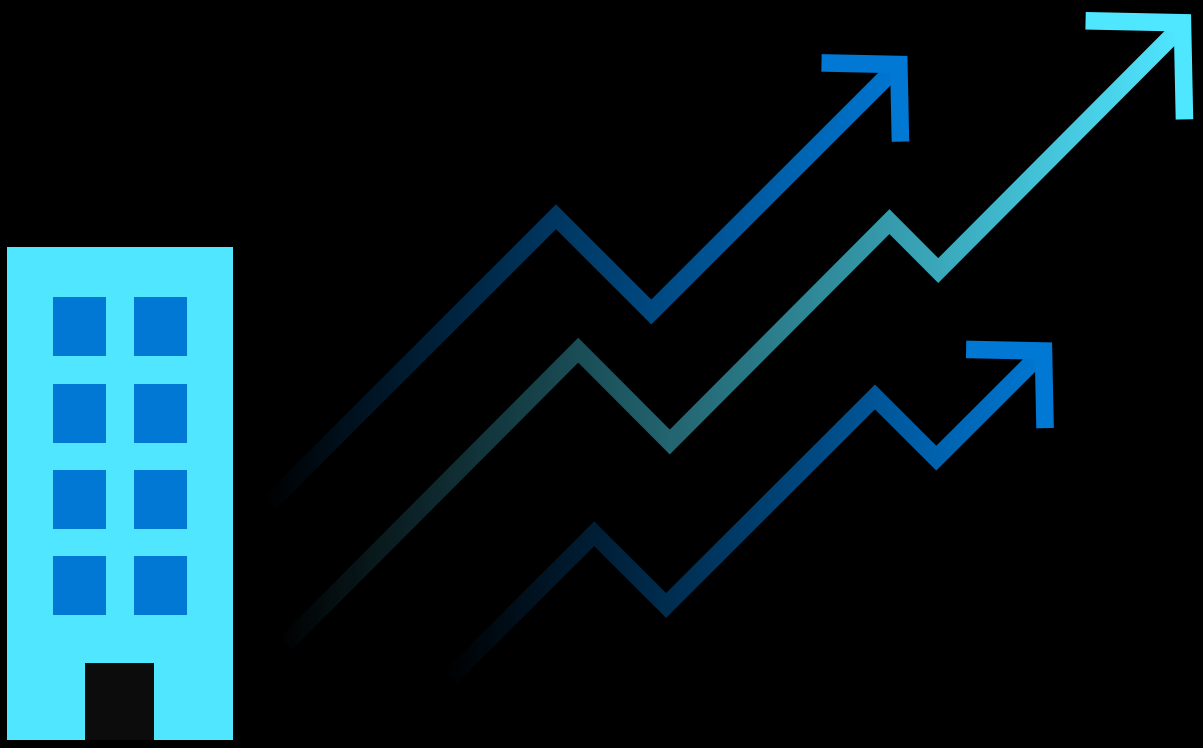
Microsoft Azure

# Modern Analytics and Governance at Scale

## A transformative approach to building a modern, end-to-end analytics ecosystem aligned to data mesh, data fabric, and data hub



# Table of Contents

**Executive summary** ..... 3

**Common challenges to digital transformation** ..... 5

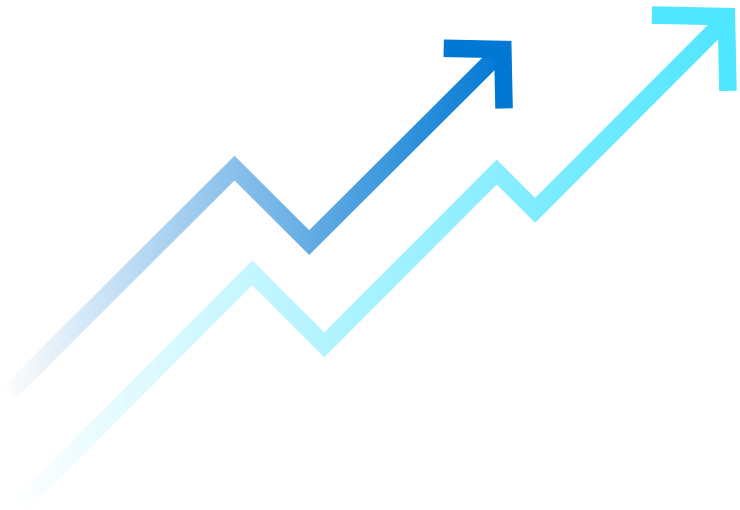
**Building an end-to-end analytics ecosystem** ..... 7

**Modern Analytics and Governance at Scale** ..... 9

**Realize the vision with Microsoft products and services** ..... 23

**Conclusion** ..... 25

**Who this is for:** Data leaders within organizations that are undergoing digital transformation who want a proven approach to building an end-to-end data analytics ecosystem.



# Executive summary

In today's organization, data is at the center of the universe, propelling digital transformations forward. But it's not just data behind transformation. It's the analytics and intelligence that consume the data and power applications, as well as the culture, people, and processes that bring them to life. Given all the interrelated pieces, organizations need clear priorities when building a unified analytics ecosystem. It comes down to having the right business goals for transformation, foundational elements to anchor the design, and a phased implementation plan that offers room to grow and evolve. This whitepaper outlines those components and simplifies the journey to a unified analytics ecosystem.

Guiding the way to modern analytics are three over-arching business priorities. First, organizations want to **accelerate the creation of value from their data** to stay competitive. To do so, they must find operational efficiencies and simplify architectures to make it easier to create new data products that drive business decisions. However, the ability to create new data products can't be a privilege held in only a few departments. Organizations need to empower all personas with the data they need, regardless of their roles. **Responsibly democratizing data**, which is the second priority, makes data and tools widely accessible for all users to foster collaboration and innovation. Of course, neither of these priorities is possible without breaking down silos between business units, which is the final objective—to **unify data**. With enterprise data governance that makes data discoverable, a hub that contains all data, and domains where business users can collaborate, different parts of the organization can share and integrate information to build meaningful insights.

# Modern Analytics and Governance at Scale

To achieve these priorities, Microsoft developed Modern Analytics and Governance (MAG) at Scale. Broken down into the following three solution pillars, this approach outlines the foundational elements that organizations need to transform, along with practical considerations for executing each section.

**Enterprise data governance:** the set of policies and practices used to discover, describe, and manage data to accelerate responsible data democratization. Data governance ties together the data and analytics stack and automates data operations, such as cataloging, classification, creating lineage, and applying security through policies. Without it, organizations limit their ability to innovate and unlock new insights.

- This section reviews key components typically involved with data governance policies to help you start planning for how to standardize across your organization.

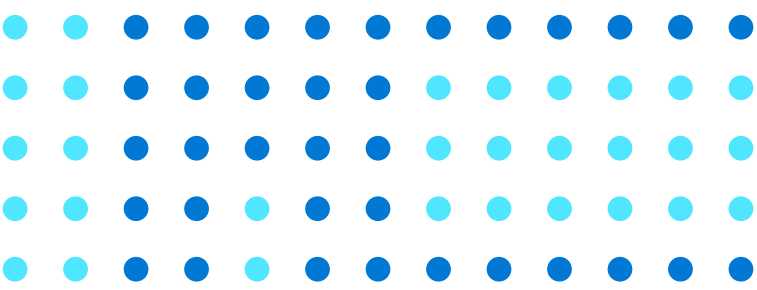
**Data management foundation:** the practices and processes that help you create efficiencies with ingesting, storing, protecting, and ultimately serving data to different domains in the organization.

- This section provides recommendations on foundational implementations to help you manage and use your data effectively and efficiently throughout its lifecycle.

**Domains and data products:** the environments and services that enable your business units to fully use their data. This looks at how to empower lines of business (LOBs) to self-serve data and analytics in a way that enables non-technical users to access, analyze, and build data insights or data products on their own.

