Project COMP 378 – Intro to Database Management

Name: Jason Bishop
Student #: 3042012

Project Module 1 (30 marks)

Study the information given below, and answer the questions.

For a long time, the Royal Victoria Hospital (RVH) worked with an information system that consisted of a mix of paper-based files and small independent databases developed within some departments. The new administration has created an information system (IS) department, and they hired you as information officer: head of the team in charge of the design and implementation of a new global information system. The following information was collected by the first team from the IS department, who conducted interviews with some of the hospital administration and staff to identify entity types for the hospital.

The hospital depends primarily on four groups of people: employees, physicians, patients, and volunteers. Of course, some common attributes are shared by all of these groups: person_ID (identifier), name, address, birth date, and phone number. Each group also has at least one unique attribute of its own. Employees have a date hired, volunteers have a skill, physicians have a specialty and a pager number, and patients have a contact date (date of the first contact with the hospital). Some people may belong to two or more of these groups at a given time (e.g., patient and volunteer).

Patient: A person who is either admitted to the hospital, or is registered in an outpatient program. One, and only one, physician is responsible for each patient. Patients are divided into two groups: resident and outpatient. Each outpatient is scheduled for zero or more visits. The entity visit has two attributes: date (partial identifier), and comments. Note that an instance of visit cannot exist without an outpatient owner entity. Only resident patients are assigned to a bed, and a bed may or may not be assigned to a patient.

Physician: A member of the hospital staff who may admit patients to the hospital, and who can administer medical treatments. A given physician can be responsible for zero or more patients at a given time.

A patient must be referred to the hospital by exactly one physician. A physician can refer any number of patients, or may not refer any patient.

Physicians may perform any number of treatments on behalf of any number of patients, or may not perform any treatment. A patient may have treatments performed by any number of physicians. For each treatment performed on behalf of a given patient by a particular physician, the hospital records the treatment date, treatment time, and results.

Employee: Any person employed as part of the hospital staff. Employees are subdivided into three groups: nurse, staff, and technician. Only nurse has the attribute certificate, which indicates a qualification (RN, LPN, etc.). Only staff has the attribute job class, and only technician has the attribute skill. Each nurse is assigned to one (and only one) care centre.

Each technician is assigned to one or more laboratories.

Care centre: A treatment centre within the hospital. Examples of care centres are maternity, emergency, and cardiology. Attributes of care centre are name (identifier) and location. A care centre may have one or more nurses assigned to it. Also, one of the nurses assigned to each care centre is appointed nurse-in-charge. A nurse cannot be appointed nurse-in-charge of a care centre unless s/he has an RN certificate.

Each hospital employee is assigned to work in one or more care centre. Each care centre has at least one employee, and may have any number of employees. The hospital records the number of hours per week that a given employee works in a particular care centre. Each physician can be assigned to one or more care centres, and a care centre can have one or more physicians assigned to it.

Laboratory: A unit in the hospital where clinical tests (i.e., blood, urine, tissue, etc.) are performed to obtain information about the health of a patient. Attributes of laboratory include name (identifier) and location. A laboratory must have one or more technicians assigned to it.

Bed: A hospital bed that may be assigned to a resident patient who is admitted to the hospital. Each bed has a bed number, a room number, and a care centre ID. There may be no bed assigned to a care centre, or a care centre may have one or more beds assigned to it.

Item: Any medical or surgical item that is used in treating a patient. Each item has an item number, description, and unit cost.

A patient may optionally consume any number of items. A given item may be consumed by one or more patients, or may not be consumed. For each item consumed by a patient, the hospital records the date, time, quantity, and total cost (which can be computed by multiplying quantity by unit cost).

Treatment: Any test or procedure performed by a physician on behalf of a patient. Each treatment has a treatment ID, which consists of a treatment number and a treatment name.

a. Is the ability to model subtype/supertype relationships likely to be important in a hospital environment such as the RVH? (4 marks)

The ability to model sub/supertype relationships is likely to be very important in a hospital environment such as the RVH. There are several relationships that can be modeled in this fashion. A general employee type can be the supertype of nurse, staff, and technician. Additionally, The patient type can be a supertype of both the resident and outpatient types. All of these can even be subtyped under a more general person entity which can also encompass physicians and volunteers. The reason for using this type of relationship is that there are characteristics that each have in common, that can be stored in a supertype, with the distinct properties of each entity encapsulated within the subtype. Additionally, each of the subtypes may participate in unique relationships, such as the resident patient needing a bed that the outpatient does not.

