

October 2006 — Features

Data Warehousing :: Too Much Information

by Justine Brown

The surge in the collection and use of data has created a new problem for K-12 administrators—where to put it all, while ensuring it stays safe and accessible. Data warehouses have the answer.



AMONG ITS MANY SWEEPING CONSEQUENCES over the past five years, the No Child Left Behind Act, with its mandates for collecting and documenting student achievement statistics, has fueled a surge in the volume of data collected by schools, districts, and state education departments. But this information explosion has gone hand in hand with technology's growing affordability and availability, and the realization by school officials that they can put data to work to improve both administrative and instructional processes.

Yet as the collection and use of data has grown, a subsequent problem has emerged—where to store it all, while ensuring it remains protected yet accessible, usable, and meaningful. That's where data warehouses come in. Born in the corporate world, data warehouses integrate data from the various operational systems a school or district uses, and when combined with a data analysis tool, enable administrators to analyze performance over time. More and more administrators at every level are using data warehouses to help manage huge volumes of data and to monitor student progress.

Why a Data Warehouse?

“The first thing that needs to be understood is that data warehouses are tools, not solutions,” says Jill Abbott, learning strategist for the [Schools Interoperability Framework Association](#), which promotes the use of a common technology framework to allow interoperability between the different applications used by schools. “In general, there is a lack of understanding about what a data warehouse truly is.”

A data warehouse is essentially a facilitator. Without one, or with one that is limited or poorly constructed, the data that a school collects often ends up in disparate silos, preventing the school from obtaining a comprehensive view of how students are

performing, how poor performance might relate to absenteeism, how effective teachers are, where the most effective instruction is taking place, and so forth.

“Without a way to manage data, educators miss opportunities to help children that they would have seen had they had access to data,” says Shawn Bay, founder of [eScholar](#), a New York-based company that has implemented hundreds of district-level data warehouses. “Not having a data warehouse also means administrators spend huge amounts of time collecting and integrating data manually to meet compliance reporting requirements—time that could have been spent helping children.”

A school without a data warehouse must cleanse and integrate data individually for each report, creating inevitable inconsistencies that ultimately destroy the credibility of the data and limit its impact.

KEEPING IT CLEAN

THE QUALITY OF YOUR DATA IS CRITICAL TO THE PROCESS OF STORING IT.

A significant challenge in implementing a data warehouse lies in the condition or “cleanliness” of data. To get the most out of a data warehouse, data must be collected completely and consistently, with minimal error.

“Data cleanliness is one of the most costly and time-consuming issues,” says Aziz Elia, chief software architect for [Computer Programs and Systems](#). “As schools start to gather data and analyze it, they’ll start to see a lot of errors pop up, and they’ll have to go back and correct their data.”

“The data cleanliness issue is always an eye-opener,” says Shawn Bay, founder of [eScholar](#). “Even if a school district thinks its data is lousy, it’s usually even worse than that. The bottom line is, you have to make sure whoever is collecting the data in the first place can collect it completely and correctly. We have cool data-cleansing processes, but we can’t fix something that isn’t there.”

Jason McCreary is director of research, evaluation, and accountability for Greenville County Schools in South Carolina, which recently implemented a data warehouse solution from [TetraData](#). McCreary says data quality is an issue that doesn’t go away. “It is a continuous process,” he says, “because one person is not the gatekeeper for all the data. There are many, many people entering data, and if they don’t see the big picture and how the accuracy of their data entry affects everyone, then everyone can get inaccurate results.”

A data warehouse solution, on the other hand, allows school officials to not merely store data in one place, but to analyze the data they collect, examine trends and statistics, and put data into the hands of teachers in near real-time. Empowered by that data, teachers and administrators can then identify individual student weaknesses and make the necessary adjustments to address them before it’s too late.

Data warehouses are typically designed to handle large amounts of data stored longitudinally over time. As a result, they can become very large. In fact, the largest databases in the world are typically data warehouses. Sometimes, data that is not needed for analysis is archived to a secondary system or stored at a less granular level of detail if space needs to be reclaimed.

A SMOOTH TRANSITION

KEEPING TO THESE RULES OF THE ROAD CAN EASE THE MOVE TO A DATA WAREHOUSE SOLUTION.

