



PULSE

Accelerating Data Warehouse Development

Best practices and tool automation
for faster business consumption
of analytics

By Philip Russom

Sponsored by:

WhereScape[®]
Data Automation

tdwi
Transforming Data
With Intelligence™



Accelerating Data Warehouse Development

Best practices and tool automation for faster business consumption of analytics

By Philip Russom

Table of Contents

- Executive Summary 3
- Drivers for the Acceleration of DW Development 4
- The Importance of Accelerating DW Development. 5
- Forms of DW Automation and Acceleration 7
 - Enablers of DW Automation 9
- Best Practices for Accelerating DW Development 10
 - Myths that Need Busting. 10
 - Real-World Success Stories 12
- Final Thoughts and Recommendations 13
- About Our Sponsor 14

© 2021 by TDWI, a division of 1105 Media, Inc. All rights reserved. Reproductions in whole or in part are prohibited except by written permission. Email requests or feedback to info@tdwi.org.

Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies. Inclusion of a vendor, product, or service in TDWI research does not constitute an endorsement by TDWI or its management. Sponsorship of a publication should not be construed as an endorsement of the sponsor organization or validation of its claims.

This report is based on independent research and represents TDWI's findings; reader experience may differ. The information contained in this report was obtained from sources believed to be reliable at the time of publication. Features and specifications can and do change frequently; readers are encouraged to visit vendor websites for updated information. TDWI shall not be liable for any omissions or errors in the information in this report.

About the Author



PHILIP RUSSOM, Ph.D., is senior director of TDWI Research for data management and is a well-known figure in data warehousing, integration, and quality. He has published more than 600 research reports, magazine articles, opinion columns, and speeches over a 20-year period. Before joining TDWI in 2005, Russom was an industry analyst covering data management at Forrester Research and Giga Information Group. He also ran his own business as an independent industry analyst and consultant, was a contributing editor with leading IT magazines, and a product manager at database vendors. His Ph.D. is from Yale. You can reach him at prussom@tdwi.org, [@prussom](https://twitter.com/prussom) on Twitter, and on LinkedIn at [linkedin.com/in/philiprussom](https://www.linkedin.com/in/philiprussom).

About TDWI Research

TDWI Research provides industry-leading research and advice for data and analytics professionals worldwide. TDWI Research focuses on modern data management, analytics, and data science approaches and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of business and technical challenges surrounding the deployment and use of data and analytics. TDWI Research offers in-depth research reports, commentary, assessments, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

About TDWI Pulse Reports

This series offers focused research and analysis of trending analytics, business intelligence, and data management issues facing organizations. The reports are designed to educate technical and business professionals and aid them in developing strategies for improvement. Research for the reports is conducted through surveys of professionals. To suggest a topic, please contact TDWI senior research directors Fern Halper (fhalper@tdwi.org), Philip Russom (prussom@tdwi.org), and David Stodder (dstodder@tdwi.org).

Acknowledgments

TDWI would like to thank many people who contributed to this report, including, our report sponsors, who reviewed outlines and report drafts. Finally, we would like to recognize TDWI's production team: James Powell, Lindsay Stares, Rich Seeley, and Michael Boyda.

Sponsor

WhereScape sponsored the research and writing of this report.

Executive Summary

Data Warehouse Professionals Must Develop More Data Solutions than Ever, Faster than Ever.

The pace of business continues to accelerate as enterprises evolve to keep up with quickly changing pandemic surges, economic roller coasters, political churn, social issues, competition, emerging technologies, and new ways of managing the modern digital enterprise. In turn, IT is modernizing continuously to support the evolving and accelerating enterprise. Data warehouses and other analytics have become priorities for understanding and predicting change as well as embracing change successfully for organizational advantage.

Given the speed of change and the urgency of managing change properly, data warehouse and analytics professionals are under renewed pressure to develop new data sets and insights—and develop them in shorter time frames than ever. In other words, data warehouse professionals must scale to handle continuously growing numbers of data solutions while also delivering each solution in record time.

As modern business evolves and accelerates, data warehouse development methods must, too.



