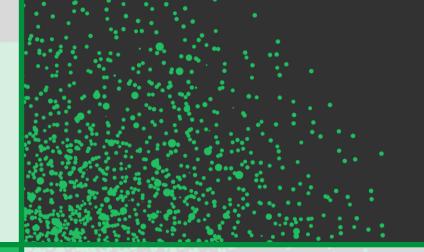


Executive Perspectives





The Future of Data Management with AI

December 2024

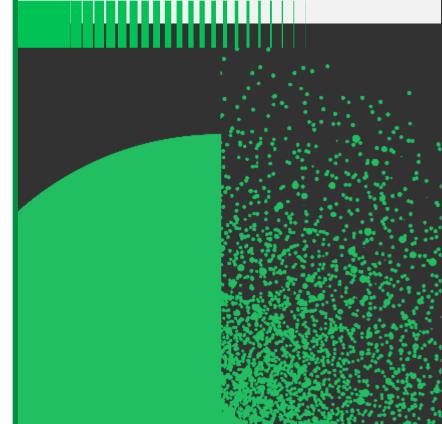
Introduction

We meet often with CEOs to discuss AI—a topic that is both captivating *and* rapidly changing. After working with over 1,000 clients in the past year, we are **sharing our most recent learning in a new series designed to help CEOs navigate AI**. With AI at an inflection point, the focus in 2024 is on turning AI's potential into *real* profit.

In this edition, we discuss the future of data management, and the role AI will play in fundamentally transforming the function. We address key questions on the minds of leaders:

- How do I keep pace with growing data regulations?
- How can I unlock cost advantages while improving my data quality?
- How can I improve my data team's experience and generate more enthusiasm around data management?
- How do I get started...and how do I get this right?

This document is a guide for CEOs and technology leaders to cut through the hype around AI in data management and understand what creates value now and in the future. In this BCG
Executive Perspective,
we articulate the vision
and value of
the future of data
management with AI



Executive summary | GenAI will industrialize the use of data, improving quality, expanding and simplifying access, and increasing productivity

Acting fast is key to tackle rising complexity and costs Data management, a manual and tedious job, is **overwhelmed** with **growing unstructured data** (>10x in 10yrs), **higher quality bars**, and **tighter regulatory oversight**

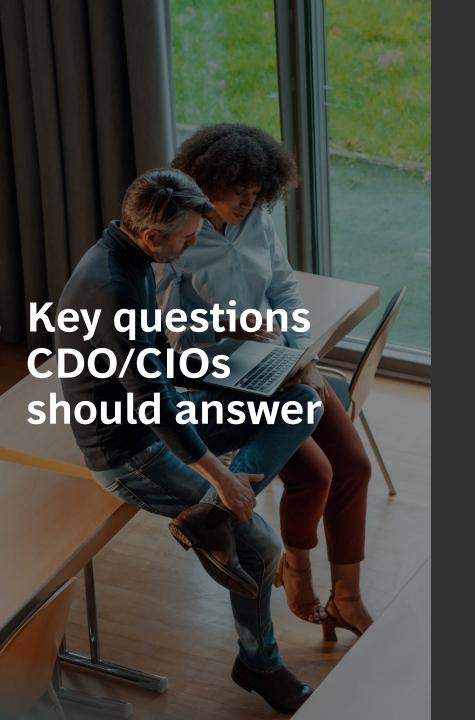
Economics are turning out to be even more challenging, with **data costs projected to grow** 80% from 2023 to 2028, and with **hidden costs** further fueling data cost growth

GenAl can support the needs of your data team... GenAl simplifies and augments data management tasks, accelerating time to value. Coupled with the right tooling, GenAl offers the potential to automate and expedite key data tasks, improving data quality and unlocking efficiencies

...by reshaping your data function for innovation and growth Al vision for data management is to **drive competitive advantage through improved data quality, expanded coverage, self-service analytics, and automated workflows,** transforming roles and democratizing data access with scalable, secure, and compliant solutions

Five key drivers ensure a successful transformation:

