LightBeam.ai

Data Privacy Automation for NoSQL databases

Reference Architecture

Table of Contents

Executive Summary

The key to meeting the 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 and application data sets, 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 procedures for technical controls designed for specific applications and functions. By working with our clients we have developed several applications and function-specific controls focusing on discovering, analyzing, and enforcing control over the use of personal information within popular applications and storage options.

Personal information is present in many forms in today's organizations. In addition to traditional storage systems like databases spreadsheets, and files, PI can exist in chat applications, ticketing systems, and images. LightBeam's Spectra engine can detect PII in almost any form it is connected to, however, databases remain the most popular place to store PI.

Traditionally, relational databases that contain tables and rows were the standard technology since the mid-80s when databases like DB4, MS Access, and Foxpro dominated the landscape. However, in 1998, the term and the concept of a NoSQL database was invented by Carl Strozz, in order to designate his lightweight and open-source relational database. The concept was then adopted and popularized by the large social media and tech companies that were processing huge volumes of data. They saw that the traditional relational database structure had become too

slow and opted for the high availability, fast execution of queries, and efficient management of big data features that NoSQL schemas brought to the table.

Our AI-driven platform engine, Spectra, can easily be configured to monitor PI used in NoSQL databases like Mongo and automatically discover, analyze, and enforce privacy policies regarding the use of sensitive information stored there. By finding and either raising alerts, redacting, or deleting files that inappropriately contain PI, organizations can reduce privacy risk and meet retention requirements for data that is no longer needed. By then automating the execution of these control policies, Privacy and Security teams can develop custom rule sets that continually scan, monitor, and control how PI is used and controlled within NoSQL datasets. The details of how this happens are discussed below.

Introduction

Audience

This document is intended for organizations that have implemented NoSQL database technologies and whose information 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 NoSQL Databases 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 NoSQL databases and how LightBeam can be used to manage the use of PI and reduce the risk posed by file duplications, inappropriate use, and long-term storage of PI.

NoSQL Overview

NoSQL databases are non-tabular databases that store data differently than relational tables. NoSQL databases allow developers to store huge amounts of unstructured data, giving them a lot of flexibility in development. NoSQL databases do not have tables and row schemas in the same rigid way that relational databases do. NoSQL databases store data in documents rather than relational tables. Accordingly, we classify them as "not only SQL" and subdivide them by a variety of flexible data models. Types of NoSQL databases include pure document databases, key-value stores, wide-column databases, and graph databases.

NoSQL databases do not have a schema in the same rigid way that relational databases have a schema. Schema-less databases are a type of NoSQL database that does not require a predefined schema to store data. Instead, they allow data to be stored in flexible and dynamic formats, such as JSON documents, key-value pairs, graphs, or columns.

NoSQL APIs provide access to the NoSQL database service, which enables the storing, processing, and consuming of data in tabular formats. Unified data models enable the use of NoSQL APIs to fetch and update any object in the system. Through the API connection developers can add items to a table, replace an existing item, create tables, update items, retrieve items or attributes, or delete tables and items. LightBeam leverages these capabilities combined with detailed control policies to monitor and control the use of PI in NoSQL databases.

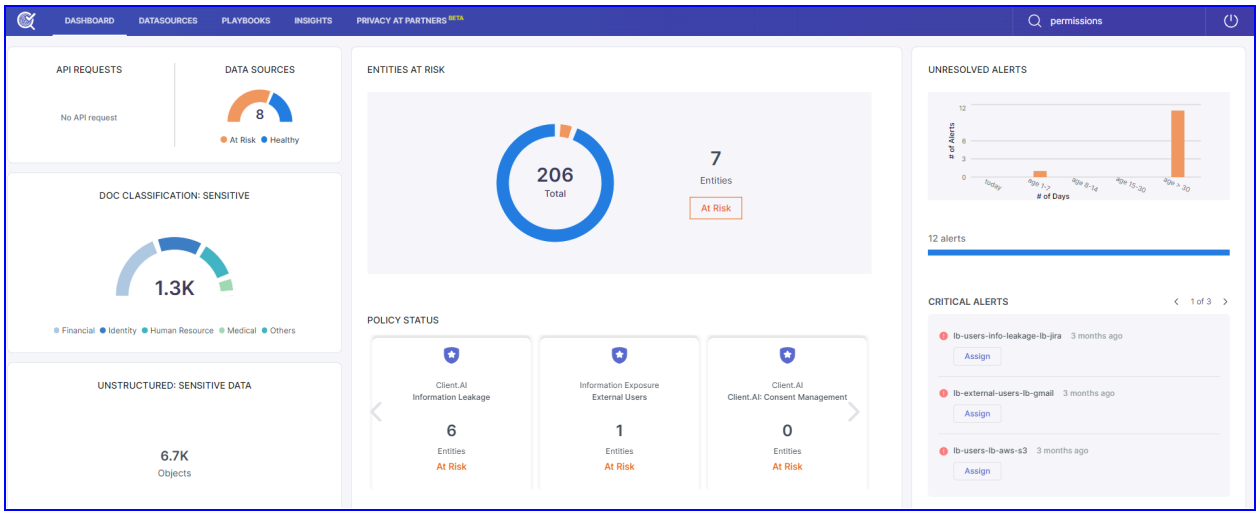
LightBeam Data Privacy Automation Platform

