



A COMPONENT-BASED
DATA MASKING SOLUTION

AVERAGE ORGANIZATIONS RISKS OF EXPOSED PII

Personally Identifiable Information sensitive and critical organizational resource

- Credit Card Numbers
- Social Security Numbers
- Names
- DOBs



PII Data in non-production environments is exposed to domestic and international development personnel, including code developers, database administrators and QA personnel.

In a study by the Software Engineering Institute of 80 fraud cases, **34 %** involved **Personally Identifiable Information**

PRIVACY COMPLIANCE

FINANCIAL

The Gramm-Leach-Bliley Act (GLBA) , US Congress 1999.

HEALTH / PHARMA

Health Insurance Portability and Accountability Act(HIPAA/HITECH), US Congress 1996.

ECOMMERCE

Payment Card Industry Data Security Standard (PCI DSS), Payment Card Industry Security Standards Council.

SOLUTION

Masking data in non-production environments.

Identity based masking in production environments and reporting.

HIGH PROFILE PRIVACY BREACHES



LATELY:

CHASE, TARGET, HOME DEPOT, APPLE, CEDARS-SINAI,
UNIVERSITY OF CALIFORNIA?

FAMOUS: MONEY GRAM

- \$100,000,000 fine
- Involved in fraud due to PII exposure in 2009, 2012
 - Now uses IBM data masking software Optim per GLBA
 - The software installation itself costs millions of \$\$\$

HEARTLAND PAYMENT SYSTEMS

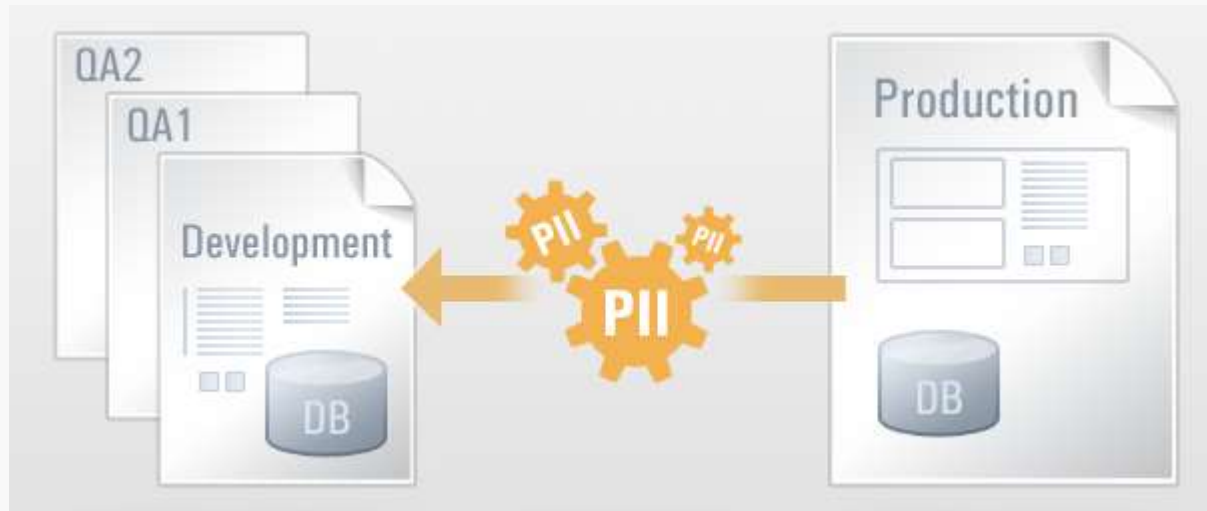
- 130M credit card numbers
- Albert Gonzalez used SQL injection in internal storage
 - Now it implements **end-to-end encryption**

ENCRYPTION IS NOT MASKING!



DIFFERENT USE CASES: EXTERNAL VS. INTERNAL FRAUD

DATA MASKING DEFINITION



The process of masking specific data elements within data store while preserving data look and feel and usability in applications.

ALGORITHMIC CHALLENGE
DATA INTEGRITY CHALLENGE
STATISTICAL CHALLENGE

ALGORITHMIC CHALLENGE – WHY?

Data Masking is NOT encryption

ENCRYPTION: CHANGE VALUE BY CALCULATING LONG UNDECIPHERABLE ENCODING



MASKING: SAME LOOK - DIFFERENT VALUE. DATA REMAINS USEFUL FOR TESTING.



