

The Road to Self-Service Bliss

Overcome Data Bottlenecks
with a Self-Service Data Platform

BY WAYNE W. ECKERSON

DECEMBER 2021

RESEARCH SPONSORED BY PROMETHIUM

 **Promethium**

THIS PUBLICATION MAY NOT BE REPRODUCED OR DISTRIBUTED
WITHOUT ECKERSON GROUP'S PRIOR PERMISSION.

About the Author



Wayne W. Eckerson has been an international thought leader in the data and analytics field since the early 1990s. He is a noted speaker, sought-after consultant, and widely read author. Eckerson has conducted groundbreaking research studies, chaired numerous conferences, and consulted with organizations around the world during his 25 years in the field. Eckerson has also written two books: *The Secrets of Analytical Leaders: Insights from Information Insiders* (2012) and *Performance Dashboards: Measuring, Monitoring, and Managing Your Business* (2005/2010). He has degrees from Williams College and Wesleyan University.

About Eckerson Group

Eckerson Group is a global research, consulting, and advisory firm that helps organizations get more value from data. Our experts think critically, write clearly, and present persuasively about data analytics.

They specialize in data strategy, data architecture, self-service analytics, master data management, data governance, and data science. Organizations rely on us to demystify data and analytics and develop business-driven strategies that harness the power of data. [Learn what Eckerson Group can do for you!](#)



About This Report

Research for this report comes from briefings with Promethium and other software vendors. The report is sponsored by Promethium, which has exclusive permission to syndicate its content.

Table of Contents

Executive Summary	4
Key Takeaways	4
The Promise and Perils of Self-Service Analytics	5
The Promise	5
The Perils	5
Bottlenecks	6
Reality Today	7
Data Analytics Workflow	8
Focus and Efficiency	8
What Do People Really Want?	9
Self-Service Data Platform	9
Step 1: Business User Asks a Question	10
Step 2: Data Analyst Examines Request	10
Step 3: Data Analyst Searches for Data	10
Step 4: Data Engineer Creates Data Set	11
Step 5: Data Analyst Prepares the Answer	11
Data Fabric Architecture	11
Upshot	12
Conclusion: A Culture of Analytics	13
Recommendations	14
About Eckerson Group	15
About Promethium	16

Executive Summary

Most organizations experience data bottlenecks that prevent the free flow of information to business users. Without ready access to data, business users can't make fact-based decisions or respond quickly to problems and opportunities. Data bottlenecks are a major obstacle for organizations that want to increase agility and embrace fact-based decision making.

Most data leaders have launched self-service strategies to help eliminate data bottlenecks. But giving business users access to powerful self-service tools is only part of the answer. It's also imperative to improve collaboration within a data supply chain so business users, data analysts, and data engineers can work efficiently to find answers to business questions.

A self-service data platform with embedded workflows and a built-in data fabric is an ideal way to improve collaboration and reduce data bottlenecks. The data platform also provides all the functionality that business users, data analysts, and data engineers need to do their jobs and resolve requests quickly. A self-service data platform improves the productivity of data analysts and data engineers tenfold while speeding the delivery of insights to business users.

Key Takeaways

- > Data bottlenecks prevent the free flow of information to business users.
- > Self-service tools don't always eliminate data bottlenecks.
- > Rather than require people to serve their own data needs, improve workflows among the constituents in a data supply chain.
- > A self-service data platform uses a data fabric to make enterprise data available to authorized business users.
- > A self-service data platform can make data analysts and data engineers ten times more productive and provides faster answers to business questions.
- > Companies will increasingly turn to self-service data platforms to support the question-to-answer lifecycle.

The Promise and Perils of Self-Service Analytics

The Promise

Liberation. Self-service analytics promises to empower business users to gather data and build their own reports and dashboards. The idea is to accelerate “time to insight” by giving people self-service tools that make it easy for them to find, profile, prepare, visualize, analyze, and share data. It also liberates the business from its dependency on the IT department, which is typically too slow, too costly, and too removed to meet business needs in an agile fashion.

Offloading. For its part, the IT department is happy to oblige. In fact, IT often instigates the self-service analytics initiative. Its goal is to offload hundreds of tickets for low-level tasks (such as data set creation and report development) so IT staff can focus on value-added activities such as supporting digital transformation and cloud migrations. Data leaders also believe that self-service analytics creates happier, more self-sufficient customers and fosters an agile, data-driven organization. In other words, self-service analytics is a proverbial win-win situation—it’s the one thing that business and IT agree on.

