


TCSS 562: SOFTWARE ENGINEERING FOR CLOUD COMPUTING

Cloud Computing Concepts and Models

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



The logo features a blue cloud with the text 'Cloud Computing' inside. Below the cloud are three green teardrop shapes, each containing the text 'IaaS', 'PaaS', and 'SaaS' respectively.

OBJECTIVES

- Term project proposal
- Cloud Computing Concepts and Models
 - Roles and boundaries
 - Cloud characteristics
 - Cloud delivery models
 - Cloud deployment models

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.2
----------------	---	------

OBJECTIVES

- Term project proposal
- Cloud Computing Concepts and Models
 - **Roles and boundaries**
 - Cloud characteristics
 - Cloud delivery models
 - Cloud deployment models

March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.3

ROLES

- **Cloud provider**
 - Organization that provides cloud-based resources
 - Responsible for fulfilling SLAs for cloud services
 - Some cloud providers “resell” IT resources from other cloud providers
 - Example: Heroku sells PaaS services running atop of Amazon EC2
- **Cloud consumers**
 - Cloud users that consume cloud services
- **Cloud service owner**
 - Both cloud providers and cloud consumers can own cloud services
 - A cloud service owner may use a cloud provider to provide a cloud service (e.g. Heroku)

March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.4

ROLES - 2

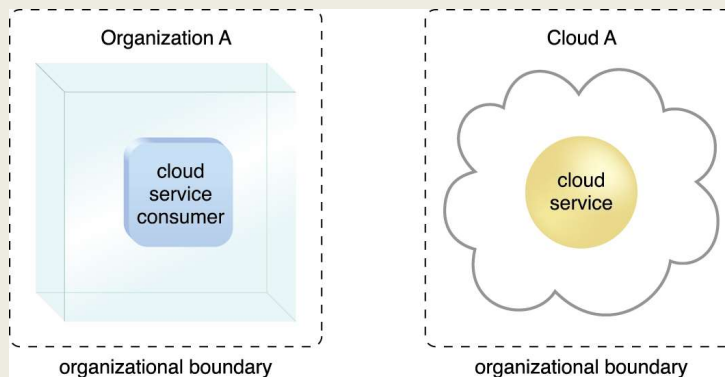
- **Cloud resource administrator**
 - Administrators provide and maintain cloud services
 - Both cloud providers and cloud consumers have administrators
- **Cloud auditor**
 - Third-party which conducts independent assessments of cloud environments to ensure security, privacy, and performance.
 - Provides unbiased assessments
- **Cloud brokers**
 - An intermediary between cloud consumers and cloud providers
 - Provides service aggregation
- **Cloud carriers**
 - Network and telecommunication providers which provide network connectivity between cloud consumers and providers

March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.5

ORGANIZATION BOUNDARY



March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.6

TRUST BOUNDARY

trust boundary

Organization A Cloud A

cloud service consumer cloud service

organizational boundary organizational boundary

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.7
----------------	---	------

OBJECTIVES

- Term project proposal
- Cloud Computing Concepts and Models
 - Roles and boundaries
 - **Cloud characteristics**
 - Cloud delivery models
 - Cloud deployment models

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.8
----------------	---	------

CLOUD CHARACTERISTICS

- On-demand usage
 - Ubiquitous access
 - Multitenancy (resource pooling)
 - Elasticity
 - Measured usage
 - Resiliency
- Assessing these features helps measure the value offered by a given cloud service or platform

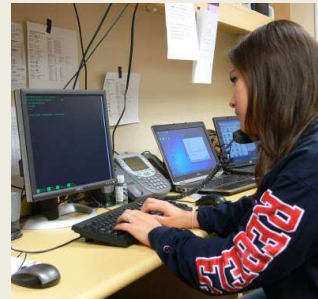
March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.9

ON-DEMAND USAGE

- The freedom to self-provision IT resources
- Generally with automated support
- Automated support requires no human involvement
- Automation through software services interface



March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.10

UBIQUITOUS ACCESS

- Cloud services are widely accessible
- Public cloud: internet accessible
- Private cloud: throughout segments of a company's intranet
- 24/7 availability

March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.11

MULTITENANCY

- Cloud providers pool resources together to share them with many users
- Serve multiple cloud service consumers
- IT resources can be dynamically assigned, reassigned based on demand
- Multitenancy can lead to performance variation

March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.12

SINGLE TENANT MODEL

The diagram illustrates the Single Tenant Model. It features a central cloud shape containing two distinct paths. On the left, 'Cloud Service Consumer A' (blue box) connects to 'Cloud Service A' (yellow circle), which in turn connects to 'Cloud Storage Device A' (teal cylinder). On the right, 'Cloud Service Consumer B' (blue box) connects to 'Cloud Service B' (yellow circle), which connects to 'Cloud Storage Device B' (teal cylinder). This represents a dedicated resource allocation for each tenant.

