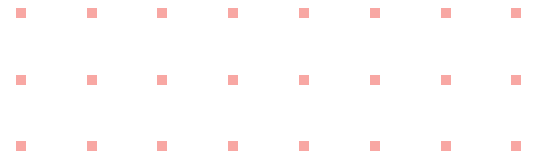
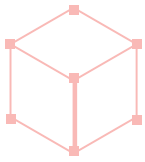


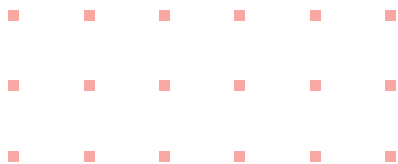
NCS DATA FABRIC WITH DENODO





Contents

Introduction	3
Complex Data Landscapes and the Modern Enterprise	4
Hybrid and Multi-Cloud Systems	4
The Complexity of Data Governance	4
What is Data Fabric	5
The Key Elements of Data Fabric	6
Data Access	6
Managing the Data Lifecycle	7
Data Governance and Privacy	7
Compliance	7
Exposing Data	7
Data Fabric and the Denodo Platform	8
An Overall Solution Architecture	11
Conclusion	13

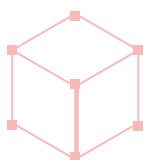




Introduction

Today's data-driven leadership thrives on how easily and quickly businesses can get data to derive meaningful insights, enabling them to respond to their dynamic ecosystem rapidly. A solution to address challenges such as complex data integration, lifecycle management, and data exposure is the necessity of the hour due to high data velocity, veracity, volume, and variety.

Data fabric is an approach to abstract such complexity and provide a mechanism for an enterprise to achieve its goals by making relevant data available to drive relevant analytics at the point of decision making.



Complex Data Landscapes and the Modern Enterprise

Organizations today span multiple countries and even continents. Also, their source data is generated by multiple departments with department-specific processes and IT systems.

Almost every company that is multi-departmental and/or spread across more than one country replicates its data multiple times, increasing cost, affecting data lineage and quality, and incurring risks that violate regional data protection rules.

Many departments within mid-to-large scale organizations use their own tools and systems to access and process the same information, which leads to inconsistent and multiple definitions of KPIs, information inconsistencies, and missing or incomplete data lineage information.

Overall, an enterprise may never use 60-70% of its data, which affects the organization's ability to make the optimum data driven decisions.

Hybrid and Multi-Cloud Systems

Increasingly, enterprises with complex data landscapes are considering the cloud as a way to extend and simplify their data analytics. Given the vast amount of data, cloud systems are used selectively as extensions, typically to drive newer use cases. Cloud systems may host newer sources of external and internal data and may also selectively copy data from on-premises systems, which is needed for common use cases that are run in the cloud.

Since multiple cloud vendors are going neck-to-neck for client business, enterprises soon have to manage multi-cloud environments with different data assets and use cases running on different cloud systems.

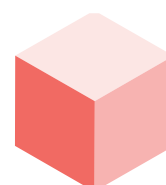
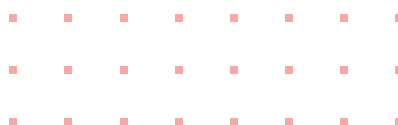
A wholesale data copy across cloud and on-premises systems is never the preferred option, due to the time it takes, the impact to data quality and lineage, the potential security and regulatory issues, and data egress charges from cloud providers.

With too many siloed tools and disparate data access points spread across systems, it becomes impossible to lay out proper data access, governance, and lifecycle management features for users.

The Complexity of Data Governance

As companies grow and datasets diversify, the number of on-premises and cloud tools and systems proliferate. Each siloed data source, or consuming system, has its own data governance and data security mechanism, and none are shared across the enterprise.

These are not easy challenges to address using traditional methods for data delivery.





What is Data Fabric

“A data fabric is a flexible, scalable, and powerful data management approach that is critical for any digital transformation, data governance, or analytics initiative. With a data fabric, enterprises can integrate, manage, and deliver their distributed data to any user, in real time, regardless of the location and format of the source data.”

