Cluster and Cloud Computing – Lectures 9-10

Cloud Computing &

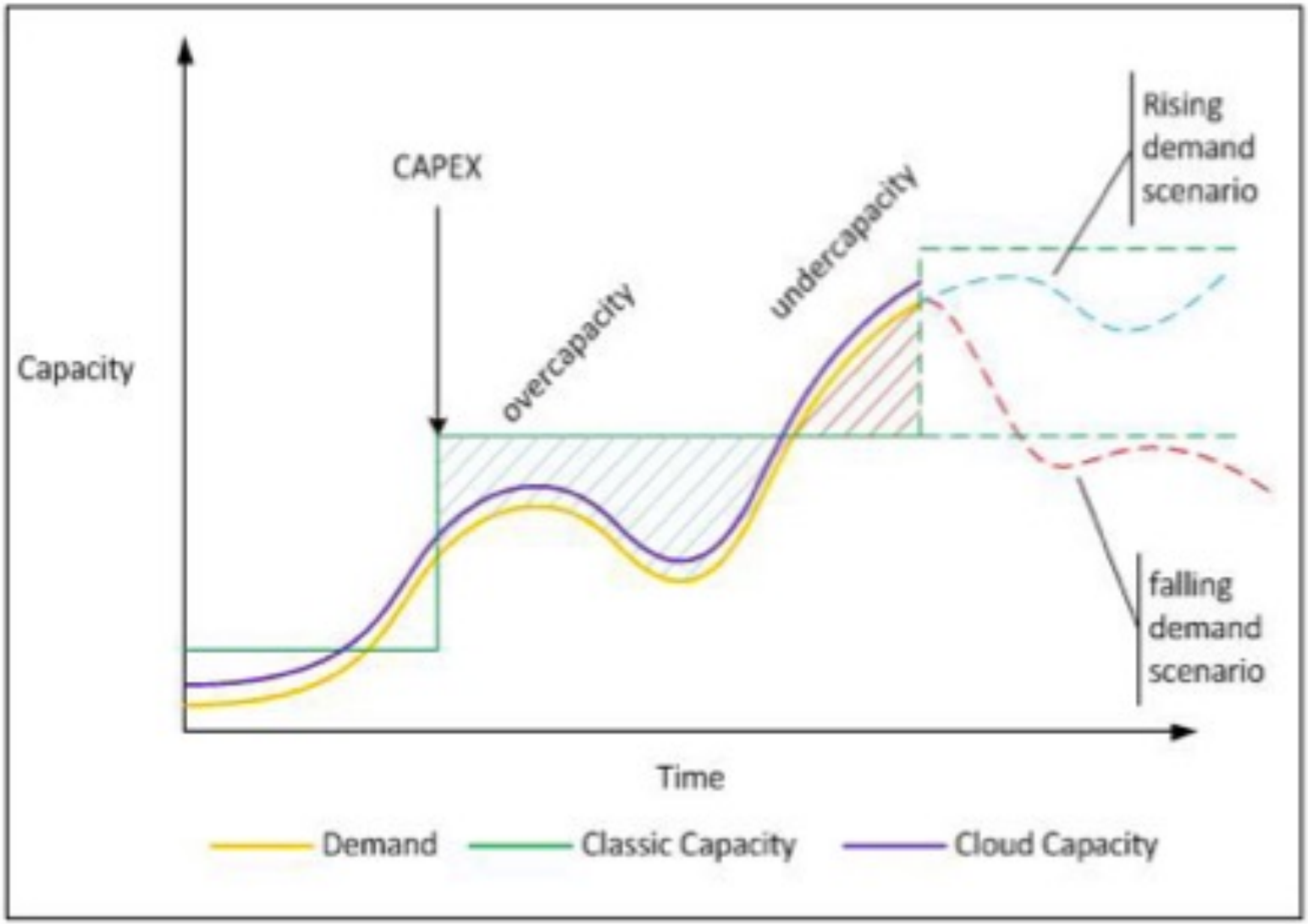
Professor Richard O. Sinnott & Yao Pan

rsinnott@unimelb.edu.au

Overview

- Cloud benefits
  - Cloud marketing!?
- The various flavours of cloud computing
  - Introduction to #aaS?
- Break
  - Demonstration of University of Melbourne Research Cloud
- Workshops on Thursday/Friday
  - Workshop: Scripting the Cloud and Ansible
  - **Pre-recorded, watch videos and Q&A in workshops**

