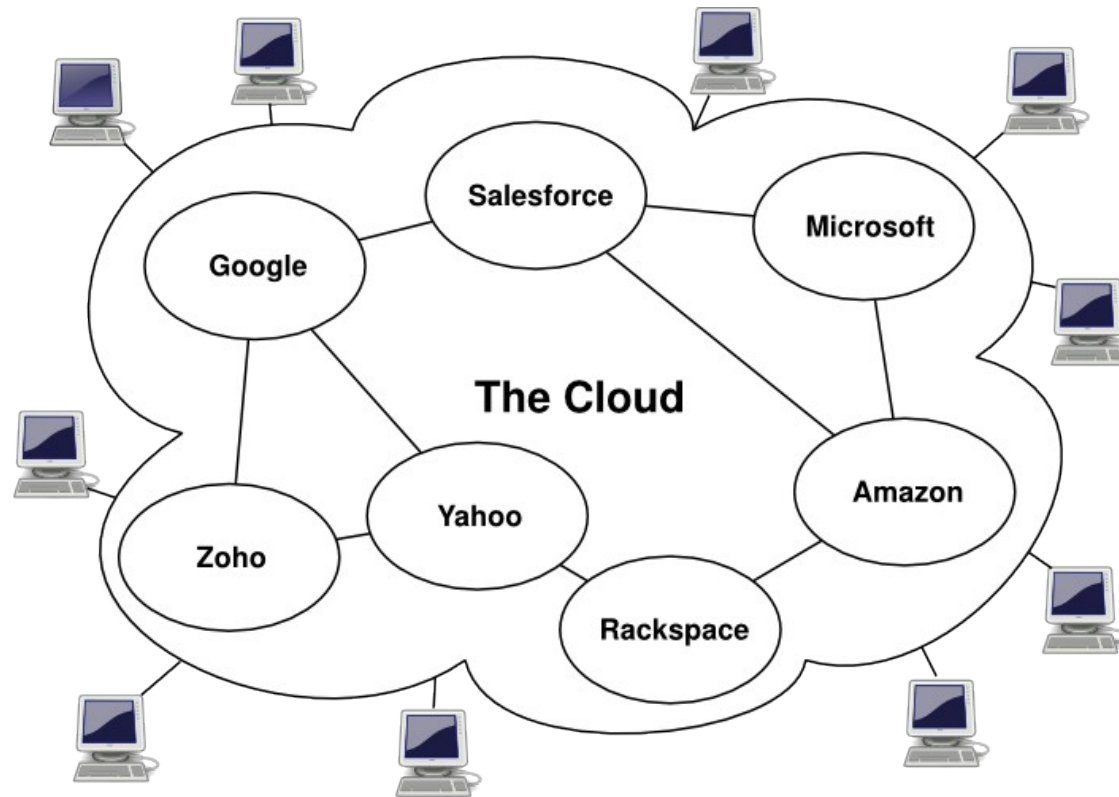# Research in the Clouds Storage and Services

UAB IT Research Computing Day
September 26, 2013

John-Paul Robinson
UAB IT Research Computing

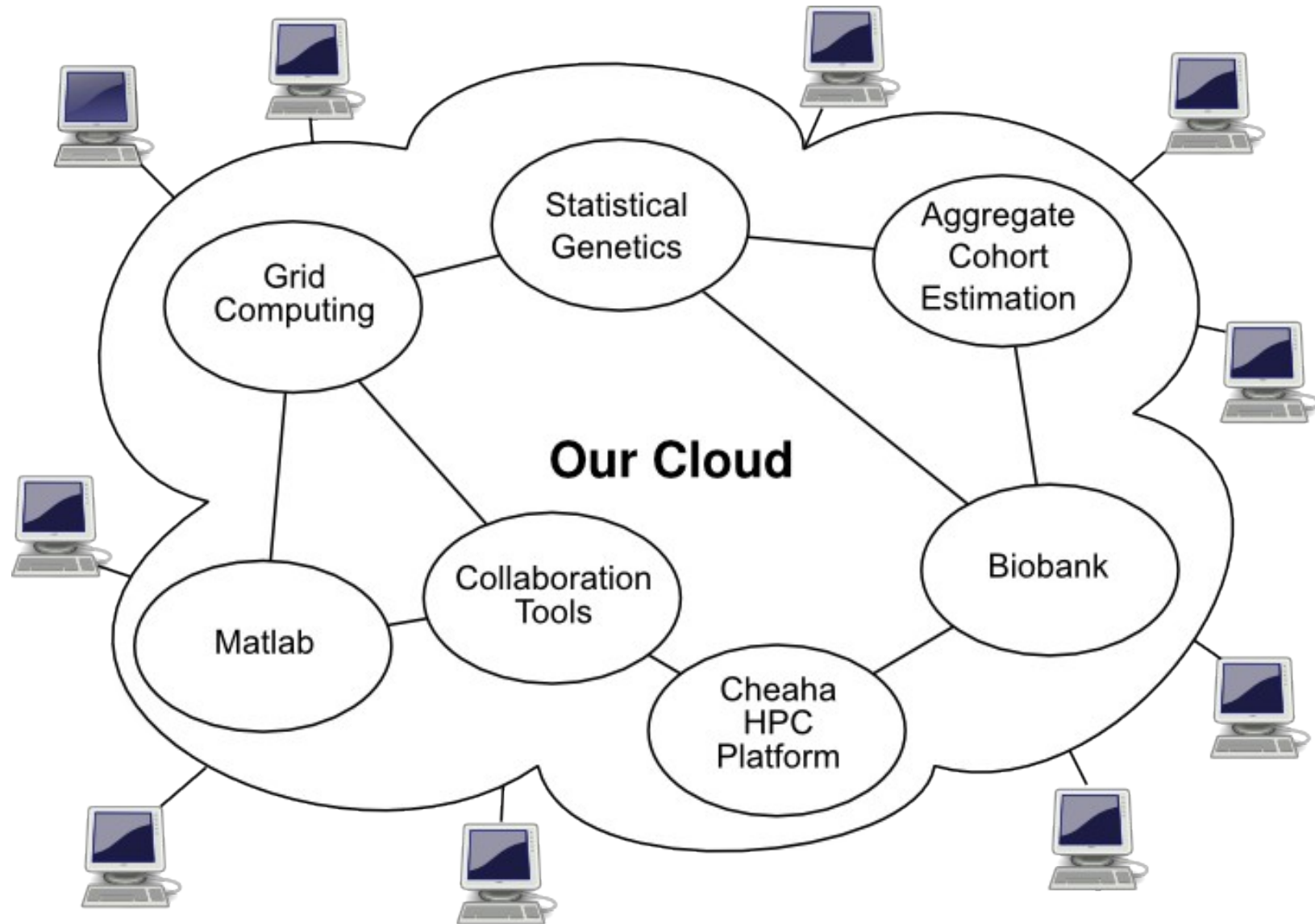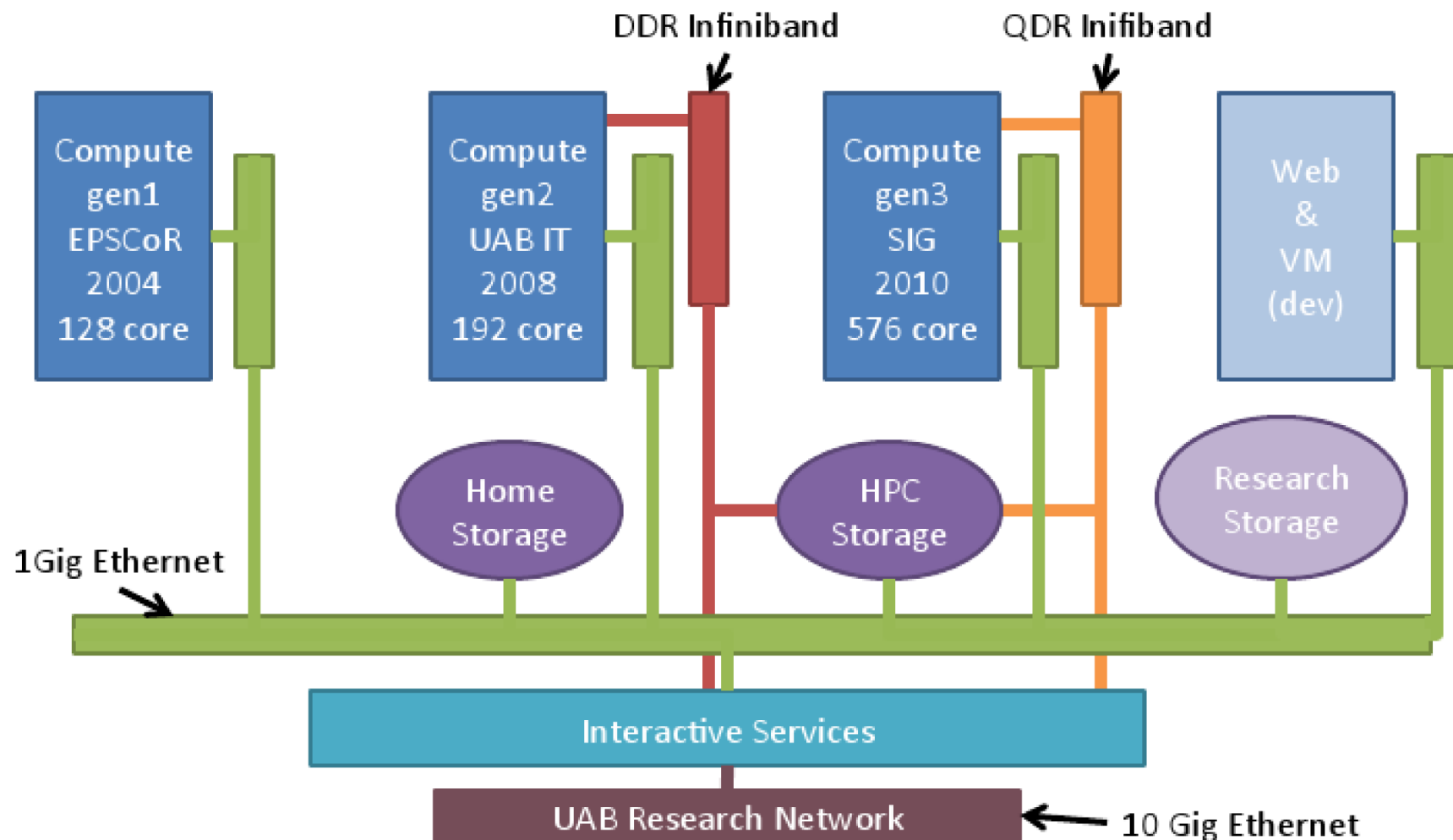# Back in 2010....

