

Paper 4761-2020

## HOW TO MAKE YOUR REPORTS FLY! THE IMPACT OF VISUALIZATION LAYER CALCULATIONS ON PERFORMANCE

Jessica Fraley, University of North Carolina at Chapel Hill

### ABSTRACT

SAS® Visual Analytics reports at the University of North Carolina Chapel Hill can be written against tables containing 500 GB of data. Reporting in an enterprise-sized environment has a specific set of requirements that were not present in the smaller environments of the past. Not paying attention to these differences can slow report visualization to an unusable crawl. However, there is one easily addressed methodology that gets the most out of your reporting performance. This paper examines the impact of visualization layer calculations (which is common in non-enterprise-sized reporting) on enterprise-sized reports. This paper also goes into the details of how some of our reports had performance increases from minutes to seconds when going against our largest tables.

### INTRODUCTION

**Change within the tech arena is fast paced and often requires 'unlearning' what was valid and functional as recently as 5 years ago;** using new processes and enhancing with new knowledge. Adoption of new techniques and implementing new software is common. As the **world's** data footprint continues to grow exponentially, it requires that technology become more and more precise in both engineering and operations. New technologies like SAS Visual Analytics™ and SAS Viya™ **push** the edge of processing speed and capability. However, from a report creation and visualization standpoint it is not uncommon for data analysts and their customers to continue working with methodologies that were effective in the past. New methodologies can improve outcomes when applied correctly.

Many authors have written papers on improving report performance covering a range of topics. For instance, parsing strings is very time consuming; so, use functions like CONTAINS sparingly. This paper will focus only on one of these topics because here at the University of North Carolina at Chapel Hill it is the one that has consistently had the biggest impact on the time it takes to visualize reports with large data sources. It is also one of the easiest to address.

This paper is a real-world example of how moving as many data calculations and data item creations as possible from the report visualization layer to the data layer effects report display times. In our tests we reduced large data report visualization times from almost 5 minutes down to 15 seconds.

**This isn't a new technique. In the IT world, this** thought process has been around for decades. It is the prime design behind most big data implementations. All data calculations and data item creation should be done at the database level if possible. The visualization layer should only be used for reading and displaying. However, this concept is not widely

known outside of the IT world. This is what we did to display the effect report level calculations have on report timings to our end users.

## PERFORMANCE ISSUES

Data volume within the Enterprise Reporting and Departmental Systems (ERDS) department at the University of North Carolina was growing at a rate of approximately 30% a year. In the past, many customers would download the data they wanted into a spreadsheet. There, they would combine data columns and generate new data items. That approach works with smaller data, but once data reaches a larger size (and reports need to be distributed to many individuals), **the spreadsheet model doesn't work anymore**. As customers move from spreadsheets to an enterprise-sized reporting solution the customer thought process must change to be effective with the new tools. For instance, one report within our SAS Visual Analytics™ environment was created against a 500GB table and contained 44 embedded data item calculations and data item creations. These calculations only began executing once the customer 'clicked' to open the report. The result was that the time to surface the visualized report to the customer was almost 5 minutes.

We examined the report and saw that there were 44 calculated items on the report. The data contained approximately 200 million rows which means that once the report was selected to be opened the system had to execute more than 8 billion calculations (as well as apply any filters) before the report could be displayed.

In order to determine the effect that these calculations were having on report surfacing times (as well as to allow us to make recommendations to the customer), we set up an experiment. Using the same data source 4 reports were created.

1. Report with no calculated data items was created using the 500GB data source
2. Report with 3 calculated data items was created using the 500GB data source
3. Report with 10 calculated data items was created using the 500GB data source
4. Report with 44 calculated data items was created using the 500GB data source

We then opened the reports in SAS Report Viewer™ and measured the time from report opening until report display for each of our experimental conditions. The within-group timings were repeated to measure variance, but the variance of timing between within-group runs was negligible from a human experience perspective.

The results of the experiment are displayed in Figure 1 below.

