Presentation Online

- https://github.com/medale/presentations
- Slides, notes and code sample with Maven

Intro

- Markus - first used AWS for Data Science MOOC in 2013 (EMR)
- Several AWS classes later...

Amazon Web Services: AWS

- Started in 2006
  - Elastic Compute Cloud (EC2)
  - Simple Storage Service (S3)
  - (Amazon EC2 history Clark 2012)
- By 2010: Amazon.com retail web services mostly moved to AWS
- 2015: Over a Million Active Customers in 190 Countries
  - e.g. Netflix, Dropbox, Airbnb, Supercell (Clash of Clans, Hay Day, Boom Beach)

Amazon Web Services (AWS) In the News

Wired Magazine July 2015

Amazon has figured out how to make cloud pay

> *$391 million profit based on $1.82 billion in revenue. That's a 407 percent increase in profit from the year before, and an 81 percent bump in revenue.*

Gartner: Infrastructure As A Service Magic Quadrant

- X-axis: Completeness of vision
- Y-axis: Ability to execute
- Four quadrants: Niche players, visionaries, challengers, leaders
- Other providers: Rackspace, Google, Microsoft - trying to catch up

AWS Regions and Edge Points of Presence

- 12 public regions (32 availability zones), 2 US government regions
  - AWS GovCloud
- Each region: 3 or more Availability Zones ($\sim$ each AZ is 1 or more independent data centers)
  - AZs connected with private fiber in a 50m radius $<$ 1ms latency
- Points of Presence - edge nodes for content delivery network (CloudFront - cache)
- How much overall? "Every day, AWS installs enough server infrastructure to host the entire Amazon e-tailing business from back in 2004" (700 million dollar revenue - Rare Peek into Massive Scale of AWS Morgan 2014)

Multiplayer Mobile Game Application

- Start-up: Mobile device, backend server and database
- Show to investors, Minimum Viable Product

Scaling Up - Getting Beefy

- Scaling up/Scale vertically - bigger equipment (sell old or use for testing)
- Must provision for peak demand? - must scale horizontally

Scaling Out/Horizontal, Content Delivery Network & Analytics

- Scale horizontal, load balancer
- Backend to NoSQL - better scaling and no fixed schema (evolutionary architecture)
- User experience - forward deploy with CDN
- Batch and realtime analytics, datawarehouse for business analysts

Redundant Environments

- Development
- Testing
- Production
  - DevOps - Blue/Green Deployments?
  - Disaster recovery (somebody didn't call Miss Utility)

Amazon Elastic Compute Cloud (EC2)

- Must provision for peak demand/Good Morning America
- Instead of capital expense - operating expense
- Multiple regions/Availability Zones (AZs)
- EC2 - Elastic Compute Cloud
  - Elastic Load Balancer service (no single point of failure)
  - Auto-scaling group - up/down (pay only for what is used)
- Route 53 - Blue/Green deployments (DNS server - also DR)

Amazon RDS, DynamoDB, ElastiCache

- Administering database is hard - RDS (redundancy, backups, read replicas)
- DynamoDB - NoSQL Key/Value store with attributes (provisioned service)
- ElastiCache Service - Redis or Memcached (faster repsonse times)

Amazon CloudFront CDN, S3

- CloudFront CDN - serve static content from Points of Presence - read S3, cache
- Simple Storage Service - Scalable Key, object store 1 byte to 5TB per object
  - Can use S3 as input to Elastic MapReduce Hadoop job!
  - Buckets, store key/objects
  - 11 9s of durability (99.999999999) - lose two facilities
  - 4 9s of availability (99.99) - 52.56 minutes per year
  - SSL and automatic encryption
  - Storage classes: General Purpose, Infrequent Access (IA), Glacier (hours to access)

Amazon Scalable Analytics - Batch, Streaming, Datawarehouse

- Elastic MapReduce (EMR) - Hadoop, Spark, Hive, Pig, Hue, Zeppelin against S3, RedShift...
  - Priced by the hour! Linear scaling - double instances?
- Kinesis Streaming - think managed Kafka? real-time analytics, scalable ETL
- RedShift - Datawarehouse, columnar, fast, petabyte-scale (1000/TB/year)

Security At Every Layer

- Identity & Access Management (IAM) - Amazon infrastructure/services
- Users, groups, roles, rich policies/conditions
- Virtual Private Cloud (VPC) - subnets w/ Network Access Control Lists, route tables, network gateways
- EC2 - Security groups (like iptables firewall)
- Web Application Firewall
- CloudTrail - all AWS service access logged
- CloudWatch - Monitor, alarm - e.g. auto-scaling up/down

Compute As A Service

- Sys admin - hard, deploy web apps via Elastic Beanstalk
- Short running code? React to environment/change/time - Lambda
- Lambda: 1 to 5 minutes (Python, Java, Node.js/JavaScript)
- Pricing: per 100 ms
- API Gateway - invoke Lambda from web URL

52 AWS Services

- Simple Queuing Service, SNS
- Simple Workflow Service (SWF)
- CloudFormation

EC2 Instances

- On Demand
- Reserved 1-3 years, up to 75% off on demand
- Reserved Instance Marketplace
- Spot Instances - up to 90% on excess, may lose if demand goes up or outbid

