InfoCepts

Using MicroStrategy and Big Data to Enhance Productivity and Save Costs for a Consumer Research Giant

Summary

Using MicroStrategy with Hadoop and other Big Data technologies, we helped our customer reduce costs by 35% and improve performance by 80%. We syndicated data for multiple countries and client into a single MicroStrategy project and built a robust security model on top of it to ensure seamless access to reports.

Industry
Retail (Market Research)

Users
Category Managers, Marketing Managers, Regional Managers

Technologies
MicroStrategy, Hive, Cloudera Impala

Team Size
8 InfoCeptians, 3 Customer Associates

InfoCepts Accelerators or Assets Used

MicroStrategy on Hadoop Fast Track, MicroStrategy – Hadoop Connector Sandbox

→ The Challenge

Our client, a global information and measurement company is in the business of providing consumer behavior insights by collating data from various sources. Their customers use those insights to make better buying, merchandizing and pricing decisions.

The client sought our expertise for syndicating their solution so client data from one specific business geography can be handled by a single schema within a single MicroStrategy project. Their global business footprint is made up of 3000+ categories across nine countries with anywhere between 100 to 150 clients in each country. Additionally, they wanted us to automate their reports and dashboards to eliminate manual efforts.

We encountered several hurdles while achieving the desired outcomes –

- Architectural flaws in the legacy system made it non-scalable and inefficient
- We had to find an automatic fix to stop analysts from manually processing flat files to produce insights in MS PowerPoint presentations
- The situation warranted us to engage in a complex exercise of customizing and integrating a MicroStrategy web application with a Big Data powered data exchange platform
- Since data across countries, categories and clients had to be housed in a single database, user data security became a major challenge

→ The Solution

We used our 'One Project for One Database' method to extend the 'category' level database for accommodating multiple clients, countries and categories. We attached a wrapper library of 'Client', 'Country' and 'Category' data to all tables within the schema to address scalability and efficiency. This helped us eliminate manual processing efforts as it provided an automatic way to feed data into reports and dashboards.

Prior to our engagement, Cloudera Impala handled analytics workload for the client's 70 nodes strong Hadoop Cluster with Cloudera Hadoop distribution. Cloudera Impala ODBC driver 2.5.33 established connectivity between MicroStrategy and the solution's Hive based data warehouse. Text file formats in the Hadoop Cluster had slowed down the query performance.

Turning the slow query performance situation around required us to change the format to Parquet with partitioning at Client, Country and Category levels. Tuning the VLDB settings as per MicroStrategy recommendations significantly reduced the processing time. We modified definitions of certain metrics to use database functions instead of MicroStrategy. This further improved performance and helped us achieve the 30 second SLA for query response time. Our implementation of a Login ID based security model addressed the security challenge posed by data about all clients, countries and categories being housed in a single schema. We introduced a dedicated table to handle secured access at the database level. This table contains client, country and category level access for all Login IDs. Every Login ID has an entry mapped within the table to enable data security as well as object and dashboard level security.

→ The Results

Our solution helped our client offer an improved, seamless and more interactive user experience to its customers. Reduction in manual efforts due to automation has yielded the following benefits:
- Brought end-to-end analysis cycle run time down by 80%.
- Reduced costs related to support staff by approximately 35%.
- Reduced maintenance costs by increasing solution scalability.

Thanks to performance tuning measures, the solution also delivered on the 30 second SLA for query response time.