|  |  |  |
| --- | --- | --- |
|  |  | *Algonquin College of Applied Arts & Technology* |

CST8276 - Assignment 02 (100 points, weight 10.0 %)

**Team members(4 persons): Huang, Zhe; Li, Xuan; Lin, Shunbiao; Zhu, Yingmei**

*Submit this assignment before the due date, via BlackBoard. Late assignments are NOT accepted and they will not be graded. Due date: Saturday April 01, at 11:59 PM.*

*Save all the answers in a word file with “****A02\_YourName\_LastName****” as a name. Use this document to answer the questions.* ***Only one submission per group****. The group could contain maximum three students. Write the names of all the students in the group at the top of your submission document.*

## Oracle General Questions

1. State the difference between partial backup and complete or offline backup?

A partial backup is any operating system backup short of a full backup, taken while the database is open or shut down, such as:

* a backup of all datafiles for an individual tablespace
* a backup of a single datafile
* a backup of a control file

An offline backup is an operating system backup of all datafiles and the control file that constitutes an Oracle database. It can be taken when the database is shut down and be used to recover to the point in time of the last complete backup. It is consistent and reliable with redo logs that are generating during backup.

Partial backup is faster but RMAN does not consider partial backups for restore and recovery. Offline backup it is usually slow but consistent and reliable.

1. Why do we use RAID? What is the difference between RAID 0 RAID 1 and RAID 5?

RAID refers to ‘Redundant Arrays of Inexpensive Disks’. We use RAID because it provides a mechanism for load balancing and securing your data across multiple disks. It offers increased performance, fault tolerance, and protect the data.

RAID 0 is also called ‘non-redundant array’; it includes two or more disk drives and provides data striping, where data is distributed evenly across the disk drives. It doesn’t offer data protection since RAID 0 does not maintain redundant data; but it provides improved I/O performance.

RAID 1 is built from two disk drives with same data stored on and one disk drive acts as a mirror of the other. Compared to independent disk drives, it provides improved performance with twice the read rate. However, the capacity of RAID 1 is only half of independent disk drives. If a RAID 1 array is built from different-sized disk drives, drive segment size is the size of the smaller disk drive.

RAID 5 is built from a minimum of three disk drives, and uses data striping and parity data to provide redundancy and improve performance. Parity data is an error-correcting redundancy that’s used to re-create data if a disk drive fails, and it is striped evenly across the disk drives with the stored data. Like RAID 1, drive segment size is limited to the size of the smallest disk drive in the array.

1. What is Flashback? Can we use this technology to recover a dropped table?

Oracle Flashback tools allow [administrators](https://en.wikipedia.org/wiki/Database_administrator) and users to view and manipulate past states of an [instance](https://en.wikipedia.org/wiki/Database_instance)'s [data](https://en.wikipedia.org/wiki/Data) without (destructively) [recovering](https://en.wikipedia.org/wiki/Data_recovery) to a fixed point in time.

Flashback refers to the collection of a group of Oracle database features which provide ways to view past states of database objects, or to return database objects to a previous state, without using traditional point-in-time recovery. It can be used to:

* Perform queries that return past data.
* Perform queries that return metadata showing a detailed history of changes to the database.
* Recover tables or individual rows to a previous point in time.

We use this technology to recover a dropped table, but can’t recover a dropped table to an earlier state across any DDL operations that change the structure of the table.

1. Is normalization a good idea to improve performance when we are dealing with data warehouse database? Why?

The answer is no for most cases. Data warehouse database are mainly read-only enterprise databases and is a time-critical for reading data. Although normalization reduces data redundancy, it costs lots of joins of smaller tables in OLAP query. OLAP query will run too long time, which reduces the performance, not improve the performance.

1. If you have a database performance issue, what are the places (in sequence) that you should look at in order to solve the problem?

I will check the following place in order:

1. Operating System level: log file, CPU/Memory usage, IOExceptions, orphan process, trace file.
2. Database level: open mode and primary role, tablespace, invalid objects, locks, statics andwait events etc.
3. SQL query level: ADDM / ASH Analytics; real-time SQL and database operation monitoring; SQL performance analyzer.
4. What are the advantages and disadvantages of partitioning a table in oracle?

Oracle partitioning means splitting table up to make them smaller. It is used as a divide-and-conquer approach to make it easier to maintain very large objects (tables and indexes) and to speed up certain types of SQL queries.

**Advantages of Database Partitioning Tables in Oracle in three main areas:**

* Performance
* Maintenance
* Availability

Performance advantages

A table is partitioned based on a criterion such as the value for a particular column. If a query requests data with a particular selection condition that would eliminate a complete partition, Oracle automatically ignores that partition in executing the query. In this way, you can partition a large table to get the advantages of a smaller table.   
  
Maintenance advantages

Most maintenance operations can be performed on a single partition. You can backup or recover a partition rather than the entire table. In this way, you can significantly reduce the time required to perform maintenance operations.

