![Neudesic, an IBM Company logo]

Reimagining data governance for a modern enterprise

Achieve unified data governance with Microsoft Azure Purview and Neudesic, an IBM Company

Executive Summary

Across all industries, companies are looking to tap into their data assets and drive intelligent decision making across the organization. At the same time, these companies are also rushing down the digital transformation path and modernizing their IT infrastructure.

This is putting an increasing pressure on the information management stakeholders, who are now having to deal with more volumes of data, more data-driven business priorities, more compliance requirements and a more complex IT environment than ever before.

According to IDC, 175 ZB zettabytes of data will be generated annually by 2025, 300% more than what was generated in 2021. If the current trends are anything to go by, this data would span the edge, multiple clouds, and the data center, and would have to be managed across both operating and analytical systems. Meanwhile, data privacy and compliance regulations would become even more demanding due to growing security concerns.

Most organizations just don't have the right infrastructure to ensure better discovery, trust, and protection of this data. Their existing data governance systems are either system-specific, or use case specific, or limited in lineage capabilities, or are not scalable.

This book shares how building your data governance foundation on Azure Purview, a unified data governance service from Microsoft, enables your data consumers to easily find valuable, trustworthy data across your on-premises, multicloud, and software as a service (SaaS) data sources. It also shares how you can take the next step in this journey with IBM Consulting and Neudesic.

Table of Contents

Trusted data is key to digital transformation

Across industries, organizations are experiencing an era of rapid change and the greatest structural transformation in our economy in a generation. Every organization today, large or small, from manufacturing and retail to education and healthcare, is accelerating its digital initiatives. At the heart of this transformation lies data and the ability to generate analytical and predictive insights from it.

But in order for this data to fuel digital transformation, organizations need to ensure that the right people are able to easily find and use this data, while staying on the right side of regulation.

So, essentially organizations must solve two problems at the same time. First, the data must be timely, because digital transformation is all about accelerating time-to-market – whether that's delivering personalized experiences to customers to providing real-time insights to businesses. Second, while speed is critical, it is not enough. Each stakeholder should be able to trust the data, be it data producers, data consumers, or compliance officers.

Data is expected to grow by 400% from 44 ZB in 2020 to 175 ZB in 2025.[1]

45% of enterprise data could be described as 'dark data', i.e. data unused for analytics initiatives.[2]

According to Forrester, only 40% of CIOs are delivering results against the speed required.[3]

Speed and trust often compete with each other and it has always been common for organizations to focus on one or the other.

Today, more than ever, it has become critical for organizations to overcome these traditional data challenges.

Sources:
1. IDC Whitepaper, sponsored by Seagate, The Digitization of the World: From Edge to Core, 2018
2. Voice of the Enterprise: Data & Analytics, Data Platforms, 451 Research, S&P Global Market Intelligence, 2021
3. Predictions 2019: Transformation Goes Pragmatic, Forrester, 2019

Time to act on data governance

Until now, data management and data governance have been highly manual tasks across siloed tools. As organizations set out to enhance discoverability of data to the right users, while ensuring its responsible use, they will need to fundamentally reimagine data governance for the modern IT ecosystem.

## Data privacy and protection is becoming more complex

Data privacy as well as the growing number of data breaches has made data protection a top boardroom priority. Further, there is a need to remain compliant with growing and stricter regulatory legislations such as European Union General Data Protection Regulation (GDPR) or the California Consumer Privacy Act (CCPA) or Canada's Personal Information Protection and Electronic Documents Act.

The number of data users are increasing

According to 451 Research by S&P Global Market Intelligence, organizations enabling departmental/LOB decision-makers to access data marketplaces will increase from 21% to 27% in 2 Years.[1] This puts even more pressure on IT and data teams to ensure broader data discoverability, while maintaining compliance.

