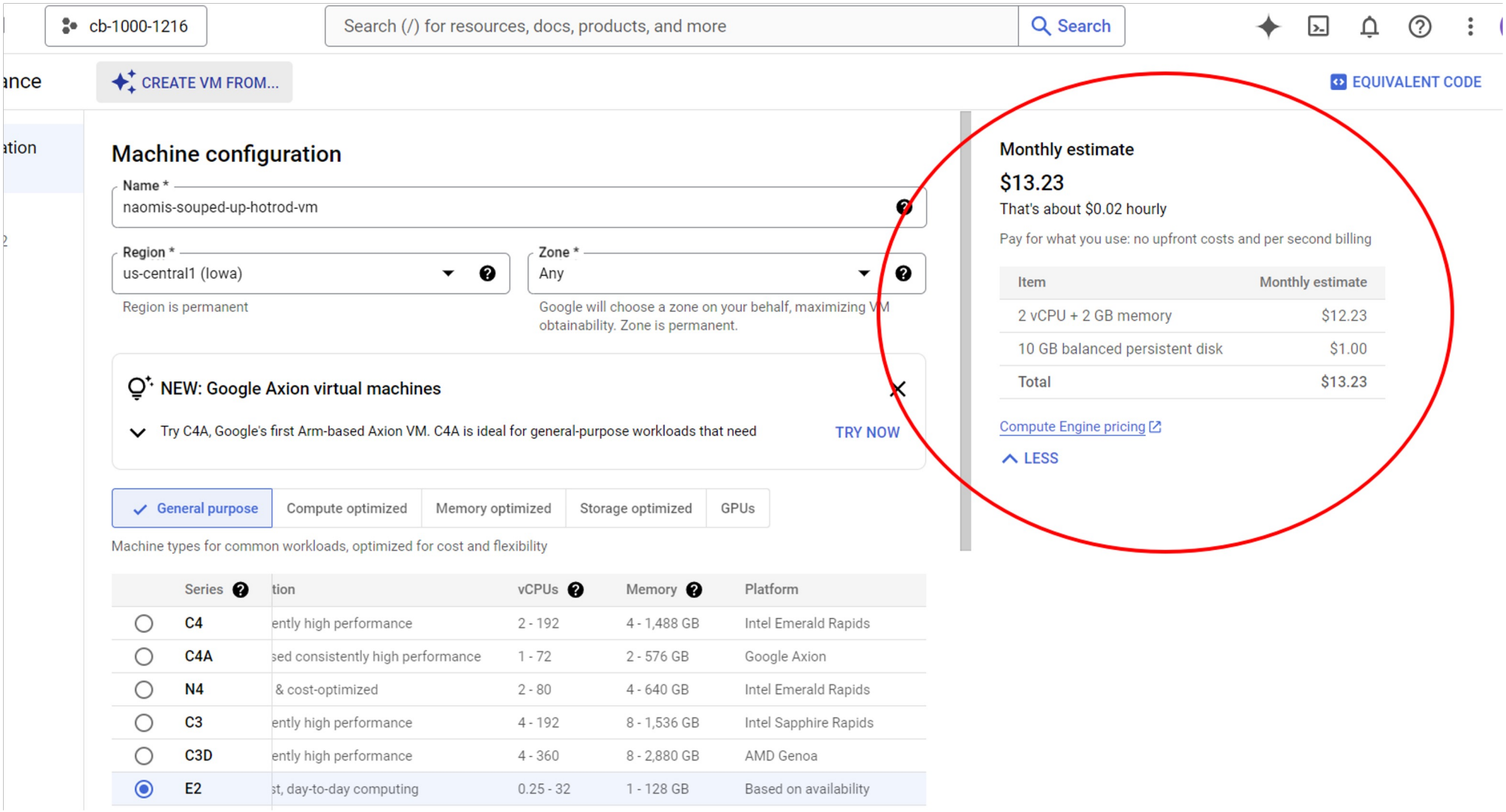
* No specific limit specified on this billing account, value is inh

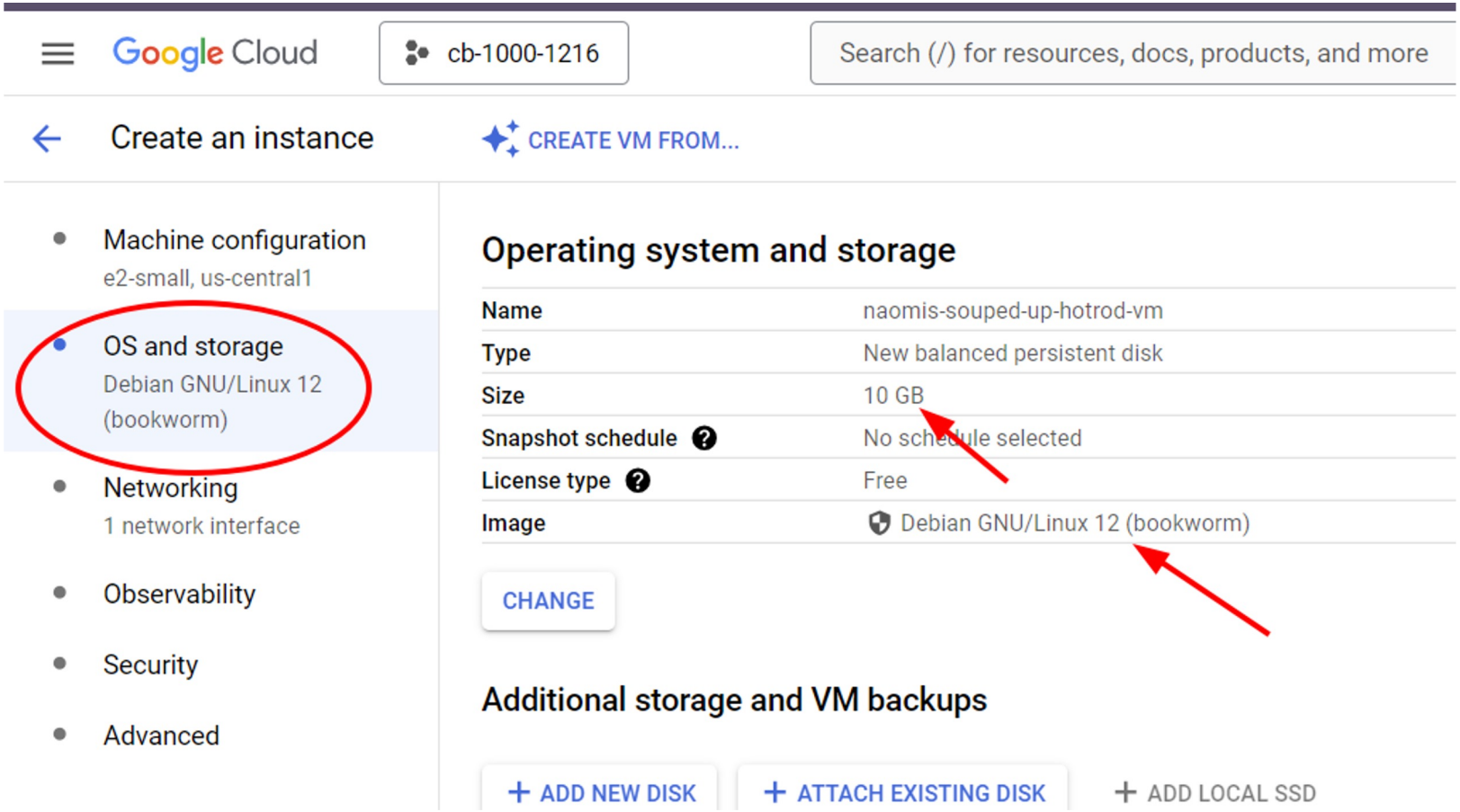


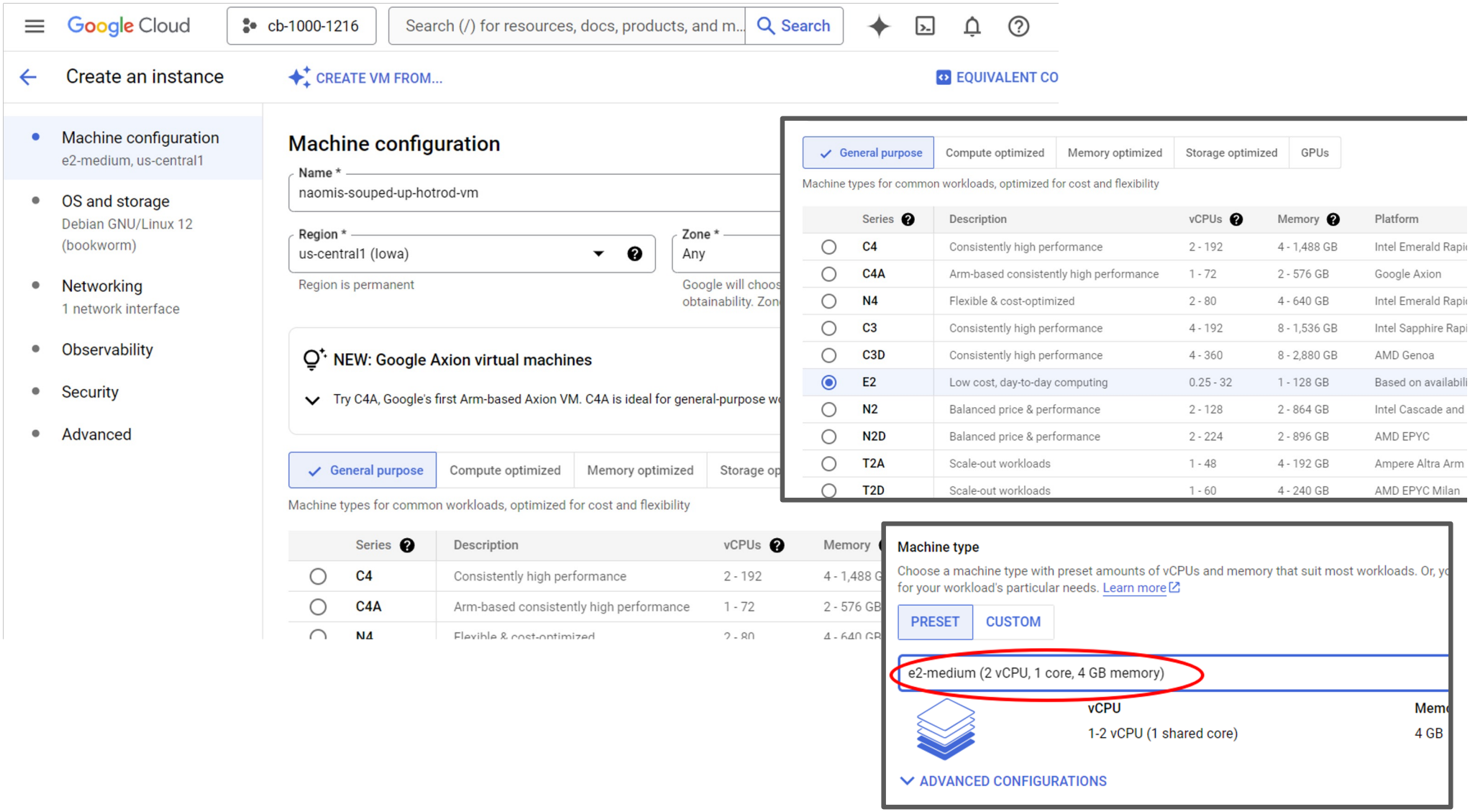

Google Cloud
cb-1000-1216
Search (/) for resources, docs
DASHBOARD
ACTIVITY
RECOMMENDATIONS
Project info
Project name
cb-1000-1216
Project number
829699090156
Project ID
cb-1000-1216
APIs
Requests (requests/sec)
No data is available for the sele
frame.

