

TCSS 562: SOFTWARE ENGINEERING FOR CLOUD COMPUTING

Cloud Computing: Intro to Cloud Computing

Wes J. Lloyd
 School of Engineering and Technology
 University of Washington - Tacoma



FEEDBACK FROM 10/10

- What is the difference between a switchboard architecture and the Monolithic all-services-in-one composition [A B C]?
- Difference is how the functions are called
- For all-services-in-one composition [A B C]
- Services are "called" internally
 - Client calls A, receives result from C
 - A calls B, B calls C, Results of C returned or published
- Objective is to reduce client/server network latency
- Eliminates network traffic

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.2

FEEDBACK - 2

- For switchboard architecture:
- All code is deployed in same package, calls remain separate
- Client calls switchboard, requests "A", requests "B", requests "C" request, etc...
- Switchboard includes case/if statement to route call
- Objective is to reduce cold starts
- Network traffic not reduced
- But call to A, initializes infrastructure for B and C

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.3

FEEDBACK - 3

- Is there a difference between PaaS and Serverless Computing (FaaS)?
- PaaS - Platform as a Service cloud
 - Deploy entire web application (e.g. WAR file) to cloud provider hosted web containers (e.g. Apache Tomcat)
 - Cloud infrastructure provisioned and managed at the APP level
 - Infrastructure not managed at microservice level
 - No need to manage web containers or servers
 - Examples: AWS Elastic Bean Stalk, Heroku, Google App Engine
 - Web container specific APIs

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.4

FEEDBACK - 4

- FaaS - Function as a Service cloud
 - Cloud infrastructure provisioned and managed to host individual microservices
 - Infrastructure managed at the microservice level
 - Deploy code for individual functions (microservices)
 - No need to manage infrastructure
 - May need to specify resource configurations (e.g. memory size)
 - No web container to be found
 - Vendor specific (or framework specific) FaaS APIs
 - Examples: AWS Lambda, Azure Functions, Google Cloud Functions, IBM Cloud Functions

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.5

OBJECTIVES - 2

- Introduction to Cloud Computing
 - Why study cloud computing?
 - History of cloud computing
 - Business drivers
 - Cloud enabling technologies
 - Terminology
 - Benefits of cloud adoption
 - Risks of cloud adoption

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.6

WHY STUDY CLOUD COMPUTING?

- LINKEDIN - TOP IT Skills from job app data
 - #1 Cloud and Distributed Computing
 - <https://learning.linkedin.com/week-of-learning/top-skills>
 - #2 Statistical Analysis and Data Mining
- FORBES Survey – 6 Tech Skills That'll Help You Earn More
 - #1 Data Science
 - #2 Cloud and Distributed Computing
 - <http://www.forbes.com/sites/laurencebradford/2016/12/19/6-tech-skills-thatll-help-you-earn-more-in-2017/>

October 15, 2018

TCCS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.7

WHY STUDY CLOUD COMPUTING? - 2

- Computerworld Magazine



October 15, 2018

TCCS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.8

A BRIEF HISTORY OF CLOUD COMPUTING

- John McCarthy, 1961
 - Turing award winner for contributions to AI
- "If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility... The computer utility could become the basis of a new and important industry..."



October 15, 2018

TCCS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.9

CLOUD HISTORY - 2

- Internet based computer utilities
- Since the mid-1990s
- Search engines: Yahoo!, Google, Bing
- Email: Hotmail, Gmail
- 2000s
- Social networking platforms: MySpace, Facebook, LinkedIn
- Social media: Twitter, YouTube
- Popularized core concepts
- Formed basis of cloud computing

October 15, 2018

TCCS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.10

CLOUD HISTORY: SERVICES - 1

- Late 1990s – Early Software-as-a-Service (SaaS)
 - Salesforce: Remotely provisioned services for the enterprise
- 2002 -
 - Amazon Web Services (AWS) platform: Enterprise oriented services for remotely provisioned storage, computing resources, and business functionality
- 2006 – **Infrastructure-as-a-Service (IaaS)**
 - Amazon launches Elastic Compute Cloud (EC2) service
 - Organization can "lease" computing capacity and processing power to host enterprise applications
 - Infrastructure

October 15, 2018

TCCS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.11

CLOUD HISTORY: SERVICES - 2

- 2006 – **Software-as-a-Service (SaaS)**
 - Google: Offers Google DOCS, "MS Office" like fully-web based application for online documentation creation and collaboration
- 2009 – **Platform-as-a-Service (PaaS)**
 - Google: Offers Google App Engine, publicly hosted platform for hosting scalable web applications on google-hosted datacenters


October 15, 2018

TCCS562: Software Engineering for Cloud Computing [Fall 2018]
 School of Engineering and Technology, University of Washington - Tacoma

L6.12

CLOUD COMPUTING
NIST GENERAL DEFINITION

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



October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.13

MORE CONCISE DEFINITION

“Cloud computing is a specialized form of distributed computing that introduces utilization models for remotely provisioning scalable and measured resources.”