Infrastructure As Code - RESTful via Management Console

- All service endpoints are REST endpoints
- GUI frontend to configure services

Infrastructure As Code - RESTful via AWS CLI

- Command line interface with Python boto backend to REST endpoints
- Script start up/shutdown, make repeatable, can version

Infrastructure As Code - RESTful via AWS SDKs

- Software Dev Kits for many languages
- Use in program to communicate with services securely
- Tonight: AWS Java SDK

Infrastructure As Code - CloudFormation JSON

- Export existing setup to different region
- Version-controlled, auditable infrastructure (no router, hardware)

AWS Marketplace

- Get instances with configured software
- Try out, pay by the hour with option to rent for year
- Preconfigured, installed, license (or bring your own)

Re:Invent 2015

- 2013: 8,000
- 2014: 13,000 attendees
- 2015: Over 20,000 attendees (TechRepublic)
- Capital One and GE as part of key note

Re:Invent 2015: Internet of Things

- From AWS IoT Tutorial right on AWS Management Console -> IoT
- Control Unit, light bulb (and mobile app)
- IoT Gateway/Message broker: X509 certs or Amazon Cognito IDs, MQTT/HTTPS
- Rules engine - transform to AWS endpoints (Kinesis Stream, S3, Lambda)
- Device shadow - keep state if device offline
- Device registry - unique ID to each thing, metadata (capabilities, attributes)

Re:Invent 2015: QuickSight

- Business Intelligence Service
- Visualizations, ad-hoc analysis
- Super-fast, Parallel, In-Memory Calculation Engine (SPICE)
- Data sources: Redshift, RDS, EMR, DynamoDB, Kinesis Streams, S3, MySQL, Oracle

Re:Invent 2015: Amazon Kinesis Firehose

- Managed service - automatically scale
- Compression, encryption to designated S3 bucket
- Run Lambda as objects arrive or EMR over bucket
- Redshift: CVS, JSON, AVRO etc. only selected columns, data type conversion

Re:Invent 2015: AWS Import/Export Snowball

- 50lbs, self-contained, tamper resistant, up to 50TB
- 100TB on 100MB/s dedicated $>$ 100 days
- AES 256-bit encryption at host, stored encrypted
- Decrypt when loaded into S3 bukcet
- Sanitize after use using NIST standard
- Send email to SNS when in bucket
- \$200 per job, plus shipping, \$15 per extra day

Re:Invent 2015: Security

- Inspector: 100s of rules mapped to common security compliance standards (e.g Payment Card Industry Data Security Standard PCI DSS), regular updates by AWS Security researchers
- WAF: Web Application Firewall - protect against SQL injection, cross-site scripting
- AWS Config Rules: cloud governance, define standards, monitor for compliance. Pre-built best practices or custom (e.g. proper tagging, use of elastic IPs etc.)
- AWS Config: All AWS resources with configurations over time

Re:Invent 2015: Other Services

- Amazon Elasticsearch Service: Document indexing (Lucene), log analytics, time series queries, works with visualization like Kibana
- Amazon Kinesis Streams Extended Retention - from 24hrs to 7 days
- AWS Database Migration Service (Preview): Oracle to Oracle, RDS etc. Data, SQL, functions, stored procedures, warn about things that cannot be auto-converted
- Amazon RDS for MariaDB - MySQL clone by creator of MySQL
- Amazon Cloudwatch Dashboards: reusable graphs of AWS CloudWatch and custom metrics
- AWS Lambda
  - Python, Versioning, Scheduled Jobs, and 5 Minute Functions

AWS SDKs - Credentials for Java

- Environment Variables: AWS_ACCESS_KEY_ID, AWS_SECRET_ACCESS_KEY
- Java System Properties: aws.accessKeyId, aws.secretKey
- Default credentials: ~/.aws/credentials
- Instance credentials (e.g. IAM role assigned to EC2 instance)

AWS Java SDK - Maven

- All inclusive or service-by-service jar

AWS Java S3 SDK

- ProfileCredProvider looks for ~/.aws/credentials
- S3: Bucket, key, object
- SLPRCELL bucket, key: damage, object: file (type=bullet target=heart - calculate damage, kevlar vest etc.)
- s3 client, PutObjectRequest
- Server exception: Problem on the server, client: no network?

AWS S3 Management Console

- view from management console: bucket, key (damage)

References I

Links

- https://d0.awsstatic.com/whitepapers/architecture/AWS_Well-Architected_Framework.pdf
- http://en.clouddesignpattern.org/index.php/Main_Page
- http://blogs.aws.amazon.com/security/post/Tx2XKTVE4JWD0F9/I Security-Services-Launched-at-AWS-re-Invent-2015-Amazon-Inspector-AWS-WAF-an
- https://aws.amazon.com/resources/gartner-2015-mq-learn-more/?trkCampaign=global_2015_ar_gartner_mq&trk=ha_ar_gart

Clark, Jack. 2012. "How Amazon Exposed Its Guts: The History of AWS's EC2." *ZDNet* (June). `http://www.zdnet.com/article/how-amazon-exposed-its-guts-the-history-of-awss-ec2/`.

Morgan, Timothy. 2014. "A Rare Peek Into The Massive Scale of AWS." *Enterprise Tech* (November). `http://www.enterprisetech.com/2014/11/14/rare-peek-massive-scale-aws/`.