DISTRIBUTED GEOSPATIAL COMPUTING (DGC)

Phil Yang
Joint Center for Intelligent Spatial Computing (CISC)
George Mason University
&
NASA Geosciences Interoperability Office (GIO)

SYNONYMS
Distributed Geospatial Information Processing; Distributed GeoComputation, Internet GIS

DEFINITION
Distributed Geospatial Computing (DGC) refers to the geospatial computing resides on multiple computers connected through computer networks. Figure 1 illustrates DGC within the C/S architecture [1]: where the geospatial computing is conducted by the geospatial components, which can communicate with each other or communicate through wrapping applications, such as web server and web browser. The Geospatial Components can communicate through application level protocols, such as HTTP or other customized protocols.

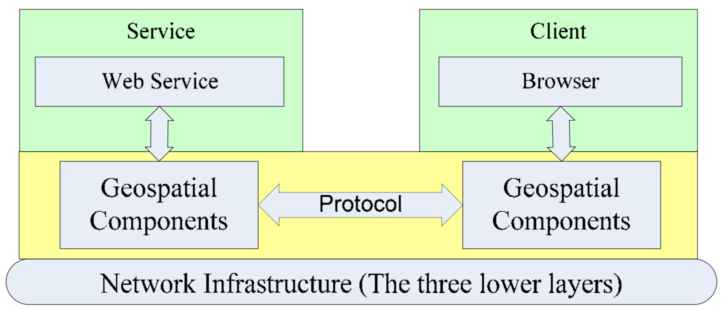

Figure 1 DGC within Computer Network Architecture

Geospatial computing includes utilizing computing devices to collect, access, input & edit, archive, analyze, render/visualize geospatial data, display within user interface, and interact with end users. Figure 2 illustrates that these previously tightly coupled components are now decoupled and distributed to a number of computers as either servers or clients across a computer network. The communications among these components are supported through different protocols: for example, the Data Exchange can be SQL [2], the Geospatial Protocol can be ArcXML [3], the message can be transmitted through pipe, and the client to web service can be HTTP [4].

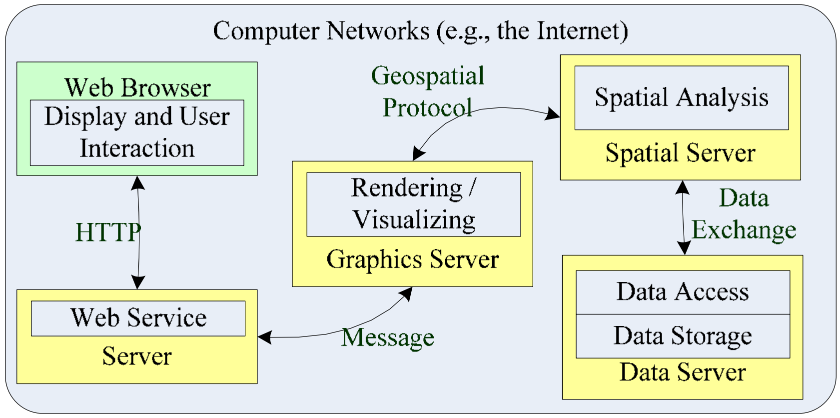

Figure 2 Distributions of DGC Components onto Servers within a Computer Network

This distribution of geospatial computing components matches the needs to integrate legacy and future components of geospatial computing deployed at different computers and hosted by different organizations [5].

Because the intensive computing component mainly resides in the geospatial analysis component, therefore, geospatial computing is focused on this component.

HISTORICAL BACKGROUND

Figure 3 illustrates the historical evolution of DGC. The major development of DGC can be traced back to the beginning of computer networks when DARPA process geospatial information across their intranet. Xerox's mapping server is recognized as the first system for processing distributed spatial information across the Internet [6]. In 1994, FGDC [7] was established to share geospatial computing across distributed platform, and OGC [8] and ISO/TC211 [9] were established to define a set of standardized interfaces to share the DGC platform.

