

# **IBM Information Protection Capabilities on z/OS**

# Ernie Mancill Executive IT Specialist mancill@us.ibm.com





# Agenda

- The need to protect data
- Information Protection entry point to Information Governance
- IBM's Information Protection capabilities
- Summary



# Organizations facing many of the following challenges

- Discovering what data needs to be secured
- How to secure your data
- Audit and separation of roles privileged user conundrum
- Encryption and data obfuscation
- Data in a test environment
- Data life-cycle management and data growth

IBM Information Management Solutions for System z – End to end Solution



# Security, Audit, and Encryption for z/OS





# Satisfy Your Auditor: Plan, Protect and Audit

#### Data Access

- Minimize the use of a superuser authorities such as SYSADM
- A different group should manage access to restricted data than the owner of the data

## Data Auditing

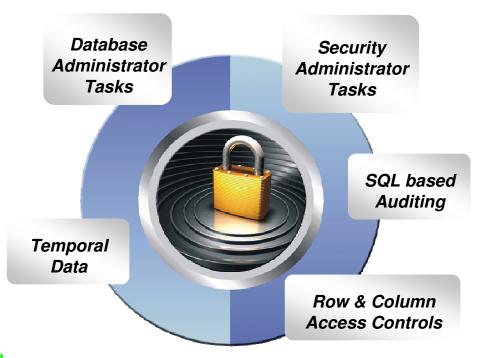
 Any dynamic access or use of a privileged authority needs to be included in your audit trail

Session A05

 Maintain historical versions of data for years or during a business period

## Data Privacy

 All dynamic access to tables containing restricted data needs to be protected



Today's Mainframe:
The power of industry-leading security,
the simplicity of centralised management



# DB2 10 for z/OS Security Enhancements

Help Satisfy Your Auditors using new features

- New granular authorities to reduce data exposure for administrators
- New auditing features using new audit policies comply with new laws
- New row and column access table controls to safe guard your data
- New temporal data to comply with regulations to maintain historical data





# Reduce your risk by minimizing use of SYSADM

## New granular system authorities and install security parameters

New in DB2 10





## Prevents SYSADM and SYSCTRL from granting or revoking privileges

**EXPLAIN** 

- New separate security install zparm parameter
- New install SECADM authority manages subsystem security
- SYSADM and SYSCTRL can no longer implicitly grant or revoke privileges

## **Control cascading effect of revokes**

- New revoke dependent privileges install parameter
- New revoke dependent privileges SQL clause



# New authority for performing security tasks without ability to change or access data

- SECADM authority
  - Allows the user to
    - Issue SQL GRANT, REVOKE statements on all grantable privileges and administrative authorities
    - Manage DB2 9 roles and trusted contexts
    - Manage DB2 10 row permissions and column masks
    - Manage DB2 10 Audit policies
    - Access catalog tables
    - Issue START, STOP, and DISPLAY TRACE commands





# New authority for managing objects without ability to access data or control access to data

- System DBADM authority
  - Allows the user to
    - Issue SQL CREATE, ALTER, DROP statements to manage most objects in the DB2 subsystem

- Issue most DB2 commands
- Execute system defined stored procedures and functions
- · Access catalog tables





## RACF and Data Servers on z/OS

- RACF and DB2
  - DB2 Subsystem Access Control (outside of DB2)
  - Control connections to the DB2 subsystem
    - CICS
    - IMS
    - CAF
    - BATCH
  - Assign identities
  - Protect the underlying DB2 data store (underlying data sets of DB2 can be protected by RACF dataset services)
  - In addition to database server-provided security, RACF can be used to control access to database objects, authorities, commands and utilities by using the RACF access control module of the database server.
- RACF and IMS
  - The IBM Information Management System (IMS™) has been enhanced to make use of RACF for controlling access to IMS resources. It is possible to use the original IMS security features, the new RACF features, and combinations of these. RACF provides more flexibility than the older security features. The normal features of RACF can be used to protect both system and database IMS data sets



## Tools from Tivoli to enhance RACF

- Tivoli zSecure Admin
  - -User friendly layer over the native RACF administration panels
  - Automatically generated RACF commands
    - Reduce complexity
    - Increased RACF administration productivity
    - Fewer errors
    - Less risk of inadvertent data exposure due to inappropriate/insufficient security
- Tivoli zSecure Visual
  - -GUI/Windows based UI
  - -Insulates security administrators from TSO/ISPF
  - Increased productivity requiring less sophistication in administration skills
- Tivoli Identity Management software
  - -Tivoli Directory Server
  - -Tivoli Identity Manager



# End User Identity Mapping - Why is this Important?

- Proper end user assignment of rights and privileges on the data server is important, but equally important:.
  - In many mult-tier implementations, to ease administration, and to influence performance through mechanisms such as connection pooling, thread reuse, etc. Shared (common) authorization IDs are used for connecting to the Data Server
  - In these types of implementations, this leads to loss of end user identification, and any associated ability to completely audit activity on the data server from these types of connections.
  - Various mechanisms can be use to preserve these credentials:
    - SQL Language Extension vis SQLESETI
    - Extended identity propagation using JDBC drivers
    - Enterprise Identity Mapping
- Support distributed identities introduced in z/OS V1R11
  - A distributed identity is a mapping between a RACF user ID and one or more distributed user identities, as they are known to application servers



#### The DB2 client information fields

DB2 allows applications to send information about them to the database with each SQL operation.

- The database externalizes this information then in its monitoring data
- The performance impact of setting them is negligible (but for DB2 on LUW V9.1 FP6 is recommended)
- The data can be set by the application itself, or via database driver properties (see next slides)
- The following information can be set:

Field	Description	Length (LUW, z/OS)
Client user ID	This user ID is for identification purposes only, and is not used for any authorization. It typically identifies the user of an application.	255, 16
Client workstation name	The workstation name of the client system. Some applications also use this field to identify the business transaction executed within an application.	255, 18
Client application name	It can be used to identify the application hosted in an application server, or to identify the business transaction within an application.	255, 32
Program name	Identifies the application running on the client. It is only supported for a connected DB2 on z/OS database.	-, 80
Accounting string	It can be used to specify charge-back information, or to add additional monitoring details about the database workload.	200, 200



## Ways to instrument your application

JDBC offers methods of class com.ibm.db2.jcc.DB2BaseDataSource^1

```
public static void main(String[] args) {
   String url = "jdbc:db2://lap1.boeblingen.de.ibm.com:50000/DEMO";
   Class.forName("com.ibm.db2.jcc.DB2Driver");
   Connection conn = DriverManager.getConnection(url, user, password);
   conn.setClientInfo("ClientUser", "xyz");
   conn.setClientInfo("ClientHostname, "my laptop");

   conn.prepareStatement("SELECT * FROM SYSIBM.SYSDUMMY1" + "WHERE 0 = 1").executeQuery();
}
CLI offers the setsqli() ^2 interface
```

```
SQL_API_RC SQL_API_FN sqleseti (
    unsigned short DbAliasLen,
    char * pDbAlias,
    unsigned short NumItems,
    struct sqle_client_info* pClient_Info,
    struct sqlca * pSqlca);

SQL_STRUCTURE sqle_client_info {
    unsigned short type;
    unsigned short length;
    char *pValue; };
```

#### 1) see

http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/index.jsp?topic=/com.ibm.db2.luw.apdv.java.doc/doc/r0021822.html

#### 2) see

http://publib.boulder.ibm.com/infocenter/db2luw/v9r5/index.jsp?topic=/com.ibm.db2.luw.apdv.api.doc/doc/r0001709.html



# Why auditing is important in a RACF controlled environment

- RACF provides significant controls to protect access to resources, but does little in the way of meaningful access reporting
  - RACF does two things:
    - Prevents people from accessing a resource that is not essential or appropriate for their jobs
    - Allows people access to the necessary data to do their jobs
  - But RACF does NOT:
    - prevent a malicious update if the user has authority to the data.
    - prevent an authorized user from accessing sensitive data that is NOT within the scope of their job.E.g. a bank teller looks up the CEOs bank balance or personal customer information
    - provide meaningful information about access to protected DB2 resources (authorized or not).
- DB2 Audit trace will do nothing to protect data, but provides data to help understand what type of access has occurred.
  - Auditing is about ensuring that the appropriate controls are in place to identify inappropriate access and use of production data
  - You need some form of audit facility to watch your privileged users who have RACF and/or DB2 authority and users that have access to sensitive data within the scope of their job
  - Understanding how trusted (privileged) users access sensitive information is essential to ensuring that data is indeed protected

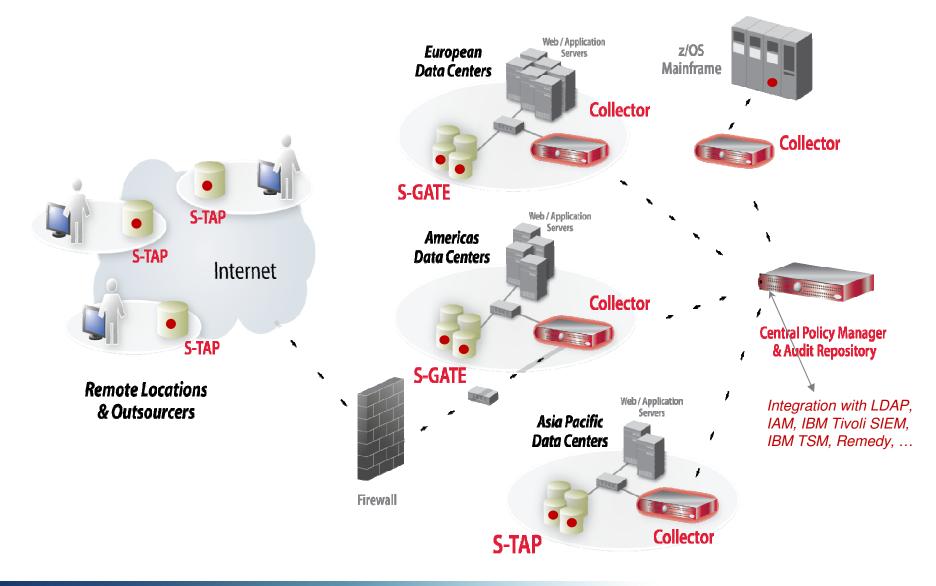


# Auditing the Privileged Database User

- DB2 trace based processes are managed by DBA's
  - -The DBA's are responsible for generating audit data with which they are in turn audited, this constitutes a significant security risk and exposure
  - -Trace data collection can be interfered with or turned off completely
    - DBA can issue –DSN Stop Trace
    - Use IFASMFDMP to selectively filter SMF data based on timestamp
    - Use DB2PM (Or Equivalent) filter such as DATE/TIME/EXCLUDE to filter selected records
  - Having the DBA involved in the collection of audit data is viewed as weak from a compliance and control perspective
- Security and Auditors with system privileges
  - Also viewed as problematic from a compliance perspective
  - Requires additional technical skills not within their core competencies
  - Misuse of privileges without coordination can result in performance and availability issues
    - Turning on traces without proper filtering to reduce overhead or quantity of trace data collected
    - Altering objects to AUDIT without ensuring that plan/package invalidation is not an issue



