# Data Governance, Protection and Privacy - Holistic Data Compliance Approach

**Sheshnarayan Agrawal**
**(shagrawal@in.ibm.com)**
**IBM India Pvt. Ltd.**

**InformationOnDemandIndia2011**

The Premier Conference for Information Management
**Manage. Analyze. Govern.**

**February 2, 2011**
Hyatt Regency I Mumbai, India

# Agenda

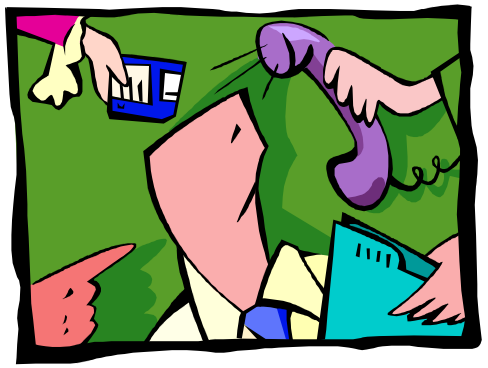
- Data Governance & Protecting Privacy
- What's at Stake?
- Considerations for a Privacy Project
- IBM Infosphere Optim
- Success Stories
- Q & A

# What is Data Governance?

*Data governance is the orchestration of people, process and technology to enable an organization to leverage information as an enterprise asset.*

*Data Governance safeguards information, keeps auditors and regulators satisfied, uses improved data quality to retain customers and constituents and drive new opportunities*

- Data Governance calls for "stewardship" of sensitive data
  - Enforced by regulations
  - Expected by customers
  - Demanded by executives



# What is Done to Protect Data Today?

**Production Environment**

- Production “Lockdown”
- Physical entry access controls
- Network, application and database-level security
- Multi-factor authentication schemes (tokens, biometrics)

**What About Non-Production Environment**

- Unique challenges in Development and Test Environment
- Replication of production safeguards not sufficient
- Need “realistic” data to test accurately

# What's at Stake?

- Fines and penalties
- Lawsuits
- Loss of customer loyalty
- Loss of revenue
- Share price erosion
- Negative publicity
- "Brand equity" damage
- Damage to company reputation
- Increased operations costs
- Intellectual property loss

## Where Data Theft Happens

- Data mistakenly left behind
  - Laptops
  - Hard drives
  - Thumb drives
- Data exposed in testing and training environments
  - Outsourcers
  - Internal employees
- Application breaches

# Consequences of Ineffective Data Privacy

## Hannaford hit by class-action lawsuits in wake of data-breach disclosur

Attorneys rush
grocer over the

By Jaikumar Vijayan

March 20, 2008 (Com
a Philadelphia law firm
lawsuits against Hann
supermarket chain tha
involving the potential

Philadelphia-based Be
U.S. District Court in
Maine-based attorney
customers in all of the

December 6, 2007

## TJX Proposes $ Inc. In the Large Cardholders

Analysis of: How TJX Becam

This analysis is solely the v

**Implications:** TJX's data bre
stretched from the U.S. to Ca
far as the Ukraine, when one
leaders was arrested for his
data card breach. Reports fir
that more than 45 million car
had been affected, however,
investigation, the number mc
doubled to over 94 million af

28 DEC, 2010, 04.51PM IST, SHAILENDRA BHATNAGAR & MOHIT BHALLA,ET NOW

## Citibank fraud: Rs 400 cr allegedly siphoned off


RELATED ARTICLES

- Banks preoccupied with working interest,deposit rates in 2010
- IOB to convert its China office into full scale branch
- Union Bank to get Rs 1,150 cr from govt by March 2011
- A one-time password to shop using phone

MUMBAI: Citibank is probing a Rs 400 crore fraud perpetrated by certain employees of the bank at its Gurgaon branch in India, atleast three people told ET Now on condition of anonymity. It is learnt that the fraud was discovered by accident and that the banks Asia Pacific fraud risk management team has been camping in India for the past two weeks conducting detailed investigations and questioning several employees who maybe suspected of being involved in prosecuting the fraud.

People familiar with the development told ET Now that the employees involved have been suspected of selling investment products to clients claiming that these would generate unusually high returns. The employees claimed that the products were authorized by the banks investment product committee and used forged bank documents and letterheads to prove the same.

The suspected employees then siphoned off the funds raised from the sale of these products into their personal bank accounts and defrauded clients to the tune of Rs. 400 crore.

