Introduction to *Cloud Computing*

Naomi Alterman
Neurohackademy 2024

UNIVERSITY of WASHINGTON

eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

Hey

- I’m Naomi
  - I’m a Technical Education Specialist and Data Science Fellow at the eScience Institute
  - Your resident computational Miss Frizzle :)

- What do I do?
  - By training, I’m an electrical engineer and computer networking researcher
  - By practice, I’m a facilitator of information flow between complex systems

Today

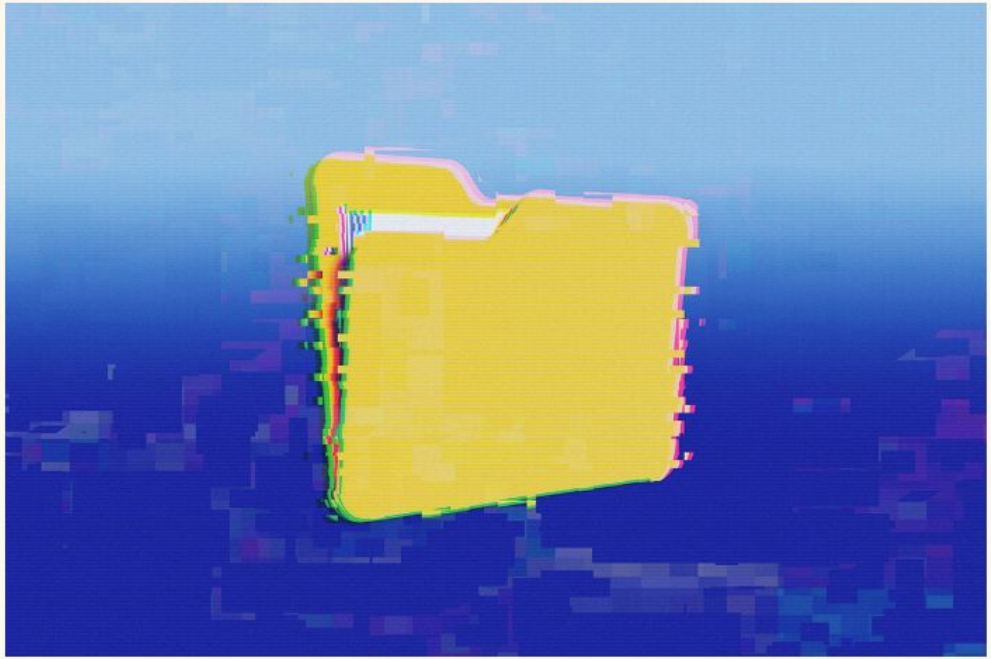
- Talk about what cloud computing is and the abstractions we use for it
- Tour common workflows for using cloud resources
- Discuss meta-skills for learning to use the cloud effectively

But first...

...what does “cloud” actually mean?

Respond in #alterman-cloud **in 3-2-1**

Context



Costly concerns: The shift could lead to prohibitive expenses for computation-heavy studies, some researchers say.
FLAVIO COELHO / GETTY IMAGES

NEWS / **COMMUNITY**

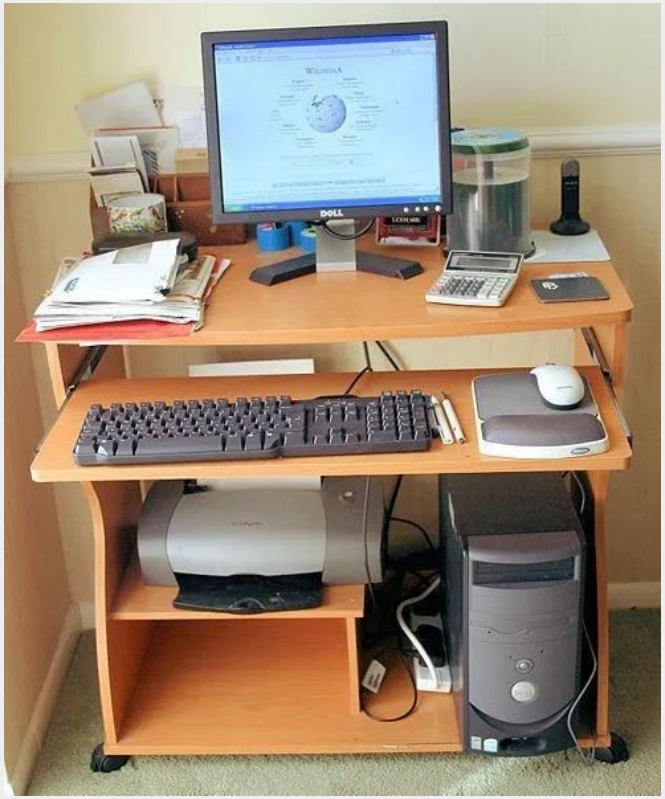
Data access changes to UK Biobank stir unease in neuroscientists

"I feel a little bit in limbo," says neuroscientist Stephanie Noble, who has paused a study using Biobank data after the repository shifted from a data download to a cloud-only access model.

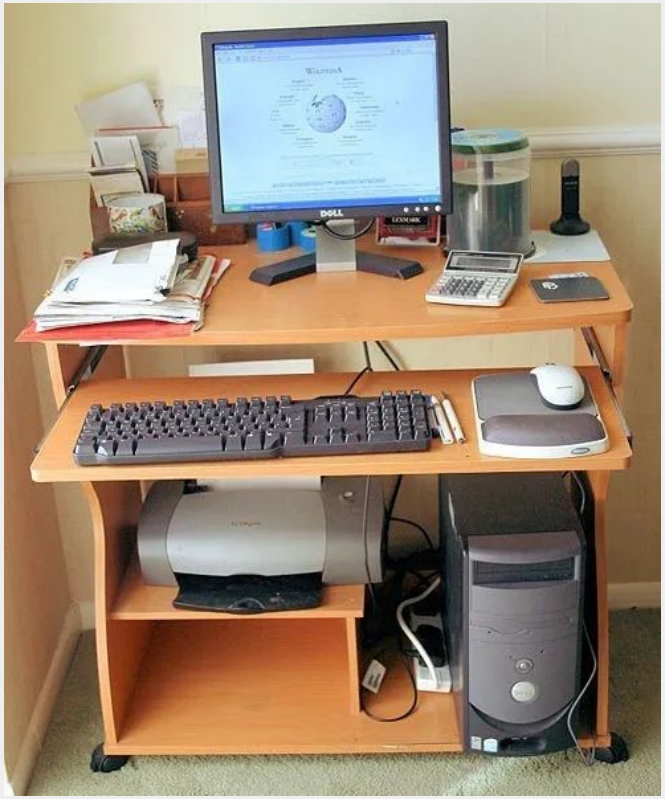
BY **CALLI MCMURRAY**
16 JULY 2024 | 7 MIN READ

