**What is Cloud computing**

* Infinite computing. Near infinite compute, CPU storate, CPU memory.
* Elimination of upfront cost for infrastructure building.
* Utility – if you use resources inefficiently, you pay a lot of money, but if you can use efficiently, you can treat it like a utility.
* Comparative advantage – you focus on things you are best at and leave other things to other people.

**Week 1**

**[1.0] [5 points] The evolution of Cloud Computing has been compared to the evolution of electricity supply as a utility. Describe specific problems that Cloud Computing solves as compared to businesses running their own data centres.**

When businesses run their own data canters, they face the following problems:

* Need up-front payment – with cloud you solved with pay as you go. Pay as you go model is cost-effective, so you pay only for resources that you use.
* Difficult to scale and it is expensive - with cloud you have instant access to any region
* Availability, reliability – if network goes down, service will suffer- with cloud you have guarantee from the provider for exactly how many minutes in a year you will have downtime.

**[2.0] [5 points] Describe the different categories of services (XaaS) cloud computing can provide with specific examples of each service.**

The most common types are Saas, Paas, Iaas, however, certain cervices within these categories can be offered such as Dass, Caas.

Saas: Software as a service is provided on substription basis and is centrally hosted. Email, Dropbox, Calendar. Software that you can rent and incorporate into your infrastructure. In short is at a large scale, renting out capabilities.

Gmail, no hosting the web server.

-Splunk, Data Dog

Iaas: Provide all the required physical infrastructure and computing resources on the cloud. AWS, Azure. Is like a big warehouse store where you get things in bulk but you still have to make sure that you can configure all that equipment and tie it together. Get things in bulk and the cost is very low.

Amazon: EC2 A lot like Costco.

You spin up virtual machines, setup the networking layer, but at a significant cost savings.

Paas: Platform as a Service is a high-level abstraction that allows the SW team to focus on the logic only and infrastructure provider will take care of the rest of the work of deploying that into production. You pay bit more.

Full-service gasoline fill up at a station vs doing it yourself.

Ex: Heroku

Google App Engine (GAE)

Amazon: Beanstalk

MaaS(Metal as a Service) is a newer term. A lot of the cloud paradigms that you use can be applied to physical hardware. This is when you want to deal with specialized situations like GPU programming or maybe large storage systems.

The ability to spin up and provision machines yourself. You can physically control servers. GPU – You may have a very specialized multi-GPU setup for, let’s say, machine learning, or specialized database and you want to control that physical hardware.

Serverless (Function as a Service Faas) newer types of service models where you do not need to worry at all about the underlying infrastructure. It has some similarities and some overlap with Paas. AWS Lambda, a function that you put into the Cloud and you can map events to it. A lot of it is around piece of logic you put on Cloud and hook it up to an event. Having a light bulb in garage when you open garage when it is night on. Abstracting business logic into a unit of work and then applying that, the work wherever you need to.

**[3.0] [10 points] An established financial company is about to launch their new banking application. Give 5 reasons why the company should use their own data centre rather than cloud computing.**

- Regulation

- Both HW, SW and personnel available already. There is already capacity available for use in the existing system. Electricity is too cheap to compare to the cost of cloud services.

- Customer requirement (perception of cloud being less secure)

- Internet network is poor. Afraid of having downtime because of the network.

- No expertise in cloud computing

- it is governmental organization, and they cannot reduce headcount when switching from traditional to cloud.

**[4.0] [20 points] Describe the concepts of vertical and horizontal scale. Describe 2 different ways in which you could scale a web application horizontally. Describe a potential architecture to scale the database to handle the scaling out of the web servers.**

Vertical : adding more power or storage space to existing machine or completely replacing machine with more powerful machine. Data sharing is less expensive as it uses reference within single machine. There is hard peak limit, once it is reached it cannot handle the load anymore.

Horizontal: is when we add more computers and use it as a network of computers to handle the load. No limit for peak handling. Redundancy is easy to create so less downtime.

There is no rule of when to choose what, but it is better to decouple the system so components can be chosen to be horizontally or vertically in scaling depending on the need.

Web server overload – Vertical scaling once it reaches its max capacity architecture remains the same but the capacity to handle more clients is increased. Single point of failure as it is on one server. Multiple server

Factoring sessions. Redis server with redundancy

Query expensive = add cache

Database failure – DB multiplication Master & slave

Queries are expensive

Sharding – multiple databases divide alphabetically.

Note: Some of the info was taken from Cloud Computing Duke University Resources