A pioneer in the data privacy automation (DPA) category, LightBeam is on a mission to empower organizations to access and manage their PI and SPI securely. Leveraging its identity-centric discovery & classification engine, Spectra, LightBeam ties data 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, entities (identities), control policies, and permission lists. The following is a quick look at how LightBeam brings an unparalleled view of PI and SPI carried in today's organizations within a myriad of data repositories.

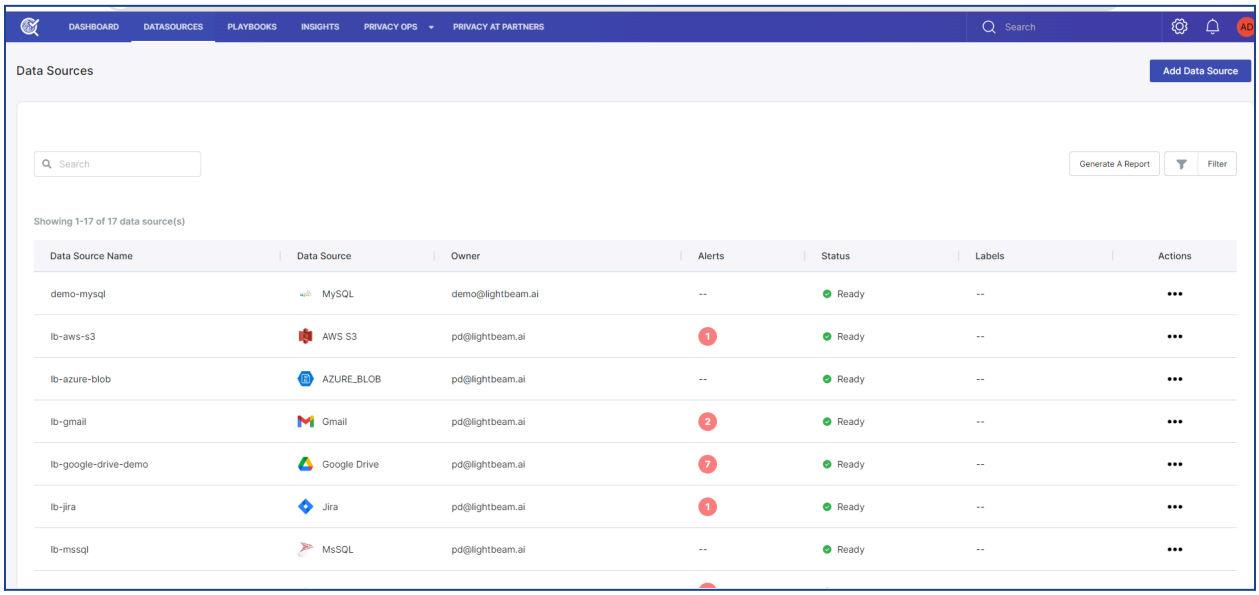
Main Dashboard

The main dashboard provides a high-level view of all data sources where sensitive data is present, the entities (customers/employees/patients et al) whose sensitive data is being carried, and any alerts that might need attention.



