University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

SEMESTER I EXAMINATION – 2015/2016

COMP 30520/41110

Cloud Computing

Prof. S. Dobson

Prof. P. Cunningham

Prof. M-T. Kechadi

Time allowed: 2 hours

Instructions for candidates

Answer any two questions. All questions carry equal marks. The paper is
marked out of 100.

.

Part A
Answer all questions of the following:

Question I (30 marks)

1. Common Cloud Service Models
    -   Cloud Service Models: Software as a service (SaaS), Platform as a service (PaaS) and Infrastructure as a service (IaaS).
    -   Software as a service (SaaS): a complete software application with a user interface.
    -   Platform as a service (PaaS): a platform within which developers can deploy their applications. A PaaS solution includes hardware (servers and disks), operating systems, development tools and administrative tools.
    -   Infrastructure as a service (IaaS): Provides machines storage and network resources that developers can manage by installing their own operating systems, applications and support resources.

2. Different types of Cloud Computing

    There are four primary Cloud Deployment Models: private-cloud, public cloud, community cloud and hybrid cloud.
    -   Private Cloud: Owned by a specific organisation and normally used only by that organisation. The underlying technology may reside on or off site.
    -   Public Cloud: Normally owned by a large organisation or company offering cloud services. It is available for use by the general public. A public cloud is usually the least expensive solution.
    -   Community Cloud: The cloud is shared by two or more organisations, typically with shared concerns.
    -   Hybrid Cloud: A cloud that consists of two or more private, public or community clouds.

3. Security disadvantages of using a cloud-based solution

    -   Country or jurisdiction: If the cloud resources reside in multiple states, questions of jurisdiction may arise in the event of a legal matter.
    -   Multitenant risks: An application error might expose one company's data to another company. If a data storage device is shared, data remnants from one company may be exposed to another company.
    -   Malicious insiders: Depending on the employee's role, a company's cloud-based data may be at risk.
    -   Vendor lock in: It may become difficult for the company to change providers latter in the event of a service-level agreement breach or other problem.
    -   Risk of the cloud-based provider failing: Companies who rely on cloud-based providers are at risk that the provider could fail.

Question II (10 marks)

In the context of Cloud Computing, answer the following questions:

1. What does "as a service" offer? (5 marks)
   - Large scalability
   - Multi-tenancy, which allows resources to be shared by many users
   - Device independence, which allows users to access the systems on different hardware.

2. What are the advantages of PaaS? Give two examples of PaaS (5 marks)
   - Lower total cost of ownership
   - Lower administrative overhead
   - More current system software
   - Increased business and IT alignment
   - Scalable solutions

   Example: Amazon Web Service (AWS) Google App Engine, Force.com solution from Salesforce, etc.

Question III (30 marks)

1. What are the most common forms of Virtualisation? Describe the characteristics of each form. (10 marks)

   The most common forms of Virtualisation are server virtualisation, desktop virtualisation, virtual networks and virtual storage.

   Server virtualisation: making one server appear as many. Each virtual server may run the same or different operating systems. Server virtualisation provides greater CPU utilisation, a smaller equipment footprint, less power consumption and support multiple operating systems.

   Desktop virtualisation: This allows a user to switch between multiple operating systems on the same computer. Desktop virtualisation provides support for multiple operating systems that is very convenient for software developers, testers and help desk support staff.

   Virtual networks: Virtual networks are sometimes called virtual private networks (VPN). Using VPN, users can connect to a network and access the network resources from any Internet-connected computer. Virtual networks also allow network administrators to segment a network making different departments appear to have their own separate networks.

   Virtual storage: This provides users with access to scalable and redundant physical storage through the use of abstract or logical disk drives or file systems or a database interface.

2. Give the most common technologies of Virtualisation and describe characteristics of each technology. (10 marks)

The most common technologies of Virtualisation are Full virtualisation, Para-virtualisation, Hardware-assisted virtualisation and OS-level virtualisation.

Full virtualisation: The host emulates a complete installation, including hardware layer, for each guest.
Para-virtualisation: Host provides a VM interface for the guests and that the guests access hardware through that host VM
Hardware-assisted virtualisation: Hardware provides support to run instructions independently for each OS.
OS-level virtualisation: Same OS in host and guests. Isolation in the user land. VM must all be running the same OS and the same version of that OS.

3. List three virtual computer devices. (5 marks)

Virtual CPU, Virtual Memory, Virtual Disk

4. List two applications that are NOT well suite for Virtualisation. (5 marks)

Application with unique hardware requirements
Graphics-intensive applications

Part B
Answer any one question of the following:

Question IVa: (30 marks)

Suppose there are two text files:
- logs that contains hostname and IP address accessed
- hostsinfo that contains hostname and username

For example:
    hostsinfo
        ccd-1, Anita
        ccd-2, Jack
        ccd-3, Anita
        ccd-4, Jim
        ….

    logs
        ccd-1, 192.168.0.1
        ccd-2, 192.168.0.5
        ccd-2, 192.168.0.1
        ccd-3, 192.168.0.1

ccd-1, 192.168.0.5
ccd-2, 192.168.0.7
ccd-2, 192.168.0.10
…….

Write a Pig script that finds the top 5 most IP address accessed by user Anita and stores this information in a file named top5sites.

```
Users    = LOAD 'hostsinfo' AS (hostname, username);
Filtered = FILTER Users BY
                 username ='Anita';
Pages    = LOAD 'logs' AS (host, ips);
Joined   = JOIN Filtered BY hostname, Pages by host;
Grouped  = GROUP Joined BY ips;
Summed   = FOREACH Grouped GENERATE group,
                 count(Joined) AS ipa;
Sorted   = ORDER Summed BY ipa desc;
Top5     = LIMIT Sorted 5;

STORE Top5 INTO 'top5sites';
```

Question IVb: (30 marks)

K-Means is a simple clustering algorithm that aims to partition a set of objects into k clusters in which each object belongs to the cluster with the nearest mean. K-Means algorithm has 4 steps:

- Step 1: Given k, partition objects into k nonempty subsets
- Step 2: Compute seed points as the centroids of the clusters of the current partition The centroid is the centre (mean point) of the cluster
- Step 3: Assign each object to the cluster with the nearest seed point
- Step 4: Go back to Step 2 until no more new assignment

Using Map/Reduce model, define the mapper and reducer of K-Mean algorithm.

nodeID: identification of each object
nodeIDs: all nodeID

centroidID: identification of each centroid point of each cluster
centroidIDs: all centroidID

Mapper:
        Input: <nodeID, [centroidIDs]>
        Emit: <centroidID, nodeID>
 Reducer:
         Input:
                <centroidID, nodeID>

Reduce -> <centroidID, [nodeIDs]>
Recompute centroid position from positions of nodes in it.
Emit:
<centroidID, [nodeIDs]>
<centroidID, [nodeIDs]> -> <nodeID, [centroidIDs]>
Repeat until no change