Assignment 3

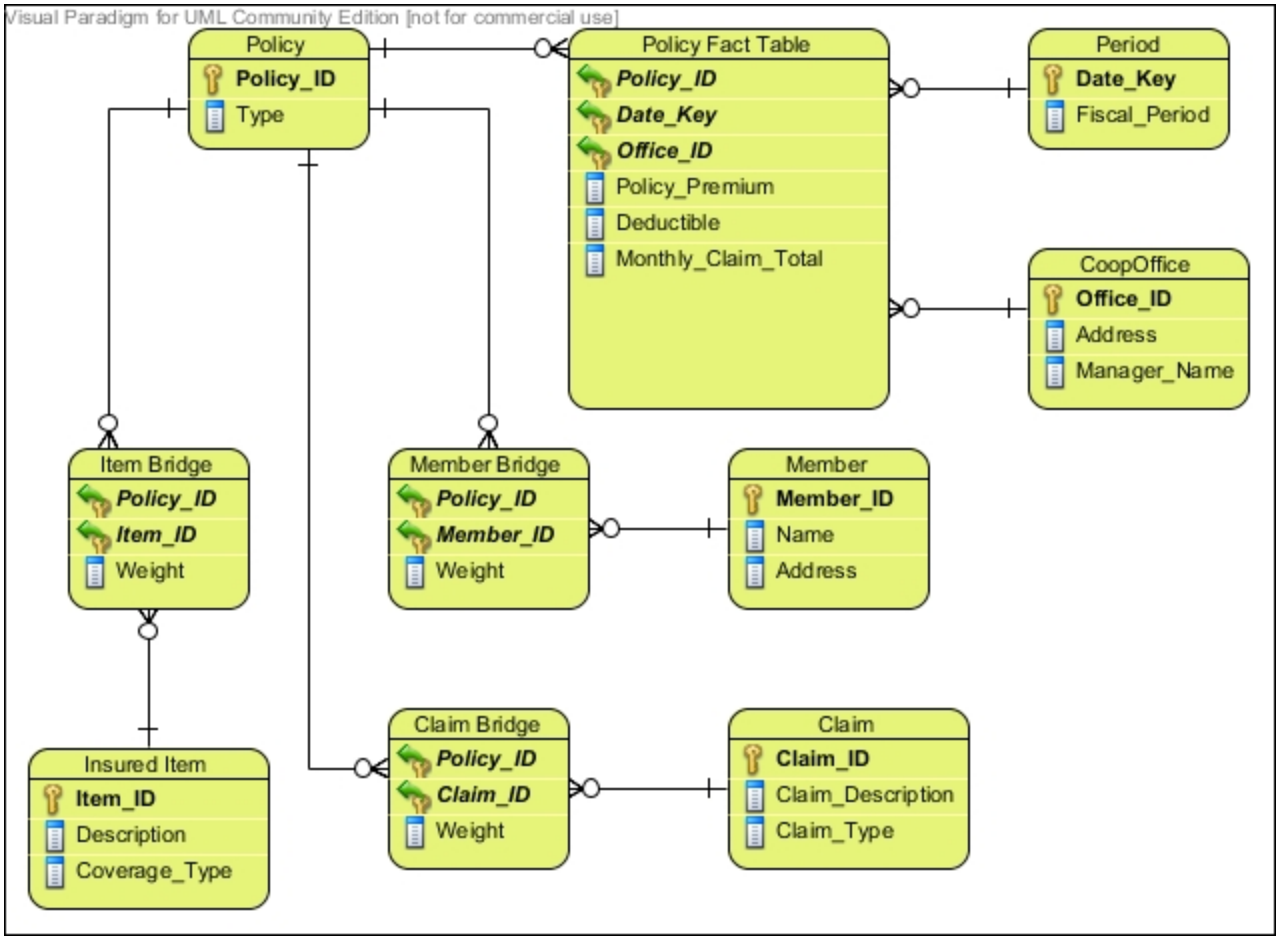
Course: COMP378
Name: Jason Bishop
Student#: 3042012

Question 1 (18 marks)
Consider the following dimensions, dimension attributes, and dimension sizes for Farm Coop-Insurance.

- Member(member_ID, Name, Address). On average, there are two members for each policy and item covered by the policy.
- InsuredItem(Item_ID, Description, Coverage_Type). There is an average of ten covered items per policy.
- CoopOffice(Office_ID, Address, Manager_name). Each policy is registered and managed by only one office of the cooperative.
- Policy(Policy_ID, Type). The company has approximately one million policies at the present time. Approximately five percent of these policies experience some change each month.
- Period(Date_Key, Fiscal_Period). The length of the fiscal period is one month. The decision system reports are supposed to be based on five years of data.
- Claim(Claim_ID, Claim_Description, Claim_Type).

