TWO SIGMA

Cloud Computing

Transitioning to a Hybrid Cloud

Scott Rich
scott.rich@twosigma.com

November 2017

Agenda/Outline

Introduction

Definition of Hybrid Cloud

Shared responsibility model
- IaaS, PaaS, SaaS

Why Hybrid Cloud, why now?

Challenges and Solutions for Moving to Hybrid Cloud

Deep Dive: Our Hybrid Simulation Farm

Adding Cloud Native Services

Important Legal Information

The views expressed herein are not necessarily the views of Two Sigma Investments, LP or any of its affiliates (collectively, “Two Sigma”). The information presented herein is only for informational and educational purposes and is not an offer to sell or the solicitation of an offer to buy any securities or other instruments. Additionally, the information is not intended to provide, and should not be relied upon for investment, accounting, legal or tax advice. Two Sigma makes no representations, express or implied, regarding the accuracy or completeness of this information, and you accept all risks in relying on the above information for any purpose whatsoever.

About Me

- 27 years at IBM, in RTP, Rome, and Zurich, building developer tools and working on IBM's Bluemix Cloud
- Founding member of the IBM Eclipse team
- Started Two Sigma's Public Cloud effort in 2015
- From one pilot project to ~1 Data Center on Public Cloud in 2 years
- Currently Two Sigma's Head of Architecture

What is Hybrid Cloud?

From Forrester: "One or more public clouds connected to something in my data center. That thing could be a private cloud, that thing could just be traditional data center infrastructure."

Why not stay in safe data centers?

- Elasticity and scale
- Compelling Cloud Native services

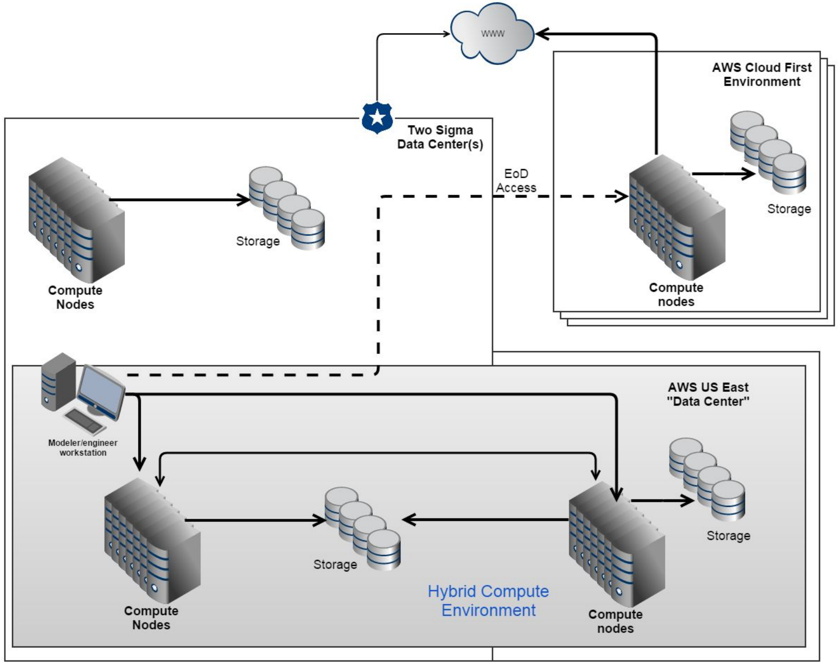
Why hybrid, why not all-in?

