Secure Data Masking for Sensitive Data

Gillian Newberry, Territory Manager, Libelle
Puneet Khatri, Sr. Technical Consultant, Libelle

SAPinsider
Las Vegas

2023

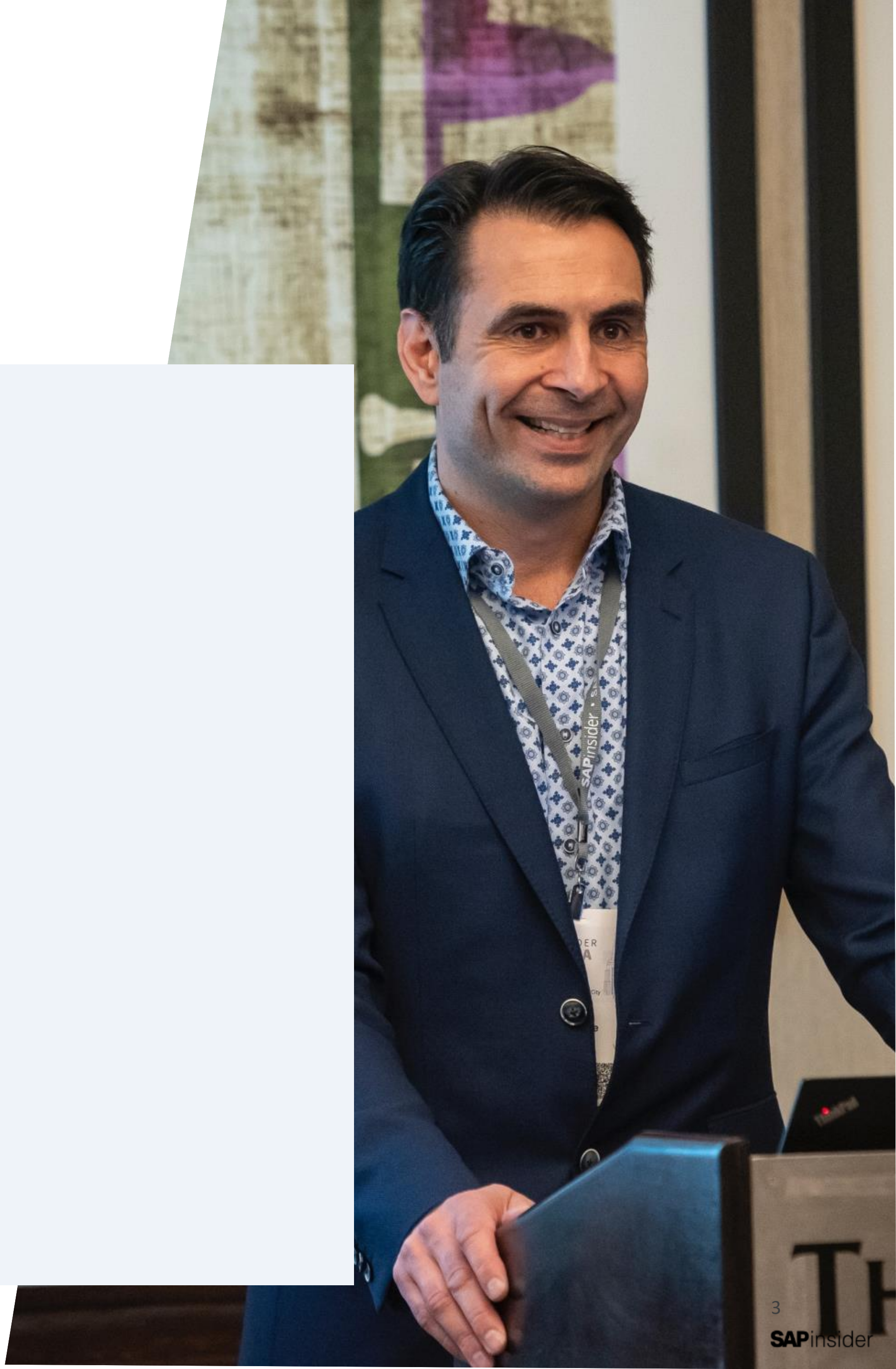
SAPinsider

In This Session

- Different types of threats to data security
- Methods for data protection
- Secure data masking – definitions and concept
- Use cases for data masking
- Software demonstration

