Exploring the Synergy between NoSQL Teaching and Research

Suneuy Kim
Department of Computer Science
San José State University
San Jose, CA, USA
suneuy.kim@sjsu.edu

***Abstract*—In this Innovative Practice Full Paper, we present our strategy for fostering synergy between NoSQL teaching and research. In response to the surging demand for NoSQL database technology, we developed the 'NoSQL Database Systems' course. The course serves as a catalyst for students to participate in NoSQL research projects, facilitated by tailored research-based courses. We demonstrate the seamless integration of NoSQL education and research within our curriculum, showcasing student-led research projects directly stemming from the course. Furthermore, we explore the positive outcomes of sustained research endeavors. By this, we mean that subsequent projects build upon prior ones, thereby enhancing the overall quality of research within a one- or two-semester timeframe allocated for each student. Emphasizing the importance of nurturing synergy between teaching and research, we elaborate on how our NoSQL curriculum benefits students in learning and engaging in research in the NoSQL field, supported by their feedback. We also discuss the mutually beneficial relationship between faculty research efforts focused on teaching topics and the resulting enhancements in teaching quality. Finally, we share the challenges encountered along the way.**

Keywords—Synergy, NoSQL Teaching, NoSQL Research, Database Curriculum

I. Introduction

Relational databases used to be the dominant solution for all data management problems. However, in the era of big data and AI, one size does not fit all. NoSQL (Not Only SQL) databases, optimized for scalability and flexible data models, are increasingly preferred in big data and AI applications where data can be diverse and constantly evolving [1]. In today's diverse database landscape, choosing the right database for an application is a key skill for database professionals. This trend imposes significant responsibility on faculty members teaching database topics: to equip students with the prevalent database technology, NoSQL, which students are likely to encounter in their professional pursuits.

To accommodate the increasing demands for NoSQL, we developed a NoSQL class in Spring 2017. The course is seamlessly integrated into our database curriculum with a relational database course as a prerequisite. This approach enables students to compare NoSQL features with relational databases, perceive NoSQL as a promising database paradigm rather than a replacement of relational databases, and develop the capacity to select the right database solution for a given problem in an impartial manner. The course has been successful, and we offer multiple sections of the course with full enrollment.

While teaching our NoSQL course, we noticed students developing a keen interest in the subject, prompting them to deepen their understanding through relevant research projects. We interpreted this student demand as an opportunity to cultivate a sustainable synergy between teaching and ongoing research projects involving students, thereby mutually enriching the academic environment. Especially, in teaching-intensive institutions where faculty members face high teaching loads alongside research expectations, fostering synergy becomes imperative to maintain teaching quality and sustainable research endeavors.

While the symbiotic relationship between teaching and research is highly valued, the simultaneous commitment to both can often lead to strain and occupational stress among academics. Without explicit strategies to align teaching and research efforts, coupled with strategic resource allocation such as faculty time, this relationship is prone to becoming adversarial [2, 3, 4, 5]. Furthermore, achieving synergy between these two pursuits requires significant effort. For instance, individuals excelling in research may find it necessary to adapt their teaching methodologies, simplify complex concepts, and enhance practical applicability. Such adjustments are essential for optimizing the interplay between teaching and research [6].

This paper presents our approach to cultivating this synergy between teaching and research in the realm of NoSQL databases. We incorporate research-tutored and research-oriented components into our teaching [7], alongside NoSQL research projects conducted in subsequent research-based courses tailored for student exploration. We recount our experiences, share findings, and address the challenges encountered in implementing this strategy. Despite the active research in the NoSQL field and existing studies on NoSQL teaching [8, 9, 10], prior investigations into the synergy between NoSQL teaching and research are lacking. We believe our approach fills this gap and offers potential for advancing both pedagogy and research in this area.

Section II outlines our strategy for achieving synergy between NoSQL teaching and research. Section III introduces the NoSQL class, detailing our efforts to seamlessly integrate this course into the curriculum and our approach to incorporating research components into relevant courses to foster synergy. Section IV presents research projects empowered by insights from the NoSQL class, highlighting three sustained projects focusing on benchmarking geospatial NoSQL systems, emphasizing the importance of continuity in research endeavors. Section V examines the benefits of the synergy between NoSQL teaching and research for both students and academics, drawing from our experiences and student feedback. In Section VI, we discuss the challenges encountered in our pursuit. Finally, Section VII concludes the paper.

