

*centralized monitoring,
data stream processing, indexing*

Michał WIDERA^{*}, Janusz WRÓBEL^{*}, Adam MATONIA^{*},
Michał JEŻEWSKI^{**}, Krzysztof HOROBA^{*}, Tomasz KUPKA^{*}

INDEXING BIOMEDICAL STREAMS IN DATA MANAGEMENT SYSTEM

We are developing Data Stream Management System that can be applied in medical monitoring system. One of the problems is how to support efficient data access methods. This paper considers indexing issues in data management system. The indexing is a commonly used method of data access acceleration. Classical and new streaming methods of indexing have been presented. The method was developed in data stream management system and first applied in a centralized neonatal monitoring system.

1. INTRODUCTION

The most important function of the medical monitoring system is the processing and recording of biomedical signals. Such signals contain information about chosen physiological quantity in time. Design of such system requires unconventional solutions – i.e. application of requirements of real time system. We are developing Data Stream Management System that can be applied in medical monitoring system and fulfil these real time requirements. One of the research problems is to support efficient data access method in such a system.

Database management system (DBMS) applied in biomedical monitoring system [1] is expected to ensure a fast data access method. Commonly used method of data access acceleration in DBMS is the indexing [2,3]. Index is usually created on table by user on demand. Created index is updated simultaneously with data set. Generally, issued query in such system is limited to a few attributes. This type of query can be supported by specific type of index that is stored in separate structure containing attributes with pointers to physical record position. Search algorithms can selectively scan the database by such index. Attributes that create this structure are called index key or if used in index context – a key.

In relational DBMS, the delete, update and insert operations are well defined. In a case of data streams set only the data append operation is used. This causes appearance of other method of index creation. The main requirement is finding process acceleration. Due to the data stream specificity, the issue of data updates and deletions is not considered.

^{*} Institute of Medical Technology and Equipment, 118 Roosevelt St, 41-800 Zabrze, POLAND,
tel./fax: (32) 2716013/2765608, e-mail: michal@widera.int.pl

^{**} Student of Informatics Department, Silesian University of Technology, Gliwice, Poland

2. METHODS OF DATA INDEXING

There are various structures able to be used as an index. In relational database management system we can find the following main examples [2,3]:

- Simple index on sorted files. This is the simplest index structure. Each record of index file contains two elements – key and pointer. If each record of data file has corresponding record in index file then this index file is called dense. Otherwise this is the sparse index. If dense index is too big to fit into memory, then sparse index is created as first layer. The radical solution is the B-tree structure of index file.
- Indexes on unsorted files. Considering ordered set of pairs we cannot assume that the both elements of the pair are in order. Therefore, considering various queries we need the set of indexes. Usually the first index of data set is the primary dense index. The second index based on other attributes refers primary index. The second, helper index is always dense.
- B-tree. B-tree index has a tree-like structure ordering data blocks into tree structure. The tree is balanced if length of all paths is the same. This is the one of the most used technique of search process speed-up in database management systems.
- Hash index. This is the one of the most effective methods of data access management. Each pointer (address) of disk block containing a desired record is computed using a function (so called hash function) and the search key. Hash function maps the set of all search keys to the set of all records or blocks. The main problem is proper choosing of hash function for each query expression.
- Bitmap index. Bitmap index provides pointers to the rows in a table that contain a given key value. Each bit in the bitmap corresponds to a possible record. A mapping function converts the bit position to block pointer, so that the bitmap index provides the same functionality as a regular index. Bitmap indexes are widely used in data warehousing environments.

2.1. DATA STREAM INDEXING

Indexing method analysis in environment of data stream processing needs to take into consideration different requirements [4,5,6]. Relations are indexed by keys; in case of sequences, an order is imposed. But timeline domain indexing in searching task is potentially helpful.

Data stream indexing is mainly considered in time series and streams of knowledge discovery research [7]. There are few papers concerning stream indexing. In literature we can find some papers about sliding windows indexing over data streams [4,5,6,8]. The problem of indexing sliding windows, stored on disk and updated on-line, has been presented by Shivakumar and Garcia-Molina [8]. The main idea is to split the index into several parts so that deletions and insertions do not affect the entire index. These algorithms were called Wave Indices. Maintaining clustered order on disk as well as temporarily storing parts of the index in main memory is also discussed.

