Optim™

Vállalati adatkezelés Optim megoldásokkal

IBM Information Management software

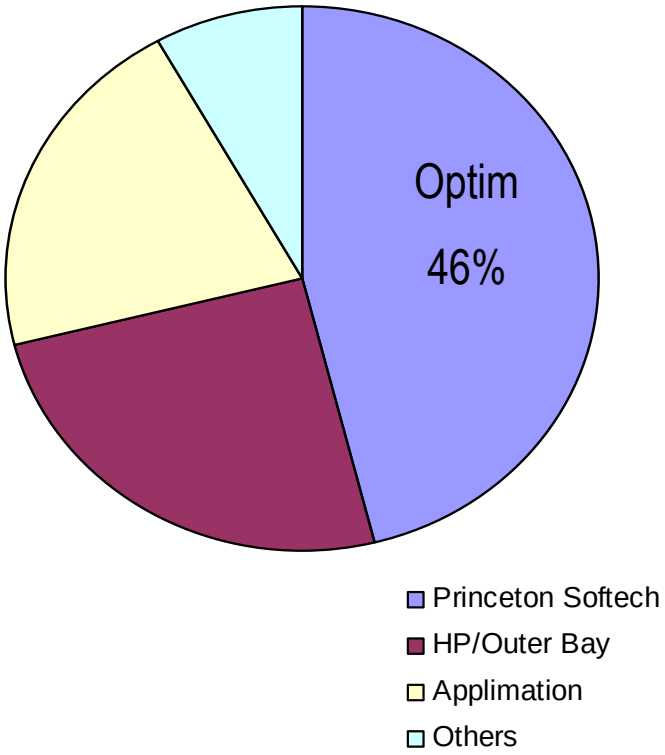
Temesi Gergely, IBM SWG

Information Management Technical Sales

gergely.temesi@hu.ibm.com

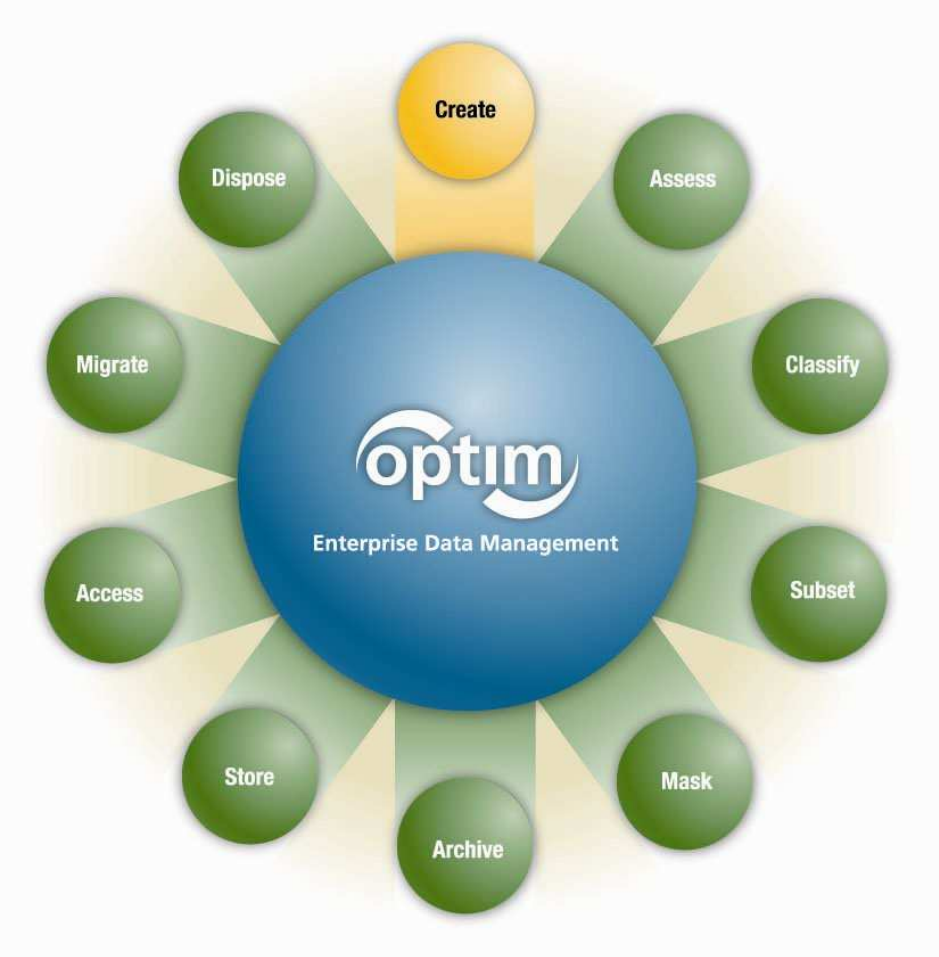
Princeton Softech

- Proven leader in **Enterprise Data Management** (EDM):
  - Data Growth
  - Retention & Discovery
  - Data Privacy
  - Test Data Management
  - Application Upgrades
  - Application Retirement
- **Solving data management issues since 1989**
- Partnered with major infrastructure and applications vendors: Oracle, IBM, EMC, Symantec, Hitachi and more
- **2400 clients worldwide; c. 50% of Fortune 500**
- Named a Rising Star Company on Deloitte's 2006 Technology Fast 500.
  - Among 25 of the fastest growing technology, media, telecommunications and life sciences companies in North America



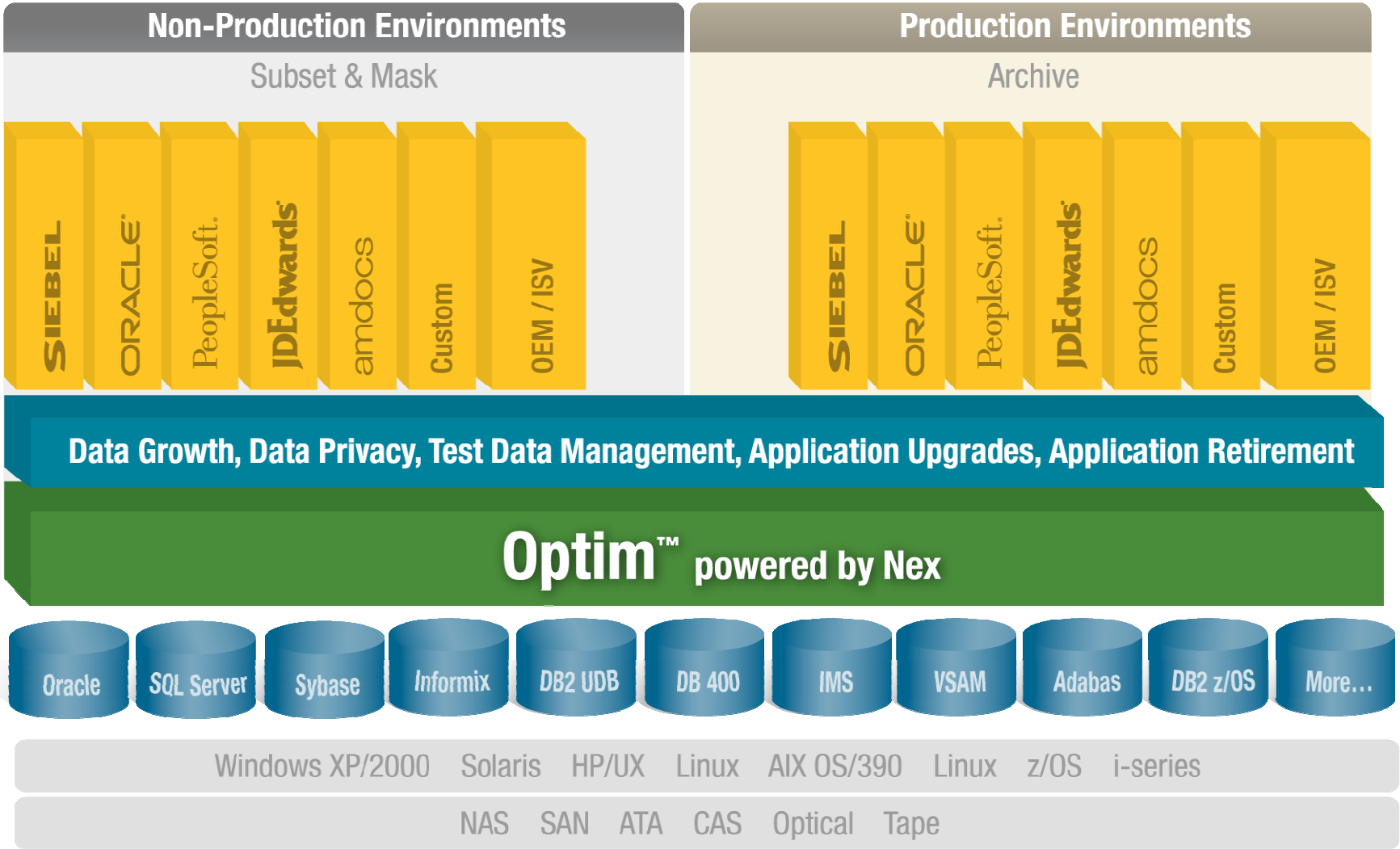
Source: Gartner, "Archiving Software Market to Experience Strong Growth Through 2010," 2006

