Lecture 1: Fundamentals of Cloud Computing

CSCI4180

Patrick P. C. Lee

Outline

- Definitions of cloud computing
- Key cloud computing concepts
  - elasticity, virtualization
- Service delivery models
  - SaaS, PaaS, IaaS
- Deployment scenarios
  - Public/private clouds
- Security issues

What is a Cloud?

- The term “cloud” is used as a metaphor for the Internet, an abstraction of computing resources
  - Computing resources are aggregated in **data centers**
- A **cloud** is a network that delivers requested virtual resources as a service [IBM’s definition]

What is Cloud Computing? Formal Definitions from NIST

- Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Peter Mell and Tim Grance, "The NIST Definition of Cloud Computing", 2011

What is Cloud Computing? Key Characteristics (from NIST)

1. On-demand self-service
   - Focuses on delivering IT services driven by user requests
   - No human interaction with the cloud provider
   - Cloud computing provides a means of delivering computing services that makes the underlying technology, beyond the user device, almost invisible
2. Ubiquitous network access
   - Focuses on delivering IT services anytime, anywhere, and through user-chosen devices
   - Users accessing services via Internet technologies expect a secure, “always-on” computing infrastructure that delivers as easily and reliably as electricity from a wall outlet

Peter Mell and Tim Grance, “The NIST Definition of Cloud Computing”, 2011

What is Cloud Computing? Key Characteristics (from NIST)

3. Resource pooling
   - Focuses on delivering IT services through resource pools that can expand and contract based on the requirements of the underlying workload and the usage characteristics
4. Elasticity
   - Resources can be rapidly and elastically scaled up or scaled down
5. Measured service
   - Focuses on delivering IT services that can be metered for usage and charged for (if needed) through pricing models including subscription, usagepricing
   - Service level agreements (SLAs) for quality of service

Peter Mell and Tim Grance, “The NIST Definition of Cloud Computing”, 2011

What is Cloud Computing? Service Model (NIST's definition)

- **SaaS (software as a service)**
  - deliver software as a service over the Internet
- **PaaS (platform as a service)**
  - provide platform for development and testing
- **IaaS (infrastructure as a service)**
  - provide virtual machine and storage and schedule resources on demand

Peter Mell and Tim Grance, "The NIST Definition of Cloud Computing", 2011

What is Cloud Computing? Formal Definitions from Berkeley

- **Cloud computing** refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services
- Cloud refers to an environment of datacenter hardware and software

Armbrust et al., "A View of Cloud Computing", Comm. of the ACM, 2010

What is Cloud Computing? My Opinions

- Cloud computing bundles existing hardware and software technologies (e.g., virtualization, multi-core, network, storage, etc.) to provide new software/computing/storage services

- Key question: How to bundle the technologies nicely to make them work well?

Why Use Clouds?

- Better capital utilization
  - Pay-as-you-go
- The unit cost of on-demand capacity may be higher than the unit cost per time unit of fixed capacity; offset by no charge when capacity is not being used
- Accelerate software development, deployment, and testing
  - Fast provisioning of resources
- Elasticity of resources
  - Scalable and flexible use of resources
- Access to complex infrastructure and resources without internal resources
  - Support for geographically distributed users
- New business opportunities

Driving Factors Towards Cloud Computing

- Traditional management of computing resources:
  - Poorly utilized resources driving up hardware and labor costs
  - Takes too long to create middleware infrastructures
  - Creating middleware infrastructures is a manual process and error prone
  - Each application must be sized to support peak load
  - Inability to use idle resources to handle extra load
- Cloud is an enabler for business and IT transformation

Concerns of Cloud Computing

- Maturity
  - Is the technology ready for production-level deployment?
- Standards
  - Still being developed
- Security concerns
  - Multiple customers sharing the same resources
- Interoperability
  - Many different vendor APIs
- Control of data
  - Organizational level of comfort with data being outside traditional IT

Grid vs. Cloud Computing

- Grid computing involves applying the resources of many computers in a network, working in concert or parallel, to solve a single problem at the same time
  - More from a high-performance perspective
- Cloud computing provides resources for many independent tasks
  - Can be viewed as an extension of grid computing with a more general service model

