

Extreme-Scale Model-Based Time Series Management with ModelarDB

Torben Bach Pedersen   

Department of Computer Science, Center for Data-intensive Systems, Aalborg University, Denmark
ModelarData, Copenhagen, Denmark

Abstract

To monitor critical industrial devices such as wind turbines, high quality sensors sampled at a high frequency are increasingly used. Current technology does not handle these extreme-scale time series well [1], so only simple aggregates are traditionally stored, removing outliers and fluctuations that could indicate problems. As a remedy, we present a model-based approach for managing extreme-scale time series that approximates the time series values using mathematical functions (models) and stores only model coefficients rather than data values. Compression is done both for individual time series and for correlated groups of time series. The keynote will present concepts, techniques, and algorithms from model-based time series management and our implementation of these in the open source Time Series Management System (TSMS) ModelarDB[2, 3, 4]¹. Furthermore, it will present our experimental evaluation of ModelarDB on extreme-scale real-world time series, which shows that that compared to widely used Big Data formats, ModelarDB provides up to 14× faster ingestion due to high compression, 113× better compression due to its adaptability, 573× faster aggregation by using models, and close to linear scale-out scalability. ModelarDB is being commercialized by the spin-out company ModelarData².

2012 ACM Subject Classification Information systems → Data management systems

Keywords and phrases Model-based storage, approximate query processing, time series management, extreme-scale data

Digital Object Identifier 10.4230/LIPIcs.TIME.2021.2

Category Invited Talk

Funding *Torben Bach Pedersen*: This work has been funded by Innovation Fund Denmark (the DiCyPS project), Independent Research Fund Denmark (the SEMIOTIC project) and Horizon 2020 (the MORE project).

Acknowledgements I want to thank Søren Kejser Jensen and Christian Thomsen for their major contributions to this work.

References

- 1 Søren Kejser Jensen, Torben Bach Pedersen, and Christian Thomsen. Time series management systems: A survey. *IEEE Trans. Knowl. Data Eng.*, 29(11):2581–2600, 2017. doi:10.1109/TKDE.2017.2740932.
- 2 Søren Kejser Jensen, Torben Bach Pedersen, and Christian Thomsen. Modelardb: Modular model-based time series management with spark and cassandra. *Proc. VLDB Endow.*, 11(11):1688–1701, 2018. doi:10.14778/3236187.3236215.
- 3 Søren Kejser Jensen, Torben Bach Pedersen, and Christian Thomsen. Demonstration of modelardb: Model-based management of dimensional time series. In Peter A. Boncz, Stefan Manegold, Anastasia Ailamaki, Amol Deshpande, and Tim Kraska, editors, *Proceedings of the 2019 International Conference on Management of Data, SIGMOD Conference 2019, Amsterdam, The Netherlands, June 30 – July 5, 2019*, pages 1933–1936. ACM, 2019. doi:10.1145/3299869.3320216.

¹ <https://github.com/skejserjensen/ModelarDB>

² <https://modelardata.com>



© Torben Bach Pedersen;

licensed under Creative Commons License CC-BY 4.0

28th International Symposium on Temporal Representation and Reasoning (TIME 2021).

Editors: Carlo Combi, Johann Eder, and Mark Reynolds; Article No. 2; pp. 2:1–2:2

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

2:2 Model-Based Time Series Management with ModelarDB

- 4 Søren Kejser Jensen, Torben Bach Pedersen, and Christian Thomsen. Scalable model-based management of correlated dimensional time series in modelardb+. In *37th IEEE International Conference on Data Engineering, ICDE 2021, Chania, Greece, April 19–22, 2021*, pages 1380–1391. IEEE, 2021. doi:10.1109/ICDE51399.2021.00123.