A data fabric is an **architecture** and **set of data services** that provide **consistent capabilities** across a **choice of data endpoints** spanning hybrid multi-cloud environments.

It is a powerful architecture that standardizes data management practices and practicalities across cloud systems, on-premises systems, and sometimes even edge devices. Among the many advantages that a data fabric affords, seamless data connectivity, the elimination of data copies, data visibility and insights, data access and control, data protection, and security quickly rise to the top.

Data fabric can be considered an enterprise data-access layer in which different data sources can be connected to provide data services to various consumers with consistent data management, security, and governance capabilities, while abstracting all data consumers from the complexities of access, including where and how the data is stored.



The Key Elements of Data Fabric

Data Access

Enterprises today have complex data estates which have evolved in the following way:

1

First Generation

Initially, every enterprise started with core applications such as enterprise resource planning (ERP), customer relationship management (CRM), and supply chain management (SCM). Such applications held important data, and each application had its own reporting and analytics systems that were built into the applications themselves. However, any cross-analysis among the data silos of such systems required this data to be copied to an enterprise data warehouse (EDW).

2

Second Generation

EDWs were built using relational database technology, which is effective for transaction processing or online transaction processing (OLTP) workloads. Subsequently, analytics processing systems were introduced for online analytical processing (OLAP) workloads. Some of those systems utilized massively parallel processing (MPP) capabilities under the hood, to provide optimum performance for analytics workloads. However, such systems are expensive and only the most relevant data was copied over to them – otherwise also known as “hot data.”

3

Third Generation

However, a lot of remaining data was deemed valuable from an analytics perspective, and organizations needed a cheap, scalable system that could hold both the “warm” and the “cold” data. These systems were the big data or data lake systems. For a while, many medium to large enterprises adopted them and started dumping data into them. Such systems facilitate this due to their “schema-on-read” orientations, they also provide file-based storage, which made data copy easy. Soon, these systems grew full of data that most had no visibility into. Today, a lot of data still lives in such systems.

4

Fourth Generation

Subsequently, enterprises also adopted many software as a service (SaaS) applications and multiple cloud providers, which created additional data silos.

Data fabric is designed to free data from the confines of such silos, offering seamless connectivity to multiple systems while eliminating or minimizing data copies. One way to achieve this is via data virtualization, which enables seamless, real-time access to data across disparate systems without replicating data, minimizing the number data copies and data movement.



Managing the Data Lifecycle

As data becomes more available via data fabric, the data lifecycle needs to be managed. This includes discovering data via a data catalog, managing metadata as well as data access, maintaining privacy by masking, tokenization, and maintaining an overall posture that is consistent with local data compliance rules.

Data Governance and Privacy



Metadata and Data Exploration

Data fabric holds a significant amount of metadata, and it is critical to govern access to this metadata as well as access to the actual data itself. Every user should have access to only the necessary metadata and data, and nothing more.



Data Catalogs

As data is made available in data fabric, it becomes essential to be able to search for, identify, and access objects. Data catalogs act as a central “watering hole” where all users can go to find the data objects they need for their work.



Governed Access

An enterprise can decide to make certain metadata accessible to all users, and users can send requests for certain objects which then can be approved by data owners. The data owners can then provide access to the actual data in the data fabric. This provides a ready mechanism with which to govern access specifically by data owners.



Privacy

Data fabric also provides a mechanism to hide, tokenize, or mask columns or hide rows of data based on the user’s role. This way, only the data that matches privacy constraints is shown to any user.

Compliance

With governed access and the privacy capabilities of data fabric, it is easier for enterprises to enforce regulatory requirements and remain in compliance with regulations such as the Personal Data Privacy Act of 1974 (PDPA), the Global Data Protection Regulation (GDPR), and the Health Insurance Portability and Accountability Act (HIPAA).

Exposing Data

Once data access is established, via mechanisms such as role-based access control (RBAC) or via the explicit granting of access to specific objects, the next capability is for users to explore the available data. If a specific view is not available, users should be able to create it via data self-service features.