Successful Case Studies

- **Clouds in Market**
  - Amazon: EC2 (Elastic Compute Cloud)
  - Google: Gmail, Google Maps, Google Calendar, …
  - IBM: Blue Cloud
  - Microsoft: Azure
  - Salesforce: Sales Cloud, Service Cloud, Custom Cloud
  - Yahoo: Yahoo Cloud Computing
  - Rackspace: Rackspace Cloud
  - Aliyun

Outline

- Definitions of cloud computing
- Key cloud computing concepts
  - elasticity, virtualization
- Service delivery models
  - SaaS, PaaS, IaaS
- Deployment scenarios
  - Public/private clouds
- Security issues

Elasticity and Scalability

- Elasticity is the ability to expand or shrink a computing resource in real time, based on the user's computing requirements
  - The ability to scale
  - Sometimes referred to as "right-sizing"
- Cloud service providers provide services based on usage
- This usage must meet **service level agreements (SLAs)** while minimizing cost
- Elasticity and scalability are used to achieve this
  - Cloud services scale up to meet demand
  - Cloud services scale down when higher demand is not required
  - Customers only pay for services used

Provisioning and Deprovisioning

- Provisioning provides resources availability to users and software
  - A provisioning system controls applications available to users
  - And controls servers resources available to applications
- Deprovisioning provides resources reduction to users and software, while deallocating back-end resources
  - Hardware
  - Software
- Self-service provisioning allows customers to request the amount of computer services without going through a lengthy process.
  - Computing
  - Storage
  - Software
  - Process
  - Other resources
- Eliminates many time delays

Virtualization

- Virtualization involves a shift in thinking from physical to logical
  - Treating IT resources as logical resources rather than separate physical resources
- With virtualization, you can consolidate the following resources into a virtual environment:
  - Processors
  - Storage
  - Networks
- With virtualization, one physical resource can be made to look like multiple virtual resources
  - Virtual resources can have functions or features that are not available in their underlying physical resources.

What Can Be Virtualized?

- Virtualization may refer to:
  - Hardware, networks, storage, operating systems, applications, desktop, data
  - E.g., data virtualization may abstract data from different physical sources into one logical source
- The main advantage of virtualization in cloud computing is that the software is decoupled from the hardware
  - Decoupling allows hosting an individual application in an environment that is isolated from underlying operating system

Characteristics of Virtualization

- Partitioning
  - Run multiple application and operating systems in a single physical machine by partitioning the available resources
- Isolation
  - Virtual machines are completely isolated from hosts and other virtual machines
- Encapsulations
  - Encapsulate the entire state of a virtual machine in hardware-independent files

Benefits of Virtualization

- Consolidation to reduce hardware cost
  - Enables you to have a single server function as multiple virtual servers
- Optimization of workloads
  - Can increase the use of existing resources by enabling dynamic sharing of resource pools
- IT flexibility and responsiveness
  - Enables you to have a single, consolidated view of, and easy access to, all available resources in the network, regardless of location

Hypervisors

- Hypervisors are virtualization software that allow multiple operating systems to run on the same computer concurrently → **virtual machines**
- Use a thin layer of code in software or firmware to achieve fine-grained, dynamic resource sharing
- Provide the greatest level of flexibility in how virtual resources are defined and managed
- Primary technology of choice for system virtualization
- May mediate access to memory, data storage, processing capacity, and network connections
- An example of a hypervisor is VMware ESX

Containers

- Unlike VMs, **containers** don't bundle a full OS, but only provide libraries that are needed to run applications

- Docker is a container management service that allows developers to **build** applications, **ship** them into containers, and **run** the containers anywhere
- http://www.docker.com

Multitenancy

- Cloud services must enable multitenancy — different companies sharing the same underlying resources
- Software as a service modes of multitenancy:
  - Simple multitenancy —each customer has his own resources, which are segregated from other customers
    - relatively inefficient
  - Fine grain multitenancy —all resources are shared, but the customer data and access capabilities are segregated within the application
    - much more efficient offering superior economies of scale
- Platform as a service modes of multitenancy:
  - This delivery model architecture allows multiple customers to run their copy separately from other customers through virtualization
  - Each customers code is isolated from each other
- The key technical challenge of multitenancy is how to support multiple client organizations from shared instances of the software solution

