Cloud Clinic

23-JAN-2024: Preemptible VMs

Rob Fatland (rob5@uw.edu), Oorjit Chowdhary, Naomi Alterman

Overview

- This presentation touches upon…
  - research infrastructure on cloud
    - baseline terminology
    - stretching your cloud budget by means of preemptible instances
      - Claim: $10k buys $20k or more of compute
- Cloud use as a means to research is a time investment
- Overburden (here) refers to learning each cloud platform's idiosyncracies
- Least Action Principle before you build a cloud solution:
  - Check if what you want already exists as a cloud **Service**
  - New services appear all the time

- Jargon for this presentation
  - VMs, Containers, Serverless, Spot, Checkpointing, Endpoints

Abstract: In today's busy world we can lose track of small details that have a big impact. Suppose you have a cloud budget of $10,000 but your computations could be scaled up beyond that limitation to produce better results. What you need is access to immutable storage (easy), access to cheap preemptible cloud VM instances (easy) and a reliable method of checkpointing your progress (easy? hard?). This one-two-three punch means you can purchase $33,333 worth of cloud computing for a mere $10,000 and get better research results as a consequence. This cloud clinic will catch you up on the how-tos and other small details of such a substantial gain in compute power. We use a CNN as our example implementation of a compute-intensive research task.

Introduction

- Research context we say there are four roles
  - Principle Investigator
  - Administrator$_{\$}$
  - Builders (who may face perceptual challenges: Cloud looks like a lot of work!)
  - Users
- We will work from an initial case study
  - Train a three-layer convolutional neural network (CNN)
    - Job: Classify10,000 small images into 10 categories (‘dog’, ‘horse’ etc)
    - The dataset and task are commonly referred to as ‘CIFAR-10’
    - Training takes 10-20 minutes on a typical workstation
    - We choose to iterate for 10 iterations or epochs
      - After each we save the resulting CNN weights: Checkpointing
- The objective is to maximize cloud value as tasks-per-$
  - Base rate: Suppose we have 10,000 tasks and a budget $10k
    - The value ratio is 1
    - If instead we can do 30,000 tasks for the same budget: v = 3
    - We do this by reducing the utility rate using preemptible instances
    - These are available as excess compute capacity, a “Spot market”

Exploring different ML tasks on Preemptible VMs

https://github.com/oorjitchowdhary/ml-training-preemptible-vms

The key program is `index.py`

- CIFAR-10: simple image classification
- ImageNet: Likewise at 100x scale
- bert: sentiment analysis

Overburden problem: Suppose these should run on any cloud…

- SkyPilot attempts to submerge cloud access under an abstraction layer
- Containers are overburden-independent

Containers

```
-v [host_path]:[container_path]
```

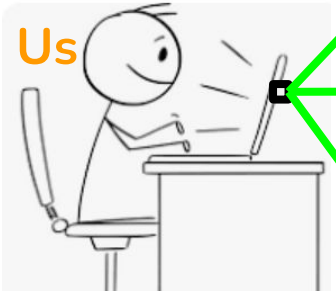
- Containers are hermetic, hosted Linux environments
  - Two popular brands: Docker and Apptainer
- Highly portable: Can run on any container-capable VM, any cloud

The Org Chart

Funding Agencies

HPC Facilities

Associated Programs



Research team
(4 roles)

