TDWI E-Book

Data Warehouse Automation: Accelerating Business Results

Sponsored by:

TIME**X**TENDER

tdwi.org

tdwi

Advancing all things data.

Q&A: HOW DATA WAREHOUSE AUTOMATION STREAMLINES SELF-SERVICE BI

Heine Krog Iversen, founder and CEO of TimeXtender, explains why data warehouse automation makes it easy to incorporate new and changing data sources and keeps existing reports consistent and correct, all while enhancing the business/IT relationship.

TDWI: What is data warehouse automation (DWA)? What does it do and what problems does it solve?

Heine Krog Iversen: DWA is an emerging category that describes the ways progressive organizations can implement, support, and operate the data access and data flows needed to support business decision making in an agile fashion. The *automation* part solves a set of fundamental problems insofar as it is now possible to establish a data warehouse in a fraction of the time, and at a fraction of the cost, than previously possible with traditional approaches. This translates directly to reduced time to value and enables business users to access, visualize, and act on relevant data in record time.

Businesses employing DWA can make decisions rapidly and easily add and modify new data sources, allowing them to stay ahead of the ever-changing market. With the efflorescence of data sources and exponential increases in the volume of relevant data to ingest, DWA is a necessary part of any business-friendly data strategy.

What are the best ways for an enterprise to incorporate new and changing data sources?

Being able to incorporate new and rapidly changing data sources is a key benefit to DWA. Traditionally, a developer would only be able to access a new or modified data source via writing code, which is a time-consuming, tedious, and labor-intensive process. With DWA, a data analyst can access, view, analyze, and include data via an intuitive interface and thereby quickly and efficiently enhance the business. This can be done in minutes, not hours or days, and can be done in increments, which allows for agility and rapid, on-the-fly

decision making. New and changing data sources are a given of the system in today's business world and DWA allows a business to quickly assimilate them.

How can business and IT work closer as a team on data analytics?

Business-IT harmony is a crucial part of organizational success, especially given the importance of IT in the modern enterprise. Too often, however, there is constant tension between IT and business users; DWA can act as a glue, a harmonizing factor bringing business and IT together. With DWA, business users can focus on the business decisions they need to make supported by the data that is relevant to the decisions at hand and IT can ensure that the data is accessible in the form needed and is secured via policies and credentials, thereby helping the business reduce time to decision.

Incorporating new data sources can make an enterprise more agile, but how should an enterprise make sure that there existing reports still work after IT upgrades an existing data source or adds a new one?

Although changing data sources and thus changing reports are inherent to all businesses, it is crucial for users to be able to access existing or historical reports even if the data sources change. Moreover, because most reports are drawn from amalgamated data from several data sources, this matter becomes complex.

Traditionally, changes in a given data source and the underlying structures mean that reports relying on that data source need to be updated and changed accordingly. Take, for instance, sales reports. Typically, they draw from different data sources including A/R, CRM, geography, segment, and pipeline repositories. This data changes frequently, so all sales reports need to be updated and maintained when a change occurs. With DWA, changes in data sources are made in the data warehouse, not in the reports themselves, which means that the amount of work maintaining the integrity of the reports will be almost completely eliminated.

Adding and modifying existing data sources is one challenge. The other is how to secure access to historical data when the data source is no long available. What do you recommend?

The value of historical data is often overlooked. Though there is no silver bullet in using historical trends to predict the future, historical data has significant value, and any business should take measures to secure and store data about its history.

Many times this is deemed unachievable due to technology challenges. For instance, when a company decides to replace a legacy line-of-business (LOB) system (such as the ERP) with a new one. Traditionally, a lot of time and effort is spent on data conversion and establishing historical data or reference data. This work involves significant analysis and mapping of old structures and master data to new ones, and then the old data is converted and loaded into the new system as reference data. This traditional approach forces business to choose how much historical data they can afford to keep because the efforts to convert are significant, and there is a limit to how much history can be stored in an LOB system.

These efforts can be completely avoided and eliminated if a data warehouse is implemented. All historical data will be stored here, and new data from the new system(s) will be loaded into the warehouse on an ongoing basis.

This way the "one version of the truth" is maintained, secured, and easily accessed both for historical and current data, as well as any future data required.

RELIEVING THE PAIN OF THE BI BACK ROOM WITH DATA WAREHOUSE AUTOMATION

By Dave Wells

Data warehouse automation is more than simply automation of ETL development. It automates the entire data warehousing life cycle from planning, analysis, and design through development and extending into operations, maintenance, and change management. Dave Wells explains the benefits of the technology.

As the business-facing capabilities of BI—dashboards, scorecards, analytics, etc.—become more demanding and the data sources more diverse and complex, it is increasingly difficult for the data warehouse to keep pace with new and changing requirements. The common and long-standing problems with data warehouses are that they take too long to build, they cost too much to build, and they're too hard to change after deployment.