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.13
----------------	---	-------

MULTITENANT MODEL

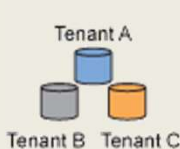
- Resource is “multiplexed” and share amongst multiple users
- Goal is to increase utilization
- Often server resources are underutilized
- There are many “sunk costs” whether usage is 0% or 100%
- Cloud computing tries to maximize “sunk cost” investments

The diagram illustrates the Multitenant Model. It features a central cloud shape containing two paths. On the left, 'Cloud Service Consumer A' (blue box) connects to 'Cloud Service A' (yellow circle). On the right, 'Cloud Service Consumer B' (blue box) connects to 'Cloud Service B' (yellow circle). Both 'Cloud Service A' and 'Cloud Service B' have arrows pointing to a single, shared 'shared cloud storage device' (teal cylinder) located at the bottom center of the cloud. A label 'shared cloud storage device' with a pointer indicates this shared resource.

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.14
----------------	---	-------

MULTITENANT DATABASE

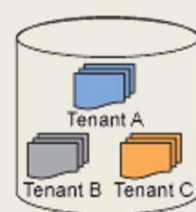
Isolated



Separate database

E1

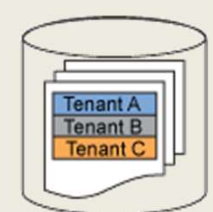
Semi-shared



**Shared database
Separate schema**

E2

Shared



**Shared database
Shared schema**

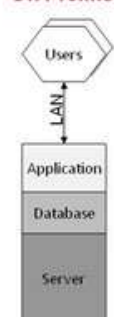
E3

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.15
----------------	---	-------

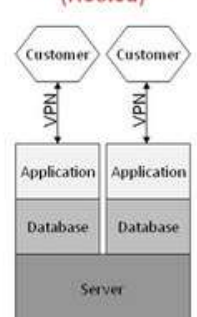
MULTITENANCY OF RESOURCES

■ Where is the multitenancy?

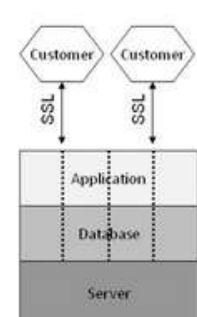
Traditional On Premise



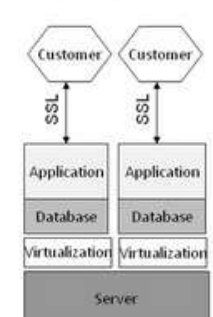
Single Tenant (Hosted)



Multi-Tenant



Virtual Appliance



March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.16
----------------	---	-------

ELASTICITY

- Automated ability of cloud to transparently scale resources
- Scaling based on runtime conditions or pre-determined by cloud consumer or cloud provider
- Threshold based scaling
 - CPU-utilization > threshold_A, Response_time > 100ms
 - Application agnostic vs. application specific thresholds
 - Why might an application agnostic threshold be non-ideal?
- Load prediction
 - Historical models
 - Real-time trends

