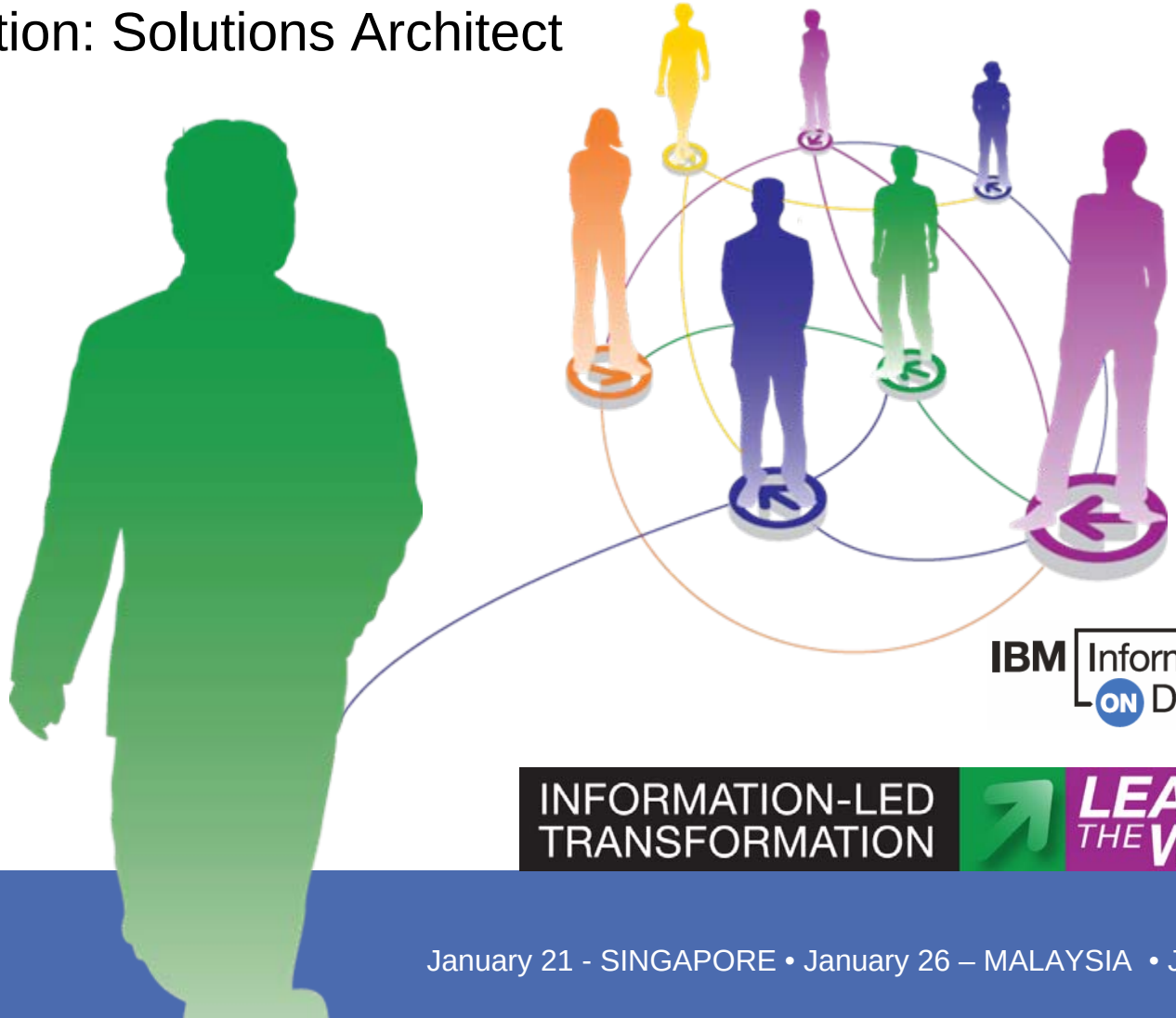
ASEAN INFORMATION ON DEMAND CONFERENCE 2010

Presenter: George Lapis

Designation: Solutions Architect



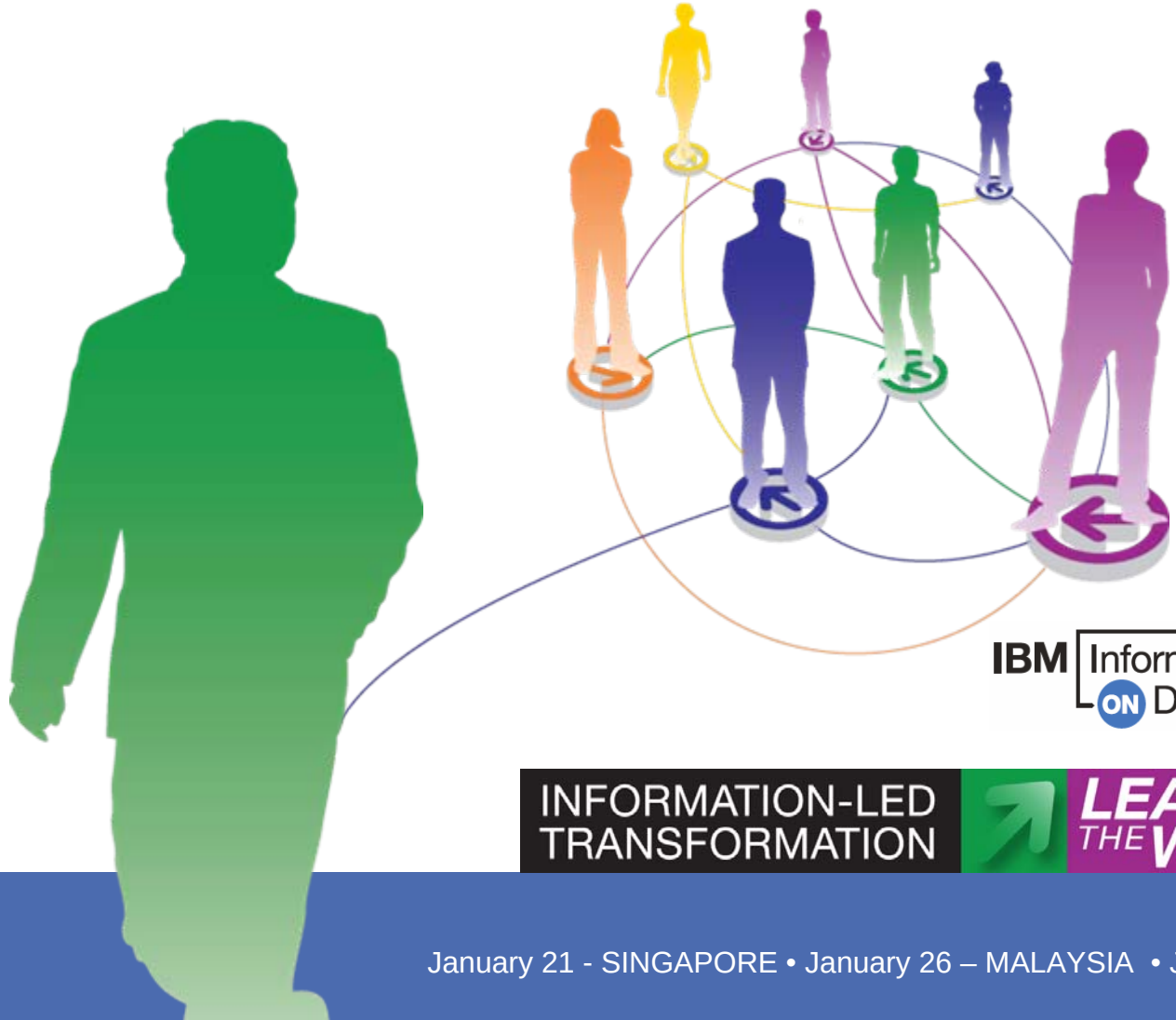
IBM Information ON Demand 2010

INFORMATION-LED TRANSFORMATION LEAD THE WAY

January 21 - SINGAPORE • January 26 – MALAYSIA • January 28 - THAILAND

IBM Optim Solutions

Introduction to Integrated Data Management

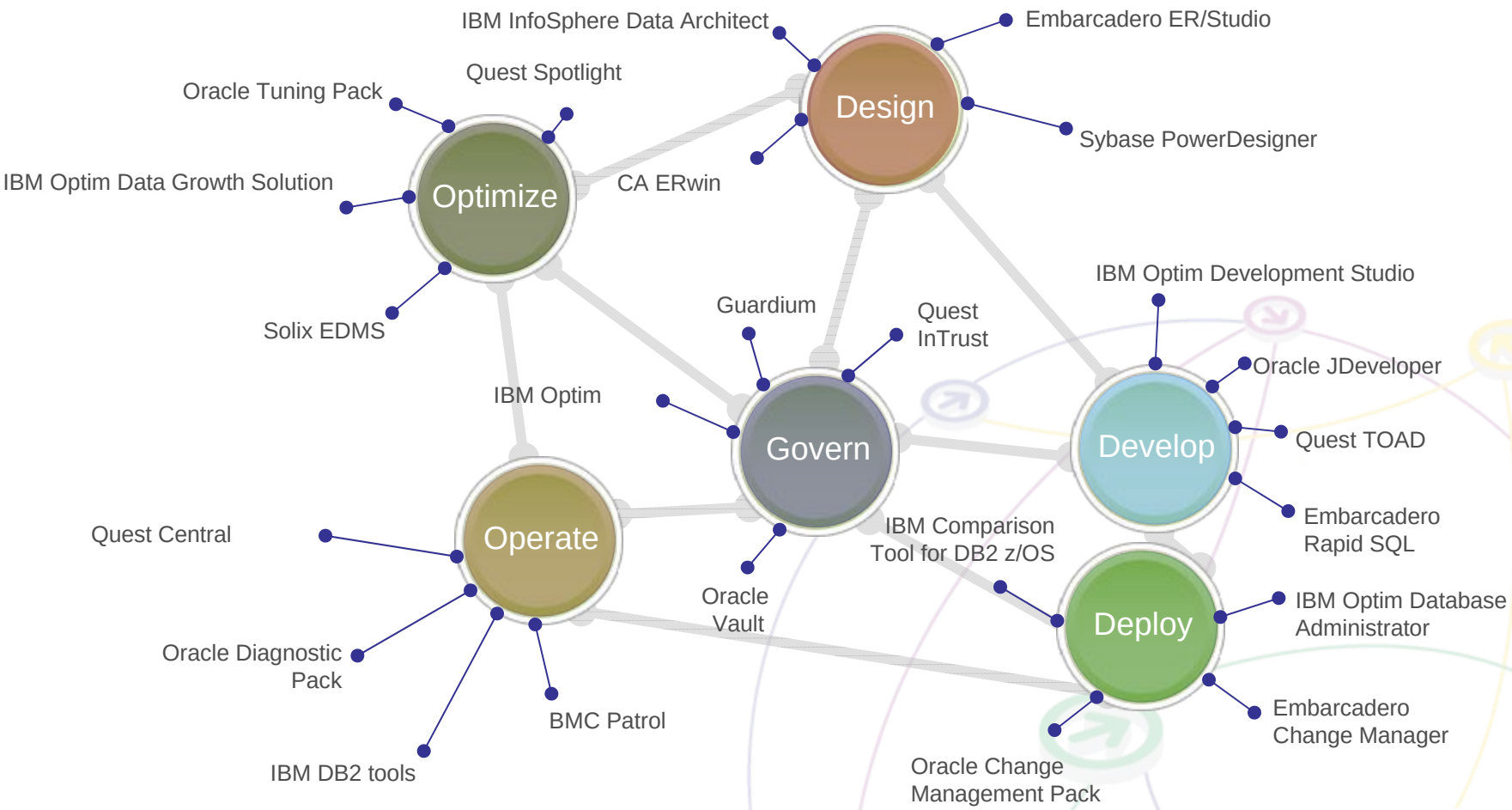


IBM Information On Demand 2010

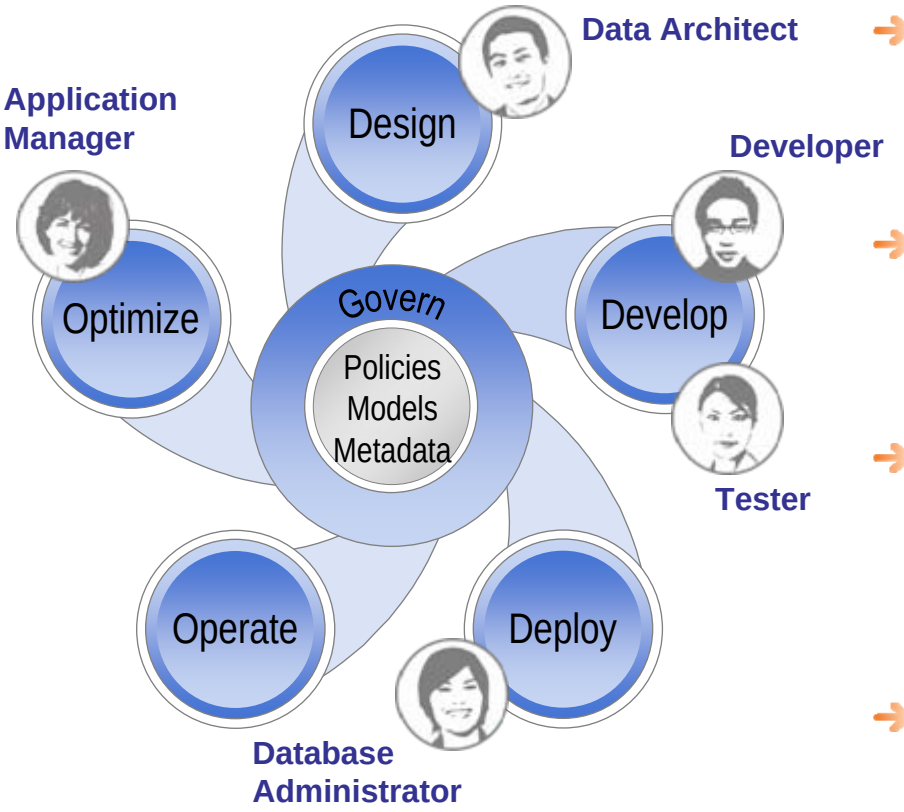
INFORMATION-LED TRANSFORMATION LEAD THE WAY

What do Businesses Have?

A Collection of Disparate, Single-Purpose Products



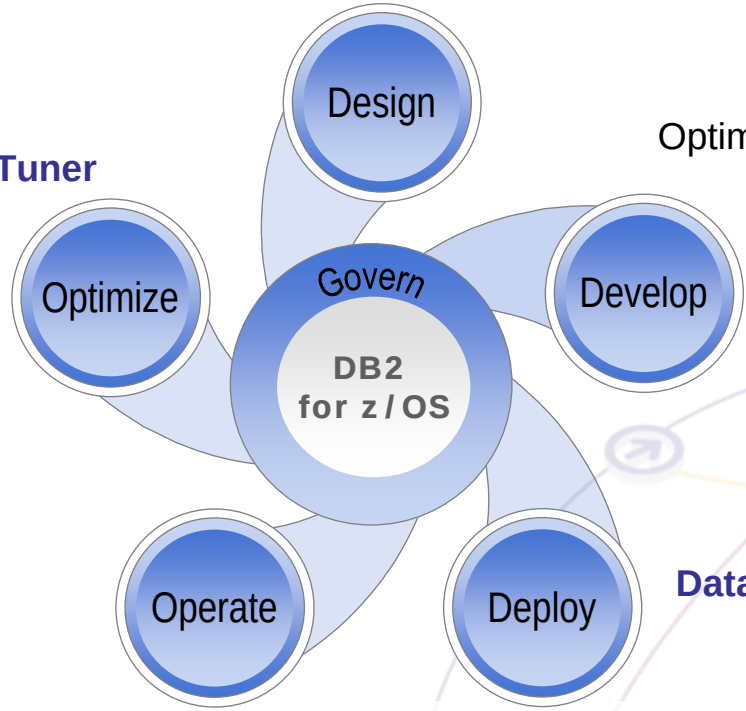
Integrated Data Management – What's Different?



- **Produce enterprise-ready applications faster**
  - Improve data access, speed iterative testing
  - Empower collaboration between architects, developers & DBAs
- **Consistently achieve service level targets**
  - Automate and simplify operations
  - Provide contextual intelligence across the solution stack
- **Support business growth**
  - Accommodate new initiatives without expanding infrastructure
  - Simplify application upgrades, consolidation & retirement
- **Facilitate alignment, consistency & governance**
  - Define business policies and standards up front; share, extend, and apply throughout the lifecycle

The broadest range of capabilities for managing the value of your data throughout its lifetime

InfoSphere Data Architect



Optim Development Studio

Optim Query Workload Tuner

Optim Data Growth Solution

DB2 Performance Management Solution

Optim Test Data Management

Optim Data Privacy Solution

DB2 Automation Tool

DB2 Recovery Solution

DB2 Utilities Suite

DB2 Change Management Solution

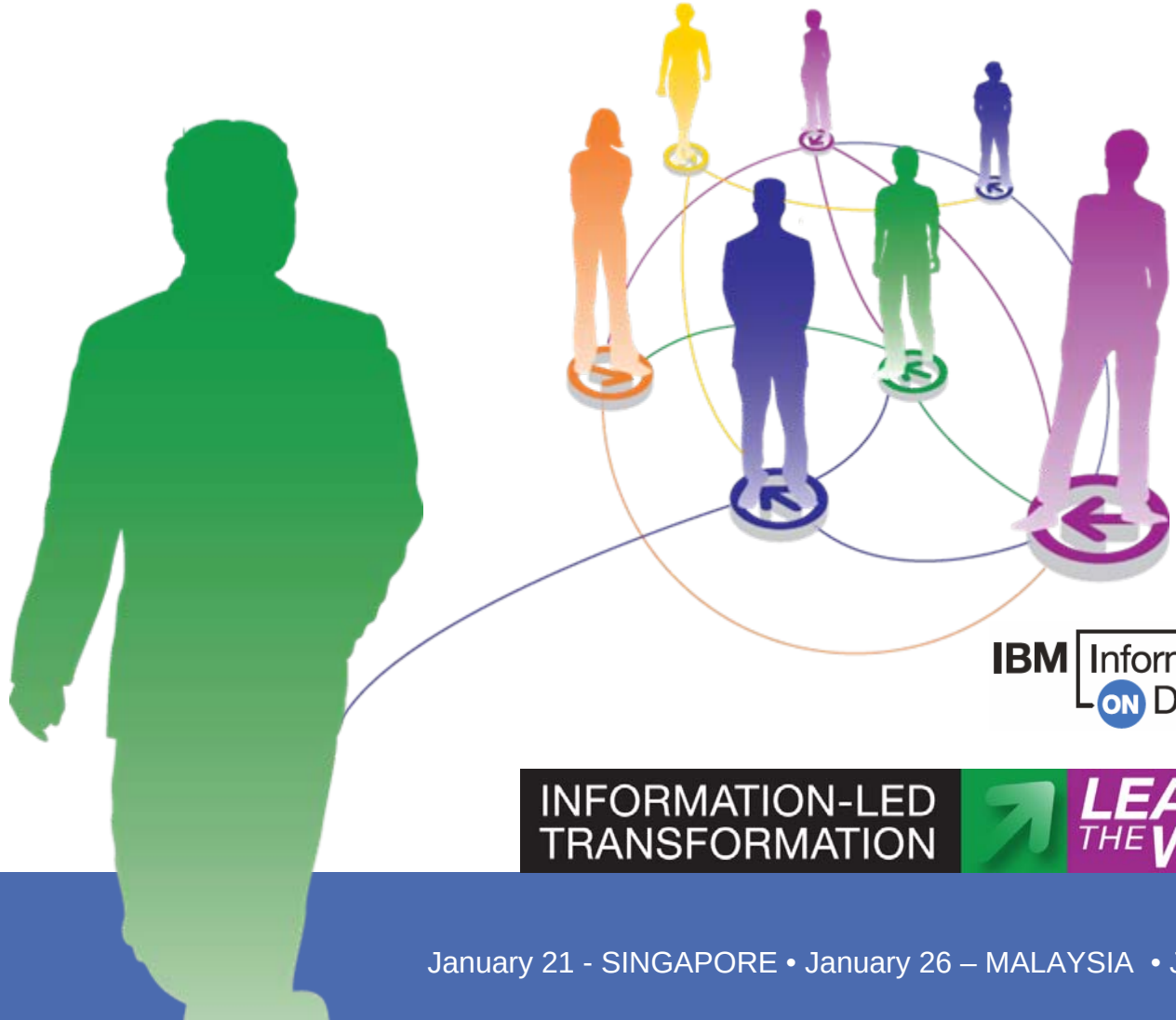
Data Studio pureQuery Runtime

DB2 Audit Management Expert

Data Encryption for DB2 and IMS

IBM Optim Solutions

Challenges around data governance



Challenges Facing Customers Today

- Mitigate Risk
  - Effectively and securely manage archived data
  - Protect data privacy
  - Accurate, prompt responses to auditing requests
- Maintain Performance in face of Data Growth
  - Improve application performance by moving historical transaction records to a safe, secure archive
  - Achieve Service Level Agreements (SLAs) consistently
- Control Costs
  - Reduce infrastructure costs; utilize cost effective tiered storage
  - Minimize cost and time for compliance
  - Improve productivity of development team

Challenges : Reduce Risk

- Insiders and hackers are targeting data for profit
- Data in and of itself has monetary value:
  - Credit Card Number With PIN - $500
  - Drivers License - $150
  - Birth Certificate - $150 *Source: USA TODAY research 10/06*
- Average cost of a data breach in 2007 was $197 USD per customer record leaked ***Source: Ponemon Institute***
- This has been a driving factor for creating data protection and privacy regulations
- How to protect Personal Identifiable Information (PII)?

