TCSS 562: SOFTWARE ENGINEERING FOR CLOUD COMPUTING

Introduction

Wes J. Lloyd
Institute of Technology
University of Washington - Tacoma

OBJECTIVES

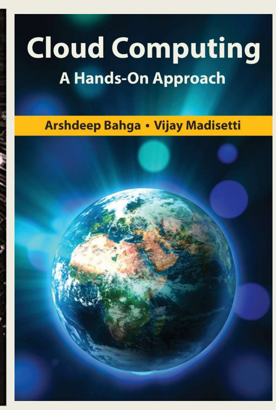
- Syllabus, Course Introduction
- Introduction to Cloud Computing
  - Why study cloud computing?
  - History of cloud computing
  - Business drivers
  - Cloud enabling technologies
  - Terminology
  - Benefits of cloud adoption
  - Risks of cloud adoption

TCSS562 – SOFTWARE ENGINEERING FOR CLOUD COMPUTING

- Syllabus
- Grading
- Schedule
- Assignments

REFERENCES

- Cloud Computing: Concepts, Technology and Architecture
- Thomas Erl, Prentice Hall 2013
- Cloud Computing: A Hands-On Approach
- Arshdeep Bahga, 2013
- Research papers



TCSS 562 – Spring 2018

- 16 in person class meetings
  - On line lectures or *project work days*: *(Instructor at conference)* Monday April 16, Wednesday April 18
  - No class: Monday April 9, Monday May 28
- No Final exam
- Project presentations – Wednesday June 6th
- Serverless Computing, Containerization

TCSS 562 SPRING 2018

March 26, 2018 | TCSS422: Operating Systems [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.5

TCS562 COURSE WORK

- Project Proposal
- Project Status Reports / Tutorials / Activities / Quizzes
  - ~ 5-8 total items
  - Variety of formats: in class, online, reading, tutorial / activity
- Midterm
  - Open book, note, etc.
- Class Presentation
- Term Project / Paper / Presentation

March 26, 2018 | TCSS422: Operating Systems [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.6

TCS562 TERM PROJECT

- Detailed project description to be posted
  – (be revised from S'17)
- Teams of 3-4, self formed, one project leader
- Proposal due: Tuesday April 10, 11:59pm (tentative)
- Last year:
  - Pick two or more cloud technology alternatives, implement a prototype service (e.g. data or application service)
  - Compare and contrast performance, cost, features for alternative cloud service implementations
- Deliverables
  - Presentation in class at end of quarter
  - Project report paper (4-6 pgs IEEE format, template provided)

March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.7

TERM PROJECT IDEAS – SPR 2017

- New for Spring 2018 – *Suggested list of projects TBA*
- From Last Spring 2017:
- Object/blob storage comparison
  - Amazon S3, Google blobstore, Azure blobstore, vs. self-hosted
- Cloud Relational Database
  - Amazon Relational Database Service vs. Self-Hosting
- Cloud Application containers
  - Amazon Elastic Beanstalk, Heroku, others
- Microservices / Serverless computing
  - Amazon Lambda, Azure Functions, Iron.io
- Cloud Load Balancer
  - Amazon Elastic Load Balancer vs. self-hosted load balancer
- Cloud VM performance comparison
  - Amazon, Microsoft, Google

March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.8

TERM PROJECT IDEAS - 2

- From Last Spring 2017:
- Storage systems evaluation
  - Amazon EBS, Amazon EFS, others
- Container Services
  - Amazon ECS, Azure Container Service
- Virtual machine imaging approaches
  - Across cloud vendors: Amazon, Google, Microsoft
- Queueing services comparison
  - Amazon SQS, others
- Application deployment comparison
  - Amazon, Google, Azure, Heroku
- Private cloud projects
  - Access NSF bare metal clouds: Chameleon, CloudLab, local UWT cluster resources

PROJECT SUPPORT

- Project cloud infrastructure support
- Sign up for the Github Student Developer Pack:
  - https://education.github.com/pack
  - Includes up to $150 in Amazon Cloud Credits
  - AWS credit extensions provided as needed
- Microsoft Azure
  - $200 free credit per account valid for 30 days
  - https://azure.microsoft.com/en-us/free/?b=17.09c
- Google Cloud
  - Education grant $50 credit coupons as needed
- Chameleon / CloudLab
  - Bare metal NSF cloud - free

TCSS562 TERM PROJECT OPPORTUNITIES

- Short papers can be submitted as a conference poster or workshop paper
  - ACM/IEEE/USENIX conferences, workshops
- Further project development:
  - Project provides preliminary work for:
    CS conference paper, UWT - MS capstone/thesis proposal
- Research publications (e.g. conference, workshop), help differentiate your resume from others
- Publications are essential if applying to PhD programs
- Research projects provide valuable practicum experience with cloud systems analysis, prototyping