CloudBank

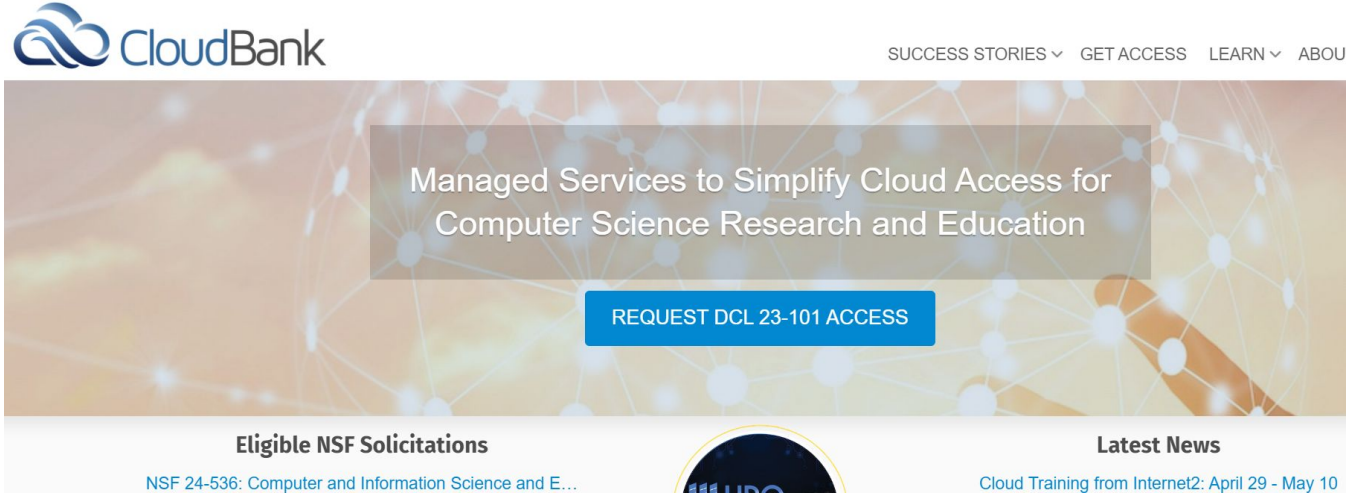
Associated Marketplaces

Third Party Technologies

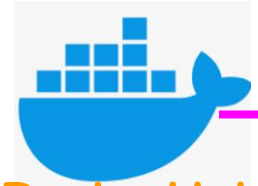
Cloud Platforms

CloudBank Support Framework

- Portal https://cloudbank.org
- Learning https://cloudbank-project.github.io/cb-resources/
- Community https://community.cloudbank.org/
- Sky Pilot study https://github.com/oorjitchowdhary/cifar-on-spot-vm



The Cloud-for-Research Ecosystem

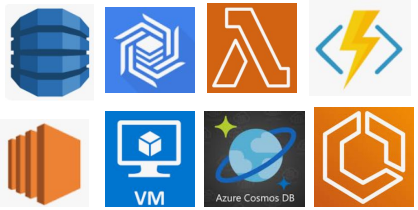


DockerHub

GitHub

The Internet

Spot Market



The Cloud

Object Storage
"Bucket", "Blob"
Capacity: Infinite
Latency: Low
URL: Yes if desired
Persistent: Yes

Preemptible Instances and Checkpointing

Cloud platforms have vast unused capacity: Computers with nothing to do. In response each cloud defines a resource pool of compute instances available at reduced rates, usually at a 50–90% *discount*. This is a “preemptible spot market”.

Spot market VMs come with a catch: Machines can be preempted (taken away) on very short notice, from 30 seconds to 2 minutes.

Enter **Checkpointing**: Storing the state of a computation so it can be interrupted and resumed at a later time from where it left off.

The task, then, is to implement checkpointing from preemptible instances

End result: $10,000 purchases $20,000 or more of compute power

This Go-Round on Preemptible Cloud VM Savings

- CloudBank team working through preemptible instance use
- This is the first round of a series
  - Grand objective: Refine methods for the research team
  - Today’s objective: Introduce terms and the central objective
- Three checkpointing modes
  - Bus mode: A preemptible instance is ‘hit by a bus’
  - Parachute mode: A pre-instance gets the signal and bails out
  - What-me-worry mode: AWS: Hibernation option

