

TCSS 562:
SOFTWARE ENGINEERING
FOR CLOUD COMPUTING

AWS OVERVIEW &
Cloud Enabling Technology

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



1

OFFICE HOURS – FALL 2025

■ **Thursdays:**
■ 6:00 to 7:00 pm - CP 229 & Zoom

■ **Friday – *** THIS WEEK *****
■ 4:30 pm to 5:30 pm – ONLINE via Zoom

■ Or email for appointment

➢ Office Hours set based on Student Demographics survey feedback

➢ * - Friday office hours may be adjusted or canceled due meeting conflicts or other obligations. Adjustments will be announced via Canvas.

November 13, 2025

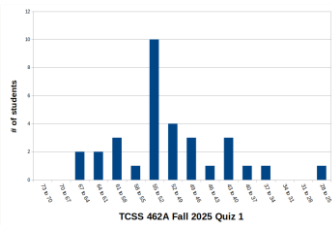
TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington – Tacoma

L13.2

2

QUIZ 1 RESULTS – TCSS 462

■ **Statistics (out of 73)**
■ Average: 50.84
■ Mode: 52
■ Median: 52
■ Min score: 26
■ Lower quartile: 48
■ 2nd quartile: 52
■ 3rd quartile: 54
■ Max score: 65
■ Standard deviation: 8.65
■ Curve: +12.0
■ Missing: 1



TCSS 462A Fall 2025 Quiz 1

April 29, 2025

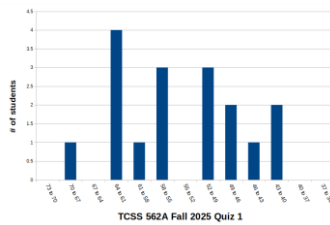
TCSS422: Computer Operating Systems [Spring 2025]
School of Engineering and Technology, University of Washington – Tacoma

L9.3

3

QUIZ 1 RESULTS – TCSS 562

■ **Statistics (out of 73)**
■ Average: 53.71
■ Mode: 61
■ Median: 55
■ Min score: 40
■ Lower quartile: 48
■ 2nd quartile: 55
■ 3rd quartile: 61
■ Max score: 68
■ Standard deviation: 7.95
■ Curve: +9.0
■ Missing: 0



TCSS 562A Fall 2025 Quiz 1

April 29, 2025

TCSS422: Computer Operating Systems [Spring 2025]
School of Engineering and Technology, University of Washington – Tacoma

L9.4

4

OBJECTIVES – 11/13

■ **Questions from 11/6**

■ Tutorials Questions

■ Class Presentations:
Cloud Technology or Research Paper Review

■ Ch. 5: Cloud Enabling Technology

■ Tutorial 5 Demo

November 13, 2025

TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington – Tacoma

L13.5

5

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 13 at 11:59pm | Due Oct 7 at 7:59pm | -10 pts

Tutorial 1 - Linux
Available until Oct 19 at 11:59pm | Due Oct 13 at 11:59pm | -10 pts

Past Assignments

TCSS 562 - Online Daily Feedback Survey - 10/5
Available until Dec 18 at 11:59pm | Due Oct 6 at 8:59pm | -13 pts

TCSS 562 - Online Daily Feedback Survey - 9/30
Available until Dec 18 at 11:59pm | Due Oct 4 at 8:59pm | -13 pts

November 13, 2025

TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington – Tacoma

L13.6

6

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 13, 2025

TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.7

7

MATERIAL / PACE

Please classify your perspective on material covered in today's class (50 respondents **100%!!**, 46 in-person, 4 online):

1-mostly review, 5-equal new/review, 10-mostly new

Average – **5.64** (↓ - previous **6.36**)

Please rate the pace of today's class:

1-slow, 5-just right, 10-fast

Average – **4.96** (↓ - previous **5.19**)

November 13, 2025

TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.8

8

FEEDBACK FROM 11/6

Regarding a Cloud Technology Class Presentation, is a demonstration required?

Yes, a demonstration is required for a cloud technology presentation, though a LIVE demonstration is **not** required

The Assignment description provides some strategies

Will Quiz 2 be similar to quiz 1 in terms of format, difficulty and/or scope?

