



The Master Data Management Dilemma: To Buy or Build, That is the Question: Benchmark Completed

Presented by: William McKnight

"#2 Global Influencer in Big Data" Thinkers360

President, McKnight Consulting Group

3 X Inc 5000

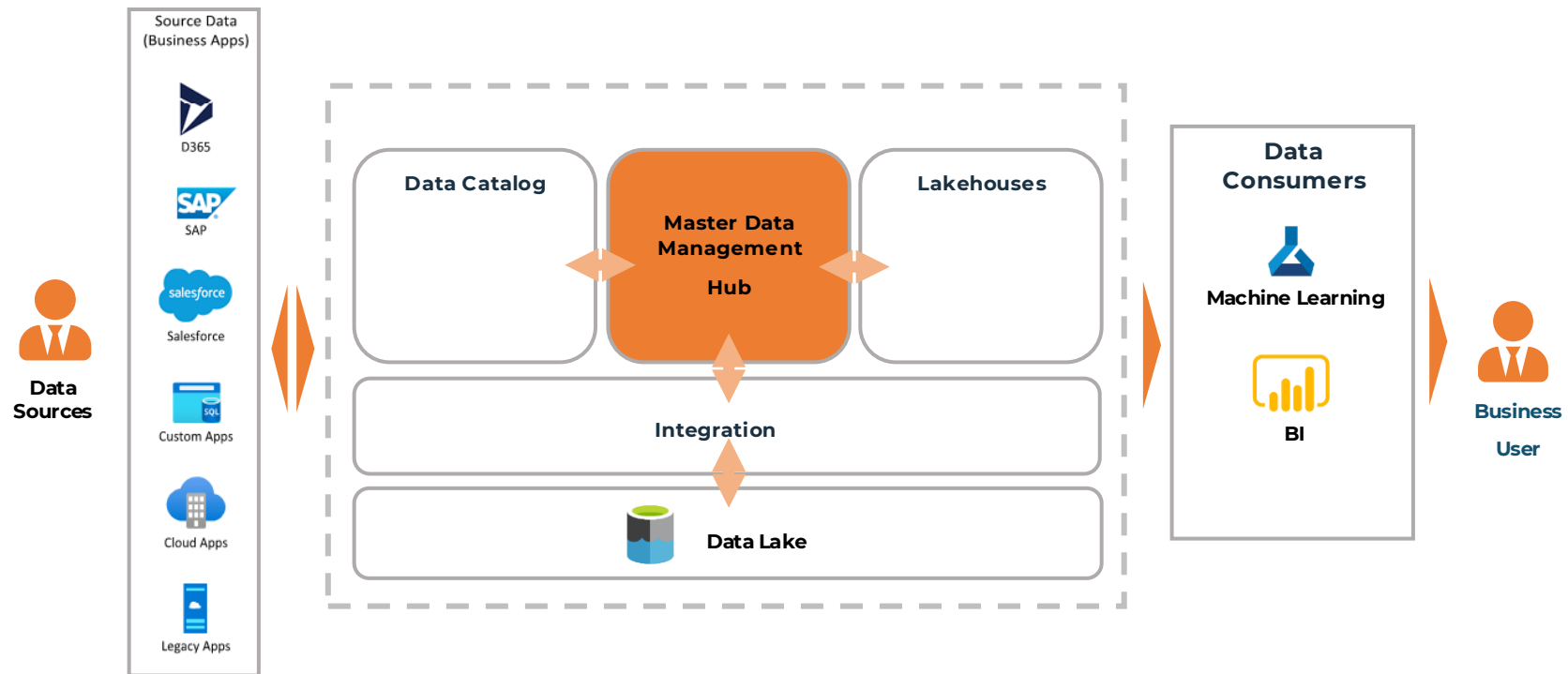


 /in/wmcknight

www.mcknightcg.com
(214) 514-1444



MDM Reference Architecture



Master Data Management Program Justifications (TCO)

- Analysis of the impact on multiple projects
- A single approach to achieving desired outcomes
- Demonstrated proficiency in the use of relevant tools and technologies
- Consolidation of expense streams to maximize efficiency
- Full curation of key enterprise subject areas which allows for maximum data utilization

