

Name: Jingqi Li

USC ID: 1207830860



$\frac{10}{10}$

INF 553 – Fall 2015 [Morning section]

Quiz 1 (10 points)

There are 7 questions in total. For questions 1–6 (each worth 1 point), circle all answers to the question to get the full credit. No partial credits will be given. Question 7 is worth 4 points.

abcd 1. In data mining, the discovered patterns and models are required to be:

- a. Valid
- b. Useful
- c. Unexpected
- d. Understandable

b 2. Which of the following unit of bytes is roughly 1000 (or 1024 to be exact) of a gigabyte (GB)?

- a. Megabyte (MB)
- b. Terabyte (TB)
- c. Petabyte (PB)
- d. Exabyte (EB)

abc 3. According to the lectures, which of the following are the major challenges to large-scale computing (i.e., processing of a large amount of data) using a distributed computing architecture?

- a. How to distribute the computation?
- b. How to make it easy to write a distributed programs?
- c. How to handle machine failures?
- d. Where to buy cheap storage devices such as hard drives?

b 4. What is the typical size of chunks (which files are divided into for distributed computation)?

- a. 16-64KB
- b. 16-64MB
- c. 16-64GB
- d. 16-32GB

ab 5. Which of the following functions are typically provided by the MapReduce programmer? Recall shuffle is to distribute the output of Map tasks among the Reduce tasks.

- a. Map
- b. Reduce
- c. Group by
- d. Shuffle