THE UNIVERSITY OF MELBOURNE

Student Number_____

Faculty/Dept. Computing and Information Systems

Subject Number COMP90024

Subject Name Cluster and Cloud Computing

Writing Time **2** **hours**

Reading Time **15** **minutes**

Open Book Status Closed Book

Number of pages (including this page) 4

Authorised Materials: None

Instructions to Students: This examination is worth 50% of your final mark

Answer 5 out of any 7 questions. Please note that only the first 5 questions will be marked.

Each question carries 10 marks.

The number in square brackets after each sub-question represents the marks allocated to it.

Instructions to Invigilators: Please provide students with standard script books

No calculators are allowed

This paper is NOT to be made available in the library after the examination

Paper to be held by Baillieu Library: Indicate whether the paper is to be held with the Baillieu Library.

Yes ☐ No X

Extra Materials required (please tick & supply)

Graph Paper ☐ Multiple Choice form ☐

Question 1:

A) Define Gustafson-Barsis' law for scaled speed-up of parallel programs. [1]

B) A parallel program takes 240 seconds to run on 24 processors. The total time spent in the sequential part of the program is 12 seconds. What is the scaled speedup? [2]

C) According to Gustafson-Barsis' law, how much faster could the application *theoretically* run if it ran across all 24 processors compared to running on a single processor? [2]

D) Why is *theoretically* italicized in Part C)? [2]

E) The message-passing interface (MPI) is often used for parallel programming on high performance computing systems. Describe four methods that are commonly found in MPI programs and explain their functionality. [3]

Question 2:

A) *Big data* is often associated with data having a range of properties including high volume, high velocity and high variety (heterogeneity). Discuss the advantages, disadvantages and suitability more generally of the following data solutions with regards to these big data properties:

   a. CouchDB [2]
   b. Apache Hadoop Distributed File System (HDFS) [2]
   c. Apache Spark [2]

B) What is the Apache Hadoop Resilient Distributed Dataset (RDD) operation type that triggers RDD evaluations? Which operation type does *not* trigger RDD evaluations? [2]

C) CouchDB views whose map part is defined using a composite key can be used to aggregate data at different levels: how can a user request different aggregation levels via the CouchDB HTTP API? [2]

Question 3:

A) Representational State Transfer (ReST) based web services are often used for creating *Resource-oriented Architectures* (ROA) whilst Simple Object Access Protocol (SOAP)-based web services are often used to implement *Service-oriented Architectures* (SOA). Discuss the similarities and differences between a ROA and a SOA. [3]

B) Discuss the advantages and disadvantages of ReST vs SOAP for web services more generally. [5]

C) HTTP methods can be *safe* or *idempotent*. What is meant by the italicized terms, and give an example of each? [2]

Please Turn Over

Question 4:

A) Popek and Goldberg laid down the foundations for computer virtualization in their 1974 paper, *Formal Requirements for Third Generation Architectures*.
   a. Identify and explain the different types of classification of instruction sets for virtualization to occur according to Popek and Goldberg. You should include the relationships between the instruction sets. [2]
   b. Describe how these principles are realized by modern hypervisors. [2]
   c. Explain the differences between *full virtualization* and *para-virtualisation*. Give an example of a hypervisor that uses full virtualization and an example of a hypervisor that uses para-virtualisation. [2]
   d. Container-based solutions such as Docker offer a lighter-weight approach to virtualization.
      i. Describe the advantages and disadvantages of using Docker over other full virtualization technologies. [3]
      ii. What is the relationship between a Docker Image and a Docker Container? [1]

Question 5:

A) Code versioning systems are frequently used in collaborative software development activities. Name three types of architectures that code versioning systems have adopted and give one example of a solution for each with their respective advantages and disadvantages. [3]
B) Give a short explanation for the following terms that are often used in a code versioning context:
   a. Commit [1]
   b. Checkout [1]
   c. Branch [1]
   d. Tag [1]
   e. Rebase [1]
C) What is the main difference between the *clone* and *checkout* commands? [2]

Question 6:

A) The NeCTAR Research Cloud is based on the openStack technology.
   a. Describe the role and features of the following openStack components:
      i. Nova [1]
      ii. Swift [1]
      iii. Glance [1]
      iv. Keystone [1]
   b. Describe the interplay between these components that allows a researcher to create an instance of a virtual machine through a pre-existing snapshot. [3]
B) The NeCTAR Research Cloud has multiple *availability zones*.
   a. What is meant by the term: availability zone? [1]
   b. What are the implications of availability zones with regards to virtual machine instance creation and data volumes offered by NeCTAR? [2]

Please Turn Over

Question 7:

A) The NeCTAR Research Cloud focuses primarily on offering Infrastructure-as-a-Service (*IaaS*) capabilities, however many research communities require Software-as-a-Service (*SaaS*).

   a. Discuss the relationship between the italicized terms: *IaaS* and *SaaS*. [3]
   b. Applications can be deployed across Clouds either through creation and deployment of virtual images (snapshots) or through scripting the installation and configuration of software applications. What are the benefits and drawbacks of these approaches? [3]
   c. Describe the approach that would be taken using Ansible for scripted deployment of SaaS solutions onto the Cloud. [2]
   d. Describe the approach that would be taken using the openStack Heat service for deployment of SaaS solutions onto the Cloud. [2]

--- END OF EXAMINATION ---