White paper

Microsoft Azure

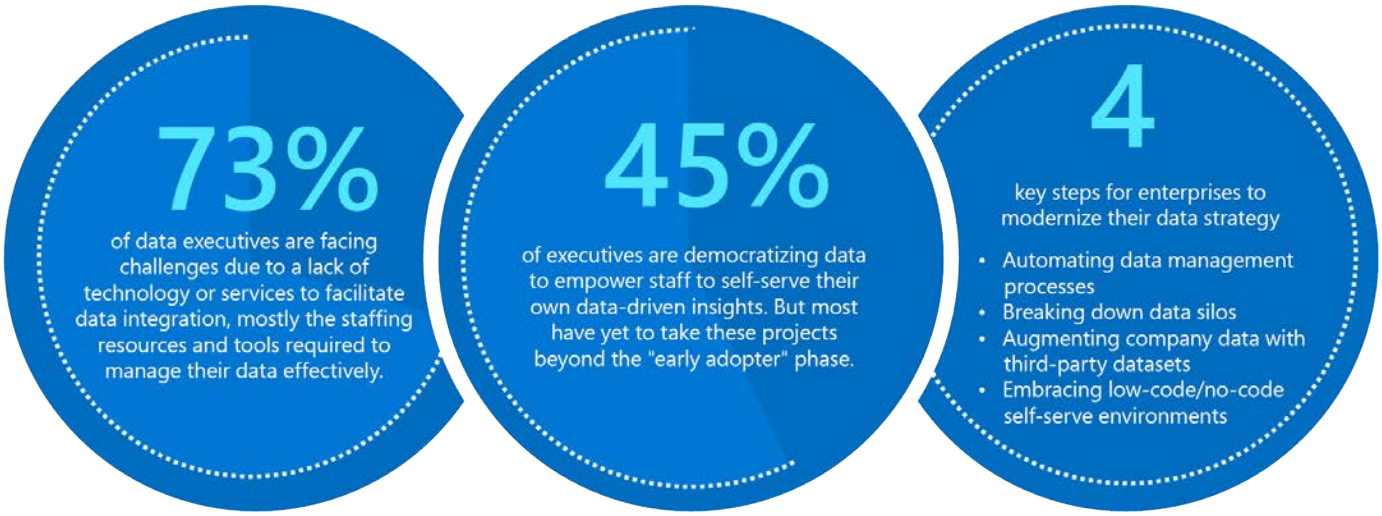
Microsoft Azure: The CDO Seat at the Cloud Table

Microsoft Azure: The CDO Seat at the Cloud Table

Overview

The global data landscape is now more complex than ever. Chief Data Officers (CDOs) across the Americas, EMEA, and Asia Pacific have experienced extraordinary opportunities to manage enterprise data assets to fuel data-driven decision-making. However, many grapple with issues created by digital disruption, cybersecurity threats, privacy and compliance, and stakeholder demand for higher levels of impact and transparency—all of which can place undue strain on IT and drain critical resources.

While most CDOs use both internal and external data for data-driven decision-making and automation, they also report significant struggles in building a data culture, where employees follow their own intuition rather than data-driven insights. According to a survey of 300+ CDOs conducted as part of Data Integrity Trends: Chief Data Officer Perspectives in 2021, there are several major trends in the digital empowerment of employees:

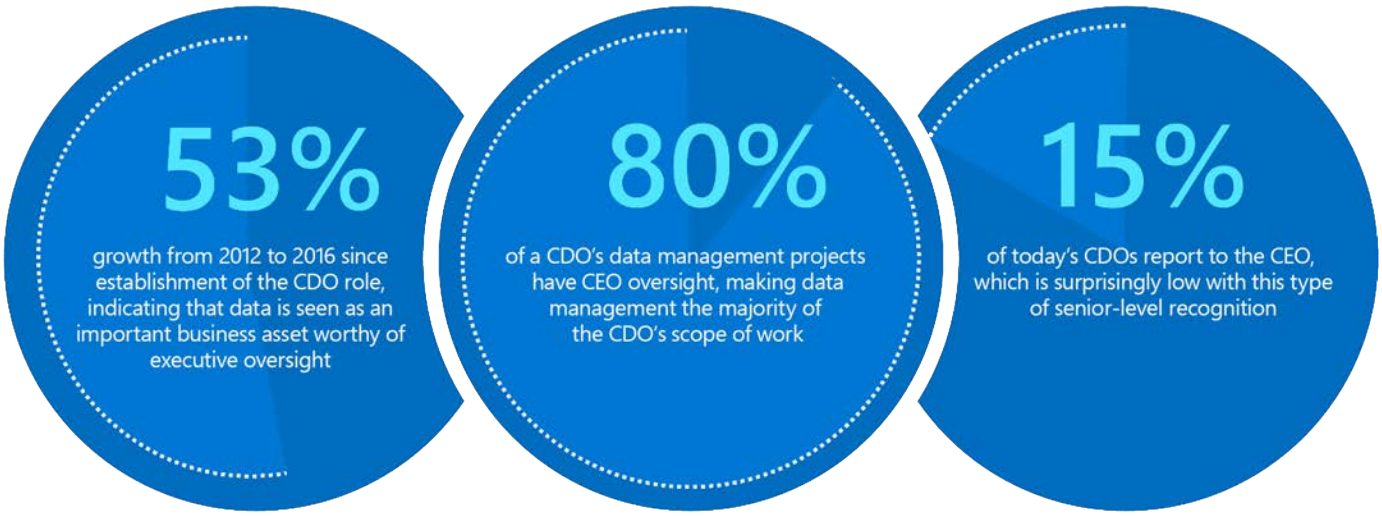


As a CDO, you're expected to defend your firm's data estate and to capitalize on it for new business opportunities and increased customer satisfaction while navigating the impact the cloud has on your organization and personal charter. Extracting trusted insights from your data estate is becoming an essential measurement of success in your cloud adoption. At the same time, you need to stay in control, comply with external regulations such as GDPR, and secure your most valuable assets. Return on digital investments was listed as #1 in globally renowned consulting firm McKinsey's top five measures of success in digital for CEOs in 2021. McKinsey emphasized the importance of focusing on how the firm uses its investment in data preparation to achieve its most strategic organizational goals.

Data management is not solely a technology problem to fix: it is how your organization prepares for the workforce of the future by building a data culture, redefining processes, and putting the right tooling in place to set the data foundation. The role of the CDO in your company's cloud strategy will be a key differentiator between your company and ones that underestimate the hurdles of building a data-driven organization to achieve strategic objectives.

The hurdles CDOs face in establishing their organizational footing are not dissimilar to the early days of the Chief Information Officer (CIO). During the 1980s mainframe era, applications fell under the remit of the CFO since organizations categorized them as an accounting-related responsibility. Information managers (precursors to the CIO) were newly integrated into the enterprise and were largely misunderstood and undervalued and lacked management direction around the expectations of duties. Turnover was high in the early stages of the CIO role's introduction, with 50% of information managers being replaced within the first 18 months. These roles were hired for their technical skills to handle the rise of computing and later the World Wide Web. The emergence of these technologies was responsible for today's incarnation of the CIO, as the CEO recognized the need for help in shaping new business models and spearheading the future direction of the organization.

