

CCA Analytics: A Scalable Solution to Facilitate Data Analysis

Chris Ramsdale, Computer Corporation of America

Let's face it: businesses today can access more data than ever before, whether from the Internet, open data-access standards, or 10 years of legacy systems. In fact, some might say there's too much data—at least, too much of what we don't want. But how do we pick out the good stuff—the golden nuggets—that can help our businesses? This challenge underscores the need for effective tools to derive business intelligence from endless mines of data. Yesterday's lightweight query, reporting, and graphing tools aren't enough.

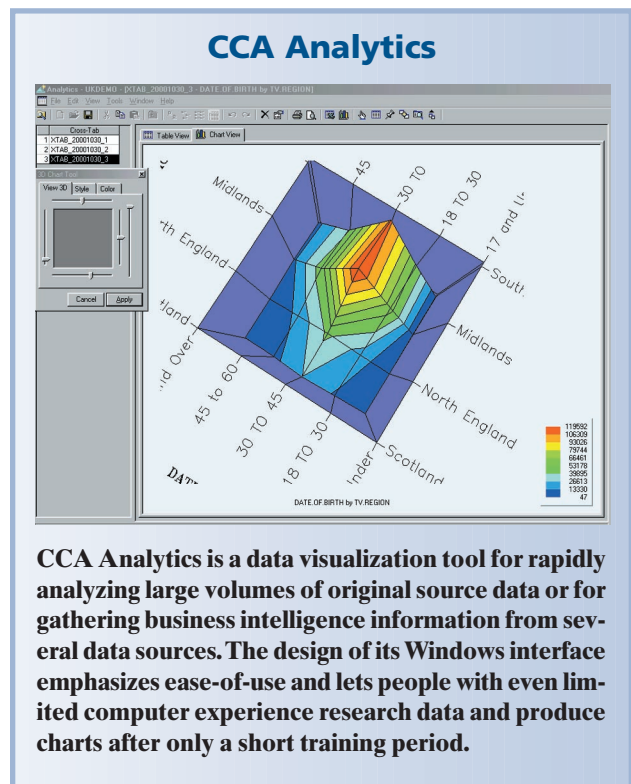
Computer Corporation of America designed its new data analysis tool, CCA Analytics, to help IT shops quickly deploy high-performance, scalable analytical applications that can be used to analyze any data source, including CCA's own Model 204, the IBM DB2, and Teradata.

QUERY AND DISPLAY FEATURES

Business users typically know what information they want from data but lack the programming skills to find it themselves. So they rely on the IT department to deliver custom reports, charts, and graphs every time they require a new piece of information. Our company designed CCA Analytics for ease of use without an in-depth knowledge of databases or the data.

Users can perform simple or highly complex queries in a point-and-click environment. They can search the data based on actual values, or macro or formula results. They can also iteratively refine queries to narrow down the data step by step. Once they find the data they're looking for, they can save the selection criteria for reuse later.

Applications for analyzing data typically involve data summaries rather than lists of detailed records. Using CCA Analytics, users can obtain such summaries—grouped by any number of fields or by computed values—including



CCA Analytics is a data visualization tool for rapidly analyzing large volumes of original source data or for gathering business intelligence information from several data sources. The design of its Windows interface emphasizes ease-of-use and lets people with even limited computer experience research data and produce charts after only a short training period.

counts, averages, and maximum and minimum values.

CCA Analytics provides numerous tools for viewing selected data, including reports, charts, and graphs. Multidimensional graphical displays help users easily visualize complex data relationships. Besides displaying the data, these graphical displays are also interactive, so users can manipulate them to form new queries, drill down to

underlying data, or view the data from different perspectives. For more specialized views, users can export the data into commonly used analytical packages such as spreadsheets or online analytical processing (OLAP) tools.

BIT-MAP INDEXING

Analytical applications typically require fast analysis and summarization of large quantities of data. Every analytical tool provides easy-to-use query and display features. However, because of performance restrictions, these tools aren't always practical for real-world analytical applications such as data warehouses.

CCA Analytics relies on bit-map indexing technology for in-memory analysis of large databases. Every data item in every database is indexable. Using the concept of *invisible data* (indexing a field without storing the actual data values in the data records), CCA Analytics lets users add new discriminators for analysis without storing actual data in the physical records. This feature has two benefits:

- It lets you prejoin entities for enhanced performance.
- It does away with the need to store any actual data records. CCA Analytics can replicate external sources of many gigabytes solely as index structures, using only a fraction of the original sources' storage.

CASE STUDY

The Board of Cooperative Education Services (BOCES) in Syracuse, New York, had an all-too-common problem: Its IT department was servicing internal clients' informational needs as well as building new applications. The Board's clients consisted of 2,500 school principals, administrators, and researchers in nearly every school district in the state. Each client request was different, involving vast information about various students and requiring diverse formats to display the data.

New development in the IT department took a back seat to information delivery, but, even then, the Board was unable to completely satisfy client requests, and it was losing customers. Therefore, BOCES decided to build a data warehouse to reduce turnaround time by letting users obtain their own answers and encouraging them to explore the data in depth. In turn, the IT department would have more time to devote to new initiatives.

A decade of data

BOCES has 10 years of research data—30 million records of 3,000 fields, with 161 different record types. All of this is stored in 58 Model 204 databases on an IBM S390 MVS enterprise server. Just figuring out what data users would need from the warehouse would take a lot of time, let alone integrating components such as data cleansing, transformation, and analytical tools. The programming staff estimated that warehouse construction would take between 8 and 18 months. But during construction, these

programmers still had to serve the Board's large user base, leaving little time for the data warehouse project. Therefore, BOCES began considering an interim, virtual data warehouse to free up IT resources.

Choosing a business intelligence tool

The first task was to choose the right business intelligence tool. BOCES chose CCA Analytics because it provided all the basic query, reporting, and analytical functions that users needed; it was from the same vendor as the database that BOCES was using, thus minimizing the learning curve for producing custom functionality; and it delivered query results far faster than the other solutions BOCES tried. "We were not naive about the performance implications of throwing large numbers of researchers at operational data," said Larry Dismore, director of BOCES's Central New York Regional Information Center. "The indexing technology inherent to CCA Analytics made this a nonissue."

Architecting the virtual data warehouse

Like building a true data warehouse, constructing a virtual one required some understanding of the users' information needs. But because the virtual data warehouse would consist mainly of index tables and logical views of the data rather than physical copies of operational data, the views did not have to be right the first time; BOCES could easily create new ones, if necessary.

BOCES constructed numerous views of all current databases, ensuring maximum flexibility for data analysis. Users could then better understand what data was available to them, which would translate into more accurate user requirements for the true data warehouse. In two weeks, the IT department completed the initial implementation of the virtual data warehouse and delivered it to users. Early results are promising, as BOCES reports that users are accessing and analyzing data relatively easily.

Moving ahead

The virtual data warehouse is not only solving a short-term business intelligence need; it is also freeing significant programming resources to design the true data warehouse. This benefit will save both time and money for the IT department.

Moreover, when the true data warehouse is up and running, the virtual data warehouse can continue to evolve. Even the most successful data warehouse implementations will reveal weaknesses and shortcomings over time. The virtual data warehouse can continue to address these problems. The virtual warehouse, because of its association with operational data, can stay one step ahead of the permanent installation of the true data warehouse and identify design change requirements. ■

Chris Ramsdale is director of strategic product planning at Computer Corporation of America. Contact him at chris_ramsdale@cca-int.com.