- Adopt an AI vision to drive sequential building of your data capability
- Rewire your end-to-end data management workflows to unlock efficiency gains
- Employ a product-centric data operating model
- Ensure a robust data governance across the life cycle
- Invest in **key partnerships** to accelerate capability builds



- Have we **started leveraging AI to reshape** our data function?
- Do we have a **refreshed data strategy,** enabled by GenAI, that is aligned with business outcomes?
- How is our data function lined up to respond to the growing data needs of the business?
- How did our **direct and indirect data costs** change in the past years?
- How are we embedding **GenAI data considerations** in technology and operating model priorities?
- How did we adapt our data-related **people**, **processes**, and organization to expanding needs of AI?
- How do we manage data-related risks and ensure adherence to evolving regulatory requirements?

Today's data management is already burdened by three key areas of friction along the data journey



Accessing data

- Processes today are more difficult to manage due to a significant increase in access groups
- Organizations include multiple levels of approval - delays of up to six weeks
- Changing policies and numerous rules require consistent management



Understanding data

- Data annotation is a labor-intensive process and, even if automated, requires human-in-the-loop
- Data annotation is the primary enabler to the remainder of the data journey
- Data stewards are either not qualified or not accountable



Governing & monitoring data

- Data usage monitoring is not actively performed
- Regulations frequently change, making compliance challenging
- Setting policies requires alignment across stakeholders with conflicting priorities
- Concerns are increasing around Intellectual Property (IP) rights and third-party data



60% of a data scientist's time is spent waiting for data

- BU CDO, Global Energy Company



Due to the manual and time-consuming nature of the job, we're seeing highest churns for data custodians and stewards

- Executive, GSIB1

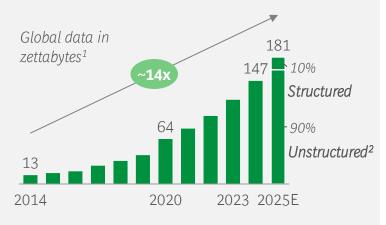


Although General Data Protection Regulation (GDPR) allows customers to request data deletion, for legal purposes we need to keep data for 6 years in case of litigation

- CDO, Global P&C insurer

In addition, exponential growth in the volume of unstructured and multimodal data in the past decade has further raised the bar on data management.

Data is growing exponentially; ~90% unstructured data by 2025









Decisions enabled by AI drive companies to collect more data than ever



With AI, **data inputs becoming**'multimodal', widening tappable landscape

This growing volume with AI will increase demand for data management activities



Data provenance

Where did data come from?

- Identification of source and history of data
- · Authenticity of model data for intended use
- Third-party training data underlying models



Data classification

How is the data labeled?

- Training data accuracy for desired output
- · Quality and consistency of labeled data
- Reduced training, improved model performance



Data lineage

What is the sequence of processing steps?

- Traceability of data transformation
- Reproduction of calculated results
- Interpretation of the data used for model



Data quality, metadata completeness

How accurate is my data?

- Greater quality of model outputs
- Higher model performance without bias
- Management of data drift and concept drift



Regulatory compliance

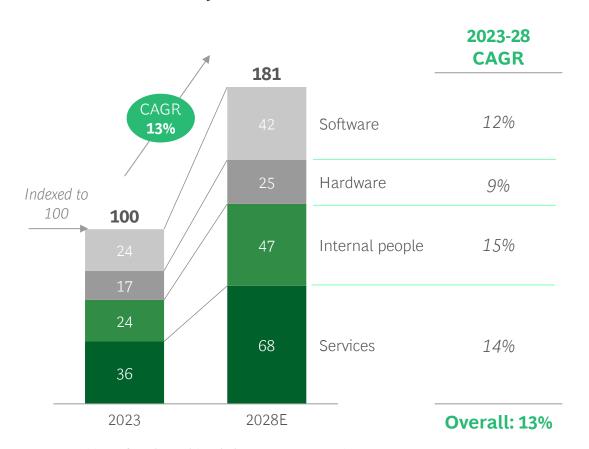
Is this usage ethical & within regulations?