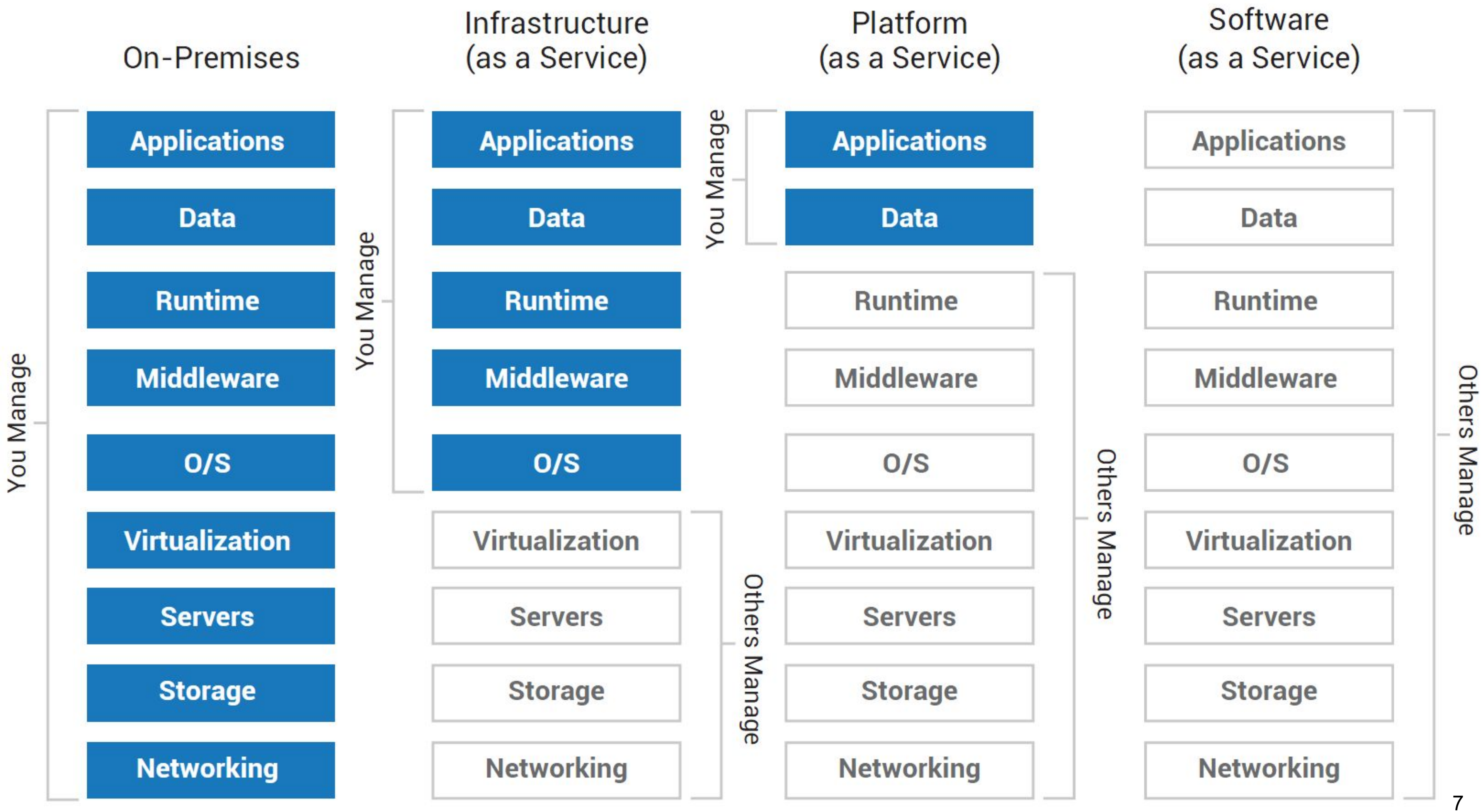
- Risk-aversion
- Existing assets, especially data, are hard to move overnight
- Cost

Ref: http://www.zdnet.com/article/hybrid-cloud-what-it-is-why-it-matters/

Hybrid Cloud Architecture

- Treat Public Cloud environments like Data Centers.
- Develop and operate a Hybrid model that spans our Data Centers and those operated by AWS, Google, Microsoft, etc.
- Extend the corporate Network/Perimeter into public clouds
- Progressively migrate applications to the Public Cloud on the basis of risk/complexity and value





On-Premises

You Manage

- Applications
- Data
- Runtime
- Middleware
- O/S
- Virtualization
- Servers
- Storage
- Networking

Infrastructure
(as a Service)

You Manage

- Applications
- Data
- Runtime
- Middleware
- O/S

Others Manage

- Virtualization
- Servers
- Storage
- Networking

Platform
(as a Service)

You Manage

- Applications
- Data

Others Manage

- Runtime
- Middleware
- O/S
- Virtualization
- Servers
- Storage
- Networking

Software
(as a Service)

Others Manage

- Applications
- Data
- Runtime
- Middleware
- O/S
- Virtualization
- Servers
- Storage
- Networking

I love this sticker by Chris Watterston

There is no cloud
it's just someone else's computer

About those computers...

It’s true, stuff happens. But they’ve got lots of computers. If one breaks, you can get another with an API call.

And if you don’t like this guy’s computer, it’s not that hard to move to someone else’s.

They’re really good at operating thousands of computers. Better than you.

