research

CHECKLIST REPORT 2019

Optimizing Data Quality for Big Data

By Philip Russom

Sponsored by:

Informatica™

tdwi TRANSFORMING DATA WITH INTELLIGENCE™

JANUARY 2019

TDWI CHECKLIST REPORT

Optimizing Data Quality for Big Data

By Philip Russom

TABLE OF CONTENTS

tdwi

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

FOREWORD

For the most part, data quality is data quality, whether data is big or small, old or new, traditional or modern, on premises or on cloud. This means that data professionals who are under pressure to get business value from new data assets can leverage existing skills, teams, and tools when ensuring quality for big data. Even so, "business as usual" is not enough.

Although data professionals must continue to protect the quality of traditional enterprise data, they must also adjust, optimize, and extend data quality and other data management best practices to fit the business and technical requirements of big data. Otherwise, they may fail to deliver the kind of analytics, operational reporting, self-service, and governance that is expected of all data assets.

The good news is that organizations can apply current data quality and other data management competencies to big data, albeit with adjustments and optimizations. In fact, familiar data quality functions are highly relevant to big data and other valuable new data sources, as seen in the following examples.

STANDARDIZATION. A wide range of users expect to explore and work with big data, often in a self-service fashion that depends on SQL-based tools. Data quality's standardization makes big data more conducive to ad hoc browsing, visualizing, and querying.

DEDUPLICATION. Big data platforms invariably end up with the same data loaded multiple times. This skews analytics outcomes, makes metric calculations inaccurate, and wreaks havoc with operational processes. Data quality's multiple approaches to matching and deduplication remediate data redundancy.

MATCHING. Linkages between data sets can be hard to spot, especially when they emanate from a variety of source systems, both traditional and modern. Data quality's data matching capabilities help to validate diverse data and identify dependencies among data sets.

PROFILING AND MONITORING. Many big data sources—such as e-commerce, Web applications, and the Internet of Things (IoT)—lack consistent standards and evolve their schema unpredictably without notification. Whether profiling big data in development or monitoring it in production, a data quality solution can reveal new anomalies and schema as they emerge. Data quality's modern business rule engines can remediate these automatically at scale.

CUSTOMER DATA. As if maintaining the quality of traditional enterprise data about customers isn't challenging enough, many organizations are now capturing customer data from smartphone apps, website visits, third-party data providers, social media, and a growing list of customer channels and touchpoints.

For these organizations, customer data is the new big data, which multiplies their data quality challenges many times over.

FOREWORD CONTINUED

THE MANY DATA DOMAINS OF BIG DATA. As big data becomes the norm, it progressively includes more information about diverse data domains beyond the customer data domain mentioned above. For example, the product data domain is commonly represented in big data from digital supply chains, data-driven logistics, customer support, and financials.

As another example, industrial use cases for IoT now generate big data about assets captured from sensors, vehicles, and almost every kind of machine, enabling new business practices in real-time business monitoring, facility management, and logistics. Each data domain within big data has unique data requirements and business use cases and therefore needs specific functionality from data quality solutions, as discussed later.

TOOL AUTOMATION. Data professionals and analysts cannot scale their work to the size and complexity of big data. Furthermore, some business users want to profile data, spot issues, and even make changes on their own, at scale and in a self-service manner. Both scenarios call for tool automation based on artificial intelligence, machine learning, and business rules for data quality.

This TDWI Checklist report will drill down into the adjustments and optimizations in data quality practices required for big data. The report will help user organizations understand technology and business requirements for big data and other new data assets, plus data quality's role in attaining maximum business value from such assets.

1 ADJUST DATA QUALITY TIMING AND OTHER DESIGN PARADIGMS TO FIT BIG DATA

Although you can apply data quality best practices and tool functions to big data, you will need to adjust and optimize these to accommodate big data's unique characteristics and use cases as well as evolving best practices in data management. This is similar to how you've treated customer data differently from data about products or how you've stored unstructured data differently from structured.

The challenge today is all about satisfying modern data storage and use case requirements; without that, your data quality solution will fail.

INGEST BIG DATA SOONER, IMPROVE IT LATER. One of the strongest trends in data management is to ensure that incoming data is stored far sooner than in the past. This is so that big data is accessible as early as possible for time-sensitive processes such as operational reporting and real-time analytics.

In these scenarios, persisting data takes priority over making material improvements. To accelerate the persistence of data to storage, up-front improvements to data are minimal or omitted, under the assumption that users and processes can make those improvements later when big data is accessed or repurposed.

