Cloud Computing Overview

Yiying Zhang
Outline

• What is cloud computing?

• Why cloud computing?

• How clouds are built?

• Quiz 1 at the end
Cloud Computing

- Datacenters that rent servers or other computing resources (e.g., storage)
- Anyone (or company) with a “credit card” can rent
- Cloud resources owned and operated by a third-party (cloud provider)

- Fine-grain pricing model
  - Rent resources by time or by I/O
  - Pay as you go (pay for only what you use)

- Can vary capacity as needed
  - No need to build your own IT infrastructure for peak needs
Cloud Computing

1. The illusion of infinite computing resources available on demand

2. The elimination of an up-front commitment by Cloud users

3. The ability to pay for use of computing resources on a short-term basis as needed

Source: Above the Clouds: A Berkeley View of Cloud Computing
XaaS (what can be rented?)

- **IaaS**: Infrastructure as a Service
  - Sell VMs or physical servers

- **PaaS**: Platform as a Service
  - Platforms (environments) for developers to develop and run applications. e.g., Google App Engine

- **SaaS**: Software as a Service
  - Offer services/applications e.g., Salesforce, Snowflake

- **FaaS**: Function as a Service

- All can be deployed at (public) cloud or local datacenters

source: https://azure.microsoft.com/en-us/overview/what-is-saas/
CLOUD MARKET REVENUE IN BILLIONS OF DOLLARS

Cloud Roles

Source: Above the Clouds: A Berkeley View of Cloud Computing
Cloud Models

- Who owns and runs the infrastructure?
  - Public clouds: resources shared by multiple organizations
  - Private clouds: resources used exclusively by one organization
  - Hybrid clouds
Cloud Infra Locations

- On-Premise (On-Prem)
  - Resources located locally (at a datacenter that the organization operates)
- Hosted
  - Resources hosted and managed by a third-party (cloud provider)
- Colocations (Colos)
  - A data center facility provides physical resources (e.g., building, cooling, power, etc.) for multiple organizations to host servers
- Private cloud can be both on-prem and hosted (virtual private cloud)
Hybrid Cloud

- Combine private (usually on-prem private) cloud and public cloud
  - Better control over sensitive data/functionalities
  - Cost effective
  - Scales well
  - Flexible
Cloud Usages

• Software/websites that serve real users
  - Netflix, Pinterest, Instagram, Spotify, Airbnb, Lyft, Slack, Expedia

• Data analytics, machine learning, and other data services
  - Databricks, Snowflake, GE Healthcare

• Mobile and IoT backend
  - Snapchat, Zynga (AWS->zCloud->AWS)

• Datacenter’s own usages
  - Google Drive/OneDrive, search, internal analytics
Cloud Providers

- Companies with large datacenters, often already running large-scale software
  - Amazon AWS
  - Microsoft Azure
  - Google Cloud Platform (GCP)
  - Alibaba Cloud
  - IBM Cloud

Amazon Web Service (AWS)

• Biggest market share, longest history

• Highest compute (and other service) options
  • >= 136 instance types in 26 families

• Storage
  • Simple Storage Service (S3)
  • Elastic Block Service (EBS)

• Many other services
  • Lambda (serverless)
  • ECS/EKS (managed containers)
  • DynamoDB, Aurora, ElastiCache (databases/key-value stores)
  • Virtual Private Cloud (VPC)
  • EMR, Redshift, many ML offerings (analytics, ML)
  • Satellite, Robotics
Cloud Services/Products/Resources
Microsoft Azure

- Moved from Windows to Linux
- Good integration with Microsoft products
  - Customers that are already using Microsoft products (e.g., having existing licenses)
- Many instance types and service types as well
Compute
Access cloud compute capacity and scale on demand—and only pay for the resources you use

Virtual Machines
Provision Windows and Linux virtual machines in seconds

Service Fabric
Develop microservices and orchestrate containers on Windows or Linux

Container Instances
Easily run containers on Azure without managing servers

SQL Server on Virtual Machines
Host enterprise SQL Server apps in the cloud

SAP HANA on Azure Large Instances
Run the largest SAP HANA workloads of any hyperscale cloud provider

Virtual Machine Scale Sets
Manage and scale up to thousands of Linux and Windows virtual machines

Mobile Apps
Build and host the backend for any mobile app

Linux Virtual Machines
Provision virtual machines for Ubuntu, Red Hat, and more

Azure CycleCloud
Create, manage, operate, and optimize HPC and big compute clusters of any scale

Azure Dedicated Host
A dedicated physical server to host your Azure VMs for Windows and Linux

Azure Kubernetes Service (AKS)
Simplify the deployment, management, and operations of Kubernetes

App Service
 Quickly create powerful cloud apps for web and mobile

Batch
 Cloud-scale job scheduling and compute management

Cloud Services
Create highly-available, infinitely-scalable cloud applications and APIs

Azure Functions
Process events with serverless code

Web Apps
Quickly create and deploy mission critical web apps at scale

API Apps
Easily build and consume Cloud APIs

Windows Virtual Desktop
Deliver a virtual desktop experience to any device at cloud scale

Azure VMware Solution by CloudSimple
Run your VMware workloads natively on Azure
Google Cloud Platform (GCP)

- Latest among the three to come in play and smallest market share, but with good growth
- Cheapest among the three
- Fewest instance types, allows customized CPU/memory sizes
  - bill based on total CPU and memory usages, not on total instance time
- TPU support
- Good support for cross geo-regions
- More open-source projects than the other two
Compute

- **Compute Engine**
  - Scalable, high-performance VMs.