Availability

You can also use partitioning to place partitions into different tablespaces to improve availability. One tablespace can go down without affecting the other tablespaces. If a tablespace becomes unavailable, the other tablespaces and their partitions are still available.  
Splitting data into different partitions is also a time saver. When you reduce the amount of data in each partition, you also reduce the amount of time required to recover that partition.

**Disadvantages:**

* additional administration tasks to manage partitions
* need more space to implement partitioning objects
* more time for some tasks, such as create non-partitioning indexes, collect  
  "global" statistics

1. What are the types of table partitioning? Give an example where a hash partitioning is better than range partitioning.

Oracle Partitioning offers three fundamental data distribution methods as basic partitioning strategies that control how data is placed into individual partitions:

* Range
* Hash
* List

**Range Partitioning**

Range partitioning maps data to partitions based on ranges of values of the partitioning key that you establish for each partition. It is the most common type of partitioning and is often used with dates.

**Hash Partitioning**

Hash partitioning maps data to partitions based on a hashing algorithm that Oracle applies to the partitioning key that you identify. The hashing algorithm evenly distributes rows among partitions, giving partitions approximately the same size.

**List Partitioning**

List partitioning enables you to explicitly control how rows map to partitions by specifying a list of discrete values for the partitioning key in the description for each partition. The advantage of list partitioning is that you can group and organize unordered and unrelated sets of data in a natural way.

Give an example where a hash partitioning is better than range partitioning.

For range partitioning, each partition has a VALUES LESS THAN clause, which specifies a non-inclusive upper bound for the partitions. Any values of the partitioning key equal to or higher than this literal are added to the next higher partition.

For example a table has a date column as the partitioning key, the January-2016 partition would contain rows with partitioning key values from 01-Jan-2016 to 31-Jan-2016. If the data for this range is big, the partition will be big and it will lower the performance and maintenance of this partitioned table as a result.

Hash partitioning is the ideal method for distributing data evenly across devices. It gives the partitions approximately the same size, and same performance, maintenance and availability as well. Hash partitioning is an easy-to-use alternative to range partitioning, especially when the data to be partitioned is not historical or has no obvious partitioning key.

## Database General Questions

RichieBank is a financial institution that offers high interest savings accounts for students. In this scenario, a Customer (Student) may have only one Account, but an Account may be shared by more than one Customer (e.g. a married couple). That is, there is a MANY to ONE relationship between Customer and Account.

The following is a partial schema of the RichieBank database.

*Customer(Customer-id: number, Lastname: string, Firstname: string, Accountnumber: number) Account(Accoutnumber: number, Date-opened: date, Balance: currency, Interest-rate: number*)

The Database Management System (DBMS) provide, amongst others the following five functions:

* 1. Uniform data administration,
  2. Reduced application development time,
  3. Data integrity,
  4. Concurrent access, and
  5. Recovery from crashes.

Explain, by using of your own examples against the database, exactly what these five functions ensure.

My answer:

1. Uniform data administration

Uniform data administration is about connectivity and controllability across multiple data sources. Most banks nowadays are located different places all over the world and multiple users may need to access the banking system at different places. For example, the couple in this study case, who have the permission to access the same bank account may travel to different cities. But when they visit online banking, the user interface they see is the same, but the IP address and the transaction behaviours differ, so this data has to be managed uniformly. A DMBS has such ability to allow DBA to administrate corresponding data.

1. Reduced application development time

This is about the application development efficiency. DBMS provides several critical functions such as concurrency control, crash recovery and query facilities, etc. These allow application developers can focus on application coding. In the banking system, developers can easily write a query to retrieve the information from the database. For example, the balance of a specific account can be retrieved by writing a select query from the account table. It significantly reduces the development time.

1. Data integrity

Data integrity is a fundamental concept of information security. It refers to the accuracy and consistency of data and is often used as a proxy for “data quality”. In the RichieBank database system, the data values have to be scandalized according to the business rules, relations, dates and so on. For example, the attribute “interest” for a bank account must be zero or positive. It does not make any sense that it is negative, So the banking system needs this validation.

1. Concurrent access

Concurrent access refers to a DBMS concept that is used to deal with the simultaneous accessing or altering of data in a multi-user system. When the simultaneous transactions occur, data integrity should be preserved. In the banking system, two users (a couple, for example) who are authorized to access the same account may request a transaction simultaneously, for example, withdraw money and deposit money at the same time without notifying each other, then concurrency control should be undertaken. Otherwise, the error may occur to this account. So a DMBS has this ability to address the conflicts.

1. Recovery from crashes

Transactions or sets of tasks against a database can be interrupted unexpectedly such as power failure on the machine, hardware failure, and operating system error, etc. If failure occurs, crash recovery, a fundamental function provided by a DBMS can move the failure state back to a consistent and usable state. In a banking database system, recovery from cashes is very critical since the users’ finance information can never be lost. For example, suppose the balance data is gone after power failure on a server, then how can the bank provide the withdraw service to the customers without their transaction record. Fortunately, a DBMS provides a sound ability to protect from these failures.

## EER Diagrams

You decided to spend a vacation in in Bora Bora, an island in the Leeward group of the Society Islands of French Polynesia.

