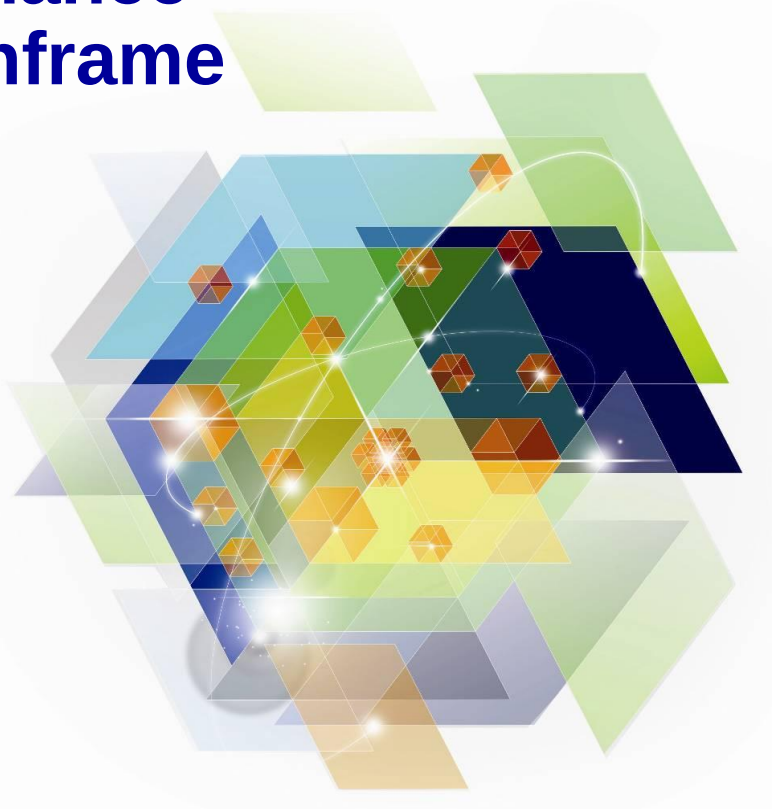
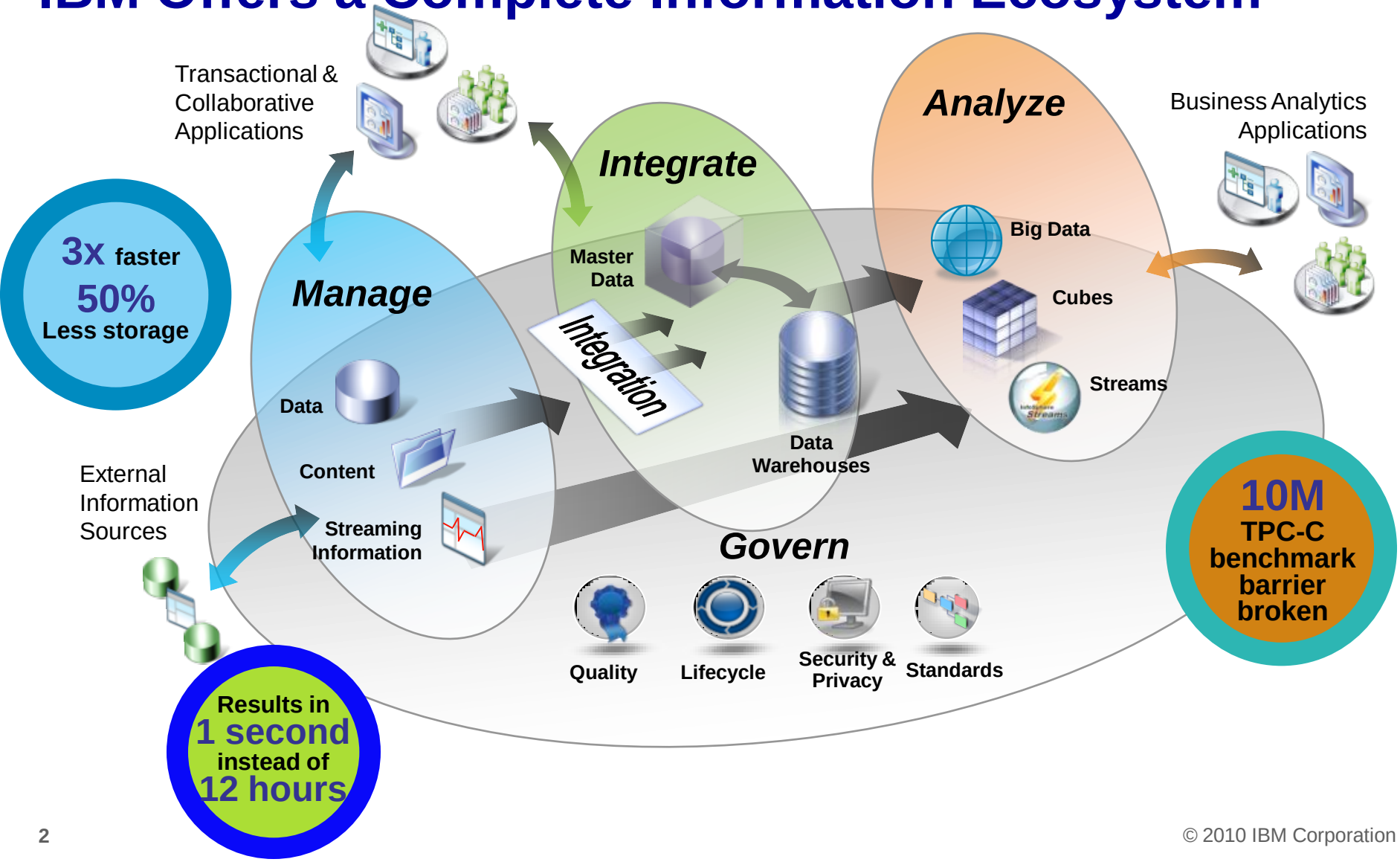
Delivering Information Governance and Data Security for the Mainframe

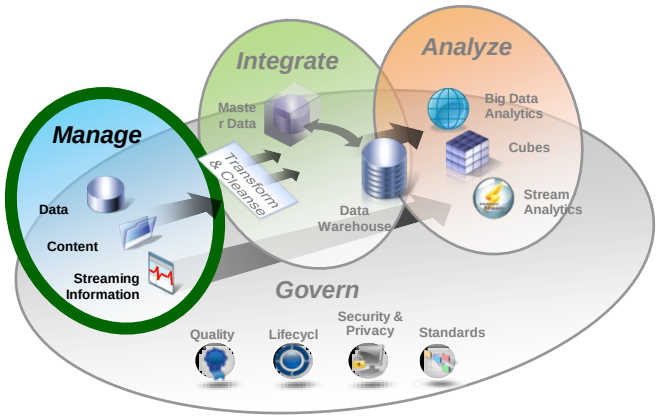


Dan Wardman,
Vice President,
Information Management for System z,
IBM Software Group

IBM Offers a Complete Information Ecosystem



Managing Information



Data Management

- Capabilities for managing structured information
- Leverage existing IT investments across a heterogeneous landscape
- Flexibility to adapt to changing business needs
- Lower total cost of ownership

Capabilities

Clustering

Data compression

Autonomic

In memory DB

Java Optimizations

SQL skins

Workload Optimization

XML

Products

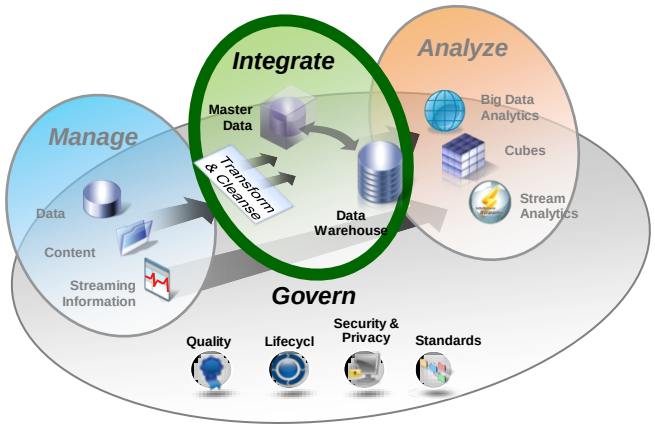
DB2

Informix

Solid DB

IMS

Integration



Data Warehouse

- Deal with TB and PB of data
- Reduced time to value, days not weeks
- Perform data mining without copying data
- Work with analytical and transactions data in real time.

Products

InfoSphere Data Warehouse

NETEZZA

Data Warehouse Appliances 2010

Capabilities

Data mining & visualization

Embedded data movement and transformation

Unstructured Data Analysis

Extreme Workload Management

Data models & sample reports

Integrated analytics systems

Mashups

Real time analytics

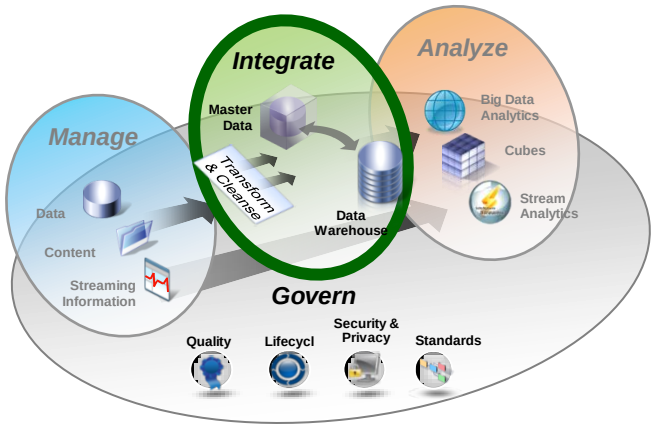
Industry models

Clinical analytics

Modeling & design tools

Hadoop

Integration



Information Integration

- Integrating and transform data and content to deliver authoritative, consistent, timely and complete information

Capabilities

Data Discovery

Data Extraction, Transformation and Loading

Data Modeling

Data Profiling

Data Quality

Metadata Management

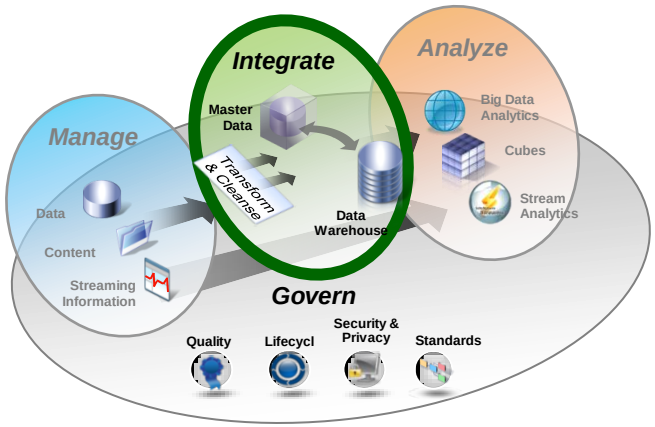
Products

InfoSphere Information Server

Change Data Capture

Replication

Integration



Products

- InfoSphere MDM
- Initiate Exchange
- Entity Analytics

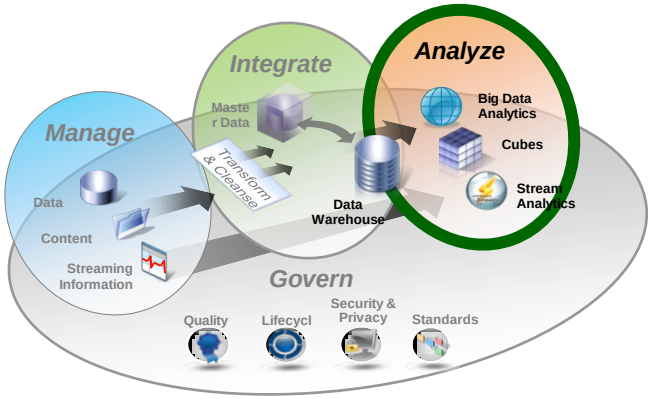
Master Data Management

Create common data about customers, suppliers, partners, products, materials, accounts and other critical entities

Capabilities

- Account
- Customer
- Product information
- Asset track and trace
- Entity Resolution & Analysis
- Name Recognition
- Healthcare Information Exchange
- Location

Business Analytics



Business Analytics

Delivers complete, consistent and accurate information that decision-makers trust to improve business performance

Capabilities

Business Intelligence

Financial Performance & Strategy Management

Analytic Applications

Predictive & Advanced analytics

Governance Risk & Compliance

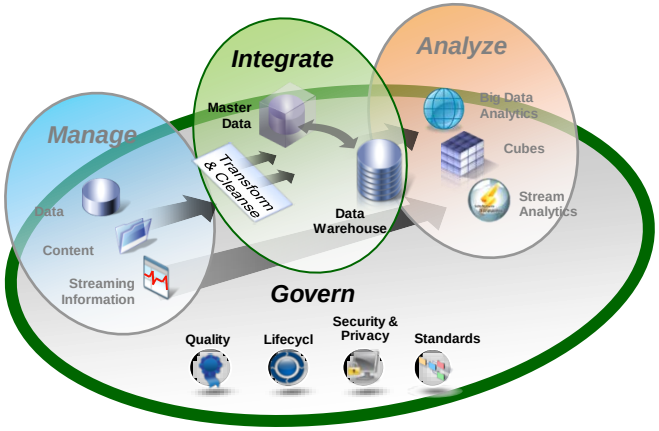
COGNOS®
Business intelligence & performance management 2008

SPSS
Predictive analytics 2009

OPENPAGES
Governance, compliance, risk management 2010

CLARITY SYSTEMS
Financial governance 2010

Information Governance



Information Governance, Risk, and Compliance

Delivers complete, consistent and accurate information that decision-makers trust to improve business performance

Products

- Optim
- Guardium
- Infosphere Discovery

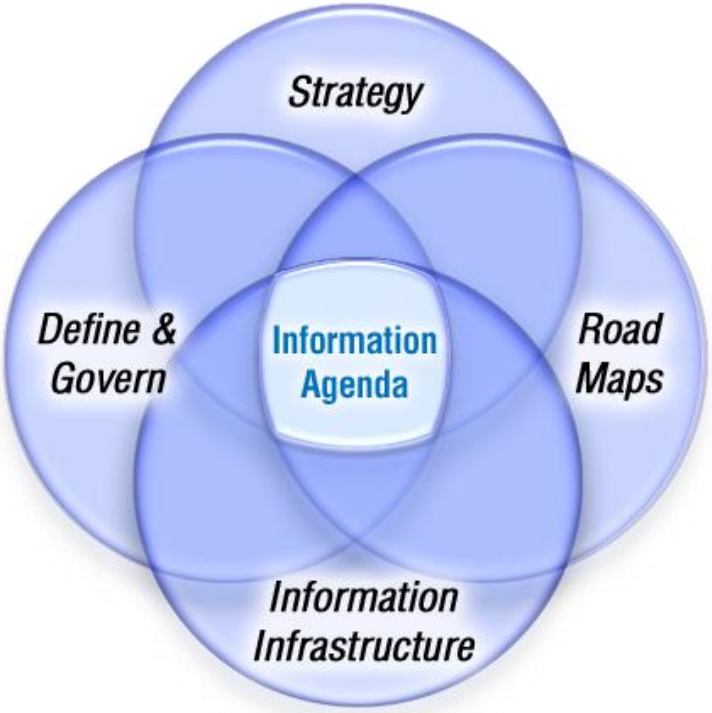
Capabilities

- Data Discovery
- Data Encryption
- Data Growth Management
- Data Masking
- Database Activity Monitoring
- Test Data Management