They’re very likely better at securing a large network of computers than you are.

Hybrid Cloud Challenges

What to move?

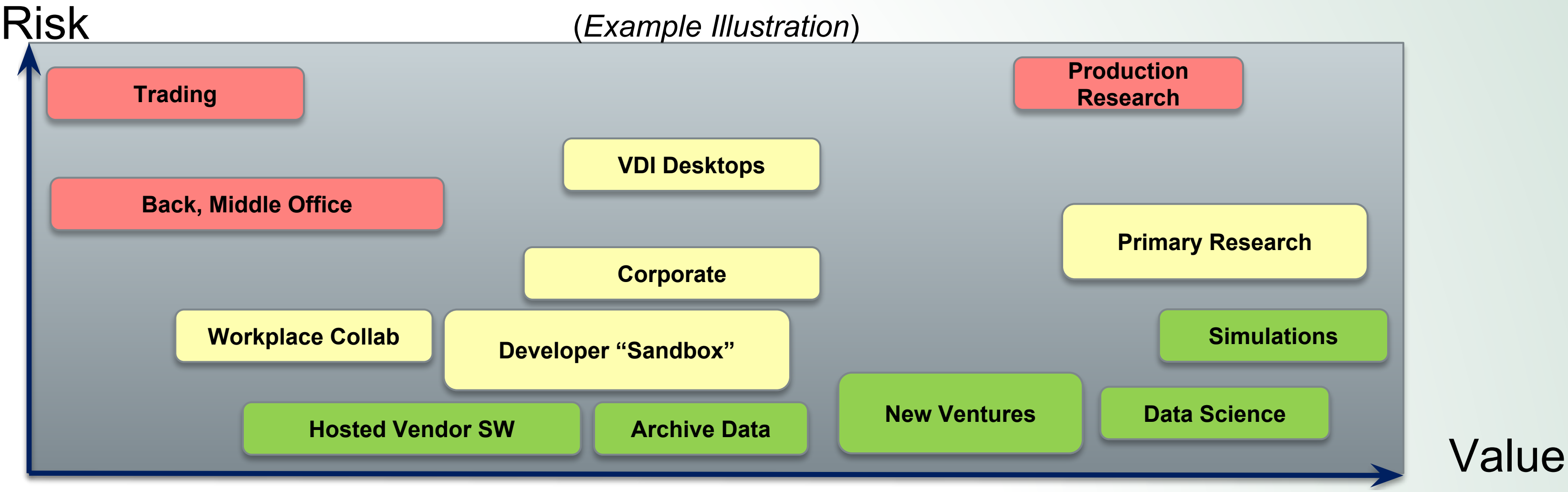
When?

How to protect data center assets while connecting to the Wild West of a Public Cloud?

Evaluating Public Cloud Workloads

Risk / Reward for Workloads

(Example Illustration)



How we Characterize “Value”

- Time to market
- Do new stuff
- Do stuff faster
- Do more stuff
- Stuff is cheaper
- Modeler/Developer Happiness

How we Characterize “Risk”

- Security
- Complexity
- Manageability
- Change / Workflow
- Cost
- Migration

Key Considerations

- What questions need to be answered?
- What dependencies need to be solved?
- What are the operating model considerations?
- What decision may be deferred to later stages?

Sidebar on Cost: How you buy in the Cloud

Most Cloud services are “pay for what you use” based on various usage metrics, e.g. machine hours, storage TB hours, TB queried

