Information Management

IBM

Data Management

Solving Complex Data Warehousing Problems with Advanced Tooling of IBM InfoSphere Warehouse Advanced Editions

Jon Lind
IBM Product Manager, InfoSphere Warehouse software
February 23, 2012

Formerly

© 2012 IBM Corporation

DB2 Tech Talk series host: Serge Rielau

SQL Architect,
DB2 for Linux, UNIX, and Windows at IBM

http://www.sqltips4db2.com/

Today's presenter: Jon Lind

Product Manager,
InfoSphere Warehouse at IBM

http://www.ibm.com/software/data/infosphere/data-warehousing/
http://www.facebook.com/pages/IBM-InfoSphere-Warehouse

Special thank you: Rav Ahuja

Senior Program Manager, IBM Cloud and Big Data

© 2012 IBM Corporation

IBM Data Warehouse & Analytics Solutions

Simplicity, Flexibility, Choice

IBM Netezza

True Appliance

IBM Smart Analytics System

Flexible Integrated System

IBM InfoSphere Warehouse

Custom Solution

Warehouse Accelerators

Information Management Portfolio
(Information Server, MDM, Streams, etc)

Simplicity — **The right mix of simplicity and flexibility** — **Flexibility**

© 2012 IBM Corporation

IBM InfoSphere Warehouse

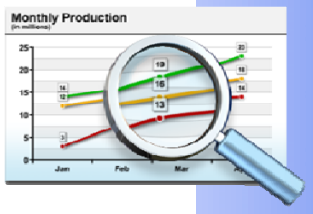
Solve problems – Create new opportunities

FLEXIBILITY

To Adapt to Your Business Needs

PERFORMANCE

For Business Agility



SOPHISTICATED ANALYTICS

For Smarter Decision-Making

© 2012 IBM Corporation

ISW Departmental Editions

Departmental Edition

- **Core server**
  - DB2 9.7 Enterprise Server Edition
  - Data Partitioning Feature
  - Workload Manager Feature
  - Storage Optimization Feature
- **ELT and Administration**
  - Design Studio
  - Replication Server
  - Data Studio
  - Performance Manager
- **Analytics**
  - Cubing
  - Intelligent Miner
  - Text Analytics
  - Cognos 10 (5 licenses)

Advanced Departmental Edition Adds

- ***Warehouse packs***
  - ***Customer Insight***
  - ***Market and Campaign Insight***
  - ***Supply Chain Insight***

© 2012 IBM Corporation

ISW Enterprise Editions

Enterprise Edition

- **Core server**
  - DB2 9.7 Enterprise Server Edition
  - Data Partitioning Feature
  - Workload Manager Feature
  - Storage Optimization Feature
  - **Label Based Access Control**
- **Design, ELT, and Administration**
  - Design Studio
  - Replication Server
  - **Federation Server**
  - Data Studio
  - Performance Manager **Extended Edition**
  - **Query Patroller**
- **Analytics**
  - Cubing
  - Intelligent Miner
  - Text Analytics
  - **Mashup Center**
  - Cognos 10 (5 licenses)

Advanced Enterprise Edition Adds

- ***Warehouse packs***
  - ***Customer Insight***
  - ***Market and Campaign Insight***
  - ***Supply Chain Insight***
- ***Design, ELT, and Administration***
  - ***InfoSphere Data Architect (10 licenses)***
  - ***High Performance Unload***
  - ***Merge Backup***
  - ***Recovery Expert***
  - ***Query Workload Tuner***

© 2012 IBM Corporation

Terabyte Pricing Makes Good Business Sense

- **Pay for what you analyze**
  - Data is the currency of the Information Age
  - The more data people can analyze the more value they receive
  - Data warehouses are a gold mine for business decision makers
- **Flexibility**
  - “Pay as you Grow”
  - Deploy on the system that best meets your business needs
- **Enterprise license planning made easy**
  - Companies know how much data they have and the growth rates
  - Companies can easily plan their warehouse license requirements
- **Lower cost of Growth**
  - Customers can provision hardware platforms for end game capacity
  - Simple lower cost of growth by simply Increasing license entitlement

© 2012 IBM Corporation

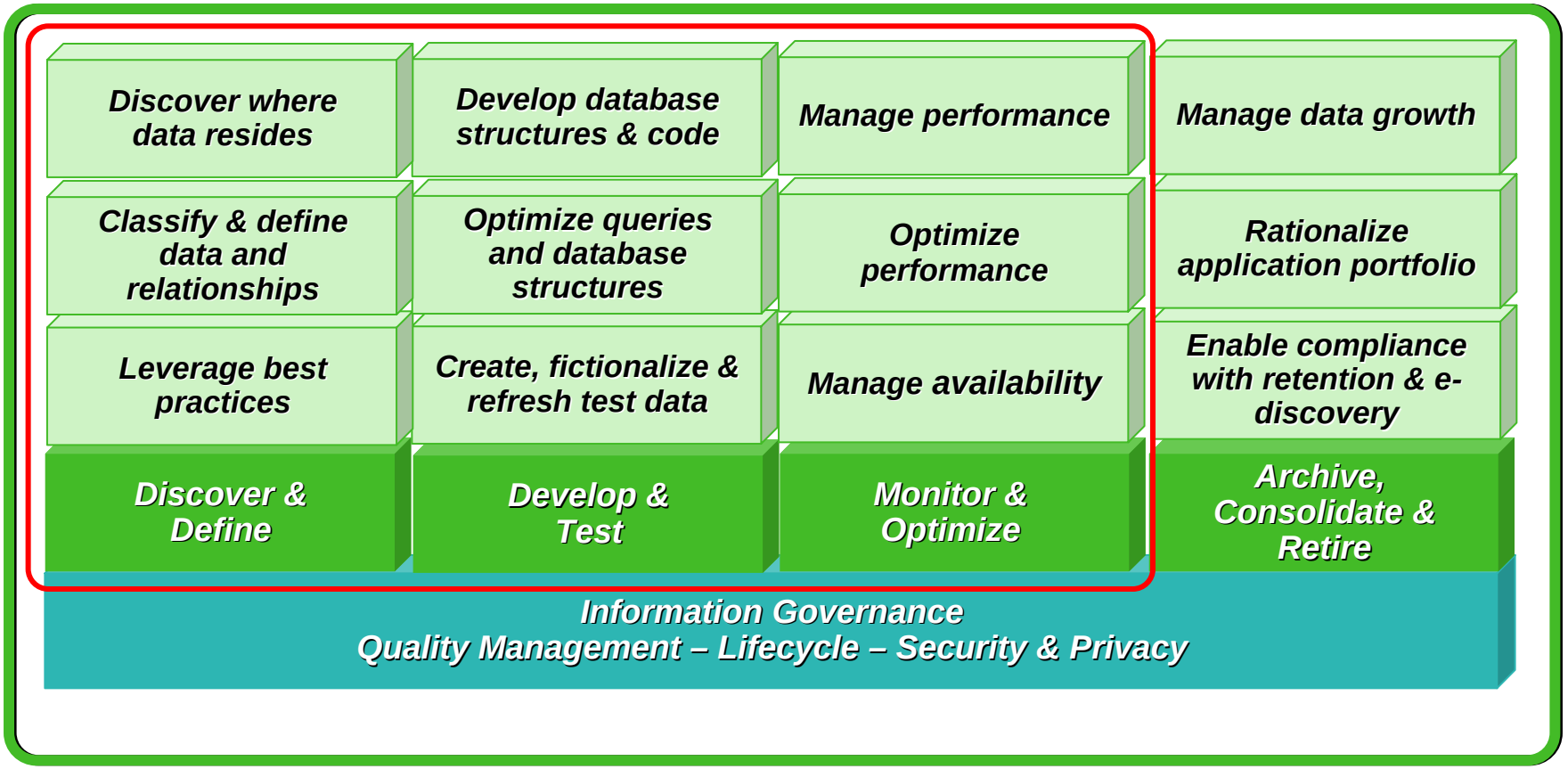
What's New in Advanced Editions?