- This section provides an overview of the different analytics patterns and services you can provide domain users to help them shape and present data in a way that is consumable and actionable for them.

Together, these three solution pillars combine to help organizations achieve MAG at Scale. By first setting business priorities as the guiding North Star, then implementing aspects from each solution pillar, your organization can shift into a whole new paradigm of doing business that empowers everyone to work toward a common goal fueled by data.



# Common challenges to digital transformation

In most cases, organizations struggle to implement an end-to-end analytics ecosystem, not because of the technology, but because they do not plan for, nor address, the role culture, people, and processes play in bringing about digital transformation. In fact, the MIT Technology Review says nearly every time—92 percent—an organization struggles with data it links back to an issue with the company's data strategy, data governance, and/or data management.

This is because many organizations approach a data challenge with a technology solution—but that only solves a fraction of the problem. Organizations need a wider and deeper focus on solutions that drive cultural changes and align people and processes with technology.

Throughout hundreds of engagements with organizations worldwide to help them become data driven, Microsoft has seen the following top challenges over and over. Here are three common problems associated with culture, people, and processes that impede MAG at Scale.

## Lack of data strategy leads to a siloed ecosystem

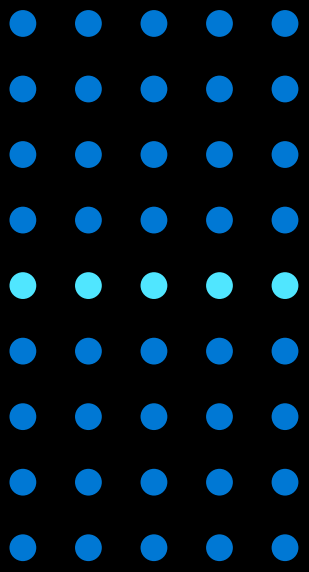
After years of implementing different analytics projects in the cloud, organizations continue to build out their ecosystem in a reactive, piecemeal way. Without a well-defined data strategy, solutions become siloed and technical debt increases, which stands in the way of bigger data and analytics innovation. The lack of an analytics foundation for all data also inhibits a thriving ecosystem.

For example, imagine a data warehouse migration that lands data in proprietary data formats. At the same time, Internet of Things (IoT) data is streamed into a data lake store. Data from these two separate projects lands in separate data stores, causing siloed data. To become data driven, the entire organization must be able to build meaningful insights from both sources, regardless of the boundaries between each business unit. LOBs need access to data of all attributes and types.

Why do organizations **struggle** with Data?

**92%**
Culture, people, processes

**8%**
Technology



## Poor governance prevents democratized data

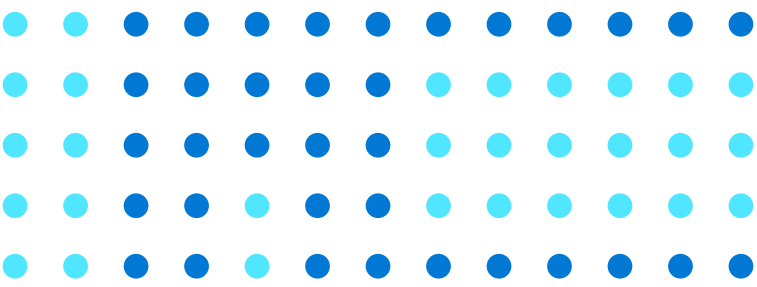
Data governance—as in a clear set of policies, processes, and controls—is critical for organizations to find, manage, and consume data across the businesses. Lack of data governance and an incomplete understanding of the data can stymie analytics projects. Failure points can run the gamut. Some projects can't access data fast enough, others can't integrate data due to the lack of data relationships, and others require additional data engineering work to be completed before being able to train machine learning models.

All these problems can be avoided if the organizations have enterprise data governance that provides the inventory and context of all data, automated processes to streamline workflows, and policies that automatically manage data access. The goal is to implement robust data governance and data management that enables different analytics projects for different LOBs.

## Manual data management slows time to insights

For data to be consumable, there's a ton of ingestion and data engineering work that must happen. It must be cleaned with approved data quality, integrated to reveal new business insights, and aggregated to become a proper data product. Most organizations tend to build manual data engineering workflows on a case-by-case basis aligned to specific projects.

However, this creates fragmented data engineering solutions that become harder to maintain as they grow to include thousands on top of thousands of pipelines. The worst part is, most data engineering tasks are manual. Many organizations use their best people to perform these manual tasks when it should be a process change instead. By implementing proper data management processes with automation, organizations can re-assign data engineers to more meaningful work of business data modelling, data aggregations, and calculations.



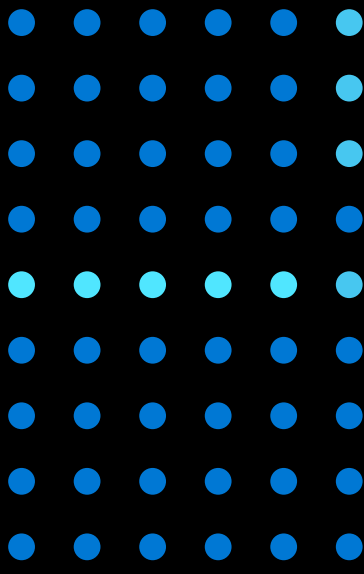
# Building an end-to-end analytics ecosystem

With a unified analytics ecosystem, organizations can streamline and optimize their analytics operations and enable all departments to share data. This accelerates the creation of value from data and enables each persona to build their own insights—from IT all the way to LOBs. It may sound like an impossible dream, as many organizations have implemented portions of this vision, but have yet to realize the full potential of MAG at Scale. On the other hand, some have had to backtrack because foundational elements weren't properly developed beforehand.

In the academic world, there is much debate over what an end-to-end analytics ecosystem looks like. Frameworks like data hub, data fabric, and data mesh have strict definitions that can put organizations down a narrow path that misses the over-arching objective—to bring analytics and insights into everyday work. Most popular among these frameworks is data mesh, which focuses on building domains that allow LOBs to access the data they need and operate autonomously to build their own data products.

While data mesh meets the requirements for LOB users, it can create extra work for the teams responsible for getting data to different domains. By applying key principles of the second framework, data fabric, organizations can implement different data services and add automation to ingest, standardize, curate, and integrate data. This helps deliver data to the domains and accelerate data value creation. The third framework, data hub, provides the foundation to store and secure data, while allowing different tenants to be implemented and aligned to each data domain. Organizations use a data hub to consolidate data sources and apply uniform governance to store all types of data at varying stages (from raw, to conformed and data products) so it can be consumed by different domains.

**Data mesh, data fabric, and data hub** are not mutually exclusive frameworks. Together, they make implementations more practical and are collectively **transformative.** Based on proven implementations, Microsoft recommends using key elements of each framework to build a durable and scalable analytics ecosystem for a data-driven organization.



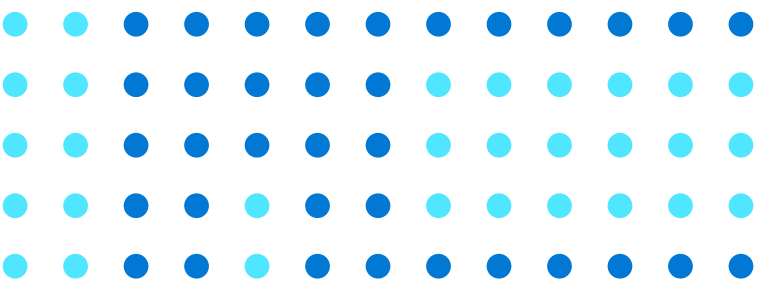
MAG at Scale uses aspects of data hub, data fabric, and data mesh to arrive at that vision. However, rather than providing an academic depiction of a modern data estate, this approach offers practical and real-world considerations for implementing an analytics ecosystem. The goal is to keep organizations focused on the foundational elements and steer away from the noise, nuances, and analysis-paralysis of what an end-to-end analytics ecosystem should look like.

