STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

Cloud Computing

Evolving Architectures

Stevens Institute of Technology

CPE 654

Vinay Purohit

Founder and CTO, CloudJuncxion, Inc.

vp@cloudjuncxion.com

Definition

Collection of Servers, Storage and Applications accessible over a Network

Types

- Public Cloud
  - Accessible to the general public over the Internet
- Private Cloud
  - Accessible only to a certain group of people
  - Over the Internet or over a private network
  - Deployed by institutions that are prohibited or have concerns with Public Clouds
    - Financial Institutions, MilGov Agencies, Some Large Corporations, …
- Virtual Private Cloud
  - Dedicated resources within a Public Cloud setting

Categories

- IaaS
  - Infrastructure as a Service
  - Customers pick CPU, Memory, Storage, OS, Applications, …
- PaaS
  - Platform as a Service
  - Less configurable than IaaS
  - Restricted to selection or deployment of specific applications
- SaaS
  - Software as a Service
  - Fixed Web applications
  - No notion of running your own applications

Examples

- Amazon Web Services (AWS), Google Cloud, MS Azure, SoftLayer, …
  - Mix of IaaS and PaaS
- Google Search, Google Docs, Google Drive, Snapchat
  - SaaS
  - Hosted on Google Cloud
- Dropbox, SalesForce, Box, FreshBooks
  - SaaS
- Amazon.com
  - Hosted on AWS

Public Cloud Advantages

- Elasticity
  - Grow and shrink compute or storage allocation dynamically
- Scalability
  - Allocate thousands of Compute Engines or Petabytes of storage quickly
- High Availability
  - Multiple regions or zones protect against disaster scenarios
- Cost Savings and Reduced IT Complexity
  - Reduces need for on-premises hardware and IT staffing
  - Cloud provider does the upgrades, backups, patches, …
  - Customers focus on deploying/running applications

Issues

- Security and Control
  - Cloud installations are highly opaque
  - No visibility into access controls
  - Possibility of rogue employees accessing customer data
  - But use of best-practices may make Cloud installations safer than private ones!
- Isolation
  - Impact of one customer's use of resources on other customers
  - Noisy Neighbor phenomenon
- Availability
  - Cloud outages are not uncommon
  - Internet connectivity

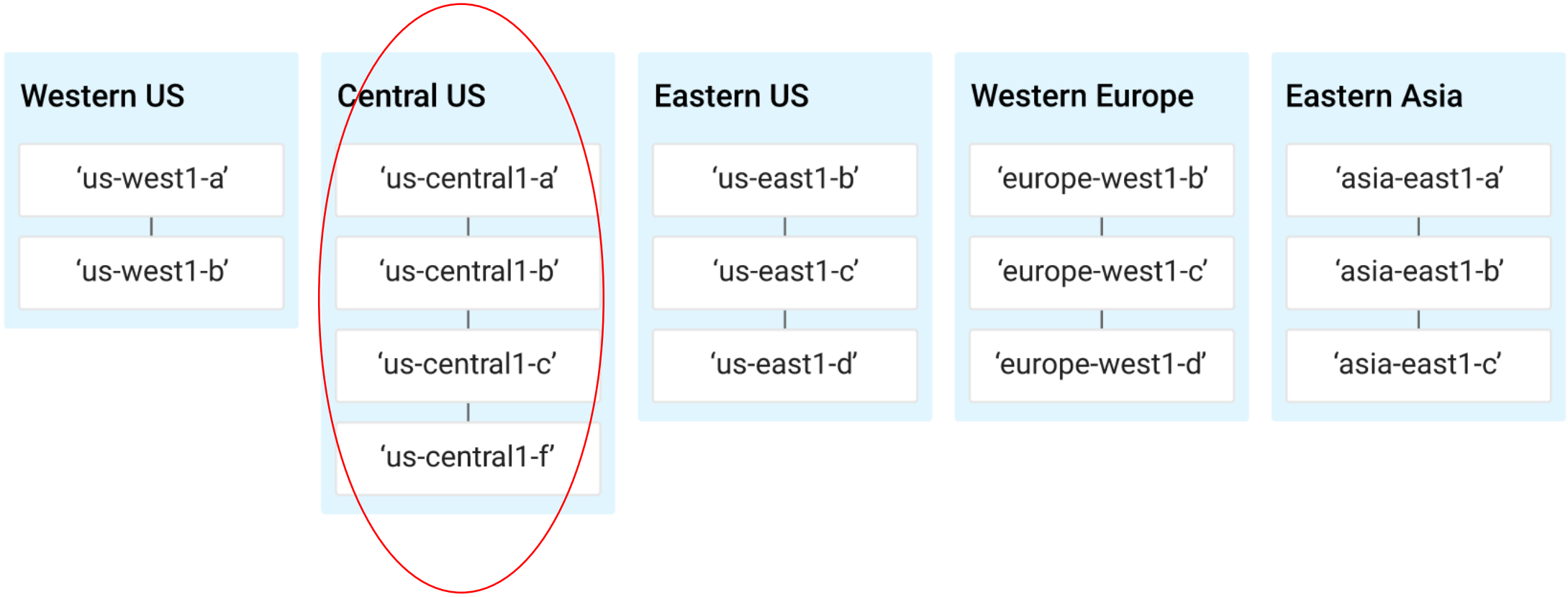
STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

