Abstract

Generating synthetic data set which is realistic as well as sufficiently large has been a cumbersome task for researchers in the past. Several models have been proposed previously, all adopting heterogeneous approaches, in this work the emphasis is on speeding up the compute time of the data set distribution. Here, Uniform, Poisson and Zipf distributions have been studied and approaches with parallel computation model have been proposed. The models have been verified for speedup using CUDA based implementation on NVIDIA Quadro 2000 GPU. A speed up in the range of 2x to 6x was observed for various range of data sets.

Refer
Generating Multi-million Data Set using GPGPU Accelerated Models

References

Index Terms
Computer Science Distributed Computing

Keywords
Data Set Generation Synthetic Dataset Zipf Poisson Uniform Distribution Gpu Cuda.