2.2. QUERY PLAN

Database management system for signal processing needs realise signal processing tasks expressed in formal query language [2,3,9]. An algorithm that creates query answer in database management system is called query plan. Methods of query plan creation and proper index usage in relational system are well presented in many papers [2,3]. But in case of queries for data stream management system there are many open problems. First, there are considered continuous query plans [10]. Second, the methods of such query optimization are not well defined and discovered. Additionally, deterministic signal processing needs alternative research.

There is no explicit method of suggested index use in official relational query language standard (SQL). Automatic index selection is used, based on developed management system rules. Some systems are using hints as suggestions for query plan. Hints enable the turn off in automatic index selection or choose better solution from available index set.

3. DEVELOPED METHOD OF BIOMEDICAL STREAM INDEXING

In the Institute of Medical Technology and Equipment an indexing method of streams that contain biomedical signals is developed. This method is implemented with a view of data stream management system for biomedical monitoring system needs.

Biomedical data stream model describe a bag of elements $(\{a_n\}, \Delta)$, where the first element is tuple sequence and the second is a rational number that determines time interval between the consecutive elements of the sequence [9]. In time domain, tuple position relies only on direct constant delay Δ and tuple position n . Additional index structure does not accelerate data finding process by using time domain.

However, in data management system we can define queries that concern unexpected, sudden events. Simple access to this data combines with full data scan necessity. Such events are recognized and recorded during on-line monitoring session.

We assume that the data access speed-up can be made by index usage. The aim of research is to minimize disk access count. Data access acceleration is performed by: maintaining all indexes in memory, lower size of index schema and fast and simple on-line data compression.

Conventional data compression solutions bring significant and unacceptable delay in data access. But in case of simple algorithms LZW [11] or Huffman [12] application, it is possible to present more effective solution. Such occurs in case of a data stream management system for biomedical monitoring system. Created index contains information about position in stream of sporadically issued events. Index as regards contents is similar to relational bitmap index [2,3]. Such data are able to compress well, and complexity of compression algorithm is linear.

Within the confines of data stream research in ITAM, the general structure of index stream was presented. Data structure $\langle \text{key}, \text{position} \rangle$ are stored in additional, separate data stream. Index stream is created on demand by `create index <key> on <stream_name>` command. The content of index is used in query plan automatically. Many continuous queries are executed simultaneously in system. Therefore the index can be used by several queries at the same time. Keeping entirely such index in main memory significantly enables speed-up of data searching process.

We have considered two types of index streams. First, (Figure 1 – Index A) was applied as simple uncompressed index but without null values (i.e. condensed). Each index tag points at data stream tuple. Such solution requires complicated method of index maintenance, especially when archive data are discarded. The other method was developed based on compressed index structure (Figure 1 – Index B). The compression method is transparent. Therefore all developed previously [1,9,13] methods of stream management are applicable. Additional delay of index maintenance is comparable in these two methods. But the algorithms of Index B creation and maintenance are simpler.

