LightBeam.ai

Data Privacy Automation for Structured Databases

Reference Architecture

Table of Contents

Executive Summary

One of the key requirements of today's privacy regulations and protecting personal information (PI) from unauthorized use and disclosure lies in understanding and managing the use of personal information within an organization's data environment. Spread across a multitude of repositories, cloud environments, and application databases, PI use can be difficult to manage through written policy alone.

We at LightBeam.ai believe the best way to implement policy across an organization is to supplement written policy with technical controls designed for specific databases, applications, and functions. By working with our clients we have developed several database, application, and function-specific controls focused on discovering, cataloging, and enforcing control over the use of personal information within popular datasets.

Traditionally a database is an organized collection of structured information. The information or data is stored electronically in hardware storage devices in personal computers, On-prem network shared drives, or cloud-based storage and processing services. A database is usually controlled by a database management system (DBMS). Together, the Data, DBMS, and the application software define the business processing activities,

Similar to spreadsheets, structured databases are typically tables made up of rows and columns. Organizing data in this fashion makes analyzing and processing data fast and efficient.

Centrally located and designed to store information in easy-to-use sharing platforms, structured databases are at the heart of business processes. However, in addition to the ease of use, the ability to create, copy, and distribute data across an environment can create new data leakage risks for an organization. The unchecked creation and duplication of files in large repositories has long been an issue with Privacy Teams trying to understand and control the use of PI in their organizations

Our AI-driven platform engine, Spectra, can easily be configured to analyze structured databases and automatically discover, analyze, and enforce privacy policies regarding the use of sensitive information. The inappropriate use of or changes to the database architecture can trigger an alert-workflow process which includes data owners and business owners who can easily either reject or approve the processing. If approved the alert is placed in a permit list for future processing authorization. Changes to database schemas and retention requirements can be easily monitored. So privacy risk is reduced and retention requirements for data that is no longer needed are identified.

By automating the execution of control policies, Privacy Officers can develop custom rule sets that continually scan, monitor, and control how PI is used and controlled within the organization's data storage systems. The details of how this happens are discussed below.

Introduction

Audience

This document is intended for organizations that have databases whose processing uses personal information. It is meant for both technical and non-technical audiences. Privacy Officers, CISOs/Security Architects, and Support leaders within organizations overseeing the use of PI and Google Drive will find this reference architecture useful in automating data privacy controls.

Purpose

This document provides greater details on the problem of processing personal information within structured databases and how LightBeam can be used to manage the use of PI and reduce the risk posed by data duplications or inappropriate and long-term storage of PI. Although this document is primarily about structured databases in general it applies in principle to the various types of databases including SQL-server, PostgreSQL, AWS S2, and other traditional databases.

Database Overview

Databases have a rich and long history across the business landscape, and they represent the primary storage and processing method for many of today's companies. As such databases are a breeding ground for duplicate copies of data. From gold source repositories, data is duplicated for a multitude of new projects that can span the entire business organization. As more and more copies of the same data are duplicated business continually creates new projects with new data. Security and privacy teams can become blind to the duplication of sensitive personal information they are responsible for without a clear view of the data behind the scenes. This kind of data sprawl creates a clear risk of data leakage. By monitoring the spread of data at the database level, a better understanding of the organization's data environment can be established.

There are many different types of databases. The best database for a specific organization depends on how the organization intends to use the data. Databases have evolved dramatically since their inception in the early 1960s and still exist in many different types including. Relational databases, Object-oriented databases, Distributed databases, Data warehouses, NoSQL databases, Graph databases, Open source databases, Cloud databases, Multi-Model databases, Document/JSON databases, Self-driving databases. LightBeam can connect to any of these types of databases through API connectors.

