**COMP624121 Querying Data on the Web**

**Lab Work QO**

Start by reading again the general instructions on lab work, in particular the section called **On Running Experiments**.

Study the optimizations that the **sqlite3** DBMS implements, as described in their online documentation:

https://www.sqlite.org/docs.html

and in particular, in this section:

https://www.sqlite.org/optoverview.html

Consider, also, studying this section in more detail:

https://www.sqlite.org/queryplanner-ng.html

But some of it very sqlite-specific and more deeply technical than needed for this lab work.

We expect you to study the documentation, i.e., to find out about facts, notions, ideas, techniques, policies, heuristics yourself. We want to gauge how much you have learned in Week 2 about query optimization and execution. Some topics (like the use of indexes) haven't been covered in detail in the lectures, but do some research. By all means, ask questions in the lab session and in the discussion board, but we are asking you to think and reflect and ask yourself questions and propose some answers and verify whether your proposed answers are indeed correct or not, and so on.

For you to get started, consider how **sqlite3** allows you to control whether or not its join ordering strategy is to be used. Finding this out will give you a head start on how to approach this lab work.

**Task 1**: Write **three pairs** of distinct SQL queries against the **Mondial** database such that

- in each query pair, the two distinct queries are semantically equivalent, i.e., they return the same result set, but
- one query in the pair is optimizable by **sqlite3** whereas the other is not

as a result, we would expect the response times in evaluating the queries in the same pair to be different. Essentially, we want to know by how much, and why.

You must:

- produce the query pair,
- explain which optimization you have in mind for each such pair and
- keep note of the execution times for each query in each pair (see the general instructions on coursework for some tips, hints and advice on how to do that).

You must use **sqlite3** to evaluate your expressions and obtain printed-out results for your submission. Use the **sqlite3** commands to read the input from a file and to redirect the output to a file as you were advised to do for last week's lab work.

**Task 2**: Summarize your investigation with a plot, accompanied by interpretation and comment, that evaluates the benefits of using the **sqlite3** optimizer.

Your submission will be a report containing the pairs, the explanation of the optimizations you have in mind, a log of the run, and the plots with interpretation and comment.

**Marking**

- Each query pair in Task 1 is worth up to 15 marks, for a total of up to 45 marks.
- Task 2 is worth up to 15 marks.
- The whole lab is worth up to 60 marks and contributes up to 6 marks to the final mark for the course unit.

**Software/Data**

You should have everything in place, and know where to find more information from last week's lab work instructions, but some of it is repeated here for ease of access.

**Mondial**

You can find documentation about Mondial in its website. There are local copies of the ER diagram, the relational schema and a diagram of the referential dependencies in

```
/opt/info/courses/COMP62421/data/Mondial/relational/mondial-ER.pdf
/opt/info/courses/COMP62421/data/Mondial/relational/mondial-RS.pdf
/opt/info/courses/COMP62421/data/Mondial/relational/mondial-abh.pdf
```

You need the database to be local to you, so that you have write permissions on it. Data is always under:

```
/opt/info/courses/COMP62421/data
```

For this lab work, you will use the Mondial relational data in

```
/opt/info/courses/COMP62421/data/Mondial/relational
```

You should copy the database and the properties files to you working directory

```
/opt/info/courses/COMP62421/data/Mondial/relational/mondial.db
/opt/info/courses/COMP62421/data/Mondial/relational/mondial.properties
```

**sqlite3**

The mondial.db file above is used in **sqlite3** too. You may benefit from passing the initialization file we prepared (setting parameters for the **sqlite3** command-line interface). This is available in

```
/opt/info/courses/COMP62421/data/Mondial/relational/sqliterc
```

If you have copied it into your working directory, you can invoke **sqlite3** as follows:

```
sqlite3 -init sqliterc mondial.db
```

For **sqlite3**, the documentation in the website is very good, but if you prefer learning from books, you have free access (from an UoM IP address) to this one:

The Definitive Guide to SQLite
Grant Allen, Mike Owens
ISBN: 978-1-4302-3225-4
Apress, 2010
http://link.springer.com/content/pdf/10.1007%2F978-1-4302-3226-1.pdf