

# IBM Optim Solutions with Designer Proof of Technology

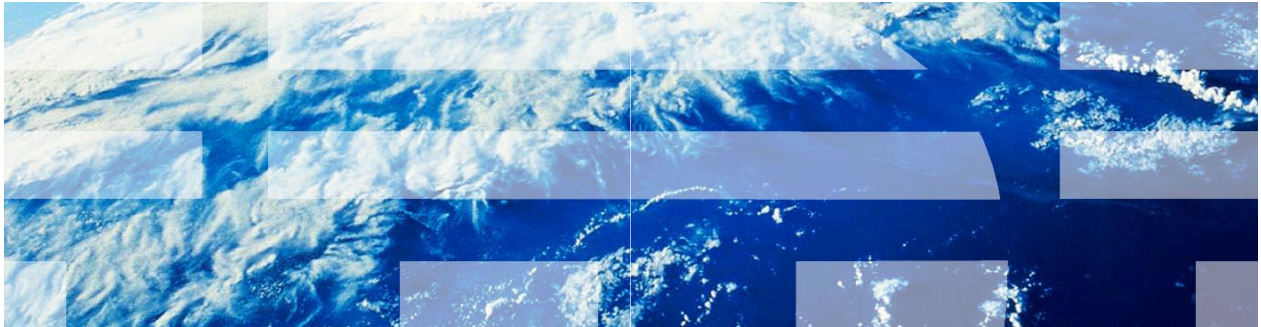
*Class Presentations*



© Copyright International Business Machines Corporation, 2008, 2012. All rights reserved.

US Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp

# Optim Solutions Proof of Technology



© 2012 IBM Corporation

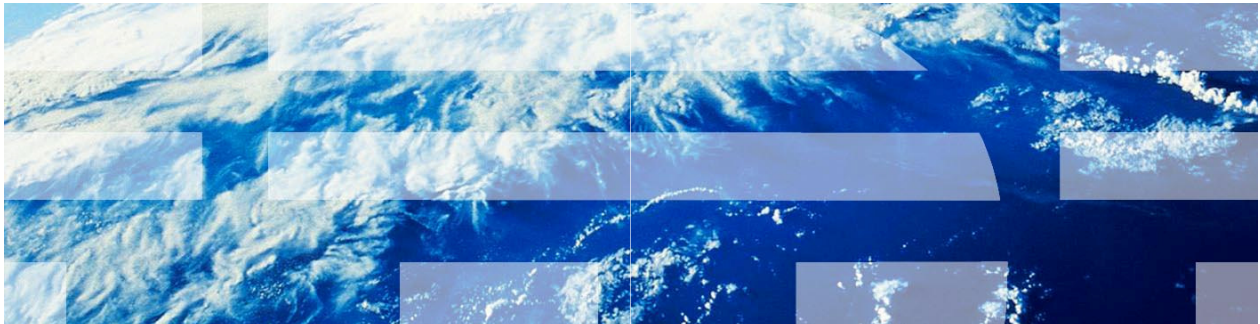
## Welcome to the Technical Exploration Center

- Introductions
- Restrooms
- Emergency Exits
- Smoking Policy
- Lunch - Special meal requirements?

## Agenda

- Data Governance Overview
- Optim Overview
- Optim Data Growth
  - Lab 1 – Discovering Optim Archive Capabilities
  - Lab 2 – Discovering Optim Delete and Restore Processing
  - Lab 3 – Adding Optim Relationships (Optional)
- Optim Test Data Management
  - Lab 4 – Discovering Optim Test Data Management Capabilities
- Optim Data Privacy
  - Lab 5 – Discovering Optim Data Privacy Capabilities
- InfoSphere Discovery
  - Lab 6 – InfoSphere Discovery Finding Relationships and Business Objects

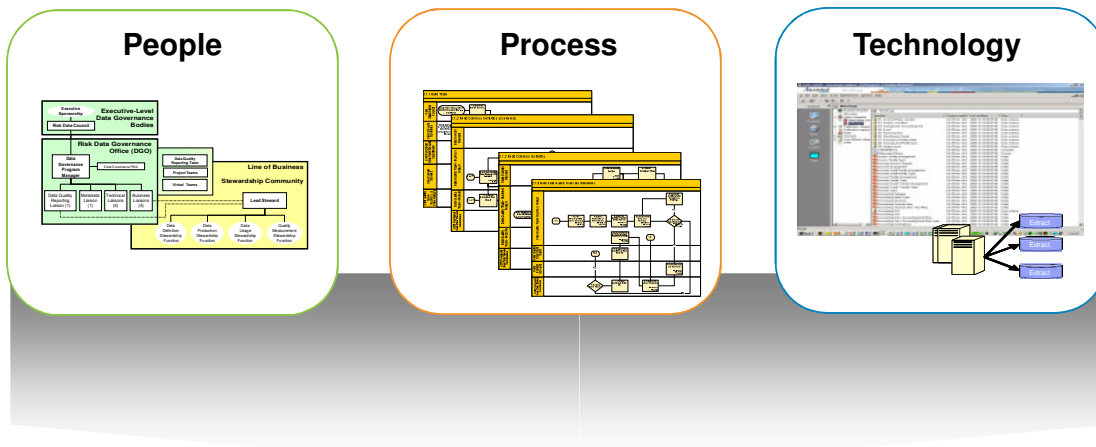
## Data Governance Overview





## It starts with Information Governance

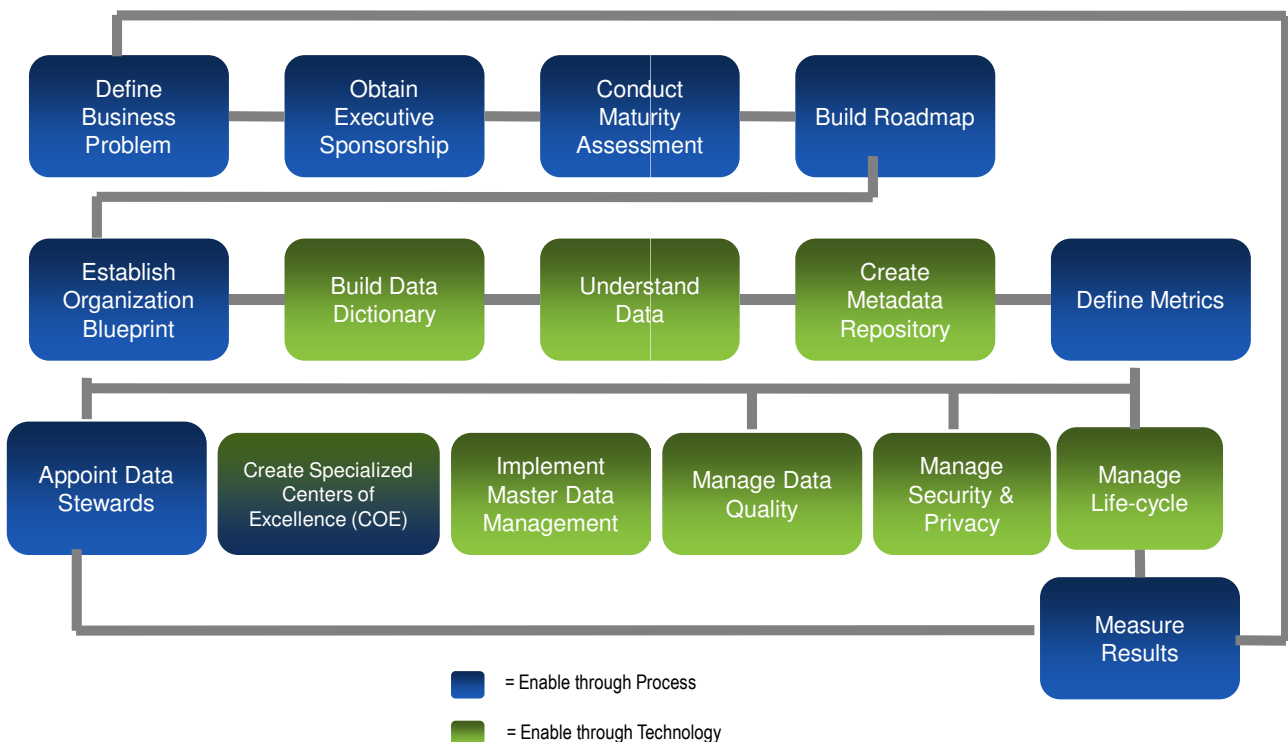
*Proactively leveraging information to unlock value and manage risk*



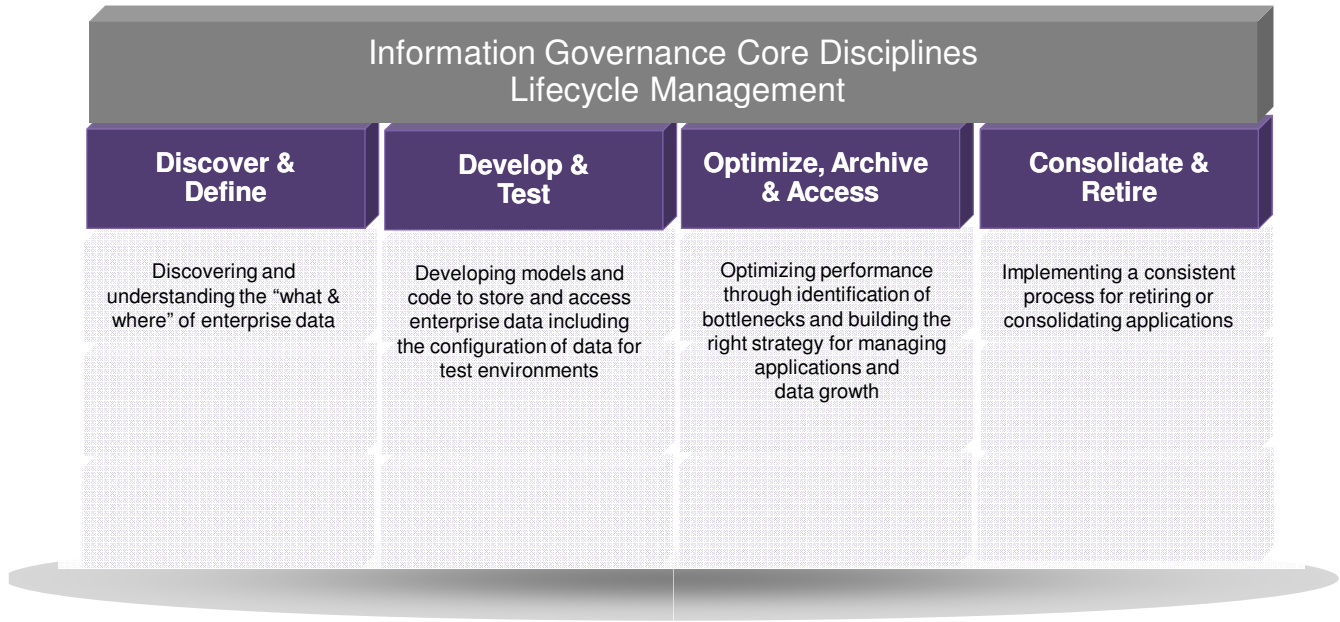
**Ensure information is understood and consistently defined.  
 Increase the use and trust of information as an enterprise asset.  
 Protect information, reduce risk and comply.**

## Enabling success

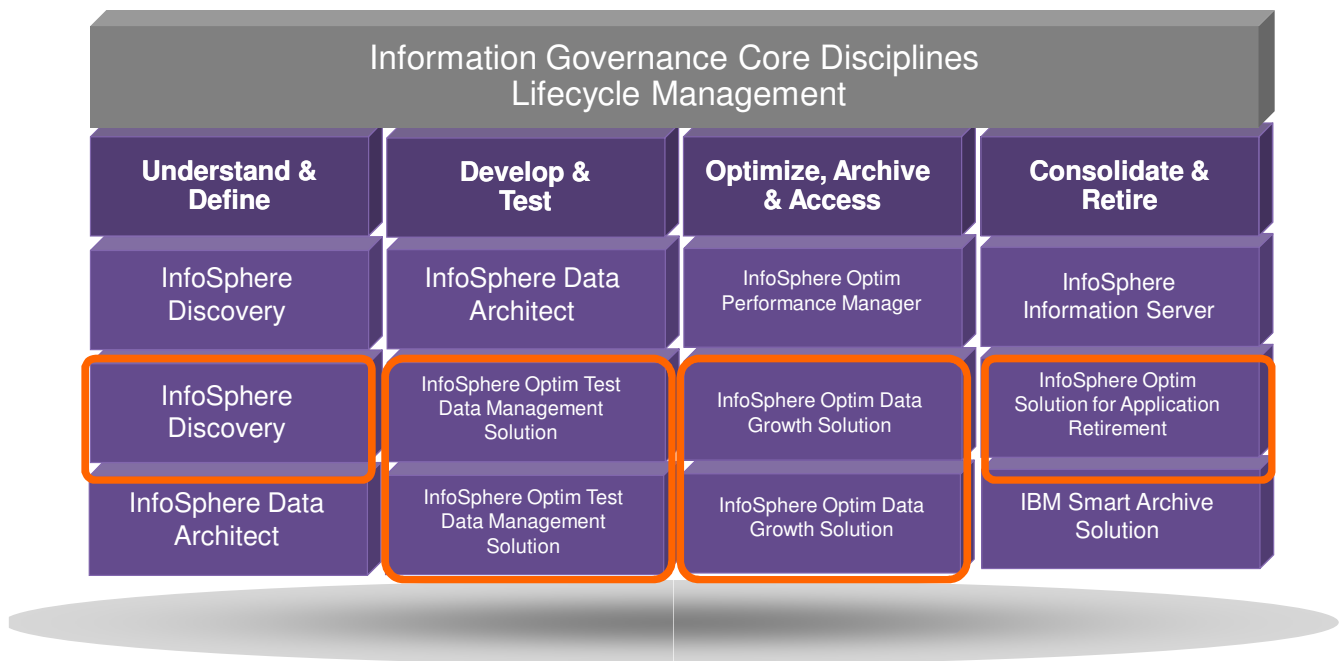
*IBM Information Governance Unified Process*



## Requirements for managing the data lifecycle



## IBM provides the solutions required to manage information throughout its lifecycle from requirement to retirement

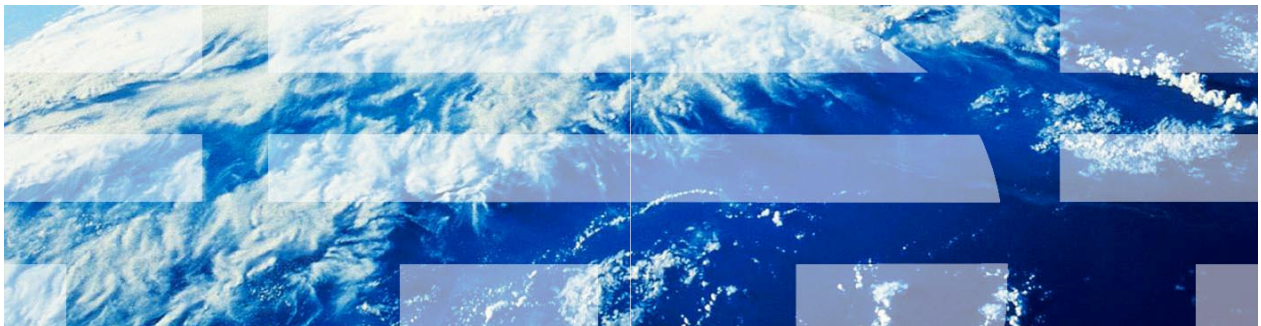


## How mature is your organization based on the core principles

- Self assessment
  - [www.infogovcommunity.com](http://www.infogovcommunity.com)
- IBM Governance Unified Process Book: (free e-Download)
  - [www.ibm.com/software/data/education/bookstore/](http://www.ibm.com/software/data/education/bookstore/)
- Free workshops and assessments: talk to your IBM representative



## Optim Overview



# The Challenges Around Data Management



## Application Redundancies

Application Portfolio Rationalization

- Sunset legacy applications
- Reduce operational IT spend



## Information Security

Protect and enable secure sharing of information

- 84% of security breaches come from internal sources, from non-production.



## Rampant Data Growth

Cost effectively support your information retention policies while controlling data growth.

- Meet SLA's
- Reduce operational IT spend

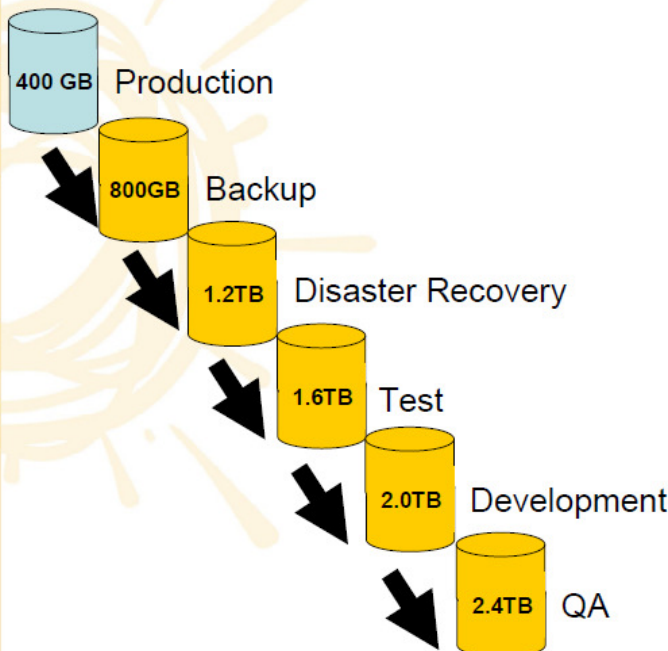


## Information Compliance

Reduce brand reputation risks and audit deficiencies

- 63% IT executives rate compliance with regulations a top challenge.

