



research

BEST PRACTICES REPORT

Q3

2021

Unified Platforms for Modern Analytics

By Fern Halper, Ph.D.

tdwi

Transforming Data
With Intelligence™

Research Sponsors

Alation

Denodo

Incorta

Matillion

SAP

SAS

Snowflake

Wavicle Data



Unified Platforms for Modern Analytics

By Fern Halper, Ph.D.

Table of Contents

Executive Summary	4
Introduction to Unified Platforms for Modern Analytics	5
The Current State of Data Management and Analytics in Organizations	6
Benefits of a Unified Platform	7
In Their Own Words	9
Characteristics of the Unified Platform for Modern Analytics	10
Minimum Characteristics of a Unified Platform	10
Important Characteristics of the Platform for Modern Analytics	11
Modern Analytics and Unification	14
Accomplishing Unification	16
Physical and Logical Consolidation	16
Organizational Considerations for Unification	18
Barriers to Unification	19
Governing the Unified Platform for Modern Analytics	21
Ethical Use of Data	23
Recommendations	24

© 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 Authors



FERN HALPER, Ph.D., is VP and senior director of TDWI Research for advanced analytics, focusing on predictive analytics, text analytics, cloud computing, and other “big data” analytics approaches. She has more than 20 years of experience in data and business analysis, and she has published numerous articles on data mining and information technology. Halper is coauthor of “Dummies” books on cloud computing, hybrid cloud, service-oriented architecture, and service management, and *Big Data for Dummies*. She has been a partner at industry analyst firm Hurwitz & Associates and a lead analyst for Bell Labs. Her Ph.D. is from Texas A&M University. You can reach her at fhalper@tdwi.org, [@fhalper](https://twitter.com/fhalper) on Twitter, and on LinkedIn at linkedin.com/in/fbhalper.

About TDWI

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 the TDWI Best Practices Reports Series

This series is designed to educate technical and business professionals about new business intelligence technologies, concepts, or approaches that address a significant problem or issue. Research for the reports is conducted via interviews with industry experts and leading-edge user companies, and is supplemented by surveys of business intelligence professionals.

To support the program, TDWI seeks vendors that collectively wish to evangelize a new approach to solving business intelligence problems or an emerging technology discipline. By banding together, sponsors can validate a new market niche and educate organizations about alternative solutions to critical business intelligence issues. To suggest a topic that meets these requirements, please contact TDWI Senior Research Directors David Stodder (dstodder@tdwi.org), James Kobieltus (jkobieltus@tdwi.org), and Fern Halper (fhalper@tdwi.org).

Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, especially those who agreed to our requests for phone interviews. Second, our report sponsors, who diligently reviewed outlines, survey questions, and report drafts. Finally, we would like to recognize TDWI’s production team: James Powell, Richard Seeley, Lindsay Stares, and Rod Gosser.

Sponsors

Alation, Denodo, Incorta, Matillion, SAP, SAS, Snowflake, and Wavicle Data sponsored the research and writing of this report.

Research Methodology and Demographics

Report purpose. As organizations evaluate more advanced analytics (such as machine learning and natural language processing) using complex data types and techniques, standalone products and the traditional warehouse are not meeting their needs. The unified platform provides an integrated set of tools that address many aspects of the data and analytics life cycle. These platforms may offer data warehouses, data lakes, governance, machine learning, model deployment, and other analytics and data services.

The purpose of this report is to help technical and business users understand new directions in data and analytics. This TDWI Best Practices Report will address adoption, use, challenges, architectures, satisfaction, and best practices for utilizing unified platforms.

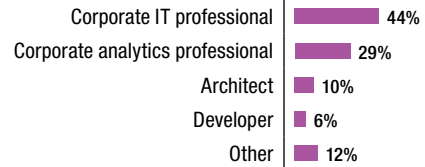
Survey methodology. In May 2021, TDWI sent an invitation via email to the analytics and data professionals in our database, asking them to complete an online survey. The survey collected responses from 474 respondents. Three hundred and ninety-eight of them completed the entire survey and met quality standards. This group is used for analysis.

Research methods. In addition to the survey, TDWI conducted telephone interviews with technical users, business sponsors, and analytics experts. TDWI also received briefings from vendors that offer products and services related to these technologies.

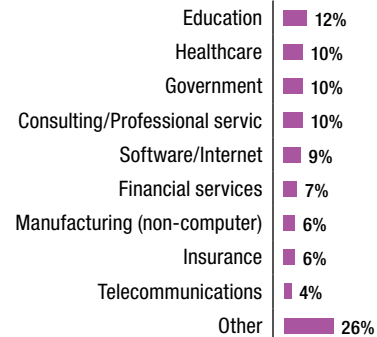
Survey demographics. Respondents act in a variety of roles. These include corporate IT professionals (44%), corporate analytics professionals (29%), architects (10%), developers (6%), and others (12%).

The education (12%), healthcare (10%), government (10%), consulting (10%), and software/internet (9%) industries dominate the respondent population, followed by financial services (7%), and manufacturing and insurance (both at 6%). Most survey respondents reside in the U.S. (82%), Asia (6%), or Europe or Canada (both 4%). Respondents come from enterprises of all sizes.

Position

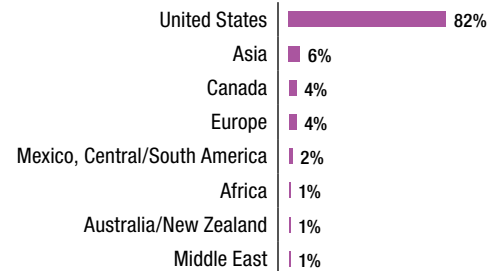


Industry



(“Other” consists of multiple industries, each represented by 3% or less of respondents.)

Geography



Company Size by Revenue



Based on 398 respondents.

Executive Summary

The market is still early for the unified platform for modern analytics although respondents to the survey for this report overwhelmingly agree that unification is an opportunity.

The demand for analytics continues to grow. At TDWI we see more organizations looking to advance beyond dashboards to self-service analytics and utilize more sophisticated algorithms such as machine learning (ML). To support this trend, companies are modernizing their data and analytics environment as well. The trend is towards a *unified environment* that provides an integrated set of tools that address many aspects of the data and analytics life cycle, including data management and development and deployment of sophisticated analytics. The platform is often cloud-based to take advantage of scale and flexibility and support massive amounts of data and compute-intensive workloads. It is automated and augmented to help address data and analytics complexity.

The market is still early for this unified platform although respondents to the survey for this report overwhelmingly agree that unification is an opportunity. The benefits of unification include the ability to capture and leverage emerging and diverse data, achieving faster time to insights, better performance and scalability, and support for advanced analytics. The challenges include vendor lock-in and the politics of ownership.

In this study, respondents identified a minimum set of important features for the unified platform. This reflects today's market. Fifty-seven percent said data warehouse services needed to be included in the unified platform. Forty-three percent said security services and 33% each said data lake and data visualization services. However, respondents all agreed that other services such as data catalogs, cloud services, augmented and automated services, advanced analytics, the ability to support open source and commercial tools, and deployment services were also important.

Organizations are accomplishing or planning to accomplish unification in different ways. On the technology front, a popular approach is to leverage an integrated stack that includes data and analytics capabilities. Another approach is to utilize a data fabric to logically integrate data across hybrid environments and integrate analytics into this fabric. New approaches continue to evolve. On the organizational front, new skills will be needed. For instance, companies will need to hire operations staff to help put analytics into production and monitor its performance—this will be key to success. This includes roles such as MLOps and data engineers.

In addition to new roles for the unified platform, there will be new governance considerations, too. Organizations will need to adopt tools to register and manage models, especially those that will be put into production. They will also need to provide explainability so users can understand output from automated tools. Data and analytics ethics will start to become more important as organizations build and deploy ML models at scale on the unified platform.

This TDWI Best Practices Report examines the adoption, use, challenges, architectures, and best practices for unified platforms for modern analytics. It examines how organizations are using these platforms today and discusses their future plans.

Introduction to Unified Platforms for Modern Analytics

Analytics continues to evolve. As organizations collect large amounts of diverse data, they often want to modernize their data and analytics environments to support new use cases, such as building machine learning models for image recognition, analyzing Internet of Things (IoT) data, and deploying conversational bots. TDWI research indicates that newer data types such as machine, text, and image data as well as other unstructured and semistructured data sources are gaining popularity for use in analytics. For instance, text data is already mainstream.

Organizations want to derive insights and take action on this data using analytics. They want to utilize self-service analytics. They want to build ML models and put them into production as part of a system or application. They want to support data scientists, business analysts, and business users in their use of analytics. Yet in many cases, the evolution of complex data and analytics (and the scale involved) has outstripped a company's ability to manage this data and analytics for business value.

Recently, organizations have started to unify the platform to support modern analytics. The idea behind the unified platform is to provide an integrated set of capabilities to manage data as well as analytics development and deployment. The modern unified platform is governed, so the data is trusted and models are transparent. It is scalable to support massive amounts of data and is often hybrid or cloud-based. It supports a range of analytics and tooling, including open source and commercial products, and supports multiple personas. Given this, it is no surprise that 83% of respondents to the survey for this report believe that a unified platform is an opportunity (not shown).

