

# TCSS 562: SOFTWARE ENGINEERING FOR CLOUD COMPUTING

## Cloud Enabling Technology II

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



1

## OFFICE HOURS – FALL 2023

- **THIS WEEK**
- Campus is closed Friday November 10, due to the Veteran’s Day holiday
- **Tuesdays:**
  - 2:30 to 3:30 pm - CP 229
- **\*\*\* Thursday \*\*\***
  - 6:00 pm to 7:00 pm – CP 229 and via Zoom
- Or email for appointment

> Office Hours set based on Student Demographics survey feedback

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.2
------------------	---	-------

2

## OBJECTIVES - 11/7

- **Questions from 11/2**
- Tutorials Questions
- Class Presentations:  
Cloud Technology or Research Paper Review
- Ch. 5: Cloud Enabling Technology

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.3
------------------	---	-------

3

## ONLINE DAILY FEEDBACK SURVEY

- Daily Feedback Quiz in Canvas - Take After Each Class
- Extra Credit for completing

- Announcements
- Assignments**
- Discussions
- Zoom
- Grades
- People
- Pages
- Files
- Quizzes
- Collaborations
- UW Libraries
- UW Resources

▼ Upcoming Assignments

- 📄 **Class Activity 1 - Implicit vs. Explicit Parallelism**  
Available until Oct 11 at 11:59pm | Due Oct 7 at 7:50pm | -/10 pts
- 📄 **Tutorial 1 - Linux**  
Available until Oct 19 at 11:59pm | Due Oct 15 at 11:59pm | -/20 pts

▼ Past Assignments

- 📄 **TCSS 562 - Online Daily Feedback Survey - 10/5**  
Available until Dec 18 at 11:59pm | Due Oct 6 at 8:59pm | -/1 pts
- 📄 **TCSS 562 - Online Daily Feedback Survey - 9/30**  
Available until Dec 18 at 11:59pm | Due Oct 4 at 8:59pm | -/1 pts

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.4
------------------	---	-------

4

TCSS 562 - Online Daily Feedback Survey - 10/5  
Started: Oct 7 at 1:13am  
Quiz Instructions

Question 1 0.5 pts

On a scale of 1 to 10, please classify your perspective on material covered in today's class:

1 2 3 4 5 6 7 8 9 10  
Mostly Review To Me Equal New and Review Mostly New to Me

Question 2 0.5 pts

Please rate the pace of today's class:

1 2 3 4 5 6 7 8 9 10  
Slow Just Right Fast

November 7, 2023 TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma L12.5

5

## MATERIAL / PACE

- Please classify your perspective on material covered in today's class (**54** respondents):
  - 1-mostly review, 5-equal new/review, 10-mostly new
  - **Average - 6.11** (↓ - *previous 6.23*)
- Please rate the pace of today's class:
  - 1-slow, 5-just right, 10-fast
  - **Average - 5.52** (↓ - *previous 5.77*)
- **Response rates:**
  - TCSS 462: 34/44 - 77.27%
  - TCSS 562: 20/25 - 80.00%

November 7, 2023 TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma L12.6

6

## FEEDBACK FROM 11/2

- **Can the API Gateway distribute different types of requests to different AWS Lambda functions using one endpoint URL?**
- **Yes** – for the resources, you can define two methods, where each method points to a different AWS Lambda function

[Delete](#) [Create method](#)

	Method type ▲	Integration type ▼	Authorization ▼	API key ▼
<input type="radio"/>	<a href="#">POST</a>	Lambda	None	Not required
<input type="radio"/>	<a href="#">PUT</a>	Lambda	None	Not required

November 7, 2023TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - TacomaL12.7

7

## TERM PROJECT PROPOSALS

- **18** Total term project proposals received
- **14** teams of 4
- **4** teams of 3
- **Round 1: 18** proposals reviewed thus far
  - **9** proposals accepted
  - **9** proposals – revisions requested
- **Application Use Cases (summary to be provided):**
  - **15** TLQ pipelines
  - **1** image generation (AI image generation model on ec2)
  - **1** NLP pipeline (sentiment analysis)
  - **1** serverless chatbot

November 7, 2023TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - TacomaL12.8

8

## AWS CLOUD CREDITS UPDATE

- AWS CLOUD CREDITS ARE NOW AVAILABLE FOR TCSS 462/562
- Credits provided on request with expiry of Sept 30, 2024
- Credit codes must be securely exchanged
- Request codes by sending an email with the subject "AWS CREDIT REQUEST" to [wllloyd@uw.edu](mailto:wllloyd@uw.edu)
- Codes can also be obtained in person (or zoom), in the class, during the breaks, after class, during office hours, by appt
  - 61 credit requests fulfilled as of Nov 6 @ 11:59p
- Codes not provided using discord

