What Are You Going To Do With All That Sensitive Data?

(do you even know where it all is?)

Barry Schrager
Xbridge Systems, Inc.
BSchrager@xbridgesystems.com
(650) 319-7345

DataSniff User Session

- David Wade, CIO of Primerica
- 12943: How to Identify Datasets Containing PCI, PII or Other Sensitive Information
- Tuesday, February 5, 2013: 11:00 AM-12:00 PM
- Golden Gate 6, Lobby Level (San Francisco Hilton)
- Visit Xbridge Systems at Booth #202 in the Expo

© 2013 Xbridge Systems, Inc.

Barry Schrager

- Formed the SHARE Security Project in 1972 which created the requirements for today's mainframe security systems
  http://www.share-sec.com/1972-2012.html
- In 1973, one of 20 customers invited to Poughkeepsie to review the new MVM (aka MVS) Operating System before it was announced
- Architect and primary author of ACF2 – developed after IBM told SHARE Security Project that protection by default and algorithmic definition of datasets and users were not achievable.
- SKK sold to UCCEL in 1986 – ACF2 had a 60% market share
- Member, Mainframe Hall of Fame
  Honored to be included with: Admiral Grace Hopper,
  T.J. Watson, Jr. and Dr. Gene Amdahl

© 2013 Xbridge Systems, Inc.

Xbridge Systems

- Xbridge co-founded by Gene Amdahl and Ray Williams the co-founders of Amdahl Corporation
- Xbridge Systems redefines data discovery with DataSniff, the world's FIRST and ONLY automated Mainframe Data Discovery tool
- DataSniff Software is a powerful search engine for discovering sensitive data stored on IBM z/OS series mainframes

© 2013 Xbridge Systems, Inc.

State of South Carolina Breach

- Announced in October 2012
- 3.6 million Social Security Numbers
- 387,000 Credit Card Numbers
- Only a portion of the data was encrypted
- "The cost is also going to be enormous" (SC Gov. Nikki Haley)

© 2013 Xbridge Systems, Inc.

5

There Are Consequences Associated with Data Breaches

- **South Carolina Official Resigns After Security Debacle**
  - *November 21, 2012 - F*ollowing a security debacle that cost South Carolina more than $14 million, Department of Revenue Director Jim Etter will resign at the end of the year, Gov. Nikki Haley announced on Nov. 20
    - http://www.govtech.com/pcio/articles/South-Carolina-Official-Resigns-After-Security-Debacle.html

- **Utah CIO Steve Fletcher Resigns, State Promises Security Reforms**
  - May 2012 - Earlier this week, Utah State Chief Information Officer (CIO) Stephen Fletcher resigned his position over a data security breach that exposed the Social Security numbers (SSNs) and other personal information of 280,000 Medicaid patients
    - www.govtech.com/policy-management/Utah-CIO-Steve-Fletcher-Resigns-State-Promises-Security-Reforms.html

- **Congress grills DHS CIO over repeated security breaches**
  - June 21, 2007 - Officials from the Department of Homeland Security faced Congressional grilling yesterday over the (lack of) security on DHS computer networks. The department charged with safeguarding America's borders, ports, and airlines itself suffered 844 known attacks or security breaches in 2005 and 2006.
    http://arstechnica.com/security/2007/06/congress-grills-homeland-security-cio-over-repeated-security-breaches/

© 2013 Xbridge Systems, Inc.

Data Security Concepts

- Identify the user
- Is the user is who he says he is? – Authenticate
- Protect the resources – datasets, transactions, etc.
- Control who can access the resources and under what conditions (time of day, location of user, program path, etc.)
- It was just assumed that the security administrator knew the users, the contents of the dataset/resource, and the business purpose of the access

© 2013 Xbridge Systems, Inc.

Storage was manageable then (late 1970's)

The IBM 3330 with 100MB!!!

© 2013 Xbridge Systems, Inc.

Scope of Assessment for Compliance with PCI DSS Requirements

Text from the Payment Card Industry, Data Security Standards:

> ***"The first step of a PCI DSS assessment is to accurately determine the scope of the review. At least annually and prior to the annual assessment, the assessed entity should confirm the accuracy of their PCI DSS scope by identifying all locations and flows of cardholder data and ensuring they are included in the PCI DSS scope."***

© 2013 Xbridge Systems, Inc.

The Problem: Finding "All" Sensitive Data

Finding "obvious" sensitive data is easy, finding "not so obvious" sensitive data amongst the hundreds of thousands of datasets is challenging. For example:

- **An end-user copies Personally Identifiable Information (PII) data from a production personnel data set to his Userid prefixed dataset to test a new application**
- **The Security System controls allow all the user's co-workers read access to the data**
- **The *userid*.PERSONEL.MASTER.DATA dataset remains on the system for several months – maybe even years.**
- **Over time, co-workers make a copy of the file, then their co-workers make a copy of that copy, and …….**

© 2013 Xbridge Systems, Inc.