# The Ripple Effect of Data Growth



- Poor performance and response times.
- Access to decision-support information is more difficult
- Increasing costs: Server, storage and CPU capacity
- System Availability: Timely completion of batch cycle
- Slower disaster recovery. Restoring large volumes of rarely accessed data will slow recovery
- Complexity of environment: Large databases yield larger problems
- Compliance Exposures

Sun Life – Estimated 4x - Reality 7x to 10x

## Have you considered your test environments?

### Challenges in Managing Test Data



*"We did not want to create 'fake' or unrealistic test data. All test data had to be created and set up manually. So it would take us a month to setup test data for 30 or more accounts."*

-- Large US Healthcare Insurer



*"We needed to improve efficiencies in development and testing environments, as well as our production environments. We can create realistic test environments that use much less disk space than we would by cloning the production database."*

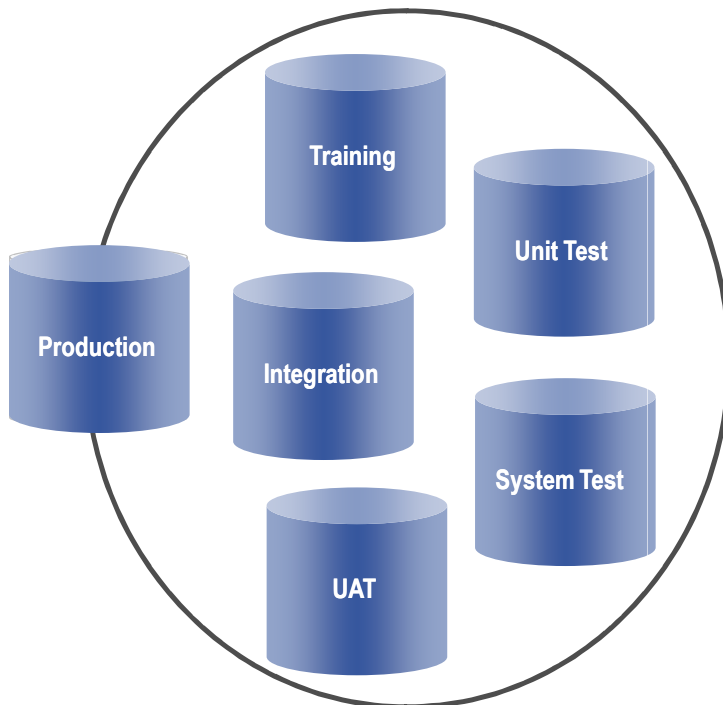
-- Allianz Seguros



*"Our staff wanted to implement more efficient and cost-effective testing processes that would shorten the time for creating and managing multiple test environments."*

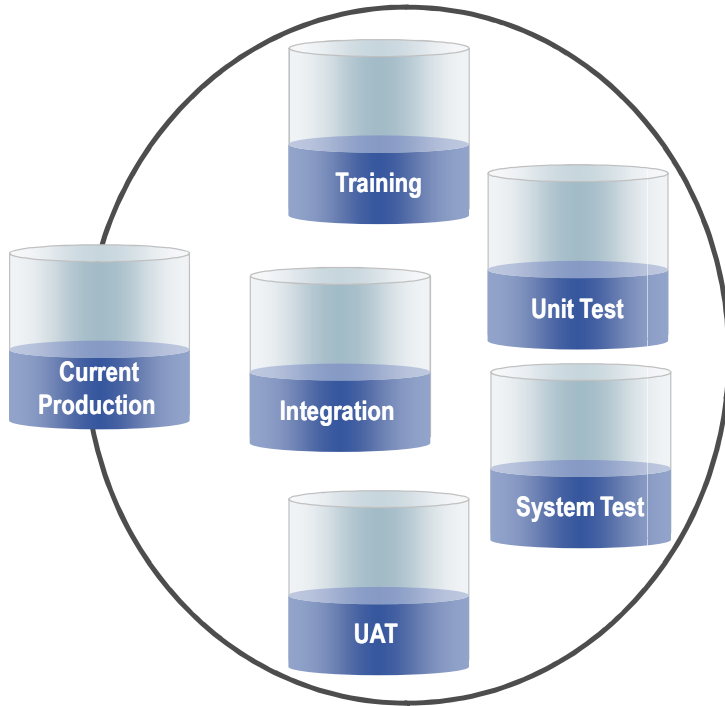
-- Cetelem

## Data Multiplier Example: Before Archiving



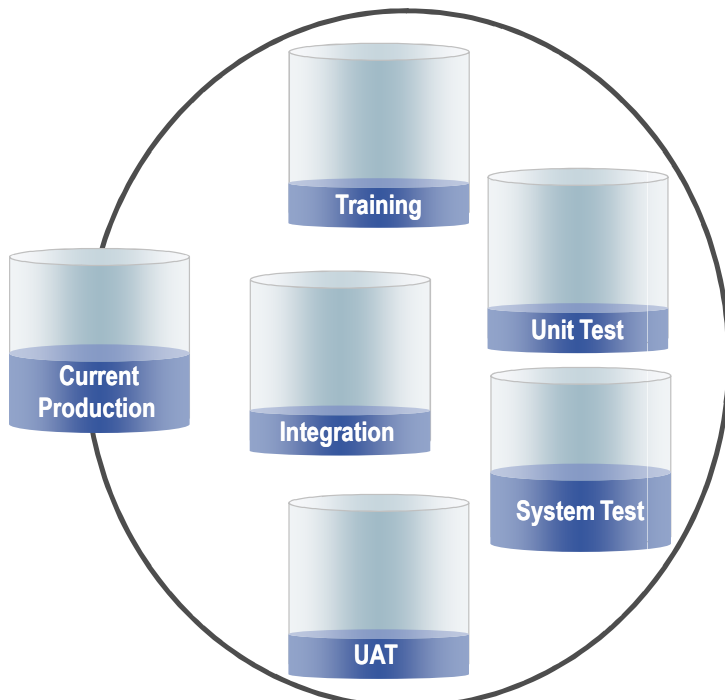
<b>Production</b>	<b>500 GB</b>
<b>Training</b>	<b>500 GB</b>
<b>Unit Test</b>	<b>500 GB</b>
<b>System Test</b>	<b>500 GB</b>
<b>UAT</b>	<b>500 GB</b>
<b>Integration</b>	<b>500 GB</b>
<hr/>	
<b>Total</b>	<b>3 TB</b>

## After Archiving: Cost and Storage Savings



<b>Production</b>	<b>200GB</b>
<b>Training</b>	<b>200GB</b>
<b>Unit Test</b>	<b>200GB</b>
<b>System Test</b>	<b>200GB</b>
<b>UAT</b>	<b>200GB</b>
<b>Integration</b>	<b>200GB</b>
<hr/>	
<b>Total</b>	<b>1.2 TB</b>
	<b>Storage reduced by 60%</b>

## Example: How to further mitigate cost with Test Data Management



<b>Production</b>	<b>200GB</b>
<b>Training</b>	<b>25GB</b>
<b>Unit Test</b>	<b>25GB</b>
<b>System Test</b>	<b>200GB</b>
<b>UAT</b>	<b>25GB</b>
<b>Integration</b>	<b>25GB</b>
<hr/>	
<b>Total</b>	<b>500GB</b>
	<b>Infrastructure reduced by 83%</b>

**Creating right-sized targeted test environments saves storage costs & speeds testing**



## The Latest Research on Test Data Usage

- **62%** of companies surveyed use actual customer data instead of disguised data to test applications during the development process
  - 50%** of respondents have no way of knowing if the data used in testing had been compromised.
- **52%** of respondents outsourced application testing
  - 49%** shared live data!
- **26%** of respondents said they did not know who was responsible for securing test data

*Source: The Ponemon Institute. The Insecurity of Test Data: The Unseen Crisis*



## The Easiest Way to Expose Private Data...

*Internally with the Test Environment*

- **70%** of data breaches occur internally (Gartner)
- Test environments use personally identifiable data
- Standard Non-Disclosure Agreements may not deter a disgruntled employee
- What about test data stored on laptops?
- What about test data sent to outsourced/overseas consultants?
- How about Healthcare/Marketing Analysis of data?



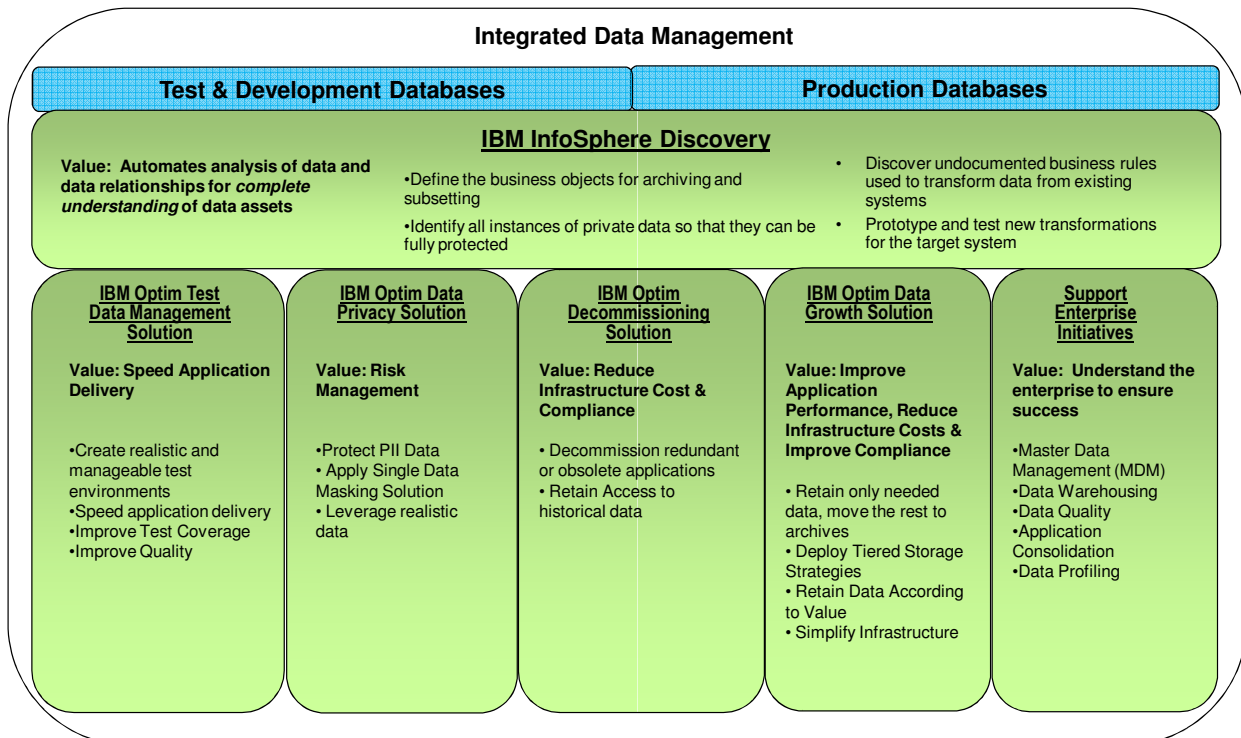
## More costs: What about those old and unused applications on 'life support'?

- Application portfolio has redundant systems acquired via mergers and acquisitions
- Line of business divested; application is no longer needed
- Legacy technologies not compatible with current IT direction
  - Old database and/or application versions no longer supported by manufacturer
- Required technical skills or application knowledge no longer available
- Budget pressures – do more with less



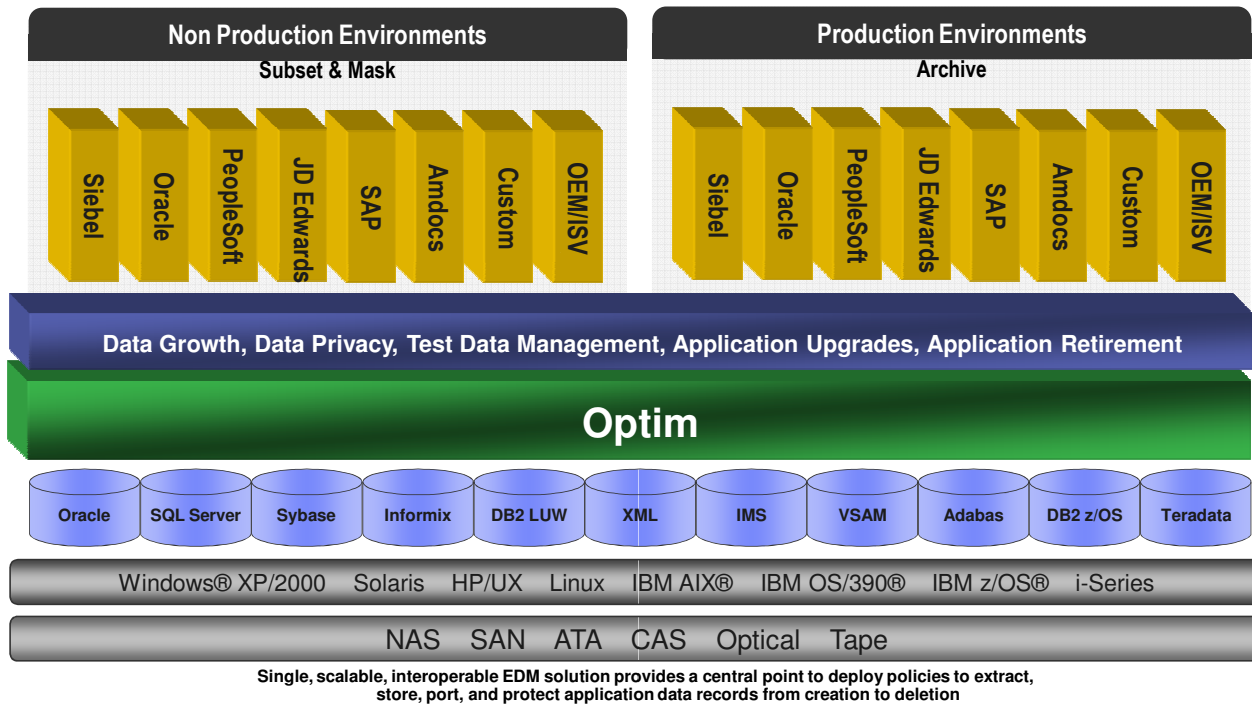
***In almost ALL cases, access to legacy data MUST be retained while the application and database are eliminated***

## Solution – IBM Optim Integrated Data Management Platform

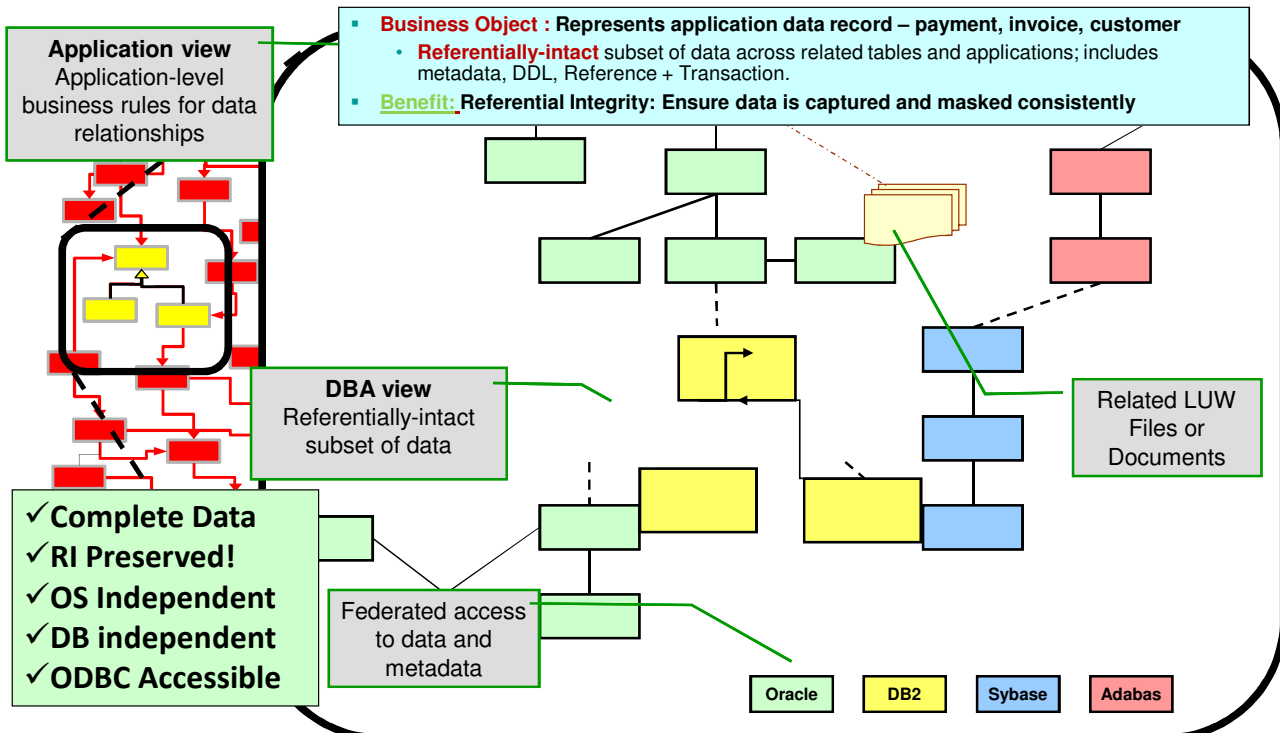




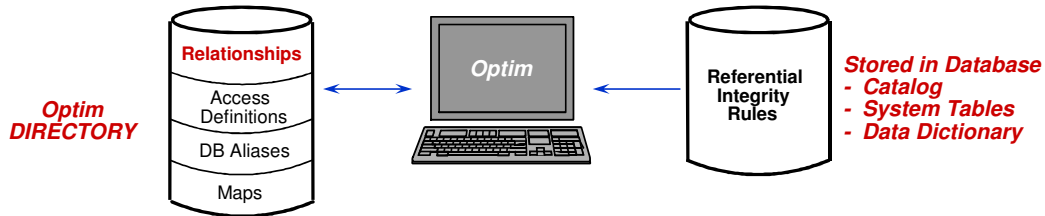
# Enterprise Architecture



# Optim Represents a Complete Business Object



## Database and Application Relationships

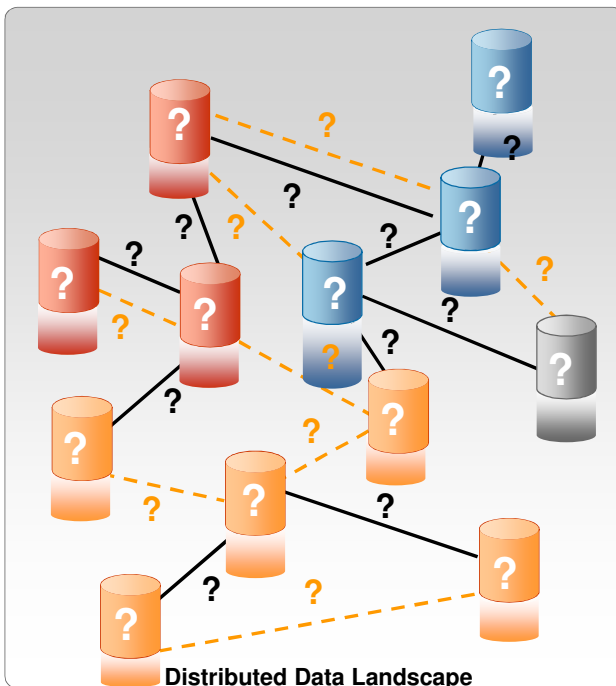


1. Database RI relationships are automatically imported
2. Application RI relationships - Known
  - Can be entered into Optim
  - Can be imported into Optim (DDL, templates)
3. Application RI relationships - Unknown
  - Use Discovery to profile data and find RI relationships

