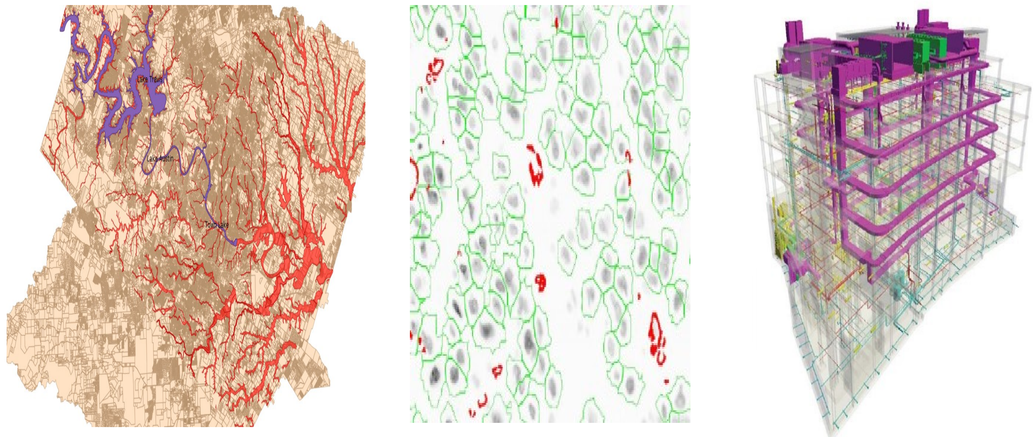
Dynamic Declustering for Parallel Spatial Databases

Tim Beauchamp, Nathaniel Brewer, Maria Patrou, Suprio Ray

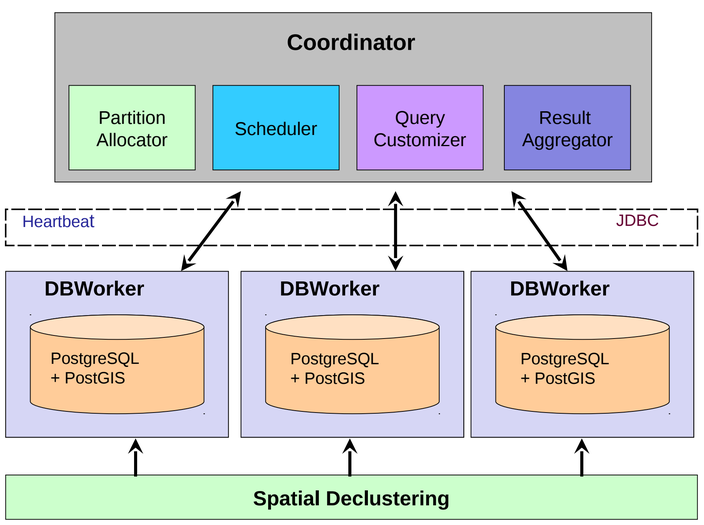
Faculty of Computer Science, University of New Brunswick, Fredericton, New Brunswick, Canada

Background

Spatial applications are becoming more and more popular in industry, commerce and research. One major challenge is that spatial join queries are long running and compute intensive.



Niharika is a spatial data analysis system for the Cloud. It is a hybrid approach inspired by HadoopDB, which takes advantage of Relational Database Management System (RDBMS) functionalities.

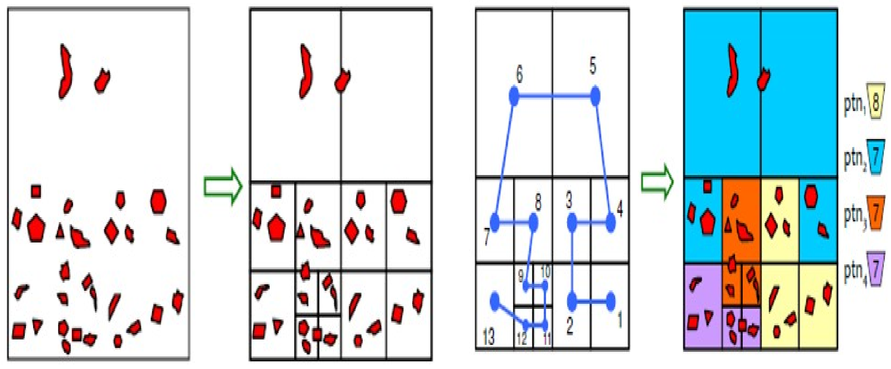


Niharika Architecture. Ref. [1]

The current implementation relies on static partitioning in a parallel database (based on PostgreSQL). However, operations that change the data require repartitioning the whole dataset and then rewriting back to each node where each worker is issued an identical query.

Problem

With static spatial declustering, tiles are divided into smaller ones from a predetermined limit in order to reduce skew. Afterwards, they are grouped into partitions for distributing the load evenly.



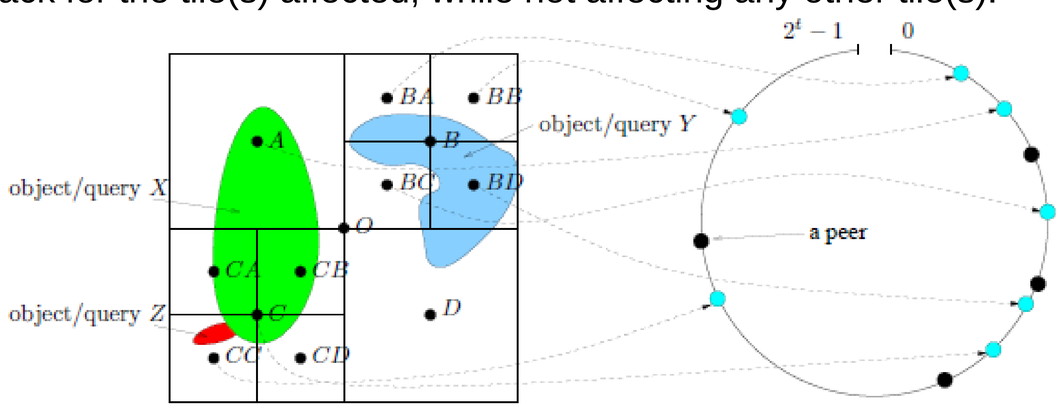
Spatial Declustering. Ref. [1]

The problem is that this process is static and hard to modify. Any change in the tiles require cleaning and re-uploading.

Approach

Our approach is to use big data tools, such as Hadoop/HDFS and a foreign data wrapper, to allow for dynamic declustering. Whenever a change is made in the dataset, only the necessary partitions should be changed.

Consistent hashing would allow involving only the necessary object(s) after an add or remove. Our system would perform retiling, and hashing back for the tile(s) affected, while not affecting any other tile(s).



Hashing Spatial Content over Peer-to-Peer Networks. Ref. [2]

Bookkeeping to assist this process will be done using Hadoop.

References: [1] Suprio Ray, Bogdan Simion, Angela Demke Brown, and Ryan Johnson. A parallel spatial data analysis infrastructure for the cloud. In Proceedings of the SIGSPATIAL International Conference (SIGSPATIAL), 2013.
[2] Aaron Harwood and Egemen Tanin. Hashing Spatial Content over Peer-to-Peer Networks, In Proceedings of the Australian Telecommunications, Networks, and Applications Conference (ATNAC), 2003.