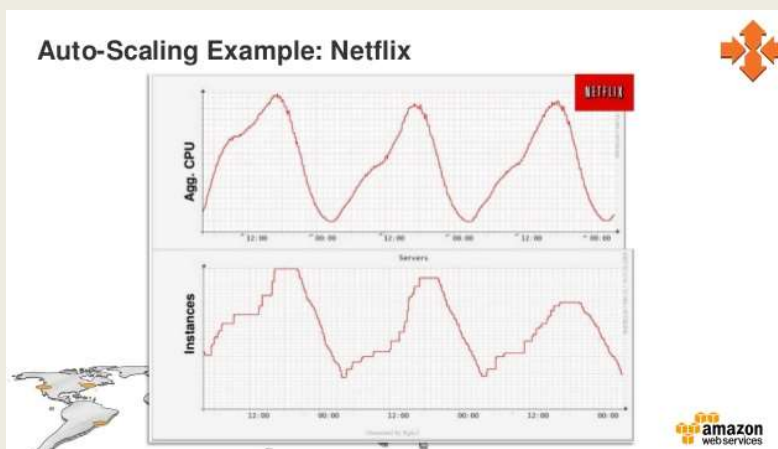
March 28, 2018

TCCS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.17

PREDICTABLE DEMAND

- Example:



March 28, 2018

TCCS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.18

MEASURED USAGE

- Cloud platform tracks usage of IT resources
- For billing purposes
- Enables charging only for IT resources actually used
- Can be time-based (minute, hour, day)
- Can be throughput-based (MB, GB)

- Not all measurements are for billing
- Some measurements can support auto-scaling
- For example CPU utilization

March 28, 2018	TCCS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.19
----------------	---	-------

EC2 CLOUDWATCH METRICS

The screenshot displays the AWS CloudWatch monitoring interface for an EC2 instance (i-1267037f). It features five line graphs showing metrics over a one-hour period from 11/16 23:30 to 11/17 00:00. The metrics are: Avg CPU Utilization (Percent), Avg Disk Reads (Bytes), Avg Disk Writes (Bytes), Max Network In (Bytes), and Max Network Out (Bytes). The CPU utilization shows a peak around 11/17 00:00. Disk reads and writes also show significant activity during this period. Network in and out metrics show a sharp spike at the end of the hour.

March 28, 2018	TCCS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.20
----------------	---	-------

EC2 CLOUDWATCH METRICS

The figure displays eight CloudWatch metrics for EC2 instances, arranged in a 2x4 grid. Each chart shows data from 4/25 00:00 to 5/4 00:00. The metrics are: CPU Utilization (Percent), Disk Reads (Bytes), Disk Read Operations (Operations), Disk Writes (Bytes), Disk Write Operations (Operations), Network In (Bytes), Network Out (Bytes), and Network Packets In (Count). The charts show various levels of activity, with Network In and Network Out showing significant data points.

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.21
----------------	---	-------

RESILIENCY

- Distributed redundancy across physical locations
- Used to improve reliability and availability of cloud-hosted applications
- Very much an engineering problem
- No “resiliency-as-a-service” for user deployed apps
- Unique characteristics of user applications make a one-size fits all service solution challenging

The image shows the cover of the book 'Resilience and Reliability on AWS' by Burg van Vliet, Flavio Paganelli, and Jasper Geurtsen. The cover features a black dog standing on a green background with the title text. The O'Reilly logo is visible at the bottom left of the cover.

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.22
----------------	---	-------

OBJECTIVES

- Term project proposal
- Cloud Computing Concepts and Models
 - Roles and boundaries
 - Cloud characteristics
 - **Cloud delivery models**
 - Cloud deployment models

March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.23

CLOUD DELIVERY MODELS



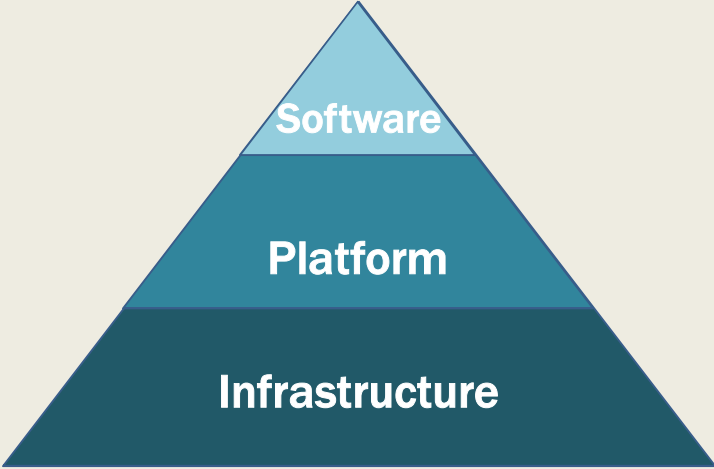
- What is the appropriate level of **abstraction**?
- How should applications be deployed?
 - IaaS, PaaS, SaaS, DBaaS, FaaS
- How do we ensure Quality-of-Service?
 - Performance, Availability, Responsiveness, Fault Tolerance
- How is **scalability** provided?
- How do we minimize hosting costs?
 - How do we estimate hosting costs?



