Information Management Optim

IBM Optim Data Privacy & Information Lifecycle Management POT for Bell Canada

March 18, 2010

Presenter: Ken Lee - kklee@ca.ibm.com

Sales Specialist: Stephen Moore – moores@ca.ibm.com

Objectives

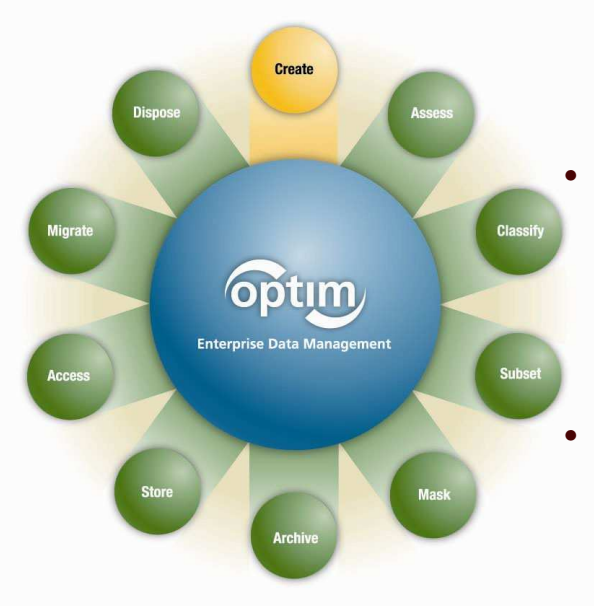
- Understand the IBM InfoSphere Discovery offerings
- Understand the IBM Optim Test Data Management offerings.
- Understand the IBM Optim Data Privacy offerings
- To obtain a basic understanding of the principles of Information Life Cycle Management (ILM).
- Obtain a basic understanding on how an IBM® Optim™ ILM software solution can manage data growth in a production environment.
- Understanding the Optim software components that support an ILM strategy.
- IBM software solutions are reinforced with hands on labs to further demonstrate product capabilities.

Agenda

- Introduction
- IBM Optim Enterprise Data Management Overview
- IBM Optim Test Data Management Overview
- IBM Optim Test Data Management Lab
- IBM Optim Data Privacy Overview
- IBM Optim Data Privacy Lab
- IBM Optim Archive/Restore Overview
- IBM Optim Archive/Restore Lab
- IBM Optim ILM Overview
- IBM Optim ILM Lab
- IBM InfoSphere™ Discovery Overview
- IBM InfoSphere™ Discovery Lab

Information Management Optim

IBM Optim High Level Overview



Optim Solutions

- **IBM Optim Test Data Management Solution:**

  Value: Speeds Application Delivery, Improves Application Quality, Reduces risk

  - Create targeted, “right-sized” subsets faster, improve operational efficiencies by shortening iterative testing cycles

- **IBM Optim Data Privacy Solution:**

  Value: Risk Mitigation

  - De-identify and mask confidential test data to close security gap, comply with privacy policies, PCI compliant.

- **IBM Optim Data Growth Solution (Archiving):**

  Value: Improve Application Performance, Reduce Infrastructure Costs & Improve Compliance

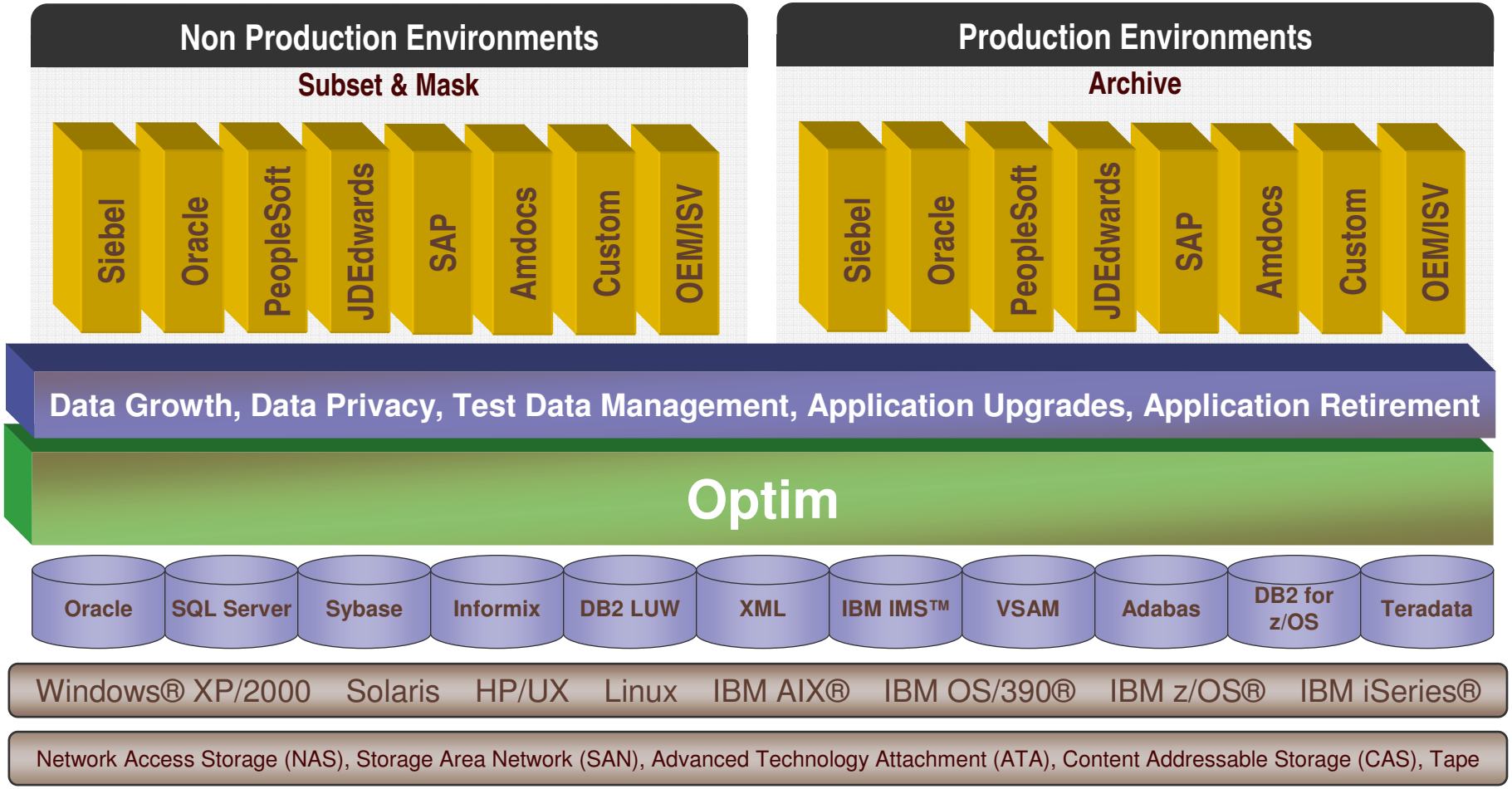
  - Segregates “inactive” historical data from current data to relieve the live database from heavy volumes of data

- **IBM Optim Decommissioning Solution:**

  Value: Infrastructure Cost Reductions & Compliance

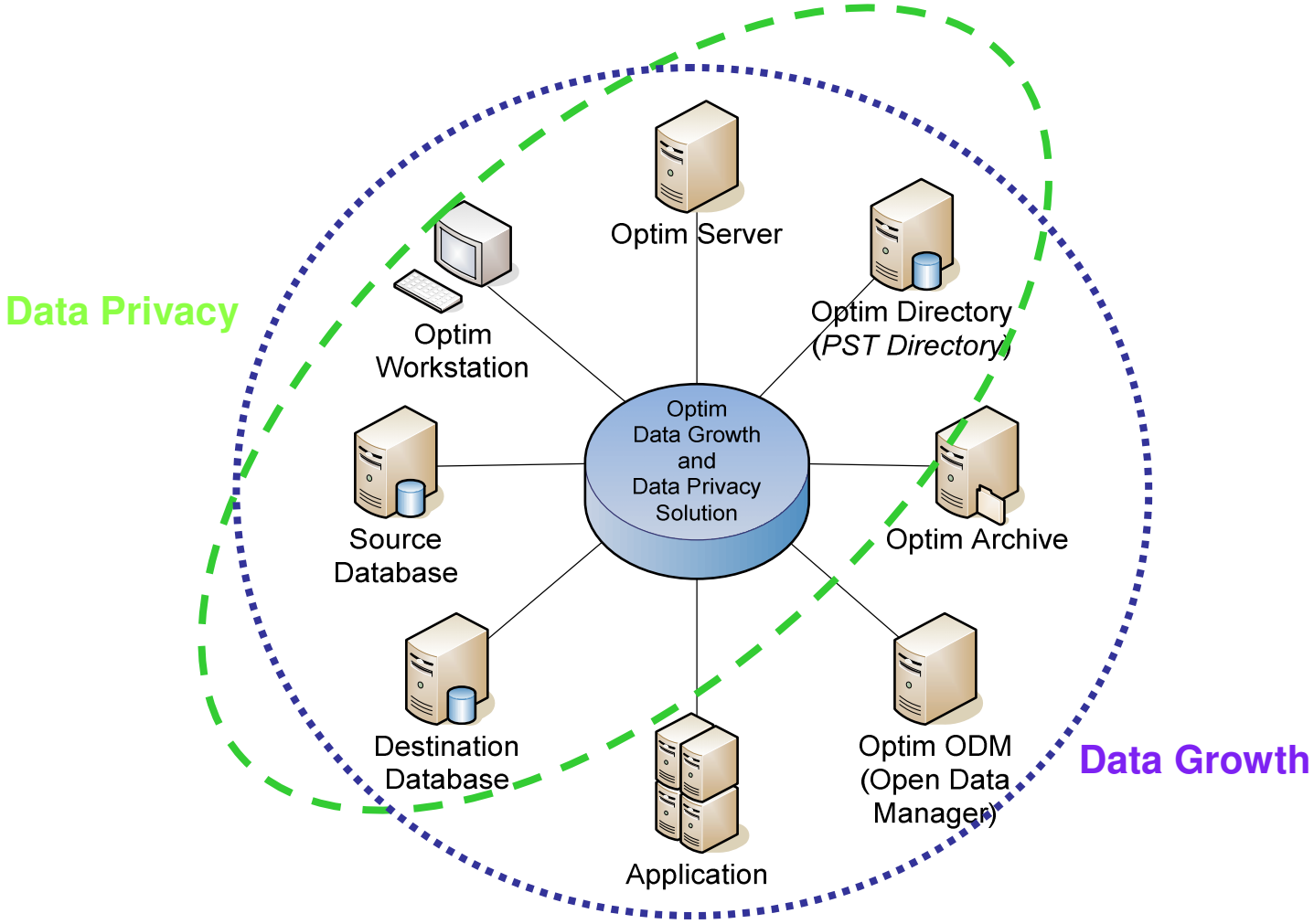
  - Enables Legacy Apps to be decommissioned, accompanying infrastructure can be re-purposed,
  - Enable access to the legacy data via standard interfaces

Enterprise Architecture

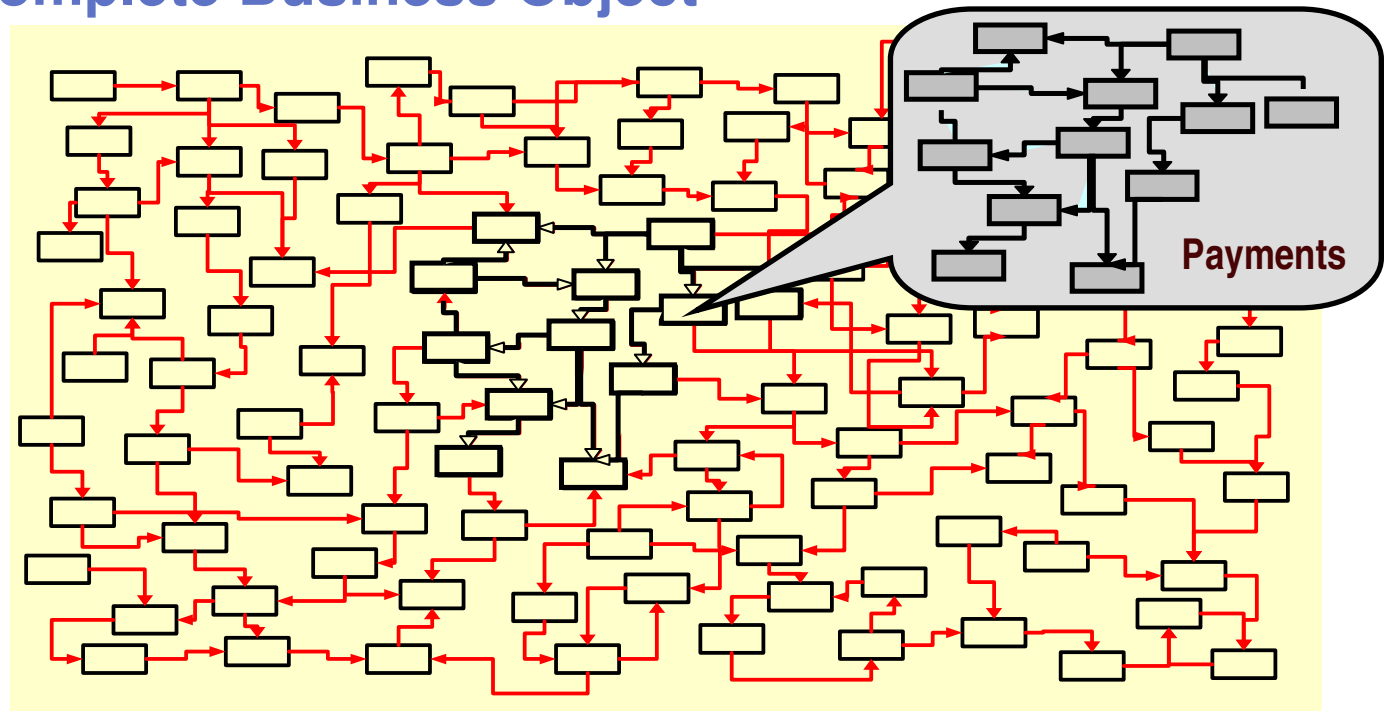


Single, scalable, interoperable Enterprise Data Management solution provides a central point to deploy policies to extract, store, port, and protect application data records from creation to deletion

Optim Concepts: Physical Components

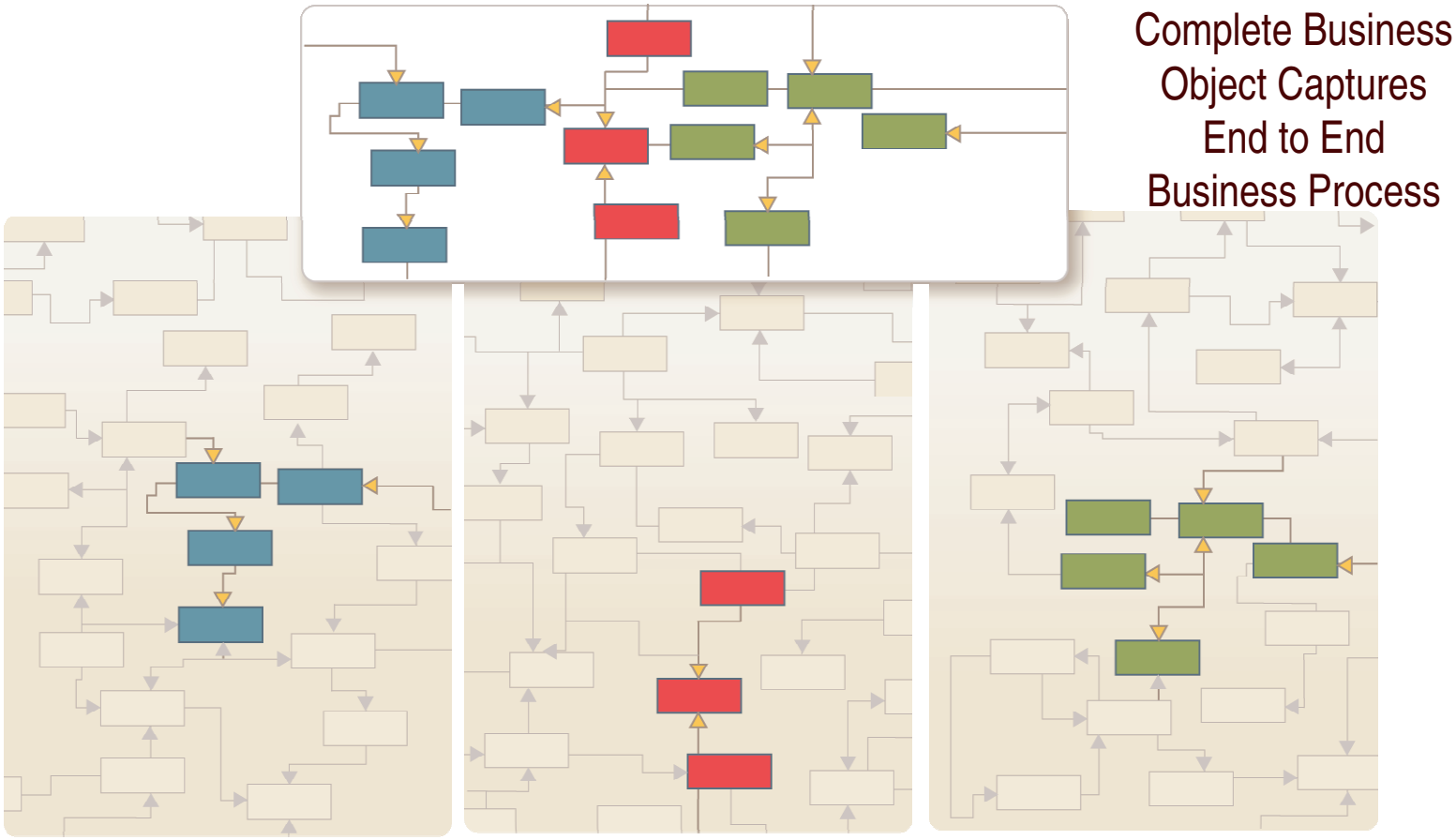


Complete Business Object



- Represents application data record – payment, invoice, customer
  - **Referentially-intact subset of data across related tables and applications; includes metadata**
- Provides “historical reference snapshot” of business activity
- Federated object support across enterprise data stores
- Relationships can be in the database or application driven relationships

Extract - Federated Data Support



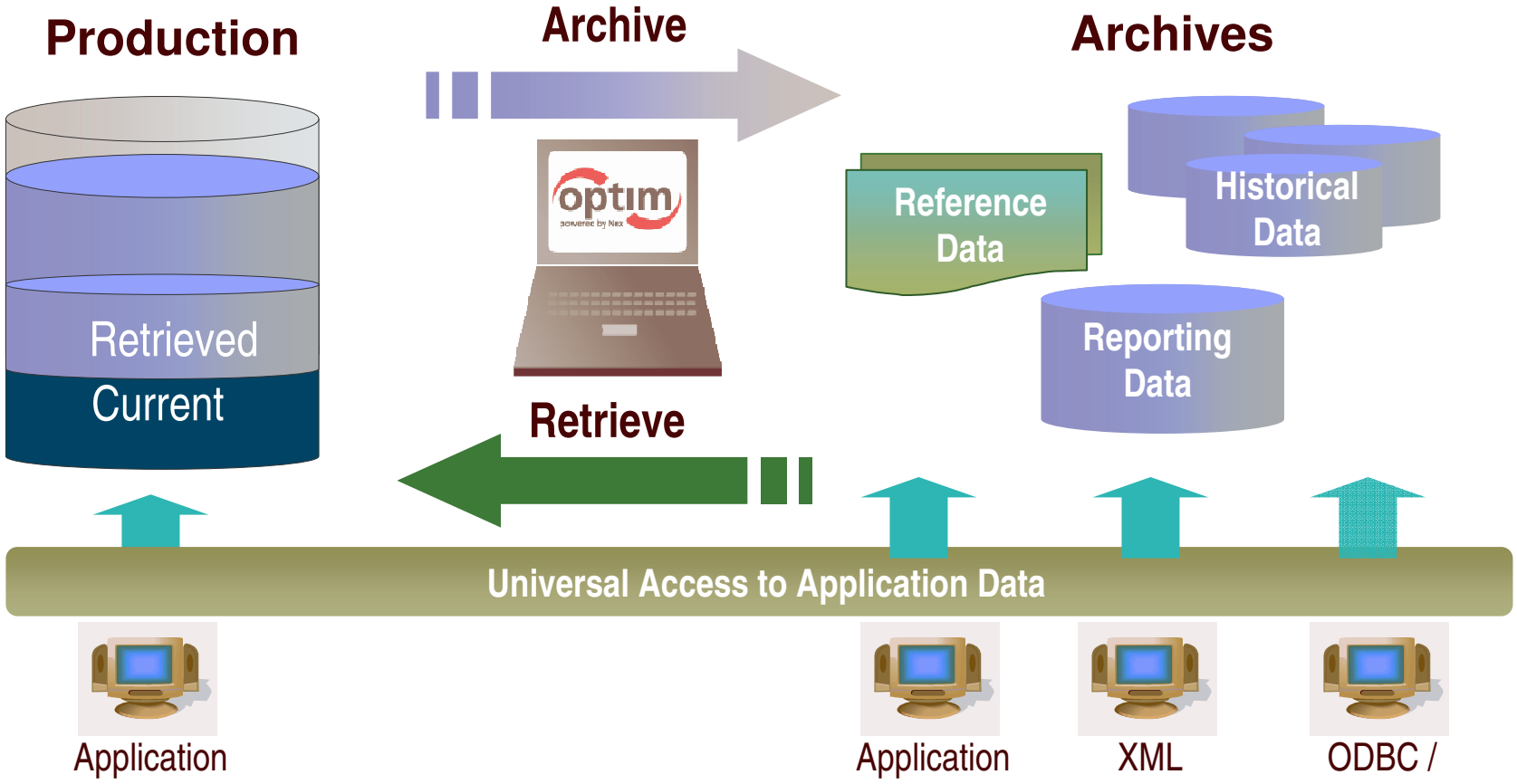
Complete Business Object Captures End to End Business Process

Retek / Oracle

Custom Inventory Management / DB2

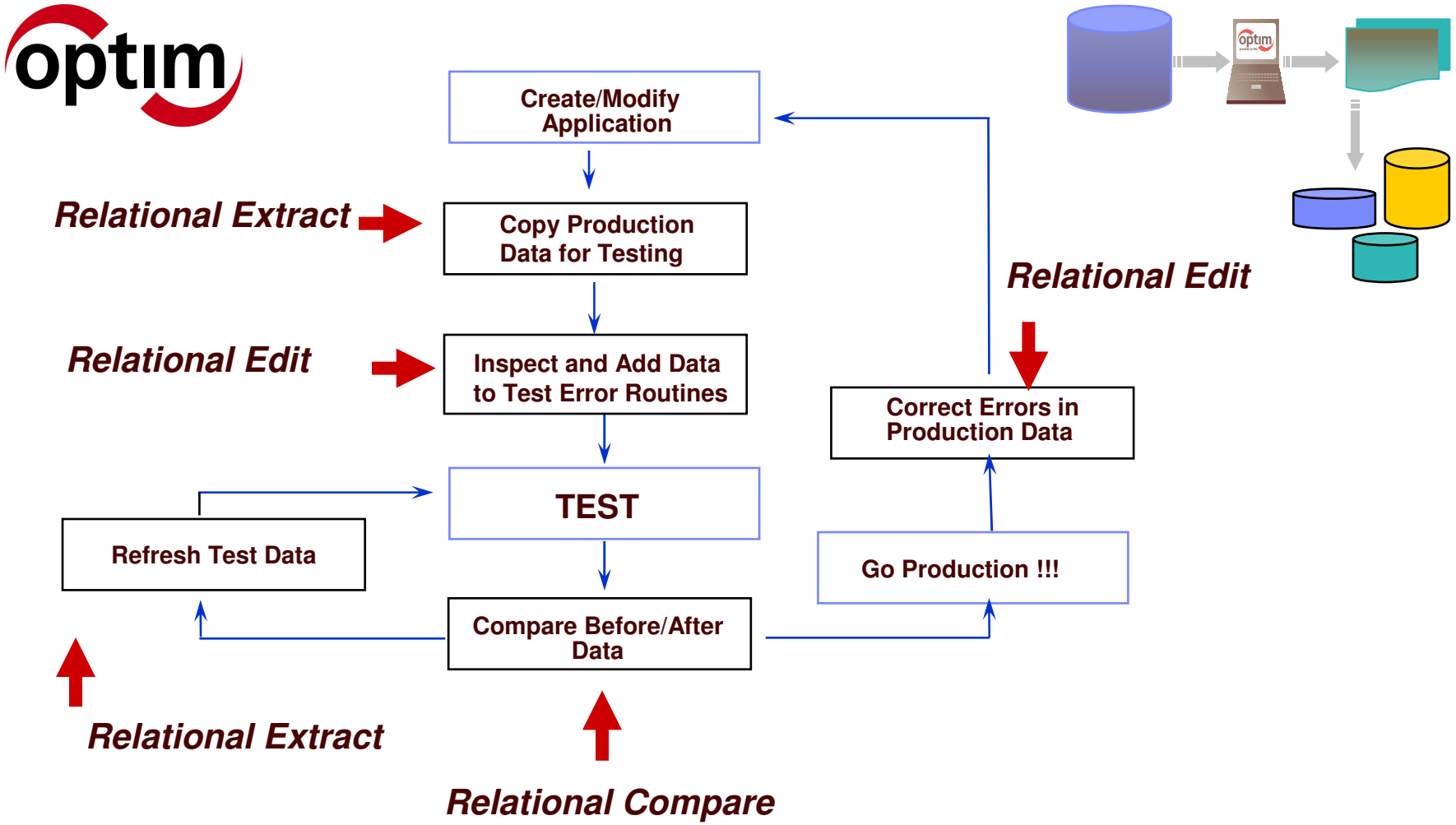
Other apps / any database management system (DBMS)

Optim Data Growth Solution

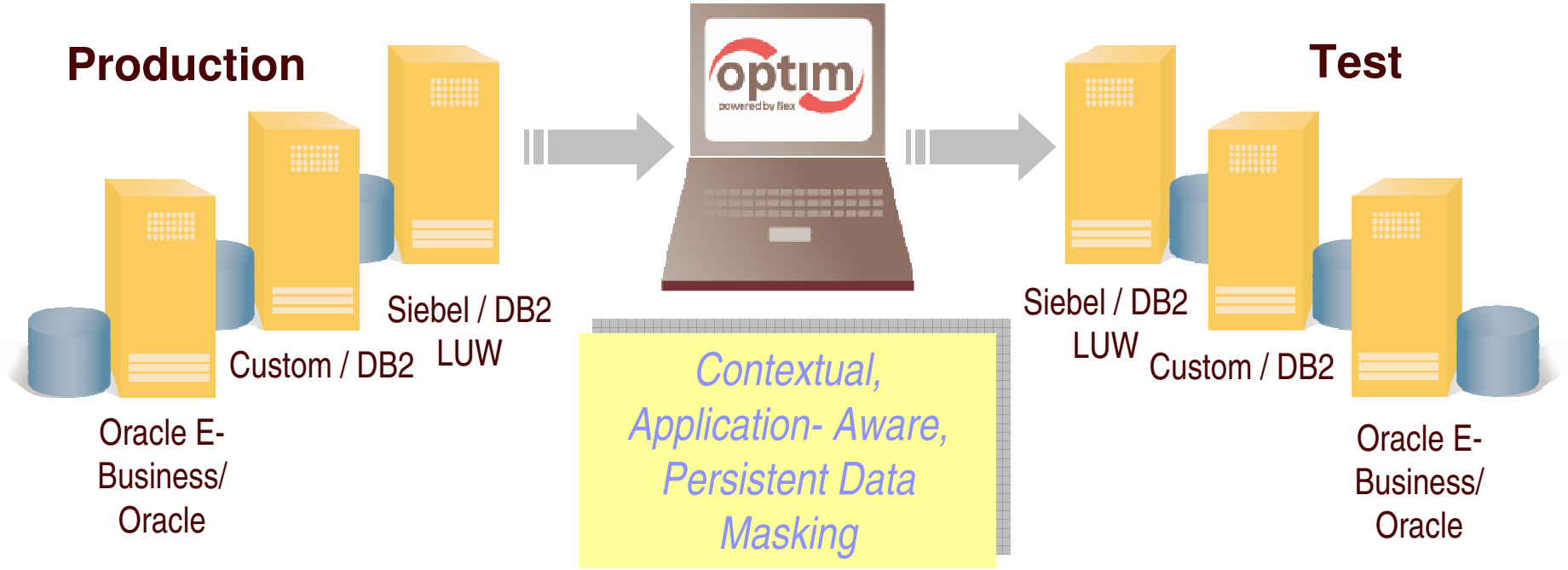


- **Complete Business Object provides historical reference snapshot of business activity**
- **Storage device independence enables ILM**
- **Immutable file format enables data retention compliance**

Product Overview : Optim Test Database Management



Optim Data Privacy Solution

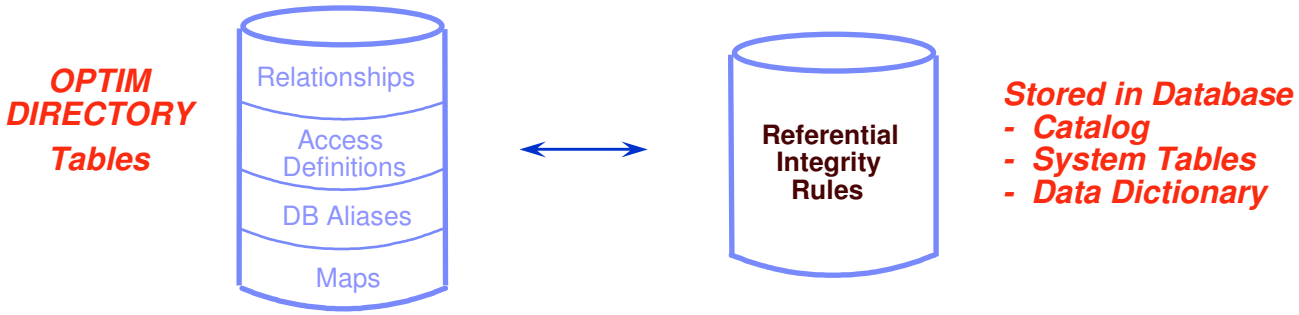


- Substitute confidential information with fictionalized data
- Deploy multiple masking algorithms
- Provide consistency across environments and iterations
- Enable off-shore testing
- Protect private data in non-production environments

Terminology

- Optim Directory
- Database Aliases
- Relationships (Native, Imported and Extended)
- Access Definitions
- Table Maps
- Column Maps
- Move
  - Extract
  - Insert/Load
- Edit
- Compare

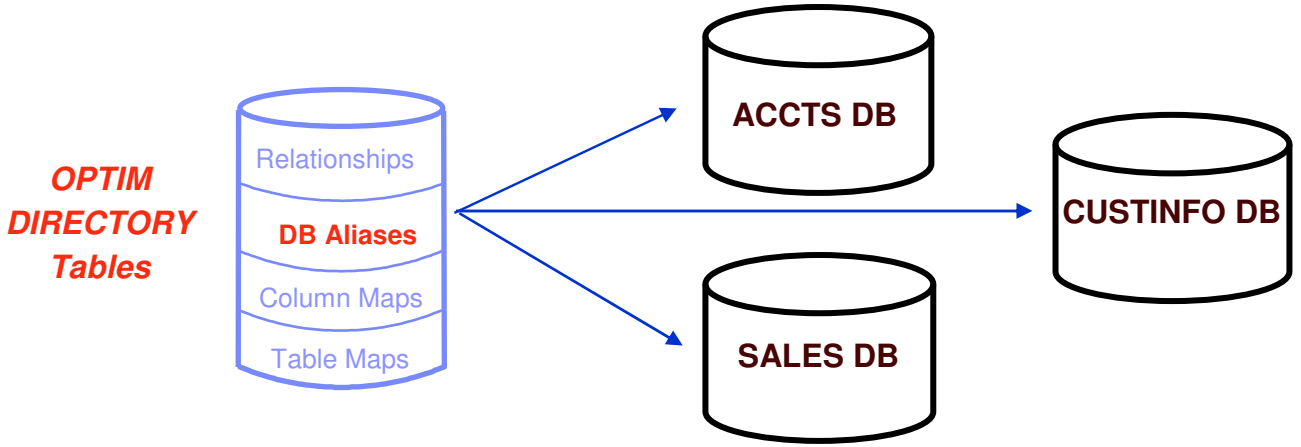
The OPTIM Directory



- **Optim catalog**
  - Supplements information stored in the database (DB)
  - Maintains product definitions and tracks processing
  - Stores database connection information (DB Aliases)
  - Stores user-defined relationships

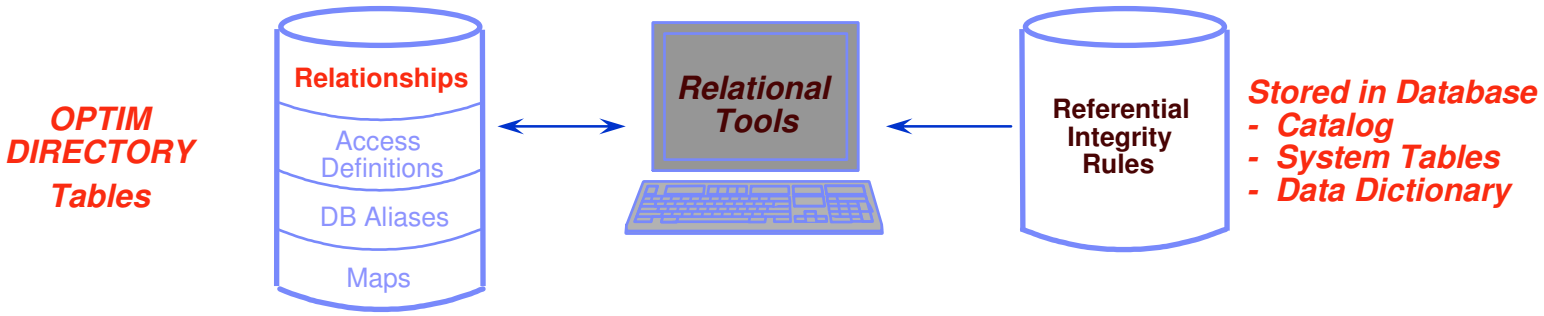
Database Alias

Establishing the Database Connection



- **Optim view of a database connection**
  - High-level qualifier for database object names
    - DBalias.creatorid.objectname
  - Enables cross-Database access
  - Saved in Optim Directory

