

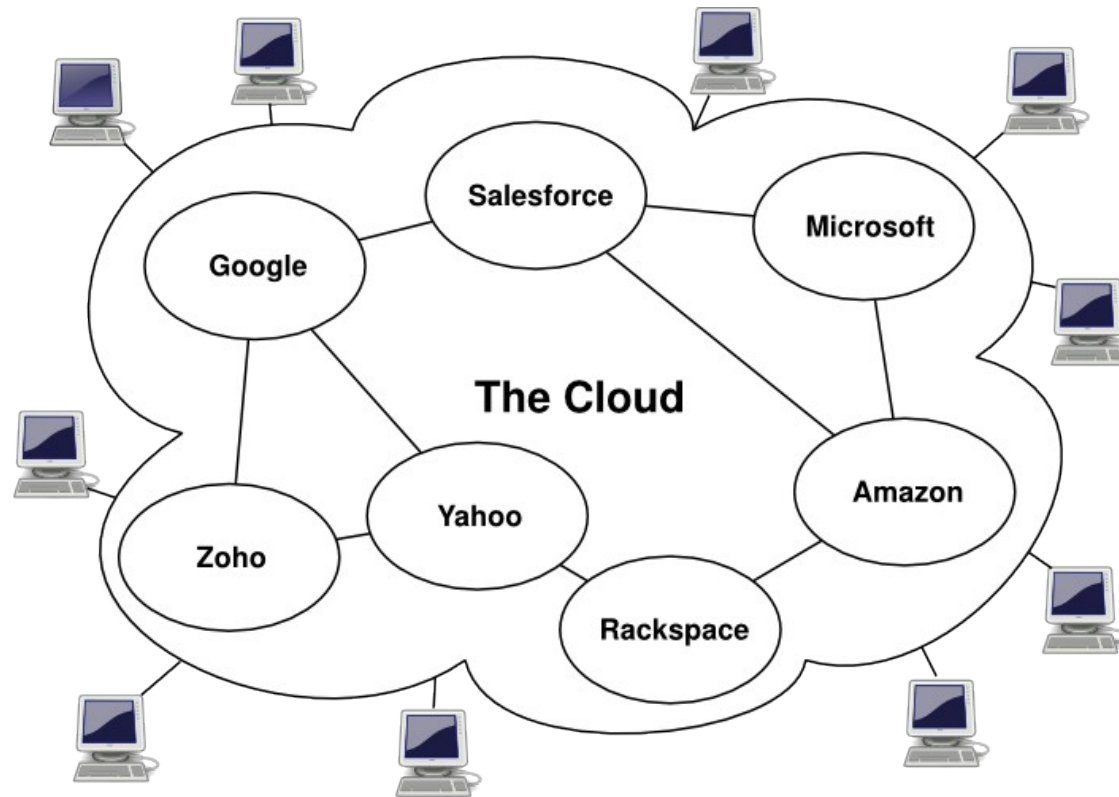
# Research in the Clouds Storage and Services

UAB IT Research Computing Day  
September 26, 2013

John-Paul Robinson  
UAB IT Research Computing

Back in 2010....

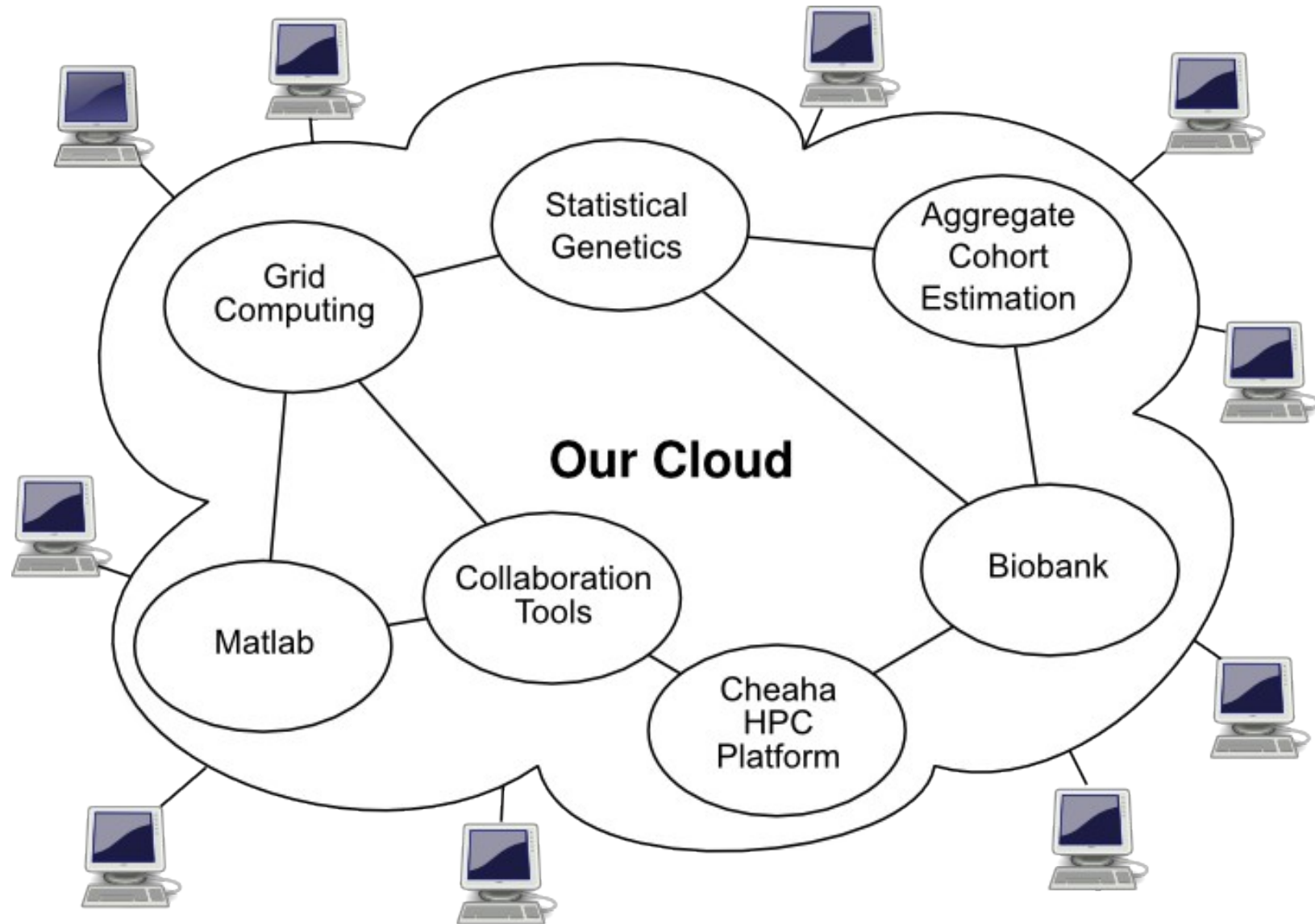
# What's a Cloud?