The IBM Information Agenda Approach

Establish end-to-end vision & business-driven value

Align people, process, & information



Accelerate projects for short & long-term ROI

Architect an extensible information infrastructure

Challenges around governing the usage, sharing and processing of massive amounts of electronic data

- **Growing infrastructure and resource management costs**
  - No policies for management of data growth – performance degradation
  - Redundancy of data
  - Disparate, complex applications and more users
- **Lack of common security and privacy requirements**
  - Risk of security breaches, compliance, audit failures
  - No ability to assess areas of vulnerability and prevent unauthorized intrusion
  - Lack of an overall protection strategy (relational / non-relational data and access controls)
- **Lack of trusted information**
  - No alignment of definitions across business and IT
  - No clear understanding of data sources & relationships
  - No standardized quality rules or threshold metrics
  - Lack of control over test data environments

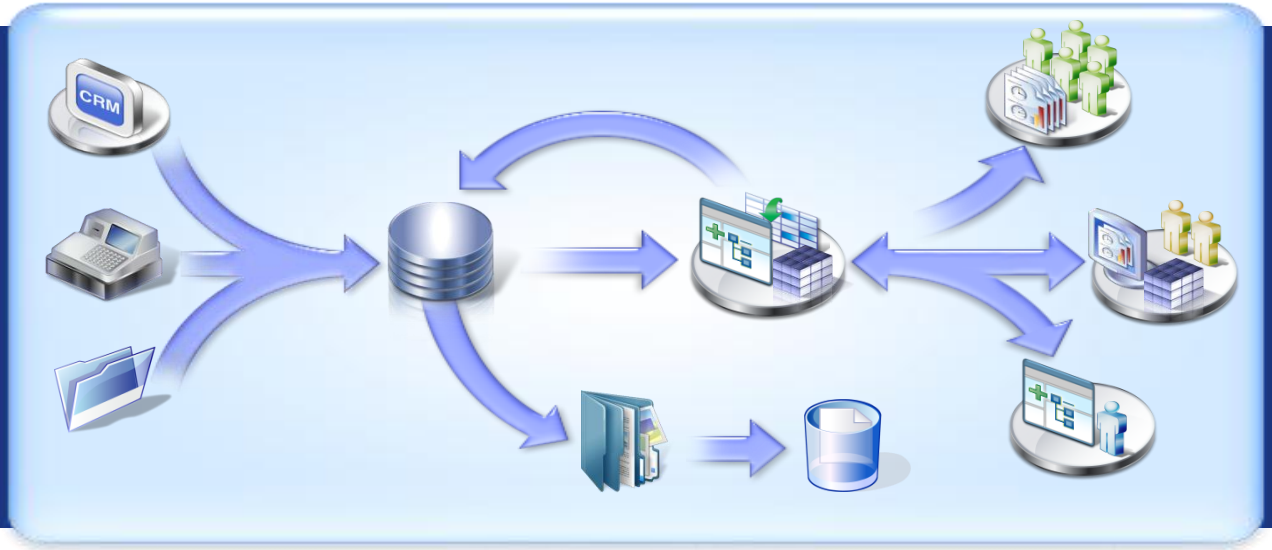
"[A]n [information management] strategy should incorporate life-cycle information governance practices [to ensure] consistent execution of ... business optimization, agility, and transformation [initiatives]."
– Forrester Research, Inc., "Refresh Your Information Management Strategy to Deliver Business Results"
Rob Karel & James G. Kobielus, August 2009

"If you are going to protect your company's most valuable asset—your data—you will begin to view data security as a component of a more comprehensive information governance strategy."
– Hurwitz & Associates,
"Why you need an information governance strategy for 2010"
Marcia Kaufman, December 2009

"By 2013, 25% of the companies in highly regulated industries will create and staff positions in accounting, human resources, compliance and audit and law that deal explicitly with the management of information via technology."
– Gartner, Inc., "Organizing for Information Governance"
Debra Logan, November 2009

What Makes the IBM Information Governance Approach Different?

- **Applied to Enterprise Information Supply Chains - not just another "process"**
  - Information Supply Chains represent the flow of information throughout an organization, much like goods through a physical supply chain
- **Validated by the Information Governance Council (top global companies, business partners and industry experts)**
- **Accelerate deployment with the Information Governance Maturity Model**
  - A framework for prioritizing actions, a starting point, a common language and a method to measure progress



Information Governance optimizes the information supply chain: Supporting compliance initiatives, reducing cost, minimizing risk and sustaining profitable growth.

InfoSphere: Collaborative Information Governance

Reusability and consistency

- Shared metadata and policies

Breadth of portfolio

- Three core information governance disciplines

Modular deployment entry points

- Supports business and IT priorities

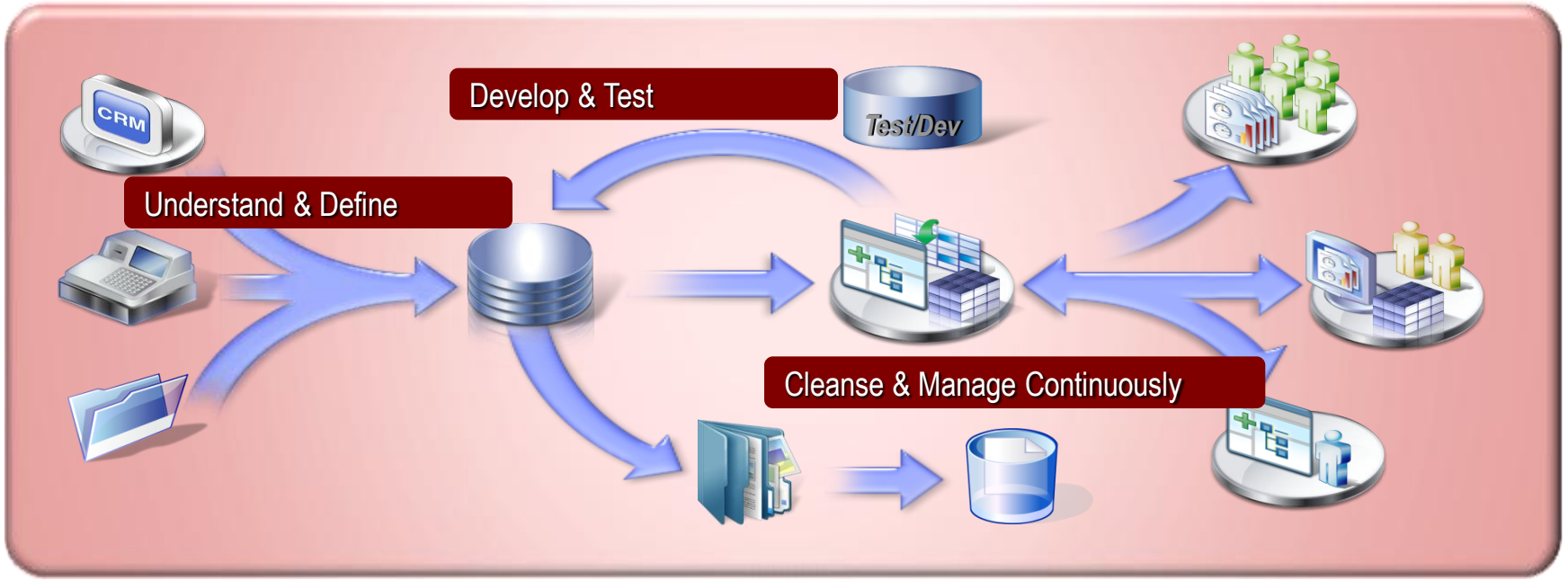
Flexible support for enterprise environments

- Open technology for heterogeneous support

Single solution provider to Optimize the Information Supply Chain

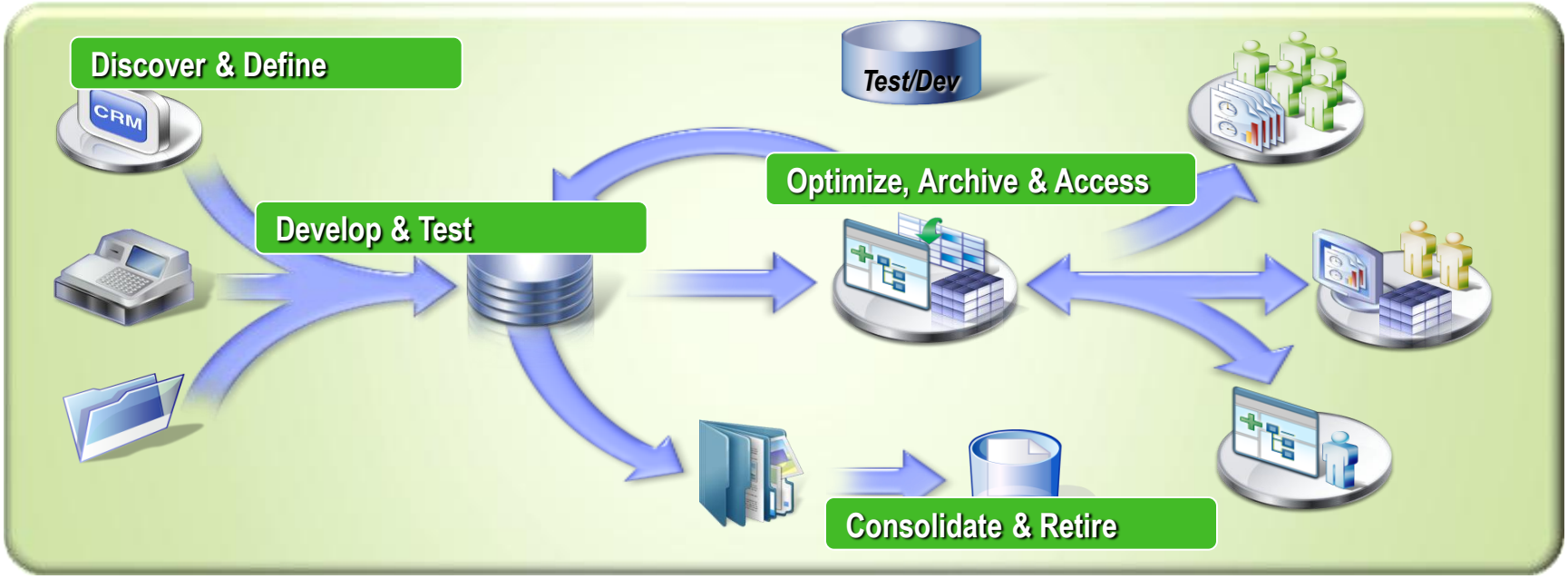
Managing & Maintaining Data Quality in the Information Supply Chain

- Define what 'high quality' data means within your business
- Develop realistic and reusable test data which adheres to privacy policies
- Standardize and cleanse data to enable strategic initiatives
- Constantly monitor for data quality issues and understand how to deal with an exception when it occurs
- Understand the lineage of your data



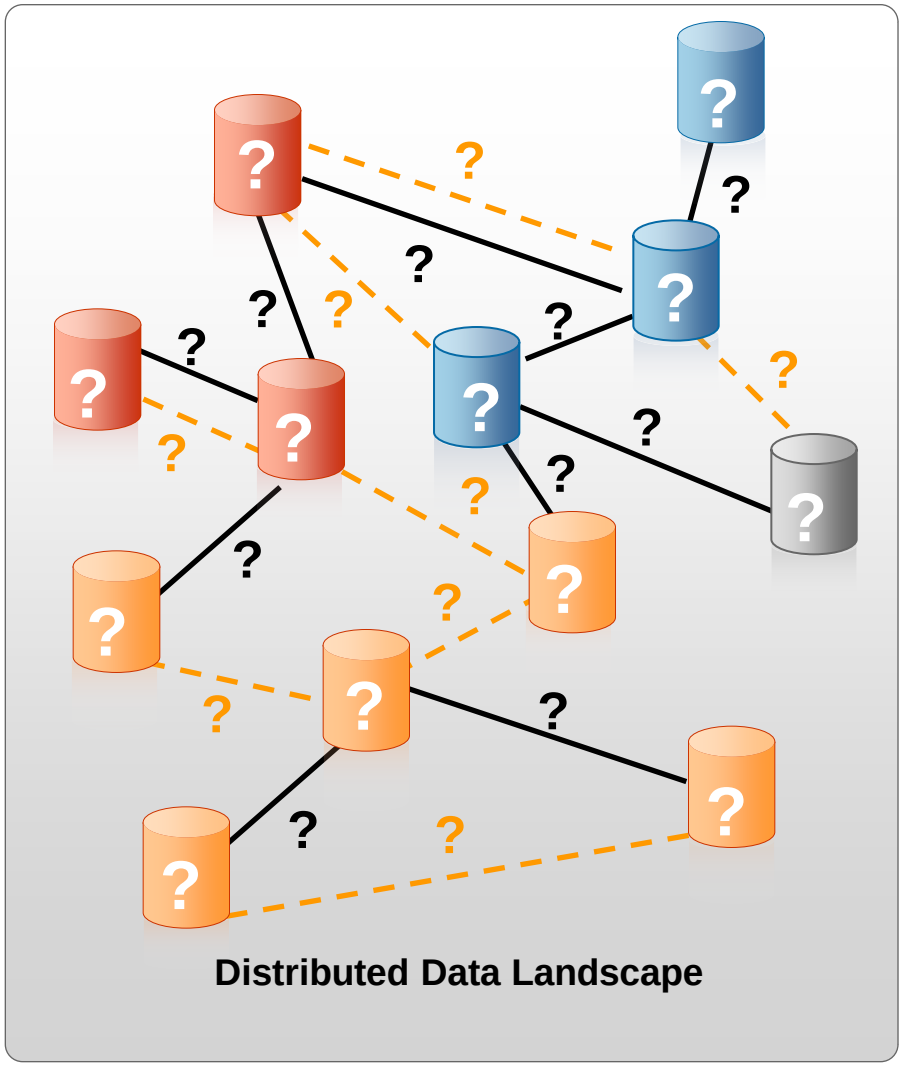
Managing the Lifecycle of Data in the Information Supply Chain

- Understanding the “what & where” of enterprise data
- Developing models and code to store and access enterprise data including the configuration of data for test environments
- Optimizing performance through identification of bottlenecks and building the right strategy for managing applications and data growth
- Implementing a consistent process for retiring or consolidating applications



