

Incremental Archiving of Database Snapshots for Smarter Use of Historical Data

Summary

With ever-growing data volumes, organisations struggle to maintain and exploit historical states of their operational databases. Current snapshot-based approaches are storage-intensive and hinder agile analytics. This PhD project tackles these challenges by designing an incremental archiving framework that merges successive snapshots into a compact temporal database, preserving full history while keeping storage and query costs low. Project context – The position is part of TalTech's research grant TEM-TA141 "SMARTER USE OF DATA VIA MACHINE#LEARNING" and will directly contribute to its goals by advancing analytics capabilities for incremental archiving of database snapshots for smarter use of historical data. The work will deliver both new algorithmic insights and a production-ready prototype, enabling intuitive time-travel analysis for data-driven decision-making while feeding empirical evidence into the wider TEM-TA141 agenda.

Research field:	Information and communication technology
Supervisor:	Dr. Innar Liiv
Availability:	This position is available.
Offered by:	School of Information Technologies
	Department of Software Science
Application deadline:	Applications are accepted between June 01, 2025 00:00 and June 30, 2025 23:59 (Europe/Zurich)

Description

The research

Archival snapshots are the de facto standard method for retaining database history, yet they duplicate unchanged data and make temporal exploration cumbersome. Temporal DBMSs, on the other hand, do not provide automated ingestion of legacy snapshots.

The overarching goal of this PhD is therefore two-fold:

- 1. Algorithmic innovation: Develop, formalise and prove an algorithm that merges $n\# \rightarrow \#n\#+\#1$ full snapshots into a single-lineage temporal store and can later extract only the changes between time points n and n + 1, producing an incremental snapshot.
- Practical impact: Build an open-source reference implementation on MariaDB (leveraging its bitemporal extensions) that adheres closely to the SQL:2011/2023 temporal standards, paving the way for future cross-platform compatibility

Main research question

Is it possible to do incremental archiving of database snapshots through merging them into an aggregate temporal database?

Goals

- Develop an algorithm for incremental archiving, incl. merging the snapshots into a temporal database and then extracting the latest increment
- Enhance the algorithm to support various temporal database platforms, then measure and optimize its performance
- Construct a set of temporal analytical queries that closely resemble traditional ones, making time-based exploration
 intuitive

Responsibilities and (foreseen) tasks

• Design, implement and mathematically analyse the incremental archiving algorithm



- Develop and harden a MariaDB-based proof-of-concept, while documenting standard-compliant interfaces for other temporal DBMSs
- Create a benchmark suite of realistic archival workloads and evaluate performance, scalability and cost
- Define a library of temporal analytical queries mirroring common business reports
- Collaborate with researchers in TEM-TA141 WP#2 & WP#4 and research the practical applications of incremental archiving of database snapshots for smarter use of historical data
- Publish results in peer-reviewed journals or at international conferences
- Contribute to project workshops and collaborate with industry or public sector organizations using snapshot data sets relevant to this research

Applicants should fulfil the following requirements:

- master's degree (or equivalent) in Computer Science, Software Engineering, Information Systems or a related field
- · solid knowledge of relational databases, their standards and practical applications
- high proficiency in SQL and preferably in at least one scripting language (e.g. Python)
- · a clear interest in the topic of the position
- excellent command of English
- strong and demonstrable writing and analytical skills
- capacity to work both as an independent researcher and as part of an international team
- · capacity and willingness to provide assistance in organizational tasks relevant to the project

(The following experience is beneficial:)

- Hands-on work with temporal extensions in PostgreSQL or MariaDB (SQL 2011 or SQL 2023 system-versioned tables, bitemporal queries).
- Experience with SIARD Suite, DBPTK, or similar database-preservation toolchains.
- Ability to write and maintain Bash scripts, and SQL stored procedures for data migration and merge tasks.
- Familiarity with change-data-capture pipelines, large-volume snapshot handling, and performance tuning on multi-terabyte datasets.

The candidate should submit a research plan for the topic, including the overall research and data collection strategy. The candidate can expand on the listed research questions and tasks, and propose theoretical lenses to be used. *We offer:*

- A 4-year fully funded PhD position.
- Integration into the TEM-TA141 research team, offering cross-disciplinary mentorship in databases and analytics.
- · Access to high-performance computing resources.
- A budget for presenting at leading international conferences.
- Professional development courses, doctoral-school seminars, and support for research commercialization.

About the department

The ambition of the Department of Software Science is to be a leading actor in software science research in the Baltic Sea region and an intermediary of top level and scientifically relevant competence between students, enterprises, public sector and researchers.

Department of Software Science is part of School of Information Technologies, which prepares specialists with bachelor's, master's and doctoral degrees in one of the fastest developing fields of science and technology, which is information and communication technology. Research and development activities at a good international level and cooperation with companies create the basis for high-quality research-based learning activities.

(Additional information)



For further information, please contact Associate Professor Innar Liiv (innar.liiv@taltech.ee) and Associate Professor Erki Eessaar (erki.eessaar@taltech.ee) https://taltech.ee/en/department-of-software-science



To get more information or to apply online, visit https://taltech.glowbase.com/positions/932 or scan the the code on the left with your smartphone.