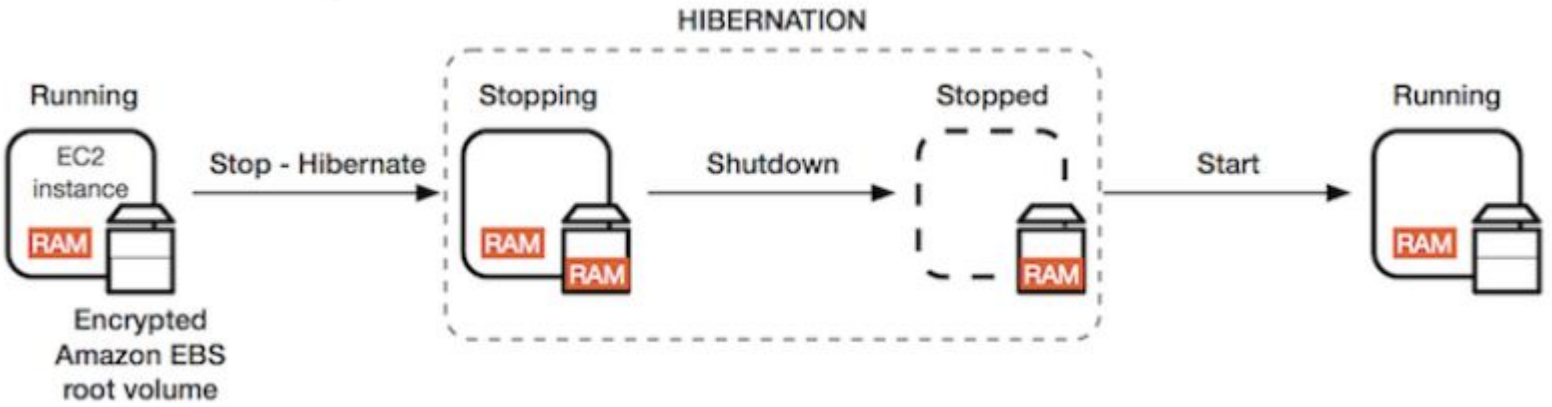
Eviction time

- Google cloud: System protocol: ACPI G2 Soft Off
  - You have 30 seconds…
- AWS*: JSON text to Event Bridge service. Recommend: Polling loop 5sec
  - You have 2 minutes…
- Azure**: Preempt signal to Schedule Events service. Polling loop 5 sec
  - You have 30 seconds… 29… 28…

* AWS features three eviction options: `Stop` or `Hibernate` or `Terminate`

**Azure features Eviction Policy: 2 options: `Stop/Deallocate` or `Delete`

AWS Hibernate advantage over Stop



- This is the (to be verified) What-me-worry? mode
- AWS will restart the VM where it left off when the price drops

From Eric Pauley: https://pauley.me/post/2022/spot-instance-pricing/

- The SPOT market (on say AWS) is not globally optimized
- VMs >> the VM you need can be available for less
- VM families, variations within each family
- From the AWS console for Spot instance selection:

Instance type requirements

Enter your compute requirements and let us choose optimal instance types to fulfill your Spot fle

Specify instance attributes that match your compute requirements

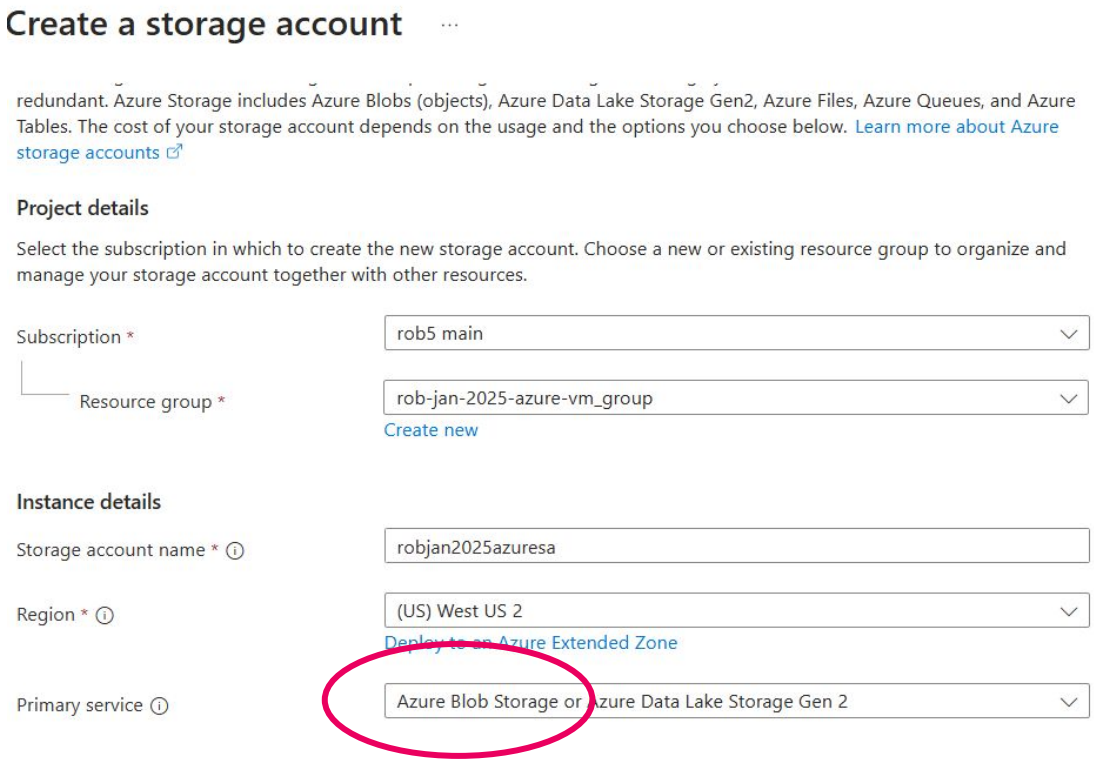
Instance type	vCPUs	Memory (GiB)
c5.18xlarge	72	144.00
c5.24xlarge	96	192.00
c5a.16xlarge	64	128.00
c5a.24xlarge	96	192.00
c5ad.16xlarge	64	128.00
c5ad.24xlarge	96	192.00
c5d.18xlarge	72	144.00
c5d.24xlarge	96	192.00
c5n.18xlarge	72	192.00
c6a.16xlarge	64	128.00