Challenges: Dealing with Data Growth

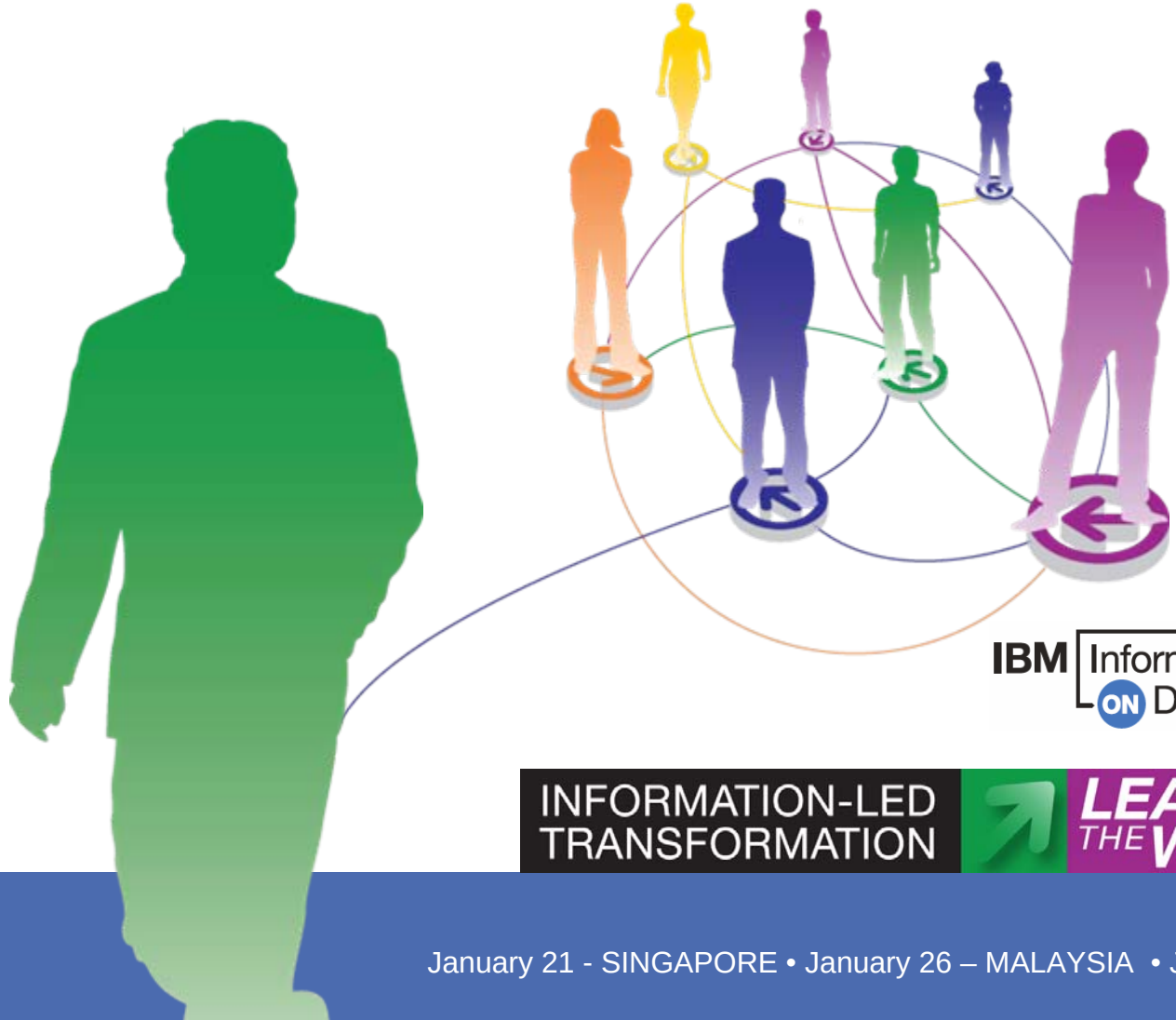
- Data is growing at a very rapid rate
  - Annual growth rates for databases exceed 125%
  - Mergers & acquisitions
  - Data Warehousing
- The Data Multiplier effect
  - OLAP cubes, data marts, and so on
  - Copies of data for test, development, quality assurance, disaster recovery, etc.
- Retention of data for compliance purposes
- Web 2.0 applications are verbose
- How to manage data growth and aging effectively?

Challenges: Control Costs

- Growing storage costs due to rapid data growth
- Cost of storing and managing the many copies of your production data
- Cost of implementing data privacy measures for compliance across different databases and applications
- Cost of archive retrievals for compliance requests and e-discovery
- Growing development time costs
- How to control storage and data management costs?

IBM Optim Solutions

How IBM Optim Can Help



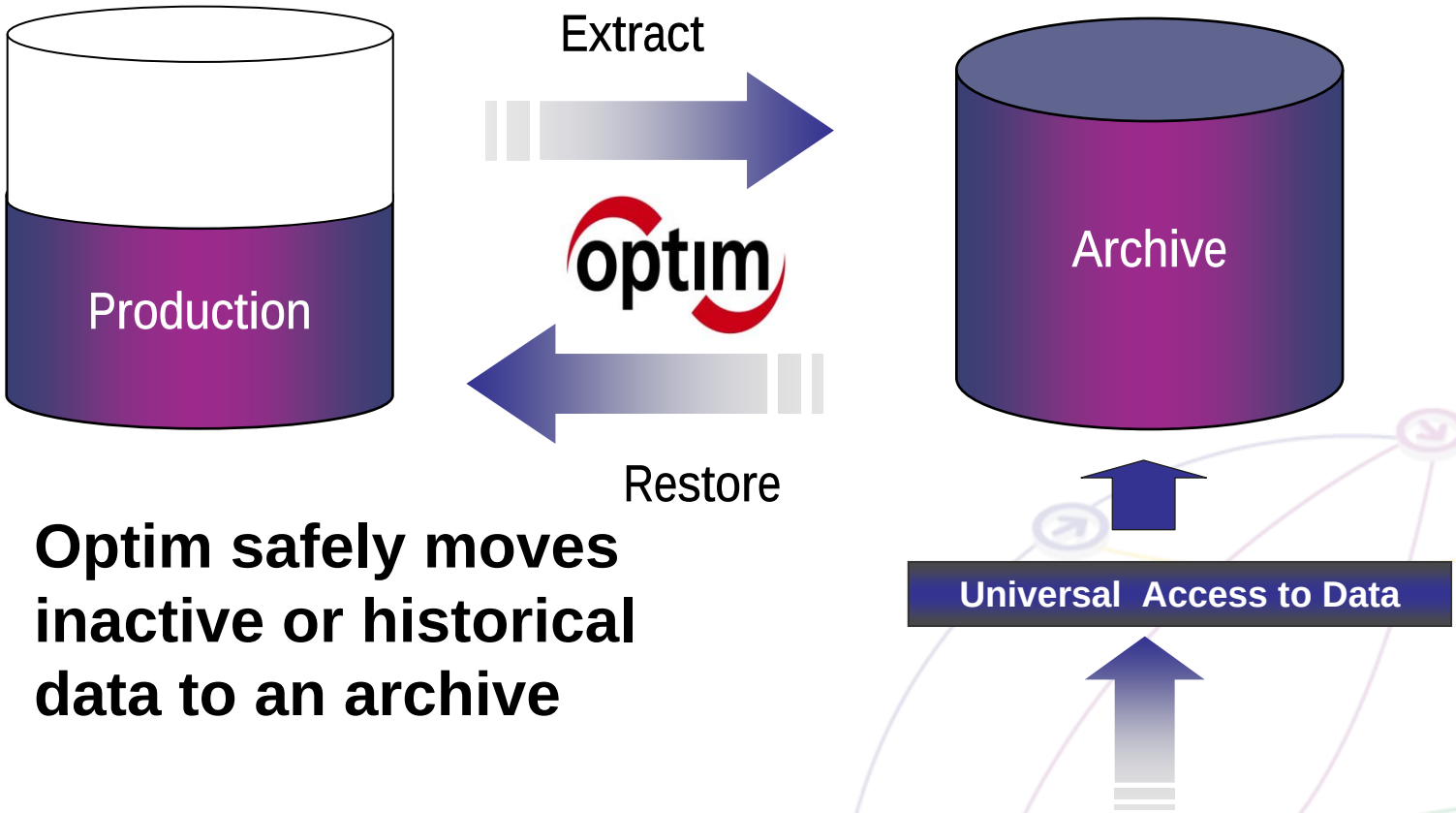
IBM Information On Demand 2010

INFORMATION-LED TRANSFORMATION LEAD THE WAY

How IBM Optim Helps

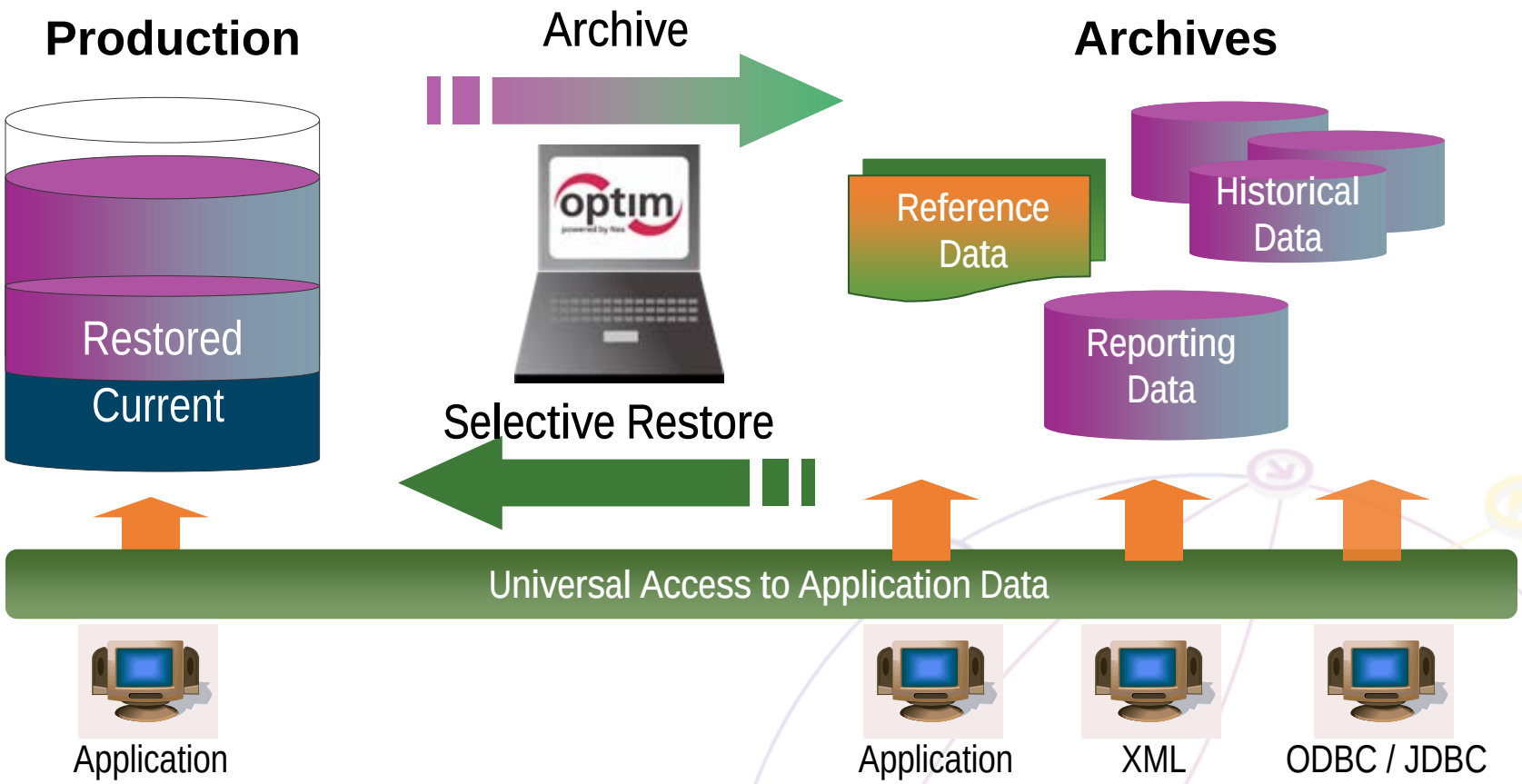
- Archive, Retention, and e-Discovery
- Test Data Management
- Application Decommissioning
- Data Privacy
- Information Lifecycle Management (ILM)

Archive, Retention, and e-Discovery



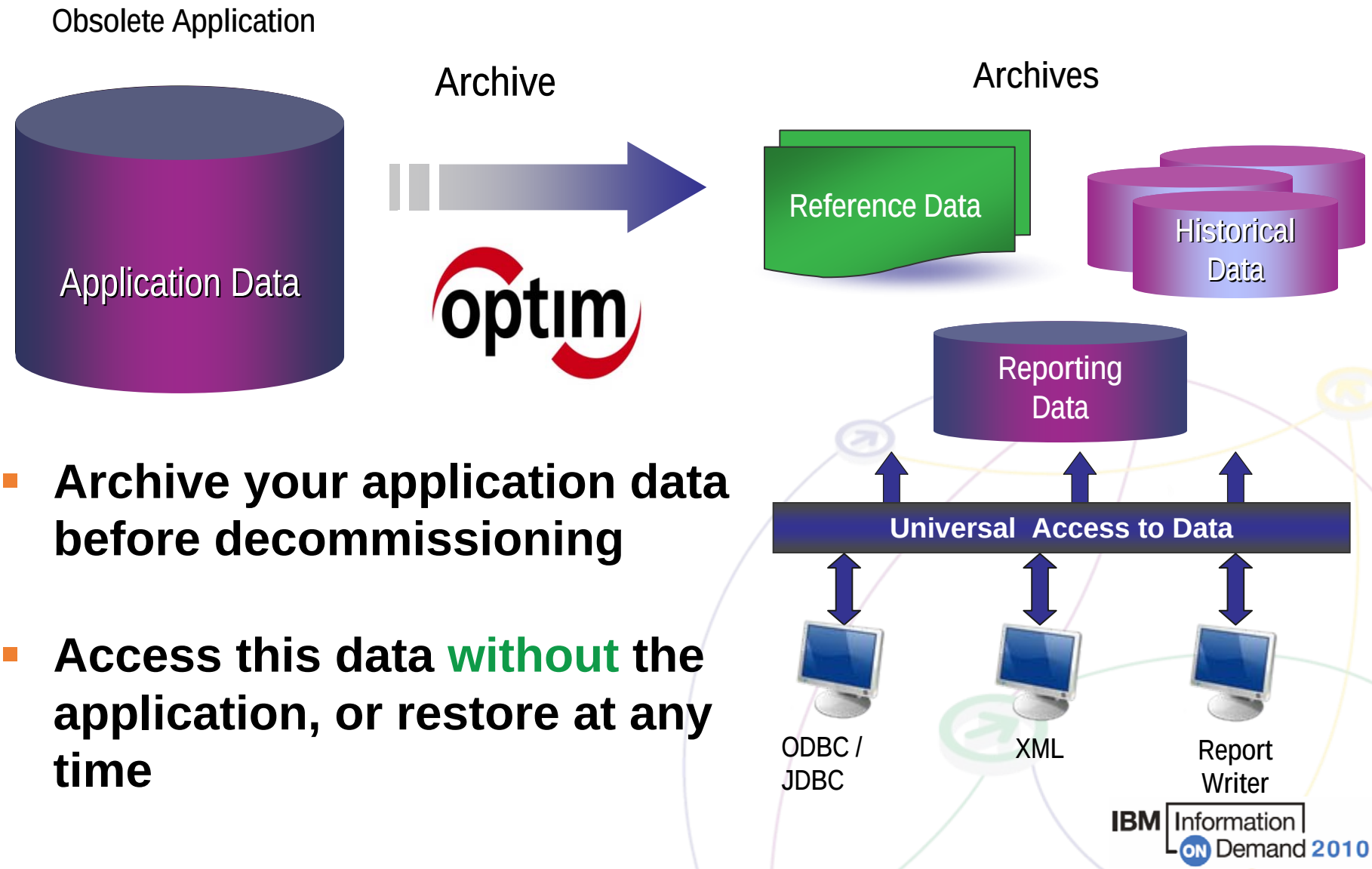
- **Optim safely moves inactive or historical data to an archive**
- **Archive can be accessed in many ways**

Optim Data Growth Solution: Archiving



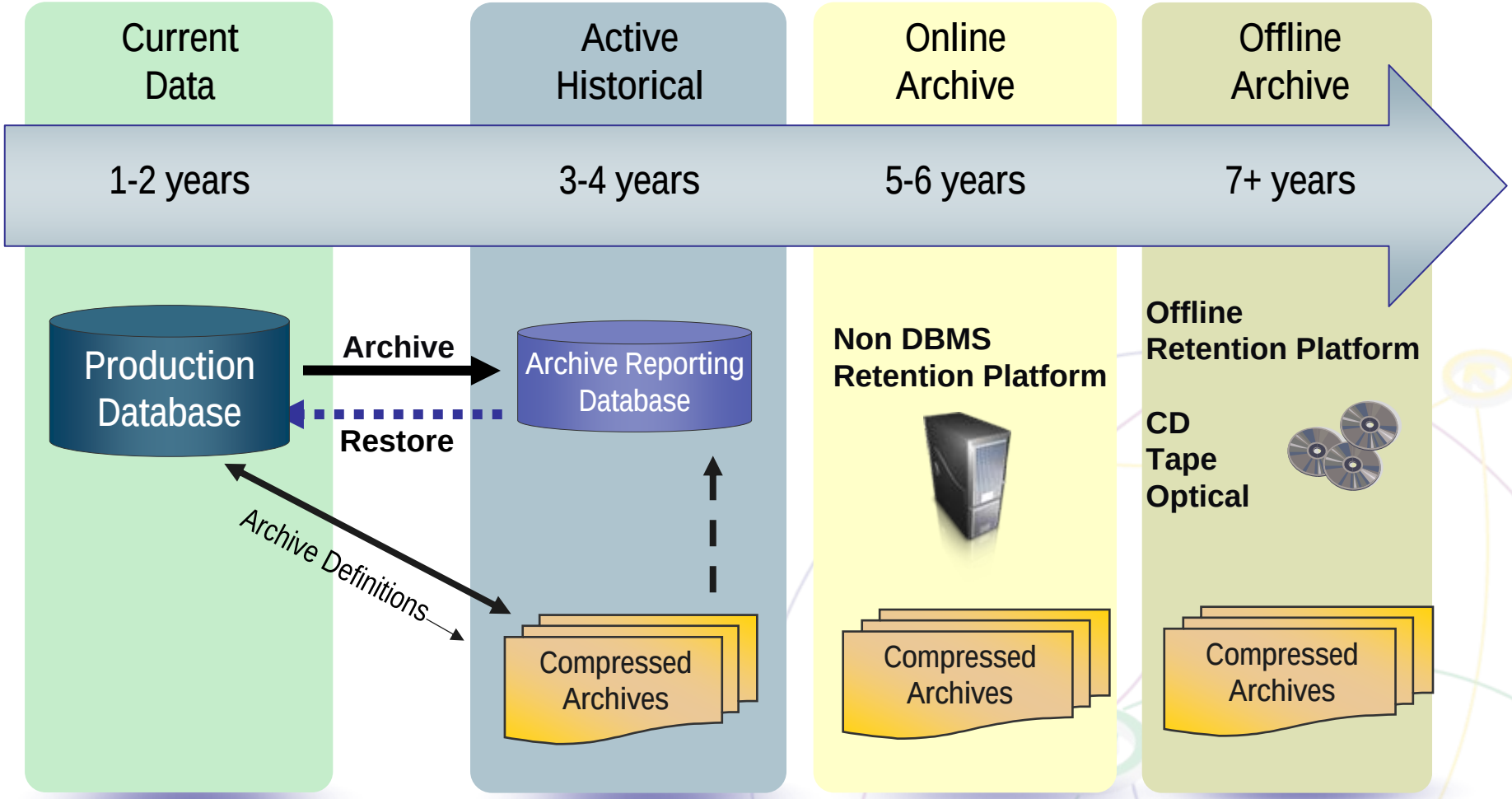
- **Complete Business Object is historical snapshot of activity**
- **Storage device independence enables ILM**
- **Immutable file format enables data retention compliance**