November 7, 2023

TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.9

9

## OBJECTIVES - 11/7

- Questions from 11/2
- **Tutorials Questions**
- Class Presentations:
  - Cloud Technology or Research Paper Review
- Ch. 5: Cloud Enabling Technology

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.10

10

## TUTORIAL 0

- Getting Started with AWS
- [http://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462\\_562\\_f2023\\_tutorial\\_0.pdf](http://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2023_tutorial_0.pdf)
- Create an AWS account
- Create account credentials for working with the CLI
- Install awsconfig package
- Setup awsconfig for working with the AWS CLI

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.11
------------------	---	--------

11

## TUTORIAL 4 – DUE NOV 7

- Introduction to AWS Lambda with the Serverless Application Analytics Framework (SAAF)
- [https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462\\_562\\_f2023\\_tutorial\\_4.pdf](https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2023_tutorial_4.pdf) (link to be posted)
- Obtaining a Java development environment
- Introduction to Maven build files for Java
- Create and Deploy “hello” Java AWS Lambda Function
  - Creation of API Gateway REST endpoint
- Sequential testing of “hello” AWS Lambda Function
  - API Gateway endpoint
  - AWS CLI Function invocation
- Observing SAAF profiling output
- Parallel testing of “hello” AWS Lambda Function with faas\_runner
- Performance analysis using faas\_runner reports
- Two function pipeline development task

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.12
------------------	---	--------

12

## TUTORIAL 5 – DUE NOV 14

- Introduction to Lambda II: Working with Files in S3 and CloudWatch Events
- [https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462\\_562\\_f2023\\_tutorial\\_5.pdf](https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2023_tutorial_5.pdf)
- Customize the Request object (add getters/setters)
  - Why do this instead of HashMap ?
- Import dependencies (jar files) into project for AWS S3
- Create an S3 Bucket
- Give your Lambda function(s) permission to work with S3
- Write to the CloudWatch logs
- Use of CloudTrail to generate S3 events
- Creating CloudWatch rule to capture events from CloudTrail
- Have the CloudWatch rule trigger a target Lambda function with a static JSON input object (hard-coded filename)
- **Optional:** for the S3 PutObject event, dynamically extract the name of the file put to the S3 bucket for processing

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.13

13

## TUTORIAL 6 – NOV 21

- Introduction to Lambda III: Serverless Databases
- [https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462\\_562\\_f2023\\_tutorial\\_6.pdf](https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2023_tutorial_6.pdf)
- Create and use Sqlite databases using sqlite3 tool
- Deploy Lambda function with Sqlite3 database under /tmp
- Compare in-memory vs. file-based Sqlite DBs on Lambda
- Create an Amazon Aurora “Serverless” v2 MySQL database
- Using an ec2 instance in the same VPC (Region + availability zone) connect and interact with the database using the mysql CLI app
- Deploy an AWS Lambda function that uses the MySQL “serverless” database

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.14

14

## OBJECTIVES - 11/7

- Questions from 11/2
- Tutorials Questions
- **Class Presentations:  
Cloud Technology or Research Paper Review**
- Ch. 5: Cloud Enabling Technology

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.15
------------------	---	--------

15

## GROUP PRESENTATION

- **TWO OPTIONS:**
- *Cloud technology presentation*
- *Cloud research paper presentation*
  - Recent & suggested papers will be posted at:  
<http://faculty.washington.edu/wlloyd/courses/tcss562/papers/>
- **Submit presentation type and topics (paper or technology) with desired dates of presentation via Canvas by:  
*Friday November 17<sup>th</sup> @ 11:59pm***
- **Presentation dates:**
  - Tuesday November 28, Thursday November 30
  - Tuesday December 5, Thursday December 7

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.16
------------------	---	--------

16



## OBJECTIVES - 11/7

- Questions from 11/2
- Tutorials Questions
- Class Presentations:  
Cloud Technology or Research Paper Review
- **Ch. 5: Cloud Enabling Technology**

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.17
------------------	---	--------

17

# CLOUD ENABLING TECHNOLOGY



18

## CLOUD ENABLING TECHNOLOGY

- Broadband networks and internet architecture
- Data center technology
- Virtualization technology
- Multitenant technology
- Web/web services technology

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.19
------------------	---	--------

19

## 3. VIRTUALIZATION TECHNOLOGY

- Convert a physical IT resource into a virtual IT resource
- Servers, storage, network, power (virtual UPSs)
- Virtualization supports:
  - Hardware independence
  - Server consolidation
  - Resource replication
  - Resource pooling
  - Elastic scalability
- Virtual servers
  - Operating-system based virtualization
  - Hardware-based virtualization