## ***Wikipedia Sez:***

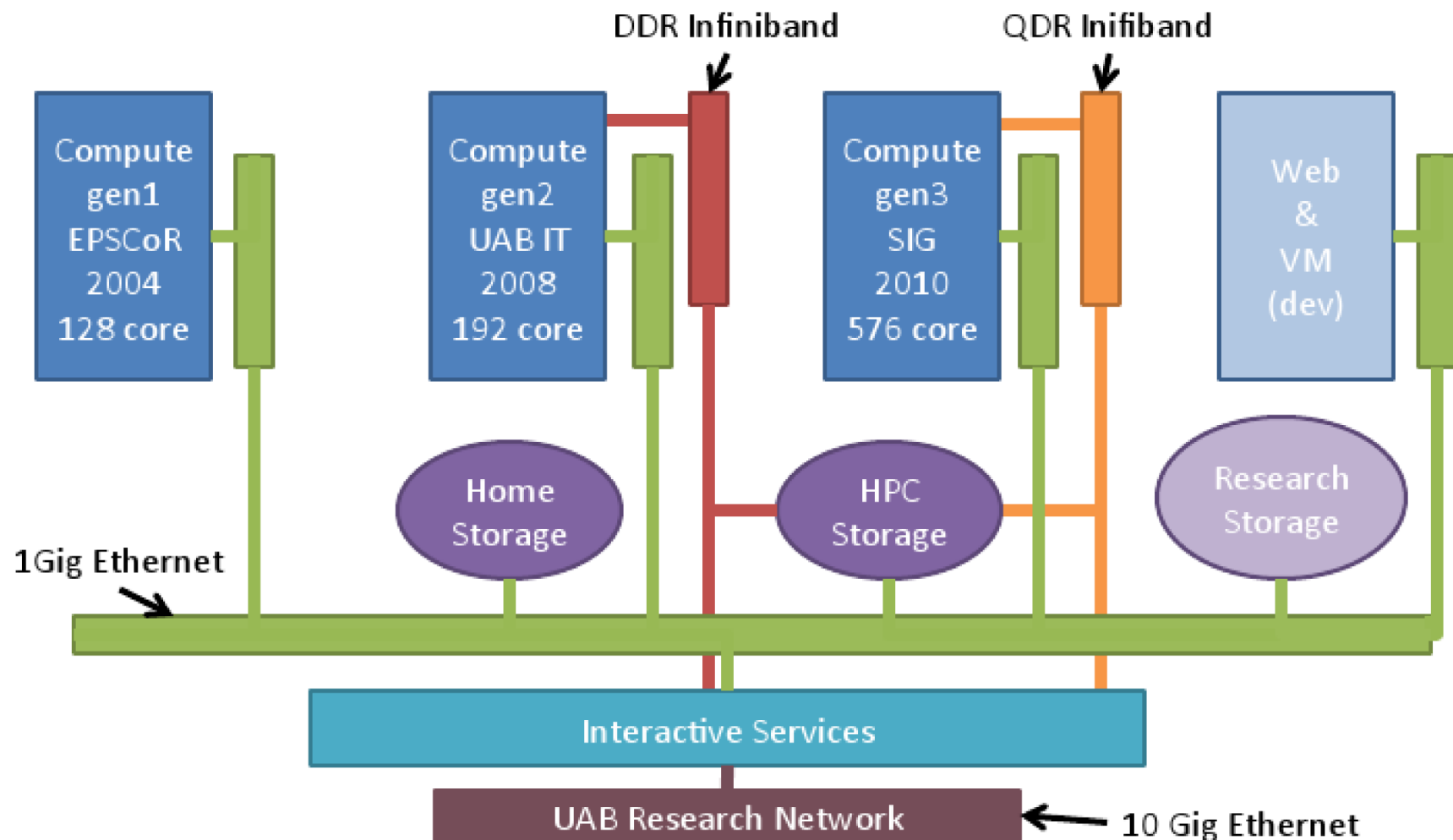
*Cloud computing is Internet-based computing, whereby shared resources, software, and information are provided to computers and other devices on demand, like the electricity grid.*

# UAB Research “Cloud” circa 2010



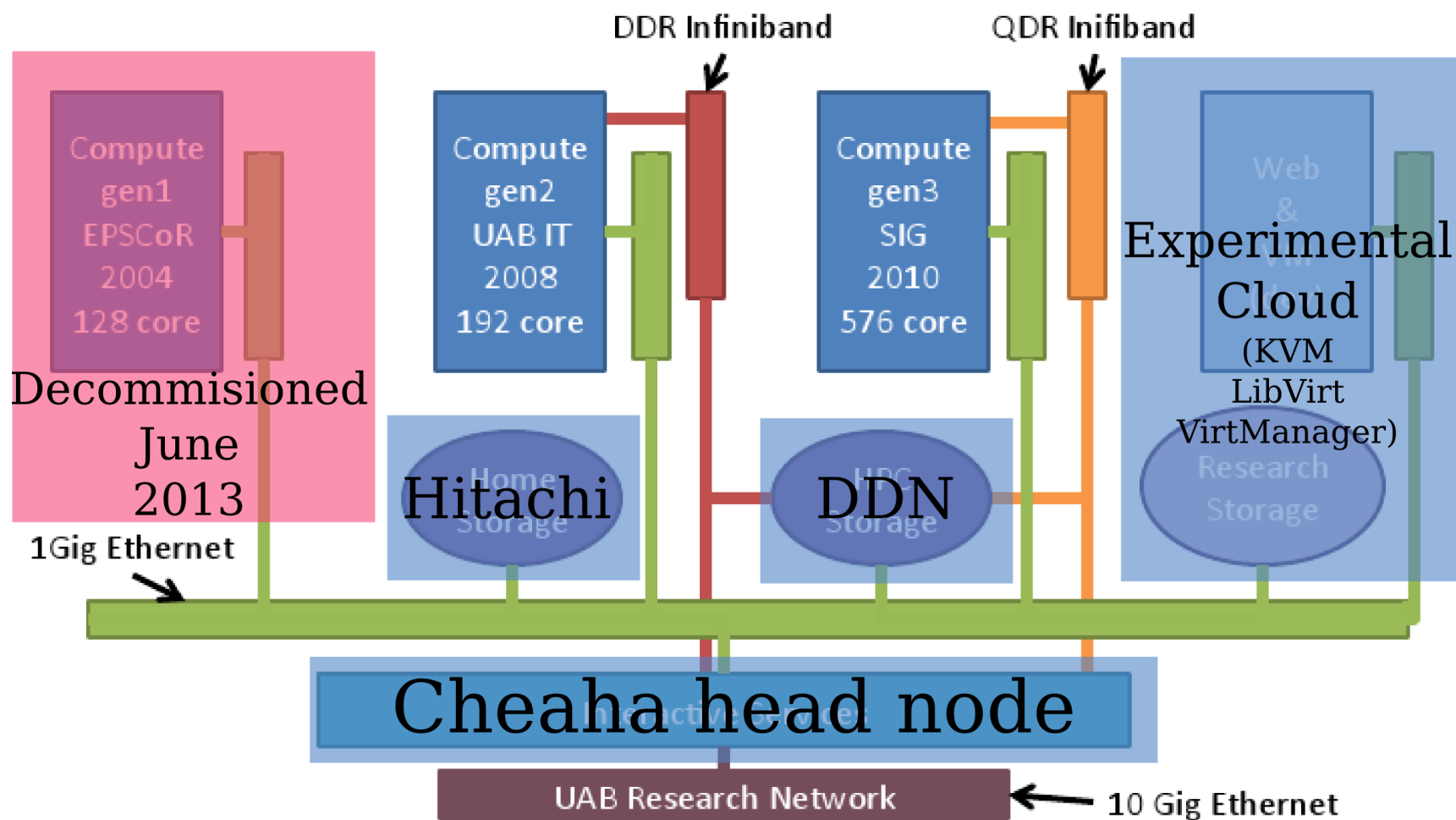
# Back in 2011...