Increasing expectations on the front-end of BI combine with expanding data sources at the back-end to amplify these problems. The data warehouse is caught in the middle and simply can't respond quickly enough using traditional methods of development, operation, and maintenance. Fortunately there is an alternative to conventional methods. Data warehouse automation—a relatively mature but underutilized technology—is a proven and effective way to resolve the top three challenges of data warehousing.

What is Data Warehouse Automation?

Data warehouse automation uses technology to gain efficiencies and improve effectiveness in data warehousing processes. Data warehouse automation is more than simply automation of ETL development. It automates the entire data warehousing life cycle from planning, analysis, and design through development and extending into operations, maintenance, and change management.

Adoption of data warehouse automation changes the way we think about building data warehouses. The widely accepted practice of extensive up-front analysis, design, and modeling is left behind as the mindset changes from "get it right the first time" to "develop fast and develop frequently." This approach fits neatly with agile development practices, yet data warehouse automation doesn't demand that you go agile. You can achieve substantial speed, quality, and cost savings without fully embracing and implementing an agile methodology.

Automation in data warehousing has many of the same benefits as in manufacturing:

- Increased productivity and speed of production

- Reduction of manual effort

- Improved quality and consistency

- Better controls and process optimization opportunities

The manufacturing parallel holds true when building a data warehouse; we can think of it as an information factory. However, data warehousing is more complex than product manufacturing. Manufactured products are typically delivered to a consumer and the job is done. Data warehouses must be sustained through a long life cycle where changes in source data, business requirements, and underlying technologies are ongoing considerations. Automation helps to implement the right changes in the right ways and as quickly as they are needed.

The Foundation

Design patterns, standards, metadata, and reuse are the cornerstones of data warehouse automation. Data warehousing offers the opportunity to apply many design patterns including:

• Architectural patterns such as hub-and-spoke, bus, and hybrid architectures

• Data structure patterns including normalized, de-normalized, and multi-dimensional

• Data management patterns such as those for key management and time variance

• Data integration patterns including ETL, ELT, virtualization, and federation

Design patterns can be coupled with architecture, design, and implementation standards and best practices to build reusable data warehousing components. The patterns, standards, and components are captured and described as metadata. Then each use of components and the relationships among them are described with more detailed and implementation-specific metadata. A good data warehouse automation platform makes it easy to cross the gap from *reusable* to *reused*.

The Technology

There are several data warehouse automation tools in the marketplace—some embedding automation functions into a broader tool suite and some built specifically for pure-play automation. I occasionally meet people who have built their own data warehouse automation functions—an approach that is certainly possible but often impractical and not the best use of developer time and skills. In general, the pure-play automation tools fall into two categories:

• *Open tools* where the generated data warehouse schema and processing logic uses standard SQL and scripting languages that can be viewed and even directly changed by developers.

• *Proprietary tools* where the generated data warehouse components are "black box" and capabilities to view and change are provided entirely and exclusively through the tool.

Dave Wells is a consultant, mentor, and teacher in the field of business intelligence. You can contact the author at dave_wells@earthlink.net.

DATA WAREHOUSE AUTOMATION DOES MORE THAN JUST ELIMINATE TEDIOUS WORK

From reducing risk to slashing development time and freeing up talent to work on generating business value, DW automation can have a major impact on managing your data warehouse.

The data warehouse isn't perfect. The traditional tools and methods we use to build and manage data warehouses are *far* from perfect. They produce data warehouse systems that take far too long to build, are too hard to manage, are difficult to change, and cost too much to operate. This isn't something that's endemic to the warehouse itself. Instead, it's a function of obsolete tools and outmoded processes. Data warehouse automation (DWA) aims to change this.

DWA refers to the category of tools used to automate and accelerate the process of designing, building, deploying, and managing data warehouse systems. It also describes an agile, business-driven methodology for building, deploying, and changing data warehouse systems. Proponents say data warehouse automation can significantly reduce costs and radically accelerate the pace of reporting and analytic development. What's more, they argue, DWA tools and methods permit organizations to adroitly respond to changing conditions, even disruptive events—such as merger and acquisition activity, expansion into new regions, the launch of new product lines, etc.—that may require drastic changes to the warehouse itself.

"The business is asking for data much faster than the classic approach of building a data warehouse can possibly respond to," argues Heine Iversen, CEO of TimeXtender, a data warehouse automation vendor that focuses on Microsoft's SQL Server database platform. "Data warehouse automation brings agility and adaptability to business intelligence and analytic development while ensuring data governance, data quality, and data security. It exposes warehouse design and development in a structured context that makes sense to the business users.