The data environment is becoming more complex

For most companies, the acceleration of digital transformation is resulting in increasingly complex IT operating environments. Today, most companies have created an operating environment that spans the edge, multiple clouds and the data center. The result is that companies are now dealing with a hybrid environment with data in multiple different data stores that are scattered across all of this landscape.
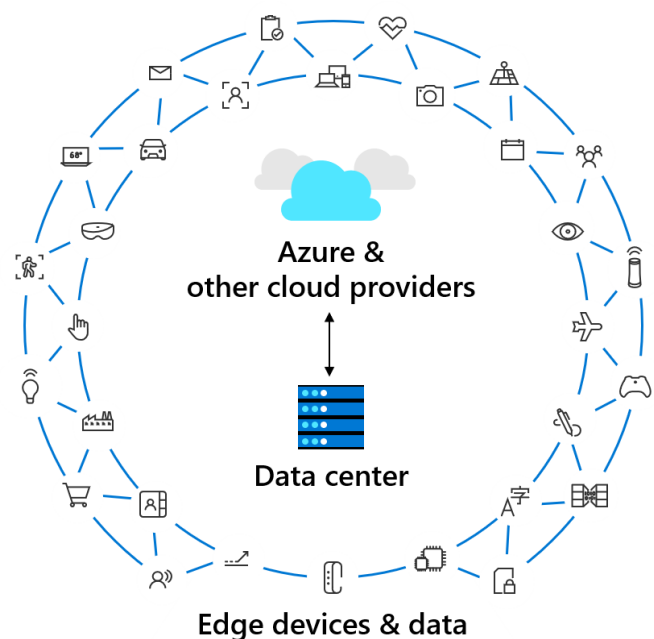
The distributed data estate

Reimagine your approach with a holistic data governance framework

For organizations to unlock the value of data, technology is key but not enough. Data governance should become a cross-functional discipline, orchestrating people, process, and technology. This allows organizations to be comprehensive, consistent, and coherent in the way they define, discuss, analyze and leverage their data as information. The topics below are the disciplines that need to be addressed in a Data Governance Program that may span one or multiple years. IBM Consulting's and Neudesic's Data Governance Framework provides a holistic approach, helping you to implement the artifacts for each discipline in a structured and complete way.

Architecture

How well does the organization design, develop, deploy, and manage data architecture?

Business Rules Management

How well does the organization define, develop, and deploy, manage, the application of business rules across the enterprise?

Change Management

Are capabilities in place to adopt required organizational change and drive business value realization?
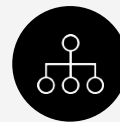
Data Quality

Is the organization able to consistently define, and measure data quality, and mitigate and data quality issues?

Metadata

How well does the organization capture, manage, and access business, technology, and operational information on key corporate data?

Organization & Stewardship

Are there formal organizations and roles for supporting, managing, and improving governance processes and capabilities?

Policy

How well does the organization define and manage organizational behavior using policy?

Privacy & Security

Are appropriate considerations in place for protections of customer privacy and data security?

Regulation & Compliance

Is the organization correctly prioritizing activities to address people, process, and technology requirements for regulatory and compliance issues?

Govern data across the entire data estate with Azure Purview

Azure Purview is a unified data governance solution that gives you a holistic, up-to-date map of all your on-premises, multicloud, and software as a service (SaaS) data and helps you establish governance at every step of the data lifecycle. This enables you to spend more time on unlocking the value from your data rather than spending time on integrating and managing infrastructure.