Relationships



- Automatically derived from database RI rules
- OR… defined within OPTIM
- OR… imported from DDL
- OR… imported from InfoSphere Discovery

Shared by all OPTIM components

Extended Relationships

Sales Table

SALESMAN_ID Char (5)	MANAGER_ID Char (7)
(NC)003	NC00123
NW012	NW00564
SC005	SC00234
SE012	SE00582

Example 1
Using Substr Function
Parent Table Sales
Substr(SALESMAN_ID,1,2)
Child Table ***District***
DISTRICT_CD

District Table

DISTRICT_CD Char (2)	MANAGER_NO Char (5)
NC	00123
SC	00564
SE	00234
NW	00582

Example 2
Using Concat Function
Parent Table **Sales**
MANAGER_ID
Child Table **District**
DISTRICT_CD || MANAGER_ID

Extended Relationships

Sales Table

AGE Integer	SEX Char (1)
45	F
56	F
18	M
35	M

Female_Rates Table

Age Integer	Rate Numeric (5,0)
32	1
35	1
45	1
50	2

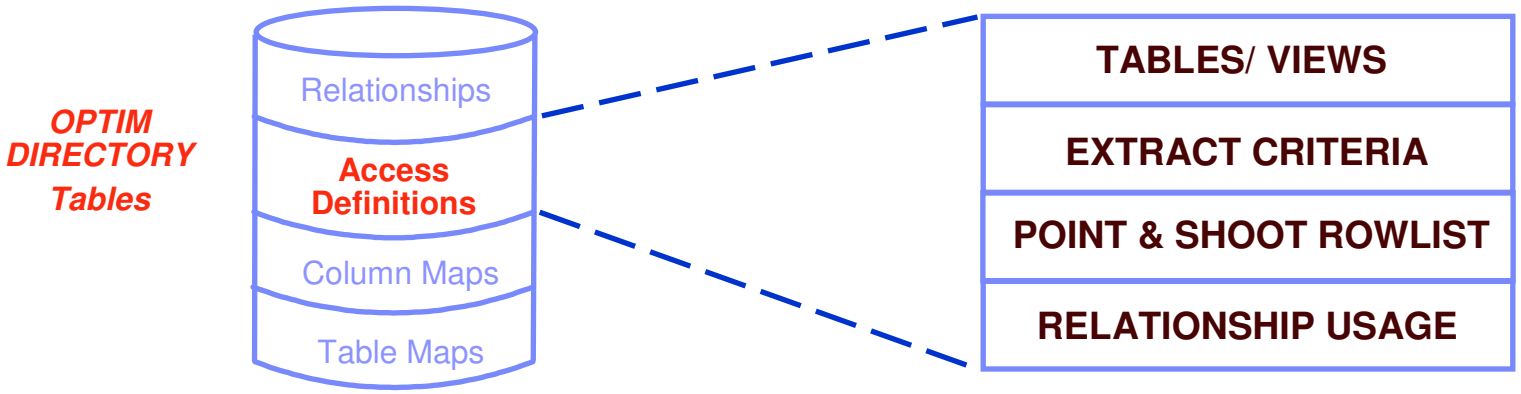
Male_Rates Table

Age Integer	Rate Numeric (5,0)
18	3
35	1
45	1
50	2

Example 3
Data Driven Relationships

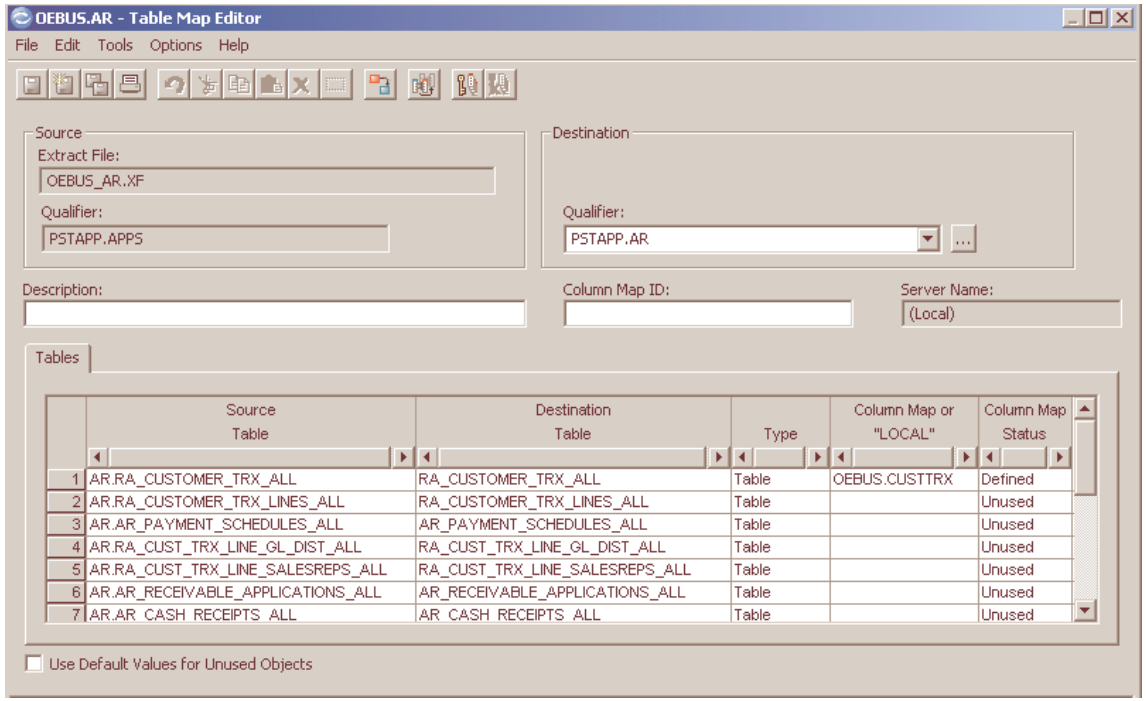
Parent Table	*Child Table*
Sales	**Male_Rates**
Sex	"M"
Age	Age
Sales	**Female_Rates**
Sex	"F"
Age	Age

The Access Definition



- Created dynamically during archive definition
- Use to re-create archive batch job when changes are needed

Table Map



- Map unlike table names, qualifiers
- Exclude individual tables from restore
- Can be saved in Optim Directory

Column Map

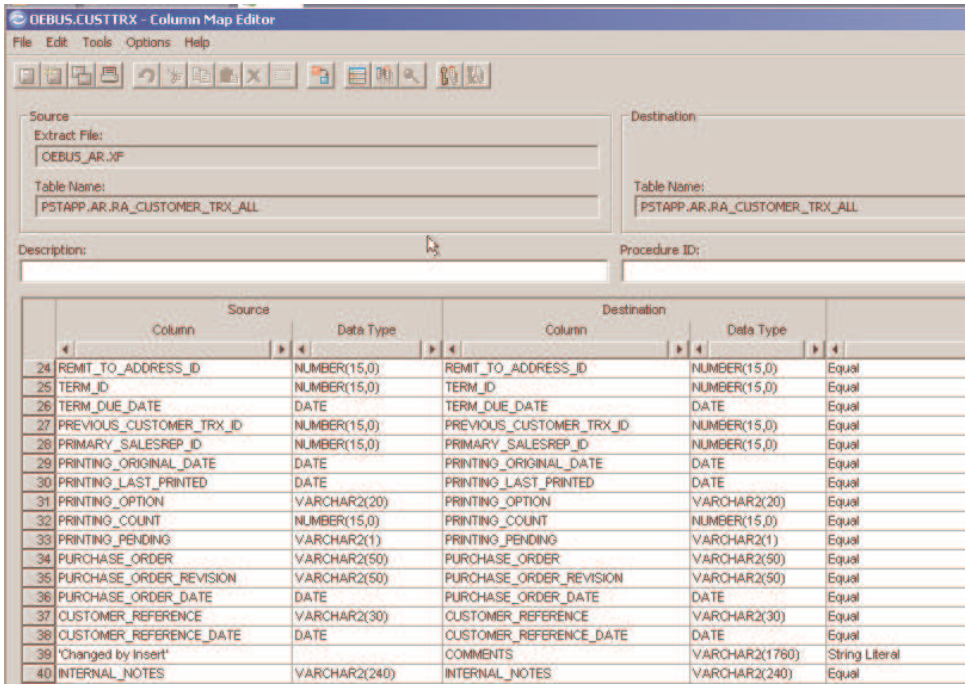
Literals

Special Registers

Expressions

Default Values

User exits



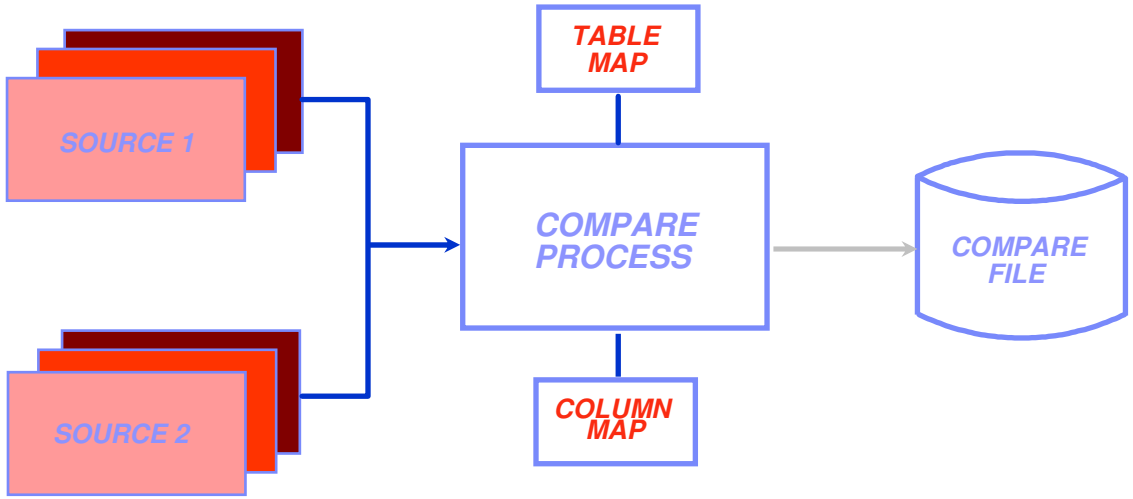
	Source Column	Source Data Type	Destination Column	Destination Data Type	
24	REMIT_TO_ADDRESS_ID	NUMBER(15,0)	REMIT_TO_ADDRESS_ID	NUMBER(15,0)	Equal
25	TERM_ID	NUMBER(15,0)	TERM_ID	NUMBER(15,0)	Equal
26	TERM_DUE_DATE	DATE	TERM_DUE_DATE	DATE	Equal
27	PREVIOUS_CUSTOMER_TRX_ID	NUMBER(15,0)	PREVIOUS_CUSTOMER_TRX_ID	NUMBER(15,0)	Equal
28	PRIMARY_SALESREP_ID	NUMBER(15,0)	PRIMARY_SALESREP_ID	NUMBER(15,0)	Equal
29	PRINTING_ORIGINAL_DATE	DATE	PRINTING_ORIGINAL_DATE	DATE	Equal
30	PRINTING_LAST_PRINTED	DATE	PRINTING_LAST_PRINTED	DATE	Equal
31	PRINTING_OPTION	VARCHAR2(20)	PRINTING_OPTION	VARCHAR2(20)	Equal
32	PRINTING_COUNT	NUMBER(15,0)	PRINTING_COUNT	NUMBER(15,0)	Equal
33	PRINTING_PENDING	VARCHAR2(1)	PRINTING_PENDING	VARCHAR2(1)	Equal
34	PURCHASE_ORDER	VARCHAR2(50)	PURCHASE_ORDER	VARCHAR2(50)	Equal
35	PURCHASE_ORDER_REVISION	VARCHAR2(50)	PURCHASE_ORDER_REVISION	VARCHAR2(50)	Equal
36	PURCHASE_ORDER_DATE	DATE	PURCHASE_ORDER_DATE	DATE	Equal
37	CUSTOMER_REFERENCE	VARCHAR2(30)	CUSTOMER_REFERENCE	VARCHAR2(30)	Equal
38	CUSTOMER_REFERENCE_DATE	DATE	CUSTOMER_REFERENCE_DATE	DATE	Equal
39	'Changed by Insert'		COMMENTS	VARCHAR2(1760)	String Literal
40	INTERNAL_NOTES	VARCHAR2(240)	INTERNAL_NOTES	VARCHAR2(240)	Equal

- Map unlike column names
- Datatype conversions
- Populate new destination columns

Information Management Optim

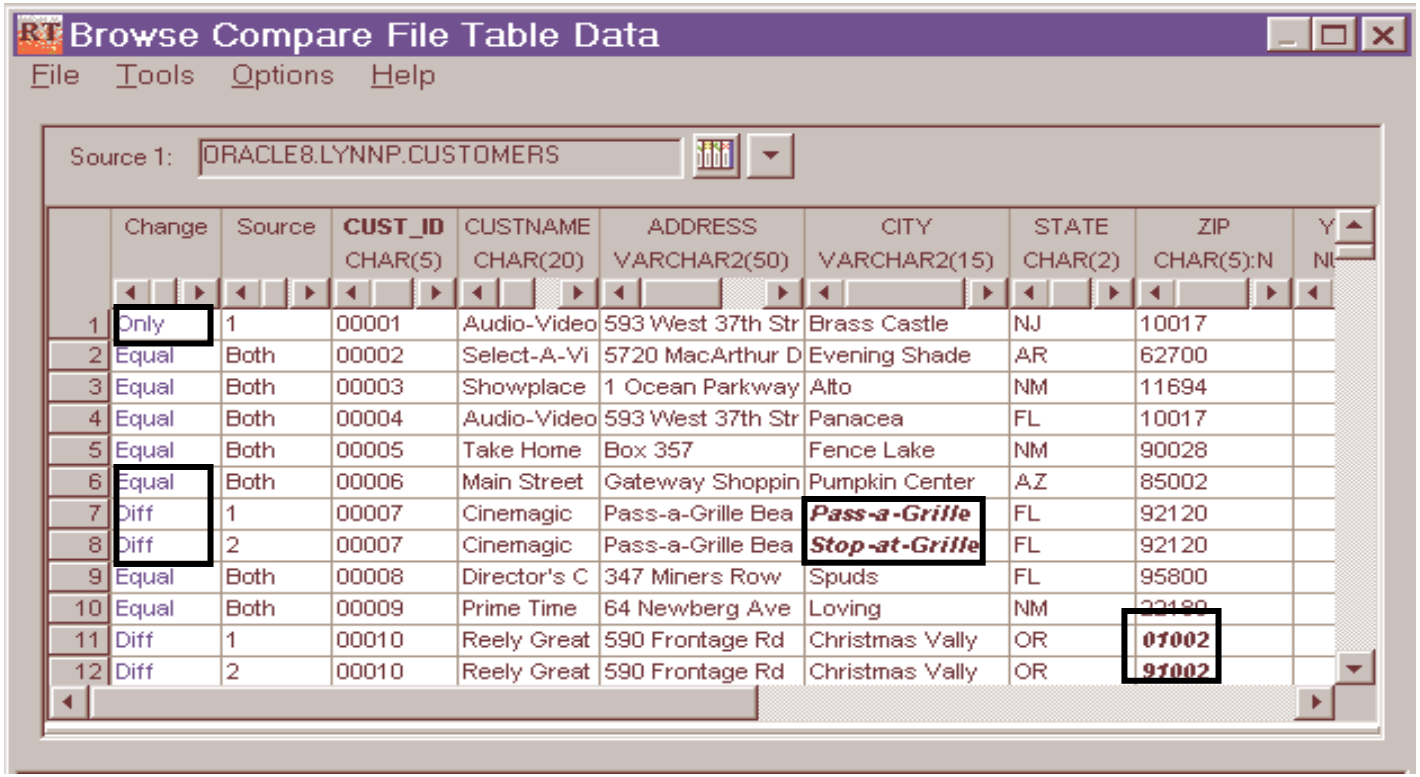
IBM Optim Compare

OPTIM Relational Compare Facility



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

Browsing the Compare File



	Change	Source	CUST_ID CHAR(5)	CUSTNAME CHAR(20)	ADDRESS VARCHAR2(50)	CITY VARCHAR2(15)	STATE CHAR(2)	ZIP CHAR(5):N
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	Cinemagic	Pass-a-Grille Bea	***Pass-a-Grille***	FL	92120
8	Diff	2	00007	Cinemagic	Pass-a-Grille Bea	***Stop-at-Grille***	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	[illegible]
11	Diff	1	00010	Reely Great	590 Frontage Rd	Christmas Vally	OR	***01002***
12	Diff	2	00010	Reely Great	590 Frontage Rd	Christmas Vally	OR	***91002***

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

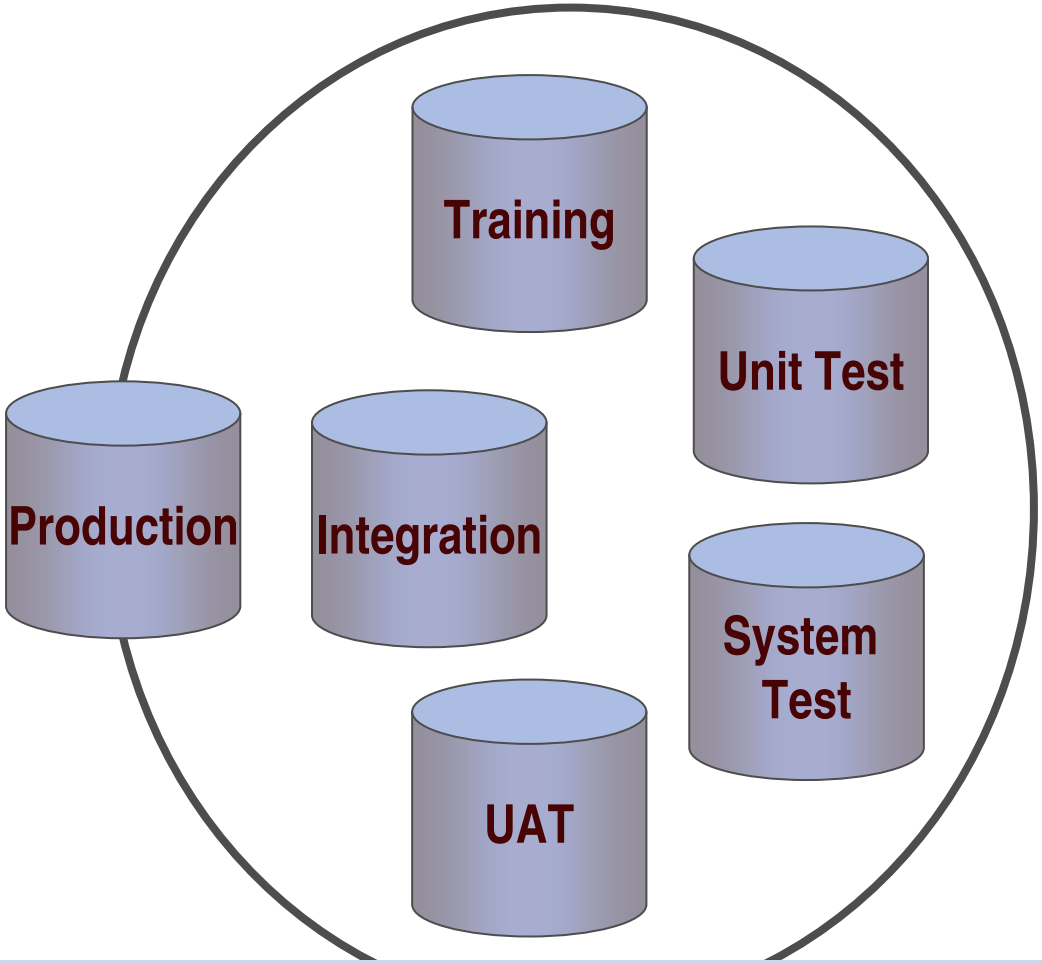
Information Management Optim

IBM Optim
Test Data Management

The Symptoms of Poor Testing Strategies

- **Management notices that new application functionality is delayed three months**
- **The business is unable to compete for customers because their software lacks "state-of-the-art" functionality**
- **The CFO is complaining over how high the IT budget has become to fix application defects**
- **Developers are sitting around waiting for their copy of the database to work with**

How Does Test Data Management Impact Cost?



Production	500GB
Training	500GB
Unit Test	500GB
System Test	500GB
UAT	500GB
Integration	500GB
Total	3 TB

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

Some Current Practices

#1 - Clone Production

Clone Production

Request for Copy

Wait

After Production Database Copy

Changes

Production Database Copy

Manual examination:
Right data?
What Changed?
Correct results?
Unintended Result?
Someone else modify?

#2 – Write SQL

Write SQL

Extract

- Complex
- Subject to Change

Extract

Changes

After

- RI Accuracy?
- Right Data?

Expensive, Dedicated Staff, Ongoing Responsibility.

Share test database with everyone else

Test Data Management – Concepts

Test Data Management (TDM) refers to the need to manage data used in various pre- production environments and is a vital part of Application Quality & Delivery

Extract production data into referentially intact data subsets to be used to support application data in other environments

De-identify (mask) extracted production data to protect privacy

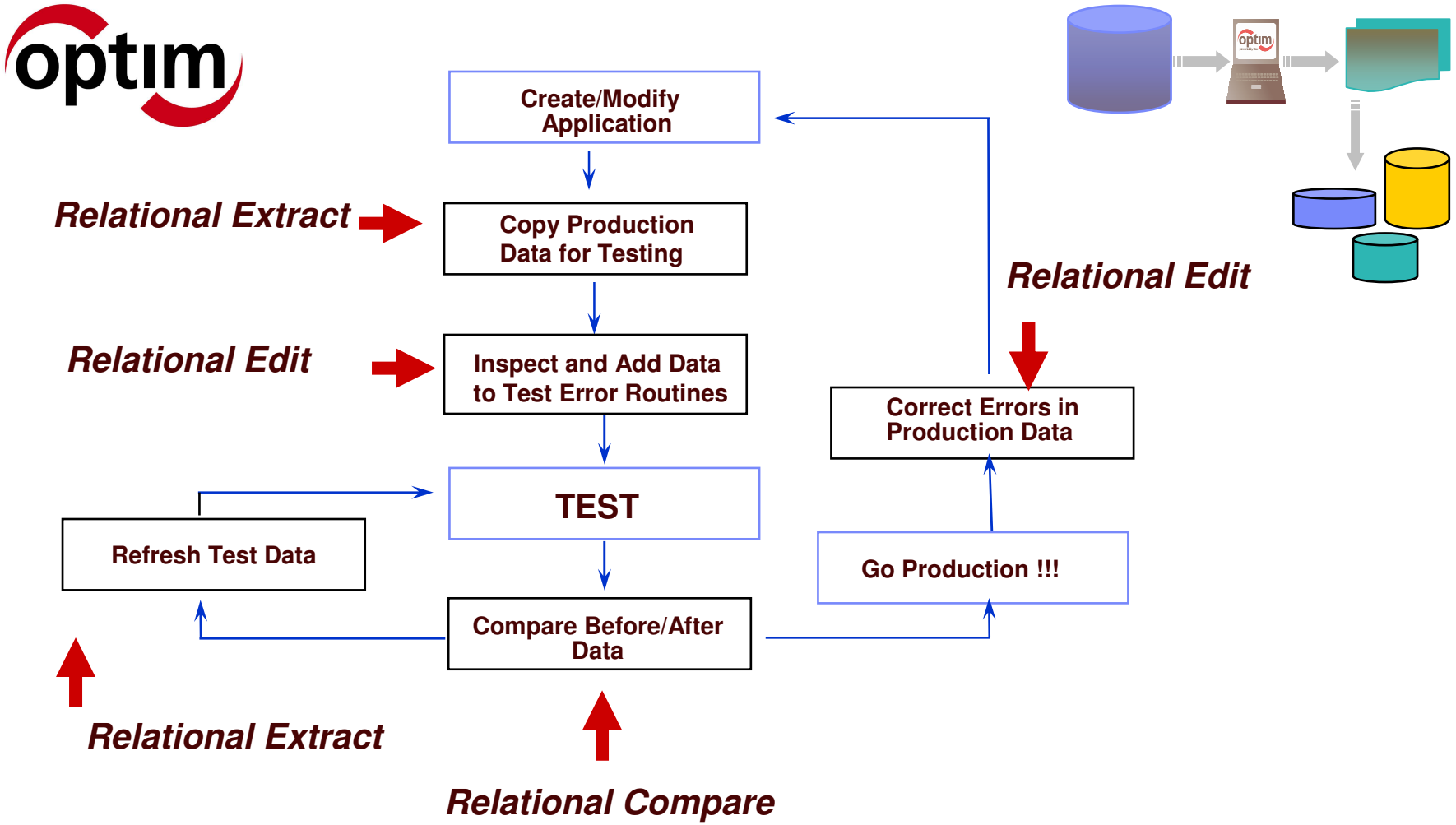
Compare “before” and “after” images of test data

Speed application quality and delivery

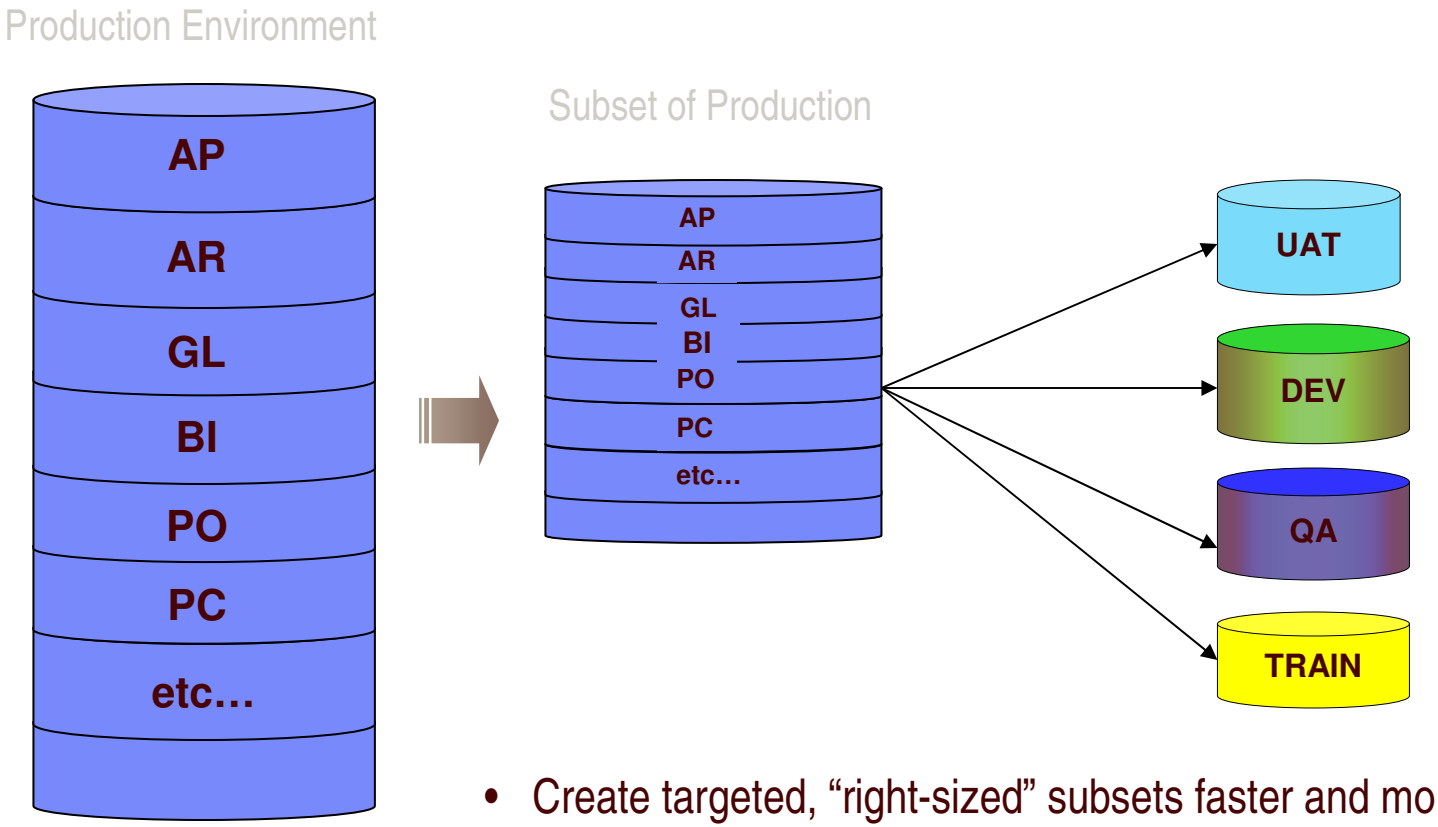
Key Requirements for a Test Data Management Solution

1. **Subset capabilities to create realistic and manageable test databases**
2. **Easily refresh test environments**
3. **Edit data to create targeted test cases**
4. **Compare ‘before’ and ‘after’ images of the test data**
5. **De-identify (mask) data to protect privacy**

Product Overview : Optim Test Data Management



Optim Test Data Management using Optim Subsetting:



- Create targeted, “right-sized” subsets faster and more efficiently than cloning
- Compare to pinpoint and resolve application defects faster
- Improve development efficiencies

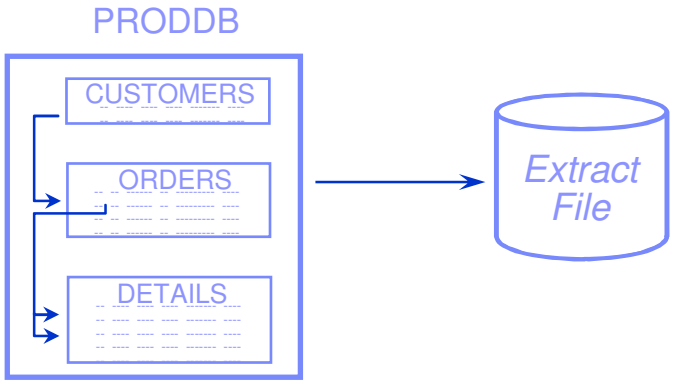
Defining the Extract…..