Sample Applications Improved with MDM

Industry	Subject Areas	Applications	Objectives with MDM	ROI
Retail	Customers, locations, Menu items, ingredients, store locations	Improve customer list management and make real-time AI-based recommendations, Speed up new menu item introduction and ensure consistency across stores	Improve average order size , Foundation for loyalty program, CCPA compliance, enable online ordering, Fraud detection, real-time recommendations	More Sales (ROI)
Healthcare	Patients, providers, locations, supplies, donors, reference data	Enable best clinical practices	Enable data sharing for research and operational efficiency	Reduced Costs (TCO)
Manufacturing	Ingredients and recipes, Financial hierarchies	Manage multiple large rollup hierarchies, Manage ingredients and variations to control the manufacturing process	Understand margin and pricing across regions; enable and expand ecommerce, Improve business processes and service while eliminating clerical errors (order fulfilment, billing, etc.), Enable nimbler FP&A (what-if scenarios, tax optimization, etc.), supply chain management	Build one, use many (TCO)
Financial	Customer, product, channel	Customer management, risk management, audit support, regulatory compliance	Anti money laundering	Reduced AML (TCO)

Agentic AI Needs Master Data

- Garbage In, Action Out
- Write-Path Query Readiness
- Unified Entity Resolution
- Continuous Context Integrity



Master Data: Not an Option

You'll need master data but without a discrete focus on it, you will not do it well

Project Focus

Focus is on an Application's Master Data Needs First

Usually a smaller work effort to get to 2nd, 3rd, etc. applications

Build to Scale!

Enterprise/Program Focus

Focus is on a Subject Area First

Higher Chance of Creating New Organizational Possibilities!

Danger: Build it and They Will Come

Either Initial Focus Needs a Secondary Focus on the Other

It's the MDM Leadership Challenge!

Robust MDM is half of the effort for Success

- Fraud Detection
- Call Center Chatbot
- Self-Driving/Transportation
- Predict Flight Delays
- Marketing - segmentation analysis, campaign effectiveness
- Smart Cities
- Retail, Manufacturing - Supply flow, Customer flow
- Oil and Gas Exploration
- Etc.

Enterprise Subject Areas

- Customer
- Employee
- Partner
- Patient
- Supplier
- Product
- Bill of Materials
- Assets
- Equipment
- Media
- Geography
- Citizen
- Agencies
- Branches
- Facilities
- Franchises
- Stores
- Account
- Certifications
- Contracts
- Financials
- Policies
- Weather

Example: Fraud Detection

Most Fraud Detection is transaction heavy

Sync MDM to Edge

Customer Attributes to include

Last n "transactions"

Avg/hi/low txn profiles

Customer State

Customer Financial profile

Enterprise Data Domains

- Customer
- Employee
- Partner
- Patient
- Supplier
- Product
- Bill of Materials
- Assets
- Equipment
- Media
- Geography
- Citizen
- Agencies
- Branches
- Facilities
- Franchises
- Stores
- Account
- Certifications
- Contracts
- Financials
- Policies
- Weather

Either Way

You'll need master data but without a discrete focus on it, you will not do it well

Do it with data specialists

Data modeling, integration, quality

Use a Tool

It's Operational and Real-Time

Let the Hub create analytical/empowering elements

Make it a discrete project

With high touchpoints with applications

Focus on Total Cost of Ownership first for Justification

Build to Scale

It doesn't take much longer to consider all known requirements

The Real Decision Points

Roadmapping Around:

- Sponsorship
- Subject Areas
- Publishers or Workflow
 - Don't Forget Third-Party Data
- Subscribers
 - Applications
 - Don't Forget "Common" Artifacts like Data Warehouse, Data Lake, and Operational Hubs
- Communications
- Data Governance/Stewardship
- Use of a Tool