From Cloud Computing Concepts, Technology, and Architecture
Z. Mahmood, R. Puttini, Prentice Hall, 5th printing, 2015

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.14

BUSINESS DRIVERS
FOR CLOUD COMPUTING

- Capacity planning
- Cost reduction
- Operational overhead
- Organizational agility

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.15

BUSINESS DRIVERS
FOR CLOUD COMPUTING

- Capacity planning
 - Process of determining and fulfilling future demand for IT resources
 - Capacity vs. demand
 - Discrepancy between capacity of IT resources and actual demand
 - Over-provisioning: resource capacity exceeds demand
 - Under-provisioning: demand exceeds resource capacity
 - Capacity planning aims to minimize the discrepancy of available resources vs. demand

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.16



Dwight, The Office TV sitcom

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.17

BUSINESS DRIVERS FOR CLOUD - 2

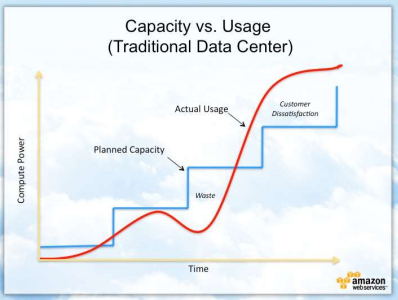
- Capacity planning
 - Over-provisioning: is costly due to too much infrastructure
 - Under-provisioning: is costly due to potential for business loss from poor quality of service
- Capacity planning strategies
 - Lead strategy: add capacity in anticipation of demand (pre-provisioning)
 - Lag strategy: add capacity when capacity is fully leveraged
 - Match strategy: add capacity in small increments as demand increases
- Load prediction
 - Capacity planning helps anticipate demand fluctuations

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.18

CAPACITY PLANNING



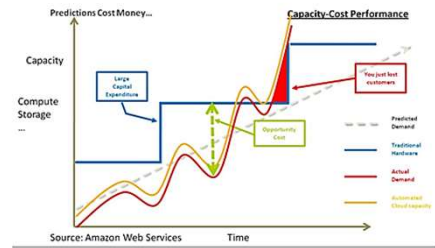
October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.19

CAPACITY PLANNING - 2

■ Ca



October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.20

BUSINESS DRIVERS FOR CLOUD - 3

■ Cost reduction

- IT Infrastructure acquisition
- IT Infrastructure maintenance

■ Operational overhead

- Technical personnel to maintain physical IT infrastructure
- System upgrades, patches that add testing to deployment cycles
- Utility bills, capital investments for power and cooling
- Security and access control measures for server rooms
- Admin and accounting staff to track licenses, support agreements, purchases

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.21

BUSINESS DRIVERS FOR CLOUD - 4

■ Organizational agility

- Ability to adapt and evolve infrastructure to face change from internal and external business factors
- Funding constraints can lead to insufficient on premise IT
- Cloud computing enables IT resources to scale with a lower financial commitment

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.22

TECHNOLOGY INNOVATIONS LEADING TO CLOUD

■ Cluster computing

■ Grid computing

■ Virtualization

■ Others

October 15, 2018


TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.23

CLUSTER COMPUTING

■ Cluster computing (clustering)

- Cluster is a group of independent IT resources interconnected as a single system
- Servers configured with homogeneous hardware and software
 - Identical or similar RAM, CPU, HDDs
- Design emphasizes redundancy as server components are easily interchanged to keep overall system running
 - Example: if a RAID card fails on a key server, the card can be swapped from another redundant server
- Enables warm replica servers
 - Duplication of key infrastructure servers to provide HW failover to ensure high availability (HA)




October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.24

GRID COMPUTING

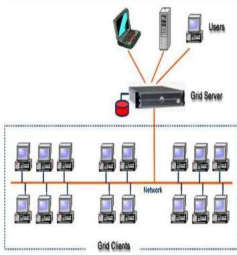


- On going research area since early 1990s
- Distributed heterogeneous computing resources organized into logical pools of loosely coupled resources
- For example: heterogeneous servers connected by the internet
- Resources are heterogeneous and geographically dispersed
- Grids use middleware software layer to support workload distribution and coordination functions
- Aspects: load balancing, failover control, autonomic configuration management
- Grids have influenced clouds contributing common features: networked access to machines, resource pooling, scalability, and resiliency

October 15, 2018	TCCS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.25
------------------	--	-------

GRID COMPUTING - 2

How Grid computing works ?

