



Extending Data Privacy to Unstructured Data

Data Security vs. Data Privacy



DATA SECURITY – IS ALL OF THE PRACTICES AND PROCESSES THAT ARE IN PLACE TO ENSURE DATA ISN'T BEING USED OR ACCESSED BY UNAUTHORIZED INDIVIDUALS OR PARTIES.



DATA PRIVACY - IS THE NECESSITY TO PRESERVE AND PROTECT ANY PERSONAL INFORMATION, COLLECTED BY ANY ORGANIZATION, FROM BEING ACCESSED BY A THIRD PARTY.



MITIGATE RISK & MAINTAIN COMPLIANCE

Mitigate Risk

Largest breaches of 2018



2 Million Records



52.5 Million Records



CATHAY PACIFIC

9.4 Million Records



100 Million
Records

facebook

29 Million Records



383 Million Records

Comparative Growth of Data Breaches by Industry

| DATA BREACH ANNUAL COMPARISON (2018 vs. 2017) | | | | |
|---|---------------|----------------------|---------------|----------------------|
| Industry | 2018 | | 2017 | |
| | # of Breaches | # of Records Exposed | # of Breaches | # of Records Exposed |
| Banking/Credit/Financial | 135 | 1,709,013 | 134 | 3,230,308 |
| Business | 571 | 415,233,143 | 907 | 181,630,520 |
| Education | 76 | 1,408,670 | 128 | 1,418,455 |
| Government/Military | 99 | 18,236,710 | 79 | 6,030,619 |
| Medical/Healthcare | 363 | 9,927,798 | 384 | 5,302,846 |
| Annual Totals | 1,244 | 446,515,334 | 1,632 | 197,612,748 |

Cost of a Data Breach 2018

Average total cost of
a data breach In US
\$7.1 million

Likelihood of a recurring material
breach over the next two years:
27.9%

Average total one-year
cost increase:
6.4%

Average cost per lost or
stolen record:
\$148

Costs are categorized as Direct, Indirect, and Opportunity as defined below:

- **Direct cost** — the direct expense outlay to accomplish a given activity.
- **Indirect cost** — the amount of time, effort and other organizational resources allocated to data breach resolution, but not as a direct cash outlay.
- **Opportunity cost** — the cost resulting from lost business opportunities as a consequence of negative reputation after the breach is reported to victims (and publicly revealed to the media).

Maintain Compliance

Regulations Driving Data Privacy



Structured vs. Unstructured Data

- **Structured** - Information with a high degree of organization, such that inclusion in a relational database is seamless and readily searchable by simple search operations.



DB2 . Oracle . SQL Server . Informix . Sybase .
Teradata ...

- **Unstructured** - information that either does not have a pre-defined data model or is not organized in a pre-defined manner.



Scanned Images . PDF . Web logs . Email . Office
docs ...

Unstructured Data Management



The **worldwide shift from structured to unstructured data is well known**. It isn't that structured data is going away. In fact, structured data continues to grow. It's just that **growth of unstructured datasets is faster – much faster**.



When you consider the analyst predictions that **90 percent of new data will be unstructured**, it becomes clear that we need to understand and prepare for how this affects us.



Unstructured datasets are growing quickly. The **typical organization reports its unstructured data growing 23% annually**, which means it will **double every 40 months**. Roughly **one fourth cite growth rates in excess of 40%**, where **total unstructured data doubles every 24 months**.



82% of (200) respondents manage **1 billion+ files and objects** and **52%** of (200) respondents manage more than **10 billion files**.



40 percent of their organization's value comes directly from their data. ***“Our data is much more valuable than we've ever thought...”That's why we save so much of it!”***



A Hacker's Dream and a Compliance Nightmare!

Unstructured Data Privacy Use Cases



TDM – COPYING DATA FROM PRODUCTION AND MOVING TO DEV/TEST ENVIRONMENTS. STRUCTURED DATA MASKING DOES NOT EXTEND TO UNSTRUCTURED DATA SO LEFT EXPOSED



