Question 1

a) Explain 3 advantages of using a database system for storing information compared with storing information in a number of files on the computer.

	Do not write in this column
	6 marks

b) The following schema describes a database application for storing student data:

```
Student (studentId, studentName, email, phone)
Module (moduleId, moduleName, moduleLeader, semester)
Module_registration(moduleId, studentId, status)
Lecturer(lecturerId, lecturerName, officeNo)
```

where `Module_registration.status` gives the status of the student with respect to that module, e.g. passed, failed, resitting, etc. `moduleLeader` in relation `Module` is a foreign key referencing `lecturerId` in relation `Lecturer`.

Express the following queries in Relational Algebra:

i) List the names and module leaders' ID for all modules with a module code higher than 100.

ii) List all modules along with the details of their module leader.

iii) List all modules led by the lecturer whose name is "John White".

iv) List the IDs of all students currently taking the "Computer Science" module.

	Do not write in this column

	10 marks

c) Explain the referential integrity rule of the Relational model and give an example of referential integrity.

	Do not write in this column
	4 marks

Question 2

a) Explain, and give an example of each of the following terms:

v) Candidate key

vi) Functional dependency

vii) Transitive dependency

viii) Multi-valued dependency

	Do not write in this column
	11 marks

b) Examine table 1 shown below. This table represents the hours worked per week for temporary staff at each branch of a company. We assume that each staff can only have one position, but can work in different branch for different hours per week.

staffNo	branchNo	branchAddress	name	position	hoursPerWeek
S4555	B002	City Center Plaza, Seattle, WA 98122	Ellen Layman	Assistant	16
S4555	B004	16 – 14th Avenue, Seattle, WA 98128	Ellen Layman	Assistant	9
S4612	B002	City Center Plaza, Seattle, WA 98122	Dave Sinclair	Assistant	14
S4612	B004	16 – 14th Avenue, Seattle, WA 98128	Dave Sinclair	Assistant	10
S4721	B003	36 Main street, Seattle, WA 98103	Peter Hoffman	Assistant	9

Table 1

i) Table 1 shown above is susceptible to update anomalies. Provide examples of how insertion, deletion, and modification anomalies could occur on this table.

	Do not write in this column
	9 marks

ii) Identify the functional dependencies represented by the data shown in table 1 and primary key. State any assumptions you make about the data (if necessary).

	Do not write in this column
	4 marks

iii) Is the relation in table 1 in Third Normal Form (3NF)? If not, decompose the table into 3NF relations. Identify the primary and foreign keys in your 3NF relations.

	Do not write in this column
	6 marks

Question 3

a) You have just been hired as a consultant for a big airplane manufacturer. Impressed by your background in databases, they want you to completely redesign their database system. Talking with the people in the company, you get the following information:

The database contains information about employees, factories and parts.
Each employee has a social security number (SSN), name and salary. An employee is uniquely identified by his or her SSN.
Each factory has an id, name and a budget. The id uniquely identifies a factory.
Each part has an id and a name. The id uniquely identifies a part.
Each employee reports to at most one other employee.
Each employee works in at least one factory.
Each part is manufactured in exactly one factory.
Each part is a component of zero or more other parts.

A partial Entity-Relationship diagram for the above application is shown in Figure 1 below.

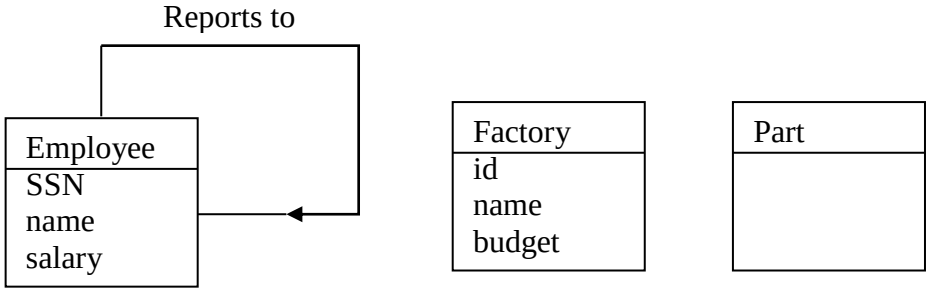

Figure 1

You are requested to do the following:

i) Add attributes for entity Part
ii) Add missing relationships
iii) Identify the primary keys
iv) Give the cardinality and participation constraints on relationships

	Do not write in this column

	16 marks

b) Consider the following schema (primary keys are underlined)

```
Student (sname, sid, grade, level, deptno)
Course (cno, cname, deptno, units)
Dept (dname, deptno)
Takes (sid, cno)
```

Answer the following:

i) Formulate the following query using SQL:

	Do not write in this column
1) List the IDs of students in the department of "EECS".	
2) List the names, grades of students who are doing "Software Engineering" course.	
3) Calculate the total number of courses in each department.	
4) For each department with more than 20 courses, find number of courses and sum of their units.	
	10 marks

ii) Which of the following queries returns the id of the student with the highest grade? **More than one choice may be correct**.

A)
```
SELECT S.sid
FROM Students S
WHERE S.grade = MAX(S.grade);
```

B)
```
SELECT S.sid, MAX(S.grade)
FROM Students S
GROUP BY S.grade;
```

C)
```
SELECT S.sid
FROM Student S
WHERE S.grade > ALL (SELECT S.grade FROM Student S);
```

D)
```
SELECT S.sid
FROM Student S
WHERE S.grade = (SELECT MAX(S.grade) FROM Student S);
```

E) None of the above

	Do not write in this column
	4 marks

Question 4

a) An ACID transaction would permit the following situations (**mark "True" or "False" for each**):

i) After a failure, an uncommitted transaction is rolled back and all of its effects are erased.

ii) After a failure, a committed transaction is rolled back and all of its effects are erased.

iii) A transaction updates a tuple, but has not committed the changes. Another transaction reads the uncommitted data.

iv) A transaction reads the same tuple twice without writing it in between and sees two different values.

	Do not write in this column
	4 marks

b) For the following schedule of two transactions T1, T2:

time	T1	T2
t1		Begin_transaction
t2	Begin_transaction	Read(X)
t3	Read(X)	X:=X+200
t4	X:=X-50	Write(X)
t5	Write(X)	Commit
t6	Commit	

Figure 2

i) Discuss the problem of updating X in the two transactions T1 and T2.

	Do not write in this column
	3 marks

ii) Explain the principles of two-phase locking (2PL). Rewrite T1 and T2 in Figure 2 using 2PL.

	Do not write in this column
	7 marks

c) 'A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision-making process'. Pick any three of the major characteristics of the data held in a data warehouse and describe them.

	Do not write in this column
	6 marks