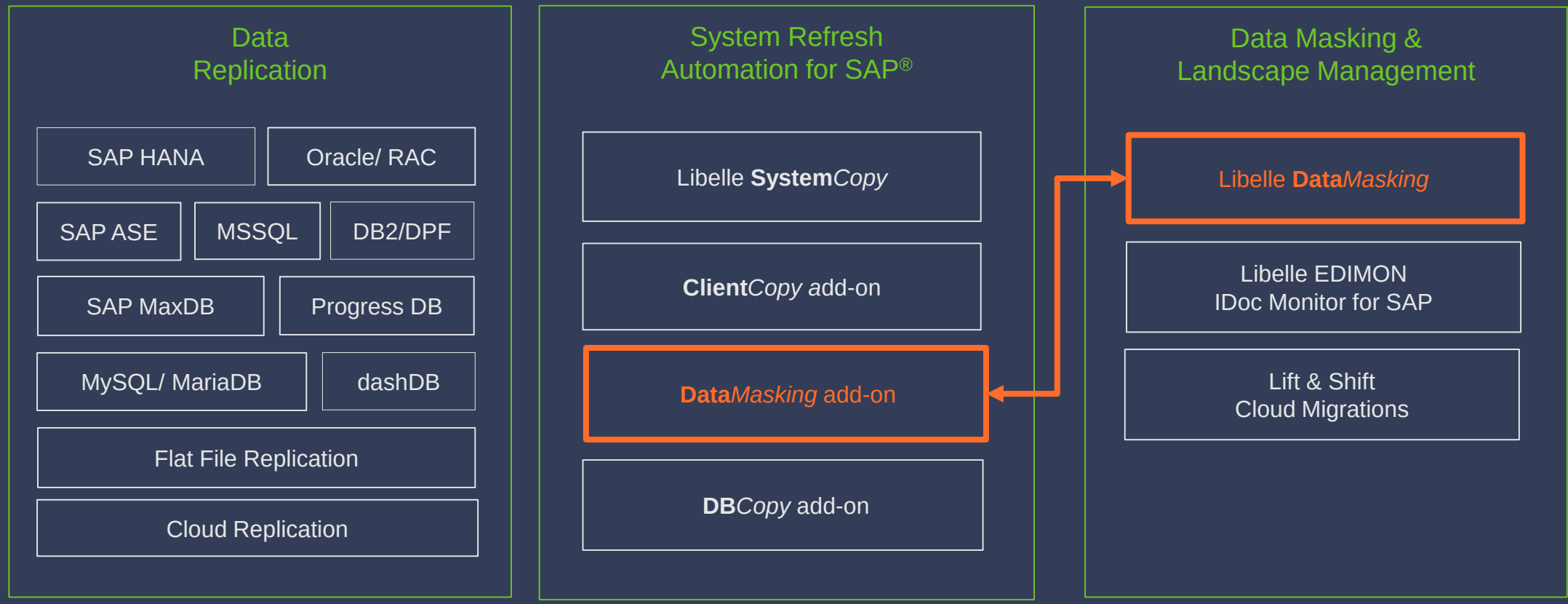
Enterprise Data Anonymization with Libelle **Data*Masking***