What We’ll Cover

- Data Security Threats
- Data Protection Methods
- Data Masking
- Use Cases
- Live Demonstration
- Wrap-Up



Data Security Threats

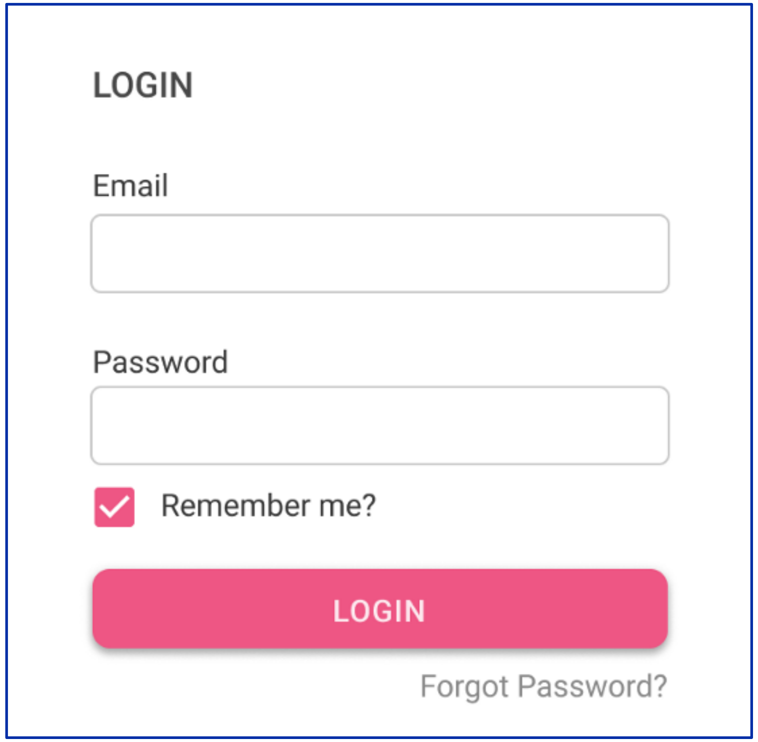
- Hacks, breaches, and leaks
- Internal vulnerabilities
- Third-party access

Hacks, Leaks, & Breaches

- **Exposure of confidential, protected, or sensitive information to an unauthorized person/organization**
- **Occurs due to weaknesses in technology or human behavior**
- **Can be intentional or accidental, internal or external**
- **Malicious attacks**
  - Phishing, brute force attacks, or malware
  - Target vulnerabilities such as weak passwords, third-party access, and carelessness
  - Can be targeted to specific organizations, or random
  - Information is either sold or held ransom