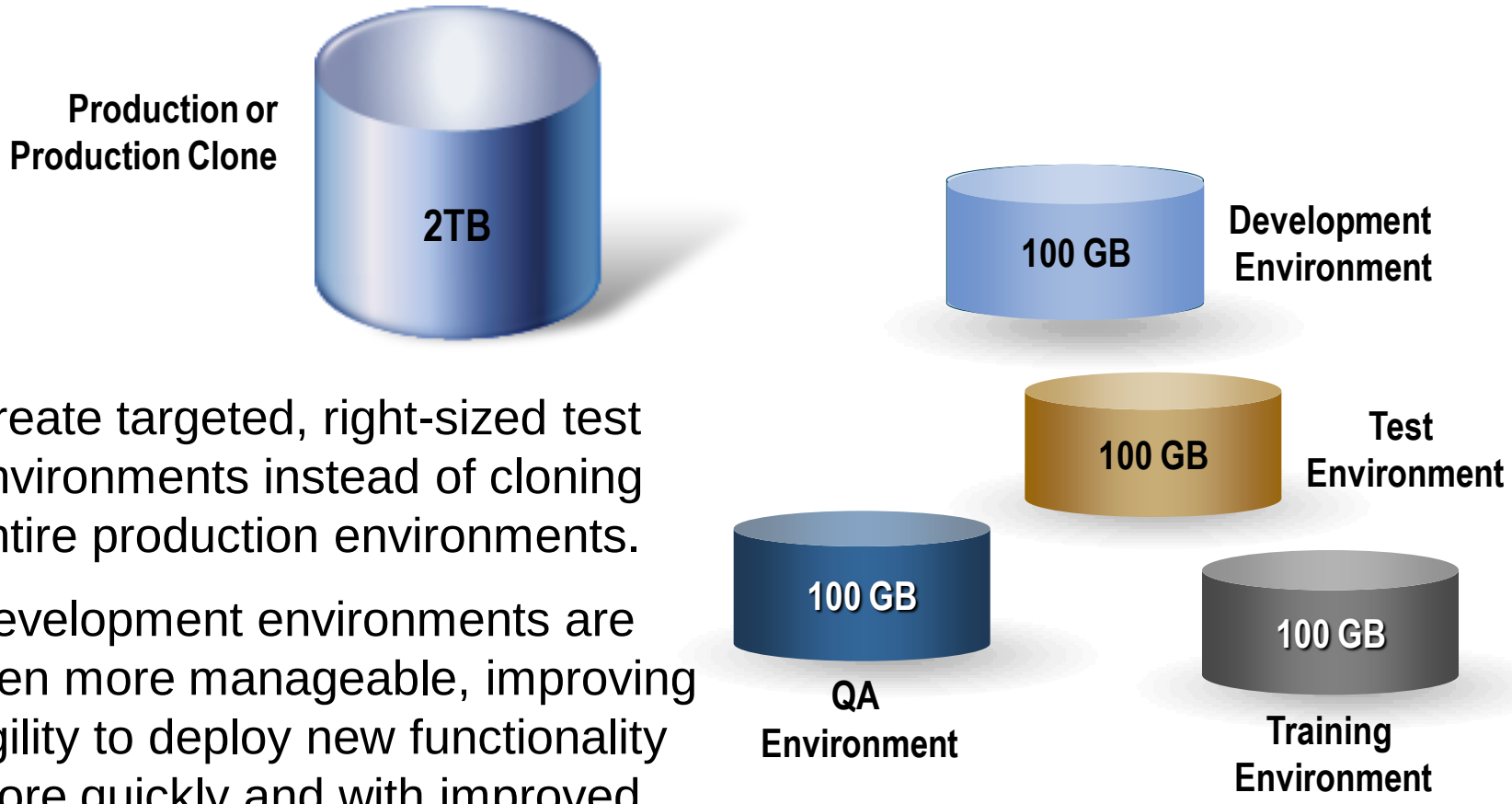
Discover and Classify

You can't manage what you don't understand



- **Data is highly distributed over multiple applications, databases and platforms**
- **Challenging to find all forms of sensitive data**
- **Complex, poorly documented data relationships**
- **Relationships not understood because:**
  - Corporate memory is poor
  - Documentation is poor or nonexistent
  - Logical relationships (enforced through application logic or business rules) are hidden

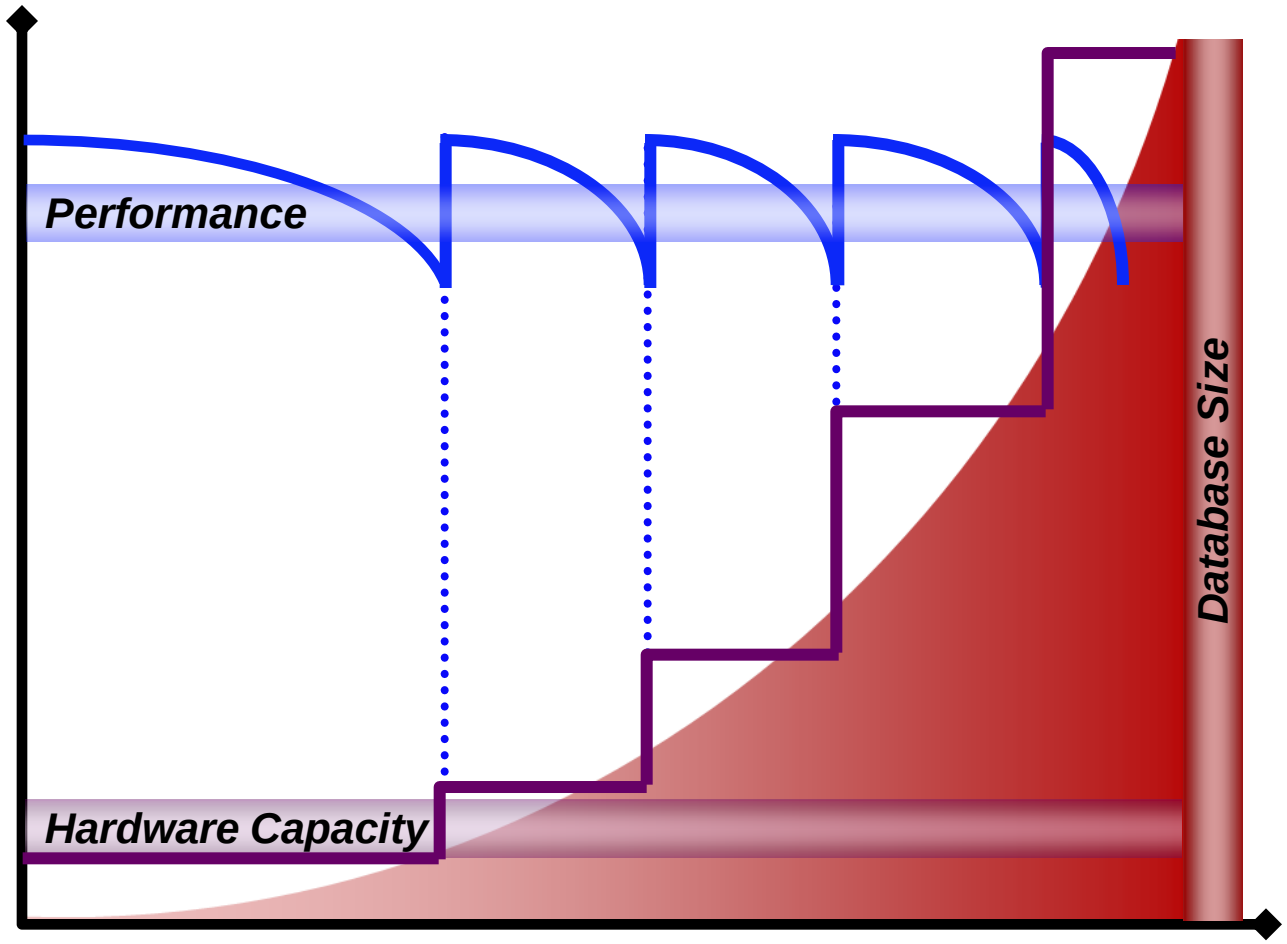
Effective Test Data Management



Create targeted, right-sized test environments instead of cloning entire production environments.

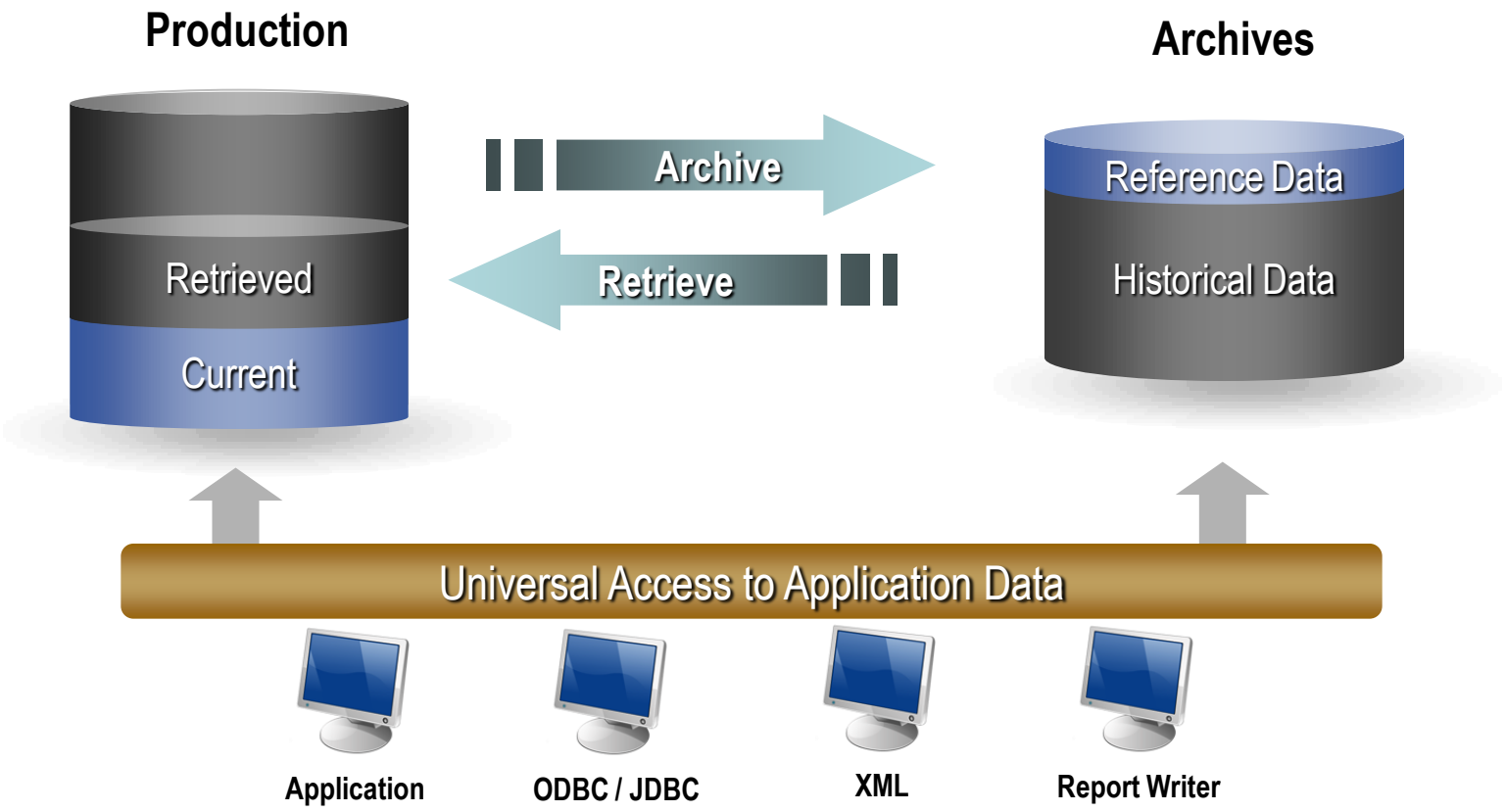
Development environments are then more manageable, improving agility to deploy new functionality more quickly and with improved quality.

How are Organizations Responding to Data Growth?



- Use database partitioning
- Use database vendor compression
- Buy more Storage & CPU hardware

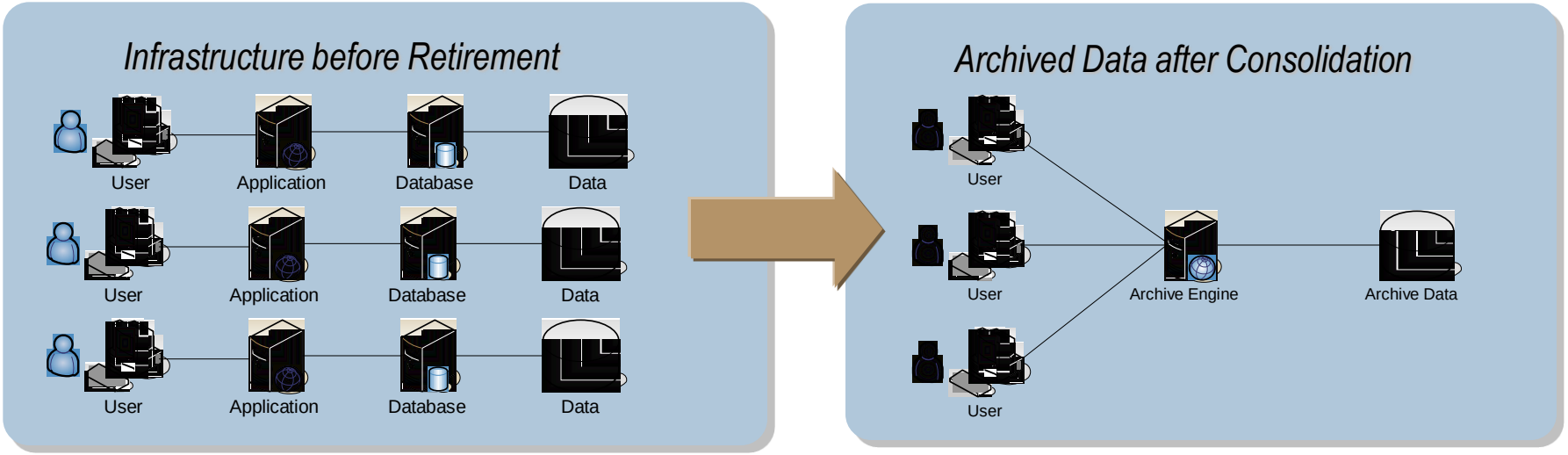
Archiving for Data Growth Management



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

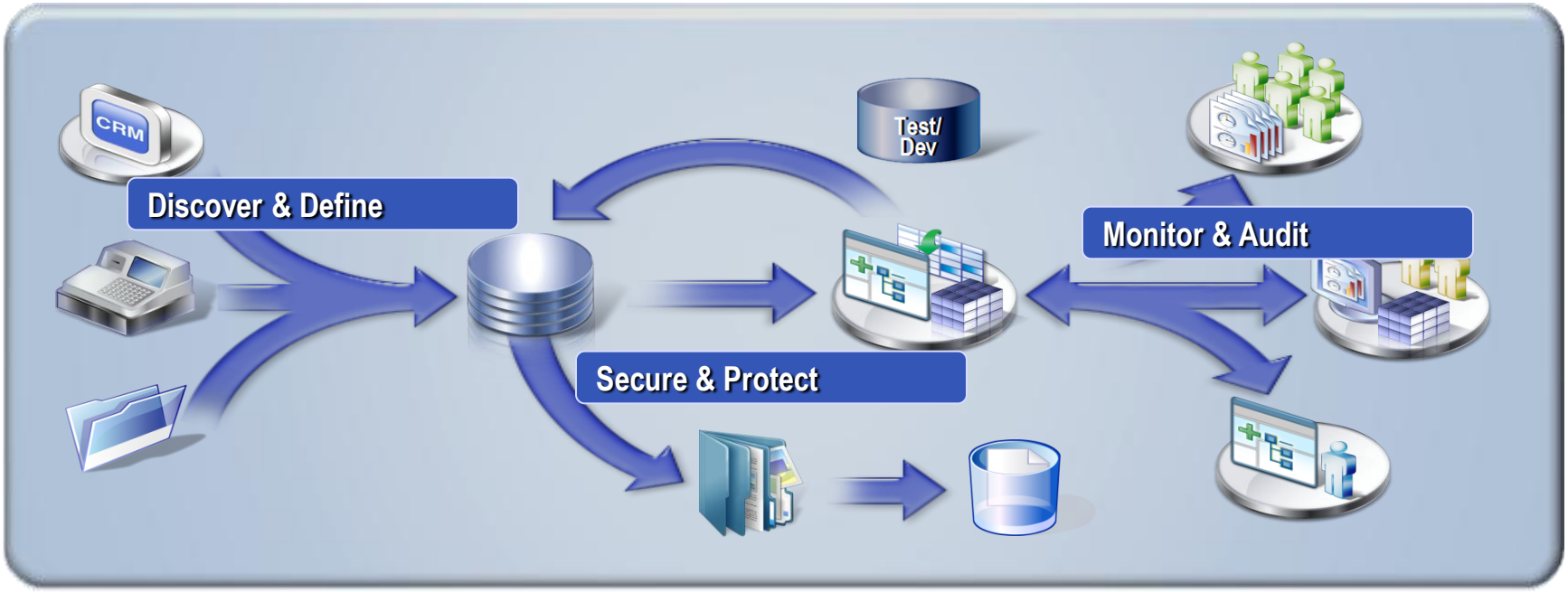
Retire redundant and legacy applications