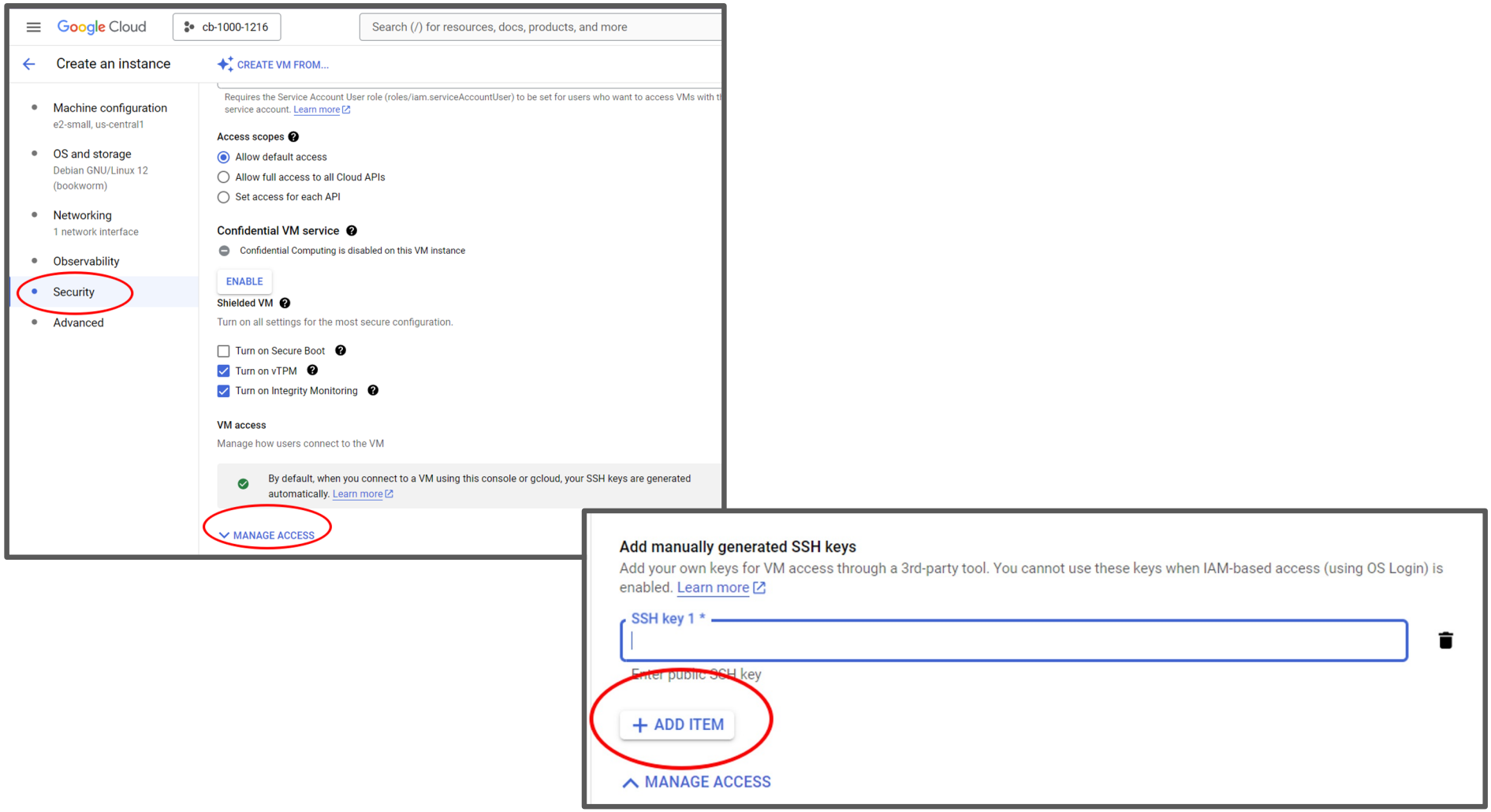

Google Cloud
Cloud overview
Solutions
PINNED PRODUCTS
APIs & Services
Billing
IAM & Admin
Marketplace
Compute Engine
Kubernetes Engine
Cloud Storage
BigQuery
VPC Network
Cloud Run
SQL
Security
VIRTUAL MACHINES
VM instances
Instance templates
Sole-tenant nodes
Machine images
TPUs
Committed use discounts
Reservations
Migrate to Virtual Machines
STORAGE
Disks
Storage Pools
Snapshots
Images
Async Replication
INSTANCE GROUPS
Instance groups
Health checks
VM MANAGER
Patch
OS policies

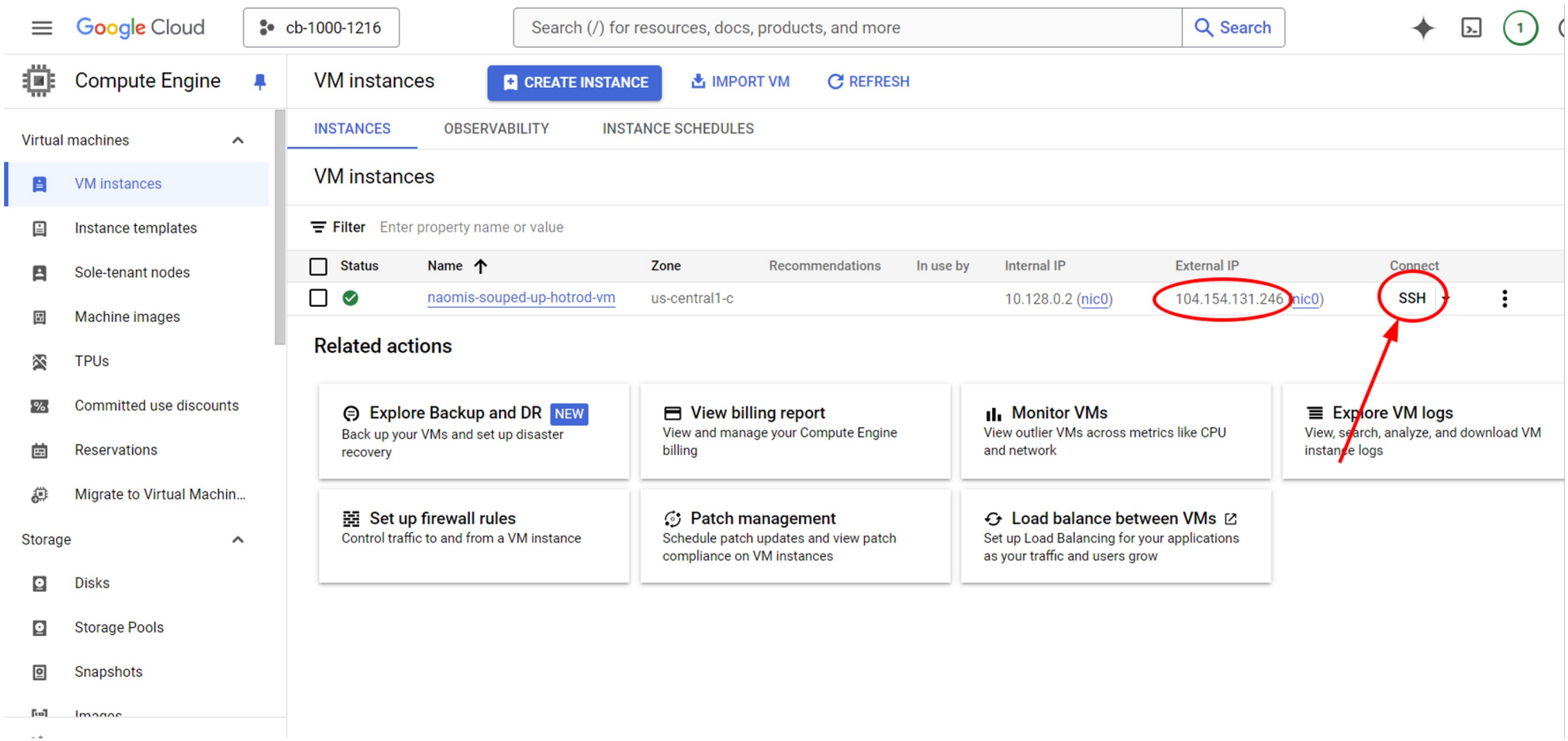

Google Cloud
cb-1000-1216
Search (/) for resources, docs, products, and m...
Search
Compute Engine
VM instances
CREATE INSTANCE
IMPORT VM
REFRESH
Virtual machines
VM instances
Instance templates
Sole-tenant nodes
Machine images
TPUs
Committed use discounts
Reservations
Migrate to Virtual Machin...
INSTANCES
OBSERVABILITY
INSTANCE SCHEDULES
VM instances
Filter
Enter property name or value
Status
Name
Zone
Recommendations
In use by
Internal IP
External IP
Connec