The Microsoft approach is based on best practices refined over more than two decades of firsthand experience in building global-scale products and services. It starts with a holistic view of the organization, factoring in the people, processes, culture, and technology. Then, by applying data governance, security, and compliance across every layer of the stack, it ensures customers have a truly innovative environment that empowers everyone to do their best work.

## Modern Analytics and Governance at Scale

The Microsoft hybrid approach to data mesh, data fabric, and data hub

| **Data mesh** | **Data fabric** | **Data hub** |
| --- | --- | --- |
| A series of domains assigned to individual LOBs that enables access to the data they need with maximum autonomy by upholding the four principles of a data mesh. | A system for automating data management tasks, such as unifying and cleaning disparate sources as well as authorizing data access, that helps a business make the most of its existing data sources without needing to migrate them. | An open and governed lakehouse platform that provides the storage foundation for serving data to multiple domains in a federated way. |



# Modern Analytics and Governance at Scale

MAG at Scale is the Microsoft-recommended implementation for an end-to-end analytics ecosystem that includes the following components:

| | |
|---|---|
| **Enterprise data strategy** | Culture transformation<br>Aligned process, people, and technology<br>Autonomous LOBs<br>Organizational change management<br>Platform and data ownership |
| **Composable solution** | Enterprise data governance<br>Data management foundation<br>Domains and data products |
| **Technical architecture** | Azure Cloud Adoption Framework<br>Intellectual property (IP) and accelerators by Microsoft and partners |
| **Implementation stages** | Discovery, MVPs, and phases |

This white paper discusses enterprise data strategy at a high level and focuses on explaining the different solution pillars that make up the composable solution. At the end, there is guidance on implementation stages.

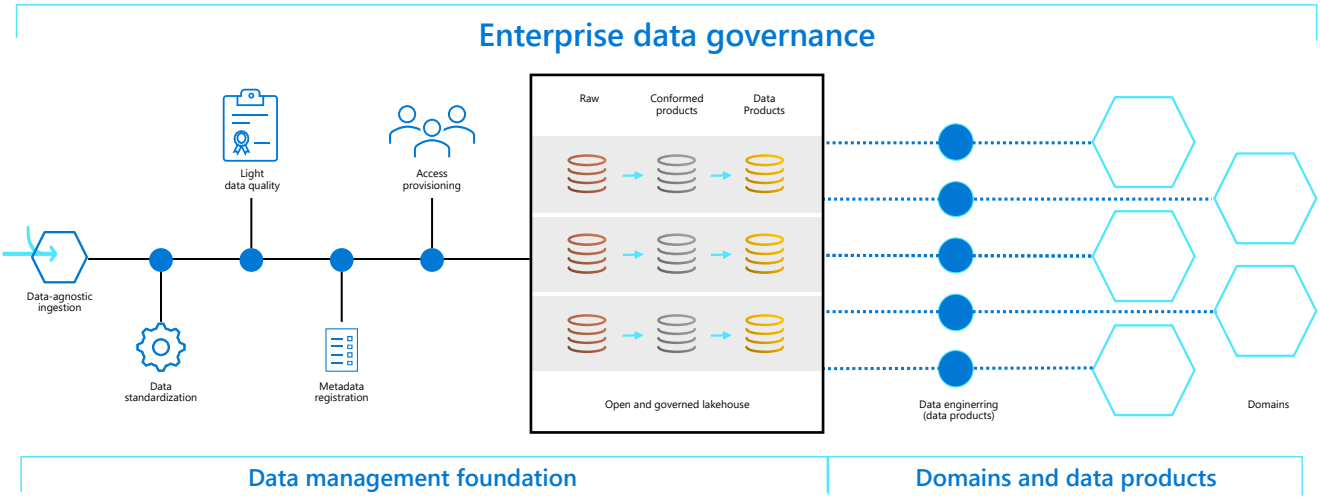



Figure 1: Three solution pillars for MAG at Scale shown in the context of a data estate, in accordance with a defined data strategy

# Enterprise data strategy

Data is the oxygen of your digital transformation, but there are so many factors that make it possible for that data to be useful. To truly drive transformation, you need an enterprise-wide data strategy that your entire organization can rally around. It comes down to your people, culture, and change management—in addition to your technology.

Here are the major pieces to consider when thinking about your enterprise data strategy:

> **Align people, process, and technology:** Simply implementing a technology solution is not enough. You need to align your people and their skills to drive change, adapt new processes, and evangelize the benefits throughout your organization.
>
> **Initiate a culture transformation:** This effort is rooted in helping employees outside of traditional technology roles develop their digital skills along with a growth mindset. This sets the stage for a culture that embraces frequent experimentation and process agility.
>
> **Enable autonomous LOBs:** To unlock the value of your data, you need to empower and enable your entire organization to use data, not just central IT teams or one or two business teams. All LOBs in the organization need true autonomy to find the data they need and execute the analytics patterns that make sense for their work.
>
> **Plan for change management:** Without a strategy for change management, employees can resist technology adoption and the organization as a whole may be misaligned. Be sure to communicate your goals, build a team who can reassure their peers, and keep the process iterative and always improving.
>
> **Clarify platform and data ownership:** Traditionally, IT owned the data, but for democratized data to take shape that ownership must shift to include domain teams. The data foundations team (potentially from IT or shared services) owns technical architecture and environment implementation including solutions that deliver data services to the domain teams. However, LOBs own the data, tools, and data services in their domains which enables them to build data products and share them with other domains.
>
> **Prioritize foundational elements versus technology features:** While it can tempting to focus on specific business needs in isolation and create solutions to address individual problems, technology choices should be grounded in foundational design principles that serve the organization as a whole. For example, traditional data warehousing solutions can lead to proprietary data silos that only serve data for descriptive analytics. This makes it difficult to use the data for real-time and machine learning use cases without incurring extra cost and overhead. By leveraging open data formats supported by enterprise-grade compute nodes, you can unblock all types of analytics for everyone in the organization.

The items above are just few of many to consider when building your enterprise data strategy. Please reach out to the Microsoft Azure Account teams supporting your organization.

# Composable solution

With an enterprise data strategy in place, your organization has a vision and roadmap for how it will tap into data and accelerate value creation. This lays the groundwork for building the three fundamental pillars of MAG at Scale:

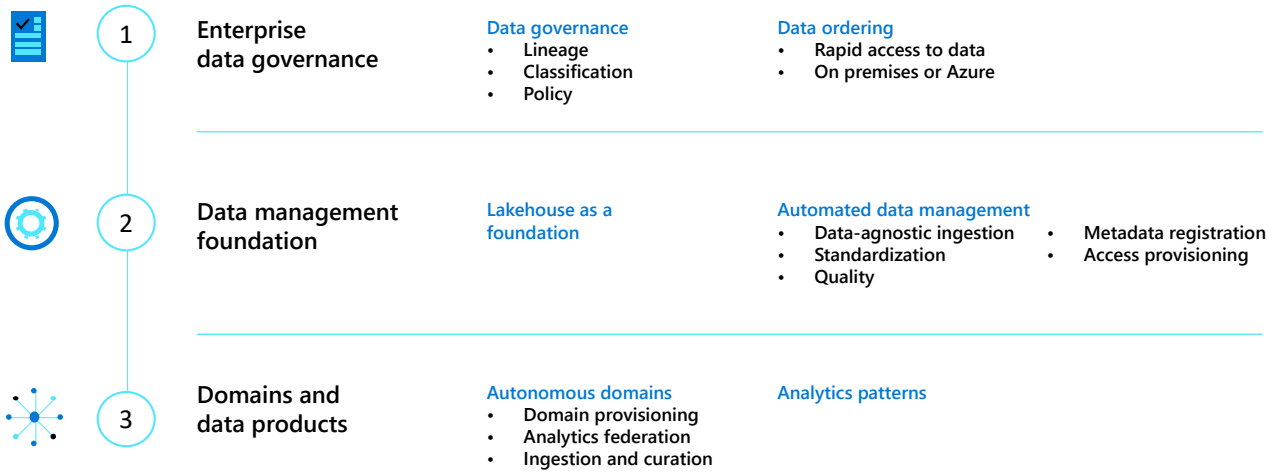



Figure 2: Three solution pillars of MAG at Scale

These pillars apply to both to the domain teams (LOBs users) and data foundation teams (IT teams) in your organization. They each consist of several components that can be addressed in phases or piecemeal depending on your needs. The flexibility inherent in the three pillars gives you agility when it comes to implementation and maximizes the use of your existing solutions. While each pillar is critical, it is the combination of the three that unlock organizations' digital transformation.