A related trend is to store data in its arrival state (discussed in the next section of this report) so that source data's rich details are captured and can be transformed and improved much later as new purposes and use cases arise.

The point of this practice is to respect big data's original state, but it has the added bonus of not impeding the fast ingestion of big data.

BIG DATA QUALITY ON THE FLY. The ramification of these trends is that data aggregation and quality improvements are done on the fly—at read time or analysis time—more often than they have been in the past. This pushes data quality execution closer to real-time. Furthermore, on-the-fly big data quality functions are sometimes embedded in other solutions, especially those for data integration, reporting, and analytics.

To enable embedding and achieve real-time performance, modern tools offer most data quality functions as services. Luckily, today's fast CPUs, in-memory processing, data pipelining, and MPP data architectures provide the high performance required to execute data quality on the fly, at big data scale.

2 ADJUST DATA QUALITY TRANSFORMATION TO IMPROVE BIG DATA WHILE STAYING TRUE TO ITS ARRIVAL STATE

Traditional data warehousing that focuses on reporting usually aggregates, transforms, and cleanses data to an extreme degree because that is required for accurate, auditable, standardized, and trusted reports for financials, sales, and operations. By comparison, analytics that processes big data barely alters incoming data (if at all) because most analytics methods depend on detailed information and unique data points that can be lost via traditional data management practices.

This reminds us that reporting and analytics can be in conflict because of their competing business goals and data requirements. Likewise, data quality practices for report-driven warehousing and analytics-driven big data can differ substantially.

PRESERVE BIG DATA ARRIVAL STATE FOR FUTURE REPURPOSING. An established best practice with big data is to preserve all the detailed content, structures, conditions, and even anomalies that big data has when it arrives from a source. Storing and protecting big data's arrival state provides a massive data store for use cases that demand detailed source, such as data exploration and discovery, as well as discovery-oriented analytics based on mining, clustering, machine learning, artificial intelligence, and predictive algorithms or models.

Furthermore, the store of detailed source data can be repurposed repeatedly for future analytics applications whose data requirements are impossible to know in advance. Data that is aggregated, standardized, and fully cleansed cannot be repurposed as flexibly or broadly.

DATA QUALITY IN PARALLEL. The best practice today with Hadoop and other big data environments is to maintain a massive store of detailed raw data as a kind of source archive. Instead of transforming the source, users make copies of data subsets needing quality improvements and apply data quality functions to the subsets. Similarly, data scientists and analysts create so-called data labs and sandboxes where they improve data for analytics. This "data quality in parallel" is necessary to retain the original value of big data while also creating a different kind of value through mature data quality functions.

CONTEXT-APPROPRIATE DATA QUALITY. Analytics users today tend to alter big data subsets as little as they can get away with because most approaches to modern analytics tend to work well with original detailed source data, and analytics often depends on anomalies for discoveries. For example, nonstandard data can be a sign of fraud and outliers may be harbingers of a new customer segment. As another example, detailed source data may be required for accurate quantification of customer profiles and performance metrics.

When implementing solutions for raw big data, technical users may build operational data stores (ODSs), marts, and master databases using Hive or Hbase in Hadoop. Though quite large, these data sets are structurally simple and so are a good fit for the straightforward row stores of Hadoop.

Such row stores definitely benefit from data quality functions, especially standardization that transforms some columns (to optimize recurring queries) but not others (to preserve unique profile traits). This greatly facilitates the query-based data exploration many users want to do, as well as query-based reporting and analytics. This kind of "context-appropriate data quality" makes users' work easier, enables apples-to-apples comparisons, and contributes to speed and scale—while leaving untouched source details that are key to other use cases.

3 AUTOMATE DATA QUALITY PROCESSES TO SURVIVE GROWING SOURCES AND DATA VOLUMES

In big data environments, data management professionals are forced to cope with:

- Exploding data volumes
- Increasing numbers of sources and targets
- Increasing numbers of users and applications
- More complex data transformations and analytics processing
- Intensifying compliance regulations
- Concentrating remediation far downstream, increasingly on read
- All the above, at scale, while doing more with less

Tool automation for data quality and other data management tasks has become imperative for many use cases involving big data. This is because big data is so big—in size, complexity, origins, and uses—that manual data remediation solely by humans is no longer an option. In fact, there are many use cases where data quality efforts, especially those for big data, benefit from tool automation, whether that automation is based on old or new technologies—or both.