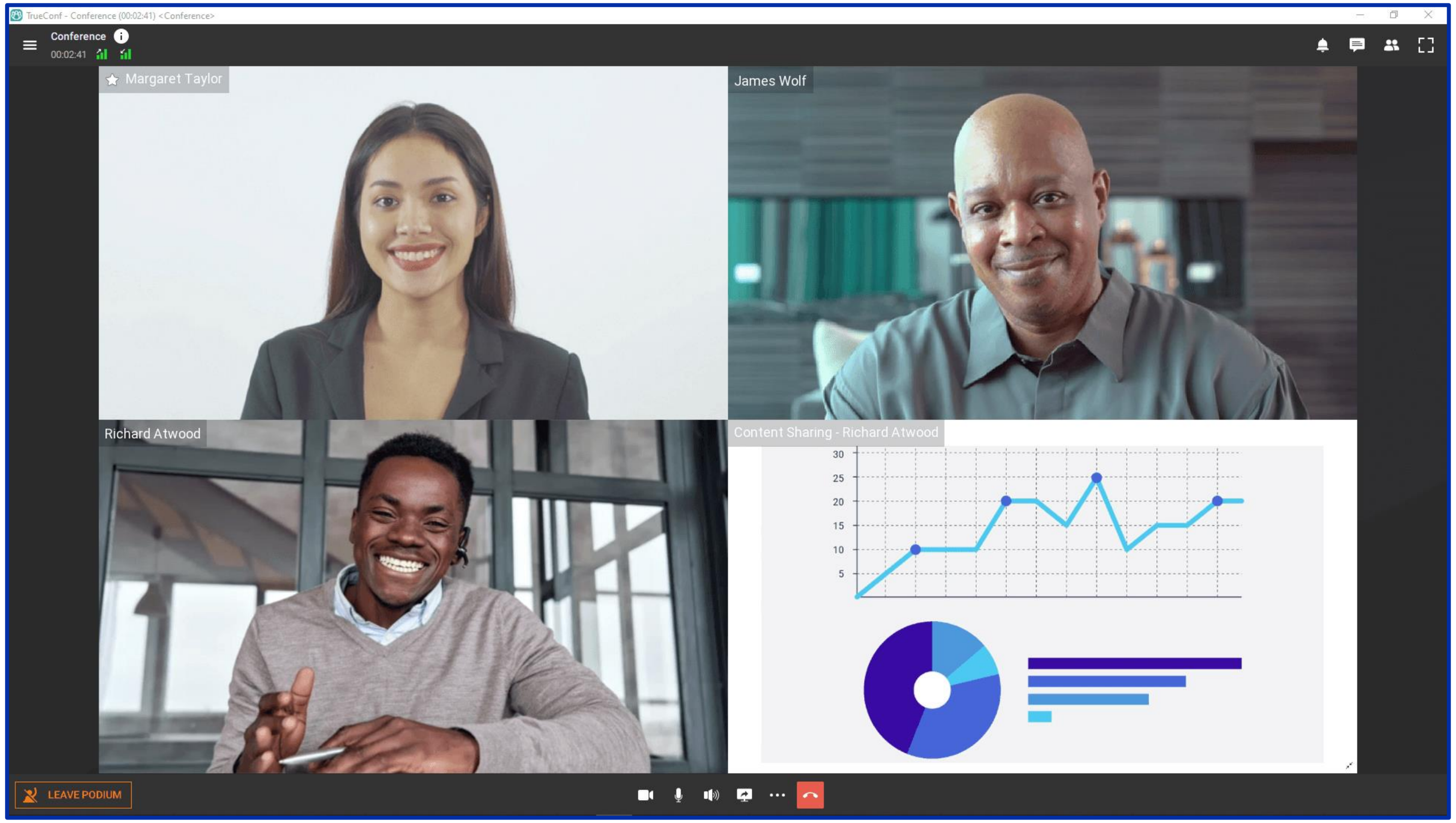
Internal Vulnerabilities

- **Employee access**
  - Standard formats for usernames
  - Low security protocols for passwords
  - Lack of continuing engagement on data security
  - Any viewing of data not meant to be seen by an individual is an exposure, regardless of the innocence of the event
- **Data store structure**
  - High security policies are often in place only in productive systems
  - Multiple support systems, more users, same data
  - Liability for consumer and employee personal data



External Vulnerabilities

- **Third-Party Access**
  - Most companies have some level of third-party access to some or all of their support systems
  - Consultants, auditors, vendors
  - Creates additional access point to data
  - Creates liability
  - Contraindicated for protection of PHI/PII

