

Visualization-Driven Data Aggregation

Uwe Jugel¹ and Zbigniew Jerzak²

¹ LOVOO GmbH, Dresden, Germany

`uwe.jugel@lovo.com`

² SAP SE, Berlin, Germany

`zbigniew.jerzak@sap.com`

Visual analysis of high-volume, numerical data is traditionally required for understanding sensor data in manufacturing and engineering scenarios. However, today the visual analysis of any kind of big data has become ubiquitous and is a most-wanted feature for visual analytics tools, required for commerce, finance, sales, and an ever-growing number of industries, whose data is prevalently stored in a relational database management system (RDBMS).

Unfortunately, contemporary RDBMS-based data visualization and analysis systems have difficulties to cope with the hard latency requirements and high ingestion rates required for interactive visualizations of big data. These systems are particularly not able to effectively sample or aggregate the data, inevitably failing to visualize the millions of acquired records. A general-purpose solution for visualization-related data reduction in RDBMS-based systems has been missing. Thereby, custom solutions are tailored to specific domains, supporting only a few custom types of visualizations, while general approaches to data reduction disregard the spatial properties of data visualizations, resulting in measurable and perceivable visualization errors.

To facilitate truly interactive visualizations of the growing volume and variety of big data, computer systems need to change the way they acquire data for the purpose of data visualizations. Visualization-Driven Data Aggregation (VDDA) facilitates up to error-free visualizations of high-volume data sets, at high data reduction rates. Defined as data reduction at the query level and leveraged in a transparent query rewriting system, VDDA is applicable to any visualization system that consumes data stored in relational databases.

This is a resubmission of previously published papers by Jugel et al. [1,2].

Keywords: data visualization, data aggregation, relational databases, query processing, sampling methods

References

1. U. Jugel, Z. Jerzak, G. Hackenbroich, and V. Markl. M4: A visualization-oriented time series data aggregation. *PVLDB*, 7(10):797–808, 2014. (VLDB Best Paper).
2. U. Jugel, Z. Jerzak, G. Hackenbroich, and V. Markl. VDDA: Automatic visualization-driven data aggregation in relational databases. *The VLDB Journal*, 25(1):53–77, February 2016.