IBM® Optim™

Strategies for Successful Data Governance

Eric Offenberg, CIPP

IBM Software Group

Agenda

- **Understanding Data Governance**
- **Controlling Data Growth**
- **Understanding the Insider Threat to Data**
- **Success Stories**

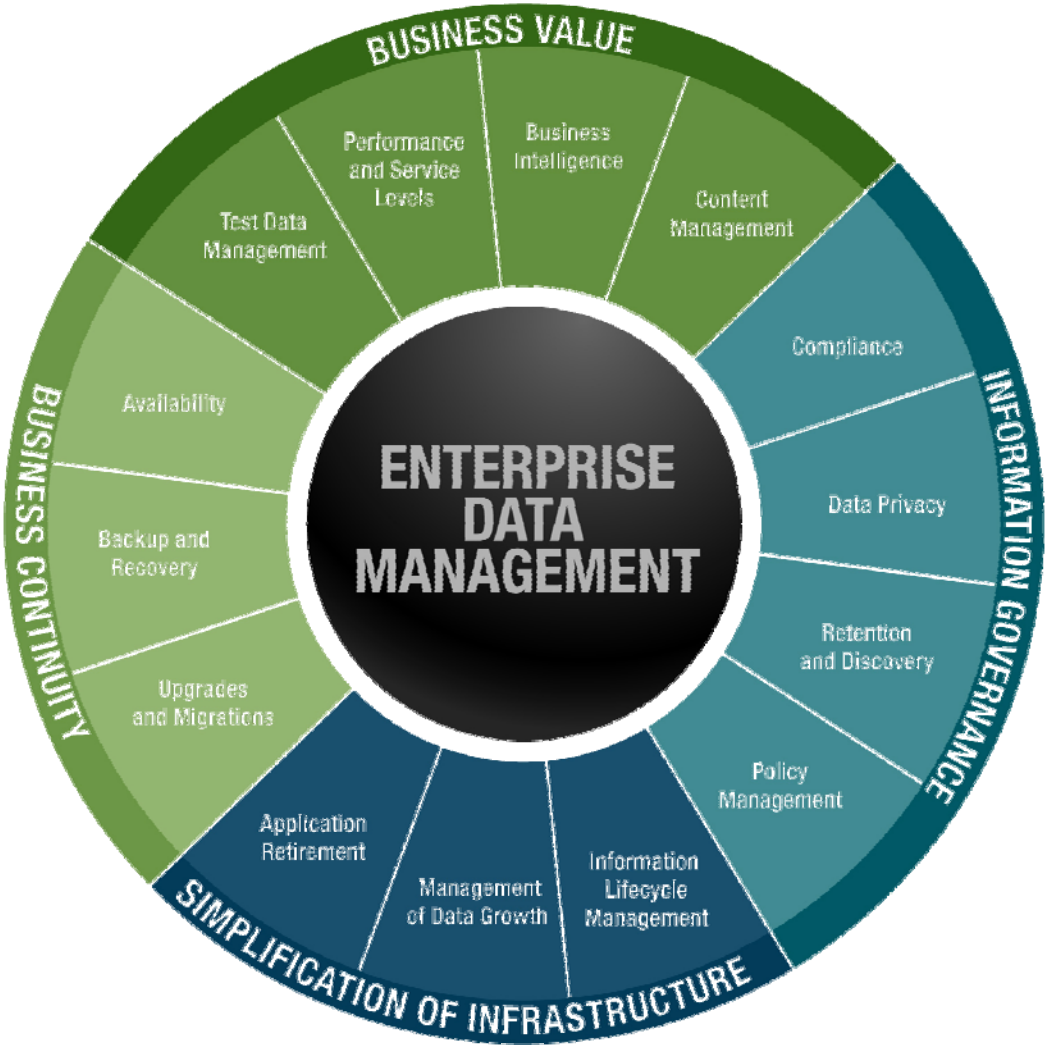
Disclaimers

IBM customers are responsible for ensuring their own compliance with legal requirements. It is the customer's sole responsibility to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws.

IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer is in compliance with any law.

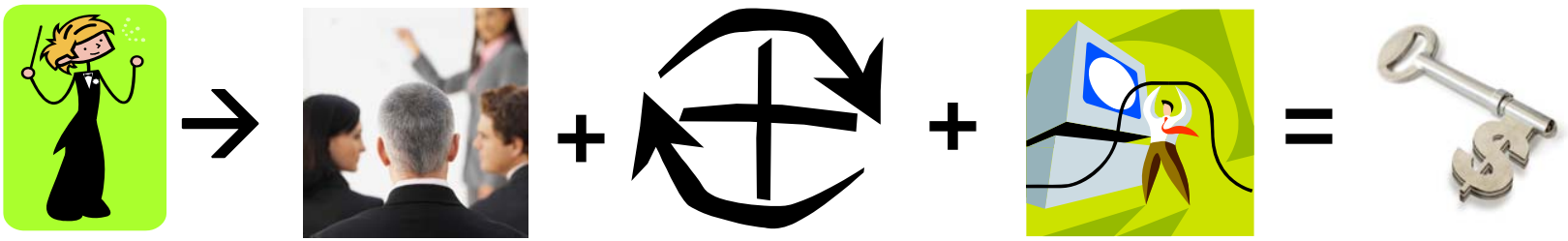
The information contained in this documentation is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information provided, it is provided “as is” without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this documentation or any other documentation. Nothing contained in this documentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM (or its suppliers or licensors), or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

Enterprise Data Management



What is Data Governance? (Strategic View)

Data Governance is the political process of changing organizational behaviour to enhance and protect data as a strategic enterprise asset



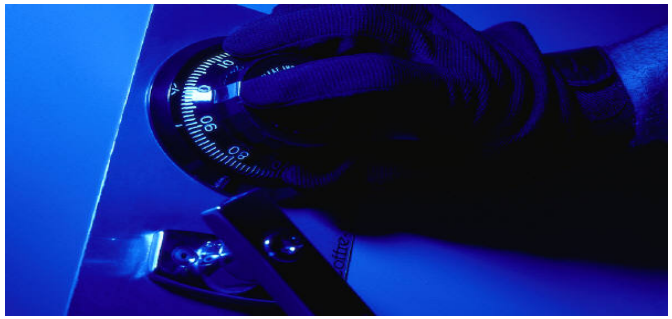
Implementing Data Governance is a fundamental change to the methods & rigor both Business and Information Technology use to define, manage and use of data

The core objectives of a governance program are:

- **Guide information management decision-making**
- **Ensure information is consistently defined and well understood**
- **Increase the use and trust of data as an enterprise asset**
- **Improve consistency of projects across an enterprise**

Why the focus on Data Governance?