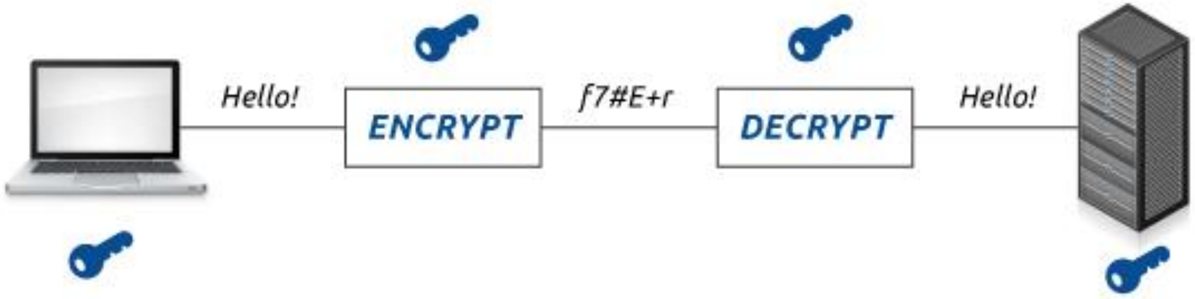


Data Protection Methods

- Encryption
- Hiding
- Anonymization/Pseudonymization
- Masking

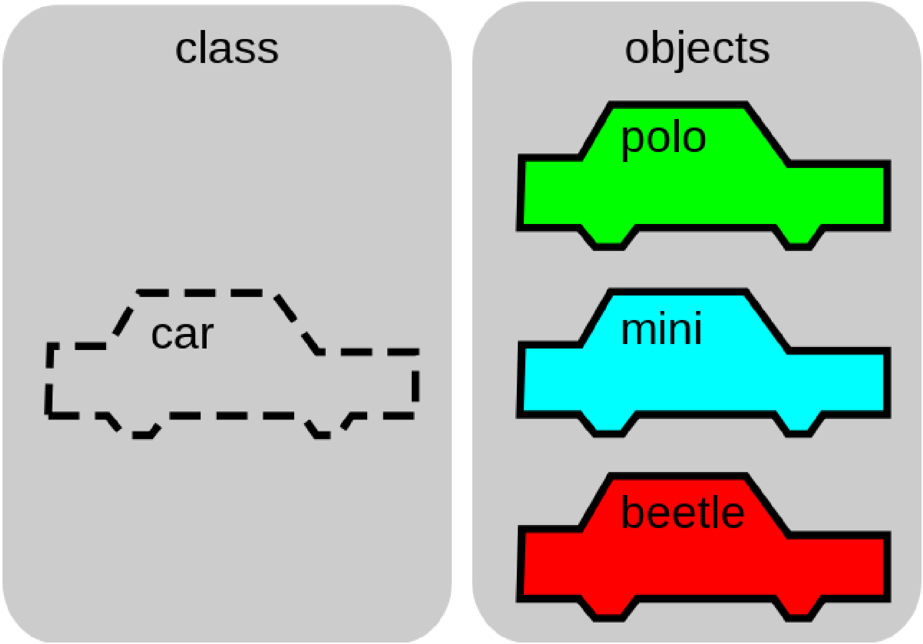
Encryption

- Process of converting data into a cypher text that can be accessed via a key
- Security depends upon the strength of encryption and access to the key
- Lacks functionality for testing and analytics — data is unreadable