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.20
------------------	---	--------

20

## VIRTUAL MACHINES

- Emulation/simulation of a computer in software
- Provides a substitute for a real computer or server
- Virtualization platforms provide functionality to run an entire operating system
- Allows running multiple different operating systems, or operating systems with different versions simultaneously on the same computer


November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.21
------------------	---	--------

21

## KEY VIRTUALIZATION TRADEOFF

■ Tradeoff space:  
**What is the “right” level of abstraction in the cloud for sharing resources with users?**

***Degree of  
Hardware  
Abstraction***



**Abstraction  
Concerns:**

- Overhead
- Performance
- Isolation
- Security

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.22
------------------	---	--------

22

## ABSTRACTION CONCERNS

- **Overhead with too many instances w/ heavy abstractions**
  - Too many instances using a heavy abstraction can lead to hidden resource utilization and waste
  - Example: Dedicated server with 48 VMs each with separate instance of Ubuntu Linux
  - Idle VMs can reduce performance of co-resident jobs/tasks
- **“Virtualization” Overhead**
  - Cost of virtualization an OS instance
  - Overhead has dropped from ~100% to ~1% over last decade
- **Performance**
  - Impacted by weight of abstraction and virtualization overhead

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.23

23

## ABSTRACTION CONCERNS - 2

- **Isolation**
  - From others:  
What user A does should not impact user B in any noticeable way
- **Security**
  - User A and user B's data should be always separate
  - User A's actions are not perceivable by User B

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.24

24

## TYPES OF ABSTRACTION IN THE CLOUD

- **Virtual Machines** – original IaaS cloud abstraction
- **OS and Application Containers** – seen with CaaS
  - **OS Container** – replacement for VM, mimics full OS instance, heavier
  - OS containers run 100s of processes just like a VM
  - **App Container** – Docker: packages dependencies to easily transport and run an application anywhere
  - Application containers run only a few processes
- **Micro VMs** – FaaS / CaaS
  - Lighter weight alternative to full VM (KVM, XEN, VirtualBox)
  - Firecracker
- **Unikernel Operating Systems** – research mostly
  - Single process, multi-thread operating system
  - Designed for cloud, objective to reduce overhead of running too many OS instances

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.25

25

## VIRTUAL MACHINES

- **Type 1 hypervisor**
  - Typically involves a special virtualization kernel that runs directly on the system to share the underlying machine with many guest VMs
  - Paravirtualization introduced to directly share system resources with guests bypassing full emulation
  - VM becomes equal participant in sharing the network card for example
- **Type 2 hypervisor**
  - Typically involves the **Full Virtualization** of the guest, where everything is simulated/emulated
- **Hardware level support** (i.e. features introduced on CPUs) have made virtualization faster in all respects shrinking virtualization overhead

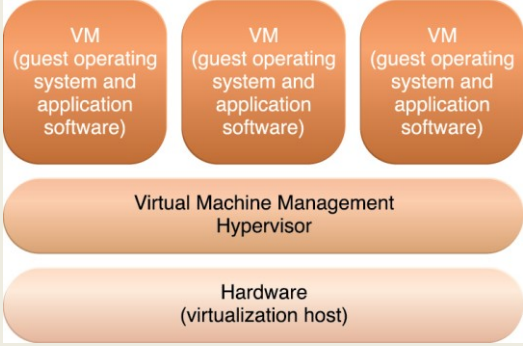
November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.26

26

## TYPE 1 HYPERVISOR



- Host OS and VMs run atop the hypervisor
- The boot OS is the hypervisor kernel
- Xen dom0

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.27
------------------	---	--------

27

## TYPE 1 HYPERVISOR

- Acts as a control program
- Miniature OS kernel that manages VMs
- Boots and runs on bare metal
- Also known as Virtual Machine Monitor (VMM)
- **Paravirtualization:** Kernel includes I/O drivers
- VM guest Oses must use special kernel to interoperate
- Paravirtualization provides hooks to the guest VMs
- Kernel traps instructions (i.e. device I/O) to implement sharing & multiplexing
- User mode instructions run directly on the CPU
- Objective: minimize virtualization overhead
- Classic example is XEN (dom0 kernel)

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.28
------------------	---	--------

