



IMS Test Data Management

Virtual IMS User Group 4 February 2014

John B Boyle

Informatica Software

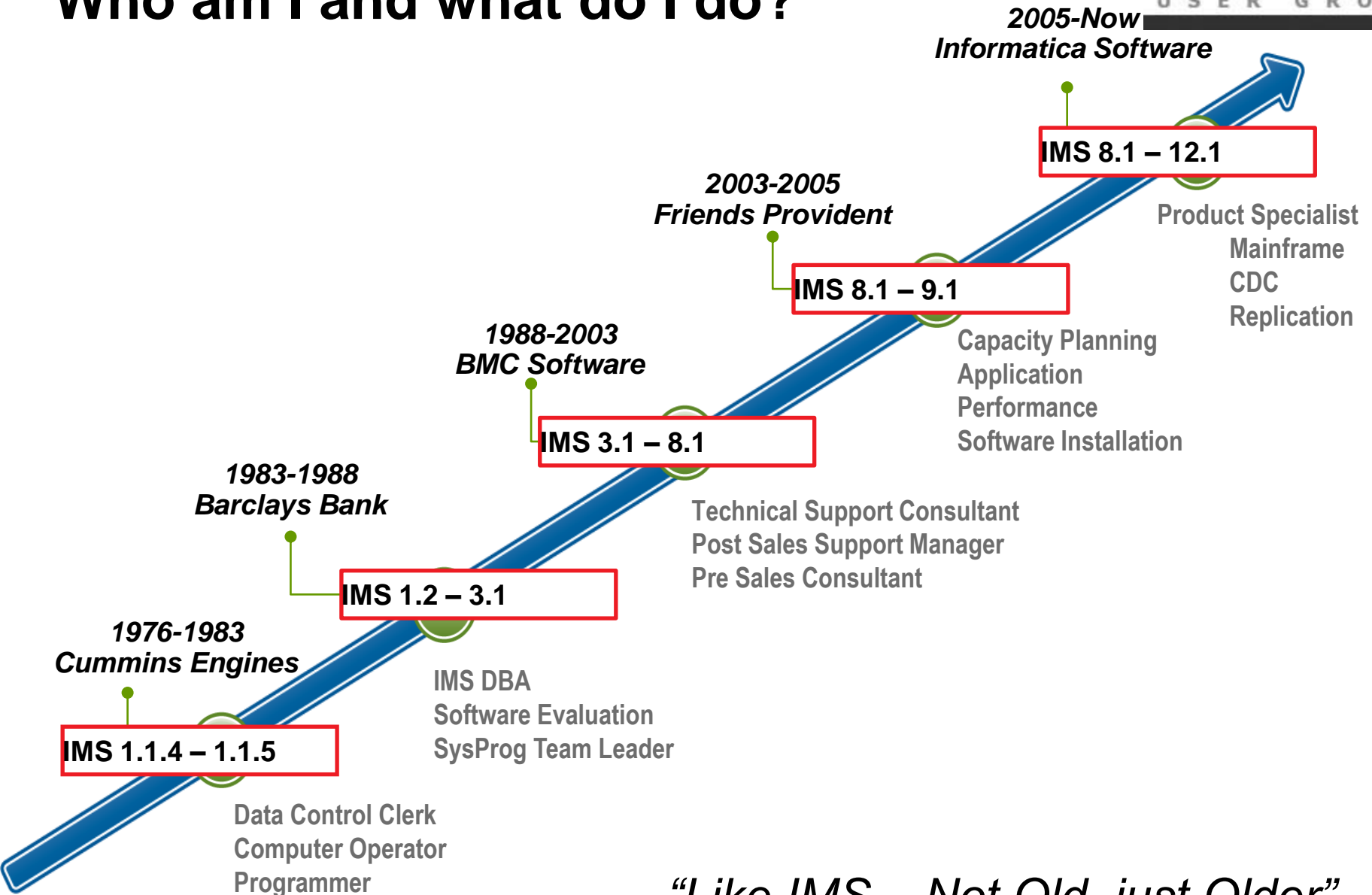
Abstract

- This session will look at the issues and challenges facing IT Development related to the creation and maintenance of appropriate test data to be used in the testing of new and modified applications. It will then provide an overview and demo of Informatica's Test Data Management Products, with specific reference to their use when IMS data forms all or part of the required Test Data. This will be supplemented by a review of two recent Proof of Concept exercises undertaken for a European Insurance Company and a European Bank, both of which included IMS Data.