Life before cloud computing

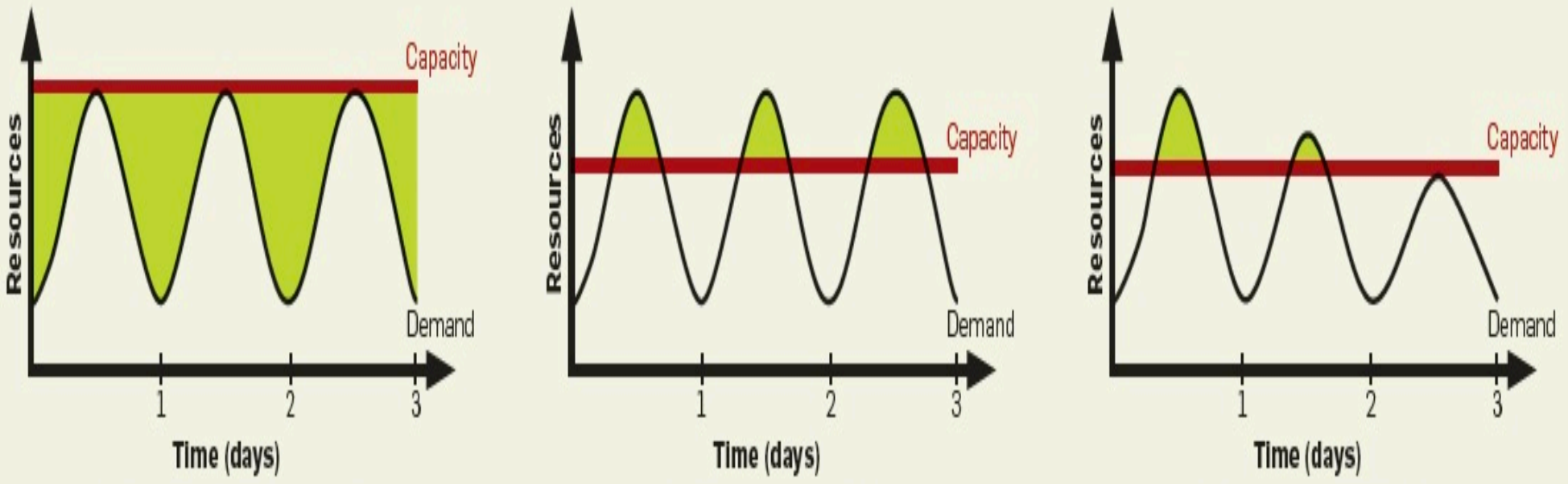


Capacity vs Utilization curves [8]

source: http://www.rackspace.com/knowledge_center/whitepaper/revolution-not-evolution-how-cloud-computing-differs-from-traditional-it-and-why-it

Life before cloud computing...ctd

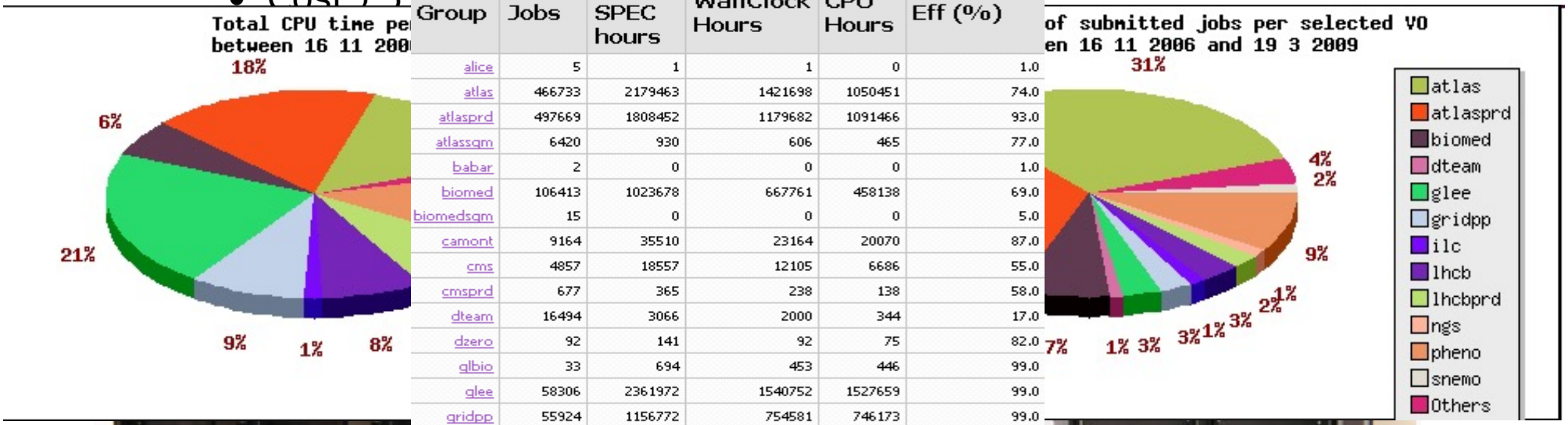
Figure 2. (a) Even if peak load can be correctly anticipated, without elasticity we waste resources (shaded area) during nonpeak times. (b) Underprovisioning case 1: potential revenue from users not served (shaded area) is sacrificed. (c) Underprovisioning case 2: some users desert the site permanently after experiencing poor service; this attrition and possible negative press result in a permanent loss of a portion of the revenue stream.