Similarly, users can leverage self-service capabilities to publish data as APIs for integration with downstream applications.

This shifts the data manipulation capability into the hands of users rather than backend IT or data engineering teams. Business users no longer have to request that certain data be made available and then have to wait, sometimes for weeks, until it is published by backend teams.

This makes data fabric an ideal candidate to go hand-in-hand with self-service business intelligence (BI) and self-service data analytics.





Data Fabric and the Denodo Platform

Denodo tackles the growing challenges of data governance, security, and compliance by provisioning a well-controlled logical data management framework.

Denodo takes a logical approach that is agile and flexible. It has proven to be very successful, particularly as organisations establish data stewards with proper ownership and accountability over their corporate data assets.

With a logical layer across of all important data, organizations can manage who can access what data, when, and how. Organizations have full visibility into how data is being used and a full audit trail of data access.

The Denodo Platform is powered by data virtualization, to offer comprehensive data and metadata discovery and management capabilities including data governance, data lineage, and change impact analysis. Data virtualization enables organizations to create central data access, data governance, and security policies across heterogeneous systems of structured and unstructured data sources.

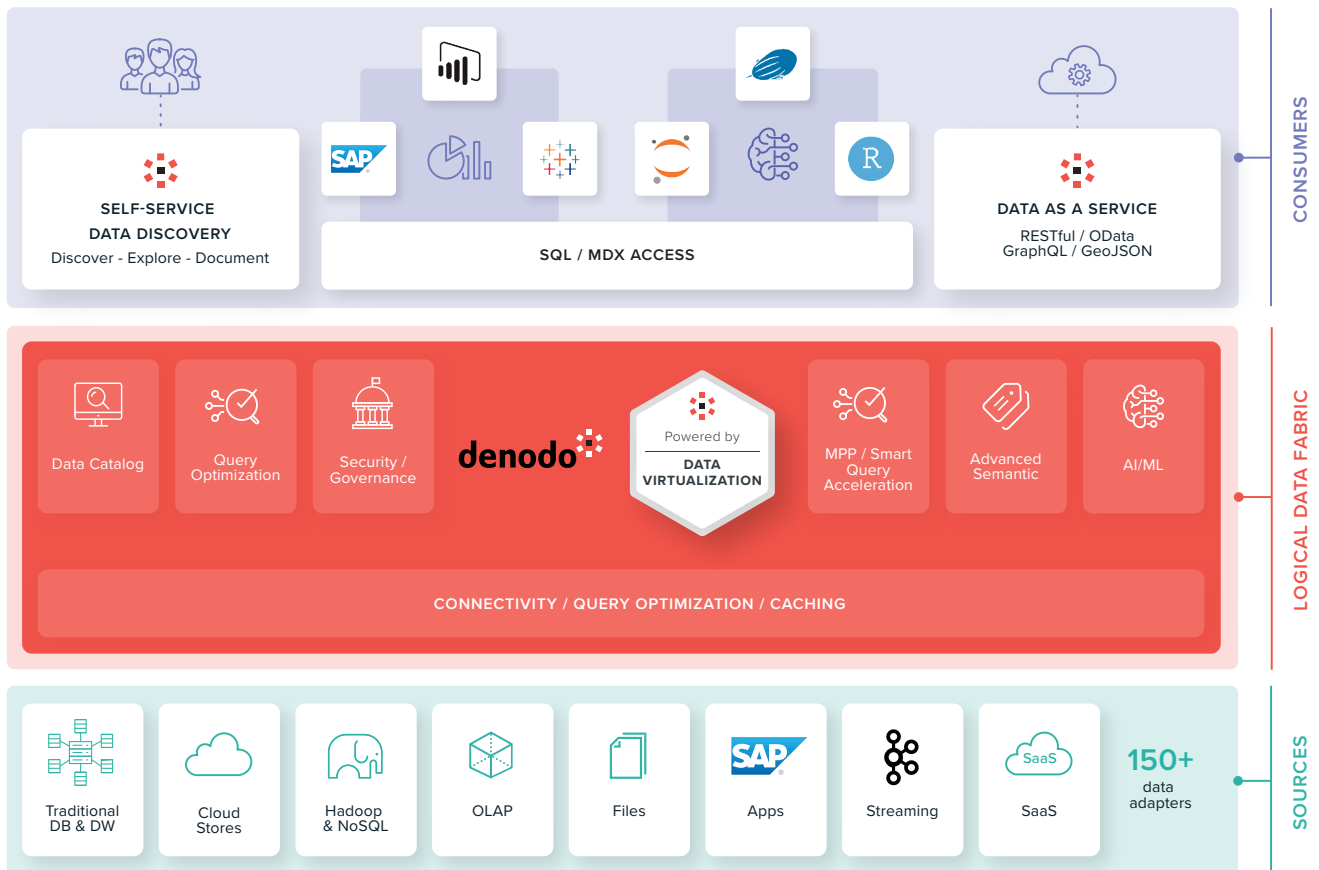
Whether some of the data sources or consuming applications are spread across geographies or divided between on-premises and cloud systems, the Denodo Platform seamlessly facilitates the central control of data governance and security.