Today, the CIO is seen as an essential role in all major organizations, but many CDOs still struggle to get support from C-level executives due to the hurdles of defining data as a strategic business asset. However, much like the rise of computers, the rise of the cloud will catapult the CDO into an essential role at the C-level table and the return on investment of proper data management will translate into a mission-critical function in the company's overall business strategy:



Traditionally, the CDO role has taken a defensive position to handle the discovery and management of data for regulatory and privacy purposes, where value has been defined by the offset of risk, fines, and reputational damage. In more recent years, and with the addition of analytics to many CDO responsibilities, a broader offensive positioning has been recognized, where value is defined by data commercialization and enhanced analytics and insight for business management. We've even seen, in a recent industry benchmark conducted by the EDM Council, how responsibility for the ethical use of data has entered the remit of the CDO, especially with the proliferation of machine learning and AI.

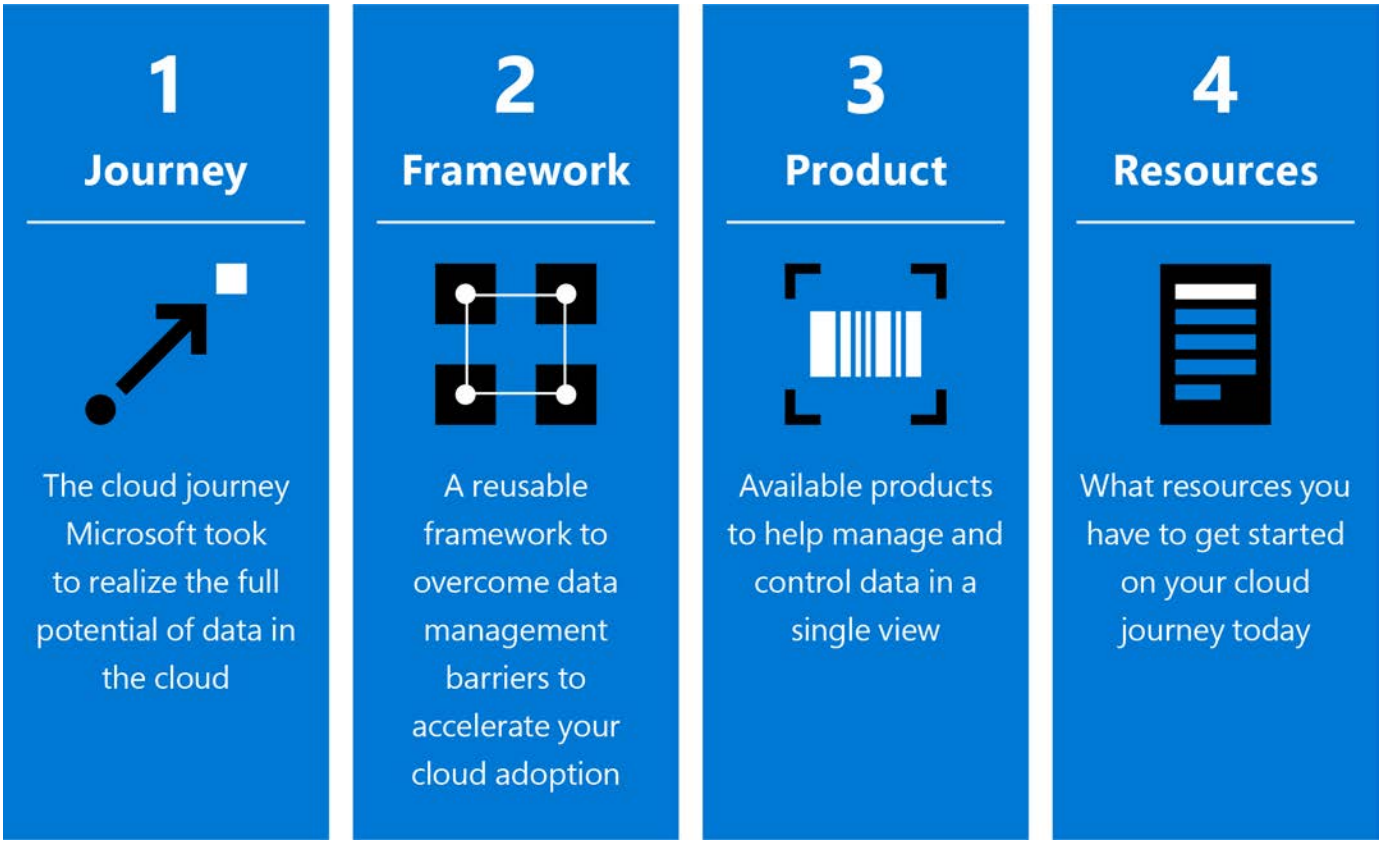
The Enterprise Data Management (EDM) Council is a global non-profit trade association dedicated to advancing data and analytics and are the authors of the DCAM (Data Management Capabilities Model) and CDMC (Cloud Data Management Capabilities) frameworks.

EDMCouncil

The scope of responsibility of the CDO will continue to be restructured while maintaining the same mantra—make data discoverable, usable, and accessible to the right people in a secure manner, while continuing to further enhance the company's data risk posture. We foresee in the coming years that CEOs will shine a spotlight on a new mission-critical seat to take on this next phase of cloud adoption—the seat of the CDO.

So, what does that mean exactly? The acceleration and optimization of cloud adoption in this next phase will largely require your organization to define a data strategy or intertwine its established data strategy with the right people, processes, and technology to build a mature and modernized data estate in the cloud.

In this white paper, we will go through:



Journey

The Microsoft data transformation journey to Azure

Instead of just telling you that the CDO is essential in delivering value on top of your organization's cloud migration, we would like to share with you the impact it had on Microsoft.

As a steward of Microsoft and our customers' data, the Enterprise Data team at Microsoft Digital searched the industry for technologies to solve similar problems to drive data democratization and power data-driven decisions. Microsoft has changed significantly over the years, with company acquisitions large and small, which has led to the propagation of data siloes across our lines of business. As a result, we have been able to traverse our data estate seamlessly, proving to us the importance of creating a common taxonomy across the organization with a renewed focus on data. Speaking with CDOs at other large organizations revealed that their data management challenges were similar to those we faced, despite nuances across industries.

We then investigated the market to find tooling to meet Microsoft's data management capabilities. With the scale of business and operations at Microsoft, data quality, privacy, integrity, and security requirements were absolute—and it soon became necessary to build our own solutions powered by Microsoft Azure. We identified some data issues with substantial impacts on business and operations that could be solved with a more modern approach to data management.