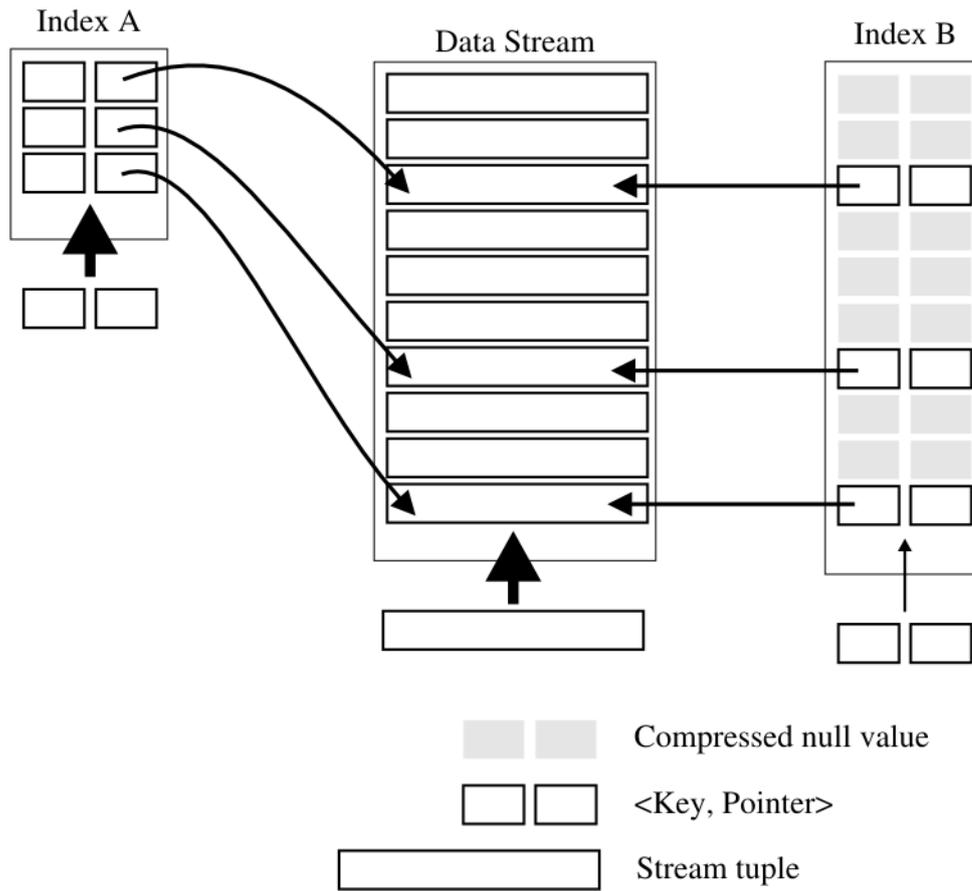


Fig. 1. Indexing of data streams

3.1. MANAGEMENT OF STREAM AND INDEX SIZE

The potential size of data stream and index size is unbounded. Data maintenance in main memory or disk drive leads to resource overload. Therefore data management requires application of sophisticated algorithms that decide what part of index can be stored or discarded. There is a need of discarding or archiving oldest and less significant parts of data. Therefore we have developed method of cyclic stream structure (CSS). This method was developed in data stream management system and first used in a centralized newborn monitoring system [14].

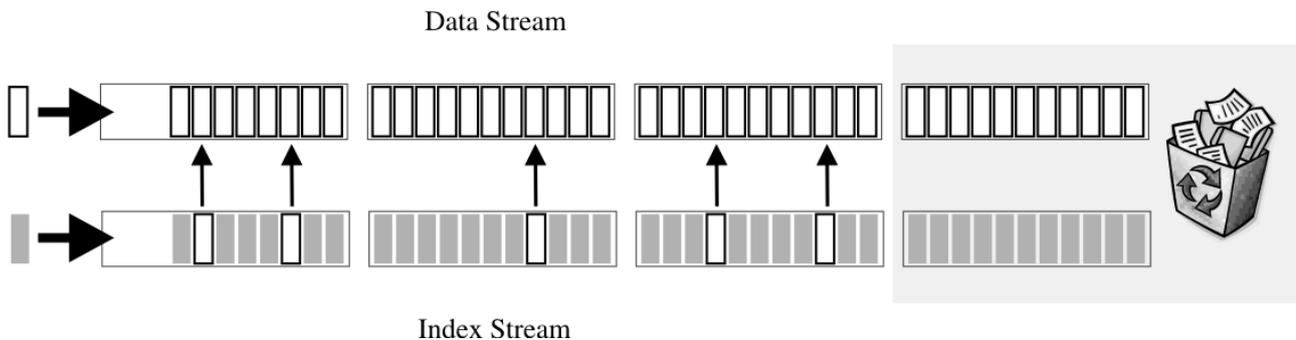


Fig. 2. Discarding archive data and index stream

CSS is based on simple data stream structure and interface. Data are stored in CSS and discarded when they reach border time. Each CSS contains a set of data files. From the user view, this is single, flat stream. In fact, CSS manages a set of data files, and destroys oldest ones

maintaining constant number of data files. This data files can be kept in memory or stored on disk. This structure supports infinite stream model, and enables maintaining in main memory the on-line compressed index structures. The process of data and index stream management by CSS is presented on Figure 2. Each CSS index is stored in memory.