II. Our Strategy for Achieving Synergy Between NoSQL Teaching and Research

This section delineates our strategy for cultivating synergy between NoSQL teaching and research. Central to the strategy is the development of a NoSQL course designed to both prepare and inspire students to engage in research, along with research-based courses where students come up with and carry out their own NoSQL projects. The developed NoSQL class is integrated into our curriculum to implement the strategy. Fig. 1 illustrates the course path within our curriculum, showcasing how students can leverage the synergy between NoSQL teaching and research.

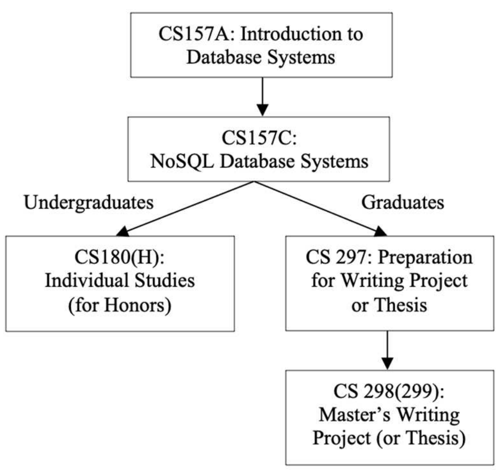

Fig. 1. NoSQL Teaching and Research Pathway

CS157A serves as the foundational course in our database curriculum, providing an introduction to database systems, particularly within the framework of relational databases. Building upon this foundation, CS157C, NoSQL Database Systems, requires CS157A as a prerequisite. In CS157C, students learn NoSQL techniques, with an emphasis on comparing them to relational database technologies whenever applicable. Crucially, the course instills in students an impartial perspective on these two prominent database paradigms by the end of CS157C. Both undergraduate and graduate students have opportunities to conduct NoSQL research through specialized coursework offered for student-led research: undergraduates can opt for the semester-long CS180(H), while graduate students can engage in the CS297/CS298(299) sequence. Students propose projects in their areas of interest, either by initiating their own projects or by undertaking successive ones in consultation with their advisors. They receive Credit/No Credit based on the assessment by their advisor (for CS180(H) and CS297) or a committee of three (for CS298(299)). Publication is not required to pass the course and is entirely up to the students' motivation. We observed that the approach of undertaking successive projects leverages sustained research efforts and tends to enhance the overall quality and depth of their research within the allotted timeframe. When funding is available to support a project, students can work on it through a research assistantship without needing to take a course, as illustrated in Section IV.B. However, such funded projects are not prevalent, and taking either CS180(H) or CS297/298(299) is the usual pathway for students pursuing research projects.

In the following section, we introduce the NoSQL class, highlighting its seamless integration into our curriculum. We will then explore our approach to incorporating research components into relevant courses, aiming to bridge teaching and research efforts.

III. The NoSQL class and its seamless Integration into the Curriculum

We developed a new course on NoSQL to equip students for the increasing demand in the industry for this emerging database technology. Additionally, we sought to stimulate their interest in NoSQL research opportunities. Initially piloted as an experimental topic course in the spring semesters of 2017 and 2018, the NoSQL class achieved significant success. Consequently, in Spring 2019, it became a permanent fixture in our curriculum, with multiple sections being offered, all of which reached full enrollment. Enrollment data gathered thus far is illustrated in Fig. 2.

