Will NoSQL Databases Live up to Their Promise?

David Alan Grier, Djaghe, LLC

The field of computing is not as unified as we like to think. The commercial and academic cohorts can have different goals and pursue different kinds of priorities. Some of Computer*'s most influential articles are those that bring these two groups together.*

Wither databases? As computing becomes more intertwined with the study of data, it is worthwhile to consider the origins and development of one of the fundamental tools of data science: the database. Neal Levitt's article, "Will NoSQL Databases Live Up to Their Promise?," (see "Article Facts") is the most prominent contribution from *Computer* to the study of databases.[1] While it is yet another example of the kind of article that *Computer* especially presents well, it is also a reminder that databases have held an anomalous position in computing. At this stage of the field, it is worth considering this position and to try to identify the direction in which database research is moving.

In his article, Levitt discusses a relatively new concept to computer science, the Not Only Structured Query Language (NoSQL) database, and asks a few pointed questions

ARTICLE FACTS

- Article: "Will NoSQL Databases Live Up to Their Promise?"
- Author: Neal Levitt
- Citation: *Computer*, vol. 43, no. 2, pp. 12–14, Feb. 2010
- *Computer* influence rank: #38 with 11,918 downloads and 97 citations

Digital Object Identifier 10.1109/MC.2021.3057443
Date of current version: 9 April 2021

0018-9162/21©2021IEEE

EDITOR **DAVID ALAN GRIER**
Djaghe; LLC; grier@gwu.edu

about this technology. He considers its origins, the different approaches to NoSQL databases, the benefits and drawbacks of each, and the possible applications of these technologies. The article introduced many members of the IEEE Computer Society to these ideas and encouraged a wider development of NoSQL databases.

Simultaneously, the article raised a fundamental question about the role of databases in computer science research. In 2010, SQL had been a mainstay of computing applications for 35 years. It was based on ideas from a paper that was foundational in computing, E.F. Codd's "A Relational Model of Data for Large Shared Data Banks," which was published in 1970.[2]

Codd's article firmly cemented the ideas of relational databases into computer science. Codd was a researcher at IBM's San Jose facility. Prior to the 1970s, he had studied a variety of topics in computer science. In the early 1960s, he had been part of the IBM team that looked at multiprogramming and scheduling. IBM had a long-standing interest in data and databases that stretched back to the 1930s. One of its largest customers was the U.S. Social Security Administration, which had built a massive database of employment records for every American citizen.

With the publication of Codd's article, relational databases became the dominant database technology in the field. In just a few years, a half-dozen companies were selling relational databases. Researchers had created rich literature on how to design and deploy these databases. SQL quickly followed. Within a decade, computer science educators were starting to teach the basics of relational databases in introductory programming classes and encouraging students to think about a programming environment that included a relational database.[3]

However, the relational model is not the only way to organize data. Certain applications, such as those connected to commerce and transactions, were better served by a different kind of structure. The point of sale terminal, which emerged in the 1980s, not only identified some of the limitations of the relational model but also pointed toward the opportunities that would eventually be found in distributed data models, large transactional models, and, ultimately, big data models.[4]

As researchers started to look at alternative models for data storage and retrieval, they encountered a structural quirk in the field of academic computer science. When academics were building the first computer science departments in the 1960s, they gave low priority to the study of databases. At the time, the largest funder of academic computer science, DARPA, pointedly excluded database research from the list of research that it was interested in funding. That list included many topics that have remained mainstays of academic computer science, including artificial intelligence, computer graphics, interactive computing, natural language recognition, and distributed computing.

It is not clear as to why DARPA did not include databases on its list of priority research. The agency leaders, who were well acquainted with the major labs and researchers in computer science, certainly knew that one of its contractors, RAND Corp., supported a major project on database research and had done some of the important early work on relational structures.[5] DARPA may have concluded that databases had already been adequately studied. They may have concluded that RAND Corp., IBM, and other companies would provide sufficient funding for this part of the field. No matter the reason, the organization kept its distance from database research, and other institutions followed its lead.

In his article, Levitt properly identified the growth of distributed computing with the rise of NoSQL databases. "Partly in response to the growing awareness of relational databases' limitations," he reported that vendors and users were "increasingly turning to NoSQL databases."[1]

He pointed to Amazon's 2007 product, Dynamo, as an important example of NoSQL technology. This product succeeded, in part, because it was integrated into the distributed computing ecosystem. Users could access the database through an application programming interface and paid for services on a cloud model that charged them only for the resources they used.[1]

There has always been a split between the academic and commercial sides of computer science. The two groups share common languages and concepts, but they tend to identify different goals and promote different priorities. "The database technology taught in standard database courses today is increasingly disconnected from reality," reported a recent study group. It is "time to rethink approaches

> While it is yet another example of the kind of article that *Computer* especially presents well, it is also a reminder that databases have held an anomalous position in computing.

to education, involvement with data consumers, and our value system and its impact on how we evaluate, disseminate, and fund our research."[6]

At its foundation, Levitt's article reminds us that the computing community is large and diverse. Like a complex database, it has large cohorts that are partially isolated from each other that don't always share their accomplishments, methodology, or aspirations. Some things that *Computer* does well is bringing these groups together and encouraging them to share their ideas.

ACKNOWLEDGMENT

For these 2021 columns, "Body of Knowledge" takes its information from a report prepared by the IEEE Publications office on 20 November 2020, and the statistics were current as of that date. Other citation services can and do provide different numbers.

REFERENCES

1. N. Leavitt, "Will NoSQL databases live up to their promise?" *Computer*, vol. 43, no. 2, pp. 12–14, Feb. 2010. doi: 10.1109/MC.2010.58.
2. E. F. Codd, "A relational model of data for large shared data banks," *Commun. ACM*, vol. 13, no. 6, pp. 377–387, June 1970. doi: 10.1145/362384.362685.
3. S. Khalifa et al., "The six pillars for building big data analytics ecosystems," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–36, Nov. 2016, Art. no. 33. doi: 10.1145/2963143.
4. A. Davoudian, L. Chen, and M. Liu, "A survey on NoSQL stores," *ACM Comput. Surv.*, vol. 51, no. 2, June 2018, Art. no. 40. doi: 10.1145/3158661.
5. R. Levien and M. E. Maron, "Relational data file: A tool for mechanized inference execution and data retrieval," RAND Corporation, Santa Monica, CA, Tech. Rep. RM-4793-PR, 1965.
6. D. J. Abadi, R. Agrawal, and A. Ailamaki, "The Beckman report on database research," *Commun. ACM*, vol. 59, no. 2, pp. 92–99. Jan. 2016. doi: 10.1145/2845915.

DAVID ALAN GRIER is a principal with Djaghe, LLC, Washington, D.C., 20003, USA. He is a Fellow of IEEE. Contact him at grier@gwu.edu.

Digital Object Identifier 10.1109/MC.2021.3066062