June 2020

Frank Küppers
Account Manager

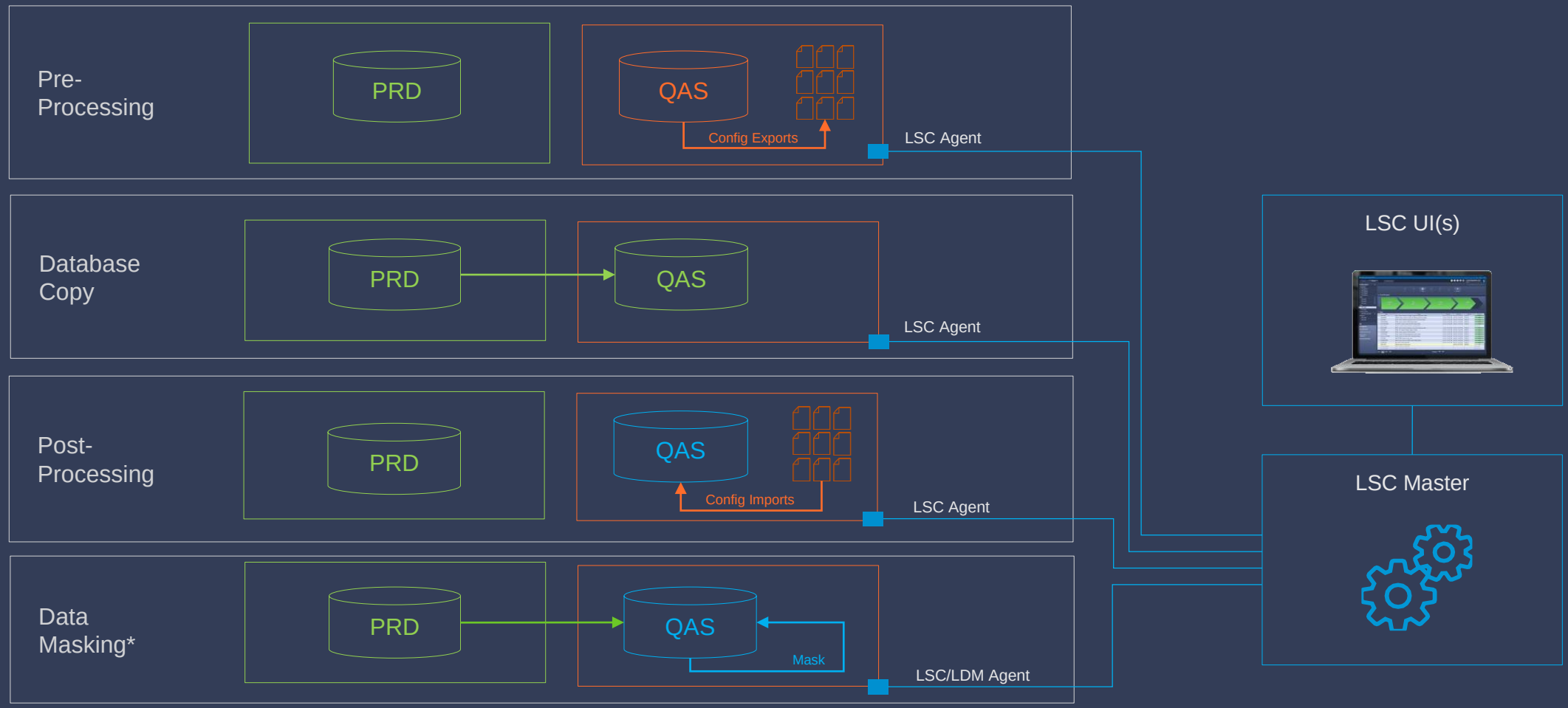
Libelle AG

Libelle Solution Portfolio



Libelle software standardizes, automates, and optimizes IT operation.

System Refresh Workflow



Config Data is exported from current QAS, followed by DB copy,
followed by config data imports and (optional) masking.

*Optional. Requires separate Libelle **Data***Masking*** tool which is integrated in LSC.

Data Masking Definition

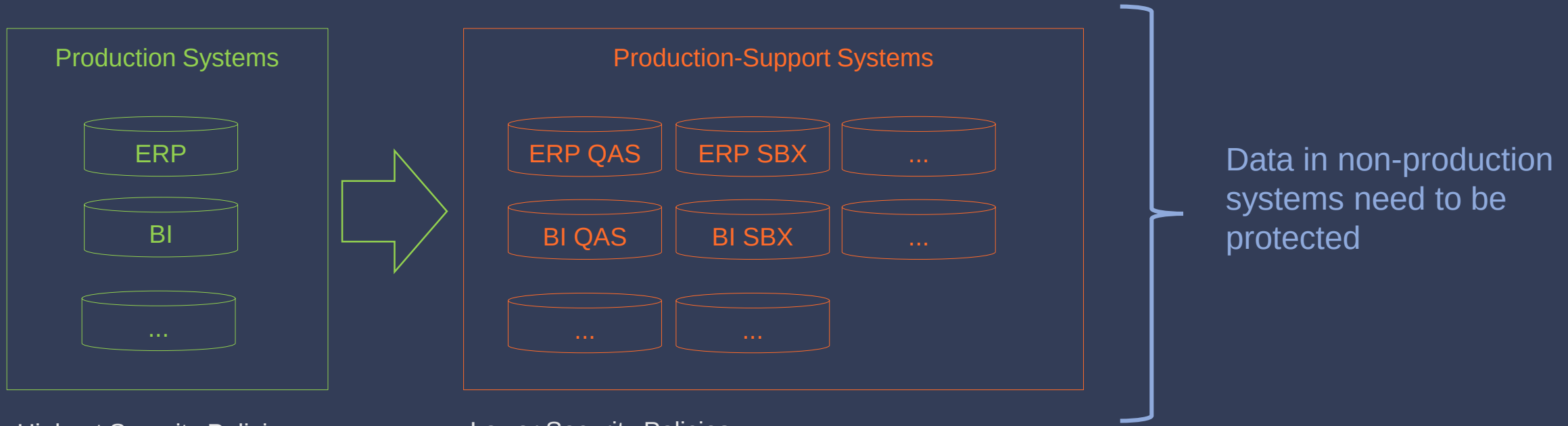
Term	Methodology	Retains Data?	Reversible?	Human Readable?
Data Masking	Replace original data with random data while maintaining original structure and readability.	No	No	Yes
Data Hiding	Provide different views on content for different users dynamically via shadow tables	Yes	n/a	Yes
Data Encryption	Convert data into cipher text typically with an encryption key that can be accessed with a key	Yes	Yes	No
Anonymization, Tokenization, or Pseudonymization	Replace original data by blanks, Asterix, or de-identified identifier (Subset of Data Masking)	No	No	No

Why Data Masking?