- Multiple data regulations globally (e.g., EU AI Act)
- Higher quality bar imposed by regulators
- Vagueness of regulations on unstructured data

at @ 2024 by Boston Consulting Groun All rights reserved

Higher standards imply increased data management direct and hidden costs, which are forecast to rise in the coming years.

Around 80% growth estimated for data management costs in the next 5 years¹



Also, hidden costs of data management will drive further increases across the board



Manual interventions driving human cost (e.g., BU analyst effort on data modifications)



Delayed analytics, use-case time to value (e.g., opportunity cost of delayed business decisions)



Remediation to regulatory inspection (e.g., issue identifications and corrections)



Fines and security breaches (e.g., due to noncompliance with regulatory guidelines)



Overspending on technology and engineering (e.g., overlapping data management tools)

^{1.} Incurred by IT function, with existing processes & tech
Source: IDC Semiannual Software Tracker; IDC Worldwide ICT Spending Guide; WEF Future of Jobs Report; Economic Intelligence Unit, Gartner Forecast Analysis: Data and Analytics Services;
BCG publication 'A New Architecture to Manage Data Costs and Complexity'

In this challenging environment, GenAI can help simplify data management and accelerate time to value across the data management value chain

GenAl can interpret and create new content, implying potential to augment or automate many key data management tasks





MATCH

data through identifying similar or related data



GENERATE

new data



ENHANCE

data traceability



DENTIFY

types of data and infer metadata



ACCELERATE

compliance and risk management



AUGMENT

data analytics and insight generation

Client example | A build-out of GenAI-led metadata labeling and lineage annotation capabilities enabled significant productivity gains

A global financial institution client aspires to accelerate its data management and governance transformation journey

- 1 Current data governance only covers **small portion of data estate** (only focusing on critical data assets)
- 2 Heavily manual processes are needed to generate business metadata (weeks per data source) with limited alignment between enterprise and BU
- 3 Low efficiency of E2E data lineage and current tooling cannot generate cross-system lineage

We helped the client by focusing five key impact levers...



Enhance data controls at scale

- **Help** accelerate identification of potential gaps/risks (e.g., code/data duplication) through lineage capturing
- Enable data-related roles to **focus on "value-added" tasks** (e.g., review of outputs)
- Streamline continued monitoring and refresh of data estate, with less manual intervention



Improve accuracy & coverage of output

• Improve **accuracy** and ensure **comprehensive** coverage of lineage and business metadata



Drive efficiency to accelerate coverage

- Augment productivity for data-related roles
 (e.g., data steward, central data governance function)
- **Boost productivity by up to 50%** for critical manual tasks



Unlock additional use cases and processes

 Create development and deployment patterns for additional use cases



 Minimize repetitive manual tasks and improve working experience for targeted users

...achieving tangible results

70%+

Human acceptance of "LLM out-of-box" business description

90%+

Accuracy

in PII¹ tagging and data lineage captured posthuman validation

50%+

Productivity boost

to accelerate coverage of data under governance

2-5 Yr.

Reduction

in compliance timeline for high-impact data assets

To control rising data-related challenges, it is imperative for companies to work toward the AI reshape vision

Al vision for data mgmt.

Empower data management to drive competitive advantage through improved data quality, expanded coverage, self-service analytics, and automated workflows, transforming roles and democratizing data access with scalable, secure, and compliant solutions

FROM ...



Data management generally considered an **afterthought in management priorities**



Data management focus **restricted to critical areas** (e.g., data under regulatory purview)



Siloed, centralized approach to data management hindering access to data



Nature of data management work **perceived as mundane** and unappealing



Core data activities (collection, cleaning, etc.)

manual, restricted to data engineers/stewards

TO ...