- **Shielded VMs**
  - Hardened virtual machines on GCP

- **Google Kubernetes Engine (GKE)**
  - Run containerized applications.

- **Anthos on-premises**
  - Build and manage modern hybrid applications on existing VMware environments.

- **Container security**
  - Secure your container environment on GCP.

- **Migrate for Compute Engine**
  - Purpose-built, enterprise-grade migration to Google Cloud.

- **App Engine**
  - Serverless application platform for apps and backends.

- **Cloud Run (beta)**
  - Run stateless containers on a fully managed environment or on Anthos.

- **Cloud Functions**
  - Event-driven serverless compute platform.

- **Cloud Functions for Firebase**
  - Run mobile backend code without managing servers.

- **Knative**
  - Components to create modern, Kubernetes-native cloud-based software.

- **Graphics Processing Unit (GPU)**
  - Leverage GPUs on Google Cloud for machine learning, scientific computing, and 3D visualization.

Data analytics

- **BigQuery**
  - A fully managed, highly scalable data warehouse with built-in ML.

- **Cloud Dataflow**
  - Real-time batch and stream data processing.

- **Cloud Dataproc**
  - Managed Spark and Hadoop service.

- **Cloud Datalab**
  - Explore, analyze, and visualize large datasets.

- **Cloud Datastream**
  - Stream data between Google Cloud sources and destinations.

- **Cloud Composer**
  - A fully managed workflow orchestration service built on Apache Airflow.

- **Cloud Data Fusion**
  - Fully managed, code-free data integration.

- **Data Catalog**
  - A fully managed and highly scalable data discovery and metadata management service.

- **Genomics**
  - Power your science with Google Genomics.
Multi-Cloud

• Use multiple clouds for an application/service

• Avoid data lock-in

• Avoid single point of failure

• Need to deal with API differences and handle migration across clouds
Outline

• What is cloud computing?
  • Why cloud computing?
  • How clouds are built?
Incentive for Cloud Users

(a) Provisioning for peak load

(b) Underprovisioning 1

(c) Underprovisioning 2

Source: Above the Clouds: A Berkeley View of Cloud Computing
Incentive for Cloud Providers

• Make a lot of money
• Leverage existing investment
• Defend a franchise
• Attack an incumbent
• Leverage customer relationships
• Become a platform
The Virtuous Cycle of Cloud Computing

1. Lower prices attract more customers
2. More customers utilize more resources
3. Greater utilization of resources requires more infrastructure
4. Buying more infrastructure in volume leads to lower unit costs
5. Lower unit costs allow for lower customer prices

Source: CloudBlogs.Micorosoft.com
The Capital One hack couldn’t have come at a worse time for Amazon’s most profitable business

Amazon’s cloud-computing unit was considered the front-runner in a race to secure a $10 billion Pentagon contract. Now that could be in trouble.

The data breach that appears to have exposed more than 100 million applications for Capital One credit cards couldn’t have come at a worse time for Amazon Web Services, which stores the bank’s data.

The profit-driving Amazon unit, which allows companies to rent out storage and computing space on massive servers, has been the favorite to win a 10-year, $10 billion contract from the Defense Department, which had been expected to be announced this month. It was thrown into question
How the Jennifer Lawrence iCloud hack really happened

In early September 2014, Apple was preparing to announce a massive iPhone upgrade: bigger iPhones than anything it launched before. But just a few days ahead of the press iPhone 6 press event, an iCloud security scandal broke out. Nude pictures belonging to Jennifer Lawrence and many other celebrities leaked online, originating from iPhone backups.

Apple explained at the time that its iCloud security was not breached and that hackers probably employed phishing schemes to obtain the usernames and passwords from their victims.

Now, 18 months later after the scandal, we finally find out what happened. And it turns out that phishing attacks were indeed used to target the celebrities.

DON'T MISS: 11 paid iPhone apps on sale for free for a limited time

According to NBC News, it’s 36-year-old Ryan Collins the person responsible for phishing login credentials from many celebrities. With usernames and passwords in hand, he was able to log into Gmail accounts and even download iCloud backups from where he extracted nude photos.
ALAS-2019-1222

Amazon Linux AMI Security Advisory: ALAS-2019-1222
Advisory Release Date: 2019-06-17 17:58 Pacific
Severity: Critical

Issue Overview:
CVE-2019-11477, CVE-2019-11478 and CVE-2019-11479 describe vulnerabilities in the Linux kernel that can be remotely exploited using a specially crafted TCP connection, crashing the targeted system.

