2021400042-İrem Nur Yıldırım

Preface

I implemented a B plus tree to speed up the search which is based on primary key search. For key-based searches, we learned that B+ tree structure is advantageous in terms of speed but has limitations in deletion and insertion operations. For our case, since insertion has one-time occurrence (because history does not change) and deletion is not a frequent operation at all, I prioritized search speed by applying a tree structure.

Limitations

Each relation has a folder with the same name, in that folder we have:

- index.py: for index structure of that relation, index structure is stored in index.pkl file
- page0.txt-page9.txt: for pages of the relation, each page has up to 10 records, a page is created when the previous page is full
- Each folder initially has 1 page
- Relations have at most 10 fields, each field name is of 20 characters length
- Maximum length of a relation name is 12
- Maximum length of a field for a relation is 20
- Maximum number of fields a type can have is 10
- Each file can have up to 100 pages, it is checked in index_template.py, so each relation

can have at most 1000 tuples.

Each node in the B plus tree is stored as a tuple of page_number, record_index, so for records, we store in the index.pkl page and offset information.

For each page, we use page number info to get the right page, each page is stored as a separate file in the file system under the folder with the same name as the relation.

For each relation, I store a directory name to reach that relation's data. In that folder, I have the index's pickle file to speed up the search for the tuples.

Disk data of relations is stored in page{number}.dat files with all their records. When I get the page number information from the index.pkl file, I direct the search into that specific dat file.

Disclaimer, Problems

- I could not conduct a search for some edge cases when the primary key is of type string
- Also due to my decision of using a tree, the deletion operation takes time and it took me days to properly implement it

Lessons learned

- Do not use a B plus tree structure unless it is explicitly required.