TDWI is confident that data warehouse professionals can rise to the challenge of *accelerating data warehouse development* as long as they have:

- **Automation and high ease-of-use** supported in development tools and data platforms that are purpose-built for data warehousing. After all, tool automation is a key enabler of developer acceleration.
- **Tools conducive to modern data development methods**—namely agile, lean, virtual, and other methods known to be fast, flexible, and efficient in rapid prototyping environments.
- **A skilled data warehouse team** that knows the appropriate methodologies for accelerating data warehouse development. Tools are important, but team competency is an equally critical success factor for accelerating data warehouse development.

TDWI recently ran a new survey about accelerating data warehouse development. The survey asked respondents about their organizations' current and future practices for data warehouse development. The point of the survey was to identify and quantify trends that can help technical and business users automate and accelerate the process of data warehouse development, the ultimate goal being to satisfy new and highly accelerated business requirements for data. Findings from that survey are presented in this report.

This report will now drill into survey responses that illuminate key areas for the acceleration of data warehouse (DW) development:

- Drivers for the acceleration of DW development
- The importance of accelerating DW development
- Forms of data warehouse automation and acceleration
- Best practices accelerating DW development
- Final thoughts and recommendations

Drivers for the Acceleration of DW Development

Most data warehouse professionals are experiencing numerous changes at the moment. What are those changes, what is driving them, and what are the desired improvements? To identify and quantify these issues, this report's survey asked: What's driving your need to improve data warehousing? (See Figure 1.) Drivers for the acceleration of DW development fall into three groups: development, business, and technology.

Data development drivers. To achieve acceleration and other goals, data warehouse teams are adopting modern methodologies and best practices. These are the highest priority for change in data warehouse development, as witnessed by the long bar at the top of the chart in Figure 1. The majority of survey respondents say agility is needed—in the sense of speed, productivity, flexibility, and innovation (70%). A number of other modern development methods are applied hand-in-hand with agile methods, namely scrum, lean, and rapid prototyping. Agile and other new methods help a data warehouse team achieve realignment so the DW can support new business directions (33%).¹

Business drivers. As businesses progress deeper into modern, digital, and data-driven practices, they are making greater and greater demands of data warehousing and analytics. For example, as the pace of business accelerates, the business demands data solutions delivered faster than ever before (55%). In a parallel trend, the business needs a broader range of analytics (47%), which is key to transformation, innovation, and competitiveness. Put those together and it becomes clear that businesses need a great number of new DW and analytics solutions (29%), thereby forcing data warehouse teams to become more agile, productive, and innovative. Yet, the business and the DW team are expected to contain costs while achieving improvements (33%).

Drivers for the acceleration of DW development fall into three groups:
development, business, and technology.



Data technology drivers. Roughly half of survey respondents say technology must process data faster for real-time analytics, integration, self-service, and data exploration (55%). This is easier said than done because supporting real-time use cases usually entails the acquisition and

¹ For more about evolving data team methods, read the TDWI Upside article "Modernizing Your Data Team and Its Best Practices," online at <https://tdwi.org/articles/2020/10/15/diq-all-modernizing-your-data-team-best-practices.aspx>

integration of additional systems that can capture streaming data and process it in subsecond time frames—as required for real-time analytics, operational reporting, and business monitoring. Other system acquisitions are needed to leverage new platforms, tools, and other modernization (37%), which often involve cloud data platforms, toolsets for DW automation, and tools purpose-built for self-service and analytics. Everyone knows technology must scale to the massive volumes of big data, but this is a relatively low priority for respondents (22%) because most organizations already have the platforms and the skills needed to deal with data at scale.

The need for speed. A common element inherent in most of the drivers just discussed is *speed*. In the context of accelerating data warehouse development, speed has multiple meanings:

- Fast and agile development methods
- Short time to use for business users so the business, not just the DW, is agile
- Quick processing of time-sensitive data for real-time analytics and operational reporting

What’s driving your need to improve data warehousing? Select five or fewer answers.



Figure 1. Based on 123 respondents.

The Importance of Accelerating DW Development

As we saw in Figure 1, there are many reasons to accelerate data warehouse development. However, doing so is not trivial; is it really worth the effort? To quantify the answer, this report’s survey asked: How important is acceleration via automation for the success of your data warehouse program? (See Figure 2.)