https://doi.org/10.53053/HZSR1572

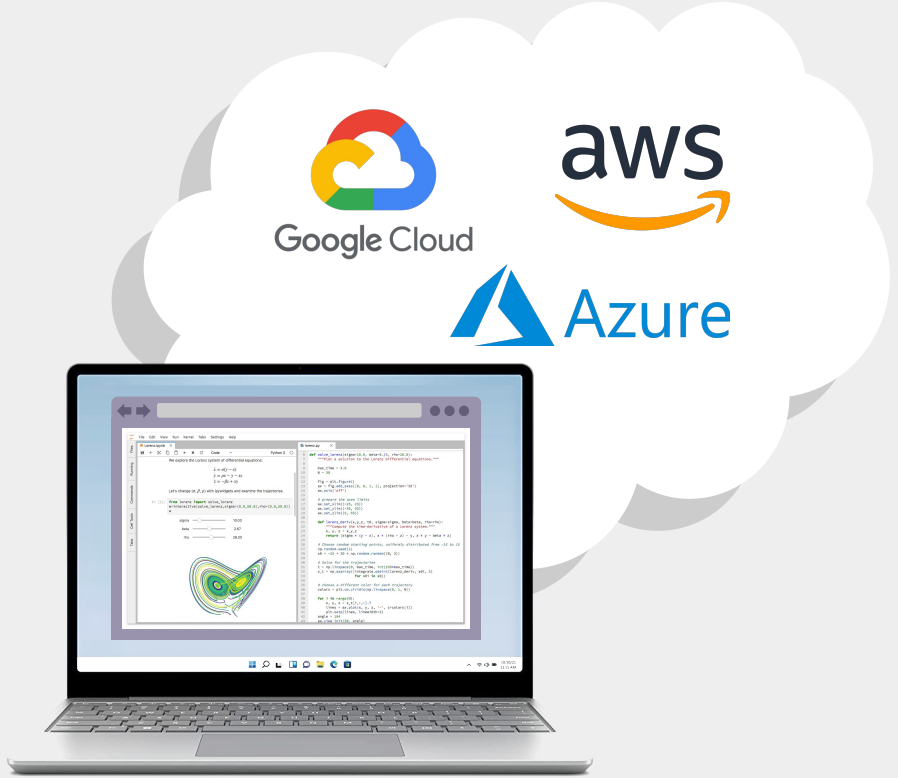
The swinging pendulum



The swinging pendulum



The swinging pendulum

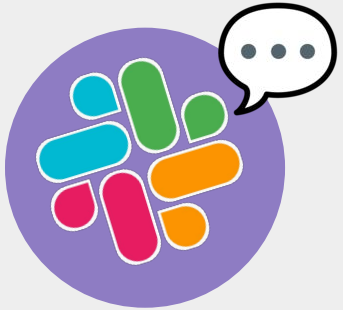


Brainstorming

What are the **pros** of paying someone to manage your computers for you? What are the **cons**?

Respond in #alterman-cloud

until Naomi is satisfied😈



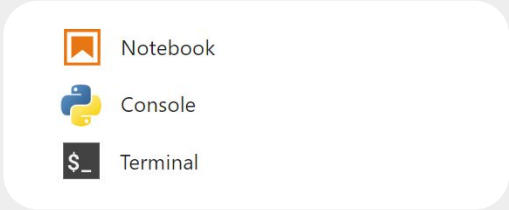
Brainstorming

How might **renting** compute (instead of buying it) **change** the way you ask **scientific questions**?

Respond in #alterman-cloud **in 3-2-1**

Building blocks

Building blocks

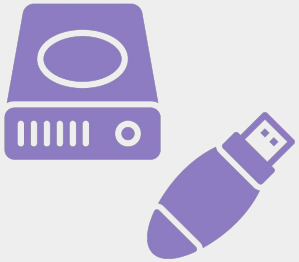


Cloud Abstractions

Compute

“Virtual machines” (VMs)

AWS: “EC2”
Azure: “Virtual Machines”



Storage

File stores, “Object storage”

AWS: “EBS”, “EFS”, “S3”
Azure: “File Stores”, “Blob Containers”

Networking

Virtual networks (vnets)

“____” as a service (_aaS)

Databases, Jupyter notebooks, etc

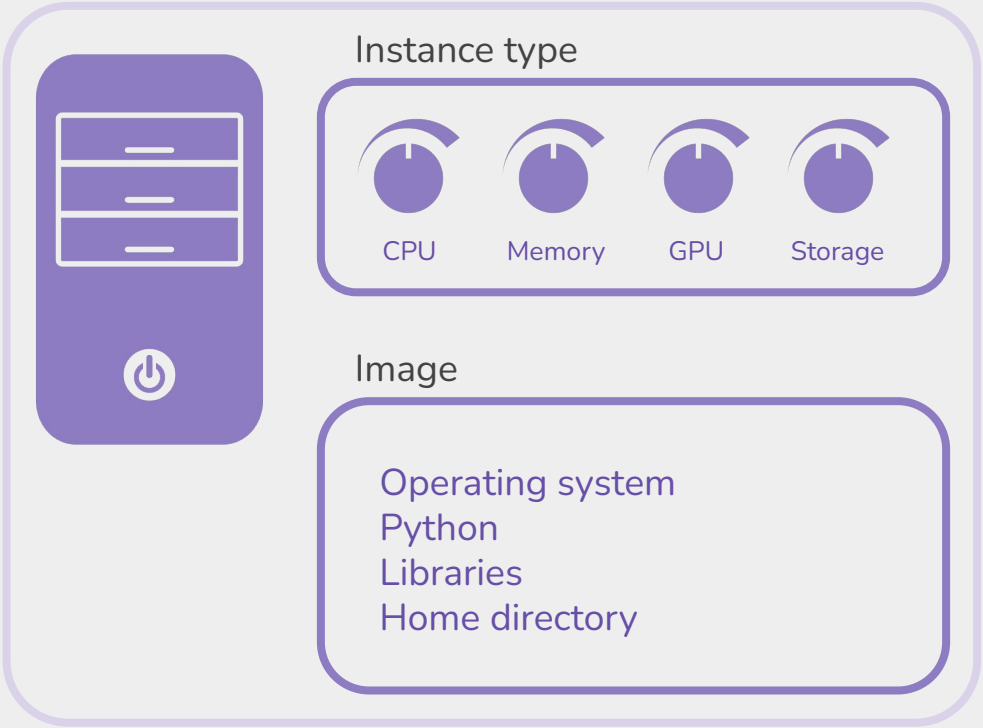
Users and security

“Identity Access Management” (IAM)