The facts to be recorded for each combination of these dimensions are: Policy_Premium, Deductible, and Monthly_Claim_Total.

a. Design a star schema for this problem.



b. Using the assumptions stated above, estimate the number of rows in the fact table.

Total number of policies = 1,000,000 total * 0.05 active = 50,000
Total number of periods = 12 mth * 5 yr = 60
Total number of offices = 1

Total rows = 50,000 policies * 60 months * 1 office = 3,000,000 rows

c. Estimate the total size of the fact table (in bytes), assuming an average of 5 bytes per field.

Total size = $7.5 * 10^{22}$ rows * 6 fields * 5 bytes/field = 90,000,000 bytes

Question 2 (20 marks)
Suggest an appropriate recovery technique that a database administrator could use to resolve each of the following situations.

a. A network disconnection occurs while a user is entering a transaction at an ATM bank machine.

If a network disconnection occurs as in this case, there is a system failure and the database should not be damaged. A good recovery technique then, would be to restart from the most recent checkpoint and roll forward any changes. This would be the safest and fastest method. The user should restart their transaction and be informed of the error.

b. A disk drive fails during regular operations while a clerk is entering data about newly registered students.

If a disk drive fails as in this case, we will have destruction of either part of the database or perhaps the entire thing. The best recovery technique would be to use a mirrored copy of the database to immediately take over and handle the clerk's transaction. However, this transaction must be restarted. If there is no mirrored copy, a backup copy can be used and rolled forward.

c. The registration office at a university entered an incorrect amount for a student tuition payment. The error was discovered by the financial services department several weeks later.

If a user input error is discovered as in this case, we have incorrect data in the database. Since it has been discovered weeks later, there may be other records that have used the data and will also be in error. Therefore, backward recovery will not be feasible. If there are only a few errors to correct, they can be done so by hand with compensating transactions to restore values to the correct state. If this is not a feasible task, then it will be necessary to restart from the checkpoint before the error, input the correct information, and process the resulting transactions with the correct information. This may also necessitate appropriate phone calls or mailings to explain the error to affected clients.

d. The database administrator of a financial institution performed a full database backup, but forgot to activate the journalizing facility. Afterwards, data entry clerks at the financial institution entered transactions for two hours before the database became corrupt. It is discovered that the journalizing facility of the database has not been activated since the backup was made.

If the journalizing facility is not functioning as in this case, the changes in the database will be lost from the 2 hours of unrecorded transactions. The best solution would be to have a mirrored database that is updated simultaeneously with the currect database. That way, the mirrored version can be switched to and the journalizing facility can be turned on immediately to begin keeping record of the transactions. If there is no mirrored version, then the backup copy of the database must be used. Unfortunately, since there is no log of the transactions, the data entry clerks will have to repeat their work and their time will be lost.

Question 3 (20 marks)

Suggest the most appropriate security measures for each of the situations described below.

a. The Western Union bank uses an electronic funds transfer (EFT) system to transmit sensitive financial data between its branches all over the world.

To secure the sensitive data transmitted between branches, the bank should begin by using encryption to code all of its transmissions. Prefereably this encryption will be with a very secure two-key method. This will not prevent the interception of data, but the intercepted data will be effectively unusable because it will be extremely difficult to decipher the data. This can be implemented using a 128-bit version of SSL. Of course, the data will need to be kept secure at each location it is transmitted to. This will necessitate other security measures such as using an authentication scheme to restrict access to unauthorized individuals. Machines with sensitive data can be kept under lock and key, and to log into systems to read or use data, a password system can be used. Additionally, the building itself can be further secured with a fingerprint scan or smart card access. The greater the number of factors used to secure access to data, the greater the security that will result.

b. A fighter jet simulation company has set up an off-site computer-based training centre for the F35. The company wishes to restrict access to the site to authorized employees. Since each employee's use of the centre is occasional, it does not wish to provide the employees with keys to access the centre.

The company can provide access to the training center to its employees by using an authentication scheme. Since the company does not want to provide keys to the site, other methods need to be used. The building can could be accessed using a unique personal characteristic such as a fingerprint or retinal scan. However, to increase the security of the site, the company should use more than one factor to grant access. In this case, the simulator should have a user identification scheme with a password or PIN number to be able to log in. This two-factor scheme would be much more secure. A three factor scheme would be even more secure, but the company is not interested in handing out keys, so would be unlikely to issue any other forms of access for people to possess such as a smart card or token.

c. A golf club uses a simple password system to protect its database. The club has created a new Web site to allow both its members and its employees to use the new Web-based system to access and update information. The club finds that it needs a more comprehensive security system to grant different privileges (such as read-only versus create or update) to different users.