- Data masking or data obfuscation removes original data and replaces it with non-sensitive, random data.
- Goal is to protect data that is classified as personal identifiable, or by any means deemed sensitive.
- Data must remain usable for valid test cycles, structurally sound, and valid as it is used for application testing or analytics.
- Masking protects sensitive data during application development, building program extensions, conducting test cycles, and data analytics.

Data masking physically replaces sensitive data following pre-defined patterns.
It ensures meaningful and structurally sound data for test systems or analytics.

Masking Production Support Systems

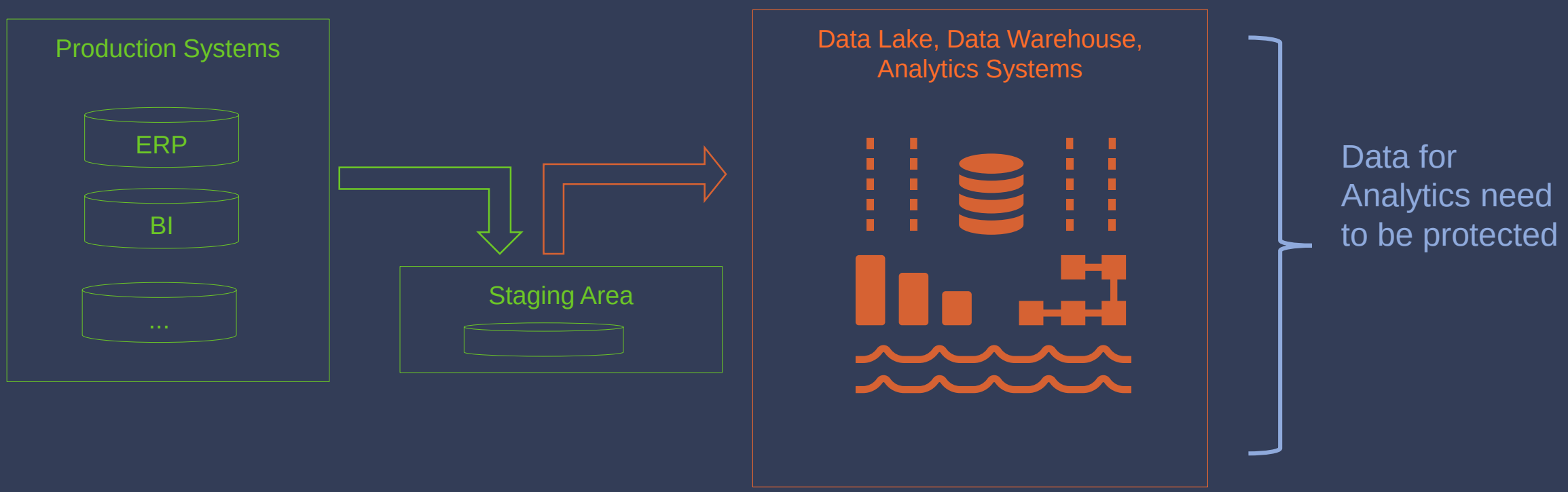


Highest Security Policies
Limited User Access
Intrusion Detection Systems
…

Lower Security Policies
Broad User Access (Developer, Offshore, etc.)
Limited Intrusion Detection Systems
…

Masking for production support systems protects sensitive and PII data
Primary goal is to keep structure of data intact for valid test cases

Masking Data for Data Analytics



Masking for Analytics protects sensitive and PII data
Primary goal is to keep essence of data intact for valid analytics results

Masking Example

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	...			