Location

Regions, zones

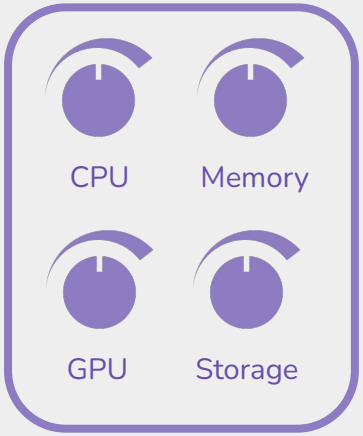
Virtual machines



- Fundamental cloud building block that “does work”
- Charged per hour of “on time”
- Easily scalable
  - Resizable
  - Duplicatable
- **“Cattle, not pets”**

VM Sizing

Instance type



Instance name	On-Demand hourly rate	vCPU	Memory	Storage	Network performance
m6idn.32xlarge	$10.18368	128	512 GiB	4 x 1900 SSD	200000 Megabit
m5.metal	$4.608	96	384 GiB	EBS Only	25 Gigabit
c6g.16xlarge	$2.176	64	128 GiB	EBS Only	25 Gigabit
c6g.metal	$2.176	64	128 GiB	EBS Only	25 Gigabit
p4d.24xlarge	$32.7726	96	1152 GiB	8 x 1000 SSD	400 Gigabit
g3.16xlarge	$4.56	64	488 GiB	EBS Only	20 Gigabit
x2gd.12xlarge	$4.008	48	768 GiB	2 x 1425 SSD	20 Gigabit
r5d.metal	$6.912	96	768 GiB	4 x 900 NVMe SSD	25 Gigabit
z1d.xlarge	$0.372	4	32 GiB	1 x 150 NVMe SSD	Up to 10 Gigabit

https://aws.amazon.com/ec2/pricing/on-demand/

VM types

	AWS	Azure	GCP
“Generic”	T, M	D	E, N
Compute	C	F	H, C
Memory	R, X	E	M
GPUs / Machine Learning	P, G	N	A, G

JUST TELL ME WHAT TO USE

😭😭😭

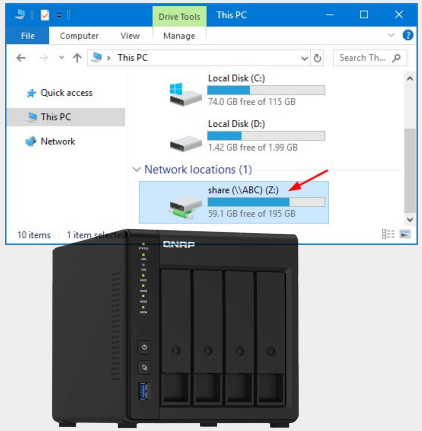
AWS:
t3.large

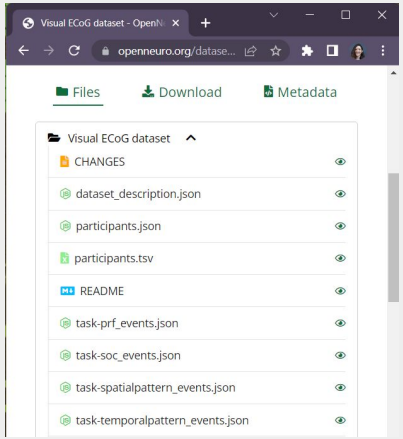
Azure:
D2 v5

GCP:
n2-standard-2

~$2-3 / day

Storage





Block Storage

Fastest, usually one per-VM,
Usually 4 GB - 512 GB
$$$ ~$0.16 / GB

AWS: EBS
Azure: “Disk Storage”

File Storage

Most flexible
< 100 TB
$$ ~$0.06 / GB

AWS: EFS
Azure: “File store”

