UMUC

University of Maryland University College

CC615x

Cloud Infrastructure

Week 2: Cloud Infrastructure Compute Part 2

Auto-scaling

- For deployments of multiple virtual servers working together in the cloud, cloud vendors usually offer some version of auto-scaling mechanism creation of a server group to which virtual servers may be added or removed depending on a set of conditions, such as:
  - Server virtual CPU load
  - Network bandwidth usage
  - Or time/calendar constrain
- For example, an electronic commerce vendor may want to automatically add virtual servers to their storefront in the cloud on weekends when more people are likely to shop, and then downscale on Monday when shopping traffic decreases.

Visit the website below for more information on the auto-scaling feature in Amazon Web Services: https://aws.amazon.com/autoscaling/

Load Balancing

- For server groups handling requests from users, cloud vendors also provide the load balancing feature; a gateway which splits incoming requests among individual servers based on their load, or some balancing algorithm (for example: “round robin” type, assigning servers in turn). Load balancing can be local and confined to a data center, or global distributing workload across data center.

- Visit the website below for a Microsoft Azure load balancing service description:

  https://azure.microsoft.com/en-us/blog/microsoft-azure-load-balancing-services/

Virtual Server Charges

- Charges for virtual servers are hourly, minute, or pay-per-use basis. When you stop a virtual server, they can be de-allocated, and then the RAM memory and local storage goes away and server charges stop; however remote storage and other charges will still be incurred. If the server makes use of remote disks, they can be retained, and the instance can be brought back to life, often on a different physical host.

- Many vendors offer several paying options, such as monthly payments if you want to use the server on continuous basis, or enterprise subscriptions for longer term. Amazon has even an option of competing for unused server capacity at a discount.

- Visit the website below for a Microsoft Azure pricing information: https://aws.amazon.com/pricing/

UMUC
University of Maryland University College

Scaling in the Cloud

- Ease of virtual server provisioning lends the cloud to utilize the horizontal scaling model, in which when the need for compute power increases, more servers are added, as in auto-scaling. This is the opposite of vertical scaling, where servers running out of capacity are replaced with bigger servers.

- Horizontal scaling is one of the tenets of cloud computing and applies to other kinds of resources too. Instead of getting bigger, which always has a limitation, you get more, which has fewer limitations and suits today's massive applications with strongly variable resource needs.

UMUC

University of Maryland University College

Bare Metal Servers

What is Bare Metal Server

- Many cloud vendors offer the capability of renting complete physical servers, without virtualization. You will not have virtual “neighbors” on such servers as they are dedicated solely to your needs. While more expensive, such servers allow clients to have full control over them. Many organizations use them for private off-premises clouds, running a dedicated hypervisor and cloud management software on them.

- Having the whole physical server at your disposal allows you to take advantage of various hardware features, such as Intel’s security enhancing Trusted Execution Technology.

- Provisioning bare metal servers may take longer that virtual servers and billing is usually completed on a monthly basis. When you stop a bare metal server and don’t de-allocate it, you continue to be charged for it.

Colocation

- Many cloud vendors offer the option of co-location, allowing customers to bring in their physical servers and install them in the vendor's data center. This way, customers can take advantage of the provider's power, cooling, physical security features, while they own and fully control the hardware. Other devices, such as network or storage devices, can also be co-located.

- There is usually a process for co-location to adhere to the provider's security measures. Cloud providers allow for co-located resources to connect to their cloud offering. A typical example is co-locating a custom network getaway to provide connectivity between on and off-premises hybrid cloud environments.

UMUC

University of Maryland University College

Containers

UMUC
University of Maryland University College

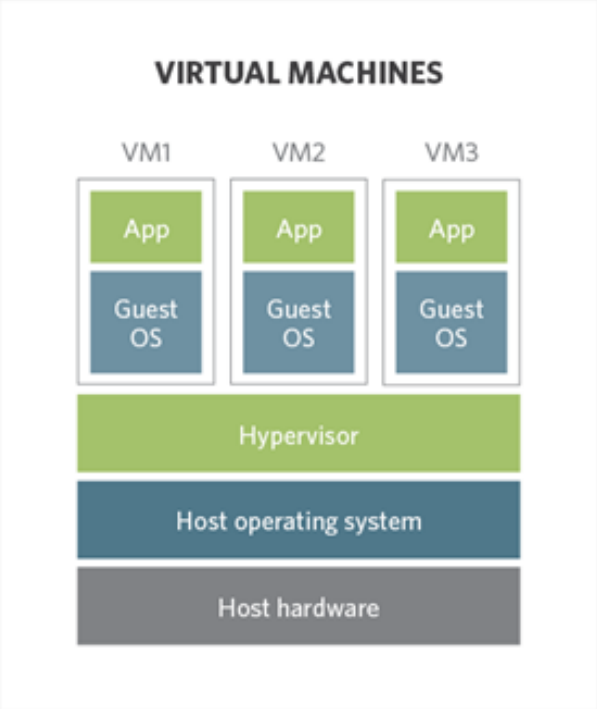
What is a Container?