What's a Cloud?
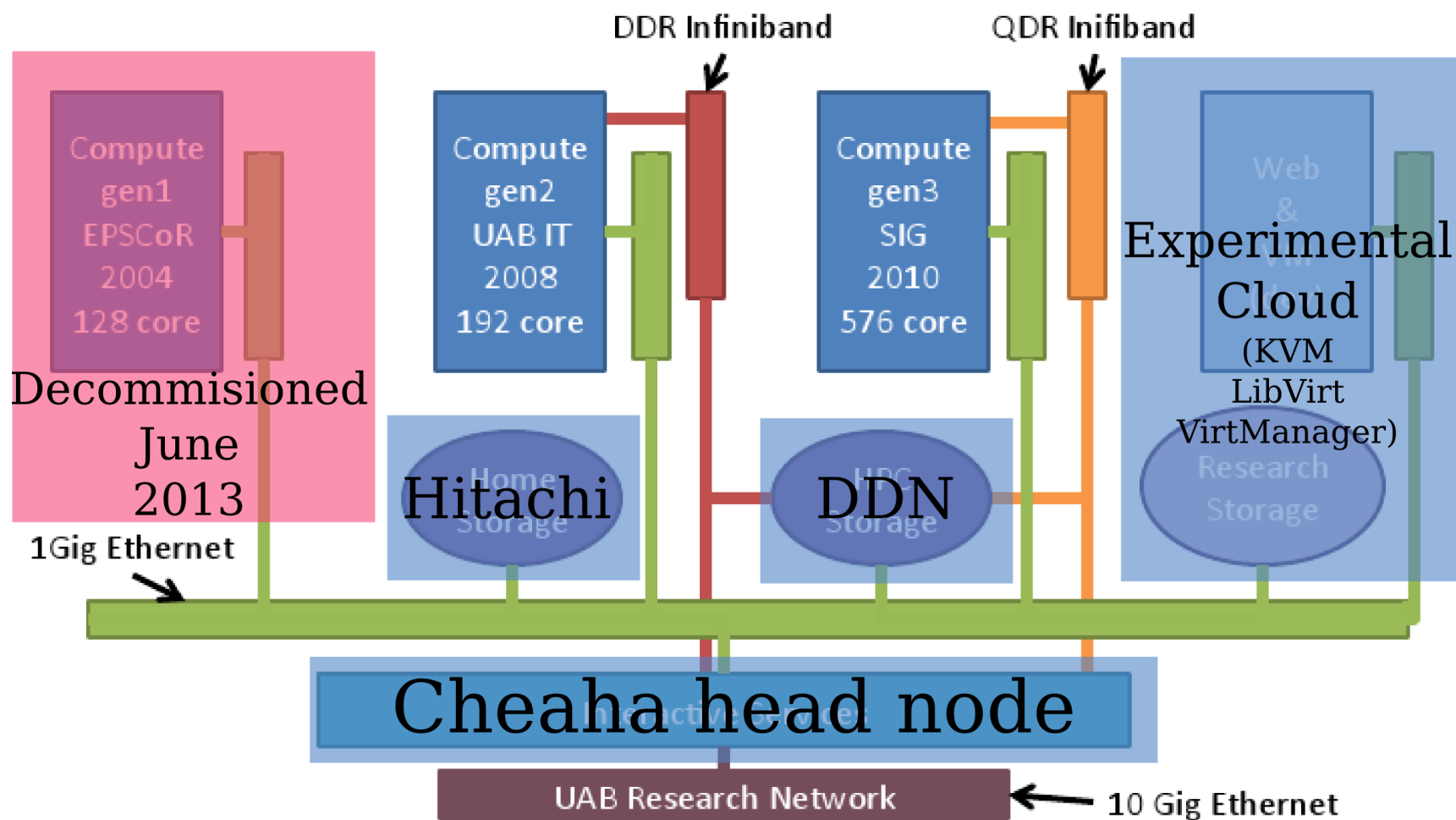
UAB Research "Cloud" circa 2010

Back in 2011...