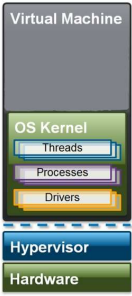


In general, a grid computing system requires:

- At least one computer, usually a server, which handles all the administrative duties for the System
- A network of computers running special grid computing network software.
- A collection of computer software called middleware

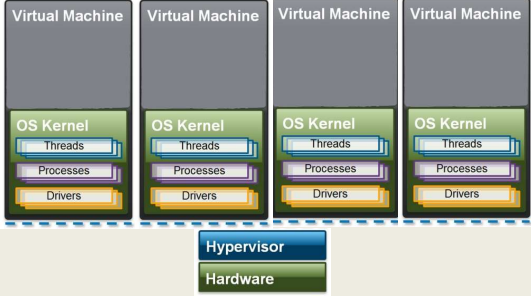
October 15, 2018	TCCS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.26
------------------	--	-------

VIRTUALIZATION



October 15, 2018	TCCS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.27
------------------	--	-------

VIRTUALIZATION



October 15, 2018	TCCS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.28
------------------	--	-------

VIRTUALIZATION

- Simulate physical hardware resources via software
 - The virtual machine (virtual computer)
 - Virtual local area network (VLAN)
 - Virtual hard disk
 - Virtual network attached storage array (NAS)
- Early incarnations featured significant performance, reliability, and scalability challenges
- CPU and other HW enhancements have minimized performance GAPS

October 15, 2018	TCCS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.29
------------------	--	-------

KEY TERMINOLOGY

- **On-Premise Infrastructure**
 - Local server infrastructure not configured as a cloud
- **Cloud Provider**
 - Corporation or private organization responsible for maintaining cloud
- **Cloud Consumer**
 - User of cloud services
- **Scaling**
 - **Vertical scaling**
 - Scale up: increase resources of a single virtual server
 - Scale down: decrease resources of a single virtual server
 - **Horizontal scaling**
 - Scale out: increase number of virtual servers
 - Scale in: decrease number of virtual servers

October 15, 2018	TCCS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.30
------------------	--	-------

VERTICAL SCALING

■ Reconfigure virtual machine to have different resources:

- CPU cores
- RAM
- HDD/SDD capacity

■ May require VM migration if physical host machine resources are exceeded

vertical scaling

A

B

2 CPUs

4 CPUs

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.31

HORIZONTAL SCALING

■ Increase (scale-out) or decrease (scale-in) number of virtual servers based on demand

pooled physical servers

virtual servers

A

A

B

A

B

C

demand

demand

horizontal scaling

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.32

HORIZONTAL VS VERTICAL SCALING

Horizontal Scaling	Vertical Scaling
Less expensive using commodity HW	Requires expensive high capacity servers

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.33

HORIZONTAL VS VERTICAL SCALING

Horizontal Scaling	Vertical Scaling
Less expensive using commodity HW	Requires expensive high capacity servers
IT resources instantly available	IT resources typically instantly available

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.34

HORIZONTAL VS VERTICAL SCALING

Horizontal Scaling	Vertical Scaling
Less expensive using commodity HW	Requires expensive high capacity servers
IT resources instantly available	IT resources typically instantly available
Resource replication and automated scaling	Additional setup is normally needed

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.35

HORIZONTAL VS VERTICAL SCALING

Horizontal Scaling	Vertical Scaling
Less expensive using commodity HW	Requires expensive high capacity servers
IT resources instantly available	IT resources typically instantly available
Resource replication and automated scaling	Additional setup is normally needed
Additional servers required	No additional servers required

October 15, 2018

TCSS562: Software Engineering for Cloud Computing [Fall 2018]
School of Engineering and Technology, University of Washington - Tacoma

L6.36

Slides by Wes J. Lloyd

L6.6

HORIZONTAL VS VERTICAL SCALING

Horizontal Scaling	Vertical Scaling
Less expensive using commodity HW	Requires expensive high capacity servers
IT resources instantly available	IT resources typically instantly available
Resource replication and automated scaling	Additional setup is normally needed
Additional servers required	No additional servers required
Not limited by individual server capacity	Limited by individual server capacity