- Regulatory Compliance
  - Consumer privacy
  - Financial Integrity
- Intellectual Property Theft
  - Confidential manufacturing processes
  - Financial information
  - Customer lists
  - Digital source code
  - Marketing strategies
  - Research data
- Economic Espionage
  - Trade secret



State sues global management consulting company over stolen backup tape. Unencrypted tape contained personal information on 58 taxpayers and nearly 460 state bank accounts.

Over 45 million credit and debit card numbers stolen from large retailer. Estimated costs $1bn over five years (not including lawsuits). $117m costs in 2Q '07 alone.

Solutions to Help with Data Governance

- **Today, information is the lifeblood of any enterprise**
- **What do you do when you have something valuable?**
  - Retain it
  - Protect it

A Definition of Archiving

Archiving is an intelligent process for ***moving*** inactive or infrequently accessed data that still has ***value***, while providing the ability to ***search and retrieve*** the data.

Why Customers Need Archiving – Drivers

Business

Compliance/Risk
- Driven by SOX, HIPAA, etc. (regulations).
- Records retention requirements.
- Business process compliance.
- Litigation support.

Cost Reduction
- Reduce overall storage costs.
- Minimize associated labor and administration costs.
- Improve disaster recovery processes.

Information Innovation
- Provide access to historical data.
- Mine information for unique value.
- Enhance business for competitive advantage or organizational improvement.

IT

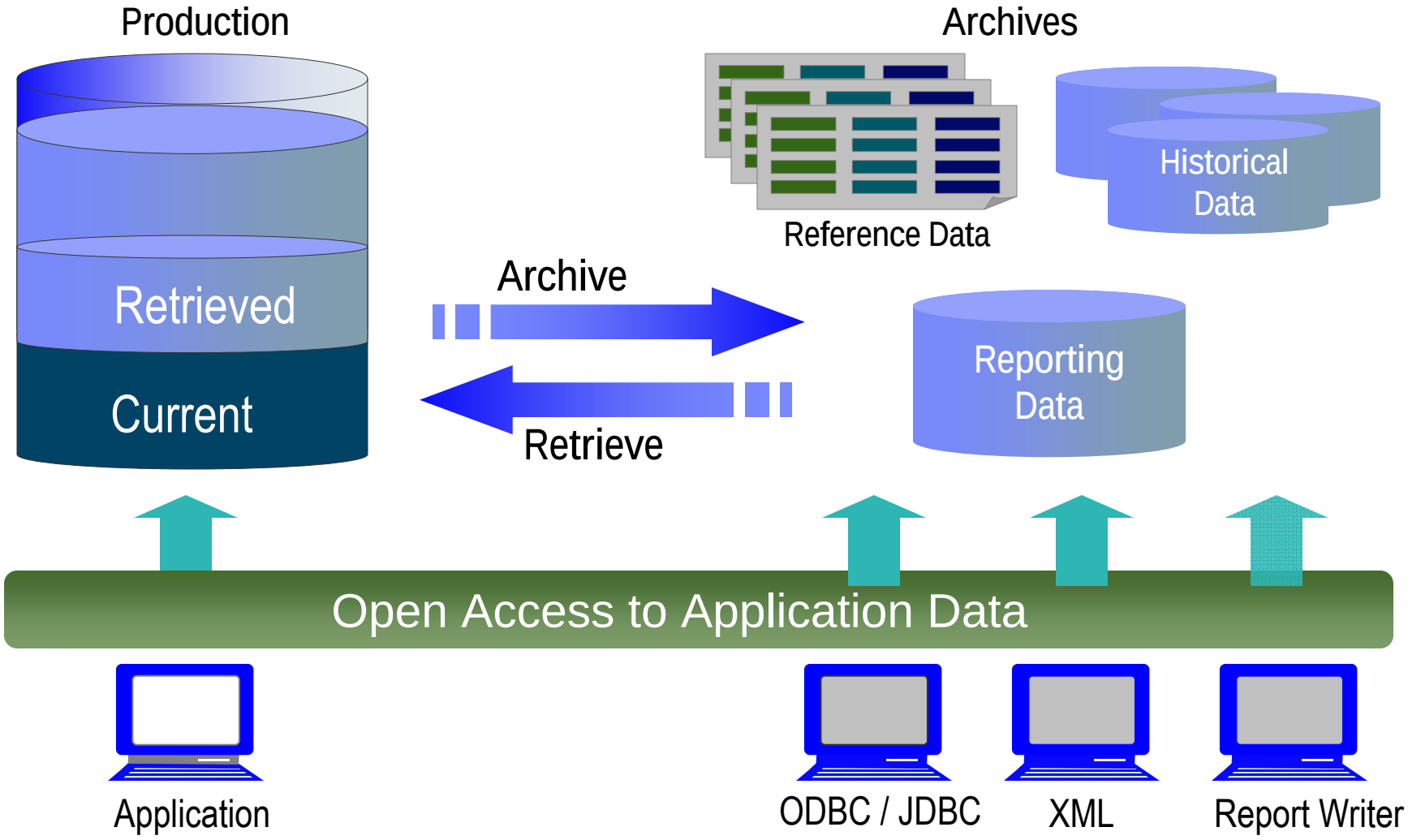
Systems Efficiency
- Reduce high cost storage.
- Reduce backup & recovery resources.

