LightBeam.ai™

Data Privacy Automation for File Repositories

Reference Architecture

Table of Contents

Table of Contents **3**
Executive Summary **4**
Introduction **5**
Audience 5
Purpose 5
LightBeam Data Privacy Automation Platform **7**
Main Dashboard 8
DataSource View 8
Attribute View 9
Entity View 10
Operational Phases **11**
Detect 11
Enforce 13
Polices 13
Permissions 13
Alerts 14
Data Privacy Automation for Cloud Based Storage Systems **14**
Connecting - AWS S3 and Microsoft Azure 15
Connecting LightBeam to AWS S3 16
Connecting LightBeam to Microsoft Azure **20**
Conclusion **23**
Appendix **24**
Revision History 24
About LightBeam 24

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. Files from desktop applications contain unstructured data of all kinds. Unstructured data can include text files like legal documents, audio files, chats, videos, images, text on a web page, spreadsheets, word files, PDF, and image files. As this data is not stored in organized or structured databases it is sometimes more difficult to monitor the presence of sensitive data elements in individual files.

The LightBeam AI driven platform engine, Spectra, can be configured to monitor PI use in local network file repositories, or in cloud-based storage systems. Spectra will automatically discover, analyze, and enforce privacy policies regarding the use of sensitive information no matter where it is stored. 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 automating the execution of these control policies through custom rule

sets to continually scan, monitor, and control how PI is used, privacy risk can be actively managed. The details of how this happens are discussed below.

Introduction

Audience

This document is intended for organizations that have network storage drives and implemented technologies 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 network storage will find this reference architecture useful in automating data privacy controls.

Purpose

This document provides greater details on the problem of processing personal information in structured and unstructured files in cloud-based environments. 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 in unsecured files.

Network Storage Overview

Most organizations provide their users some kind of storage area to store their working files in. Traditionally, desktop system files were stored in local drives or locally hosted Network Area Storage devices (NAS). Network storage consisted of shared file folder systems that were organized in groups allowing teams to easily manage access to departmental level files. However, the advent of cloud computing has created a variety of new storage systems that do not store files in local repositories. The leading cloud providers; Amazon, and Microsoft provide cloud storage systems with many capabilities that go far beyond traditional data storage.

Amazon Simple Storage Services or S3 and Microsoft Azure Files are cloud-based products that more and more organizations are leveraging for their enterprise computing and storage needs which provide both cost savings and heightened security along with additional utility services.

Amazon S3

Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. Customers of all sizes and industries can use Amazon S3 to store and protect any amount of data for a range of use cases, such as data lakes, websites, mobile applications, backup and restore, archive, enterprise applications, IoT devices, and big data analytics. Amazon S3 provides management features so that you can optimize, organize, and configure access to your data to meet your specific business, organizational, and compliance requirements.

Microsoft Azure Files

Azure Files offers fully managed file shares in the cloud that are accessible via the industry standard protocols. Azure Files can be used to replace or supplement traditional on-premises file servers or network-attached storage (NAS) devices. Popular operating systems such as Windows, macOS, and Linux can directly mount Azure file shares wherever they are in the world. Azure Files makes it easy to "lift and shift" applications to the cloud that expect a file share to store file application or user data.

Google Cloud Storage

Google Cloud Storage is a service for storing data objects in a Google Cloud. Any amount of data can be stored and retrieved as often as is needed. An object is an immutable piece of data consisting of a file of any format. Objects are stored in containers called buckets. Buckets can also contain managed folders, which you use to provide expanded access to groups of objects with a shared name prefix. Once established additional security services can be added to the cloud storage system.

Blob Storage

A BLOB (binary large object) is a varying-length binary string that can be up to 2,147,483,647 characters long. Like other binary types, BLOB strings are not associated with a code page. In addition, BLOB strings do not hold character data. Common examples of files stored in a BLOB data type field include: Images (JPG, JPEG, PNG, GIF, HEIC, WEBP, raw binary data) Videos (MP4, AVI, MOV, MKV) Audio files (MP3, WAV, AAC)

AWS object storage comes in the form of Amazon S3. Azure object storage is available with Azure Blob Storage. Google Cloud Storage also includes Blob storage. Amazon S3 and Azure Blob Storage, and Google Cloud Services are massively scalable object storage services for unstructured data.