## Scalable Multi-Tier Architecture



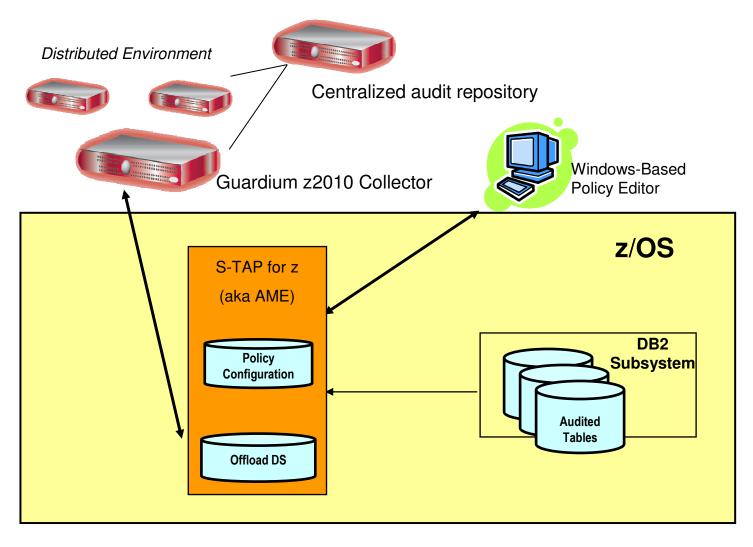


# Guardium for z - Components

- S-TAP for System z (Aka AME)
  - Mainframe probe
  - Collects audit data for Collector appliance
  - Leverages existing IBM DB2/z collection technology
  - When used in the Guardium context, we're NOT calling it AME, we're calling it S-TAP for z. But, we're STILL ordering AME (5655-T67)
  - NOT z-TAP, that's another solution
- Guardium z2010 Collector appliance
  - Securely stores audit data collected by mainframe probe
  - Provides analytics, reporting & compliance workflow automation
    - Offloads audit data processing from mainframe
  - Integrated with Guardium enterprise architecture
    - Centralized, cross-platform audit repository for enterprise-wide analytics and compliance reporting across mainframe & distributed environments
      - (Oracle, SQL Server, DB2, Informix, Sybase, MySQL, Teradata)



## Guardium for z – Architecture





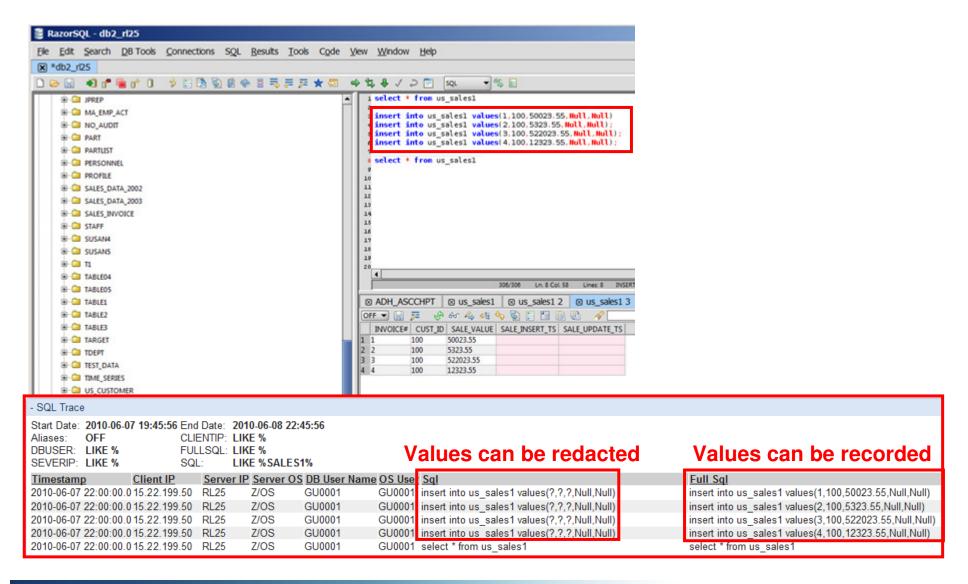
# Database Activity Monitoring using Guardium

- Guardium ensures separation of roles and audit data integrity
  - No reliance on native DBMS trace or logging facilities for audit data collection
  - Audit repository placed on hardened Linux appliance
    - No Root Access
    - No "native" database access allowed
  - Control of collection and audit policies by Security administrators with no DBA involvement and without any DBMS or server privileges needed





## Reports - Inserts...





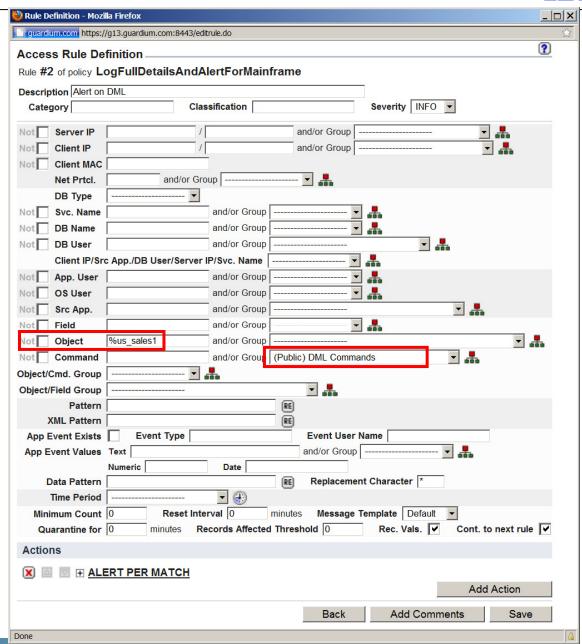
# Sample Report

- SQL Trace						
Start Date: 2010-06-0 Aliases: OFF DBUSER: LIKE % SEVERIP: LIKE %	CI	nd Date: 2010- UENTIP: LIKE ULLSQL: LIKE QL: LIKE	%	:45		
Timestamp	Client IP	Server IP	Server O	S DB User Name	OS User	Sql
2010-06-08 03:11:24.0	15.22.19.50	RL25	Z/OS	GU0002	GU0002	REVOKE EXECUTE ON PROCEDURE SYSIBM.SQLTABLEPRIVILEGES FROM PUBLIC BY ALL RESTRIC
2010-06-07 22:12:28.0	015.22.19.50	RL25	Z/OS	GU0001	GU0001	INSERT INTO udt_table VALUES(CAST(? AS udt1), CAST(? AS udt2), CAST(? AS udt3))
2010-06-08 03:04:29.0	015.22.19.50	RL25	Z/OS	GU0001	GU0001	INSERT INTO udt_table VALUES(CAST(? AS udt1), CAST(? AS udt2), CAST(? AS udt3))
2010-06-07 22:14:09.0	15.22.19.50	RL25	Z/OS	GU0001	GU0001	delete from camp_roster where NAME like ?
2010-06-08 03:12:13.0	15.22.19.50	RL25	Z/OS	GU0002	GU0002	GRANT CREATEIN, ALTERIN, DROPIN ON SCHEMA va_test_schema TO QA_TEST
2010-06-08 03:11:10.0		RL25	Z/OS	GU0002	GU0002	REVOKE EXECUTE ON PACKAGE NULLID. SYSSN101 FROM PUBLIC BY ALL
2010-06-08 02:29:05.0	15.22.19.50	RL25	Z/OS	GU0002	GU0002	GRANT ALL ON TABLE VA TEST.EMP TO VA TEST



## Alerts

- Processed audit data can create alerts
- Alert on any component within the policy
- In this example, US\_SALES1 with DML Commands





# Vulnerability Assessment - Why is this important?

- <u>Database Vulnerability Assessment</u> –scans the database infrastructure for vulnerabilities and provides evaluation of database and data security health, with real time and historical measurements.
  - The Guardium Vulnerability Assessment application enables organizations to identify and address database vulnerabilities in a consistent and automated fashion.
  - Guardium's assessment process evaluates the health of the database environment and recommends improvement by:
  - Assessing system configuration against best practices and finding vulnerabilities or potential threats to database resources, including configuration and behavioral risks.
     Some examples are:
    - identifying all default accounts that haven't been disabled; checking public privileges and authentication methods chosen, etc.
    - Finding any inherent vulnerabilities present in the IT environment, like missing security patches
    - Recommending and prioritizing an action plan based on discovered areas of most critical risks and vulnerabilities. The generation of reports and recommendations provide guidelines on how to meet compliance changes and elevate security of the evaluated database environment
- VA (Vulnerability Assessment) for DB2 on z/OS (Phase 1) in Guardium V8
  - Based on DB2 Development at SVL, DISA STIG and CIS security standards
    - Server defaults
    - Patch levels
    - OS and DBMS Vulnerability Assessment

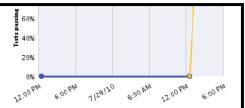


## VA for DB2 z/OS

#### Tests passing: 97%\*

\*Percentage does not take into account any current filtering

Based on the tests performed under this assessment, data access of the defined database environments conform to best practices. You have a controlled environment in terms of the tests performed. You should consider scheduling this assessment as an audit task to continuously assess these environments.



Showing 59 of 59 results (0 filtere

Result Summary						Sh	owi	ing !	59 o	f 59	res	ults	(0)	filte	red)
	Critical		Major		Minor		Caution		Info						
Privilege	48p	2f		8p					-						
Authentication		-			-				-						
Configuration	1p	-							-						
Version		-			-				-						
Other		-		-	-				-						

#### Current filtering applied:

Test Severities: - Show All -Datasource Severities: - Show All -Scores: - Show All -

Types: - Show All -

Assessment Test Results

#### Test / Datasource

Result
SYSADM privilege has been granted to unauthorize users.

z/OS Restrict system privilege - SYSADMAUTH

Test category: Priv. Severity: Critical

The SYSADMAUTH privilege grants the authority to a grantee with system administration authority. It is recommended that SYSADMAUTH privilege be granted to authorize users only. This test exclude grantee who is a member of SYSADM.

