

Parallel GIS Processing

Thesis Proposal

Nathan Kerr

April 2009

1 Intentions

Develop and explore methods for decreasing turn around time for processing and analyzing GIS datasets. Focus is given on the execution environment and not the user interface, though implementation difficulty of adhoc operations will also be considered.

2 Value

Exploration and analysis of GIS datasets is made difficult because of the length of time between starting an analysis or exploration routine and the return of results. Exploration and analysis abilities would be increased by enabling near realtime (minutes instead of hours) interaction with GIS capabilities.

3 Plan

Compare turn around time and difficulty of implementing a selection of GIS operations on several different environments.

3.1 GIS Operations

3.1.1 Loading and Preprocessing

Data set loading and indexing time (for all data)

3.2 Filtering

Find all the records that intersect a given geometry

3.3 Search

For a set of points representing employers, find the nearest parcel of land (polygon data) with compatible zoning codes.

3.4 Environments

3.4.1 PostGIS

PostgreSQL extended with PostGIS (OGC SFS compliant)

3.5 Hadoop

Map/Reduce environment with GIS extensions (Java Topology Suite)

3.6 Distributed PostGIS

Datasets distributed between multiple PostGIS instances connected with an MPI application

3.7 Paradise???

Distributed object relational database with GIS extensions (If I can figure out how to run it)

3.8 Serial JTS

Use the Java Topology Suite in a serial java program

3.9 Serial GEOS

3.10 Parallel GEOS

4 Completion Criteria

1. A comparison of turn around times for the given GIS operations in the different environments
2. A comparison of implementation difficulty levels for the GIS operations in each environment
3. Recommendations on which environment to further develop and use for processing GIS operations

5 Timeline

Dates are approximate.

- End of April 2009 - Define queries and run in PostGIS
- Mid May 2009 - Hadoop and Serial JTS
- End of May 2009 - Distributed PostGIS
- Mid June 2009 - Serial and Parallel GEOS
- Mid July
- End of July 2009 - Defend

6 Bibliography