
Managing Storage in the Cloud

Challenges in embracing cloud storage

Author:

Randy Bias (VP Technology Strategy of GoGrid/ServePath)

January 2009

INTRODUCTION

Backgrounds

- Randy Bias, VP Technology Strategy, GoGrid
 - Specializing in IT infrastructure since 1990
 - 3 ISPs, 3 Internet Datacenters, 2 MSSPs, 2 Cloud providers
 - Pioneered world's 1st multi-platform, multi-cloud management system (CloudScale Networks, Inc)
- GoGrid
 - Division of ServePath, LLC, an established hosting provider
 - Pioneer in cloud computing
 - First 'cloudcenter' type Infrastructure-as-a-Service cloud in U.S.

Today's Agenda

- ⌚ Cloud Computing Overview
- ⌚ Why Storage in the Cloud?
- ⌚ Storage Today
- ⌚ Management Challenges
- ⌚ Future / Vision

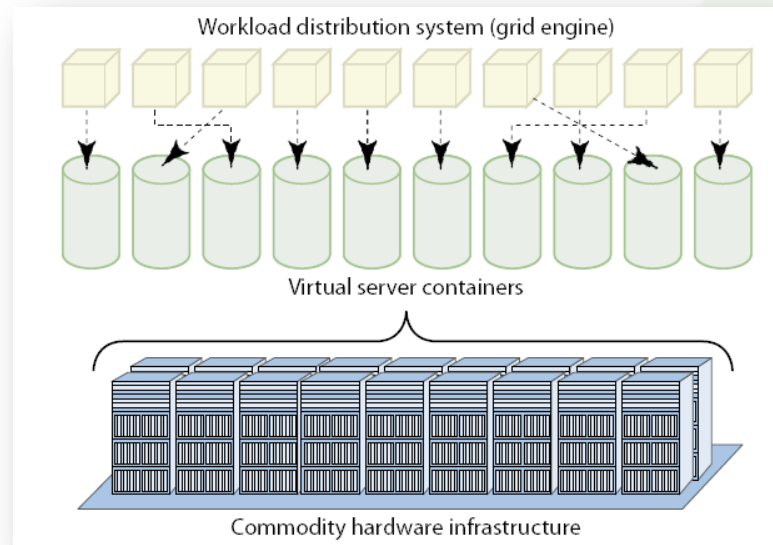


What is cloud computing?

CLOUD COMPUTING

Forrester Research

“A pool of abstracted, highly scalable, and managed compute infrastructure capable of hosting end-customer applications and billed by consumption¹”



1- “Is Cloud Computing Ready for The Enterprise?” Forrester Research, Inc.

Other Definitions

“Cloud computing is an emerging approach to shared infrastructure in which large pools of systems are linked together to provide IT services.”

– IBM press release on “Blue Cloud”

“...a hosted infrastructure model that delivers abstracted IT resources over the Internet”

– Thomas Weisel Partners LLC from “Into the Clouds: Leveraging Data Centers and the Road to Cloud Computing”

“Cloud computing describes a systems architecture. Period. This particular architecture assumes nothing about the physical location, internal composition or ownership of its component parts.”

– James Urquhart, Wisdom of the Clouds blog

Cloud Computing is...

ABSTRACTION
OVER THE INTERNET
INTERFACES

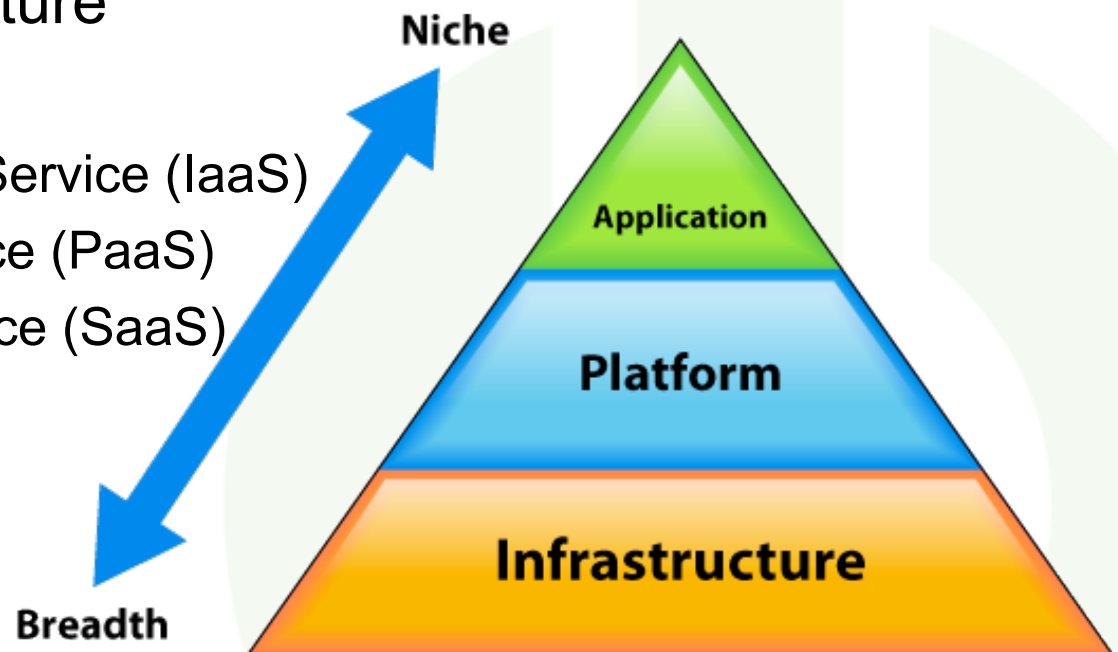
PROGRAMMATIC
USAGE BILLING
SCALABLE

SELF-SERVICE
SHARED
ON-DEMAND

Self-service Internet infrastructure where you pay for what you use, use only what you need, and managed using a web browser or API.