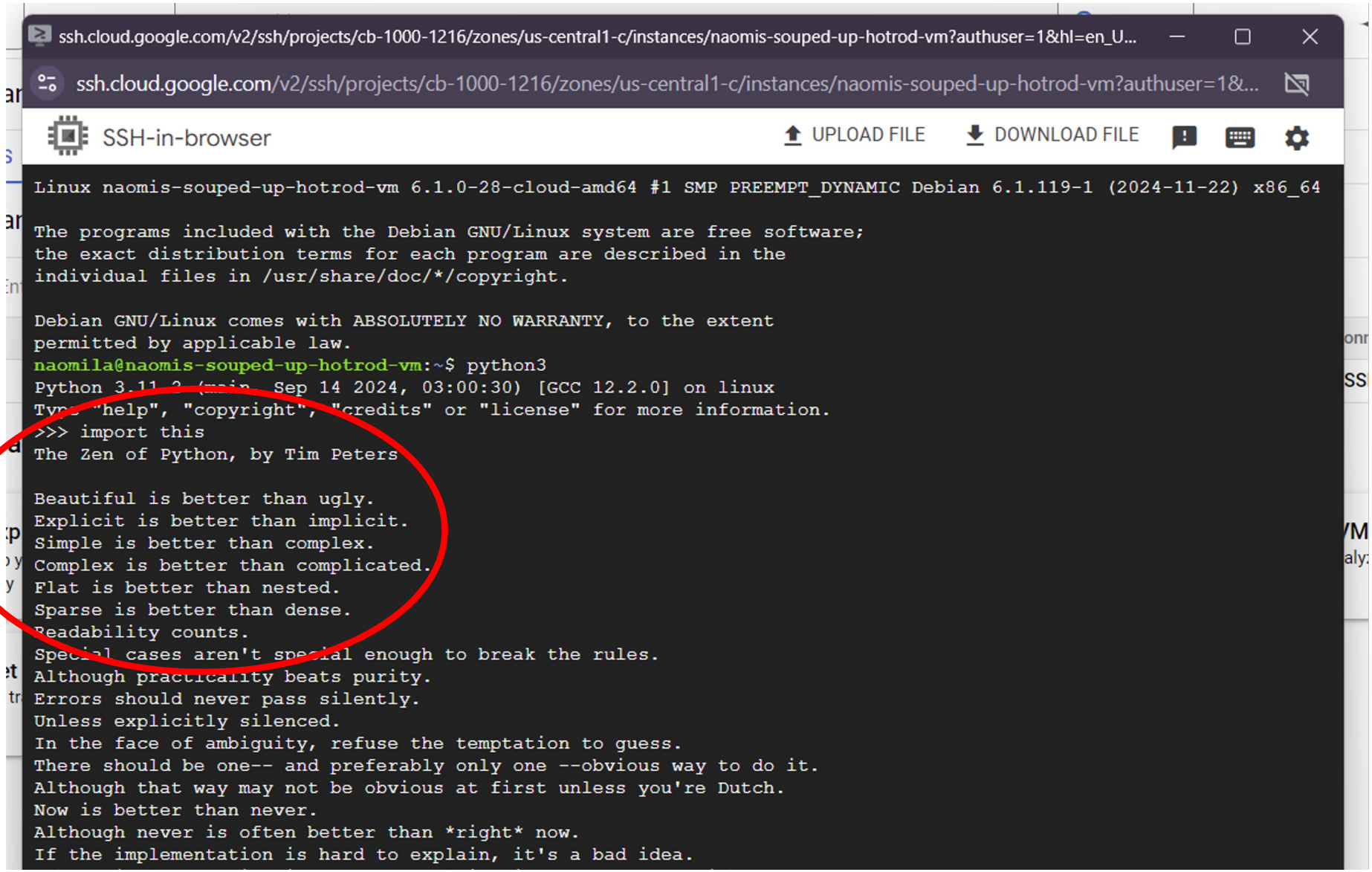

cb-1000-1216
Search (/) for resources, docs, products, and more
Search
CREATE VM FROM...
EQUIVALENT CODE
Machine configuration
Name *
naomis-souped-up-hotrod-vm
Region *
us-central1 (Iowa)
Region is permanent
Zone *
Any
Google will choose a zone on your behalf, maximizing VM obtainability. Zone is permanent.
NEW: Google Axion virtual machines
Try C4A, Google's first Arm-based Axion VM. C4A is ideal for general-purpose workloads that need
TRY NOW
General purpose
Compute optimized
Memory optimized
Storage optimized
GPUs
Machine types for common workloads, optimized for cost and flexibility
Series
vCPUs
Memory
Platform
C4
2 - 192
4 - 1,488 GB
Intel Emerald Rapids
C4A
1 - 72
2 - 576 GB
Google Axion
N4
2 - 80
4 - 640 GB
Intel Emerald Rapids
C3
4 - 192
8 - 1,536 GB
Intel Sapphire Rapids
C3D
4 - 360
8 - 2,880 GB
AMD Genoa
E2
0.25 - 32
1 - 128 GB
Based on availability
Monthly estimate
$13.23
That's about $0.02 hourly
Pay for what you use: no upfront costs and per second billing
Item
Monthly estimate
2 vCPU + 2 GB memory
$12.23
10 GB balanced persistent disk
$1.00
Total
$13.23
Compute Engine pricing
LESS

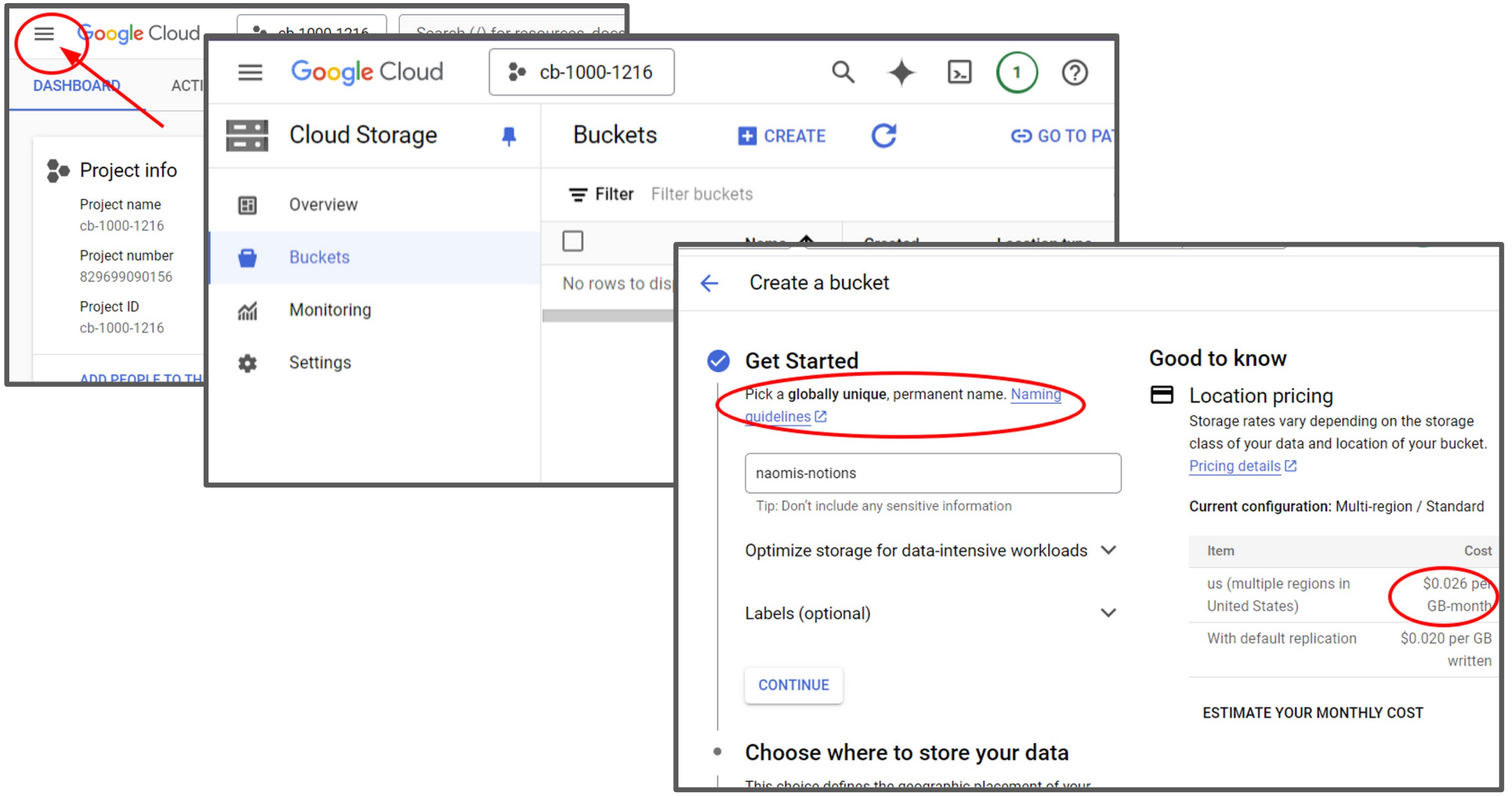

Google Cloud
cb-1000-1216
Search (/) for resources, docs, products, and more
Create an instance
CREATE VM FROM...
Machine configuration
e2-small, us-central1
OS and storage
Debian GNU/Linux 12 (bookworm)
Networking
1 network interface
Observability
Security
Advanced
Operating system and storage
Name
naomis-souped-up-hotrod-vm
Type
New balanced persistent disk
Size
10 GB
Snapshot schedule
No schedule selected
License type
Free
Image
Debian GNU/Linux 12 (bookworm)
CHANGE
Additional storage and VM backups
+ ADD NEW DISK
+ ATTACH EXISTING DISK
+ ADD LOCAL SSD