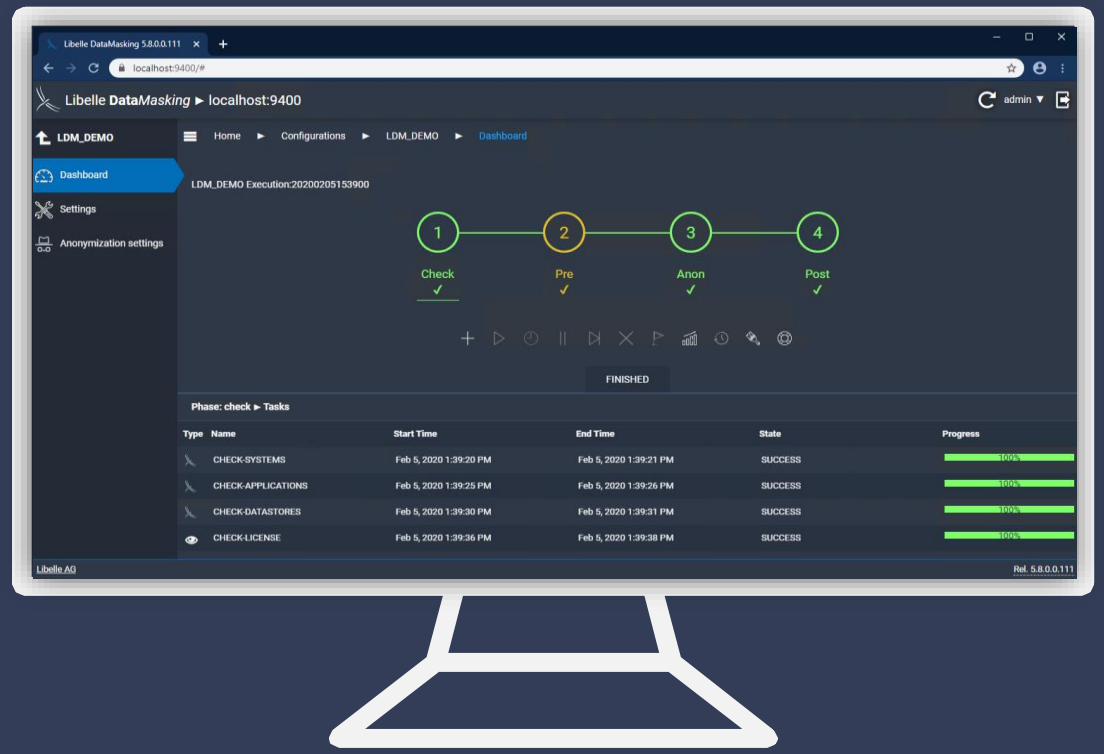
LDM

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	...			

LDM replaces data with human-readable random data while following pre-defined patterns such as masking different geographical regions, pre-defined structures or ranges, defined character sets, etc.

What is Libelle Data*Masking* (LDM)?



- LDM is a standard software solution from Libelle.
- LDM installs on a masking server in a customer data center;
- Once installed, connections to data stores are configured.
- Masking typically runs directly on a DB level for maximum performance.

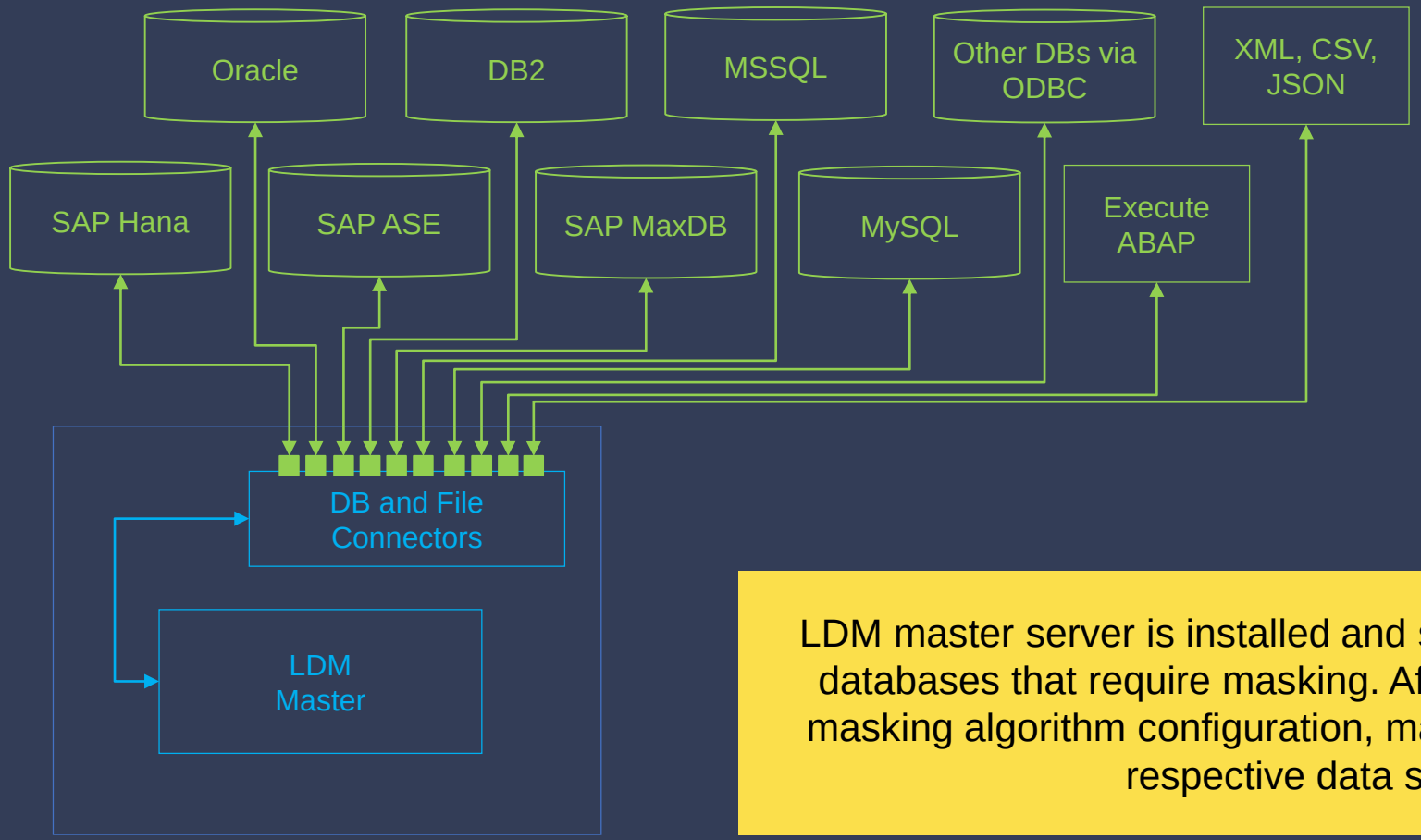
LDM is a software solution from Libelle to obfuscate data.
LDM installs in a customer data center and is under customer's sole control.

