CloudBank Onboarding Protocol

Appendix 1: Cloud Operations For Researchers

[CloudBank Onboarding Protocol (parent doc)](#)
[CloudBank Onboarding Protocol Appendix 2: Security and Cost Management](#)

Introduction

CloudBank provides a means for research teams to gain access to the public cloud for research computing and curriculum support. This includes cloud 'mechanics' (account management and access through the portal) as well as learning pathways and technical support. The [CloudBank Onboarding Protocol](#) document is the parent of this document. It is the starting point for understanding responsibilities, working securely and managing cost on the cloud. This document goes into further detail on certain aspects of moving research computing to the public cloud. Experienced cloud users may find this material extraneous; but newer cloud users may find it instructive.

The Inconvenience Principle

As a collection of learning curves and new technical details, learning to work on the cloud can be frustrating. Stumbling blocks and road blocks will crop up; and our perspective is: Overcome roadblocks quickly but also *properly* to avoid further problems down the road.

- Cloud adoption / migration is non-trivial for most research teams
- Challenges and points of inconvenience will arise
- CloudBank as a 'guide organization' is here to help you address these
    - Contact us when stuck (help@cloudbank.org)
    - Contact us when you solve a problem so we can document it
    - When possible: Write up detailed notes intended "for your team in 12 months"
    - Go to the [CloudBank community forum](#) and search for or create a Topic
    - Ask CloudBank for pointers to learning resources and/or technical consulting
        - The various cloud providers are eager to help you out

Obstacles and inconveniences are part of computing and cloud computing is no exception. The nuance here is that some obstacles have simple solutions that are straightforward to find with a browser search. Tackling bigger obstacles can sometimes really benefit from consulting with experts with a broad perspective. The cloud is really a collection of services, a complex of abstractions. As a researcher your time is valuable and we recognize you probably do not want to spend it sorting through this maze of abstract services; so that's what we are trying to avoid. At the same time: We often find that cloud services enable researchers to do more (particularly collaboratively) than they had realized was possible. It comes down to an optimization of your

time spent mastering this new technology against the benefits to your research, 'the important stuff'.

Cloud Building Framework

We present two ways of looking at building a research computing environment on the cloud. First we have "what the research team is looking for". The team may plan to build compute infrastructure from the Virtual Machine up; for example adapting a containerized workflow to a cloud VM cluster. We'll refer to this as Infrastructure as a Service or IaaS. Alternatively the team may intend to use "platform" tools such as managed databases, serverless computing or integrated templates. This is Platform as a Service or PaaS. Finally the team may have a turnkey solution in mind; so that the team members do comparatively little "on the cloud" and focus on simply using the built infrastructure to continue the research.

For the cases of IaaS and PaaS we can define some roles that help articulate typical cloud-based tasks.

- **Principle Investigator (PI)**: Takes on overall responsibility for the project
- **Administrator**: Covers identity and access management; monitors spend and security
- **Builder**: Creates cloud infrastructure; possibly including data publication
- **Researcher**: Focus on research; so needs little beyond an access mechanism

Admin / PI Skills

- Create a spend plan: Cloud resources used commensurately with the project budget
- Log in to the Cloudbank portal, activate cloud accounts
- Track spending through the portal and/or tracking tools such as Nutanix Beam

Admin Skills

- Create Identity and Access Management accounts for team members
- Ensure team members are briefed on best practices
- Track the allocated resources (often a *profusion* of resources)
- Ensure billing tracks spend

Admin / Builder Shared Skills

- Familiarity with cloud components and the utility approach to traditional computing
- Understand the distinction between block and object storage
- Understand the use of machine images in relation to Virtual Machines
- Log in to the browser-based portal
- Manage account description and basic information

- Working familiarity with the cloud jargon set: What is a VM called? What is object storage called?
- Look up resource costs, calculate cost of design choices
- Familiarity with cloud services above and beyond the basics: Jargon, purpose, possible use by the team

Builder Skills

- Locate a machine image and launch a Virtual Machine
- Create machine images from Virtual Machines
- Move data into and within the cloud (object storage, block storage)
- Configure Virtual Machines as working environments
- Identify, learn and use cloud services that simplify management tasks
- (Optional) Install and use the cloud command line utility
- (Optional) Install and use the cloud API
- (Optional) Install and manage a database on the cloud
- (Optional) Create data access mechanisms: Team and/or external use

Researcher Skills

- Data science, domain expertise
- Responsibilities may include developing software for large computational tasks
  - Should be aware that not all local code translates easily to the cloud
  - Working from small to large can help ensure success
  - May need to learn checkpointing to take advantage of preemptible instances

Period of Performance (PoP) Actions

'What you will be doing with your cloud account' amounts to adding team members, building a work framework, importing data, running analysis, and hopefully working towards publication of your research. As the PoP draws to a close you will also want to chart out an exit plan, possibly in relation to your Data Management Plan. Here are some of the key activities in bullet form.

- Add team members, aka Identity and Access Management or IAM
  - Educating team members on safe and proper use of credentials
- Creating daily Stops and Starts: Ensuring VMs are stopped when not in use
- Understanding machine images: Flexibility matching compute power to task
- Understanding cost drains: How unattended resources can reduce your cloud budget
- Build using gradualism: Take small incremental steps to develop confidence in your cloud infrastructure. If you plan to run a large compute task: Run smaller ones first to establish that they are doing what you expect.

○

CloudBank Consulting and Community
CloudBank is not a cloud engineering service. We can and regularly do consult with research teams on finding the best paths forward, often including cloud vendor architects.

- Case studies
- Experience
- Relationships with cloud providers
- Optimization: Yes we can help
- Feasibility: Yes we can helop
- Engineering: No; out of scope

Learning from cloud vendor walk-throughs and tutorials

Working with vendor-provided resources can be the most efficient way to get going.

- Learn terminology / jargon / concepts specific to that particular cloud
- Learn to navigate and make use of their browser interface (console / portal)
- Follow the logical layout of the cloud vendor's teaching resources
- Learn about choosing which services and resources will work best for you
    - ...and learn how to not pay for them when you are not using them

Open Questions

- How do I handle licensing of specialized software on the cloud? MATLAB for example.