And often quite legitimate!

- How are you going to develop modifications to production application programs?
  - With copies of production data, of course
- How are you going to quality assurance test new production versions of applications?
  - With copies of production data, of course
- But, are these copies always deleted?
- And what security controls were in place when they were being used? Could the copies have been copied?
- IBM, at the 2012 System z Technical University, stated that up to 75% of the data on z systems was duplicated!!

© 2013 Xbridge Systems, Inc.

The world was different before

- No one had heard of Data Masking
- Or Tokenization
- Or Format Preserving Encryption
- Many places even used Social Security Numbers as identifiers
- Massive numbers of copies of production data sets created prior to recent encryption initiatives remain stored in clear-text form
- And some companies are still not using these techniques all the time!

© 2013 Xbridge Systems, Inc.

Example of Unknown Unknown Datasets:

- DataSniff customer scanned TSO Users' datasets
- Found many datasets which had been migrated
- These datasets contained name, Medical Diagnosis, Social Security Number, address, etc.
- They had not been referenced for years
- Were the results of queries against their databases
- All done in the line of business
- But not deleted
- The Data Security team did not have a clue that they existed

© 2013 Xbridge Systems, Inc.

Insiders Are Your Biggest Risk

- Verizon Data Breach Investigation Report
  - The report showed that insiders accounted for 48% of data breaches for the year. (2010 Report)
  - 66% of breaches involved data that the victim did not know was on the system (2008 Report)
- 2010 PacketMotion Survey -- US Government Agencies surveyed
  - 59% said employees are the biggest threat
- 2012 Information Week Strategic Security Survey
  - Which of these possible sources of breaches pose the greatest threat to your company in 2012?
    - 52% -- Authorized Users or Employees
- To eliminate unnecessary data and properly protect what is left, you must find it first (Verizon 2010 Data Breach Report)

© 2013 Xbridge Systems, Inc.

Outsourced?

- Don’t forget, that even though you may outsource part of your sensitive data processing, or even all of it, You are still responsible
- What measures is your outsourcer taking to protect your data?
- Do you have a commitment in writing?
- Do your own employees still have development and other access to the datasets on the outsourced systems?
- What about the outsourcer’s employees?
- Do you even know who they are?
- What datasets do they have access to?

© 2013 Xbridge Systems, Inc.

What kind of Data?

- Compliance and Regulatory issues
  (consumer protection)
  - Payment Card Industry – Data Security Standard (PCI-DSS)
  - Health Insurance Portability and Accountability Act (HIPAA)
  - Health Information Technology for Economic and Clinical Health (HiTech Act)
  - Sarbanes-Oxley Act (SOX)
- Legal/Financial Information
- Corporate Data Plans, Secrets, etc. –
  - Your company's Intellectual Property!!

© 2013 Xbridge Systems, Inc.

Accumulated Data as a result of M&A

- A large Banking Organization acquires a brokerage house after the 2008 financial meltdown
- Mainframe data is copied from brokerage house to the Bank – 400,000 datasets
- No one knows the data, where it is, or who should have access to what
- Sysprog assigned to analyze the data and validate access permissions

© 2013 Xbridge Systems, Inc.

Is Manual An Option?

Sysprog charged with discovering sensitive data obtained as a result of an acquisition.

> ***"Finding PII data manually in a catalog full of production data sets would take us decades. I wouldn't even scratch the surface before I'm eligible for Social Security. DataSniff could do it in weeks."***

Systems Programmer
Major Banking Institution
March, 2012

© 2013 Xbridge Systems, Inc.

Incorrect Access Permissions

- The system programmer found over 2 million inappropriate access permissions to datasets he reviewed manually for sensitive contents
- Couldn’t get management attention
- Called COO at home to get his attention and told him 3,500 people have access to his personal information

© 2013 Xbridge Systems, Inc.

Recent South Carolina Data Breach

South Carolina Inspector General message to all Agencies:

Locate all high risk data, primarily personal identifying information and protected health information. Additionally, request help on any personal identifying information or protected health information not sufficiently secured.

© 2013 Xbridge Systems, Inc.

Automation Alternatives?

- For Open Systems:
  - Identify Finder
  - Symantec Vontu
  - McAfee DLP

- For Mainframe z/OS
  - Xbridge DataSniff

© 2013 Xbridge Systems, Inc.

Why is the Mainframe such a challenge?

- 40 years of production datasets routinely copied for development testing and quality assurance
- Data accumulated from mergers and acquisitions
- Datasets are seldom deleted
- Migrated Data – possibly millions of datasets

© 2013 Xbridge Systems, Inc.

Why is the Mainframe such a challenge?

