Project 1 : Implementing a Simple Zone Map

Due: Friday 20th September, 2024 at 23:59

Background

A zone map is a light-weight sparse index that maintains the minimum and maximum values of one (or more) attribute(s) of data that is stored across multiple contiguous blocks or pages. Theses blocks or pages are also referred to as *zones* [1]. A zone map helps in pruning the number of zones qualifying for a range or point query, and thereby, reduces the number of accesses (or I/Os) to slower storage. Typically, a zone map is read once from storage, and if possible, pinned in memory during workload execution. The zone map may be updated in memory as an when required as new data in ingested and moved to the storage. When realizing (point or range) queries, first we probe the in-memory zone maps. The target key (for point query) or target key range (for range queries) is first compared with the maps (i.e., min/max values) of every block or zone, and a block is read from the secondary only if the zone map probe return true. However, if the zone map probe returns false, we can avoid performing unnecessary I/Os to slower storage, and thus, improve the query performance.

Objective

The objective of this project is to implement a simple zone map and evaluate its performance on both point and range queries. You will benchmark the zone map performance for both when (i) the data is stored as a *heap file* and (ii) the data is stored as a *sorted file*.

Language of Implementation. It is strongly suggested that you use C/C++ to do your implementation if you are familiar with the languages. If you are unfamiliar with C/C++, but want to learn it while working on the project, you can find the learning resources here. If neither of the above is suitable for you, we can do the implementation in Java.

Workflow. If you are implementing the project in C/C++, click on this link. If you plan to implement it in Java, click here. The general workflow for the project is as follows.

1. Once you are inside the correct folder (based on which language you choose for implementation), you will see an API that contains a header file with basic function definitions for a zone map. You are free to modify certain components to improve performance.

0.1 Submitting your assignment

Please submit as pdf on Gradescope. Also make sure to submit as group submission. The due date for Project1is Friday 20th September, 2024 at 23:59.