- **Preserve application data in its business context**
  - Capture all related data, including transaction details, reference data & associated metadata
  - Capture any related reference data - may reside in other application databases
- **Retire out-of-date packaged applications as well as legacy custom applications**
  - Leverage out-of-box support of packaged applications to quickly identify & extract the complete business object
- **Shut down legacy system without a replacement**
  - Provide fast and easy retrieval of data for research and reporting, as well as audits and e-discovery requests

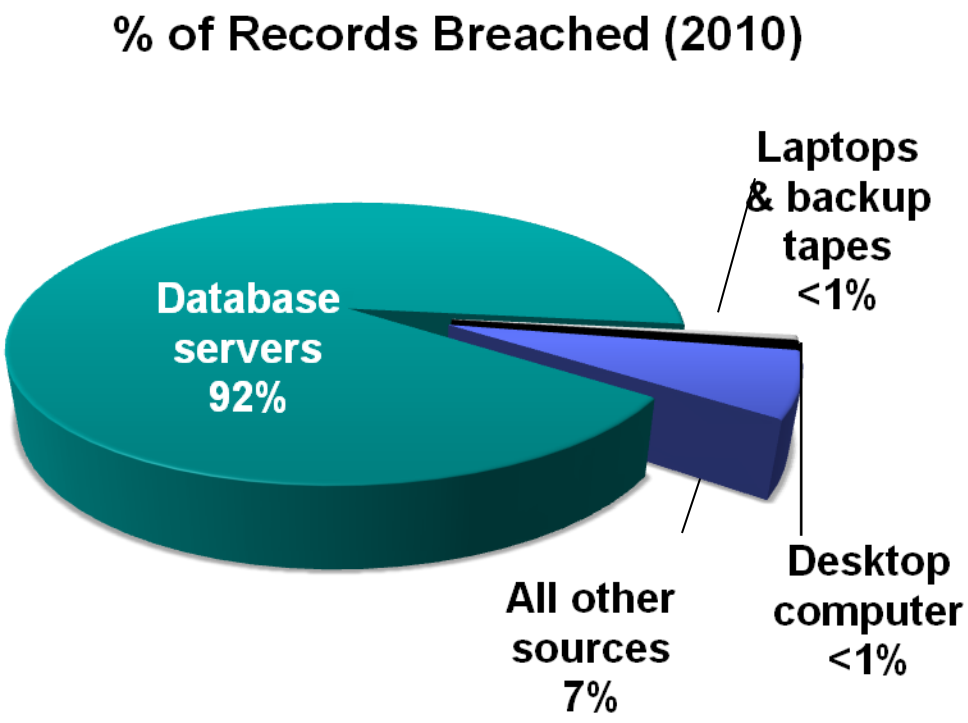


Securing and Protecting Your Information Supply Chain

- Understanding the “what & where” of enterprise data
- Protecting the data across the enterprise, both internal and external threats
- Knowing who’s accessing your data when, how and why
- Monitoring and reporting on database access for audit purposes



Database Servers
The Primary Source of Breached Data



…up from 75% in 2009

"Although much angst and security funding is given to **offline data, mobile devices, and end-user systems,** these assets **are simply not a major point of compromise."**

- 2009 Data Breach Investigations Report

Source: http://www.verizonbusiness.com/resources/security/reports/2009_databreach_rp.pdf Note: multi-vector breaches counted in multiple categories

Why?

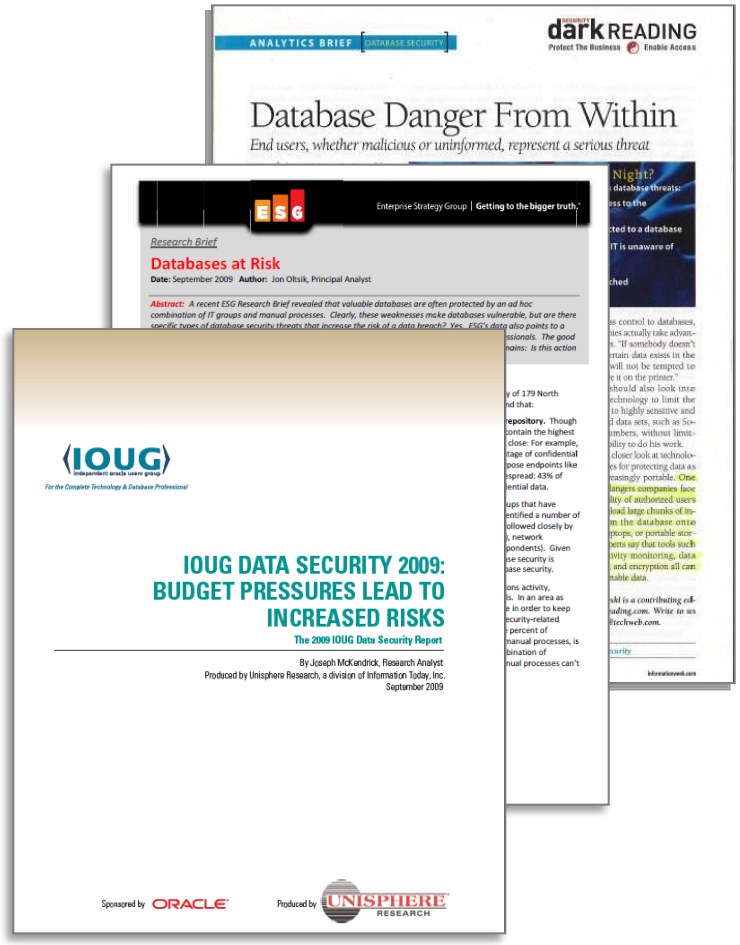
- **Database servers contain your most valuable information**
  - Financial records
  - Customer information
  - Credit card and other account records
  - Personally identifiable information
- **High volumes of structured data**
- **Easy to access**

"Because that's where the money is."

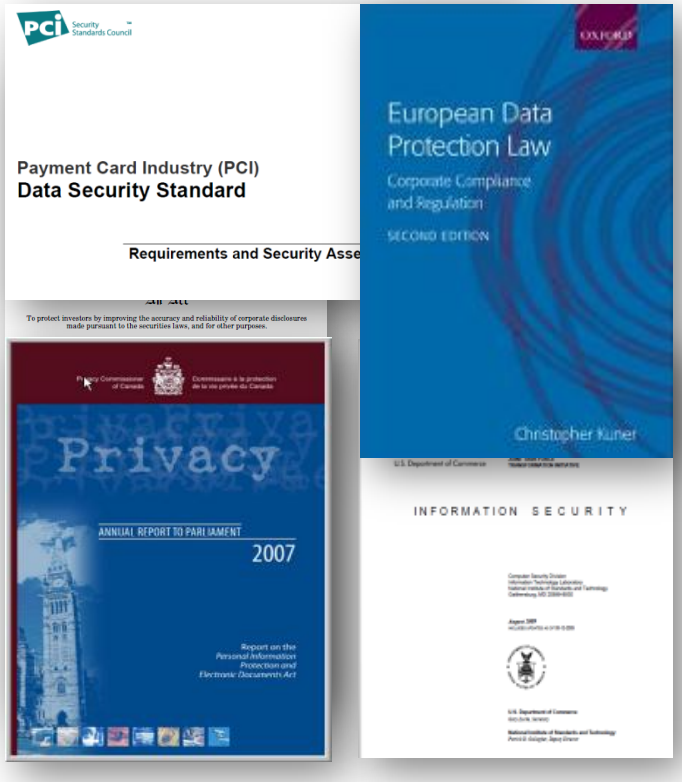
- Willie Sutton

Database Danger from Within

- “Organizations overlook the most imminent threat to their databases: authorized users.” (Dark Reading)
- “No one group seems to own database security … This is not a recipe for strong database security” … 63% depend primarily on manual processes.” (ESG)
- Most organizations (62%) cannot prevent super users from reading or tampering with sensitive information … most are unable to even detect such incidents … only 1 out of 4 believe their data assets are securely configured (Independent Oracle User Group).



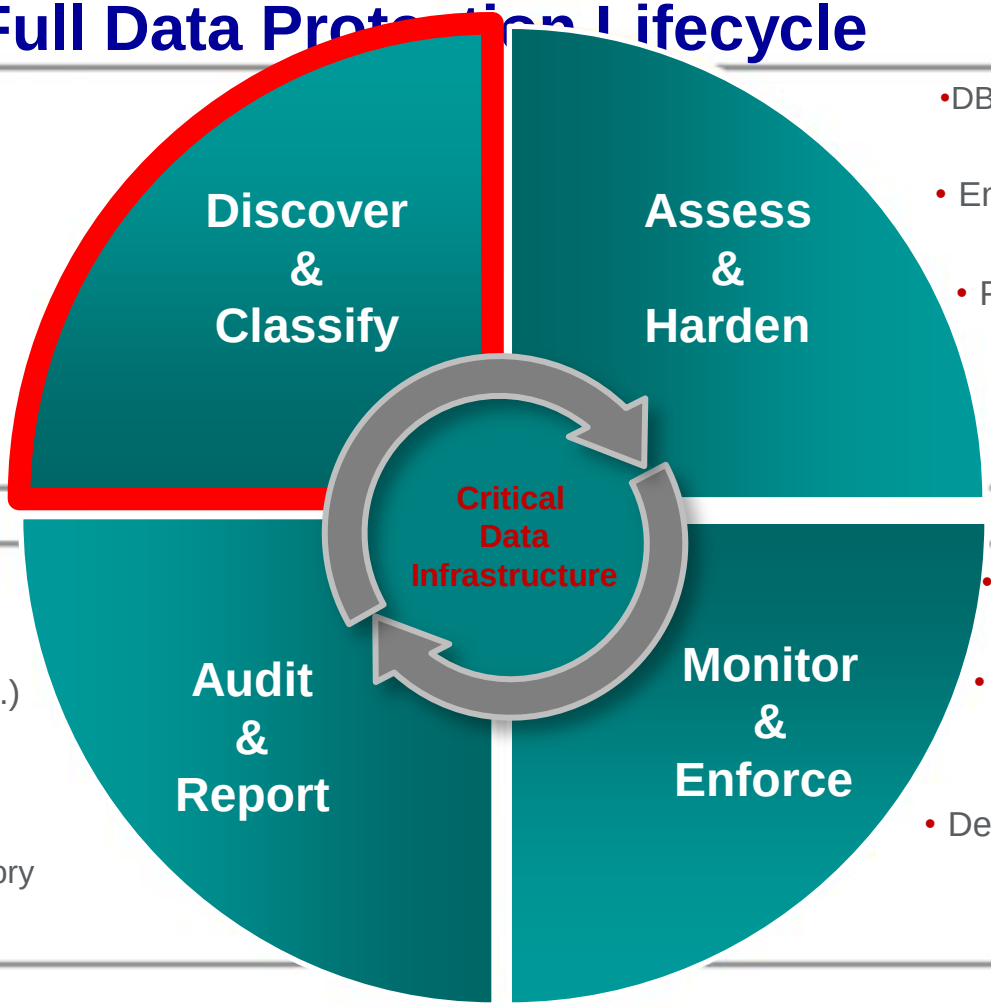
Growing Compliance Mandates



- **Explosion in successful breaches has resulted in growing regulation of sensitive data in North America**
  - SOX
  - HIPAA
  - PCI DSS
  - 46 state-specific data privacy laws
  - Gramm-Leach-Bliley
- **Many EU and Asian countries have enacted similar regulations**
  - EU Data Privacy Directive and supporting local laws
  - C-SOX
  - FIEL
  - PCI DSS
  - etc.

Address the Full Data Protection Lifecycle

- Discover your DBMSs
- Discover & classify sensitive data
- Continuously update security policies



- DB vulnerability assessments
- Masking sensitive data
- Encryption of sensitive data
- Archive un-needed data
- Preconfigured tests based on best practices and standards

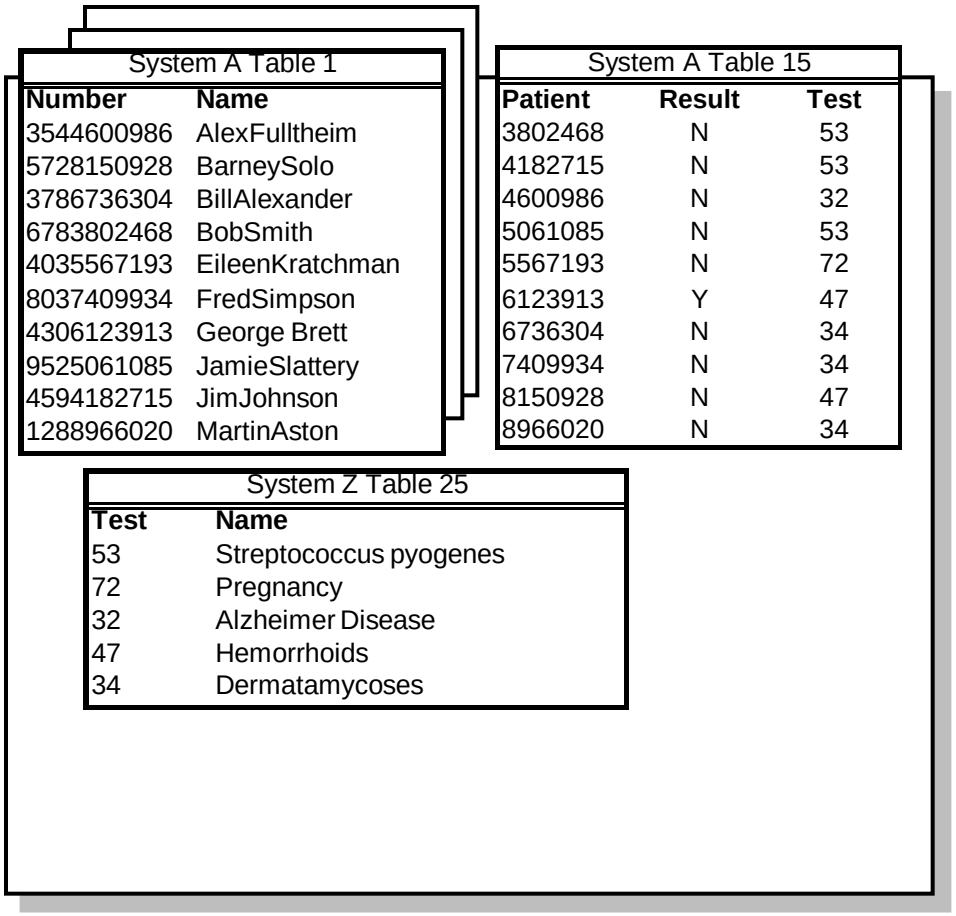
- Cross-DBMS policies
- Pre-built compliance reports (SOX, PCI, etc.)
- Enterprise integration
- SIEM integration
- Sign-off management
- Centralized audit repository
- No database changes