In its simplest form, the unified platform is capable of ingesting diverse data from multiple sources, has pipeline services, a data storage and/or virtual/semantic layer (the data environment), an analytics layer, and a consumption layer (see Figure 1). The platform is well governed and secure and can be deployed in the cloud, on premises, or in a hybrid fashion. In practice, it may include many other services.

The idea behind the unified platform is to provide an integrated set of tools to manage data as well as analytics development and deployment.

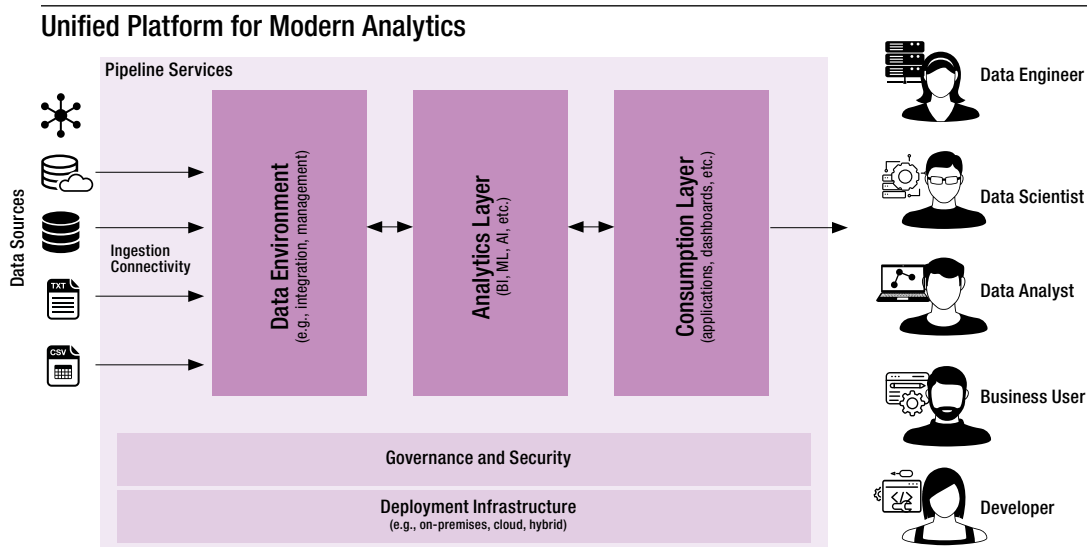


Figure 1. The unified platform for modern analytics

TDWI has recently written about unifying the data warehouse and data lake environment.¹ A unified data and services layer may form the data foundation for the unified platform for modern analytics. This data layer may consist of the data warehouse and/or the data lake. In many cases the data warehouse and data lake are converging to form what some call a lakehouse or a data cloud. In other cases, the data warehouse and data lake are unified via a data fabric. A data fabric maps and connects relevant application data stores with metadata to describe data assets and their relationships. In any scenario, the data is trusted for analytics use.

The Current State of Data Management and Analytics in Organizations

At TDWI, we've seen the traditional enterprise data warehouse environment evolve to include cloud-based platforms. In fact, cloud data warehouses and cloud data lakes are already mainstream. In this survey, for instance, 48% of respondents were already using a data warehouse in the cloud (not shown). Thirty-one percent were already deploying a data lake in the cloud, compared to 18% of organizations that have implemented data lakes on premises (not shown). Drivers for modernization include the need to support new use cases using sophisticated analytics such as machine learning and artificial intelligence (AI). To get a sense for where respondents are now in their data management implementations and if they are meeting their needs, we asked respondents, "Which statement best describes your current data and analytics environment?"

Many organizations are still on premises. As illustrated in Figure 2, 28% of respondents state that their current environment is still largely on premises and that this is working for them. The on-premises warehouse is a staple for many organizations and we do not expect that to change any time soon. Yet, we are seeing more organizations moving to other kinds of environments when the need to support more complexity arises. For instance, of those on premises, the majority (78%, not shown) are either thinking about building a unified platform or building one currently.

Data silos and shadow IT still plague other organizations. Another 21% of respondents are dealing with silos and shadow IT. In this case, respondents note that nothing is unified and close to 50% of these organizations are thinking about building a unified platform for analytics (not shown). It is hard to compete when utilizing a siloed data environment because data may not be consistent across silos and therefore not trusted.

The hybrid environment is dominant. Figure 2 also illustrates that the largest percent (36%) of respondents are using some sort of hybrid environment to manage their data. This may include a data warehouse on premises and a data lake in the cloud. The organization may run its standard reports from the on-premises warehouse and run more advanced analytics in the cloud data lake. This architecture may evolve out of necessity rather than choice (e.g., because of dealing with legacy systems), and it may be fine in their current situation. At TDWI, we often see that the hybrid or multiplatform environment is the predominant choice for organizations on their analytics journey.

A small percentage of companies are completely in the cloud. Thirteen percent of respondents stated that their environment was completely cloud-based. These companies were more likely to be high-tech focused (e.g., software/internet) and smaller in size than the overall respondent pool. However, at TDWI we see a growing number of organizations moving to the cloud as well as implementing a cloud-first model where all new data and analytics work goes into the cloud.

At TDWI, we often see that the hybrid or multiplatform environment is a popular step for organizations on their analytics journey.

Regardless of whether the organization is on premises, in the cloud, or in a hybrid situation, the majority believe that it is extremely important to provide an integrated set of tools to manage data and analytics development and deployment (not shown).

Which statement BEST describes your current data and analytics environment?

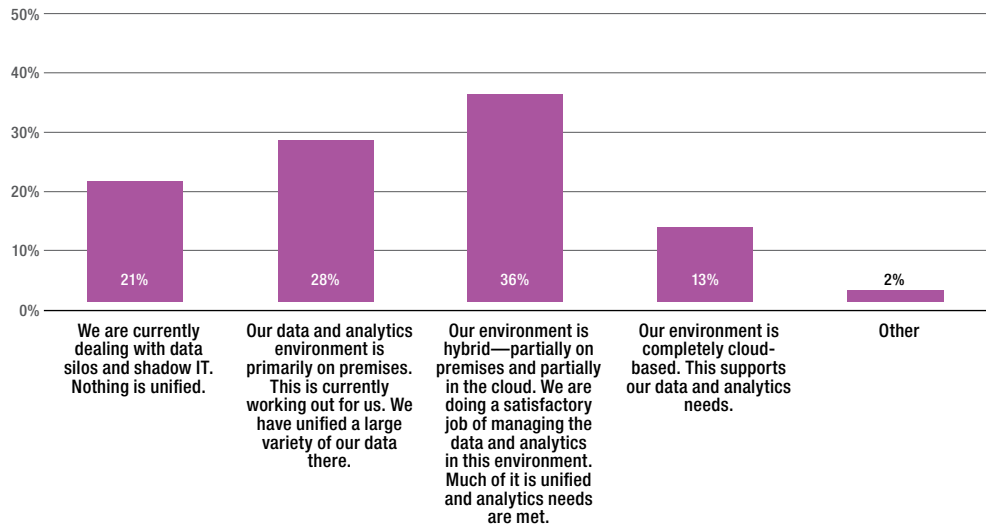


Figure 2. Based on 398 respondents.

Benefits of a Unified Platform

As noted, the central idea behind the unified platform for modern analytics is to provide an integrated set of capabilities that address many aspects of the data and analytics life cycle, including data management and analytics. In this survey, we asked respondents what benefits a unified platform for modern analytics would provide (see Figure 3).

Capture and leverage emerging and diverse data types. At the top of the list was the need to capture and leverage new and emerging data types and sources, including data from the Internet of Things (IoT) as well as text, image, semistructured, and streaming data for use across the organization. Thirty-five percent of respondents cited this as a top benefit of the platform. Twenty-eight percent cited utilizing more diverse data for analytics as a benefit of the unified platform. For instance, organizations are utilizing IoT data for predictive maintenance or in new data products and services. They are marrying text data (such as sentiment data) derived from natural language processing together with structured data to gain better insights into customers. A unified platform that includes the capabilities of the data warehouse (DW) and the data lake (DL) for managing this diverse data can be critical. Advanced analytics such as machine learning often makes use of large amounts of diverse data for model training. The data lake or converged DW/DL is becoming a popular solution for storage of newer data types.

The top benefit of the unified platform for modern analytics is to capture and leverage emerging data types.

A unified platform for analytics can help achieve faster time to insight because a trusted data source is easily accessible.

Achieve faster time to insights and support decision intelligence. Analytics tooling is a key component of the unified platform. This includes self-service tools as well as more advanced tools such as machine learning that are integrated as part of the platform. A unified analytics platform can help to achieve faster time to insight because a trusted data source is easily accessible as are tools for analytics depending on different skill levels. In this survey, 29% cited faster time to insights using available platform tools as a top benefit. Twenty-two percent stated a top benefit is support for decision intelligence.

