Michał Kukowski

Dyscyplina naukowa: informatyka techniczna i telekomunikacja
Numer albumu: 119010034
Numer PESEL: 95042513016

The database indexing using the modern types of memories

Abstract

The Phd thesis is devoted to indexing databases on various modern types of memory, such as single flash memory chips, solid-state drives (SSDs), and phase-change memory (PCM). Nowadays, an index is not just a simple data structure equipped with basic procedures, but often a large, complex system that takes into account many factors, including memory characteristics. An efficient way of indexing records should not only minimize the time required to execute a query but also extend the lifespan of new memory media. In data warehouses, terabytes of information are written every day. In this context, minimizing memory usage is essential not only from an economic point of view but also from an ecological one.

The first main goal of the paper is to analyze popular algorithms and systems used in database engines for indexing records. The second goal is to develop new indexing methods and implement designed data structures. Several new algorithms and data ordering systems were proposed in the paper, realizing record indexing. Each of the algorithms was analyzed, including theoretical computational complexity analysis and proof of correctness.

Chapter 5 discusses algorithms for indexing records on flash memory. A new indexing algorithm, FA-Tree (Flash Aware Tree), is also presented, which achieved a 20-fold improvement in query handling time over the popular B+ tree and used six times less memory.

Chapter 6 discusses ways of indexing data stored on SSDs. New data indexing algorithms, such as row-based FALSM-Tree (Flash Aware LSM-Tree), are presented, which significantly improve database performance when inserting multiple elements into an index stored on an SSD. During experiments,

FALSM-Tree achieved six times better time and wrote six times fewer bytes than the classical LSM index. The same chapter also presents a completely new column-based CFT (Columned FD-Tree) indexdespite its column-based layout, the CFT index achieved modification time comparable to the time of modification using the original row-based approach with the use of the FD tree, while significantly improving the query search time. Chapter 6 also presents a new partial indexing mechanism called LAM (Lazy Adaptive Merging). This system is fully adapted to the characteristics of SSDs. It not only achieved twice as good results as the AM algorithm when creating the index itself but also executed query modifications to the table on average 15% faster.

Chapter 7 discusses indexing databases using PCM to store data. A new data structure, BB+-Tree (Buffered B+-Tree), and a new partial indexing method adapted to PCM memory - PAM (PCM Adaptive Merging) - are also presented. The results of the experiments showed a huge advantage of the new system over the original Adaptive Merging algorithm. The PAM algorithm created an index even five times faster, while modifying five times less memory, significantly extending the lifespan of this memory medium.

20.06.2023 Michał Krakowski