Agenda

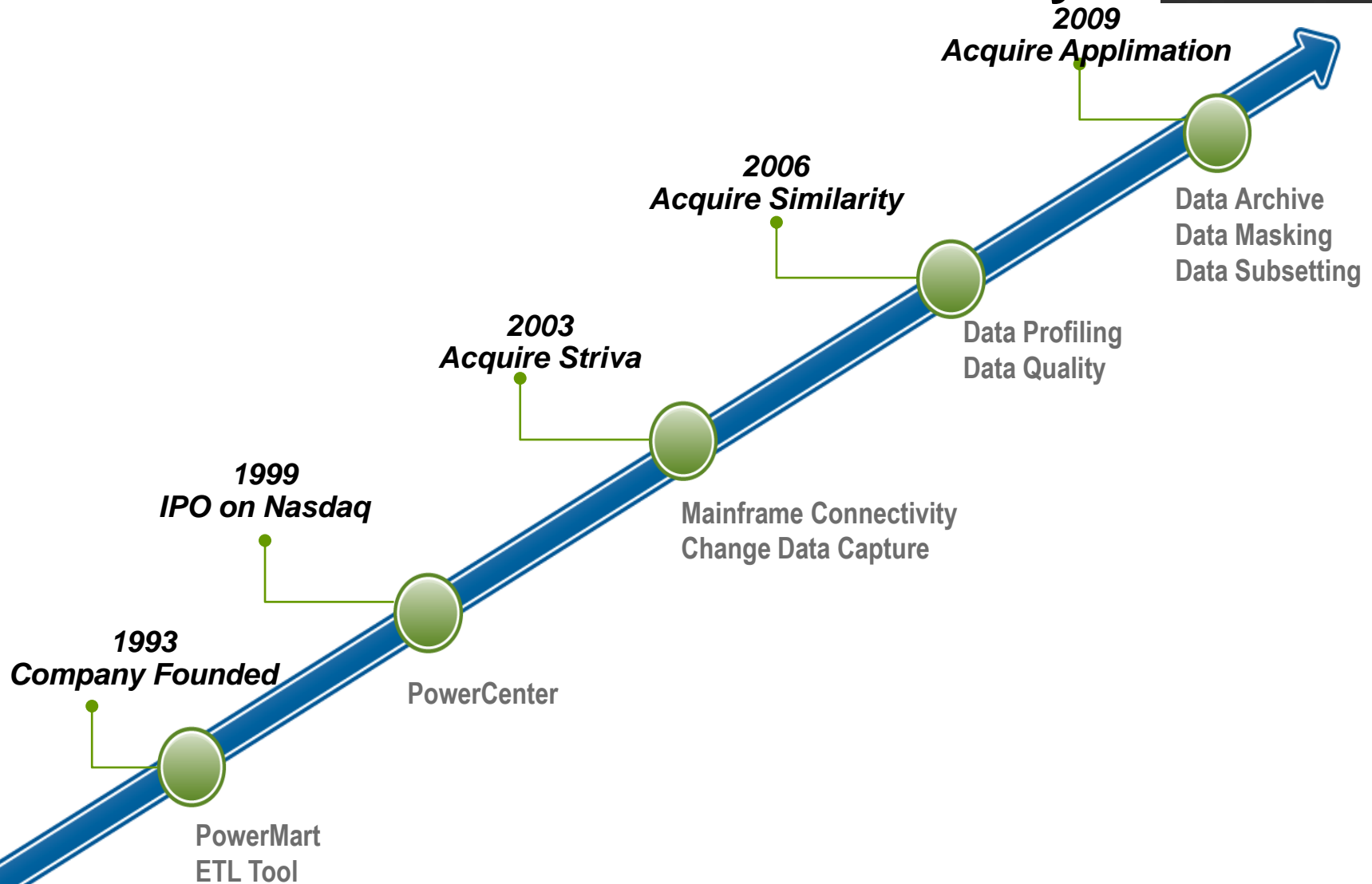
- Who am I and what do I do?
- Who are Informatica?
- A little history
- Issues and Challenges
- Informatica Test Data Management for IMS
- Case Studies

Who am I and what do I do?