Once defined, relationships shared by all Optim components

## Where to start?

*You can't manage what you don't understand...*

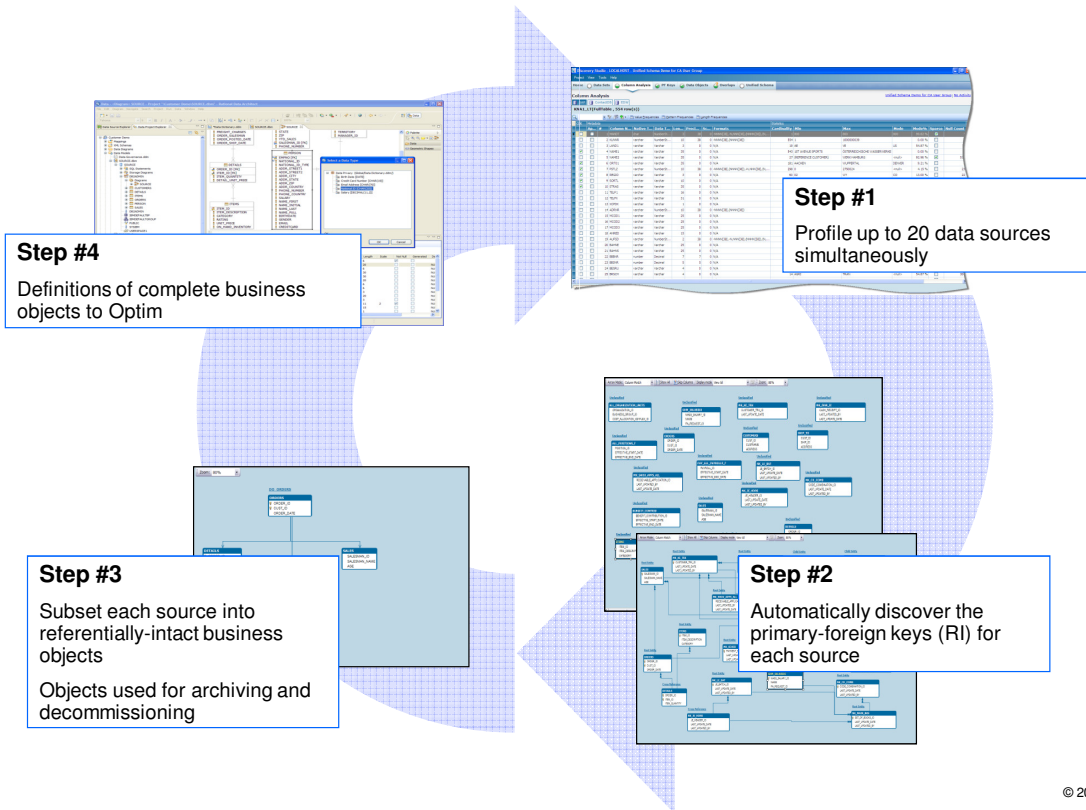


- Highly distributed over multiple applications, databases and platforms
- Complex, poorly documented data relationships
  - Which clients are eligible for the new sales promotion?
  - Which version of the data should we use for the ERP consolidation?
- Relationships not understood because:
  - Corporate memory is poor
  - Documentation is poor or nonexistent
  - Logical relationships (enforced through application logic or business rules) are **hidden**

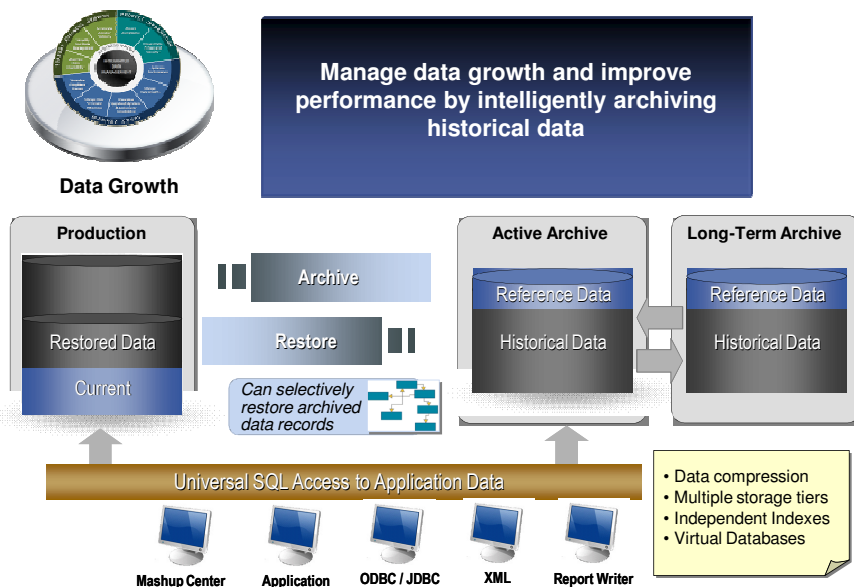


# How does it work?

Discover Complete Business Objects by Analyzing Data



# IBM InfoSphere Optim Data Growth Solution



## Requirements

- Archive, manage and retain application data according to business policies
- Minimize downtime during application upgrades
- Consolidate application portfolio and retire legacy applications
- Enable compliance retention and defensibility

## Benefits

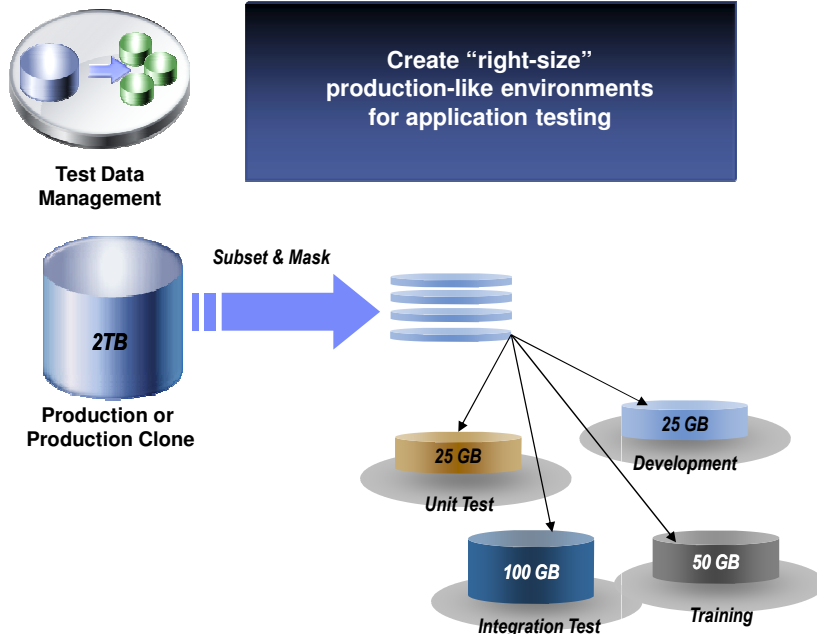
- Reduce hardware, storage and maintenance costs
- Streamline application upgrades and improve application performance
- Safely retire legacy & redundant applications while retaining the data

InfoSphere Optim Data Growth supports data on distributed platforms (LUW) and z/OS.  
Out-of-the-box archiving support for packaged applications available for ERP/CRM solutions:



# Optim Test Data Management & Data Privacy Solution

Manage your Test Database Environments



Create "right-size" production-like environments for application testing

**Requirements**

- Create referentially intact, "right-sized" privatized test databases
- Automate test result comparisons to identify hidden test errors and confirm privatization
- Shorten iterative testing cycles and accelerate time to market
- Mask Sensitive Data

**Benefits**

- Deploy new functionality more quickly and with improved quality
- Easily refresh & maintain test environments
- Reduce storage and operational costs

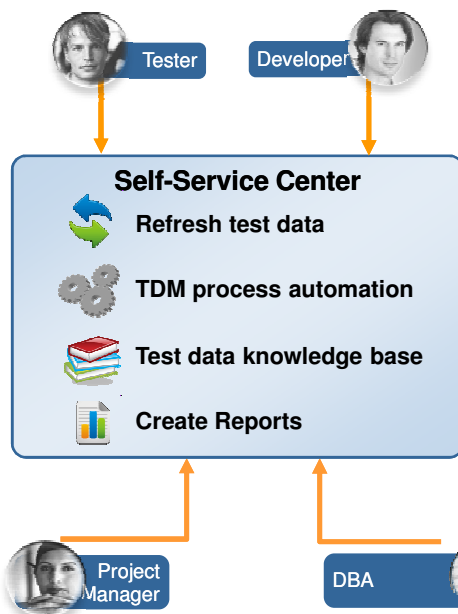
InfoSphere Optim TDM supports data on distributed platforms (LUW) and z/OS.

Out-of-the-box subset support for packaged applications ERP/CRM solutions as well as :



# Optim Self Service Center for Test Data Management

Streamline test data delivery



**Reduced software development time and increased operational efficiency**

**Improved predictability and repeatability of testing efforts**

**Improved visibility into TDM process**

# IBM InfoSphere Guardium Data Redaction

(Separate Product – Not Covered Today)



Data Redaction

Protect sensitive unstructured data in documents and forms

Finresearch LLC  
934 Fifth Ave  
New York, NY 00124

September 19, 2008

James McDonald CEO  
Financial National Bank  
111 Massachusetts Ave  
Boston MA 02140

Re: Preliminary Anti-Trust Pre-Acquisition Investigation

Finresearch LLC has conducted research of the market and legal situation in preparation for an acquisition of Northern Investments Inc. by Financial National Bank Inc., scheduled for Jan. 21, 2009. The assignment was to determine the risk of civil and/or criminal action from the Attorney General of the United States under Section 15 of the Lombard Act, 15 U.S.C. § 19 to enjoin the acquisition of Northern Investments. We were asked to assess if such an acquisition would substantially affect competition in the housing

Before

[Organization]  
[Address]  
[Address]  
[Date]

[Person] [Org.]  
[Organization]  
[Address]  
[Address]

Re: [Organization] Pre-Acquisition Investigation

[Organization] has conducted research of the market and legal situation in preparation for an acquisition of [Organization] [Person] by [Organization], scheduled for [Date]. The assignment was to determine the risk of civil and/or criminal action from the Attorney General of the [Location] under Section 15 of the Lombard Act, 15 U.S.C. § 19 to enjoin the acquisition of Northern Investments. We were asked to assess if such an acquisition would substantially affect competition in the housing

After

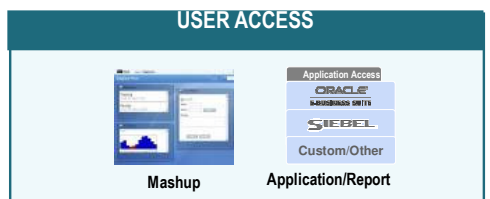
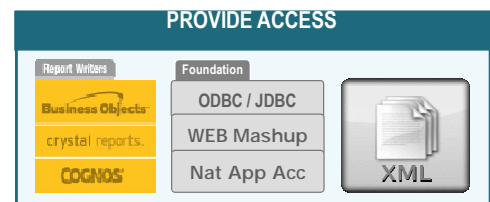
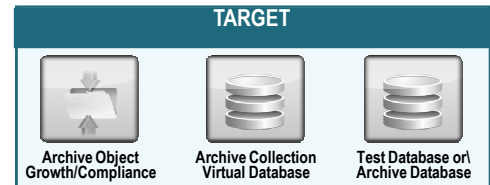
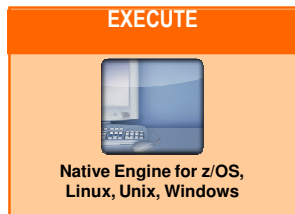
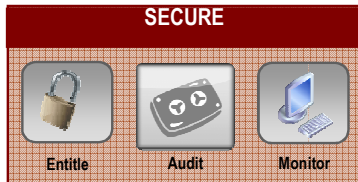
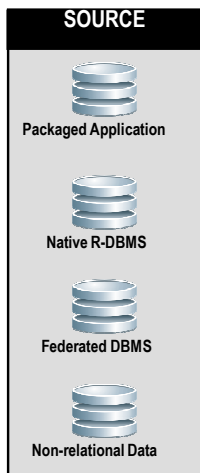
### Requirements

- Protect unstructured data in textual, graphical and form based documents
- Control data views with user role policies
- Automate batch workflow process with optional human review

### Benefits

- Prevent unintentional data disclosure
- Comply with regulatory and corporate compliance standards
- Increase efficiency and reduce risk via automation

# Optim – One Engine for Structured Data

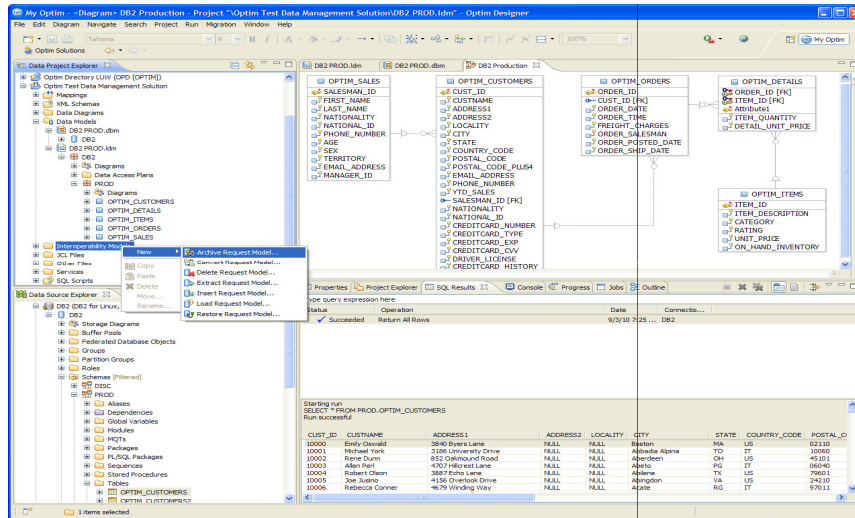




# Optim Designer



**Design, Build, Test and Validate Optim Services**



## Requirements

- Design, build, test and validate Optim Test Data Management, Data Growth and Data Privacy services
- Integrate with other IBM Optim and InfoSphere solutions
- Leverage existing Optim assets

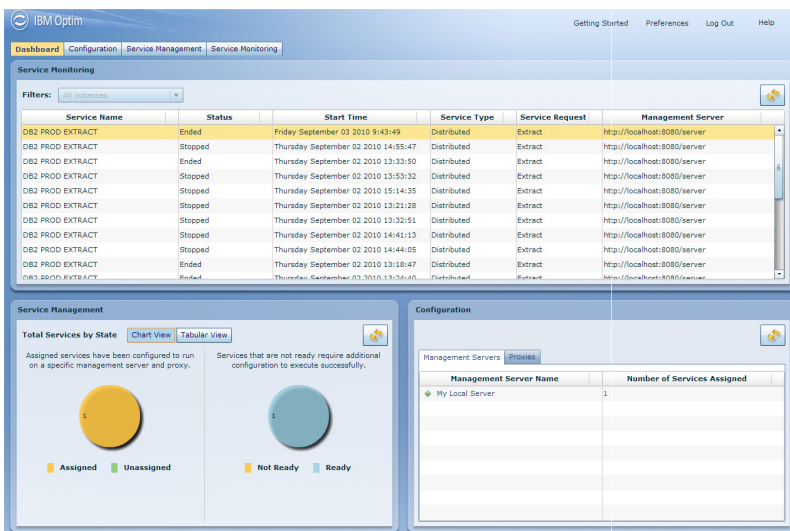
## Benefits

- A common design time interface for all Optim solutions
- Independent of Optim run time environment; supports Optim distributed and z/OS requests and services

# Optim Manager



**Deploy, Manage and Monitor Optim Services**



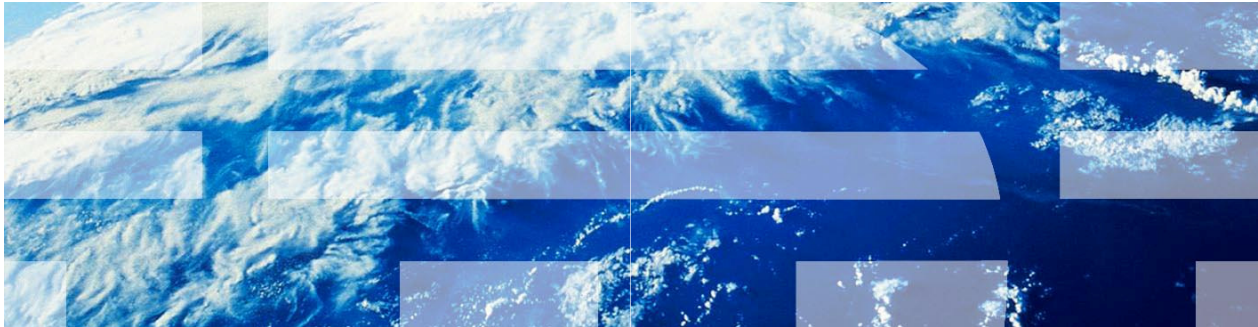
## Requirements

- Manage the execution of Optim services independent of their runtime environment
- Validate results from a common interface
- Access and report on operational metadata

## Benefits

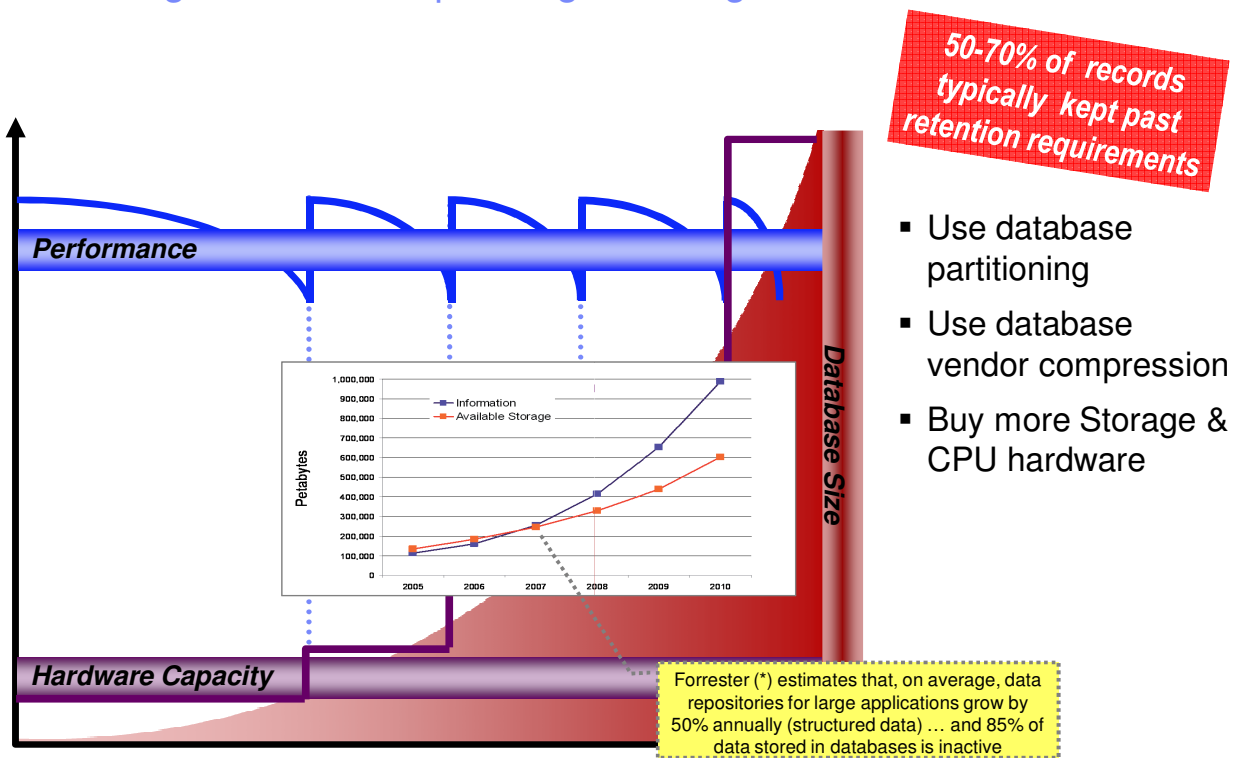
- A common run time interface for all Optim solutions
- Manage Optim services from anywhere at anytime
- Separation of design and run time roles and responsibilities

# Optim Data Growth



© 2012 IBM Corporation

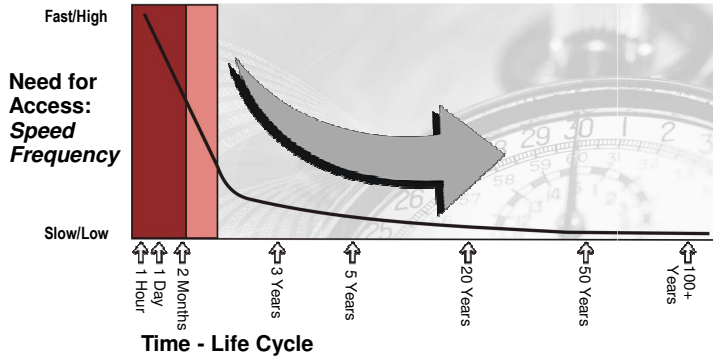
## How are organizations responding to data growth?