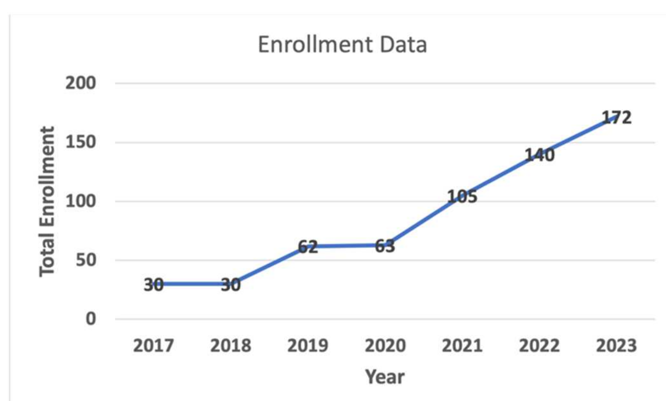

Fig. 2. Student Enrollment Data

The course comprises lectures, hands-on assignments (including research-oriented tasks), projects, and research paper presentations. We have meticulously designed the course content to seamlessly integrate it into our database curriculum. To achieve this, we first identify key concepts and techniques from prerequisite courses that directly relate to teaching NoSQL topics, such as schema, index, normalization, entity relationships, and ACID transactions. These concepts are then linked with corresponding NoSQL concepts and techniques and taught through comparative methods, emphasizing both similarities and differences. MongoDB and Cassandra, representing document store and wide-column store respectively, serve as exemplars for teaching representative aggregate-oriented NoSQL databases [11]. (Further details about the course design are provided in [10].) Additionally, we have included a presentation component in the course to ignite students' interest in NoSQL research. Many students have shown enthusiasm for NoSQL research, and we believe that the research papers presented during the presentation sessions have greatly contributed to this enthusiasm. Consequently, they pursue research projects for their undergraduate studies and Master's projects (or theses) by enrolling in CS180(H) and CS297/CS298(299), respectively.

The role of students and the types of research projects they engage in are crucial factors in connecting teaching and research within curriculum design. A framework, initially proposed by Griffiths [12] and further developed by Healey [7], outlines four approaches for academics to assess their methods of integrating research into teaching. The application of this framework to the computing discipline is detailed in [13]. Fig. 3 illustrates the four approaches, and Healey's matrix is utilized to demonstrate the effectiveness of incorporating research components into the involved courses. Our method of integrating research components into the courses is in close accordance with this framework.

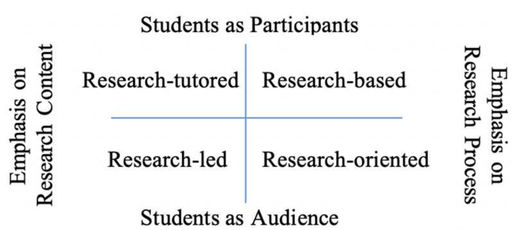

Fig. 3. Healey's Matrix about Curriculum Design and the Research-Teaching Connection (Source: based on Healey *[7]*)

In a research-tutored approach, students engage in small group discussions with the instructor to explore research findings. This method is emphasized during the final weeks of CS157C, allowing students to leverage their acquired fundamental knowledge and techniques in NoSQL for comprehending and discussing research papers. Groups of three students each present a selected research paper, followed by a class-wide discussion under the instructor's guidance.

In a research-oriented approach, students are given tasks through which they can understand and practice research skills. In our case, these tasks are strategically woven throughout the course and include activities such as comparing sharding and replication methods among various NoSQL systems, identifying shards responsible for specific queries, and analyzing the performance disparities of different query solutions addressing the same problem.

In a research-based approach, students gain insights into the research process by either individually or collaboratively conducting research projects. This approach is facilitated through courses like CS180(H) and the CS297/CS298(299) sequence.

In a research-led approach, students delve into cutting-edge research concepts, with the curriculum reflecting the current research interests of the discipline. While our database curriculum currently lacks a NoSQL course adopting this approach, plans are underway to develop a graduate-level NoSQL class likely to align with this method.

In summary, the synergy between NoSQL teaching and research is attained by developing and seamlessly integrating a dedicated NoSQL class into the database curriculum, while also incorporating research-tutored, research-oriented, and research-based approaches across relevant courses.

IV. NoSQL Research Projects

After completing the NoSQL course, students have the necessary knowledge and skills in NoSQL and are ready to start research projects. Here is a list of NoSQL research projects conducted by students who took the NoSQL class.

- **Benchmarking Scalability of NoSQL Databases for Geospatial Queries**
- **Performance Evaluation of MongoDB using GeoYCSB Macro-benchmark**
- **Performance Evaluation of Accumulo on top of GeoMesa using GeoYCSB**
- R*-Tree index in Cassandra for Geospatial Processing
- Schema Migration from Relational Databases to NoSQL Databases with Graph Transformation and Selective Denormalization
- Developing a MongoDB Monitoring System using NoSQL Databases for Monitored Data Management
- Benchmarking MongoDB Multi-document Transactions in a Sharded Cluster
- NewSQL Monitoring System
- Performance Evaluation of Postgres under GeoSpatial Workloads (with comparison to NoSQL)
- NoSQL or NewSQL: Which is better for Time Series Data Analysis
- NoSQL Databases in Kubernetes

Now, let's explore the first three highlighted projects from the list above, which were conducted sequentially, with each subsequent project building upon the research findings of its predecessor. Expecting undergraduate and master's students to engage in long-term projects can be challenging. For instance, in our case, undergraduate students typically undertake one semester-long project in CS180(H), while graduate students pursue two semester-long projects or theses in CS297/CS298(299). Despite this constraint, participants can explore their research more thoroughly by progressively advancing related tasks, leveraging findings from preceding projects. Our experience shows that results and findings from such incremental projects tend to meet publishable quality standards more effectively.