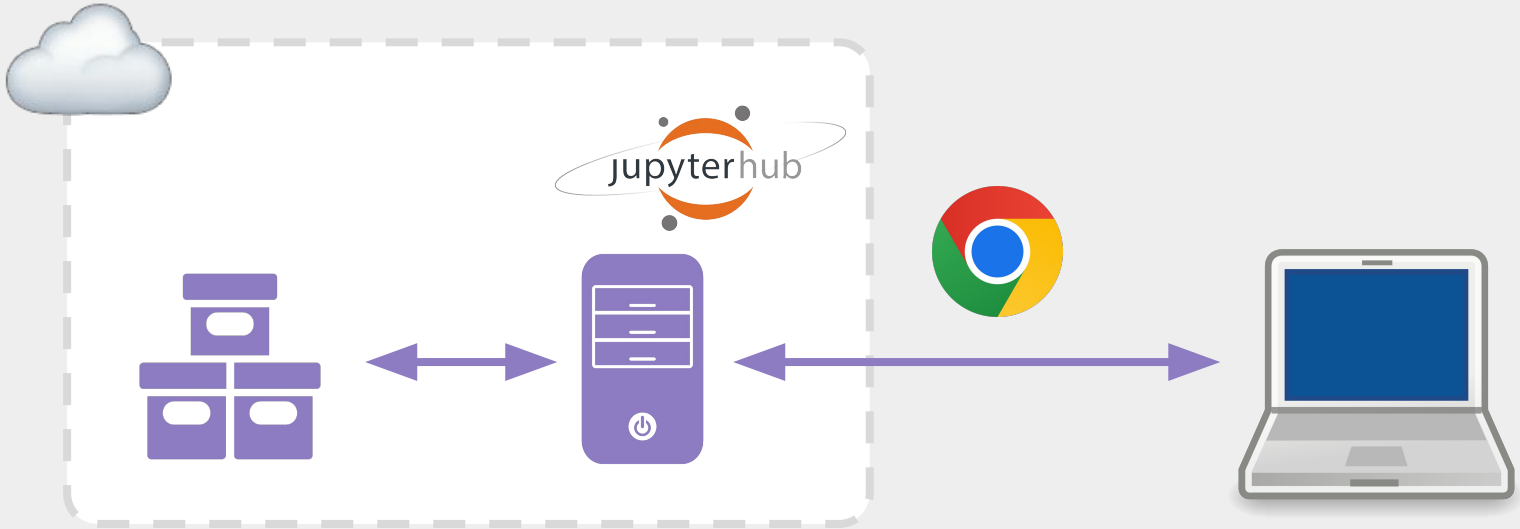
Object Storage

Most cost-effective
Unlimited
$ ~$0.01 / GB

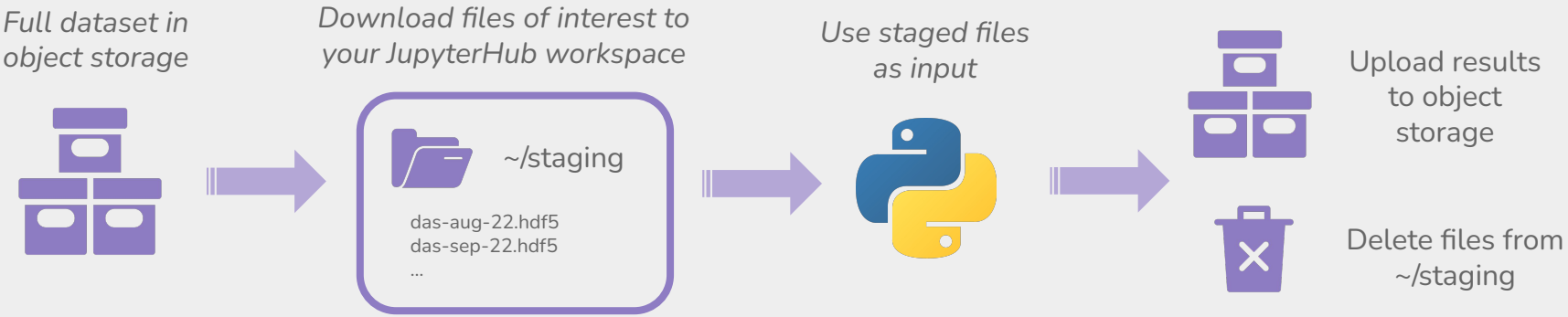
AWS: S3
Azure: “Blob container”

