GIGAOM

Image credit: Funtap

Andrew J. Brust

Sep 24, 2021

GigaOm Radar for Data Governance Solutions v1.0

Data Governance, Data Infrastructure, AI & Analytics

GigaOm Radar for Data Governance Solutions

Table of Contents

1. Summary

In the era of digital transformation, organizations employ analytics to gain and maintain a competitive edge by constantly improving the value of the insights they derive from data. To achieve this strategic objective, it's not enough to simply implement high-powered analytics software. IT leadership must "evangelize" data throughout the organization, transforming their institutional culture into one that is data-driven. In such an organization, data informs decision-making—and becomes a familiar, trusted concept—for personnel at all levels of the company.

At the same time, organizations find themselves striving for another strategic objective: establishing an effective framework for protecting their data. Pressure to do so has been brought on by the passing of global privacy regulations such as GDPR and CCPA, the additional scrutiny of and heightened awareness around the management of user data stemming from recent high-profile data breaches, and the sudden proliferation of online transactions resulting from the pandemic and the spike in user data generated by them.

Data governance has catapulted to prominence as the key for organizations trying to balance these two seemingly contradictory strategic objectives. Many platforms and solutions have been developed to assist organizations with modernizing their approach to data governance.

As data governance platforms rise in importance within an organization's data management strategy, the act of selecting one requires careful deliberation. It can be difficult to wade through the sea of options, as many products describing themselves as "data governance" solutions do not necessarily provide the core functionalities described here. At the same time, other platforms that do provide robust data governance features are not marketed as such. This report can help you navigate these murky waters, informing you in the process and guiding you to the platform that best suits your needs.

This GigaOm Radar report builds on insight provided in the Key Criteria Report for Data Governance Solutions. That report establishes a decision framework for IT leaders assessing governance solutions, describing the decision criteria for evaluating the different vendor offerings: the table stakes common to virtually all the offerings in the report, the key criteria that differentiate offerings in the space, and the evaluation metrics that describe high-level characteristics that determine the impact a particular solution can have on your organization. We round out the decision criteria description with a discussion of the technologies emerging in the landscape, which represent a preview of where the industry will likely be headed in the future.

In this Radar report, we apply informed, technical insight to these criteria to assess and score the value that available solutions can have to the organization.

HOW TO READ THIS REPORT

This GigaOm report is one of a series of documents that helps IT organizations assess competing solutions in the context of well-defined features and criteria. For a fuller understanding consider reviewing the following reports:

Key Criteria report: A detailed market sector analysis that assesses the impact that key product features and criteria have on top-line solution characteristics—such as scalability, performance, and TCO—that drive purchase decisions.

GigaOm Radar report: A forward-looking analysis that plots the relative value and progression of vendor solutions along multiple axes based on strategy and execution. The Radar report includes a breakdown of each vendor's offering in the sector.

Solution Profile: An in-depth vendor analysis that builds on the framework developed in the Key Criteria and Radar reports to assess a company's engagement within a technology sector. This analysis includes forward-looking guidance around both strategy and product.

2. Key Criteria Comparison

The GigaOm Radar report identifies key differentiating features and capabilities for solutions in the sector. It further identifies evaluation metrics—top-line characteristics that help define the impact that a solution may have on an organization. These key criteria and evaluation metrics are described here, and each vendor then scored in terms of its support for them.

Key Criteria

Single-View Control Panel: Some platforms feature an integrated cockpit/dashboard that functions as a single "window" or "pane of glass" through which data professionals can view the controls and policies for the entire organization. This control panel provides a unified, holistic approach, simplifying access management across the organization while allowing data to remain physically close to the sources that generate it for optimal performance.

Audit/Logging Capabilities: This is an important capability in which the platform is configured to record the activity performed on data or metadata. This activity can include access requests, queries to the data, and changes to the policies applied to the data. The significance of this capability lies in how it helps display patterns in the activity and determine any anomalies in normal usage that may point to unauthorized access or misuse. Such alerts can be a prelude to data breaches, and therefore an important tool for helping to avoid them.

Data Lineage Capabilities: Data lineage capabilities generate a viewable trail for a data asset, allowing it to be tracked from its source, through transformations and movement to other data repositories, and out to assets such as reports and visualizations. Platforms that include automated data lineage capabilities are given special note in this report.

Self-Service Approach: A number of the platforms in this report are designed to provide self-service data governance solutions to organizations. With self-service platforms, business professionals in particular, as well as data stewards, are enabled to perform data governance tasks without the support of IT professionals. In other words, the platforms are designed to free IT professionals from having to create and manage the policies and data access controls on top of their other duties by providing a designated space and specialized tools for other members of the organization to perform these tasks.

