(1)

A Hadoop Extension for Analysing Spatiotemporally Referenced Events

Mohamed S Bakli, Mahmoud A Sakr, Taysir Hassan A Soliman

Abstract:

A spatiotemporally referenced event is a tuple that contains both a spatial reference and a temporal reference. The spatial reference is typically a point coordinate, and the temporal reference is a timestamp. The event payload can be the reading of a sensor (IoT systems), a user comment (geo-tagged social networks), a news article (gdelt), etc. Spatiotemporal event datasets are ever growing, and the requirements for their processing goes beyond traditional client-sever GIS architectures. Rather, Hadoop-like architectures shall be used. Yet, Hadoop does not provide the types and operations necessary for processing such datasets. In this paper, we propose a Hadoop extension (indeed a SpatialHadoop extension) capable of performing analytics on big spatiotemporally referenced event dataset. The extension includes data types and operators that are integrated into the Hadoop core, to be used as natives. We further optimize the querying by means of a spatiotemporal index. Experiments on the gdelt event dataset demonstrate the utility of the proposed extension.

Keywords:

Spatiotemporal data, Hadoop, Geo-events, Movement analysis

Published In:

International Conference on Advanced Intelligent Systems and Informatics. , (Vol 639) , (pp.905-914)

(2)

A spatiotemporal algebra in Hadoop for moving objects

Mohamed S. Bakli, Mahmoud A. Sakr, Taysir Hassan A. Soliman

Abstract:

Spatiotemporal data represent the real-world objects that move in geographic space over time. The enormous numbers of mobile sensors and location tracking devices continuously produce massive amounts of such data. This leads to the need for scalable spatiotemporal data management systems. Such systems shall be capable of representing spatiotemporal data in persistent storage and in memory. They shall also provide a range of query processing operators that may scale out in a cloud setting. Currently, very few researches have been conducted to meet this requirement. This paper proposes a Hadoop extension with a spatiotemporal algebra. The algebra consists of moving object types added as Hadoop native types, and operators on top of them. The Hadoop file system has been extended to support parameter passing for files that contain spatiotemporal data, and for operators that can be unary or binary. Both the types and operators are accessible for the MapReduce jobs. Such an extension allows users to write Hadoop programs that can perform spatiotemporal analysis. Certain queries may call more than one operator for different jobs and keep these operators running in parallel. This paper describes the design and implementation of this algebra, and evaluates it using a benchmark that is specific to moving object databases.

Keywords:

Spatiotemporal algebra, Hadoop, MapReduce, moving objects

Published In:

Geo-spatial Information Science , (Vol 21 - No 2) , (PP.102-114)

(3)

HadoopTrajectory: a Hadoop spatiotemporal data processing extension

Mohamed Bakli, Mahmoud Sakr, Taysir Hassan A. Soliman

Abstract:

The recent advances in location tracking technologies and the widespread use of location-aware applications have resulted in big datasets of moving object trajectories. While there exists a couple of research prototypes for moving object databases, there is a lack of systems that can process big spatiotemporal data. This work proposes HadoopTrajectory, a Hadoop extension for spatiotemporal data processing. The extension adds spatiotemporal types and operators to the Hadoop core. These types and operators can be directly used in MapReduce programs, which gives the Hadoop user the possibility to write spatiotemporal data analytics programs. The storage layer of Hadoop, the HDFS, is extended by types to represent trajectory data and their corresponding input and output functions. It is also extended by file splitters and record readers. This enables Hadoop to read big files of moving object trajectories such as vehicle GPS tracks and split them over worker nodes for distributed processing. The storage layer is also extended by spatiotemporal indexes that help filtering the data before splitting it over the worker nodes. Several data access functions are provided so that the MapReduce layer can deal with this data. The MapReduce layer is extended with trajectory processing operators, to compute for instance the length of a trajectory in meters. This paper describes the extension and evaluates it using a synthetic dataset and a real dataset. Comparisons with non-Hadoop systems and with standard Hadoop are given. The extension accounts for about 11,601 lines of Java code.

Keywords:

Spatiotemporal, Hadoop, 3DR-tree, Trajectory data management, Big data

Published In:

Journal of Geographical Systems , NULL , NULL