Part 2: Online quiz (15 marks, 1 hour)

Instructions: Answer all questions.

1. Identify a paragraph with more than 30 words from any article online. State the source of the paragraph, and demonstrate how term extraction work to generate index from the identified paragraph.

(5 marks)

2. Discuss two factors that can reduce the chances for an HDFS to suffer data lost with real life examples.

(5 marks)

3. Discuss a practical solution to ensure data provenance and trustworthiness in a big data solution.

(5 marks)

Submissions:
The student is expected to submit answers to each question individually, and submit the document in PDF format. The student can include online materials, screenshots and/or codes to support your answer.