

Kramer Johnson

Homework 5

Database Systems 5021 02

Transactions / Warehouse

a. The main reason that someone would want to relax an isolation level in their database is to increase the speed of concurrent transactions. Perhaps in this database the data integrity is not of the utmost importance but the ability for many users to access data extremely quickly is. The database administrator could lower the transaction isolation level which would allow concurrent transactions to access the same data at the same time.

The downside of this approach is the reduction in data integrity. If concurrent transactions can access the same data at the same time then one transaction may read a row then update it but after it reads the row, a separate, concurrent, transaction updates that same row. After the first transaction completes, its read will no longer be accurate. If this happens often, users could begin to have many inconsistencies in their queries.

b. The “A” in the ACID acronym for databases refers to the atomicity requirement. The requirement states that each transaction is atomic in that the transaction is one part and all the transaction is performed, or none of it is performed. For example, if a transaction is making an update in the INVOICE table and it then is supposed to make an update in an INVOICE_LINE table, but it is interrupted, perhaps due to a system fault, then the entire transaction is scrapped. This way the INVOICE data does not remain with an update that is, in essence, only half of a transaction.

c.

Script 1:

```
START TRANSACTION;
```

```
UPDATE actor SET first_name = 'bob' WHERE actor_id = 25;
```

```
SELECT SLEEP(10);
```

```
UPDATE actor SET first_name = 'frank' WHERE actor_id = 25;
```

```
SELECT first_name FROM actor WHERE actor_id = 25;
```

```
COMMIT;
```

Script 2:

```
START TRANSACTION;
```

```
SELECT SLEEP(5);
```

```
SELECT first_name FROM actor WHERE actor_id = 25;
```

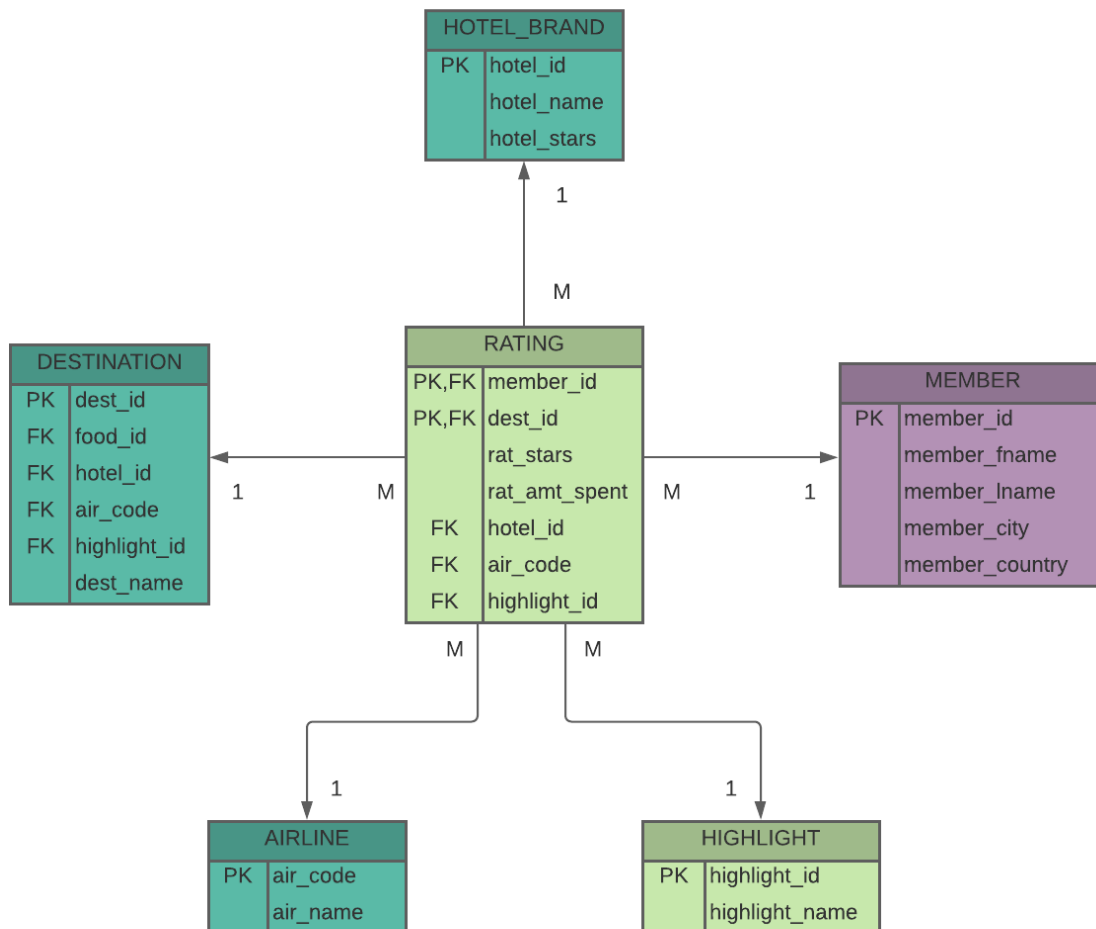
```
COMMIT;
```

d. The main reason someone would prefer to use READ COMMITTED instead of READ REPEATABLE is the same reason a user would prefer to use any lower isolation level, to increase transaction concurrency. READ REPEATABLE will result in no non-repeatable reads, which is desirable, but it does so by increasing the scope of locks placed on data. With more locks, you may end up with more concurrent transactions being queued while they wait for locks to be removed.

e. The most important value of a transaction log is the ability to restore a database after corruption. A transaction log stores all data about every transaction including the type of transaction, the objects affected, the before and after values, and pointers to the previous and next transaction log entries for the same transaction. With this data in hand, if some kind of server error causes the database to crash before some data is written to file (but has been committed), the database can roll forward the updated, but unwritten data. Similarly, if some data has not been committed prior to the server failure that database can be rolled back to a state prior to the update. Since failures can regularly happen in a database, the ability to restore the database is essential and the transaction log is essential to that process.