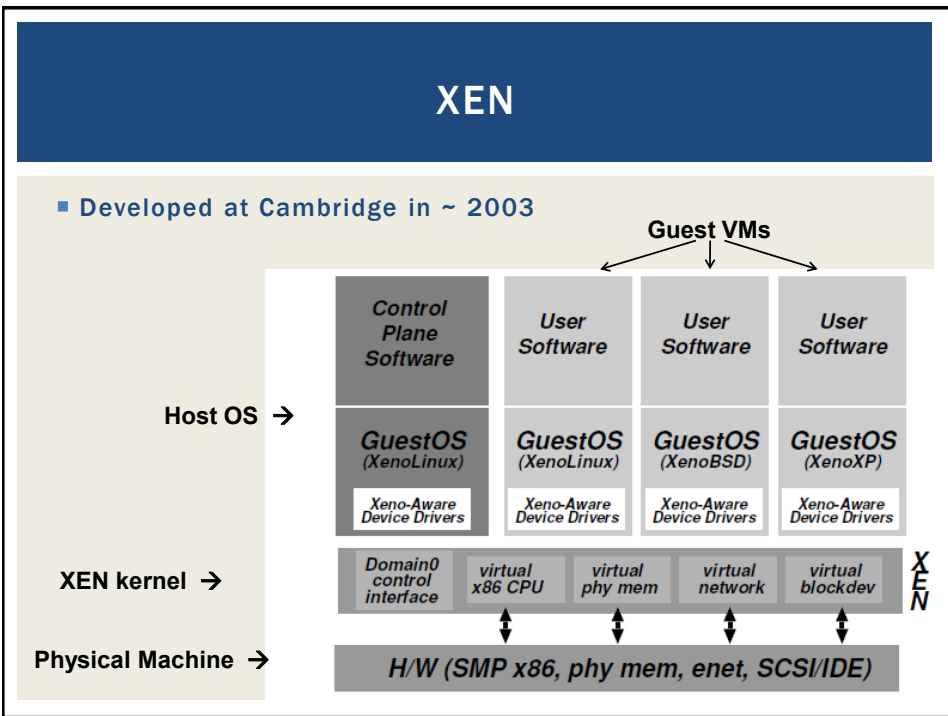
28

## COMMON VMMS: PARAVIRTUALIZATION

- **TYPE 1 Hypervisor**
- XEN
- Citrix Xen-server (a commercial version of XEN)
- VMWare ESXi
- KVM (virtualization support in kernel)
  
- Paravirtual I/O drivers introduced
  - XEN
  - KVM
  - Virtualbox

November 7, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma
L12.29

29



30

## XEN - 2

- VMs managed as “domains”
- Domain 0 is the hypervisor domain
  - Host OS is installed to run on bare-metal, but doesn't directly facilitate virtualization (*unlike KVM*)
- Domains 1..n are guests (VMs) – not bare-metal

```

xentop - 17:53:48 Xen 3.1.2-398.e15
3 domains: 1 running, 2 blocked, 0 paused, 0 crashed, 0 dying, 0 shutdown
Mem: 8379564k total, 8377876k used, 1688k free CPUs: 4 @ 2400MHz

```

NAME	STATE	CPU(sec)	CPU(%)	MEM(k)	MEM(%)	MAXMEM(k)	MAXMEM(%)	VCPUS
NETS	NETIX(k)	NETRX(k)	VBDS	VBD OO	VBD RD	VBD WR	SSID	
centos	--b---	46	0.0	532352	6.4	1064960	12.7	1
1	27960	885	1	0	6313	37119	0	
centos-2	--b---	17	0.0	1056640	12.6	2113536	25.2	1
1	50	0	1	0	3981	541	0	
Domain-0	-----r	2979	19.3	6568960	78.4	no limit	n/a	4
4	1057374	290072	0	0	0	0	0	

November 7, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma
L12.31

31

## XEN - 3

- Physical machine boots special XEN kernel
- Kernel provides paravirtual API to manage CPU & device multiplexing
- Guests require modified XEN-aware kernels
- Xen supports full-virtualization for unmodified OS guests in hvm mode
- Amazon EC2 largely based on modified version of XEN hypervisor (EC2 gens 1-4)
- XEN provides its own CPU schedulers, I/O scheduling

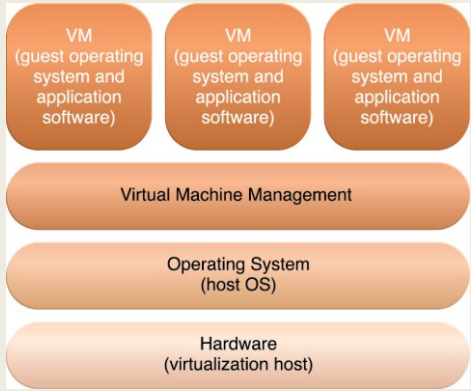
November 7, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma
L12.32

32



## TYPE 2 HYPERVISOR

- Adds additional layer



The diagram illustrates the architecture of a Type 2 Hypervisor. It consists of three layers of software stacked on top of hardware. At the top, there are three separate Virtual Machines (VMs), each containing a guest operating system and application software. Below these VMs is a layer for Virtual Machine Management. This layer sits on top of the Operating System (host OS), which is installed on the Hardware (virtualization host).

