EE 660: Computer Architecture
Cloud Architecture: IaaS

Yao Zheng
Department of Electrical Engineering
University of Hawai‘i at Mānoa

Based on the slides of Prof. Roy Campbell & Prof Reza Farivar
Agenda

- Infrastructure as a Service
- IaaS: OpenStack
- IaaS Providers: Amazon
- Serverless Architecture
- Amazon S3
- Dropbox
Xen and the Linux Kernel

- Xen was initially a university research project
- Invasive changes to the kernel to run Linux as a paravirtualized guest
- Even more changes to run Linux as dom0
- Xen support in the Linux kernel; no upstream
- Great maintenance effort on distributions
- Risk of distributions dropping Xen support
- Xen hard to use
Xen Concepts

• Control Domain a.k.a. Dom0
  • Dom0 kernel with drivers
  • Xen management toolstack
  • Trusted computing base

• Guest Domains
  • Your apps
  • For example, your cloud management stack

• Driver/Stub/Service Domain(s)
  • A "driver, device" model or “control service in a box"
  • De-privileged and isolated
  • Lifetime: start, stop, kill
Mirantis Fuel and OpenStack Architecture

Images = controller, compute, boot server, network OS, storage OS

Dashboard -> Fuel controller -> Boot server
Boot server -> OpenStack compute
OpenStack compute -> Node
Node -> Node

Images
OpenStack Architecture

Dashboard

OpenStack controller

Network

Compute

Storage

User applications

Shared resources
Mirantis Example

- Mirantis brings up a fuel controller
- The fuel controller allows you to bring up OpenStack controller, networks, compute servers, and storage
- https://training.mirantis.com/online-resources/fuel-video
Agenda

- Infrastructure as a Service
- IaaS: OpenStack
- IaaS Providers: Amazon
- Serverless Architecture
- Amazon S3
- Dropbox
OpenStack

“The OpenStack project has been created with the audacious goal of being the ubiquitous software choice for building cloud infrastructures.”
— Ken Pepble

“OpenStack is a collection of open source software projects that enterprises/service providers can use to set up and run their cloud compute and storage infrastructure.”
— docs.openstack.org
OpenStack Architecture

Dashboard

OpenStack controller

Network

Compute

Storage

User applications

Shared resources
OpenStack's Core Components

• **Compute ("Nova")**
  • Orchestrates large network for virtual machines (VMs)
  • Responsible for VM instance lifecycle, network management, and user access control

• **Object Storage ("Swift")**
  • Provides scalable, redundant, long-term storage for things like VM images, data archives, and multimedia
OpenStack's Core Components

• Image Service (“Glance”)
  • Manages VM disk images
  • Can be a stand-alone service
  • Supports private/public permissions, and can handle a variety of disk image formats
Agenda

• Infrastructure as a Service
• IaaS: OpenStack
• IaaS Providers: Amazon
• Serverless Architecture
• Amazon S3
• Dropbox
Amazon Web Services (AWS)

- AWS provides a collection of services for building Cloud applications
- Services for
  - Storage: S3, EBS
  - Database: RDS, DynamoDB, ElastiCache
  - Coordination: Simple notification service, simple workflow framework
- All services are paid depending on use
Amazon Web Services

Amazon EC2
Resizable compute capacity in the Cloud.

Amazon DynamoDB
Fast and flexible NoSQL database with seamless scalability.

AWS Lambda
Compute service that runs your code in response to events and automatically manages the compute resources.

Amazon S3
Highly scalable, reliable, and low-latency data storage infrastructure.
Nine Regions

- US East (N. Virginia) (135)
- US West (Oregon) (135)
- US West (N. California) (135)
- EU (Frankfurt) (60)
- EU (Ireland) (135)
- Asia Pacific (Singapore) (134)
- Asia Pacific (Tokyo) (134)
- Asia Pacific (Sydney) (133)
- South America (São Paulo) (130)
Six Types of Instances

• Micro instances (free tier)
• General purpose
• Memory optimized
• Storage optimized
• Compute optimized
• GPU instances
Amazon versus Competition

• Wall Street estimates of AWS revenue are typically in the $4 billion to $5 billion range for 2014, representing a doubling of revenue over a 2-year period, despite continuous decreases in prices.

