Universidade de Brasília
Faculdade de Tecnologia
Departamento de Engenharia Elétrica

Efficient Query Planning for Sensor Time-Series Databases

Camila Duarte

Master's Dissertation submitted to the Graduate Program in Electrical Engineering of the University of Brasília as part of the requirements for the degree of Master.

Advisor: Prof. Dr. Ricardo Almeida

Brasília
2026

Universidade de Brasília
Faculdade de Tecnologia
Departamento de Engenharia Elétrica

Efficient Query Planning for Sensor Time-Series Databases

Camila Duarte

Master's Dissertation submitted to the Graduate Program in Electrical Engineering of the University of Brasília as part of the requirements for the degree of Master.

Trabalho aprovado. Brasília, 20 de maio de 2026:

Prof. Dr. Ricardo Almeida
UnB/FT/ENE
Advisor

Prof. Dr. Fernanda Lima
UnB/FT/ENE
Internal examiner

Prof. Dr. Tiago Novaes
UnB/CIC
External examiner

To everyone who ever waited patiently for a query to finish.

Acknowledgements

I thank my advisor for the guidance and patience shown throughout this project, and my committee members for their careful reading of this work. Thanks also to the members of the Distributed Systems Lab for many helpful discussions, and to the university's High-Performance Computing center for providing access to the compute resources used in the evaluation.

Abstract

Sensor time-series databases grow rapidly in volume, which makes efficient query planning essential for real-time monitoring applications. This dissertation proposes a query planning strategy that exploits the temporal regularity of such data to reduce scan cost, evaluated on three large-scale environmental sensing datasets. The proposed planner reduces average query latency substantially relative to a generic range-scan baseline, with the largest gains on queries that aggregate across many sensors over long time windows.

Keywords: Database; Time series; Query planning; Sensing.

Resumo

Bancos de dados de séries temporais de sensores crescem rapidamente em volume, o que torna o planejamento eficiente de consultas essencial para aplicações de monitoramento em tempo real. Esta dissertação propõe uma estratégia de planejamento de consultas que explora a regularidade temporal desses dados para reduzir o custo de varredura, avaliada sobre três conjuntos de dados de sensoriamento ambiental de grande escala.

Palavras-chave: Banco de dados; Séries temporais; Planejamento de consultas; Sensoriamento.

List of figures

List of tables

Contents

1 INTRODUCTION 10
1.1 Motivation 10
1.2 Contributions 10
1.3 Dissertation Outline 10

2 RELATED WORK 11
2.1 Time-Series Storage Engines 11
2.2 Query Planning over Temporal Data 11
2.3 Gap Addressed by This Dissertation 11

REFERENCES 12

1 Introduction

Sensor deployments for environmental and infrastructure monitoring routinely produce time-series data at a scale that makes naive range-scan query execution impractical. A query asking for the average temperature across a thousand sensors over the last month may need to scan billions of individual readings if the underlying storage engine has no way to exploit the temporal and sensor-wise structure of the data (Jensen; Lin; Ooi, 2017).

1.1 Motivation

Existing time-series database systems typically rely on generic indexing structures borrowed from relational systems, which do not exploit the fact that sensor readings arrive at regular intervals and are naturally partitioned by sensor identity and time window (Pelkonen *et al.*, 2015).

1.2 Contributions

This dissertation makes the following contributions:

- A query planning strategy that partitions scans along both the sensor and time dimensions, exploiting known sampling regularity to prune irrelevant partitions before execution.
- An implementation evaluated against three large-scale environmental sensing datasets with heterogeneous sampling rates.
- An analysis of when the proposed planner's gains are largest, showing the biggest improvements on queries aggregating across many sensors over long time windows.

1.3 Dissertation Outline

Chapter 2 reviews prior work on time-series database systems and query planning. Later chapters (omitted from this starter document) would present the planning algorithm, implementation, and experimental results in detail.

2 Related Work

2.1 Time-Series Storage Engines

Purpose-built time-series databases such as Gorilla store data in compressed, append-only blocks keyed by sensor and time window, which improves both write throughput and compression ratio relative to general-purpose storage engines (Pelkonen *et al.*, 2015).

2.2 Query Planning over Temporal Data

Query optimization for temporal data has a long history in the database literature, though most classical work targets transactional workloads rather than the high-cardinality, append-mostly workloads typical of sensor telemetry (Jensen; Lin; Ooi, 2017).

2.3 Gap Addressed by This Dissertation

Prior time-series query planners generally treat sensor identity and time as independent dimensions to be indexed separately. This dissertation explores planning jointly over both dimensions when sampling rate is known in advance.

References

JENSEN, C. S.; LIN, D.; OOI, B. C. Query and update efficient b+-tree based indexing of moving objects. **Proceedings of the 30th International Conference on Very Large Data Bases**, p. 768–779, 2017. Cit. on pp. 10 and 11.

PELKONEN, T.; FRANKLIN, S.; TELLER, J.; CAVALLARO, P.; HUANG, Q.; MEZA, J.; VEERARAGHAVAN, K. Gorilla: A fast, scalable, in-memory time series database. *In*: **Proceedings of the VLDB Endowment**. [*S.l.*: *s.n.*], 2015. v. 8, n. 12, p. 1816–1827. Cit. on pp. 10 and 11.