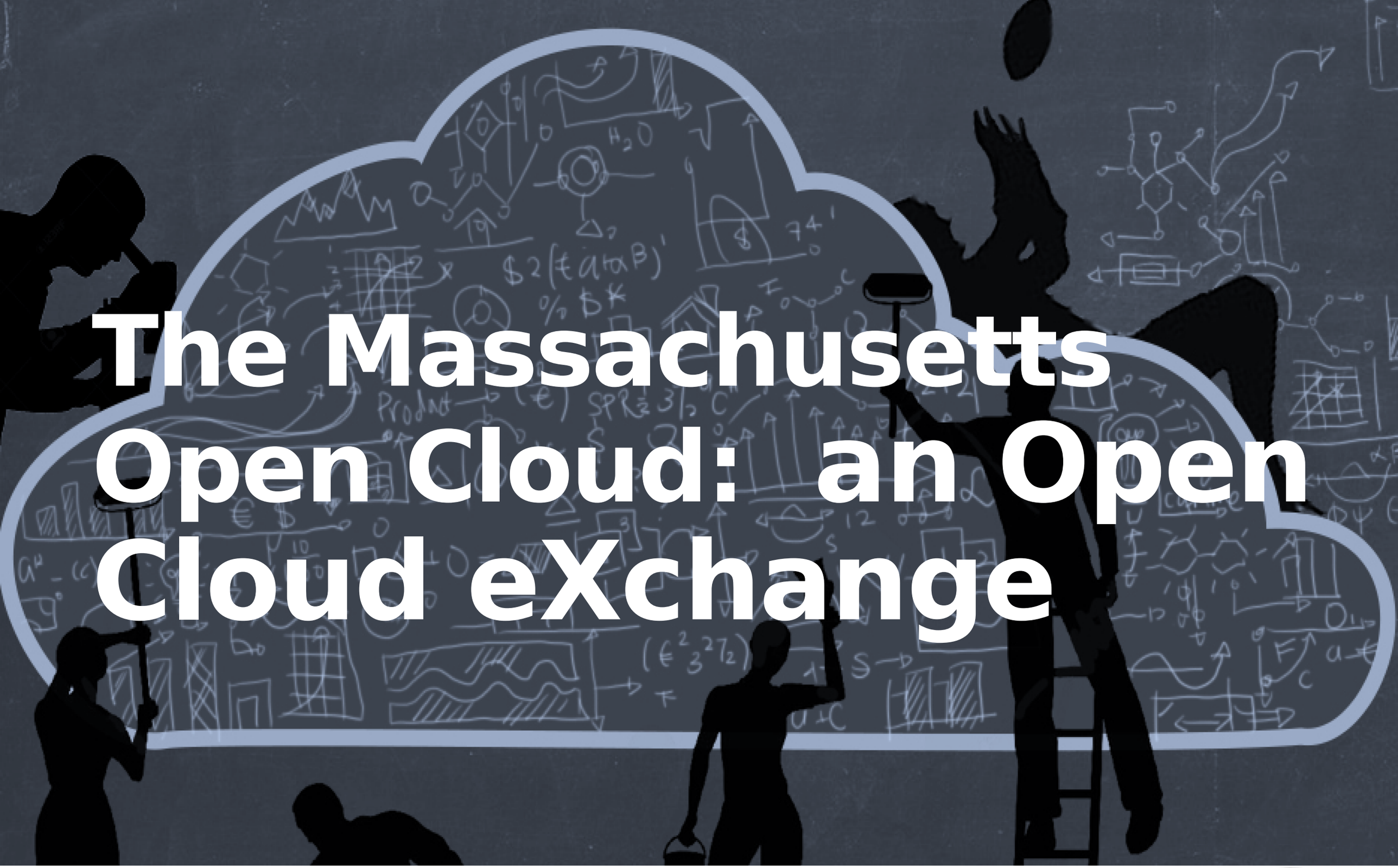
The Massachusetts Open Cloud: an Open Cloud eXchange



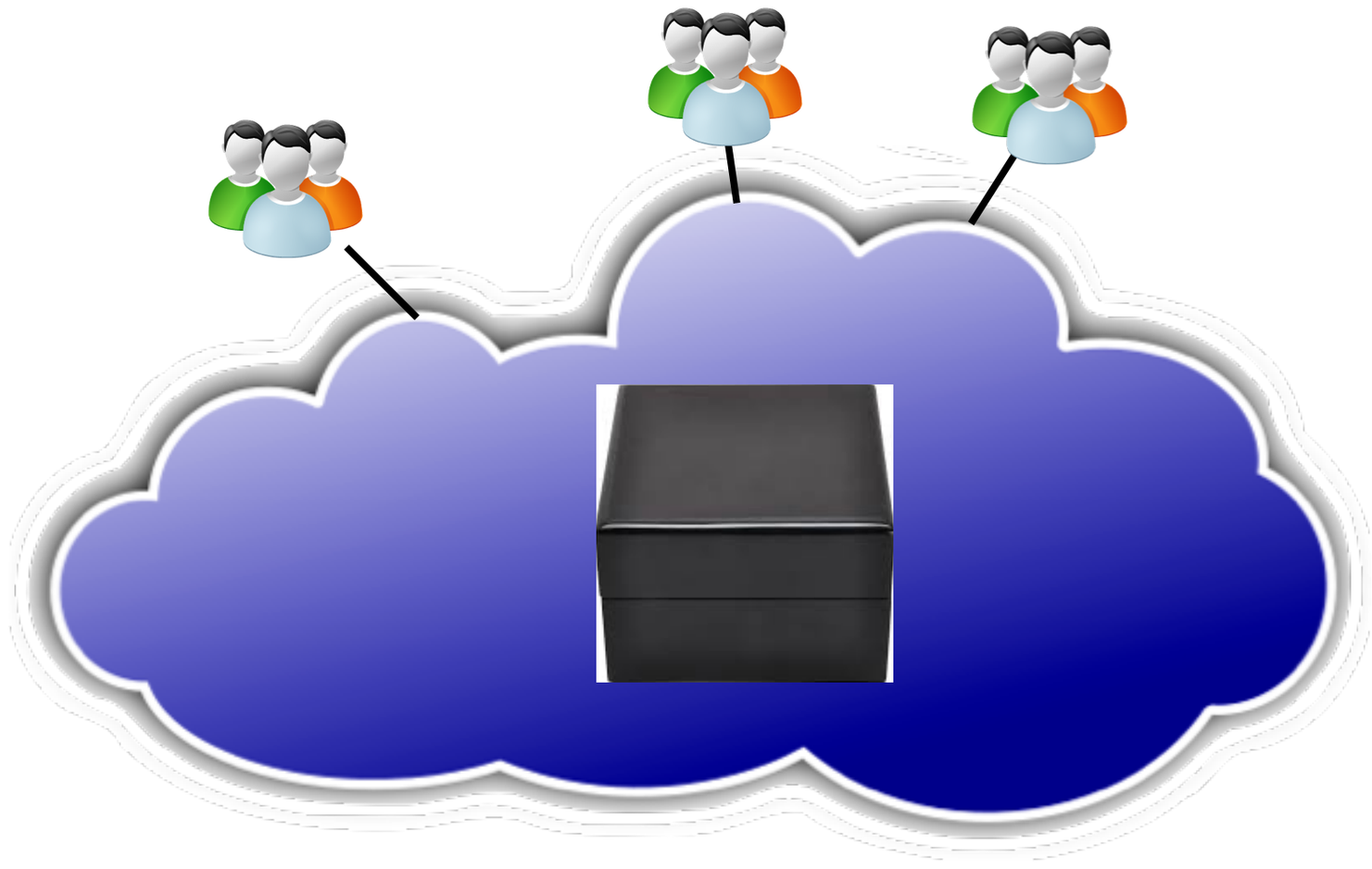
CLOUD COMPUTING IS HAVING A DRAMATIC IMPACT

- On-demand access
- Economies of scale

All compute/storage will move to the cloud?

Today’s IaaS clouds

- One company responsible for implementing and operating the cloud
- Typically highly secretive about operational practices
- Exposes limited information to enable optimizations

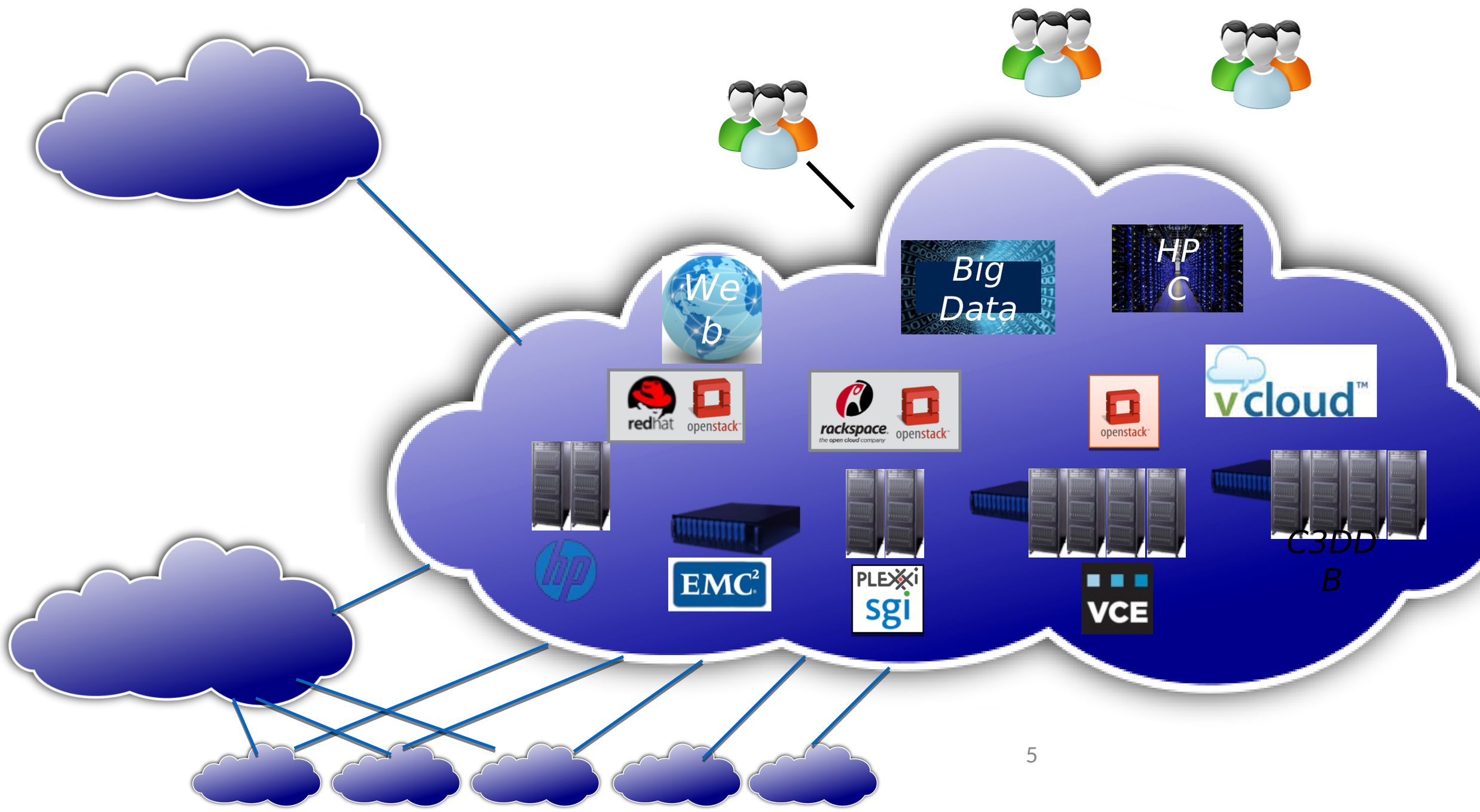


What's the problem

- Lots of innovation above the IaaS level… but
  - consider Ente
- Lots of differen
  - bandwidth b
  - offerings inco em
  - price challenges to moving
- No visibility/auditing internal processes
- Where is your data!
- Price is terrible for computers run 24x7x365