***Self-service analytics is a proverbial win-win situation—
it’s the one thing that business and IT agree on.***

The Perils

Unfortunately, self-service analytics has proven much more difficult to implement than expected. There are numerous reasons. (See **“Self-Service: What Could Possibly Go Wrong?”**)

Eye of the beholder. First, “self-service” means different things to different people. An executive thinks self-service is the ability to drill into a dashboard and save the view; a department head thinks it’s the ability to customize a report with new columns populated with data from a spreadsheet; a data analyst thinks it’s the ability to manipulate data sets and develop sophisticated visualizations.

Too many tools. For data analysts, self-service requires five different tools, each with unique graphical interfaces and workflows:

1. **A data catalog** to search, find, and profile data
2. **A query tool** to extract, filter, and load data
3. **A data prep tool** to clean, combine, and massage the data
4. **A visualization tool** to analyze and visualize the data
5. **A reporting tool** to distribute the results to colleagues

It's challenging, time-consuming, and expensive for data analysts to master all five tools and figure out how to pass data among them.

Over- and under-use. Finally, self-service paradoxically fosters both over-use and under-use of analytical tools. Over-use occurs when data analysts use new self-service tools to proliferate dashboards without governance. This creates conflicting metrics and reports, which erode user confidence in data and trust in reports. With too many dashboards and reports containing partial or conflicting information, business users don't know where to turn. They either give up or ask IT to create a custom report, thus recreating the problem that self-service was supposed to solve.

