

Introducing Amazon S3 and EC2



Justin Mason

<http://jmason.org/>
<http://taint.org/>

Introducing Amazon S3 and EC2

- Two new web services from Amazon Web Services (AWS)
 - there are others, but not talking about them
 - These two are pretty revolutionary
- Not an Amazon employee – just think it's way cool
 - following virtualisation for a few years
 - EC2 is a major virtualisation win
 - turning out handy for my SpamAssassin work

S3: Simple Storage Service

- a hard disk in the cloud
 - also, a web server, if you set the files to be visible
- essentially infinite -- limited by your wallet ;)
- 99.99% availability; no single points of failure
- great parallel scalability
- all files offered as BitTorrent, too



EC2: Elastic Compute Cloud

- "Hardware As A Service"
- create Linux "servers" on the fly
 - really Xen virtual machine instances running on AMD x86; each instance has 2GB RAM and 150GB disk
- create/destroy from the command line
- very competitive with "real" hosting

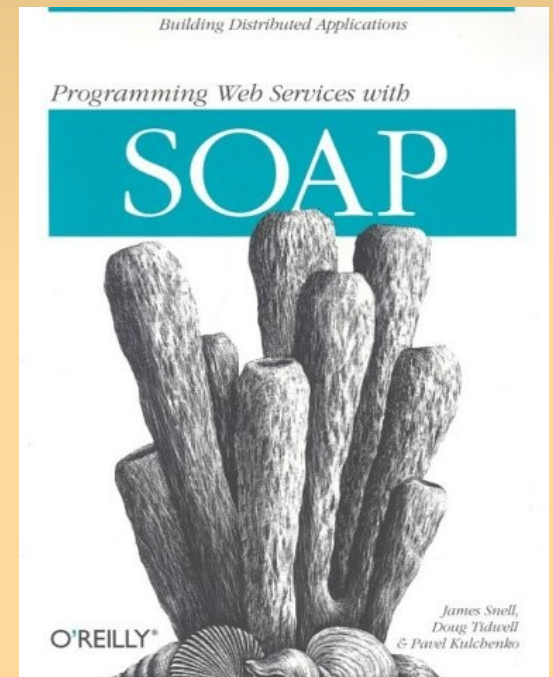


Pricing

- S3 is really quite cheap; \$0.20 per GB of data transferred, plus \$0.15 per GB-Month of storage used
 - (That's a good price for bandwidth, as far as I know)
 - There are better deals around, but this has other features...
- EC2: a bit pricier: \$0.10 per instance-hour used
 - plus \$0.20 per GB of data transferred outside Amazon; but traffic to/from S3 is *free*

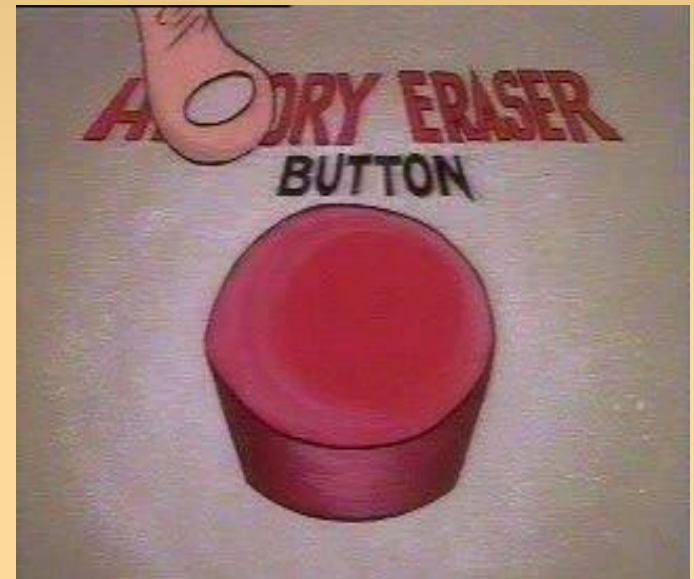
Usability for Developers

- super-easy -- just give them address and credit card number
 - S3 immediately usable
 - EC2 has a beta program with a waiting list :(
- SOAP and REST APIs -- very usable and easy to hack with
- Billing in small increments, no big upfront charges or monthly fees (“paid by the drink”)
- all done via the web



Reliability

- no need to:
 - worry about RAID, hardware
 - visit the data centre to hit the big red button
 - pay for data centres, full stop!
- S3 is in production use with Amazon's products
- S3-hosted data has one copy in at least 2 data centres (apparently)



S3 Gotchas

- it appears that their hosting location diversity is not great
 - diverse across the US, but apparently not further, e.g. Asia
 - not a replacement for a full CDN like Cachefly or Akamai
- only serves static content via HTTP
- reportedly "*extended and unannounced periods of downtime*", according to one unhappy user

S3 Gotchas (contd.)

- Quite hacky to use directly as a network filesystem
 - (OpenFount S3InfiDisk -- free-as-in-beer product)
 - doesn't have real POSIX semantics, anyway
- no atomic filesystem semantics
 - however, there is an interesting “rename” hack using the md5sum metadata
- no rsync support
 - (although s3sync is close)

EC2 Gotchas

- *Big* queue to get on the beta program
 - took 1.5 months for my account to come through
- Not very cheap for low-end users; \$0.10 per instance-hour adds up quickly
 - (\$67 per month, per running instance)
- billed by the clock-hour, not the CPU-hour
 - so a 20%-utilised server costs the same as a 100%-busy one

EC2 Gotchas (contd.)

- Local storage is non-persistent
 - When you shut down, your data is lost
 - Need to write it elsewhere; but S3 is free!
 - You can "freeze" a running instance's "disks" to S3, as an "AMI" (Amazon Machine Instance), then duplicate that to as many servers as you like

EC2 Gotchas (contd. 2)

- DHCP IP address assignment
 - So IP changes when instance reboots
 - Hard to use as a public server
 - HTTP is still usable with a reverse proxy, such as Pound or Apache's mod_proxy
- Also geographically non-distributed
 - East-coast US datacenters

Things To Do With EC2

- on-demand gaming servers
 - turn 'em off when you're finished!
- on-demand spam-filtering backend servers, using spamd
 - handle spam load spikes
- other kinds of on-demand backend, to handle spikes
 - easy to horizontally scale with EC2



URLs

- <http://www.amazonaws.com/>
 - The Amazon site for both services
- <http://del.icio.us/jm/aws>
 - My bookmarks on the topic
- <http://taint.org/tag/aws>
 - My plans for an EC2-hosted SpamAssassin backend
- **Questions?**