Visualizing Database-Performance Through Shape, Reflecting the Development Opportunities of Radar Charts

Verena Lechner and Karl-Heinz Weidmann[(✉)]

User Centered Technologies Research Institute, University of Applied Sciences Vorarlberg,
Achstrasse 1, 6850 Dornbirn, Austria
uct@fhv.at

Abstract. At a time where databases contain millions of data records and their organization gets complex, the visualization especially of metadata gets a necessity to get an overview of the database performance. In this paper we'll provide an insight into a research project commissioned by Crate Technology GmbH, who developed an elastic SQL Data Store that is massively scalable [1]. The aim of our efforts in the UCT Research Institute is to find different ways to visualize the data state in such a cluster. In particular we investigated the development opportunities of radar charts in database metadata visualization and the visual appearance of the developing shapes. During the paper, the primary challenges of the project will be displayed and comparable products will be investigated. In the last part, we'll give a short insight to our work in progress, deal with the issue of form perception and also present further required efforts.

1 Introduction

When thinking of visualizing database-performance today, our thoughts are most likely with a large amount of meta data records, which are updated regularly. To keep track of these meta data seems to be difficult. Moreover, upcoming problems, which a bad condition of the database may trigger in future seem to be invisible. But especially for employees dealing with big data and being responsible for the optimal functional capability of databases, finding the sources of problems in advance in order to counteract could be absolutely helpful. Therefore providing an easily comprehensible overview about the real-time condition of databases could be a particularly forward-looking development. In this paper we will propose the use of radar charts and the resulting shapes as a possible solution and investigate the potential of these shapes.

2 Definition of Metadata

Before we can start with the description of the initial situation of the project, it's necessary to define the term "metadata" because there exist innumerable definitions [2] that could have a confusing effect.

Metadata can be used in different matters and targets, but mostly have the following functionalities: supporting data discovery, enabling the dealing of data by humans and

facilitating the automated handling of data in the subdivisions of data discovery, ingestion, processing and analysis [3].

Moreover, in all metadata schemas semantics can be found as a characteristic. So they give data elements a meaning and help people working with metadata to understand the individual elements [4].

The term "metadata" comes from the field of computer science, where the prefix "meta" means "about", so the aim of metadata is to describe other data. Two main requirements that metadata have to fulfill can be defined: firstly the information has to be structured and therefore recorded according to a documented metadata approach. Secondly a resource of information has to be described by the metadata, where the question will arise what exactly could be an information resource, especially when we think of the numerous fields metadata play a key role [5].

Looking at database metadata in a more precise sense, it can be defined as "data about database data" as for example a list of all tables of a database, their sizes and their number of rows. So metadata (also seen more generally) has the task to describe data, but it isn't the data itself [6].

After having defined metadata in particular for our purpose of database-performance visualization, some general challenges in this field as well as some of the main challenges especially in the context of the research project for Crate will be described.

3 Challenges of Visualizing Database-Performance in a Big Data Environment

When working in a big data environment probably the most obvious fact may be that either the data collections or the data objects itself are big [7]. Regarding our research task, big data collections but as well the wide variability of users and therefore the variability of the size of their data collections represent a challenge for finding ways to visualize the data state. Moreover, the visualization needs to take into account the fact that some databases are under constant transformation, therefore the metadata change constantly as well, depending at which intervals the metadata are queried.

A Cluster running on Crate contains a certain amount of nodes, defined by the user, which according to the experience values of Crate Technology GmbH can be assumed with a number between 8 and 300 nodes. As a result, when starting with the visualization work, the capability of the visual sense has to be kept in mind and a presentation method has to be found which should communicate the content precisely and let it be perceptible in an intuitive level [8]. So in our specific task situation where the main goal is to give an overview over the condition of the whole amount of nodes (putting aside the second goal for the time being, which is to provide further information about selected individual nodes) we are talking not only about communicating the condition of up to approximately 300 nodes, but also about taking into account the simultaneous view of different points in time. So the number of objects to be detected can rise sharply. And as the user of the visualization should be able to make decisions about manual optimization measures, any incorrect representation and any resulting wrong interpretation may have far-reaching consequences.

