Shubhangi Dutt
CS-524A

CS-524A : Introduction to Cloud Computing
Assignment-2

Answer 2

All Operating Systems (OS) make use of clock timers in order to perform preemptive multitasking.Whenever the timer gives a notice the OS decides which task should run on it. The timer needs to run even if the machine is idle. This however is not the case in paravirtual OS.

Paravirtualization is a type of virtualization that recompiles and installs an OS on a virtual machine that is operating on a host OS. Paravirtualization executes commands by hypercalls to virtual layers. The virtual machine code is updated to send notices of timers at set time instead of regular predefined intervals. With this update the hypervisor no longer needs to send notices when the machine is idle therefore improving efficiency.

Answer 3

Consider the case of parallel processing which makes use of SMP or Symmetric Multi-Processing where multiple interconnected processors run simultaneously on a shared memory. Scheduler and interrupt handlers are used to redirect calls to the required processor. In order to facilitate SMP, Advanced Programmable Interrupt Controller (APIC) is used which sets priorities to interrupts. Because all processors share the same primary memory, the size of memory required is higher, meaning that it is more expensive and multiple processes running simultaneously require a complex structure which makes each transition costly. This is where paravirtualization cuts our costs. Since paravirtualization has a full view of the code it doesn't need multiple calls and hence replaces them with one single hypercall.

Answer 4

User-mode and kernel-mode are two distinct modes in Linux. The user stack is utilized when running a process in user mode. The kernel stack is part of the Kernel space and hence unavailable to processes operating in user mode. In user mode, the mode bit is set to 1. When transitioning from user mode to kernel mode to execute a Syscall, it is changed from 1 to 0. As a result, there are two distinct stacks for user and kernel modes, each of which may be accessed by various processes.

Answer 5

To give the processor information about a certain segment's location, size or access control we make use of a data structure called Segment Descriptor. In order to determine the size of the segment we make use of the Segment Limit Field. A Segment Limit Field interprets the G flag; if the granularity (G flag) is clear the segment size's range is from 1 byte to 1 MByte in Byte increments, else if the G-flag is set, the segment size from 4 KBytes to 4 Gbytes in 4 Kbyte increments.
The Unscrambled Limit here is the limit that is scaled according to the setting of the G flag in this Segment Descriptor.

Answer 6

1)
Advantages of using I/O MMU:-

- "Decoupling provides for flexible mappings between logical and physical devices, facilitating seamless portability."
  Consider a number of heterogeneous systems with different I/O devices and configurations, all of which can be linked from logical I/O devices to physical devices, provided they have different but compatible interfaces. This can be achieved by the I/O Virtualization Layer that uses the same VM image to provide the conversion.

- Another advantage is “Live Migration”. This enables us to relocate a running VM across real computers. All active I/O devices are first unlinked to the machine then recoupled once the VM is active.

- “Device Aggregation” provides an edge by integrating multiple single physical devices into a single logical device. For example combining multiple smaller storage disks into a larger storage disk.

- “Many I/O virtualization enhancements are designed to improve system security.” This, for example, can be achieved upon enabling a transparent disk encryption by running an encryption function over I/O.

- As cited in the paper “The Price of Safety” , the CPU utilization increases significantly , “CPU utilization, however, could be as much as 60% more in a hypervisor environment and 30% more in a bare-metal environment, when the IOMMU was enabled”

Disadvantages of using I/O MMU:-

- Throughput remains the same with or without I/O MMU as presented in the paper “The Price of Safety.” In the common case throughput remained the same whether the IOMMU was enabled or disabled”

- So far the CPU Utilization costs are too high and reducing them is a work in progress. “The main CPU utilization cost came from too-frequent map and unmap calls (used to create translation entries in the DMA address space)”

- Scheduling also has an impact on VM performance by generating issues like latency while utilizing TCP networking due to CPU congestion. This problem can be overcome by employing Network Interface Controllers, although the sophistication is yet in the works.

2)

Hardware Virtualization is accomplished by running a hypervisor on bare metal , this enables multiple workloads or OS instances to run concurrently on a single physical platform. A Carrier Grade Virtualization is a software component that ensures the qualities and performances of the existing carrier grade solutions. This is ensured by maintaining a certain number of features in the VM.
These features include but are not limited to the following:-

- Availability - on a physical device there exists a one-to-one configuration between OS, middleware and applications. A hypervisor boosts this configuration with HA middleware that enables monitoring of apps. It provides an external view of guest OS health/resources as well as a generic framework for controlling the guest OS lifetime.

- Scalability - Hypervisor augments the spec with virtual clustering because application software and operating systems are incapable of scaling to a high number of cores. Embedded virtualization relieves the OS and applications of the difficulties of multi core scalability.