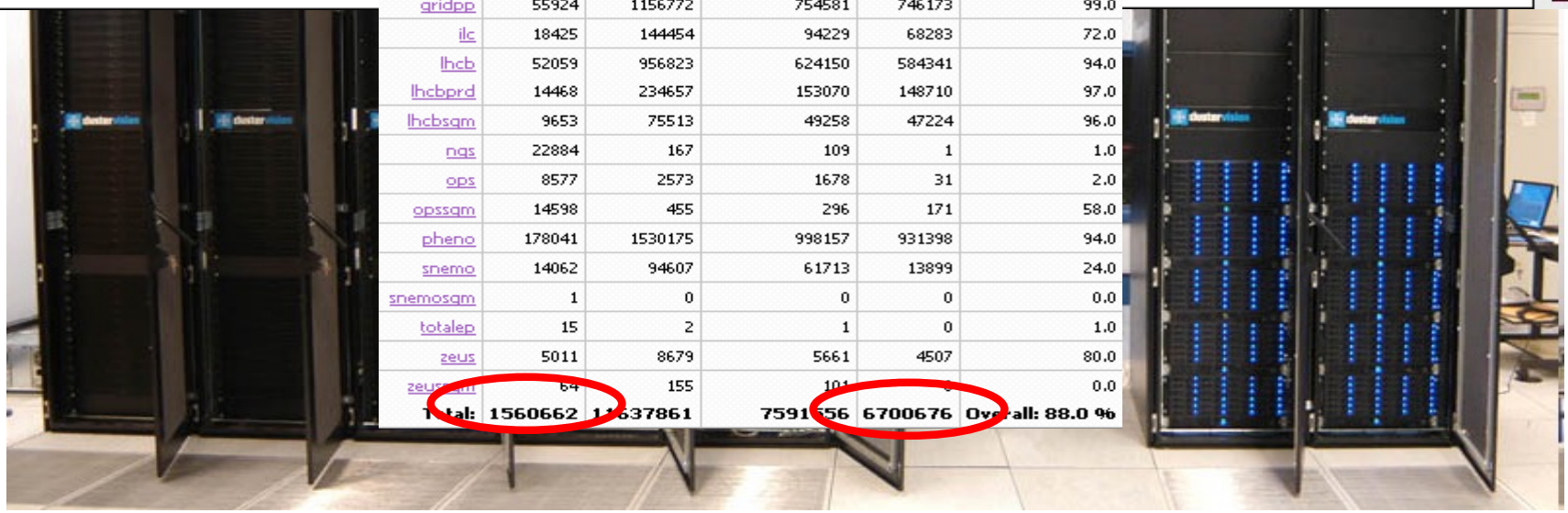
source: Ambrust et al, doi:10.1145/1721654.1721672

Cloud-busting? (~2009)

- To buy or not to buy that is the question…?
  - ScotGrid (www.scotgrid.ac.uk)
    - Cost £5



