Introduction to Cloud Computing

Dr. James Broberg (brobergj@csse.unimelb.edu.au)
Australian Postdoctoral Fellow
University of Melbourne

Introduction to Cloud Computing

- Programme:
  - Part I - An Introduction to Cloud Computing.
  - Part II - "Get your head in the Clouds" - MetaCDN Cloud Case Study.
  - Part III - Open Research Issues in Cloud Computing.

Cloud Computing - Part I

- What is "cloud computing"?
- What type of clouds exist?
- How does cloud computing differ from Cluster, Grid & Utility Computing?
- Examples of Computational Clouds
- Examples of Storage Clouds
- Examples of Platform/Application Clouds

Cloud Computing - Part II

- Case Study: MetaCDN
- Cloud CDN Overlay application that uses Compute and Storage Clouds combined with traditional physical resources.

Cloud Computing - Part III

- Open Research Issues in Cloud Computing
  - Interoperability
  - Portability
  - Reliability & Monitoring
  - Data Security and Jurisdiction
  - Innovative Cloud Pricing Models
  - Innovative Service Models

An Introduction to Cloud Computing

What is Cloud Computing?

- How can we define Cloud Computing?
  - What does (and does not) constitute a "cloud" platform?
- Enabling technologies for cloud computing
  - Virtual Machines
  - Virtualised Storage
  - Web Services

Cloud Computing defined...

- ".. a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet" - **Wikipedia**

Cloud Computing defined...

- "Clouds are hardware-based services offering compute, network and storage capacity where: Hardware management is highly abstracted from the buyer, Buyers incur infrastructure costs as variable OPEX, and Infrastructure capacity is highly elastic" - **McKinsey & Co. Report: "Clearing the Air on Cloud Computing"**

Cloud Computing defined...

- "Cloud computing has the following characteristics: (1) The illusion of infinite computing resources… (2) The elimination of an up-front commitment by Cloud users… (3). The ability to pay for use…as needed…"– **UCBerkeley RADLabs**

Cloud Computing defined...

- "... a pay-per-use model for enabling available, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." **- National Institute of Standards and Technology (NIST)**

Cloud Computing defined...

- "Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically re-configured to adjust to a variable load (scale), allow-ing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs." - **Paper by Vaquero et. al.: "A break in the clouds: towards a cloud definition"**

Common Ground?

- Pay-per-use (no commitment, utility prices).
- Elastic capacity - scale up/down on demand.
- Self-service interface.
- Resources are abstracted / virtualised.

Enabling Technologies

- Several key components have matured and are essential building blocks of many Cloud services:
  - Virtual Machines
  - Virtualised Storage
  - Web Services

Virtual Machines

- In the last 5 years VM technology has become pervasive and “cheap” / free
- The main players are:
  - VMWare [ESX/GSX Server, Workstation]
  - XenSource [XenEnterprise/Server/Express]
  - SWsoft/Parallels [Virtuozzo, Desktop]
  - Microsoft [Virtual Server]

Virtual Machines

- Factors contributing to rising profile:
  - Commodity multi-core, 64bit machines [dual/quad core, >>4 gig memory]
  - Hardware assisted virtualisation - using Intel VT(tm) and AMD Pacifica(tm) - runs unmodified OS at near native speeds
  - Integration of VM tech. into mainstream server OS's (RHEL, Ubuntu, Windows)

Virtual Machines

- Improves utilisation by multiplexing many VMs on one physical host (consolidation)
- Allows agile deployment and management of services.
- On demand cloning, (live) migration & checkpoint improves reliability.
- VM can be a self-contained unit of execution and migration!

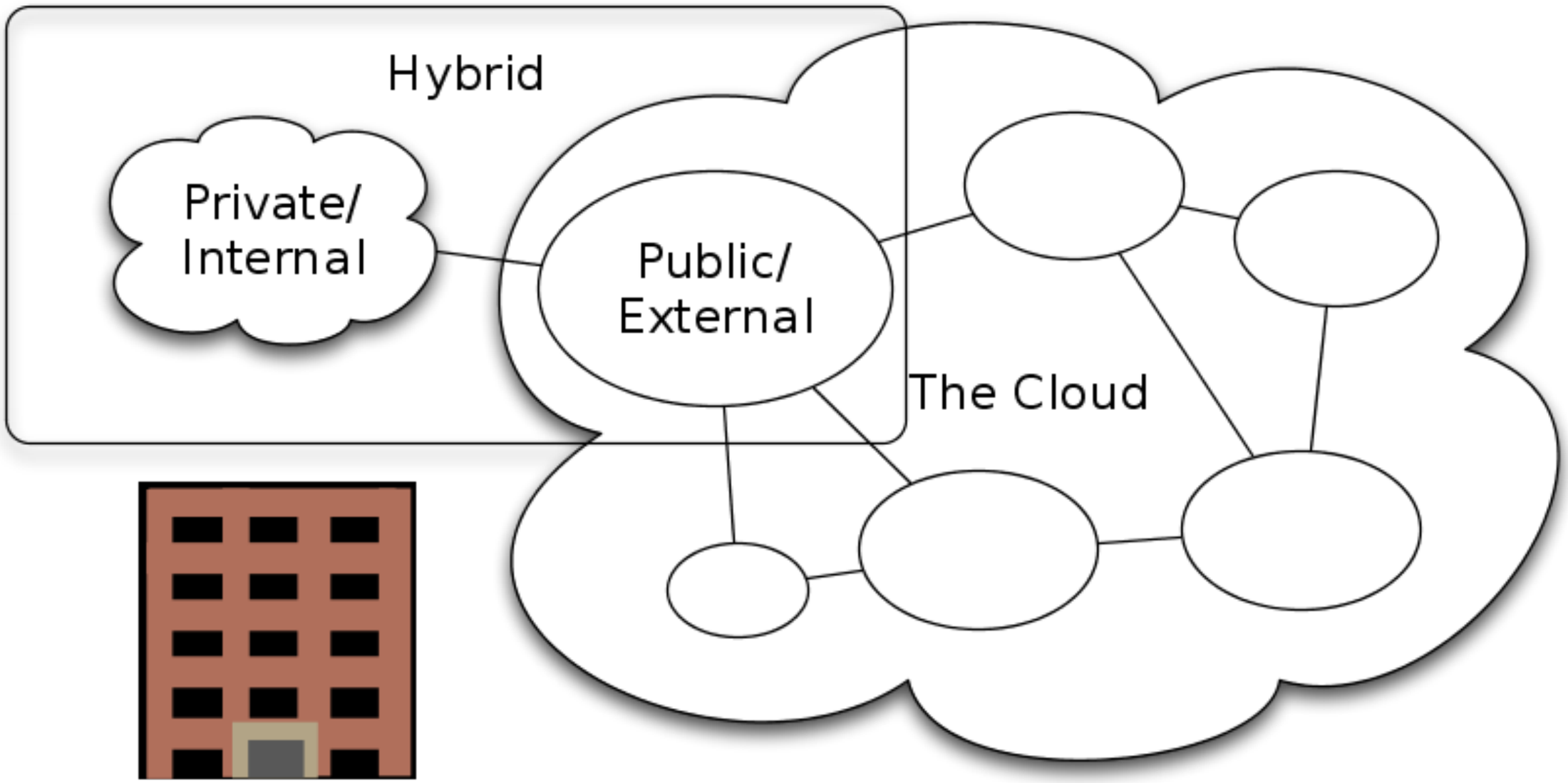
Virtualised Storage

- Distributed File Systems:
  - Google File System (GFS)
  - Hadoop Distributed File System (HDFS)
- Cluster File Systems:
  - VMware vStorage VMFS
  - XenServer Storage Pool

Web Services

- SOAP
  - XML Messages
  - Web Services Description Language (WSDL)
  - WS-*
- REST / RESTful
  - GET, POST, PUT, DELETE for HTTP (crud)
  - More lightweight HTTP / JSON schemes

Types of Clouds?



Cloud Computing Types

Public Cloud

- Publicly accessible, self-service model:
  - Access via well-defined & published Web Services (i.e. SOAP or REST).
  - Access via management portal.
- Free or pay-per-use.
- No ongoing contract or commitment.

Private Cloud

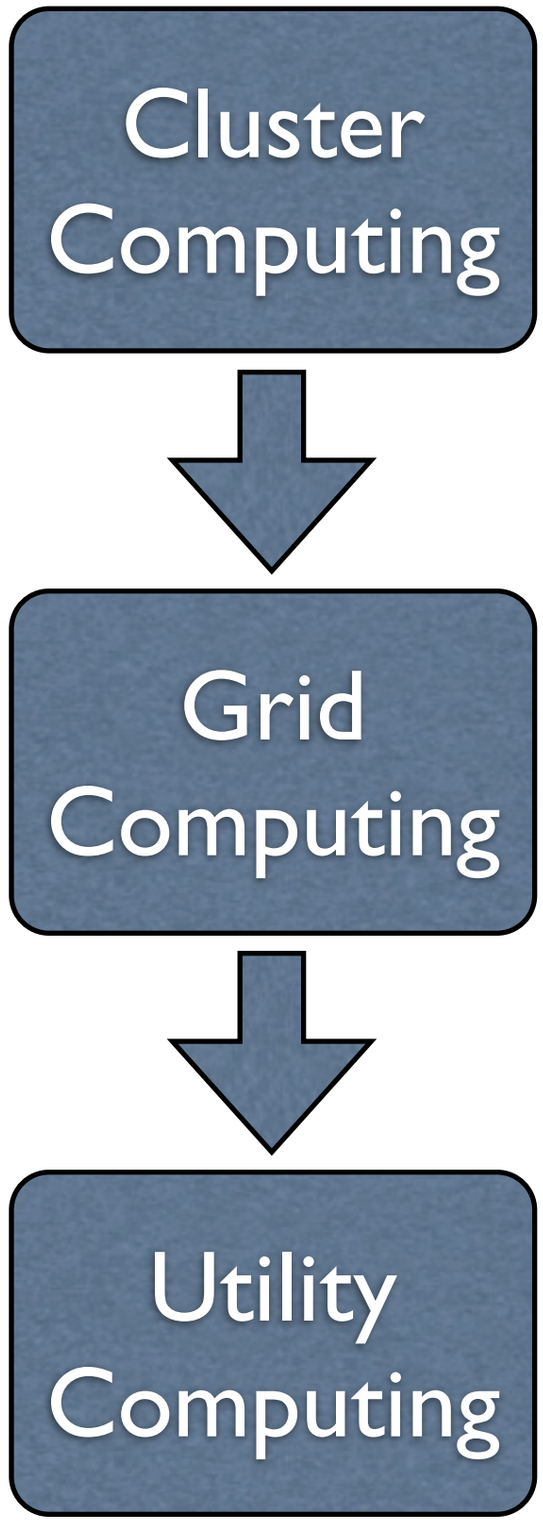
- Emulate Public Cloud on private / internal resources.
- Gives benefit of Clouds (elasticity, dynamic provisioning, multiplexing) while:
  - Maintaining control of resources (security).
  - Meeting Corporate/Regulatory req. (governance).
  - Option to scale out to Public Cloud ...