- Better Error recovery - Since the guest OS is isolated from the core machine and the Hypervisor it makes for a better fault tolerance. Carrier Grade Virtualization takes advantage of this inherent characteristic by allowing VM specialization and devoting them to execute native device drivers, as well as exporting (para)virtualized devices to other VMs. Because device drivers are isolated in VMs, the impact of a device driver failure is limited to only the virtual machines that use a virtualized device.

- Security - Carrier Grade Hypervisor must safeguard vital resources from unauthorized access by other VMs. Carrier Grade Virtualization must provide each tenant the ability to administer a part of the system while restricting them from accessing the resources or information flows of other tenants. The overall physical platform administration must also be delegated to a platform administrator rather than any of the tenants.

Some products that are available include VMware , Carrier Grade Linux.

Answer 7

Nitro hypervisor - Currently amazon makes use of Nitro Hyeprvisor in its Nitro system.

- The Nitro Hypervisor enables Amazon's EC2 to segregate certain operations and offload them to specialized hardware and software. This saves money by transferring the bulk of a server's resources to a user's instance.
- Nitro is a lightweight hypervisor.

- It has faster provisioning time with virtualized hardware.
- It is based on the KVM core kernel module.

Xen Hypervisor

- Before Nitro , Xen Hypervisor was used by Amazon. It is an open-source Hypervisor.
- It focused on hardware virtualization for EBS volumes.

Answer 8

a)

Amazon uses "EC2 Compute Units," or ECUs, as a measure of virtual CPU power. One ECU is similar to a 2007 Intel Xeon or AMD Opteron CPU running at 1 GHz to 1.2 GHz. EC2 Compute Units are calculated by multiplying the number of compute units by the number of virtual cores allotted to each instance, allowing comparisons between instances. Since one virtual CPU core = one physical CPU core the power can be calculated by determining the number of cores and multiplying ECUs with cores.

b)

General Purpose - General-purpose instances have a good combination of computation, memory, and networking capabilities and can handle a wide range of applications. These instances are appropriate for applications like web servers and code repositories that require these resources in equal amounts. This is facilitated by Intel i7 core processors and 32 Gib memory.

Compute optimized - If a user needs a machine which can support high multiprocessing then a Compute optimized instance is the right way to go. It provides high performance CPUs that are ideal for high performance computing such as gaming servers or computationally heavy applications.
Memory optimized - If the user is interested in Big Data analytics or utilizes databases that demand large amounts of Memory, a regular instance may fail during periods of high traffic. In this situation, a Memory optimized instance with up to 25 Gbps network speed and 19000 Mbps EBS capacity is employed. This provides rapid delivery of workloads with high data utilization. It is powered by the AWS Nitro System.

Accelerated optimized - For applications of Machine Learning HPCs are required however not just CPUs but also co-processors or hardware accelerators that can perform calculation heavy tasks with more efficiency. In this case Accelerated optimized instances are used. This is facilitated by robust core GPUs such as Nvidia A100 Tensor. Applications include drug discovery, speech recognition among others.

Storage optimized - Consider a business with a requirement of a medium sized database(like Relational database , MySQL , PostgresQL) but fast I/O transactions to and from local storage and database. This requires high performance and fast network throughput which is provided by Storage optimized instances. The two most important factors to have enabled this are; up to 100 Gbps of network bandwidth using Elastic Network Adapter (ENA)-based Enhanced Networking and support for Elastic Fabric Adapter on im4gn.16xlarge.

.

c)

The following Operating Systems are available on the above mentioned systems:-

1. Linux based systems that include
   Ubuntu 18.04
   CentOS7
   Red Hat Enterprise
2. Windows Based Servers

d)

An AMI or amazon machine image is a disk image that is ready to run on a hypervisor that is used to create a virtual machine in the EC2. An AMI provides the information that is required to launch an instance or multiple instances. The disk image must contain all the root information to run an instance such as OS , storage ,memory etc.

e)

Components of an AMI include:-

- EBS backup to recover data loss or migration , root device data, boot instances.
- Permissions to launch instances of the machine on AMI
- A block device that maps the root device data onto AMI inorder to incorporate it into the instance launched.

Answer 9

- On Demand Pricing- Pay-as-you-go is the most essential benefit of cloud services, and on-demand pricing offers it. As the name implies, the user only pays for the time the instance is in use, which is invoiced per 'instance-hour' from the time it starts until the time it ends. This saves the user time and

money by avoiding the need to plan and acquire large sums of money for resources they may or may not utilize in the future.
For example for a Compute Optimized, Linux-2 core ,16 GB memory, network performance of upto 10 GB in New York region the charges are $0.48/hour.

- Compute Savings Plan - A savings plan is useful for users who are not willing to invest a lot immediately but can in the long term(1-3 years). The Compute Savings Plan provides a discounted rate of upto 66% on payments of dollars/ hour to the user.
  For example, for a standard reserved instance with a 1 year term agreement in the US East region, Linux OS dedicated instance the per hour usage rate is $0.025. This Is significantly less than on-demand pricing.

