Data Governance & Security Information Day

Information Life Cycle Management

Dave Welch—Manager: Optim & Data Governance Center of Excellence
3 September 2014
Agenda

- Understanding the Solution Area
- Trends
- Benefits
- Select Your ILM Partner
## Trends in an Information Driven Economy

*Organizations are treating data as an enterprise asset to grow & compete by leveraging key technology trends*

### Big Data
![Big Data icon](image)

Every day, we create so much data that **90% of the data in the world today has been created in the last 2 years**. Data comes from everywhere: sensors, social media sites, pictures, purchase transactions, cell phone GPS signals, etc.

### Cloud
![Cloud icon](image)

Deliver on-demand computing resources. • Deliver real-time personalized services • Gain intelligence by connecting networks • Become nimble and agile • Capture new markets

### Cyber Security
![Cyber Security icon](image)

The Open Security Foundation reported 1,088 events for 2011 that cover loss, theft, and exposure of personally identifiable information. In 2012, there were 1,502 documented incidents – a rise of nearly 40%.

### Mobile
![Mobile icon](image)

71% of companies interviewed by Forrester in 2013 **invest in mobile development** with even stronger agile development & testing requirements **demanding accelerate test data provisioning**.
## Delivering Value in Strategic Client Initiatives

<table>
<thead>
<tr>
<th>Trends</th>
<th>Pain points</th>
<th>Success Proof Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agile / Mobile / Continuous Development</td>
<td>Faster time to market requires faster testing</td>
<td>Nationwide®</td>
</tr>
<tr>
<td>Outsourcing Testing (&amp; Development)</td>
<td>Protect sensitive test data for outsourced testers &amp; to reduce risk</td>
<td>44% fewer untested cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42% less labor cost</td>
</tr>
<tr>
<td>Big Data</td>
<td>Protect data residing in Hadoop or during “ingestion” (e.g. in ETL jobs)</td>
<td>ALBRIDGE™ AN AFFILIATE OF PERSHING</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Protect 4TB with 60% faster masking</td>
</tr>
<tr>
<td>Cloud</td>
<td>Address privacy concerns that are the primary inhibitor for cloud adoption</td>
<td></td>
</tr>
<tr>
<td>Mergers &amp; Acquisitions / Consolidation</td>
<td>Retire legacy applications that are replaced / consolidated</td>
<td>Citi saved</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$275k/month per retired app</td>
</tr>
<tr>
<td>Operational Efficiency</td>
<td>Free up people + process + hardware to focus on new strategic initiatives/workloads</td>
<td></td>
</tr>
</tbody>
</table>

© 2013 IBM Corporation
Most approaches to data security and compliance miss the mark, and doing nothing is not optional

$5.5M average cost per data breach in 2011

$3M cost of losing customer loyalty (lost business) following a data breach

Source: The True Cost of Compliance, The Cost of a Data Breach, Ponemon Institute, 2011

$3.5M Yearly average cost of compliance

Source: Aberdeen Group. Why Information Governance Must be Addressed Right Now, 2012

<table>
<thead>
<tr>
<th>Company Data Security approach</th>
<th>Audit events/year</th>
<th>Average cost/audit</th>
<th>Data loss events/year</th>
<th>Average cost/data loss</th>
<th>Total cost (adjusted per TB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o data security</td>
<td>6.3</td>
<td>$24K</td>
<td>2.3</td>
<td>$130K</td>
<td>$449K/ TB</td>
</tr>
<tr>
<td>w/ data security</td>
<td>1.7</td>
<td></td>
<td>1.4</td>
<td></td>
<td>$223K/ TB</td>
</tr>
<tr>
<td>Annual Cost of not implementing data security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$226K/ TB</td>
</tr>
<tr>
<td>Total annual cost of doing nothing:</td>
<td>(for average Big Data organization with 180 TB of business data)</td>
<td></td>
<td></td>
<td></td>
<td>$40+ M</td>
</tr>
</tbody>
</table>
IBM Solves Key Data Challenges

Identify Relevant & Sensitive Data
Find what data must be retained, protected or removed

Optimize Test Data
Automate and optimize the application test processes that rely on data to enable continuous testing & DevOps

Dispose of Unnecessary Data
Remove unnecessary data from critical transactional or analytics applications

Retain Essential Data
Historical inactive data is safely retained while easily accessible for reports and compliance

Protect Sensitive Data
Private Data: Customer IDs, credit cards and financial data are masked or redacted

InfoSphere Optim Archive Enterprise Ed.
Requirements for managing data across its lifecycle

Information Governance Core Disciplines

Lifecycle Management

Discover & Define
- Discover where data resides
- Classify & define data and relationships
- Define policies

Develop & Test
- Develop & test database structures/code
- Create & refresh test data
- Capture & replay production workloads

Optimize & Archive
- Enhance performance
- Manage data growth
- Report & retrieve archived data

Consolidate & Retire
- Integrate into single data source
- Move only the needed information
- Enable compliance with retention & e-discovery
Automates discovery to help you understand your data

You can’t govern what you don’t understand

• Acceleration of Test Data Management projects
  • Identification of Relationships
  • Identification of Business objects
  • Identification of Sensitive data

• Answers the Questions:
  • How is the data related?
  • What are the related business objects?
  • Where is my sensitive data?