LightBeam Data Privacy Automation Platform

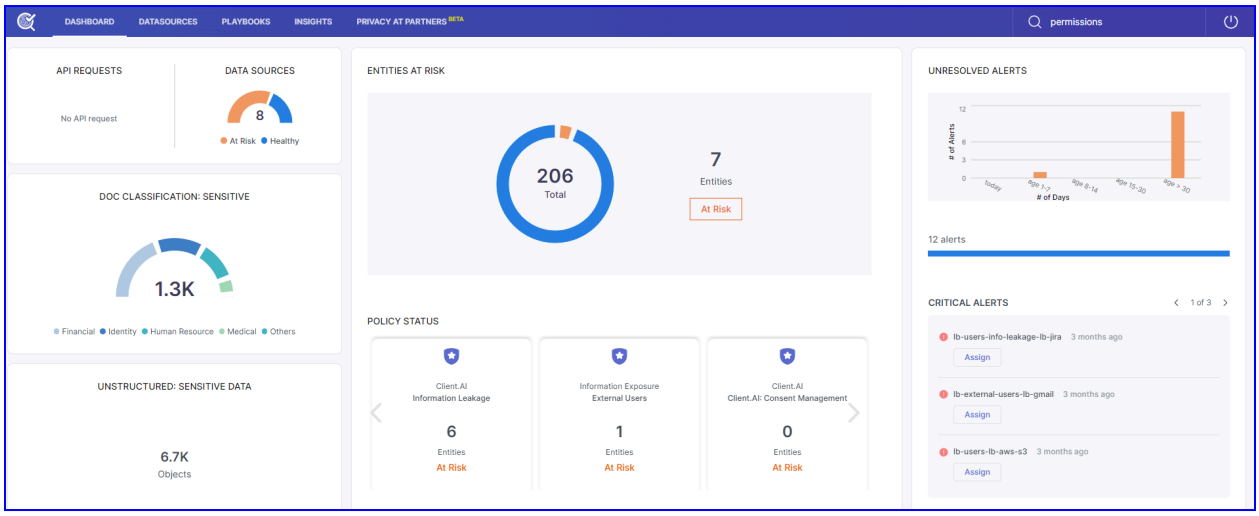
A pioneer in the data privacy automation (DPA) category, LightBeam is on a mission to empower organizations to access and manage their PI and SPI securely. Leveraging its identity-centric discovery & classification engine, Spectra, LightBeam ties data cataloging, access, and sharing into a unified privacy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

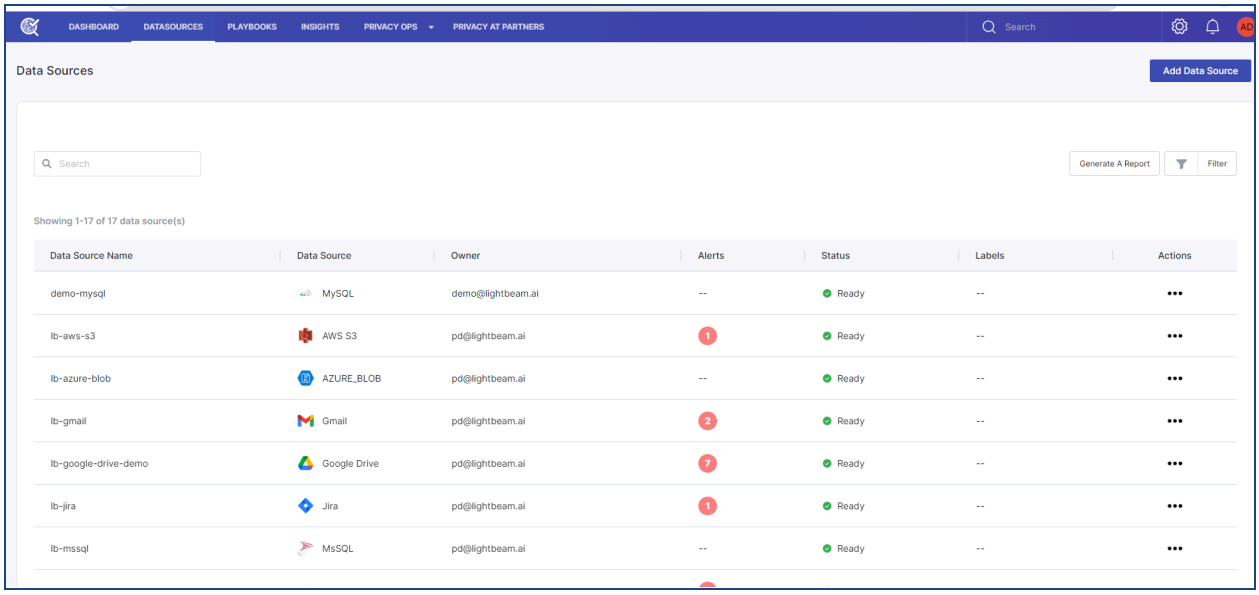
Main Dashboard

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.



