

Manage the Data Lifecycle of Big Data Environments

Sambasivam Sampathnathan  
IBM Corporation - Senior Product Manager



# Our World today



# Big Data!





## Some facts

- In 2009, amid the "Great Recession," the amount of digital information grew 62% over 2008 to 800 billion gigabytes (0.8 Zettabytes).
- 90% of the world's total data has been created just within the past two years
- By 2020, IT departments will be looking after 10 x more servers, 50 x more data and 75 x more files. Meanwhile, the number of IT administrators keeping track of all that [data growth](#) will increase by 1.5 times.
- By 2020, the percent of digital information requiring security beyond baseline levels will grow from 30% to 50%.
- 35% more digital information is created today than the capacity exists to store it. This number will jump to over 60% over the next several years.



twitter: Follow @ANZ\_IM or mention #IIGS

## So when will this data storm end?

- The answer is never...
  - Data will get bigger
  - Data will become more complex
- You cannot build a dam big enough to hold all of your data
- You need to be smarter
- You need to be prepared

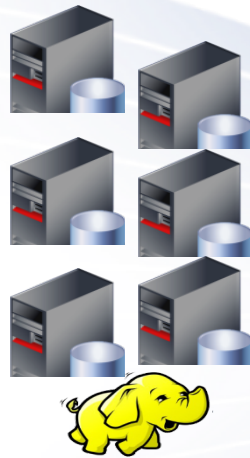


twitter: Follow @ANZ\_IM or mention #IIGS

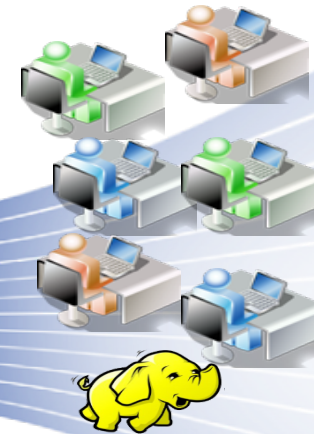
In the New Era of Computing, IIG is more important than ever before



### Sources



### Consumers



**Informatic  
Integration**

### Big Data Explosion

Data volumes are growing in every application

Performance and scalability of IIG and lifecycle management to retire old data is more important than ever before

### Increasing Complexity

New Technologies bring more source and consuming systems = greater integration complexity

IIG must integrate with big data, cloud, and mobile technologies

### Need for Agility

Need to act at the point of usage

IIG must be proactive and embedded in the consuming application

twitter: Follow @ANZ\_IM or mention #IIGS



## Strategies to maximise your investment



twitter: Follow @ANZ\_IM or mention #IIGS



Understand & discover your data



Develop a Data Management lifecycle strategy



Test Data management practices



twitter: Follow @ANZ\_IM or mention #IIGS



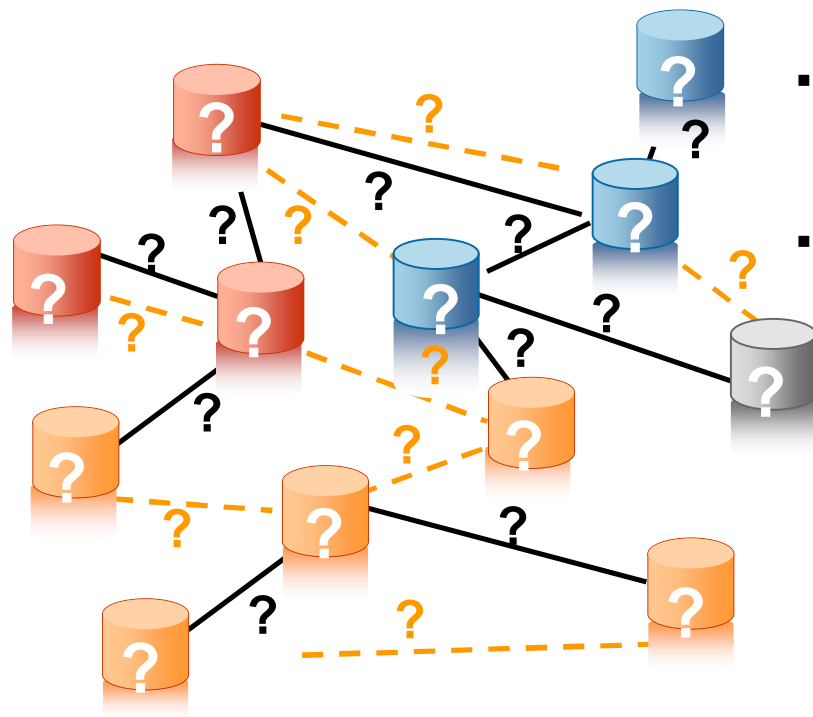


Understand & discover your data



twitter: Follow @ANZ\_IM or mention #IIGS

# You can't govern what you don't understand



**Distributed Data Landscape**

- Data can be distributed over multiple applications, databases and platforms
  - Where are those databases located?
- Complex, poorly documented data relationships
  - Which data is sensitive, and which can be shared?
  - Whole and partial sensitive data elements can be found in hundreds of tables and fields

