CASE STUDY

NATIONAL STUDENT CLEARINGHOUSE

CC PACE

THE PROBLEM

The National Student Clearinghouse® won a multi-million-dollar grant from the Bill and Melinda Gates Foundation to improve the quality of longitudinal studies on high school students. This would provid high schools access to NSC's vast data warehouse of student information to reliably measure the outcomes of educational efforts.

NSC decided to use part of the grant to improve their data collection and warehousing system called Student Tracker for High Schools, or STHS. STHS is the central system that manages all the data used for longitudinal studies. According to NSC's projections, a surge of demand for high school request records would occur in tandem with the roll-out of this new system, reflecting an increase of nearly 2,000% from previous levels to over 47 million annual requests. Faced with the need to support this much greater access to their data warehouse of over a billion records, and with tight deadlines to meet fall enrollment data requests, NSC hired CC Pace to help deliver the new system. Several challenges faced the effort. Each challenge was successfully met by the CC Pace team.

On the NSC side, research specialists needed to constantly tune and tweak algorithms to improve both the frequency and quality of matches. However, the legacy system used "hard-wired" matching algorithms: the application code itself contained the various strategies used to assemble data from the warehouse and the rules applied to those data. Only skilled programmers could modify the rules, and the system had to go through a release process when these rules changed. The result was that the specialists' need for better algorithms couldn't be quickly met. As the volume of data in the warehouse grew to over a billion records, the legacy system also faced performance problems. Matching took an ever-increasing chunk of time as the algorithms had to sift through huge volumes of student data.

THE SOLUTION

CC Pace designed and built a data intake workflow system, a matching engine, and reporting functionality. Colleges and high schools can send their files in a secure manner. The system automatically ingests the files, processing several files in parallel if necessary, and runs batteries of complex validation rules to ensure data quality. Two kinds of validations are run: syntactic validations ensure that the data formats are correct, and semantic validations ensure that the data make sense.

NSC data intake specialists then view the validation results at a website that features innovative interactivity. If there are errors, the specialists fix them and re-submit the data. Once the data are clean, the specialists import them into the central NSC data warehouse. Data that pass through this gauntlet of validations and corrections are much cleaner and more consistent than previously. In turn, research-based on these data has lower margins of error.

Then CC Pace built a matching engine capable of running complex student matching algorithms against the warehouse. The engine does not contain any matching algorithms within its code. Instead, it lets analysts design complex algorithms and, using simple web interfaces, load them into the engine as XML files. Multiple algorithms, each supporting a different research objective, can run simultaneously. Researchers can now modify their matching algorithms without any programmer intervention and while the system remains up and running.

Finally, a reporting function was built using a commercially available reporting server and designer to provide reports for thousands of participating high schools and districts that can be processed in parallel. Both the database and report server were designed to scale by offloading the report processing to dedicated report servers, allowing NSC to scale horizontally by simply adding another report server node and registering that node with the application.

THE RESULTS

NSC's business goals for STHS have been met.

1. High schools' benefit from the new system's capabilities. They get higher quality research made possible by the improved and more flexible matching algorithms. The reports they receive from the system are more visually appealing, but also contain a lot more information than previously. All of this allows the schools to concentrate on using the reported data to improve their own educational outcomes.
2. The improved intake workflow and modern user experience sharply reduces the time data analysts spend on cleaning up data. The high-performance matching engine improves the quality of longitudinal studies and reduces the time that researchers spend resolving ambiguous matches. In peak season, data ingestion is many times faster than it used to be, freeing the organization to work on more research projects from the same data.

ABOUT NSC

Founded in 1993 by the higher education community, the National Student Clearinghouse® relieves the administrative burdens and costs related to student data reporting and exchange. The NSC is a nonprofit and nongovernmental organization and the leading provider of educational reporting, data exchange, verification, and research services. NSC's work — performed in a trusted, secure, and private environment — provides numerous time and cost-saving benefits to students, schools, administrators, and requestors. In addition, education leaders rely on the Clearinghouse's unique national student data resources to better understand student pathways and outcomes in order to help students succeed.

ABOUT CC PACE

CC Pace is a business and technology consulting firm headquartered in Fairfax, Virginia. For over 40 years, CC Pace has built its reputation on an unwavering commitment to customer satisfaction and is nationally recognized for solving business problems and delivering high-impact results. CC Pace specializes in the mortgage banking and financial services industries and our clients appreciate that we have a deep understanding of their business. Our domain experience allows us to design solutions to not only meet their needs, but achieve both measurable bottom-line results and significantly improve business execution.