The suspected employees had access to High Net Worth Individual (HNI) clients of the bank and were in roles

# Data Privacy

What do we mean by Data Privacy?

- Information about a personally identifiable individual or other potentially sensitive organizational data like Name, address, phone number, bank record, SSN, etc.

What do we mean by Protecting Data Privacy?

- Rendering Data Unusable
  - Encryption
  - Data Masking

# Drivers for Privacy of non-production Data

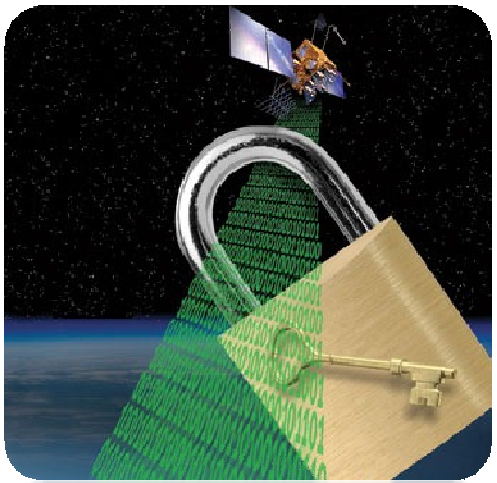
- Regulatory & Compliance

| | | |
|---|---|---|
| HIPAA<br>(CMS) | Organizations that handle patient health information | Confidentiality, integrity and availability of patient health information |
| Gramm-Leach-Bliley<br>(SEC, FTC, FDIC…) | Credit card issuers<br>All financial services. | Protection of consumer information |
| PCI<br>(Visa, MC, Discover, AMEX) | Major retailers and processors | Protection of credit card data |

- Offshoring test
- Subcontracting test & dev.
- Sensitive data
- Training Environment
- Good business practice

# Rendering Data Unusable - Encryption

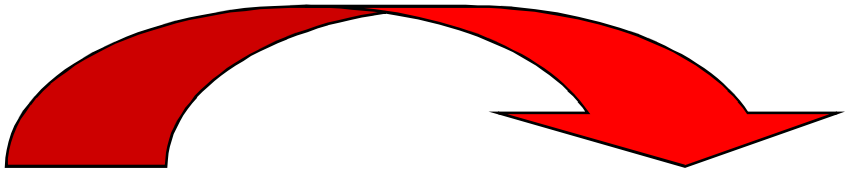
- Encryption:  Protecting data where it lies
- Definition:  Coding or scrambling of information so that it can only be decoded and read by someone who has the correct decoding key
  - Must be transparent to application, database and storage environments
  - Provide unified policy and key management for protecting data in both online (file system) and off-line (backup) environments
  - Protects against unauthorized attempts to view files…and attacks against the database operating environment
  - Protects stolen/lost media



# Encryption is NOT enough

- DBMS encryption protects DBMS theft and hackers
- Data decryption occurs as data is retrieved from the DBMS
- Application testing displays data
  - Web screens under development
  - Reports
  - Date entry/update client/server devices
- If data can be seen it can be copied
  - Download
  - Screen captures
  - Simple picture of a screen

# Masking




Before Masking


After Masking

# Data Masking Considerations

- **Understand Application and Business Requirements**
  - Where do applications exist?
  - What is the purpose of the application's?
  - How close does replacement data need to match the original data?
  - How much data needs to be masked?

- **Determine what you need to mask**
  - Customer Information
  - Employee Information
  - Company Trade Secrets
  - Other