Empowered data management function transforming data as a driver of competitive advantage for organization



Expanded data management coverage, to significantly enhance data quality and utilization across functions



Departments empowered with self-service analytics and insights, unlocking data potential across organization



Reinvented roles with engaging, joyful, and strategic responsibilities, adding visible value to organization



Core data activities **automated and streamlined** with business owner involvement

Five key drivers can ensure that companies are on a path in line with the vision for data management

×250

Translate your Al-driven data management vision to drive sequential build

Support your AI-enabled vision for data by starting with building a platform and foundational capabilities, ensuring business alignment and enabling more advanced capabilities



Fully rewire data management workflows, minimizing manual iterative loops via automated processes and AI interfaces to accelerate time to value

Employ a product-centric model to data, as part of org-wide platform operating model

Establish efficient and effective ways of working to drive faster throughput via transforming from a reactive, service-based data model to a more proactive 'business partnership' product model



GenAI-led
"RESHAPE"
of data
management



Accelerate Al data governance



Fundamentally transform AI data governance from the sidelines of IT into a core, daily business practice—embedding standards, control, and governance culture across business units





Invest in key partnerships to address capability gaps

Identify and drive partnerships with technology providers to address gaps and accelerate capability builds across the data value chain

Value-centric approach starts and delivers value early from AI, and matures organizational capabilities and data/tech foundations as you go

Plan your journey

Define vision, business opportunity, and key workflows impacted

Experiment with AI

Resolve initial issues and pain points, train core teams, activate key squads

Mature your foundations

Establish teams, **expand data platforms**, achieve economies of scale

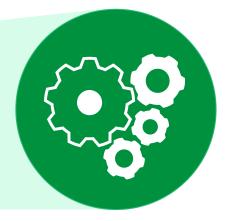
Consolidate and scale

Accelerate delivery of next wave of capability builds









Value creation through successive waves of capabilities

Governance coverage of data domains

Democratizing access and ease of data use across organization

But to unlock value from data as capabilities are deployed, it's key to sequence initiatives by building capabilities in waves



Start with fundamental 'no-regret' initiatives

Metadata and lineage annotation are logical starting points, to accurately describe/catalog data consistently, and are critical for other GenAI capabilities to be effective (e.g., data quality related)



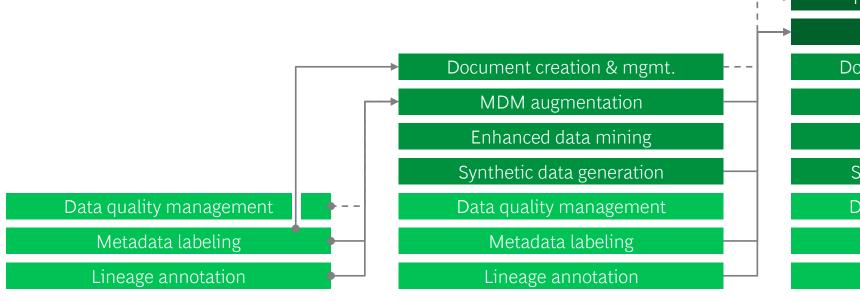
Build advanced capabilities for regulatory compliance

High-quality metadata and lineage enable consistent tagging of data sets, implying better discoverability and improved document cataloging & mgmt. operations for better compliance



Activate 'insight generating' capabilities for business

High-quality metadata, lineage, MDM, and synthetic data will **better support** discovery and utilization, and generate better insights for business (e.g., through augmented analytics)



Policy compliance mgmt.

Augmented analytics

Document creation & mgmt.

MDM augmentation

Enhanced data mining

Synthetic data generation

Data quality management

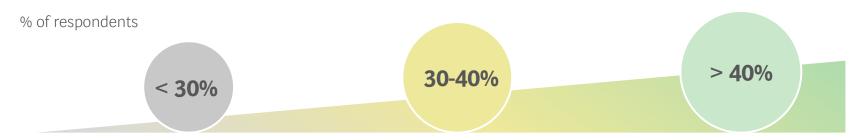
Metadata labeling

Lineage annotation

12

Al is anticipated to have the largest impact on data analytics and data management workflows, necessitating focus across people and processes

In which of the following IT processes and workflows do you anticipate AI technologies having a transformative impact? [Multi-select]