It is likely to be similar

Quiz 2 will have less new content, and more review, since there are just 4 lectures before December 2nd

Lectures 12, 13, 14, 15, 16

Tutorials 5, 6, 7

November 13, 2025

TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.9

9

OBJECTIVES – 11/13

Questions from 11/6

Tutorials Questions

Class Presentations:
Cloud Technology or Research Paper Review

Ch. 5: Cloud Enabling Technology

Tutorial 5 Demo

November 13, 2025

TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.10

10

Don't Forget to Terminate (Shutdown)
all EC2 instances for Tutorials 3

Spot instances:
c5d.large instance @ ~3.2 cents / hour

\$0.78 / day
\$5.48 / week
\$23.78 / month
\$285.42 / year

AWS CREDITS → → → → → → → →



11

TUTORIAL 4 – DUE NOV 11

Introduction to AWS Lambda with the Serverless Application Analytics Framework (SAAF)
https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2025_tutorial_4.pdf

Set up Java development environment

Introduction to Maven build files for Java

Create and Deploy "hello" Java AWS Lambda Function

Create API Gateway REST endpoint

Sequential testing of "hello" AWS Lambda Function

- API Gateway endpoint, AWS Lambda CLI Function invocation, AWS Function URL

Profiling function performance with SAAF

Concurrent function testing with faas_runner

Performance analysis using faas_runner reports

Two function pipeline development task: Caesar Cipher

November 13, 2025

TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.12

12

TUTORIAL 4 - NOTES

- Sometimes students get unexpected results when adding **Thread.sleep(10000)** to the hello Lambda function:
- It is possible that:
 - Results from the SAAF Report Generator were from a test run before the **Thread.Sleep()** statement was added to the code
 - OR -
 - The **Thread.Sleep()** statement was added in the incorrect location of the code
 - OR -
 - When opening the CSV output from the Report Generator, the file separator characters were set incorrectly.
- The only separator for a CSV file is the comma ",". Be sure to correctly open the CSV file in the spreadsheet. Columns can be offset resulting in the wrong answers being provided for Question 6.

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.13

13

TUTORIAL 4 – NOTES - 2

- The sleep statement must go between the START FUNCTION and END FUNCTION comments in the `handleRequest()` method specified as the AWS Lambda function's handler under runtime settings in the AWS Lambda GUI.

```
//*****START FUNCTION IMPLEMENTATION*****
try
{
    Thread.sleep(10000);
}
catch (InterruptedException ie)
{
    System.out.println("Interruption occurred while sleeping.");
}
//*****END FUNCTION IMPLEMENTATION*****
```
- Code must be recompiled and redeployed after modification

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.14

14

TUTORIAL 4 – NOTES - 3

- SANITY CHECK:** consider that adding 10 seconds of sleep to your AWS Lambda function will cause the function to run for at least 10 seconds. This will impact the outputs requested for Question 6:
- avg_runtime** is the server-side (cloud) runtime of the function
- This is the time it takes for the function to run on AWS Lambda (cloud)
- Adding sleep of 10 seconds should increase a function's **avg_runtime**
- avg_roundTripTime** is the total time for a request from a client (laptop?) to travel to the server (cloud), make the function call, and return.
- If trying to make 50 calls at once on a laptop with a small # of CPU cores this time may be slow
- Adding sleep of 10 seconds should increase a function's **avg_roundTripTime**

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.15

15

TUTORIAL 4 – NOTES - 4

- avg_cpulidleDelta** time is the amount of time the Lambda function's Firecracker two vCPUs are idle during the function call on the server measured in centiseconds:

100 centiseconds = 1 second
100 centiseconds = 1000 milliseconds
- By default, AWS Lambda functions with 512 MB run in a runtime environment with access to two vCPU cores
- This is the total vCPU idle time for both cores (it is doubled)
- Adding sleep of 10 seconds should increase your function's **avg_cpulidleDelta**
- How much should **avg_cpulidleDelta** increase ?

