Technical Overview and Demonstration

THE POWER OF
ENTERPRISE DATA
MANAGEMENT

© 2005 Princeton Softech, Inc.

princeton softech

Why are we here?

- Data Growth volumes having a negative impact on performance of online transactions and reporting.

- Increase Performance of Application, Reports and Faster Response Times = consistent user experience and increased customer satisfaction.

- Enterprise Data Management:
  - Database Archiving
  - Test Data Management
  - Data Privacy

© 2005 Princeton Softech, Inc.

Why is Privacy Protection Important?

Top Incidents

Organization	People Impacted	Date of Disclosure Incident
CardSystems	40 million	June 17, 2005
DSW Shoe Warehouse	1.4 million	March 8, 2005
Bank of America	1.2 million	Feb. 25, 2005
Wachovia, Bank of America, PNC Financial Services Group, Commerce Bancorp	676,000	April 28, 2005
Time Warner	600,000	May 2, 2005
Georgia Department of Motor Vehicles	465,000	April 2005
LexisNexis	310,000	March 9, 2005
University of Southern California	270,000	July 19, 2005
Marriott International	206,000	Dec. 28, 2005

Note: As of March 2006
Data: Privacy Rights Clearinghouse, InformationWeek

softech

© 2005 Princeton Softech, Inc.

Why I$ Privacy Protection Important?

Just read the headlines!

Organization	Financial Impact
ChoicePoint	FTC Fine, $15mm
Providence Health & Services	$7 - $9mm, excluding litigation
BJ's WHOLESALE CLUB® Where values come to life.®	$10mm plus 20 years of audits
DSW	20 years of audits

Source: Information Week, March 20, 2006

princeton softech

© 2005 Princeton Softech, Inc.

Protecting Test Environments

Test Data Management

from

Forrester Research:

"…IT's own access to customer and personnel data must be examined – strictly speaking, none should actually be necessary. *Test data must be "anonymized*…. "

Information Week:

"The search for consumer data and its uses doesn't stop at large production databases -- *it extends to application test data* and Web applications."

© 2005 Princeton Softech, Inc.

Data Privacy – The Needs

- No longer confidential; acceptable to use in open test system
- Address issue across the entire organization
  - Variety of data sources and platforms
- Intelligent: Syntactically valid / Appropriate “Look and Feel”
- Transformations
  - Generic (sequential, random, arithmetic, string)
  - Specific to field type
- Consistency – Within row, across rows, across tables
  - Same field, related fields
- Replacement data
- Performance
- Application Aware
- Services (Experience, Best Practices)

© 2005 Princeton Softech, Inc.