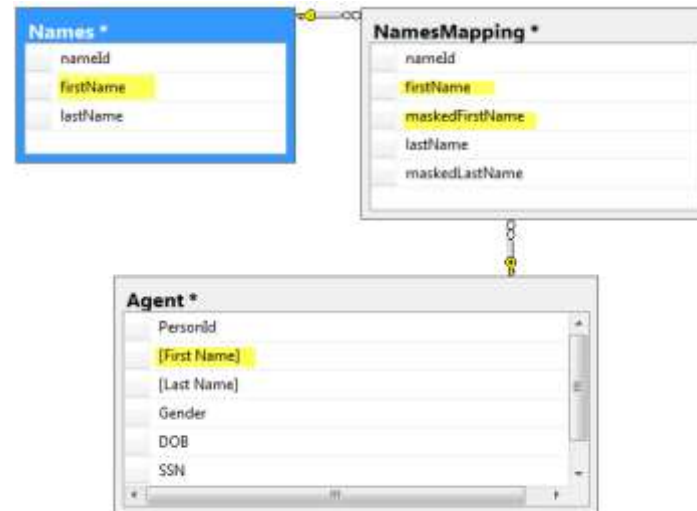
DATA INTEGRITY CHALLENGE

CREATING AND MAINTAINING MAPPING TABLES

It is a tedious process that requires creating a lot of data on disk and a lot of structures and code to maintain primary –foreign key relationships. Here are some excerpts from the Oracle manual:

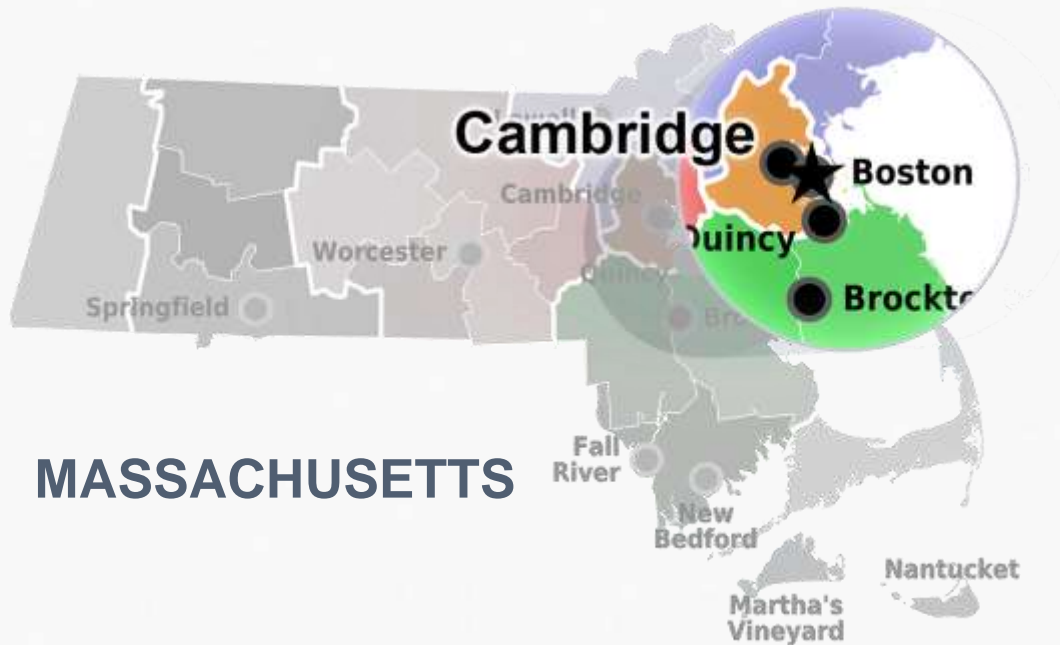
“The typical data masking process... usually involves...

- 1. clone non-production database*
- 2. build the tables and columns chosen for masking*
- 3. tables selected for masking are processed in the optimal order. Typically, the tables with the primary keys get masked first, followed by the dependent tables containing foreign keys.”*



PROBLEM: STATISTICS AS AN ENEMY

How difficult is it to find William Weld who lives in Cambridge, MA, city of 54,000 residents with 7 zip codes?