Application performance monitoring



Project and portfolio management



IT infrastructure (data center, networks) operations



IT service management

Software development life



Data analytics



Example to follow

70%

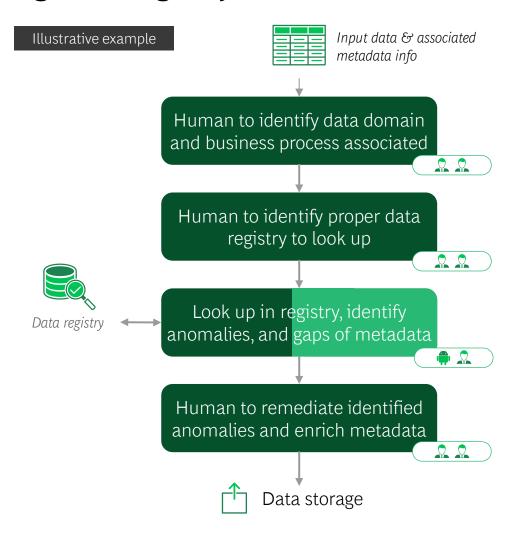
of AI transformation effort should be invested in people and processes

Drive change management and other processes related to people

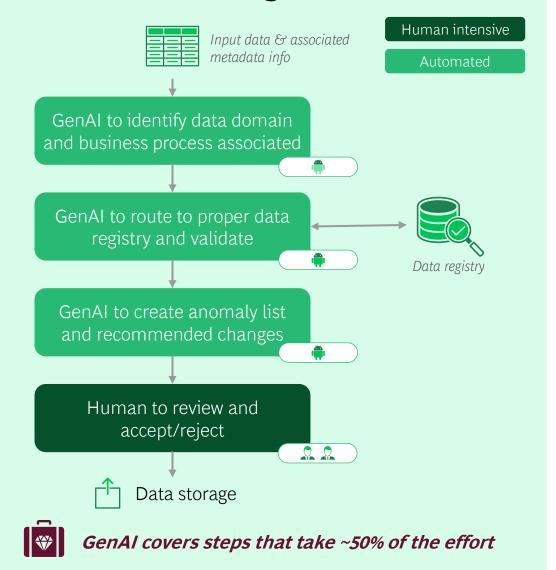
IT asset management and maintenance

Source: BCG Build for the Future 2024 survey (n=1,000 respondents)

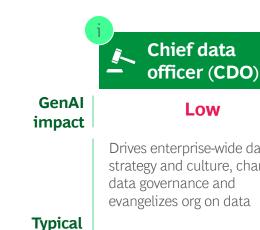
TODAY: Enrich metadata manually against registry



TOMORROW: ~50% effort reduction in metadata labeling with GenAI



In addition to process, key data management roles are being transformed to be more interesting, interactive, and productive with GenAl



Data governance office



Data domain owner (DDO)



Data steward



Data custodian

Drives enterprise-wide data strategy and culture, champions data governance and evangelizes org on data

Coordinates prioritization of actions / data quality remediation plans

Medium

Operationally supports rollout of enterprise-level data governance and data culture

Prepares requirements to rollout governance of data domains, including taxonomy, identifying roles in organization

Trains employees on data management roles

Medium

Is responsible for a specific data domain (global/ regional)

Refines data domain taxonomy and glossary

Produces data domain heatmap, defines data quality targets and measures of data quality

Very High

Operationally supports DDOs with set of data families

Proposes remediation actions and roadmap

Implements data governance policies and processes

Aligns with data custodian on IT needs

High

Determines strategic direction for data platforms

Ensures implementation of data strategy on data platforms and other relevant IT systems

Manages IT architecture for data platforms





response

Augments and facilitates improvements to data supply chain (e.g., data quality assessment and remediation planning)

Allows employees to explore policies and data management, through chatbots, improving efficiency and reducing ad hoc support and training

Augments data domain creation and maintenance of taxonomy, ontology, and glossary; automates reporting (e.g., heatmap and data quality)