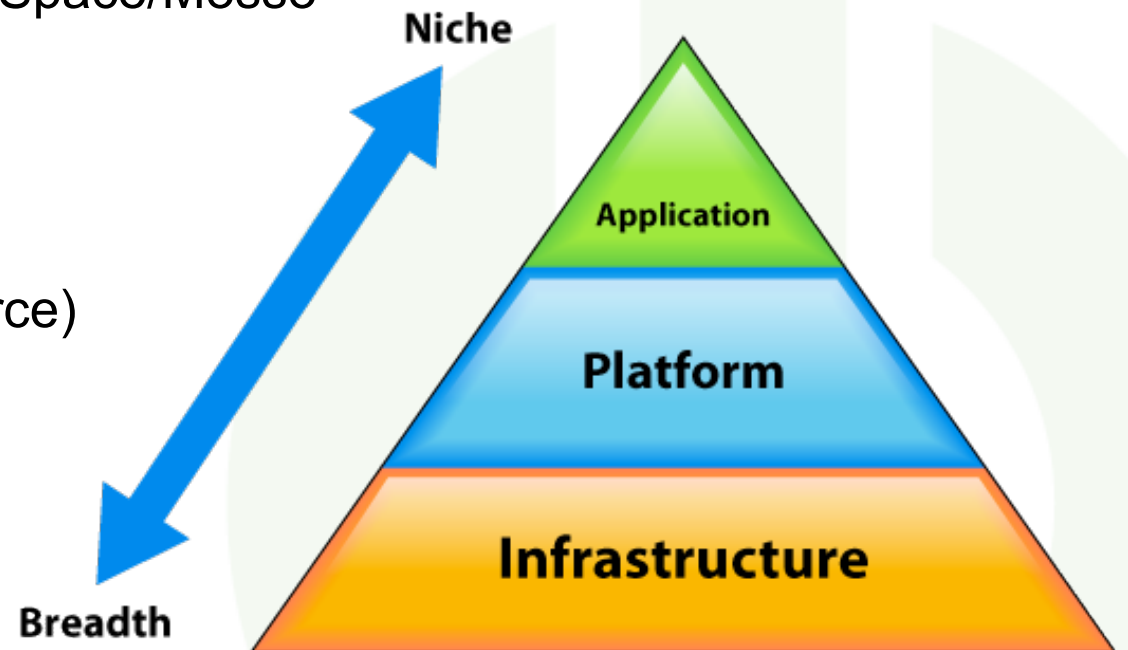
The “Cloud Pyramid”

- Cloud Services economy is a pyramid
- Layers equate structure
- Building blocks:
 - Infrastructure-as-a-Service (IaaS)
 - Platform-as-a-Service (PaaS)
 - Software-as-a-Service (SaaS)
- Breadth vs. Niche

