Survey of DBMS Development Over the Past 30 Years

California Polytechnic State University
Nick Alereza
Stephen Calabrese
Nick Clarke
Ethan Frame
Grant Frame
Luke Plewa

Abstract

The field of DBMSs has grown considerably since its inception in 1962. Periodically, the brightest minds of the field will publish papers that discuss, and potentially guide, the next steps of database development. Analyzing these papers in retrospect reveals a generally high level of aptitude. Here we surveyed five different papers to cover the range of expertise. After review and inspection, we have found the community to be trustworthy in their present experimentation and precise in their future expectations.

1 Introduction

This paper looks to survey the claims made in each of the documents listed in the bibliography. Each of the five documents is written by a different set of researchers and thus provides a different set of viewpoints. However, they all discuss the ensuing future of database management systems (DBMSs). This allows an ample comparison to be made. In addition, because these papers were written approximately thirty years ago, they can be analyzed through the perfect sense of hindsight.

Instead of going through each paper separately, it is more prudent to break them down into a series of categories. Each category will be a question that probes at a different angle of DBMS development. Therefore the following contents of the paper will be broken down into seven short sections followed by a conclusion to summarize.

2 Past Successes

Over the past 20 years, which obstacles have researchers been able to progress through? Which problems have been solved? This section's question inquires about the overarching effectiveness of the last 20 years in perspective of the authors. Some of the problems the authors stated were completely solved while others have been neglected. By analyzing to what degree has the database community been able to eradicate the complex difficulties of perfecting a DBMS we can chart both the technological development and the accuracy of these researcher's predictions, which we will delve into more next section.

It is useful to jump briefly back to the inception of databases, over 30 years age, to get the entire picture of database development. They originated in the 1960s with graph-based databases that were able to avoid data replication by utilizing standalone data applications. As the structure of data changed in the 1970s, most databases became relational. This changed again in the 1980s with improvements in object oriented databases. In the decades after that, the database community saw rapid advances. A few of these research breakthrough were seen in queries, both language evaluation and optimization, general data management, structural transactions, and systems were distributed and scalable.

Over the past 30 years database researchers have sought to solve the important problems of their time by designing databases which incorporate new advancements and ideas in the field. From object oriented relational databases [Neuhold and Stonebraker, 1988] to query optimizations [Silberschatz *et al.*, 1995], made possible by the enhancement of hardware in computers, these problems have caused a redesign of almost every aspect of database management systems in order to better meet the needs of users. This includes such ingenious applications of new data types, and enhanced transaction processing for multiple concurrent users. Problems that have been solved completely or significantly are those that have been around since the beginning of database design. A language which is powerful and can be used to answer any question, mechanisms which provide ACID transactions and efficient buffer managers are all problems that seem to have been solved for the traditional relational database. It seems that the problems have not been solved yet are those that are relatively new and thus have not been developed as much.

3 Legitimacy of Achievements

To what degree do you agree with the validity of the author's achievements? Have their claims been well founded or were some of their assessments incorrect? Each set of researchers claimed that different goals had been achieved by the database community and speculated on the significant challenges in the near future. The previous section presented an objective breakdown of that work and the benefits that ensured. This section subjectively evaluates it. With the assistance of retrospection, we can identify if indeed these

dilemmas were monumental as well as if any new obstructions cropped up unexpectedly.

For the most part the claims made by the authors are true. However there are some instances where their assessment was wrong, for instance in regards to search engine companies not using databases that much [Bernstein and et. al., 1998]. These types of companies have used databases extensively, and even developed new database solutions to meet the needs of their business. Some areas where their assessments were right were in regards to internet and the large amount of concurrent users that would need to be accessing a database at a single time.

Another such incorrect claim that I disagree with is that search engine type companies have not embraced database technologies. Companies like Google, Microsoft and Facebook rely heavily on databases for their data. Feeding off this, is the claim that all but the largest of databases will be in main memory. There are reasons why some databases are in memory and others are not. Despite these, they correctly predicted that millions and billions of gadgets will be using databases.They realized that each gadget is going to hold data and therefore need a database.

4 Overarching Problems

On the long term scale, how did the authors rank the problems preventing the perfect DBMS? Over time, which challenges did they mark as especially relevant as opposed to solved or even impossible? Although the idea of database management is relatively new, the field has had enough time to develop through multiple cycles. Evidently, there were different problems in the earlier years as opposed to those in the later years. this is not only because solutions were discovered to earlier problems, but because the way databases were used was continually mutating. The researchers in these papers, scattered throughout the years, each identified notable problems as their primary objective and the foremost challenge for those who were to follow them. In the early 1990s the problems that were most significant were the adaptation of database systems to the increasing hardware speeds and capacities. This was coupled with the need to store more complex types to match the burgeoning multimedia data [Silberschatz *et al.*, 1995]. In the early 2000s there was the rise of personal devices, which would cause databases to exist everywhere. This would cause consistency problems unless databases were adapted to communicate with each other [Bernstein and et. al., 1998]. In the late 2000s handheld personal devices were being introduced and becoming the norm. These devices had limited space and compute power, thus databases could not really be stored on them. Thus databases had to be moved to the cloud, and needed to be adapted to work in this environment [Bernstein and et al., 2003].

Another set of researchers identified other issues as being paramount. While still identifying application specific DBMS regardless of consumer hardware, they also specified the need for both object oriented databases, active databases, and especially distributed databases. Another important segment of the field that required research was the front end interfaces and the connections of a database to other software currently residing within the operating system. As more users make increased use of cloud computing, there is increased pressure on the structure of internet available databases as well as databases that can store complex data types, such as multimedia, while maintaining proper security measures.

