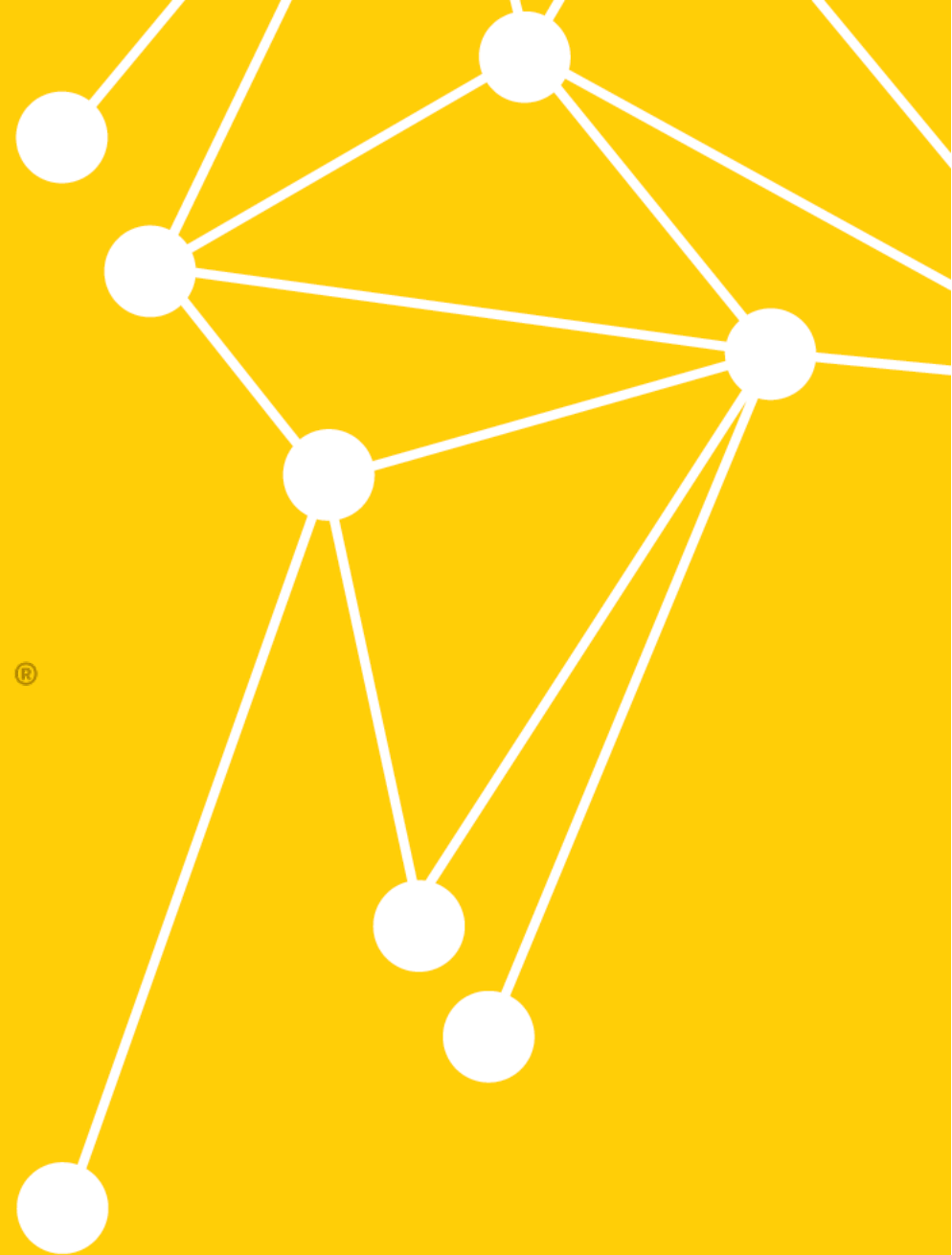




Gold Partner

# Data Masking Unmasked

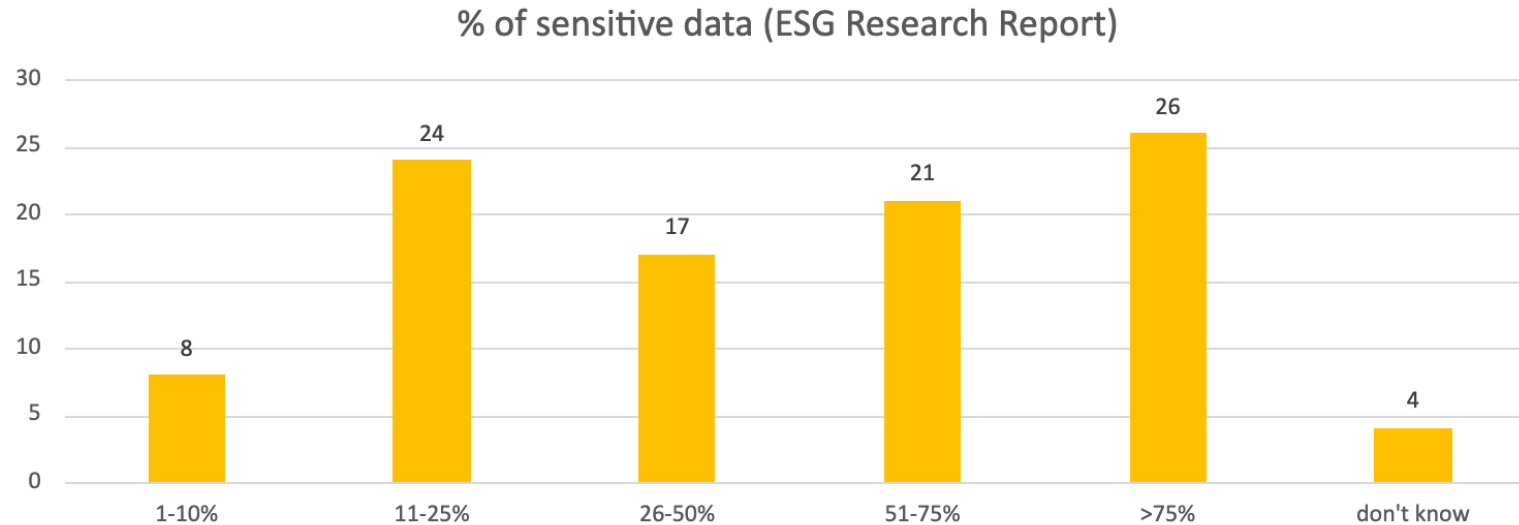
Šta se krije iza zaštite  
podataka?