Hybrid Cloud

- Combination of private/internal and external Cloud resources.
- "Cloudbursting" to handle "Flash Crowds"
  - Provision additional resources from Public Clouds on-demand.
  - Release resource when no longer needed.
- Can outsource non-critical functions.

How did we get here?

- What about:
  - Cluster Computing?
  - Grid Computing?
  - Utility Computing?



Cluster Computing

- Clusters are linked computer systems that can co-operate to perform computations and deliver services:
  - Often function / appear as single server.
  - Typically linked over LAN.
  - Offers scalability over single-server.
  - Used for high-availability (redundant), load balancing, shared compute (w/ IPC).

Grid Computing

- Grids are autonomous and dynamic distributed resources contributed by multiple organisations.
  - Can offer computing, network, sensor and storage resources.
  - Resources are loosely coupled, heterogeneous, geographically dispersed.
  - Used in diverse fields: climate modelling, drug design and protein analysis to solve “Grand Challenges”.

Grid Computing

- Traditional Grid resource management techniques did not ensure fair and equitable access to resources in many systems:
  - Traditional metrics (throughput, waiting time, slowdown) failed to capture the more subtle requirements of users.
  - No incentives for users to be flexible about resource req. or job deadlines.
  - No provisions to accommodate users with urgent work.

Utility Computing

- Users assign a "utility" value to their jobs:
  - Utility is a fixed or time-varying valuation that captures various QoS constraints (deadline, importance, satisfaction)
  - Valuation is amount they are willing to pay a service provider to satisfy demands.
- Service providers attempt to maximise their own utility:
  - Utility may directly correlate with their profit.
  - Prioritise high yield (i.e. profit per unit of resource) user jobs.
- Shared Grid systems are then viewed as a marketplace, where users compete for resources based on the perceived utility / value of their jobs.

Utility Computing

Table 1 Summary of price setting and negotiation

System name	Price setting	Acceptance criteria	Penalty
Nimrod-G [1]	Fixed	Posted price	No
G-commerce [39]	Fixed	Current price > *mean_price*	N/A
Bellagio [4]	Fixed	'Threshold'-based	N/A
Mirage [12]	Fixed	'Winner determination problem'	N/A
Tycoon [21, 23]	Fixed	First/second price	No
Libra [35]	Fixed	Minimum cost	No
Libra+$ [41]	Fixed	Deadline/budget feasible	Yes (LibraSLA [42])
Li [24]	Fixed	*n*th Price	No
FirstPrice [13]	Variable	None	No
FirstReward [20]	Variable	Risk versus reward	Yes
FirstProfit [32]	Variable	Risk versus per-job reward	Yes
FirstOpportunity [32]	Variable	Affect on profit	Yes
Aggregate utility [5]	Variable	Contract feasibility/profitability	Yes

J Grid Computing (2008) 6:255–276
DOI 10.1007/s10723-007-9095-3

Utility Computing

Table 2 Summary of utility-driven resource management systems

System name	Allocation	Scheduler	Adm. control
Bellagio [4]	Combinatorial auction-based	Proportional share	N/A
Mirage [12]	Auction-based	Proportional share	N/A
Tycoon [21, 23]	Auction-based	Auction share	No
Libra [35]	N/A	Proportional share	Yes (basic)
Libra+$ [41]	N/A	Proportional share	Yes
Li [24]	Double auction	Proportional share	N/A
FirstPrice [13]	N/A	Highest unit value first	No
FirstReward [20]	N/A	Highest unit value first[a]	Yes
FirstProfit [32]	N/A	Highest per-job profit first	Yes
FirstOpportunity [32]	N/A	Highest total profit first	Yes
Aggregate utility [5]	N/A	Highest contract profit first	Yes

[a]Risk-aware.

J Grid Computing (2008) 6:255–276
DOI 10.1007/s10723-007-9095-3

Public Compute Clouds - Infrastructure as a Service

- Amazon EC2
- GoGrid
- Slicehost
- Mosso Cloud Servers

Amazon Elastic Compute Cloud (EC2)

- Provides resizable compute capacity in the cloud.
- Can be leveraged via:
  - Web Services (SOAP or REST).
  - AWS Management Console.
  - EC2 Command Line Tools.
- Provides hundreds of pre-made AMIs.

Amazon Elastic Compute Cloud (EC2)

- Standard Instances (S, L, XL)
  - Prop. more RAM than CPU
- High CPU Instances (M, XL)
  - Prop. more CPU than RAM
- Choice of Windows or Linux Images
  - US-East and EU-West Regions ($EU>$US)
  - "Value-added" AMIs cost more

Amazon EC2 On-Demand Instances

United States | Europe

Standard On-Demand Instances	Linux/UNIX Usage	Windows Usage
Small (Default)	$0.10 per hour	$0.125 per hour
Large	$0.40 per hour	$0.50 per hour
Extra Large	$0.80 per hour	$1.00 per hour
High CPU On-Demand Instances	**Linux/UNIX Usage**	**Windows Usage**
Medium	$0.20 per hour	$0.30 per hour
Extra Large	$0.80 per hour	$1.20 per hour

United States | **Europe**

Standard On-Demand Instances	Linux/UNIX Usage	Windows Usage
Small (Default)	$0.11 per hour	$0.135 per hour
Large	$0.44 per hour	$0.54 per hour
Extra Large	$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 EC2 Reserved Instances

United States | Europe

Linux/UNIX	One-time Fee		
Standard Reserved Instances	**1 yr Term**	**3 yr Term**	**Usage**
Small (Default)	$325	$500	$0.03 per hour
Large	$1300	$2000	$0.12 per hour
Extra Large	$2600	$4000	$0.24 per hour
High CPU Reserved Instances	**1 yr Term**	**3 yr Term**	**Usage**
Medium	$650	$1000	$0.06 per hour
Extra Large	$2600	$4000	$0.24 per hour

United States | **Europe**

Linux/UNIX	One-time Fee		
Standard Reserved Instances	**1 yr Term**	**3 yr Term**	**Usage**
Small (Default)	$325	$500	$0.04 per hour
Large	$1300	$2000	$0.16 per hour
Extra Large	$2600	$4000	$0.32 per hour
High CPU Reserved Instances	**1 yr Term**	**3 yr Term**	**Usage**
Medium	$650	$1000	$0.08 per hour
Extra Large	$2600	$4000	$0.32 per hour

Amazon EC2 Windows Server

Pricing for Instances Running Windows Server

SQL Server Express, Microsoft IIS and ASP.NET can be used on any Amazon EC2 instance running Windows Server for no additional cost.

United States | Europe

Standard Instances	**Windows**	**Windows with Authentication Services**
Small (Default)	$0.125 per hour	$0.25 per hour
Large	$0.50 per hour	$0.75 per hour
Extra Large	$1.00 per hour	$1.50 per hour
High CPU Instances	**Windows**	**Windows with Authentication Services**
Medium	$0.30 per hour	$0.50 per hour
Extra Large	$1.20 per hour	$2.00 per hour

Pricing for Instances Running Windows Server with SQL Server Standard

United States | Europe

Instance Type	**Windows**	**Windows with Authentication Services**
Standard Large	$1.10 per hour	$1.35 per hour
Standard Extra Large	$2.20 per hour	$2.70 per hour
High CPU Extra Large	$2.40 per hour	$3.20 per hour

Amazon EC2 IBM Platforms

Pricing for Instances Running IBM Lotus Web Content Management Server Standard Edition

Instance Type	US Region	EU Region
Standard Large	$2.48 per hour	$2.52 per hour
Standard Extra Large	$4.82 per hour	$4.90 per hour
High CPU Extra Large	$8.71 per hour	$8.79 per hour

Pricing for Instances Running IBM WebSphere Portal Server and IBM Web Content Management Server Standard Edition

Instance Type	US Region	EU Region
Standard Large	$6.39 per hour	$6.43 per hour
Standard Extra Large	$12.64 per hour	$12.72 per hour
High CPU Extra Large	$24.35 per hour	$24.43 per hour

Amazon EC2 Success Stories

- SmugMug - Online Photo Sharing
  - EC2 instances handles batch ops: upload, rotate, transcode, watermark, etc...
  - Hybrid approach: Own datacenter + Cloud, autoscale up/down as needed.
- Animoto - Creates Cool Videos from Photos & Music
  - Scaled from 50 instances of EC2 up to 3,500 instances of EC2 within days.

GoGrid Cloud Hosting

- Provides pre-made images:
  - Windows and Linux.
  - “Value-added” stacks on top.
- Range of (fixed) instance sizes.
- Offers hybrid dedicated / cloud server infrastructure.
- Free hardware LB and autoscaling.

Slicehost VPS Hosting

- Fixed size "slices".
  - Slice size measured by RAM, not CPU.
- Pre-made Linux images.
- No SLA, simply "best effort".
- Monthly billing (not per-hour).
  - Does that fit "Cloud" definition?

Mosso Cloud Servers

- Fixed size Cloud Servers.
  - Server size measured by RAM, not CPU
- Offers range of pre-made Linux instances
- Offers hybrid dedicated / cloud server infrastructure.
- Offers fixed IPs and A-DNS by default.

Compute Cloud Comparison

	Amazon EC2	GoGrid	Slicehost	Mosso Cloud Servers
Instance Cost	$0.10-$1.28 / hr	$0.095-$1.32 / hr	$20-$80 / month	$0.015-$0.96 / hr
Linux Support	Yes	Yes	Yes	Yes
Windows Support	Yes	Yes	No	No
Cores/CPU	1-20	1-6	4 (weighted)	4 (weighted)
Memory	1.7GB -15GB	0.5GB - 8GB	256MB - 15.5GB	256MB - 15GB
Persistent Storage	No (req. S3/EBS)	Yes	Yes	Yes
Default Public IP	No (req. Elast. IP)	Yes	Yes	Yes
Managed DNS	No	Yes	No	Yes
Support Cost	$400 / 20% Spend	Free	Free	$100
Hybrid Cloud?	No	Yes	No	Yes
SLA	99.95%	100%	N/A	100%
Running instances	20*	?	?	Unlimited*
API	Yes	Yes	Yes	Yes

Public Storage Clouds - Storage as a Service

- Amazon Simple Storage Service (S3)
- Amazon CloudFront
- Nirvanix Storage Delivery Network
- Mosso Cloud Files
- Microsoft Azure Storage Services

Amazon Simple Storage Service (S3)

- Amazon S3 was launched US in March 2006, and EU in November 2007.
- REST/SOAP interfaces to storage resources
- Users can read, write or delete an unlimited # of objects, 1 byte to 5 GB each.
- Files stored in flat "bucket" structure.
- BitTorrent can be used to minimise cost.