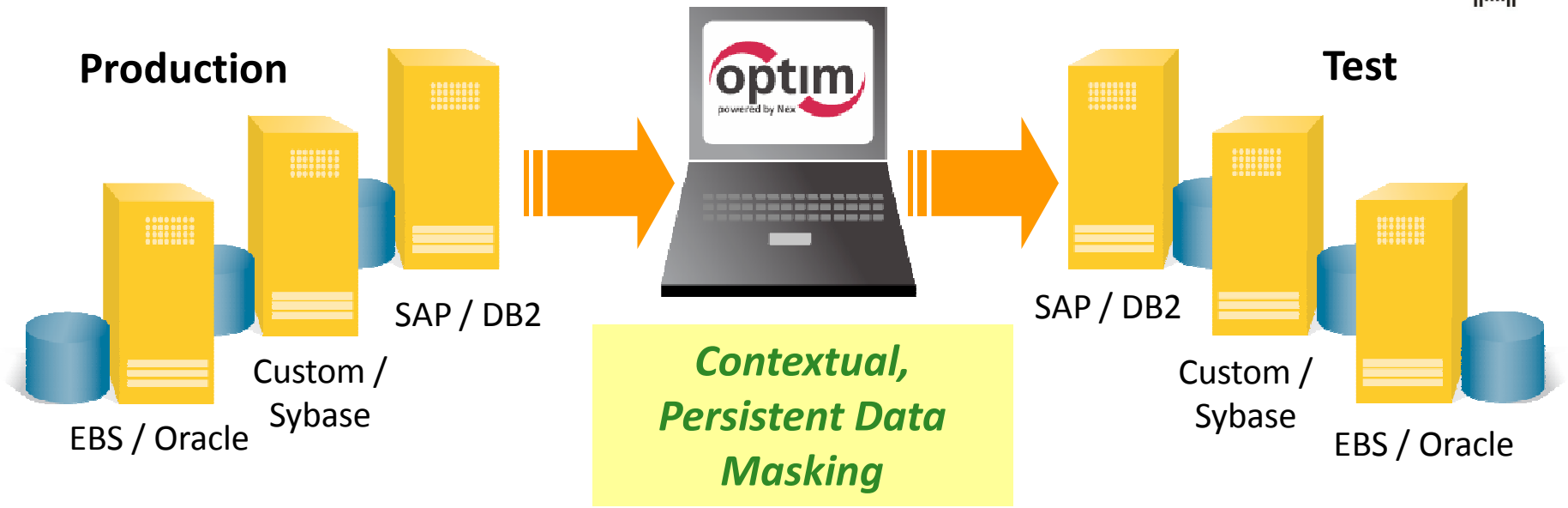
# Rendering Data Unusable - Masking

- Data Masking:  Protecting non-production data
- Definition:  Removing, masking or transforming elements that could be used to identify an individual
  - Name, address, telephone, SSN / National Identity number, credit card #…
- Key Components of Masking
  - Masked data should look “realistic”
  - Masked data must be appropriate to the context
  - Should give persistent results
  - Should not break referential integrity of data
- Some other names you may see for masking
  - Obfuscation, Scrambling, Data de-identification, Privacy

# What is Optim

- Joined IBM family as a part of Acquisition
- Proven & Tested over the years (A 10+ year mature product)
- Enterprise Data Management Solution
- Provides 3 Core solutions for…
  - **Data Growth Management**
  - **Data Privacy Protection**
  - **Test Data Management**
- Available in:
  - **Mainframe/zOS**
  - **Client Server/LUW**
  - **AppAware (i.e. SAP, Siebel, Oracle E-business, JD Edwards etc)**
- Supports most of the standard Databases & OS
- Application Aware

