The Rise of Data Access Governance in the Insurance Industry

Whitepaper
July 2021

Content

The explosive proliferation of data has transformed the way business operates today. Data supports decision making, performance reporting, predictive analysis, process improvement, and makes a case for strategic and innovation initiatives, among others. Using data as an asset, it is no longer fictional for businesses to achieve the ultimate objective of being "brilliant in every moment" in their operations. However, with great power comes great responsibility, and many would add, even greater challenges. Data is generated at unprecedented speed; figuring out the what, how, and where data is stored and accessed is a challenge that led to the rise of big data and cloud computing in the past decade. Until now, data has sprawled across on-premises data lakes and cloud services, along with integrations to analytical databases, data visualization, and business intelligence tools. Therefore, the coordination of people, processes, and tools that enables an organization to discover and understand its data, control access to data, and ensure that data is trusted, compliant, and fit for consumption becomes of the utmost importance. That's where data governance comes into play.

1   A Closer Look into the Insurance Industry

Insurance providers that practice data governance focus primarily on data related to structured policy, claims, underwriting, and finance. These types of data are fundamental to the business with respect to claims processing, risk assessment, and policy administration and pricing. At the same time, most insurance organizations overlook effective data governance for both unstructured and third-party/external data.

According to a study by Novarica[1], only 15% of survey respondents indicated that unstructured data from documents and internal systems are part of the data governance program in their respective organizations. As a result, the lack of a comprehensive data governance strategy represents missed opportunities and growing concerns in the following areas for the insurance industry:

1.1   Lack of Self-Service Analytics Impedes Business Agility

Self-service analytics refers to the ability that allows business analysts and lines of business to access data and perform analysis and reporting without the active involvement of IT. This is important in today's insurance landscape where customers demand convenience, similar to an Amazon- or Netflix-type experience, where products and services are only a few clicks away. To create stellar

---

1 Data Governance: Current State, Objectives, And Challenges, Novarica, October 2019

customer experience, insurers need to achieve business agility across their insurance operations by enabling data to the right persons at the right time for the right use, such as allowing:

⊕ claims processors to access claims and historical information freely

⊕ underwriters to utilize third-party/external data to assess risk and make informed decisions

⊕ actuaries and product management to acquire necessary data to determine risk tolerance and build profitable insurance products

⊕ data scientists to have the visibility of available data to perform exploratory analysis, train machine learning models, and enable straight-through processing and automation

The self-served nature accelerates the time-to-value of analytics initiatives and helps transform insurance organizations to make data-driven decisions. The challenge lies with implementing an effective access governance framework to ensure authorized access by users to produce accurate analyses and reports. For example, it is a common enterprise practice for the same data to be copied multiple times by different groups. These copies are then manipulated and transformed according to the needs of each group. Over time, copies of data become the "source of truth" for each of the groups, leading to varying definitions of data, conflicting reporting results, an overall distrust of data sources, and worst yet, contradicting or imprecise data insights resulting in poor decision making. Without fine-grained access controls in place, data administrators do not have the visibility into groups, users, or access privileges, resulting in the organization losing trust in the integrity of its data.

1.2 Ungoverned Use of Sensitive Data Results in Violations of Regulations

The enforcement of privacy regulations such as GDPR in Europe, LGPD in Brazil, CCPA and NYPA in the US, and more to come across the state, federal, and international level, has become a major driver of insurers' data governance strategies to control the use of sensitive information. Some of the sensitive data can offer risk-differentiating insights that lead to a significant improvement on loss ratio. For example, in the small business segment of the Property and Casualty (P&C) insurance market, underwriting has been increasingly

challenging, as many small business risks are loss-free and mostly look the same based on the meager information provided in standardized application forms. The good news is that the use of external and sensitive data has served as the leading indicator that paves the way for better underwriting of small business risk. Depending on the different insurance lines, external data sources such as the following can be highly predictive of loss:

General Liability Insurance		Workers' Compensation Insurance		Employment Practices Liability Insurance	
Search Engine Activity	Area Lifestyle e.g., Smoking	Payment Data	Prior Safety Violations	Prior Wage Practice Violations	Social Media Sentiment
Back Wages Owed	Healthcare Affordability	Customer Ratings	Policy Data	Employee Ratings	Civic Engagement
OSHA Inspections	Social Vulnerability	Reserve Data	Web Reputability	Voting Patterns	Attorney Prevelance

The privacy mandates formalize insurers' responsibility for the use of high-value sensitive data as a resource for analysis. Insurers require improved identification and access control of all sensitive data, of which personally identifiable information (PII) is only a subset. A comprehensive data governance solution that encompasses the following aspects would serve the need of understanding the existence, classification, usage, location, and protection of sensitive data:

⊕ Automated data discovery to detect sensitive data for proper classifications

⊕ Granular access controls based on usage, roles, and attributes

⊕ Scalable data masking and encryption to protect PII and other sensitive information

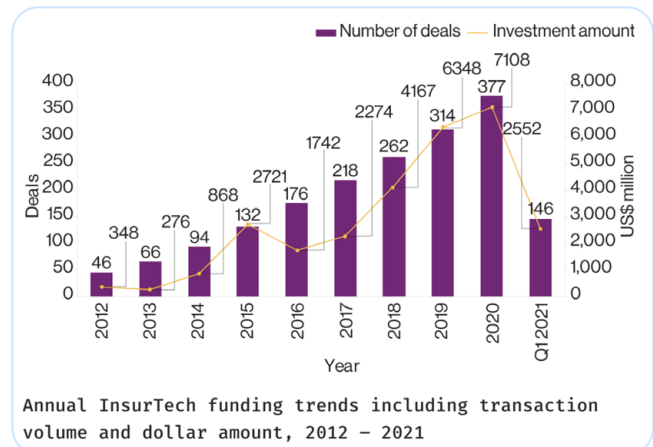**Without effective data governance, the consequences are often twofold:**

1 Data or security officers employ an overly cautious approach to managing privacy and compliance policies, resulting in a strict and often unnecessary restriction on the use of data. The conservative approach can be an immediate roadblock to short-term data analytics projects and may have a negative effect on long-term innovation initiatives.

2 Violations of data privacy and usage regulations result in severe penalties that not only damage the company's reputation and brand, but also generate customer mistrust.

Either scenario represents a major setback for any insurance organization and highlights the importance of a comprehensive data governance framework.

1.3 The Great Migration to the Cloud Introduces New Complexity

Digital transformation is a driving force of change in the insurance industry. Starting out as an IT strategy to reduce infrastructure costs and increase efficiency, cloud computing is the game-changer that transforms the insurance industry, where legacy systems dominate and rigid technology systems prevent

Annual InsurTech funding trends including transaction volume and dollar amount, 2012 – 2021

Source: Quarterly InsurTech Briefing Q1 2021, Willis Towers Watsons

insurers to meet customer demand and truly innovate. Moreover, the continued investments in insurance technology have spurred new competitions and intensified internal rivalries on customer reach, pricing, and service offerings. For example, the public offering of Lemonade, a mobile-based insurance provider built on a new concept that uses artificial intelligence and chatbots to process claims in seconds, was one of the top IPO debuts in 2020 with an impressive 139% increase on its initial stock price. This indicates the public's acceptance and confidence in up-and-coming insurance technology while putting immense pressure on insurers that are traditionally slow to adapt to the change of technology.

questions@privacera.com | privacera.com | 510.413.7300   privacera   @privacera

Moving to the cloud is the start of digital transformation, but not nearly the end. Many insurers pace themselves by gradually incorporating the concept into their IT stack, such as building their own software on cloud technologies or hosting applications on cloud infrastructure. Executives are dazzled by the quick return on investment as a result of the significant reduction of capital expenditure on building and maintaining data centers. The increase in business agility and cost-efficiency validates the value of cloud and propels a migration of more critical data and business-essential functions to this "promise land," also known as the cloud migration. However, the benefits don't come without a cost. Cloud migration introduces new and growing complexity, and insurers' IT teams are tasked with the conquest.

For example, Amazon Web Services (AWS) – the pioneer and leader in cloud computing – started out with three foundational services: Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Storage Service (Amazon S3), and Amazon Simple Queue Services (SQS). It now offers more than 212 services for storage, computing, databases, analytics, application services, and Internet of Things (IoT). When migrating to the cloud, insurers can quickly lose sight of their data landscape when data spans across endpoints, on-premises infrastructure, one or multiple cloud environments, and across regions in different countries with varying data privacy mandates. In addition, the distinct identity and access management (IAM) tools and processes varied by cloud providers and cloud-native solutions create another layer of complexity to navigate data access. Without a consolidated view of data and unified access governance across different platforms, an insurer's cloud migration strategy is ripe for failure.