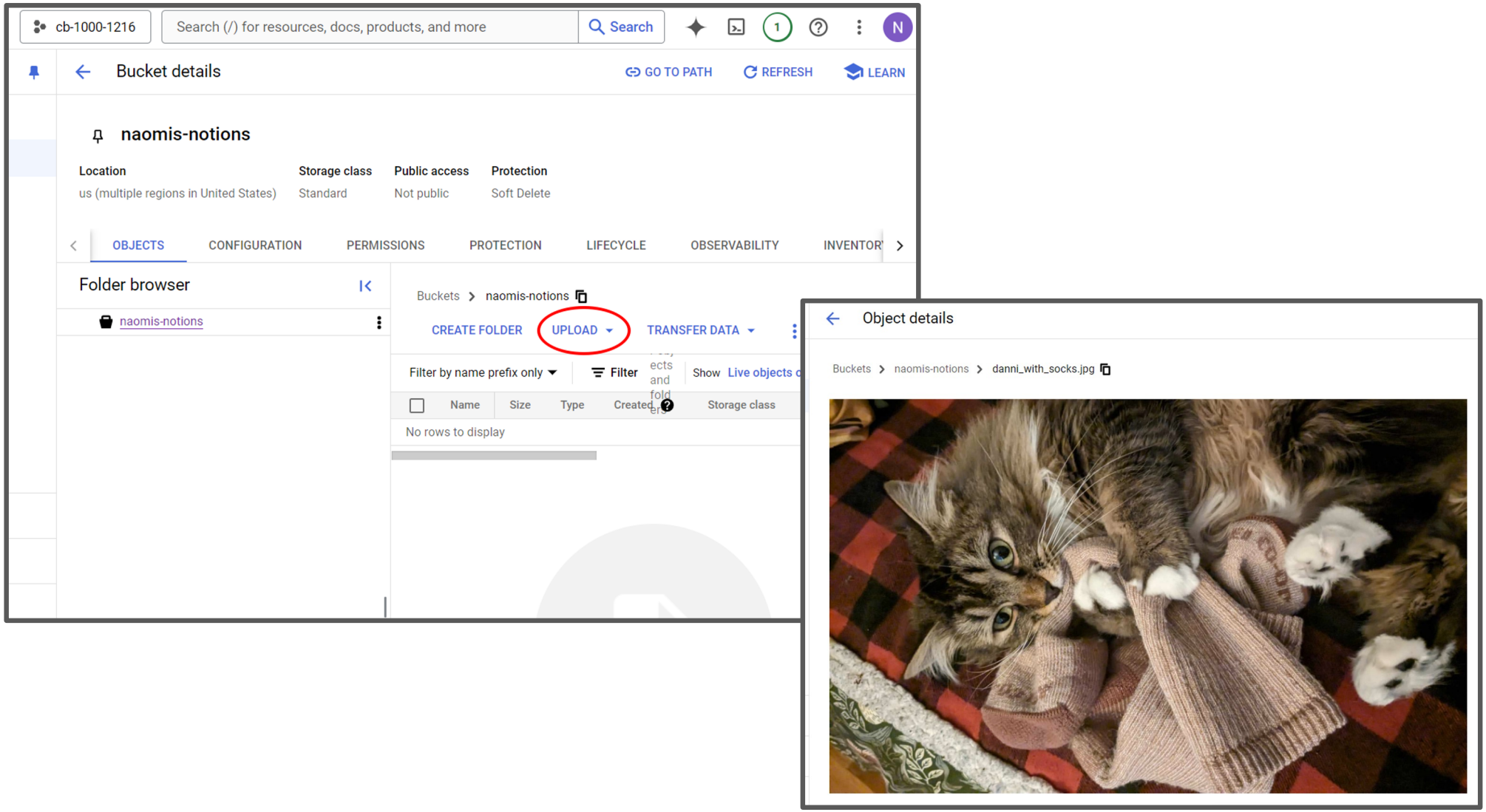

Google Cloud
cb-1000-1216
Search (/) for resources, docs, products, and m...
Search
Create an instance
CREATE VM FROM...
EQUIVALENT CO
Machine configuration
e2-medium, us-central1
OS and storage
Debian GNU/Linux 12 (bookworm)
Networking
1 network interface
Observability
Security
Advanced
Machine configuration
Name *
naomis-souped-up-hotrod-vm
Region *
us-central1 (Iowa)
Region is permanent
Zone *
Any
NEW: Google Axion virtual machines
Try C4A, Google's first Arm-based Axion VM. C4A is ideal for general-purpose w
General purpose
Compute optimized
Memory optimized
Storage op
Machine types for common workloads, optimized for cost and flexibility
Series
Description
vCPUs
Memory
C4
Consistently high performance
2 - 192
4 - 1,488 G
C4A
Arm-based consistently high performance
1 - 72
2 - 576 GB
N4
Flexible & cost-optimized
2 - 80
4 - 640 GB
General purpose
Compute optimized
Memory optimized
Storage optimized
GPUs
Machine types for common workloads, optimized for cost and flexibility
Series
Description
vCPUs
Memory
Platform
C4
Consistently high performance
2 - 192
4 - 1,488 GB
Intel Emerald Rapi
C4A
Arm-based consistently high performance
1 - 72
2 - 576 GB
Google Axion
N4
Flexible & cost-optimized
2 - 80
4 - 640 GB
Intel Emerald Rapi
C3
Consistently high performance
4 - 192
8 - 1,536 GB
Intel Sapphire Rapi
C3D
Consistently high performance
4 - 360
8 - 2,880 GB
AMD Genoa
E2
Low cost, day-to-day computing
0.25 - 32
1 - 128 GB
Based on availabili
N2
Balanced price & performance
2 - 128
2 - 864 GB
Intel Cascade and
N2D
Balanced price & performance
2 - 224
2 - 896 GB
AMD EPYC
T2A
Scale-out workloads
1 - 48
4 - 192 GB
Ampere Altra Arm
T2D
Scale-out workloads
1 - 60
4 - 240 GB
AMD EPYC Milan
Machine type
Choose a machine type with preset amounts of vCPUs and memory that suit most workloads. Or, yo
for your workload's particular needs. Learn more
PRESET
CUSTOM
e2-medium (2 vCPU, 1 core, 4 GB memory)
vCPU
1-2 vCPU (1 shared core)
Mem
4 GB
ADVANCED CONFIGURATIONS


Google Cloud
cb-1000-1216
Search (/) for resources, docs, products, and more
Create an instance
CREATE VM FROM...
Machine configuration
e2-small, us-central1
OS and storage
Debian GNU/Linux 12 (bookworm)
Networking
1 network interface
Observability
Security
Advanced
Requires the Service Account User role (roles/iam.serviceAccountUser) to be set for users who want to access VMs with t
service account. Learn more
Access scopes
Allow default access
Allow full access to all Cloud APIs
Set access for each API
Confidential VM service
Confidential Computing is disabled on this VM instance
ENABLE
Shielded VM
Turn on all settings for the most secure configuration.
Turn on Secure Boot
Turn on vTPM
Turn on Integrity Monitoring
VM access
Manage how users connect to the VM
By default, when you connect to a VM using this console or gcloud, your SSH keys are generated automatically. Learn more
MANAGE ACCESS
Add manually generated SSH keys
Add your own keys for VM access through a 3rd-party tool. You cannot use these keys when IAM-based access (using OS Login) is enabled. Learn more
SSH key 1 *
Enter public SSH key
ADD ITEM
MANAGE ACCESS


Google Cloud
cb-1000-1216
Search (/) for resources, docs, products, and more
Search
Compute Engine
VM instances
CREATE INSTANCE
IMPORT VM
REFRESH
Virtual machines
VM instances
Instance templates
Sole-tenant nodes
Machine images
TPUs
Committed use discounts
Reservations
Migrate to Virtual Machin...
Storage
Disks
Storage Pools
Snapshots
INSTANCES
OBSERVABILITY
INSTANCE SCHEDULES
VM instances
Filter
Enter property name or value
Status
Name
Zone
Recommendations
In use by
Internal IP
External IP
Connect
naomis-souped-up-hotrod-vm
us-central1-c
10.128.0.2 (nic0)
104.154.131.246 (nic0)
SSH
Related actions
Explore Backup and DR
NEW
Back up your VMs and set up disaster recovery
View billing report
View and manage your Compute Engine billing
Monitor VMs
View outlier VMs across metrics like CPU and network
Explore VM logs
View, search, analyze, and download VM instance logs
Set up firewall rules
Control traffic to and from a VM instance
Patch management
Schedule patch updates and view patch compliance on VM instances
Load balance between VMs
Set up Load Balancing for your applications as your traffic and users grow


ssh.cloud.google.com/v2/ssh/projects/cb-1000-1216/zones/us-central1-c/instances/naomis-souped-up-hotrod-vm?authuser=1&hl=en_U...
ssh.cloud.google.com/v2/ssh/projects/cb-1000-1216/zones/us-central1-c/instances/naomis-souped-up-hotrod-vm?authuser=1&...
SSH-in-browser
UPLOAD FILE
DOWNLOAD FILE
Linux naomis-souped-up-hotrod-vm 6.1.0-28-cloud-amd64 #1 SMP PREEMPT_DYNAMIC Debian 6.1.119-1 (2024-11-22) x86_64
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
naomila@naomis-souped-up-hotrod-vm:~$ python3
Python 3.11.2 (main, Sep 14 2024, 03:00:30) [GCC 12.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import this
The Zen of Python, by Tim Peters
Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.
In the face of ambiguity, refuse the temptation to guess.
There should be one-- and preferably only one --obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.
Now is better than never.
Although never is often better than *right* now.
If the implementation is hard to explain, it's a bad idea.


Google Cloud
cb-1000-1216
DASHBOARD
Project info
Project name
cb-1000-1216
Project number
829699090156
Project ID
cb-1000-1216
Cloud Storage
Overview
Buckets
Monitoring
Settings
Buckets
CREATE
Filter
Filter buckets
Create a bucket
Get Started
Pick a globally unique, permanent name. Naming guidelines
naomis-notions
Tip: Don't include any sensitive information
Optimize storage for data-intensive workloads
Labels (optional)
CONTINUE
Choose where to store your data
Good to know
Location pricing
Storage rates vary depending on the storage class of your data and location of your bucket.
Pricing details
Current configuration: Multi-region / Standard
Item
Cost
us (multiple regions in United States)
$0.026 per GB-month
With default replication
$0.020 per GB written
ESTIMATE YOUR MONTHLY COST

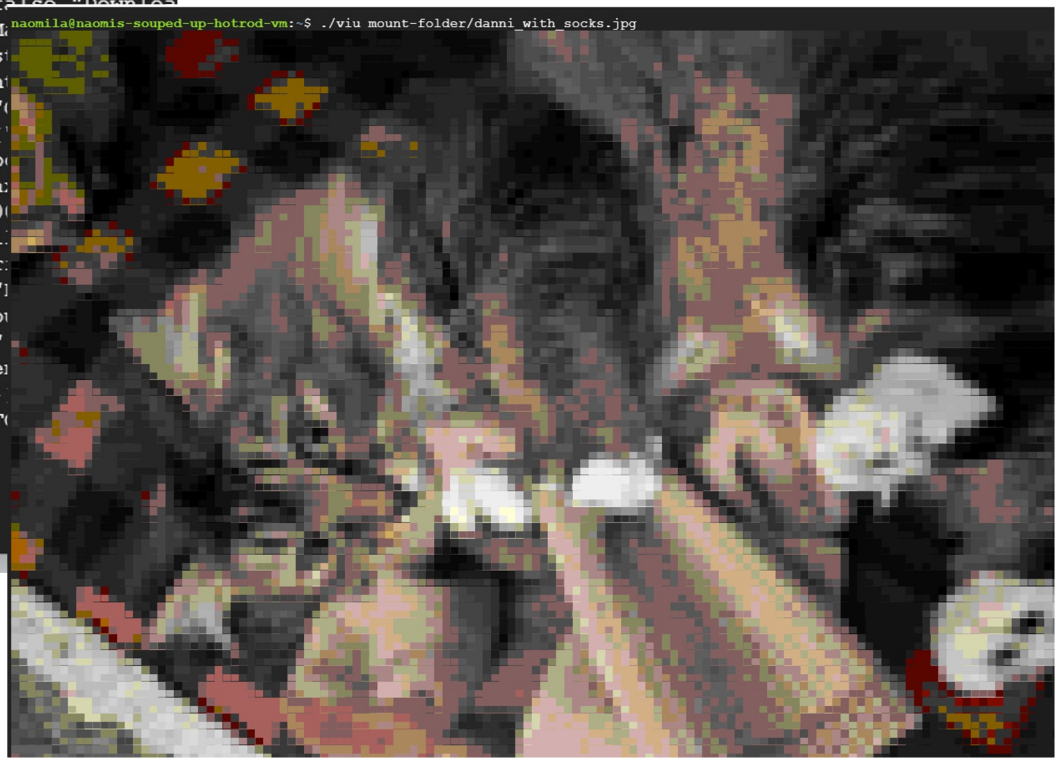

cb-1000-1216
Search (/) for resources, docs, products, and more
Search
Bucket details
GO TO PATH
REFRESH
LEARN
naomis-notions
Location
us (multiple regions in United States)
Storage class
Standard
Public access
Not public
Protection
Soft Delete
OBJECTS
CONFIGURATION
PERMISSIONS
PROTECTION
LIFECYCLE
OBSERVABILITY
Folder browser
naomis-notions
Buckets
naomis-notions
CREATE FOLDER
UPLOAD
TRANSFER DATA
Filter by name prefix only
Filter
Name
Size
Type
Created
Storage class
No rows to display
Object details
Buckets
naomis-notions
danni_with_socks.jpg

