1/1/2020

Arslan Hussain

LSDDA Report

Introduction

One of the largest technological challenges in software systems research today is to provide mechanisms for storage, manipulation, and information retrieval on large amounts of data. This report will contain the advantages and disadvantages of using NoSQL and relational databases. The databases discussed will have the positive and negatives of why they are used.

NoSQL Databases

This database is created to sacrifice the restrictions of the traditional relational databases. It does not require complex data models and schemas. The database is more appealing to business the deal with large amount data and may need an option to increase the capacity servers to use this database is due to the databases being scaled horizontally which allows the users to increase the database scale by attaching additional machines. There are three common and basic variations. Starting with document based, key value stores and wide column stores. The document base store is designed to store and query data as Json-like documents. It is vital for developers and simplifies storing and querying data in database by using the same document used in an earlier stage. The key value store lets a user store data without using a predefined schema. That data used in the database is saved as an array in map or dictionary of format using keys. The key is a string which can be a file name or hash and the data can contain an image, or document. This can only be stored in pairs of keys and values, as well as retrieve values that are known. The wide column stores allows for information to be stored in columns together. It also stores data in records with an ability to hold very large numbers of dynamic columns. The column names as well as the record keys are not fixed, and since a record can have billions of columns, it can be stored as a two dimensional key-value stores. The most popular examples used for this variation are Cassandra and HBase that are optimised for queries over large data sets.

The advantages of these databases that NoSQL is easy to rescale when needed. This makes good use of new nodes and are designed with expansion in mind. Large data is another advantage as it easier to manage volumes of data when compared to relational database. It is also containing easy maintenance as NoSQL does not require much hand on management with the inclusion of data distribution and auto repair functionality. The disadvantage is that there are only small amount advanced users of NoSQL meaning that there is less help when its needed.

Relational database

A relational database allows users to store a collection of data which is organised as a set of tables. The use of this type of database, the data can be used or changed in many ways the worry of adjust the database. Data that is inserted in a table must contain and match the data type that is with in the column of the table that it is being inserted to. the relations in every row a primary key is defined to make it unique and a foreign key is added which reference to its primary key in another table. These two keys show a relationship between multiple tables and is seen as one to one or one to many or many to many relationships. The advantage a relation database is that it is easy to use when the information stored in the correct manner that data is easy to find and understand. There is also more security when using a relational database, as data is split into multiple tables and to make it more secure some tables can request some credentials of the users before viewing or modifying data. The disadvantage of this is when depending on the size of the database it can affect the speed of processing data significantly, with difficult queries requiring more processing power.

Conclusion

In conclusion the database used for this project is NoSQL due to its being more flexible, and it is faster than a relational database when it needs to be adapted to different circumstances. With a large amount of organisations using a NoSQL database due to data being retrieved faster and more data being stored it simplifies the data found with the queries used.

PHP

The programming choice is PHP which is the most popular scripting language. PHP is cross platform programming language that can be used in many operating systems. The process in which PHP work is that it need some kind of script in order for to work fully. PHP is main source of connection between a HTML page and a database. In addition, the programming language can help change data on a database and call specific data that it is called instead of the whole collection the data. PHP is not just limited to HTML and can produce anything from an image to a flash movie.

Normalisation

This is a process of taking data from a problem and reducing it to a set of relations along with making sure there is data integrity and elimination data redundancy. Data integrity is that all the data in the database are consistent and satisfy all integrity constraints. Data redundancy is when data in the database can be found in two different tables or if data can be calculated from other data items then the data is said to contain redundancy. This will ensure the data is accurate and there will not be any confusion when there is a search for data. In order to find data through query’s the data has to be normalised which will help create relationships between the tables, for this to happen primary and foreign key have to be used in normal formal. The primary key identifier cannot be changed in other tables and foreign key is used to reference the primary key in another table. The redundancy can be shown in the table by multiple normal forms which go from the first normal form to the second and third and so on.

The first normal form ensures that every column in the table must be unique and separate tables must be created for each set of related data. The table will be identified by a unique column called the primary key. The rows and columns will not be duplicated along with no row or columns intersection contain a null value or multivalued fields.

The second normal form is like the first normal form in which every non key attribute is fully functionally dependent on the primary key. For example, in a table, if attribute B is functionally dependent on A, but it is not functionally dependent on a proper subset A, then B is considered fully functional dependent on A. however, in a second normal form table, all non-key attributes cannot be dependent on a subset of the primary key. This shows that a table that is in first normal form and contains only a single key as the primary key is automatically in second normal form.

