BCDE214 Database Administration

Assignment 3 – Portfolio

Appendix C

By Daria Prianishnikova

# Contents

[Contents 2](#_Toc56543181)

[Recovery 3](#_Toc56543182)

[Transaction Management 5](#_Toc56543183)

[Distributed Databases 7](#_Toc56543184)

[Data Warehousing 9](#_Toc56543185)

[Communication Time Calculation 11](#_Toc56543186)

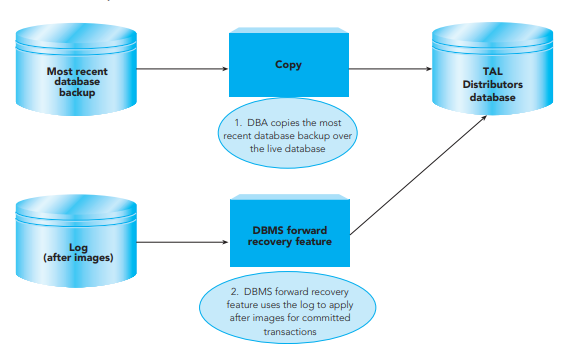
[Spatial Query 13](#_Toc56543187)

[References 15](#_Toc56543188)

# Recovery

1. When does a DBA perform forward recovery? What are the forward recovery steps? Explain in detail.

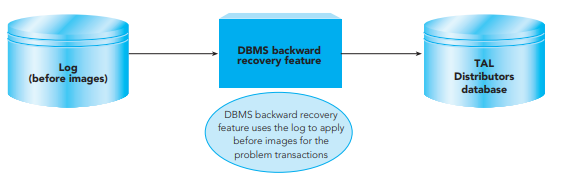
(5 marks)

Forward recovery is a process of copying the backup over the live database and applies the after images of committed transactions from the log to keep the database up to date. A DBA performs forward recovery if a database is destroyed (Pratt & Last, 2015). Forward recovery consists of two steps (Pratt & Last, 2015):

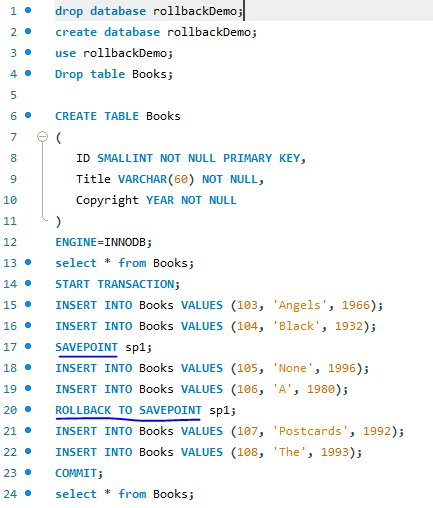
* The DBA copies the most recent database backup over the live database, which is destroyed.
* The DBMS forward recovery feature uses the log that is kept after each backup to apply after images of each record in the log for committed transactions and bring the database up to date.

1. If a transaction is interrupted, what steps does the DBMS take to correct the database? Explain in Detail. Show example with Roll back and Save point in MySQL.

(5 marks)

Backward recovery, or rollback, can be used to recover the database to a valid state by undoing the problem transactions (Pratt & Last, 2015). The DBMS accomplishes the backward recovery by reading the log for the problem transactions and applying the before images to undo their updates.

In MySQL, it can be achieved using SAVEPOINT and ROLLBACK statements. For example, this script shows the use of these commands:

Only the data added on line 15, 16 and 21, 22 will be present in the database.

Lines 15-16 were added before creating a savepoint, and lines 18-19 will be erased due to rollback recovery performed.

Lines 21-22 are added after performing the recovery, and on line 23 the transaction is finished or committed.

Therefore, after committing the transaction, 2 rows of data will be lost.

# Transaction Management

1. Briefly explain what is meant by a “lost update” and give an example to illustrate your answer.

(3 marks)

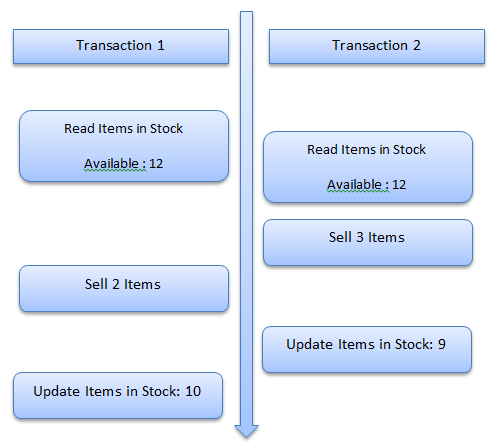
A lost update could occur in a concurrent update situation when two users attempt to update the same data at the same time, and the DBMS does not have concurrency update features such as two-phase locking or timestamping (Pratt & Last, 2015).