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.33
------------------	---	--------

33

## TYPE 2 HYPERVISOR

- **Problem: Original x86 CPUs could not trap special instructions**
- **Instructions not specially marked**
- **Solution: Use Full Virtualization**
- **Trap ALL instructions**
- **“Fully” simulate entire computer**
- **Tradeoff: Higher Overhead**
- **Benefit: Can virtualize any operating system without modification**

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.34
------------------	---	--------

34

## CHECK FOR VIRTUALIZATION SUPPORT

- See:  
<https://cyberciti.biz/faq/linux-xen-vmware-kvm-intel-vt-amd-v-support>
- # check for Intel VT CPU virtualization extensions on Linux  
`grep -color vmx /proc/cpuinfo`
- # check for AMD V CPU virtualization extensions on Linux  
`grep -color svm /proc/cpuinfo`
- Also see 'lscpu' → "Virtualization:"
- Other Intel CPU features that help virtualization:  
`ept vpid tpr_shadow flexpriority vnmi`

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.35

35

## KERNEL BASED VIRTUAL MACHINES (KVM)

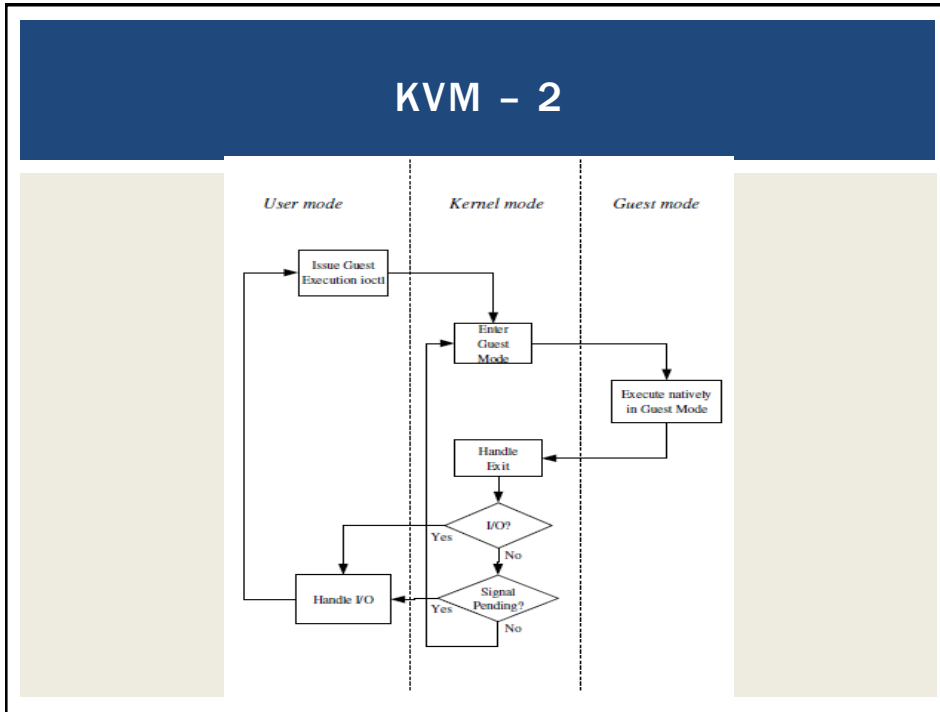
- x86 HW notoriously difficult to virtualize
- Extensions added to 64-bit Intel/AMD CPUs
  - Provides hardware assisted virtualization
  - New "guest" operating mode
  - Hardware state switch
  - Exit reason reporting
  - Intel/AMD implementations different
    - Linux uses vendor specific kernel modules

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.36

36



37

## KVM - 3

- **KVM has /dev/kvm device file node**
  - **Linux character device, with operations:**
    - Create new VM
    - Allocate memory to VM
    - Read/write virtual CPU registers
    - Inject interrupts into vCPUs
    - Running vCPUs
- **VMs run as Linux processes**
  - Scheduled by host Linux OS
  - Can be pinned to specific cores with “taskset”

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.38
------------------	---	--------

38

## KVM PARAVIRTUALIZED I/O

- KVM – Virtio
  - Custom Linux based paravirtual device drivers
  - Supersedes QEMU hardware emulation (full virt.)
  - Based on XEN paravirtualized I/O
  - Custom block device driver provides paravirtual device emulation
    - Virtual bus (memory ring buffer)
    - Requires hypercall facility
    - Direct access to memory