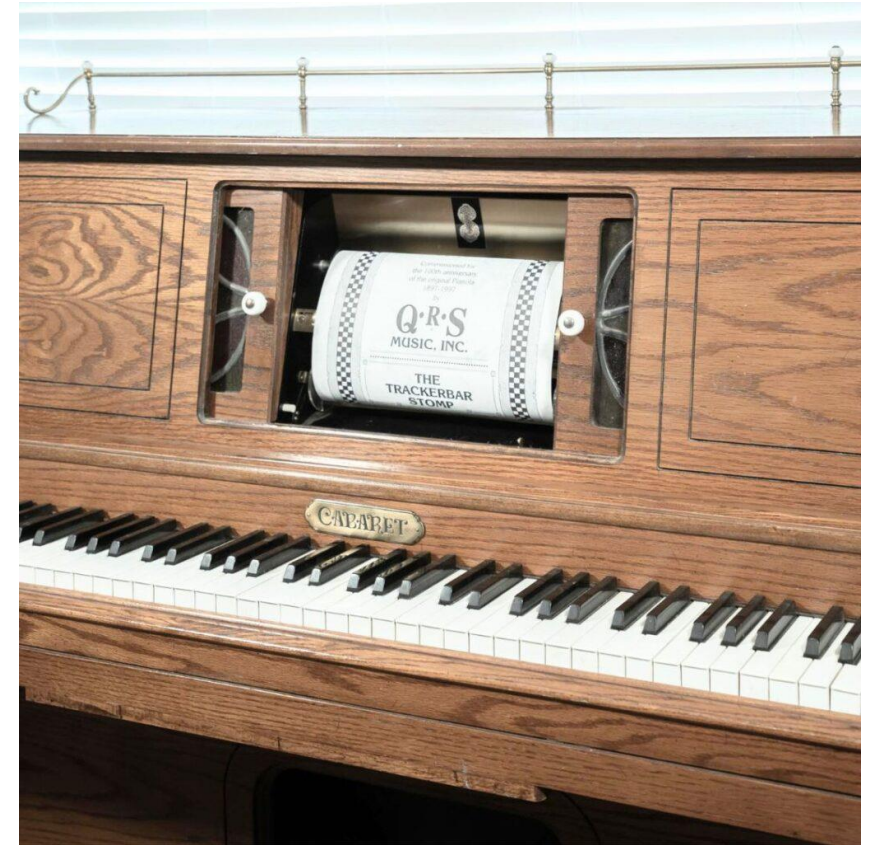


MDM Misconceptions

- MDM is only customer/product data
- MDM is unnecessary with ERP/CRM
- MDM is just one thing (e.g., data integration)
- MDM is only for IT
- MDM is a one-time project, rather than an ongoing enterprise strategy
- MDM is expensive
- MDM projects can be staffed without business involvement.
- Implementing an MDM tool automatically fixes data quality, which requires business input
- MDM is just another data warehouse
- If you have master data, you must use MDM workflow, or conversely, MDM doesn't have workflow
- MDM is not for syndicated data
- MDM is only for operational or analytical data
- All MDM subject areas take as long as the first to implement
- MDM doesn't change job roles
- A "half job" or other systems are as effective as dedicated MDM for core master data
- MDM can succeed without business buy-in and sponsorship
- MDM is optional for becoming a data-driven organization

Automation and AI

- Automated Corrections in Data Verification
- Automation in Data Quality
- AI/ML Enhanced Search
- Data Discovery and Domain Identification
- Entity Resolution
- Lineage, Modeling and Mapping
- API-First Approach
- No-Code/Low-Code Development
- Workflow Automation Features
- Automated Recommendations for Data Stewards
- Completion of Text Fields
- Prompt Engineering



Human Oversight



- AI offers significant automation capabilities, but human involvement remains crucial.
- Outlier detection often requires human review to validate anomalies.
- Active learning enables non-technical users to train and retrain ML models.
- Technology alone is insufficient to replace good data management processes.
- Human oversight and judgment are necessary to ensure accurate and reliable data management.

Test Setup



Platform Summary

- DIY
 - Implementing data management with a UDAP in a Do-it-Yourself (DIY) fashion involves designing data management tools for a UDAP, while using its tools or in-house solutions to handle key MDM capabilities such as entity resolution, survivorship, and governance.
- With an MDM Tool
 - We used an MDM Tool to measure what happens to overall TCO.