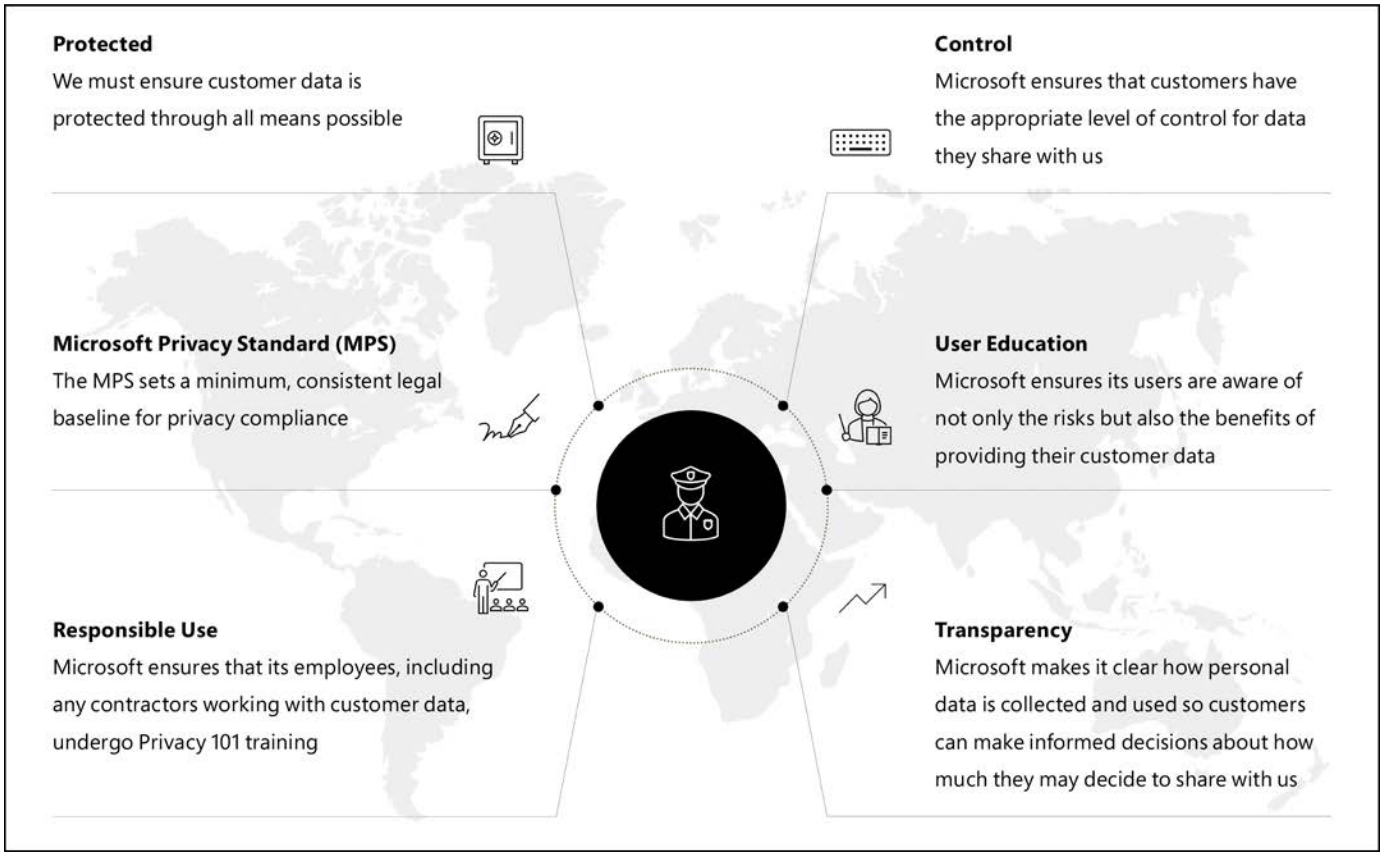
Data issues

- **Fragmented data estate** with siloed data driving up capital expenses and infrastructure redundancies in addition to the complexity of the hybrid cloud model
- **Redundancies and inconsistencies** in data management, multiple data masters and copies causing Shadow IT risks and inefficiencies
- **Poor data quality** causing delays in insights generation, proactive fraud detection, partner incentives, and global trade optimizations
- **Difficulty tracking end-to-end data lineage** causing supply chain and workflow issues
- **Manually managed data policies** or data standards causing compliance and data exposure risks

Among the data issues faced, we appreciated the importance of privacy regulations like GDPR. Recent financial penalties levied on other Fortune 100 companies for lack of compliance were a constant reminder of the importance of protecting personal data. We quickly realized that losing sight of this could not only have financial implications but also dent Microsoft's global reputation and trust.

Thus, maintaining compliant data foundations became the cornerstone for enabling responsible data democratization. We built custom data solutions for GDPR and Privacy Command Feed (PCF) based at scale over our data foundations. Having a common foundation made it easier to stay compliant with the ever-changing regulations in this area, while simultaneously enabling our partner teams with access to the right standards and programs to remain rigorous when building federated applications.

The following visual depicts the pillars we defined for our common foundation:

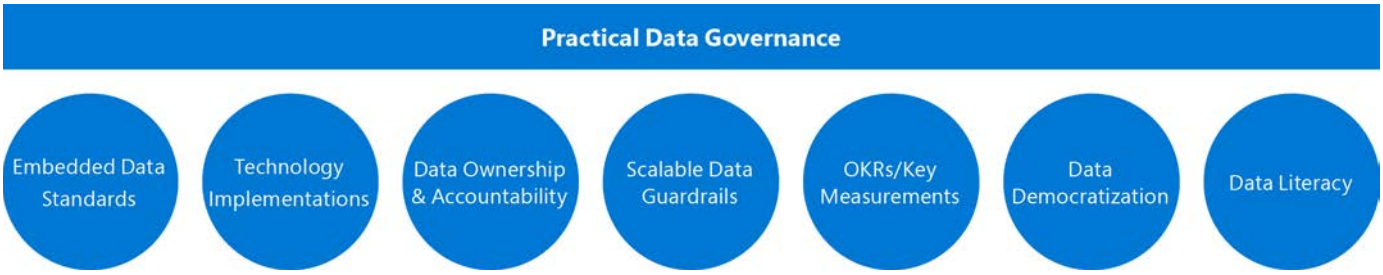


Creating a culture of data governance

The next important step in Microsoft's modernization journey was the practical, evidence-based culture of data governance. The evidence-based approach helped identify priorities for data governance at Microsoft, which were then mapped to data governance capabilities embedded into technology and processes. A culture driven from the top-down of evangelism, training, and data steward community outreach resulted in:

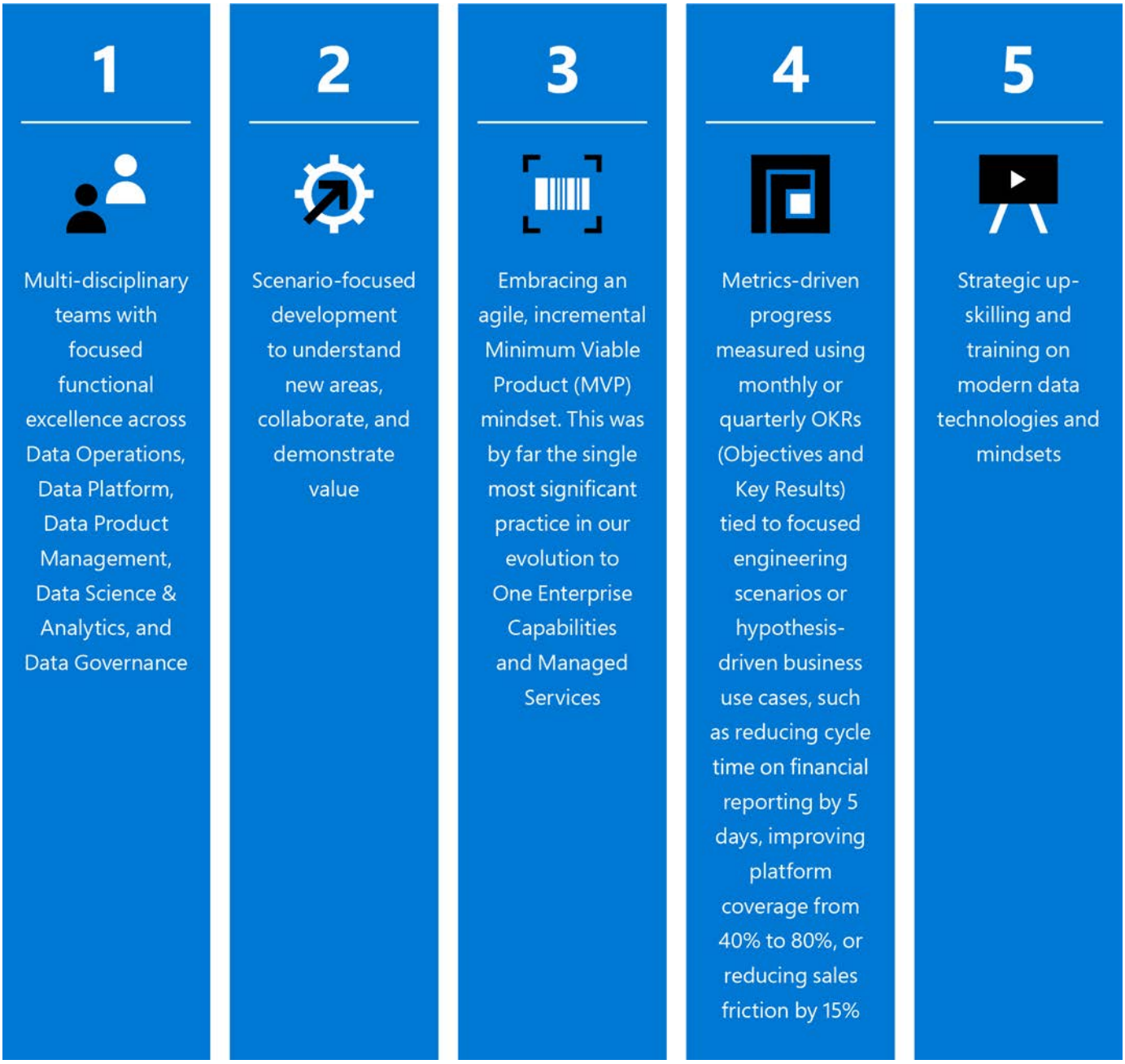
- Formalized data standards integrated into modern engineering processes.
- Systematic data management controls, scanning, measurement, and access/authorization.
- A single Enterprise Data Estate (EDE) engine for a company-wide inventory of data assets, metadata, catalog, and data custodianship.
- Scalable, evidence-based data guardrails to detect and prevent data problems and fixes.

The following figure shows the scorecards we used to drive visibility of the impact and adoption of data management and data governance processes and technologies:



Forming a cross-disciplinary team of data professionals diverse in age, gender, skills, geographical location, and area of professional training was foundational. To really excel, we needed deep cross-functional integration across multiple disciplines: platform engineers, program managers, data scientists, and governance experts all had significant roles to play in delivering and operating impactful data solutions.

Data democratization is the way forward for making use of the true value of data. People are at the heart of this transformation. In the end, five practices were fundamental to our evolution:

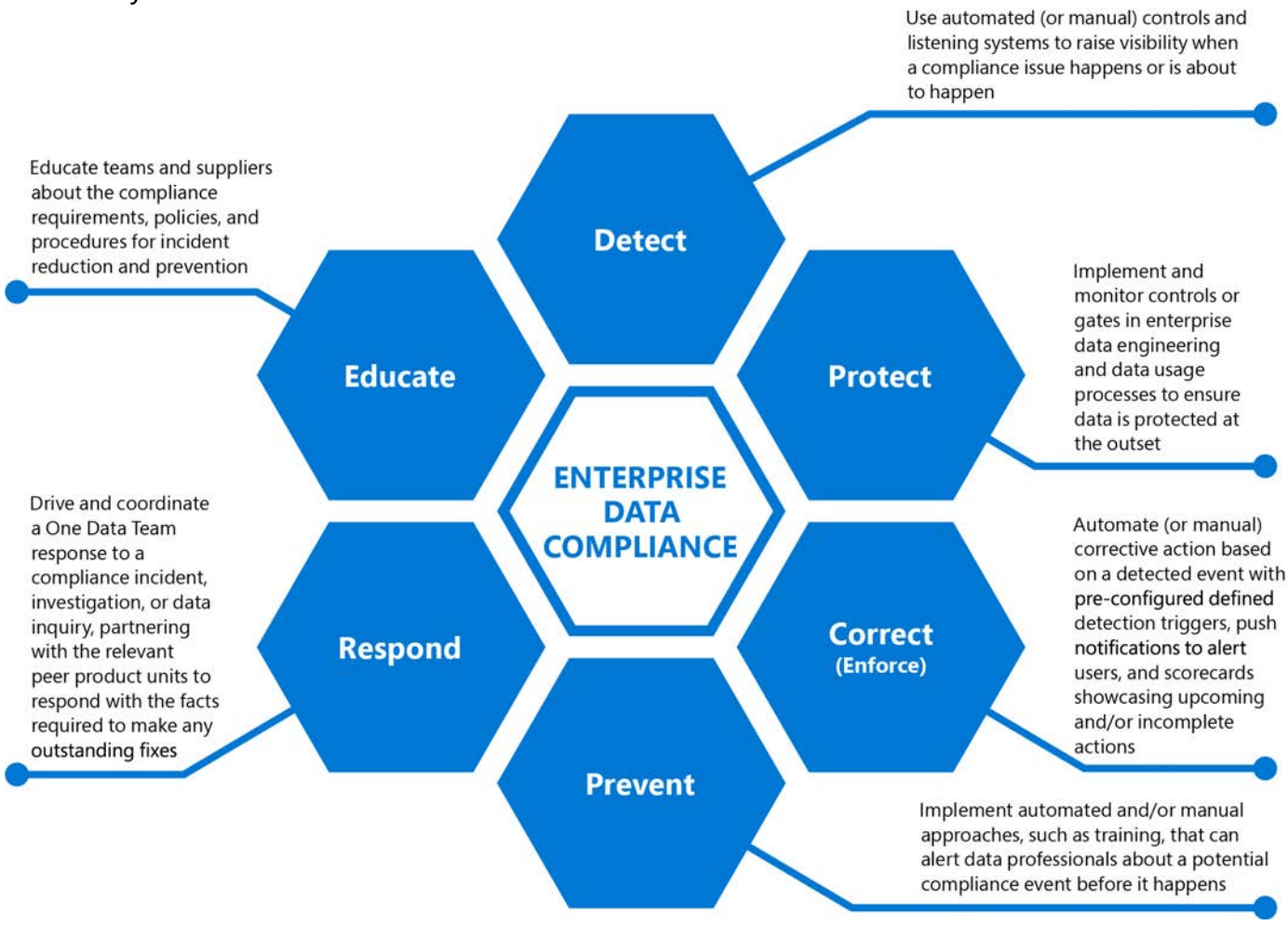


Standardizing data management with corporate standards and compliance programs

To standardize data management across Microsoft, we created a holistic compliance program that was intentionally more *policy consumption*-oriented than *policy creation*-oriented. In other words, the strategy is about the robust implementation of existing policies, standards, and controls defined by peer compliance organizations (for example, Privacy, Security, and Accessibility).

Our strategy for this compliance program consists of a six-point plan that progressively moves our organization to embody the most advanced compliance partner possible. The core compliance hallmarks of detection *controls*, preventative *measures*, compliance *education*, and incident *response* are still captured and underpin the strategy.

The six key elements are defined as follows:

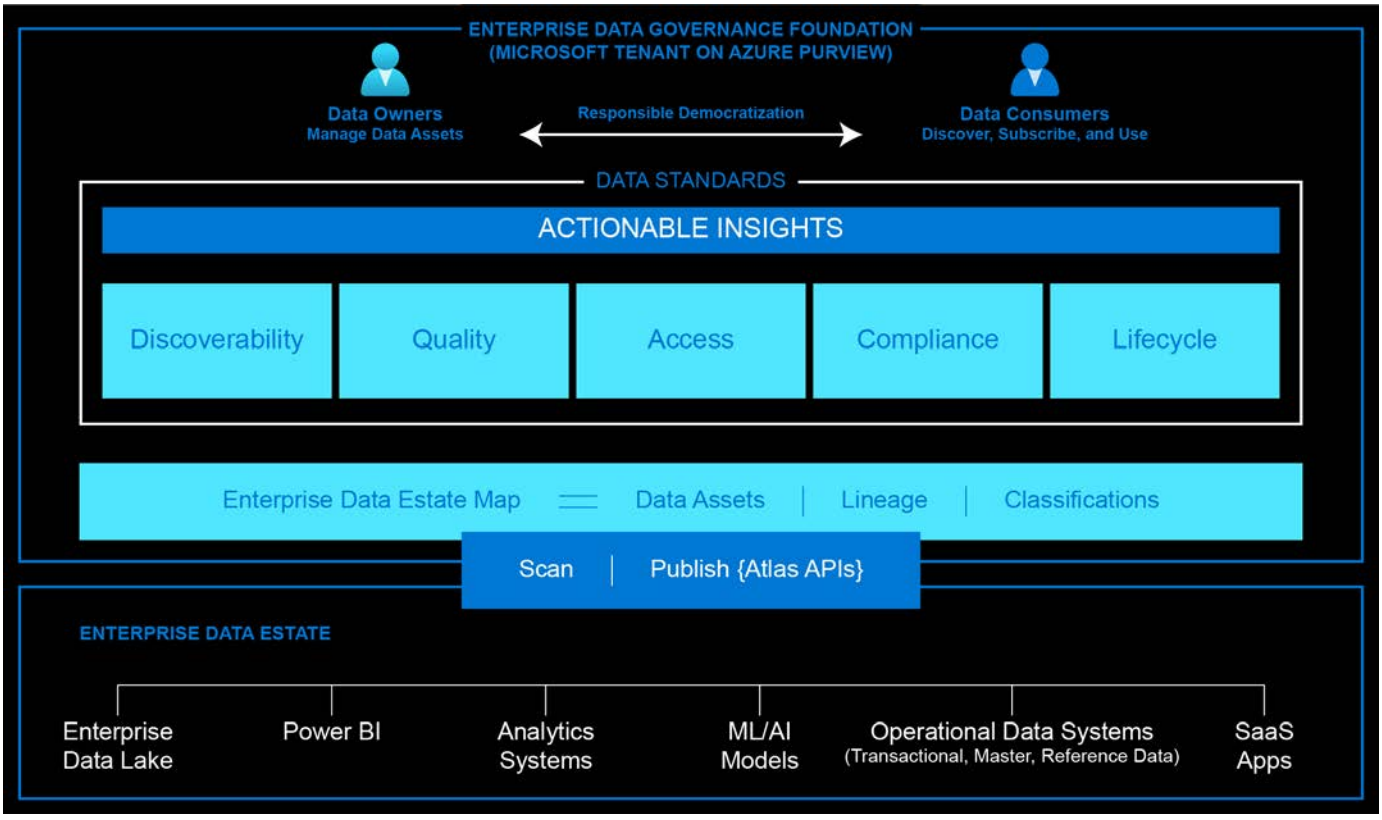


Corporate Compliance Program

Scaling with a technology foundation

Our approach to scaling data management and governance is anchored in our technology foundation. With this foundation, we can scale prior manual effort-intensive processes for data management and governance, with intelligent cloud-scale automation reinforced by essential human data stewardship. Cloud-scale automation of the essential controls to operate a secure, compliant, and quality Enterprise Data Estate (EDE) is fundamental to being able to responsibly democratize an organization's data assets to accelerate applied innovation and digital transformation. Such automation also entails seamless integration of essential human data stewardship processes and reinforcement to continually optimize an intelligent automation cycle for enterprise data management and governance.