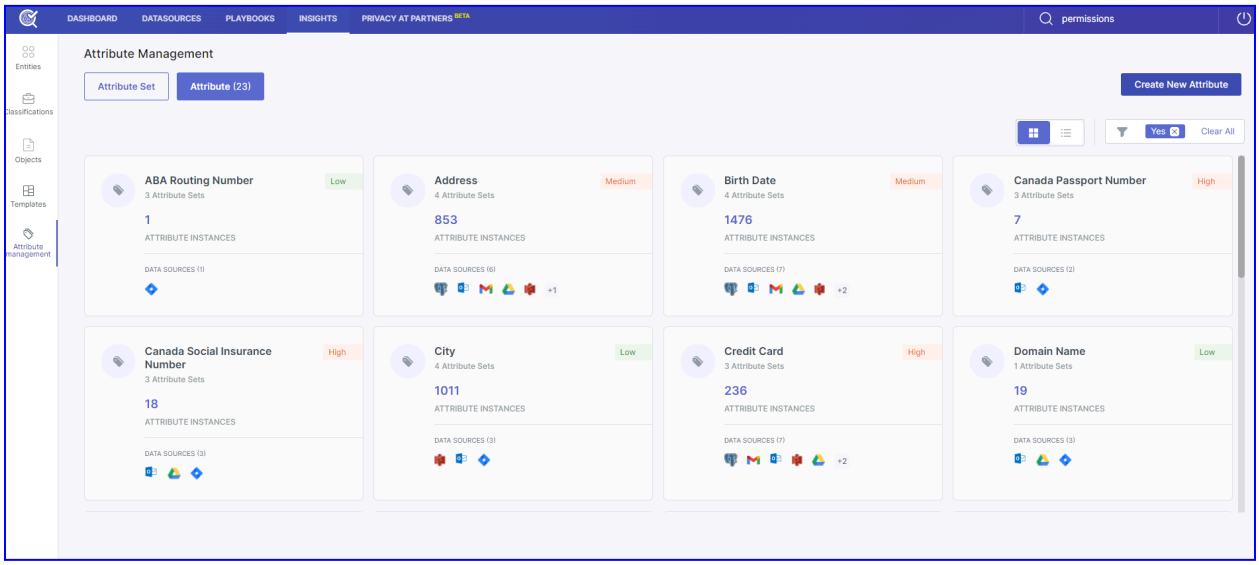
DataSource View

LightBeam can connect to a large number of data sources. Network storage, Cloud Storage, applications, databases, and NoSQL databases all can be connected and scanned for the presence of PI. The DataSource view provides a complete picture of the PI attributes, their sensitivity levels, and the current status of the PI contained in any select dataset. Further policies can be created to scan the data sources and raise workflow alerts or take actions to secure PI



