

## DATABASE SYSTEM – Final Exam

1. (20%) Assume an “account” relation:

A	B	C	D	E	F
---	---	---	---	---	---

And we have the following dependencies:

$\{A \rightarrow \{B, C, D, E, F\}, BC \rightarrow \{E, F\}, E \rightarrow B, C \rightarrow D\}$

a) (10%) find a minimal set of functional dependency.

suppose we want to decompose into the following 3 relations:

A	B	C
---	---	---

B	C	E	F
---	---	---	---

B	D
---	---

b) (10%) Please verify if this is a decomposition with non-additive join property or not.

2. (30%) 3 Provides cases or examples (along with possible schema) that insert/delete/update anomalies or other difficulties may happen when

- (1) There are partial dependencies within a relation.
- (2) There are transitive dependencies within a relation.
- (3) There are non-trivial multi-value dependencies within a relation.

3. (25%) Suppose a financial organization A want to build a relational database. Another consulting company B get the contract and is responsible for implementing this project. The entire work is divided into 4 steps of processes --- “Requirement analysis”, “Model design”, “Data preparation (cleaning, extraction, transformation, loading...)”, “Testing and confirmation”. There are 4 different groups of people to work together --- “Manager” (from A, who supervises this project, but do not have the background of computer science), “Expert” (from B), “Clerks” (from A, who will use the Database), and “Local IT” (the IT people in A, who maintain most of the systems, including this new database). Please answer the following questions:

- a) (15%) For each of the processes, which groups of people are likely to get involved? Please explain.
- b) (10%) Which process(es) is/are the most critical step(s) for “Manager”, “Clerk”, and “Local IT” to trustfully work around the database in the long run? Please Explain.

4. (25%) For a very complicated data model, suppose there are functional dependencies between attributes in separated tables, and tables are maintained

distributed by multiple sites(each sites have roughly same computation and communication power). Suppose data transmission between different sites are about 50 times slower than local transmission(within same sites). Considering all possible referential constraints (especially when CASCADE option is used) which DBMS must enforce, discuss the overheads and how the tables should be distributed in how many different sites for efficiently enforcing constraints.