EXAMINE AND COMPARE DATABASE TECHNOLOGIES TO TIME SERIES DATA

Yosafe Fesaha Oqbamecail

Master's Thesis in Applied Computer Science and Engineering

Engineering Data Science Specialisation

Faculty of Technology, Environmental and Social Sciences
Western Norway University of Applied Sciences

28th August 2025

ABSTRACT

This is where you write the abstract.

ACKNOWLEDGMENTS

I would like to thank . . .

Contents

CHAPTER 1

INTRODUCTION

The seas and oceans play a key role in the operation of global ecosystems, climate regulation, food security, and energy production [3]. Monitoring of such large and complex environments requires advanced technology. Among these, underwater sensor networks (UWSNs) have become valuable resources for continuous real-time data collection in marine ecosystems, subsea structures, and oceanic processes [4].

The SFI Smart Ocean project is designed to create an autonomous and flexible wireless marine observation program. The system will allow for large-scale, long-term monitoring of underwater spaces as well as installations by combining UWSNs and cloud-based big data solutions. The infrastructure is intended to deal with multi-parameter observations, ensuring reliable data for both scientific research as well as industrial applications. [7, 8].

Limitations of acoustic communication, long-range UWSN deployments are impacted by environmental and technical challenges. These consist of power restrictions, minimal data rates and the inability to recalibrate once deployed at depth. Sensors are subjected to extreme conditions. Data quality can be affected by pressures, biofouling, corrosion, along with electronic drift with time. Moreover, environmental measurements are affected by parameters such as temperature and salinity can introduce additional uncertainty. [10]. Once data reaches surface nodes, there is a need for efficient data handling. The storage infrastructure needs to be able to deal with large volumes of inbound data and maintain longterm availability.

The thesis investigates techniques for managing time-series data produced by the Smart Ocean sensor network. The emphasis is on evaluating the development of database technologies that can deal with large-scale storage and effective querying of time-stamped observations. The work contributes to the wider picture of the Smart Ocean which aims to promote data-driven decision making in marine operations. This research aligns with the ACM Computing Classification System (CCS) under Information systems → Data management systems → Database administration → Database performance evaluation [1]. The study focuses on evaluating and comparing time-series database technologies to improve the efficiency and reliability of underwater sensor data management.

1.1 Context and Motivation

1.1.1 *The SFI Smart Ocean Platform*

The Norwegian Research Council has provided funding for the SFI Smart Ocean project. Research institutes as well as industry partners work together on key issues in building smart ocean systems [9]. The initiative concentrates on 3 main areas:

1. **Underwater Sensor and Measurement Technology**: Focuses on developing autonomous sensors and methods for real-time monitoring of underwater environments. These sensors include features like data collection, acoustic communication, and energy-efficient operations.
2. **Underwater Wireless Sensor Networks Based on Acoustic Communication**: Addresses the establishment of reliable communication networks for data transmission from underwater sensors.
3. **Smart Ocean Platform for Cloud-Based Data and Application Services**: Consists of developing the Smart Ocean platform to incorporate, process, and visualize ocean information. The platform uses Standard APIs & efficiently formats to deal with data from diverse underwater sensors.

1.1.2 Motivation

The deployment of Smart Ocean's sensor network will generate a large volume of time series data. The underwater sensors continuously measure parameters over extended periods, resulting in a continuous stream of timestamped readings. The data volume rapidly reaches big-data scales with dozens or possibly hundreds of sensors reporting in real time (each measuring several variables). The system needs to store years of data reliably, deal with high-frequency ingestion, and support queries for real-time alerts and long-term analysis a major challenge.

The selection of suitable data management technologies is crucial given these requirements. Standard relational databases, although robust, are not specifically designed to deal with high-volume time-series workloads where data arrives sequentially and is mainly queried by time. In recent years, specialized Time-Series Management Systems (TSMSs) have emerged in response to the high-volume, high-velocity nature of data generated by IoT devices and sensors. Unlike general-purpose databases, TSMSs are architected to efficiently store, query, and process time-stamped data streams, making them uniquely suited for real-time sensor workloads [6]. This systems are designed to efficiently ingest and index time-stamped data and provide built-in functions for time-window queries, downsampling, and time-centric analytics. Dedicated TSDBs are increasingly considered a natural fit for fast sensor data streams, yet with a wide variety of database options available, it is not obvious which technology is most suitable for the Smart Ocean platform.