Data Map
Automate and manage metadata at scale

Data Catalog
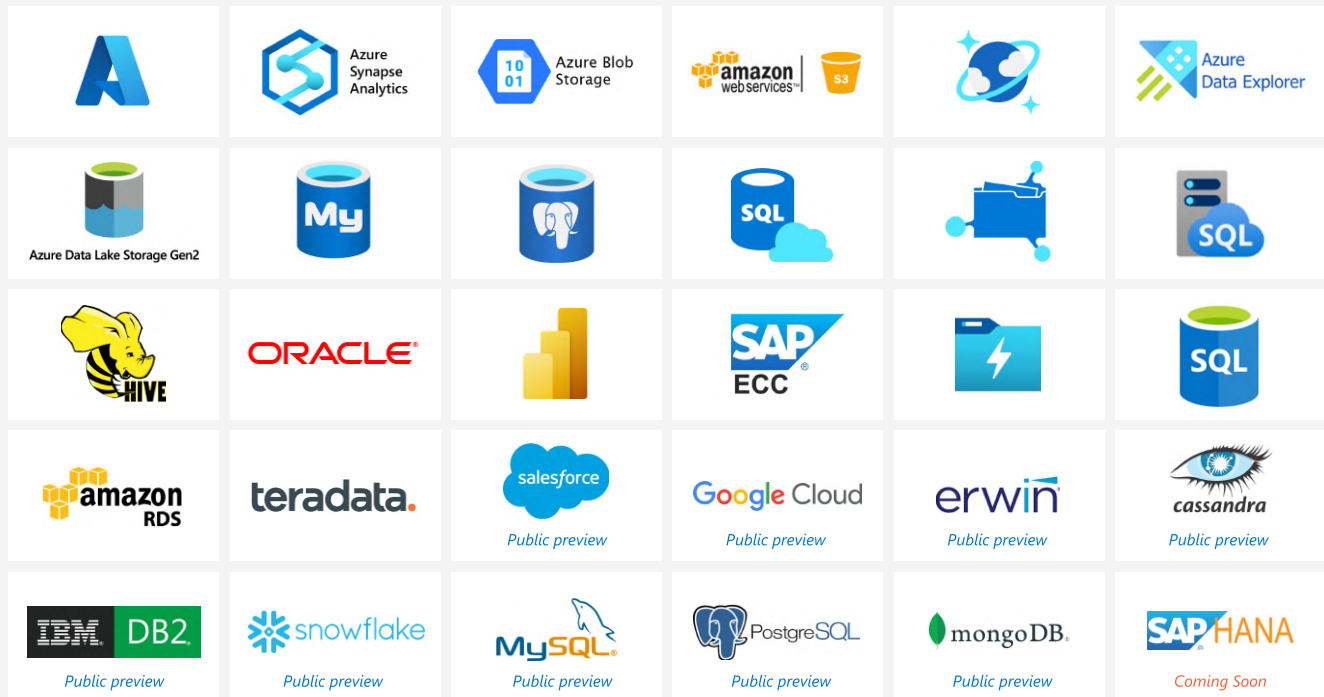Enable effortless discovery for data consumers

Data Insights
Assess data usage across your organization

Azure Purview

Data Insights

Data Catalog

On-prem

Data Map

Multicloud

Saas

01　Automate data discovery

Out of the box support for 35 data sources and growing...[1]

The platform automates today's highly manual data discovery processes by providing data scanning as a fully managed service. This enables data engineers to scan petabyte-sized data estates across on-premises and multi-cloud data sources, to build a data map, that captures key pieces of information such as technical metadata, semantic metadata, and business metadata.

As of today, Purview supports native support for 35 data sources covering on-premises, cloud and SaaS sources, with others in Preview. In addition, Purview provides you the capability to integrate the non-native data systems using the Apache Atlas APIs. At IBM Consulting and Neudesic, we are also developing "Custom Scanners" to collect characteristics of these non-native sources to make it easier for you to integrate your data estate in Azure Purview.