Optim Solves the EDM Challenge



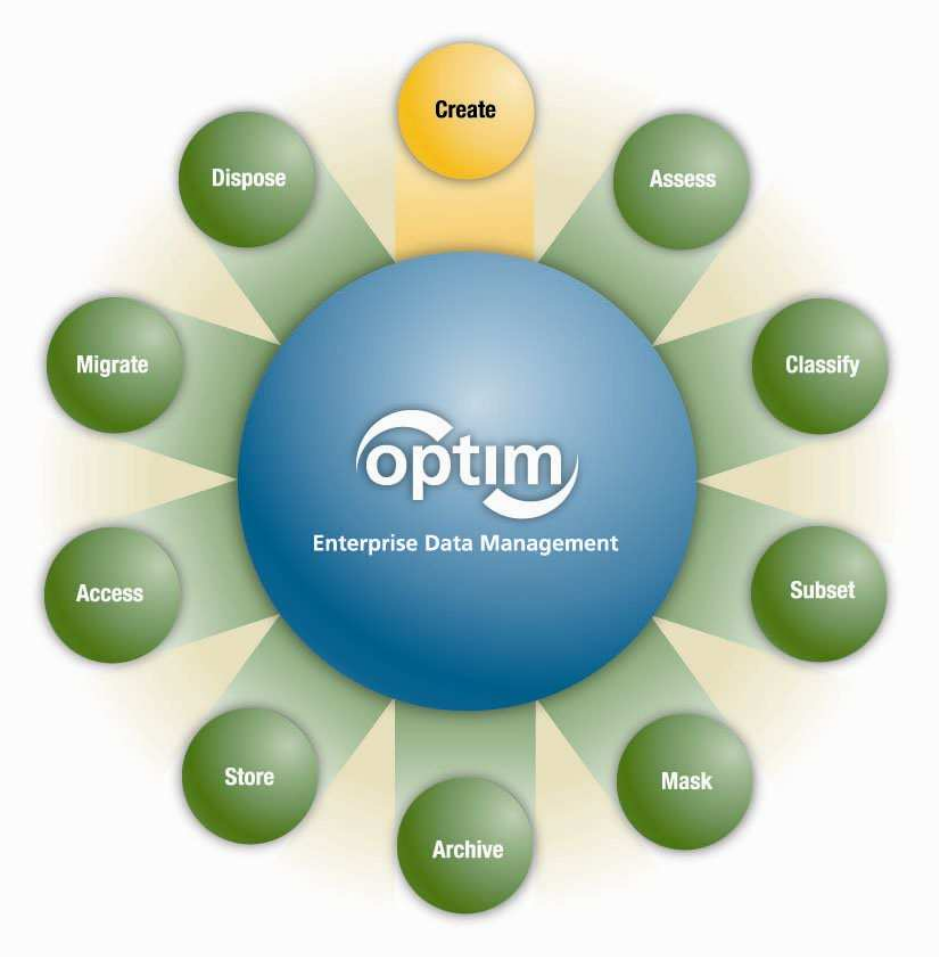
- **Test Data Management**
  - Create targeted, right sized test environments
  - Improve application quality
  - Speed iterative testing processes
- **Data Privacy**
  - Mask confidential data
  - Comply with privacy policies
- **Archiving**
  - Improve performance
  - Control data growth, save storage
  - Support retention compliance
  - Enable application retirement
  - Streamline upgrades

Optim Enterprise Architecture



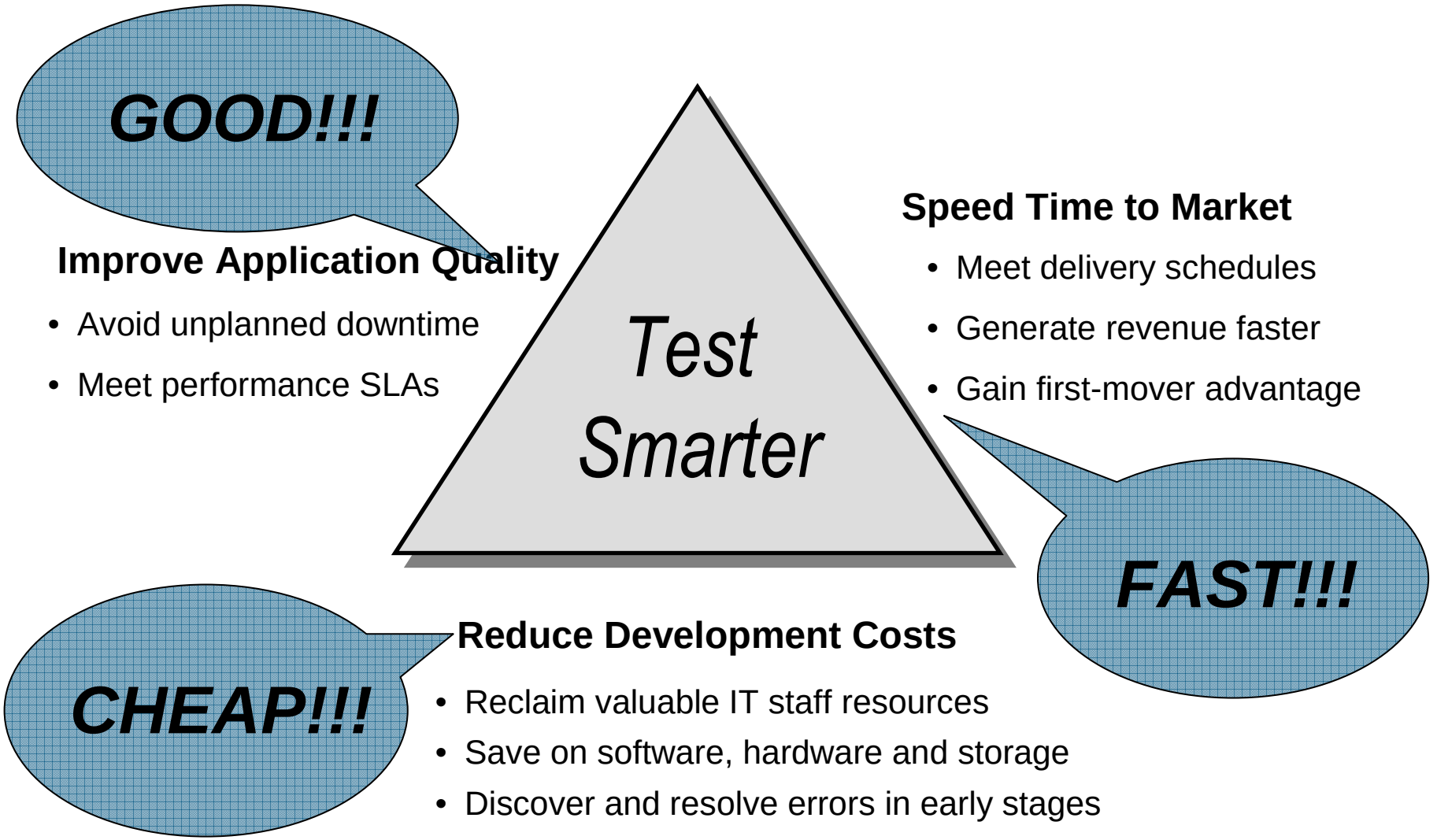
- Single, scalable, interoperable EDM solution provides a central point to deploy policies to extract, store, port, and protect application data records from creation to deletion

Optim Solves the EDM Challenge



- **Test Data Management**
  - Create targeted, right sized test environments
  - Improve application quality
  - Speed iterative testing processes
- **Data Privacy**
  - Mask confidential data
  - Comply with privacy policies
- **Archiving**
  - Improve performance
  - Control data growth, save storage
  - Support retention compliance
  - Enable application retirement
  - Streamline upgrades

Software development and testing…



Current Practice?

#1 - Clone Production

Repeat ?*%$!

Clone Production

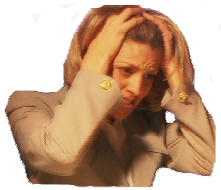
Request for Copy

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

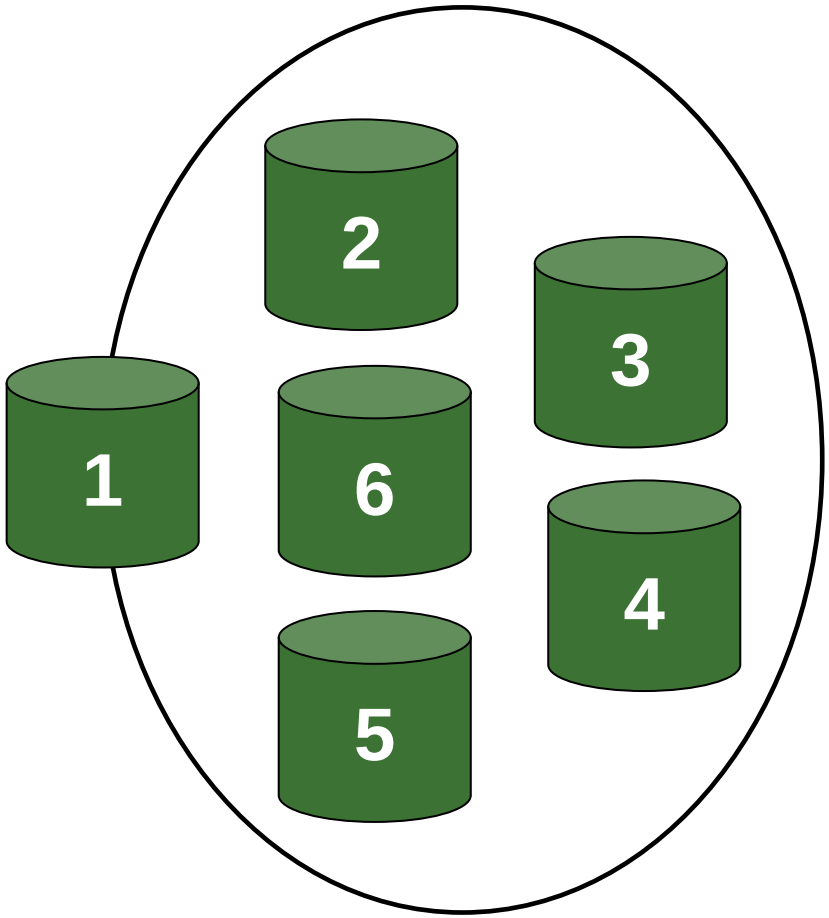
- RI Accuracy?
- Right Data?

Expensive, Dedicated Staff, Ongoing Responsibility.