• 10× capacity of its nearest 145 competitors combined (Gartner)
Amazon Technology

- Xen Hypervisor – proprietary
- Micro instances are oversubscribed, but others are not
- Storage and local area networks are shared – multi-tenant
- Billed by hour
- Provisional Amazon machine image instances – licenses are a user problem
- Operating system is a customer problem
- Can use third-party consoles to control Cloud VMWare, Microsoft
Storage

- Transient, instance-specific storage
- Persistent, instance-independent Elastic Block Store (EBS) storage (SSD and encryption options)
- Object-based Simple Storage Service (S3)
- Data restricted to region
Networking

• Virtual Private Cloud (=VPN)
• Private routing between VPCs
• VPN tunnels can connect your enterprise to Amazon
• DirectConnect allows a customer to use a carrier for private WAN services
Gain free, hands-on experience with AWS for 12 months

Learn more about the AWS Free Tier »

AWS re:Invent 2015
Registration is now open. Full conference pass $1299

WHAT IS CLOUD COMPUTING?
Learn about the benefits of on-demand IT resources and pay-as-you-go pricing

GET STARTED WITH AWS
Learn how to start using AWS in minutes

SOFTWARE FREE TRIALS
Try top software products from AWS Marketplace for free
Cloud Computing Applications – Roy Campbell

Resources

You are using the following Amazon EC2 resources in the US West (Oregon) region:

- Running Instances: 0
- Volumes: 0
- Key Pairs: 0
- Placement Groups: 0
- Elastic IPs: 0
- Snapshots: 0
- Load Balancers: 0
- Security Groups: 1

Automate application deployments to EC2 with CodeDeploy.

Create Instance

To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

Launch Instance

Note: Your instance will launch in the US West (Oregon) region.

Service Health

Service Status:
- US West (Oregon): This service is operating normally

Availability Zone Status:
- us-west-2a: Availability zone is operating normally
- us-west-2b: Availability zone is operating normally
- us-west-2c: Availability zone is operating normally

Scheduled Events

US West (Oregon):
No events

AWS Marketplace

Find free software trial products in the AWS Marketplace from the EC2 Launch Wizard. Or try these popular AMIs:

- Brocade 5400 Virtual Router/Firewall/VPN
  Provided by Brocade
  Rating 5 stars
  Pay by the hour for software and AWS usage
  View all Networking

- Alert Logic Threat Manager for AWS
  Provided by Alert Logic, Inc.
  Rating 5 stars
  Pay by the hour for software and AWS usage
  View all Security Software

- TIBCO Spotfire Analytics Platform (Hourly)
  Provided by TIBCO Software, Inc.
  Rating 5 stars
  Pay by the hour for software and AWS usage
  View all Security Software
# Cloud Computing Applications

*Roy Campbell*

## Step 1: Choose an Amazon Machine Image (AMI)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.

### Quick Start

<table>
<thead>
<tr>
<th>AMI Details</th>
<th>AMI</th>
<th>AMI Details</th>
<th>AMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Linux AMI 2015.03 (HVM), SSD Volume Type</td>
<td>ami-7527ed7</td>
<td>Red Hat Enterprise Linux 7.1 (HVM), SSD Volume Type</td>
<td>ami-4db9e7d</td>
</tr>
<tr>
<td>Free tier eligible</td>
<td>64-bit</td>
<td>Free tier eligible</td>
<td>64-bit</td>
</tr>
<tr>
<td>The Amazon Linux AMI is an EBS-backed, AWS-supported image. The default image includes AWS command line tools, Python, Ruby, Perl, and Java. The repositories include Docker, PHP, MySQL, PostgreSQL, and other packages.</td>
<td></td>
<td>Red Hat Enterprise Linux version 7.1 (HVM), EBS General Purpose (SSD) Volume Type</td>
<td></td>
</tr>
<tr>
<td>Root device type: ebs</td>
<td>Virtualization type: hvm</td>
<td>Root device type: ebs</td>
<td>Virtualization type: hvm</td>
</tr>
<tr>
<td>SUSE Linux Enterprise Server 12 (HVM), SSD Volume Type</td>
<td>ami-d7450be7</td>
<td>Ubuntu Server 14.04 LTS (HVM), SSD Volume Type</td>
<td>ami-5189e881</td>
</tr>
<tr>
<td>Free tier eligible</td>
<td>64-bit</td>
<td>Free tier eligible</td>
<td>64-bit</td>
</tr>
<tr>
<td>Root device type: ebs</td>
<td>Virtualization type: hvm</td>
<td>Root device type: ebs</td>
<td>Virtualization type: hvm</td>
</tr>
<tr>
<td>Free tier eligible</td>
<td>64-bit</td>
<td>Root device type: ebs</td>
<td>Virtualization type: hvm</td>
</tr>
</tbody>
</table>

*Are you launching a database instance? Try Amazon RDS.*

---

© 2005-2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
### Step 2: Choose an Instance Type

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. Learn more about instance types and how they can meet your computing needs.

