

PERFORMANCE ANALYSIS OF MONGODB VS. POSTGIS/POSTGRESQL DATABASES FOR

Line Intersection and Point Containment Spatial Queries

Sarthak Agarwal, KS Rajan

September 16, 2015

Lab for Spatial Informatics

International Institute of Information Technology Hyderabad

FOSS4G Seoul, South Korea

PROBLEM STATEMENT

Why do we need a Internet connection to know directions from one point to another?

Why not deploy the server on the mobile device itself?

Why to query heavy SQL servers everytime?

- Spatial databases currently are primarily based on RDBMS. eg PostGIS
- They have a great potential to store, manage and query very large dataset.
- SQL databases face scalability and agility challenges.
- Spatial applications do not always have a fixed schema every time.

- PostgreSQL is an open source, object-relational database management system (ORDBMS).
- PostGIS adds support for geographic objects to the PostgreSQL object-relational database.
- The functions of PostGIS can be divided into 5 broad categories-
 - Management.
 - Conversion.
 - Retrieval.
 - Comparison.
 - Generation.

- **NoSQL** or Not only Sql databases for non-relational data stores.
- They have a great potential to store, manage and query very large dataset.
- They performs better in cases
 - where there is a need to improve the query response time
 - can handle the rise in the data storage and frequency at which it is accessed and processed.
- Spatial applications deals with problems like over time evolution of schema and data size

- Document oriented datastore.
- High performance and retains friendly properties of SQL.
- GeoJSON objects.
- Multiple geospatial indexes per collection - 2d, 2dsphere .
- Data can be imported from CSV files by converting it into GeoJSON objects.
- No support for R trees.

COMPARISON

SQL

- Not designed for distributed System
- Good for structured data
Unstructured data, points and lines not very suitable.

NoSQL

- Distributed Databases spread over multiple servers,
- Schema less databases where multiple geometries can be stored in the same column.

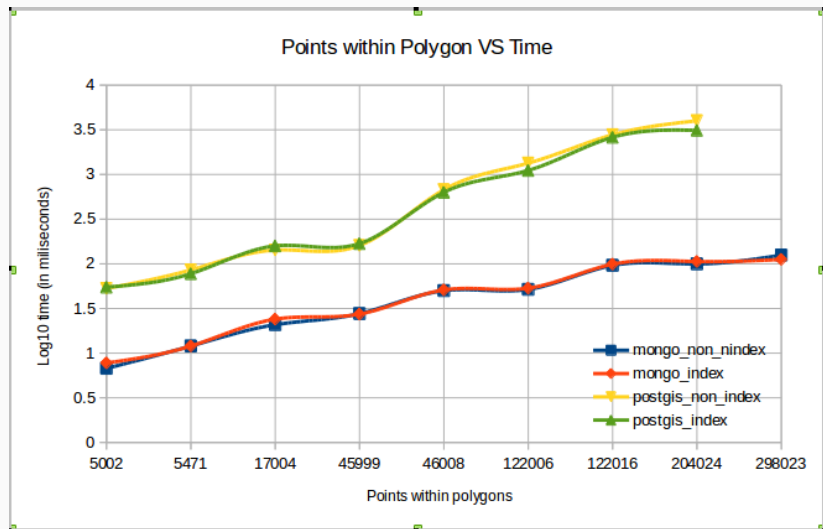
Does NoSQL hold a promise in the context of Spatial Databases and Spatial Queries?

HOW TO COMPARE

- Conventional Databases.
- Geometry Databases.
 - Point in a polygon
 - Line Intersection

- Synthetic dataset created for all cases.
- Test best case scenarios.
- Small data size to very big data size
- All the data in the analysis was processed using In-memory and no secondary memory was used.

PERFORMANCE



RESULTS

- MongoDB performs better as the dataset size increases.
- PostGIS fails at very large datasets.
- Indexing increases the performance of both datasets.
- PostGIS time increases exponentially as size of dataset increases, whereas MongoDB still performs within some bounds.

- These results suggest that MongoDB performs better by an average factor of 25x.
- Factor increases exponentially as the data size increases in both indexed and non-indexed operations.
- After observing these results NoSQL databases can be stated better suited for simultaneous multiple-user query systems including Web-GIS and mobile-GIS.

CONCLUSION

- non-relational databases are more suited to the multi-user query systems
 - potential to be implemented in servers with limited computational power.
- Further studies are required
- In future we are planning on expanding our study to other spatial query functions as well as spatial algorithms

QUESTIONS?