Application Decommissioning

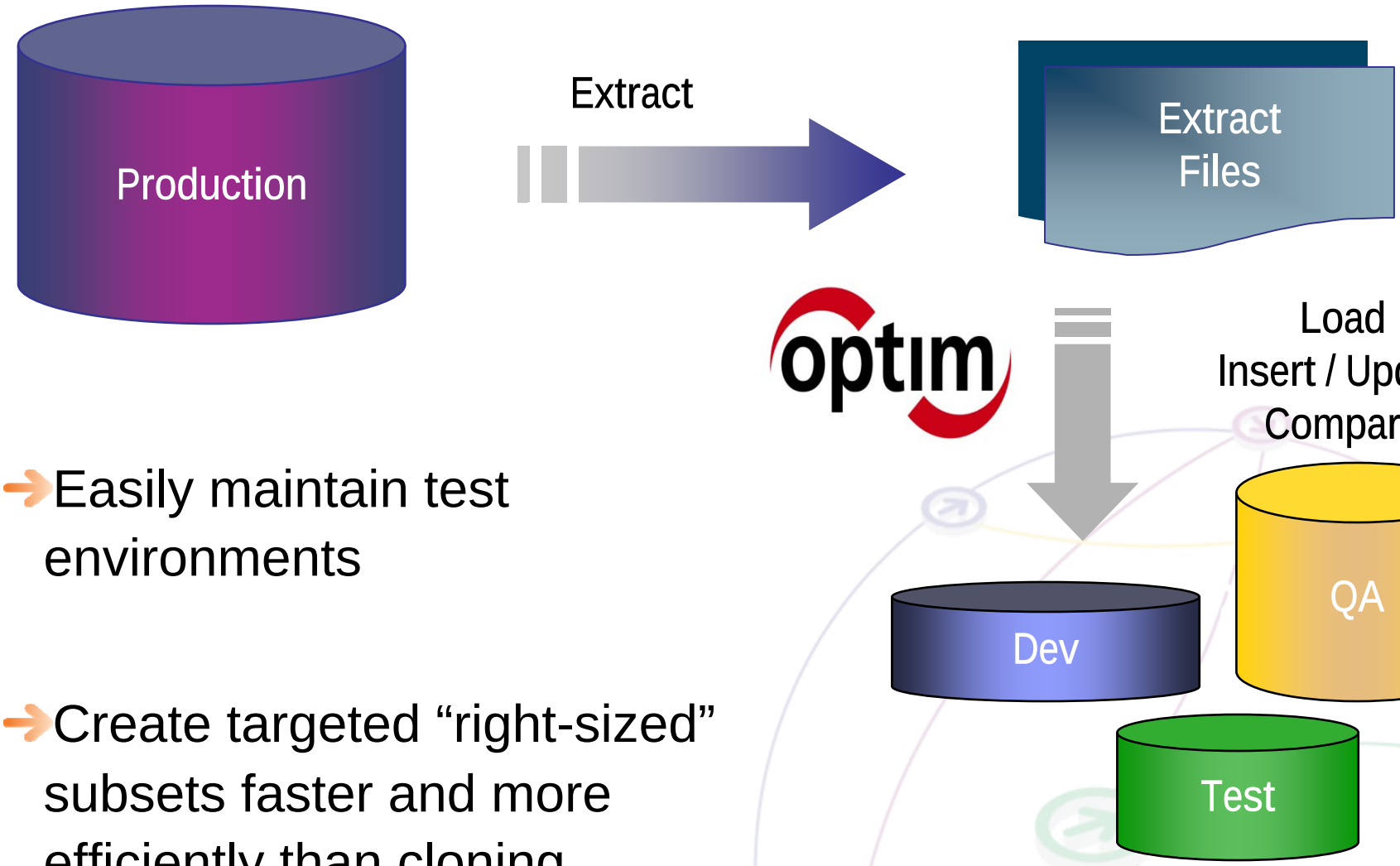


- **Archive your application data before decommissioning**
- **Access this data without the application, or restore at any time**

Optim Enables Information Lifecycle Management

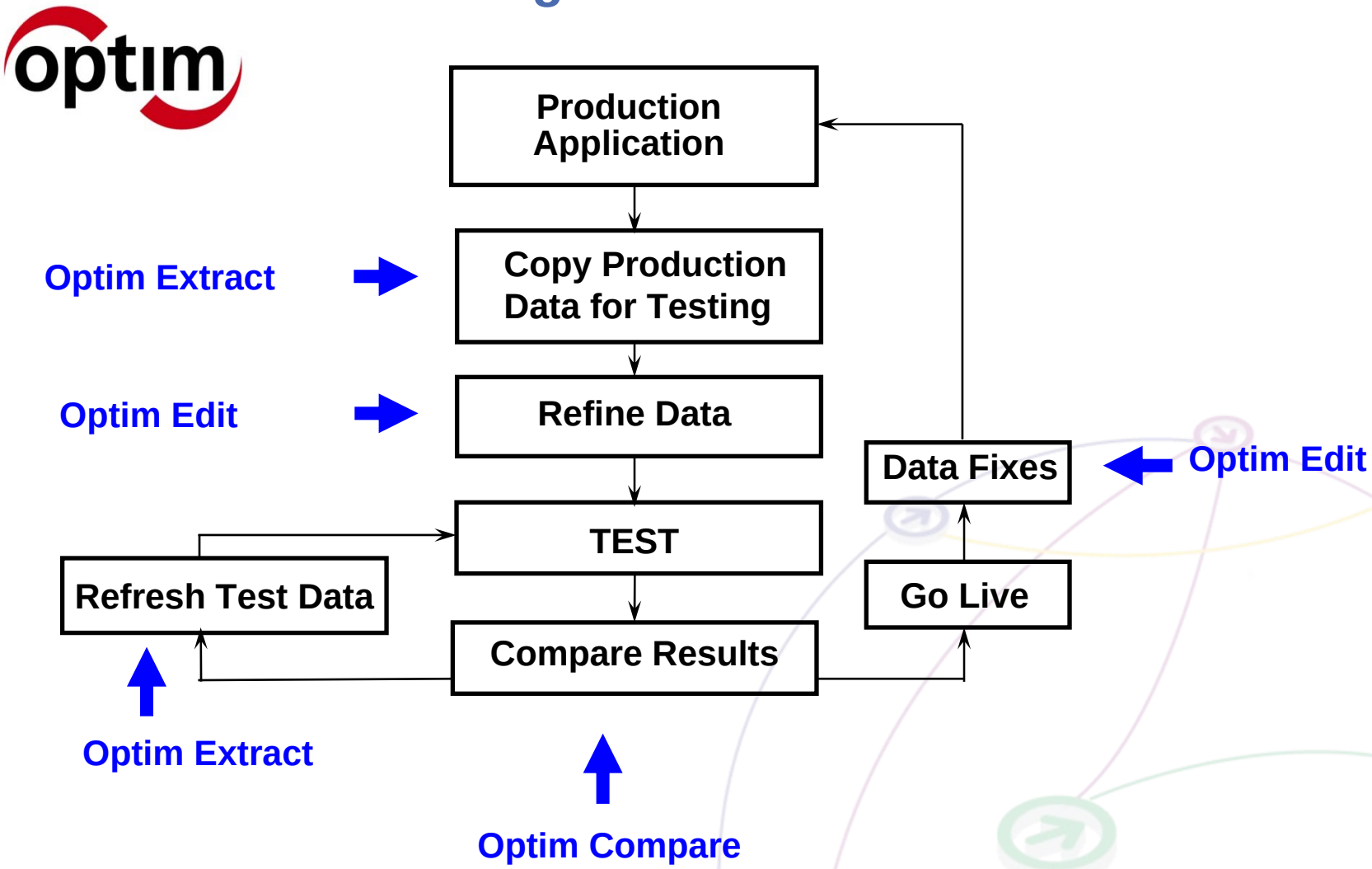


Test Data Management

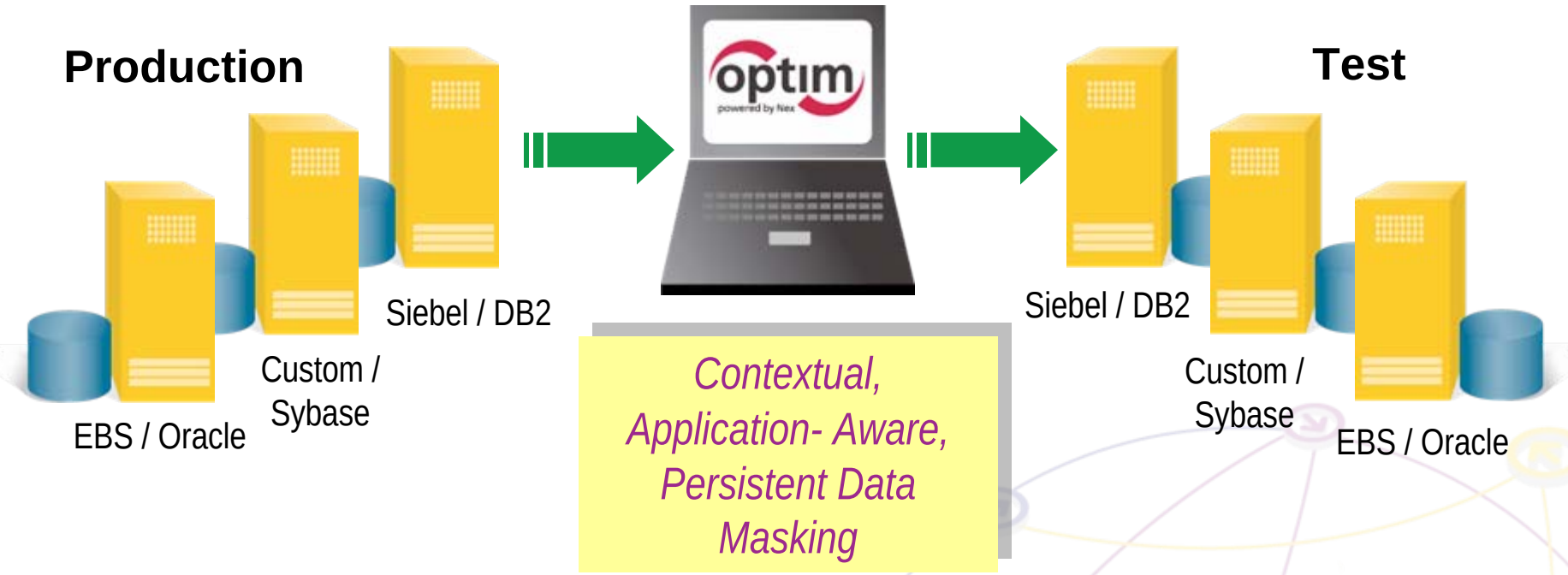


- Easily maintain test environments
- Create targeted “right-sized” subsets faster and more efficiently than cloning

Optim Test Data Management



Optim Data Privacy



- Substitute confidential information with fictionalized data
- Deploy multiple masking algorithms
- Provide consistency across environments and iterations
- Enable off-shore testing
- Protect private data in non-production environments

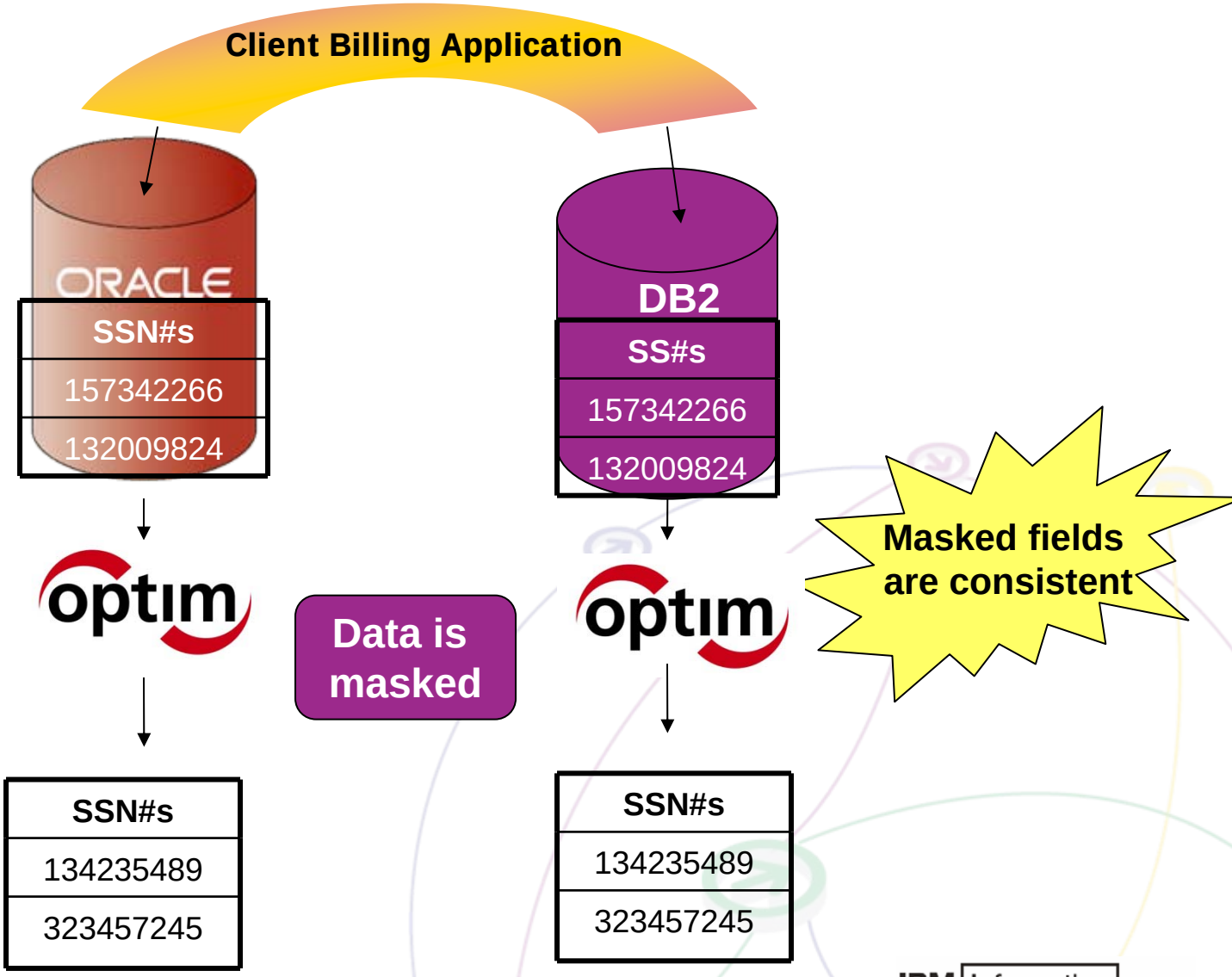
Data Privacy

- A comprehensive set of data masking techniques to de-identify data
- Replaces (masks) confidential data with contextually accurate but fictionalized data

De-Identify test data

- Can Be Performed
  - During Extract Process from DB
  - During Insert/Load Process to DB
  - Or as a Standalone Convert Process
- Transform or mask sensitive data using :
  - Standard rules: Literals, Special Registers, Expressions, Default Values, Look-up tables
  - Intelligent transformation rules: PCI, Addresses etc.
  - Custom mapping rules: user exits
- Converted extract file is safe to share – sanitized data

Consistent mapping Across the enterprise

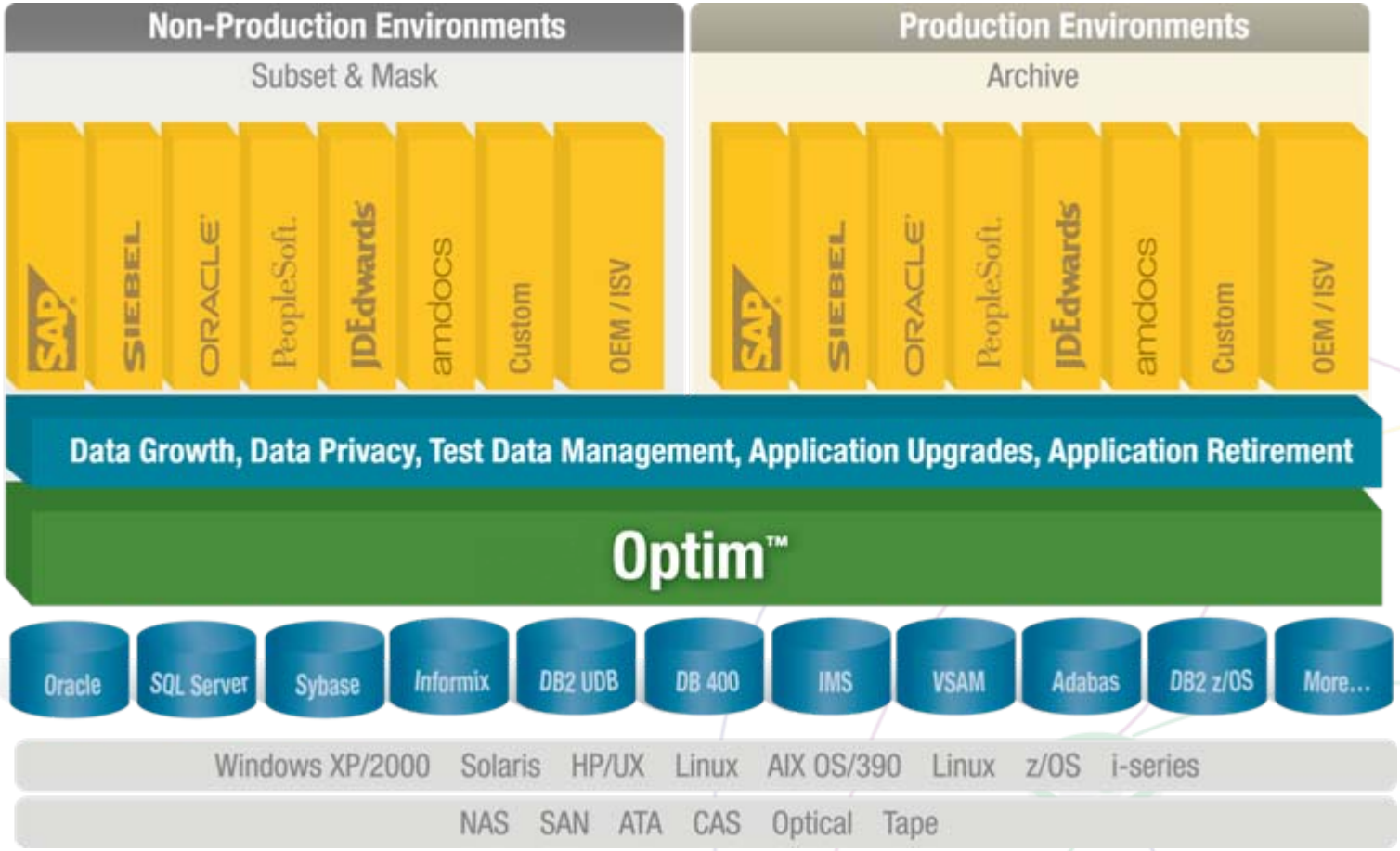


Optim is Designed for the Enterprise

- Built on Enterprise Architecture
  - Support for all popular platforms and data sources
  - Built-in support for many popular applications
  - Scalable
- Complete Business Object
  - A referentially intact subset of data
  - Maintains relationships across your tables, databases, and applications throughout its operations
- Extract, Store, and Restore
  - Clone, subset, move, and restore your data throughout your enterprise
- Universal Access

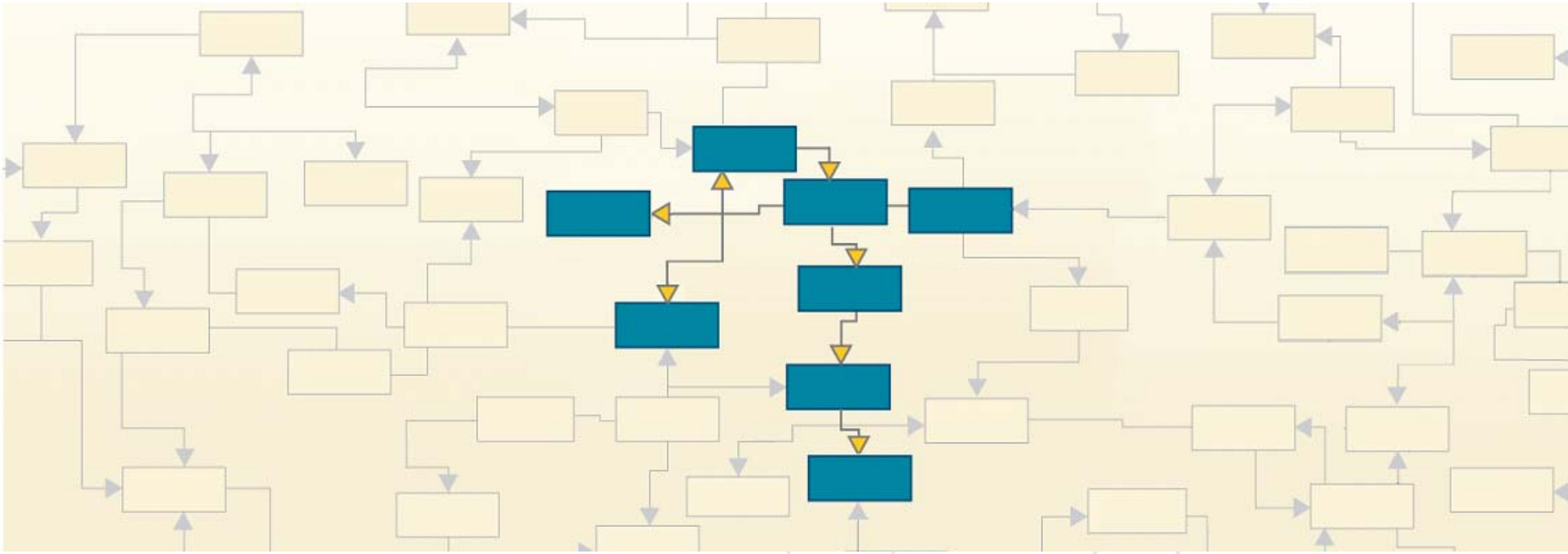
Optim's Enterprise Architecture

- **Single scalable architecture provides central point to extract, store, restore, and protect (mask) application data records**



Complete Business Object

- Referentially-intact subset of data across related tables and applications; includes metadata
- Provides “historical reference snapshot” of business activity



Extract, Store, and Restore

- Extract
  - Identify and extract business objects across multiple related applications, databases, and platforms
- Store
  - Store immutable business objects independent of infrastructure – any hardware device or platform
- Restore
  - Restore complete business objects from any database, application or version to any other database, application, or version
  - Choose exactly what items to restore – a single transaction record, a year's worth of data, or any other range

Universal Access

- Access any record, any time, anywhere!
- Native application access
- Application independent access
  - Industry standard methods (e.g. SQL, ODBC/JDBC, XML)
  - Portals
  - Report writers (e.g. Cognos)
  - Desktop formats (e.g. Excel, CSV, MS Access)
  - Database formats

Optim Products

- IBM Optim Test Data Management
- IBM Optim Data Growth
- IBM Optim Data Privacy
- For Custom Applications or Packaged Application Solutions
  - Amdocs CRM
  - JD Edwards Enterprise One
  - Oracle E-Business Suite
  - PeopleSoft Enterprise
  - Siebel Applications
- IBM Optim Data Privacy for SAP

Optim Terminology

- **Referentially-Complete**

  Optim extracts data based on primary/foreign key relationships (“parent/child”) between tables.