D a t a V o l u m e	H i g h	<p style="text-align: center;"><u>Scaled Chaos</u></p> <p style="text-align: center;"><i>Multiple domains and sources but immature or ad-hoc entity resolution</i></p> <ul style="list-style-type: none"> ✓ Entity Resolution (core) ✓ Multiple Sources ✓ Multiple Domains ✗ NO Stewardship UI ✗ NO Workflows ✗ Limited Data Distribution ✗ NO DevOps or CI/CD ✗ Immature and ad-hoc 	<p style="text-align: center;"><u>Enterprise-Scale MDM</u></p> <p style="text-align: center;"><i>Fully mature, multi-domain with workflows, automation, DevOps</i></p> <ul style="list-style-type: none"> ✓ Entity Resolution (core) ✓ Multiple Sources ✓ Multiple Domains ✓ Stewardship UI ✓ Workflows ✓ Data Distribution ✓ DevOps and CI/CD ✓ Mature and robust
	L o w	<p style="text-align: center;"><u>Proof of Concept</u></p> <p style="text-align: center;"><i>Experimentation with entity resolution on a small, simple domain</i></p> <ul style="list-style-type: none"> ✓ Entity Resolution (core) ✗ Single Source ✗ Single Domain ✗ NO Stewardship UI ✗ NO Workflows ✗ Limited Data Distribution ✗ NO DevOps or CI/CD ✗ Immature and ad-hoc 	<p style="text-align: center;"><u>Over-Engineered Data Silo</u> <input type="checkbox"/></p> <p style="text-align: center;"><i>Robust with pipelines, DevOps, & complexity only for one domain</i></p> <ul style="list-style-type: none"> ✓ Entity Resolution (core) ✗ Single Source ✗ Single Domain ✓ Stewardship UI ✓ Workflows ✓ Data Distribution ✓ DevOps and CI/CD ✓ Mature and robust
		Low	High
		Maturity & Complexity	

Sizing Scope

- *Infrastructure, database, software and labor costs*
- We considered a number of variables for project scoping that can be applied to the overall time, effort, and complexity calculations. A model was developed to track these variables as inputs.
 - Number of master data domains (e.g., customer, product, supplier, employee)
 - Complexity and size of each domain (record volume, attribute count, data model complexity)
 - Number of source systems feeding master data (ERP, CRM, HRIS, etc.)
 - Degree of source system heterogeneity/complexity factor (relational, NoSQL, legacy, etc.)
 - Number of data stewardship teams (by domain, region, business unit)
 - Number and type of stewardship workflows (match review, merge approval, exception handling)
 - Perceived quantity data quality issues as a % of overall data (e.g., 10% duplicates, missing values, or inconsistent formats)
 - Number of and complexity of consuming systems (BI tools, operational apps, APIs)
 - Number of environments to support (dev, test, prod)
- Total cost (time, effort, and complexity of building an MDM solution with an MDM Tool and the DIY approach and maintaining them once they are in production, given the input project scoping variables described above)

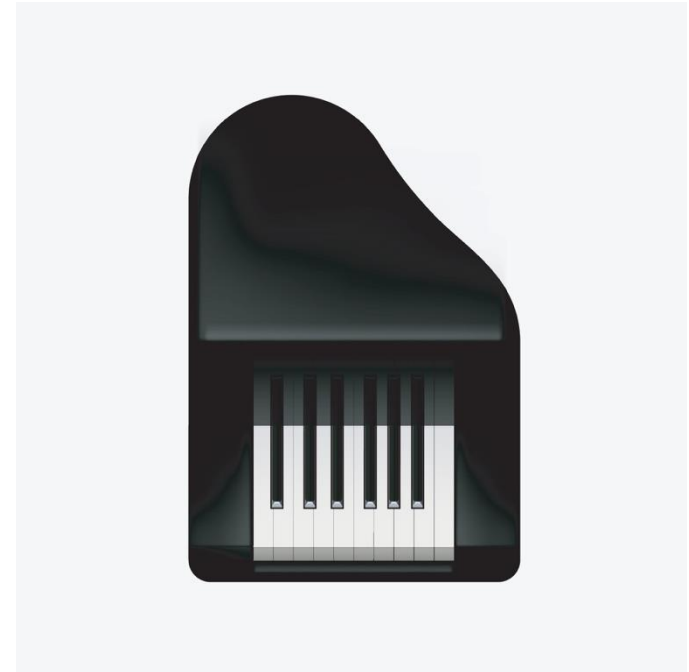
Project Sizing Groups

For our analysis, we created the following project sizing groups of common MDM implementations we have seen over the years:

Sizing Parameter	Small	Medium	Large
Data Domains	3	4	6
Customer Records	100,000	1,000,000	5,000,000
Data Stewards	3	6	12
Source Systems	4	7	10
Source Complexity	6.5	11.5	16.5
Workflows	9	15	21
Workflow Complexity	12.25	18.5	25.25

DIY MDM Short Changes...