- 40 years of database query results and report output being stored in datasets
- Almost all datasets with an unknown structure
- Petabytes of Rogue Data that are unknown, unclassified and uncategorized
- Remember: If someone has READ access to a dataset they can download it to their PC
- **READ MEANS COPY!**

23
© 2013 Xbridge Systems, Inc.
23

Why is the Mainframe such a challenge?

- Almost all mainframe data is structured, but the structure is not known or easily associated with the datasets
- DataSniff will induce the structure:
  - *E.g. Credit Card numbers found in 1000 of 1000 records in packed decimal format in column 20*
- Mainframe applications often stored data in "packed decimal" format which saves about half the space
- Any analysis engine must recognize packed decimal, expand it and then look for PCI, PII, etc.

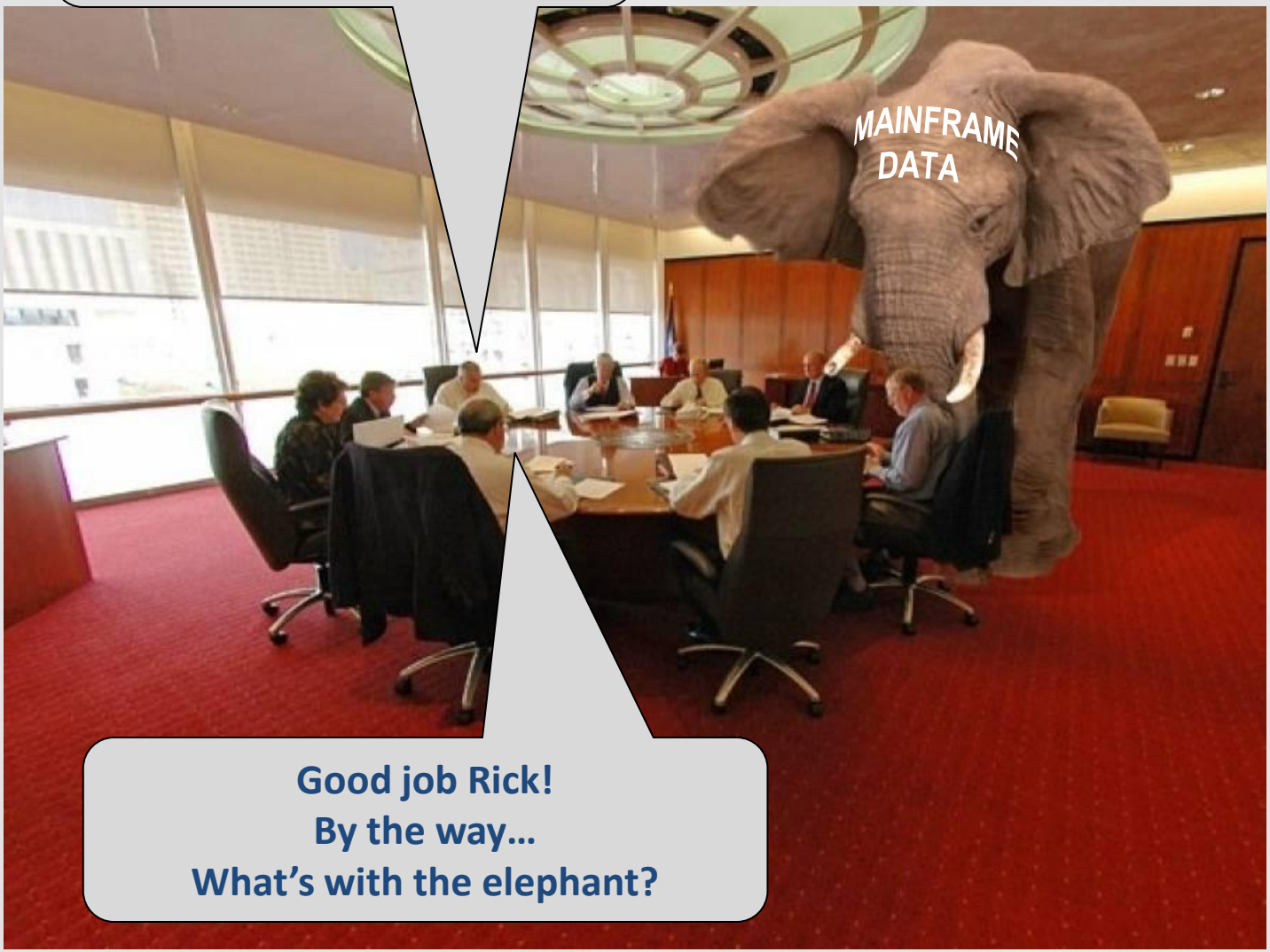
© 2013 Xbridge Systems, Inc.