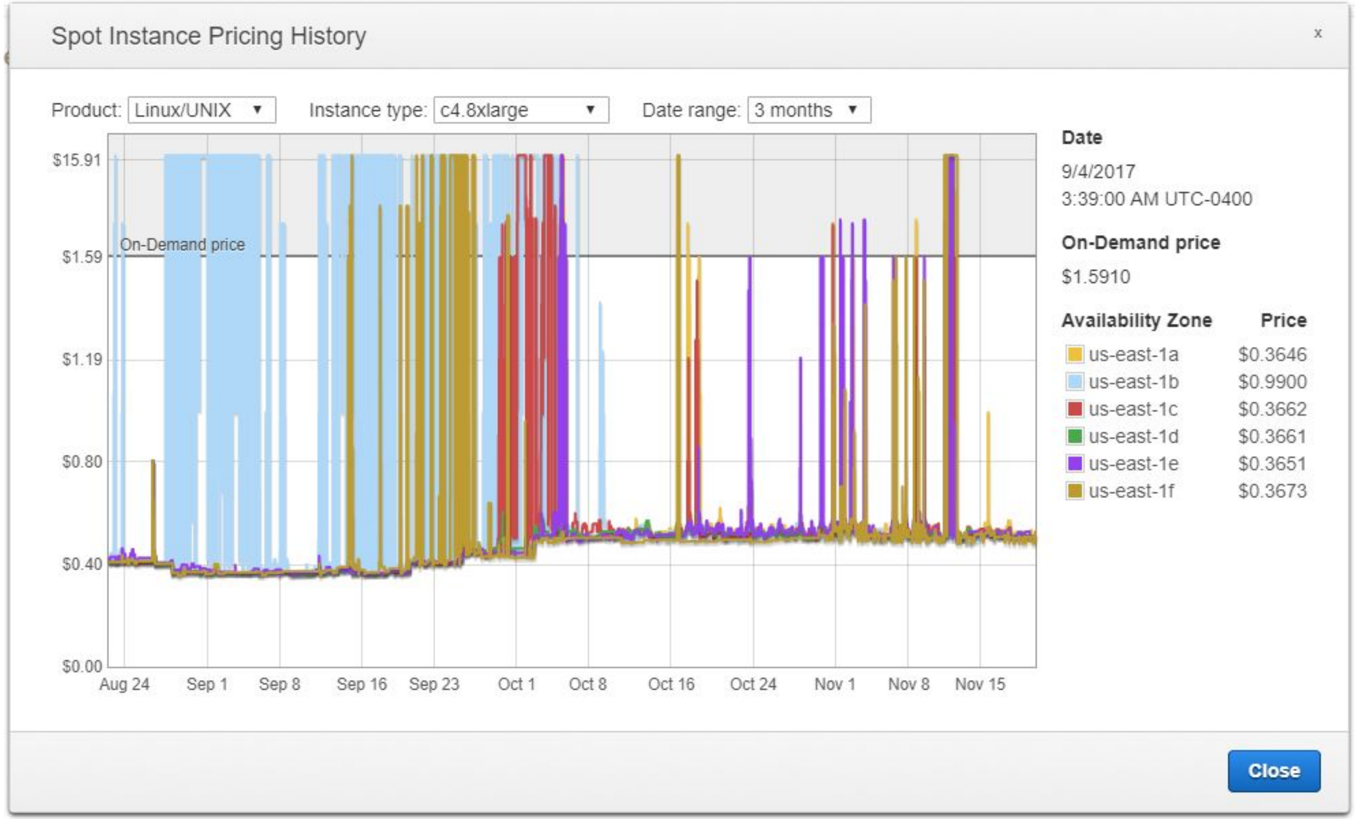
Ideally, you rent what you need and then turn it off. You can rent fractional machines.

At AWS, for example, you can buy the same VM capacity in three different ways:

- On demand - full retail price, e.g. C4.8XLARGE =$1.59/hr
- Reserved - commit to a volume, earn a discount = $0.94/$0.62/hr
- Spot Instances! - bid for excess capacity, aim for $0.40/hr

