CS 524 Homework #1

1. Using the formulae for the first software business model, find the year where the cumulative support expense equals that of the initial licensing fee p, where p = $12,000 per user, and c = 0.40. In how many years will the initial cost of software becomes 5% of the overall expenditure?

   Given, cumulative support expense (CSE) = initial licensing fee
   where p = $12000
   c = 0.4
   CSE = pcm
   ie, p = pcm
   ie, 1 = cm ; c = 0.4
   therefore m = 1/0.4 = 2.5 years

   Now, given that Initial Cost = 5% of Overall expenditure
   ie, IC = 1/20 * OE
   ie, 20 * IC = OE
   We know that IC = pn, which implies OE = 20pn - I

   Now, using first software business model, we know that overall expenditure = np(1+cm) – II

   using I & II, we get, 20pn = np(I + cm)
   ie, 20 = 1 + cm
   ie, 19 = cm
   ie, 19 = 0.4m
   m = 47.5

   Therefore, in **47.5 years**, the initial software cost will become 5% of the overall expenditure.

2. Give three examples of each, SaaS, PaaS, and IaaS.

   SaaS: Gmail, Stevens Outlook, Google Docs
   PaaS: Office 365, Google App Engine, Heroku
   IaaS: Netflix (Built on AWS), Instagram (Built on AWS), VMWare

   (References: Cloud Computing Business Trends and Technologies, Class 02)

3. In the definition of Hybrid Cloud, a term “Cloud bursting” is mentioned. Search the Web for its definitions. Do these definitions agree? If so, provide what you think is the best definition (you can rephrase it as you see fit). If not, explain the differences between the definitions.

   The mix of at least one public and private cloud condition is called a Hybrid cloud. Although public and private cloud situations make up a hybrid cloud, they are one of a kind but separate elements, changing between them is encouraged by encrypted application programming interfaces that help transmit assets and remaining tasks at hand. The design is isolated however yet associated, this permits numerous ventures to run critical tasks in the private cloud, remaining burdens that are less delicate to be in the public cloud and pull resources from either environment when needed. When there is a peak in IT requests, few configuration steps have been set-up between private cloud and public cloud.
   The deployment model in which application runs in the private cloud, and later blasts into the public cloud when the demand in computing capacity increments is known as Cloud Blasting.
   After reading a few articles on the internet, I feel that Hybrid cloud and Cloud Bursting are related to each other.
   (References: https://www.redhat.com/en/topics/cloud-computing/what-is-hybrid-cloud)

4. What are the essential differences between the public and private cloud that have made CIOs worry about legal consequences of Shadow IT? Read the original text of the US Government acts mentioned in the text (HIPAA and SOX) and summarize each in one paragraph.

   Shadow IT refers to Information Technology (IT) systems within an organization that is neither monitored nor supported by the organization’s IT department whereas in case of a private cloud, all the tasks are monitored. It needs explicit approval from the IT department. Resources cannot be accessed with just permissions. The employees and the departments accessing the resources need to follow the set of rules and guidelines while accessing the resources over the cloud. Shadow IT can easily be compared to private cloud. Whereas in the case of a public cloud, each department and member can access the resources without any intervention from the IT department after the budget approval of the project to fulfil their requirements.

   HIPAA is the acronym for the Health Insurance Portability and Accountability Act that was passed by Congress in 1996. The portion of HIPAA addressing the ability to retain health coverage is overseen by the California Department of Insurance and the California Department of Managed Health Care. HIPAA was established for privacy & security provisions for safeguarding medical data. HIPAA has 4 purposes:

- Provides the ability to transfer and continue health insurance coverage for millions of American workers and their families when they change or lose their jobs.
- Reduces health care fraud and abuse.
- Mandates industry-wide standards for health care information on electronic billing and other processes.
- Requires the protection and confidential handling of protected health information.