Research Hardware Platform

RCS Implementation Details

Research Computing System 2011

Research Computing System ^{NextGen}

DDR Infiniband   QDR Inifiband

Compute
gen2
UAB IT
2008
192 core

Compute
gen3
SIG
2010
576 core

Cloud
Storage
(Ceph)

HPC
Storage

Cloud Services
(OpenStack, Swift, S3)

UAB Research Network   10 Gig Ethernet
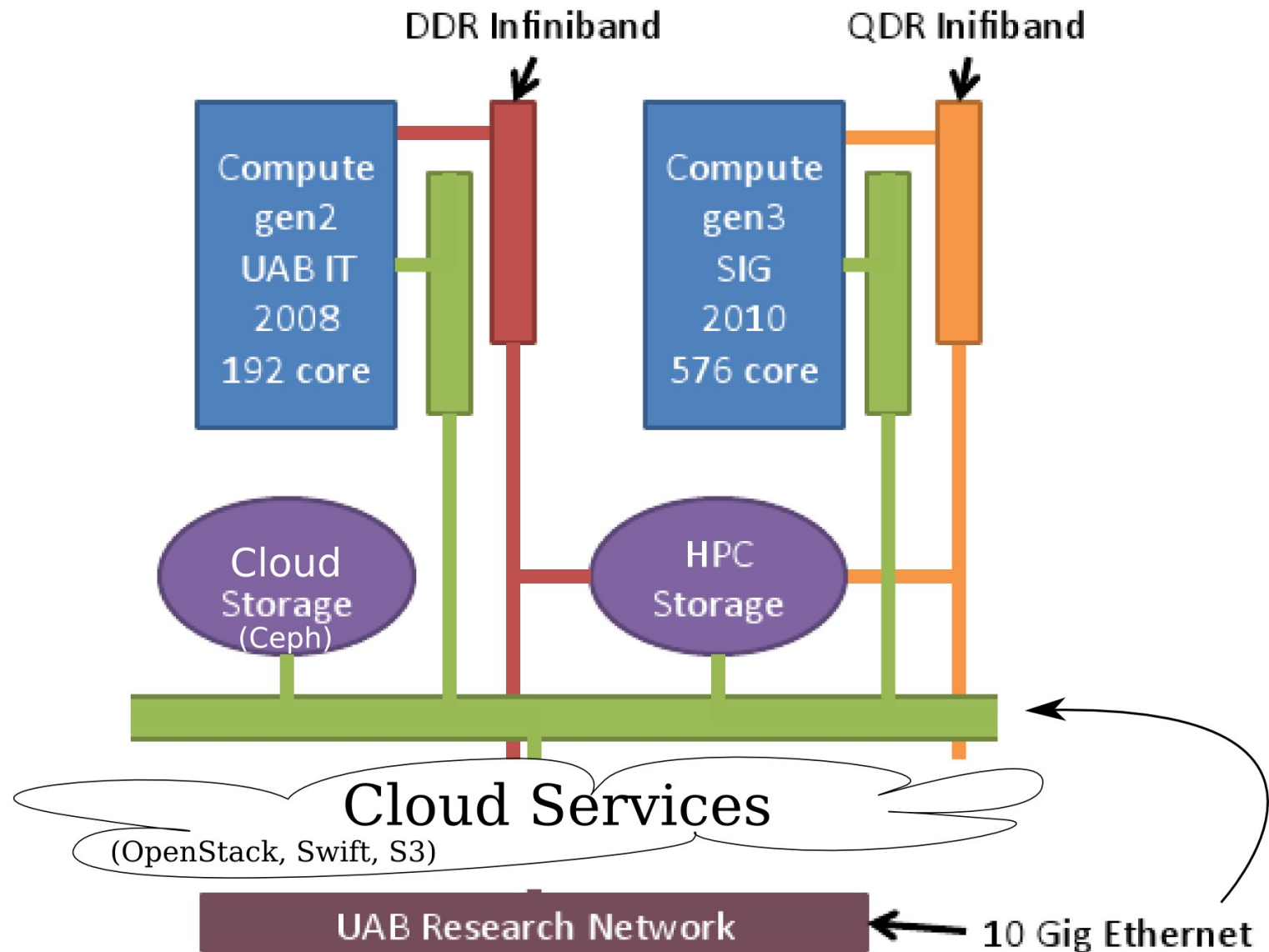
What does it mean to use the cloud today?

- It's not always about being "out there"

- It's about autonomy

- It's about choice

- It's about self direction

- It's about a lot of things that we do already

- It's about how we can work *with* technology

What are the Apps?

What motivates me...

- Search Me
  - My life online
  - Search all my data: research notebook, email, web history, command history, document history, listening history,...
  - Lot's of data, lot's of computing
- Whatcha' Doin'?
  - GIS data adds new temporal and spatial dimensions
- Learning Landscapes
  - Construct a topology of my knowledge space
  - A structure for my narrative
  - AI bots help me find and fill in holes and organize my learning

I really rather buy this...

Like DropBox...

Wherever you are

Put your stuff in Dropbox and get to it from your computers, phones, or tablets. Edit docs, automatically add photos, and show off videos from anywhere.

Share with confidence

Share photos with friends. Work with your team like you're using a single computer. Everything's automatically private, so you control who sees what.

Safe and secure

Even if your phone goes for a swim, your stuff is always safe in Dropbox and can be restored in a snap. Dropbox secures your files with AES-256 bit encryption and two-step verification.

DURACLOUD™

Overview Solutions Features Learn More

ONE CLICK TO THE CLOUD: MULTIPLE COPIES, LOCATIONS AND PROVIDERS

Sharing
ENABLE COLLABORATION WITH ONE CLICK

▶ NO WAITING: TRY IT NOW

▶ FIND OUT MORE

▶ PURCHASE

▶ **NEW** LOW-COST STORAGE

VISIT SPECTRUM
OF DURACLOUD
USE CASES

See comparisons of costs vs benefits
Find a use case similar to your institution
Look at simple vs robust use cases

UAB Galaxy Platform for NextGen Sequencing

✉ Newsletter

CLC bio

CLC GENOMICS SERVER

A flexible enterprise level infrastructure and analysis backbone for next generation sequencing data analysis.

Overview
GENOMICS SERVER

Centralized Bioinformatics Analyses

CLC Genomics Server is an advanced high-throughput sequencing focused three-tier solution offering secure, powerful, and flexible bioinformatics computing on a server-architecture located centrally in your organization.
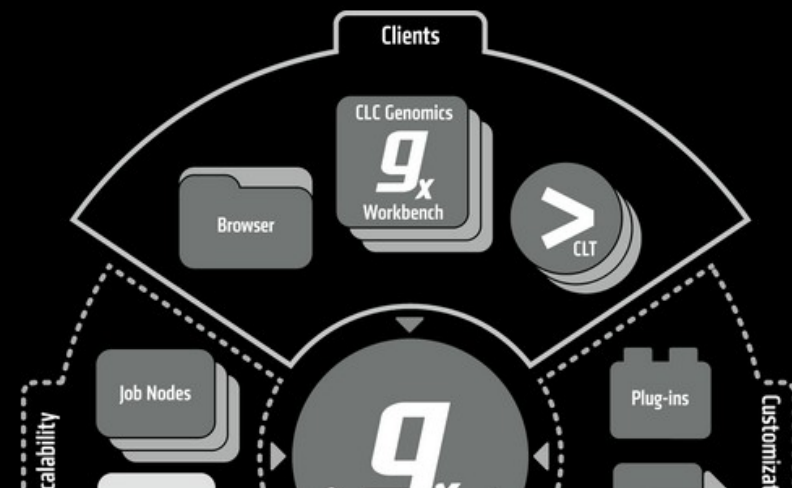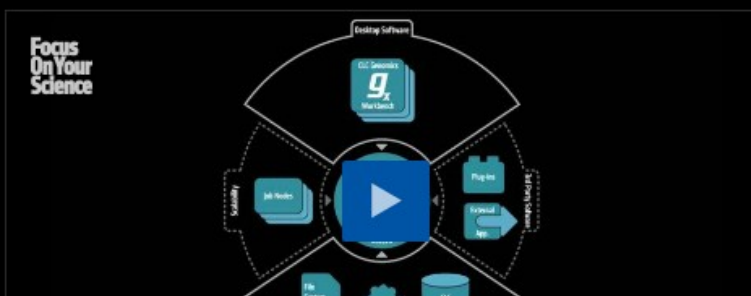
Overview

This is the home page of the Section on Statistical Genetics (SSG) Mobile Lab Virtual Machine we call *ssglabvm*. If you are unfamiliar with the concept of virtualization, you can think of this "virtual" machine as a custom computer for statistical genetics research, analysis, and demonstration that you can run on your own computer.