Data Warehouse

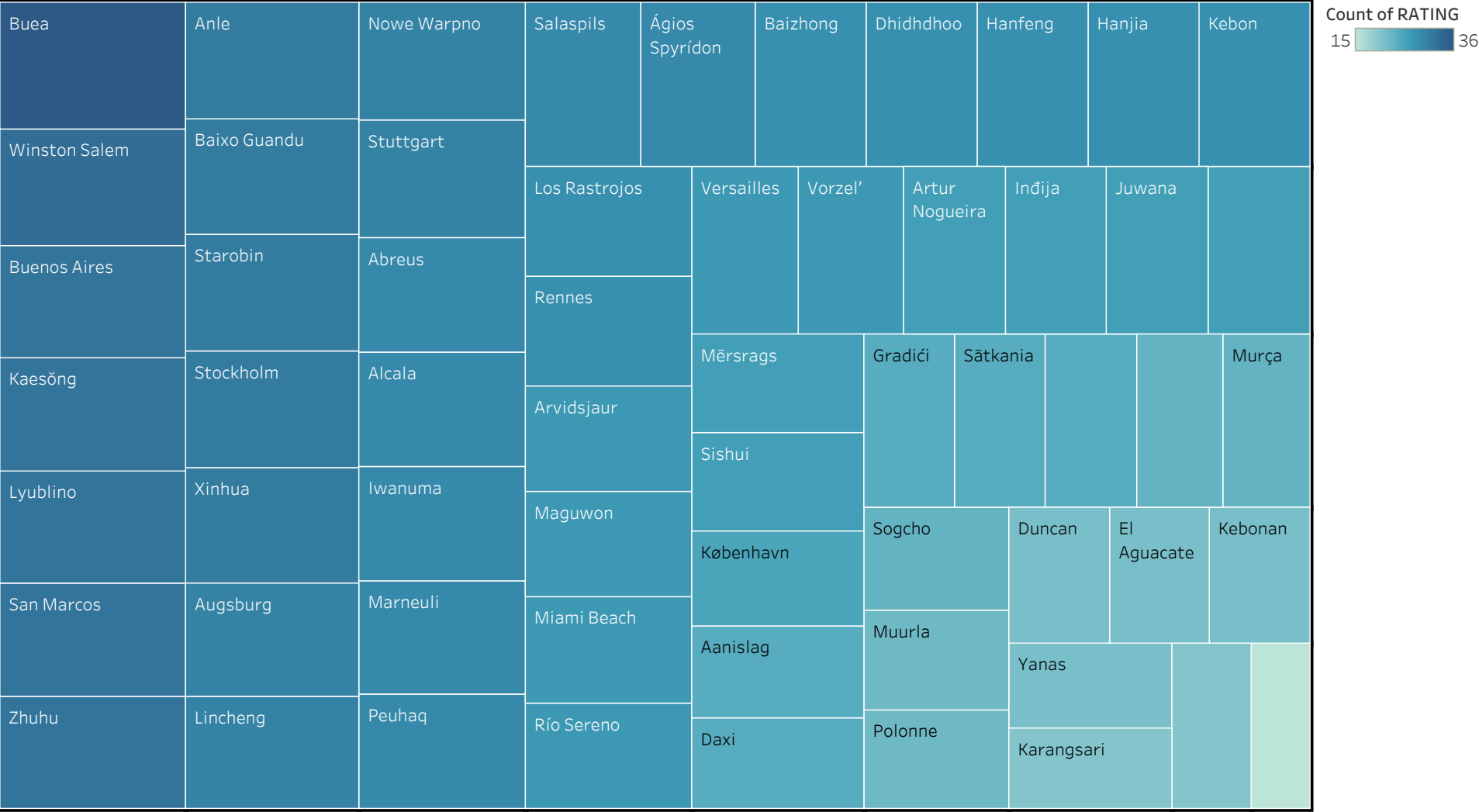
a.



b. A data warehouse is a database that stores historical data from an operational database that is used for analysis to gain business insight. Data is taken from the operational (day-to-day) database and integrated to the data warehouse. Once there, users can run all kinds of analytical queries on it without having to be concerned with bogging down the operations of the business. For example, data could be gathered about a website's users' social media activity throughout the day. Then, each night during the lowest activity of the site, the daily data could be integrated into the data warehouse.

c. Data analytics is the process of combing through data to discover meaningful insights and patterns from the data and use that information to make decisions and predictions. Data analytics is mostly performed on data in a data warehouse so that forecasts and decisions can be made, and businesses can learn useful information about their data. For example, analytics could be run to find out how likely a website's user is to post pictures of food when they are a user who posts pictures on a daily basis.

Most Visited Destinations



Dest Name. Color shows count of RATING. Size shows count of RATING. The marks are labeled by Dest Name.