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.16

16

TUTORIAL 5 – DUE NOV 16

- Introduction to Lambda II: Working with Files in S3, Cloud Trail, and Amazon Event Bridge Rules
- https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2025_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 Event Bridge rule to capture events from CloudTrail
- Have the Event Bridge rule trigger a 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 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.17

17

TUTORIAL 6 – NOV 23

- Introduction to Lambda III: Serverless Databases
- https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2025_tutorial_6.pdf
- Create and use Sqlite databases using `sqlite3`
- 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 the AWS CloudShell in the same VPC (Region + availability zone) connect and interact your Aurora serverless database using the `mysql` CLI app
- Deploy an AWS Lambda function that uses the MySQL "serverless" database

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.18

18

TUTORIAL 7

- Introduction to Docker
- To be posted
- Must complete using Ubuntu 24.04 (for cgroups v2)
- Use DOCX file for copying and pasting Docker install commands
- Topics:
 - Installing Docker
 - Creating a container using a Dockerfile
 - Using cgroups virtual filesystem to monitor CPU utilization of a container
 - Persisting container images to Docker Hub image repository
 - Container vertical scaling of CPU/memory resources
 - Testing container CPU and memory isolation

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.19

19

TUTORIAL COVERAGE

- Docker CLI → Docker Engine (dockerd) → containerd → runc
- Working with the docker CLI:
 - docker run create a container
 - docker ps -a list containers, find CONTAINER ID
 - docker exec --it run a process in an existing container
 - docker stop stop a container
 - docker kill kill a container
 - docker help list available commands
 - man docker Docker Linux manual pages

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.20

20

Docker CLI

```
attach      Attach local standard input, output, and error streams to a running container
build       Build an image from a Dockerfile
commit      Create a new image from a container's changes
cp          Copy files/folders between a container and the local filesystem
create      Create a new container
deploy      Deploy a new stack or update an existing stack
diff        Inspect changes to files or directories on a container's filesystem
events      Get real time events from the server
exec        Run a command in a running container
export      Export a container's filesystem as a tar archive
history      Show the history of an image
images      List images
import      Import the contents from a tarball to create a filesystem image
inspect     Display system-wide information
info        Return low-level information on Docker objects
kill        Kill one or more running containers
load        Load an image from a tar archive or STDIN
login       Log in to a Docker registry
logout      Log out from a Docker registry
logs        Fetch the logs of a container
pause       Pause all processes within one or more containers
port        List port mappings or a specific mapping for the container
ps          List containers
pull        Pull an image or a repository from a registry
push        Push an image or a repository to a registry
rename      Rename a container
restart     Restart one or more containers
rm          Remove one or more containers
rmi         Remove one or more images
run         Run a command in a new container
save        Save one or more images to a tar archive (streamed to STDOUT by default)
search      Search the Docker Hub for images
start       Start one or more stopped containers
stats       Display a live stream of container(s) resource usage statistics
stop        Stop one or more running containers
tag         Create a tag TARGET_IMAGE that refers to SOURCE_IMAGE
top         Display the running processes of a container
unpause     Unpause all processes within one or more containers
update      Update configuration of one or more containers
version     Show the Docker version information
wait        Block until one or more containers stop, then print their exit codes
```

21

TUTORIAL 7

- Tutorial introduces use of two common Linux performance benchmark applications
 - stress-ng
 - 100s of CPU, memory, disk, network stress tests
- Sysbench
 - Used in tutorial for memory stress test

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.22

22

OBJECTIVES – 11/13

- Questions from 11/6
- Tutorial Questions
- Class Presentations:
Cloud Technology or Research Paper Review
- Ch. 5: Cloud Enabling Technology
- Tutorial 5 Demo

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.23

23

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:
Tuesday November 18th @ 11:59pm
- Presentation dates:
 - Tuesday November 25
 - Tuesday December 2*, Thursday December 4
 - * - day of quiz 2. only 1 presentation slot

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.24

24

EARN ADDITIONAL AWS CREDITS

- [Explore AWS](#)
[AI Agent Marketplace](#) [Amazon Quick Suite](#) [OpenAI open weight models](#)

- <https://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/free-tier-plans-activities.html>

L13.26

- AWS Management Console
- Elastic Compute Cloud (EC2)
- Instance Storage: Virtual Disks on VMs
- Elastic Block Store: Virtual Disks on VMs
- Elastic File System (EFS)
- Amazon Machine Images (AMIs)
- EC2 Paravirtualization
- EC2 Full Virtualization (hvm)
- EC2 Virtualization Evolution
- (VM) Instance Actions
- EC2 Networking
- EC2 Instance Metadata Service
- Simple Storage Service (S3)
- AWS Command Line Interface (CLI)
- Legacy / Service Specific CLIs
- AMI Tools
- Signing Certificates
- Backing up live disks
- Cost Savings Measures
- Disk images and S3