  Handling data this way reduces errors and allows data to be moved without breaking application software

- **Subsetting**

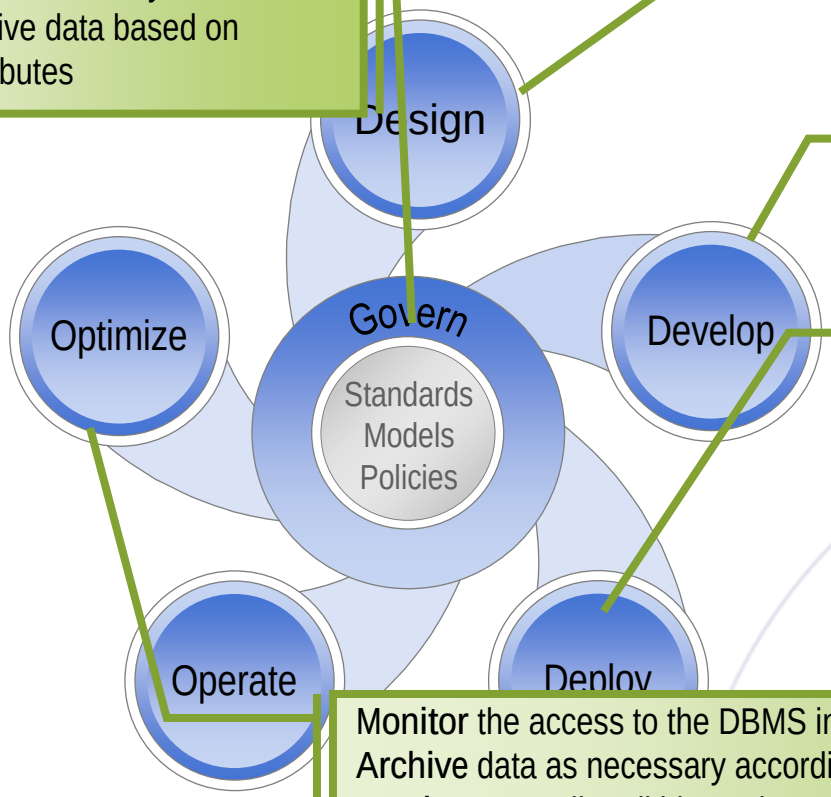
  Using Optim to create a reduced size but referentially complete copy of a database for development or test.

- **Masking**

  Changing sensitive data before testing by replacing it with false but equally valid data.

Integrated Data Management

Data Governance Direction



Audit, **report**, and **alert** any issues with access to sensitive data based on governance attributes

Discover your databases
Discover where your sensitive data is
Define governance attributes in the data model

Mask your sensitive production data when copied to development or test based on governance attributes

Protect the data based on governance attributes
Encrypt the data to address threats outside the control of the DBMS
Control access to address threats from within the DBMS

Monitor the access to the DBMS in real-time
Archive data as necessary according to retention policies
Retain your audit trail history based on governance attributes

data governance attributes

t

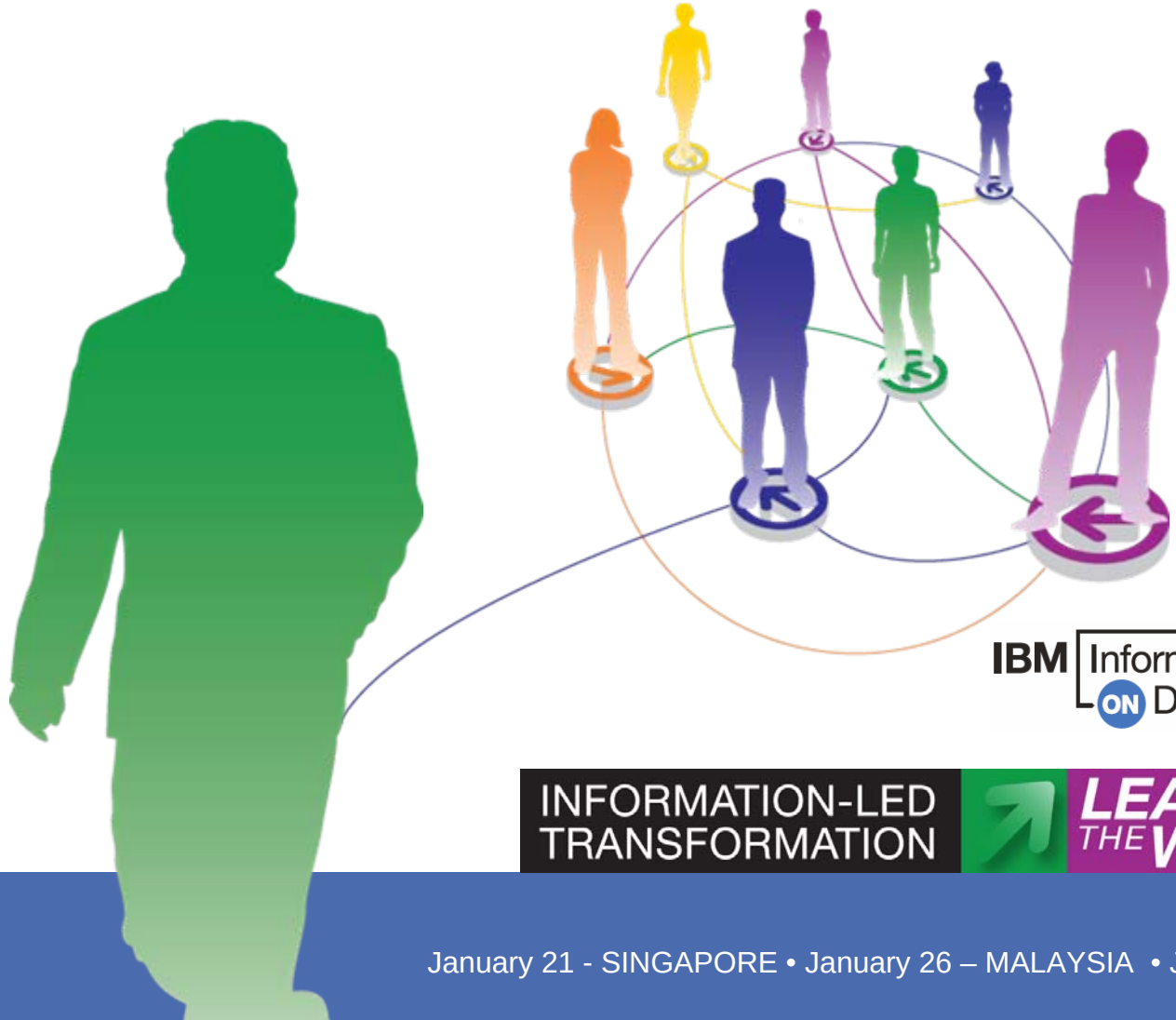
Retention

extend, share,

and implement policies

IBM Optim Solutions

Optim Development Studio and PureQuery



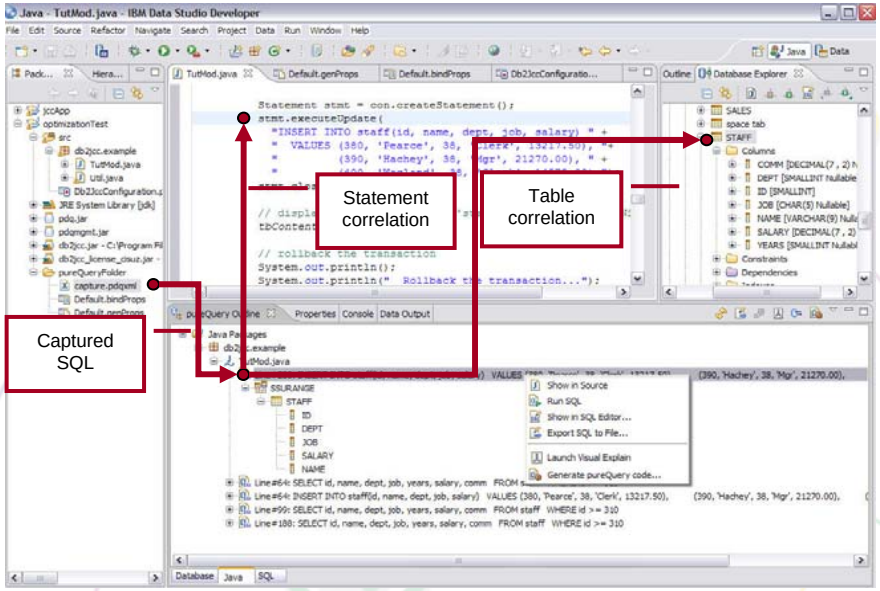
IBM Information On Demand 2010

INFORMATION-LED TRANSFORMATION LEAD THE WAY

Optim Development Studio and pureQuery Runtime

Development Studio is an integrated database development environment that speeds application design, development, and deployment while increasing data access performance and manageability.

- Ease JAVA coding
  - pureQuery code assistance
- Improve predictability and manageability with static SQL
  - Switch dynamic to static SQL without changing code
  - Eliminate SQL Injection Risk by approving which statements can be executed
  - Replace existing SQL with more efficient SQL without changing the source code
- Visualization of JAVA request to SQL code
  - SQL Outline correlates SQL to JAVA code and the associated objects
  - Provide developers with: No. of executions, elapsed time, min./max executions
- Impact Analysis
  - Display lines of code associated to a DB2 object to understand the impact of an object change.
- PureQuery for JAVA, Open JPA and .NET



Problem Statement

High cost of database application development

- No support for SQL development within Java IDE
- No way to generate and customize optimized SQL code and test cases

Improve developer productivity by bridging the gap between data and Java

Unexpected and uncontrolled application behavior

- Unpredictable performance due to runaway, ad-hoc and unexpected queries
- No link between application and executed SQL makes troubleshooting time consuming
- Difficulty meeting Service Level Agreements (SLA) and availability targets

Improved Quality of Service using static SQL

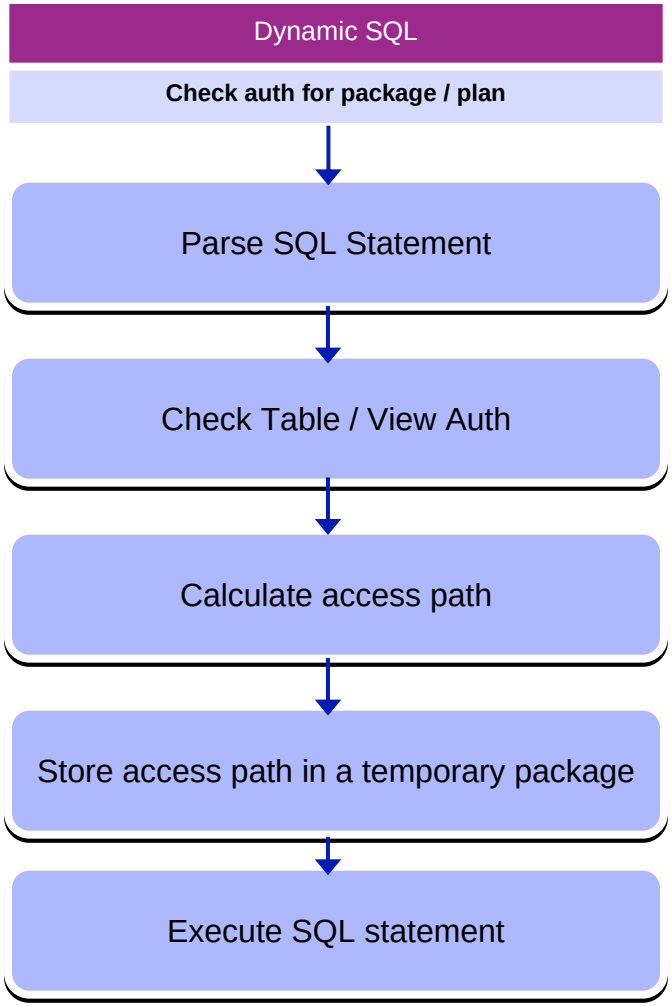
Introducing pureQuery

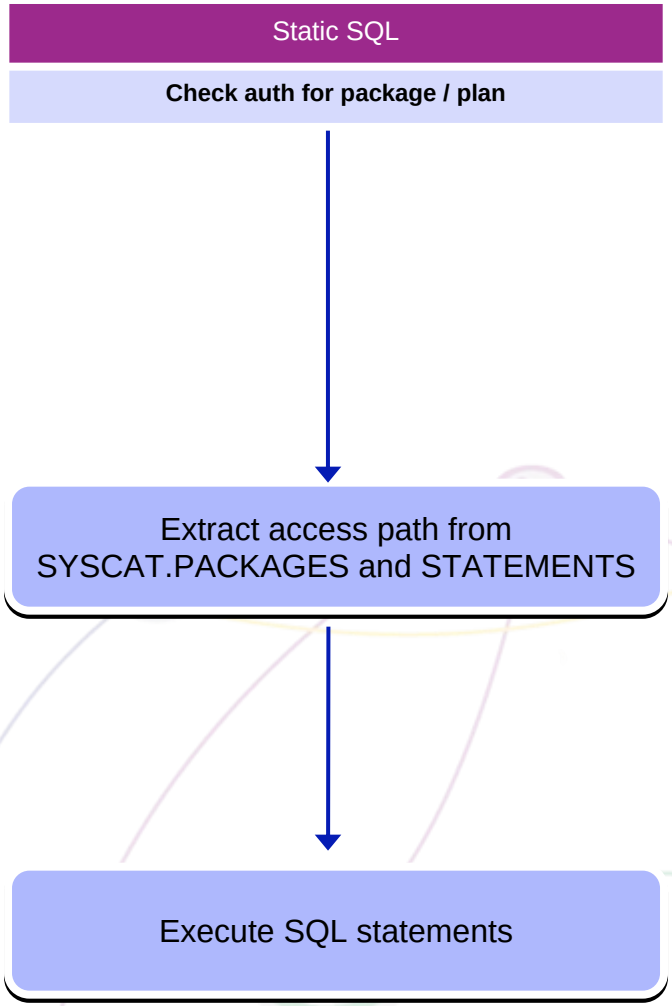
A high-performance, data access platform to simplify developing, managing, securing, and optimizing data access.

pureQuery Components:

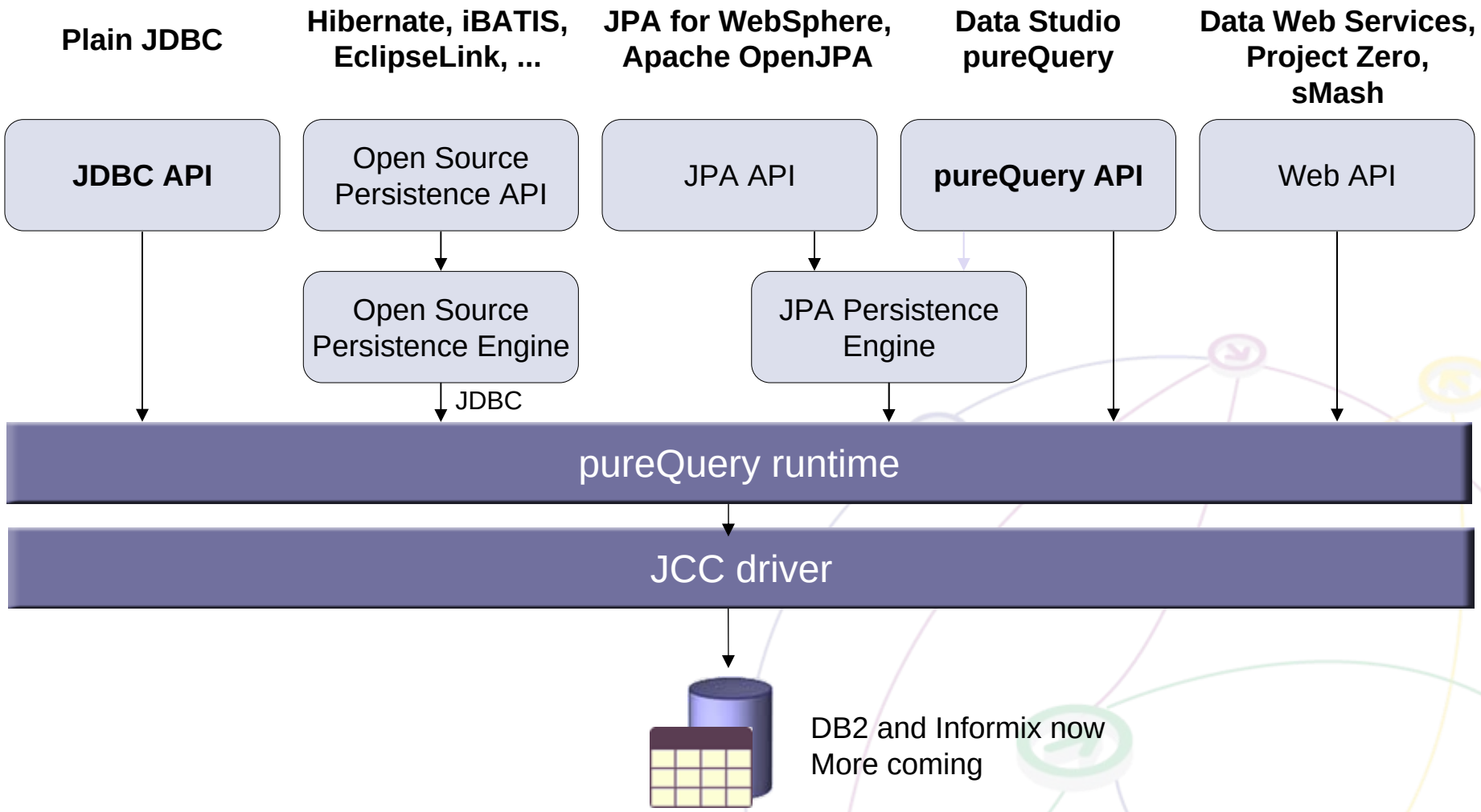
- Simple and intuitive API
  - Enables SQL access to databases or in-memory Java objects
  - Facilitates best practices
- Optim Development Studio
  - Integrated development environment with Java and SQL support
  - Improve problem isolation and impact analysis
- Optim pureQuery Runtime
  - Flexible static SQL deployment for DB2