Application Programming Interfaces (API)

- Cloud services should have standardized application programming interfaces (API)
- The interface defines how two or more applications and data sources can communicate with each other
  - Multiple applications communicating
  - Multiple data sources communicating
- The cloud API allows customers (companies) infrastructure or application to plug into the cloud
- Currently, different cloud vendors are developing different APIs
- Cloud APIs have not been standardized yet
  - Beware of vendor API lock-in
  - API integration may include SOAP and REST APIs

Billing and Metering of Services

- To calculate the customer charge, cloud usage is tracked via metered services
  - The billing service is automated
  - Customer should be able to monitor usage
- Billing services normally track:
  - Number of users
  - Capacity used
  - Services leveraged
- Metered services normally provide:
  - A dashboard providing insight into application and services running in the cloud
  - SLA being met in the cloud

Questions

- Name two differences between public clouds and traditional data centers.

- Multitenancy is a necessary component for cloud computing, but could be problematic. Explain one potential problem due to multi-tenancy.

- Explain one solution that can relieve the lock-in problem.

Outline

- Definitions of cloud computing
- Key cloud computing concepts
  - elasticity, virtualization
- Service delivery models
  - SaaS, PaaS, IaaS
- Deployment scenarios
  - Public/private clouds
- Security issues

Cloud Service Models

- Software as a service (SaaS)
  - Use of software or applications that are delivered via a network
- Platform as a service (PaaS)
  - The middleware platform and solution stack are accessible on the cloud
- Infrastructure as a service (IaaS)
  - Provision servers, storage, and networking resources

Peter Mell and Tim Grance, “The NIST Definition of Cloud Computing”, 2009

Cloud Computing Stack

By Nick Barcet, “What is Ubuntu Cloud”, Nov 2009

IaaS

- An infrastructure provider (IP) makes an entire computing infrastructure available “as a service”
- IPs manage a large pool of computing resources and use virtualization to assign and dynamically resize the resources required by customers
- Customers rent processing capacity, memory, data storage, and networking resources that are provisioned over a network
- Examples: Amazon EC2, S3, Azure storage

IaaS

Infrastructure services
Virtualized infrastructure –
Server, storage, network, facilities
Infrastructure for hosting cloud services,
dynamic provisioning
Servers
Storage
Networks

PaaS

- Service provider (SP) supplies the software platform or middleware where the applications run
- Service user is responsible for the creation, updating, and maintenance of the application
- The sizing of the hardware required for the execution of the software is made in a transparent manner
- Examples: Facebook, Google App Engine

PaaS

Middleware services
Database
Messaging
Application server
Java runtime
Web 2.0 runtime
Business process management

SaaS

- Service provider (SP) is responsible for the creation, updating, and maintenance of software and application
- Service user accesses the service through Internet-based interfaces
- Examples: Google’s Drive, iCloud, Dropbox

SaaS

Application services
Collaboration
Enterprise applications
Business processes
Industry applications
Analytics

Other Cloud Service Models

- Storage as a service
  - part of the IaaS
- Data as a service
- Testing as a service
- Integration as a service

- In short, use the cloud to offer a “utility” as a service

Serverless Computing

- Allows a cloud user to simply write the code and leaves all the server provisioning and administration tasks to the cloud provider

- Serverless computing = FaaS + BaaS
  - FaaS (Function-as-a-Service): cloud functions
  - BaaS (Backend-as-a-Service): serverless frameworks

- Examples: Amazon Lambda, Google Cloud Functions, Azure Functions

"Cloud Programming Simplified: A Berkeley View on Serverless Computing", 2019

Outline

- Definitions of cloud computing
- Key cloud computing concepts
  - elasticity, virtualization
- Service delivery models
  - SaaS, PaaS, IaaS
- Deployment scenarios
  - Public/private clouds
- Security issues

Cloud Deployment Models

- Public cloud
  - Service provider lets clients access the cloud via the Internet
  - Made available to the general public or a wide industry group
- Private cloud
  - The cloud infrastructure is used solely by the organization that owns it
  - May reside in-house or off premises
- Hybrid cloud
  - Composed of two or more clouds (private, public, or community) that remain unique entities, but that can interoperate using standard or proprietary protocols
- Community cloud
  - Shared by several organizations that have a common mission

Peter Mell and Tim Grance, “The NIST Definition of Cloud Computing”, 2009

Public Clouds

- Multitenant infrastructure
- Anyone may use
- Functions vary
- Fee arrangements vary

Private Clouds

- Secure, dedicated infrastructure
- User buys or leases the cloud

- These types of clouds are not burdened by network bandwidth and availability issues or potential security exposures that may be associated with public clouds. Private clouds can offer the provider and user greater control, security, and resilience.