\* Source: Noel Yuhanna, Forrester Research, Database Archiving Remains An Important Part Of Enterprise DBMS Strategy

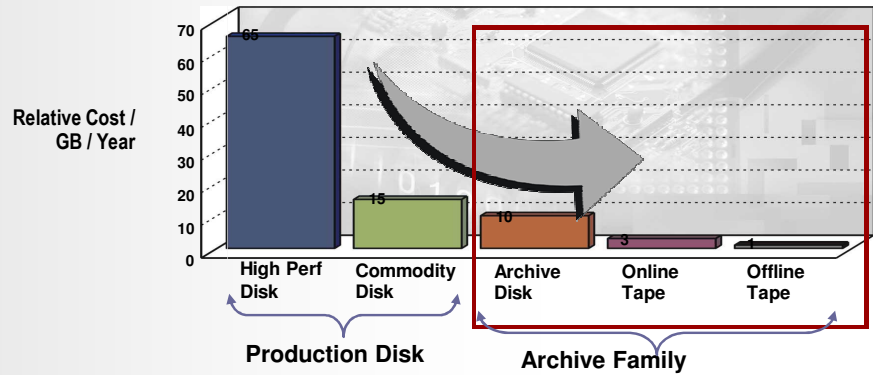


## Why consider an archiving and retention solution?



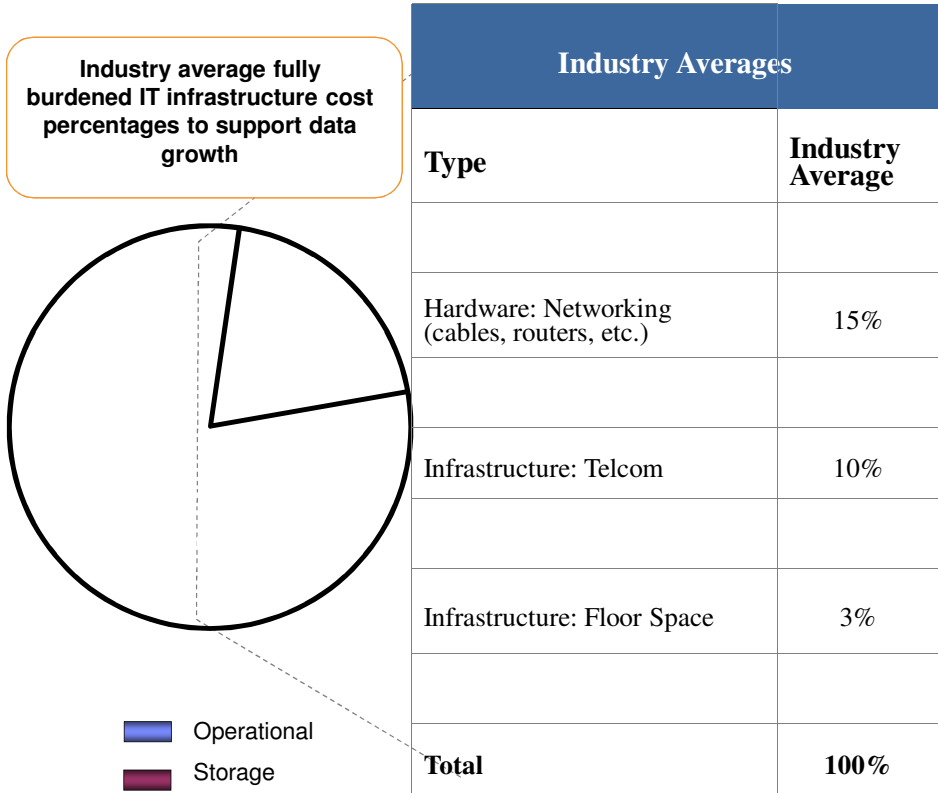
### The Time Value of Information

### Fully Managed Costs: Storage Options



© 2012 IBM Corporation

## The results can have significant financial impact



**For every 20% that is spent on storage, 80% cost is spent on the operational elements of managing that stored information.**

## What about applications that are no longer needed?

- Application portfolio has redundant systems acquired via mergers and acquisitions
- Line of business divested; application is no longer needed
- Legacy technologies not compatible with current IT direction or supported by vendor
- Required technical skills or application knowledge no longer available
- Budget pressures – do more with less

**In almost ALL cases, access to legacy data MUST be retained while the application and database are eliminated**



## Gartner Selection Criteria

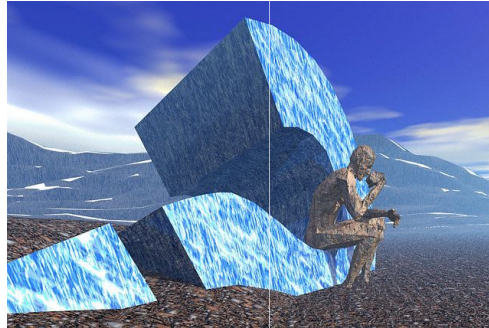
*Database Archiving Solutions “Must Haves”*

- Archive to an alternate (nonproduction) database
- Archive to file format
- Maintain referential integrity (even for the most complex data models)
- Seamless access to archived data from original application and/or via alternate methods (search, reporting)
- Security, access control and audit logs

For additional selection criteria, download this report at : [https://www14.software.ibm.com/webapp/iwm/web/signup.do?lang=en\\_US&source=sw-infomgt&S\\_PKG=optim\\_archiving\\_criteria\\_analyst](https://www14.software.ibm.com/webapp/iwm/web/signup.do?lang=en_US&source=sw-infomgt&S_PKG=optim_archiving_criteria_analyst)

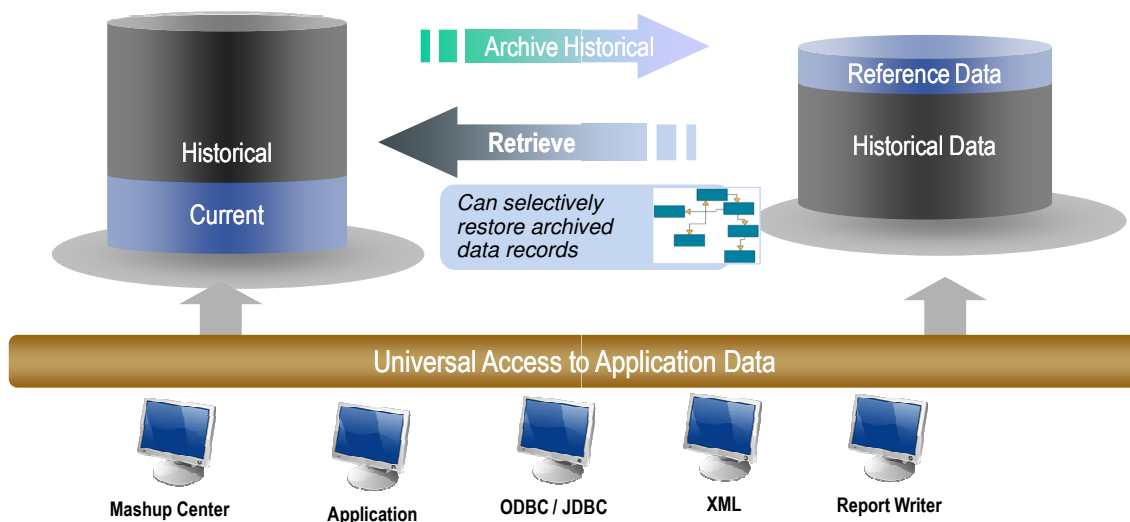
Source: Gartner, Inc., “Selection Criteria for Database Archiving Products”, S. Childs, February 2011

## What are the Drivers for Implementing Data Archiving?



- 1) Control Costs
- 2) Improve Performance
- 3) Mitigate Risks

## What is Active Archiving?

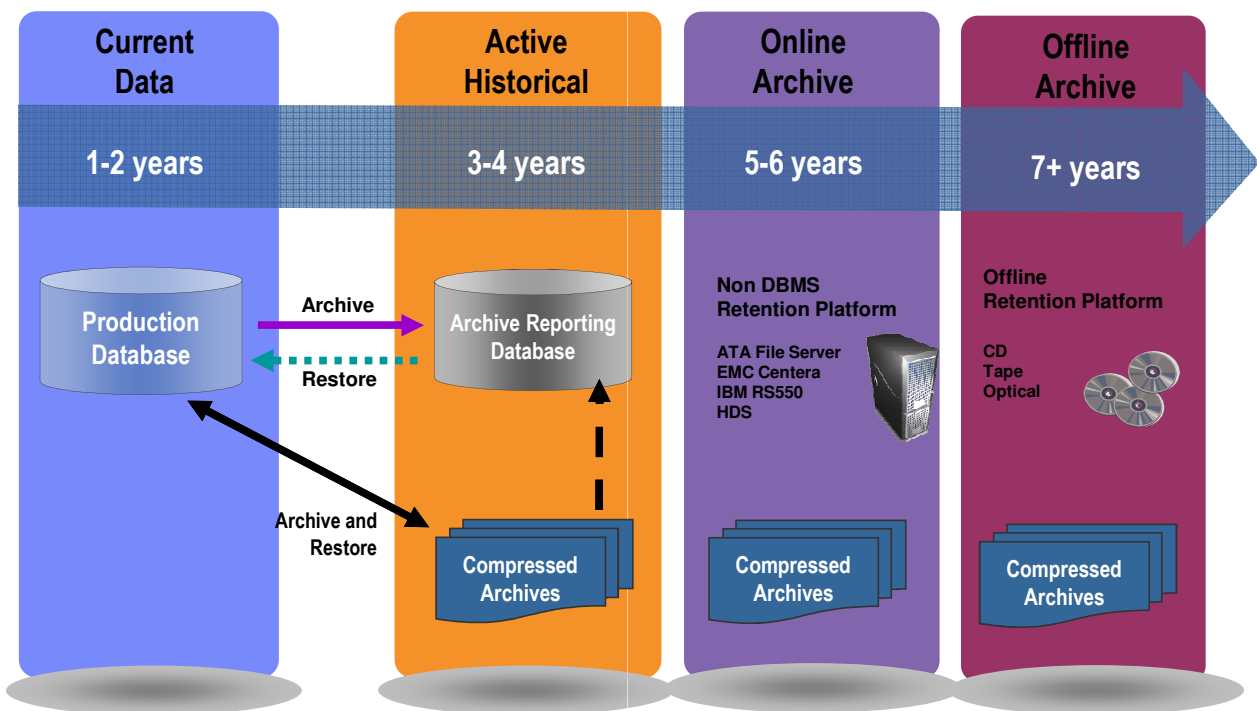


Active Data Archiving is an intelligent process for moving inactive or infrequently accessed data that still has value, while providing the ability to search and retrieve the data.

## Optim Archive Files Characteristics

- Application not required
- DBMS not required
- Complete business object
- Compressed (up to 90%)
- Indexed
- Secured
- Any storage device
- Access via any SQL Based tool
- Frozen in time
  - Immutable
  - Auditable
  - Snapshot of historical events
- “Future-proofed”

## Leverage cost-effective storage alternatives

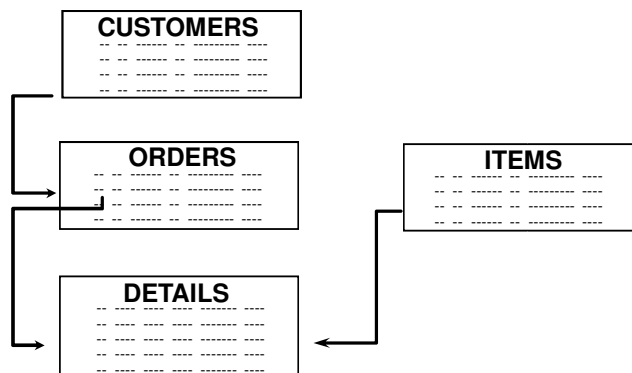


## Steps for Archiving Data

1. Identify the data to be archived
2. Define the data to be deleted
3. Create the archive
4. Review the validity of the archive
5. Delete the data
6. Find Data in the Archives
7. Browse, Report or Restore Fully or Selectively

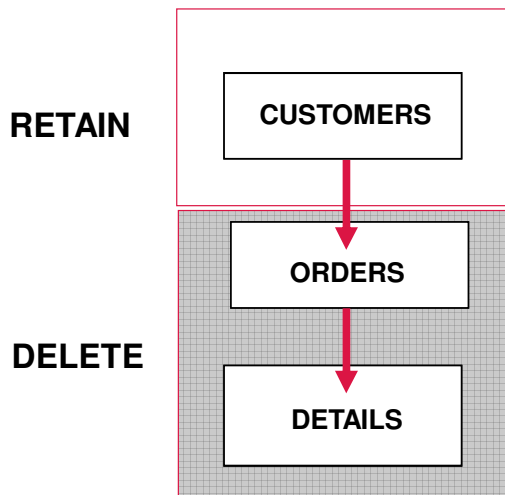
## Identify the data to be archived

### Data Access Plan/Access Definition Defines a subset of relational data- the Complete Business Object



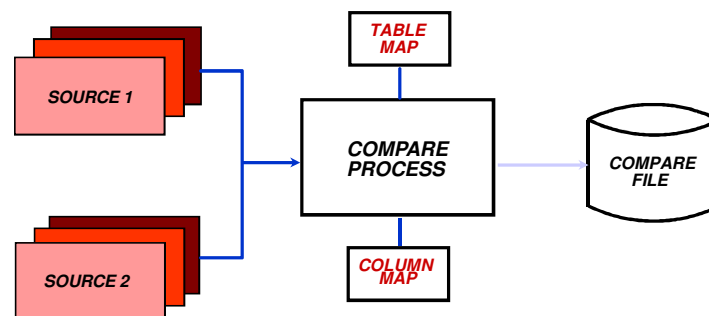
- Start table
- Associated data
- Relationships
- Extraction rules
- Index specifications
- Archive Actions
- Attachments

## Define the data to be deleted



- Archive all data
- Delete orders and details after they are safely archived
- Preserve semantic intelligence

## Optim Compare Feature



- Single-table or multi-table compare
- Creates compare file of results
- Displays results on screen
- For archiving validation, application testing and QA
- Enhances productivity by finding unexpected changes in the data

## Browsing the Compare File

**Browse Compare File Table Data**

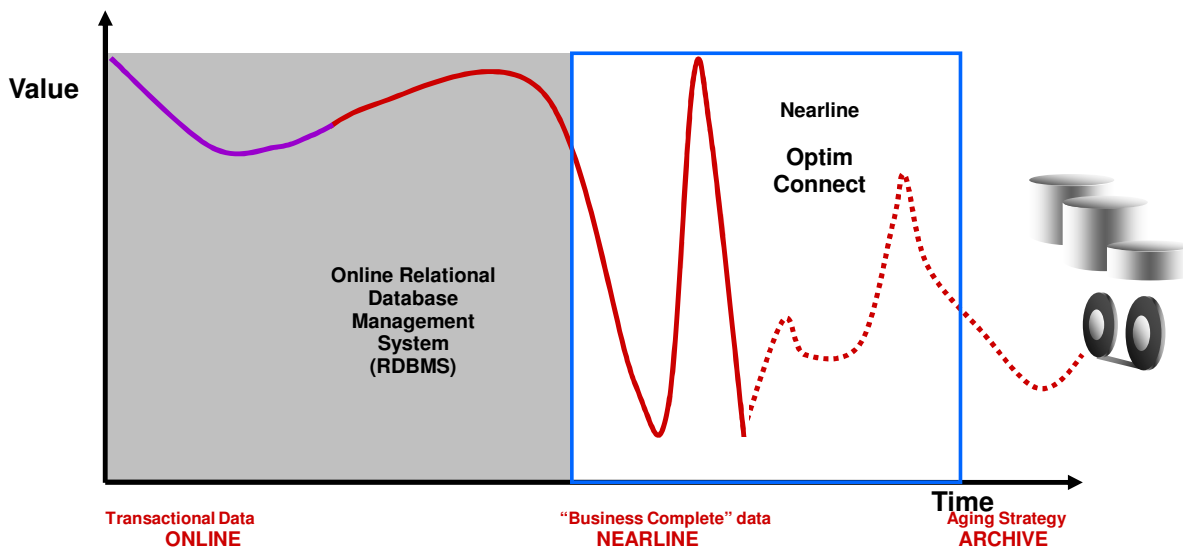
File Tools Options Help

Source 1: ORACLE8.LYNNP.CUSTOMERS

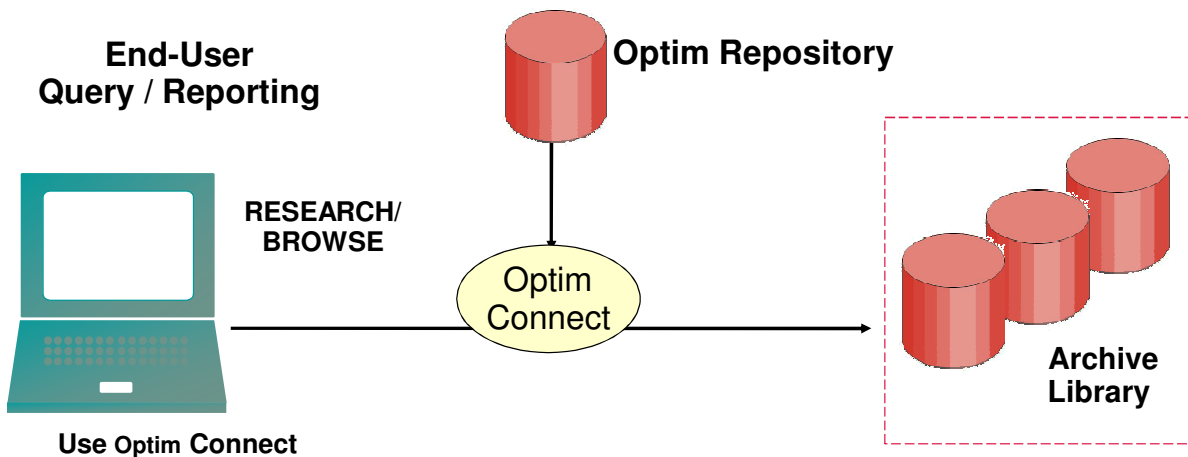
	Change	Source	CUST_ID CHAR(5)	CUSTNAME CHAR(20)	ADDRESS VARCHAR2(50)	CITY VARCHAR2(15)	STATE CHAR(2)	ZIP CHAR(5):N	Y
1	Only	1	00001	Audio-Video	593 West 37th Str	Brass Castle	NJ	10017	
2	Equal	Both	00002	Select-A-Vi	5720 MacArthur D	Evening Shade	AR	62700	
3	Equal	Both	00003	Showplace	1 Ocean Parkway	Alto	NM	11694	
4	Equal	Both	00004	Audio-Video	593 West 37th Str	Panacea	FL	10017	
5	Equal	Both	00005	Take Home	Box 357	Fence Lake	NM	90028	
6	Equal	Both	00006	Main Street	Gateway Shoppin	Pumpkin Center	AZ	85002	
7	Diff	1	00007	Cinematic	Pass-a-Grille Bea	<b>Pass-a-Grille</b>	FL	92120	
8	Diff	2	00007	Cinematic	Pass-a-Grille Bea	<b>Stop-at-Grille</b>	FL	92120	
9	Equal	Both	00008	Director's C	347 Miners Row	Spuds	FL	95800	
10	Equal	Both	00009	Prime Time	64 Newberg Ave	Loving	NM	22180	
11	Diff	1	00010	Reely Great	590 Frontage Rd	Christmas Vally	OR	<b>07002</b>	
12	Diff	2	00010	Reely Great	590 Frontage Rd	Christmas Vally	OR	<b>91002</b>	