Group	Jobs	Kilo SPEC hours	WallClock Hours	CPU Hours	Eff (%)
alice	5	1	1	0	1.0
atlas	466733	2179463	1421698	1050451	74.0
atlasprd	497669	1808452	1179682	1091466	93.0
atlassgm	6420	930	606	465	77.0
babar	2	0	0	0	1.0
biomed	106413	1023678	667761	458138	69.0
biomedsgm	15	0	0	0	5.0
camont	9164	35510	23164	20070	87.0
cms	4857	18557	12105	6686	55.0
cmsprd	677	365	238	138	58.0
dteam	16494	3066	2000	344	17.0
dzero	92	141	92	75	82.0
glbio	33	694	453	446	99.0
glee	58306	2361972	1540752	1527659	99.0
gridpp	55924	1156772	754581	746173	99.0
ilc	18425	144454	94229	68283	72.0
lhcb	52059	956823	624150	584341	94.0
lhcbprd	14468	234657	153070	148710	97.0
lhcbsgm	9653	75513	49258	47224	96.0
ngs	22884	167	109	1	1.0
ops	8577	2573	1678	31	2.0
opssgm	14598	455	296	171	58.0
pheno	178041	1530175	998157	931398	94.0
snemo	14062	94607	61713	13899	24.0
snemosgm	1	0	0	0	0.0
totalep	15	2	1	0	1.0
zeus	5011	8679	5661	4507	80.0
zeussgm	64	155	101	[illegible]	0.0
Total:	1560662	11637861	7591[illegible]56	6700676	Overall: 88.0 %



Cloud-busting? (~2009)

On-Demand Instances

United States | Europe

Standard On-Demand Instances	Linux/UNIX Usage	Windows Usage
1.7Gb memory/160GB disk	$0.11 per hour	$0.135 per hour
7.5Gb memory/850GB disk	$0.44 per hour	$0.54 per hour
15Gb memory/1690GB disk	$0.88 per hour	$1.08 per hour
High CPU On-Demand Instances	Linux/UNIX Usage	Windows Usage
Medium	$0.22 per hour	$0.32 per hour
Extra Large	$0.88 per hour	$1.28 per hour

Amazon Elastic Block Store

United States | Europe

Amazon EBS Volumes

- $0.11 per GB-month of provisioned storage
- $0.11 per 1 million I/O requests

Amazon EBS Snapshots to Amazon S3 (priced the same as Amazon S3)

- $0.18 per GB-month of data stored
- $0.012 per 1,000 PUT requests (when saving a snapshot)
- $0.012 per 10,000 GET requests (when loading a snapshot)

Data Transfer

Internet Data Transfer

The pricing below is based on data transferred "in" and "out" of Amazon EC2.

Data Transfer In	
All Data Transfer	$0.10 per GB

Data Transfer Out	
First 10 TB per Month	$0.17 per GB
Next 40 TB per Month	$0.13 per GB
Next 100TB per Month	$0.11 per GB
Over 150 TB per Month	$0.10 per GB

- $1=£0.69 (back then!)
  - £0.30*6,700,676 CPU hours
    - = £2,010,202 for just compute on-demand + data + networking + …?
- Now…???
  - https://aws.amazon.com/ec2/pricing/

AWS EC2 Pricing (just EC2!)

- **On demand**
  - pay for compute capacity used; no long-term commitments; increase/decrease as required
- **Spot pricing**
  - when have flexibility in using resources, e.g., get resources to run my jobs when price drops below…
- **Saving plans**
  - Longer term arrangements, need to use regularly
- **Dedicated Hosts**
  - Physical dedicated servers; can use own licensed software
- **Reserved instances**
  - for applications with known usage patterns, e.g., fixed reserve resources in given availability zone

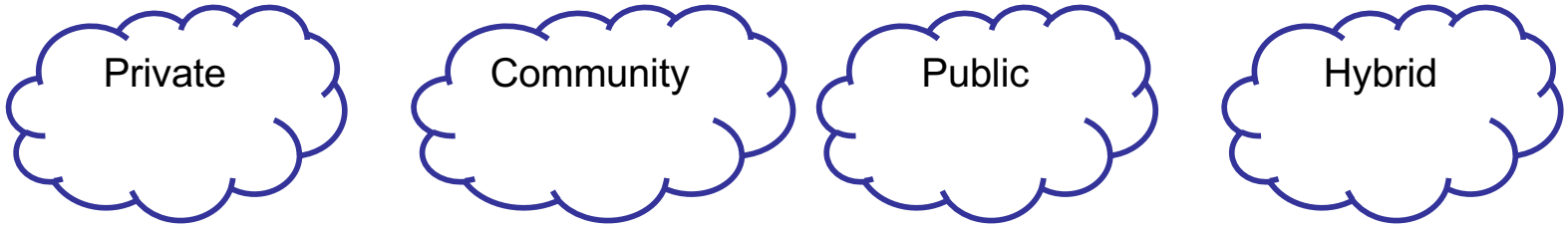
Cloud Computing: A Definition

- NIST definition: "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."
  - » National Institute of Standards and Technology (http://dx.doi.org/10.6028/NIST.SP.800-145)