For example, a scenario where two people initiating transactions at the same time and working on the same table (Richardson, 2017).

User 2 updates the stock of an item by selling 3 items. However, User 1 then updates the stock at the same time and their update overwrites the User 2’s update. Therefore, data is incorrect and lost.

Instead of 10 items in stock it should be 7 items left, but data from transaction 2 is now lost and the stock number is wrong by 3 sold but “lost” items.

Figure 1 Lost Update scenario. Source: (Richardson, 2017)



**Lost update**

1. **In what circumstances can the problem of “deadlock” or “deadly embrace” arise? Support your answer with an example.**

(3 marks)

Deadlock or deadly embrace can occur when two or more users are each waiting for the other(s) to release a lock for their transactions before they can proceed and commit the changes (Pratt & Last, 2015).

Waiting to lock Table B record already locked by user 1

Waiting to lock Table A record already locked by user 2

**Deadlock Scenario**

For example, user 1 locked Table B, and user 2 locked Table A. Now User 1 is waiting for table A to commit changes made to Table B, but at the same time user 2 needs to lock Table B to commit changes to Table A. Both users are deadlocked while they wait for the release of records they need to lock next.

1. **Outline two (2) ways in which the problem of “deadlock” or “deadly embrace” can be overcome by a database management system.**

(4 marks)

From (Pratt & Last, 2015):

1. Let the deadlock happen first, and then allow the DBMS to detect and break it. The DBMS should keep track of all records it has locked in each transaction and records that are waiting to be locked. After detecting that there are locked records held by other locked records, the DBMS chooses a “victim” user, for whom it will undo all changes, release all locks, and reschedule the transaction. While the deadlock is resolved, the user will only feel a slight delay.
2. Recovery through rollback:

* checkpoint a process periodically
* use this saved state
* restart the process if it is found deadlocked

# Distributed Databases

1. **Briefly describe the concept of location transparency, using an example to support your answer.**

(5 marks)

Distributed databases – single logical database that is physically divided among multiple computers at several sites on a network – all share the same characteristics, one of them being location transparency (Pratt & Last, 2015).

Location transparency stands for users being able to access data at the remote site from the local site as if the entire database is stored at the local site (it could take longer, but nevertheless without any restrictions) (Pratt & Last, 2015). A user must be unaware of the location of data in a distributed database, all this information must be hidden.

An example could be creating a view using two tables that are stored at two different sites (Fogel, 2001). There are two servers and two physical databases: a remote sales database and a local HQ database. Table emp is stored in a local database and table dept is stored in a remote database. To make these two tables locally transparent to a user, a “COMPANY” view is created, and a user accesses it from a PC. When users access this view, they do not need to know where the data is physically stored, or if data from more than one table is being accessed (Fogel, 2001).

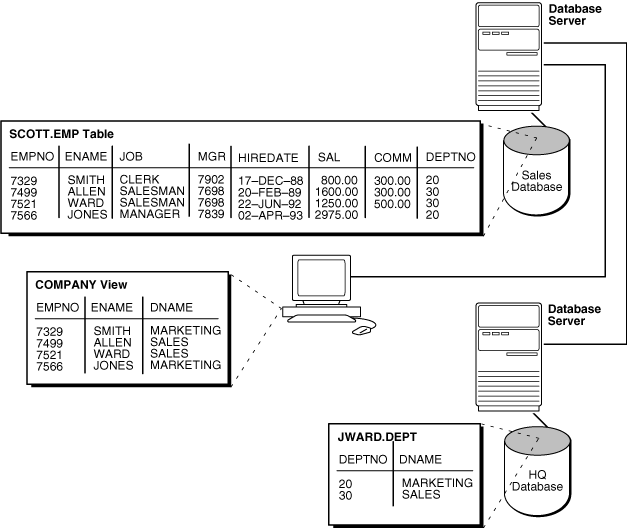


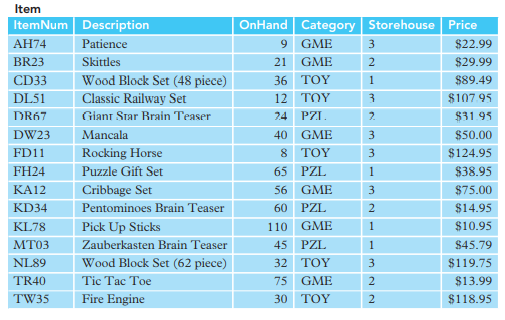
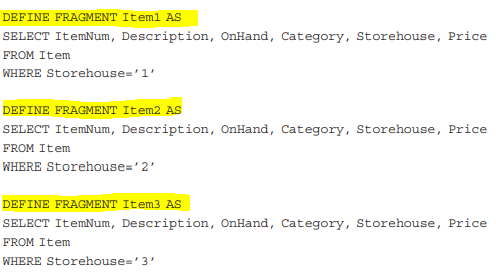
Figure 2. Creating a View using Location Transparency characteristic. Source: (Fogel, 2001)