Tables

Views

Synonyms

Aliases



Required:

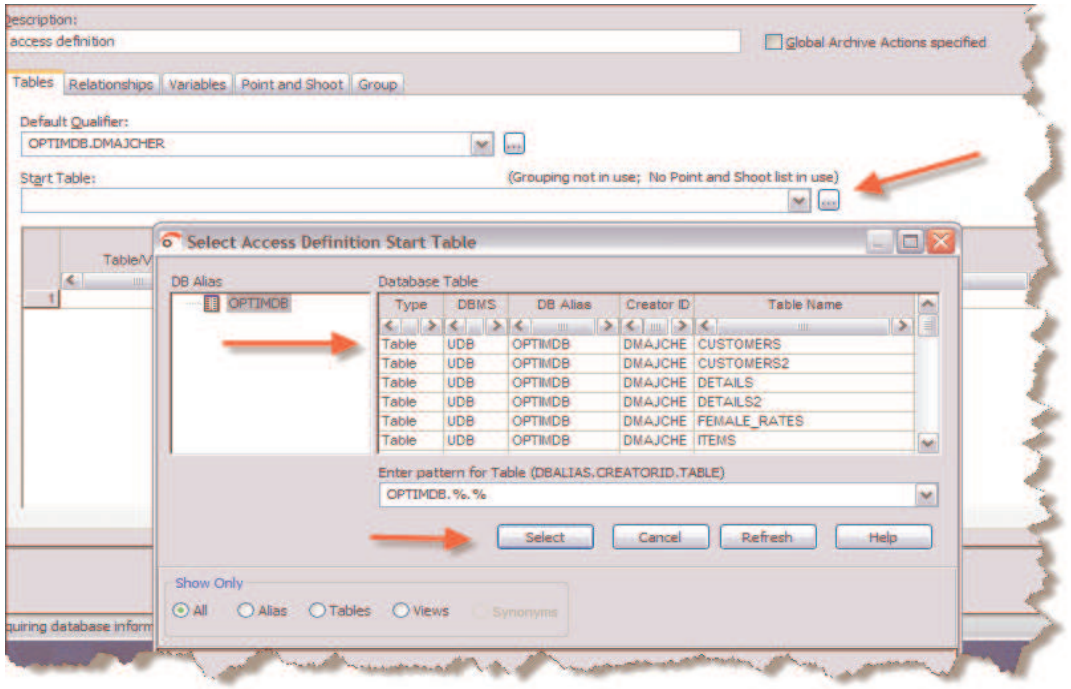
- **Start Table**
- **Set of Tables**

Optional:

- Selection Criteria
- Data Sampling
- Data Grouping
- Point and Shoot
- Relationship Usage

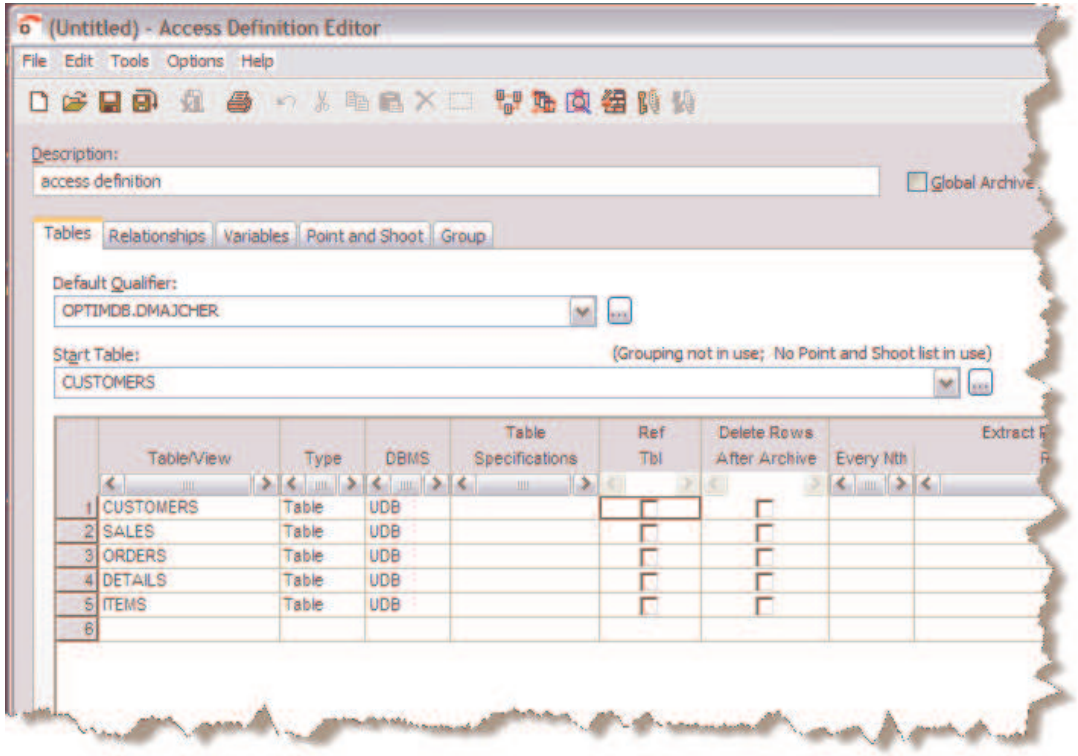
Extract Process

- **Identify the Start Table**
- **Choose from a list or type in a known table name**



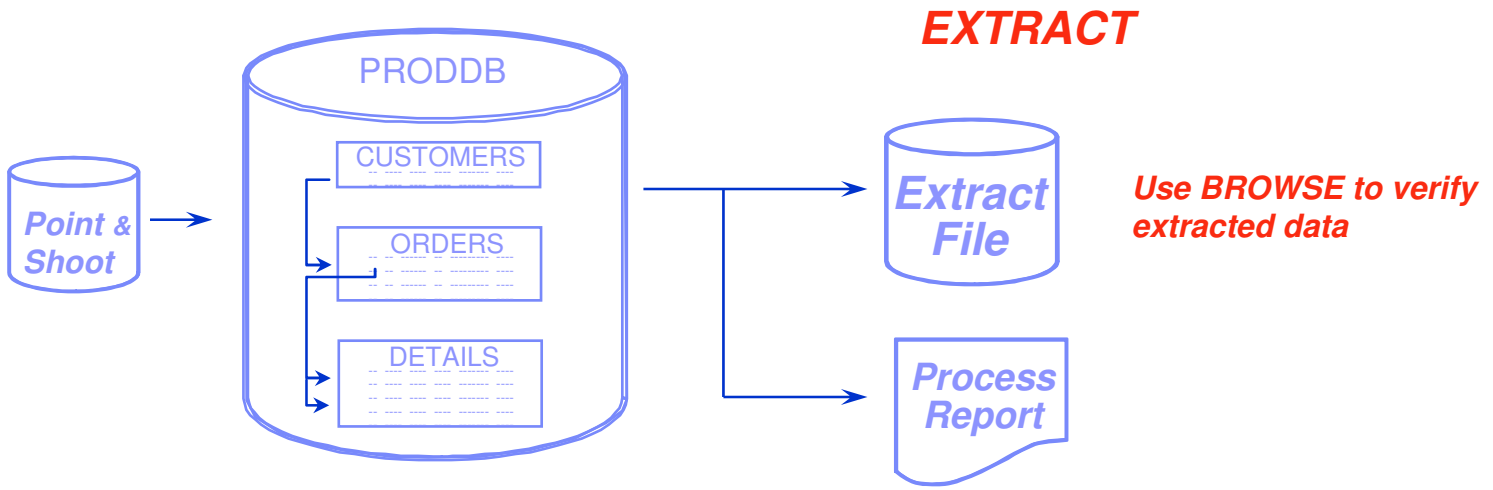
Extract Process

Defining the Access Definition



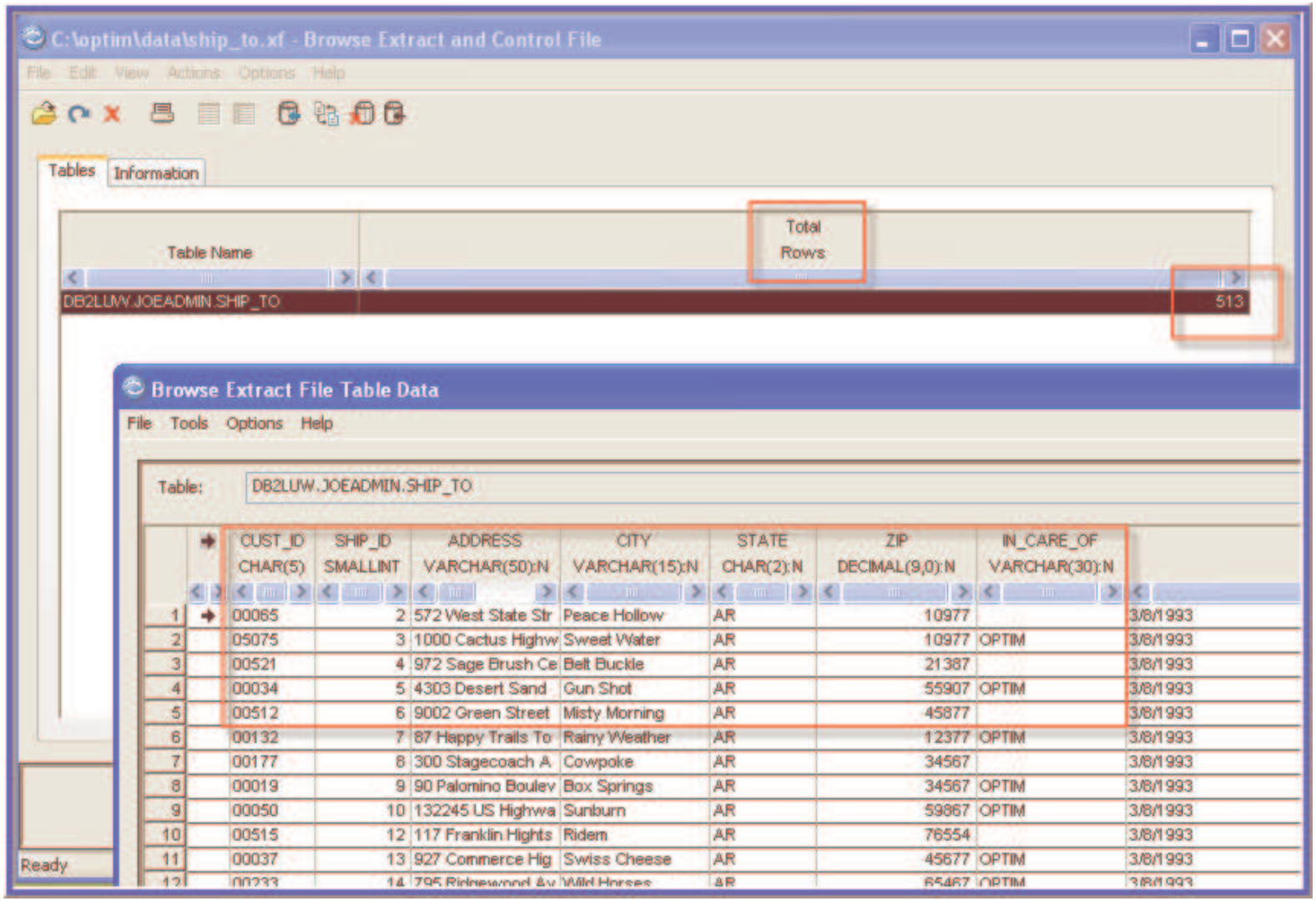
- **Include random selection factor, extract limits and selection criteria**
- **Use the RELATED functions to populate list with other tables**

Extract Process



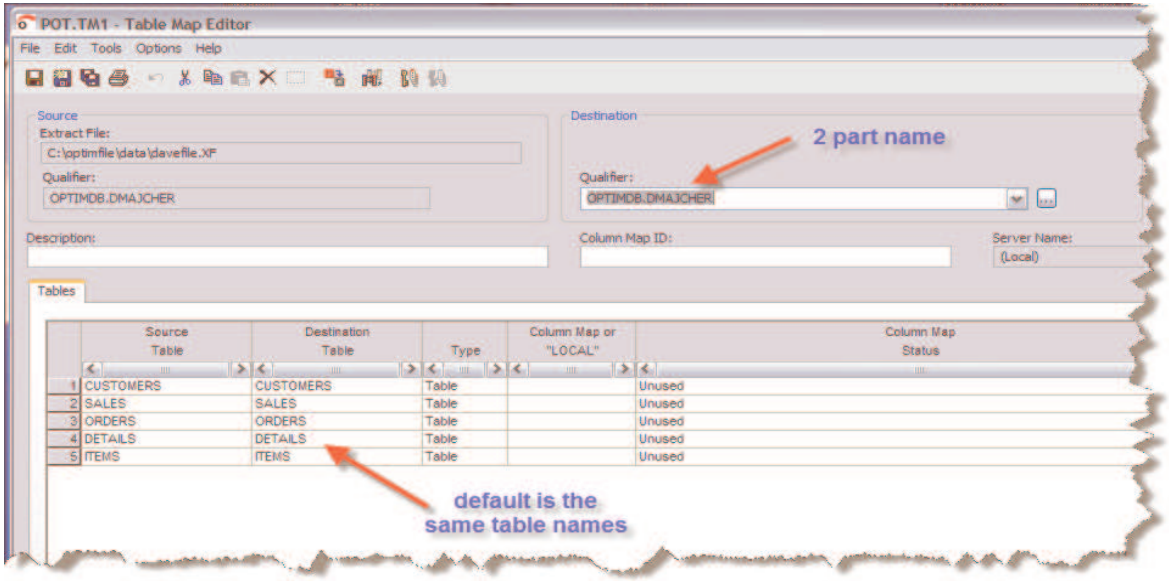
- **Extract from source tables**
  - using dynamic SQL
- **Extract data and/or object definitions**

Browse Extract file



Populate Destination Tables

- **Table Map**
  - Table names need not match
  - Change qualifier and/or table name
  - Can be saved in Optim Directory



Populate Destination Tables

- **Column Map**
  - Map unlike column names
  - Transform/mask sensitive data
  - Datatype conversions
  - Column-level date aging

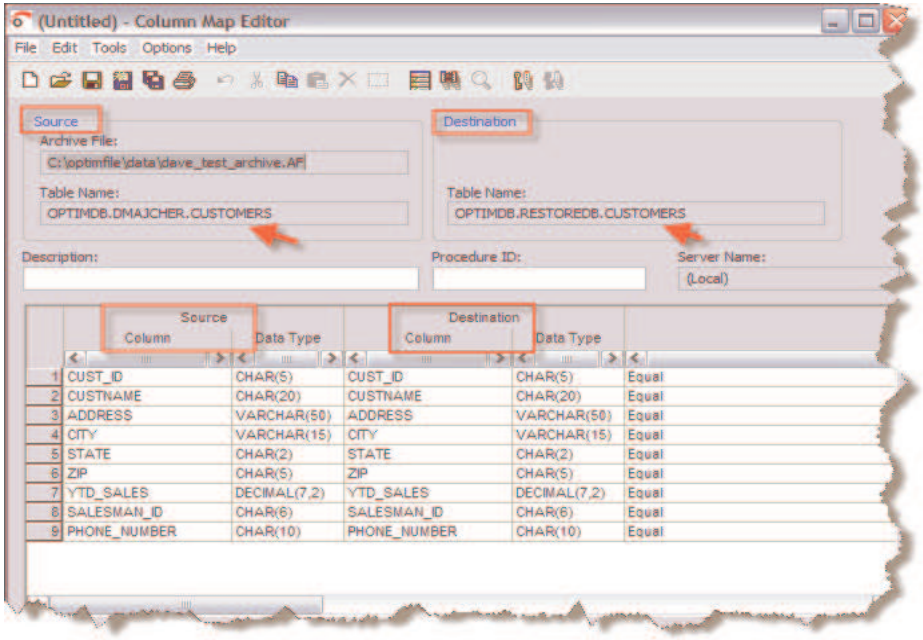
Literals

Special Registers

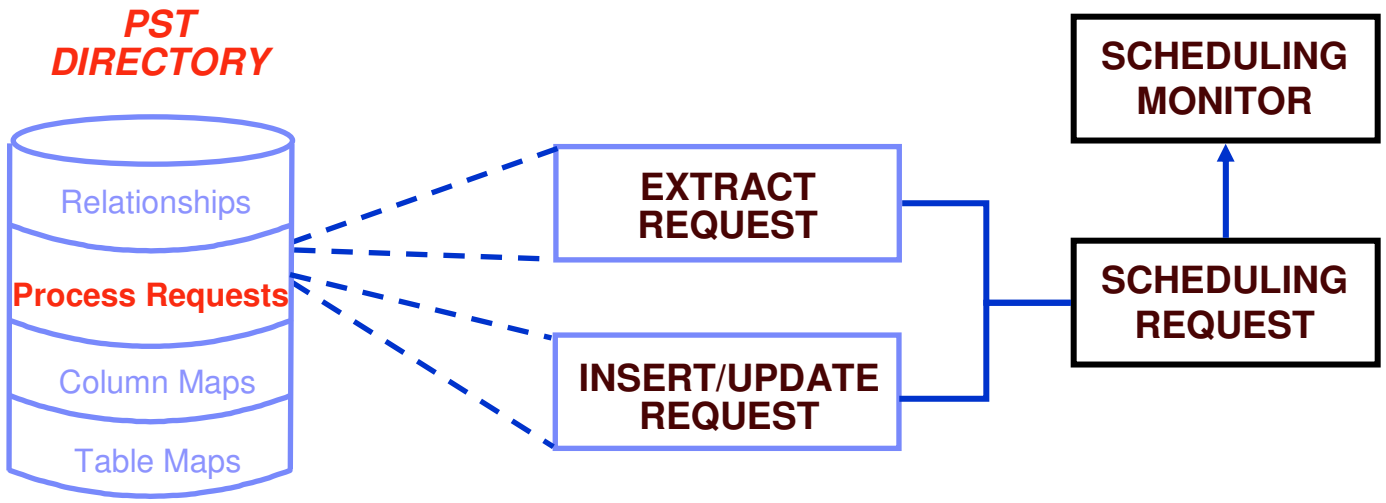
Expressions

Default Values

User exits



Scheduling



- **Package saved Process Requests for a complete job**
- **Schedule requests for automated operation**
- **Command line interface available**

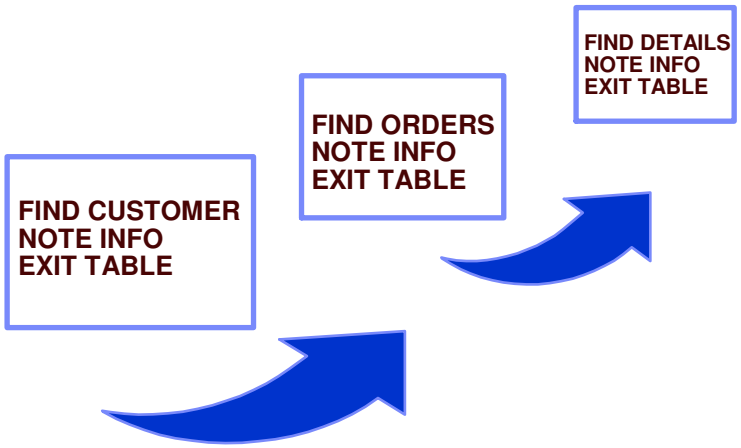
Information Management Optim

IBM Optim Editor

Traditional vs. Relational Tools

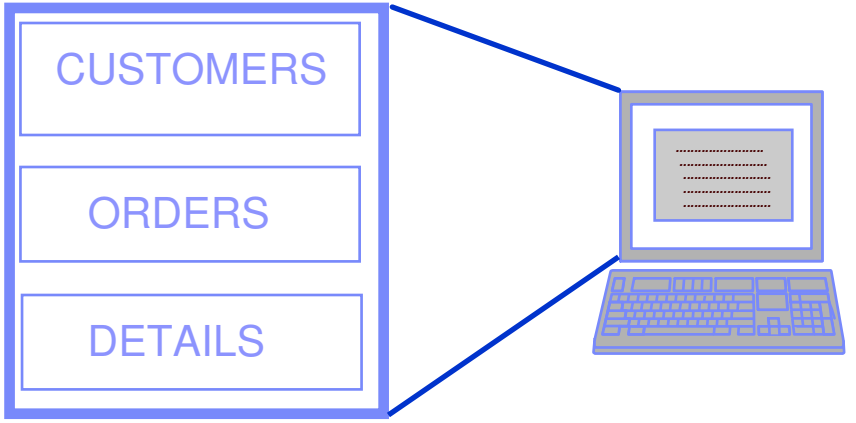
Single Table Editors

- One table/view at a time
- No edit of related data from multiple tables

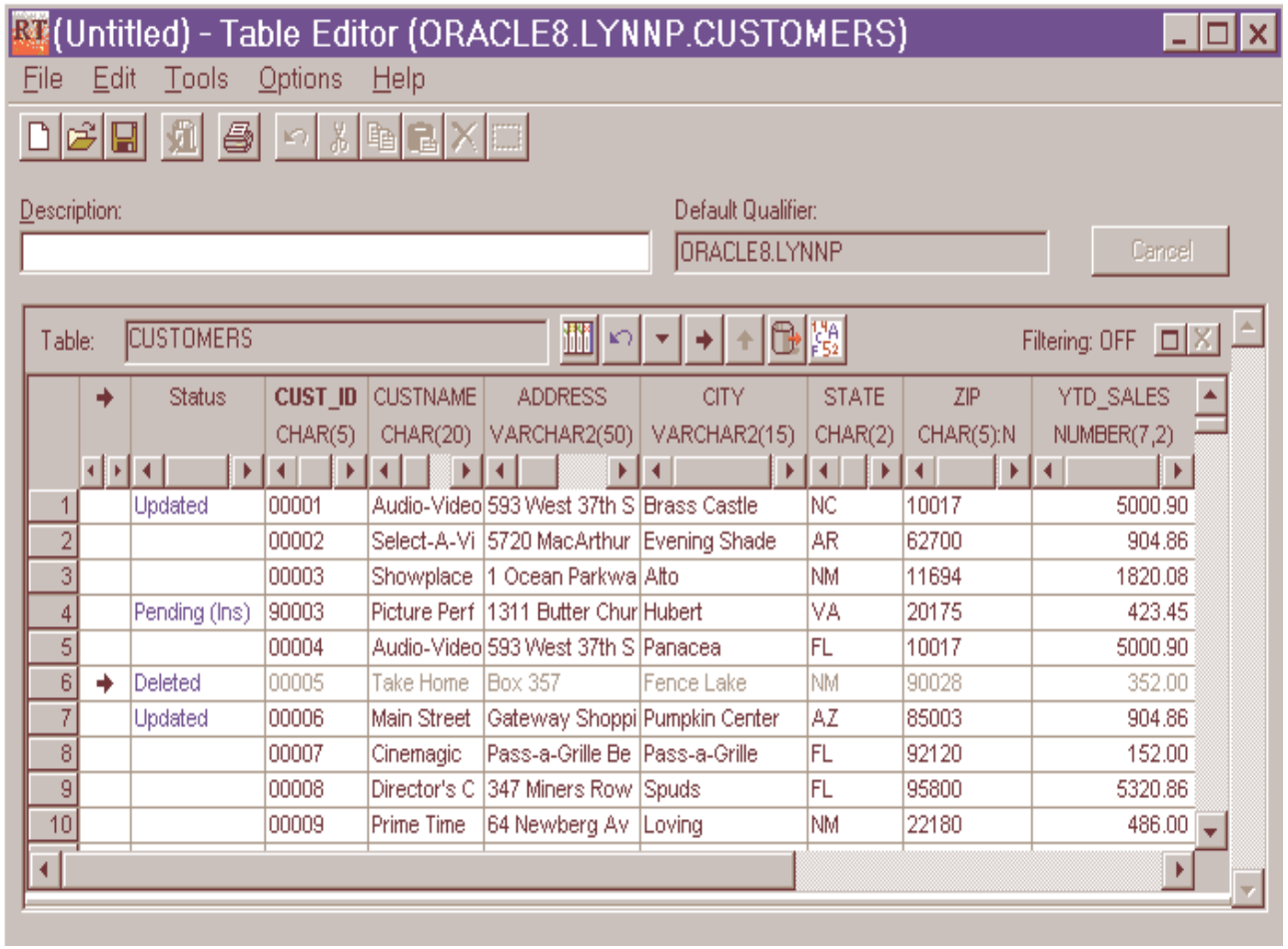


The Relational Editor

- ***Simultaneous*** **browse/edit of related data from multiple tables**



Editing Data



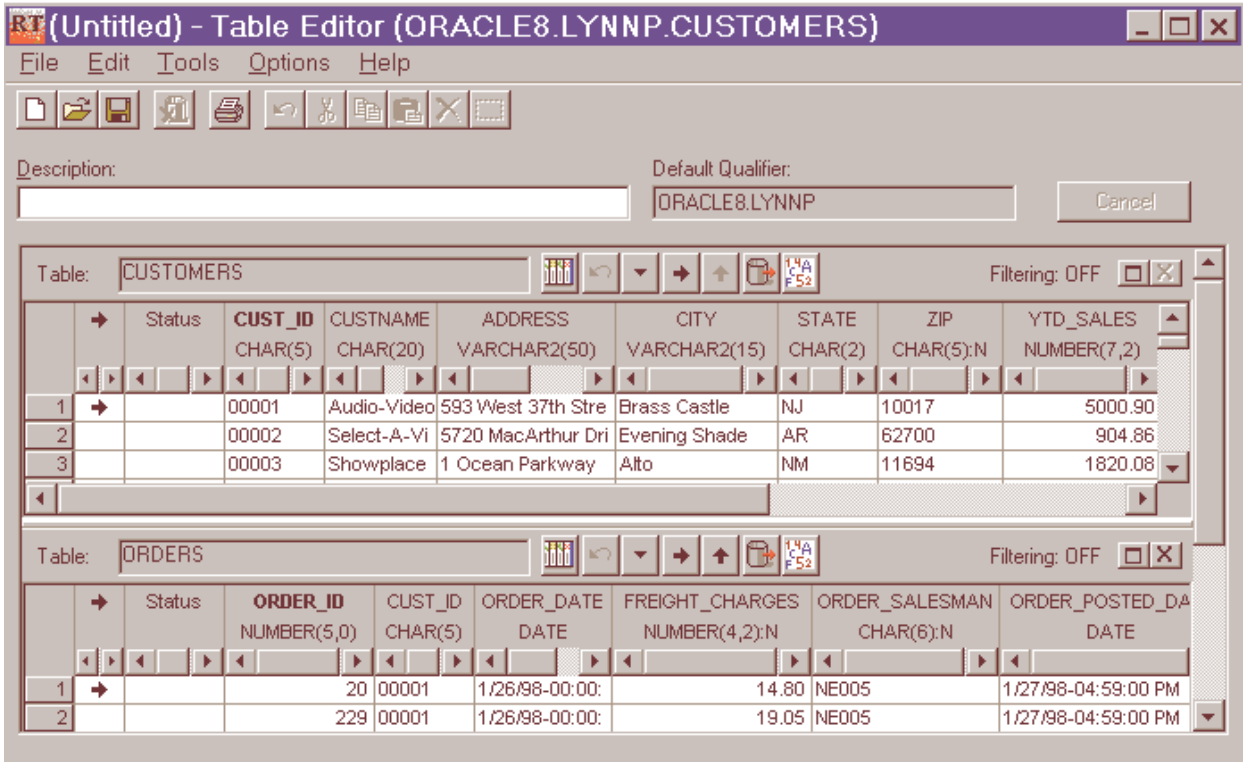
	Status	CUST_ID CHAR(5)	CUSTNAME CHAR(20)	ADDRESS VARCHAR2(50)	CITY VARCHAR2(15)	STATE CHAR(2)	ZIP CHAR(5):N	YTD_SALES NUMBER(7,2)
1	Updated	00001	Audio-Video	593 West 37th S	Brass Castle	NC	10017	5000.90
2		00002	Select-A-Vi	5720 MacArthur	Evening Shade	AR	62700	904.86
3		00003	Showplace	1 Ocean Parkwa	Alto	NM	11694	1820.08
4	Pending (Ins)	90003	Picture Perf	1311 Butter Chur	Hubert	VA	20175	423.45
5		00004	Audio-Video	593 West 37th S	Panacea	FL	10017	5000.90
6	Deleted	00005	Take Home	Box 357	Fence Lake	NM	90028	352.00
7	Updated	00006	Main Street	Gateway Shoppi	Pumpkin Center	AZ	85003	904.86
8		00007	Cinemagic	Pass-a-Grille Be	Pass-a-Grille	FL	92120	152.00
9		00008	Director's C	347 Miners Row	Spuds	FL	95800	5320.86
10		00009	Prime Time	64 Newberg Av	Loving	NM	22180	486.00

Edit data to:

- Insert Rows
- Delete Rows
- Update Rows

Relationally Joined Data

- Browse or edit related rows
- Scroll of higher-level table automatically synchronizes all lower-joined tables



Information Management Optim

Optim Test Data Management Lab

Information Management Optim

IBM Optim Data Privacy

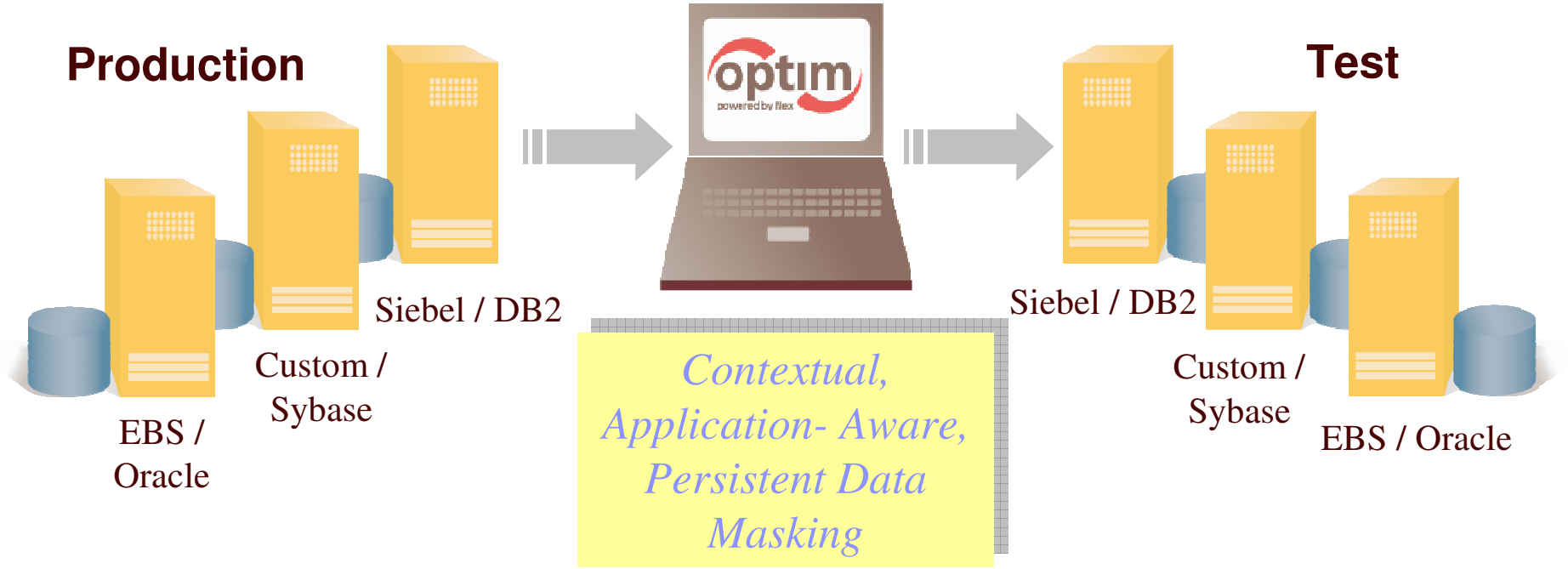
Challenges of Enterprise Data Privacy

- **Multi-platforms**
- **Relational database applications in the enterprise**
  - Complex data model
  - Multiple databases
  - Legacy data components
  - Interconnected applications
- **Distributed work teams**
  - Employees and contractors
  - Global 24 x 7 operations

How does Optim Protect Privacy?

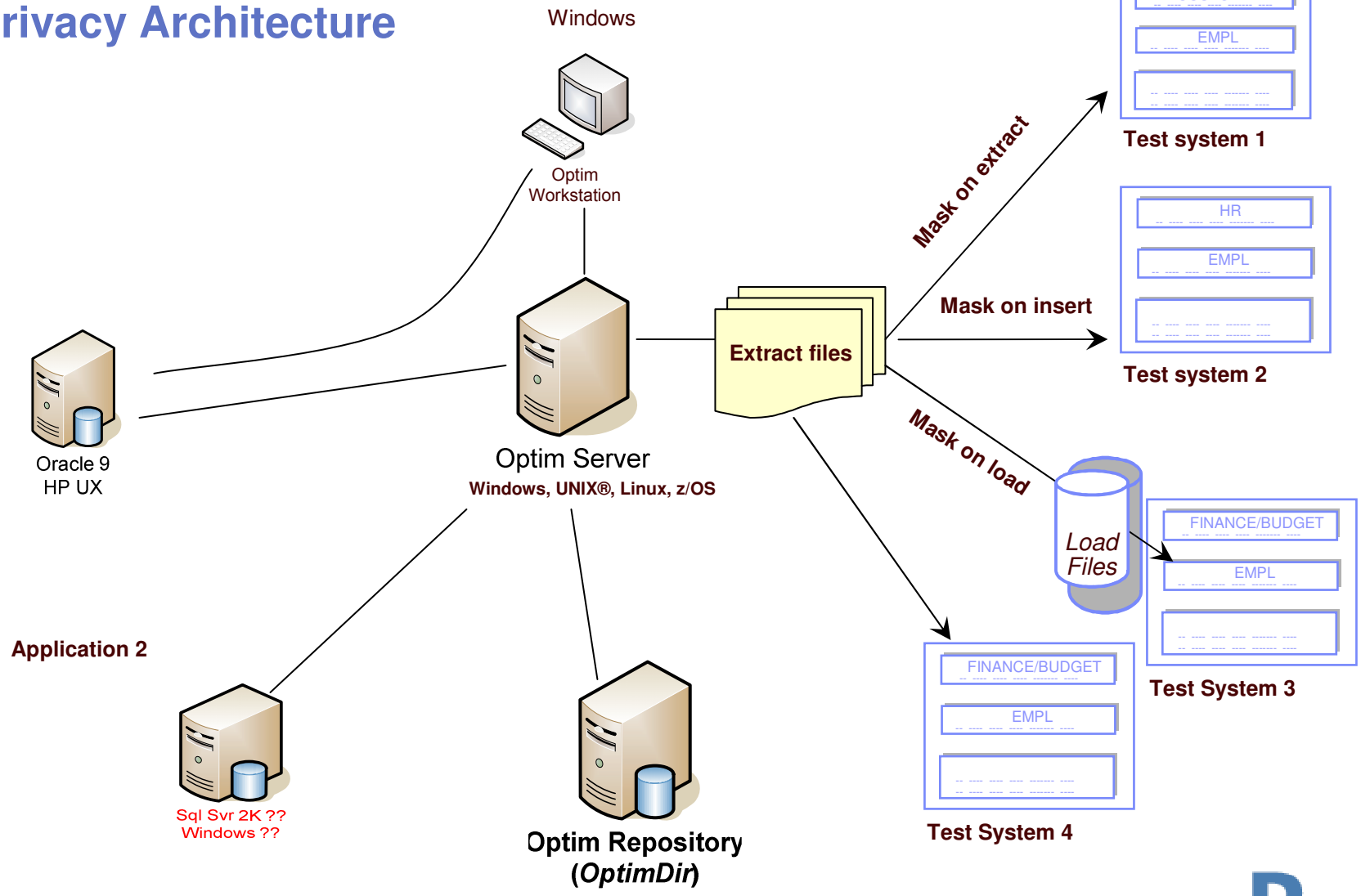
- **IBM Optim provides the fundamental components of test data management and enables organizations to** *de-identify, mask and transform* **sensitive data**
- **Companies can apply a range of transformation techniques to substitute customer data with** *contextually-accurate but fictionalized data* **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**