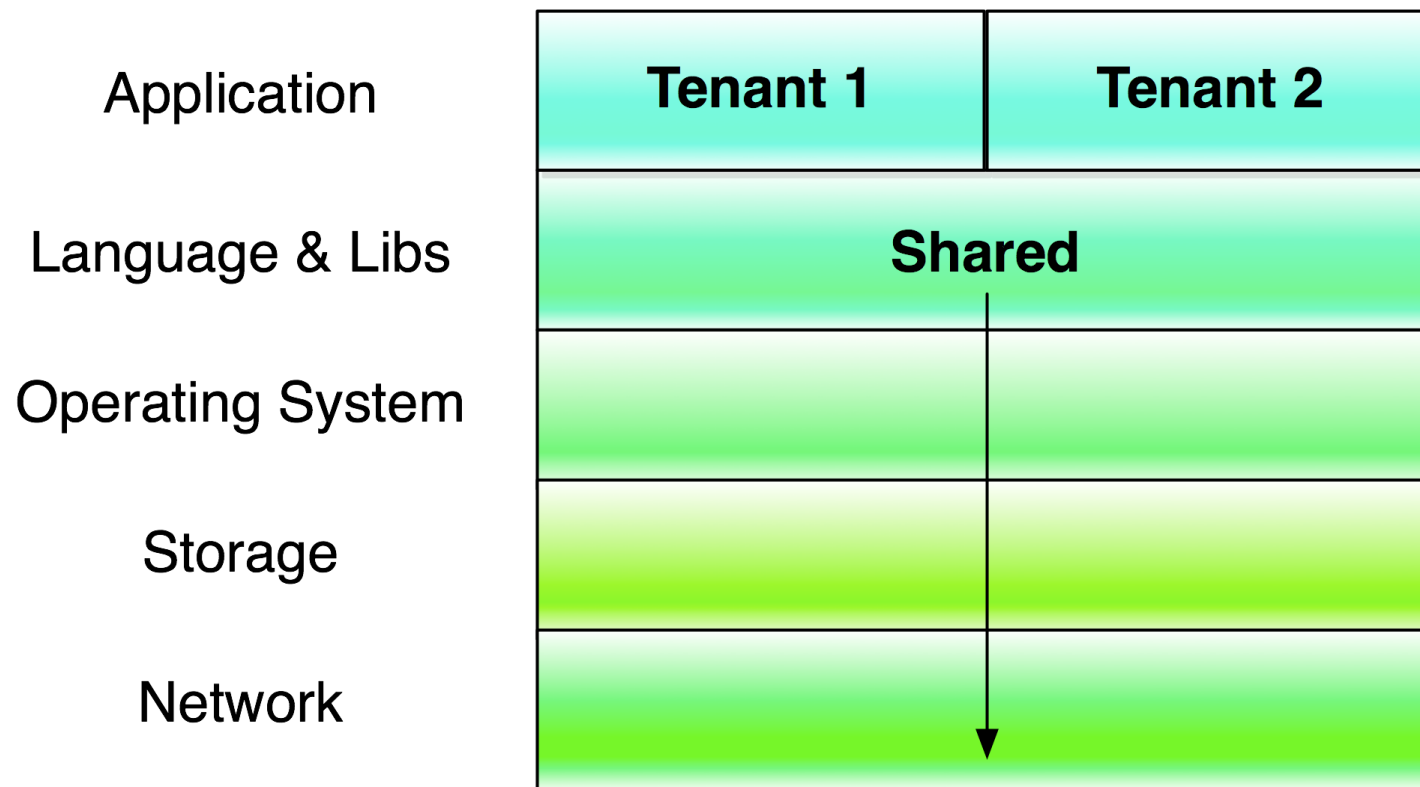


Cloud Service Economy

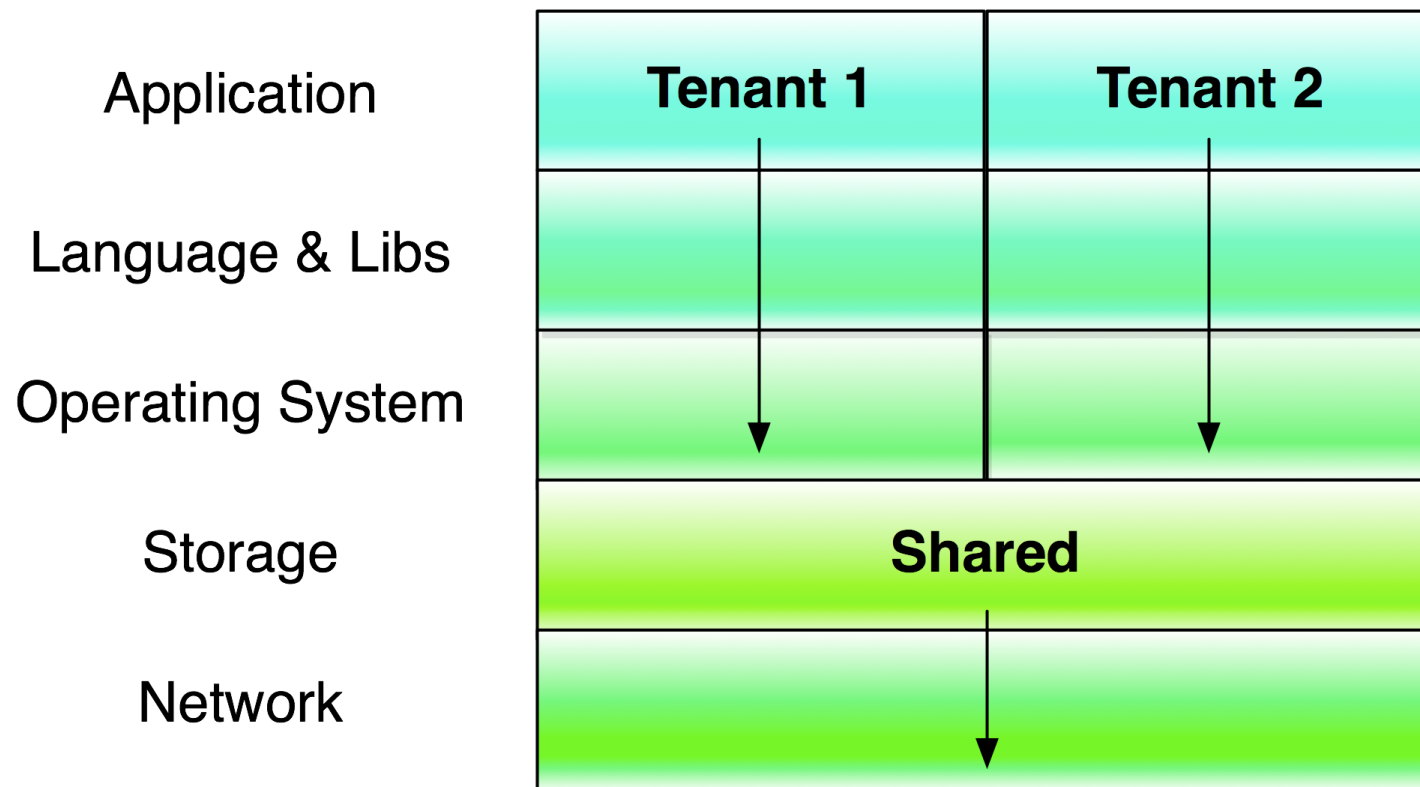
- 🔌 IaaS (Infrastructure)
 - AWS, GoGrid, RackSpace/Mosso
- 🔌 PaaS (Platform)
 - Google App Engine
 - RightScale
 - Force.com (Salesforce)
- 🔌 SaaS (Application)
 - Salesforce (CRM)
 - Gmail (Mail)