One set of researchers suggested XML as a major disruption within the database system. This did not come to fruition as NOSQL emerged instead. An increase in the number of devices would bring communication complications, and with so many different databases, maintenance is going to need to be made almost automatic. Along with this self-tuning, these researchers noted the need for databases to manipulate their data in addition to storing it.

5 Forecasting Troubles

How well were the authors able to predict the roadblocks and obstacles confronting the modern database systems as the years progressed? As we previously evaluated the researchers' achievements through a backward facing lens, the question of this section seeks a subjective analyses of the author's ability to state and predict the problems their successors would be facing. Fortunately, the available knowledge of the past twenty years provides an ample scale on which to measure such accuracy. Notably, we will not delve into the casuistic circumstance which supposes that exclaiming the importance of a problem actually makes it more important, as it attracts the focused attention of other researchers. We will merely investigate their claims compared to reality and leave the realm of which problems 'could have been' more or less serious to another peer.

They have been able to correctly address the issues of their time and predict the problems coming in the near future. While there were some instances where the author's assessments were incorrect, like regarding XML causing huge problems [Bernstein and et. al., 1998], the authors claims have mostly been right on track. Choice of which topics to present within the papers was done systematically. When a conference was held for around 15 leading members of the database research community, each member would present a few topics of research that they would like to do in the future and the projects they are currently pursuing that they feel will have a large impact in the future. The members then vote on the topics to decide which were the most important and those were presented. With this process, or a similar method, persisting throughout these papers, I feel that the significance of the topics presented were generally in line with their true significance.

6 Missing Predictions

Which challenges were the authors' unable to predict during the last twenty years? Were these significant challenges and what was the effect of neglecting them? Although the researchers' were at the forefront of their field, there were undoubtedly certain troubles that they could not have predicted. Focussing on solving particular problems always comes at the cost of ignoring others. This sections attempts to ask if the spotlight was aimed in the correct direction; if

the most important problems were solved and the unexpected problems, although critical, were indeed worthy of only secondary attention.

Most of the authors seemed to have missed the NOSQL databases. They mentioned that XML would cause problems for databases but did not mention a way to fix it. A few of the researchers seemed to express expectations that were far too ambitious, such as assuming everyone would be allocated and actively using RAM on the scale of terabytes, a prediction which is still far off today.

7 Overall Assessment

Based on the previous sections and consequently the previous 20 years of research, how confident are you in the database community? How accurately can it label its achievements, overcome challenges, and predict where future obstacles may arise? Now that we have looked in depth at the community of researchers of database management systems, we can come to grasp with a general opinion of their aptitude regarding self knowledge. The ability to accurately assess one's self is directly related to the ability to infer future situations. It is also necessary to accurately state the problem before attempting to solve it. Therefore, we can use these connections to provide a reasonable evaluation.

Based on these papers, the database community has a firm grasp on the state of the industry and are able to adapt to its needs. They have, for the most part, been able to foresee problems and direct research so that they stay relevant. They realized that the webpages were going to move from static to dynamic page. The authors are were aware of the huge growth of people connecting and posting to the internet. Since they were able to predict the scale of what going to happen they did a decent job at predicting the future.The also were right about the need to accept more papers. They failed to correctly foresee a few parts of the industry, such as the emergence of in-memory databases. They also could be considered wrong that significant changes will come from only the academic community. Companies in industry, Google and Facebook for example, have been practically motivated to funds significant advancements in database technology.

8 Comparison to Present

How do you view the current forecasting papers, the Big Data white paper and the Claremont report? Do you trust the accuracy with which they rate the problems the database community will be facing? Now that we have established a general understanding of the DBMS community, we can come to a conclusion regarding their current output. By applying that standard to these papers, we gain a new perspective with which to glean information. This is highly valuable and can reduce wasted time while allowing current research to focus on the most important issues at hand.

They are definitely challenges that the database community and the computer science community will face in the coming years. We are generating a lot of data. All that data needs to be efficiently acquired, stored, and processed. These steps will require better tools and very specific designs to accomplish.Nevertheless, although the authors correctly predicted that there would be an abundance of data available, they did not say how to solve it or exactly how much of it there would be.They were only able to identify that there is going to need to be more metadata in order to process and understand all the data. This expansion of databases can already be seen in current research and advancements

9 Conclusion

The database community has grown immensely through its creation approximately 60 years ago. It has encountered and overcome numerous pitfalls to create its current state. Even now, incredible advancements are being made in efficiently and capabilities. In the past, the community has been able to accurately predict its failures and correctly define its achievements. There are a few exceptions that can be accounted to rapid or unexpected breakthroughs. Therefore, it is with great confidence that we can trust the modern collection of database researchers in both their current experimentation and their future foretelling.

References

[Bernstein and et. al., 1998] Phil Bernstein and et. al. The asilomar report on database research. September 1998.

[Bernstein and et al., 2003] Phil Bernstein and et al. The lowell database research self assessment. June 2003.

[Neuhold and Stonebraker, 1988] Erich Neuhold and Michael Stonebraker. Future directions in dbms research. May 1988.

[Silberschatz and Zdonik, 1996] Avi Silberschatz and Stan Zdonik. Stategic directions in database systems - breaking out of the box. 1996.

[Silberschatz *et al.*, 1995] Avi Silberschatz, Mike Stonebraker, and Jeff Ullman. Database research: Achievements and opportunities into the 21st century. 1995.