Optim Data Privacy Solution

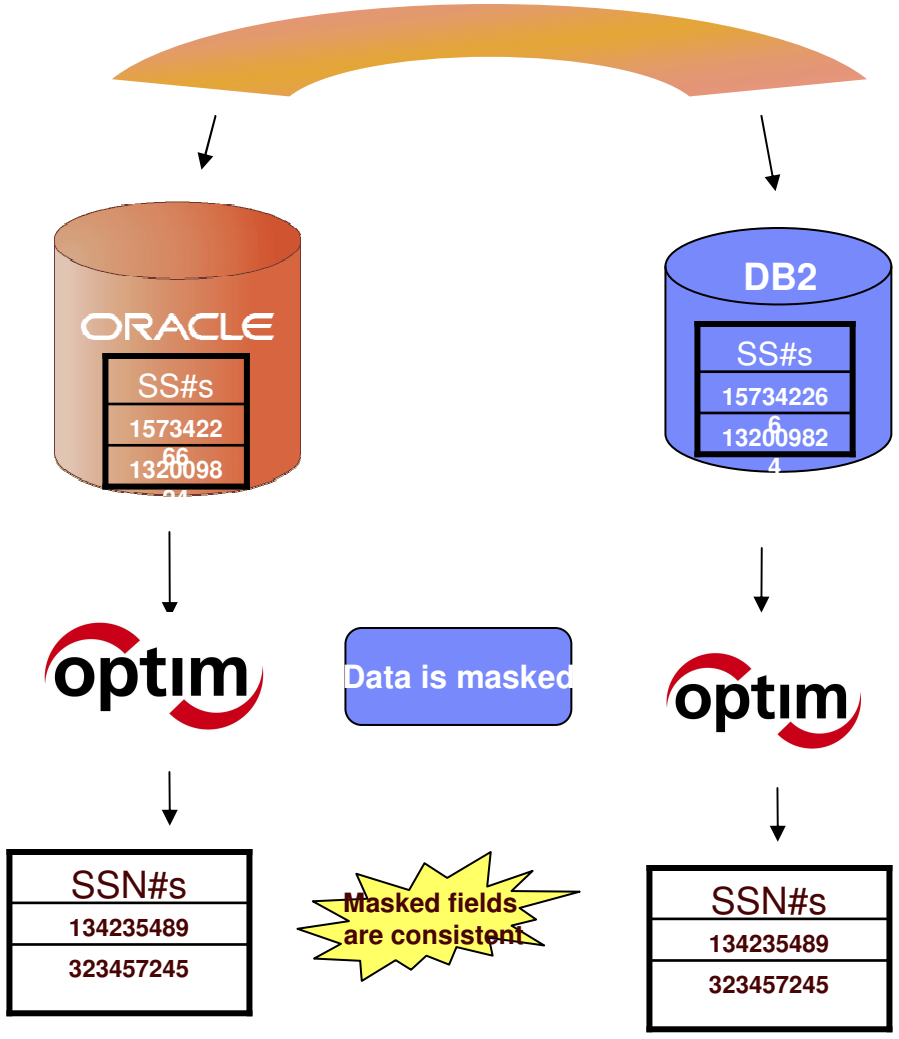


- **Substitute confidential information with fictionalized data**
- **Deploy multiple masking algorithms**
- **Provide consistency across environments and iterations**
- **Enable off-shore testing**
- **Protect private data in non-production environments**

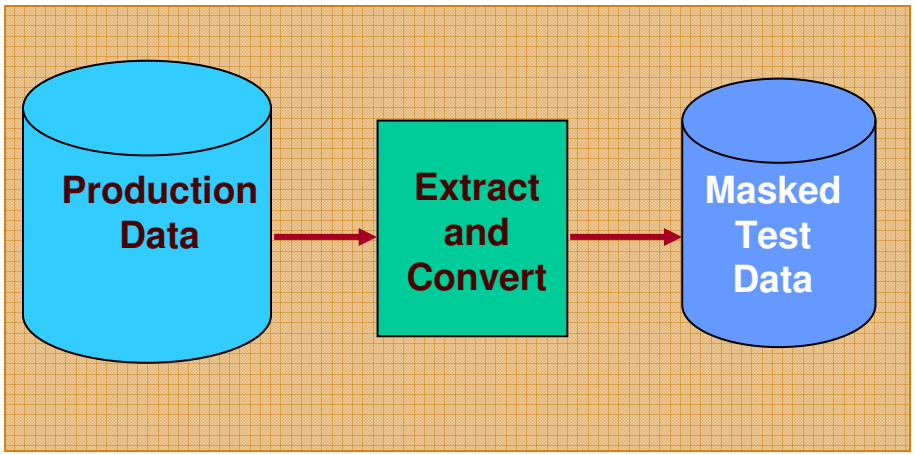
Conceptual Optim Test Data Management/Optim Data Privacy Architecture



Consistent Masking across the Enterprise



De-Identify test data



During Extract Process

Or

Standalone Convert Process

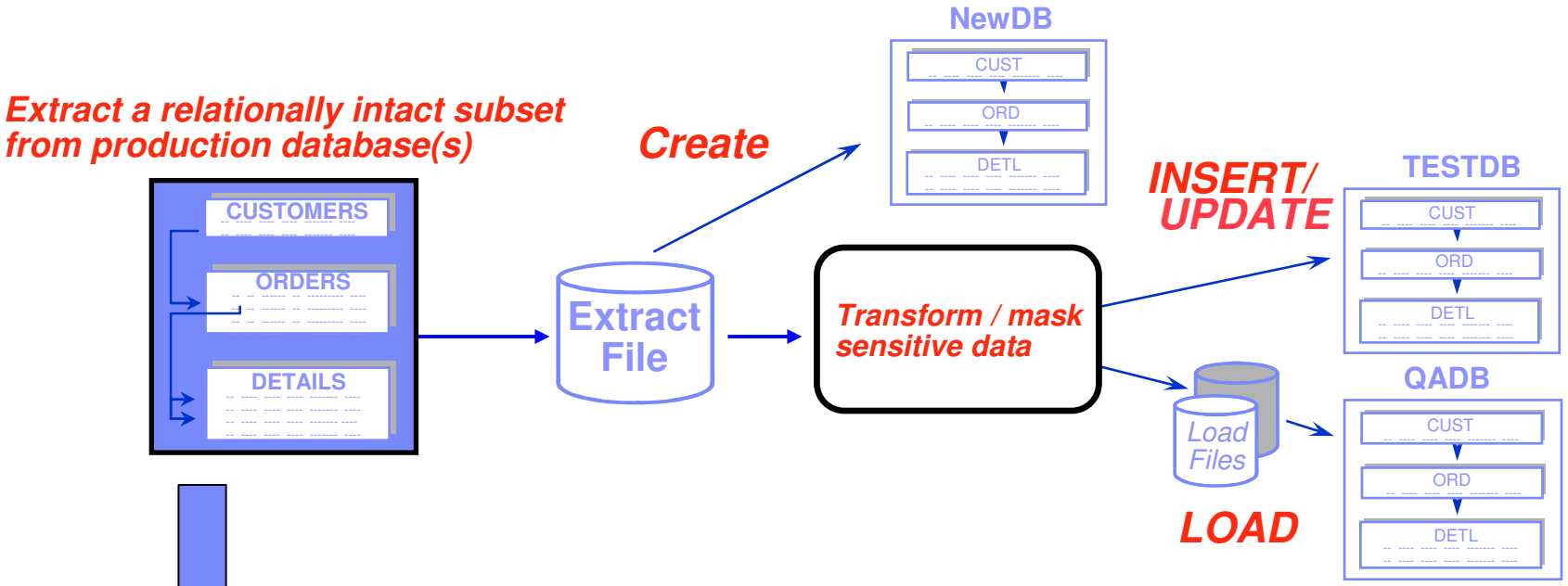
Or

During Insert/Load Process

Transform or Replace sensitive data using

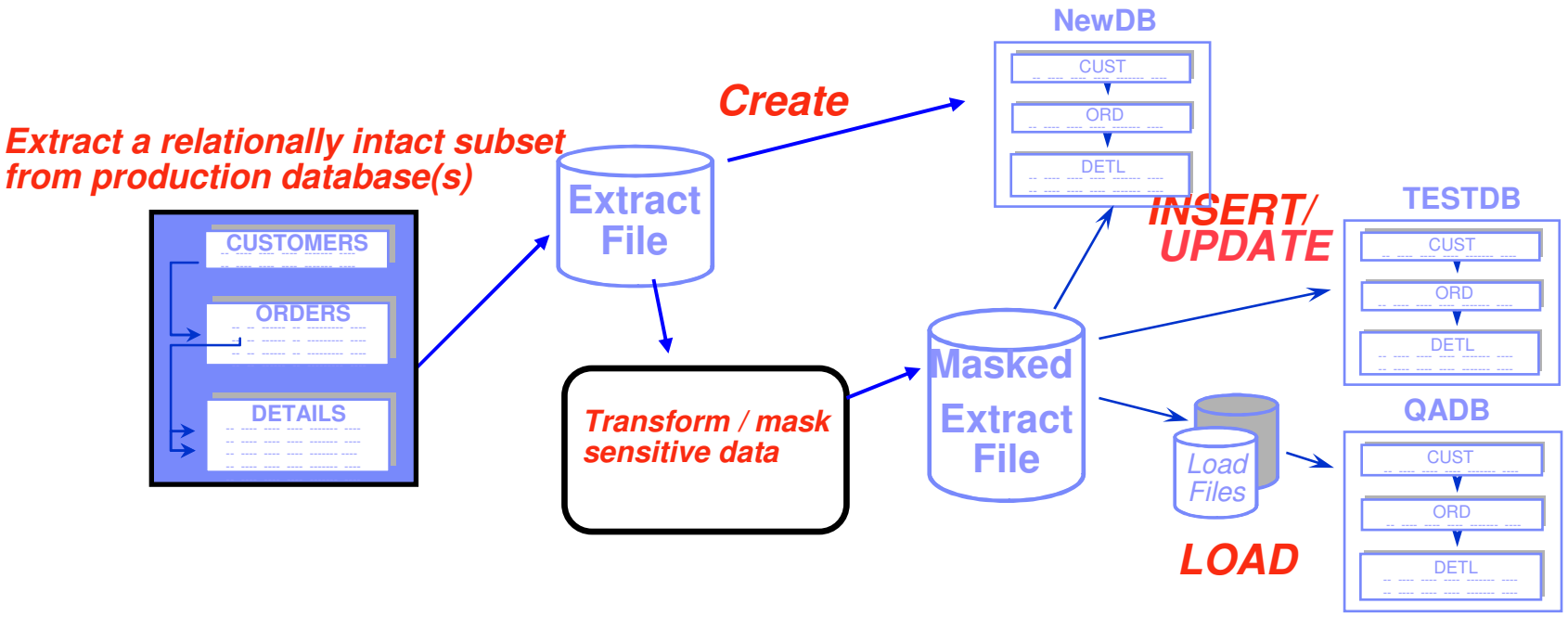
- Standard mapping rules: Literals, Special Registers, Expressions, Default Values, Look-up tables
- Complex mapping rules: User exits

Optim Data Privacy in Application Testing



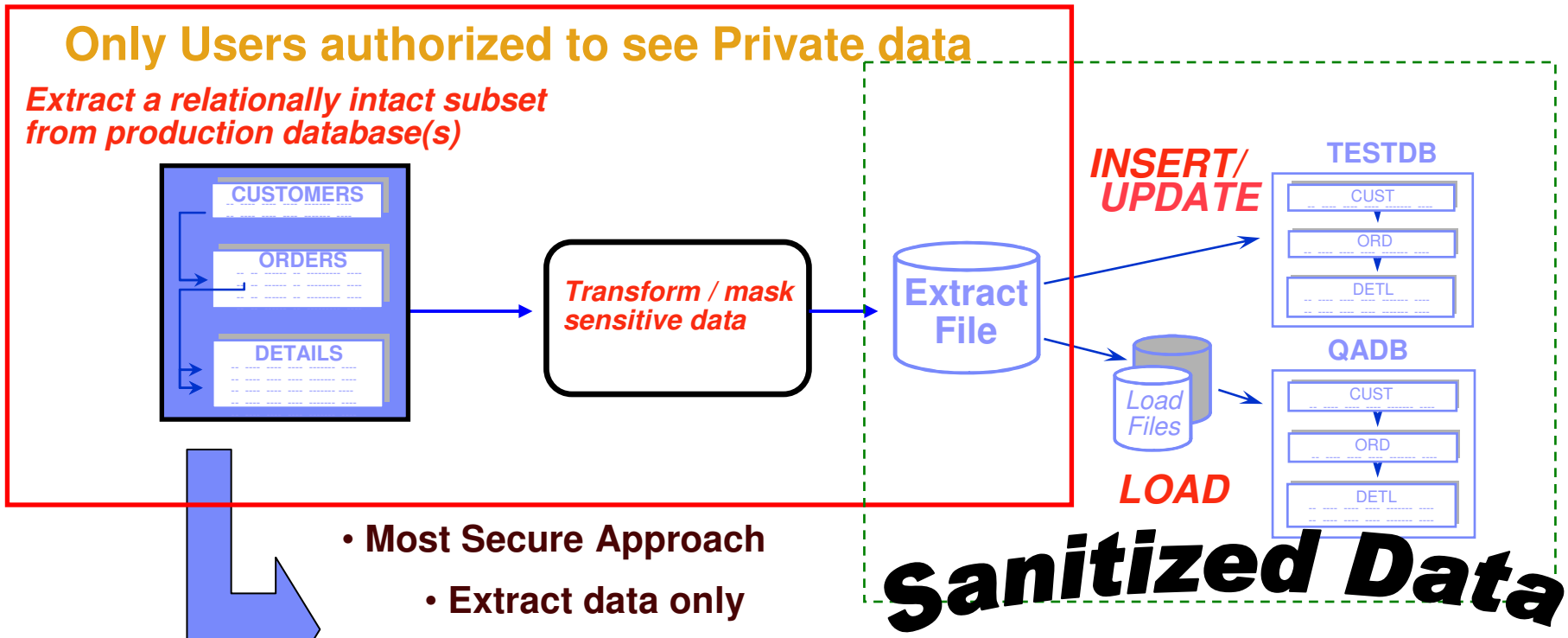
- Extract data and/or object definitions
- Define a new set of test tables
- Apply masking during population process
- Extract file may be reused but contains un-Masked data
- Good practice for testing masks

Optim Data Privacy in Application Testing



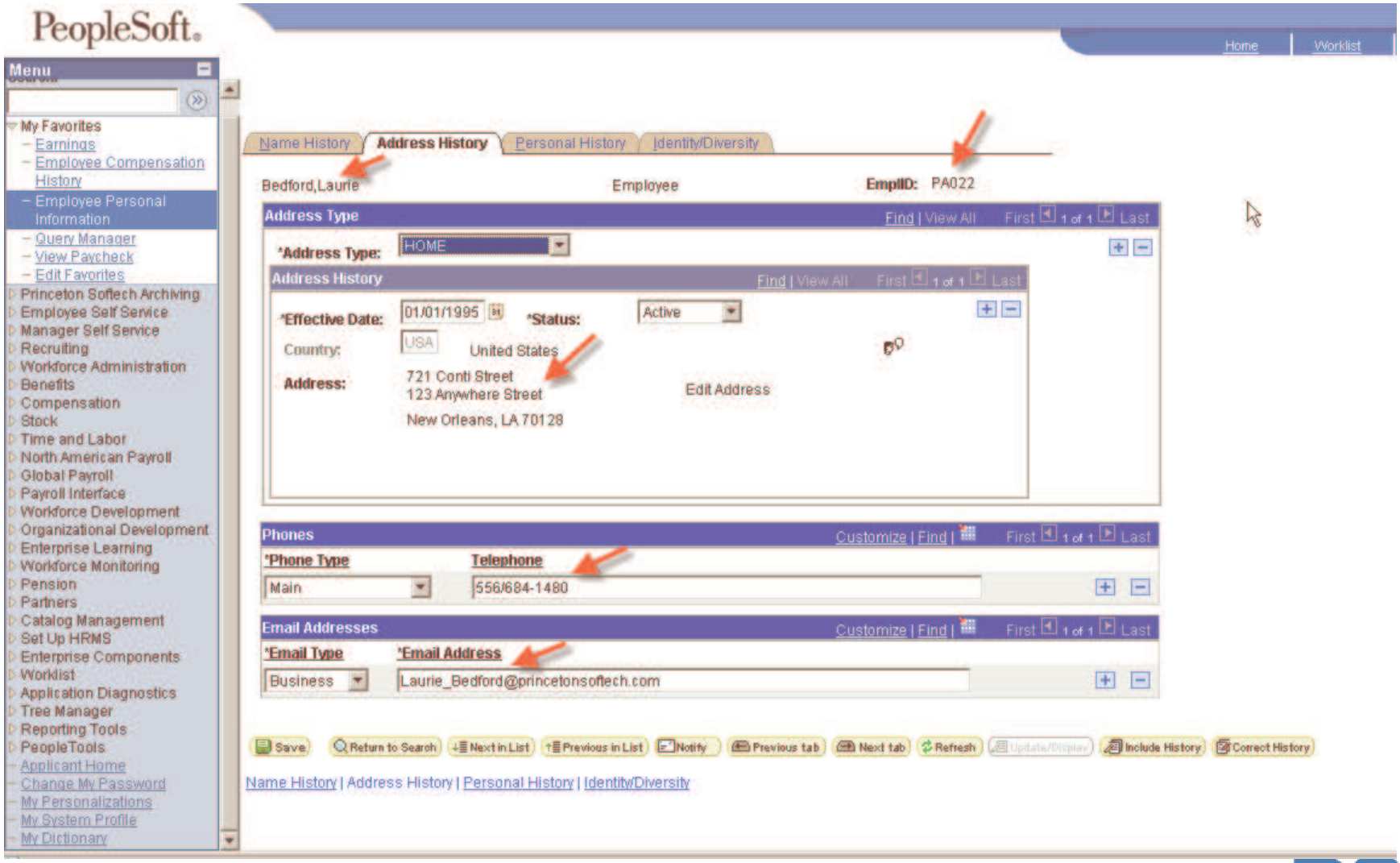
- Extract data and/or object definitions in pre-masked file
- Use pre-masked Extract file to create new set of tables
- Convert Pre-masked extract file data into second masked extract file
- Share masked extract file to be reused for population step
- Good practice for testing masks using COMPARE

Optim Data Privacy in Application Testing

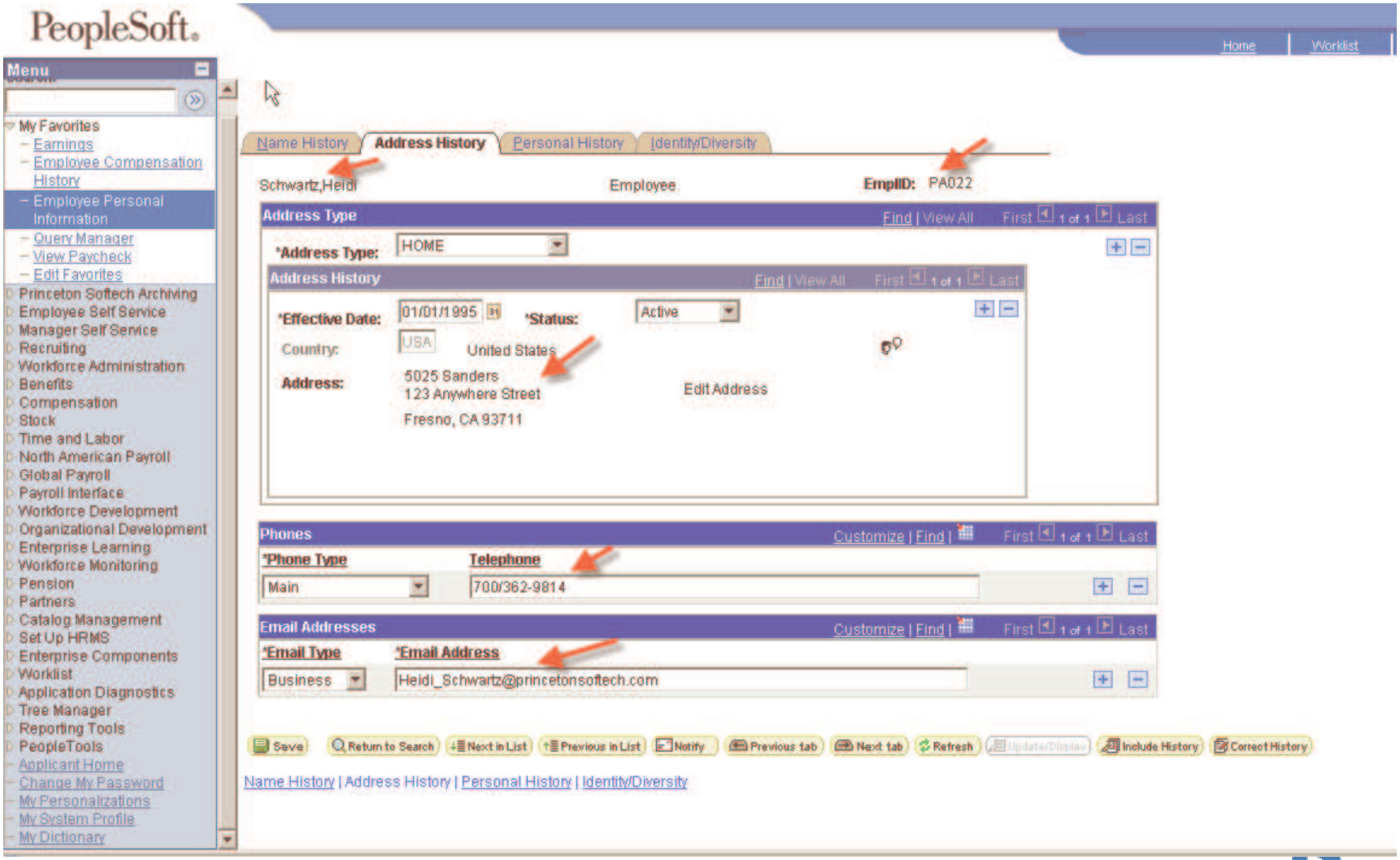


- Most Secure Approach
  - Extract data only
  - Convert during extract
- Extract file already contains masked data
  - Can be shared with testers to reuse

Before Data Masking

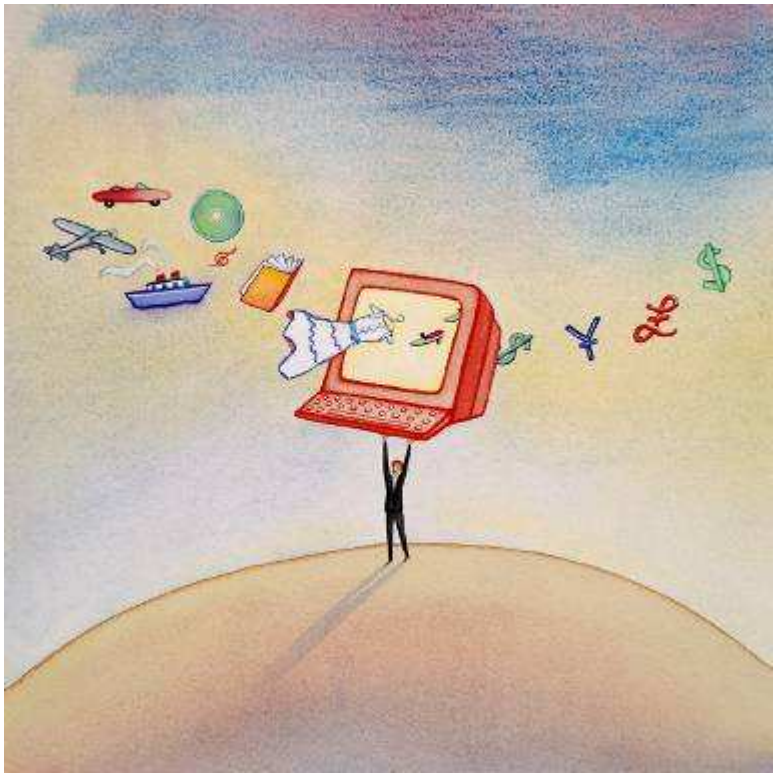


After Data Masking



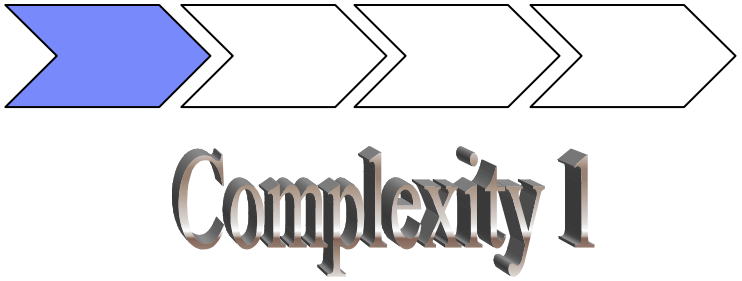
Transformation Techniques

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

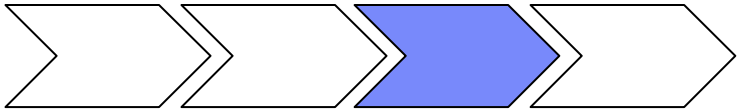


Example: Bank Account Numbers

- **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"



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:**

  "11110 Campus Drive Princeton, NJ 08541"

  becomes…

  "1223 E. 12th Street NY, NY 10079"

  - **Optim ships with over 100,000 valid addresses**

Complexity 2

Street Address/City/State/Zip Code Data Sets

Total Assets	Customers	Street	City	State	Zip Code
$534,674,233	**54,999**	**12 Buttercup Ln**	**Cleveland**	**OH**	**44101**
$8,777,733,811	**105,333**	**6767 Rte 10 S**	**Princeton**	**NJ**	**08594**

1) Client is a Bank who wishes to mask its assets by location

Address Lookup Table

288 Helm St	**Milwaukee**	**WI**	**53201**
12 Roden Dr	**Los Angeles**	**CA**	**90001**
3526 Diamond Rd	**Seattle**	**WA**	**98101**
12 Street Road	**Las Vegas**	**NV**	**89101**
2 Applegarth Ln	**Brunswick**	**ME**	**04011**

2) Optim provides corresponding Street Address/City/State/Zip Codes for masking

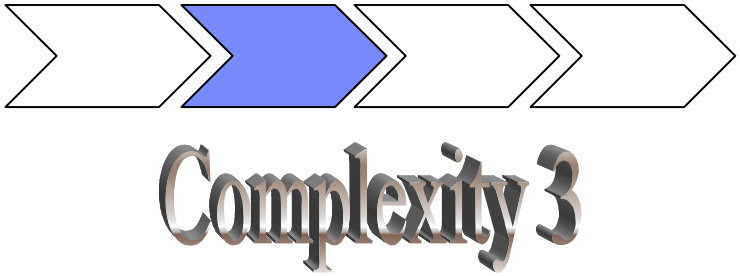
3) Leverage Multiple Column Replacement. Entire address row can be masked with a valid Coding Accuracy Support System (CASS) address using enhanced random lookup function

New Table with Masked Data

Total Assets	Customers	Street	City	State	Zip Code
$534,674,233	**54,999**	3526 Diamond Rd	Seattle	WA	98101
$8,777,733,811	**105,333**	21 Street Rd	Las Vegas	NV	89101

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**



First Names and Last Names Data Sets

Production Database

First Name	Last Name	GPA	High School	Advisor	State
Paul	Smith	3.2	Princeton	Johnson	NJ
Kate	Jones	2.7	Albany	Kline	NY

First Name Lookup Table

John
Bob
Danielle
Dave
Stacey

Last Name Lookup Table

Newton
Nelson
Kline
Howell
Reese

1) Client is a University who wishes to mask the first and last name fields in their admissions database

2) Optim now has a first name lookup table with over 5,000 male/female names and a last name lookup table with over 80,000 names

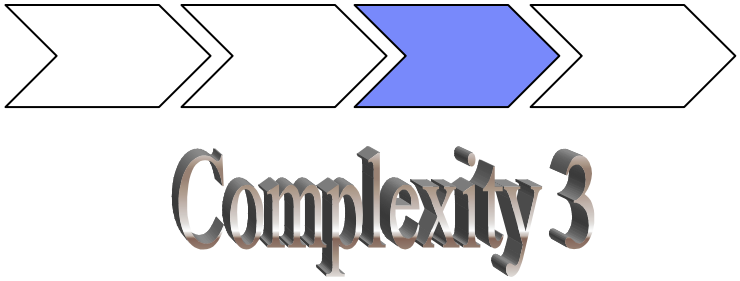
3) Use Lookup Tables to randomly replace table first and last names

Test Database

First Name	Last Name	GPA	High School	Advisor	State
Stacey	Nelson	3.2	Princeton	Johnson	NJ
Dave	Reese	2.7	Albany	Kline	NY

Example: Semantic Transformation

- **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*



Social Security Numbers and Credit Cards

Production Database

F. Name	L. Name	Credit Card#	SSN#
John	**Jones**	**5298774132478855**	**254-77-6644**
Vanessa	**Jones**	**4324115574123654**	**154-74-7788**

Data before Masking

Valid

Valid

Test Database

F. Name	L. Name	Credit Card#	SSN#
John	**Jones**	5326458711224956	854-77-6644
Vanessa	**Jones**	4972584612457744	258-74-7788

Data after Masking… Masked with Valid CC# and SS#

How are these numbers valid?

For Social Security Numbers	For Credit Card Numbers
A Social Security Number (SSN) consists of nine digits. The first three digits is called the "area number'. The central, two-digit field is called the "group Number". The final four-digit field is called the "serial Number". All numbers must fit the latest available criteria for each section.	Most credit card numbers are encoded with a "Check Digit". A check digit is a digit added to a number (either at the end or the beginning) that validates the authenticity of the number. A simple algorithm is applied to the other digits of the number which yields the check digit.

Propagating Masked Data

Customers Table

Cust ID	Name	Street
08054	Alice Bennett	2 Park Blvd
19101	Carl Davis	258 Main
27645	Elliot Flynn	96 Avenue

Orders Table

Cust ID	Item #	Order Date
27645	80-2382	20 June 2004
27645	86-4538	10 October 2005

- **Key propagation**
  - Propagate values in the primary key to all related tables
  - Necessary to maintain referential integrity

Masking with Key Propagation

Original Data

Customers Table

Cust ID	Name	Street
08054	Alice Bennett	2 Park Blvd
19101	Carl Davis	258 Main
27645	Elliot Flynn	96 Avenue

Orders Table

Cust ID	Item #	Order Date
27645	80-2382	20 June 2004
27645	86-4538	10 October 2005

Referential integrity is maintained

De-Identified Data

Customers Table

Cust ID	Name	Street
10000	Auguste Smith	Mars23
10001	Claude Jones	Venus24
10002	Pablo Adams	Saturn25

Orders Table

Cust ID	Item #	Order Date
10002	80-2382	20 June 2004
10002	86-4538	10 October 2005

Without Key Propagation…

Original Data

Customers Table

Cust ID	Name	Street
08054	Alice Bennett	2 Park Blvd
19101	Carl Davis	258 Main
27645	Elliot Flynn	96 Avenue

Orders Table

Cust ID	Item #	Order Date
27645	80-2382	20 June 2004
27645	86-4538	10 October 2005

Without Key Propagation

Customers Table

Cust ID	Name	Street
10000	Auguste Smith	Mars23
10001	Claude Jones	Venus24
10002	Pablo Adams	Saturn25

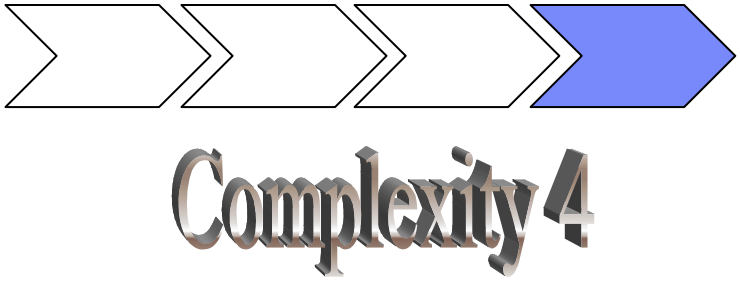
Now these are Orphans!

Orders Table

Cust ID	Item #	Order Date
27645	80-2382	20 June 2004
27645	86-4538	10 October 2005

Using Custom Masking Exits

- **Apply complex** data transformation algorithms **and populate the resulting value to the destination column**
- **Selectively** include or exclude rows **and apply logic to the masking process**
- **Valuable where the desired transformation is beyond the scope of supplied Column Map functions**
- **Example: Generate a value for CUST_ID based on customer location, average account balance, and volume of transaction activity**



Why Do Something?