In 1995, mapquest and other DGC applications are released and eventually gain a great success by leveraging a single geospatial computing, such as routing, to serve the public. In 1996, ESRI, Intergraph, and other GIS companies began to participate in the DGC effort by fully implementing geospatial computing components in the distributed environment. [6, 10]

In 1998, Vice President Al Gore proposed the Digital Earth vision to integrate all geospatial resources to support a virtual environment that could facilitate all walks of human life from research, development, to our daily life. In 2004, Google Earth was announced, and provided a milestone to gradually implement such a vision. In 2005, Microsoft started Virtual Earth. Within these two implementations, limited functions of geospatial computing are addressed, but they focus on the massive data and friendly user interaction, and solved many problems in dealing with massive simultaneous users by using thousands to millions of computers [11].

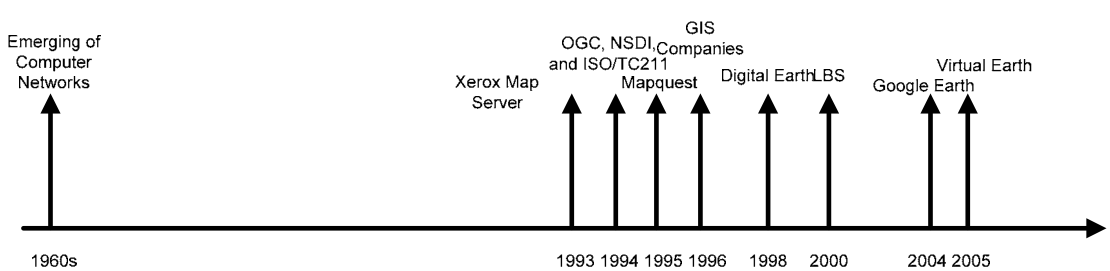

Figure 3 Evolution of DGC

Accompany these events that have profoundly impact DGC, many scholars also addressed issues on how to distribute geospatial computing [12], improve performance of DGC[13], parallelize geospatial computing[14], leverage grid platform [15] and agent-based environment [16] for distributed geospatial computing. Several centers/labs (such as CISC [17] and PTL [18]) are established to address the research needs in this field.

SCIENTIFIC FUNDAMENTALS

The scientific fundamentals of DGC rely on geometry, topology, statistics, and cartography principles to design and implement algorithms. Geospatial recognition and physical science provides conceptualization and model foundation for DGC. Earth sciences provide support for the DGC application design. Other relevant sciences, such as human recognition, provide support for other DGC aspects, such as graphical user interface.

The architecture of DGC also relies on the combinational scientific research, and is mainly driven by the distributed computing advancements, for example, the Client-Server, 3-tier architecture, N-tier architecture, tightly coupled, and Peer-to-Peer categories.

DGC is heavily depending on the concurrency process with the processing tasks increase in earth science and emergency or rapid response systems. Therefore, multiprocessor systems, multicore systems, multicomputer systems, and computer clusters, as well as grid computing is being researched to provide support to DGC.

KEY APPLICATIONS

DGC are used in many application domains, most notably the sciences and domains needing process of distributed geospatial resources, such as Oceanography.

Sciences

The first domain to use DGC is the Sciences. The science domains discussed here are geography, oceanography, geology, health.

Geography

DGC can be used to help integrate widely geographically dispersed geospatial resources and provide a comprehensive overview about the earth surface [19].

Oceanography

DGC can be used to help integrate the in-situ and satellite observation system and the modeling system to monitor Tsunami, sea level change, and coastal disasters [20].

Geology

DGC can help to integrate the observed earth surface heat flux and the in-situ sensor’s observation to monitor and possibly predict earth quakes [21].

Heath

DGC can help to integrate the health information and environment observations to find correlation between the environment changes and human health.

National and GEOSS Applications

NASA identified 12 application areas of interest to the national level [22] and GEO identified 9 application areas of interest to the global level [23]. Figure 4 illustrates the integration of earth observations, earth system models, with decision support tools to support decision or policy making [22]. All these areas require the using of DGC to integrate distributed earth observations, earth system models, to support decision support tools hosted by government agencies or other organizations. The 12 application areas are Agricultural Efficiency, Air Quality, Aviation Safety, Carbon Management, Coastal Management, Disaster Management, Ecological Forecasting, Energy Management, Homeland Security, Invasive Species, Public Health, and Water Management. The 9 application areas are Human Health & Well-Being, Natural & Human Induced Disasters, Energy Resources, Weather Information & Forecasting, Water Resources, Climate Variability & Change, Sustainable Agriculture & Desertification, Ecosystems, and Oceans.

