

Abstract

Many devices such as sensors, measuring stations or even servers produce enormous amounts of data in the form of time series, which are then processed and stored for later analysis. A huge role in this process takes data processing on graphics cards in order to accelerate calculations. To efficiently use the GPGPU (General Purpose Graphical Processing Unit) a number of solutions has been presented, that use the GPU as a coprocessor in a databases. There were also attempts to create a GPU-side databases. It has been known that data compression plays here the crucial role. Time series are special kind of data, for which choosing the right compression according to the characteristics of the data series is essential.

This paper is a research and presents a new approach to compression of time series on the side of the GPU, using a planner to keep building the compression scheme based on statistics of incoming data, in the incremental manner. The solution compresses columnar data using lightweight and lossless compressions in CUDA technology. The aim of the work is to create an optimizer with high performance in terms of obtained compression ratio for data of variable characteristics.

The beginning of the document is a description of the problem, along with an analysis of existing solutions and research, under the direction of compression using SIMD (Single Instruction Multiple Data) architectures. Further it describes adopted technology developed by NVIDIA and implemented algorithms of light compression for GPU. The following sections describe the implementation of the optimizer algorithm, along with created environment and a program for parallel compression of data columns. At the end are the results of experiments demonstrating the usefulness of such a solution and a description of further work that will be conducted in the topic. This method is applicable in all types of columnar data warehouses, which will be the subject further research.