- **Accelerate value delivery**
  - Warehouse packs with industry models and Cognos reports
    - Markets and Campaign Insight
    - Customer Insight
    - Supply Chain Insight
- **Improve governance**
  - InfoSphere Data Architect
- **Improve performance**
  - InfoSphere Optim Query Workload Tuner
- **Improve availability**
  - InfoSphere Optim High Performance Unload
  - DB2 Merge Backup
  - DB2 Recovery Expert

© 2012 IBM Corporation

InfoSphere Optim Offerings

Streamlining the Management of Data Across its Lifecycle



© 2012 IBM Corporation

InfoSphere Warehouse Solutions to Discover and Define

Standardize and collaborate

- **Data Architect**
  - Design and manage enterprise models
  - Enforce model conformance to enterprise standards
  - Leverage industry data models for best practices

Accelerate projects

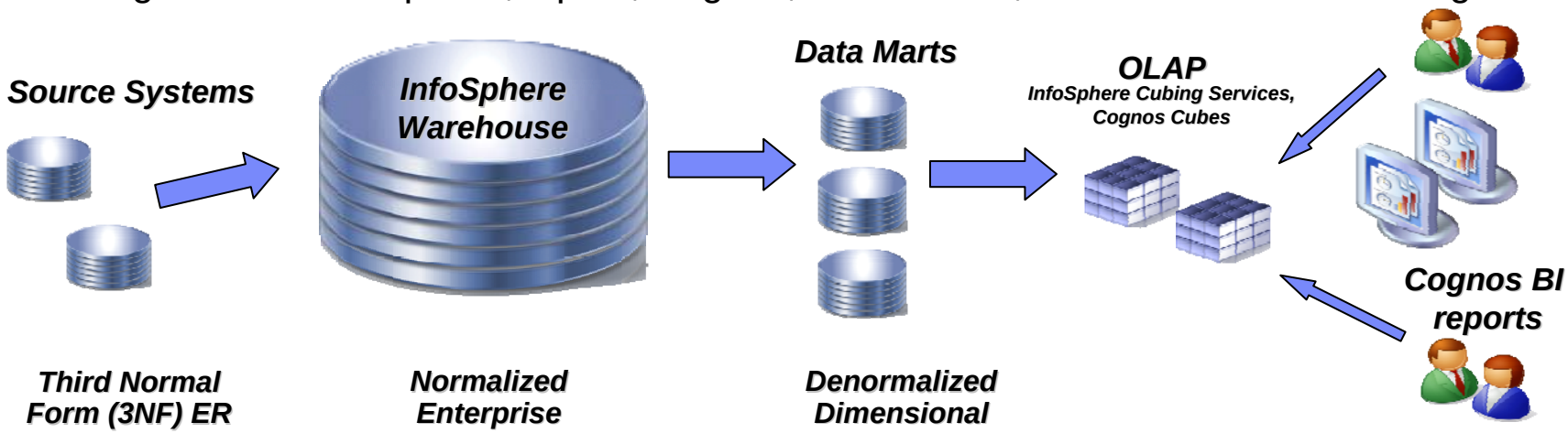
- **Warehouse Packs**
  - Data models
  - Business reports

© 2012 IBM Corporation

Managing Your Information Models

InfoSphere Data Architect

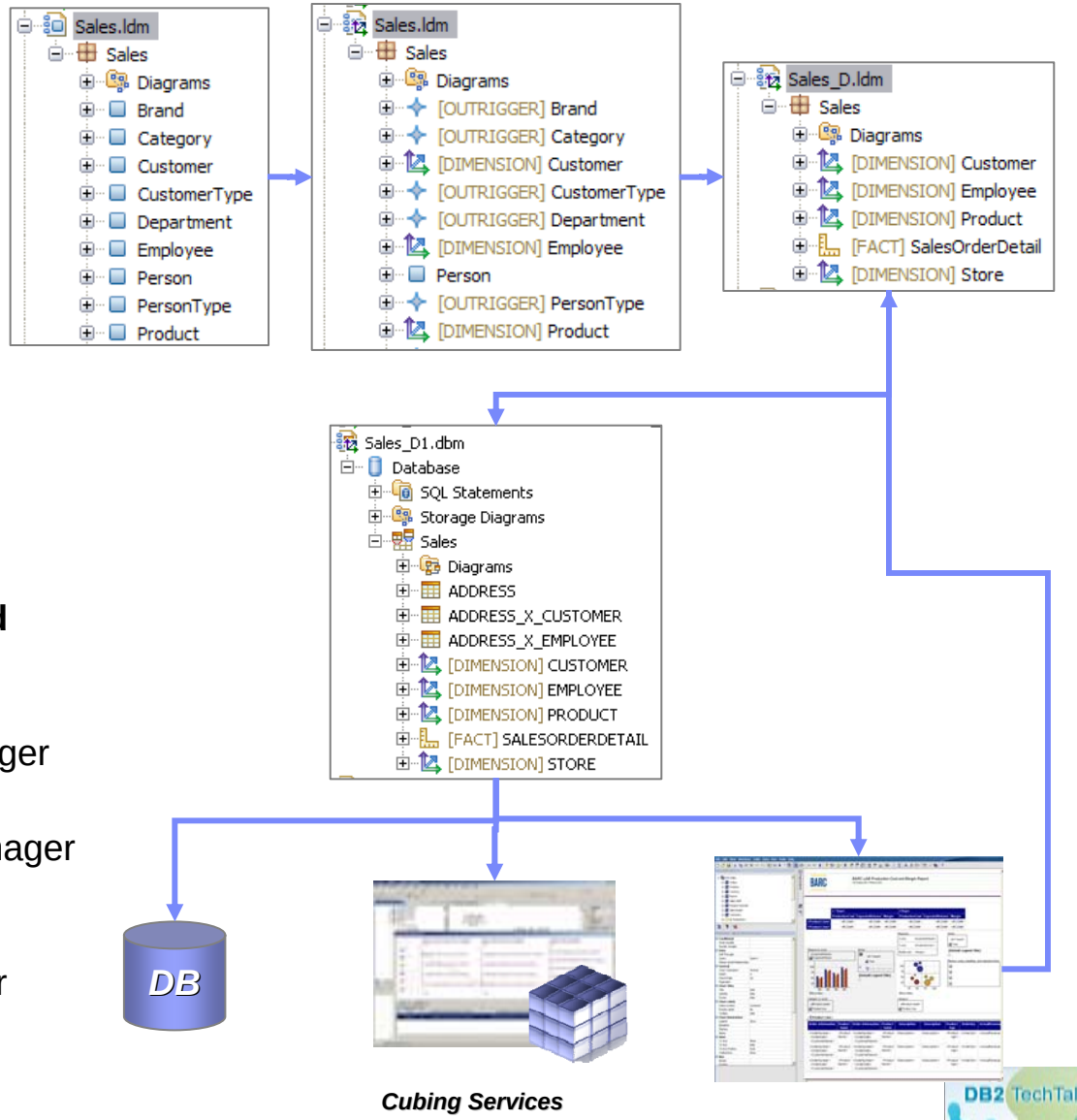
- **Full service data modelling**
  - Create logical, physical, and dimensional models
- **Validate models against enterprise standards**
  - Standardize names, terms, data types, transformations, etc
- **Leverage existing assets**
  - Forward engineer, reverse engineer, and import from source or other
- **Collaborate across teams**
  - Integrate with InfoSphere, Optim, Cognos, Data Studio, and Rational and offerings



© 2012 IBM Corporation

Multi-Dimensional Modeling

- **Automatic Discovery**
  - Facts and measures
  - Dimensions and outriggers
- **Dimensional Notation**
- **Automatic De-normalization**
  - Star schema, snowflake, starflake
- **Lossless integration with Cognos and Cubing Services**
  - Transform dimensional-physical data model to a Cognos Framework Manager model or Cubing Services model
  - Transform a Cognos Framework Manager dimensional model to a dimensional-logical data model
  - Support Cognos Framework Manager model data source “re-targeting”



© 2012 IBM Corporation

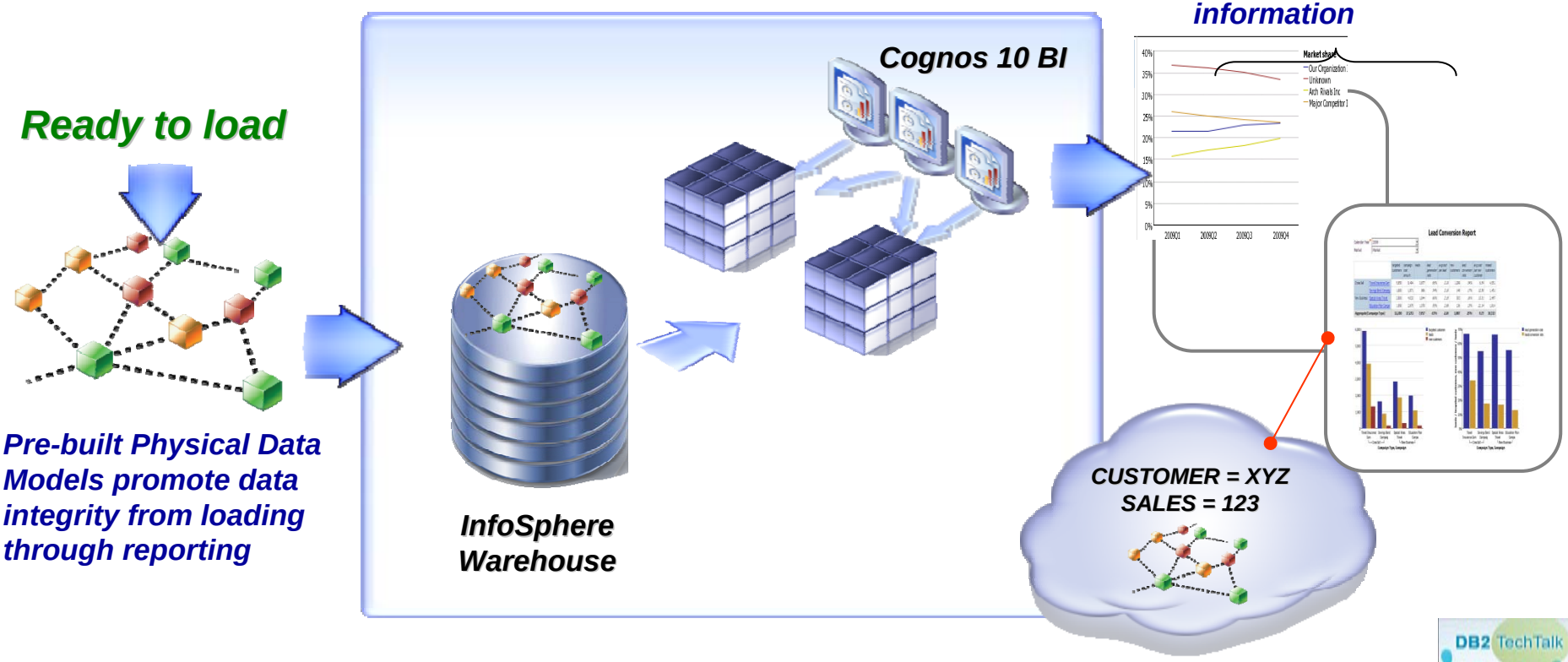
Business Models – InfoSphere Warehouse Model Packs

Solution acceleration brings immediate value – the work is done!

