



Personal Data Privacy and Protection platform integration with Google Cloud Spanner

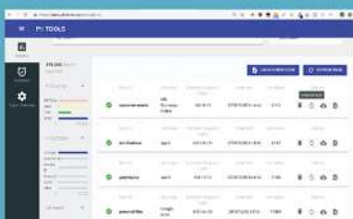


Cloud Spanner

Cloud Spanner returns the requested data for the scan

Connector app authenticates the user using OAuth and makes a call to Cloud Spanner to run SQL Query

PII data Scanning Application



Connector app retrieves the requested data for a certain timeline, processes the response data, sample it based on predefined sampling techniques, parse the response and sends it to the Customer Platform



Connector App

Connector triggers a call to the connector app to initiate a scan on the Spanner Data



Customer

Customer is a leading Personal Data Privacy and Protection provider.

It enables organizations to discover and map all types of data from all enterprise data sources; automatically classify, correlate, and catalogue identity & entity data into profiles; manage and protect enterprise data with advanced data intelligence; and automate data privacy and protection.

It identifies all PII across structured, unstructured, cloud & Big Data.

Customer requested to build a Connector app to integrate its platform with Google Cloud Spanner to scan the data present in the Spanner for finding the PII information.



Requirement



Technology Solution

- Google Cloud Spanner is a fully managed, mission-critical, relational database service on the Google Cloud Platform that offers transactional consistency at a global scale, schemas, SQL and automatic, synchronous replication for high availability.
- Sacumen developed the Connector app to integrate Cloud Spanner using Java SDK 8 version. The Connector app performs the following actions:
 - Set up the OAuth consent screen in Cloud Console and get the Client Id and Client Secret to authenticate the API.
 - Connects to Cloud Spanner using the ProjectId, Cloud Spanner instance and the OAuth flow.
 - Retrieves the data using SQL query statement execution on the database.
 - Performs a strong read, a stale read, a read using an index on the database.
 - Performs retrieving of data in parallel.
 - Gets the requested data within a certain timeline.
 - Samples the fetched data using predefined sampling techniques.
 - Formats the received data in the required format and pass it to the customer.

