

# M4: A Visualization-Oriented Time Series Data Aggregation

Uwe Jugel, Zbigniew Jerzak,  
Gregor Hackenbroich  
SAP AG  
Chemnitzer Str. 48, 01187 Dresden, Germany

Volker Markl  
Technische Universität Berlin  
Straße des 17. Juni 135  
10623 Berlin, Germany  
volker.markl@tu-berlin.de

## I. INTRODUCTION

Visualization of large scale time series data is a crucial need of modern exploratory bigdata analysis [1]. But the huge size of the data is a barrier to visualization [2], [3], [4]. To address this challenge of bigdata different data reduction and sampling strategies are used to overcome the barrier [5], [6]. But for preserving the semantics of trend line of time series data these sampling strategies show huge limitations [7].

In this review paper we present a review of the paper [7] which address this issue of preserving the semantic of time series data and present some related works in the line. The paper appeared in the Proceedings of the VLDB Endowment, 2014.

The authors present M4, an aggregation based time series data reduction strategy that guarantees error free visualization of time series data as line chart as well as higher rate of data reduction. The approach is generic to any visualization system as long as the visualization systems uses RDBMS as data source.

## II. CONTRIBUTIONS OF THE PAPER

The authors of the paper rewrite visualization queries  $Q$  using data reduction operator  $M_R$  such that the visualization of the original data from query  $Q$  and the visualization from the query  $Q_R = M_R(Q)$  are similar and error free. As

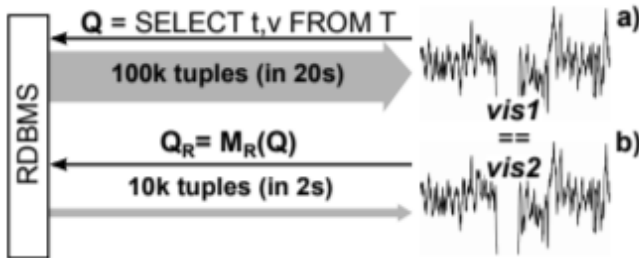


Figure 1. Time series visualization: a) based on original data; b) Using data reduction operator;

shown in Figure 1  $Q_R$  produced the same visualization as  $Q$  with almost 10 times less tuples and 10 times reduced time.

## A. Query Rewriting

TODO Query rewriting

## B. Time Series Visualization

TODO

## C. Data Reduction Operators

TODO

## D. Time Series Data Reduction

## III. EVALUATION

## IV. RELATED WORKS

TODO

## V. OUR PROPOSAL

TODO

## VI. CONCLUSION

The conclusion goes here. this is more of the conclusion

## REFERENCES

- [1] T.-c. Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181, 2011.
- [2] A. Labrinidis and H. V. Jagadish, "Challenges and opportunities with big data," *Proceedings of the VLDB Endowment*, vol. 5, no. 12, pp. 2032–2033, 2012.
- [3] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," *National science review*, vol. 1, no. 2, pp. 293–314, 2014.
- [4] C. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Information Sciences*, vol. 275, pp. 314–347, 2014.
- [5] G. Cormode and N. Duffield, "Sampling for big data," 2014.
- [6] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE transactions on knowledge and data engineering*, vol. 26, no. 1, pp. 97–107, 2014.
- [7] U. Jugel, Z. Jerzak, G. Hackenbroich, and V. Markl, "M4: a visualization-oriented time series data aggregation," *Proceedings of the VLDB Endowment*, vol. 7, no. 10, pp. 797–808, 2014.