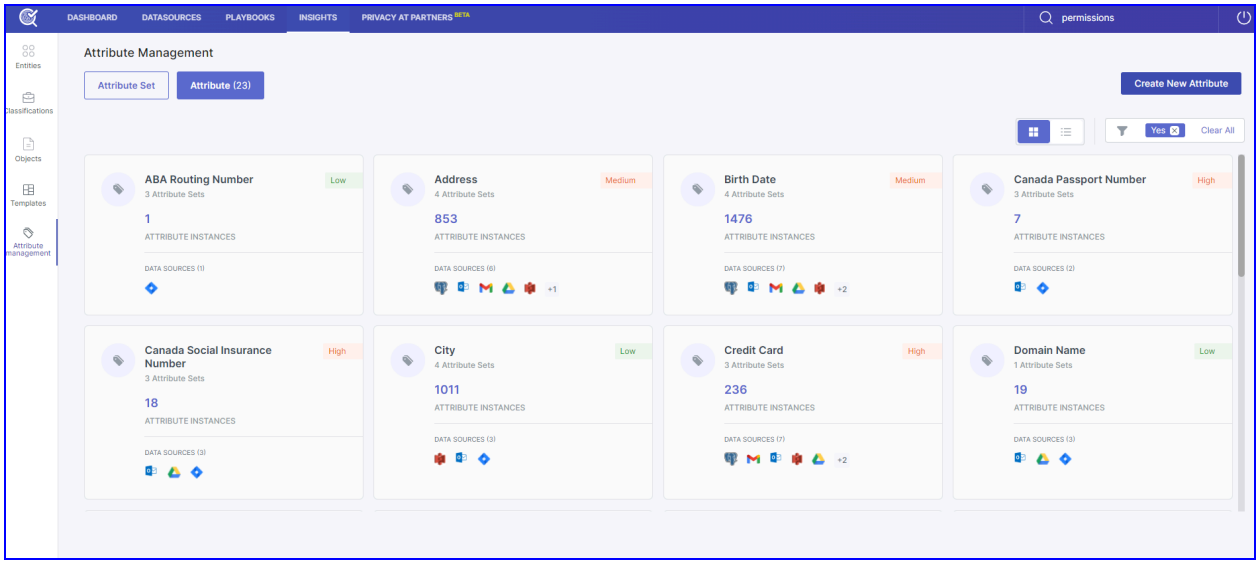
DataSource View

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



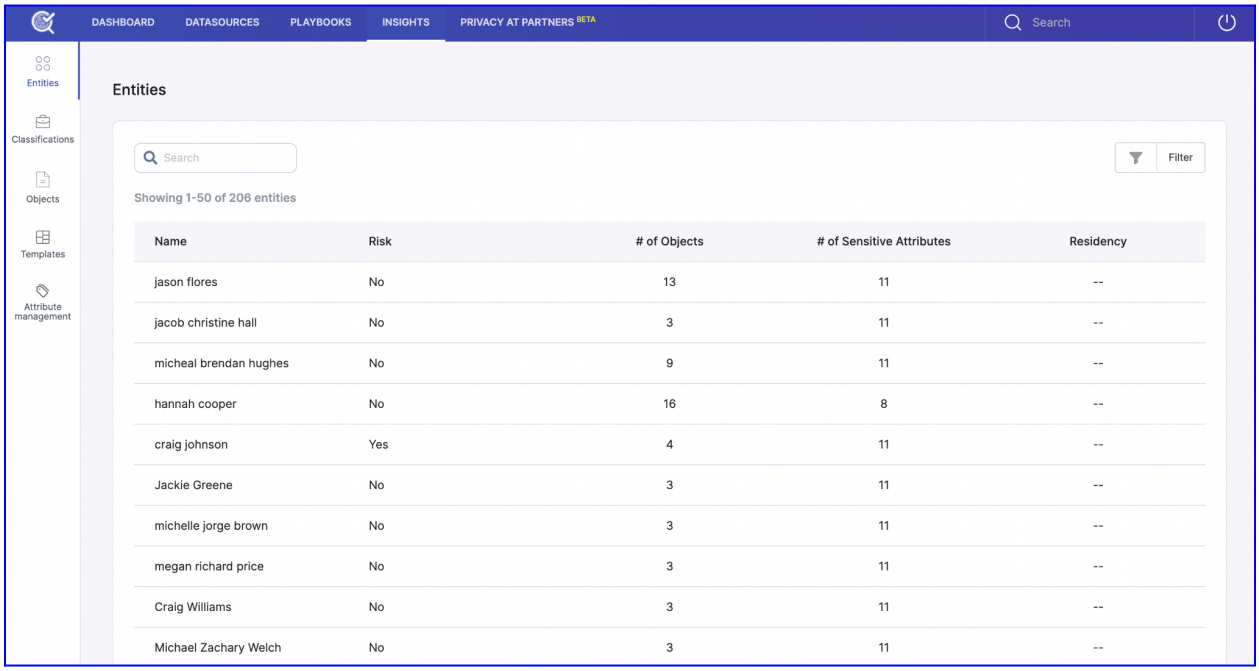
Attribute View

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 setting, 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 who an alert should be sent to.
- The setting of 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 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.