“Like IMS – Not Old, just Older”

Who are Informatica and what do they do?



Maximize & Unleash Information Potential

Your
Data

- Understand it
- Integrate it
- Cleanse it
- Relate it
- Secure it
- Act on it

Your
Architecture

- Across infrastructure, applications & devices
- Fast, flexible, free of lock-in and future-proof

Your
Business

- Deliver agility
- Reduce risk
- Unleash potential

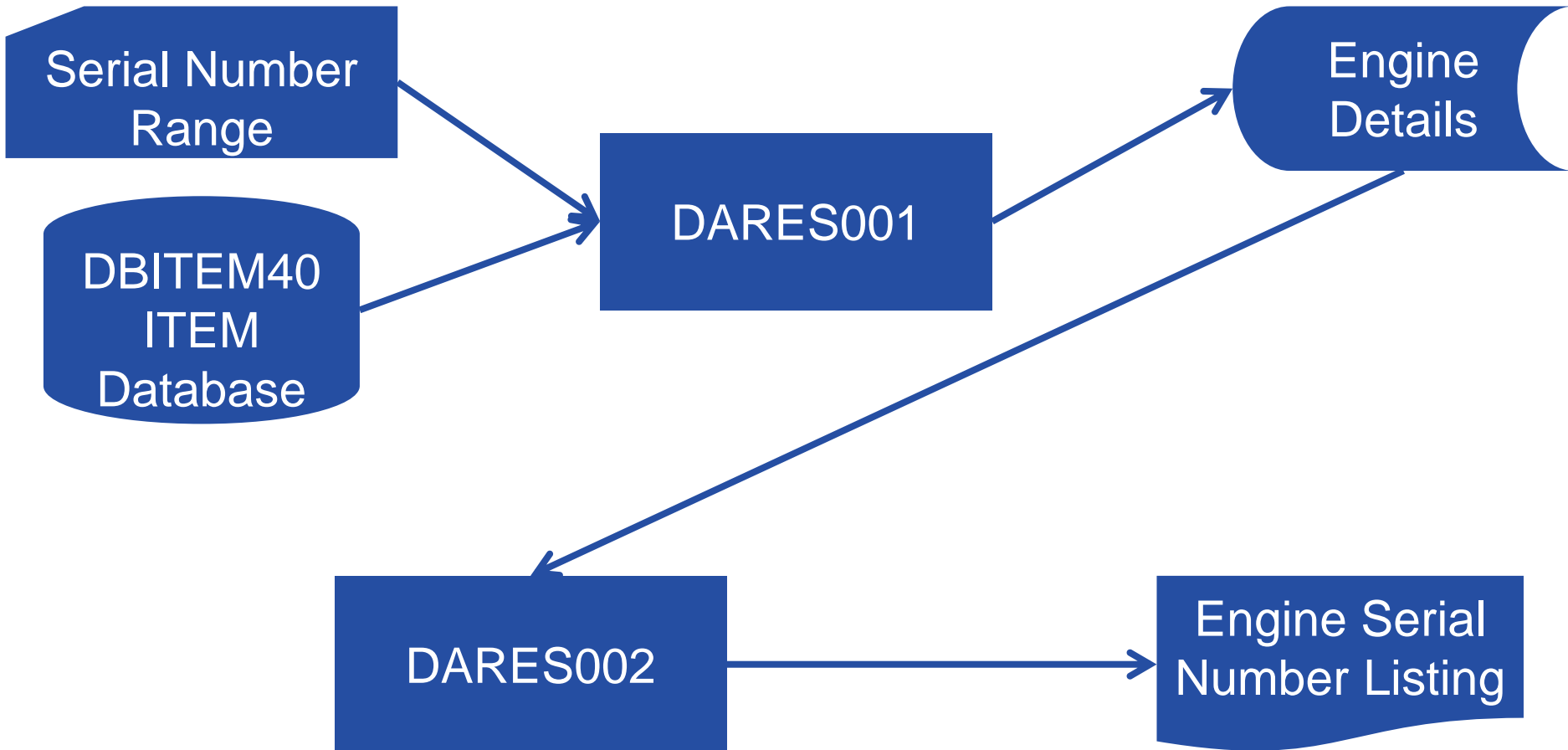


Why does Test Data Management matter?