Preemption and Checkpointing Path Build

- To build this path we need…
  - …a fairly fast 10-iteration (epoch) CNN task: CIFAR-10
  - …a VM image on one or more clouds
    - …pre-configured to run the task
      - …including a run-at-start script
    - …adjacent idea: Use a Container image
- Need a fake interrupt: `touch halt.txt`
- Need an external `{watcher/launcher}`
  - A lightweight always-on master process
  - Notices preemption
  - Re-start from last checkpoint

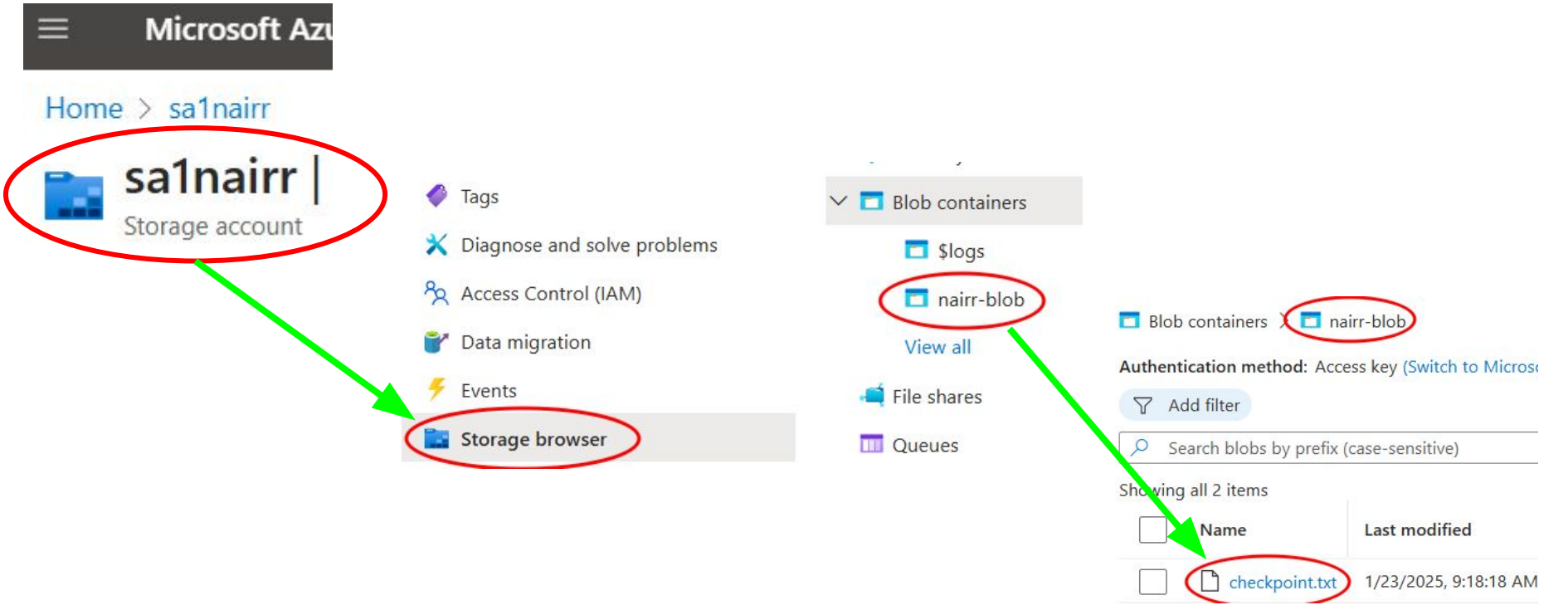
Azure storage account

A logical container for persistent storage: objects are “blobs”