Share test database with everyone else

Managing Application Data Growth

The “Data Multiplier Effect”

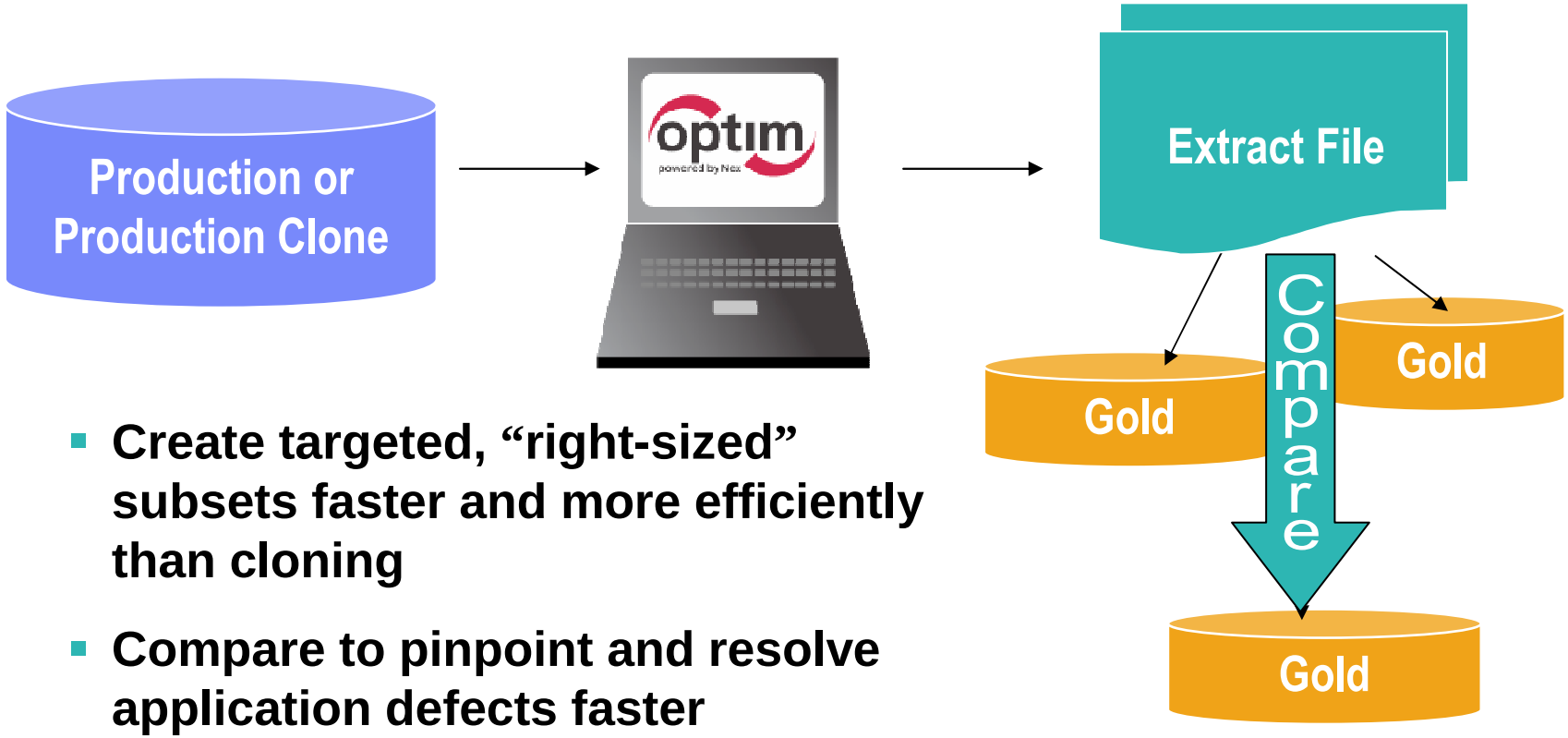


Production	200 GB
Training	200
Unit	200
System	200
UAT	200
Integration	200
Total Size	1200 GB

Test Data Management Projects

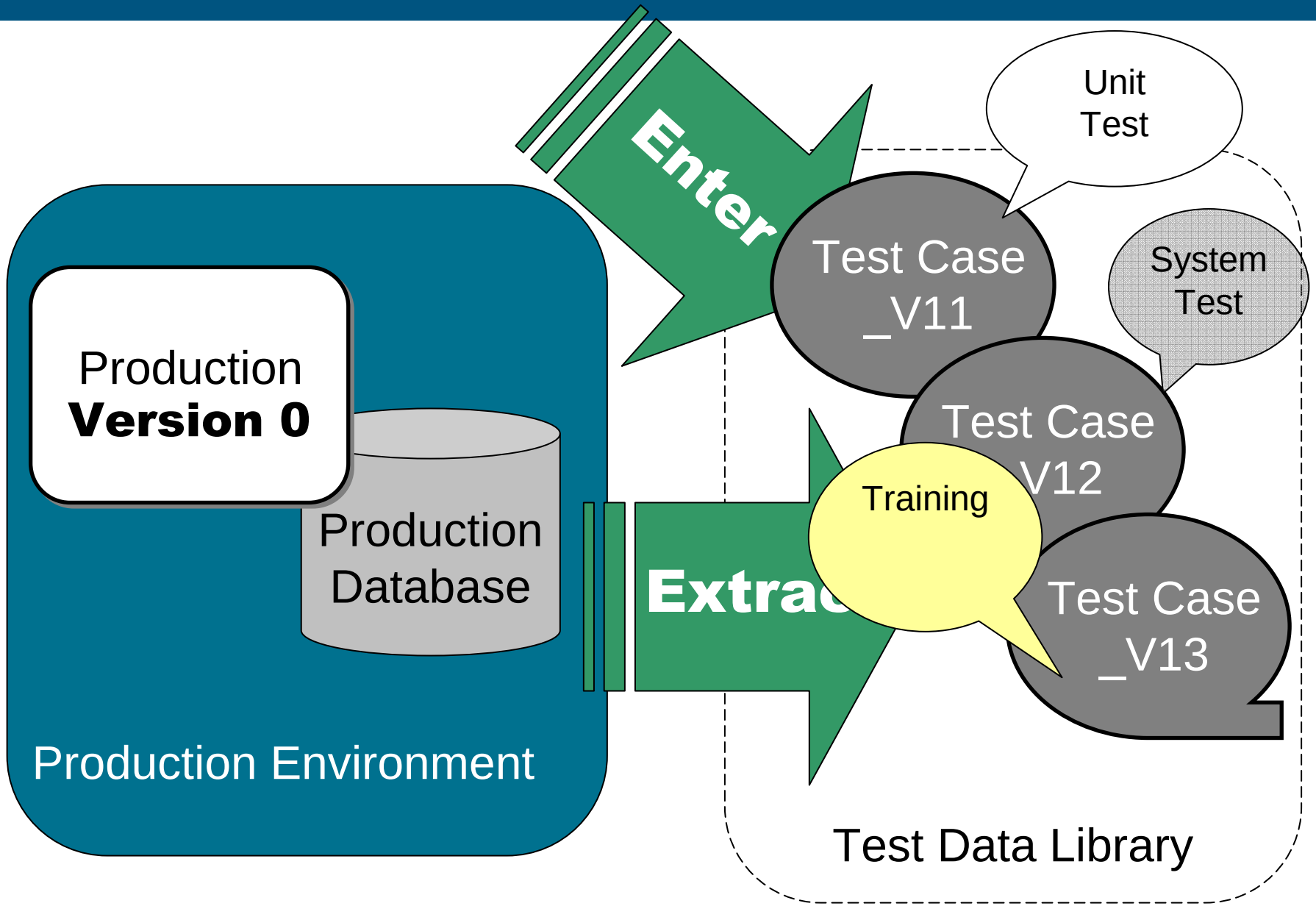
- Characteristics for Test Data Management Projects
  - Subset capabilities to create realistic and manageable test databases
  - Quickly refresh test environments
  - Edit data to create targeted test cases
  - Compare ‘before’ and ‘after’ images of the test data
  - Improve test coverage and quality
  - De-identify (mask) data to protect privacy

Optim Test Data Management Solution

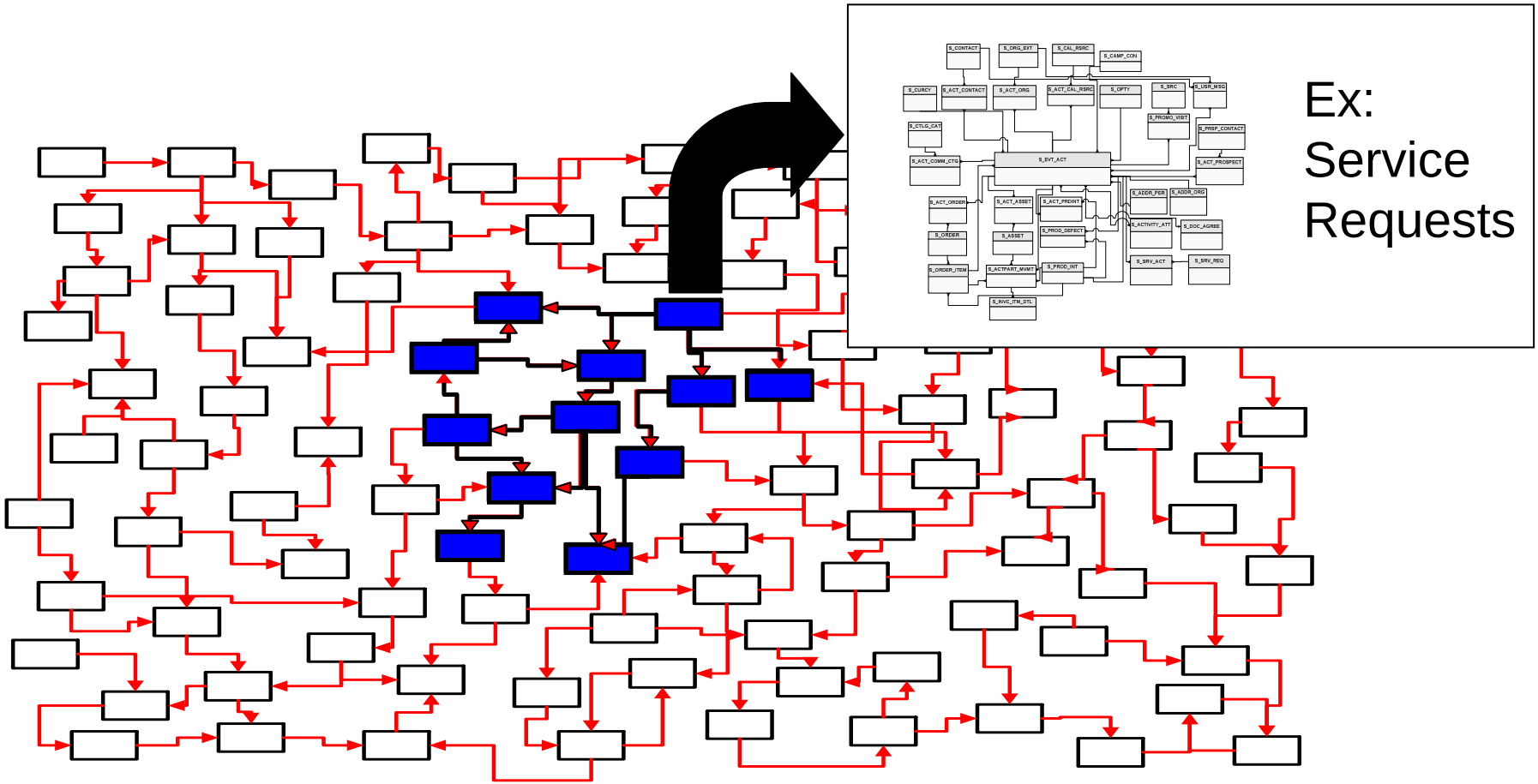


- **Create targeted, “right-sized” subsets faster and more efficiently than cloning**
- **Compare to pinpoint and resolve application defects faster**
- **Improve development efficiencies**