Issues and Challenges

My very first system

- Engine Serial Number Reporting



My Very First Program

- DARES002 (Engine Serial Number Listing)
 - We'd just started using IBM's Data Dictionary
 - DAR = Darlington Systems
 - ES = Engine Serial Number Project
 - 002 = Second program registered.
- Written on coding sheets
- Keyed (to disk) by Punch Room
- Compiled, Linkedited
 - Not quite first time!
- Desk Checked
- Ready to test

My first Testing Problem

- DARES001 'not quite finished'
 - I need the file it produces to test my program
- Need to Generate some Test Data
- Use a Utility – IEBGENER (was IEBPTPCH ??)

```
//S01      EXEC PGM=IEBGENER
//SYSPRINT DD SYSOUT=*
//SYSUT1   DD *
1111111111
2222222211
3333333311
4444444411
/*
//SYSUT2   DD DISP=(,CATLG,DELETE),SPACE=(TRK,(1)),
//          DSN=JBOYLE.TEST.ESLDATA,
//          UNIT=SYSDA,
//          DCB=(RECFM=FB,LRECL=30,BLKSIZE=3600)
//SYSPRINT DD SYSOUT=*
//SYSIN    DD *
GENERATE  MAXFLDS=3,MAXLITS=26
RECORD   FIELD=(10,1,ZP,1),
          FIELD=(20,'ENGINE DESCRIPTION X',,7),
          FIELD=(6,'300177',ZP,27)
```

Now I have Data

- Took almost as long as writing the program!

```

EDIT          JBOYLE.TEST.ESLDATA
Command ==>
=COLS>  ----+----1----+----2----+----3
*****  *****
000001          ENGINE DESCRIPTION X  "
          011111CDCCDC4CCECDCDECDD4E0017
          11111F55795504523997396507307F
-----
000002          ENGINE DESCRIPTION X  "
          022221CDCCDC4CCECDCDECDD4E0017
          22221F55795504523997396507307F
-----
000003          ENGINE DESCRIPTION X  "
          033331CDCCDC4CCECDCDECDD4E0017
          33331F55795504523997396507307F
-----
000004  ààà  ENGINE DESCRIPTION X  "
          044441CDCCDC4CCECDCDECDD4E0017
          44441F55795504523997396507307F

```

Things my Test Data Didn't Test

- Empty Input File
- Invalid Build Date
- Invalid Serial Number
- New Page
- More than 9 Pages
- Lots of other things...

Do you test your IMS Applications?

 No

 Wrong Answer

 Look for a new job

 You're going bust

 Yes

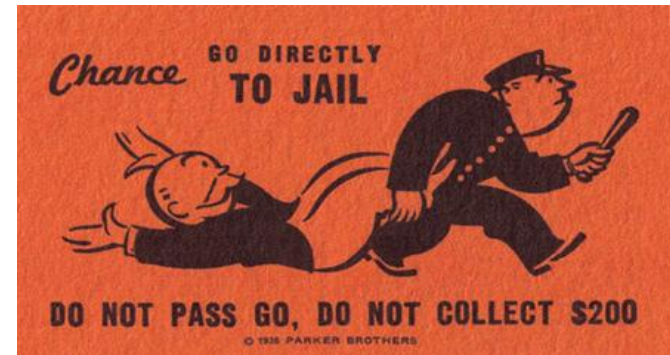
 Right Answer

 Proceed to next slide

What data do you test them with?

☹️ Copy of Production Data

- ☹️ Wrong Answer
- ☹️ Look for a new job
- ☹️ You're going to jail!



😊 A subset of Production Data which has been de-personalised

- 😊 Right Answer
- 😊 Proceed to next slide



What do we need and Why?

- A Subset of Production Data for Testing
 - DASD is cheap, but it isn't free
 - Test runs take too long on live sized datasets and use too much system resource
 - Everyone wants their own test environment
- Data which does not contain Sensitive Information
 - What is Sensitive?
 - PII – Personally Identifiable Information
 - PHI – Personal Health Information
 - CCI – Credit Card Information
 - Most Countries now have Data Privacy Legislation – and Penalties!
 - UK Data protection act
 - High profile 'data losses' = bad publicity
 - Offshore testing and Cloud-based Applications

Issues with creating a subset

- High Volumes of data to process when building a subset of production data
 - Subset needs to be kept current – so this is not a one time process
- Subset must be 'consistent' with respect to
 - Database defined RI
 - Application defined RI (invisible from Database Metadata)
 - Cross Database RI (even less visible)
 - Cross Platform RI Applications span platforms
- Packaged Applications
 - Understanding a data model containing 80,000 tables?