**Domain teams** include LOB groups and individual data consumers. They use the data to create data products.

**Data foundation teams** typically consist of IT or shared data services. These groups own the architecture and deployment of the environments, enabling domain teams to use data and implement their own data products.

The following diagram illustrates the capabilities of the end-to-end ecosystem, which organizations can achieve by implementing the three solution pillars of MAG at Scale. Note there are more services and capabilities enabled through this implementation that are not shown on this diagram.

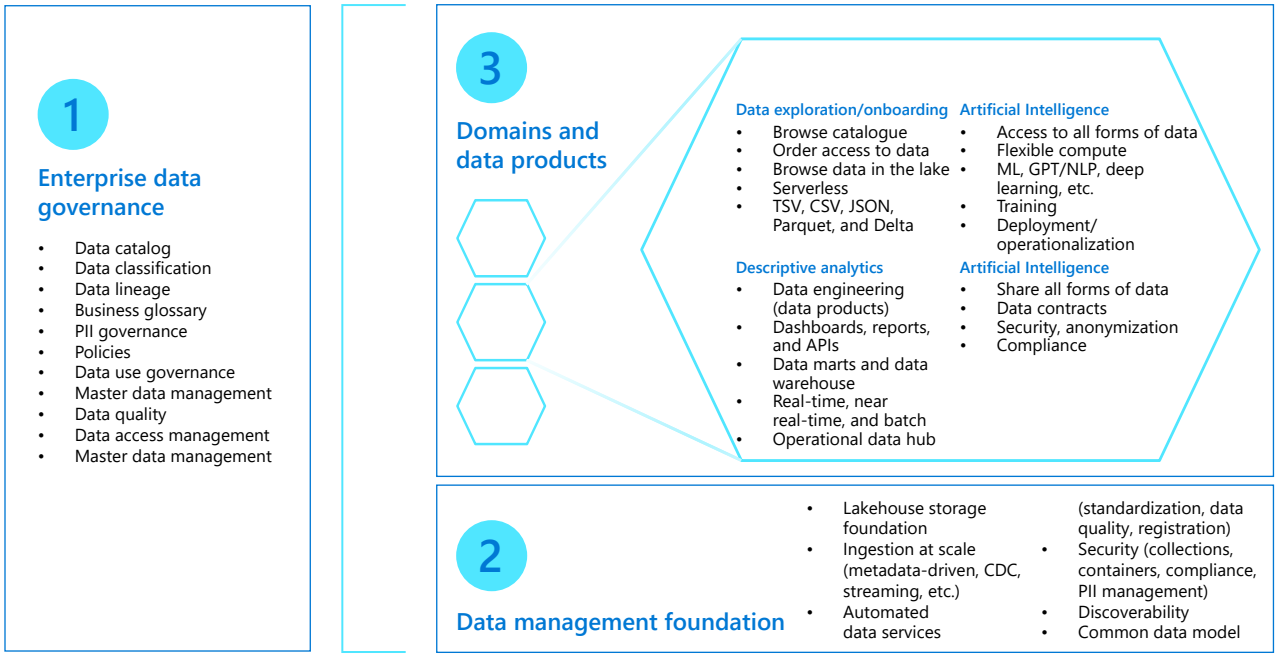



Figure 3: Capabilities enabled by MAG at Scale

## Enterprise data governance

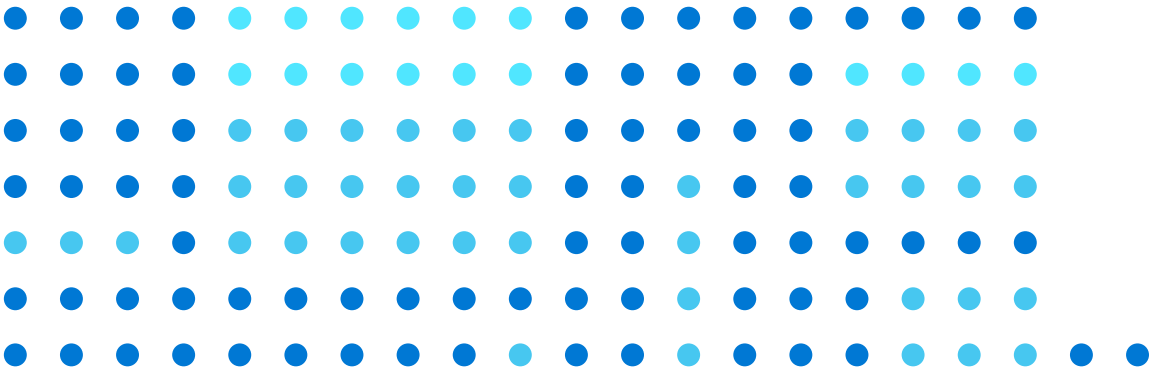
Governance is essential to every organization—regardless of the framework or solution implemented—because it lays the bedrock for responsibly democratizing data. Data governance translates your data strategy into data ownership, rules, and policies that improve data discoverability, confidence, security, compliance, and operational efficiencies.

Over the years, data access and use has spread to all corners of a company with individuals making critical business decisions affecting organizations, customers, and shareholders. The range of disparate uses, abuses, and copies of data is leading to widespread confusion and risk. You might see the finance department visualizing billions of rows of risk data, while analysts in marketing identify customers for a new product. Data governance provides the glue that ties together all the data in the analytics stack and ensures the right data is easily accessible by the right people.

Data governance provide solutions for the following objectives:

> **Discover and describe data** across business units and provide pertinent information, such as metadata, lineage, profile, quality, business context, and classification to help different personas decide if the data is useful or not.
>
> **Automatically secure data** at storage layers by using data classification and policies to implement security instead of using IT to manually apply controls to every data asset—this reduces errors and greatly improves efficiency.
>
> **Democratize data access** by allowing all personas to discover and request access to data. There are two major pieces to democratization, discoverability and having an efficient way to request access, for instance, through a catalog. The data catalog should inventory all your data whether it is raw, conformed, mastered, or already built as a data product.
>
> **Provide data leaders an understanding of their data.** With all data assets properly mapped, cataloged and classified, data leaders have a better sense of their total number of data assets, as well as their profiles, attributes, and confidentiality.

The following sections discuss how your team can define, implement, and automate rules so your data sources and data users are as well understood and auditable as you need them to be.



## Lay the groundwork for better governance

Data foundation teams play a crucial role in establishing better data governance throughout an organization. It's up to them to organize data in a way that drives accountability and ownership, associates data with business processes, and simplifies the compliance and overall execution of data governance tasks.

Here are four key ways data foundation teams can improve enterprise data governance.

**Creating metadata:** Well-governed data includes information about its origin and quality so that it's trusted and useful for all end parties. Metadata should capture the data's lineage, profile, quality, business context, and classification to help LOBs navigate various datasets and determine what they need.

