Introduction to Database Systems

Individual Homework 4: SQLite Code Base Exploration

1. Introduction

Throughout this semester, you had learned different aspects of a database system, and should thus have basic concept on how a database system works. Now it is time for some real case study. In this homework, you are required to explore a widely-used, open source relational database management systems - SQLite. Answer the questions based on what you learned in the lectures and homework, along with your findings in the SQLite code base.

2. Tasks

2-A. Questions Answering (85 points)

In this part, you need to answer several questions related to the concepts mentioned in the lectures and implementation in SQLite code base. Go to the SQLite Source Repository and download the source code, see the figure below for instruction.

The source codes you need are in the folder "sqlite/src". Once you have the source code of SQLite, read the following questions and answer based on your knowledge on database systems and the SQLite source code:

1. Describe the general (high-level) structure of SQLite code base (20 points)
    a. draw a diagram of SQLite structure
    b. list the components of SQLite and describe their functionality
        i. hint: you can define yourself what is "a component", generally speaking, a component is some part of code (some functions) that together form a particular functionality
        ii. in this question you are not required to do detail explanation on code, just list the related function names and briefly describe them is enough

    c. recall the lecture of Database System Concepts, compare components of SQLite to the components mentioned in the lecture slides, what are the special features of SQLite
2. Describe how SQLite process a SQL statement (25 points)
    a. list the components included in the procedure, describe their roles in the procedure (a brief explanation is enough here)
    b. describe how SQLite optimize the query execution in detail
        i. explain how each term (where, like, between etc.) are optimized, how indexes are used in the optimization
        ii. explain the query planner as detail as possible
3. Describe the interaction between SQLite and OS (20 points)
    a. describe how SQLite stores files in file system (format, pages, headers etc.) in detail
    b. describe how SQLite control file read, write, close etc. (a brief explanation is enough here)
4. Describe the concurrency control of SQLite (20 points)
    a. describe how SQLite handle concurrency control (file locking, journal files etc.) in detail
    b. use examples to explain concurrency control/isolation in SQLite
        i. when is the change done by an operation visible to other operations (which operations)
        ii. when will nondeterministic happen (that is to say, in what condition we can not know what will happen in advance)

2-B. Code Base Exploration (15 points)

In the last part you are required answer questions related to topics mentioned in the lectures. There are still lots of components in the database not included in the questions. In this part, you need to select a component (or several components if you want) you are interested in (and not the component we asked for detail explanation in the last part) then describe the component in detail. Include the following points in your explanation:

- Description of the component you choose (its functionality and role in SQLite)
- Related source code files (in “sqlite/src/”)
- Detail explanation of how it works (code explanation, description in documentation)

3. Writing Guide and Hints

Here are some advices and hints to help you understand how to do this homework and what we expect to see in your report, **please read the following points carefully BEFORE you start to do this homework**:

- **Start from the SQLite documentation**, do not start to read the source code without having any idea on SQLite.
- Try to find external resources online (e.g. stackoverflow, blogs etc.) when you are confused or have no idea where to find answers.
- When we say "in detail", it means you should answer as detail as possible, do not assume we know something unless it is super basic or trivial.
- Remember to explain code when answering the question (except the questions regarding SQLite structure: 1-a, 1-b, 2-a in part 2-A), you can, for example, paste part of code and explain with text.
- Always justify what you write, **DO NOT leave statement without any support**. You can write like "As mentioned in the SQLite documentation about XXX(link)" or paste part of code to justify your answer.
- We encourage you to make connections of your findings in SQLite code base to what you learned in the lecture and previous homework. Try to compare SQLite to database systems mentioned in the lecture (does it match the description in the slides?) and homework (what are the differences/similarities between SQLite and your implementation)

4. Rules

- This is an individual homework, **please do not share your answer with other students and do not copy paste from other students or online resources**. You will get huge penalty if we find you do that.
- We expect at least 5 pages of report in this homework, this is just the minimum number required. You are encouraged to write more details.
- There is no "coding/implementation" part in this homework, **we expect you to do this homework individually and independently**. **Please do not discuss your answer/findings with other students.**
- The total score of this homework is 100 points, 85 points for 2-A and 15 points for 2-B. The points of each questions are marked in section 2.
- You can write this homework in Chinese or English, both are accepted.

5. Discussion

TAs had opened a channel **HW4 討論區** on Microsoft Teams of the course, you can ask

questions about the homework in the channel. TAs will answer questions in the channel as soon as possible.

Discussion rules:

- Do not ask for answer of the homework.
- Check if someone had asked the question you have before asking.
- **Unlike previous homework, in this homework we expect you to complete the homework without discussion with other students, so please do not discuss homework content with other students in the channel**.
- Since we have this discussion channel, do not send email to ask questions about the homework unless the questions are personal and you do not want to ask publicly.

6. Submission

- The deadline of this homework is **07/01 (Wed.) 23:59:59**, **no late submission accepted this time.**
- You only need to submit one file, that is your report in this homework. The report should named as "**HW4_XXXXXXX.pdf**, where XXXXXXX is your student ID.
- We **only accept one pdf file**, wrong format or naming format cause -10 points to your score.
- **Again, no late submission accepted this time**.

If there is anything you are not sure regarding submission, ask in the forum.