**Ocean Lu**

Exam #2 (100 points)

This is a close-book test with one 8”x11” page (double side okay) of cheat sheet. Must take it in the classroom (must sign in at the class roster) and complete within the allocated time. Use a computer to type, however, cannot reference to any online material. Use blackboard to start and submit the exam, use MSWord to type, no other software should be used. You will automatically receive a score of 0 if any form of plagiarism found. Save the answer in a pdf file (e.g. Exam2\_lyang.pdf) and upload it onto blackboard along the Exam 2 link.

**Part I: Simple questions (True/False, Multiple Choice etc.) (40 points)**

1. Which of the following did NOT place an influence on the evolution of cloud computing?
2. utility computing
3. parallel/distributed computing
4. grid computing
5. **mobile computing**
6. I don’t own a computer but a game store near me hosts a powerful computer cluster with 8 desktops each equipped with 8 cores and a Nvidia GPU. I pay them $5 an hour to play a game and my game makes full use of these 8 desktop resources. This involves the following computing concepts (choose all apply)

a. distributed computing b. cluster computing c. grid computing **d. utility computing**

1. General speaking, cloud computing is most popular for \_\_\_\_\_ computing.

**a. data intensive**

b. computation intensive

c. network communication intensive

d. other, please specify \_\_\_\_\_\_\_\_\_\_\_\_\_

1. Which of the following is NOT an advantage of cloud computing?

a. Can access from anywhere b. keep data safe and secure

c. Easy to store/backup data **d. Unlimited scalability**

1. Which of the following is NOT an essential characteristic of cloud computing as defined by NIST?

a. On-demand self-service b. Resource pooling **c. Resource collaboration**

d. Rapid elasticity e. Measured service

1. True or **False**: PaaS is a subset of middleware.

**False**

1. The \_\_\_\_\_\_\_\_ cloud infrastructure is the best choice for the exclusive and joint use of organizations such as Boys Scout and Girls Scout.

a) Public b) Private c) Hybrid **d) Other, please specify** \_**Community Cloud**

1. **True** or False: Community and Hybrid are both cloud deployment models.
2. **True** or False: Hadoop is a new programming paradigm for parallel/distributed computing.
3. **True** or False: In Google cloud services, Chubby is a lock service for distributed coordination.
4. What is the architecture model of Hadoop?

(a) Peer-to-peer **(b) Master-Slave** (c) Distributed (d) none of above

1. HDFS mostly deals with one type of file, called “WORM”. What does WORM stand for?

WORM: \_\_\_**Write once read many**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. When a client writes a file, the system needs to update the replica too. When updating the replica what kind of structures used in linking the replica?

(a) master-slave **(b) pipelined**  (c) tree-structure (d) none of above

1. Spark is more efficient than Hadoop in supporting

a) Recursive data mining

b) Iterative and interactive MapReduce data processing

**c) TB sized data set processing**

d) none of above

15. True or False.

* 1. Clients of a PaaS service can manage the data on the server but not the application nor the operating system. **True**
  2. If a business doesn’t go with cloud services, it has to manage everything including application, data, middleware, networking, … all by itself. **True**

16. What level of service is NOT provided by AWS S3?

(a) SaaS (b) PaaS **(c) IaaS**  (d) other, please specify\_\_\_\_\_\_\_\_\_\_

17. True or False.

1. A hypervisor is a type of VMM. **True**
2. Full virtualization is more efficient than paravirtualization. **True**

**Part II: Short answer questions. (40 points)**

1. For each of the following cloud services, match it with the cloud service levels (SaaS, PaaS, …)
   1. Amazon AWS EC2

**IaaS**

* 1. MicroSoft Azure DevOps

**SaaS, Paas, IaaS**

* 1. Google App Engine

**PaaS**

1. A data center is full of computers. A WSC also contains many computers. What is the key difference between a (traditional) data center and a WSC in terms of the machines (e.g. not to talk about difference in networks etc.).

**The key difference between a traditional data center and a WSC in terms of the machine is that a traditional data center typically hosts many relatively small/medium sized applications, each running on a dedicated hardware infrastructure. A WSC usually belongs to a single organization, use a relatively homogenous hardware and system software platform, and share common systems management layer.**

1. Briefly describe the main functionality of the following:
   * + - 1. Google’s MegaStore

**Google’s MegaStore’s functionality is a scalable storage for online services and handles tons of transactions. It makes use of Bigtable.**

* + - * 1. AWS DynamoDB.