Leading North American Financial Institution –
*Eliminated downtime associated with rebuilding test environments - savings of up to US**$250,000 per year.** Achieved more than US**$100,000 annual savings** collectively for 10 to 15 projects.*

Large International Financial Services Group –
*Reduced the time needed to create a test environment by up to **90% (from 20 days to just 2 days).** Improved time-to-deployment of new application functionality, contributing to critical business/financial initiatives.*

Leading Banking & Payment Technology Solutions –
*Reduced **operational cost** and **improved efficiencies** by reducing the size of test database from 1.2TB to 24GB*

Information Management Optim

Data Privacy Lab

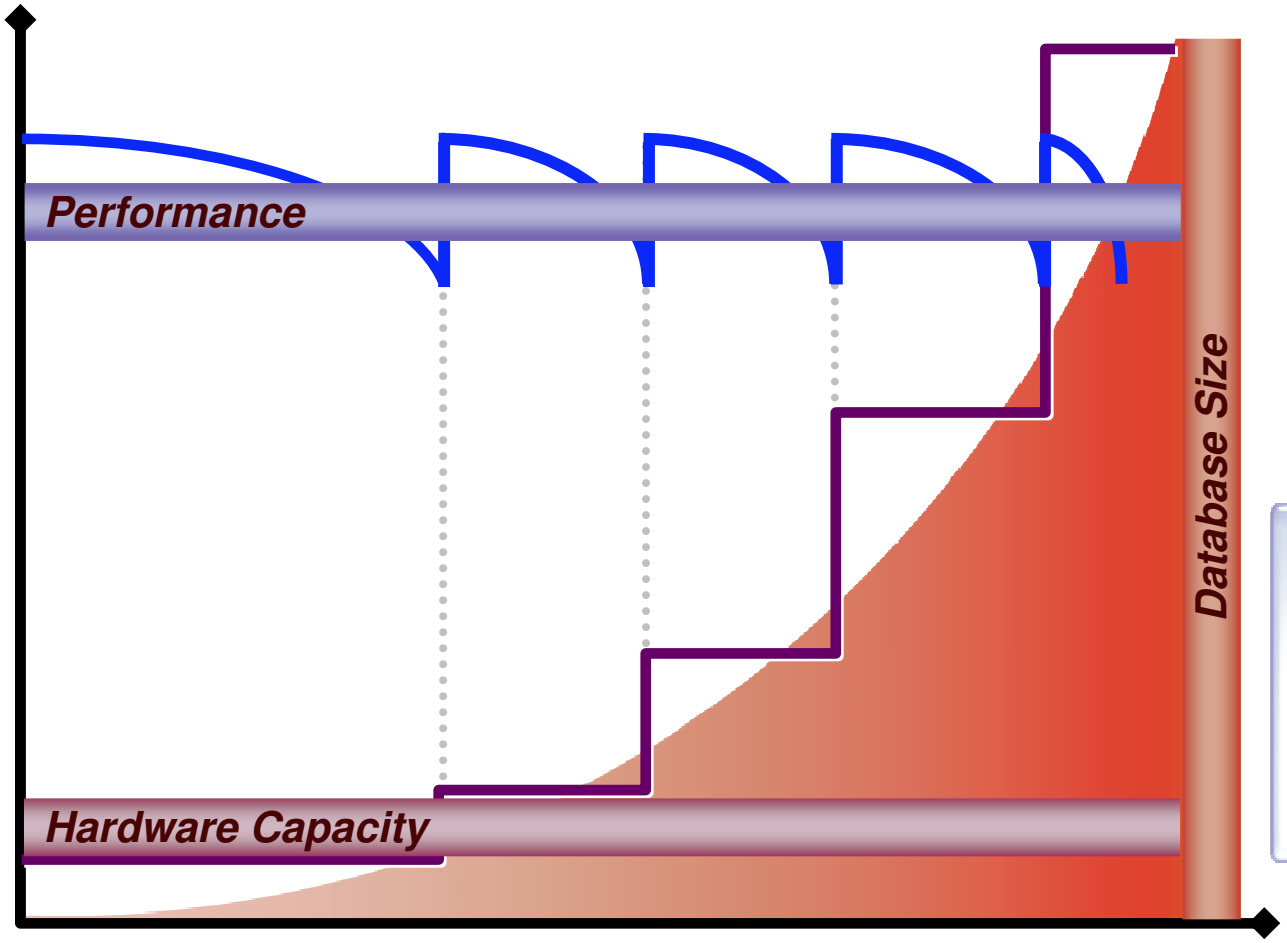
Information Management Optim

IBM Optim Data Growth

The Symptoms of Data Growth Problems

- **Applications perform slowly**
  - Service level agreements are being missed
  - Customer satisfaction declining
- **Backups seem to take forever**
- **Batch jobs run into working hours**
- **Increased infrastructure & storage costs**
  - “Every time I turn around, we are buying more storage”
- **Data Retention Compliance**

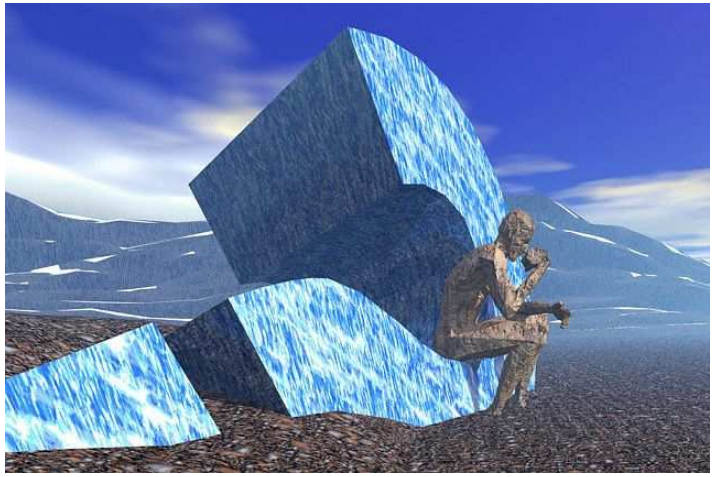
Hardware is Not the Answer



Forrester estimates that, on average, data repositories for large applications grow by 50% annually (structured data).[1]

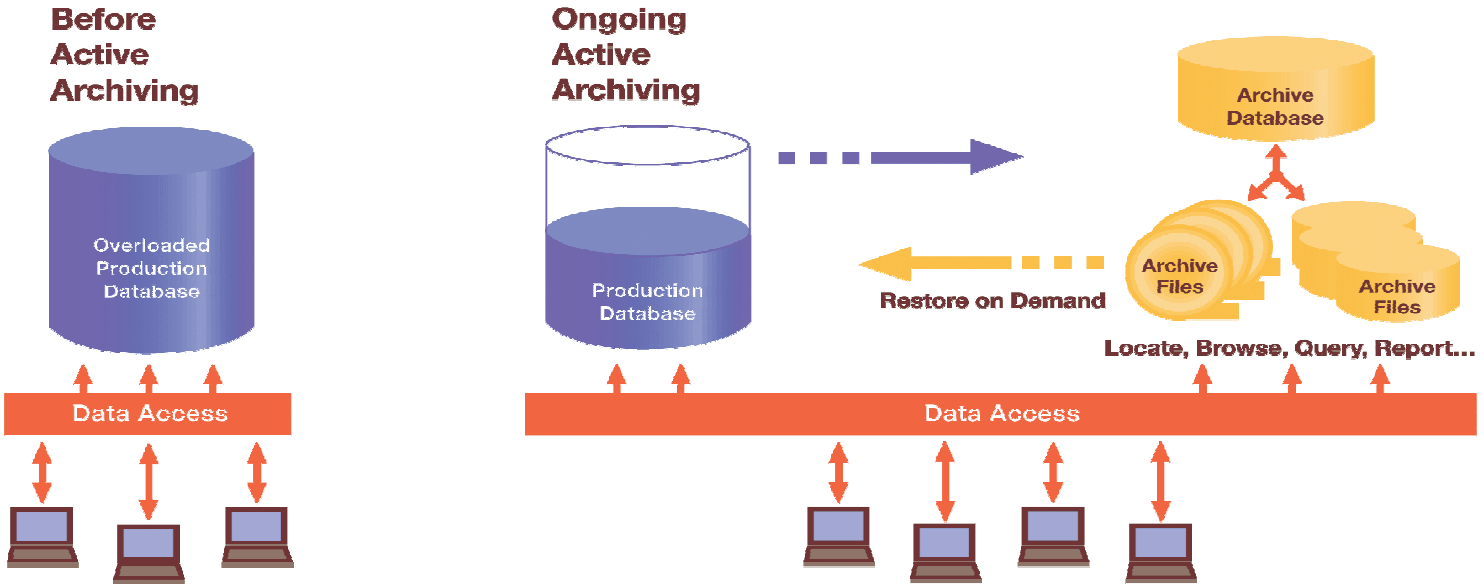
1 Source: Noel Yuhanna, Forrester Research, Database Archiving Remains An Important Part Of Enterprise DBMS Strategy, Q3 2007

What are the Benefits of Data Archiving?



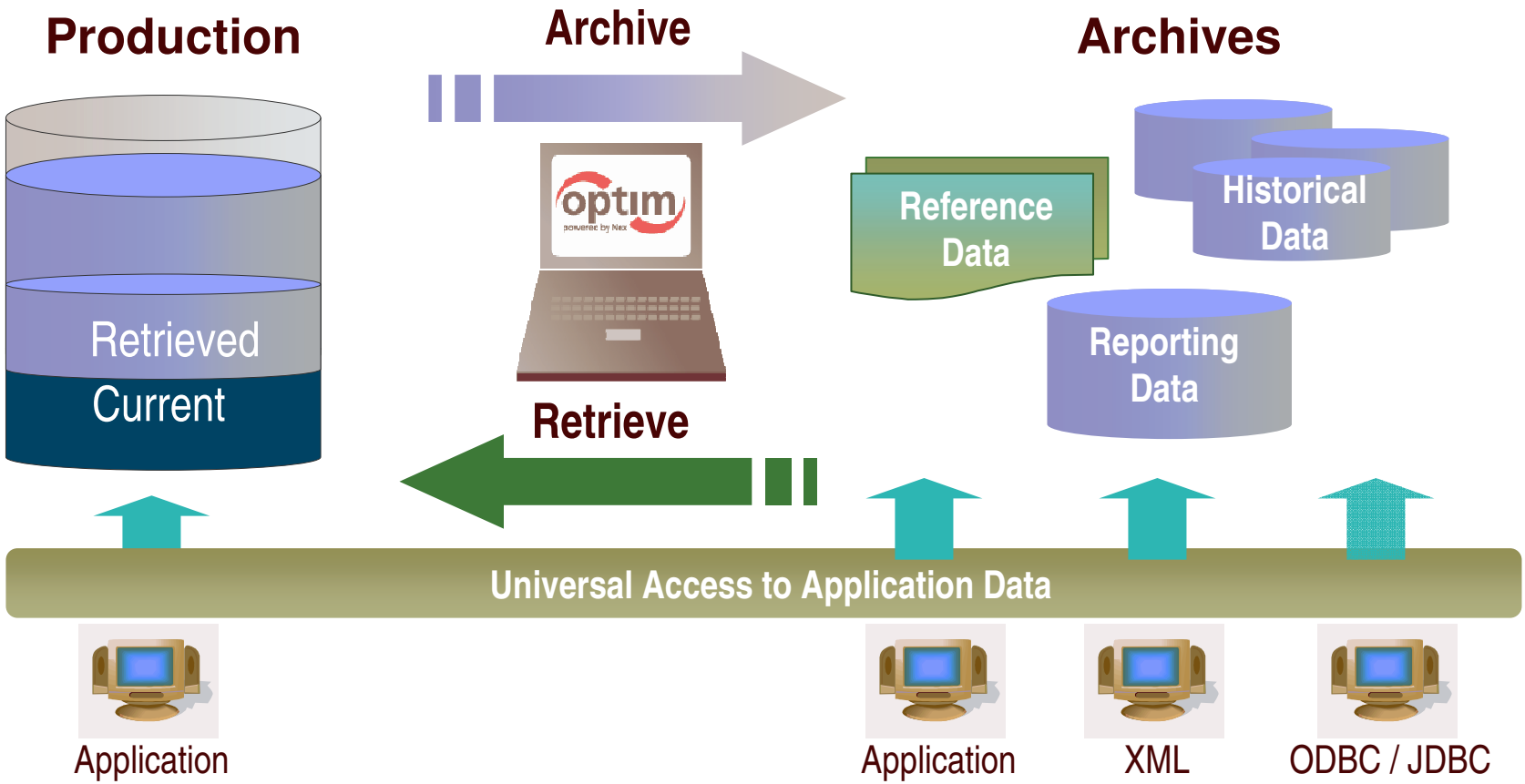
1) **Control Costs**
2) **Improve Performance**
3) **Mitigate Risks**

Active Archiving Defined



- **Reduce the amount of data in the application database by:**
  - Separating infrequently accessed data from transactional data
  - Preserve metadata and relationships of archived data outside db
  - Archive relational subsets vs. entire files
- **Enable easy user access to archived information**
  - View, research and restore as needed
- **Complementary to Information Life Cycle Management (ILM)**

Optim Data Growth Solution: Archiving



- **Complete Business Object provides historical reference snapshot of business activity**
- **Storage device independence enables ILM**
- **Immutable file format enables data retention compliance**

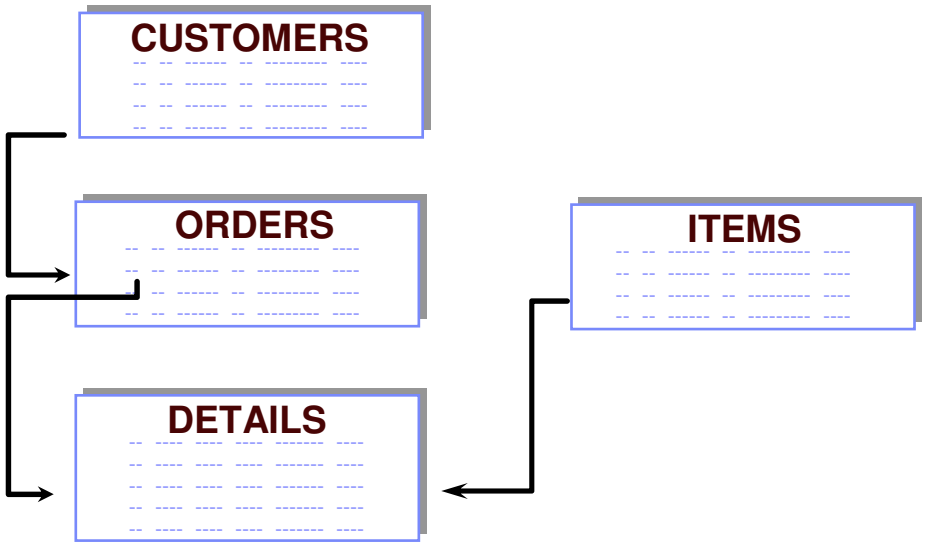
Steps for Archiving Data

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

Identify the data to be archived

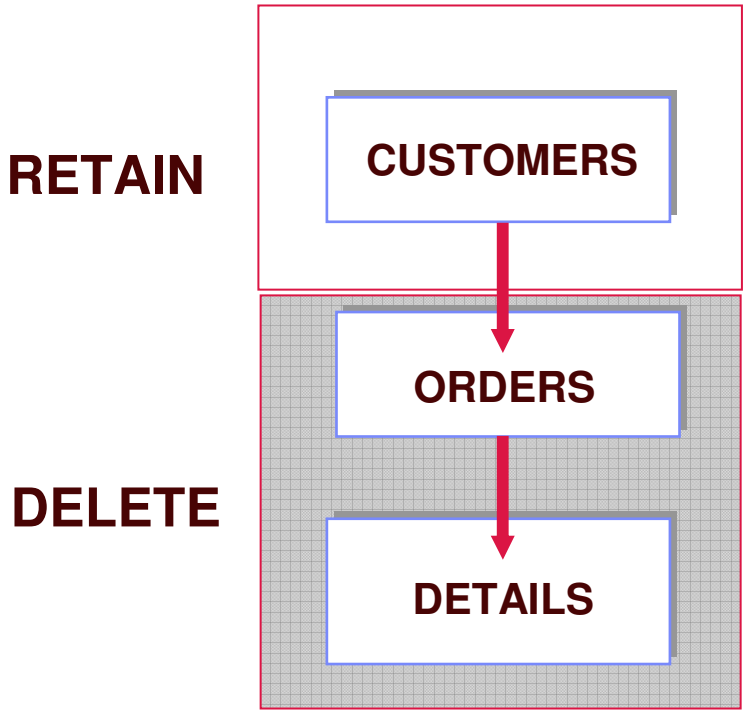
Access Definition

Defines a subset of of relational data



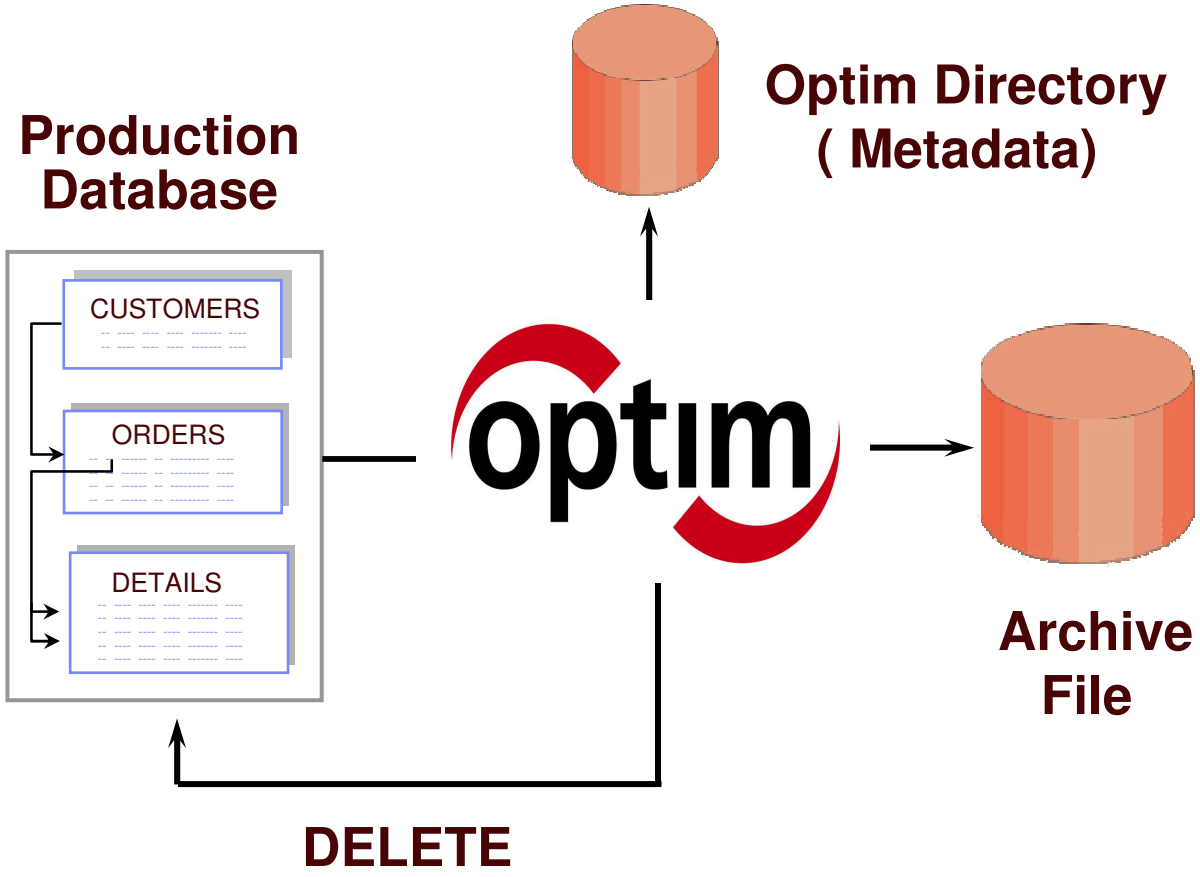
- Start table
- Associated data
- Relationships
- Extraction rules
- Index specifications
- Archive Actions
- Attachments

Define the data to be deleted



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

Create the archive



Archive Options

- **Optim Archive**
  - Compressed (7 – 25%)
  - Indexed for fast retrieval
  - Secured including hiding columns
  - Flexible access options
- **Databases and Applications Support**
- **Heterogeneous tiered storage support for ILM**
  - Storage area network (SAN), network-attached storage (NAS), content-addressed storage (CAS), tape, optical
  - HP, IBM, EMC, NetApp, StorageTek

Repository-Managed Archiving

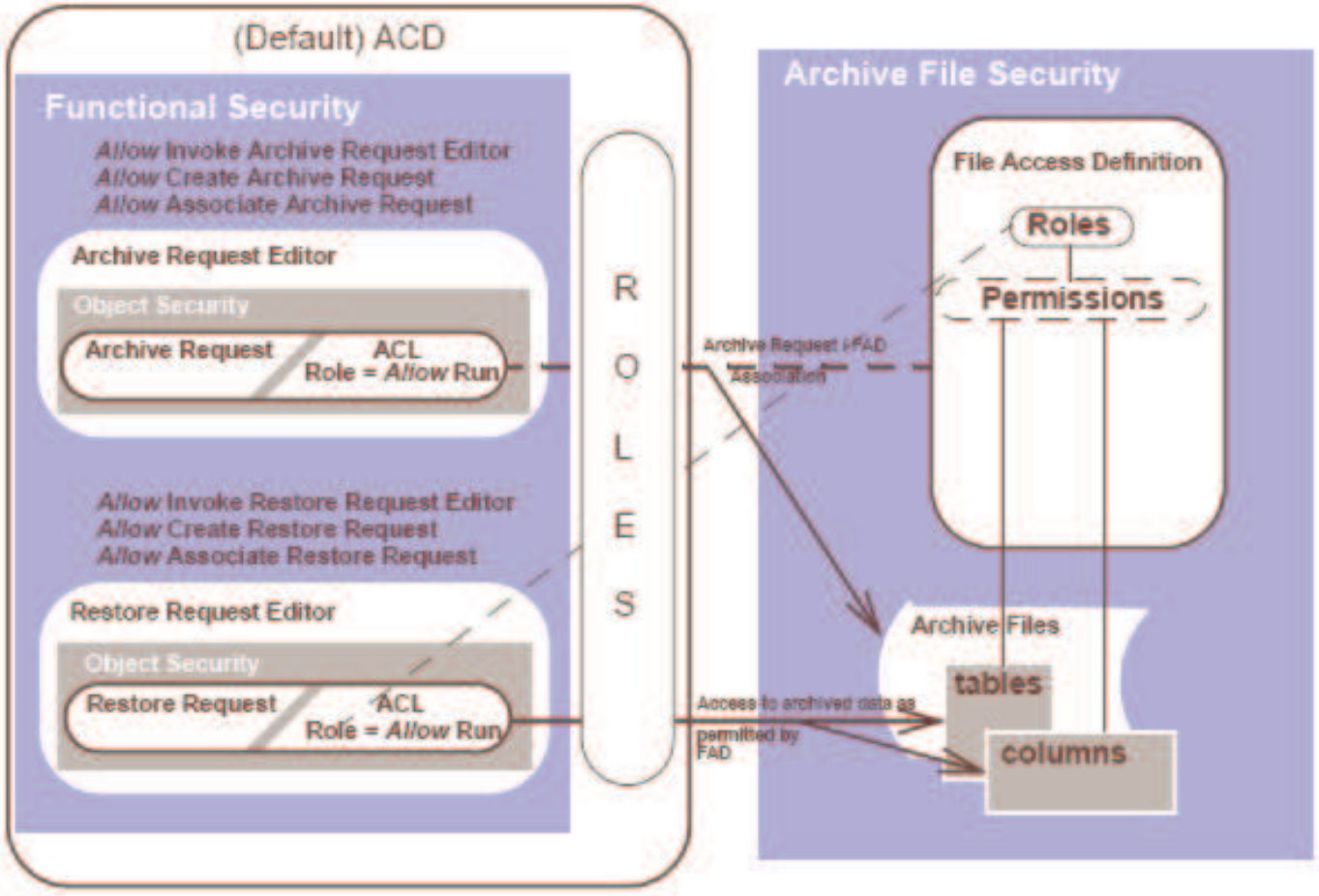
- **Maintain a record of all archive activities**
- **Actively manage archives**
- **Optimize access to archives**
- **Manage metadata**

Optim Security

***Optim* provides three types of security. For each Optim Directory, you may establish any or all of the following 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 PST 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 PST Directory.

Archive File Security



Researching the Archives

Restore archived data only when you need to

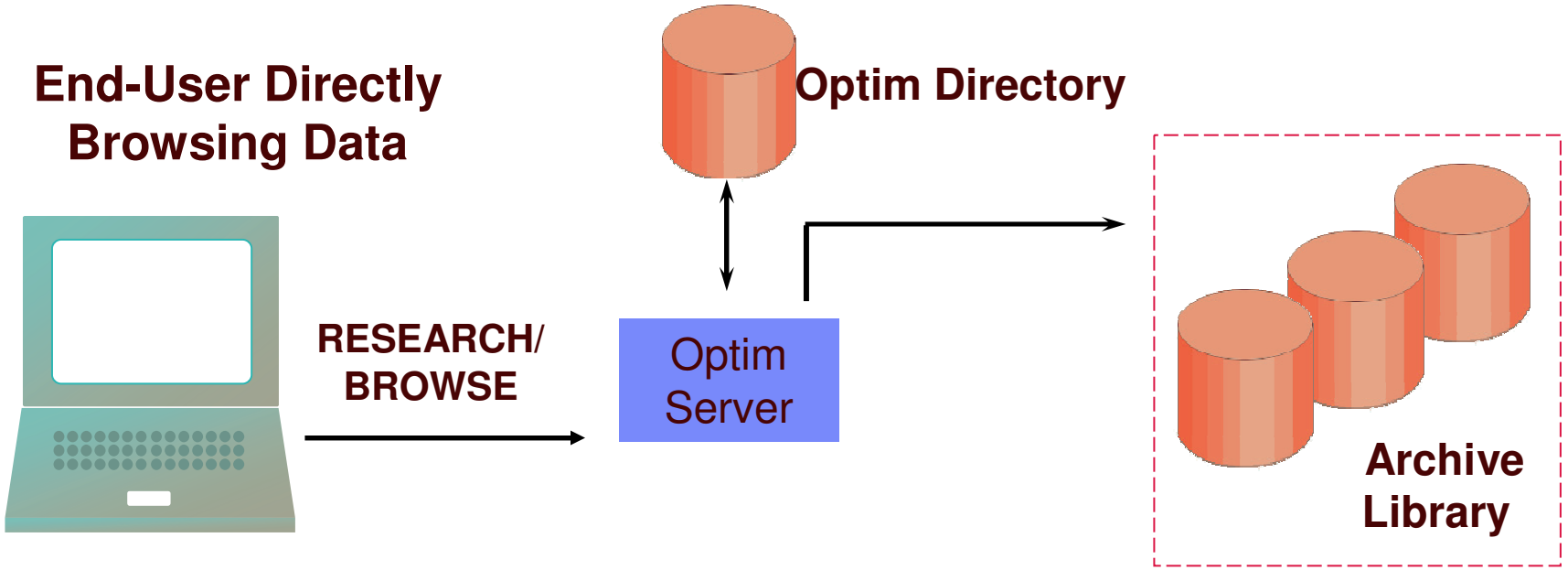
Direct access to archived data:

- **User maintainable indexes**
- **Global searches**
- **Simple or complex criteria**
- **Intelligent browse**

Restore Archived data

- **Selective Restore**
- **Full Restore**

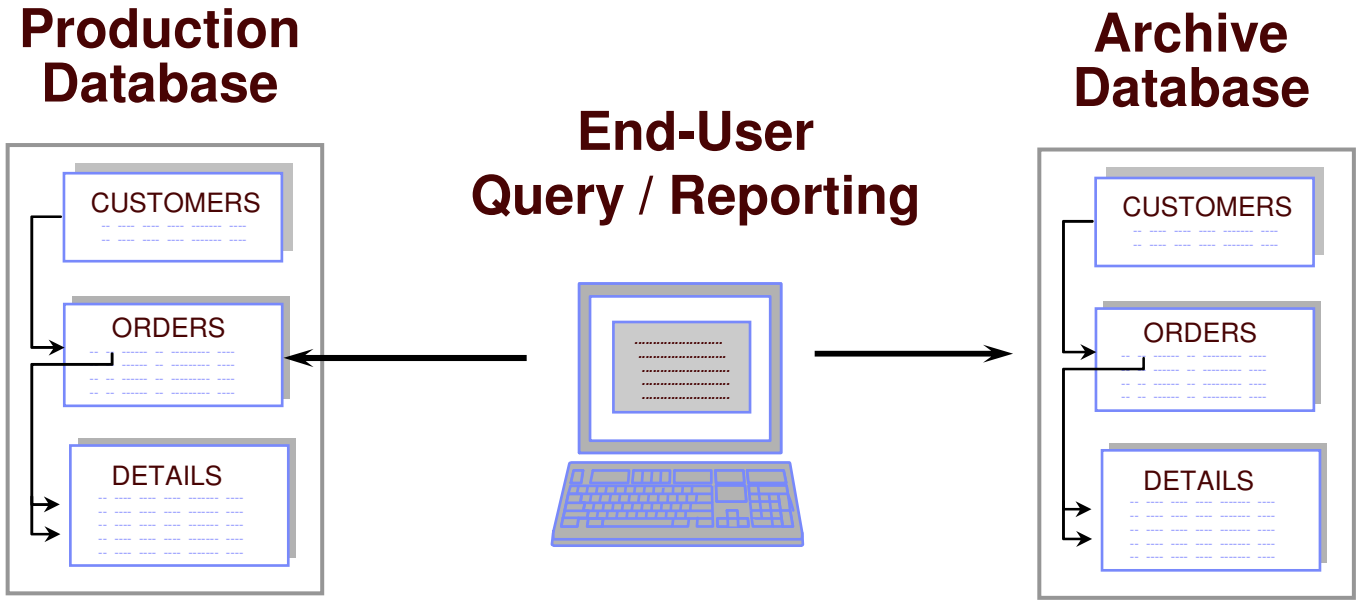
Option 1: Browsing the Archive Files



Option 1: Use the Optim Relational Browse facility

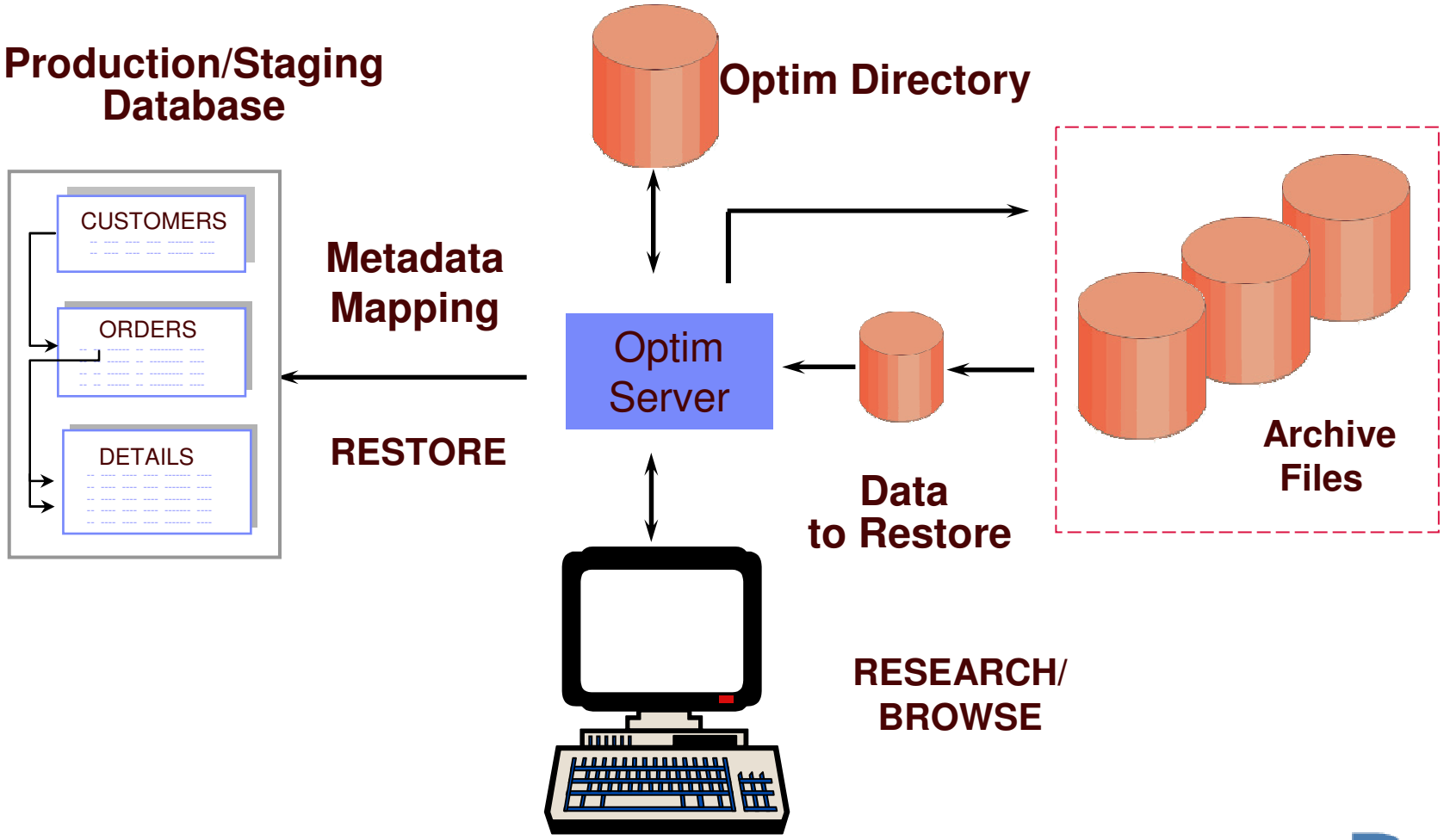
- Full table or apply Find criteria
- JOIN to view related archive data
- Create hardcopy reports

Option 2: Accessing the Archive Database



- **Minor or no changes to Application Code**
  - **Application can point to one or both databases**
- **Must LOAD, INSERT or RESTORE from the Archive File**
- **Archive File Reporting requires Open Data Manager (ODM) (Option 3)**

Selectively Restoring Archived Data



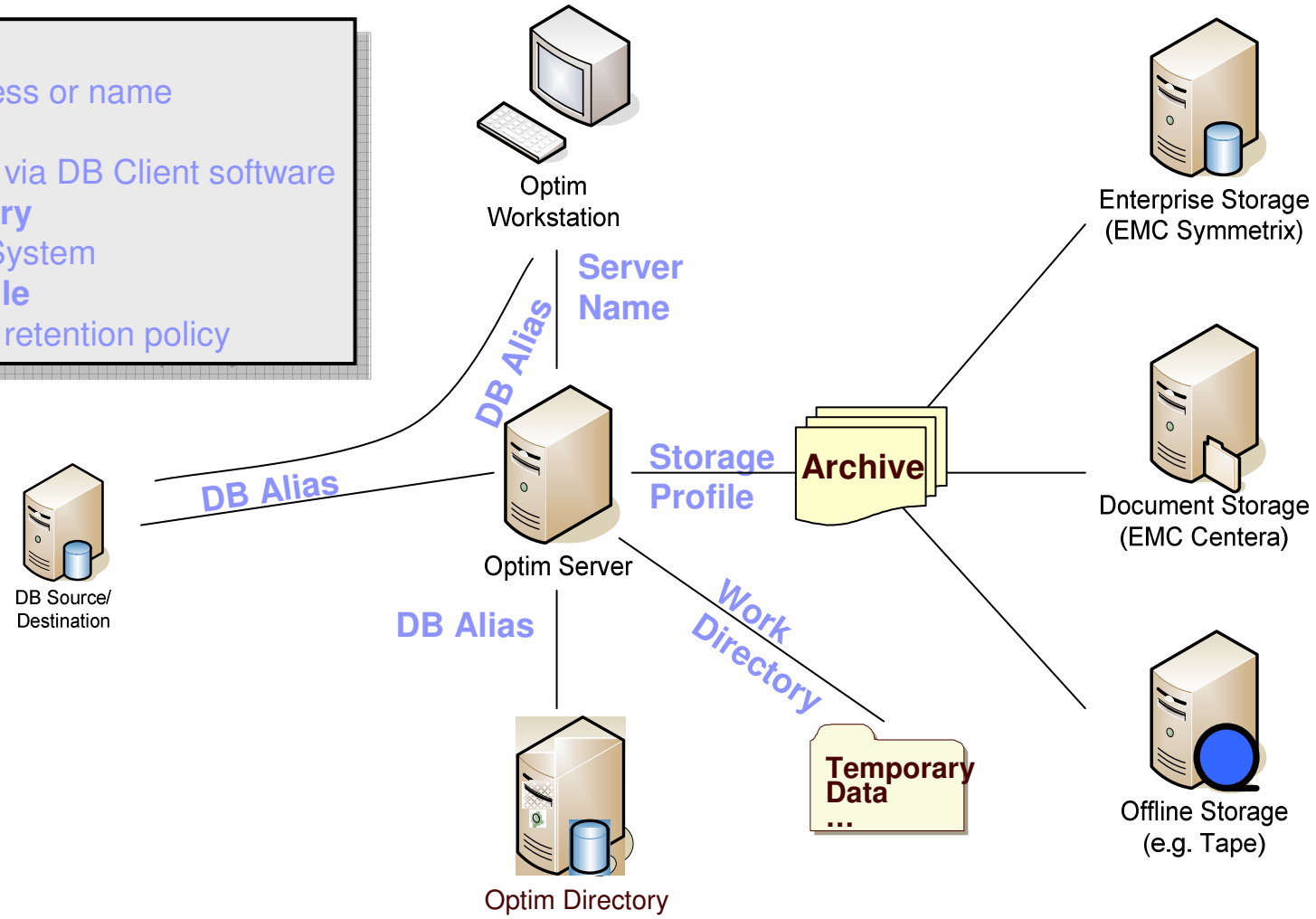
Conceptual Data Growth Architecture

Server Name
- Server address or name

•DB Alias
- Connectivity via DB Client software

Work Directory
- Server File System

Storage Profile
- Storage and retention policy



Reduce Costs with Archiving

1. Storage

- Production level data is typically one of the most expensive storage platforms
- Migrate and store data according to its evolving business value (ILM)
- Use tiered storage strategies to your advantage to maximize cost efficiencies
- Utilize the storage you already have (including tape!)