# Optim™ Data Privacy Solution



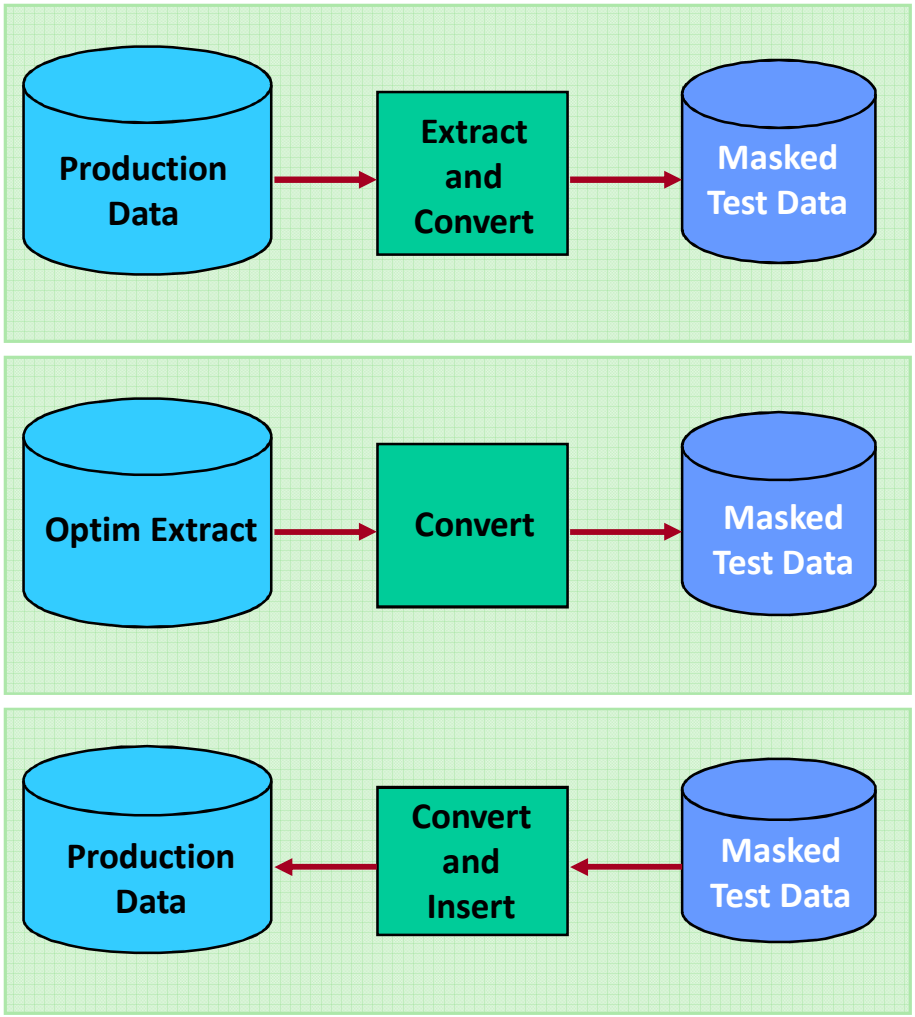

- Substitute confidential information with fictionalized data
- Deploy multiple masking algorithms
- Provide consistency across environments and iterations
- Enable off-shore testing
- Protect private data in non-production environments

# Features: IBM Optim Data Privacy

- Move and mask
  - Move data into a test environment and mask as it is moved using convert.
  - Insert data into database and mask while inserting using convert.
- Persistent Masking
- Contextual Masking