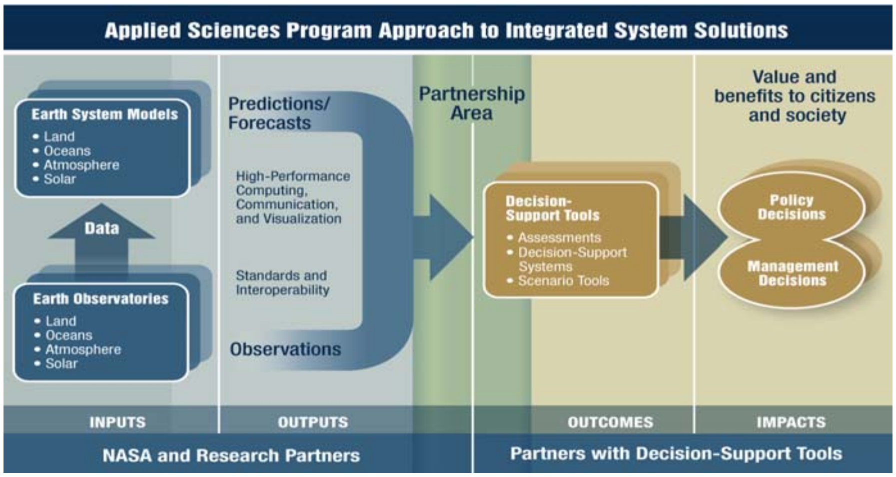

Figure 4 DGC is embedded in the integrated system solutions (Courtesy of NASA Applied Sciences Program).

Routing

DGC can be used to integrate roads network datasets, dispatch routing request to different servers to select the shortest or fastest path. This can be used in 1) driving directions, such as mapquest, yahoo map, 2) rapid response, such as routing planning after an emergency event, 3) operation planning, such as coordinating super shuttle, schedule FedEx package pick up.

FUTURE DIRECTIONS

More than 80% data collected are geospatial data. DGC is needed to integrate these datasets to support comprehensive applications from all walks of our life as envisioned by Vice President Gore.

The utilization of DGC to facilitate our daily life requires further research on 1) massive data management, such as Petabytes data archived by NASA, 2)intensive computing, such as real-time routing, 3) intelligent computing, such as real-time automatically identify objects from in-situ sensors, 4) quality of services, such as intelligent geospatial service searching engine, 5) interoperability, such as operational integration of DGC components in a real-time fashion, and 6) Cyber Infrastructure, such as the utilization of massive desktops and other computing facilities to support intensive computing or massive data management.

CROSS REFERENCE

Computer Environments for GIS, Spatial Information Retrieval

RECOMMENDED READING

1. C. Yang, D. Wong, M. Kafatos, R. Yang, Network GIS, *Earth Science Satellite Remote Sensing*, Qu, Gao, Kafatos, Murphy and Solomonson (Eds), Springer and Tsinghua University Press, Vol 2, pp. 252-269, 2006
2. ISO/IEC 13249. Information Technology-Database languages – SQL Multimedia and Application Packges, Part 3: Spatial, 2005