**Mapping data assets:** With a data map, data consumers can easily and visually inspect all data assets across all domains whether it's physically stored on premises or in the cloud. The data map describes physical data assets, their relationships to other artifacts, and associated properties like personal identifiably information (PII), protected health information (PHI), or other regulatory attributes, business tags, and associated stakeholders. Domain groups can also be defined to logically group data assets that align to domain nodes.

**Cataloging data assets:** Data should be discoverable and accessible to all users through a data catalog. This not only organizes data, but it also democratizes access for LOBs. With a catalog, data consumers can surface the location of all datasets and see whether or not they're complete. A catalog provides a business glossary allowing data product owners to define the business context associated with the data asset. For data stewards, the catalog allows them to curate and manage their data inventories to ensure accurate input for data discovery and compliance processes. When changes are made or an approval is requested, workflows send out automated notifications to streamline the process.

**Automating governance and security:** Instead of manually implementing data security for each data asset—which is prone to errors and highly inefficient—automated governance keeps data secure while still ensuring access of data to the right users. Automation works by implementing policies based on data classification or attributes. For instance, human resources (HR) classified datasets can automatically be made accessible only to the HR team. While IT teams are tasked with empowering LOB teams with access to the data they need, they must also ensure security audits remain a top priority. That includes understanding how data assets are used across an organization to ensure they're used appropriately.

Gain an aggregated view with up-to-date information and alerts about your data estate from the Data Estate Insights application from **Microsoft Purview.** The platform also allows data domains. This way, data can be organized to drive accountability and ownership, associating data with business processes, and simplifying the compliance and overall execution of data governance tasks. As part of this, data policies can be articulated and automated.

## Enable data democratization through self-service

To truly democratize data, organizations need to implement a solution where data is discoverable through a data catalog and domain users can request access without having to go through the traditional manual process of opening a ticket, reviewing the permissions with different teams, and then implementing security changes.

Members of the domain teams (data analytics, data products developers, data owners, etc.) should be able to browse the data catalog to discover new data relevant to their use case. Displaying information about the data, such as metadata with classification, lineage, business terms, related assets, and data owners, helps domain teams validate the data and make good decisions about which sources to explore. Once they have selected data sources, users should be able to order access to it if they do not have pre-established permission.

Leverage Data Use Governance within Microsoft Purview to implement self-service for data already stored in some data stores in Azure. The solution uses an automated workflow to send the request to the data owner. Approving the request automatically executes the workflow to provision access. **Learn more about requesting access for a data asset.**

You can implement this manner of data sharing with data sources hosted outside of Azure as well. In the case of an on-premises data source, metadata from the data catalog (server names, table names) can be used to onboard the data into an ingestion framework so the data flows from the data source to the data lake in Azure. Once the data is in Azure, access can then be granted as part of the Data Use Governance solution, thereby furthering the goal of data democratization. Additionally, this workflow helps organizations comply with data policies they have that prohibit performing analytics on top of operational data stores due to governance, security, and performance issues.

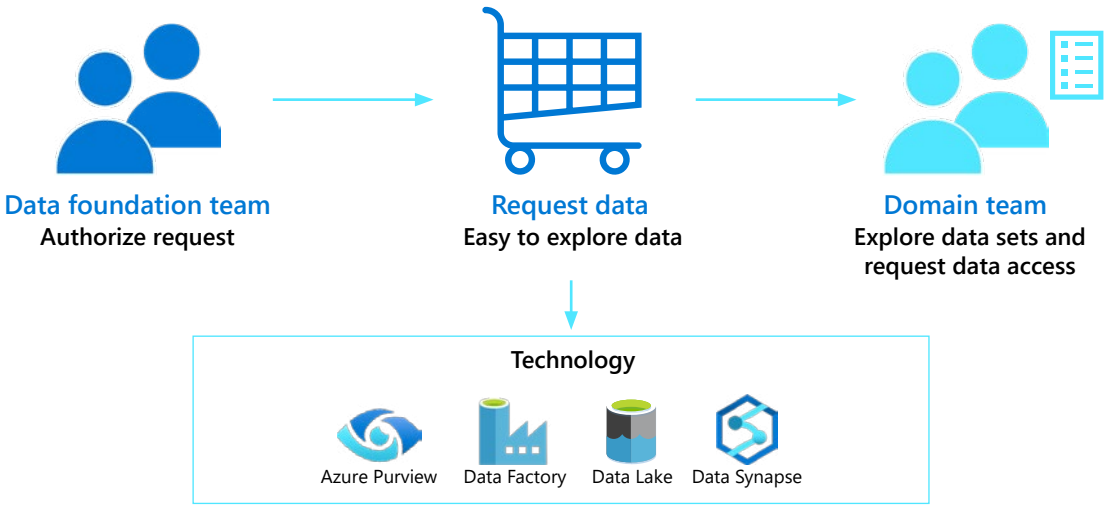
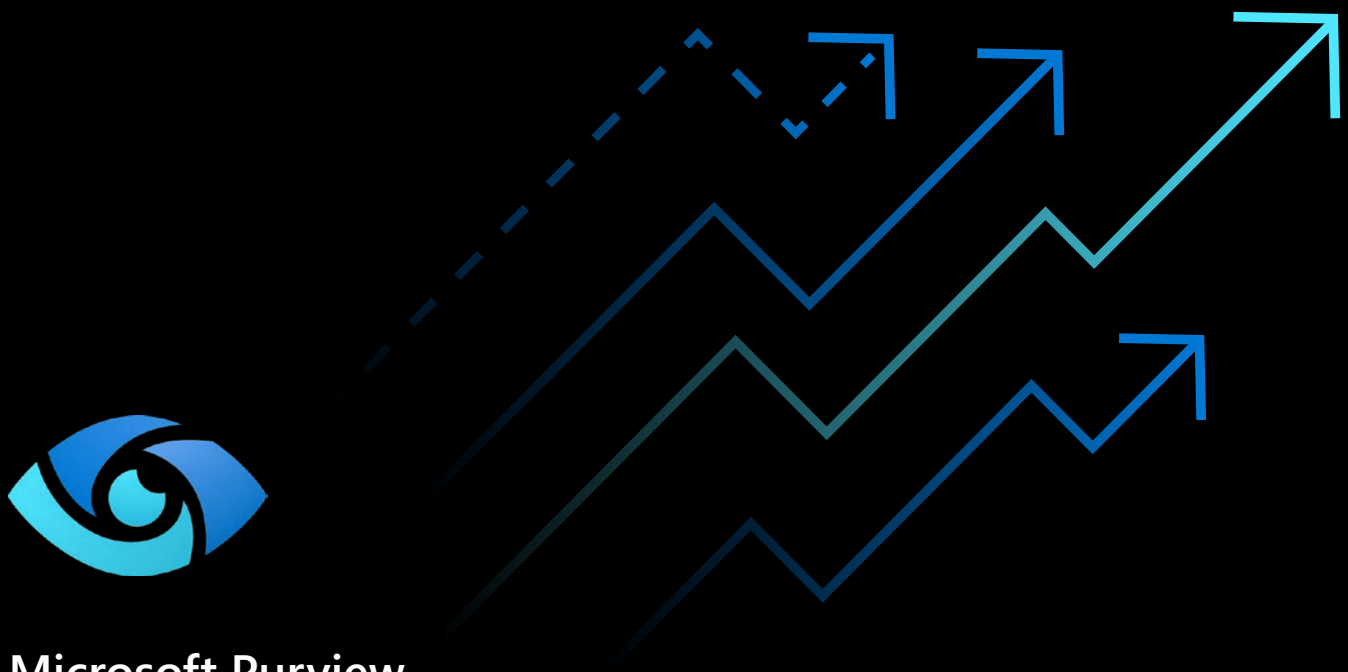



Figure 4: Using data governance to enable self-service



# Microsoft Purview

**Microsoft Purview** provides a unified data governance solution to help manage and govern your on-premises, multi-cloud, and software-as-a-service (SaaS) data. You can easily create a holistic, up-to-date map of your data landscape with automated data discovery, sensitive data classification, and end-to-end data lineage. Enable LOBs to access valuable, trustworthy data management. With Purview, you can take advantage of the following capabilities:

- Automatically scan and catalog all datasets across the organization, whether they're on premises, in Azure, or running on other public clouds
- Automatically capture metadata and data lineage to help personas to decide if data is usable
- Automatically classify data with pre-defined and custom rules
- Automatically apply data policies based on classification or attributes of the data (ex: masking data that was classified as confidential)
- Allow different LOB users to request access to specific data via the data catalog
- Catalog analytics assets such as machine learning models and Power BI reports
- Enrich data with business glossary, descriptions, and data owners
- Profile data at the source to indicate attributes like min, max, average, and thresholds
- Measure and implement data quality as part of the data pipelines

## Data management foundation

The purpose of data management is to ensure that data is properly collected, stored, processed, analyzed, and used in a secure and efficient manner to support an organization's goals and objectives. According to MIT research, 44 percent of leaders indicated that a lack of data management holds their organizations back from becoming data driven. Likewise, 39 percent of leaders attribute this setback to slow data processing.