TCSS562 TERM PROJECT - 3

- Project status report / term project check-ins
  - Written status report
  - Three times in quarter
  - Part of: ***"Project Status Reports / Tutorials / Activities / Quizzes"*** category
  - 20% of grade
- Project meetings with instructor
  - After class, end half of class, office hours

CLASS PRESENTATION

- Each student will make one presentation in a team of 2-3
- Technology sharing presentation
  - PPT Slides, demonstration
  - Provide technology overview of one cloud service offering
  - Present overview of features, performance, etc.
- Cloud Paper Presentation
  - PPT slides, identify research contributions, strengths and weaknesses of paper, possible areas for future work

OBJECTIVES

- 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

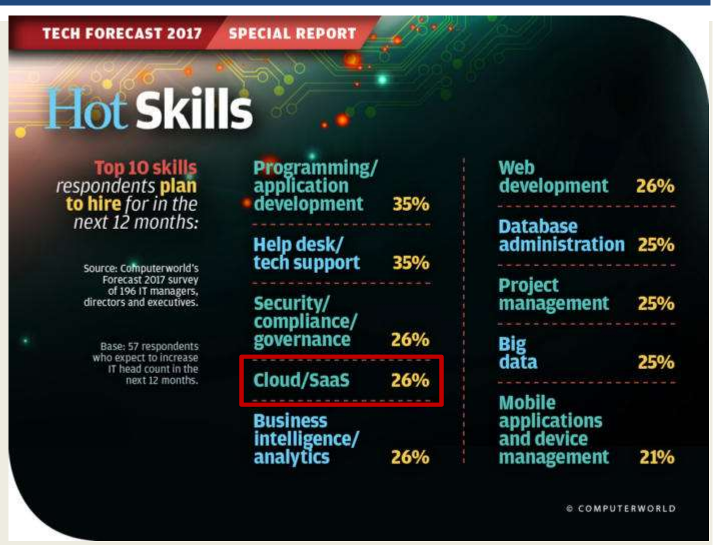
WHY STUDY CLOUD COMPUTING?

- LINKEDIN - TOP IT Skills (October 2016), 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 – 6 Tech Skills That'll Help You Earn More in 2017 (December 2016)
  - #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/

March 26, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L1.15

WHY STUDY CLOUD COMPUTING? - 2

- Computerworld December 2016



March 26, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L1.16

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…"

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

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

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

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"…

MORE CONCISE DEFINITION

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

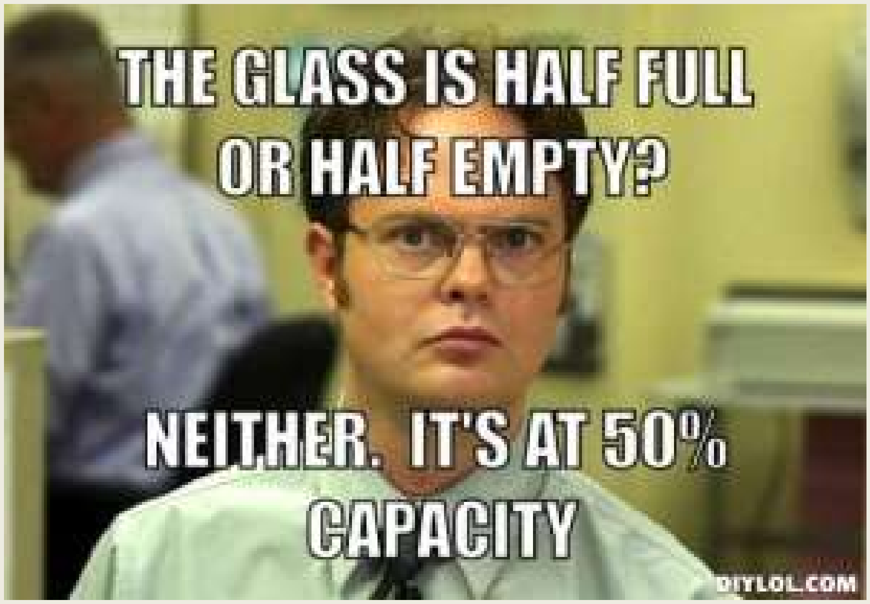
From Cloud Computing Concepts, Technology, and Architecture
Z. Mahmood, R. Puttini, Prentice Hall, 5th printing, 2015

