A Novel Approach for Geospatial Computational Task Processing in Grid Environment

Zhou HUANG, Yu FANG
Institute of Remote Sensing & GIS, Peking University
Beijing, P. R. China 100871

Geographic Information System (GIS) software architecture goes through the age of single-user, the age of multi-user integrating commercial DBMS to manage attribute data, and the age of Internet, focusing on data and accomplishing component reconstruction. Nowadays, GIS software is entering a new era of Grid GIS. Represented by Grid GIS, the next generation GIS has become the frontline and hot issue in both the academic community and industrial sector. However, implementing Grid GIS is confronted with a great deal of challenges, among which the grid-based geospatial computational task processing, i.e. the mechanism which efficiently processes the grid geospatial computational task submitted by users and obtains reliable results so as to improve the geospatial information sharing and cooperative computation capability. The paper is mainly focused on this problem.

The present researches and practices on grid-based geospatial computational task processing have several problems, especially in theoretical background, in practice and in feasibility. In order to solve these problems, we systematically propose the Problem Oriented Framework of Grid-based Geospatial Computational Task Processing (POFGGCTP). Based on analyzing POFGGCTP’s components and designing algorithms of its key technologies, the thesis introduces Nebula, a POFGGCTP-based grid system, analyzes Nebula’s test results, and discusses its advantages, scenarios as well as developing trends.

POFGGCTP has two meanings: in order to implement POFGGCTP, on the one hand, the grid framework and service environment on which the geospatial computational task depends are indispensable; on the other hand, the technologies of expressing, interpreting, distributing and result fetching should also be developed. According to its meaning, POFGGCTP consists of five key technologies: geospatial grid node architecture, geospatial grid resource catalog, grid-based geospatial computational task description language, global geospatial query
parsing and geospatial sequence execution management. Geospatial grid node architecture
defines the grid framework on which the geospatial computational task processing relies;
geospatial grid resource catalog stores the metadata of grid service environment which is
necessary for geospatial computational task; geospatial computation task description language
is to describe the grid-based geospatial computational task; global geospatial query parsing is
the parsing mechanism of grid-based geospatial computational task which is global when
submitted by users and has to be split into distributed geospatial sequence in order to be
executed on different grid nodes; geospatial sequence execution management includes
dispensing, result fetching and execution control mechanism of distributed geospatial
sequence. These five key technologies serve as supporting components of grid-based
deospatial computation task processing.

As to geospatial grid node architecture, the existing node architecture is not designed for
geospatial applications, so it cannot adapt fairly well to grid sharing and geospatial data
computation. Therefore, based on analyzing geospatial operation requirement and merging
Peer-to-Peer (P2P) node architecture into the traditional grid, this thesis brings forward a
domain-based geospatial grid node architecture. In this architecture, grid system is composed
of a series of ‘domain’ i.e. a node set which is formed via splitting according to certain
geospatial principles such as administrative areas; the nodes in a ‘domain’ can be classified
into two categories: domain manager and resource node. Grid user uses a domain manager to
launch a geospatial computation requirement. Then the thesis makes comparisons between the
proposed node architecture and the traditional ones in the following four aspects:
communication cost, updating complexity, reliability and marching degree with geospatial
operation requirement. Consequently, POFGGCTP grid node architecture performs better in
grid-based geospatial computation task processing, matches geospatial operation requirement
better, and has better reliability as well as lower communication cost.

As for grid-based geospatial computation description language, the current research is
almost procedure-oriented. The software with such mode lacks user-friendly interfaces, so is
inconvenient for users. The solution to this problem is designing a geospatial query language,
which is not only problem-oriented, but also corresponds closely to natural language for
organizing the grid-based geospatial computational task. Therefore, we develop Grid
Global geospatial query parsing is a process which translates a global grid-based geospatial computational task into a distributed geospatial sequence. The current distributed query optimizing algorithms, however, does not take into account the characteristics of geospatial data and query, so they cannot be applied well in the global geospatial query parsing. Hence, this thesis puts forward a methodology of global geospatial query parsing based on hybrid join strategy. In order to formalize the distributed geospatial sequence, we designed a descriptive language: Equivalence Distributed Program (EDP). We defined the grammar and procedure structure for it. We also combine the direct join optimization strategy with the semi-join one, developing a new algorithm—Hybrid Heuristic Optimization Algorithm (HHOA) for global geospatial query parsing. HHOA can translate the global geospatial computational task submitted by users into an optimized EDP which can be executed by the geospatial sequence execution management engine. Besides, we proved the validity and accuracy of HHOA through formal approaches as well as experiments. On the one hand, we use equivalence rules in relational algebra to prove the equivalence between global GGQL and the corresponding EDP; on the other hand, we carried out experiments which showed that the efficiency of HHOA is much better than some traditional parsing algorithms such as MST and SSD-1.

As to geospatial sequence execution management, the existing work concentrated in using centralized controlling mode to allocate, execute and manage computational task. This kind of mode not only usually leads to network congestion or “single point of failure”, but also fails to make full use of grid’s abundant computational resources. Consequently, this thesis developed a new mechanism of distributed geospatial sequence execution management which can support the dynamic task-migration. In the aspect of structure, query processor,
node communication processor and data transmission processor constitute the POFGGCTP
geospatial sequence execution engine. In the aspect of management strategy, the engine
divides the distributed geospatial sequence into several transaction stages. In each stage, the
engine assigns different management node to take charge of distributed task scheduling and
disseminating. In order to ensure the self-adaptability of the engine, we also design a rational
task compensation mechanism. Besides, through quantitative comparison and analysis, we
infer that POFGGCTP with a lower communication cost and better reliability can effectively
avoid the defects brought about by centralized controlling mode.

In the end, we carries out all the key technologies of POFGGCTP and implements
Nebula, a grid-based geospatial computational task processing prototype system. Via the test
for Nebula and the analysis of test data, we comes to the conclusion: POFGGCTP can adapt
to the cooperative processing of mass storage volume geospatial information in grid
environment, can make effective use of dispersed computational resources, can suit for the
geospatial computational task, especially those with higher selectivity, and can matches
dynamic characteristic of grid due to its well reliability.

**Short Bibliography**


Geographic Information System. *Science in China, Series E: Technological Sciences*,

[3] Di L, Chen A, Yang W, etc. The development of a geospatial data Grid by integrating
OGC Web services with Globus-based Grid technology. *Concurrency and

TeraGrid for *G*(i)(d) analysis. *Concurrency and Computation: Practice and