Snowflake

Data Masking

snowflake

dataZense™ Whitepaper

Objective

This paper provides an insight into Best Practices for protecting and securing the Snowflake sensitive PII data by implementing the Data Profiling and Data Remediation solution, thereby assuring the companies to ensure compliance with the GDPR, SOX Compliance and similar data privacy laws worldwide and maximize the business value.

Ganapathy Arumugam

Senior Enterprise Architect
ChainSys Corporation

Senior Enterprise Architect in ChainSys with 20+ years of experience specialized in Data Security and Compliance. Heavily engaged with the Global Clients to secure their sensitive data. Published several ML Algorithms for sensitive data identification and remediation. My current focus areas are Data Security and Compliance, Data Migration, Data Integration, Big Data, Master Data Management, Business Intelligence and Artificial Intelligence.

PII Snowflake
Data Masking

BUSINESS
EXPECTATIONS

Personal information or Personally identifiable information (PII) is any data that could potentially identify a specific individual. No company can claim 100% compliance, but they have to make "Good Faith Effort" to cover most of the ground. The new rules grant people more rights regarding how companies handle their personally identifiable information (PII), and it imposes heavy fines for non-compliance and data breaches--up to 4 percent of a company's yearly revenue.

For security teams, this means making sure that PII is adequately protected and that the proper reporting processes are in place. Companies are at high risk of significant fines, data breach, brand reputation, and potential loss of customers if they do not mitigate the risk of adhering to the proper compliance levels in all regions.

Making 'Zense' of data is a challenge for most companies. dataZense is an end-to-end analytics platform that integrates all of your company's data, secures it, and brings it to life through easy-to-use Dashboards. The Data Catalog helps look for needed data in an organization. dataZense® provides an end-to-end capability for Data Profiling, Data Remediation (Masking, Scrambling, and Encryption), and Data Governance.

PII Data Categories

Personal data are grouped in to two different categories Critical and Confidential Data Elements or Categories.

Critical Data Elements/Categories

Critical data elements are those elements by which you can directly identify individual/person such as their social security number, National Identifier, Visa Number, Passport Number, Employee Number etc.

Confidential Data Elements/Categories

Confidential data elements are information which when combined other personal data elements can able to identify individual. E.g.: First name, Last Name, Date of Birth.

Snowflake PII Data Profiling and Data Masking

The PII Data Profiling and Data Masking within the dataZense application is broken out into the following three high level process

- PII Data Identification

- PII Business Review and Approval

- PII Data Remediation and Governance

1   **2**   **3**

PII Data Identification → PII Data Business Review and Approval → PII Data Remediation and Governance
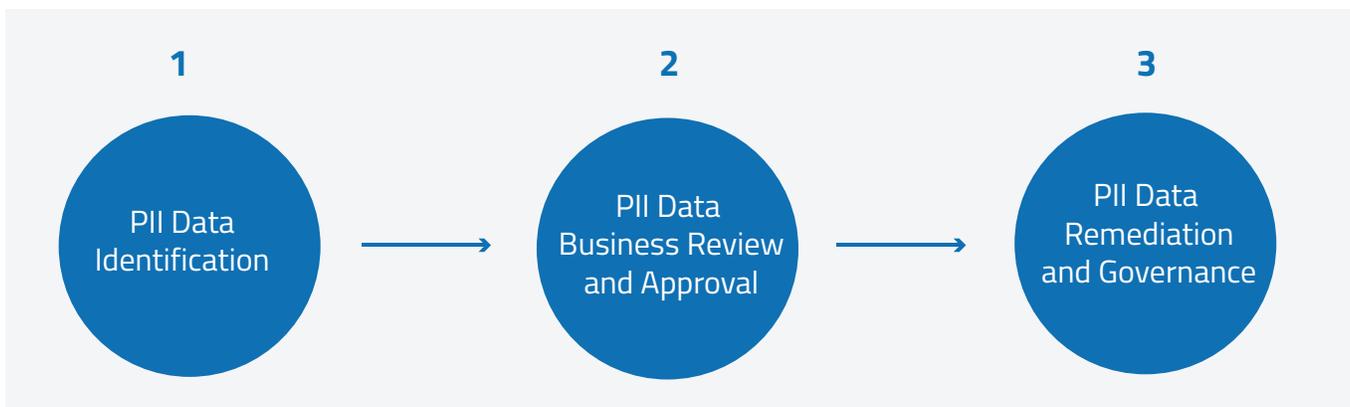
PII Identification

Identify Systems considered for Sensitive Data Identification (PII) and run the Data profiling process to identify the PII Attributes. Provides the complete visibility of the identified PII data. i.e., Server, Instance, Database, Schema, Table, Column level details.