Moreover, as the database-performance can be assessed according to many different criteria, there has to be found a way to group them to logical categories that answer upcoming questions of the employees being responsible for their databases. As a result, the number of necessary parameters for answering these questions can vary widely. So when trying to find a way to visualize them, a very flexible form concerning the representable parameters has to be found.

Summing up what has been said so far, the main challenges for projects working with the visualization of database-performance in a big data environment are:

- the enormous amount of data objects itself,
- the constant transformation of the database,
- the variety of the number of nodes that have to be presented and perceived intuitively,
- the variety of criteria that determine the performance and therefore the visualization of answers to upcoming questions that consist of a different amount of parameters.

As the baseline situation and the major challenges of such projects has been described, the following section takes a closer look on other tools and products that try to visualize database performance or other comparable data.

4 Analysis of Comparable Tools and Products

4.1 Planning of the Comparison Procedures

For the analysis of tools that visualize comparable issues as Crate does, eight products on the market have been chosen and investigated. The main challenges but also characteristics of these will be discussed and examples will be given in order to illustrate the development potential for our research project.

4.2 Limited Number of Parameters Shown per Visualization

As already mentioned previously, database performance depends on various parameters, therefore when trying to visualize the condition of a database cluster, many different criteria have to be taken into account. When examining for example the user interfaces of Zabbix [9], cacti [10], Ganglia [11], NuoDB [12] or Pivotal HD [13], it can be noted that in one single visualization they mostly use only two dimensions when producing bar charts or line charts. In order to be able to present more than these few dimensions, they add more charts or other visualization types, resulting in a dashboard-like collection of different views. As a result, the situation for the user could be confusing, because associated dimensions or criteria aren't displayed in the same visualization.

4.3 Necessary Prior Knowledge for the Usage of Modular Systems

Kibana, for instance, allows the user to adapt its queries and also the visualization types and attributes in order to offer a totally personalized visualization [14]. But this assumes that the user is able to combine different parameters in a way that the output makes sense

and decisions can be made. As in our research project Crate Technology GmbH requests a solution, that not only offers an optimal integration into existing company processes but also a maximum of the ease of use, we decided to specify the research questions and therefore the shown parameters the user could be interested in when observing the cluster state.

4.4   Difficult Determinability of the Condition of the Elements Shown

Whenever producing a visualization of parameters that affect the condition of elements in order to make human decisions, in the end it lies in the hands of the observer to determine whether the viewed visualization contains negative or improvable content. With the help of human intuition, the viewer should get a contextual understanding of the data condition, so it is necessary to offer an intuitive human-machine interface to integrate intuition into the analysis process [15].

In the example of software diagnostics [16], a spatial model was found to present answers to different questions concerning software systems. Individual code units are represented as blocks ordered in a way that intuitively let the user think of a city view consisting of skyscrapers with different heights and base areas. The system detects bad code units, lets them appear in red and therefore give hints to possible improvement potential. In comparison to software diagnostics, MongoDB MMS [17] (a cloud service that can be used to support the usage of MongoDB) offers more possibilities to specify individual queries and shows the selected metadata of a database as a chart, but the system doesn't interpret the chart nor gives a hint to a dangerous state of data. So the user has to decide on his own whether the presented data could contain a problematic issue or not, and this could develop difficulties when thinking of visualizations of really large metadata sets. For that reason, what we have been looking for in our project is a visualization, that lets the user get a quick and intuitive impression of the database performance and highlight problematic issues in any way to facilitate the improvement work.

So with these three challenges the main potential of our visualization development has been formulated. This leads us to a detailed presentation of our approach using radar charts and the shapes developed through an intuitive handling of radar charts.