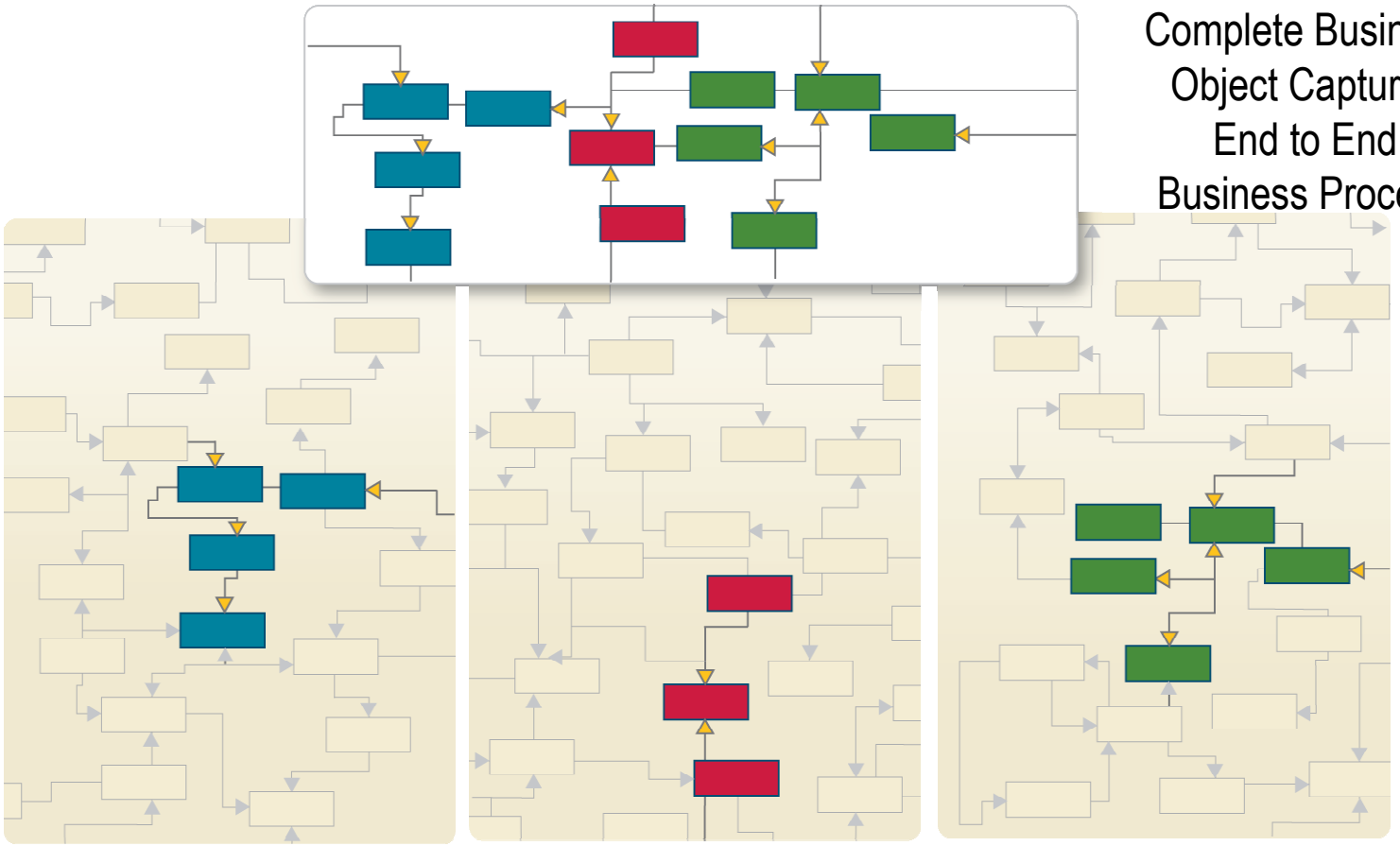
A Test Data Library



What's in a Test Case?



Federated Data Support



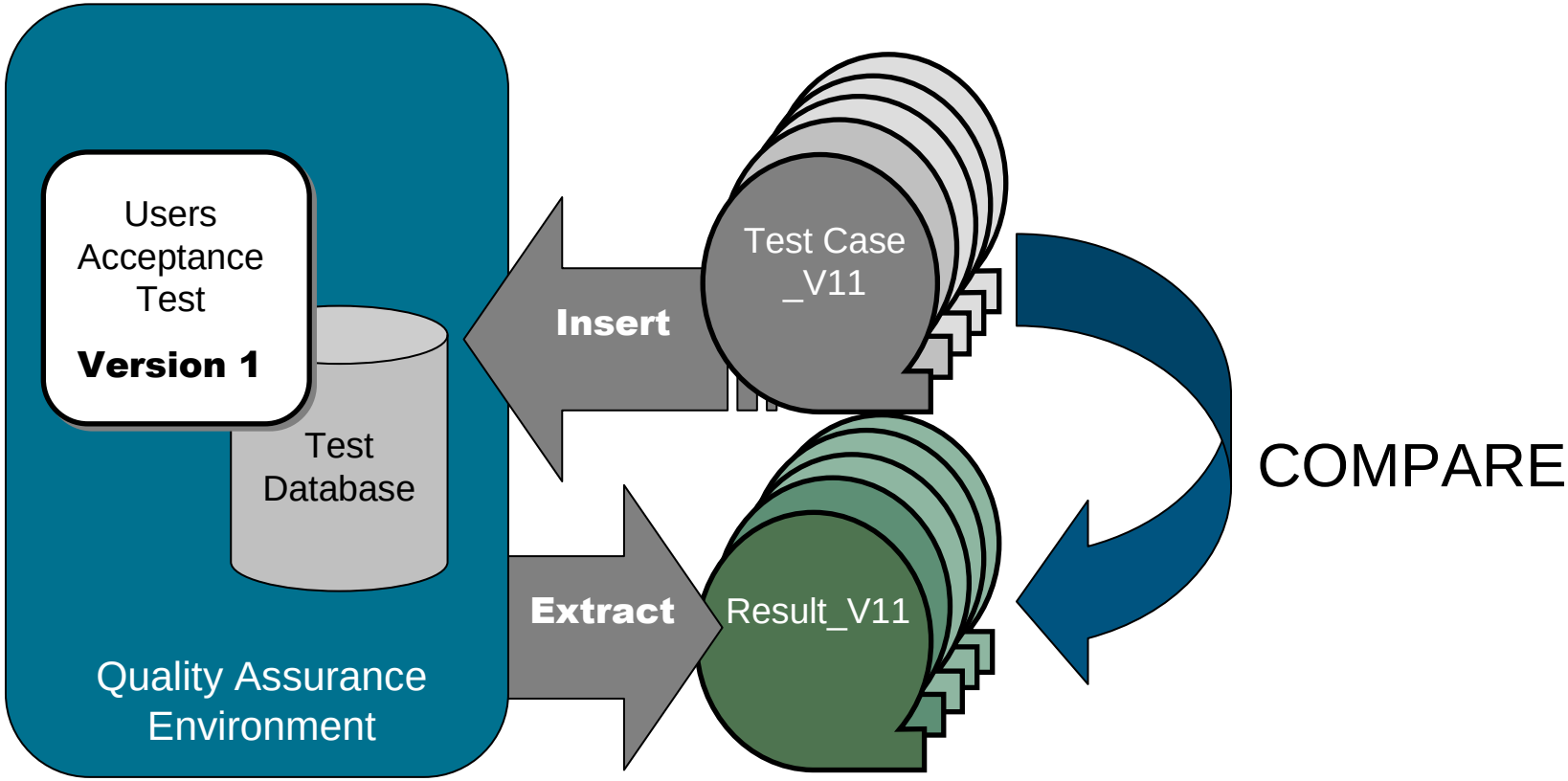
Complete Business Object Captures End to End Business Process