Data Privacy Automation for Cloud-Based Storage Systems

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 - AWS S3 and Microsoft Azure

This section covers connecting and configuring LightBeam to previously established and connected AWS S3 and Microsoft Azure instances. For additional information on setting up cloud services refer to additional LightBeam installation documentation.

Connecting cloud services to LightBeam

Connecting any new data source 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 AWS S3 bucket and continue with the steps below.

Connecting LightBeam to AWS S3

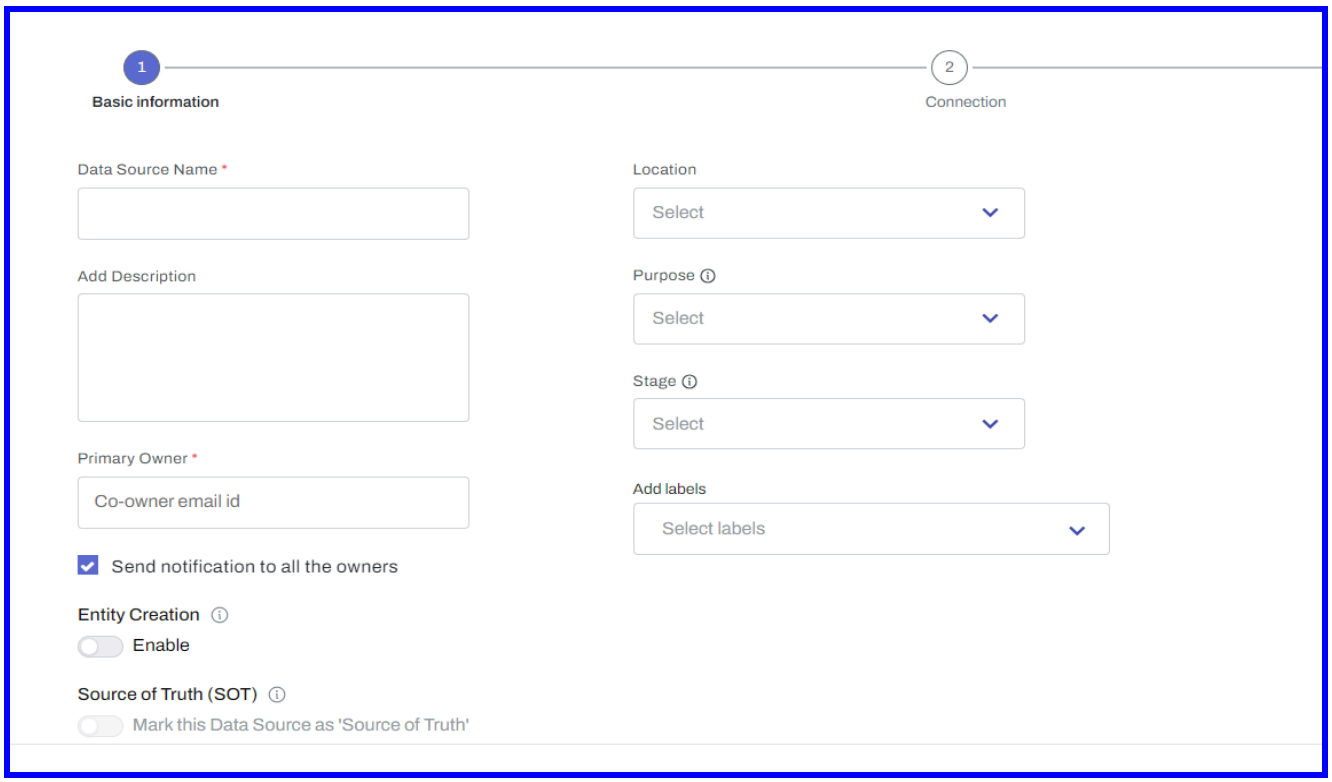
Start: From the Add New Data Source screen scroll to find data source (AWS S3), Select AWS S3.

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., LB_AWSS3_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.