User Productivity
- Remove inactive data to improve application performance.
- Reduce backup & recovery time.
- Improve application availability.
- Easy access to historical/enterprise data.

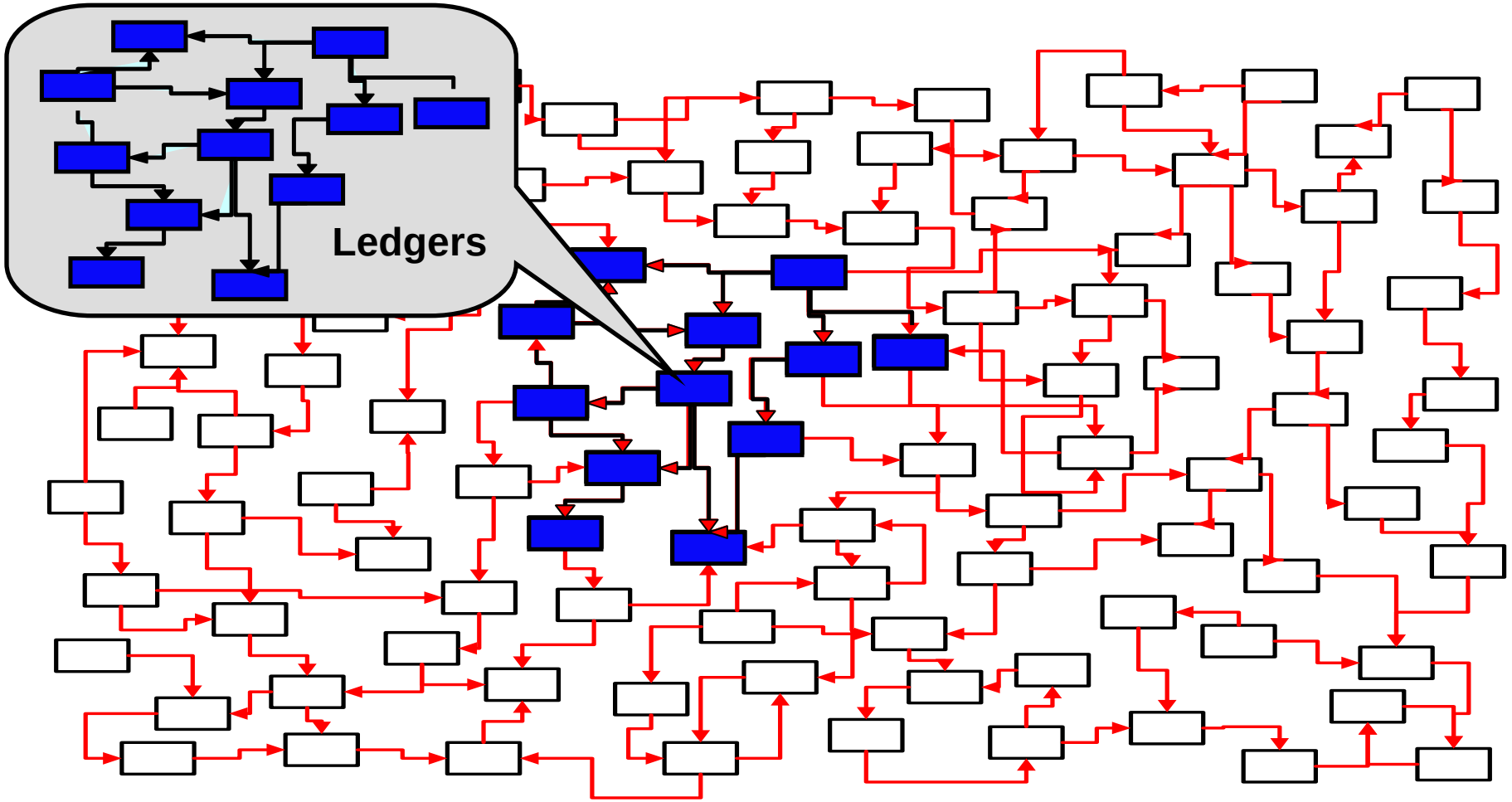
Customers Are Asking Archiving Questions

- What data should I be saving, for how long and for what reasons?
- What data should I be deleting?
- How am I going to find the data when I need it?
- What do I do with the data when I no longer need it?
- What is the most appropriate solution to meet my archiving needs?
- What is the cost/benefit analysis to support an archiving solution acquisition?

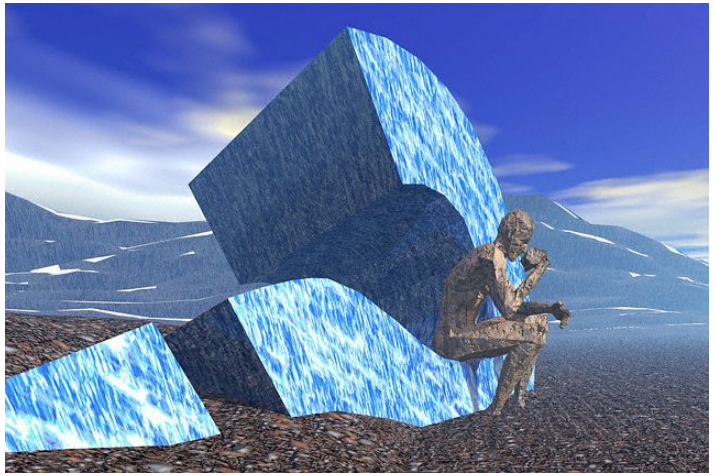
How does Archiving Work?



Archiving a Complete Business Object



What Benefits Exist from Data Archiving?