Retek / Oracle

Custom Inventory Management / DB2

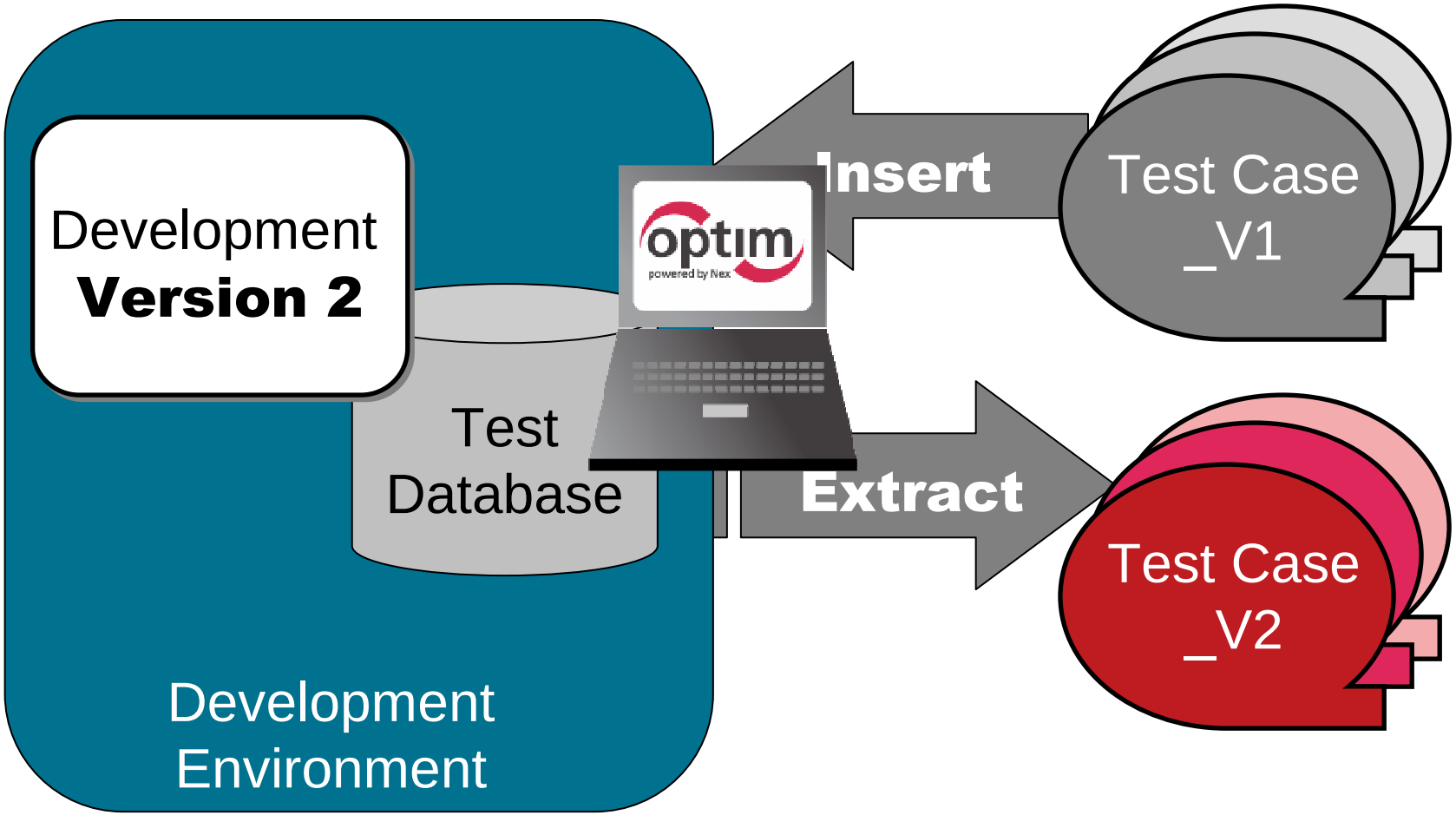
Other apps / any DBMS

Tracking the Results

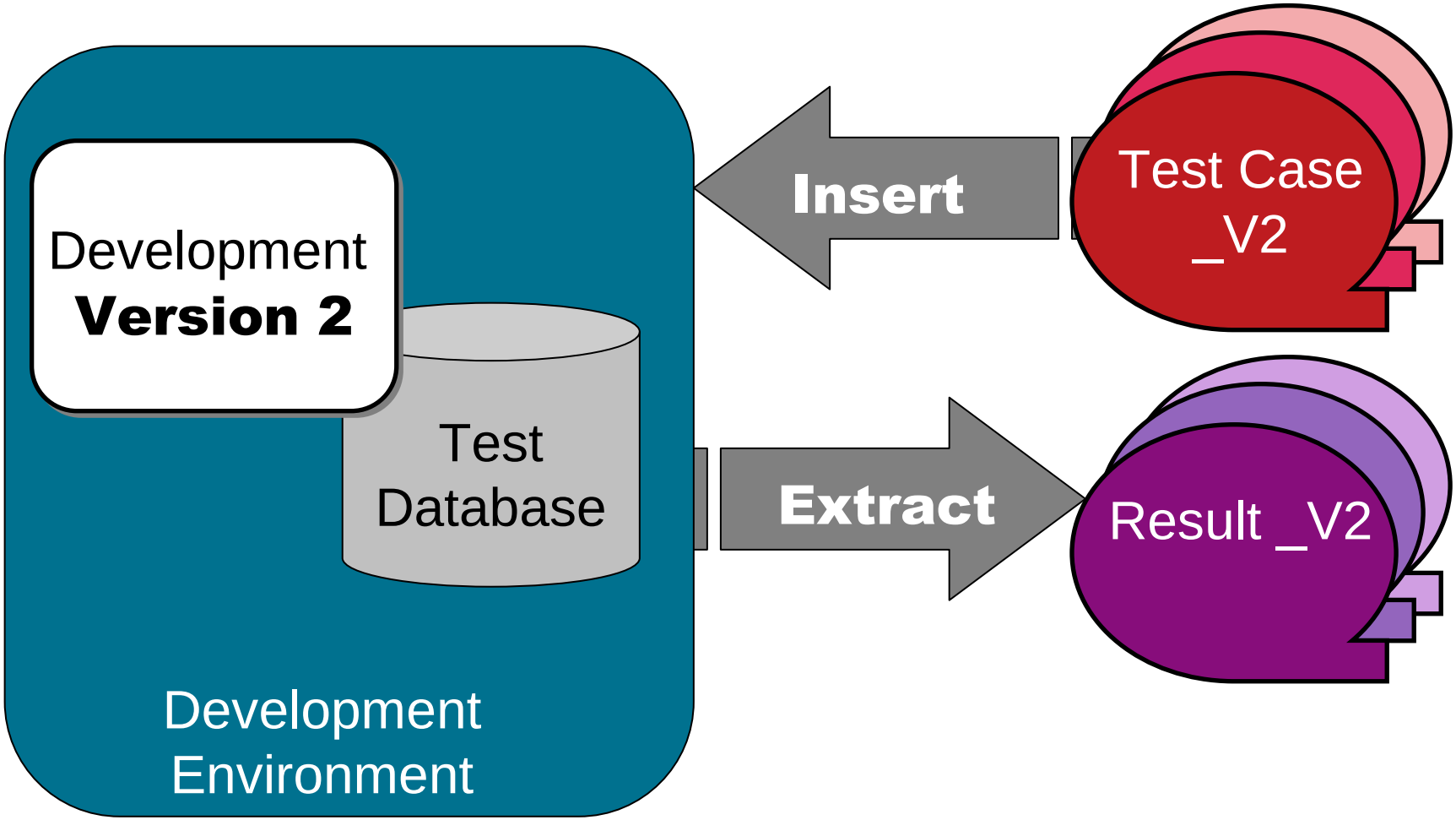


- Take snapshots of process results for later comparison
- Reuse Test Cases from “Test Case Library”

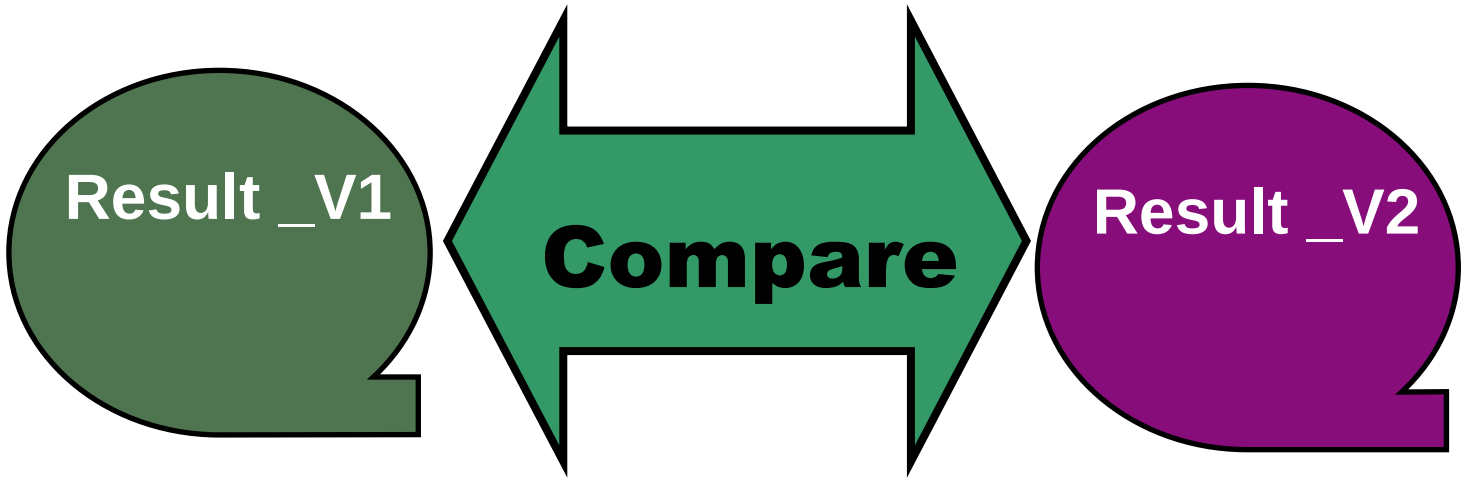
Version 2: Unit Test



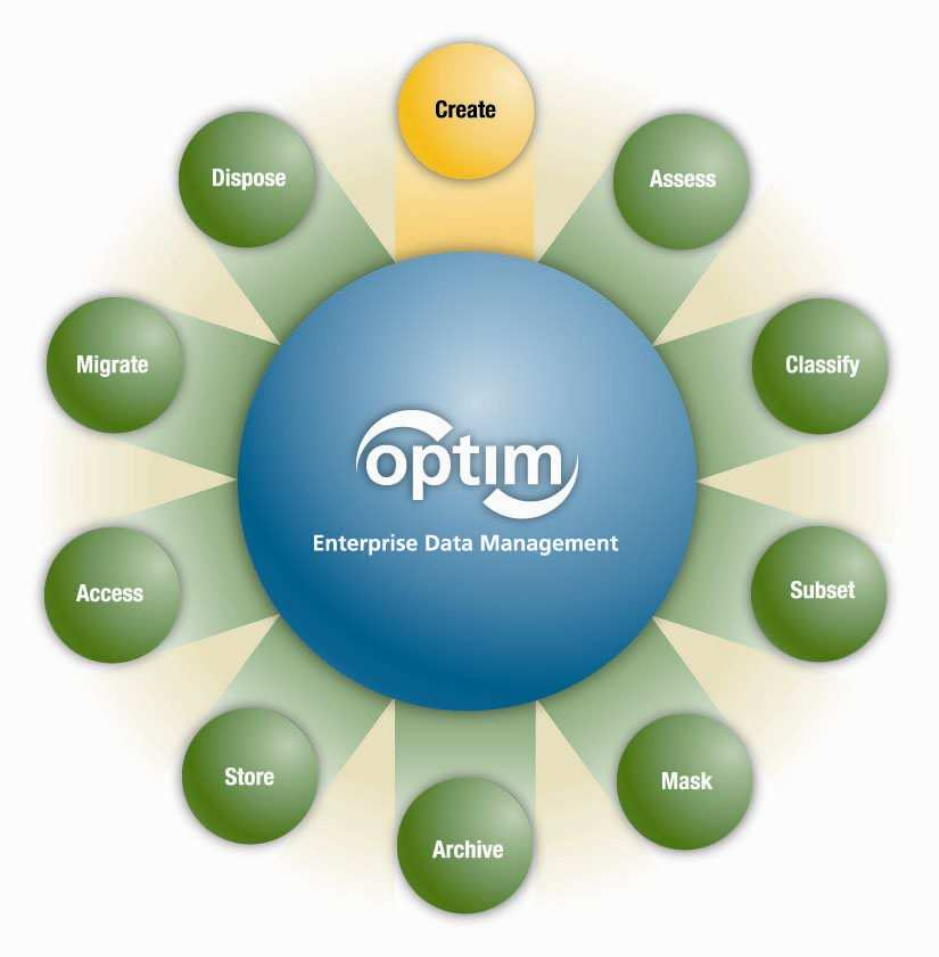
Version 2: Regression Test



Version 2: Regression Test of Process “A”



Optim Solves the EDM Challenge



- **Test Data Management**
  - Create targeted, right sized test environments
  - Improve application quality
  - Speed iterative testing processes
- **Data Privacy**
  - Mask confidential data
  - Comply with privacy policies
- **Archiving**
  - Improve performance
  - Control data growth, save storage
  - Support retention compliance
  - Enable application retirement
  - Streamline upgrades

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?
- Payment Card Data Security Industry Reg. 6.3.4 states, **"Production data (real credit card numbers) cannot be used for testing or development"**
  - HIPAA, GLBA, PIPED, DDP, NPP, others

The Solution is Data De-Identification

In the News….