We are in the equivalent of the pre-Internet world, where AOL and CompuServe dominated on-line access

Is a different model possible? An “Open Cloud eXchange (OCX)”



BIG BOX STORE

SHOPPING MALL

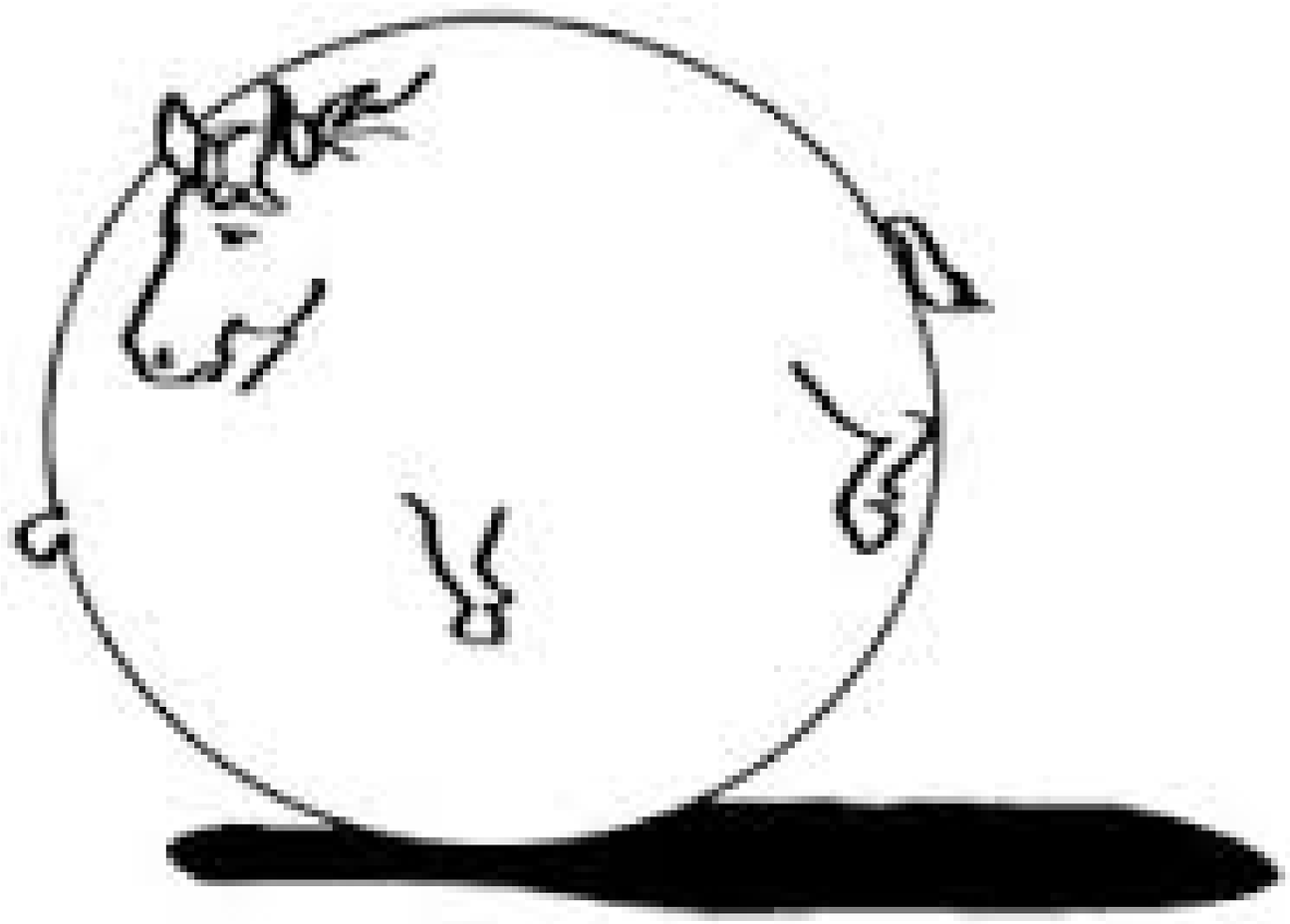
CATHEDRAL

BAZAAR

Why is this important

- Anyone can add a new service and compete in a level playing field
- History tells us the opening up to rich community/marketplace competition results in innovation/efficiency:
  - “The Cathedral and the Bazaar” by Eric Steven Raymond
  - “The Master Switch: The Rise and Fall of Information Empires” by Tim Wu
- This could fundamentally change systems research:
  - access to real data
  - access to real users
  - access to scale

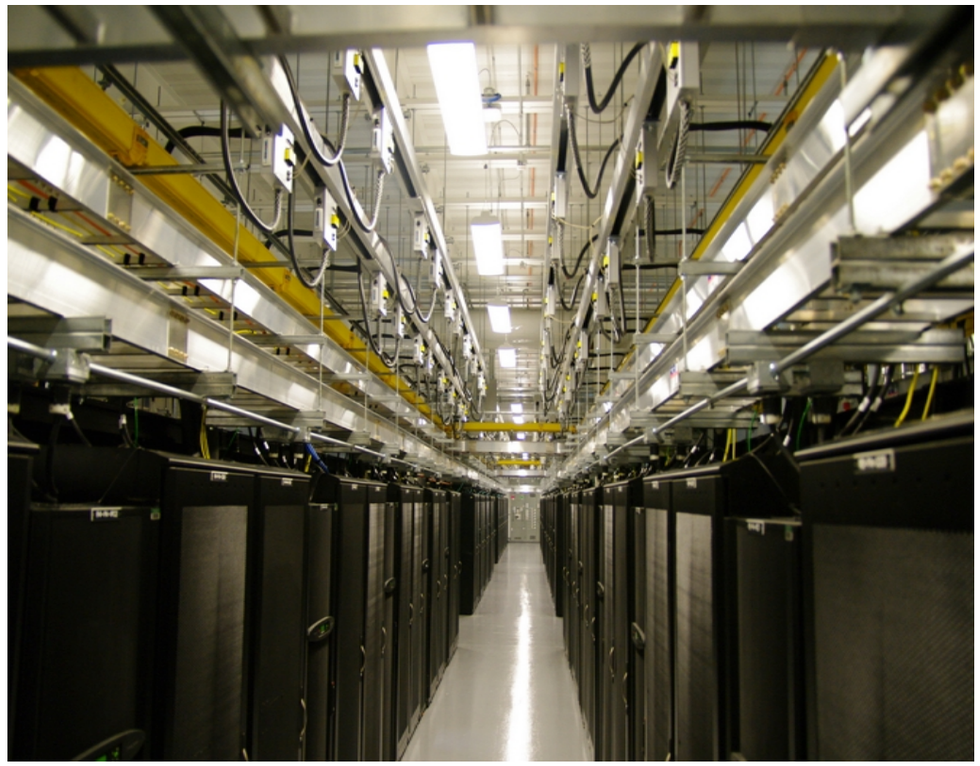
Without that…solving the spherical horse problem…



This isn’t crazy… really

- Current clouds are incredibly expensive…
- Much of industry locked out of current clouds
- lots of great open source software
- lots of great niche markets; markets important to us…
- lots of users concerned by vendor lock in…
- this doesn’t need to be AWS scale to be worth it
  - “Past a certain scale; little advantage to economy of scale” — John Goodhue

MGHPCC



15 MW, 90,000 square feet + can grow

10s of thousand HPC users, potentially many more cloud users

The Massachusetts Open Cloud

ADVERTISEMENT

Governor Patrick Announces Funding to Launch Massachusetts Open Cloud Project

Mon, 04/28/2014 - 12:07pm

by Mass Open Cloud Project

Get the latest news in High Per
more - Sign up now!

BOSTON UNIVERSITY

THE MASSACHUSETTS COLLABORATORS

CANONICAL

CAMBRIDGE Computer

MathWorks

EnterpriseDB The Enterprise Postgres Company

riverbed

OAK RIDGE National Laboratory

Mellanox TECHNOLOGIES

EMC²

JUNIPER NETWORKS

PLEXXI

intel

U.S. AIR FORCE

sgi

DataDirect NETWORKS Information in Motion

DELL

Lenovo

2σ TWO SIGMA

UMASS

BROCADE

MIT

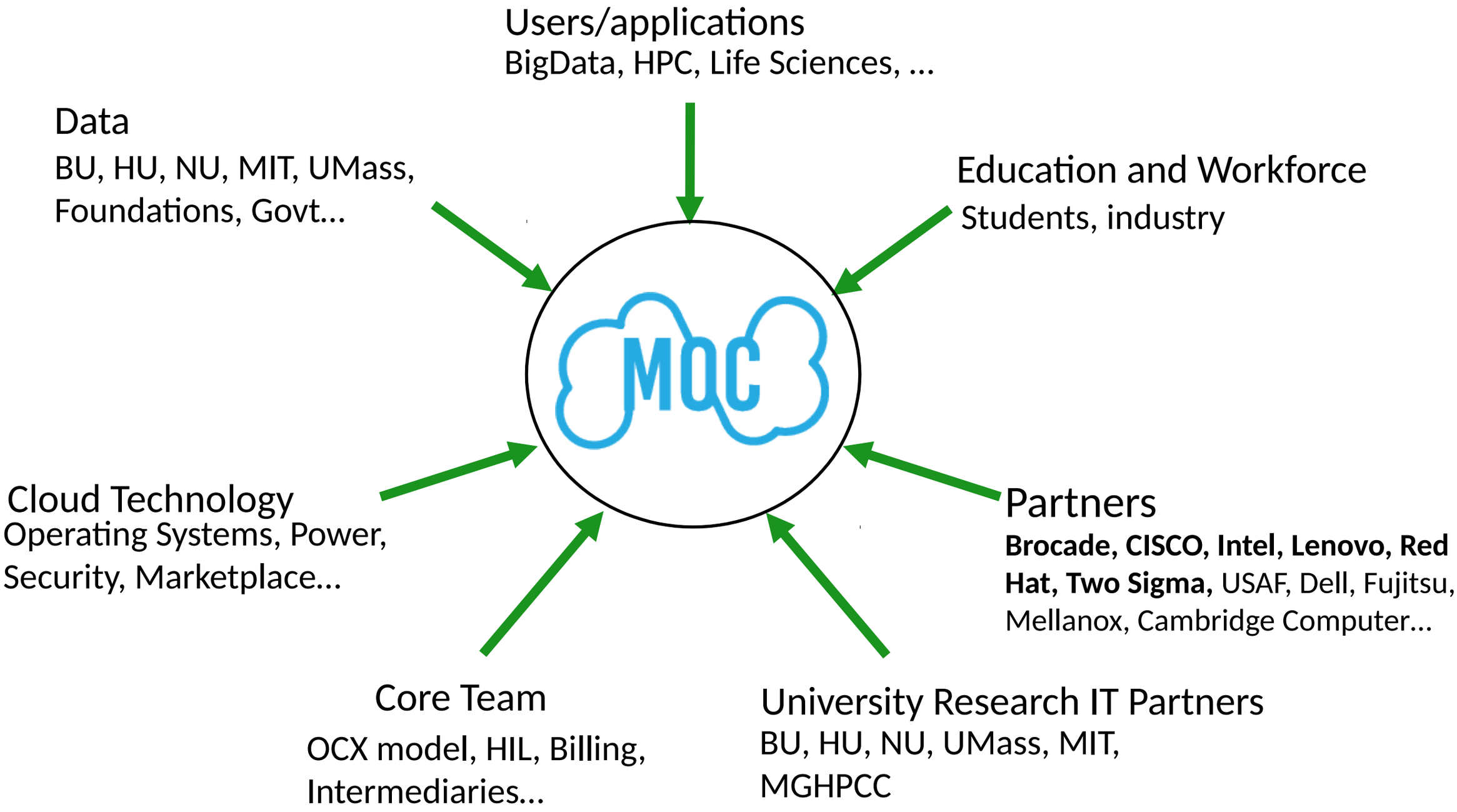
CISCO

Northeastern

redhat

BOSTON UNIVERSITY

MOC Ecosystem



It’s real…

- Available now: Production OpenStack services…
  - Small scale, but growing (couple of hundred servers, 550 TB storage), 200+ users
  - VMs, on-demand Big Data (Hadoop, SPARK…),