## Research Hardware Platform

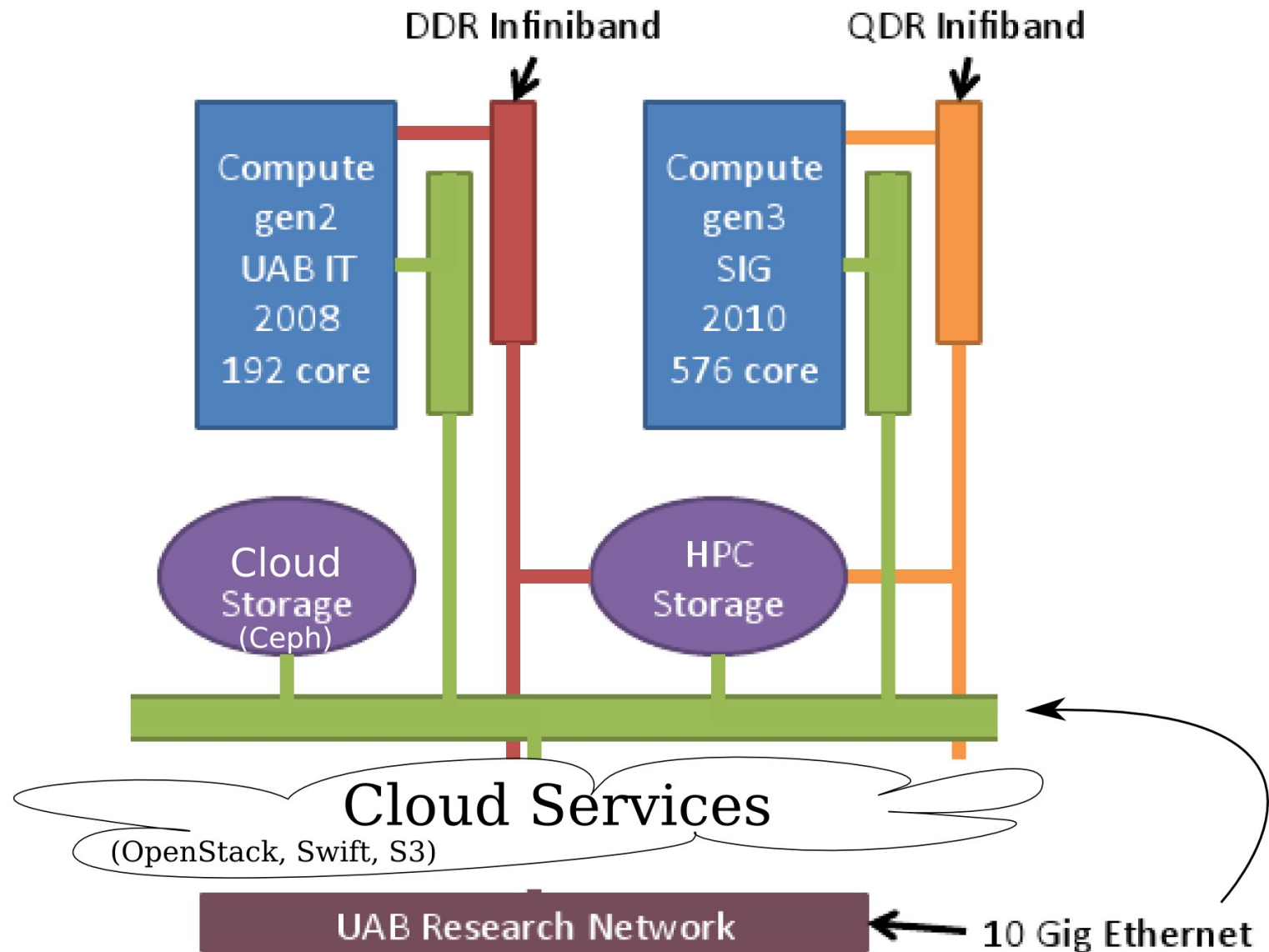


# RCS Implementation Details

## Research Computing System 2011



# Research Computing System<sup>NextGen</sup>



# 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...

drobo

Global | Resources | Partners | Corporate | News | Blogs | Forums |

How it Works **Products** Solutions Customers Support

Buy

Products

Overview

Drobo For Business

Drobo For Professionals

Drobo Dashboard

Capacity Calculator

Drobo Home > Products > Overview



## Drobo FS

### Easiest-to-Use Network Attached Storage

The revolutionary Drobo FS was designed with one purpose in mind: to deliver the best file sharing experience ever. From the moment you plug it in and it instantly appears on your desktop, you'll know there's nothing like it. The all-in-one Drobo FS is perfect for any connected home, home office, or small office environment needing a simple, safe device for sharing and backing up files over the network.

Like all Drobo storage products, Drobo FS provides redundant data protection without the complexities of RAID. Just plug in your Drobo FS and your data protection is all set up with no configuration required. We call this **BeyondRAID™** technology. You'll call it peace of mind.

The Drobo FS storage array is available from \$699 US.

Live Chat with

# 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.





[Overview](#) [Solutions](#) [Features](#) [Learn More](#)

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



▶ 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

**Galaxy / UAB** Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

- [Get Data](#)
- [Send Data](#)
- [Demo Tools](#)
- [ENCODE Tools](#)
- [Lift-Over](#)
- [Text Manipulation](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [Convert Formats](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Get Genomic Scores](#)
- [Operate on Genomic Intervals](#)
- [Statistics](#)
- [Wavelet Analysis](#)
- [Graph/Display Data](#)
- [Regional Variation](#)
- [Multiple regression](#)
- [Multivariate Analysis](#)
- [Evolution](#)
- [Motif Tools](#)
- [Multiple Alignments](#)
- [Metagenomic analyses](#)
- [FASTA manipulation](#)
- [NCBI BLAST+](#)
- NGS TOOLBOX BETA**
- [NGS: QC and manipulation](#)
- [NGS: Picard \(beta\)](#)
- [NGS: Assembly](#)
- [NGS: Mapping](#)
- [NGS: Indel Analysis](#)

**Welcome to UAB Galaxy!**  
Where all you need is a *BlazerId* and a web browser to run NGS analyses on the UAB Cheaha Cluster!

**Local Resources**

- UAB Galaxy Wiki: [Overview](#), [Data Import](#)
- UAB Mailing Lists
  - UAB Galaxy-users ([search archive](#); [subscribe](#)) discuss with other UAB users
  - UAB Galaxy-help ask the UAB admins for help!
- UAB Cheaha Computing Cluster
  - Cluster Hardware ([wiki](#))
  - Request a [cheaha account](#) (needed only for command-line access and bulk data upload)

**Internet Resources**

- [Learn Galaxy](#) - tutorials
- [Galaxy Project](#) user mailing list ([searchable archives](#); [subscribe](#); [post](#))
- [Galaxy Toolshed](#) plug-ins for additional tools that you can request for installation at UAB
- Public Galaxy Server at Penn State (PSU): [UseGalaxy.org](#) (more tools, but small disk quotas)

**Brought to you by**

- UAB IT Research Computinga under the Office of the Vice President for Information Technology at UAB
- UAB CCTS (Center for Clinical and Translational Science under grant UL1 RR025777 from the NIH National Center for Research Resources)
- The [Galaxy Platform](#) is developed by Penn State and Emory University

**Live Quickies**

- Mapping against custom genome  
Galactic quickie # 10
- Illumina mapping: Single Ends  
Galactic quickie # 11
- Illumina mapping: Paired Ends  
Galactic quickie # 12
- Basic fastQ manipulation:  
Galactic quickie # 13
- Advanced fastQ manipulation:  
Galactic quickie # 14
- 454 Mapping: Single End  
Galactic quickie # 15
- Uploading Data using FTP  
Galactic quickie # 17
- Managing account histories  
Galactic quickie # 19