Leverage 'pre-built' data models targeted at specific business issues

- ***Customer Insight***
- ***Market and Campaign Insight***
- ***Supply Chain Insight***

Pre-built reports illustrate IMMEDIATE representation of your information

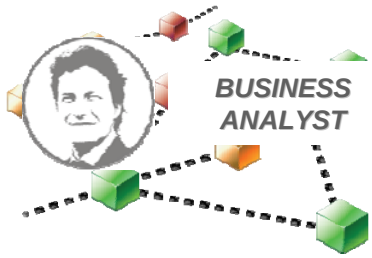


Pre-built Physical Data Models promote data integrity from loading through reporting

© 2012 IBM Corporation

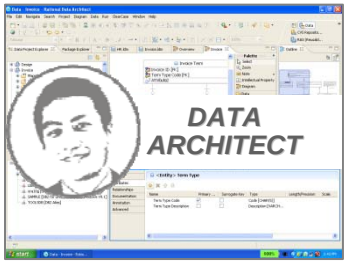
Managing Database Objects to Align Business and IT

Leverage best practices



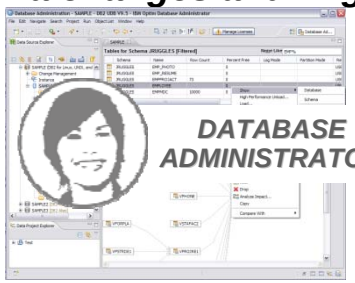
Warehouse pack models

Design, extend, and govern data



InfoSphere Data Architect

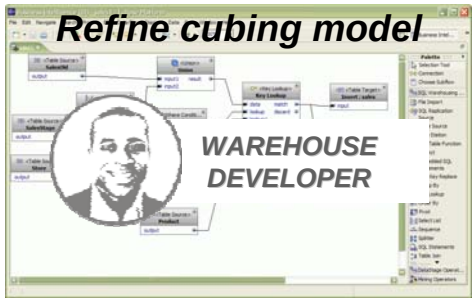
Compare, generate, and deploy schema changes and migrate



Data Studio

Define ETL transformations

Refine cubing model



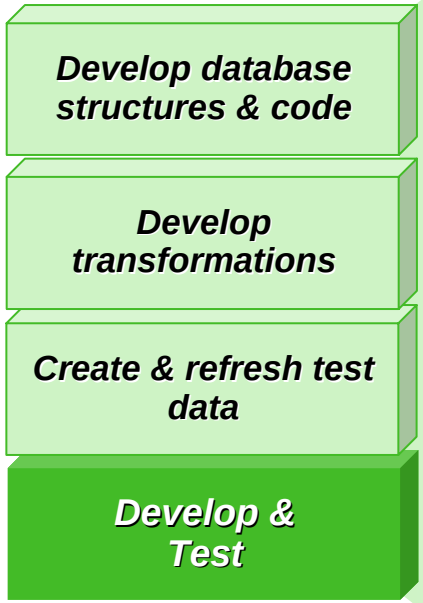
Design Studio

Develop business reports

Cognos

© 2012 IBM Corporation

InfoSphere Warehouse Solutions to Develop and Test



Accelerate development
Save time, reduce errors

- **Data Studio**
  - Manage databases and database changes
  - Develop database routines and access methods
  - Develop SQL
  - Copy and edit data

Optimize query performance

- **Design Studio**
  - Develop in-database transformations
  - Leverage cubing services

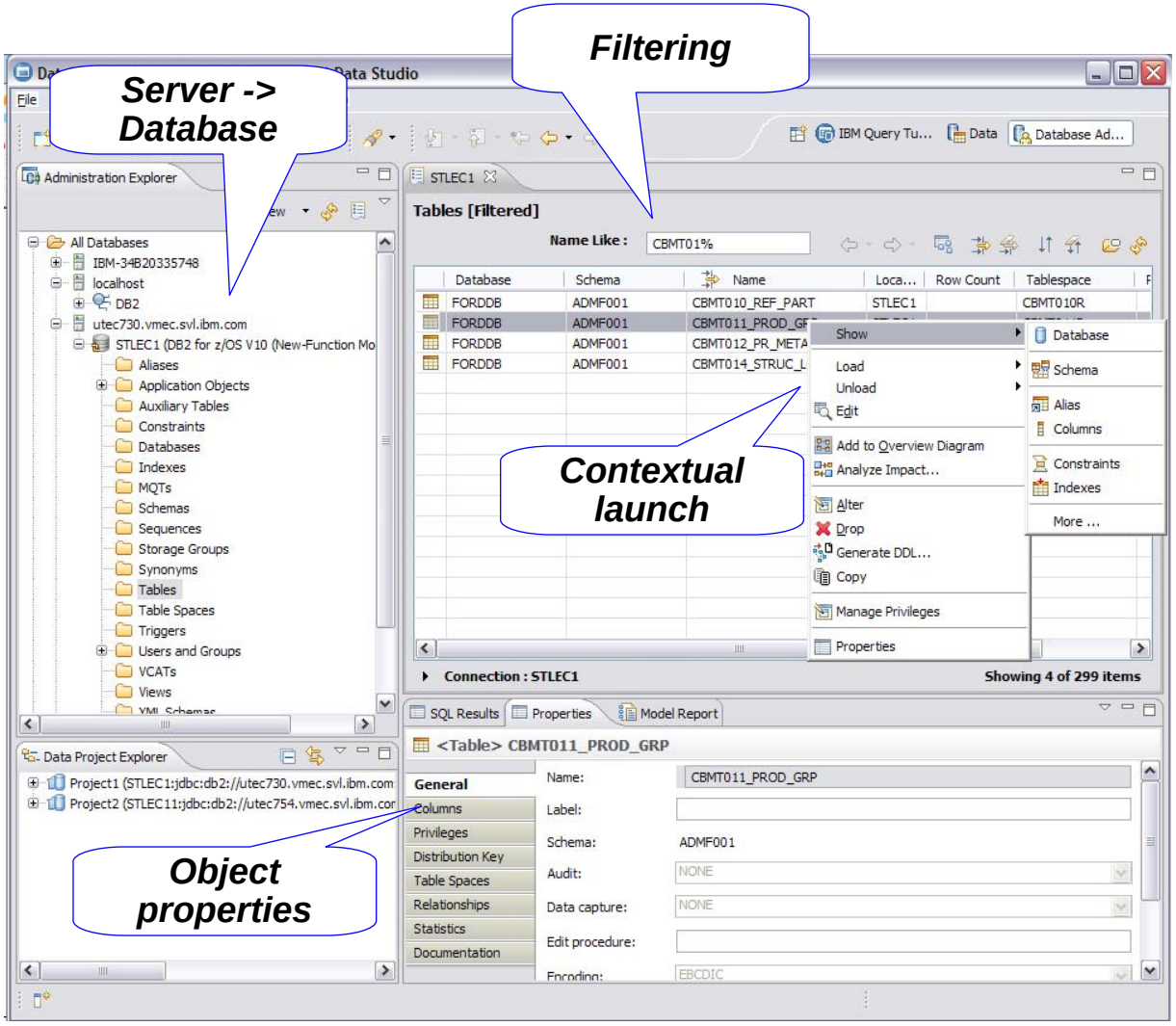
Install Data Architect, Data Studio, and Design Studio into the same Eclipse workspace

© 2012 IBM Corporation

Strategic replacement for Control Center

Data Studio 3.1

- **Object navigation, filtering and management**
- **Complex change management**
- **Configuration and maintenance assistants**
- **Task scheduling**
- **SQL and routine development**
  - Query analysis
  - Oracle and Sybase Skin compatibility
- **Data browsing and editing**
- **Shell sharing with Design Studio**



