Amazon Web Services: Building a Web-Scale Computing Infrastructure

Jeff Barr Senior Web Services Evangelist





Survey Says...



- Are you an Amazon retail customer?
- Have you heard of the Amazon Web Services?
- Have you used them?
- Which ones (Associates, EC2, S3, SQS)?
- i Have you seen me speak before?
- Could you give this talk?

Things to Remember



AWS Site: aws.amazon.com

AWS Blog: aws.typepad.com

My Email: jbarr@amazon.com

Who am I?

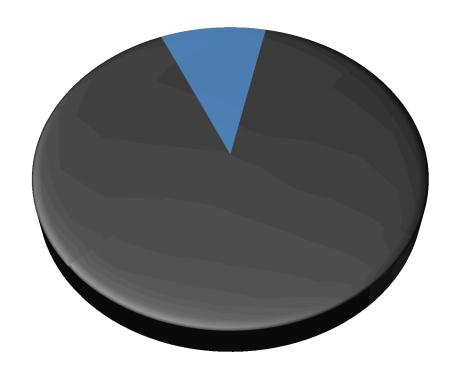


- Philly native!
- Software development background
- Farly RSS: Headline Viewer & Syndic8.com
- Programmable applications and web sites
- Microsoft Visual Basic and .Net Teams
- Startup / venture consultant
- Over 5.5 Years with Amazon:
 - Senior Developer
 - Senior Web Services Evangelist
 - World traveler Wiki powered



Today's Talk....





- Technical
- Business
- Customer
- Code

Amazon Web Services



- Application building blocks
- Stable APIs
- Proven Amazon infrastructure
- Focus on innovation and creativity
- Long-term investment

Issues Facing Developers



- 70% of Web Development Effort is:
 - Data Centers
 - Bandwidth / Power / Cooling
 - Operations
 - Staffing



- Scaling is Difficult and Expensive:

 - Invest Ahead of Demand
 - Load is Unpredictable



Dream or Nightmare?



- Rapid, unexpected customer demand/growth
- Unpredictable load
- Slashdot/Digg/TechCrunch
- "Success disaster"
- Seasonal spikes



Reality: Cloud Computing



- Mature Amazon infrastructure
- Utility pricing
- Scale capacity on demand
- Turn fixed costs into variable costs
- Clean APIs
- Simple conceptual models
- Always available
- Cost-effective



Utility Computing Services Menu



- Amazon Simple Queue Service
- Amazon Simple Storage Service
- Amazon Elastic Compute Cloud
- Amazon Flexible Payments Service
- Amazon SimpleDB
- Amazon DevPay





Amazon Simple Storage Service

S3

Amazon Simple Storage Service



- Object–Based Storage
- 1 B 5 GB / object
- Fast, Reliable, Scalable
- Redundant, Dispersed
- 99.99% Availability Goal
- Private or Public
- Per-object URLs & ACLs
- US & European Locations

Amazon Simple Storage Service



- Object-Based Storage
- 1 B 5 GB / object
- Fast, Reliable, Scalable
- Redundant, Dispersed
- 99.99% Availability Goal
- Private or Public
- Per-object URLs & ACLs
- US & European Locations

\$.15 per GB per month storage

\$.01 for 1000 to 10000 requests

\$.10 - \$.18 per GB data transfer

Amazon S3 Concepts



Objects:

- Opaque data to be stored (1 byte ... 5 Gigabytes)
- Authentication and access controls

Buckets:

- Object container any number of objects

Keys:

- Unique object identifier within bucket
- Flat object storage model

Standards-Based Interfaces:

- REST and SOAP
- URL-Addressability every object has a URL
- BitTorrent seed



S3 API



Service:

■ ListAllMyBuckets

Buckets:

- CreateBucket
- DeleteBucket
- ListBucket

Objects:

- PutObject
- PutObjectInline
- GetObject
- GetObjectExtended
- DeleteObject
- GetObjectAccessControlPolicy
- SetObjectAccessControlPolicy