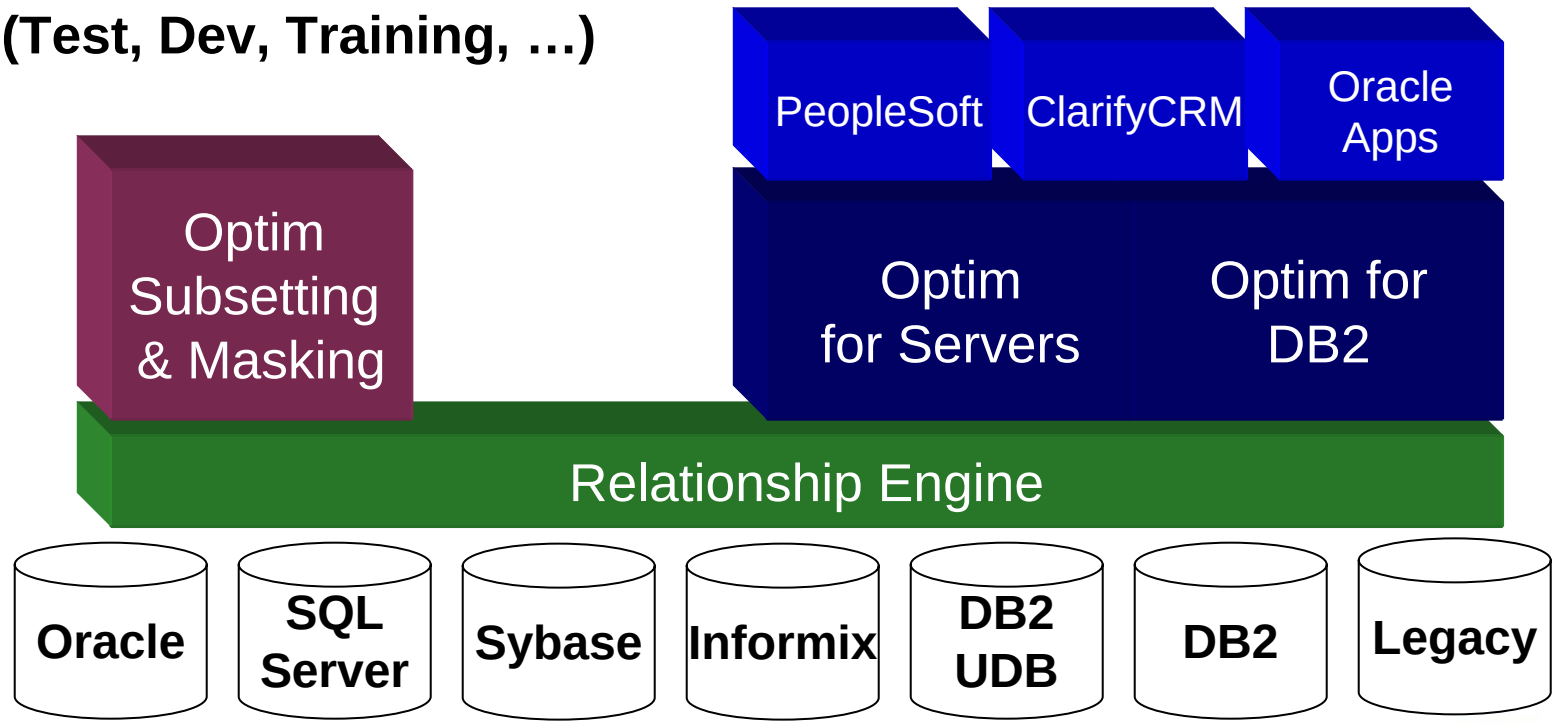
Manage Your Enterprise Data Smarter

Optim Subsetting

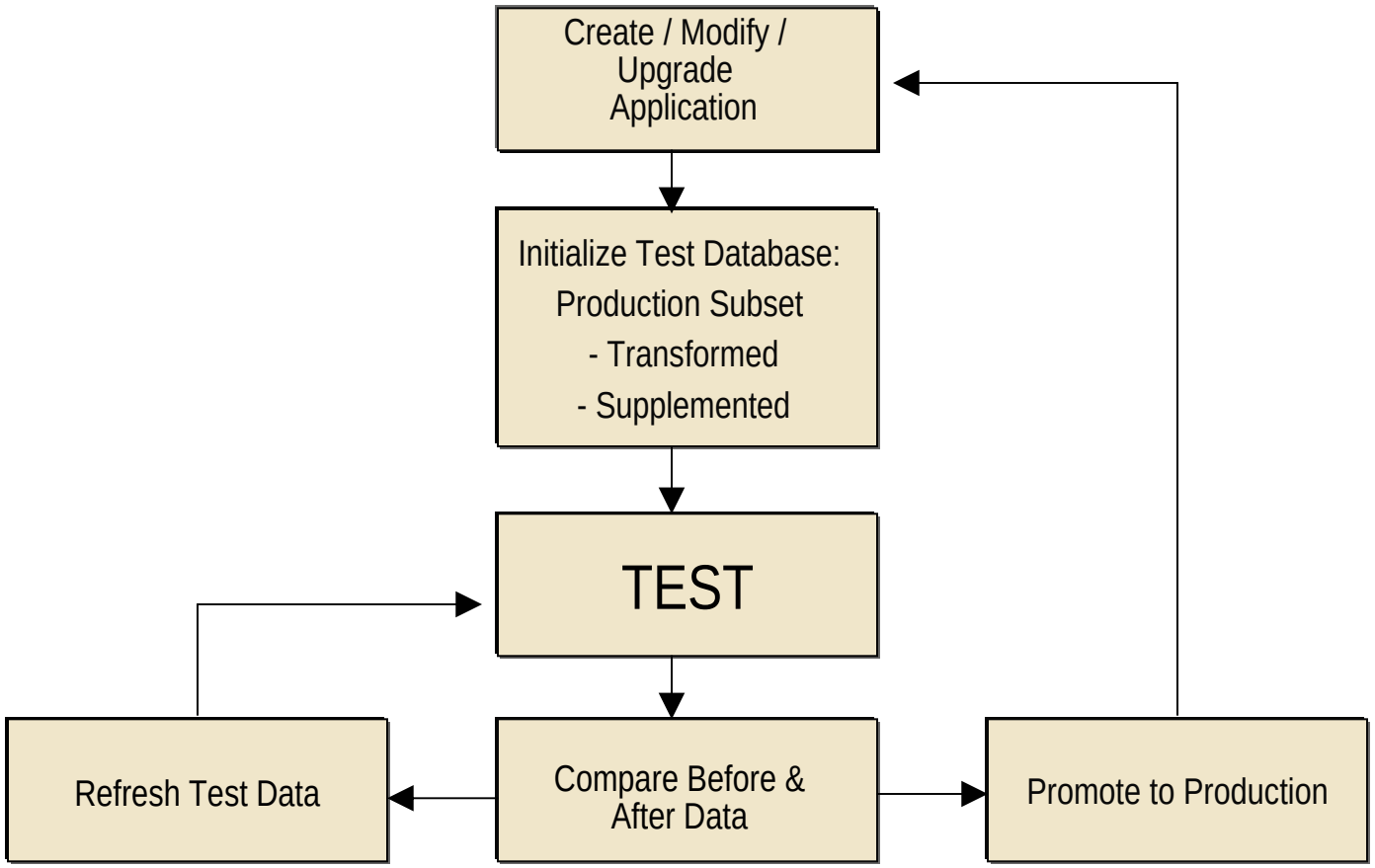
Optim Archiving

Pre-Production
(Test, Dev, Training, …)