© 2012 IBM Corporation

Accelerate Development

Query editing

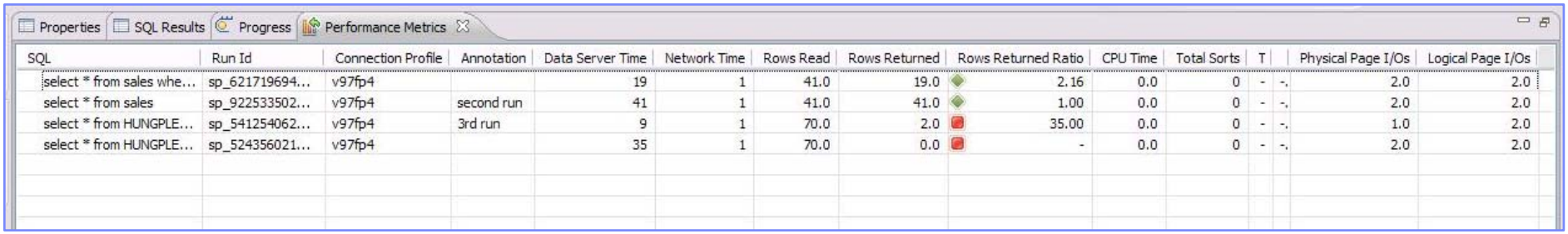
- **Template-enabled**
- **Content assist with connected database**
- **Platform specific syntactic and semantic validation**
- **SQL formatting**
- **Control on commit and error processing and execution environment**
- **Access path visualization**
- **Statistics advice**
- **Performance visibility and validation**
- **Deployment management**

Routine development

- **Template enabled**
- **SQL/PL, PL/SQL, Sybase ANTS, Java**
- **Routine debugging**
- **Deployment management**

Java development

- **Query editor features**
- **Facilitate collaboration with SQL Outline**
- **Identify SQL hot spots**
- **Simplify problem isolation to source code**



SQL	Run Id	Connection Profile	Annotation	Data Server Time	Network Time	Rows Read	Rows Returned	Rows Returned Ratio	CPU Time	Total Sorts	T		Physical Page I/Os	Logical Page I/Os
select * from sales whe...	sp_621719694...	v97fp4		19	1	41.0	19.0	2.16	0.0	0	-	-.	2.0	2.0
select * from sales	sp_922533502...	v97fp4	second run	41	1	41.0	41.0	1.00	0.0	0	-	-.	2.0	2.0
select * from HUNGPLE...	sp_541254062...	v97fp4	3rd run	9	1	70.0	2.0	35.00	0.0	0	-	-.	1.0	2.0
select * from HUNGPLE...	sp_524356021...	v97fp4		35	1	70.0	0.0	-	0.0	0	-	-.	2.0	2.0

© 2012 IBM Corporation

Save Time and Reduce Errors

Complex change management

- **Multiple objects at a time**
- **Automatic dependency handling**
- **Compare objects side by side**
- **Data and privilege migration**
- **Undo scripts**

Maintenance

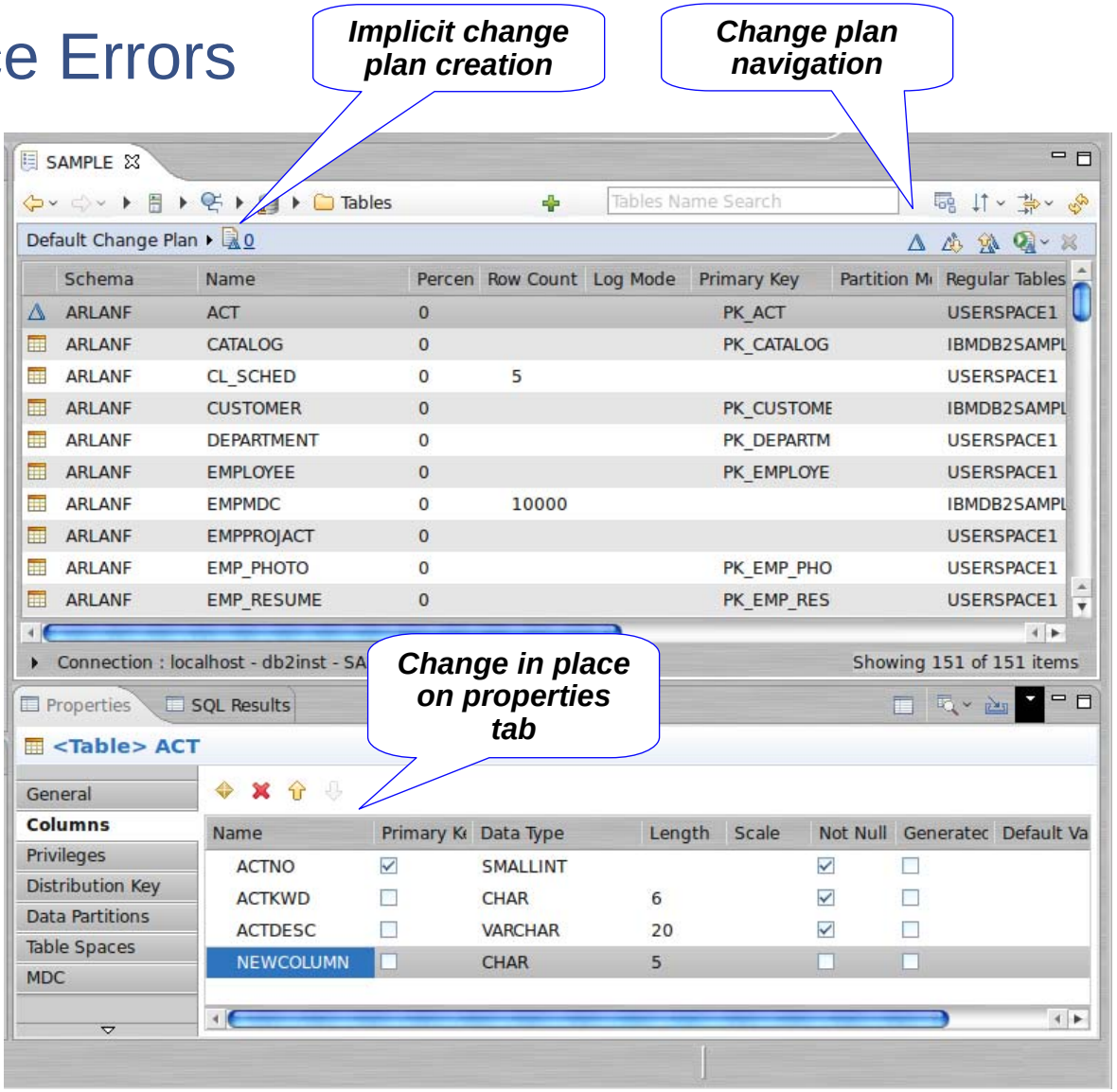
- **Configuration and utilities task assistants**