Cloud Architecture

Cloud Physical Attributes

- One or two servers in a rack (cloudlet)
- One data-center with hundreds of servers and storage farms
- Multiple data-centers in a Zone
- Multiple Zones in a Region, Regions across countries.

Google Cloud



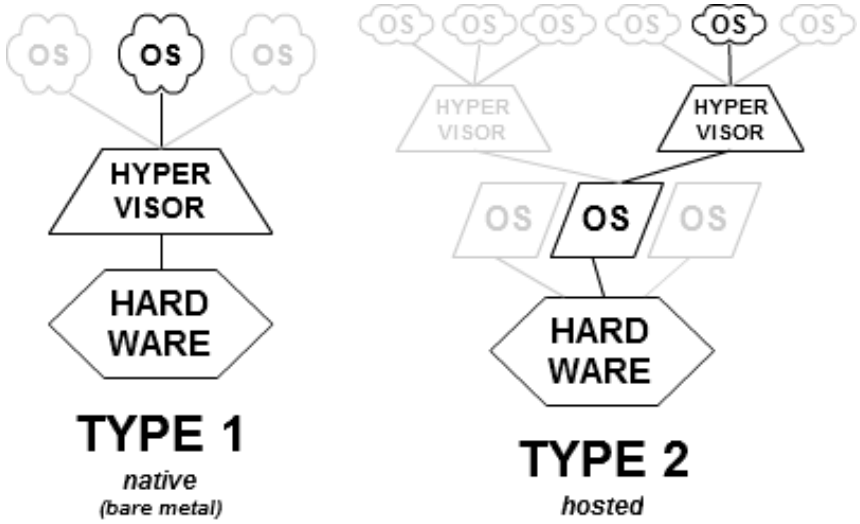
Region

Key Developments

- Hypervisors
- Virtual Machines (VMs)
- Bigger and Cheaper Data Pipes
  - Faster Internet Connectivity for Customers
  - 10/40/100G pipes between data centers
- Software Defined Networking (SDN)
  - Reconfigure networks with flexible public/private addressing
- Containers (Kubernetes, Docker)
- Object Storage