SQL Execution – Dynamic vs. Static

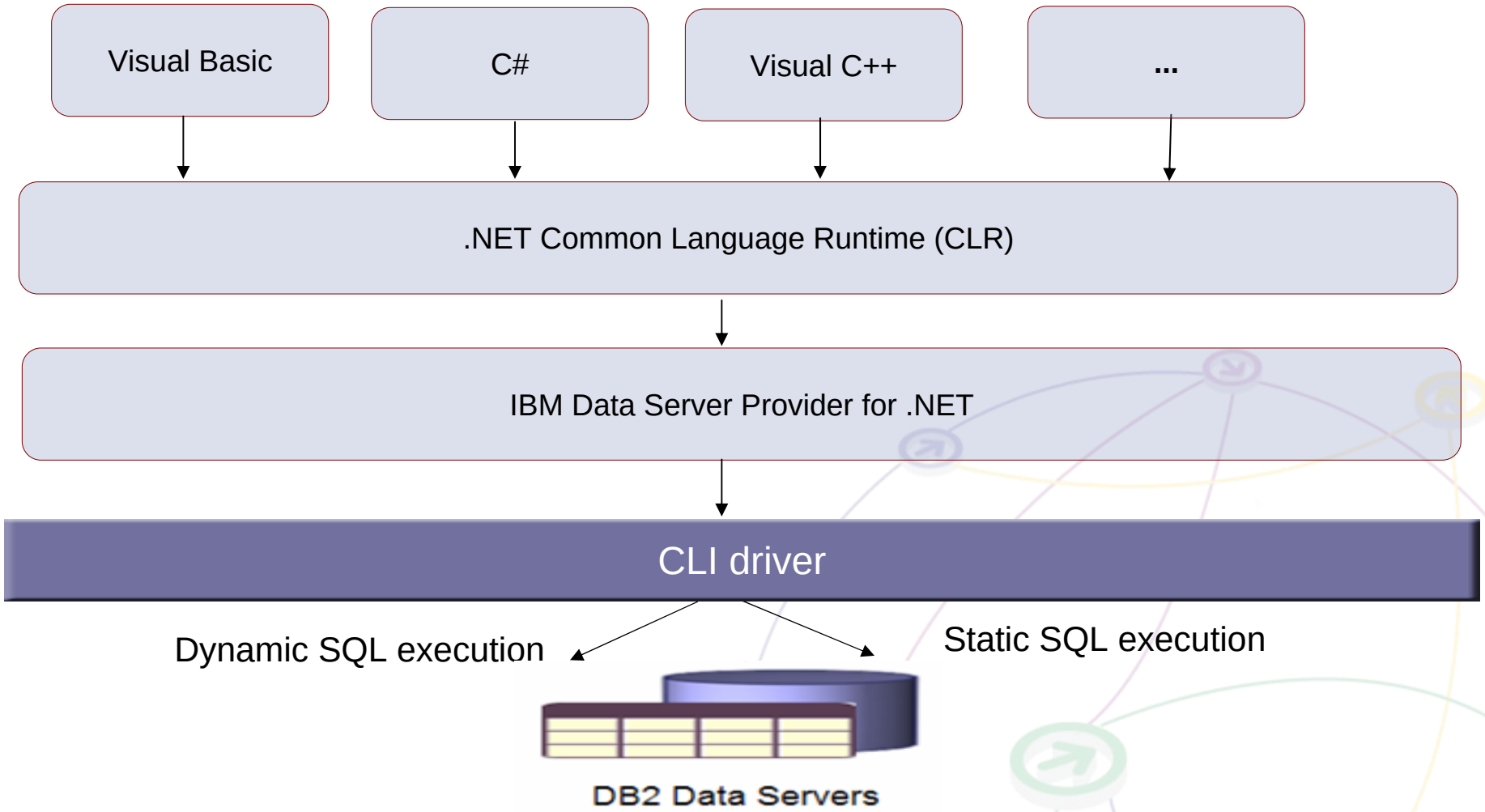




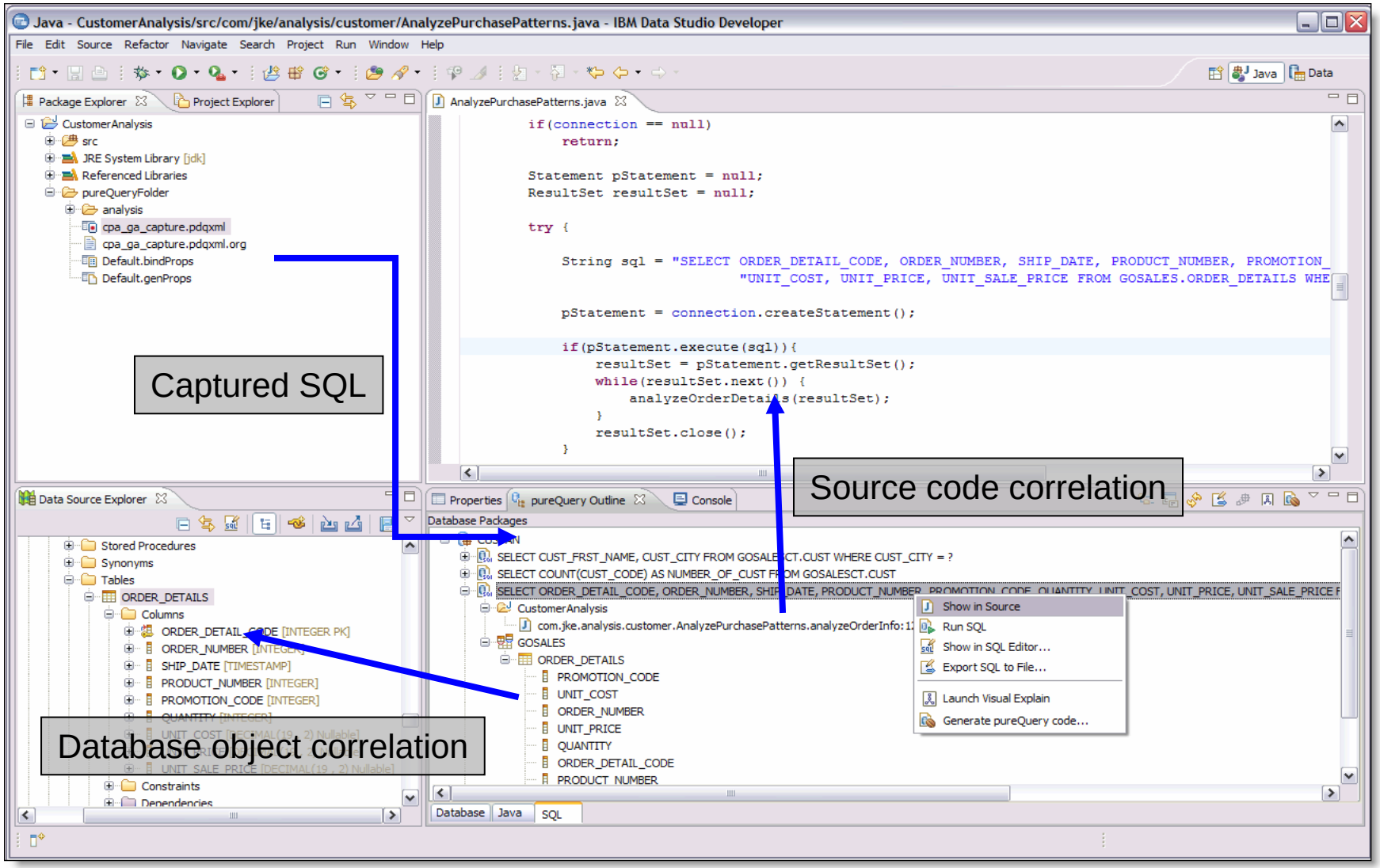
Java Database Access and pureQuery



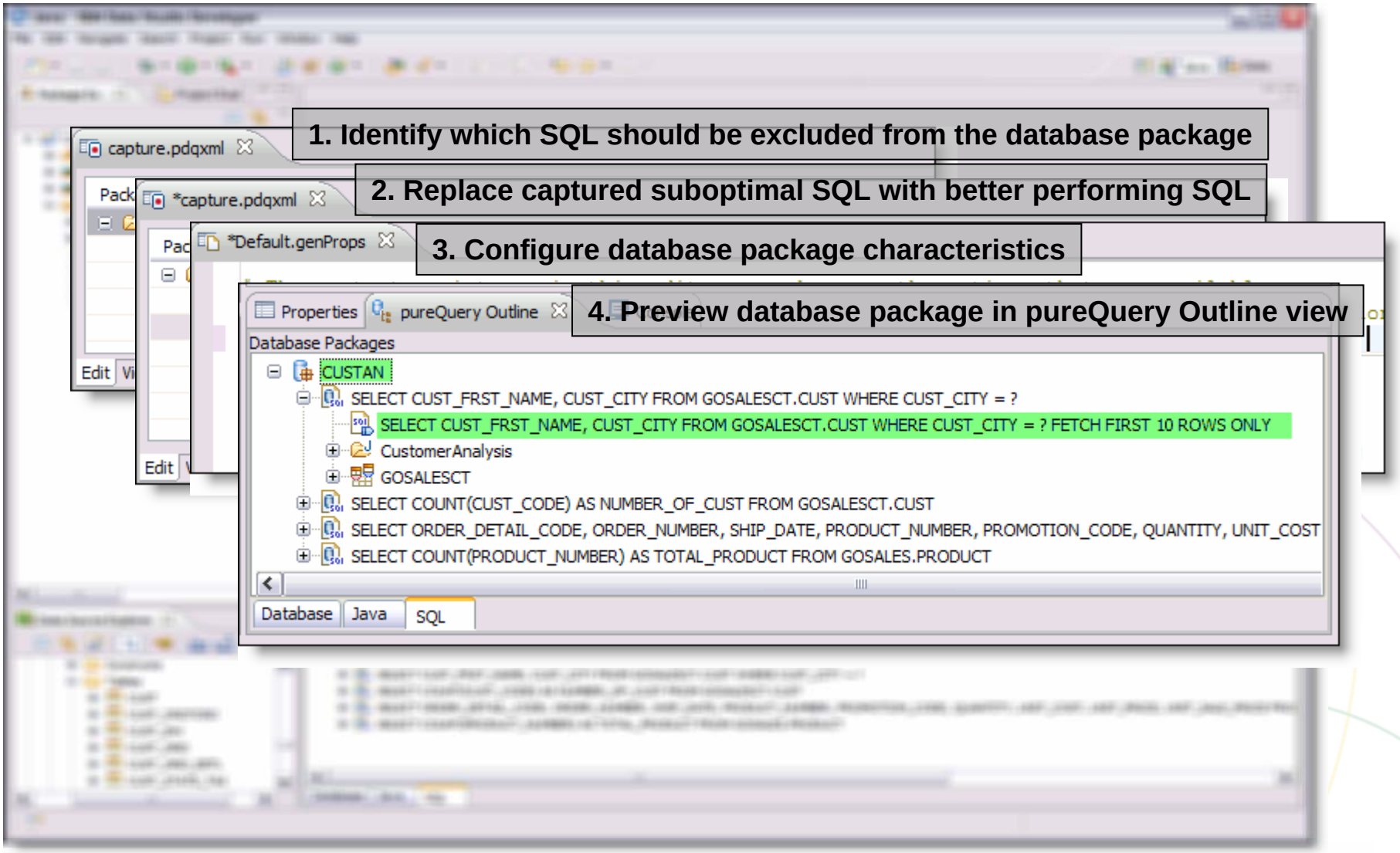
Optimizing .NET Database access using pureQuery



Step 1: Browse the captured metadata



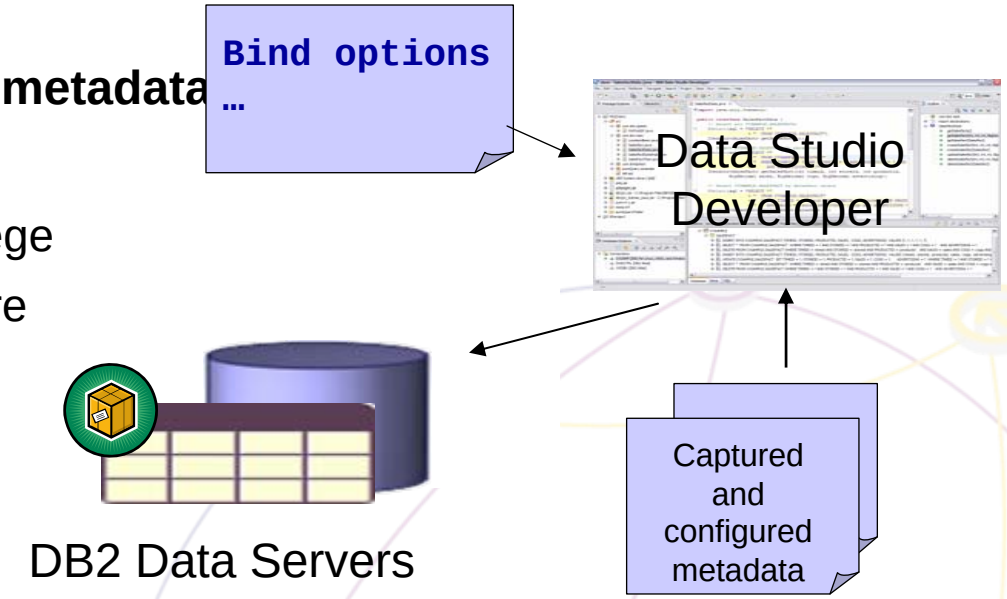
Step 2: Configure database package



Optimize an existing application in four easy steps

Step 3: Bind database package to target database

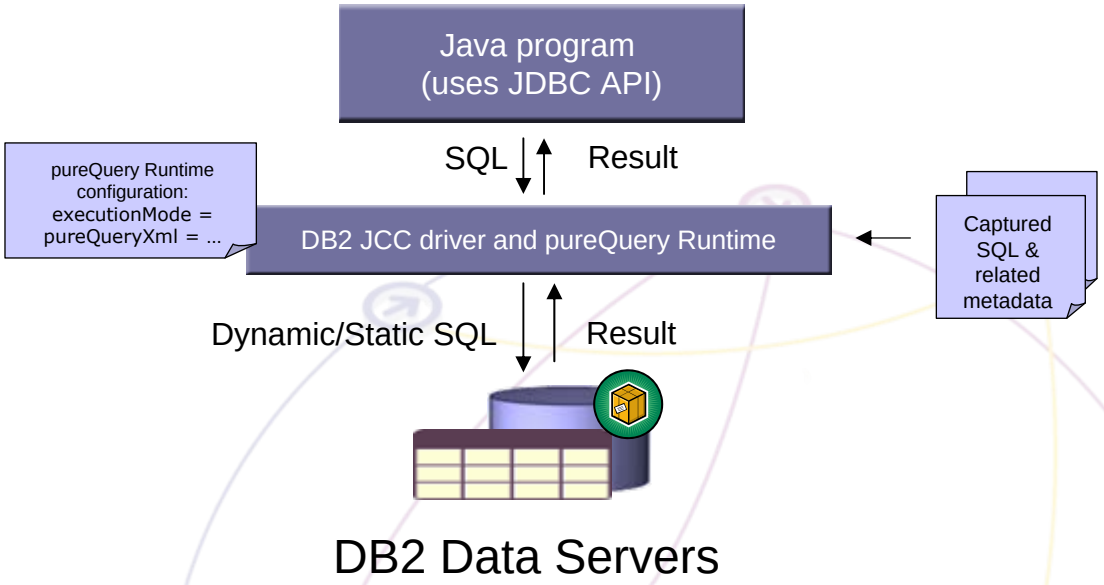
- **Bind captured and configured metadata**
  - Optionally, grant execute privilege for database package to one or more users



Optimize an existing application in four easy steps

Step 4: Execute application

- **Application runtime environment configuration determines SQL execution behavior**
- **Configured metadata file can be loaded via class loader to simplify application deployment**
- **Run SQL statically, dynamically or in mixed mode**
  - Choose execution mode based on life-cycle phase
- **Prevent execution of SQL that has not been approved**
  - Improve security by preventing SQL injection
- **Run applications original SQL or (equivalent) optimized SQL**
- **Improve performance of problem queries without source code changes**



SQL Concentrator

- SELECT employee FROM hr WHERE serial = 123456
- SELECT employee FROM hr WHERE serial = 563293
- SELECT employee FROM hr WHERE serial = 432345
- SELECT employee FROM hr WHERE serial = 243223

- SELECT employee FROM hr WHERE serial = ?

- Lock SQL and protect against SQL Injection
- Convert to static and save CPU cycles
- Save space in dynamic statement cache

pureQuery Improves Performance, Security, and Manageability for DB2 *…Without Changing a Line of Code*

Three steps

1. Capture the SQL
   - Use pureQuery API, generate from WebSphere JPA, or capture while executing
   - Use with custom-developed, framework-based, or packaged applications
2. Bind SQL to DB2
   - Use tooling in Data Studio Developer, WAS console, or command line
3. Choose execution mode
   - Dynamic or static
   - Choose at deployment time instead of development time

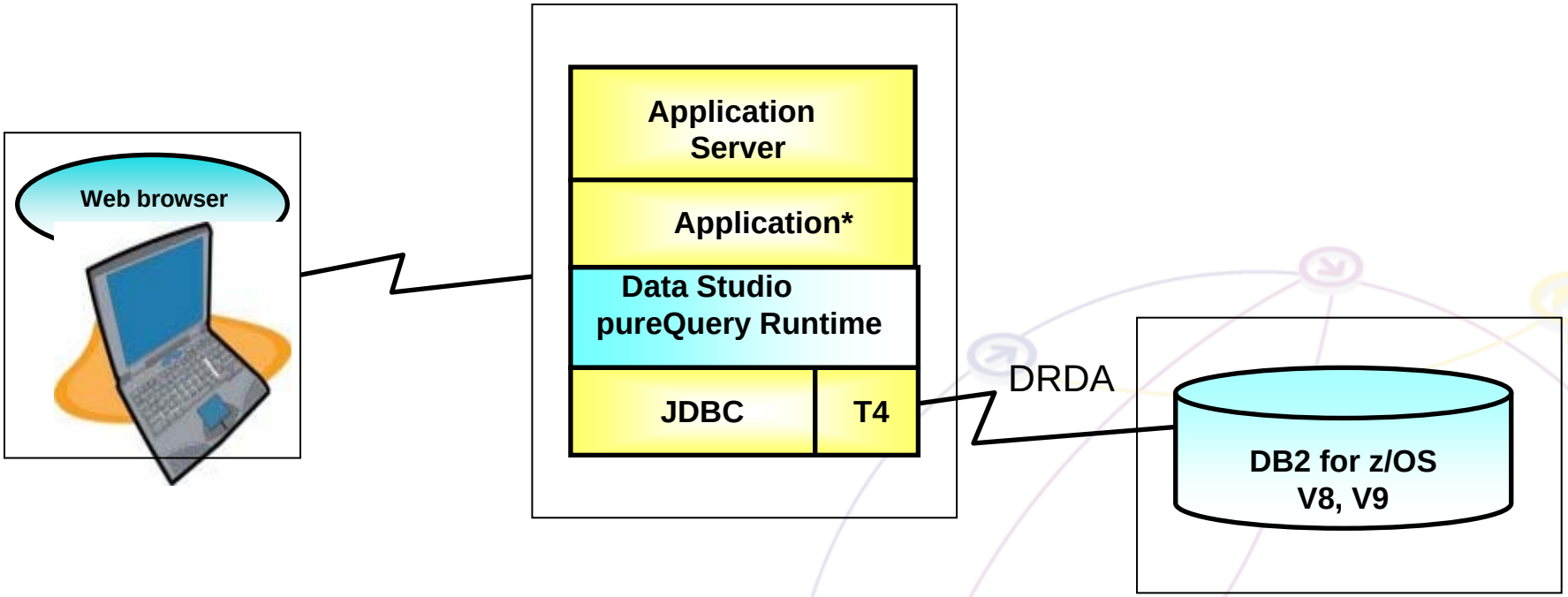
Static SQL value

- Make response time predictable and stable
  - By locking in the SQL access path pre-execution, rather than re-computing at access time
- Limit user access and reduce SQL injection
  - By granting execute privileges on the query packages rather than access privileges on the table
- Accelerate problem resolution
  - By tracing SQL execution to a specific package and the originating source
- Increase system capacity
  - By driving down DB cycles

80-90% of pure Query - Reference Architecture

Configurations Supported:

Application on Linux, UNIX, Windows, or z/Linux accessing DB2 for z/OS

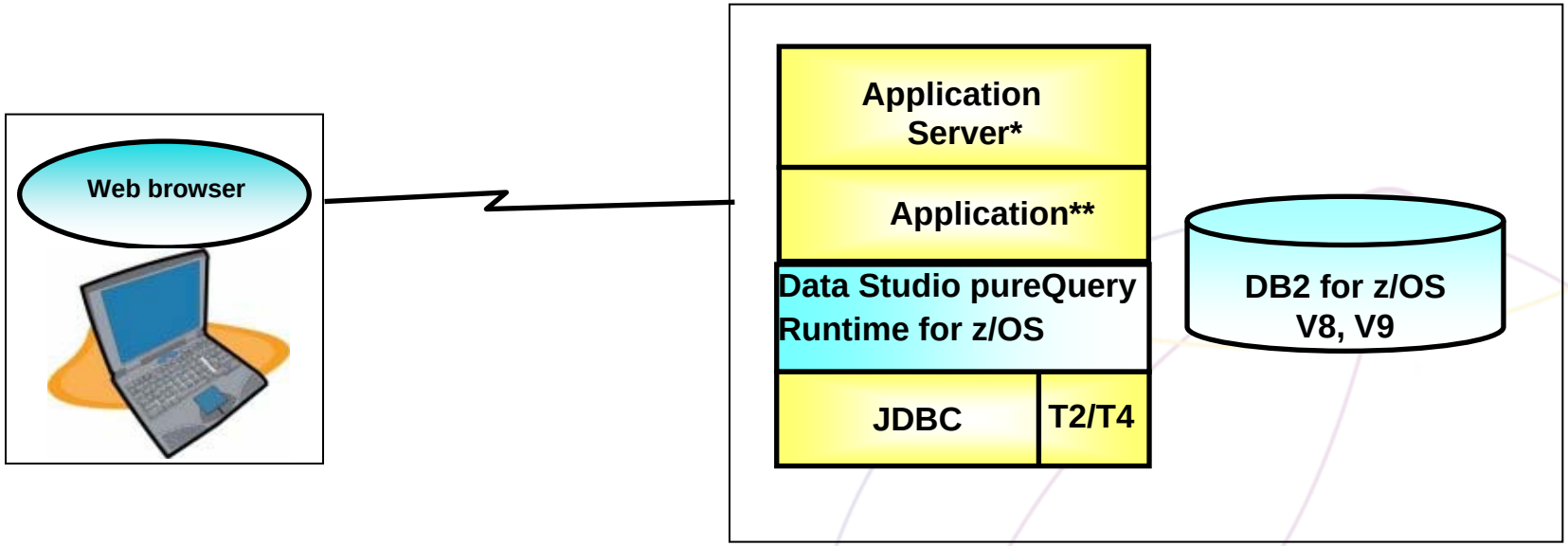


*** - Application running on AIX, Linux, Solaris, Windows and z/Linux**

5-10% of pureQuery - Reference Architecture

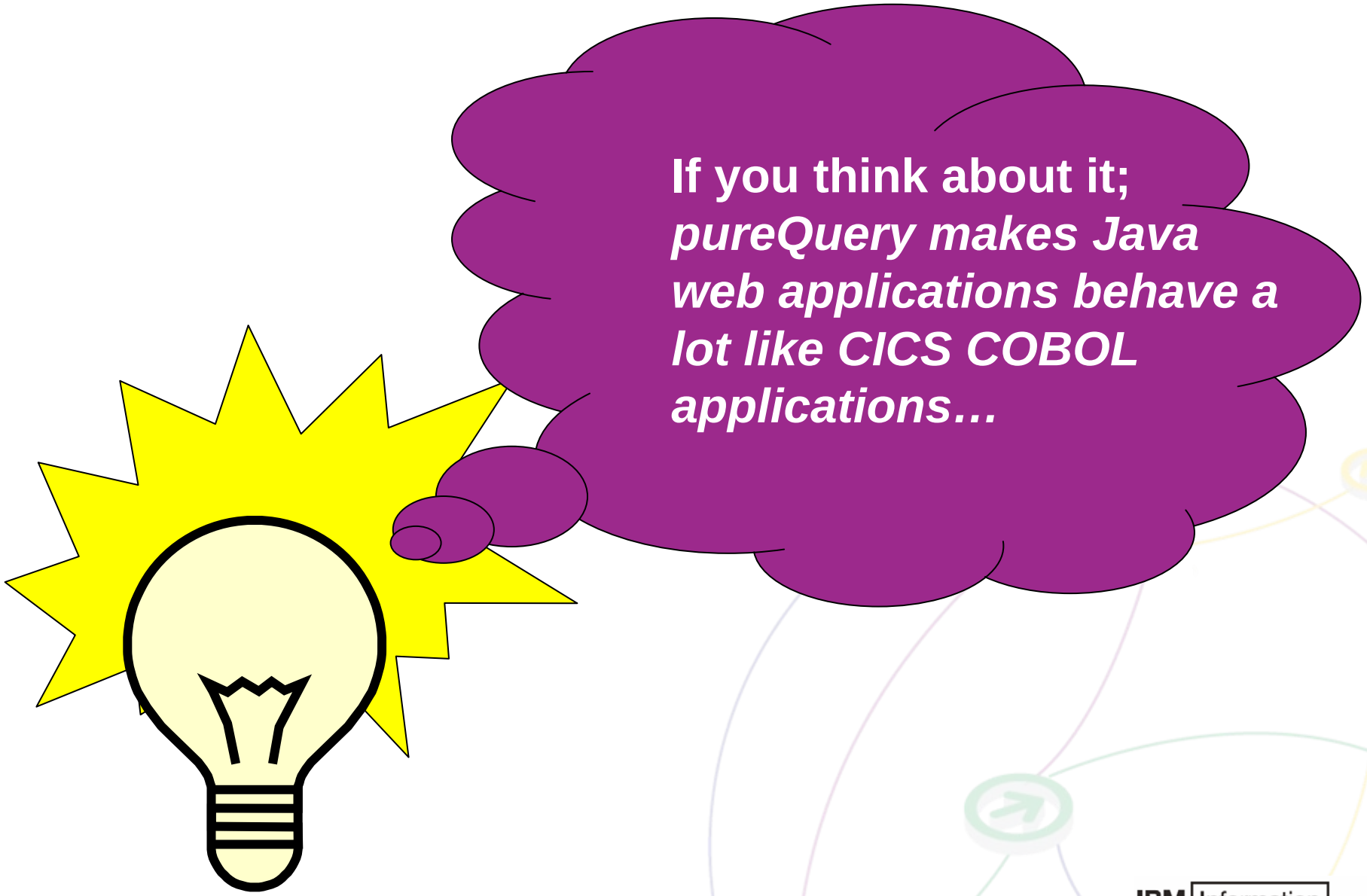
Configurations Supported:

Application running natively on z/OS accessing DB2 for z/OS

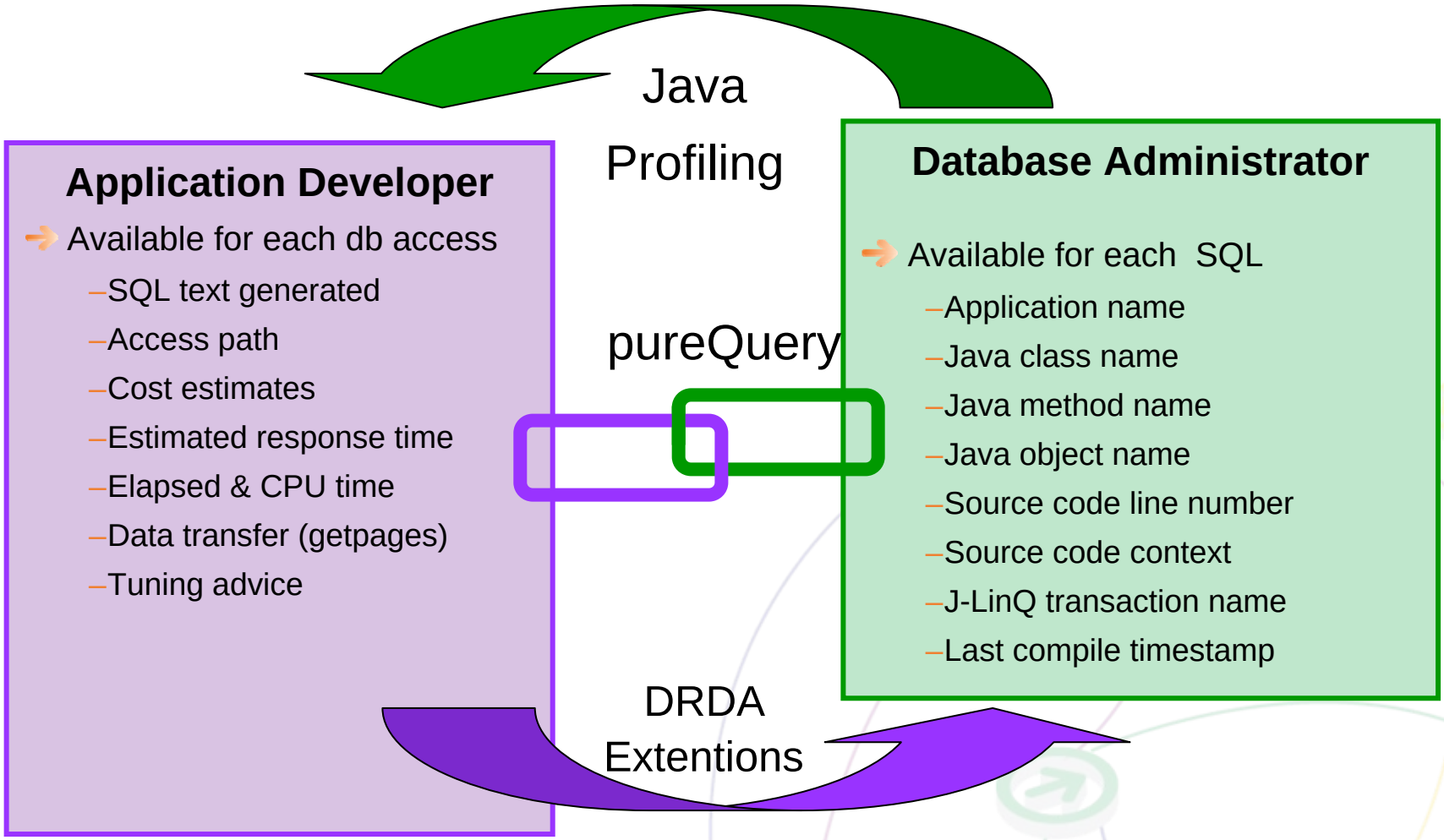


*** - App. Server on z/OS**

**** - Application running on z/OS**

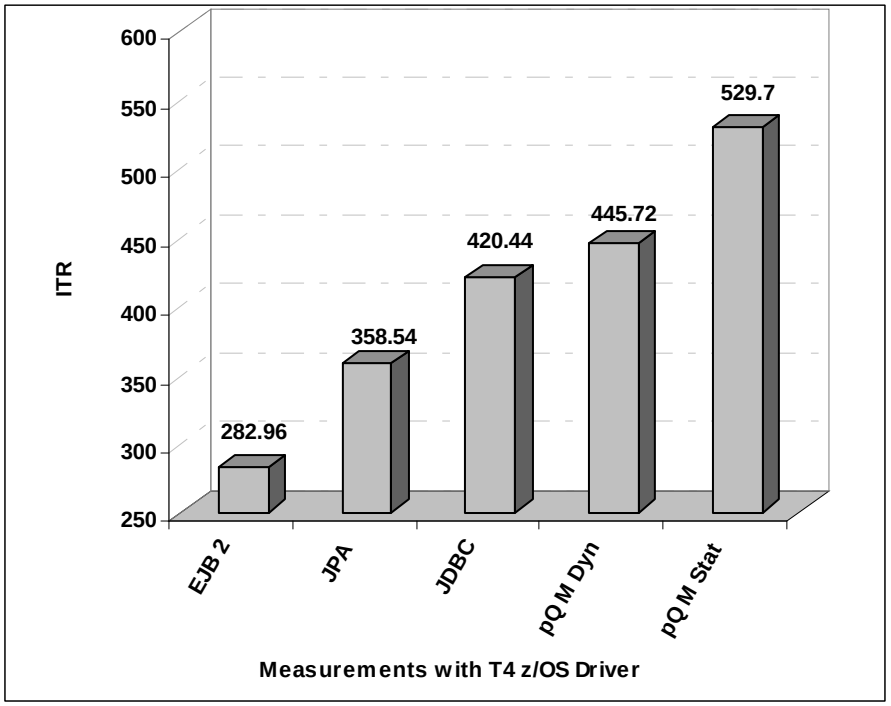

If you think about it;
pureQuery makes Java web applications behave a lot like CICS COBOL applications…

Simplifying Problem Determination Scenario

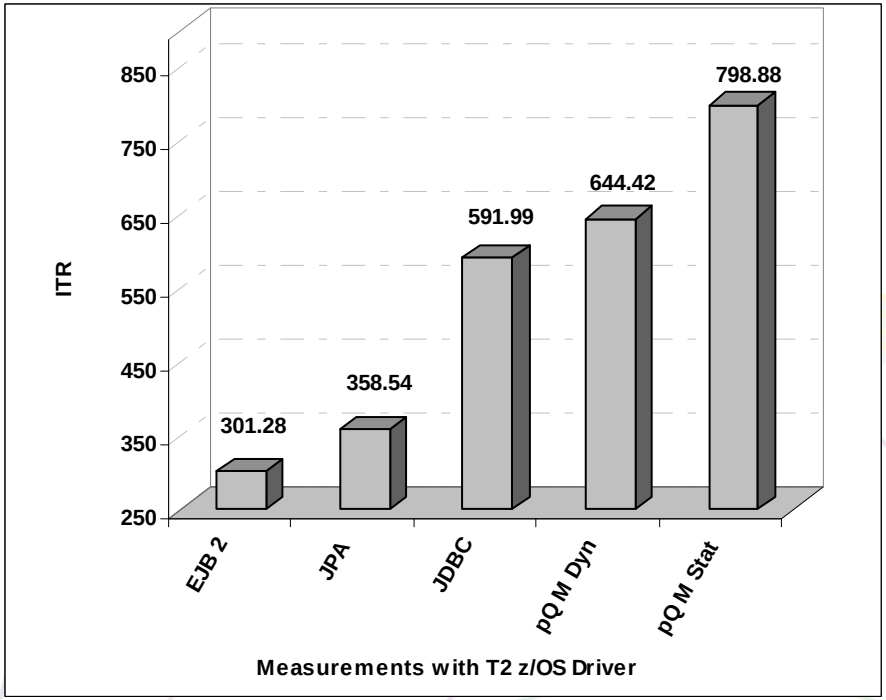


pureQuery improves throughput

(pre-release numbers for v1.2 on IRWW benchmark, ~75% Dynamic cache hit)



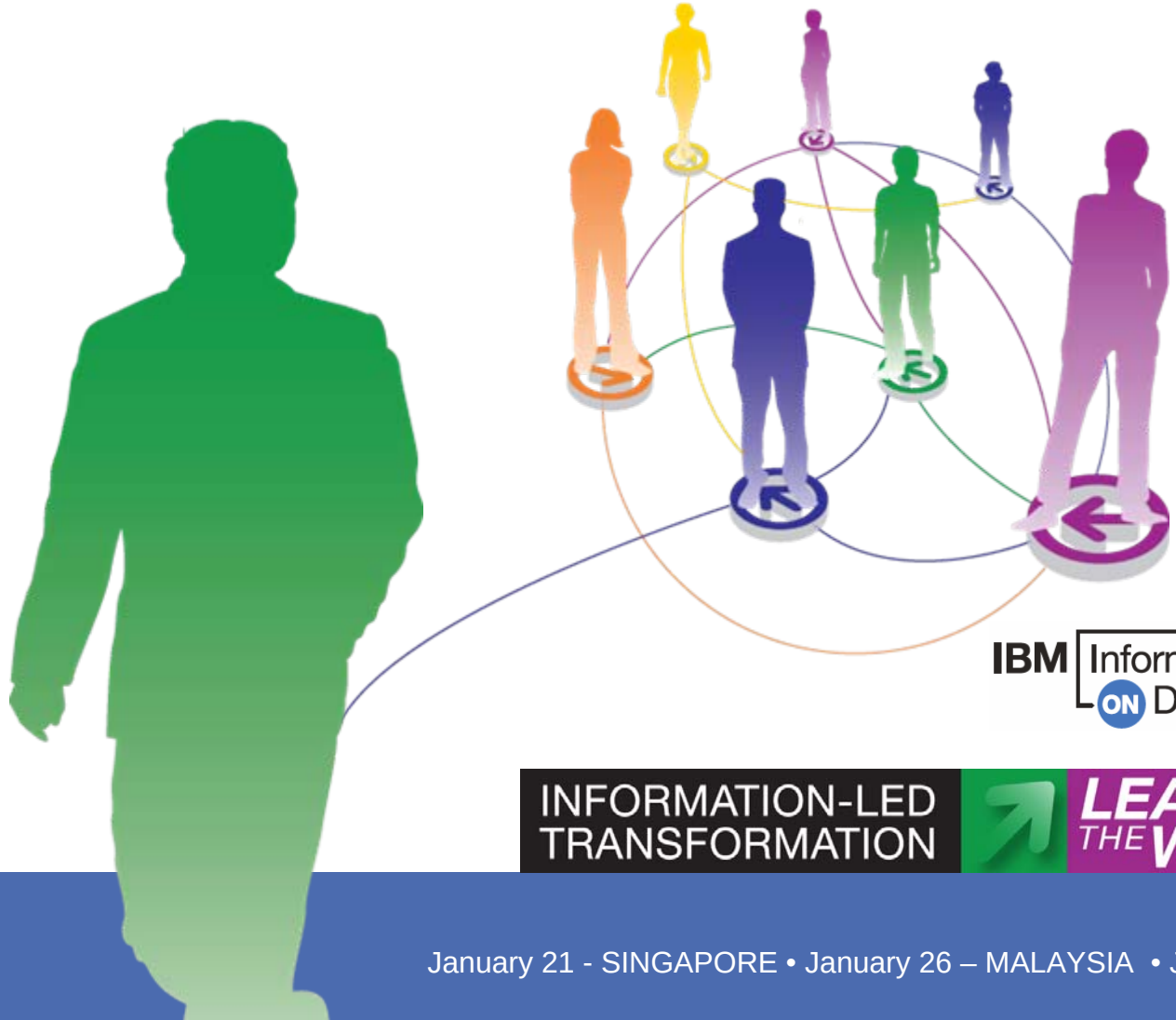
pQ M Static 26% faster than JDBC



pQ M Static 35% faster than JDBC

IBM Optim Solutions

Optim Query Tuner



IBM Information On Demand 2010

INFORMATION-LED TRANSFORMATION LEAD THE WAY

Why is Query Tuning Such a Pain?

- Development
  - Lack of query tuning skills
  - Use of frameworks that generate SQL
  - Inadequate testing environments to drive the workload and data scale
  - Disconnect between application and data groups
- Production
  - Performance problems appear without warning
  - Aggregate information from multiple places
  - Complexity of analysis requires specialized skills
  - Need to consider entire workloads, not just single queries

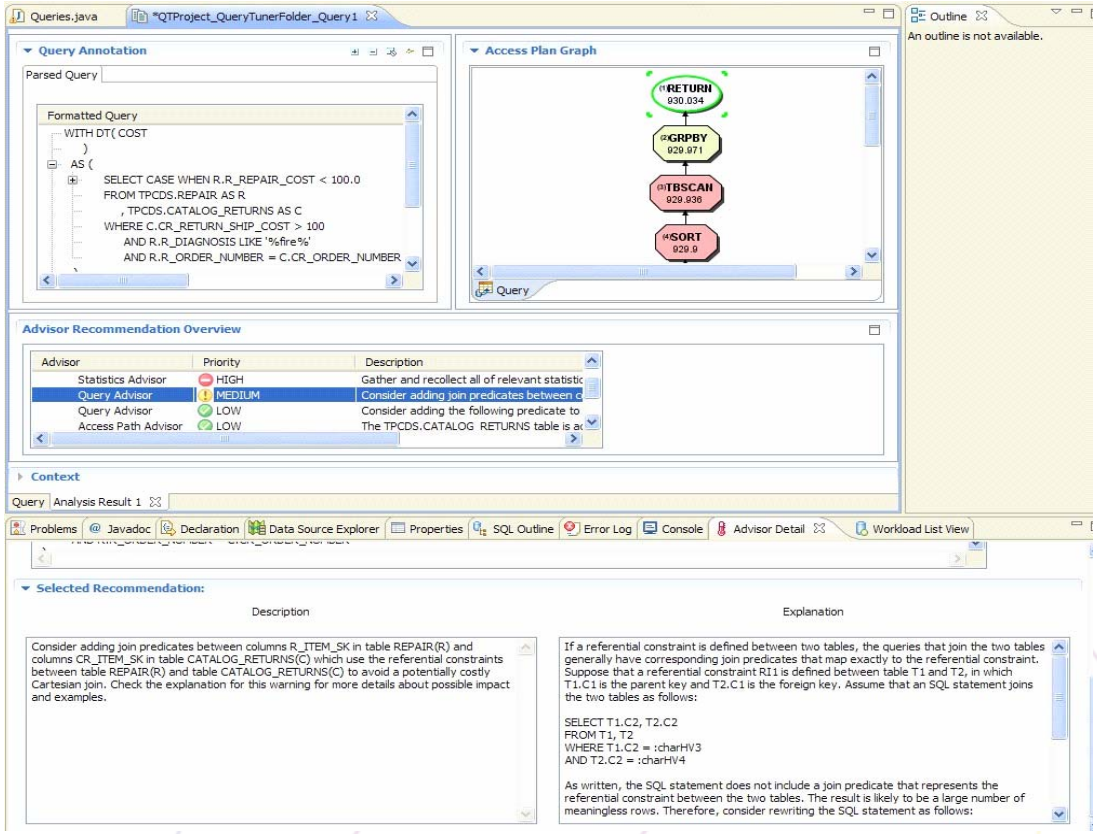
IBM Optim Query Tuner Offerings

Optimize performance and reduce costs

- Optimize performance
  - Improve customer satisfaction
  - Improve organizational productivity
  - Satisfy service level objectives
  - Fit workloads into maintenance windows
- Reduce costs
  - Increase overall system capacity
  - Reduce chargebacks
  - Simplify analysis for DBAs, developers, designers, and others
  - Improve statistics collection efficiency
  - Reduce time to respond

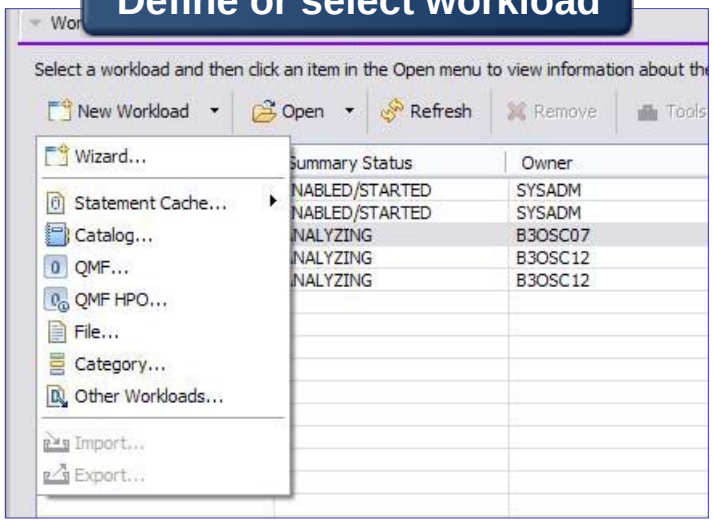
Expert-Enabled Query Tuning Out of the Box

- **Identify query candidates**
  - DB2 catalog
  - Dynamic statement cache
  - Development Studio hot spots
  - Query or performance monitors
- **Facilitate analysis**
  - Query formatting
  - Query annotation
  - Access path visualization and annotation
- **Get expert tuning advice**
  - Improve query design
  - Improve statistics quality
  - Improve database design

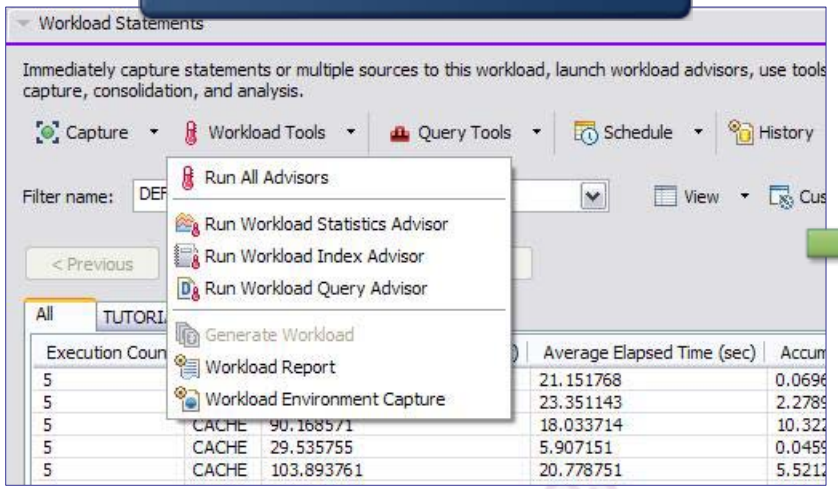


Streamlined Analysis

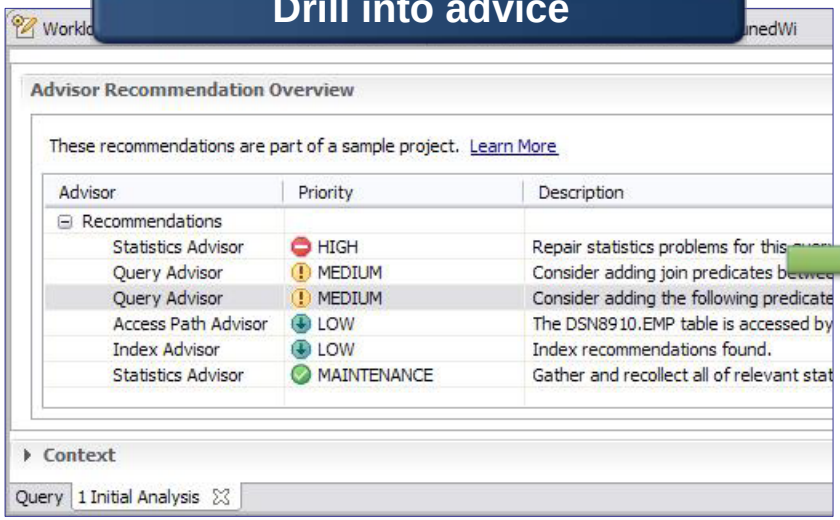
Define or select workload



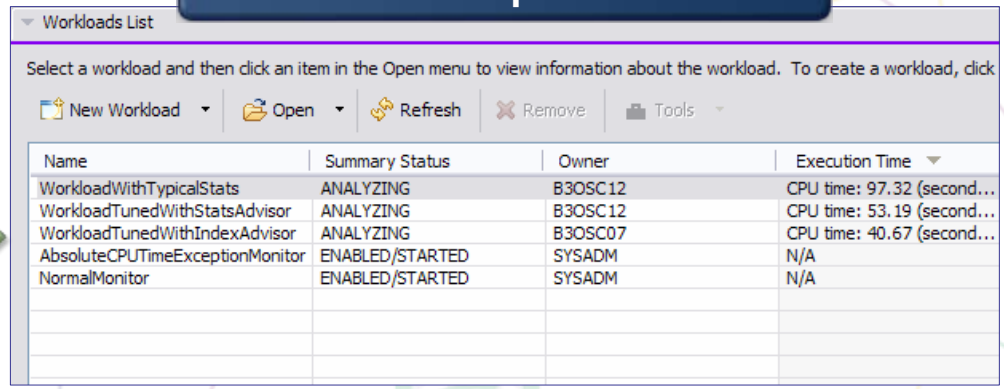
Execute advisors



Drill into advice



Validate improvement



Gather High Cost Queries and Workloads

DB2

- Plan table
- Statement table
- Catalog plan or package
- Statement cache
- Profile monitor