Data management

- **Browsing, editing, copying, loading, unloading**

Job Management

- **Scheduling jobs, view applications, view history**



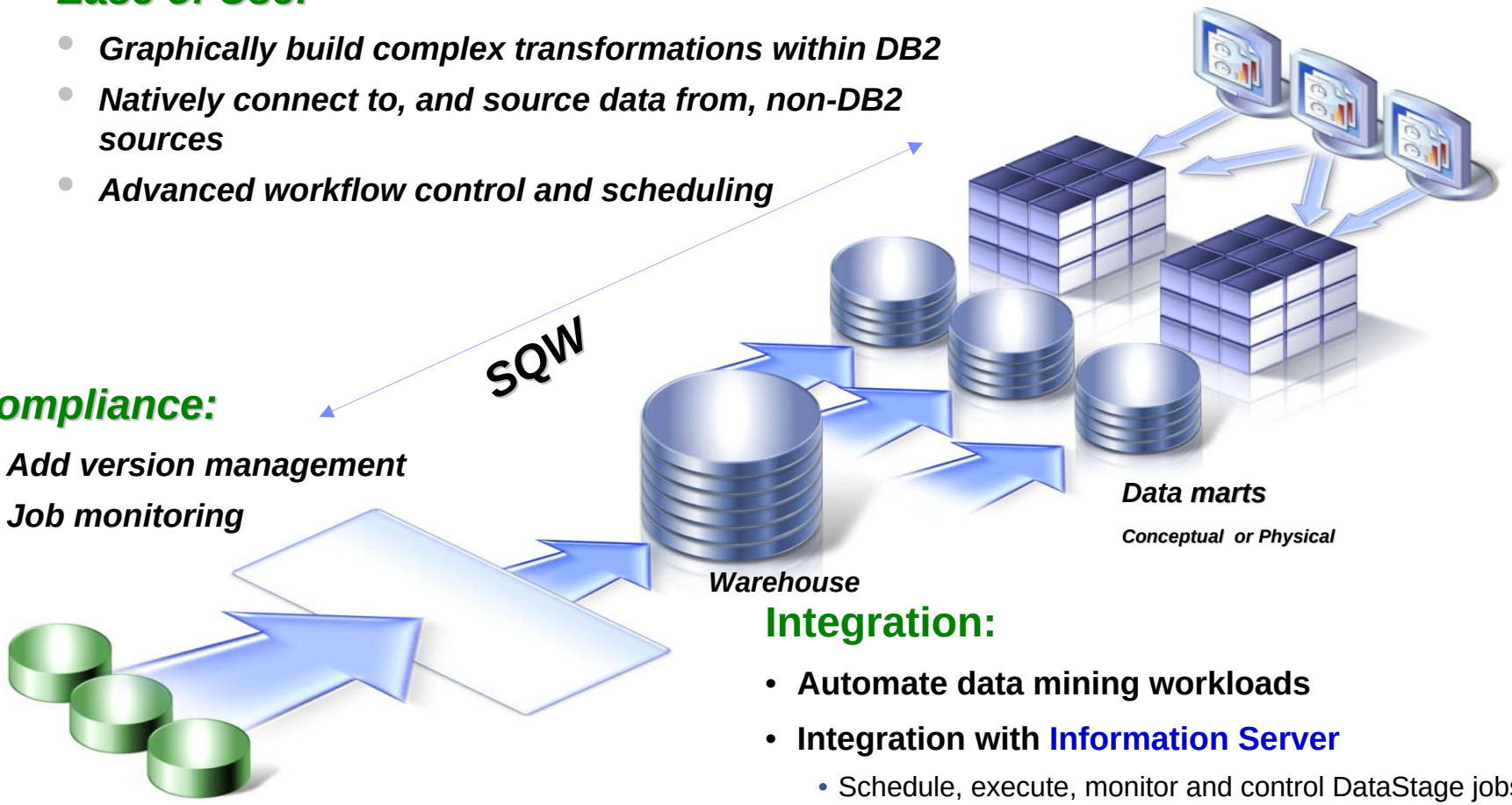
© 2012 IBM Corporation

For data movement & transformation (ELT)

Design Studio includes the SQL Warehousing Tool

Ease of Use:

- ***Graphically build complex transformations within DB2***
- ***Natively connect to, and source data from, non-DB2 sources***
- ***Advanced workflow control and scheduling***



Compliance:

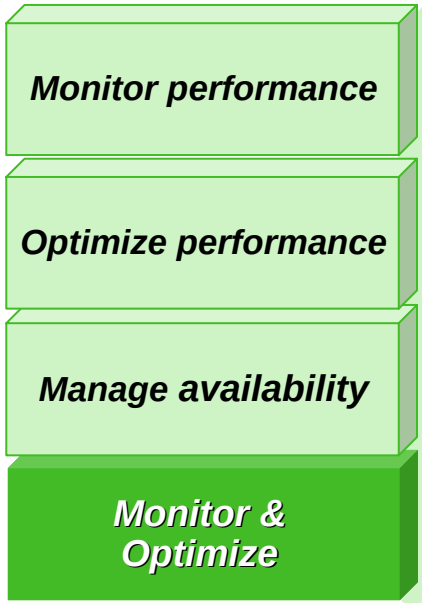
- ***Add version management***
- ***Job monitoring***

Integration:

- **Automate data mining workloads**
- **Integration with Information Server**
  - Schedule, execute, monitor and control DataStage jobs
  - Integrate DataStage work streams

© 2012 IBM Corporation

InfoSphere Optim Solutions to Monitor and Optimize



Meet service level objectives
Improve service
Reduce costs

- **Performance Manager**
  - Monitor database performance and health
  - Align resource use with business priorities
  - Get expert advice on queries, indexes, statistics, and access plans for entire workloads
  - Optimize and stabilize performance of Java applications

Optimize query performance

- **Query Workload Tuner**
  - Get expert advice on queries, indexes, statistics, and access plans
  - Leverage monitoring data to select high cost queries

Optimize recovery time

- **DB2 Advanced Recovery Solutions**
  - Minimize impact of database backups
  - Minimize recovery time

Support high availability scenarios

- **Replication Server**
  - Disaster recovery
  - Active - active

© 2012 IBM Corporation

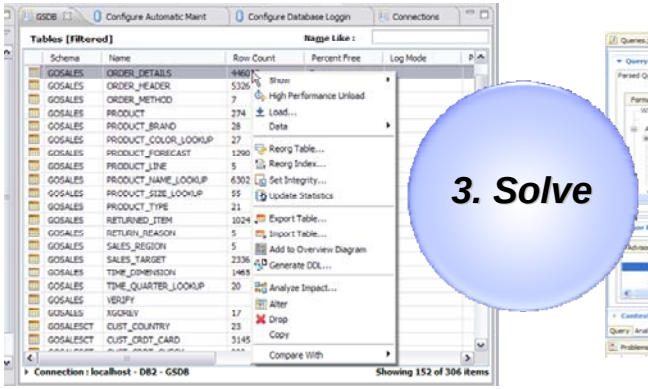
Monitoring and Optimizing Performance

Get early warning of potential problems

Diagnose database problems with resource specific

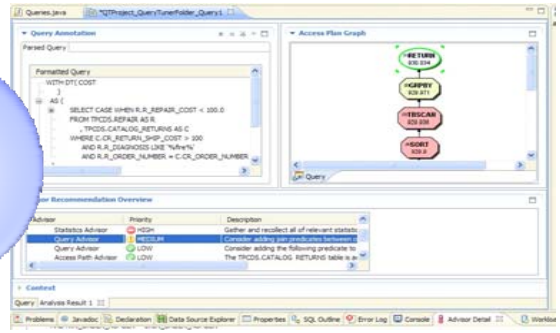
InfoSphere Optim Performance Manager Extended Edition