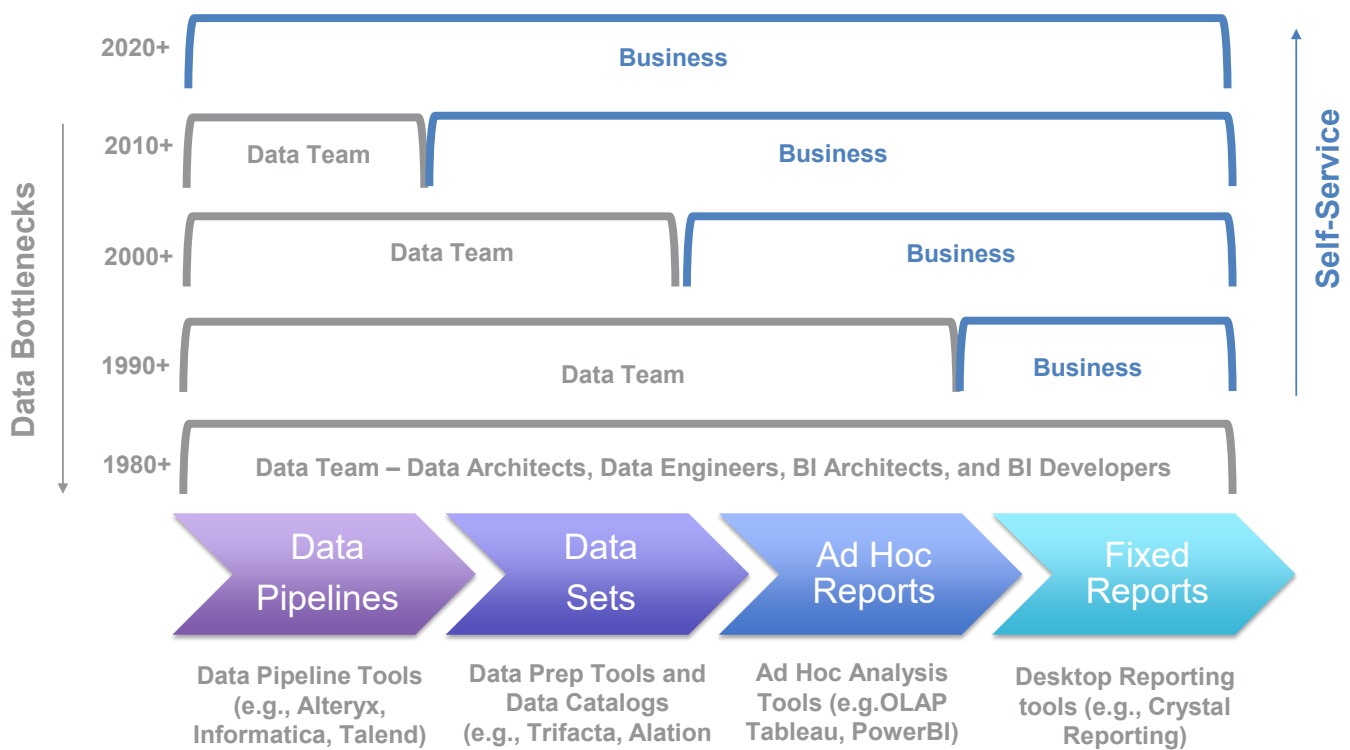
Bottlenecks

A Moving Target

Self-service analytics is designed to eliminate bottlenecks that prevent the free flow of data from data sources to business users. Although we've made significant progress in the past 20 years, data bottlenecks still remain. For the most part, we've simply moved them further up the data supply chain. (See figure 1.)

IT bottleneck. In the 1980s (and still today for many departments), IT or enterprise data teams supported the entire data supply chain, from sourcing data (e.g., data marts and data sets), to building reports and dashboards, to training and supporting users. Here, IT was the bottleneck.

Figure 1. The Evolution of Data Bottlenecks



Data analyst bottleneck. In the 1990s, new business intelligence tools such as Crystal Reports enabled ambitious data analysts to create static reports without assistance from IT. But then these data analysts became a bottleneck for business users who wanted new reports and insights from data.

Data engineering bottleneck. Starting in 2000, new OLAP (online analytical processing) and point-and-click query and visualization tools enabled those same analysts to analyze data iteratively without having to request a static report each time they had a new question. The bottleneck now centered on data engineers who worked on tickets submitted by data analysts and others requesting data sets for analysis.

Data governance bottleneck. Starting in 2015, new data prep tools and data catalogs and glossaries enabled data analysts to create custom data sets from authorized data sources. This shifted the bottleneck to data stewards and governance teams, who document assets in a data catalog and authorize access. Agile governance is critical for keeping up with emerging self-service workflows.

Architecture-as-a-service. Today, we are on the cusp of eliminating data bottlenecks altogether. New point-and-click data pipeline development tools enable data analysts to build data pipelines to feed reports, data marts, or other targets. In essence, they become their own data architects.

Today, we are on the cusp of eliminating data bottlenecks altogether.

There is still a role for the enterprise data team. It needs to manage the central repository of shared data, review the development of user-driven pipelines to ensure they adhere to core architectural and governance standards, and develop cross-functional or enterprise analytic applications that no single department can or wants to build on its own. But, theoretically at least, the data team no longer stands between business users and data.

Reality Today

The reality, of course, is that almost every organization still has all the bottlenecks described above. For example, finance may have its own data engineer who creates a local database, but it still relies on IT to create the data pipeline that feeds that database. Marketing creates its own dashboards and develops complex analytics, but it relies on IT to create custom data sets. Finally, the human resources department still relies on IT for everything, from building data marts and data sets to developing custom reports and dashboards.

So despite the advent of self-service analytics, the enterprise data team—with its data architects, data engineers, quality assurance testers, and BI developers—still has much work to do. But newer self-service

tool sets promise to make it easier for data analysts and data engineers to keep up with demand without IT assistance.

Too few data engineers. Today, the data engineer is the most common bottleneck. Most data leaders complain that they are woefully short of data engineers. The demand for their services comes largely from data analysts and data scientists, who need data engineers to source new and existing data sets. (The project management office also competes for data engineering resources.)

Today, the data engineer is the most common bottleneck.

Although data analysts know how to use Tableau or PowerBI, fewer know SQL or Python or have been given carte blanche access to data sources. Most data scientists know how to source data, but would rather focus on building machine-learning models than querying data. Consequently, most data analysts and data scientists still rely on data engineers to source, clean, and model data.

Data Analytics Workflow

Faced with these bottlenecks, organizations have two options: (1) They can hire more data engineers, or (2) they can teach business users to source and prepare data for themselves.

Most companies find it hard to justify doubling or tripling the number of data engineers on staff. Most also can't find data analysts who are capable of sourcing their own data. This will change gradually over time as more data-savvy analysts enter the workforce and more data-driven executives run organizations.

