**Database Management, 2020 Final**

**Q1. (8%)**

Consider the following ER diagram for part of an Olympic Game database. Each Olympic Committee holds multiple games and gives out several medals. Every athlete must participate at least one game. Each game is participated by at least two athletes. Each medal is won by a single athlete. (1) Specify the cardinality ratio of each relationship in this ER diagram. (2) Derive the relational database schema from this ER diagram using ER-to-Relational Mapping.



**Q2.**

1. (4%) Mapping the EER diagram of the following Figure to relations by using option 8d.
2. (4%) Draw the set diagram of the following figure.

FARM

Captive

Herding

Agriculture - Fruit

Agriculture - Vegetable

d

O

Livestock farming

**Q3.**

1. (8%) Use examples to explain why transitive dependencies and partial dependencies are considered bad in a relational schema.
2. (3%) Discuss the problem of spurious tuples and how we may prevent it.
3. (3%) X is a key in a relation R. Prove that X 🡪 A for any attribute A in R, according to the definition of functional dependency.
4. (3%) Given the functional dependencies K 🡪 Y and Y 🡪 Z, where K is a candidate key, Z is a non-prime attribute and Z is transitively dependent on K. Explain why the dependencies are problematic when Y is not a candidate key. Explain why the dependencies are not problematic when Y is a candidate key.

**Q4.**

(a) (2%) Define the general definition of 3NF, which considers all keys of a relation.

(b) (6%) Derive the undesirable dependencies when a functional dependency X 🡪 A violates 3NF. You need to clearly explain how the undesirable dependencies are derived from X 🡪 A violating the conditions of 3NF.

(c) (4%) How does **Boyce-Codd Normal Form (BCNF)** differ from 3NF? Explain why 3NF adopts a weaker condition than BCNF does.

**Q5**. (8%) Consider the relation: R(A, B, C, D, E, F, G, H, I, J, K, L,M,N,O,P,Q). {A, B} is the primary key. {C,D,E} forms a candidate key. The following functional dependencies exist among the attributes of the relation:{A} → {F,G}; {C,D} →{H,I}; {F} → {J,K}; {D} → {L,M}; {M} → {N,O,P}

Transform this relation into 2NF, then 3NF relations.

**Q6.** (8%) Consider the relation: R(A, B, C, D, E, F, G, H, I, J, K, L, M, N, O). {A, B, C} is the primary key. {E,F} forms a candidate key. The following functional dependencies exist among the attributes of the relation: {B, C} →{D}; {C} → {G, H}; {D} → {K,L}; {G} → {I,J}; {E} → {M,N}

Transform this relation into 2NF, then 3NF relations.

**Q7.** Consider the relation R = {A, B, C, D, E, F, G, H, I, J, K, L, M, N} and the set of functional dependencies F = {{A, B, C} → {L}, {B} → {D, E, F}, {A, C} → {M, N}, {D, E} → {I, J, K},

{D}→ {G, H}}.

(a) (2%) What is the key for R?

(b) (6%) Decompose R into 2NF, then 3NF relations.

**Q8**

1. (4%) Briefly explain the differences between operational database systems and data warehouses.
2. (3%) Draw an example diagram to briefly explain the snowflake schema
3. (4%) Give an example to explain the transaction states - partially committed and committed.

**Q9.** Suppose that we have an ordered file with r = 60000 record stored on a disk with block size,

B = 1024 bytes. File records are of fixed size with record length R = 100 bytes.

1. (6%)

(1)What's the blocking factor for the file? What's the number of blocks needed for file?

(2)How many block accesses would be needed to do a binary search for ordering field on the data file?

(3)How many block accesses in average would be needed to do a linear search for non-ordering field on the data file?

(b)(2%) Assume that we have constructed a primary index for the file that the ordering key field of the file is 15 bytes long and a block pointer that is 10 bytes long.

How many block accesses would be needed to search for a record using the primary index?

(c)(6%) Assume that we have constructed a secondary index on a non-ordering key field of the file that is 15 bytes long and a block pointer that is 10 bytes long.

(1) What's the total number of index entries for the file? What's the total number of blocks needed for the index?

(2) How many block accesses would be needed to search for a record using the secondary index?

The above index, which is constructed based on a non-ordering key field, is called a first level index. Assume that a second-level index has been constructed based on the first-level index to create a multi-level index.

(3) (2%) What's the total number of index entries for the second-level index.

(4) (2%) How many block accesses would be needed to search for a record using the multilevel index?

**Q10.** Explain the following.

1. (3%) What is the time performance for searching an element in a search tree? What is the desirable property of a search tree that can have better search performance? Give an example to illustrate the property.
2. (3%) Given the same tree height of B-tree and B+tree, which one can be used to index more data records? Why ?
3. (2%) Explain the differences between Heap files and Sequential files.