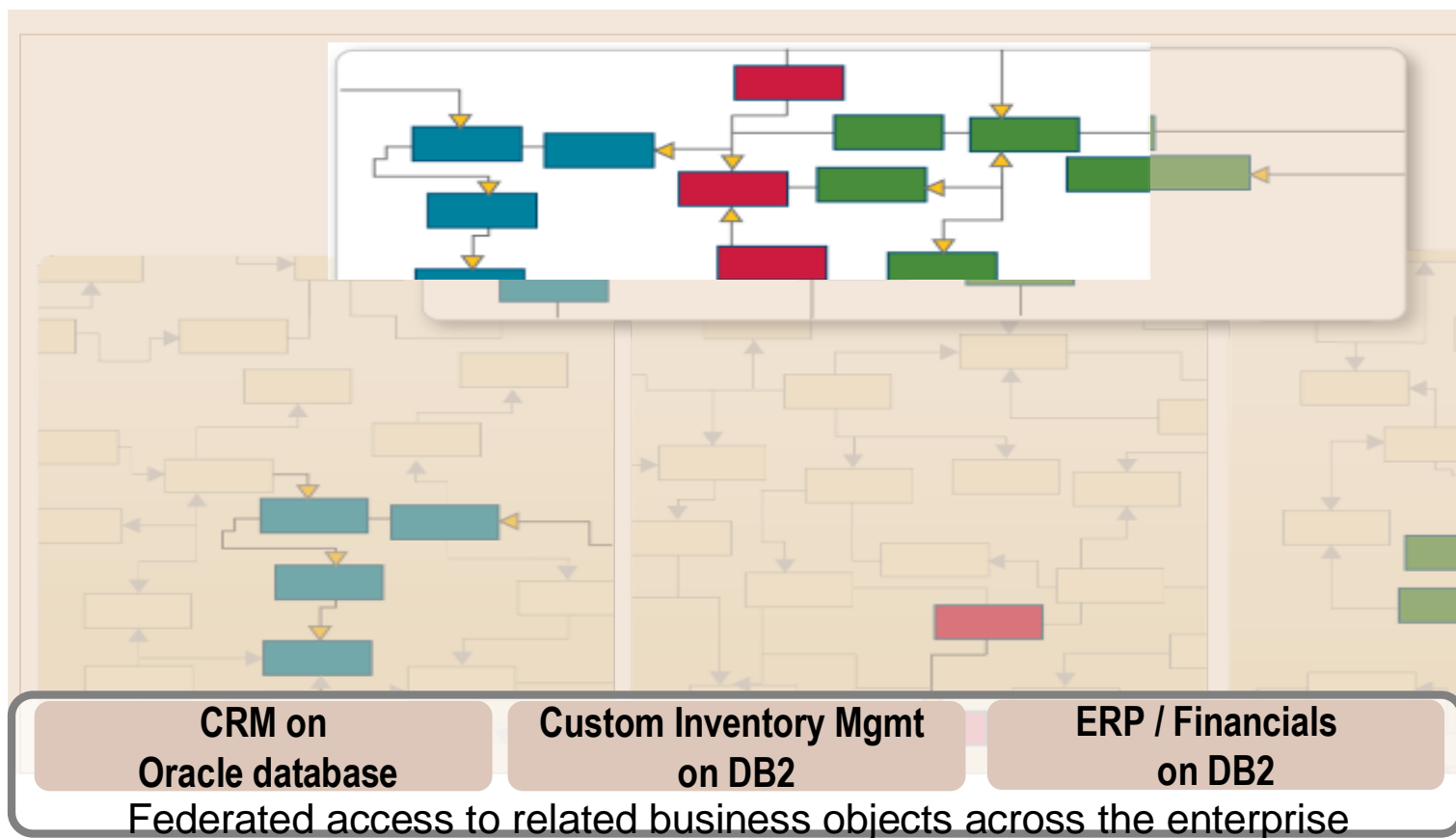
Data relationships not understood because:

- Corporate memory is poor
- Documentation is poor or nonexistent
- Logical relationships (enforced through application logic or business rules) are hidden

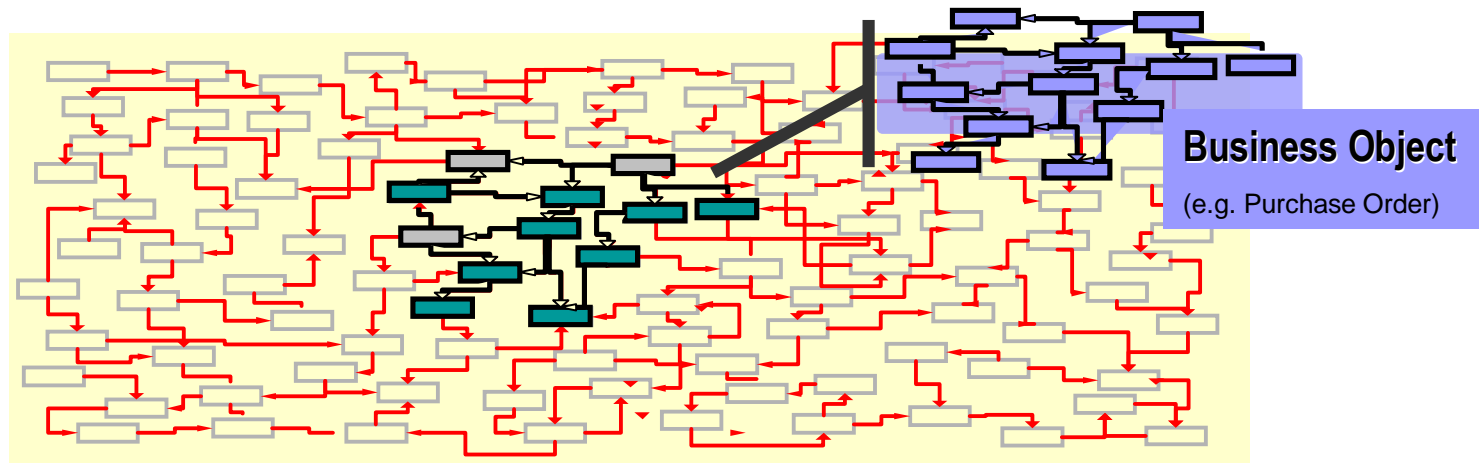




## Understanding Business Objects Across Enterprise Applications



## Oracle Applications: The Challenge



- Database schemas with thousands of tables & complex relationships
- Almost No PK or FK Constraints
- Database System Catalog does not hold useful metadata
- Proprietary ERP/CRM metadata holds 'Logical View' of data
- Customers extending the standard data model
- Scoping the tables for data mgmt projects

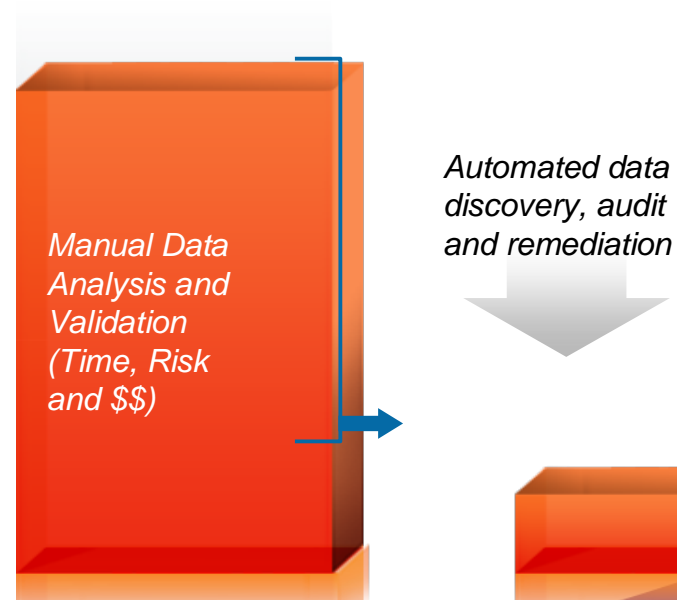
## Value to Our Customers

Accelerate understanding your existing distributed data landscape for:

- Data Archiving
- Test Data management
- Sensitive Data
- Application/Data Consolidation, Migration and Retirement
- Master Data Management and Data Warehousing