History

Unnamed history 0 bytes

Your history is empty. Click 'Get Data' on the left pane to start

Logged in as jpr@uab.edu

- Saved Histories
- Saved Datasets
- Saved Pages
- API Keys
- Public Name





# CLC GENOMICS SERVER

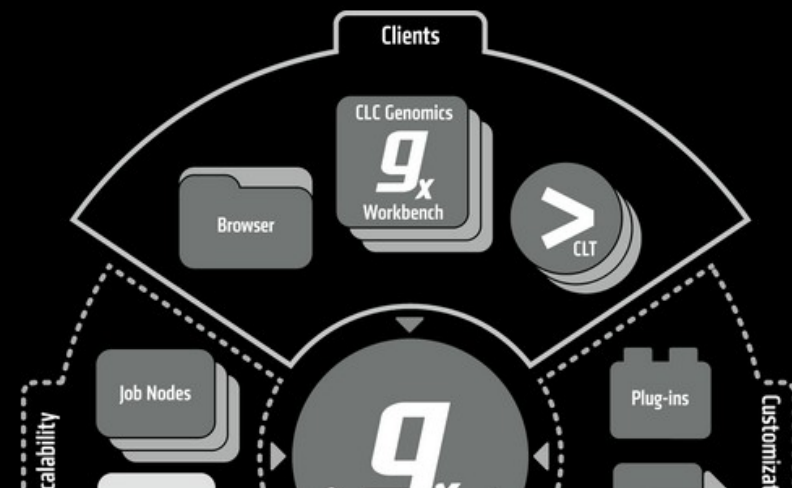
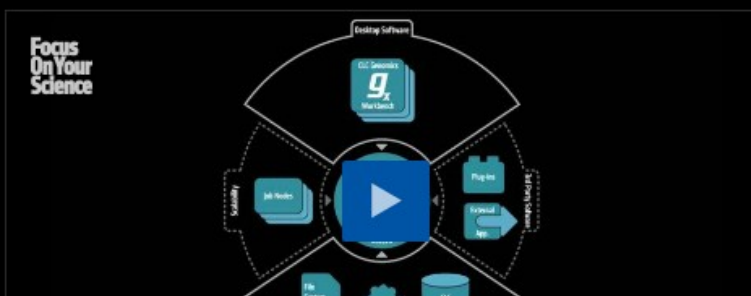
A flexible enterprise level infrastructure and analysis backbone for next generation sequencing data analysis.

[GET STARTED](#)
[DESKTOP SOFTWARE](#)
[ENTERPRISE PLATFORM](#)
[PRODUCTS](#)
[CONSULTING](#)
[SUPPORT](#)
[Genomics Server](#)
[Features](#)
[Clients](#)
[Customization](#)
[Data Management](#)
[Scalability](#)
[Latest Improvements](#)
[Download](#)
[To Top](#)

## 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

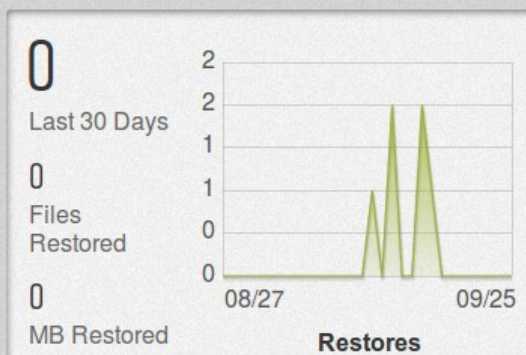
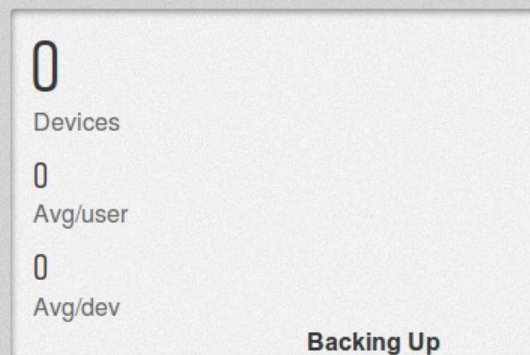
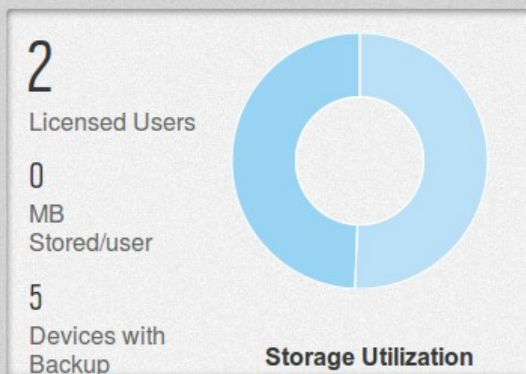
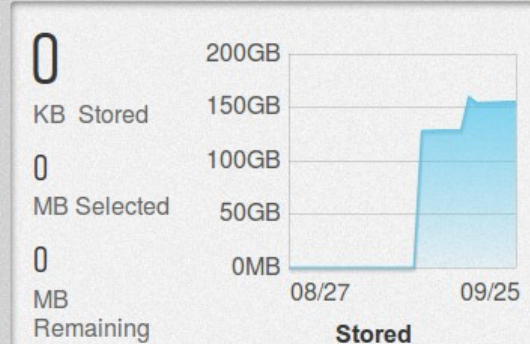
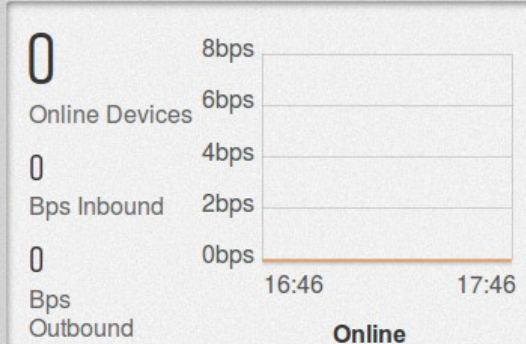
## Why a virtual machine?

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.