March 28, 2018

24

CLOUD DELIVERY MODELS



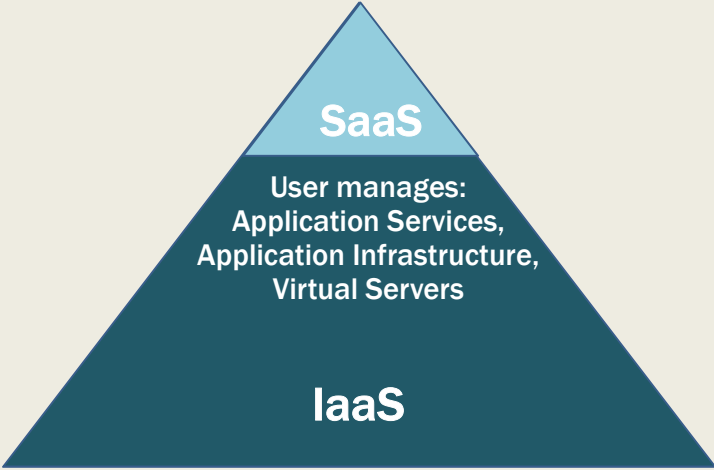
Software

Platform

Infrastructure

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.25
----------------	---	-------

CLOUD DELIVERY MODELS



SaaS

User manages:
Application Services,
Application Infrastructure,
Virtual Servers

IaaS

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.26
----------------	---	-------

CLOUD DELIVERY MODELS

SaaS
User manages:
Application Services
PaaS
IaaS

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.27
----------------	---	-------

CLOUD DELIVERY MODELS

SaaS
PaaS
IaaS

March 28, 2018	TCSS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.28
----------------	---	-------

INFRASTRUCTURE-AS-A-SERVICE

- Compute resources, on demand, as-a-service
 - Generally raw “IT” resources
 - Hardware, network, containers, operating systems
- Typically provided through virtualization
- Generally not-preconfigured
- Administrative burden is owned by cloud consumer
- Best when high-level control over environment is needed
- Scaling is generally **not** automatic...
- Resources can be managed in bundles
- AWS CloudFormation: Allows specification in JSON/YAML of cloud infrastructures

March 28, 2018

TCCS562: Software Engineering for Cloud Computing [Spring 2018]
 Institute of Technology, University of Washington - Tacoma

L2.31



Bell's Number:

k: number of ways n components can be distributed across containers

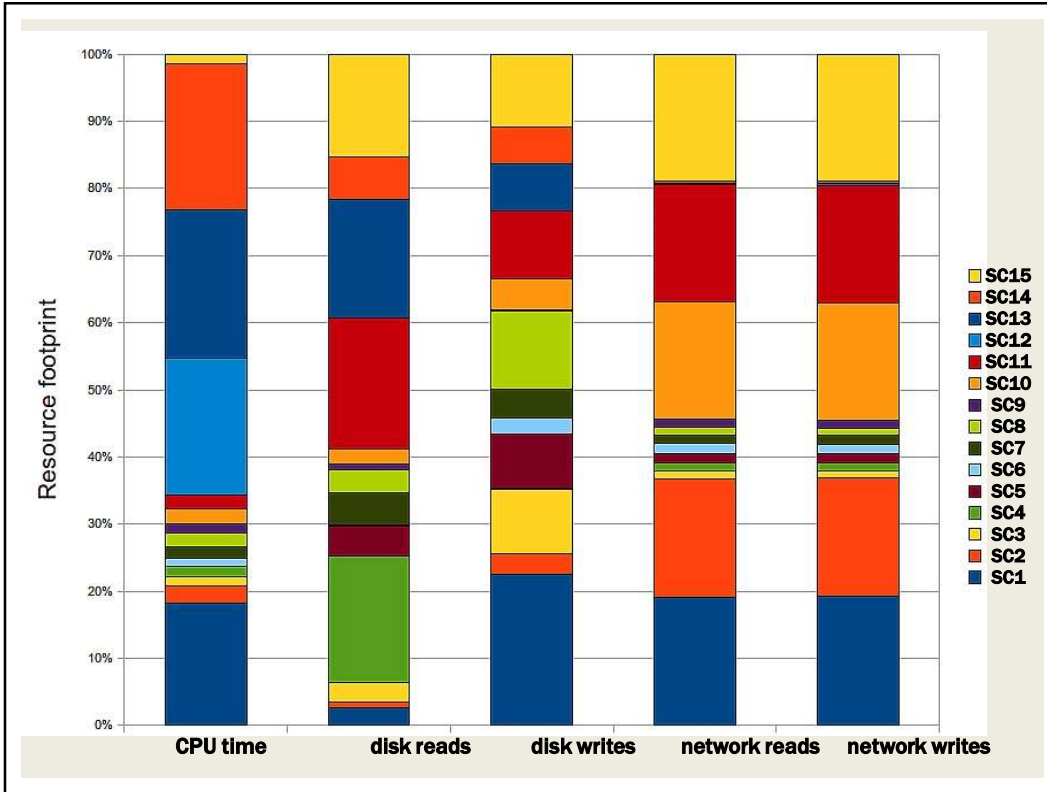
n	k
4	15
5	52
6	203
7	877
8	4,140
9	21,147
n	...

