PREGEL

By: Phillip Picinic

5/6/15

PREGEL IDEA

- To distribute an algorithm over a large scale graph database via distributed systems similar to how MapReduce works. Since, MapReduce can have suboptimal solutions in graph databases, a more graph-centric approach is needed.

PREGEL IMPLEMENTATION

- Pregel uses an abstract API that is expressive in order to simply define algorithms to distribute over several workers on a large graph database. It divides work between partitions and vertices using combiners, aggregators, messaging, and fault tolerance to minimize latencies and maximize scaling. It provides a general framework to handle the distribution of work and allow high scaling of algorithms with even very simplistic algorithms.

PREGEL ANALYSIS

- I find that Pregel idea and implementation proves successful. With an abstract API there is very easy access for users. The distributed system scales well and provides very fast performance even simplistic algorithms giving the possibility of greater performance with high tuning and tweaking.

COMPARISON PAPER IDEA

- Parallel DBMSs are capable of outperforming MapReduce implantation of Hadoop through various scaling of nodes. Parallel DBMSs are also much easier to program than MapReduce.

COMPARISON PAPER IMPLEMENTATION

- Hadoop and two parallel DBMSs, Vertica and DBMS-X, are benchmarked fairly between five tasks via several node scales. The results show that Vertica larger outperforms both, and that both parallel DBMSs still outperform MapReduce.

COMPARISON PAPER ANALYSIS

- The benchmarks show convincing evidence that parallel DBMSs are capable of performing tasks, that MapReduce was designed for, better. However, given the small sample of benchmarks I doubt it can be said for certain that parallel DBMSs are the sole solution to ever problem. There also exists the possibility of improving both technologies using the advantages of others as well as seeing how they might work together to create unique solutions to more complex problems.

COMPARISON OF BOTH PAPERS

- The comparison paper shows that parallel DBMSs can outperform MapReduce, but Pregel shows promising benchmark results to graph based problems. A major problem with comparing the two is if relational databases can represent graph based data and at what cost is incurred.

STONEBRAKER TALK IDEA

- Stonebraker's talk covers that the traditional row-store DBMSs having become obsolete and are not good for any of the current markets. Instead he covers a number of various new database technologies each which has a place in a single or few markets.

PREGEL PROS & CONS

- Pregel seems to be a very good solution to a specific problem. It shows good benchmark results for large graph based problems and provides a generic approach to cover a multiple number of graph problems with very little coding. The problem behind Pregel is given that is not open-source there is not much comparison to already existing technologies and whether or not it is indeed the best solution for graphs. Given its similarities to MapReduce, it stands to question whether it suffers from the same problems.