

Ultimate Guide to Master Data Management

Six Criteria to Evaluate Products

BY KEVIN PETRIE AND JAY PISCIONERI MARCH 2024

RESEARCH SPONSORED BY





About the Authors



Kevin Petrie is the VP of Research at Eckerson Group, where he manages the research agenda and writes about topics such as data integration, generative AI, data observability, machine learning, and cloud data platforms. For 25 years Kevin has deciphered what technology means for practitioners as an industry analyst, instructor, marketer, services

leader, and tech journalist. He launched a data analytics services team for EMC Pivotal in the Americas and EMEA and ran field training at the data integration software provider Attunity (now part of Qlik). A frequent public speaker and coauthor of two books about data management, Kevin most loves helping start-ups educate their communities about emerging technologies.



Jay Piscioneri has over 25 years of experience in data technologies, including data warehousing, business intelligence, data quality, and data governance. He's worked with organizations in various industries, helping them plan and build their next-generation data platforms. As a leader of many initiatives, Jay brings extensive experience in the

nuts and bolts of implementing data solutions and the organizational challenges of adopting new priorities, processes, and tools.

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

This report is sponsored by Semarchy, which has exclusive permission to syndicate its content.



Table of Contents

Executive Summary	. 4
Introduction	. 5
Evaluation Criteria	11
Recommendations and Next Steps	17
About Eckerson Group	19
About the Sponsor	20



Executive Summary

The need for master data management has never been greater. Without consistent descriptions of business entities such as locations, customers, products, partners, and employees, companies cannot operate efficiently or win in competitive markets. Master data management helps by creating "golden records" that streamline operations and reduce risk.

This report defines five criteria that data leaders and their teams can use to evaluate MDM products: breadth of functionality; ecosystem support; price performance and scalability; ease of use; and vendor profile. For each criterion we recommend specific questions to consider and pose to MDM vendors. We also explore use cases (organized by entity, function, and project type) and architectural patterns (registry, consolidation, coexistence, and centralization.)

As a data leader you should take the following steps to start or extend your MDM journey.

- > **Define your business objectives.** These objectives will vary by industry, company, team, and so on. Enlist an executive sponsor within a key line of business, and define their highest-priority strategic initiatives over the next three years. Then build your MDM plan, define your milestones, set your KPIs, and secure your budget.
- > **Prioritize use cases.** You'll have no shortage of potential use cases for MDM. The key is to assess their alignment with business objectives as well as their feasibility and risk. Identify major sources of pain and look for ways to notch up an early win. Those use cases can be your highest priority.
- > **Define your MDM domains.** Define your data domains, including data sources, estimated rate and volume of change in each system, and data stewards. Also include the systems and users that will consume golden records and an estimate of their rate and volume of MDM data retrieval.
- > Create a roadmap. You can't do everything all at once. Use your business needs, data management maturity, and MDM domain inventory to create a roadmap for MDM initiatives. You must start small by tackling an initial project that balances high-impact results, technical challenges, and your company's readiness for change.
- > Let people and processes guide your product selection. Assign a technical lead for a POC that understands the applications, pipelines, and platforms involved. Then assign a business lead that understands the functional domain—finance, customer service, etc. These two experts must collaborate to align the skills, tools, and processes required for MDM success.



Introduction

The need for master data is nothing new. Eighty-five years ago the **National Geographic Society** gave President Franklin Roosevelt and Prime Minister Winston Churchill detailed maps of the world. This "master" view of countries, islands, and oceans helped them synchronize military operations and win World War II.

By the same token, the need for master data has never been greater. Today companies need master data to synchronize business operations. Without consistent descriptions of business entities such as locations, customers, products, partners, and employees, they cannot operate efficiently or win in competitive markets. The need for consistency rises each year as the complexities of modern business accumulate—and as companies seek to control the risks of artificial intelligence/machine learning (AI/ML) applications. In fact, 80% of AI/ML projects will fail without reliable master data.

80% of AI/ML projects will fail unless they have master data

Master data can help. It comprises standard attributes and terms that data engineers, data stewards, and business analysts use to describe key entities in a consistent, trustworthy way. Master data management (MDM) products match and merge records from various source systems to create master data and assemble it into "golden records." This much-vaunted single source of truth reduces duplicates, resolves discrepancies, and strengthens data governance programs. It enables applications to operate efficiently, and analytics projects to generate more accurate results with lower risk. Six primary MDM vendors are PiLog, Informatica, Semarchy, TIBCO, Stibo Systems and Profisee.