Amazon CloudFront

- Launched in November 2008.
- CDN service that adds 14 edge locations (8 in the US, 4 in EU, and 2 in Asia).
- CloudFront does not offer persistent storage.
  - Analogous to proxy cache, with files deployed to and removed from CF locations based on demand.

Nirvanix Storage Delivery Network (SDN)

- Launched on September 2007.
- Has 3 US locations (San Diego, Seacaucus, Houston), 1 EU (Frankfurt), 1 Asia (Tokyo)
- Launched with SLA-backed uptime guarantee when S3 was only best-effort.
- Offers automatic file replication over sites in the SDN, and supports file sizes up to 256 GB.

Mosso CloudFiles

- Launched late 2008.
- Base storage location in Dallas Texas.
- Files stored in flat “bucket” structure.
- Offers integration with Limelight CDN.
  - Adds more than 70 locations.
  - No transfer costs from origin to CDN.

Microsoft Azure Storage Service

- CTP launched in late 2008.
- Hosts Blobs, tables, and queues.
- Blobs (files) stored in flat "bucket" structure.
- Simple language agnostic REST interfaces.

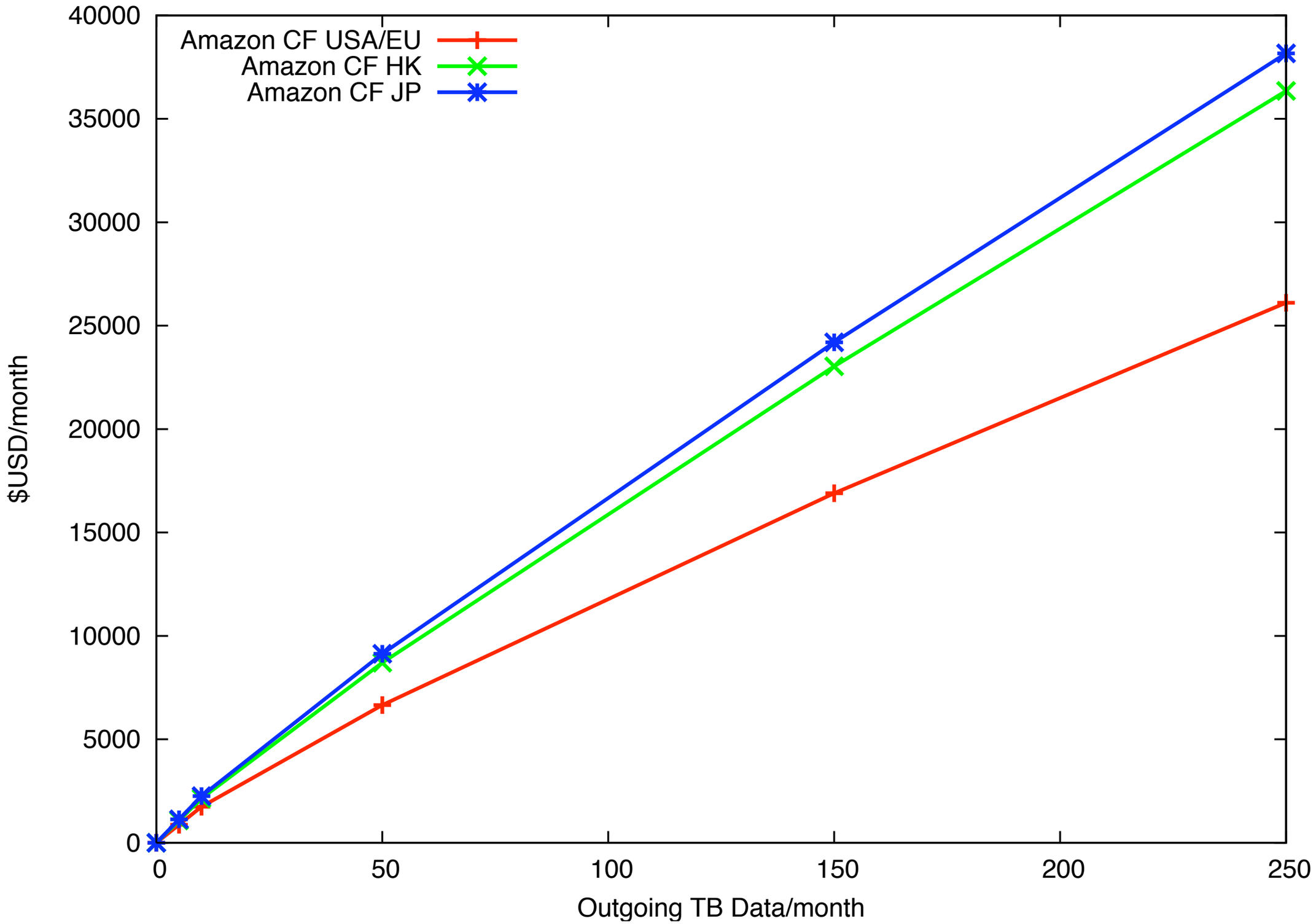
Cost Structures

	Nirvanix Global SDN < 2TB	Amazon S3 USA < 10TB	Amazon S3 Europe < 10 TB	Amazon CloudFront < 10 TB	Mosso CloudFiles < 5TB
Incoming data ($/GB)	0.18	0.1	0.1	N/A	0.08
Outgoing data ($/GB)	0.18	0.17	0.17	0.17-0.21	0.22
Storage ($/GB/month)	0.25	0.15	0.18	N/A	0.15
Requests ($/1,000 PUT/POST/LIST)	0.00	0.01	0.012	N/A	0.02
Requests ($/10,000 GET)	0.00	0.01	0.012	0.010-0.013	0.00

Feature Comparison

	Nirvanix Global SDN	Amazon S3	Amazon CloudFront	Mosso CloudFiles
SLA	99.9	99-99.9	99-99.9	99.9
Max File Size	256GB	5GB	5GB	5GB
US PoP	Yes	Yes	Yes	Yes
EU PoP	Yes	Yes	Yes	Yes
Asia PoP	Yes	No	Yes	Yes
AUS PoP	No	No	No	Yes
Per file ACL	Yes	Yes	Yes	Yes
Automatic replication of files	Yes	No	Yes	Yes
Developer API's / Web Services	Yes	Yes	Yes	Yes

Pricing comparison



Public App Clouds - Platform as a Service

- Google AppEngine
- Microsoft Azure Services Platform
- Salesforce.com / Force.com
- Mosso Cloud Sites
- Heroku

Google AppEngine

- Python stack
- Java stack (J2EE 6, JDO, JPA, JSP) [new]

- App Engine serving architecture allows real time auto-scaling without virtualization.
  - ...provided you use limited native APIs.
  - ...provided you use Google APIs.
    - URLFetch, Datastore, memcache, etc.

Google AppEngine - Free vs Billed

Resource	Free Default Quota		Billing Enabled Quota	
	Daily Limit	Maximum Rate	Daily Limit	Maximum Rate
Requests	1,300,000 requests	7,400 requests/minute	43,000,000 requests	30,000 requests/minute
Outgoing Bandwidth (billable, includes HTTPS)	10 gigabytes	56 megabytes/minute	10 gigabytes free; 1,046 gigabytes maximum	740 megabytes/minute
Incoming Bandwidth (billable, includes HTTPS)	10 gigabytes	56 megabytes/minute	10 gigabytes free; 1,046 gigabytes maximum	740 megabytes/minute
CPU Time (billable)	46 CPU-hours	15 CPU-minute/minute	46 CPU-hours free; 1,729 CPU-hours maximum	72 CPU-minute/minute

Source: http://code.google.com/appengine/docs/quotas.html

Google AppEngine - Billed

- $0.10 per CPU core hour (actual CPU time an app uses to process requests and any Datastore usage).
- $0.10 per GB bandwidth incoming, $0.12 per GB bandwidth outgoing. (in/out data, data usage of URLFetch API and Email API).
- $0.15 per GB of data stored per month.
- $0.0001 per email recipient for emails sent

Azure Services Platform

- Hosted .NET Stack (C# ,VB.Net, ASP.NET).
- Java & Ruby SDK for .NET Services also available.
- Windows Azure Fabric Controller:
  - Provides Scaling and reliability.
  - Manages memory resources and load balancing.
- .NET Service Bus registers & connects applications together.
- .NET Access Control identity providers, including enterprise directories and Windows LiveID.
- .NET Workflow allows construction and execution of workflow instances.

Salesforce.com

- Force.com PaaS allows developers to create add-on functionality that integrates into main Salesforce CRM SaaS app.
- Hosted Apex/Visualforce app
  - Apex is a proprietary Java-like language.
  - Visialforce is a XML-like syntax for building UIs in HTML, AJAX or Flex.
- AppExchange paid & free app directory.

Mosso Cloudsites

- Windows and Linux stacks:
  - Linux - PHP 5, Perl, Python, MySQL 5
  - Win - .Net 2.0, 3+3.5, ASP, MS SQL 2k8
- Runs application with very little (if any) modification (no custom APIs)
- Transparent load balancing and autoscale.
- Can run PHP, ASP, HTML, Python, and Perl pages concurrently from 1 site!

Heroku

- A platform for instant deployment of Ruby/ Rails web apps.
- Servers are invisibly managed by the platform, and are never exposed to you.
- Apps are dispersed across different CPU cores, maximizing performance and minimizing contention.
- Logic can route around failures.

Private / DYI Clouds

- Many platforms exist to run your own Cloud:
  - OpenNebula
  - Eucalyptus
  - Nimbus / Globus Workspaces
  - Others ...

OpenNebula.org

Eucalyptus Systems

the globus alliance

OpenNebula

- Open-source Virtual Infrastructure management software.
- Supports dynamic resizing, partitioning and scaling of resources
- Supports Private / Public / Hybrid Cloud models:
  - Turn cluster into private cloud.
  - Can expose service to public.
  - Cloud plugins (EC2, GoGrid) enable hybrid model.
- Offers XML-RPC Web Services

Eucalyptus

- Open-source (BSD) software infrastructure for implementing "cloud computing" on clusters.
- Public or Private Cloud
- Web Service is fully AWS API compliant:
  - EC2, S3 and EBS are all emulated.
  - Enables a Hybrid approach - mix and match EC2 and Eucalyptus resources.