Only 17% of respondents cited advanced analytics as a benefit of the unified platform, but this may reflect the current state of maturity of the respondents. In another survey question, over 70% of respondents are planning to build and deploy advanced analytics models on their unified platform. About a quarter weren't sure (not shown). In other words, this is the direction the unified platform is headed. If organizations are collecting more diverse data, they will ultimately want to analyze it for competitive advantage.

Provide better performance and scalability. Twenty-seven percent of respondents cited better performance and scalability as a benefit of the unified platform for analytics. As previously mentioned, close to half of the respondents to the survey are already using a cloud platform in their current environment. The cloud can provide a scalable and extensible platform for data and analytics. These are often hyperscale environments where a large number (e.g., thousands) of servers are networked together and scale horizontally (scale out) to achieve massive scale when required. In some deployments, services such as ETL or analytics natively use the computational power of the cloud. The cloud may house the unified platform entirely or be part of a hybrid environment unified by other means to leave data in place, such as data virtualization.

What is the benefit of a unified platform for analytics?

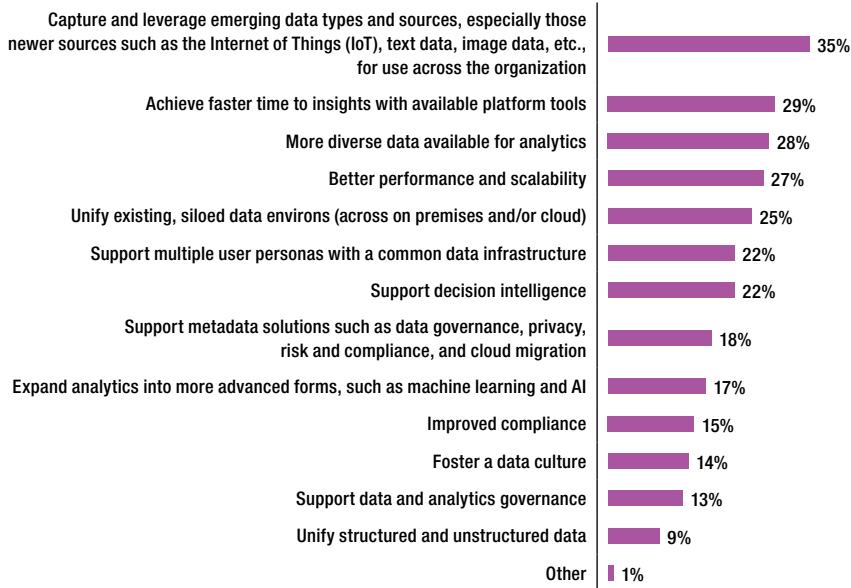


Figure 3. Based on 398 respondents. A maximum of three responses allowed.

In Their Own Words

In a separate question, we asked respondents to state, in their own words, the benefits (as well as some risks) of the unified platform for modern analytics. Responses include:

“It provides you with a holistic view of the organization and the ability to integrate external sources to aggregate value.” Architect, consulting/professional services

“Unified platforms empower competent decision making across the organization due to everyone analyzing the same data points/trends.” Architect, government (federal, state, local)

“It is necessary to assure the governance around the data models, metadata, etc. This will allow a unified and consistent definition of data and metrics used by the enterprise.” Director, data governance, wholesale/retail/distribution

“It provides more options for managing an increasingly diverse range of use cases, data types, end users, and analytics needs.” Business analyst, healthcare

“It is beneficial because it helps in getting a diverse range of data sets.” Developer, software/internet

“Everyone needs to pull from one source of truth.” CIO, real estate

“It is beneficial because all the tools and software will be readily available and we don't have to wait for any licenses.” Business analyst, education

“It allows you to get rid of data silos that give way to ineffective analytics and it allows for consistency in data mining and/or analytical results.” Data analyst, education

"[The unified platform] provides more options for managing an increasingly diverse range of use cases, data types, end users, and analytics needs."

Although respondents cited many benefits for the unified platform, it is important to note that many said they weren't using a unified platform yet. Some also noted they wouldn't unify all their data into one platform. Others felt that the unified platform might be a risk because it would lead to people not utilizing certain data types or “forced conformity” by eliminating silos. Still others felt it would be too costly or hard to build. Some did not yet have the use cases they believed would warrant the need for a unified platform. Yet, for those using a unified platform, the primary use cases were dashboards and visualization (e.g., leaning towards self-service) as well as advanced analytics such as machine learning.

USER STORY: MAKING THE MOVE TO THE CLOUD

According to Serge Billiouw, director of data, analytics, and digital transformation at Atlantic Broadband, “We had a classic data warehouse. However, it wasn't completely meeting our business needs.”

Atlantic Broadband's parent company in Canada, Cogeco, had moved to a cloud platform that provided a cloud data warehouse, analytics capabilities, and other services. Billiouw decided his group should also move in that direction for a number of reasons. The platform was more cost-effective, provided new features weekly, and supported Python. His team could also leverage the work their partner had done, be consistent across the company, and better collaborate across groups. Additionally, the cloud provider offered tools for analyzing voice files for call centers—an important application for telecommunications companies—that were easier to integrate into the platform.

The company is continuing to build out its data and analytics capabilities. “We've built out the infrastructure,” said Billiouw. “Now each department is getting a business analyst.” Analysts will be trained on the software. As part of building up its data governance strategy, the company is working on a catalog to define what each term means. “We are developing a single source of truth and making sure everyone uses the same data and understands what it means.” In this way, the company hopes to help more people analyze data and provide consistency in analysis.

Characteristics of the Unified Platform for Modern Analytics

If a unified platform provides an integrated set of services for data and analytics, what services are most important? We asked respondents, “What minimum set of services should the unified platform for analytics provide?” A maximum of six responses were allowed.

Minimum Characteristics of a Unified Platform

Top platform services cited by respondents as necessary are primarily focused on data management, with the exception of data visualization.

As illustrated in Figure 4, top platform services are primarily data management-focused, with the exception of data visualization. This is consistent with where the market is today. Although most organizations use self-service data visualization and look for trusted data, the use of more advanced analytics such as machine learning or natural language processing (NLP) is in the early stages of mainstream adoption. That isn’t to say that these advanced analytics services aren’t important—they most certainly are, as we will discuss later in the report. However, many organizations are still early in their analytics journey. Therefore, it is not surprising that respondents chose a current minimum set of services focused on the following:

Data warehouse and data lake services. Not surprisingly, data warehouse services ranked at the top of the list of a minimum set of services for the unified analytics platform. Fifty-seven percent of respondents stated this is an important service. The data warehouse is a mainstay of analytics, whether on premises or in the cloud. The data warehouse ingests multiple sources of data and makes them available for analysis. Organizations looking to build a unified analytics stack, for instance, look to the data warehouse as a core component. As organizations strive to derive value from their data, they are often modernizing their DW environments to support self-service, advanced analytics, and data sharing.

Likewise, organizations are also moving towards deploying data lakes, more often in the cloud than on premises. Previous TDWI research indicates that primary use cases for data lakes include supporting source data staging, advanced analytics, and extending the data warehouse to process and store newer data types (such as unstructured data).² It is not surprising that 33% of respondents cited this as an important service for the unified platform. Some organizations might use data virtualization (a semantic layer that integrates data across platforms) to help unify the data warehouse and the data lake. Others may use a unified cloud environment that offers data warehouse and data lake services.

Security services. Organizations also want their data to be secure in the unified environment. This may include data access and authorization services along with data obfuscation services such as encryption or masking. Forty-three percent of respondents cited this as an important service for the unified platform. Data protection is discussed in more detail later in this report.

Data visualization services. Traditional reporting and dashboards are still a popular way for organizations to gain insight and information. However, the movement in analytics is towards self-service discovery and more advanced analytics such as predictive analytics and machine learning. In a recent TDWI survey, for instance, self-service analytics was cited as the top priority for organizations.³ Self-service analytics enables nontechnical users to be productive with data because it is easier to use (thanks to a graphical user interface (GUI), automated business glossary, and natural language search), does not require coding, and does not require IT developers to set up all data access, queries, visualizations, and preparation routines. Data visualization enables organizations to derive insights themselves through easy-to-use visualization tools. Today’s data visualization tools are often augmented and automated (see

The movement in analytics is towards self-service discovery and more advanced analytics such as predictive analytics and machine learning.

below) as part of the unified platform. In this survey, 33% of respondents cited data visualization services as a must-have for the unified platform for analytics.

ETL and connectivity services. Organizations look to extract, transform, and load (ETL) to help enable integration. ETL is a mainstay of the data warehouse environment. Thirty-three percent of respondents cited ETL as being important for the unified platform. Thirty-two percent cited connectivity services. Both of these make sense because the platform needs to have ingestion and processing capabilities. Organizations are looking to connect to multiple data sources as they mature including databases, enterprise applications, and real-time data. However, many organizations are moving to modern pipelining services that are augmented and automated.

What minimum set of services should the unified platform for analytics provide?

