Intersection of Geospatial Big Data, Geocomputation and Cloud Computing

Pouria Amirian*, Anahid Basiri**, Francois Van Loggerenberg*, Terry Moore**, Trudie Lang* and Margaret Varga*

* The Global Health Network, University of Oxford

** Nottingham Geospatial Institute, University of Nottingham

Extended Abstract

Big data is defined as high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and processes optimization (Laney 2012). Data component of big data can be broken down into two broad categories: human-generated and machine data. Machine data are generated automatically by machines without direct intervention of humans. Sensor generated data (like data generated by jet engines for monitoring the status and performance of engines) falls into the machine data category. Utilizing the mentioned machine data by using efficient and automatic approaches at large scale for research purposes largely ignored until recently. To be more specific, machine data has always been used for research purposes (like manual image processing of satellite images) and the automatic analysis of huge amount of machine generated data using supercomputers or volunteer computing has been done since late nineties. However the automatic analysis of machine data using new model of computation is new in big data world. The mentioned computation model is based on new advances in parallel processing and distributed computing and is very different from the traditional distributed computing. In addition, this new model of computation can be used to store, process and analysis of dark data which are operational data generated by sensors without any explicit intension for extract valuable insights. In fact, machine (dark) data has vast and as yet untapped potential as a source of highly valuable information, as they contain important insights about the system as well as users of the system. Often data component of big data has a positional component as an important part of it in various forms. Given the above definition of big data, many researchers believe that geospatial data has always been big data! In contrast, as it defined by Pouria Amirian (Amirian et al. 2014), "geospatial big data" term should be used only if the positional components in big data extensively used in storage, retrieval, analysis, processing, visualization and knowledge discovery. In other words, if some data in a big data system have positional components and/or used for simple visualization it is not necessarily considered as geospatial big data. Based on this definition, geospatial big data systems need certain type of techniques, algorithms for efficient management, analytics and sharing. In fact the mentioned needs are one of the challenges of GeoComputation (Amirian et al. 2014). As it mentioned by many researchers, management and analysis of geospatial data is complex and requires specific storage, processing, analysis and publication mechanisms (Taniar et al. 2013, Mahboubi et al. 2013). In fact management and analysis of geospatial data have been always revealed the limitations of information systems and

computational frameworks. Some researchers agreed that geospatial data may represent the biggest big data challenge of all (Davenport 2014).

The focus of our ongoing project is on finding useful insights from data generated by Point-Of-Care diagnostics sensors using specific techniques of geocomputation in the cloud. However by including location of the sensors in the analysis and by inclusion of other geographic layers, whole new set of useful questions can be addressed and therefore many new useful insights can be found. In addition using elastic computational resources provided by cloud computing and new computation model of big data technologies, it is possible to overcome one of challenges of geocomputation which is obtaining computability on geographical analysis problems. This paper will explain the challenges, opportunities and findings of a research project currently being done at the Oxford University in collaboration with Stanford University.

References

Amirian P, Basiri A, Winstanely A (2014) Evaluation of Data Management Systems for Geo-spatial Big Data, Computational Science and Its Applications–ICCSA 2014, 678-690.

Amirian P, Van Loggerenberg F, Lang T, Varga M (2014). Geospatial Big Data for Finding Useful Insights from Machine Data, GISRUK 2015

Davenport T (2014). Big Data at Work: Dispelling the Myths, Uncovering the Opportunities. Harvard Business Review Press.

Laney D (2012). The Importance of Big Data: A Definition, https://www.gartner.com/doc/2057415/importance-big-data-definition

Mahboubi H, Bimonte S, Deffuant G, Chanet J, Pinet F (2013). Semi-Automatic Design of Spatial Data Cubes from Simulation Model Results. IJDWM 9(1): 70-95.

Minelli M, Chambers M, Dhiraj A (2013). Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Wiley.

Taniar, D. and Rahayu, W. (2013). A taxonomy for nearest neighbour queries in spatial da¬tabases, Journal of Computer and System Sciences 79(7): 1017-1039.