SOX is the acronym for the Sarbanes–Oxley Act, also known as the "Public Company Accounting Reform and Investor Protection Act" and "Corporate and Auditing Accountability, Responsibility, and Transparency Act". SOX bill contains eleven sections, was enacted as a reaction to several major corporate and accounting scandals. The sections of the bill cover responsibilities of public corporation's board of directors and add criminal penalties for certain misconduct.

Many a Chief Information Officer (CIO) has observed this trend and understood that it is not enough just to implement virtualization in their data centers (often called Private Cloud, but really they were not that). The risks of Shadow IT are many, among them the loss of control over personnel. There are also significant security risks, since critical company data are now replicated in the Cloud. The matter of access to critical data (which we will address in detail in the Appendix) is particularly important, as it often concerns privacy and is subject to regulatory and legal constraints. For instance, the US

These considerations, under the threat of Shadow IT, lead CIOs to take new approaches. One is called Virtual Private Cloud, which is affected by obtaining from a Cloud provider a secure area (a dedicated set of resources). This approach allows a company to enjoy all the benefits of the Cloud, but in a controlled manner, with the company's IT being in full control of the security as well as costs. The service-level agreements and potential liabilities are clearly defined here.

The second approach is to build true private Clouds in the company's own data centers. The technology enabling this approach has evolved sufficiently, and so the vendors have started offering the full capabilities of a Cloud in software products. Example is the open-source project developing Cloud-enabling software, OpenStack. With products like that the enterprise IT departments can advance their own data center implementation, from just supporting virtualization to building a true Cloud, with services like those offered by a Cloud provider. These private Clouds provide internal services internally, with most of the benefits of the public Cloud (obviously with limited scale), but under full control and ultimately lower costs, as the margin of the Cloud provider is eliminated.

(References: Cloud Computing Business Trends and Technologies,
https://searchhealthit.techtarget.com/definition/HIPAA ,
https://www.dhcs.ca.gov/formsandpubs/laws/hipaa/Pages/1.00WhatisHIPAA.aspx &

https://en.wikipedia.org/wiki/Sarbanes%E2%80%93Oxley_Act)

5. Consider the case of the Instagram as described in the textbook. How many employees and customers did it have at the time of the purchase by Facebook? How much did Facebook pay for it? What was the value that the purchased business has generated in the first two years, and what were the factors that enabled generating this value?

   Instagram initially had 11 employees managing 30 million people. Facebook bought Instagram for one billion dollars. Instagram had no physical infrastructure, and only three individuals were employed to manage the infrastructure within the Amazon Cloud. There was no capital expense required, no physical servers needed to be procured and maintained, no technicians paid to administer them, and so on. This enabled the company to generate one billion dollars in value in two years, with little or no upfront investment in people or infrastructure. Most company expenses went toward customer acquisition and retention. The Cloud allowed Instagram to scale automatically as more users came on board, without the service crashing with growth.

   (References: Cloud Computing Business Trends and Technologies)

6. Familiarize yourself with the description of the Amazon Elastic Cloud Computing (http://aws.amazon.com/ec2/). What kind of a service model does it provide (i.e., SaaS, PaaS, IaaS, or a combination of these)? Please list the features that support your answer.

   Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud. It is designed to make web-scale cloud computing easier for developers using simple web service interface.
   Amazon Elastic Cloud Computing provides Infrastructure as a service (IaaS) model, then Platform as a Service.

   Amazon EC2 has the following features:
   Bare Metal instances: Amazon EC2 bare metal instances provide your applications with direct access to the processor and memory of the underlying server. These instances are ideal for workloads that require access to hardware feature sets (such as Intel VT-x), or for applications that need to run in non-virtualized environments for licensing or support requirements.
   Pause and Resume Your Instances: You can hibernate your Amazon EC2 instances backed by Amazon EBS, and resume them from this state at a later time. Applications that take a while to bootstrap and persist state into memory (RAM) can benefit from this feature.