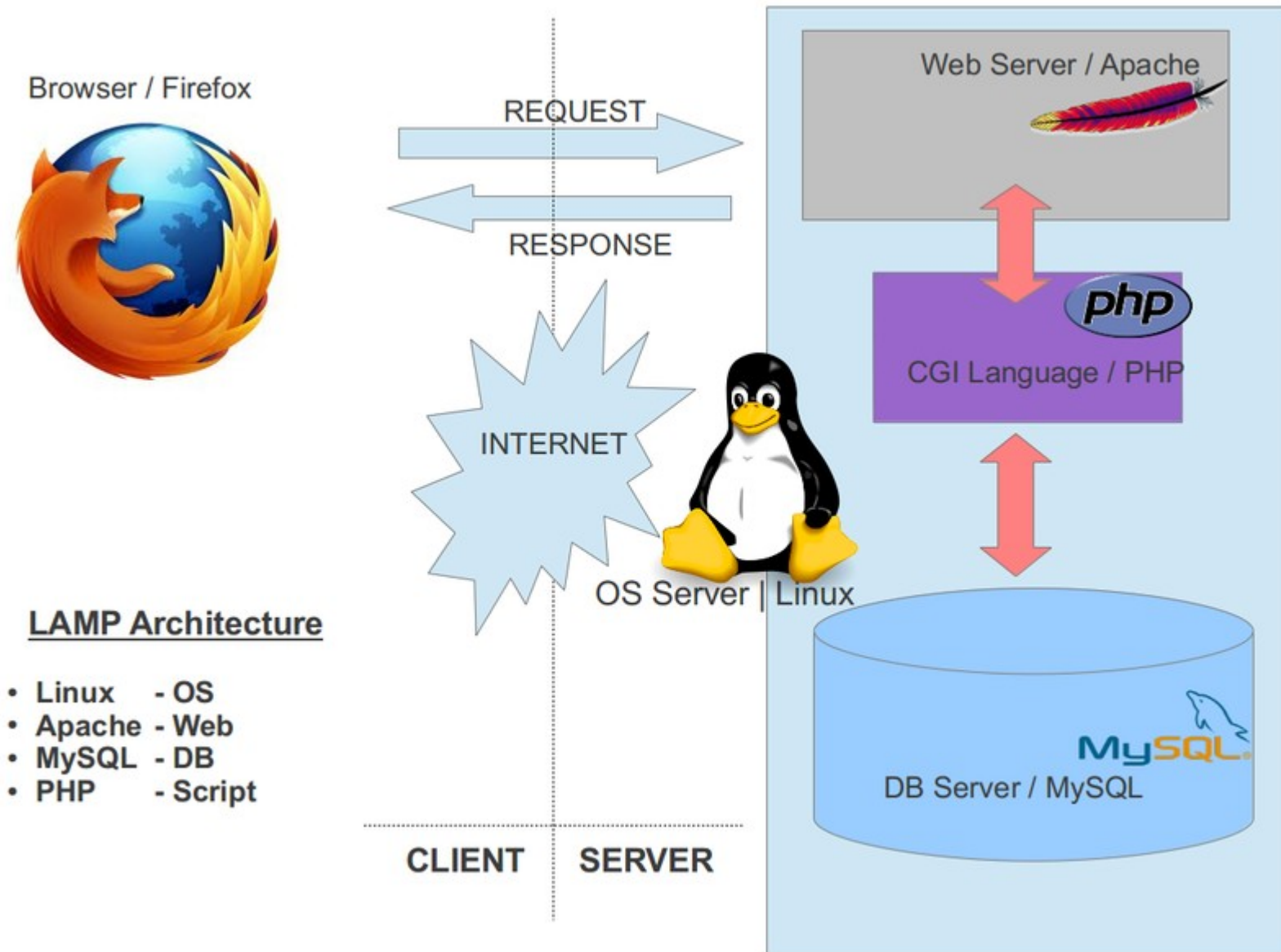


Smart Search

- Dashboard
- Alerts 4 1 3
- Organizations
- Users
- Devices
- Destinations
- Settings
- Licensing
- My Profile



# Building and Using LAMP Stacks



# Tools to Build Cloud-ready Systems



[VMWARE INTEGRATION](#)

[DOWNLOADS](#) [DOCUMENTATION](#) [SUPPORT](#) [ABOUT](#)

**Development  
environments  
made easy.**

Create and configure lightweight, reproducible,  
and portable development environments.

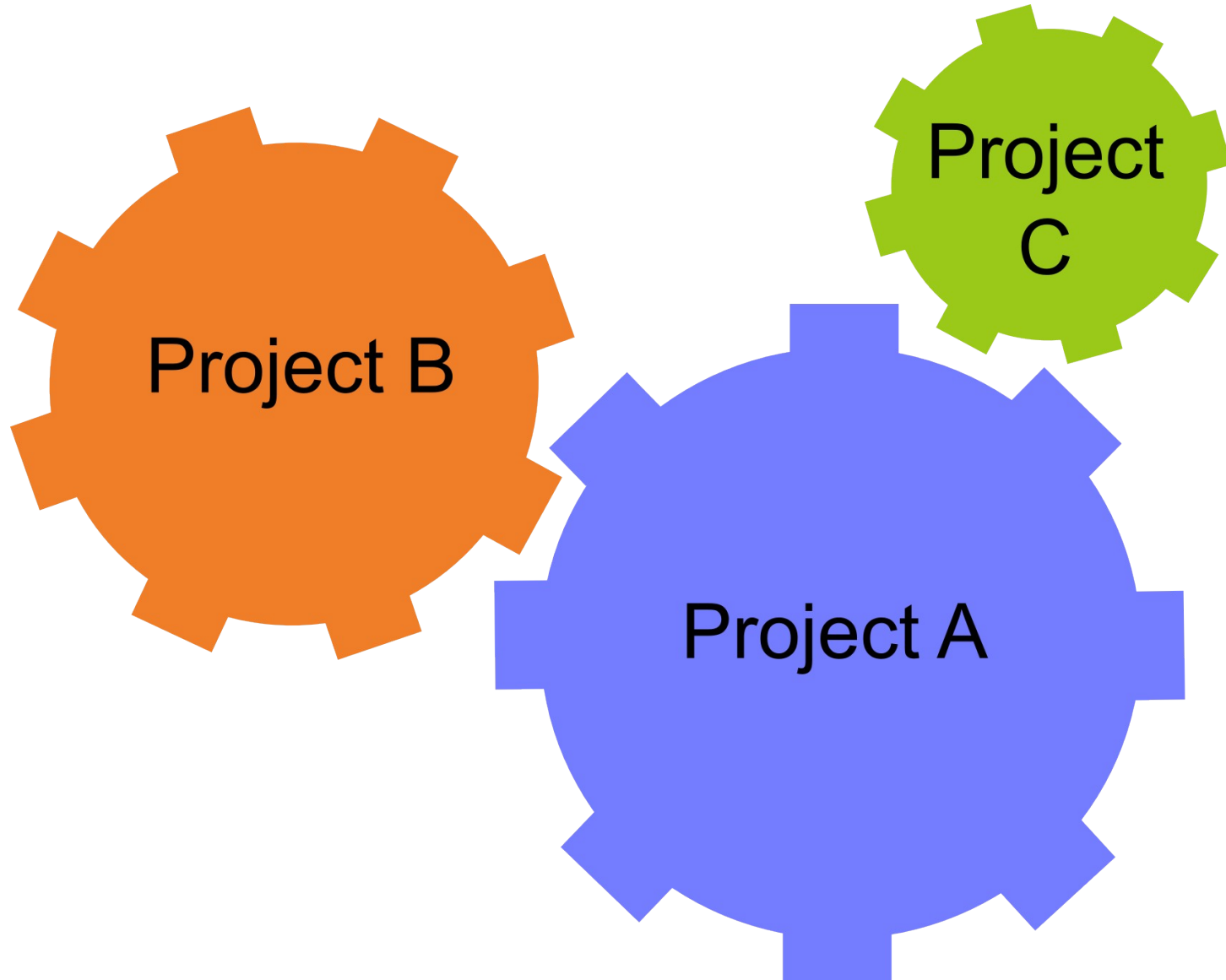
[DOWNLOAD](#)

[GET STARTED](#)

**VAGRANT WILL CHANGE HOW YOU WORK**



# Free Downstream Developers to Build at the Speed of Research

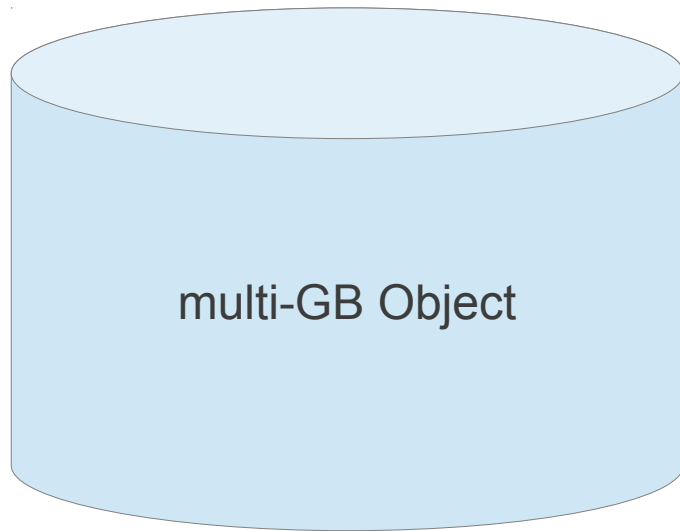


# What Apps are we building?

# What Apps are we building?

## Research Storage!

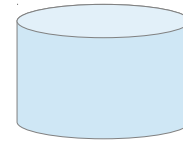
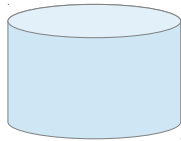
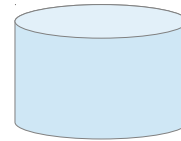
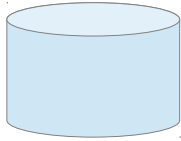
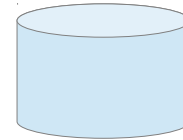
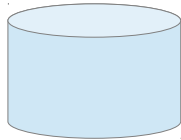
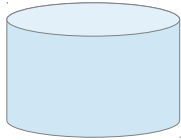
# Research Storage!