10x reduction in risk, time and effort for the discovery phase projects

- Automated discovery of business entities, cross-source business rules and anomalies
- Increased repeatability
- Verifiable results



twitter: Follow @ANZ\_IM or mention #IIGS

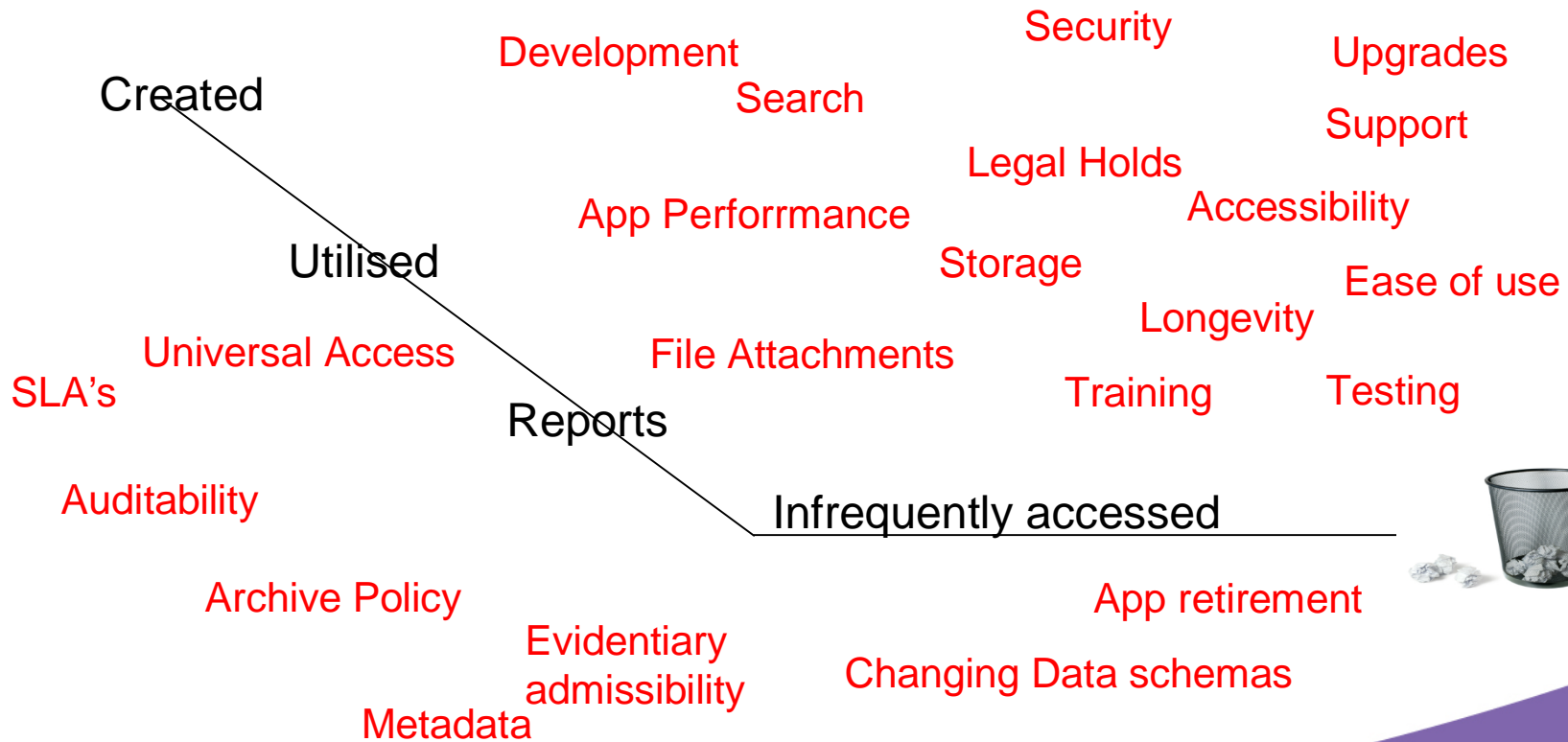


Develop a Data Management lifecycle strategy



twitter: Follow @ANZ\_IM or mention #IIGS

# The Lifecycle of data



twitter: Follow @ANZ\_IM or mention #IIGS

## Organizations have been increasingly challenged with successfully managing data growth



### Increasing Costs

**3-10x**

Cost of managing storage over the cost to procure<sup>a</sup>

**\$1.1 billion**

Amount organizations will have spent in 2011 on storage<sup>b</sup>

### Poor Application Performance

**80%**

The time DBA's spend weekly on disk capacity issues<sup>c</sup>

**250 hours**

The amount of time needed to run "daily" batch processes<sup>d</sup>

### Manage Risk & Compliance

**50%**

of firms retain structured data for 7+ years<sup>e</sup>

**57%**

of firms use Back-up for data retention needs<sup>e</sup>

(a) Merv Adrian, IT Market Strategies, "[Data Growth Challenges Demand Proactive Data Management](#)", November 2009