The Challenges of Migrated Data

- Resides on secondary storage or tape
- Must be restored to primary disk storage
- Analyzed
- Re-migrated (not needed for Innovation's FDRABR)
- And not change the Last Referenced Date information

© 2013 Xbridge Systems, Inc.

All of our sensitive data is identified and protected.



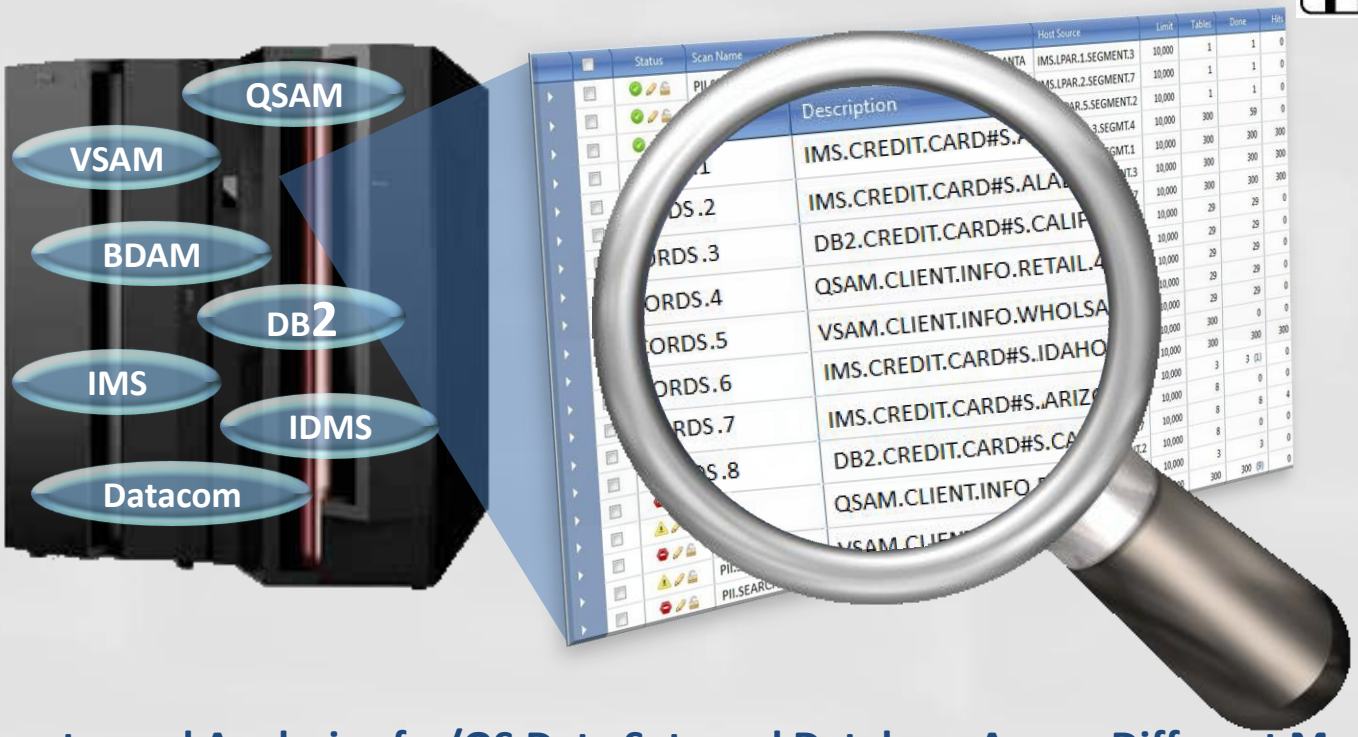
Good job Rick!
By the way…
What's with the elephant?

© 2013 Xbridge Systems, Inc.