BUSINESS DRIVERS FOR CLOUD COMPUTING

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

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



Dwight, The Office TV sitcom

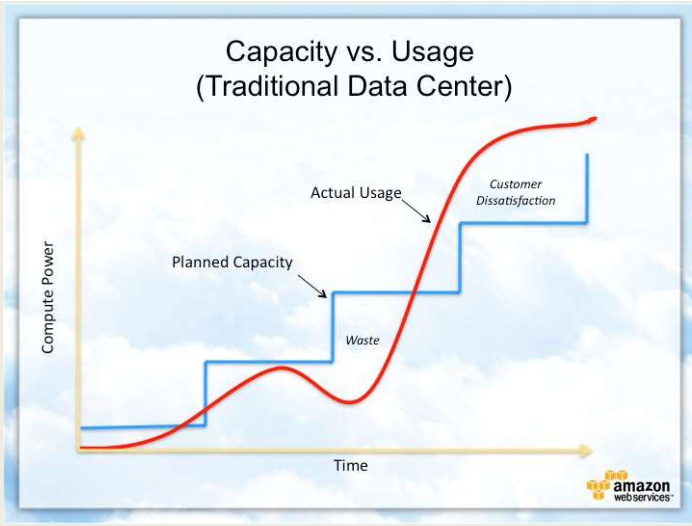
March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.25

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 flucations

March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.26

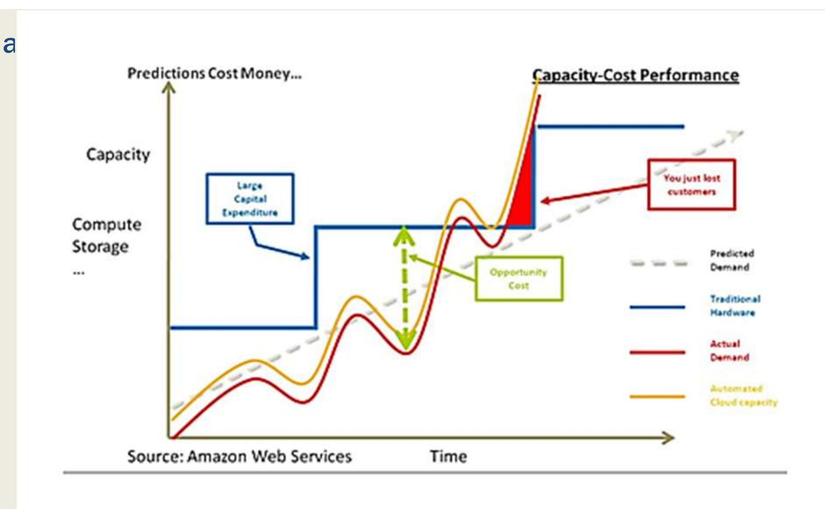
CAPACITY PLANNING



March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.27

CAPACITY PLANNING - 2

- Ca



March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.28

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

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

TECHNOLOGY INNOVATIONS LEADING TO CLOUD

- Cluster computing
- Grid computing
- Virtualization
- Others

CLUSTER COMPUTING

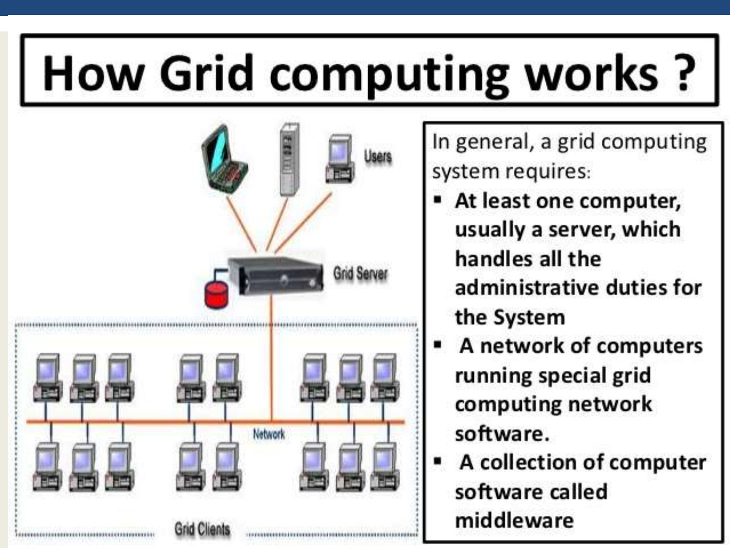
- Cluster computing (clustering)
  - Cluster is a group of independent IT resources interconnected as a single system
  - Servers configured with homogeneous hardware and software
    - Identical or similar RAM, CPU, HDDs
  - Design emphasizes redundancy as server components are easily interchanged to 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)