1. **Briefly describe the concept of fragmentation transparency and explain why fragmentation transparency is done, using an example to support your answer.**

(5 marks)

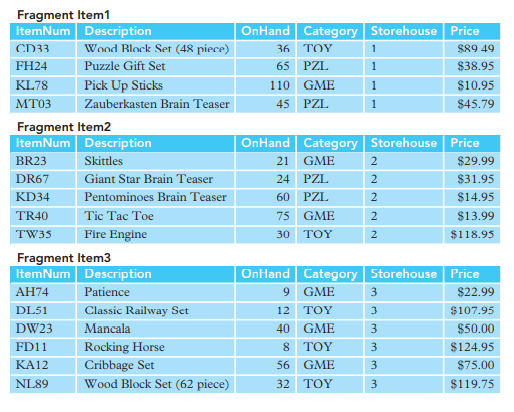
A distributed database management systems support data fragmentation, which is a process of dividing and managing a logical object among the various locations under its control (Pratt & Last, 2015).

To illustrate an example of fragmentation, I will use the following example (Pratt & Last, 2015):



**SQL-type commands**

**Original table**



**Result**

A “Item” table has been fragmented into three pieces, each fragment corresponding to a physical warehouse that has those items in stock. The data from one centralised database is fragmented and placed into three local databases of warehouses, where this data is mostly often used.

The principle of fragmentation transparency, same as with location transparency, is that users should not be aware of the fragmentation, and should feel like they are using a single central database (Pratt & Last, 2015).

# Data Warehousing

1. **What is a data warehouse and what are its main characteristics?**

(2 marks)

Data warehouse is a system with a large storage used for reporting and data analysis for companies (Pratt & Last, 2015). Main characteristics of a data warehouse are:

* Subject-oriented
* Integrated
* Time-variant
* Non-volatile.

1. **What does it mean when a data warehouse is non-volatile?**

(2 marks)

Data warehouse is non-volatile due to data being read-only (Pratt & Last, 2015). Users cannot upload, create, or delete data to a data warehouse directly. Data is loaded into a data warehouse periodically.

1. **Why would an organization invest in a data warehouse?**

(2 marks)

An organisation would invest in a data warehouse for:

1. RDBMS and OLTP often suffer from severe performance problems, which makes them unsuitable for deeper data analysis (Pratt & Last, 2015).
2. Data warehouses’ functionality allows to clean data, transform, catalogue, and make available for use by managers and other business professionals for data mining, online analytical processing, market research and decision support.
3. **What is a fact table in a data warehouse?**

(2 marks)

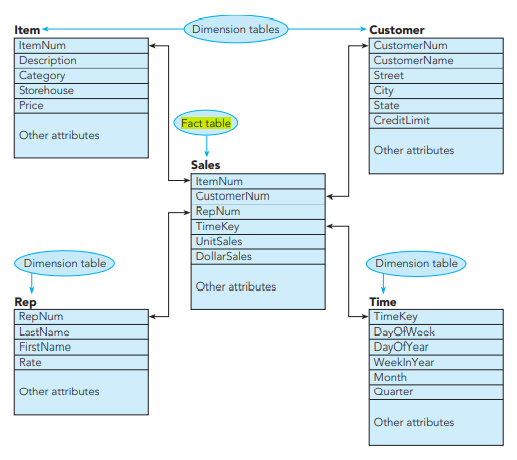
A fact table consists of rows that contain consolidated and summarized data. The fact table contains a multipart primary key, each part of which is a foreign key to the surrounding dimension tables (Pratt & Last, 2015). The overall structure shown in the figure below is called a star schema because of its conceptual shape (Pratt & Last, 2015):

Figure 3. Source: (Pratt & Last, 2015)

1. **What is a dimension table in a data warehouse?**

(2 marks)

A dimension table is a table in a star schema of a data warehouse. A dimension table stores attributes, or dimensions, that describe the objects in a fact table (Rouse, 2012). The fact table contains a multipart primary key, each part of which is a foreign key to the surrounding dimension tables (Pratt & Last, 2015).

Communication Time Calculation

A Toy Company wants to list all items in TOY category with a price that is more than $100.00. For this query assume the following set of assumptions:

1. The Product table contains 1000 rows and is stored in a remote site;
2. Each record in the Product table is 500 bits long;
3. There is no special structure, such as an index, that would be helpful in processing this query faster; and
4. Only 10 of the 1000 rows in the Product table satisfy the conditions.

