Compunnel
Digital

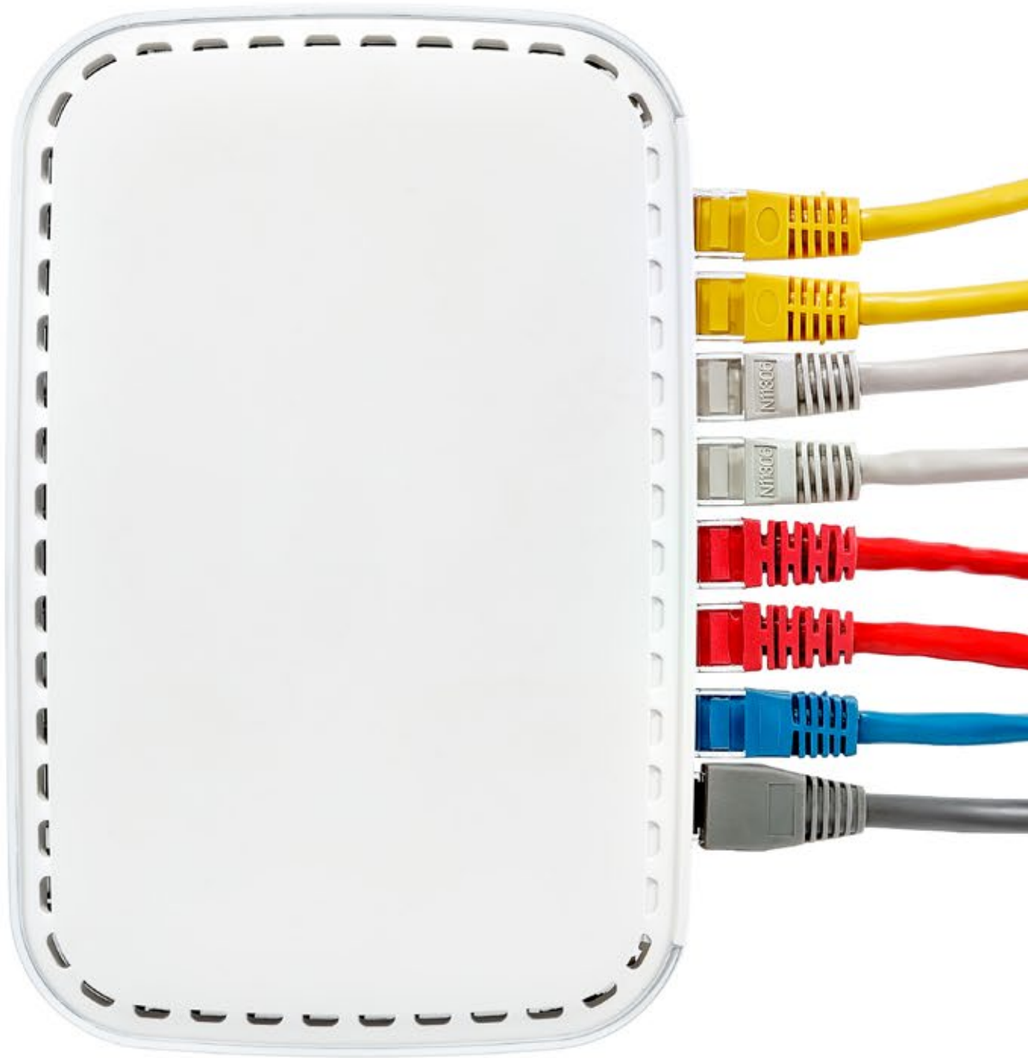
Data Governance

A Foundation for Modern Data Catalog

Introduction

For any business seeking competitive edge in data, enterprise data governance has become a strategic priority. It ensures that data used to make critical business decisions is trustworthy, governed & protected. However, it is not possible until businesses know where the data is stationed, what is the quality of the data as well other data related attributes.

This is where a modern data catalog comes into picture. As the complexity of data ecosystems and the volume of data grows, the traditional approach to a data catalog needs to evolve.



Moving Beyond the Traditional Data Catalog

Let us circle back to the reason for a data catalog in the first place. A Gartner survey revealed that finding and identifying siloed data is one of the primary challenges for any data user and is the first and crucial step. Consequently, the need for data identification gave rise to the use of data catalogs, which enable users to perform actions such as:

- ***Collecting metadata***
- ***Building a data dictionary***
- ***Creating data profile***

However, they failed to provide a single source of truth because of their lack of ability to update automatically and bring metadata alongside data, which created data silos. And this is where the AI-ML enabled data catalog took centre stage.