Options for Creating a Subset

- In-House developed Programs and Scripts
 - As Part of initial application design?
 - Developed some time later?
 - Maintained by?
 - What about Packaged Applications like SAP, Oracle E-Biz?
 - Enterprise wide consistency?
- Vendor Supplied Test Data Management Tool
 - Support for All Platforms, Databases, Files used by the Enterprise?
 - License and Maintenance Cost?
 - Ease of use?

Issues with Data Masking

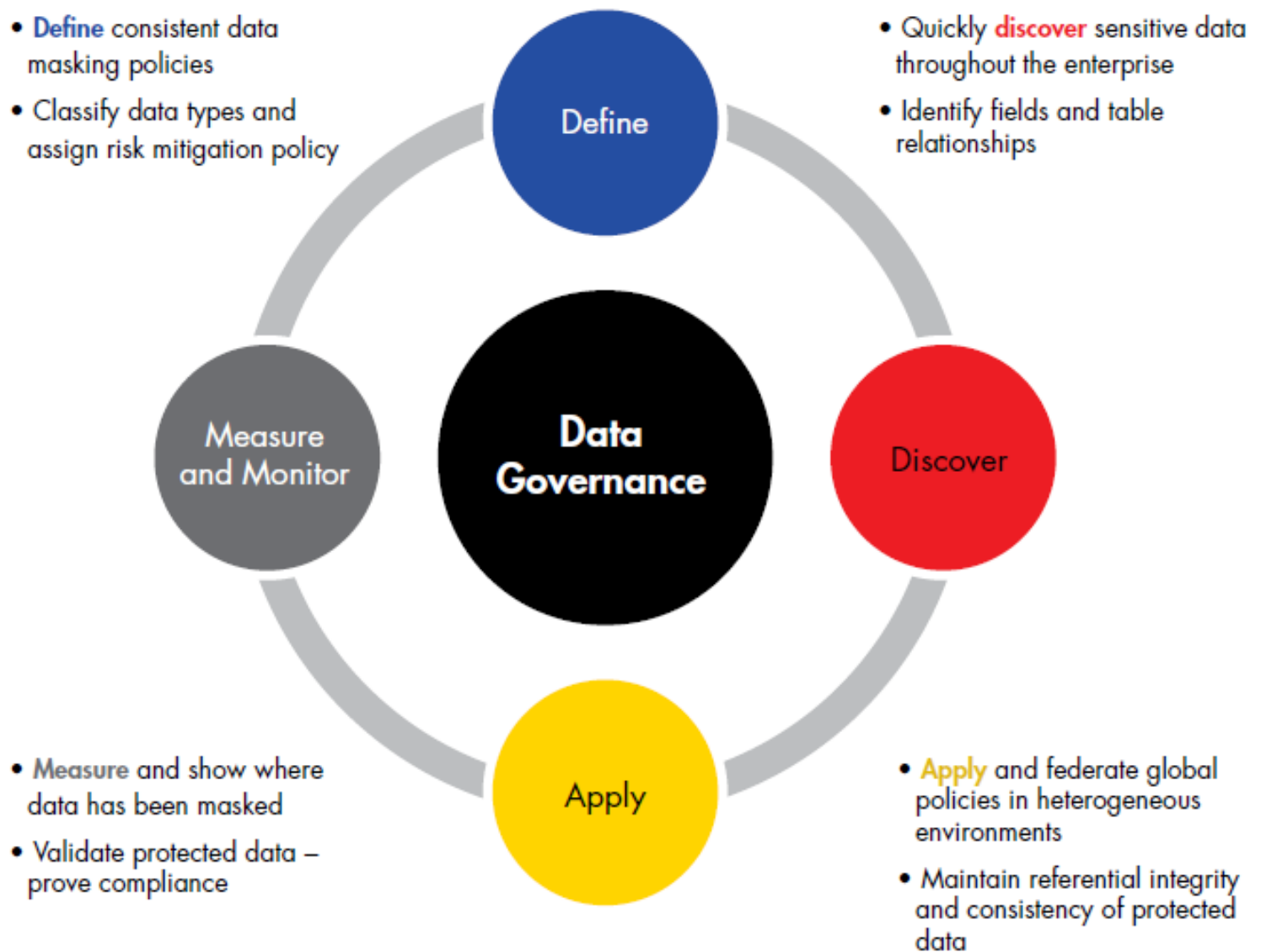
- Identifying what needs to be masked
 - Masking all data is counter productive
- Data masking must be consistent
 - Same issues as for consistent Subsetting – data is inter-related by Database and Application RI
- Masked Data must be representative
 - Testing with Random data will not exercise application logic
- Masked data must be secure
 - Prevent 'reversal' of masking to give original data
- Masking must be repeatable
 - Need to refresh test data with new data periodically