Workflows

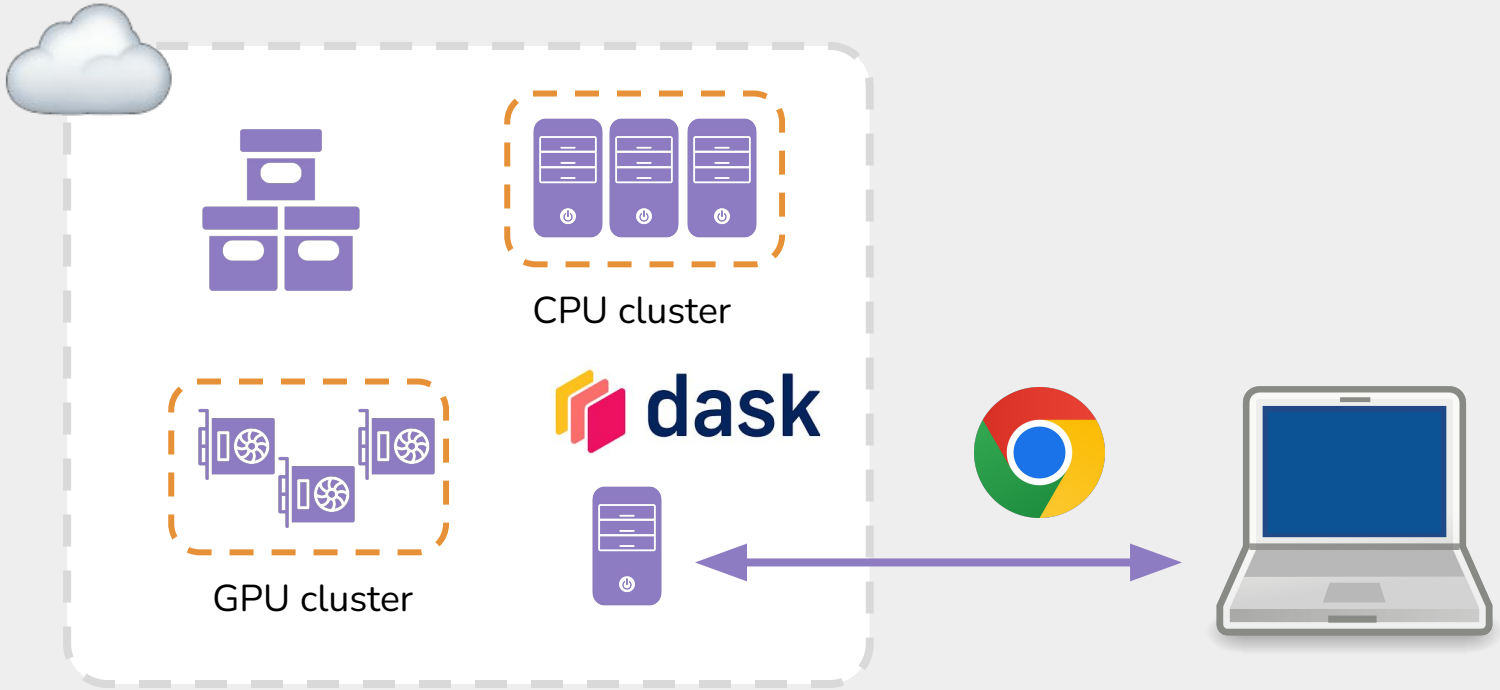
Interactive data exploration



Day in the life of object storage



Data pipeline



Organizational structures

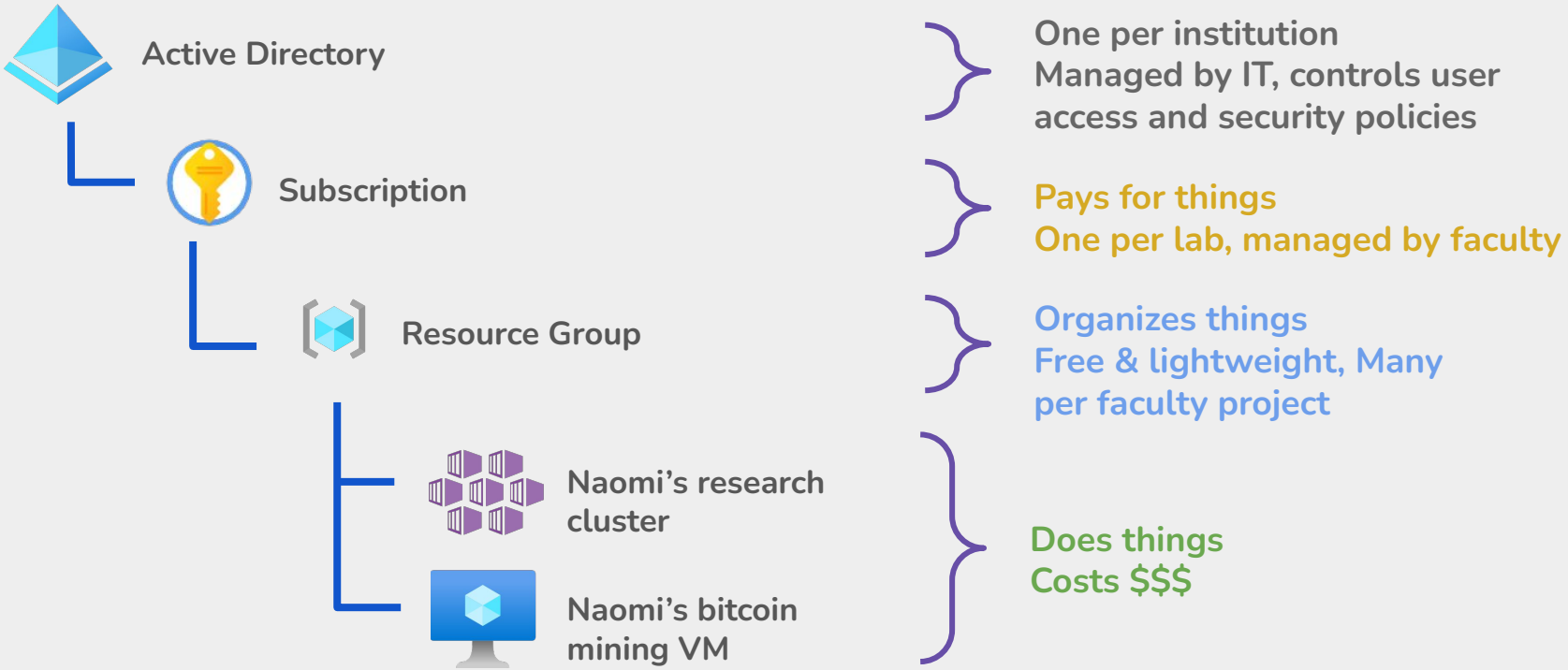
Prompt

Think about all your *stuff* (clothes, books, toothbrush, etc..)

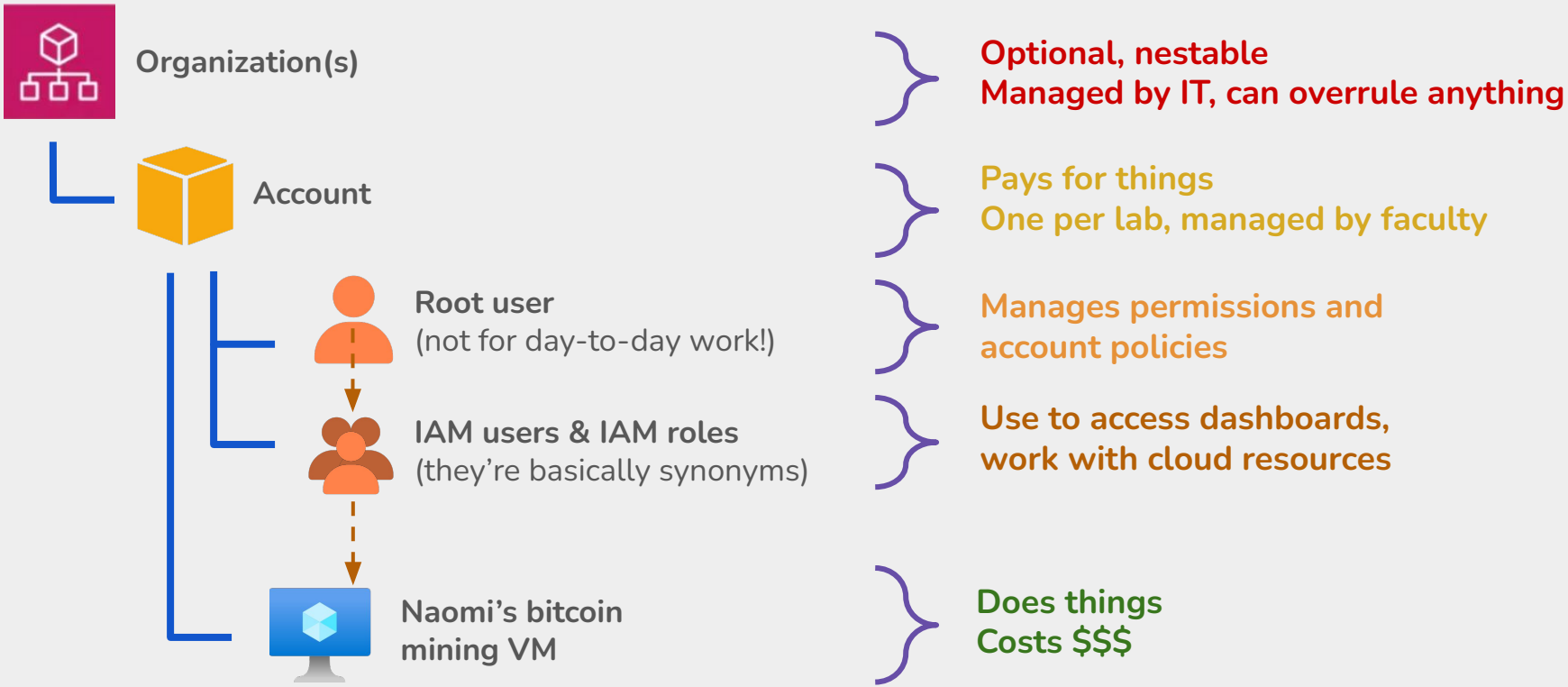
How do you normally organize it?
How about when you're travelling?