The Mainframe DLP Solution

DataSniff

DataSniff: The Only Mainframe DLP Solution

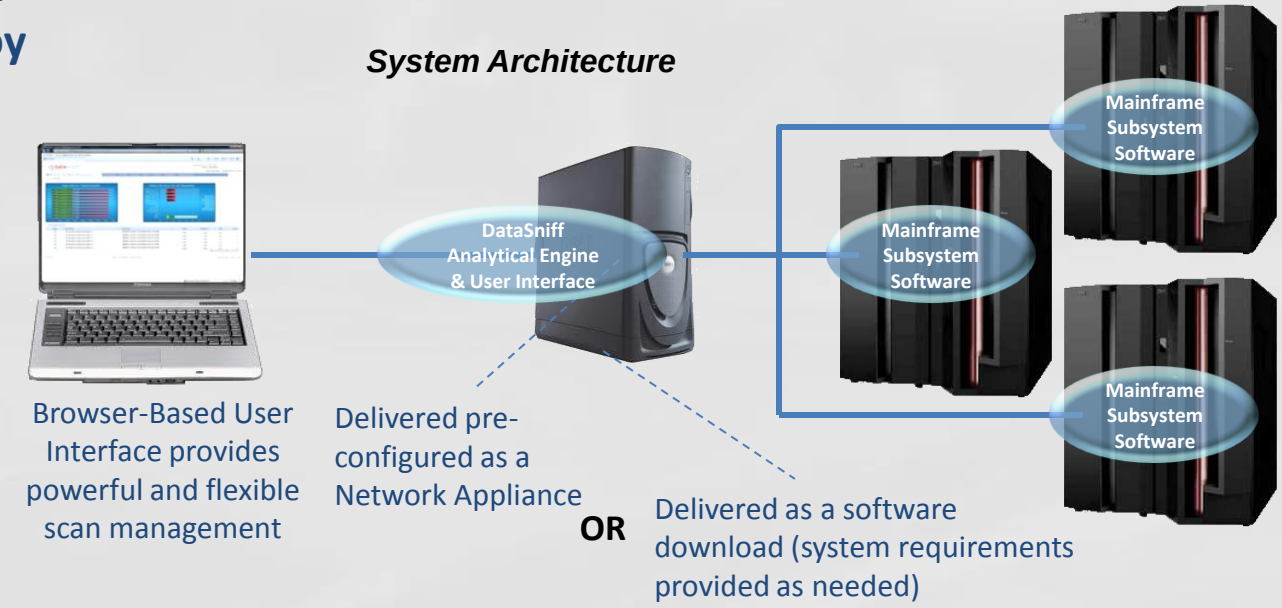


- **Access to, and Analysis of, z/OS Data Sets and Database Across Different Media**
  - IMS, DB2 & IDMS databases
  - VSAM, QSAM/BSAM, BDAM, PDS, GDG data sets
  - Migrated and Tape (physical and virtual)
- **Easily Scalable to Handle The Largest Data Storage Environments**
  - DataSniff is designed to maximize the efficient use of all components within its architecture to handle databases and data stores of all sizes and complexity

© 2013 Xbridge Systems, Inc.

DataSniff: The Only Mainframe DLP Solution

- Intelligent Architecture
  - z/OS resources highly controlled – Limited impact on production
  - Off load CPU intensive analysis
  - Understands Media Types – Not all data the same
  - Secure
    - Data Never Persisted
    - SSL Encryption
  - Easy to Deploy



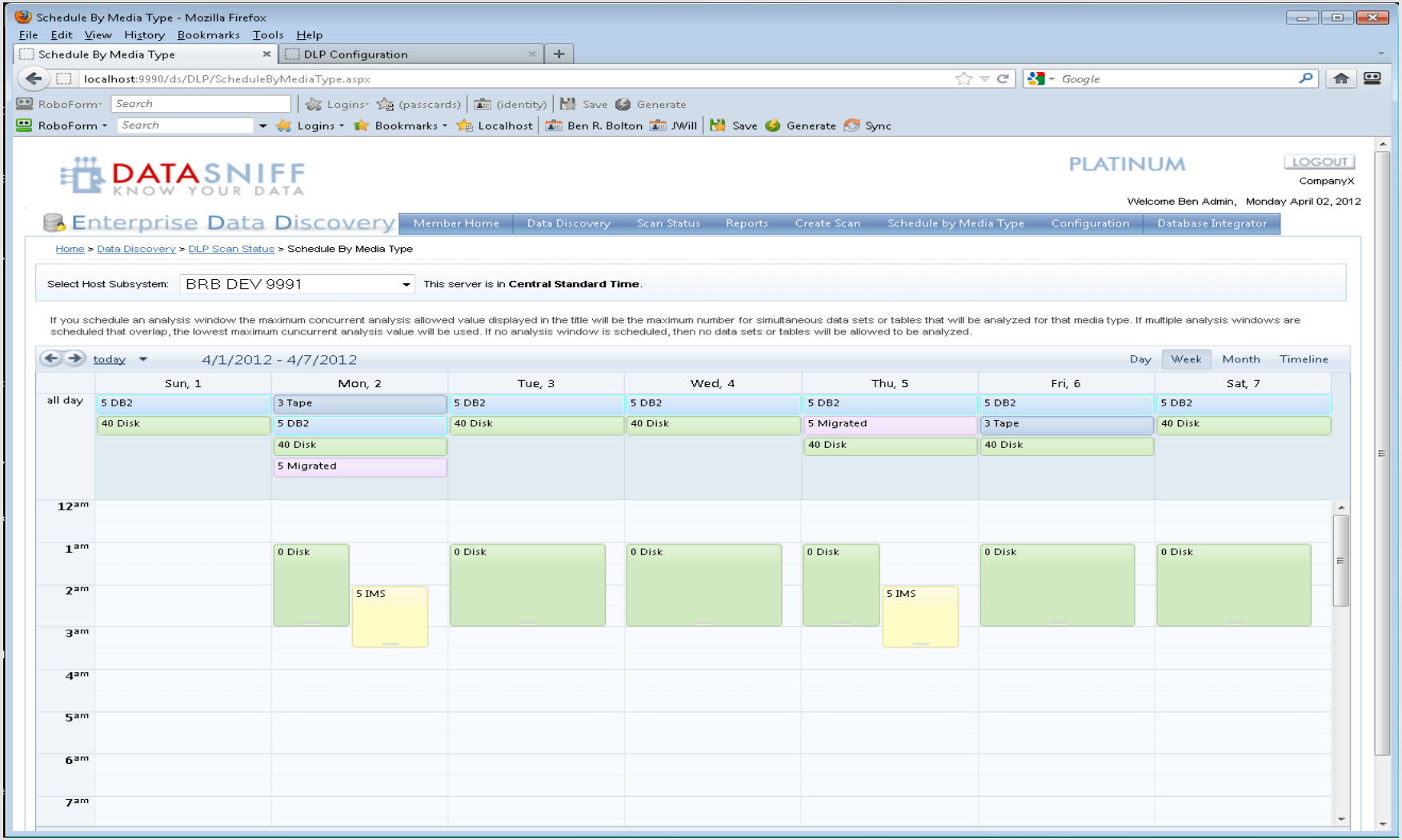
© 2013 Xbridge Systems, Inc.