1) **Improve Performance**
2) **Control Costs**
3) **Mitigate Risks**

How Does Archiving Improve Performance?

- **Improved Availability**
  - No downtime caused by batch process overruns
  - Uptime during crunch time
  - Meet SLAs
- **Speeding Backup and Recovery**
  - Bring up important/recent data first
  - Bring up older/reference data as conditions permit
- **Improved Application Performance**
  - One of the most understated benefits to archiving
  - Longest and most lasting benefit

Let's start with the Analysts

"Moving inactive data to another instance or archive system not only makes production databases more efficient, but it also lowers cost."

"Large databases also drive up hardware cost, database license cost, and general administration effort."

Noel Yuhanna, Forrester Research, **Database Archiving Remains An Important Part Of Enterprise DBMS Strategy, 8/13/07**

"Improved database and application performance, as well as reduce infrastructure cost, can be achieved through database archiving."

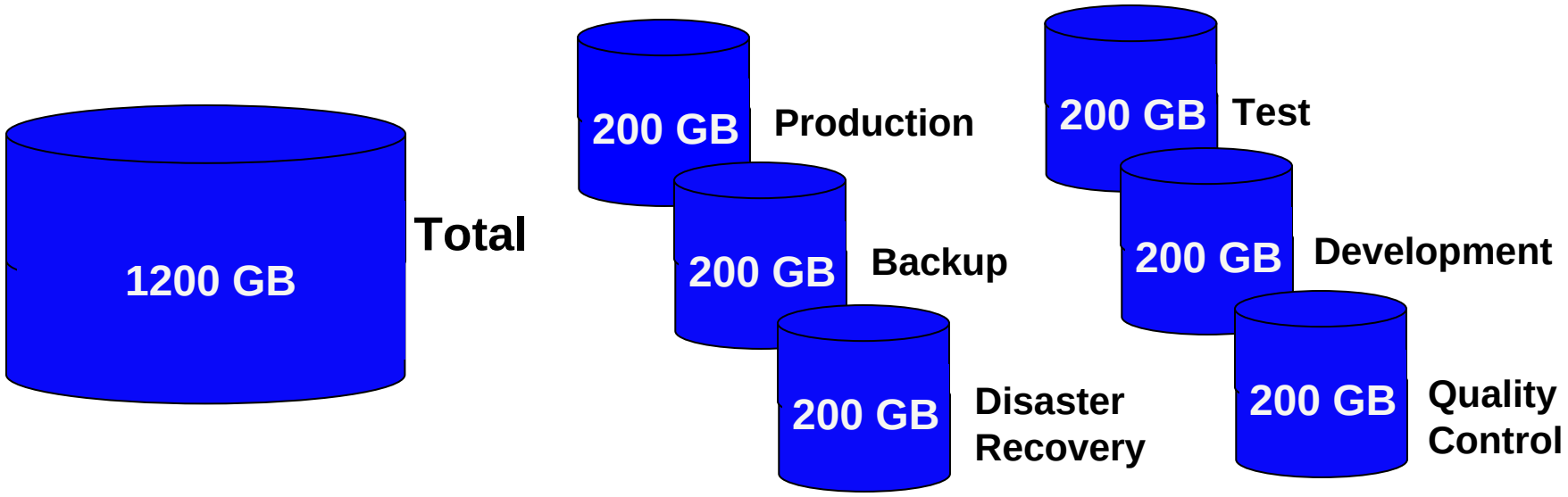
Carolyn Dicenzo and April Adams, Gartner, **Archiving Technology Overview 2/6/07**

How can I save money by Archiving Data?

- **Storage**
  - Production level data is typically one of the most expensive storage platforms
  - Migrate and store data according to its evolving business value (ILM)
  - Use tiered storage strategies to your advantage to maximize cost efficiencies
  - Utilize the storage you already have (including tape!)

Data Multiplier Effect

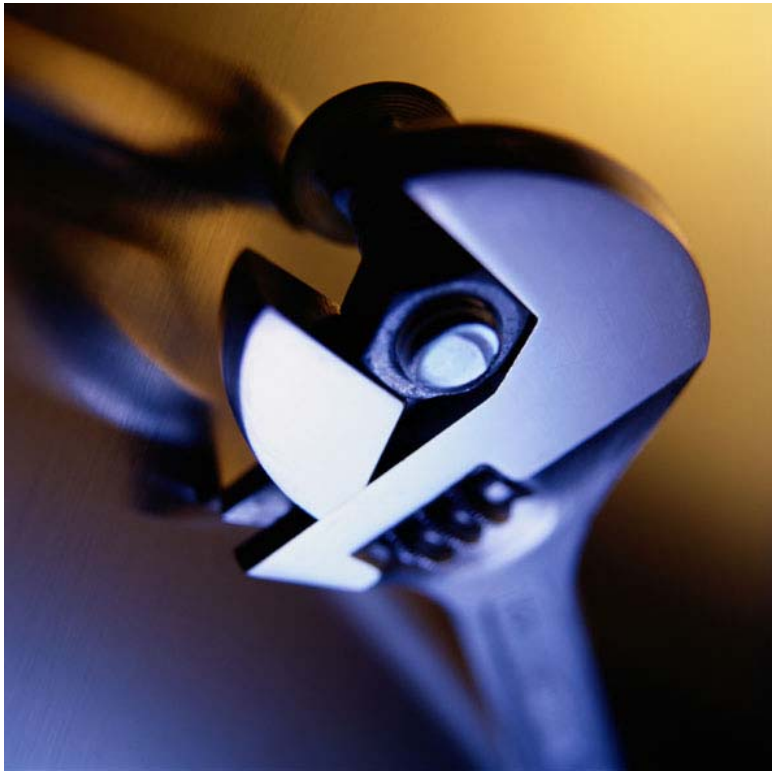
Actual Data Burden = Size of production database + all replicated clones



How can I save money by archiving data?