Focus and Efficiency

A bigger issue is whether organizations want data analysts to spend valuable time engineering data. Perhaps data analysts should focus their time on preparing, analyzing, and visualizing data rather than sourcing, combining, and cleaning data, which a data engineer can do much more efficiently. Likewise, perhaps business users should focus on asking questions and reviewing answers rather than preparing, analyzing, and visualizing data, which a data analyst can do much more efficiently.

In other words, why do we want people to perform tasks they weren't hired to do and aren't very good at? In our rush to upskill everyone's capabilities in the name of self-service and data literacy, maybe we're missing the point. Rather than force people to perform functions they aren't equipped for, why don't we simply build more efficient workflows among them? Instead of giving people more data tasks, why don't we give them what they really want? (See figure 2.)

Rather than force people to perform functions they aren't equipped for, why don't we simply build more efficient workflows among them?

What Do People Really Want?

But what do data engineers, data analysts, and business users really want?

- > **Data engineers** want a data platform that uses artificial intelligence to make it easy to find, profile, and query various data sources, join the results, and, if required, turn the query into a data pipeline that delivers data to a target on a scheduled basis.
- > **Data analysts** want a data platform that gives them easy access to a flattened data set that they can massage and manipulate with a data preparation tool and analyze, visualize, and publish with a data visualization tool to answer a business question.
- > **Business users** want a data platform that enables them to ask questions and get answers. Although some will readily query data to get those answers, most would rather not. If a data analyst did the data dirty work for them, they'd be quite happy.

Maybe the goal of self-service should be to liberate people to do their own jobs and not someone else's.

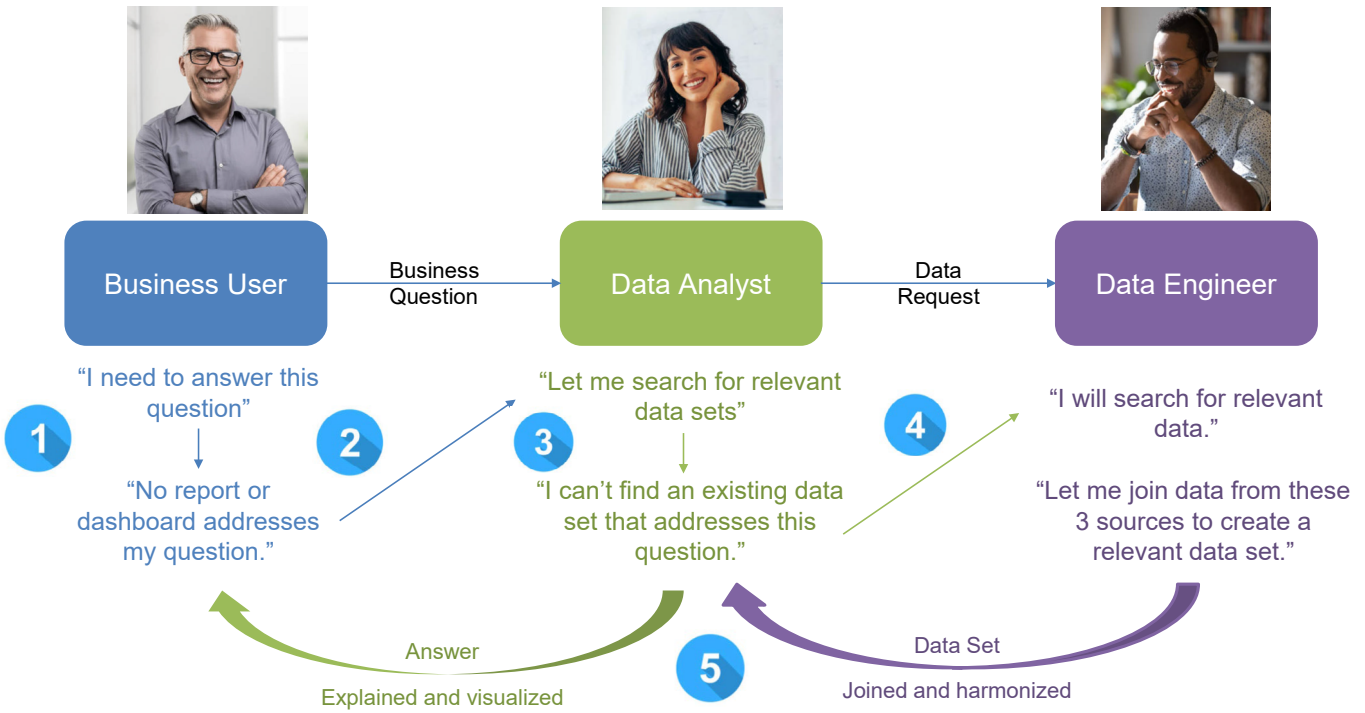
So perhaps the goal of self-service shouldn't be to turn business users into data analysts and turn data analysts into data engineers. Maybe the goal of self-service should be to liberate people to do their own jobs and not someone else's.

