COMP4057 Distributed and Cloud Computing

Written Assignment

Due date: 21 April, 2020 (23:59)

Answer ALL questions.

Please add this declaration in your submission: "I hereby declare that all the work done in this COMP4057 written assignment is of my independent effort."

Question 1. The transactions **A** and **B** are defined as follows:

A: *write(x, 5); write(y, 6);*

B: *x = read(x); write(y, 4);*

x and *y* are integer variables of an object in a distributed system.

(a) Draw an interleaving of the transactions **A** and **B** in which locks are released early with the effect that the interleaving is NOT serially equivalent.

(b) Draw an interleaving of the transactions **A** and **B** using two phase locking.

(c) Explain all possible outcomes that transaction **A** crashes after executing *write(x, 5)*.

(d) Use transactions **A** and **B** to illustrate optimistic concurrency control. Explain the best case and the worst scenario of optimistic concurrency control.

Question 2. The email servers for @comp.hkbu.edu.hk are an on-premise resource of the Department. With respect to the characteristics of the cloud, namely, on-demand usage, ubiquitous access, multitenancy, elasticity, measured usage and resiliency, write a small essay (fewer than 600 words) to support or argue against hosting the email servers or email service on Google cloud.

Remarks: not all of the six cloud characteristics are relevant to the answer of this question.

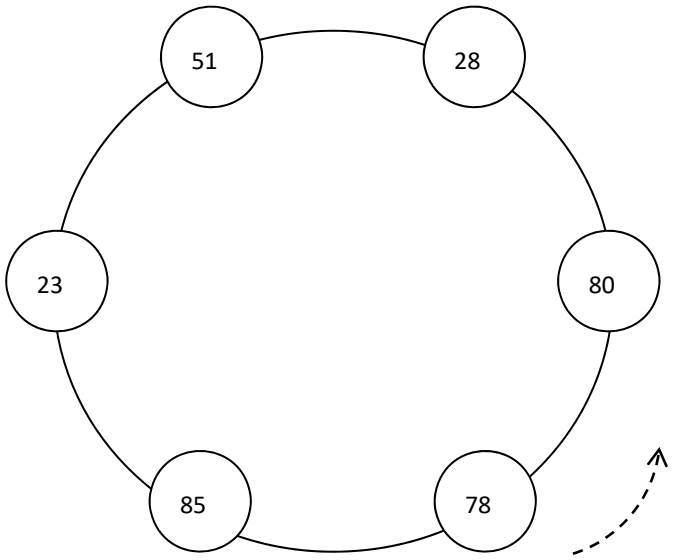
Question 3. Answer the following short questions with respect to cloud characteristics.

(a) You are asked to design the architecture of a Massively Multiplayer Online Game. What scaling method(s) would you adopt? Justify your answer.

(b) You are asked to run a grabber program on the cloud that collects climate data hourly. Would you recommend to take the failover option to run the program? Justify your answer.

(c) You are asked to set up a database server for a small/medium enterprise. Explain three advantages and three disadvantages of a cloud solution in contrast to an on-premise solution.

Question 4. A Ring-based Election Algorithm is used to elect a master/coordinator from a group of processes arranged in a logical ring. Suppose messages can only be sent in an anti-clockwise direction along the ring.



The process ID's are shown in the above diagram. Assume that the process with the highest ID is going to be selected. If process 28 begins the election process by sending out an election message, how many messages are needed to finish the election algorithm. When you count the number of messages, you must explain your steps involved.

(*Hint: these terms may help: coordinator, participant, non-participant, election message, elected message*)

Question 5. The environmental protection department of HKSAR has released air quality monitoring data. In particular, the air pollution index (1995 – 2013) is publicly available [1].

Tasks: You are required to write map-reduce *pseudo-codes* to finish the following 2 tasks. Explain the logic clearly.

(a) You are given the raw data of the monitoring stations of each district from 1995-2013, e.g., xxx.txt, where xxx is a district (e.g., Sham Shui Po). The format of xxx.txt is shown Figure 1 below. Each line starts with the district xxx, a date and the Air Quality Index (AQI) of each hour (from hour 0 to hour 23) of the day. The value **-1** indicates the index is not valid. Write map-reduce pseudo-code to convert the input file to the corresponding air pollution level, according to Table 1. Assume to put the outputs of your pseudo-code into xxx_lv.txt.

```
ShamShuiPo:19990701 21    21    … 14   14
ShamShuiPo:19990702 14    13    … 20   20
…
ShamShuiPo:20131231 87    85    … -1   58
```

Figure 1. The format of Sham Shui Po.txt

Table 1 AQ1 vs Air Pollution Level

AQI	Air Pollution Level
0-50	Good
51-100	Moderate
101-150	Unhealthy for Sensitive Group
151-200	Unhealthy
201-300	Very Unhealthy
300+	Hazardous
-1	Invalid data

(b) The stations at Sham Shui Po and Kwun Tong are the nearest stations of the districts Kowloon City and Wong Tai Sin. We want to know whether or not the AQI of these stations mostly agree with each other. Write map-reduce pseudo code to output the respective percentages of the following four cases between Sham Shui Po and Kwun Tong. Assume the input is the xxx_lv.txt files output from Question 5(a).

(1) no discrepancy,
(2) discrepancies of 1 level,
(3) discrepancies of >=2 levels, and
(4) either or both locations have invalid reading

References:

[1] Environmental Protection Department, the Government of the Hong Kong SAR. Air Pollution Index (1995 -2003). Available at http://www.aqhi.gov.hk/en/related-websites/air-pollution-index.html. Last accessed at Feb, 2015.