- Monitor & alert on attacks
- Monitor privileged users
- Monitor changed behavior
- Real-time alerts
- Prevent cyberattacks
- Detect application-layer fraud
- Enforce change controls
- Forensics data mining

Sensitive Data Discovery - Infosphere Discovery

- **Common PII data element discovery**
  - Pre-Defined Scanning
- **Custom sensitive data discovery**
  - Supply Discovery with “descriptions/examples”
    - *Patterns*
    - *Data examples*.
  - Discovery will scan for matching columns
- **Hidden sensitive data discovery**
  - Sensitive data embedded in free text columns
    - Scan by “floating” patterns
  - Sensitive data that is partial or hidden
    - Use Transformation Discovery to find data that are “transformed”

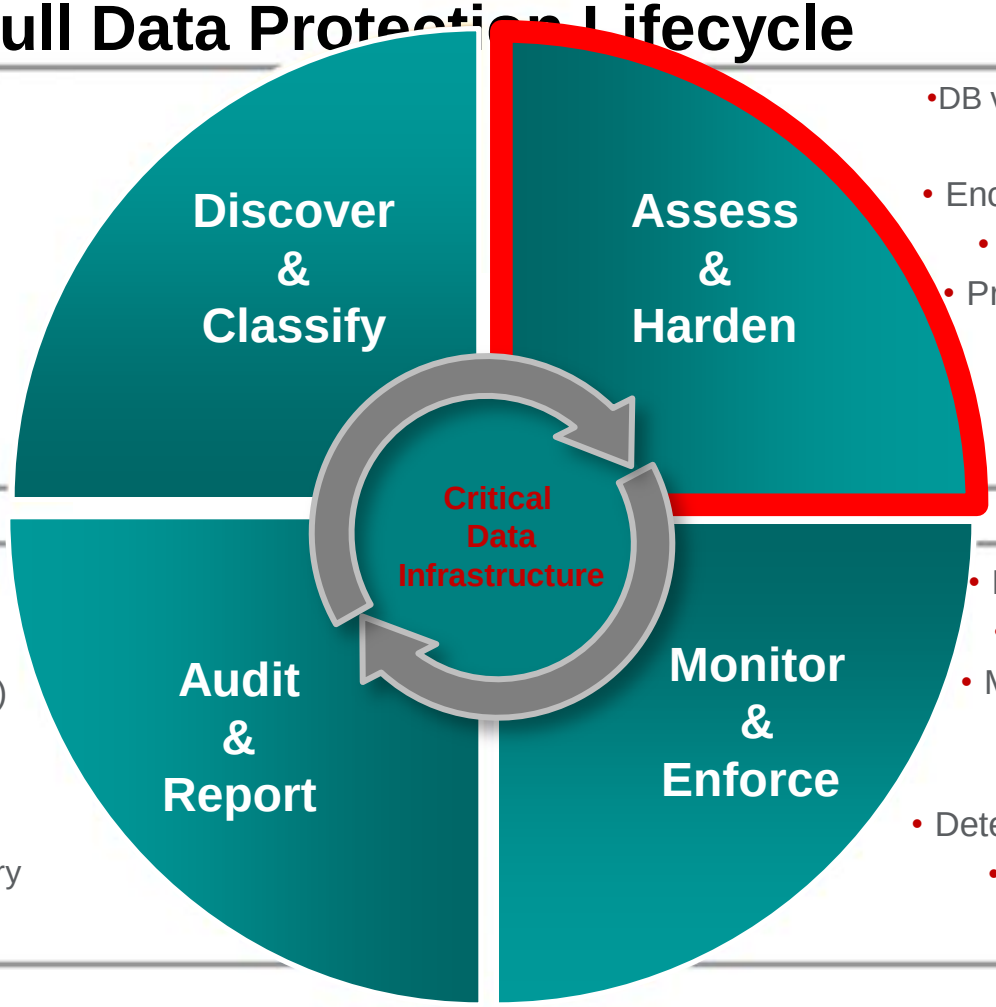
Why is Sensitive Data Discovery so Data Difficult?



- **Sensitive data can't be found just by a simple data scan.**
  - Must connect tables and lookup tables
  - Hidden within larger fields (substring)
  - Hidden across fields (concatenations)
  - Represented differently (lookup tables and case statements)
- **"Corporate memory" is poor**
  - Documentation is incomplete
  - SME's and Data Analysts are only knowledgeable of one or two systems
- **Hundreds of tables and millions of rows:**
  - Complex
  - Difficult to verify
- **Data quality problems make discovery even more difficult**

Address the Full Data Protection Lifecycle

- Discover your DBMSs
- Discover & classify sensitive data
- Continuously update security policies



- DB vulnerability assessments
- Masking sensitive data
- Encryption of sensitive data
- Archive un-needed data
- Preconfigured tests based on best practices and standards

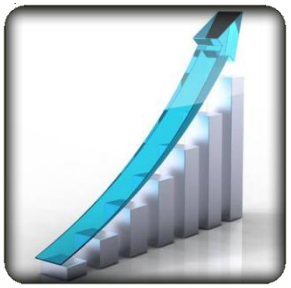
- Cross-DBMS policies
- Pre-built compliance reports (SOX, PCI, etc.)
- Enterprise integration
- SIEM integration
- Sign-off management
- Centralized audit repository
- No database changes

- Monitor & alert on attacks
- Monitor privileged users
- Monitor changed behavior
- Real-time alerts
- Prevent cyberattacks
- Detect application-layer fraud
- Enforce change controls
- Forensics data mining

Assess and Harden - Guardium Vulnerability Assessment

Based on best practices

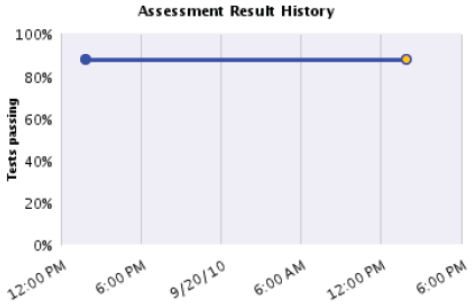
- **New capability enabling customers to cost effectively improve the security of mainframe environments by conducting automated database vulnerability assessment tests**
  - Packaged tests to detect vulnerabilities including inappropriate privileges, grants, default accounts, etc..
  - Capabilities enabling the development of custom tests
- **Based on industry standards such as STIG and CIS**
- **Management of mainframe VA testing from central InfoSphere Guardium console for enterprise-wide control**
  - Configuration and scheduling of mainframe tests
- **Integrated with other InfoSphere Guardium elements for improved process efficiency, including Compliance Workflow Automation and audit repository**



IBM® InfoSphere™ Guardium®

Results for Security Assessment: **VA test for system Z**

Assessment executed **2010-09-20 13:55:27.0**
From: **2010-09-19 13:55:27.0**
To: **2010-09-20 13:55:27.0**
Client IP or IP subnet: **Any**
Server IP or IP subnet: **Any**



Tests passing: **88%***
*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.

Result Summary — Showing 73 of 73 results (0 filtered)

	Critical		Major		Minor		Caution		Info	
Privilege	46p	4f	7p	1f	–	–	–	–	–	–
Authentication	–	–	–	–	–	–	–	–	–	–
Configuration	1p	–	–	–	–	–	–	–	–	–
Version	–	–	1p	–	–	–	–	–	–	–
Other	1p	–	3p	2f	2p	1f	–	–	3p	1f

Current filtering applied:
Test Severities: - Show All -
Datasource Severities: - Show All -
Scores: - Show All -
Types: - Show All -

Assessment Test Results — Showing 73 of 73 results (0 filtered)

Test / Datasource	Result
z/OS Grant option - Resauth Test category: **Priv.** Severity: **Critical** This test check for privileges on various resources that has been granted with the grant option. These resource include: Buffer pool, Collection, Distinct type, Table space, Storage group and JAR file. Grant option is not a good practice and should be avoid where possible. When privileges are granted with the grant option, a user can grant privileges on that resource to other users. We do not recommend granting resource privilege with grant option. This test exclude grantee who is a member of SYSADM and SYSIBM user. Ext. Reference: Guardium, Test ID 2179 **System Z Datasource** Datasource type: **DB2** Severity: **None** Details: Grantee causing failure: Grantee=ADMIN_A: Obtype=D: Qualifier=GU0003: Name=CANADIAN_DOLLAR Grantee=ADMIN_A: Obtype=D: Qualifier=GU0002: Name=CANADIAN_DOLLAR	**Fail** One or more resources privileges has been granted with the grant option. **Recommendation:** We recommend that you revoke resources privileges granted with the grant option. Please redo your resource privilege so that you are using grant instead of grant option. If you need to exclude certain grantee or resource that must have grant option, you can create a group then populate it with authorize grantee and or resource name and link your group to this test.
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	**Fail** One or more object privileges has been granted with the grant option. **Recommendation:** We recommend that you revoke schema privileges granted with the grant option. Please redo your schema privilege so that you are using grant instead of grant option. If you need to exclude certain grantee or objects that must have grant option, you can create a group then populate it with authorize grantee and or objects name and link your group to this test.

System Z Vulnerability Assessment - Remediation

IBM InfoSphere Guardium: Security Assessment Results - Mozilla Firefox

https://9.70.147.47:8443/saResultsViewer.do?method=view&viewerType=testResultDetails&viewedTaskId=-1&selectedTestResultId=20158&saResultId=20001

IBM® InfoSphere™ Guardium®

Results for Security Assessment: **VA test for system Z**

Assessment executed **2010-09-20 13:55:27.0**
From: **2010-09-19 13:55:27.0**
To: **2010-09-20 13:55:27.0**
Client IP or IP subnet: **Any**
Server IP or IP subnet: **Any**

z/OS Grant option - Resauth
Test category: **Priv.** Test severity: **Critical**

System Z Datasource
Datasource type: **DB2** Datasource severity: **None**

Fail

One or more resources privileges has been granted with the grant option.

Short Description: This test check for privileges on various resources that has been granted with the grant option. These resource include: Buffer pool, Collection, Distinct type, Table space, Storage group and JAR file. Grant option is not a good practice and should be avoid where possible. When privileges are granted with the grant option, a user can grant privileges on that resource to other users. We do not recommend granting resource privilege with grant option. This test exclude grantee who is a member of SYSADM and SYSIBM user.

External Reference: Guardium, Test ID 2179

Recommendation: *We recommend that you revoke resources privileges granted with the grant option. Please redo your resource privilege so that you are using grant instead of grant option. If you need to exclude certain grantee or resource that must have grant option, you can create a group then populate it with authorize grantee and or resource name and link your group to this test.*

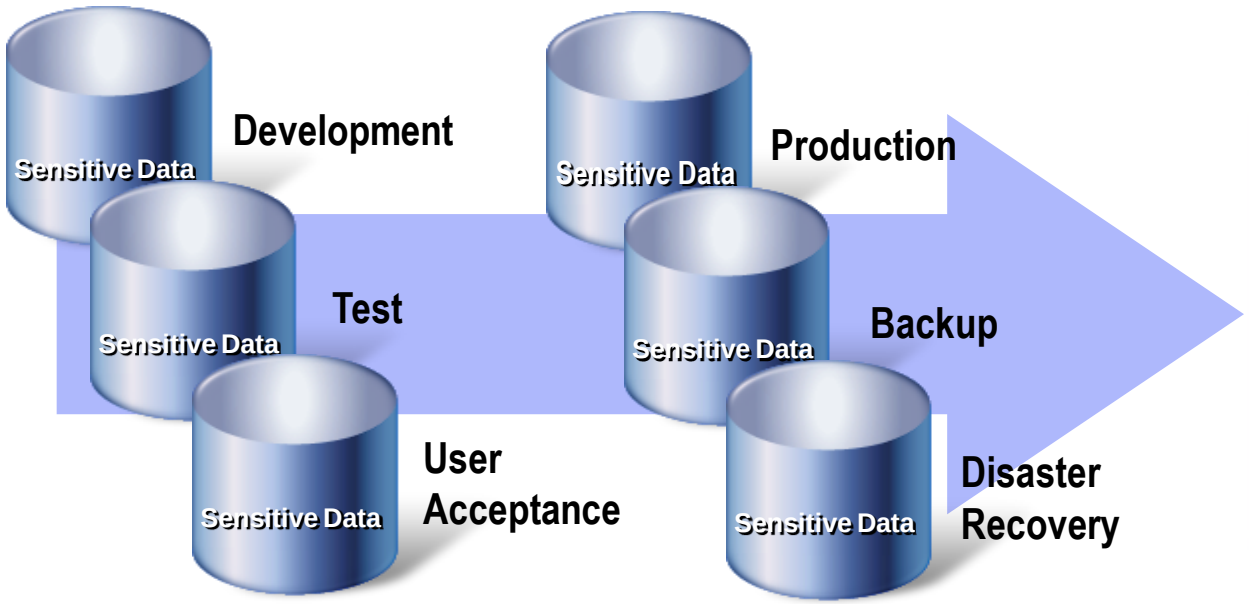
Details: Grantee causing failure: Grantee=ADMIN_A: Obtype=D: Qualifier=GU0003: Name=CANADIAN_DOLLAR
Grantee=ADMIN_A: Obtype=D: Qualifier=GU0002: Name=CANADIAN_DOLLAR

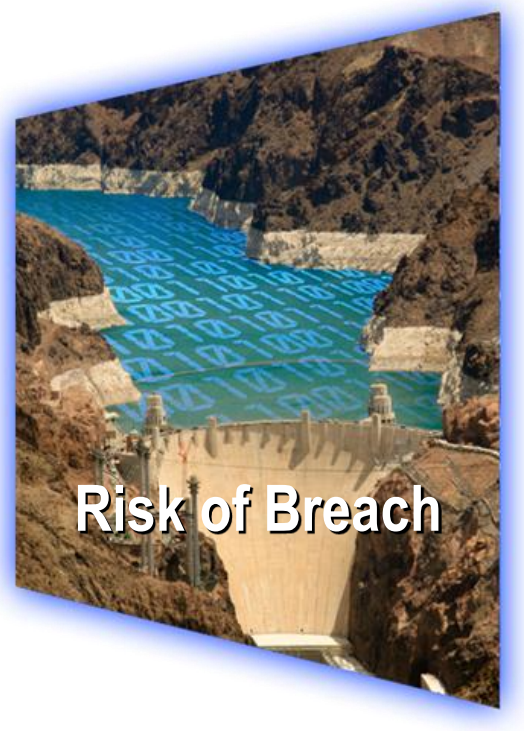
Close this window

Done

Assess and Harden

Sensitive Data Proliferation





Actual risk and compliance burden =
Original production data + all derived clones

Assess and Harden - Data Masking

Masked or transformed data must be appropriate to the context:

- –Consistent formatting (alpha to alpha)
- –Within permissible range of values
- –Context and application aware
- –Maintain referential integrity