Implementing automation, frameworks, and services can help organizations bolster their data management practices. When creating a data management foundation, the goal is twofold. On one hand, it's about implementing an open and governed data lakehouse foundation for storing data. On the other hand, it's about automating processes that enable self-service, so that requested data is sent to domain teams efficiently without overburdening data foundation/IT teams.

### Establish a storage foundation

At a high level, an opened and governed storage foundation allows you to unify your data estate by integrating data from different data sources including on-premises and multi-cloud locations. And the keywords here are open and governed because everything in the data lakehouse foundation is based on open data formats and standards. This enables durability that's not possible with other closed-standards or closed-format-based analytics foundations.

Many organizations today implement analytics solutions that store data in a proprietary storage format, which causes data silos and inefficiencies. For example, consider a proprietary data warehouse platform that serves data for descriptive analytics. When the organization wants to use this data for other analytics practices such as machine learning, it will incur dual compute costs for the data warehouse and machine learning compute nodes. There are also cases when the data will need to be extracted to a data lake before training machine learning models. A lakehouse solution solves this problem by storing data once with OneCopy and serving it to all types of analytics, including but not limited to, business intelligence, machine learning, streaming analytics, and data exchange. This approach of having OneCopy of data provides the most cost-effective way to store and share data.

What holds companies back on becoming data-driven organizations?

**44%**
Lack of data management

**39%**
Slow data processing

**29%**
Lack of collaboration on analytics

**22%**
High data duplication

## Automate data management

Running on top of the storage foundation are automated data services that help improve efficiency around data ingestion, standardization, quality, metadata registration, and access provisioning to benefit both the data foundation and domain teams. This results in less repetitive, manual work for IT and it lays the foundation for enabling domain teams to self-serve data.

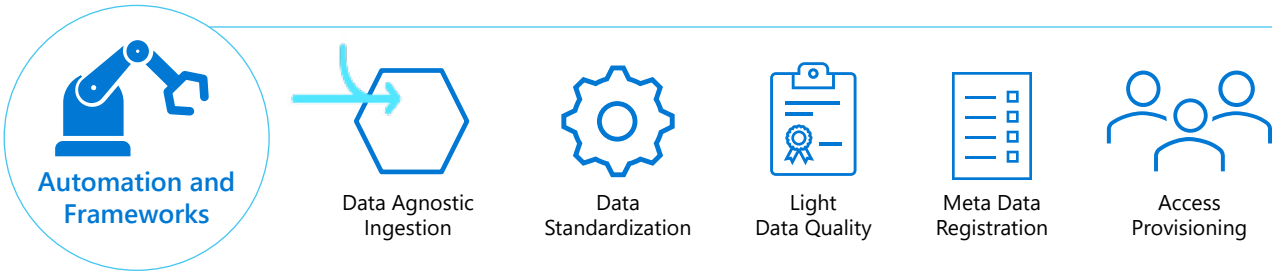



Figure 5: Automated data services

To implement foundational data management practices, the following aspects are necessary.

### Data-agnostic ingestion

Data-agnostic ingestion allows you to automatically ingest data regardless of its attribute, format, and the domain it belongs to. Organizations can automatically push or pull data from different sources then process it. Metadata-ingestion frameworks or Kafka-based solutions are sample solutions that can be implemented to automate this process. Several partners and solution providers have developed intellectual property (IP) in the form of solution accelerators that can be deployed in your organization.

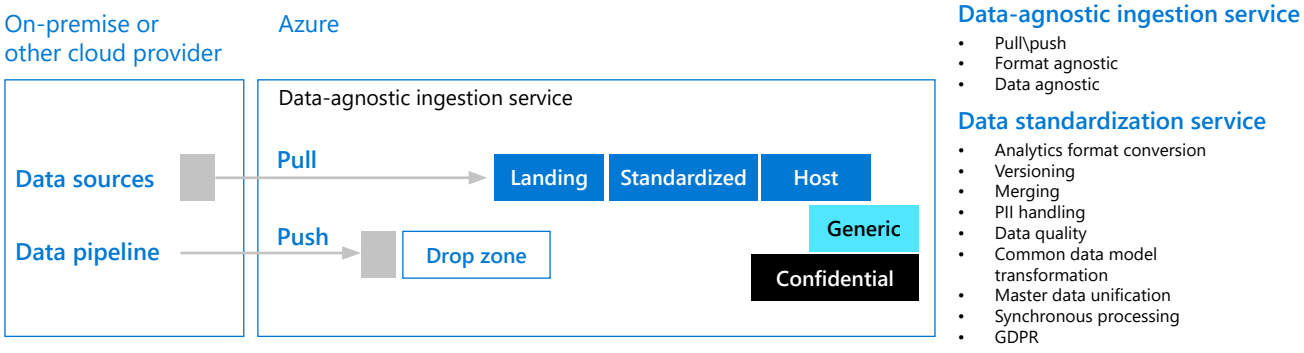



Figure 6: Data-agnostic ingestion

Implement an ingestion tool such as **Azure Data Factory** as part of the metadata ingestion framework. Learn more through **this documentation**. Alternatively, you can opt for a Kafka-based solution that provides a distributed publish-subscribe messaging system that can handle high volumes of data streams in real time or batch.

### Data standardization and quality

As the data gets ingested into the ecosystem, you can improve the efficiencies of your data operations by automating processes related to data standardization and quality. These processes include format conversions, versioning, merging, PII handing, and master data management. Additional services related to data quality management address issues such as deduplication, threshold identification, and alignment with master data. Without proper checks on data quality, you run the risk of slowing down time to insights.

> Use Spark notebooks in **Azure Synapse Analytics** or **Azure Databricks** to quickly implement data standardization practices.

### Metadata registration and access provisioning

As data is ingested and standardized in the storage foundation, the next step is to register these new data assets in the catalog to ensure they become instantly discoverable. As a safety net, data governance scheduled scanning of the data hub should register these new assets. Once data is provisioned in the ecosystem, another automated service can provision access according to the data classification of data being ingested.

## Domains and data products

Organizations are shifting their approach from running a center of excellence—where everything is centrally controlled—to being champions of enablement who use federated domains to provide LOB teams more control and autonomy.

**Domains:** The simplest definition of a data domain is a grouping of data aligned to a specific business function in an organization. As a starting point, organizations can create data domains aligned to their LOBs, such as finance, marketing, and operations, then and assign data stewards within each department. The goal is to empower LOB teams with self-serve access to data, analytical tools, and data services so they can develop data products for their own use.