(b) IDC, "Worldwide Archival Storage Solutions 2011–2015 Forecast: Archiving Needs Thrive in an Information-Thirsty World", October 2011

(c) Simple-Talk, "[Managing Data Growth in SQL Server](#)", January 2010

(d) IBM Client Case Study: [Toshiba TEC Europe](#); archiving reduced batch process time by 75%

(e) IDC Quick Poll Survey 2011, "Data Management for IT Optimization and Compliance", November 2011



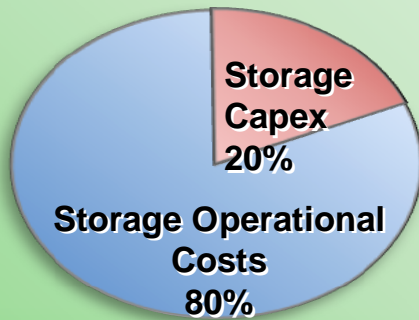
twitter: Follow @ANZ\_IM or mention #IIGS



Organizations have been increasingly challenged with successfully managing data growth

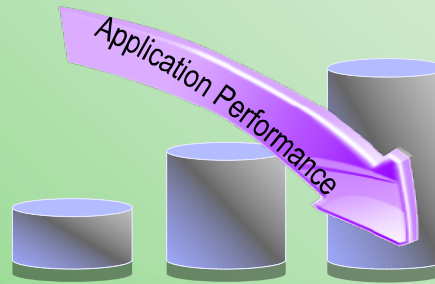


### Increasing Costs



Buying more storage is not a “cheap” fix when you add the operational burden

### Poor Application Performance



Business users & customers wait for application response; DBA's spend majority of time fixing performance issues