Ext. Reference: Guardium, Test ID 2164

Recommendation: We recommend you revoke SYSADMAUTH from unauthorize grantee. You can use this command to revoke: REVOKE SYSADM FROM -grantee BY ALL. To exclude authorize SYSADMAUTH grantee, you can create a group then populate it with authorize grantee and link your group to this test.

#### z/OS Grant option - Routine

Test category: Priv. Severity: Critical

This test check for object privilges on routines that has been granted with the grant option. A routine can be a user-defined function, cast function, or stored procedure. Grant option is not a good practice and should be avoid where possible. When object privileges are granted with the grant option, a user can grant privileges on that object to other users. We do not recommend granting objects privilege with grant option. This test exclude grantee who is a member of SYSADM and SYSIBM user. Granteetype 'P' is also excluded from this test.

Ext. Reference: Guardium, Test ID 2180

#### DPS: DB2 z/OS 9.1 rocket

Datasource type: DB2 Severity: None

Details: Grantee causing failure:

#### z/OS Grant option - Schema

Test category: Priv. Severity: Critical

This test check for schema privilges that has been granted with the grant option. Grant option is not a good practice and should be avoid where possible. When object privileges are granted with the grant option, a user can grant privileges on that object to other users. We do not recommend granting objects privilege with grant option. This test exclude grantee who is a member of SYSADM and SYSIBM user.

Ext. Reference: Guardium, Test ID 2181

#### DPS: DB2 z/OS 9.1 rocket

Datasource type: DB2 Severity: None

Details: Grantee causing failure:

#### z/OS Grant option - Sequence

Test category: Priv. Severity: Critical

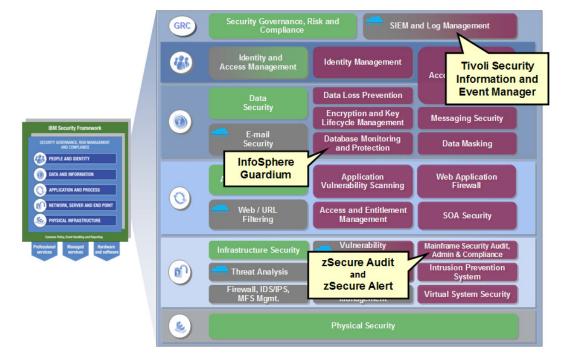
This test check for object privilges on sequences that has been granted with the grant option. Grant option is not a good practice and should be avoid where possible. When object privileges are granted with the grant option, a user can grant privileges on that object to other users. We do not recommend granting objects privilege with grant option. This test exclude grantee who is a member of SYSADM and SYSIBM user. Granteetype 'P' is also excluded from this test.

Ext. Reference: Guardium, Test ID 2182



# Why is Guardium Needed With TSIEM and zSecure?

- Traditional approaches to database security and compliance rely upon native logging, which:
  - Does not meet auditors requirements for separation of duties
  - Can easily be circumvented by DBAs
  - Imposes a higher performance overhead on database servers
  - Does not provide real-time alerting or blocking
  - Does not enable enforcement of consistent policies in heterogeneous environments





# Encryption and "Data at Rest" Protection

- Key requirement for most of the "popular" data protection initiatives
- Main requirement is to protect "data at rest" to ensure that only access if for business needto-know, and through mechanisms which can be controlled by the native security mechanisms (such as RACF)
- Consider the following scenario:
  - DB2 Linear VSAM datasets are controlled via RACF from direct access outside of DB2 via dataset access rules
  - DBA or Storage Administrator has RACF authority to read VSAM datasets in order to perform legitimate storage administration activities.
  - Administration privileges can be abused to read the linear VSAM datasets directly and access clear-text data outside of DB2/RACF protections.
- Now consider the above scenario, but with the underlying Linear VSAM datasets encrypted
  - When DBA or Storage Administrator uses their RACF dataset authorities in a manner which is outside of business need-to-know, the data retrieved is cybertext and thus remains encrypted and protected.
  - Only way to access and obtain clear-text data will be via SQL which can be protected via DB2/RACF interface

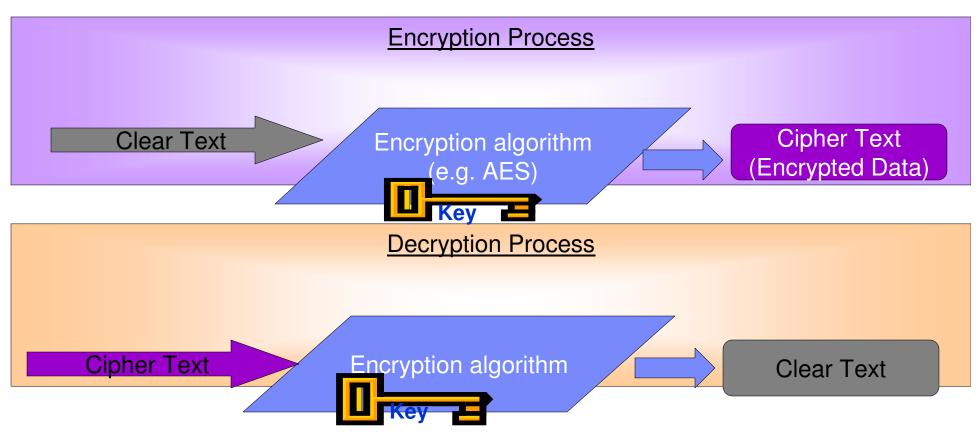


# Encryption and DB2 for z/OS

- IBM Data Server Drivers starting in V9.5 support SSL protocol and AES encryption.
- Starting with Fix Pack 2, non-Java clients supports the Secure Sockets Layer (SSL) protocol. All DB2 Version 9.5 clients now support SSL. In addition, Java and CLI clients now support 256-bit AES encryption.
- SSL connectivity and AES user ID and password encryption requires Communication's AT-TLS configured and ICSF started. AES support requires PK56287 to be applied on DB2.
- Starting with DB2 for z/OS V8, column level encryption implemented via SQL primitives is supported
- Row level encryption implemented for all supported releases of DB2 for z/OS using the IBM Encryption Tool for IMS and DB2 databases
- DS8000 DASD Based Encryption
- TS1120/TS1130 Tape Based Encryption
- TKLM (Tivoli Key Lifecycle Manager) Required for DS8000 and recommended for TS1120/TS1130

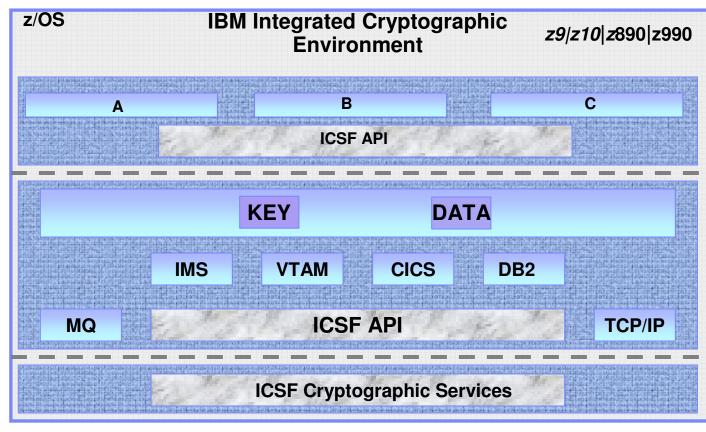


## Encryption is a technique used to help protect data from unauthorized access



- Data that is not encrypted is referred to as "clear text"
- Clear text is encrypted by processing with a "key" and an encryption algorithm Several standard algorithms exist, include DES, TDES and AES
- Keys are bit streams that vary in length
  - For example AES supports 128, 192 and 256 bit key lengths







Clear and Enciphered User Keys Master Key Verification Pattern



**CP Assist for Cryptographic Functions** 

- Problem State Instructions
- Clear Keys Only
- DES/TDES Encryption
- AES (128 Bit)
- SHA-1 (256 on z9)

#### **Crypto Express 2 Coprocessor**

- ICSF Access Only (Key 0)
- Master Key Stored Within Boundary of Crypto Express 2 Feature
- Secure Key DES/TDES Encryption
- SSL Accelerator
- Tamper Resistant



## System z9/z10 Cryptographic Support Summary

## **CP Assist for Cryptographic Function (CPACF) "free"**

- •Supports DES, TDES and SHA-1
- Standard on System z9/z10 (feature code 3863)
- Standard on every CP and IFL
- Advanced Encryption Standard (AES)
- •Secure Hash Algorithm 256 (SHA-256)
- Pseudo Random Number Generation (PRNG)

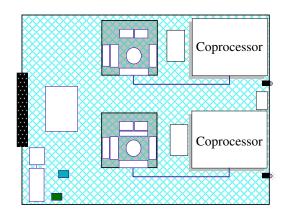
## Crypto Express2 (feature code 0863) "fee"

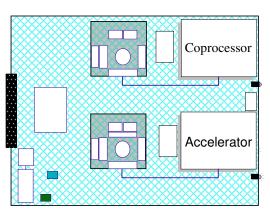
### Crypto Express3 (feature code 4863) "fee"

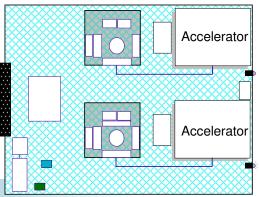
- •Two configuration modes
- Coprocessor (default)
- •Federal Information Processing Standard (FIPS) 140-2 Level 4 certified
- •"Tamper Resistant"
- (Secure Key) "Exclusive"
- SSL Accelerator (Handshake offload)

### Three configuration options

- Default set to Coprocessor (1)
- •SSL Acceleration (3)
- Mixture of configuration (2)