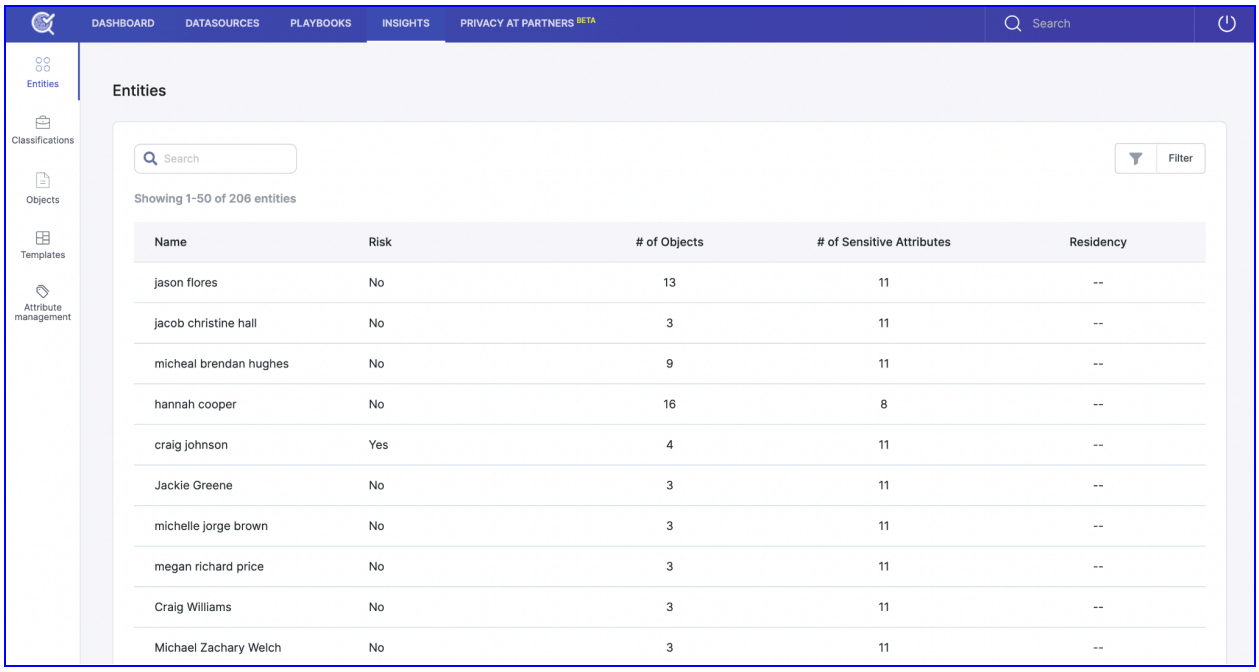
Attribute View

LightBeam has over 200 pre-configured sensitive attributes (sometimes called fields/columns) in its system and is capable of recognizing their identifiers (sometimes called Values) from all the data sources; moreover, users can also add their own attributes to the system and make it learn from various sources.
These attributes have 3 sensitivity levels based on their weight 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.



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, and if there are any known associated risks. This view supports GDPR, CCPA, and other individual rights requests.



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 repositories and discovering sensitive data elements called "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. Then, an Entity is resolved from the data related to an identified individual or “entity”.

By understanding the data source and entity data that exists in the environment the LightBeam platform 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:

- Types of policies. policy types include Internal, External, and Leakage.
- The contact information for whom an alert should be sent.
- The setting of the permissions list for whitelisting approved data sources.

White Listed data sources marked as gold source repositories are helpful in architecting a complete data schema.

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 rule sets inside policies.
    - Alerts can be set for specific applications or all connected applications.
    - Alerts can trigger a workflow to drive a review and approve cycle

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 alerts raised. By automating the execution of enforcement controls like alerting and redacting 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.

Data Privacy Automation for NoSQL

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 on a continual basis, 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.

Connecting a NoSQL database to LightBeam

Connecting any NoSQL Database to LightBeam is a simple 3-step process.

1. Create a new data source instance.
2. Connect to the new data source.
3. Configure scan preferences

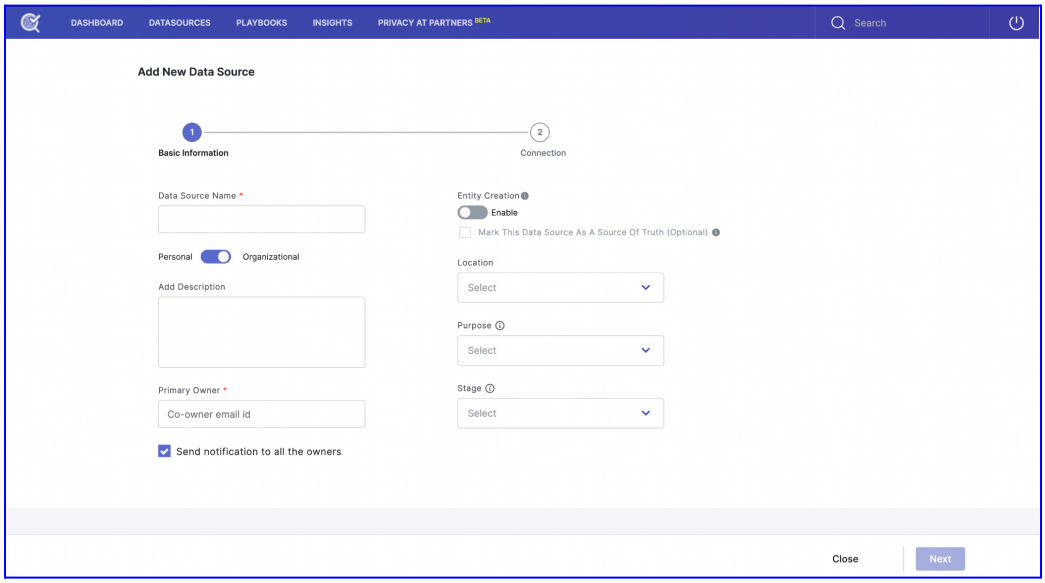
To begin from the LightBeam dashboard, select DATA SOURCES. From the DataSource home screen select Add Data Source from the top right of the page. From the application list select New NoSQL database and continue with the steps below.