Nimbus / Globus

- Open-source framework to turn your cluster into an IaaS cloud.
- Offers two Web Service interfaces:
  - Amazon EC2 WSDLs.
  - Grid community WSRF.
- Several test "Science Clouds" deployed:
  - U. of Chicago, Florida, Perdue & Marsarky U.

Get your head in the Clouds

Cloud Case Study - MetaCDN

MetaCDN: Enabling High Performance, Low Cost Content Storage and Delivery via the Cloud

Dr. James Broberg (brobergj@csse.unimelb.edu.au)
http://www.csse.unimelb.edu.au/~brobergj

http://www.metacdn.org

Australian Government
Australian Research Council

THE UNIVERSITY OF MELBOURNE

MELBOURNE SCHOOL OF ENGINEERING

RMIT UNIVERSITY

meta CDN

Content Delivery Networks (CDNs)

- What is a CDN?
  - *Content Delivery Networks (CDNs) such as Akamai, Mirror Image and Limelight place web server clusters in numerous geographical locations to improve the responsiveness and locality of the content it hosts for end-users.*

Existing CDN providers

- Akamai is the clear leader in coverage and market share (approx. 80%), however...
  - Price is prohibitive for SME, NGO, Gov...
    - Anecdotally 2−15 times more expensive than Cloud Storage, and require 1−2 year commitments and min. data use (10TB+)
- Academic CDNs include Coral, Codeen, Globule, however...
  - No SLA / QoS provided, only ‘best effort’

Major CDN pricing

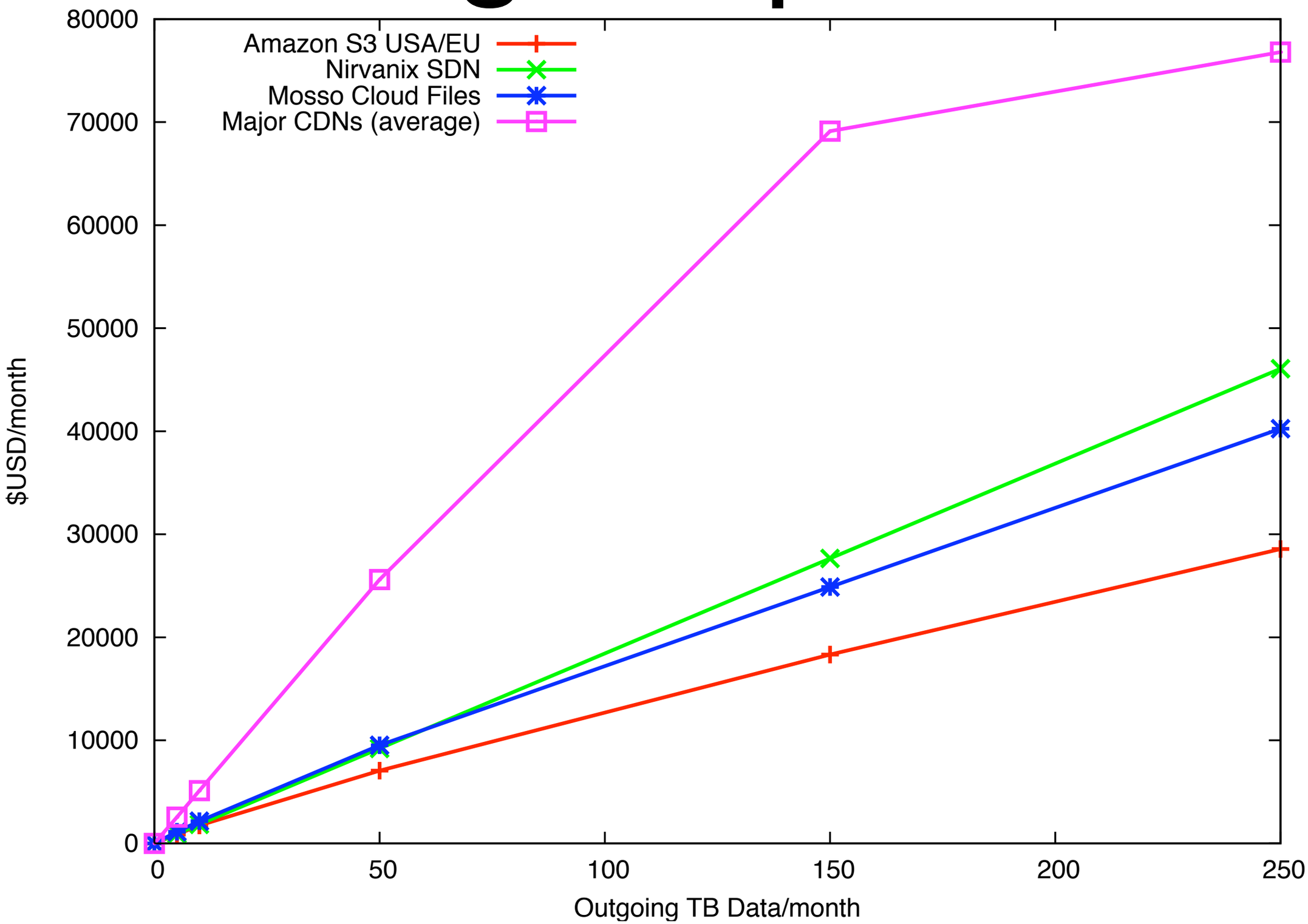
- Most major CDN providers do not publish prices
- "Average" prices from the 4-5 major CDNs in the market:
  - 50TB: $0.40 - $0.50 per GB delivered
  - 100TB: $0.25 - $0.45 per GB delivered
  - 250TB: $0.10 - $0.30 per GB delivered
- Information taken from Dan Rayburn @ www.cdnpricing.com, StreamingMedia.com

Storage Clouds
'Storage as a Service'

Now!

??????