Production

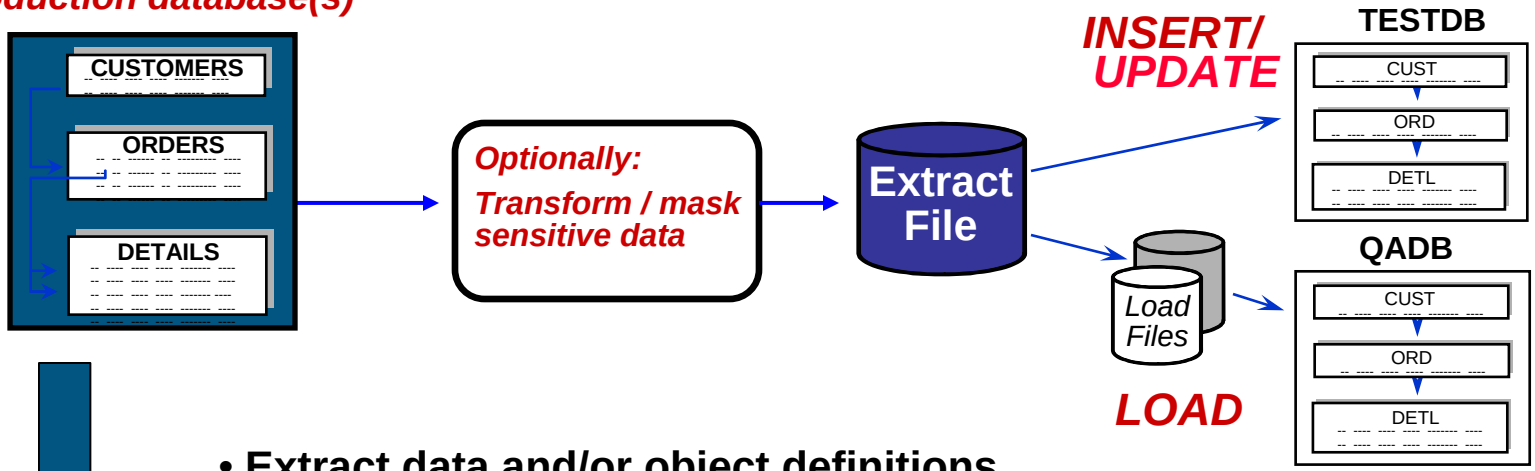


Application Development Cycle



Migrating Data

Extract a relationally intact subset from production database(s)

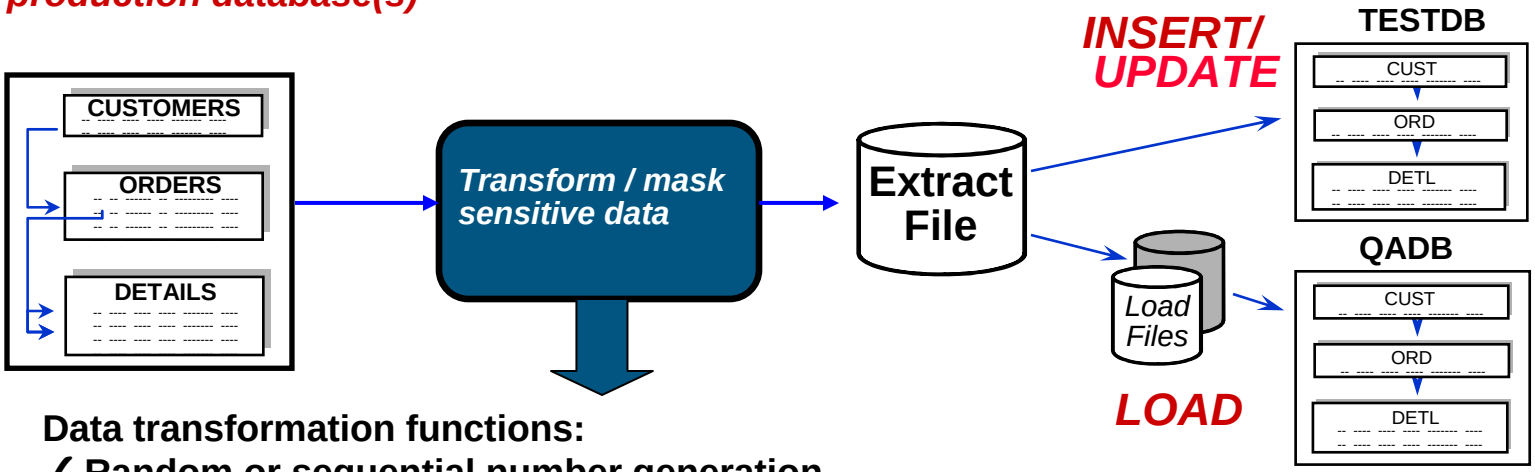


- **Extract data and/or object definitions**
- **Extract from source tables using dynamic SQL**
- **Select start table or define a set of tables**
- **Enter selection criteria**
  - **SQL where-clause**
  - **random rows, data partitioning & sampling**
- **Control traversal; Show steps**

© 2005 Princeton Softech, Inc.

Data Privacy in Application Testing

Extract a relationally intact subset from production database(s)



Data transformation functions:

- ✓ **Random or sequential number generation**
- ✓ **Hard-code literals, special registers such as date, time**
- ✓ **Substring and concatenation of values**
- ✓ **String sequencing**
- ✓ **Arithmetic calculations**
- ✓ **Lookup tables (sample Names, CASS addresses, custom)**
- ✓ **Intelligent masking (SSN, Credit Card Numbers)**
- ✓ **Access to client-defined exit routines to apply complex algorithms**
- ✓ **Propagation of masked primary keys to dependent foreign keys**

princeton softech

© 2005 Princeton Softech, Inc.

Column Map Functions

- Mask alphanumeric data to substitute confidential data with contextually-accurate but fictionalized data
- Character Substrings
  - Returns a substring of the contents of the named column
  - SUBSTR(column_name, start, length)
    - Example: SUBSTR(CUST_ID,1,3)
- Random or Sequential Numbers
  - RAND(low, high)
    - Example: RAND(1000,10000)
  - SEQ(start, step)
    - Example: SEQ(100,5)

© 2005 Princeton Softech, Inc.

Column Map Functions

- Sequential String function to a field (key or not key)
  - e.g., Salesman_ID column: Use the column name to generate values like Salesman_ID1, Salesman_ID2, Salesman_ID3, etc.

© 2005 Princeton Softech, Inc.

Column Map: Numeric Expressions

- Numeric Expressions
- Operand1 (operator) Operand2
  - Supports addition, subtraction, multiplication or division
  - One value must be a column name
    - Example: FREIGHT_CHARGES * 1.1

© 2005 Princeton Softech, Inc.

Column Map: Concatenated Expressions

- Example:
  - 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 number.
  - Mask account numbers for testing, using the actual first three digits, plus a sequential four-digit number.
  - SUBSTR(AC_NO,1,3) + '-' + SEQ(1111,1)

© 2005 Princeton Softech, Inc.

Column Map: Concatenated Expressions

- Concatenated Expressions
  - Example:
    - ‘PST’ concat SUBSTR(CUST_ID,4,3) concat SEQ(1,1)
    - Assume CUST_ID is (‘000146’, ‘003586’); generated values are PST1461 and PST5862

© 2005 Princeton Softech, Inc.