- What’s coming:
  - Simple GUI for end users
  - OpenShift – Red Hat
  - Federation across universities
  - Rapid/secure Hardware as a Service
  - 20+ PB NESE DataLake
  - Cloud Dataverse

- Platform for enormous range of research projects across BU, NEU, MIT & Harvard

Red Hat Collaboratory

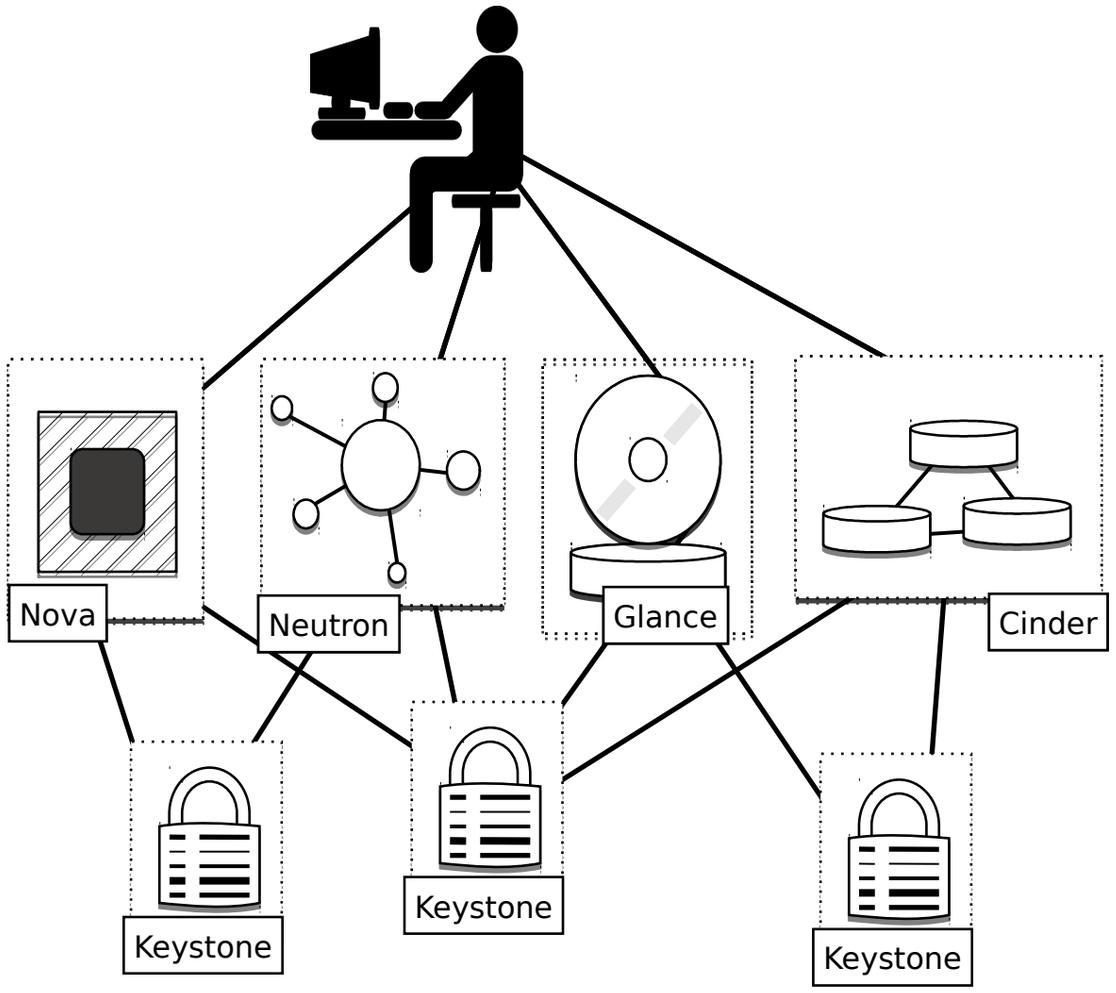
- Mix & Match
- HIL & BMI (and QUADS integration)
- Big Data Analytics and Cloud Dataverse
- Datacenter-scale Data Delivery Network (D3N)
- Monitoring, Tracing, Analytics …
- OpenShift on the MOC
- Accelerator Testbed

End-to-end POC: Radiology in the cloud targeting OpenShift with accelerators

Mix & Match:
Resource Federation

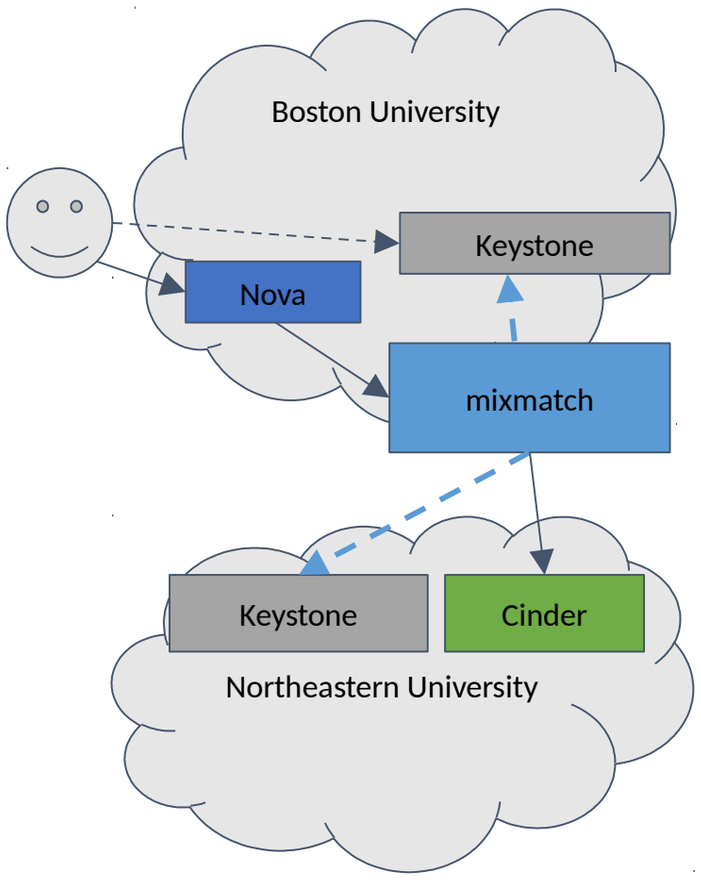
OPENSTACK FOR AN OCX

- OpenStack is a natural starting point
- Mix & Match federation



Mix and Match (Resource Federation)

- Solution
  - Proxy between OpenStack services
- Status of the project
  - Hosted upstream by the OpenStack infrastructure
    - https://github.com/openstack/mixmatch
  - Production deployment planned for Q1 2017
- Team:
  - Core Team: Kristi Nikolla, Eric Juma, Jeremy Freudberg
  - Contributors: Adam Young (Red Hat), George Silvis, Wjdan Alharthi, Minying Lu, Kyle Liberti
- More information:
  - https://info.massopencloud.org/blog/mixmatch-federation/



MOC Bare Metal Cloud Projects

Jason Hennessey (henn@bu.edu)

Why Bare Metal?

Useful for different workloads:

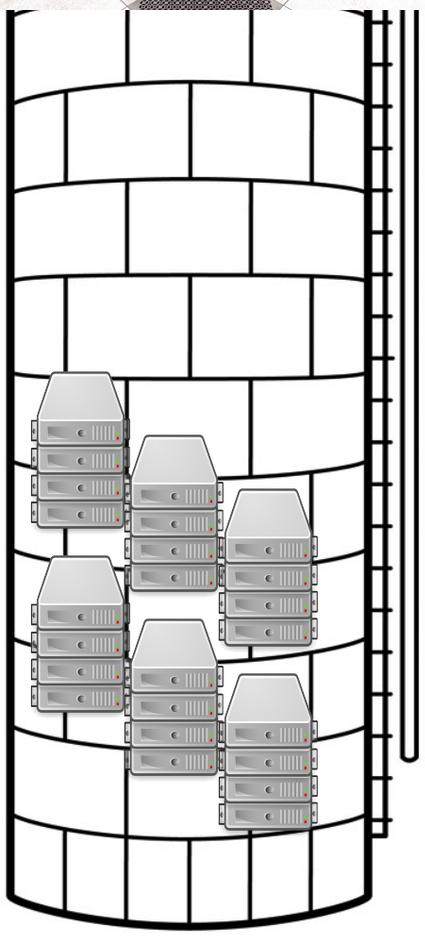
- Staging, testing, production
- HPC + Cloud
- Max / predictable performance
- Run VMs
- Non-virtualizable hardware
- Increased Security
- Less trust in the provider

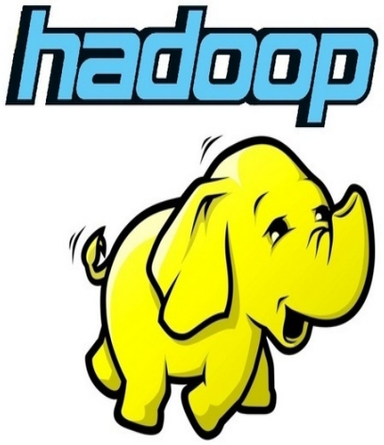
MOC Projects in Bare Metal

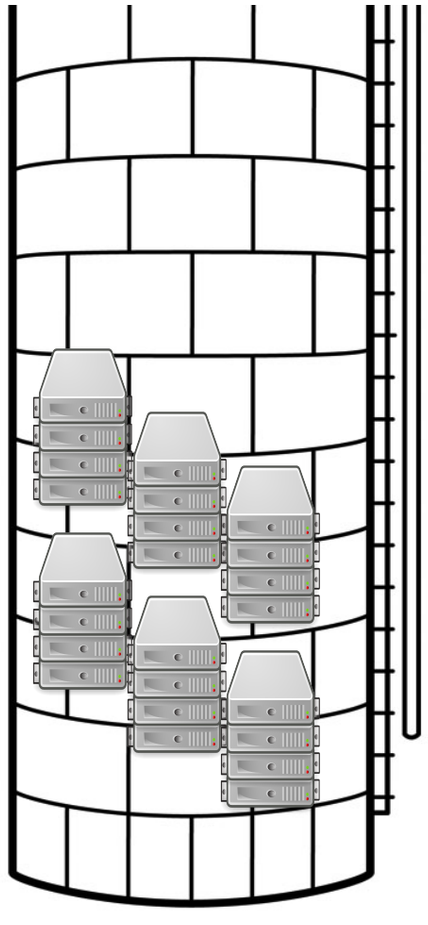
- Bringing software configuration advantages of virtualization to bare metal:
  a) Hardware Isolation Layer
     - Allocate & configure nodes and networks
  b) Bare Metal Imager
     - Image management: fast provisioning, cloning and snapshotting of disks
  c) Secure Cloud
     - Checks that each machine is in pristine / untampered