2007… largest off-price apparel retailer in the United States…

- TJ Maxx hack exposes consumer data
- 45.7 million accounts
- Data Breach Will Cost TJX $1.7B
- Payment Card Industry Data Security Standard

http://blog.wired.com/27bstroke6/2007/03/data_breach_wil.html

What is Data Masking?

- AKA depersonalization, desensitization, or 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

Component A - Consistency

- Masking is a repeatable process
- Subsystems need to match originating
- The same mask needs to be applied across the enterprise
  - Predictable changes
  - Random change will not work
- Change all ‘Jane’ to ‘Mary’ again and again

Example: First and Last Name

- Direct Response Marketing, Inc. is testing its order fulfillment system
- To fictionalize customer names, use the a random lookup function to pull first and last names randomly from the Customer Information table:
  - "Gerard Depardieu" becomes "Ronald Smith"
  - "Lucille Ball" becomes "Elena Wu"

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 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

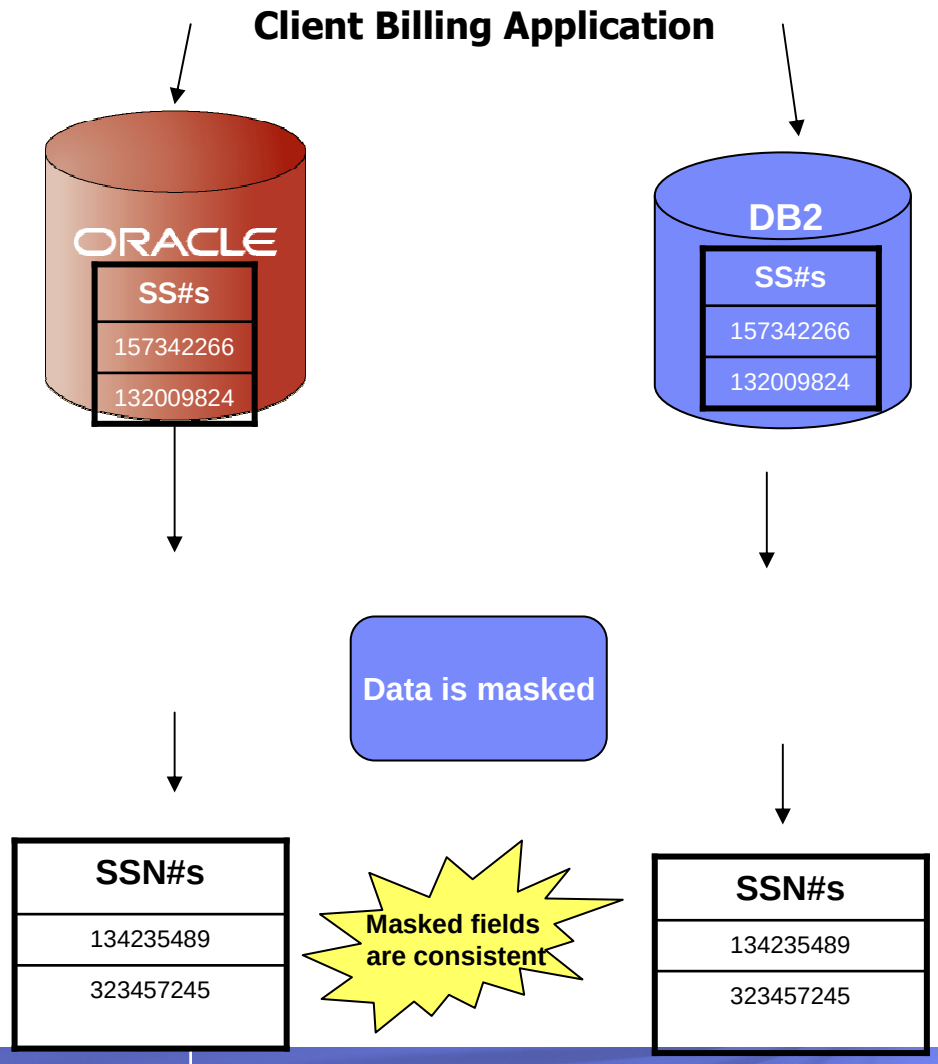
Customers Table

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

Component B - Context

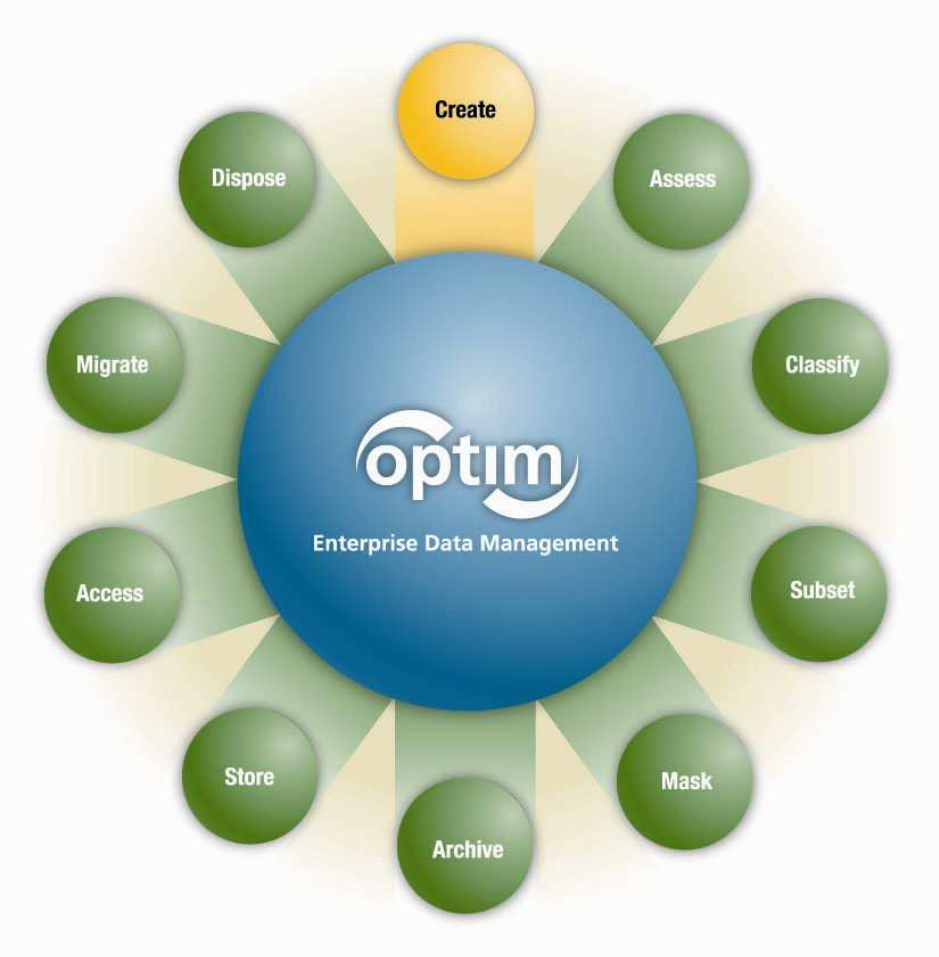


- A single mask will affect 'downstream' systems
- Column/field values must still pass edits
  - SSN
  - Phone numbers
  - E-mail ID
- Zip code must match
  - Address
  - Phone area code
- Age must match birth date

Component C - Flexibility

- Laws being interpreted
- New regulations being considered
- Change is the only certainty
- ERPs being merged
- Masking routines will change, frequently
- Quick changes will be needed

Optim Solves the EDM Challenge



- **Test Data Management**
  - Create targeted, right sized test environments
  - Improve application quality
  - Speed iterative testing processes
- **Data Privacy**
  - Mask confidential data
  - Comply with privacy policies
- **Archiving**
  - Improve performance
  - Control data growth, save storage
  - Support retention compliance
  - Enable application retirement
  - Streamline upgrades

What are the Key Drivers of Data Growth?

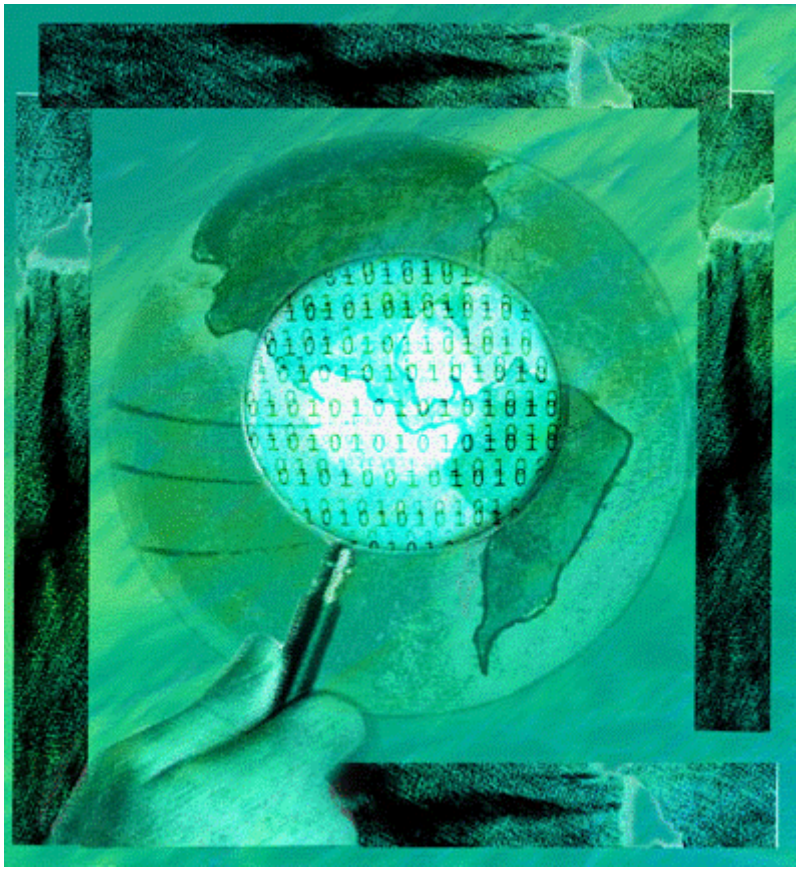
- **Mergers & acquisitions**
- **Organic business growth**
  - eCommerce
  - ERP/CRM
- **The digital revolution**
- **Records retention**
  - Basel II
  - SOX
  - Euro-SOX
- **Data multiplier effect**