Automated Data Classification: Automated data classification helps organizations discover information about the kinds of data contained within its data sources. This capability automatically scans large numbers of datasets and identifies the type of data they contain. Classifications of data include not only basic information about the type (numeric, text, date/time data, and so forth), but also categories such as geographic region, customer names or IDs, and most importantly for the scope of this report, personally identifiable information (PII), although this last category is just one subset of data classification.

Evaluation Metrics

These evaluation metrics are designed to be applied simultaneously with the key criteria outlined above to better help organizations identify their ideal solutions.

Elasticity: Some platforms are structured with optimizations that allow them to be scaled up or down depending on the volume of the workload the customer needs to handle. One example: a platform that runs on dedicated clusters and can add or remove clusters depending on the customer's workload.

Data Source Connectivity: Because data governance platforms enable organizations to manage access to data across the entire organization, support for a wide variety of data sources is crucial. The platforms in this report meet this customer need by offering support for data lakes, data warehouses, database platforms, enterprise applications, and related APIs.

Collaboration: This metric measures the degree to which the platform supports collaboration among business users by means of features that encourage users to explore and govern data together. For example, some platforms include dedicated project workspaces where users can collaborate around data securely. Other platforms might encourage collaboration through the sharing of governed data among users within the same organization, as well as secure sharing of data with third parties, by including data-sharing features.

Maturity: A number of the vendors in this report are startups of varying degrees of maturity, specializing only in data access management and governance tools, while others are venerable players with strengths in many areas of data management beyond the scope of this report. The good news is that this heterogeneity in the landscape means there is something for everyone. This metric measures the maturity of the vendor's offering as it pertains to the specific scope of this report.

Ease of Use: This metric assesses the degree to which the platform's user interface is appealing and accessible to business users. This includes the overall feel of the graphics and layout as well as the features within the platform. Hence, there will be some overlap with other ratings in this report.

Building on the findings from the GigaOm report, "Key Criteria for Evaluating Data Governance Solutions," **Table 1** summarizes how each vendor included in this research performs in the areas that we consider differentiating and critical in this sector. **Table 2** follows with insight into each product's evaluation metrics—the top-line characteristics that define the impact each will have on the organization. The objective is to give the reader a snapshot of the technical capabilities of available solutions, define the perimeter of the market landscape, and gauge the potential impact on the business.

Table 1. Key Criteria Comparison

	KEY CRITERIA				
	Single-View Control Panel	Audit/Logging Capabilities	Data Lineage Capabilities	Self-Service Approach	Automated Data Classification
BigID	++	+++	++	+++	+++
Collibra	+++	++	+++	++	+++
Delphix	+++	++	++	++	+
Global IDs	+++	++	+++	+++	+++
Immuta	+++	+++	++	+++	++
Informatica	+++	++	+++	++	+++
Okera	+++	+++	++	+++	++
Privacera	+++	+++	++	+++	+++
Talend	+++	++	+++	–	–

+++ Exceptional: Outstanding focus and execution
++ Capable: Good but with room for improvement
+ Limited: Lacking in execution and use cases
– Not applicable or absent

Source: GigaOm 2021

Table 2. Evaluation Metrics Comparison

	EVALUATION METRICS				
	Elasticity	Data Source Connectivity	Collaboration	Maturity	Ease of Use
BigID	++	+++	++	++	+++
Collibra	++	++	+++	+++	++
Delphix	++	+++	++	++	+++
Global IDs	+++	+++	+	+++	++
Immuta	+++	+++	+++	++	+++
Informatica	++	+++	+++	+++	++
Okera	+++	++	+	++	+++
Privacera	++	+++	+	++	+++
Talend	++	+++	–	+++	–

+++ Exceptional: Outstanding focus and execution
++ Capable: Good but with room for improvement
+ Limited: Lacking in execution and use cases
– Not applicable or absent

Source: GigaOm 2021

By combining the information provided in the tables above, the reader can develop a clear understanding of the technical solutions available in the market.

3. GigaOm Radar

This report synthesizes the analysis of key criteria and their impact on evaluation metrics to inform the GigaOm Radar graphic in **Figure 1**. The resulting chart is a forward-looking perspective on all the vendors in this report, based on their products' technical capabilities and feature sets.