b. Can the business rules paradigm, and the ability to easily define, implement, and

maintain business rules, be used as a competitive advantage in a hospital environment such as the RVH? (4 marks)

The business rules paradigm and the definition, implementation, and maintenance of business rules can certainly be used as a competitive advantage in a hospital environment such as the RVH. The main reason for this is that the enforcement of business rules can be automated by the system developed by the RVH. This can lead to all sorts of competitive advantages. For example, patient processing times may be faster because of automation. More patients processed means more revenue, and it will likely attract more patients because of the efficiency of service in a hectic hospital environment. Also, if business rules are easy to implement, then when new treatments become available, they can quickly and easily become incorporated into the system. This allows patients to receive new treatments sooner, which is highly attractive to patients meaning they will be more likely to seek out the RVH for their treatments. These types of competitive advantages will surely help the RVH succeed.

c. Do there appear to be any weak entities in the description of the data requirements in this project module? (4 marks)

The entity visit appears to be a candidate for a weak entity in the description of the data requirements. The existence of a weak entity type depends on another entity type. In this case, the entity type would be an outpatient. No visit can occur without an outpatient, and thus, the outpatient is the identifying owner on which a visit depends. We are also given a clue to the weak entity status of visit because the attribute date is mentioned as a partial identifier. This can be combined with the person(supertype) identifier to form a full identifier for the entity, or a surrogate identifier can be used in its place. These factors lead to the use of the entity visit as a weak entity.

d. Draw an EER diagram to accurately represent this set of requirements. State any assumption you had to make in developing the diagram. (14 marks)

*see attached diagram *Module1-Partd.jpg*
Assumptions:

- A care centre can have 0 beds assigned to it – as when it is under construction or renovation
- The bed number is a unique number which can serve as an identifier for a bed
- Each person can be uniquely identified in the system with a unique identification number, and there is no need for specific employee or patient numbers.
- Each employee can only be of one type and not two at the same time. For example, there are no nurses who are also doctors.
- A patient can be either a resident or outpatient but also both. For example, a heart attack victim may be admitted to the hospital, but follow ups after release will classify them as an outpatient. To reduce redundancy it seems better to allow overlap.

e. Are there any universal data models that can be reused as a starting point for modeling

RVH's data requirements? (4 marks)

There is a universal data model that can be reused as a starting point for modeling RVH's data requirements: the "Person" model. In the original model, this concept covers the 4 categories of employee, physician, patient and volunteer. In the ERD, the model is used to represent the same categories as entities in the diagram. The person entity holds the general information common to each person, such as their name, address, birth date and phone number. Additionally, it also holds the primary key of Person_ID to give each person a unique identifier, regardless of the subtypes. The universal model can help kickstart the modeling process because it provides a starting point for figuring out requirements and it can make the diagram easier to read because people are familiar with the concept of a person. Additionally, the universal model can adapt to changes that can be made in the future and it can make it easier to share data for interorganizational systems that may be developed.

Project Module 2 (30 marks)

Use the relational schema of the EER diagram you developed in Module 1 to answer the following questions.

a. Should the RVH use normalization when designing its database? (3 marks)

The RVH absolutely should use normalization when designing its database. Normalization minimizes the amount of redundancy in the database. This helps when deleting or making changes to the data because these changes will ideally need to be made in one place only. Additionally, normalized data can be accessed and manipulated more quickly and efficiently.

b. Why are entity integrity and referential integrity constraints of importance to the hospital?(3 marks)

Integrity constraints are important to the hospital for a variety of reasons. Referential integrity ensures foreign key values match primary key values in another relation. For example, it would ensure that treatments are assigned by a qualified physician and assigned to an existing patient. Problems could possibly occur without these constraints as when a treatment is assigned to no patient, then valuable medical inventory could be stolen or misused. Also, having physicians attached to treatments helps the hospital when accidents occur. For malpractice reasons, the hospital will want to make sure they have an accurate record of who assigned and performed treatments on which patients. Entity integrity ensures that each relation has a non-null primary key. This is also important for keeping accurate records to avoid misuse and malpractice issues as in the previous examples.

c. Map the EER diagram to a relational schema, and transform the relation into 3NF. (10 marks)

I have laid out the relation schema below and color coded values to show the references among attributes.