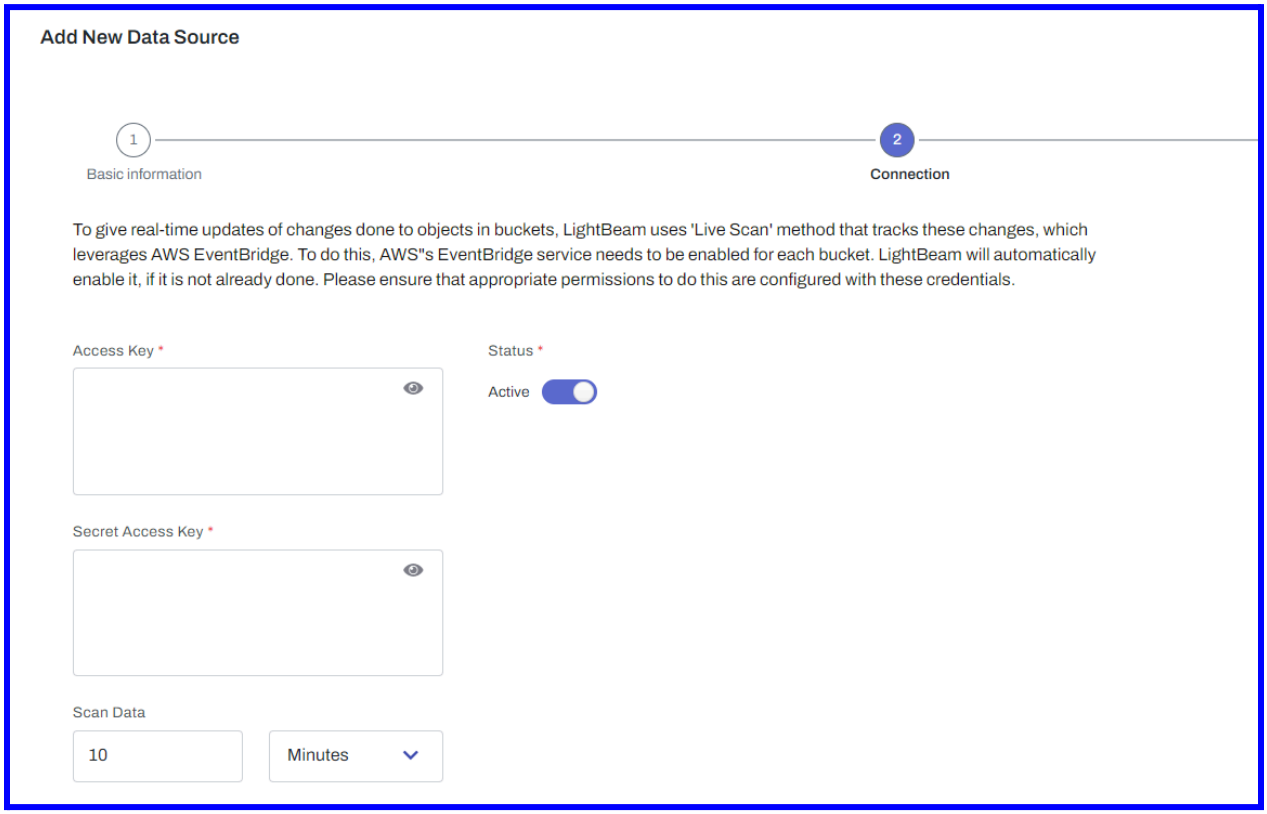
4. You can add another email for notifications using the co-owner button. Remember to check the ‘send notification to all owners’ box.
5. Entity Creation: Enable to create entities out of this data source.
6. Location tells where the data source is located.
7. Purpose tells for what purpose is this data source storing or processing data for.
8. Stage tells what stage the data belongs in could be source, processing, transactional, archival, etc.
9. Click on Next to proceed