The vast majority of survey respondents consider DW acceleration important. In fact, almost two-thirds (63%) find it extremely important. Additionally, more than a quarter (28%) find it moderately important.

Conversely, almost no one considers acceleration unimportant. A mere nine percent feel that it is not a pressing issue currently.

How important is acceleration via automation for the success of your data warehouse program?

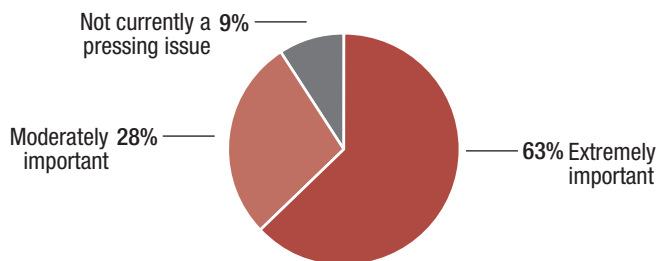


Figure 2. Based on 123 respondents.

Acceleration and automation in development are clearly important to most DW professionals, but in which situations? What are the benefits for the DW team and the whole organization? To gather answers and ideas, this report’s survey asked respondents an open-ended question: In your own words, why is accelerating DW development important (or not)?



DW professionals list many good reasons for accelerating DW development.

Figure 3 reproduces 10 representative quotations drawn from text responses entered by 79 survey respondents. As you can see, there are many good reasons for implementing DW acceleration and automation, and these reasons reveal a number of benefits for data warehousing and business use cases. Note that the positive views expressed by respondents range across different industries and geographies, plus many job titles. In other words, the needs and benefits of DW acceleration are nearly universal—as well as top of mind for DW professionals.

In your own words, why is accelerating DW development important (or not)?

- “Our BI backlog is long and ever-expanding. Without some level of accelerating DW development—especially the repeatable mundane tasks—we will never succeed in realizing all the organization’s BI outcomes.” Data architect, insurance, U.S.
- “Business change happens fast. The DW needs to move in sync or the DW-driven data is out of date and/or not related to what the strategy and leadership team needs to be analyzing.” Data analyst, education, Australia
- “Anything repetitive is automatable and so should be, to reduce time and improve accuracy.” Senior product manager, internet, U.S.
- Accelerating DW development “reduces time-to-market. My stakeholders don’t want to wait for answers to their questions because IT is still implementing the data sets.” BI engineer, law, Europe

- “We use a pseudo-agile methodology for development, with quick sprints and iterations, knowing that everything may not be there the first go-around.” BI director, insurance, U.S.
- We need “to use BI resources efficiently to deliver on the promise of supporting a data-driven organization. DataOps is a key part of this.” Head of data and insight, media, Europe
- We can “meet business needs faster while keeping the same team size by utilizing tools that accelerate development.” Reporting and analytics director, insurance, U.S.
- “Our business clients are demanding access to new data sources constantly. So we need to quickly ingest the new data elements into our DW ecosystem for our data scientists and our business clients.” Data architect, financial services, U.S.
- “Better documentation and lineage—enforced by model-driven code generation—enables the ‘no manual coding’ for ingestion that’s required.” Data leader, advertising, Europe
- “If we use a traditional approach that’s developing for months on end, the business and its requirements will change, and we will be behind before we even finish.” BI director, insurance, U.S.

Figure 3. Based on text responses from 79 respondents.

Forms of DW Automation and Acceleration

This report’s survey asked: Among the following forms of DW automation and acceleration, which would be the most valuable to you and your organization? Judging by responses to this question, data warehouse professionals know the kinds of automation and acceleration they need and where to use them. (See Figure 4.) Here follows a discussion of their responses.

Forms of DW automation and acceleration fall under four categories: data quality, data integration, general development, and metadata. Note that all the functions discussed here (and others as well) can be accelerated by a modern toolset purpose-built for data warehouse automation.²

Data quality functions. At the top of Figure 4, the most selected answers involve data quality (DQ), which makes DQ a high priority for DW automation and acceleration. Data warehouse professionals want their DQ issues automatically recognized and remediated (64%), and they want automation for data validation and verification (51%). Potentially, software tool automation could help with other DQ functions such as standardization, cleansing, deduplication, profiling, and data monitoring.