- **Upgrades and Migrations**
  - Important for packaged applications space (Siebel, Peoplesoft Enterprise, Oracle E-Business, JD Edwards EnterpriseOne, Amdocs, etc.) and for Databases
  - Reduce time allocated for database conversion
  - Reduce downtime during transition
    - One recent client stated 1 hour downtime = $5M
  - Deploy new version quickly
    - Revenue recognition
    - Competitive Advantage

Possible Alternatives to Archiving



- **Tune or partition the database**
- **Add capacity**
  - Processors, storage
- **Back up the database**
- **Purge data**
- **Alleviate symptoms temporarily, but…**
  - ***Inflate costs***
  - ***Do not address underlying data growth***

Creating and Managing Archived Data

1. **Identify the data to archive**
2. **Define the data to delete**
3. **Select Archive File storage**
4. **Create the archive**
5. **Research, report, retrieve**

How can I save money by archiving data?

- **Administrative costs of data management**
  - Software license fees
  - Hardware costs
  - Labor to manage data growth
    - DBA
    - System Admin
    - Storage Admin
- **Reduction in processor upgrades**
  - More MIPS/processors required to process large data repositories
  - Example: **1 TB database that supports 500 concurrent users might require an eight-processor server with 4 GB of memory to achieve optimal performance. The same application that runs a database half that size might require only six processors and 2 GB of memory.**

Noel Yuhanna, Forrester Research, Database Archiving Remains An Important Part Of Enterprise DBMS Strategy, 8/13/07

How Does Data Archiving Mitigate Risk?

- **Data is stored in an immutable format that cannot be altered**
- **Data is indexed following archiving for easy retrieval**
- **Data can be retrieved either from the application it was archived or in various other formats (ex. Excel Spreadsheet, XML, Reporting tools)**

Lawyers ... Ya Gotta Love 'em

Legal Costs of E-Discovery

Identify Appropriate Data	**$200/hour**
Preserve the Data	**$100-$300/hour**
Collect the Data	**$200-$300/hour**
Review the Data	**$120-$350/hour**
Produce the Data	**$1000-$2100/hour**

Debra Logan, "Mapping Technology and Vendors to the Electronic Discovery Reference Model," GartnerResearch, ID Number: G00153110, November 9,2007.

The latest on E-Discovery

- **Electronic discovery (also called e-discovery or ediscovery) refers to any process in which electronic data is sought, located, secured and searched with the intent of using it as evidence in a civil or criminal legal case.**
- **In the process of electronic discovery, data of all types can serve as evidence. This can include text, images, calendar files, databases, spreadsheets, audio files, animation, Web sites and computer programs.**

Example

- **E-Discovery Issues go way beyond just email**
  - Retail organization had contract dispute with partner over provisions in an agreement struck in the late 1990s providing for some collaboration as they expanded into the online world.
  - Sales transaction data became central to the case.
  - Reviewers analyzed details of every sales transaction the retailer completed over a six-year period—more than 250 million in all—to study the sales patterns of different categories of products.
  - Analysis ultimately concluded no violation of agreement. Had the large volume of sales transaction data not be reviewable, the retailer would have been at risk of losing millions of dollars.

Source: FTI Consulting/Forrester Research

Success: Data Retention

About the Client:
Telecommunications, $13 Billion

- Application:
  - Siebel Application
- Challenges:
  - Need for data cleanse and purge records older than 7 years from Siebel databases
  - Preparing for corporate-wide data management effort to sustain goal of keeping only “what’s needed for the right amount of time”
  - Maintain operational efficiencies and reduce cost of maintenance
- Solution:
  - IBM® Optim™ Data Growth Solution for Siebel

- Client Value:
  - Satisfied long-term data retention requirements by archiving for secure and readily accessible information
  - Ensured support for SOX and auditor compliance requirements by implementing archiving capabilities to locate and access historical financials data when needed for audit and discovery requests
  - Established a consistent methodology for managing and retaining historical data using Optim across applications, databases and hardware platforms

Success: Data Growth and Upgrades

About the Client:
Marketing Services, $1.1 Billion Annually

- Application:
  - Oracle E-Business Suite
- Challenges:
  - Managing the 20 to 25% annual data growth rate in Oracle E-Business Suite and managing the expected data growth of 40 to 50% in the next year for the projected upgrade from 10.7 to 11i.
  - Reducing costs for the additional hardware and storage required to support continued data growth
  - Meeting compliance requirements for retaining historical data for 3 to 10 years, while keeping data accessible
  - Reducing the time, effort and downtime associated with upgrading Oracle E-Business Financials
- Solution:
  - Optim Oracle E-Business Suite Solution

- Client Value:
  - Controlled data growth by implementing database archiving for Oracle E-Business Suite
  - Projected a savings of $2million in IT capacity expansion costs over 5 years, and provided the capability to move archived data to a less expensive storage options
  - Supported compliance requirements by providing access to archived data and the capability to report against this data
  - Projected a reduced cutover time to upgrade from Oracle E-Business 10.7 to 11i implementation

Does This Define Your Privacy Strategy?

