Chapter 4
High–Level Languages for Geospatial Analysis of Big Data:
Strengths and Weaknesses

Symphorien Monsia
LTSIRS Laboratory, National Engineering School, Tunis, Tunisia

Sami Faiz
LTSIRS Laboratory, National Engineering School, Tunis, Tunisia

ABSTRACT

In recent years, big data has become a major concern for many organizations. An essential component of big data is the spatio-temporal data dimension known as geospatial big data, which designates the application of big data issues to geographic data. One of the major aspects of the (geospatial) big data systems is the data query language (i.e., high-level language) that allows non-technical users to easily interact with these systems. In this chapter, the researchers explore high-level languages focusing in particular on the spatial extensions of Hadoop for geospatial big data queries. Their main objective is to examine three open source and popular implementations of SQL on Hadoop intended for the interrogation of geospatial big data: (1) Pigeon of SpatialHadoop, (2) QLSP of Hadoop-GIS, and (3) ESRI Hive of GIS Tools for Hadoop. Along the same line, the authors present their current research work toward the analysis of geospatial big data.

INTRODUCTION

Over the last few years, mega-data or big data has become a major concern for many organizations. The term 'Big Data' refers to data sets that become so large that they become difficult to work with conventional database management systems. These massive data come from several sources among them the Web, sensor networks, satellites, drones, radars, cameras, connected devices (such as smartphones, tablets, etc.), geolocation practices and social networks (such as Twitter, Facebook, Google+, LinkedIn, etc.) online that bring together billions of users.

These phenomena considerably add to the challenges of big data for many organizations and have led to the emergence of Geospatial Big Data, which represents the application of big data issues to geographic data. Geospatial Big Data is therefore an essential component of the larger phenomenon of big data in that geographic data is an important part of the data collected and processed (Lee and Kang 2015). Franklin (1992) estimates that 80% of business data is geographic. An illustrative example is the LP DAAC (Land Processes Distributed Active Archive Center), an archive of terrestrial information originating from space borne sensors aboard NASA (National Aeronautics and Space Administration) satellites, which contains more than 1 petabyte of data and increases every day with new data.

This explosion of geographic data compels the community of researchers and developers of the geospatial domain to store and process them using traditional Big Data frameworks such as Spark (Zaharia et al., 2010), Flink (Carbone et al., 2015), MapReduce (Dean and Ghemawat 2004), Dryad (Isard et al., 2007), Hyracks (Borkar et al., 2011) and Hadoop (White 2015). Although these conventional Big Data systems can handle both geographic and non-geographic data, they display significantly lower performance compared to Geospatial Big Data processing. In fact, the only way to have Geospatial Big Data processed by traditional Big Data platforms is to either treat it as non-spatial data or to write a set of methods or functions as wrappers around existing non-spatial systems. However, doing so does not take any advantage of the properties of spatio-temporal data, which will lead to performance degradation (Eldawy and Mokbel 2016).

As a result, several extensions of traditional Big Data frameworks have emerged in recent years, many of which overcome this limitation by integrating geospatial functionality in a variety of ways among them HadoopGIS (Aji et al., 2013a,b), SpatialHadoop (Eldawy and Mokbel 2015), ESRI GIS Tools for Hadoop (Whitman et al., 2014), STARK (Hagedorn et al., 2017), SpatialSpark (You et al., 2015), GeoTrellis (Kini and Emanuele 2014), Simba (Xie et al., 2016), MD-HBase (Nishimura et al., 2013), GeoSpark (Yu et al., 2015), and GeoMesa (Hughes et al., 2015). In addition, some Geospatial Big Data frameworks are also implemented from-scratch among

18 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/high-level-languages-for-geospatial-analysis-of-big-data/279250

Related Content

Using a GIS and the American Community Survey to Address Community Health Problems
J.ane L. McCall, Amy K. Pasiniand Richard B. Wait (2003). *Geographic Information Systems and Health Applications (pp. 43-56).*
www.irma-international.org/chapter/using-gis-american-community-survey/18834

When Competitive Intelligence Meets Geospatial Intelligence
Christophe Othenin-Girard, Manon G. Guillemette, Eric Foleyand Claude Caron (2019). *Geospatial Intelligence: Concepts, Methodologies, Tools, and Applications (pp. 1809-1830).*
www.irma-international.org/chapter/when-competitive-intelligence-meets-geospatial-intelligence/222976

What We Know and Do Not Know About Mobile App Usage and Stickiness: A Research Agenda
Christopher P. Furner, Pradeep Racherlaand Jeffrey S. Babb (2016). *Geospatial Research: Concepts, Methodologies, Tools, and Applications (pp. 117-141).*
www.irma-international.org/chapter/what-we-know-and-do-not-know-about-mobile-app-usage-and-stickiness/149491

BIM as a Learning Tool in Design Studio
Türkan Uzunand Hülya Soyda Çakr (2022). *International Journal of Digital Innovation in the Built Environment (pp. 1-14).*
www.irma-international.org/article/bim-as-a-learning-tool-in-design-studio/306239

Usage of BIM in Smart Cities
Didem Ugurluand Begum Sertyesilisik (2019). *International Journal of Digital Innovation in the Built Environment (pp. 17-27).*
www.irma-international.org/article/usage-of-bim-in-smart-cities/245733