Executing SQL over Encrypted Data in the Database-Service-Provider Model

By Chulakorn Aritajati

Summary:
“Database as Service” benefits users from data that is available in the Internet, so users can edit or access to it in the Internet. However, the service has a problem that data need a protection from hackers and providers to access to it. Researchers proposed techniques that use “the bulk of the work of executing the SQL queries” and “coarse index” in the server site without encrypting data. The client decrypt the data, so data is privacy, and clients do not have to manage a server. Author tested these techniques that have different number of buckets in two experiments with query and join operations. If number of buckets increases, network communication cost and query execution time in a client will decrease. Increasing of number of bucket also cause a server screens out false roles.

Comments:
It is the well-written paper. Authors used a lot of victual aids for readers, and every table and figures were necessary and useful. It is a good literature review. The study is well design that has both experiments to test performance of this technique.

Clients keep a key in their machine, but they may face security problems and data lost. Clients have to maintain their facilities and back up their key. If users lost their data, they will not decrypt their data in a server, so they lost data forever. Moreover, hackers can hack clients to gain keys because clients may have lower securities than a provider. I disagreed authors’ claim. Researchers claimed that users implement this technique and benefits from the database service provider.

I have a concern that this technique processes all queries in the server side, so the figure 11 presented that all execute time was in a server. This technique uses much performance of a server, so users have to invest much money to implement this technique. Researchers may have to provide alternative methods for users.

Researchers mentioned that the cost of query were I/O of a client and server, CPU of a client and server, and network communication, but they did not present performance of I/O and CPU performance in a client and server from both experiments. I have a concern on the client side because this technique in a client does not save data, so it may use much memory.

Authors should write a future work section to guide others to implement this algorithm.

Question:
What is a concern that the server side processes all query?
What technique can use with this technique to decrease load on a server side?
Is encryption technique equal or better than the industry standard?
Human-Powered Sorts and Joins

By Chulakorn Aritajati

**Summary:**
Researchers found that workers in Amazon’s Mechanical Turk did similar tasks to database tasks, so they can optimize Amazon’s Mechanical Turk by database techniques. Authors revamped workers’ performance to compare items for sorting and joining data. They used task batching, comparisons, rating, and filtering to decrease cost of Amazon’s Mechanical Turk. They tested all techniques to join two sets of celebrities’ pictures. Filtering and Smart Batching is the best joining operation, and Rating is the best ordering operation. Both techniques decrease cost of Amazon’s Mechanical Turk from $67 to $3 with high accuracy.

**Comments:**
It was the well thought paper. Researchers designed the study very well, because they presented performance of each technique. Its introduction covered many areas related to crowdsourcing. Authors created useful metaphor that crowdsourcing is similar to database process, and it can be optimized by database theories. However, a paper has minor issues. Researchers had few citations in “language overview and system”, “join operator”, and “sort operator”. Authors did not provide reasons that why they select these techniques over others. For example, they did not explain that why they create smart batching and what their ideas support it.

Researchers fixed the wage to one cent because Mechanical Turk could not create high quality work. However, Heimerl and others presented that local experts did sophisticate job and cost less than Mechanical Turk. Therefore, wage did not relate to quality of work, but quality of work based on expertise of workers. The value of one cent depends on economics. The one-cent in a developing country has more power of purchase than the United States. Inflation and deflation impacts on money. Thus, Smart Batching that has 5 roles was the best in 2002, but it might not work in 2012 because workers may think that one-cent was too low payment.

Researchers ignored performance of Smart Batching that was larger than 5 roles. They mentioned that workers did not this task, but they should present its results.

**Question:**
If we change workers to be experts, will performance of these techniques increase?
If we change workers to be experts, can we increase wage and batches?
Why did researcher not use ratio between wage and batches to optimize the technique?

Works Cited
Heimerl, K., Gawalt, B., Chen, K., Parikh, T., & Hartmann, B. (2012). CommunitySourcing: engaging local crowds to perform expert work via physical kiosks. (pp. 1539-1548). New York, NY, USA: ACM.