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

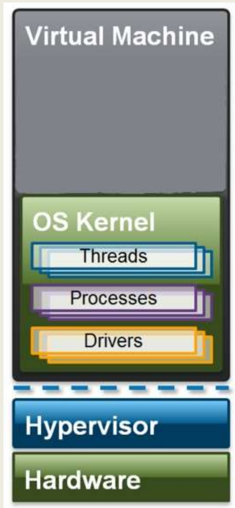
March 26, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L1.33

GRID COMPUTING - 2



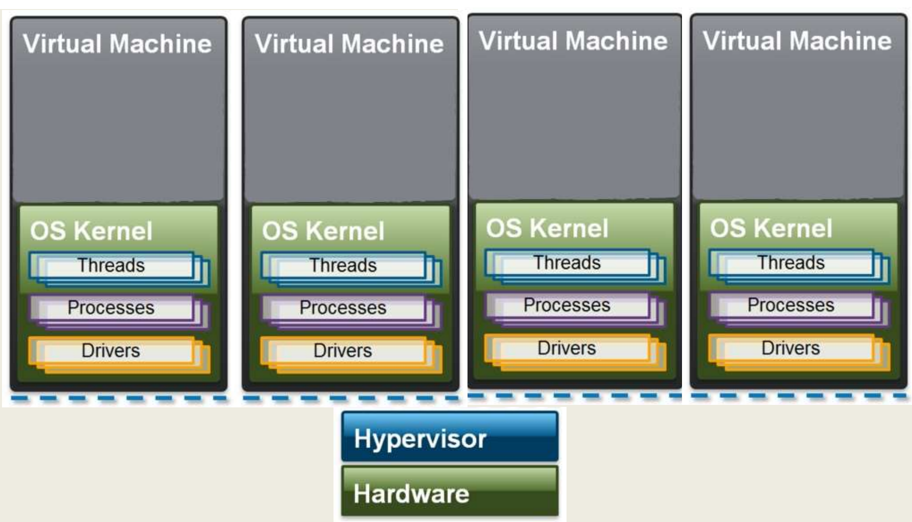
March 26, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L1.34

VIRTUALIZATION



March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.35

VIRTUALIZATION



March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.36

VIRTUALIZATION

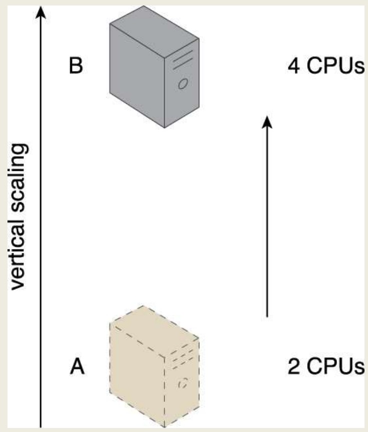
- Simulate physical hardware resources via software
  - The virtual machine (virtual computer)
  - Virtual local area network (VLAN)
  - Virtual hard disk
  - Virtual network attached storage array (NAS)
- Early incarnations featured significant performance, reliability, and scalability challenges
-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

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

VERTICAL SCALING

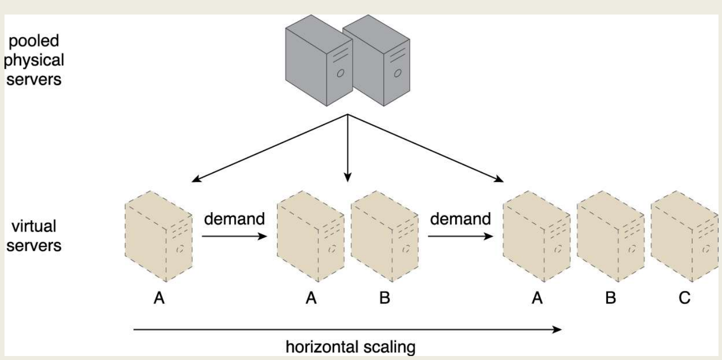
- Reconfigure virtual machine to have different resources:
  - CPU cores
  - RAM
  - HDD/SDD capacity
- May require VM migration if physical host machine resources are exceeded



March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.39

HORIZONTAL SCALING

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



March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.40

HORIZONTAL VS VERTICAL SCALING

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

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

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

March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.43

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

March 26, 2018 | TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma | L1.44