What Makes DataSniff Special ?

- Schedule by Media Type (SMT)
  - It Ain't a PC!
    - Production Systems must not be affected
    - Migrated Data Sets Must Be Recalled
    - Primary Disk is Limited
    - Tape Heads are Limited
    - Resource Availability Varies with Time/Schedule
- Schedule as part of scan definition
  - Non-production data during off-hours
  - Production data when not running production

© 2013 Xbridge Systems, Inc.

Schedule by Media Type (SMT)



© 2013 Xbridge Systems, Inc.

What Makes DataSniff Special ?

- Create a scan definition
  - Where to Start – Organize the Chaos
  - Where's the Risk?
    - Test, Archives, Users, Acquisitions, Applications
  - Leverage the Catalog with Filters
    - Include What You Need
      - PROD.PAYMENT1.**
      - TEST.PAY*.**
      - DEV*.**
    - Exclude with Dispositions
      - Regular Expression
      - Manual
  - When is it best to run the scan?
    - Production data different than user data

© 2013 Xbridge Systems, Inc.

Suggestion by a GAO Auditor

- Best Practice – Don't copy sensitive production data to test systems.
- The Reality – Enforcing this Best Practice on mainframes is challenging.
- When searching for undiscovered / sensitive data;

1) Start with scanning non-production data
   - At least we know the production data is protected, but, it might not be properly categorized – meaning access permissions may not be totally correct
   - But, we don't know what is outside of production
   - The unknowns will hurt you

2) Then, scan production data to make sure nothing is miss-categorized

© 2013 Xbridge Systems, Inc.

What Makes DataSniff Special ?

- Enough is Enough!
  - Define Limits (Sample Size)
  - Define Thresholds
    - 100 of 100 – it's enough! Stop!
    - 569 of 10,000 – It's enough! Stop!
    - 11 of 10,000 – It's enough Maybe
- Rescan later
  - Rescan Transient Errors (security issues, in use)
- Analyze deep after you go wide
  - Rescan "At Limit"
  - Re-drive to Avoid Re-Analyze