multi-GB Object

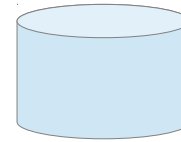
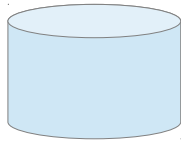
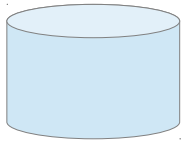
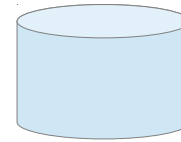
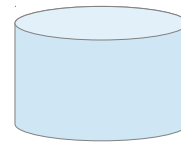
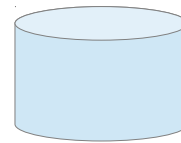
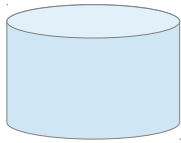
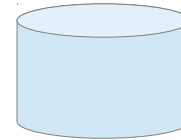
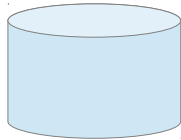
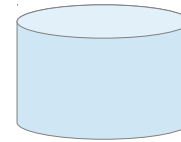
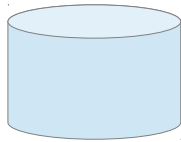
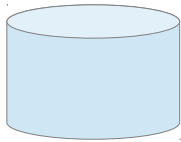
We like containers...

# Research Storage!

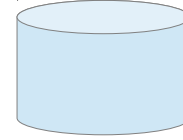
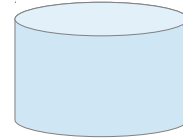
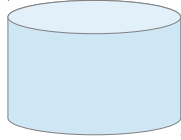
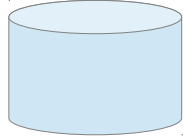
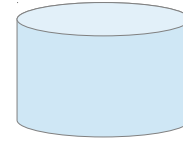
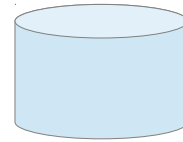
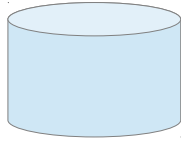
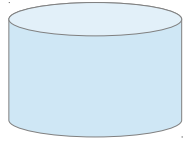
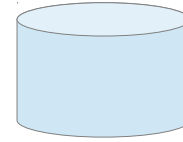
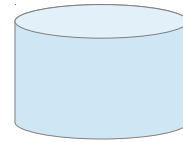
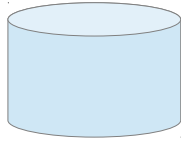
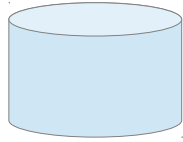
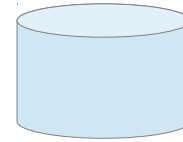
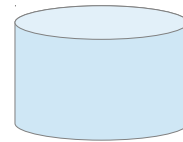
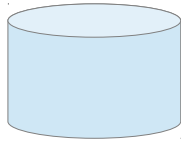
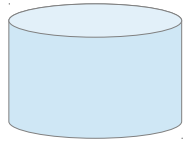
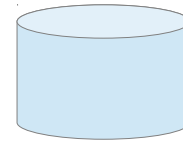
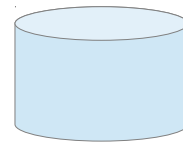
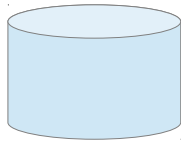
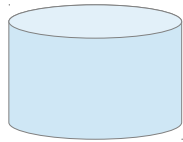




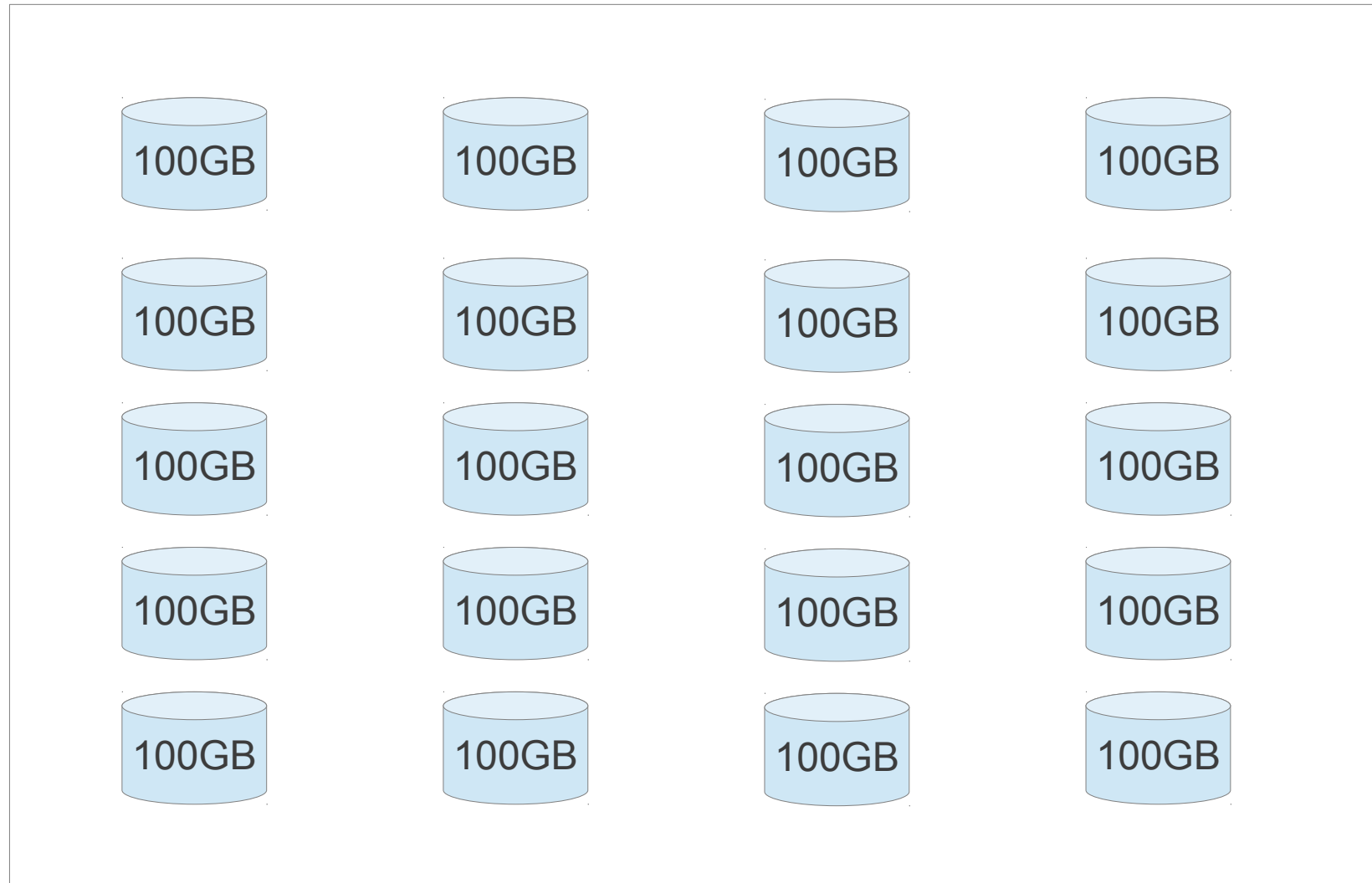
# Research Storage!