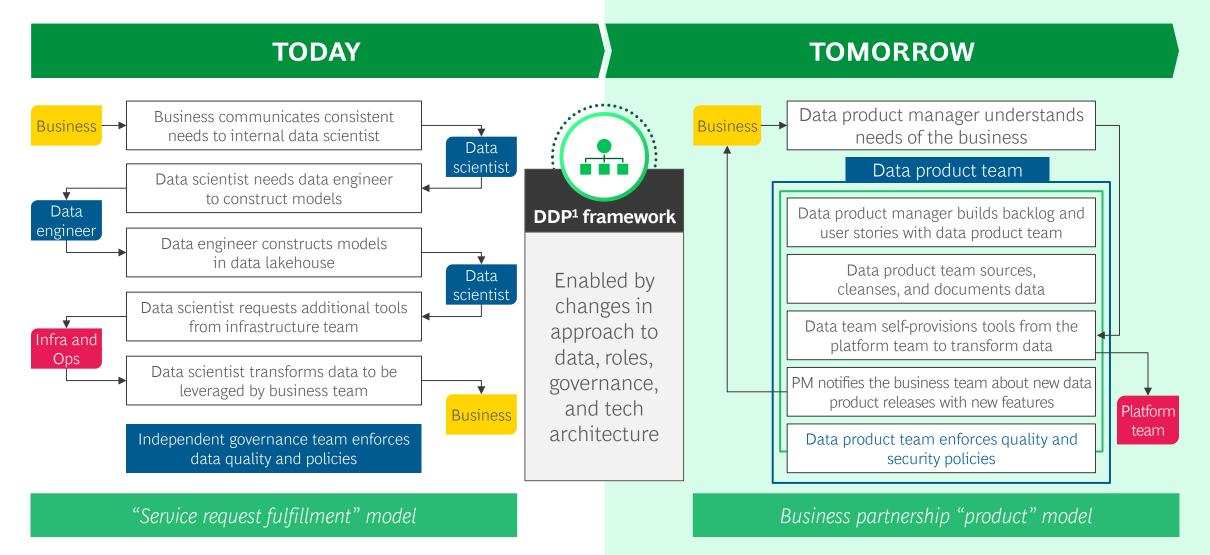
Augments data glossary, dictionary, ontology generation and coherence; automates data map creation and data glossary updates



Automates tagging and labeling, monitors compliance of data strategy on data platform and other systems

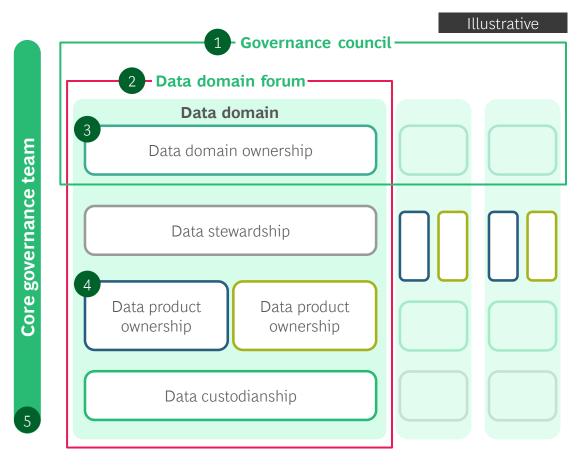
As technology enables bionic features, the burden of these roles will be reduced, enabling them to scale to higher-value tasks

A move from today's siloed structure toward a product-centric approach will improve data accessibility and consumption



Establish a centralized system to fundamentally transform Al data governance through integrated standards and control across business units

Federated data governance organization to ensure standards, control, and governance culture across BUs



- **Governance council,** formed of data domain owners, steers strategic priorities and establishes global policies (including responsible use of data), embedding standards across business units
- **Data domain forum** sets the overall strategy and priorities for the data domain to ensure that the domain's data product portfolio meets consumer requirements
- Data domain ownership ensures that global priorities are enforced in the data domain and data quality issues are escalated to the council, instilling data governance culture in BUs
- **Data product ownership** manages the data product's E2E life cycle and ensures data availability, data usage, and consumer satisfaction
- S Core governance team drafts policies for the governance council and supports data domains in the implementation of data governance capabilities