- Key Propagation to retain Referential Integrity
- Advanced masking using Data Transformation Library

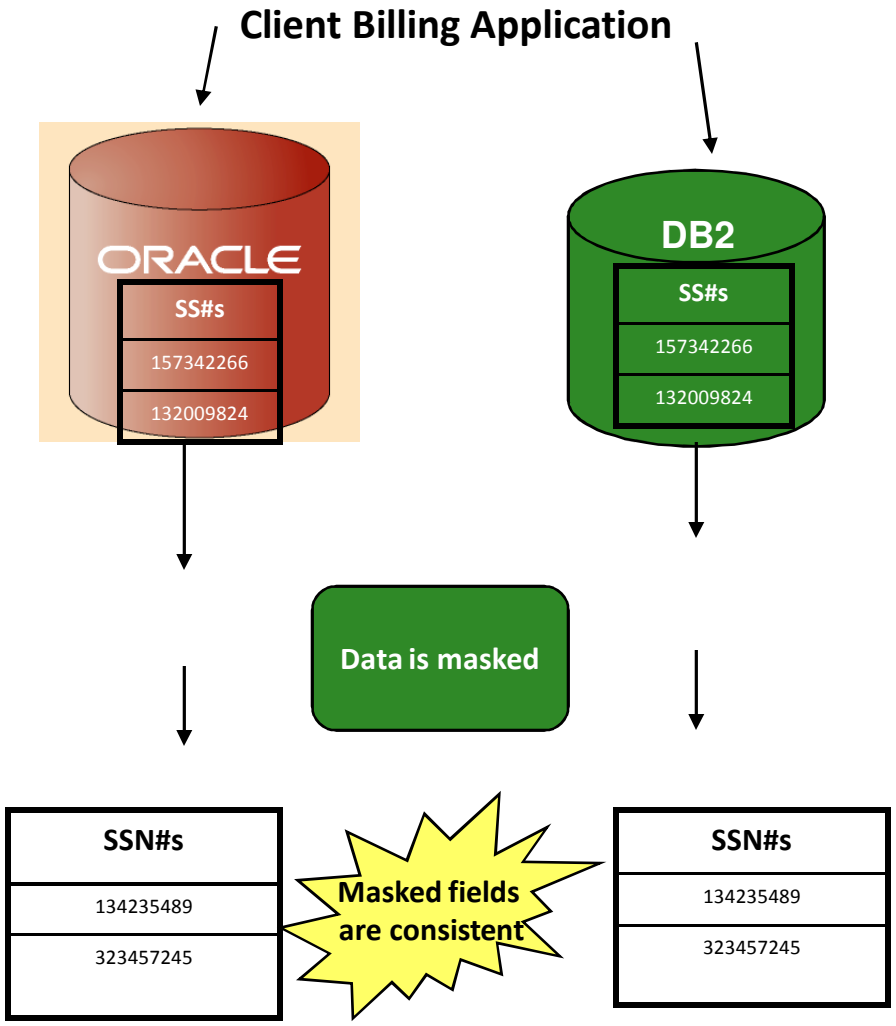
- Replacement via lookup (Uses lookup tables for masking personal details like name, address etc…)
- Simple Transformations using literals, special registers, expressions, arithmetic functions, random numbers, sequential numbers
- Application Aware masking
  - Customized pre-packaged masking routines for SAP, Oracle E-Business Suite, PeopleSoft Enterprise, JD Edwards Enterprise, Siebel Application, Amdocs CRM