TOOL AUTOMATION CAN ASSIST USERS IN SEVERAL CONTEXTS. No matter how smart tool automation gets, human exception handling and solutions designed by developers are still needed and won't go away. Luckily, these manual tasks can be assisted by predictive algorithms and other forms of artificial intelligence now embedded in some tools. Smart algorithms can suggest mappings, standard schema, metadata descriptions, merge strategies, catalog or glossary tags, and other quality transformations.

IN SOME SITUATIONS, TOOL AUTOMATION MAY REPLACE USERS. When updated via machine learning, data quality algorithms and their predictive models get broader and more accurate over time as they see more data patterns and more variances among data types and structures.

These enhancements continuously improve user productivity as the algorithms' recommendations become richer and better targeted. The enhancements also bring the tool closer to autonomy, where it can be trusted to select and execute remediation processing without human intervention.

TDWI expects tool intelligence and automation to steadily increase—enabled by artificial intelligence, machine learning, smart algorithms, and older approaches to automation—such that the development and production behavior of data quality and other data management solutions will increasingly rely on advanced tool automation.

SMART ALGORITHMS ASIDE, MATURE APPROACHES TO TOOL AUTOMATION REMAIN RELEVANT. Business rules for data quality continue to be indispensable for automated, productive, reusable, and scalable data quality development and production, even with big data and other new data assets. Technical users should continue to look for tools that include large libraries of canned business rules.

Even more imperative is that data quality tools facilitate the creation and sharing of new user-defined business rules and related artifacts (e.g., profiles, quality metrics, domain definitions, and solution components).

3 AUTOMATE DATA QUALITY PROCESSES TO SURVIVE GROWING SOURCES AND DATA VOLUMES CONTINUED

Shared library approaches to data management automation are powerful and useful for a broad range of both technical and business users, and many users have already built out large libraries that they fully intend to maintain and leverage. The shared library approach can also simplify maintenance (fewer artifacts to maintain) and governance (the same rule and related definitions applied everywhere).

NEW BIG DATA SOURCES COME ONLINE REGULARLY. One of the things that makes big data so big is that new sources of it come online regularly. IoT is the most dramatic example because some firms deploy a new type of sensor or device almost daily. Onboarding new sources needs automation to quickly establish connectivity, profile the new data to assess its quality needs, and prototype a solution.

DATA MONITORING IS KEY TO LONG-TERM SUCCESS WITH BIG DATA. Big data is highly variant in that many big data sources lack standards for data types and structures, and they change unpredictably and unannounced. In-production monitoring of a source's quality metrics and schema is a form of data quality tool automation that has been with us for years. Given the volatility of big data, this mature approach to automation is more relevant than ever to detecting change in big data and managing the change.

4 LOOK FOR DATA QUALITY TOOLS THAT SERVE THE NEEDS OF MULTIPLE USER ROLES

The noblest aspect of data quality as a practice discipline is that it brings a wide range of diverse people together. Both business and technology people are needed to identify and design data improvements that align with business goals so that data quality programs give an organization meaningful impact and bang per buck. This has been true for decades, and the collaboration is now key to surviving the onslaught of big data from analytics, IoT, multichannel marketing, smartphones, and new apps.

Accordingly, data quality development environments and end-user solutions are employed by users with diverse roles, talents, and use cases. These include technical people such as data quality developers, data warehouse professionals, and other data management specialists. Some users specialize in analytics such as data analysts, data scientists, and report designers.

Furthermore, data quality user types also include a growing number of businesspeople who serve as data stewards, curators, and governors. Other business users include marketers, brand managers, product designers, procurement specialists, and line-of-business managers.

Ideally, a single data quality tool should serve the needs of all these users. This is challenging given the great diversity of user roles. However, leading tools are meeting the challenge in innovative ways.

SEPARATE ROLE-BASED AREAS WITHIN THE TOOL. Most functionality supports technical data specialists. However, a few areas may target stewards and other business users who need straightforward, workflow-driven functionality so they can explore data, identify problems, mark data elements for developer attention, and perform manual remediation and exception processing.

With so many new sources of big data coming online, technical users need to profile new data to determine its technical condition. Stewards need to explore new data to determine how business entities are represented in it and whether governance policies apply.

Although profiling and exploration are similar, their user types and use cases differ such that separate tool areas may be appropriate. Other functions may demand separate tool areas for technical and business users, such as metadata management, business rule creation, and monitoring.