### Manage Risk & Compliance

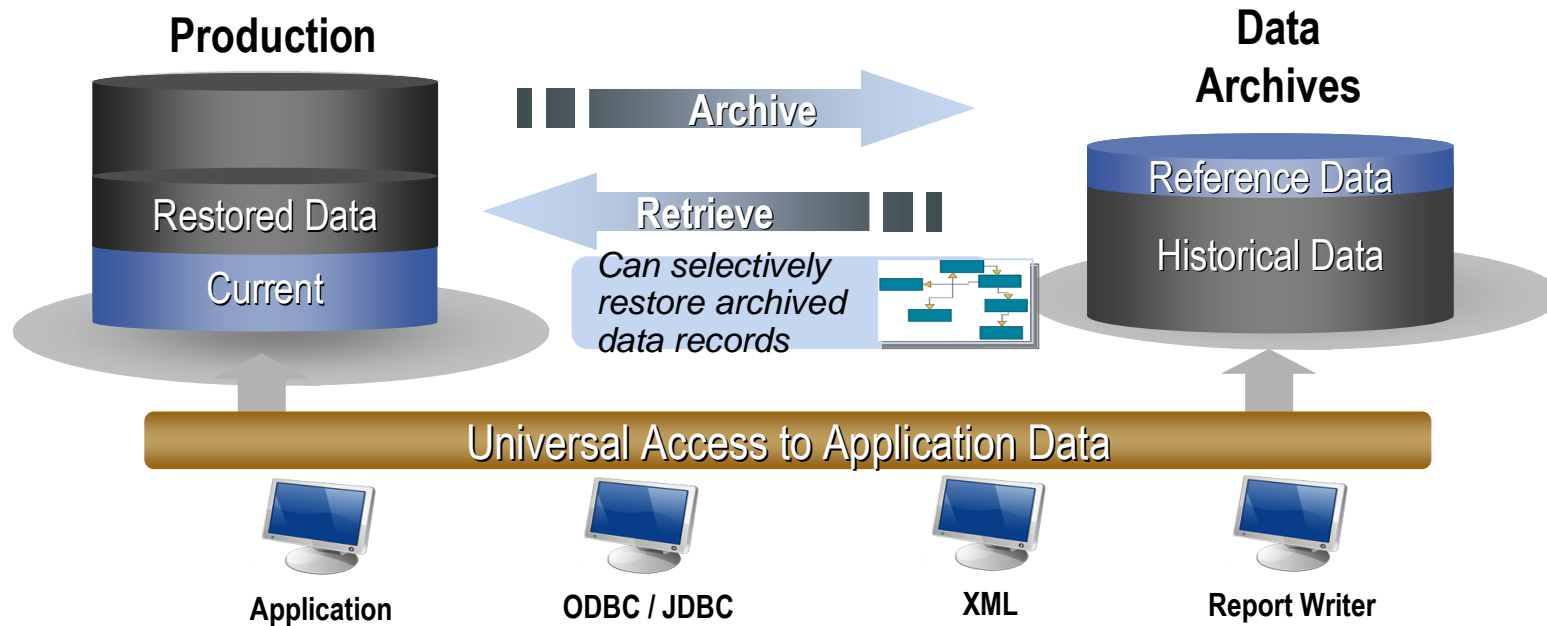


The “keep everything” strategy can impact disaster recovery and data retention & disposal compliance

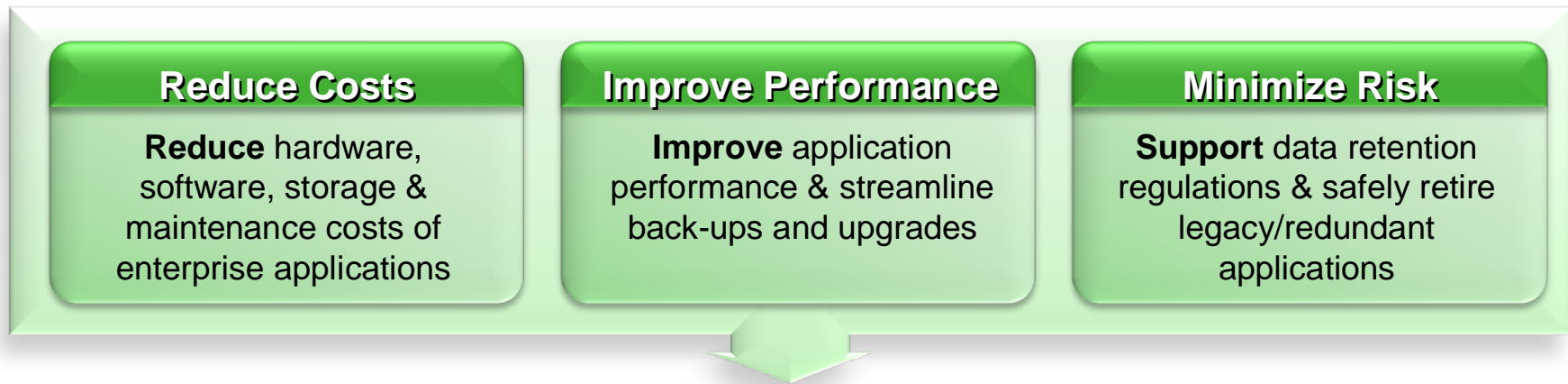


twitter: Follow @ANZ\_IM or mention #IIGS

# Archive historical data for data growth management



# Effectively Archive and Manage Data Growth



- Discover & identify data record types to archive across heterogeneous environments
- Intelligently archive data to improve application performance and support data retention
- Capture & store historical data in its original business context
- Define & maintain data retention policies consistently across the enterprise
- Ensure long-term, application-independent access of archived data via multiple access methods
- Support for custom & packaged ERP applications in heterogeneous environments