Reduce Costs with Archiving

2. Administrative costs of data management

- Software license fees
- Hardware costs
- Labor to manage data growth
  - Database Admin (DBA)
  - System Admin
  - Storage Admin

Reduce Costs with Archiving

3. Upgrades and Migrations

- Important for packaged applications space (Siebel, PeopleSoft Enterprise, Oracle E-Business, JD Edwards EnterpriseOne)
- Reduce time allocated for database conversion
- Reduce downtime during transition
  - One recent client stated 1 hour downtime = US$5M
- Deploy new version quickly
  - Revenue recognition
  - Competitive Advantage

Reduce Costs with Archiving

4. Application Decommissioning/Consolidation

- Archiving allows you to move only the data needed, but maintain access
  to the original data in its business-object form without the original application.
  You can then:
  - Retire an application
  - Consolidate redundant systems into a single “enterprise standard”
  - Migrate portfolio to lower-cost platform
  - Consolidate and eliminate unsupported databases and versions
- Benefits
  - Reduce IT infrastructure costs (hardware, software, labor costs)
  - Reduce infrastructure complexity (eliminate confusion)
  - Reclaim assets

Additional Benefits of Archiving for Applications

- **Improved Availability**
  - No downtime caused by batch process overruns
  - Uptime during crunch time
  - Meet service level agreements
- **Speeding Backup and Recovery**
  - Bring up important/recent data first
  - Bring up older/reference data as conditions permit
- **Improved Application Performance**
  - One of the most understated benefits to archiving
  - Longest and most lasting benefit

Before You Archive

- **Identify business parameters that will drive the archive**
- **Establish service levels for archive access by functional users**
- **Appropriate storage medium for archive data**
  - (Archive files, Archive DB, external storage)
- **Determine the appropriate archive access interface**
  - (ODBC/JDBC, Reporting tools)

Information Management Optim

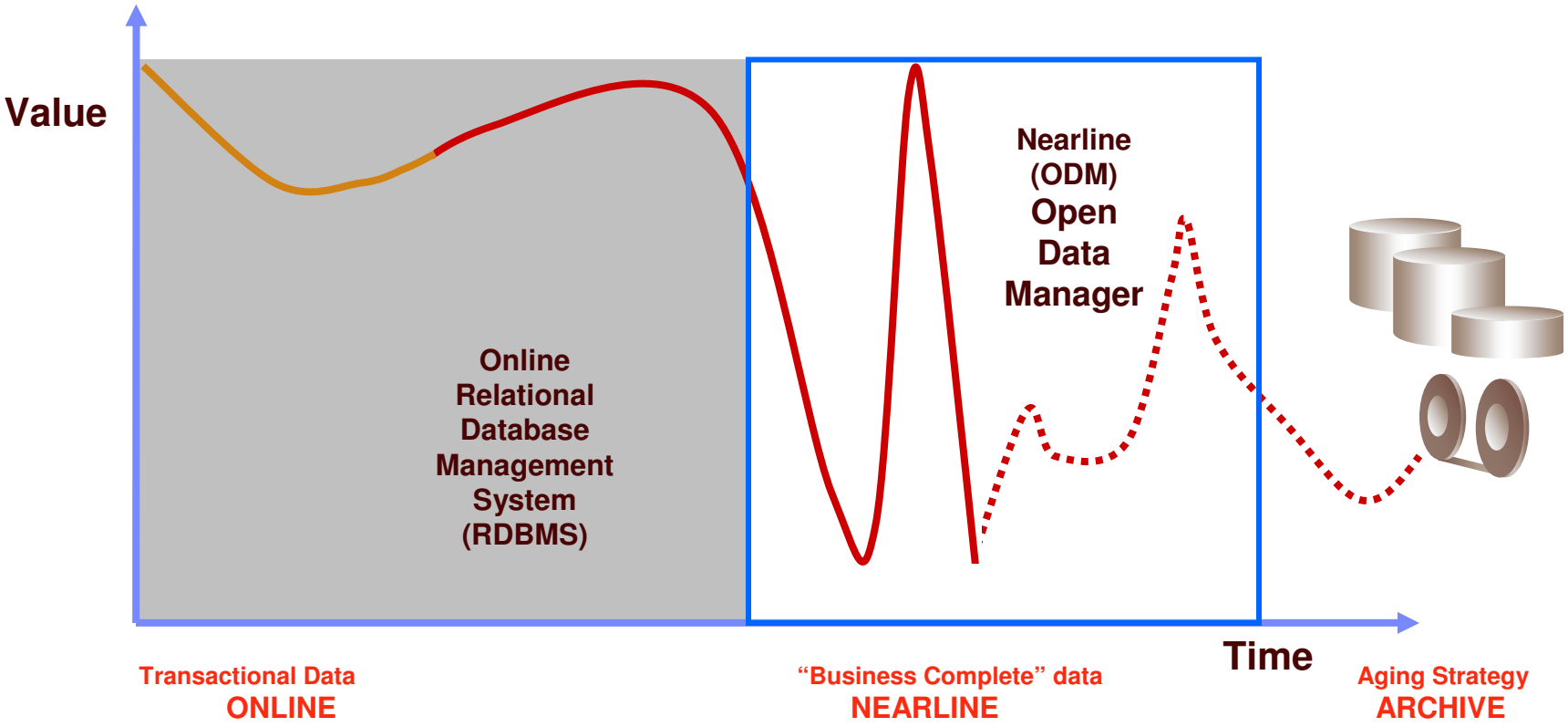
Optim Archive and Restore Lab

Information Management Optim

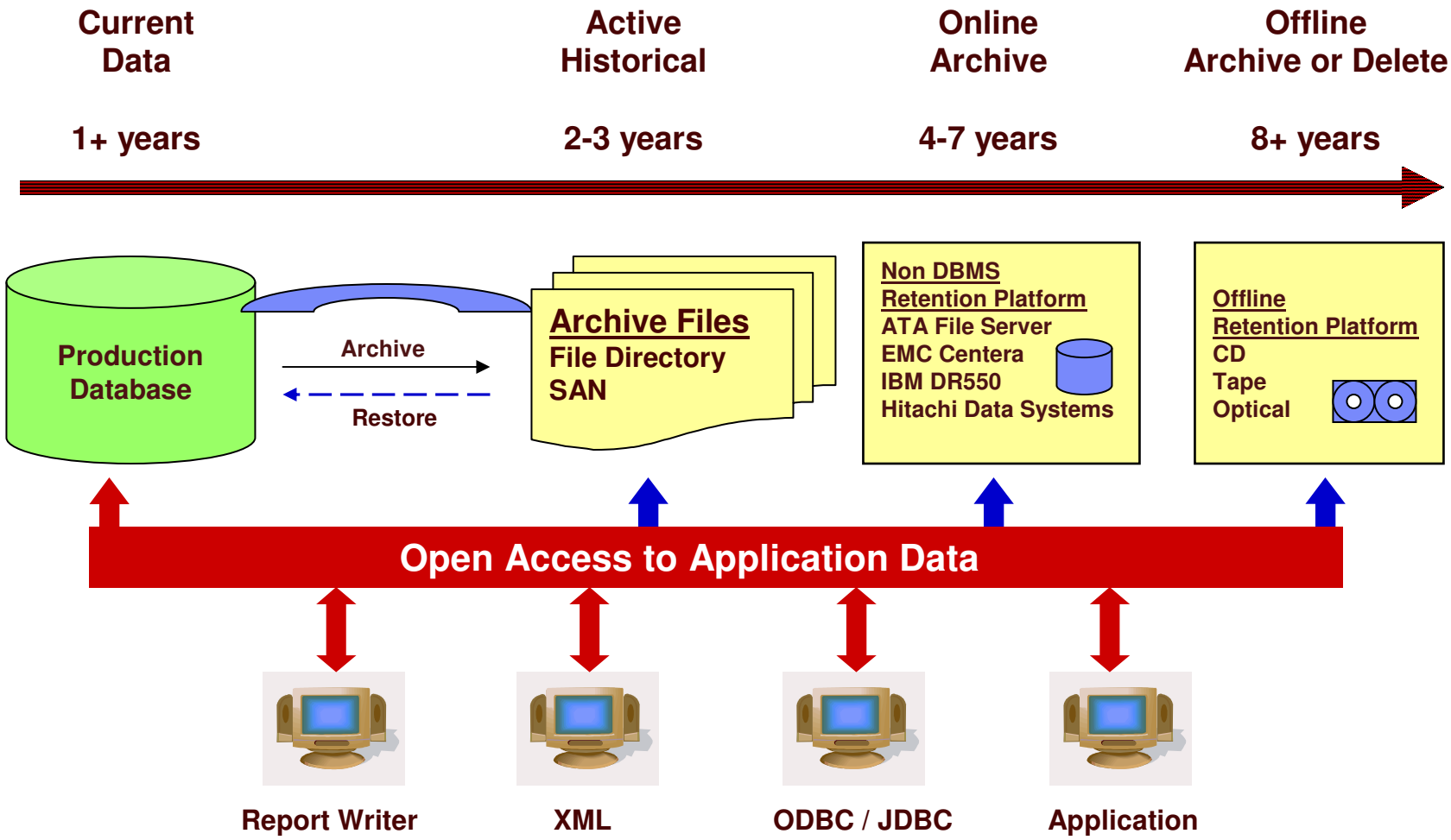
IBM Optim
Information Life Cycle Management (ILM): Working with Archive files

ILM Review

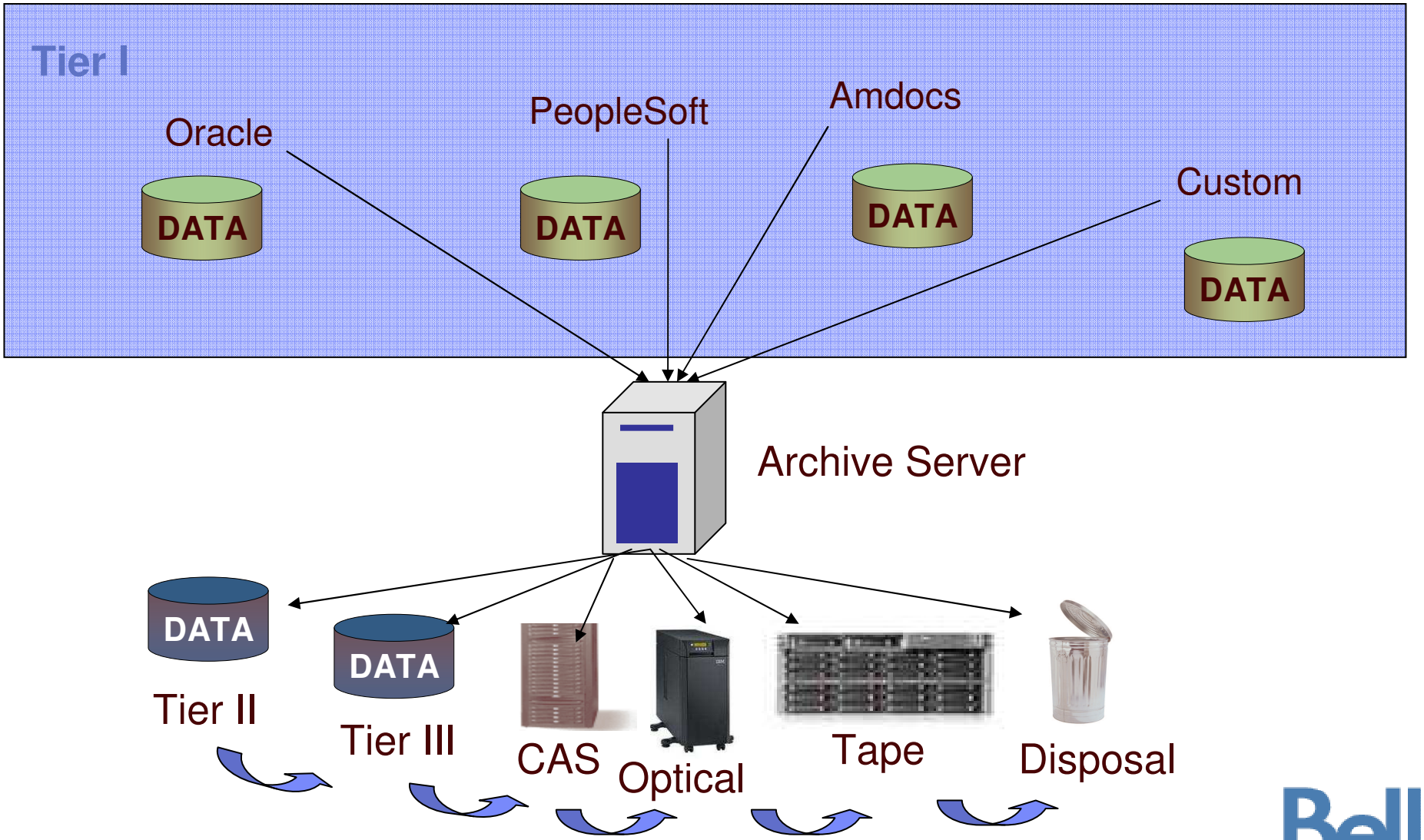
Information Life Cycle Management (ILM)



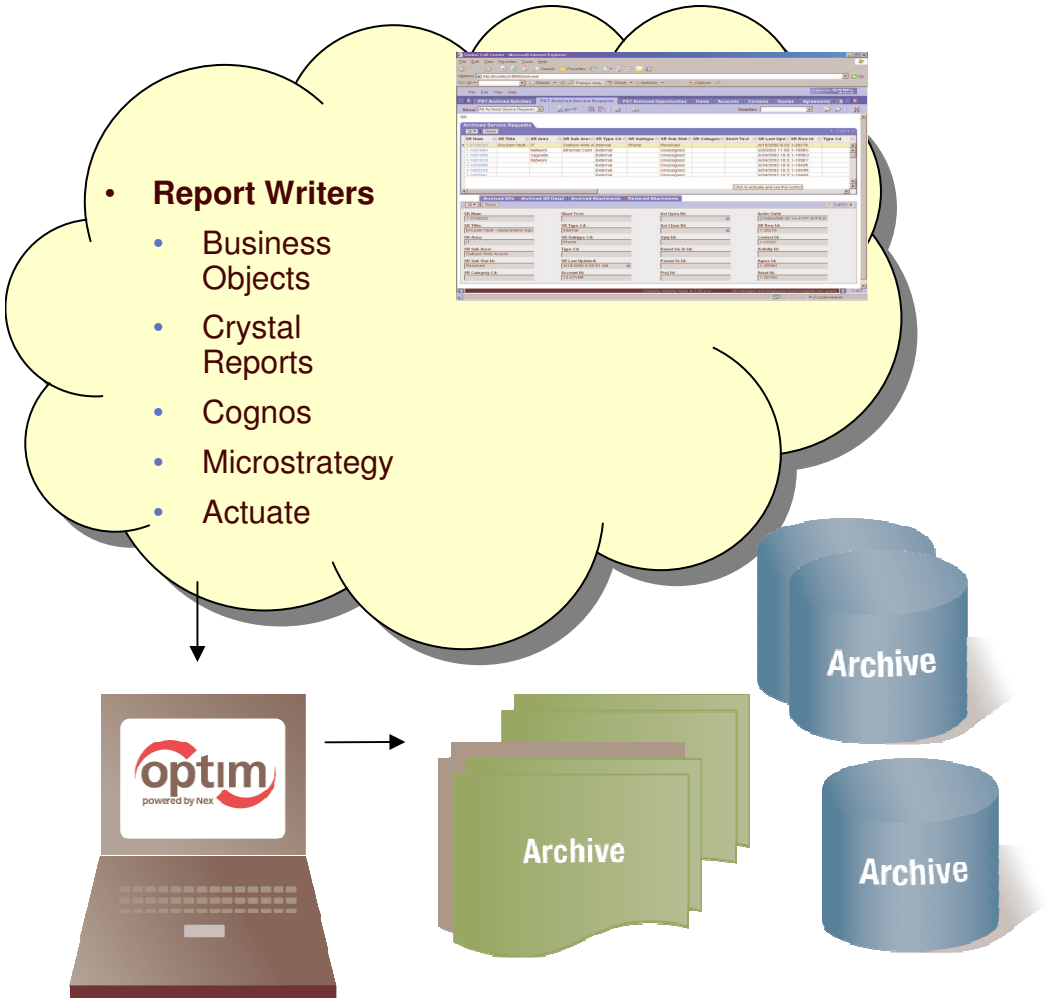
Information Life Cycle Management – Data Retention Strategy



One Example of an ILM Infrastructure

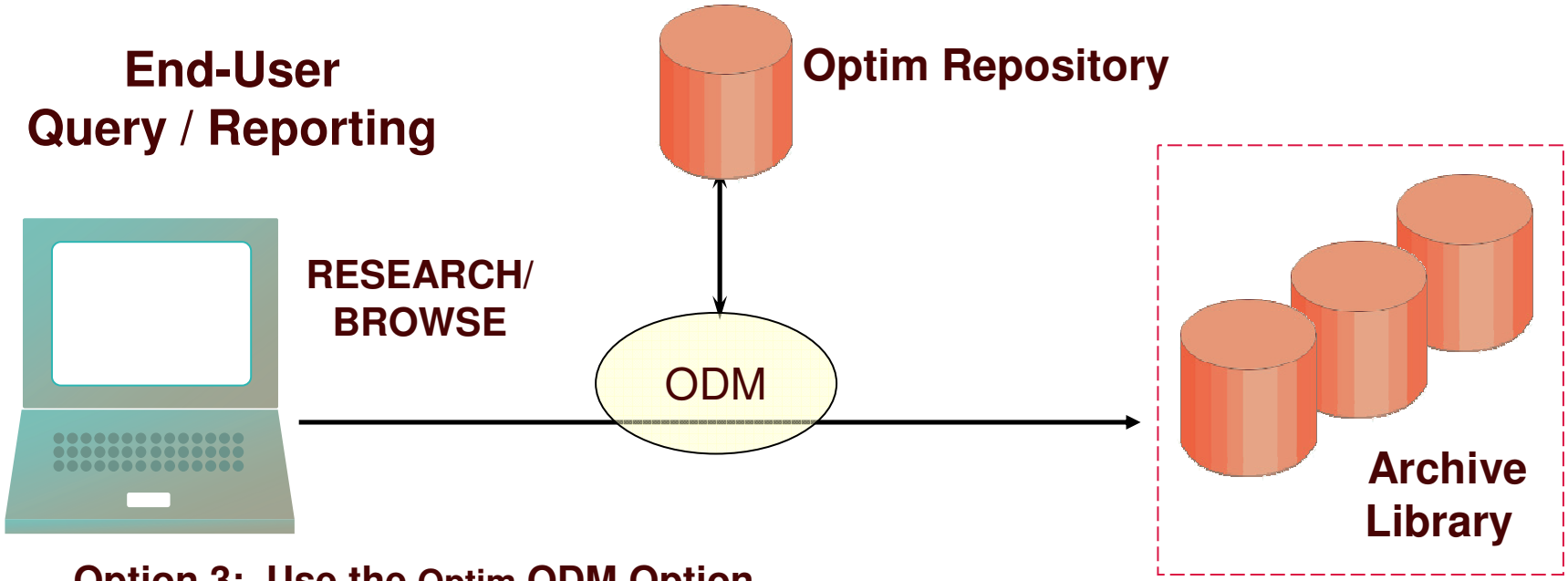


Universal Access to Archived Data



- **Native application access**
  - Familiar screens and processes
- **Application independent access**
  - Industry standard methods: SQL, ODBC/JDBC, XML
  - Portals
  - Report writers: Crystal Reports, Cognos, Business Objects, Discoverer, Actuate
  - Desktop formats: Excel, CSV, Microsoft® Access
  - Database formats

Applications Accessing the Archive Files



Option 3: Use the Optim ODM Option

- ✓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

Open Data Manager (ODM) Concepts

What does ODM do?

- Provide “nearline” access to the Archive repository
  - Make it possible to query archived data
  - Support collections
- Facilitate reporting and usage
  - Standard interfaces
- Increase the value of archived data
- Increase the value of business systems

ODM Concepts

But what is ODM, really?

- Software
  - Attunity Connect
  - Optim “connector”
  - Interfaces and application programming interfaces (APIs) for the real world
- Documentation
- Best Practices
  - Integration
  - Federation
  - Transparency

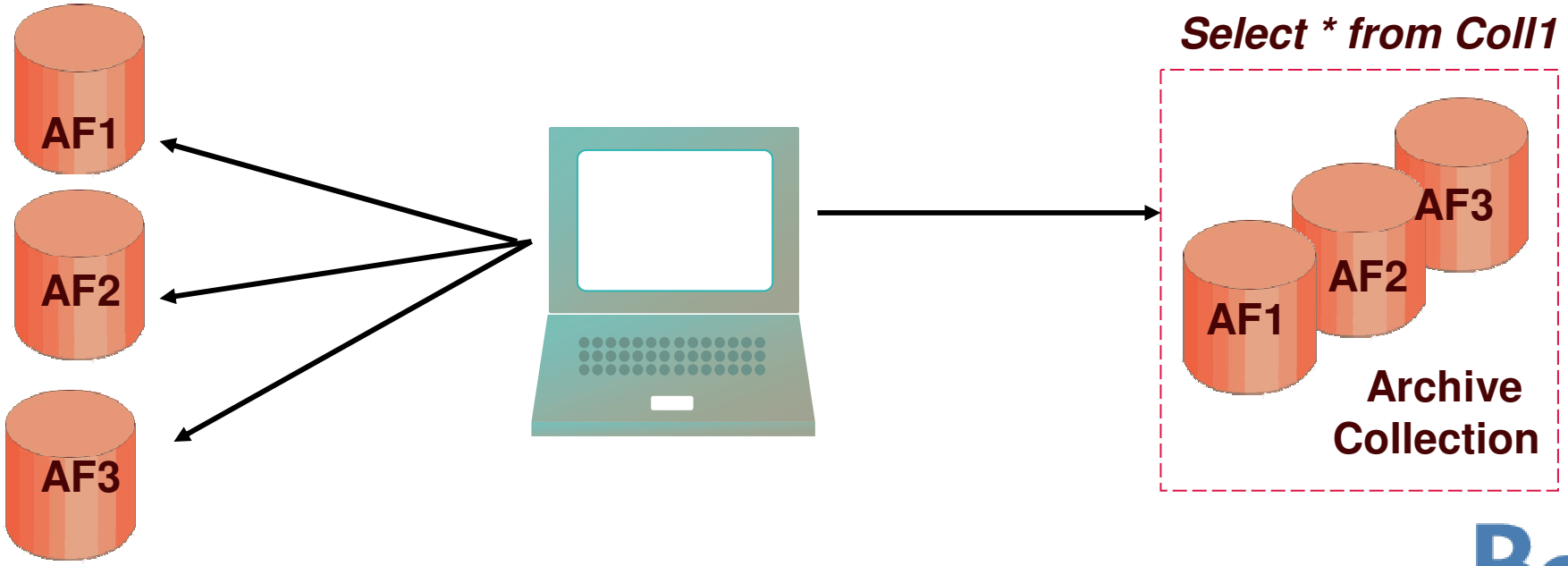
ODM Integration

- **All ODBC-enabled applications including:**
  - Excel, Business Objects, Crystal, Cognos…
- **All JDBC-enabled applications including:**
  - App servers, Portals
- **Federation with Oracle, IBM DB2, and SQL Server.**
- **Runs everywhere that Optim runs.**
- **Supports automatic Collections**

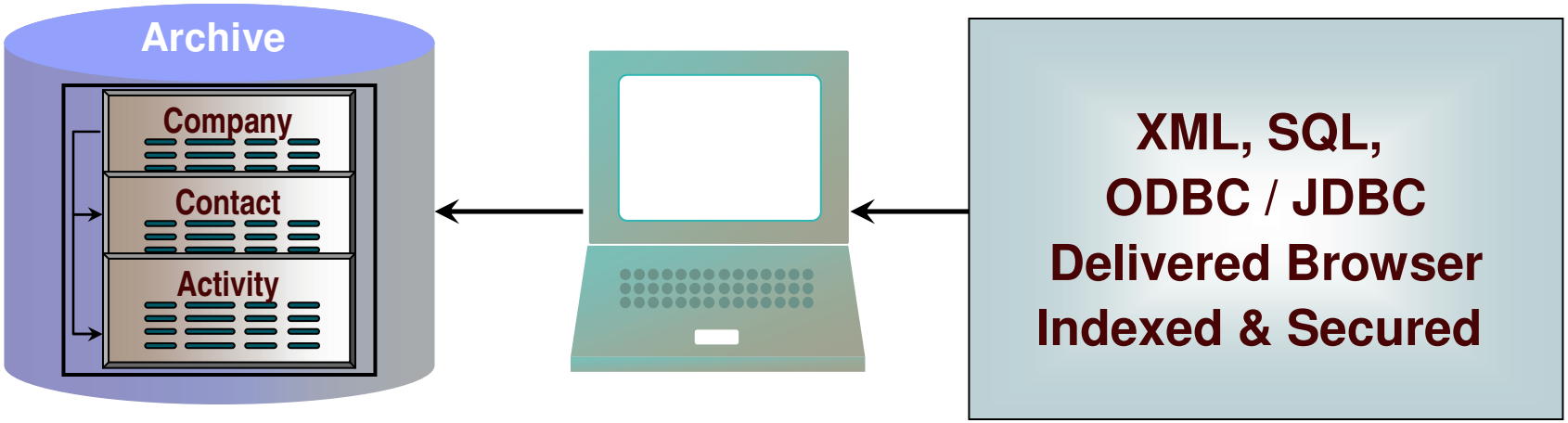
Archive File Collections

- **Manage Archive Files as a single data source using** Open Data Manager (ODM)**.**
- **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



Application-Independent Access to Archives

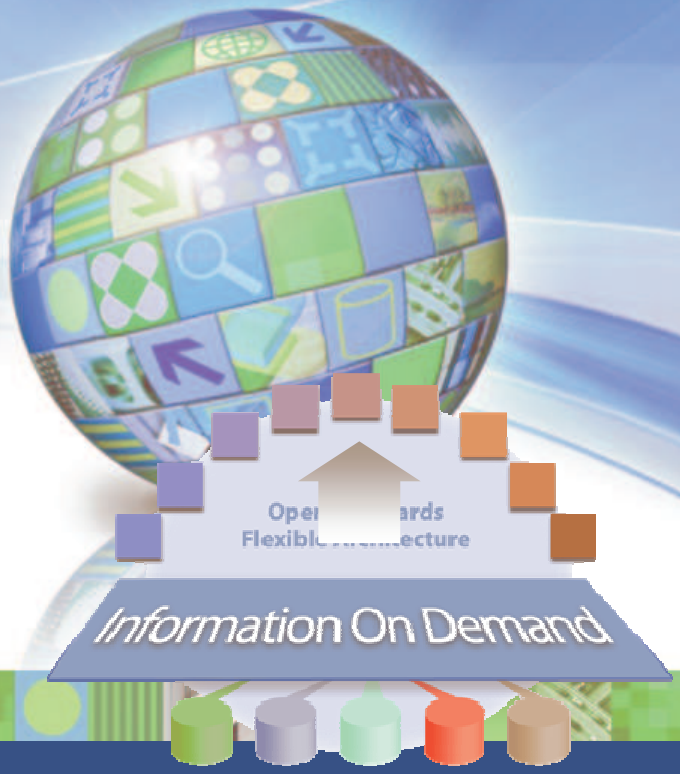


- Application not required
  - Enables decommissioning of obsolete or redundant applications and versions
  - No reliance on application vendor
- Online transaction processing (OLTP) not required
  - Offline access available
  - Archive access does not impair production processing

Information Management Optim

Optim ILM Lab

Information Management Optim



Building IBM Optim Archiving Solutions Using IBM InfoSphere Discovery

Agenda

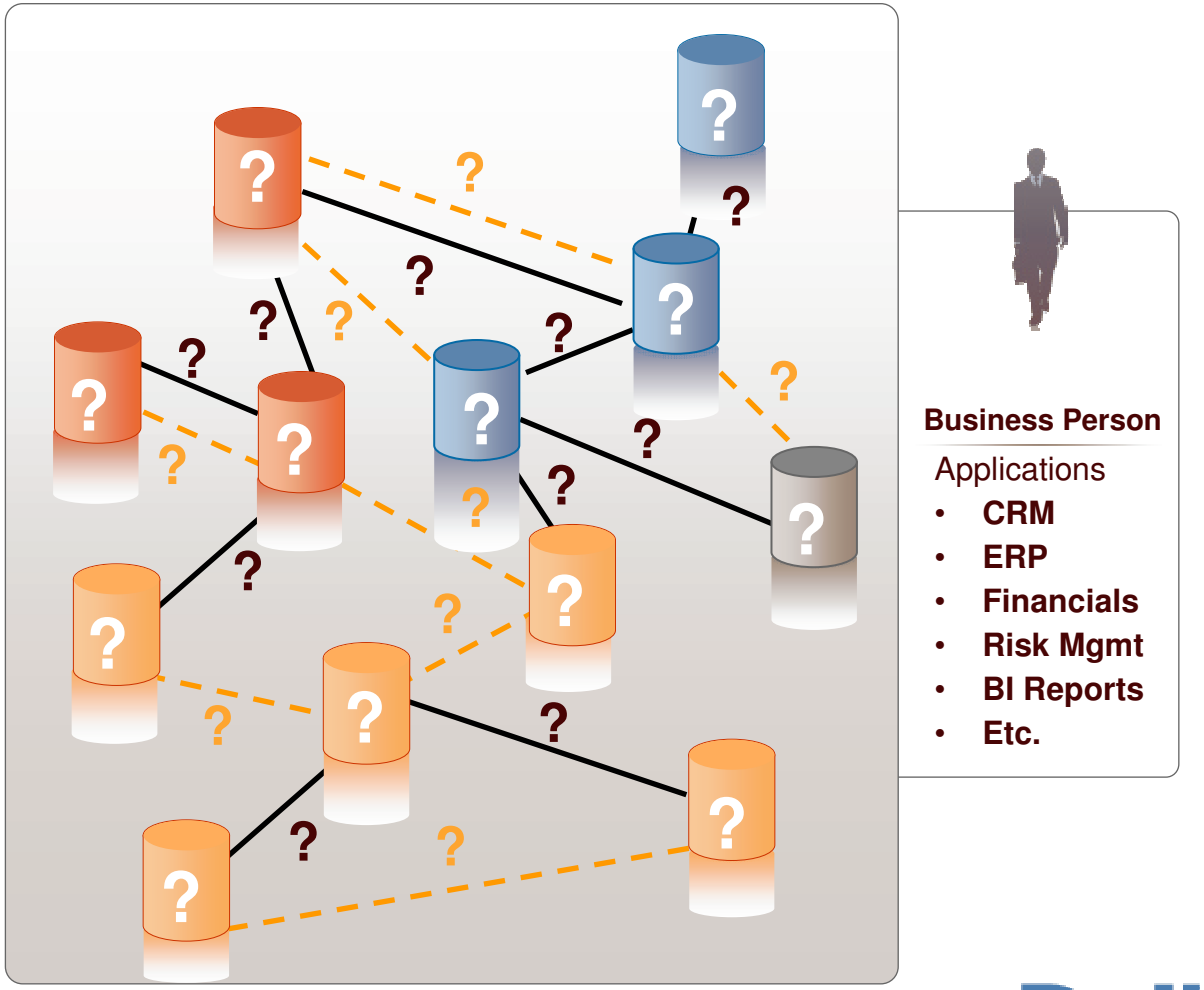
- **Overview of IBM InfoSphere Discovery and InfoSphere Validator products**
- **Overview of how InfoSphere Discovery works with IBM Optim technology**
- **Example: using InfoSphere Discovery to jump start an Optim solution**

Overview of InfoSphere Discovery and InfoSphere Validator

You Can't Manage What you don't understand!