LDM Components

- **LDM Master Server:** Installed on Linux, Windows, or UNIX. Master holds configuration data, templates, connection data, etc.
- **LDM UI:** A web-based console to setup configuration and execute masking. Users authenticate with the master via username/password combinations or via LDAP/ Active Directory.
- **LDM Masking Profiles**: A list of tables, columns, and fields that are included for masking.
- **LDM Masking Algorithms**: Pre-configured algorithms to anonymize data such as replacing numbers, names, etc.
- **LDM Reference Database**: A Libelle-or customer provided reference database with names, addresses, and other data that is used to replace original data..

LDM is an obfuscation framework with predefined profiles and algorithm.
It supports enterprise-level data masking efforts.

Master Server Installation



LDM master server is installed and set up to connect to the databases that require masking. After masking fields and masking algorithm configuration, masking executes on the respective data store.

LDM Masking Algorithms

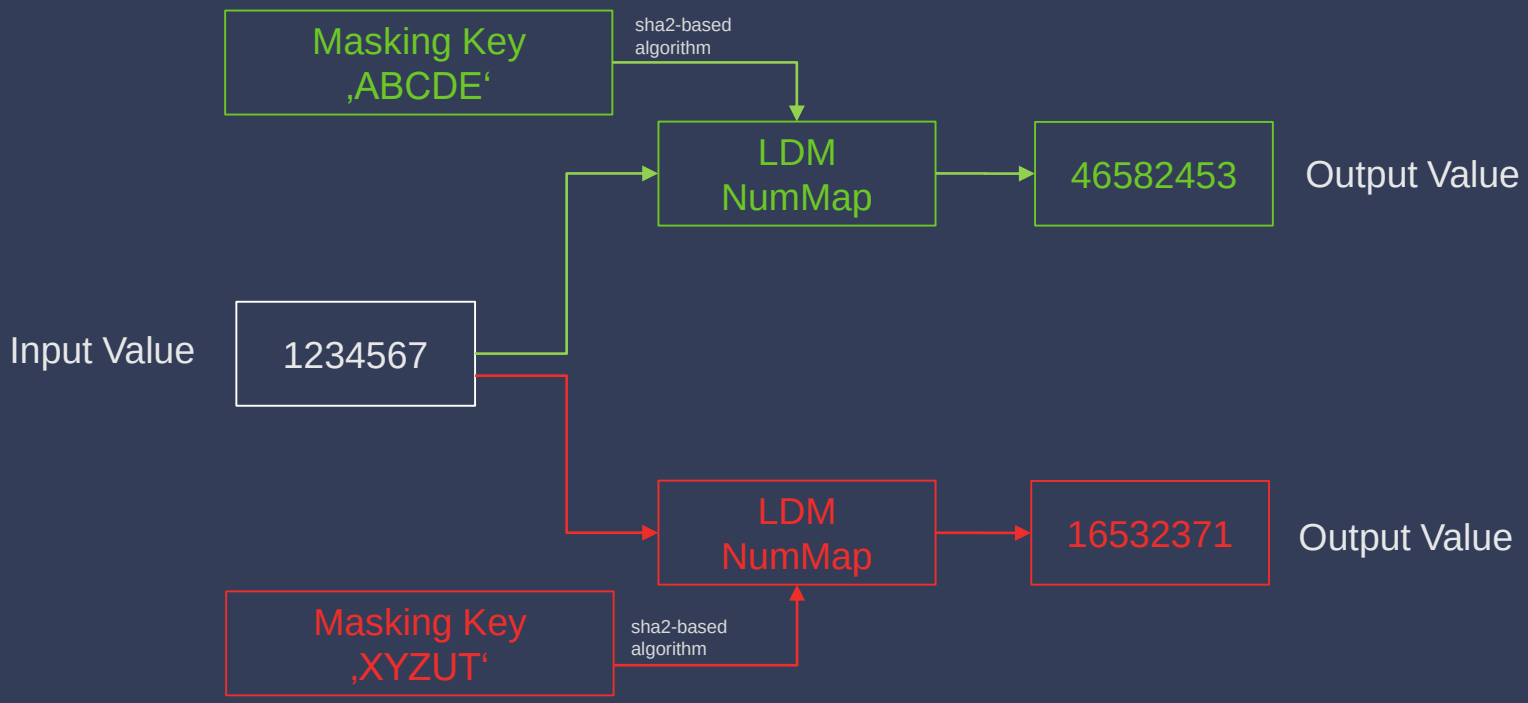
- LDM comes with numerous standard masking algorithms,
- Customers can adjust existing algorithms or develop their own.
- Generally, all LDM default algorithms ignore input values defined as ‘White Space’, ‘Empty String’, and ‘Nil’.
- Masking algorithms are extended on an ongoing basis and there are 40+ algorithms available.
- A variety of options, such as conditional masking or masking groups, accommodate specific requirements for post-masking data consistency.