Hybrid Clouds

- Allows applications and data to flow across clouds
- Requires interoperability, visibility, and management
- Supports a very flexible performance model

Community Clouds

- Used and controlled by a group of organizations with a shared interest
- Private cloud purchased by a single user to support a community of users
- Fees may be charged to subsidiaries
- Functions vary
- Common functions
  - Computer ower
  - Storage
  - Elasticity
  - Community-wide sharing of data and applications

Virtual Private Clouds

- A virtual private cloud (VPC) is dedicated to a single user within a public cloud.
- The virtual private cloud extends the customer network into the cloud provider's "space", making the additional resources available on demand

Migration Paths for Cloud Adoption

- Use public clouds
  - Smaller organizations can use resources provided by larger cloud service providers
- Develop private clouds
  - Build or procure private clouds
  - Metering so as to determine the costs
- Build or procure community clouds
  - For organizations that share common goals
  - Shared infrastructure or sandbox environment
- Use hybrid clouds
  - Balance workloads between clouds

Outline

- Definitions of cloud computing
- Key cloud computing concepts
  - elasticity, virtualization
- Service delivery models
  - SaaS, PaaS, IaaS
- Deployment scenarios
  - Public/private clouds
- Security issues

Understanding Security Risks

- Security challenges in cloud computing:
  - Security is trusted to the cloud provider; therefore, if the provider has not done a good job, there may be problems
  - Security is difficult to monitor, so problems may not be apparent until there is a problem
  - Measuring the quality of the cloud provider'ssecurity approach may be difficult because many cloud providers do not expose their infrastructure to customers
- Many security breaches are caused by insiders
  - The security approach must deal with internal and external threats
- Cloud service agreement (contract) is often crafted to protect the service provider, not the cloud customer
  - Cloud customers must have a deep level of understanding the contract

Virtualization and Multitenancy

- Virtualization and multitenancy technologies were not designed with strong isolation in place
  - Hypervisors have extended these risks, potentially exposing the operating system
  - Creating an environment where attackers can gain access at the operating system level (hypervisors) and higher level services (functionality and data)
- Side channel attacks:
  - VMs infer behaviors of other VMs

Reference: “Hey, You, Get Off of My Cloud: Exploring Information Leakage in. Third-Party Compute Clouds”, ACM CCS 2009

Nonstandard and Vulnerable APIs

- Cloud API are not standardized, forcing users of multiple cloud providers to maintain multiprogramming interfaces, increasing complexity and security risk
- Since an API offers access to the internals of a system, a weak API exposes consumers to a variety of security issues encompassing all of the operational exposure the of the compromised API's functionality

Internal Security Breaches

- The IT industry has well documented that over 70% of security violations are internal
  - This threat is amplified in cloud computing as both IT providers and consumers are under a single management domain

Data Corruption or Loss

- Data corruption or loss is amplified since the cloud provider is the source for a companies data, not the company itself
- These operational characteristics of the cloud environment, at the PaaS and SaaS layers, amplify the threat of data loss or leakage increase

User Account and Service Hijacking

- User account and service hijacking occurs when an attacker obtains your cloud services information and uses it to take over your cloud access
- If attackers gain access to a cloud user's credentials, they can eavesdrop on activities and transactions, manipulate or steal data, return falsified data, and redirect clients to illegitimate sites

References

- Required:
  - IBM, "*Fundamentals of Cloud Computing",* 2010
- Optional:
  - Armbrust et al., "A View of Cloud Computing", Comm. of the ACM, 2010
  - Peter Mell and Tim Grance, "The NIST Definition of Cloud Computing", 2011 version
  - Jonas et al., "Cloud Programming Simplified: A Berkeley View on Serverless Computing", 2019