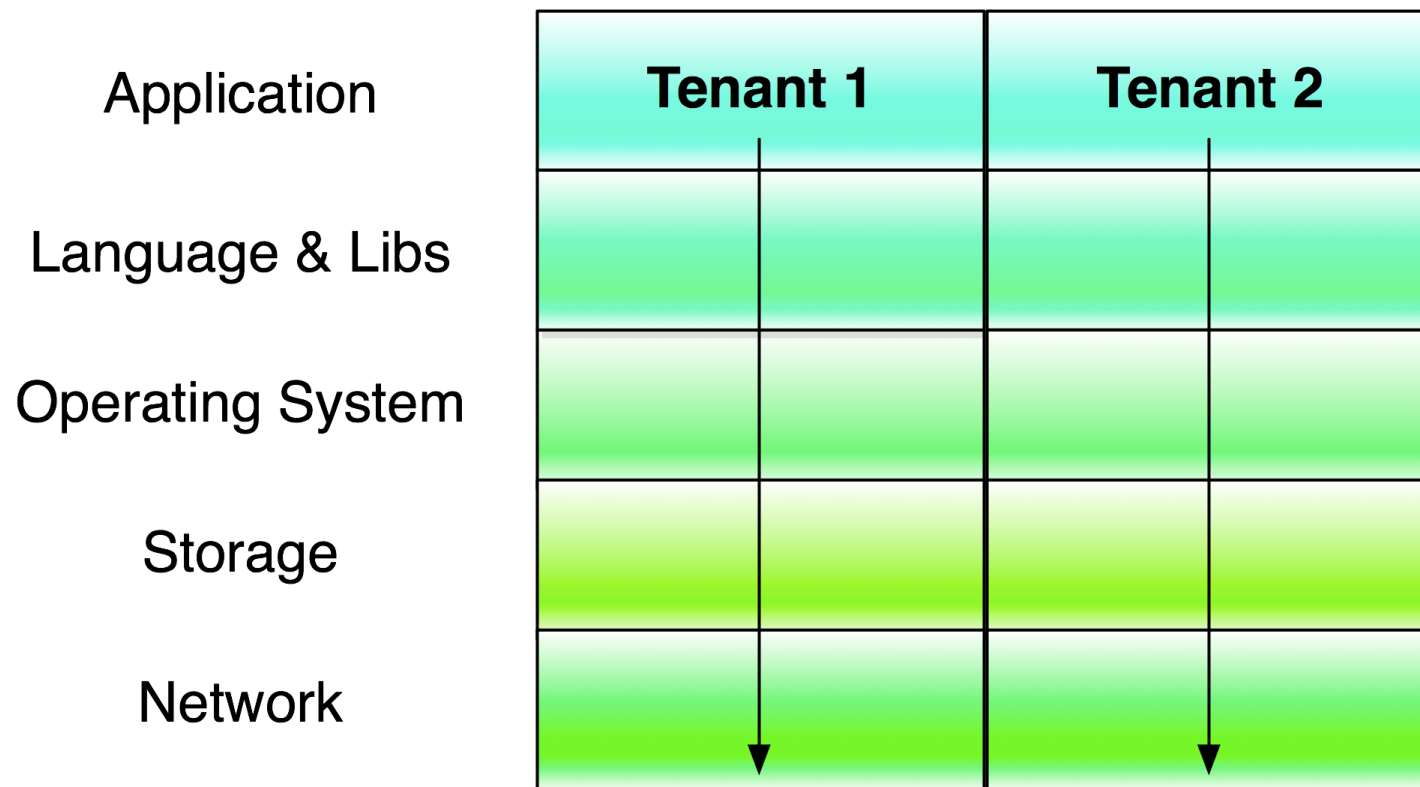
SaaS Tenancy Model



PaaS Tenancy Model



IaaS Tenancy Model



Two Faces of IaaS

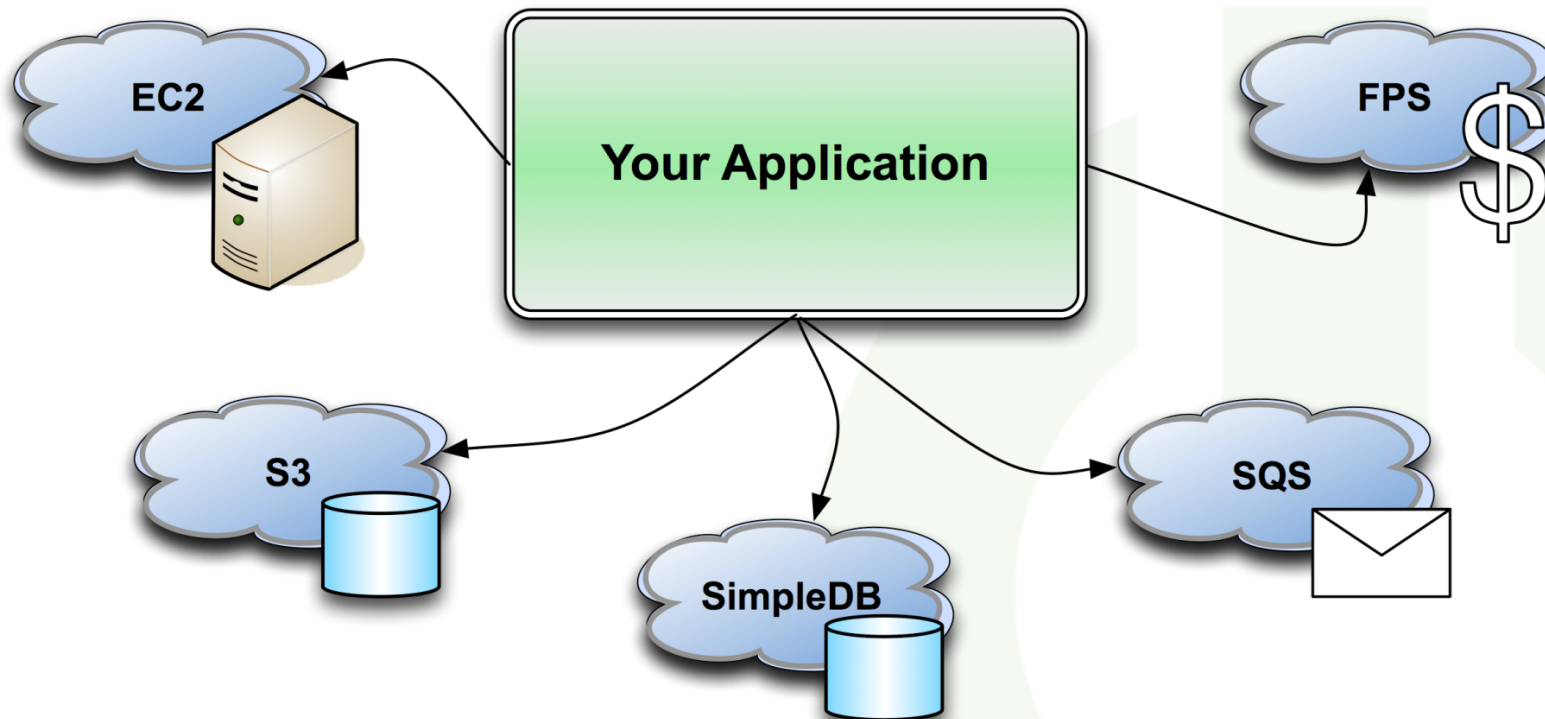
⚙ Infrastructure Web Services

- AWS is canonical example
- Web service model
- Pick and choose 'a la carte' web services
- Proprietary interfaces