S3.RB - Establish Connection



```
require 'S3'
AWS ACCESS KEY = '<your key>'
AWS SECRET ACCESS KEY = '<your key>'
conn = S3::AWSAuthConnection.new
(AWS ACCESS KEY ID,
AWS SECRET ACCESS KEY,
false)
```

S3.RB - Create Bucket



```
BUCKET_NAME = 'assets'
conn.create_bucket(BUCKET_NAME)
```

S3.RB – Upload File



```
datafile = File.open(path)
Key = path.basename;
conn.put(BUCKET NAME, key, datafile.read,
{"Content-Type" => mime,
 "Content-Length" => File.size(path).to s,
 "x-amz-acl" => "public-read" } )
http://assets.sample.com/...
```

Common Amazon S3 Use Cases



- Media Sharing
 - NRK, Smugmug, Blingee, Ringo
- mailtrust^{*}



- Media/Software Distribution
 - Microsoft, Liberated Syndication, Linden Lab





- Backup (Server and PC)
 - Mailtrust
- Online Storage
 - MyDataBus, SendAlong
- Application Storage
 - 37 Signals, Xerox Global Services



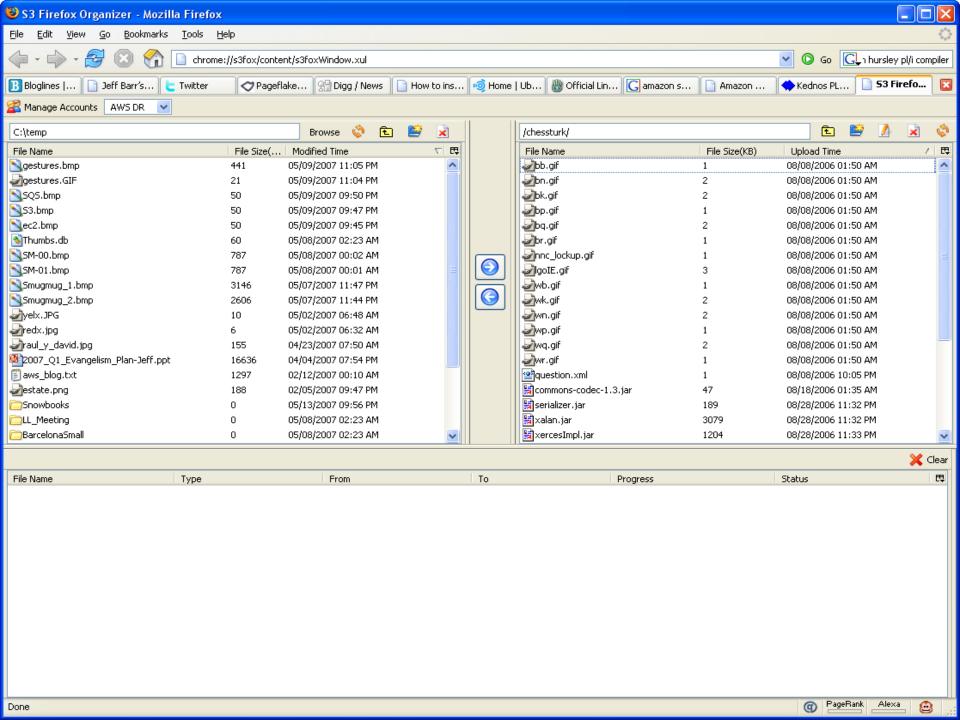














Amazon Elastic Compute Cloud

EC2



Amazon Elastic Compute Cloud



- Virtual Compute Cloud
- Root-level System

Access

- Elastic Capacity
- Management API
- Scale in Minutes
- Multiple Instance Sizes
- Network Security Model

\$.10-\$.80 per server hour

Amazon Elastic Compute Cloud



- Virtual Compute Cloud
- Root-level System

Access

- Elastic Capacity
- Management API
- Scale in Minutes
- Multiple Instance Sizes
- Network Security Model