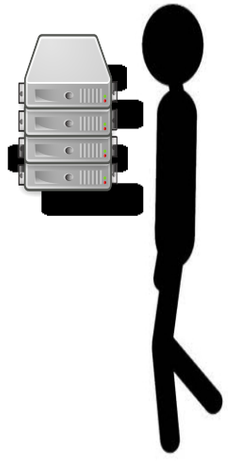
HIL: Hardware isolation layer

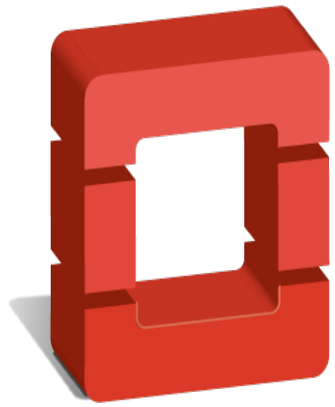
Datacenter has isolated silos



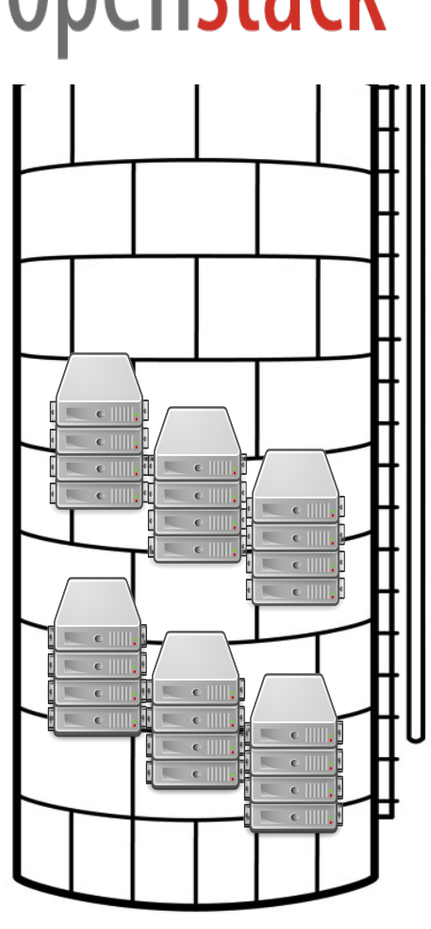




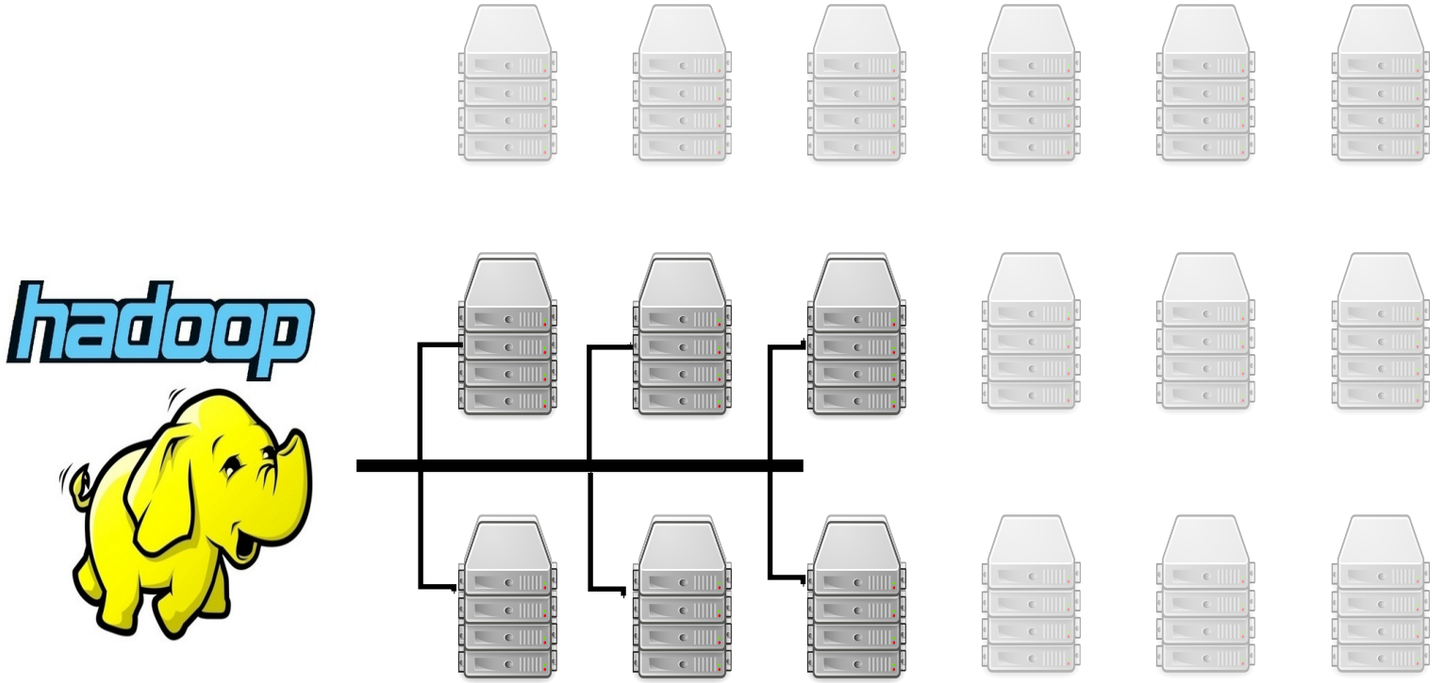




openstack™



Hardware isolation layer



Connect nodes and networks

Status

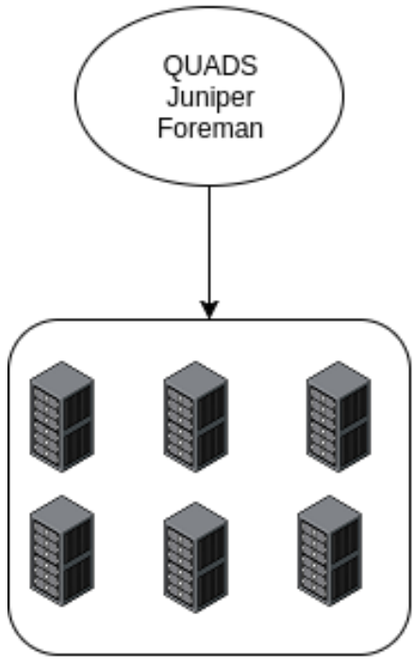
- In production at the Mass Open Cloud: production OpenStack environment, staging area(s), OS research, Big Data on-demand
- Supports variety of provisioning systems: Foreman, MaaS, Ironic, home brewed research (EbbRT)

HIL and QUADS integration

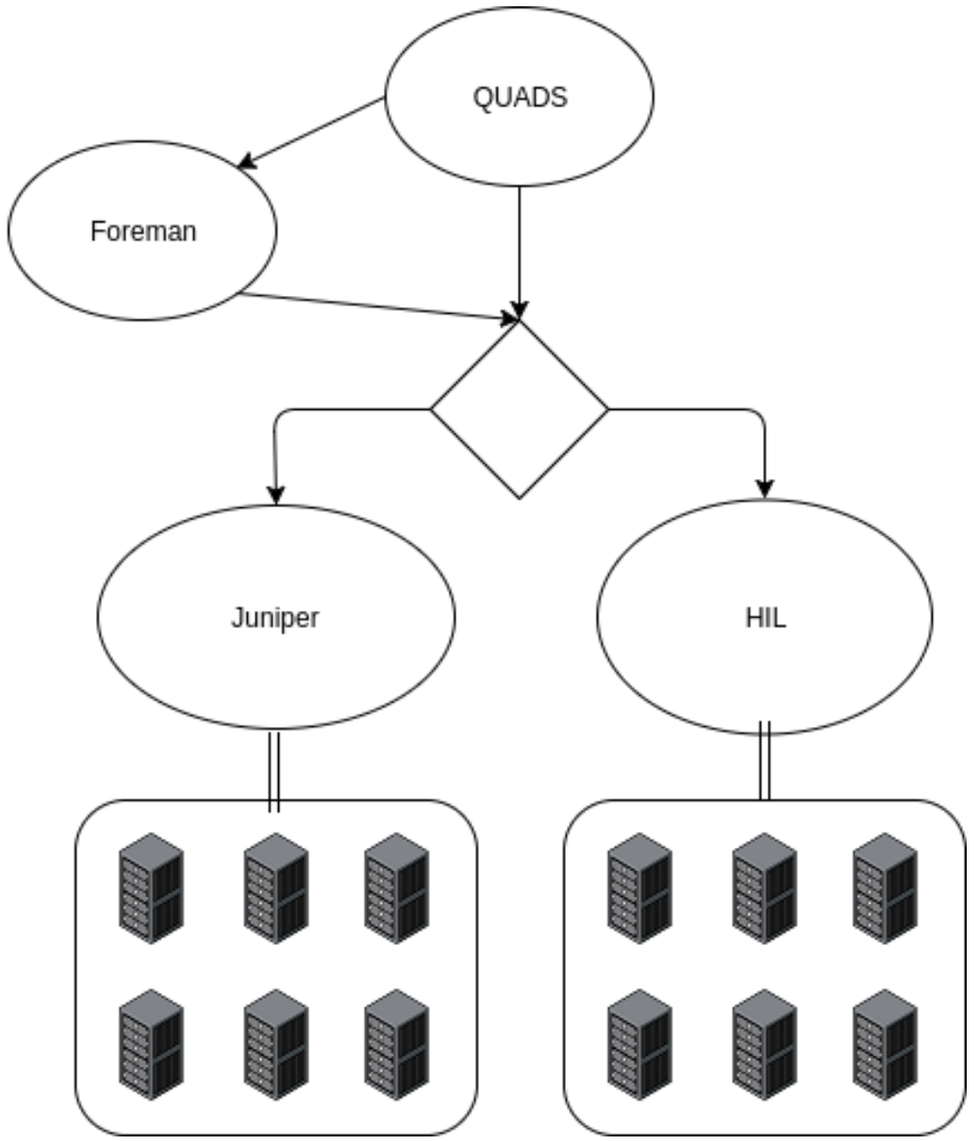
Goal: Extend Red Hat's QUADS (Quick and Dirty Scheduler) to be able to use the MOC's HIL to manage hardware isolation

Reasons/Motivation: Enhances QUADS portability, and endows HIL the ability to dynamically schedule

