

Text Analysis on GPU Enhanced Desktop

Computational Sciences & Engineering Division



Figure 1: Expensive Cluster Computer vs GPU Enhanced Desktop

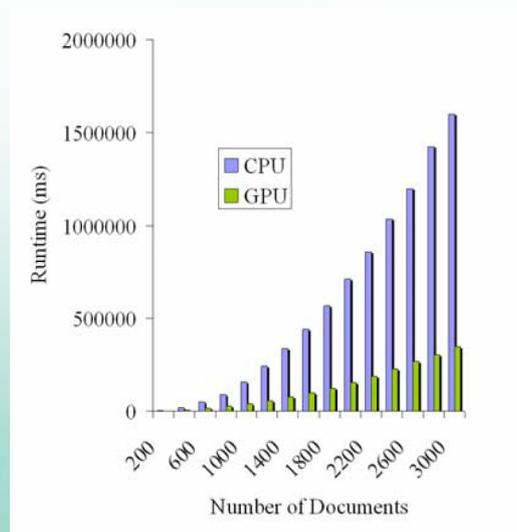


Figure 2: Document Clustering Runtime, CPU vs. GPU

Problem Statement:

- In the last decade, an explosion in the amount of available digital text resources has occurred. It is estimated that the Internet contains hundreds of terabytes of text data, a sizable amount of which is in an unstructured format. We will soon reach a point where terabyte-scale text corpora are routinely used on personal desktops for the purposes of research and decision making. However, most current text processing algorithms work well only on small corpora and are difficult to be scaled to the terabyte level on desktops because of the lack of enough computing power. Even running some simple text analysis tasks can take days or weeks of computer time to process a relatively large collection of data.

Technical Approach:

- In the last few years, the Graphics Processing Unit (GPU) has received much attention for its ability to solve highly parallel and semi-parallel problems much faster than the traditional sequential processor (CPU). New algorithms and methods are being developed that will allow utilizing a GPU as a general purpose co-processor for intensive computing in terms of text analysis (term/feature extraction, dimension reduction, similarity calculation, and clustering).

Benefit:

- Currently, GPUs are mainly assembled as graphic cards and utilized only for computing graphics in desktop computers. They are relatively non-expensive (\$300 - \$600 for a high-end GPU graphic card) and can be found in almost every PC. The use of GPUs as acceleration devices is a much more cost effective choice for desktop users in terms of speeding up the desktop rather than using a high performance supercomputer. GPUs can perform general computing algorithms many times faster than conventional CPUs. A GPU enhanced desktop system will have a huge commercial market in the massive scale text mining field.

Point of Contact:

1-12

Xiaohui Cui
(865) 576-9554
cuix@ornl.gov