\$.10-\$.80 per server hour

\$.10 - \$.18 per GB data transfer

EC2 Instance Types



S

1.7 GB RAM
1 EC2 Compute
Unit
160 GB
32-bit platform

L

7.5 GB RAM
4 EC2 Compute
Units
850 GB
64-bit platform

XL

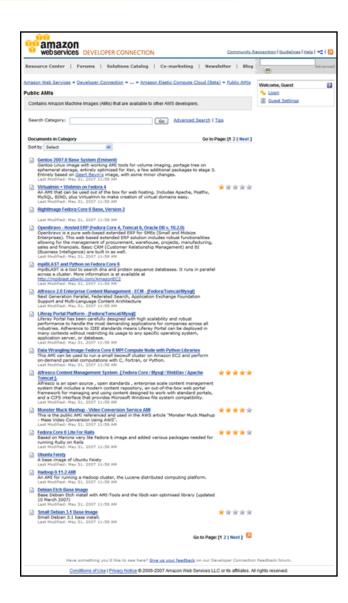
15 GB RAM 8 EC2 Compute Units 1690 GB storage, 64-bit platform

\$0.10/ Hr \$0.40/ Hr \$0.80/ Hr

Amazon EC2 Concepts



- Amazon Machine Image (AMI):
 - Bootable root disk stored in S3
 - Pre-defined or user-built
 - Catalog of user-built AMIs
 - OS: Fedora, Centos, Gentoo, Debian, Ubuntu, Windows Server
 - App Stack: LAMP, mpiBLAST, Hadoop
- Instance:
 - Running copy of an AMI
 - Launch in less than 2 minutes
 - Start/stop programmatically
- Network Security Model:
 - Explicit access control
 - Security groups
- Inter-service bandwidth is free



EC2 API



images:

- RegisterImage
- Describelmages
- DeregisterImage

instances:

- RunInstances
- DescribeInstances
- TerminateInstances
- GetConsoleOutput
- RebootInstances

Keypairs:

- CreateKeyPair
- DescribeKeyPairs
- DeleteKeyPair

Image Attributes:

- ModifyImageAttribute
- DescribelmageAttribute
- ResetImageAttribute

Security Groups:

- CreateSecurityGroup
- DescribeSecurityGroups
- DeleteSecurityGroup
- AuthorizeSecurityGroupIngress
- RevokeSecurityGroupIngress

Newest EC2 Features (Today)



i Elastic IP Addresses:

- AllocateAddress
- ReleaseAddress
- AssociateAddress
- DisassociateAddress

Availability Zones:

- DescribeAvailabilityZones
- RunInstances
- User Selectable Kernels

Amazon EC2 At Work



- Startups
 - Cruxy Media transcoding
- Science / Research:

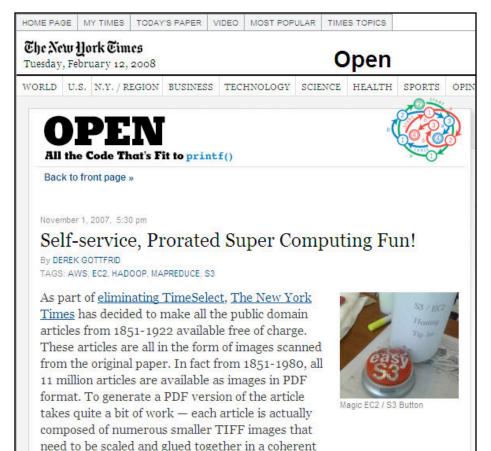
 - mpiBLAST
- Load-Management and Load Balancing Tools:
 - Rightscale
 - Scalr Scalr
- Corporate:
 - New York Times
 - High-Impact, S hort-Term Projects
 - Development /test host





