Database Management, Final Practice2

Q1.

(a) (8%) Use examples to explain why transitive dependencies and partial dependencies are considered bad in a relational schema.

(b) (3%) Discuss the problem of spurious tuples and how we may prevent it.

Q2.

(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.

Q3.

(a) (4%) What is **Deletion** Anomaly? Illustrate with an example.

(b) (2%) Prove that X → A, if A is a subset of X, according to the definition of functional dependency.

(c) (4%) Give an example of a functional dependency X → Y in a relation R where X is not a superkey in R. Give an example of a non functional dependency where X cannot decide Y.

Q4. (8%) Consider the relation R(A, B, C, D, E, F, G, H, I, J, K, L). (A, B) forms the primary key. C is a candidate key. The following functional dependencies exist among the attributes of the relation:
{A }→ {E, F}; {F}→{G};{G} → {H}; {B}→ {I, J}; {I}→ {K, L}
Transform this relation into a set of 1NF, 2NF, 3NF relations.

Q5. (8%) Consider the relation R (A, B, C, D, E, F, G, H, I, J, K, L, M, N, O). {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}→{M, N}, {D, E}→{G, H}, {E}→{I, J}, {J}→{K, L}, {M}→{O}.
Transform this relation into 2NF, then 3NF relations.

Q6. Q6. 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}, {A} → {D, E, F}, {B, C} → {M, N}, {D, E} → {I, J, K},
{F}→ {G, H}}.

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

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

Q7.

(a) (4%) Briefly explain the differences between operational database systems and data warehouses.

(b) (3%) Draw an example diagram to briefly explain the star schema

Q1

(a)

A relation schema R is in third normal form (3NF) if whenever a FD X → A holds in R, then X is a superkey of R catches violations : where a non-prime attribute functionally determines a non-prime attribute. This catches 3NF violations due to a transitive dependency.

Unlike BCNF, some redundancy is possible with 3NF.
The problems associate with partial and transitive dependencies persist if there is a nontrivial dependency X → A and X is not a superkey, even if the relation is in 3NF because A is part of a key.

(b)

spurious tuples:

when 2 relations are natural joined and if the resulting relation has more tuples than the original set of tuples then those tuples.
a record produced as a result of a join on two or more tables where the joining fields are neither a primary or foreign keys. This could lead to an integrity issue since primary keys are guaranteed to be unique.
a record in database that get created when two tables are joined badly.
created when two tables are joined on attributes that are neither primary keys nor foreign keys.

Spurious tuples can be avoided by joining the 2 relations on equality condition on foreign keys or primary keys that makes sure that no spurious tuples are generated.

No spurious tuples should be generated by doing a natural-join of any relations.

Q2

(a)

3NF: Third Normal Form is the normal form database normalization use, asking all the non primary key only relate to non condidate key，that is, primary keys should be independent for one another. In the condition that 3NF satisfy 1NF and 2NF，all non primary keys are dependent on primary keys → non primary keys are independent for one another

(b)

A relation schema R is in third normal form (3NF) if whenever a FD X → A holds in R, then X is a superkey of R catches violations : where a non-prime attribute functionally determines a non-prime attribute. This catches 3NF violations due to a transitive dependency.

(c)

3NF (Third Normal Form):

only eliminate non primary attribute's transitive dependencies on primary attribute
non primary attribute to be determined by another non primary attribute is ot allowed, but primary attribute to be determined by another non primary attribute is allowed.

BCNF (Boyce-Codd Normal Form):

eliminate all attribute's transitive dependencies on primary attribute

any attribute be determined by non primary attribute is not allowed

There exist relations that are in 3NF but not in BCNF

BCNF is considered a stronger form of 3NF

Q3

(a)

Delete Anomaly:

when you delete a record that may contain attributes that shouldn't be deleted

Consider the relation:

EMP_PROJ(Emp#, Proj#, Ename, Pname, No_hours)

Delete Anomaly:

When a project is deleted, it will result in deleting all the employees who work on that project. Alternately, if an employee is the sole employee on a project, deleting that employee would result in deleting the corresponding project.