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

March 26, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L1.47

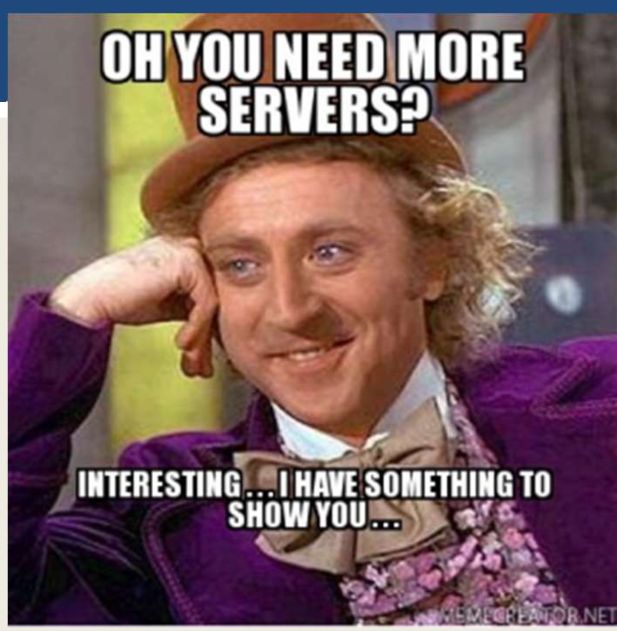
CLOUD BENEFITS - 2

- 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...

March 26, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L1.48

CLOUD BENEFITS - 3

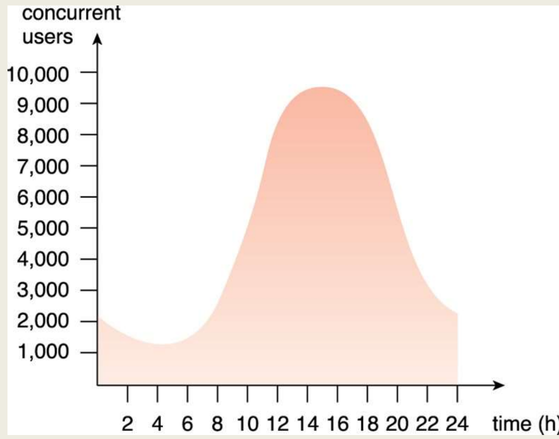
- Example: Using 100 servers for 1 hour costs the same as using 1 server for 100 hours
- Rosetta Protein Folding: Working with a UW-Tacoma graduate student, we recently deployed this science model across 5,900 compute cores on Amazon for 2-days…
- *What is the cost to purchase 5,900 compute cores?*
- 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



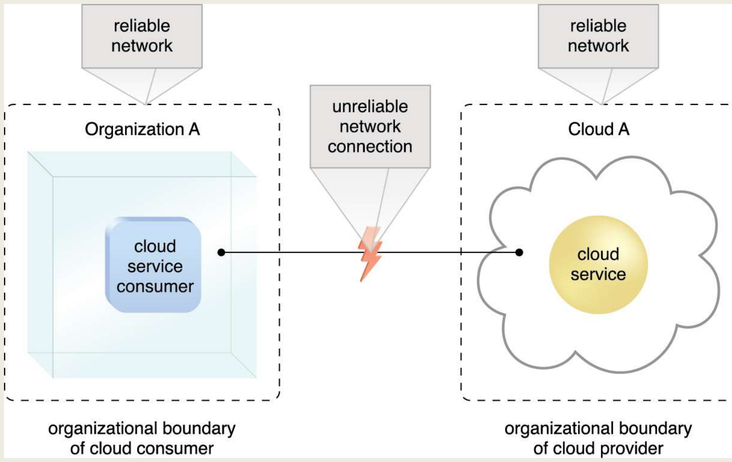
March 26, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L1.51

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

March 26, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L1.52

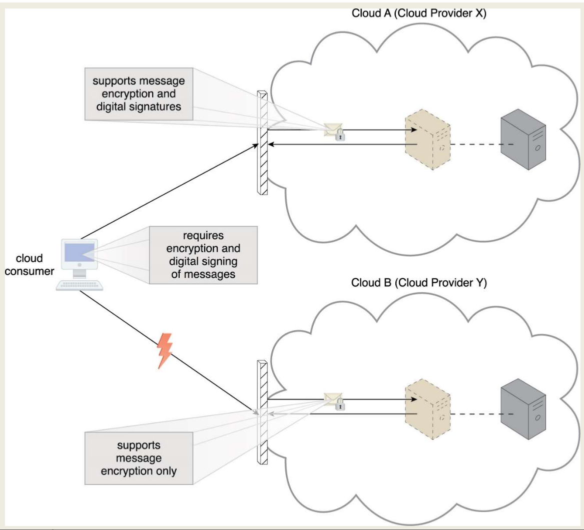
NETWORK LATENCY COSTS



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

CLOUD: VENDOR LOCK-IN



QUESTIONS