MORE CLOUDY DAYS – AN UPDATE

Jeffrey Schiller
MIT App Inventor
Information Systems and Technology
ITPartners Conference – June 2019

INTRODUCTION

- Three years ago I talked about how we use Cloud Computing with MIT App Inventor
- Usage has roughly doubled since then
- We are shifting from using PaaS to IaaS
- That is part of what this talk is about
    - Including how we did a recent migration from one vendor to another while all services stayed operational throughout

SAAS: SOFTWARE AS A SERVICE

- Software provided over the Internet, typically via a web browser
- Examples: Gmail, Service Now, Concur, Salesforce, Google Apps

IAAS: INFRASTRUCTURE AS A SERVICE

- Virtual Servers “in the cloud”
- Typically Linux servers, though Windows servers can also be purchased
- You manage software, including the OS, vendor manages hardware
- Typically pay for bandwidth, wall clock time and storage

PAAS: PLATFORM AS A SERVICE

- Vendor provides a programming SDK. You write to the SDK and they handle everything else
- Almost all PaaS systems are proprietary, with vendor lockin

SAAS: THE GOOD, THE BAD

- Easy! The end-user service is offered.
  - Never have to perform a software upgrade
- Have zero control
  - System is updated and changed based on the whims of the vendor
  - (though often as a result of end-user feedback)

IAAS

- You don’t have to maintain hardware
- You **do** have to maintain the OS and all software
- Can allocate resources from anywhere (just use a browser!)
    - Better protect your credentials!
- Cost is usually easy to manage, you are charged for measurable usage

PAAS

- Vendor maintains the OS and platform
- **Vendor handles the scaling to handle load**
- You just code to the vendor supplied APIs
- Those API’s can (and do) change, you have to keep up
- Cost is harder to calculate. Chargers are often made for API Calls and operations that are hard to track, measure and control

THERE IS NO CLOUD

- It’s just someone else’s computer
    - And they make the rules
- And so is a social network

TERMS OF SERVICE

- Its not a legal system
- Commercial organizations most value expedience
    - Formal “fair” processes are expensive
    - No Upside for a commercial organization
- Larger providers are using automated tools
    - Which don’t always work well
    - Appeal processes are usually poor
        - If there are humans involved, they may be low level employees or dis-empowered employees

NEW PUBLIC DEMANDS

- With recent events putting the spotlight on bad behavior on display through social networks have caused a public outcry
    - Demands have included moderation of all uploaded video content
        - This is **not** physically possible
- No Good Answers here. We don't really want government censors, but private senors may be worse. But we really cannot have murder live-streamed.

SO WHAT IS THE MESSAGE?

- SaaS is risky
    - You are usually tied to one vendor
        - Their terms of service can be very one-sided
- If your business depends on a SaaS provider, you are quite literally at their mercy
- PaaS is almost as bad
    - Vendor lock-in. May even be worse then SaaS because you have investment in the particular PaaS vendor

ANSWER: IAAS (MAYBE)

- IaaS can be reasonable, if you are careful
- Use only “common” features available on multiple providers
    - Provider agility is key
- If you buy Linux servers from an IaaS provider, the “API” you have is effectively the x86 platform. It doesn’t change much, at particularly not at the whim of the vendor!

NEW APP INVENTOR ARCHITECTURE REQUIREMENTS

- Based on Linux Servers (IaaS)
- Needs to scale (horizontally)
- Minimize “hairy” components

NEW APP INVENTOR ARCHITECTURE

- Based on Linux servers running docker (swarm)
- Uses CEPH Storage system
- Still a Java servlet back-end, using jetty
- Minimal use of H2 database
  - not a critical component and all data can be recovered from the CEPH cloud
- Memcache for performance

WHY CEPH

- Provides a simple replicated horizontally key/value store
  - It provides other features, but we just use the key/value store the oldest and most stable layer
- CEPH is open source (ok, that's my bias, but it means I can debug CEPH based issues)
- But it is hairy…

WHY DOCKER SWARM

- MIT App Inventor runs in two types of containers
  - Containers running CEPH
    - These do not use Docker Swarm
  - Stateless containers that process transactions
- Swarm is a simple almost turn-key system

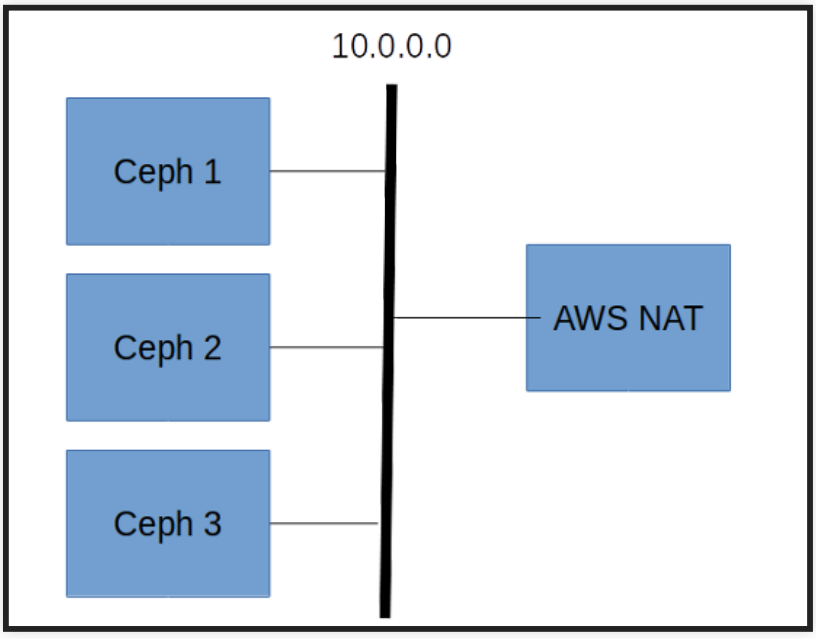
KUBERNETES, SWARM COMPEITOR

- Much more complicated, in particular when you want access to raw hardware for the storage service
- Different vendors offer slightly different flavors

