INFS3200 Advanced Database Systems

Tutorial 9: Advanced Database Applications

Semester 1, 2021

Question 1: What kind of values can be used as the key in key-values storage? **(Choose one answer)**:

a) Any kind of values
b) Values that must be globally unique
c) Values that must not be empty (i.e., values cannot be optional)
d) Above b) and c).
e) Must be unique numerical values

Question 2: In data processing **(choose one answer)**:

a) The higher the data dimensionality is, the better for data processing algorithms.
b) The lower data the dimensionality is, the better for data processing algorithms.
c) We must consider the application problems to decide if we need higher or lower data dimensionality (i.e., feature selection).

Question 3: What is the "curse of dimensionality"? What is the impact of dimensionality to the data processing?

Question 4: For each of the following algorithms, answer: what kind of Big Data problems can it be applied? (1) "from-small-to-big", (2) from-big-to-small", (3) "Knowing unknown". In order to answer these questions, you may need to Google the meanings of these algorithms.

a) Pattern Recognition algorithms
b) Data Linkage algorithms
c) Classification algorithms
d) Clustering algorithms
e) Ranking algorithms
f) Recommendation algorithms
g) Skyline database query
h) PCA (Principle Component) algorithms
i) Time series algorithms
j) Ontology discovery algorithms
k) Graph / Topology algorithms
l) Relationship discovery algorithms: (e.g., correlation, association, causality, dependency)
m) Network centrality algorithms
n) Random Walk algorithms
o) Simulated Annealing Algorithms

Question 5:

Discuss the differences of SQL and NOSQL languages and their applicability respectively.

Question 6: Given the follow relational table Sensors, an SQL statement:

```
SELECT avg(temp), time
FROM sensors GROUP BY time;
```

is used to find out the average temperature that the sensors have detected within an environment. However, one may notice that there are some abnormal temperatures which have caused the average temperature becoming unexpectedly high.

Sensors

Tuple id	Time	SensorID	Voltage	Humidity	Temp.
T1	11AM	1	2.64	0.4	34
T2	11AM	2	2.65	0.5	35
T3	11AM	3	2.63	0.4	35
T4	12PM	1	2.7	0.3	35
T5	12PM	2	2.7	0.5	35
T6	12PM	3	2.3	0.4	100
T7	1PM	1	2.7	0.3	35
T8	1PM	2	2.7	0.5	35
T9	1PM	3	2.3	0.5	80

(1) Construct a predicate expression ***P*** to be used in an SQL WHERE clause in order to find out the outliers that have caused the unexpected high average temperature of sensors. So this SQL query is able to explain an abnormal phenomena. The SQL statement is in the format of:

```
SELECT avg(temp), time
FROM sensors GROUP BY time
WHERE P;
```

(2) Describe a step-by-step procedure that can be applied to automatically give the explanations for outliers in aggregation queries of a relational database.

Question 7: What are the differences of the Row-based and Column-based database storages? Describe the advantages of each of them respectively.

Question 8: What is the Scale-Free networks? List some examples for the scale-free networks.

Question 9: List the possible computations that can be applied on complex networks.

---ooo000O000ooo---