Resolve database problems



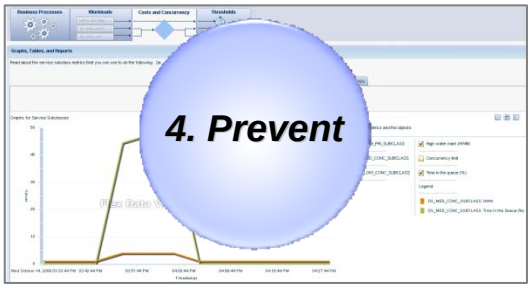
Data Studio

Resolve query problems



InfoSphere Optim Query Tuner

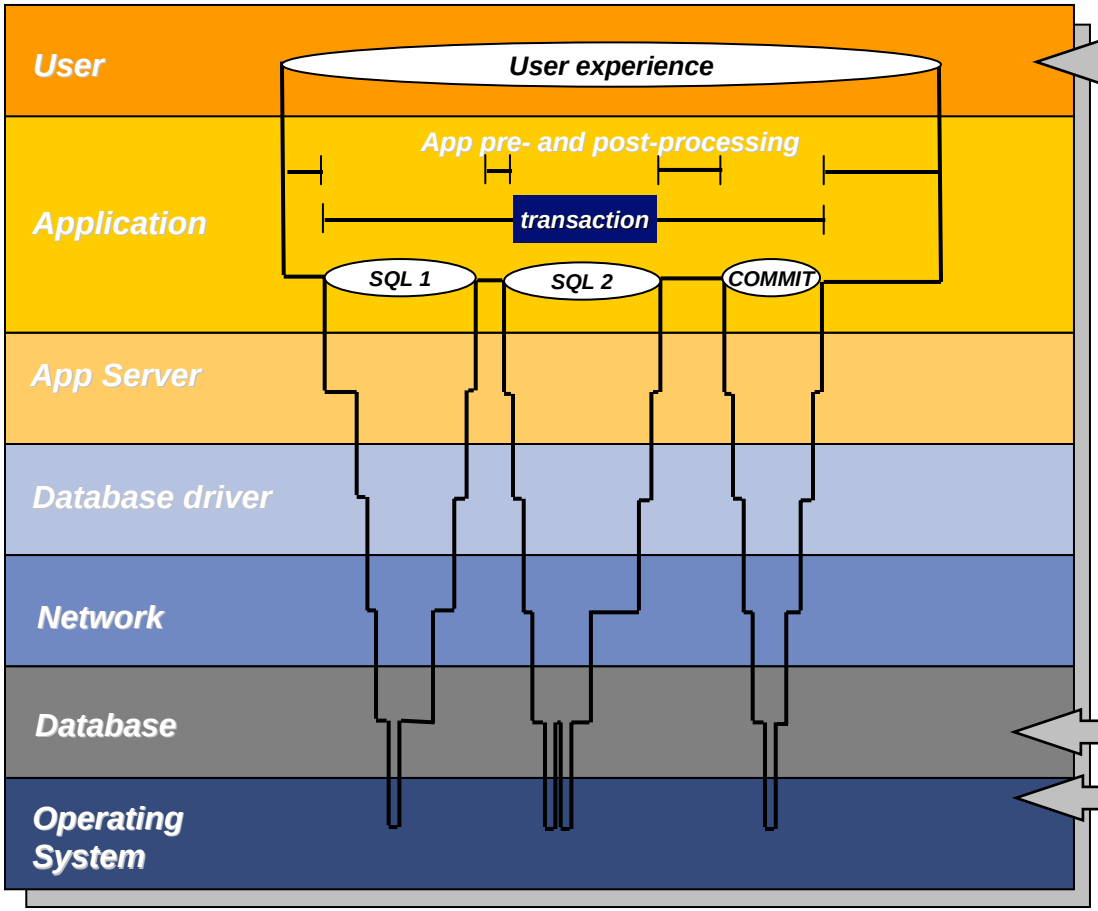
Prevent problems



DB2 Workload Manager

© 2012 IBM Corporation

Database Monitor Limitations



User's experience

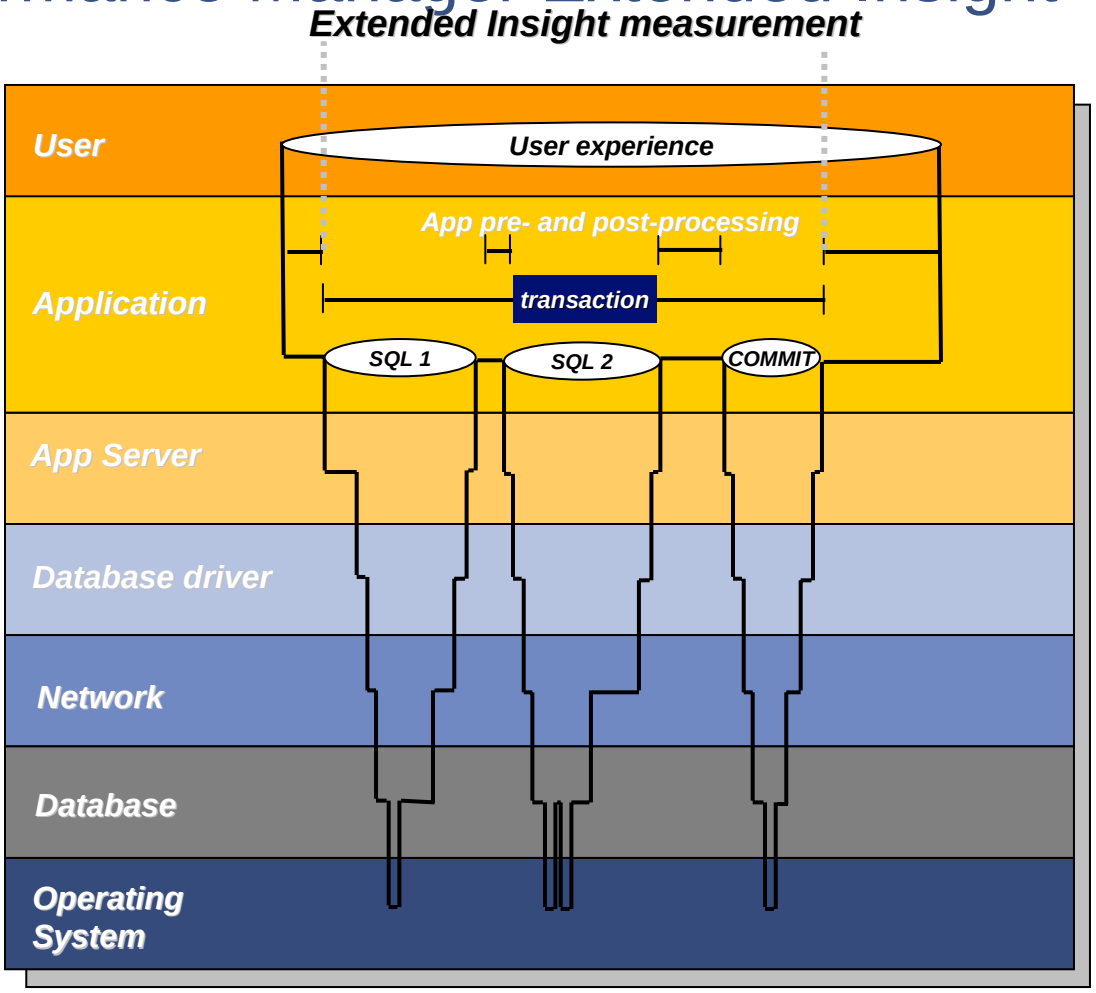
- **Request flows through many layers of the stack**
- **Performance bottlenecks can occur in any layer**
- **The request loses detail about its origin as it moves through the stack**

What database monitors typically monitor

© 2012 IBM Corporation

InfoSphere Optim Performance Manager Extended Insight

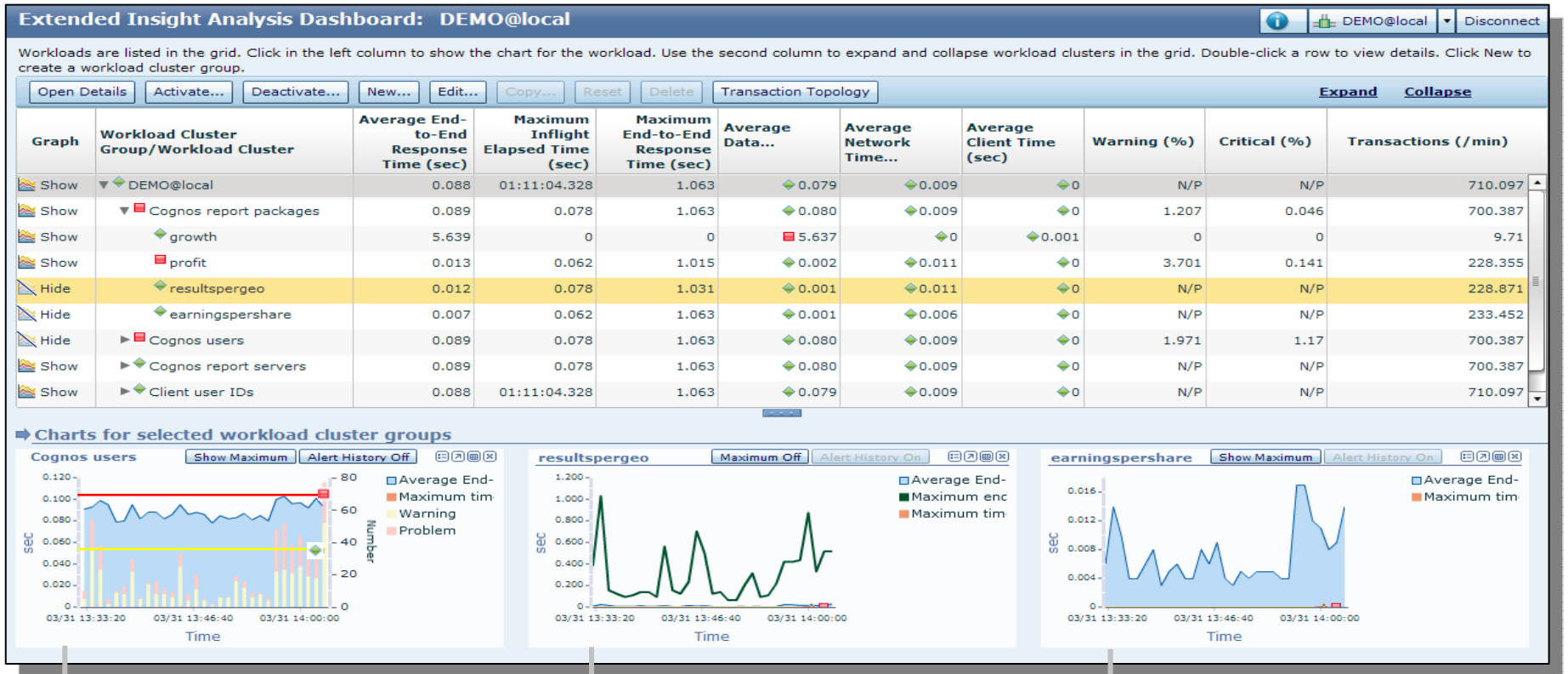
- **Manage response time SLAs**
  - By workload
- **Pinpoint the problem**
  - Problem workload
  - Problem layer
    - App server, driver, network, database, operating system
  - Problem timeframe
  - Problem SQL
- **Down to database component**
  - Breakdown database time by transaction or statement*
    - Compilation, workload manager queue, lock wait, I/O, sort, routine, transaction end,...
    - * with DB2 9.7