- Focus of today is to get you up and running on the Cloud and explore the technologies related to the underlined
  - Later lecture will do compare/contrast with AWS
    - …and then how do Clouds actually work (hypervisors etc)

The Most Common Cloud Models

Deployment Models



Delivery Models

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS)

Essential Characteristics

- On-demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service

Note also XYZ-as-a-Service

Public Clouds

- Pros
  - Utility computing
  - Can focus on core business
  - Cost-effective
  - “Right-sizing”
  - Democratisation of computing
- Cons
  - Security
  - Loss of control
  - Possible lock-in
  - Dependency of Cloud provider continued existence

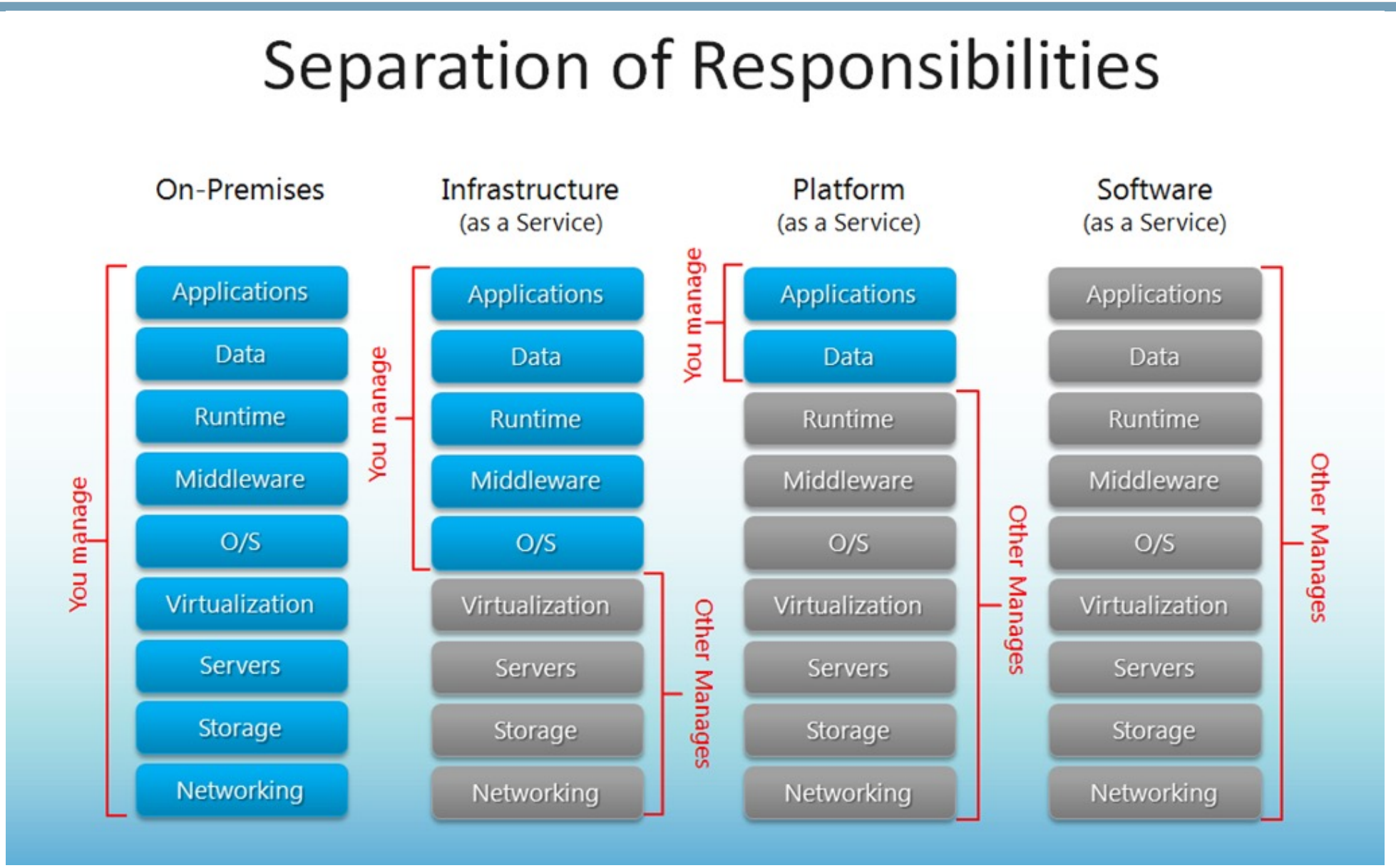
Private Clouds

- Pros
  - Control
  - Consolidation of resources
  - Easier to secure
  - More trust
- Cons
  - Relevance to core business?
    - e.g., Netflix to Amazon
  - Staff/management overheads
  - Hardware obsolescence
  - Over/under utilisation challenges

Hybrid Clouds

- Examples
  - Eucalyptus, VMWare vCloud Hybrid Service
- Pros
  - Cloud-bursting
    - Use private cloud, but burst into public cloud when needed
- Cons