# Encryption - key forms on z/OS

	zSeries 900	System	System z10			
	CCA Secure Key	Clear Key	CCA Secure Key	CPACF Protected Key		
Key Wrapping: Host Storage	CCA Master Key – Key material is never visible in the clear outside the tamper resistant hardware boundary	None – Key material is visible in the clear in system and application storage.	CCA Master Key – Key material is never visible in the clear outside the tamper resistant hardware boundary	CPACF Wrapping Key – Key material is not visible in the clear in operating system or application storage.		
Key Wrapping: Key Store	CCA Master Key – Key material is never visible in the clear outside the tamper resistant hardware boundary	None – Key material is visible in the clear key store.	CCA Master Key – Key material is never visible in the clear outside the tamper resistant hardware boundary	CCA Master Key – Key material is never visible in the clear outside the tamper resistant hardware boundary		
Key Store	CKDS or application key file	CKDS or application key file	CKDS or application key file	CKDS only		
Encryption Engine	CCF	CPACF or software	CEX2C	CPACF		
Symmetric Encryption Algorithms	DES and TDES	DES, TDES and AES	DES, TDES and AES	DES, TDES and AES		
Benefits	High Security	High Performance	High Security	High Performance High Security		

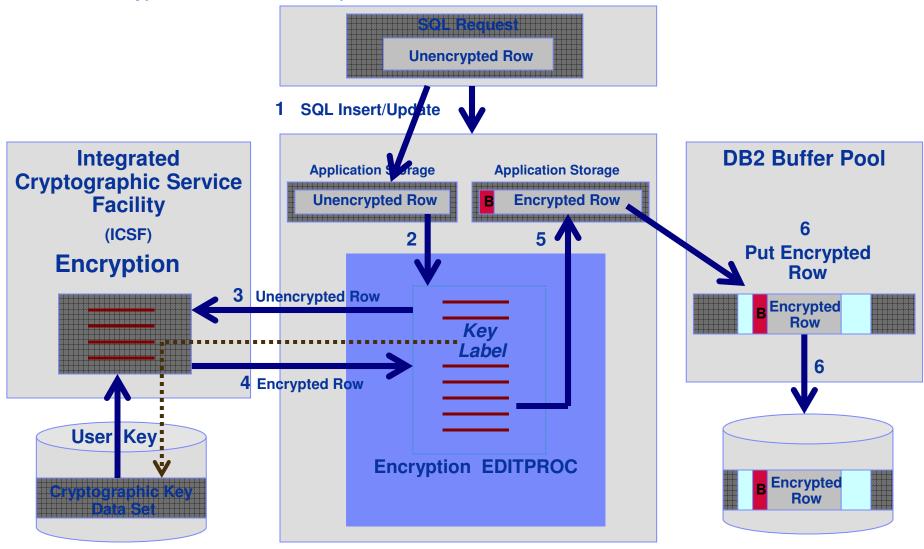


# Encryption Tool for IMS and DB2 Databases

- Generates standard DB2 EDITPROC for Accessing Cryptographic Functions
- All Supported DB2 Versions
- Member of IBM IMS | DB2 Tools Family of Products
- Pre-coded EDITPROC for encryption of DB2® Data
- Encryption/Decryption occurs at the DB2 Row Level
- Unique EDITPROC can be defined for each DB2 Table
- Exploits z/OS Integrated Cryptographic Service Facility (ICSF)
- Exploits zSeries CPACF Cryptographic Hardware Directly
- Requires no changes to your applications
- Fast implementation
- Edit Procedures (EDITPROC) are Programs that:
- Transform Data on INSERT | UPDATE | LOAD
- Restore Data to Original Format on SELECT
- Transformations on Entire ROW
- Supported by Utilities
- Implemented via Create Table specification
- Requires unload/load of data



#### DB2 Data Encryption Flow – Insert / Update





# zIIP Assisted IPSec (VPN) on z/OS

#### Benefits of having secure channel end-point on z/OS

- No clear-text data on any network segments
  - Security regulations compliance
- End-to-end authentication of secure channel end-points
  - Both end-point authentication and message authentication
- Key management and storage done on System z by z/OS
- Compliance with end-to-end security regulations

#### System z CPU cost is a concern

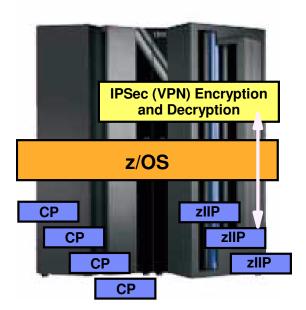
- Encryption/decryption CPU cost can be a significant percentage of overall CPU cost for a given application
- Especially the case for streaming workloads (file transfer type of workload)

#### zIIP processors

- Specialty processor on System z9 or later hardware
- zIIPs priced lower than general purpose processors
- No IBM software charges on zIIPs

#### zIIP Assisted IPSec

- Use zIIP processors for most IPSec encryption/decryption
- Lower the cost of doing IPSec processing on z/OS



System z9 or later z/OS CS V1R8 + PTFs z/OS CS V1R9

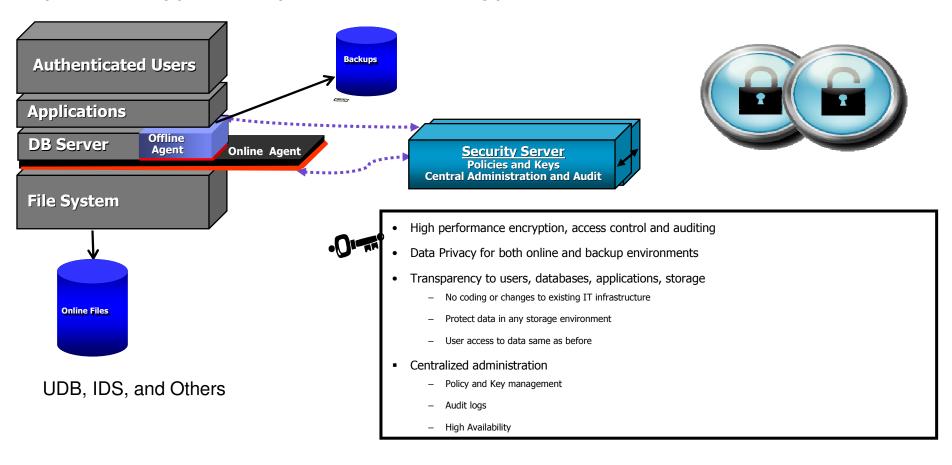


# IBM DS8000 Disk Encryption - Characteristics

- Customer data at rest is encrypted
  - Data at rest = data on any disk or in any persistent memory
- Customer data in flight is not encrypted
  - Data in flight = on I/O interfaces or in dynamic memories (Cache, NVS)
    - If you can read/write to disk, you get access to clear-text data.
- Uses Encrypting Disk
  - Encryption hardware in disk (AES 128)
  - Runs at full data rate
  - 146/300/450 GBs 15K RPM
    - No measurable performance impact
- Integrated with Tivoli Key Lifecycle Manager (TKLM)
  - DS8000 automatically communicates with TKLM when configuring encryption group or at power on to obtain necessary encryption keys to access customer data
  - Each disk has an encryption key
    - Data is always encrypted on write and decrypted on read
    - Encryption key is wrapped with access credential and maintained within the disk
    - Access credential maintained by TKLM
    - Establishing a new encryption key causes cryptographic erasure
- Key attack vectors prevented:
  - Disk removed (repair, or stolen)
  - Box removed (retire, or stolen)



### Optim Encryption Expert – Data Encryption





# Discovery, Test Data Management/Obfuscation, and Data Growth





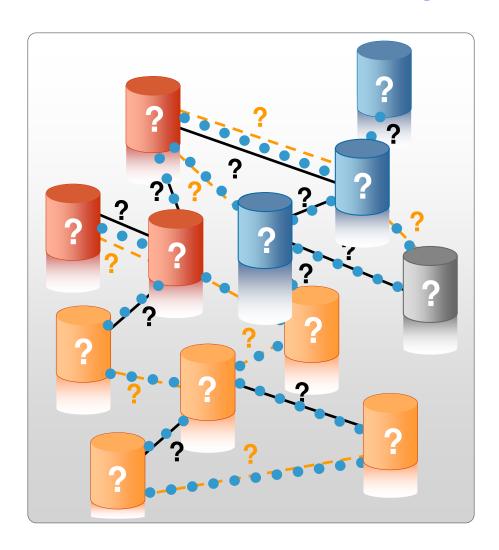
### Automate Discovery and Accelerate Information Understanding

- Significant Acceleration of Information Agenda projects

  – Data Growth Management

  - Test Data ManagementSensitive Data De-identification
  - Application/Data Consolidation, Migration & Retirement
  - Master Data Management and Data Warehousing
- Why is this Different?

  - <u>Data-based</u> discovery
     <u>Automate</u> discovery of business entities, cross-source business rules & transformation logic
  - Evaluate multiple data sources simultaneously
  - Identify & <u>remediate</u> cross-system rules and inconsistencies



© 2010 IBM Corporation 13



## InfoSphere Discovery Components

#### **Cross-Profiler**

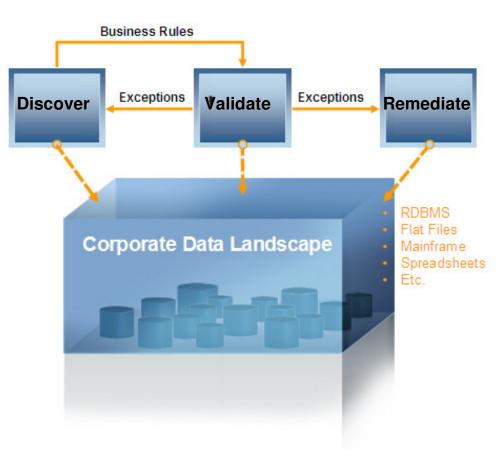
 Basic profiling plus automated primaryforeign key, business entity & crosssource overlaps discovery

#### **Unified Schema Builder:**

 Prototype empty targets from the combination of many data sources

#### **Transformation Analyzer:**

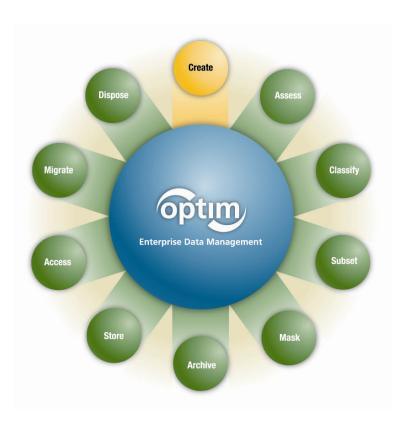
 Discover complex business rules and transformation logic between two data sources