PERSON

PersonID	Name	Address	BirthDate	PhoneNo

VOLUNTEER

PersonID	Skill

PHYSICIAN

PersonID	Specialty	PagerNo

PHYSICIANASSIGNMENT

PersonID	CareCentreName

PATIENT

PersonID	ContactDate	ReferredPersonID	AttendingPhysicianID

RESIDENT

PersonID	BedNo

EMPLOYEE

PersonID	DateHired

TECHNICIAN

PersonID	Skill	LabName

STAFF

PersonID	JobClass

NURSE

PersonID	Certificate

LABORATORY

Name	Location

VISIT

PersonID	Date	Comments

ITEM

ItemNo	Description	UnitCost

TREATMENT

TreatmentNo	TreatmentName	PhysicianID	PatientID	ItemNo	Date	Time	Results

WEEKLYHOURS

PersonID	CareCentreName	HoursWorked

NURSEINCHARGE

PersonID	CareCentreName	RNCertNo

d. Besides the 3NF relations, what additional types of information are required to create a physical database design? (3 marks)

Along with the relations provided, there is much more information needed to create a physical database design. For each attribute, definitions are needed such as type, size, min/max values, etc. Descriptions of data usage and database technology implemented are also required. Furthermore, requirements for data security, response time and similar needs must also be provided.

e. Are there opportunities for horizontal or vertical partitioning of the database? Are there other opportunities to denormalize the relations of this database? If no, explain why? If yes, how might you denormalize the database? (3 marks)

Horizontal partioning distributes the rows of a relation into separate tables. One opportunity for horizontal partitioning is in the Treatment relation. This can be done along the TreatmentName or ItemNo column to be able to process specific types of treatments or treatments with specific items more quickly and easily. Also, the relation can be partitioned along the Date column to process the daily records. A partion along the Date column of the Visit relation would be similar in function. However, the Treatment relation would probably be best served by vertical partioning, which distributes the columns of the relation into separate tables, repeating the primary key. In this way, the Treatment can be broken up into tables with medical information and accounting information separately.

f. Suppose the date treatment performed was not entered. What procedures are required to handle the missing data? (3 marks)

Handling the missing data if a date for the treatment performed was not entered depends on the length of time before the error is discovered. If it is discovered right

away, the database can be rolled back and the information can be entered, but care must be taken to ensure that any transactions involving the missing data are also reversed and reprocessed. The more likely situation is that the error is not discovered until several days or weeks later. In this case, a series of compensating transactions may be required to correct any errors the missing data may have caused. For example, a patient bill may have to be readjusted to reflect the treatment they received, or a monthly status report of treatments may need to be reissued.

g. Consider the following query against the RVH database.

For each treatment performed in the past two weeks, list the physicians performing the treatment (grouped by treatment), and the number of times this physician performed that particular treatment, on that particular day. Order the list by treatment ID, and by reverse chronological order for each treatment ID.

Create secondary key indexes to optimize the performance of this query. State any assumptions. (5 marks)

There can be several indexes used to optimize the performance of this query. Indexes on the physician ID number and the treatment ID will be beneficial, as well as one for the date value of the treatment relation. However, because the physician ID is the primary key of the Person relation, and the treatment ID is the primary key for the Treatment relation, I would assume that the DBMS has already made an index for these attributes. Therefore, I would create the final date index with the following SQL command:

CREATE INDEX DateIndex ON Treatment_t(Date);

Project Module 3 (40 marks)

Consider the following relations:

- Patients(pid, name, address, telephone, care_centre_id)
- Care_centres(cid, name, location, nurse_charge_id)
- Treatments(tid, patient_id, physician_id, treatment_name, date)
- Nurses(nid, name, care_centre_id, certificate_type, telephone, salary)
- Physicians(phid, name, pager_number, specialization, salary).