    - How do you move data/resources when needed?
    - How to decide (in real time?) what data can go to public cloud?
    - Is the public cloud compliant with PCI-DSS (Payment Card Industry – Data Security Standard)?

Delivery Models



Public SaaS examples

- Ubiquitous to the world we live in…
  - Gmail
  - Sharepoint
  - Salesforce.com CRM
  - On-live
  - Gaikai
  - Microsoft Office 365
  - Some definitions include those that do not require payment, e.g., ad-supported sites

Public PaaS Examples

Cloud Name	Language and Developer Tools	Programming Models Supported by Provider	Target Applications and Storage Options
Google App Engine	Python, Java, Go, PHP + JVM languages (scala, groovy, jruby)	MapReduce, Web, DataStore, Storage and other APIs	Web applications and BigTable storage
Salesforce.com's Force.com	Apex, Eclipsed-based IDE, web-based wizard	Workflow, excel-like formula, web programming	Business applications such as CRM
Microsoft Azure	.NET, Visual Studio, Azure tools	Unrestricted model	Enterprise and web apps
Amazon Elastic MapReduce	Hive, Pig, Java, Ruby etc.	MapReduce	Data processing and e-commerce
Aneka	.NET, stand-alone SDK	Threads, task, MapReduce	.NET enterprise applications, HPC

Infrastructure As A Service (IaaS)