Data integration functions. After DQ, data integration is a pressing priority. For example, process automation for onboarding new data sources (46%) is important for scenarios where new data sources arrive regularly, as they do with the Internet of Things (IoT), multichannel marketing, and business-to-business (B2B) partnerships. In these and other scenarios, data warehouse professionals and others need automation that works with various data structures and latencies (39%). Data integration work regularly involves time-consuming and somewhat tedious tasks; developers can be freed from much of that via modern automation for ETL code generation (43%) and automated mappings from sources to targets (41%).

² For detailed definitions of data warehouse automation, replay the TDWI webinar “Automation First as a Strategy for Data Warehouse Modernization,” online at <https://tdwi.org/webcasts/2019/09/diq-all-automation-first-as-a-strategy-for-data-warehouse-modernization.aspx>

General development tasks. Many DW professionals want a unified developer experience from their tools based on seamless development, testing, scheduling, and deployment (39%), combined with automation for various types of testing (28%) and all the other steps of the seamless development process. In a related area, agile and lean development methods openly discourage solution documentation because it is time-consuming, low-value, and impossible to keep up to date or accurate when manual. These problems are solved by tools that support documentation generation and automatic update (36%) and support versioning for change management (31%).

Forms of DW automation and acceleration can mostly be categorized under four areas: data quality, integration, general development, and metadata.



On the leading edge of automation, DW developers look forward to tools supporting intelligent assistants that guide developers (19%), possibly based on artificial intelligence and machine learning built into data development tools. The point of intelligent assistance is for the tool to make useful recommendations automatically for models, metadata, and transformations (30%). In some cases, tool automation can apply recommendations automatically without human intervention.

Metadata functions. Increasingly, DW professionals must work with data pushed to them from data streams, sensors, web applications, and so on. In some of these scenarios, the structures within pushed data evolve unpredictably, and data regularly comes from sources that do not expose metadata about their source data. Hence, there is a growing need for automatic metadata generation based on smart scans of source data (33%). In related trends, DW professionals also need more functions built atop metadata such as impact analysis (33%), business metadata (for self-service data access), data glossaries, data catalogs, and data lineage.

Among the following forms of DW automation and acceleration, which would be the most valuable to you and your organization? Select eight or fewer.

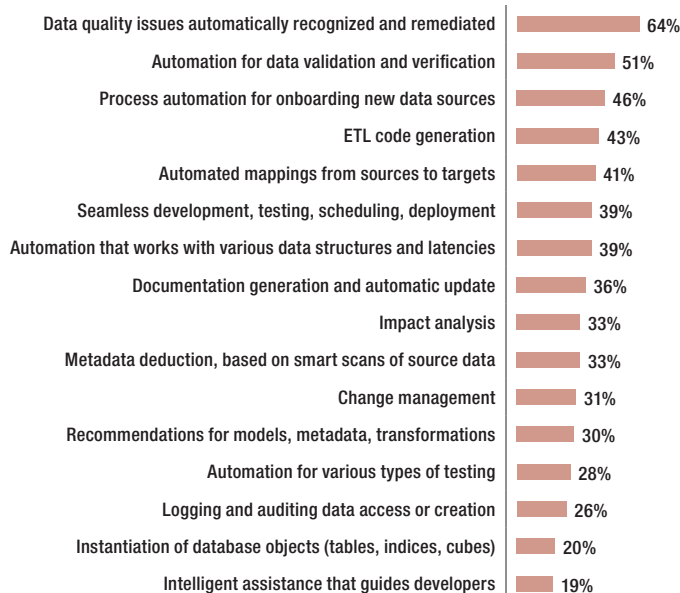


Figure 4. Based on 114 respondents.