"The business doesn't always know exactly what they're looking for, but by working with [developers] and asking questions, they can quickly figure it out. To support this kind of [question-driven] experience, you need a very agile approach, a very agile platform that's more like modeling than development."

Minimizing Risk while Making the Most of What You Have

Iversen and other proponents stress that data warehouse automation is not a strategy for replacing human beings. Instead, DWA tools and methods should be used to automate the most repetitive, tedious, or time-consuming aspects of warehouse development, starting with the creation and maintenance of documentation, which is every developer's least-favorite activity.

DWA tools likewise automatically generate metadata, track and monitor data lineage, generate optimized SQL code (in TimeXtender's case, SQL Server-specific T-SQL), and exploit platform-specific features or optimizations, such as Microsoft's SQL Server Integration Services (SSIS).

In this way, Iversen argues, DWA helps to mitigate risk. Tool-generated documentation and metadata definitions are always consistent and maintained. Automated code generation produces standard, consistent SQL code and eliminates the risk of developer mistakes. Tool-driven orchestration and management make it possible to back out of (or to step back through) changes.

Data warehouse automation tooling provides some degree of platform abstraction, too. DWA tools generate structures that are optimized for specific platforms and they make it easier to move from one platform or from one *version* of a platform to another. In TimeXtender's case, this means an organization running, say, SQL Server 2008 will be able to automatically upgrade that warehouse to SQL Server 2016. The DWA tool handles feature dependencies and deprecations and also makes intelligent choices about which new features or functions to exploit.

The rationale for automating isn't just to minimize risk or to eliminate manual human dependencies, Iversen says. Rather, he argues, automation frees up human talent to focus on "value generation."

"Instead of writing code, you build a metadata model that has proven to cut out roughly 80 percent of the manual workload in building and maintaining your data warehouse. This [manual human] expertise can be better used [elsewhere]. There's a very high demand for skilled people in the business who can analyze data, who understand data models and data modeling, who understand how data relates [across different systems and domains]. We can take the resources that we use today to prepare data and to do all of these manual, time-consuming things—we can take them and move them into the business and have them spend much more valuable time building out the analyses and reports that the business needs to remain competitive," he argues.

Surfing (Not Foundering in) the Big Data Wave

Big data is as much a function of data source heterogeneity as it is of increasing (or exploding) data volumes. Data warehouse automation doesn't just help with the big data deluge, Iversen

argues; it's essential for managing it. "What we see today is that the number of data sources is growing day by day. *Everything* today is a data source depending on your industry and what you're doing. Google Analytics is a data source. Twitter is a data source. Facebook, too. Organizations are actually bringing this data into the warehouse today and combining it with data from CRM systems."

The goal is the elusive "Big Picture" view of customers, products, and business operations. The rub is that organizations have too few resources with which to develop, interpret, and "frame" this picture.

Self-service data preparation tools can provide some relief, but data scientists and business analysts can't be expected to acquire, prepare, and produce cleansed, consistent data sets for *an entire organization*. In most cases, they're busy preparing data for use in their own, one-off, idiosyncratic, esoteric analyses. The data warehouse, by contrast, is an engine for reliably managing and delivering the information that drives day-to-day decision making in an organization. Managed, repeatable, and governed data integration workflows are essential for producing this information. "IT has the people with the bigger and better understanding of how to tap into these big data data sources and actually pull out the relevant information and put it into the data warehouse in a structured way. Data warehouse automation can standardize these data integration processes," Iverson explains.

To the extent that DWA helps to eliminate manual, tedious, repetitive, and time-consuming activities, he argues, it frees up IT practitioners to work on profiling, integrating, and managing data from big data (i.e., non-traditional) sources. "A lot of the work in the old-fashioned way of doing data warehousing is about optimizing your load, optimizing your ETL processes, and optimizing your load times. From our perspective, that is something you can fully 100-percent automate. We use machine learning [to help with this]. We can have the data warehouse constantly monitor the load, constantly monitor how many threads can be used against specific data sources, and so on," he says.

Data warehouse automation technologies can also be used to automate—or, at the very least, to accelerate—the orchestration of integration data flows from big data sources, Iversen points out. These include streaming events or messages. Typically, this data is bulk-loaded in batch operations or rapidly loaded at intervals of several minutes to a few seconds. At present, the mechanisms for

ingesting, transforming, and persisting streaming data are largely script-driven.

Data warehouse automation tools give you a means to manage, monitor, and audit—in other words, *to govern*—these processes. "You need to have an automated way of adding all of these data sources very quickly and in a very agile way to the platform," Iversen points out.

"You need to much more frequently refresh than you did five years ago. If you look at what we're able to do with an automated data warehouse, we can constantly load data [in] microbatch into the data warehouse so that the information analysts are seeing in an OLAP cube is actually newer than [the information they can get] from a report directly against an ERP or CRM system. It's pre-aggregated, it's pre-calculated, we're just constantly bringing in new data."