1.2 Problem Description and Research Questions

the best way to efficiently store and manage the large amounts of time-series data generated by the Smart Ocean platform in the above mentioned context is to determine what database technology (or combination of technologies) best meets the platform's demands for scalability, functionality as well as dependability in dealing with underwater sensor data. It involves comparing various approaches. For example traditional relational databases with time-series extensions versus purpose-built time-series databases to figure out their relative advantages and disadvantages for Smart Ocean's use case.

The investigation is guided by the following research question and sub-questions:

- **Main Research Question:** *Which database technology is most suitable for managing large-scale time-series data*
- **RQ1:** What are the key requirements and challenges in managing sensor data (e.g., data volume, frequency of data, query patterns, real-time access needs, and deployment constraints)?
- **RQ2:** Which existing database systems (relational, NoSQL, and dedicated time-series databases) are potential candidates for this task, and what are their expected advantages or limitations in the Smart Ocean context?
- **RQ3:** How do selected candidate databases perform ingestion rate, query performance, scalability, storage efficiency, and other relevant metrics?

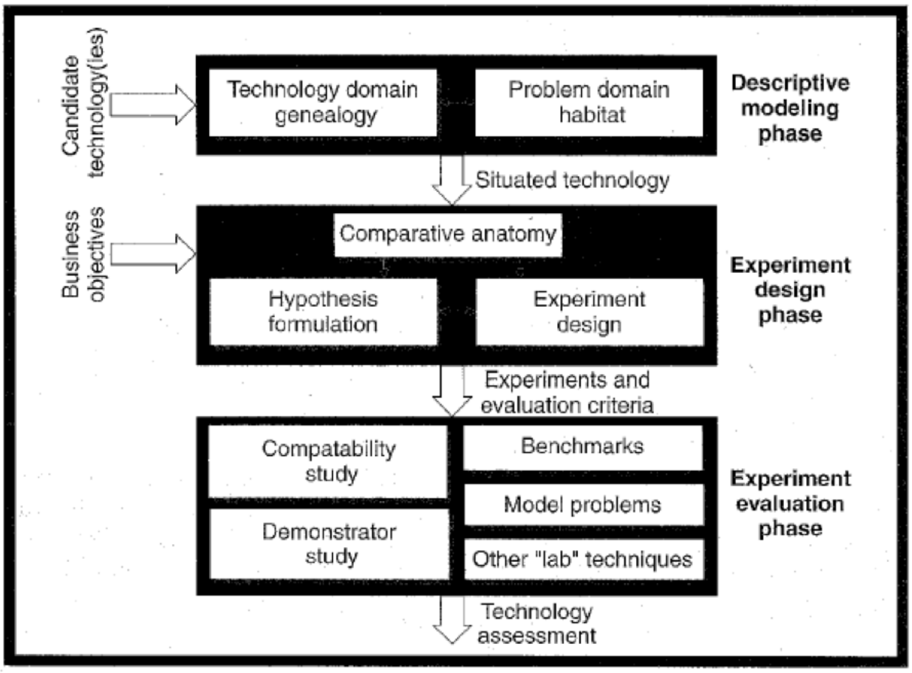

Fig. 1.1: Software technology evaluation framework.

1.3 Research Method

To answer the research questions, this thesis applies the evaluation methodology proposed by Brown and Wallnau [2]. Their framework provides a structured way to assess software technologies through the definition of measurable benchmarks or technology deltas, that highlight how one technology performs relative to another. The methodology is organized into three phases: descriptive modeling, experiment design, and experiment evaluation. Figure 1.1 illustrates the overall structure of this framework.

The descriptive modeling phase consists of identifying the relevant features. This involves determining which requirements are most relevant for evaluation. These requirements then form the basis for defining performance benchmarks such as ingestion throughput, query latency, scalability, and storage efficiency.

The experiment design phase serves as the planning stage of the evaluation. In this phase, hypotheses are formulated about how different database technologies are expected to behave under representative workloads. Workload models and test harnesses are developed to simulate large-scale time-series data, ensuring that the experiments are both realistic and repeatable.

Finally, in the experiment evaluation phase, the selected databases are tested using the defined benchmarks. The resulting measurements are collected, analyzed, and compared to highlight the deltas between technologies. This process enables a systematic assessment of strengths, weaknesses, and trade-offs, ultimately guiding conclusions about which database technologies are most suitable for handling time-series data at scale.