- Data Integration
- Data Modeling
- Data Quality
- 3rd Party Data
- Workflow
- Data Governance



The Absence of Data Stewardship Tools is the Ultimate DIY Bottleneck

- The DIY Bottleneck: In-house builds typically skip building UIs for business users, creating a massive operational roadblock where simple data corrections require manual backend developer intervention.
- A tool reduces the development effort of building Data Stewardship UIs by 80%.
- Business Empowerment: Enables non-technical stakeholders to directly resolve data quality issues and enforce rules.
- IT Efficiency: Eliminates constant maintenance toil, freeing up critical engineering resources for higher-value projects.

Factors that push the effort beyond POC

1. Increasing Data Volume and Source Diversity
2. Custom Code Maintenance Becomes Unsustainable
3. Governance and Business Stewardship Requirements Grow
4. Complexity of Hierarchies and Rollups Increases
5. Pressure for Faster Time-to-Value
6. Bidirectional Integration Needs



Test Results





A Tool Costs Less

3.8x less for Small Projects

3.0x less for Medium & Large Projects



Ongoing Staffing Needs

DIY: 5 to 9 team members

A Tool: 1 to 3 team members



Weeks to Production

DIY: 28 to 52 weeks

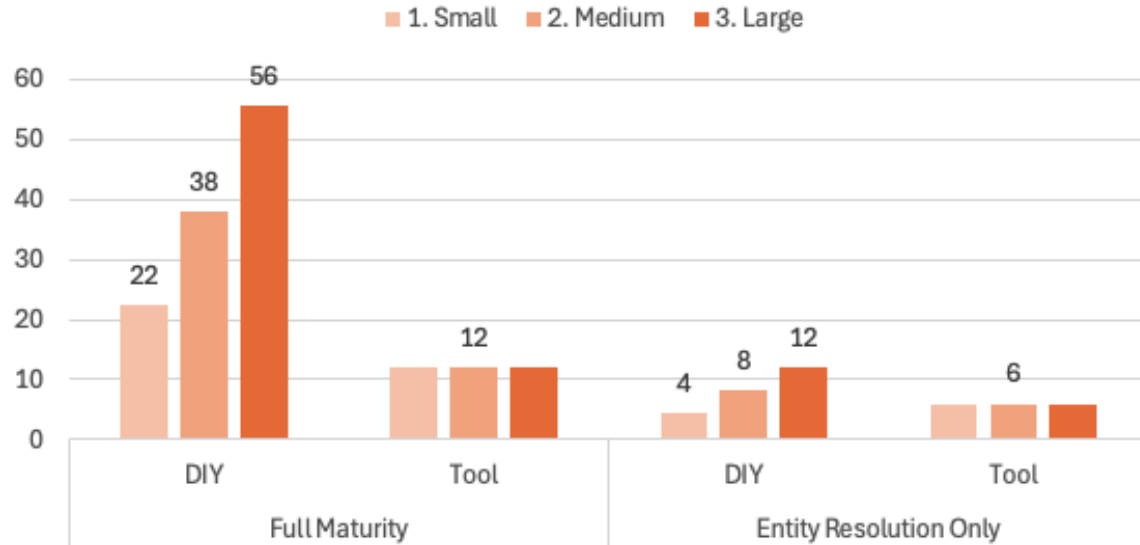
A Tool: 12 weeks+

The Trap of Over-Engineered Silos

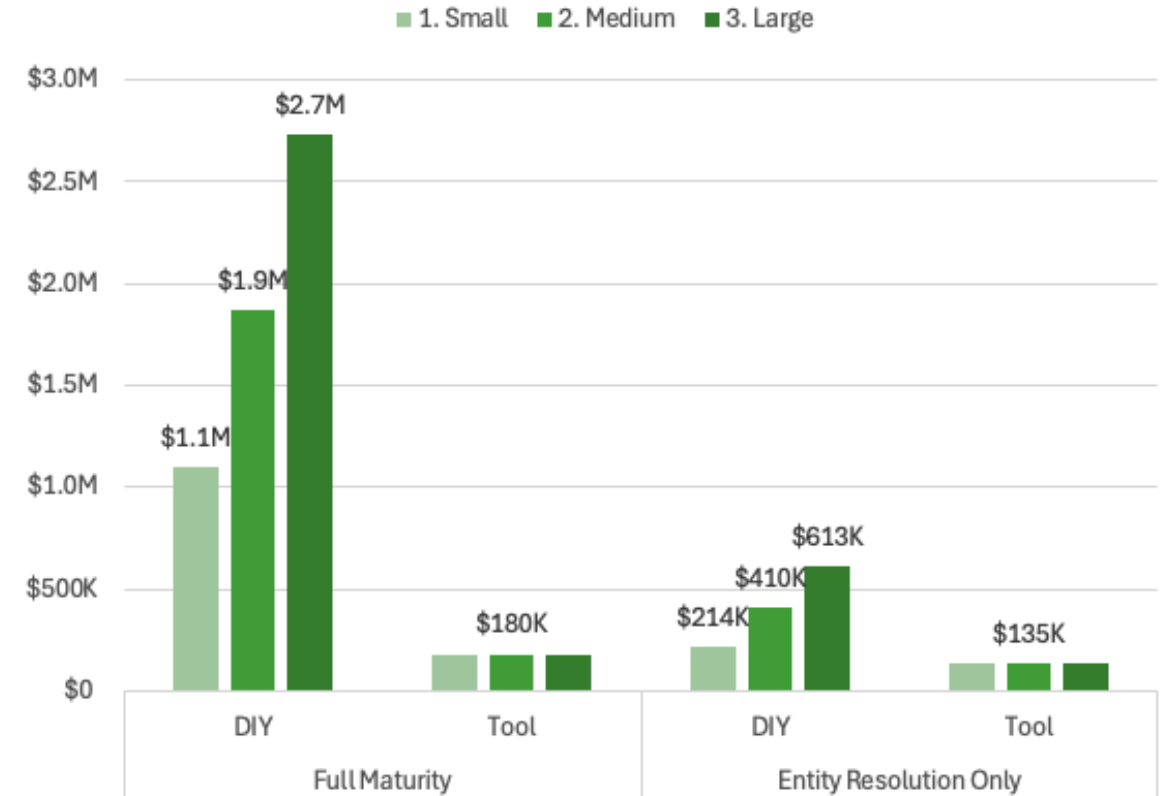
- **Narrow Scope Success:** Teams build high-maturity pipelines and UIs that are successfully functional but restricted to a single, low-volume domain.
- **Disproportional Effort:** Vast amounts of time and budget are consumed without delivering proportional enterprise-wide business value.
- **Funding at Risk:** Low-impact results make it difficult to maintain stakeholder motivation or secure critical future funding.
- **The Strategic Pitfall:** Focusing on engineering excellence in a vacuum results in an expensive silo rather than a scalable data asset.