In the third normal form is like it’s in the second normal form and no non key attribute is transitively dependent on the primary key. Due to being transitively dependent, by this it is meant that there is the following relationship in the table. A is functionally dependent on B, and B is functionally dependent on C. in this case, C is dependent on A and B and contains the foreign key.

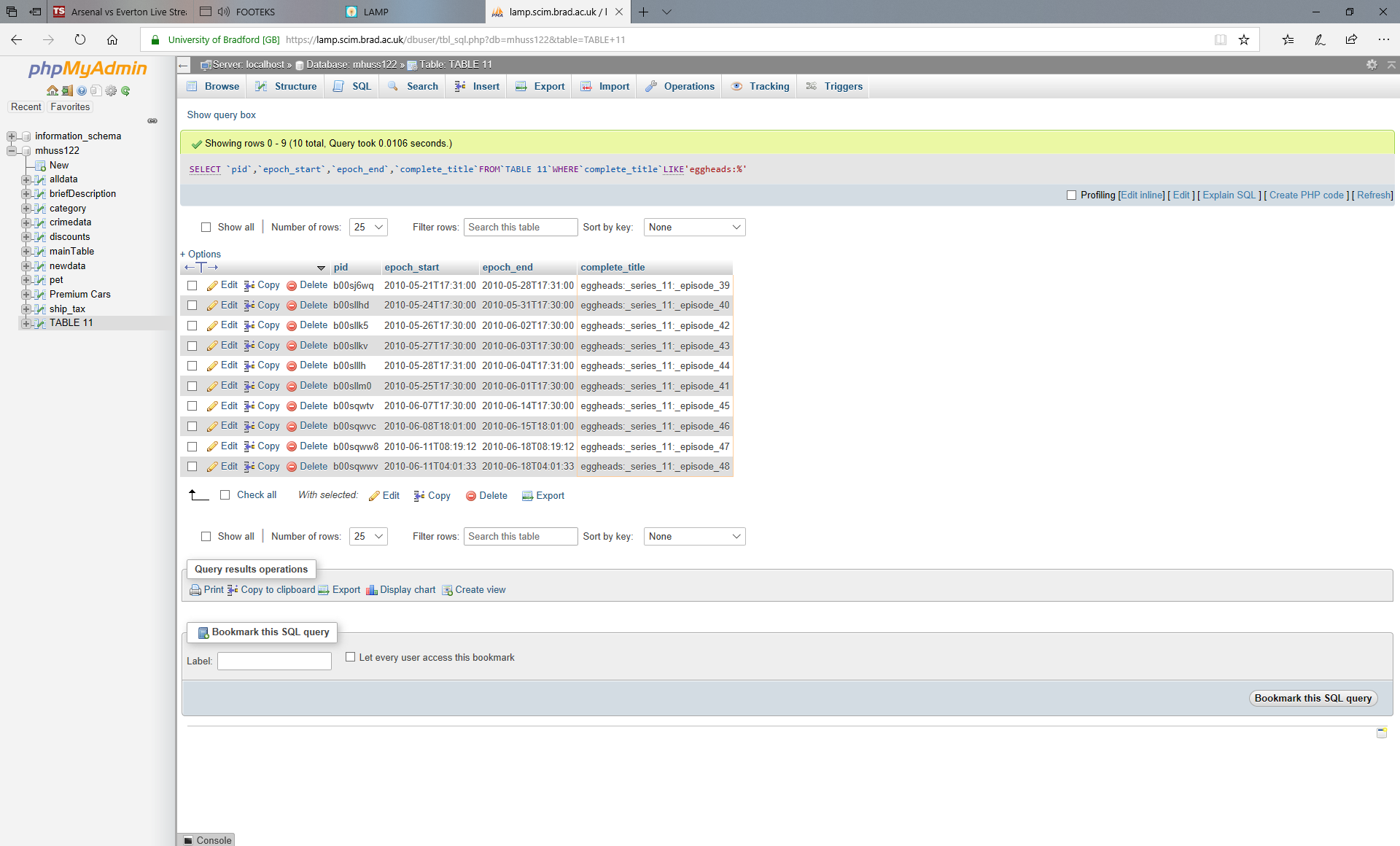
ERD

The entity relationship diagram (ERD) represents the relationship of the entity set stored in a database. the diagram is often used to design or debug relational databases. It focuses on the relationships between entities themselves.