Respond in #alterman-cloud
in 3-2-1

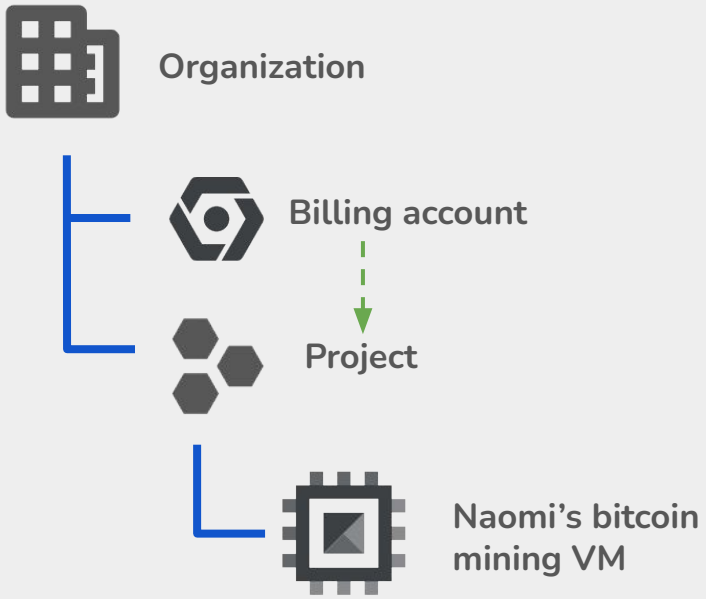
Azure-specific Structures & Jargon



AWS-specific Structures & Jargon



GCP-specific Structures & Jargon



Optional
Managed by IT, can overrule anything

Pays for things
One per lab, managed by faculty

Grouping cloud resources together for billing and sharing purposes. Generally one per lab.

Does things
Costs $$$

Cloud Access

- **Web Consoles**
  GUI in your web browser

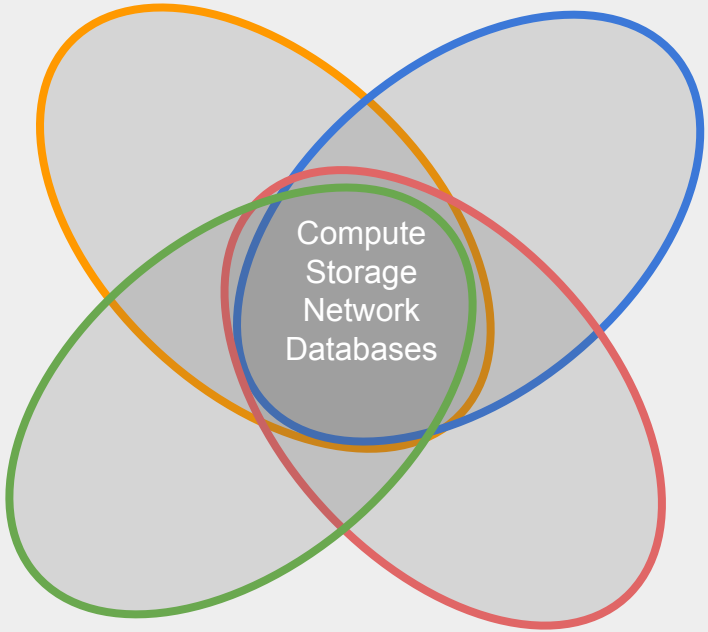
- **Command Line Interfaces (CLIs)**
  Commands from a terminal on your laptop

- **Software Development Kits (SDKs)**
  Libraries for Python / R / Julia / Matlab scripts

- **“Infrastructure-as-Code” (IaC)**
  Custom programming languages for describing your cloud setup)

“But which platform should I use?”

- Where they’re the same:
  - Cost and technical capability
- Where they differ:
  - Which workflows are a smooth polished experience, and which take extra work
- Real differentiators:
  - Funding opportunities
  - Where your colleagues have experience
  - Dataset locality



Some meta-discussion about Learning To Cloud

Prompt

What is something taught in Neurosci 101 that you think they should leave out?

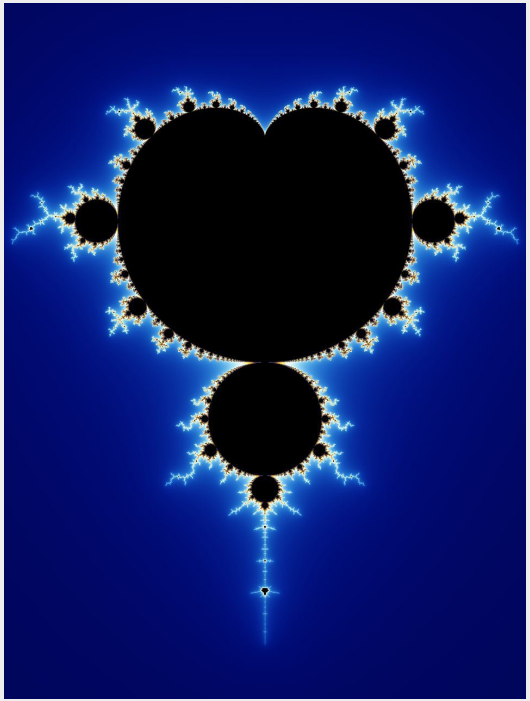
What about something you think they should include?

Respond in #alterman-cloud

in 3-2-1

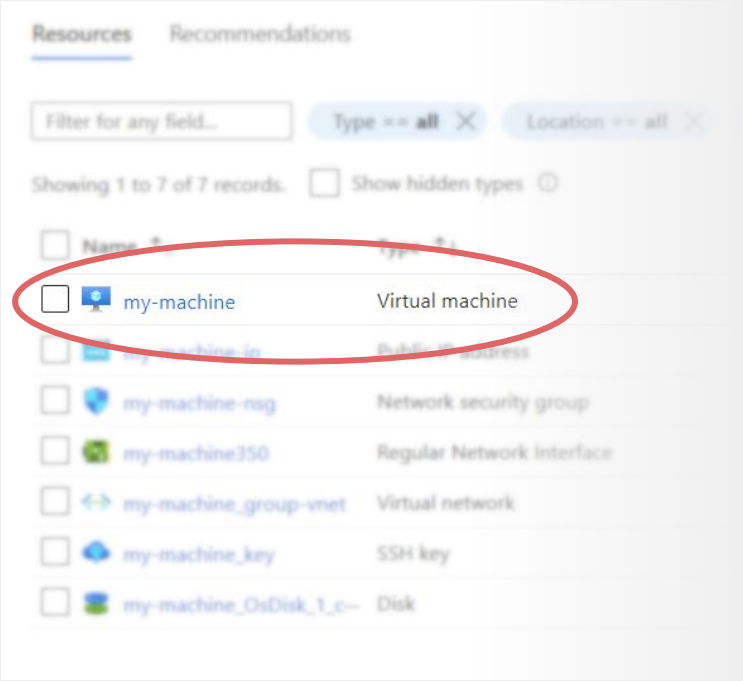
Don’t get discouraged by the rabbit hole