1. **DECIDE WHAT YOU WANT TO COLLECT.** Understand the uses to which you'll be putting the data before you begin to gather it. "Once you know which questions you want to answer, it's a smoother process to figure out the right tools and architecture," says Jonathan D. Harber, CEO of [SchoolNet](#), which implements instructional management solutions.
2. **LAY DOWN THE LAW.** Implementing business-practice rules and strictly enforcing them goes a long way toward safeguarding the value of the data you collect. If you set rules and practices for all people entering data into your systems, you'll end up with a more coherent system that is easier to manage and validate. "It's important to make sure there is a common vocabulary and that there are business rules and governances on data," says Laurie Collins, project strategist for the [Schools Interoperability Framework Association](#). "Who is doing what? That's something districts haven't really thought about. We're now seeing that that changes because of the demand for quality data. You need to make sure you are engaging people at all levels of the administration— from leaders to data entry clerks—as well as considering the other applications that feed the data warehouse."
3. **KNOW YOUR AUDIENCE.** There are often several different audiences within a school district or state that will view and use data from your data warehouse: researchers, statisticians, teachers, principals, and others. Customizing the user interface can help ensure that each group gets what it needs and understands what it is looking at. For example, parents have different needs, different skills, and different technology available to them than statisticians at a central office.
4. **PLAN AN INTERNAL MARKETING CAMPAIGN.** Keep your eventual end users in the loop as your project moves along. Doing a little internal marketing can build readiness and help get the buy-in you'll need in order to make your new system successful. Be sure to educate users on what you are doing, why you are doing it, how it will make a difference in their work, and so on.
5. **CREATE A FEEDBACK LOOP.** Creating a feedback loop for those who are collecting and using data can help ensure cleaner data is collected. "Once educators start using the data and seeing where there are problems with it, and that there are things they did with the data where they were incomplete or inconsistent, then they have a reason to improve it," says Shawn Bay, founder of [eScholar](#). "All the methodologies and technology for improving quality of data are great, but the most important is that educators actually have a reason to do it. When you create that feedback loop, and users see that they are getting something valuable that they can use if the data is clean, then they suddenly have a

- reason to get it clean the first time—or clean it up if they messed up when entering it initially.”
6. **CONSULT YOUR PEERS.** Enough school districts have implemented data warehouses that best practices have been established, and states and districts can confer on issues they run into and learn from one another. Don’t be afraid to ask.
 7. **MAKE SURE YOUR INFRASTRUCTURE IS UP FOR IT.** A large data load can put your computing infrastructure to the test. “Stress testing is definitely advisable for school districts prior to implementing a solution like this,” says Katherine Conoly, CIO of the Corpus Christi Independent School District in Texas.
 8. **DON’T OVERLOOK THE IMPORTANCE OF PROFESSIONAL DEVELOPMENT.** Thoughtful professional development that accounts for the various levels of technology adoption cannot be underemphasized. “Not all users are ready to take on a new technology system,” Conoly says. “Be sure to take into consideration the laggards as well as the innovators.”

The data is stored in a structure, often called a data model, created in a relational database. This data model is important. If it is not well designed, a school or district may be stymied in its efforts to analyze the data. Also, if the data model is strictly fixed, the district may not be able to add the custom data that it wants to analyze.

But there’s much more to data warehousing than just a database. Data must be collected, transformed, verified, and analyzed in order to be useful for decision-making. Ultimately, a data warehouse readies data for analysis, wherein the greatest benefits are derived.

“A data warehouse alone doesn’t help anybody do anything,” says Peter Waldschmidt, CTO for [TetraData](#), a South Carolina-based provider of education solutions. “It’s the analysis portion that provides the value.”

“When you can look at data longitudinally and at a detailed level, you begin to see trends that teachers can take action on and that they wouldn’t otherwise see,” says eScholar’s Bay. “To see trends emerging ahead of time and to implement corrective action early allows teachers to provide much faster and more effective student intervention.”

Data Warehouses at Work

McKinney Independent School District in McKinney, TX, realized that the volume of data it was collecting was continuing to grow and needed to be consolidated in one location so it could be effectively analyzed and used to improve instruction. “We had many systems for storing data throughout the district,” says Joe Miniscalco, senior director of secondary education at McKinney ISD. “Teachers had spreadsheets on their desktops. We had our student information system and our assessments. We needed all of this data to be clean, in the same format, and in one place so we could get a 360-degree view of a student.”