M: Tomcat ApplicationServer
D: Postgresql DB
F: nginx file server
L: Log server (Codebeamer) ³³

Component Composition Example

- An application with 4 components has 15 compositions
- One or more component(s) deployed to each VM
- Each VM launched to separate physical machine

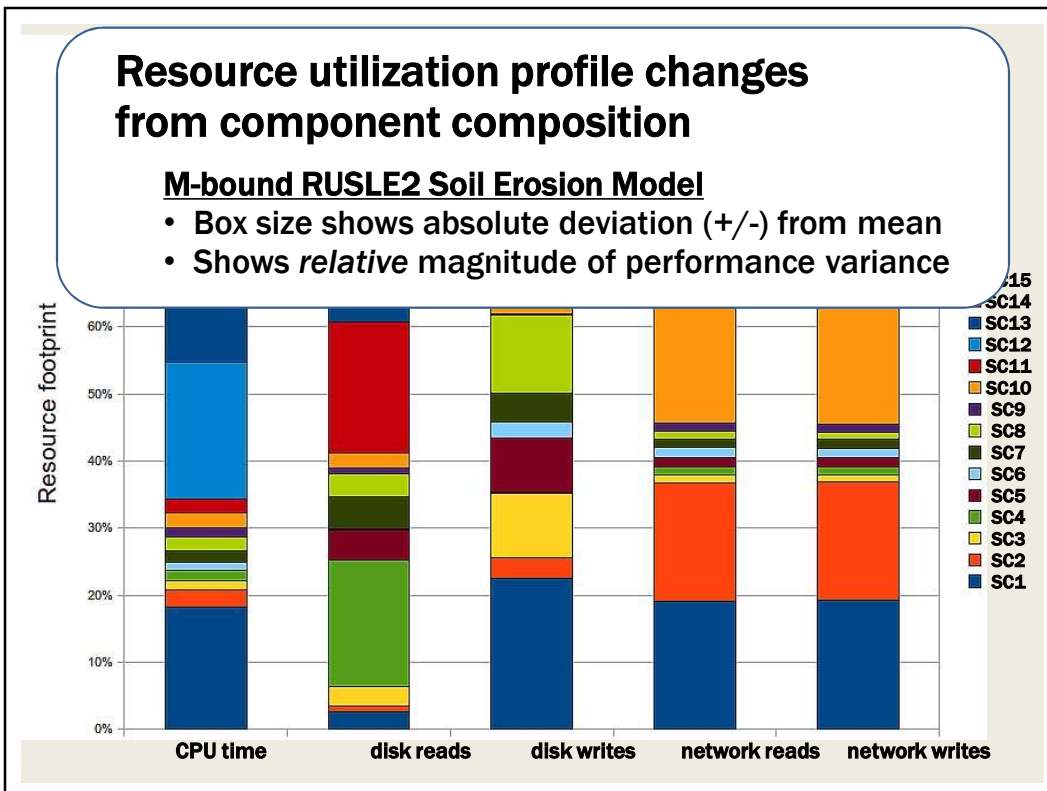
M: Tomcat ApplicationServer
D: Postgresql DB
F: nginx file server
L: Log server (Codebeamer) ³⁴