L13.27

[illegible]

- L13.29

- L13.30

30

INSTANCE STORAGE

- Also called ephemeral storage
- Persisted using images saved to S3 (simple storage service)
 - ~2.3¢ per GB/month on S3
 - 5GB of free tier storage space on S3
- Requires "burning" an image
- Multi-step process:
 - Create image files
 - Upload chunks to S3
 - Register image
- Launching a VM
 - Requires downloading image components from S3, reassembling them... is potentially slow
- VMs with instance store backed root volumes not pause-able
- Historically root volume limited to 10-GB max- **faster Imaging...**

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.31

31

ELASTIC BLOCK STORE

- EBS provides 1 drive to 1 virtual machine (**1 : 1 (not shared)**)
- EBS cost model is different than instance storage (uses S3)
 - ~10¢ per GB/month for General Purpose Storage (GP2)
 - ~8¢ per GB/month for General Purpose Storage (GP3)
 - 30GB of free tier storage space
- EBS provides "live" mountable volumes
 - Listed under volumes
 - Data volumes:** can be mounted/unmounted to any VM, dynamically at any time
 - Root volumes:** hosts OS files and acts as a boot device for VM
 - In Linux drives are linked to a mount point "directory"
- Snapshots back up EBS volume data to S3
 - Enables replication (required for horizontal scaling)
 - EBS volumes not actively used should be snapshotted, and deleted to save EBS costs...

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.32

32

EBS VOLUME TYPES - 2

- Metric: I/O Operations per Second (IOPS)
- General Purpose 2 (GP2)**
 - 3 IOPS per GB, min 100 IOPS (<34GB), max of 16,000 IOPS
 - 250MB/sec throughput per volume
- General Purpose 3 (GP3 - new Dec 2020)**
 - Max 16,000 IOPS, Default 3,000 IOPS
 - GP2 requires creating a 1TB volume to obtain 3,000 IOPS
 - GP3 all volumes start at 3000 IOPS and 125 MB/s throughput
 - 1000 additional IOPS beyond 3000 is \$5/month up to 16000 IOPS
 - 125 MB/s additional throughput is \$5/month up to 1000 MB/s throughput

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.33

33

EBS VOLUME TYPES - 3

- Provisioned IOPS (IO1)**
 - Legacy, associated with GP2
 - Allows user to create custom disk volumes where they pay for a specified IOPS and throughput
 - 32,000 IOPS, and 500 MB/sec throughput per volume MAX
- Throughput Optimized HDD (ST1)**
 - Up to 500 MB/sec throughput
 - 4.5 ¢ per GB/month
- Cold HDD (SC1)**
 - Up to 250 MB/sec throughput
 - 2.5 ¢ per GB/month
- Magnetic**
 - Up to 90 MB/sec throughput per volume
 - 5 ¢ per GB/month

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.34

34

ELASTIC FILE SYSTEM (EFS)

- EFS provides 1 volume to many client (**1 : n**) **shared storage**
- Network file system (based on NFSv4 protocol)
- Shared file system for EC2, Fargate/ECS, Lambda
- Enables mounting (sharing) the same disk "volume" for R/W access across multiple instances at the same time
- Different performance and limitations vs. EBS/Instance store
- Implementation uses abstracted EC2 instances
- ~ 30 ¢ per GB/month storage - **default burstable throughput**
- Throughput modes:**
 - Can modify modes only once every 24 hours
- Burstable Throughput Model:**
 - Baseline - 50kb/sec per GB
 - Burst - 100MB/sec per GB (for volumes sized 10GB to 1024 GB)
 - Credits - .72 minutes/day per GB

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.35

35

ELASTIC FILE SYSTEM (EFS) - 2

Burstable Throughput Rates

Information subject to revision

- Throughput rates: baseline vs burst
- Credit model for bursting: maximum burst per day

File System Size (GiB)	Baseline Aggregate Throughput (MiB/s)	Burst Aggregate Throughput (MiB/s)	Maximum Burst Duration (Min/Day)	% of Time File System Can Burst (Per Day)
10	0.5	100	7.2	0.5%
256	12.5	100	180	12.5%
512	25.0	100	360	25.0%
1024	50.0	100	720	50.0%
1536	75.0	150	720	50.0%
2048	100.0	200	720	50.0%
3072	150.0	300	720	50.0%
4096	200.0	400	720	50.0%