What programs are installed?

Here is a short, representative, but non-exhaustive list:

- PLINK
- R
- Bioconductor
- Bowtie
- SAMtools
- Haploview
- R/maanova
- minimac
- PLINK/SEQ
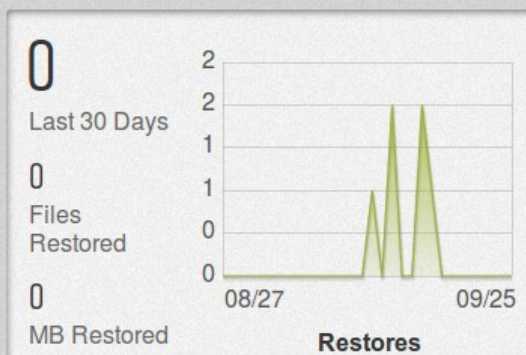- MACS
- CpGassoc
- R/qtlbim
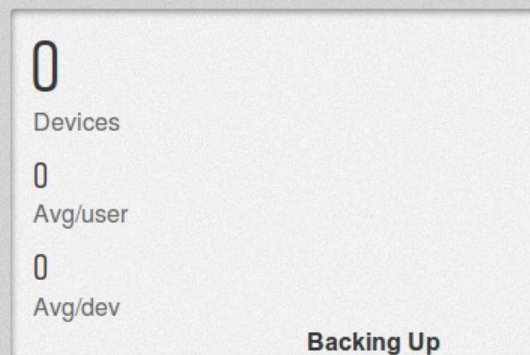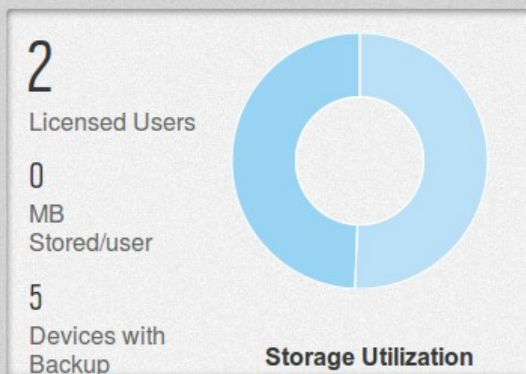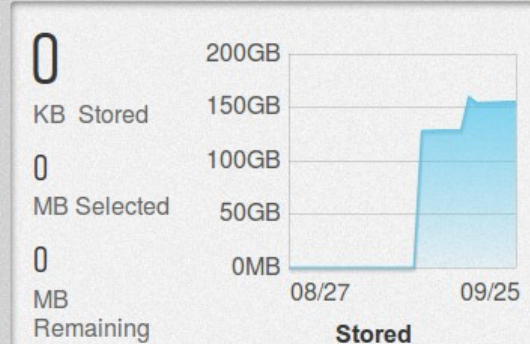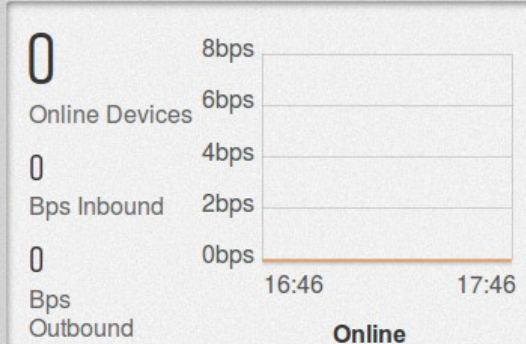- HDBSTAT
- The Power Atlas

Why a virtual machine?

While it is true that anyone can download and install these freely available programs themselves on their own computers, doing so takes time and in some cases, specialized system administration or programming expertise. Though many software come with pre-compiled binaries for different operating systems, some statistical genetics programs may require non-trivial effort to compile and install. Furthermore, after the initial installation, the software should ideally be rigorously tested, by both the system administrator and by the data analyst or statistical programmer. These challenges make the creation and availability of an installed, tested, and reliable collection of statistical genetics programs a worthwhile exercise for us.
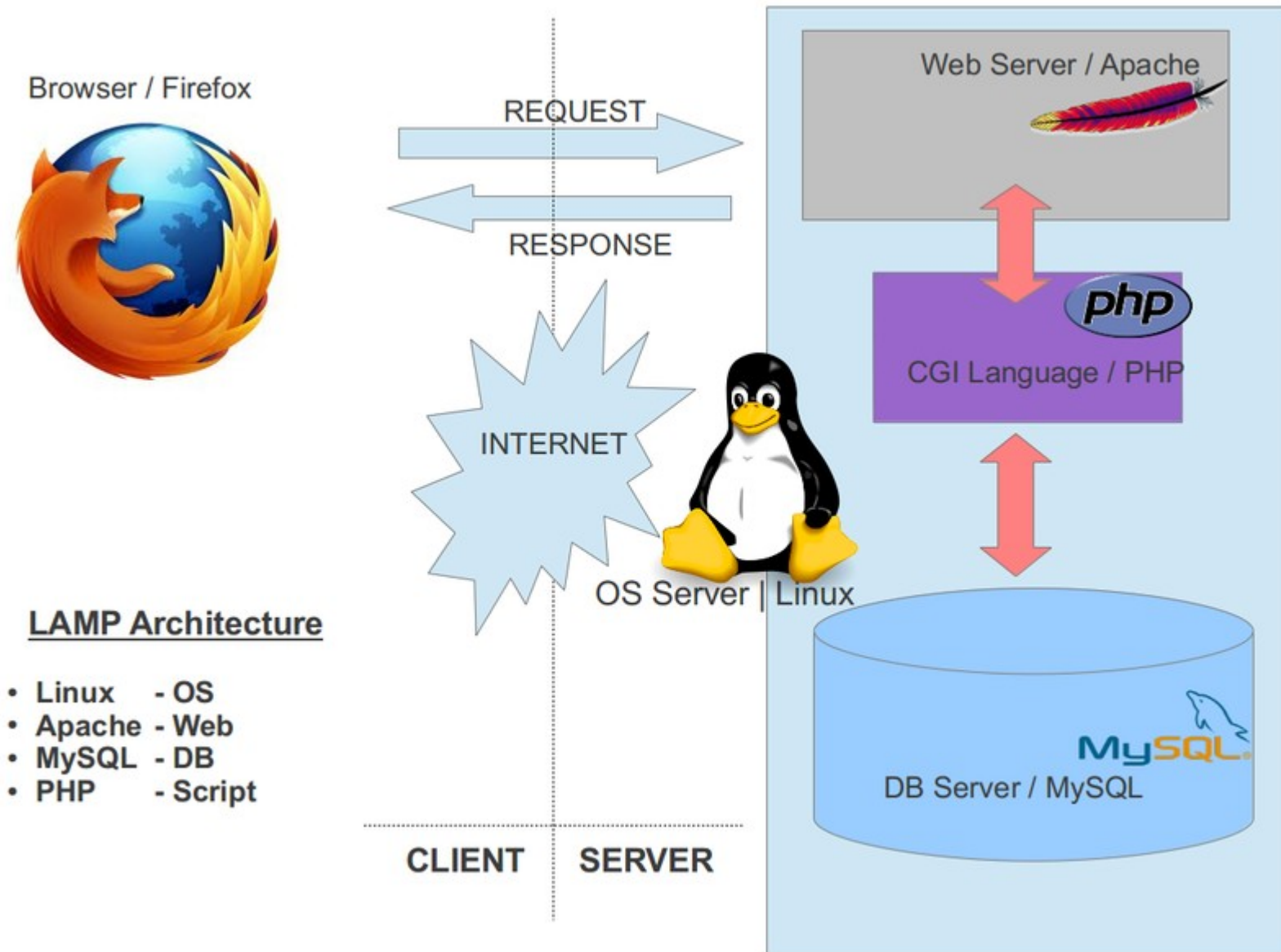
Dashboard - CrashPlan PROe f...