4. CONCLUSIONS

Application of central monitoring system significantly simplifies the structure of the monitoring systems [14,15]. The DSMS takes over the responsibility for several crucial functions and processes that have been developed so far within the monitoring workstation. Most of these functions may be initially incorporated into the data management system in order to reduce the workload of software developers and designers of monitoring systems. One of the problems is to support efficient data access method in such system.

Commonly used method of data access acceleration in DBMS is indexing. Index is created on database table by user on demand and updated simultaneously with data set. We have assumed that the data access speed-up can be realized by index usage. We have considered two types of index streams – compressed and condensed. Finally, after research we have decided to use compressed one. Data access acceleration was accomplished by: maintaining all indexes in memory, lower size of index schema as well as fast and simple on-line data compression.

The potential size of data stream and index size is unbounded. Data maintenance in main memory or disk drive leads to resource overload. Therefore we have developed method of CSS. This method was developed in the data stream management system and first used in a centralized newborn monitoring system. Compressed on-line index data under control of CSS maintained in memory was fast enough method to support all real-time requirements for developed database management system and centralized neonatal monitoring system.

BIBLIOGRAPHY

- [1] WIDERA M, WRÓBEL J., WIDERA A., A.GACEK: System zarządzania danymi dla potrzeb medycznych systemów monitorujących, Współczesne Problemy Systemów Czasu Rzeczywistego, WNT, pp. 448-458, 2004
- [2] GARCIA-MOLINA H., ULLMAN J.D., WIDOM J.: Implementacja systemów baz danych, WNT, 2003
- [3] ULLMAN J.D., WIDOM J. Podstawowy wykład z systemów baz danych, WNT, 2003
- [4] GOLAB L., M.T. OZSU: Issues in data stream management, ACM SIGMOD Record vol.32, pp.5-14, 2003
- [5] GOLAB L., PRAHLADKA P., OZSU M.T. Indexing the Results of Sliding Window Queries. University of Waterloo Technical Report CS-2005-10, 2005
- [6] GOLAB L., GARG S., OZSU M.T. On Indexing Sliding Windows over On-Line Data Streams. In Proc. 9th Int. Conf. on Extending Database Technology (EDBT), pp. 712-729, 2004
- [7] KEOGH E., CHAKRABARTI K., PAZZANI M., MEHROTRA S.: Locally adaptive dimensionality reduction for indexing large time series databases Proceedings of the 2001 ACM SIGMOD international conference on Management of data ACM Press, pp: 151-162, 2001
- [8] SHIVAKUMAR N., GARCIA-MOLINA H. Wave-indices: indexing evolving databases. Proc. ACM SIG-MOD Int. Conf. on Management of Data, pp. 381-392, 1997
- [9] WIDERA M., JEŻEWSKI J., WINIARCZYK R., WRÓBEL J., HOROBA K., GACEK A. : Data stream processing in fetal monitoring system: I. Algebra and query language, Journal of Medical Informatics & Technologies, vol.5, pp.83-90, 2003
- [10] ARASU A, WIDOM J.: A Denotation Semantics for Continuous Queries over Streams and Relations Sigmod Record vol. 33, Nr. 3 2004, pp. 6-11, 2004
- [11] NELSON M. LZW Data Compression, Dr. Dobb's Journal, 1989. (US Patent. 4558302)
- [12] HUFFMAN D.A. A method for the construction of minimum-redundancy codes, Proc. of the I.R.E., pp. 1098-1102, 1952

- [13] WIDERA M., WRÓBEL J., WIDERA A., MATONIA A.: A method of the ensuring data integrity in the data stream management system, *Journal of Medical Informatics & Technologies* vol.8, pp.141-148, 2004
- [14] WIDERA M., WRÓBEL J., JEŻEWSKI J., HOROBA K., MATONIA A., KUPKA T. System For Centralized Newborns Monitoring, , *Journal of Medical Informatics & Technologies*, 2005 /in press/
- [15] WROBEL J., JEZEWSKI J., HOROBA K., GACEK A., GRACZYK S. (1998): System for centralized fetal monitoring, *Proc. of the 2nd IMACS CESA '98*, pp.772-775, 1998

This study is financed by the State Committee for Scientific Research resources in 2003-2005 years as a research project KBN Grant No. 4 T11F 002 25. Adam Matonia is a fellowship holder of the Foundation for Polish Science.