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.39

39

## KVM DIFFERENCES FROM XEN

- KVM requires CPU VMX support
  - Virtualization management extensions
- KVM can virtualize any OS without special kernels
  - Less invasive
- KVM was originally separate from the Linux kernel, but then integrated
- KVM is type 1 hypervisor because the machine boots Linux which has integrated support for virtualization
- Different than XEN because XEN kernel alone is not a full-fledged OS

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.40

40

## KVM ENHANCEMENTS

- Paravirtualized device drivers
  - Virtio
- Guest Symmetric Multiprocessor (SMP) support
  - Leverages multiple on-board CPUs
  - Supported as of Linux 2.6.23
- VM Live Migration
- Linux scheduler integration
  - Optimize scheduler with knowledge that KVM processes are virtual machines

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.41
------------------	---	--------

41

## FIRECRACKER MICRO VM

*From <https://firecracker-microvm.github.io/>*

The following diagram depicts an example host running Firecracker microVMs.

The diagram illustrates the Firecracker microVM architecture. On the left, a stack of microVMs is shown within the **HOST KERNEL SPACE** of a **BARE METAL SERVER**. The host space includes **KVM** and **I/O** components. A callout states: "Firecracker scales to thousands of multitenant microVMs." and "Configurable microVMs across CPU and memory, running as user space processes." On the right, a detailed view of a microVM shows its internal structure. It is divided into a **CONTROL PLANE** (top) and a **DATA PLANE** (bottom). A **VIRTUALIZATION BARRIER** separates the control plane from the data plane, and a **JAILER BARRIER** separates the data plane from the host. The **CONTROL PLANE** contains a **RESTful API** and **Rate limiting**. The **DATA PLANE** contains **Network**, **Storage**, and **Metadata Service**. A **Client** is shown interacting with the RESTful API. The top of the data plane is labeled **Guest OS & Container workload**.

42

## FIRECRACKER MICRO VM

- Provides a virtual machine monitor (VMM) (i.e. hypervisor) using KVM to create and manage microVMs
- Has a minimalist design with goals to improve security, decreases the startup time, and increases hardware utilization
- Excludes unnecessary devices and guest functionality to reduce memory footprint and attack surface area of each microVM
- Supports boot time of <125ms, <5 MiB memory footprint
- Can run 100s of microVMs on a host, launching up to 150/sec
- Is available on 64-bit Intel, AMD, and Arm CPUs
- Used to host AWS Lambda and AWS Fargate
- Has been open sourced under the Apache 2.0 license

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.43

43

## FIRECRACKER - 2

- **Minimalistic**
- MicroVMs run as separate processes on the host
- Only 5 emulated devices are available: virtio-net, virtio-block, virtio-vsock, serial console, and a minimal keyboard controller used only to stop the microVM
- Rate limiters can be created and configured to provision resources to support bursts or specific bandwidth/operation limitations
- **Configuration**
- A RESTful API enables common actions such as configuring the number of vCPUs or launching microVMs
- A metadata service between the host and guest provides configuration information

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.44

44

## FIRECRACKER - 2


- **Security**
- Runs in user space (***not the root user***) on top of the Linux Kernel-based Virtual Machine (KVM) hypervisor to create microVMs
- Lambda functions, Fargate containers, or container groups can be encapsulated using Firecracker through KVM, enabling workloads from different customers to run on the same machine, without sacrificing security or efficiency
- MicroVMs are further isolated with common Linux user-space security barriers using a companion program called “jailer” which provides a second line of defense if KVM is compromised

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.45
------------------	---	--------

45

## UNIKERNELS

- Lightweight alternative to containers and VMs
  - Custom Cloud Operating System
  - Single process, multiple threads, runs one program
  - Launch separately atop of hypervisor (XEN/KVM)
  - Reduce overhead, duplication of heavy weight OS
- OSv is most well known unikernel
- Several others exist has research projects
- More information at: <http://unikernel.org/>
- Google Trends OSv →



November 7, 2023	TCSS462/562 School of Eng	
------------------	------------------------------	--

46

**WE WILL RETURN AT  
~4:50 PM**



47

## VIRTUALIZATION MANAGEMENT

- Virtual infrastructure management (VIM) tools
- Tools that manage pools of virtual machines, resources, etc.
- Private cloud software systems can be considered as a VIM
  
- Considerations:
- Performance overhead
  - Paravirtualization: custom OS kernels, I/O passed directly to HW w/ special drivers
- Hardware compatibility for virtualization
- Portability: virtual resources tend to be difficult to migrate cross-clouds

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.48
------------------	---	--------

48



## VIRTUAL INFRASTRUCTURE MANAGEMENT (VIM)

- **Middleware to manage virtual machines and infrastructure of IaaS “clouds”**
  
- **Examples**
  - **OpenNebula**
  - **Nimbus**
  - **Eucalyptus**
  - **OpenStack**

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.49
------------------	---	--------