# When can Optim Mask Data




- Mask Data during an Extract Process
- Mask Data before an Insert or Load Process
- Mask Data during an Insert Process
- Mask Data during a Load Process

# Persistent Masking




- Irrespective of number of times data is masked output should always be the same for similar data.
  - It is important in an enterprise environment because there could be multiple data sources containing similar data
  - Similar test data set might be required multiple times for testing

# Example: Bank Account Numbers

- First Financial Bank's account numbers are formatted "123-4567" with the first three digits representing the type of account (checking, savings, or money market) and the last four digits representing the customer identification number
- To mask account numbers for testing, use the *actual first three digits*, plus a *sequential four-digit number*
- The result is a fictionalized account number with a valid format:
  - "001-9898" becomes "001-1000"
  - "001-4570" becomes "001-1001"

# Propagating Masked Data

- Key propagation
  - Propagate values in the primary key to all related tables
  - Necessary to maintain referential integrity

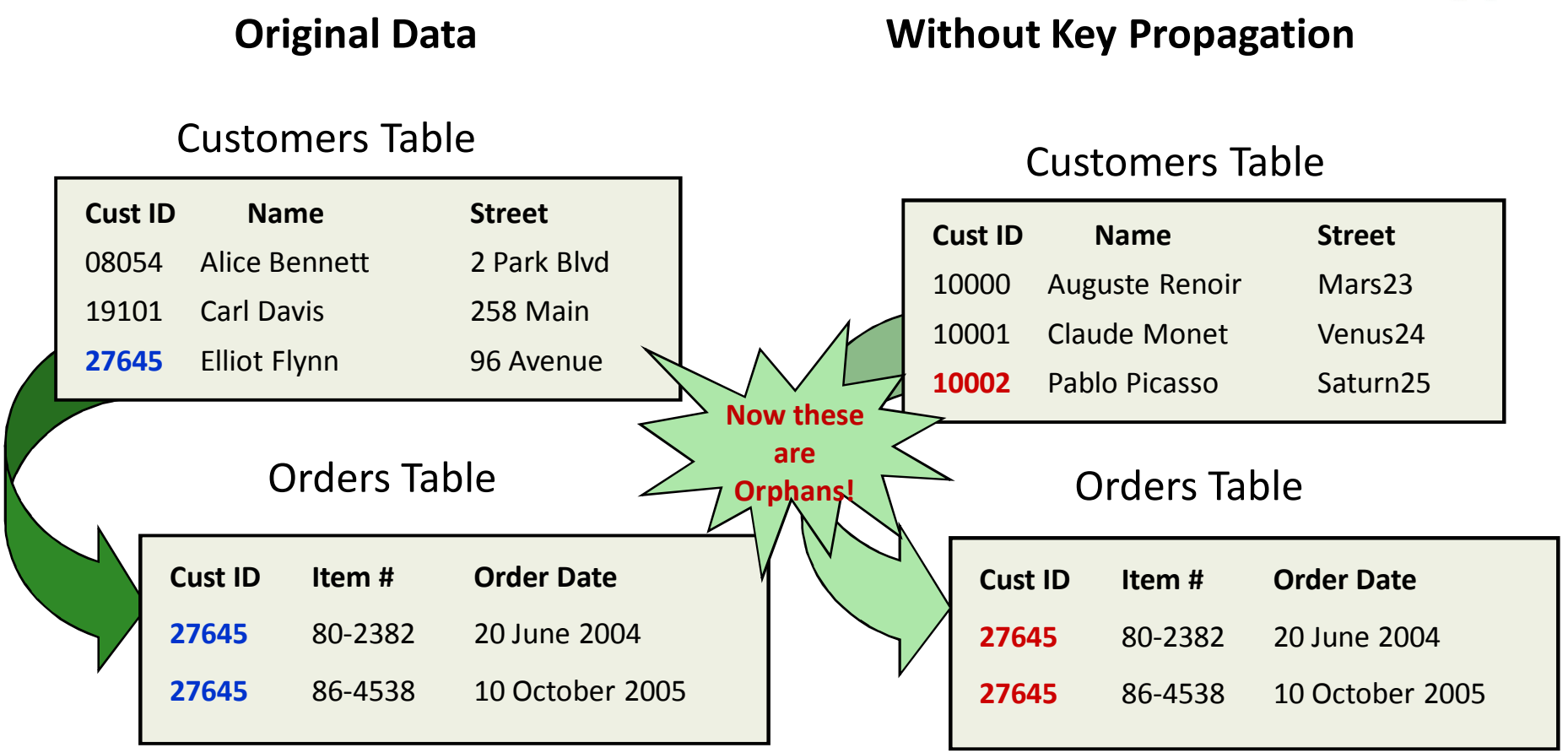
**Customers Table**

| Cust ID | Name | Street |
|---|---|---|
| 08054 | Alice Bennett | 2 Park Blvd |
| 19101 | Carl Davis | 258 Main |
| 27645 | Elliot Flynn | 96 Avenue |

**Orders Table**

| Cust ID | Item # | Order Date |
|---|---|---|
| 27645 | 80-2382 | 20 June 2004 |
| 27645 | 86-4538 | 10 October 2005 |

# Without Key Propagation…



# Masking with Key Propagation

## Original Data

### Customers Table

| Cust ID | Name | Street |
|---|---|---|
| 08054 | Alice Bennett | 2 Park Blvd |
| 19101 | Carl Davis | 258 Main |
| **27645** | Elliot Flynn | 96 Avenue |

### Orders Table

| Cust ID | Item # | Order Date |
|---|---|---|
| **27645** | 80-2382 | 20 June 2004 |
| **27645** | 86-4538 | 10 October 2005 |

**Referential integrity is maintained**

## De-Identified Data

### Customers Table

| Cust ID | Name | Street |
|---|---|---|
| 10000 | Auguste Renoir | Mars23 |
| 10001 | Claude Monet | Venus24 |
| **10002** | Pablo Picasso | Saturn25 |

### Orders Table

| Cust ID | Item # | Order Date |
|---|---|---|
| **10002** | 80-2382 | 20 June 2004 |
| **10002** | 86-4538 | 10 October 2005 |