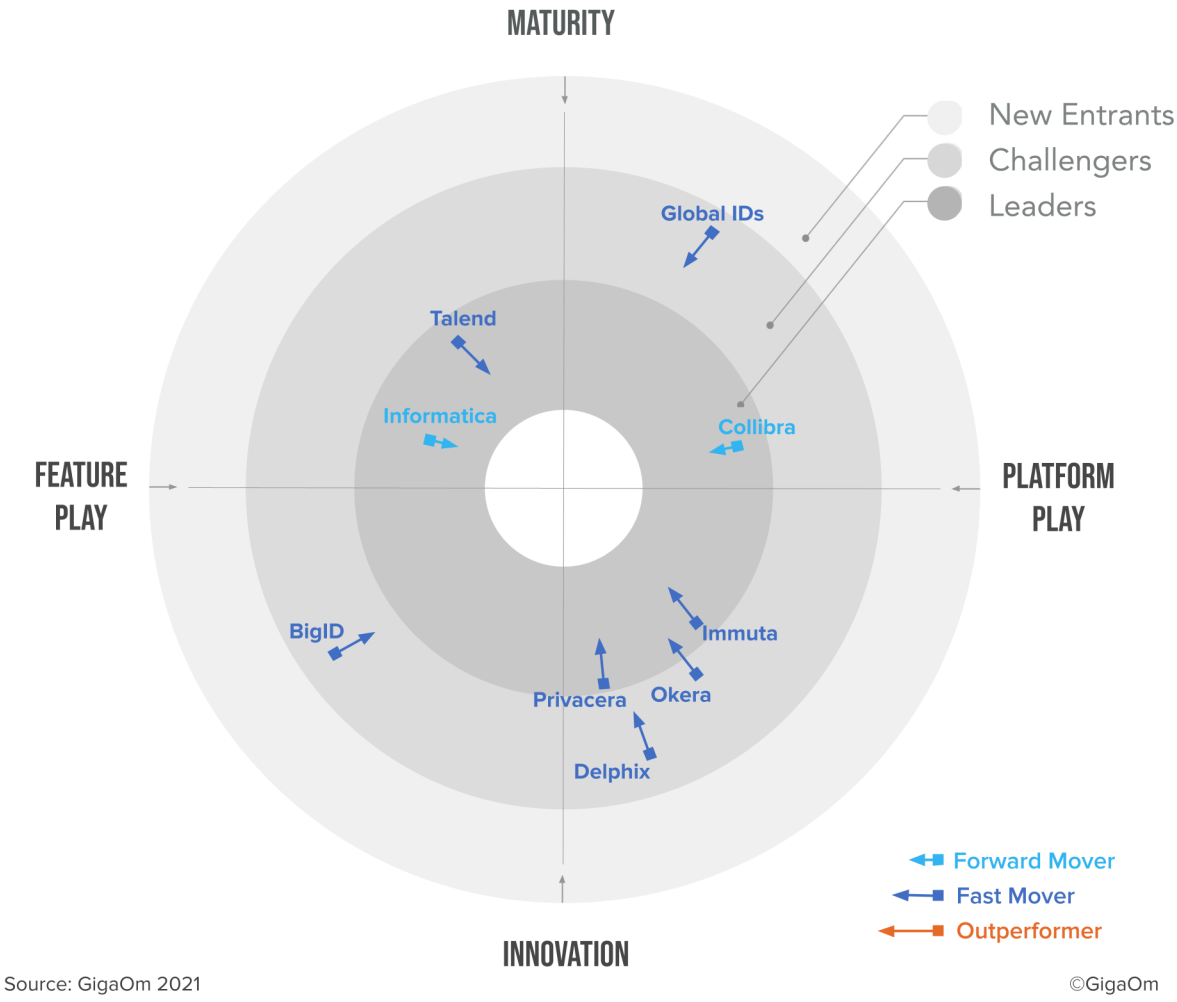

Figure 1. GigaOm Radar for Data Governance

The GigaOm Radar plots vendor solutions across a series of concentric rings, with those set closer to center judged to be of higher overall value. The chart characterizes each vendor on two axes—Maturity versus Innovation, and Feature Play versus Platform Play—while displaying an arrow that projects each solution's evolution over the coming 12 to 18 months.

As you can see, the data governance sector offers a balance of solutions, with vendors ably attacking the market from all four quadrants. A pair of companies hail from the Maturity and Feature Play quadrant—overall Leader Informatica followed not far behind by Talend—reflecting the state-of-the-art

data governance capabilities embedded within each vendor's respective stack.

Also notable is the crowd of vendors approaching the sector from the Innovation and Platform Play quadrant. Included among these are two Leaders (Privacera and Immuta) and a pair of Challengers (Okera and Delphix). All have shown a willingness to be innovative and forward-thinking in their effort to bring cutting-edge solutions to the challenges of data governance.

Finally, Collibra bears mention as a stable, platform-minded solution with a long and strong pedigree across the data management space. For organizations focused on ensuring consistency and compatibility across the data estate, Collibra stands out as a compelling Leader.

INSIDE THE GIGAOM RADAR

The GigaOm Radar weighs each vendor's execution, roadmap, and ability to innovate to plot solutions along two axes, each set as opposing pairs. On the Y axis, **Maturity** recognizes solution stability, strength of ecosystem, and a conservative stance, while **Innovation** highlights technical innovation and a more aggressive approach. On the X axis, **Feature Play** connotes a narrow focus on niche or cutting-edge functionality, while **Platform Play** displays a broader platform focus and commitment to a comprehensive feature set.