Options for Data Masking

- In-House developed Programs and Scripts
 - As Part of initial application design?
 - Developed some time later?
 - Maintained by?
 - What about Packaged Applications like SAP, JD Edwards?
 - Enterprise wide consistency?
- Vendor Supplied Test Data Management Tool
 - Support for All Platforms, Databases, Files used by the Enterprise?
 - License and Maintenance Cost?
 - Ease of use

Same as Subsetting

Data Masking Solution – Basic Requirements



What Vendor?

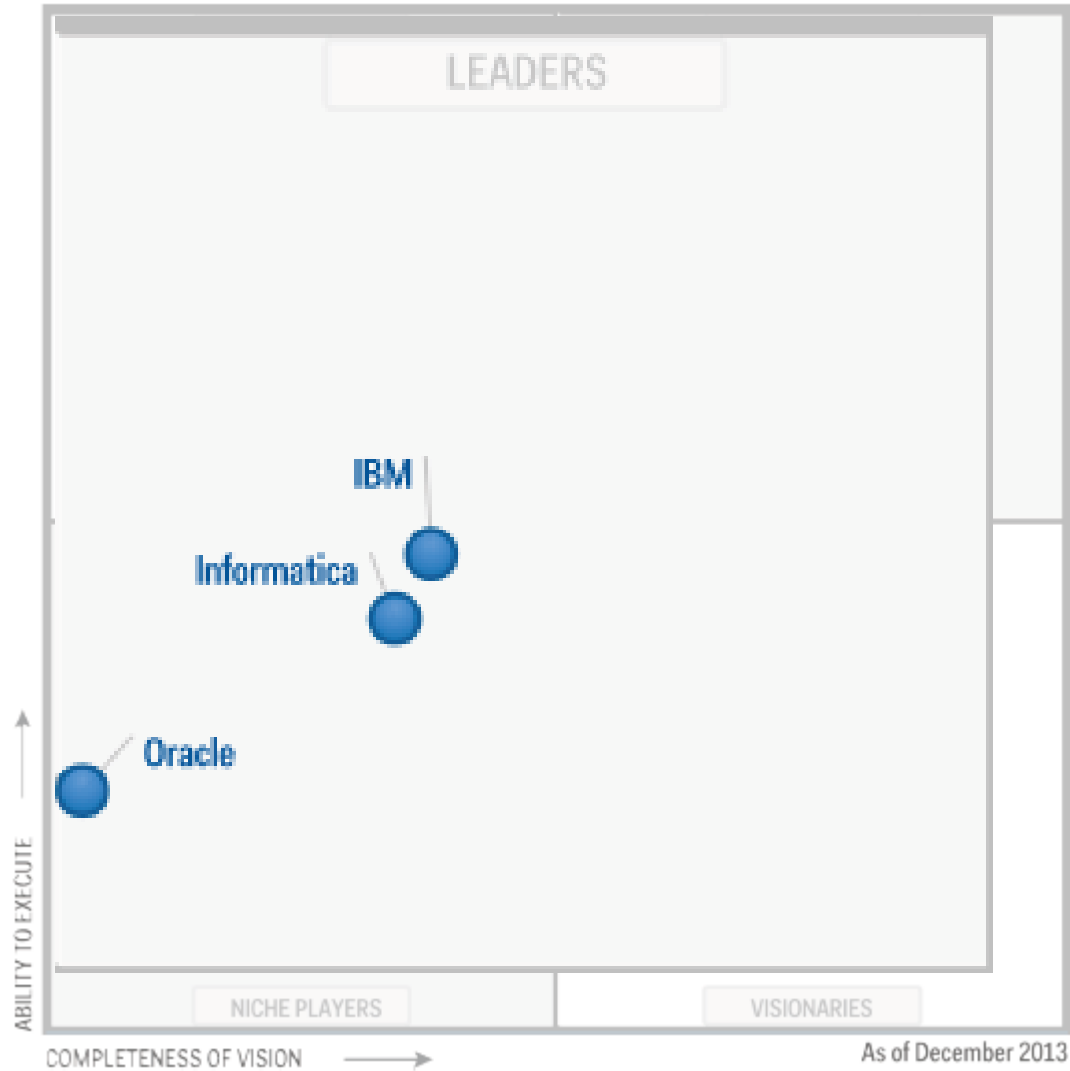
Magic Quadrant

Figure 1. Magic Quadrant for Data Masking Technology



What Vendor?

Figure 1. Magic Quadrant for Data Masking Technology



Source: Gartner (December 2013)

As of December 2013

Test Data Management - Enterprise

1. Establish what data items needs to be masked
2. Establish how these data items should be masked
3. Use results of 1 and 2 to establish and document Enterprise Data Masking Policy
4. Build Data Masking Rules for the identified data items
5. Establish where these data item are stored
6. Build Data masking processes to mask these data items
7. Establish Inter-Object relationships and dependencies
8. Build subsetting process to limit volume of data written to 'test set' when we mask the data
9. Execute!

Informatica Test Data Management for IMS

And other less important things

TDM Overview

- Browser based 'Workbench' used to Define the entire Test Data Generation Process
- Subsetting
 - Which records to move (retain)
- Masking
 - How to 'de-personalise' the data
- Generates PowerCenter Mappings
 - Contain the logic to move 'Source' to 'Target'
- Generates PowerCenter Workflows
 - Physical Connection to Source and Target
- Workflows use PowerExchange to read and write the data
 - PowerExchange provides access to Mainframe Data, Applications/Packages, and Relational Databases on LUW platform

PowerCenter Overview

- General Purpose Data Integration Tool
- Moves data from any source to any target
- Transforms the data on the way through
- Graphical Design tool with extensive range of pre-defined 'Transformations'
- All 'objects' stored in Metadata Repository
 - Metadata Manager provides full lineage analysis