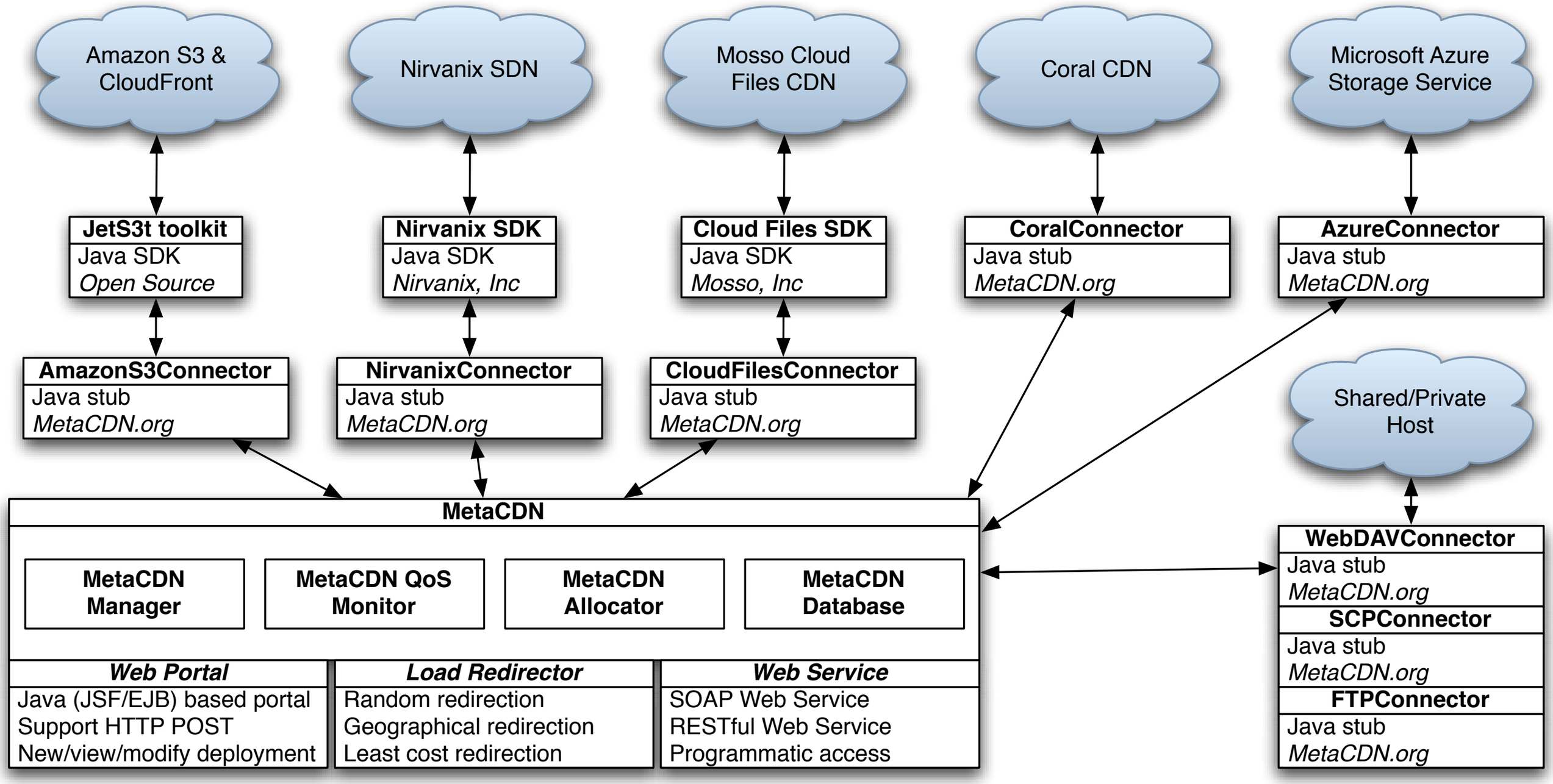
Pricing comparison



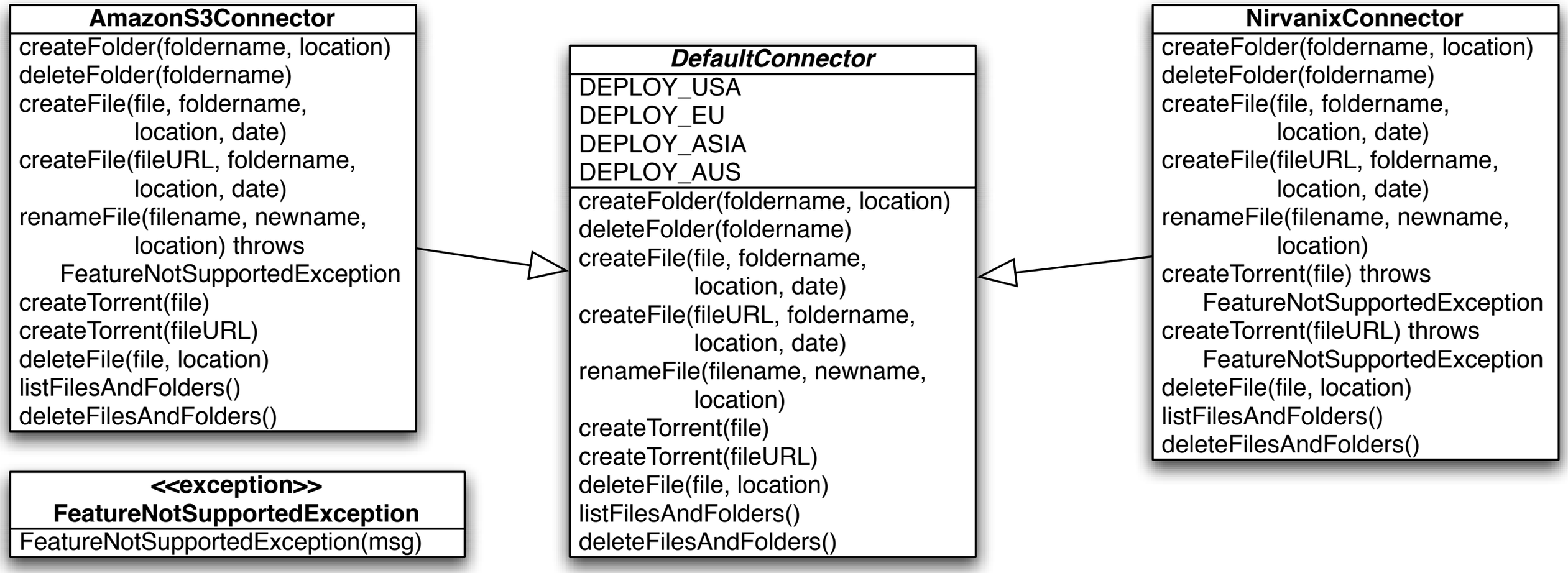
Introducing MetaCDN

- What if we could create a low-cost, high performance overlay CDN using these storage clouds?
    - MetaCDN harnesses the diversity of multiple storage cloud providers, offering a choice of deployment location, features and pricing.
    - MetaCDN provides a uniform method / API for harnessing multiple storage clouds.
    - MetaCDN provides a single namespace to cover all supported providers, making it simple to integrate into origin sites, and handles load balancing for end-users.
    - The MetaCDN system offers content creators (who are not necessarily programmers!) a trivial way to harness the power of multiple cloud storage providers via a web portal

How MetaCDN works



How MetaCDN works



MetaCDN Allocator

- The MetaCDN Allocator allows users to deploy files either directly or from a public *origin* URL, with the following options:
  - Maximise coverage and performance
  - Deploy content in specific locations
  - Cost optimised deployment
  - Quality of Service (QoS) optimised

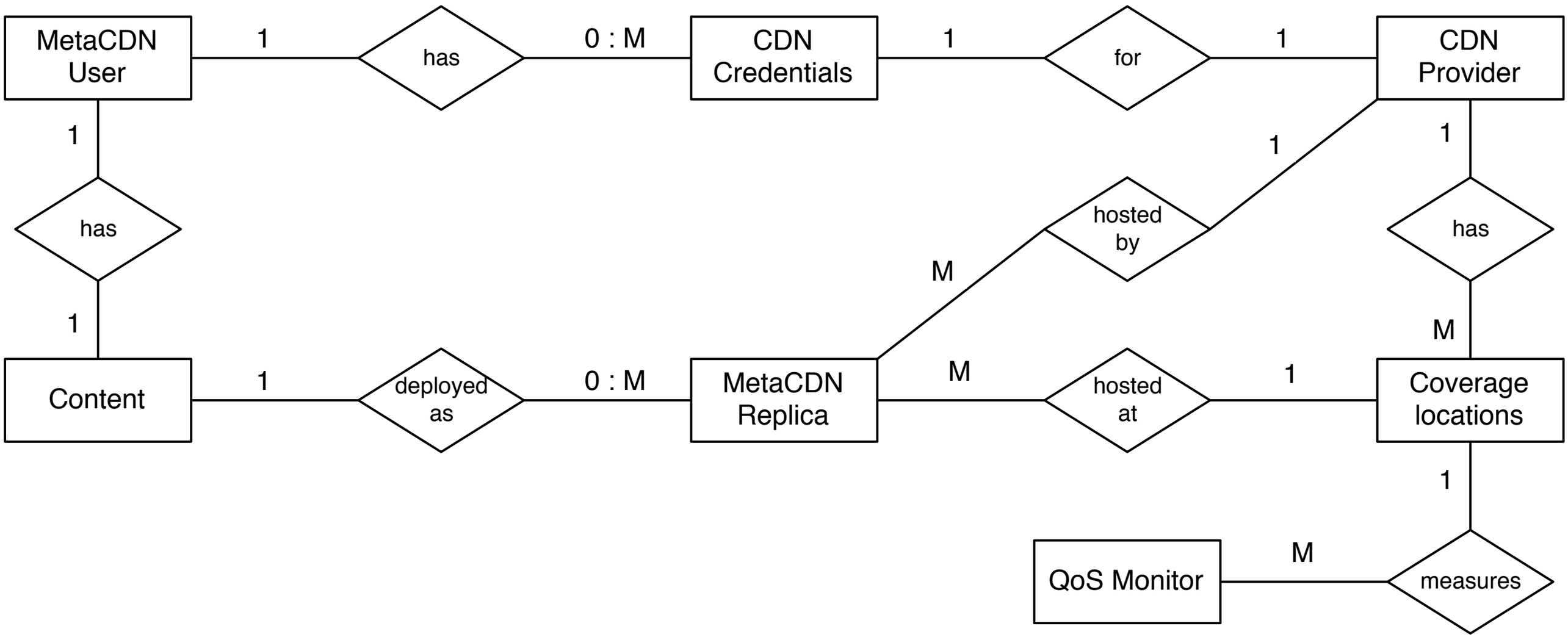
MetaCDN Manager

- The MetaCDN Manager ensures that:
  - All current deployments are meeting QoS targets (where applicable)
  - Replicas are removed when no longer required (minimising cost)
  - A users' budget has not been exceeded, by tracking usage (i.e. storage/download)

MetaCDN QoS Monitor

- The MetaCDN QoS Monitor:
  - Tracks the performance of participating providers at all times
  - Monitors and records throughput, response time and uptime from a variety of locations
  - Ensures that upstream providers are meeting their Service Level Agreements

MetaCDN Database



MetaCDN Web Portal

- Developed using Java Enterprise and Java Server Faces (JSF) technologies
- JPA/MySQL back-end to store persistent data
- Web portal acts as the entry point to the system **and** application-level load balancer
- Most suited for small or ad-hoc deployments, and especially useful for less technically inclined content creators.

Don't have a MetaCDN account?

Register

Sign in

Username: master

Password: ••••••

Login

This project is supported by the:

Australian Government
Australian Research Council

Register new MetaCDN account:

Username:	joebloggs	Full Name:	Joe Bloggs
Password:	••••••	Email:	joe@blogs.com

Preferred Providers:

- [x] Nirvanix SDN
- [x] Amazon S3
- [x] Mosso Cloud Files
- [ ] Microsoft Azure Storage Service
- [ ] Shared/Private Host

Nirvanix SDN | **Amazon S3** | Mosso Cloud Files | Microsoft Azure Storage Service | Shared/Private Host

Enter your Amazon S3 Credentials:

AWS Access Key:	**********
AWS Secret Key:	**********
Enable CloudFront:	[x]

Register Cancel

This project is supported by the:

Welcome to the MetaCDN Portal!

Deploy content | View existing content | Deployment Map

View existing replicas | MetaCDN Analytics

Edit Account | Logout

This project is supported by the:

Australian Government
Australian Research Council

Sideload Content:

Choose URL: http://pitchfork.com/media/frontend/images/header_log

Deployment region(s):
- [x] North America
- [x] Europe
- [x] Asia
- [ ] Australasia

Host Until: 30/05/2009

Deploy Cancel

This project is supported by the:

View Content:

ID	Origin URL	GUID	MetaCDN URL	File Name	Host Until?	Downloads	Last Access?	# Replicas	Modify	Delete
1	http://www.dtvforum.info /uploads/av-42774.jpg	15AA390F-29E3-13EE-5334-EA21C6D8F918	http://www.metacdn.org:8080/MetaCDN /FileMapper?guid=15AA390F-29E3-13EE-5334-EA21C6D8F918	av-42774.jpg		12	Thu Apr 16 12:04:57 EST 2009	10	Modify	Delete
2	http://pitchfork.com/media /frontend/images /header_logo.gif	A3ADEB4A-D395-E492-E792-00F958F0D63D	http://www.metacdn.org:8080/MetaCDN /FileMapper?guid=A3ADEB4A-D395-E492-E792-00F958F0D63D	header_logo.gif		0		10	Modify	Delete

Back

This project is supported by the:

Australian Government
Australian Research Council

View Replicas:

|< << < > >> >|

Replica ID	Provider ID	Location	Replica URL	Download Count	Last Access Time	Host Until?	Modify	Delete	Show on map
1	Nirvanix SDN	DEPLOY_USA	http://node1.nirvanix.com/MetaCDN/brobergj/metacdn-us-master/av-42774.jpg	1	Thu Apr 16 12:02:09 EST 2009		Modify	Delete	Map
2	Nirvanix SDN	DEPLOY_USA	http://node4.nirvanix.com/MetaCDN/brobergj/metacdn-us-master/av-42774.jpg	0			Modify	Delete	Map
3	Nirvanix SDN	DEPLOY_USA	http://node5.nirvanix.com/MetaCDN/brobergj/metacdn-us-master/av-42774.jpg	1	Thu Apr 16 12:02:16 EST 2009		Modify	Delete	Map
4	Amazon S3	DEPLOY_USA	http://metacdn-us-master.s3.amazonaws.com/av-42774.jpg	2	Thu Apr 16 12:04:39 EST 2009		Modify	Delete	Map
5	Nirvanix SDN	DEPLOY_EU	http://node2.nirvanix.com/MetaCDN/brobergj/metacdn-eu-master/av-42774.jpg	0			Modify	Delete	Map
6	Amazon CloudFront	DEPLOY_EU	http://d5z0tvntvsn23.cloudfront.net/av-42774.jpg	2	Thu Apr 16 12:04:51 EST 2009		Modify	Delete	Map
7	Amazon S3	DEPLOY_EU	http://metacdn-eu-master.s3.amazonaws.com/av-42774.jpg	1	Thu Apr 16 12:04:54 EST 2009		Modify	Delete	Map
8	Nirvanix SDN	DEPLOY_ASIA	http://node3.nirvanix.com/MetaCDN/brobergj/metacdn-asia-master/av-42774.jpg	2	Thu Apr 16 12:04:57 EST 2009		Modify	Delete	Map
9	Amazon CloudFront	DEPLOY_ASIA	http://d5z0tvntvsn23.cloudfront.net/av-42774.jpg	1	Thu Apr 16 12:04:48 EST 2009		Modify	Delete	Map
10	Mosso Cloud Files	DEPLOY_AUS	http://cdn.cloudfiles.mosso.com/c52122/av-42774.jpg	2	Thu Apr 16 12:04:35 EST 2009		Modify	Delete	Map

Back

Deployment Map:

Back

MetaCDN Web Service

- Makes all functionality available via Web Services (SOAP & REST/HTTP)
  - Web interface is useful for novices and for ad-hoc deployments, but doesn't scale
  - Larger customers have 1,000's - 10,000 - 100,000s of files that need deployment
  - Let them automate their deployment and management via Web Services!
  - Perfect for Mashup developers!!!