At Google, it’s much simpler. Preemptible VMs are a fixed 80% discount

Current state of the High-end Spot market



Spot Instance Pricing History

Product: Linux/UNIX

Instance type: c4.8xlarge

Date range: 3 months

Date

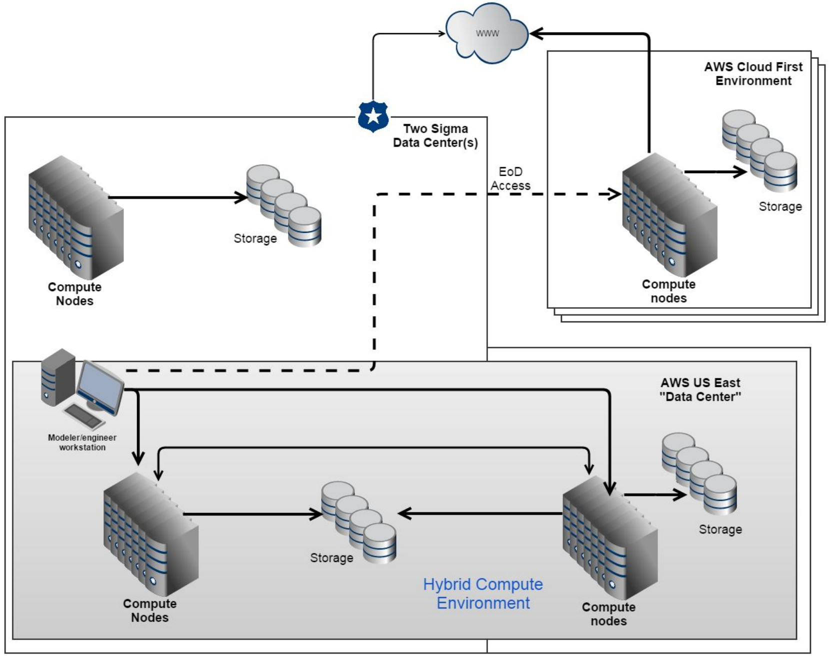
9/4/2017
3:39:00 AM UTC-0400

On-Demand price

$1.5910

Availability Zone	Price
🟨 us-east-1a	$0.3646
🟦 us-east-1b	$0.9900
🟥 us-east-1c	$0.3662
🟩 us-east-1d	$0.3661
🟪 us-east-1e	$0.3651
🟫 us-east-1f	$0.3673

The Dream: Infinitely-scalable Research Environment



Researchers have nearly transparent access to nearly infinite resources

Simulation developers code to the same platform everywhere

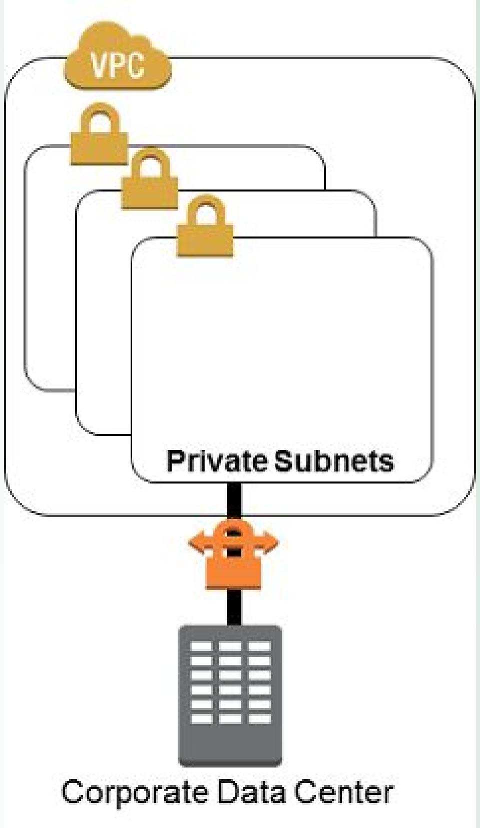
Only the Public Cloud platform team is exposed to AWS vs Google variance

Networks and Virtual Private Clouds

AWS Virtual Private Cloud or Google Private Network

No connection to the Internet. All resources accessed via corporate network

Direct connect from our network to AWS/Google, with redundant firewalls



Managing Permissions

The VPC is a new perimeter for the corporate network

No Internet gateway!   How do we keep it that way?

- Organizations -> Accounts/Projects provide isolation of resources, and permissions.
- The permission to grant permissions is the key to everything. This permission is only available on an exceptional basis.
- Federated login with Corporate identity helps ensure that access originates from our corporate network