This report defines five criteria that data leaders and their teams can use to evaluate MDM products: breadth of functionality; ecosystem support; price performance and scalability; ease of use; and vendor profile. For each criterion we recommend specific questions to consider and pose to MDM vendors. We start by defining MDM use cases and architectural patterns.

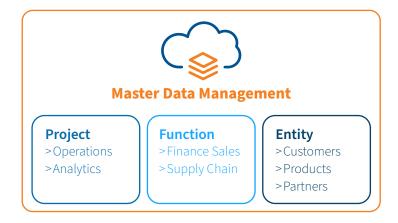
This report defines five evaluation criteria: breadth of functionality; ecosystem support; price performance and scalability; ease of use; and vendor profile

Use Cases

Data teams organize the use cases for MDM according to three dimensions: entity, function, and project type. Let's explore each, working from the bottom up in Figure 1. These dimensions apply across industry verticals.



Figure 1. Dimensions of Master Data Management



- > **Entity**. As described above, master data describes entities such as customers, products, components, offices, and so on. Companies often choose their entity labels based on reference data—i.e., slow-changing categories based on external standards. This might include customer demographic groups, industry product definitions, component metrics, or geographic regions.
- > **Function**. MDM supports a wide range of business functions. For example, finance teams use it to standardize profit metrics; sales managers use it to assign customers and territories to individual reps; and supply chain managers use it to track component shipments by supplier or geography. The more complex the function, the greater the need for MDM.
- > **Project**. Both operational and analytical projects depend on MDM. Operational projects focus on one-time events such as cloud migrations or application upgrades; periodic events such as quarterly earnings reports or annual compliance audits; and ongoing processes such as synchronization of CRM and HR records. Analytical projects range from business intelligence reporting or dashboards to data science.

Data teams organize the use cases for MDM by entity, function, and project type

Know Thyself

Socrates said "To know thyself is the beginning of wisdom." So, before getting into the questions to ask vendors about their products, let's consider the questions that data leaders, engineers, and stewards should ask of themselves. These questions center on business needs, data management maturity, and data environment.

Business needs

The first and most important question to consider: what are your company's business needs for MDM? For example, if you need up-to-the-minute master data to help salespeople improve customer interactions,



then you'll need an MDM product that supports real-time data integration. Or if you need enterprise-level analytics spanning divisions with different transaction systems, you'll need an MDM product that standardizes disparate reference data such as transaction classification codes and status codes.

Data management maturity

It's critical to understand your company's overall data management maturity. For example, more mature companies have designated data owners and stewards, and automated tools for monitoring and remediating data quality issues. They need an MDM product that can be integrated with their existing processes and tools. Less mature companies will need a flexible MDM product to evolve with their program. They also might need help filling other gaps in their data management toolkit with modules such as data cataloging, pipeline development, and quality monitoring.

Data environment

To select the right MDM product, you must understand your data environment, including your sources, their data formats, and the volumes they require for both initial loads and regular updates. Also, consider the tools you already have in place. For example, if you have a well-established data catalog, you need an MDM product that can integrate with it.

Architectural Patterns

Data teams implement MDM products using one of four architectural patterns: registry, consolidation, coexistence, and centralization. Here is how each pattern creates and manages golden records. Companies often choose different patterns for different use cases, and migrate from one to another as they mature and requirements evolve. Keep the advantages and drawbacks in mind—along with your objectives and use cases—as you evaluate potential MDM products.

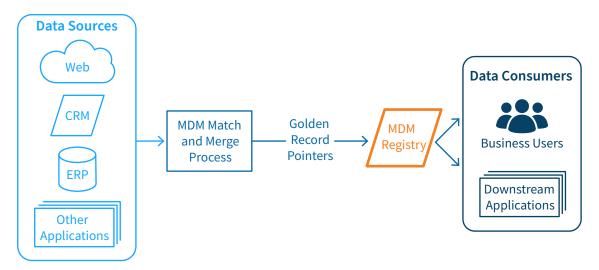
MDM architectural patterns include registry, consolidation, coexistence, and centralization

Registry

An MDM registry comprises "virtual" golden records that point back to source data rather than copying and storing the values themselves. When analytical or operational applications request master data, the MDM system uses pointers stored in the virtual golden records to retrieve data values from source systems. Because the registry gathers and unifies data without altering the sources, you can integrate it with your data environment with minimal disruption to existing processes. This makes the registry faster and less costly to implement than other patterns.

