Big Data Paper Summary:

- Pregal: A System for Large-Scale Graph Processing
- Choosing A Cloud DBMS: Architectures and Tradeoffs
- Michael Stonebraker on his 10-Year Most Influential Paper Award

Reisha Puranik 10/29/2020

Main Idea of Pregal

- proposed a large-scale graph processing system, called Pregal
  - vertex-based and will be able to show various algorithms
- when graph is initialized, has a series of supersteps where it invokes vertices
- these vertices can send messages to each other and then outputs the set of values by the vertices

How Pregal Is Implemented

- system has a Pregal library where it divides a graph into partitions where each has a set of vertices and their edges
- uses fault tolerance to detect if something has failed, and if something does fail, the master will reassign graph partitions and system will recompute the missing supersteps
- model is well suited for distributed implementations

Analysis of Pregal

I thought this paper proposed an interesting idea of a large scale graph processing system. I think that the graph also seemed flexible enough to be able to function in various places. It is also interesting how the vertices are sending messages in order to get the output. I also learned a lot about bipartite matching and semi-clustering and how the vertices interact with those topics. I thought this paper was really fascinating and proposed a really interesting way of outputting data and I learned a lot in the process.

Main Idea of Choosing A Cloud DBMS

- talks about how companies are choosing to move their applications to the cloud
- compares and contrasts various DBMS and their storage types as well as how costly they are on the cloud
- went into their experiments, testing which one is more efficient
  - found that while they explored architectural tradeoffs, it is extremely important to select a configuration that will fit the DBMS's purpose
  - some of the different versions will have some benefits but also might not have what the DBMS is looking for

How Choosing A Cloud DBMS is Implemented

- to test all the systems and accurately portray the results of how each system is running, they did three runs on different days and found that some did not support their complex queries
- also ran warm cache's and cold cache's to test each of their performances and how fast each was
- compared the storage costs as well as the data compatibility and the scalability
- All of these comparison techniques were used to decide how to pick and choose what system would be best for a specific DBMS

Analysis of Choosing A Cloud DBMS

I thought this paper was interesting in how to choose the best system. I think their experiment was an accurate representation of how to choose the best system and what they did to test each system. I liked this paper because it gave me more insight into other types of systems I had never heard of. I got to learn a lot about cloud storage and how to accurately pick which one would be better as well.

Comparison of Pregal and Choosing A Cloud DBMS

These two papers were very different but very interesting. Where Pregal talked about one specific system that they created, Choosing A Cloud DBMS talked about popular systems that were used. I would wonder if they included the Pregal system, how complex the queries would be compared to the other systems tested as well as storage costs, speed, scalability, etc. I think that both of these papers addressed their thoughts about how their specific system, or way of choosing a system, works perfectly for what they intend for it which was one of their main goals, to be able to pick the perfect system that works for what they plan to do with it.

Main Ideas of Stonebraker Talk

- In the 80s and 90s researchers had the idea that the "one size fits all" phrase applied to DBMS
- thought that the traditional row stores could be applied to everything
- as new technology and new ideas arose, it became clear that with all the new updates to technology, that one size does not fit all
- went into talking about different kinds of markets and how their idea back in the 80s and 90s could not be applied to that market
- with new technology coming up, it is important to note that different systems need different things and will not be able to have a "one size fits all" model

Advantages and Disadvantages of Pregal in Context of Stonebraker Talk and Choosing A Cloud DBMS

- disadvantages with Pregal is that, like in the Stonebraker talk, this idea may not be suitable for applications in a decade or two because new ideas will come about
  - Although systems can be altered, you can never expect what will happen and what new technology will offer
- advantages of Pregal, is that this system is what is working now and is altered and specified to what the group who made Pregal wanted it to
  - like how choosing what will work for the specific DBMS is discussed in Choosing A Cloud DBMS