IDC MarketScape

IDC MarketScape: Worldwide Data Catalog Software 2022 Vendor Assessment

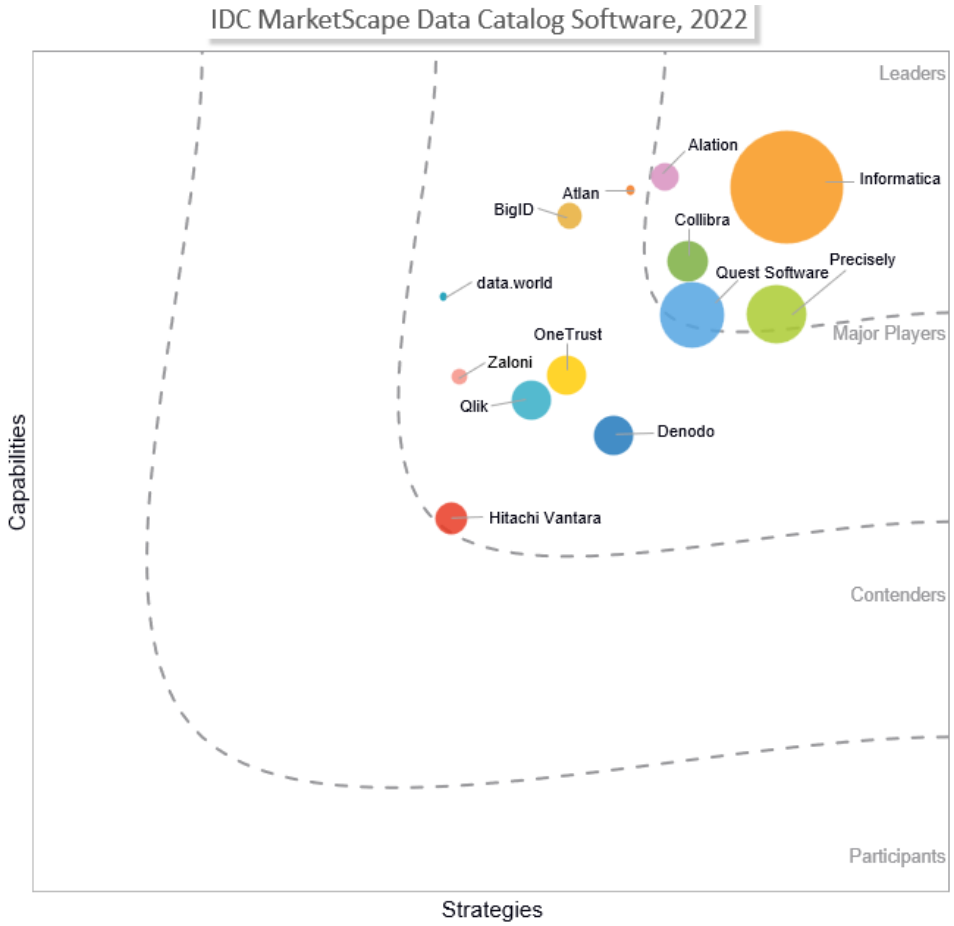
Stewart Bond

THIS IDC MARKETSCAPE EXCERPT FEATURES INFORMATICA

IDC MARKETSCAPE FIGURE

FIGURE 1

IDC MarketScape Worldwide Data Catalog Software Vendor Assessment



Source: IDC, 2022

Please see the Appendix for detailed methodology, market definition, and scoring criteria.

August 2022, IDC #US48395622e

IN THIS EXCERPT

The content for this excerpt was taken directly from IDC MarketScape: Worldwide Data Catalog Software 2022 Vendor Assessment (Doc # 48395622). All or parts of the following sections are included in this excerpt: IDC Opinion, IDC MarketScape Vendor Inclusion Criteria, Essential Guidance, Vendor Summary Profile, Appendix and Learn More. Also included is Figure 1.

IDC OPINION

A digital-first world is emerging amid disruption in the global, regional, and local economies. Digital-first applies to any company, government, or person that is always asking: "Is there some digital-based capability or enhancement that could improve our lives and desired outcomes?" To succeed at becoming digital first, organizations need to put data first.