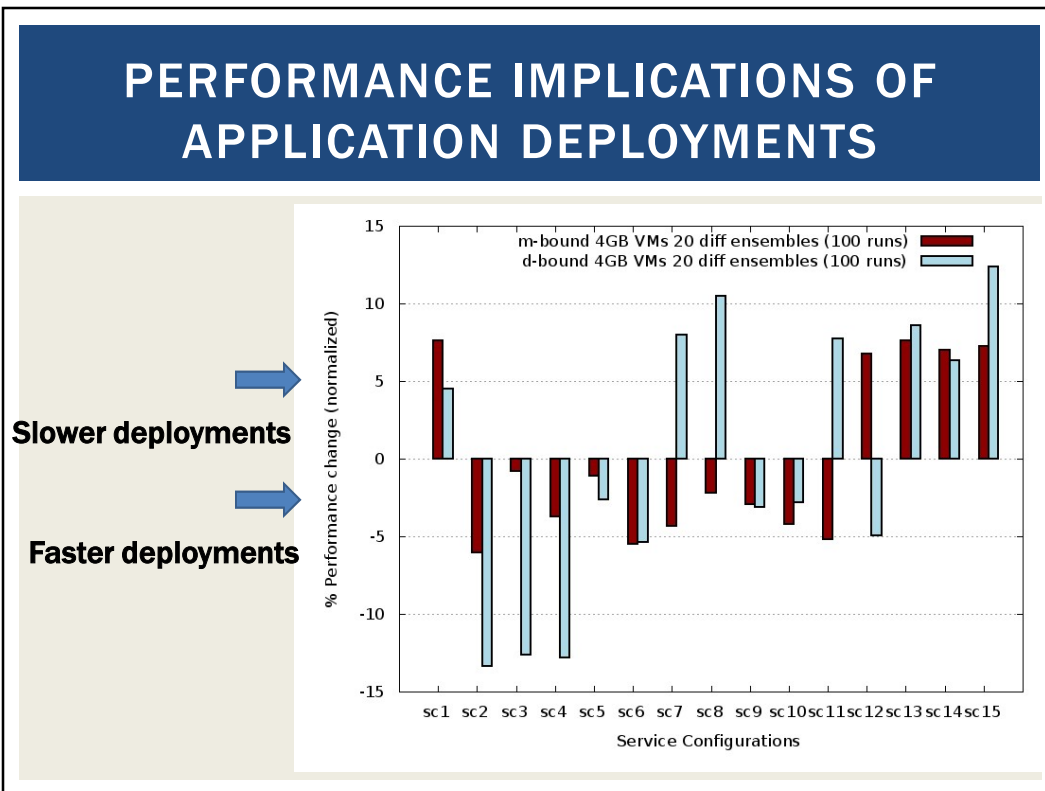
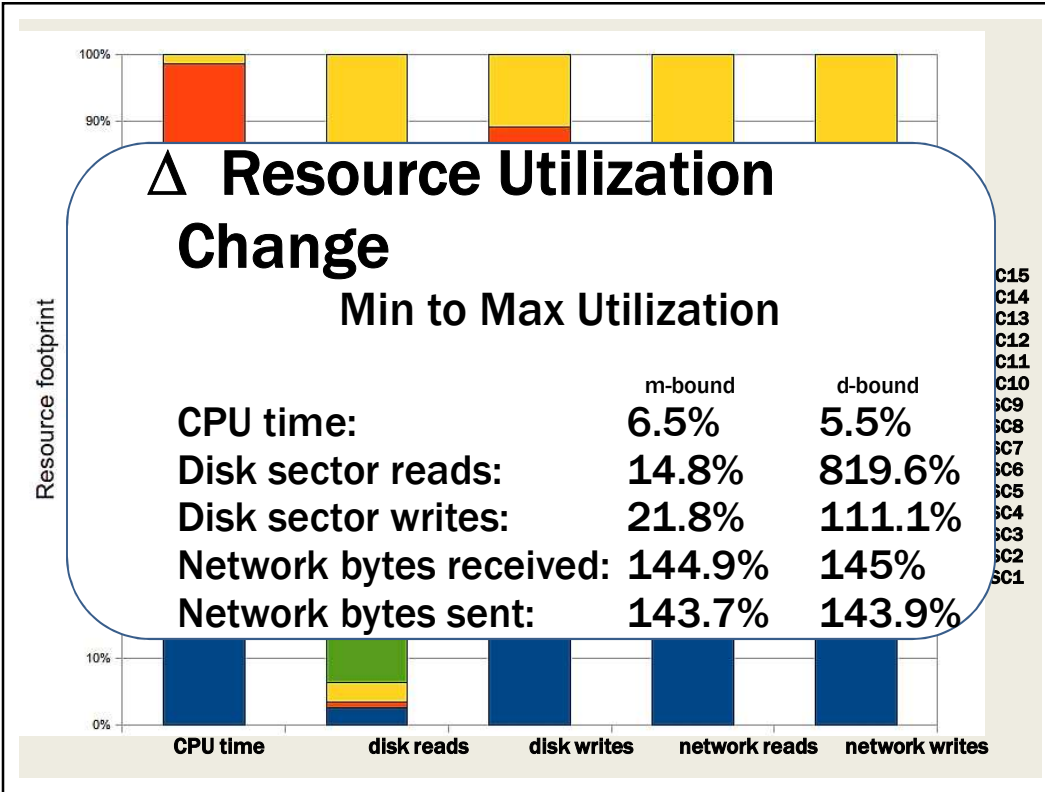


