COMP9313
Template and Example Questions

Note: This is just a template for the final exam, including example questions for each part of the exam. The final exam will include more questions (which may be different from the ones below) covering topics presented/included in lectures, labs and assignments. The final exam is lab-based.

PART I – Multiple choice questions

In this part, choose one option that best answers the question.

Question 1

Which of the "Vs" listed below were the three first Vs proposed to characterize big data?
[A] Value, Volume, Velocity.
[B] Volume, Velocity, Viability.
[C] Validity, Variability, Volume.
[D] Velocity, Volume, Variety.

Question 2

The Big Data Process involves a number of phases and activities. The success of a Big Data project can be guaranteed only when such process is executed end-to-end and sequentially (no iterations involved) from acquisition and recording, to interpretation.
[A] True.
[B] False.

Question 3

Which of the statements below is false about Data Curation?
[A] Data Curation is a one-off task that happens early in the big data process.
[B] Data Curation typically deals not only with removing/fixing errors from data but also with other tasks such as data enrichment.
[C] In the context of Big Data, Data Curation needs to take scalability issues into account.
[D] All of the above

PART II – Short answer questions

In this part, write down a short answer to each question.

Question 4

List 3 (three) NoSQL technologies / tools that follow the document-based model.
_____
_____
_____

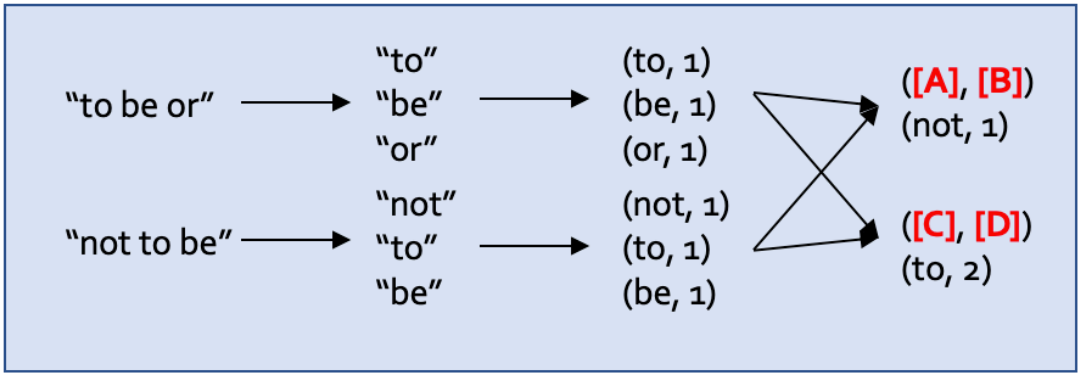
Question 5

Consider the Scala / Apache Spark code snippet below and a text input file containing the sentences "*to be or*" and "*not to be*". Provide the values for [A], [B], [C] and [D] in the image (output) shown below:

```
val file = sc.textFile("hdfs://...")

val counts = file.flatMap(line => line.split(" "))
    .map(word => (word,1))
    .reduceByKey(_ + _)

counts.saveAsTextFile("hdfs://...")
```



(*provide the values for [A], [B], [C] and [D] in the lines below*)

_____
_____
_____
_____

PART III – Written answer questions

Keep your answers clear and coherent. You should directly address the question

Question 5

In the context of Big Data Management, explain the reasons why it is important to take Big Data Security into account.

_____
_____
_____
_____
_____
_____
_____

Question 6

In the context of Apache Spark, explain what Lineage Graphs are and what they are useful for.