Optim Development Studio

QMF and QMF HPO

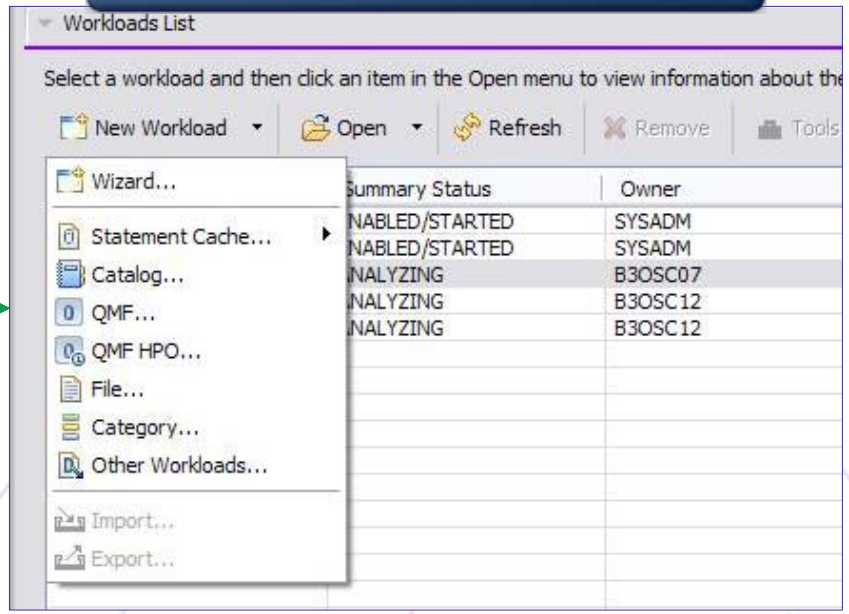
DB2 Query Monitor

OMEGAMON XE for DB2

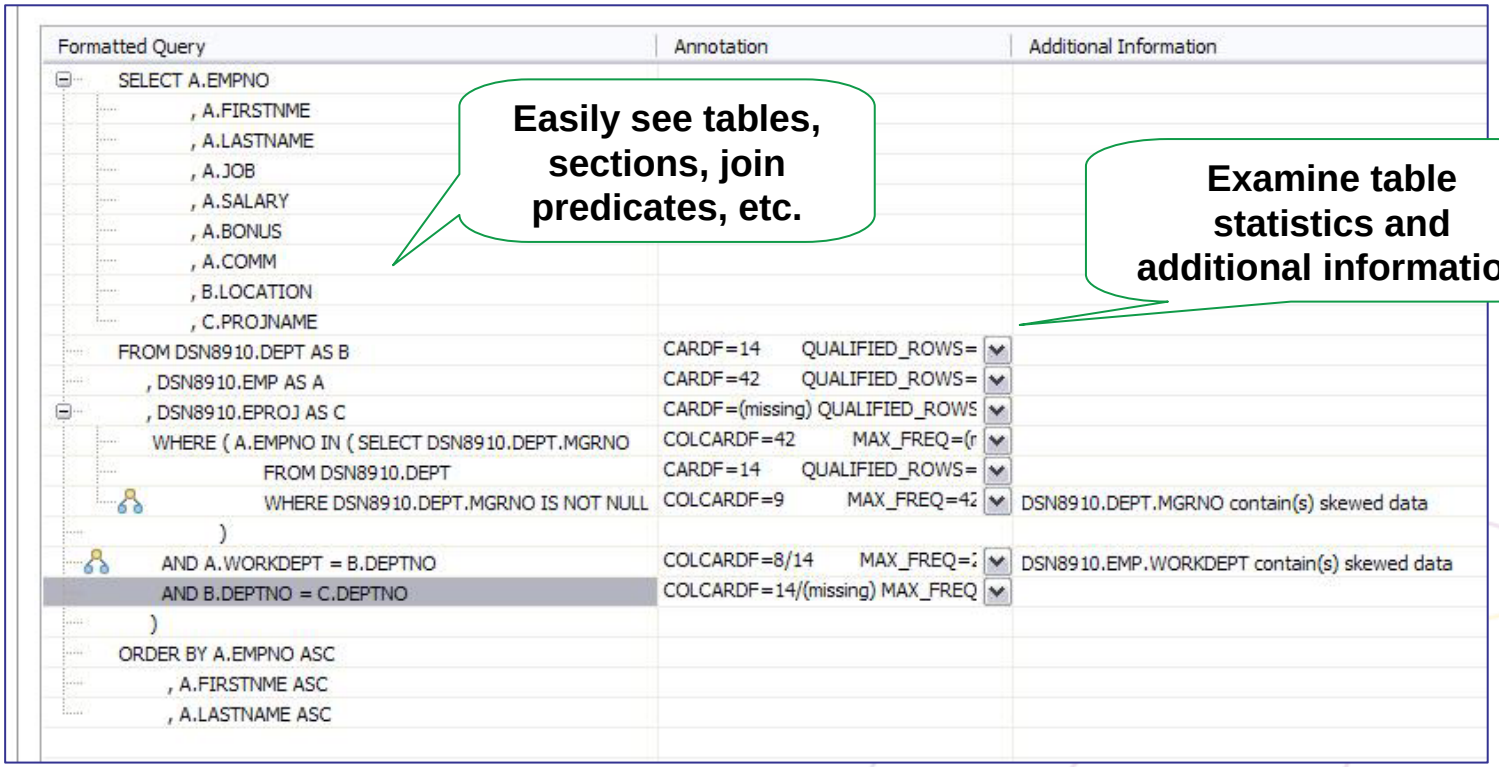
File, text, or exported workload

User defined category

Define or select workload

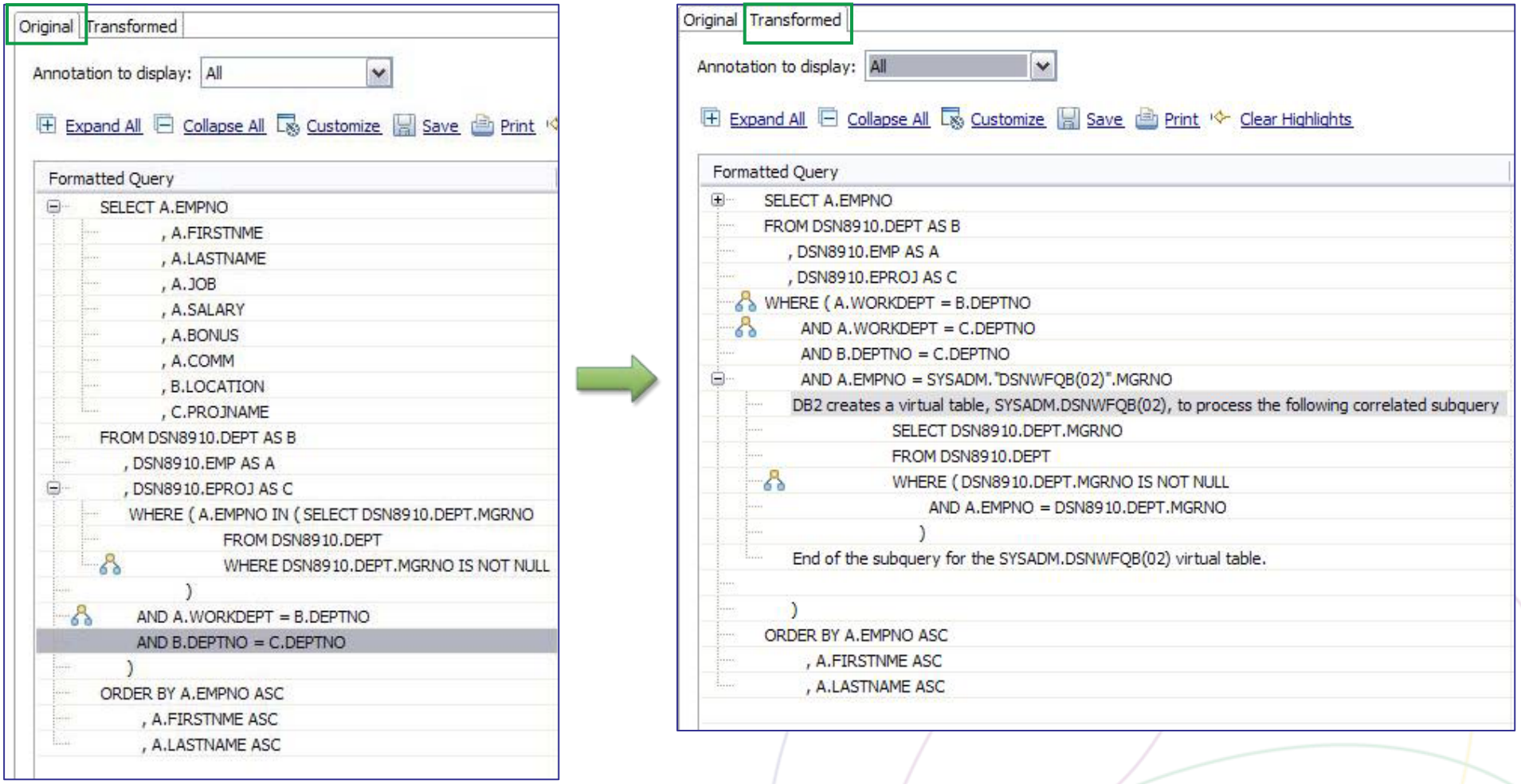


Visualize Queries and Costs to Speed Analysis



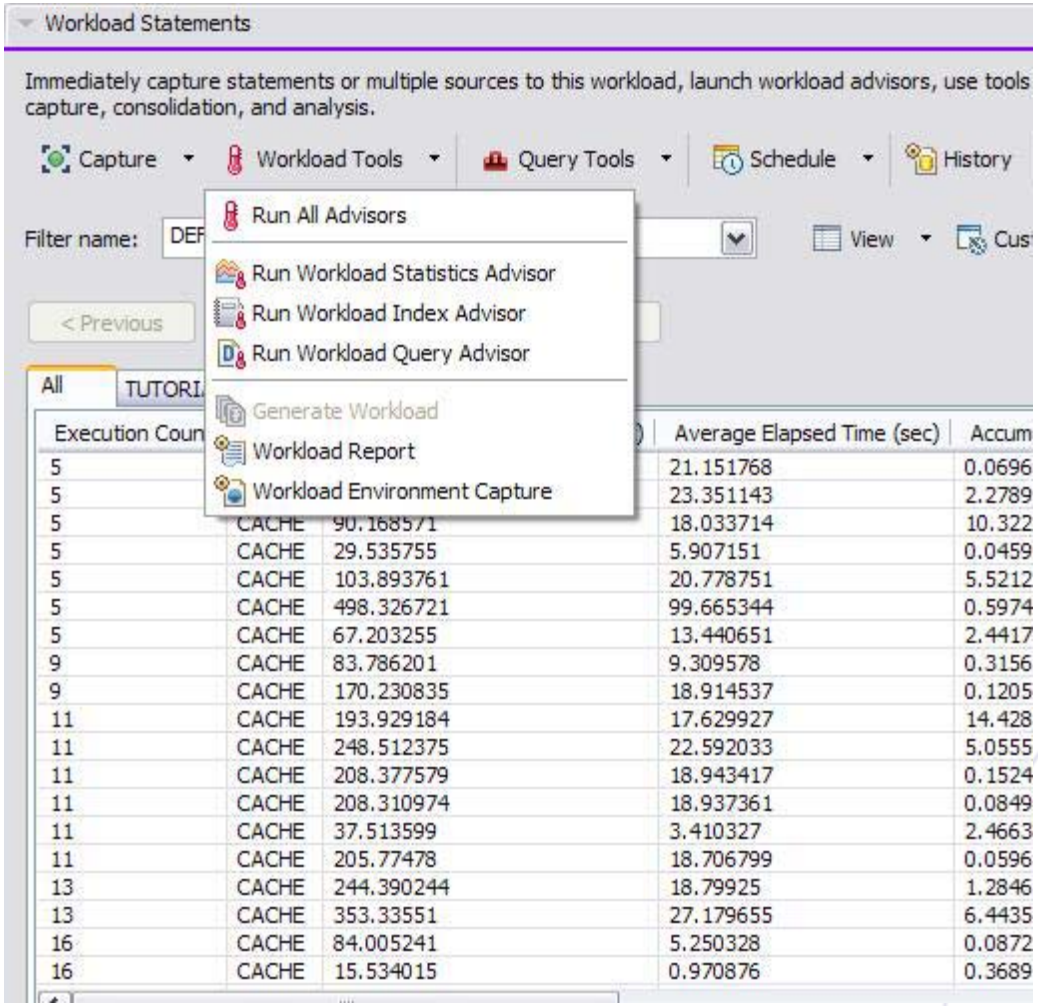
- Accelerate analysis, reduce downtime
  - Spot human errors
  - Identify where filtering should occur

View Optimizer Transformations



- Accelerate analysis, reduce downtime
  - Spot transformations which will occur

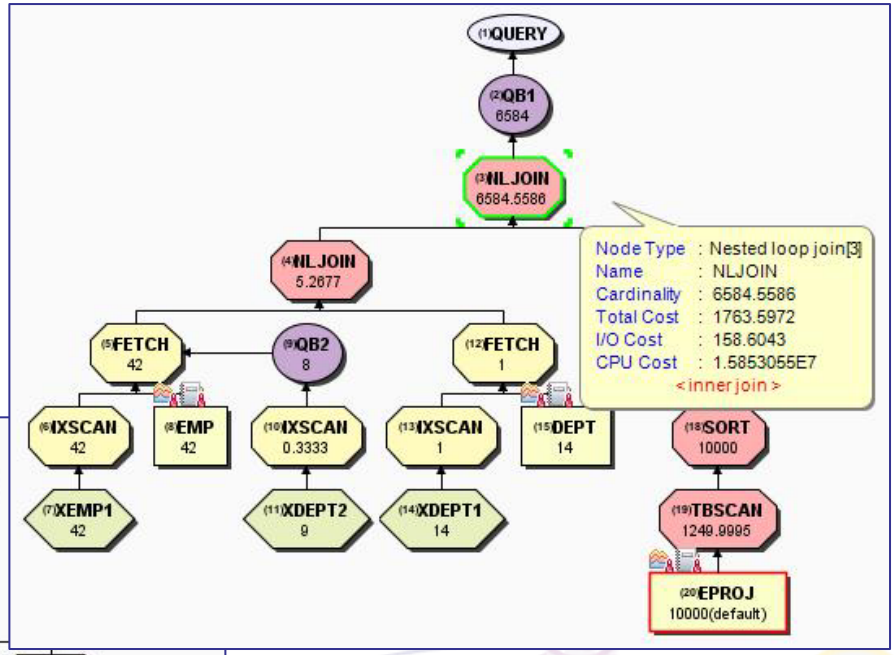
Execute Advisors

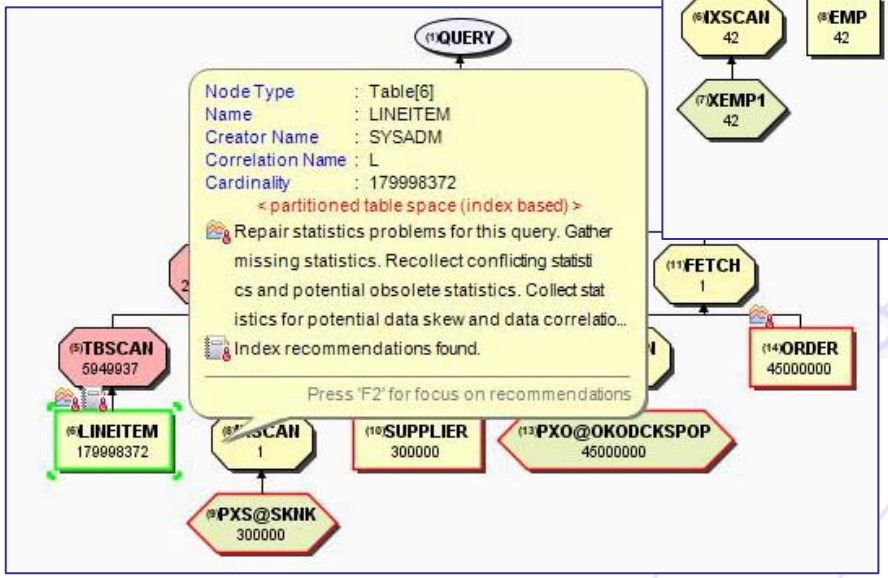


- **Statistics**
  - Get recommendations on the best statistics to capture to influence access path selection
- **Query**
  - Get recommendations regarding how to rewrite the query for better efficiency
- **Index**
  - Get recommendations on additional indexes that can reduce database scans

Analyze Access Plans

- Visualize access path
  - See flow of query processing
  - See indexes and operations
  - See optimizer rationale



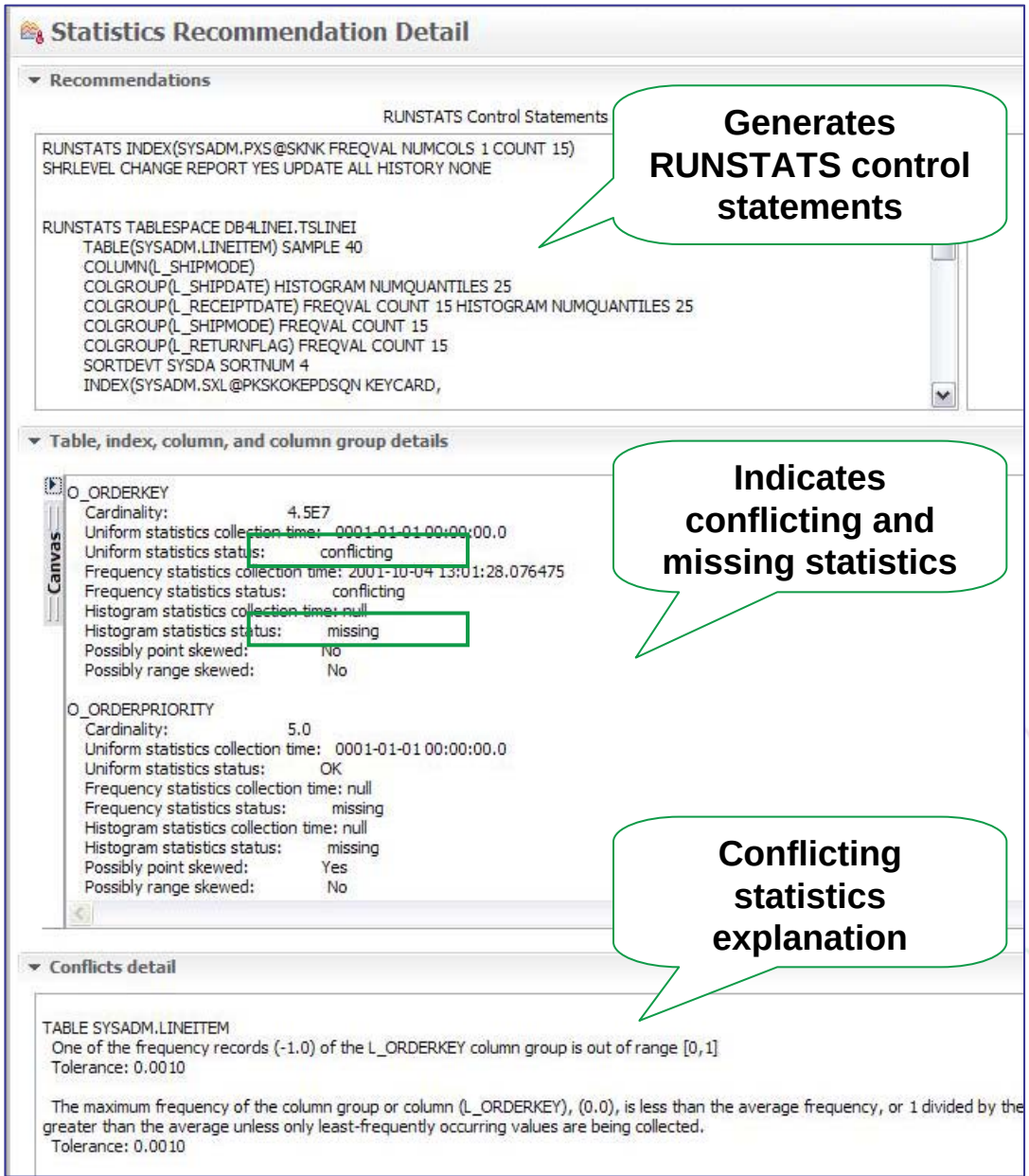


- Assess access path stability to reduce risk of performance regression
  - Is the optimizer able to apply the filtering early?
  - Are there indexes that support an efficient path?
  - Do statistics allow distinction between the choices?