GPU Compute Instances: Customers requiring massive floating point processing power will benefit from the next-generation of general-purpose GPU compute instances from AWS, Amazon EC2 P3 instances with up to 8 NVIDIA® V100 Tensor Core GPUs. P3 instances provide up to 1 petaFLOPS of mixed-precision, 125 teraFLOPS of single-precision and 62 teraFLOPS of double-precision floating point performance.
GPU Graphics Instances: Customers requiring high graphics capability will benefit from GPU graphics instances. The current generation GPU graphics instance, G3 instance, provides access to NVIDIA Tesla M60 GPUs, each with up to 2,048 parallel processing cores, 8 GiB of GPU memory and a hardware encoder supporting up to 10 H.265 (HEVC) 1080p30 streams and up to 18 H.264 1080p30 streams.
High I/O Instances: Customers requiring very high, low latency, random I/O access to their data can benefit from High I/O instances. High I/O instances are an Amazon EC2 instance type that can provide customers with random I/O rates over 3 million IOPS.
Dense HDD Storage Instances: Customers requiring very high storage density per instance, and high sequential I/O for data-intensive applications like Massively Parallel Processing (MPP) data warehouse, MapReduce and Hadoop distributed computing, and log and data processing can benefit from Dense Storage instances.
Optimized CPU Configurations: The Optimize CPUs feature gives you greater control of your Amazon EC2 instances on two fronts. First, you can specify a custom number of vCPUs when launching new instances to save on vCPU-based licensing costs.
Flexible Storage Options: Amazon EBS provides persistent, highly available, consistent, low-latency block storage volumes for use with Amazon EC2 instances. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability.
Flexible Scaling: Amazon EC2 enables you to scale resources up/down within minimal timelines via the web service that the API calls depending on the application requirements.
Full control: Root access to each instance enables you to have complete control over them.
Flexible hosting: Amazon EC2 allows you to choose your configurations so that are optimal to your configurations. You can pick from a wide range of instance types, operating systems and software packages.
Compatible with other AWS: Amazon EC2 in conjunction with other AWS such as S3 or RDS will shape up a complete computing solution around different areas.
Reliable: The environments are highly reliable with 99.95 percent availability commitment.
Secure: The environments are highly secure and robust. The inbound/outbound network access is controllable via the network ACLs and security groups.
Cost-effectiveness: Amazon EC2 allows you to pay as you go, thereby reducing upfront investments.

(References: http://aws.amazon.com/ec2/)

7. Consider the example of the Zing Interactive Media and explain how you would launch the same service today using Amazon EC2. Specifically list the steps (and costs) you would avoid by doing so.

   These are the steps involved in launching Zing media in the recent times:
   - Rent space on hosting site.
   - Purchase and install server
   - Lease dedicated T1 lines.
   - Purchase networking gear
   - Purchase and install software (OS, DB, etc)
   - Purchase networking gear for installation in the CAGE
   - Purchase and install load balancer
   - Hire an IT team of networking experts, system administrators, database administrator to maintain the same.

     (References: Cloud Computing Business Trends and Technologies)

8. Explain what CPU pinning is and how Intel supports it with API.

   CPU pinning otherwise known as Processor affinity enables the binding and unbinding of a certain process or thread to a particular CPU or a range of CPUs because of which the tasks will be executed on the designated CPU.

   An application needs a predictable amount of CPU power. Until recently, in the Cloud it could not be assured with fine granularity what an application would receive, which could be a major problem for real-time applications. Intel is contributing to solving these problems. Intel is providing API that allows the host to guarantee a certain percentage of the CPU to a given virtual machine. This capability, effected by assigning a virtual machine to a given processor or a range of processes—so-called **CPU pinning**—is exposed via the hypervisor and the Cloud provider's systems, and it can be consumed by the application.

   (References: Cloud Computing Business Trends and Technologies, Wikipedia)

9. Study the Amazon EC2 SLA. What service commitment (in percentage) does it guarantee? What is the bound on the downtime in a year?