© 2012 IBM Corporation

Transaction Response Time Monitoring

By Workload



Response time objective violations

Maximum transaction response time

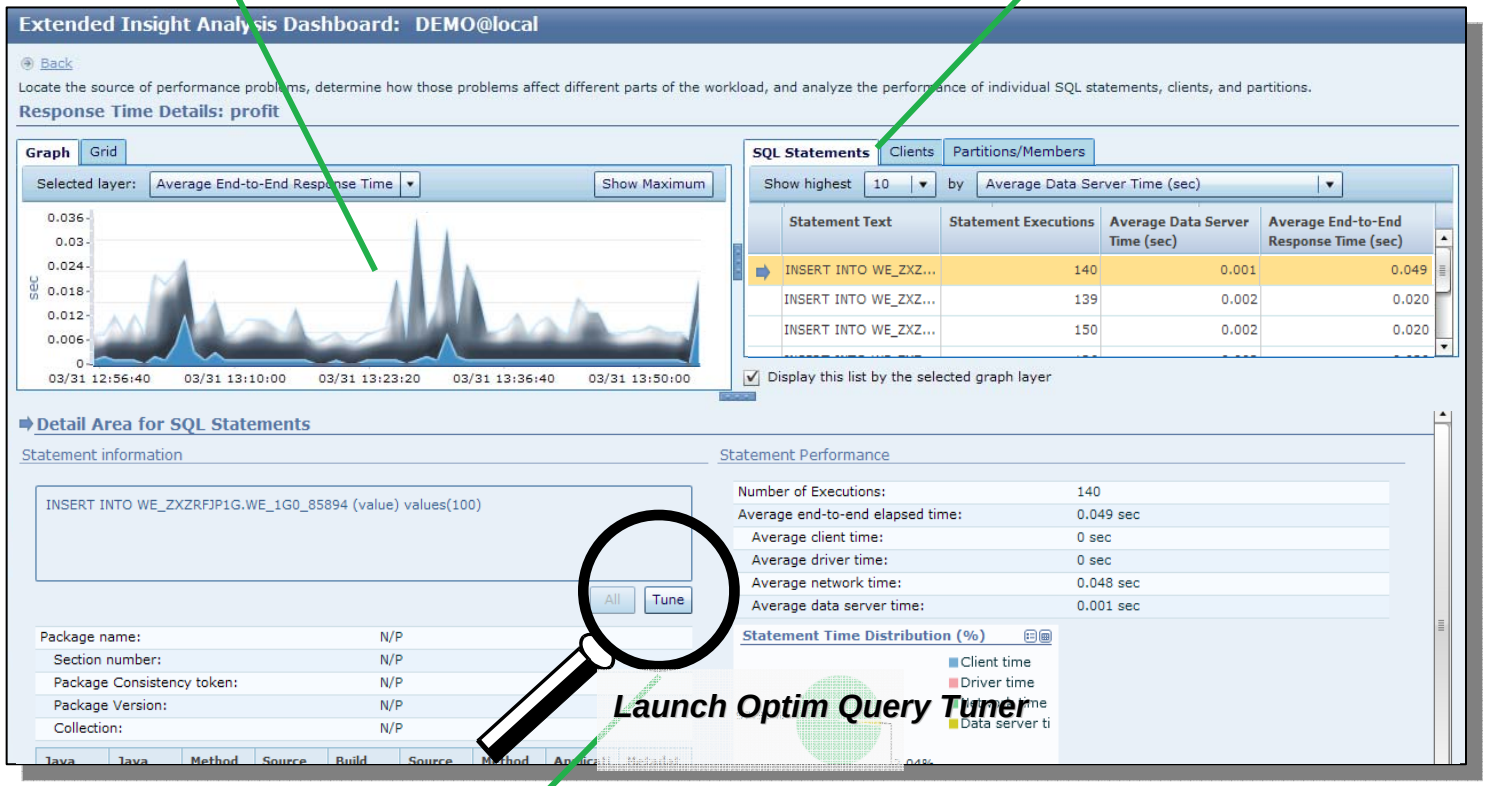
Average transaction response time

© 2012 IBM Corporation

Analyze Workload Response Time

Shows where workload is spending time

Provides detail on queries, clients, and partitions



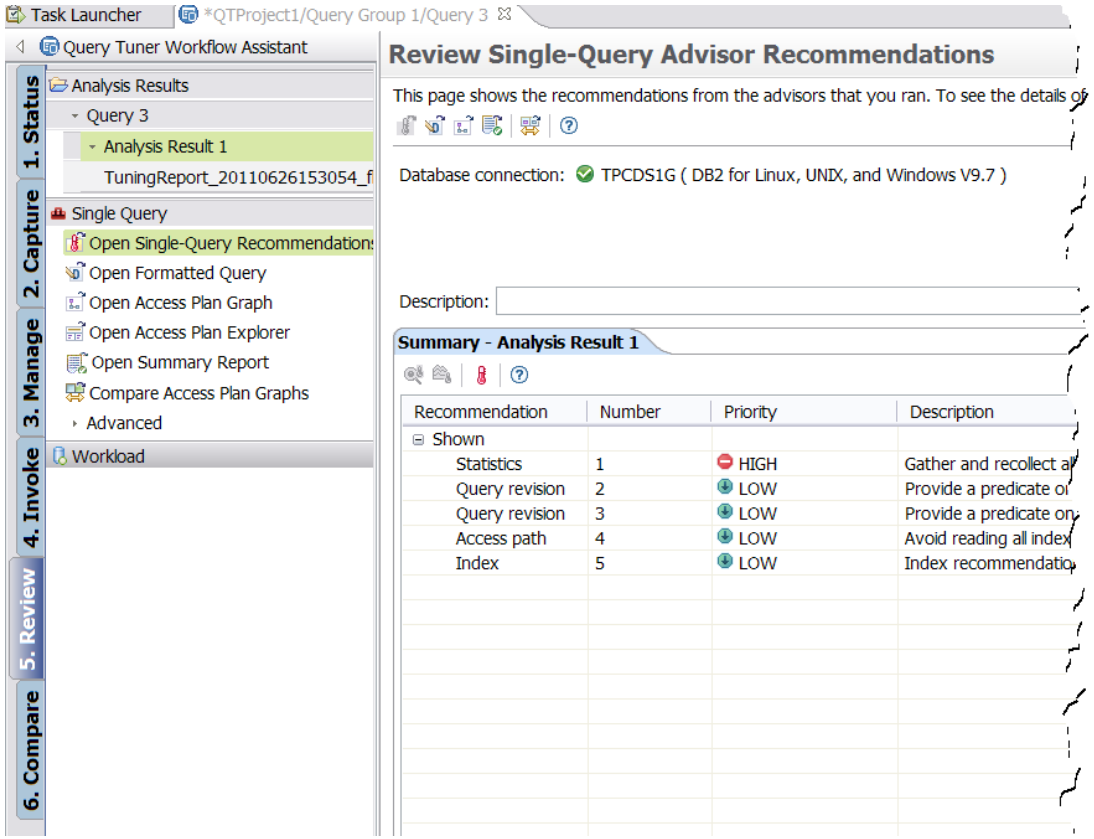
Provides detail on selected area

© 2012 IBM Corporation

InfoSphere Optim Query Workload Tuner

Expert-Enabled Query Tuning Out of the Box

- **Identify query candidates from numerous sources**
  - DB2 catalog
  - Package cache
  - Development Studio hot spots
  - Optim Performance Manager
- **Facilitate analysis**
  - Query formatting
  - Query annotation
  - Access path visualization, annotation, and warnings
- **Get expert tuning advice**
  - Improve query design
  - Improve statistics quality
  - Improve database design



Analyze single queries or query workloads

© 2012 IBM Corporation

Improve Query Design

Guard against errors and oversights

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

Summary - 1 Initial Analysis

Recommendation	Number	Priority	Description
Shown			
Query revision	1	LOW	Provide a predicate on column SR_RETURNED_DATE_SK.
Access path	2	LOW	Avoid table space scan (Operator ID = 6) on table TPCDS.STORE_RETURNS.
Access path	3	LOW	Avoid table space scan (Operator ID = 7) on table TPCDS.DATE_DIM.
Index	4	LOW	Index recommendations found.
Statistics	5	MAINTENANCE	Gather and recollect all of relevant statistics for this query.

Filter recommendations by severity

Summary - 1 Initial Analysis | Query Revision

Recommendation 1: Provide a predicate on column SR_RETURNED_DATE_SK.

SQL Text

```
SELECT TPCDS.STORE_RETURNS.SR_CUSTOMER_SK AS CTR_CUSTOMER_SK
     , TPCDS.STORE_RETURNS.SR_STORE_SK AS CTR_STORE_SK
     , SUM(TPCDS.STORE_RETURNS.SR_RETURN_AMT) AS CTR_TOTAL_RETURN
FROM TPCDS.STORE_RETURNS
   , TPCDS.DATE_DIM
WHERE TPCDS.DATE_DIM.D_YEAR = 2000
   AND TPCDS.STORE_RETURNS.SR_RETURNED_DATE_SK = TPCDS.DATE_DIM.D_DATE_SK
GROUP BY TPCDS.STORE_RETURNS.SR_CUSTOMER_SK,
  TPCDS.STORE_RETURNS.SR_STORE_SK
```

Highlight relevant components of the query

Recommendation and rationale

Query Advisor checks for

- ***Missing join predicate for referential constraint***
- ***Predicates that can be rewritten as indexable***
- ***Use local predicates for joining tables or single table.***

Recommendation Details

Provide a predicate on column SR_RETURNED_DATE_SK.