Step 2: Connection Details

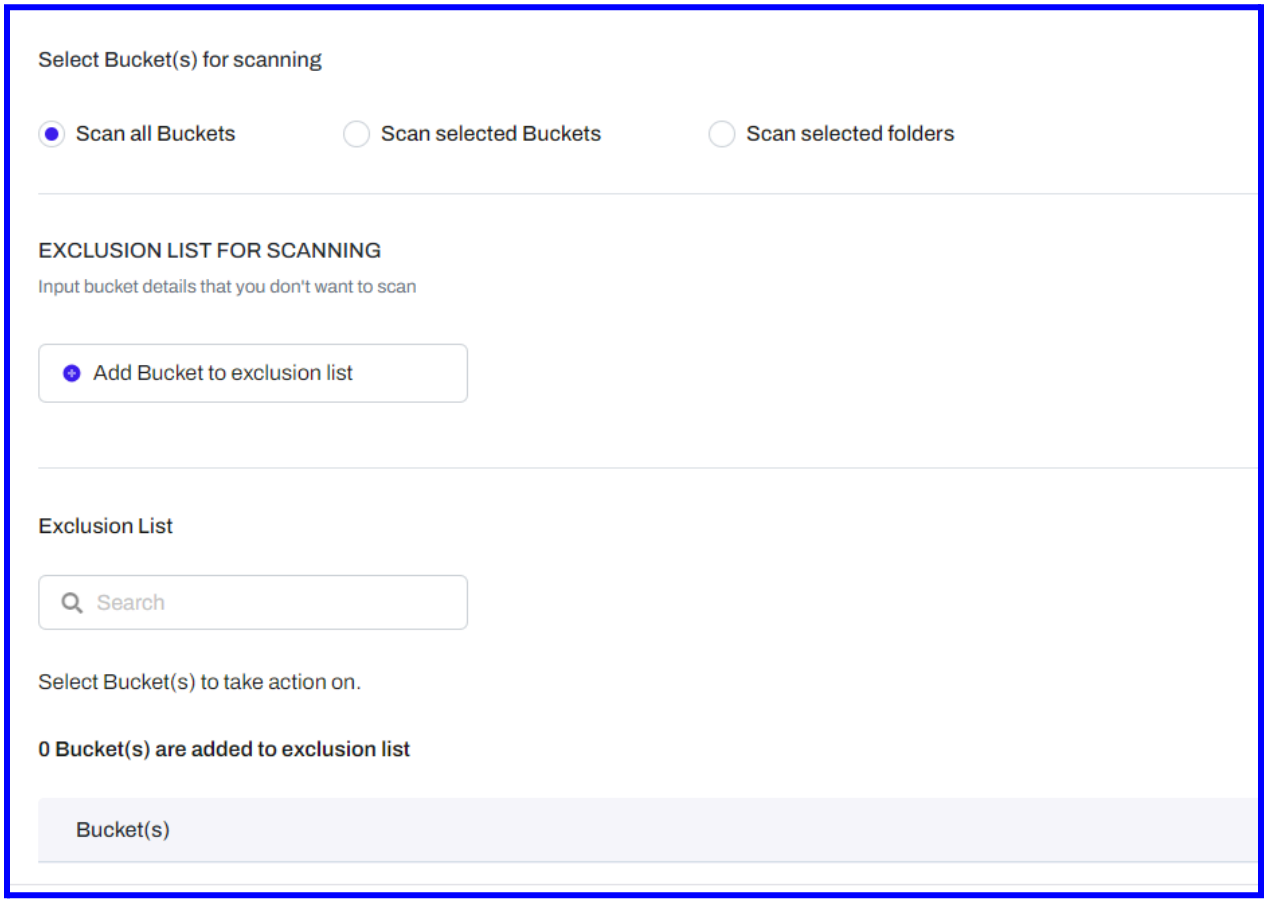
1. Input Access Key
2. Input Secret Access Key
3. Select time to scan the repository
4. Set status to active
5. Click on Next to proceed

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



Step 3: Scan Settings

1. Scan all buckets
2. Scan selected buckets
3. Exclusion list
    a. Add/Exclude buckets for scanning
4. Click on Validate and Save