Getting Up and Running

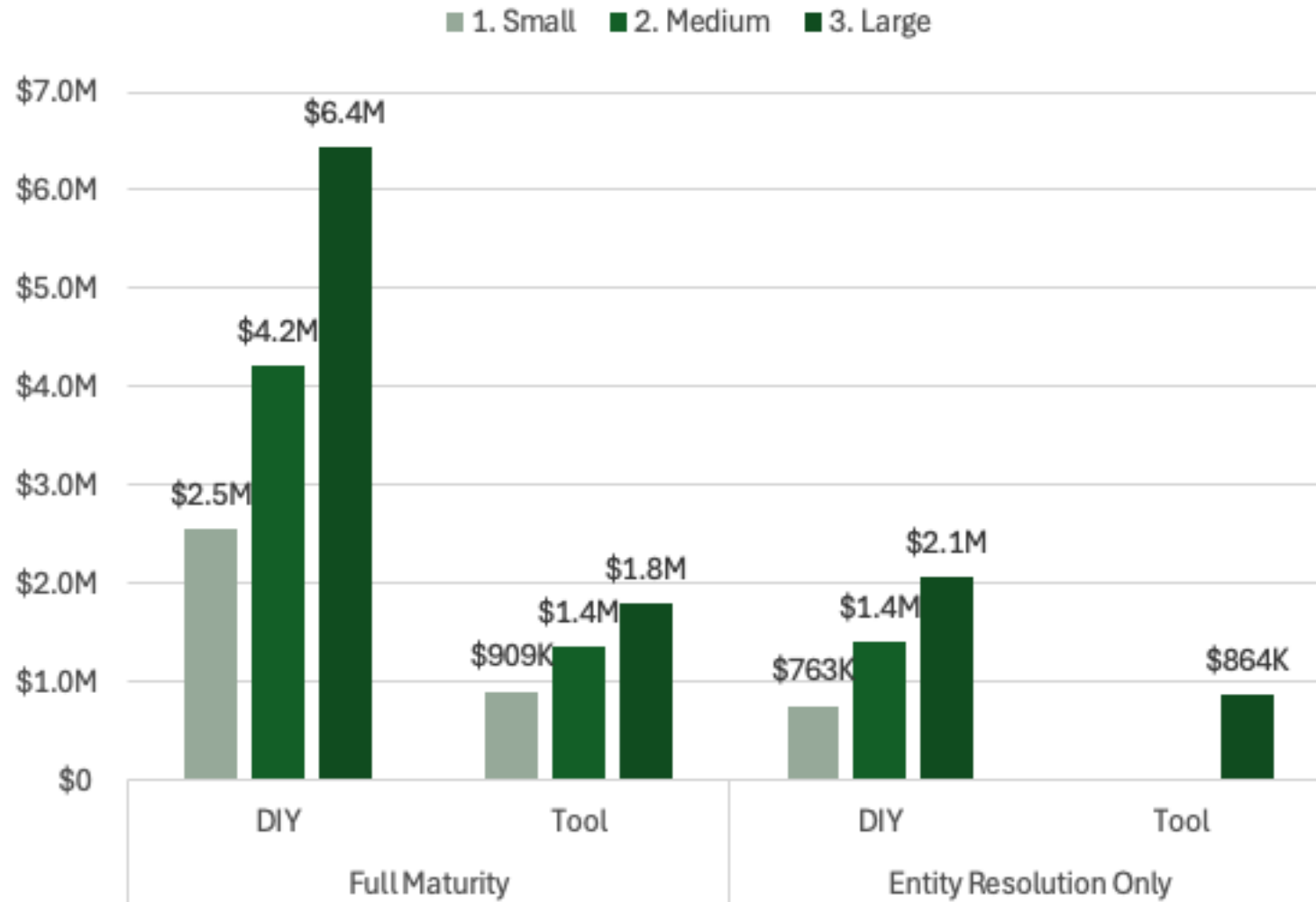
Weeks to Full Production Release



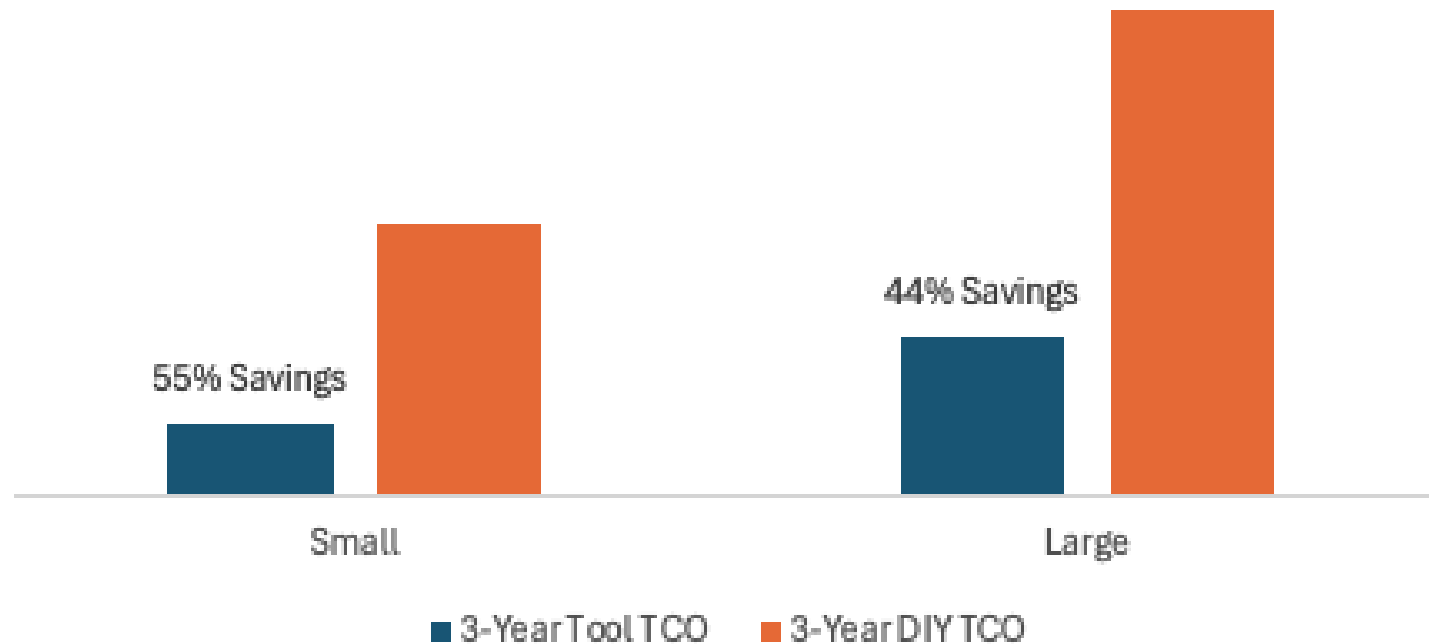
Upfront Development Costs



TCO Over 3 Years



Comparison of Tool and DIY MDM Solutions

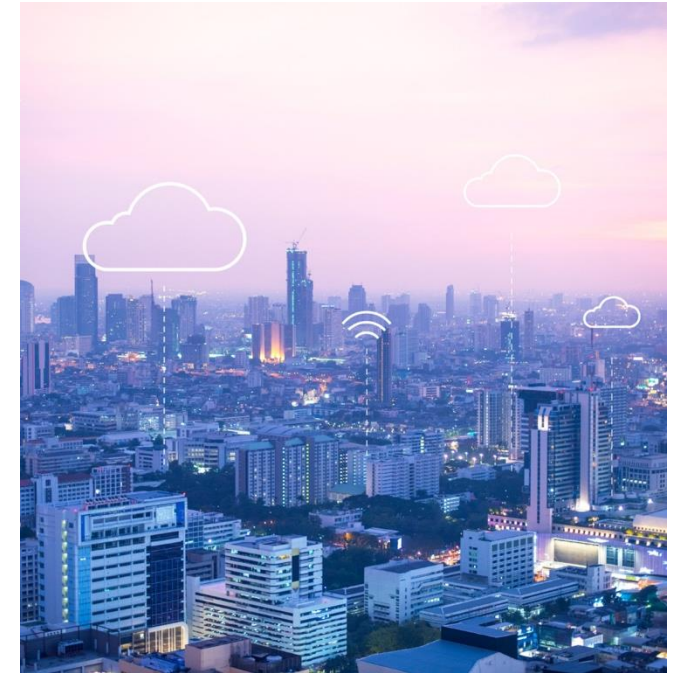


Key Findings

- While **DIY approaches may be sufficient and cost-effective for proof-of-concept projects that won't be scaled up**, an MDM Tool offers substantial cost and time advantages for full-scale deployments and mission-critical operations.
- We identified several indicators that signal the need to move beyond POC-level efforts to a full MDM tool, including **increasing data volume and source diversity, unsustainable custom code maintenance, growing governance and business stewardship requirements, increasing complexity of hierarchies, pressure for faster time-to-value, and bidirectional integration needs**.
- Without disciplined MDM (rigorous data quality, governance, stewardship, and entity resolution), a UDAP promise of an integrated analytics ecosystem can erode due to **duplicated, inconsistent, or poorly controlled master data**.
- An MDM Tool provides **visual rule-building, native governance support, and scalable matching/survivorship logic**, empowering business users and freeing data scientists from low-value preparation tasks.

Summary

- Master Data is not an Option; Robust MDM is half of the effort for project successes
- Automation is abundant now in MDM
- What works for a single domain or POC can't keep up with enterprise needs
- DIY builds often lack integrated stewardship, auditability and workflow; An MDM Tool delivers integrated stewardship, workflows and DevOps maturity out-of-the-box, reducing the operational burden.
- Every enhancement requires more developer time, more integration points and more maintenance
- An MDM Tool reached production in just 12 weeks across all project sizes, compared to 9–52 weeks for DIY builds
- An MDM Tool reduced 3-year total cost of ownership (TCO) by 55% for small projects and 44% for large projects compared to DIY
- For medium projects, an MDM Tool's TCO was more than 50% lower than DIY, and for large projects, DIY costs were 2–3.5x higher
- DIY has a ceiling. It may be cost-effective for limited-scope POCs, but beyond that, it becomes slower and more expensive to maintain.





The Master Data Management Dilemma: To Buy or Build, That is the Question: Benchmark Completed

Presented by: William McKnight

"#2 Global Influencer in Big Data" Thinkers360

President, McKnight Consulting Group

3 X Inc 5000



 /in/wmcknight

www.mcknightcg.com
(214) 514-1444