PII Business Review and Approval

The identified PII attributes will go through False-Positive / False-Negative analysis and approved for remediation

PII Remediation and Governance

The business approved PII attributes will be considered for remediation.

Remediation Options:

- Scramble the Data

- Mask the Data

- Remove Data (With capture of logs) – remove from all production and non-production environments

- Keep the Data (Encrypt in all Non-Production environments)

## High Level Architecture
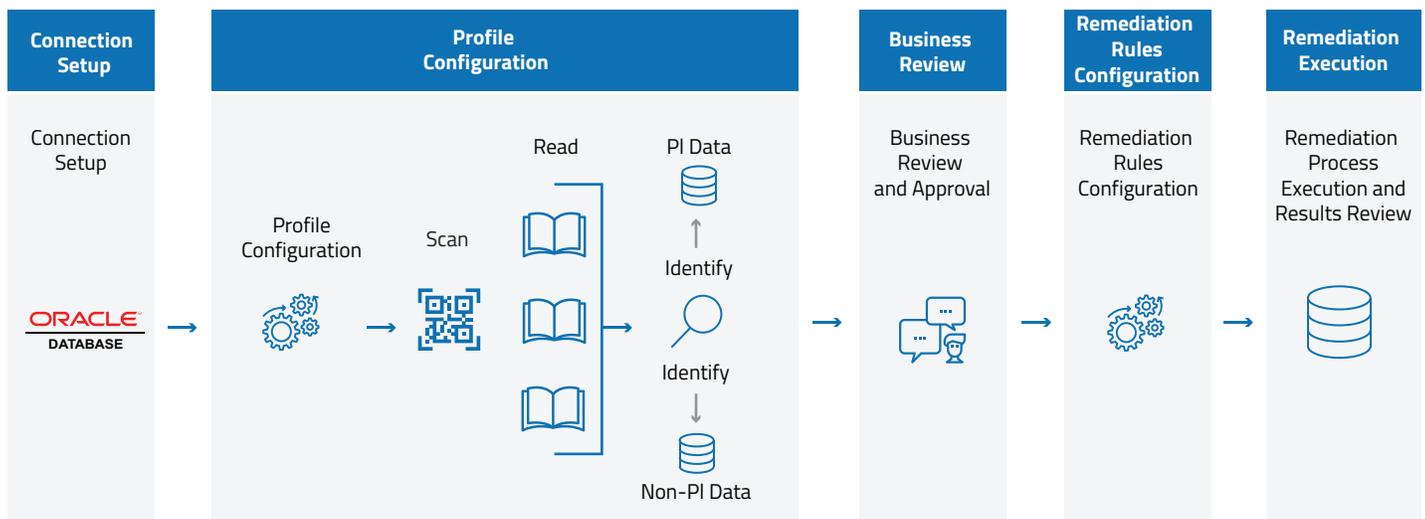
Data Masking Process Steps

Steps to configure:

- Create a Connection to the Salesforce instance.

- Create a Data Security profile for scanning the Salesforce Instance.

- Business Review of scanned results and complete the Business validation of Approve/ Reject the PII Data.

- Create a Remediation Profile to configure the Remediation Rules for masking the data.

- Execute the configured Remediation Profile to mask the data and Review the masked results from the Salesforce Application and database.

Connection

Verifying the Source Connection:

Ensure that the respective Snowflake source connection is available. If not, create a new connection.
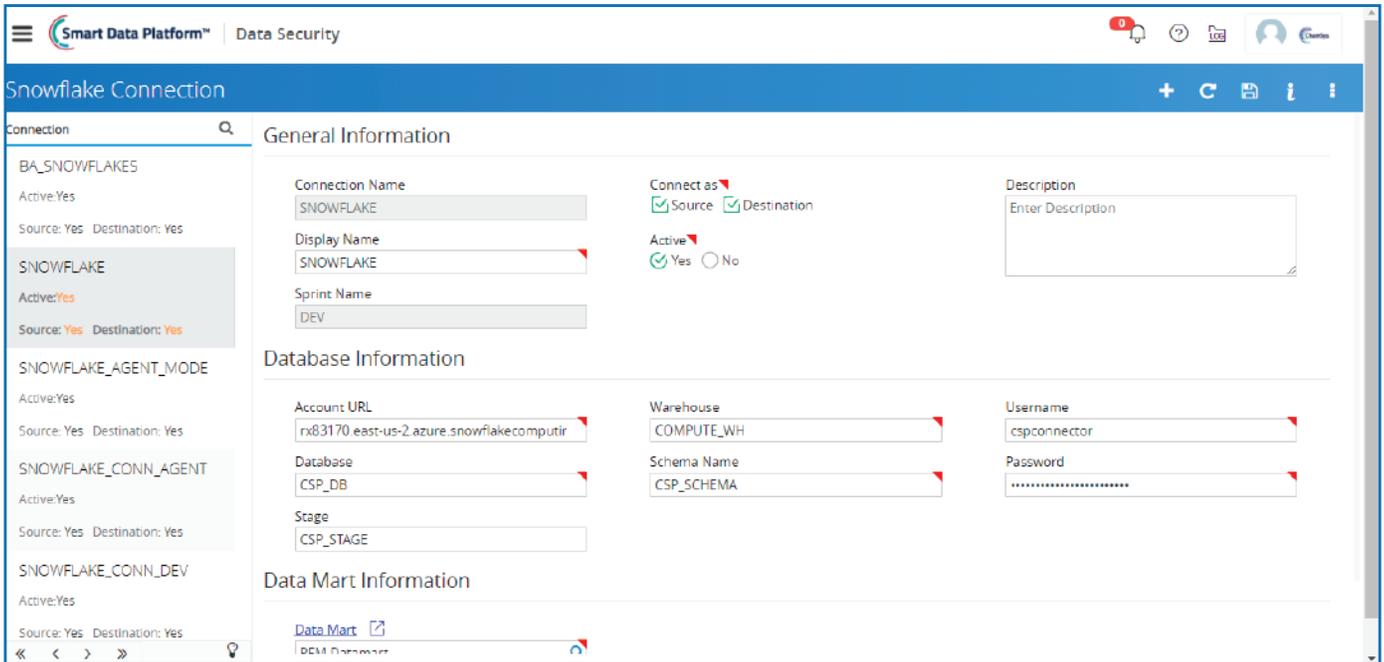
Profile Creation and Execution

Create a profile for scanning the Snowflake Database and execute
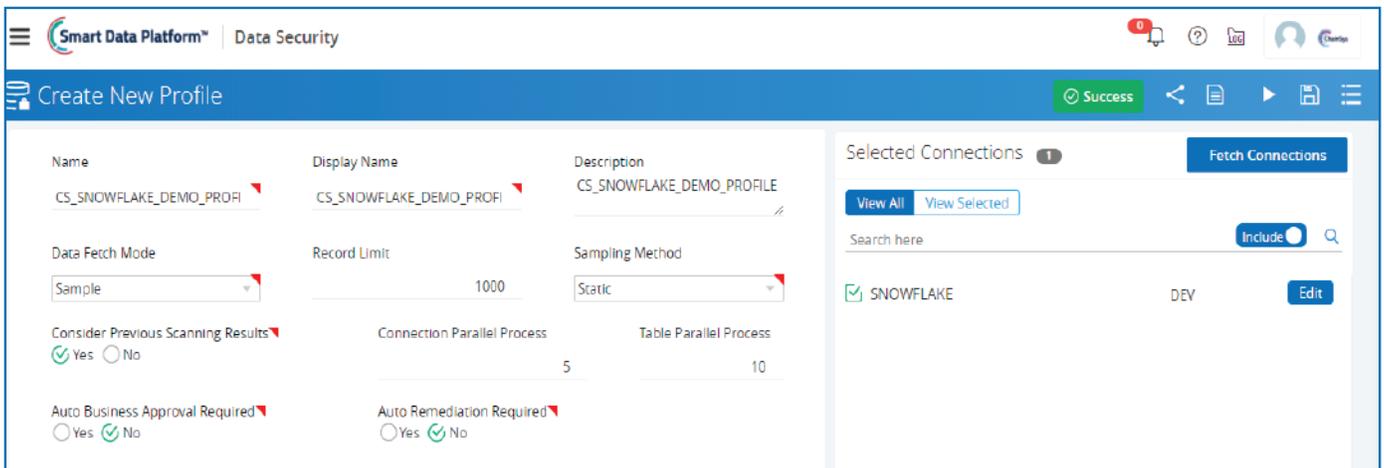
Business Review

Review the scanned results and complete the Business validation by Approving or Rejecting the PII Data.

Remediation Profile

Create a Remediation Profile to configure the Remediation Rules for masking the data.