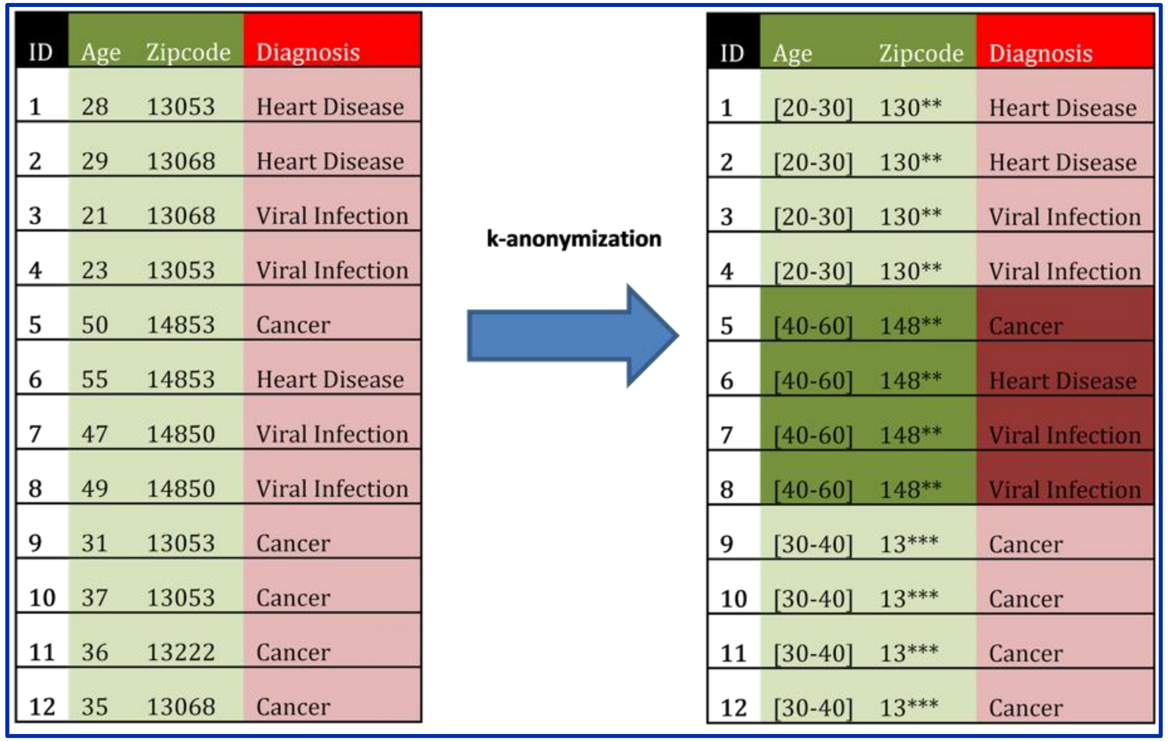
Hiding

- Provides different views for different users based on preset access
- Uses shadow tables
- Limited information about the full data is available
- Data remains in place and access to the correct username means access to the data



Anonymization/Pseudonymization

- Replacement of all or part of data with blanks, asterisks, or other characters
- Random, unreadable, and unusable for testing/analytics
- Irreversible — more secure
- Original data is not retained in full or at all
- Functional for analytics in certain aspects, but not functional for full test/development cycles



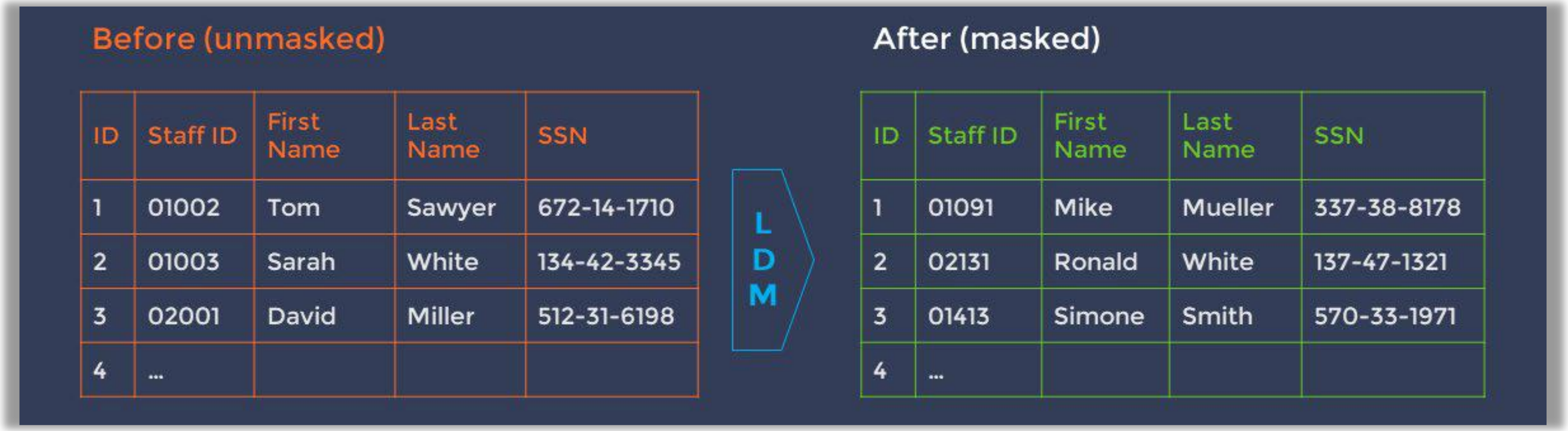
ID	Age	Zipcode	Diagnosis
1	28	13053	Heart Disease
2	29	13068	Heart Disease
3	21	13068	Viral Infection
4	23	13053	Viral Infection
5	50	14853	Cancer
6	55	14853	Heart Disease
7	47	14850	Viral Infection
8	49	14850	Viral Infection
9	31	13053	Cancer
10	37	13053	Cancer
11	36	13222	Cancer
12	35	13068	Cancer

k-anonymization

ID	Age	Zipcode	Diagnosis
1	[20-30]	130**	Heart Disease
2	[20-30]	130**	Heart Disease
3	[20-30]	130**	Viral Infection
4	[20-30]	130**	Viral Infection
5	[40-60]	148**	Cancer
6	[40-60]	148**	Heart Disease
7	[40-60]	148**	Viral Infection
8	[40-60]	148**	Viral Infection
9	[30-40]	13***	Cancer
10	[30-40]	13***	Cancer
11	[30-40]	13***	Cancer
12	[30-40]	13***	Cancer

