Lucene Spatial Index

Alix Bernard

March 5, 2021

Overview

Main indexing methods

In DBMS, data is stored as records where very records has a key field that helps recognizing it uniquely. Indexing is a data structure technique which aims to efficiently retrieve those records based on some attributes on which the indexing is done. Multiple types of indexing exists, the main ones are display in figure 1.

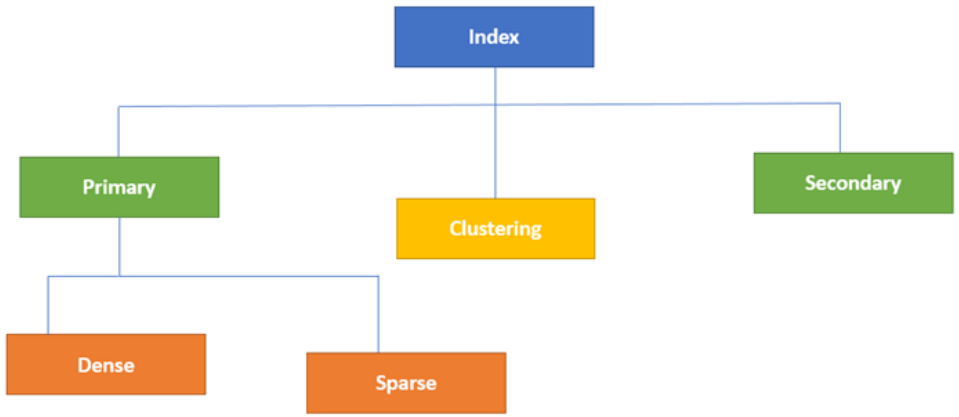

Figure: 1 Main indexes

Spatial indexing

When dealing with spatial data, it is apparent that traditional indexing is not fitted for the task, therefore spatial indexing has been created.
Spatial data (aka geospatial data or geographic information) is data identifying geographic location of features and boundaries — usually on earth. It is usually stored as coordinates and topology, and is data that can be mapped.

Spatial indexing

Spatial data is composed of three general data types: point, line, and region; furthermore they are differentiated in two classes: simple and complex.

Figure: 2.1 Simple spatial type

Figure: 2.2 Complex spatial type

Spatial indexing

One way to index those data is to index their bounding boxes; therefore when performing a search such as "which lines cross this area?", only the question "which bounding boxes overlap this bounding box?" is answer at first using the indexing since it is a really fast operation, afterwards an exact calculation is performed to answer the initial question accurately.

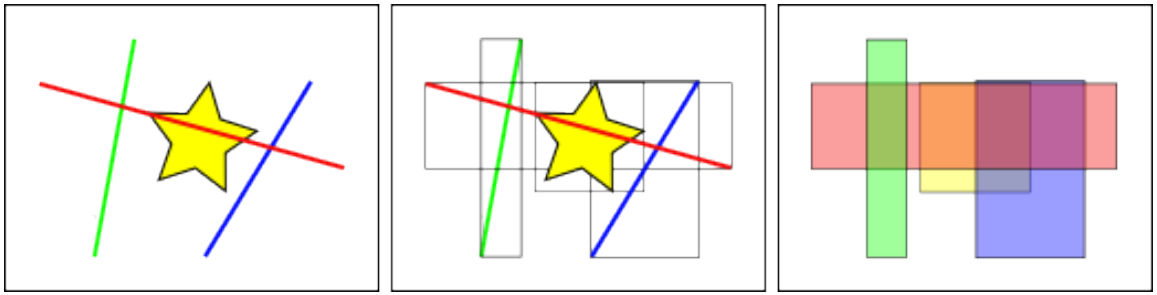

Figure: 3 Example of bounding boxes

Lucene Spatial indexing

One way to index spatial data is to use the Lucene Spatial indexing that uses the Lucene engine to perform a spatial search.
The Lucene engine (Apache LuceneTM), is a full-featured text search engine library written entirely in JAVA. This engine has a module allowing the implementation of a spatial strategy which encapsulates an approach to indexing and searching based on shapes. Different implementation of the spatial strategy will support different features.

For more details and examples:

- https://lucene.apache.org/core/4_0_0/spatial/
- https://orientdb.com/docs/3.0.x/indexing/Spatial-Index.html
- https://postgis.net/workshops/postgis-intro/indexing.html#how-spatial-indexes-work