Lebanese University
Faculty of Sciences 1

INFO448 E

Cloud Computing

Final 2023 - 2024
Session 1
Duration: 1h30

Question I – Cloud Computing [50/100 pts]

1. You have a company and you have an information system to deploy. When do you choose to move to the cloud? (5 pts)
2. Assume now that you have a server and two virtual machines are created on this server using VMWare.
    a. Why do we call it a virtual machine? (5 pts)
    b. In order for the server to be virtualized and accessed by several VM, what is the main condition for the server to satisfy? (5 pts)
    c. Is the above virtualization called OS or Hardware virtualization? Explain. (5 pts)
    d. Given that the virtualization is hosted – full virtualization. How does an instruction access the CPU hardware? (5 pts)
    e. In case the instruction to be run is critical, how does the call access the hardware, explain the whole mechanism. (5 pts)
    f. What is the part of the server responsible of managing the memory access of a virtual machine? (5 pts)
3. Give an example of one PaaS, one IaaS, one SaaS. When will you need to use each one of them? For each one of them, who is the Cloud provider and who is the Customer/user. (5 pts)
4. Cloud Storage:
    a. List three kinds of risks that exist in the architecture of cloud storage models. How you can protect your cloud against them? (6 pts)
    b. What are the two types of storage that exist in in the cloud? Explain them briefly. (4 pts)

Question II – Big Data [15/100 pts]

1. Explain the key components of the Hadoop ecosystem and their roles in the data processing pipeline.
2. Discuss the MapReduce programming model and its advantages in processing large-scale data.
3. Illustrate the concept of data shuffling and sorting in MapReduce. Explain why it is a crucial step in the MapReduce process.

Question III – Big Data [15/100 pts]

Consider a scenario where you have been given a large dataset containing information about customer transactions for an e-commerce platform. The dataset consists of records with the following attributes: transaction ID, customer ID, product ID, purchase quantity, and timestamp.

Using Hadoop MapReduce, design and implement a solution to Calculate the average purchase quantity for each product.

Question IV – Big Data [20/100 pts]

1. Explain the purpose and functionality of Hadoop Distributed File System (HDFS). How does it provide fault tolerance and data reliability?
2. Discuss the role of NameNode and DataNode in HDFS architecture. How do they collaborate to manage the file system?
3. Describe the Hadoop cluster setup process, including the configuration of master and worker nodes. What factors should be considered when determining the cluster size?
4. Discuss the benefits and challenges of processing unstructured data using Hadoop MapReduce.