   Amazon Compute Service Level Agreement is a policy governing the use of the Included Services and applies separately to each account using the Included Services. In the event of a conflict between the terms of this SLA and the terms of the AWS Customer

Agreement or other agreement with us governing your use of our Services (the "Agreement"), the terms and conditions of this SLA apply, but only to the extent of such conflict. Capitalized terms used herein but not defined herein shall have the meanings set forth in the Agreement.
AWS will use commercially reasonable efforts to make the Included Services each available for each AWS region with a Monthly Uptime Percentage of at least 99.99%, in each case during any monthly billing cycle (the "Service Commitment"). In the event any of the Included Services do not meet the Service Commitment, you will be eligible to receive a Service Credit as described below.
The included services are EC2, EBS, ECS & Amazon fargate for ECS. "Monthly Uptime Percentage" is calculated by subtracting from 100% the percentage of minutes during the month in which any of the Included Services, as applicable, was in the state of Unavailability. Monthly Uptime Percentage measurements exclude Unavailability resulting directly or indirectly from any Amazon Compute SLA Exclusion.

The bound on the downtime in a year is less than 1 hour and 12 minutes.

(References: https://aws.amazon.com/compute/sla/)

10. What is the "telecom-grade" service commitment? Who were the ETSI NFV Industry Specifications Group founders? List the areas where the NFV is expected to act. (Optional recommended reading: the ETSI NFV White Papers.)

    "Telecom grade" means that the hardware is (1) specifically engineered for running in telecommunications networks, (2) designed to live in the network for over 15 years, and (3) functional 99.999% (the "five nines") of the time (i.e., with about 5 minutes of downtime per year). This comes with a high cost of installation and maintenance of customized equipment. Especially when considering Moore's "law," according to which the computing power doubles every 18 months, one can easily imagine the problems that accompany a 15-year-long commitment to dedicated hardware equipment.
    Seven of the world's leading telecom network operators which are AT&T, BT, CenturyLink, China Mobile, Colt, Deutsche Telekom, KDDI, NTT, Telecom Italia, Telefonica,Telstra, and Verizon joined together to create a set of standards that were to become the framework for the advancement of virtualizing network services. On October 12, 2012, the representatives of 13 network operators worldwide published a White Paper outlining the benefits and challenges of doing so and issuing a call for action.

    The Network Functions Virtualization (NFV) movement is about radically transforming the "hardware-box-based" telecom world along Cloud principles. 52 other network operators—along with telecom equipment, IT vendors, and technology consultants—formed the ETSI NFV Industry Specifications Group (ISG) The areas where action was needed can be summarized as follows.

First, operational improvements. Running a network comprising the equipment from multiple vendors is far too complex and requires too much overhead (compared with a Cloud operator, a telecom network operator must deal with the number of spare parts—which is an order of magnitude higher).
Second, cost reductions. Managing and maintaining the infrastructure using automation would require a tenth of the people presently involved in "manual" operations. With that, the number of "hardware boxes" in a telecom network is about 10,000(!) larger than that in the Cloud operator.
Third, streamlining high-touch processes. Provisioning and scaling services presently require manual intervention, and it takes 9 to 18 months to scale an existing service, whereas Cloud promises instant scaling.
Fourth, reduction of development time. Introducing new services takes 16 to 25 months. Compare this to several weeks in the IT industry and to immediate service instantiation in the Cloud.
Fifth, reduction of replacement costs. The respective lifespans of services keep shortening, and so does the need to replace the software along with the hardware, which is where the sixth—and last—area comes in.
Sixth, reduction of equipment costs. (The hint lies in comparing the price of the proprietary vendor-specific hardware with that of the commodity off-the-shelf x86-based servers.

To deal with the above problem areas, tried-and-true virtualization and Cloud principles have been called for. To this end, the NFV is about integrating into the telecom space many of the same Cloud principles discussed earlier. It is about first virtualizing the network functions pertinent to routing, voice communications, content distribution, and so on and then running them on a high-scale, highly efficient Cloud platform.

(References: Cloud Computing Business Trends and Technologies)