Enablers of DW Automation

Our survey also asked: Among the following enablers of automation, which do you think would be the most appropriate to how you and your team work? Responses to the question (as seen in Figure 5) reveal that data warehouse professionals depend equally on both old and new enablers of automation. For example:

Business rules for user-defined automation. Some forms of automation for DW development are “black boxes” that don’t allow much customization. However, business rules (56%) continue to be a popular enabler for automation because they enable technical users to design their own unique logic for tool automation. In turn, the logic can trigger automatic responses in development, test, and production environments. There is a long tradition of business rules supported in tools for data quality, profiling, monitoring, ETL, and modeling. Modern cutting-edge technologies, such as machine learning, may also invoke business rules.

Older enablers for general productivity. Anything—old or new—that boosts a developer’s productivity can contribute to the overall acceleration of DW development. In this regard, there is still a clear need for thoughtfully designed user interfaces and well-organized, seamless tool environments. Hence, DW professionals still value drag-and-drop tools for design, modeling, and workflow creation (54%), high ease-of-use in graphical user interfaces (50%), and wizards and well-written dialog boxes (25%).

Newer enablers for automation based on embedded analytics. An emerging area for tool automation in data warehousing and other data disciplines concerns deep intelligence built into data management tools. Think of predictive analytics, which has become very powerful for predicting customer behaviors, churn, fraud, market directions, equipment maintenance issues, and so on. Imagine those techniques applied to predicting appropriate solutions for mapping data

sources to targets, remediating data quality problems, modeling data, and deducing metadata from data scans.

Survey respondents recognize that predictive analytics can predict successful design patterns (47%) as just described. They also know that artificial intelligence can turn correlations into recommendations (42%), as when the data domains discovered in a new data source should probably map to target systems known to manage those domains. We are approaching a time when data management tools will observe how successful technical users work, then apply machine learning that mimics developers' previous actions (31%). It's a brave new world for advanced automation in data warehouse tools and related systems. TDWI anticipates that this kind of intelligent assistance and predictive abilities will be commonplace in data-driven development within a few years.

Among the following enablers of automation, which do you think would be the most appropriate to how you and your team work? Select four or fewer.

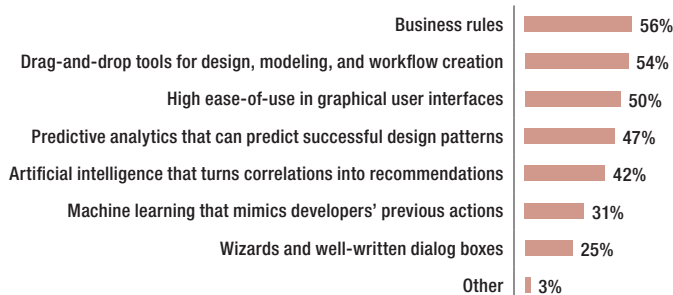


Figure 5. Based on 114 respondents.

Best Practices for Accelerating DW Development

Myths that Need Busting

As with so many disciplines in IT, the acceleration of data warehouse development suffers from a few myths and misconceptions. Beware, these myths can inhibit an organization's success or eagerness to adopt new tools and practices. This report's survey addressed the myths by telling respondents to select one "truth value" for each row of a table describing various points of confusion. (See Figure 6.)

The responses show that data warehouse professionals have a fairly clear and positive understanding of the strengths and weaknesses of DW acceleration and automation, as well as how these pan out in the practical world. For example, consider their responses to these statements from the survey:

Developers lose their jobs due to software automation. False! Over two-thirds of respondents (69%) believe this statement to be false, and they are correct. This is the most dangerous myth of software automation because caving in to this fear can prevent an organization from adopting progressive and modern practices.

On the contrary, software automation *helps* developers by:

- Accelerating the development process to deliver data solutions sooner
- Boosting developer productivity for greater numbers of solutions
- Freeing developers from many repetitive, manual, and administrative tasks
- Elevating developers into higher-level design tasks
- Enabling developers to adopt new and more agile development methods

Development is the only DW aspect that needs automation and acceleration. False! A whopping 83% of respondents feel this statement is false, and they are right. This is because a comprehensive toolset for data warehouse professionals will automate a wide range of tasks, including deployment, testing, monitoring, documentation, and administration—not just development. Furthermore, many of these tasks become more accurate, repeatable, compliant with data standards, and conducive to collaboration when automated by special software.