The closer to center a solution sits, the better its execution and value, with top performers occupying the inner Leaders circle. The centermost circle is almost always empty, reserved for highly mature and consolidated markets that lack space for further innovation.

The GigaOm Radar offers a forward-looking assessment, plotting the current and projected position of each solution over a 12- to 18-month window. Arrows indicate travel based on strategy and pace of innovation, with vendors designated as Forward Movers, Fast Movers, or Outperformers based on their rate of progression.

Note that the Radar excludes vendor market share as a metric. The focus is on forward-looking analysis that emphasizes the value of innovation and differentiation over incumbent market position.

4. Vendor Insights

BigID

Founded in 2016, BigID is a data intelligence company that aims to help companies manage their sensitive data. BigID offers a data intelligence platform that combines a data discovery foundation consisting of ML-based classification, cataloging, correlation, and cluster analysis with modular "apps," enabling organizations to govern their data and enforce data privacy and security policies for compliance with global regulations.

BigID's automated, ML-based platform provides an end-to-end data intelligence solution driven by automation and machine learning. It supports deployment in the cloud or on-premises, and broad data source connectivity. It is structured as an extensible platform, within which customers can use both native modular apps and partner apps. These apps support various data management capabilities, allowing users to take action for governance, security, and privacy. Customers can build their own apps to extend functionality for custom use cases as well.

In May of 2021, BigID announced the arrival of its Data Governance Suite, a set of apps that can perform a variety of data governance, privacy, and security functions. The apps included in this suite are listed as Data Stewardship, Data Quality, Data Retention, and Business Glossary—all leveraging the foundational ML-augmented data catalog.

With the Data Stewardship app, an organization's data stewards can collaborate, gain context about data, and improve data quality. The app provides a single-view control panel from which data stewards can assign owners to data, assign workflows for remediation, and track audit requests. The Data Quality app lets users search and view data quality across attributes and provides a central place from which administrators can define, and later add or update, rules to measure data quality. The Data Retention app allows administrators to manage data retention by applying policies, automating workflows, and managing policy violations.

BigID also offers an abundance of apps for data privacy and data security. The platform's discovery foundation enables administrators to define and enforce specific policies based on data content and context. The Data Risk Mitigation app allows users to manage and score risk consistently across a variety of parameters, like data type, content, policy, and residency. The Data Processes & Sharing app helps users manage, monitor, and validate third party data transfers and sharing, while automating data mapping as well. The File Access Intelligence app highlights overexposed data and over-privileged users, while the Data Remediation app enables enterprise-wide remediation orchestration.

Strengths: BigID brings a unique, app-first approach to data management, data security, and governance, and leverages ML technologies to power its platform. Features for automated data discovery and data classification are also strong.

Challenges: A number of capabilities offered by the BigID platform are delivered through apps. These are de facto modules and they, along with third-party authored applications available on the BigID marketplace, deliver a wide range of capabilities. However, this means that functionality and features may be somewhat fragmented across this modular architecture.

Collibra

Founded in 2008, Collibra is a well-established name in the data industry, providing a number of data intelligence solutions.

Collibra's solution for access management and data protection approaches the matter from a data intelligence perspective that centralizes, automates, and guides workflows. A single platform enables teams across departments to collaborate, operationalize privacy, and address global regulatory requirements. Collibra takes a persona-based approach to user experience, with differing levels of access to data assigned to different personas.

Product highlights include ML-powered data discovery, which runs an ML model in the background to locate, identify, and classify PII and other sensitive information. A data sharing capability allows users to share trusted and governed data to support privacy programs. Regulatory dashboards monitor compliance progress for CCPA and GDPR readiness with easy-to-understand reports and dashboards. Audit trails are generated whenever data is modified within the platform, showing the impacted change, responsible user, and time stamp.

A processing activities workflow allows users to understand how sensitive data is used throughout an intelligent workflow. Data lineage capabilities dynamically generate a visual graph of data flowing throughout the organization. Customizable pre-built templates allow privacy assessments to be conducted while mitigating data risks.

Strengths: Collibra's data intelligence platform leverages strong ML-based data discovery and data classification capabilities, as well as detailed reporting and dashboards.

Challenges: Collibra, as a whole, is a mature company with a portfolio of robust offerings in many areas of the data management space. Data privacy and access management are only one facet of a larger whole. Therefore, implementing the data privacy solution really requires adoption of the wider Collibra suite.

Delphix

Founded in 2008, Delphix assists its customers on their journey to digital transformation through custom application development, self-service provisioning, and enterprise resource planning solutions. The Delphix DevOps Data Platform is the company's API-oriented technology set for automating data operations. Its capabilities include the delivery of compliant data to those who need it, helping organizations accelerate their data governance initiatives with an automated approach to

data protection, policy management, sensitive data discovery, and data masking.