Use Oracle to complete the following tasks.

a. Create the tables that correspond to these relations in your Oracle home database. (6 marks)

```
CREATE TABLE Physicians_t
        (phid                   INT             NOT NULL,
```

```
        name            VARCHAR(30),
        pager_number    VARCHAR(15),
        specialization  VARCHAR(20),
        salary          INT,
CONSTRAINT phid_PK PRIMARY KEY (phid));

CREATE TABLE Treatments_t
        (tid            INT             NOT NULL,
        patient_id      INT             NOT NULL,
        physician_id    INT             NOT NULL,
        treatment_name  VARCHAR(30)     NOT NULL,
        treatment_date  DATE,
CONSTRAINT treatmentID_PK PRIMARY KEY (tid, treatment_name)
CONSTRAINT phid_FK FOREIGN KEY (physician_id) REFERENCES
Physicians_t(phid));

CREATE TABLE Care_centeres_t
        (cid            INT         NOT NULL,
        name            VARCHAR(30),
        location        VARCHAR(20),
        nurse_charge_id INT         NOT NULL,
CONSTRAINT cid_PK PRIMARY KEY (cid));

CREATE TABLE Patients_t
        (pid            INT         NOT NULL,
        name            VARCHAR(30),
        address         VARCHAR(100),
        telephone       VARCHAR(15),
        care_centre_id  INT         NOT NULL,
CONSTRAINT pid_PK PRIMARY KEY (pid)
CONSTRAINT tid_FK FOREIGN KEY (care_centre_id) REFERENCES
Care_centres_t(cid));

CREATE TABLE Nurses_t
        (nid            INT             NOT NULL,
        name            VARCHAR(30),
        care_centre_id  INT             NOT NULL,
        certificate_type VARCHAR(5),
        telephone       VARCHAR(15),
        salary          INT,
CONSTRAINT nid_PK PRIMARY KEY (nid)
CONSTRAINT tid_FK FOREIGN KEY (care_centre_id) REFERENCES
Care_centres_t(cid));

ALTER TABLE Care_centres_t ADD CONSTRAINT nid_FK FOREIGN KEY
(nurse_charge_id) REFERENCES Nurses_t(nid)
```

b. If not automatically created by the DBMS, create indexes corresponding to the primary and foreign keys. (6 marks)

These were created in part a when creating the tables (see above).

c. Populate these tables with some sample data, and write SQL queries that show the content of each table after entering the data. (6 marks)

```
SELECT *
        FROM Physicians_t
```

phid	name	pager_number	specialization	salary
1	A. A.	416-123-4567	cardiology	100000
2	B. B.	416-123-4568	neurology	120000
3	C. C.	416-123-4569	podiatry	70000
4	D. D.	416-123-4560	oncology	100000
5	E. E.	416-123-4561	proctology	100000

```
SELECT *
        FROM Treatments_t
```

tid	patient_id	physician_id	treatment_name	treatment_date
1	1	1	EKG	2015-01-01
2	2	2	Lobotomy	2015-01-02
3	3	3	Left foot amputation	2015-01-01
4	4	4	Chemotherapy	2015-01-03
5	5	5	Fecal transplant	2015-01-04
6	6	1	Angioplasty	2015-01-03
7	7	2	Lobotomy	2015-01-06
8	8	3	Wart removal	2015-01-07
9	9	4	consultation	2015-01-05
10	10	5	endoscopy	2015-01-08
11	11	1	triple bypass	2015-01-06
12	12	2	Lobotomy	2015-01-07
13	13	3	pedicure	2015-01-09
14	14	4	biopsy	2015-01-01
15	15	5	endoscopy	2015-01-01
16	16	1	lidocane injection	2015-01-02
17	17	2	Lobotomy	2015-01-10
18	18	3	Wart removal	2015-01-06
19	19	4	chemotherapy	2015-01-07
20	20	5	fecal transplant	2015-01-08

```
SELECT *
        FROM Care_centres_t
```

cid	name	location	nurse_charge_id
1	cardiology	building 1	1
2	oncnology	building 2	2
3	emergency	building 1	3
4	neurology	building 2	4
5	icu	building 1	5
6	long-term care	building 3	6

```
SELECT *
        FROM Patients_t
```