14 © 2010 IBM Corporation



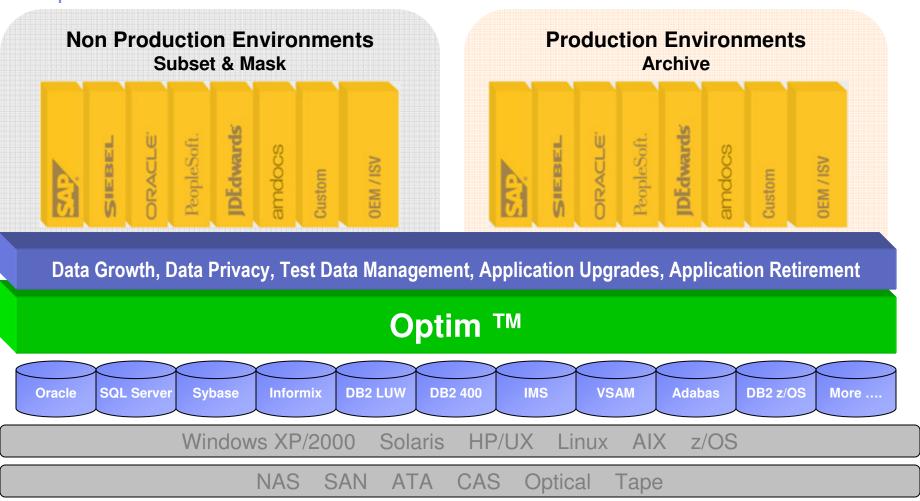
#### Optim™ Solutions



- Optim<sup>™</sup> Data Growth Solution (Archiving)
  - Improve performance
  - Control data growth, save storage
  - Support retention compliance
  - Enable application retirement
  - Streamline upgrades
- Optim<sup>™</sup> Test Data Management Solution
  - Create targeted, right sized test environments
  - Improve application quality
  - Speed iterative testing processes
- Optim<sup>™</sup> Data Privacy Solution
  - Mask confidential data
  - Comply with privacy policies
- Enterprise Capabilities
  - Single, scaleable solution for complex multi-DB application environments



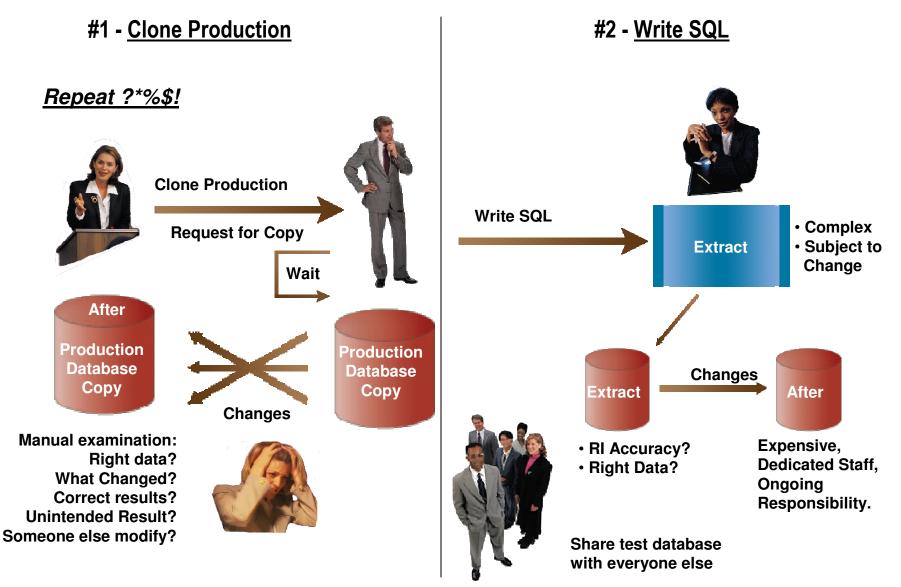
#### **Enterprise Architecture**



 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



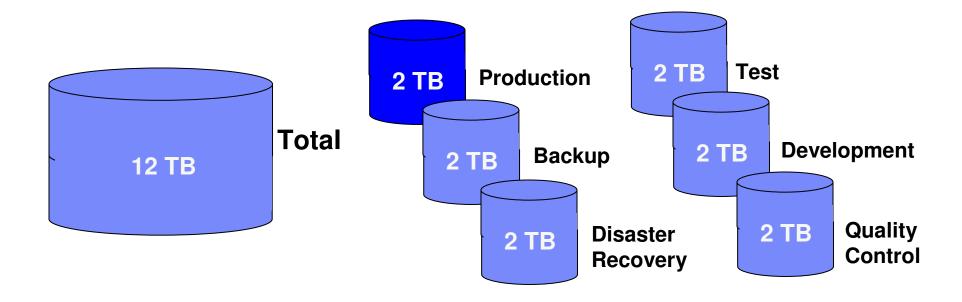
#### **Current Practices?**





#### Data Multiplier Effect

#### **Actual Data Burden = Size of production database + all replicated clones**





#### Data Obfuscation and Data Relationships

- Data Obfuscation
  - AKA data masking, depersonalization, desensitization, obfuscation 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

Obfuscation of Key Columns: When sensitive data is a key column involved in a relationship, then care must be taken to ensure that the obscured value is changed in all rows of data that reference that changed value, this is referred to as key propagation.



© 2010 IBM Corporation

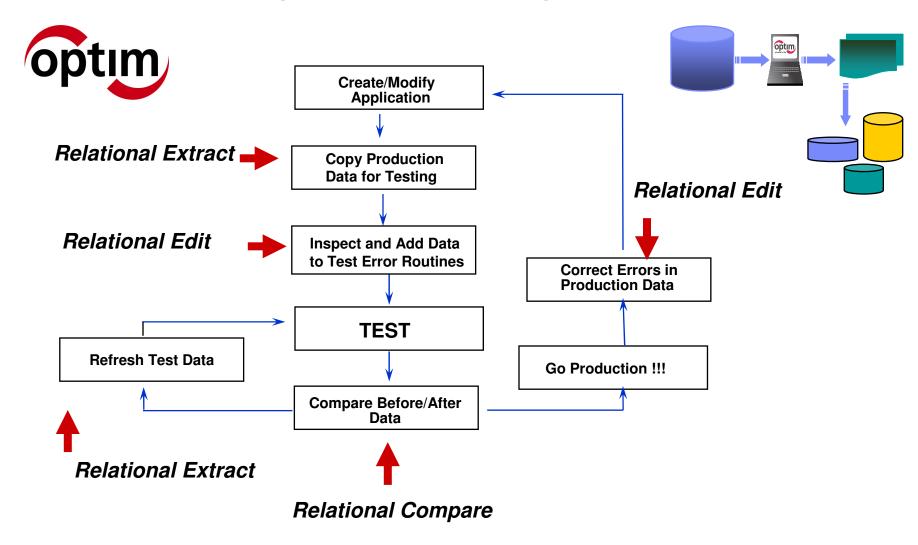


# Optim Overview Relational Extract Facility / Test Data Management



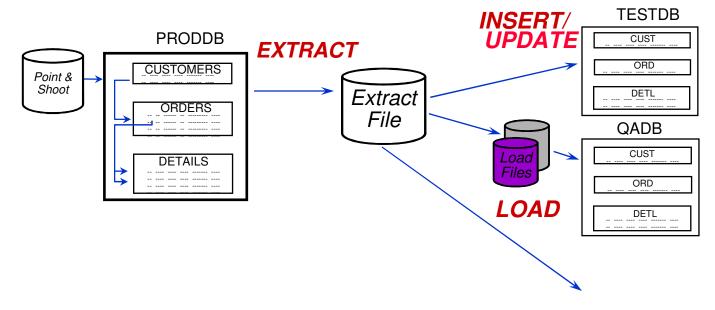


#### Product Overview : Optim Test Data Management

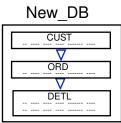




## **OPTIM Relational Extract Facility**



- Creating and maintaining test data bases
- Migrating data
- The data and/or the object metadata can be extracted



Create



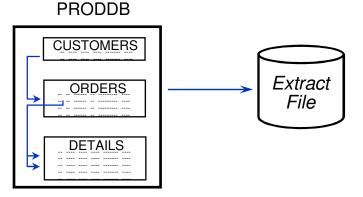
#### Defining the Extract.....

**Tables** 

**Views** 

**Synonyms** 

Aliases



## Required:

- Start Table
- Set of Tables

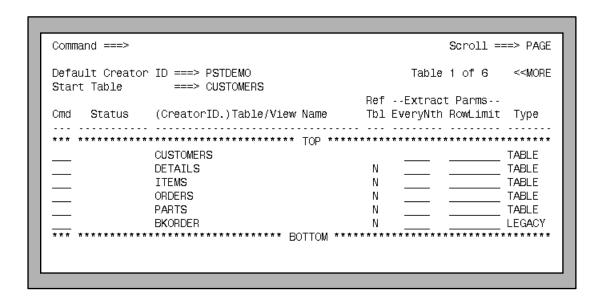
## Optional:

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



# **Extract Process**

The Table List



- Identify the Start Table
- Use the RELATED functions to populate list
- Include random selection factor, extract limits and selection criteria



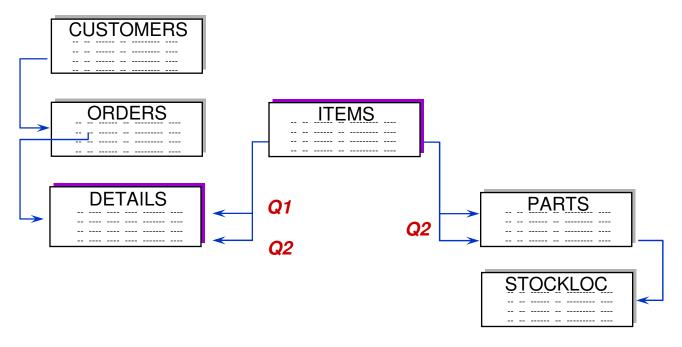
## **Extract Process** Relationship Usage

Command ===>		S	Goroll ===>	PAGE
For Each Relationsh	ip Indicate:	F	el 1 of 3	
		de its Parent Row to Satis tisfy any RI Rule, Include	_	
Q Q Chil	.d		Relat	ion
Cmd Status 1 2 Limi	t Parent Table	e Child Table	Name	Туре
*** ********	*****	** TOP ************	*****	****
SELECT Y N	CUSTOMERS	ORDERS	RCO	DB2
UNSEL Y N	ITEMS	DETAILS	RID	DB2
SELECT Y N	ITEMS	PARTS	RIP	PST
SELECT Y N	ITEMS	BKORDER	RIB	PST
SELECT Y N	ORDERS	DETAILS	ROD	DB2

- Select relationship paths
  Defined to DB2 catalog or PST Directory
- Designate relationship traversal
- Limit number of child rows extracted



# Extract Process Relationship Traversal



- Q1 Only ITEMS that are parents of DETAILS
- Q2 All other DETAILS for those ITEMS ... Each of the PARTS for those ITEMS



# Extract Process Show the Extract Steps

Command ===> Scroll ===> PAGE

Step 1: Extract Rows from Start Table PSTDEMO.CUSTOMERS. Row List is used and Determines the Rows Selected.