November 13, 2025

TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma

L13.36

36

ELASTIC FILE SYSTEM (EFS) - 3

- **Throughput Models** *Information subject to revision*
- **Provisioned Throughput Model**
- For applications with:
high performance requirements, but low storage requirements
- Get high levels of performance w/o overprovisioning capacity
- \$6 MB/s-Month (Virginia Region)
 - Default is 50kb/sec for 1 GB, .05 MB/s = 30 ¢ per GB/month
- If file system metered size has higher baseline rate based on size, file system follows default Amazon EFS Bursting Throughput model
 - No charges for Provisioned Throughput below file system's entitlement in Bursting Throughput mode
 - Throughput entitlement = 50kb/sec per GB

November 13, 2025 TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma L13.37

37

ELASTIC FILE SYSTEM (EFS) - 4

Performance Comparison, Amazon EFS and Amazon EBS *Information subject to revision*

	Amazon EFS	Amazon EBS Provisioned IOPS
Per-operation latency	Low, consistent latency.	Lowest, consistent latency.
Throughput scale	10+ GB per second.	Up to 2 GB per second.

Storage Characteristics Comparison, Amazon EFS and Amazon EBS

	Amazon EFS	Amazon EBS Provisioned IOPS
Availability and durability	Data is stored redundantly across multiple AZs.	Data is stored redundantly in a single AZ.
Access	Up to thousands of Amazon EC2 instances, from multiple AZs, can connect concurrently to a file system.	A single Amazon EC2 instance in a single AZ can connect to a file system.
Use cases	Big data and analytics, media processing workflows, content management, web serving, and home directories.	Boot volumes, transactional and NoSQL databases, data warehousing, and ETL.

November 13, 2025 TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma L13.38

38

EC2

- **EC2 Spot Instance Advisor:**
- <https://aws.amazon.com/ec2/spot/instance-advisor/>
- Provides sortable list of ec2 instance types with interruption (termination) frequencies
- Helps you choose an instance type that is less likely to be terminated
- Best practices for using spot instances:
- <https://docs.aws.amazon.com/whitepapers/latest/cost-optimization-leveraging-ec2-spot-instances/spot-best-practices.html>

November 13, 2025 TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma L13.39

39

EC2 - 2

- **On Amazon EC2, what is a "metal" instance?**
- A bare metal server is not shared with anyone
- There is no virtualization hypervisor
(program the contextualizes and hosts virtual machines)
- The operating system is installed directly on the root disk and the machine is booted directly like a laptop or desktop computer
- The user can install any operating system and make configurations changes to the machine's base operating system
- The user can then install and control a virtualization hypervisor on bare metal servers
- Bare metal servers were offered on AWS starting in ~2017

November 13, 2025 TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma L13.40

40

AMAZON MACHINE IMAGES ★

- AMIs
- Unique for the operating system (root device image)
- Two types
 - Instance store
 - Elastic block store (EBS)
- Deleting requires multiple steps
 - Deregister AMI
 - Delete associated data - (files in S3)
- Forgetting both steps leads to costly "orphaned" data
 - No way to instantiate a VM from deregistered AMIs
 - Data still in S3 resulting in charges

November 13, 2025 TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma L13.41

41

EC2 VIRTUALIZATION HYPERVERSORS

- 1st, 2nd, 3rd, 4th generation → XEN-based
- 5th+ generation instances → AWS Nitro virtualization
- XEN (pre-nitro) - two virtualization modes:
- XEN **Para-virtualization (PV)** "paravirtual"
 - 10GB Amazon Machine Image - base image size limit
 - Addressed poor performance of old XEN HVM mode
 - I/O performed using special XEN kernel with XEN paravirtual mode optimizations for better performance
 - Requires OS to have an available paravirtual kernel
 - PV VMs: use common **AKI** files on AWS - **Amazon kernel Image(s)**

November 13, 2025 TCSS462/562: Software Engineering for Cloud Computing [Fall 2025]
School of Engineering and Technology, University of Washington - Tacoma L13.42