- **Distributed Data Situation:**
  - Grows exponentially
  - Increasingly distributed
  - Poorly understood

- **Problems:**
  - Poor understanding = Poor IT Agility
  - Poor understanding = Poor Data Governance
  - Poor understanding = Bad data = Bad business decisions



Business Person

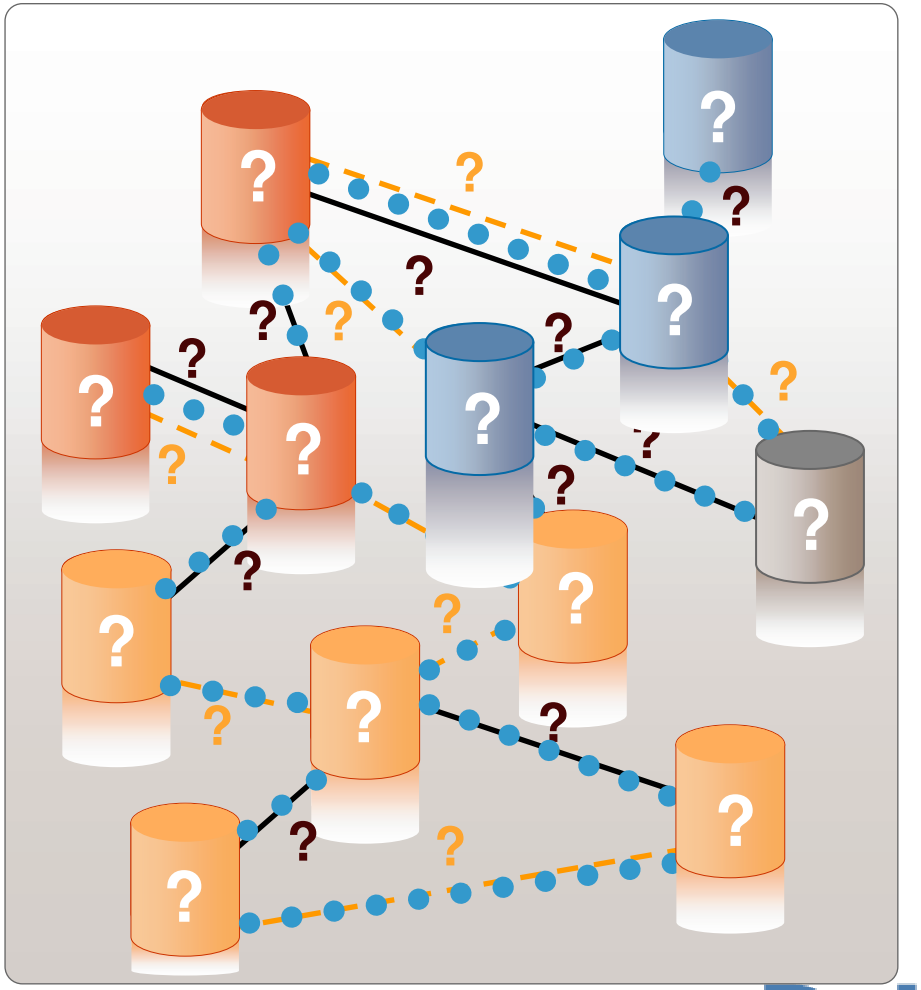
Applications
- **CRM**
- **ERP**
- **Financials**
- **Risk Mgmt**
- **BI Reports**
- **Etc.**

Distributed Data Landscape

Automate Discovery and Accelerate Information Understanding

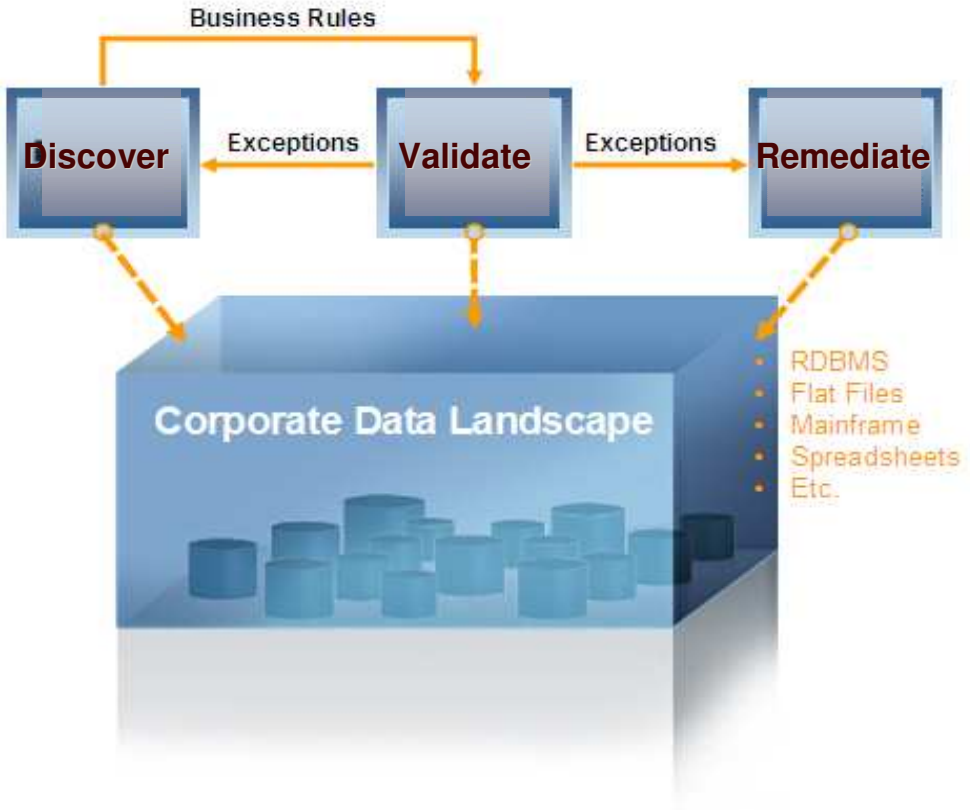
- **Significant Acceleration of Information Agenda projects**
  - Data Growth Management
  - Test Data Management
  - Sensitive Data De-identification
  - Application/Data Consolidation, Migration & Retirement
  - Master Data Management and Data Warehousing

- **Why is this Different?**
  - Data-based discovery
  - Automate discovery of business entities, cross-source business rules & transformation logic
  - Evaluate multiple data sources simultaneously
  - Identify & remediate cross-system rules and inconsistencies



Discover, Audit, Remediate

- InfoSphere Discovery
  - Basic Discovery:
    - Basic profiling plus automated primary-forei key, business entity & cross-source overlap discovery
  - Unified Schema Builder:
    - Prototype empty targets from the combinati of many data sources
  - Transformation Analyzer:
    - Discover complex business rules and transformation logic between two data sources

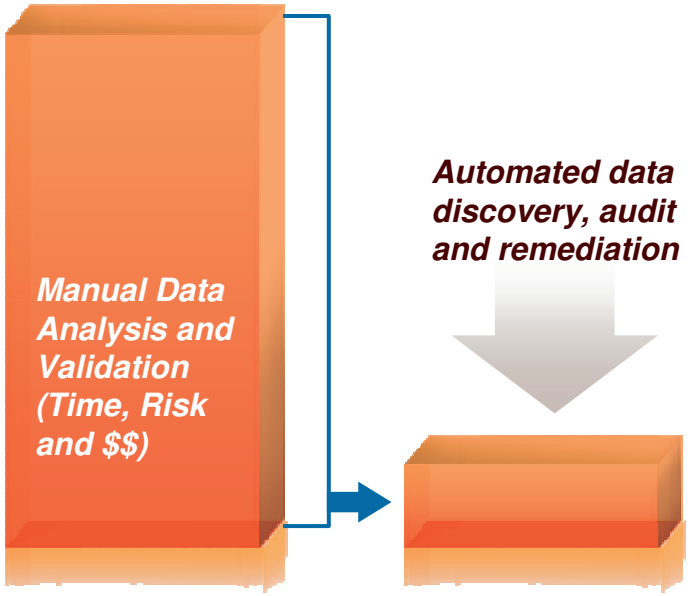


- InfoSphere Validator
  - Operational platform that provides ongoing audit and remediation of business rules across a distributed data landscape

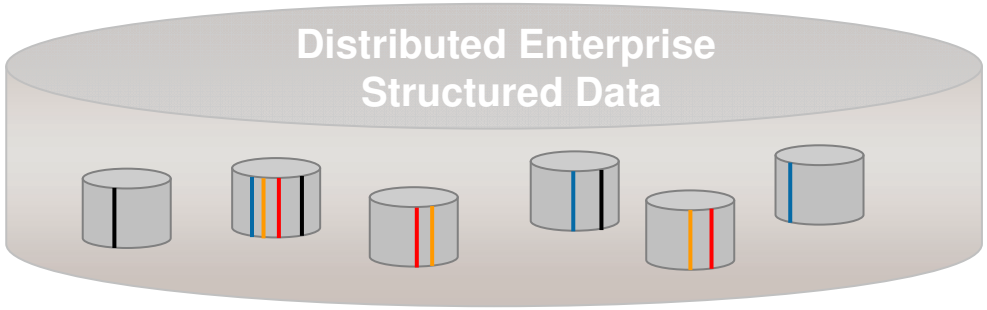
Value to Our Customers

- **Accelerate understanding your existing distributed data landscape for:**
  - Data Archiving
  - Test Data management
  - Sensitive Data
  - Application/Data Consolidation, Migration and Retirement
  - Master Data Management and Data Warehousing

- **10x reduction in risk, time and effort for the discovery phase of your project**
  - Automated discovery of business entities, cross-source business rules & anomalies
  - Increased repeatability
  - Verifiable results



InfoSphere Discovery Basic Discovery

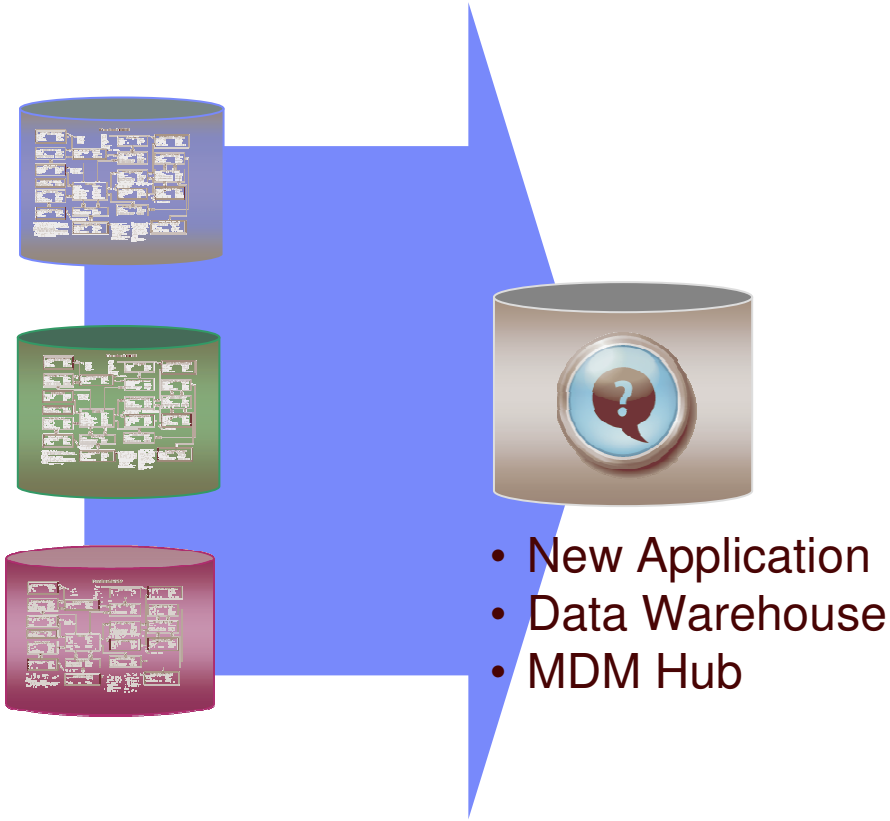


- Basic Discovery:
  - Data profiling and cross-system overlap analysis
  - Automated PF (Primary-Foreign) Key and Business Object discovery
  - Extremely easy to install and use

What is unique?
Only solution on the market that automatically discovers primary foreign keys, business entities, and performs cross-source analysis

- Applicability
  - Data Archiving
  - Test Data Management
  - Sensitive Data Discovery
  - Application retirement Analyzer)
  - MDM (Master Data Management)
  - Data Quality

InfoSphere Discovery - Unified Schema Builder



- New Application
- Data Warehouse
- MDM Hub

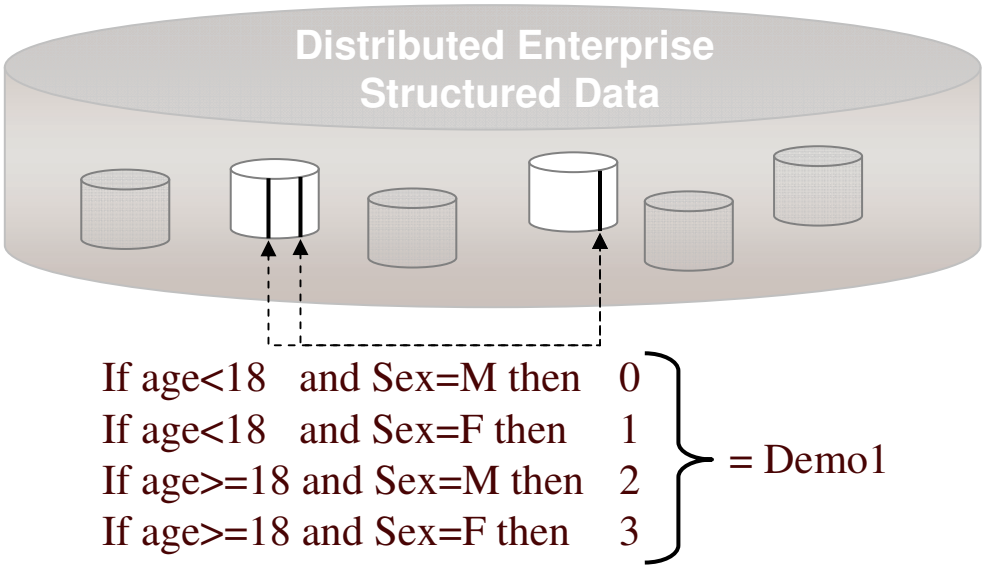
What is unique?

Prototypes empty targets from existing source data (MDM – Master Data Management, EDW – Enterprise Data Warehouse, data migration)

- **Unified Schema Builder:**
  - Data analyst workbench for data consolidation projects
    - Profile data sources
    - Perform overlap analysis
    - Unified data models
    - Unified data profiles
    - Analyze Matching keys
    - Propose conflict resolution precedence
  - Cross source trouble-shooting workbench

- **Applicability**
  - Application/Data Consolidation, Migration & Retirement

InfoSphere Discovery Transformation Analyzer

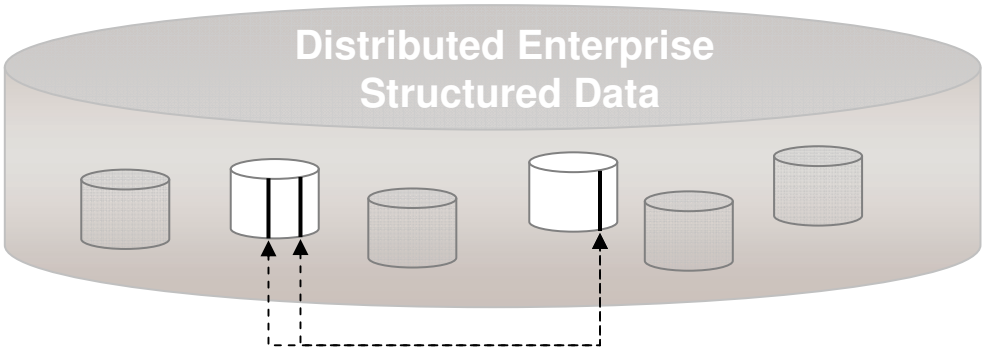


What is unique?

Discovers cross-system business rules, transformations and data exceptions by examining data values

- Transformation Analyzer:
  - Automates discovery of:
    - cross-system business rules and transformations
    - data inconsistencies
  - Detailed data mapping between 2 data sources
  - Discrepancy discovery
  - Cross source trouble-shooting workbench
- Applicability
  - Application de-commissioning and retirement
  - Accurate archiving of packages applications.
  - ETL (Extract, Transform, Load) migration
  - Metadata repository

InfoSphere Validator



If age<18 and Sex=M then 0
If age<18 and Sex=F then 1
If age>=18 and Sex=M then 2
If age>=18 and Sex=F then 3

= Demo1

1% of Rows Map Incorrectly

- Row 6: Demo1: Expected Value = 3 Actual Value = 1
- Row 125: Demo1: Expected Value = 0 Actual Value = 3
- Etc…

What is unique?

Complete remediation environment for establishing ongoing data audit and governance

- InfoSphere Validator:
  - Operational platform that provides ongoing audit and remediation of single-source and cross-source business, metadata and data
  - Business user interface for data remediation
- Applicability
  - Validate that data migrated from legacy applications to new application is equivalent before retiring legacy application
  - Sensitive data monitoring

How does InfoSphere Discovery work together with Optim Technology?

How does this relate to Optim?

- Data Growth Management
- Test Data Management
- Sensitive Data De-identification

EDM Solution Requirements – The Four Pillars

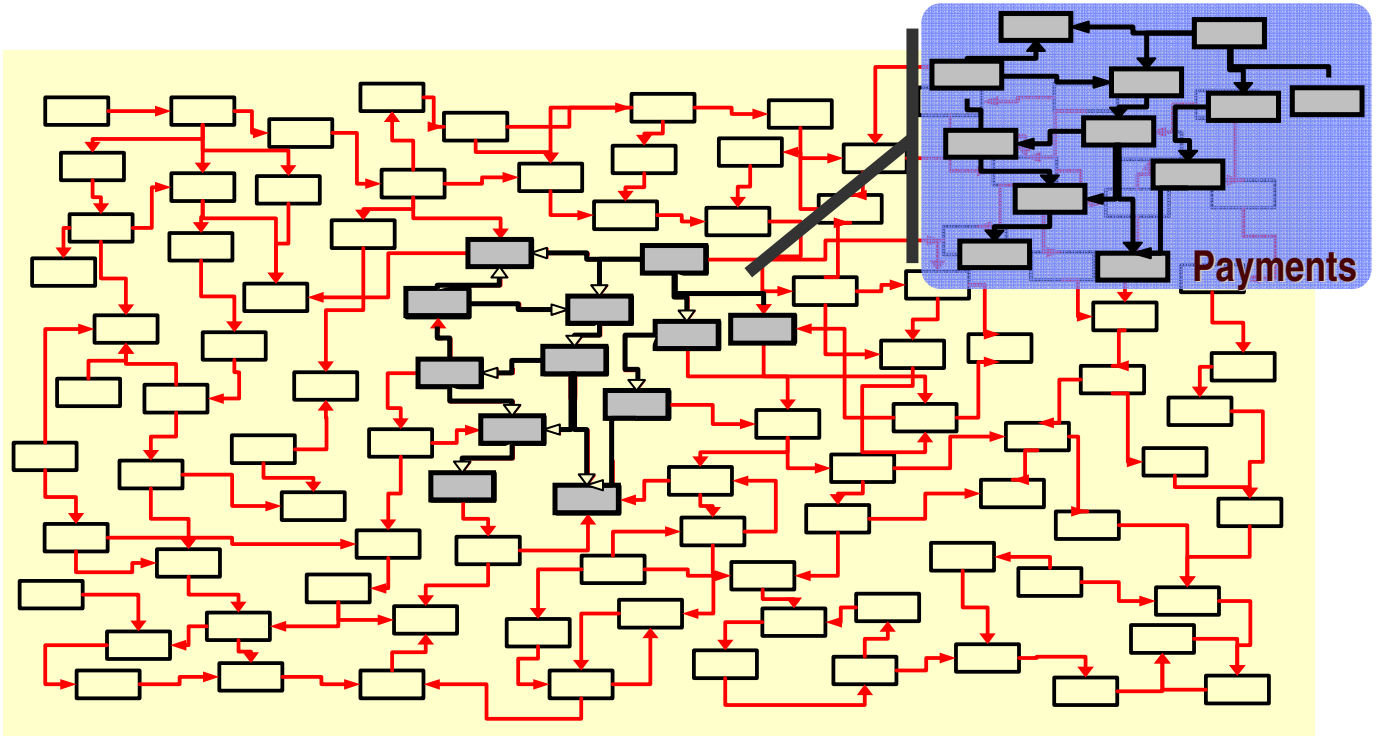
Enterprise Architecture

Complete Business Object

Extract, Store & Restore

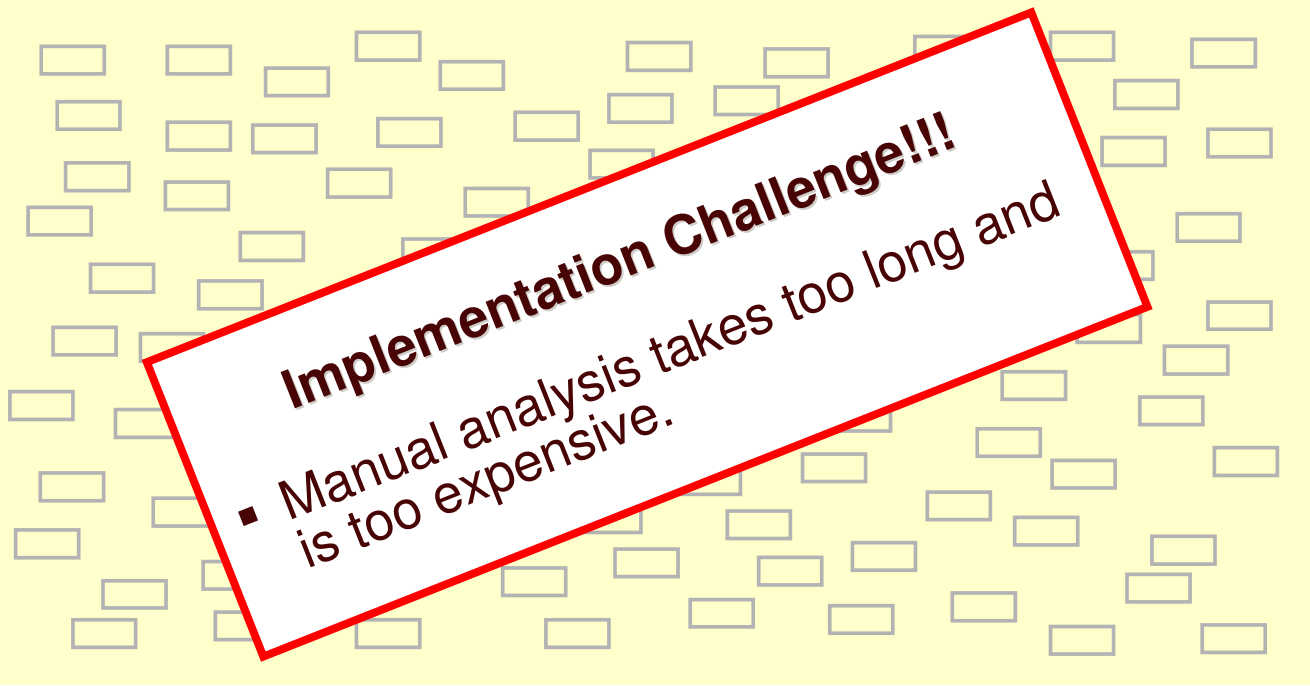
Universal Access

Complete Business Object



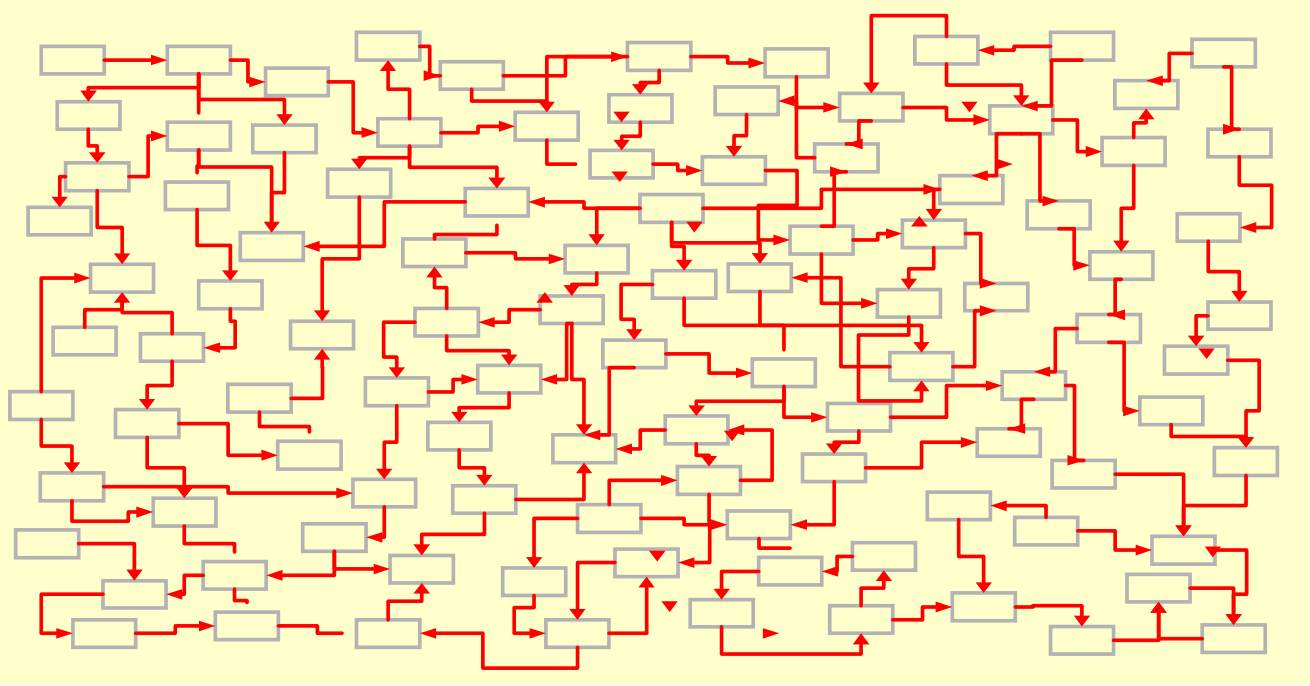
- **Represents application data record – payment, invoice, customer**
  - Referentially-intact subset of data across related tables and applications; includes metadata
- **Provides "historical reference snapshot" of business activity**
- **Federated extract support across enterprise data stores**

Complete Business Object: The Challenge



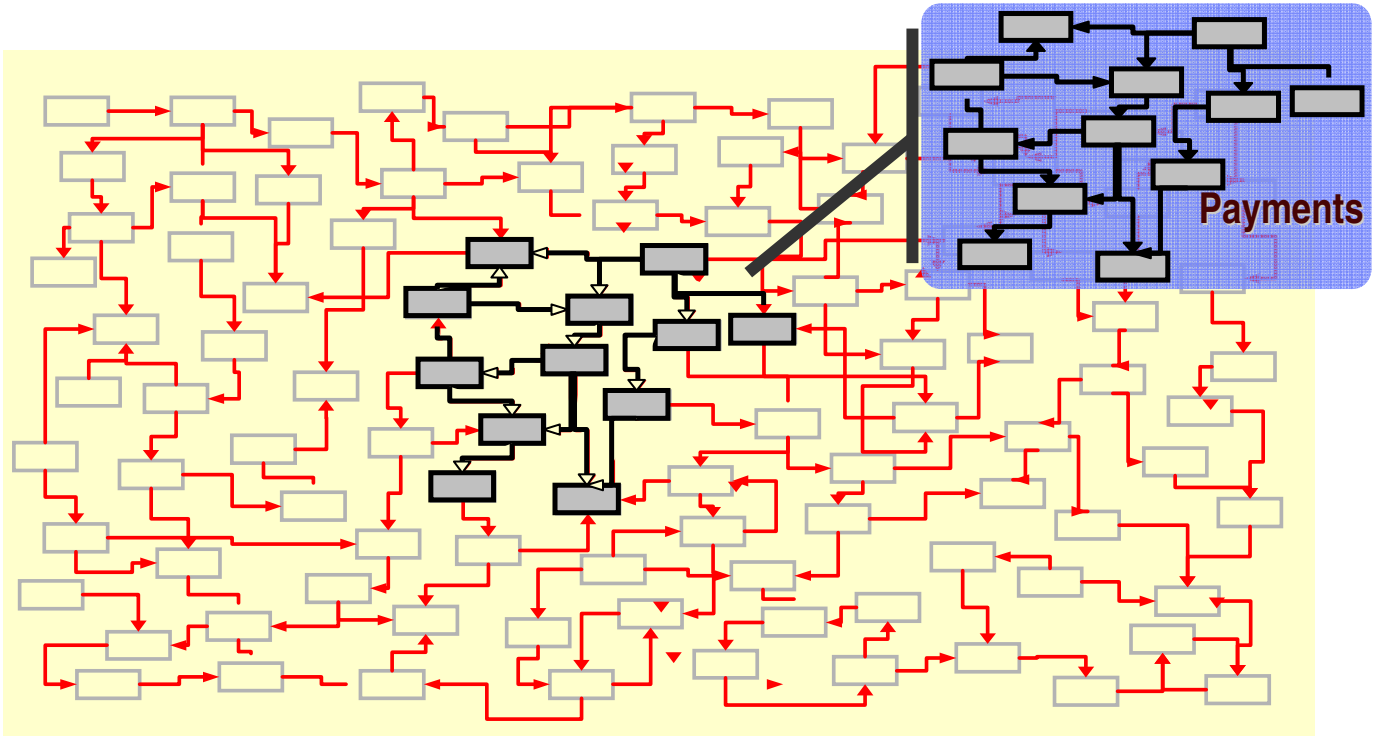
- **Where are they?**
- **What are they?**
- **How do I find them?**

Complete Business Object: Automated Discovery Solution



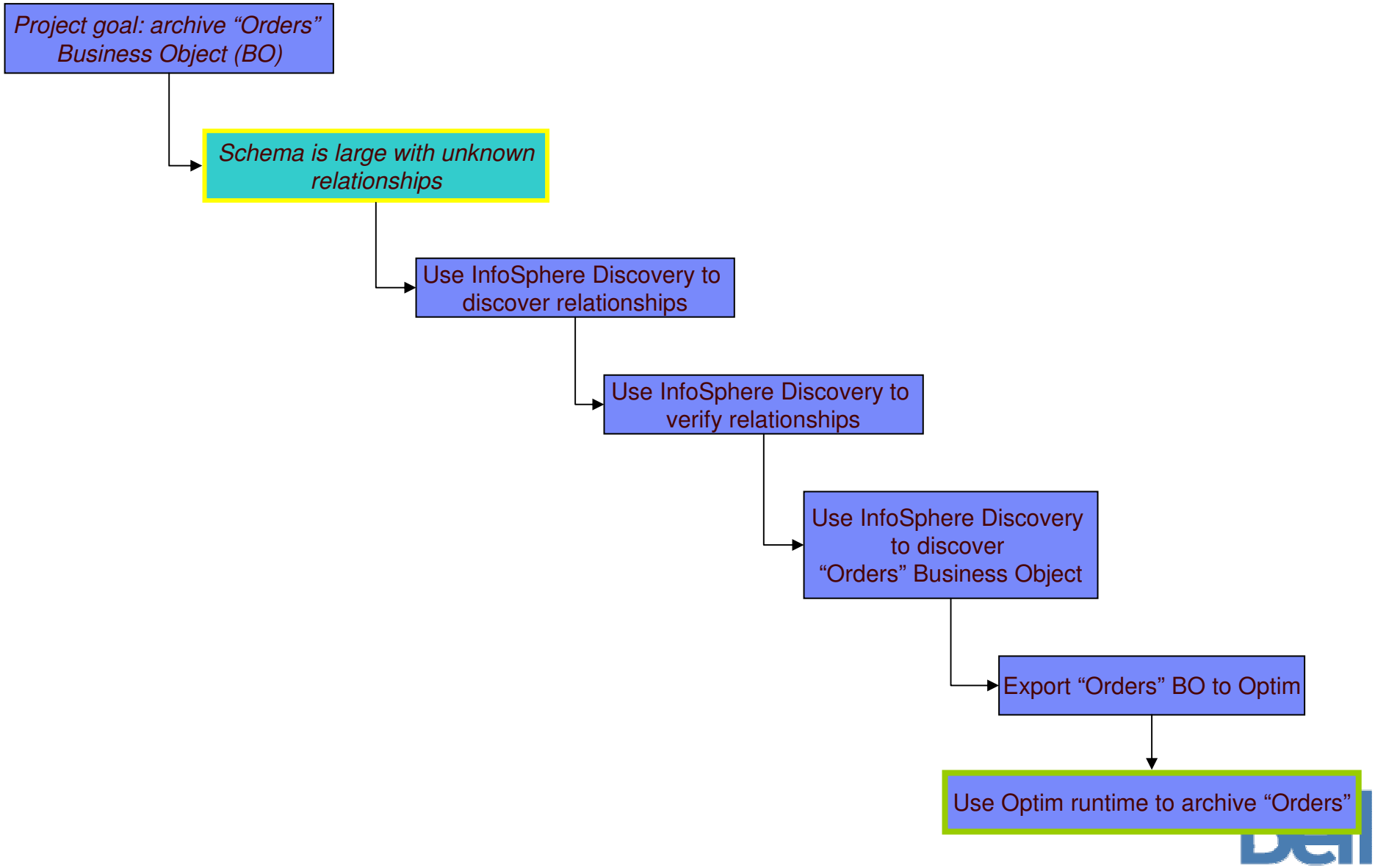
- **Automated discovery of Primary Foreign Keys**

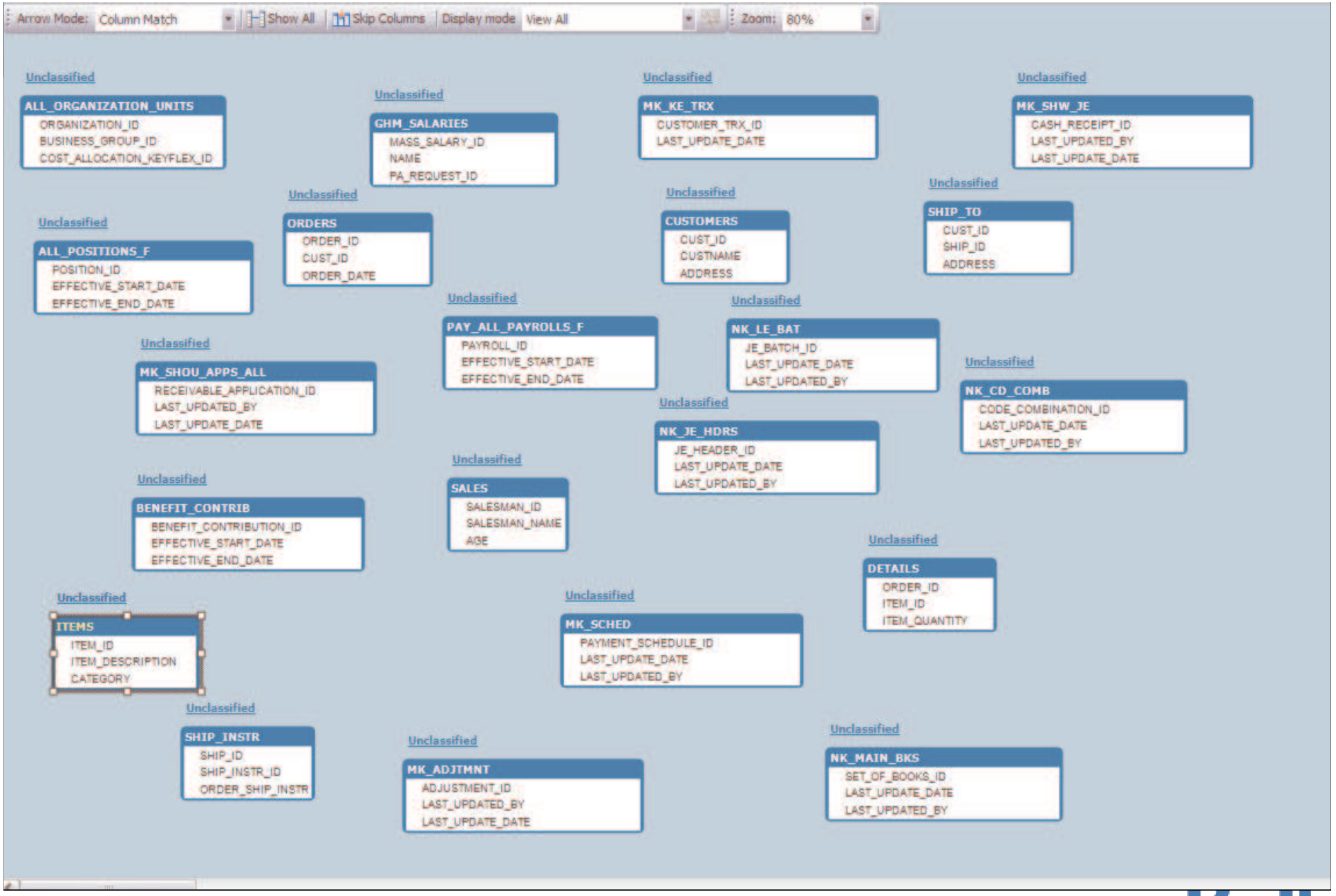
Complete Business Object: Automated Discovery Solution



- **Automated grouping of tables into business entities**
  - Optim will automatically generate service definition/requests based on these entities.

