Research of possible Databases

M. Arzul

26 June 2013

1 Introduction

To select an appropriate database management system (DBMS), we have to consider what its requirements are. For our purposes, we will need to store historical data of social media queries which are turned into sets of coordinates. These are stored with the intent of using them to create retrospective heatmaps to monitor progress, trends, and patterns.

2 Relational or non-relational?

Since the sets of coordinates are of unknown and variable length and do not need to be in related tables, it would make sense to use a non-relational model, so that we can store self-describing, semi-structured data (we can assume a certain general shape with abstract optionalities and multiplicities, but can ask the data what it contains).

3 MongoDB or CouchDB?

If one looks at a developer's view comparison of NoSQL databases[1], among the popular choices that seem relevant are MongoDB 2.2 and CouchDB 1.2. Both are document-oriented[2] but give precedence to different aspects of the CAP theorem (of consistency, availability and partitioning, choose two). Both support partitioning (obviously), but Mongo supports consistency whereas Couch supports availability. Also, Mongo has geospatial indexing, which could be useful. Couch has very thorough version control, which will be unnecessary since our database will essentially be a warehouse.

---

[1] Cassandra vs MongoDB vs CouchDB vs Redis vs Riak vs HBase vs Couchbase vs Neo4j vs Hypertable vs ElasticSearch vs Accumulo vs VoltDB vs Scalaris comparison, Kristóf Kovács (http://kkovacs.eu/cassandra-vs-mongodb-vs-couchdb-vs-redis)

[2] http://blog.nahurst.com/visual-guide-to-nosql-systems

4   Conclusion

MongoDB has all the features we will need, and even some we can experiment with to make our system more efficient. It is well documented (see below) and is relatively simple to set up and use. This will be our choice of DBMS.

5   Links for MongoDB installation

To install MongoDB on the server (or for use with a localhost database):
http://docs.mongodb.org/manual/tutorial/install-mongodb-on-ubuntu/
To install PyMongo driver:
http://api.mongodb.org/python/current/installation.html
For a basic PyMongo tutorial:
http://api.mongodb.org/python/current/tutorial.html
For PyMongo API:
http://api.mongodb.org/python/current/