3. ESRI. ArcXML Programmer's Reference Guide, ArcIMS 4, 2002
4. C. Yang, D. Wong, M. Kafatos, R. Yang, Implementing computing techniques to accelerate network GIS, Proc. SPIE Vol. 6418, 64181C, *Geoinformatics 2006: GNSS and Integrated Geospatial Applications*; Deren Li, Linyuan Xia; Eds. 2006
5. C. Yang, V. Tao, 2005, Chap.5: Distributed Geospatial Information Service, *Frontiers of Geographic Information Technology*, Edited by Sanjay Rana and Jayant Sharma, London: Springer, 113-130.
6. B. Plewe. *GIS Online: Information retrieval, Mapping, and the Internet* (Santa Fe: OnWord Press) pp 6-14. 1997
7. FGDC website has information for clearinghouse, NSDI, GOS portal, and funding opportunities, http://www.fgdc.gov/
8. Open Geospatial Consortium (OGC) has information on WMS, WFS, WCS, CS-W, and other OGC specifications; http://www.opengeospatial.org
9. International Standards Organization, Technical Committee 211: GeoInformation, Geomatics, http://www.isotc211.org/
10. Z. R. Peng, M. H. Tsou, *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks* (Hoboken, New Jersey: John Wiley & Sons, Inc.) pp 2-5, 2003.
11. V. Tao, Microsoft Virtual Earth: New Horizons in on-line mapping, keynote at GeoInformatics'2006, Wuhan, China October 28-29, 2006
12. C. Yang, D. Wong, B. Li, *Computing and Computational issues of Distributed GIS, special issue of GIScience*, 10(1), 2005.
13. C. Yang, D. Wong, R. Yang, M. Kafatos, and Q. Li, WebGIS performance improving techniques, *International Journal of Geographic Information Science* (19):319-342, 2005
14. R. Healey, S. Dowers, B. Gittings, M. Mineter, *Parallel Processing Algorithms for GIS*. Bristol: PA: Taylor & Francis, 1998.
15. G. Aydin, A. Sayar, H. Gadgil, M. S. Aktas, G. C. Fox, S. Ko, H. Bulut,and M. E. Pierce. Building and Applying Geographical Information System Grids 12 January 2006, *Special Issue on Geographical information Systems and Grids based on GGF15 workshop, Concurrency and Computation: Practice and Experience*, 2006
16. J. Nolan. *An agent-based architecture for distributed imagery and geospatial computing*, Ph.D. Dissertation at George Mason University, Fairfax, VA, 2003
17. Joint Center for Intelligent Spatial Computing at George Mason University, http://landscan.scs.gmu.edu:8080/, 2005
18. Pervasive Technology Labs at Indiana University, http://grids.ucs.indiana.edu/ptliupages/, 2004
19. Executive Office of the President (EOP). "Coordinating geographic data acquisition and access: the National Spatial Data Infrastructure", http://www.fgdc.gov/publications/documents/geninfo/execord.html
20. L. Mayer, K. Barbor, P. Boudreau, T. Chance, C. Fletcher, H. Greening, R. Li, C. Mason, K. Metcalf, S. Snow-Cotter, and D. Wright, *A Geospatial Framework for the Coastal Zone: National Needs for Coastal Mapping and Charting*, Washington, DC: National Academies Press, 149 pp. 2004.
21. Cervone G., Singh R. P., Kafatos M., Yu. C.., Wavelet maxima curves of surface latent heat flux anomalies associated with Indian earthquakes, Natural Hazards and earth System Sciences, 5: 87–99, 2005.
22. Birk R., Frederick, M., Dewayne L. C., and Lapenta M. W., 2006, NASA's Applied Sciences Program: Transforming Research Results into Operational Success, *Earth Imaging Journal*, 3(3): 18-23.
23. GEO, The Global Earth Observation System of Systems (GEOSS)10-Year Implementation Plan, http://earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf, 2005

ACRONYMS

ArcXML, Arc Extensible Markup Language

CISC, joint Center for Intelligent Spatial Computing

C/S, Client/Server

DGC, Distributed Geospatial Computing

DARPA, the Defense Advanced Research Projects Agency

ESRI, the Environmental System Research Institute

FGDC, Federal Geographic Data Committee

GEO, Global Earth Observation

GEOSS: the Global Earth Observation System of Systems

GGF, Global Grid Forum

GIO, Geosciences Interoperability Office

GIS, Geographic/Geospatial Information System

GOS, Geospatial One Stop

HTTP, HyperText Transfer Protocol

ISO/TC211, International Standards Organization/Technical Committee 211

NSDI, National Saptial Data Infrastructure

NASA, National Aeronautics and Space Administration

OGC, the Open Geospatial Consortium

PTL, Pervasive Technology Lab

SQL, Structured Querying Language