**Data products:** A data product shapes and presents data in a form that is consumable and actionable by its target persona(s). A data product embodies all the attributes expected of a quality product for the contextual purpose(s) that it serves. The following are sample data products produced by sample personas within an organization:

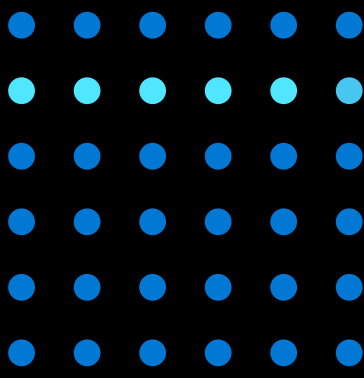
| Persona | Data products |
|---|---|
| Data Analyst<br>Data Scientist | **Quality Datasets** that eliminate the need to wrangle data and enable focus on creating insights and intelligence for applied use cases. |
| Application Developers | **Data and ML/AI APIs** to integrate data, insights, and intelligence (from datasets, analytics models, and ML/AI models) in applications. |
| Business Analyst / User / Decision Maker | **Metrics, Reports, Dashboards, Models,** Easy to use self-serve **Querying Interfaces,** and **Applications** that present data in contextual and actionable domain forms. |
| Compliance Office | Data compliance health metrics and actionable insights. |
| Security Office | Data security health metrics and actionable insights. |

Depending on the maturity of your organization and your analytics vision, you'll have to decide whether a federated or shared model of domains is the right approach.

> In a **shared** model, there is only one domain that is shared by multiple business units.
> In a **federated** model, there are multiple domains that may be aligned to business units, products, or data producers/consumers.

In practice, smaller companies often do not implement federation across multiple domains because the cost benefit will not be realized. Instead, taking a solutions-related approach that establishes enterprise data governance and data management foundation can help improve efficiencies. For medium to large organizations, there should be a plan to evolve the analytics ecosystem toward federation.

**Flexibility to evolve to a federated platform**. For organizations that start with a monolith domain that different LOBs use, there are patterns you can follow to evolve into a federated model.



Regardless of being shared or federated, the purpose of domains remains the same—to serve your LOBs. Domains should provide LOBs with end-to-end capabilities, including all the data and tools they need, to perform any type of analytics.

## Build templates for fast, repeatable domain deployment

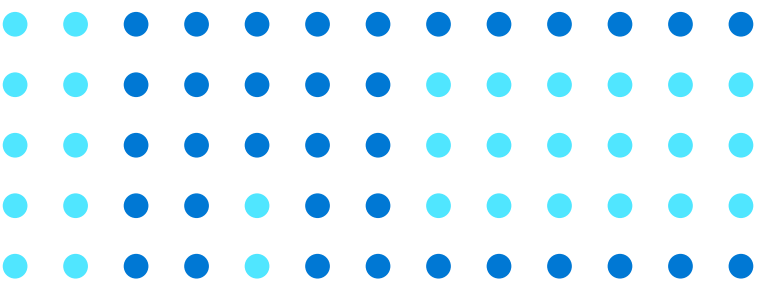
Domain teams need their own environments with end-to-end capabilities that allow them to integrate, orchestrate, engineer, and provision data products. To deploy these environments, data foundation teams use templates (infrastructure as a code) to make this process repeatable ensuring all the networking, security, vault, monitoring services, and other infrastructure services are deployed correctly. Templates can include platform-as-a-service (PaaS) and software-as-a-service (SaaS) solutions that provide essential services as well as other and components built in and ready to use.

Data foundation teams also deploy data services to enable domains teams to operate within their own environments. Data services that data foundation teams provide for domain users include:

> **Search for data:** The ability for domain users to explore the data catalog and find new data within the organization regardless of its location (on-premises or in any public cloud).
> **Request data access:** The ability for domain teams to get access to the data once it's in the data hub or data lake.
> **Self-serve data ingestion:** The ability to copy data from other sources to the data hub without having to create a data pipeline.

## Execute all analytics patterns in a domain

The benefit of implementing domains is that LOBs or domain teams are able to execute the full range of analytics patterns on their own—including but not limited to data engineering, real-time analytics, machine learning and data science, SQL-based analytics, and enterprise BI.



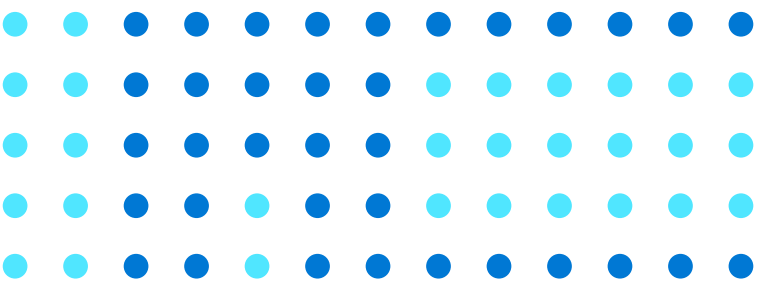
Domain teams use data from the data hub and associated data services to create their data products. The top four analytics patterns used to create data products are:

1. **Data exploration/onboarding:** This pattern helps domain teams look for new data they can leverage for new use cases. It usually starts with users browsing the data catalog and requesting access. Once the request is granted, data can be interrogated by running SQL scripts regardless of the data format (TSV, CSV, Parquet, Delta) in the storage foundation. If the data is deemed useful for the use case, it can then be used in the following patterns.

2. **Descriptive analytics:** This pattern allows teams to build data products like dashboards and reports. It makes it possible for data to be consumed by these products regardless of the data's current stage (integrated, aggregated, cleansed). Some of the data is processed in batch, near-real time, or real time, and all of it is made available within the domain.

3. **Artificial intelligence:** Another way to build data products is by implementing machine learning practices. This involves wrangling data from the data hub, training and deploying machine learning models, and then consuming the produced machine learning APIs for scoring data.

4. **Data exchange:** Once the data products are built, these can be shared across other domains or organizations. Within an organization, access can be requested for data products stored in the data lake. This is the simplest and most practical implementation of sharing data products within the organization. For cross-organization data sharing, multiple data sharing capabilities can be leveraged such as Azure Data Share and Databricks Delta Share.



# Realize the vision with Microsoft products and services

Bringing together elements from the three solution pillars of Modern Analytics and Governance at Scale provides a roadmap to achieving a unified analytics ecosystem. Microsoft offers services and products that align with essential components of each pillar to help you build a full end-to-end analytics platform. As with all Microsoft offerings, the open and extensible framework underpinning the platform helps you make the most of what you're using today.