The Latest on Data Privacy

- **2007 statistics**
  - **$197**
    - Cost to companies per compromised record
  - **$6.3 Million**
    - Average cost per data breach “incident”
  - **40%**
    - % of breaches where the responsibility was with Outsourcers, contractors, consultants and business partners
  - **217 Million**
    - TOTAL number of records containing sensitive personal information involved in security breaches in the U.S. since 2005

* Sources”: Ponemon Institute, Privacy Rights Clearinghouse, 2007

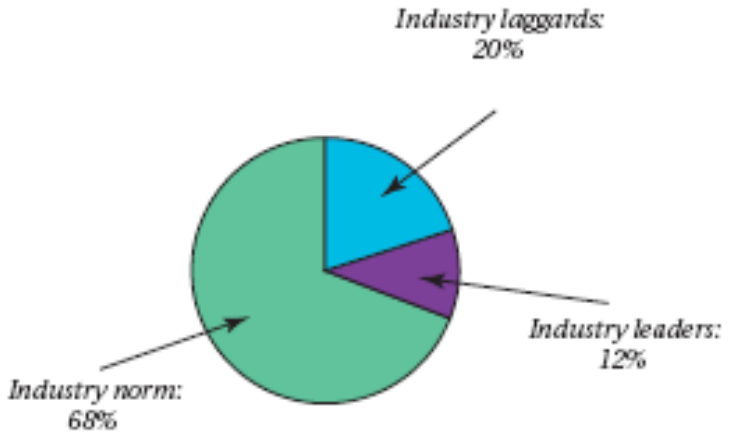
Did You Hear?

- Hannaford Supermarket chain (165 stores in New York and New England) recently confirmed a data intrusion of 4.2 million credit/debit cards
- Included were Sweetbay stores in Florida (106 stores)
- 1800 reported cases of fraud thus far
- This merchant claimed PCI compliance!

Where do F1000 Corporations Stand today?

	Performance classification	Confirmed annual losses of sensitive data
●	Industry laggards	22
■	Industry norm	6
◆	Industry leaders	Less than 2

N: 201



Population experiencing losses

Figure 1: Sensitive data loss results

Source: IT Policy Compliance Group, 2007

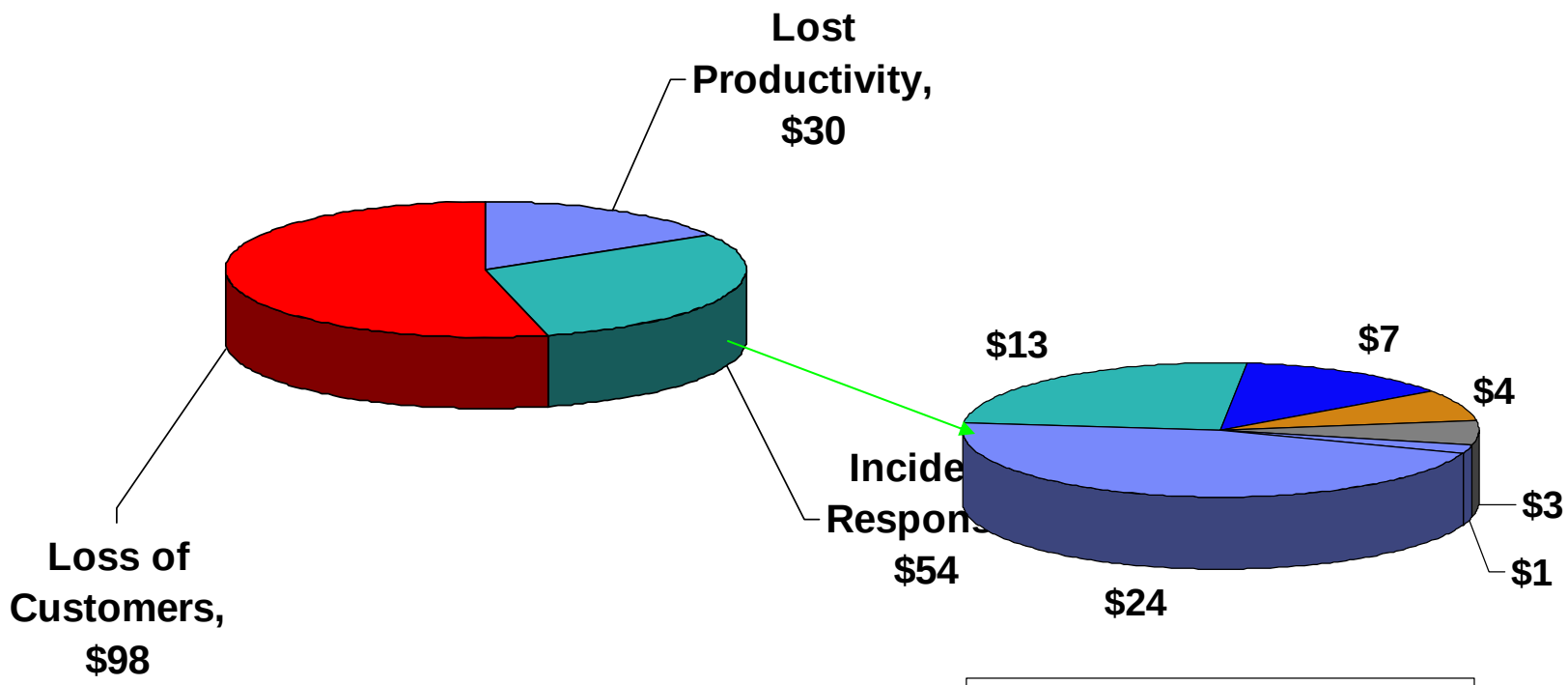
How much is personal data worth?

- **Credit Card Number With PIN - $500**
- **Drivers License - $150**
- **Birth Certificate - $150**
- **Social Security Card - $100**
- **Credit Card Number with Security Code and Expiration Date - $7-$25**
- **Paypal account Log-on and Password - $7**

Representative asking prices found recently on cybercrime forums.

Source: USA TODAY research 10/06

Cost to Company per Missing Record: $197



Over 100 million records lost at a cost of $16 Billion.

Source: Ponemon Institute

What is Done to Protect Data Today?