Remediation Execution

Execute the configured Remediation Profile to mask the data.

Remediation Execution

Review Masked Results

Review the masked results from the Snowflake database.

Dashboards

PII Home Dashboard

Provides the high-level summary of Profiled and Remediated Databases
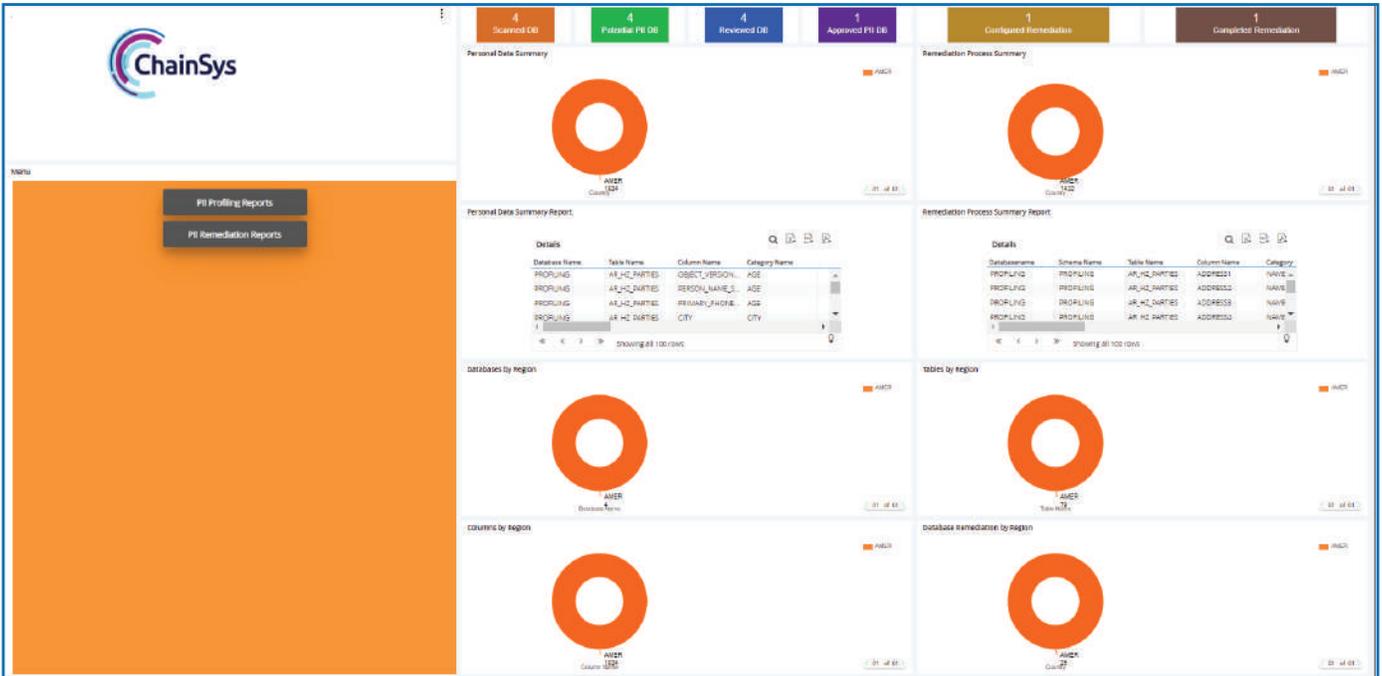
Personal Data Count by PII Category

Provides the PII data count by each PII Category

Supported Endpoints (Partial)