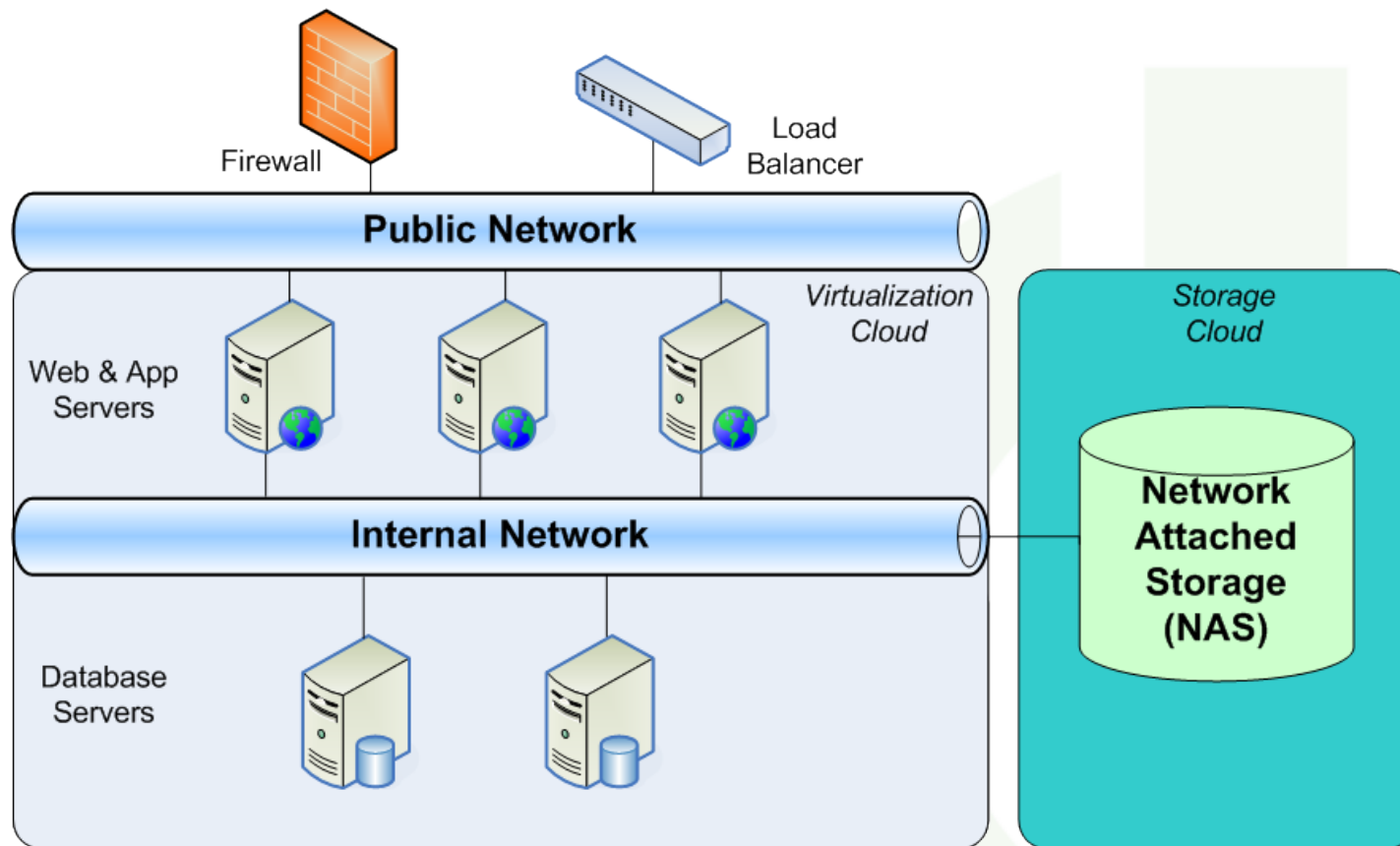
⚙ Cloudcenters

- GoGrid is canonical example
- Very common: FlexiScale, RackSpace/Mosso, & AppNexus
- Datacenter model 'in-the-sky'
- Standards-based datacenter-like infrastructure

Infrastructure Web Service



Cloudcenters



WHY STORAGE IN THE CLOUD?

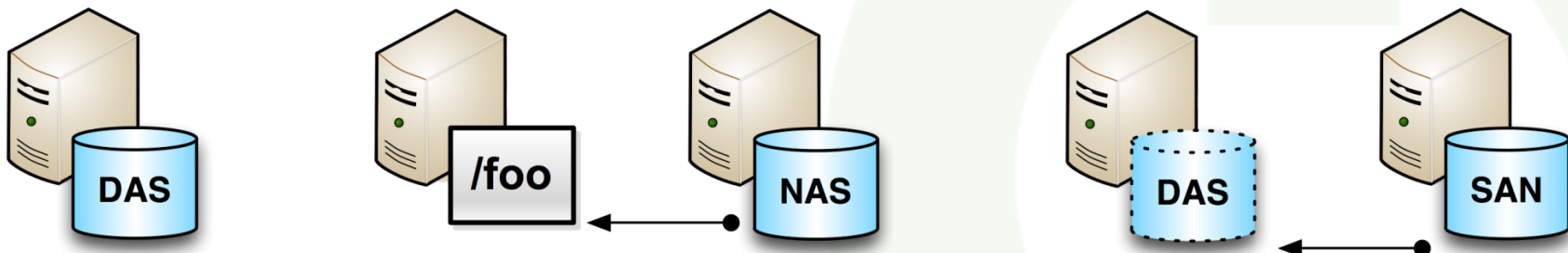
Many Drivers

- ⌚ Abstraction (hardware & network)
- ⌚ Cost efficiencies (at smaller scale only?)
- ⌚ Use cases demand it (DR, offsite backup, etc)
- ⌚ Not an either/or proposition
- ⌚ Data proximity (to compute)
 - Storage without compute?
 - Compute without storage?

STORAGE TODAY

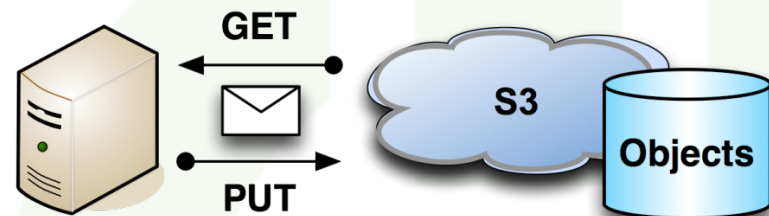
Traditional Datacenters

- ⌚ You choose: NAS, SAN, DAS, or some of each
- ⌚ High performance
- ⌚ High degree of control