Microsoft's Enterprise Data Governance Technology Foundation is powered by Azure Purview:



The Azure Purview Scanners and Metadata Publishing APIs are used to register Microsoft's EDE in an EDE map. The EDE map serves as the foundation to operationalize scalable automation with essential human data stewardship processes and reinforcement for data management controls and responsible data democratization. Related capabilities need to be included in a data governance program:

- **Data Discoverability** using an Enterprise Data Catalog with built-in capabilities for defining and managing a data glossary, data taxonomies, and data classification rules
- **Data Quality Management** to configure and maintain shared enterprise and domain-specific data quality standards
- **Data Access Management** for data owners to configure and manage data access policies, automated workflows for data consumer access request processing in compliance with data owner-defined access policies, and data access auditing and reporting
- **Data Compliance** for regulatory and corporate data compliance standards such as Privacy, GDPR, SOX, and the Data Governance Act
- **Data Lifecycle Management** for data retention in compliance with configurable data retention policies for regulatory and data use purposes
- **Data Health Scorecards** to automate measurements for data health standards and observability for timely actions in maintaining a healthy EDE

The human data stewardship is anchored in an Enterprise Data Governance Council comprising data stewards and domain subject-matter experts from Microsoft's central Data Governance team and domain teams. Azure Purview as the Enterprise Data Governance Foundation enables balcony and drilldown views of the EDE with observability and actionable insights to guide human steward processes and actions in maintaining a healthy data estate. Data consumers use the Data Catalog in Purview to discover, subscribe to, and gain access to enterprise data assets in compliance with data owner- and steward-configured data management, access, and use policies.

Collectively, the Enterprise Data Governance Foundation in Azure Purview is the single destination for data owners and data consumers in scaling the essential data management controls and applied data innovation for Microsoft's Digital Transformation.

Key learning and business impact

With increasing data maturity, our customer and partner experiences improved consistently. In a short period of 18 months, with a consolidated data estate and transformative enterprise applications, compelling product experiences, sustainable innovation, and high operational efficiencies were delivered.

Here are some examples of the automations and modernizations that we achieved:

Automations and modernizations

- **Accountable data owners and responsible data stewards** who manage and govern the quality and compliance of enterprise data subject areas.
- **Central control of permanent datastores**, which regulates the number and use of persistent datastores.
- **Single data masters (MDM)/sources of truth** for the most important enterprise master data.
- **Centralized data provisioning/authorization/access processes and tools** that enforce common data authorization standards no matter where the data resides.
- **End-to-end data tracking/lineage** allows the complete tracing of data consumption no matter where it gets consumed.
- **Common enterprise data catalog** with discoverable, certified sources of truth.
- **Governed information architecture** that centrally controls the most important data schema elements and data links (keys).
- **Shared systems of intelligence infrastructure** that bring together all data in one place for ease of discovery, lifecycle management, and compliance enforcement.
- **Data governance by design** gets institutionalized into the engineering DNA and is applied to data at every touchpoint (systems of intelligence, systems of engagement, or systems of record).

Here are some examples of the resulting business impact:

Framework

Microsoft Azure data management framework

Microsoft has utilized an automated and unified data governance approach internally that is now a core part of the Microsoft Azure data management framework used by customers. Our implementation experience with customers has informed a location-agnostic, hybrid, and multi-cloud data management product offering and framework.

Requirements for governing data in a modern enterprise

The aim of Microsoft's data management framework is to help customers understand their organizational maturity and define policies, processes, and architecture to improve and automate their unified data governance posture. This pairs with guidance to ensure that as data is moved to the cloud, it is automatically brought under a level of control in line with your organization's requirements.

The business case for data governance in the cloud

Due to the scale and automation in Azure's cloud data offerings, CDOs have a significant opportunity to modernize their data governance strategy as part of the cloud adoption journey. Whether a company's data strategy is tailored toward ensuring compliant use of data or driving value through insight, it is important to ensure the holistic benefit of data management in the cloud is understood.

Satisfying digital transformation and increasing demand for data

The inability to discover and access trusted data impacts business agility and the quality of insights, with 80% of analysts' time being spent simply discovering and preparing data[1]. As data workloads are moved to the cloud, there is an opportunity to discover and describe the entire data estate (on-premises and the cloud) in a single, accessible, and business-friendly data catalog—with 48% of companies reporting a better understanding of their data for insight after implementing data catalogs[2].

Solving data issues in an increasingly complex data landscape

As the amount and types of data processed by organizations in the cloud increase, the complexity of the technology, security, and policy landscape needed to manage it grows. The unified data governance approach used by Microsoft takes an automated, metadata approach to discover and classify data across multi-cloud and hybrid estates to apply data governance policies automatically. This allows data consumers to access data more quickly and safely, by reducing the time and cost of manually performing data governance processes.

Reducing the business impact of ungoverned data

Data regulations are increasingly complex and globalized, especially as governments and regulators look at the impact of cloud adoption on data privacy and risk. Over $190m in penalties were applied under GDPR alone in 2020[3]. Data governance, data protection, and data privacy professionals are looking at increasingly complex solutions to satisfy regional regulations evolving from GDPR, CCPA, COPPA, HIPAA, and more. Setting your data governance up in a unified, metadata-driven way across multiple locations and technologies will give you a significant head start in responding to these challenges and ultimately avoiding significant fines and reputational damage.

1 The 2 Types of Data Strategies Every Company Needs (hbr.org)
2 Forrester - Machine Learning Data Catalogs Put the Entire Business in Full View 2019
3 20 Biggest GDPR Fines of 2020 & 2021 (So Far) | Updated 2021 | Tessian

Microsoft's data governance approach

Microsoft supports customers adopting the cloud at different levels of data governance maturity and has developed a unique approach for customers on the unified data governance journey. Azure's metadata-driven foundation enables scalability and automation across the technical estate, but it is also important to treat business adoption as a key part of the data governance approach.

Set the scope of data governance for your organization

Be pragmatic, not theoretical, in identifying what data governance capabilities are required for your organization as part of your data governance strategy. Microsoft can support this by running a data strategy review and industry-defined data management maturity assessments.

Set enterprise data governance requirements through policies and standards

Data governance policies should be understandable, easily accessible, and syndicated with the business for buy-in. Microsoft can provide guidance on unified data governance policies; however, an organization's policies should represent the business and ideally be written by those responsible for enforcing them.

Set ownership and accountability for data governance

Data governance requires ownership at the business, technical, and compliance levels as a minimum, and many organizations will include even more stakeholder functions. Remember that data governance ultimately helps the business to safely manage and consume data, so identify the people that drive the most value through data and implement data governance in a way that serves them.

Start with a unified, metadata-driven vision with automation

Microsoft positions Azure Purview at the center of its unified data governance approach for all stakeholders. Purview is powered by a multi-cloud and hybrid cloud metadata and policy management estate that supports significant automation. Data governance professionals will agree that the biggest killer of data governance programs is a lack of automation, and this is compounded across multi-cloud and hybrid cloud environments.

Iterate, not big bang

When deploying a data catalog, it can be daunting to identify, scan, catalog, and classify across a significant data estate. Identify areas of most value or risk and iterate through assets. There will be a tipping point for your catalog as the number of assets under management drives data owners and consumers to the catalog for discovery and governance.