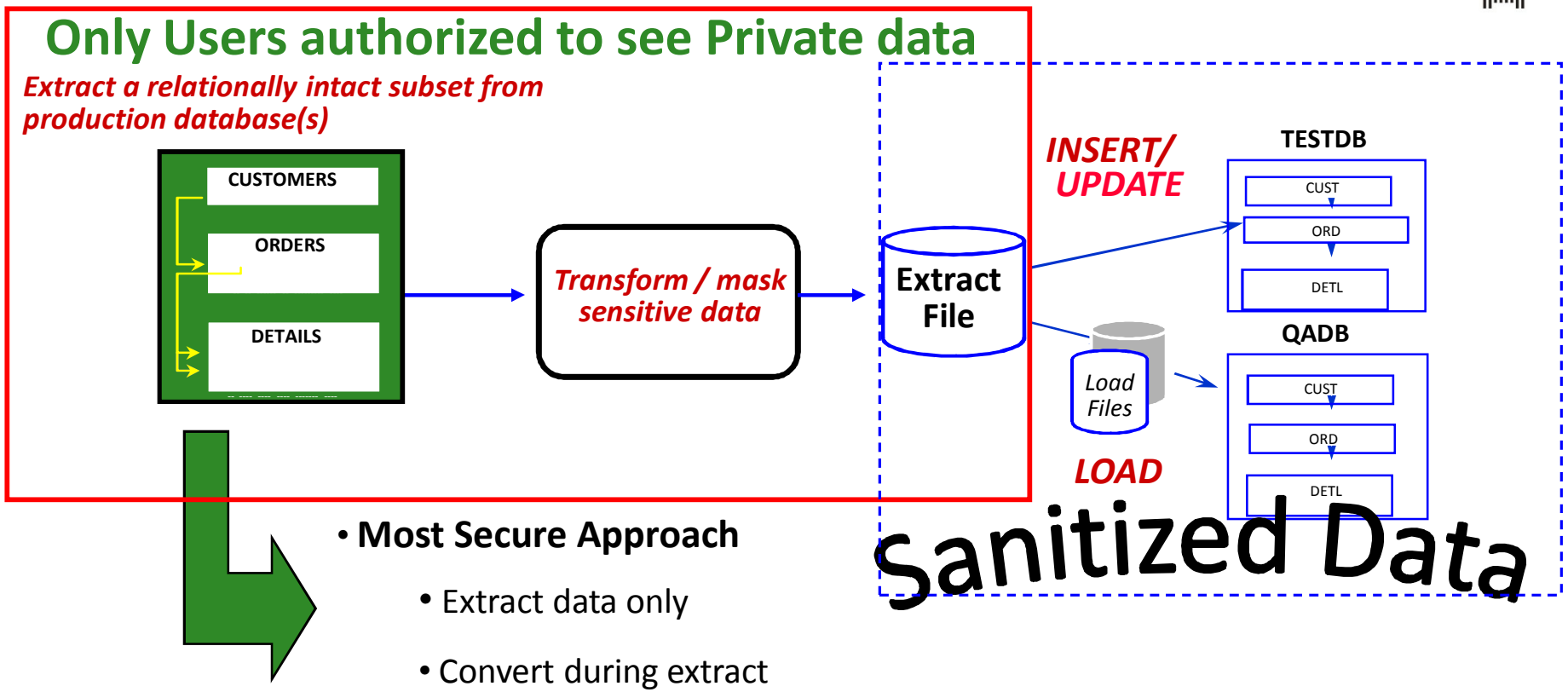
# Example: First and Last Name


- Direct Response Marketing, Inc. is testing its order fulfillment system
- To fictionalize customer names, use the a random lookup function to pull first and last names randomly from the Customer Information table:
  - **“Gerard Depardieu” becomes “Ronald Smith”**
  - **“Lucille Ball” becomes “Elena Wu”**

# Data Privacy Transformation Library

- **Specialized Transformation functions to mask**
  - National Identifiers
  - Credit Card Numbers
  - Email addresses
- **What it can do**
  - Generate Unique Masked values
  - Contextual masking.
  - Masking part by part
  - Validate the Input data
  - Use of Column maps allow
    - Masking Multiple Columns in one go
    - Ease in mapping source columns to destination columns
    - Ease in specifying masking routines

# Data Privacy in Application Testing




- **Most Secure Approach**
  - Extract data only
  - Convert during extract
- **Extract file already contains masked data**
  - Can be shared with testers to reuse

# The Market Need

- Corporations have a duty to protect confidential customer information and have gained an understanding that vulnerabilities exist both in the Production and Test Environments
- Companies have begun implementing basic privacy functionality but are requiring more specific and application aware masking capabilities that can be applied across applications
  - *IT organizations require that development databases provide realistic and valid test data (yet not identifiable) after it is masked. This includes: Valid social security #'s, credit card #'s, etc.*
  - *Enterprises require the option to mask data consistently across several different applications, databases, and platforms*

## Success with Data Masking

- " Today we don't care if we lose a laptop"