Before



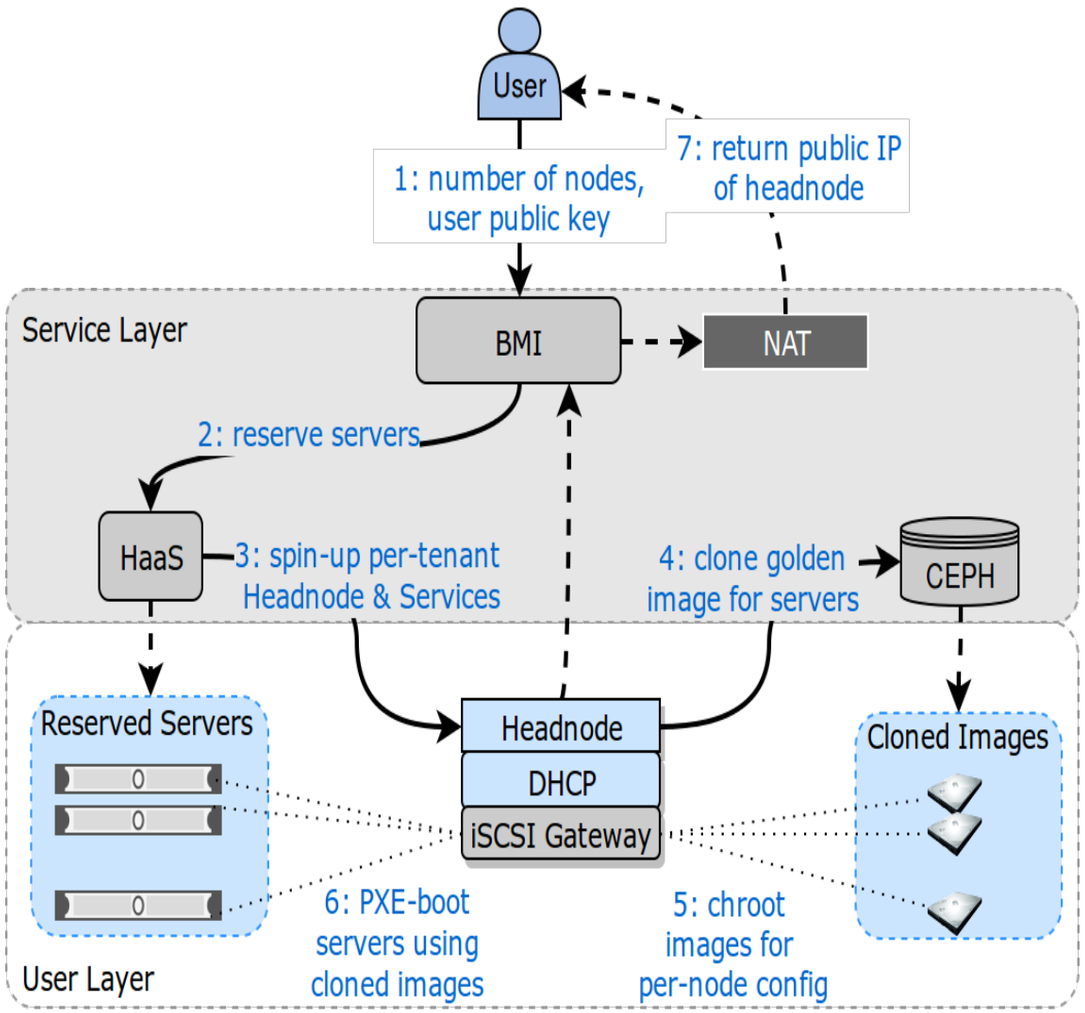
After



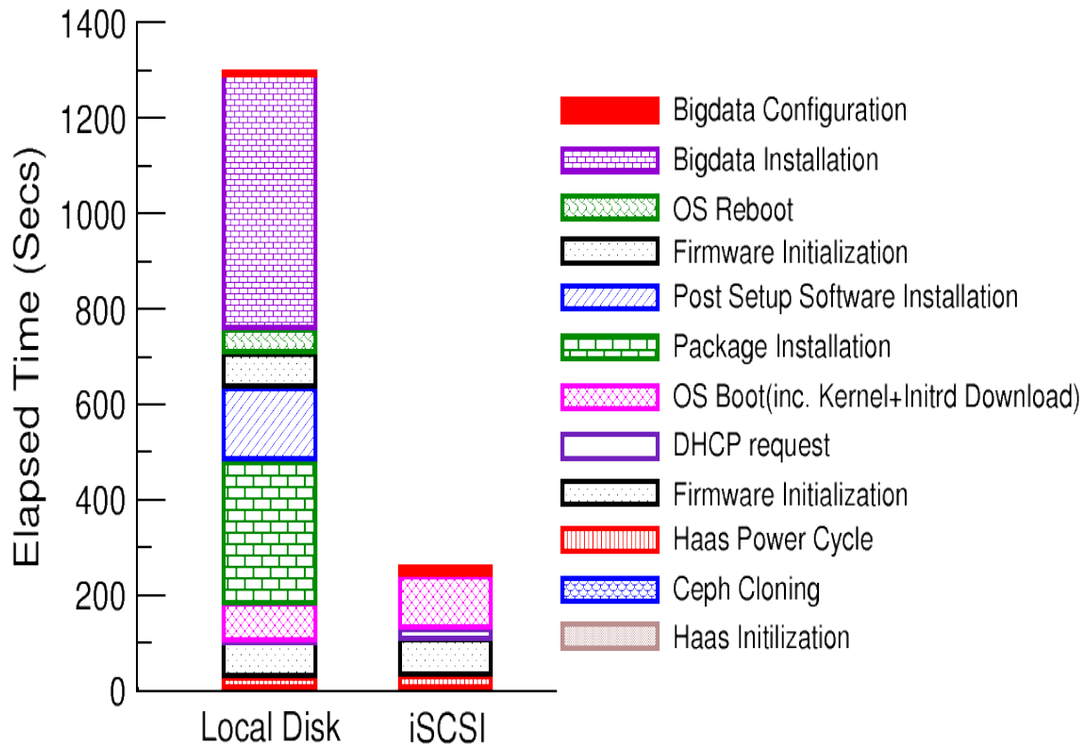
BMI: Bare Metal Imager

Bare Metal Imager: VM-like disk image management

iSCSI-based



Able to provision + boot in < 5 min

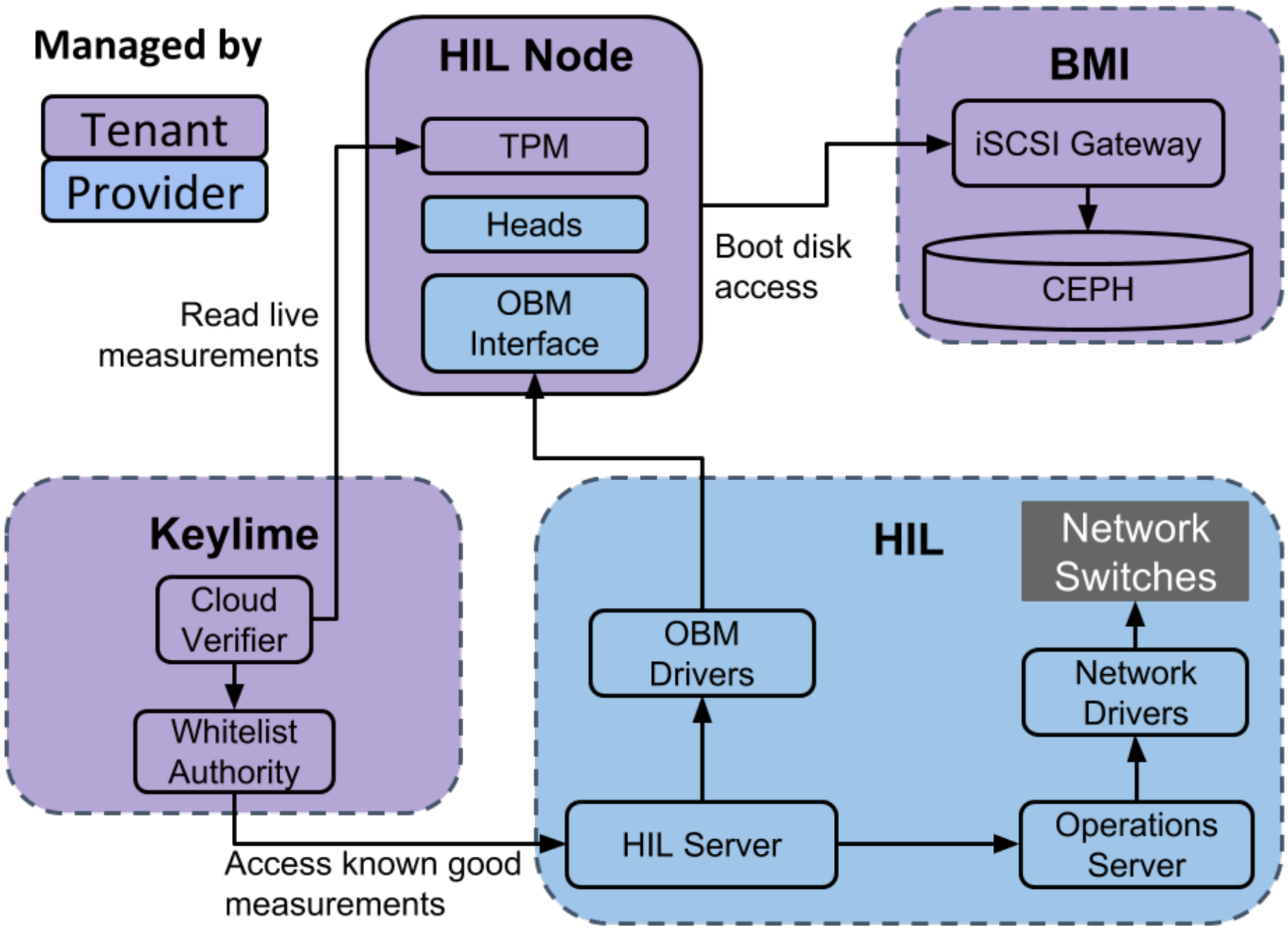


Turk, A., Gudimetla, R. S., Kaynar, E. U., Hennessey, J., Tikale, S., Desnoyers, P., & Krieger, O. (2016). An Experiment on Bare-Metal BigData Provisioning. In 8th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 16).

Secure Cloud

Secure Cloud

- Goals
  - Increase confidence in the firmware
  - Minimal amount of provider-trusted changes
  - Transition nodes quickly
- Firmware integrity (system & peripheral):
  - **Measure**
  - Protect
  - Replace
  - Maintain/Audit
- Vendor survey
- Hardware specification


Managed by
Tenant
Provider
HIL Node
TPM
Heads
OBM Interface
BMI
iSCSI Gateway
CEPH
Boot disk access
Read live measurements
Keylime
Cloud Verifier
Whitelist Authority
HIL
OBM Drivers
Network Switches
Network Drivers
HIL Server
Operations Server
Access known good measurements

People/contacts:

- **HIL (Hardware Isolation Layer)** github.com/CCI-MOC/hil
  - Contact: haas-dev-list@bu.edu
  - Core team: Jason Hennessey (BU), Sahil Tikale (BU), Ian Denhardt (BU), Peter Desnoyers (NEU), Orran Krieger (BU), Jethro Sun (BU), Kristi Nikolla (BU), Nicholas Matsuura (USAF), Naved Ansari (BU), Kyle Hogan (BU), Mengyuan Sun (MIT), Gwen Faline Edgar (MIT)
  - Contributors (some were past affiliations): George Silvis III (BU), Yue Zhang (BU), Apoorve Mohan (NEU), Ravisantosh Gudimetla (NEU), Minying Lu (BU), Zhaoliang Liu (NEU), Ryan Abouzahra (USAF), Jonathan Bell (BU), Jonathan Bernard (BU), Rohan Garg (NEU), Andrew Mohn (BU), Abhishek Raju (NEU), Ritesh Singh (NEU), Ron Unrau and Valerie Young (BU)

- **BMI (Bare Metal Imager)** github.com/CCI-MOC/ims
  - Contact: Gene Cooperman <gene@ccs.neu.edu>
  - Core team: Gene Cooperman (NEU), Naved Ansari (BU), Apoorve Mohan (NEU), Pranay Surana (NEU), Ravi Santosh Gudimetla (Redhat, formerly NEU), Sourabh Bollapragada (NEU)
  - Contributors: Jason Hennessey (BU), Ata Turk (BU), Ugur Kaynar (BU), Sahil Tikale (BU), Orran Krieger (BU), Peter Desnoyers (NEU)

- **Secure Cloud**
  - Contact: Jason Hennessey <henn@bu.edu>
  - Core team: Jason Hennessey (BU), Nabil Schear (MIT LL), Trammell Hudson (Two Sigma), Orran Krieger (BU) Gerardo Ravago (BU), Kyle Hogan (BU), Ravi S. Gudimetla (NEU), Larry Rudolph (Two Sigma), Mayank Varia (BU)

Cloud Dataverse

Dataverse

- Dataverse is an open-source software platform for building data repositories
- It provides an incentive to share data
  - Gives credit through data citation
- Provides mechanisms for control over data access
- Builds a community:
  - To foster new research in data sharing
  - To define new standards and best practices
  - Installed in 20 repositories world wide
  - Hosting dataverses from > 500 institutions

Cloud Dataverse

Cloud Dataverse

- A dataset repository solution for cloud
- Extends Dataverse
  - Store datasets in Object Store (Swift)
  - Harvest datasets from all Dataverses
  - Compute button that enables on-cloud computation
    - No need for download