McKinney implemented a data-warehouse reporting and analysis solution from eScholar. Today, the district stores all of its data in a comprehensive warehouse, which has allowed it to take a more proactive approach in academic intervention and student preparation.

“Our administrators are now able to apply filters to isolate specific cross-sections of student data,” Miniscalco says. “By following students over time, looking across many data domains, we are able to have more effective student placement and early identification of students needing intervention. We are also able to predict staffing needs and prepare administrators for anticipated challenges.”

A data warehouse component is included in the [SchoolNet](#) instructional management solution recently implemented by Katherine Conoly, CIO of the Corpus Christi Independent School District. Conoly says that many schools she’s talked to are looking at employing data warehouses primarily as tools for managing data requirements prescribed by the No Child Left Behind Act. “I think many school districts are looking for a solution for compliance rather than instruction, and that’s unfortunate,” she says. “At **Corpus Christi ISD**, we are trying to create a culture where we are using data for continuous improvement. We are bringing different data sets together so they can give us an indication of why certain students are not performing.”

Conoly says the instructional management solution allows her district to see where it is, where it’s been, and where it’s going. “But most importantly, it gives teachers the power to chart a course—to see what they’ve missed and to ensure they have covered all the bases in lesson planning and instruction.”

Managing Implementation

Because data warehouses were traditionally developed for large-scale corporate implementations, adapting them to smaller-scale school districts can be tricky. At first, districts were forced to buy custom-built data warehouses, which carried high price tags.



A [data management solution] gives teachers the power to chart a course—to see what they’ve missed and to ensure they have covered all the bases in lesson planning and instruction.

Katherine Conoly, Corpus Christi Independent School District

“In K-12,” eScholar’s Bay says, “we realized pretty quickly that custom-building data warehouses wasn’t going to work for the majority of districts because the cost was high, and school districts for the most part are pretty small.”

In response to this, Bay says, eScholar has concentrated on establishing data standards that can be reapplied over a great number of districts. “For us, it’s worked out great,” he says. “Our users can use what is in effect the same data warehouse to meet their unique needs, and we are spreading the development costs across a huge number of school districts.”

Waldschmidt says TetraData takes a similar approach. “We standardize large portions of the process so all our customers have the same type of support requirements, yet each school has an individual, custom warehouse,” he says. “We’ve tried to do a little bit of both—gain the efficiencies while still allowing a school or district to have a custom data warehouse. If you are looking for maximum efficiency and low cost, you have to accept

standardization to some degree, but there is an opportunity to expand into custom warehousing if the needs and requirements are there.”

Planning Is Key

School leaders who are considering a data warehouse solution should expect to spend significant time planning prior to jumping in. Planning should include making decisions about both “understructure” issues, such as data cleansing and refresh frequencies, as well as analytical issues. For example, are you going to deliver data to the teachers through a portal? Are you going to employ advanced analytical tools that will allow for longitudinal analysis and the ability to track students through different situations, through time, and through mobility?

“There are really two different sectors to the solution—the analysis portion and the whole data-warehouse portion—and they really can’t be designed independently of one another,” says Waldschmidt. “You have to design a data warehouse so it can be analyzed. Similarly, your analytical tools have to understand the structure of the data and be well suited to that structure in order to be effective. And you have to understand that reviewing data accuracy and processes will be ongoing tasks.”

Miniscalco at McKinney ISD knows this all too well. While McKinney has experienced many successes through its increased application of technology, Miniscalco is quick to point out that there’s more work to be done. The district is currently considering an enhanced student information system and reviewing location-level practices focused on data accuracy. By providing its teachers with direct access to data, McKinney hopes they will use that data to drive instruction, and in turn identify additional areas where technology can increase their effectiveness.

“Our district is focused on getting the right tools in the hands of those who need them,” says Miniscalco. “We want to seamlessly integrate technology to provide a better education for all our students.”

“Our district is focused on getting the right tools in the hands of those who need them,” says Miniscalco. “We want to seamlessly integrate technology to provide a better education for all our students.”

“Our district is focused on getting the right tools in the hands of those who need them,” says Miniscalco. “We want to seamlessly integrate technology to provide a better education for all our students.”

:: web extra :: For more information on this topic, visit www.thejournal.com. In the Browse by Topic menu, click on **Data Management**.

Justine Brown is based in Cool, CA, and specializes in writing about technology, education, and government.