49

## VIM FEATURES

- **Create/destroy VM Instances**
- **Image repository**
  - **Create/Destroy/Update images**
  - **Image persistence**
  
- **Contextualization of VMs**
  - **Networking address assignment**
    - **DHCP / Static IPs**
  - **Manage SSH keys**

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.50
------------------	---	--------

50

## VIM FEATURES - 2

- Virtual network configuration/management
  - Public/Private IP address assignment
  - Virtual firewall management
  - Configure/support isolated VLANs (private clusters)
  
- Support common virtual machine managers (VMMs)
  - XEN, KVM, VMware
  - Support via libvirt library

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.51
------------------	---	--------

51

## VIM FEATURES - 3

- Shared “Elastic” block storage
  - Facility to create/update/delete VM disk volumes
    - Amazon EBS
    - Eucalyptus SC
    - OpenStack Volume Controller

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.52
------------------	---	--------

52

## CONTAINER ORCHESTRATION FRAMEWORKS

- Middleware to manage Docker application container deployments across virtual clusters of Docker hosts (VMs)
- Considered Infrastructure-as-a-Service
  
- **Opensource**
  - Kubernetes framework
  - Docker swarm
  - Apache Mesos/Marathon
  
- **Proprietary**
  - Amazon Elastic Container Service

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.53
------------------	---	--------

53

## CONTAINER SERVICES

- **Public cloud container cluster services**
  - Azure Kubernetes Service (AKS)
  - Amazon Elastic Container Service for Kubernetes (EKS)
  - Google Kubernetes Engine (GKE)
  
- **Container-as-a-Service**
  - Azure Container Instances (ACI - April 2018)
  - AWS Fargate (November 2017)
  - Google Kubernetes Engine Serverless Add-on (alpha-July 2018)

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.54
------------------	---	--------

54

## CLOUD ENABLING TECHNOLOGY


- Adapted from Ch. 5 from *Cloud Computing Concepts, Technology & Architecture*
- Broadband networks and internet architecture
- Data center technology
- Virtualization technology
- Multitenant technology
- Web/web services technology

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.55
------------------	---	--------

55

## 4. MULTITENANT APPLICATIONS

- Each tenant (like in an apartment) has their own view of the application
- Tenants are unaware of their neighbors
- Tenants can only access their data, no access to data and configuration that is not their own
- Customizable features
  - UI, business process, data model, access control
- Application architecture
  - User isolation, data security, recovery/backup by tenant, scalability for a tenant, for tenants, metered usage, data tier isolation



November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.56
------------------	---	--------

56

## MULTITENANT APPS - 2

- Forms the basis for SaaS (applications)

The diagram illustrates a multi-tenant architecture. At the top, two separate boxes represent 'Organization A' and 'Organization B'. Inside each organization box is a 'cloud service consumer'. Arrows from these consumers point down to a central cloud icon. Inside the cloud, there is a 'multitenant application' (represented by a grid of squares) and a 'hosting virtual server' (represented by a stack of blocks). A dashed line connects the multitenant application and the hosting virtual server, indicating their interaction.

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.57
------------------	---	--------

57

## CLOUD ENABLING TECHNOLOGY

- Adapted from Ch. 5 from *Cloud Computing Concepts, Technology & Architecture*
- Broadband networks and internet architecture
- Data center technology
- Virtualization technology
- Multitenant technology
- Web/web services technology**

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.58
------------------	---	--------

58

## 5. WEB SERVICES/WEB

- Web services technology is a key foundation of cloud computing’s “**as-a-service**” cloud delivery model
- SOAP - “Simple” object access protocol
  - First generation web services
  - WSDL - web services description language
  - UDDI - universal description discovery and integration
  - SOAP services have their own unique interfaces
- REST - instead of defining a custom technical interface REST services are built on the use of HTTP protocol
- HTTP GET, PUT, POST, DELETE

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.59
------------------	---	--------

59

## HYPERTEXT TRANSPORT PROTOCOL (HTTP)

- An ASCII-based request/reply protocol for transferring information on the web
- HTTP request includes:
  - request method (GET, POST, etc.)
  - Uniform Resource Identifier (URI)
  - HTTP protocol version understood by the client
  - headers—extra info regarding transfer request
- HTTP response from server
  - Protocol version & status code →
  - Response headers
  - Response body

**HTTP status codes:**

2xx — *all is well*

3xx — *resource moved*

4xx — *access problem*

5xx — *server error*

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.60
------------------	---	--------

60

## REST: REPRESENTATIONAL STATE TRANSFER

- Web services protocol
- *Supersedes SOAP* – Simple Object Access Protocol
- Access and manipulate web resources with a predefined set of stateless operations (known as web services)
- Requests are made to a URI
- Responses are most often in JSON, but can also be HTML, ASCII text, XML, no real limits as long as text-based
- HTTP verbs: GET, POST, PUT, DELETE, ...