Currently selected: t2.micro (Variable ECUs, 1 vCPUs, 2.5 GHz, Intel Xeon Family, 1 GiB memory, EBS only)

<table>
<thead>
<tr>
<th>Family</th>
<th>Type</th>
<th>vCPUs</th>
<th>Memory (GiB)</th>
<th>Instance Storage (GiB)</th>
<th>EBS-Optimized Available</th>
<th>Network Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>General purpose</td>
<td>t2.micro</td>
<td>1</td>
<td>1</td>
<td>EBS only</td>
<td>-</td>
<td>Low to Moderate</td>
</tr>
<tr>
<td>General purpose</td>
<td>t2.small</td>
<td>1</td>
<td>2</td>
<td>EBS only</td>
<td>-</td>
<td>Low to Moderate</td>
</tr>
<tr>
<td>General purpose</td>
<td>t2.medium</td>
<td>2</td>
<td>4</td>
<td>EBS only</td>
<td>-</td>
<td>Low to Moderate</td>
</tr>
<tr>
<td>General purpose</td>
<td>t2.large</td>
<td>2</td>
<td>8</td>
<td>EBS only</td>
<td>-</td>
<td>Low to Moderate</td>
</tr>
<tr>
<td>General purpose</td>
<td>m4.large</td>
<td>2</td>
<td>8</td>
<td>EBS only</td>
<td>Yes</td>
<td>Moderate</td>
</tr>
<tr>
<td>General purpose</td>
<td>m4.xlarge</td>
<td>4</td>
<td>16</td>
<td>EBS only</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>General purpose</td>
<td>m4.2xlarge</td>
<td>8</td>
<td>32</td>
<td>EBS only</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>General purpose</td>
<td>m4.4xlarge</td>
<td>16</td>
<td>64</td>
<td>EBS only</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>General purpose</td>
<td>m4.10xlarge</td>
<td>40</td>
<td>160</td>
<td>EBS only</td>
<td>Yes</td>
<td>10 Gigabit</td>
</tr>
<tr>
<td>General purpose</td>
<td>m3.medium</td>
<td>1</td>
<td>3.75</td>
<td>1 x 4 (SSD)</td>
<td>-</td>
<td>Moderate</td>
</tr>
</tbody>
</table>
Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click Launch to assign a key pair to your instance and complete the launch process.

**Warning:**

Improve your instances' security. Your security group, launch-wizard-1, is open to the world.

Your instances may be accessible from any IP address. We recommend that you update your security group rules to allow access from known IP addresses only.

You can also open additional ports in your security group to facilitate access to the application or service you're running, e.g., HTTP (80) for web servers. [Edit security groups](#)
Launch Status

✅ Your instances are now launching
The following instance launches have been initiated: i-30fc14f9  View launch log

📢 Get notified of estimated charges
Create billing alerts to get an email notification when estimated charges on your AWS bill exceed an amount you define (for example, if you exceed the free usage tier).

How to connect to your instances
Your instances are launching, and it may take a few minutes until they are in the running state, when they will be ready for you to use. Usage hours on your new instances will start immediately and continue to accrue until you stop or terminate your instances.

Click View Instances to monitor your instances' status. Once your instances are in the running state, you can connect to them from the Instances screen. Find out how to connect to your instances.

▼ Here are some helpful resources to get you started
- How to connect to your Linux instance
- Amazon EC2: User Guide
  - Amazon EC2: Discussion Forum

While your instances are launching you can also
Create status check alarms to be notified when these instances fail status checks. (Additional charges may apply)
Create and attach additional EBS volumes  (Additional charges may apply)
Manage security groups
Introduction to Serverless Architecture

• “Applications where some amount of server-side logic is still written by the application developer but unlike traditional architectures is run in stateless compute containers that are event-triggered, ephemeral (may only last for one invocation), and fully managed by a 3rd party”

• ‘Functions as a service / FaaS’

• AWS Lambda is one of the most popular implementations of FaaS at present, but there are others
Introduction to Serverless Architecture

Static web content served from Amazon S3

Web Client

Amazon API Gateway

AWS Lambda + ImageMagick

Mobile Client
Desktop Platform

Microsoft

Operating System

HAL

Dell

Hardware
Desktop Platform

Dell

Hardware

HAL

Microsoft

Operating System

Applications

QoS

Events

SDKs

Provision & Deployment

System Administrators

Software Developers

Runtime
AWS Cloud Platform, 2010

AWS (IaaS)

Hardware (S3, EC2, ELB etc.)

Cloud Operating System (Elasticity, VPN, DNS etc.)

