

High Performance Visualization of Spatial Interaction Data in Data-Intensive Computing Environment

Danhui Guo¹, Kaichao Wu¹, Jianhui Li¹, Yuanmin Chen², Zhenghua Zhang¹ and Delu Zhu¹

¹ Computer Network Information Center, Chinese Academy of Sciences

² China Internet Network Information Center, DNSLAB

Spatial interaction data visualization, also named as flow mapping, has become widely used for exploratory spatiotemporal data analysis to understand complex spatial phenomena such as human migration, commercial trading, subsurface flow of water shed, disease propagation, and social networks. With the sharp growth of spatial interaction data type and their scale, the flow mapping meet two big challenges: growing data scale and growing computing intense followed up. To solve the problem caused by big data throughput, computation efficiency of complex algorithm and flexibility of user interaction in data exploring, we build a three-tiers flow mapping services framework in data-intensive computing environment. Each tier employs different architecture to fulfill different demand in different stage. The first tier, names as data tier, which is responsible for data import/export and data preprocessing, is a computer cluster with scalable high throughput capacity. The second tier, named as model tier, is a hybrid computation framework integrated with CPU and GPU core aiming to different application and different algorithms to produce middle stage graph data through complex computation such as classification, clustering and visual clutter prevention. The last tier, named as visualization tier, is a service-oriented architecture based on OGC® Web Map Service and achieved through browser.

The .CN DNS log files is employed to testify effectiveness and efficiency of our framework. The record amount of .CN DNS visiting exceed 10 billion per day; and the cumulative size of log file can be exceed to 100 GByte. The mapping of interactive visiting among 70,000+ IP address is a big challenge to traditional flow mapping toolkits and other GIS software. The findings from this case studies suggest that the hybrid flow mapping framework significantly enhances flow mapping analytics that would otherwise not be achievable based on individual standalone methods and tools.

Keywords: flow mapping, data-intensive environment, visualization, spatial interaction data