2 Conquer Your Data Access Governance Challenge with Privacera

At the prime intersection of data privacy, security, access control and governance, Privacera is the access governance platform that liberates secure data sharing to the universe of available data across all your data environments. Built by the same team that developed Apache Ranger, a proven centralized policy administration that manages security policies across the compute engines of Hadoop data lakes, Privacera has extended Ranger's foundational access control capability to cloud services and emerged as a leading access governance platform built for the cloud. Integrating data discovery, access control, anonymization, and encryption capabilities in one, Privacera delivers the most comprehensive access governance platform that empowers:

⊕ data administrators to discover, define, and enforce data access policies across heterogeneous cloud and on-premises data services from a single interface

⊕ data security staff to increase protection for zero-trust networks and reduce risk of data breaches

⊕ data consumers, such as business analysts and data scientists, to quickly access reliable and trusted data for analytical projects and innovation initiatives

⊕ data privacy officers to gain a centralized view of sensitive data across the organization to ensure enterprise-wide privacy for data regulations such as GDPR, CCPA, LGPD, etc.

questions@privacera.com | privacera.com | 510.413.7300   privacera   @privacera
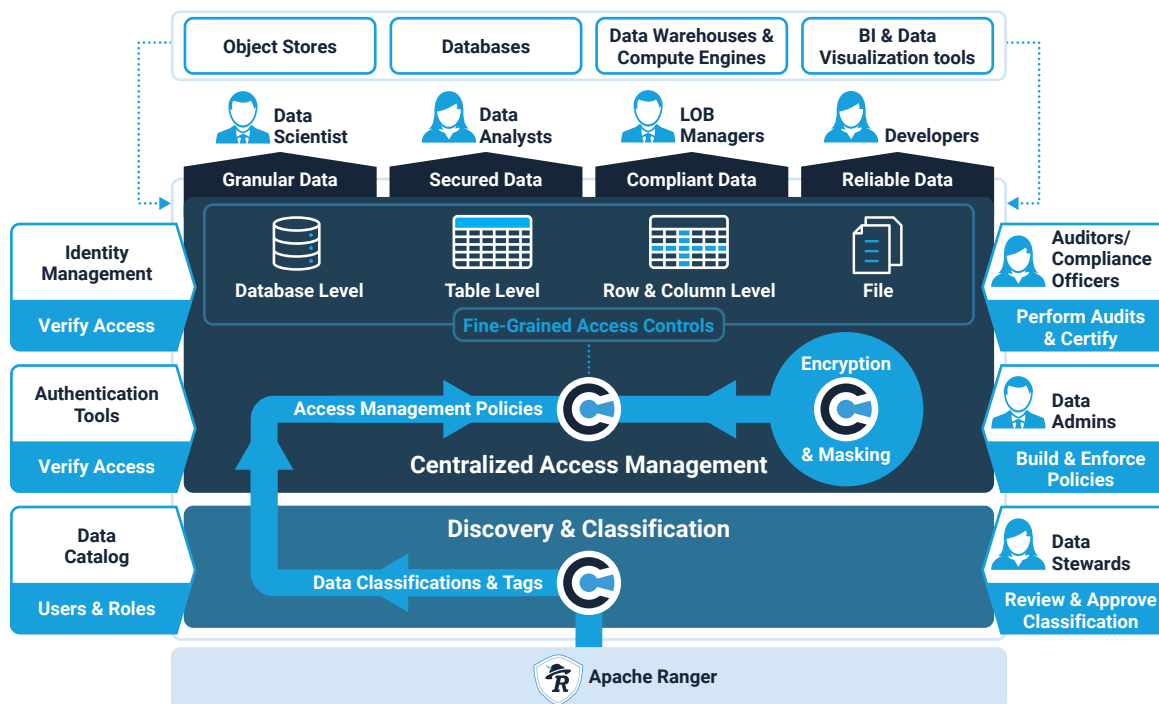
2.1 Uncover Sensitive Data to Ensure Compliance with Automated Data Discovery

Privacera's Data Discovery module leverages sophisticated rules, pattern matching, dictionaries, algorithms, and machine learning models to understand the context of sensitive data and accurately classify it. Sensitive data is scanned, identified, and tagged in real-time as it is uploaded to the cloud or object storage. With Privacera's out-of-the-box reporting, customizable data usage reports and auditing, and alerts when sensitive data is accessed or moved, infrastructure teams have instant visibility of their data assets.

2.2 Administer Universal Access Authorization with Centralized Policy Management

Once data is classified, the right type of controls must be implemented. These controls are more effective if they are administered from a central location. It is extremely important for administrators to consider if a single platform can be used to administer access control policies across services.

