UNIVERSITY OF SOUTH FLORIDA

Defense of a Doctoral Dissertation

Computing Group-By and Aggregate in Massively Parallel Systems

by

Chengcheng Mou

For the Ph.D. degree in Computer Science and Engineering

The semiconductor industry has mainly exploited two routes for designing microprocessors. The multi-core route aims to speed up the performance of latency-oriented processing. In contrast, the many-thread route concentrates on throughput-oriented improvement of parallel processing. Many-thread microprocessors, such as Graphics Processing Units (GPUs), are leading the computing capability for this past a decade. According to the current hardware market, at the similar price range, the ratio of peak computing power between multi-core CPUs and many-thread GPUs is up to 15X. This large performance gap on data processing has motivated many practitioners in database community to exploit computation-intensive parts on GPU for query execution. Group-By and aggregate operations are very often used together to summarize data, such that data scientists and domain experts could quickly gain analytical insights over possibly massive amounts of data. They play fundamental and critical roles in data visualization community and contribute large part of the user experience in the interactive visualization analysis. In this research, we investigate the low-level computing features of GPUs, and we exhibit in-depth study of design, implementation, and optimization of Group-By/Aggregate algorithms on GPUs. We primarily focus on the design and implementation of hash-based Group-By/Aggregate algorithms. We then introduce an adaptive and dynamic radix-hash algorithm, which is insensitive to input cardinality (number of distinct groups). We present a performance model which guides us to pick an optimal set of bits to proceed radix-hash in each pass, such that the overall group-by operation could achieve maximum throughput. We also reproduce the-start-of-art hash-based implementation on both modern CPUs and GPUs. Our experiments verify that our adaptive and dynamic algorithm chooses the optimal solution and deliver highest throughput.

Examining Committee

Xiaoming Liu, Ph.D., Chairperson
Yicheng Tu, Ph.D., Major Professor
Srinivas Katkoori, Ph.D.
Paul Rosen, Ph.D.
Yu Zhang, Ph.D.
Feng Cheng, Ph.D.

Tuesday, March 29, 2022
11:00 AM
Online (MS Teams)
Please email for more information
[chengcheng@usf.edu]
THE PUBLIC IS INVITED

Publications

1) **C Mou**, S Chen, YC Tu, “A comparative study of dual-tree algorithms for computing spatial distance histograms”, The Computer Journal 62 (1), 42-62, 2019
2) **C Mou**, S Chen, YC Tu, “A comparative study of dual-tree algorithm implementations for computing 2-body statistics in spatial data”, 2016 IEEE International Conference on Big Data (Big Data), 2676-2685
3) H Li, **C Mou**, N Pitaksirianan, R Rui, Z Nouri-Lewis, M Eslami, R Sheng, S Lei, J Wang, YC Tu, “CheetahDB: A System for High-Throughput Database Processing on GPUs", Graphic Technology Conference 2020, Poster, San Jose, CA, USA., Mar 23-26, 2020
4) C Li, **C Mou**, MD Swartz, B Yu, Y Bai, Y Tu, X Liu, “dbMTS: a comprehensive database of putative human microRNA target site SNVs and their functional predictions” Human mutation 41 (6), 1123-1130, 2020
5) X Liu, C Li, **C Mou**, Y Dong, Y Tu, “dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs” Genome medicine 12 (1), 1-8, 2020

Robert Bishop, Ph.D.
Dean, College of Engineering

Dwayne Smith, Ph.D.
Dean, Office of Graduate Studies

Disability Accommodations:

If you require a reasonable accommodation to participate, please contact the
Office of Diversity & Equal Opportunity at 813-974-4373 at least five (5) working days prior to the event.