There are 40+ default masking algorithms available, with various options for additional parametrizations such as setting ranges for output values, conditional masking, or groups that are masked together (e.g. a masked country code and country) to ensure consistency.

Masking Algorithms Overview

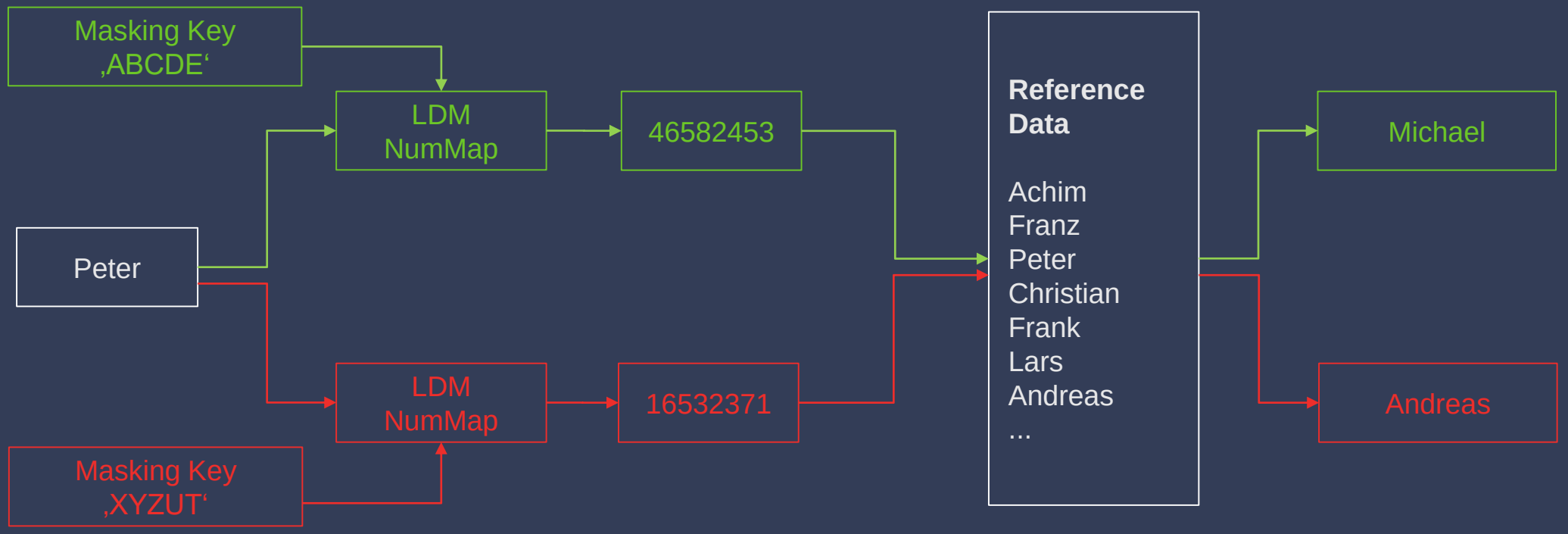
- Generic Number Masking
- Generic Alphanumeric Masking
- Masking Algorithms for Names via Reference DB
- Masking via Mapping Input with predefined Output values
- Masking Sets (group-masking multiple connected field)
- Conditional Masking
- Email Masking
- Credit Card Masking
- Constants, Empty, and Truncates
- Date & Time Masking
- Etc.