**AWS DynamoDB is a NoSQL database service for applications that need consistent access.**

4. In a typical MapReduce programming model on Hadoop,

(a) where is the input data to the Map task usually stored?

**Input data is stored in nodes on HDFS, with pre-loaded local input data.**

(b) where is the intermediate result (i.e. the output of the Map task) usually stored?

**The intermediate result (output of the Map task) is usually stored in a <key, value> format, set to their reducer.**

(c) where is the output of the Reduce task to be stored?

**The output of the Reduce tasks are stored locally, in output directory of HDFS.**

5. Spark vs. Hadoop.

(a) Name a drawback of Hadoop that Spark is designed to solve?

**A drawback of Hadoop that Spark is designed to solve is the processing speed.**

(b) Name the key feature introduced by Spark.

**Spark introduced in-cluster computing key feature.**

(c) Name a scenario that Spark normally performs better than Hadoop?

**Spark performs better than Hadoop in batch processing and streaming workloads, interactive queries, and machine-based learning.**

6. Given the figure showing service layers, what layer(s) can clients manage if using a

(a) DaaS service (Data as a service)

**Applications, Networks.**

(b) IaaS service

**Applications, Data, Runtime, Middleware, O/S.**



7. The following are a number of cloud architecture components.

(a) “A Specialized, high-speed network for data block transfers between computer systems and storage elements.” What is this called?

**Storage Area Network (SANs).**

(b) Name a popular topology used for DCN (Data Center networks).

**3 tier, Multiple-rooted tree.**

(c) Name one advantage of CDN (Content Delivery Networks).

**An advantage of CND is that it can deliver static content or live/on-demand streaming media.**

8. Virtualization.

(1) What is the purpose of virtualization?

**The purpose of virtualization is to allow users to operate in environment they are familiar with, and to simplify the management of physical resources for the abstractions.**

(2) What interface(s) does the virtualization simulate?

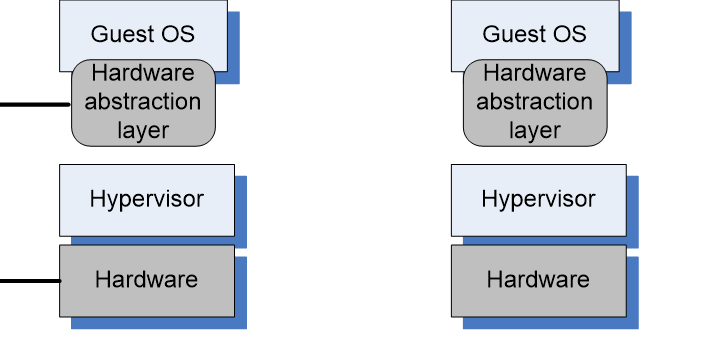
**Virtualization simulates multiplexing, aggregation, emulation, multiplexing and emulation. Has the interfaces Instruction Set Architecture (ISA), Application Binary Interface (ABI), and Application Program Interface (API).**

(3) Identify the type of VMS as illustrated in (b), (c), (d) of the following figure.

**The types of VM illustrated are (b) Traditional VMs (c) Hybrid VMs (d) Hosted VM**



(4) Identify each of the Virtualization methods as illustrated in the following figure.



**^Full virtualization**   **^Paravirtualization**

**Part III: Analysis questions or case studies. (20 points)**

1. Cloud security.
2. Discuss a scenario that involves security abuse/breach that particularly aimed to cloud services.

**A scenario that involves security abuse/breach aimed to cloud services is no service available without the internet.**

1. For the case in (b) briefly explain why non cloud services won’t suffer from such problem.

**Non cloud services won’t suffer from such problem because the data is accessible without the internet, although physical access is usually restricted.**

1. Describe a solution that could prevent the situation as discussed in above (b).

**Make sure the environment has a safe and consistent Wi-Fi availability.**

1. Cloud services and big data application. (Note: questions (a) to (c) are connected.
2. Assume we’d perform a big data analytics project on AWS cloud service. Where (on which cloud storage service) would you choose to store the big data set?

**I will choose to store the big data set on Amazon AWS DynamoDB.**

1. Assume you’ve just received an AWS account and the first step of your task is to put a huge data set on the cloud storage identified in (b), describe the step-by-step actions for transferring the big data set to the cloud storage.

**To create a table, add to the query, insert data and run queries, create indexes.**

1. What AWS service you’d use to perform Hadoop/Spark like operations on the big data set?

**AWS service I’d use to perform Hadoop/Spark like operations would be AWS EMR.**