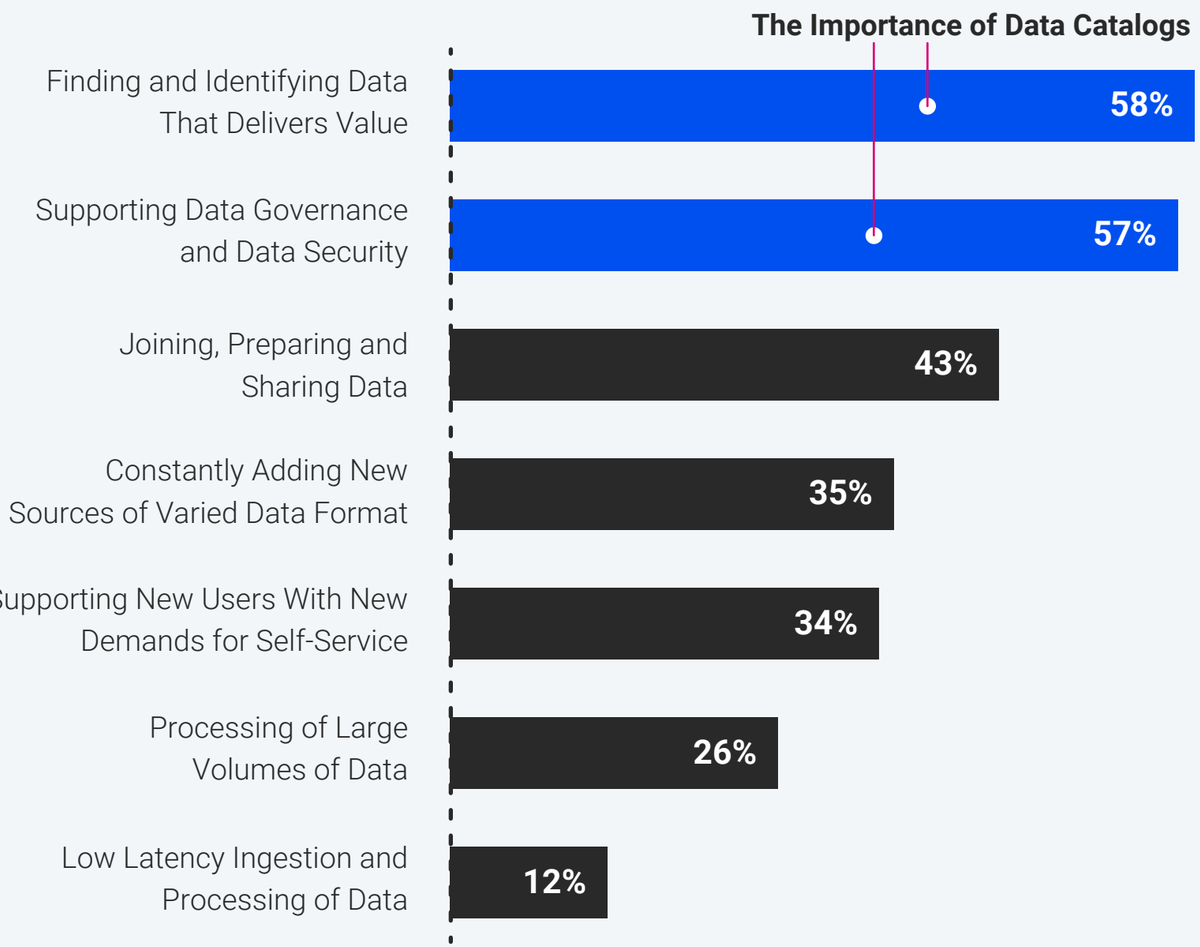
> ***By 2022, over 60% of traditional IT-led data catalog projects that do not use ML to assist in finding and inventorying data distributed across a hybrid/multi-cloud ecosystem will fail to be delivered on time, leading to derailed data management, analytics and data science projects.***
> ***- Gartner***

Features of a modern data catalog platform:

- Acts as a single window for all data sets
- Ensures verified metadata
- Offers advanced analytics capabilities
- Allows data revision and rollback
- Allows easy shareability of each data point

The Biggest Challenges for Data Management

Top Two Centered Around Data Catalogs



n=113

Source: Gartner Data Management Strategy Survey (November 2017)

Introduction to

Modern Data Catalog

Today's data is fragmented across various sources— on-premises, in the cloud, or in multi-cloud environments, and it often exists in silos (walled off by boundaries such as business function, geography, or source.) Because of these siloed and fragmented data environments, data governance becomes a challenge. Data analysts and other users are not able to discover data, nor process the data set, which diminishes the value of data as an asset. Unlike traditional data catalogs, which were used to manage data inventory, the modern data catalog has grown in its functionality and has expanded its reach through machine learning (ML) augmentation. Modern data catalogs are now crucial to data stewardship, data curation, and data governance, and they assist in automating four broad categories of metadata management tasks:

1 Discover Metadata

Modern data catalogs can provide native connectors and REST-based APIs to scan for and extract metadata from various data sources, e.g. enterprise data warehouses, operational databases, enterprise applications, cloud data stores and many more. They embed ML algorithms that allow users to perform a semantic search over metadata to find the most relevant datasets, and then filter the derived datasets as needed.