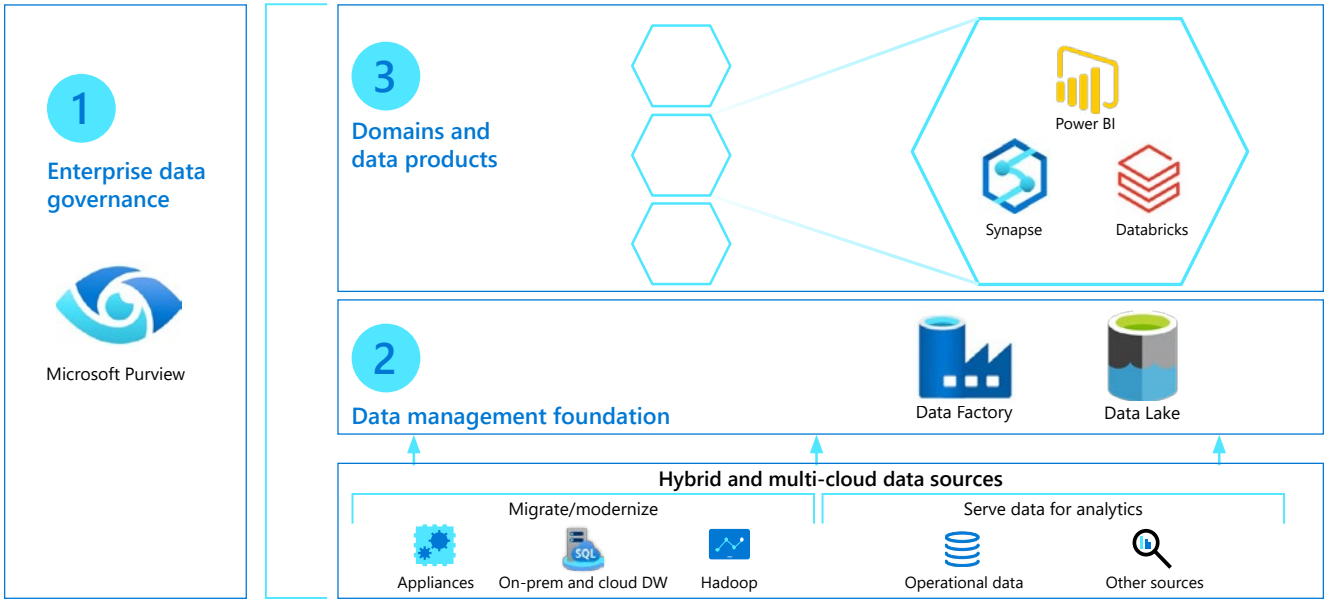



Figure 7: MAG at Scale supported by Microsoft products

**Microsoft Purview** provides a unified data governance solution to help manage and govern your on-premises, multi-cloud, and SaaS data.

**Azure Data Factory** enables hybrid data integration through a fully managed, serverless service.

**Azure Data Lake** includes all the capabilities required to make it easy for developers, data scientists, and analysts to store data of any size, shape, and speed, and do all types of processing and analytics across platforms and languages.

**Azure Synapse Analytics** accelerates time to insight across data warehouses and big data systems. It brings together the best of SQL technologies used in enterprise data warehousing, Apache Spark technologies for big data, and Azure Data Explorer for log and time series analytics.

**Azure Databricks** enables an open data lakehouse in Azure built on top of an open data lake to power a variety of analytical workloads while allowing for common governance across your entire data estate.

**Microsoft Power BI** is a unified, scalable platform for self-service and enterprise business intelligence.

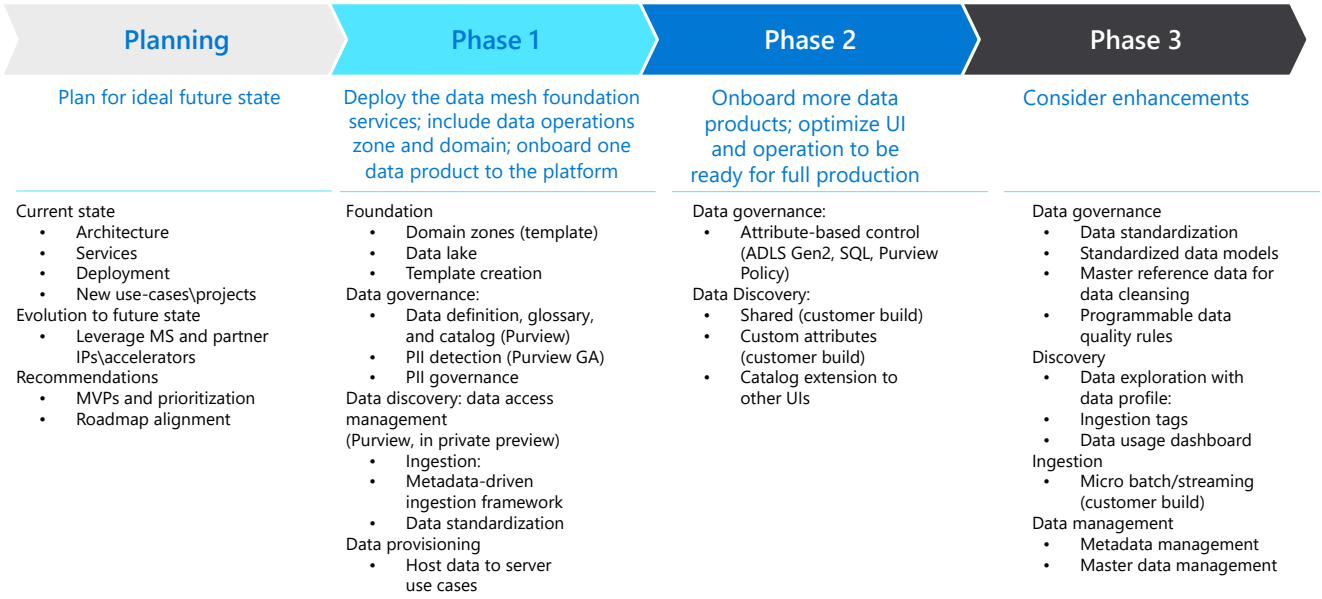
# Four phases for implementation

When your organization is ready to start implementing MAG at Scale, Microsoft has developed common implementation patterns through its work deploying multitudes of these projects with organizations globally. The first phases, Planning and Phase 1, create the foundation for transformation in each area. Continuing to work through Phase 2 and Phase 3 encourages ongoing iteration and exploration as your team learns and your needs evolve.

**Planning:** In the planning phase, a Microsoft partner helps assess the current state of your environment (architecture, services, deployment framework, etc.) and builds a plan for evolving to the future state. The plan includes a series of projects. Each project involves deploying solutions to solve well-defined use cases that have tangible business outcomes (ex. supply chain analytics). As each use case is addressed, your organization gains foundational wins and moves closer to realizing the full value of MAG at Scale.

**Phase 1, 2, and 3:** Phase 1 focuses on deploying different minimum viable products (MVPs) either sequentially or in parallel, depending on your timelines. The first MVP typically relates to the foundation (domains and storage). The second MVP tends to tackle data governance issues like data maps, cataloging, classification, and policies. The third MVP focuses on automating data services and implementing data-agnostic ingestion, data standardization, and light data quality. The succeeding phases provide enhancements and deployment of new capabilities. Microsoft has recommended solutions packaged in MAG at Scale that serve as the North Star for implementation.

Whether you want to implement 100+ projects at once or are starting off with a small-scale objective such as one domain or business issue, the implementation stages can flex to meet your needs.



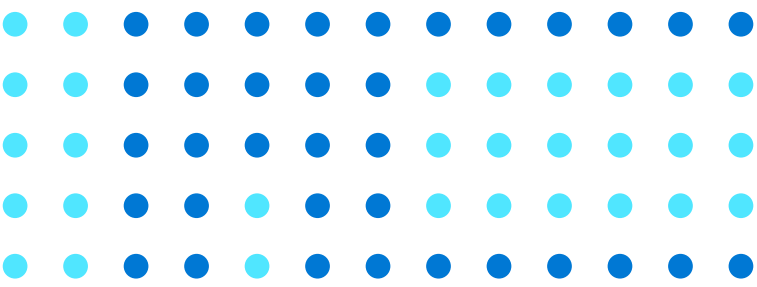
# Conclusion

As the world of data and analytics continues to exceed all perceived limits of speed, intelligence, and business use cases, organizations must find a way to democratize insights across LOB teams as quickly and easily as possible to stay competitive. The traditional approach to building an analytics ecosystem, which relied heavily on adopting now and integrating later, is getting far too complex to sustain and it's too costly from a time and talent perspective in the race to real-time insights.

The approach defined in this whitepaper shifts the paradigm many organizations follow today and offers a more sustainable and unified approach. It starts by stepping back to better understand the people, processes, and culture that make up the ecosystem and using those insights, along with desired business goals and experiences as the North star for moving forward.

## Reach out to learn more

Interested in learning more about the Composable Solution described above or other components of the Microsoft packaged solution such as enterprise data strategy, technical architectures, IP and accelerators, or the implementation stages? Reach out to an **Azure sales specialist** or your Microsoft sales representative for best practices on analytics, help setting up a proof of concept, and more. Or ask about visiting a local Executive Briefing Center or Microsoft Training Center.



Microsoft Azure