However, retrieving golden record data inline can add unacceptable latency for transactional applications, or any applications that process large datasets. Because the registry does not save data values, it must fetch them from disparate source systems each time, which can drive up compute costs and slow workloads. Source systems don't benefit from MDM-generated data improvements because data updates don't flow back to the source. However, this can be an advantage if you require source system data to remain unchanged. Figure 2 illustrates the registry pattern.

Figure 2. Registry Pattern



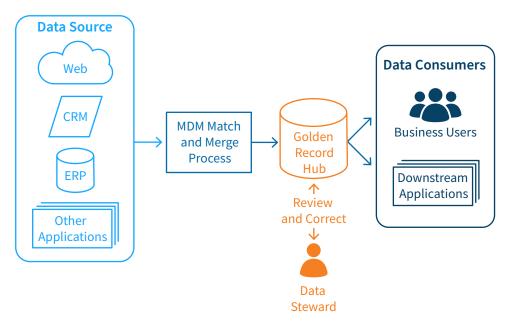
Consolidation

Similar to the registry, the consolidation pattern matches and merges data from various source systems to create a golden record. Unlike the registry, the consolidation pattern writes the golden record to a central MDM repository, also called a hub, where data stewards review the system's updates and either approve or correct them. Downstream applications and users consume the persisted golden record directly from the hub, which eliminates the functional latency inherent to a registry-based pattern.

The human review of data in the hub improves data quality, although it can introduce process latency depending on how quickly data stewards complete their reviews. The consolidation pattern leaves source data unchanged. However, it can be more costly to implement because it requires data storage for the hub as well as a user interface for stewards to update and approve golden records. Figure 3 illustrates the consolidation pattern.



Figure 3. Consolidation Pattern



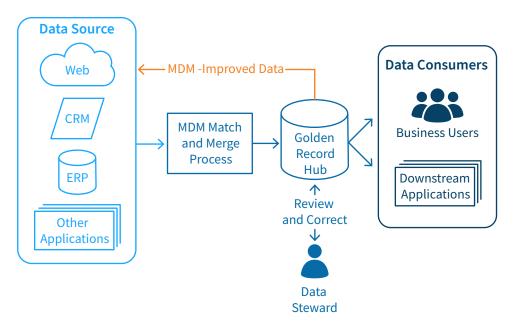
Coexistence

The coexistence pattern is an extension of the consolidation pattern. It maintains a human-reviewed hub of mastered data and adds a feedback loop that distributes golden records to source systems. This avoids the risk of letting the MDM hub become a data quality silo. It also puts higher-quality data into the hands of users in the business applications where they do most of their work.

However, creating the processes that feed data back to source systems can be complicated. Data teams must define the conditions for allowing or preventing the MDM hub from updating source data. Also, for many business systems such as ERP, CRM, and financial systems, data can only be updated reliably through an API, as opposed to writing directly to their respective data stores. This makes the coexistence pattern more costly to implement and maintain. For this reason many companies implement the consolidation pattern first, then upgrade later to add the coexistence feedback loop. Figure 4 illustrates the coexistence pattern.



Figure 4. Coexistence Pattern

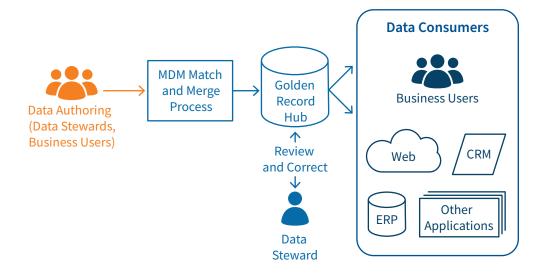


Centralization

In all three of the previous models, master data originates in upstream business applications. The MDM system gathers and updates source data to create a golden record, then sends it downstream to consuming applications and users.

The centralization pattern turns this paradigm on its head by making the MDM hub the point of origin for master data. In other words, teams enter and maintain master data in the MDM hub itself. Operational applications, analytical systems and tools, and data users consume master from the hub. This pattern enables data teams to create a single source of truth. But the time and cost of providing a full-function user interface for authoring master data, and maintaining many complex application integrations, can be prohibitive. Figure 5 illustrates the centralization pattern.