"Prorated Super Computing Fun"The New York Times





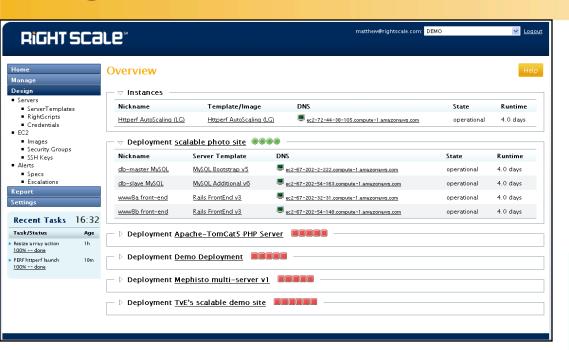
Previously we had generated all the PDFs dynamically. This approach had worked reasonably well, but with the strong possibility of a significant traffic increase we started to rethink things. Clearly, pre-generating all the articles and statically serving them would be a great option. Pretty quickly I thought about how we could do this (and have some fun along the way,

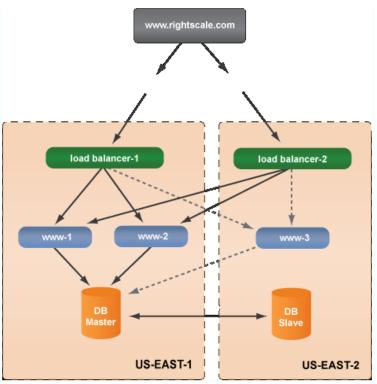
fashion.

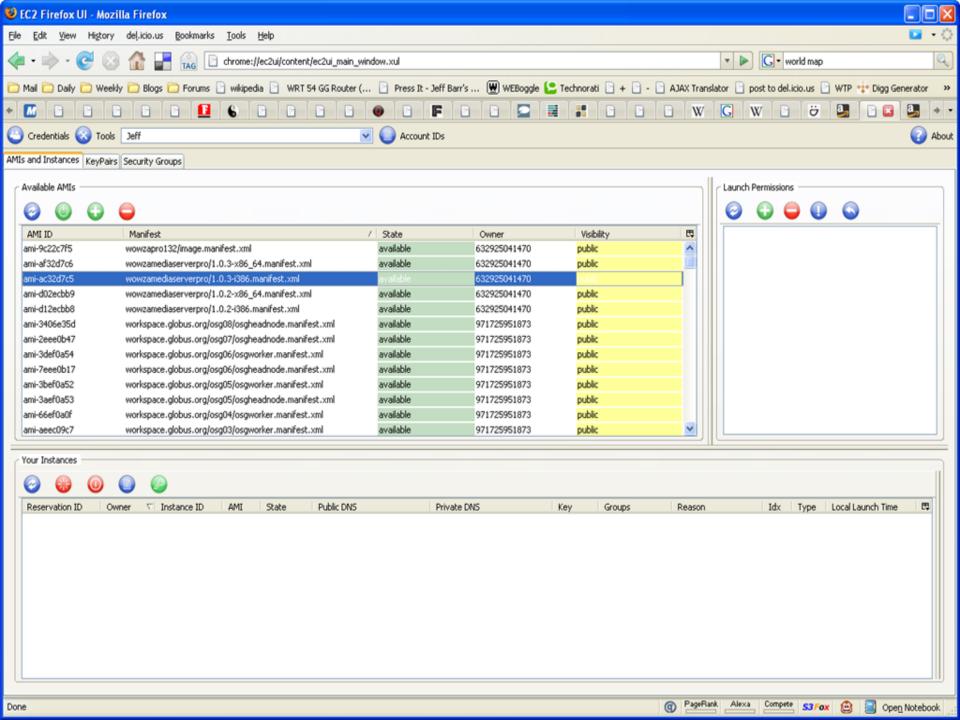
- 1851-1922 Articles
- TIFF -> PDF
- Input: 11 Million Articles (4TB of data)
- What did he do?
 - 100 EC2 Instances for 24 hours
 - All data on S3
 - Output: 1.5 TB of Data

RightScale Platform











Amazon Simple Queue Service

SQS

Amazon Simple Queue Service



- Scalable Queuing
- Elastic Capacity
- Reliable, Simple,
 Secure

Inter-process messaging, data buffering, scalable architecture component \$.01 per 10000 messages

Amazon Simple Queue Service



- Scalable Queuing
- Elastic Capacity
- Reliable, Simple,
 Secure

Inter-process messaging, data buffering, scalable architecture component \$.01 per 10000 messages

