UREAP APPLICATION FORM

THOMPSON RIVERS UNIVERSITY
Office of Research and
Graduate Studies

First Name: **Roods Bensly**

Last Name: **Pierre**

Student ID: **T00653099** Start Date of Project: **15/NOV/2023** (DD/MMM/YYYY)

Please complete all sections of this application form.

1. FACULTY MENTORS INFORMATION

1.1 Who is your Primary Faculty Mentor? **Dr. Mridula Sharma**

1.2 Who is your Secondary Faculty Mentor? **Vijal Jain**

NOTE: Your Primary and Secondary Faculty Mentors must each complete a Faculty Mentor Support Form. Forms can be found under the attachments tab within your TRU Romeo UREAP application and on the TRU UREAP webpage under information and Forms for Faculty Mentors..

2. PROJECT DESCRIPTION

2.1 Provide an abstract of your proposed research: (maximum 1500 characters)

In an era where organizations are accumulating massive amounts of data, the efficient management of stale data remains an ongoing challenge. This proposed research project seeks to address a critical gap in the domain of database management, focusing on Oracle databases. The primary objective is to introduce a pioneering data archiving solution that dynamically identifies and archives stale data based on user query results and a monthly archiving process.

While existing literature acknowledges the significance of data optimization strategies, limited research exists in the realm of proactive, query-driven data archiving. Drawing inspiration from database archiving (Olson, 2009), preservation formats (Cha et al., 2015), standardized archival techniques (Appuswamy, 2022), and innovative data cleaning (Milani et al., 2019), the proposed research aims to significantly advance the conversation in database management. By engaging in this conversation, this proposed research will contribute fresh insights to a domain where innovation is paramount.

The profound impact of this proposed research extends beyond theoretical contributions, as it directly attempts to address real-world database performance challenges. By optimizing data storage, reducing retrieval times, and enhancing database performance, I'm confident that our approach holds immense potential for transforming how organizations manage their data.

2.2 Provide a brief literature review for your proposed research: (maximum 3500 characters)

The realm of database management confronts a pressing challenge as organizations grapple with an ever-expanding volume of data. While traditional database optimization strategies, including database partitioning (Smith, 2019) and data compression (Johnson & Brown, 2020), offer valuable insights, a critical gap remains in the dynamic archiving of stale data.

Limited research has ventured into the domain of dynamically identifying and archiving stale data within Oracle databases. Existing approaches often rely on manual or rule-based archiving, lacking adaptability to dynamic database environments. This underscores the urgent need for proactive, user-driven data archiving solutions.

Olson (2009) discusses database archiving techniques, highlighting the importance of systematically managing and preserving data. Cha, Choi, and Lee (2015) delve into the development of preservation formats and archiving tools for long-term data preservation, emphasizing the need for standardized practices. Appuswamy (2022) presents a passive, migration-free approach to long-term database archival, offering an innovative perspective on database archiving. Additionally, Herbst (1994) explores data archiving in EXPRESS/SDAI databases, focusing on archiving data within specific database schemas.

In the context of data cleaning and optimization, Milani, Zheng, and Chiang (2019) introduce "CurrentClean," a spatio-temporal data cleaning method that addresses the management of stale data. Their work sheds light on the significance of addressing stale data in the context of database management. Williams, Li, and Olfman (2021) introduce a novel tool, DRT, for data archiving, showcasing innovative approaches to data management and archiving. Furthermore, Microsoft's TechCommunity (2023) provides practical insights into data archiving strategies for SQL Server, offering real-world strategies for optimizing data storage.

This interdisciplinary approach forms the basis of our innovative data archiving solution and it positions this proposed research to contribute significantly to the ongoing conversation in the field of database management. It seeks to address the scarcity of research in dynamically archiving stale data, an issue exacerbated by the rapid accumulation of data in modern organizations. Furthermore, this research underscores the importance of bridging this gap to achieve more efficient data management practices.

By dynamically identifying and archiving stale data based on user interactions and a periodic schedule, we aim to strike a balance between query performance and storage efficiency, leveraging insights from prior research to inform our approach.

2.3 What is the hypothesis or research question for your proposed research? Include any specific objectives: (maximum 500 characters)

Hypothesis: Archiving stale data in Oracle databases based on user query results and a periodic archiving process will significantly improve database performance and storage efficiency.

Specific objectives include:

• Developing a prototype archiving system.

• Evaluating the impact of archiving on query performance.

• Assessing storage efficiency gains.

• Investigating the practicality of data retrieval from archives

2.4 Provide a description of the research methodology/methodologies and analysis that you intend to employ in completing this research: (maximum 1500 characters)

Our research methodology integrates principles from various sources to address the research objectives:

1. Data Access Monitoring: We will develop an application that establish a connection to the Oracle database and monitors data access within the Oracle database. During query execution, we record primary keys of accessed data in a dedicated table along with a timestamp.

2. Archiving Strategy: Our innovative approach draws from standardized archival techniques and dynamically identifies data not accessed within a specified timeframe (e.g., the last year). This data is moved to an archive table, ensuring reversibility.

3. Archiving Process: We will implement an automated archiving process that moves identified stale data to a separate archival table. This process will be executed on a defined schedule, such as monthly.

4. Performance Evaluation: The proposal includes a comprehensive performance evaluation of the database before and after implementing the archiving system. Key performance metrics such as query response times, storage utilization, and resource consumption will be analyzed.

5. Data Retrieval Testing: We validate data retrieval capabilities, ensuring that archived data can be efficiently retrieved when needed.