IDC research regularly shows that executives openly articulate the need for their organizations to be more data driven, to be "data companies." Leveraging data in decision making to drive better business outcomes requires trust, and trusting data requires intelligence. Intelligence about data provides transparency to assure that the highest-quality data is being used in the right context within appropriate governance controls. Data intelligence is fueled by metadata, that is, data about data, such as where is the data, where did it come from, what level of quality does it have, who is accountable for it, what business context should it be used in, and what compliance controls is it subject to. Data intelligence can also help organizations balance security and compliance with data innovation because the more that is known about data, the easier it is to ensure it's used compliantly.

Data catalogs are central to the collection and activation of data intelligence. Gathering and curating intelligence about data is not trivial in highly distributed, diverse, and dynamic modern data environments. Modern data catalogs are characterized as automated and intelligent in the collection, curation, and analytics of technical and business metadata, across many different types of data formats and organization technologies, in hybrid and multicloud environments. Data catalogs also enable the capture of tribal knowledge through crowdsourcing.

Data catalogs gather and manage metadata, the intelligence about data, and the value of metadata becomes exponentially higher once it is activated. Data intelligence can be a point of data control by providing data classification, ownership, compliance, and security intelligence to inform policy. Data intelligence can help data stewards enforce policy and prioritize data quality efforts, and by setting data elements within the context of business processes, intelligence can be used to understand the business value of data. Lineage intelligence is beneficial in assuring data is being used appropriately, and helpful in troubleshooting data exceptions. Data intelligence also enables self-service "shop for data" experiences in data marketplaces.

IDC has nine research practices focused on characteristics of the future enterprise: future of intelligence, future of work, future of digital infrastructure, future of connectedness, future of customer experience, future of digital innovation, future of trust, future of industry ecosystems, and the future of operations. Data is the common element than runs through each characteristic in the future enterprise, and intelligence about data provides context, confidence, and provenance to improve data-driven decision making and business outcomes. As the world continues to face global, regional, and local disruptions, gaining control of data by leveraging data intelligence will help organizations build digital resiliency.

Every vendor in this evaluation activates data intelligence differently, but how it is activated was not a focus of this evaluation. However, it has become clear that data catalogs are becoming part of broader platforms that activate the data intelligence in data governance, quality, DataOps, self-service, collaboration, and control activities. This IDC MarketScape evaluated core data catalog capabilities, focusing on how metadata is captured, curated, and presented, available for activation — to help organizations put data, first, in the goal to be digital first.

IDC MARKETSCAPE VENDOR INCLUSION CRITERIA

Vendor products evaluated in this IDC MarketScape meet the following inclusion criteria:

- The product's primary functionality is that of an enterprise data catalog (EDC).
- The data catalog captures intelligence about multiple disparate data sources and repositories, that is, it cannot be limited to one technology platform or database management system.
- The product can be used standalone, can be licensed separately from a vendor's corresponding data processing platform, and/or is a core capability of a data intelligence platform.
- The data catalog at a minimum must have native capabilities to capture the following types of metadata:
  - Technical
  - Lineage
  - Business
- The data catalog can automate collection of metadata.
- The product is sold and/or being used across multiple global regions.
- The product in its current scope of portfolio coverage and integration is the strategic data catalog product for the software vendor.

ADVICE FOR TECHNOLOGY BUYERS

The data catalog software market is continuing to evolve and mature. Vendors are bringing new innovations forward to meet the demands of organizations trying to control highly distributed, diverse, and dynamic data, operating within the constraints of regional data regulations and balancing the need to accelerate digital innovation. Vendors included in this study range from start-ups to some that have been in business for decades. However, most (if not all) of the data catalog software products evaluated in this study have been in the market for less than a decade.

Each vendor and product have unique origins based on vendor legacy, regardless of the vendor being a market incumbent or a young company. There are three primary categories of catalogs based on vendor legacy:

- **Data integration:** Catalogs emerged from the need to harvest and curate technical metadata for the purpose of managing data integration and engineering workloads. Vendors began adding end-to-end lineage, data stewardship, privacy, and business metadata management into product portfolios to round out data intelligence capabilities.

- **Data governance:** Catalogs emerged from the need to gain control over data in the business, focusing on policy creation and implementation to manage data privacy and security within the constraints of data regulations. Vendors have added data quality, collaboration, lineage, and technical metadata harvesting and curation to round out data intelligence capabilities.
- **Data team collaboration:** Catalogs emerged from the need to support data-native workers in daily activities, facilitating data knowledge capture, management, and collaboration within and across data teams. To collaborate around data, catalogs needed to harvest technical and business metadata. Vendors have added lineage, quality, and automation to round out data intelligence capabilities.