\$.10 - \$.18 per GB data transfer

SQS API



Queues:

- CreateQueue
- **ൂ** ListQueues
- DeleteQueue

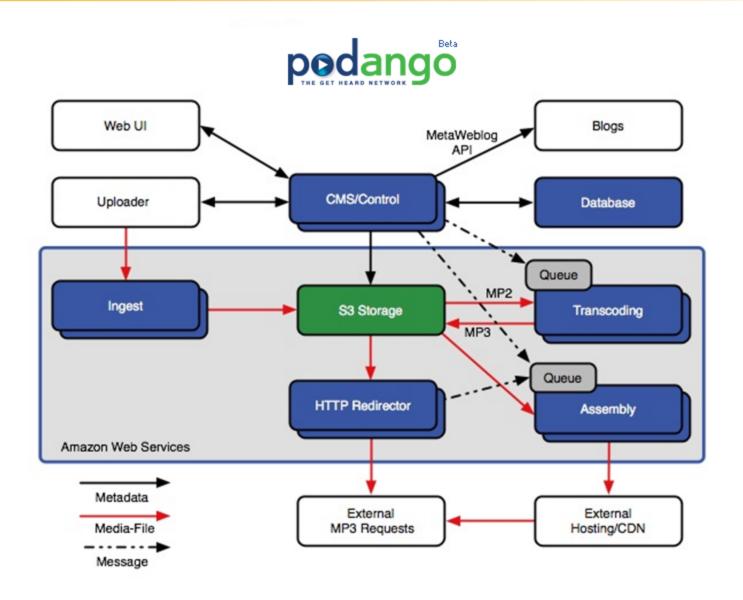
Messages:

- ReceiveMessage
- DeleteMessage



Amazon SQS At Work







Amazon SimpleDB

SDB

Amazon SimpleDB Service



- Distributed Data Store
- Structured Storage
- Fully Indexed
- Redundant
- Scalable
- Elastic Capacity
- Query Language

\$1.50 per GB per month

Amazon SimpleDB Service



- Distributed Data Store
- Structured Storage
- Fully Indexed
- Redundant
- Scalable
- Elastic Capacity
- Query Language

\$1.50 per GB per month

\$.10 - \$.18 per GB data transfer

\$.14 per CPU hour (query processing)

SimpleDB Concepts



Domain:

- Collection of similar items
- Query language
- Any number of items per domain (10 GB beta limit)

i Item:

- Collection of key-value pairs (attributes)
- Multiple values per attribute
- Up to 256 attributes per item

Billing:

- Data storage
- CPU utilization
- Data storage

itemID	description	color	material
123	Sweater	Blue, Red	
456	Dress shirt	White, Blue	
789	shoes	Black	Leather

SimpleDB API



i Domains:

- CreateDomain
- ListDomains
- DeleteDomain

items:

- PutAttributes
- **Query**
- GetAttributes

Query Language (samples):

- if Title' = 'The Right Stuff']
- **■** ['Number of Pages' < '00310']
- ('Rating' = '***' or 'Rating' = '*****']
- i ['Year' > '1950' and 'Year' < '1960' or 'Year' starts-with '193' or 'Year' = '2007']
- [Keyword' = 'Frank Miller'] union ['Rating' starts-with '****']

Other Amazon Infrastructure Services



Flexible Payments Service:

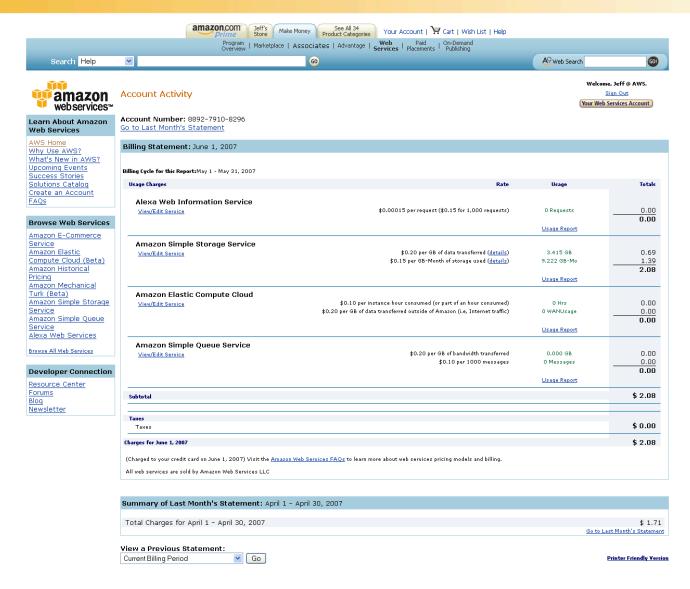
- Move money between any two people or systems.
- Credit cards, bank accounts, Amazon Payments.
- One-time, multiple, or recurring transactions.
- Payment processing language (Gatekeeper).
- Aggregated transactions (micropayments).

DevPay:

- Wrap custom business models around S3 and EC2.
- Set custom prices for each charging unit.
- The Charge for your applications.

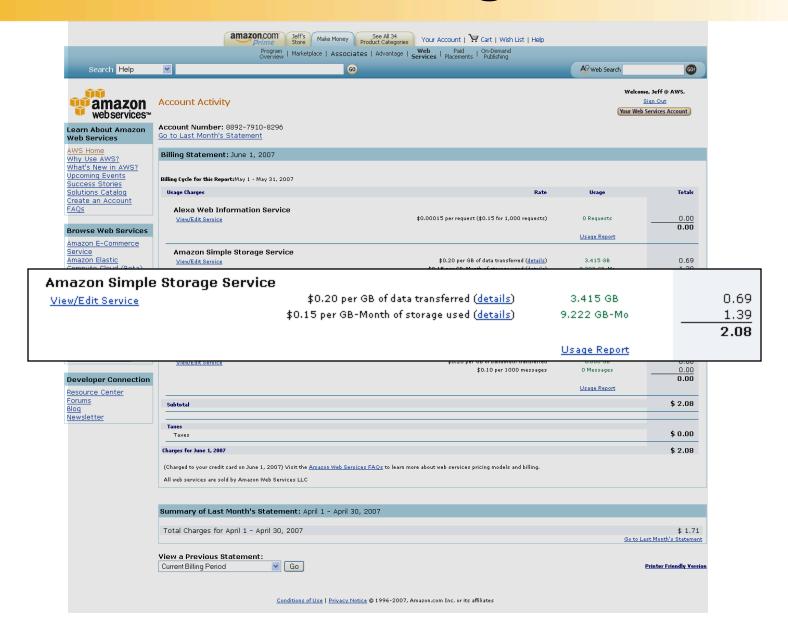
Web Services Billing





Web Services Billing

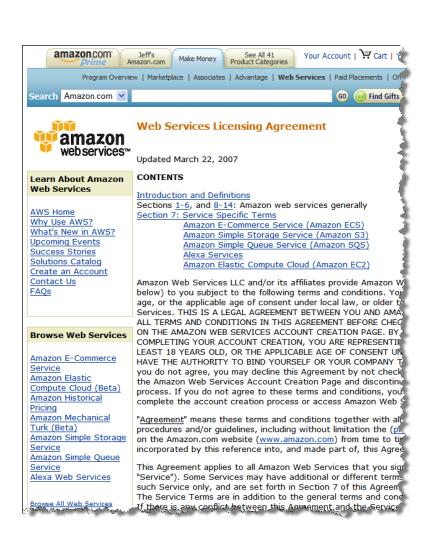




The Commercial Side



- Standard licensing terms
- Commercially usable
- Monthly credit card billing
- Self-serve model:
 - Sign up as developer
 - Choose services
 - Agree to service licenses
 - Enter payment info
 - Start coding





Q&A



Things to Remember



AWS Site: aws.amazon.com

AWS Blog: aws.typepad.com

My Email: jbarr@amazon.com