Number Masking via cryptographic Hashing



LDM masking keys create SHA2-based lookup tables.
Each unique input masks to the same output with the same masking key.
Numbers are generally masked bijective (generate unique output values)

Consistent Name Masking via Masking Keys

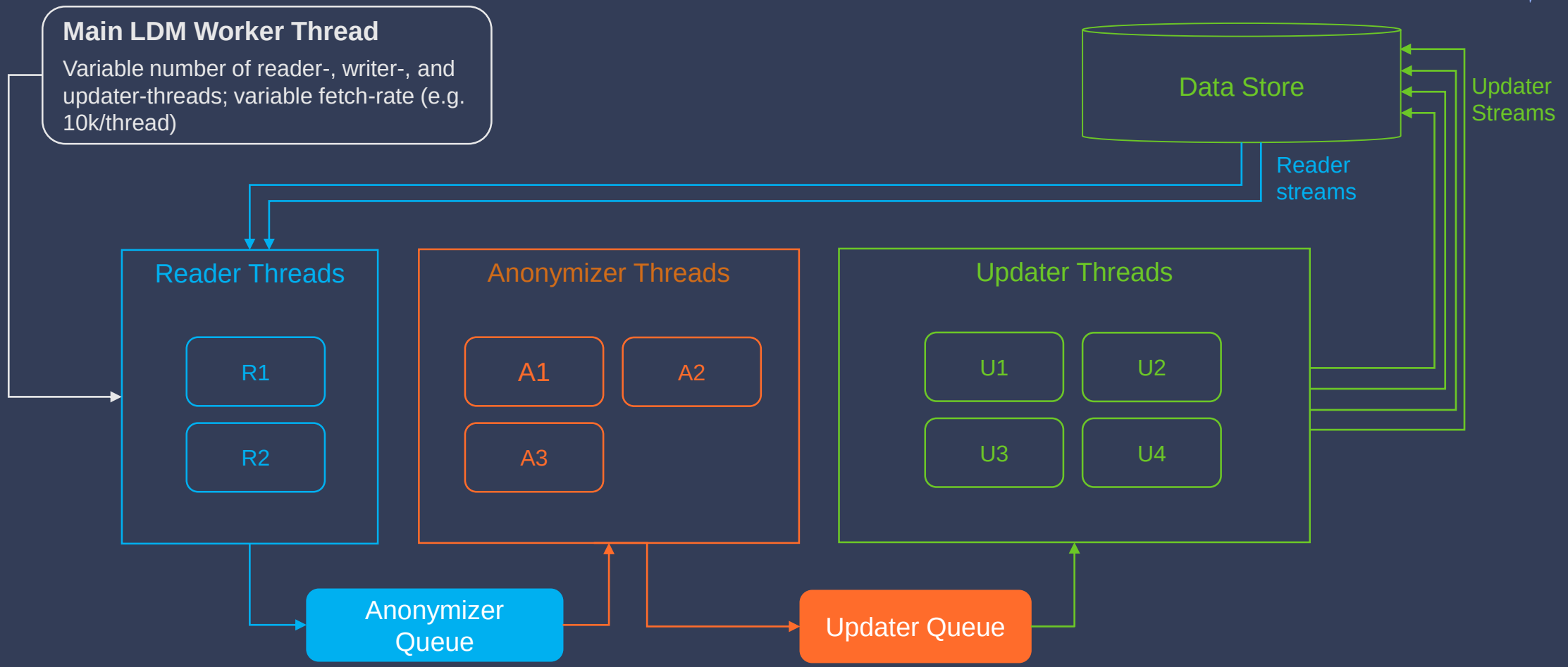


LDM masking keys create SHA2-based lookup tables.
Each unique input masks to the same output with the same masking key.
Names are masked surjective (each input has the same output, but outputs may repeat).