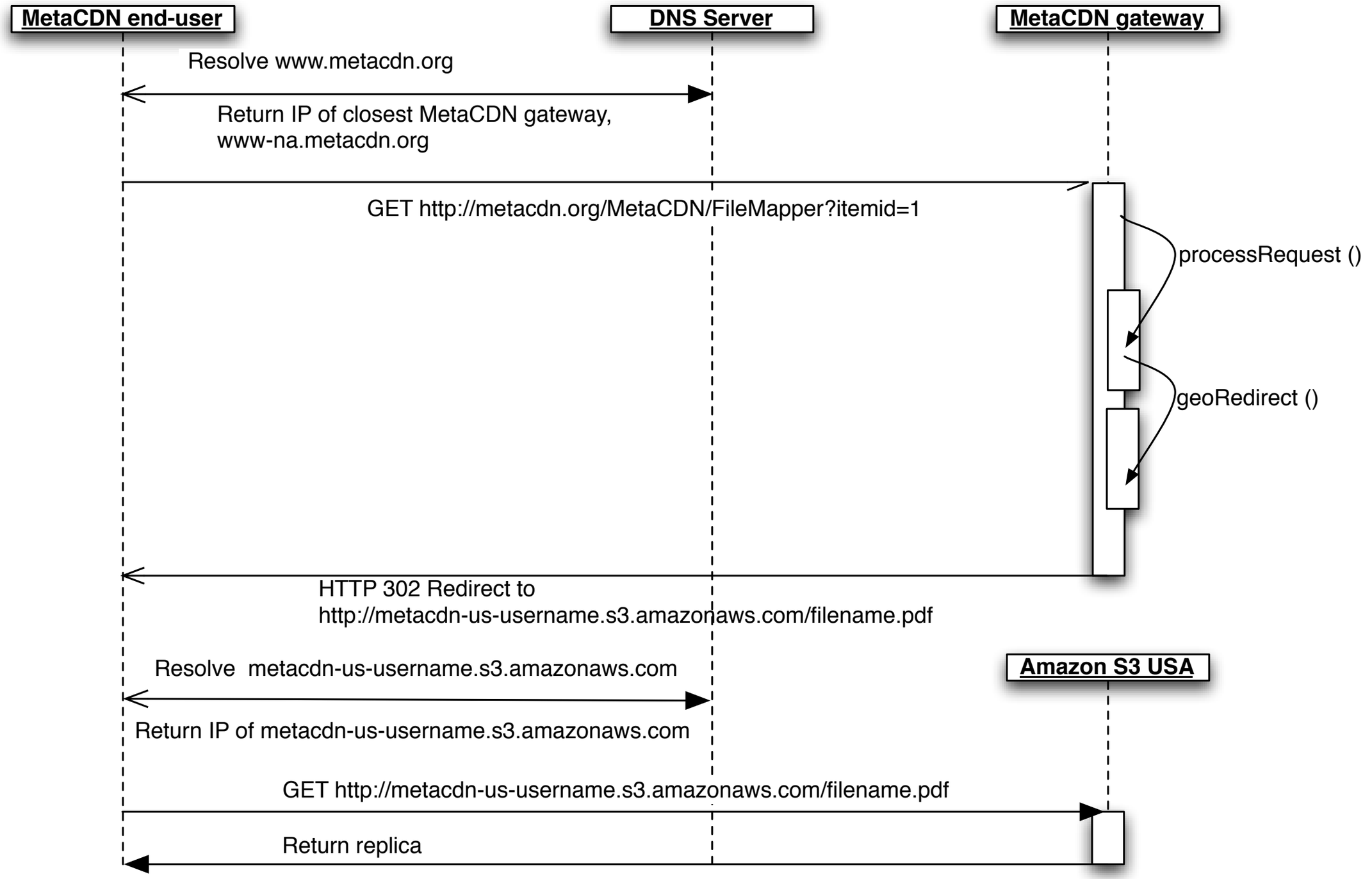
MetaCDN Load Redirector

- We have created a unified namespace to simply deployment, routing, management
- Currently, file deployment results in multiple URLs, each mapping to a replica
  - http://metacdn-eu-user.s3.amazonaws.com/myfile.mp4
  - http://metacdn-us-user.s3.amazonaws.com/myfile.mp4
  - http://node3.nirvanix.com/MetaCDN/user/myfile.mp4
- Single namespace is created for fine control
  - http://www.metacdn.org/FileMapper?itemid=2

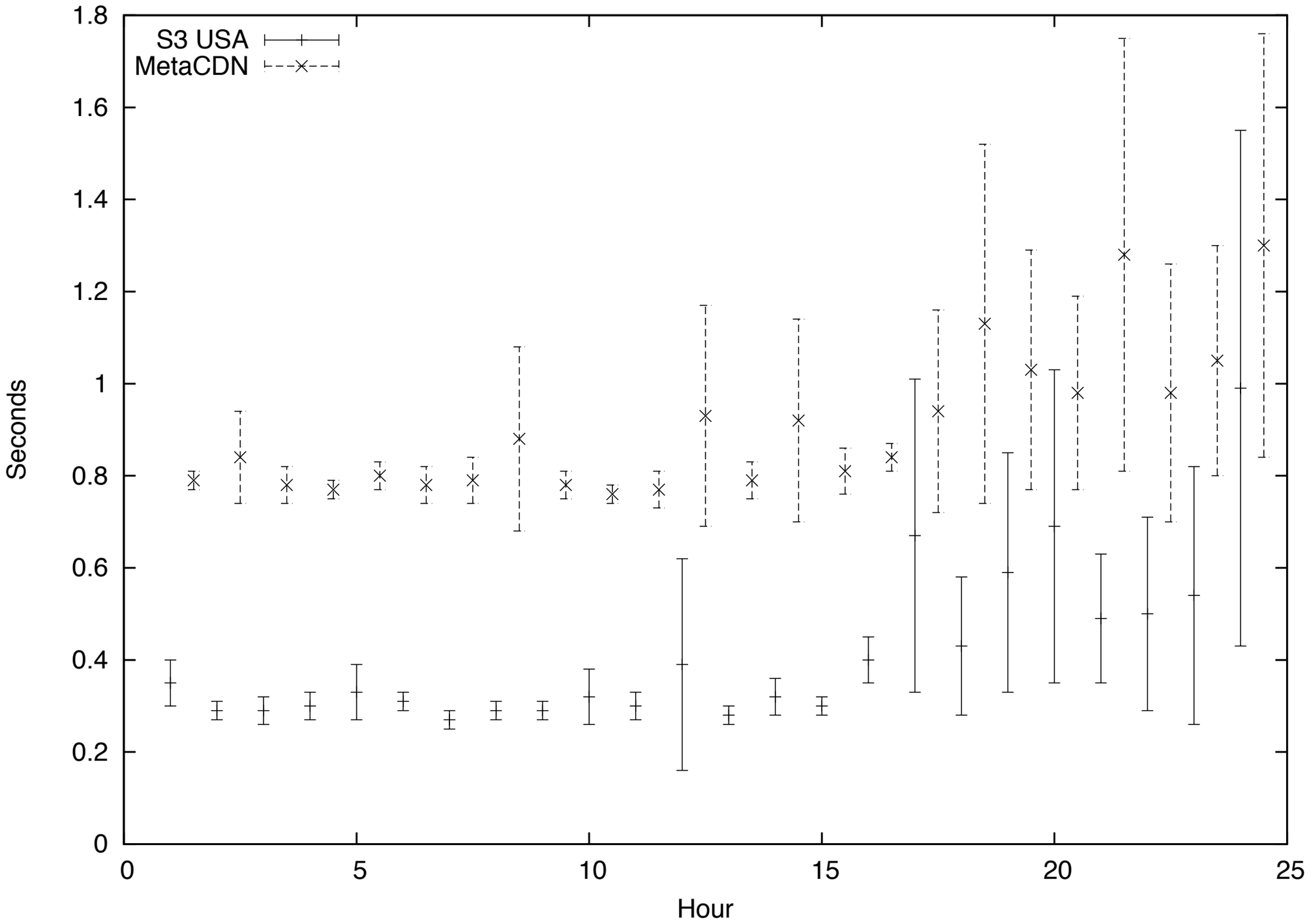
MetaCDN Load Redirector (cont.)