Figure 5. Centralization Pattern



Evaluation Criteria

Now we explore five criteria to help data teams evaluate candidate MDM products: breadth of functionality; ecosystem support; price performance and scalability; ease of use; and vendor profile. Consider each of these criteria in the context of their company's business objectives, use cases, and of course available budget. Ask the following questions of MDM vendors, keeping in mind your current and likely future requirements. To corroborate vendor responses and deepen your knowledge, be sure to study vendor documentation, vendor user forums, and the MDM sections of peer review sites such as **Gartner Peer Insights, G2, SoftwareReviews, SourceForge.**

Criterion 1. Breadth of Functionality

An MDM product's breadth of functionality depends on the architectural patterns, processing methods, and additional data management capabilities it supports.

What MDM architectural patterns does this product support?

Vendors should describe how they support each of the architectural patterns we reviewed above. They should also describe how they help you migrate from one pattern to another. This is important if you plan to start with registry and grow into consolidation and then coexistence as your MDM program matures. If you plan on using the centralization pattern, you need to review the user interface and workflows that products provide for authoring master data. Vendors should also describe how they integrate with downstream applications that need to consume master data from the hub.



What are its data modeling capabilities?

Before they start making golden records, data teams build master data models that define how business entities relate to one another. These models illustrate product categories, organizational hierarchies, inter-relationships between groups of entities, and so on. Ask your vendor how its MDM product helps configure the rules for such models, then visualize the result and share it with colleagues.

Does the product support multiple processing methods, including batch, micro batch, and streaming?

Batch processing is the most common way to update data in MDM systems. It involves collecting source data at regular intervals (e.g., hourly) and integrating it with existing MDM data through the match and merge process. Micro-batches do the same thing but in smaller chunks at more frequent intervals (e.g., every 5 minutes). Some MDM systems support real-time data streaming. This is beneficial for fast-changing data domains such as customer behavior, and for operational systems that require up-to-the-moment master data such as customer 360 applications. Vendors should explain which processing methods they offer for each architectural pattern they support.

Does the vendor have additional data management functions that we need?

Some vendors offer a full data management suite; others are focused on the core functions of MDM. Since you've evaluated your company's data maturity and data environment, you have a good idea of what additional functions you might need. So, focus on MDM products that fill critical data management gaps. Look for vendors that offer modules or products for functions such as data governance workflows, data pipeline development, and data quality management, that you can implement separately.

This gives you the flexibility to introduce new data management platform capabilities and processes at a pace your company can handle. Vendors should provide details on how their modules work together and how you can substitute other tools. For example, you might use a suite's data catalog module for now but move to a robust enterprise data catalog in the future. Be sure you understand what's involved in swapping out the MDM vendor's data catalog module for a best-of-breed product.

An MDM product's breadth of functionality depends on its architectural patterns, processing methods, and additional data management platform capabilities

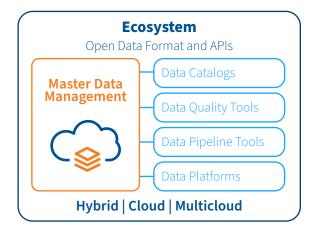
Criterion 2. Ecosystem Support

You can view MDM as the transformation stage of data pipelines that connect sources, registries, hubs, and applications in your data ecosystem. To support this ecosystem, you need your MDM product to



integrate with data platforms, pipeline tools, quality tools, and data catalogs across hybrid, cloud, and multi cloud environments (see figure 6). Evaluate what APIs or custom scripting are required for integration in each of these categories.

Figure 6. MDM Ecosystem



- > Data platforms span data warehouses such as Teradata or Snowflake, lakehouses such as Databricks or Delta Lake, and databases such as Oracle or IBM Db2. They also include platforms in each of these categories from the hyperscalers AWS, Microsoft Azure, and Google Cloud. MDM products must consume data from these sources and provide master data back to them.
- > **Data pipeline tools** include ingestion tools such as **Fivetran**, transformation tools such as **dbt**, and pipeline suites such as **Qlik**, **Matillion**, or **Semarchy Data Integration**. MDM products must deliver master data in open formats that these tools can extract, load, and transform without extra effort by the data team.
- > Data quality tools assess the accuracy, completeness, and timeliness of data that feeds both operational and analytical applications. These types of tools include Monte Carlo, Acceldata, and Anomalo, as well as a number of tool suites that also offer MDM. Whether as individual tools or a suite, these capabilities must complement one another. Data teams should evaluate how a candidate MDM product would contribute to these combined workflows.
- Data catalogs centralize metadata such as file names, database schemas, category labels, and more to help data teams vet inputs for all types of analytics projects. They include tools such as Alation or Atlan, and form part of suites from vendors such as Talend (now Qlik), Semarchy, or Microsoft Azure. MDM products should provide metadata for golden records that data catalogs can incorporate.

