Peer review for project

Reducing workload and optimizing Data processing in a Data Warehouse

By Group 2 (Mohamed Gabr and Óttar Guðmundsson) – 2018.9.7

So a few points...

What set of processing tools are you going to use? Would be nice to know what libraries and APIs you intend to use. Will you compare/benchmark them against other libraries to make sure which one of them is the best for your implementation?

As you mention that your system will handle multiple requests, will you only consider basic requests or requests that are dependent on one another? Is a request a simple as "I need this" or can it get more complicated?

In warehouses some processes can not be completed unless some others have finished. This would probably be harder to manage as systems get more distributed rather than having it centralized.

Be aware that you have some restrictions on how to query databases, transactions/requests need to guarantee ACID properties which we don´t understand how it can be done if requests are handled in parallel for a same database, but it seems to look like the banking transactions problem. Distributing the task does not mean it will be done faster, it means it can use multiple machines seen as a single one. If you distribute the tasks on one machine, it may even be slower due to reconciliation between nodes. Consider the CAP theorem.

How would you define fairness against the public? If you were to be introduced to some company secrets that did indeed harm the public, wouldn't you consider whistle blowing a form of damage reduction? What do you define as economic damage? You shouldn't jump predict what is going to happen in the future. Also, you will most likely have to sign some sort of NDA contracts so I don't think this is a problem.

> "This also means that processes and/or algorithms which handle requests will execute in
>
> short amount of time thus saving on energy consumption."

Does it? How you will measure it? Isn't this defined by the size of the warehouse?

If you had a warehouse that was only handling 100 requests an hour a single computer would simply be enough but having multiple computers running and processing, simply waiting for requests would be over tuning for a simple company. Also note that shorter amount of time does not indicate saving energy consumption. A fast but heavy algorithm might drain a battery of a computer faster than slow and light algorithm.

Finally, we don't want to be picky on your English but our professor recommended us to use Grammarly for both syntax and grammar check. Not only did a lot of your text include a few syntax errors but in some cases, sentences were a little bit to long. They could be split into simpler phrases, to make sure you drag out the most important info.

Interesting idea. Looking forward to seeing the results 😊