Ikhlas Attarwala
MSAI 339 – Data Science Seminar
10-02-18 – Background Reading

Summarizing the "Redbook - Database Background" by Stonebraker

Michael Stonebraker, the winner of the 2014 Turing Award and author of this segment in the Red Book, seems rather frustrated over the repeated and failed attempts to invent lasting architectural fronts of data modeling by newer generations of scientists. Stonebraker began by critiquing and finding fault in society's ever-changing expectations in what would push the field of data science faster and forward, claiming that the same mistakes are made over and over again. He started by discussing how people had placed their hopes in XML when working with relational engines, but it was soon replaced by JSON. He moved on to similarly disparage Map-Reduce as being a data model that did not fit the bill; its users have turned toward using a relational SQL engine through Hadoop (HDFS). Nonetheless, the author does claim that while the architectural format hasn't changed, the details of the work definitely have over the last decade.

A positive note in this background chapter is that the author has expansive knowledge of, and has gone into great detail in recounting the history of repeated faults by analysts. Secondly, as he traverses one relational data model or format to another, he discusses the pros and cons of each, even offering thoughts on what he believes works and doesn't work depending on the input and need. One such example is how the author says that row stores (instances) in JSON are not suitable for sparse data. He says that while JSON is not useful for making general hierarchical data formats, it works well in column stores (values of instances), or as a data type. Finally, the author makes a good point in ending with a concise conclusion while offering advice from his experience of the situation, suggesting that newer researchers begin to understand the working systems of now before jumping in to attempt to reinvent the hierarchical structure before it's left in failure, as before.

Unfortunately, Stonebraker is far from being the perfect author, and while he's passionate about this segment, he introduces a great deal of bias and personal opinion, which never comes across as professional or impressive work. In addition, the overuse of quotation marks introduces confusion, with words the reader may not understand, such as his use of "legs" and "elephants". A second issue I find in his approach is that he does not explicitly state why new attempts will follow old mistakes, just that they will; this is likely due to his intense feelings toward repeated behavior. Lastly, building off of the first two, it would appear his rebuking of, say, ways to work around the hierarchical structure, limits himself from even introducing or being open to the idea that he could be wrong, and that times can introduce change.

I think that the author should channel his excitement of scoffing at "newbies" (as he calls them) for potentially repeating mistakes in the future into a direction that better develops on the advice he made at the end, where he suggested how newer students should understand the architecture of the 'dominant' systems for now. If he delved further into *why* these older hierarchical systems have lasted longer than newer proposals, that may better help correct the wrong he sees in repeated endeavors.