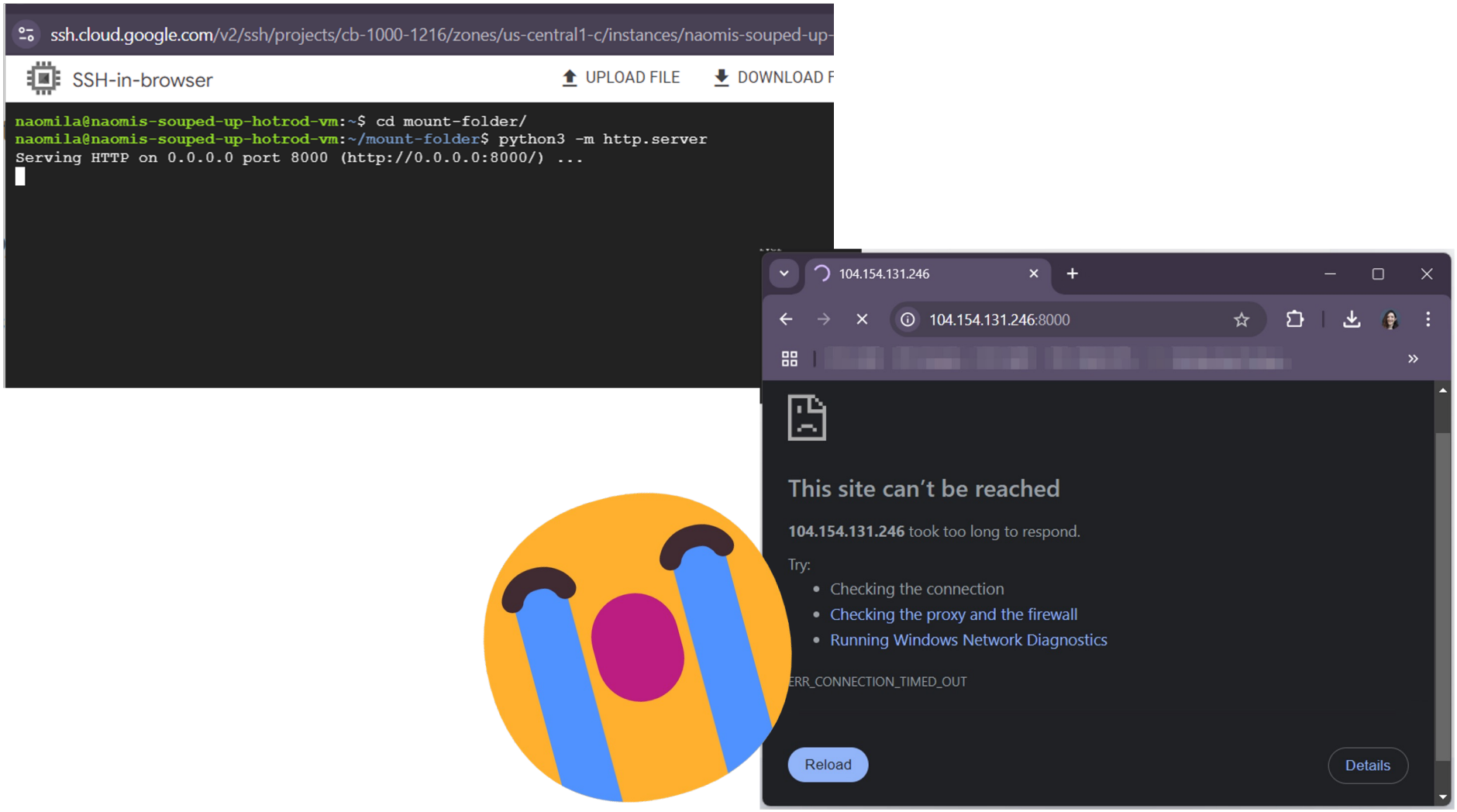
ssh.cloud.google.com/v2/ssh/projects/cb-1000-1216/zones/us-central1-c/instances/naomis-souped-up-hotrod-vm?

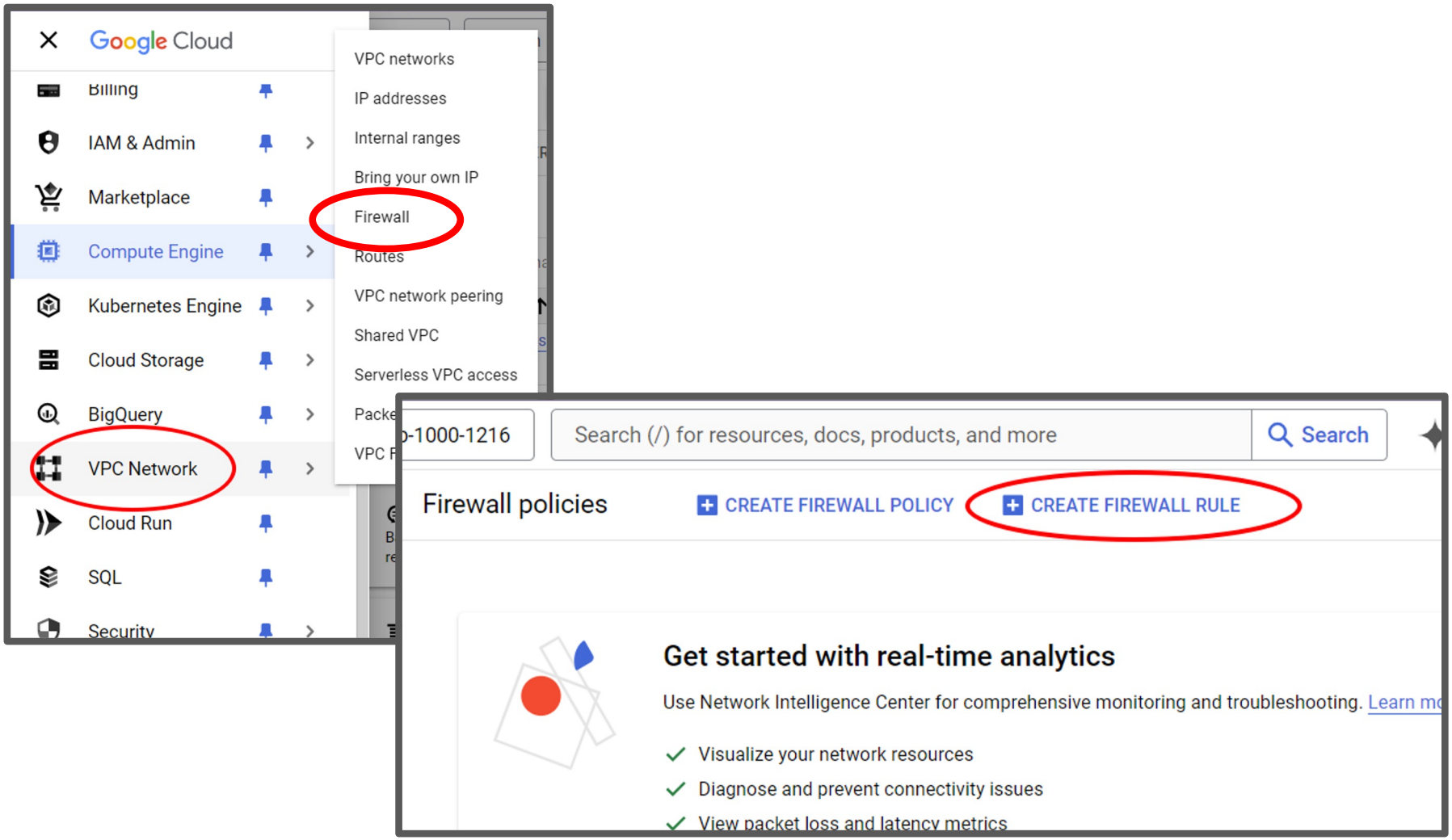
ssh.cloud.google.com/v2/ssh/projects/cb-1000-1216/zones/us-central1-c/instances/naomis-soup