5   Using Radar Charts as a Form Generator and Intuitive Observation Tool

5.1   General Information to the Work with Shapes in this Project

As already described earlier, a Cluster running on Crate consists of a variable amount of single nodes, the state of which has to be monitored. Just as to mention a single example, overloaded servers could cause unintended waiting periods, so the parameters "average load", "used disk", "used heap" and "number of shards" could be interesting when observing all the nodes.

What we started to do in the beginning of the research project is to find a way to give an overview over all nodes that is flexible regarding the number of parameters that have

to be taken into account, shows the user all the problematic information in an intuitive way and doesn't require much specialist knowledge to work with. For this reason we combined the flexible properties of radar charts with findings of form perception and developed an opportunity to show various values of different parameters and let them build a shape, that deviates more or less of an ideal circle (Fig. 1).
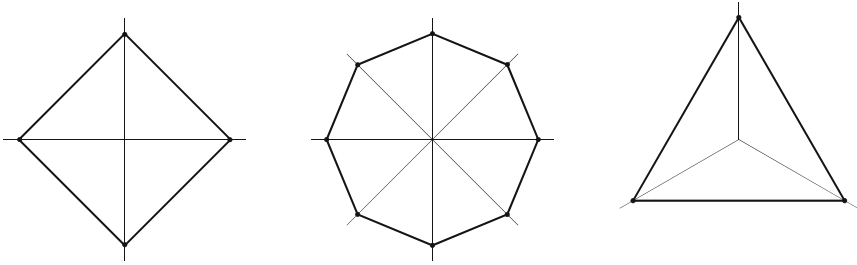

Fig. 1. Radar charts with a flexible number of axes or displayed parameters

In Fig. 2 the transformation from a radar chart to a shape, which could represent the state of one single node can be seen. Visible in the background of the shape on the right is the contour of the ideal circle. For the development of the shape in the foreground the four vertices of the ideal circle have been moved along the four axes to or away from the center of the circle, maintaining the rounding of the four vertices.
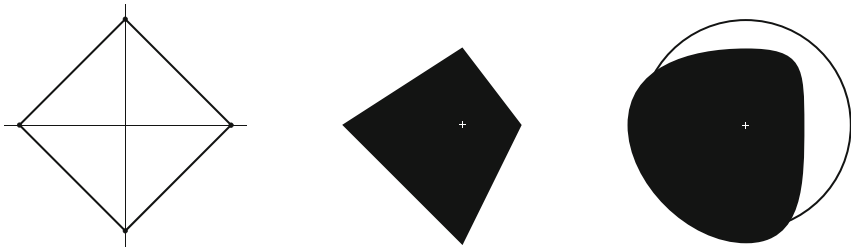

Fig. 2. Transformation of a radar chart with four axes to a shape that deviates from the ideal circle

When talking of a circle as the ideal form (the form where all the values on the axes are on the ideal position), the reason why we didn't use a regular polygon in this place has to be explained.

5.2   Form Perception of Primitives

Perception of Circles and Polygons. Considering how circles and polygons are perceived, the principle arises very quickly that round forms are perceived as soft, continuous, fluent and natural, whereas angular ones are perceived as unnatural and lifeless [18, 19]. The reason for the above named associated characteristics isn't only

the fact that we associate them with other shapes known to us, but also with the visual process itself, where the shapes are patted down by the eyes [19].

In other words, forms are findings of the sense of movement. For instance when we view a circle, the eye tracks the whole form bit by bit, until we experience the completely enclosed circle [20].

Because of its centric symmetry it doesn't exist any direction when perceiving a circle, it is the simplest visual pattern [21]. Also for Gestalt psychologists the circle played a special role: Its percept is totally stable and it is the most regular and simple 2D shape because its surface area is enclosed by the minimal-length. Therefore the circle looks very compact. [22]

So, the perception of circles and polygons and the characteristics associated with them as well differ much. Moreover different types of polygons are also associated with different characteristics: Whereas the square has a very static, quiet aura, the triangle appears as very active, especially if it stands on one of its tips [23].