CRASH**PLAN**PROe

Admin Admin   Sign Out

Smart Search

Dashboard

Alerts   4 1 3

Organizations

Users

Devices

Destinations

Settings

Licensing

My Profile

Online

0 Online Devices

0 Bps Inbound

0 Bps Outbound

8bps
6bps
4bps
2bps
0bps
16:46   17:46

Stored

0 KB Stored

0 MB Selected

0 MB Remaining

200GB
150GB
100GB
50GB
0MB
08/27   09/25

Storage Utilization

2 Licensed Users

0 MB Stored/user

5 Devices with Backup

Backing Up

0 Devices

0 Avg/user

0 Avg/dev

Restores

0 Last 30 Days

0 Files Restored

0 MB Restored

2
1
0
08/27   09/25

User License Coverage

14 Days Left In Free Trial

2 Licensed Users

10
8
6
4
2
0
15 days   30 days   90 days

Building and Using LAMP Stacks

Browser / Firefox

REQUEST

RESPONSE

INTERNET

OS Server | Linux

LAMP Architecture

- Linux   - OS
- Apache - Web
- MySQL   - DB
- PHP   - Script

CLIENT   SERVER

Web Server / Apache

CGI Language / PHP

DB Server / MySQL

Tools to Build Cloud-ready Systems

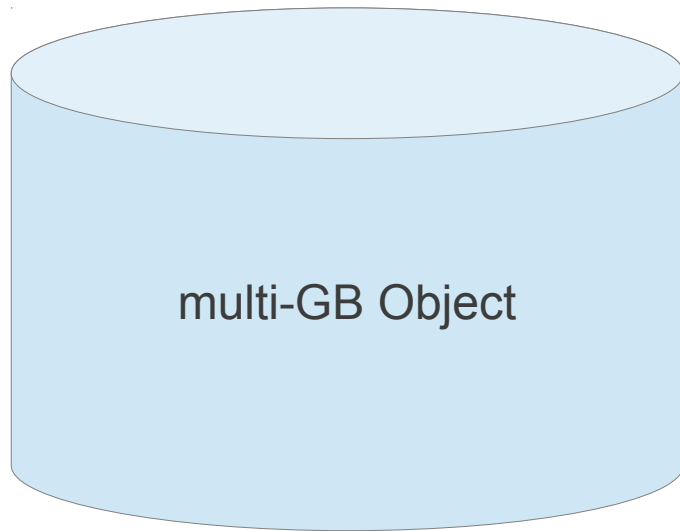# Free Downstream Developers to Build at the Speed of Research
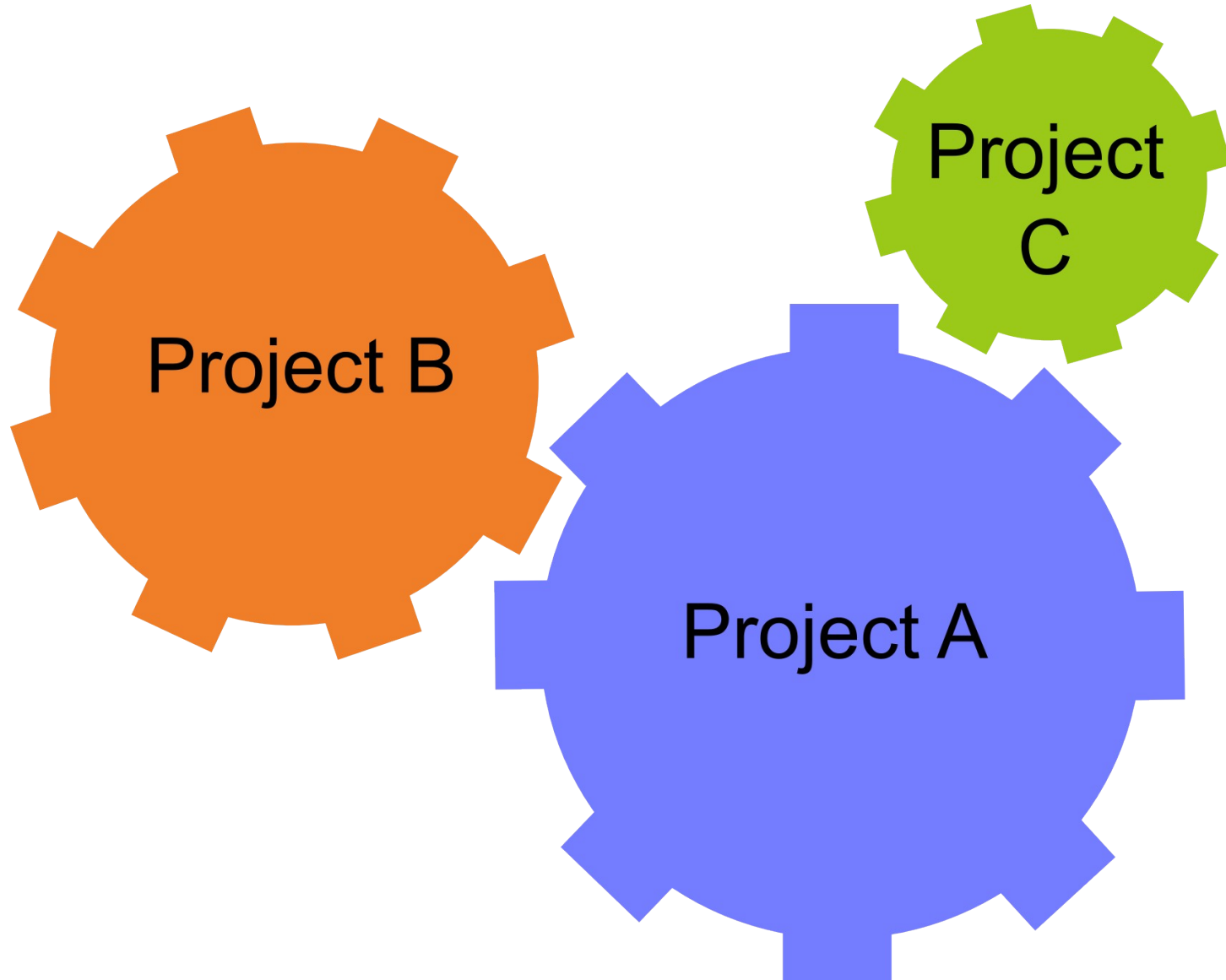
What Apps are we building?
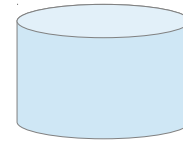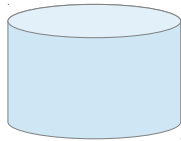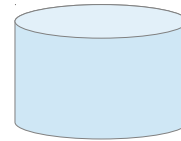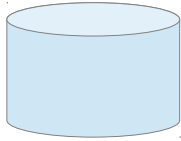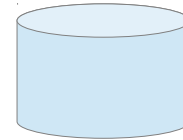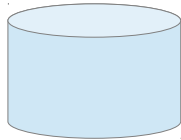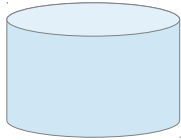
What Apps are we building?