• Why is this Different?
  • Analysis done by actually reading the data
  • Automates the discovery of business entities cross-data-store business rules & sensitive data
  • Can analyze multiple data sources simultaneously
Effective Test Data Management
Subsets data while preserving referential integrity

Producing complete business objects that contain the data and business relationships to accurately test and re-test applications with integrity

Retrieval of referentially-intact subsets of data across related applications, including the metadata.

Customer Management in DB2 LUW
Shipping and Logistics in Oracle
Sales Management in DB2 LUW

Federated access to related business objects across the enterprise
Employ effective test data management practices

Production or Production Clone

- Create targeted, right-sized test environments
- Substitute sensitive data with fictionalized yet contextually accurate data
- Easily refresh, reset and maintain test environments
- Compare data to pinpoint and resolve application defects faster
- Accelerate release schedules
Effective Test Data Management
Consistently masks data in non-production environments

- Age
- Name
- Address
- Birth Date
- Phone Number
- Email Address
- Account Numbers
- National Identifiers
- Payment Card Numbers
- And many more...

- Mask data in non-production environments such as development, QA and training...
- Substitute sensitive data with fictionalized but contextually accurate data, preserving integrity
- Ensures that masked data resembles the structure and characteristics of the original data
- Preserves referential integrity, propagates masked key values consistently across relationships
- Provides out of the box, pre-packaged postal service Lookup data for different countries for realistic replacement of Names, Addresses, National Identifiers, Credit Card Numbers etc...
- Mask custom and packaged ERP/CRM applications
Understanding Dormant & Related Data for Archiving

- **Forrester: “Only 12% of the data is utilized”** (December 2013)
  - How do you know which data is active and which is **dormant**
- Most of the information is duplicated and spread across multiple systems
  - How do you detect and archive **business objects** across databases

![Diagram of active and dormant data]

*Report of dormant data showing which data fields have not been accessed in a time period (powered by Guardium technology)*

*Capture business objects of related data within and across databases in order to archive complete & consistent data*
Optimize Operational Efficiency & Manage Data Growth

**Use cases**
- Application retirement (e.g. consolidating/migrating applications)
- Reducing cost of high data volumes in transactional systems
- Retaining data over extended period for compliance

**Optim**
- Policy-based retention
- Business object oriented
- Audit capabilities
- Dormant data analysis

**Universal Access**
- Applications
- ODBC/JDBC
- Data Explorer
- Report Tools
- XML
OPTIM ILM Archive Strategy – One Example

Current Data
1-2 years
Production Database

Active Historical
3-4 years
Archive Reporting Database
Compressed Archives

Online Archive
5-6 years
Non DBMS Retention Platform
ATA File Server
EMC Centera
IBM RS550
HDS
Compressed Archives

Offline Archive
7+ years
Offline Retention Platform
CD Tape
Optical
Compressed Archives

Tier 3 Infrastructure
Tier 1 Infrastructure
Tier 2 Infrastructure
Tier 4 Infrastructure
Homogeneous / Interoperable / Heterogeneous Support

- **Homogeneous**
  - Production Source (e.g. DB2) → Archive → Restored (e.g. DB2)

- **Interoperable**
  - Production Source (e.g. ORCL) → Archive → Restored (e.g. DB2)

- **Heterogeneous & federated**
  - Source (e.g. ORCL) → Archive → Source (e.g. DB2) → n:m → Source (e.g. ORCL)
  - Source (e.g. DB2) → Archive → Source (e.g. SQL Srv)

**Enterprise Environments**
- Custom, Siebel, Oracle, PeopleSoft, JD Edwards, Amdocs, SAP, Maximo, Partner-delivered Solutions
- Oracle, SQL Server, Sybase, Informix, DB2, IMS, Teradata, VSAM, Adabas, Netezza, More...

