Optimizing Data Visualization

**Summary**

Great visualizations can make data come alive, leading to better insights and decisions. Data consumers of every kind rely on dashboards to perform their work, but they have a common complaint, slow loading times.

Academic research indicates that human attention tends to drift after a period of ten seconds. If data consumers need to wait more than ten seconds for a dashboard or report to render, they will often turn their attention to another task, leading to less productive employees.

Our client, a Fortune 500 entertainment company, manages approximately 13,500 titles available in 190 countries for 148 million paid customers. It needed to improve the response performance of its Tableau dashboards and reached out to Starschema to help optimize their data visualizations.

**Practice Areas**

- Data visualization
- Data engineering

**Challenges**

- Dashboard optimization
- Slow dashboard rendering due to large data sets
- Dashboards only displayed one month of data limiting their value

**Business Impact**

- Deeper insights into the state of the business
- More responsive data consumer experience
- Reduced operational costs

**Technologies**

- Tableau

**Challenge**

To measure a title’s performance, our client’s data experience team built a complex exploratory dashboard for both data analysts and senior management. This “Title Performance” dashboard relied on data from an Amazon Redshift table with 600m rows, more than 40 columns, and more than 30 plus complex calculated fields in Tableau (Level of Detail (LOD), Table Calcs).

Due to the size of the data set, loading extracts took hours and both extract and live query rendering took 30-40 seconds for each filter interaction. To overcome these performance issues, our client built several dashboards, each limited to one year and a subset of countries, to limit the underlying data. While these new dashboards had acceptable performance, they did not provide a view spanning multiple years or geographies, limiting our client's ability to see long-term trends across multiple geographies.

Not only did this solution lack a single view of title performance, it was difficult to maintain. Any dashboard changes required modification of the 10 to 20 workbooks within the filtered data set.
Solution

Having deep expertise and extensive experience with Tableau but also data engineering of back end systems, our team knew that the underlying causes of latency could exist across the whole data platform. Starschema approached the optimization process in different, parallel ways.

After confirming that the dashboard loading time were equally slow on both the Desktop and Server versions, our team focused on the in-scope workbooks instead of looking for Tableau Server capacity bottlenecks. The team next checked how the datasources were created in Tableau – this was a crucial step as it determines the type of query Tableau’s SQL engine generates and how it performs in the underlying database. Using Starschema’s custom tooling, we discovered that Tableau was generating queries that had unnecessary joins in them, slowing queries in the database and transferring unnecessary data over the network.

By setting up the correct parameters in Tableau like join culling and proper database key management, we eliminated the unnecessary overhead from the queries. For fixed LOD calculations the team pre-calculated the fixed values in a normalized table to further reduce the overall number of queries. In addition to filter and query optimization, our engineering team redesigned the workbook data sources to leverage Tableau’s query fusion engine to server multiple visualizations from single queries even if the views data aggregation levels were different.

We then analyzed the physical layout of the data in the underlying database. Once we understood the generic structure of the queries, we modified the data model to run these queries faster. The modifications included creating compound and interleaved sort keys based on the most frequently used filters. This could be further fine-tuned by using Starschema’s Tableau Tracker extension to find hot and cold data within the database tables and use database facilities to ensure that the most commonly used data was kept in cache for faster retrieval.

Results

By switching from multiple, fragmented dashboards containing a subset of data stored in extracts to a live connection as well as applying proven best practices in data engineering we extended the time horizon one month to multiple years of data, successfully visualizing billions of rows and improved response times by an order of magnitude - from minutes to seconds. In addition to the extended scope of analysis and Workbook’s improved responsiveness, our client saves approximately two full-time equivalent (FTE) of resources per month by maintaining only one Dashboard that serves all use cases. By eliminating three to four hours of Tableau Server backgrounder jobs, the team saved $30,000 annually on license and capacity.

Our client’s leadership team recognized Starschema’s efforts and the Data Experience Team and Starschema presented the solution and methodology at the San Francisco Tableau User Group.