Step 2: Extract Rows from PSTDEMO.ORDERS which are Children of Rows Previously Extracted from PSTDEMO.CUSTOMERS in Step 1 using Relationship RCO.

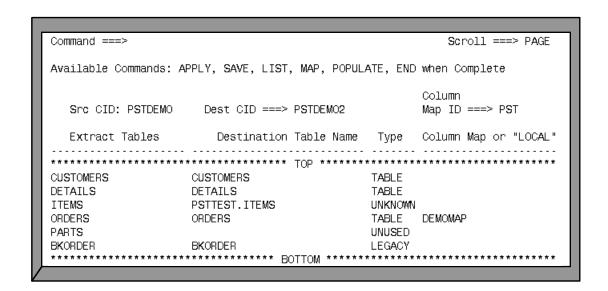
Step 3: Extract Rows from PSTDEMO.DETAILS which are Children of Rows Previously Extracted from PSTDEMO.ORDERS in Step 2 using Relationship ROD.

Untraversed Table(s): PSTDEMO.ITEMS PSTDEMO.PARTS PSTDEMO.BKORDER

- Steps required to perform extract
- Cycles processed
- Untraversed tables



# Populate Destination Tables Table Map



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



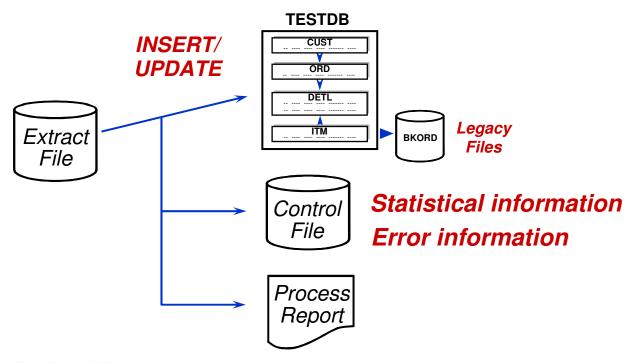
# Populate Destination Tables Creating New Tables

	Command ===>			Scroll ===> PAGE
	Cmd Status	Type	Object Name	Database Tablespace
	EXISTS EXISTS EXISTS	TABLE INDEX PK(DB2)	PSTDEMO2.CUSTOMERS PSTDEMO2.XCUSTPK	DSOFTECH SSOFTECH
	EXISTS EXISTS EXISTS	TABLE INDEX PK(DB2)	PSTDEMO2.DETAILS PSTDEMO2.XORDETPK	DSOFTECH SSOFTECH
Missing _	EXISTS SELECT SELECT SELECT	FK(DB2) TABLE INDEX PK(DB2)	ROD PSTTEST.ITEMS PSTTEST.XITEMPK	DSOFTECH SSOFTECH
destination object(s)	SELECT SELECT SELECT	VIEW LEGACY DATASET	PSTTEST.V_ITEMS PSTDEMO2.BKORDER PST.ADB2.BKORDERS	

- Select destination object(s) to be created from source table definitions
- Functions include DROP, key conversion, and display of SQL



# Populate Destination Tables Control File



#### If INSERT/UPDATE errors occur:

- 1. BROWSE the control file for error information
- 2. RETRY/RESTART the INSERT/UPDATE



#### Extract Parameters and Execution - Data Source

```
Specify EXTRACT Parameters and Execute
                                                     SUBSYS: DSNC
Current AD Name : OPTZOS.SYS248.HPUEXT
                                           B-Both)
 Limit Number of Extract Rows ===> 30000 (1-9999999, Blank-Site Limit)
Perform Convert with Extract ===> N
 If Batch, Review or Save JCL ===> R
Command ===>
```

IBM HPU as source



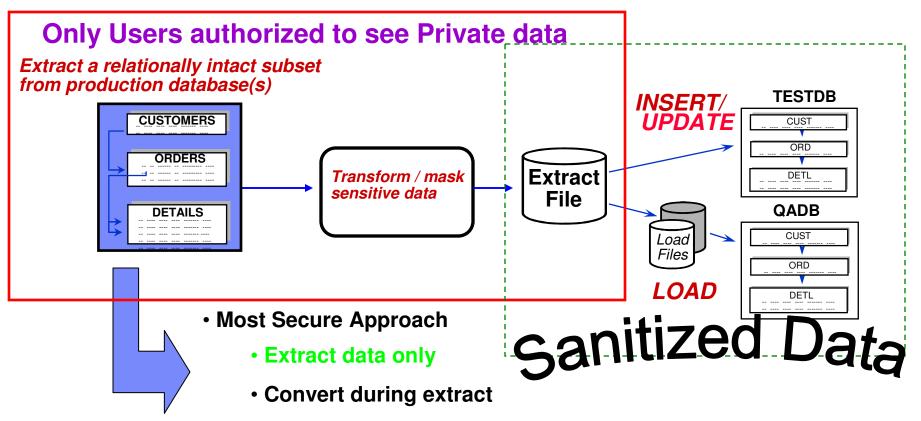
### Extract Data Parameters - Image Copy Opions

```
Specify EXTRACT Parameters and Execute ------
                                                   SUBSYS: DSNC
                  --Specify Unload Program Parameters-----
                                   B-First On or Before Date/Time,
     Date (YYYY-MM-DD)
    Time (HH.MM.SS)
Command ===>
```

Different Image Copy Input Options



# Data Privacy in Application Testing



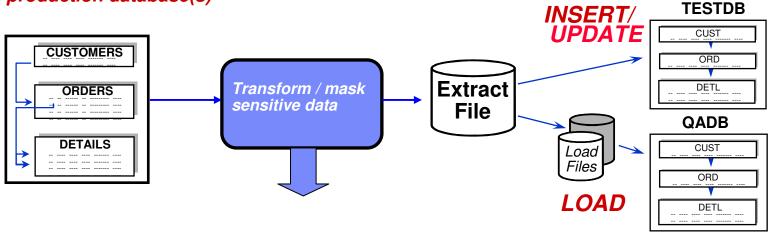
- Extract file already contains masked data
  - Can be shared with testers to reuse

© 2010 IBM Corporation



# **Data Privacy in Application Testing**

Extract a relationally intact subset from production database(s)



#### **Data transformation functions:**

- **Hard-code literals.**
- special registers such as date, time
- **Arithmetic calculations**
- Sequential number generation
- Random number generation
- Substring and/or concatenation of values
- Lookup Table Functions Random, Specific or HASH
- Intelligent TRANSformation Library SSN, CCN, email,...
- Access to client-defined exit routines to apply complex algorithms, encryption, ...
- Propagation of masked primary keys to dependent foreign keys

© 2010 IBM Corporation



#### **Propagating Keys**

# CUSTOMERS

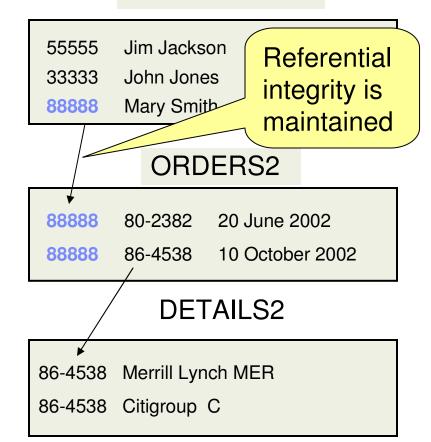


27645 80-2382 20 June 200227645 86-4538 10 October 2002

#### **DETAILS**

86-4538 Merrill Lynch MER 86-4538 Citigroup C

#### **CUSTOMERS2**





## 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

#### Orders Table

Now these

**Orphans** 

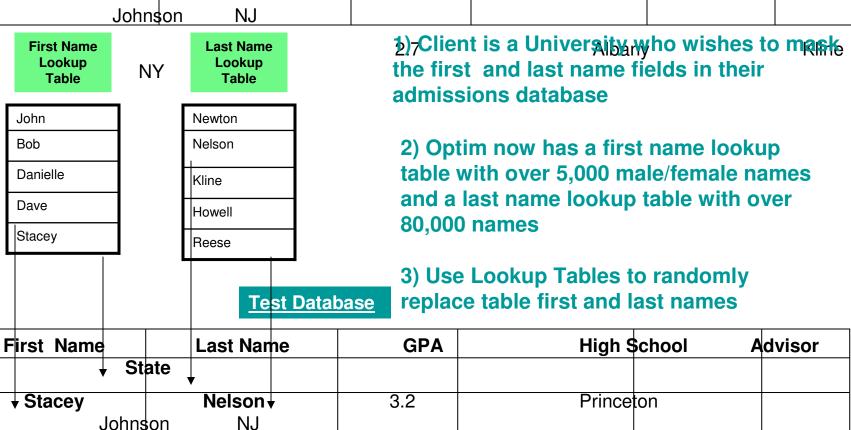
Cust ID	Item #	Order Date
27645	80-2382	20 June 2004
27645	86-4538	10 October 2005



#### First Names and Last Names Data Sets

#### **Production Database**

First Name	Last Name	GF	PA High	School
Advisor Sta	te			
Paul	Smith	3.2	2 Princ	ceton
Johns	on NJ			



DaveReese2.7AlbanyKline

NY



#### 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	ОН	44101
\$8,777,733,811	105,333	6767 Rte 1 S	Princeton	NJ	08540

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

Address Lookup Table

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

#### **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

- 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 CASS
    address using
    enhanced random
    lookup function



### Intelligent Masking Capability

#### **Production Database**

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

Data before Masking

#### **Test Database**

F. Name	L. Name	Credit Card#	SSN#
John	Denver	5326458711224956	300111 0011
Vanessa	Jones	4972584612457742 Valid	154-74-7788 Valid

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

#### How are these numbers valid?

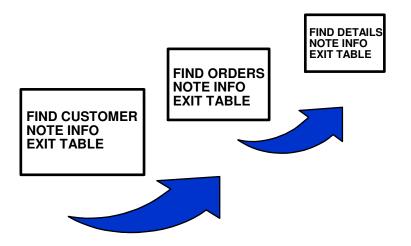
For Social Security Numbers	For Credit Card Numbers
A Social Security Number (SSN) consists of nine digits. The first three digits is called the "area number". The central, two-digit field is called the "group Number". The final four-digit field is called the "serial Number". All numbers must fit the 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.