The acceleration of data warehouse development suffers from a few myths but still enjoys success in a broad range of use cases.

Automation tools are affordable compared to hiring and consultants. True! A third of respondents say this statement is definitely true (33%), while almost another third say it is maybe true (32%). Traditionally, data warehouse teams have increased the scale of solutions work by augmenting staff with new hires and/or consultants. Because payroll costs are usually higher than software licensing, acquiring modern tools is comparably affordable.

Automation inhibits creativity, flexibility, and unique customization. False! Even when a tool has its own approach to automating a task or process, there is still room for developers to add their own value and adapt tool usage to the organization’s unique requirements.

Automation only works with traditional data structures in moderate volumes. False! It depends on the tool, but most are adapting to the wide range of data structures, sources, types, containers, and latencies that are increasingly common nowadays, as well as the burgeoning volumes of big data.

Please select one “truth value” for each row of the following table.

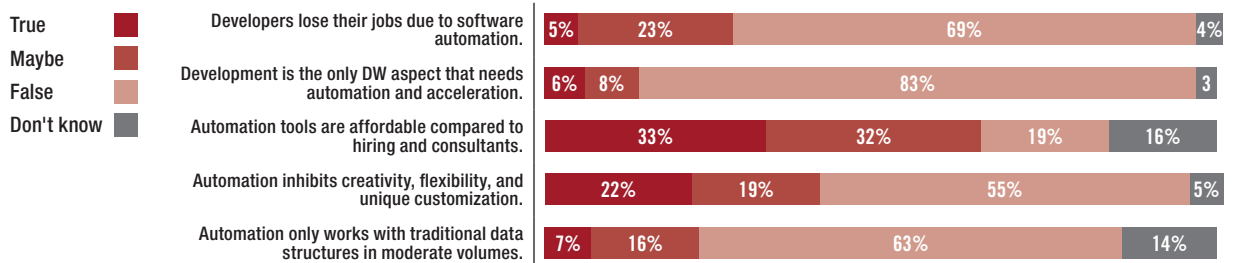


Figure 6. Based on 106 respondents.

Real-World Success Stories

This report's survey (once again) gave respondents an empty text field to express their views in an open-ended way. The survey told them: Briefly list some areas where data warehouse solutions succeeded in your organization due to tool automation. In response, almost all the DW professionals surveyed pointed to successful practices in DW automation, drawn from their real-world experiences. These successes range across many areas, scenarios, and specific tasks in data warehousing. They also cover many nations, industries, and professional job titles. (See Figure 7.) Hence, multiple forms of data warehouse automation are firmly established, and they are providing value to data teams and businesses.

Briefly list some areas where data warehouse solutions succeeded in your organization due to tool automation.

- Tool automation has helped us with “source-to-target mapping, orchestration, and workflow management.” Managing director of cloud services, financial services, U.S.
- “Using views and standing up a virtual DW rather than [a] traditional approach has sped up the process and enabled projects to deliver on expectations.” Data analyst, education, Australia
- We are now “able to do a lot with a small team.” VP of information management and analytics, financial services, U.S.
- “Validation for source files and the loading of DW tables.” Senior advisor, financial services, Canada
- “We have automated a backup and restore process for various DW processes.” Data architect, insurance, U.S.
- “Automation tools have helped self-service practices and thus resulted in faster decisions.” Applications manager, financial services, Middle East
- “Machine learning, advanced analytics, model scoring.” Enterprise data architect, insurance, U.S.
- “Documentation generation and updating.” Principal, consulting, Europe
- Via automation, “we have built repeatable load patterns. This has reduced our effort for developing ETL processes significantly.” Data architect, insurance, U.S.
- “We have had good success using [a vendor tool] to automate our import of data sources and the creation of star-schema data marts” automatically based on the imports. Senior DW developer, healthcare, U.S.

Figure 7. Based on text responses from 53 respondents.