Research Storage!

Research Storage!

multi-GB Object

We like containers...

Research Storage!
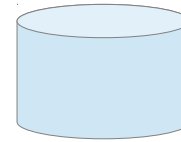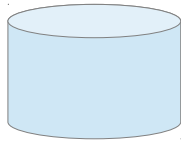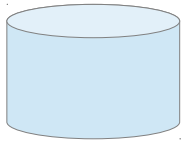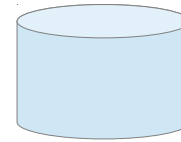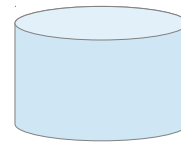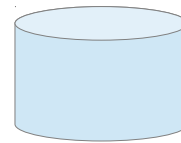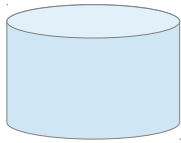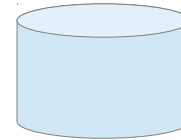
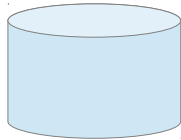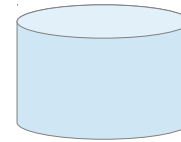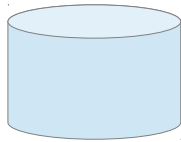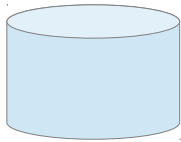# Research Storage!
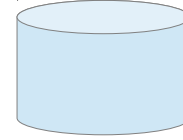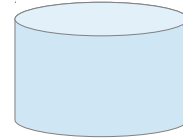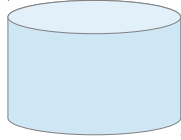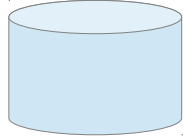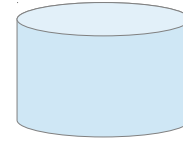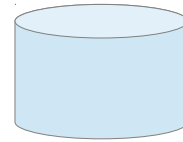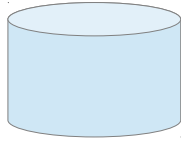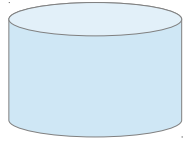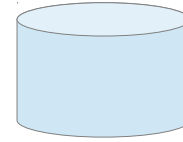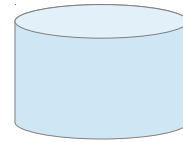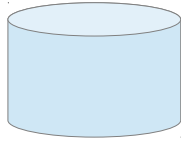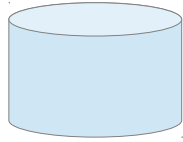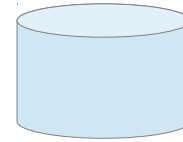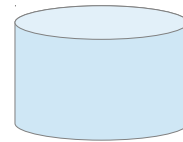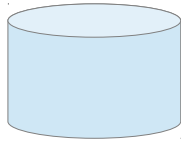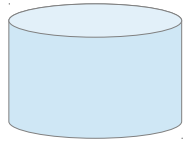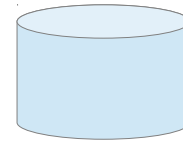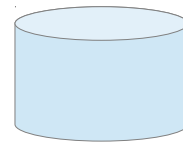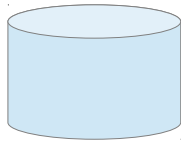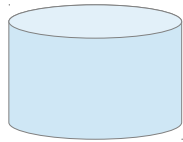
Research Storage!
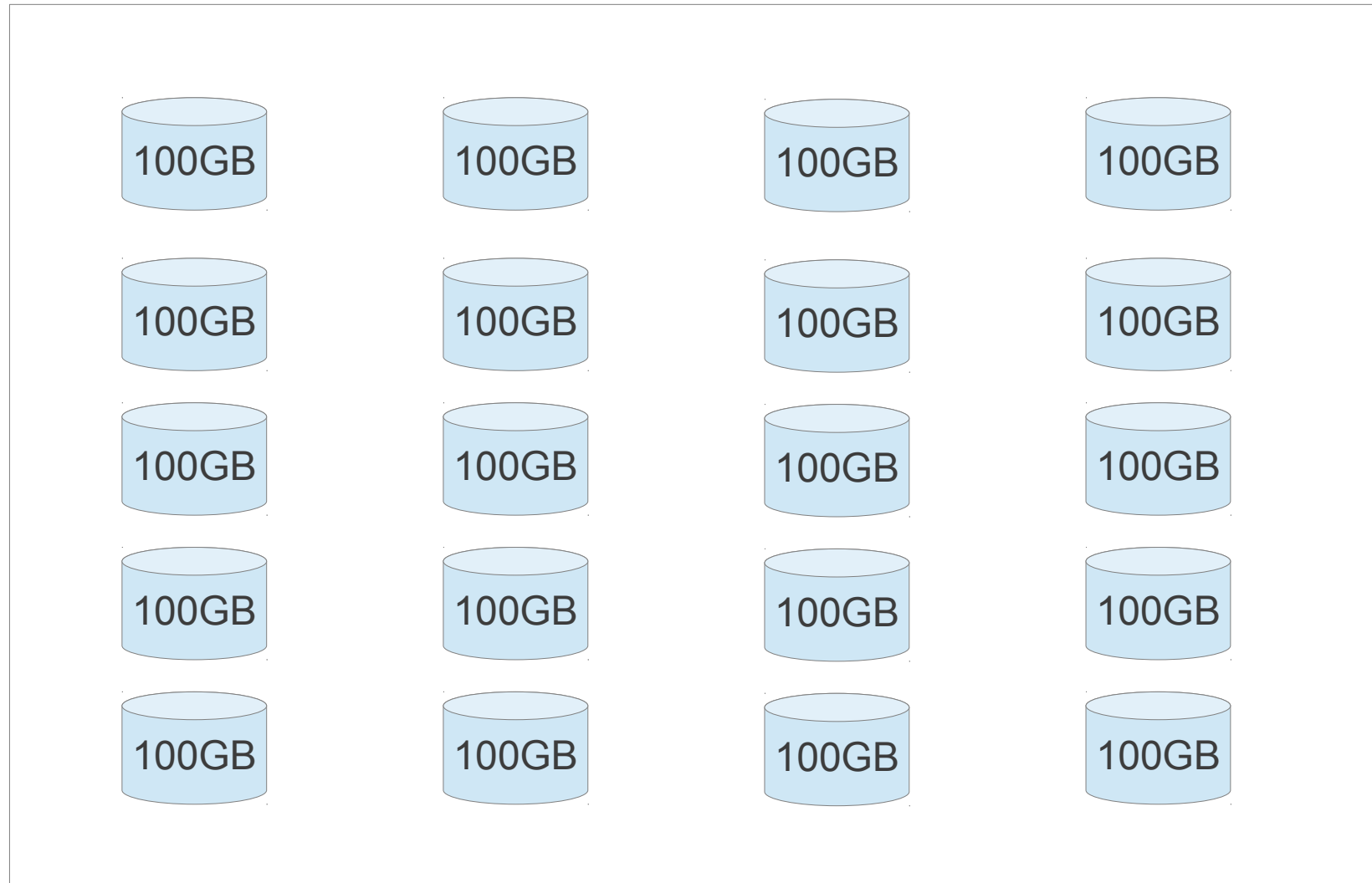
Research Storage!