- Ec2 Instance Savings Plan - The instance savings plan provides an even higher discount, upto 72% to the user provided the region is fixed. The instance can be switched within the family, size, OS and tenancy but not across regions.

- Ec2 Reserved Instances Pricing - When a user requires capacity reservation it enables them to get upto 72% discount on the prices. They can switch between availability zones and instance size however the Instance family, OS and tenancy remain fixed.
  For example, for a standard reserved instance with a 1 year term agreement in the US East region, Linux OS dedicated instance the per hour usage rate is $0.018 which is significantly discounted.

- Ec2 Reserved Convertible Instance Pricing- In order to get a flexibility option for future this option comes handy. It allows switching in instance size , family, OS , tenancy and availability zones at a reduced discount of 66%.
  For example, for a standard reserved instance with a 1 year term agreement in the US East region, Linux OS dedicated instance the per hour usage rate is $0.019 which is slightly less discounted than standard reserved.

- Dedicated Instances Pricing - This term is used for prices on dedicated instances instead of shared ones and are priced for upto 70% in on-demand pricing and upto 90% in spot pricing.

- Spot Instances Pricing - The unused Ec2 capacity can be bid on by the users just like in on-demand however in this case the discounts go upto 90%. Amazon EC2 sets Spot Instance rates, which fluctuate over time depending on long-term patterns in supply and demand for Spot Instance capacity.
  For example for a Linux based general purpose instance in the US East region the charges go as low as $0.0035 per hour.

Answer 10

a)
A service-level agreement (SLA) specifies the amount of service that a customer expects from a supplier, as well as the metrics that are used to assess that service, as well as any remedies or penalties that may be imposed if the agreed-upon service standards are not fulfilled. SLAs are most commonly used between firms and external suppliers, although they may also be used between two divisions inside the same organization.

Key Objectives include :-

- Service Agreement
- Service Management
- Performance Level
- Reporting Issues
- Resolution of issues
- Repercussion of issues

In order to maintain this service for free, user needs to register for the free tier service.

b)

The AWS Free Tier allows for 750 hours of total consumption and 30 GB-months of storage each month.

To create a Windows EC2 instance that is eligible for the AWS Free Tier, follow these steps:

1. Open the Amazon EC2 console and select Launch Instance in the Launch Instance Section
2. In the navigation pane, select Free Tier only.
3. Choose Select next to a Windows AMI that's labeled "Free tier eligible."
4. Select an instance type that's labeled "Free tier eligible."
5. Choose Next: Configure Instance Details.
6. On the Configure Instance Details page, select the following:-
   'Instance Tenancy' is set to Shared,
   'Request Spot instances' is not selected
7. Choose Next: Add Storage. Choose up to 30 GB of storage. Under Storage Type, choose either General Purpose SSD (GP2) or Magnetic.
8. Choose Next: Add Tags.
9. Choose Next: Configure Security Group.
10. Configure a security group that allows only trusted traffic in and out.
11. Choose Review and Launch.
12. Review your configuration. If it matches your specifications, choose Launch.

c)

Yes, we can construct a machine instance of our own PC and transfer the disk image to the server. This may be accomplished by setting up an EC2 instance and hosting it as a server, to which we will connect our own PC to transfer the image.

Resources

https://en.wikipedia.org/wiki/Paravirtualization
https://searchservervirtualization.techtarget.com/definition/paravirtualization
https://www.tutorialspoint.com/User-Mode-vs-Kernel-Mode
https://www.intel.com/content/dam/support/us/en/documents/processors/pentium4/sb/25366821.pdf
https://cacm.acm.org/magazines/2012/1/144808-i-o-virtualization/fulltext
https://www.kernel.org/doc/ols/2007/ols2007v1-pages-9-20.pdf
https://cacm.acm.org/magazines/2012/1/144808-i-o-virtualization/fulltext
https://www.kernel.org/doc/ols/2007/ols2007v1-pages-9-20.pdf
https://aws.amazon.com/blogs/hpc/bare-metal-performance-with-the-aws-nitro-system/
https://www.cloudhealthtech.com/blog/brief-guide-and-comparison-amazon-ec2-instance-types#:~:text=EC2%20Compute%20Units%20represent%20the,different%20Amazon%20EC2%20Instance%20types.
https://aws.amazon.com/ec2/pricing/
https://aws.amazon.com/ec2/pricing/on-demand/
https://aws.amazon.com/savingsplans/compute-pricing/
https://aws.amazon.com/ec2/pricing/reserved-instances/pricing/
https://aws.amazon.com/ec2/spot/pricing/
https://aws.amazon.com/ec2/pricing/dedicated-instances/