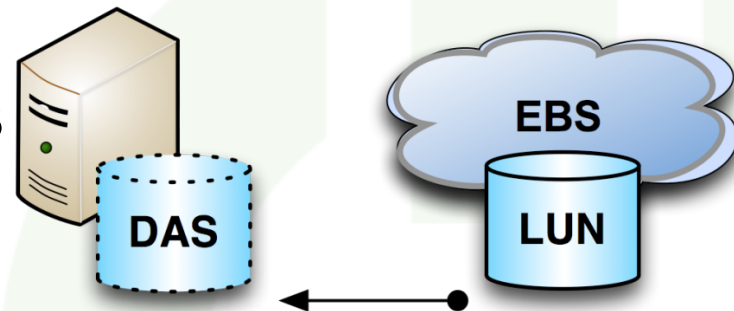
Amazon's S3

- 🔌 'Object-store'
 - URL PUT and GET
 - Simple usage
- 🔌 Proprietary & unique
 - Coding required
- 🔌 Variable performance
- 🔌 Infinitely scalable
- 🔌 Best for archives/SoR



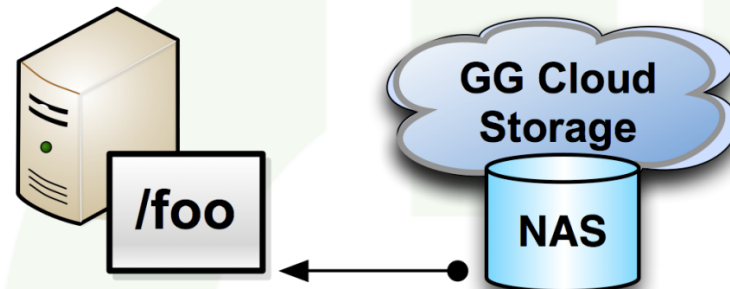
Amazon's Elastic Block Storage (EBS)

- ⌚ DAS (iSCSI?)
- ⌚ S3 integration
- ⌚ Performance is OK
- ⌚ Manage disks & servers
- ⌚ Overbuy problem



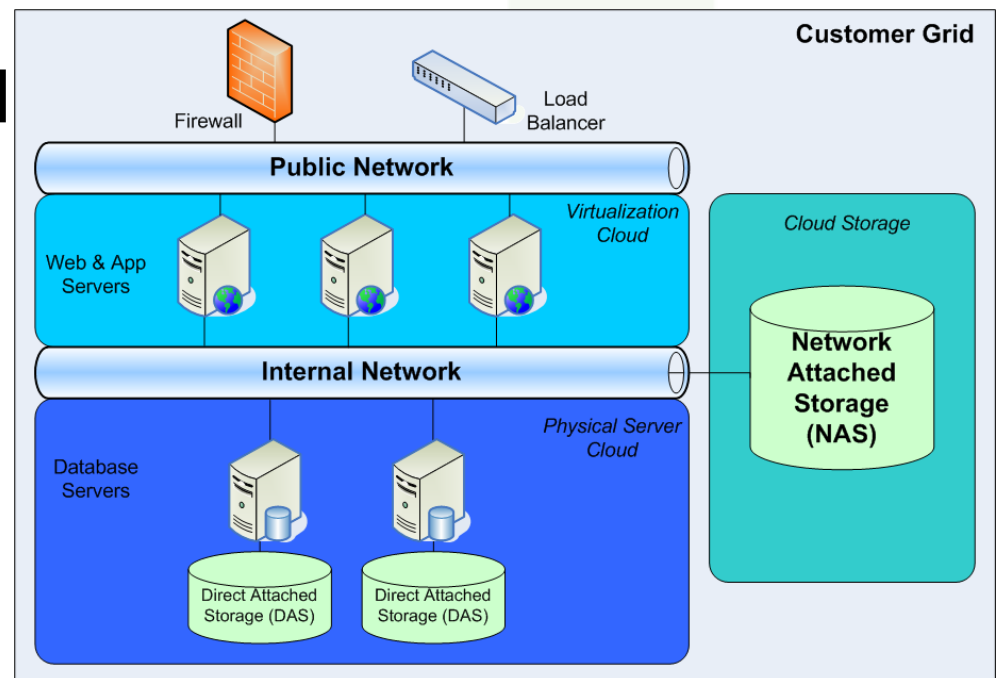
GoGrid's Cloud Storage (NAS in the Sky)

- 🔌 NAS
- 🔌 Performance == EBS
- 🔌 Infinitely scalable multi-tenant NetApp
- 🔌 Just works; pay-as-you-go



GoGrid's CloudConnect (DAS on-demand)

- 🔌 DAS + Servers
- 🔌 Physical server cloud
 - v1: manual
 - v2: fully automated
- 🔌 High speed spindles
- 🔌 High Performance
 - Fileservers or databases



Storage Type Comparison Matrix

	S3	EBS	GG Cloud Storage	GG Cloud Connect
Standards-based	No	Yes	Yes	Yes
Managed	Yes	No	Yes	No
Performance	Low	Med	Med	Very High
Scalable Capacity	High	Med	High	Low
Overbuy	No	Yes	No	Yes