# 02   Make data easily discoverable with Purview Data Catalog
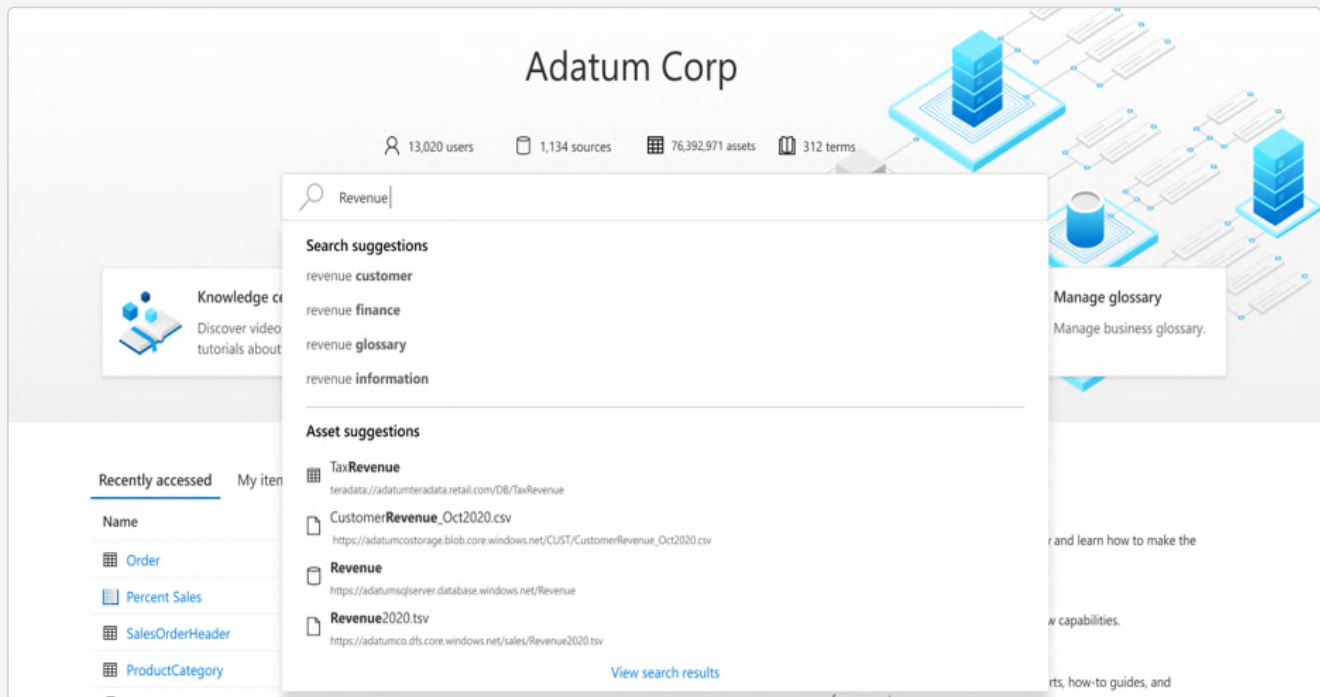
Azure Purview Data Catalog with Semantic Search

Most data stewards or data scientists don't know how their data estate is structured. Purview provides a browser experience with powerful data search capabilities, empowering all data consumers to easily find trusted data.

Azure Purview also provides an enterprise-grade integrated business glossary. This allows your subject matter experts to provide the business context associated with their data assets and empowers business users to easily discover data assets using the language and terms that are most familiar to them.

## 03 Trace lineage of data assets across the data estate
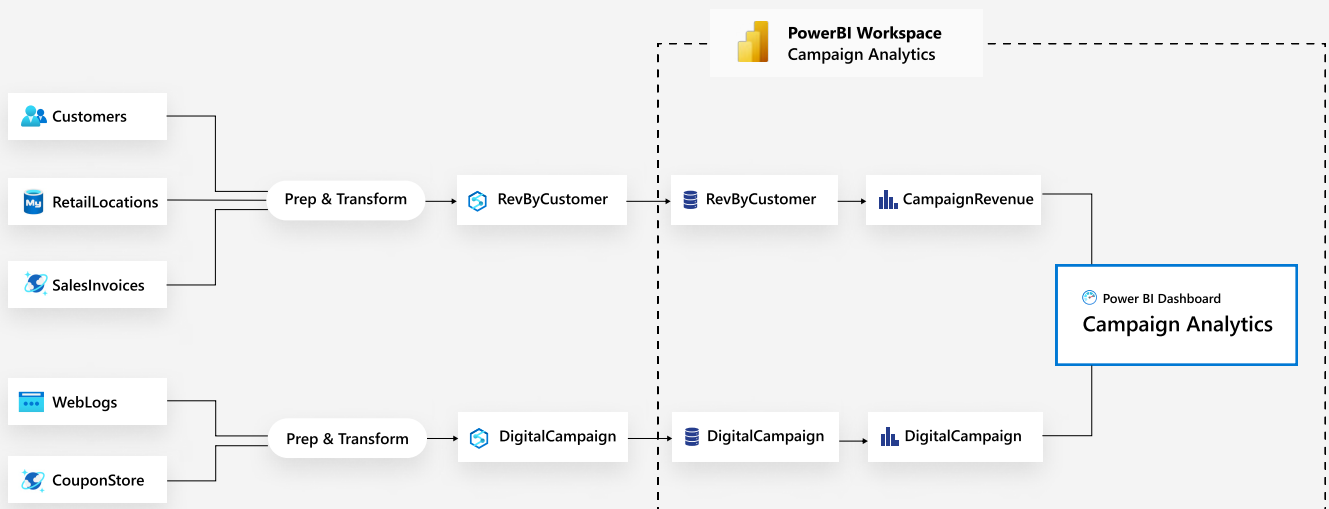
Your data estate may include systems doing data extraction, transformation (ETL/ELT systems), analytics, and visualization systems. Each of the systems captures rich static and operational metadata that describes the state and quality of the data within the systems boundary. For your entire data teams, be it data analyst, data engineer, data scientist or data officer, clearly understanding data lineages can provide critical insights into your data and simplify regulatory compliance.