**Operating Systems & Storage**
- Windows, Solaris, HP/UX, Linux, AIX, z/OS, IBM i, NAS, SAN, ATA, CAS, Optical, Tape
Efficient Data Retirement

Preserve application data in its business context
  • Capture all related data, including transaction details, reference data & associated metadata
  • Capture any related reference data that may reside in other application databases

Retire out-of-date packaged applications as well as legacy custom applications
  • Leverage out-of-box support of packaged applications to quickly identify & extract the complete business object

Shut down legacy system without a replacement
  • Provide fast and easy retrieval of data for research and reporting, as well as audits and e-discovery requests
# Requirements for managing data across its lifecycle

<table>
<thead>
<tr>
<th>Information Governance Core Disciplines</th>
<th>Lifecycle Management</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discover &amp; Define</strong></td>
<td><strong>Develop &amp; Test</strong></td>
</tr>
<tr>
<td>Discover where data resides</td>
<td>Develop &amp; test database structures/code</td>
</tr>
<tr>
<td>Classify &amp; define data and relationships</td>
<td>Create &amp; refresh test data</td>
</tr>
<tr>
<td>Define policies</td>
<td>Capture &amp; replay production workloads</td>
</tr>
<tr>
<td></td>
<td>Report &amp; retrieve archived data</td>
</tr>
<tr>
<td></td>
<td>Enhance performance</td>
</tr>
<tr>
<td></td>
<td>Manage data growth</td>
</tr>
<tr>
<td></td>
<td>Integrate into single data source</td>
</tr>
<tr>
<td></td>
<td>Move only the needed information</td>
</tr>
<tr>
<td></td>
<td>Enable compliance with retention &amp; e-discovery</td>
</tr>
</tbody>
</table>
Designing Masking and Redaction Solutions

1. **Extract production data for test and development**
   - Mask sensitive data from production database when copied to test and dev

2. **Masking “Providers” inside the database**
   - Build masking into the database

3. **Direct database connections (Privilege Users)**
   - Redact or block unauthorized access

4. **Build masking into custom applications**
   - Build masking into the application for unauthorized users

5. **Applications that request “clear text” data from DB**
   - Redact sensitive data from the database will break the application
   - If you can’t modify the application, then redact in transit from application server to the browser
A solutions for each step:

InfoSphere solutions protect structured and unstructured sensitive data

**Discover**
- **Guardium Base**
  - Discovery & Classification
  - Queries & Reports
  - Compliance Workflow
  - Group Management
  - Integrations
  - Incident Management
  - Self Monitoring
- **Guardium VA**
  - Assessment reports
  - Subscription
  - Configuration Changes
  - Entitlement Reporting

**Harden**
- **Optim Data Privacy**
  - Mask sensitive data in test, publishing in databases and Big Data environments

**Mask**
- **Guardium Data Encryption**
  - File-level encryption
  - Policy-based Access control
- **Guardium Data Redaction**
  - Redact sensitive documents

**Monitor**
- **Guardium DAM**
  - Activity Monitoring
  - Real-time alerts
  - Compliance Reporting
  - Federate large deployment
  - Central control
  - Central audit collection

**Block**
- **Guardium VA**
  - Activity Monitoring
  - Dynamic Masking
  - Users Quarantine

---

**Data Privacy for Hadoop**
- Packaged discovery, masking, and monitoring for Hadoop
A solutions for each step: InfoSphere solutions protect structured and unstructured sensitive data

**Base functionality**
- Discovery & Classification
- Queries & Reports
- Compliance Workflow
- Group Management
- Integrations
- Incident Management
- Self Monitoring

**VA**
- Assessment reports
- Subscription
- Configuration Changes
- Entitlement Reporting

**Data Encryption**
- File-level encryption
- Policy-based Access control

**Data Privacy**
- Mask sensitive data in test, publishing in databases and Big Data environments

**Database Activity Monitoring**
- Activity Monitoring
- Real-time alerts
- Compliance Reporting
- Federate large deployment
- Central control
- Central audit collection

**Data Redaction**
- Redact sensitive documents

**Data Privacy for Hadoop**
- Packaged discovery, masking, and monitoring for Hadoop

**Commands & Solutions**
- **Check**
  - Base functionality
  - VA
  - Data Encryption
  - Data Privacy
  - Database Activity Monitoring
  - Data Redaction

**IBM**
IBM Optim Data Privacy Differentiators

- **Deepest understanding sensitive data**
  - Uniquely discover automatically new databases that may contain sensitive data (bundled Guardium technology)
  - Uniquely align to business terms and policies (bundled Business Glossary / Business Information Exchange)

- **Uniquely mask any data in any method in any place**
  - Uniquely support unstructured data (redact, mask)
  - Uniquely enable masking in the database and through APIs
  - Uniquely provide semantic masking

- IBM is **THE leader in Gartner’s Magic Quadrant** since first publication (2 years in a row)
InfoSphere data security solutions value proposition

Protect, monitor, and audit access to sensitive data in databases, data warehouses, Big Data environments and file systems to:

1. Prevent data breaches
   - Mitigate external and internal threats & prevent disclosure or leakages of sensitive/privacy data (Guardium)
   - Enable secure use of privacy data