ANALYTICS – SENDING (SCANNED) DOCUMENTS TO 3RD PARTY VENDORS FOR ANALYTICS OR UNDERWRITING OR WEB/MARKETING ANALYTICS



DYNAMIC DATA MASKING – MASK PRODUCTION DATA “ON THE FLY” BASED ON CREDENTIALS

Test Data Management: Path of Least Resistance



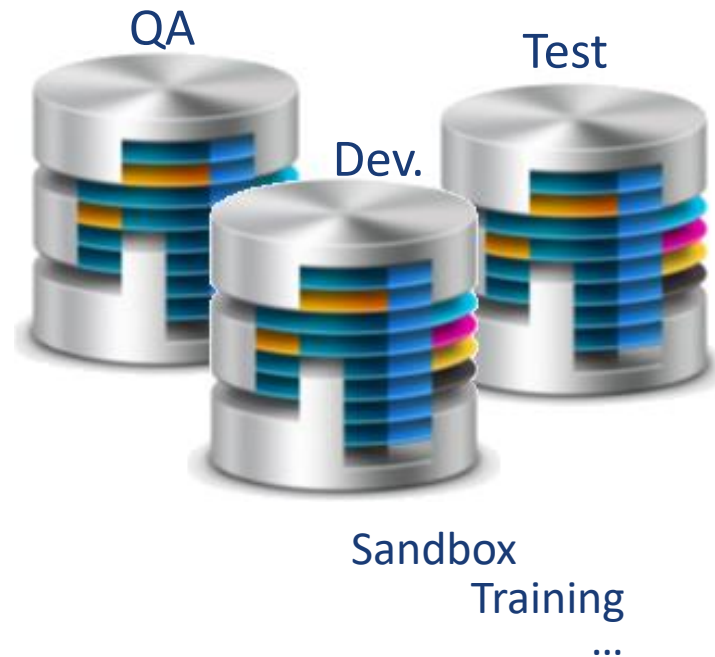
Production



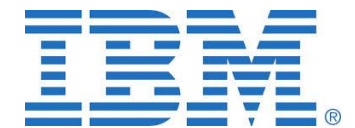
- Secured
- Locked down
- Access Controlled
- Stable / Static



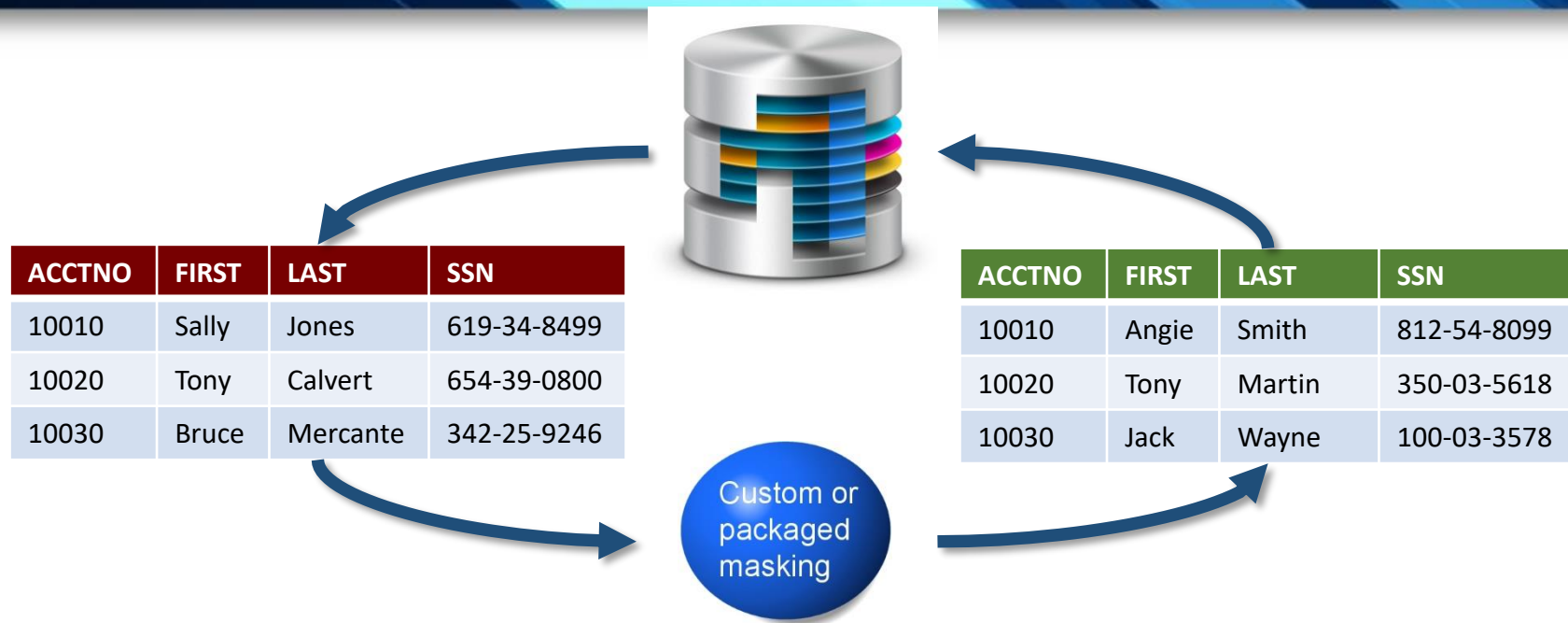
Test



- Exposed
- Outsource Development/IT
- Greater access/permissions
- Stale copies..



Structured



protected

protected?

unstructured



UMask

Unstructured Data Protection

Unstructured Data Masking

- Mask over 65 different file types including Images, PDF, CSV, Email, Images, Office docs ...
- Maintain referential integrity ensuring data is masked consistently both within the database and within the files
- Mask and replace data with like-for-like fictitious data : SSN with fictitious valid SSN, credit card number with fictitious valid credit card number
- Run standalone or in conjunction with other packaged masking and discovery solutions

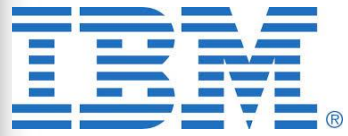
Structured

| ACCTNO | FIRST | LAST | SSN |
|--------|-------|----------|-------------|
| 10010 | Sally | Jones | 619-34-8499 |
| 10020 | Tony | Calvert | 654-39-0800 |
| 10030 | Bruce | Mercante | 342-25-9246 |

| ACCTNO | FIRST | LAST | SSN |
|--------|-------|--------|-------------|
| 10010 | Angie | Smith | 812-54-8099 |
| 10020 | Tony | Martin | 350-03-5618 |
| 10030 | Jack | Wayne | 100-03-3578 |

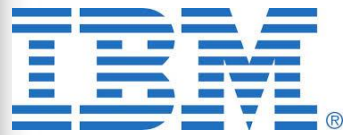


& other IBM governance tools ...