The result is differentiating capabilities, user experiences, strengths, challenges, and licensing models across products in the market that can make the evaluation and procurement process difficult. When evaluating data catalog solutions, organizations should use the following questions that reflect the evaluation criteria in this IDC MarketScape:

- Is the metadata capture an automated process? How much manual effort is required to capture metadata in the catalog?
- What is the scope of lineage metadata that can be shown in the catalog? Some catalogs can only report on lineage as data has been processed within the broader vendor portfolio of services, whereas others can harvest lineage from sources outside of the platform.
- Is business metadata captured by the catalog, and can business classification of data be automated?
- Is behavioral metadata important, that is, do you need to understand who is accessing data, when and how often, and how data sets are being joined in queries? If yes, ask about the vendor's ability to capture it and how it is captured. As with lineage, some can capture it from external sources such as database query logs and others leverage intelligence about how the catalog is being used and how data is processed through services in the broader platform.
- What is the scope of relationship metadata captured in the catalog? Is it reflective of lineage or does it also capture and render semantic relationships among data elements? Are relationships to business terms and business processes included?
- Can the catalog identify where sensitive information resides within the data landscape such as personally identifiable information (PII)? Does the catalog also capture who has access to the sensitive information?
- Can the catalog capture metadata about analytic dashboards, visualizations, reports, models, and other information assets? Data cataloging is not just about the data, but it needs to include analytic assets too.
- Can the catalog integrate with my business intelligence, dashboarding, and analytics tools?
- Can the catalog support multiple user personas?
- Does the catalog support crowdsourcing and collaboration?
- Can the catalog integrate with the collaboration technologies we already use?
- What types of automation capabilities exist in the catalog and for what functions? Is it AI/ML, rules, or regular expression based, or is there a combination of methods being used? What effort is required by users or administrators to maintain automation rules?
- What metadata analytics and reporting capabilities are available in the catalog?
- Is the catalog a cloud-native solution, and how does it connect to on premises and across multicloud data sources for metadata harvesting?

- How many metadata connectors are available as part of the standard offering, and do they align with the types of data sources and data tools we have in our environment?
- What security and compliance standards can the catalog support?
- Are there any prebuilt metadata schemas, templates, definitions, and business terms specific to the vertical industry my organization competes in?
- How is the solution priced? Are there any extra costs associated with additional connectivity or advanced features? What is the pricing relationship to value-added capabilities that may be available in the larger portfolio of capabilities available from the vendor?
- How often will updates be issued, and what is the vendor's commitment to research and development?
- Does the vendor offer customer advisory boards or councils that can help shape the future of the catalog and associated products?
- What does the vendor's innovation and product road map look like for the next three to five years?

The evaluation of data catalog software within this IDC MarketScape is based on the criteria that IDC created, informed by research. Buyers need to perform their own evaluation against their own requirements within the constraints of their own IT, procurement, and industry environmental constraints. Every evaluation process should also include a proof of concept or what is often referred to as a "bake-off." These evaluations of the software need to be done against data used by your organization, in the context of your organizational structures, and within the constraints of your organization's technology environment.

Once the selection process has completed, the hard work of implementation begins. Change management will be key to a successful implementation because successful use of data intelligence will change the way people work. Adoption is critical. Data catalogs can automate a lot of the effort in capturing and curating data intelligence, but like anything else, the value that can be achieved is equally proportionate to the amount of effort that goes into it.

VENDOR SUMMARY PROFILES

This section briefly explains IDC's key observations resulting in a vendor's position in the IDC MarketScape. While every vendor is evaluated against each of the criteria outlined in the Appendix, the description here provides a summary of each vendor's strengths and challenges.

Informatica

After a thorough evaluation of Informatica's strategies and capabilities, IDC has positioned the company in the Leaders category in this IDC MarketScape for worldwide data catalog software.

Informatica has transformed its products and business to meet the demands of data integration and intelligence use cases across modern data environments. Informatica was founded in 1993 and is headquartered in Redwood City, California. Informatica's Intelligent Data Management Cloud is a cloud-first and cloud-native solution powered by AI for end-to-end data management that connects, manages, and unifies data across multicloud and hybrid environments. Informatica's data catalog is part of the Intelligent Data Management Cloud, empowering modernization and advancement of enterprise data strategies.