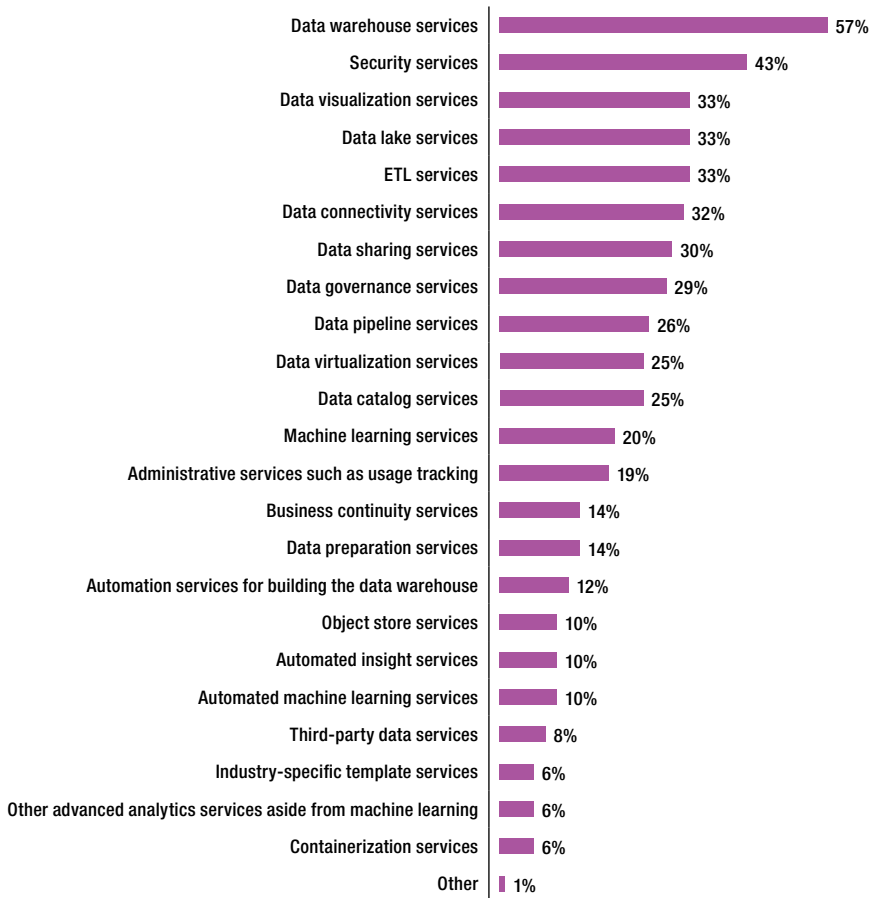


Figure 4. Based on 398 respondents. A maximum of six responses allowed.

Important Characteristics of the Platform for Modern Analytics

Respondents didn't necessarily rank more modern services such as augmented and automated tools, data cataloging, or data pipelines as part of the basic set of functions for a unified platform for data and analytics, but they did agree that these and others (such as MLOps services) were important (see Figure 5). At TDWI, we see vendors already supporting much of this additional functionality. In fact, these are critical components of a platform for modern analytics.

At TDWI, we see vendors supporting additional functionality that is important for the unified platform for modern analytics.

Support for multiple data sources. As mentioned, organizations are collecting more diverse data types including structured, semistructured, and unstructured data, which they plan to use for analytics. The modern platform must support the ingestion and processing of new data types, such as streaming data or image data. It should also support ingestion of traditional data types, such as data from ERP systems. Eighty-five percent of respondents agreed with the statement, “The unified platform for analytics should connect modern and traditional data sources; therefore, it should connect to a broad range of data sources.” This data will be used to enrich structured data for analytics purposes. It is critical that the platform support it.

Support for modern pipelines. Two-thirds (66%) of respondents to this survey agreed with the statement, “Our pipelines are complex. Given this, we will need pipeline tools that include automation and augmentation” (see next item). Manually built complex pipelines can cause bottlenecks and be error-prone. Automation can help. Others felt that it made sense to “push down” ETL processing to the cloud layer and use the computational power of the cloud (65%, not shown) for ETL processing. Still others felt that visual cues as guidance were needed to see where data comes from and how it changes (78%, not shown). These are modern pipeline services. Pipeline vendors are also providing other services to process new kinds of data, such as text data. This data can be processed to extract entities (people, places, things) as well as concepts (words and phrases that indicate a particular idea), themes (groups of co-occurring concepts), or sentiments (positive, negative, neutral) and intent (predictions of future behavior). That information can be used by itself or fed back into a platform and used for analytics.

Augmentation and automation. Augmented intelligence refers to software that has AI technologies such as machine learning or natural language processing infused into the product. Augmentation and automation are important trends occurring across the data and analytics life cycle. For example, machine learning can operate behind the scenes to identify poor-quality data in a pipeline or learn patterns associated with sensitive data to automatically classify it. Data mapping services, where ML is used behind the scenes to map data from different sources together, is another area that utilizes augmentation and automation. Some organizations will use automated tools to surface insights in data visualization tools. Others provide a natural language interface to make it easier to query a data set. Still others will use autoML tools with machine learning infused into the product to assist with such tasks as identifying features, tuning hyperparameters, or even building models.

In our survey, 81% of respondents agreed with the statement, “Automation and augmentation tools and services are critical across the whole data and analytics life cycle on the unified analytics platform.”

Collaboration services. Previous TDWI research indicates that collaboration is important for successful analytics. Data science teams need to collaborate with data engineering teams. Business analysts may want to collaborate with data science teams. Business should collaborate with IT. All teams should collaborate internally. In particular, as remote work becomes more prevalent, it is important for teams to be able to collaborate. Eighty-one percent of respondents agreed with the statement, “As remote work has become a norm, the unified platform for analytics should provide a place to collaborate on data.” This can include collaborating via the data catalog on rating or recommending data sets and sharing information via conversations and wiki-like articles. It can include sharing data across partners. It might include a team workspace, for instance, to get input from others about a machine learning model and then build a data product.

Data catalog and other governance services. In Figure 4, a quarter (25%) of respondents stated that the data catalog was necessary for the unified platform. In other words, although not in the top six, it was still important. This is reflected in Figure 5, where 76% of respondents agreed that “a data catalog is critical for the unified platform [to provide] one source of truth.” Those performing analytics want to understand *and* trust their data because otherwise analytics efforts will fade on the vine. TDWI has seen data catalogs grow in importance in organizations, especially for data governance (discussed later in this report).

Modern data catalogs have numerous features. Some will let you know who has recently used a certain data set. Some are augmented and automated. Some have a built-in intelligent SQL editor. For instance, if an individual started to query sales numbers, the catalog could then surface queries that have already been created for similar questions or provide suggestions for queries. Some modern catalogs can tell you if particular reports have been used recently. Others provide collaboration services.

Modern data catalogs have numerous features that are often automated or augmented.

Support for open source and commercial tooling. Depending on talent and the maturity of its data and analytics, an organization will combine its tools, techniques, and toolkits. That often means combining open source and commercial software. Many data scientists and others like open source because it is a free, collaborative community of innovation that attracts large numbers of people to contribute to and use it. Open source tools for analytics have been available for decades. Popular toolkits include R and Python. Likewise, other organizations prefer commercial tools because they provide support, are often easier to use, and are compliant. The unified platform should support both because organizations will continue to use multiple tools, perhaps standardizing for certain kinds of problems, such as regulatory reporting. Sixty-nine percent of respondents agree that the unified platform must support both open source and commercial analytics tools.

Works on multiple clouds (AWS, Google, Microsoft, etc.). As mentioned, the unified analytics platform can be cloud-based. Many organizations want the scalability and flexibility of the cloud. Vendors are providing tools that leverage the computing power of the cloud. Some are specifically cloud-native. Many work across multiple cloud environments. Seventy-four percent of respondents agreed that “The unified platform for analytics should be scalable and flexible; therefore, it should be cloud-based.”

Machine learning and MLOps services. Organizations are looking for more advanced analytics services, such as machine learning, to be part of the platform. At TDWI, we see demand for machine learning increasing. As noted, automated machine learning (autoML) is also becoming popular. Sixty-two percent of respondents agreed with the statement, “Augmented and automated tooling for advanced analytics such as machine learning is an important part of the unified platform. This feature functionality is/will be an important purchase criteria.”

Additionally, as organizations scale the number of models they put into production, they will need MLOps services. Only 18% are using model deployment services today in the cloud and 16% are using them on premises (not shown). However, these services (including connectivity, registration, model validation, and deployment and monitoring services) are critical for organizations looking to operationalize modern analytics, put the best models into production, and do so in a timely fashion. Otherwise, models run the risk of becoming stale before they are even put to use.

Microservices. Almost half of respondents (48%) agreed that microservices are a key part of the architecture for the unified platform. This is an evolution of a services-oriented architecture (SOA), where applications are assembled using a set of building blocks (components). One cloud

service might be responsible for compilation, authorization, authentication, and metadata. Another might be responsible for distributed query optimization.

One benefit of the services approach is that services can be updated easily. Microservices are also scalable and easier to understand and maintain. A modern cloud architecture often utilizes a microservices approach along with containers and Kubernetes, the hallmark of a cloud-native design. Key to this platform is the ability to integrate with the growing list of cloud-provider commodity services that interoperate with Kubernetes (e.g., identity management, monitoring, logging, etc.).