PowerCenter Objects

- Source – Source Metadata Definition
- Target – Target Metadata Definition
- Transformation – Pre-defined processes sorter, joiner, union, router etc.
- Mapping – Links Source via Transformation Logic to Target
- Session – Controls Execution of a mapping – specifies physical connection properties to read source and write target
- Workflow – Controls execution of a series of sessions

PowerExchange Overview

- Provides PowerCenter processes with access to sources & targets where no native client access is available
- Relational Source - DB2
 - Allows Import of Metadata from DB2 Catalog
- Non-Relational Sources - IMS, IDMS, ADABAS, DATACOM, VSAM, SEQ
 - Requires creation of Datamap to provide Relational View

What does TDM provide

- A profiling tool
 - Find relationships between data sources
 - Find sensitive data based on format and metadata
- A workbench
 - To define how to select data to be included in a subset operations
 - To define what to mask
 - To define how to mask specific data items
- An execution framework
 - To manage the execution of the Masking and Subsetting processes
- A Dashboard
 - To view the status of your various Test Data Management Projects

TDM Profiling

- Exploits Informatica's Data Explorer product (also part of our Data Quality Solution)
- Report on unique/non-unique columns
- Establish potential primary key and foreign key relationships between objects.
- Suggest groups of related objects (Entities) which could be processed together to provide consistent subset

Data Subsetting Objects

- Entity – Basic building block for subset creation
 - Contains Main Table, Component Table and Filter Criteria
- Group – Additional Tables where all rows are to be selected
- Template – Collection of Entities and Groups
- Plan – Executes the Subsetting Components
 - Specifies the actual source and target

Data Masking Objects

- Rules – Basic building block for masking
 - Define how a particular data element is to be masked
- Policy – Contain one or more rules
- Plan – Executes the Masking Policy
 - Specifies the actual source and target
 - Can execute Subsetting Template and Masking Policy in the same Plan

What does all this mean?

