Transforming Test Data Management with Copy Data Virtualization

May 2017

MAVERIC
Transforming Assurance

Agenda

Test Data Management

- Why TDM
- TDM Focus Areas
- TDM using Data Masking, Synthetic Data Generation & CDV

Source: Maveric-Systems

Test Data Management

Why TDM

4 KEY AREAS THAT DEMANDS TEST DATA MANAGEMENT

Distributed Environment

- Different phases of SDLC are executed in multiple places and the corresponding data environment may not be controlled by the banks
- Time to Market should be minimal without impacting the quality of the deliverables

Data Complexity

- Data is often stored in legacy back-end and scattered across different DBs
- The problem is compounded by lack of sufficient documentation supporting the often decades-old legacy system

Types of Testing

- Different types of testing (QA, UAT, SIT, Performance Testing, etc.) requires different sets of data
- Data coverage should be sufficient enough for testing requirements
- Effort spent by testers to prepare test data should be minimized
- Data for complex test cases may not be readily available in the correct format in the database

Data Security

- Sensitive & confidential data in production shouldn't be copied to test environments
- Data copied from production should follow specific Regulation & Compliance laws
- Rising incidents of data breach, forces banks to limit access to production data for cloning purposes

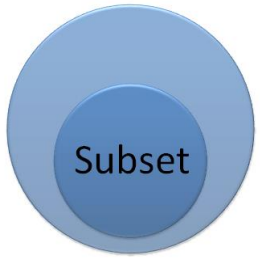
TDM Focus Areas

Data Masking

- Ability to wash the data using techniques such as lookup, scrambling, shuffling, etc.
- Ability to wash the data using different techniques for different data releases without any code change
- Ability to wash the data without losing the business context

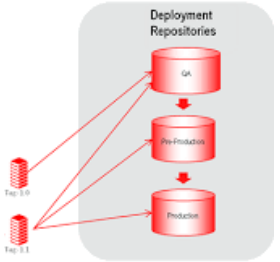
Data Manufacture

- Ability to mine the production data to identify the data set for each requirement in hrs.
- Ability to manufacture the data in minutes
- Ability to manufacture the data that resembles production data



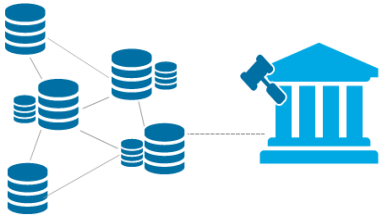
Data Sub-set

- Identify data relationships through 'Data Discovery' and 'Data Lineage'
- Classify data
- Refresh data from production database(s) using native tools/methods
- Maintain 'Golden Copy' to reduce TDM related work load on production database(s)



Data Deployment

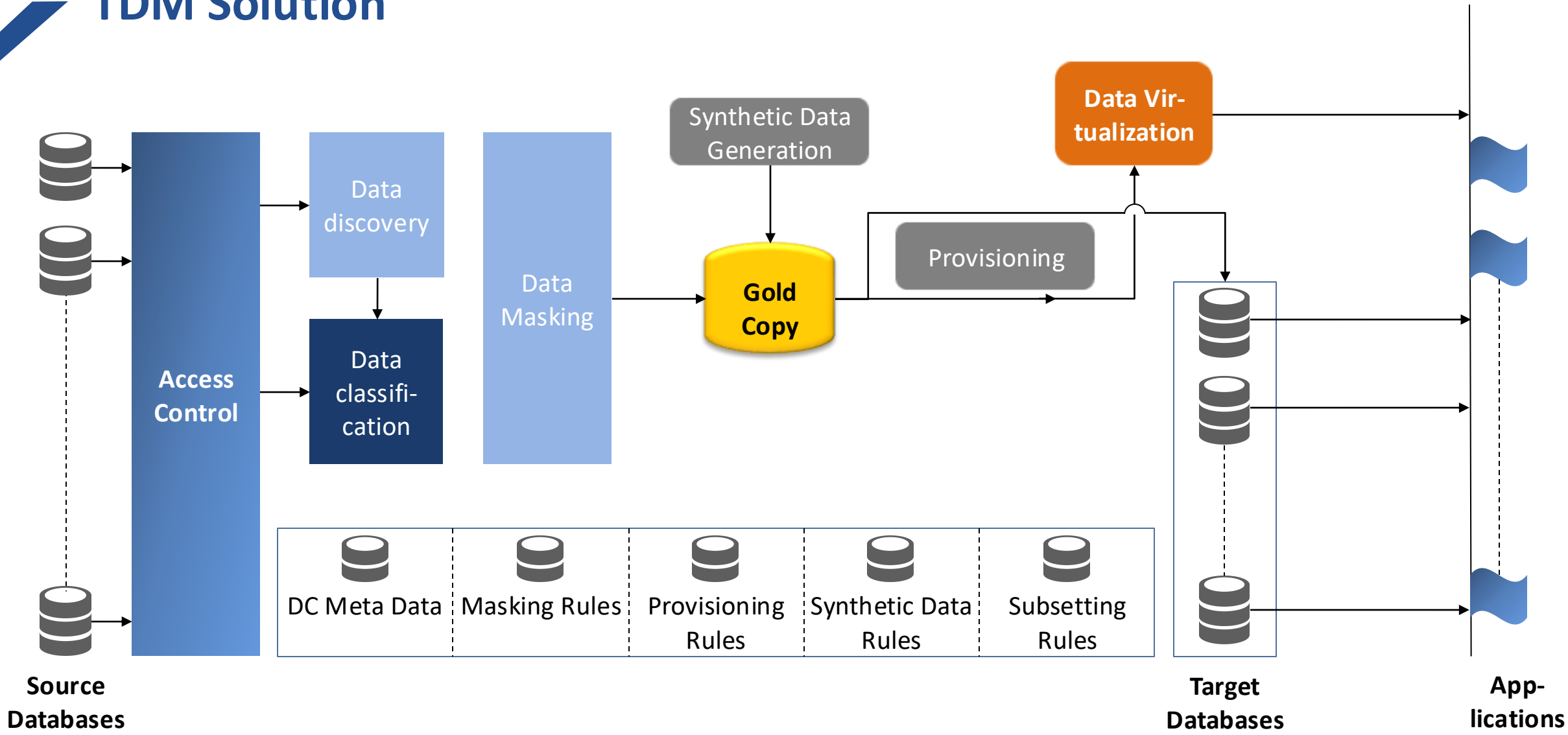
- Ability to deploy data by creating virtual databases for each test environment based on a single golden copy
- Ability to copy data to test databases on need basis (ex: performance testing)