twitter: Follow @ANZ\_IM or mention #IIGS



## An effective data archiving policy

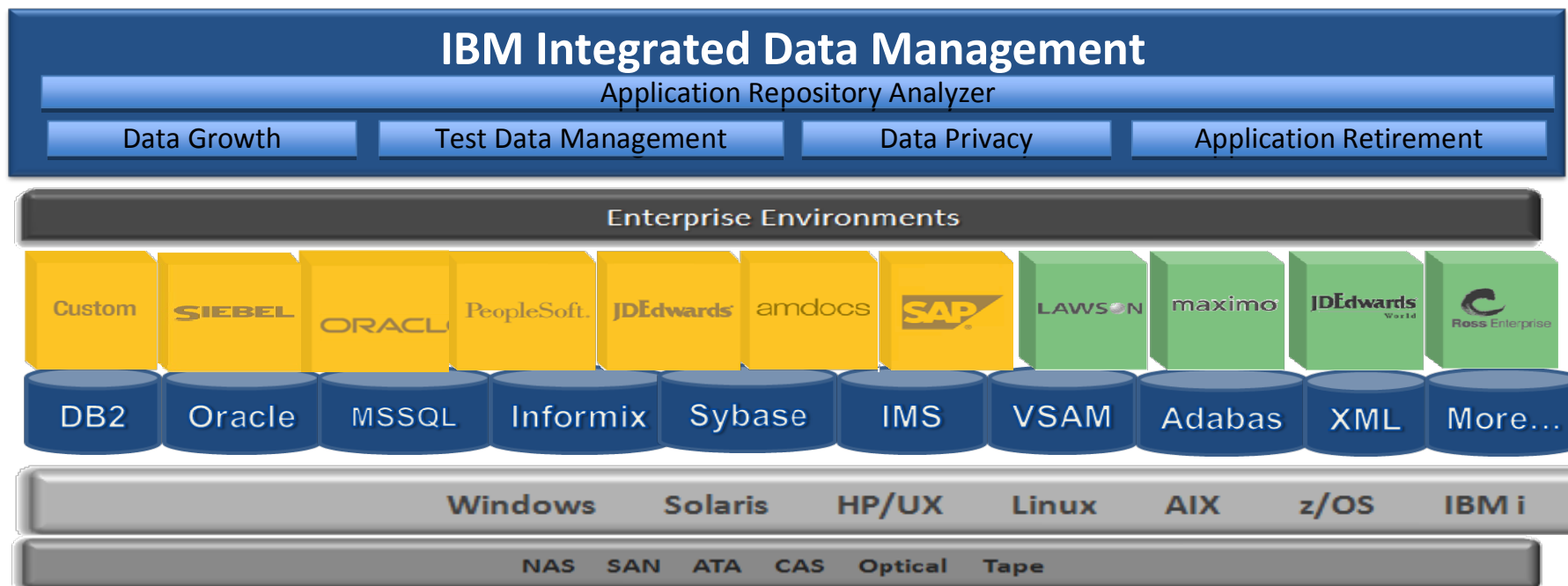
- Starts out small and targeted
  - Archive data object and related entities as a referentially intact set of data
  - Ensures archived data is still accessible
  - Ties into corporate records keeping requirements
- 
- IS an evolving project that **MUST** be flexible as requirements change



twitter: Follow @ANZ\_IM or mention #IIGS



# InfoSphere Optim™ Enterprise Architecture



Single, Scalable, Enterprise Data Management Solution



twitter: Follow @ANZ\_IM or mention #IIG\$1



## Test Data management practices



## Organizations continue to be challenged with building quality applications



### Increasing Risk

**45,000+**

Number of sensitive records exposed to 3<sup>rd</sup> party during testing<sup>c</sup>

**62%**

companies use actual customer data to test applications<sup>a</sup>

### Time to Market

**37%**

Satisfied with speed of software development<sup>f</sup>

**30-50%**

Time testing teams spend on setting up test environments, instead of testing<sup>b</sup>

### Increasing Costs

**\$300 billion**

Annual costs of software-related downtime.<sup>d</sup>

**32%**

Low success rate for software projects<sup>e</sup>

a. The Ponemon Institute. The Insecurity of Test Data: The Unseen Crisis

b. NIST, Planning Report. The Economic Impacts of Inadequate Infrastructure for Software Testing

c. Federal Aviation Administration: Exposes unprotected test data to a third party <http://fcw.com/articles/2009/02/10/faa-data-breach.aspx>

d. The Standish Group, *Comparative Economic Normalization Technology Study*, CHAOS Chronicles v12.3.9, June 30, 2008

e. The Standish Group, *Chaos Report*, April 2009

f. Forrester Research, "Corporate Software Development Fails To Satisfy On Speed Or Quality", 2005



twitter: Follow @ANZ\_IM or mention #IIGS



## Organizations continue to be challenged with building quality applications



### Increasing Risk



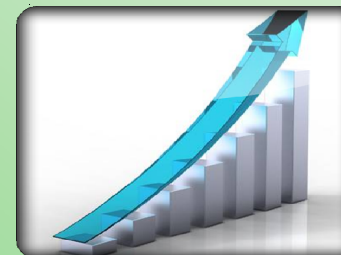
Mandatory to protect data and comply with regulations

### Time to Market



Lack of realistic test data and inadequate environments

### Increasing Costs



Defects are caught late in the cycle

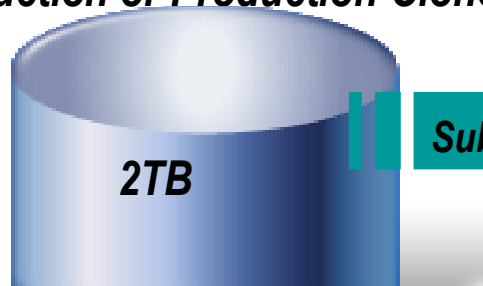




# Employ effective test data management practices



**Production or Production Clone**



**Subset & Mask**



**-Compare  
-Refresh**

**25 GB**

**Unit Test**

**25 GB**

**Development**

**100 GB**

**Integration Test**

**50 GB**

**Training**



twitter: Follow @ANZ\_IM or mention #IIGS

- Create targeted, right-sized test environments
- Substitute sensitive data with fictionalized yet contextually accurate data
- Easily refresh, reset and maintain test environments
- Compare data to pinpoint and resolve application defects faster
- Accelerate release schedules