Hypervisors and VMs

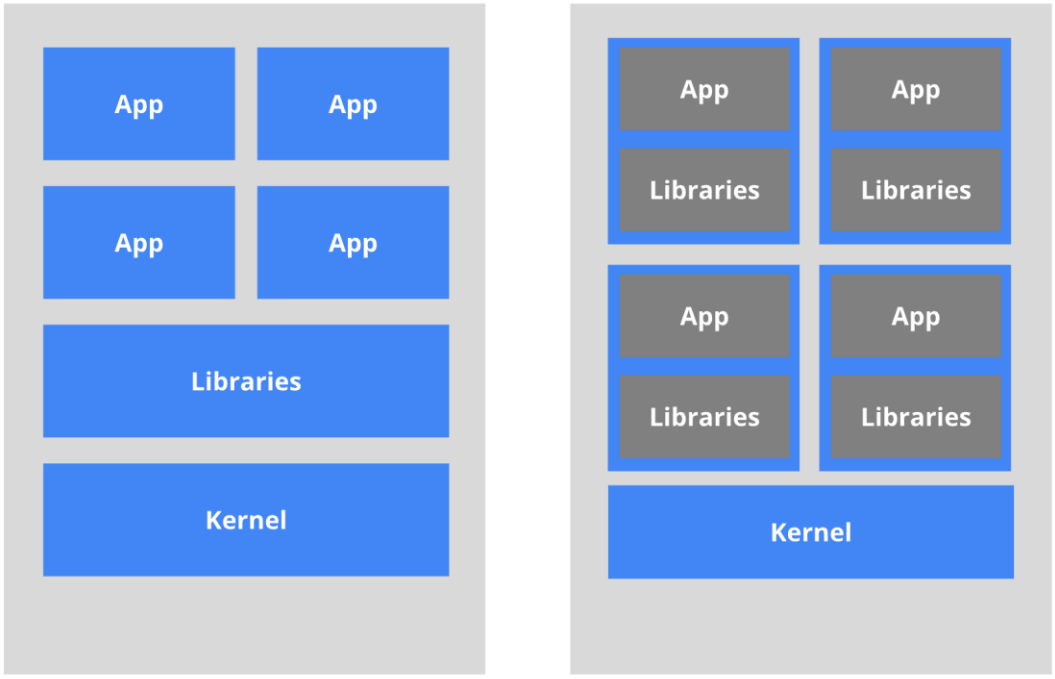
- Run multiple Virtual Machines (VMs) on a single hardware platform
- Each VM has its own OS, File System, IP Address, Processes
- VMs are isolated from each other
- Type 1 (Baremetal) and Type 2 (OS-Hosted) Hypervisors



Source: Wikipedia

Containers

- Useful for deploying and managing applications across a cluster
- E.g.: Kubernetes, Docker
- One app per image
- More lightweight than VMs
- Have their own load-balancers!



Heavyweight, non-portable
Relies on OS package manager

Traditional

Small and fast, portable
Uses OS-level virtualization

Containerized

Containers vs. VMs

- Two separate camps of thought
- VMs
  - Complete OS + multiple applications
- Containers
  - One application per image (includes bare minimum necessary to get application working)
- Kubernetes (Google's Container Management system)
  - Deploying across a large cluster of VMs
  - Focusing on the "What" instead of the How
  - Automatic scaling and transferring of application images
  - Starting, stopping, services
  - Built-in load balancers

Software Defined Networking (SDN)

STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

- Virtualization of Networks
- Separate Control Plane and Data Plane
- Control Plane
  - Determine how traffic is routed
  - Create rules and forwarding entries for the Data Plane
- Data Plane
  - Actual routing of packets and flows based on rules provided by the control plane

Object Storage

- Cloud Providers offer storage Buckets
- Instead of files and directories, you store Objects
- No limit on number of Objects per Bucket
- No notion of a directory hierarchy or file system
  - A, A/B, A/B/C are three different objects
  - A appears as a file and a directory at the same time!
  - The ‘/’ character is not a path separator, but part of the object name!
- You can store metadata with each object

Access Mechanisms

- Web Applications
  - Good for human interaction
- REST API
  - For interaction between client and cloud applications
  - HTTPS-based
- SSH
  - Secure Shell
  - Command Line Interface (CLI) access to VMs in the cloud
  - Good for code development, debugging, management of VMs, etc.