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.61

61

```
// SOAP REQUEST

POST /InStock HTTP/1.1
Host: www.bookshop.org
Content-Type: application/soap+xml; charset=utf-8
Content-Length: nnn

<?xml version="1.0"?>
<soap:Envelope
xmlns:soap="http://www.w3.org/2001/12/soap-envelope"
soap:encodingStyle="http://www.w3.org/2001/12/soap-
encoding">
<soap:Body xmlns:m="http://www.bookshop.org/prices">
  <m:GetBookPrice>
    <m:BookName>The Fleamarket</m:BookName>
  </m:GetBookPrice>
</soap:Body>
</soap:Envelope>
```

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.62

62

```
// SOAP RESPONSE
POST /InStock HTTP/1.1
Host: www.bookshop.org
Content-Type: application/soap+xml; charset=utf-8
Content-Length: nnn

<?xml version="1.0"?>
<soap:Envelope
xmlns:soap="http://www.w3.org/2001/12/soap-envelope"
soap:encodingStyle="http://www.w3.org/2001/12/soap-
encoding">
<soap:Body xmlns:m="http://www.bookshop.org/prices">
  <m:GetBookPriceResponse>
    <m:Price>10.95</m:Price>
  </m:GetBookPriceResponse>
</soap:Body>
</soap:Envelope>
```

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.63

63

```
// WSDL Service Definition
<?xml version="1.0" encoding="UTF-8"?>
<definitions name="DayOfWeek"
targetNamespace="http://www.roguewave.com/soapworx/examples/DayOfWeek.wsdl"
xmlns:tns="http://www.roguewave.com/soapworx/examples/DayOfWeek.wsdl"
xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns="http://schemas.xmlsoap.org/wsdl/">
  <message name="DayOfWeekInput">
    <part name="date" type="xsd:date"/>
  </message>
  <message name="DayOfWeekResponse">
    <part name="dayOfWeek" type="xsd:string"/>
  </message>
  <portType name="DayOfWeekPortType">
    <operation name="GetDayOfWeek">
      <input message="tns:DayOfWeekInput"/>
      <output message="tns:DayOfWeekResponse"/>
    </operation>
  </portType>
  <binding name="DayOfWeekBinding" type="tns:DayOfWeekPortType">
    <soap:binding style="document"
transport="http://schemas.xmlsoap.org/soap/http"/>
    <operation name="GetDayOfWeek">
      <soap:operation soapAction="getdayofweek"/>
      <input>
        <soap:body use="encoded"
namespace="http://www.roguewave.com/soapworx/examples"
encodingStyle="http://schemas.xmlsoap.org/soap/encoding"/>
      </input>
      <output>
        <soap:body use="encoded"
namespace="http://www.roguewave.com/soapworx/examples"
encodingStyle="http://schemas.xmlsoap.org/soap/encoding"/>
      </output>
    </operation>
  </binding>
  <service name="DayOfWeekService" >
    <documentation>
      Returns the day-of-week name for a given date
    </documentation>
    <port name="DayOfWeekPort" binding="tns:DayOfWeekBinding">
      <soap:address location="http://localhost:8090/dayofweek/DayOfWeek"/>
    </port>
  </service>
</definitions>
```

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.64

64



## REST CLIMATE SERVICES EXAMPLE

```
■ USDA // REST/JSON
Lat/Long // Request climate data for Washington
Climate
Service {
Service "parameter": [
Demo {
"parameter": [
{
"name": "latitude",
"value": 47.2529
},
{
"name": "longitude",
"value": -122.4443
}
]
}
}
}
```

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.65

65

## REST - 2

- App manipulates one or more types of resources.
- Everything the app does can be characterized as some kind of operation on one or more resources.
- Frequently services are CRUD operations (create/read/update/delete)
  - Create a new resource
  - Read resource(s) matching criterion
  - Update data associated with some resource
  - Destroy a particular a resource
- Resources are often implemented as objects in OO languages

November 7, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L12.66

66


## REST ARCHITECTURAL ADVANTAGES

- **Performance:** component interactions can be the dominant factor in user-perceived performance and network efficiency
- **Scalability:** to support large numbers of services and interactions among them
- **Simplicity:** of the Uniform Interface
- **Modifiability:** of services to meet changing needs (even while the application is running)
- **Visibility:** of communication between services
- **Portability:** of services by redeployment
- **Reliability:** resists failure at the system level as redundancy of infrastructure is easy to ensure

November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.67
------------------	---	--------

67

## QUESTIONS



November 7, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L12.68
------------------	---	--------

68