Galactica - Accelerated Queries Processing

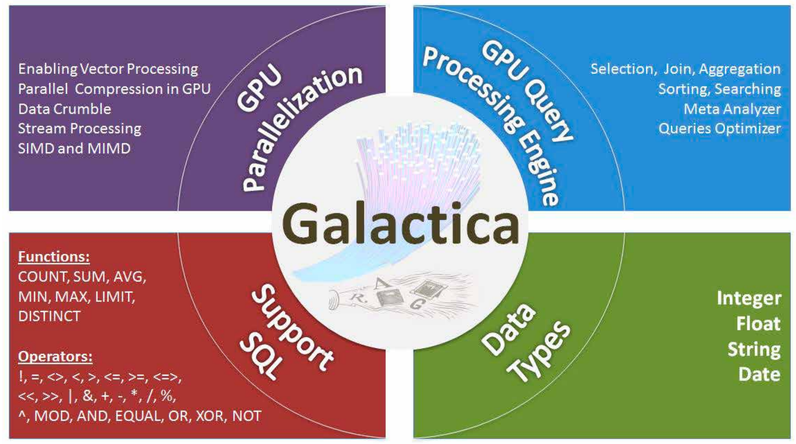
Malaysian Institute of Microelectronics System (MIMOS Berhad), Malaysia.

Motivation

Traditional databases ***aren't equipped*** to efficiently compute massive volume of data, and there are related costs for ***storing***, ***moving***, ***computing*** and ***developing*** a fast database operation system. These areas observe quick growth of large data which needs either timely analytics or batched processing. Thus, the challenge is to analyze and mine these ***big data*** in order to effectively exploit the information to improve efficiency and quality of service for consumers and producers alike. With size, it comes ***performance issues***.

Galactica is an emerging GPU database engine that accelerating analytical computation with parallelizing queries processing and exploiting NVIDIA high performance Tesla GPUs.

Introduction



Development Status: 20% completed on standard SQL query support.

1. **Preprocessor**
   - Profiling hardware and data
   - Crumbling the input data into segments
   - Processing data into a disk storage
2. **GPU Query Processing Engine**
   - Providing the analysis and optimization for the input multi queries
   - Parsing, transforming and executing the generated queries into a parallel instruction sets
3. **Load Balance Controller**
   - Managing the streams which is required APIs based on the instruction sets for parallel execution
4. **Daemon Server**
   - Communicating Galactica to the outside world application through network
   - Queuing and managing queries execution

Evaluation 1: Vs PostgreSQL

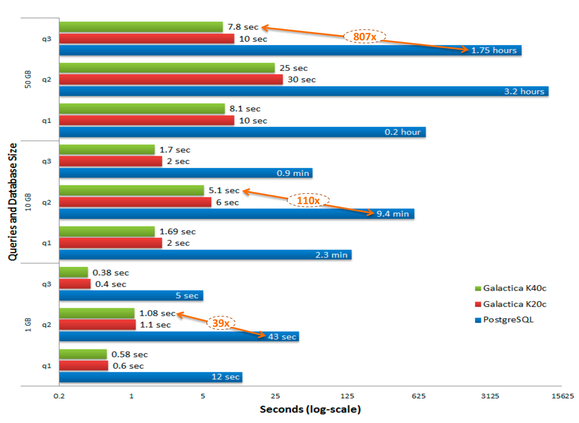
Test Data

- Using a standard and popular TPC-H benchmark
- Performing 3 different sets of query in 1GB, 10GB and 50 GB sizes.
- Comparing Galactica against one of the standard and powerful open source object-relational database system (PostgreSQL 9.3)

Environment

- **Dual** Intel(R) Xeon(R) CPU X5680 @ 3.33GHz with **6 cores processors**
- 22 GB RAM, WD HDD 1TB
- NVIDIA Tesla **K20c** and **K40c**
- Microsoft Windows 7 (64 Bits)
- CUDA 5.5 and 6.0