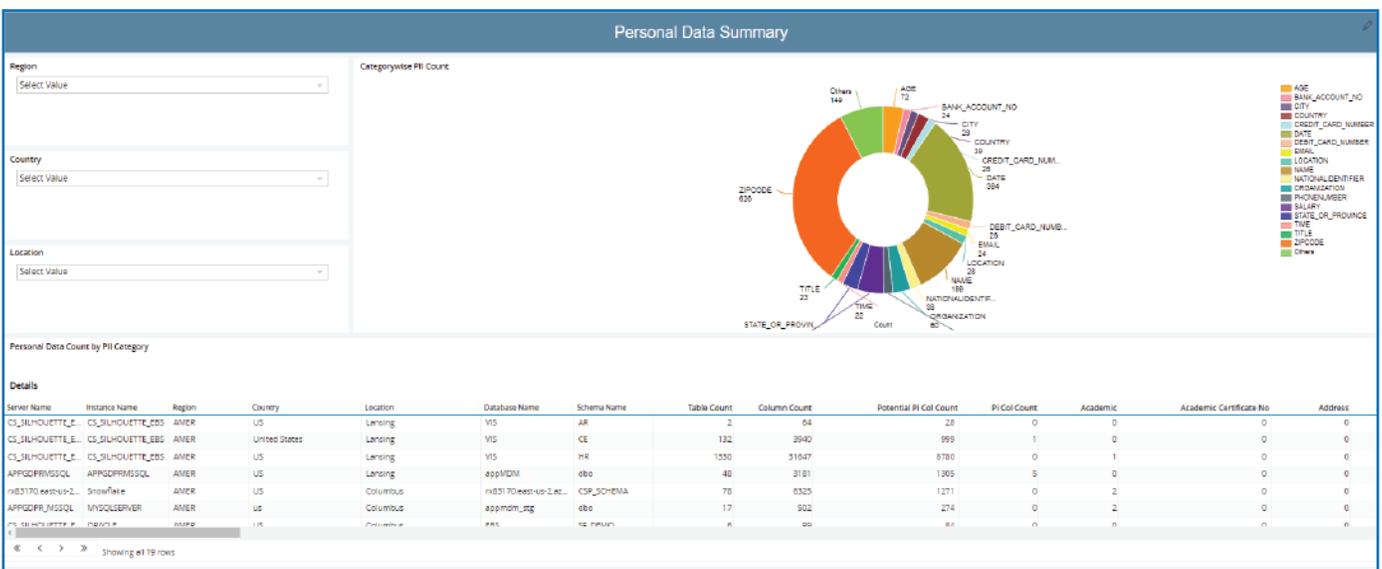|	
Oracle Sales Cloud, Oracle Marketing Cloud, Oracle Engagement Cloud, Oracle CRM On Demand, SAP C/4HANA, SAP S/4HANA, SAP BW, SAP Concur, SAP SuccessFactors, Salesforce, Microsoft Dynamics 365, Workday, Infor Cloud, Procore, Planview Enterprise One	**Cloud Applications**
Oracle E-Business Suite, Oracle ERP Cloud, Oracle JD Edwards, Oracle PeopleSoft, SAP S/4HANA, SAP ECC, IBM Maximo, Workday, Microsoft Dynamics, Microsoft Dynamics GP, Microsoft Dynamics Nav, Microsoft Dynamics Ax, Smart ERP, Infor, BaaN, Mapics, BPICS	**Enterprise Applications**
Windchill PTC, Orale Agile PLM, Oracle PLM Cloud, Teamcenter, SAP PLM, SAP Hybris, SAP C/4HANA, Enovia, Proficy, Honeywell OptiVision, Salesforce Sales, Salesforce Marketing, Salesforce CPQ, Salesforce Service, Oracle Engagement Cloud, Oracle Sales Cloud, Oracle CPQ Cloud, Oracle Service Cloud, Oracle Marketing Cloud, Microsoft Dynamics CRM	**PLM, MES & CRM**
Oracle HCM Cloud, SAP SuccessFactors, Workday, ICON, SAP APO and IBP, Oracle Taleo, Oracle Demantra, Oracle ASCP, Steelwedge	**HCM & Supply Chain Planning**
Oracle Primavera, Oracle Unifier, SAP PM, Procore, Ecosys, Oracle EAM Cloud, Oracle Maintenance Cloud, JD Edwards EAM, IBM Maximo	**Project Management & EAM**
OneDrive, Box, SharePoint, File Transfer Protocol (FTP), Oracle Webcenter, Amazon S3	**Enterprise Storage Systems**
HIVE, Apache Impala, Apache Hbase, Snowflake, mongoDB, Elasticsearch, SAP HANA, Hadoop, Teradata, Oracle Database, Redshift, BigQuery	**Big Data**
mangoDB, Solr, CouchDB, Elasticsearch	**No SQL Databases**
PostgreSQL, Oracle Database, SAP HANA, SYBASE, DB2, SQL Server, MySQL, memsql	**Databases**
IBM MQ, Active MQ	**Message Broker**
Java, .Net, Oracle PaaS, Force.com, IBM, ChainSys Platform	**Development Platform**

One Platform for your

End to End Data Management needs

Smart Data Platform™

dataZap™
Data Migration
Data Reconciliation
Data Integaration

dataZen™
Data Quality Management
Data Governance
Analytical MDM

dataZense™
Data Analytics
Data Catalog
Data Security & Compliance

www.chainsys.com