LightBeam Data Privacy Automation Platform

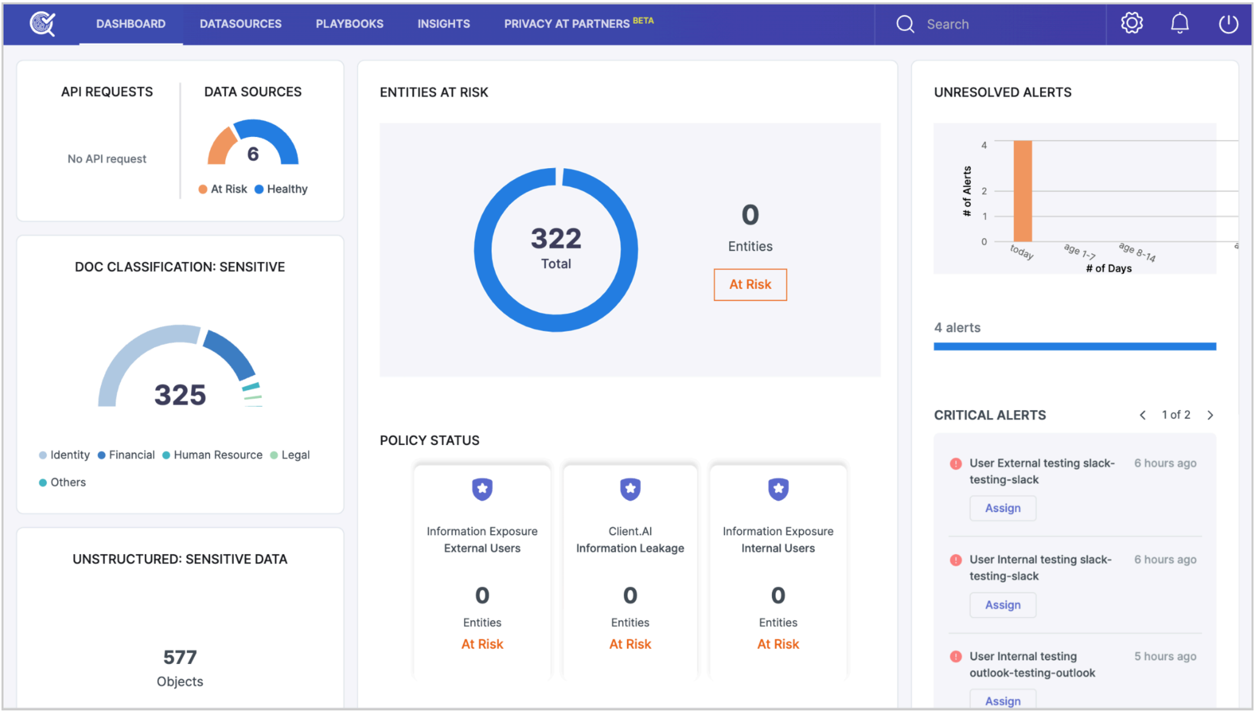
A pioneer in the data privacy automation (DPA) category, LightBeam is on a mission to empower organizations to manage and access their sensitive data securely. Leveraging its identity-centric discovery & classification engine, Spectra, LightBeam ties data discovery and cataloging, access, and sharing into a unified privacy control platform.

LightBeam empowers privacy and compliance executives to keep their organizations under continuous compliance for GDPR, CCPA, HIPAA, and PCI-DSS among others, while information security executives can finally rest assured that sensitive data is being used, accessed, and stored securely. LightBeam's 360 view of the data environment provides an up-to-date accurate dashboard of data sources, data attributes, identities, control policies, and permission lists.

Here is a quick look at how LightBeam brings unparalleled observability to the data an organization carries within its myriad data repositories.

Main Dashboard

The main LightBeam dashboard provides a high-level view of all structured and unstructured databases, and other data sources where sensitive data is present, The Entities (customers, employees, patients et al) whose sensitive data, defined here as The dashboard includes the data attributes or specific elements which are being processed. Processing rules and alerts representing data usage, and violations that need attention.