To help customers protect and govern their sensitive data, the Delphix platform includes an array of features. An organization's data can be governed from a single point of control, which Delphix calls a "data control tower." This access point also contains all of Delphix's API-driven operations, and can be accessed from a single interface, where administrators manage access privileges, create data policies, modify or create new algorithms, and more. A robust set of APIs integrates masking and virtualization into enterprise development and CI/CD workflows.

Delphix scans data values and metadata, automatically detecting sensitive data values. The platform includes prepackaged masking algorithms, and application-specific accelerators, which identify known schemas for known applications, resulting in a default understanding of where sensitive data resides. This allows data from those applications to be governed out of the box, by virtue of having Delphix in place. Options also exist for building custom masking algorithms to address company- or industry-specific policies and regulations. An algorithm SDK enables partners and customers who might be in the same industry to create and share data masking algorithms. The low-code/no-code approach to masking provided by the profiling templates and automated masking framework ensures the platform's capabilities are accessible to all users, including non-technical ones.

Delphix's platform allows organizations to deliver masked data copies to downstream environments. Recently Delphix released an enhanced masking functionality called "continuous masking," which allows transactional data to be masked on-the-fly, in near real time, and delivered in Delphix's virtualized form. Another capability, called "hyperscale masking," provides special connectors that mask large amounts of file data using horizontal scaling. A connector SDK allows partners and customers to custom-configure JDBC drivers for optimized operation with corresponding data platforms. Audit logs record user actions taken in the platform, and can be accessed directly from the UI or REST APIs.

Delphix is a comprehensive, open, and extensible platform, including API or UI-based operation; broad data source support from conventional to NoSQL databases, mainframe platforms, files, and more; support for all clouds that includes PaaS, cloud-native DBs, object storage, and APIs.

Strengths: Delphix brings a differentiating automated approach to data masking, including its new continuous masking and hyperscale masking functionalities. Additionally, its integration of data masking with data virtualization enables both data protection and ready access. It is a highly extensible platform, with its single-view control panel called the "data control tower," prepackaged apps for sensitive data detection and discovery, and a low-code/no-code approach to masking to ensure usability at all technical levels.

Challenges: For Delphix, there remains room to grow in the realm of data lineage capabilities: for example, it lacks detailed drill-down views of individual data assets, allowing the tracing of these assets through various transformations and out to visualizations and reports. And while Delphix offers automated sensitive data detection, broader data classification capabilities for other types of data are not yet in place.

Global IDs

Founded in 2001, Global IDs aims to help its customers solve data management problems. To this end, it offers software that can assist organizations with data management, data governance, data privacy, and regulatory compliance.

The vendor offers both the Global IDs Data Ecosystem Evolution Platform (DEEP) and the Global IDs Enterprise Data Automation (EDA) Platform. DEEP is an integrated solution that lets organizations structure and understand their data to drive business insights and solve business problems. It includes metadata management, data lineage and reporting, data discovery and rules management, encryption, anonymization, monitoring, policy management, and governance workflow features.

EDA is an inclusive enterprise data platform that provides a strong foundation for a wide set of data solutions: data governance, compliance, cloud migration, application rationalization, privacy, and more. The platform is scalable and automates many complex data processes.

Global ID's data governance solution leverages the features of its platforms to help its customers create a strong foundation for their data governance framework and comply with global data protection regulations. These features include data discovery, data quality, data classification, and lineage that analyzes how data flows through the enterprise. These features help organizations identify, manage, and protect sensitive information using machine learning, automated matching, and pattern recognition.

The platform also provides data privacy features, including sensitive data classification using machine learning and business rules, identification of PII data, generation of custom reports for compliance with GDPR and CCPA, federated search queries for individuals and their key data across the enterprise, and 200+ predefined GDPR and CCPA entities and domains.

Strengths: Global IDs's platform leverages a strong ML and automation foundation to generate a holistic view of an enterprise's data, which can be used as a starting point from which to define a strong data governance framework. Data classification is especially strong.

Challenges: Details are sparse in the company's literature. The permissions and access controls are not always as fine-grained as those of other offerings in this report.

Immuta

Founded in 2015, Boston-based vendor Immuta's mission is to "enable the legal and ethical use of data." The vendor aims to help organizations modernize their approach to data governance and solve two of the main issues that organizations face regarding their data: access and governance. With this focus, Immuta seeks to enable data professionals with complete access to their data and to use it in compliance with data protection regulations.

Toward this end, the Immuta platform functions as a single, unified access point for data throughout an organization. It provides fine-grained access control at the column, row, and cell level. Native access support is provided for Databricks, Databricks SQL Analytics, Snowflake, Amazon Redshift, Apache Spark, Trino (formerly PrestoSQL), Starburst, Azure Synapse, Google BigQuery, Amazon S3, Azure Data Lake Storage (ADLS) Gen 2, as is access to dozens of other data sources using Immuta Query Engine.