Masking

- Replacement of data with readable, testable data from data repository
- Selective — certain data can be left in original form as selected
- Irreversible — more secure
- Original data is not retained



Before (unmasked)

ID	Staff ID	First Name	Last Name	SSN
1	01002	Tom	Sawyer	672-14-1710
2	01003	Sarah	White	134-42-3345
3	02001	David	Miller	512-31-6198
4	...			

After (masked)

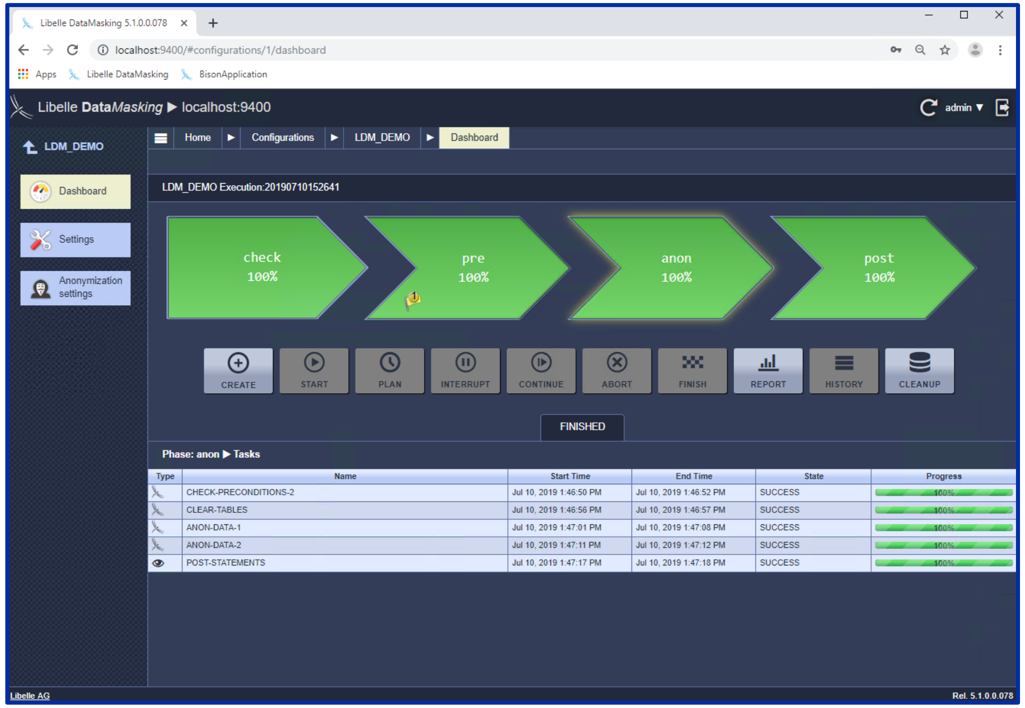
ID	Staff ID	First Name	Last Name	SSN
1	01091	Mike	Mueller	337-38-8178
2	02131	Ronald	White	137-47-1321
3	01413	Simone	Smith	570-33-1971
4	...			