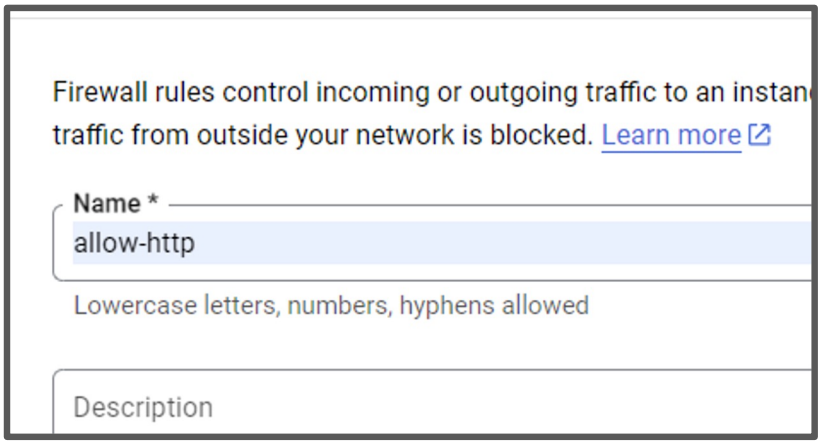
SSH-in-browser UPLOAD FILE DOWN

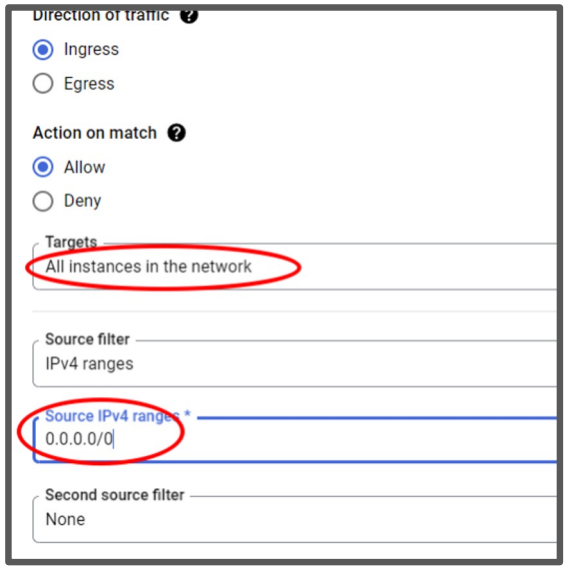
```
naomila@naomis-souped-up-hotrod-vm:~$ gcsfuse naomis-notions mount-folder/
{"timestamp":{"seconds":1733874421,"nanos":903160256},"severity":"INFO","message"
Go version go1.23.0) for app \"\" using mount point: /home/naomila/mount-folder\n
{"timestamp":{"seconds":1733874421,"nanos":903219046},"severity":"INFO","message"
ig":{"AppName":"","CacheDir":"","Debug":{"ExitOnInvariantViolation":false,"Fuse":
utex":false},"EnableHns":true,"FileCache":{"CacheFileForRangeRead":false,"Downloa
eCrc":false,"EnableODirect":false,"EnableParallelDownloads":false,"M
-1,"ParallelDownloadsPerFile":16,"WriteBufferSize":4194304},"FileSys
elDirops":false,"FileMode":"644","FuseOptions":[],"Gid":-1,"IgnoreIn
cs":0,"RenameDirLimit":0,"TempDir":"","Uid":-1},"Foreground":false,"
eyFile":"","ReuseTokenFromUrl":true,"TokenUrl":""},"GcsConnection":{
:"http1","CustomEndpoint":"","ExperimentalEnableJsonRead":false,"Grp
:0,"LimitBytesPerSec":-1,"LimitOpsPerSec":-1,"MaxConnsPerHost":0,"Ma
adSizeMb":200},"GcsRetries":{"MaxRetryAttempts":0,"MaxRetrySleep":30
irs":false,"List":{"EnableEmptyManagedFolders":false},"Logging":{"Fi
e":{"BackupFileCount":10,"Compress":true,"MaxFileSizeMb":512},"Sever
catedStatCacheCapacity":20460,"DeprecatedStatCacheTtl":60000000000,"
"EnableNonexistentTypeCache":false,"ExperimentalMetadataPrefetchOnMo
:32,"TtlSecs":60,"TypeCacheMaxSizeMb":4},"Metrics":{"PrometheusPort"
onitoring":{"ExperimentalOpentelemetryCollectorAddress":"","Experime
cingSamplingRatio":0},"OnlyDir":"","Write":{"CreateEmptyFile":false}
{"timestamp":{"seconds":1733874422,"nanos":65201023},"severity":"INF
uccessfully mounted."}
naomila@naomis-souped-up-hotrod-vm:~$ ls mount-folder/
danni_with_socks.jpg
naomila@naomis-souped-up-hotrod-vm:~$
```



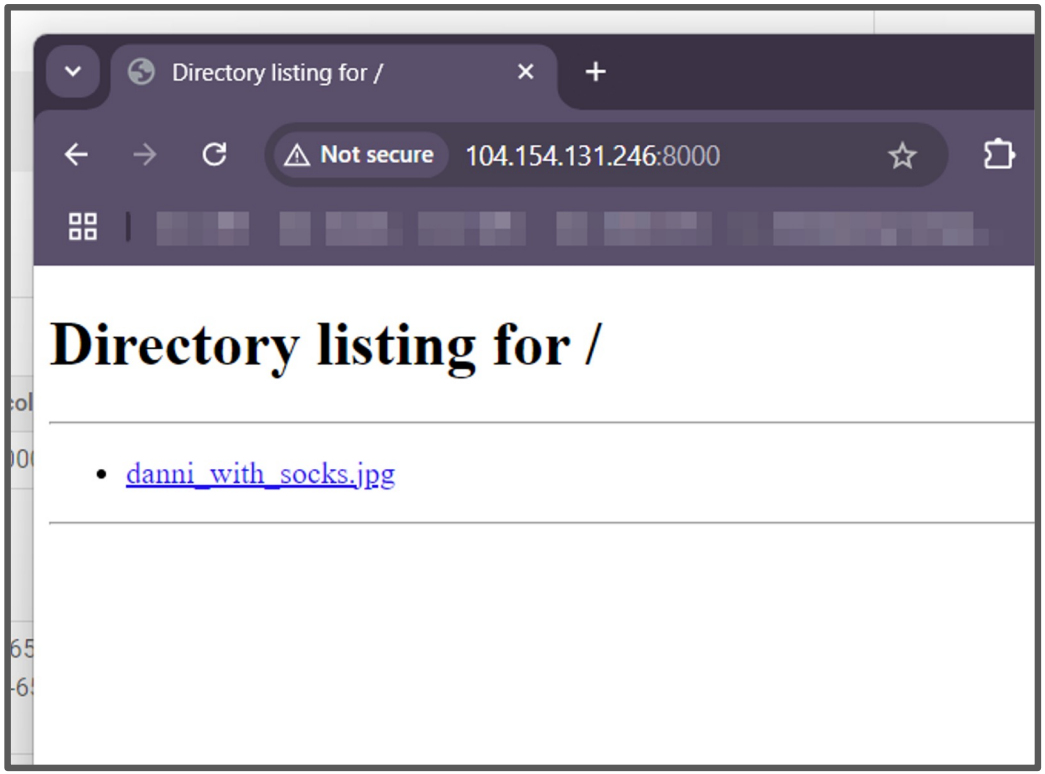

ssh.cloud.google.com/v2/ssh/projects/cb-1000-1216/zones/us-central1-c/instances/naomis-souped-up-
SSH-in-browser
UPLOAD FILE
DOWNLOAD F
naomila@naomis-souped-up-hotrod-vm:~$ cd mount-folder/
naomila@naomis-souped-up-hotrod-vm:~/mount-folder$ python3 -m http.server
Serving HTTP on 0.0.0.0 port 8000 (http://0.0.0.0:8000/) ...
104.154.131.246
104.154.131.246:8000
This site can't be reached
104.154.131.246 took too long to respond.
Try:
Checking the connection
Checking the proxy and the firewall
Running Windows Network Diagnostics
ERR_CONNECTION_TIMED_OUT
Reload
Details

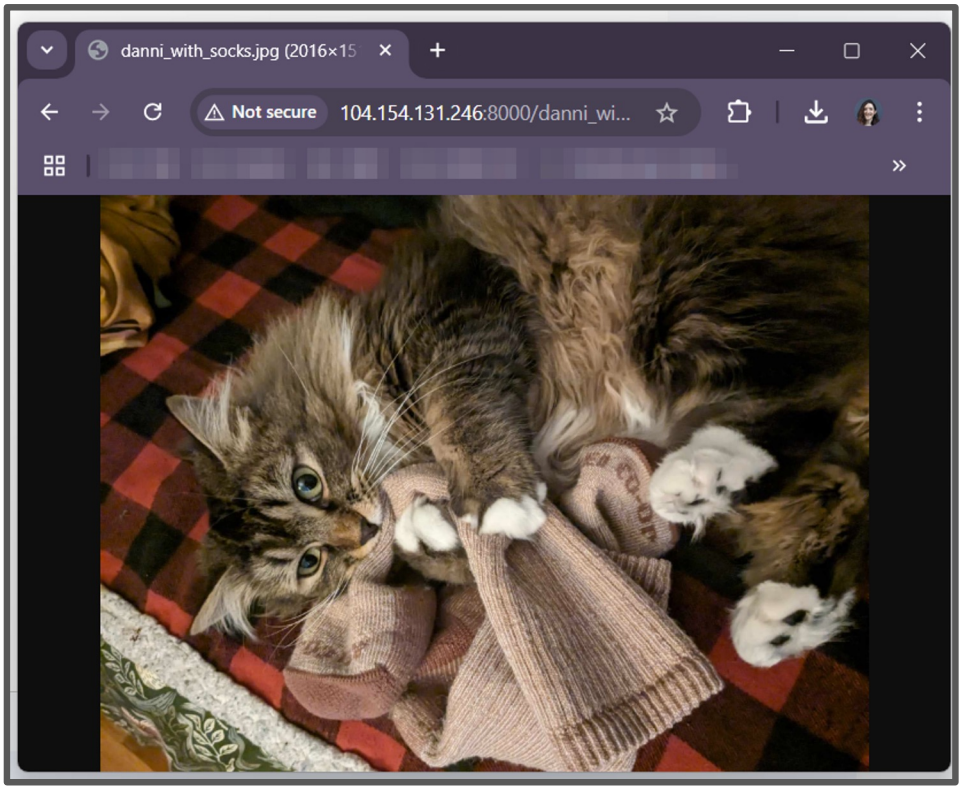

Google Cloud
Billing
IAM & Admin
Marketplace
Compute Engine
Kubernetes Engine
Cloud Storage
BigQuery
VPC Network
Cloud Run
SQL
Security
VPC networks
IP addresses
Internal ranges
Bring your own IP
Firewall
Routes
VPC network peering
Shared VPC
Serverless VPC access
Search (/) for resources, docs, products, and more
Search
Firewall policies
CREATE FIREWALL POLICY
CREATE FIREWALL RULE
Get started with real-time analytics
Use Network Intelligence Center for comprehensive monitoring and troubleshooting.
Visualize your network resources
Diagnose and prevent connectivity issues
View packet loss and latency metrics


Firewall rules control incoming or outgoing traffic to an instan
traffic from outside your network is blocked. Learn more
Name *
allow-http
Lowercase letters, numbers, hyphens allowed
Description


Direction of traffic
Ingress
Egress
Action on match
Allow
Deny
Targets
All instances in the network
Source filter
IPv4 ranges
Source IPv4 ranges *
0.0.0.0/0
Second source filter
None


Protocols and ports
Allow all
Specified protocols and ports
TCP
Ports
8000
E.g. 20, 50-60
UDP
Ports


Directory listing for /
Not secure
104.154.131.246:8000
Directory listing for /
danni_with_socks.jpg


danni_with_socks.jpg (2016×15
Not secure
104.154.131.246:8000/danni_wi...