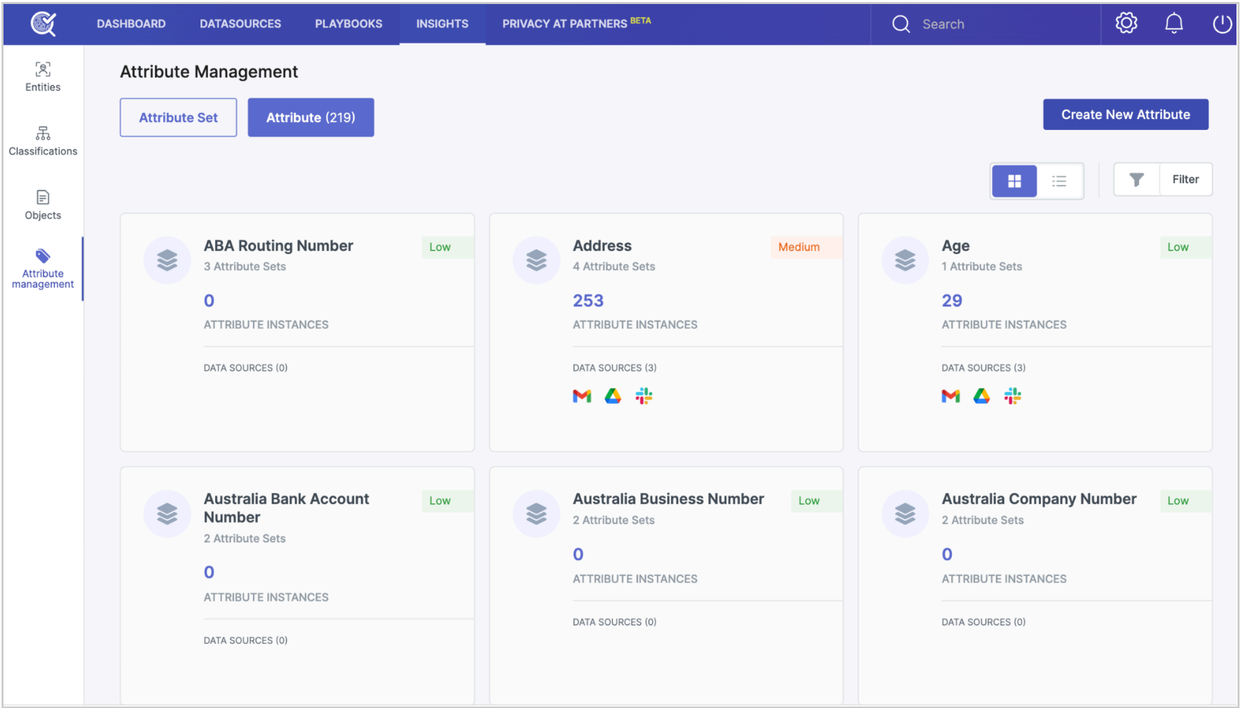
User External testing slack-testing-slack 6 hours ago

User Internal testing slack-testing-slack 6 hours ago

User Internal testing outlook-testing-outlook 5 hours ago

Attribute View

LightBeam recognizes over 200 pre-configured sensitive data elements, or attributes, in the databases, and is capable of recognizing the identifiers, which are called values that have been discovered in connected data sources. Additional client-specific proprietary attributes that are unique to their organizations can be added and learned by the system. New attributes are assimilated into the system contextually based on interactions with other connected data sources. Attributes carry sensitivity levels based on their risk and type in the system (i.e. High, Medium & Low.) Examples of attributes are U.S. Social Security Number, Loan Account Number, Medical Record Number, and so on. The attribute screen also shows in which database or application the attribute was found



Attribute Management

ABA Routing Number — Low
3 Attribute Sets
0
ATTRIBUTE INSTANCES
DATA SOURCES (0)

Address — Medium
4 Attribute Sets
253
ATTRIBUTE INSTANCES
DATA SOURCES (3)

Age — Low
1 Attribute Sets
29
ATTRIBUTE INSTANCES
DATA SOURCES (3)

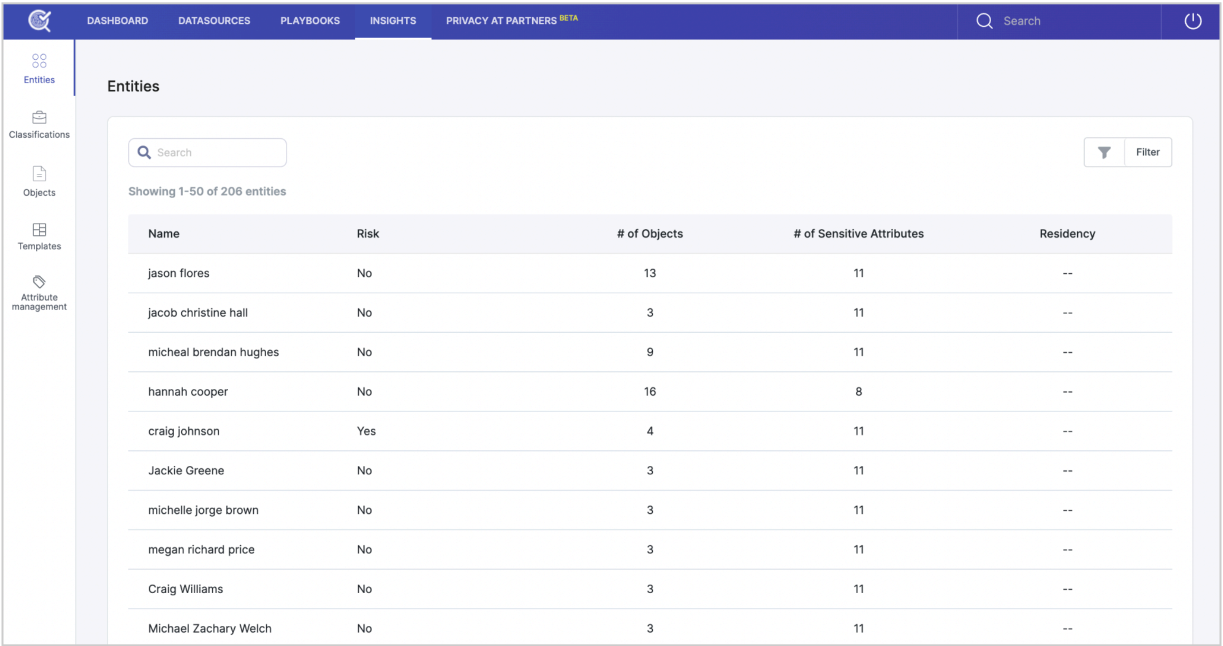
Australia Bank Account Number — Low
2 Attribute Sets
0
ATTRIBUTE INSTANCES
DATA SOURCES (0)