42

EC2 VIRTUALIZATION - HVM

- XEN HVM (hardware virtualization) mode:
 - HVM just meant that virtualization used CPU extensions (Intel VT-x or AMD-V) to run an unmodified OS directly on the VM
 - Today HVM usually implies there is more than just native-CPU virtualization support, but also I/O-device support, etc.
 - Full virtualization - no special OS kernel required (no AKI)
 - Computer entirely simulated (... i.e. slow)
 - MS Windows pre-nitro ran in "HVM" mode
 - Allows work around: 10GB instance store root volume limit
 - Kernel is on the root volume (under /boot)
 - No AKIs (kernel images)
 - Commonly used today (EBS-backed instances)

November 13, 2025

TCSS462/562 (Software Engineering for) Cloud Computing (Fall 2025)
School of Engineering and Technology, University of Washington - Tacoma

L13.43

43

XEN PV VS HVM ON EC2

Feature	Paravirtual (PV)	Hardware Virtual Machine (HVM)
Guest OS Modification	Requires the guest OS to be modified (aware of the hypervisor).	Allows the use of unmodified guest OS, as if running on bare metal hardware.
Boot Method	Boots with a special boot loader called PV-GRUB, which then chain loads the kernel.	Boots by executing the master boot record of the root block device, like a physical machine.
Hardware Access	Relies on a patched kernel and specialized drivers (hypercalls) for I/O operations and direct communication with the hypervisor.	Emulates a full set of hardware and takes advantage of hardware virtualization extensions (Intel VT-x, AMD-V) for fast access to the underlying physical hardware.
Performance (Historically)	Historically offered better performance than early HVM instances due to the use of direct drivers for network/storage I/O.	Early HVM had more overhead due to hardware emulation.
Performance (Modern Legacy)	Performance is limited and lacks modern enhancements.	With the availability of PV drivers for HVM AMIs and hardware advancements, HVM instances provide better or superior performance for most workloads.
Feature Support	Does not support modern features like enhanced networking or GPU processing.	Required for advanced features such as enhanced networking, GPU instances, and specific instance types (e.g., I3 instances).

November 13, 2025

TCSS462/562 (Software Engineering for) Cloud Computing (Fall 2025)
School of Engineering and Technology, University of Washington - Tacoma

L13.44

44

EC2 VIRTUALIZATION - NITRO

- Nitro based on Kernel-based-virtual-machines
 - Stripped down version of Linux KVM hypervisor
 - Uses KVM core kernel module
 - I/O access has a direct path to the device
- Goal: provide indistinguishable performance from bare metal
- There are 5 Nitro versions (v2 to v6)
- Article describes features and provides links to instance families (m, c, r, etc.) where it is possible to check the Nitro version for specific instance types to understand feature evolution
- <https://docs.aws.amazon.com/ec2/latest/instancetypes/ec2-nitro-instances.html>

November 13, 2025

TCSS462/562 (Software Engineering for) Cloud Computing (Fall 2025)
School of Engineering and Technology, University of Washington - Tacoma

L13.45

45

EVOLUTION OF AWS VIRTUALIZATION

From: <http://www.brendangregg.com/blog/2017-11-29/aws-ec2-virtualization-2017.html>

AWS EC2 Virtualization Types

#	Tech	Type	With	Importance				
				Most	Least	Most	Least	
1	VM	Fully Emulated		VS	VS	VS	VS	
2	VM	Xen PV 3.0	PV drivers	P	P	P	VS	
3	VM	Xen HVM 3.0	PV drivers	VS	P	P	VS	
4	VM	Xen HVM 4.0.1	PVHVM drivers	VS	P	P	VS	
5	VM	Xen AWS 2013	PVHVM + SR-IOV(nets)	VS	VS	P	VS	
6	VM	Xen AWS 2017	PVHVM + SR-IOV(nets, dist)	VS	VS	VS	VS	
7	VM	AWS Nitro 2017		VS	VS	VS	VS	
8	HW	AWS Bare Metal 2017		H	H	H	H	
		Bare Metal		H	H	H	H	

VM: Virtual Machine; HW: Hardware.
VS: VS in software; VS: VS in hardware; P: Paravirt; Not all combinations shown.
SR-IOV(nets): accelerated driver; SR-IOV(storage): native driver.

November 13, 2025

TCSS462/562 (Software Engineering for) Cloud Computing (Fall 2025)
School of Engineering and Technology, University of Washington - Tacoma

L13.46

46

QUESTIONS

November 13, 2025

TCSS462/562 (Software Engineering for) Cloud Computing (Fall 2025)
School of Engineering and Technology, University of Washington - Tacoma

L13.130

130