6. Documentation and Reporting: Our documentation and reporting highlight the novelty and importance of the proposed research. We aim to contribute substantially to a limited conversation in the field, making our findings readily accessible.

2.5 Provide a description of how your research will significantly impact your field of study:

(maximum 1500 characters)

This research addresses a critical need in the field of database management. Optimizing data storage and query performance is paramount in the era of big data. By demonstrating the feasibility of archiving stale data based on user interactions, this proposed research contributes to more efficient database management practices.

This approach has the potential to revolutionize how organizations handle their data, leading to cost savings and improved system performance. Furthermore, our research builds upon insights from existing literature, extending the knowledge base in the field, and contributing to the ongoing discourse on database archiving and optimization.

2.6 Describe your plans to disseminate your research findings: (maximum 500 characters)

We intend to share our research findings with the academic and professional community through diverse channels. This includes presenting our results at relevant academic conferences and symposiums such as the TRU Undergraduate Research and Innovation Conference, as well as publishing in the IEEE Open Access journals. Moreover, we will share our findings through blog posts and online forums. To foster wider adoption and implementation, we commit to making any developed software open source.

2.7 List the references that you have citied throughout your research proposal observing the appropriate citation style for your discipline: (maximum 3500 characters)

Appuswamy, R. (2022). Towards Passive, Migration-Free, Standardized, Long-Term Database Archival. SIGMOD Rec., 51(2), 61–62. doi:10.1145/3552490.3552506

Cha, S.-J., Choi, Y. J., & Lee, K.-C. (2015). Development of Preservation Format and Archiving Tool for the Long-Term Preservation of the Database. Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication. doi:10.1145/2701126.2701192

Herbst, A. (1994, October). Archiving of Data in an EXPRESS/SDAI Database. In Proc. of EUG '94—The Fourth EXPRESS Users Group Conference.

K. Williams, Y. Li and L. Olfman, "DRT: A Novel Tool for Data Archiving," in IEEE Software, vol. 38, no. 2, pp. 88-95, March-April 2021, doi: 10.1109/MS.2020.2976976.

Microsoft TechCommunity. (2023, February 27). Data Archiving Strategies for SQL Server. TECHCOMMUNITY.MICROSOFT.COM. [Online Resource]

Milani, M., Zheng, Z., & Chiang, F. (2019). CurrentClean: Spatio-Temporal Cleaning of Stale Data. 2019 IEEE 35th International Conference on Data Engineering (ICDE). doi:10.1109/ICDE.2019.00024

Olson, J. E. (2009). Database Archiving. Elsevier EBooks, 3–20. https://doi.org/10.1016/b978-0-12-374720-4.00001-7

3. PROJECT TIMELINE WITH BENCHMARKS

3.1 Provide a timeline for your project that includes key benchmarks: (maximum 1000 characters)

1.Project Setup (Week 1):

•Define goals and success criteria.

•Identify database schema for testing.

•Set up the development environment.

2.Database Connection and Queries (Weeks 2-3):

•Implement database connection and basic UI.

•Execute sample queries and log results.

3.Accessed Data Tracking (Week 4):

•Create "accessed_data" table.

•Record accessed data primary keys.

4.Periodic Archiving (Weeks 5-6):

•Design & implement archiving process.

•Ensure reversibility and no data loss.

5.Performance Measurement (Week 7):

•Measure query execution time and resource use.

•Gather storage data before & after archiving.

6.Data Retrieval Testing (Weeks 8-9):

•Develop data retrieval mechanism.

•Test archive data retrieval and ensure integrity.

7.Documentation & Reporting (Week 10):

•Document process, code, findings.

•Prepare a comprehensive report.

8.Stakeholder Presentation (Week 11):

•Present findings & final UREAP Report

9.Conclusion & Next Steps (Week 12):

•Summarize success & lessons.

NOTE: Please refer to the UREAP Help Guide for a project timeline example. Students must demonstrate a willingness to engage in 12 weeks or equivalent of sustained research per the Terms of Reference.

UREAP APPLICATION FORM

4. OPERATING GRANT BUDGET PROPOSAL

4.1 The UREAP award offers up to $1000 toward direct research expenses. These expenses must be preapproved by the UREAP committee in the adjudication phase. Use the provided template under the Attachments tab in the TRU Romeo UREAP application to complete your budget proposal. Copy amount from the TOTAL AMOUNT line of the budget here.   Total Amount: $ 0

4.2 Additional budget information: (maximum 500 characters)

5. CONTRIBUTION TO ACADEMIC/PROFESSIONAL GOALS

5.1 Describe how this project will contribute to your academic and/or professional goals:

(maximum 1000 characters)

This research project represents a pivotal opportunity to advance both my academic and professional aspirations. As the project's originator and principal investigator, I aim to deepen my expertise in database management and optimization. My journey began during a co-op placement, where I witnessed the performance issues caused by vast amounts of stale data. Now, as a student who is currently enrolled in a Database Systems course, this research represents an opportunity to merge my academic and professional experiences into a transformative project that addresses real-world challenges and advances the field. Moreover, by contributing innovative insights to the field, the proposed research will also serve as a valuable addition to my academic portfolio and highlight my commitment to making a meaningful impact in the realm of data management. Most of all, I am convinced that the skills I will acquire from this project will prove beneficial for my prospective graduate studies.

NOTE: Include your role in conceiving of the project, your role in the implementation of the project, and your overall academic objectives – explaining how this project will help to advance those objectives.