Purview automates extraction of the movement, transformation, and operational metadata from each data system at the lowest grain possible. Parallelly, many of Microsoft data products*, such as Azure Data Factory, Power BI and Synapse pipelines also automatically report data lineage to the data map. The data engineers can easily run the data pipelines and verify if lineage is complete and not broken due to absence of metadata of components or links between components.

Note: Purview does not yet support lineage for Azure Databricks. However, IBM Consulting and Neudesic can create a custom lineage for when Databricks is used for transformation & propagation of data (ETL).

04 Classify data at scale
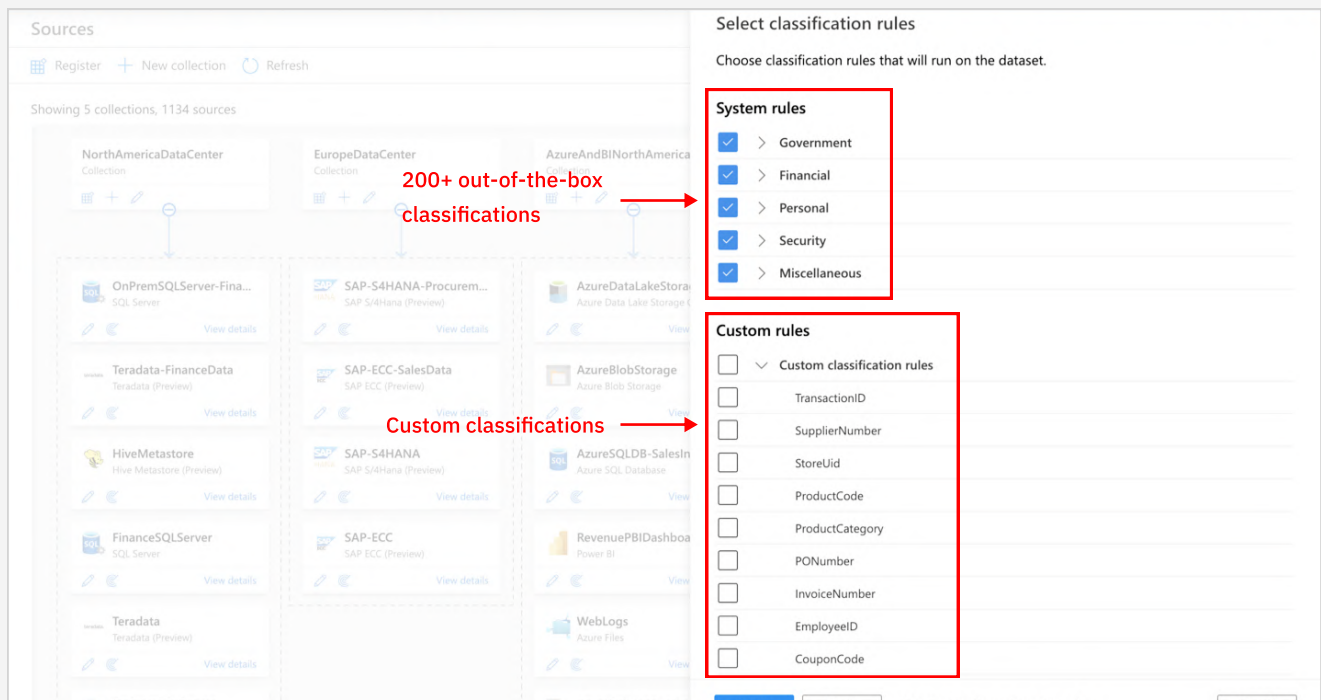
Classifications in Azure Purview

Azure Purview doesn't just automate data discovery. It also provides data classification as a fully managed service. There are 200+ built-in classifiers to help data stewards and data engineers to automate the classification of data to specify sensitivity, compliance, industry, business and company-specific value.