Libelle Data*Masking*

- Software overview
- Key technologies
- Algorithms
- Masking Profiles
- Masking Keys

Software Overview

- Libelle **Data***Masking* (LDM) is a standard software solution from Libelle IT Group for masking sensitive data in non-productive data stores
- LDM installs in the customer's data center and is under the sole control of the customer
- Masking is configured once for each data store in scope for masking, and then masking runs can be executed on a schedule or ad-hoc



LDM_DEMO Execution:20190710152641

FINISHED

Phase: anon ▶ Tasks

Type	Name	Start Time	End Time	State	Progress
	CHECK-PRECONDITIONS-2	Jul 10, 2019 1:46:50 PM	Jul 10, 2019 1:46:52 PM	SUCCESS	100%
	CLEAR-TABLES	Jul 10, 2019 1:46:56 PM	Jul 10, 2019 1:46:57 PM	SUCCESS	100%
	ANON-DATA-1	Jul 10, 2019 1:47:01 PM	Jul 10, 2019 1:47:08 PM	SUCCESS	100%
	ANON-DATA-2	Jul 10, 2019 1:47:11 PM	Jul 10, 2019 1:47:12 PM	SUCCESS	100%
	POST-STATEMENTS	Jul 10, 2019 1:47:17 PM	Jul 10, 2019 1:47:18 PM	SUCCESS	100%

Key Technologies

- **LDM Master Server**
  - Holds configuration data, templates, connection data, etc.
- **LDM UI**
  - Console to configure and execute masking, authentication via username/password
- **LDM Masking Algorithms**
  - Pre-configured algorithms to anonymize data in multiple formats – e.g., alphanumeric, alphabetic, UTF8, etc.
- **LDM Masking Profiles**
  - Preset lists of grouped data types identified for masking – e.g., bank data, address data, etc.
- **LDM Reference Database**
  - Standard repository of names, addresses, etc., for replacing original data provided by Libelle
  - Can also use existing database inside customer data store

Algorithms

- 40+ standard algorithms included out-of-the-box
- Adjustable and customizable
- White space, nil, empty string ignored by default
- Conditional masking available
- Can be set to certain ranges for output results
- Geographical relevancy for names
- Geographical consistency for addresses

Profile	Attributes	Example (Input/Output)	
aAlphabetic	Anonymize only Latin characters in ASCII between. Numbers will be ignored.	S-L-1234 ABB 1234	M-P-1234 JVV 1234
aAlphanumeric	Combines aNumber and aAlphabetic, so that both numbers and characters are masked.	S-L-1234 Alp12 1234	M-P-0356 Khj79 0356
aAlphanumeric_UTF8	aAlphanumeric_UTF8	广文字第03086 073 ἈΔ'ΘΕΣΣΑΛΟΝΙ 装文字第081928 عبد الملك	巾女八元32124 004 φ̓Α'ΑνάάΘΰΖζη 弓女八元255451 غيۏ بُؤه
aSerial	Anonymize - ignore leading zeros and anonymize like aNumber the rest of value.	00123 0012S3 1234 0	00301 0030S1 0356 0
...	,,,		

Masking Profiles

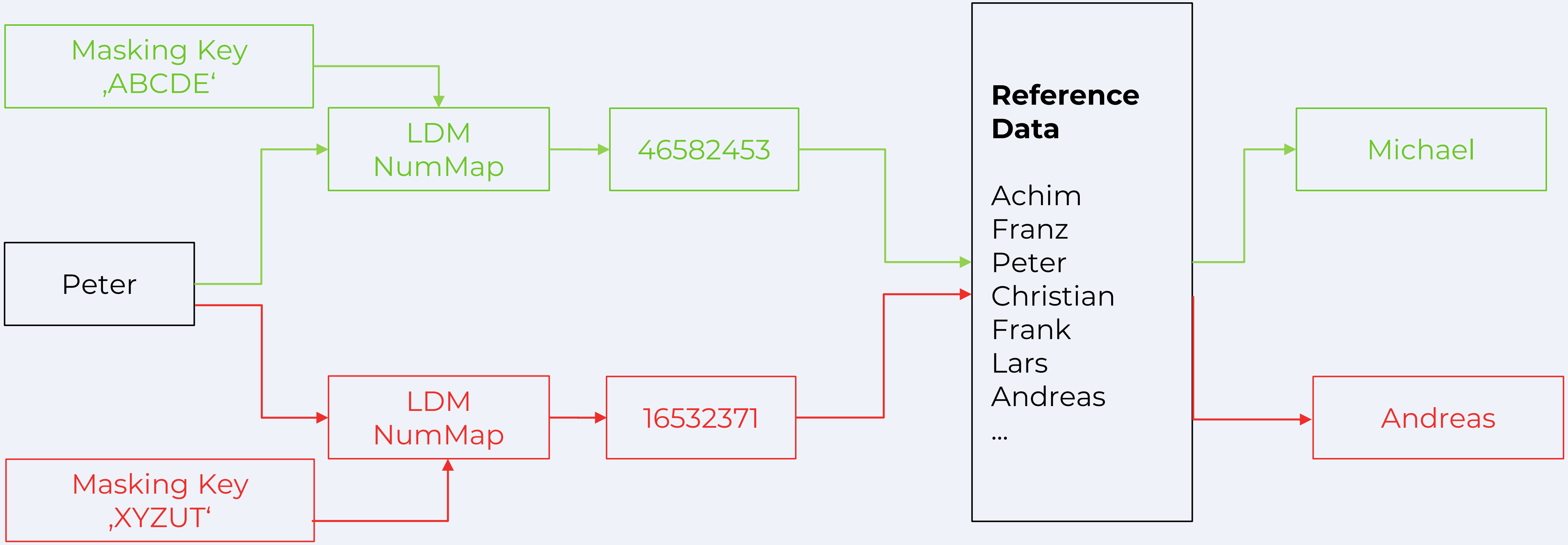
- Generic objects that include attributes with characteristics of certain data
- Group together connected fields, such as address data, bank data, etc.
- Allows for quick and thorough selection of data to be masked