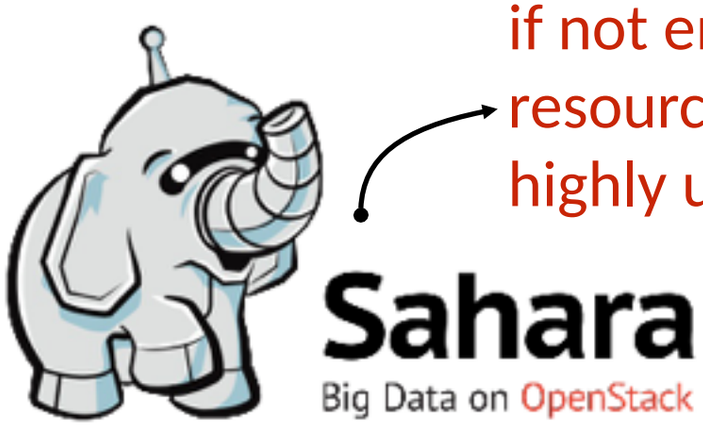
BDaaS @ MOC

Cloud Dataverse
Dataset Repository

BigData clusters spinned via **OpenStack Sahara** to compute over datasets

BDaaS @ MOC

Cloud Dataverse
Dataset Repository



if not enough resources / highly utilized

BigData clusters spinned via **OpenStack Sahara** to compute over datasets

BDaaS @ MOC

Cloud Dataverse
Dataset Repository

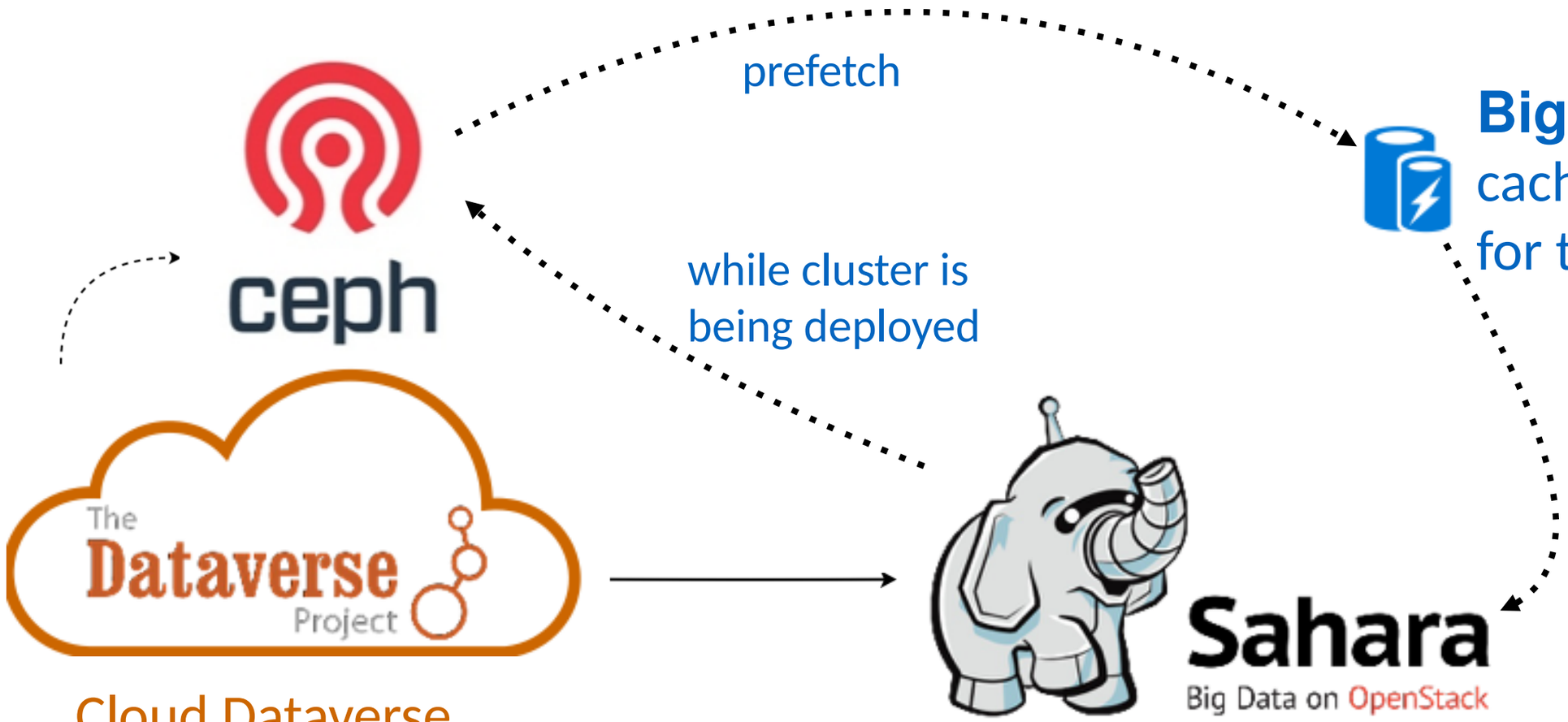
Sahara
Big Data on OpenStack

BigData clusters spinned via
OpenStack Sahara to
compute over datasets

if not enough
resources /
highly utilized

Expand
OpenStack
cluster using
HIL and BMI

BDaaS @ MOC



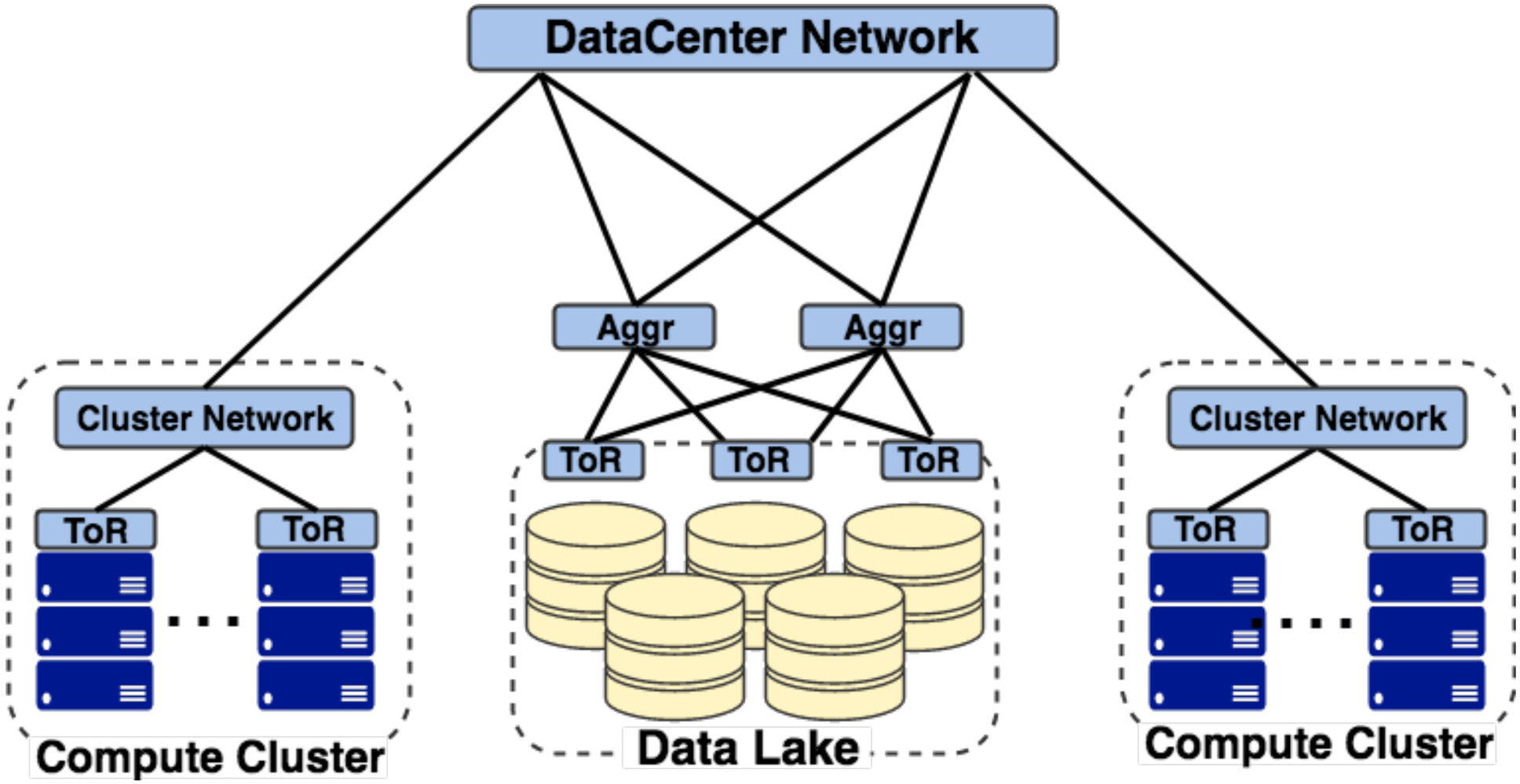
Cloud Dataverse
Dataset Repository

BigData clusters spinned via
OpenStack Sahara to
compute over datasets

Datacenter-scale Data Delivery Network (D3N)