EASY TO USE GRAPHICAL USER INTERFACES. High ease of use makes data quality tools more accessible for less technical users such as the stewards just mentioned. It also boosts productivity for all users, including the technical ones.

ROLE-SPECIFIC SEMANTICS. Easy graphical user interfaces aside, users also get a boost from business-friendly semantics in the form of business metadata, business glossaries, and data cataloging. Role-specific semantics hide the complexities of big data and other new assets as well as Hadoop, clouds, and other big data platforms. This way, data quality users of all types can be up and running quickly without having to learn the ins and outs of big data and its platforms.

5 EXPECT BIG DATA TO GREATLY EXPAND SEVERAL DATA DOMAINS AND RELATED BUSINESS OPPORTUNITIES

THE CUSTOMER DATA DOMAIN AS REPRESENTED IN BIG DATA

MANY FIRMS RUN ON CUSTOMER DATA. This is a common business model in financial services, insurance, retail, hospitality, and other customer-centric industries. Data is instrumental in acquiring, retaining, and growing customers. It also enables modern digital practices such as multichannel marketing, online marketing campaigns, e-commerce recommendations, insightful analytics, and superlative customer service.

DESPITE THE OPPORTUNITIES, CUSTOMER DATA FACES QUALITY CHALLENGES. Customers move, change jobs, buy new phones, change their names after marrying, and so on, which degrades data accuracy. Furthermore, customer data is notoriously incomplete, making it difficult to convert a prospect into a customer, cross-sell for account growth, deliver a package efficiently, or construct a complete view of each customer.

BIG DATA IS GENERATING EVEN MORE INFORMATION ABOUT CUSTOMERS. That's because customer entities and domains are regularly represented in big data, as seen in social media, website visitor behavior, third-party demographics, acquired leads, consumer-oriented IoT, and e-commerce transactions, shopping carts, and recommendations.

Big data about customers is also a burgeoning component of data-intensive enterprise practices, especially multichannel marketing and digital campaign management. Furthermore, as firms go deeper into advanced analytics, the customer is the business entity most often studied for organizational advantage.

DATA PRIVACY MUST NOT BE AN AFTERTHOUGHT. Trusted customer relationships are based on the transparency you provide around what data you collect, how you plan to use it, who you will share it with, and how you will protect it. Companies often underestimate the privacy risk that big data initiatives bring.

Today's growing list of global data privacy regulations and consumers' backlash against companies that don't protect their data is forcing every organization to be aware of whose data it has, where it's located, how it's used, how it's secured, and who's responsible for governing it. Without this knowledge, an enterprise cannot use its data compliantly and ethically in its big data initiative.

DATA QUALITY TOOLS AND SERVICES ARE OPTIMIZED FOR CUSTOMER DATA. They can help you improve customer data and reap the business benefits while operating on both big data and traditional enterprise data. In addition, tools have modernized in recent years. They are still strong when cleansing customer addresses, consolidating redundant household records, and standardizing customer contact data.

However, modern tools can also verify the existence and accuracy of a customer's identity, address, phone number, email address, and demographics. This removes bogus, unwanted, and fraudulent records while making customer interactions richer and marketing campaigns better targeted.

Furthermore, such investments in customer data quality almost always yield a return for the business because customers, prospects, and accounts are intrinsically linked to revenue and costs.

5 EXPECT BIG DATA TO GREATLY EXPAND SEVERAL DATA DOMAINS AND RELATED BUSINESS OPPORTUNITIES CONTINUED

THE LEADING BENEFIT OF BIG CUSTOMER DATA IS THE COMPLETE CUSTOMER VIEW. Synonyms for this include customer 360, single customer view, and customer master. No matter what it's called, the goal is to quantify, record, analyze, and operationalize as many characteristics of each customer as possible. This comprehensive data set, in turn, assists with customer conversion, retention, and growth.

Big data, when its quality is assured properly, can quantify customer characteristics that went unrecorded previously. For example, big data can capture valuable customer behaviors on new platforms, such as website browsing, online shopping, social media, and mobile device apps.

Note that a customer view is not fully complete without data from other data domains. This typically includes data about a customer's financials, product preferences, demographics, and activity locations (which is an opportunity to augment customer data via geocoding). When big data feeds customer views, the views themselves become "big" and are increasingly managed on big data platforms, which in turn demands modern data quality solutions that integrate with big data platforms such as Hadoop and clouds.