KEY TERMINOLOGY - 2

- Cloud services
 - Broad array of resources accessible “as-a-service”
 - Categorized as Infrastructure (IaaS), Platform (PaaS), Software (SaaS)
- Service-level-agreements (SLAs):
 - Establish expectations for: uptime, security, availability, reliability, and performance

GOALS AND BENEFITS

- Cloud providers
 - Leverage economies of scale through mass-acquisition and management of large-scale IT resources
 - Locate datacenters to optimize costs where electricity is low
- Cloud consumers
 - Key business/accounting difference:
 - Cloud computing enables anticipated capital expenditures to be replaced with operational expenditures
 - Operational expenditures always scale with the business
 - Eliminates need to invest in server infrastructure based on anticipated business needs
 - Businesses become more agile and lower their financial risks by eliminating large capital investments in physical infrastructure

CLOUD BENEFITS - 2

- On demand access to pay-as-you-go resources on a short-term basis (less commitment)
- Ability to acquire “unlimited” computing resources on demand when required for business needs
- Ability to add/remove IT resources at a fine-grained level
- Abstraction of server infrastructure so applications deployments are not dependent on specific locations, hardware, etc.
 - The cloud has made our software deployments more agile...



CLOUD BENEFITS - 3

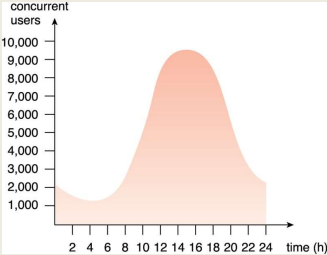
- Example: Using 100 servers for 1 hour costs the same as using 1 server for 100 hours
- Rosetta Protein Folding: Working with a UW-Tacoma graduate student, we recently deployed this science model across 5,900 compute cores on Amazon for 2-days...
- What Is the cost to purchase 5,900 compute cores?
- Recent Dell Server purchase example:
20 cores on 2 servers for \$4,478...
- Using this ratio 5,900 cores costs \$1.3 million (purchase only)



Gene Wilder, Charlie and the Chocolate Factory

CLOUD BENEFITS

- Increased scalability
 - Example demand over a 24-hour day →
- Increased availability
- Increased reliability



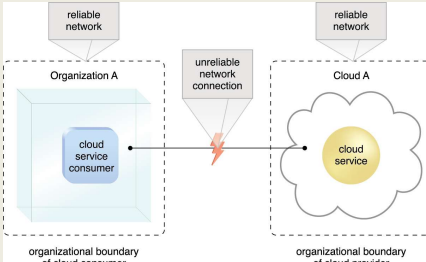
October 15, 2018	TCSS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.43
------------------	--	-------

CLOUD ADOPTION RISKS

- Increased security vulnerabilities
 - Expansion of trust boundaries now include the external cloud
 - Security responsibility shared with cloud provider
- Reduced operational governance / control
 - Users have less control of physical hardware
 - Cloud user does not directly control resources to ensure quality-of-service
 - Infrastructure management is abstracted
 - Quality and stability of resources can vary
 - Network latency costs and variability

October 15, 2018	TCSS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.44
------------------	--	-------

NETWORK LATENCY COSTS



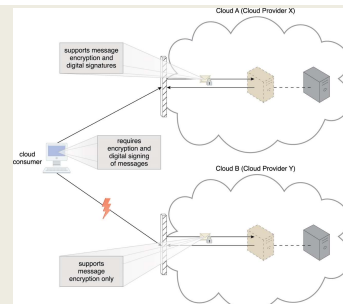
October 15, 2018	TCSS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.45
------------------	--	-------

CLOUD RISKS - 2

- Performance monitoring of cloud applications
 - Cloud metrics (AWS cloudwatch) support monitoring cloud infrastructure (network load, CPU utilization, I/O)
 - Performance of cloud applications depends on the health of aggregated cloud resources working together
 - User must monitor this aggregate performance
- Limited portability among clouds
 - Early cloud systems have significant "vendor" lock-in
 - Common APIs and deployment models are slow to evolve
 - Operating system containers help make applications more portable, but containers still must be deployed
- Geographical issues
 - Abstraction of cloud location leads to legal challenges with respect to laws for data privacy and storage


October 15, 2018	TCSS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.46
------------------	--	-------

CLOUD: VENDOR LOCK-IN



October 15, 2018	TCSS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.47
------------------	--	-------

QUESTIONS



October 15, 2018	TCSS562: Software Engineering for Cloud Computing [Fall 2018] School of Engineering and Technology, University of Washington - Tacoma	L6.48
------------------	--	-------