Step 1:

Complete the Basic Information page to create a new data source. Keep the following guidance in mind when creating a new data source:

1. Make sure the name is not repeated, as it acts as metadata for the data source and must be unique. E.g., NoSQL_Sandbox_Alpha
2. Use the description to explain the kind of information the data source contains. E.g., All HR related documents stored here
3. LightBeam uses the email ID stated as the Primary Owner to send alert notifications.
4. You can add another email for notifications using the co-owner button. Remember to check the 'send notification to all owners' box.

5. Entity Creation: Enable to create entities out of this data source.
6. Location tells where the data source is located.
7. Purpose tells for what purpose is this data source storing or processing data.
8. Stage tells what stage the data belongs in could be the source, processing, transactional, archival, etc.



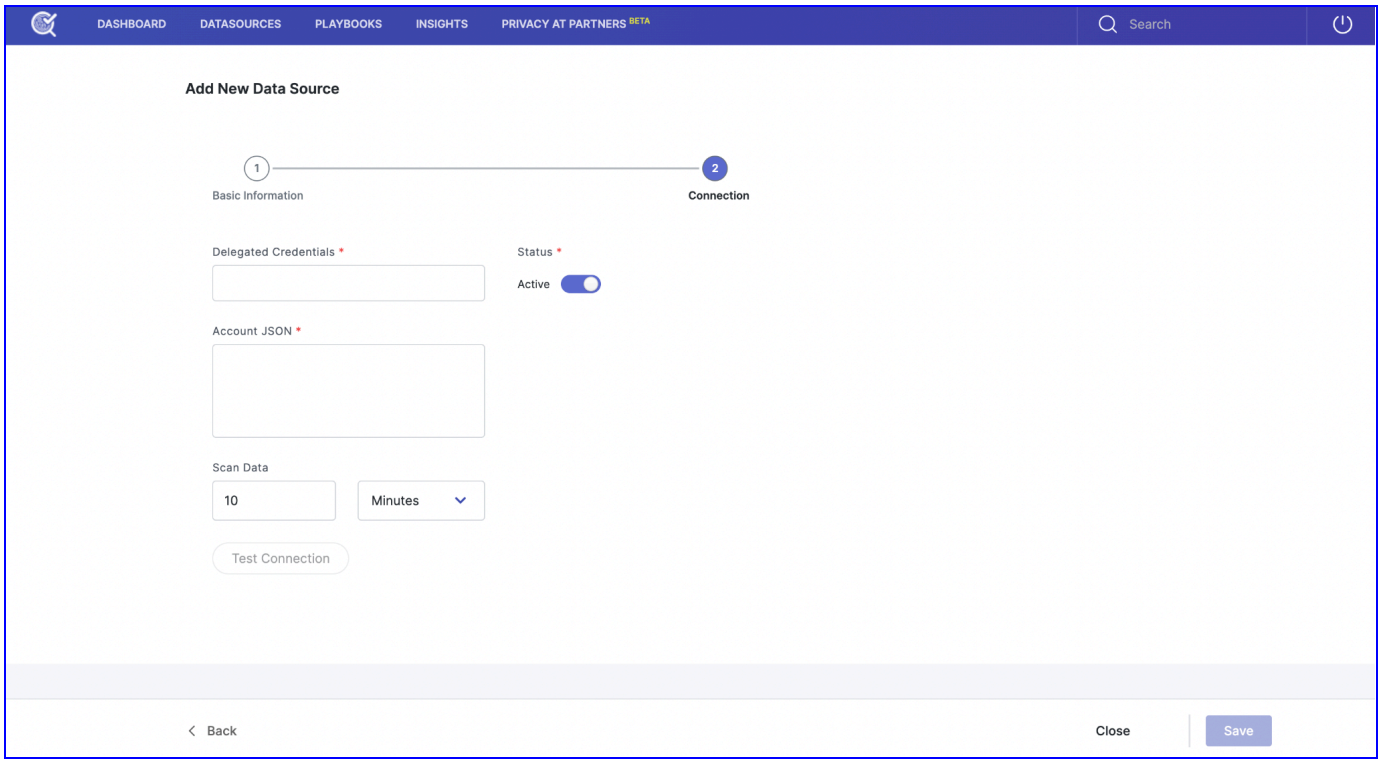
Step 2: Connection Details

1. Select the start date from when you want to start scanning the tickets
2. Add subdomain name
3. Configure API Token for authentication