Provision & Deployment
System Administrators
AWS Cloud platform, 2014

Hardware (S3, EC2, ELB)

AWS (IaaS)

Provision & Deployment

System Administrators => “DevOps”

AWS (PaaS)

Applications (Elastic Beanstalk, custom etc)

QoS

SDKs

Cloud Operating System (Elasticity, VPNs etc)

AWS (PaaS)

Software Developers

Runtime
AWS Cloud Platform 2016

* Lambda is a Game Changer

Hardware (S3, EC2, ELB) -> CAL

AWS (IaaS)

Provision & Deployment

System Administrators => “DevOps”

AWS (TCaaS)

Kinesis, EMR, Red Shift

Cloud Operating System (Elasticity, VPNs etc)

Runtime

Software Developers

QoS, Lambda, SDKs
AWS Elastic Beanstalk

• Deploy and scale web applications easily

• Languages: Java, .NET, PHP, Node.js, Python, Ruby, Docker

• Servers: Apache, Nginx, Phusion Passenger, IIS

• Simply upload your code; AWS handles:
  - Deployment
  - Capacity Provisioning
  - Auto scaling
  - Load balancing
  - Health Monitoring
AWS Lambda Event-driven Compute

• Runs stateless, request-driven code called **Lambda functions in Java, NodeJS & Python**
• Triggered by events (state transitions) in other **AWS services**
• Pay only for the requests served and the compute time
• Focus on business logic, not infrastructure.
• Just upload your code; AWS Lambda handles:
  
<table>
<thead>
<tr>
<th>Capacity</th>
<th>Monitoring</th>
<th>Fault Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaling</td>
<td>Logging</td>
<td>Security Patching</td>
</tr>
<tr>
<td>Deployment</td>
<td>Web service front end</td>
<td></td>
</tr>
</tbody>
</table>
AWS Lambda Event Sources

- Amazon S3
- Amazon DynamoDB
- Amazon Kinesis Streams
- Amazon Simple Notification Service
- Amazon Simple Email Service
- Amazon Cognito
- AWS CloudFormation
- Amazon CloudWatch Logs
- Amazon CloudWatch Events
- Scheduled Events (powered by Amazon CloudWatch Events)
- AWS Config
- Amazon Echo
- Amazon API Gateway
- Other Event Sources: Invoking a Lambda Function On Demand
- Sample Events Published by Event Sources
AWS Lambda Execution Environment

- State-less functions
- You can use multi-threading, etc.
- 500 MB of /tmp storage space
- You set how much memory you need:
  - From 128 MB to 1.5GB
  - 64GB increments
  - CPU scales accordingly
- Function should finish in a certain time
  - Default 3 seconds, up to 300 seconds
AWS Lambda Pricing

- You pay per use of your function
- $0.20 per 1 million function call
- Also, $0.00001667 for every GB-second used
Agenda

• Infrastructure as a Service
• IaaS: OpenStack
• IaaS Providers: Amazon
• Serverless Architecture
• Amazon S3
• Dropbox
Definition

A Binary Large OBject (BLOB) is a collection of binary data stored as a single entity in a database management system.

(The blob was a science fiction movie featuring Steve McQueen.)

https://en.wikipedia.org/wiki/Binary_large_object
Use case

- Store unstructured object data like text or binary data
- Images
- Movies
- Audio, Signal data
- Large queue of messages
- Example is LinkedIn data in a user page (Uses Ambry)
- Usually accessible over the web
Examples

• Windows Azure Blob Storage
• Ambry – LinkedIn
• Facebook’s Warm BLOB Storage System
• Amazon Simple Storage Service (S3)*
• Apache Open Stack Blob Service (SWIFT)
Goals

• Data growth ~ 50% a year
• 50%-70% data is unstructured or archival
• RESTful API (HTTP)
• High availability (no single point of failure)
• Agile data centers
• Open Source
• Multi-region, geographic distribution of data
• Storage policies
• Erasure Coding
Swift API

https://swift.illinois.edu/version2/auth_account/container/object

PUT /version2/roy/myblobs/classvideo1

GET /version2/reza/hisblobs/yesterdaysdataforhadoop
Swift Components

Web requests

Proxy

Account | Container | Object

Erasure Encoding
Write Requests - Load Balancer and Proxy

Client Requests → Load Balancer → Proxy → Use Quorum → Object Nodes

Cloud Computing Applications - Roy Campbell
Read Requests - Load Balancer and Proxy

Client Requests → Load Balancer → Proxy → Object Nodes

Cloud Computing Applications - Roy Campbell
Details

• MD5 Checksums with each object
• Auditing and active replication
• Any sized disks
Swift Partitions