In the following sections, we examine each of the three projects, demonstrating the direct application of concepts from the NoSQL class and illustrating how these projects evolve incrementally to facilitate students' in-depth research efforts.

A. *Benchmarking Scalability of NoSQL Databases for Geospatial Queries*

This project, spanning two semesters and undertaken by a graduate student in both CS297 and CS298, focused on the development and assessment of a benchmark framework named GeoYCSB. The student utilized this framework to evaluate the geospatial capabilities of MongoDB and Couchbase, two prevalent NoSQL document stores. Leveraging MongoDB knowledge acquired from coursework, which covered its data model, CRUD operations, indexing, sharding, and replication mechanisms, the student conducted experiments using the Graffiti Abatement Incidents dataset of Tempe [14].

During the NoSQL class, the student practically deployed sharding and replication across a cluster comprising a minimum of three AWS EC2 instances, gaining essential experience for scalability experiments. This hands-on familiarity significantly contributed to the student's proficiency in conducting horizontal scalability experiments. Moreover, a thorough understanding of MongoDB acquired from the NoSQL class accelerated the student's grasp of Couchbase for this research project.

The research findings and outcomes of this project are detailed in [15], with sample experimental results presented in Fig. 4 and Fig. 5. Fig. 4 illustrates an experiment where the workload consists of 100% specific geospatial read operations, and the performance metric is the maximum throughput the system can achieve. In Fig. 5, where data are sharded across a cluster of specified nodes, the experiment highlights the effectiveness of NoSQL systems in achieving horizontal scalability, addressing the scalability challenges posed by big data management, which traditional relational database systems struggle to overcome.

It's worth noting that benchmarking projects offer substantial benefits to students, particularly those with limited prior research experience, especially when completed within a short timeframe. Beginning with small-scale benchmarking experiments, students can progressively advance toward comprehensive evaluations, making the project more manageable. Moreover, acquiring tangible intermediate results, such as data points, allows students to track their progress toward project objectives. The iterative nature of benchmarking not only enhances students' understanding of target systems but also fosters the development of critical thinking skills through result analysis. Also, we believe that starting this project within the realm of geospatial computing opens doors for sustainable research opportunities. As a significant portion of big data consists of geospatial data, with its volume growing by at least 20% annually [16], there is an increasing demand for geospatial applications. Therefore, we anticipate ample opportunities for project expansion to address this demand.

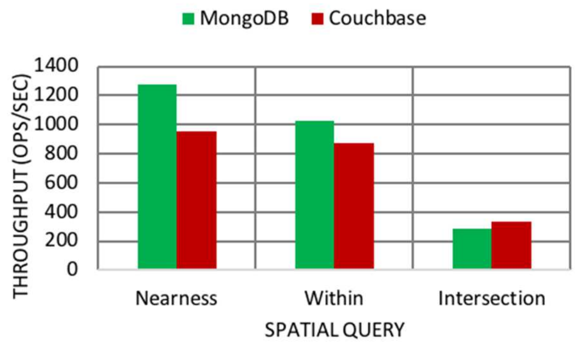

Fig. 4. Throughput of Individual Geospatial Queries from MongoDB and Couchbase

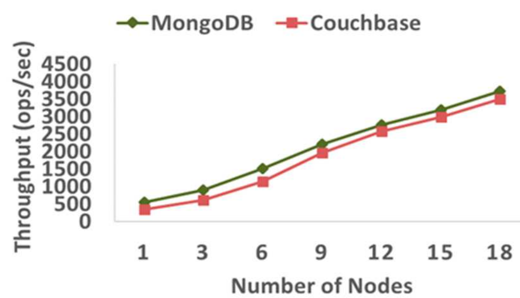

Fig. 5. Horizontal Scalability Comparison between MongoDB and Couchbase

In the next section, we elaborate on extending this project to incorporate the macro-benchmarking capability of GeoYCSB and its automation features for streamlined evaluation.