2 Understand and Enrich

Once the data is collected and scanned, it can be then be used to enrich the collected metadata for further usage. The traditional data catalogs allowed users to view data profiling statistics, so that the user can understand the level of data quality and trust that the data can be used further, whereas modern ML-augmented data catalogs allow auto-tagging and annotation of metadata based on defined metrics. It also allows users to create and maintain business glossaries within the data catalog.

Introduction to

Modern Data Catalog

3 Govern

AI data catalog also helps drive data governance by building controls into the catalog's data-discovery process to ensure that only people with the appropriate need and authority can access sensitive data. It also requires the data requestor to fill out additional forms or sign policies to understand data sharing and usage before user access is granted. Modern data catalogs are also capable of incorporating traceability, tracking who accesses what, how often the data is accessed, and how it's being used.

4 Consume

Once the enriched metadata is available in the data catalog, it allows business users to consume this data by querying across datasets through embedded query editors that highlight sensitive attributes and promote the usage of a certified schema. This enables users to begin their analytics workflows with trusted data and reduces redundancies and errors. Finally, data catalogs include REST-based APIs that enable users to integrate the catalog into their environment and consume cataloged content across different applications.

You Can’t Govern What You Are Not Aware Of

The more widely data is used across the organization, the more strictly it needs to be governed. This can be possible only if the users understand what data is available, where to find it and who to approach with questions. If users don’t know data’s origin, they probably won’t trust it, and they might be reluctant to use it. A business glossary, data dictionaries and data lineage form the foundation of data governance.

What is

Data Governance?

According to DAMA International, Data Governance is defined as, “the exercise of authority, control, and shared decision-making (planning, monitoring, and enforcement) over the management of data assets.” A well-crafted data governance strategy is fundamental for any organization. It underpins how businesses benefit from consistent, standard processes and responsibilities, what data needs to be carefully controlled in the data governance strategy, and the benefits expected from this effort. These strategies form the basis of any data governance framework. Two significant data privacy and security compliance regulations are The EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

Benefits of Data Governance

Regardless of the type of data an organization is managing, whether in data warehouse, data lake, big data, etc., a strong data governance capability is paramount. Some of the benefits of having a data governance are:

Increases value of the data

Allows data transparency

Allows standardized data systems, policies and procedures

Helps to solve data related issues

Emerging Role of Data Governance in

Modern Data Catalog

Data may reside anywhere, it could be on-premise, in the cloud or in a multi-cloud environment. It could be anywhere, which makes data governance a challenging task to accomplish. To implement a successful data governance framework it is important to include these 2 critical processes: discovering and understanding the data. The degree to which an organization is successful at discovery and inventorying data underpins the success of the data governance program. Let’s dig deeper into these two processes more closely.

Benefits of Data Governance

- ***Where all does my data reside?***
- ***What other entities are associated with my data?***
- ***What are the definitions of the data fields?***
- ***Who accesses the data?***
- ***Where does it comes from?***

Data catalogs maintain the necessary business metadata to answer these questions and many more. It gives visibility into wherever the data resides and more. Considering the huge volume of data, it is important that the data catalog answers these questions, and this isn’t possible without leveraging AI & ML in the data catalog (as discussed above.)

Understanding Data

- *How is the term defined?*
- *Who is the owner of this term?*
- *Where did this term originate?*

Data should be defined and classified for it to be searchable and usable. This can be done by having a business glossary, which can answer the above questions. A business glossary is a resource that defines terms across a business domain, providing an authoritative source for all business operations, including its Database Systems (DATAVERSITY). Having a business glossary can help in improving organizational collaboration, removing conflict, and encouraging data stewardship initiatives.

Conclusion

The Need to Replace Traditional Data Governance

Traditional data governance offers a rigorous and strategic framework for designing roles, data standards, and responsibilities, along with procedures and policies for data management throughout the organization. However, today's industry is full of unstructured data sources like IoT and big data, and traditional data governance can't keep up. Therefore, to unleash the value of your data assets, it's important to lay the right foundation and is crucial to have a robust data catalog tool with the right set of capabilities that help establish data governance programs.

About

Compunnel Digital

Compunnel Digital is a leading global IT services and consulting company, providing a wide range of services and solutions in digital, technology, strategy, consulting, and operations. Integrating unparalleled experience and specialized skills across industries, Compunnel Digital strives to work at the intersection of design, innovation, and strategy to ensure digital transformation success. Compunnel Digital helps businesses deliver real results by enabling them to run better, change faster, and grow bigger. Compunnel Digital capitalizes on highly flexible business processes, seamless Global Delivery Network, and an unparalleled domain and industry expertise. Headquartered in New Jersey, Compunnel Digital has been ranked in Inc. 5000 as one of the fastest growing companies. Compunnel Digital has been consistently ranked as one of the “Great Places to Work”, according to Inc.

Compunnel Digital

CONTACT US

Mail us info@compunneldigital.com

Visit us compunneldigital.com

Call us 609-606-9009

FOLLOW US