SIMPLIFIED VIEW OF AWS BASED SYSTEM

- We have two smaller deployments of App Inventor using new architecture
  - One is in AWS us-east-1 Region
    - Used for “Hour of Code”
  - One is in AWS Singapore Region
    - Used for our collaboration with the Coolthink project in Hong Kong

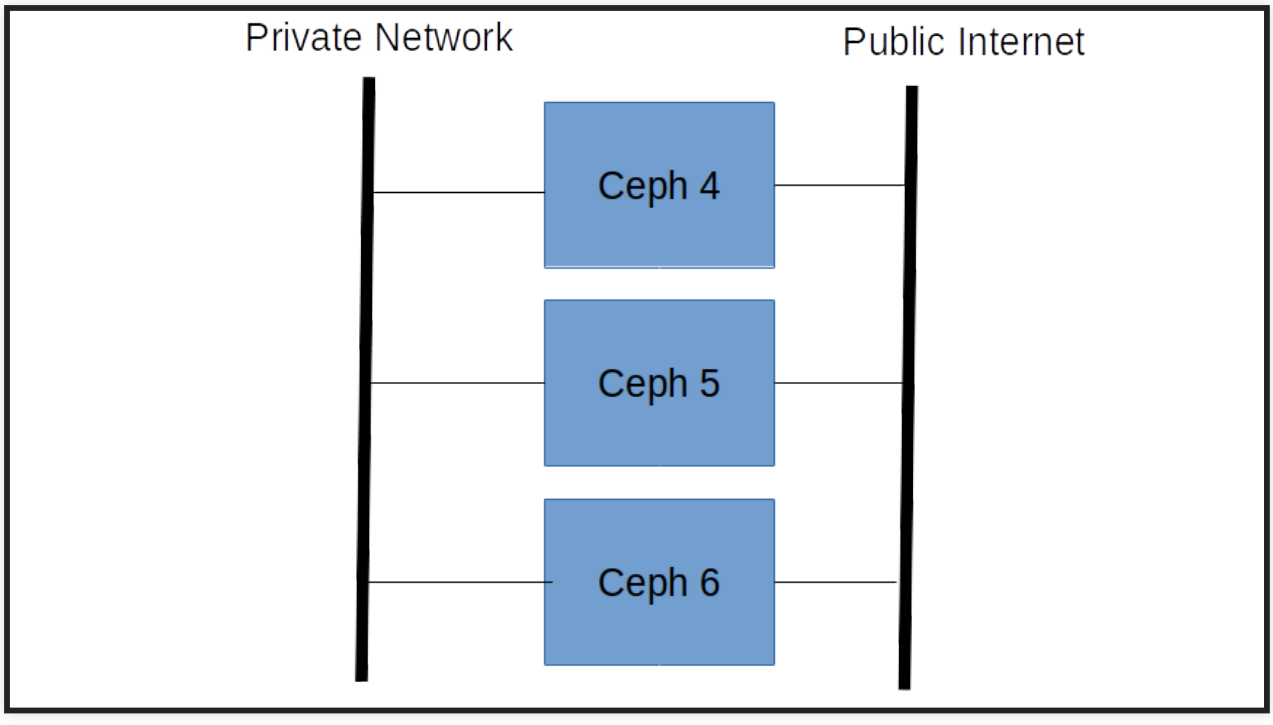
SIMPLIFIED VIEW OF AWS BASED SYSTEM



IBM CLOUD

- More traditional Data Center
  - VM's directly on public IP Addresses
  - Backend private network, but it doesn't span data centers

IBM CLOUD



TO MOVE TO THE IBM CLOUD

- We are going to move the “Hour of Code”
- Why the IBM Cloud?
    - We have free credits!
- Goal, to move service, complete with 100Gb of data without any downtime

THE MAGIC TOOL!

- TINC
    - An old Linux program, “tinc” lets you setup a mesh network so you can make an arbitrary network of hosts appear to be on the same network segment
        - In our case it is 192.168.55.0/24 and 10.0.0.0/16
- “There Is No Cabal” (tinc) reference to an old Usenet catchphrase

SO HOW TO MOVE

- Use TINC to setup a virtual network between AWS and IBM
- Setup new CEPH servers on the IBM hosts (using IBM SAN Storage)
- Setup a Docker Swarm network that has containers both in AWS land and in IBM
    - Connecting to a port on any Swarm hosts routes to those containers, a sort of built in load balancer

SO HOW TO MOVE

- Use CEPH tools to migrate data in the background to new CEPH servers, decommission AWS versions as they are emptied
- Point code.appinventor.mit.edu DNS to IBM side
- Shutdown AWS hosts
- Profit!

DID IT WORK?

- Yes. The whole process took a weekend, mostly because I throttled the migration traffic to avoid overloading the VMs.

QUESTIONS