B. *Performance Evaluation of MongoDB using GeoYCSB Macro-benchmark*

Funded by the University Undergraduate Research Grant and RSCA (Research, Scholarship, Creative Activity) Grant, this project was undertaken by an undergraduate student who had completed the NoSQL class. In its initial phase, presented in the previous section, the GeoYCSB framework supported only micro-benchmarks, evaluating basic spatial operations in isolation. Recognizing the need for a more comprehensive assessment, this project aimed to incorporate macro-benchmarking capabilities into GeoYCSB. Macro-benchmarks simulate real-world application scenarios, providing a holistic view of system performance. Previously, benchmarking experiments were set up manually, requiring significant effort and time to test the systems for various factors. For example, setting up a cloud for an increasing number of nodes for the horizontal scalability test was time-consuming. This experience motivated the succeeding student to initiate the automation of the benchmarking process. Consequently, this second phase of the GeoYCSB project laid the groundwork for automating the benchmarking process, incorporating automated cloud resource allocation and database management system (DBMS) deployment using Ansible [17]. An industry expert provided comprehensive training in Ansible to facilitate this transition.

The student first established the GeoYCSB benchmark environment, ensuring its validity by recreating certain micro-benchmarking experiments conducted in the previous project. To develop macro-benchmarks, we identified representative use cases for a geospatial application utilizing graffiti data. These use cases served as the basis for formulating corresponding workloads based on MongoDB query implementations. Sample experimental results from this phase are illustrated in Fig. 6.

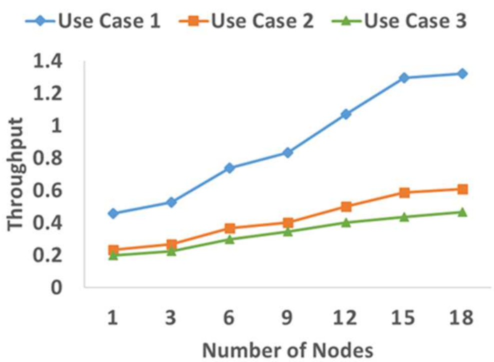

Fig. 6. Horizontal Scalability Evaluation with Various Use Cases

In the following section, we outline the expansion of GeoYCSB to accommodate various NoSQL systems with diverse data models, datasets containing intricate geospatial objects, and a broader spectrum of geospatial queries.

C. *Performance Evaluation of Accumulo on top of GeoMesa using GeoYCSB*

In this project, we aimed to show how GeoYCSB could be adapted for new NoSQL systems with different data models, larger datasets containing complex geospatial objects, and a wider variety of geospatial operations. To do this, we chose Apache Accumulo, a wide-column store, and applied the GeoMesa framework on top. The project was carried out by an undergraduate student in CS180H, with input from the previous student who had worked on GeoYCSB's macro-benchmarking. For our datasets, we used a combination of two new datasets from the Japanese Ministry of Land, Infrastructure, Transport, and Tourism [18, 19], which offered a larger spatial range and included complex geometries like polygons and line strings. We scaled these datasets up to a total of 40GB.

GeoMesa enhances the geospatial query capabilities of the underlying NoSQL system, such as Accumulo, enabling us to benchmark Accumulo with a broader range of geospatial operations than what MongoDB or Couchbase can support. Specifically, our workloads encompassed nine spatial predicate functions based on the DE-9IM (Dimensionally Extended Nine-Intersection Model) [20].

The student's familiarity with Cassandra from the NoSQL class expedited the learning process for Accumulo, given their shared features. Furthermore, leveraging accumulated research insights and findings from previous projects significantly propelled our progress. Fig. 7 illustrates a sample experimental result from this project.

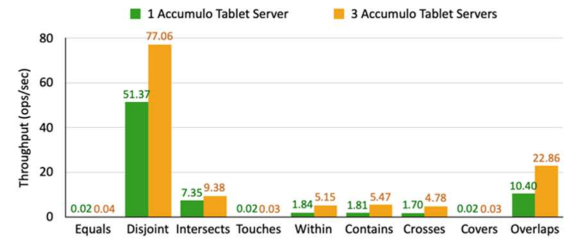

Fig. 7. GeoMesa Accumulo Throughput with 40GB data

The author collaborated with three students, each engaged in different phases of the GeoYCSB project, to publish a research paper that consolidates the integrated findings and outcomes in a journal [21].

D. *Upcoming Student Research Projects*

The NoSQL class has attracted significant interest among students in our computer science department. Being located in Silicon Valley, our students experience firsthand the intense industry demand for NoSQL techniques, and consequently, we anticipate that the demand for this course and for students pursuing research opportunities will continue to rise.

