Explore the Challenges for Building Big Data Statistics and Visualization system

LIU,HONGYANG

March 18, 2020

1 Choose one domain / research area.

big data visualization system

2 State the research title.

Explore the challenges for building big data statistics and visualization system

3 You are also required to provide a minimum one page introduction of the research domain and the topic that you have selected.

3.1 Abstract

Since 2005, the big data terminology has appeared for the first time. People are facing the problem that the data is growing exponentially. As more and more people see the information behind the data, the data is changing the way people when they make decisions. However, due to the characteristics of big data like variety and the different sources data comes from, it is quite difficult for people to build a unified platform for data statistics and visualization.

Besides, a large amount of data also affects the performance of the platform showing better functionality. In this article, we try to focus on exploring and comparing the challenges of implementing big data visualization system to discover which methodologies are critical and also their limitations.

3.2 Introduction

The general objective of this article is to explore and compare the prerequisites and challenges of building a big data visualization system to help system developers or big data architectures to find better measures to develop an interactive system for non-experts to analyze big data. Interactive systems can not only help people figure out data, they can also facilitate users to find potential knowledge through data exploration(Bikakis and Sellis 2016). The visual system can provide users with an intuitive sense of understanding of the meaning and patterns of the data.

Traditional data platforms struggle to handle large amounts of data sets(L. Wang, G. Wang, and Alexander 2015). Considering the scalability and computing power, we will face more challenges in addressing problems related to big data and also improving the visualization system performance. Therefore, the modern big data visualization system should consider more limitations like responsive time, scalability and data format issues. In this article, we emphasize the instruments like various programming languages or data warehouses which may influence the big data system on performance, structured cost and responsive time.

According to the research conducted by previous researchers(Bikakis, Papastefanatos, and Papaemmanouil 2019), the writers introduced that there are several challenges building big data analytic system: One challenge is that the large size of dynamic data may reduce the system performance when people retrieve or query data from database or data warehouse. To accomplish the retrieval task, the system should have the ability to efficiently access the database and process data sets. Meanwhile, We should also consider the hardware performance because the large raw data formats make a single computer busy to handle it. A single computer has the limited hardware resources like ARM, CPU to handle large datasets and we should ensure that the system could run on the scalable computing nodes so that the performance could be scaled in to meet the users' demands and reduce

the companies' cost. Another challenge is the variety of data formats. Because the big data may be collected through many different resources like web crawling or public database. The data structure may also be different. Therefore, users could choose different data formats according to their tasks. This challenge forces the system developers to consider which way to present the data; The last challenge is that the developers should also recognize, is that the techniques of implementing a big data system. With the development of computer science and data science subjects, there are amounts of choices for developers to build a big data platform. But we still need to consider the cost and complexity of the system. In this article, we will compare different techniques such as programming languages or frameworks, database choice or data warehouses, and big data tools to discover their advantages and disadvantages in terms of the money or complexity of developing interactive systems.

To test the performance or scalability of the system. we would use some metrics to measure the system performance. A survey like SurveyMonkey would be utilized to non-expertise users for collecting their questionaries about evaluating the system. The goal of the data visualization system is to help average people without data science knowledge to process big data and show the visual results. Therefore, the developer should ensure that the interactive interface is user-friendly which requires little background knowledge to use it and contribute to users exploring the meaning or pattern behind the big data(Caldarola and Rinaldi 2017).

4 Please discuss and state the research problem, the research objectives and research questions.

4.1 Please discuss and state the research problem

Nowadays, with the advent of the big data era. there is an increasing number of companies deciding to collect transaction records to do data statistics and analysis. Data visualization is one of the best useful ways to do data storytelling and helps people in the company to analyze customer's actions and improve their sales strategy to increase companies' revenue. However, considering the characteristics of big data and the amounts of techniques for implementing big data visualization systems, it is very hard for the software company to choose a better way to efficiently develop a big data visualization system. In this article, we will explore the challenges, such as programming languages, big data software tools. The research problem of the study is that the comparison of the challenges while building the big data statistics and visualization system. Because of the complexity of the big data, it makes it difficult for developers to handle the massive data and this needs higher requirements than a traditional software system. The problem in this study is exploring a more economical way to develop a big data visualization for software developers to help users easily analyze or understand the behind meaning of data to help them to do decision-making(Bikakis, Papastefanatos, and Papaemmanouil 2019).

4.2 the research objectives

The general objective of the research is to explore the challenges of building big data statistics and visualization systems, such as data volume and data format and software techniques, to help software system developers find better ways to develop interactive systems for non-professional users to analyze data and make business decision-making.

4.3 research questions

What challenges do the data characteristics and technologies influence the performance and scalability of the big data system?

5 Discuss your research significant and your research motivation.

We will explore the limitations and challenges in building the statistic and visualization system in the previous research like graph visualization system, they mainly focus on how to solve the graph problem(Abello, Van Ham, and Krishnan 2006). But they lack the comparison of the challenges of different methods and research on how other solutions of developing a big data system platform. Throughout the article, it will contribute to demonstrate which way could help software developers to develop a better interface for the user. For example, we will compare the different frameworks like Flask, Springboot, Shiny to notice the advantages and disadvantages of building a big data system. In the results, we will also give a survey to evaluate the performance of the system to non-experts to use it. The collection of data may also become metrics to help their development of software development.

6 What is the research gap?

In the study, we read related previous articles about big data visualization system and the data visualization products developed by companies to find how they cope with the large data volume and complexity of big data problem. Through the comparison of different big data tools and their implementation methods, we could find their benefits and weakness regarding processing big data. Because the previous study mainly focuses on traditional data visualization tools, we decide to find new progress since 2005. This method or experience would ensure the software developers could learn from the previous developing schema and experience. On account of the broad field, we mainly discuss the challenges of big data characteristics and techniques about the data visualization tools. Because of the big data is changing with time, the new methodologies are improved, the article would focus on the challenges for building big data statistics and visualization system.

References

Abello, James, Frank Van Ham, and Neeraj Krishnan (2006). "Ask-graphview: A large scale graph visualization system". In: *IEEE transactions on visualization and computer graphics* 12.5, pp. 669–676.

Bikakis, Nikos, George Papastefanatos, and Olga Papaemmanouil (2019). *Big Data Exploration, Visualization and Analytics*.

Bikakis, Nikos and Timos Sellis (2016). "Exploration and visualization in the web of big linked data: A survey of the state of the art". In: *arXiv preprint arXiv:1601.08059*.

Caldarola, Enrico G and Antonio M Rinaldi (2017). "Big Data Visualization Tools: A Survey". In: *Proceedings of the 6th International Conference on Data Science, Technology and Applications*. SCITEPRESS-Science and Technology Publications, Lda, pp. 296–305.

Wang, Lidong, Guanghui Wang, and Cheryl Ann Alexander (2015). "Big data and visualization: methods, challenges and technology progress". In: *Digital Technologies* 1.1, pp. 33–38.