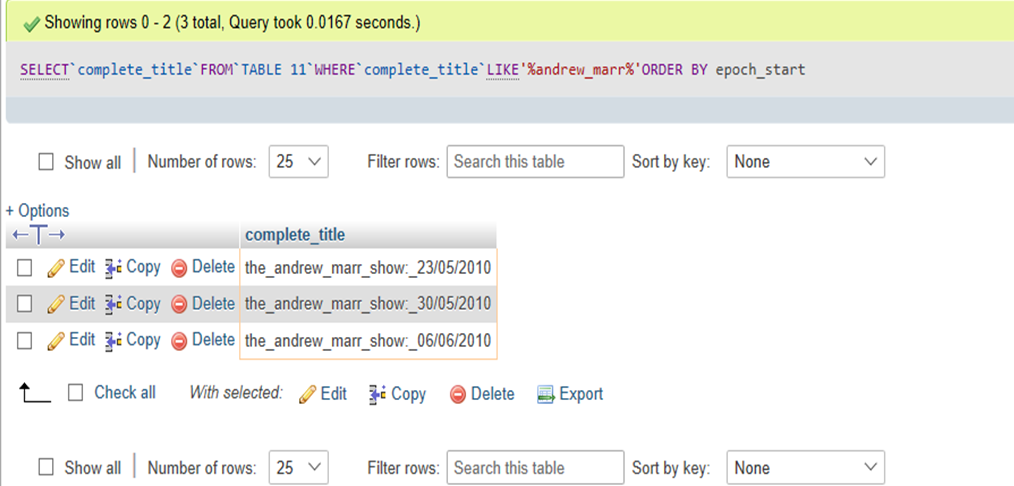
This is the ERD diagram used.