Self-Service Data Platform

Rather than giving people a variety of self-service tools, we should provide a self-service data platform that supports every person in the data supply chain and uses flexible, built-in workflows to move data and requests among them.

What should such a data platform look like? It would have separate modules for data engineers, data analysts, and business users and support the sample workflow in figure 2.

Figure 2. Sample Self-Service Data Workflow



Step 1: Business User Asks a Question

Business users begin the workflow by asking a business question. Ideally, they type natural language into a search bar. The data platform then checks whether the question or something similar has already been asked and answered. If so, the platform displays related links, and business users get a qualified answer instantly.

Step 2: Data Analyst Examines Request

If the question is new, the platform places it in a queue for a data analyst. A built-in chat mechanism enables the data analyst to ask the business user clarifying questions and provide progress updates. The data platform also displays the status of the request, such as when it was last updated, how many people have looked at it, and so on.

Step 3: Data Analyst Searches for Data

The data analyst then searches for data sets or queries they are authorized to view. If they find something appropriate, they use the platform to massage the data, visualize the results, and post to the system, which automatically alerts the user. If they don't find a relevant data set, they place a request into a queue for a data engineer.

Step 4: Data Engineer Creates Data Set

A data engineer takes the request from the queue and searches for relevant data using the data platform, which crawls data sources, extracts the metadata, and displays it in a searchable catalog. The engineer then queries relevant data from one or more sources and joins it to create a table. Ideally, the data platform automatically creates the SQL query and generates the table.

Step 5: Data Analyst Prepares the Answer

Once the data engineer has created a data set, the data platform notifies the data analyst via an alert built into the data platform. The data analyst then uses the data platform to analyze and visualize the data and notify the business user that their answer is ready.

This is the most common workflow to support self-service analytics. But organizations can create variations to support business requirements. For instance, authorized data analysts may also fill the role of a data engineer and handle those steps in the workflow. Or a data engineer might create a data set without a request from a business user or data analyst. With a flexible platform, organizations can support multiple types of workflows.

Data Fabric Architecture

To support the sample workflow described above, the data platform needs to embody a data fabric that connects with every data source in an organization and provides a unified view. It also needs to support the full stack of data and analytics capabilities geared to self-service and ad hoc reporting. (See “The Self-Service Triumvirate: The New Analyst Workbench.”) (See figure 3.)

The architectural components of a self-service data platform include the following:

- 1. Natural language processing (NLP).** This enables business users to submit queries as natural language questions and find answers in a search-like interface.
- 2. A question-and-answer index.** The index stores and displays previous questions and answers so business users can search, tag, and annotate them.
- 3. Workflow engine.** The engine supports collaboration among business users, data analysts, and engineers via chat, automated alerts, and social media and usage metrics.
- 4. Data connectors.** Connectors access data from any data source internal or external to the company.
- 5. A metadata engine (data catalog).** This crawls data sources and extracts and tags relevant metadata about source data so it can be easily searched and profiled by data analysts and engineers.

6. **A scalable data fabric.** The fabric supports distributed queries across multiple data sources in real time, enabling data users to discover, query, and prepare data sets from across the enterprise.
7. **A scheduling engine.** This engine enables users to turn a distributed query into a data pipeline that feeds one or more targets with the result set of a query.
8. **A data preparation tool.** The preparation tool enables data analysts and data engineers to clean, combine, and aggregate data.
9. **A data visualization tool.** This tool helps data analysts and engineers analyze, visualize, and share data with business users.
10. **AI engine.** Artificial intelligence generates smart suggestions for searches and automatically builds queries and data sets.
11. **Access control.** This manages user access to data based on governance rules and integrates with enterprise authentication and authorization tools.
12. **Integrations.** Integrations let data analysts and business users access third-party analytics tools to analyze and visualize data sets as desired.