• 1 Node, 8 Disks, 16 Partitions per disk, 8\*16 = 128 partitions
• 2 Nodes, 8 Disks each, 8 Partitions per disk, 8\*16 = 128 partitions
• Use Hash into Ring to map objects into storage partitions
Agenda

• Infrastructure as a Service
• IaaS: OpenStack
• IaaS Providers: Amazon
• Serverless Architecture
• Amazon S3
• Dropbox
Definition

Online file storage web service offered by Amazon Web Services. Amazon S3 provides storage through web service interfaces REST, SOAP, BitTorrent. (wikipedia)

https://aws.amazon.com/s3/
Use case

• Scalability, high availability, low latency – 99.99% availability
• Files up to 5 terabytes
• Objects stored in buckets owned by users
• User assigned keys refer to objects

• Amazon Machine Images (exported as a bundle of objects)
• SmugMug, Hadoop file store, Netflix, reddit, Dropbox, Tumbler
Simple Storage Service (S3)

- A **bucket** is a container for objects and describes location, logging, accounting, and access control.
  - A bucket has a name that must be **globally unique**.
    - [http://bucket.s3.amazonaws.com](http://bucket.s3.amazonaws.com)
    - [http://bucket.s3-aws-region.amazonaws.com](http://bucket.s3-aws-region.amazonaws.com).

- A bucket can hold any number of **objects**, which are files of up to 5TB.
  - [http://bucket.s3.amazonaws.com/object](http://bucket.s3.amazonaws.com/object)
Fundamental operations corresponding to HTTP actions:

http://bucket.s3.amazonaws.com/object

- POST a new object or update an existing object.
- GET an existing object from a bucket.
- DELETE an object from the bucket
- LIST keys present in a bucket, with a filter.

A bucket has a **flat directory structure**
S3 Weak Consistency Model

• “Updates to a single key are atomic....”
• “Amazon S3 achieves high availability by replicating data across multiple servers within Amazon's data centers. If a PUT request is successful, your data is safely stored. However:
  • A process writes a new object to Amazon S3 and immediately attempts to read it. Until the change is fully propagated, Amazon S3 might report "key does not exist."
  • A process writes a new object to Amazon S3 and immediately lists keys within its bucket. Until the change is fully propagated, the object might not appear in the list.
  • A process replaces an existing object and immediately attempts to read it. Until the change is fully propagated, Amazon S3 might return the prior data.
  • A process deletes an existing object and immediately attempts to read it. Until the deletion is fully propagated, Amazon S3 might return the deleted data.”
S3 Command Line Interface

aws  s3  mb  s3://bucket
...  cp  localfile  s3://bucket/key
mv  s3://bucket/key  s3://bucket/newname
ls  s3://bucket
rm  s3://bucket/key
rb  s3://bucket

aws  s3  help
aws  s3  ls  help

Cloud Computing Applications - Roy Campbell 7
Agenda

- Infrastructure as a Service
- IaaS: OpenStack
- IaaS Providers: Amazon
- Serverless Architecture
- Amazon S3
- Dropbox
Cloud Storage

• One interesting case study is Dropbox
• Dropbox offers cloud file storage
  • Easily synced across multiple devices
  • Accessible through web interface, mobile apps, and directly integrated with the file system on PCs
• Dropbox itself uses clouds!
  • Metadata stored in Dropbox servers
  • Actual files stored in Amazon S3
  • Amazon EC2 instances run the logic
Dropbox Architecture
Dropbox API

• Two levels of API access to Dropbox
  • Drop-ins
    • Cross-platform UI components that can be integrated in minutes
    • *Chooser* allows instant access to files in Dropbox
    • *Saver* makes saving files to Dropbox easy
  • Core API
    • Support for advanced functionality like search, revisions, and restoring file
    • Better fit for deeper integration
Drop-In API

- Simple objects
  - *Chooser* available for JavaScript, Android and iOS
  - *Saver* on web and mobile web
- Handles all the authentication (OAuth), file browsing
- *Chooser* object returns the following:
  - Link: URL to access the file
  - File name
  - File Size
  - Icon
  - Thumbnails
- *Saver*
  - Pass in URL, filename and options
Core API

- Many languages and environments
  - Python, Ruby, PHP, Java, Android, iOS, OS X, HTTP
- Based on HTTP and OAuth
  - OAuth v1, OAuth v2
- Low-level calls to access and manipulate a user's Dropbox account
  - Create URL schemes
  - Upload files
  - Download files
  - List files and folders
  - Delta
  - Metadata access
  - Create and manage file sharing