#### Traditional vs. Relational Tools

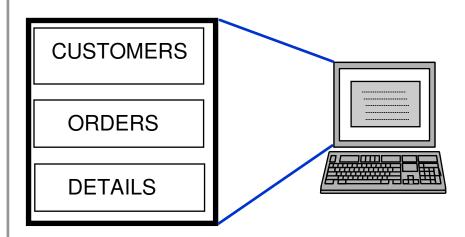
## Single Table Editors

- 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





# Browsing or Editing Data

Command ===>			Scroll ===> PAGE		
	CUST_ID	CUSTNAME	ADDRESS	CITY	STATE
*** *		**************************************	****** TOP **********	. * * * * * * * * * * * * * * * * * * *	MN
			823 Chestnut St	_	
	00049		120 Central Avenue		CA
	00094		Aramingo Place		
	00041	Prime Time Video		_	CA
	10051	Take Home Movies		Coyote	CA
	01150	Rick's Flicks	823 Chestnut St	Forked River	NJ
	00203	Movies-R-Us	1772 Bridge St	Brigantine	NJ
	00191	Popcorn	15 Crystal Park	Green Pond	NJ
	00 <b>26</b> 0	Five Star Videos	123 Howe Lane	Норе	NJ
	00189	Showtime	322 Rt 28 ;	Little Ferry	NJ
	00160	Reely Great Videos	590 Frontage Rd	Pellettown	NJ
	00156	Prime Tyme	982 Upper State St	Hackensack	NJ
	00015	Director's Chair	347 Miners Row	Нарру Сатр	CA
	00141	Showcase II	57 Rock Hollow	Brick	NJ

- User can define how data is displayed
  - SORT, HEX, sidelabel/columnar format
- All DB2 access authority enforced



### Joining to Another Table

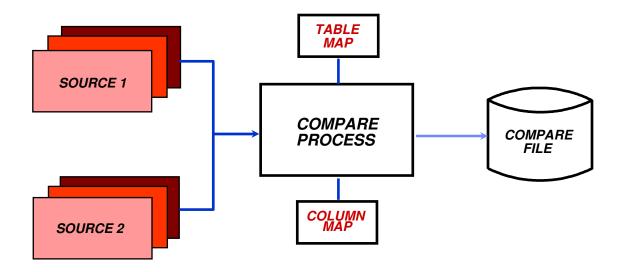
#### JOIN [table]

```
Command ===>
                                                         Scroll ===> PAGE
Cmd F == Table: PSTDEMO.CUSTOMERS(T1) ========== 1 OF 36 === MORE>>
     CUST_ID
                  CUSTNAME
                                      ADDRESS
                                                        CITY
                                                                  STATE
                             593 West 37th Street Angels Camp
      00068 Audio-Video World
Cmd F == Table: PSTDEMO.ORDERS(T2) =========== 1 OF 4 === MORE>>
     ORDER_ID CUST_ID ORDER_DATE ORDER_TIME FREIGHT_CHARGES ORDER_SALESMAN
              00068 12/02/1997 08.16.09
                                              14.80
                                                            WE005
         222
              00068 12/31/1997 14.22.31
                                              19.05
                                                            WE005
         278
              00068 02/02/1998 11.51.47
                                              21.97
                                                            WE005
                                                                           wer-
       30013 00068 01/12/1998 15.23.04
                                              33.85
                                                            WE005
```

© 2010 IBM Corporation



## **OPTIM TDM 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



### What are the Key Drivers of Data Growth?

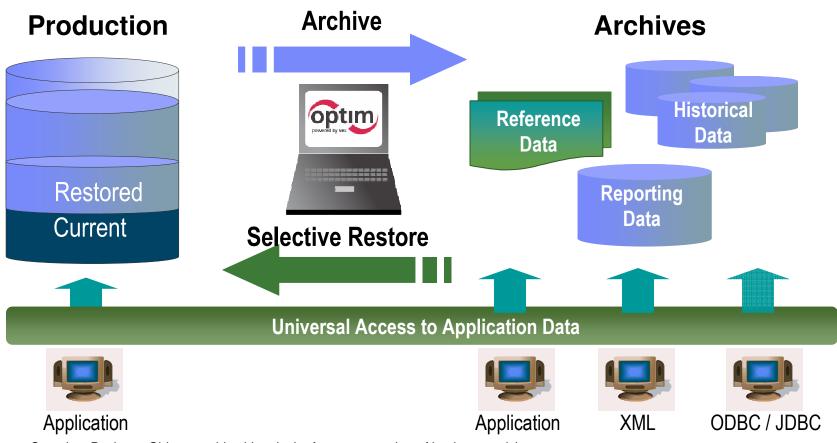
- Mergers & acquisitions
- Organic business growth
  - -eCommerce
  - -ERP/CRM
- The digital revolution
- Records retention
  - -Basel II
  - -SOX
  - -Euro-SOX
- Data multiplier effect
- Forrester estimates that 85% of data stored in databases is inactive



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



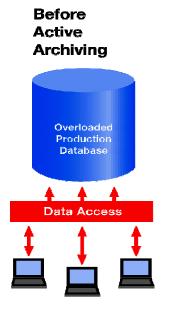
### Optim<sup>™</sup> Data Growth Solution: Archiving

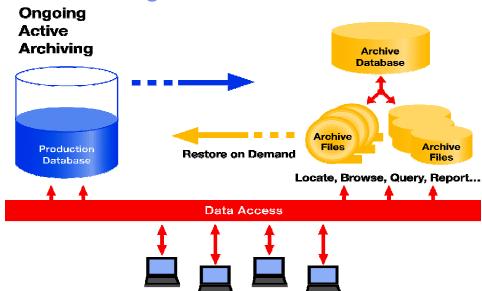


- Complete Business Object provides historical reference snapshot of business activity
- Storage device independence enables ILM
- Immutable file format enables data retention compliance



#### **Active Archiving Defined**

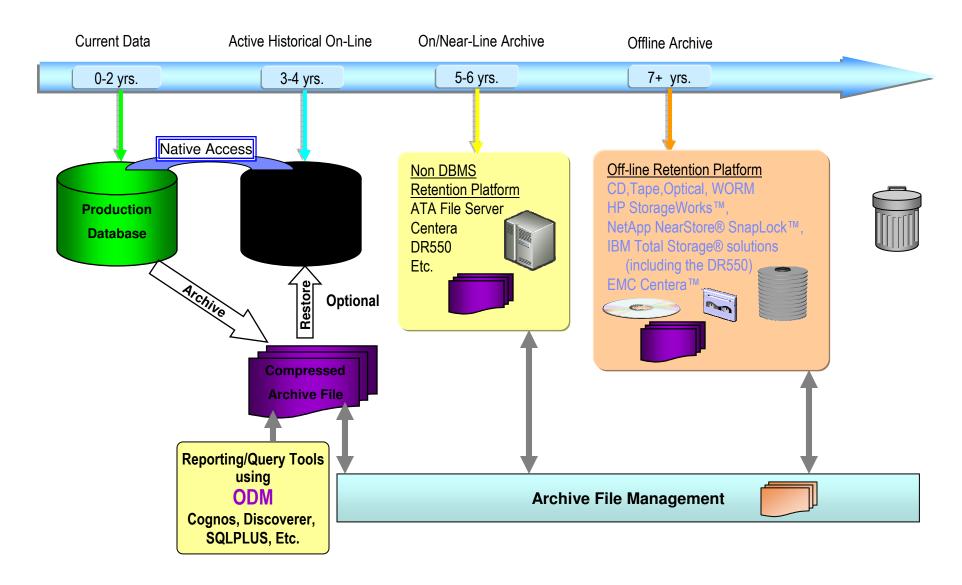




- Reduce the amount of data in the application database by:
  - Separating infrequently accessed data from transactional data
  - Preserve metadata and relationships of archived data outside db
  - Archive relational subsets vs. entire files
- Enable easy user access to archived information
  - View,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