Self-Servicing Users the Right Way

By any reasonable standard, self-service BI tools give short shrift to critical data management issues such as data governance and information security. The situation is improving, but compared to the data warehouse-driven BI model, self-service tools tend to have impoverished metadata management and data lineage tracking capabilities, to cite two prominent shortcomings.

To a degree, these issues are baked into the self-service BI model, which eschews the data warehouse and enables users to extract data directly from ERP, CRM, and other upstream sources. In practice, this has a non-trivial impact on the performance of these upstream sources.

By contrast, data warehouse architecture consolidates information from multiple sources into a single system and staggers batch extraction jobs in order to avoid such problems. Data warehouse automation tools schedule, monitor, and audit these batch jobs as well as perform other critical tasks. For example, DWA tools automate and standardize metadata management and data lineage tracking, ensuring that metadata definitions are updated and consistent and that lineage—i.e., how data changes from the point of ingestion to any subsequent updates and manipulations—is automatically tracked and recorded. By automating warehouse operations and accelerating the rate at which new data sources can be identified, integrated, and managed, data warehouse automation technologies have the potential to make self-service BI even better—or so proponents claim.

"In the future, IT will have the responsibility to supply the data warehouse, the nerve center, with quality, cleansed data in a governed manner. This is and will remain IT's primary role. However, to a large degree, IT is not going to be involved in building out the analyses and building out the reports. The business users can do this themselves because they have these self-service [BI] tools. The difference is that [with the data warehouse] they have a solid foundation on which to do it," Iversen says.

In this self-serving future, IT still has a critical role to play, Iversen asserts.

For one thing, IT will continue to develop and support the data warehouse, integrating new sources of data and instantiating new warehouse structures as business users identify them. IT will likewise continue to build and manage repeatable, high-volume reports and analytical assets (such as scorecards and dashboards), although in many cases, these will first have been developed by business users. By instantiating them (as stored procedures in the data warehouse), they can be disseminated to a much bigger pool of potential consumers. Finally, IT will have its hands full with big data technologies, which lack meaningful self-service features, to say nothing of management-automation capabilities. For these and other reasons, Iversen argues, data warehouse automation isn't a luxury, it's a *necessity*: IT must automate or fail in its mission to support the business.

"You will realize cost reductions in operations because you're automating most if not all of the work [that human beings used to have to do] to make sure that everything is just running. This [work] includes manual tasks, such as scanning logs to find out how and where a job failed, writing scripts to fix new problems, or generating documentation," he comments.

"You can actually cut 80 percent of that [time-consuming] workload every time you want to add a data source. You're going to need that [labor that you save] to deal with the growing number of data sources, the growing volumes of data, the requests from the business to integrate more and different types of data. The real savings will be the ability to do more with the resources you have."

TIMEXTENDER

www.timextender.com

TimeXtender is a world-leading data warehouse automation vendor dedicated to Microsoft SQL Server.

The TimeXtender software, TX DWA, revolutionizes the way a data warehouse is developed and maintained by automating all manual data warehouse processes—from design to development, operation, and maintenance to change management. TimeXtender ensures an improved and inexpensive solution that is fully documented.

The TX DWA software enables medium and large data-driven enterprises to get business intelligence done faster, more efficiently, and with less stress by providing "one truth" to improve decision-making processes and overall business performance that reduces costs and saves valuable time. TX DWA turns a business intelligence project into a business intelligence process with a flexible solution that expands as the business evolves and grows.

TimeXtender collaborates with VAR and OEM partners across six continents, providing more than 2,600 customers in 61 countries all over the world with its advanced data warehouse automation software.

Why wait days before taking action? With TimeXtender, data is available at your fingertips in mere hours!

- **What Can DWA Do For You?**
  Sign up for the report from The Eckerson Group

- **Why DWA?** (Infographic)

tdwi

Advancing all things data.

tdwi.org

TDWI is your source for in-depth education and research on all things data. For 20 years, TDWI has been helping data professionals get smarter so the companies they work for can innovate and grow faster.

TDWI provides individuals and teams with a comprehensive portfolio of business and technical education and research to acquire the knowledge and skills they need, when and where they need them. The in-depth, best-practices-based information TDWI offers can be quickly applied to develop world-class talent across your organization's business and IT functions to enhance analytical, data-driven decision making and performance.

TDWI advances the art and science of realizing business value from data by providing an objective forum where industry experts, solution providers, and practitioners can explore and enhance data competencies, practices, and technologies.

TDWI offers five major conferences, topical seminars, onsite education, a worldwide membership program, business intelligence certification, live Webinars, resourceful publications, industry news, an in-depth research program, and a comprehensive website: tdwi.org.