The club can easily grant different priveleges to different users of its database. They should start by granting different views of data to different users. This way, each user can see only the data they need. However, this does not necessarily prevent unauthorized manipulation or access. To accomplish this, they will need to implement authorization rules restrict actions on data and access to it. For example, Web users may be granted privleges to read data only unless updating user information. Employees on the intranet can be granted greater privleges based on their role within the organization. Additionally, integrity controls should be in place to protect the data from unauthorized manipulation. For example, assertions and triggers can be used to check for unauthorized actions. Finally, because this system can be accessed through the Web, additional Web security features will be needed such as a firewall

and properly configured routers. Sensitive information can be placed in a location that is not accessible from outside the organization's local intranet.

d. A training centre at the Northern Technical University has experienced considerable difficulty with unauthorized users who access files and databases by appropriating passwords from legitimate users.

The university needs a way to stop unauthorized use of files and databases, but would still want authorized users to have the necessary access to these same resources. Unfortunately, the password system in place does not seem to be secure enough. Using passwords is only a one-factor authentication scheme and the university can significantly increase security by implementing a two or three factor scheme. This can be done by issuing smart cards or using biometric scans as in the previous question with the fighter jet simulation. Additionally, the university should re-examine the authorization rules it has in place and consider being more restrictive to what data can be accessed by whom. Changes can be implemented to restrict views, changes, and access to data as in the previous example with the golf club.

Question 4 (12 marks)

The UBS broker company has a database server with three disks. Both the accounting and stock exchange applications share the same disk, and they are experiencing performance problems. Discuss potential reasons for the performance problem, and suggest how to reduce I/O contention.

Since all the information is on the same disk, it is unlikely that the performance problems are due to CPU usage. The disk read/write speeds are almost always much slower than the CPU processing. Therefore, the company should begin by placing the applications and data on separate disks. This will allow for parallel processing, which should dramatically increase performance. Each application will likely access different data, so the data that is most often accessed by a particular application should be placed with similar data on its own disk. This way, each application will have less effect on the other. Furthermore, each application will need to be tuned to take advantage of this situation. Tuning can be done by rewriting SQL queries to be more efficient and denormalizing data to prevent costly join operations. Heavy jobs can be deferred to more quiet usage times to prevent problems as well. Finally, the users of the system should have realistic expectations of the performance of the system so that they write better queries and take advantage of the system during periods of lower usage.

Question 5 (10 marks)

Consider the concurrent execution of the following transactions.

T1	T2
Read A	Read B

Read B	Write A
Write C	Read C
Write A	Write B
Commit	Commit

List two problems that may occur from the concurrent execution of these two transactions.

One problem that may occur from concurrent execution of these transactions is a lost update. Transaction T1 reads A and gets a specified value. Afterward, transaction T2 writes a value to A, thus the information read in T1 is no longer valid. Later, when T1 tries to write to A based on the previous read and will record over the information written by T2. The update by T2 is lost due to interference between the transactions.

Another problem that may occur is an inconsistent read. This happens when partially updated information is read. When transaction T1 is writing to C, transaction T2 is attempting to read at the same time. If the write has not finished before the read, then T2 may only be reading some of the updated information and some of the information that has yet to be updated. If any transaction, including T2, were to attempt to read from C again later, it would not be able to reproduce the results obtained from the inconsistent read, because T1 will have written data that changes the results.

Question 6 (20 marks)

The Edmonton International Airport would like to implement a database that will be used to keep track of airplanes, their owners, airport employees, and pilots. From the requirements for this database, the following information was collected.

- Each airplane has a registration number, is of a particular plane type, and is stored in a particular hangar.
- Each hangar is managed by an employee who supervises the maintenance services performed in that hangar.
- Each plane type has a model number, a capacity, and a weight.
- Each hangar has a number, a capacity, and a location.
- The database keeps track of the owner of each plane, and the employees who have maintained each plane.
- The database keeps track of each airplane's purchase date.
- Each maintenance service record is identified by a work code, and includes the employee who performed the service, the date and time of the service, and the number of hours the maintenance service required.
- Each plane undergoes service many times, and all its service records are kept.
- An owner can be either a person or a corporation.
- A person can be an owner, a pilot, or an employee of the airport.
- Each pilot has specific attributes, including license number and restrictions.
- Each employee has specific attributes, including salary and shift worked.

• The database stores social insurance number, name, address, and telephone number for all person entities.
• The database stores name, address, and telephone number for all corporation entities.
• The database also keeps track of the types of plane each pilot is authorized to fly, and the types of plane each employee is qualified to service.

Draw an object-oriented diagram for the Edmonton International Airport database.