Australia Business Number — Low
2 Attribute Sets
0
ATTRIBUTE INSTANCES
DATA SOURCES (0)

Australia Company Number — Low
2 Attribute Sets
0
ATTRIBUTE INSTANCES
DATA SOURCES (0)

Entity View

Centered on the individual, the entity view provides a precise breakdown of what data is being held for any individual, in what data sources it is stored, and if any known associated risks have been identified. The entity view is central in supporting GDPR, CCPA, and other regulations where the ability to gather and provide an individual's data to fulfill an individual rights request is a requirement.



Entities

Showing 1-50 of 206 entities

Name	Risk	# of Objects	# of Sensitive Attributes	Residency
jason flores	No	13	11	--
jacob christine hall	No	3	11	--
micheal brendan hughes	No	9	11	--
hannah cooper	No	16	8	--
craig johnson	Yes	4	11	--
Jackie Greene	No	3	11	--
michelle jorge brown	No	3	11	--
megan richard price	No	3	11	--
Craig Williams	No	3	11	--
Michael Zachary Welch	No	3	11	--

LightBeam Operational Phases

LightBeam's Spectra DPA platform employs a three-phase approach to managing privacy risk. These phases include Detect, Enforce, and Automate. Each of these phases builds on the previous phase to create a fully automated privacy management system that can;

1. Understand the existence and use of PI
2. Create control policies with resulting actions.
3. Create automated tasks to execute control policies.

Detect

LightBeam's initial step is to gain an understanding of the data environment. This includes connecting to the applications and databases and discovering sensitive data elements also known as "attributes." Attributes are contained in applications and repositories and are duplicated across the environment based on the relevant business processes. LightBeam uses API connections to analyze structured and unstructured repositories and identify the data attributes, attribute types, the related sensitivity levels, and connects all data to an identified individual or "entity".

By understanding which data exists in the environment, LightBeam learns which data is important to an organization and its business processes. With this understanding as a foundation, LightBeam is able to then set policies as to how that data is stored, shared, and viewed.

During the detect phase, LightBeam natively recognizes and classifies

- 200+ common attributes including the common identifiers from a variety of countries.
- Industry attribute type sets like (Financial, Healthcare, Identity…)
- Unlimited client-specific attributes - every LightBeam customer is unique and may carry sensitive data that is unique to them. LightBeam enables customers to add custom attributes. (e.g. Customer account number, employee number, member numbers, SKUs, and other values)
- Document classification of type and sensitivity of data contained in attachments in a service ticket.
- Sensitive attributes detected across multiple data repositories are linked using a machine learning algorithm to see if they belong to a single entity. The cross-linking of fragments of information to a central identity is a unique capability that helps customers understand not only if sensitive data is at risk but more importantly, whose data it is that might be at risk.
- The DETECT phase helps create data maps, RoPA reports and a 360-degree view of all information that's present about customers within an organization's systems.

Enforce

The enforce phase is used to establish the rules for data usage. There are four primary control components in the enforce phase. These include Policies, Permissions, Alerts, and Redaction.

Policies

LightBeam Policies are configured to track both internal and external data sources. Each policy may contain multiple rule sets that define the search criteria and details about the data including; attribute sets and types, data sources, alert level settings, and the associated relevant regulations, and are configured via a query selection screen.
Policies include:

- Policy types of Internal, External, and Leakage.
- The contact information for whom an alert should be sent to.
- The setting of the permissions list and no scan for whitelisting of approved data sources.

Permissions

- Permission lists (also sometimes referred to as Permit Lists) establish and maintain an inventory of approved repositories and uses of PI
  - Approved repositories are added to a permission list and used to compare new scan results.
  - Alerts can be raised when a new instance is not found on a permission list.
  - Workflows are initiated by an alert to approve new instances and update the permission list so future findings will not raise an alert.