What is Data Virtualization?

Data virtualization is a style of data integration and delivery. Most data integration solutions move a copy of the data from disparate sources into a new, consolidated source system, but data virtualization offers a completely different approach. It's a logical approach that can provide significant efficiencies.

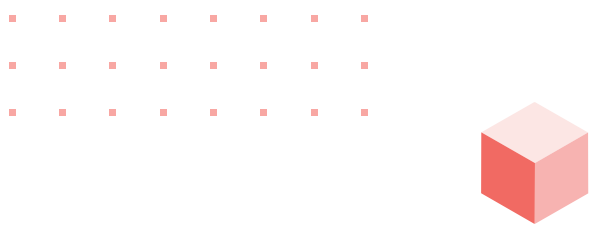
Rather than moving data, data virtualization provides a view of the integrated data, in real time, leaving the source data exactly where it is. This means that companies do not have to pay the costs of moving and housing the data, and yet they still gain all of the benefits of data integration.



Using virtualized views, the complexities of accessing data are taken away from the data consumer. These complexities are defined and managed in a single place, the logical layer. This means that business users can easily find data, understand its content, understand its currency, its lineage and its quality level and validity before using the data for their specific business purposes.

Data virtualization is also data and technology agnostic. The Denodo Platform will work with practically all data sources and deliver the required results through any tool. Data is consumed in many different ways, so making sure that data is delivered consistently and accurately is imperative for businesses.

Data virtualization is a straightforward way to modernize an organization's data architecture, leverage its existing technology investments, and meet ever-changing business demands.



What benefits does this approach provide?

Because data virtualization is deployed as a layer above an organization's disparate sources, serving as a single enterprise-wide data access layer, it can enable significant benefits, such as:

Streamlined data security, privacy, and governance

Protocols can be administered across the disparate sources from a single point of control.

Data ownership and trust

Data access control can also be managed from one single point of control. Data can be categorised, tagged, endorsed, and recommended. This builds trust back into data.

Simplified data sharing, data collaboration, and data recommendations

Data can be easily identified by such factors as data currency, data lineage, data ownership, data quality, and data usage.

Accelerated data delivery

Service the business faster from a data perspective. Accelerate data delivery initiatives and deliver in days, not weeks.

Future-proof the data Architecture

Introduce a true self-service model to meet the data needs of the business. Encourage data to be shared and re-used, but most of all, trusted.