Educate and enable change

Data governance will require new roles in the organization to properly own, curate, and manage data assets. Ensuring these people are properly empowered and trained is a crucial part of a successful data governance process. However, training and enablement should not just stop with your data owners and stewards. Ensuring your data engineers, scientists, and analysts are aware of the governance requirements for the data they use, as well as training them on how to effectively use the catalog, will improve adoption and overall data literacy.

Monitor and revise

The Purview data catalog provides analytics and insights on the usage and management of data across the estate, enabling the CDO to make decisions around which data is proving the most valuable and where policies or standards might need to be revised. The usage of data is constantly evolving, so ensure there is a regular review of the policies, standards, and processes that govern it.

Core capabilities for unified data governance

Microsoft supports customers with an understanding of how managing data in cloud and hybrid environments may be different from on-premises. While not all capabilities will be relevant to all organizations, a holistic view is important for organizations looking to responsibly democratize their data at scale. It is equally important for CDOs to understand the impact the cloud provides with significant automation of cumbersome data management tasks and to reprioritize the enterprise data strategy budget.

These capabilities are closely aligned with the EDM Council's Cloud Data Management Capabilities (CDMC) industry standard, which Microsoft was a core contributor to.

The CDMC, a global collaborative industry effort, was developed by over 100 companies and 300 subject matter experts as an auditable certification framework.

The following is a holistic list of capabilities highlighted in the CDMC. The full CDMC model is maintained by the EDM Council and is available for free download here:

1. **Data Cataloging and Discovery** – The automatic identification and physical record of data assets in a unified manner to enable logical search, description, and discovery of an organization's data.
2. **Data Classification** – Tagging data with appropriate information, privacy, or other sensitivity classifications to secure onward use and protection.
3. **Data Ownership** – Ensuring data is owned for protection, description, access, and quality by accountable and empowered agents within the organization.
4. **Data Security** – Ensuring data is encrypted, obfuscated, tokenized, or has other appropriate security measures applied in line with its classification. Includes capturing evidence of security application and management of data loss prevention.
5. **Data Sovereignty and Cross-Border Data Sharing** – Ensuring data is being stored, accessed, and processed according to jurisdictional rules and prohibitions.
6. **Data Quality** – Ensuring data is fit for purpose according to the core measures of data quality—accuracy, completeness, consistency, validity, relevance, and timeliness.

7. **Data Lifecycle Management** – Ensuring data is sourced, stored, processed, accessed, and disposed of in line with its legal, regulatory, and privacy lifecycle requirements, which are often defined in a retention schedule.
8. **Data Entitlements and Access Tracking** – Data must only be accessible to those that are intending to access it. Auditing this access is an important part of evidencing and ensuring control.
9. **Data Lineage** – Ensuring it is possible to identify where data has originated, the steps it has undertaken, and where it is being used at a granularity and frequency that is relevant.
10. **Data Privacy** – Define a framework for the protection of the privacy of data subjects that reflects the regulatory and privacy laws governing your organization. Ensure processes and technology are employed to ensure the privacy framework is actively applied.
11. **Trusted Source Management and Data Contracts** – Large organizations may have similar data originating from or processed through a number of sources. Identifying and managing trusted sources and defining consumption data contracts is important to ensure data is being sourced from an agreed source of truth and the overall data architecture is being managed effectively.
12. **Ethical Use and Purpose** – Increasingly, the ethical use of data is being questioned beyond privacy laws and data subject rights. As the use of AI and machine learning increases, it is important to ensure data is being processed in a way that customers would expect according to your company's code of ethics.
13. **Master Data Management** – Master data is the most commonly used and duplicated data within an organization. It is often the data that describes the core operational aspects of a company (for example, product, customer, employees, and company structure). Ensuring there is a single consistent view of this data is fundamental to accurate and reliable data usage.

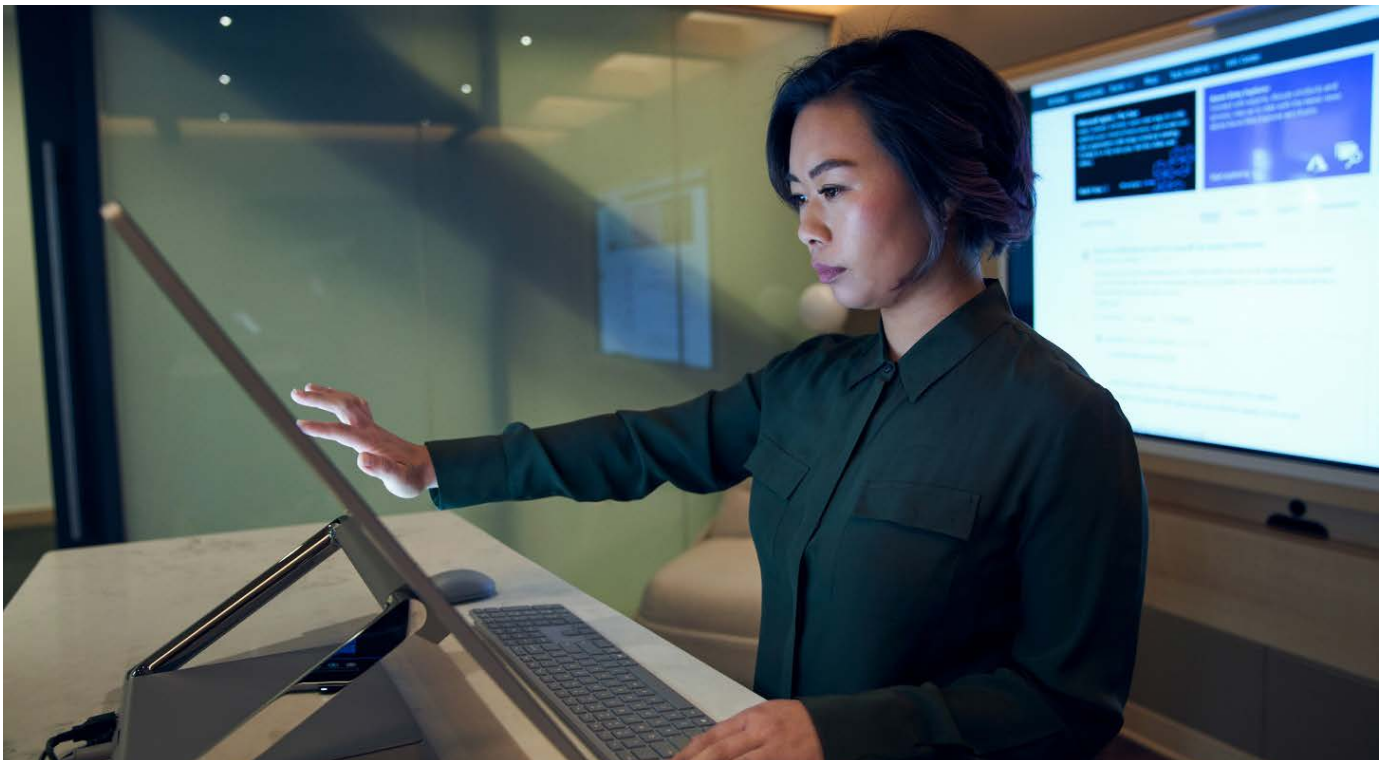
To satisfy the capabilities listed in the CDMC framework, Microsoft provides best-practice guidance on people, processes, policies, and technology. It is important to recognize that data governance is not satisfied by technology solutions alone, but in an increasingly hybrid and multi-cloud world, an interoperable data governance architecture is becoming a more important part of any solution.