Quick facts about Informatica are:

- **Product name:** Enterprise Data Catalog
- **Employees:** 5,200+
- **Global footprint**: Customers in 30 countries
- **Top industry areas:** Financial services, manufacturing, retail, healthcare, government, life science, education, telecommunications, and energy
- **Ideal customer size:** Medium-sized to very large enterprises
- **Average implementation time:** One to two months
- **Cloud:** Available as SaaS on AWS and also available in the AWS, Azure, and Google Cloud Marketplaces
- **Pricing model:** Subscription, based on number of data sources cataloged for on-premises deployments; consumption based for the SaaS offering
- **Partner ecosystem:** Global partner network
- **Interesting fact:** A hedge fund, working to break down historical silos and accelerate value for both data producers and consumers, is leveraging Informatica EDC on top of Databricks to cut data discovery time by 75%. Data preparation time has also been cut by 50–60%, helping connect data subject matter experts, data producers, and data scientists to better serve portfolio analysts in the business through automation that streamlines data delivery. The number of emails between consumer and producer has dropped significantly now that EDC is the front end to data.

Strengths

- **Market presence:** Informatica has the largest share of revenue in the overall data integration and intelligence software market as defined by IDC. Informatica also has the large share of revenue in the metadata management and data intelligence submarkets, as reported in *Worldwide Data Integration and Intelligence Software Market Shares, 2021: Accelerated Growth in a Digital-First World* (IDC #US47920522, June 2022).
- **One common metadata foundation:** Informatica's Intelligent Data Management Cloud portfolio of capabilities leverages one common foundation of metadata, captured and generated within the platform whenever data is moved, transformed, cleansed, controlled, and managed throughout its life cycle. Informatica's core central intelligence engine (CLAIRE) learns from this common foundation of metadata to automate data intelligence, integration, management, and governance activities.
- **Cloud:** Informatica supports multiple deployment patterns across all major hyperscale platforms. The Intelligent Data Management Cloud is available as a service, including some serverless functions, managed by Informatica. It can also be deployed and managed by a customer on its cloud infrastructure of choice, or it can be deployed on premises.

Challenges

- **Long-tail connectivity:** On-premises deployments at large customers with complex legacy and custom data environments can be a challenge in some customer situations.
- **Market perception:** Informatica is often faced with stereotypes and competitive assertions that it is a legacy ETL company with legacy technology. Informatica's transformation over the past several years illustrates it has a broad and comprehensive data solution, from integration to intelligence, supporting data engineering to data governance workflows all available as part of an integrated cloud platform. An innovative commercial model, which includes serverless options, gives customers the opportunity to optimize spend.

Consider Informatica When

Consider Informatica if you are a midsize to very large global enterprise with a data strategy that prefers a data catalog to be part of a comprehensive modern data platform and a data catalog capability that is included in the same platform that manages all data integration and intelligence workloads in the enterprise, reduces integration effort, and increases the intelligence of the catalog. Although the catalog benefits from the broader platform, it can also be used independently of data movement, transformation, and cleansing functions deployed in alternative solutions.

APPENDIX

Reading an IDC MarketScape Graph

For the purposes of this analysis, IDC divided potential key measures for success into two primary categories: capabilities and strategies.

Positioning on the y-axis reflects the vendor's current capabilities and menu of services and how well aligned the vendor is to customer needs. The capabilities category focuses on the capabilities of the company and product today, here and now. Under this category, IDC analysts will look at how well a vendor is building/delivering capabilities that enable it to execute its chosen strategy in the market.

Positioning on the x-axis, or strategies axis, indicates how well the vendor's future strategy aligns with what customers will require in three to five years. The strategies category focuses on high-level decisions and underlying assumptions about offerings, customer segments, and business and go-to-market plans for the next three to five years.

The size of the individual vendor markers in the IDC MarketScape represents the market share of each individual vendor within the specific market segment being assessed. For this IDC MarketScape, vendor size was determined using IDC's 2021 Software Tracker, which represents an estimate of each vendor's 2021 software revenue. Functional markets include data integration and intelligence software, governance, risk, and compliance software.

IDC MarketScape Methodology

IDC MarketScape criteria selection, weightings, and vendor scores represent well-researched IDC judgment about the market and specific vendors. IDC analysts tailor the range of standard characteristics by which vendors are measured through structured discussions, surveys, and interviews with market leaders, participants, and end users. Market weightings are based on user interviews, buyer surveys, and the input of IDC experts in each market. IDC analysts base individual vendor scores, and ultimately vendor positions on the IDC MarketScape, on detailed surveys and interviews with the vendors, publicly available information, and end-user experiences in an effort to provide an accurate and consistent assessment of each vendor's characteristics, behavior, and capability.