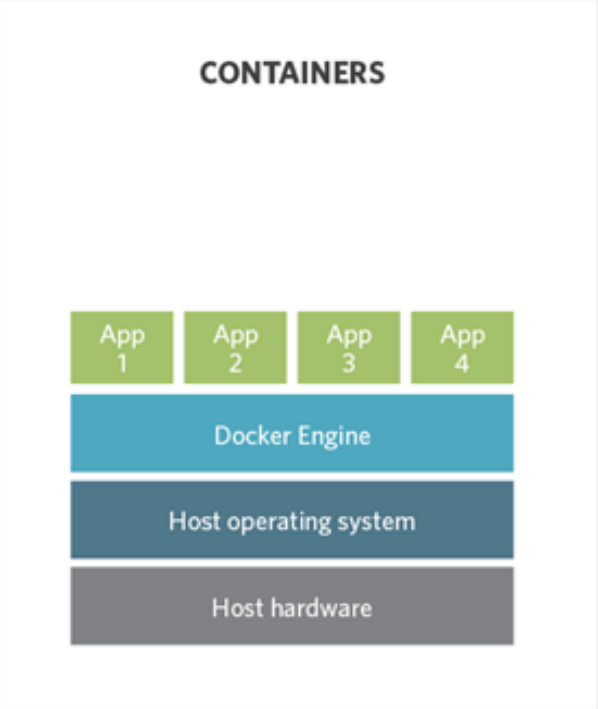
- In recent years, another form of cloud compute started gaining popularity "containers."

- While each virtual server carries its guest operating system, containers run on top of the single operating system installed on the host server. Containers are computing units offering a high level of isolation from each other. Each container is confined to its own "namespace" and it sees only its file system and memory.

- Containers are provisioned from container images, containing libraries and runtimes needed for running a particular application. They do not contain the operating system binaries. Cloud vendors offer libraries of standard container images, as well as support for creation and storage of custom images. One particularly popular container implementation open-source project is called **Docker**.

Virtual Machines Versus Containers





Advantages and Disadvantages of Containers

- Creating a container does not involve installing and booting the operating system; they can be brought up and down much faster than virtual servers. Container images are smaller and easier to create and use.

- They only have one operating system: the host. Since containers are isolated environments, and each can have its own set of libraries and runtimes, they can run without version conflict even with a shared OS.

- The fact that there is a single shared OS is a limitation. You can't run Windows and Linux apps in containers on a single host. Also, the security holes in the shared OS can influence all containers on a host.

UMUC
University of Maryland University College

Advantages and Disadvantages of Containers, cont.

- Container technology is gaining popularity and is being offered by most cloud vendors. It is particularly well suited for deployment of micro-services independent application components, each with its development and deployment cycle.

- Visit the following websites for information about container services available in AWS and for Azure containers:

  https://aws.amazon.com/ecs/

  https://azure.microsoft.com/en-us/services/container-service/

UMUC

University of Maryland University College

Server-less Computing

What is Server-less Computing?

- Last but not least, we need to mention the most lightweight compute option – server-less computing. If you just have a piece of software which you want to run in the cloud as a reaction to some external event, and you don't want to care about provisioning and maintaining servers or containers, you may take advantage of it. It is called FaaS (Function as a Service). You will have to adhere to a particular way of writing and deploying your code, but the cloud vendor will supply the infrastructure to run it without your intervention, and release it upon completion.

- Code to be deployed on server-less compute has to be stateless; which means that no data can survive across its invocations unless it is persisted externally. Applications best suited for server-less compute are systems handling events coming from a large number of sources which do not require complicated processing, such as Internet of Things (IoT) software. See this website for more information about server-less computing: https://en.wikipedia.org/wiki/Serverless_computing

UMUC
University of Maryland University College

Server-less Computing? (Cont.)

- Most cloud vendors offer server-less computing in some form. AWS offering is called AWS Lambda.

- Visit the website below to learn more:

  https://aws.amazon.com/serverless/

- On the other hand, Microsoft Azure calls it Functions. Visit the website below for more information:

  https://azure.microsoft.com/en-us/services/functions/

Summary

This week you learned about:

Virtual Servers	Bare Metal Servers	Containers	Server-less Computing
• Virtual Server Implementation • Virtual Server Provisioning • Auto-scaling • Load Balancing • Virtual Server Charges • Scaling in the Cloud	• What is a Bare Metal Server? • Colocation	• What is a Container? • Advantages and Disadvantages of Containers	• What is Server-less Computing?