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

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

Example 1

Patient Information

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

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

Example 2

Personal Info Table

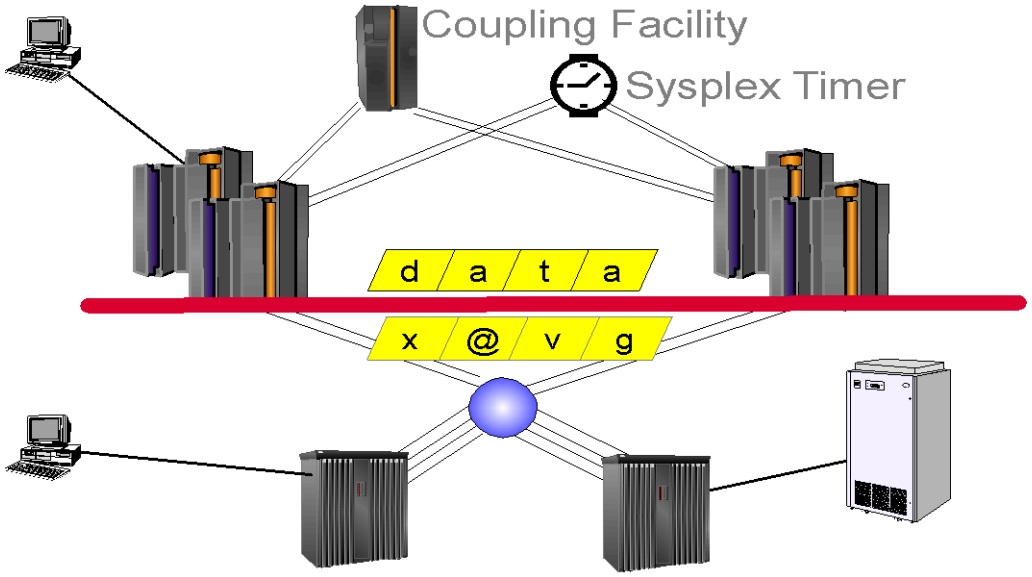
PersNbr	FirstName	LastName
10000	Jeanne	Renoir
10001	Claude	Monet
10002	**Pablo**	**Picasso**

Referential integrity is maintained with key propagation

Event Table

PersNbr	FstNEvtOwn	LstNEvtOwn
10002	**Pablo**	**Picasso**
10002	**Pablo**	**Picasso**

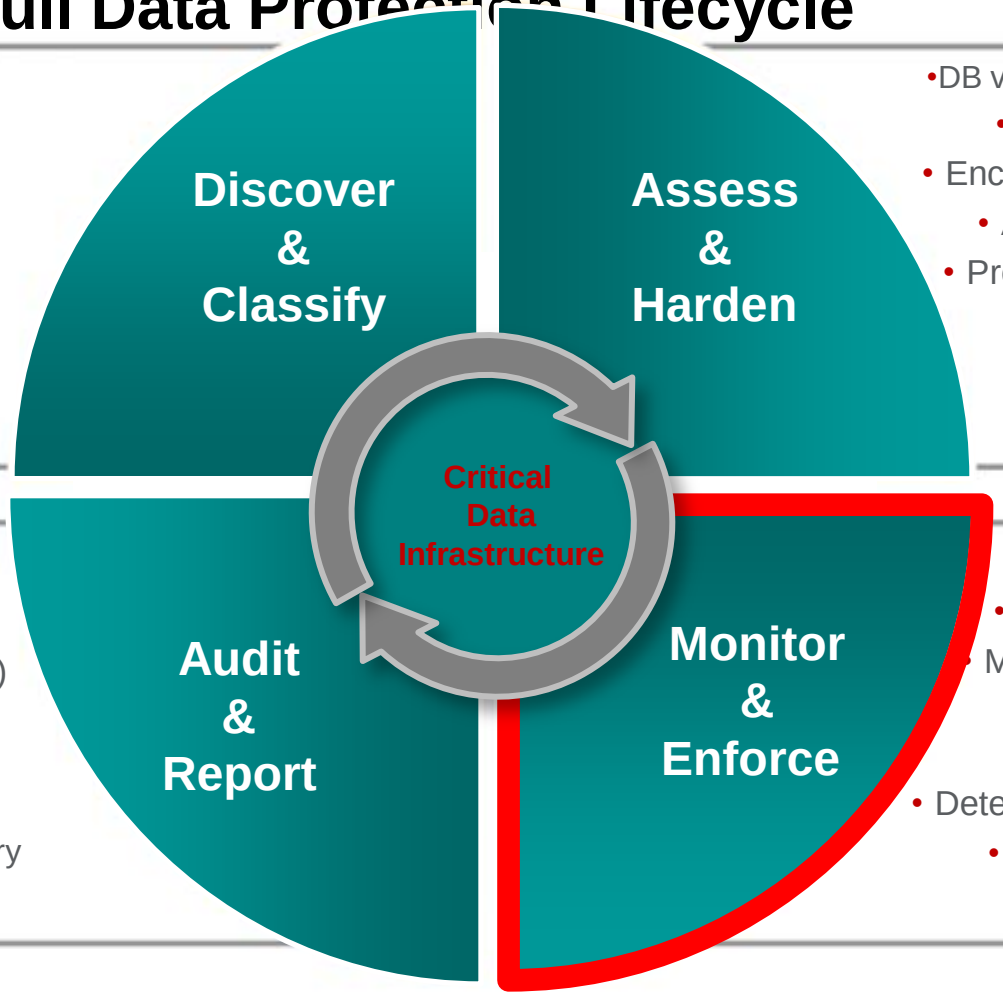
Assess and Harden - Data Encryption for DB2/z and IMS



- **Row/Segment level encryption**
- **Data encrypted on disk and through channel**
- **Pre-coded EDITPROC for DB2 and segment edit routines for IMS**
  - Can be used for multiple tables / segments
  - Fast Implementation
- **Low Performance overhead**
- **Key management – use System/z facilities**
  - No new key management facility learning curve
  - Uses existing ICSF facility to manage encryption keys in one central repository
- **Supports all levels of DB2**
- **No application changes**
  - System administration changes only to the segment or table definition
- **Compatible with DB2 Load/Unload utilities and DB2 Tools**
- **Encryption and compression can be combined**
- **Supports both secure key and clear key encryption**

Address the Full Data Protection Lifecycle

- Discover your DBMSs
- Discover & classify sensitive data
- Continuously update security policies

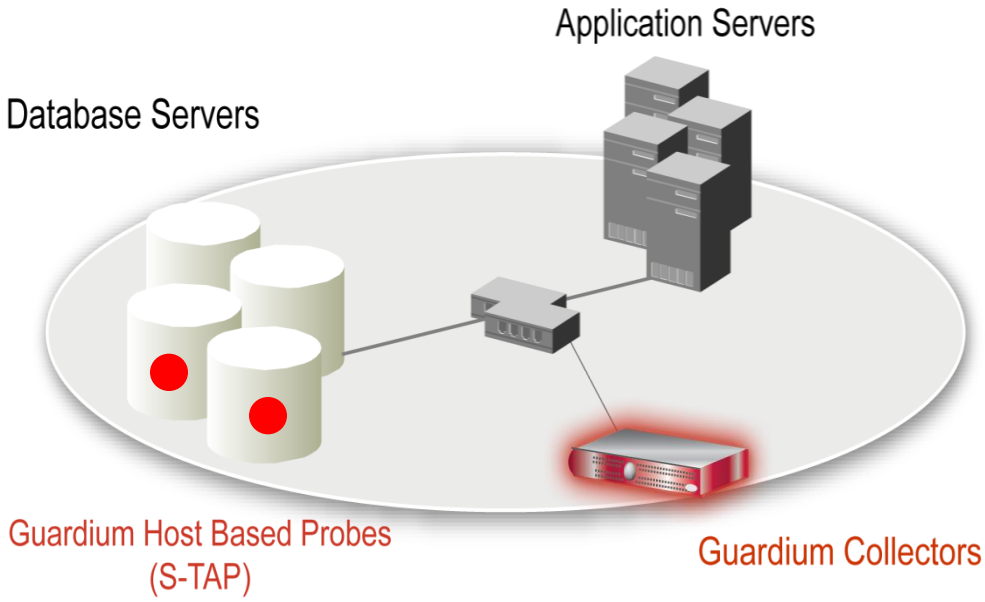


- DB vulnerability assessments
- Masking sensitive data
- Encryption of sensitive data
- Archive un-needed data
- Preconfigured tests based on best practices and standards

- Cross-DBMS policies
- Pre-built compliance reports (SOX, PCI, etc.)
- Enterprise integration
- SIEM integration
- Sign-off management
- Centralized audit repository
- No database changes

- Monitor & alert attacks
- Monitor privileged users
- Monitor changed behavior
- Real-time alerts
- Prevent cyberattacks
- Detect application-layer fraud
- Enforce change controls
- Forensics data mining

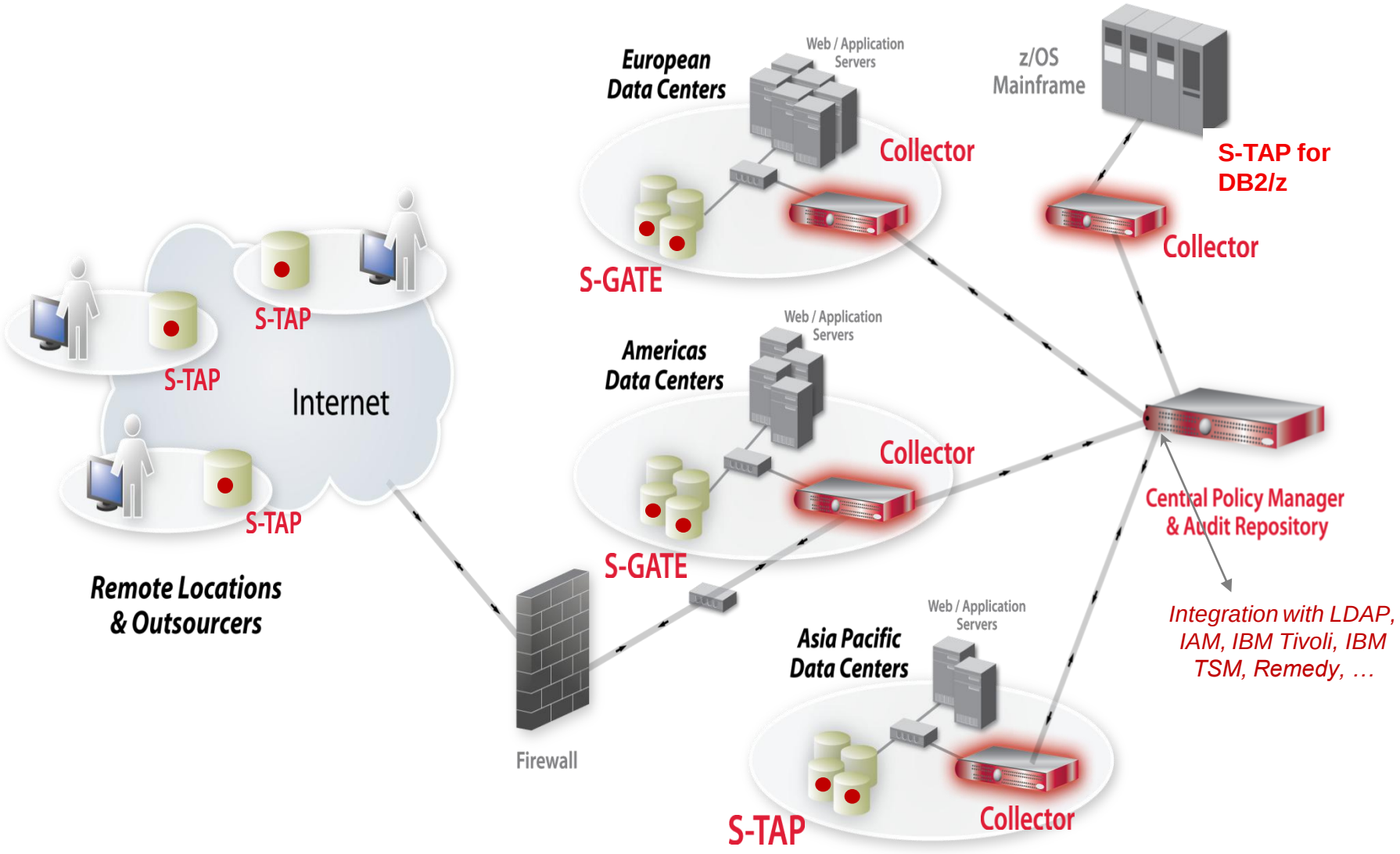
Monitor and Enforce – Database Activity Monitoring



- Non-invasive architecture
  - Outside database
  - Minimal performance impact
  - No DBMS or application changes
- Cross-DBMS solution
- 100% visibility including local DBA access

- Enforces separation of duties
- Does not rely on DBMS-resident logs that can easily be erased by attackers, rogue insiders
- Granular, real-time policies & auditing
  - *Who, what, when, how*
- Automated compliance reporting, sign-offs & escalations (SOX, PCI, NIST, etc.)

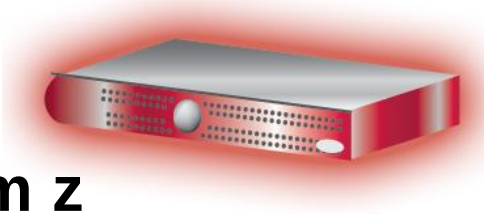
Scalable Multi-Tier Architecture



Guardium for z

- **Provides a unified view and secure audit trail of all database activities**
  - Across *both* mainframe and distributed environments
  - Enterprise-wide compliance reporting, alerting, analytics & forensics
- **Best Database Activity Monitor Technology from Guardium**
  - Leverage all Guardium functionality off host
- **Best DB2/z event capture technology**
  - Lightweight deployment
    - Audited data streamed to Guardium appliance, not stored in DB2/z
  - DB2 trace not used for high volume SQL events
    - Class 4 / Class 5 audit traces NOT used
  - Ongoing performance and collection enhancements
  - Optimal performance for customers using IBM Query Monitor
    - Query Monitoring and Audit requirements leverage a single collector process
- **Exploring support for IMS and VSAM**

