

Chapter 2

BigGIS With Hadoop in MapReduce Environment: Towards an M2BG Framework

Nada M. Alhakkak



<https://orcid.org/0000-0002-5435-0730>

Baghdad College of Economic Sciences University, Iraq

ABSTRACT

BigGIS is a new product that resulted from developing GIS in the “Big Data” area, which is used in storing and processing big geographical data and helps in solving its issues. This chapter describes an optimized Big GIS framework in Map Reduce Environment M2BG. The suggested framework has been integrated into Map Reduce Environment in order to solve the storage issues and get the benefit of the Hadoop environment. M2BG include two steps: Big GIS warehouse and Big GIS Map Reduce. The first step contains three main layers: Data Source and Storage Layer (DSSL), Data Processing Layer (DPL), and Data Analysis Layer (DAL). The second layer is responsible for clustering using swarms as inputs for the Hadoop phase. Then it is scheduled in the mapping part with the use of a preempted priority scheduling algorithm; some data types are classified as critical and some others are ordinary data type; the reduce part used, merge sort algorithm M2BG, should solve security and be implemented with real data in the simulated environment and later in the real world.

INTRODUCTION

Over the last years, big data got a great interest almost from all internet users, especially from the academic people, government and some organizations. Mainly, Big data includes multiple huge amount of unstructured data that made its storing process hard and complicated; when using the traditional techniques because it needs more real time analysis. According to the rapid growth of services presented by internet companies the big data increased, i.e. Google deals with data greater than a petabyte (PB).

Big data resources are cloud computing, mobile internet, and GIS services. Big data is categorized by 5 Vs, i.e. huge volume, high velocity, high variety, low veracity, high value. The high variety is one

DOI: 10.4018/978-1-5225-9238-9.ch002

of the main challenges in dealing with big data. Consequently, any publication would need to deal with Structured data and semi-structured data or unstructured data, like text, image, video, and voice. However, dealing with big data faced a great challenge, related to data: capturing, storing, searching, sharing, visualization, and analyzing that includes inconsistency and incompleteness, scalability, timeliness, and security. Moreover, in order to improve the quality of data and the used analysis tools; developers should understand how to preprocess their data in an efficient way. This work is organized as follows: a review on Big GIS and its relation to Hadoop to manage its storing and processing with the best warehousing, then the most related works that have been done in this area, to be ended with the proposed framework and the conclusion with future work (Giachetta, 2015; Pasha, 2016).

BACKGROUND

The background of the present research handles some key terms that are widely and newly discussed as urban tools. These tools are Big GIS, Warehousing, Hadoop, Cloud computing with big geo-information, Intelligent GIS services, Intelligent GIS services, Parallel GIS based on Hadoop Cluster.

Generally speaking, the term ‘urban analysis’ is related to collecting geographical information about cities (Elshater & Abusaada, 2016) and towns and analyzing it. GIS is one of the main tools used for storing the collected data in databases and analyzing it for better decision making. This improves the required work with big geospatial data generated from GIS applications (Yeh, 1999; Elshater, 2015).

According to the vision of data holders in the GIS area for big data in managing, processing, visualization, and analyzing; Big GIS has been introduced. Mostly, most traditional GIS software is limited in dealing with big data challenges, discussed before. As a result, for developing the traditional GIS software that is related to big data issues and challenges, BIG GIS have been introduced; this new term used to manage and process big geospatial data (Yue & Jiang, 2014). The faster growing of all data types used by nowadays applications requires more flexible data warehousing software with the focus on some factors, i.e. data format and volume, varieties in data sources, unstructured data’s integration, and data analysis tool. With the faster change in Bigdata term, it’s important to solve the issues and challenges related to data storing and its related warehouses and use more flexible tools, like Hadoop and Hive (Sebaa, Chick, Nouicer, & Tari, 2017)

For Hadoop, it’s a Java-based programming framework with open source feature that supports managing big data in distributed an environment; solved the failure issues because of supporting clustering process in master-slave structure; which is all presented in Hadoop Distributed File System (HDFS). Meanwhile, there are two foremost programming tools that Hadoop depends on, i.e. Map Reduce and Spark. Each one has its own benefits and limitations. This work uses Map Reduce which covers three main factors. These factors are scheduling, monitoring, and re-executing failed tasks (Zhao, Chen, Ranjan, Choo, & He, 2016). However, Map Reduce leading idea can be summarized as follows:

- Divide the data into individual blocks, which are processed by map jobs in parallel.
- The output of the maps sorted by the framework is then inputted to the reduce tasks.

Cloud computing makes urban analysis possible to process geo-information in an effective and efficient way. SOLAP is one of its tools. However, adopting a cloud-based strategy improves the efficiency of geo-processing tasks. However, adopting those strategies means that the need is to rent commercial

6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/biggis-with-hadoop-in-mapreduce-environment/230521

Related Content

Public Utility Systems in the Republic of Serbia: Current State and Future Perspectives With Particular Regard to Economic Issues and SDG 6 Targets

Dejan Vasovi, Tamara Raenoviand Snežana Živkovi (2023). *Transformation and Efficiency Enhancement of Public Utilities Systems: Multidimensional Aspects and Perspectives* (pp. 1-26).

www.irma-international.org/chapter/public-utility-systems-in-the-republic-of-serbia/325683

Transformation Rooms: Building Transformative Capacity for European Cities

Gudrun Rita Haindlmaier, Petra Wagnerand Doris Wilhelmer (2021). *International Journal of Urban Planning and Smart Cities* (pp. 48-69).

www.irma-international.org/article/transformation-rooms/280153

Designing a Semantic Tool to Evaluate Web Content of Government Websites

Manoj A. Thomasand Jaffar A. Alalwan (2016). *International Journal of Public Administration in the Digital Age* (pp. 19-36).

www.irma-international.org/article/designing-a-semantic-tool-to-evaluate-web-content-of-government-websites/146805

Earthquake and Fukushima Nuclear Crisis in 2011 with Gender View

Yoiko Andô (2013). *Healthcare Management and Economics: Perspectives on Public and Private Administration* (pp. 190-202).

www.irma-international.org/chapter/earthquake-fukushima-nuclear-crisis-2011/75018

The Vietnam Capital Management and Budgeting Case Study

Lien Nguyen, Meagan M. Jordanand Thuy Thi Nguyen (2019). *Capital Management and Budgeting in the Public Sector* (pp. 280-309).

www.irma-international.org/chapter/the-vietnam-capital-management-and-budgeting-case-study/222859