Authentication

- Standard Username/Password
- OAUTH 2.0
  - E.g., Login using your Google or Facebook account
  - RFC6749
  - Three-party handshake (User, Vendor, Identity Manager)
- HMAC SHA-1/SHA-2 Signatures (for REST transactions)
- Dual-Factor Authentication
  - Receive a security code on your mobile phone after you specify your password

REST

- REpresentational State Transfer
- A stateless way for client server interaction
- Provides a uniform mechanism for interaction between various nodes
- Self-descriptive Messages

REST Example (Object GET)

Request

```
GET /ObjectName HTTP/1.1
Host: BucketName.s3.amazonaws.com
Date: date
Authorization: authorization string
Range:bytes=byte_range
```

Response

```
HTTP/1.1 200 OK
x-amz-id-2: eftixk72aD6Ap51TnqcoF8eFidJG9Z/2mkiDFu8yU9AS1ed4OpIszj7UDNEHGran
x-amz-request-id: 318BC8BC148832E5
Date: Wed, 28 Oct 2009 22:32:00 GMT
Last-Modified: Wed, 12 Oct 2009 17:50:00 GMT
ETag: "fba9dede5f27731c9771645a39863328"
Content-Length: 434234
Content-Type: text/plain
Connection: close Server: AmazonS3
[434234 bytes of object data]
```

Load Balancers

- Machines that distribute requests across a farm of servers that actually do the processing
- Load Balancers usually have public IP addresses
  - Single name (google.com) but multiple IP addresses in their DNS entries
  - Clients pick one of the IP addresses at random when connecting with a Load-balancer
- Server farms may use Private IP addresses
- Benefits
  - Scalability
  - Fault-Tolerance
  - Reduced need for public IP addresses

STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

Cloud Applications

Examples

- Cloud-Based Accounting and Payroll Software
- Document management
- Big-Data Processing and Analytics
  - Use of MapReduce, Cloud DataFlow
- Machine Learning
  - Image recognition
  - Speech-to-Text conversion
  - Document Insights
- Deep Learning (Neural Networks)

Machine Learning (ML)

- Requires large compute and storage resources
- Ideal use case for Cloud Computing
- Large datasets fundamental to Machine Learning
  - Necessary for build Training Models
  - Type and nature of data critical for effective models

ML Examples

- Image Recognition (Tagging your smartphone photos)
- Natural Language Processing
- Meaning-based Document Search
- Email Spam Filtering
- Product Recommendations
  - Amazon.com
- Sentiment Analysis
  - Tone of a letter, article, speech
- Autonomous Driving (Waymo/Google, Tesla, Uber)

Transfer Learning

- What if you don’t have large datasets at your disposal?
- Use pre-trained models offered by the cloud providers
  - For e.g., Google’s Tensor Flow offers the ImageSet object-recognition model
- Enables rapid prototyping using new object types

Cost Implications

- Compute Engine Usage
  - Type of CPU, number of cores, Memory
  - Compute Time
- Object Storage
  - Total size of objects stored
  - Number of GET/PUT operations
- Networking
  - Egress Costs (the hidden surprise!)

Creating your own Cloud

- Open Stack
  - Open source software for creating your own IaaS installations
  - Runs on standard hardware
  - Components
    - Nova (Compute)
    - Neutron (Networking)
    - Swift (Object Storage)
    - Keystone (Identity Management)
    - ….
- Used by Rackspace, Oracle and HP for their own clouds

Review Questions

- Different types of clouds?
- IaaS vs. PaaS vs Saas
- What were the enablers of Cloud Computing?
- Difference between a Type 1 vs. Type 2 Hypervisor
- OAuth vs standard login/password
- How is Object Storage different compared to regular file systems?
- Difference between a VM and a Container
- What is a REST API and what is it used for?
- What are the key contributors to the costs of using a Cloud Provider?
- Why are the large Public Clouds well-suited for ML?