A no-code policy builder allows policies to be authored using plain English, providing accessibility to non-technical users. APIs allow policies to be extended to other tools in the data stack. Through these policies, data is hidden, masked, redacted, and anonymized, depending on either the privileges associated with the role of the user within the organization or contextual factors and environmental attributes. Additionally, Immuta offers a Policy-as-Code capability, enhanced with a command line interface (CLI), enabling a DevOps framework to be applied to policy creation and management.

Audit logging capabilities allow users to view what actions have been taken and what data has been accessed down to the exact SQL query that was written. Audit/monitoring data can be exported to logging or SIEM tools. Immuta "fingerprints" allow users to save a snapshot of a dataset as it existed at any point in time to compare to past and future versions, which is useful for tracking changes.

Immuta's "projects" feature enables data sharing and user collaboration around the organization's data. Projects are designated workspaces that unite users and data sources under a common defined purpose. This purpose is then used to restrict access to data and enable secure collaboration. A project equalization feature allows users working under the same project to see the same data, regardless of their varying levels of access. Additional capabilities of the project workspaces enable users to write data back to the platform within the workspaces, and to share their analysis with other users.

The Immuta platform is a fully containerized solution that runs in the cloud, on-premises, or both, depending on customer needs. In the public cloud, the platform supports storage and analytics services from Amazon Web Services, Microsoft Azure, and Google Cloud Platform. Immuta Managed Cloud, available via the AWS Marketplace, is a fully managed option that is deployed directly into the customer's AWS account, where deployment, infrastructure, configuration, maintenance, and backups are handled automatically for the customer. The hybrid cloud model allows Immuta to be deployed across a customer's on-premises and cloud infrastructure. The on-premises deployment model supports customers operating exclusively in their own data centers.

Strengths: Pluses include Immuta's single-view control panel (unified platform); easy-to-use user interface; no-code policy builder; and project workspaces, which are designed to encourage use by those without a high level of technical expertise. Additionally, Immuta's dynamic ABAC model results in optimizations for scalability and time-to-implementation.

Challenges: Immuta is limited to highly structured tabular data; therefore, unstructured or blob-based storage and NoSQL databases do not have many options for complex policy enforcement in Immuta. Also, Immuta's primary focus is cloud compute and cloud data warehouses, rather than on-premises

databases.

Informatica

Founded in 1993 and headquartered in Redwood City, California, Informatica is a long-time player in the data management space and a leading data software and services vendor. Informatica's Intelligent Data Management platform provides numerous services to help organizations manage their data at scale.

Specific to the scope of this report, Informatica's platform includes integrated data governance and privacy management capabilities: the Axon Data Governance hub, Informatica Data Privacy Management, and Informatica Data Masking. Axon Data Governance is integrated with the suite of Informatica products and provides a single point from which an organization's governance team can manage the organization's data.

Axon Data Governance draws its metadata store from the Informatica Enterprise Data Catalog. It integrates with Informatica Data Privacy Management to help users define and apply privacy management policies to their enterprise data, as well as apply data encryption and data masking to classify and protect sensitive data. Integration with Informatica Data Quality provides governance teams insight into the trustworthiness of the data, and data and system lineage viewing capabilities. Dashboards track and display costs and revenues relating to specific data items so that analysis of business returns on a particular item can be generated.

Axon Data Governance also provides a unique approach to access controls. It works as follows: from Axon Data Governance, administrators can create preapproved sets of governed data and grant specific users or groups access to these sets. To the end user, the interface appears as a retail-like marketplace called Axon Data Marketplace, displaying pre-packaged sets of data along with their context and owners, as well as the user's specific access privileges to this data. If the user has been granted access to the dataset, they can "order" it and begin working on it.

Informatica also recently announced a new cloud data governance product: Informatica Cloud Data Governance and Catalog, a cloud-native data governance solution that enables Cloud Analytics Governance use cases through collaboration of data professionals and business users.

Strengths: Informatica's strong points include its single-view control panel and the unique marketplace experience of the Axon Data Marketplace, which makes governed data accessible to business users in a retail-like experience.

Challenges: While the integration of Informatica's data governance capabilities with other components of Informatica's larger platform for data intelligence extends the range and robustness of the capabilities offered, this means that data governance capabilities are an embedded functionality within the data intelligence platform rather than a standalone product.

Okera

Okera was founded in 2016 with the idea of helping organizations address the challenges facing them in data access and data governance. The Okera platform enables organizations to derive more insights from their data through improved data access management across hybrid and multi-cloud environments.

The platform provides a consistent user experience that lets data owners and stewards manage data access. Okera's core structural components consist of the Okera Metadata Services, Okera Data Access Service (ODAS), and the Okera Portal.