Standing up the Environment

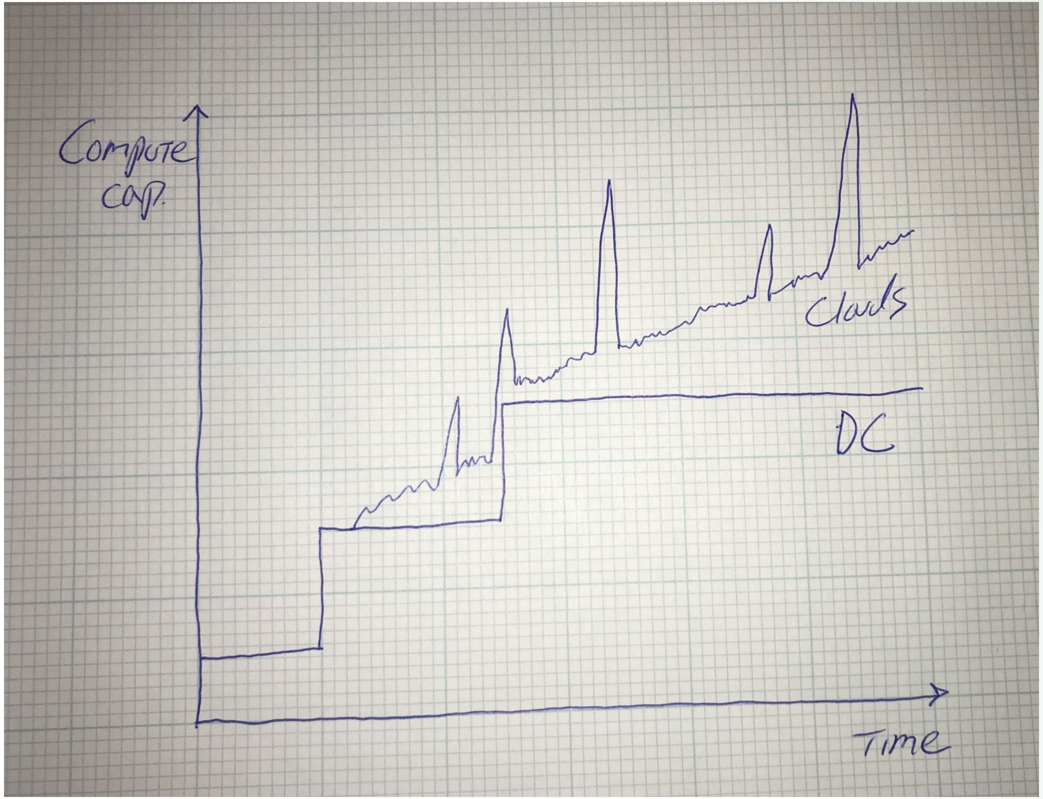
CloudFormation to define Infrastructure (Python scripts for Google)

(-> Terraform which is less verbose)

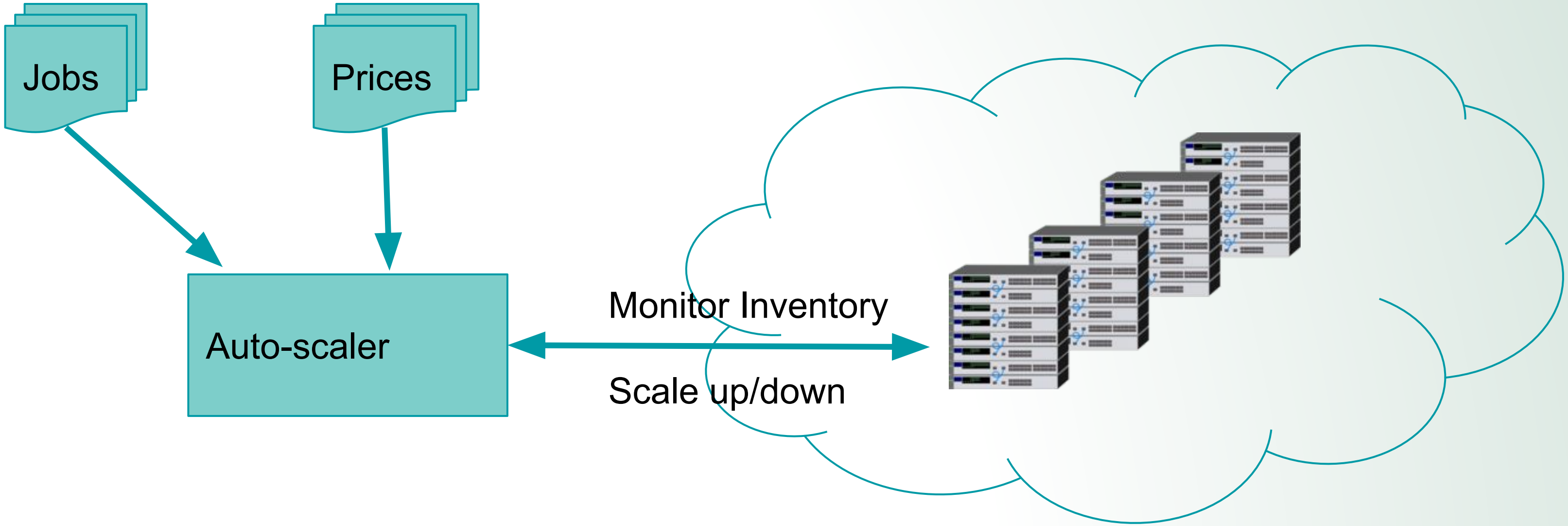
VM Images and Cloud init for Mesos slaves (x2 for GCP)