To support this ecosystem, you need your MDM product to integrate with data platforms, data pipeline tools, data quality tools, and data catalogs across hybrid, cloud, and multi cloud environments (see figure 6).



You need your MDM product to integrate with data platforms, pipeline tools, data quality tools, and data catalogs

Criterion 3. Price-Performance and Scalability

Much of MDM's value centers on its ability to support operational and analytical processes at scale without delays or disruptions. This capability matters especially for the registry pattern, in which transactional and high-throughput applications must wait for inline retrieval of golden record values from source systems. Regardless of the MDM pattern you choose, you should ask the following questions about service level agreements (SLAs), scalability metrics, and resource requirements.

Can this product meet performance SLAs for target use cases?

Devise proofs of concept that test MDM products' ability to support their most rigorous use cases with low enough latency and high enough throughput. Vendors' documentation and implementation/ support teams should provide guidance about what performance ranges to expect for their environment. Ask each vendor for proof points about the SLAs they met for other customers with similar use cases and environments. These proof points should involve similar applications, data platforms, and cloud or on-premises infrastructure.

Does this MDM product support the necessary user counts, applications, and data platforms?

Scalability also matters. Based on your use cases and architectural pattern(s), scope the user counts, applications, and data platforms your MDM product must support. These dimensions will guide the proof of concept and overall evaluation. You need evidence—test results, documentation, and customer proof points—that the MDM product will support initial requirements and scale sufficiently in coming years.

What infrastructure resources does this product require?

Like all software, an MDM product requires compute, memory, and storage resources. On premises MDM products might consume these resources from a shared pool, impacting other processes, or require the purchase of dedicated servers. A cloud-based MDM product might not impact other processes, but incur variable usage fees from the cloud provider. Forecast your resource requirements in either scenario to understand their impact on business processes, capital expenditure, and operating expenses. To reduce the risk of bottlenecks or cost overruns, you should get granular visibility into CPU/memory utilization metrics and KPIs for similar customer environments.

Evaluate the MDM product's ability to meet SLAs, scalability metrics, and resource requirements



Criterion 4. Ease of Use

To understand an MDM product's ease of use, evaluate its training requirements, level of automation, flexibility, and data integration requirements.

How much training does this product require?

Vendors should provide sufficient training to make users productive during hands-on testing in a proof of concept (PoC). By the time you complete the POC and decide to purchase, experienced team members should feel comfortable handling basic tasks and bringing their less experienced colleagues up to speed. Keep in mind that training time reduces the net productivity benefit of the product.

What level of automation does it provide?

Users should be able to discover and connect to source data, then configure basic match and merge rules by manipulating a mouse and drop-down menus more than typing SQL commands. Your MDM product should guide them through this process with intuitive prompts and recommendations to fix record errors, eliminate duplicates, and reconcile discrepancies. It also should "learn" from user decisions by recognizing patterns and recommending improvements over time. Wherever possible, it should provide different teams with the same consistent, consolidated view of their environment.

How flexible is this product?

Master data and its environment always change. Your MDM product must adapt to changes in data sources, entities, functions, and projects, as well as the architectural approach itself. Scope likely implementation scenarios over the next three years and evaluate the level of effort your MDM product requires to support those changes. This should include both planned project expansions—i.e., adding datasets or teams—and potential exogenous changes such as mergers and acquisitions.

How does it help integrate data?

The creation and dissemination of golden records complements other data transformation processes as part of larger pipelines within your environment. An MDM product therefore must do more than just connect to data sources and consuming applications. It must also share transformation tasks with data pipeline tools (or pipeline modules within suite offerings). Vendos should explain what these combined pipelines look like. How does a data engineer or steward manage the MDM product alongside other pipeline tools? Does the MDM product perform related transformation tasks, such as the formatting and filtering of records, or rely on a separate pipeline tool? Answering these questions helps you assess the overall ease of use.



Ease of use depends on training requirements, level of automation, flexibility, and data integration requirements

Criterion 5. Vendor Profile

A vendor is an important member of your MDM implementation team. To assess their suitability, you need information about their growth trajectory, financial stability, market position, and user community.

How old is the company, and what is its growth trajectory?

A vendor's evolution tells you about its priorities, resilience, and adaptability. Vendors that show steady organic growth demonstrate their solid understanding of market demand and effective research and development. To understand this, you should evaluate how each vendor grew in terms of employees and revenue since inception.