Add the Email ID from which the API token has been created

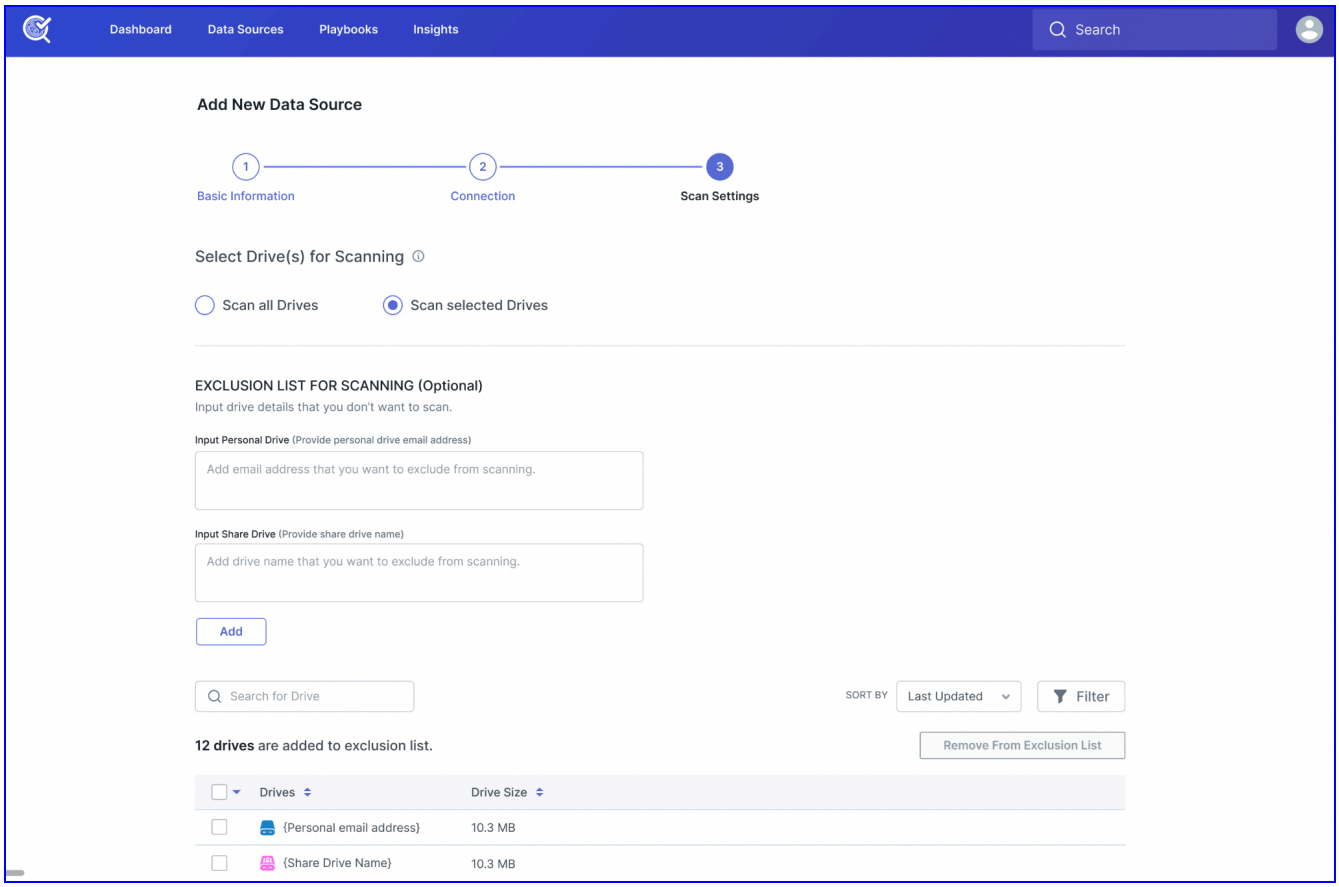
Add API token

You can test the connection and see if the connection really went through



Step 3: Scan Settings

1. Scan all drives
2. Scan selected drives
3. Exclusion list
    a. Add personal drives
    b. Add shared drives



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 with new accurate pictures of their sensitive data and how it 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

Reference Architecture Update	Date	Author
First generally available version.	11/30/2023	Bill Schaumann

About LightBeam

With its focus on Data Privacy Automation (DPA), LightBeam is pioneering a unique identity-centric and automation-first approach to the 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.