- Because of the Direct Connect, most data center services are accessible
- Boot up much like in our Data Center, even join the same Mesos clusters

Provisioning and Auto-Scaling



Backlog, prices, and capacity



Scaling Up The Hybrid Research Environment

Scaling up

- Initially, it just works
- But slowly… There is limited bandwidth back to the data center
- Docker images are huge, optimize deployment with disk snapshots
- Storage is the next bottleneck, deploy a cache
- On-premise services aren't engineered for elastic scaling, servers normally arrive slowly, over months, not in seconds
  - e.g. Databases, deploy DB in the Cloud (even better, use RDS)
- Preserve optionality by coding to "neutral" APIs wherever possible.
  - e.g. Cook and Mesos for Research jobs

And now just copy/paste that to another Cloud provider, like Google

- New VPC, new permission model
- New provisioner impl, work around several gaps in GCP services
- Eventually "just works", and we can load-balance across the clouds

Cloud Native Services (~SaaS)

High-value, but vendor-specific Services, such as:

- Big Database: Google BigQuery, AWS Athena and Aurora
- Deep/Machine Learning: Google Cloud ML, AWS Machine Learning, IBM Watson
- Vertical Deep Learning models-as-a-Service (Speech, video, sentiment)

Evaluate cost/benefit carefully

Pick an Open Source API to preserve portability, e.g. Apache Beam, Mesos, Kubernetes, SQL

BigQuery as the Ultimate Cloud Native Service

Insanely-scalable SQL Database in the sky

Pay as you go for storage and queries

Accessed by REST API, CLI, Jupyter notebook

Imagine operating your own Internet-scale distributed database

Benchmarking BigQuery

See http://tech.marksblogg.com/billion-nyc-taxi-rides-bigquery.html

1.1 billion taxi rides, 104GB data

Loaded in 24 minutes

Typical queries run in ~2 seconds: “select count of all rides grouped by taxi type”

This obviously justifies paying some premium and accepting some degree of lock-in