MOC, Red Hat, Intel, Brocade, Lenovo, 2Sigma

Data Lake in a typical DC

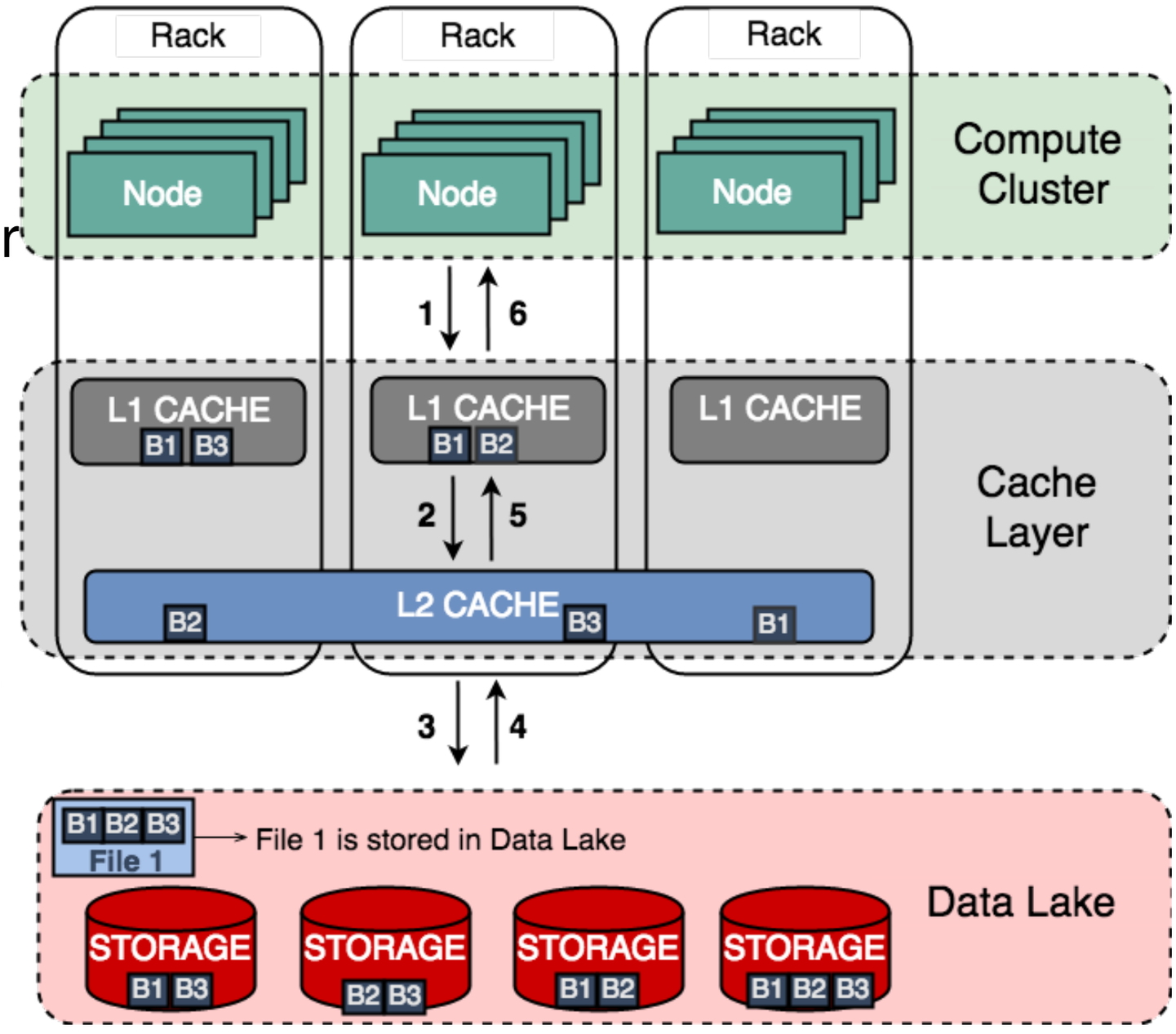


North Eastern Storage Exchange (NESE): 20+PB Harvard, NEU, MIT, BU, UMass

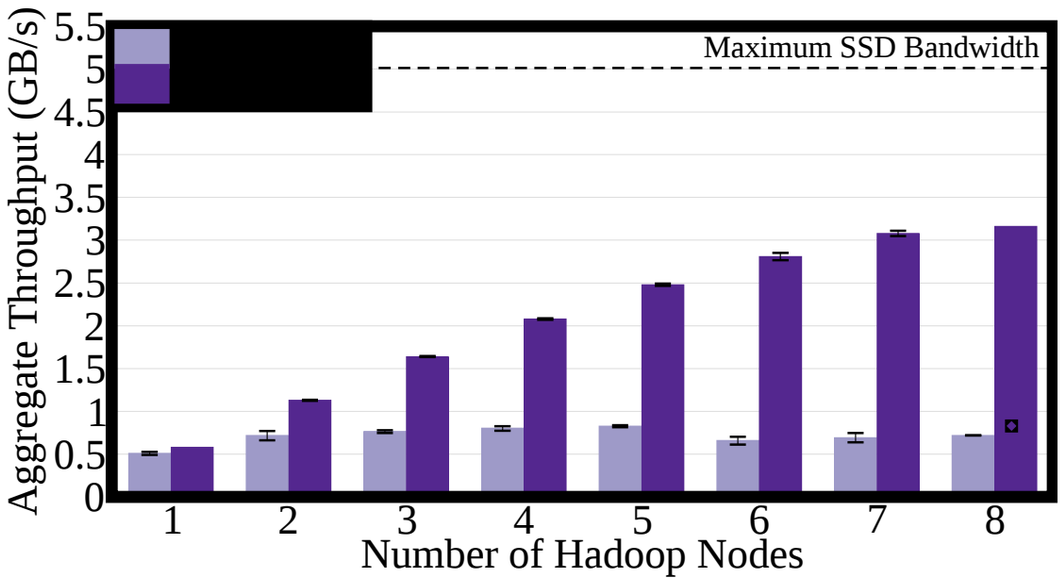
Datacenter scale Data Delivery Network (D3N)

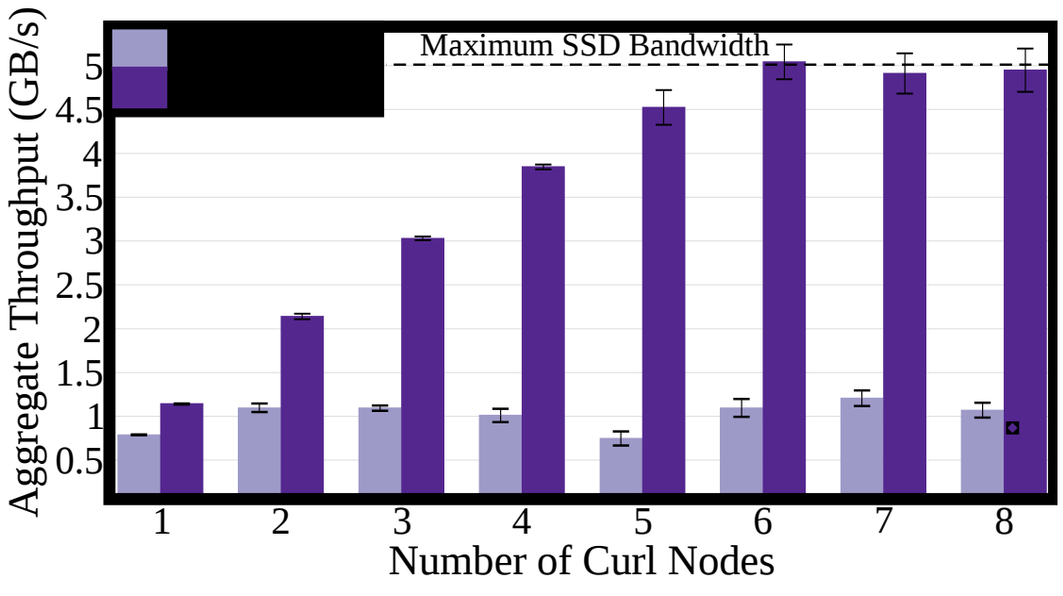
Simple deployment:

- Dedicated cache servers per rack
- **L1 :** Rack Local
  - reduce inter rack traffic
- **L2 :** Cluster Local
  - reduce clusters and back-end storage traffic
- Implemented by modifying **CEPH Rados Gateway**



D3N Results





- Exceeds maximum bandwidth Hadoop
- Demonstrates makes sense to share expensive SSDs – faster than local disk
- With extreme benchmark can saturate SSD & 40 Gb NIC
- Will be of enormous value with NESE data lake

Red Hat Collaboratory

- Mix & Match
- HIL & BMI (and QUADS integration)
- Big Data Analytics and Cloud Dataverse
- Datacenter-scale Data Delivery Network (D3N)
- Monitoring, Tracing, Analytics …
- OpenShift on the MOC
- Accelerator Testbed

End-to-end POC: Radiology in the cloud targeting OpenShift with accelerators

Monitoring, Tracing, Analytics

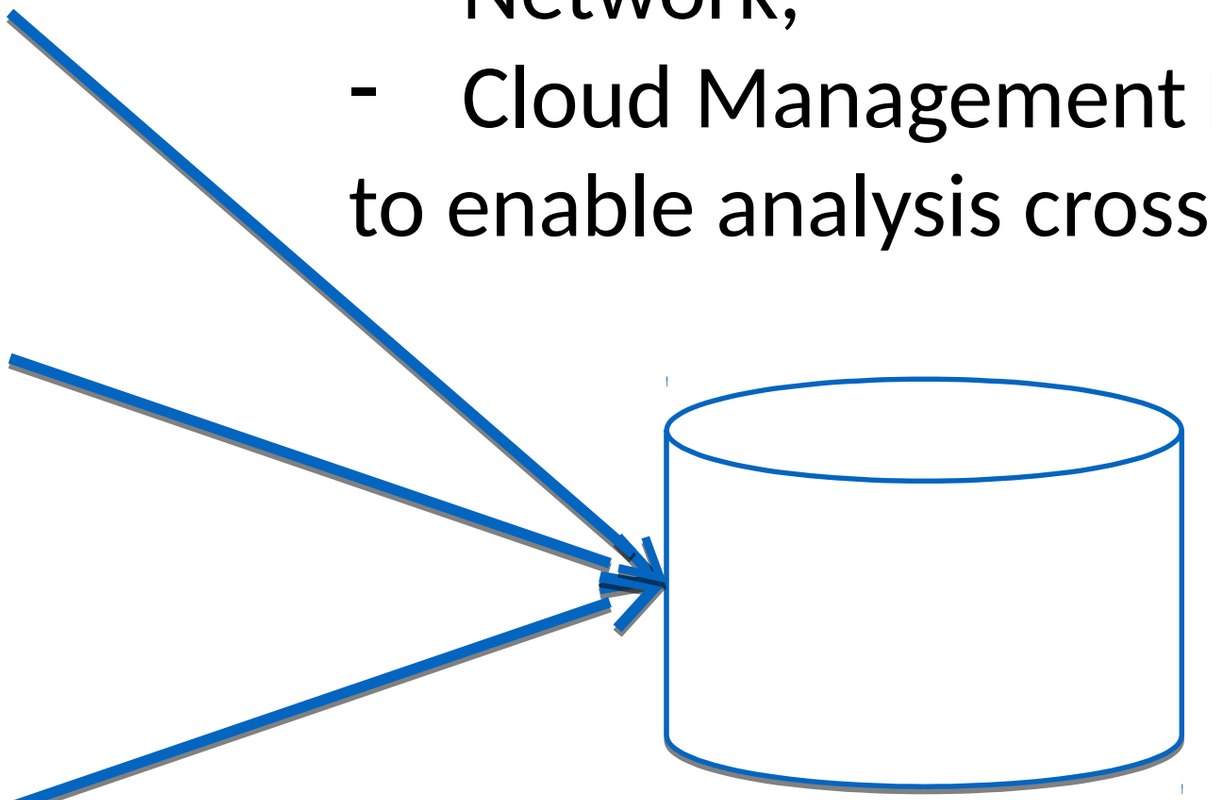
- Problem
  - Complexity; distributed systems
- Solved … with data
  - We need data … good data, to help us deal with complex, distributed systems
  - We need help distilling that data for human consumption

MOC Monitoring Infrastructure

Collect & Consolidate

- Datacenter,
- Physical,
- Network,
- Cloud Management Layer

to enable analysis cross layers



Servers

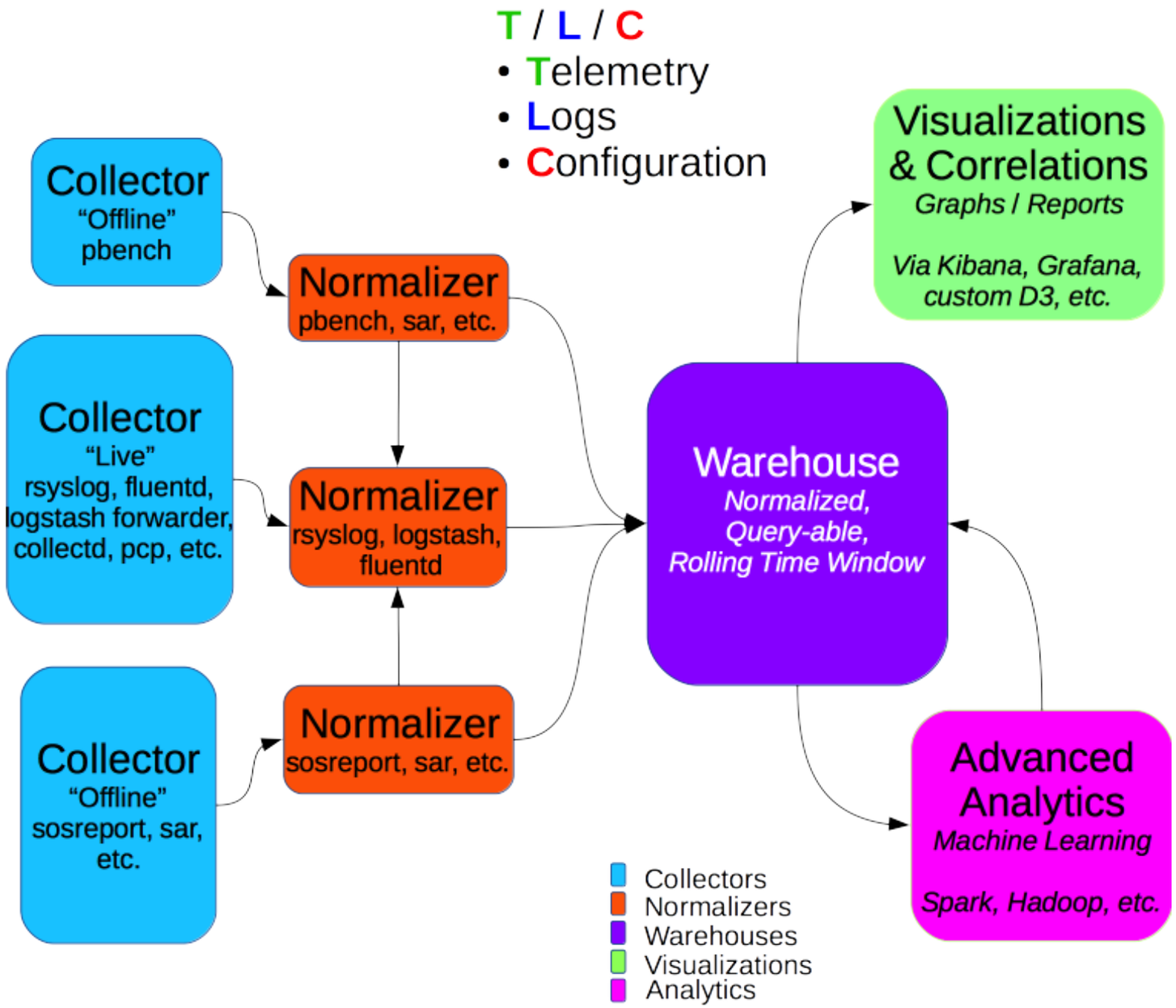
Switches

Datacenter Layer

Water Pumps IRCs Chilling Towers

Approach

- Working on data collection
  - Sprinkle TLC (Telemetry, Logs, Configuration)
    - Fluentd, rsyslog, collectd, prometheus, etc. (See pbench)
  - Data Model, Observability and automation lightning talks
  - OpenShift Aggregated Logging and RHV, OpenStack to follow
- Analytics
  - See OpenShift and the insightful appli