Connecting LightBeam to Microsoft Azure

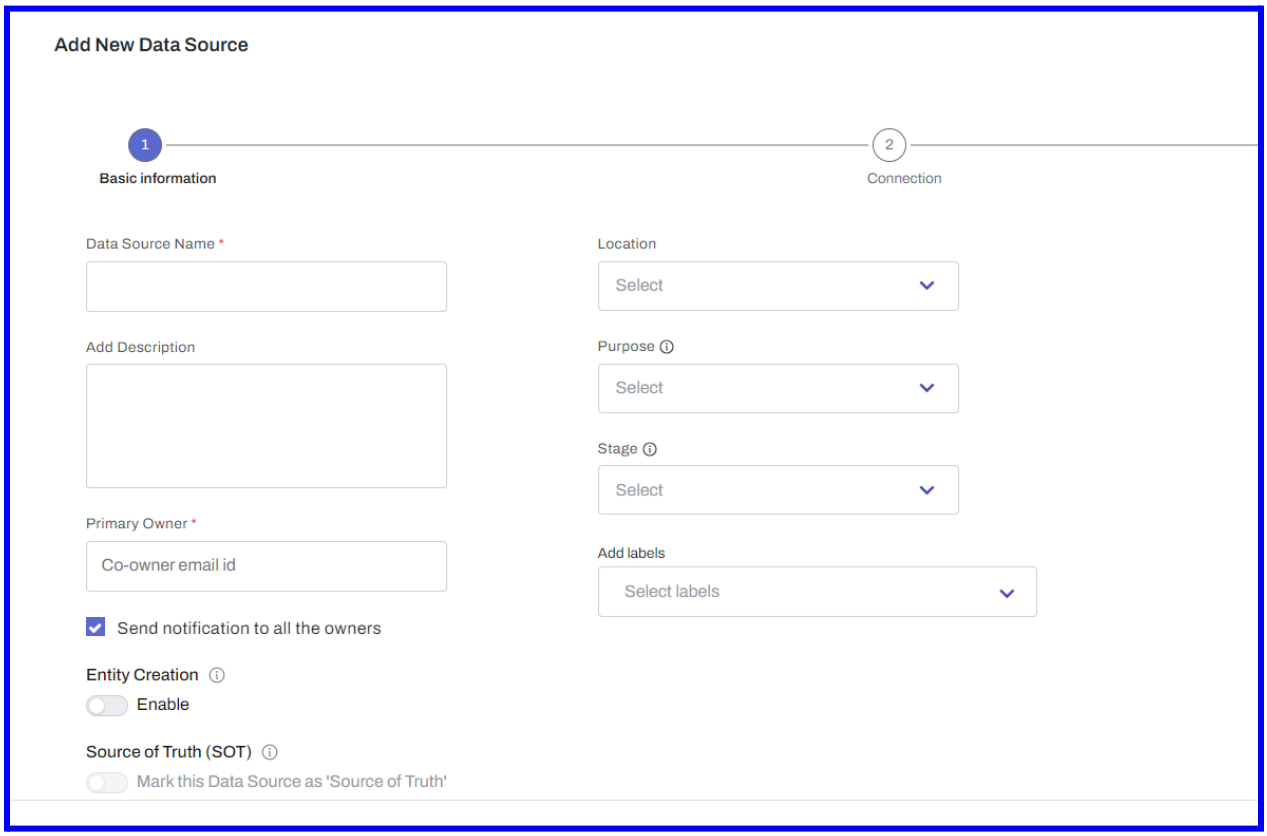
Start: From the Add New Data Source screen scroll to find data source (AWS S3), Select AWS S3.

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:

10. Make sure the name is not repeated, as it acts as metadata for the data source and must be unique. E.g., LB_AWSS3_Sandbox_Alpha
11. Use the description to explain the kind of information the data source contains. E.g., All HR related documents stored here
12. LightBeam uses the email ID stated as the Primary Owner to send alert notifications.
13. You can add another email for notifications using the co-owner button. Remember to check the 'send notification to all owners' box.
14. Entity Creation: Enable to create entities out of this data source.
15. Location tells where the data source is located.
16. Purpose tells for what purpose is this data source storing or processing data for.
17. Stage tells what stage the data belongs in could be source, processing, transactional, archival etc