In a recent study conducted by Forrester on the Total Economic Impact (TEI) of using data virtualization, Forrester discovered that data virtualization resulted in:

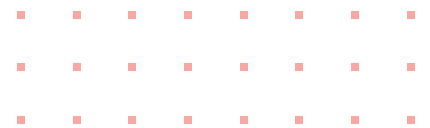
83% faster
in time-to-value

65% faster
data delivery

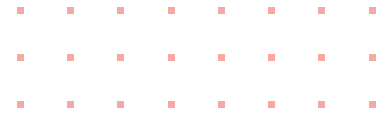
Data Projects taking
6 weeks now 1 week

67% reduction
in data preparation

Knowledge Workers able to work
50% faster



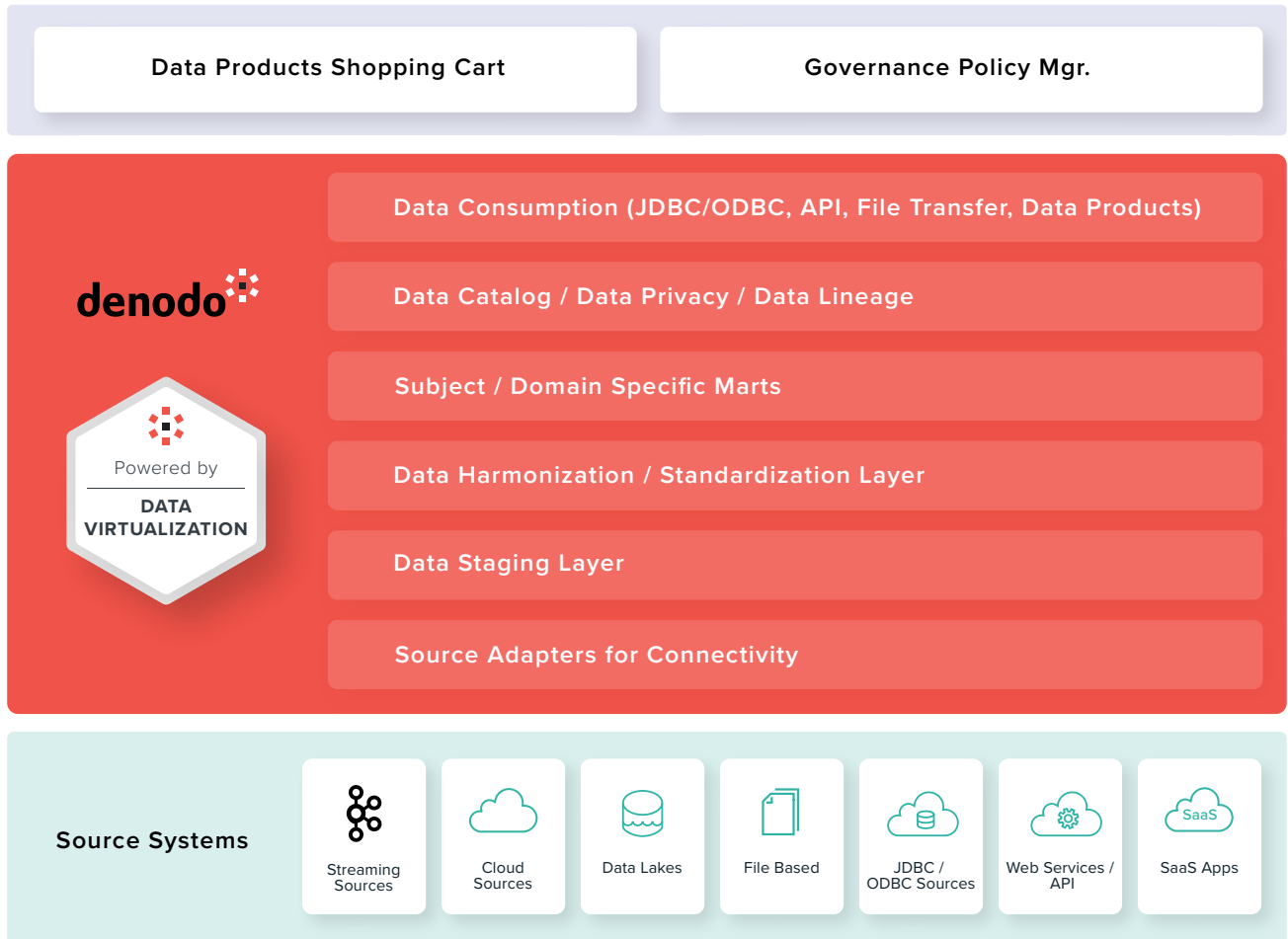
An Overall Solution Architecture



A solution architecture is shown below for reference. It has the following components:

- **Source Systems:** Source systems are various operational, analytics systems, or SaaS applications already in place across the enterprise in both on-premises and cloud systems. These contain data that is exposed through a single governed layer, the data fabric.
- **Source Connectivity Adapters:** Source connectivity adapters are part of the data fabric, enabling it to seamlessly connect to a wide variety of data sources.
- **Data Staging Layer:** The data staging layer is a local representation of source system data in the data fabric. It typically provides a one-to-one representation of source system data sets exposed through the fabric. The data staging layer may also use consistent data types across all sources. So, raw-data-type harmonization may take place within the staging layer itself.
- **Data Harmonization Layer:** Above the staging layer is a harmonization layer that provides light data standardization as well as the data links to the various data sets that enable cross-data-set analysis.
- **Subject/Domain Specific Marts:** Subject-specific data marts are collections of processed data that is relevant to a particular set of use cases. For example, a finance-specific data mart will include data assets processed through the fabric to provide finance-specific data and computations, organized with relevant dimensions to support the required analysis.
- **The Data Catalog, Data Privacy, and Data Access Control:**
 - The data catalog in the data fabric is the enterprise catalog, which provides all users in the organization with search and ‘shop for data’ capabilities. It is an essential component in the data fabric, the primary source for business content in the entire architecture.
 - As the data fabric exposes the entire enterprise data for consumption, it is essential that it complies with data protection regulations and protects personal and sensitive data. Data Privacy controls should be managed and enforced at the data fabric layer and it should apply to all data sources exposed through the data fabric.
 - The data fabric, as the single point of entry, will enforce data access control across all data sources exposed through the fabric. This ensures consistent access to all sources across the fabric, and a single place from which to manage access control.
- **Data Consumption:** This is the layer that serves data from the data fabric to analytics systems and applications through multiple consumption protocols and services. This is typically through JDBC/OCBC, but it can also be through API services.
- **Governance Policy Manager:** With data fabric, data governance is an essential component to all data management deployments across the enterprise. The governance policy manager component becomes an essential ingredient in the data fabric architecture pattern, in which all data governance policies and rules are defined and managed. Data protection rules are integrated with the consumption layer in which they are enforced. The policy manager is potentially an existing data governance tool in the enterprise, but can also be integrated into the data fabric, to bring consistency into data governance rules and policies applied across not only the data fabric but also other data processing systems.
- **Data Products Shopping Cart:** Users want to ‘shop for data’ with the same experience they have with Amazon or Lazada. They want to search for data sets, add them to their shopping cart, and request access. Pre-curated, use-case specific data assets or data products may also be created by data owners and listed for consumption on the catalog. The shopping-like experience provides an interactive workflow with which users can request access and data owners can grant access, provisioning the data in the data serving layer.

The orange box in the architecture below shows solution elements that may be implemented with Denodo as of the writing of this document.