- Change column identifies the type of change
- Source column identifies input source row
- Data differences are highlighted

## Information Life Cycle Management (ILM)



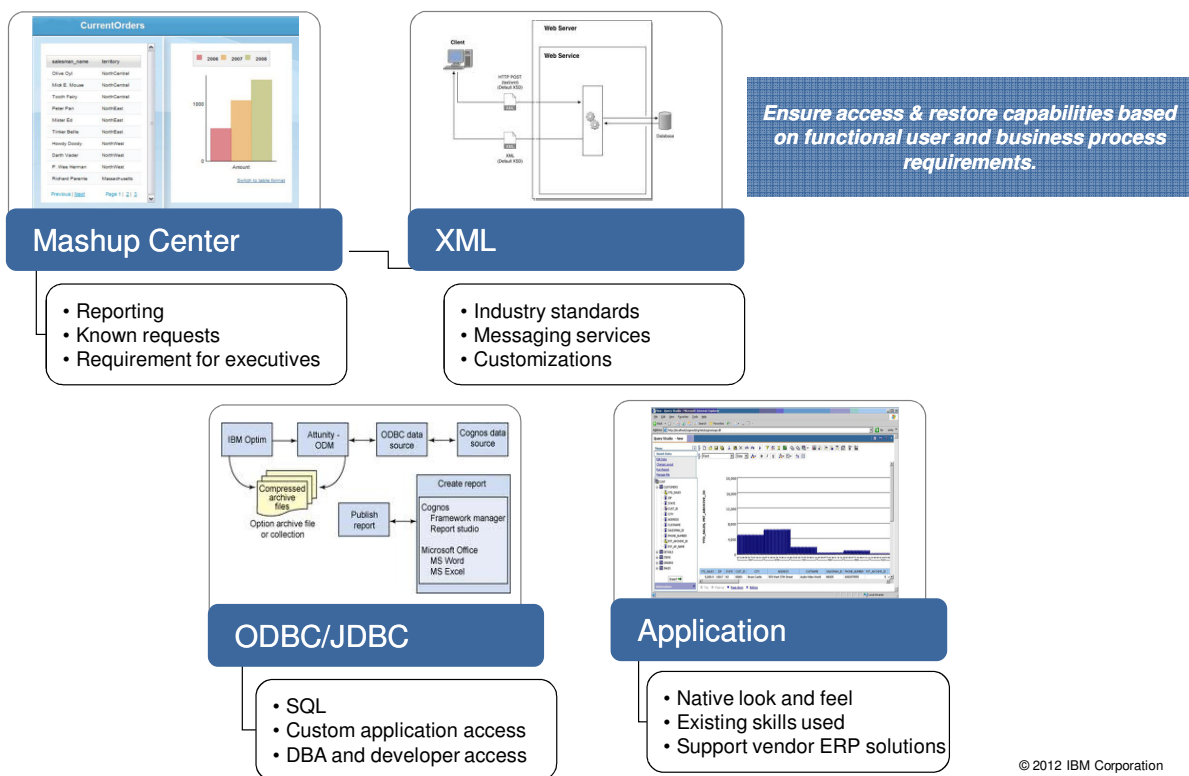
## Accessing the Archive Files



### Use Optim Connect

- ✓ Direct Access within Your Application using standard SQL
- ✓ Defines data-sources for any ODBC or JDBC application
- ✓ Archive Collections
- ✓ Joins between multiple data-sources
- ✓ Archive files and database tables

## Flexible access methods for archived data

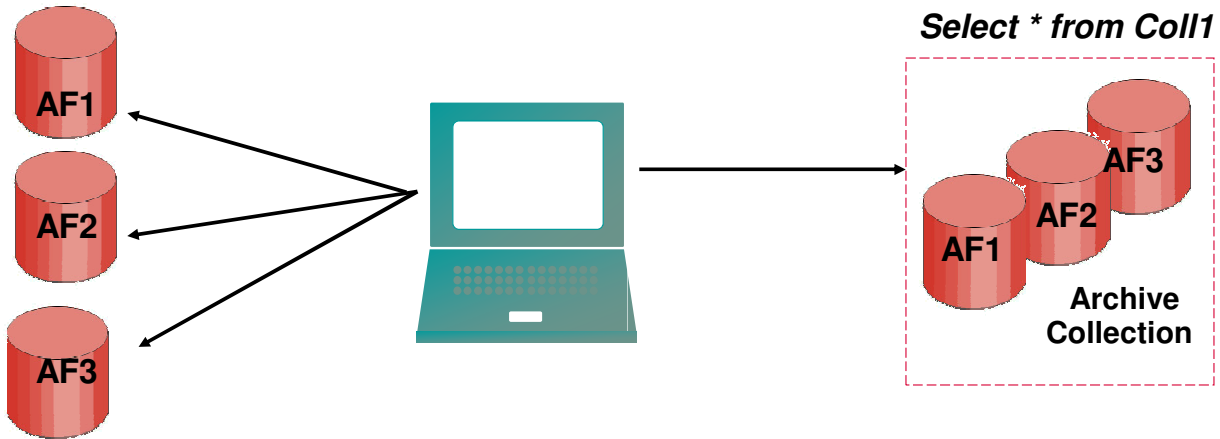




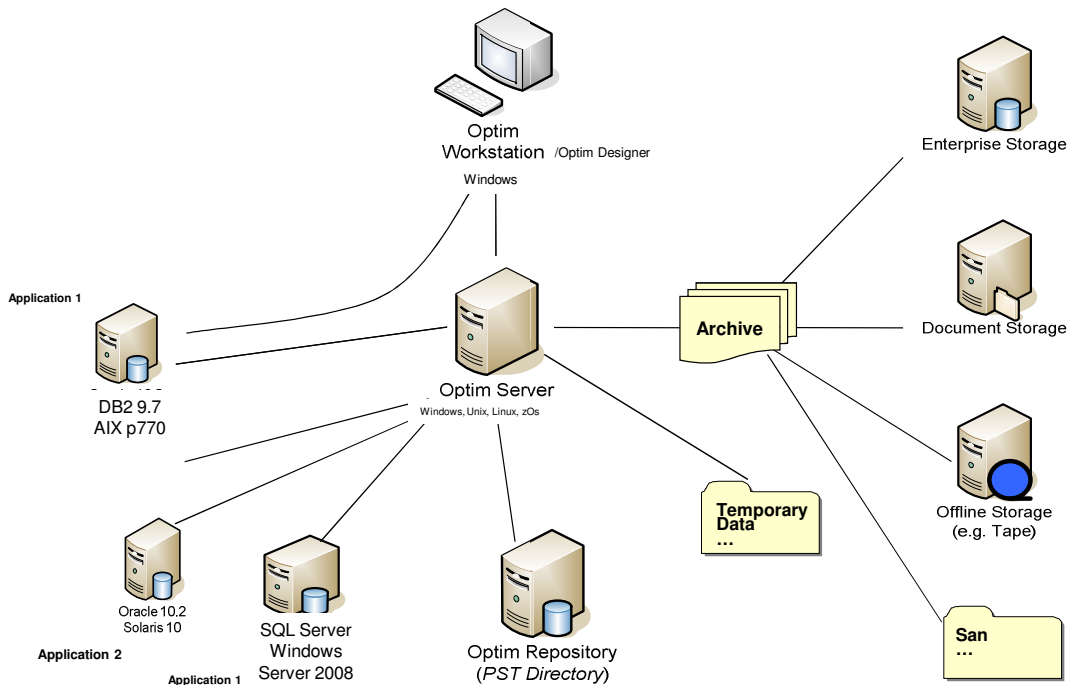
## Archive File Collections

- Manage Archive Files as a single data source using **Optim Connect**
- Create an Archive File Collection that logically references data in multiple Archive Files.
- Join tables in separate Archive Files contained in an Archive File Collection.

**Select \* from AF1,AF2,AF3**



## Conceptual Data Growth Architecture

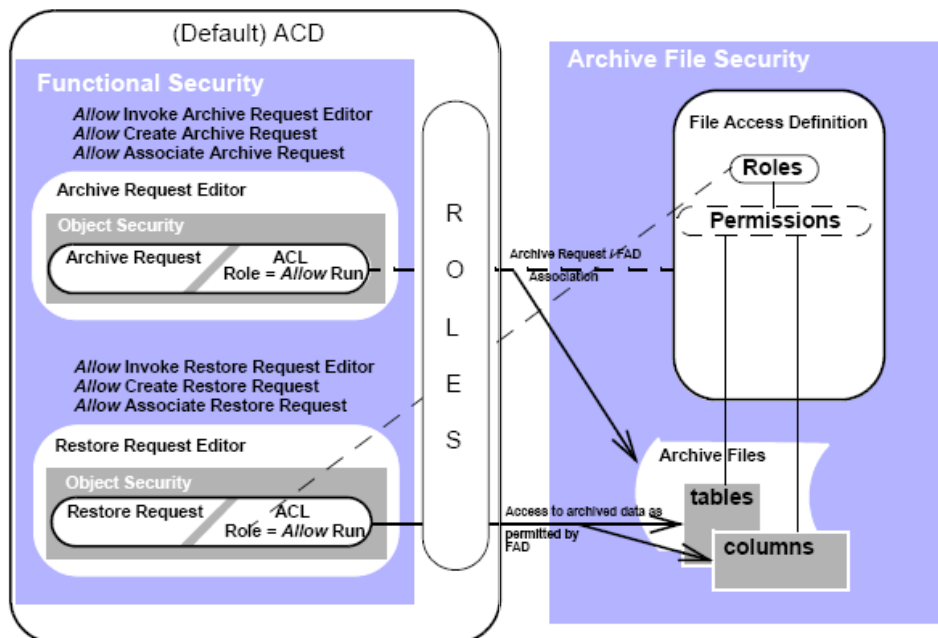


## Optim Security

### Optim provides three types of security

- **Archive File Security**  
 Archive File Security allows you to control access to data in Archive Files. For example, you might use Archive File Security to prevent any access to data in a specific table or column for most users while granting access to members of selected roles for the same data.
- **Functional Security**  
 As the most general level of Optim Security, Functional Security allows you to control user access to the interface for functions provided by Optim.
- **Object Security**  
 Object Security allows you to control access to specific objects in the Optim Directory.

## Archive File Security

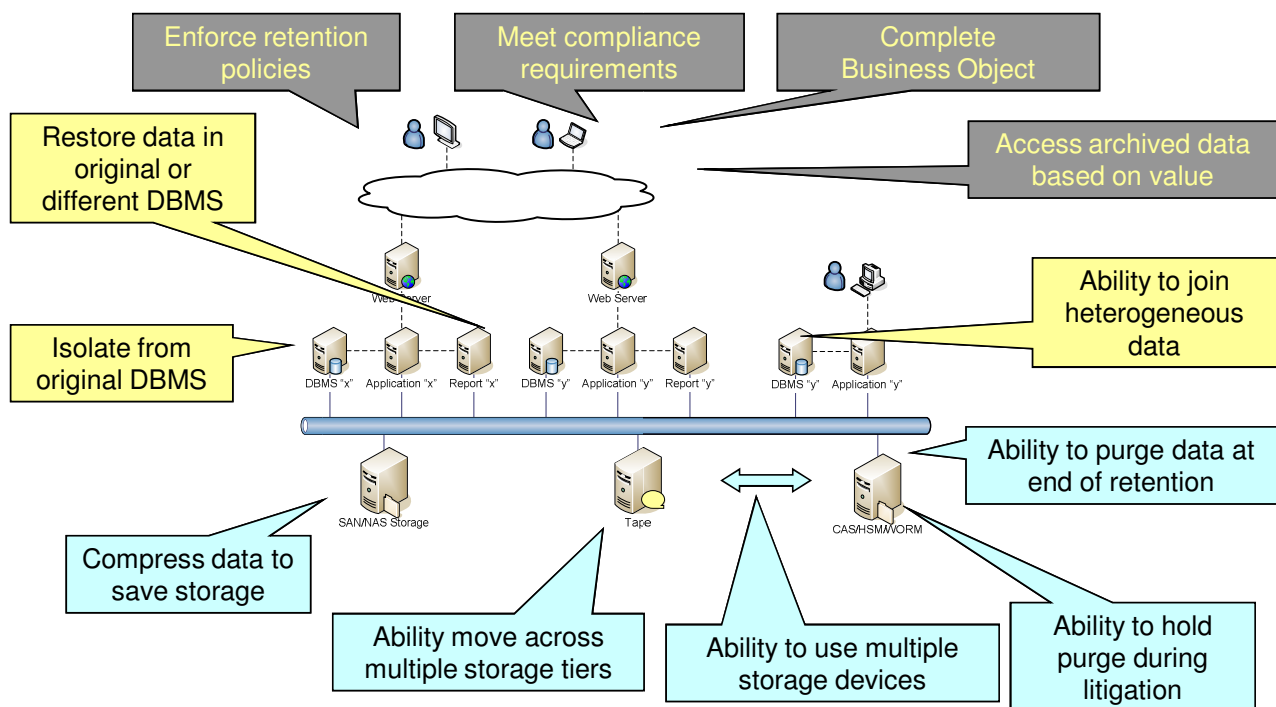


## Benefits of Optim Data Growth Management

- Facilitate Compliance with Data Retention and Disposal Regulations
  - ☑ Automate Data Retention and Disposal policies
  - ☑ Business Rule Driven approach
  - ☑ Maintain Security via audit-ready immutable format archives
  - ☑ Industry standard methods to access archives to respond quickly to audit and discovery requests.
  
- Improve Application Performance and Availability
  - ☑ Improve On-Line Performance & Meet SLA's
  - ☑ No downtime caused by batch processing overruns
  - ☑ Improve performance with queries, batch processing and reporting
  - ☑ One of the most understated and most lasting of benefits
  
- Speed Back-up and Recovery
  - ☑ Bring up important most current data first
  - ☑ Bring up older data as conditions permit
  - ☑ Increase accuracy of testing through fresher Data
  
- Reduce Infrastructure Spend
  - ☑ Enable Lifecycle Management and Tiered Storage Strategies
  - ☑ Archive historical data to reclaim capacity
  - ☑ Reduce Hardware Capacity Upgrades



## Archiving Summary



## Lab Preparation - Optim Terminology and Concepts



## Optim Terminology and Concepts

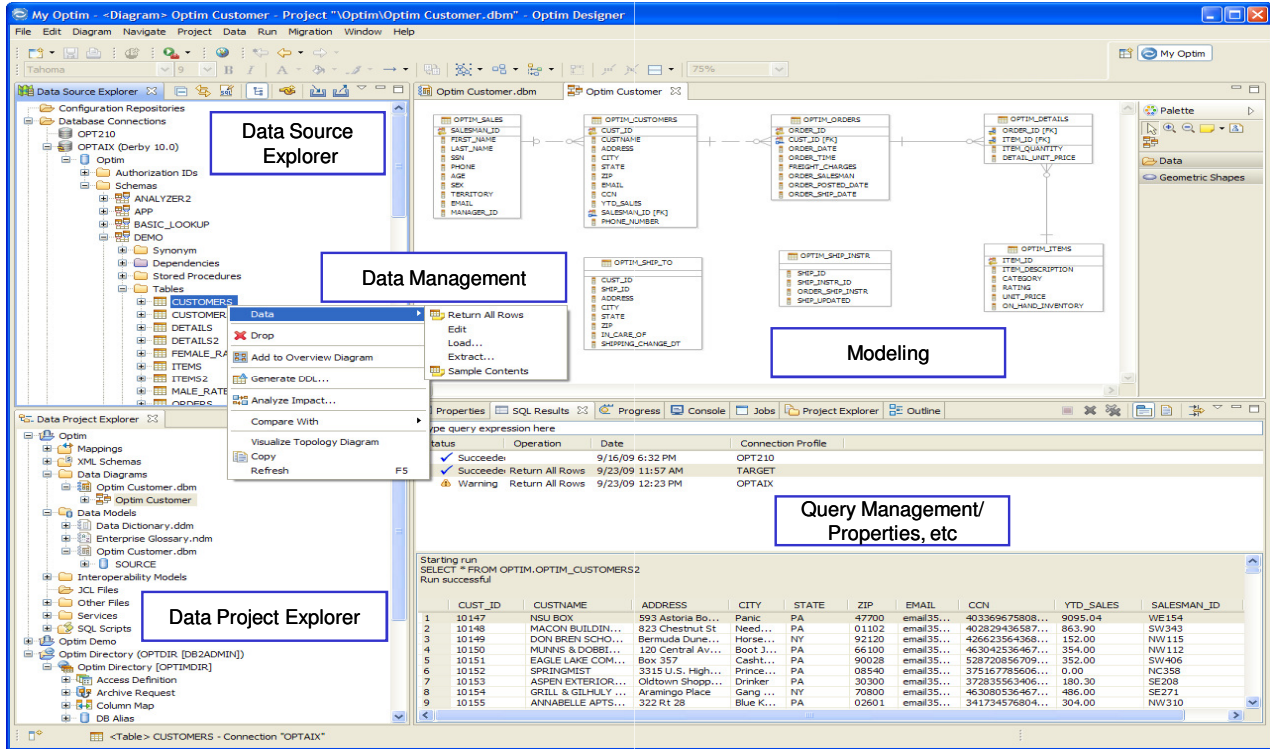
### Optim Designer:

- Physical Data Model
- Logical Data Model with Optim Properties
- Data Access Plan
- Optim Service
- Optim Directory
- Database Aliases
- Relationships (Native, Imported and Extended)

