Exploration of Data Warehousing and Graph Applications with GPUs
Ifrah Saeed, Se Hoon Shon, Haicheng Wu, Jeffrey Young, Sudhakar Yalamanchili
School of Electrical and Computer Engineering, Georgia Institute of Technology

Application Space: Data Warehousing

- Current applications process 1 to 50 TBs of data [1]
- Not a traditional domain for GPU acceleration, but parallel queries experience good speedup on GPUs [2]

Base Primitives and Data Structures

- Primitives
  - Relational Algebra: PROJECT, PRODUCT, SELECT, JOIN
  - Math: Arithmetic, String, Datetime

OpenCL Backend for Red Fox

- Port of Red Fox to new accelerator platforms using OpenCL
- Initial performance is evaluated using 16-bit key-value store and TPC-H micro-benchmarks

TPC-H Micro-benchmarks

- Micro-benchmarks represent frequently occurring patterns of operators from the 22 queries in TPC-H [3].

Graph Applications – BFS (Ongoing Work)

- Breadth First Search (BFS) is important for large-scale analysis of social networks and linked datasets such as Wikipedia
- Limited PCIe bandwidth makes it difficult to map this algorithm to clusters of accelerators
- Exchange of edge lists requires low-latency transfer
- Current work involves the design of an efficient partitioning scheme that maps across a cluster
  - Building on single-node work done by Merrill [4] as well as CPU-based Graph500 implementations
  - Optimized OpenCL and CUDA versions will allow for high performance with different accelerators

References