Let's Start With Some Basics

Understand the fundamentals of what is a Server, Virtualization, Hypervisor. What are the fundamental components that encompass a Data Center ? This section will then go into the reasons why organizations move to the cloud and the technical benefits of moving to the cloud.

The below is a glossary of terms that you need know. The reason for brining this right in the start of the course is to help you to understand the fundamental terminologies before proceeding through the course.

Servers

Think of them as computers that are responsible for the storage of data, computation (where your application lives and provides some value), perform some networking functions. Inside a server you will find a processor, motherboard, RAM, a networking card, a hard disk and a power supply.

Servers are expensive and they require a lot of power and people to keep them operational. They are normally procured by the IT team and are kept running as long as possible so that the organization gets the maximum return on their investment. The hardware on the servers fail more often than you think and requires maintenance.

Physical servers are great because they can be configured however you want. But, physical servers lead to waste because it is difficult to run multiple applications on the same server. Software conflicts, network routing, and user access all become more complicated when a server is maximally utilized with multiple applications. Hardware virtualization solve some of these problems.

Virtualization

Virtualization emulates a physical server's hardware in software. A Virtual Machine (VM) can be created on demand, is entirely programmable in software. Hypervisors increases these benefits because you can run multiple virtual machines (VMs) on a physical server.

Hypervisors

Hypervisors allows applications to be portable because you can move a VM from one physical server to another. One problem with running your own virtualization platform, is that VMs still require hardware to run. Companies still need to have all the people and approvals required to run physical servers, but now capacity planning becomes harder because they have to account for VM overhead too.

What is a Data Center ?

A data center is a physical facility that is used to house all the hardware that is needed for applications to function.

Data Center Components

Data centers are often referred to as a singular thing, but in actuality they are composed of a number of technical elements such as routers, switches, security devices, storage systems, servers, application delivery controllers and more. These are the components that are needed to store, run and manage applications. Due to the needs of the application and the decreasing life of the hardware managing a data center is a time consuming task.

Data Center Infrastructure

In addition to technical equipment a data center also requires a significant amount of facilities infrastructure to keep the hardware and software up and running. This includes power subsystems, uninterruptible power supplies (UPS), ventilation and cooling systems, backup generators and cabling to connect to external network operators i.e a connection to the internet where applicable.

Data Center Architecture

Any company of significant size will likely have multiple data centers possibly in multiple regions. This gives the organization flexibility in how it backs up its information and protects against natural and manmade disasters like floods, storms, earthquakes etc. How the data center is architected can be some of the most difficult decisions because there are almost unlimited options.

Reasons To Move To The Cloud

Capital Expense vs Variable Expense !

Instead of having to invest heavily in building data centers and acquiring servers before you know how you are going to use them, you pay only when you consume resources.

Benefit from massive economies of scale:

By using cloud computing, you can achieve a lower variable cost than you can get on your own. Because usage from hundreds of thousands of customers are aggregated in the cloud, providers like AWS, Azure, Oracle can achieve higher economies of scale, which translates into lower pay as-you-go prices for the end IaaS customers - you !

Stop guessing about capacity:

Eliminate guessing on your infrastructure capacity needs. When you make a capacity decision prior to deploying an application, you often end up either sitting on expensive idle resources or dealing with limited capacity. With cloud computing, these problems are negated. You can access as much or as little capacity as you need, and scale up and down as required.

Increase speed and agility:

In a cloud computing environment, new IT resources are only a click away, which means that you reduce the time to make those resources available to your developers from weeks to just minutes. This results in a dramatic increase in agility for the organization, since the cost and time it takes to experiment and develop is significantly lower.

Stop spending money on running and maintaining data centers:

Focus on projects that differentiate your business, not the infrastructure. Cloud computing lets you focus on your own customers, rather than on the heavy lifting of racking, stacking, powering and maintaining your servers.

Go global in minutes

Easily deploy your application in multiple regions around the world with just a few clicks. This means you can provide lower latency and a better experience for your customers at minimal cost.

Almost zero upfront infrastructure investment:

If you have to build a large-scale system it may cost a fortune to invest in real estate, physical security, hardware (racks, servers, routers, backup power supplies), hardware management (power management, cooling), and operations personnel. Because of the high upfront costs, the project would typically require several rounds of approvals before the project could even get started. Now, with utility-style cloud computing, there are no fixed or startup cost.

Just-in-time Infrastructure:

In the past, if your application became popular and your systems or your infrastructure did not scale you became a victim of your own success. Conversely, if you invested heavily and the app did not get popular, your investment in infrastructure is lost. By deploying applications in-the-cloud with just-in-time self-provisioning, you do not have to worry about pre-procuring capacity for large-scale systems. This increases agility, lowers risk and lowers operational cost because you scale only as you grow and only pay for what you use.

More efficient resource utilization:

System administrators usually worry about procuring hardware (when they run out of capacity) and higher infrastructure utilization (when they have excess and idle capacity). With the cloud, they can manage resources more effectively and efficiently by having the applications request and relinquish resources ondemand.

Usage-based costing:

With utility-style pricing, you are billed only for the infrastructure that is being used. You are not paying for allocated or unused infrastructure. This adds a new dimension to cost savings. You can see immediate cost savings (sometimes as early as your next month's bill) when you deploy an optimization patch to update your cloud application.

Moreover, if you are building and selling a platform on the top of the cloud, you can pass on the same flexible, variable usage-based cost structure to your own customers.

Reduced time to market:

Parallelization is the one of the great ways to speed up processing. If one

compute-intensive or data-intensive job that can be run in parallel takes 500 hours to process on one machine, with cloud architectures, it would be

possible to spawn and launch 500 instances and process the same job in 1 hour. Having available an elastic infrastructure provides the application with the ability to exploit parallelization in a cost-effective manner reducing time to market.

Technical Benefits of Cloud Computing

Some of the technical benefits of cloud computing include:

Automation"Scriptable infrastructure":

You can create repeatable build and deployment systems by leveraging programmable (API-driven) infrastructure. This comes back to the point of deploying in a new geographical region in less than 10 minutes. You are not hunting for data centers or buying hardware. You simply run your cloud formation script to create your entire fleet of servers along with all your network configurations.

Auto-scaling:

You can scale your applications up and down to match your unexpected demand without any human intervention. Auto-scaling encourages automation and drives more efficiency.

Proactive Scaling:

Scale your application up and down to meet your anticipated demand with proper planning understanding of your traffic patterns so that you keep your costs low while scaling.

Efficient Development lifecycle:

Production systems may be easily cloned for use as development and test environments. Staging environments may be easily promoted to production.

Improved Testability:

Never run out of hardware for testing. Spin-up and spin down testing environments as and when you need them. Whether you want to run automation tests, load test, longevity tests you can use the infrastructure and return it when you are done with them.

Disaster Recovery and Business Continuity:

The cloud provides a lower cost option for maintaining a fleet of DR servers and data storage. With the cloud, you can take advantage of geo-distribution and replicate the environment in other location within minutes.

"Overflow" the traffic to the cloud:

With a few clicks and effective load balancing tactics, you can create a complete overflow-proof application by routing excess traffic to the cloud from your existing on-prem infrastructure.