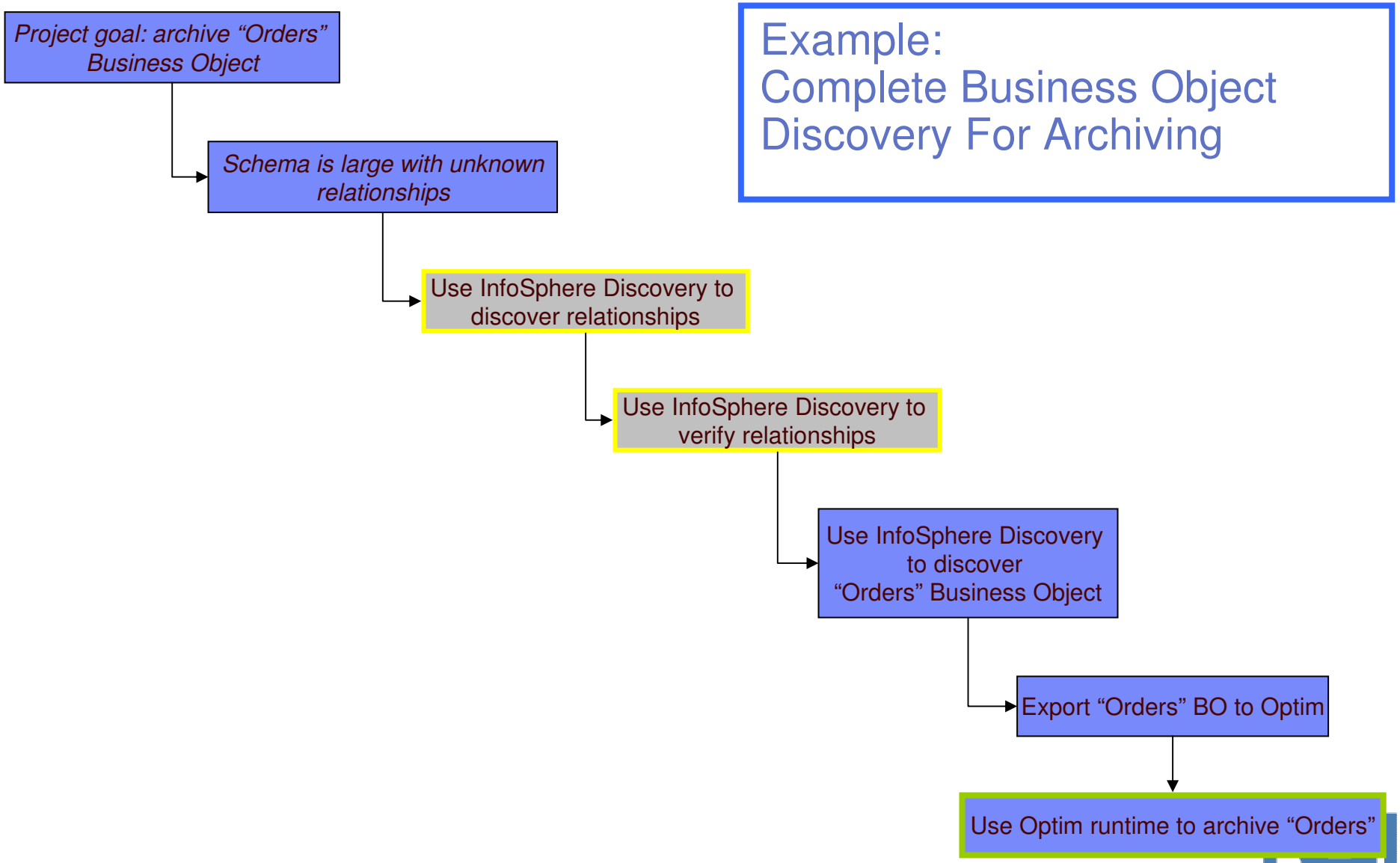
Example:
Using InfoSphere Discovery to Perform Complete Business Object Discovery For Archiving

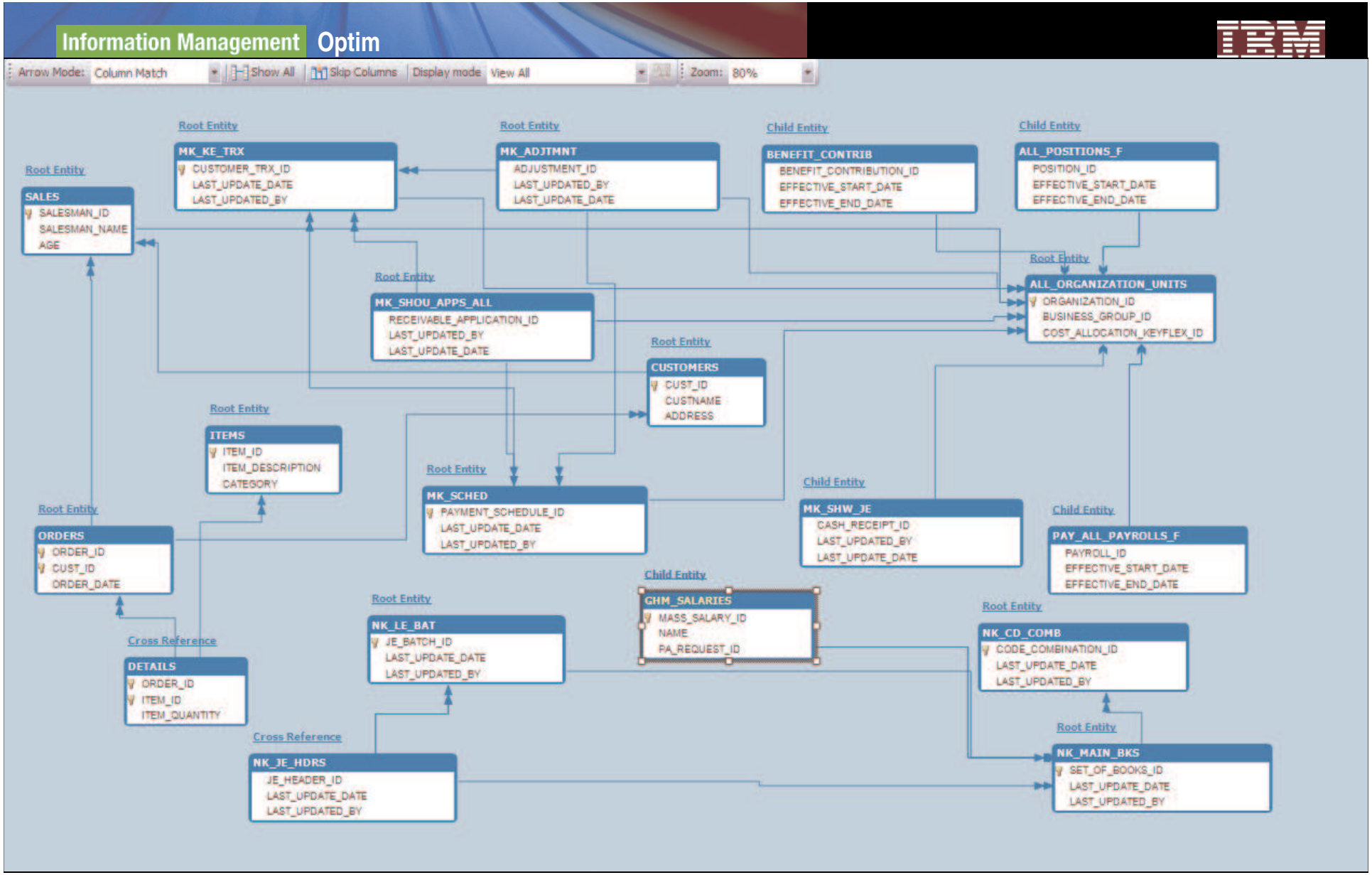




How many tables to archive for Orders, where are they?

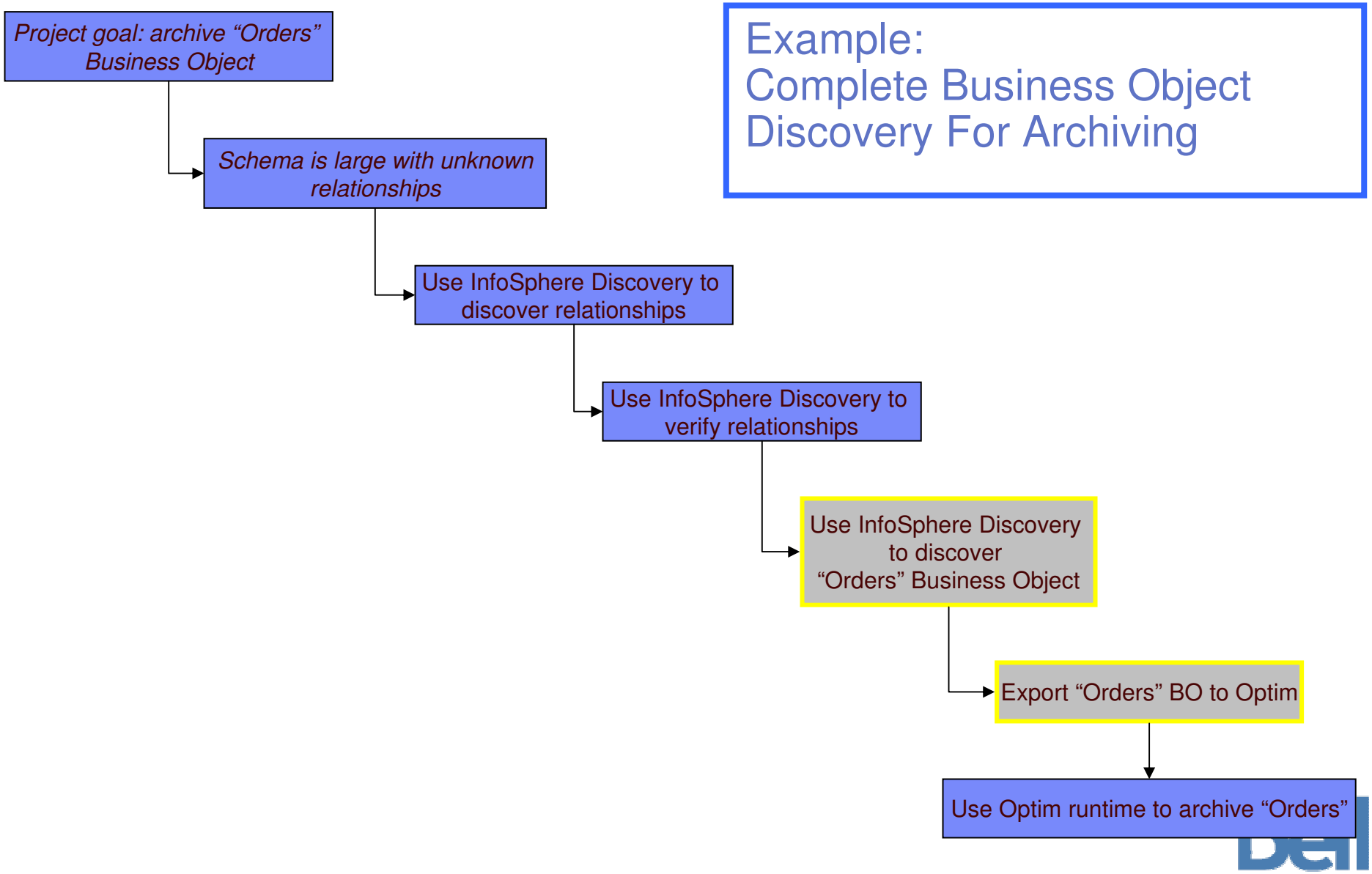
Example: Complete Business Object Discovery For Archiving

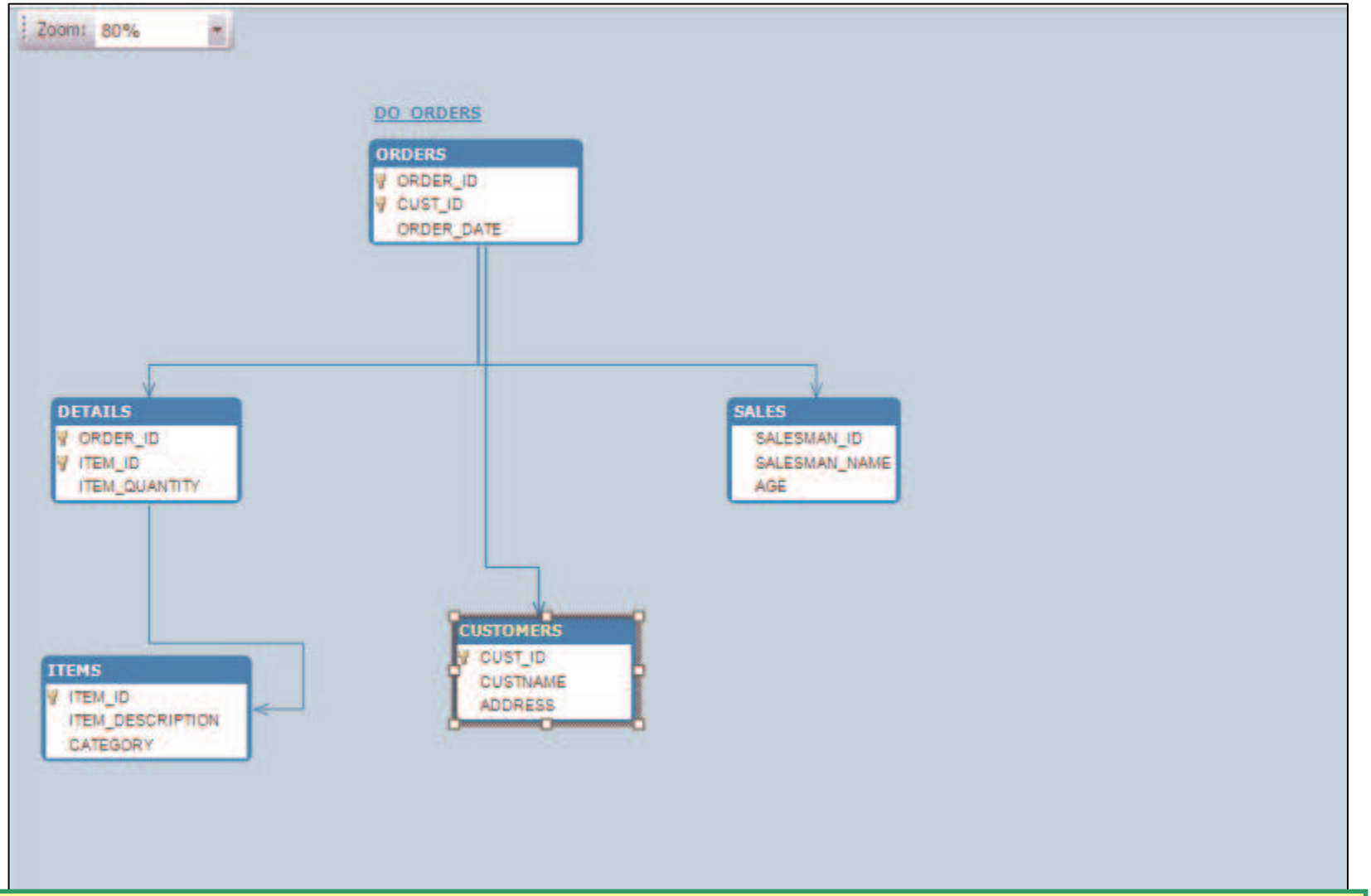




After PFkey discovery…still a lot to work with. Use zoom-and-focus features to review and confirm relationships around Orders table

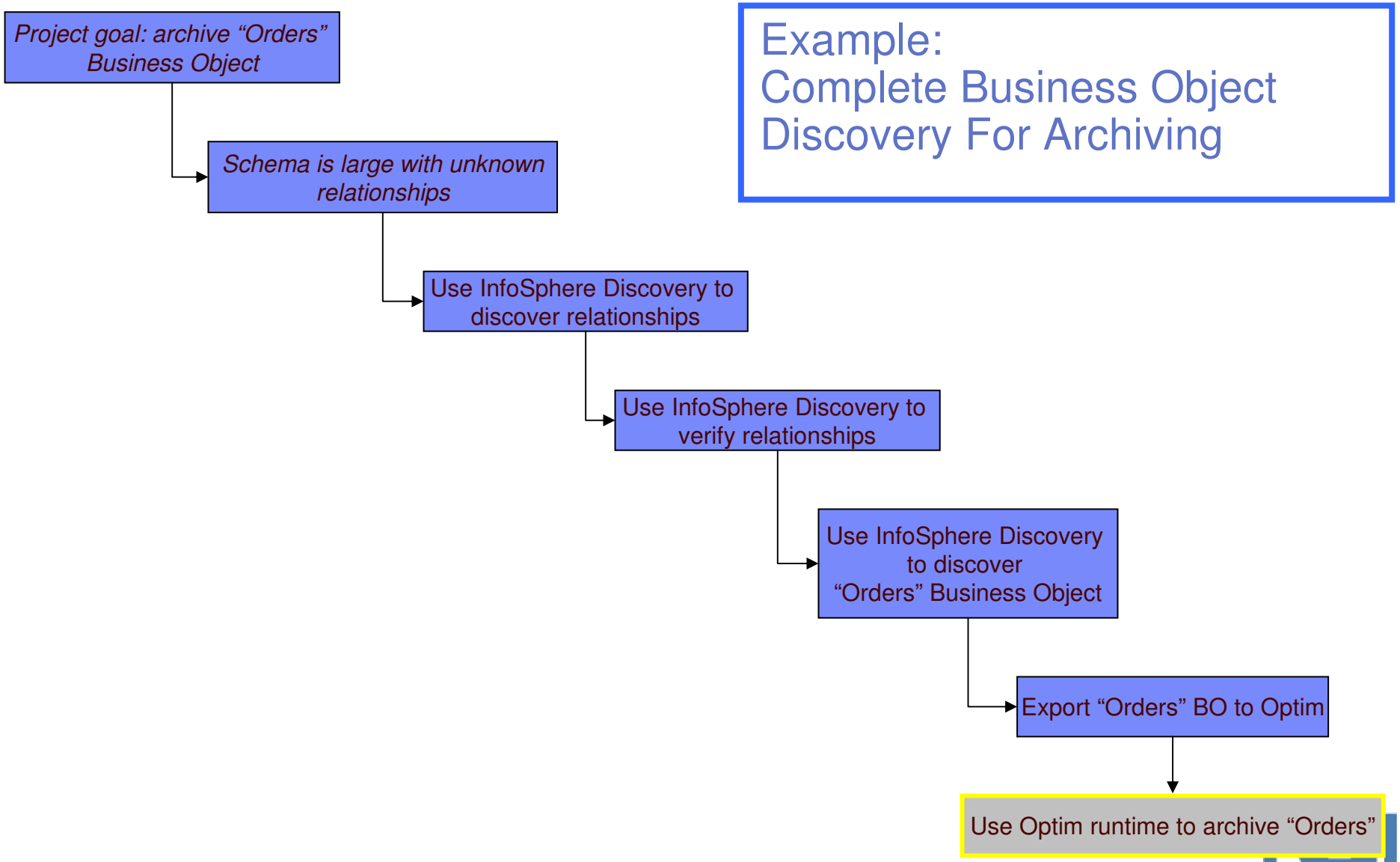
Example: Complete Business Object Discovery For Archiving

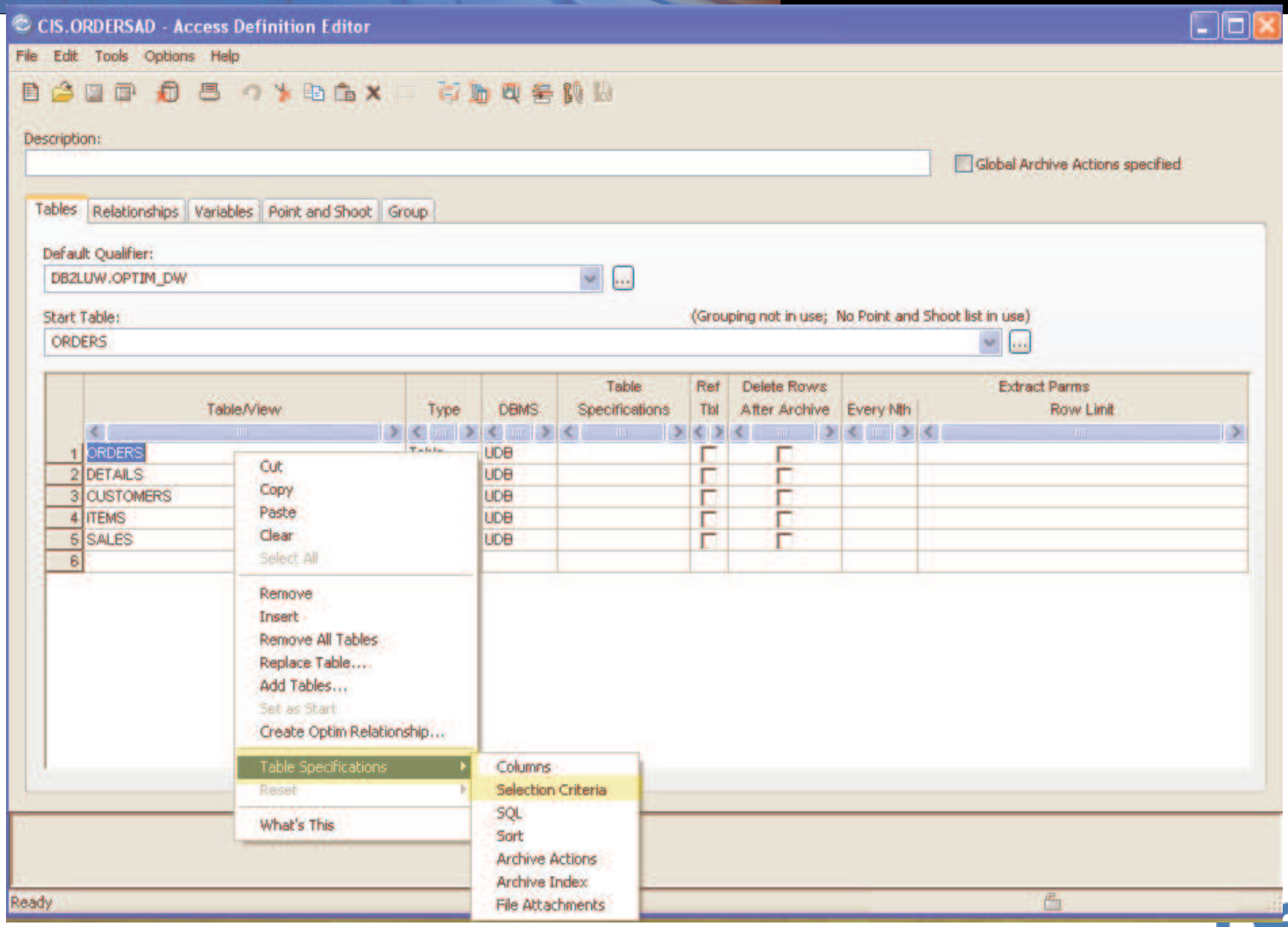


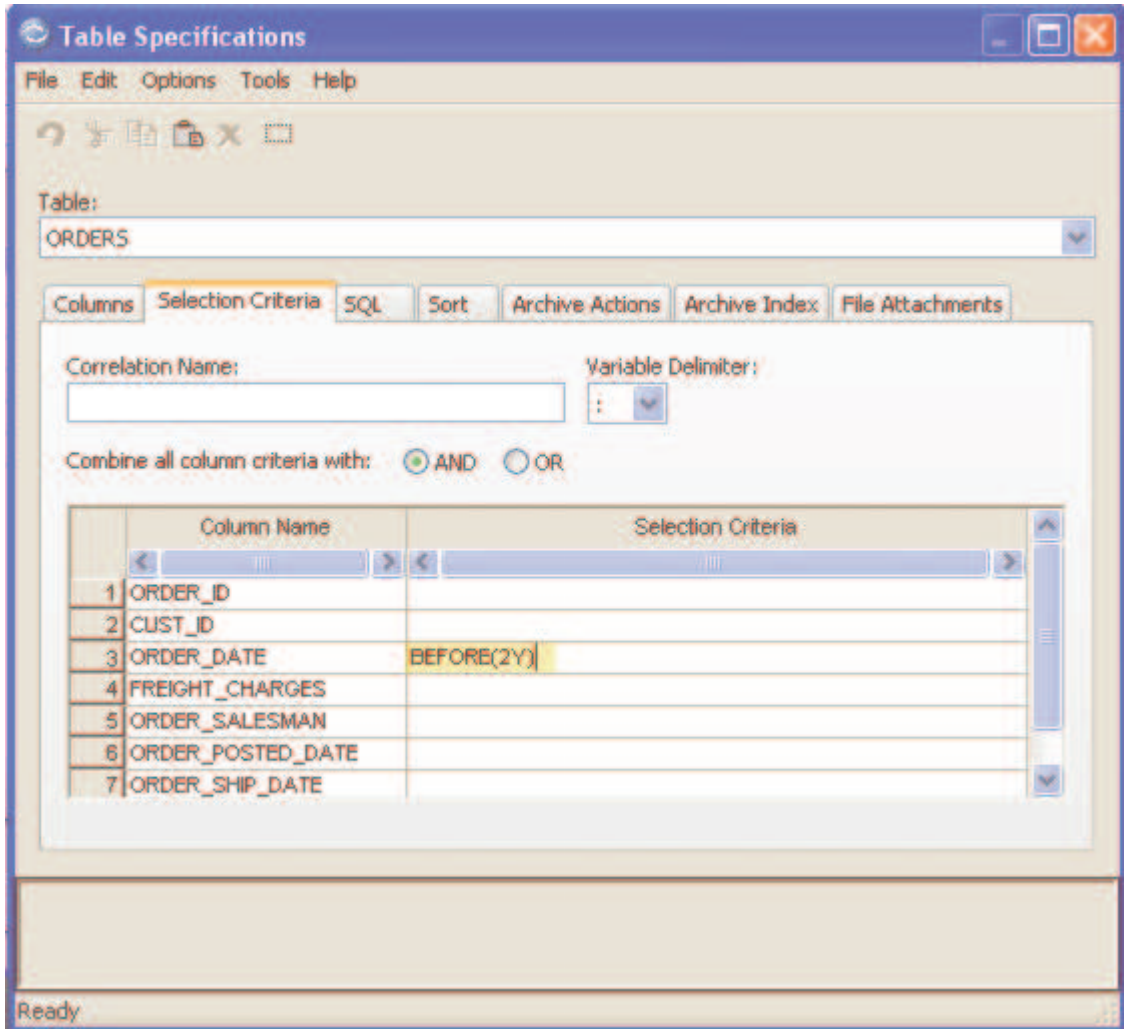


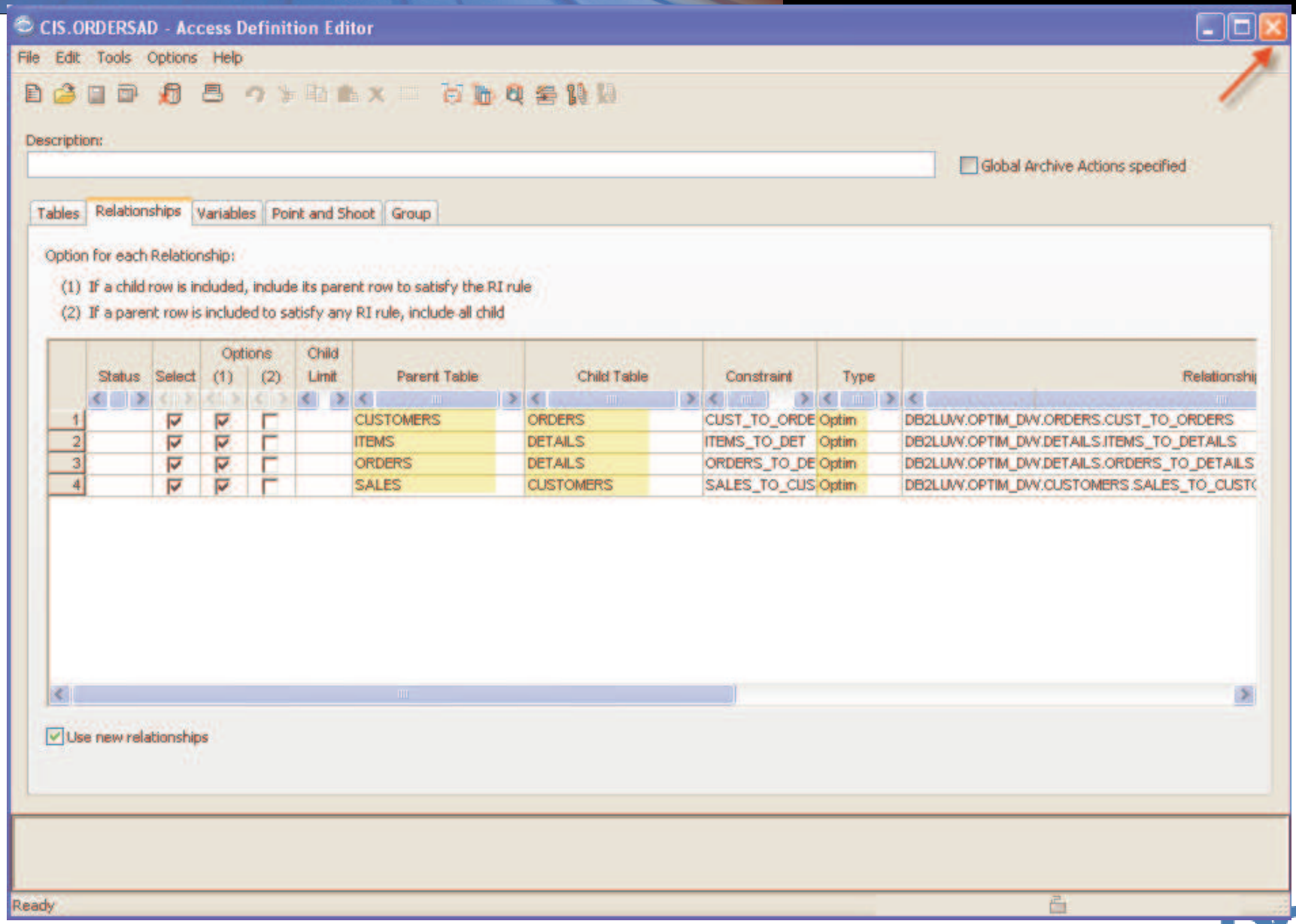
Once confirmed all relevant keys, use Data Object Discovery to produce business object. Export this object to Optim.

Example: Complete Business Object Discovery For Archiving









Information Management Optim

InfoSphere Discovery Lab

Bell

Additional Resources

Education Options

- **IBM Optim**
  - Optim Data Growth for Open Systems
    - http://www-304.ibm.com/jct03001c/services/learning/ites.wss/us/en?pageType=course_description&courseCode=DT200
  - Optim Data Privacy for Open Systems
    - http://www-304.ibm.com/jct03001c/services/learning/ites.wss/us/en?pageType=course_description&courseCode=DT210
  - Optim Test Data Management for Open Systems
    - http://www-304.ibm.com/jct03001c/services/learning/ites.wss/us/en?pageType=course_description&courseCode=DT220
- **IBM Federation Server**
  - Federation Server Fundamentals
    - http://www-304.ibm.com/jct03001c/services/learning/ites.wss/us/en?pageType=course_description&courseCode=DX900

IBM Federation Server Links

- **IBM Federation Server Documentation**
  - http://www-1.ibm.com/support/docview.wss?rs=845&uid=swg27011375
    - Configuration Guide for Federated Data Sources
      - SC19-1034-01
    - Installation Guide for Windows, Linux, Unix
      - GC19-1017-01
    - Administration Guide for Federated Systems
      - SC19-1020-01