Final Report for Period: 09/2011 - 08/2012
Submitted on: 08/30/2012
Principal Investigator: Shi, Xuan .
Award ID: 1048162
Organization: Georgia Tech Research Corp
Submitted By:
Shi, Xuan - Principal Investigator

Title:
CiC: EAGER: Collaborative: GIS Vector Data Overlay Processing on Azure Platform

Project Participants

Senior Personnel

Name: Shi, Xuan
Worked for more than 160 Hours: Yes
Contribution to Project:

Post-doc

Graduate Student

Undergraduate Student

Technician, Programmer

Other Participant

Research Experience for Undergraduates

Organizational Partners

Georgia State University
This is a collaborative research project led by Dr. Sushil Prasad at Georgia State University.

Other Collaborators or Contacts

Activities and Findings

Research and Education Activities:
This research explores solutions to utilize the Azure cloud platform to improve the performance of the overlay computation over large scale GIS vector data sets. The education activities were conducted in Georgia State University (GSU), the lead institute of this collaborative research project. In general, GSU students in computer science could hardly understand the concepts and algorithms in GIS. It took very long time to train such students to have initial domain knowledge about spatial computation.

Findings:
1. Traditional data structure designed for desktop or standalone GIS is not appropriate for the efficient parallelization of vector overlay computation over the cloud platform. This conclusion is the same to other parallel computing platforms, such as MPI/grid and GPU.
2. The key to the resolution is not about how to develop new algorithm, but about how to use multiprocessing resources to do the same job that

is traditionally implemented on a single machine. Thus the solution can be used on other parallel computing platforms as well.
3. The solution is prototyped on a certain type of overlay computation, while the quality of the output needs to be further refined and examined. Since we use a third party component to do overlay calculation, thus the quality issue is not relevant to my solution.
4. In the case of topological operations, five (5) different solutions are identified to cover four (4) use cases, i.e. Union, Intersect, Difference, and XOR.
5. Most students in computer science could hardly understand the fundamental functions and concepts in GIS, while most students in GIS program could hardly implement HPC solutions. It will be great if a joint degree program can be initiated to train the next generation of scientists.

Training and Development:

This research offered excellent opportunities to those who worked on the project to learn fundamental GIS, parallelizing vector data for overlay computation over Azure cloud, and Web development. All these topics are new to those students involved in this project development.

Outreach Activities:

By introducing my solution for this project to Microsoft's researchers at Microsoft's Open Data for Open Science workshop, Microsoft's researchers understood and acknowledged the research challenge and the value of my solution in partitioning multi-layer spatial data for parallelized overlay computation.

A patent application entitled 'System and Methods for Parallelizing Polygon Overlay Computation in Multiprocessing Environment' was submitted on 6/14/2012 to claim both the new data structure and the solution for parallelizing overlay computation via the new data structure. The Patent Office has assigned this application Serial No. 13/523,196.

The outcome of this research will be submitted to professional journals and conferences or workshops after the patent is granted.

Journal Publications

Books or Other One-time Publications

Web/Internet Site

Other Specific Products

Product Type:

Other inventions

Product Description:

U.S. Patent entitled "System and Methods for Parallelizing Polygon Overlay Computation in Multiprocessing Environment" was submitted by GTRC on June 14, 2012. The Patent Office has assigned this application Serial No. 13/523,196.

Sharing Information:

If the patent can be granted, this invention will be published by the Patent Office. Other researchers can access the details from the online patent document. A journal paper is under preparation, in which the background information and research challenges will be covered. The paper will be submitted after the patent is granted.

Contributions

Contributions within Discipline:

The findings and solutions from this research are significant to both GIScience and computer science. Since this is the first time that GIS vector overlay computation is implemented in the Cloud, my findings and solutions will help to develop the next generation of GIS software based on the improved data structure to support parallel data processing and analytics over Cloud/Grid/GPU. This research also contributes to computer science in general to help understand that the key in multiprocessing environment is how to utilize the resource by effectively and efficiently partition the data on to the computing nodes for data processing, while any algorithm can be implemented, although improvement in algorithm may be helpful in general.

Contributions to Other Disciplines:

My findings and solutions discussed in this report are critical to the successful accomplishment of this research project. This project could not make any substantial progress and achievement until I proposed my solutions in March 2011. Student researchers from GSU had tried different approaches but their methods could not be successful until now. They made substantial progress and achievement in two weeks by following my solutions using the new data structure that I proposed and preprocessed.

Contributions to Human Resource Development:

This research has provided several research assistant positions to student researchers in which underrepresented groups are included. Students are exposed to GIScience and computer science and engineering areas, covering the research topics on spatial computation, high performance computing, cloud computing and parallelism.

Contributions to Resources for Research and Education:

Contributions Beyond Science and Engineering:

Vector overlay computation is the foundation to determine the spatial relationships between geographic features and has been implemented in a variety of applications, such as FEMA's HUZAS and the MAEviz module. The successful solution on parallelizing vector overlay computation can significantly accelerate the geospatial computation and improve the performance in decision making and emergency response and thus contribute to the public welfare beyond science and engineering.

Conference Proceedings

Categories for which nothing is reported:

Any Journal

Any Book

Any Web/Internet Site

Contributions: To Any Resources for Research and Education

Any Conference