### Optim Legacy Client:

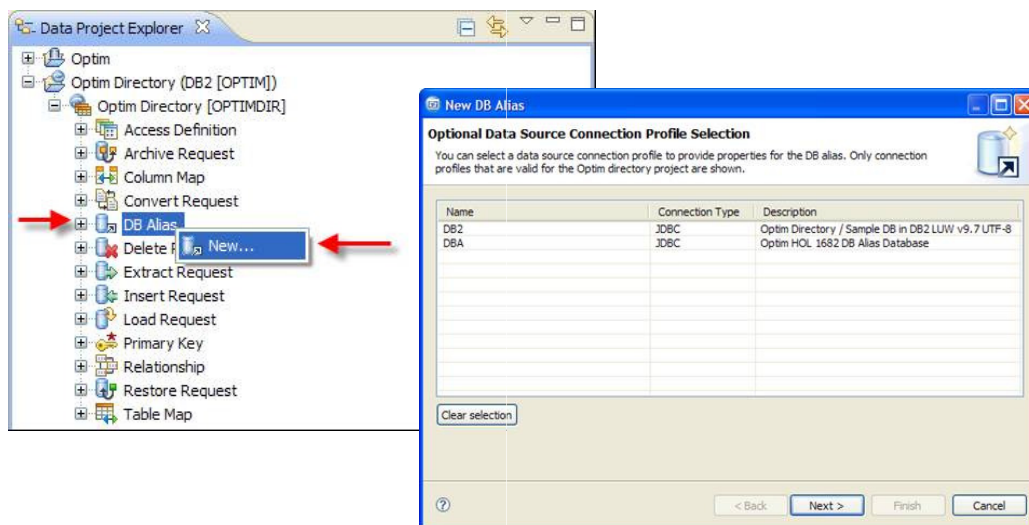
- Access Definitions
- Table Maps
- Column Maps

# Optim Designer Landscape- Terminology



# DB Alias

- Logical name for Optim physical connection to database
- High-level qualifier for database object names: DBalias.creatorid.objectname
- Specific to database/data driver
- Client driver must be installed, configured, and tested with native access
- Make sure you have connection information (server, port #, authorized userid/password)



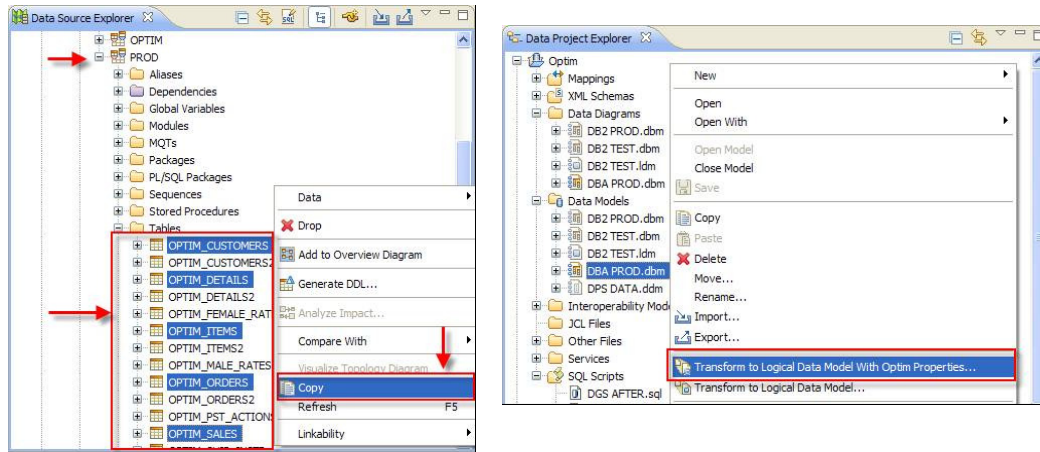
# Physical and Logical Data Models

## Physical Data Model:

- Physical model ties back database schema/tables/columns, structure to physical database connection
- May be created via reverse engineering, DDL, or drag/drop elements such as tables from a physical database connection (Data Explorer)

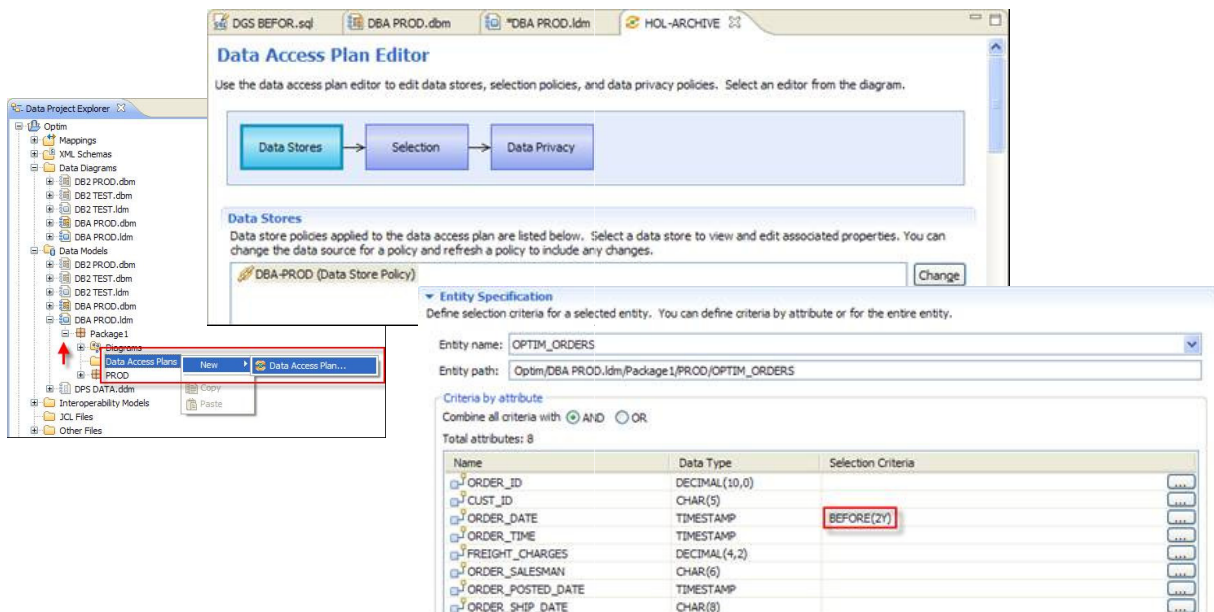
## Logical Data Model with Optim Properties:

- Logical representation of the database model
- Can selectively include elements (does not have to include entire schema or database)
- Required to generate Optim requests
- Can be imported from InfoSphere Discovery or generated from Physical Data Model or imported



# Data Access Plan

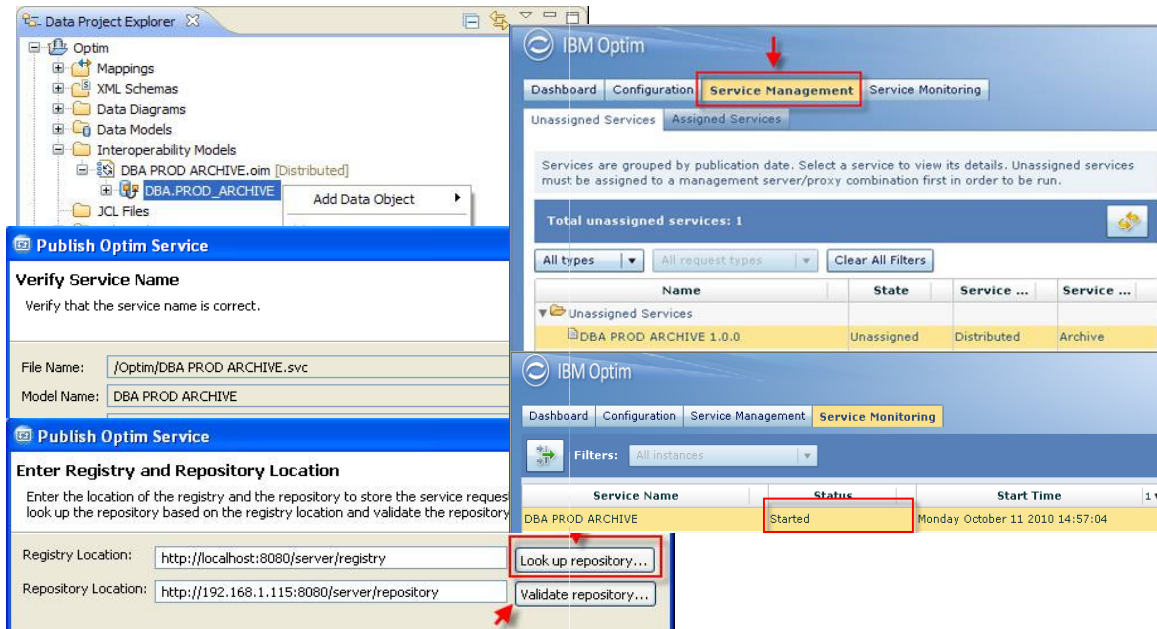
- Specify selection criteria and privacy policies
- Specify start table, relationships, and reference tables



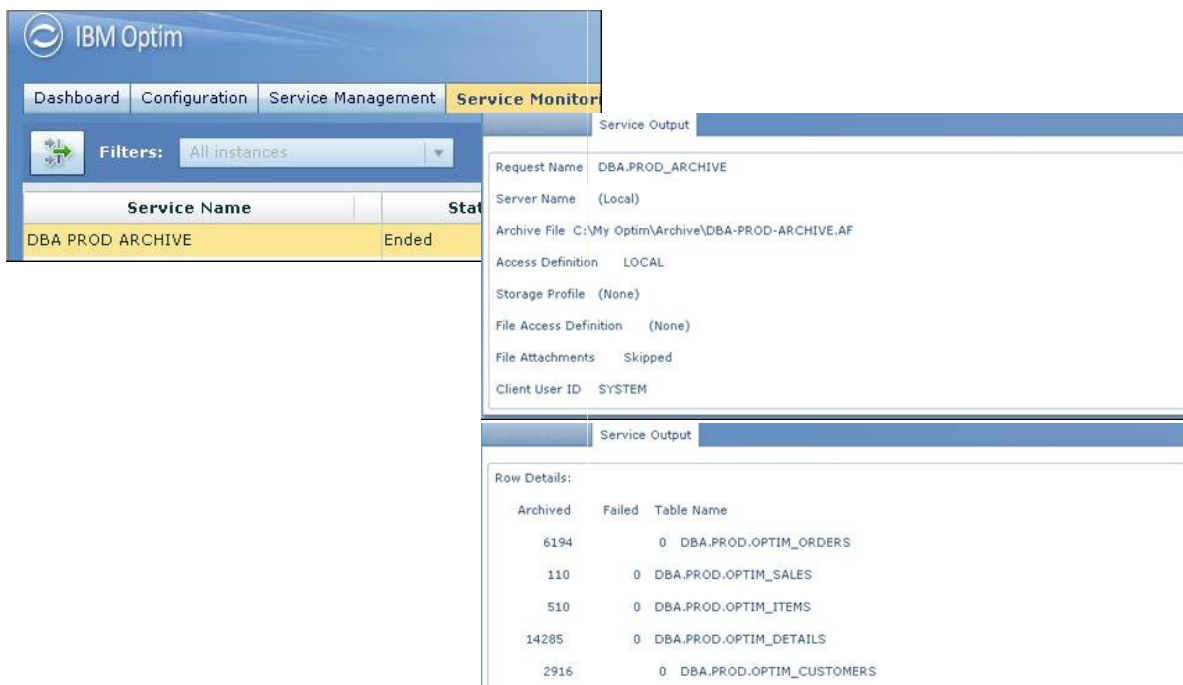


## Optim Service- Execute Optim Functions (Archive, Restore, etc.)

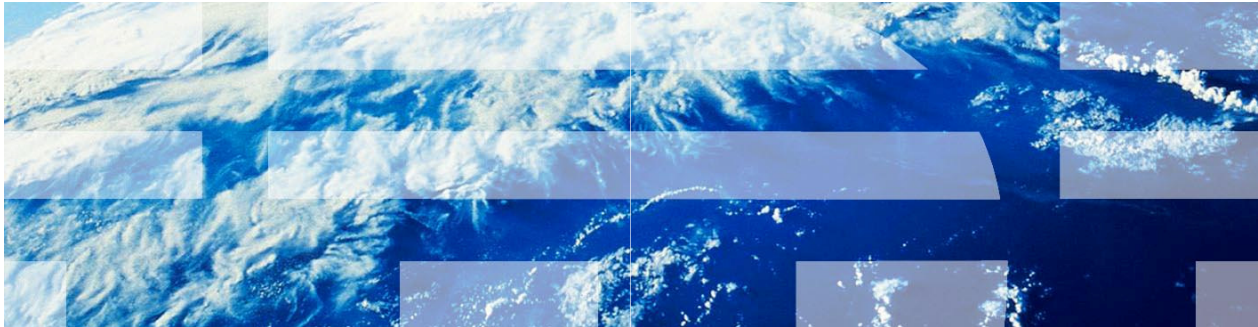
- Service can be created, executed, and monitored directly from Designer
- Service can also be published to a registry



## Sample Service Execution Output

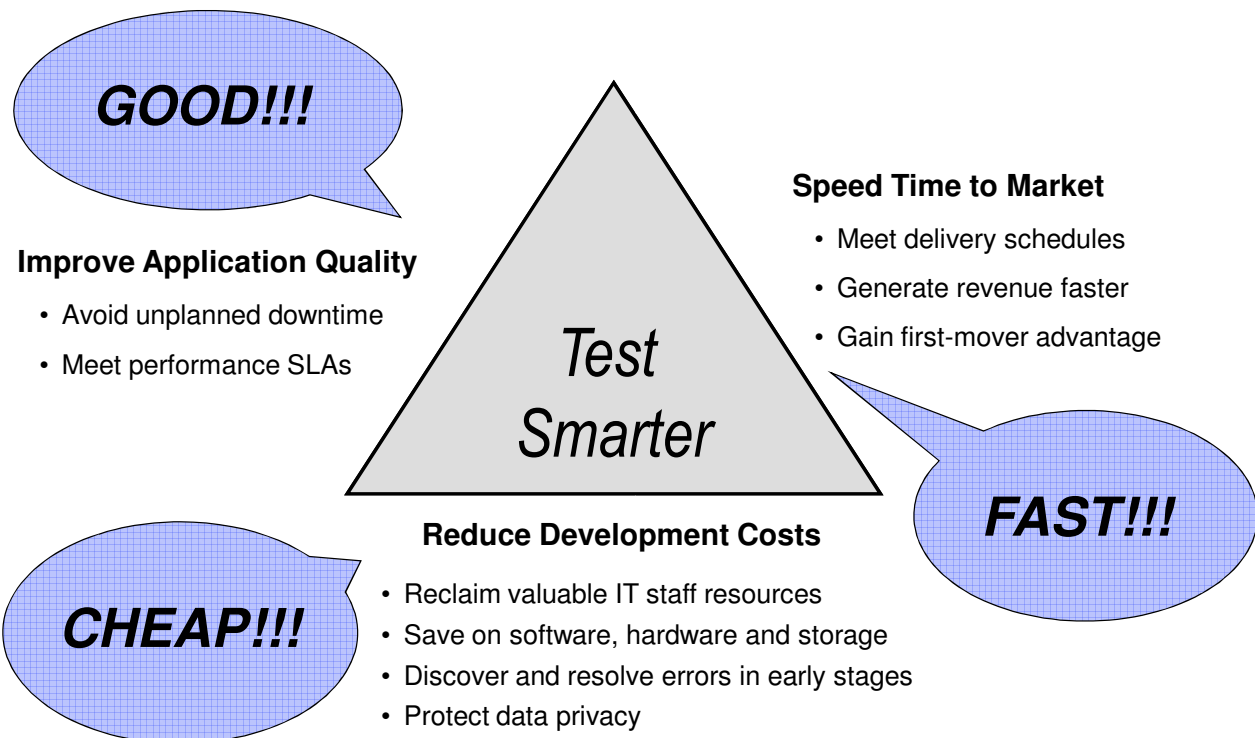


# Optim Test Data Management



© 2012 IBM Corporation

## A Picture to Explain





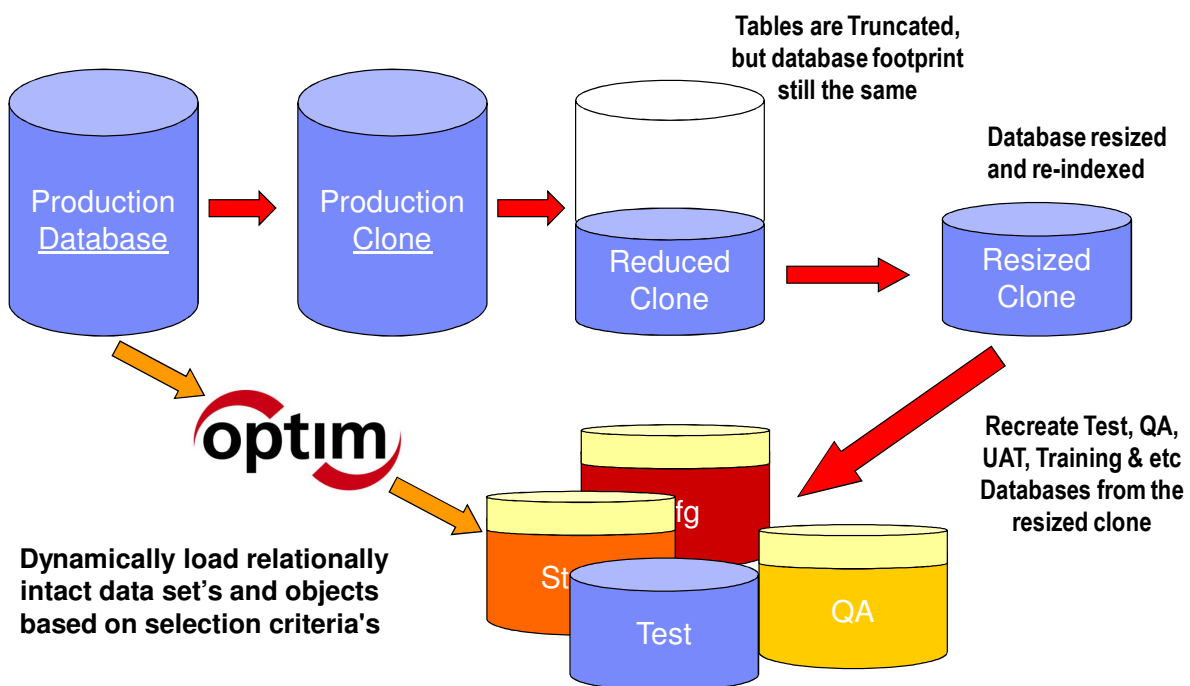
## Test Data Management: Best Practices

TDM refers to the need to manage data used in testing and other non-production environments

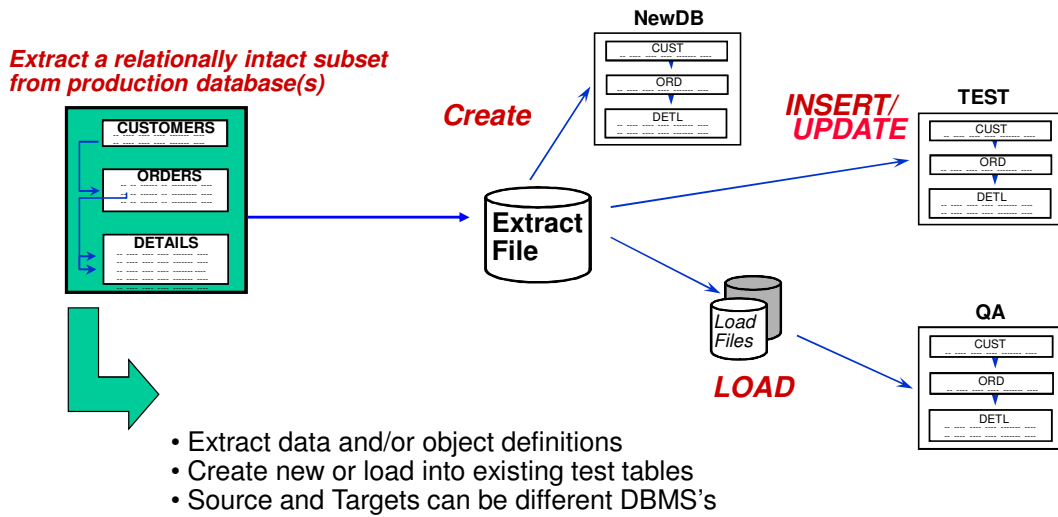
- Extract related subsets of production data that are targeted to functionality under test
- De-identify / mask related test data to protect privacy
- Quickly and easily refresh test environments
- Edit data to create error and boundary conditions
- Compare “before” and “after” images of test data