Column Map: Date Aging

- Date Aging Function
  - Processes any character, numeric, DATE or TIMESTAMP column
  - Ages to an explicit date or relative to a target date
  - Semantically ages data based on your business rules

© 2005 Princeton Softech, Inc.

Column Map: Lookup Values

- Lookup Function obtains the value for a destination column from a lookup table, according to the value in the source column:
  - LOOKUP(*source,table_name(search,value)cache*)

Example:

LOOKUP(STATE,STATE_LOOKUP(CODE,NAME)), where:

STATE	=	Source column to be masked
STATE_LOOKUP	=	Lookup table name to be searched
CODE	=	Value in the STATE_LOOKUP table that matches the source column value
NAME	=	Value inserted into the destination column

© 2005 Princeton Softech, Inc.

Column Map: Random Lookup

- Random Lookup Function obtains the value for a destination column from a value selected at random from a column in a lookup table:
  - RAND_LOOKUP(*table_name,columnname,limit*)

Example:

RAND_LOOKUP(CUSTOMERS,STATE,500), where:

STATE	=	Lookup table column name for substitution data
CUSTOMERS	=	Lookup table name to be searched
500	=	(optional) Limit on number of rows used to select column values

© 2005 Princeton Softech, Inc.

Column Map: Pre-Packaged Lookups

- Names Lookup Table
  - list of more than 5,000 male/female first names
  - list of more than 88,000 last names
  - taken from US Census Bureau list of names

- Company Names Lookup Table
  - contains over 11,000 values
  - taken from US Postal CASS list

- Street Address/City/State/Zip Code Lookup Table
  - contains corresponding Street / City / State / Zip Codes for over 111,000 US locations
  - based on US Postal CASS list

© 2005 Princeton Softech, Inc.

Column Map: Multi-Column Assignment Lookups

- Multi-column assignment on Lookups (Key and Random)
  - Ensure columns within row are consistent
  - Scenario: Name – First, Last
    Address – City, State, Zip

- Examples:

```
LOOKUP(SOC_SEC,DEST=(FIRST_NAME,LAST_NAME),
  NAME_LOOKUP(SSN,SOURCE=(FIRST_MASK,LAST_MASK)

RAND_LOOKUP(ADDR_LOOKUP,DEST=(CITY,STATE,ZIP),
  SOURCE=(CITY,STATE,ZIP))
```

© 2005 Princeton Softech, Inc.

Column Map: Email address

- Email address format: name@domain
  - Mask left or right side of "@"
  - Generated from destination name values or
  - Generated by using literal and concatenating a sequential number
    - Domain can be copied from source value or specified by literal
    - Flags can specify delimiter to include; use initial; specify case

© 2005 Princeton Softech, Inc.

Column Map: SSN Intelligent Masking

- Intelligent Masking
  - Generate valid social security numbers (as defined by the US Social Security Administration)
    - aaa-gg-ssss where
      - aaa is an Area
      - gg is a Group Number (Odd 01-09, Even 10-98, Even 02-08, Odd 11-99)
      - ssss is a sequential Serial Number
  - Processing
    - Random: use random area assignment in same state, valid (descending) Group Number & Serial Number
    - Consistent: keep same Area; alter / scramble Group Number & Serial Number

© 2005 Princeton Softech, Inc.

Column Map: Credit Card Intelligent Masking

- Intelligent Masking
  - Generate valid credit card numbers as defined by credit card issuers – ISO 7812 format
  - iiiiiinnnnnnnnnnnnc where
    - iiiiii is a 6-digit issuer identification (ex. 34nnnn is AMX T&E)
    - nnnnnnnnnnnn is a 7-12 digit Account Number defined by issuer
    - c is a Check Digit using the Luhn algorithm
- Processing:
  - Random: use random Issue Identifer, ascending Account Number and Check Digit
  - Consistent: keep same 4-digit Issuer Identifier,: alter/scramble remainder & Account Number; Assign Check Digit

© 2005 Princeton Softech, Inc.

Column Map: Key Propagation

- Key Propagation
  - Propagates values in the primary or foreign key to all related tables
  - Necessary to maintain referential integrity
- PROP(*value,column-name*)
  - *Value*: literal, expression, special register, exit routine
  - *Column-name*: specifies source column containing value that is subject of the function

EX: PROP(EXIT PX0SSN)

PROP(SEQ(10000,1),CUST_ID)

© 2005 Princeton Softech, Inc.

Maintaining Referential Integrity

Customers Table

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

Cust_ID is primary key.

Orders Table

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

- Customers is parent table; CUST_ID is Primary Key
- Mask CUST_ID using SEQ function
- Propagate masked value to foreign key of child table (Orders)
- PROP(SEQ(10000,1))

© 2005 Princeton Softech, Inc.