Example: Alphanumeric Masking Algorithms

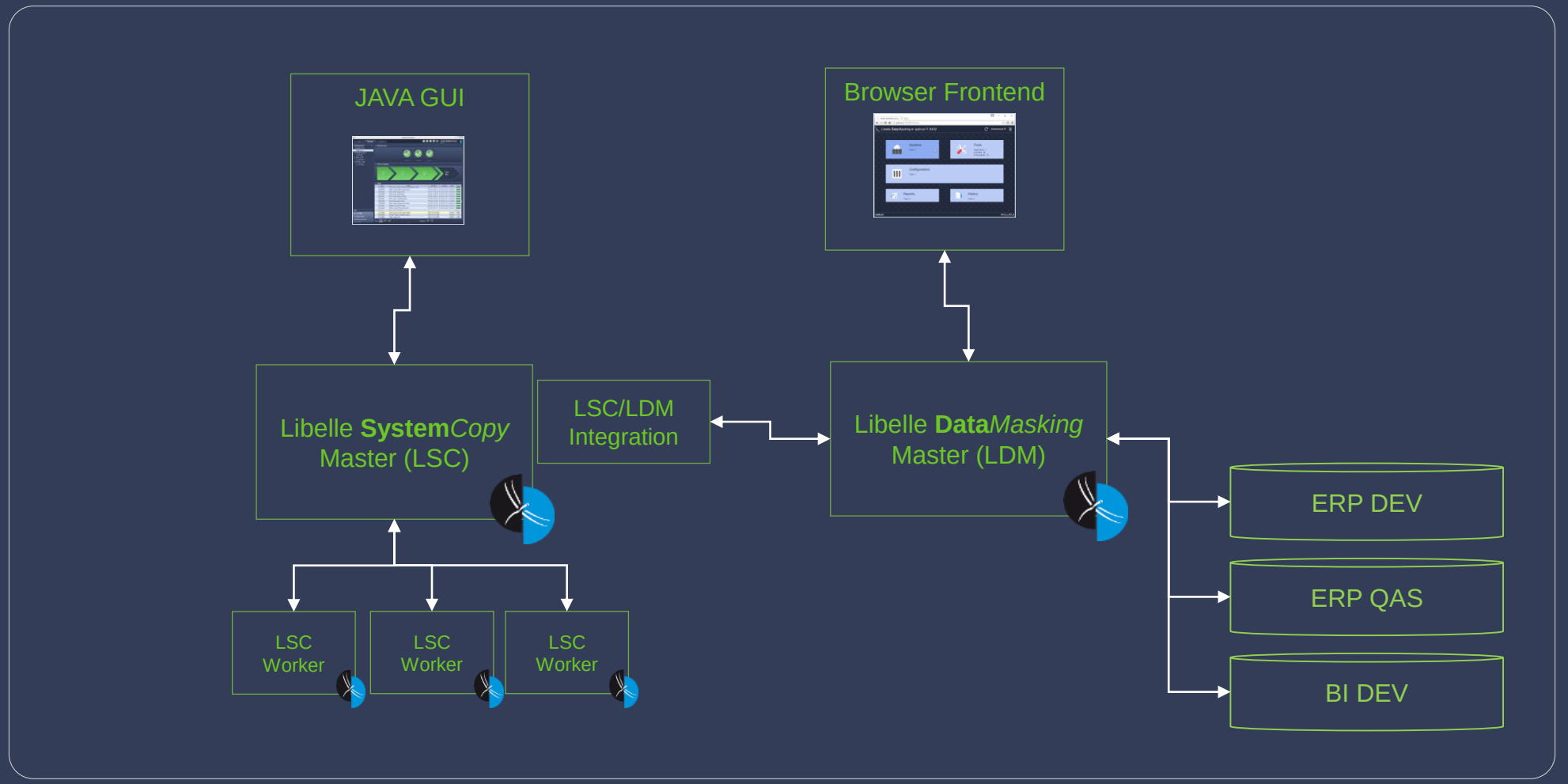
Profile	Attributes	Example (Input/Output)	
aAlphabetic	Anonymize only Latin characters in ASCII between. Numbers will be ignored.	S-L-1234 ABB 1234	M-P-1234 JVV 1234
aAlphanumeric	Combines aNumber and aAlphabetic, so that both numbers and characters are masked.	S-L-1234 Alp12 1234	M-P-0356 Khj79 0356
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 ᾠA'AνάάΘΰZζη 弓女八元255451 غيؤ بـُؤه
aSerial	Anonymize - ignore leading zeros and anonymize like aNumber the rest of value.	00123 0012S3 1234 0	00301 0030S1 0356 0
…	,,,		

Masking Execution via Multi-Threading



30,000 – 50,000 rows updated per second

LSC/LDM Integration



Summary

- Libelle provides standard software solution for masking
- Clearly defined Masking Profiles and Standards in the organization
- Standard software enforces standard workflows. Documents by reports
- Provide technical foundation for organizations to work improving their compliance goals for data
- Centralizes technical masking execution and enforces structured methodological application of masking
- Intelligent Masking
  - LDM is toolset build for masking to help customers mask data consistently, safely, efficiently.
  - Methodological approach to avoid de-identification of masked data by defined algorithms
  - Mask data while retaining the essence of data for Research and Clinical trials