Market Definition

IDC defines the data catalog software market as a submarket of what IDC defines as a data intelligence software market. Data catalog software is a cornerstone of data intelligence, providing users with technical, business, lineage, and behavioral intelligence about data being managed by the organization. As part of the data intelligence software market, data catalog software leverages metadata to enable people, processes, and technology with trustworthy and reliable data in delivering better business outcomes, controlling data in and across the enterprise, and enabling self-service access to data. Metadata management software isn't new, but modern data catalogs are separating themselves from legacy metadata management solutions in their nontechnical user experiences, automation, crowdsourcing data knowledge, inherent intelligence, and internet-like search for data capabilities. More about how IDC defines data intelligence can be found in this blog post.

Strategies and Capabilities Criteria

The assessment criteria are divided into two primary categories of strategies and capabilities for the success of the data catalog evaluation. IDC analysts look at how well a vendor is building/delivering capabilities that enable it to execute its chosen strategy in the market. In the strategy category, IDC evaluates whether a software vendor's strategies in various areas are aligned with customers' requirements (and spending) over a defined future time period and business and go-to-market plans. In this IDC MarketScape, the most influential criteria within the strategy category is financial/funding/market presence, measured by the number of years the company has been in business, IDC estimated revenue for the company in related software markets, the number of global customers, and the percentage of customers running data catalog software in the cloud. For key capabilities, IDC not only evaluated the native capabilities offered in the data catalog application but also considered third-party and partner-extended capabilities that were natively connected and supported by the vendor. These extensions were either provided as native integration, modular add-ons, or configurable options in the administrative user interface or available for download and self-service install from the vendor's marketplace.

This IDC MarketScape evaluated essential capabilities of metadata ingestion and capture, evaluating how well each vendor captures different types of metadata required for data intelligence, usability by multiple personas, crowdsourcing and collaboration capabilities, and the level of AI/ML-based automation being used to help deliver insights at scale. Data catalog configurability and security capabilities were evaluated for data protection and regulatory compliance support. Architectural aspects of the solution were evaluated against cloud-native and hybrid reference architectures. Strategic evaluation focused on the vendor's dedication to customer success, innovation accelerators, opportunities for growth including partner and geographic strategies, dedication to research and development, how long the company has been in business, IDC-estimated revenue for the company in the data integration and intelligence software market for 2021, and the current number of customers combined with renewal rates and new customer growth rates. Vendors provided reference customers that participated in interviews and an evaluation survey, responding to questions about vendor strategies and product capabilities.

LEARN MORE

Related Research

- *IDC TechBrief: Operational Data Intelligence Software* (IDC #US49374422, August 2022)
- *Worldwide Data Integration and Intelligence Software Market Shares, 2021: Accelerated Growth in a Digital-First World* (IDC #US47920522, June 2022)
- *Worldwide Data Integration and Intelligence Software Forecast, 2022-2026* (IDC #US47920222, May 2022)
- *Controlling Data in the Future of Operations* (IDC #US48972622, April 2022)
- *In Data We Trust: Or Do We?* (IDC #DR2022_T5_SB, March 2022)
- *IDC Market Glance: Data Control Plane, 1Q22* (IDC #US47919921, March 2022)

Synopsis

This IDC study provides an assessment of data catalog software applications and presents the criteria most important for companies to consider when selecting a data catalog application as part of a data intelligence solution. This assessment discusses both quantitative and qualitative characteristics that explain success in data cataloging and intelligence in organizational data enablement. The evaluation is based on a comprehensive and rigorous framework that assesses vendors relative to the criteria and one another. The study highlights the factors expected to be the most influential for success in the market now and into the future.

"Data catalogs are central to the collection and activation of data intelligence, but gathering and curating intelligence about data is not trivial in modern data environments characterized as having data that is highly distributed, diverse, and dynamic," says Stewart Bond, research vice president, Data Integration and Data Intelligence Software research at IDC. "Every vendor in this evaluation approaches data intelligence differently, and it has become clear that data catalogs are becoming part of broader platforms that capture and curate data intelligence to activate it in data governance, quality, DataOps, self-service, collaboration, and control activities. Buyers will need to carefully evaluate each vendor's solution capabilities against data intelligence requirements within organizational, regional, environmental, and technology constraints."

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com