For example, if there is a file named multiple.docx and it has a National ID number in its content, Azure Purview automatically adds the classification 'EU National Identification Number' to the file asset's detail page.

In some scenarios, the data or compliance teams might want to manually add data classifications. Purview also enables them to create these custom classifiers on a file, table, or column asset level.

05 Get a bird's eye-view of sensitive data
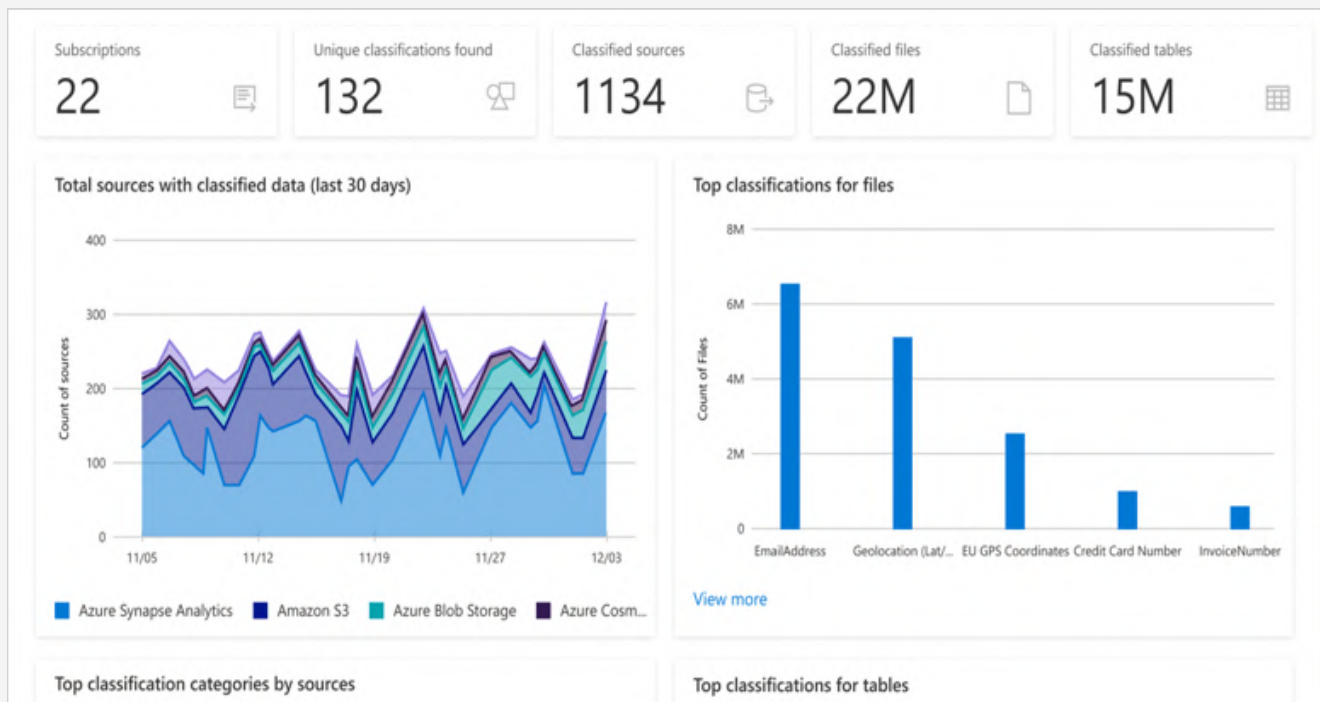
Azure Purview Data Insights

Finally, Purview Data Insights provides a holistic understanding of your data estate, along with the ability to drill down so you can pinpoint issues and maximize the compliant use of data.

This application aims to provide specific insights to the data source administrators, business users, data stewards, data officer and, security administrators.

Asset insights: Distribution of your data estate by source type, by classification and by file size, as some of the dimensions.

Scan insights: Overall health of the scans - how many succeeded, how many failed, how many canceled.

Glossary insights: Distribution of glossary terms by status: how many glossary terms are attached to assets and those not yet attached.

Classification insights: Classifications found during a scan, location of classified data, and a drill-down to the classified files themselves.

Sensitivity labeling insights: Details about the sensitivity labels found during a scan, as well as a drill-down to the labeled files themselves.