The latest Amazon Linux AMIs as available in AWS EC2 already contain these kernels and are not vulnerable.

Affected Packages:
kernel

Issue Correction:
Run yum update kernel and reboot to update your system.
The Impact of GDPR on Cloud Computing

22 May 2018

The EU General Data Protection Regulation (GDPR) is the most significant piece of privacy legislation to come into effect across Europe in a generation.

It will apply to any organisation who handles EU citizen data, even if they’re not from the EU. This means any European company with employees or a US firm with European customers will have to comply or face the consequences.

**GDPR** places new obligations on businesses, ones which will affect how they use cloud services. With cloud adoption now around 90% in the UK, it is important to ensure that the cloud services you use are compliant and that the systems and applications you design do not expose you to risk.

**Strengthening privacy**

GDPR strengthens user privacy in two main ways. Firstly, it increases the obligations on organisations to protect user data and secondly it grants citizens major new powers over how their information is collected, used and stored.

For example, businesses must ensure that all reasonable steps are taken to secure data, train staff and disclose breaches and must be clear and transparent to citizens about how they use personal data. Citizens can demand to see what data an entity is held on them and can also request that this is deleted at any time.

## Obstacles and Opportunities

<table>
<thead>
<tr>
<th>Obstacle</th>
<th>Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability of Service</td>
<td>Use Multiple Cloud Providers; Use Elasticity to Prevent DDOS</td>
</tr>
<tr>
<td>Data Lock-In</td>
<td>Standardize APIs; Compatible SW to enable Surge Computing</td>
</tr>
<tr>
<td>Data Confidentiality and Auditability</td>
<td>Deploy Encryption, VLANs, Firewalls; Geographical Data Storage</td>
</tr>
<tr>
<td>Data Transfer Bottlenecks</td>
<td>FedExing Disks; Data Backup/Archival; Higher BW Switches</td>
</tr>
<tr>
<td>Performance Unpredictability</td>
<td>Improved VM Support; Flash Memory; Gang Schedule VMs</td>
</tr>
<tr>
<td>Scalable Storage</td>
<td>Invent Scalable Store</td>
</tr>
<tr>
<td>Bugs in Large Distributed Systems</td>
<td>Invent Debugger that relies on Distributed VMs</td>
</tr>
<tr>
<td>Scaling Quickly</td>
<td>Invent Auto-Scaler that relies on ML; Snapshots for Conservation</td>
</tr>
<tr>
<td>Reputation Fate Sharing</td>
<td>Offer reputation-guarding services like those for email</td>
</tr>
<tr>
<td>Software Licensing</td>
<td>Pay-for-use licenses; Bulk use sales</td>
</tr>
</tbody>
</table>

*Source: Above the Clouds: A Berkeley View of Cloud Computing*
10 years later

- What’s your view on Cloud Computing?
Outline

• What is cloud computing?
• Why cloud computing?
• How clouds are built?
Virtualization

- Traditional: applications run on physical servers
  - Manual mapping of apps to servers
    - Apps can be distributed
    - Storage may be on a SAN or NAS
  - IT admins deal with “change”

- Modern: virtualized data centers
  - App run inside virtual servers; VM mapped onto physical servers
  - Provides flexibility in mapping from virtual to physical resources
Virtualization Benefit

• Resource management is simplified
  • Application can be started from preconfigured VM/container images
  • Virtualization layer / hypervisor permits resource allocations to be varied dynamically
• VMs can be migrated without application down-time
Virtual Datacenter

• A cluster of machines, each running a set of VMs
  • drive up utilization by packing many VMs onto each cluster node
  • fault recovery is simplified
    • if hardware fails, copy VM image elsewhere
    • if software fails, restart VM from snapshot
  • can safely allow third parties to inject VM images into your data center
    • hosted VMs in the cloud, commercial computing grids
Recent Trend: Container

- Light-weight virtualization
  - Running multiple isolated user-space applications on one OS
  - Virtualization layer runs as an application within the OS
  - Focusing on performance isolation

- Example: Docker, LXC, Kubernetes, Xen Unikernel
Software-Defined Data Center

- All infrastructure is virtualized and delivered as a service & the control of this datacenter is entirely automated by software
## Traditional Data Center

<table>
<thead>
<tr>
<th>WINDOWS</th>
<th>LINUX</th>
<th>DATABASE</th>
<th>MISSION CRITICAL</th>
<th>BIG DATA</th>
<th>HPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Windows Icon]</td>
<td>![Linux Icon]</td>
<td>![Database Icon]</td>
<td>![Mission Critical Icon]</td>
<td>![Big Data Icon]</td>
<td>![HPC Icon]</td>
</tr>
</tbody>
</table>

Source: EMC
Software-Defined Data Center

Source: EMC
Quiz 1

• Closed book

• Good luck!