

Homework B

Due by Midnight (IST) of Fri 10 Apr, 2015. Total of 100 points.

Submit the homework solution as a PDF file by email with the subject line (and file name)

SE252_JAN2015_HW-B_StudentName.

Be succinct. The points allocated to each question gives you a sense of the depth of explanation required.

State assumptions you are making. Always provide citations/footnotes for external sources of information.

Always provide citations for external sources of information and avoid verbatim use of external content!

(I) ILO3: Algorithms and Programming Patterns [35 points]

- 1) Give the pseudocode for finding the shortest path between a given source and a destination using Giraph/Pregel? Assume each vertex has a unique ID and edge has a weight of 1. [15 points]
- 2) Discuss the types and the degrees of parallelism that MapReduce and Pregel exploit for processing tuple and graphs, respectively. Characterize their potential speedup when running on multiple Cloud VMs, and what the limiting factors would be in achieving perfectly linear speedup. [20 points]

(II) ILO4: Scheduling [65 points]

- 3) Similar to job scheduling, where jobs need to be mapped to physical machines, Cloud service providers have to provision VMs on physical machines in their data centre (Sec. 4.5 from textbook). Discuss some of the resource provisioning approaches that are taken by Cloud providers. [15 points]
- 4) You are a software architect at an IaaS Cloud service provider, VM4Less Inc., which offers end-users with several VM Instance types, each with a specific number of cores. Your job is to develop an algorithm to provision these VM instance requests from the end-users onto physical machines in your data centre, each machine having a fixed number of cores available. How would you use a List Scheduling algorithm to decide which VM request to place on which machine? What is the impact of this algorithm on the machine utilization for the data centre, operational costs of your company, and the VM performance for the end-user? [20 points]

- 5) Grid computing introduced the notion of utility computing, and shares similarities with Clouds. "Workflow Scheduling Algorithms for Grid Computing", Yu, et al¹, offers a taxonomy for scheduling techniques for Grids. Which of these are also relevant for Clouds? What characteristics of Clouds are not covered by the scheduling techniques presented in the paper? [15 points]

- 6) A scientist has a dataflow application, composed as a Directed Acyclic Graph. She has been running this on a cluster using the HEFT scheduling algorithm² to allocate resources, but now wants to run the application on a private Cloud. Would you recommend that she continues to use HEFT to schedule her DAG on the Cloud? Why/Why not, and if not, what alternatives would you recommend? [15 points]

¹ Workflow Scheduling Algorithms for Grid Computing, Jia Yu, Rajkumar Buyya and Kotagiri Ramamohanarao, in *Metaheuristics for Scheduling in Distributed Computing Environments*, 2008, www.cloudbus.org/papers/MHS-Springer-Jia2008.pdf

² Performance-effective and low-complexity task scheduling for heterogeneous computing, H Topcuoglu, S Hariri, M Wu, *Transactions on Parallel and Distributed Systems*, 2002.