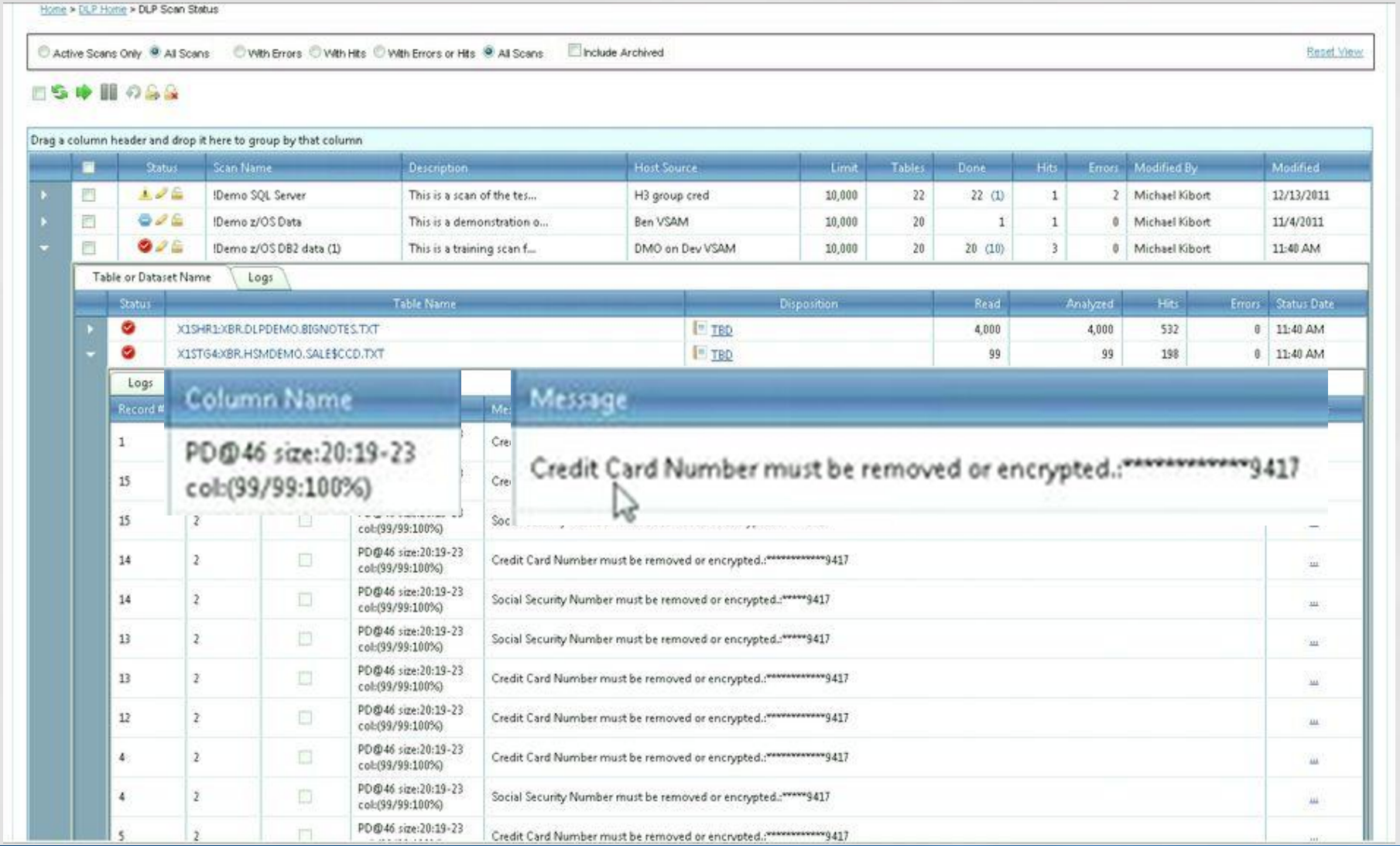
© 2013 Xbridge Systems, Inc.

What makes DataSniff Special ?

- Ability to automate process of inferring structure

© 2013 Xbridge Systems, Inc.

View “Hits” with Location Info



© 2013 Xbridge Systems, Inc.

DataSniff

- Locates sensitive data, such as CC#'s, SSNs, etc.
- By sampling a certain amount of data from datasets and database tables
- Creates an SQL database with information about each dataset including the Last Referenced Date
- Does not maintain the actual data on the DataSniff SQL Database, only the summary information
- Produces a report containing this information with the specific data obscured
- Allows the installation to locate information they did not know they had, so it can be remediated – deleted, encrypted, tokenized, etc.

© 2013 Xbridge Systems, Inc.

Mainframe Audit News

"DataSniff from Xbridge is a tool to find sensitive data on your mainframe, so you know what you need to protect. Protecting sensitive data is easy once you know where it is. Finding it is the hard part. DataSniff from Xbridge has just the tool you need to locate sensitive data throughout your z/OS system."

Stu Henderson, Sept 2011

© 2013 Xbridge Systems, Inc.

Continuous Monitoring

- Initial Scans:
  - All datasets selected by dataset name pattern
- Periodic Scans:
  - All new datasets since the last scan
  - All datasets that could have possibly been modified since they were last scanned

© 2013 Xbridge Systems, Inc.
40

What do I do now?

- How do I deal with the many datasets that DataSniff has located?
- Am I putting my organization at risk by finding all these locations of sensitive data?
- Is there a plan?

© 2013 Xbridge Systems, Inc.

Remediation /De-Identification Hurdles

- There could be thousands of datasets and database tables located by DataSniff
- Are they online or migrated?
- For datasets z/OS maintains the Last Referenced Date in the DSCB and the migration products save this
- DataSniff requests z/OS not to modify the Last Referenced Date or resets it
- DataSniff maintains this in its SQL database along with number of hits, etc. – can query for datasets not referenced in “x” months, etc.

© 2013 Xbridge Systems, Inc.