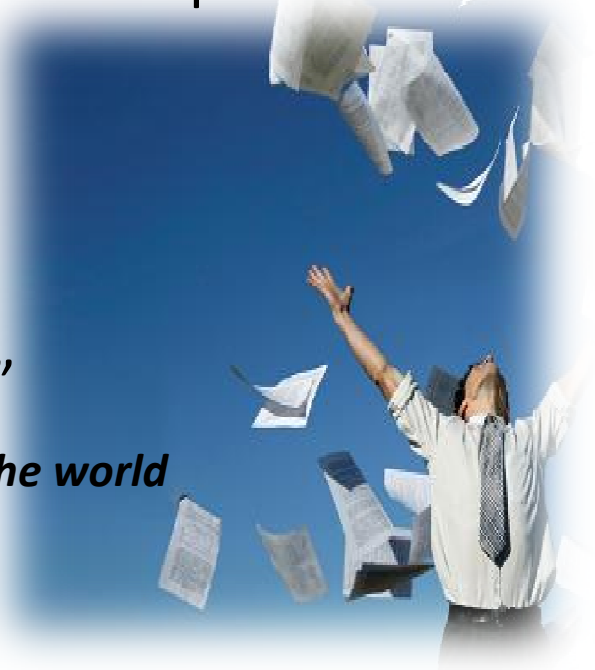
  ***Large Midwest Financial Company***

- " The cost of a data breach is exponentially more expensive than the cost of masking data"

  ***Large East Coast Insurer***

- "This corporation is the only large retailer to state full compliance with PCI regulations"

  ***News article about the largest retailer in the world***



# Success: Data Privacy

About the Client:
$300 Billion Retailer
*Largest Retailer in the World*
*Largest Informix installation in the world*

- **Application:**
  - Multiple interrelated retail transaction processing applications
- **Challenges:**
  - Comply with Payment Card Industry (PCI) regulations that required credit card data to be masked in the testing environment
  - Implement a strategy where Personally Identifiable Information (PII) is de-identified when being utilized in the application development process
  - Obtain a masking solution that could mask data across the enterprise in both Mainframe and Open Systems environments
- **Solution:**
  - IBM Optim™
- **Client Value:**
  - Satisfied PCI requirements by giving this retailer the capability to mask credit data with fictitious data
  - Masked other PII, such as customer first and last names, to ensure that "real data" cannot be extracted from the development environment
  - Adapted an enterprise focus for protecting privacy by deploying a consistent data masking methodology across applications, databases and operating environments

## Success: Global Financial Business Solutions Provider Creates Privatized Test Data Environment with IBM Optim Solutions

- **Application:**
  - Over 200 Applications
  - Mix of Mainframe and LUWs
- **Challenges:**
  - Responsiveness
  - Stale data, Multiple Environments, Excessive Data
  - Inappropriate & Invalid Test Cases
  - Development Teams unable to create test data
  - Industry and corporate data privatization policies
- **Solution:**
  - **IBM Optim Test Data Management Solution**
  - **IBM Optim Data Privacy Solution**
- **Results:**
  - Creates a secure environment for processing customer information
  - Delivers an improved and requested solution to our customers
  - Mitigates the risks of regulatory and legal impact
    - Fines up to $200,000 per incident; $90 per lost account record
  - Reflects adherence to Banks' Information Security policies
  - Provided savings in processing by reducing volume of Test Data – CPU, DASD
  - Allows better management, control of test data cases

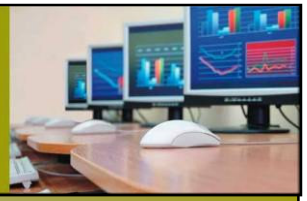
# Data Privacy Protects Your Organization

## Business

**Protect data privacy in both production and non-production environments**

- Reduce risk of exposure during data theft
  - Fines and lawsuits
  - Avoid the negative publicity
  - Customer loss
  - Loss of intellectual property

## IT



**Easily protect data privacy to meet data governance requirements**

- Mask in Place
- Move data and mask when creating test environments
- No changes required to current backup/recovery processes

# How does Data Masking Protect Privacy?

- Comprehensive enterprise data masking provides the fundamental components of test data management and enables organizations **de-identify, mask and transform** to sensitive data across the enterprise

- Companies can apply a range of transformation techniques to substitute customer data with **contextually-accurate but fictionalized data** to produce **accurate test results**

- By masking personally-identifying information, comprehensive enterprise data masking protects the **privacy and security** of confidential customer data, and **supports compliance** with local, state, national, international and industry-based privacy regulations

# Summary

- Data governance - regulations & customer expectations
- Data Privacy & why it is important
  - Non-Production environment is an ideal environment for data exposure
  - Protection of Non-Production Data is as important as Production Data
- Stakes for not protecting Data Privacy are very high
- IBM Optim Data Privacy masking solution makes data hard to use

***We're not going to solve this by making data hard to steal. The way we're going to solve it is by making the data hard to use...***

# Thank You

## InformationOnDemandIndia2011

The Premier Conference for Information Management

**Manage. Analyze. Govern.**

**February 2, 2011**

Hyatt Regency I Mumbai, India