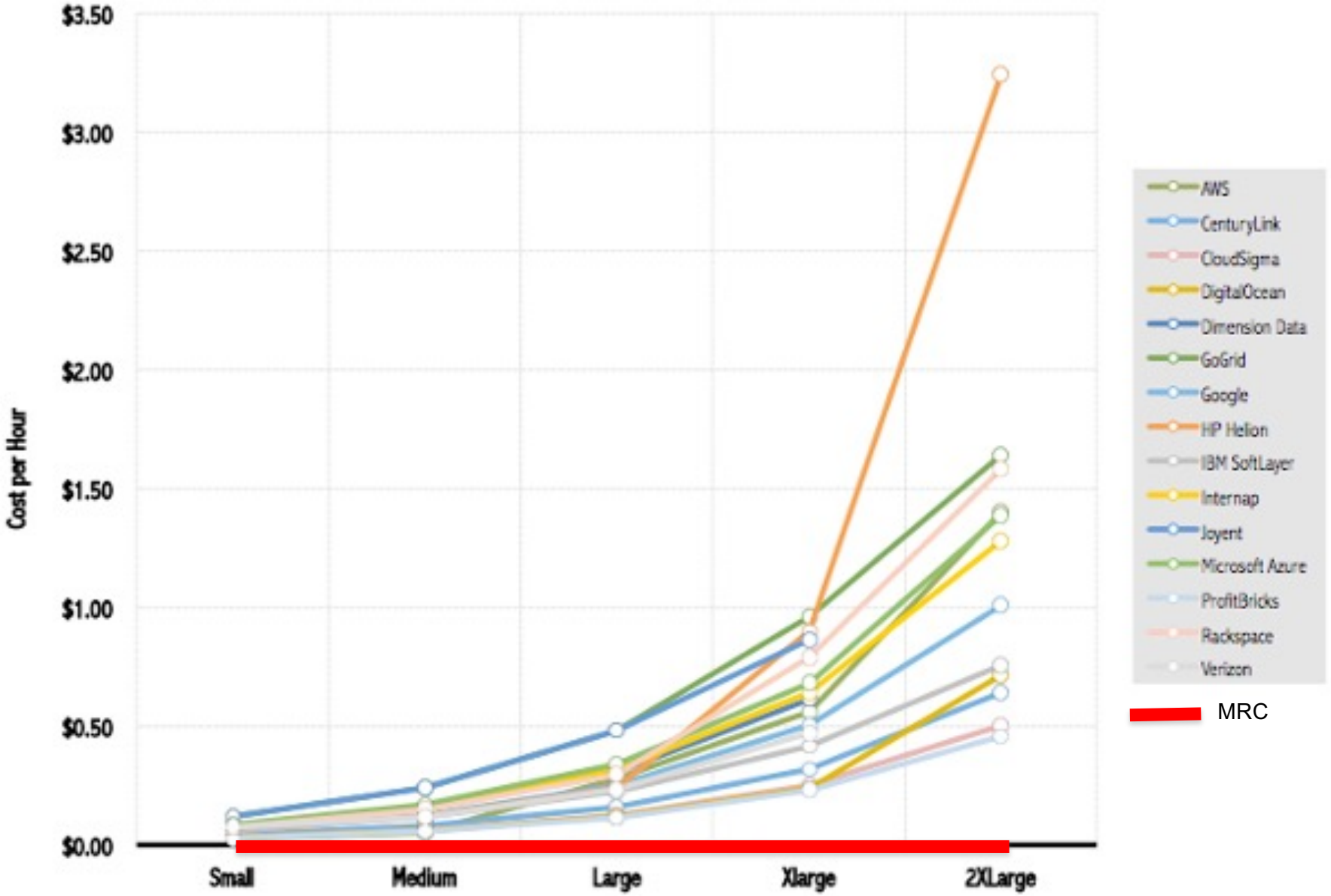
- Primary focus of this course…
- Many providers
  - Amazon Web Services (Market leader)
    - http://aws.amazon.com
  - Oracle Public Cloud
    - https://cloud.oracle.com/
  - Rackspace Cloud
    - www.rackspace.com
  - *CenturyLink, CloudSigma, DigitalOcean, DimensionData, GoGrid, Helio, Internap, Joyent, ProfitBricks, Verizon, …*
  - Melbourne Research Cloud/Openstack
    - https://dashboard.cloud.unimelb.edu.au/

UniMelb Research Cloud

- Based on OpenStack
  - Open source cloud technology (more later lecture)
- Many associated/underpinning services
  - Compute Service (code-named Nova)
  - Image Service (code-named Glance)
  - Block Storage Service (code named Cinder)
  - Object Storage Service (code-named Swift)
  - Security Management (code-named Keystone)
  - Orchestration Service (code-named Heat)
  - Network Service (code-named Neutron)
  - Metering Service (code-named Ceilometer)
  - …

https://docs.openstack.org/rocky/

Cost Comparison



See also https://www.infoworld.com/article/3237566/cloud-pricing-comparison-aws-vs-azure-vs-google-vs-ibm.html

NeCTAR/ARDC/UniMelb Research Cloud

- National eResearch Collaboration Tools and Resources (NeCTAR – www.nectar.org.au)
  - $50m+$10m+$10m+$72m (Australian Research Data Commons)… federal funding
  - Originally lead by University of Melbourne
  - Had four key strands
    - ~~National Servers Program~~
    - Research Cloud Program
      - OpenStack IaaS
      - 4Gb-64Gb (mostly Linux flavours)
      - 30,000 physical servers available across different availability zones
        - » Being upgraded continually!
    - ~~eResearch Tools Program~~
    - Virtual Laboratories Program
      - Astro,
      - Genomics,
      - Humanities,
      - Climate,
      - Nano-,
      - …endocrine genomics