- There is effectively an infinite amount to learn about using the cloud
- Keep your focus on the *science* you want to do
  - Learn the minimum subset of sysadmin skills necessary to do your work
  - Incrementally dive deeper into inner workings as you optimize your workflow



Ignore most of what you see

- These aren't tax forms, not every field needs your attention
  - Assume things are "set up to work", until they don't
- Understand what gets billed and when
- Understand what actually "does the work"



Reduce head banging

- Find your “cloud TAs”
  - Identify colleagues who have done similar work and are willing to answer questions
- Time-limit your attempts and be ready to pivot
  - There are lots of ways to do computation in the cloud. If a guide turns out to be more trouble than it’s worth, try something else

What questions do you have?

Come up with one, even if you hadn’t already ;)

Send it to #alterman-cloud in 3-2-1

Stay in touch

naomila@uw.edu

;)

Profiling workloads

VM Sizing

- Bigger ≠ Better
- Broadly, computers are either **memory optimized** or **compute optimized**
    - Which of these two resources limit your work?
    - How *much* of that limiting resource does it need?
    - Answering these questions is called **profiling** your workload

Basic profiling

- Linux commands

```
top
```

-> live list of applications sorted by CPU usage

```
ps aux
```

-> cpu+memory usage of all currently running applications

- Jupyter / Python
  (put these at the top of individual notebook cells)

```
%%timeit
```

-> time how long a cell takes to run

```
%%memit
```

-> measure how much RAM a cell uses when running

(you need to first run `%load_ext memory_profiler` in an earlier cell)

Spot instances

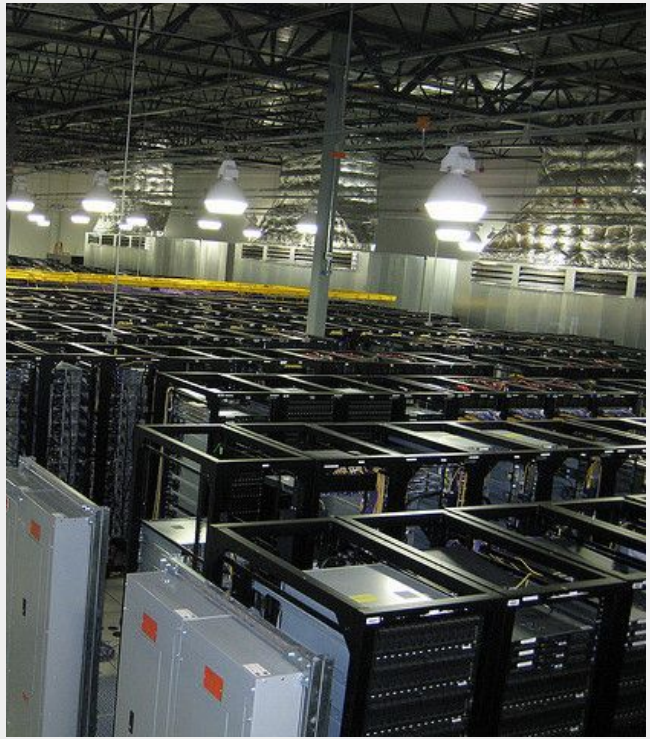
Prompt

When you're really focused on work, how do you recover from interruptions?

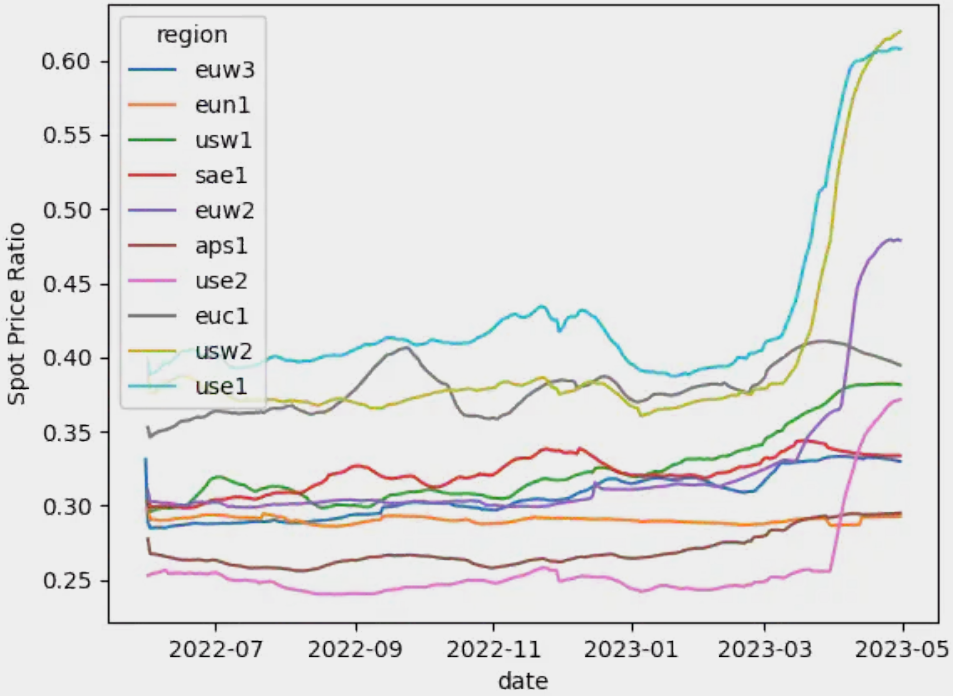
Respond in #alterman-cloud **in 3-2-1**

Preemptible / Spot instances

- Our machines might be virtual, but the data center is very very real
  - At any given moment, some % of a data center will be sitting unused
- We can host our VMs on underutilized server racks for a discounted rate
- **The catch:**
  - **We have to bid on the rate**
  - **If someone outbids us, or there isn't enough capacity, our VM gets evicted**