- **Production "Lockdown"**
  - Physical entry access controls
  - Network, application and database-level security
  - Multi-factor authentication schemes (tokens, biometrics)
- **Unique challenges in Development and Test**
  - Replication of production safeguards not sufficient
  - Need "realistic" data to test accurately

The Easiest Way to Expose Private Data … Internally with the Test Environment

- **70% of data breaches occur internally (Gartner)**
- **Test environments use personally identifiable data**
- **Standard Non-Disclosure Agreements may not deter a disgruntled employee**
- **What about test data stored on laptops?**
- **What about test data sent to outsourced/overseas consultants?**
- **How about Healthcare/Marketing Analysis of data?**
- **Payment Card Data Security Industry Reg. 6.3.4 states,** "Production data (real credit card numbers) cannot be used for testing or development"

*** The Solution is Data De-Identification ***

The Latest Research on Test Data Usage

- **Overall application testing/development**
  - 62% of companies surveyed use actual customer data instead of disguised data to test applications during the development process
  - 50% of respondents have no way of knowing if the data used in testing had been compromised.
- **Outsourcing**
  - 52% of respondents outsourced application testing
  - 49% shared live data!!!
- **Responsibility**
  - 26% of respondents said they did not know who was responsible for securing test data

Source: The Ponemon Institute. The Insecurity of Test Data: The Unseen Crisis

What is Data De-Identification?

- **AKA - data masking, depersonalization, desensitization, obfuscation or data scrubbing**
- **Technology that helps conceal real data**
- **Scrambles data to create new, legible data**
- **Retains the data's properties, such as its width, type, and format**
- **Common data masking algorithms include random, substring, concatenation, date aging**
- **Used in Non-Production environments as a Best Practice to protect sensitive data**

Optim

Masking is transparent to the outside world

Card Holder and Card Number have been masked

Failure Story – A Real Life Insider Threat

- **28 yr. old Software Development Consultant**
- **Employed by a large Insurance Company in Michigan**
- **Needed to pay off Gambling debts**
- **Decided to sell Social Security Numbers and other identity information pilfered from company databases on 110,000 Customers**
- **Attempted to sell data via the Internet**
  - Names/Addresses/SS#s/birth dates
  - 36,000 people for $25,000
- **Flew to Nashville to make the deal with…..**
- **The United States Secret Service (Ooops)**

Results:

- Sentenced to 5 Years in Jail
- Order to pay company $520,000

The Top 3 Reasons Why Insiders Steal Data

1. **Greed**

2. **Revenge**

3. **Love**

Source: US Attorney General's Office, Eastern PA District

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

Strategic Issues for Implementing Data Privacy

Data Masking Considerations

- **Establish a project leader/project group**
- **Determine what you need to mask**
- **Understand Application and Business Requirements**
- **Top Level Masking Components**
- **Project Methodology**

Data Masking Consideration – Step 1

- Establish a Project Leader/Group
  - Many questions to be answered/decisions to be made
  - Project Focus
  - Inter-Departmental Cooperation
  - Use for additional Privacy Projects

Data Masking Consideration – Step 2

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

Data Masking Consideration – Step 3

"My computer doesn't understand me !"

- 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?

Data Masking Consideration – Step 4
Masking Components (Top Level)

- **Masking is not simple!**
  - Many DBMS
  - Legacy Files
  - Multiple platforms
- **Needs to fit within existing processes**
- **Not a point solution – consider the enterprise**
- **Not a one time process**

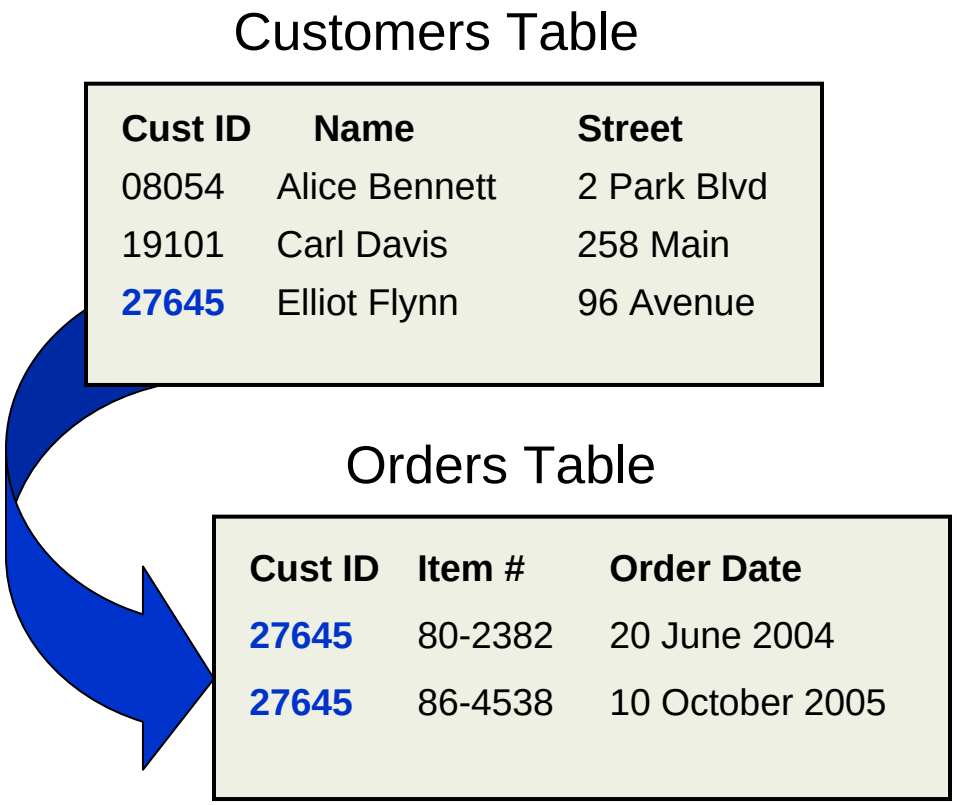
Component A - Consistency

- **Masking is a repeatable process**
- **Subsystems need to match originating**
- **The same mask needs to be applied across the enterprise**
  - Predictable changes
  - Random change will not work
- **Change all ‘Jane’ to ‘Mary’ again and again**