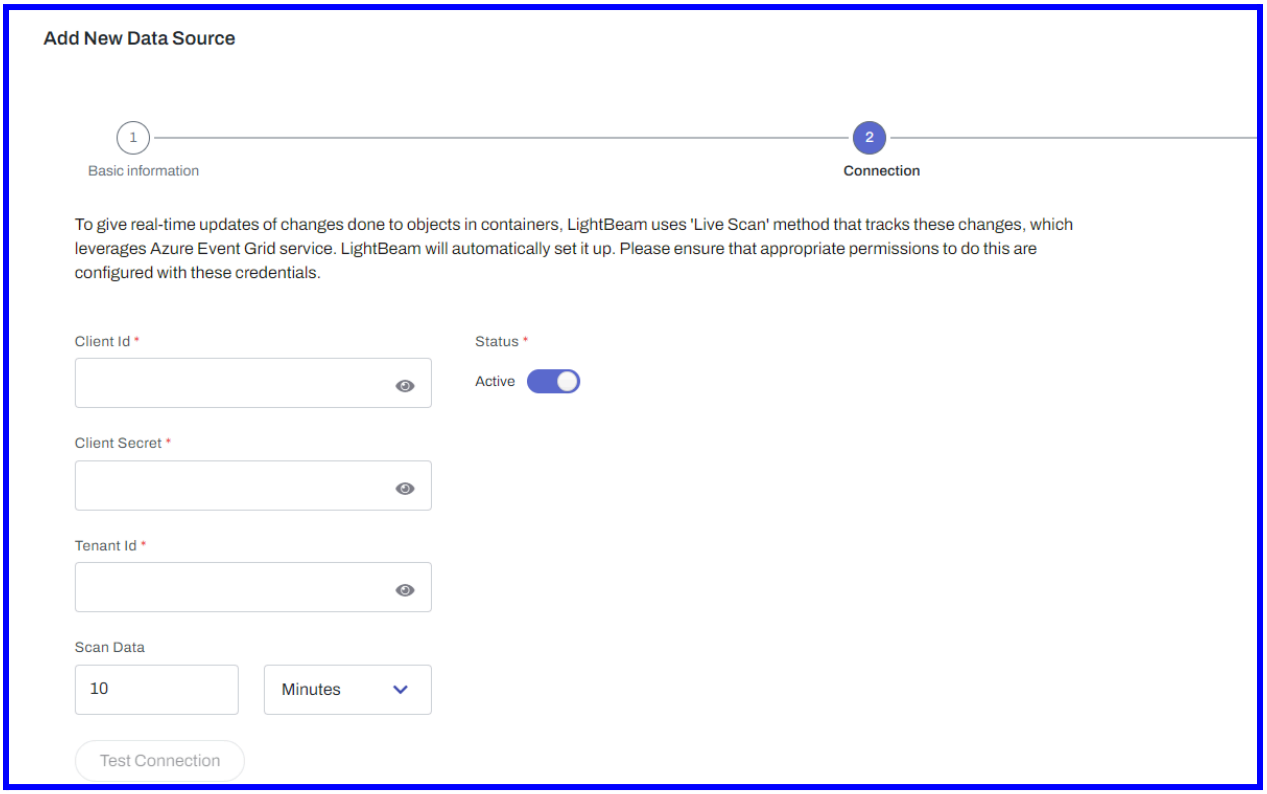
Basic Information Screen



Step 2: Connection Details

6. Client ID
7. Client Secret
8. Tenent ID
9. Set scan Time
10. Set status to active

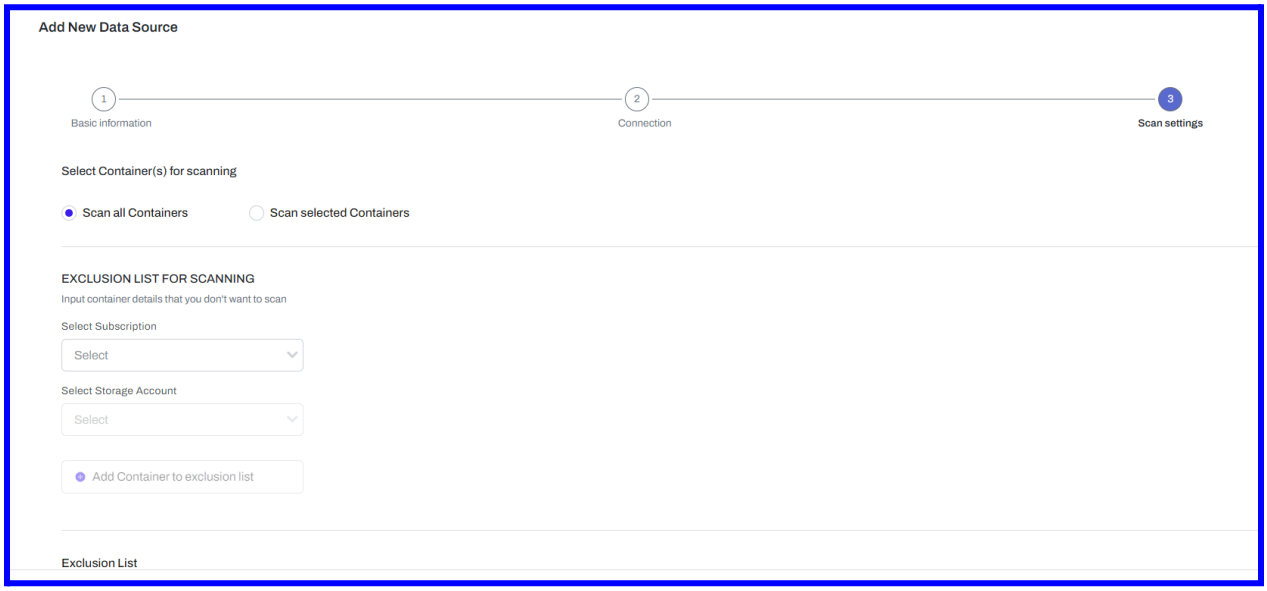
You can test the connection and see if the connection really went through Connection Screen



Step 3: Scan Settings

5. Scan all containers
6. Scan selected containers
7. Exclusion list
    a. Add/Exclude buckets for scanning

Scan Configuration ScreenFor



For more information on configurations in LightBeam see LightBeam technical documentation here; https://docs.lightbeam.ai/lxqobxw6ak7CTnsQjikH

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 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.

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.