Partnerships with technology providers are key to address gaps and accelerate capability development across the data value chain

Partners/Tools¹ Sample processes # Collibra **Data quality automation** ■ lightup ataccama Al tools identify and correct errors, inconsistencies, and inaccuracies in data, as well as enrich data precisely **DQ**LABS OpenRefine with additional information **CLEAN** TOMIC **%** baffle **PR**TEGRITY **Data anonymization** Anonymize data to protect privacy while preserving utility and integrity **PRIVATE**AL brīghter AI **<2**VIeW gretel **MDCLONE** Synthesis.ai Synthetic data generation Generate synthetic data that resembles real data to protect privacy and facilitate safe data sharing MOSTLY AI TOMIC **Code generation M**OpenAl Sourcegraph **GENERATE** tobnine **GitHub** Copilot Automatic functional code generation, optimization, and rewriting using trained LLMs results in codium° rapid prototyping and speeds up development cycles Labelbox SambaNova[®] scale **Data labeling** Snorkel Al automates process of labeling data/suggesting answers for human review and verification appen □ Superb Al SambaNova[®] Forcepoint **IDENTIFY Data classification** Automate the process of classifying data into different categories such as sensitivity, type Labelbox gretel Nightfall Al **MDM** automation **PiLog** Informatica Ol ataccama Auto MDM (master data management) powered by AI intelligently identifies and reconciles data, **TSTIBO** SYSTEMS reducing manual effort and improving data quality **Data lineage Qlik Talend MATCH** Trace the path and transformations of data from its source to destination, ensuring transparency and accountability in data handling, which helps in verifying data quality and compliance As of Q3 2024

Copyright © 2024 by Boston Consulting Group. All rights reserved

povight © 2024 by Boston Consulting Group. All rights reserved.

How to get started | Four key activities can activate the building of strong foundations and accelerate time to value

Develop core capabilities & strategies to effect a data-driven transformation

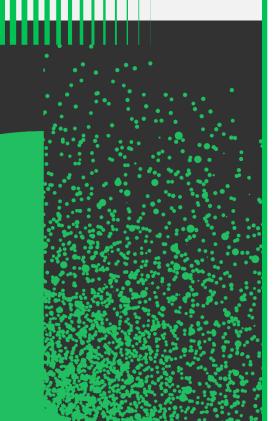
Extend capabilities across data management opportunities

Fully integrate AI into data management at org-wide level

- Develop your data strategy: Adopt a value-centric strategy to prioritize opportunities and align business outcomes, data platforms, and assets to unlock outcomes over time
- Assess existing setup: Review how your data function is positioned along the five drivers, understanding impact on workflows and key AI opportunities that map to it
- Focus on fundamental capabilities: Introduce and pilot AI with lineage annotation, metadata labeling and data quality management, leveraging third-party vendors as needed
- Fill talent gaps: Upskill existing workforce and recruit new data-related talent that is needed to support transformation, while protecting upskilling time required

Supported by embedding an AI culture organization-wide, driven by leadership, to empower teams in leveraging AI for sustainable and scalable growth, and to boost adoption

BCG experts Key contacts for Al data management transformation



NAMR



Dylan Bolden



Vladimir Lukic



David Martin



Amanda Luther



Matthew Kropp



Sesh Iyer



Julie Bedard



Beth Viner



Djon Kleine



Steve Mills



Benjamin Rehberg



Renee Laverdiere



Helen Han



Daniel Martines



Bo Xu



Vikram Sivakumar



Tauseef Charanya

EMESA



Nicolas de Bellefonds



Jessica Apotheker



Marc Schuuring



APAC

Jeff Walters



Romain de Laubier



Dan Sack



Andrej Levin



Marcus Wittig



Julian King

Akira



Aparna Kapoor





Nipun Kalra