19 down

11 to go

End GCP
Start AI Overview

Artificial Intelligence Stacks

- **AI** defined in the broad sense
  - ~Equivalent to Machine Learning, hence **AI/ML**
  - Generative modes are referred to as **genAI**: LLMs and related
- Each major cloud has an AI “laboratory environment”
  - AWS SageMaker
  - Google Colab and AI Studio
  - Azure AI Studio (genAI-specialized) and ML Studio (data science AI/ML)
- These paid services connect to scalable GPU resources

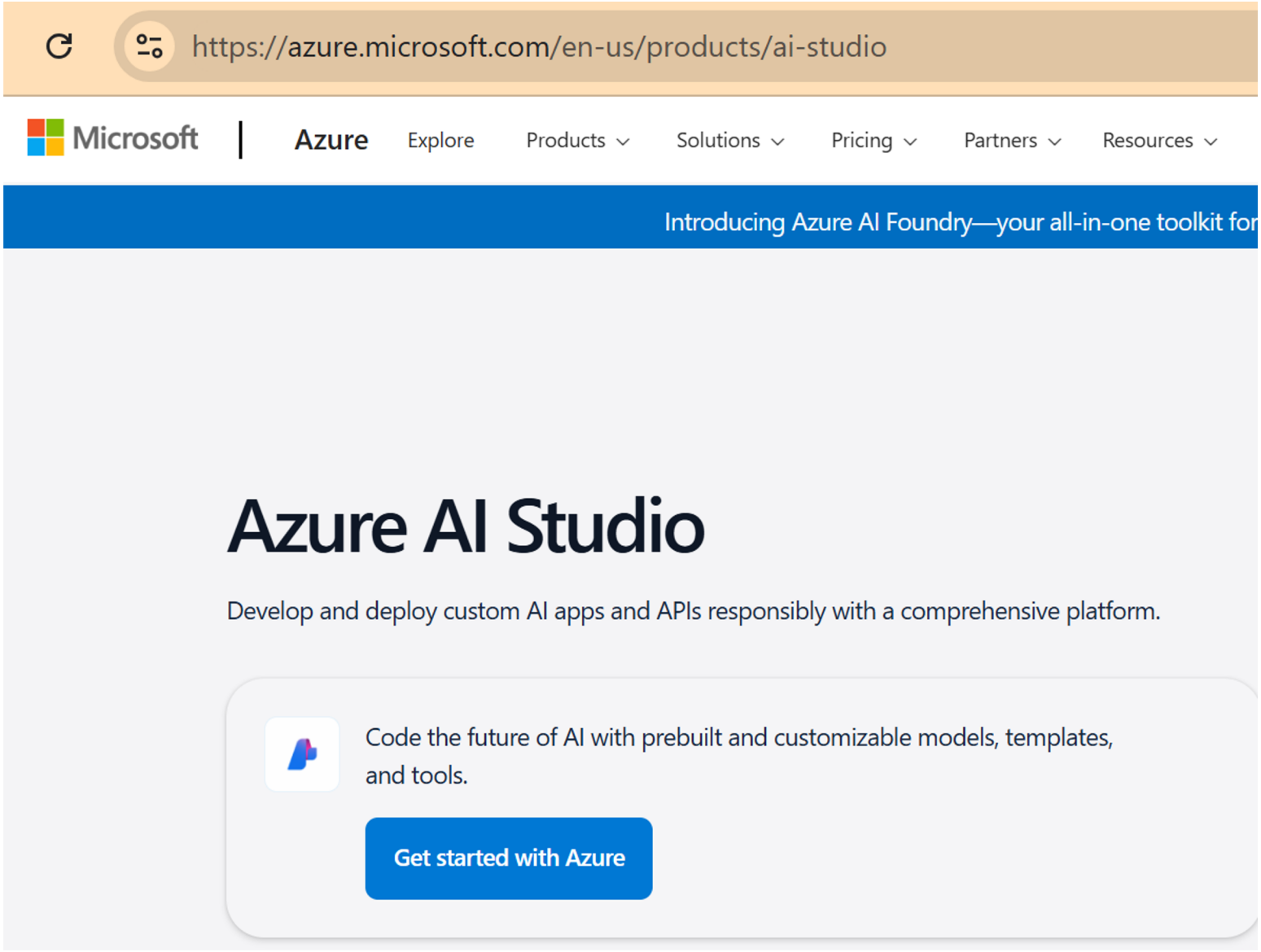
1 down

5 to go

AI 3 vendors:
10 slides / vendor

End AI Overview
Start Azure AI

Azure…


https://azure.microsoft.com/en-us/products/ai-studio
Microsoft
Azure
Explore
Products
Solutions
Pricing
Partners
Resources
Introducing Azure AI Foundry—your all-in-one toolkit for
Azure AI Studio
Develop and deploy custom AI apps and APIs responsibly with a comprehensive platform.
Code the future of AI with prebuilt and customizable models, templates, and tools.
Get started with Azure

Azure *services*

- Search ‘Azure AI services’ for latest
- Examples —->
  - “building apps”

Available Azure AI services

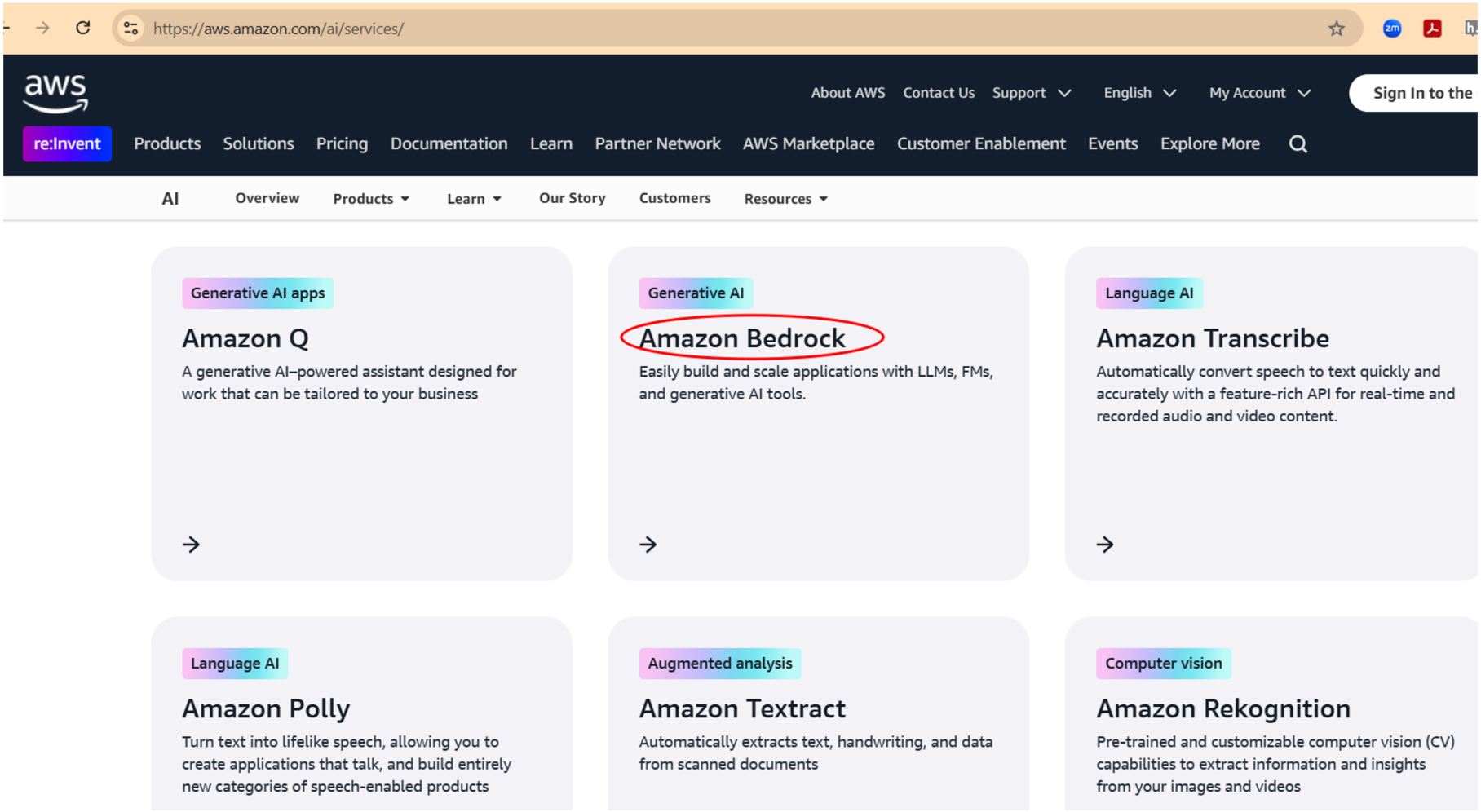
When building AI applications, use the following Azure AI services:

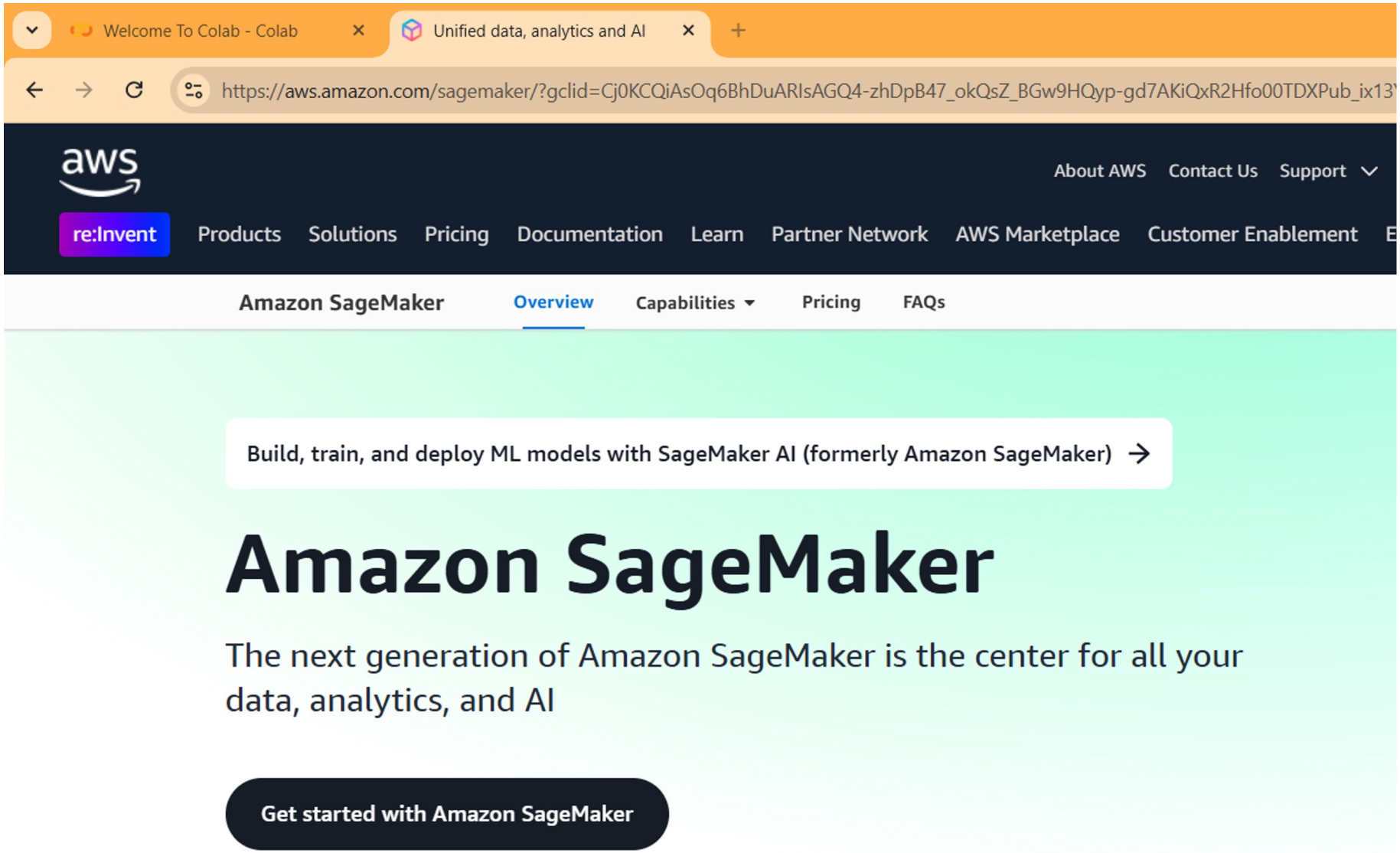
Expand table

Service	Description
Azure AI Search	Bring AI-powered cloud search to your mobile and web apps.
Azure OpenAI	Perform a wide variety of natural language tasks.
Bot Service	Create bots and connect them across channels.
Content Safety	An AI service that detects unwanted contents.
Custom Vision	Customize image recognition for your business.
Document Intelligence	Turn documents into intelligent data-driven solutions.
Face	Detect and identify people and emotions in images.