ZIP,
DOB,
SEX,
PUBLIC DATA

MASSACHUSETTS

WILLIAM WELD
MEDICAL
HISTORY

Only six people in Cambridge share the governor's birth date, only three of them men, and of them, only he lives in his ZIP code. 87 percent of all Americans could be **uniquely identified using only three bits of information**: ZIP code, birthdate, and sex.

KNOW YOUR ENEMY!

SOLUTION: PROPER DATA MASKING



VARIETY OF ALGORITHMS

Unique Elements:

- Stay Unique. FPE-like algorithm.

Maintaining Integrity:

- Newly masked values always map consistently. FPE-like algorithm.

Same Ol' Random:

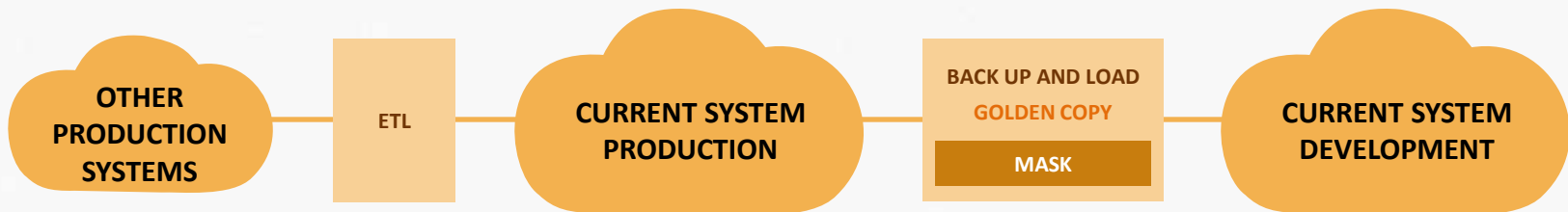
- Randomly created mapping tables.

Resolving Statistical Distribution:

- **Patent Pending.** We approach from a different angle.

THE USUAL DATA LIFECYCLE

Enough “real” PII data must be used to create enough use cases to develop the new system properly



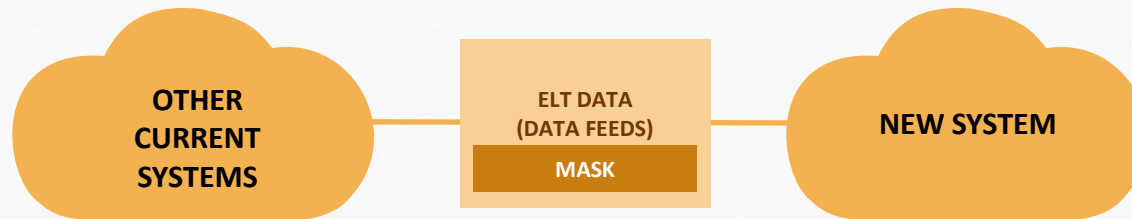
However, this real data should be masked to avoid identity, other theft

Masking is usually done on the backup “golden copy” of the data base. Then data is reduced and applied to non-production.

FOR DEVELOPING A NEW SYSTEM WITHIN AN EXISTING ORGANIZATION

Enough “real” PII data must be used to create enough use cases to develop the new system properly

However, this real data should be masked to avoid identity, other theft



Masking within ETL prevents data breach from the beginning

In the new system, realistic data is used to develop the system with no chance of recovering the production data.

WHY HUSH HUSH FOR DATA MASKING?



