Pregel: A System for Large-Scale Graph Processing

Mark Lozinski

3/14/17

Work cited:
cs.brown.edu/~debrabant/cis570-website/slides/pregel.pptx
Pregel A System for Large-Scale Graph Processing
A Comparison of Approaches to Large-Scale Data Analysis

What's the big idea: Pregel

- Pregel is trying to create a graphing system that is advanced enough that it can be able to track things like common newspaper articles, disease outbreaks and social media feeds to better be able to create data analytics. Doing so will, in a way, be able to make things like the internet much easier to understand.

How does it work: Pregel

- To address distributed processing of large scale graphs, Pregel built a scalable and fault-tolerant platform with an API that is flexible enough to express arbitrary graph algorithms.
- Pregel works similarly to MapRequest except for the fact hat it only uses one superstep. For each superstep every vertex of the graph calculates a certain function that is given by the programmer. This enables one to process large graphs which are distributed over several machines.

My thoughts

- I think that if this is as good as the article says it is that this is going to revolutionize the way that we use the internet and how the shopping is going to be advertised.
- I think that the way this is run will be the way that all graph processing will be done just due to the simplicity of its design.

What's the big idea: MapRequest

- It works in two parts, the Mapping part and the Reduce part.
- MapReduce has created the simplest and most efficient way to make advanced graphing possible.
- It set the framework for most programs like it, such as Pregel.

How does it work: MapRequest

- Each of the map instances are assigned a distinct portion of the input file by the MR scheduler to process.
- They are then organized based on what hash they got assigned and sent to the Reduce part.
- Each Reduce processes or combines the records assigned to it in some way, and then writes records to an output file, which forms part of the computation's final output.

My thoughts

- MapRequest may not be in the forefront of its field for very long but I think what they have made will impact the industry due to how efficient they have made it at the very base of the program.

What's the difference?

- Overall I don't think that you can really do a comparison on the 2 systems due to the fact that Pregel took the idea and backbone of MapReduce and made more advancements and features to make it be able to do more and be able to track information at a more efficient rate.

What's the big idea: Stonebreaker talk

- In his talk, Stonebreaker talks about how back in the 1990s it was possible and almost common idea that you could make a DBMS be able to do anything and that it was the end all be all of processing and storing information. He then goes on to explain that in the age of evolving technology that this just cant be the way things work anymore. In todays market you need a much more versatile DBMS that can be used for more than one thing, as he says it needs to be able to be used as a graphing processor, but also be used by data scientists so that I can be universally accepted.

Pros and Cons

Pros

- Pregel is looking to the future and is constantly doing updates and working on feature to keep their product up to date.
- 4 years after its release it is still one of the best programs for running graph analytics.

Cons

- It is having issues with being able to become more versatile and be able to do advanced regression formulas for data mining, and as stonebreaker talked about everything needs to be able to do a lot of different things.