When deciding for a circle as the basic form for our visualization of a cluster state, we therefore solved the problem, that questions, which need a variable amount of parameters to be answered (and already in it's ideal form lead to polygons with a different amount of vertices) are associated with different characteristics. Using a circle and modify it according to the number of parameters (and therefore axes) leads to a much more constant perception, because in it's ideal state it always has a circle as a result (Fig. 3).
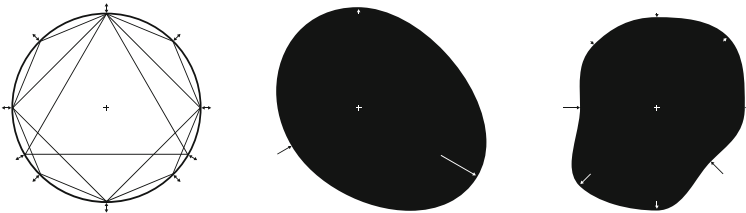

Fig. 3. Examples of the construction of shapes out of a circle with three, four and eight axes

The Circle and its Reason to Describe an Ideal State. As a unique exception of the basic geometric shapes, the circle doesn't consist of different directed lines, but of a uniform curvature [23]. It is completely directionless, has no beginning nor end and is a symbol of perfection, having already been described by Plato (in Timaeus), but also appears in other cultures like the Chinese or the Hindu culture [24].

When perceiving a shape, we start looking for a structural framework [19, 21]. Examples for elements of this framework include symmetry, angle or side length. Regular shapes in general include such a framework that puts in order the single parts of it. The brain is able to quickly complete the form, without determining it over a long time. By comparison, irregular forms don't have this framework but get associated with something already known [19].

So when using a regular shape (e.g. the circle) as the basic shape and modify it according to the angles of the axes allows the viewer of many of these shapes (in our purpose that represents the state of the whole amount of nodes of a database-cluster) to

quickly identify the irregular shapes. This leads to a fast detection of nodes that are in a bad condition and need improvement measures, which only a human being is able to deduce.

5.3   Difficulties of the Concept

During the first tests of this concept with real data (with four parameters as described in Sect. 5.1) received from Crate Technology GmbH, we detected a few difficulties that need to be taken into account when thinking of adapting this concept to various questions that could interest the employees responsible for their databases. Some of the most important ones will be clarified below.

Firstly, the size of a single node presented as one piece of a big group of nodes has to decrease in order to give an overview about the whole cluster. This results in a large number of very small elements (nodes), the curvatures of which may be difficult to detect, depending on which end-device the database is observed. So in the further planning of the interface, a suitable order has to be found, a very contrasting coloration of the nodes and the background as well as enough space around each element has to be ensured.

Another fact we have to take into account is that the ways to define the ideal value of the different parameters may vary. For example when one considers the parameters "average load", "used disk", "used heap" and "number of shards" of each node, the arithmetical average of all nodes may be the ideal value for the used disk, used heap and number of shards as these values should be as balanced as possible. But for the average load it could be more interesting to see just the values that exceed a specific value (for example 1). As already described previously, the curvatures necessary to produce a circle when the values are ideal have to be taken with the vertices moving along the axes.

If the values of all axes are on the same level but not ideal (too high or too low values), this principle produces either too big or too small completely round circles. So if the variation of the size of the circle doesn't stand out regarding the whole cluster, another variation of the curvatures has to be developed in order to avoid misunderstandings.

6   Further Tasks and Conclusion

Our next steps will be to tackle the previously described difficulties, to perform tests with a broader range of participants selected from the expected customer group of Crate Technology GmbH and to revise the visualization according to the results. Subsequently we plan to extend the visualized answers to a broader range of answers to questions that may be valuable as well and to think of opportunities to integrate these visualizations into the user interface of Crate. In this process, the necessary functions of the visualization interface will reveal, such as a possibility to see and compare the cluster state at different points in time, or the zooming from the view of the whole cluster to a single node view with further details.

First budget usability tests of the visualization of single nodes with real-time data have already shown that the participants are able to deduce the state of the node via a shape developed through the merging of radar charts and circles. This suggests further research in the simplicity and intuitiveness of the perception and assessment of shapes, in particular when working in the environment of big data.