**Benefits:** Improving application quality & customer satisfaction

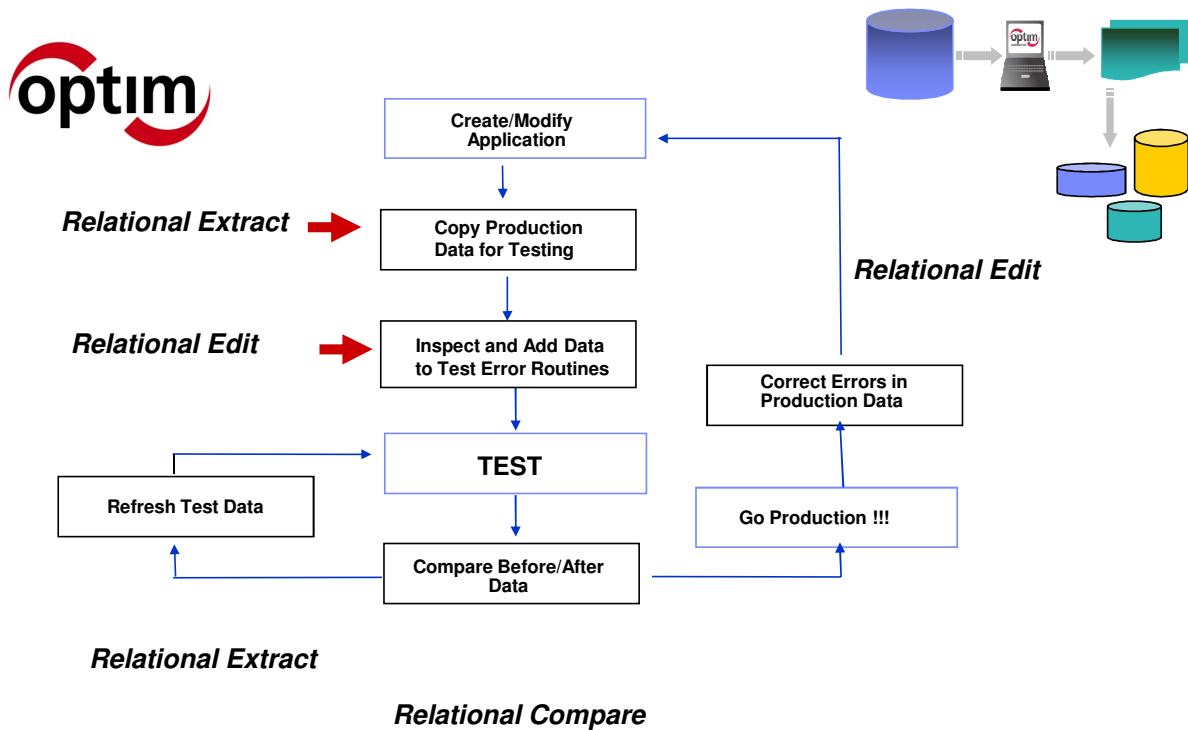
## Subsetting Conceptual Overview



# Creating Test Environments



# Test Data Management Process

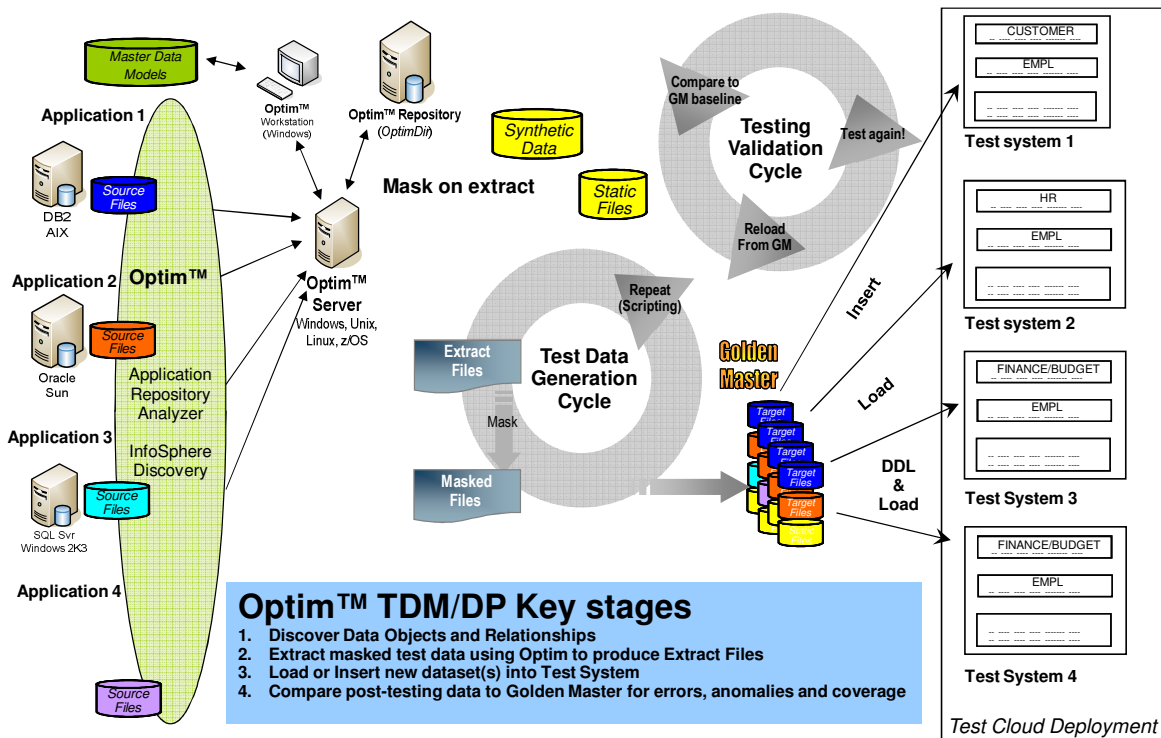


## Optim Test Data Management Steps

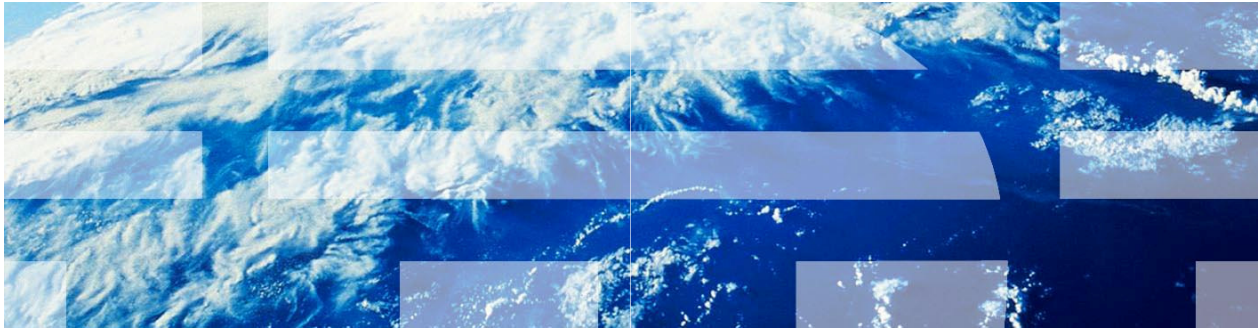
1. Identify the data to be extracted
2. Optionally define the data to be masked
3. Create the extract file
4. Review the validity of the extract
5. Identify the target non-production database
6. Insert or load the data into the target system
7. Browse or Compare target with source

## Conceptual Test Data Management / Data Privacy Architecture

*Test Data Factory generates from models and/or masked Production extracts*



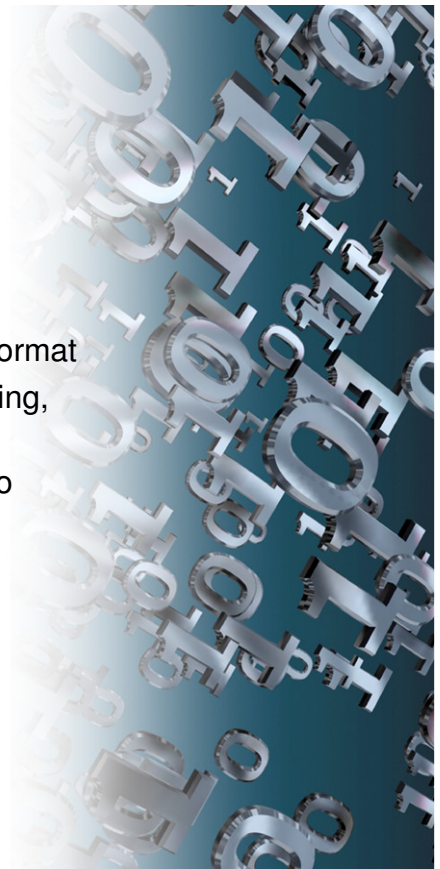
# Optim Data Privacy



© 2012 IBM Corporation

## What is Data Masking?

- Also known as: data de-identification, depersonalization, desensitization, obfuscation, data scrubbing
- Technology that helps conceal real data
- Scrambles data to create new, legible data
- Retains the data's properties, such as its width, type and format
- Common data masking algorithms include random, substring, concatenation, date aging
- Used in non-production environments as a Best Practice to protect sensitive data



## Optim Data Privacy

- IBM Optim provides the fundamental components of test data management and enables organizations to **de-identify, mask and transform** sensitive data
- Companies can apply a range of transformation techniques to substitute customer data with **contextually-accurate but fictionalized data** to produce **accurate test results**
- By masking personally-identifying information, Optim protects the **privacy and security** of confidential customer data, and **supports compliance** with local, state, national, international and industry-based privacy regulations

## Intelligent Data Masking

A comprehensive set of data masking techniques to transform or de-identify data, including:

- String literal values
- Character substrings
- Random or sequential numbers
- Arithmetic expressions
- Concatenated expressions
- Date aging
- Lookup values
- Intelligence
- Key propagation

### Example 1

Patient Information	
Patient No.	123456 SN 333-22-4444
Name	Erica Schafer
Address	12 Murray Court
City	Austin State TX 78704

Data is masked with contextually correct data to preserve integrity of test data

### Example 2

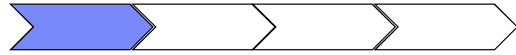
Personal Info Table		
PersNbr	FirstName	LastName
10000	Jeanne	Renoir
10001	Claude	Monet
<b>10002</b>	<b>Pablo</b>	<b>Picasso</b>
	⋮	

Referential integrity is maintained with key propagation

Event Table		
PersNbr	FstNEvtOwn	LstNEvtOwn
<b>10002</b>	<b>Pablo</b>	<b>Picasso</b>
<b>10002</b>	<b>Pablo</b>	<b>Picasso</b>

## Example: Bank Account Numbers

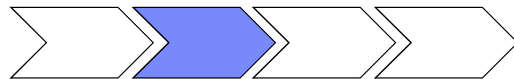
- First Financial Bank's account numbers are formatted "123-4567" with the first three digits representing the type of account (checking, savings, or money market) and the last four digits representing the customer identification number
- To mask account numbers for testing, use the *actual first three digits*, plus a *sequential four-digit number*
- The result is a fictionalized account number with a valid format:
  - ▶ "001-9898" becomes "001-1000"
  - ▶ "001-4570" becomes "001-1001"



### Masking Example #1

## Example: First and Last Name

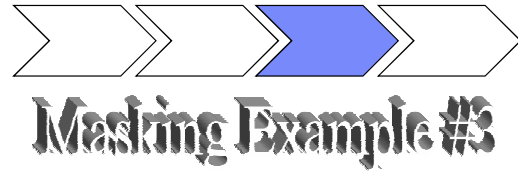
- Direct Response Marketing, Inc. is testing its order fulfillment system
- Fictionalize customer names to pull first and last names randomly from the Customer Information table:
  - ▶ "Adam Adams" becomes "Ronald Smith"
  - ▶ "Anna Adams" becomes "Elena Wu"
  - ▶ **Optim ships with over 5,000 male/female names and over 80,000 last names**



### Masking Example #2

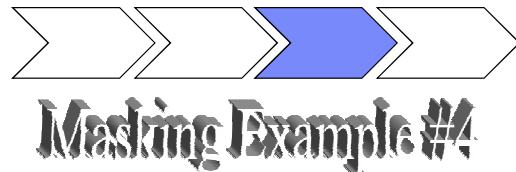
## Example: Addresses

- Direct Response Marketing, Inc. is testing its order fulfillment system
  - Fictionalize customer addresses to pull an entire address from the Customer Information table:
    - “111 Campus Drive Princeton, NJ 08540 ”
    - becomes...*
    - “1223 E. 12<sup>th</sup> Street NY, NY 10079”
- ▶ **Optim ships with over 100,000 valid CASS addresses**

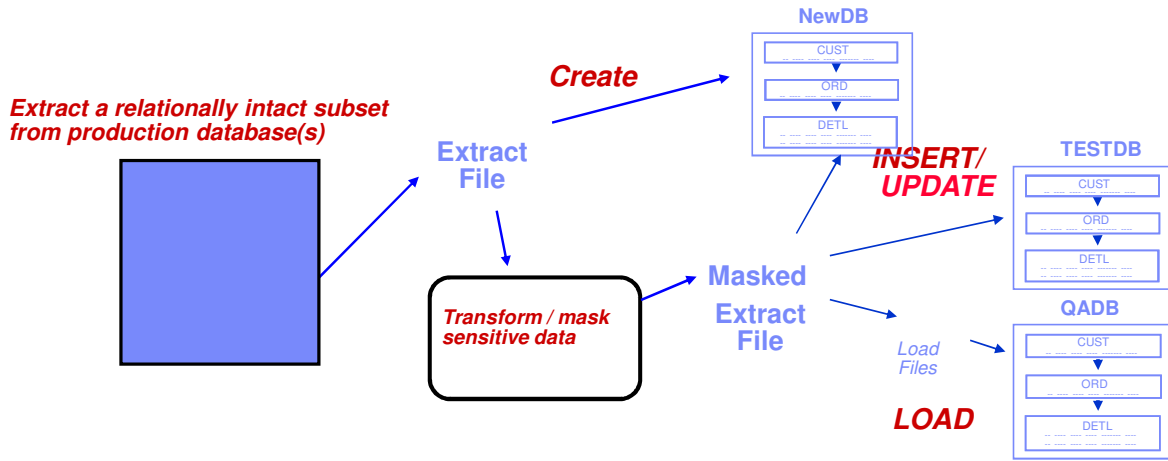


## Example: Intelligent Masking

- Generating valid **social security** numbers (as defined by the US Social Security Administration)
- Generate valid **credit card** numbers (as defined by credit card issuers)
- Generate **desensitized e-mail** addresses
  - *Generate Email address based on format: name@domain*

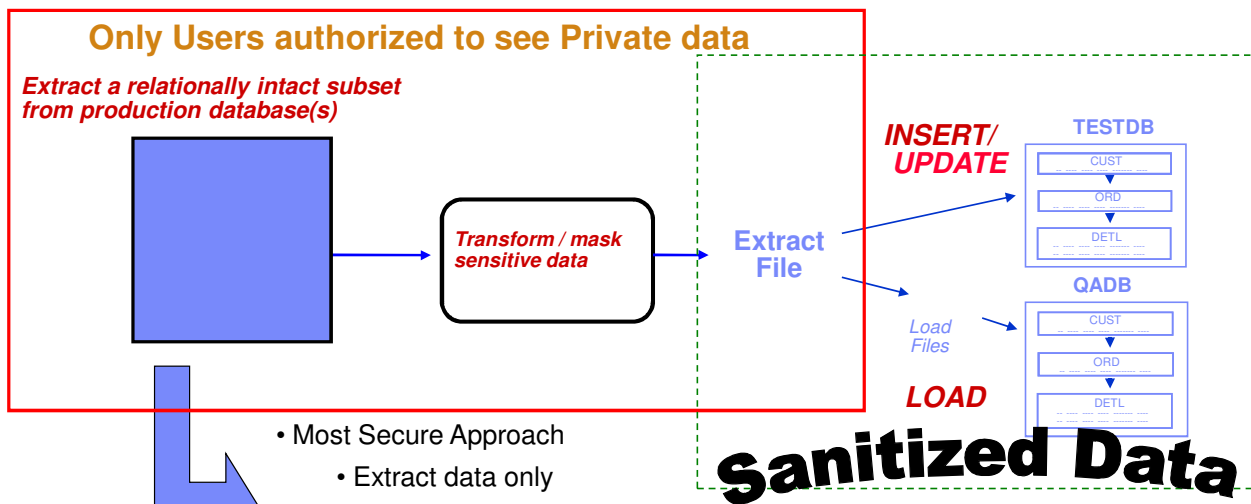


## Data Privacy in Application Testing



- Extract data and/or object definitions in pre-masked file
- Use pre-masked Extract file to create new set of tables
- Convert Pre-masked extract file data into second masked extract file
- Share masked extract file to be reused for population step
- Good practice for testing masks using COMPARE

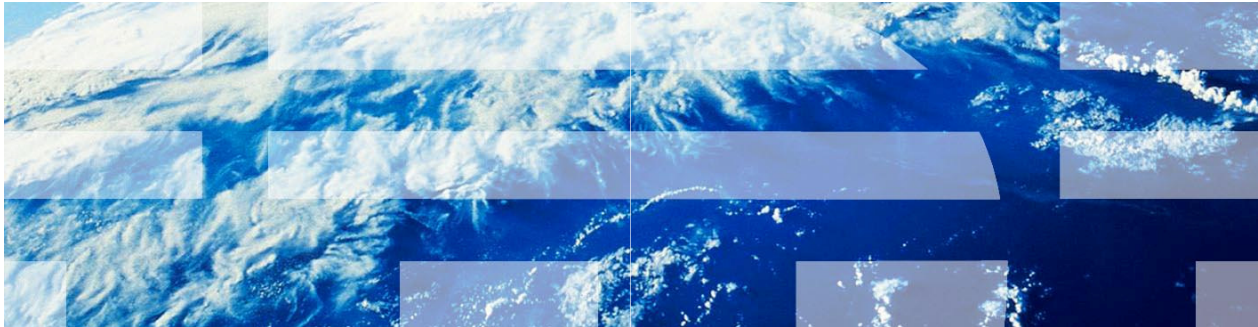
## Data Privacy in Application Testing



- Most Secure Approach
  - Extract data only
  - Convert during extract
- Extract file already contains masked data
  - Can be shared with testers to reuse