- Actual load redirection logic depends on deployment option used by MetaCDN user:
  - Maximise coverage and performance? - Find closest physical replica
  - Deploy content in specific locations? - Find closest physical replica
  - Cost optimised deployment? - Find cheapest replica, minimises cost to maximise lifetime
  - Quality of Service (QoS) optimised? - Find best (historically) performing replica

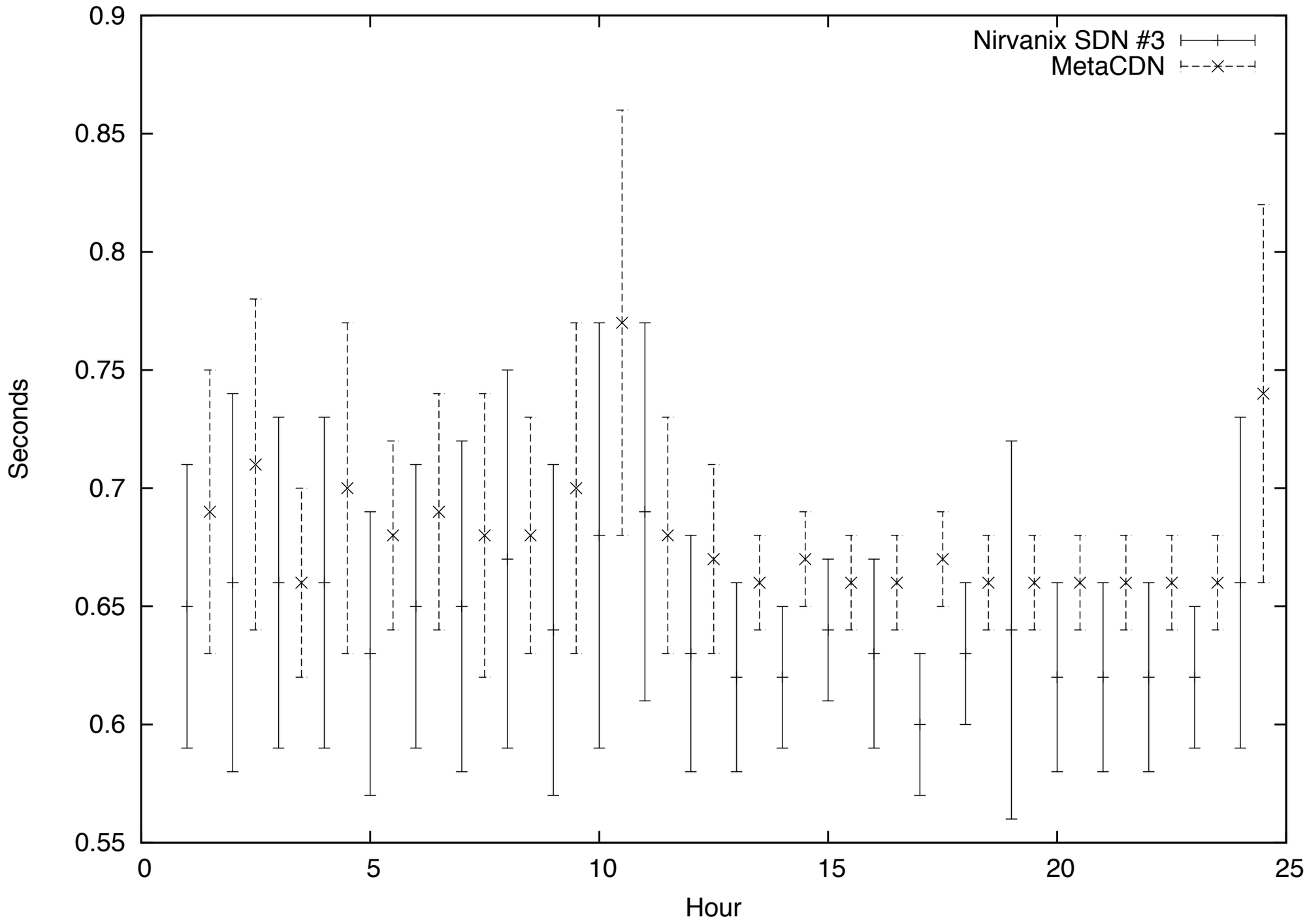
MetaCDN Load Redirector (cont.)



Redirector Overhead from USA



Redirector Overhead from Australia



Evaluating MetaCDN performance

- Ran tests over 24 hour period (mid-week)
- Downloading test replicas (1KB, 10MB) 30 times per hour, take average and conf. inter.
  - 10MB - Throughput, 1KB - Response Time
- Ran test from 6 locations: Melbourne (AUS), Paris (FRA), Vienna (AUT), San Diego & New Jersey (USA), Seoul (KOR)
- Replicas located across US, EU, ASIA, AUS

Summary of Results - Throughput (KB/s)

	S3 US	S3 EU	SDN #1	SDN #2	SDN #3	SDN #4	Coral
Melbourne Australia	264.3	389.1	30	366.8	408.4	405.5	173.7
Paris France	703.1	2116	483.8	2948	416.8	1042	530.2
Vienna Austria	490.7	1347	288.4	2271	211	538.7	453.4
Seoul South Korea	312.8	376.1	466.5	411.8	2456	588.2	152
San Diego USA	1234	323.5	5946	380.1	506.1	820.4	338.5
New Jersey USA	2381	1949	860.8	967.1	572.8	4230	636.4

Summary of results - Response Time (Sec)

	S3 US	S3 EU	SDN #1	SDN #2	SDN #3	SDN #4	Coral
Melbourne Australia	1.378	1.458	0.663	0.703	1.195	0.816	5.452
Paris France	0.533	0.2	0.538	0.099	1.078	0.316	3.11
Vienna Austria	0.723	0.442	0.585	0.099	1.088	0.406	3.171
Seoul South Korea	1.135	1.21	0.856	0.896	1	0.848	3.318
San Diego USA	0.232	0.455	0.23	0.361	0.775	0.319	4.655
New Jersey USA	0.532	0.491	0.621	0.475	1.263	0.516	1.916

Summary of results (cont)

- Clients benefited greatly from local replicas
- Results are consistent in terms of response time and throughput with previous studies of dedicated CDNs
- Back-end cloud providers have sufficient performance and reliability to be used to host replicas in the MetaCDN system
- Adding "CDN" Cloud Providers (Amazon CloudFront, Mosso Cloud Files) will likely significantly improve results.

How can we make MetaCDN scale?

- Already grown from 1 gateway to multiple gateways:
  - 1 gateway in Melbourne, Australia (physical)
  - 1 gateway in Seattle, WA (EC2 "USA-East")
  - 1 gateway in Dublin, IRE (EC2, "EU-West")
- Deploy more gateways on-demand as required:
  - Based on locality of requests, add gateways
  - As demand decreases, remove gateways.

How do we make MetaCDN scale?

- Use DNS-based redirection to closest gateway.
  - Dynamically update A-DNS entries using Web Service API.
- Fill Cloud black holes with seamlessly integrated non-“Cloud” storage:
  - Add shared / private host support.
  - FTP, WebDAV, SCP accessible.

MetaCDN features in planning / development

- Support as many providers as possible
  - Windows Azure Storage Service support will be finished very shortly.
- Integrated Flash video and MP3 audio streaming with embeddable players for customers to place on their origin sites.

MetaCDN features in planning / development

- Autonomic deployment management (expansion/contraction) based on demand
- Autonomic deployment management (expansion/contraction) based on QoS
- Security / ACL framework that spans all cloud storage providers
- “One time” MetaCDN URLs

Collaborations

- Always looking for people to collaborate on the project
- Work to be done on:
  - Load balancing / redirection algorithms
  - Intelligent Caching / replication algorithms
  - Security / Access Control of content
  - Improving MetaCDN Web Services
- Please contact me if you are interested...

Publications

- J. Broberg and Z. Tari. MetaCDN: Harnessing Storage Clouds for High Performance Content Delivery. In Proceedings of The Sixth International Conference on Service-Oriented Computing [Demonstration Paper] (ICSOC 2008), LNCS 5364, pp. 730–731, 2008.

- J. Broberg, R. Buyya and Z. Tari. Creating a 'Cloud Storage' Mashup for High Performance, Low Cost Content Delivery, Second International Workshop on Web APIs and Services Mashups (Mashups'08), In Proceedings of The Sixth International Conference on Service-Oriented Computing Workshops, LNCS 5472, pp. 178–183, 2009.

- J. Broberg, R. Buyya, and Z. Tari, MetaCDN: Harnessing 'Storage Clouds' for high performance content delivery, Journal of Network and Computer Applications (JNCA), To appear, 2009

- R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility, Future Generation Computer Systems. To appear, 2009.

Open Research Issues in Cloud Computing

Research Issues

- Interoperability
- Portability
- Reliability & Monitoring
- Data Security and Jurisdiction
- Innovative Cloud Pricing Models
- Innovative Service Models

Interoperability

- Many ongoing "open" standards efforts:
  - Cloud Computing Interoperability Forum
  - Open Grid Forum (OGF) Open Cloud Computing Interface Working Group
  - DMTF Open Cloud Standards Incubator
  - GoGrid API (CC licensed)
  - Sun Cloud API (CC licensed)
  - Amazon Web Services as "de-facto"

Portability

- Amazon Machine Image (AMI)
- VMWare Virtual Machine Disk Format (VMDK)
- Citrix XenServer / Microsoft Hyper-V Virtual Hard Disk (VHD)
- Open Source Xen Virtual Block Device (VBD) - File or LVM based

Portability - Solutions?

- Virtual to Virtual (V2V):
  - VMware vCenter Converter
  - Citric XenConvert
- DMTF Virtualization Management Initiative
  - Open Virtualization Format (OVF)
  - Initial support in XenServer / VMWare
    - Project Kensho (Xenserver/Citrix)
    - OVF Tool Technology Preview (VMWare)

Reliability & Monitoring

- Outages happen!
  - Amazon AWS: 8hrs (7/08), 2hrs (2/08), 48hrs (10/07)
  - Google GMail - 2hrs (2/09)
  - Google - 1.5hrs (5/09) - "An error in one of our systems caused us to direct some of our web traffic through Asia."

Reliability & Monitoring

- Full transparency is needed:
  - Cloudstatus.com (Hyperic)
  - http://trust.salesforce.com/ (Salesforce.com)
  - http://status.aws.amazon.com/ (Amazon)

CloudStatus BETA

powered by HYPERIC

Learn More about CloudStatus

- Amazon Service Summary
- Elastic Compute Cloud (EC2)
- Simple Storage Service (S3)
- Simple Queue Service (SQS)
- Simple DB (SDB)
- Flexible Payment Service (FPS)

Sign-up for CloudStatus Updates.

Updated 17:10:06. Updates in 44

Amazon Web Services Health Summary

These charts display real-time health status and the last **twenty four** hours of health history for key Amazon Web Services. Click a Service in the left panel for detailed service health status, metrics, and more history. More information on CloudStatus.

Healthy Service Issues Service Failure

EC2 — CURRENT

7/20/08 12:09 AM — 7/21/08 12:09 AM

- Hyperic: EC2 deployment latency affected by SQS outage. Excessive amounts of EC2 zombies created @ 7/20/08 8:46 AM

S3 — CURRENT

7/20/08 12:09 AM — 7/21/08 12:09 AM

- Hyperic: S3 reports numerous internal server errors @ 7/20/08 8:46 AM

SQS — CURRENT

7/20/08 12:09 AM — 7/21/08 12:09 AM

- Hyperic: SQS reports numerous internal server errors @ 7/20/08 8:46 AM

SDB — CURRENT

7/20/08 12:09 AM — 7/21/08 12:09 AM

FPS — CURRENT

7/20/08 12:09 AM — 7/21/08 12:09 AM

trust.salesforce.com

salesforce.com

Home | Security | System Status | Customer Login

View Upcoming Maintenance Schedule

Wednesday May 6, 2009 | 9:02 am PDT

Service System	AP0 (Japan)	AP1 (APAC)	EU0 (EMEA)	NA0 (SSL)	NA1	NA2	NA3	NA4	NA5	NA6	CS0	CS1	CS2	CS3
Status	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

All Systems Operational
No issues reported.