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

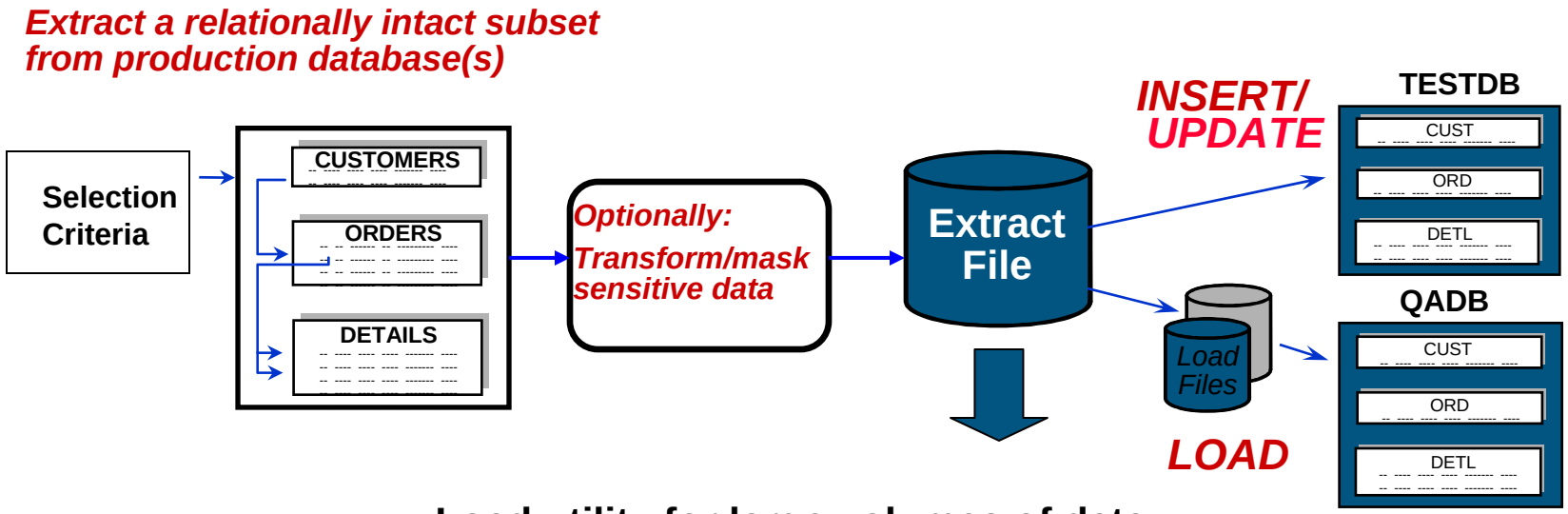
Cust ID	Name	Street
10000	Auguste Renoir	Mars23
10001	Claude Monet	Venus24
10002	Pablo Picasso	Saturn25

Orders Table

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

© 2005 Princeton Softech, Inc.

Migrating Data



- Load utility for large volumes of data
- Dynamic SQL
- Insert new rows
- Update existing rows; insert others
- Refresh from the Extract File
- Extract File maintains consistent baseline

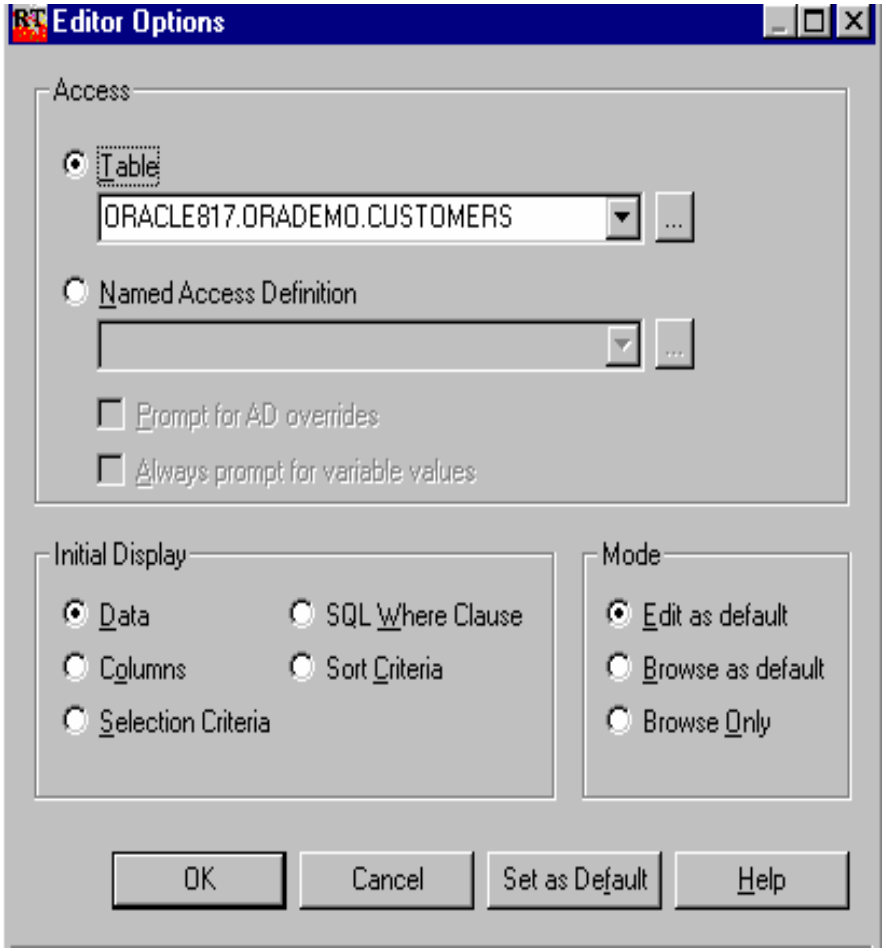
© 2005 Princeton Softech, Inc.

Editing Test Data

- Understand the data your application is to process
- Create data values to test program logic
- Inspect and correct data that is causing problems
- Verify execution results

© 2005 Princeton Softech, Inc.