Data Management

- Configurable data sub-set requirements
- Configurable data patterns to identify and classify sensitive data
- Configurable business rules and data requirements for each release
- Configurable data deployment requirements
- Web UI for better usability with built-in access control

TDM Solution



Application Data Modeling

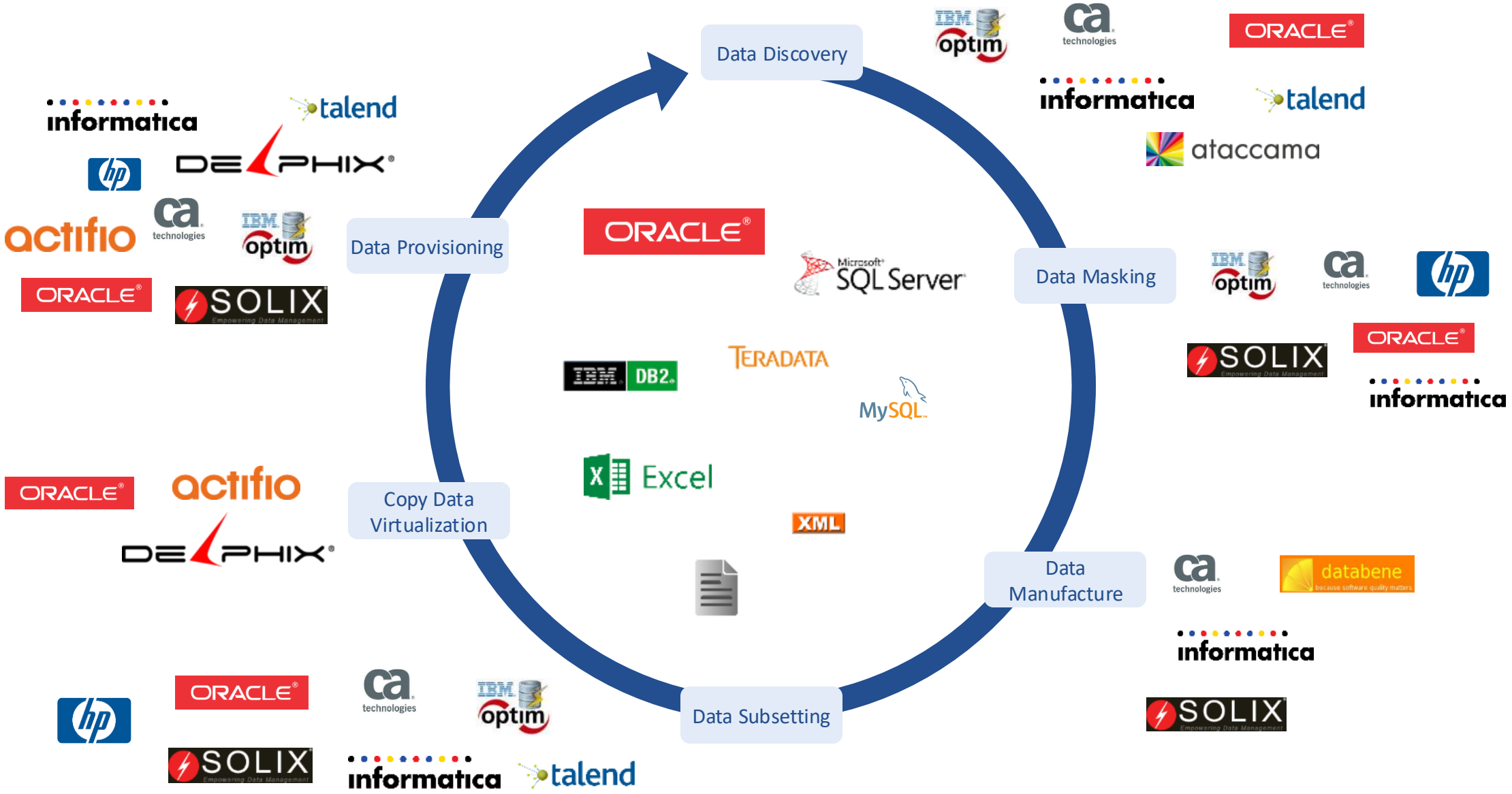
Single Data Source with known data model:

- Automated profiling process that involves the analysis of:
  - Data model
  - Referential integrity parameters
  - Table hierarchy
  - Completeness of data
  - Uniqueness of field values
  - Data Pattern and sensitivity analysis

Single Data Source without data model or **Multiple Data Sources**:

- **Data Discovery** is the process of automating:
  - Single source data profiling
  - Comparing the data set of each column or set of columns (from the outcome of data profiling) against other columns for Primary-foreign key discovery
  - Based on the identified data relationships, discovering business objects
  - Using the same approach for cross-source data overlap discovery
- **Data Lineage** is identifying data relationships by:
  - Parsing the 'Stored procedures' and the SQL queries of complete 'SQL history log'
  - This also provides table hierarchy
  - Limited to Single source only

Tools


Data Discovery
Data Masking
Data Manufacture
Data Subsetting
Copy Data Virtualization
Data Provisioning
ORACLE
Microsoft SQL Server
TERADATA
IBM DB2
MySQL
Excel
XML
informatica
talend
DELPHIX
actifio
ca technologies
IBM optim
SOLIX
hp
ataccama
databene

Case Study

UK based world-leading provider of outsourcing and technology solutions to the financial services industry needed a TDM solution to overcome their test data challenges

Requirement

- Optimize Test data provisioning as they have 30+ test environments
- Create Application Data Model as the data relationships/constraints are maintained at application level
- Effective data sub-setting for moving data to Dev team in Australia and for QA testing
- Mask sensitive Data
- Generate Test data to get the complete data set for testing

Solution

- Virtualize databases using Actifio to reduce the no. of test databases
- Reverse engineering the production database thru ‘Data Discovery’ and ‘Data Lineage’ to get the application data model
- Tool based data masking, data sub setting and data provisioning

Benefits

- Ability to provision data to test environments virtually (from a single copy of actual data) – Upto 70% savings in effort and storage
- Systematic test data provisioning – 25% effort savings for testers
- Estimated ROI of £340K at the end of 3 years with break-even at the 2nd year

Case Study

UK based Tier 1 bank needed to implement a TDM solution to overcome their test data challenges

Requirement

- Generate Test data needed by Digital Team to provision the complete data set required for end to end testing (70+ applications/ interfaces, heterogeneous databases with 3 million+ records)
- Reduce data provisioning time from 4 weeks to same day
- Create Application Data Model as the data relationships/constraints are maintained at application level
- Prevent data over stepping
- Mask sensitive Data

Solution

- Reverse engineering the production database thru ‘Data Discovery’ and ‘Data Lineage’ to get the application data model
- Data subset of 100,000 records representing the production data generated using CA TDM toolset
- CA ARD tool used to generate a dynamic GUI to enable testers to query the subset data and select the required datasets
- Lock test data selected by testers using CA TDM resolving data over stepping issues
- Tool based data masking, data sub setting and data provisioning

Benefits

- Data subset created within a single sprint. Scripts to refresh the Gold copy of the data subset periodically/on need basis
- Data provisioning to a new test environment takes less than 6 hours
- Resolved data over stepping issues

Corporate Headquarters
Lords Tower, Block 1,
2nd Floor, Plot No. 1&2 NP,
Jawaharlal Nehru Road,
Thiru Vi Ka Industrial Estate
Ekkaduthangal, Chennai – 600 032

www.maveric-systems.com

Global Locations

SINGAPORE | UK | US | DUBAI | RIYADH | MALAYSIA | MEXICO