Research on Parallel Buffer Analysis

with Grided based HPC technology

Lv Pang[1,2], GuoQing Li[1] and Yunxuan Yan[1,2]

[1]Centre for Earth Observation and Digital Earth, CAS, Beijing, China

[2]Graduated Sschool of CAS

Buffer analysis is one of the most popular GIS functions to take spatial analysis. It generates area around geospatial entity with a given width. This paper takes such operation as example to study how it can be paralleled with cluster and published on a grid environment. This approach can guide the study of new GIS architecture modes: Parallel GIS and Gird GIS.

The geospatial entity of vector data format usually includes point, line and area. So, the buffer generation mainly is based on these entities. Point buffer is simply operation comparing with line buffer, and the line buffer area combines the arcs with the parallel. Although there are many kinds of vector data format in different software, only shapefile is selected as the vector data format in this study, because it is the most simply and widely used vector data format. Two different parallel strategies are studied the paralleling of vector data structures. One is partitioning of the vector data, just like parallel algorithm of raster data. Each block of the vector data is processed, and then the master node gather all the computing nodes' data together to generate the final output. This strategy deals with the vector data with topology. The other way is to take the vector data's characteristics. One vector data usually contains a lot of entities and it can be divided and distributed to each computing node. For the case with entities N, computing nodes M, the data can be separated and mapped to 0-(N/M-1), N/M-(2*N/M-1)……((M-1)*N/M)-(N-1). With such mechanism, each computing node can get the entities' number, processes these entities, which is typical data parallel. The promoted mechanism also has been studied to use cache pool to keep load balance of computing nodes. The processed result will be directly send back to master node and assemble to be the final result buffered data.

Parallelization of buffer analysis provides high performance capability, and it also can be demonstrated as web service. SIG (Spatial Information Grid) research is national level grid research project in China and it can provide basic grid platform for GIS web-services. Under such grid structure, new web service nodes were created -- SIG vector data service node and SIG spatial analysis service node. With OGC WFS interface, SIG vector data service node offers user application the vector data according to user's request in form of key words. The parallel buffer processing programme in form of C++ class (also can be in form of Java Class)

can be easily assembled into OGC WPS (Web Processing Service) container and published as both standard web-service and enhanced WPS service.

To prove the performance of the above mechanism, necessary experiments are needed to be designed. One kind of experiment is for the accelerating performance with parallel cluster, while another kind experiment is for the tolerance of efficiency lost causing from web service cost. The well deigned experiments gave encouraging result to above two performances. In condition of the normal cluster platform with 10 nodes, the performance with more than 0.5 parallel speedup efficiency can be taken. In another experiment, the parallel buffer web service brought 7~8 seconds delay. However, comparing with the buffer processing time of more than 100 seconds, the service level cost is less than 5% in generally.

With the service assembling of GIS spatial analysis functions, the normal GIS architecture is changing to be a distributed grid GIS system. Because the mechanism of transferring to be independent high performance spatial analysis function web service can be used in any kind of GIS software platform with its develop interface, it can cooperate different spatial analysis function with web service. Overcoming the difference of GIS platform will great benefit to unique GIS system.