Conclusion

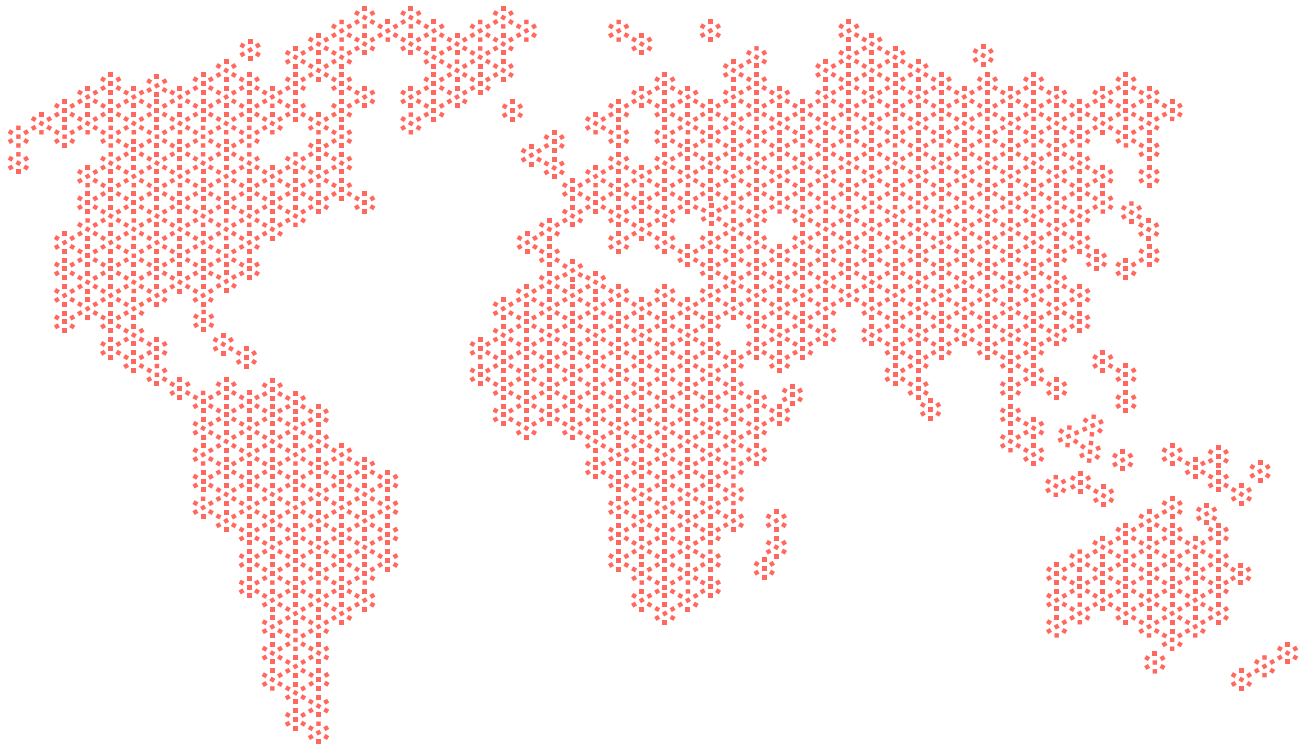
As data estates grow in complexity, a newer, more agile approach is needed to leverage that data and drive decisions across the enterprise. The approach should enable an agile “fail fast” culture that in turn enables users to experiment with many ideas and focus on those with the greatest potential to benefit the enterprise.

Users should also be unhindered by delays and organizational complexities while accessing any information necessary for their projects. This “self-serve” culture puts the power of data analytics into the hands of analysts and business users. Data fabric offers a way to achieve this. It is a great starting point for enterprises embarking on such a journey.

NCS has helped numerous public sector and enterprise clients on their journeys to increased adoption of data and analytics. We have successfully helped a client deploy a data marketplace to enable industry partners to collaborate and share data nationwide. We have deployed an enterprise data platform to enable analytics of sensitive data across multiple institutions without compromising privacy regulations through sophisticated data privacy technologies.

We can jumpstart organizations on their chosen journeys through our methodology and assets. The graphic below shows how we can help.





Denodo is a leader in data management. The award-winning Denodo Platform is the leading data integration, management, and delivery platform using a logical approach to enable self-service BI, data science, hybrid/multi-cloud data integration, and enterprise data services. Realizing more than 400% ROI and millions of dollars in benefits, Denodo's customers across large enterprises and mid-market companies in 30+ industries have received payback in less than 6 months.

Visit www.denodo.com



Make extraordinary happen

Since 1981, we have been part of Singapore's remarkable technological transformation.

With 40 years of experience in digitalising public services and enterprises, we are now #1 in technology services in South East Asia.

Together, we make extraordinary happen.

Visit www.ncs.co