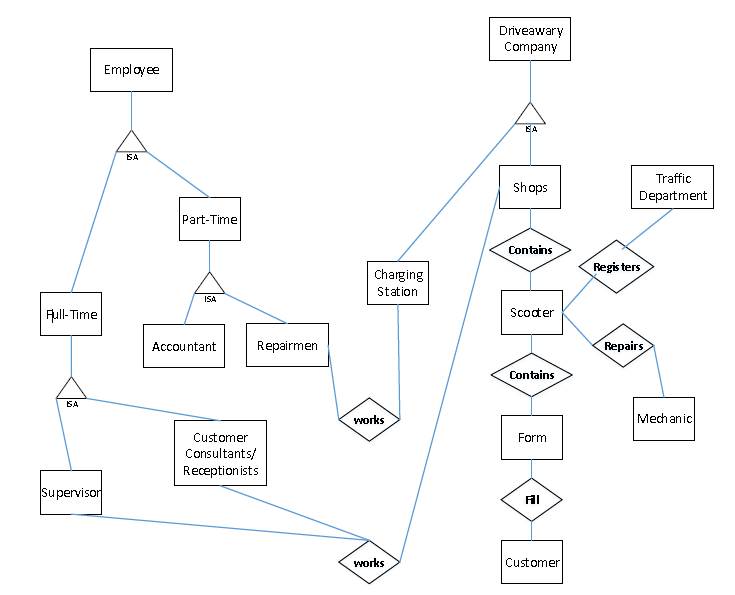
In order to pay from your trip, you accepted a position as a database application programmer at the DriveAway electric scooter rental company on the island. company. This is the only electric scooter rental company on the island.

As a first step, you are asked to create an EER diagram that captures the following business rulesTypically, DriveAway’s customers are tourists that visit the island for periods ranging from one week to 6 months. Customers are identified by their client number, and we also record their addresses while on Bora Bora. In addition, the passport information and home addresses of all customers are stored.

1. A customer rents one or more scooters from the company and a scooter may be rented by many customers, obviously for rental periods which do not overlap.
2. The shortest period that a scooter may be rented is for 8 hours, and the longest contract is for 1 month.
3. There are three locations (shops) on the island where a customer may rent a scooter from. The customer is always required to return a scooter to the location where he rented it from.
4. DriveAway has a fleet of 50 electric scooters. For each scooter, we keep information about the color, the year purchased, the current kilometer reading, the brand name, the model, the power, the voltage, the charging time and the range per charge.
5. Each scooter needs to be registered and the registration has to be renewed every year, at the traffic department of French Polynesia, which is located in Papeete (the capital of French Polynesia).
6. A scooter is serviced at a regular interval, usually after every 800 kilometers of travel. Sometimes a scooter may break down and then it needs to be repaired. There is one mechanic on the island who services and repairs all the scooters. Note that she is not employed by DriveAway.
7. DriveAway also maintains 50 charging stations, throughout the island. The customers may use these stations to recharge their scooters. This service is free of charge and available 24/7.
8. An employee is responsible for completing all the necessary forms when a customer rents a scooter.
9. There are 15 employees that work for DriveAway, including two (full‐time) supervisors and two (part‐time) charging station repairmen. The company further employs one (part‐time) accountant who works from home. The other employees are full‐time customer consultants that also act as receptionists. These employees are deployed at any one the three shops, as need be. Full‐time employees receive a salary and benefits such as health care, while part‐time employees are paid by the hour and do not receive any benefits. All employees are required to pay income tax.

## Your task

Create an EER diagram that models the DriveAway database. Your model should include entities, relationships as well as all attributes. Remember to state all your assumptions.



**Here are the tables corresponding to this diagram**

1. Employee(EID,Fullname,Salary,type[FullTime(Supervisor,Customer Consultant(Receptionist))/PartTime(Repairmen,Accountant))]

2. Customer(ClientNumber,ClientName,IslandAddress,PassportInformation,HomeAddress)

3. DriveAwayCompany(Shops,ChargingStation)

4. Scooter(ScooterID,BelongShop,Color,PurchasedYear,Kilometer,BrandName,Model,Power,Voltage,ChargingTime,ChargingRange,RenewDate,RepairInformation,AvailableOrNot)

5. Form(FormNumber, EID,ClientNumber,ScooterID,RentPeriod)

6. Mechanic(RepairRecordID,RepairDay,RepairInformation,ScooterID)

7. Bill(ScooterID,BrokenInfo,Amount)

8. Traffic Department(ScooterID,FirstRegisterYear,RenewDate)

**Here are the design choices:**

9 If a customer delay to return scooter, another form will be filled in by employee with the rentPeriod with pending, and the scooter status will show notAvailable. Until the customer return the scooter, RentPeriod will be filled in and the status show available.

10 If a scooter is broken, DriveAwayCompany will send it to Mechanic. Then the bill will send back to company. Amount percent will be charged for customer dependent on the broken reason.

11 If the customer do not return the scooter or lost, amount will be charge for customer dependent on the scooter PurchasedYear. Then DriveAwayCompany will purchase another one.