### Comparison of report visualization times against number of calculated data items in the report

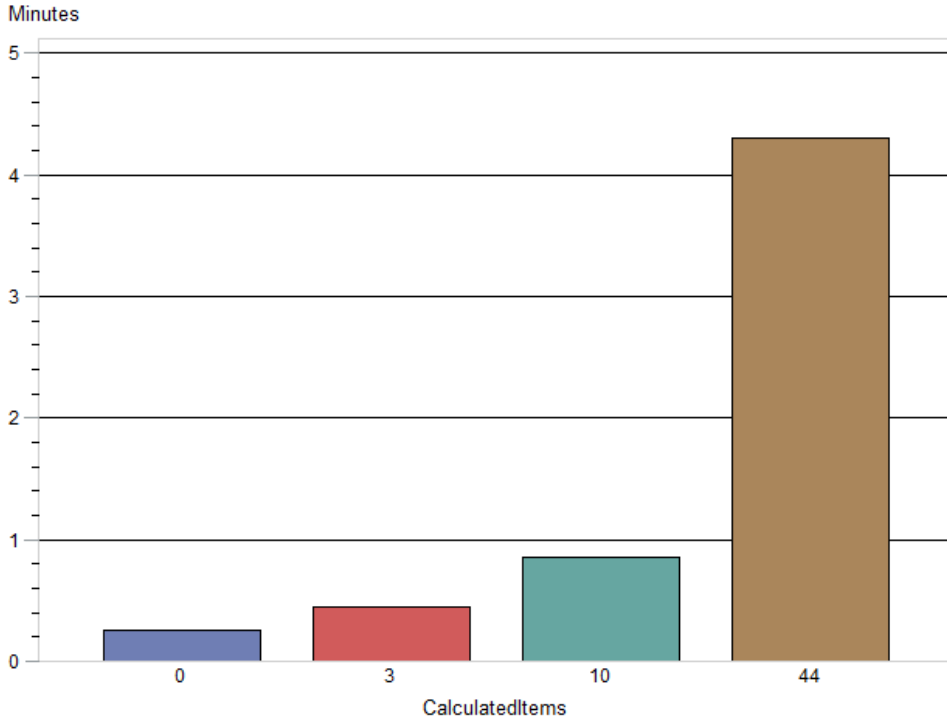


Figure 1. Time to open reports with a 500GB data source versus the number of calculated items in the reports

The result of the runs was as follows (the times are being rounded to the nearest 5 seconds because from the human experience the few second time variance was a non-factor):

1. Opening the report with no calculated data items averaged approximately 15 seconds.
2. Opening the report with 3 calculated data items averaged approximately 25 seconds.
3. Opening the report with 10 calculated data items averaged approximately 50 seconds.
4. Opening the report with 44 calculated data items averaged approximately 4 minutes and 20 seconds.

## RESULTS

As you can see from Figure 1 above, the time to display a report increases drastically with an increase in report defined calculations. A 500GB data source displays in approximately 15 seconds if no calculations are required at visualization time. And the amount of time that it takes to display a report increases with additional display time calculations. These results are not specific to **any SAS™** product but would apply to any reporting product by any manufacturer.

## CONCLUSION

Moving data calculations and data item creation out of the reporting layer and into the data layer is really an obvious solution to aid report performance issues. However, from our experience this was a solution that report creators and customers were hesitant to adopt until the size of the effect it would have on report performance was shared with them. Once we demonstrated the size of the performance improvement that doing this would have, the customers began embracing this idea. By emphasizing this one simple step to our report creators we have been able to improve our end user experience drastically.

The hardest aspect of working with larger data is the mindset change that is required. Big data solutions are focused on fast data extraction and display. Data preparation is most often done at the ETL/Data layer because data is now too large to effectively perform complex calculations at display time. When we create reports against larger data, if we want those reports to display in a timely fashion, the report creators must keep in mind that each of the calculations and data item created will require the report to do work which will lengthen display **times**. **It's a balancing act because as the data size increases the constraints of efficient report creation become more and more important.**

## ACKNOWLEDGMENTS

I would like to thank Mimi Bennett for her help in preparing this paper.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jessica Fraley  
University of North Carolina at Chapel Hill  
[Jessi.gc@gmail.com](mailto:Jessi.gc@gmail.com)  
Jessica\_Fraley@unc.edu