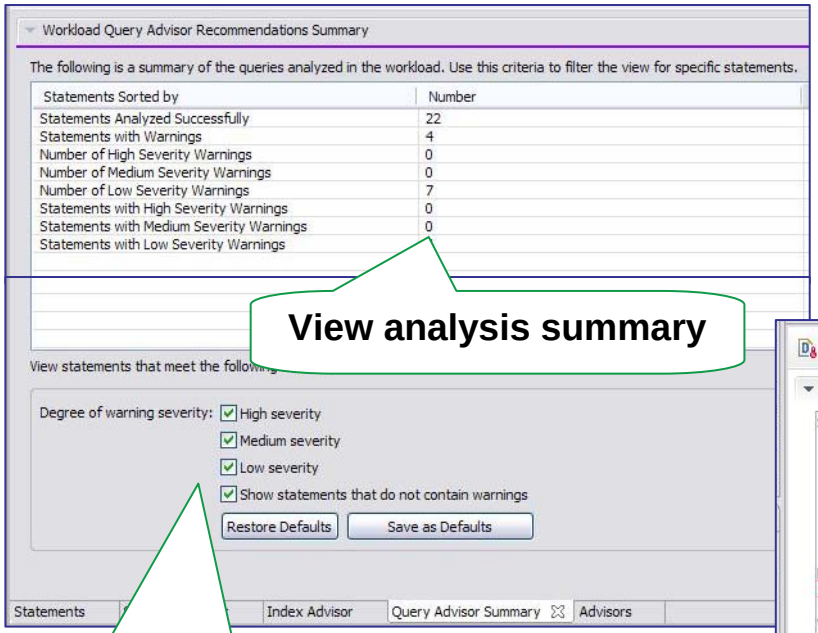
Improve Statistics Quality and Collection

- Provides advice on
  - Missing statistics
  - Conflicting statistics
  - Out-of-date statistics
- Results
  - Accurate estimated costs
  - Better query performance
  - Less CPU consumption
  - Improved maintenance window throughput

"Half of access path PMRs could be resolved by statistics advisor before calling IBM support."
– IBM Support

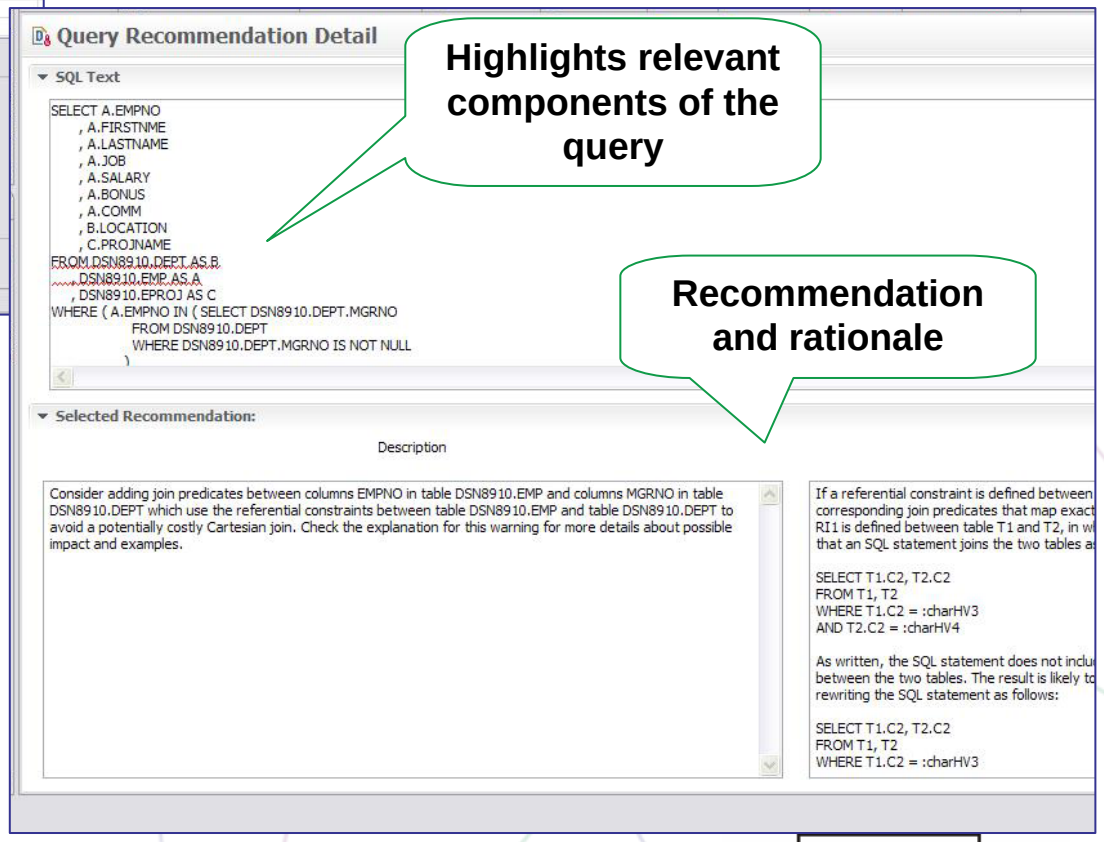


Improve Query Design



Guard against errors and oversights:

- Further constrain query, increase index utilization, and reduce data reads

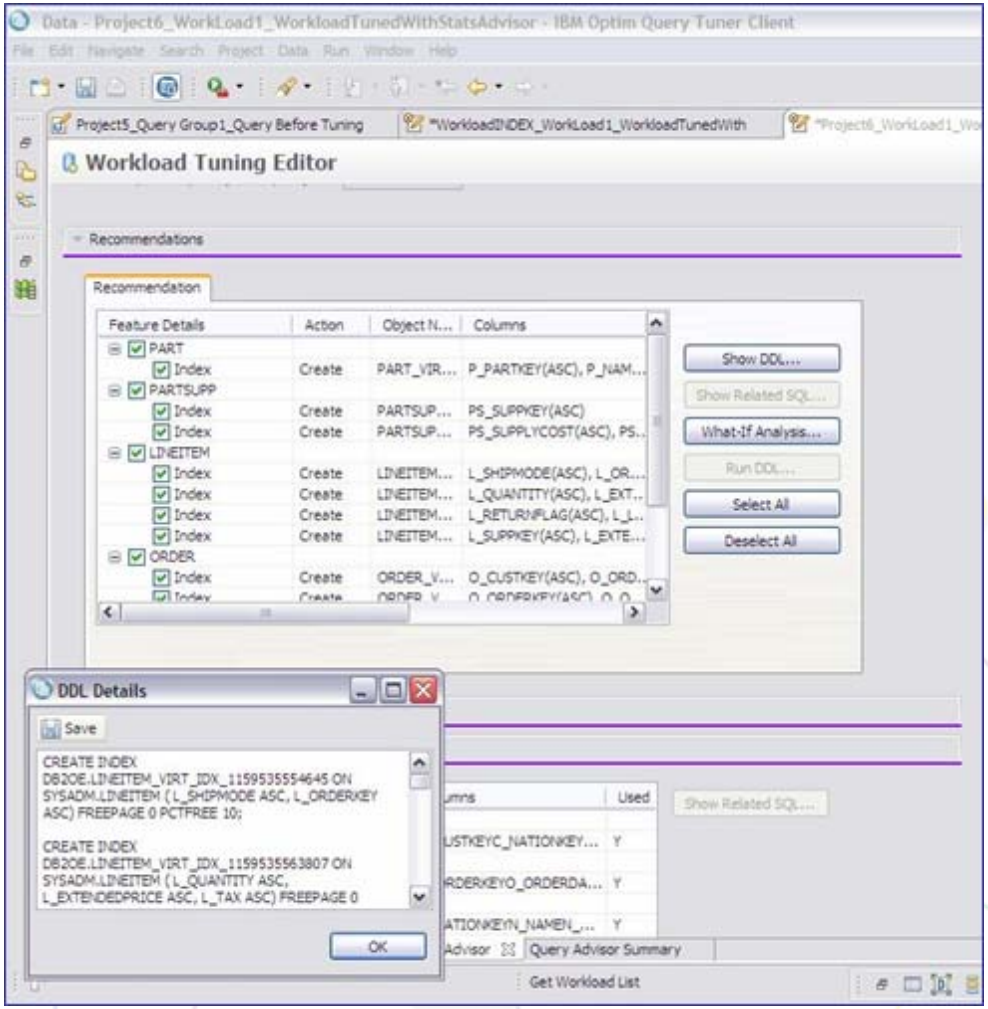


Query Advisor checks for

- Missing join predicate for referential constraint
- Predicates that can be rewritten as indexable
- Stage 2 predicates that can be rewritten as stage 1 predicates

Indexing Advice to Improve Query Efficiency

- Improve query efficiency
  - Indexing foreign keys in queries that do not have indexes defined
  - Identifying index filtering and screening
  - Support for index only access
  - Indexing to avoid sorts
- Simplify use
  - Consolidate indexes and provide a single recommendation
  - Enables what-if analysis
  - Provides DDL to create indexes
  - Run immediately or save



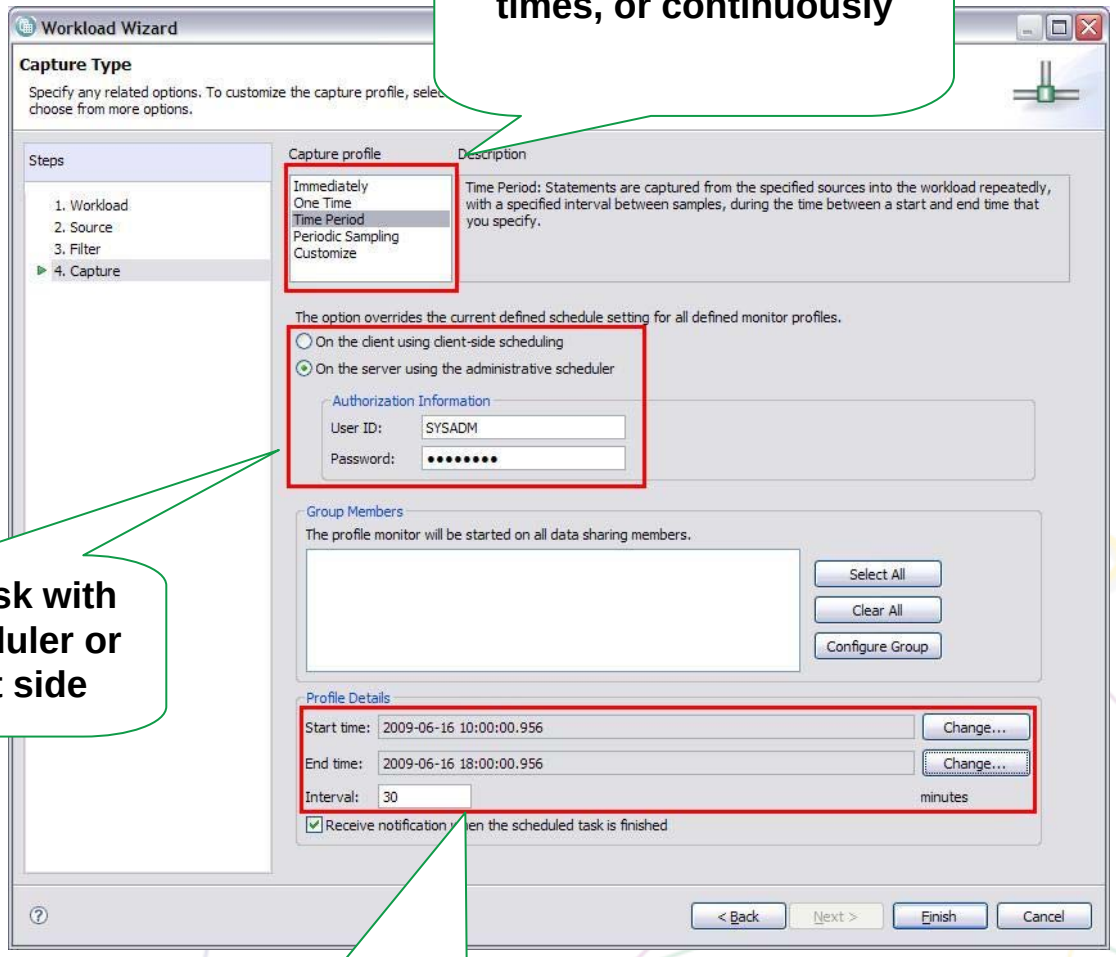
Workload Tuning vs. Query Tuning

- **Optimization decisions are based on tradeoffs**
  - Statistics – CPU costs vs. query savings
  - Indexing – query speed vs resource and transaction
- **Workload tuning speeds up analysis**
  - Analyzes multiple queries at once
- **Workload tuning consolidates and optimizes recommendation for overall workload**
  - Statistics recommendations
  - Index recommendations

Speed up analysis, optimize design, and balance resource usage

Scheduling

- Easy task scheduling
  - Capture of workloads
  - Profile monitor activity
  - Explain of workloads
  - Advisors for workloads



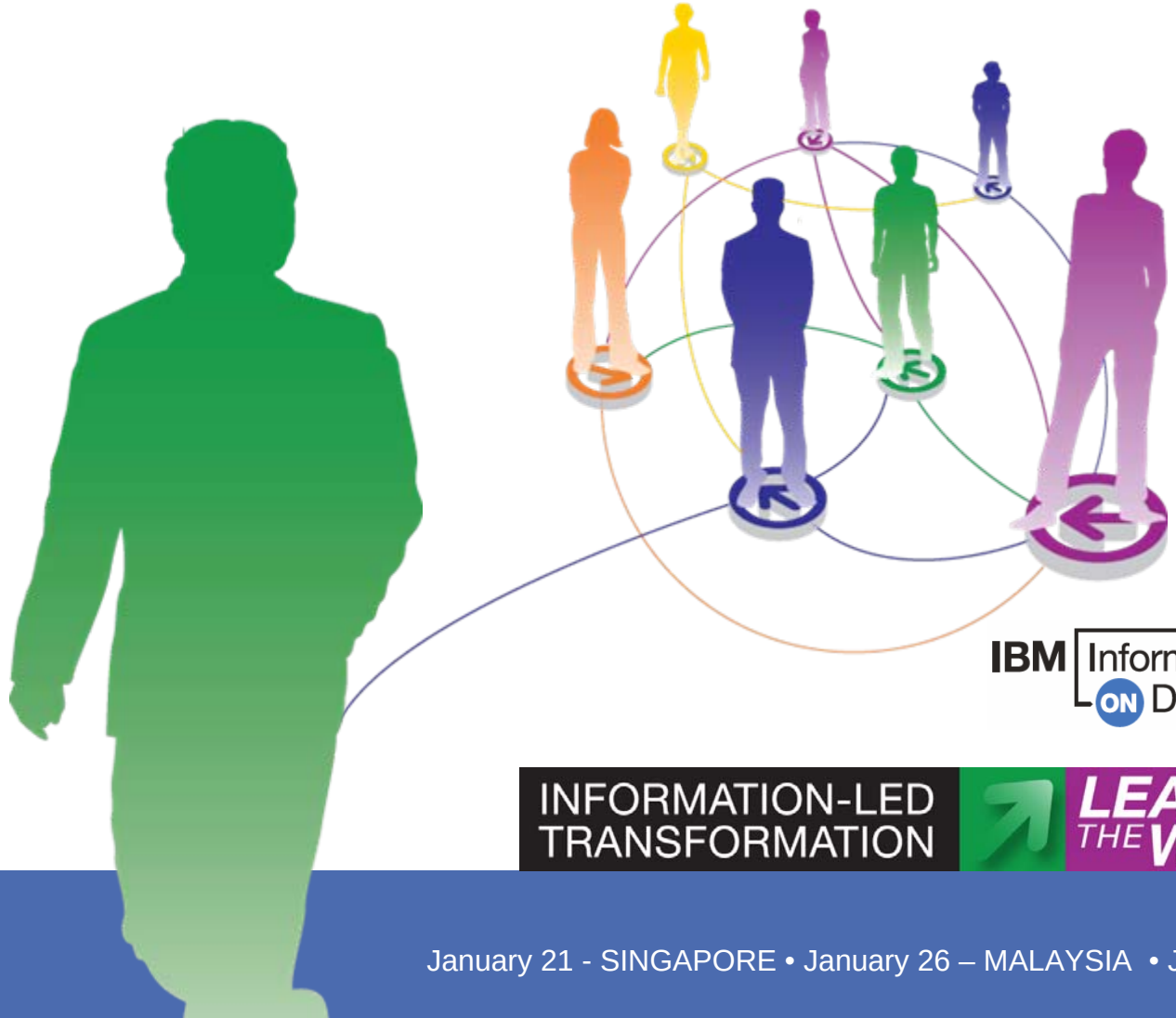
Schedule the task to be executed once, multiple times, or continuously

Schedule the task with the admin scheduler or from the client side

Specify the start time, end time, and interval

IBM Optim Solutions

What Customer are Saying



IBM Information On Demand 2010

INFORMATION-LED TRANSFORMATION LEAD THE WAY

"Optim has already yielded the returns we were looking for, and we have just touched the tip of the iceberg…"

Chiquita!

"We can easily comply with data privacy regulations without incurring additional costs…"

DISCOVER FINANCIAL SERVICES

"The bottom line is that Optim's archiving capabilities will help us exceed our SLA…"

RCI Banque groupe Renault

MidAmerican ENERGY
OBSESSIVELY, RELENTLESSLY AT YOUR SERVICE.

…And Many Other Happy Optim Customers

CORNING
TOSHIBA
BOEING
HONDA
The Power of Dreams
UCSF
DIRECTV
Rabobank
Advance Auto Parts
Countrywide
HOME LOANS
Safeco Insurance
NAUTILUS Inc
newgistics
CHIVAS REGAL 12
TD Bank Financial Group
BlueCross BlueShield Association
AEP
AMERICAN ELECTRIC POWER
BOSTONIA CONDITA A.D. 1630
City of Boston
ART IN MOTION
ROGERS
Your World Right Now
Limitedbrands
EOS
Puerto Rico Federal Affairs Administration
EAGLE Family Foods, Inc.
Bausch & Lomb

…And More


Gap Inc.


PRIMERICA


Wendy's
OLD FASHIONED
HAMBURGERS.


LendingTree
WHEN BANKS COMPETE, YOU WIN®


Metavante


BT


MetLife


Telstra


BMW


Sun
Life Financial


CANAL+


Ford


DIRECTV


ANZ


CapitalOne


Caixa Geral
de Depositos


verizon


MERCK


GEICO


LOWE'S


WELLS
FARGO


BOEING


DAIMLERCHRYSLER


TARGET


TOSHIBA


T · · · Systems ·


ebay


Rabobank

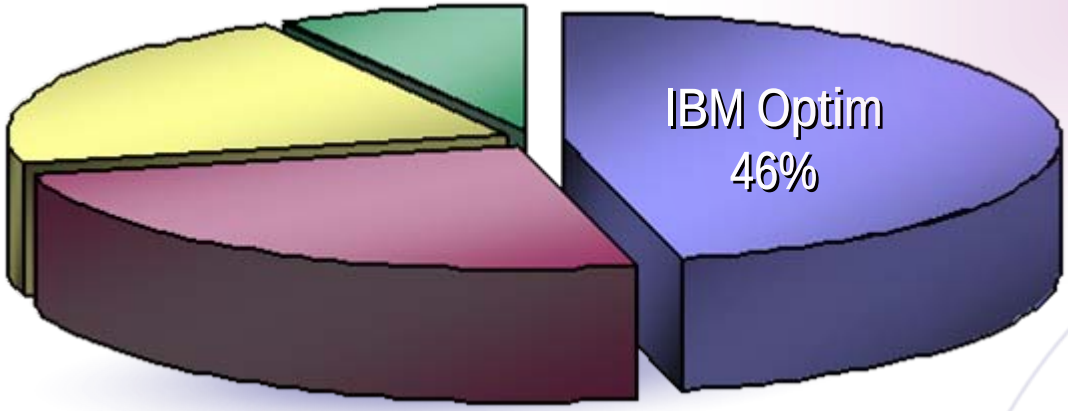

Johnson & Johnson


HSBC
The world's local bank

Customers Have Made Optim The Market Leader

"[Optim] customers spoke highly of the reliability and high performance of its archiving solution, which sets it apart from other vendors in this space."

-- Forrester Research

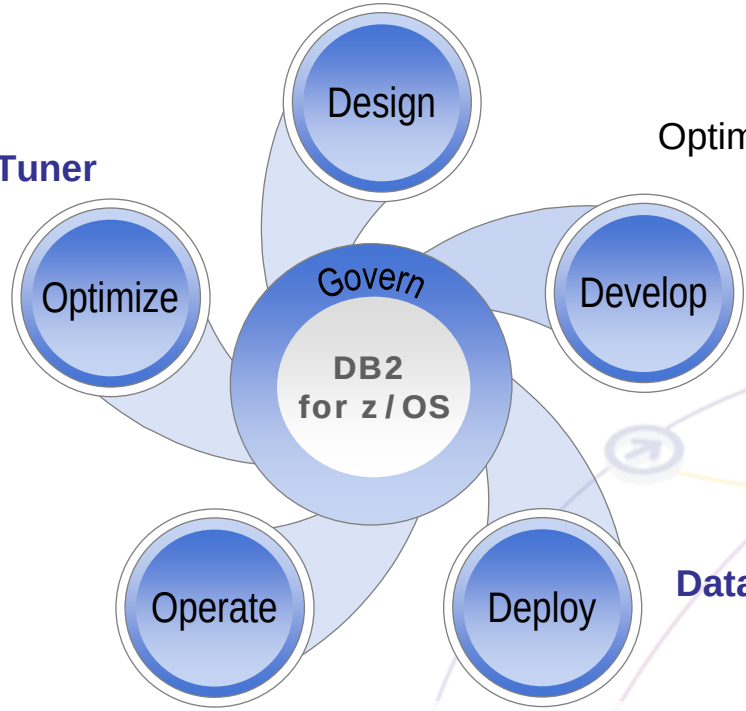


Gartner Research

Sources:
Gartner, "Archiving Software Market to Experience Strong Growth Through 2010" 2006;
Forrester Research, "Database Archiving Remains An Important Part Of Enterprise DBMS Strategy," 2007

The broadest range of capabilities for managing the value of your data throughout its lifetime

InfoSphere Data Architect



Optim Development Studio

Optim Query Workload Tuner

Optim Data Growth Solution

DB2 Performance Management Solution

Optim Test Data Management

Optim Data Privacy Solution

DB2 Automation Tool

DB2 Recovery Solution

DB2 Utilities Suite

DB2 Change Management Solution

Data Studio pureQuery Runtime

DB2 Audit Management Expert

Data Encryption for DB2 and IMS