

Chapter 17

Framework for GeoSpatial Query Processing by Integrating Cassandra With Hadoop

S. Vasavi

V. R. Siddhartha Engineering College, India

Mallela Padma Priya

V. R. Siddhartha Engineering College, India

Anu A. Gokhale

Illinois State University, USA

ABSTRACT

We are moving towards digitization and making all our devices, such as sensors and cameras, connected to internet, producing bigdata. This bigdata has variety of data and has paved the way to the emergence of NoSQL databases, like Cassandra, for achieving scalability and availability. Hadoop framework has been developed for storing and processing distributed data. In this chapter, the authors investigated the storage and retrieval of geospatial data by integrating Hadoop and Cassandra using prefix-based partitioning and Cassandra's default partitioning algorithm (i.e., Murmur3partitioner) techniques. Geohash value is generated, which acts as a partition key and also helps in effective search. Hence, the time taken for retrieving data is optimized. When users request spatial queries like finding nearest locations, searching in Cassandra database starts using both partitioning techniques. A comparison on query response time is made so as to verify which method is more effective. Results show the prefix-based partitioning technique is more efficient than Murmur3 partitioning technique.

1. INTRODUCTION

Companies that use big data for business challenges can gain advantage by integrating Cassandra with Hadoop. Hadoop distributed file system framework can process voluminous of data generated from various sources. Out of various NoSQL databases, Cassandra supports linear scalability and high availability for ensuring fault tolerance. As such, when integrated, Cassandra and Hadoop together increase the pro-

DOI: 10.4018/978-1-5225-8054-6.ch017

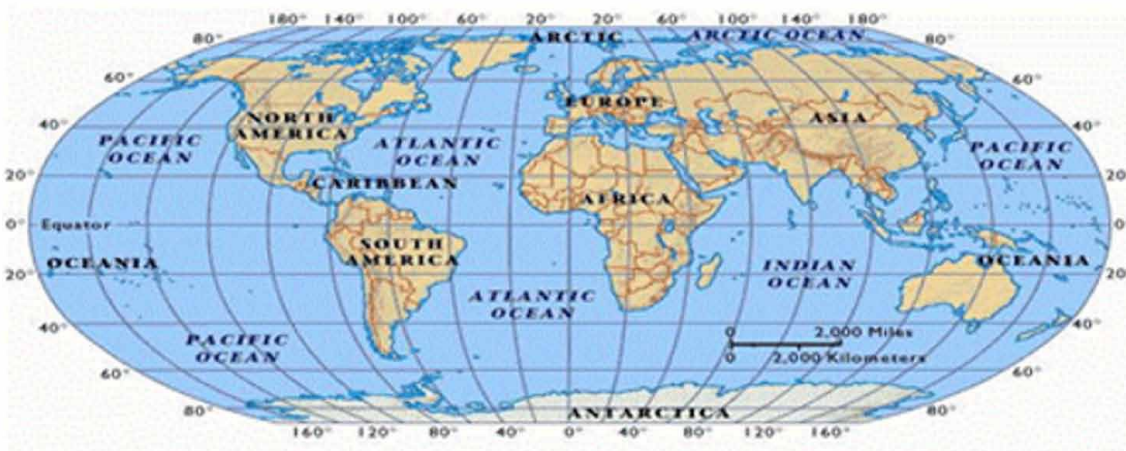
cessing capabilities to manage big data efficiently. Geospatial data helps in identifying the geographic location of an object, its features and boundaries on Earth. Such data can be analyzed to serve various purposes such as tourism, health care, geo marketing and intelligent transportation system. There are two data types for spatial data, vector and raster. Both data types stores object reference as latitude and longitude (Vertices/paths or grid cells) as shown in figure 1. Raster data includes remote sensing, photogrammetric, and vector data includes Geographical Positioning System (GPS), digitizing. Cassandra integration with Hadoop helps in Querying spatial data efficiently by reducing the query response time.

Raster data can be represented at its original resolution and form without generalization. But the location of each vertex needs to be stored explicitly. Advantage of Vector data is the geographic location of each cell is implied by its position in the cell matrix. Disadvantage is it is difficult to adequately represent linear features depending on the cell resolution. The following figure 2 and figure 3 presents example for vector data and raster data.

Motivation

Traditional databases (Relational database) are suitable for storing and querying structured data that guarantees ACID properties. With the emergence of internet, large amount of unstructured data is being produced. NoSQL databases, that guarantees CAP properties are suitable for storing such unstructured data. Dynamo, MongoDB, BigTable, HBase, Cassandra are designed to handle the data storage and processing with less response time. Even though MongoDB suits for complex queries such as social networking applications where we have to be optimise for latency, HBase and Cassandra when integrated with Hadoop are equivalently good in such a scenario. Cassandra DB and its query language CQL supports queries such as indexing, search libraries but not spatial queries. Even though some works are reported for labeling and retrieving Cassandra database, are not efficient. This chapter aims at adding the functionality of spatial querying for Cassandra database by integrating Cassandra with Hadoop.

Figure 1. Longitude and Latitudes of Earth (Geohash and its format, 2016)



34 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/framework-for-geospatial-query-processing-by-integrating-cassandra-with-hadoop/222907

Related Content

An Interoperability Mobile Application for Location and Travel Planning Services in Ghana: myTroski

Gift Dumedah, Charles Ofori Marfo and Samuel Ato Andam-Akorful (2022). *International Journal of Applied Geospatial Research* (pp. 1-17).

www.irma-international.org/article/an-interoperability-mobile-application-for-location-and-travel-planning-services-in-ghana/298299

The Image of Historic Urban Landscapes: Representation Codes

Giuseppe Amoroso (2016). *Geospatial Research: Concepts, Methodologies, Tools, and Applications* (pp. 344-372).

www.irma-international.org/chapter/the-image-of-historic-urban-landscapes/149500

On Predicting the Future Locations of Moving Objects: The State of the Art

Nicola Corona, Fosca Giannotti, Anna Monreale and Roberto Trasarti (2016). *Geospatial Research: Concepts, Methodologies, Tools, and Applications* (pp. 1899-1917).

www.irma-international.org/chapter/on-predicting-the-future-locations-of-moving-objects/149582

Drought and Flood Risk, Impacts and Adaptation Options for Resilience in Rural Communities of Uganda

Shuaib Lwasa (2018). *International Journal of Applied Geospatial Research* (pp. 36-50).

www.irma-international.org/article/drought-and-flood-risk-impacts-and-adaptation-options-for-resilience-in-rural-communities-of-uganda/190663

Historical GIS as a Platform for Public Memory at Mammoth Cave National Park

Katie Algeo, Ann Epperson and Matthew Brunt (2011). *International Journal of Applied Geospatial Research* (pp. 19-37).

www.irma-international.org/article/historical-gis-platform-public-memory/58625