Final Thoughts and Recommendations

Before closing, let's consider some recommendations that summarize this report's findings, but also indicate steps to be taken when accelerating data warehouse development:

Realize that DW development must keep up with an accelerating and evolving world. In particular, you must keep pace with and align to the changes your employer is reacting to or implementing at the business management level.

Despite the changes, do what data warehouses have always done: support business goals. You still have to support mature goals for business reporting, dashboards, and OLAP. However, you must also support modern goals for advanced analytics, big data, and real-time operations. If you take these goals and their requirements seriously, you will be compelled to accelerate and automate data warehouse development.

The Pulse: Data warehouse professionals must develop more data solutions faster than ever.



Depend on accelerated DW methods and automated DW tools. These will enable you to keep pace with business change, make the data warehouse relevant in a modern world, and modernize the DW and analytics in general.

Modernize your development methods, not just data platforms and tools. The highest priority should be the adoption of agile methods (originally designed for application development) that have been adapted to data-driven development. Note that agile is often supported by related methods for scrum, sprint, and rapid prototyping—and sometimes lean and self-service practices.

Broaden DW acceleration and automation to cover many data disciplines. It's not just the central data warehouse. Many supporting data disciplines around it need acceleration and automation, too, including data integration, quality, modeling, reporting, and analytics.

Adjust your DW software portfolio to include tools that support automation and agility. Success with the three A's—acceleration, automation, and agility—assumes modern tooling. Your best efforts for accelerating data warehouse development will be hamstrung without a proper toolset.



About Our Sponsor

WhereScape data automation software accelerates the design, build, documentation, and management of complex data ecosystems and ensures the delivery of data—wherever it is stored and however it is presented—to the people who need it.

Thousands of users worldwide leverage WhereScape automation to remove reliance upon repetitive aspects of data infrastructure projects and deliver data warehouses, vaults, lakes, and marts in days or weeks rather than in months or years.

WhereScape sits on top of and orchestrates your existing data infrastructure using automated ETL. This data automation enables developers to produce data architectures in a fraction of the time it would take with hand coding, without human error. It enables individual staff members to be more productive and innovative by eliminating repetitive manual tasks.

What does it do?

WhereScape software automates the data warehouse life cycle from end to end. The traditional model of data warehouse development involves many different tools and teams to progress from requirements through design, architecture, and ETL to deployment and documentation. With WhereScape, this can all be done in one highly automated methodology with one suite of tools and the same team of developers maintaining control. This eradicates human error, makes change management much simpler, and crucially shortens the life cycle from many months to just a few days, saving time and money in equal measure.

The products

WhereScape's tools are metadata-driven. They do not touch your data, but sit on top of it and orchestrate what you already have by generating and altering metadata to generate and change data infrastructure. Once your metadata is connected to WhereScape, your infrastructure is effectively future-proofed because you can switch to and from any target database, cloud or on premises, much easier and faster. This means companies do not have to stick to the decisions they make now for the next 10–20 years, providing a level of agility that is essential today given the pace of advances in database technology.

- **WhereScape 3D** (the 3D stands for Data-Driven Design) scopes and profiles any data source and allows architects to quickly spin up prototypes for approval by the business before they build, so architects can check all requirements have been met in one hour rather than waiting weeks or months as with waterfall methodologies. 3D enables agile principles by encouraging collaboration and short iterations.
- **WhereScape RED** is an integrated development environment in which developers can build data structures with a drag-and-drop interface. At a click of a button, the tool physicalizes the model by generating thousands of lines of native SQL code in seconds, saving time and freeing up developers to focus on higher value work.
- **Data Vault Express** uses a combination of RED, 3D, and best practice templates to shorten Data Vault 2.0 implementation times and ensure that the often very complex Data Vault architectures can scale effectively and stand the test of time.

The many stages of the data warehouse development life cycle are automated with WhereScape 3D, WhereScape RED, and Data Vault Express, all fully documented at the click of a button.



research

TDWI Research provides research and advice for data professionals worldwide. TDWI Research focuses exclusively on business intelligence, data warehousing, and analytics issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of business intelligence, data warehousing, and analytics solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.



**Transforming Data
With Intelligence™**

555 S. Renton Village Place, Ste. 700
Renton, WA 98057-3295

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org