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

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

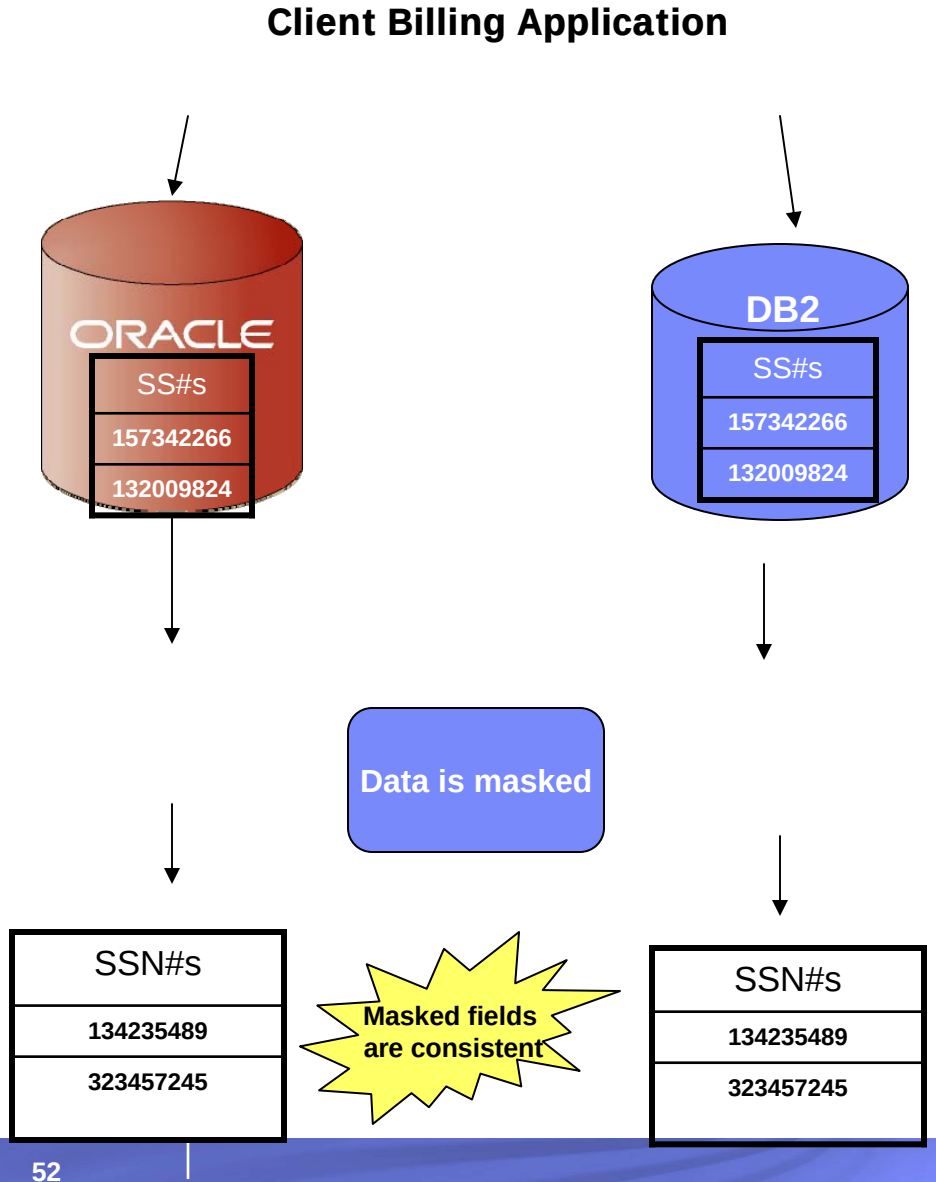
Customers Table

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

Component B - Context



- **A single mask will affect 'downstream' systems**
- **Column/field values must still pass edits**
  - SSN
  - Phone numbers
  - E-mail ID
- **Zip code must match**
  - Address
  - Phone area code
- **Age must match birth date**

Component C - Flexibility

- **Laws being interpreted**
- **New regulations being considered**
- **Change is the only certainty**
- **ERPs being merged**
- **Masking routines will change, frequently**
- **Quick changes will be needed**

Data Masking Consideration – Step 5
Project Methodology

- Determine Base Directives
- Compile Data Sources List
- Design Transformation Strategy
- Develop Transformation Process
- Implement Testing Strategy

How does Data De-Identification Protect Privacy?

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

Concluding Thought #1

“It costs much less to protect sensitive data than it does to replace lost customers and incur damage to the image of the organization and its brand—an irreplaceable asset in most cases.”

IT Compliance Group Benchmark Study 2/07

Concluding Thought #2

“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.”

Bruce Schneier, author of "Beyond Fear: Thinking Sensibly About Security in an Uncertain World"

Success: Data Privacy

About the Client:
UK Government

DWP Department for Work and Pensions

- Application:
  - **Siebel Application (largest in the world)**
- Challenges:
  - Supporting compliance initiatives mandated by the Data Protection Act 1998 to protect privacy in the application development and testing environments.
  - Managing realistic, right-sized development and test databases and preserving the referential integrity of the test data.
  - Employ a ‘best practice’ solution that can be applied across the Department for Work and Pensions four Siebel enterprise
- Solution:
  - **Optim™ Siebel Solution for TDM and Archiving**
- Client Value:
  - Satisfied DWP requirements to de-identify citizen data through ‘masking’
  - Delivered a Seibel solution for ‘live extract’ guaranteeing referential data integrity
  - Commercially ‘ring-fenced’ Pension Transformation Programme (PTP) to open up downstream revenue in 3 further Siebel environments as the ‘defacto’ best practice solution

Success: Data Privacy

About the Client:
$300 Billion Retailer
Largest Company 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 Data Privacy Solution™
- 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

Questions?

- **For more information:**

Eric Offenberg
ERICO@US.IBM.COM

www.OPTIMSOLUTION.COM