One immediate project we plan to assign to the next student(s) from the NoSQL class is the automation of the GeoYCSB benchmarking process. Automation of the evaluation process entails tasks such as cloud resource allocation, database system deployment, workload execution, and cloud adaptation. This project can be divided into manageable sub-projects, with each focusing on automating a specific aspect and assigned to an individual student. Students can utilize GeoYCSB with the partially automated capabilities they developed to conduct benchmarking experiments. This hands-on approach allows students to explore how specific automation stage influences the portability and extensibility of GeoYCSB. Another project can be dedicated to integrating these automated stages into a unified framework. Continued involvement in projects like this is likely to foster deeper research engagement among students, especially considering the limited timeframe typically available for each project.

V. Benefits of Synergy between NoSQL Teaching and Research

A. *Benefits to Students*

Incorporating research components into teaching should be carefully executed to ensure that the course content is delivered effectively without causing distractions. In our case, we enhance course materials through research-tutored, research-oriented, and research-based approaches, all within the boundaries of the curriculum. This approach enables all students enrolled in the course to engage with current research content, develop relevant skills, and familiarize themselves with research processes. Furthermore, learning from instructors actively involved in researching the course topic area offers numerous benefits to students. For instance, they gain insights into the latest advancements in the field, and instructors can impart intellectual curiosity and critical thinking skills.

We recently surveyed to gather feedback from students, receiving 7 responses out of the 11 surveys distributed. The response rate was lower than anticipated, possibly because we proactively reached out to graduates whose contact details may have become outdated. To address this issue, we plan to repurpose the survey as an exit questionnaire for departing students. This approach will enable us to methodically collect feedback from all intended participants. Please refer to TABLE I for the quantitative survey questions and responses.

TABLE I. SURVEY QUESTIONS AND RESPONSES

Q1	Did the course adequately cover foundational theories and principles essential for conducting your research?	7 Strongly Agree 0 Agree 0 Disagree 0 Strongly Disagree
Q2	Were the teaching materials provided in the course beneficial in understanding the foundational concepts necessary for future research?	6 Very 1 Moderately 0 Not very 0 Not at all
Q3	Did the course effectively lay the groundwork for understanding advanced concepts and methodologies for your research?	6 Strongly Agree 1 Agree 0 Disagree 0 Strongly Disagree
Q4	Did the course foster an interest in pursuing further research in NoSQL?	5 Significantly 2 Moderately 0 Did not 0 Decreased
Q5	Did you feel adequately supported by the instructor during your research endeavors?	7 Strongly Agree 0 Agree 0 Disagree 0 Strongly Disagree
Q6	Did you feel that involvement in research projects has deepened your understanding of the subject matter covered in the NoSQL class?	7 Significantly 0 To some extent 0 Not really 0 Not at all
Q7	How would you rate your overall experience in the NoSQL class in terms of preparing you for your research?	6 Excellent 1 Good 0 Fair 0 Poor
Q8	Would you recommend this course to other students as a foundational step for research in NoSQL?	7 Definitely 0 Probably 0 Probably not 0 Definitely not

The quantitative survey findings indicate that the course effectively stimulates students' interest in research, motivating them to actively seek out research opportunities and gain valuable insights gleaned from the NoSQL class. This prepares them to navigate the research steps and excel in their research projects, leveraging the knowledge acquired from the course.

We also conducted a qualitative survey to identify challenges encountered during research participation and to gather suggestions for improving the integration of research projects with teaching. Participants highlighted issues such as the lack of documentation for deploying and analyzing target database systems, the considerable scale disparity between the practiced database clusters in class and clusters they had to build in projects, and dealing with database systems not covered in class. Moreover, Docker integration was suggested by a student who completed the course a few years ago. We are committed to finding solutions for these challenges, understanding that they largely stem from constraints beyond the instructor's influence, notably limited time and resources. Despite these hurdles, Docker has already been successfully incorporated into the curriculum for several recent semesters. Additionally, we are actively pursuing plans to enhance the curriculum further by incorporating Kubernetes, aligning it with current industry trends.

Supplementary to the survey, we interviewed with the undergraduate student who led the second phase of the GeoYCSB project, acknowledging the student's significant contribution and extended engagement. The transcript of this interview is available in the Appendix.