Accelerate your journey with IBM Consulting's and Neudesic's additional tools and capabilities

Methodology
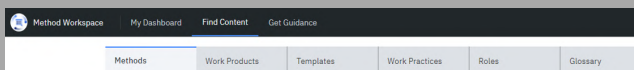
A method workspace to use a structured approach for all deliverables in a project. Enables content customization based on applicable Growth Platform & Service Line.

Ideation and Agile

An end-to-end model to chart your transformation journey with an iterative framework and equip you with the experts, practices, and technologies to bring that journey to life.
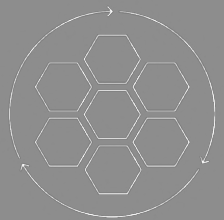
Frameworks

Organizational and technical approaches to drive consistency and implement best practices across the end-to-end data governance lifecycle:

- Data Governance Framework
- Data Quality Framework
- Data Catalog Use Cases

Architectures

Technical approaches to meet both your current as well as future data governance needs:

- Reference Platform Architecture
- Metadata Architecture; patterns and integration and exchange of MTD between various Data Management components

Add-ons

1. Data Quality

Capability to check data integrity based on your project specific data quality rules as input

2. Master Data Management (via Profisee)

Out-of-the-box tooling to ensure alignment to standards for master data

3. Custom lineage for Azure Databricks

4. Automation via Atlas

Automations to upload glossary terms, build custom lineages, and import sensitivy labels

5. Cognitive Classifier

Algorithms to map source data attributes and assess classifications
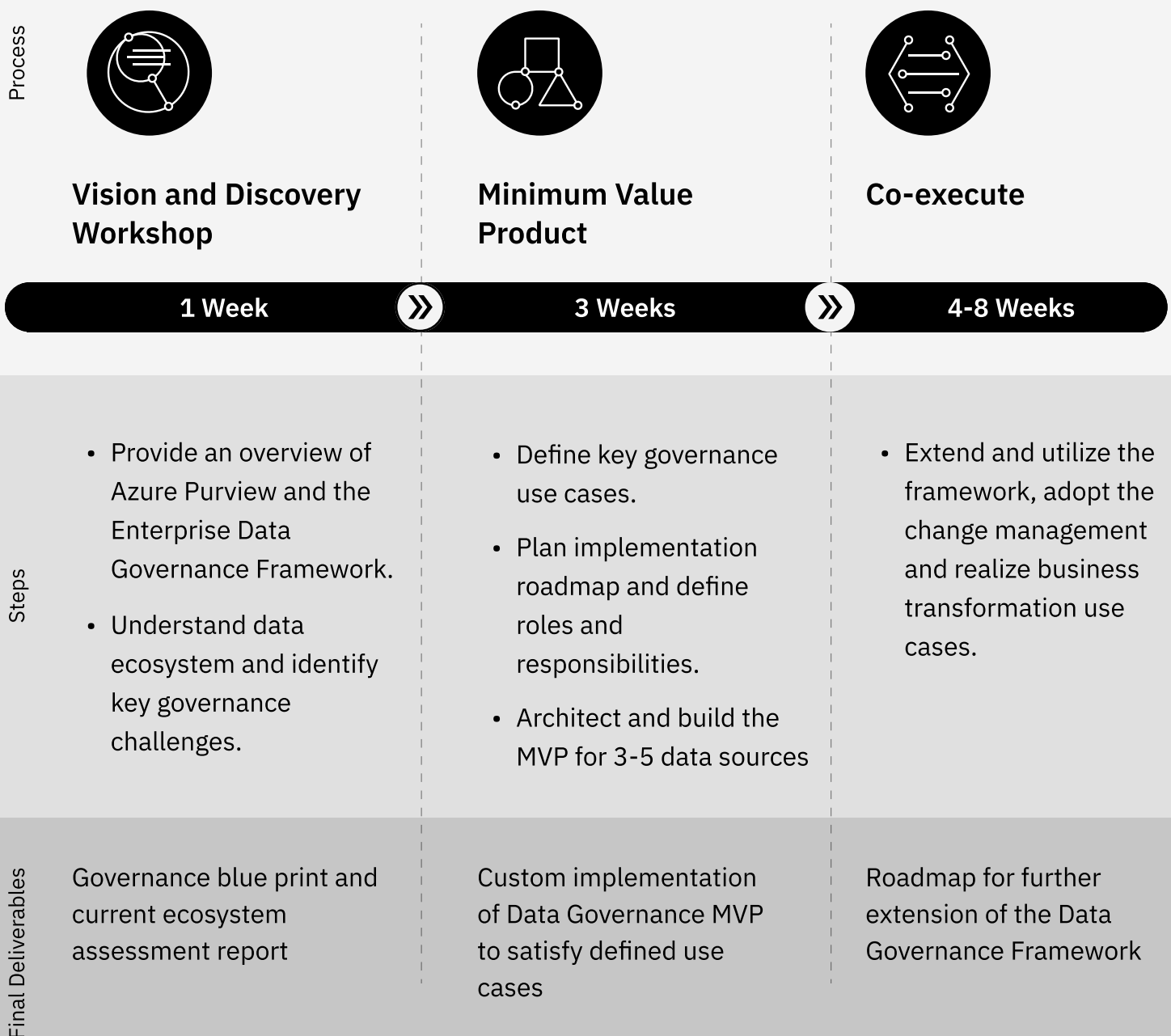
6. Custom scanners for data sources not available out of box from Purview