References

1. CRATE Technology GmbH: Crate. Your Elastic Data Store. https://crate.io/
2. Olson, J.E.: Database Archiving: How to Keep Lots of Data for a Very Long Time. Morgan Kaufmann, Burlington (2009)
3. Sicilia, M.-A., Lytras, M.D. (eds.): Metadata and Semantics. Springer Science & Business Media, New York (2009)
4. Taylor, A.G.: Much about metadata. The Whole Library Handbook 4: Current Data, Professional Advice, and Curiosa About Libraries and Library Services, pp. 303–313. American Library Association, Chicago (2006)
5. Caplan, P.: Metadata Fundamentals for all Librarians. American Library Association, USA (2003)
6. Parsian, M.: JDBC Metadata, MySQL, and Oracle Recipes: A Problem-Solution Approach. Apress, Berkeley (2006)
7. Cox, M., Ellsworth, D.: Managing Big Data for Scientific Visualization. Siggraph 97, Computer Graphics Annual Conference Series, Course 4, Exploring Gigabyte Datasets in Real-Time: Algorithms, Data Management, and Time-Critical Design. Assn for Computing Machinery, Los Angeles (1997)
8. Hahn, B., Zimmermann, C.: Visueller Atlas des Spitalalltags - Visualisierung organisatorischer und kommunikativer Abläufe im Patientenprozess. In: Mareis, C., Joost, G., and Kimpel, K. (eds.) entwerfen wissen produzieren. Designforschung im Anwendungskontext, pp. 271–291. Transcript, Bielefeld (2010)
9. Zabbix SIA: Zabbix. The Enterprise-class Monitoring Solution for Everyone. www.zabbix.com
10. The Cacti Group, Inc.: Cacti. the complete rrdtool-based graphing solution. http://www.cacti.net/
11. Ganglia Monitoring System. http://ganglia.info/
12. NuoDB, Inc.: NuoDB. http://www.nuodb.com/
13. Pivotal Software, Inc.: Pivotal HD. http://www.pivotal.io/big-data/pivotal-hd
14. Elasticsearch: Kibana. Visualize logs and time-stamped data. http://www.elasticsearch.org/overview/kibana/
15. Schulz, H.-J., Nocke, T.: Maschinelle Datananalyse im Informationszeitalter - Können oder müssen wir ihr vertrauen? Ordnungen des Denkens. Debatten um Wissenschaftstheorie und Erkenntniskritik, pp. 121–132. LIT Verlag, Berlin (2007)
16. Software Diagnostics Presents the Effectiveness Platform (2014)
17. MongoDB, Inc.: MongoDB MMS. Cloud Managed MongoDB. https://mms.mongodb.com/
18. Grillo, P.J.: Form, Function, and Design. Original Title: What Is Design? Paul Theobold and Company, Publisher, Chicago 1960. Dover Publications Inc., New York (1975)
19. Bernhard, T.: Wahrnehmung der Schönheit in der Architektur. epubli, Berlin (2012)
20. Naumann, A.: Raimer Jochims: FarbFormBeziehungen. Anschauliche Bedingungen seiner Identitätskonzeption. Königshausen & Neumann, Würzburg (2005)

21. Arnheim, R.: Art and visual perception: a psychology of the creative eye. 50th Anniversary Printing. Original edition: Art and Visual Perception: A Psychology of the Creative Eye. Universitiy of California Press, Berkeley and Los Angeles 1954. University of California Press (2004)
22. Pizlo, Z.: 3D Shape: Its Unique Place in Visual Perception. MIT Press, Massachusetts (2008)
23. Alexander, K.: Kompendium der visuellen Information und Kommunikation. Springer, Heidelberg (2013)
24. Neumann, E.: The Origins and History of Consciousness. Original Edition: Ursprungsgeschichte des Bewusstseins. Rascher Verlag Zurich 1949. Karnac Books, London (1989)