## Improve application quality and delivery efficiency

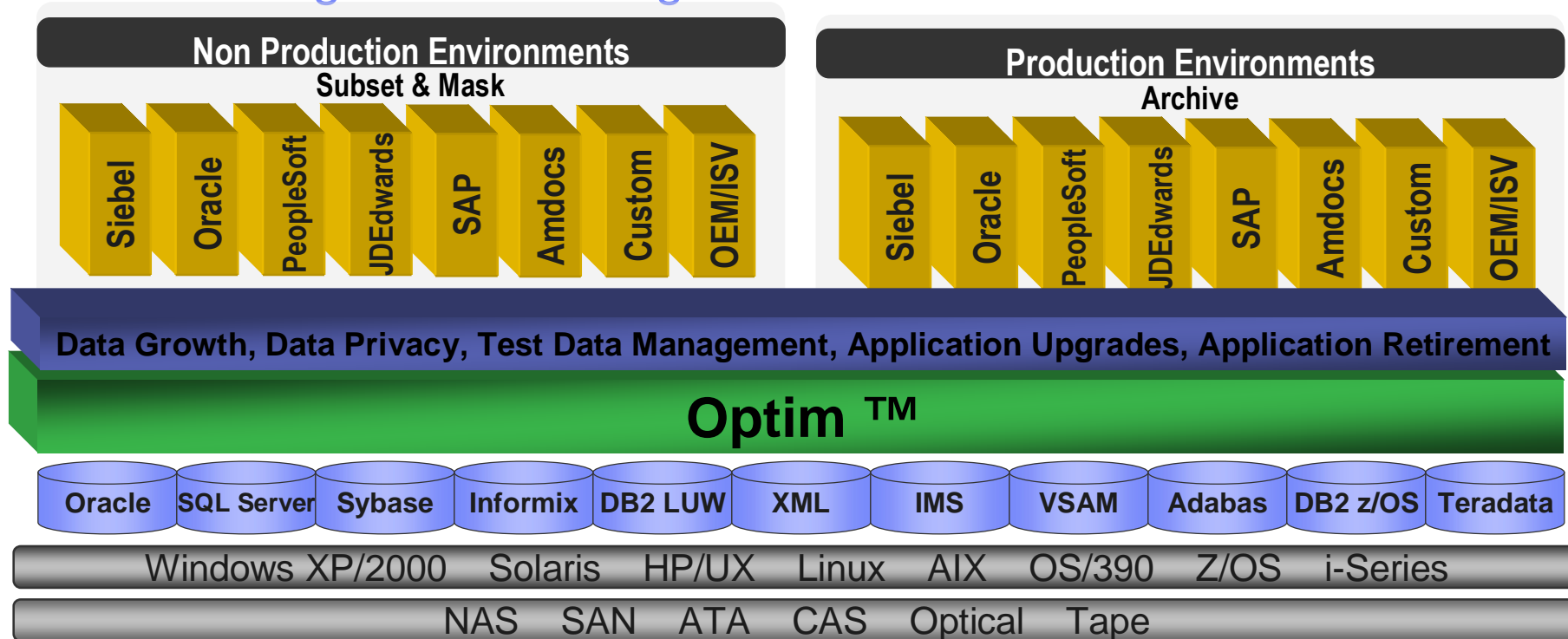


- Understand what test data is needed for test cases
- Create “right-sized” test data by subsetting
- Ensure masked data is contextually appropriate to the data it replaced, so as not to impede testing
- Easily refresh & maintain test environments through self service access by developers and testers
- Automate test result comparisons to identify hidden errors
- Support for custom & packaged ERP applications in heterogeneous environments



twitter: Follow @ANZ\_IM or mention #IIGS

## IBM meeting these challenges



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



# Questions?

## **Further Information**

[www.optimsolution.com](http://www.optimsolution.com)



twitter: Follow @ANZ\_IM or mention #IIGS



# Team You



twitter: Follow @ANZ\_IM or mention #IIGS

This document was created with Win2PDF available at <http://www.win2pdf.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.  
This page will not be added after purchasing Win2PDF.