**FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE**

Supervisor's statement of a final thesis

Supervisor: Ing. Tomáš Oberhuber, Ph.D.
Student: Bc. Jan Groschaft
Thesis title: Parallel algorithms for data hashing on GPUs
Branch / specialization: Computer Science
Created on: 30 May 2022

Evaluation criteria

1. Fulfillment of the assignment

[1] assignment fulfilled
[2] assignment fulfilled with minor objections
▸ **[3] assignment fulfilled with major objections**
[4] assignment not fulfilled

Unfortunately, the third point of the assignment was not fully resolved. It was also the most important part of the assignment. The student was supposed to propose and test several approaches of "dynamically growing hash table". Instead, the thesis describes only one approach which is allocating of empty slots in the hash table. These empty slots are used when additional data are inserted into the table. However, once these empty slots are filled completely, one cannot add new data anymore. Though the implementation of this method was not trivial and the author had to deal with associated difficulties, the capacity of the hash table is still limited and must be known a priori. The goal of this thesis was supposed to be design of methods which would allow the hash table to grow. During the work on this thesis we were discussing several such approaches with the student and I am quite sure that the student has implemented some of them. Therefore I am surprised that they are not mentioned in the thesis even if they would not give the best performance. In my opinion, the third point of the assignment was fulfilled by 30%.

2. Main written part   95 / 100 (A)

The thesis is written very well. It is well structured, the English is on a very good level. I discussed all errors and inaccuracies that I found in the text with the student before submission and as far as I know, the student fixed all of them. The text is rich in references. The only thing that could be improved typographically, in my opinion, are the code listings.

3. Non-written part, attachments 95 /100 (A)

The author has implemented new parallel algorithms based in HashGraph with a number of unit tests checking the proper functionality. He has used adequate technology which means C++ compilers lika g++ and nvcc for CUDA, existing algorithms in TNL library and Gtest library for unit tests. Unfortunately, he does not specify what kind of data where used for testing of the hashing algorithms. The author only says that he used 4-byte integers (page 41) but it is not clear whether the data are synthetic or real. If real, what is the origin?

4. Evaluation of results, publication outputs and awards 70 /100 (C)

The author got familiar with the state-of-the-art hashing algorithms on GPUs and he compared probably the three most popular of them - HashGraph, WarpCore and SlabHash. I see it as a very valuable comparison of these methods. If the comparison was even more comprehensive, it might be published probably as a scientific paper. I think that the algorithms implemented by the authors could be well used by other users. The implementation made by the author seems to be more robust than the original implementation of the HashGraph. Unfortunately, the hash table cannot grow larger when all slots are used. In this case, the user just gets an error message and the algorithm fails. This can be limiting in some applications when one cannot estimate the number of elements that need to be hashed. Dealing with this was one of the main goals of this thesis.

5. Activity of the student

    [1] excellent activity
▸ **[2] very good activity**
    [3] average activity
    [4] weaker, but still sufficient activity
    [5] insufficient activity

Most of the time, we had regular consultations with the student once a week. Unfortunately there were two longer breaks when the student was not available for consultations. The second was just before finishing the thesis and my guess is that the student just did not have enough time to describe the results he obtained with dynamic resizing of the hash table.

6. Self-reliance of the student

    [1] excellent self-reliance
▸ **[2] very good self-reliance**
    [3] average self-reliance
    [4] weaker, but still sufficient self-reliance
    [5] insufficient self-reliance

The student was able to work independently and to come up with his own ideas. Our consultations had a form of discussions and ideas exchange between us rather than just the supervisor telling the student what to do.

The overall evaluation

75 /100 (C)

The author has done a lot of valuable work and he has written a very good thesis. If he has described even results obtained with the real dynamical hash table resizing this would have been a perfect thesis. Unfortunately, real hash table resizing is not mentioned in the thesis and therefore the third point of the assignment cannot be considered as successfully fulfilled.

Instructions

Fulfillment of the assignment

Assess whether the submitted FT defines the objectives sufficiently and in line with the assignment; whether the objectives are formulated correctly and fulfilled sufficiently. In the comment, specify the points of the assignment that have not been met, assess the severity, impact, and, if appropriate, also the cause of the deficiencies. If the assignment differs substantially from the standards for the FT or if the student has developed the FT beyond the assignment, describe the way it got reflected on the quality of the assignment's fulfilment and the way it affected your final evaluation.

Main written part

Evaluate whether the extent of the FT is adequate to its content and scope: are all the parts of the FT contentful and necessary? Next, consider whether the submitted FT is actually correct – are there factual errors or inaccuracies?

Evaluate the logical structure of the FT, the thematic flow between chapters and whether the text is comprehensible to the reader. Assess whether the formal notations in the FT are used correctly. Assess the typographic and language aspects of the FT, follow the Dean's Directive No. 52/2021, Art. 3.

Evaluate whether the relevant sources are properly used, quoted and cited. Verify that all quotes are properly distinguished from the results achieved in the FT, thus, that the citation ethics has not been violated and that the citations are complete and in accordance with citation practices and standards. Finally, evaluate whether the software and other copyrighted works have been used in accordance with their license terms.

Non-written part, attachments

Depending on the nature of the FT, comment on the non-written part of the thesis. For example: SW work – the overall quality of the program. Is the technology used (from the development to deployment) suitable and adequate? HW – functional sample. Evaluate the technology and tools used. Research and experimental work – repeatability of the experiment.

Evaluation of results, publication outputs and awards

Depending on the nature of the thesis, estimate whether the thesis results could be deployed in practice; alternatively, evaluate whether the results of the FT extend the already published/known results or whether they bring in completely new findings.

Activity of the student

From your experience with the course of the work on the thesis and its outcome, review the student's activity while working on the thesis, his/her punctuality when meeting the deadlines and whether he/she consulted you as he/she went along and also, whether he/she was well prepared for these consultations.

Self-reliance of the student

From your experience with the course of the work on the thesis and its outcome, assess the student's ability to develop independent creative work.

The overall evaluation

Summarize which of the aspects of the FT affected your grading process the most. The overall grade does not need to be an arithmetic mean (or other value) calculated from the evaluation in the previous criteria. Generally, a well-fulfilled assignment is assessed by grade A.