The Privacera Platform provides administrators with a centralized management interface to define and administer data access policies for on-premises data lakes, public cloud services, as well as third-party cloud-native services such as Databricks,

AWS	Azure	GCP	3rd Party Services
Amazon Athena	Azure Data Lake Storage	Google Cloud Storage	Databricks
Amazon Redshift	Azure Data Bricks	Google Big Query	Starburst
Amazon EMR	Azure HDInsight	Google Big Table	Snowflake
Amazon DynamoDB	Azure Synapse Analytics	Google Data Proc	Qubole
Amazon Glue	Power BI on Azure		Cazena
Amazon Kinesis	Azure SQL Server		Dremio
Amazon RDS	Azure Postgres		Talend
Amazon S3			Streamsets

Snowflake, and others from a single console. Privacera scripts and utilities provide easy configuration of enforcement points across cloud and on-premises data services.

questions@privacera.com | privacera.com | 510.413.7300 | privacera | @privacera

2.3 Govern Data Sharing to Enable Self-Service Analytics with Fine-Grained Access Controls

The richer the access control platform's ability to administer policies to finer grains of data, the easier it is for infrastructure administrators to grant access to the precise data users need to do their jobs. Privacera's Access Control module provides administrators the flexibility to define access policies at a database, table, column, or file level. With Privacera, administrators can build access policies based on roles, attributes, and assigned tags. Users from lightweight directory access protocol (LDAP) or active directories can be associated with specific organizational roles, which can then be assigned access privileges or permissions. Rules-based on dynamic conditions, such as time or geography, can also be added to an existing policy rule.

2.4 Protect Sensitive Data with Dynamic Data Masking and Encryption

It is not enough for a data governance platform to simply discover sensitive data and apply access control policies to secure it. Data infrastructure teams must also provide mechanisms for data scientists and analysts to extract insights from regulated data. This requires masking sensitive data, prior to making it available for analysis and restricting users' access to specific rows based on organizational role or attribute. The Privacera Platform provides dynamic data masking capabilities to protect sensitive content in a variety of flexible formats. This capability enables only authorized users to see data they are permitted to see, while the same data is masked or anonymized for other users or groups. Masking policies can be used to define which specific data fields are masked and how to anonymize or pseudonymous specific data.

Enterprise data must be protected while it is in motion or at rest. Privacera encryption gateway (PEG) is a robust, scalable application programming interface (API) gateway that protects customers' sensitive data and personally identifiable information, without the need for manual processes or operational burden. PEG provides flexible mapping schemes and policy-based encryption and decryption using NIST standards-based encryption algorithms, such as AES-128, AES-256, hashing, and format preserving encryption (FPE).

Fortune 100 Life Insurance Company Processes 500,000 Requests Per Day Across 200 EMR Nodes with Privacera

One of the largest global life insurance companies had a vast amount of on-premises and cloud data stored across its AWS infrastructure (S3 and EMR) that contained sensitive customer data and personally identifiable information (PII). When the insurer initiated its on-premises to cloud data migration, it needed a centralized view of all sensitive cloud data and automated fine-grained access controls. This would ensure sensitive data was protected against unauthorized access and to decrease the manual complexity for its data teams. Extending EMR's native Apache Ranger integration, the Privacera Platform offered the customer a seamless way to leverage its existing Apache Ranger investments in the cloud, securely migrate data without exposing sensitive elements, and implement consistent fine-grained data access controls across its EMR and S3 environments from a single, centralized location.

The Challenges

(+) Migrate Apache-based security to AWS environment

(+) Support data policies compliant with International Financial Reporting Standards (IFRS)

(+) Keep PII safe with encryption or masking

(+) Automate assignment of data policies and discovery and intake of customer data

The Solution

(+) The insurer began using Privacera for security, compliance, and data management across 200 Amazon EMR nodes in cloud and on-prem environments

(+) Privacera's user-friendly admin console helps with monitoring and management tasks

The Results

(+) Privacera currently processes 500,000 requests per day for the customer, invoking about 4,000 policies as needed, for highly granular data access control and masking of data distributed across multiple cloud instances

(+) Instead of taking months to manually recreate existing data governance policies in the cloud, the insurer can use Privacera to seamlessly manage a unified approach across both on-prem Hadoop and cloud

(+) The Privacera platform allows the customer to extend its existing Apache Ranger foundation to the cloud as a unified solution across both environments

questions@privacera.com | privacera.com | 510.413.7300 privacera @privacera

3   Looking at the Big Picture by Developing a Plan to the End

Having secure data access has never been as critical to an enterprise as it is today. It is an involved project that spans multiple teams, systems, and years to complete. Before embarking upon any accelerated implementation, it is critical to clearly understand the full scope of requirements across geographies, systems, and teams, without adversely affecting end-user requirements for scalable system performance.

Proven at scale in 2000+ production environments, Privacera can help you develop a holistic approach to identify where sensitive data resides, its characteristics, and the life cycle and end-users' use cases for the data to prevent any missteps in implementing a scalable, enterprise-wide solution.