B. *Benefits to Academics*

The synergy between teaching and research can yield rich rewards for academics. This symbiotic relationship is particularly beneficial in teaching-intensive universities where the demands of teaching remain high while expectations for research output continue to rise. In such academic settings, achieving synergy between research and teaching is paramount for fulfilling professional duties competently. Also, we have found that motivated students with a passion for research often offer valuable insights and pose thought-provoking questions, enabling faculty to enhance their critical thinking skills and pursue research endeavors in depth and breadth that yield significant and publishable outcomes.

VI. CHALLENGES

One challenge we face is that students can participate in short-term (one- or two-semester-long) projects only. Most students enrolled in the NoSQL class are senior undergraduates or first-year graduates who typically graduate one or two semesters after taking the course. Addressing this situation requires substantial additional effort and careful guidance during and after the research project to ensure that the research outcomes reach a publishable level.

Another challenge arises from insufficient resources, particularly faculty time, to supervise the numerous students seeking research opportunities. Establishing a positive and synergistic relationship between teaching and research hinges on implementing explicit management strategies that integrate teaching and research while efficiently managing faculty time.

Moreover, NoSQL techniques are undergoing rapid evolution to address the demands of modern big data and AI applications. Instructing students in these subjects adds substantial pressure and responsibility to academics, who must continually update course materials to keep pace with the rapid advancements in NoSQL fields.

VII. Conclusion

This paper explored the synergy between NoSQL teaching and research in a teaching-centric university facing heightened research expectations. We developed a comprehensive NoSQL course seamlessly integrated into the database curriculum, imparting a well-rounded understanding of both NoSQL and relational databases. Moreover, we embedded research elements, enabling students to sample recent advancements in NoSQL research while staying within the curriculum's scope.

The NoSQL course served as a catalyst, motivating students to explore NoSQL research further in subsequent research-based classes within our curriculum. These students were then able to apply their knowledge to excel in research endeavors, benefiting from collaborative efforts between academics and students, potentially leading to publishable research results. Active involvement in research within the instructional domain empowers academics to infuse cutting-edge domain knowledge into their teaching practices.

In conclusion, we posit that the synergy between NoSQL instruction and research not only enriches students' learning experiences but is also essential for academics to uphold teaching standards and sustain research endeavors amid resource constraints.

REFERENCES

[1] A. Davoudian, L. Chen and M. Liu, "A survey on NoSQL stores," *ACM Computing Surveys,* vol. 51, no. 2, pp. 1-43, 2018.

[2] K. Coate, R. Barnett and G. Williams, "Relationships Between Teaching and Research in Higher Education in England," *Higher Education Quarterly,* vol. 55, no. 2, pp. 158-174, 2001.

[3] M. Daumiller and M. Dresel, "Subjective Perceptions of the Teaching-Research Nexus and Occupational Stress at Universities," *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie,* vol. 50, pp. 126-138, 2018.

[4] M. Turk and J. Ledic, "Between Teaching and Research: Challenges of the Academic Profession in Croatia," *Center for Educational Policy Studies Journal,* vol. 6, no. 1, pp. 95-111, 2016.

[5] J. Humphreys, "Reflection," *Psychology Learning and Teaching,* vol. 8, no. 1, pp. 51-53, 2009.

[6] V. H. Vroom, "On the Synergy between Research and Teaching," *Journal of Management Education,* vol. 31, no. 3, pp. 365-375, 2007.

[7] M. Healey, "Linking Research and Teaching to Benefit Student Learning," *Journal of Geography in Higher Education,* vol. 29, no. 2, pp. 180-201, 2005.

[8] B. Fowler, J. Godin and M. Geddy, "Teaching Case Introduction to NoSQL in a Traditional Database Course," *Journal of Information Systems Education,* vol. 27, no. 2, pp. 99-104, 2016.

[9] S. Mohan, "Teaching NoSQL Databases to Undergraduate Students: A Novel Approach," in *SIGCSE '18: Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, February 2018.

[10] S. Kim, "Seamless Integration of NoSQL class into the Database Curriculum," *ITiCSE '20: Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education,* pp. 314-320, 2020.

[11] P. J. Sadalage and M. Fowler, "Chapter 2. Aggregate Data Models," in *NoSQL Distilled A Brief Guide to the Emerging World of Polyglot Persistence*, Addison-Wesley, 2012.

[12] R. Griffiths, "Knowledge production and the research–teaching nexus: the case of the built environment disciplines," *Studies in Higher Education,* vol. 29, no. 6, pp. 709-726, 2007.