Upshot

A self-service data platform will soon become an essential component in the portfolio of every data and analytics team afflicted with nagging and persistent bottlenecks, especially those that pool around data analysts and data engineers. Such platforms reduce time to insight and help organizations meet the growing demand for data and answers among business users.

10x productivity. A self-service data platform dramatically increases the productivity of highly skilled data professionals, erasing bottlenecks and improving worker morale. Early evidence shows that these platforms can improve the productivity of data analysts and data engineers tenfold. In other words, if you have five data engineers who adequately support 10 data analysts and data scientists today, a self-service data platform will enable them to support 100 data analysts and scientists. And those 10 data analysts who adequately support 50 business users today will be able to support 500 business users.

A self-service data platform effectively eliminates the need to hire dozens of data analysts and data engineers.

In other words, a self-service data platform effectively eliminates the need to hire dozens of data analysts and data engineers to meet demand and remove bottlenecks. But don't be surprised if the platforms are expensive. Per-user subscription prices may exceed those of traditional BI and visualization tools by ten times or more. However, if the platform works as promised, it will be well worth the cost.

The right focus. The best part of a self-service data platform is that it places the emphasis on business questions and answers rather than on queries and data sets. Rather than requiring business users to frame questions in the language of data, these platforms let users ask questions in the language of business. The platform, with the help of data analysts and data engineers, then transforms those questions into answers, shielding business users from the complexity of having to query data systems.

A self-service data platform places the emphasis on business questions and answers rather than on queries and data sets.

Data partners. And rather than casting business users onto a self-service island where they are solely responsible for performing their own data engineering and analysis, self-service data platforms connect business users with data-savvy professionals through embedded workflows. By giving each person the proper tools to carry out their part of the question-to-answer lifecycle, self-service data platforms eliminate data bottlenecks, accelerate insights, and liberate business users and data professionals from the dirty work of finding and massaging data.

Conclusion: A Culture of Analytics

Although self-service data platforms are relatively new, we expect them to play an important role in helping organizations generate value from data assets. Once business users recognize that the platform reduces the time required to get answers to ad hoc business questions, they will ask more questions. This ultimately fosters a culture of analytics.

With greater use, a self-service data platform becomes smarter and faster. With a large repository of existing questions and answers and an AI engine to sort through them, the platform will provide answers to more business questions without involving a data analyst or data engineer. This further accelerates time to insight and lays the foundation for a data-driven organization.

Recommendations

- > Rather than hiring more data analysts and data engineers to address data bottlenecks, purchase a self-service data platform.
- > Rather than giving data analysts and engineers five different tools to meet business requests, provide a self-service data platform that integrates all the functionality they need along with embedded workflows and a data fabric to support the question-to-answer lifecycle.
- > Rather than require business users to learn new data skills, create workflows among business users, data analysts, and data engineers to streamline requests.

About Eckerson Group



Wayne Eckerson, a globally known author, speaker, and consultant, formed **Eckerson Group** to help organizations get more value from data and analytics. His goal is to provide organizations with expert guidance during every step of their data and analytics journey.

Eckerson Group helps organizations in three ways:

- > **Our thought leaders** publish practical, compelling content that keeps data analytics leaders abreast of the latest trends, techniques, and tools in the field.
- > **Our consultants** listen carefully, think deeply, and craft tailored solutions that translate business requirements into compelling strategies and solutions.
- > **Our advisors** provide competitive intelligence and market positioning guidance to software vendors to improve their go-to-market strategies.

Eckerson Group is a global research, consulting, and advisory firm that focuses solely on data and analytics. Our experts specialize in data governance, self-service analytics, data architecture, data science, data management, and business intelligence.

Our clients say we are hard-working, insightful, and humble. It all stems from our love of data and our desire to help organizations turn insights into action. We are a family of continuous learners, interpreting the world of data and analytics for you.

Get more value from your data. Put an expert on your side. [Learn what Eckerson Group can do for you!](#)



About Promethium

Always be ready to answer tomorrow's questions today.

Promethium is on a mission to help every business to be data driven by enabling every employee to make

data driven decisions in real time without the technical complexity of data management. Promethium is the only SaaS-based data management solution that uses AI to answer questions for analytics.

So you can spend your time taking action on insights, Promethium automates data discovery, preparation, query, and visualization without the need to move, migrate, or ETL data. Founded in 2018 and headquartered in Menlo Park, Calif. For more information visit www.pm61data.com.