1.4 Thesis Outline

The remainder of this thesis is organized as follows:

- **Chapter 2 – Background and Related Work:** Provides the necessary background and context. Reviews time-series data characteristics and challenges, and surveys existing database technologies for time-series data management. Discusses related work and prior evaluations of database performance in IoT and sensor data scenarios.
- **Chapter 3 – Design and Analysis:** Outlines the design of the experimental evaluation, including the selection of database technologies, the design of test workloads, and the metrics used for assessment.
- **Chapter 4 – Implementation and Prototypes:** Describes the implementation of the experimental setup, including the deployment of database systems and the creation of test environments. Details the development of prototypes or models used for evaluation.
- **Chapter 5 – Evaluation and Results:** Presents the empirical results of the comparative evaluation of database technologies. Includes ingestion throughput, query performance, storage efficiency, and other findings.
- **Chapter 6 – Conclusion and Future Work:** Summarizes key findings and contributions, provides recommendations for the Smart Ocean platform, and suggests directions for future work.

CHAPTER 2

BACKGROUND AND RELATED WORK

This is another chapter. . .

CHAPTER 3

DESIGN AND ANALYSIS

CHAPTER 4

IMPLEMENTATION AND PROTOTYPES

```java
public static void main(String[] args) {

  int b, h, d;
  String btext, htext, dtext;

  [ ... ]

  int volum = b * h * d;

  String respons =
      "Volum [" + htext + "," + btext + "," + dtext + "] = " + volum;

}
```

CHAPTER 5

EVALUATION AND RESULTS

Table 5.1 gives an example of how to create a table.

Config	Property	States	Edges	Peak	E-Time	C-Time	T-Time
22-2	A	7,944	22,419	6.6 %	7 ms	42.9%	485.7%
22-2	A	7,944	22,419	6.6 %	7 ms	42.9%	471.4%
30-2	B	14,672	41,611	4.9 %	14 ms	42.9%	464.3%
30-2	C	14,672	41,611	4.9 %	15 ms	40.0%	420.0%
10-3	D	24,052	98,671	19.8 %	35 ms	31.4%	285.7%
10-3	E	24,052	98,671	19.8 %	35 ms	34.3%	308.6%

Table 5.1: Selected experimental results on the communication protocol example.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

BIBLIOGRAPHY

[1] ACM computing classification system. ACM Digital Library website, 2012. [Data set; ontology]. Retrieved from https://dl.acm.org/ccs, accessed 2025-08-28. 1

[2] A. W. Brown and K. C. Wallnau. A framework for evaluating software technology. *IEEE Software*, 13(5):39–49, 1996. Available at: https://doi.org/10.1109/52.536457 (accessed: 26.08.2025). 1.3

[3] Copernicus Marine Service. Why ocean health matters for food security and human health. Available at: https://marine.copernicus.eu/explainers/why-ocean-important/food-security. (accessed: 27.08.2025). 1

[4] F. P. F. Domingos et al. Underwater communication systems and their impact on aquatic ecosystems. *Electronics*, 14(1):7, 2024. Available at: https://www.mdpi.com/2079-9292/14/1/7 (accessed: 27.08.2025). 1

[5] P. Grzesik and D. Mrozek. Comparative analysis of time series databases in the context of edge computing for low power sensor networks. In *Computational Science - ICCS 2020: 20th International Conference, Proceedings, Part V*, volume 12141 of *Lecture Notes in Computer Science*, pages 371–383. Springer, 2020. Available at: https://link.springer.com/chapter/10.1007/978-3-030-50426-7_28 (accessed: 26.08.2025).

[6] S. K. Jensen, T. B. Pedersen, and C. Thomsen. Time series management systems: A survey. arXiv preprint arXiv:1710.01077. Available at: https://arxiv.org/abs/1710.01077 (accessed: 28.08.2025). 1.1.2

[7] Research Council of Norway. SFI Smart Ocean - prosjektbanken. https://prosjektbanken.forskningsradet.no/project/FORISS/309612, 2020. (accessed: 26.08.2025). 1

[8] SFI Smart Ocean. About sfi smart ocean. https://sfismartocean.no/about/. (accessed: 26.08.2025). 1

[9] SFI Smart Ocean. Sfi smart ocean hub. https://smartoceanplatform.github.io/. (accessed: 26.08.2025). 1.1.1

[10] A. M. Skålvik, C. Sætre, K.-E. Frøysa, R. N. Bjørk, and A. Tengberg. Challenges, limitations, and measurement strategies to ensure data quality in deep-sea sensors. *Frontiers in Marine Science*, 10, 2023. Available at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1152236/full (accessed: 27.08.2025). 1

APPENDIX A

SOURCE CODE

Where to find the source code . . . (if applicable)

APPENDIX B

RESEARCH DATA

Where to find the research data . . . (if applicable)

List of Figures

List of Tables