[13] S. White and A. Irons, "Relating research and teaching: learning from experiences and beliefs," *ITiCSE '09: Proceedings of the 14th annual ACM SIGCSE Conference on Innovation and Technology in Computer Science Education,* pp. 75-79, July 2009.

[14] "Graffiti Abatement Incidents Data Set," https://catalog.data.gov/dataset/graffiti-abatement-incidents.

[15] S. Kim and Y. S. Kanwar, "GeoYCSB: A Benchmark Framework for the Performance and Scalability Evaluation of NoSQL Databases for Geospatial Workloads," *Proceedings of 2019 IEEE International Conference on Big Data,* p. 3666 – 3675, 2019.

[16] J.-G. Lee and M. Kang, "Geospatial Big Data: Challenges and Opportunities," *Big Data Research,* vol. 2, pp. 74-81, 2015.

[17] "Red Hat Ansible," [Online]. Available: https://www.ansible.com..

[18] "Japan's Counties Dataset," [Online]. Available: https://nlftp.mlit.go.jp/ksj/gml/datalist/ksjtmplt-n03-v24.html.

[19] "Japan's Routes Dataset," [Online]. Available: https://nlftp.mlit.go.jp/ksj/gml/datalist/ksjtmplt-n07.html.

[20] C. Strobl, "Dimensionally Extended Nine-Intersection Model (DE-9IM)," in *Encyclopedia of GIS*, Springer, 2008, pp. 240-245.

[21] S. Kim, Y. Hoang, T. T. Yu and Y. S. Kanwar, "GeoYCSB: A Benchmark Framework for the Performance and Scalability Evaluation of Geospatial NoSQL Databases," *Big Data Research,* vol. 31, 2023.

APPENDIX: Interview Transcript

We interviewed the undergraduate student who made the most significant contributions to the GeoYCSB project. During the interview, she discussed the key insights gained from the NoSQL course and their application in the research. Additionally, she reflected on the impact of continued research efforts and shared her overall experience.

Question 1: How did your NoSQL class prepare you for your research?

Answer: Since I was working with MongoDB & AWS EC2 instances in my research project, the NoSQL class I took laid a good foundation for working with MongoDB and setting up AWS through the assignment. We also heavily focused on the different parts of the database system, such as mongos, config servers, and replica sets in sharding. It was also relevant when I read other pre-existing papers on benchmarking NoSQL systems that did not necessarily use MongoDB for further research or discussed GeoMesa's benchmarking process. It also proved valuable in analyzing the experimental results, which required a thorough knowledge of MongoDB and how queries worked in general for distributed systems.

Question 2: What were some of the key takeaways you gained from your research on NoSQL databases and their practical implementation?

Answer: There are multiple things I've taken away over the course of this research, the most obvious being a more practical understanding of NoSQL databases. The class, by nature, was heavily knowledge-based, which was necessary to learn the terminology to discuss the roles of a system's architecture within only a semester's amount of time. We did have some practice setting up the various databases and writing queries for them. However, I didn't feel it strongly connected with the high-level knowledge we discussed in class until it had consequences, such as directly affecting my test results. For example, it is a recommended practice to run each mongod process on a separate machine to separate their CPU, RAM, and hard disk usage. Still, the effects of this can only be seen when you are stressing your system, which was never a situation in class. Working with big data has its challenges in not only setting up the database system and its shards but also understanding the limitations of the machines holding those shards. I learned even more about the value of diagnostic tools, using the command line, and how to automate with Ansible. More broadly, I feel I've become more confident in my skills and the value of my judgment. At the same time, I've learned there is a lot to learn, and the speed at which technology develops ensures there's no such thing as knowing everything; it's more important to understand how to seek those answers out.

Question 3: What are some of the advantages and challenges you encountered while continuing someone else's research project as an undergraduate student?

Answer: I think continuing research has its pros and cons. The advantages are that there's something to build from and usually someone who can help you get started, like the previous research student. On the other hand, continuing someone else's work may have some issues, such as difficulty understanding the code, bugs, or lack of clarification for something that appeared in the previous research. A portion of your research relies on someone else's, and admittedly, that's a little scary, but I think the challenges are similar to working in any team environment. Despite the cons, I think it was overall more helpful to be continuing research as an undergraduate student. In school, it's an unfamiliar situation to be working on a pre-existing project, and even the process of just reading someone else's code was new and helped me learn a lot. Also, as I had little experience with research papers, it was a huge benefit to read, discuss, and understand something to extend it. Continuing research was a good mix of both practical experience and learning different ways to learn.