Guardium for z - Components

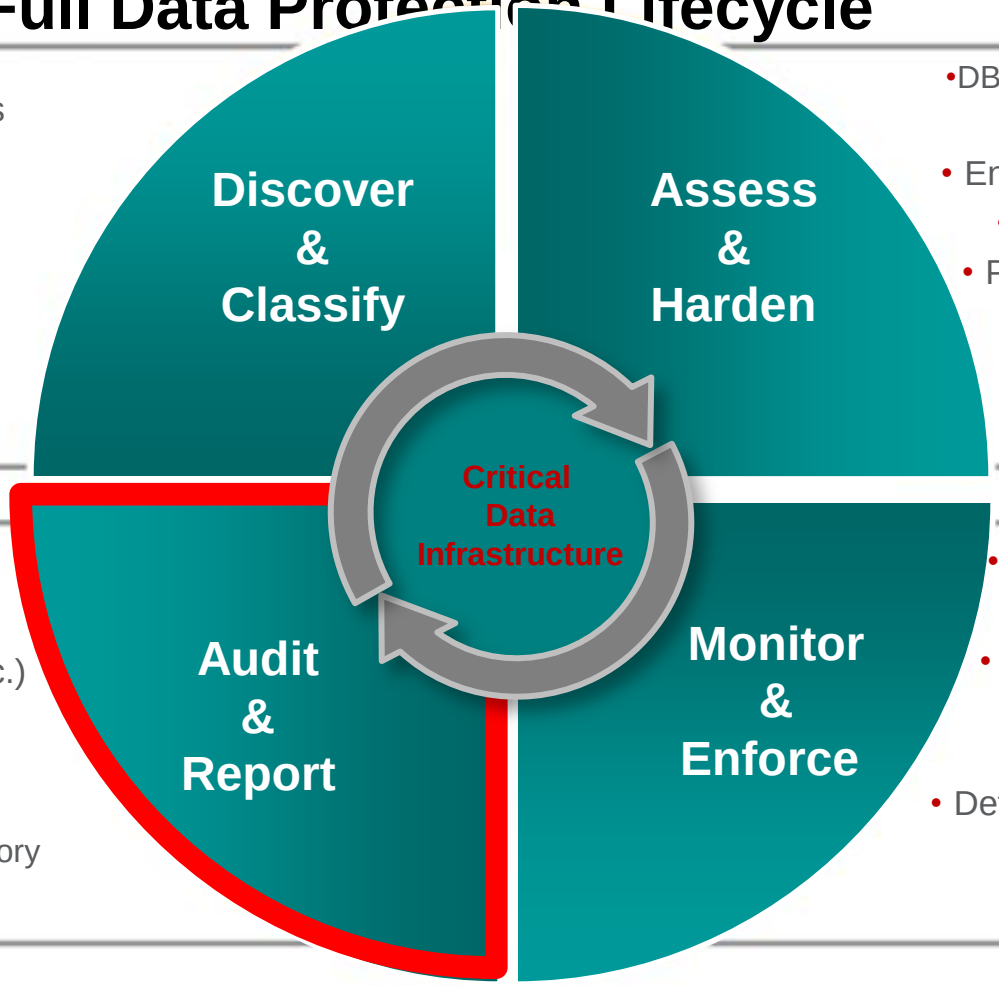


- **Guardium Collector appliance for System z**
  - Securely stores audit data collected by mainframe tap
  - Provides analytics, reporting & compliance workflow automation
  - Integrated with Guardium enterprise architecture
    - Centralized, cross-platform audit repository for enterprise-wide analytics and compliance reporting across mainframe & distributed environments

- **S-TAP for DB2 on z/OS event capture**
  - Mainframe tap
  - Collects audit data for Guardium appliance
  - Leverages existing IBM DB2/z collection technology
  - DB2/z event capture policy editor

Address the Full Data Protection Lifecycle

- Discover your DBMSs
- Discover & classify sensitive data
- Continuously update security policies



- DB vulnerability assessments
- Masking sensitive data
- Encryption of sensitive data
- Archive un-needed data
- Preconfigured tests based on best practices and standards

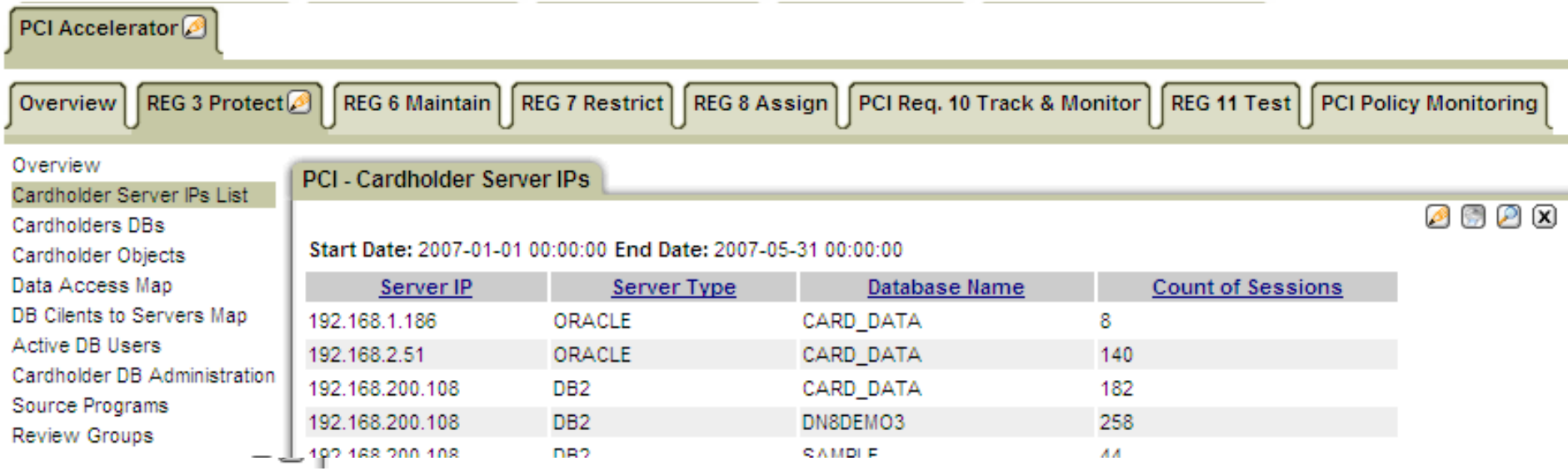
- Cross-DBMS policies
- Pre-built compliance reports (SOX, PCI, etc.)
- Enterprise integration
- SIEM integration
- Sign-off management
- Centralized audit repository
- No database changes

- Monitor & alert on attacks
- Monitor privileged users
- Monitor changed behavior
- Real-time alerts
- Prevent cyberattacks
- Detect application-layer fraud
- Enforce change controls
- Forensics data mining

Audit and Report

Custom and Pre-Built Compliance Reports

- Custom reporting
- SOX and PCI accelerators
  - Financial application monitoring (EBS, JD Edwards, Peoplesoft, etc)
  - Authorized application access only
  - Automated compliance reporting, sign-offs & escalations (SOX, PCI, NIST, etc.)



Disclaimer/Trademarks

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements, or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

The information on the new products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information on the new products is for informational purposes only and may not be incorporated into any contract. The information on the new products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. The development, release, and timing of any features or functionality described for our products remains at our sole discretion.

This information may contain examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious, and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Trademarks The following terms are trademarks or registered trademarks of other companies and have been used in at least one of the pages of the presentation:

The following terms are trademarks of International Business Machines Corporation in the United States, other countries, or both: DB2 Universal Database, eServer, FlashCopy, IBM, IMS, iSeries, Tivoli, z/OS, zSeries, Guardium, IBM Smart Analytics Optimizer, Data Encryption Tool for IMS and DB2 Databases, DB2 Administration Tool / DB2 Object Compare for z/OS, DB2 Audit Management Expert for z/OS, DB2 Automation Tool for z/OS, DB2 Bind Manager for z/OS, DB2 Change Accumulation Tool for z/OS, DB2 Cloning Tool for z/OS, DB2 High Performance Unload for z/OS, DB2 Log Analysis Tool for z/OS, DB2 Object Restore for z/OS, DB2 Path Checker for z/OS, DB2 Query Management Facility for z/OS, DB2 Query Monitor for z/OS, DB2 Recovery Expert for z/OS, DB2 SQL Performance Analyzer for z/OS, DB2 Table Editor for z/OS , DB2 Utilities Enhancement Tool for z/OS, DB2 Utilities Suite for z/OS, InfoSphere Change Data Capture, InfoSphere Data Event Publisher, InfoSphere Replication Server, Optim Data Growth Solution for z/OS, Optim Development Studio, Optim pureQuery Runtime, Optim Query Workload Tuner, Optim Test Data Management Solution for z/OS, Tivoli OMEGAMON XE for DB2 Performance Expert on z/OS
EMC and TimeFinder are trademarks of EMC Corporation
Hitachi is a traademark of Hitchi Ltd

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Other company, product, or service names may be trademarks or service marks of others.

Evolving
IBM Smart Analytics Optimizer

Dan Wardman

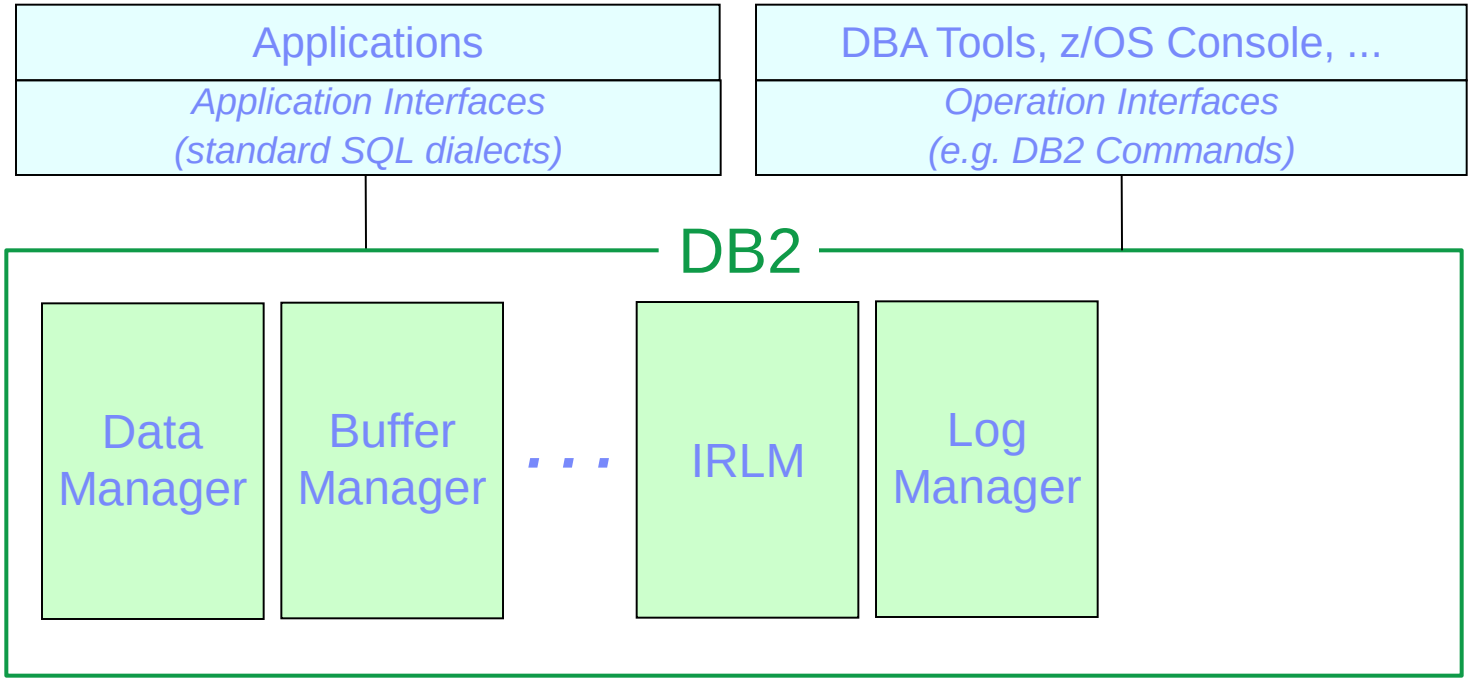
Santa Barbara

August 2011

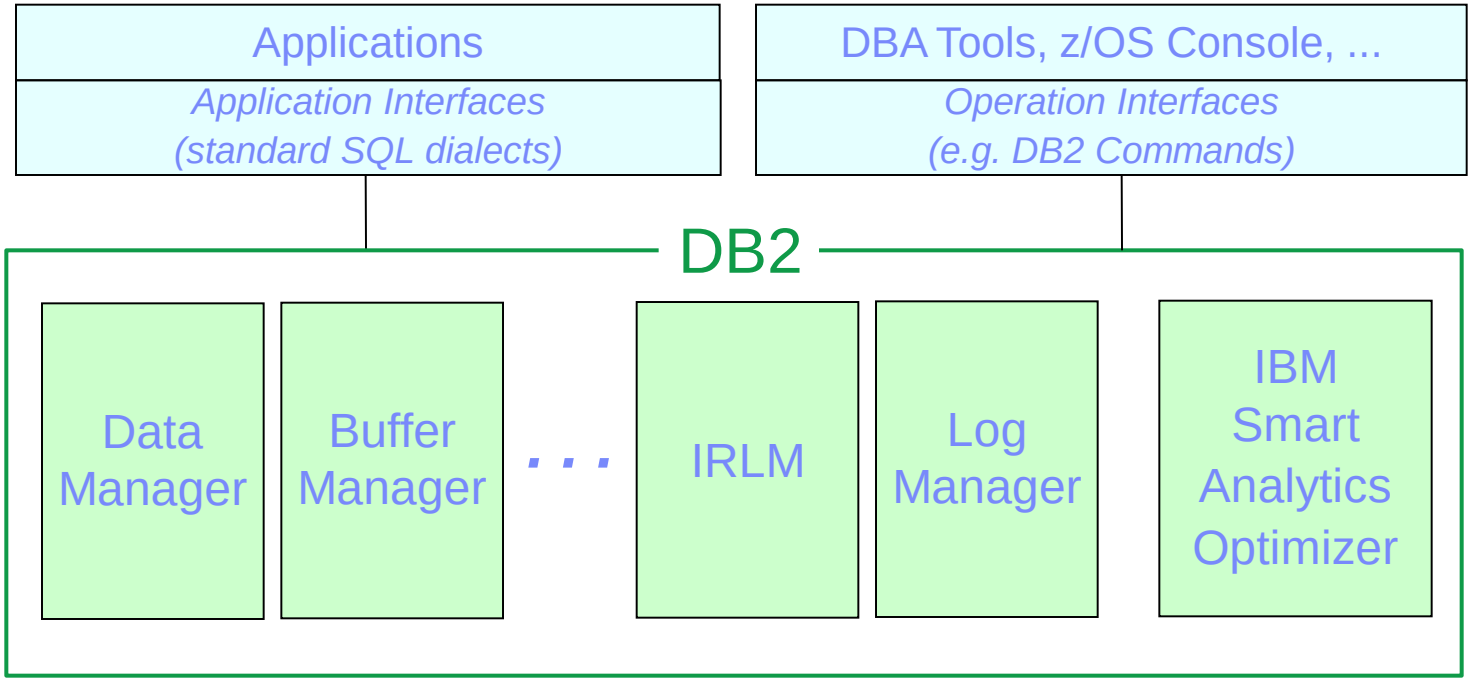
Agenda

- **V1**
  - Architecture
  - Value propositions
- **V1.2**
  - New query engine
  - Preserving value propositions

DB2 Components



IBM Smart Analytics Optimizer as a Virtual DB2 Component



Deep DB2 Integration within zEnterprise

Applications

Application Interfaces (standard SQL dialects)

DBA Tools, z/OS Console, ...

Operation Interfaces (e.g. DB2 Commands)

DB2

Data Manager

Buffer Manager

. . .

IRLM

Log Manager

IBM Smart Analytics Optimizer

Superior availabi
reliability, securi
workload managem

Abundance
of resources

z/OS on System z

10‘s of processors

100‘s GB of memory

Linux on zBX

100‘s of processors

1000‘s GB of memory

Enabling Technology – IBM Research Project BLINK

- **Various Compression Techniques**
  - Enables in-memory database
  - Order-preserving
  - Frequency partitioning
- **Register-store: a combination of row- and column-based stores**
- **Multi-core friendly scans**
  - Massive scale-out parallelism
  - Scans on compressed data
  - Vector processing

ISAOpt V1 Query Off-load Limitations

- **Multiple query blocks are processed one by one resulting in less than optimal performance of:**
  - Subselects in quantitative predicates (SOME, ANY, ALL)
  - EXISTS or IN predicates with subselects
  - UNION, INTERSECT, EXCEPT
- **Outer query block that contains a subselect cannot be accelerated**
  - Only the leaf-blocks in an access path tree apply for acceleration
- **Limited support for very large dimension tables**
  - Especially if the predicates on them are not selective
- **No full outer join, no right outer join**
- **Only equi-joins (no range join predicates)**
- **No queries that do not include at least one fact table**
- **Fact table cannot appear at the null-padding side of the outer join**
- **No queries that spread across multiple marts**
- **No static SQL**
- **No support for Unicode**
- **Not all DB2 data types such as LOBs, ROWID, XML.**
- **Not all DB2 functions**
  - No mathematical functions such as SIN, COS, TAN.