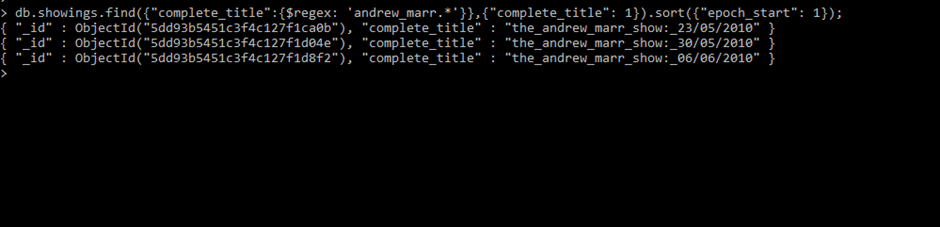
First Query

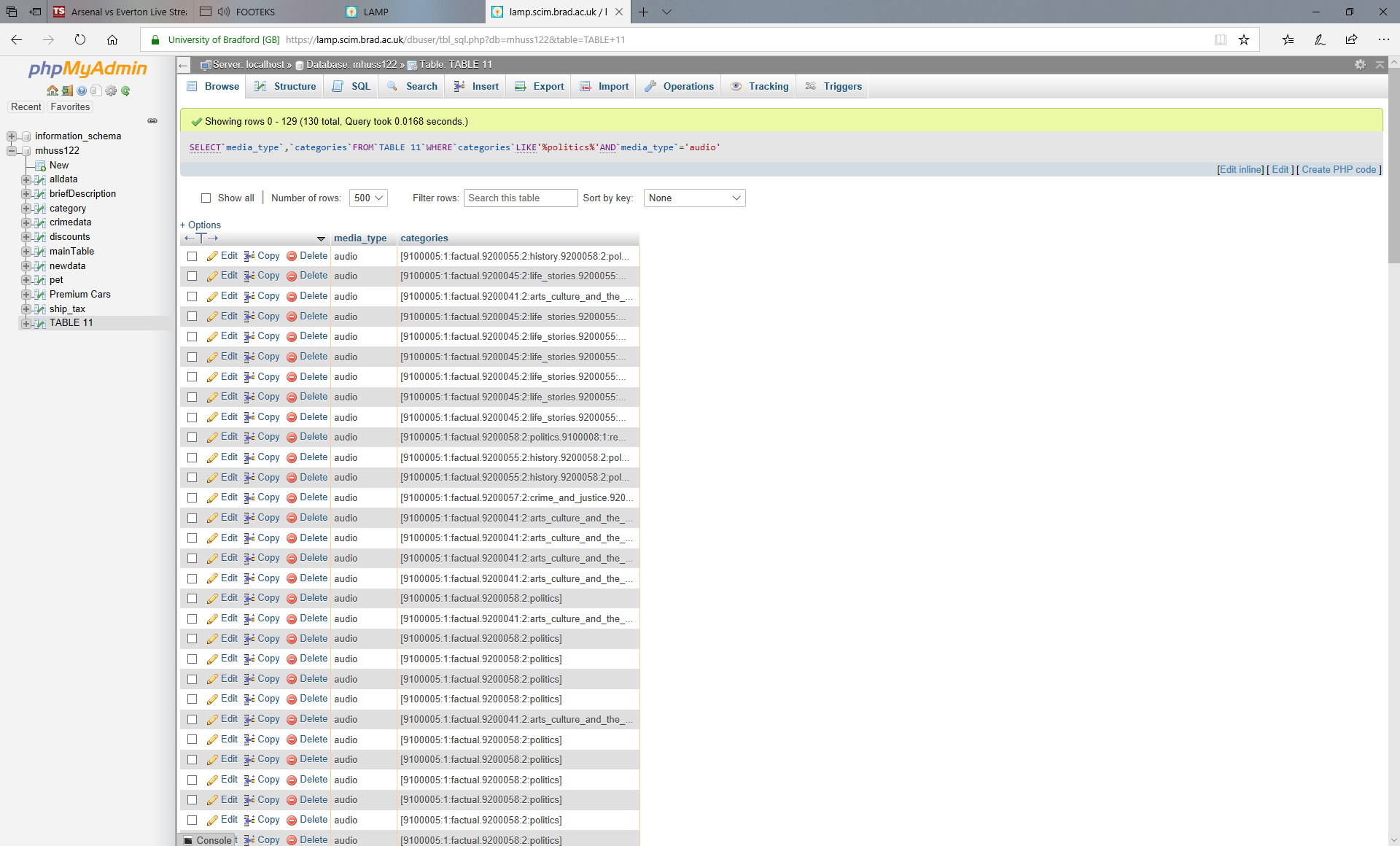


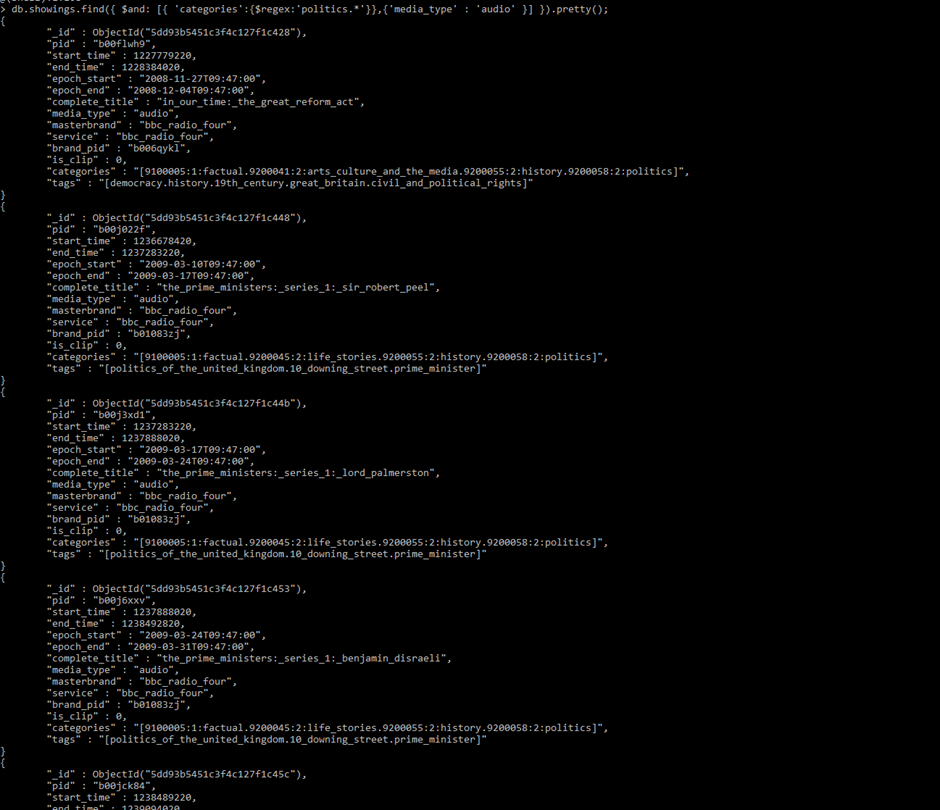


Second query

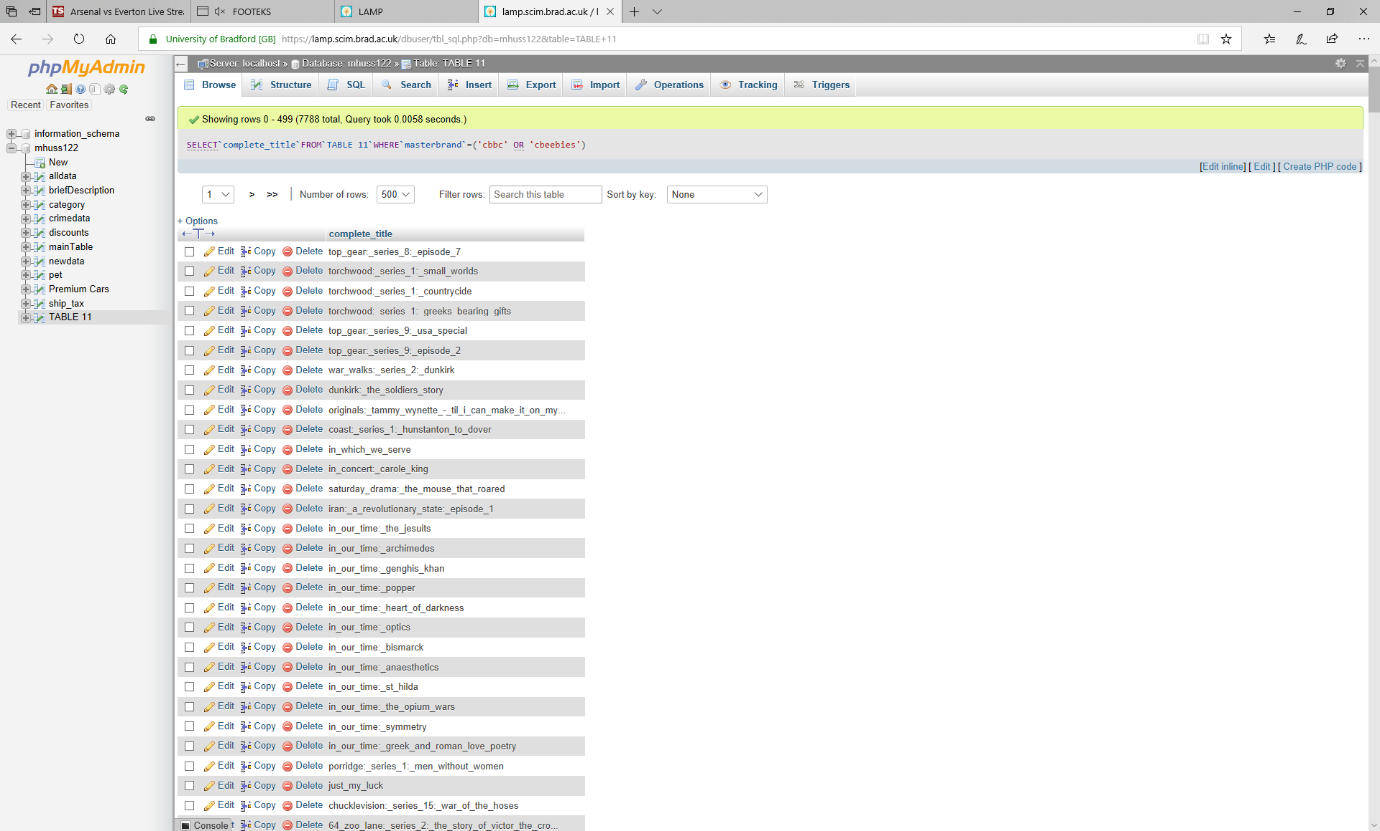


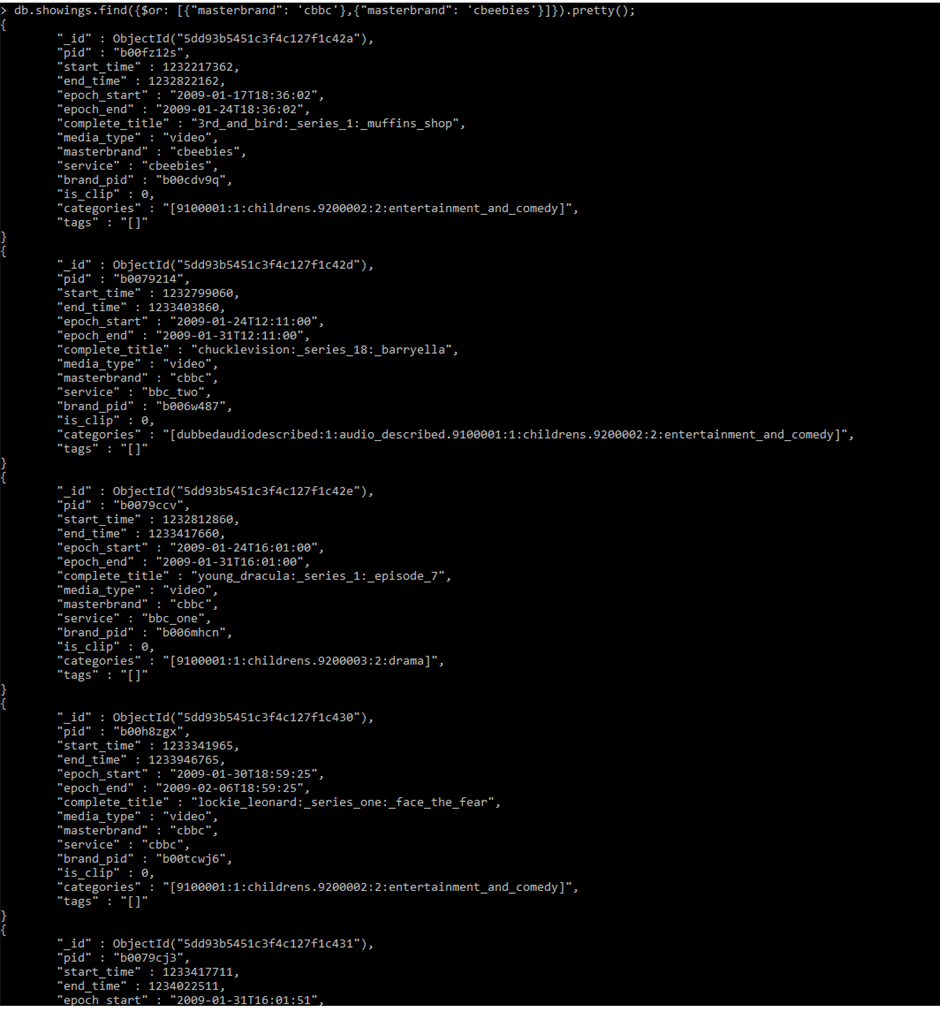


Third Query

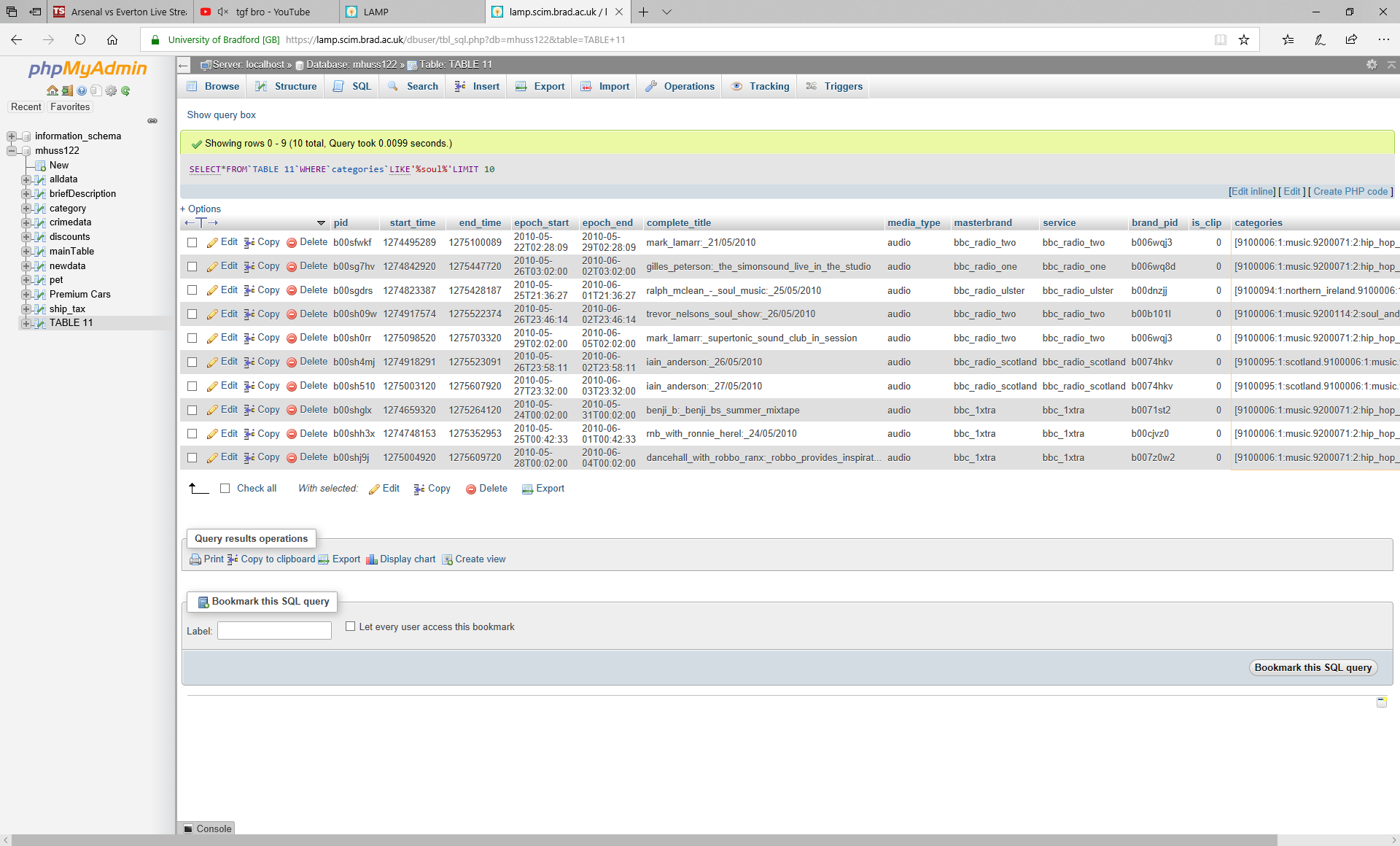