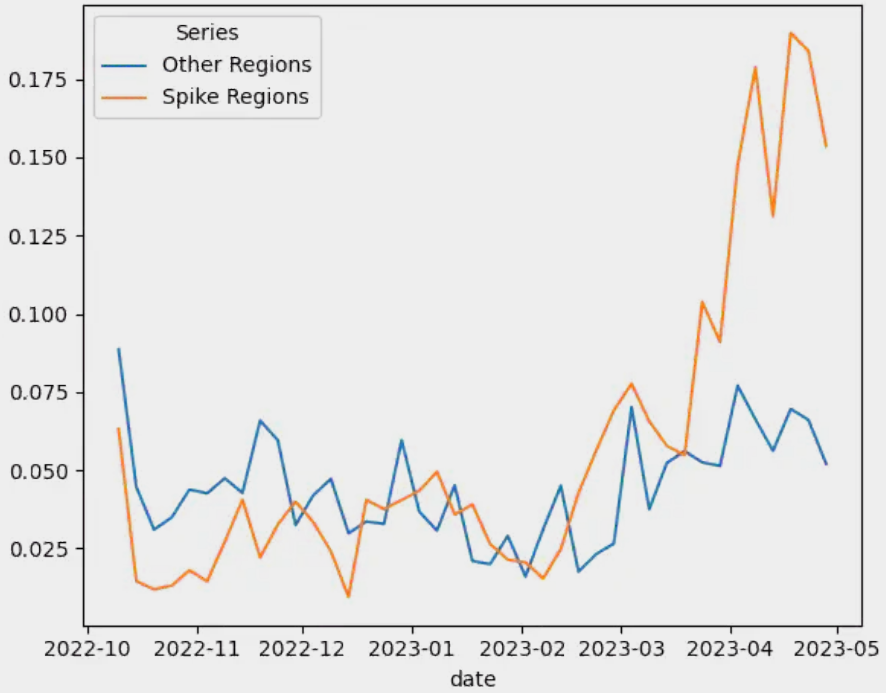
Spot pricing (AWS)



https://pauley.me/post/2023/spot-price-trends/

Spot eviction (AWS)

% of 5 million machines that were evicted within 10 minutes of boot



https://pauley.me/post/2023/spot-price-trends/

Spot instances

- Practically, how can we make use of these?
- One way: A cloud HPC manager like Open OnDemand (https://openondemand.org/)
- Another way: Distributed computing libraries like:
  - Skypilot (https://skypilot.readthedocs.io/en/latest/index.html)
  - Dask (https://www.dask.org/)

Costing

Official cost calculators

- Can create budgets and share with perma-links

- **AWS:** https://calculator.aws/
  - Example budget
- **Azure:** https://azure.microsoft.com/en-us/pricing/calculator/
  - Example budget

Budgeting process

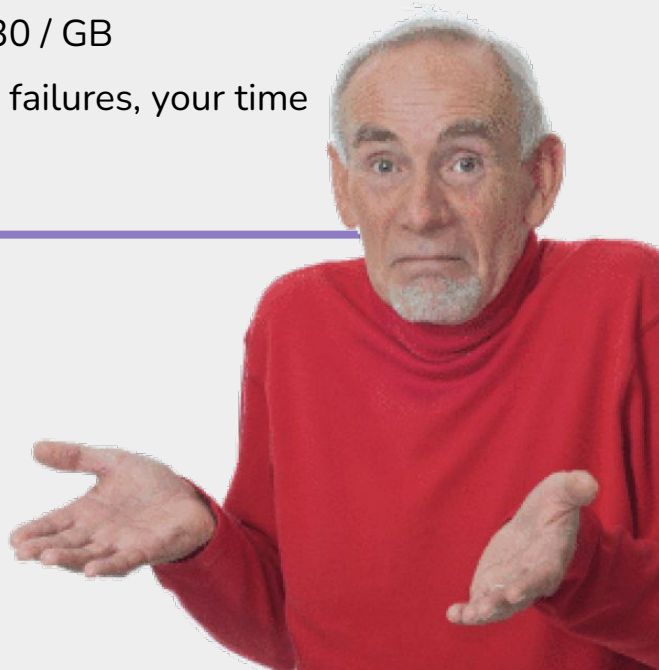
- It can be really hard to predict cloud costs
- How we do it:
  - Set aside $500 for a “tire kicking fund”
  - Use said fund to run a sample workflow
  - Retrospectively look at the bill
  - Translate the services you see getting billed over to the official cost calculator
  - Use the calculator to extrapolate the cost for your full project

Storage costs

$0.010 / GB x 3 disks (RAID 5) = $0.030 / GB

+ power, disk failures, your time

$0.023 / GB / month



Compute costs

16 cores, 2 GPUs, 128 GB ram

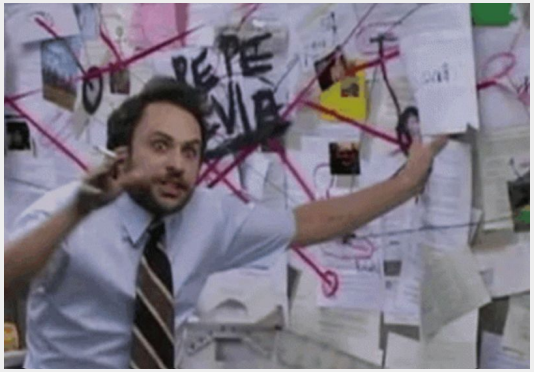
$8924.04

$1.74 / hr

Compute-optimized:
$1.15 / hr

Memory-optimized:
$0.67 / hr

VM Usage estimation



Interactive workstations need to be on for 40 hr/week

HPC machines need to be on for 72 hours straight, twice a month

Servers need to be on 24/7

Reserve pricing and savings plans

- Normally we're charged for **on-demand** VM pricing
  - VM turns off, we stop paying

- If we commit to paying for a VM 24/7, we can get a discount
  - Commitments are usually for 1 or 3 years
  - Can be paid in part or upfront
  - Vary in savings based on machine type

- **⅔ rule**: If you're going to be running a VM for more than ⅔ of the hours in a given month, it might be cost effective to use a savings plan

Hands-on activities

Setting up a JupyterHub

- Students can get a $100 credit on Azure if they sign up with their school email address:
  - https://azure.microsoft.com/en-us/free/students

- Setting up a single-machine JupyterHub
  - https://cloudbank-project.github.io/tljh-tutorial/azure/