Autumn 2019

Homework 3

Big Data Management

1. Refer to the book at http://gephi.michalnovak.eu/Mastering%20Gephi%20Network%20Visualization.pdf
Explain the meaning for the following different layouts in Gephi. (2.5 marks)

- Force Atlas
- Fruchterman-Reingold
- Radial Axis
- Yifan Hu
- ARF

2. Select the correct HDFS commands. (Select 3 choices) (0.5 mark)

A. List files in / *hdfs dfs –ls /*

B. Make a directory *hdfs dfs –mkdir /user/test*

C. Runs the HDFS filesystem checking utility *hdfs fsck /user/test*

D. Copy the file from a local file system to HDFS *hdfs dfs -copyToLocal <hdfs source> <localdst>*

3. Check answers that apply when replication in HDFS is lowered. (Select 3 choices) (0.5 mark)

A. HDFS is less robust

B. Less likely that data will be local to more workers

C. Aggregate I/O rate will be worse

D. HDFS will have more space available

4. Consider the following Student table: (4 marks)

CREATE TABLE Student

(

StudentId int NOT NULL PRIMARY KEY,

Firstname varchar(255),

Lastname varchar(255),

Major varchar(255),

GPA int

)

(1) Write a SQL query to display the first name and lastname of all students.

(2) Write a SQL query to display all columns of students whose Major is not "CS".

(3) Write a SQL query to display all columns of students whose firstname starts with "J" and lastname starts with "S".

(4) Write a SQL to change student 101's major to "CS"

5. Given the relational schemas R(ABC) and S(CDE), let r(R) and s(S) be the relations corresponding to R and S respectively as the following. Write the results for the following different operators. (2.5 marks)

R

A	B	C
a1	b1	c1
a3	b3	c3

S

C	D	E
c1	d1	e1
c2	d2	e2

a) Cartesian product of R and S, R X S

b) Results of R join S, $R \bowtie_{R.C = S.C} S$

c) Results of R semi join S, $R \ltimes_{R.C = S.C} S$

d) Result of R left outer join S, $R \rtimes_{R.C = S.C} S$

e) Results of R full outer join S, $R \bowtie_{R.C = S.C} S$