Also learn how they've grown through acquisitions rather than organically. Dig into their product's best features to understand whether they were developed by the vendor or added by acquiring another company. Have them describe how they integrate acquired technology—and read customer reviews to see whether they agree. Avoid vendors with a history of disruptive acquisitions and product releases.

Is the vendor financially stable?

Examine the vendor's financial health through revenue growth, venture capital funding, and profitability. These metrics indicate business viability and market relevance. Venture capital funding is important for growing companies that need time to establish themselves in the market. The degree and stage of funding is a sign of investor confidence that help you evaluate your risk in partnering with an innovative but less proven vendor.

What is their position in the market?

Assess the vendor's market position in terms of their differentiators and installed base. A robust product with forward-looking features indicates a vendor's ability to anticipate demand. Also, look for a broad customer list with compelling implementation stories, especially in your company's industry, and a growing number of product installations.

Does it have an engaged user community?

The vibrancy of a vendor's user community is a testament to the product's value and the company's customer-centric approach. A large and active user community not only provides validation but also fosters an ecosystem of support and continuous learning. Engagement in forums, user groups, and events are strong indicators of a healthy community.



By analyzing these characteristics and selecting the right vendor, you can establish a partnership that's both beneficial and sustainable over the long haul.

To understand an MDM product's ease of use, evaluate its training requirements, level of automation, flexibility, and data integration requirements.

Assess each vendor's growth trajectory, financial stability, market position, and user community

Recommendations and Next Steps

Master data management, while never easy to implement, offers a time-tested way to make business data more consistent and trustworthy. It creates and disseminates "golden records" that streamline operations, increase analytical value, and strengthen data governance programs. Data leaders and their teams should evaluate MDM products by their breadth of functionality; ecosystem support; price performance and scalability; ease of use; and vendor profile.

As a data leader you should take the following steps to start or extend your MDM journey.

- > Define your business objectives. These objectives will vary by industry, company, team, and so on. Enlist an executive sponsor within a key line of business, and define their highest-priority strategic initiatives over the next three years. They might need to improve profitability, modernize on the cloud, or support advanced analytics such as generative AI. Then build your MDM plan, define your milestones, set your KPIs, and secure your budget.
- > Prioritize use cases. You'll have no shortage of potential use cases for MDM. The key is to assess their alignment with business objectives as well as their feasibility and risk. Identify major sources of pain and look for ways to notch up an early win. Those use cases can be your highest priority.
- > **Define your MDM domains**. Create an inventory of the data domains you plan to bring under MDM management. Include the data sources that will contribute to each domain, estimated rate and volume of change in each system, and the assigned stewards who will check and approve golden records. Also include the systems and users that will consume golden records and an estimate of their rate and volume of MDM data retrieval.
- **Create a roadmap**. You can't do everything all at once. Use your business needs, data management maturity, and MDM domain inventory to create a roadmap for MDM initiatives. This roadmap should have two guiding principles.



- **Start small and grow**. Balance high-impact results and their attendant technical challenges with your company's readiness for change. For example, real-time customer 360 data can make compelling impacts on revenue and customer satisfaction. But your company needs a higher level of data maturity and technical skill to realize those benefits. If that's not you, start with smaller, less volatile domains such as reference data.
- **Pick the right pattern**. Be sure to determine the right architectural patterns for your initiatives.
- > Let people and processes guide your product selection. Your data engineers, stewards, analysts, and scientists—as well as your business analysts—should contribute to your evaluation process. Assign a technical lead for a POC that understands the applications, pipelines, and platforms involved. Then assign a business lead that understands the functional domain—finance, customer service, etc. These two experts must collaborate to align the skills, tools, and processes required for MDM success.

18



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 the Sponsor

Semarchy, a leader in the data integration and master data management markets, enables organizations to rapidly generate business value from their data. Its unified



platform empowers organizations of any size, to quickly discover, govern, manage, integrate, and report critical information scattered across applications. Semarchy's focus on ensuring successful data initiatives is why they're:

- > A Gartner® Magic Quadrant Leader for Master Data Management Solutions
- > The most recommended MDM solution on Gartner® Peer InsightsTM in the last 12 months
- > 100% of customers met their implementation goals in 2023

Semarchy's unique value proposition to data leaders is its agile approach, customer success organization and flexible platform allowing businesses to build custom data apps in days and measure ROI in under 12 weeks. Their commitment to ensure successful data initiatives is why over 350+ clients trust them to manage over 1 trillion consolidated master records of mission-critical business data.