Profile	Attributes
Bank Data	BIC, Bank Codes, IBAN Numbers, Bank Account Numbers, Country Code ISO, SWIFT Codes, …
Location Data	UTM Coordinates, WGS Coordinates, Country, Place of Birth, Post Office Box, Postal Code, Phone, Street Address, City, …
Date & Time Data	Year of Birth, Month of Birth, Day of Birth, SAP Date Format YYYYMMDD, …
…	,,,

Masking Keys

- NOT analogous to an encryption key
- Small file created at inception of masking run — unique to individual masking run
- Masking keys create SHA2-based lookup tables
- For interdependent systems, users can input the same masking key for subsequent runs on different systems
- Ensures data masks in the same way for each system with the same masking key

Masking Key Example



Input Value

Calculate Hash Depending on LDM NumMap

Find Position in reference DB

Output Value

Use Cases

- **Analytics**
  - Allows for necessary data to be retained, while PII or PHI is removed
  - Can keep certain ranges on data for accuracy, while protecting sensitive information
- **Testing**
  - Keeps data in similar structure to original data for accuracy during testing cycles
- **Development**
  - Allows developers access to data which is structured like productive data, meaning development work is performed on data most closely resembling productive data

Demonstration

Wrap Up

- Libelle **Data***Masking* provides secure, compliant masked data for non-productive systems, compatible with analytics, test, and development cycles
- Masking is one of the preferred ways to secure data, as it provides for the least disturbance to business processes, while providing a higher level of data protection
- All businesses and organizations are at risk of data breach — it is more “when,” not “if”

Where to Find More Information

Product Landing Page - https://www.libelle.com/products/datamasking/

Detailed Whitepaper - https://www.libelle.com/whitepapers/datamasking/

Reference Story - https://www.libelle.com/references/roland-rechtsschutz-anonymization-testdata/

Webinar – Integrated Data Masking with System Refresh - https://www.youtube.com/watch?v=lPTg7ejnspA

Demo System - https://demo.libelle.com/?product=datamasking

Key Points to Take Home

- Data security must be a high priority for all organizations
- Multiple methods of data protection exist, with various pros and cons
- Secure data masking is the most effective way to ensure protection while maintaining functionality
- Libelle **Data***Masking* is one answer to the question of data security
- While no organization is ever fully threat-proof, organizations who secure their data face fewer risks

Thank You! Any Questions?

Gillian Newberry

https://www.linkedin.com/in/gillian-newberry-39019461/

Puneet Khatri

https://www.linkedin.com/in/puneet-khatri-42969114/

Please remember to complete your session evaluation.

SAPinsider

SAPinsider.org

PO Box 982Hampstead, NH 03841

SAPinsider comprises the largest and fastest growing SAP membership group worldwide, with more than 600,000 members across 205 countries.