End Azure AI
Start AWS AI

AWS…

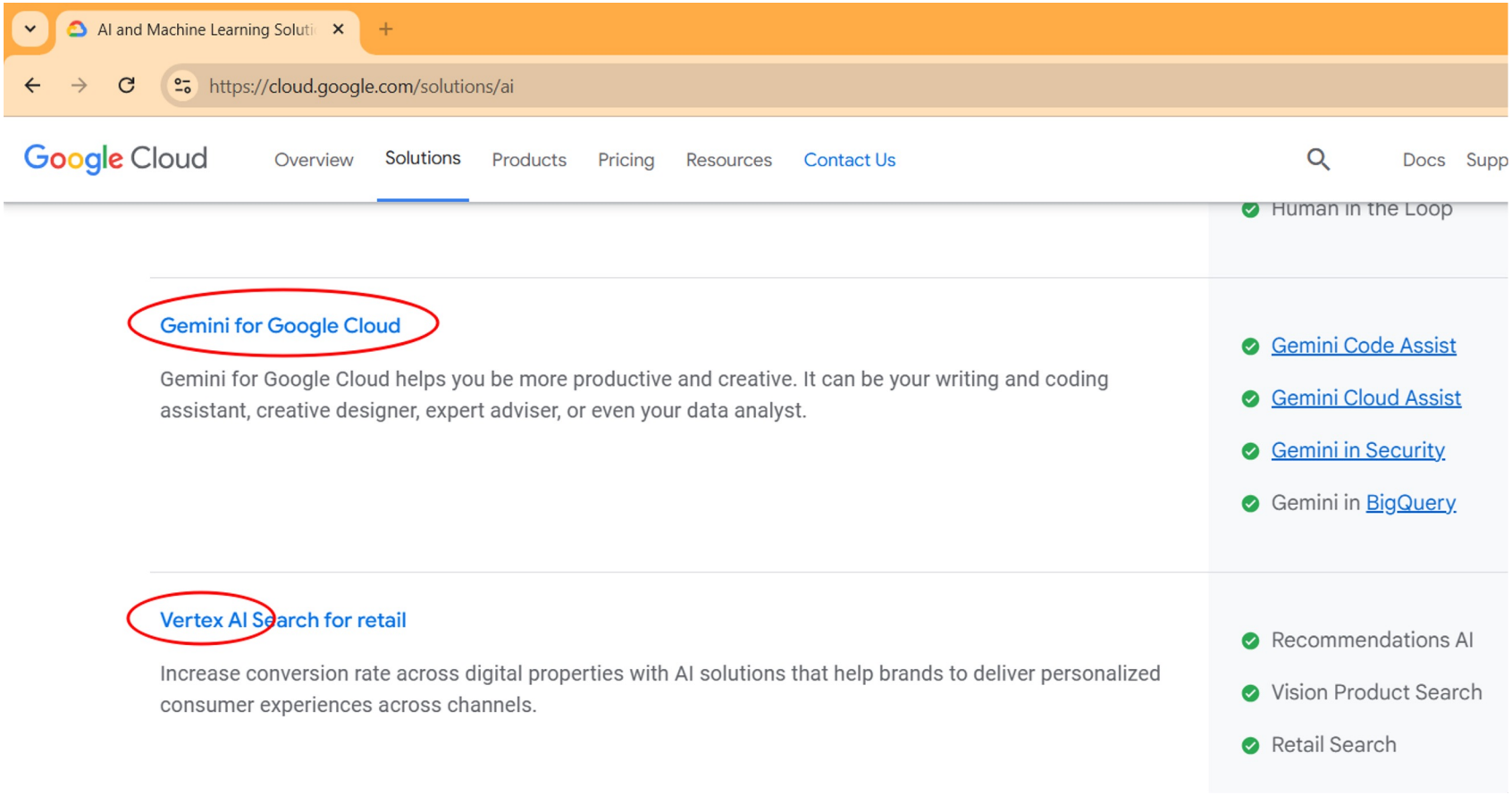

https://aws.amazon.com/ai/services/
aws
About AWS
Contact Us
Support
English
My Account
Sign In to the
re:Invent
Products
Solutions
Pricing
Documentation
Learn
Partner Network
AWS Marketplace
Customer Enablement
Events
Explore More
AI
Overview
Products
Learn
Our Story
Customers
Resources
Generative AI apps
Amazon Q
A generative AI–powered assistant designed for work that can be tailored to your business
Generative AI
Amazon Bedrock
Easily build and scale applications with LLMs, FMs, and generative AI tools.
Language AI
Amazon Transcribe
Automatically convert speech to text quickly and accurately with a feature-rich API for real-time and recorded audio and video content.
Language AI
Amazon Polly
Turn text into lifelike speech, allowing you to create applications that talk, and build entirely new categories of speech-enabled products
Augmented analysis
Amazon Textract
Automatically extracts text, handwriting, and data from scanned documents
Computer vision
Amazon Rekognition
Pre-trained and customizable computer vision (CV) capabilities to extract information and insights from your images and videos

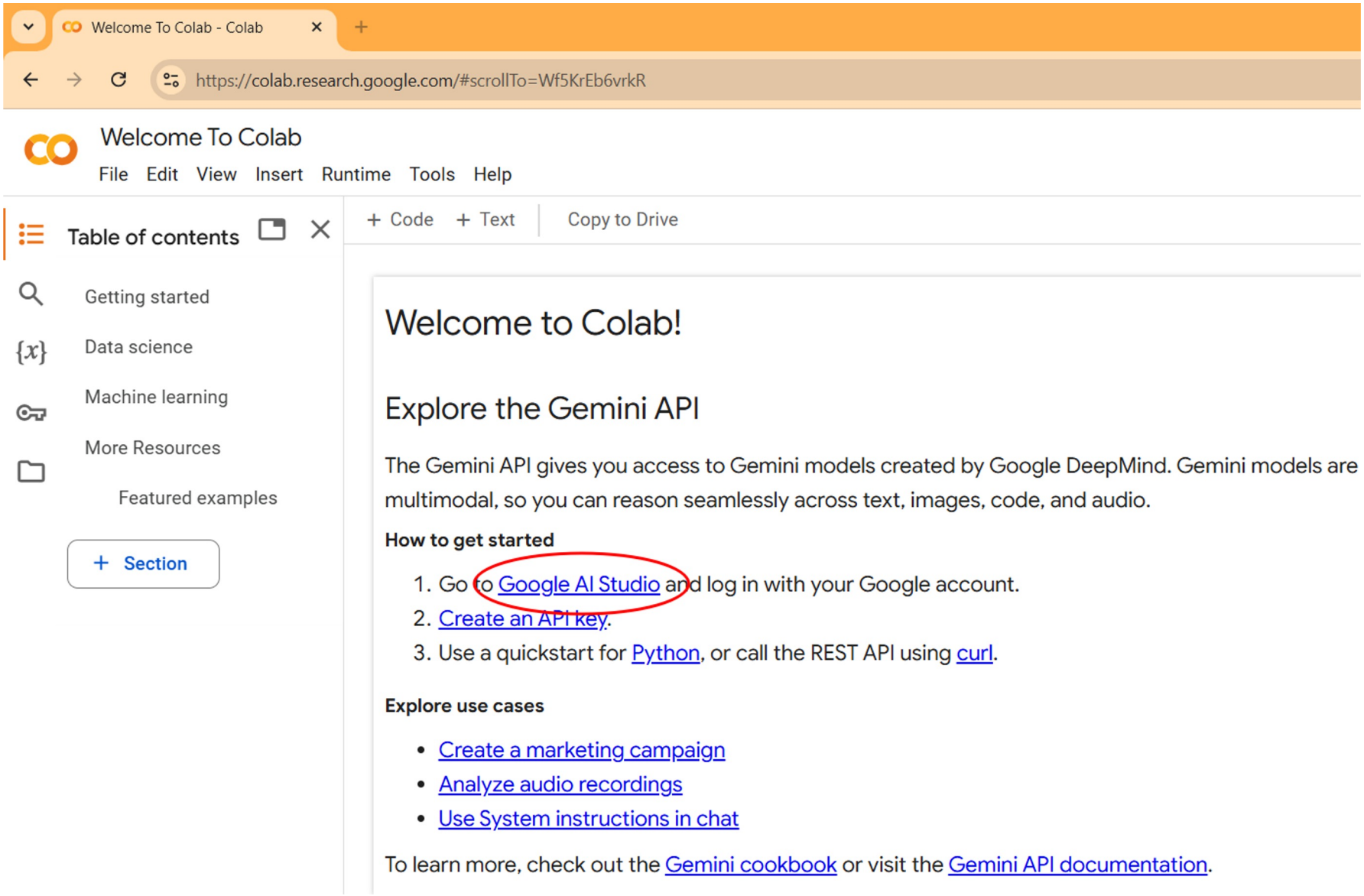

Welcome To Colab - Colab
Unified data, analytics and AI
https://aws.amazon.com/sagemaker/?gclid=Cj0KCQiAsOq6BhDuARIsAGQ4-zhDpB47_okQsZ_BGw9HQyp-gd7AKiQxR2Hfo00TDXPub_ix13
aws
About AWS
Contact Us
Support
re:Invent
Products
Solutions
Pricing
Documentation
Learn
Partner Network
AWS Marketplace
Customer Enablement
Amazon SageMaker
Overview
Capabilities
Pricing
FAQs
Build, train, and deploy ML models with SageMaker AI (formerly Amazon SageMaker)
Amazon SageMaker
The next generation of Amazon SageMaker is the center for all your data, analytics, and AI
Get started with Amazon SageMaker

End AWS AI

Start GCP AI

Google Cloud…




Welcome To Colab - Colab
https://colab.research.google.com/#scrollTo=Wf5KrEb6vrkR
Welcome To Colab
File Edit View Insert Runtime Tools Help
Table of contents
Getting started
Data science
Machine learning
More Resources
Featured examples
+ Section
+ Code + Text Copy to Drive
Welcome to Colab!
Explore the Gemini API
The Gemini API gives you access to Gemini models created by Google DeepMind. Gemini models are multimodal, so you can reason seamlessly across text, images, code, and audio.
How to get started
1. Go to Google AI Studio and log in with your Google account.
2. Create an API key.
3. Use a quickstart for Python, or call the REST API using curl.
Explore use cases
• Create a marketing campaign
• Analyze audio recordings
• Use System instructions in chat
To learn more, check out the Gemini cookbook or visit the Gemini API documentation.

End of slide deck