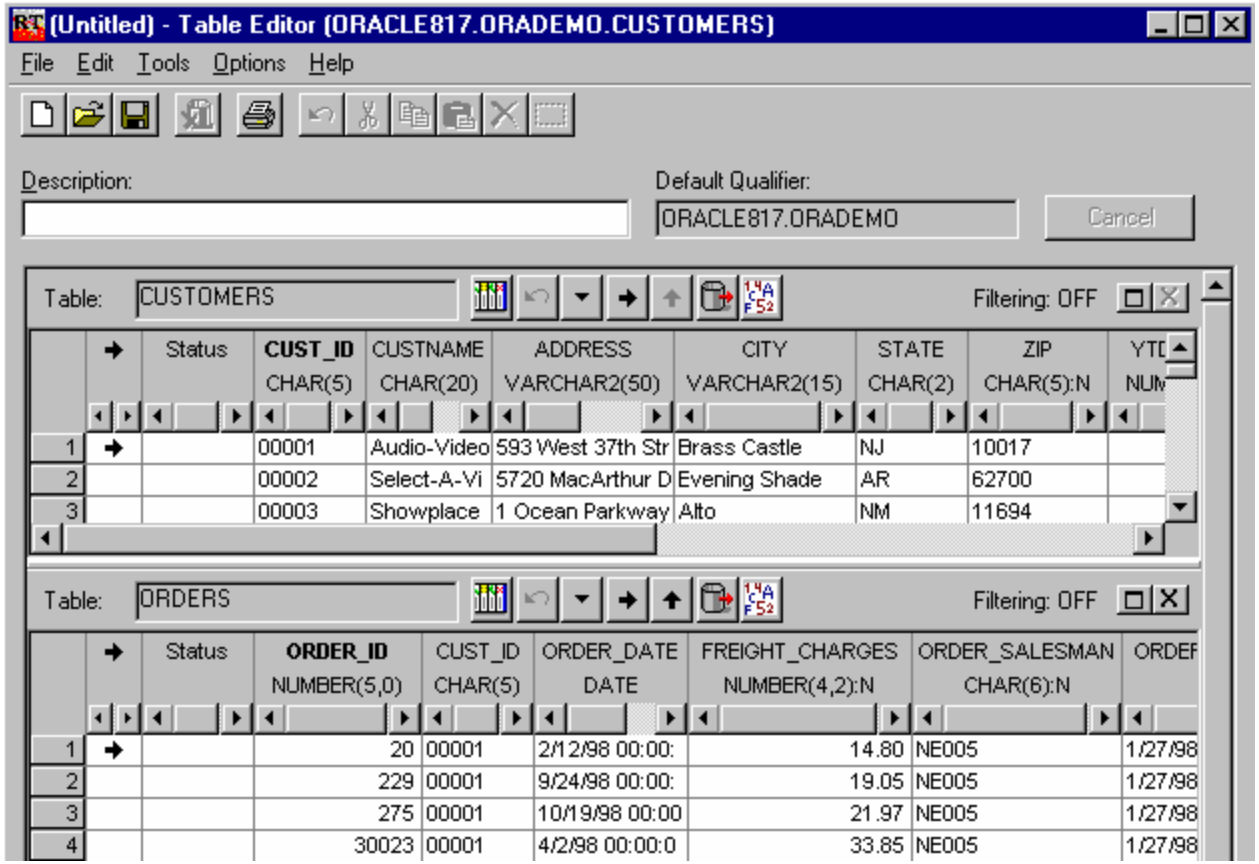
Beginning the Edit/Browse Session



- Select Browse or Edit Mode
- Display Options:
  - ✓ Data
  - ✓ Columns
  - ✓ Selection/SQL Criteria
  - ✓ Sort Criteria

© 2005 Princeton Softech, Inc.

Editing Data: Insert, Update and Delete



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

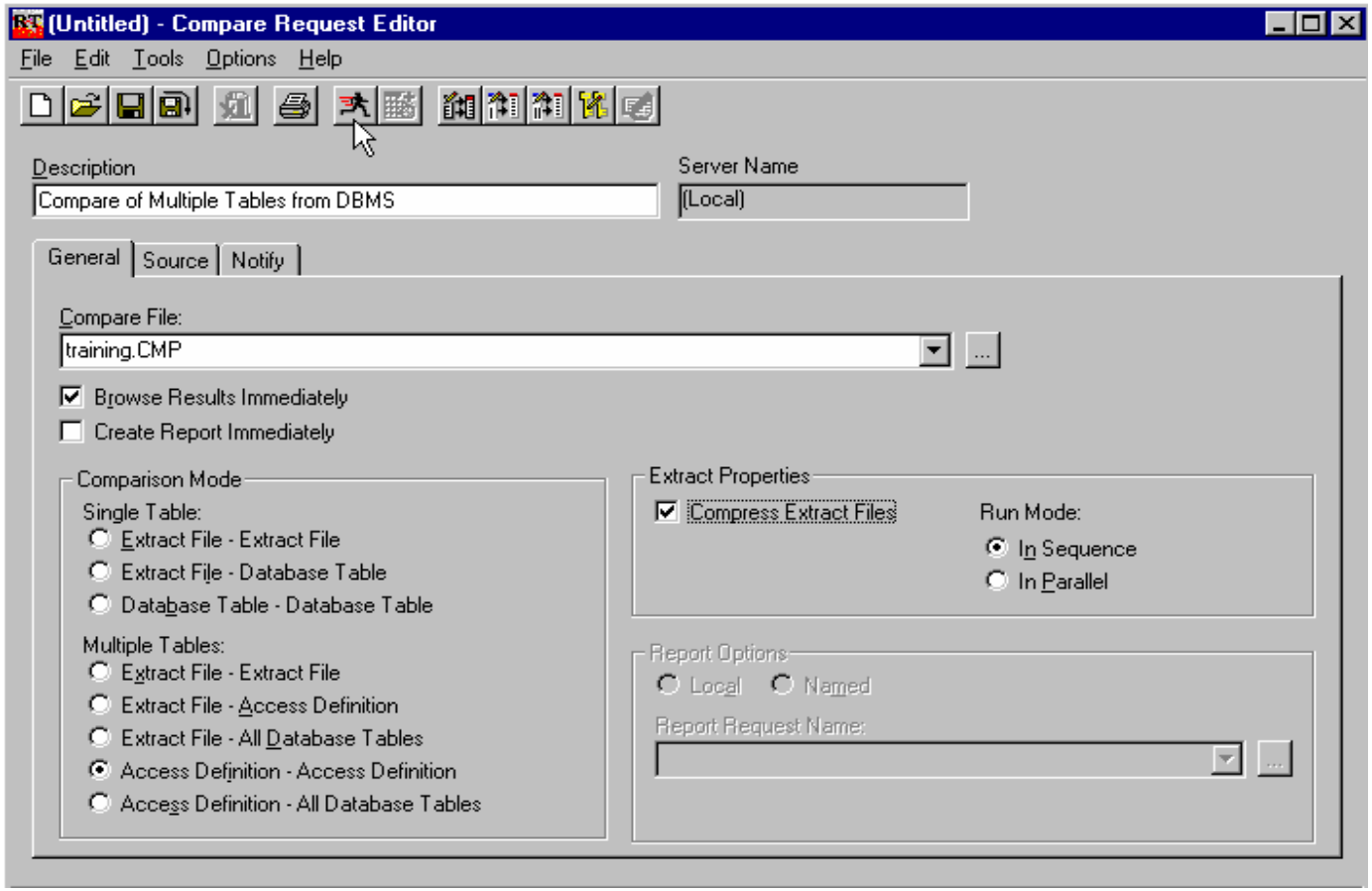
© 2005 Princeton Softech, Inc.