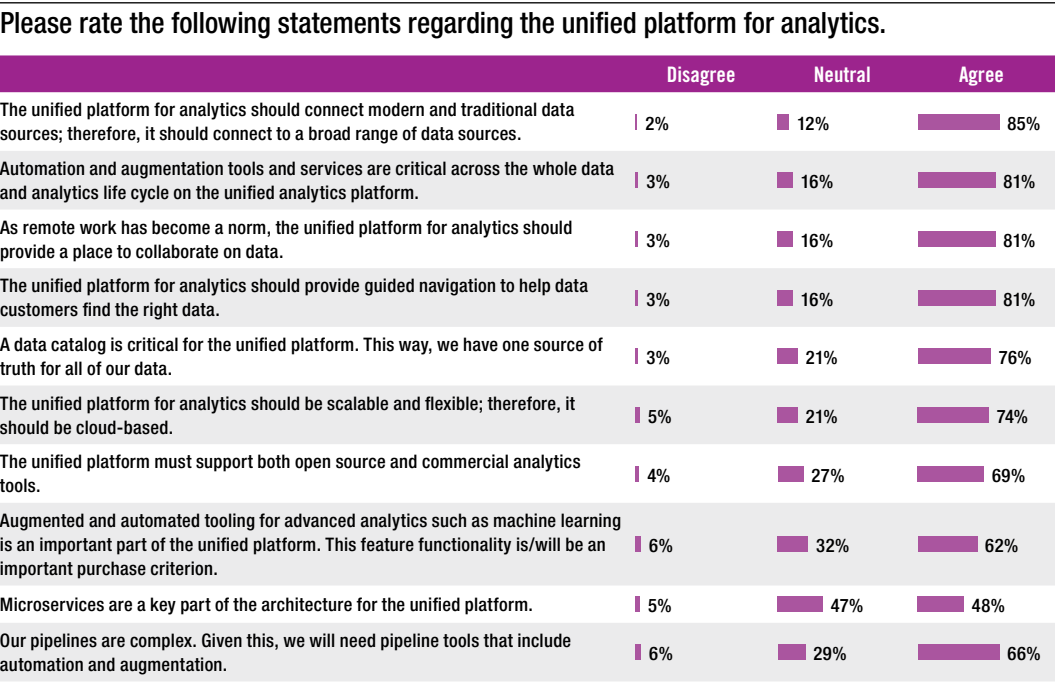


Figure 5. Based on 398 responses. “Completely disagree” and “disagree” responses combined. Completely agree” and “agree” responses combined. Rounding accounts for totals that do not equal 100%.

Modern Analytics and Unification

According to the majority of respondents, an important goal for the unified platform is to support advanced analytics.

We’ve mentioned some of the modern analytics tools that should be part of the unified platform. These include self-service capabilities for data visualization as well as predictive analytics/ machine learning algorithms, NLP, other AI technologies and augmented tools for surfacing insights and building models, and model validation, testing, and deployment tools. According to three-quarters (76%) of respondents, an important goal of the unified platform is to build advanced analytics, such as machine learning (not shown).

There are important issues organizations should consider as they move past data visualization to more advanced forms of analytics. These include skills for both building out and deploying advanced analytics such as ML. For instance, organizations will need to hire data scientists to build models. They may want to upskill business analysts as well. Often, however, organizations spend an inordinate amount of time worrying about who will build models and don’t consider how they will be operationalized in production. They don’t necessarily consider how they will ultimately scale out and grow advanced analytics deployments.

Another important issue is that many organizations think primarily about using structured data for analytics, rather than also utilizing text or image data or other unstructured data that they are *already collecting*. Finally, many organizations haven't yet considered how they will govern advanced analytics.

Operationalizing models. This means putting models into production, in either a system or an application. If you don't put a model into production, it is difficult to get the full value from it. Operationalizing models includes validating that the model will work in the production environment. It includes deploying models as well as monitoring them in production and retraining them when needed. It relies on the fact that you've registered the models somewhere so you capture their metadata, including who built the model, the version number, when it was built, the attributes used in the model, and so on.

As noted, 76% of respondents believe an important goal of the unified platform is to build advanced analytics. Although the software to build, validate, test, deploy, and monitor models should be part of an analytics platform, the actual deployment/execution may happen against production data in an operational system or application, at the edge (i.e., on or close to where the data is created or collected), or elsewhere. Where deployment happens is dependent on a number of factors, including the amount of data feeding the model, whether the data is batch or real time, and the model use case. About a third (39%, not shown) of the respondents stated that they would run the model on the platform.

As the number of models starts to scale, a data science team can't operationalize them all manually. Some organizations may hire members of a MLOps, DevOps, or ModelOps team. This team is responsible for putting models into production and monitoring them there. They are also often involved in the data pipeline process. To empower limited resources with more capabilities, vendors are offering tools to help with deployment. This might include one-click deployment. It may include the capability to deploy models in containers, which involve packaging up software code and all of its dependencies in a container so the software can run on practically any infrastructure.

In our survey, the majority (56%) of respondents agreed that the unified platform should have the ability to deploy models in containers (not shown). The Ops team can work with tools and automated dashboards that collect information about models to manage, monitor, and update hundreds of models (or more) in production. All of these services can help to keep operations manageable.

Governing advanced analytics. Of course, in order for models to succeed in production, the business needs to trust them. Just as organizations realize how important trust in data is, so too must they trust their models. This means that there should be a governance process in place. Individuals must be able to understand the model that they built and the output of the model. They must be able to ensure that it is not biased and meets compliance obligations. To use models effectively, explainability and interpretability are important. Organizations should look for tools that provide explainability, especially in the case of autoML. Governance and data ethics for the unified platform are described in more detail later in this report.

Supporting multiple personas. Support for multiple tools in the unified platform will most likely result in support for multiple personas involved in data and analytics. On the data front, this may include data engineers or DevOps who are building pipelines and putting models into production. On the analytics front, this can include the data scientist as well as others who deal in analytics, including the business analyst/data analyst and even the business user, perhaps using some of the automated tooling described previously. Eighty percent of respondents (not shown) agreed with the statement, "The unified platform for analytics should support all user

As the number of models starts to grow, organizations will often need to build out their team to include an operations group (sometimes called MLOps, DevOps, or ModelOps).

Just as organizations realize how important trust in data is, so too must they trust their models.

A unified platform for modern analytics should support multiple personas.

personas including (but not limited to) data engineers, business users, business analysts, data scientists, application developers, CAO/CDO, or citizen data scientists.” The interface should support these multiple personas and the data underneath should be common and trusted.

USER STORY: UNIFIED ENVIRONMENTS FOR ADVANCED ANALYTICS

According to a data architect from a medical manufacturing company, “Our data science efforts were growing rapidly, especially with unstructured data from social media and IoT devices. This led to a huge amount of data.” The company wanted a unified environment and merged its data warehouse and data lake, including data management, quality processes, master data management (MDM), and data governance. The process took about a year. The platform now supports more advanced analytics, such as machine learning for predictive maintenance.

“We wanted to bring everyone together and move into predictive and preventive maintenance with new data types,” according to the architect. However, the platform supports multiple users. For instance, there is a GUI used by multiple personas, including business analysts and data scientists. “Our company is data-driven and top executives were on board. They are interested in radical thoughts and wanted people to collaborate.”

Accomplishing Unification

How can organizations accomplish unification? There are several different approaches (see Figure 6).

Physical and Logical Consolidation

Organizations are using a range of approaches to accomplish unification.

Tightly integrated stack or full stack. Thirty-four percent of respondents said they will utilize a tightly integrated stack, making this choice the top contender. In an integrated stack solution, vendors typically form partnerships to make sure their products are tightly integrated. Partnerships also help users by adding more capabilities. These partnerships should be technical partnerships rather than sales and marketing partnerships. There should be a reference architecture in place for multitool solutions. For example, a pipeline vendor may work with certain cloud data warehouse vendors at a technical level to enable pushing compute-intensive tasks to the cloud layer. In a separate question, 52% of respondents stated that their organization would only use a stack if there were technical partnerships in place so they don’t have to worry about tools and services not integrating easily (not shown).

Another 17% stated that they would use one single platform from one vendor for unification. This full stack is another way to go. Some vendors might provide an end-to-end analytics platform consisting of products and offerings that are already tightly integrated and allow integration with open source frameworks, orchestration, collaboration, hand-offs, repeatability, and automation across the different stacks/layers/stages of the unified platform. These vendors might also have partnerships with cloud providers or DW/DL vendors. Some may use a completely different approach, such as data mapping and semantic layers.

The reality, however, is that a single stack will involve multiple vendors, technologies, and partnerships working together. It’s about bringing natural components together and making compromises where it makes sense.

Build it themselves. Thirty-four percent of respondents were going to build the platform themselves—either by using open source tools (17%) or using commercial and open source tools (17%). Some of these respondents are consultants who would want to build a platform for their clients. Others have more of a developer focus and are looking to build a platform to primarily support more advanced analytics. For instance, they may be ramping up their data science effort and building more applications. They may have the skills and feel they need to build the platform themselves using tools their organization likes.