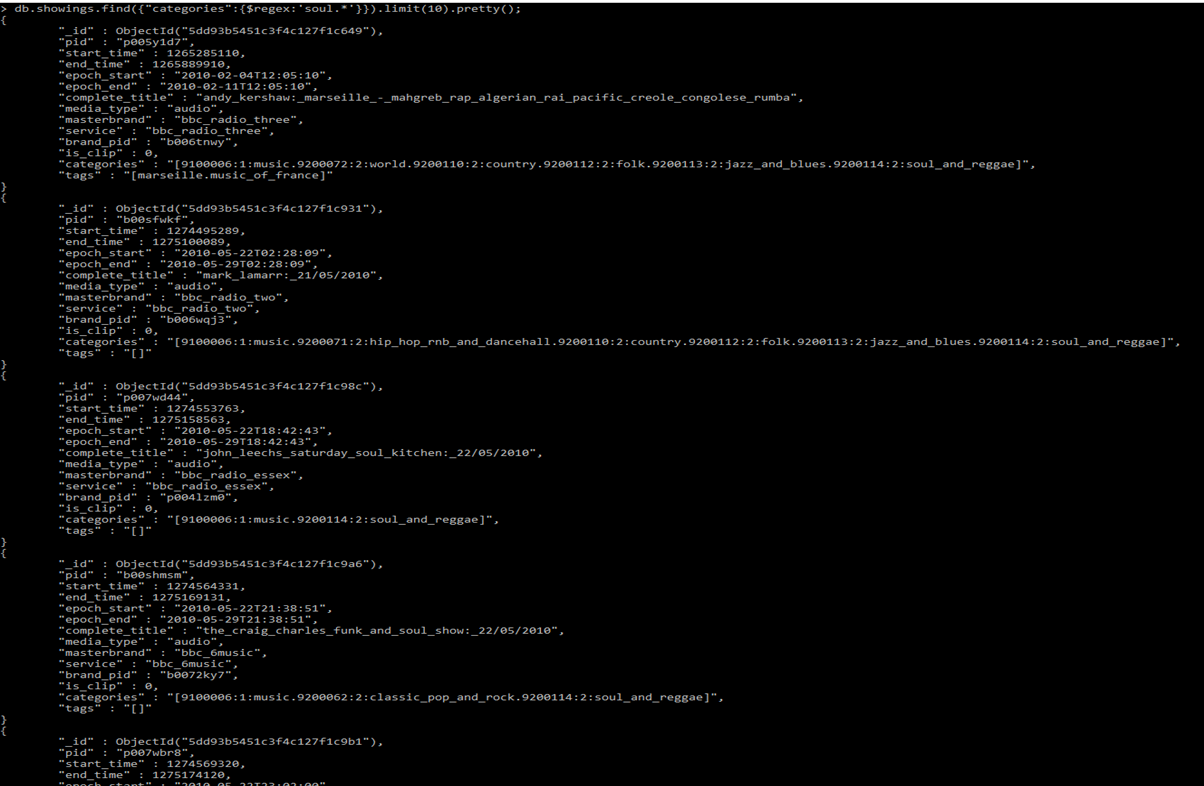
Fourth Query

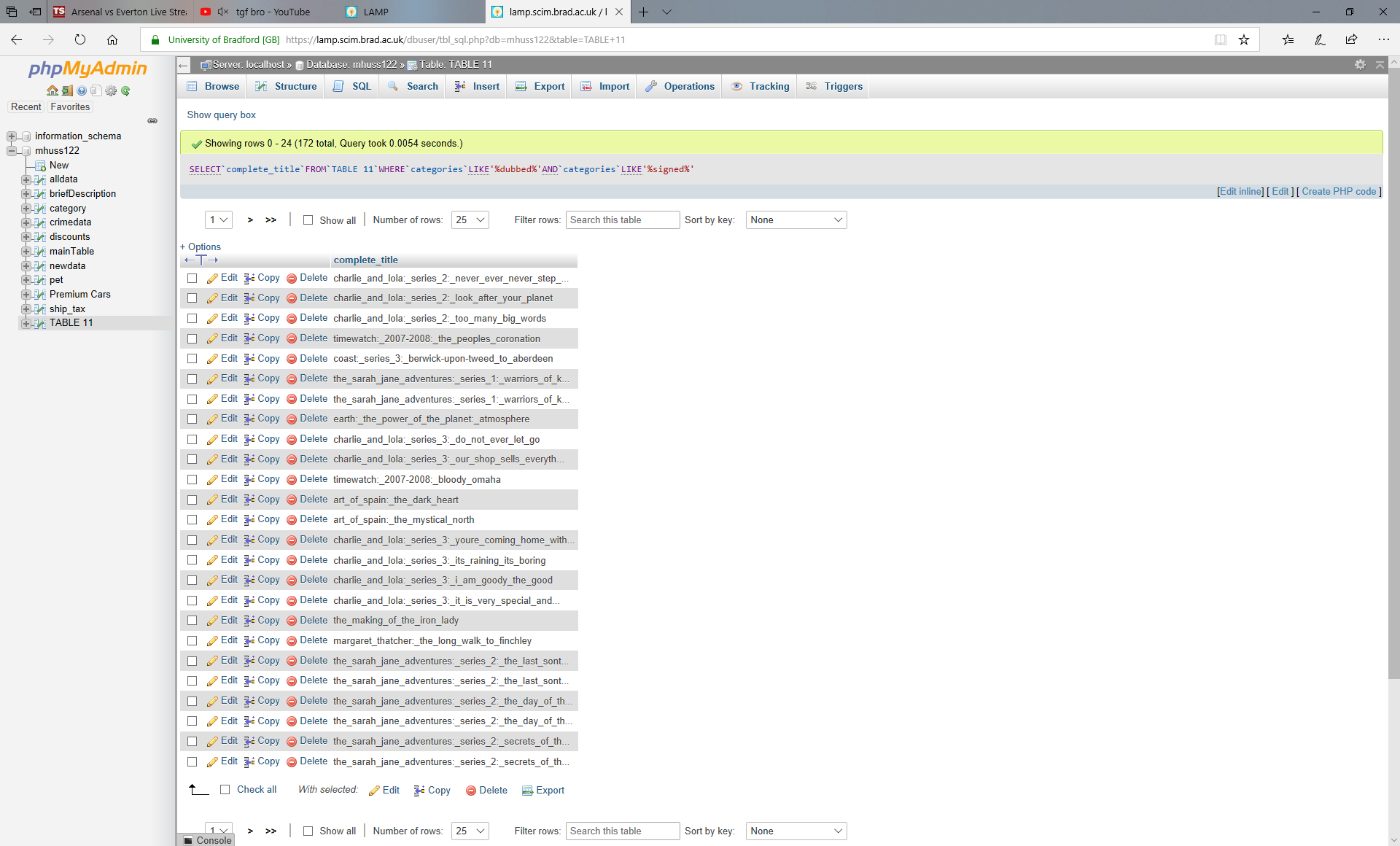




Fifth Query





Sixth Query



Bibliography

1keydata.com. (2020). *Second Normal Form (2NF) - Database Normalization*. [online] Available at: https://www.1keydata.com/database-normalization/second-normal-form-2nf.php [Accessed 6 Jan. 2020].

1keydata.com. (2020). *Third Normal Form (3NF) - Database Normalization*. [online] Available at: https://www.1keydata.com/database-normalization/third-normal-form-3nf.php [Accessed 6 Jan. 2020].

Db-engines.com. (2020). *Key-value Stores - DB-Engines Encyclopedia*. [online] Available at: https://db-engines.com/en/article/Key-value+Stores [Accessed 6 Jan. 2020].

Lucidchart.com. (2020). [online] Available at: https://www.lucidchart.com/pages/er-diagrams [Accessed 6 Jan. 2020].

MongoDB. (2020). *NoSQL Databases Explained*. [online] Available at: https://www.mongodb.com/nosql-explained [Accessed 6 Jan. 2020].

Oracle.com. (2020). *What is a relational database?*. [online] Available at: https://www.oracle.com/uk/database/what-is-a-relational-database/ [Accessed 6 Jan. 2020].