- Using PowerCenter you can define reusable processing logic which can be applied to different physical sources/targets
- TDM further extends this re-usability by allowing the creation of re-usable rules, entities, policies which can also be applied to multiple different sources/targets – even in different DBMS structures – like DB2 and Oracle
- Higher Re-use → Lower Development Costs

Out of the box masking rules

- Substitution
 - Repeatable or non-repeatable
 - Unique or Non-Unique
 - Dictionaries supplied or use your own
- Specific Data Types
 - Phone Number/Credit Card Number/email/url/IP Address
- Random
 - Blurring
 - Ranges
- Shuffle
- Advanced
 - Write your own in PowerCenter
 - Implemented as a re-usable Mapplet

Test Data Management

1. Profile your data to establish inter-relationships
2. Build your Masking Rules
3. Build your subsetting criteria
4. Apply masking rules to the appropriate columns
5. Build execution plan
6. Execute plan to read production data and write test data
7. Repeat step 6 as required to build additional test sets and refresh existing test sets

Does it work for IMS Data?

😊 YES!

- 😊 Create an Unload File containing production Data
- 😊 Perform steps 1 to 7 on previous slide
- 😊 Execution Plan reads the Production Unload and writes a subsetted and masked Unload File
- 😊 Use this unload file to load your test databases
- 😊 Keep your job and stay out of jail

Why use Unload Files

- You don't have to!
 - Can also read and write from/to IMS Database
 - BMP, Batch DLI or ODBA
 - Issues with Unkeyed Segments – RULES=FIRST/LAST
- Clean 'point in time' copy
 - Create multiple subsets from same source
 - Repeatable process
- Minimises impact on Production Systems
- Better performance / lower overhead
 - Everyone has high performance Unload/Load utilities
 - If you don't talk to IBM/BMC/CA/???

Case Study – Swedish Bank

- Sophisticated Test Data Management Project
- Extract production data from IMS, DB2 and Oracle
 - Mask it as required
 - Store it in Oracle.
- Create application specific ‘test sets’ by extracting stored data from Oracle and writing it back to ‘source’ DBMS
- Initial project started before Informatica acquired Applimation, now migrated to TDM
- Control application built to facilitate extraction of data

Case Study – German Insurance Company

- Multiple Source on different platforms
 - DB2 and IMS on z/OS
 - Oracle on Linux
 - SAP on Unix
- Combination of Custom masking rules for specific data items and standard masking techniques
- Some new SAP module specific Accelerators created
- Entities consisting of IMS DB2 and Oracle objects are required

Case Study: Medium Enterprise Insurance Company

Overview

A recent audit of this insurance company's data privacy and protection processes revealed that existing methods for procuring data for testing purposes and manual methods to mask sensitive information were non-compliant with existing PCI, PHI, and Sarbanes Oxley (SOX) data privacy requirements. In addition, these processes resulted in higher testing and development costs for new and existing IT investments and significantly increased their risk of an unwanted data breach.

Solution

In response, the company adopted Informatica Test Data Management to streamline acquisition of realistic and purposeful data to avoid copying entire data sets from production systems for testing purposes. Packaged data masking policies and rules compliant with PCI, PHI, and SOX were also applied. Masked data was validated against required policies before using it for testing purposes

Results

The company realized greater than 50 percent in time savings using Informatica Test Data management vs. previous methods. The number of defects in testing processes were reduced by 30 percent or more. Usage of Informatica Data Subset increased in time savings over 50 percent to capture optimal test cases.

Want more Information?

Contact your friendly Informatica Sales Rep

Or email me at jboyle@informatica.com

Or visit the Informatica Web site

<http://www.informatica.com/us/solutions/application-information-lifecycle-management/test-data-management/>

Bibliography

- Securosis Research: Understanding and Selecting Data Masking Solutions
 - https://securosis.com/assets/library/reports/UnderstandingMasking_FinalMaster_V3.pdf
- Gartner Magic Quadrant
 - <http://www.gartner.com/technology/reprints.do?id=1-1DC8KNJ&ct=121221&st=sb>
- Enterprise Strategy Group Report on Informatica Data Masking
 - http://www.informatica.com/Images/02162_esg-enterprise-data-masking_ar_en-US.pdf
- Informatica Case Study
 - http://www.informatica.com/Images/02494_accelerating-insurance-legacy-modernization_wp_en-US.pdf

● ● ● ● ● ● ● ● ● ●
informatics

Thanks for Listening

*“We cannot hold a torch to light another's path
without brightening our own”*