Of course, there are risks associated with building a platform from the ground up, including time spent building the platform versus analyzing data, time spent by data scientists trying to access data, and governance across the entire life cycle with pure open source tools. In some cases, organizations may feel they have no choice but to build it themselves because of their particular requirements.

Data fabric approach. A data fabric stitches together multiple data sources. Data virtualization—a data fabric approach for data integration—integrates heterogeneous and distributed data across multiple platforms without replicating it. It creates a single, virtual data layer that unifies data and supports multiple applications and users. Data virtualization can create logical views in which the data looks consolidated, though the data has not been moved or physically altered.

In this survey, 15% of respondents were using/going to use data virtualization, although 68% felt it was a good choice for the unified platform because it gets rid of data replication issues (not shown). A good data virtualization platform supports analytics, a catalog, self-service, and strategies for optimizing cross-platform performance (such as dynamic query optimization, caching capabilities, summary tables, or in-memory computing), even across multiple cloud providers. The virtualization layer connects to multiple BI and analytics tools and other consuming applications such as data science tools (for example, Apache Zeppelin or Jupyter).

How is your company accomplishing unification?

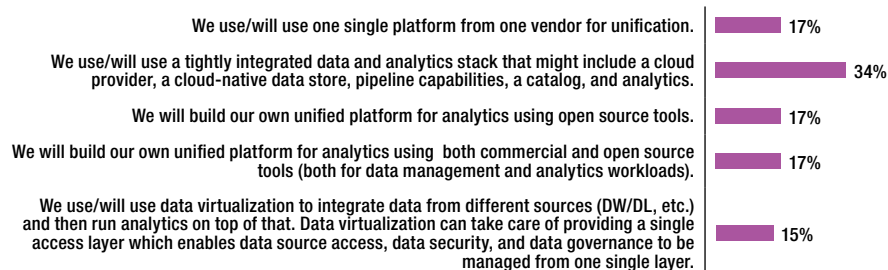


Figure 6. Based on 398 respondents.

USER STORY: USING OPEN SOURCE AND COMMERCIAL TOOLS TO BUILD A UNIFIED PLATFORM

“Several years ago, we had a data warehouse that wouldn’t scale,” said a data analyst at a financial services company, “We wanted to expand our scope to include other data sources such as streaming data and semistructured data.” At the same time, the company’s data warehouse was near its end of life. The company decided to build a new on-premises platform to support consolidated data for decisions, risk computations, and consolidated audit-trail reporting. The platform included a unified and centralized data lake.

A big driver for the platform was to support advanced and real-time analytics as well as AI/ML initiatives. “We wanted an open architecture and the ability to use self-service analytics tools. We didn’t just want one analytics tool to work with the platform because we know each tool has limitations.”

The company ultimately built a platform that included Spark clusters, Kafka for message capture, and Impala for querying. “Unification is critical for fundamental parts of the platform,” according to the business analyst. “The platform has worked well in some use cases. It is important to have broad choices for multiple constituents in the organization and freedom of movement.” In the future, the organization plans to incorporate a cloud-native data warehouse as part of the platform.

Organizational Considerations for Unification

Sixty-nine percent of respondents agreed that IT should own the unified platform.

We’ve explained that MLOps should be responsible for putting models into production once the organization starts to scale its efforts. However, there are other organizational considerations for unification. For instance, who owns the unified platform? The jury is out on this. In this survey, we asked whether respondents agreed that IT should own the unified platform. Although 69% agreed they should, 10% disagreed and 21% were on the fence (all not shown).

Ownership will need to be determined. If IT owns the platform, professionals in that department will have to be flexible enough to allow the business to get the most value from it. For instance, in a separate question we asked who would use the platform on a day-to-day basis. Data analysts (58%), business analysts (55%), and data scientists (45%) were among the top users (all not shown). That means that the business should be a stakeholder. In some cases, it may be that an analytics group in a center of excellence (CoE) within the business owns the platform. In other TDWI research, respondents cite a CoE (or some sort of expert team) as important for analytics success. The platform may be owned by the CAO/CDO organization, if the company has one. In any case, the owner of the platform will need to ensure governance of the platform, including responsible analytics use.

Additionally, if the organization will use advanced analytics, culture and mindset will be important for success. Achieving an analytics mindset will involve a number of factors. For instance, it is important to initially target a high-value use case and deliver results in a short amount of time. That demonstrates an end-to-end use case and gets it up, running, and maintained. Then tackle the next case, then the next, and so on. Old processes are ultimately replaced with new ones. TDWI has referred to this virtuous circle in other reports. As organizations begin to reap the benefits of analytics, they tend to put more advanced analytics in place. The success builds on itself.

Another consideration is to evangelize the value of the platform analytics so people truly understand how it can help. This also means providing appropriate training. Individuals in the organization will require different levels of training. For instance, a model may be put into production on the operations floor or in a call center. Those consuming the results of the model will need to understand how to trust the model’s results and understand how to use the output.

Barriers to Unification

In the perceptions of survey respondents, there are a number of potential barriers for the unified platform for modern analytics (see Figure 7). These barriers aren't necessarily about the lack of analytics or big data skills or lack of security or even multicloud anxiety. The top barriers are more business-focused and include vendor lock-in, politics, and cost.

Vendor lock-in. Often, organizations do not want to be tied to one vendor or face substantial costs or technical incompatibilities when they look to change vendors/providers. Twenty-six percent of respondents chose this as a top barrier to the unified platform. In other words, they like the idea of trusted data and easy-to-access analytics, but some respondents are concerned about being dependent on a vendor and incurring substantial switching costs if they want to move. Some organizations believe the benefits outweigh the risks.

In some cases, this is why organizations might use open source and build the platform themselves. In other cases, organizations will look to vendors that support multicloud environments (e.g., the ability to run software in one of many cloud environments) so they are not tied into one cloud provider (although they may still need to pay for data migration).

You can reduce your reliance on a single vendor by ensuring those that you contract with have flexible technology (API-driven, open-source programmable, runs across multiple cloud providers, etc.) and are stable organizations so you expect they'll be around for the next several years.

The politics of ownership. Many companies are used to data silos or shadow IT where organizations own their analytics and data sources. Politics within an organization can hinder efforts for a unified platform because politics are based on opinions, experience, and bias. Twenty-four percent cited the politics of platform ownership as a challenge. For instance, if an individual was responsible for building an on-premises data warehouse, they are invested in this platform and it may be hard to convince them to move to a cloud platform. Some individuals will fear job loss if the platform moves to the cloud. Some want to maintain or increase their power. To hold onto power, they may say the project is too expensive or isn't needed yet. In this survey, close to a quarter of the respondents cited this as a challenge.

In some cases, the politics of ownership can be overcome with an executive on board who can articulate the business value of the unified platform and set the strategy to allay fears. In other cases, overcoming politics may involve many conversations and meetings.

Cost and time of migration. Almost a quarter (23%) of respondents pointed to the time and cost of migration. Migrating data to another environment can take time and does incur cost. Data movement is expensive. The respondents who selected this barrier were more likely to consider the data fabric option because it enables the organization to leave its data in place. Typically during a migration there can be issues with poor-quality data (which is also cited as a top concern), workloads need to be matched to the right platform, and so on. At TDWI, we typically advise organizations to perform a deliberate phased migration and to use the migration effort to improve the data.

Vendor lock-in, the politics of ownership, and the cost and time of migration were the top perceived barriers to migrating to a unified platform for modern analytics.

What barriers to implementing a unified platform for analytics have you/do you expect to face?

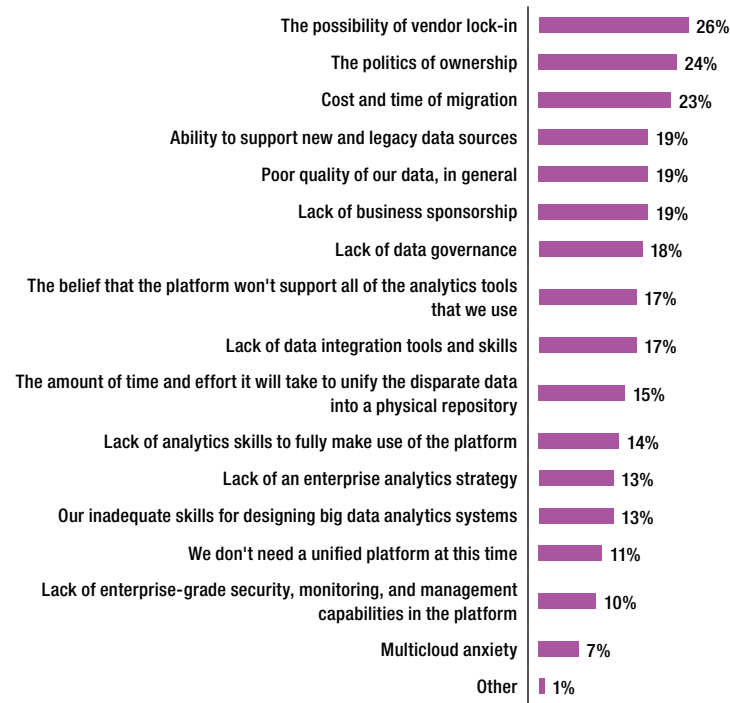


Figure 7. Based on 398 respondents. A maximum of three responses allowed.