7. Customized scanners for selective data extraction and cost savings

Take the next step to better discover, understand, and trust your data

Gain an understanding of IBM Consulting's and Neudesic's framework and technology solution for all three aspects (People, Process and Technology) of your Data Governance journey, get a thorough demonstration in your environment by starting with 3-5 data sources, and implement the framework at scale, with Azure Purview as the key technology component.

Our delivery process

Process

Vision and Discovery Workshop	**Minimum Value Product**	**Co-execute**
1 Week ≫	3 Weeks ≫	4-8 Weeks

Steps

- Provide an overview of Azure Purview and the Enterprise Data Governance Framework.

- Understand data ecosystem and identify key governance challenges.

- Define key governance use cases.

- Plan implementation roadmap and define roles and responsibilities.

- Architect and build the MVP for 3-5 data sources

- Extend and utilize the framework, adopt the change management and realize business transformation use cases.

Final Deliverables

Governance blue print and current ecosystem assessment report

Custom implementation of Data Governance MVP to satisfy defined use cases

Roadmap for further extension of the Data Governance Framework

Case study: Automating data discovery and enhancing data governance for Grundfos

Situation

Founded in 1945, Grundfos manufactures pumps that provide critical services such as heating, cooling, and clean water to people around the world. Over time, the company started using Interent of Things (IoT) solutions to add value.

As data volumes grew, the company's data became hard to track and manage across the hybrid environment. To find data, Grundfos employees had to know that it existed in the first place, then identify the right colleague and send an email or knock on their office door to ask for the information.

Solution

To drive better data discoverability and governance, Grundfos implemented Azure Purview and integrated it with SAP HANA environment and with other Microsoft Data sources, including Azure Synapse Analytics, Power BI, Azure Data Lake Storage, and SQL Server.

Impact

By implementing Azure Purview, Grundfos is reducing cumbersome manual tasks, boosting productivity, simplifying compliance, and making it fast and easy for employees to gain full value from the company's data.

"Adopting Azure Purview helps us free up time for our resources to be used for more purposeful, proactive work. We'll accelerate our ability to track and interpret a wide range of data, which we'll use to create solutions at scale that truly help improve the environment and people's lives."

Thomas Asger Hansen,
Senior Manager, Enterprise Data and AI, Grundfos

Contact Us

Lisa Schnitzius

Offering Leader, IBM Consulting

lisa.schnitzius@ibm.com

Tyler Suss

Director, Data & AI Solutions Marketing & Alliances, Neudesic, an IBM Company

tyler.suss@neudesic.com

Why IBM Consulting and Neudesic

Global Practice focused on Azure Data, AI, and IoT

Microsoft's fastest growing GSI

2,000+ Certified Azure Data and AI Specialists

1,000+ Cloud Migrations in the past year

30 global centers, 55 languages, 24x7 operations

17 Microsoft Gold Competencies

Analytics on Azure Advanced Specialization

AI and Maching Learning on Azure Advanced Specialization

neudesic
an IBM Company