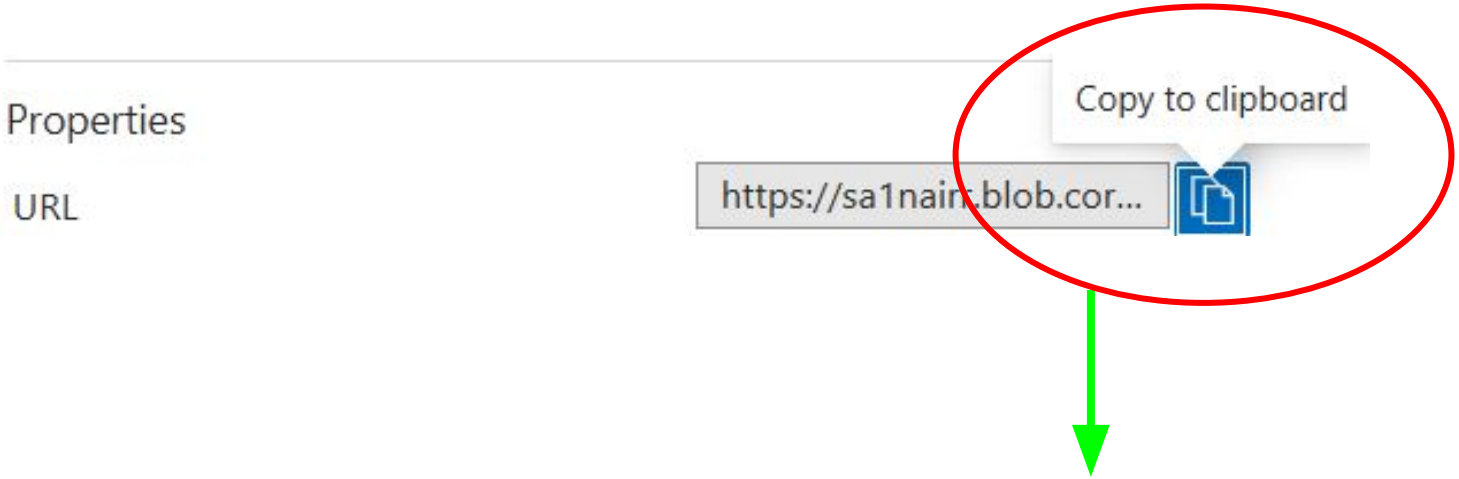


`azcopy` example of `bash` command line ifc to cloud

```
azcopy copy ./checkpoint.txt https://sa1nairr.blob.core.windows.net/nairr-blob/checkpoint.csv
```



`azcopy` continued: verify




```
>>> import pandas as pd
>>> d = pd.read_csv('https://sa1nairr.blob.core.windows.net/nairr-blob/checkpoint.txt')
>>> d
  cifar10progress   4
0          epochs  10
>>>
```

checkpointing key concepts (bus model)

- **`checkpoint.txt`** is written…
  - …by a script (for example Python + bash)
  - …running on a preemptible VM
  - …at regular intervals
    - (not responding to a 2-minute warning)
- **`checkpoint.txt`** was retrieved by laptop Python code
  - Authentication is the important detail
  - Each cloud has similar but unique approach to credentials
- To cash in on preemption: { Watcher ⇒ Launcher }

Automating console action as CLI code

This is a Google `gcloud` example

```
gcloud compute instances create my-useful-VM \
    --project=robs-project-123456 \
    --zone=us-west1-a \
    --machine-type=c3-standard-4 \
    --network-interface=etcetera \
    --maintenance-policy=MIGRATE \
    --provisioning-model=STANDARD \
    --service-account=12345678901234-compute@developer.gserviceaccount.com\

--scopes= … etcetera …

--create-disk=… etcetera …
```

Concluding…

- This presentation and deck will be available; and will update
- The goal here is efficient use of cloud
  - Ultimately CloudBank will demystify these topics
  - The Research team Builder will have a usable path forward
  - The Research team will get to “$30k of compute for $10k”
- Your contributions and / or questions: Contact us help@cloudbank.org

Thanks!

-The CloudBank team