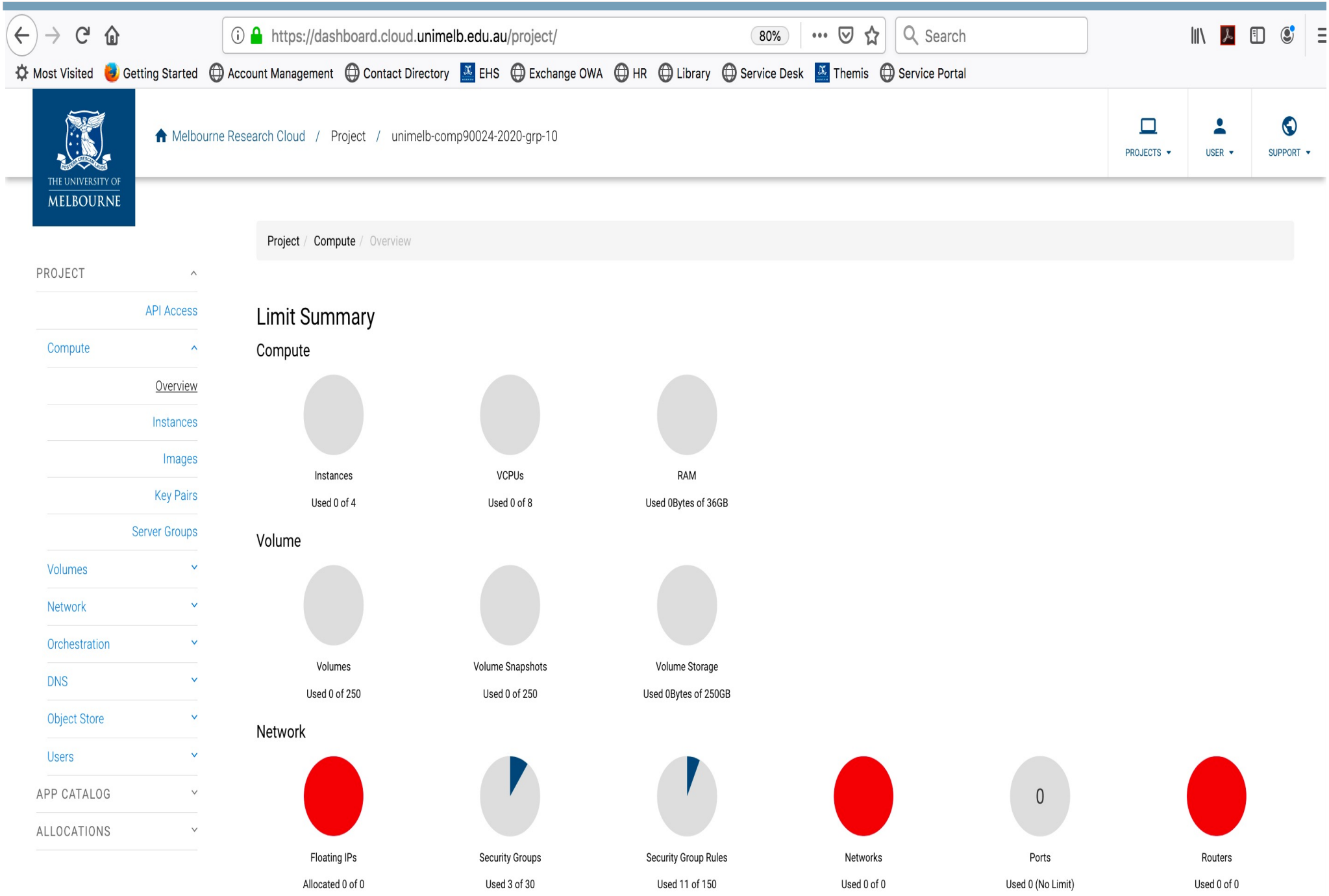
Data Infrastructures

- Research Data Services (RDS) project to establish data storage resources across Australia
  - ~100 Petabytes national data storage
  - Victoria Node (VicNode)
    - UniMelb, UniMonash for Vic-wide "nationally significant data sets"
      - Used by many diverse communities

UniMelb Cloud Activities

- Petascale Campus Initiative
  - https://research.unimelb.edu.au/infrastructure/petascale-campus-initiative
- Research Computing Services
  - https://research.unimelb.edu.au/infrastructure/research-computing-services
- Melbourne eResearch Group
  - https://www.eresearch.unimelb.edu.au

UniMelb Research Cloud



BREAK