1 TB-year for RCS Users

Research Storage!
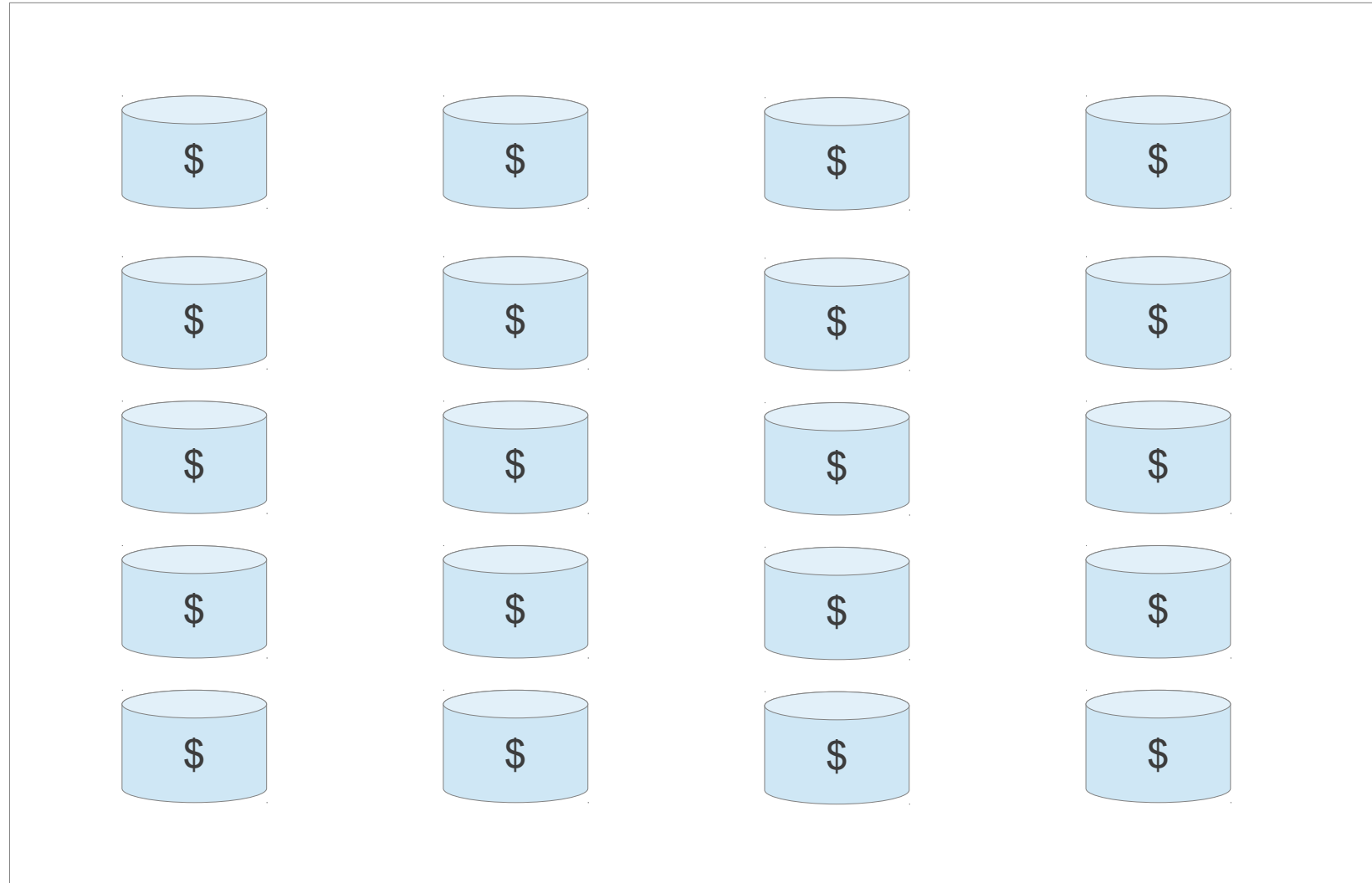
They're like Research Bucks!

Research Storage!

- You can buy more @ $395 per TB-year

- Accounting for use in Gigabyte-hours

- Annual price review

- Designed to scale and trend with public cloud prices

- Additional storage solutions can be built

What Apps are we building?

Research Storage!

Research Backup!

Research Backup!
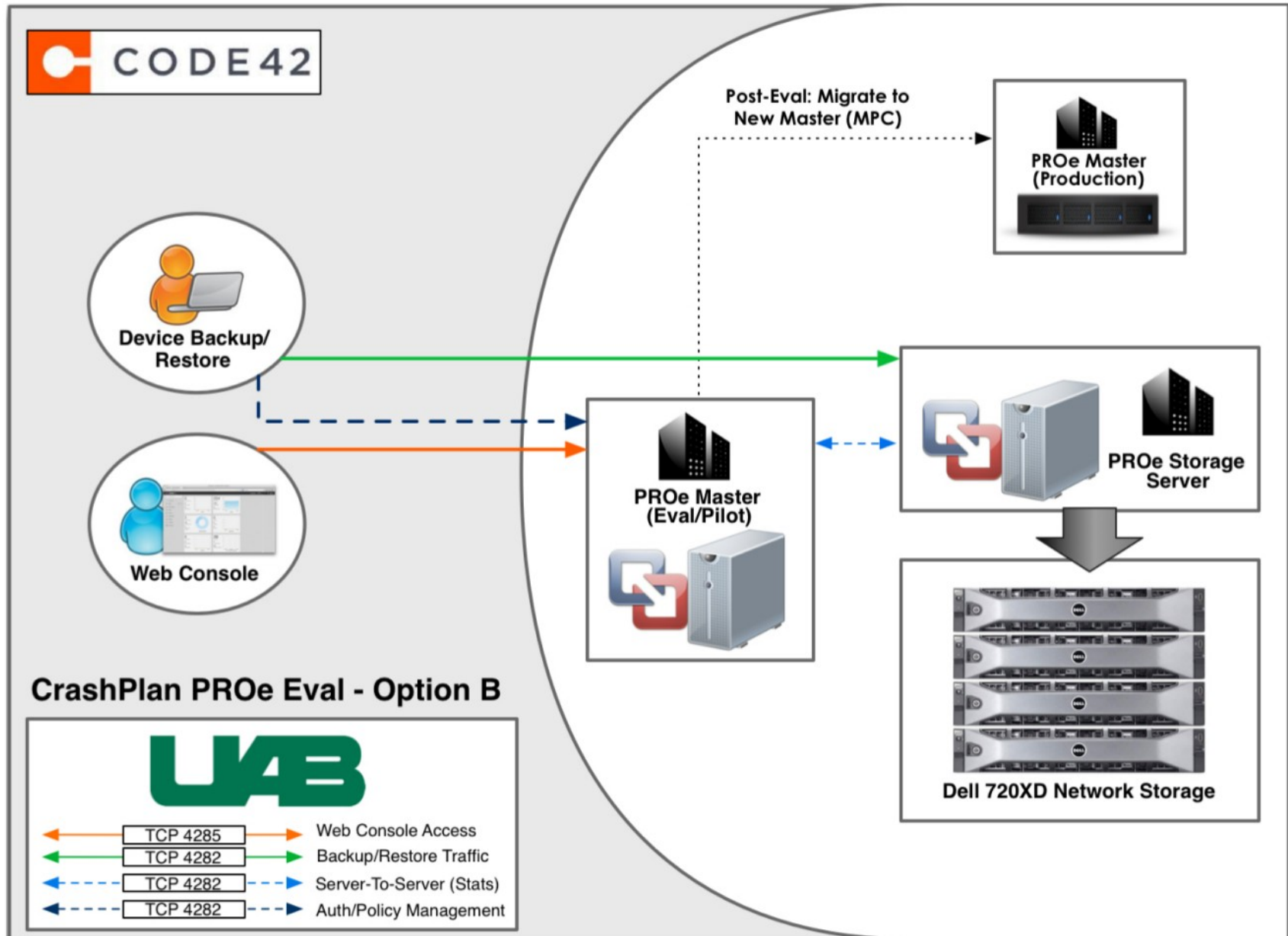
Research Backup!