Alerts

- Alerts are used to notify system owners and others that a policy has been violated and that action may be needed.
  - Alerts are triggered based on individual policy rule sets.
  - Alerts can be set for specific applications or all connected applications.
  - Alerts can trigger a workflow to drive a review and approve cycle

Automate

Utilizing LightBeams DPA technology to execute privacy process controls to continually scan, monitor, and control data usage is a powerful tool for today's Privacy, Compliance, IT, and IT Security teams. Automated monitoring of IT system controls has long been a part of most modern IT and security programs. Now the monitoring of sensitive data usage through automated processes greatly expands visibility, control, and understanding of an organization's sensitive data use across the data lifecycle.

LightBeam ties together sensitive data cataloging, control, and compliance across structured and unstructured data applications providing 360-visibility, redaction, self-service DSRs, and automated ROPA reporting ensuring ultimate protection against ransomware and accidental exposures while meeting data privacy obligations efficiently.

Guided by LightBeam’s established policies, the scanning engine, Spectra, continuously scans the data environment looking for changes to the data. New copies and uses can quickly be identified and either added to a permissions list or raised for review By automating the execution of enforcement controls of alerting continually, an always-on accurate inventory of personal information is created. The process maintains an identity-centric index that can be used to facilitate the retrieval of an Individual's PI and aid in the processing of Individual Rights Requests. Additionally, DPA allows for duplicated data to be easily monitored and controlled reducing data leakage, And changes that add new databases of PI, PII, and SPI are identified.

LightBeam Data Privacy Automation for Structured Databases

LightBeam customers can configure policies to trigger actions to manage the use of PI stored in structured databases. The use case is centered on database tables and the PI stored in them. For connected Databases, LightBeam can display a view of the database architecture including views for Overview, Tables, Columns, Clusters, Schemas, and the sensitive data residing there.

Spectra can discover PI in Dbs tables, columns, and rows. LightBeam identifies over 200 common national attributes as well as any proprietary client custom attributes stored in structured databases(i.e.SKU, Part numbers, etc) LightBeam will scan all tables containing PI and based on the automation policy and rule set that is in force the systems will take steps to raise an alert and initiate a review and approve workflow when changes are made to tables. Adding new sensitive attribute fields, or if new columns are added to a table that contains sensitive attributes a PIA can be raised.

Connecting a Database to LightBeam

Connecting databases to LightBeam is a simple 3 step process. LighBeam has created connection processes for most common databases including. Oracle, Snowflake, MS SQL, MySQL, PostgreSQL, and many others. Refer to the Data Source page for a complete listing. The three connection steps include Basic Information,

Connection Information, and Configuration settings.

1. Basic Information
   The Basic Information page includes the instance names, names of database owners, data location, and data purpose fields.

2. Connection
   The connection screen sets parameters for the actual connection including
the
   Login credentials, domain information, and API token connection information.

3. Configuration

The Configuration screen sets the scanning parameters including what to scan and what to exclude from scanning.

For a detailed description of the database connection information refer to the connection documentation for the specific database type you are using.

Conclusion

Managing the appropriate use of personal information is challenging for any organization. Administrative controls like policies, procedures, and employee training are only as good as their execution which is often an afterthought in many organizations. By using the power of AI to diligently scan and monitor data applications and repositories, data officers can apply significant technical control over how data is stored and processed. This in turn provides privacy teams new accurate pictures of how sensitive data is used in the organization. By understanding all sensitive data in an organization LightBeams 360 degree view allows for more advanced reviews and proactive actions to be taken to manage privacy operations and reduce overall privacy risk.

Appendix

Revision History

Document Update	**Date**	**Author(s)**
Generally available. Updates for database	July 22, 2022 11/30/2022 2/6/2024	Bill Schaumann bill@lightbeam.ai Deepak Jha deepak@lightbeam.ai

About LightBeam

With its focus on Data Privacy Automation (DPA), LightBeam is pioneering a unique identity-centric and automation-first approach to data privacy and data security markets. Unlike siloed solutions, LightBeam's Data Privacy Automation (DPA) ties together sensitive data discovery, cataloging, access, and data loss prevention (DLP), and makes the right (sensitive) identity-centric data available to the right people and teams. It becomes the privacy control tower providing a 360-degree view of PII/PHI sensitive data sprawl. LightBeam enables privacy officers to set policies to automate their enforcement, while information security executives can finally rest assured that sensitive data is being used and accessed securely.

For any questions or suggestions, please contact us at: sales@lightbeam.ai.