**QuadTree Index for Spatial Objects on Apache MapReduce**

**Brief Documentation**

Donghan Miao

1. **Source and Compile**

File List

src

├── io

│   ├── QuadTreeFileOutputFormat.java

│   └── QuadTreeInputFormat.java

├── mapReduceTasks

│   ├── QuadTreeIndexer.java

│   └── QuadTreeQuery.java

└── quadIndex

├── FileLoc.java

├── InputParser.java

├── Point.java

├── Polygon.java

├── QuadTree.java

├── QuadTreeTest.java

├── Rectangle.java

└── SpatialObj.java

Packages

* io

This package includes implementation for customized input format and output format interface of Apache hadoop MapReduce.

* mapReduceTasks

This package defines two MapReduce jobs, namely, index building job, query search job.

* quadIndex

this package is an overall wrap-up for different spatial objects and quadtree structure.

Additional libraries

* libs
* hadoop.core

Minimum subset of Apache hadoop standard library are included for compilation.

Latest source code is available on Github.

>> git clone <https://github.com/miaodonghan/QuadTreeHadoop.git>

1. **Deploy hadoop**

The hadoop version we are using is 1.2.1 stable release. Please see Apache Hadoop official website for deployment instructions.

Single node deployment

<http://hadoop.apache.org/docs/r1.2.1/single_node_setup.html>

Cluster deployment

<http://hadoop.apache.org/docs/r1.2.1/cluster_setup.html>

1. **Prepare the data source**

Default Location

The source data could be placed under **/user/username/src** directory on hdfs.

Format

Raw data should be in plain text format. Each line indicates a spatial Object. Points, Rectangles and Polygons are supported.

Point:

1, x, y

Rectangle:

2, x, y, width, height

Polygon:

numberOfpoints, x1, y1, x2, y2, …, xn, yn

Compatibility with spatialhadoop

Now the latest version seamlessly support spatialhadoop data format, you can directly use the data source from spatialhadoop.org

1. **Indexing the spatial data**

Command

hadoop jar tb.jar mapReduceTasks.QuadTreeIndexer

Or alternatively, you can specify your input path and output path

hadoop jar tb.jar mapReduceTasks.QuadTreeIndexer in out

Output

Default Output Path is /user/username/out

**Run a query on mapreduce**

Query format

queryID, x, y, width, height

where queryID is a unique integer

Command

hadoop jar tb.jar mapReduceTasks.QuadTreeQuery queryID, x, y, width, height

Or alternatively, you can specify your input path and output path

hadoop jar tb.jar mapReduceTasks.QuadTreeQuery queryID,x,y,width,height in out

Output

By default, All output results are produced into /user/username/result