MANAGEMENT CHALLENGES

Traditional Datacenter Challenges

- ❶ Over or under buying
 - Either ahead or behind your data growth curve
- ❷ Tends towards vendor/hardware lock-in
- ❸ Data replication costs
- ❹ Data backups, retention, and offsite storage
- ❺ Significant staffing requirements
 - Storage experts
 - Networking experts

IaaS Challenges

- ⌚ Security & data integrity
- ⌚ Data movement
- ⌚ Performance
- ⌚ Multi-tenancy impacts
- ⌚ Vendor scalability unknown
- ⌚ More expensive 'at scale'?
- ⌚ Vendor lock-in

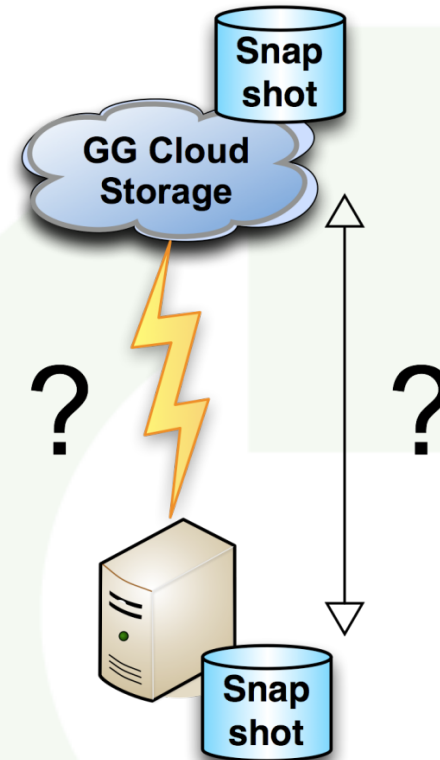
IaaS Challenges: Security

- ⚙️ How is your data handled?
- ⚙️ Auditability: Who stored what where?
- ⚙️ Data destruction guarantees
- ⚙️ Compliance and privacy measures?



IaaS Challenges: Data Movement

- 🔌 Migration & Synchronization
- 🔌 B/W is constrained
- 🔌 Standards?
 - rsync
 - Block-level replication
 - Snapshot replication
 - NetApp <-> GGCS ?
 - EMC <-> EBS ?



IaaS Challenges: Performance

- 🔌 S3: highly variant
 - 10x swings
 - Archives only?
- 🔌 EBS/GGCS
 - Good perf
 - Multi-tenancy impacts
 - EBS needs admin?
- 🔌 CloudConnect
 - Premium cost
 - Storage admin required?

IaaS Challenges: Multi-tenancy

- ❶ Rogue applications or customers consume too many resources
- ❷ Traditional datacenter similar, but lower scale
- ❸ 10^2 more tenants == more impacts
- ❹ Performance guarantees?

IaaS Challenges: Vendor Scalability

- ❏ Unproven scale from vendors
- ❏ AWS most proven, but still:
 - S3 performance issues gave rise to EBS
 - EBS itself limited
 - 1TB per volume (aka LUN)
 - 20 volumes max (default)
 - ~15 volumes mountable per EC2 instance
- ❏ Only time will tell

IaaS Challenges: Expensive at Scale

- ❶ Cloud isn't always cheaper
- ❷ Sufficient scale == sufficient sunk cost in:
 - Storage systems
 - People
 - Network
- ❸ Where's the line?
 - 20TB? 200TB? 1PB? More? Less?
 - Business case dependent
 - SmugMug 300+TB, 500M images
- ❹ Probably need both Cloud and DC long term anyway

IaaS Challenges: Vendor Lock-in

- ⌚ High level of 'stickiness' once significant storage in place
 - Cost of moving high
 - Need to keep data near my compute
- ⌚ S3 and similar are proprietary standards
 - Custom solutions not portable to other clouds
- ⌚ New value added features create new lock-in?
- ⌚ Need standards that make sense

Standards

- ⌚ Need to bridge DC data to cloud data
- ⌚ Smart synchronization & replication
- ⌚ Some standards exist we should reuse:
 - Filesystems: CIFS / NFS / ZFS?
 - Replication: NDMP / RSYNC
 - Network LUNs: iSCSI / FCoE
- ⌚ Need more:
 - Snapshots, incremental snapshots
 - Safe data encryption & destruction
 - Auditability: who put what where and when?

Vision

Your data, safe and secure, available anywhere at any time and performing to your applications' needs

Future

- ⌚ Current hardware lock-in replaced by open software
 - ZFS
 - NetApp functionality in open software
- ⌚ ZFS the great equalizer?
 - Nexenta, Solaris, Linux (via FUSE), and FreeBSD
- ⌚ Cross-cloud interoperability via open standards
- ⌚ Data-centric storage management tools
 - Not vendor-centric; storage-centric
- ⌚ Next generation of DR & offsite solutions
 - Block replication of virtual machines

Q&A

EMAIL

randyb@gogrid.com

GOGRID INFO

<http://www.gogrid.com>

BLOG

<http://neotactics.com/blog>

WORK BLOG

<http://blog.gogrid.com>

TWITTER

<http://twitter.com/randymbias>

GOGRID PROMO CODE

GGRB (\$100 credit)