Data Retention Guidelines

- Determine organization's data retention guidelines for different categories of data
- If none, meet with company management and legal department and develop guidelines
- Some datasets you'll find have not been referenced for 20 years – you can't keep everything? Well, maybe …

© 2013 Xbridge Systems, Inc.

If it hasn't been used in a long time

- Beyond Company data retention guidelines?
- Talk to the people in charge:
  - Option 1 – Just delete the data
  - Option 2 – encrypt the data, archive the data to removable media and store the media in a vault
- Objective is to make intervention by Legal and/or Compliance Departments mandatory
- You want human intervention before access

© 2013 Xbridge Systems, Inc.

If It Has Not Been Used Recently But Still Within Retention Limit

- Was the last access attempt over 6 months or a year?
- If so, consider encrypting the dataset
  - Use a migration product to migrate it
  - It will still be in the catalog
  - When the user tries to access the dataset, it will be unreadable
  - Users will complain and you can deal with it on a case by case basis
  - Is the sensitive data necessary?

© 2013 Xbridge Systems, Inc.

If it has been used recently

- Use the IEBEYEBALL utility ☺
- This dataset/table may be a false positive
- It also may have credit card numbers or other sensitive information placed there in comment fields with no residual business purpose – if so, just clean it up
- Should it be moved to the production data environment? Why is it not there now?
- Sensitive data will be better protected if it is treated as production data

© 2013 Xbridge Systems, Inc.

Who's using the datasets?

- Scan SMF to report on all users of a dataset
  - How far back do you go? 3 months, 15 months?
  - How many tapes is that?
  - What programs are being used to access them?
- Or, ask ACF2, RACF or Top Secret who has access
  - ACF2 and Top Secret have built-in commands for this
    - Or EKC's ACF2/RACF Access Analysis product
    - Or Compliance Information Analysis component in ACF2 & Top Secret
  - For RACF, use Vanguard or zSecure products
- Contact all the authorized users to see who the owner is, if they are using it, and how they are using it

© 2013 Xbridge Systems, Inc.

How are they using it?

- Reading as text?
  - Do they need the sensitive fields?
- Encrypt, Data Mask or Tokenize the fields
- Or encrypt the entire dataset
  - And change processing programs to decrypt as each record is read
  - Be careful about adding a preliminary step to decrypt the dataset into a temporary scratch dataset
  - It might get left around – use Virtual I/O (VIO) instead

© 2013 Xbridge Systems, Inc.

Journal All Accesses

- Change ACF2 Access from ALLOW to LOG
- Add AUDIT to RACF Profiles
- Add to AUDIT Profile for Top Secret
- Or use CA Chorus or Compliance Manager to monitor access for ACF2 or Top Secret

© 2013 Xbridge Systems, Inc.

Set Classification

- ACF2 – use DCLASS Infostorage classification
- RACF – add LEVEL to PROFILES
- Top Secret – use DCLASS
- For ACF2 & RACF, use EKC's Security Reporting Facility for resource classification and reporting
- For ACF2 & Top Secret, use CA Chorus or Compliance Manager

© 2013 Xbridge Systems, Inc.

Reporting on Accesses

- Use ACF2 Reports with DCLASS parameter
- RACF – use Vanguard & zSecure
  - use LEVEL parameter to select based upon profile value
- RACF & ACF2 – use EKC's Security Reporting Facility
- ACF2 & Top Secret – Use CA Compliance Information Analysis, CA Chorus or Compliance Manager

© 2013 Xbridge Systems, Inc.

Summary

- Mainframes have decades of data collected
- Unknown datasets are dangerous
- Which ones are dangerous?
- What's in them?
- A Manual Process will take decades, employ people for their lifetimes and the organization will be vulnerable until the process is completed
- Automation is the answer
- DataSniff is the only z/OS Data Discovery Product
- Then the datasets & tables containing the sensitive information can be deleted, encrypted, etc. and have their access authorities reviewed

© 2013 Xbridge Systems, Inc.

DataSniff User Session

- David Wade, CIO of Primerica
- 12943: How to Identify Datasets Containing PCI, PII or Other Sensitive Information
- Tuesday, February 5, 2013: 11:00 AM-12:00 PM
- Golden Gate 6, Lobby Level (San Francisco Hilton)
- Visit us at Booth #202 in the Expo

© 2013 Xbridge Systems, Inc.

Questions?

Visit us at EXPO Booth # 202

Xbridge Systems

1941 Landings Drive

Mountain View, California, 94043

Please feel free to contact us with specific questions about DataSniff, Mainframe Data Discovery or any questions regarding Xbridge Systems, or mainframe compliance controls – thank you.

For more information about our products, please send an e-mail to sales@xbridgesystems.com or phone toll-free 1-866-356-1515.

For customer support, please send e-mail to support@xbridgesystems.com or phone toll-free 1-877-209-1515.

© 2013 Xbridge Systems, Inc.