SIMPLE

- Fast “Time to Market”
- Drag-and-Drop encapsulated algorithms

EFFORTLESS ENTERPRISE-WIDE RI

HIGHLY CUSTOMIZABLE

INTRODUCES JUST IN TIME PRIVACY PROTECTION

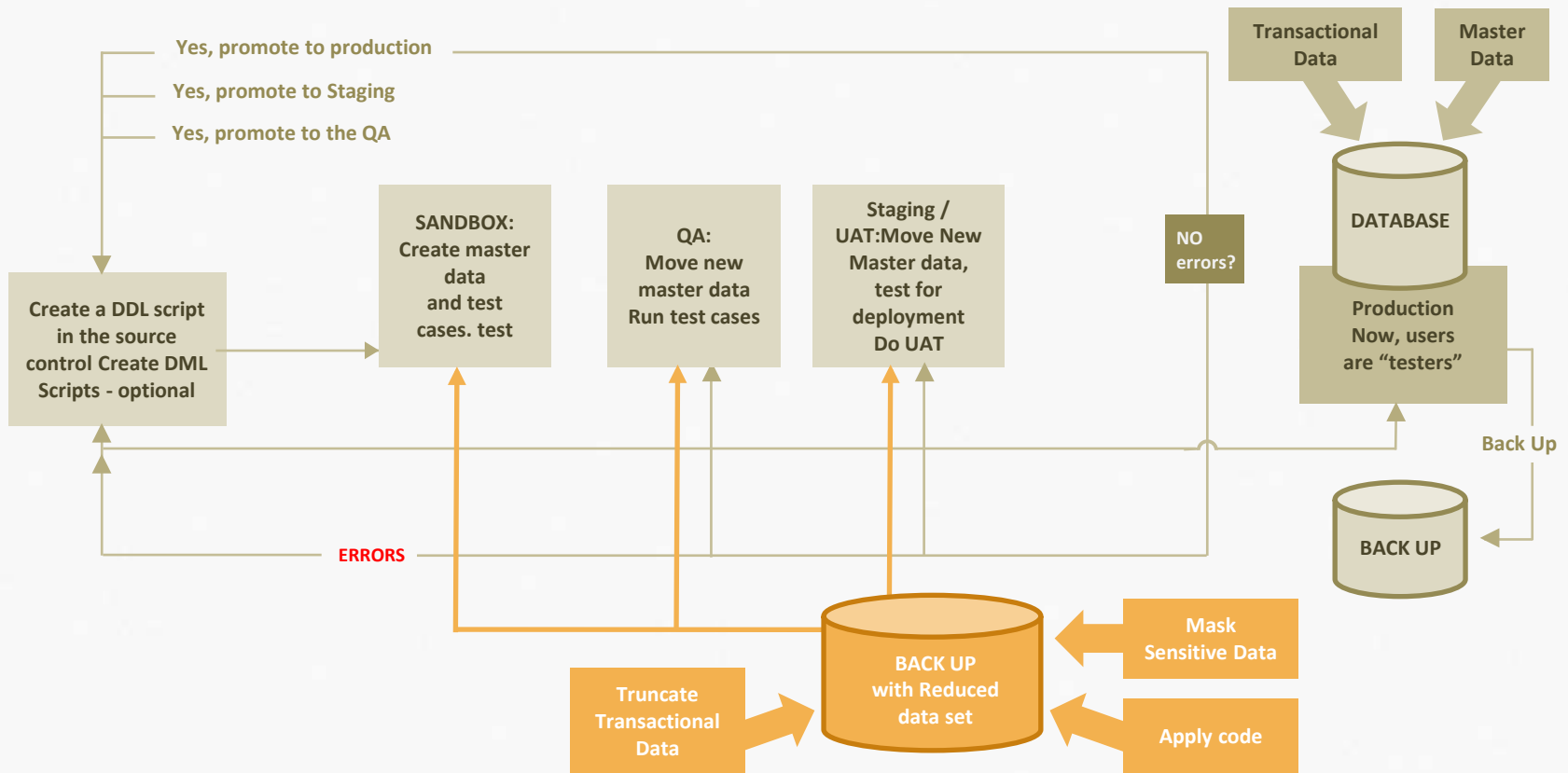
- In reporting and in real-time solutions, policy-based
- Extremely scalable and available in the cloud

The screenshot shows a data flow diagram in Microsoft Visual Studio. The diagram includes an OLE DB Source, a Masking SSN Info task, and OLE DB Destination and Flat File Destination tasks. Below the diagram is a report titled "HUSHHUSH PERSON REPORT" with the following data:

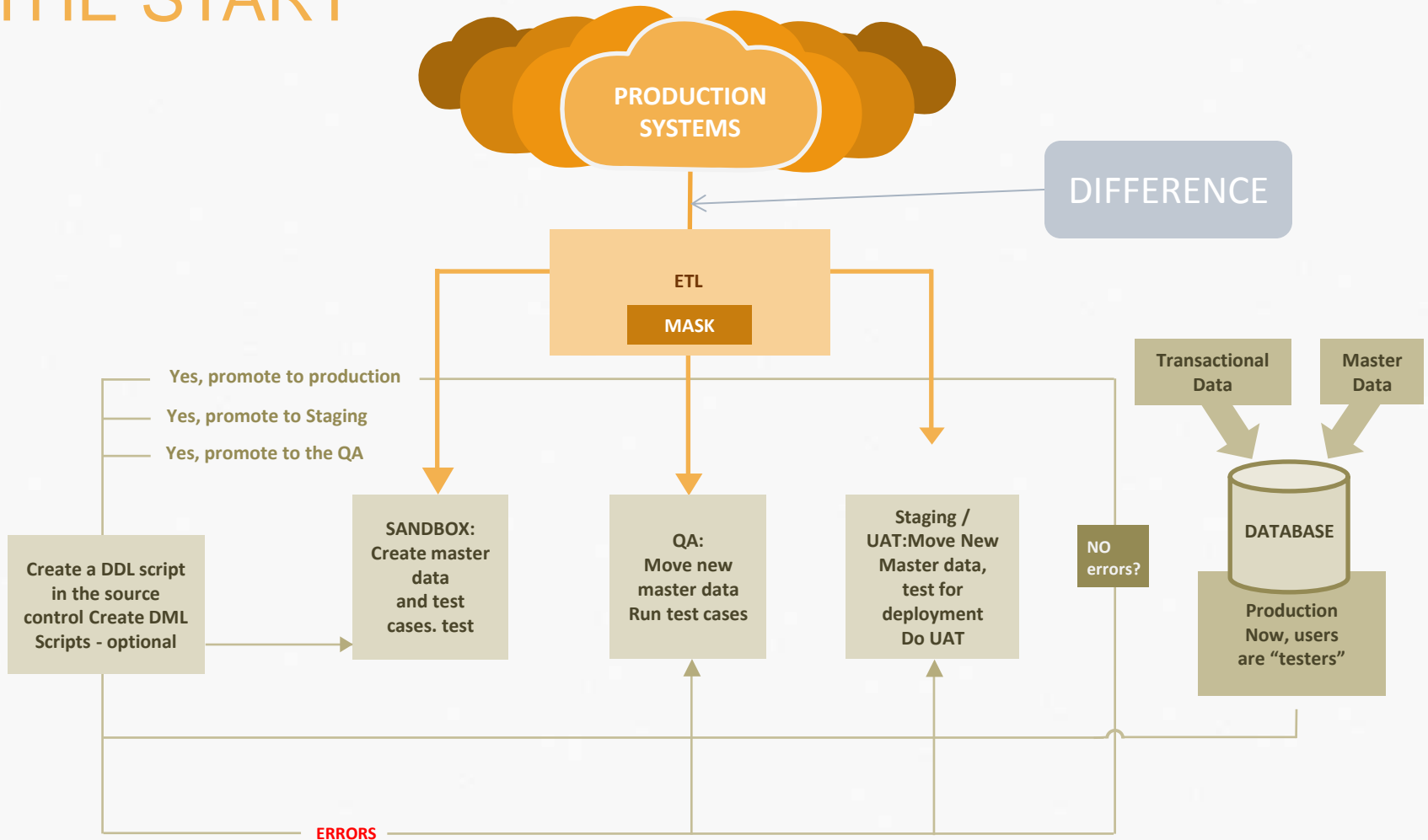
Person Id	First Name	Last Name	DOB	SSN
	Catherina	Reyes	8/22/1900 12:00:00 AM	123-45-6789
	Irvin	Harrington	8/22/1900 12:00:00 AM	123-12-3456
	Catherina	WADE	8/22/1900 12:00:00 AM	iam-an-error
	Irvin	REYES	8/22/1900 12:00:00 AM	XXX-XX-XXXX
	Catherina	BOWERS	8/22/1900 12:00:00 AM	XXX-XX-XXXX

APPENDIX

SOLUTIONS: CURRENT WAY



APPENDIX OFFER: BE COMPLIANT WITH HH FROM THE START



TEAM:

- **Virginia Mushkatblat**

Chief of Everything in HushHush

- In the past performed multiple roles in the data domain, including enterprise architecture, development, and management in the fortune 500 companies. She has consulting and business development experience. Industry verticals include financial, technology, advertisement, entertainment, e-commerce and engineering. She possesses B.S. In Engineering and M.S. In Computer Science. Patents filed on her behalf by TV Guide International.

- **Yury Lobzakov**

Algorithms Architect

- Former Software Architect for Research in Motion, owner of multiple patents, MFTI (Russian MIT) graduate has extensive experience in technology companies in multiple roles

- **G6G Consulting Group**

- Marketing and Sales advisory and help in creating marketing materials. The consultants qualifications include MBA degrees and experience in the sales and marketing roles of increasing level of responsibility in multibillion dollar companies.