OTHER DATA DOMAINS REPRESENTED IN BIG DATA

The customer data domain aside, many other data domains are amply represented in big data. For example, when the Internet of Things focuses on industrial use cases (logistics and facility monitoring, as opposed to consumer activity), the machine data domain explodes with information from sensors, devices, vehicles, and shipping pallets. IoT data sources are infamous for their evolving schema, anomalies, redundancy, and lack of standards, which a scalable data quality solution can remedy.

As another example, many firms in manufacturing and retail are modernizing their supply chains, which makes business-to-business (B2B) transactions and communication more data-driven than ever. Of course, traditional enterprise applications are producing greater volumes of traditional data about financials, products, assets, and many types of parties (e.g., prospects, partners, and employees).

All data domains found in big data need quality assurance. The challenge is that these domains need a data quality platform that supports both traditional and big data platforms with the speed and scale demanded of modern solutions. Furthermore, a fully modern data quality platform will provide comprehensive support for all data domains with the ability to recognize a domain automatically, then catalog data appropriately and apply domain-aware quality rules.

6 MANAGE AND GOVERN ALL DATA HOLISTICALLY

This report began by saying that "data quality is data quality" in most situations, because modern and traditional data have quality requirements that are remarkably similar, albeit with data set-specific tailoring. Likewise, big data requires data governance because its misuse may lead to compliance infractions, security breaches, or failed data standards, just as with enterprise data.

MAKE BIG DATA USAGE COMPLIANT AS YOU DO WITH ALL OTHER DATA. We can also say that "data governance is data governance" when big data is governed by the same governance body that governs other data as well as when each governance policy is written broadly (and with context options) to apply broadly across all data sets. This kind of holistic data governance is an emerging requirement for user organizations that need full business value from all data assets—but with compliance.

After all, "compliance is compliance" in most data usage scenarios, relative to regulations (e.g., HIPAA and the GDPR), certain data domains (consumers and patients), data privacy policies, and enterprise requirements for stewardship and curation.

EXTEND YOUR DATA GOVERNANCE PROGRAM TO ENCOMPASS BIG DATA, TOO. "All data" means traditional enterprise data, big data, and new data assets such as those from IoT and emerging customer channels. Data this diverse is often managed on multiple diverse data platforms and other IT systems, such that data may exist on premises, in the cloud, or both (as is common in today's hybrid multiplatform data architectures). We've already seen that data quality practices and tools can be adapted to big data. Likewise, business best practices and tool functionality for data governance (and related disciplines such as data quality and integration) can be extended from enterprise data to big data, despite the extreme diversity of "all data" and the numerous data platforms required to manage it.

GOVERN BIG DATA'S STANDARDS, NOT JUST ITS COMPLIANT USE. Governing data usage is critical for regulatory compliance, and that is the highest priority for most data governance programs. However, a truly holistic data governance program will also govern an enterprise's standards for all data quality, modeling, and interfacing. Enterprise standards give data the consistency it needs to be mastered, improved, and shared across multiple IT systems and business units, which is increasingly difficult as hybrid data architectures proliferate. That's important because sharing data helps the modern business succeed with single views of customers, complete data for reports and analyses, and up-to-date status information for transactions, shipments, and other multistep operational processes.

MAKE DATA GOVERNANCE HOLISTIC, FOR CONSISTENT POLICIES AND STANDARDS ACROSS ALL DATA. Holistic data governance—which provides both compliance policies and data standards for all data—should be a goal for user organizations, especially as they assimilate big data with enterprise data. This assimilation, when governed carefully, contributes to data-driven but compliant business goals such as rich analytics correlations across all data sources, broad data exploration, enterprisewide visibility via data, and next-level business monitoring, performance management, and operational agility.

7 LOOK FOR TOOLS THAT SUPPORT ALL DATA QUALITY FUNCTIONS FOR DATA BIG AND SMALL IN A UNIFIED PLATFORM

We say "data quality" as if it is a single monolith, but it isn't. Instead, data quality is a collection of a dozen or more related techniques and practices, including standardization, contact data cleansing, profiling, monitoring, verification, merging, and enhancement via third-party data. You'll need all these techniques with big data—just as you needed them all with enterprise data. This is because big data has similar problems with standardization and redundancy, plus similar opportunities for augmentation.