pid	name	address	telephone	care_centre_id
1	A.B.	Toronto, ON	416-555-5555	1
2	B.C.	Toronto, ON	416-555-5554	2
3	C.D.	Toronto, ON	416-555-5553	3
4	D.E.	Toronto, ON	416-555-5552	4
5	E.F.	Toronto, ON	416-555-5551	5
6	F.G.	Toronto, ON	416-555-5550	6
7	G.H.	Mississauga, ON	905-777-8899	1
8	H.I.	Mississauga, ON	905-777-8898	2
9	I.J.	Mississauga, ON	905-777-8897	3
10	J.K.	Mississauga, ON	905-777-8896	4
11	K.L.	Mississauga, ON	905-777-8895	5
12	L.M.	Mississauga, ON	905-777-8896	6
13	M.N.	Brampton, ON	905-666-8896	1
14	N.O.	Brampton, ON	905-666-8895	2
15	O.P.	Brampton, ON	905-666-8894	3
16	P.Q.	Brampton, ON	905-666-8893	2
17	Q.R.	Brampton, ON	905-666-8892	1
18	R.S.	Brampton, ON	905-666-8891	2
19	S.T.	Brampton, ON	905-666-8890	3
20	T.U.	Brampton, ON	905-666-6666	6

```
SELECT *
        FROM Nurses_t
```

nid	name	care_centre_id	certficate_type	telephone	salary
1	Z.Z.	1	RN	416-897-6543	50000
2	Y.Y.	2	RN	416-897-6542	45000

3	X.X.	3	RN	416-897-6541	40000
4	W.W.	4	RN	416-897-6540	55000
5	V.V.	5	RN	416-897-6549	50000
6	U.U.	6	RN	416-897-6548	45000
7	T.T.	2	AOCN	416-897-6547	40000
8	S.S.	3	CRNA	416-897-6546	30000
9	R.R.	5	TNS	416-897-6545	35000
10	Q.Q.	6	NPC	416-897-6545	30000
11	P.P.	1	NPC	416-897-6545	40000
12	O.O.	4	NPC	416-897-6545	35000
13	N.N.	3	HACP	416-897-6666	30000
14	M.M.	2	GPN	416-897-5555	30000
15	L.L.	6	CTRS	416-897-4444	30000

d. For some strategic decisions, the president of the hospital needs summary data about the care centres. For each care centre, s/he needs to know the number of nurses holding an RN certificate, as well as their total and average salaries. Does the following view answer the president's request? If not, write the correct view that will satisfy the president's request. (7 marks)

```
CREATE VIEW NURSE_SUMMARY (D, C, TOTAL_S, AVERAGE_S)
AS SELECT cid, COUNT (*), SUM (salary), AVG (salary)
        FROM Care_centres, Nurses
        WHERE nurse_charge_id = nid and certificate_type like 'RN'
        GROUP BY cid;
```

This view does not satisfy the president's request. Here is the correct SQL command to generate the correct view:

```
CREATE VIEW NURSE_SUMMARY_v (D, C, TOTAL_S, AVERAGE_S)
AS SELECT cid, COUNT (*), SUM (salary), AVG (salary)
        FROM Care_centres_t, Nurses_t
        WHERE cid = care_centre_id and certficate_type = 'RN'
        GROUP BY cid;
```

Here is the view generated:

D	C	TOTAL_S	AVERAGE_S
6	3	105000	35000
5	1	50000	50000
3	1	40000	40000
2	1	45000	45000
1	2	90000	45000

4	2	90000	45000

e. State which of the following queries and updates would be allowed in this view. If a particular query or update would be allowed, show what the corresponding query or update on the base relations would look like, and give its result when applied to the database. (5 marks)

Q1. SELECT *
FROM NURSE_SUMMARY;

This query would be allowed. See the results above in part d.

Q2. SELECT D, C
FROM NURSE_SUMMARY
WHERE TOTAL_S > 100000;

This query would be allowed. Here are the results:

D	C
6	3

Q3. SELECT D, AVERAGE_S
FROM NURSE_SUMMARY
WHERE C > (SELECT C FROM NURSE_SUMMARY WHERE D=4);

This query would be allowed. Here are the results:

D	AVERAGE_S
6	35000

Q4. UPDATE NURSE_SUMMARY
SET D=3
WHERE D=4;

This update would not be allowed.

Q5. DELETE FROM NURSE_SUMMARY
WHERE C > 4;

This update would not be allowed.

f. Use Oracle Forms, or write a program (using PL/SQL and/or Java) to display the following report for a given patient (the user will only enter the patient number). (10 marks)

Patient Number: _____

Patient Name: _____

Care Centre Name: _____

Name of Nurse-in-Charge: _____

Treatment ID	Treatment Name	Physician ID	Date
_____	_____	_____	_____
. . .	. . .	. . .	. . .

I developed a program in Java using JDBC to connect to the database and get the patient number and related info to display. See the file *OracleReport.java*