EXPERT OPINION

According to Dr. Prashanth Southekal, founder and managing principal of the DBP-Institute, there are four main building blocks in data analytics: data capture, data integration, data science, and data visualization. Southekal sees today's data capture issues as an obstacle for realizing unified data platforms. "Complexity is only going to increase as we collect more data with different formats and frequencies. For instance, before 2010, much of the customer data was captured on ERP and CRM systems. Today we have customer data coming from mobile devices, transactional systems, websites, call centers, bots, and more. How do you stitch all of that customer data together to do 360-degree customer analytics and offer personalized products, services, and experiences to your customers?"

Southekal believes these specialties within the digital, data, and analytics industry are not talking the same language. He points out that, "Basically, in the field of data capture, we talk of best-of-breed decentralized solutions over integrated monolithic ERP platforms, but in data science we talk about unification (i.e., a unified platform for data engineering and data science). Unless we address data issues in the source systems (that is, in the data capture systems), we will continue to have issues in deriving insights because garbage-in leads to garbage-out."

In terms of a unified platform for analytics, Southekal believes that a single source of truth is hard to come by. "BI and analytics vendors talk about the unified platform as a single source of truth. MDM and ERP vendors have been talking about this for a long time. It is not going to be easy because it is an organizational change management issue more than a technology issue. Going forward, with new data sources, and increased business disruption (due to mergers, acquisitions, and divestures), achieving a single source of truth will be very hard. We need leadership on data capture and data integration."

The goal, according to Southekeal, is to “get good quality data into whatever canonical data store you want to use, be it a data warehouse, data lake, or data mart. Focusing on activities in the early stages of the data life cycle will result in better data quality and ultimately better insights.”

Governing the Unified Platform for Modern Analytics

Data governance is critical in the unified platform for modern analytics. This new platform will house modern data types and support modern analytics. It may be in the cloud. Someone must own this new data and be responsible for it and the analytics that comes from it. Someone will need to determine, for instance, what data quality means for new data types such as text or image data. This role involves ensuring the data is trusted and in compliance. The role also governs the ethical use of data for analytics.

Today, data governance involves people and processes as well as technology. The environment is simply too complex for manual governance alone. We asked respondents about the governance tools they are using or plan to use for the unified platform (see Figure 8).

Data catalogs are critical for governing the unified platform. The top answer was the data catalog (52%). We saw earlier that the data catalog was an important feature of the unified platform. In terms of governance, modern catalogs can help improve users’ trust in the data on the platform. For instance, some data catalogs have rating, review, and recommendation features so users can see how others rated the data. Some allow the data steward to certify or “badge” the data as trustworthy. Others allow users to post comments about the data and the use cases for which they worked with it.

Modern catalogs help improve users’ trust in the data on the unified platform.

All of this can help build trust in the data. Catalogs also assist governance by providing metadata about the data—who created it, when it was created, how it is structured, where it originated, how it has been used, and what policies are enforced. Some catalogs provide visibility into data use across the cloud data platform and help track data lineage—metadata that describes where data originated and how it has been transformed, consumed, and shared. This is critical for risk management in a data governance program. Of course, model governance is also important (see below).

Data protection is important. Respondents also felt that data protection was important in the unified platform (42%). A key to protecting sensitive data—at rest, in motion, and in use—is to make it unreadable or unusable to potential intruders. If it is unusable, then it is not valuable. Encryption is a method of protecting sensitive data using algorithms that scramble the data into an unreadable form, which is only reversible with the right key.

Controls over data access are also important to prevent unauthorized data access. Authorization verifies what you are authorized to do with the data (or application). It looks at issues such as who can access the data and what permissions or privileges does that user or role have regarding different data elements. Organizations will need to put a process in place to address identities, groups, and roles and assign access rights to establish a level of managed access. Some organizations might be interested in fine-grained access-control policies such as table- or row-level access. Thirty-eight percent stated that tools to manage user permissions and role-based access were in place or would be in place.

Along with encryption and access, disaster recovery plans should be implemented, and data should be tracked and monitored so an enterprise can understand who accessed the data and when they accessed it, among other information.

Over a third (36%) of respondents cited data discovery and automated data quality tools as important for data governance.

Data discovery and automated data quality. Organizations are often considering data discovery tooling and automated data quality. In this survey, for instance, 36% cited data discovery and automated data quality tools as important tools for data governance. TDWI has seen more interest in automating data quality using techniques such as machine learning. A simple example of this would be training an ML algorithm to learn that “10 Main Street” is the same as “10 Main St.” to reduce data duplication. Other algorithms might identify poor-quality data.

Model governance. Twenty-one percent of respondents plan to have or already have a centralized catalog and storage for all types of analytics/models (for search/query, updates, backup and restore capabilities, overwrite protection, and event logging). For example, models will need to be registered to collect data about who built the model, when they built it, who has touched it, important attributes in the model, and so on. This will help keep track of the models and information about the models to ensure the correct model is in production and is being managed.

What governance tools have you/are you planning to implement as part of the unified platform for data and analytics?

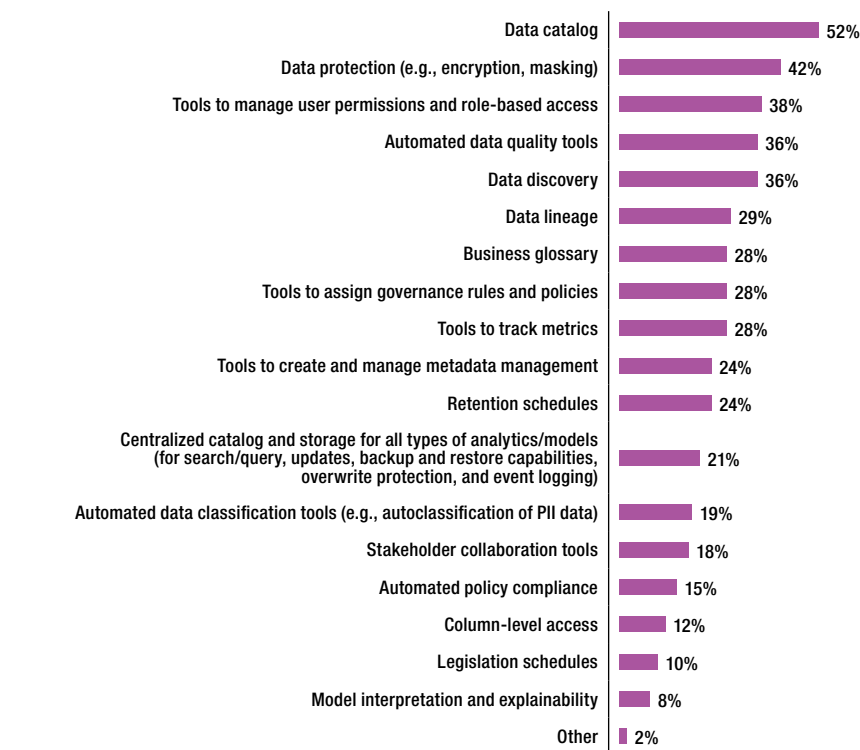


Figure 8. Based on 398 respondents. Multiple responses allowed.

Model interpretability and explainability. This answer ranked at the bottom of the list. However, at TDWI we believe this is extremely important, especially with the use of more automated tools to help build machine learning models. Many ML models act as black boxes, meaning their inner workings are not available to the user. A neural network is a good example. This can mean that there is no way for the user to understand how predictions are derived, what they mean, or how to interpret and explain the results. Interpretability is about understanding an algorithm and how changes to the algorithm can change the output. Explainability is about understanding the *why* behind an ML prediction the way a human can. Both are important in the unified platform for modern analytics because often the platform will produce models automatically. However, someone will need to be able to defend the output.

Ethical Use of Data

Hand in hand with interpretability and explainability is the ethical use of data. As organizations begin to perform more advanced analytics on their unified platform, they will need to be careful to use data in a responsible way. Data ethics involves moral and ethical issues for data and analytics. For example, machine learning may be used to determine who gets a credit card or who is selected for a job interview. Models learn from data and often social bias already exists in that data. An important form of bias in machine learning is discriminatory bias based on the data used to train a model. It is important to understand the bias inherent in data to ensure fair predictions across all groups.

Organizations will need to carefully determine how to train their employees in moral and ethical issues associated with data that consider responsible data use as well as analytics needs. Training will include all parts of the organization that use data, such as analytics, IT, marketing, and other stakeholders. For example, some companies are developing a data ethics office that might be found in the CIO or CDO organization. Although it is early days, this is something that organizations need to consider.

Governance will involve the ethical use of data as well as ensuring it is in compliance and is trusted.