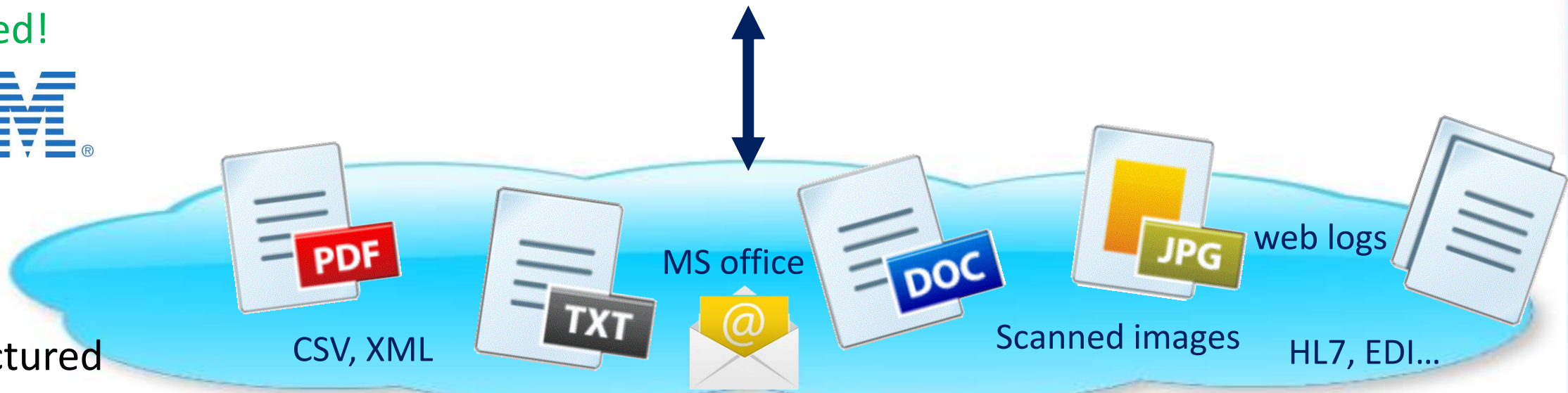


protected

protected!



unstructured



Challenges of Masking Unstructured Data



Images stored as pixels – JPEG, TIFF, PNG, GIF, RAW...
Insurance claims, DICOM/MRI, Bank checks, X937...



Adobe proprietary format, compressed, embedded images, freeform text



MS Office Proprietary format , XLSX, XLS, XLSB, XLSM
Excel 2007, 2013, ...



Increasing the Value of Unstructured Data Discovery

Discovery Solution

- Delete
- Quarantine
- Move



- Delete, Quarantine, Move
- Mask
- Redact

| Name | Path | vol_name | objtype | match Type | Value | Location | Action | Umask Scanned |
|-----------------|-------------|----------|---------|------------|---------------------|-------------|--------|--------------------|
| 2018 Class.docx | C:\SIQ\Docs | removed | Word | SSN | (XXXXXX)-3800 | Paragraph 2 | R | 9\10\2018 12:28:32 |
| Weblog.txt | C:\SIQ\Docs | removed | TXT | Visa | (XXXXXXXXXXXX) 3407 | Line 127 | R | 9\10\2018 12:28:33 |

Standard Discovery report

Enhanced Discovery report

- ✓ Match type: SSN, Credit Card...
- ✓ Sample Value: return part of found value
- ✓ Location: PDF paragraph 2, Excel Sheet 4, line 23...
- ✓ Action: Redact, Ignore, Delete, Browse file
- ✓ Date Scanned

Functions to Mask Unstructured Data

Direct: Find and Replace

Bruce Mercante | Wayne Andrews

To: **Lt Bruce Mercante** from the U.S...

To: **Lt Wayne Andrews** from the U.S...

RegEx: Pattern matching

`^\d{3}-\d{2}-\d{4}$` | 123-45-6789

...Social Security Number: **619-44-3800**...

...Social Security Number: **123-45-6789**...

Region: Area or Region

`<from:>` | `</from:>` | 'REDACTED'

`<from>` **Angela Mace** `</from>`...

`<from>` **REDACTED** `</from>`...

Optim: Consistent

`\d{3}-\d{2}-\d{4}$` | ODPD:NID

Social Security Number: **619-44-3800**...

Social Security Number: **619-21-9217**...

Array: Consistent

Angela | ARRAY:FIRSTNM

Angela Mace...

Brianna Mace...