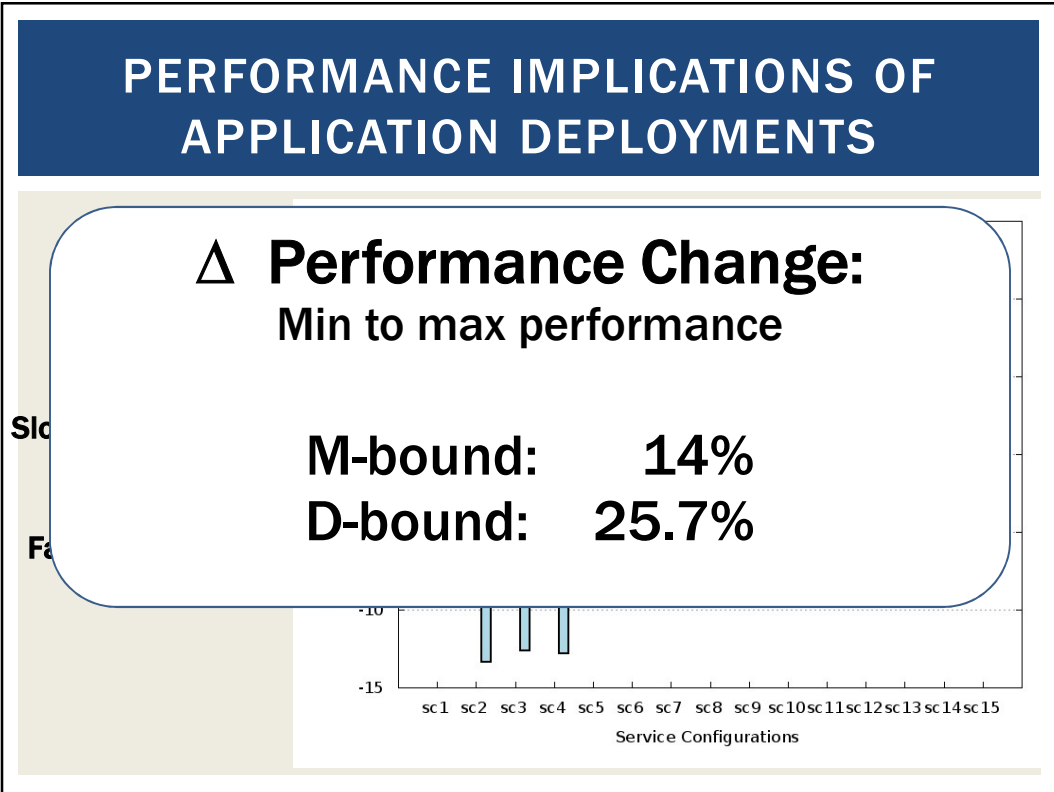
Resource utilization profile changes from component composition