Within Okera Metadata Services, a no-code policy definition feature allows users of any technical level to write and enforce fine-grained role-based or attribute-based control policies across structured or unstructured data. The policies can be applied on the file, column, row, or cell level. Data de-identification types supported include masking, redaction, tokenization, anonymization, and differential privacy (a function that obfuscates individual values for a column but still allows approximate aggregation of that column based on the actual values), all of which is applied to the data at query time. The platform also provides comprehensive audit logs of all activity performed on data or metadata, such as query requests, policy changes, and the like. Audit data is persisted as JSON files, which can be integrated with SIEM tools like Splunk or Sumologic for real-time monitoring and data breach alerts.

ODAS enables third-party applications and analytics tools to interact with data through the platform. It authorizes queries on-demand, transforming the queries as needed, to allow dynamic masking, tokenization, hiding, and more, depending on the user and data context. The platform allows integration with various retrieval, streaming, and analytics tools like Spark, Python, SQL engines, and notebooks, as well as business intelligence tools like Tableau and Power BI.

The Okera Portal is a web-based user interface from which users can discover datasets, browse dataset content, inspect permissions, and write SQL queries. Okera runs on dedicated clusters, and can be deployed on-premises, in a hybrid cloud model, or in any of the three public clouds. A typical Okera setup involves a single set of shared metadata services and one or more instances of ODAS that are deployed as needed.

Strengths: Okera offers a unified experience for access management; audit logging capabilities; no-code policy definition experience; query brokering; and deployment flexibility.

Challenges: While Okera does provide lineage view capabilities for data assets, its lineage capability currently shows only direct parents and children, and root base table information. Okera states it is improving its lineage capabilities so this deficit may dissipate in time.

Privacera

Founded in 2016 by the creators of Apache Ranger, Privacera extends the Apache Ranger open source technology to provide organizations with solutions to automate their data access control and policy management across multiple cloud services.

The Privacera Platform is the vendor's centralized, integrated data governance and security platform that runs either as a SaaS offering, or in the customer's on-premises or private cloud environment. The platform extends the Apache Ranger functionality across multiple cloud environments. A unified UI is layered above this foundation, resulting in a platform that provides centralized access control, automated data discovery/classification, and data encryption/masking.

Access to data resources can be controlled via roles, resources and tags. And because the Privacera platform allows for the association of attributes with tags, tag-based access provides, in turn, a form of attribute-based access control (ABAC) policy. These fine-grained access management policies can be applied down to the row, column, or file level. Access control policies also can be defined based on physical and logical metadata as well as classifications and tags (business metadata). The solution generates built-in reports and includes dashboards for access governance, audit, and compliance. It also allows users to anonymize and mask sensitive data to further enhance privacy and ensure compliance with global data privacy regulations.

In January 2021, the vendor introduced PrivaceraCloud, a SaaS-based data security and governance platform, which enables faster cloud onboarding and data access governance for hybrid and multi-cloud data services. PrivaceraCloud is a fully managed service, reducing time spent on managing an underlying on-premises infrastructure.

PrivaceraCloud is ideal for enterprises looking to migrate data and analytical workloads from their on-premises Hadoop-based data lakes or from private clouds to the public cloud. PrivaceraCloud covers major cloud-native query engines and data warehouses, including Databricks, Amazon EMR (Hive and Presto), Starburst, Snowflake, Azure Data Lake Storage (ADLS), Azure Synapse Analytics, Amazon Redshift, Amazon Athena, and Amazon S3. PrivaceraCloud supports access control within data engines regardless of where they are being run, as long as there is connectivity between the data engine and PrivaceraCloud. This solution is easy to get up and running quickly: it can be deployed without leveraging container technology, configuring the services, scaling the portal software, or managing operation and uptime of the environment.

A new release of collaboration and governed data-sharing features was imminent as this report went to publication. These features implement a data domain-based approach that gives individual teams authority over data that they own and the sharing of it with data consumers. The platform functionality facilitates this sharing through a dedicated interface or API where users can request access directly from owners.

Strengths: Privacera has a lot going for it, with a foundation based on Ranger open source technology, a single-view, unified platform, automated data discovery and data classification,

deployment flexibility, and the easy-to-use PrivaceraCloud.

Challenges: While PrivaceraCloud sports impressive technology, it is still a relatively new platform, having been introduced at the beginning of 2021, and it will need time to mature. That said, updates (including availability on Azure) have been forthcoming and others, including availability on additional public clouds, are on the roadmap. We await PrivaceraCloud's future developments with interest.

Talend

Founded in 2005 and headquartered in Redwood City, California, Talend produced the first commercial open source data integration software. Since then, Talend has grown into an industry leader in data integration and data quality with its enterprise data management platform.

Talend Data Fabric is the vendor's data platform, which combines data integration and governance capabilities to help organizations ensure their data's trustworthiness. Talend TrustScore is a feature that instantly evaluates a dataset and assigns a score to it, rating the quality and trustworthiness of the data within.