- **Forrester estimates that 85% of data stored in databases is inactive**

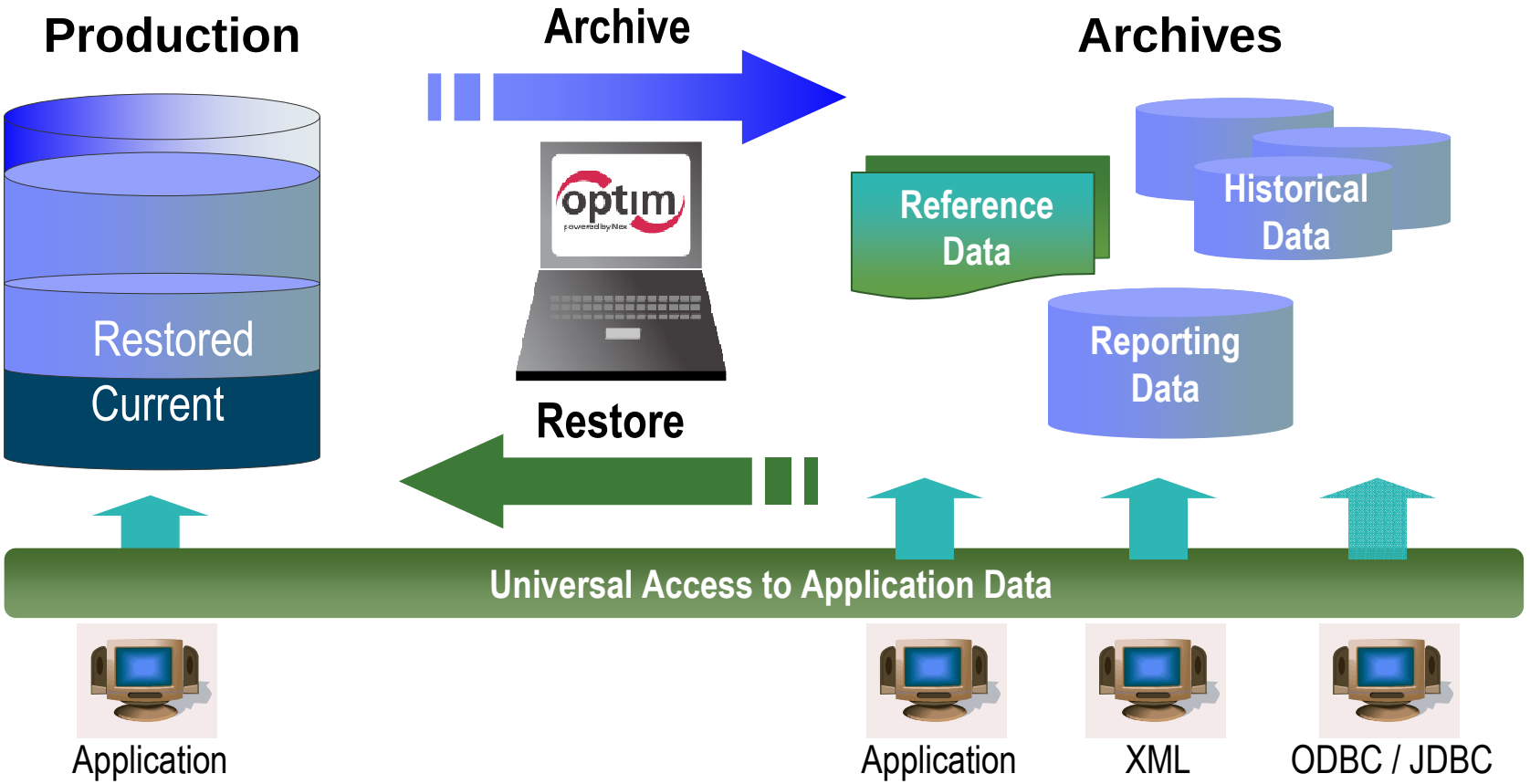
* Source: Noel Yuhanna, Forrester Research, Database Archiving Remains An Important Part Of Enterprise DBMS Strategy, 8/13/07

The Symptoms

- **Applications perform slowly**
  - SLAs are being missed
  - Customer satisfaction declining
- **Backups seems to take forever**
- **Batch jobs run into working hours**
- **Legal costs are soaring**
- **“Every time I turn around we are buying more storage”**