# Research Storage!

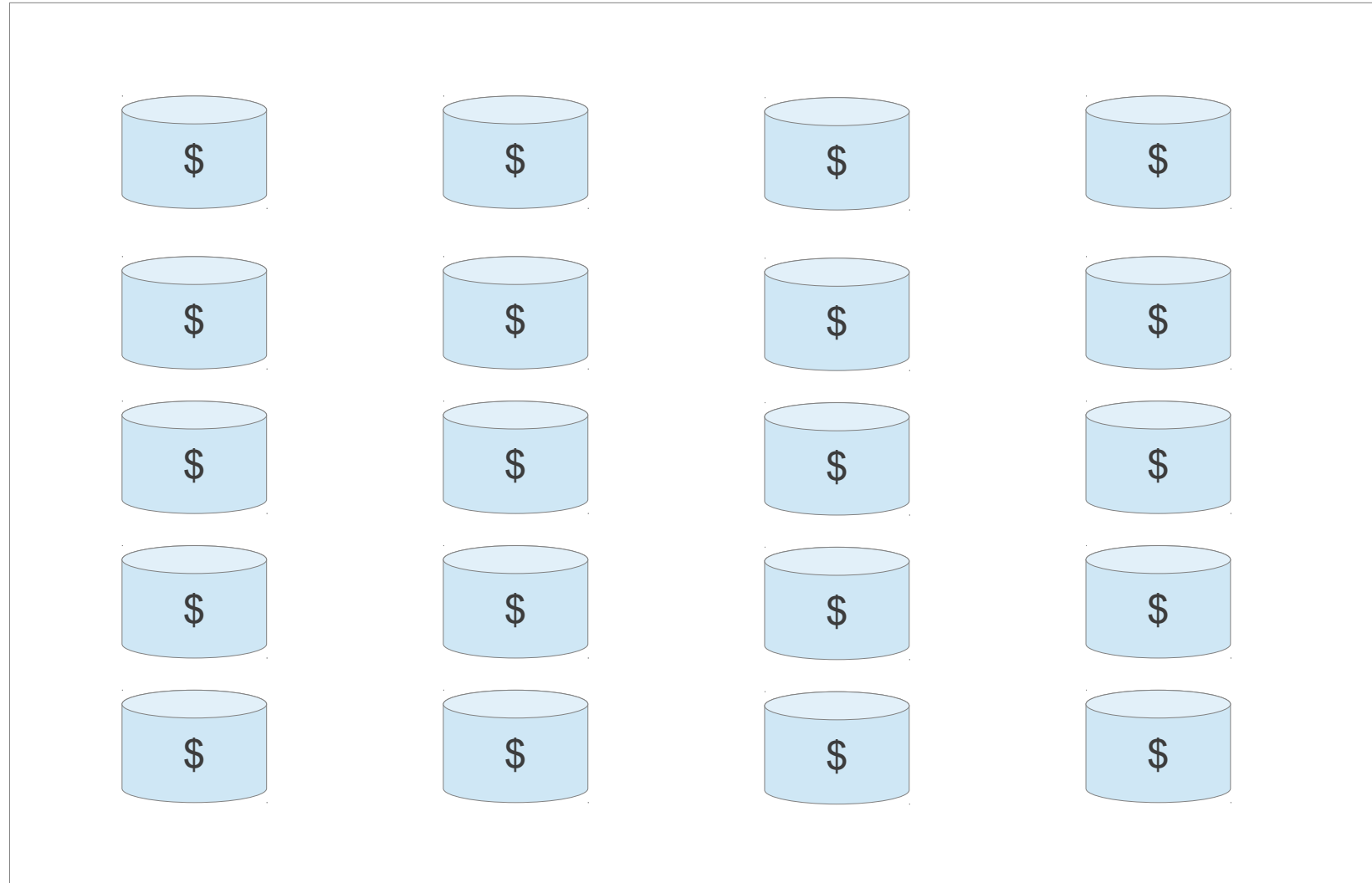


# Research Storage!



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!



Contact Sales | Support

Free Trial

Buy

CrashPlan PROe | Cloud Destinations | News & Events | Resources | Partners | Company

## Back up everyone, everywhere.

Enterprise endpoint backup IT teams trust and end users treasure.

Learn More →



Trusted by leading, brand-name organizations worldwide, including:

SanDisk

TaylorMade

TARGET

Adobe

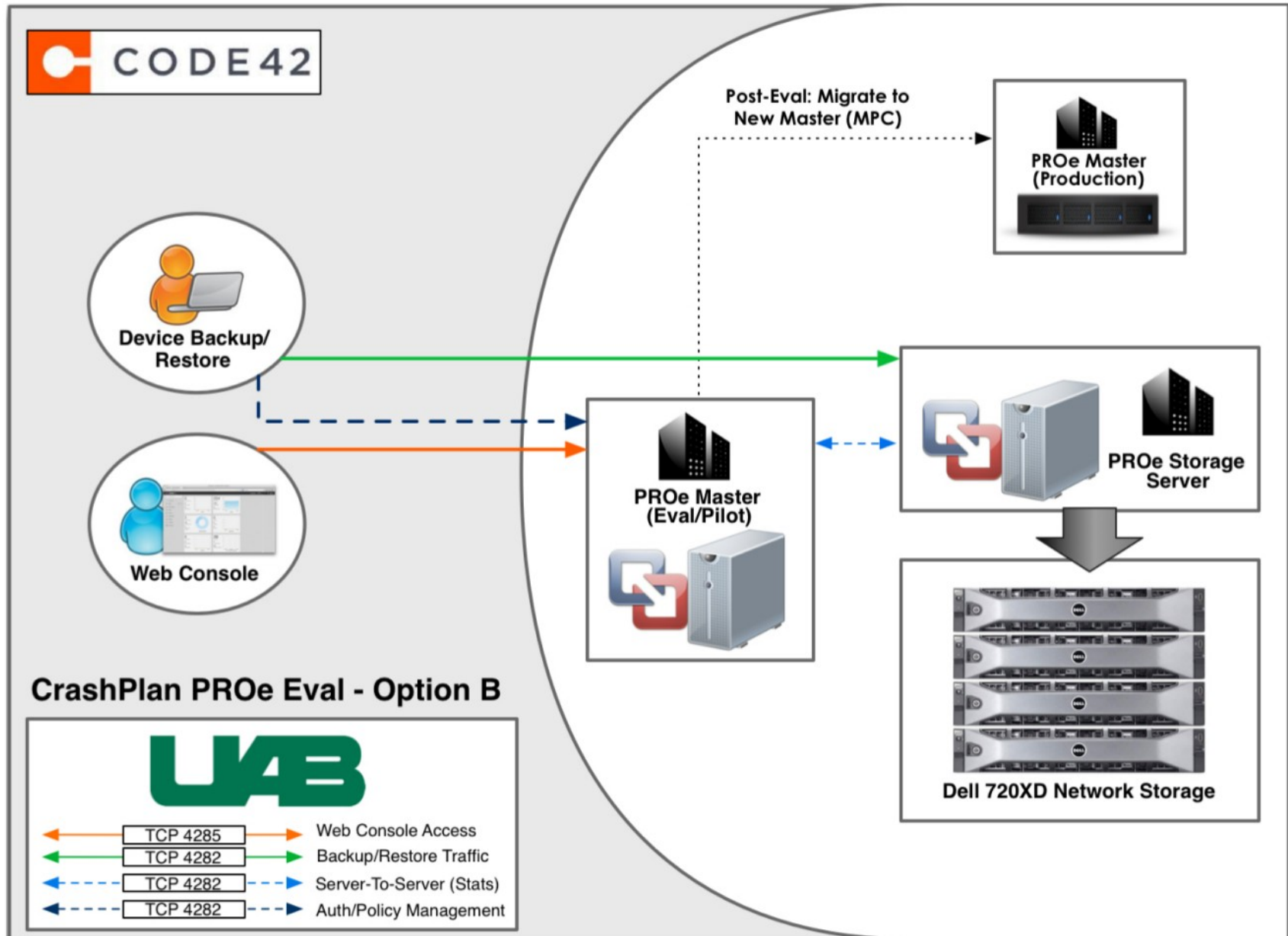
P&G



Leave a message



# 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!

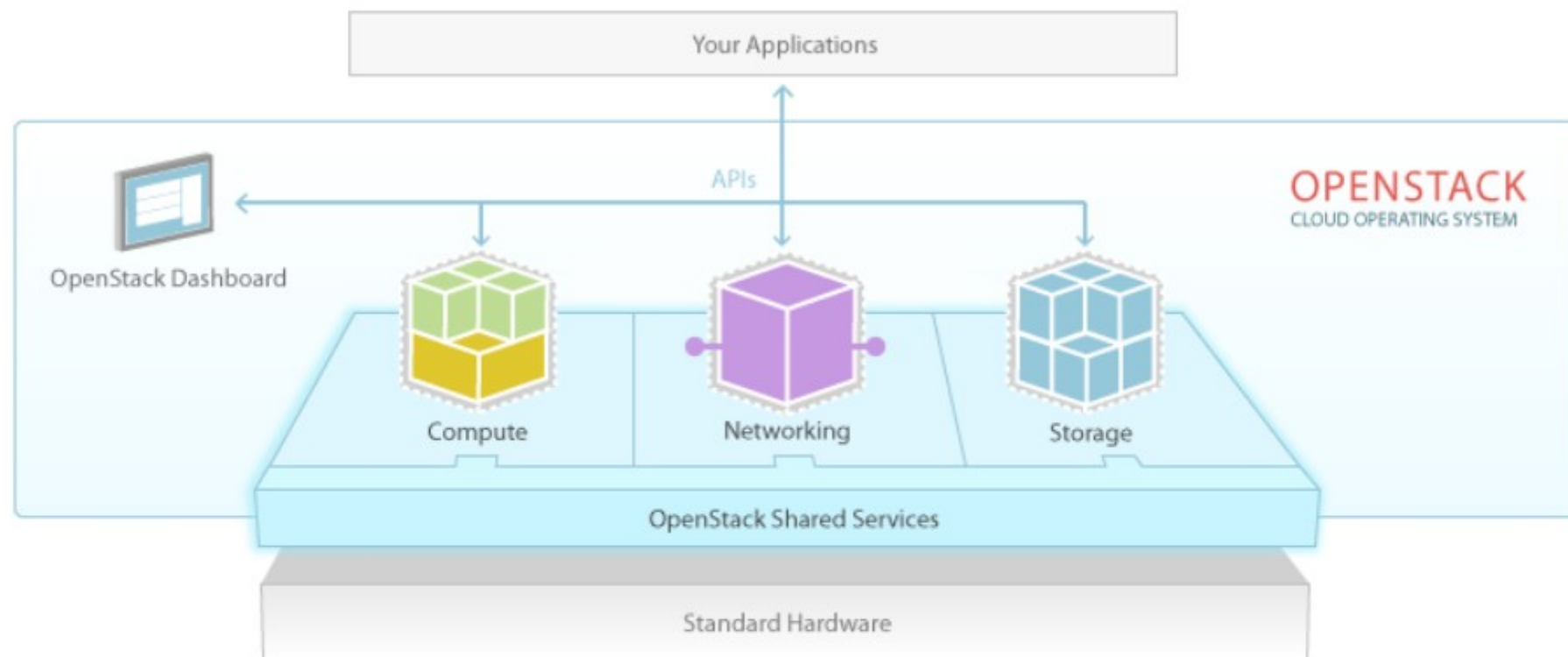
Research Whatever!

# 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

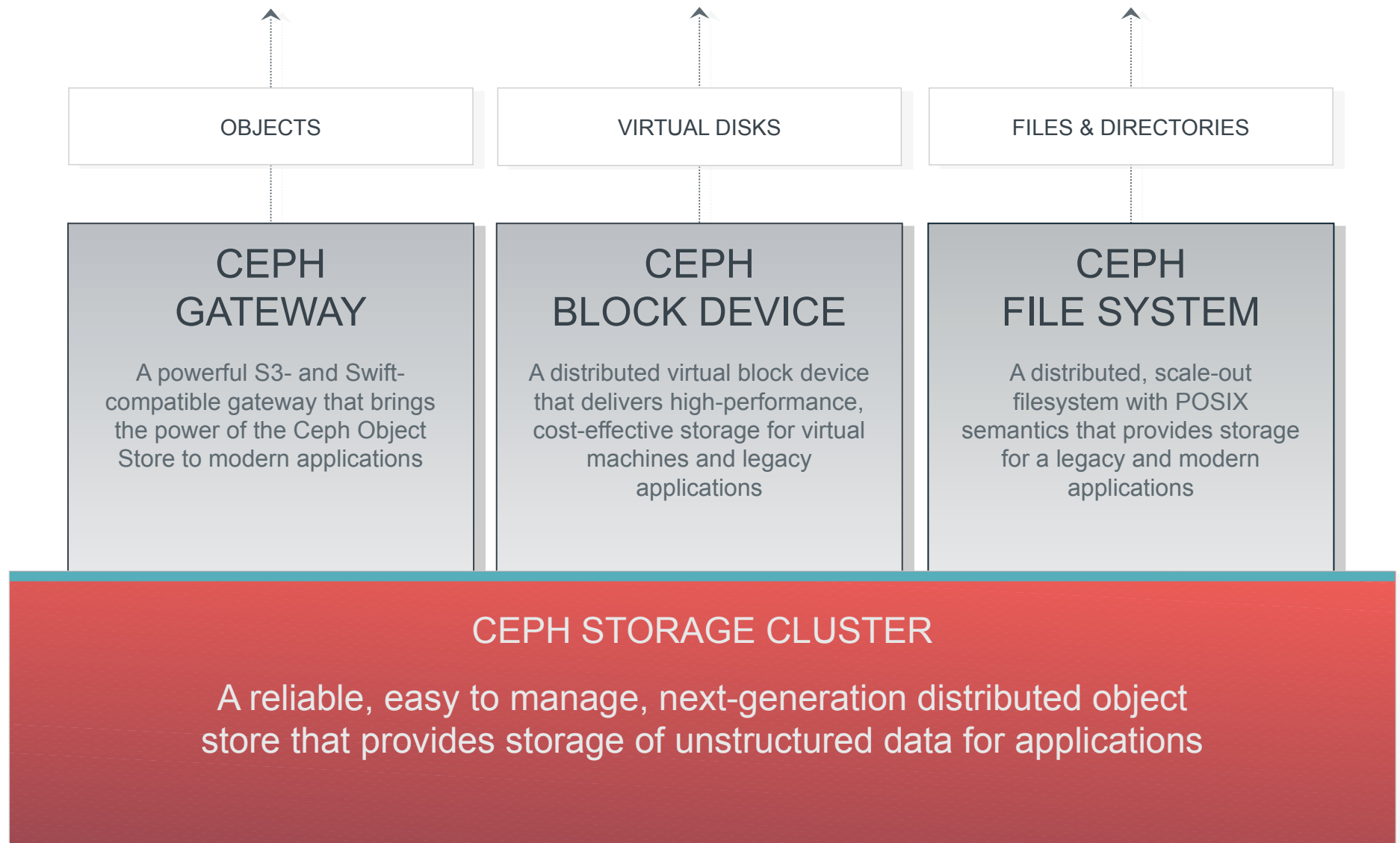
OpenStack is a cloud operating system that controls large pools of compute, storage, and networking resources throughout a datacenter, all managed through a dashboard that gives administrators control while empowering their users to provision resources through a web interface.

THE SEVENTH OPENSTACK RELEASE

# GRIZZLY

A QUICK VIDEO OVERVIEW

# Ceph Unified Storage Platform



# 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