Talend Data Fabric gives organizations a single view across all of its data sources, whether these reside on-premises or in the cloud. Talend's TrustScore helps organizations comply with global privacy regulations, as well as industry-specific regulations. The platform's data masking capabilities anonymize or pseudonymize data.

Strengths: Talend's Data Fabric platform leverages its open source data integration foundation, compatibility with big data platforms including Apache Hadoop and Spark, and strong ML-based data quality capabilities to help its customers achieve their goals.

Challenges: Talend Data Fabric is a complete platform with many strongly integrated data integrity and governance capabilities. However, its data governance capabilities are not presented as a standalone product, but rather as embedded functionalities within the larger Data Fabric Platform.

5. Analyst's Take

This report leverages the decision framework established in the Key Criteria Report for Evaluating Data Governance to provide a technical and operational assessment of vendor solutions in this space. The report incorporates assessment of key criteria and evaluation metrics to inform decision making and provides a forward-looking assessment of products in the category.

Of course, vendors are constantly seeking to improve their offerings—mergers and acquisitions offer the benefit of powerful platform fusion, and the space is actively evolving and innovating. Numerous vendors continue to hone their ML-powered sensitive-data detection capabilities, broaden their capability set overall, introduce more features into their platforms, and open them up to ever more integrations.

All the vendors in the landscape provide robust solutions, and all have differentiating characteristics in which they excel. This report is intended to be used as a guide to help you inform yourself about the scope of available data governance products and technologies, relate these to your organization's needs, and decide on the evaluating criteria and metrics that are most relevant to your organization. Then you can use what you have learned to review the vendor write-ups with a keen and critical eye, focusing on offerings that best match your requirements.

Key Takeaways

- Data governance is an extremely broad topic, encompassing many subtopics and overlapping greatly with the concept of data management. To narrow it down, the scope of this report includes only those aspects of data governance that deal with capabilities such as access controls, data protection, and data lineage.
- The topic of data governance has catapulted to prominence in recent years; first, with the passing of privacy regulations such as GDPR and CCPA, and now with the additional scrutiny of, and heightened awareness around, the management of user data. This scrutiny has intensified due to high-profile data breaches, and the huge growth of online transactions and the user data generated by them, resulting from the pandemic.
- Data governance platforms assist organizations in finding the balance between two competing needs: encouraging data-enabled culture transformation within the organization and ensuring data protection.
- A number of vendors now offer platforms and products for data governance that include automated and ML-based functionality to allow data professionals and business users to manage data access and data protection across the organization, freeing them from legacy manual processes.
- Because of the increased attention surrounding global data privacy regulations in recent years, data governance was seen as a mechanism to reduce risk and exposure. However, most recently, there has been a repositioning of data governance as a value-add, based on the principle that

better-governed data drives even higher business value.

Data governance is becoming increasingly relevant to a growing number of organizations and is evolving rapidly. The solutions assessed in this report deliver important capabilities for complying with regulations, organizing data, and positioning companies to derive value from the data they produce, transact, and store.

6. About Andrew Brust

Andrew Brust has held developer, CTO, analyst, research director, and market strategist positions at organizations ranging from the City of New York and Cap Gemini to Gigaom and Datameer. He has worked with small, medium, and Fortune 1000 clients in numerous industries and with software companies ranging from small ISVs to large clients like Microsoft. The understanding of technology and the way customers use it that resulted from this experience makes his market and product analyses relevant, credible, and empathetic.

Andrew has tracked the Big Data and Analytics industry since its inception, as Gigaom's Research Director and as ZDNet's lead blogger for Big Data and Analytics. Andrew co-chairs Visual Studio Live!, one of the nation's longest-running developer conferences. As a seasoned technical author and speaker in the database field, Andrew understands today's market in the context of its extensive enterprise underpinnings.

7. About GigaOm

GigaOm provides technical, operational, and business advice for IT's strategic digital enterprise and business initiatives. Enterprise business leaders, CIOs, and technology organizations partner with GigaOm for practical, actionable, strategic, and visionary advice for modernizing and transforming their business. GigaOm's advice empowers enterprises to successfully compete in an increasingly complicated business atmosphere that requires a solid understanding of constantly changing customer demands.

GigaOm works directly with enterprises both inside and outside of the IT organization to apply proven research and methodologies designed to avoid pitfalls and roadblocks while balancing risk and innovation. Research methodologies include but are not limited to adoption and benchmarking surveys, use cases, interviews, ROI/TCO, market landscapes, strategic trends, and technical benchmarks. Our analysts possess 20+ years of experience advising a spectrum of clients from early adopters to mainstream enterprises.

GigaOm's perspective is that of the unbiased enterprise practitioner. Through this perspective, GigaOm connects with engaged and loyal subscribers on a deep and meaningful level.

8. Copyright