Service Performance History †

Updated 5/6/2009 9:01 am PDT

			System Status													
Date	**Number of Transactions**	**Avg. Speed*** (seconds)	**AP0 (Japan)**	**AP1 (APAC)**	**EU0 (EMEA)**	**NA0 (SSL)**	**NA1**	**NA2**	**NA3**	**NA4**	**NA5**	**NA6**	**CS0**	**CS1**	**CS2**	**CS3**
05/06/09	59,249,466	0.342	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
05/05/09	192,641,081	0.326	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
05/04/09	178,749,530	0.327	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
05/03/09	66,264,679	0.251	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
05/02/09	70,840,532	0.260	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
05/01/09	165,414,591	0.311	✓	✓	✓	⚠	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/30/09	194,557,330	0.328	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/29/09	191,071,439	0.303	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/28/09	198,498,961	0.323	✓	✓	⚠	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/27/09	194,860,414	0.321	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/26/09	66,081,408	0.227	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/25/09	70,805,190	0.231	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
04/24/09	182,156,924	0.285	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/23/09	191,111,300	0.282	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/22/09	190,279,196	0.334	✓	✓	✓	✓	✓	✓	✓	⚠	✓	✓	✓	✓	✓	✓
04/21/09	192,991,907	0.334	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚠
04/20/09	189,647,185	0.331	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/19/09	61,745,868	0.220	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/18/09	68,755,196	0.252	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/17/09	177,727,343	0.323	✓	✓	✓	✓	✓	✓	✓	⚠	⚠	✓	✓	✓	✓	✓
04/16/09	188,910,455	0.310	✓	✓	✓	✓	✓	✓	✗	✓	⚠	✓	✓	✓	✓	✓
04/15/09	191,104,119	0.340	✓	✓	✓	✓	✓	✓	✓	✓i	✓	✓	✓	✓	✓	✓
04/14/09	190,953,644	0.353	✓	✓	✓	✓	✓	✓	✓	⚠	✓	✓	✓	✓	✓	✓
04/13/09	166,227,451	0.319	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/12/09	63,256,822	0.237	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/11/09	69,538,950	0.242	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/10/09	139,355,992	0.284	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
04/09/09	179,285,342	0.297	✓	✓	✓	✓	⚠	✓	✓	✓	✓	✓	✓	✓	✓	✓

✓ Instance available ⚠ Performance issues ✗ Service disruption i Informational message ? Status not available

† Service Performance History is updated four times daily. Today's date may reflect partial day's data.
* Salesforce.com service transaction speed. Your on-site page load times may vary based on your bandwidth and network latency.

May 6, 2009 Report an Issue

Current Status	Details	RSS
Amazon CloudFront	Service is operating normally.	
Amazon Elastic Compute Cloud (EU)	Service is operating normally.	
Amazon Elastic Compute Cloud (US)	Service is operating normally.	
Amazon Elastic MapReduce	Service is operating normally.	
Amazon Flexible Payments Service	Service is operating normally.	
Amazon Mechanical Turk (Requester)	Service is operating normally.	
Amazon Mechanical Turk (Worker)	Service is operating normally.	
Amazon Simple Queue Service	Service is operating normally.	
Amazon Simple Storage Service (EU)	Service is operating normally.	
Amazon Simple Storage Service (US)	Service is operating normally.	
Amazon SimpleDB	Service is operating normally.	
AWS Management Console	Service is operating normally.	

Service is operating normally Performance issues Service disruption Informational message

May 13, 2009

Report an Issue

Current Status		Details	RSS
✓	Amazon CloudFront	Service is operating normally.	
✓	Amazon Elastic Compute Cloud (EU)	Service is operating normally.	
✓	Amazon Elastic Compute Cloud (US)	Elevated API Error Rates less ⌃	

8:14 PM PDT We are currently investigating increased API error rates for the DescribeVolumes and DescribeSnapshots calls in the US-EAST-1 region.
8:43 PM PDT We are currently investigating high API error rates for the EBS API method calls (CreateVolume, DescribeVolumes, CreateSnapshot, DescribeSnapshots, DeleteSnapshot) in the US-EAST-1 region.
9:22 PM PDT We are continuing to investigate elevated EBS API error rates in the US-EAST-1 Region.
10:06 PM PDT EC2 EBS API error rates have recovered. We are investigating potential impact to existing EBS volumes.
11:08 PM PDT EC2 EBS is currently capacity constrained in a single Availability Zone in the US-EAST-1 Region.

✓	Amazon Elastic MapReduce	Service is operating normally.	
✓	Amazon Flexible Payments Service	Service is operating normally.	
✓	Amazon Mechanical Turk (Requester)	Service is operating normally.	
✓	Amazon Mechanical Turk (Worker)	Service is operating normally.	
✓	Amazon Simple Queue Service	Service is operating normally.	
✓	Amazon Simple Storage Service (EU)	Service is operating normally.	
✓	Amazon Simple Storage Service (US)	Service is operating normally.	

Status History

Amazon Web Services keeps a running log of all service interruptions that we publish in the table below for the previous 35 days. Mouse over any of the status icons below to see a detailed incident report (click on the icon to persist the popup). Click on the arrow buttons at the top of the table to move forward and backwards through the calendar.

	Apr 21	Apr 20	Apr 19	Apr 18	Apr 17	Apr 16	Apr 15	
Amazon CloudFront	✓	✓	✓	✓	✓	✓	✓	
Amazon Elastic Compute Cloud (EU)	✓	✓	✓	✓	✓	✓	✓	
Amazon Elastic Compute Cloud (US)	✓	✓	✓	✓	✓	✓	✓	
Amazon Elastic MapReduce	✓	✓	✓	✓	✓	✓	✓	
Amazon Flexible Payments Service	✓	✓	✓	✓	✓	✓	✓	
Amazon Mechanical Turk (Requester)	✓	✓	✓	✓	✓	✓	✓	
Amazon Mechanical Turk (Worker)	✓	✓	✓	✓	✓	✓	✓	
Amazon Simple Queue Service	✓	✓	✓	✓	✓	⛔	✓	
Amazon Simple Storage Service (EU)	✓	✓	✓	✓	✓	✓	✓	
Amazon Simple Storage Service (US)	✓	✓	✓	✓	✓	✓	✓	
Amazon SimpleDB	✓	✓	✓	✓	✓	✓	✓	
AWS Management Console	✓	✓	✓	✓	✓	✓	✓	

Google Apps Status Dashboard

This page offers performance information for Google Apps services. Unless otherwise noted, this status information applies to consumer services as well as services for organizations using Google Apps.

Check back here any time to view the current status of the services listed below. For all other information or to report a problem, please visit the Google Apps Help Centers.

On May 14, some users experienced a network issue that affected multiple Google services for about an hour. We're sorry for the inconvenience, and we appreciate your patience as we resolved this issue. For more information, please see our blog post.

	Today's Status	5/14/09	5/13/09	5/12/09	5/11/09	5/10/09	5/9/09
Google Mail	Information available	ℹ	✓	✓	✓	✓	✓
Google Calendar	No Issues	ℹ	✓	✓	✓	✓	ℹ
Google Talk	No Issues	✓	✓	✓	✓	✓	✓
Google Docs List	No Issues	✓	✓	✓	✓	✓	✓
Google Documents	No Issues	✓	✓	✓	✓	✓	✓
Google Spreadsheets	No Issues	✓	✓	✓	✓	✓	✓
Google Presentations	No Issues	✓	✓	✓	✓	✓	✓
Google Sites	No Issues	ℹ	✓	✓	✓	✓	✓
Google Video for business	No Issues	✓	✓	✓	✓	✓	✓
Admin control panel	No Issues	✓	✓	✓	✓	✓	✓
							Older »

The following symbols signify the most severe issue (if any) encountered during that day. Click a symbol in the table above to view detailed information.

RSS Feed

✓ No issues Service disruption ✖ Service outage ℹ Information available

Why is monitoring important?

- “To be eligible, the credit request must (i) include your account number in the subject of the e-mail message [..]; (ii) include, in the body of the e-mail, the dates and times of each incident of non-zero Error Rates that you claim to have experienced; (iii) include your server request logs that document the errors and corroborate your claimed outage [..]; and (iv) be received by us within ten (10) business days after the end of the billing cycle in which the errors occurred.” - SLA for S3

Why is monitoring important?

- "To receive a credit, you must contact your The Rackspace Cloud account team within thirty (30) days following the end of the downtime. You must show that your use of the Cloud Servers was adversely affected in some way as a result of the downtime to be eligible for the credit." - SLA for Mosso Cloud Servers

Why is monitoring important?

- “Customer must open a support case (a "Case") during the Failure in question. Customer will open all Cases through the Customer Portal. In opening a Case, Customer will provide complete information regarding the nature of the problem, including any information reasonably necessary for diagnosis and correction, by following the Case opening procedures at the Customer Portal. [...]” - SLA for GoGrid

Why is monitoring important?

- Service credits are not automatic!
- Requires immediate user action:
  - Claim within certain time period
  - Note specific dates and times of outage
  - Provide proof (i.e. logs) of outage
- Automated monitoring & alert services are invaluable in this environment.

Data Security and Jurisdiction in the Cloud

- Cloud Servers abstract where processing occurs and where your data is stored.
- Software/data are subject to laws & regs. of where they are executed & stored.
- Consider:
  - Digital Millennium Copyright Act (DMCA)
  - Cryptography Export laws
  - Other regs (SEC, Sarbanes-Oxley, HIPPA)

Innovative Cloud Pricing Models

- Pricing for Cloud services is very static
- Lets take some cues from traditional utilities:
  - Peak / Off Peak pricing.
  - Spreads demand, making capacity planning more predictable for providers.
  - Flexible Cloud Consumers can save money.

Innovative Service Models for Clouds.

- Create framework for Cloud Consumers to be Cloud Providers.
- Again, cues from utilities like power:
  - End-users & businesses can sell back electricity generated by solar energy to grid
  - Enable sell back of under-utilised resources using same tech. cloud providers' use!
- Win-win situation as companies could resell these resources and expand their footprint.

Introduction to Cloud Computing

- What we covered:
  - Part I - An Introduction to Cloud Computing.
  - Part II - "Get your head in the Clouds" - MetaCDN Cloud Case Study.
  - Part III - Open Research Issues in Cloud Computing.

Thanks

Any questions?