2. Ensure the integrity of sensitive data
   - Provide realistic but not real data for testing, analytics and “monetizing data” (i.e. selling data)

3. Reduce cost of compliance
   - Automate and centralize controls
   - Simplify audit review processes

4. Meet scalability and TCO requirements for enterprise-wide deployments
   - Increase operational efficiency
   - No degradation of infrastructure or business processes
Value Proposition & Compelling Reason to act

- **Demonstrate Compliance**
  - Regulatory retention requirements / federal mandates, e.g. HIPAA 5010,
  - Only 12% of data is utilized (Forrester, Dec 2013)

- **Reduce Cost**
  - $277k per month per app
  - $11.6M over 3 years
  - $450k over 3 years & >50% DBA cost
  - 80% less storage costs
  - 40-50% in total costs
  - 50% in DB maintenance

- **Improve Performance**
  - 66% faster access to customer record
  - 90 min faster batch processing
  - Reports ready in 10sec vs. weeks
  - 200% better performance
### Addressing Trends and Client Initiatives

<table>
<thead>
<tr>
<th>Agile / Mobile / Continuous Dev.</th>
<th>“Getting test data takes too long”</th>
<th>Test Data on Demand (in 2 clicks)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“My DB upgrades take too long”</td>
<td>Optim Workload Replay</td>
</tr>
<tr>
<td>Outsourcing Testing (&amp; Dev.)</td>
<td>“Testers must not access sensitive data”</td>
<td>Embedded Optim Data Masking</td>
</tr>
<tr>
<td></td>
<td>“Costs for testing are too high”</td>
<td>Automation and sub-setting</td>
</tr>
<tr>
<td>Big Data</td>
<td>“How do we preserve analytical patterns and protect sensitive data”</td>
<td>Optim Data Masking</td>
</tr>
<tr>
<td>Cloud</td>
<td>“How do I address privacy concerns for data on the cloud”</td>
<td>Optim Data Masking</td>
</tr>
<tr>
<td>M &amp; A / Consolidation</td>
<td>“How do I reduce cost of legacy app”</td>
<td>Application de-commissioning</td>
</tr>
<tr>
<td></td>
<td>“Legal: still need access to retired apps”</td>
<td>Archive/retain &amp; restore data</td>
</tr>
<tr>
<td>Operational Efficiency</td>
<td>“High data volume, unclear need”</td>
<td>Dormant data analysis &amp; archiving</td>
</tr>
<tr>
<td></td>
<td>“Data volumes impact perform. &amp; cost”</td>
<td>Policy-based archiving</td>
</tr>
</tbody>
</table>
Compelling Reason to Act

- **Nationwide**: reduced untested scenarios by 44% → **higher software quality**
- **Travelport**: Improved test processes → **90% reduced test data storage cost**
- **JN Data**: shortening testing time by 75% → **4x more frequent releases**
- **Boeing**: protecting sensitive (outsourced) data → **reduced risk**
- **Albridge**: Optim vs manual masking → **60% higher scalability**
- **<private>**: Responding to breach → **damage control**
- **Dignity Health**: Federal mandates (HIPAA 5010, HER, …) → **demonstrate compliance**
- **Citi**: Exploding data volumes → **$275k per month per app reduced cost**
- **Progressive**: Growth of key tables exceeded 20% per year → **improve performance**
The Market & Analysts Confirm: Optim is THE Leader

Healthcare providers: 5/5 top  
Pharmaceutical companies: 5/6 top  
Telecomm. companies: 2/3 top  
Banking institutions: 4/5 top  
Retailers: 2/3 top  
Insurance companies: 3/3 top

Vendor Market Share by Revenue

- IBM: 76%
- Informatica: 16.2%
- Other: 1.7%
- HP: 2.6%
- Solix: 3.5%

Gartner MQ 2012 (& 2013): Data Masking Technology

- IBM
- Oracle
- Informatica
- Compudata
- Camouflage Software
- Axis Technology
- Deltagise
- Solix Technologies
- Voltage Security
- Privacy Analytics
- Grid-Tools
- Mentis

Gartner MQ 2013: Integrated Software Quality Suites

- IBM
- HP
- Microsoft
- Oracle
- Red Hat
- Soasta
- Borland
- CA Technologies
- Parasoft
- SmartSoft
- Automation Anywhere
- TriciNet
- Tricentis
- Original Software
- Worksoft
- SmartBear
- Talend
Main Business Drivers

- **Security**: prevent data breaches
- **Privacy**: protect PII
- **Compliance**: pass the audit

Confidence

ANY DATA
The Need for Continuous Deployment & Testing due to Cloud, Mobile & DevOps Trends

FASTER TIME TO MARKET
to deploy app & system releases

Higher degree of AGILITY
to respond to app & system changes

© 2013 IBM Corporation
Effective Data Growth Management

Data 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.