Adding Cloud Native Services to your Hybrid Cloud

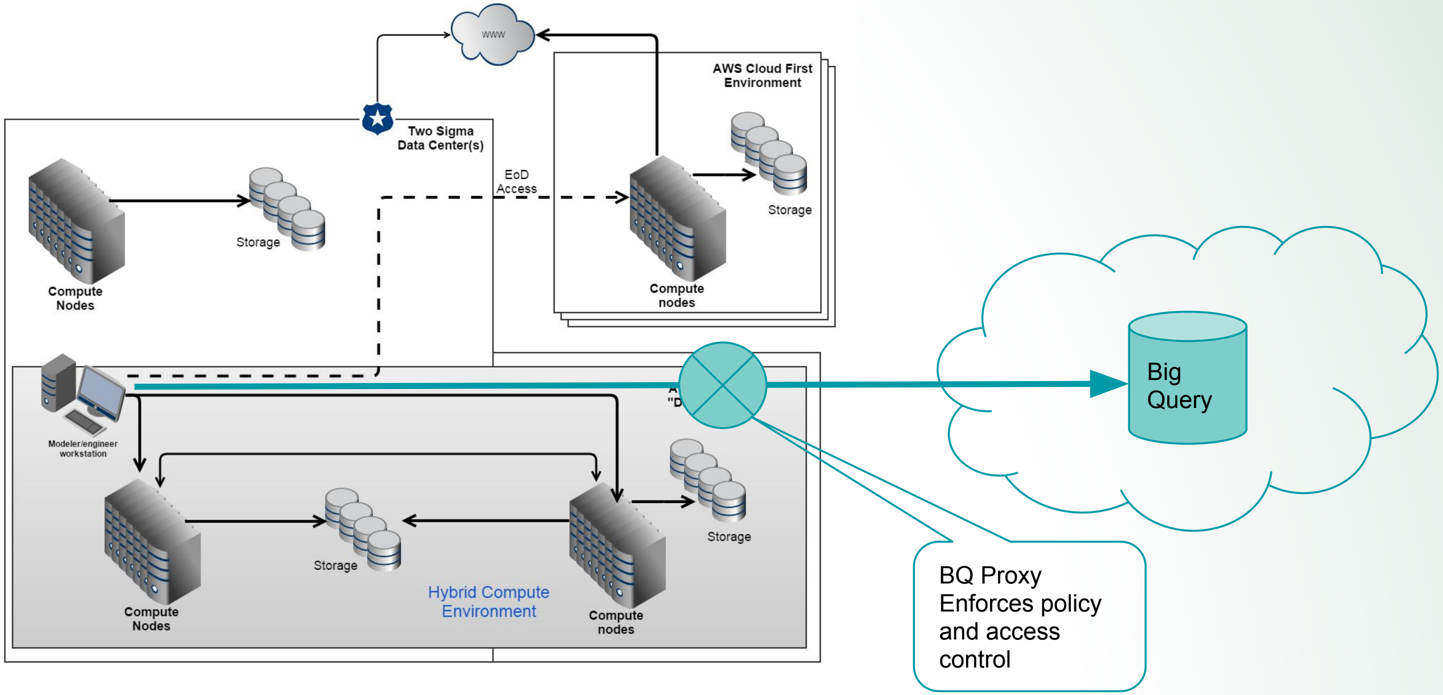
Security considerations

IAM/permissions

Network paths

Proxying can maintain control

Proxying access to BigQuery



BQ Proxy
Enforces policy
and access
control

Conclusion

Hybrid Cloud(Data Center + Public Cloud(s)) is the reality for many enterprises at the moment.

You can take advantage of Public Cloud resources, but at some cost.

Decide what your goals are for adoption, and be strategic about it.

Be prepared to exploit Cloud Native services as appropriate.

Thank you!

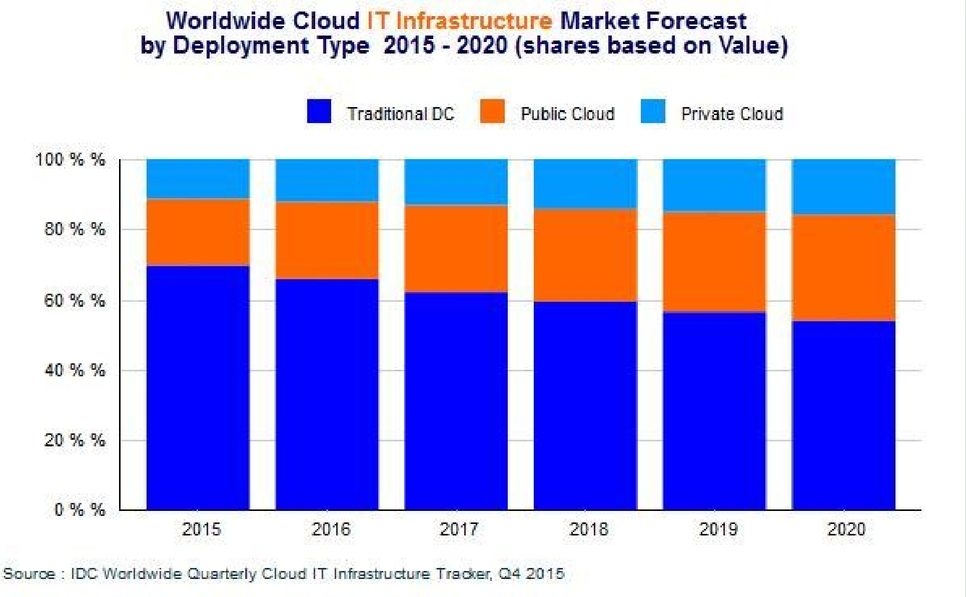
Backup/Reference

Why now?

Now: It’s clear we’ve reached mainstream acceptance for Cloud Computing.

Why?

- Time -> trust
- AWS success stories
- If you don’t, your competitor will



Worldwide Cloud IT Infrastructure Market Forecast
by Deployment Type 2015 - 2020 (shares based on Value)

Source : IDC Worldwide Quarterly Cloud IT Infrastructure Tracker, Q4 2015

Suggested reading

- A View of Cloud Computing
  - http://cacm.acm.org/magazines/2010/4/81493-a-view-of-cloud-computing/fulltext
- NIST definition of Cloud Computing - 2011
  - http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf
- The origin of Amazon Web Services:
  - https://blog.hackerrank.com/how-amazon-web-services-surged-out-of-nowhere/
- Netflix migration story (GOTO 2012 • Globally Distributed Cloud Application at Netflix • Adrian Cockcroft)
  - https://www.youtube.com/watch?v=Mn0_Xmw4rQs
  - http://gotocon.com/dl/goto-aar-2012/slides/AdrianCockcroft_GloballyDistributedCloudApplicationsAtNetflix.pdf