Recommendations

This report has detailed many best practices for the unified platform for modern analytics. In closing, we summarize the report by listing the top best practices for successful unification, along with comments about why each is important. Think of the best practices as recommendations that can guide your organization into successful unified platform implementations.

Know why you're unifying. As evidenced in this report, not all companies need a unified platform. Some organizations are well served by their data warehouse on premises against which they run dashboards or other reports. Other enterprises are concerned about vendor lock-in. Yet, as many organizations mature, they realize the need for more advanced analytics to solve business problems, and the unified platform for modern analytics may be an important part of their journey. It will be a matter of weighing the benefits and the challenges along with the use cases to determine if a unified platform is right for your organization.

Understand the different unification approaches. As illustrated in this report, organizations can use multiple approaches to unifying the data and analytics to support modernization. This may include a cloud-based stack, if it makes sense to migrate data to the cloud for its scale and computational power. It may include a data fabric approach if your organization has a hybrid environment and wants to keep it that way. The approach your organization uses will depend on its specific circumstances and business needs.

Stay up to date on new tools. Respondents to this survey selected a basic set of services for the unified platform, but they agreed that many others were important. These include augmented and automated services for data profiling, quality, data mapping, and analytics; data catalogs to build trust in data; modern pipeline tools to handle numerous pipelines; autoML; and MLOps. These services will be important to handle the complexity and scale of a modern environment. Stay up to date on new tools so your organization is ready when the need arises.

Plan to put analytics into production. As organizations mature and start to build more advanced analytics, such as machine learning models, they should put them into production. Some platforms will support scoring models inside the platform. Some support containers. To scale the number of models in production, organizations must consider having MLOps tools (and team members) in place to help version, validate, and monitor models in production.

Don't forget about model governance. Organizations are used to thinking about data governance; they don't necessarily think about analytics and model governance. Yet this will be critical as organizations start to build and scale machine learning and other models in production environments and build business applications that utilize models. The only thing worse than not trusting a model is thinking it is accurate when it is not. Model governance cannot be an afterthought. Plan for it in the unified platform for modern analytics.

Support multiple personas. The unified platform for modern analytics should support multiple individuals including data scientists, developers, data engineers, business analysts, and business users and give them the appropriate GUI they need. A data scientist may want a notebook GUI. A business user may want a natural language interface. This is important as more users across the organization use analytics.

The only thing worse than not trusting a model is putting an inaccurate one into production.

Support open source and commercial tools. The reality in today's environment is that individuals in the organization may want to use either open source or commercial products or a combination of the two—especially for analytics. The unified platform should support both.

Embrace a culture of collaboration. Bring all stakeholders together to collaborate, regardless of who owns the platform. Internally, collaboration is important for trusting data and for building out the best analytics and applications. Collaboration can also be important for partners.

Organize to execute. As organizations mature and scale out their advanced analytics, they will need to train users to leverage these tools. They will also likely need new staff, such as MLOps and/or data engineers. Ownership of the platform will need to be determined.

Sponsors



Alation is a leader in enterprise data intelligence solutions including data search and discovery, data governance, data stewardship, analytics, and digital transformation. Alation's initial offering is a popular choice in the data catalog market. Thanks to its powerful behavioral analysis engine, inbuilt collaboration capabilities, and open interfaces, Alation combines machine learning with human insight to successfully tackle even the most demanding challenges in data and metadata management.

More than 250 enterprises drive data culture, improve decision making, and realize business outcomes with Alation, including AbbVie, American Family Insurance, Cisco, Exelon, Finnair, Munich Re, New Balance, Pfizer, Scandinavian Airlines, and US Foods. Headquartered in Silicon Valley, Alation was named to *Inc. Magazine's Best Workplaces* list and is backed by leading venture capitalists including Blackstone, Costanoa, Data Collective, Dell Technologies, Icon, ISAI Cap, Riverwood, Salesforce, Sanabil, Sapphire, and Snowflake Ventures. For more information, visit alation.com.



Denodo is a leader in data virtualization providing agile, high-performance data integration, data abstraction, and real-time data services across a broad range of enterprise, cloud, big data, and unstructured data sources at half the cost of traditional approaches. Denodo's customers across every major industry have gained significant business agility and ROI by enabling faster and easier access to unified business information for agile BI, big data analytics, web and cloud integration, single-view applications, and enterprise data services.

The Denodo Platform offers broad access to structured and unstructured data residing in enterprise, big data, and cloud sources, in both batch and real time, exceeding the performance needs of data-intensive organizations for both analytical and operational use cases, delivered in a much shorter time frame than traditional data integration tools.

The Denodo Platform drives agility, faster time to market, and increased customer engagement by delivering a single view of the customer and operational efficiency from real-time business intelligence and self-serviceability.

For more information visit www.denodo.com, follow Denodo via [Twitter](https://twitter.com/denodo) (@denodo), or contact us to request an evaluation copy at info@denodo.com.

Sponsors



Incorta is the data analytics company on a mission to help data-driven enterprises be more agile and competitive by resolving their most complex data analytics challenges. Incorta's Direct Data Platform gives enterprises the means to acquire, enrich, analyze, and act on their business data with unmatched speed, simplicity, and insight. Backed by GV (formerly Google Ventures), Kleiner Perkins, M12 (formerly Microsoft Ventures), Telstra Ventures, and Sorenson Capital, Incorta powers analytics for some of the most valuable brands and organizations in the world. For today's most complex data and analytics challenges, Incorta partners with *Fortune* 5 to Global 2000 customers such as Broadcom, Vitamix, Equinix, and Credit Suisse. For more information, visit <https://www.incorta.com/>.



Matillion makes the world's data useful with an easy-to-use, cloud-native data integration platform. Matillion's low-code and no-code platforms deliver rapid returns on cloud investments for global enterprise customers, helping them wield data as their most strategic asset.

Matillion solves a key challenge for enterprises, helping them overcome the struggle to capture and innovate with data to quickly unlock the insights they need. Matillion's cloud-native platform enables enterprise organizations to transform raw data into the refined, analytics-ready data required to support business intelligence, visualization, artificial intelligence, and machine learning. Matillion helps data teams achieve faster time to insight with modular low-code design patterns to speed up development without compromising sophistication. Its visual transformation logic enables cross-team involvement and self-documenting workflows reduce overall maintenance work.

Optimized for modern enterprise data teams, Matillion is built on native integrations to cloud data platforms such as Snowflake, Delta Lake on Databricks, Amazon Redshift, Google BigQuery, and Microsoft Azure Synapse to enable new levels of efficiency and productivity in data programs. This ensures data teams get maximum performance without the need for complex custom code.

Learn how Matillion delivers rapid returns on cloud investments for global enterprise customers at www.matillion.com.

Sponsors



SAP's strategy is to help every business run as an intelligent enterprise. As a market leader in enterprise application software, we help companies of all sizes and in all industries run at their best: 77% of the world's transaction revenue touches an SAP system. Our machine learning, Internet of Things (IoT), and advanced analytics technologies help turn customers' businesses into intelligent enterprises.

SAP helps give people and organizations deep business insight and fosters collaboration that helps them stay ahead of their competition. We simplify technology for companies so they can consume our software the way they want—without disruption. Our end-to-end suite of applications and services enables business and public customers across 25 industries globally to operate profitably, adapt continuously, and make a difference. With a global network of customers, partners, employees, and thought leaders, SAP helps the world run better and improve people's lives. For more information, visit www.sap.com.



SAS is a leader in analytics. Through innovative software and services, SAS empowers and inspires customers around the world to transform data into intelligence. SAS gives you "the power to know."

sas.com

Sponsors



Snowflake delivers the Data Cloud—a global network where thousands of organizations mobilize data with near-unlimited scale, concurrency, and performance. Inside the Data Cloud, organizations unite their siloed data, easily discover and securely share governed data, and execute diverse analytics workloads. Wherever data or users live, Snowflake delivers a single and seamless experience across multiple public clouds.

Snowflake's platform is the engine that powers and provides access to the Data Cloud, creating a solution for data warehousing, data lakes, data engineering, data science, data application development, and data sharing. Join Snowflake customers, partners, and data providers already taking their businesses to new frontiers in the Data Cloud.

snowflake.com



Wavicle Data Solutions, a leading data and analytics consulting firm, empowers companies to make smart decisions with smart data. We help businesses across multiple industries leverage cloud-native technology, machine learning, and AI to capture, analyze, and act on their growing volumes of data.

Trusted for our technical expertise, innovative solutions, and partnership, our rapidly growing roster of clients relies on us to solve their most complex business issues quickly and cost-effectively. Wavicle has been recognized by *Inc. 500* as one of the fastest-growing private companies in America in 2020 and 2019 and by *Crain's Chicago Business* as one of Chicago's fastest-growing companies in 2021 and 2020.

<https://wavicledata.com/>



TDWI Research provides research and advice for data professionals worldwide. TDWI Research focuses exclusively on data management 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 data management 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.



A Division of 1105 Media
6300 Canoga Avenue, Suite 1150
Woodland Hills, CA 91367

E info@tdwi.org

tdwi.org