Comparing Data

- Compare the "before" and "after" data from an application test
- Compare results after running modified application during regression testing
- Identify differences between separate databases
- Audit changes to a database

© 2005 Princeton Softech, Inc.

Beginning the Compare



- Identify if a single table or multiple table compare
- Identify sources

Browsing the Compare File

c:\rt4s\rtdata\training.CMP - Browse Compare File

File Edit View Actions Options Help

Tables | Information

Source	Table Name	Total Rows	Unmatched Rows	Equal Rows	Different Rows	Rows with Duplicate Match Keys
1	ORACLE817.ORADEMO.CUSTOMERS	704	2	698	4	0
2	ORACLE817.ORADEMO.CUSTOMERS2	703	1			0
1	ORACLE817.ORADEMO.ORDERS	1717	4	1711	2	0
2	ORACLE817.ORADEMO.ORDERS2	1714	1			0
1	ORACLE817.ORADEMO.DETAILS	3619	10	3607	2	0
2	ORACLE817.ORADEMO.DETAILS2	3609	0			0
1	ORACLE817.ORADEMO.ITEMS	102	0	98	4	0
2	ORACLE817.ORADEMO.ITEMS2	102	0			0

- Generated for each pair of tables
- Identifies tables containing unmatched rows
- Identifies tables containing duplicate match keys

© 2005 Princeton Softech, Inc.

Browsing the Compare File

Browse Compare File Table Data

File Tools Options Help

Source 1: ORACLE817.ORADEMO.CUSTOMERS

	Change	Source	CUST_ID CHAR(5)	CUSTNAME CHAR(20)	ADDRESS VARCHAR2(50)	CITY VARCHAR2(15)	STATE CHAR(2)	ZIP CHAR(5):N
1	Equal	Both	00001	Audio-Video	593 West 37th Street	Brass Castle	NC	10017
2	Only	1	00002	Select-R-Vi	5720 MacArthur Drive	Evening Shade	AR	62700
3	Only	1	00003	Showplace	1 Ocean Parkway	Alto	NM	11694
4	Diff	1	00004	Audio-Video	593 West 37th Street	***Panacea***	***FL***	10017
5	Diff	2	00004	Audio-Video	593 West 37th Street	***Hollywood***	***CA***	10017
6	Equal	Both	00006	Main Street	Gateway Shopping C	Pumpkin Center	AZ	85002
7	Diff	1	00008	Director's C	***347 Miners Row***	Taters	FL	95800
8	Diff	2	00008	Director's C	***1986 PanHandlers***	Taters	FL	95800
9	Equal	Both	00009	Prime Time	64 Newberg Avenue	Loving	NM	22180
10	Diff	1	00010	***Reely Grea***	590 Frontage Rd	Christmas Vally	OR	01002
11	Diff	2	00010	***Ted's Excel***	590 Frontage Rd	Christmas Vally	OR	01002
12	Diff	1	00011	Director's C	347 Miners Row	Kiester	MN	***95800***
13	Diff	2	00011	Director's C	347 Miners Row	Kiester	MN	***10455***

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

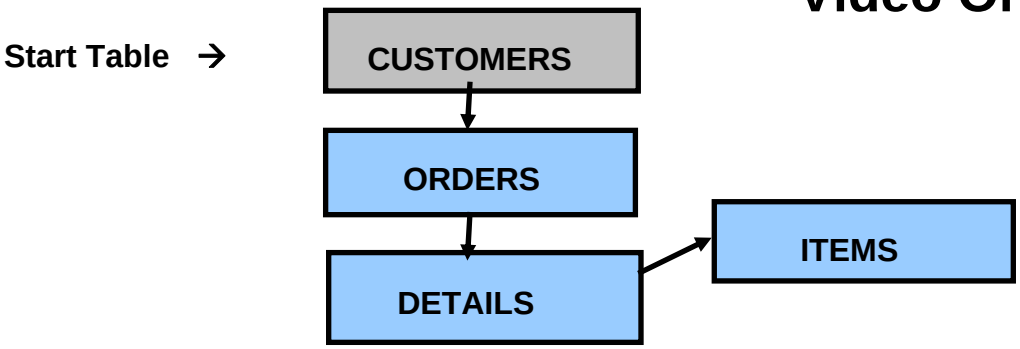
© 2005 Princeton Softech, Inc.