Product

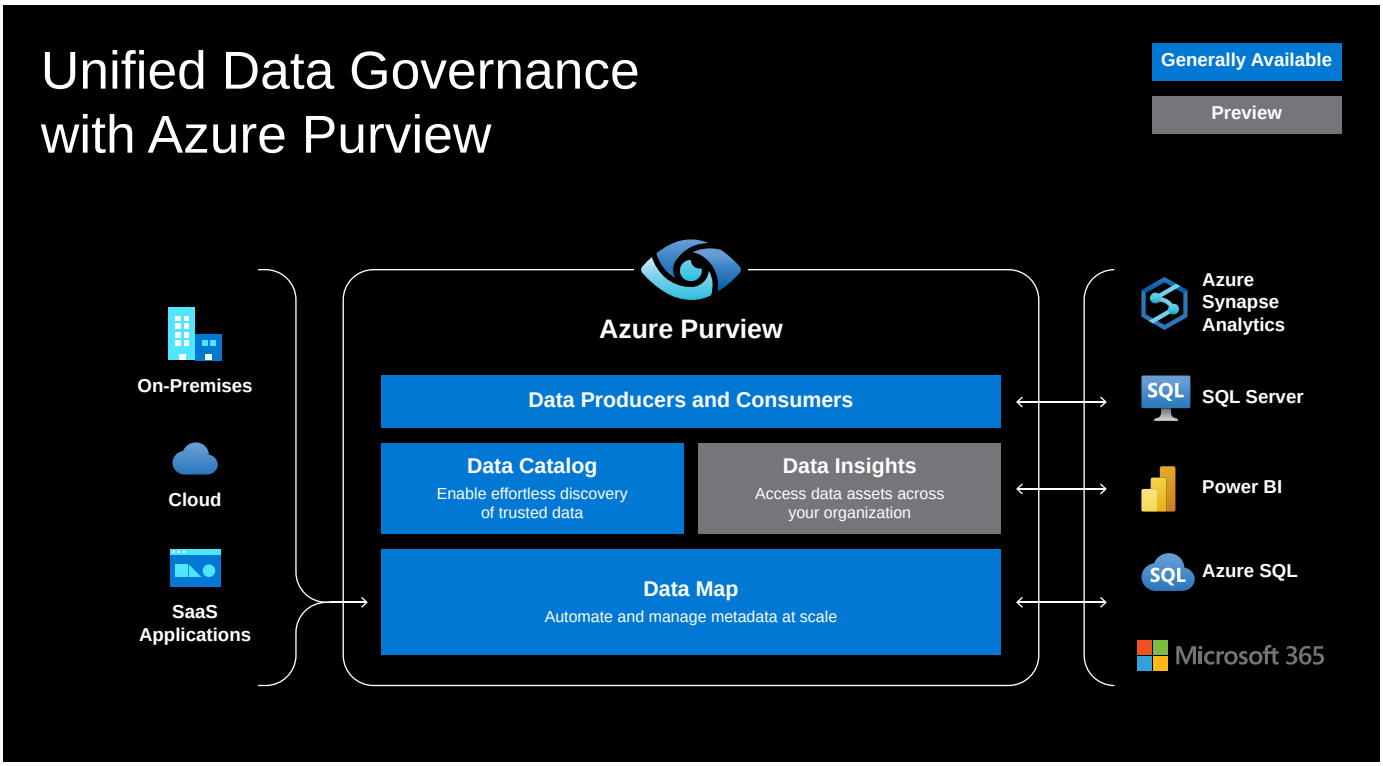
Products and architecture

In this white paper, we have emphasized the important role the CDO has in your organization's cloud adoption to achieve business objectives. From Microsoft's own experience in searching for a tool to solve its data management challenges, we realized there is no single effective tool, and introducing multiple niche tools presents interoperability issues where the cost of maintaining multiple tools may outweigh the value they bring. So, we decided to build one based on the challenges that we faced, those of our customers, and our deep relationship with regulators.

Currently, organizations are faced with the adoption of not just a single cloud, but multi-cloud environments. We wanted to build not just a product that helped organizations with their on-premises and Azure environments, but one that would empower CDOs to properly discover and control data in a single view and not worry about the interoperability between the cloud provider or SaaS solution—the result was Azure Purview.



Azure Purview simplifies data governance by enabling organizations to create a holistic, up-to-date map of all the data assets in the organization with automated scanning and classification. We created Purview with APIs built on open-source Apache Atlas that enable our partners to innovate on top of the data map, enriching it with various business, quality, relationships, and other contexts useful for business data consumers to easily find and use data:



Resources

Getting started with Azure data governance

Thank you for reading this white paper. We hope you enjoyed it. Microsoft offers many resources to get started today no matter where you are in your cloud journey. The following links can help you get started today.

References

Product

Looking for an introduction to Microsoft's data governance product capabilities?

- Azure Purview for unified data governance | Microsoft Azure

Microsoft Journey

Learn more about Microsoft's internal data modernization journey:

- Powering digital transformation at Microsoft with Modern Data Foundations

Self-Paced Skilling

- Please visit Microsoft Certified: Azure Data Fundamentals – Learn | Microsoft Docs

Cloud Adoption Framework for Data Governance

- Azure data management and analytics scenario

Industry Standard (CDMC)

- Cloud Data Management Capability (CDMC) framework

Customer Stories:

- How Microsoft helped Walmart in accelerating digital transformation
- Humana and Microsoft announce multiyear strategic partnership to reimagine health for aging populations and their care teams
- ABN AMRO embraces an Azure-first data strategy to drive better business decisions

Further reading

- A brief history of the CIO
- Harvard Business Review: Why do Chief Data Officers have such short tenures?
- The CDO reason, role, and responsibilities for 2017
- EDM Council Benchmark Report
- How do you measure success in digital? Five metrics for CEOs

Acknowledgements

We'd like to acknowledge the following contributors to this white paper:

Karthik Ravindran
Ria Sankar
Shafiq Mannan
Santosh Kumar Ravindran
Todd Singleton
Mike Flasko
Srikanth Venkat
Pratik Verma
Effie Kilmer
Anna Fisher
Nick Lewins
Kevin Zhang
Damon Buono
Divya Polavaram
Piethein Strengholt
Alex Posar
Jubair Patel
Samuel Madiki
Suresh Yadav
Mike Boswell
Angus Foreman
Ben Wright-Jones
Tillmann Eitelberg