RESULT 1



LESSONS LEARNED

- Galactica shows the speedup of over two orders of magnitude over the same operation that had done in PostgreSQL on a multicore machine
- Galactica gained better speedup by having larger data for parallel processing
- Galactica performs on commodity data operation, also, it is an effective co-processor for performing database and query operations.
- Low cost GPU workstation has a competitive result against high end server with a distributed Hadoop system
- Parallel processing massive volume of string objects are given challenges

Future Work

- Advancing the parallel compression engine by optimizing data transfer time
- Supporting more standard SQL queries
- Exploring distributed queries execution across GPU cluster

Evaluation 2: Vs Hadoop

Test Data

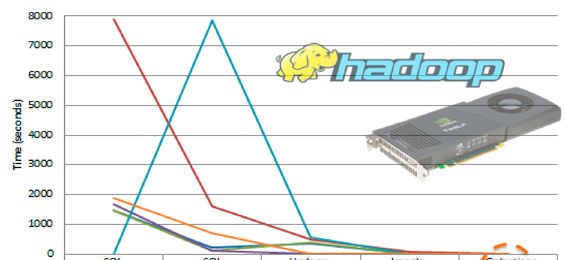
- Using Health Care Database with 32GB data size
- Performing 7 different sets of query
- Comparing Galactica against PostgreSQL 9.3, Hadoop Hive and Impala Hive

Environment

Hadoop: HP DL380p G8 Server with 48 cores and 96GB RAM, 7 VM (12 cores for master node and 6 cores each for worker nodes)

Galactica: Intel® Xeon® E5630 @ 2.53GHz with Quad Core Processor , NVIDIA Tesla K20c, 12 GB RAM, WD HDD 1 TB, Windows Server 2008 R2 Enterprise 64-bit

RESULT 2



	SQL (8GB/4Core)	SQL (96GB/48Core)	Hadoop-Hive	Impala-Hive	Galactica
Q1	1466.7	218.7	347.6	3.7	0.3
Q2	7901	1612	505	64.2	N/A
Q3	1464.7	103.6	383.5	3.5	3
Q4	1688.7	102.7	N/A	2.9	1.6
Q5	Failed	7878	557.3	10.1	N/A
Q6	1893	704	N/A	3.7	6.3

N/A – Not Available – Query not supported.

ACKNOWLEDGEMENT

This research was done under Joint Lab of "NVIDIA - HP - MIMOS GPU R&D and Solution Center". This is the first GPU solution center in South East Asia established in October 2012. Funding for the work came from MOSTI, Malaysia. The authors would like to thank Prof. Simon See and Pradeep Gupta from NVIDIA for the supports.

CONTACT INFORMATION

Author : Yong Keh Kok
: Chua Meng Wei
: Dr. Ettikan Kandasamy Karuppiah
Company : MIMOS Berhad (Malaysia)
Accelerative Technology Lab - ATL

Email : kk.yong@mimos.my
Website : http://gpu.mimos.my

Innovation for Life™

1. P. K. Chong, Ettikan K. Karuppiah and K.K. Yong, "A Multi-GPU Framework for In-Memory Text Data Analytics", The 27th IEEE International Conference on Advanced Information Networking and Applications (AINA), Barcelona, Spain, Mar. 2013.
2. Ettikan K. K., K.K. Yong and Keeratpal Singh., "A Middleware Framework for Programmable Multi-GPU based BigData Applications", Chapter of GPU Computing and Applications, Springer. (Expected publishing June 2014)
3. K.K. Yong and Ettikan K. K., "Hash Match on GPU", Open Systems (ICOS), 2013 IEEE Conference on , Dec. 2013
4. W.M. Ong, V. M. Baskaran, P.K. Chong, K.K. Yong and Ettikan K. K., "A Parallel Bloom Filter String Searching Algorithm on a Many-core Processor", Open Systems (ICOS), 2013 IEEE Conference on , Dec. 2013