Product Demonstration

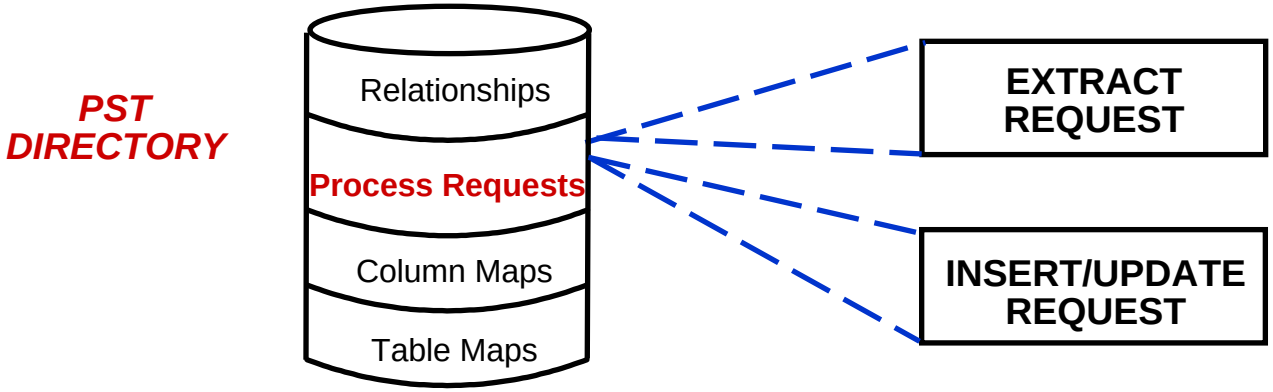
- **Extract a subset of Customers and their related information**
- **Convert data using mapping rules**
- **Populate a second set of test tables**
- **Edit test tables and compare changes**

Application Database:

Video Ordering System



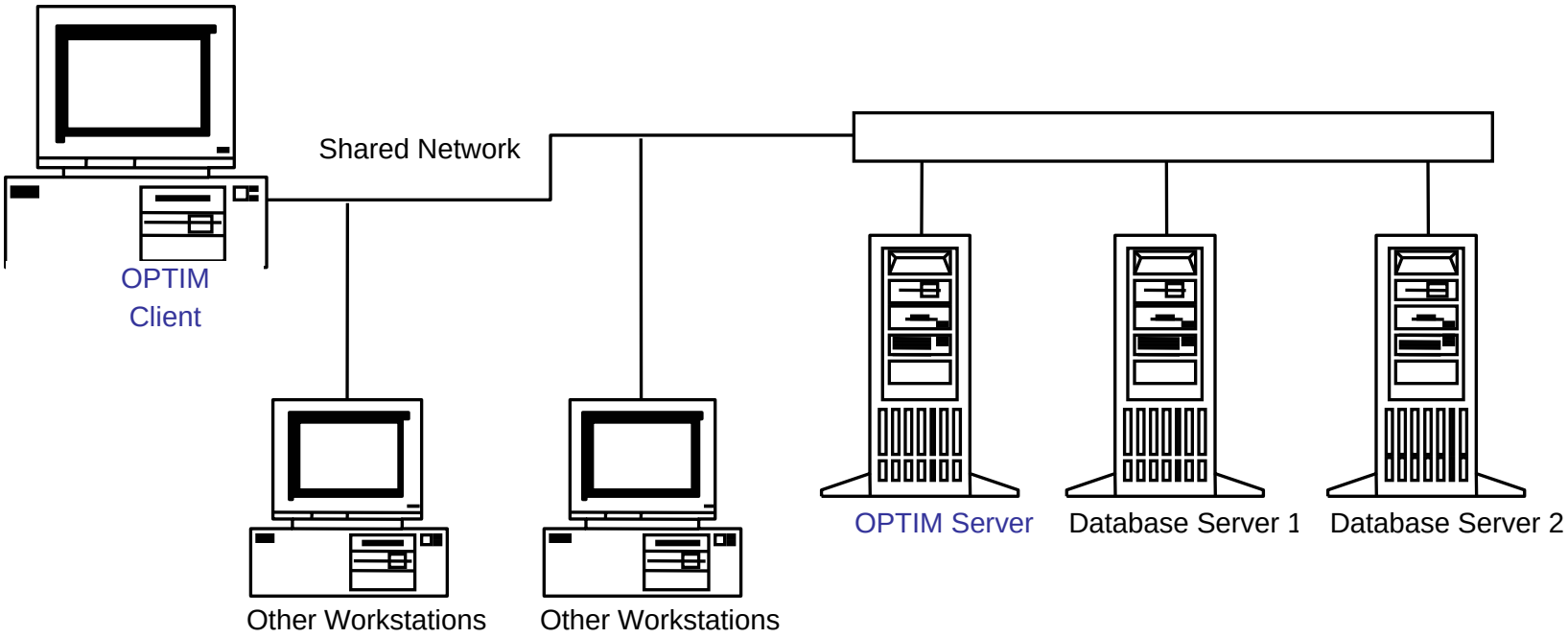
Process Requests



- Named ‘package’ of specifications for a process
  - Archive, Delete, Restore, Extract, Convert, Insert/Update or Load
- Can be saved and reused; can be scheduled

© 2005 Princeton Softech, Inc.

Implementation



- Separation of UI and execution
- OPTIM Server Installed on separate server(s) or on database servers
- Single OPTIM Server can connect to multiple database servers running different database types

© 2005 Princeton Softech, Inc.

Princeton Softech: Customers

Princeton Softech: Partners

VERITAS

EMC² where information lives

IBM

PeopleSoft

hp invent

SYMANTEC

BEACON

BROOKS-PRI AUTOMATION

STORAGETEK

NetworkAppliance

SYBASE

ORACLE

OSI

amdocs

btrg

SIEBEL Strategic Platform Partner

MERCURY INTERACTIVE

princeton softech

© 2005 Princeton Softech, Inc.

Questions?

princetonsoftech.com

111 Campus Drive
Princeton, NJ
08540-6400

800.457.7060
609.627.5500
609.627.7799 Fax

© 2005 Princeton Softech, Inc.