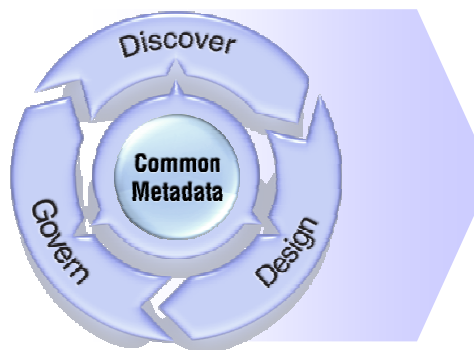
# InfoSphere Discovery



© 2012 IBM Corporation

## Information Server Life Cycle Tools

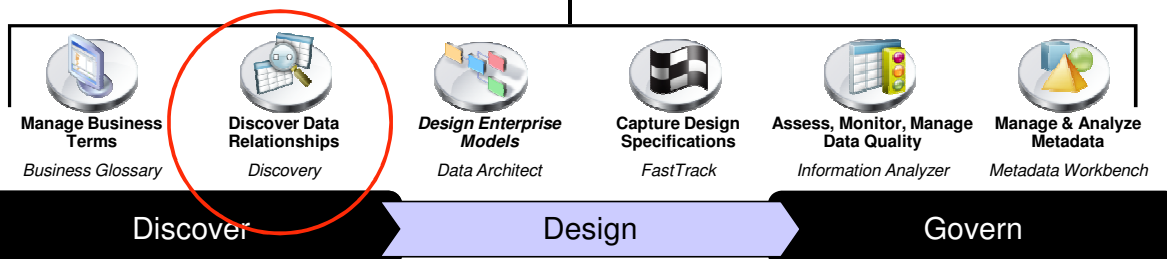
- **Discover** and understand the data across heterogeneous systems
- **Design** trusted information structures for business optimization
- **Govern** that information over time



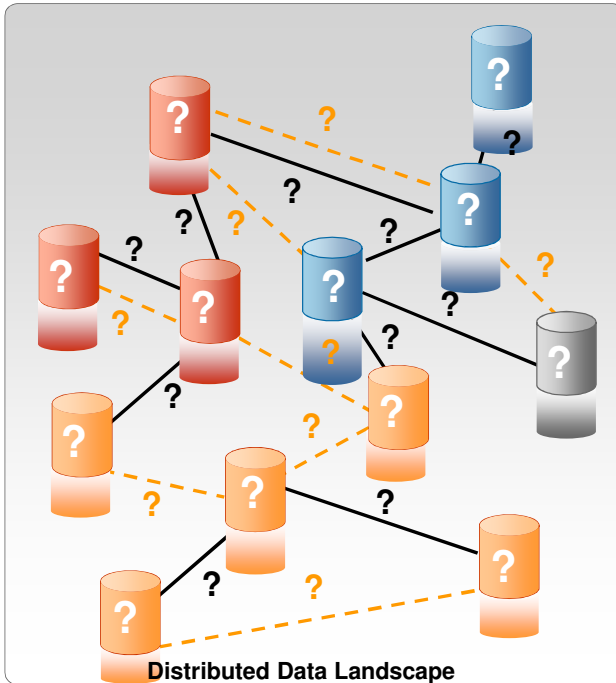
InfoSphere Foundation Tools

### Enterprise Projects

- Test Data Generation
- Application Retirement & Consolidation
- Data Archival
- Data De-identification
- Data Quality
- Data Integration
- Master Data Management
- Data Warehousing

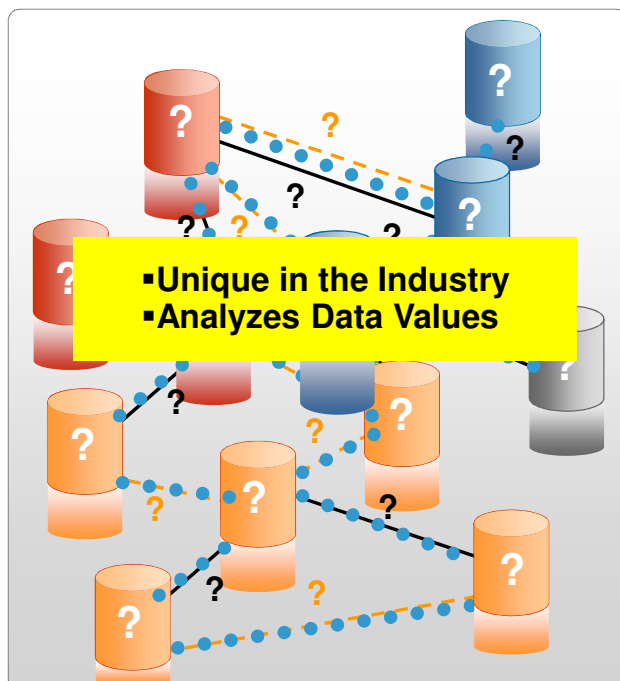


## You can't ----- what you don't understand



- Increasingly distributed
- Complex, poorly documented data & relationships within & across sources
- Data not understood because:
  - Corporate memory is poor
  - Documentation is poor or nonexistent
  - Logical relationships (enforced through application logic or business rules) are **hidden**

## InfoSphere Discovery: An Invaluable Data Analysis Tool

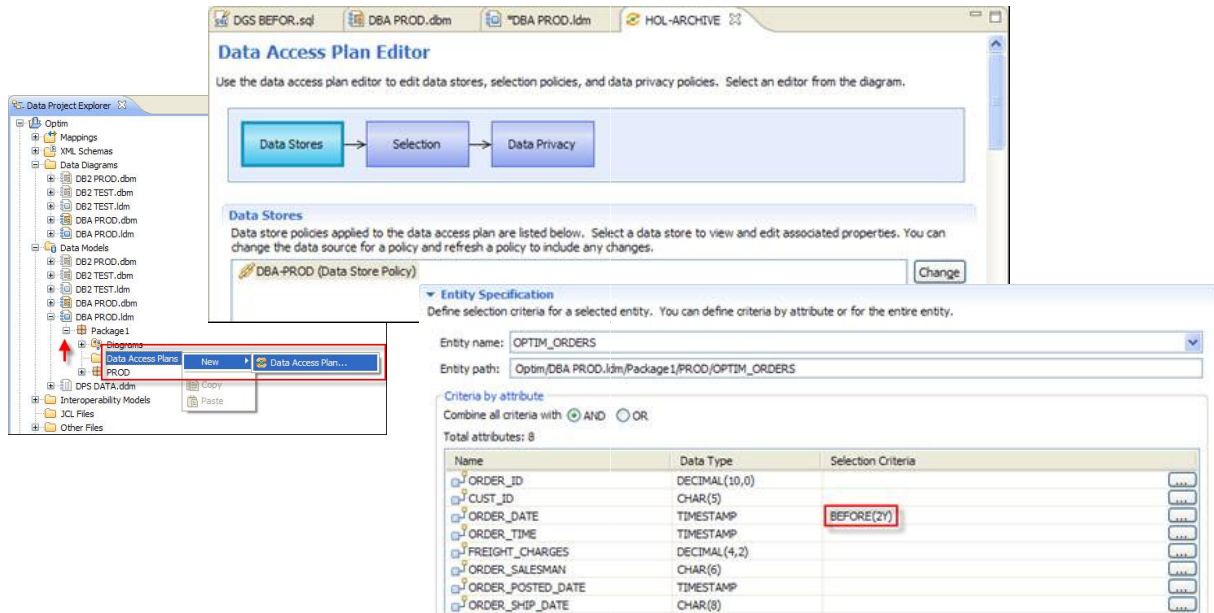


- Automated analysis of data and relationships for **complete understanding** of data assets:
  - **Characterizes** data elements **within** a Source
  - Identifies **relationships** that link data elements into “business entities” **within** a source
    - Customer, counterparty, invoice
  - Identifies **complex logic** that relates **multiple** sources



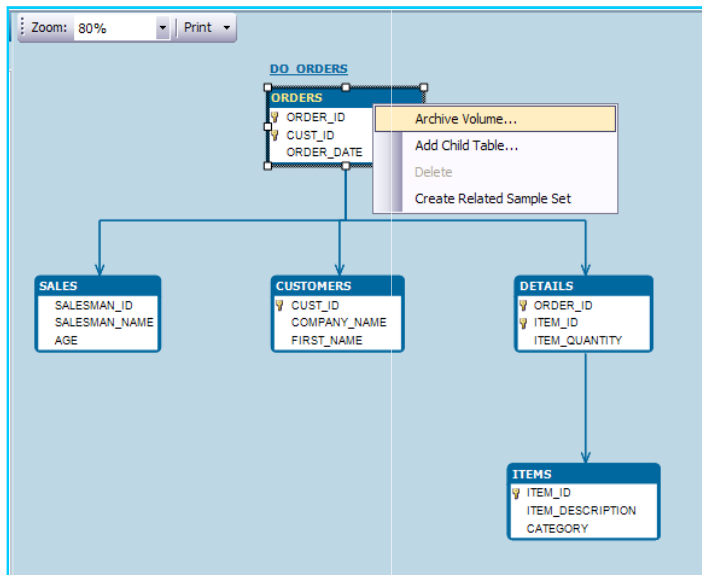
# Import to Optim as Logical Data Model

## Create Data Access Plan: Tables, Columns, Relationships Identified!



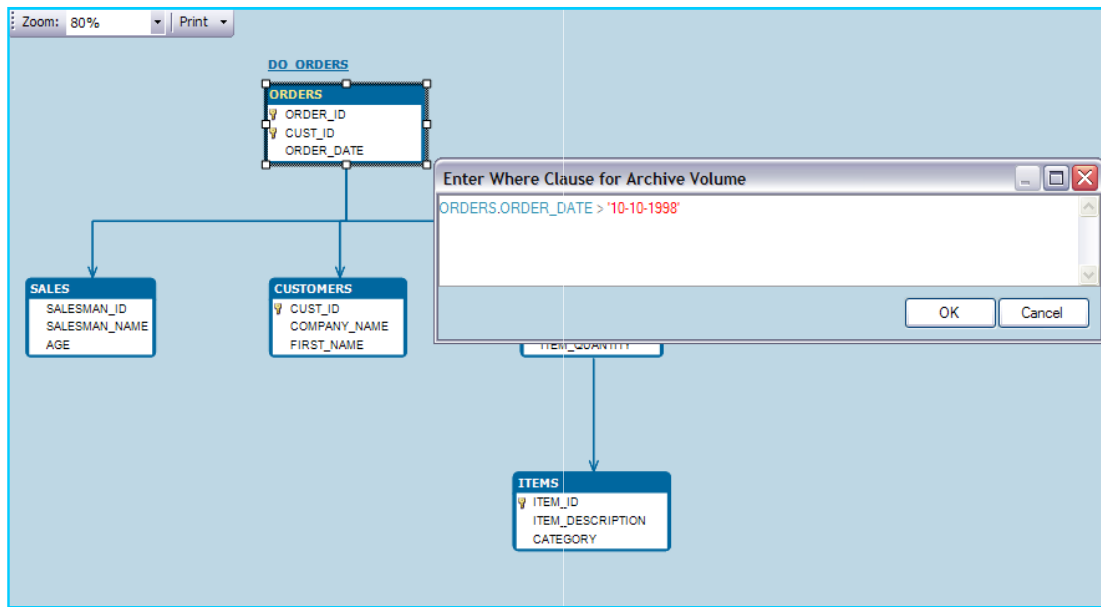
# Data Object Volume Projection

*Assess volume reduction of archiving strategy:  
 "How much do I really save if I archive away all data older than 7 years?"*



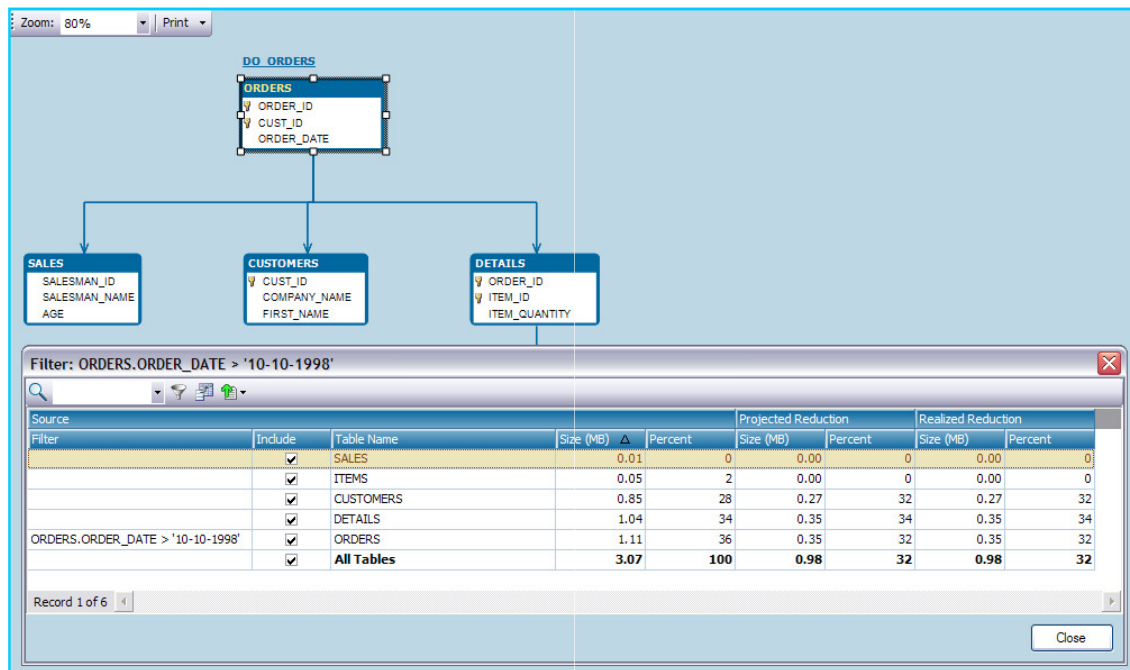
Volume projections for a filter entered in any table in the Object diagram. Here I will enter a filter on the ORDERS table.

# Data Object Volume Projection



Type in the filter and select OK.

# Data Object Volume Projection



A tabular display showing the volumes. You can check/uncheck the "Include" flag to calculate the volume.

# Pattern Based Sensitive Data Discovery Example: SSN

## InfoSphere Discovery Classified Columns View

MC	MasterCard	Yes	Yes	No
NEA	Notes Email Address	No	Yes	No
NIF	Spanish NIF	No	Yes	No
NINO	UK NINO	No	Yes	No
PAN	Passport Number	No	Yes	No
PN	Personal Name	No	Yes	No
SIN	Canadian SIN	No	Yes	No
SSN	US Social Security Number	Yes	Yes	No

Column Metadata			Column Classification			Statistics		
Name	Table	Data Set	Hit Rate	Ap...	Method	Cardinality	Selectivity	Null Count
COMMENTS	ALL_POSITIONS_F	Data Set 1	50%	<input type="checkbox"/>	Discovered	4	0.00	3641
MRC_ACCTD_AMOUNT	MK_ADJTMNT	Data Set 1	35.2941%	<input type="checkbox"/>	Discovered	12	0.04	257
MRC_EXCHANGE_RATE	MK_KE_TRX_70	Data Set 1	52.5%	<input type="checkbox"/>	Discovered	45	0.00	9397
MRC_EXCHANGE_RATE	MK_SCHED	Data Set 1	33.2075%	<input type="checkbox"/>	Discovered	47	0.00	9590
MRC_ACCTD_EARNED_DISC_TAKEN	MK_SHOU_APPS_ALL	Data Set 1	100%	<input type="checkbox"/>	Discovered	1	0.00	200
MRC_ACCTD_UNEARNED_DISC_T...	MK_SHOU_APPS_ALL	Data Set 1	100%	<input type="checkbox"/>	Discovered	1	0.00	200
ID_PIN	MK_KE_HDRS	Data Set 1	100%	<input type="checkbox"/>	Discovered	23	0.00	0
BILLING_ID	PRFL_3_CUST	Data Set 1	39.777%	<input type="checkbox"/>	Discovered	269	1.00	0

URL	URL	No	Yes	No
USPHN	US Phone Number	Yes	Yes	No
USSC	US State Code	No	Yes	No

## Hidden Sensitive Data Discovery

Row	Member	SS #	Age	Phone	Sex
1	595846226	123-45-6789	15	(123) 456-7890	M
2	567472596	138-27-1604	8	(138) 271-6037	F
3	540450091	154-86-4196	22	(154) 864-1961	M
4	514714372	173-44-7900	55	(173) 447-8996	F
5	490204164	194-26-1648	4	(194) 261-6476	F
6	466861109	217-57-3046	66	(217) 573-0453	M
987,623	444629628	243-68-1812	25	(243) 681-8107	F
987,624	423456789	272-92-3629	87	(272) 923-6280	M

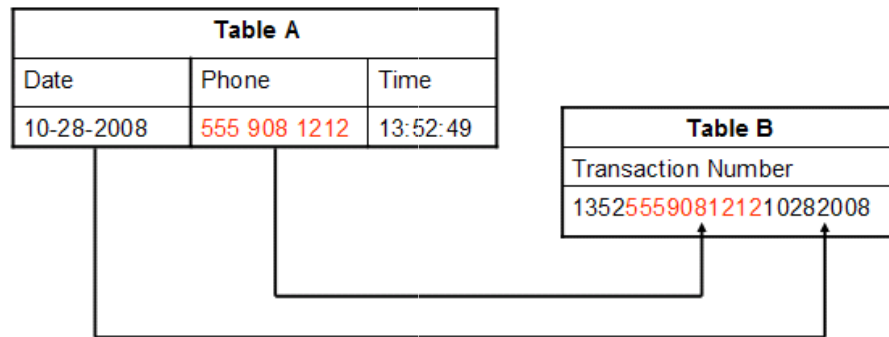
  

**TYPES**

- Hidden within longer fields  
- SSN hidden in a 46 digit routing number
- Divided up across multiple columns  
- SSN divided into three separate columns
- Transformed  
- Items converted into codes

- Finding Sensitive Data Elements (SDE) in each system can take days
- Whole and partial SDE's can be found across the enterprise

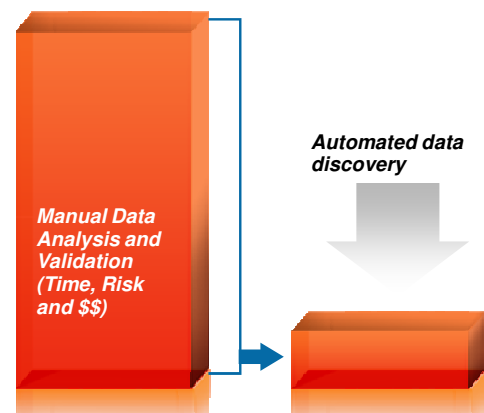
## Hidden Sensitive Data Discovery Example



**Hidden instances of private data represent a potential compliance risk**

## IBM InfoSphere Discovery – Invaluable Data Analysis Tool

- Accelerate deployment of your information agenda projects:
  - Improves accuracy, predictability and repeatability
  - Speeds project data analysis by as much as 10 times
  - Minimize SME Time





---

*Thank You*









© Copyright IBM Corporation 2012.

The information contained in these materials is provided for informational purposes only, and is provided AS IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, these materials. Nothing contained in these materials is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software. References in these materials to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. This information is based on current IBM product plans and strategy, which are subject to change by IBM without notice. Product release dates and/or capabilities referenced in these materials may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way.

IBM, the IBM logo and [ibm.com](http://ibm.com) are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at [ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

Other company, product and service names may be trademarks or service marks of others.



Please Recycle

---