**How would you process this query?**

**Calculate the communication time for your query strategy**.

Assumed numbers:

Access delay: 3 seconds

Transmission rate: 100,000 bits per second

Communication time = access delay + (data volume / transmission rate)

Retrieving all records from the table: 3 + ((1000\*500) /100,000)) = 8 seconds

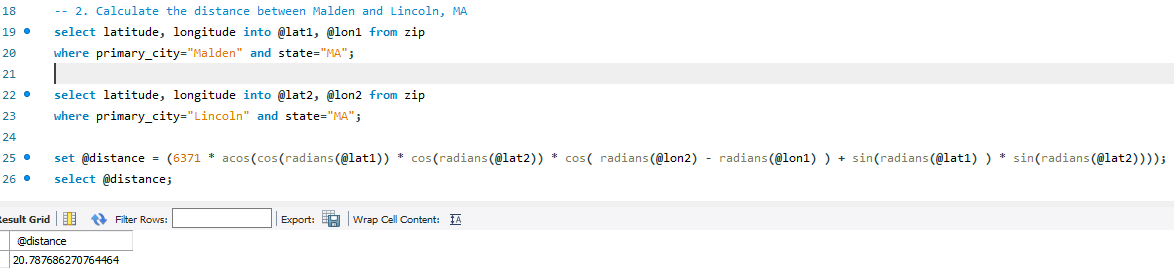
Retrieving only 10 rows that satisfy the conditions: 3 + ((10\*500) /100,000)) = 3.05 seconds

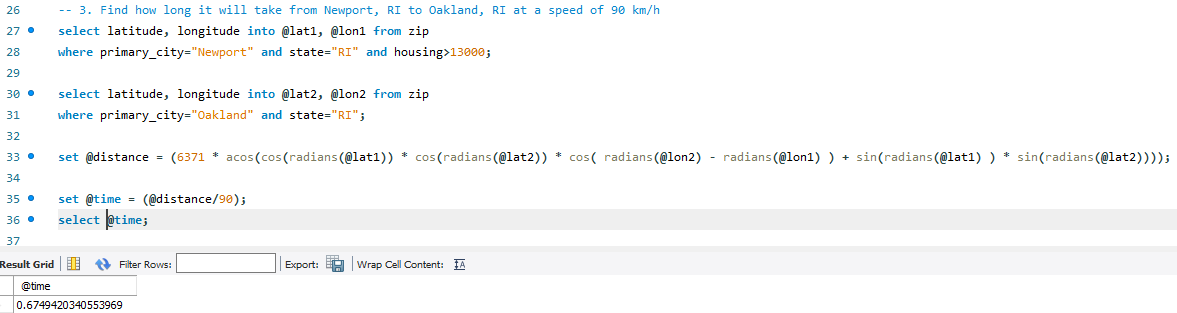
I would retrieve only the 10 rows that satisfy the conditions as it is 4.95 seconds faster.

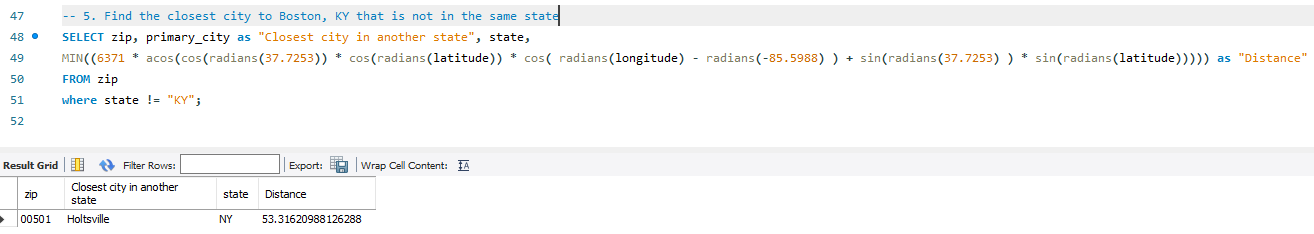
Spatial Query  **(Marks 5)**

Write a query that will:

Calculate Distance between any **two** locations (in Kms.) as provided in SQL file. Test your query by providing at least **5** suitable test scenarios. Use the Zip.SQL file for this task.







# References

Fogel, S. (2001). *Oracle Database Administrator's Guide.* Oracle Corporation. Retrieved from https://oracle.su/docs/11g/server.112/e10595/ds\_admin006.htm

Pratt, P., & Last, M. (2015). *Concepts of database management.* Cengage Learning.

Richardson, B. (2017, September 22). *The Lost Update Problem in Concurrent Transactions*. Retrieved from Coding Sight: https://codingsight.com/the-lost-update-problem-in-concurrent-transactions/

Rouse, M. (2012, April). *What is dimension table?* Retrieved from Seach Data Management TechTarget: https://searchdatamanagement.techtarget.com/definition/dimension-table