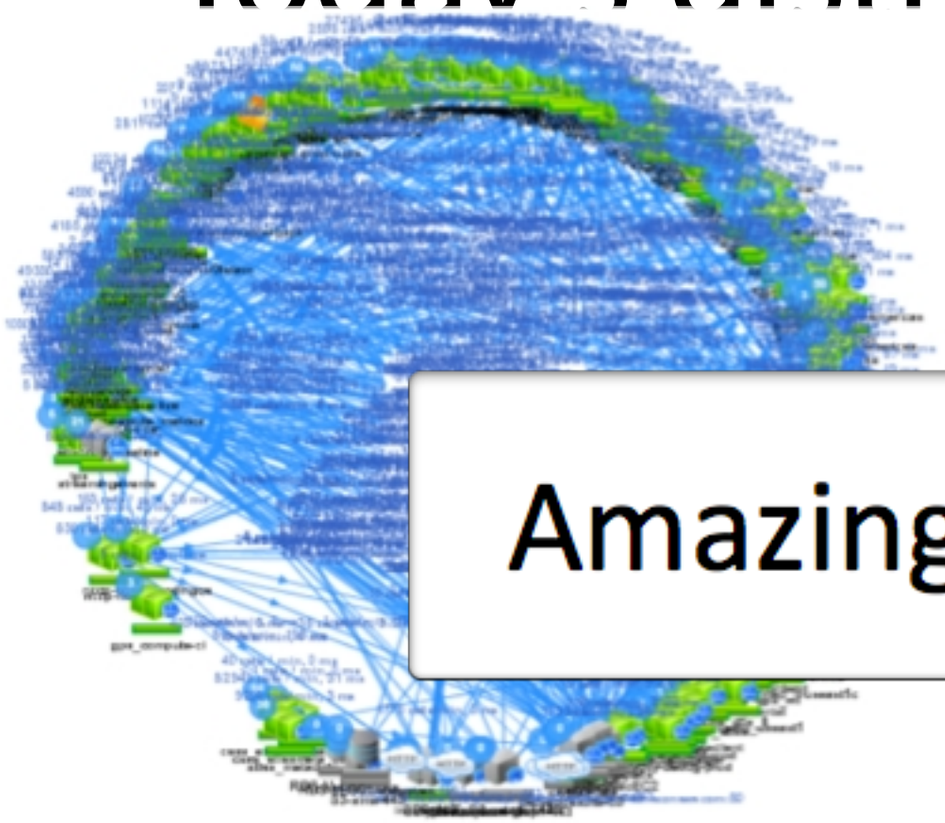


Workflow-centric tracing in OpenStack

- **Raja Sambasivan**

Ata Turk, Joe Talerico, Peter Portante, Orran Krieger

Today's distributed systems



Amazingly **complex**

E.g., Netflix

E.g.,Twitter

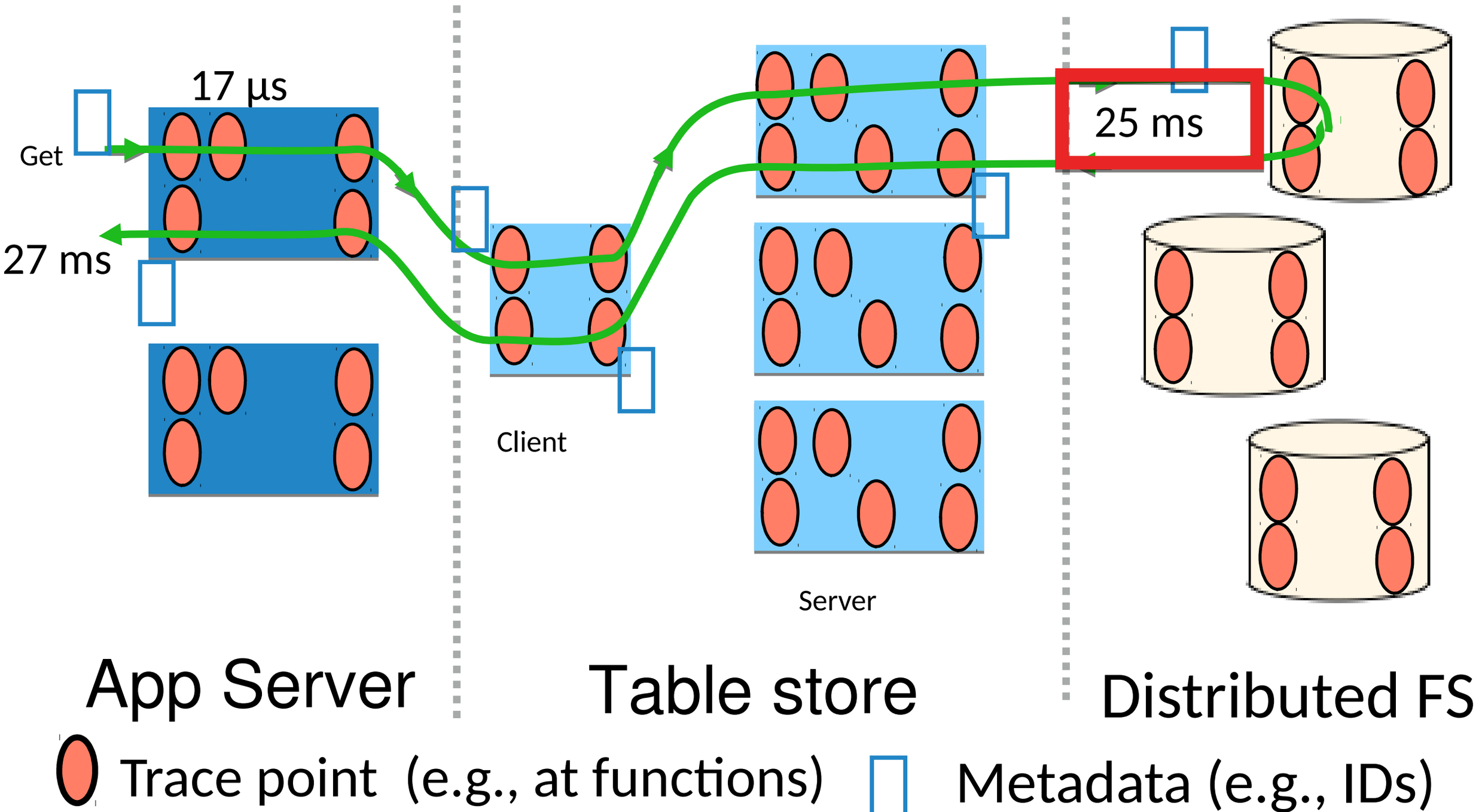
Machine-centric tools **insufficient**

{ GDB, gprof, strace, linux perf. counters

Twitter "death star": https://twitter.com/adrianco/status/441883572618948608
Netflix "death star": http://www.slideshare.net/adriancockcroft/fast-delivery-devops-israel

Workflow-centric tracing

Provides the needed coherent view



Explore tracing's potential in OpenStack

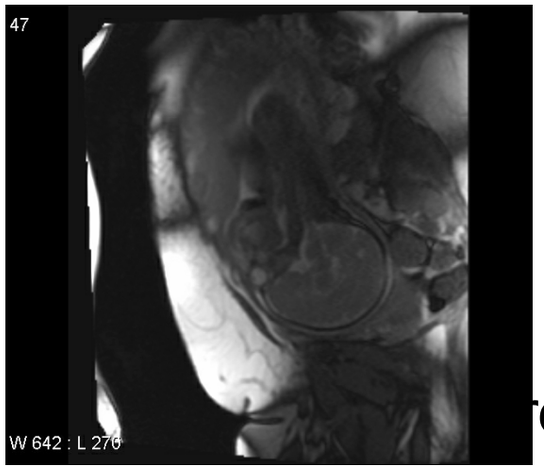
- Implement tracing in OpenStack
  - Use OSProfiler as a starting point
- Explore applicability of existing diagnosis tools (e.g., Spectroscope [NSDI'11])
- Explore new tools for new classes of problems and different operational tasks

Red Hat Collaboratory

- Mix & Match
- HIL & BMI (and QUADS integration)
- Big Data Analytics and Cloud Dataverse
- Datacenter-scale Data Delivery Network (D3N)
- Monitoring, Tracing, Analytics …
- OpenShift on the MOC
- Accelerator Testbed

End-to-end POC: Radiology in the cloud targeting OpenShift with accelerators

Radiology in the cloud workflow

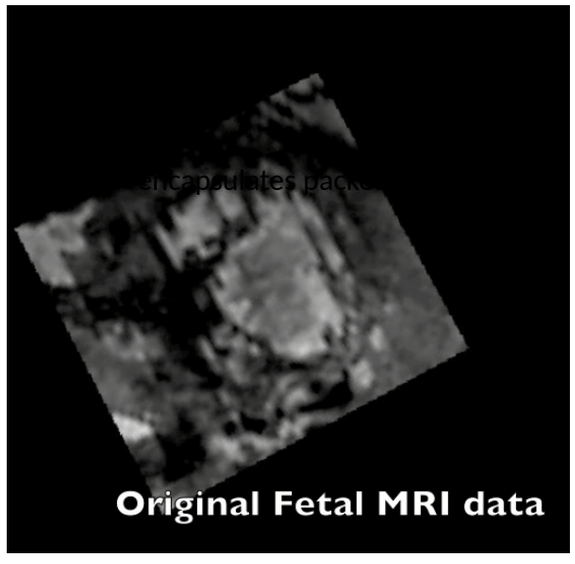


resized,
cropped
96x96
50 slices

APP

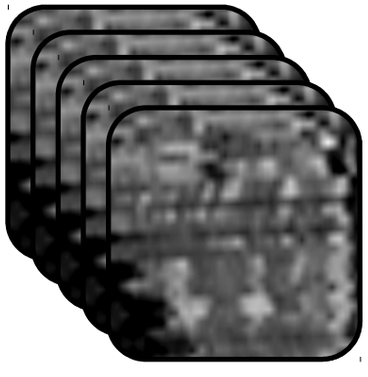
IRTK

24hrs



Fetal Image Reconstruction

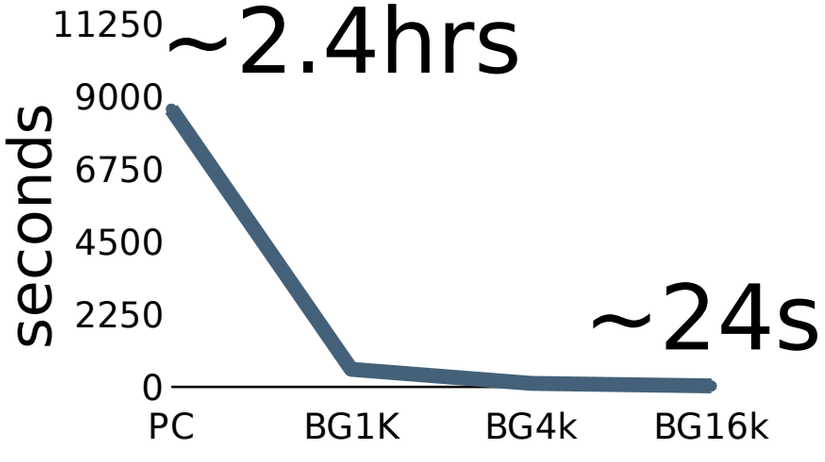
synthetic
1024x1024
200 slices



Exemplar

AFNI

MPI



Concluding remarks

- MOC a functioning small scale cloud for region today:
  - http://info.massopencloud.org/blog/user-account-request-form/
- Key driver is the OCX Model:
  - Key enablers going on in OpenStack
  - Could become important component of clouds
  - Major research challenge & opportunities: presented a small sampling
  - Enabling research to co-exists with production:
    - real data, real users, real scale
- Combining innovation open source, research, cloud (CI/CD)