(b)

X → A, 如果有一個 relation R, X 與 A 為 R 的屬性的 subset，若且唯若 R 的 X 值可以唯一決定 A 的值，若 A 是 X 的 subset，則 X 便符合 X → A 之定義，X → A 成立因為 Armstrong's Axioms 裡的反射性

Q4

1NF:

A, B is primary key, and the others are same

C is candidate key

{A}→{E, F}; {F}→{G}; {G}→{H}; {B}→{I, J}; {I}→{K, L}

2NF:

A → {E, F, G, H}

B → {I, J, K, L}

3NF:

A → {E}, {F}

B → {I}, {J}

{F} → {G}

{G} → {H}

{I} → {K}

{J} → {L}

Q5

A, B is primary key

C, D, E is candidate key

2NF:

{A} → {M, N}

{M} → {O}
{A, B} → {C, D, E, F}
{C, D, E} → {A, B, E}
{D, E} → {G, H}
{E} → {I, J}
{J} → {K, L}

3NF:

R1-1:

{A} → {M, N}

R1-2:

{M} → {O}

R2-1:

{A, B} → {C, D, E}
{C, D, E} → {A, B}

R2-2:

{A, B} → {F}

R3:

{D, E} → {G, H}

R4-1:

{E} → {I, J}

R4-2:

{J} → {K, L}

Q6

(a) (A, B, C)

(b)

2NF:

R1:

{A, B, C} → {L}

R2:

{A} → {D, E, F}
{F} → {G, H}
{D, E} → {I, J, K}

R3:

{B, C} → {M, N}

3NF:

R1:

{A, B, C} → {L}

R2:

{A} → {D, E, F}

R3:

{D, E} → {I, J, K}

R4:

{F} → {G, H}

R5:

{B, C} → {M, N}

Q7

(a)

data warehouses:

mainly optimized for appropriate data access

emphasize more on historical data as their main purpose is to support time-series and trend analysis

operational database systems:

commonly used for intermediate form of databases before they are cleansed, aggregated and transformed into a warehouse

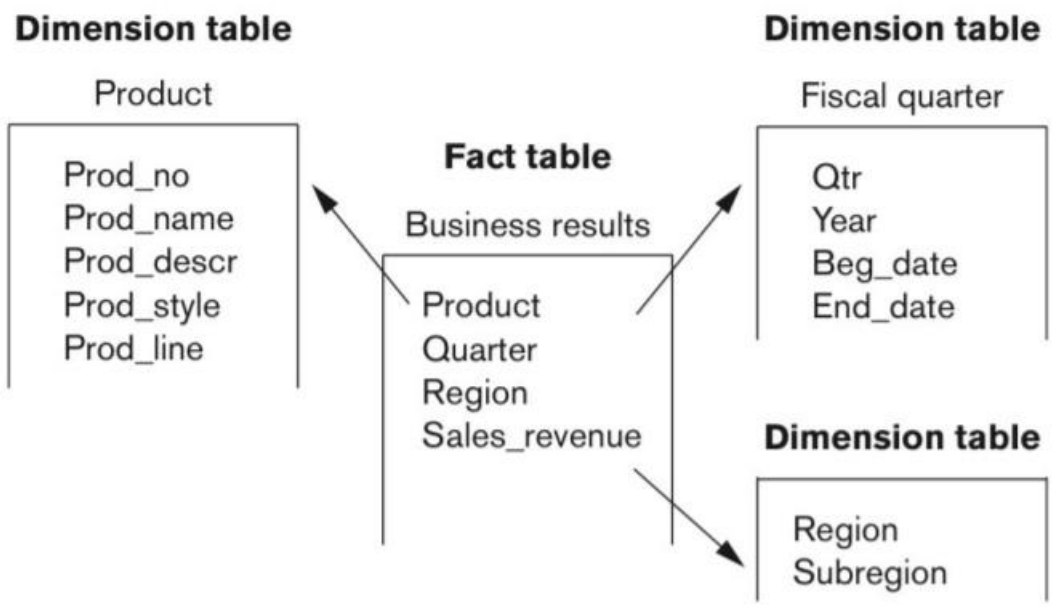
Some major differences between Operational Database Systems and Data Warehouses are:

Operational Database Systems	Data Warehouses
designed to support high-volume transaction processing.	designed to support high-volume analytical processing. (i.e. OLAP).
focuses on Data in.	focuses on Information out.
data is stored with a functional or process orientation.	data is stored with a subject orientation.
Performance is for low analysis queries.	Performance is for high analysis queries.
It is used for Online Transactional Processing (OLTP)	It is used for Online Analytical Processing (OLAP)
represent current transactions.	reads the historical data.
Data within operational systems are generally updated regularly.	Data within a data warehouse is non-volatile, meaning when new data is added old data is not erased so generally updated
Complex data structures.	Multi dimensional data structures.

(b)

Star schema:

Consists of a fact table with a single table for each dimension


Dimension table
Product
Prod_no
Prod_name
Prod_descr
Prod_style
Prod_line
Fact table
Business results
Product
Quarter
Region
Sales_revenue
Dimension table
Fiscal quarter
Qtr
Year
Beg_date
End_date
Dimension table
Region
Subregion