# Data Privacy

Data Privacy is relationship between Technology and Privacy

Almost every organization have sensitive data



**54%**



of data stored in  
Database is  
confidential

**40%**



of Electronic  
documents is  
confidential

**31%**



of E-mail and  
attachments is  
confidential

# Data Privacy – Sensitive data

Efforts in protecting the sensitive data are driven mostly by cost of data breach and government/industry regulations

## Sensitive data types:

- ✦ Personally identifiable information (PII) Data – Data that uniquely identifies a person, such as: full-name, address, phone, email, ID number, passport, date of birth, driver's licence number
- ✦ Criminal Offence Data
- ✦ Payment card industry (PCI) - Credit card number, expiration date, security code.
- ✦ Protected Health Information (PHI) - Health Insurance Portability and Accountability Act of 1996 (HIPAA) defines the protection requirements of PHI.
- ✦ Custom confidential business data



## Compliance

GDPR, HIPAA, PCI, PII, Gramm-Leach-Bliley Act, Data Protection Act (UK)

# Neglecting Data Privacy

## Reasons:

- ✦ Easy to use production database mirror in non-production environments
- ✦ Hard to generate data for non-production envs which will be suitable for real-world test cases and performance benchmarks
- ✦ Complex data structure, a lot of triggers, multiple applications working together, referential integrity, integrations with production systems, etc.
- ✦ Underestimating internal threats
- ✦ Confusing Data Privacy with Data Encryption

## Consequences:

- ✦ Legal, costs
- ✦ Loss of trust (customers, vendors, partners)
- ✦ Negative publicity, damage to reputation
- ✦ Executive dismissal

# Data in non-production environments

- ✦ Mirror of production database (usually, database sync job done over the weekend)
- ✦ Subset of the production data goes to non-production environments
- ✦ Generated data (seed, faker libraries, etc.)
- ✦ Data masking – creating non-production data by applying data masking to production data. Also known as data obfuscation, sanitization, scrambling, de-identification, etc.

# Solutions

## IBM Optim Data Privacy & TDM

- ✦ Data masking
- ✦ Data discovery and classification
- ✦ Test data management
- ✦ Audit and compliance reporting
- ✦ Application and report masking

## IBM Guardium

- ✦ Compliance and audit reporting
- ✦ Data protection features
- ✦ Vulnerability assessment
- ✦ Data risk analytics & anomaly detection
- ✦ Data discovery & classification
- ✦ Real-time database activity monitoring

# Example of masked data

Ime i Prezime	JMBG	Fizicka aktivnost sati nedeljno	BMI	Nivo holesterola u krvi mmol/l
Dragana Spasojevic	1207978150230	25	28	4.02
Ivan Dragovic	0412998716094	10	34	5.12
Nemanja Zdravkovic	2208980713129	20	27	3.68

Ime i Prezime	JMBG	Fizicka aktivnost sati nedeljno	BMI	Nivo holesterola u krvi mmol/l
Milica Perisic	2105978150237	25	28	4.01
Marko Jovanov	1104998716091	10	34	5.11
Nikola Milutinovic	0102980713129	20	27	3.70

# Data masking in non-production envs

## Challenges:

- ✦ Data utility – masked data must look and act like the real data
- ✦ Data relationships – must be maintained after masking
- ✦ Data masking needs to fit into current business processes
- ✦ Ease of use – must balance ease of use with need to intelligently mask data
- ✦ Working with clusters of applications – applications which work together, have separate databases but still need to have cross-consistent data. It's like referential integrity but between different databases
- ✦ Vendor software with complex DB schema – for example SiebelCRM, SAP, etc.
- ✦ Identifying which data should be masked, and which must not (for example integration IDs used to pull data from external source which is not part of the masking process).

# Data masking system expectations

Different Data masking solutions offer different set of capabilities, but there is a common ground which needs to be covered

- ✦ Keeps referential integrity
- ✦ Data consistency (same original value gets replaced in the same way in all places)
- ✦ Format validation
- ✦ Identification of Business Object, so that functional analysis can take broader look at the data, enabling the easier changes
- ✦ Supports automation
- ✦ Automatically account for database changes
- ✦ Simulation of masking, comparing the data



# Collaborative implementation

Successful implementation of data masking requires collaboration and expertise from different teams.

- ✦ Compliance Officer
- ✦ Enterprise Architect
- ✦ Network Administrator
- ✦ Storage and DB Administrator
- ✦ QA

Essential: Planning and proper methodology





Thank you!

[SparkAnalytics.net](https://SparkAnalytics.net)