Furthermore, this rich collection of data quality functionality should be complemented by an equally rich collection of data integration functionality, including ETL/ELT, replication, data synchronization, federation, virtualization, and event processing.

Together, the many techniques of data quality and data integration enable a broad range of multifunctional solutions for diverse data, both traditional and modern. Likewise, data quality and data integration functionality must support all big data platforms and traditional data platforms, whether on premises, on multiple clouds, or a hybrid combination of these.

Ideally, you should demand all of that functionality in one toolset, with all the functions interoperating and unified as an integrated tool platform. The toolset should be unified via a single console, and it should make certain artifacts shared across all data quality and all data integration functions, including metadata, profiles, quality metrics, business rules, development artifacts, libraries of canned solutions, and interfaces.

To achieve this level of integration, the entire toolset must come from one vendor.

A unified toolset for data quality and data integration techniques offers advantages for both traditional data and big data:

PRODUCTIVITY BOOST. Users—whether business or technical people—can move in a quick and controlled manner from one set of functionality to another, instead of ping-ponging among vendor brands. Productivity is key when big data and other new data assets increase human workloads.

COLLABORATION AMONG MANY USER TYPES. Technical developers can borrow from each other's prior work instead of reinventing the wheel. Stewards, curators, and other businesspeople can see the work of developers as well as suggest directions for data quality work that would benefit the business. This collaboration is more important than ever as organizations incorporate new big data sources and integrate big data with enterprise data.

SIMPLER BUT BROADER GOVERNANCE. As discussed earlier, a unified toolset must also provide functionality for less technical users, especially data stewards, curators, governors, and data owners.

Furthermore, a unified toolset makes most data management solutions and the data sets they manage visible from a single console. This simplifies, empowers, and unifies governance tasks by providing visibility and stewardship assistance across all data, both traditional and modern.

7 LOOK FOR TOOLS THAT SUPPORT ALL DATA QUALITY FUNCTIONS FOR DATA BIG AND SMALL IN A UNIFIED PLATFORM CONTINUED

MODERN DATA FLOW DESIGN. One of the strongest design trends among data professionals is to create a single data flow, workflow, or pipeline that calls many diverse data management functions. After all, many data sets need quality improvements, enhancements, and verification as they are ingested and integrated; this includes big data.

A multifunctional data flow is relatively easy to deploy and manage compared to the older practice of developing a plague of small routines that need elaborate scheduling and coordination. This kind of unified data quality/integration solution is best implemented with a unified toolset.

ENABLING NEW DATA ARCHITECTURES. Big data and the advanced analytics that usually comes with it are forcing organizations to adopt new data architectures and data platforms. Modern data quality tools and techniques must evolve to support these. The data lake is a case in point.

Research by TDWI shows that data lakes may be deployed on Hadoop, clouds, relational databases, or a combination of these. Big data's quality processing is complicated because much of it should be done "in situ," i.e., inside Hadoop, a cloud, or a database. This is because big data and lake data are too voluminous to move from platform to platform.

In such cases, a data quality toolset must interface with multiple big data and traditional platforms to execute and manage in situ processing, even when big data is hybrid in the sense of distributed across on-premises and cloud systems.

ABOUT OUR SPONSOR

Informatica™

informatica.com

Digital transformation changes expectations: better service, faster delivery, with lower cost. Businesses must transform to stay relevant and data holds the answers.

As a world leader in enterprise cloud data management, we're prepared to help you intelligently lead—in any sector, category, or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities, or create new inventions. With 100 percent focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

Informatica accelerates data-driven digital transformation. Informatica enables companies to fuel innovation, become more agile, and realize new growth opportunities, resulting in intelligent market disruptions. Over the past 25 years, Informatica has helped more than 9,000 customers unleash the power of data. For more information, call +1 650-385-5000 (1-800-653-3871 in the U.S.), or visit www.informatica.com. Connect with Informatica on LinkedIn, Twitter, and Facebook.

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, having published over 550 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 was a product manager at database vendors. His Ph.D. is from Yale. You can reach him at prussom@tdwi.org, @prussom on Twitter, and on LinkedIn at linkedin.com/in/philiprussom.

ABOUT TDWI RESEARCH

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on analytics and data management issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data management solutions. TDWI Research offers reports, commentary, and inquiry services via a worldwide membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.

ABOUT TDWI CHECKLIST REPORTS

TDWI Checklist Reports provide an overview of success factors for a specific project in business intelligence, data warehousing, analytics, or a related data management discipline.