ISAOpt V1 Needs Following Enhancements

1. **Increase applicability by relaxing current off-load restrictions**
2. **Increase applicability by supporting larger amount of data**
3. **Support concurrent query execution**
4. **Improve data currency**
5. **Support disaster recovery**
6. **DB2 10 support**

Agenda

- V1
  - Architecture
  - Value Propositions
- **V1.2**
  - New query engine
  - Preserving value propositions

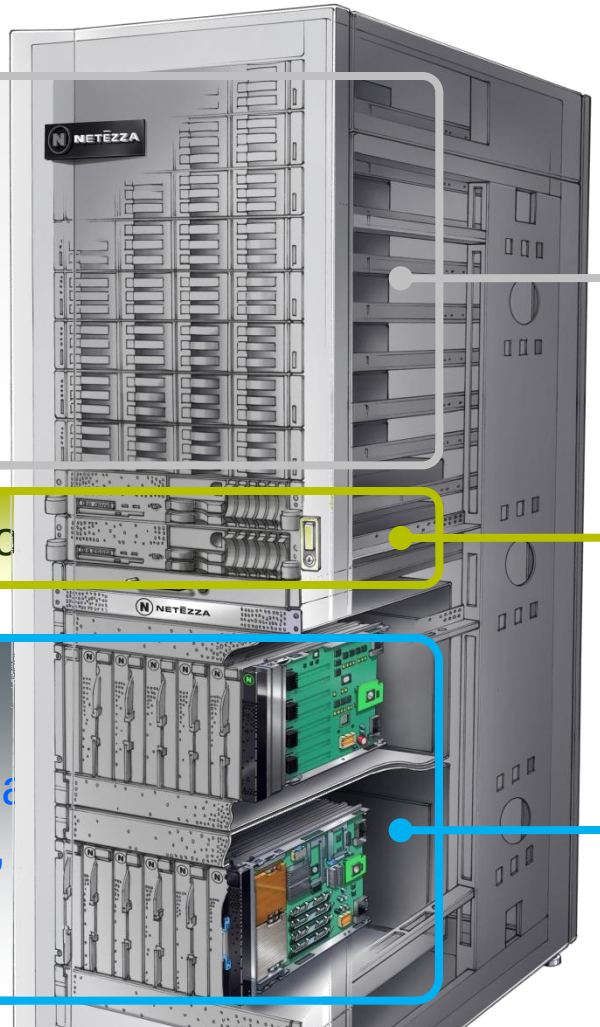
Enter Netezza

ISAO V1.2 Powered by Netezza Twinfin™ Appliance

Disk Enclosures

SMP Ho

Snippet Bla
(S-Blades,



Slice of User Data
Swap and Mirror partitions
High speed data streaming
High compression rate

EXP3000 JBOD Enclosur

12 x 3.5” 1TB, 7200RPM

max 116MB/s (200-500

ISAO Server
SQL Compiler, Query Plan, Optimize
Administration

e.g. TF12:

8 enclosures → 96 HD

2 front/end hosts, IBM 36

32TB uncompressed us

clustered active-passive

Processor & streaming Snippet
High-performance database engine streaming joins, aggregations, sorts, etc.

2 Nehalem-EP Quad-co

e.g. TF12: 12 back/end S

(more details on followi

Deep DB2 Integration

Applications

Application Interfaces (standard SQL dialects)

DBA Tools, z/OS Console, ...

Operation Interfaces (e.g. DB2 Commands)

DB2

Data Manager

Buffer Manager

. . .

IRLM

Log Manager

IBM Smart Analytics Optimizer

z/OS on System z

10‘s of processors

100‘s GB of memory

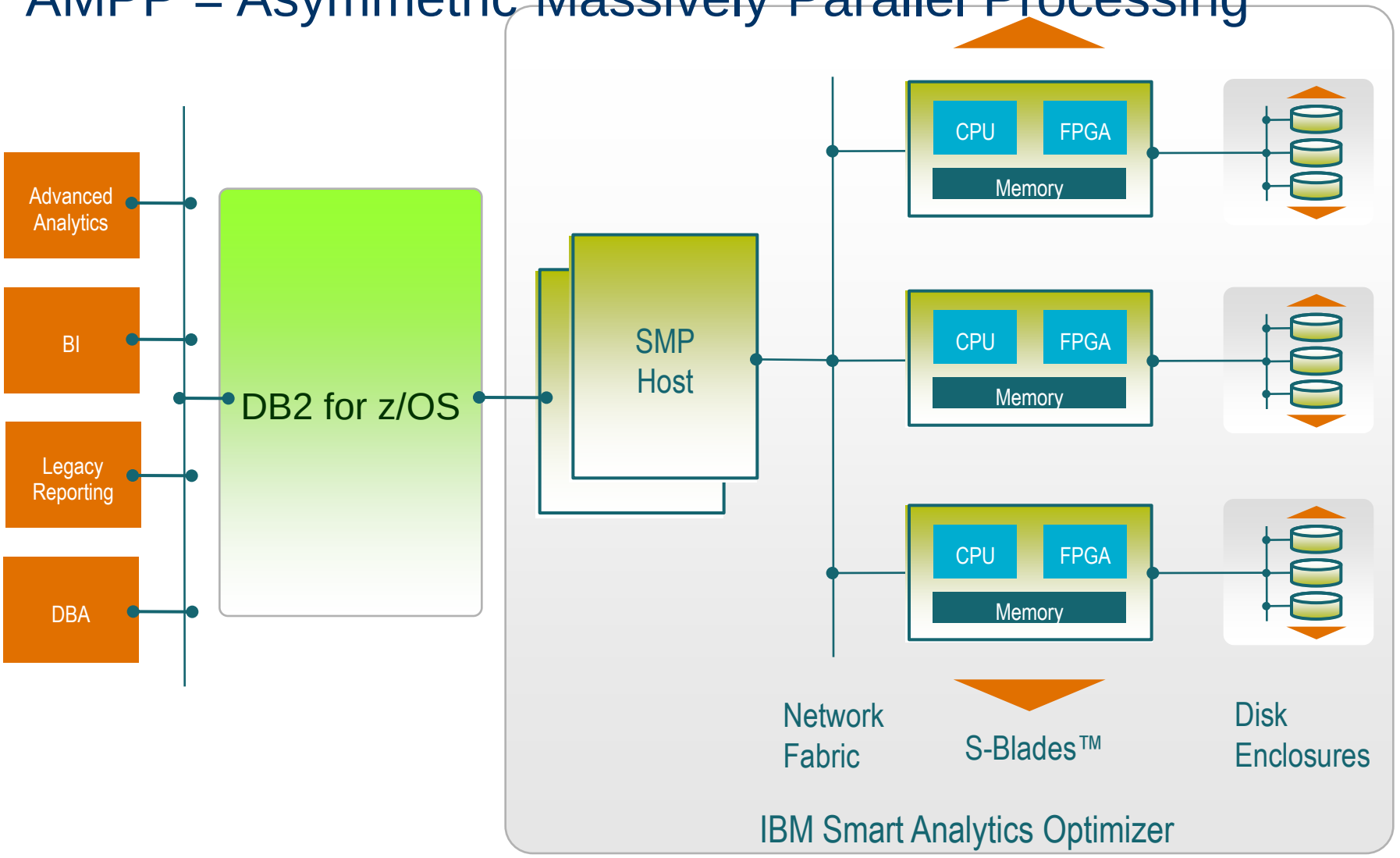
Netezza

BI/DW Queries Accelerator

Superior availabi… reliability, securi… workload managem…

Bringing Netezza AMPP™ Architecture to DB2

AMPP = Asymmetric Massively Parallel Processing



Asymmetric Massively Parallel Processing

Database tables spread across all disks on all S-Blades

Distribution done during loading and during query processing (hash and random supported)

The SMP Host

Receives incoming SQL from DB2

Compiles SQL into AMPP optimized plan and distributes plans step-by-step to S-Blades

Receives, consolidates and transmits the results to DB2

Provides system management and oversight when problems detected, and reports back to DB2

The S-Blades

Work in parallel on up to 8 data streams (data 'slices') from associated disks

Perform complex operations on their local data (joins, aggregating, sorting, etc)

Communicate data with each of the other S-Blades across internal Ethernet fabric

For each table accessed

Data is read into table cache buffers for all disks on each S-Blade

ZoneMap ™ statistics are used to skip reading data extents that can't have needed data

FPGAs decompress data, validate transactional visibility, apply restrict and project rules and write resulting data subset into DRAM for further processing by DBOS software

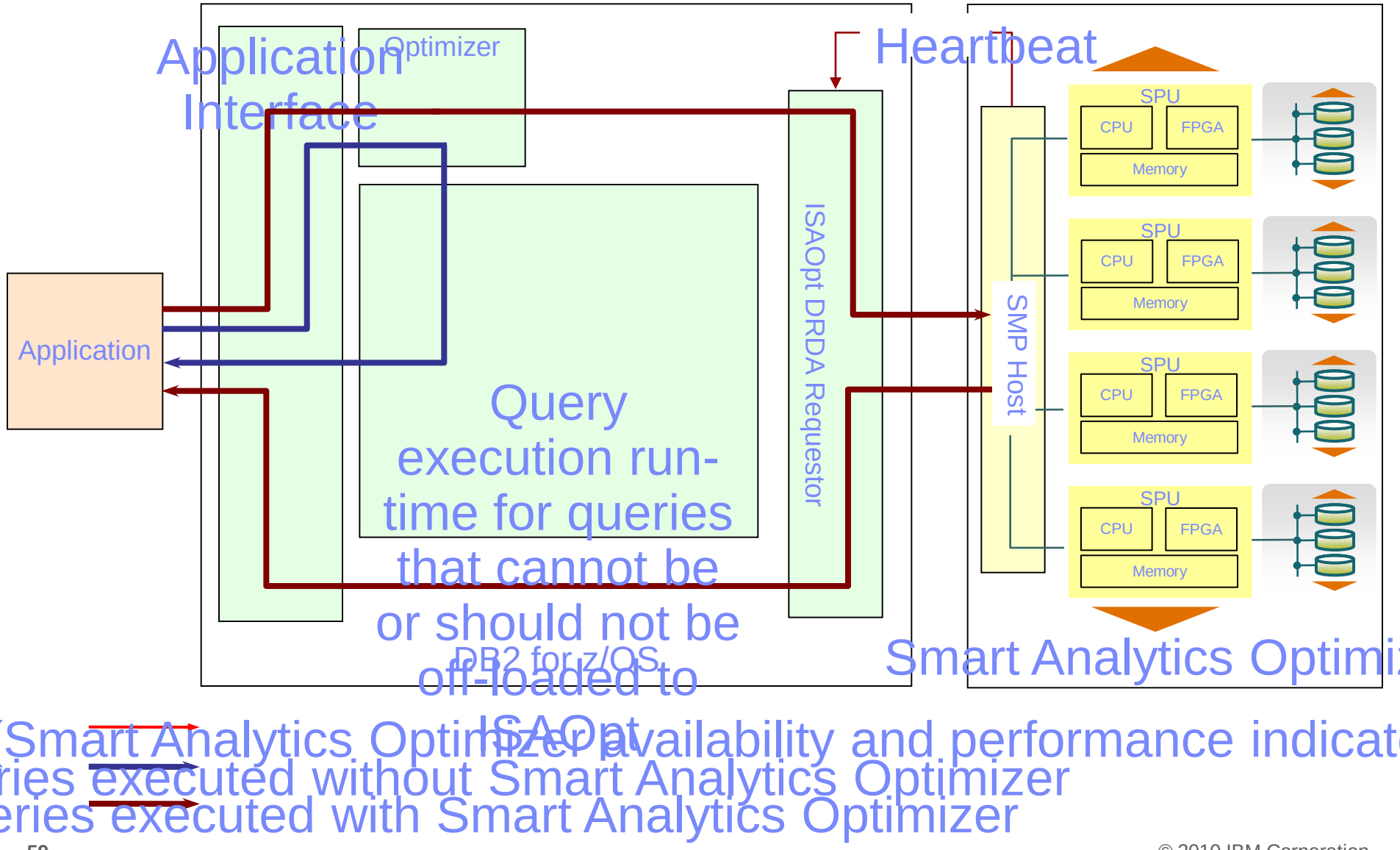
The NPS architecture achieves massively parallel processing by

Using all S-Blades in all operations whenever possible

Eliminating unnecessary data by reading and processing only the data required to satisfy the query

Processing the data on the S-Blades close to the disks where it resides

Query Execution Process Flow



ISAOpt V1 Needed Following Enhancements

ISAOpt V1.2 addresses all of them !

1. **Increase applicability by relaxing current off-load restrictions**
   - See next chart
2. **Increase applicability by supporting larger amount of data**
   - Up to 32TB of uncompressed data
   - e.g. with 1:4 compression ratio, up to 128TB of user data
3. **Support concurrent query execution**
   - Exploiting Netezza workload management capabilities
4. **Improve data currency**
   - Partition-scope update
5. **Support disaster recovery**
   - Building blocks provided
6. **DB2 10 support**
   - ISAOpt V1.2 supports both DB2 9 and DB2 10
   - Requires zEnterprise (z196 or z114)

ISAOpt V1.2 Lifting Query Off-load Limitations

- **Multiple query blocks are processed one by one resulting in less than optimal performance of:**
  - Subselects in quantitative predicates (SOME, ANY, ALL)
  - EXISTS or IN predicates with subselects
  - UNION, INTERSECT, EXCEPT
- **Outer query block that contains a subselect cannot be accelerated**
  - Only the leaf-blocks in an access path tree apply for acceleration
- **Limited support for very large dimension tables**
  - Especially if the predicates on them are not selective
- **No full outer join, no right outer join**
- **Only equi-joins (no range join predicates)**
- **No queries that do not include at least one fact table**
- **Fact table cannot appear at the null-padding side of the outer join**
- **No queries that spread across multiple marts**
- **No static SQL**
- **No support for Unicode**
- **Not all DB2 data types such as LOBs, ROWID, XML.**
- **Not all DB2 functions**
  - No mathematical functions such as SIN, COS, TAN.

ISAOpt V1.2 Preserves V1 Key Value Propositions

- **DB2 continues to own data (both OLTP and DW)**
  - ➔ Access to data (authorization, privileges, …)
  - ➔ Data consistency and integrity (backup, recovery, …)
  - ➔ **Enables extending System z QoS characteristics to BI/DW data as well**
- **Applications access data (both OLTP and DW) only through DB2**
  - ➔ DB2 controls whether to execute query in DB2 mainline or route to ISAOpt
  - ➔ DB2 returns results directly to the calling application
  - ➔ **Enables mixed workloads and selection of optimal access path (within DB2 mainline or ISAOpt) depending on access pattern**
- **ISAOpt continues to be implemented as DB2 internal component**
  - ➔ DB2 provides key ISAOpt status and performance indicators as well as typical administration tasks by standard DB2 interfaces and means
  - ➔ No direct access (log-on) to ISAOpt
  - ➔ **Enables operational cost reduction through skills, tools and processes consolidation**
- **Investment protection**
  - ➔ **The full value of ISAO V1 will be applied to ISAO V1.2**