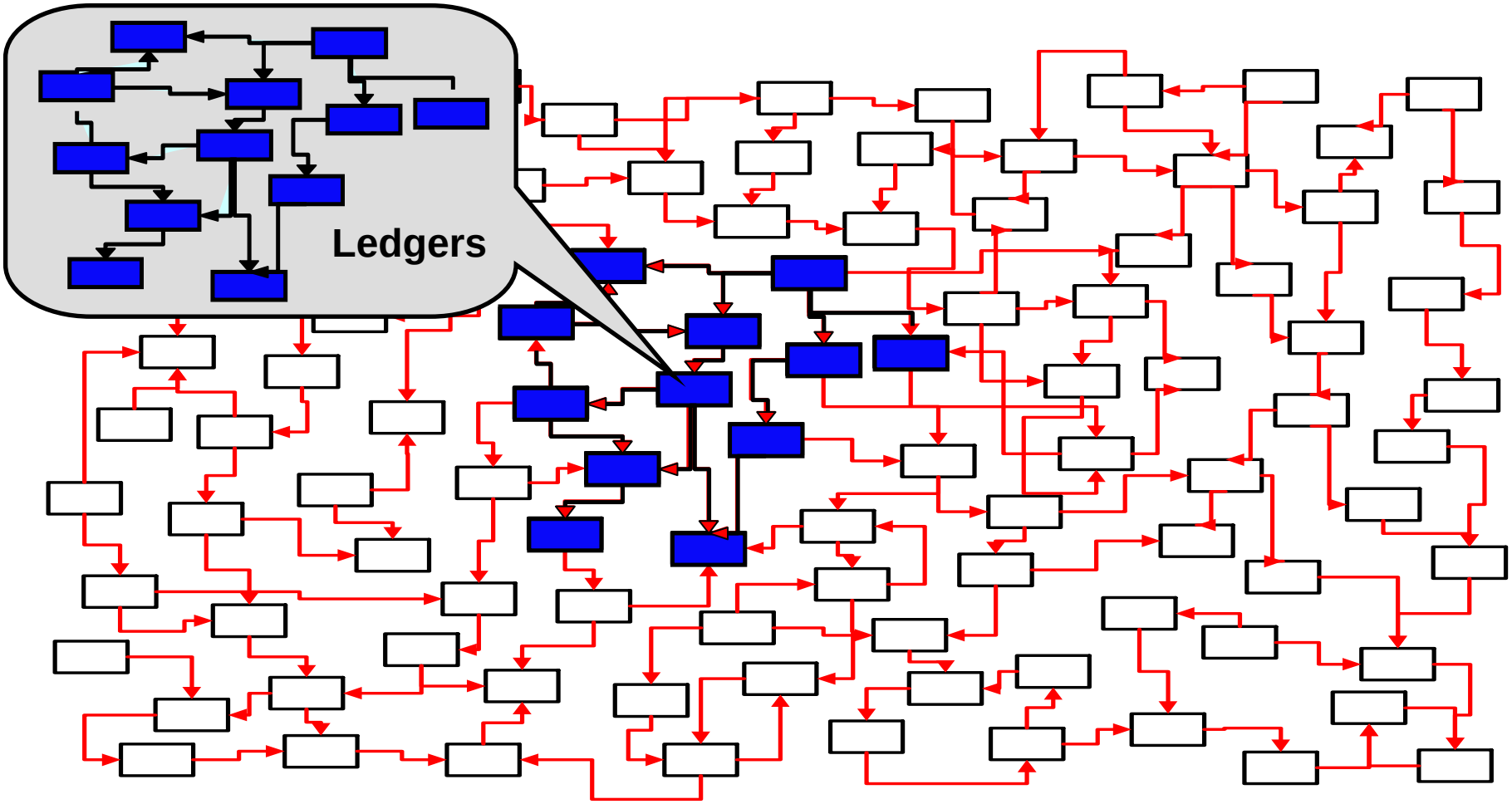


Optim™ Data Growth Solution: Archiving

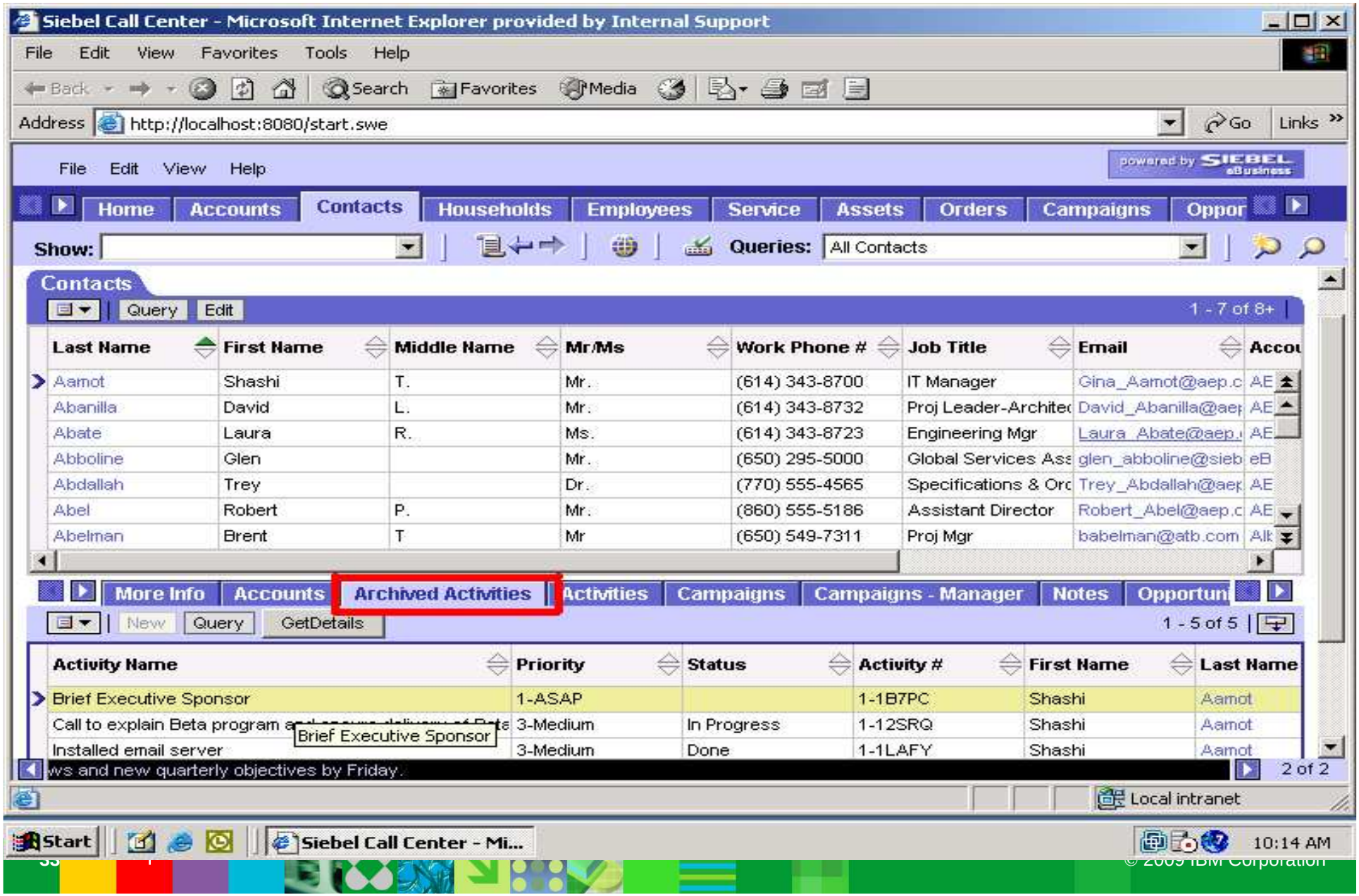


- **Complete Business Object provides historical reference snapshot of business activity**
- **Storage device independence enables ILM**
- **Immutable file format enables data retention compliance**

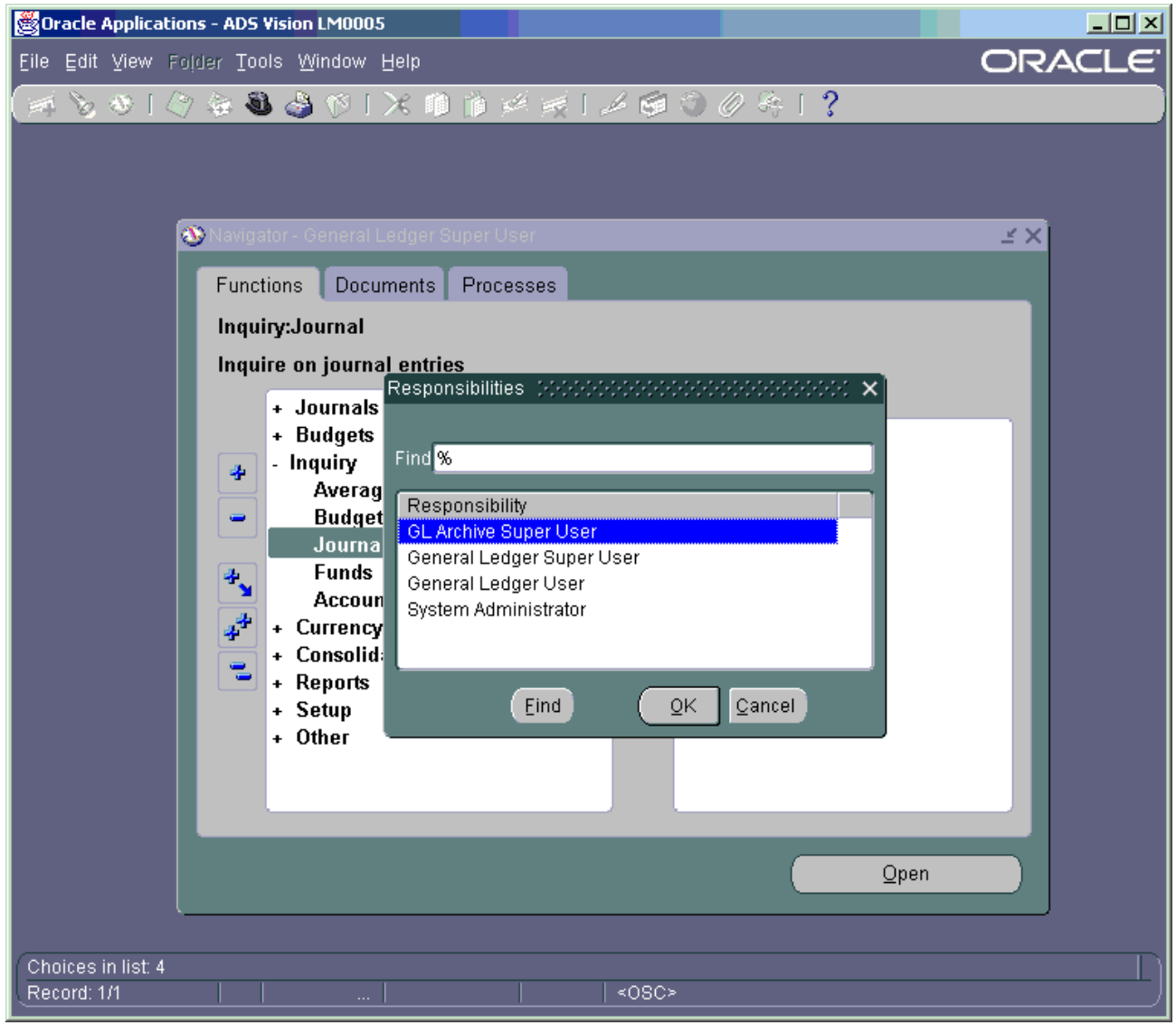
Archiving a Complete Business Object



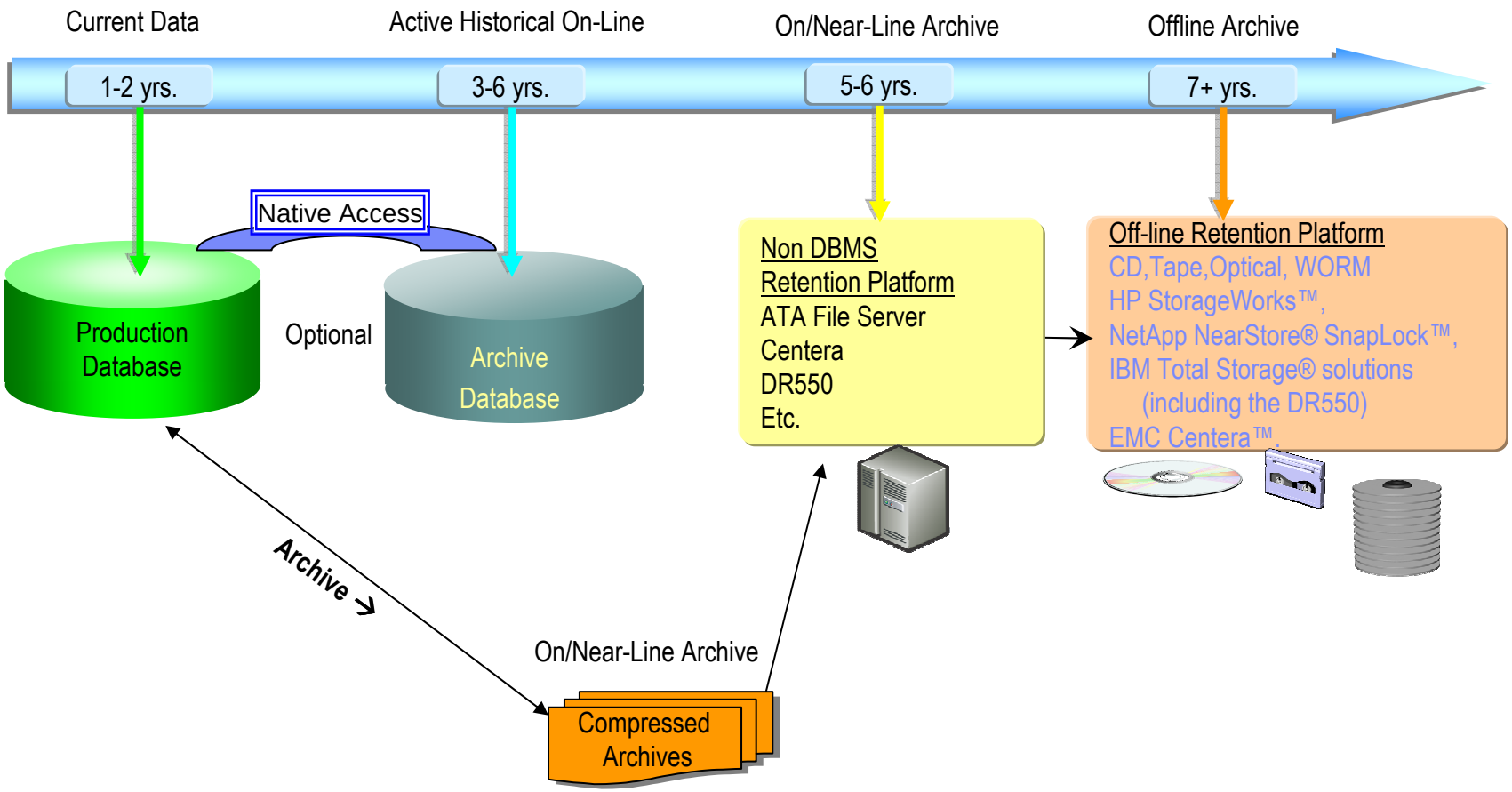
Optim Integrated Within Siebel



Optim Switch to GL Archive User



Store - Data Retention Strategies



What Benefits Will You See from Data Archiving?

1) **Improved Performance, Shorter Outages**

   Faster applications, Less outage = More Revenue

2) **Reduction of Costs**

   Less Data in Production Environments
   = Significant Savings

3) **Mitigation of Risks**

   Data Retention policies and procedures
   = Compliance with audit and e-discovery requests

Questions