Research Backup!

- Piloting CrashPlan Enterprise solution using OpenStack Platform

- Street price is $60/user/year for 4 clients

- Cost will also accrue $395/TB/year Research Storage rate for backed up data

- CrashPlan scales to on-site managed service or Internet2 NetPlus unlimited cloud storage

What Apps are we building?

Research Storage!

Research Backup!

Research Whatever!

Research Whatever!

- The sky's the limit

- Custom cloud desktops

- Scale your work to compute environments

- Support your science with tools you control

A peak under the hood..

OpenStack: The Open Source Cloud Operating System
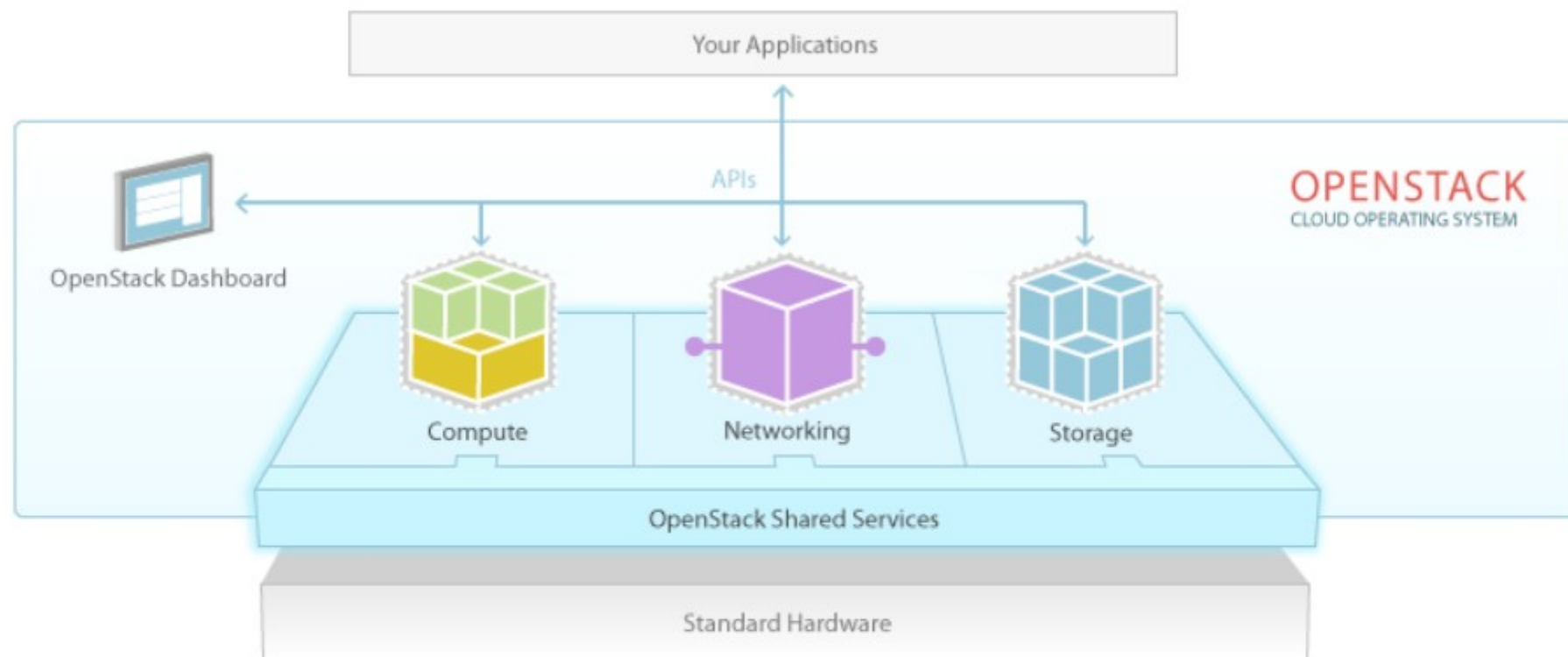
About OpenStack

OpenStack is a cloud operating system that controls large pools of compute, storage, and networking resources throughout a datacenter, all managed through a dashboard that gives administrators control while empowering their users to provision resources through a web interface.

THE SEVENTH OPENSTACK RELEASE
GRIZZLY

A QUICK VIDEO OVERVIEW

Ceph Unified Storage Platform

OBJECTS	VIRTUAL DISKS	FILES & DIRECTORIES

CEPH GATEWAY

A powerful S3- and Swift-compatible gateway that brings the power of the Ceph Object Store to modern applications

CEPH BLOCK DEVICE

A distributed virtual block device that delivers high-performance, cost-effective storage for virtual machines and legacy applications

CEPH FILE SYSTEM

A distributed, scale-out filesystem with POSIX semantics that provides storage for a legacy and modern applications

CEPH STORAGE CLUSTER

A reliable, easy to manage, next-generation distributed object store that provides storage of unstructured data for applications

A peak under the hood...

- OpenStack extends the flexibility of our compute capabilities

- Ceph lets us scale storage at commodity prices

- Developer licenses to tune the platform

- Extend our skills network through open source community

- Vendor support to help us scale

Going Forward

- UAB's research cloud is here

- It's ready for the explorers

- It'll be easier to get to once we map to campus IPs

- Our first cloud App is "research storage" – now for beta testers

- Join us – google "uab openstack plus ceph"

Thanks All Around IT!

- VPIT – still funding :)
- Enterprise Infrastructure Services
    - Computers and Networks
    - People
    - Projects
    - Desktop Support
- Business & Resource Management
    - Storage billing
    - Research Associate position
- Enterprise Information Security
    - Early engagement
    - Patience with our experiments :)
- Academic Technology & Relationship Management
    - MATLAB contract help
    - Cloud contracts review
- Dell and Inktank for their support and partnership