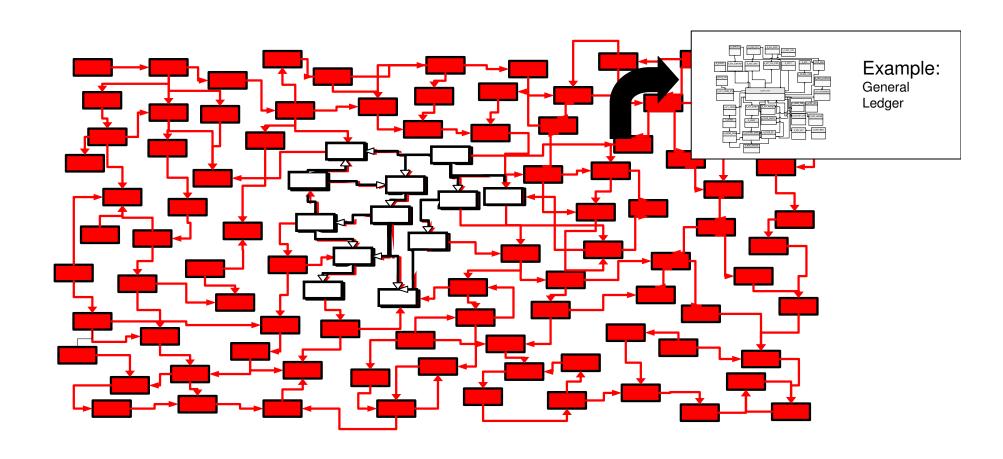


#### Information Lifecycle



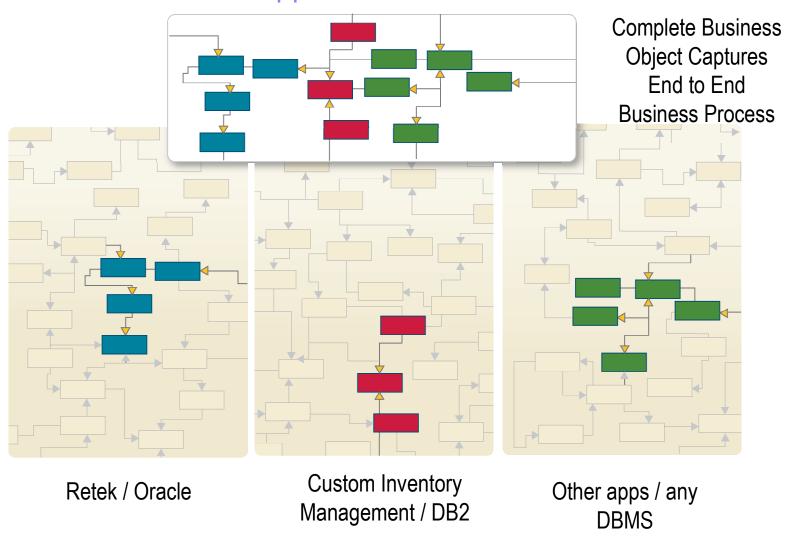


# Our Unique Capability: Complete Business Object



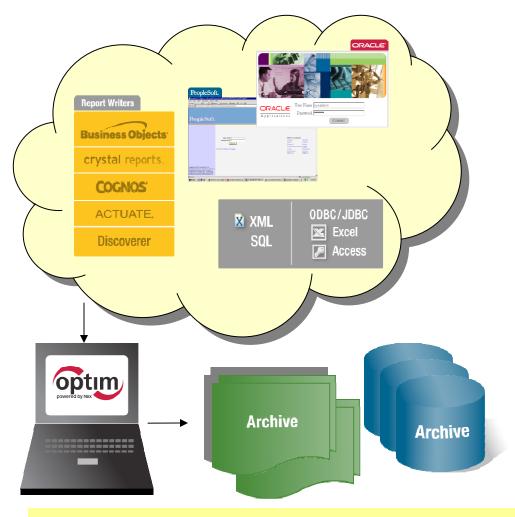


## **Extract - Federated Data Support**





#### **Universal Access**



- Native application access
  - Familiar screens and processes
- Application independent access
  - Industry standard methods: SQL, ODBC/JDBC, XML
  - -Portals
  - Report writers: CrystalReports, Cognos, BusinessObjects, Discoverer, Actuate
  - Desktop formats: Excel, CSV, MS Access
  - Database formats

Access Any Record, Anytime, Anywhere!



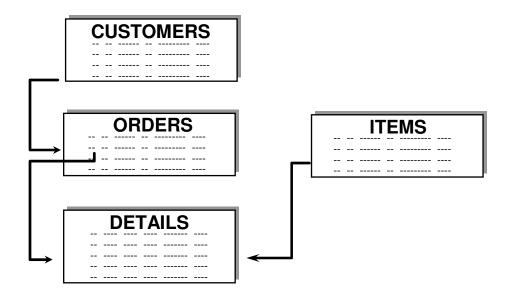
## Steps for Archiving Data

- Identify the data to be archived
- Define the data to be deleted
- Choose a delete method
- Create the archive & Delete the data
- Find Data in the Archives
- Browse or Restore



## Identify the data to be archived

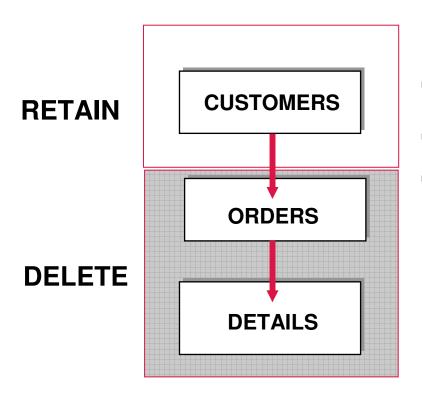
# Access Definition Defines a subset of of relational data



- Start table
- Associated data
- Relationships
- Extraction rules
- Index specifications



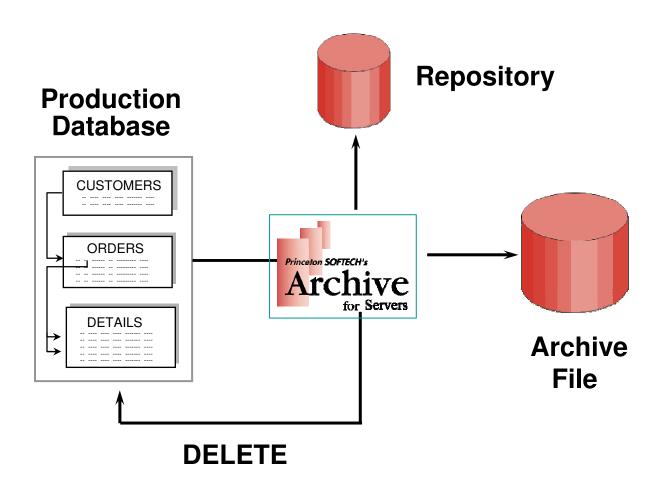
#### Define the data to be deleted



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



#### Create the archive





## Researching the Archives



#### Direct access to archived data:

- User maintainable indexes
- Global searches
- Simple or complex criteria
- Intelligent browse
- ODBC Access
- ODM Access
- Save as CSV

# Restore archived data only when you need to



## Why Restore?



#### Browse archived data for:

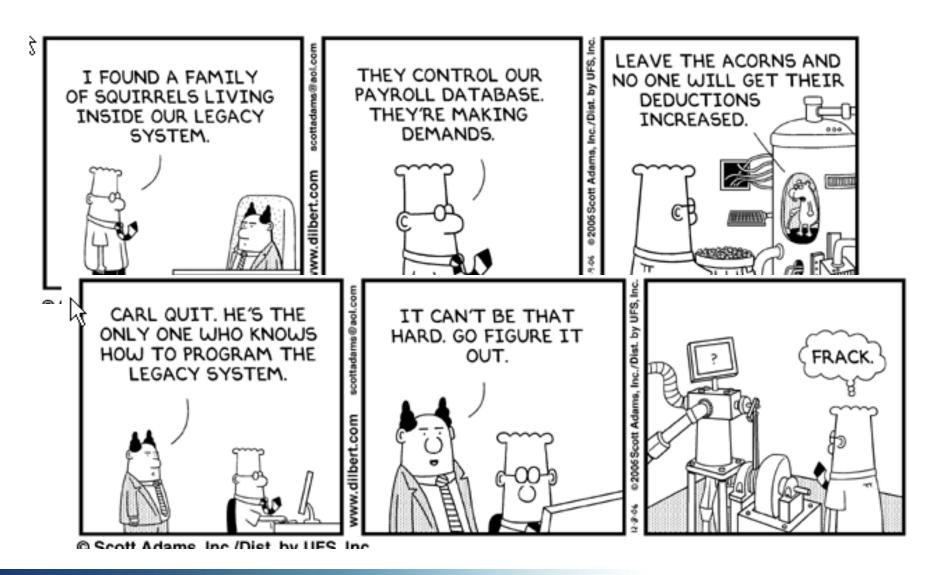
- Customer service
- Answering questions
- Archive research

#### Restore archived data for:

- Audit situations
- Application-generated reports



### Untold stories of legacy applications





#### Issue: Retire obsolete applications

#### CIO:

- Reduce risk and cost by sunsetting obsolete or redundant technologies
- Reduce IT expenses (software, hardware, personnel)
- Preserve access to legacy system data for retention compliance

#### **Business Line Executive:**

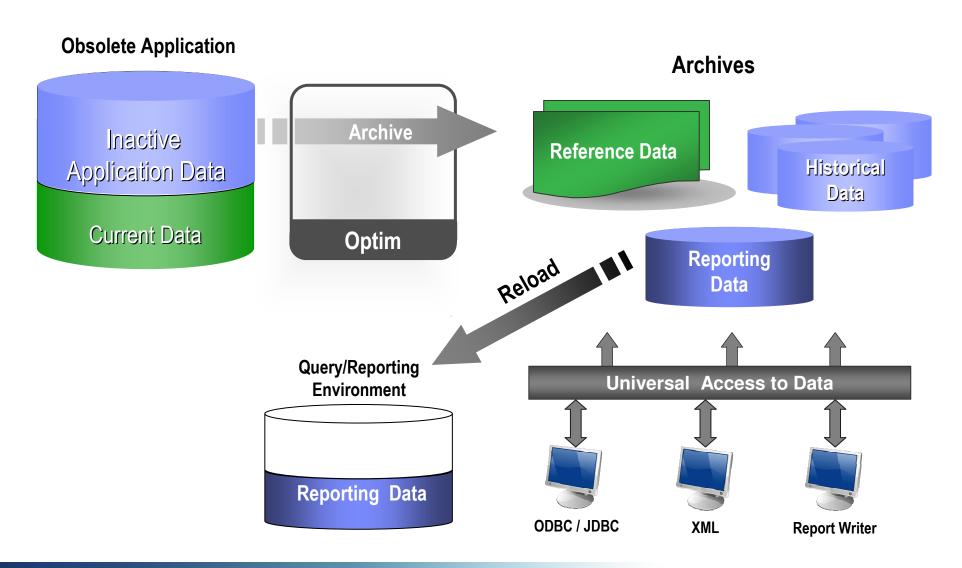
- Minimize negative budget variances and reduce IT charge-backs to line of business
- Allocate scarce resources for priority business needs
- Maintain access to historical business data for retention compliance

#### **Technical Management:**

- Eliminate expenses associated with underperforming assets
- Enable application-independent access to legacy system data for retention compliance
- Reduce risks from dependence on specialized labor and no longer supported vendor products



# Enterprise Challenge: Application Retirement Optim Supports Application Retirement Strategies





### Summary

#### Take Back Control with IBM Information Protection solutions on System z:

- Transform your information from a Liability into your most strategic, valuable Asset
- Help manage business risk by enforcing security, audit, privacy and policy controls
- Lower operational costs by optimising data management, retention and archiving

#### Software, Hardware and Expertise.

- Information Management the most complete end-to-end Information Protection software solutions
- Information Protection Entry point as part of your wider Information Governance strategy
- System z the ultimate platform to for security
- Clear ROI business cases for each area of Information Protection.

#### For more information visit

- www.ibm.com/software/data/db2imstools/solutions/data-governance.html
- Download the Information Protection white paper now.