M-bound RUSLE2 Soil Erosion Model

- Box size shows absolute deviation (+/-) from mean
- Shows *relative* magnitude of performance variance







PLATFORM-AS-A-SERVICE

- Predefined, ready-to-use, hosting environment
- Infrastructure is further obscured from end user
- Scaling and load balancing may be automatically provided and automatic
- Variable to no ability to influence responsiveness

■ Examples:

- Google App Engine
- Heroku
- AWS Elastic Beanstalk
- AWS Lambda (FaaS)

```
graph LR; C1[Client] --> LB[Load Balancer]; C2[Client] --> LB; C3[Client] --> LB; LB --> S1[Server]; LB --> S2[Server]; LB --> S3[Server];
```

March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.40

USES FOR PAAS

- **Cloud consumer**
 - **Wants to extend on-premise environments into the cloud for “web app” hosting**
 - **Wants to entirely substitute an on-premise hosting environment**
 - **Cloud consumer wants to become a cloud provider and deploy its own cloud services to external users**
- **PaaS spares IT administrative burden compared to IaaS**

March 28, 2018

TCS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.41

SERVERLESS COMPUTING

What is serverless?

Build and run applications without thinking about servers



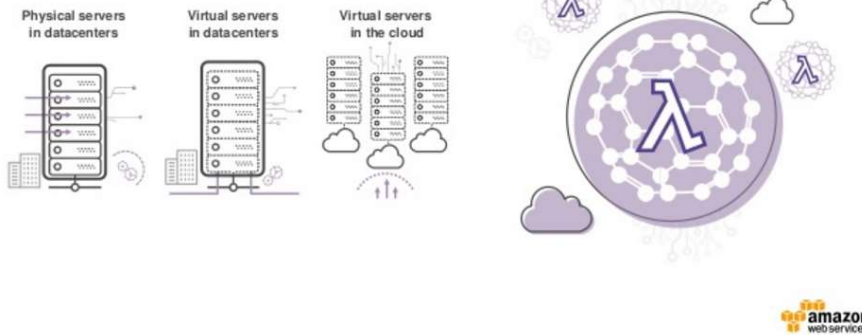
March 28, 2018

TCS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.42

SERVERLESS COMPUTING - 2

Evolving to serverless



March 28, 2018

TCCS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.43

SERVERLESS COMPUTING - 3

- New cloud platform for hosting application code
- Every cloud vendor provides their own:
 - AWS Lambda, Azure Functions, Google Cloud Functions, IBM OpenWhisk
- Similar to platform-as-a-service
- Replace open-source web container (e.g. Apache Tomcat) with abstracted vendor-provided **black-box** environment

March 28, 2018

TCCS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.44

SERVERLESS COMPUTING - 4

- Many challenging features of distributed systems are provided automatically
- **Built into the platform:**
- Highly availability (24/7)
- Scalability
- Fault tolerance

March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.45



SERVERLESS COMPUTING

AWS
Lambda
Demo

March 28, 2018

46

SERVERLESS SOFTWARE ARCHITECTURE

- Every service with a different pricing model

Example: Weather Application

Front-end code for weather app hosted in S3 User clicks on link to get local weather information App makes REST API call to endpoint Lambda is triggered Lambda runs code to retrieve local weather information and returns data back to user.

March 28, 2018	TCS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.47
----------------	--	-------

IAAS BILLING MODELS

- Virtual machines as-a-service at ¢ per hour
- No premium to scale:

$$= \begin{matrix} 1000 \text{ computers} & @ & 1 \text{ hour} \\ 1 \text{ computer} & @ & 1000 \text{ hours} \end{matrix}$$

- Illusion of infinite scalability to cloud user
- As many computers as you can afford
- Billing models are becoming increasingly granular
 - By the minute, second, 1/10th sec
- Auction-based instances:
Spot instances →

March 28, 2018	TCS562: Software Engineering for Cloud Computing [Spring 2018] Institute of Technology, University of Washington - Tacoma	L2.48
----------------	--	-------

QUESTIONS

March 28, 2018

TCSS562: Software Engineering for Cloud Computing [Spring 2018]
Institute of Technology, University of Washington - Tacoma

L2.49