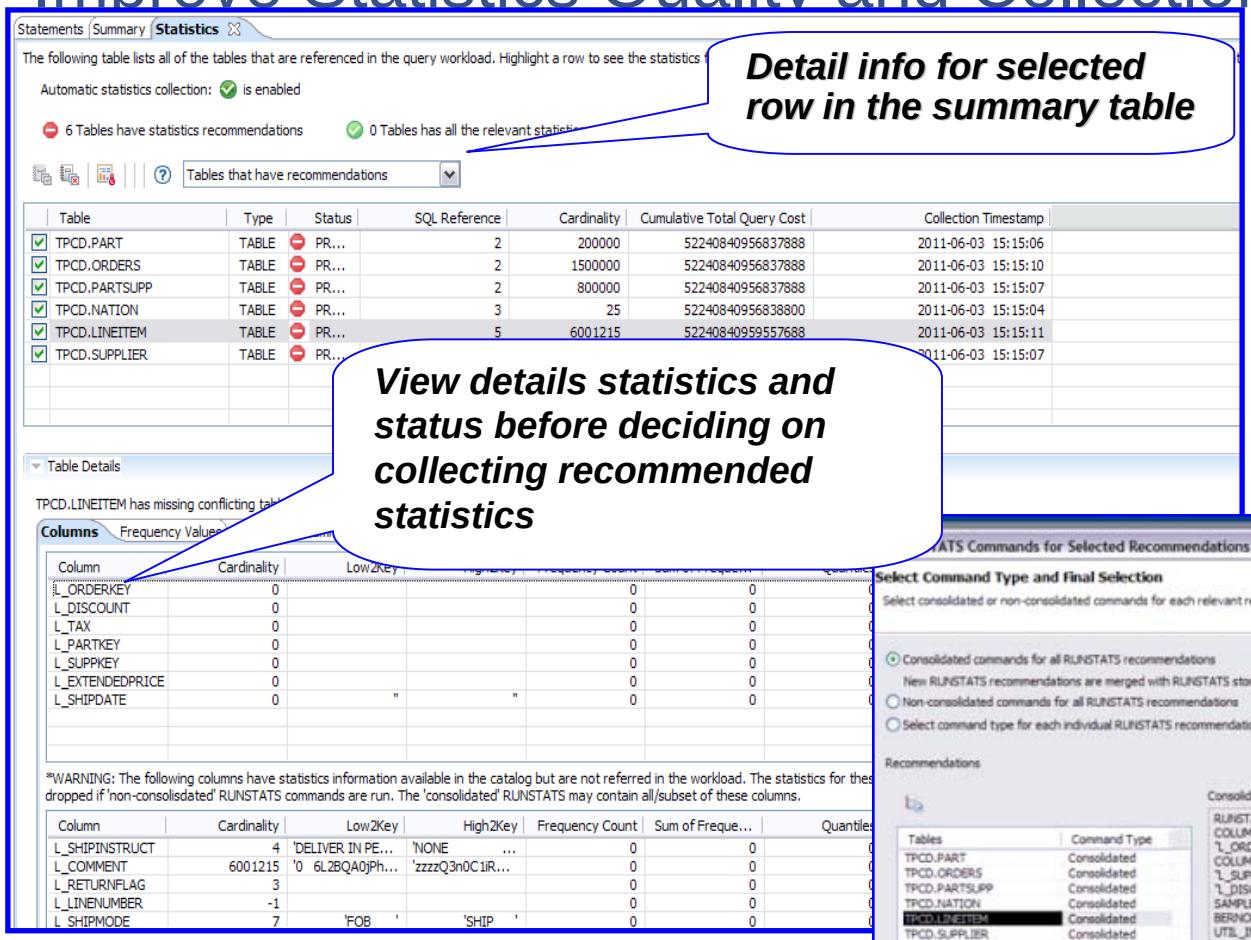
Consider adding the following predicate to column SR_RETURNED_DATE_SK in table STORE_RETURNS:
SR_RETURNED_DATE_SK IS NOT NULL
to filter the table earlier and unlock more possible join sequences.

Explanation

If a predicate filters out NULL columns and data type mismatches earlier in processing, performance can be improved.

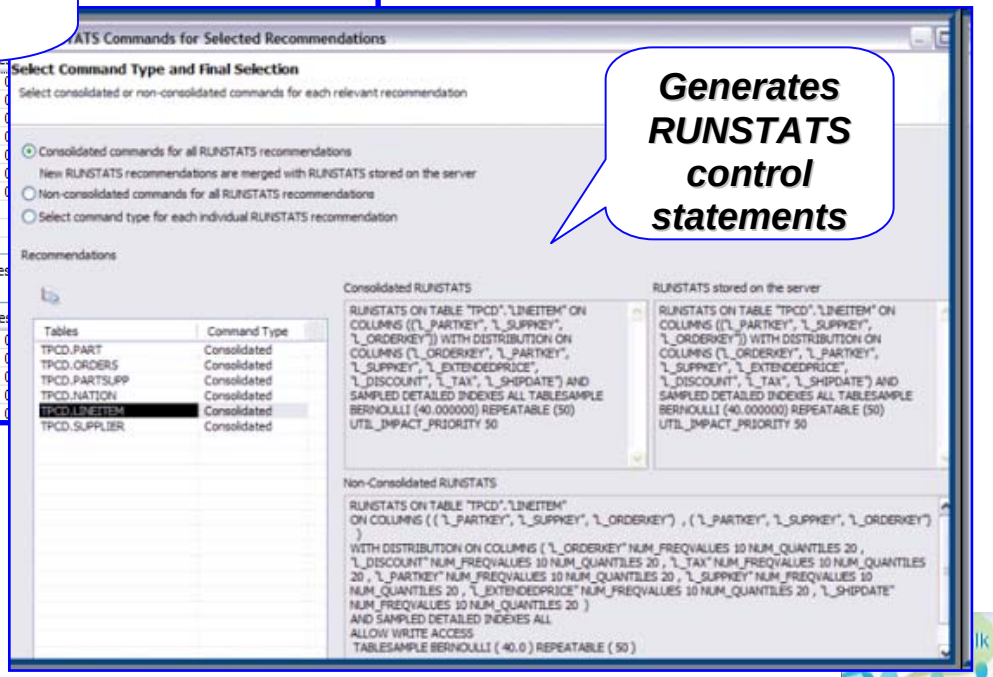
© 2012 IBM Corporation

Improve Statistics Quality and Collection



Detail info for selected row in the summary table

View details statistics and status before deciding on collecting recommended statistics



Generates RUNSTATS control statements

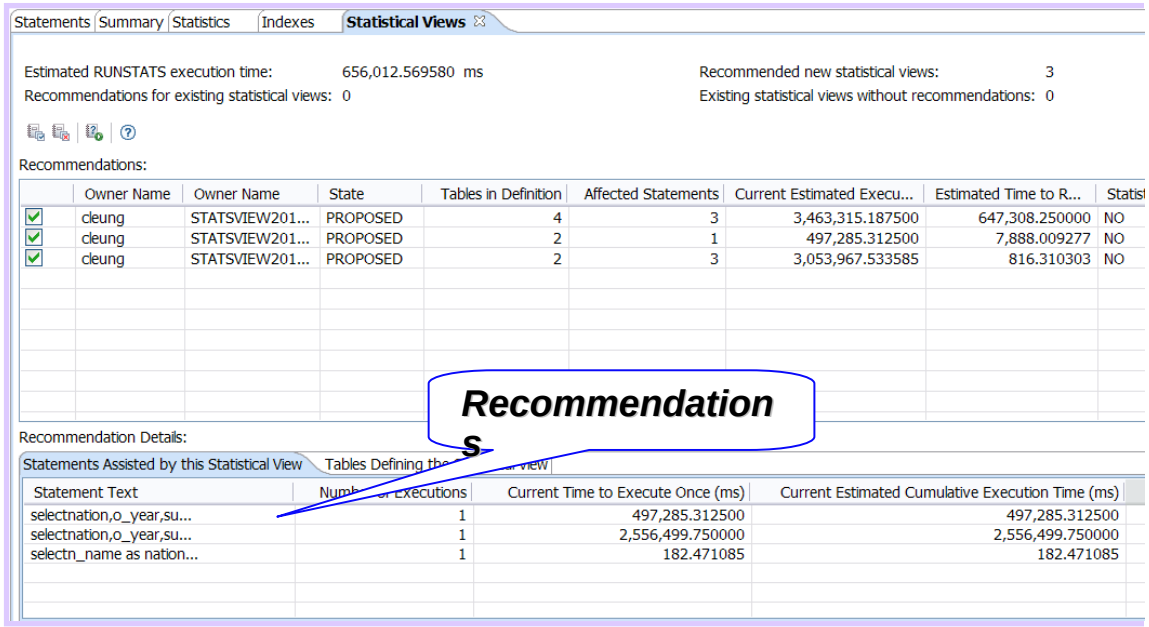
- *Provides advice on*
  - *Missing statistics*
  - *Conflicting statistics*
  - *Out-of-date statistics*
  - *Statistical views*
- *Simplify use*
  - *Provides Runstats control statements*
  - *Run immediately or save*

- *Better query performance*
- *Reduced CPU consumption*
- *Improved maintenance window throughput*

© 2012 IBM Corporation

Workload Statistics Advisor and Statistical Views

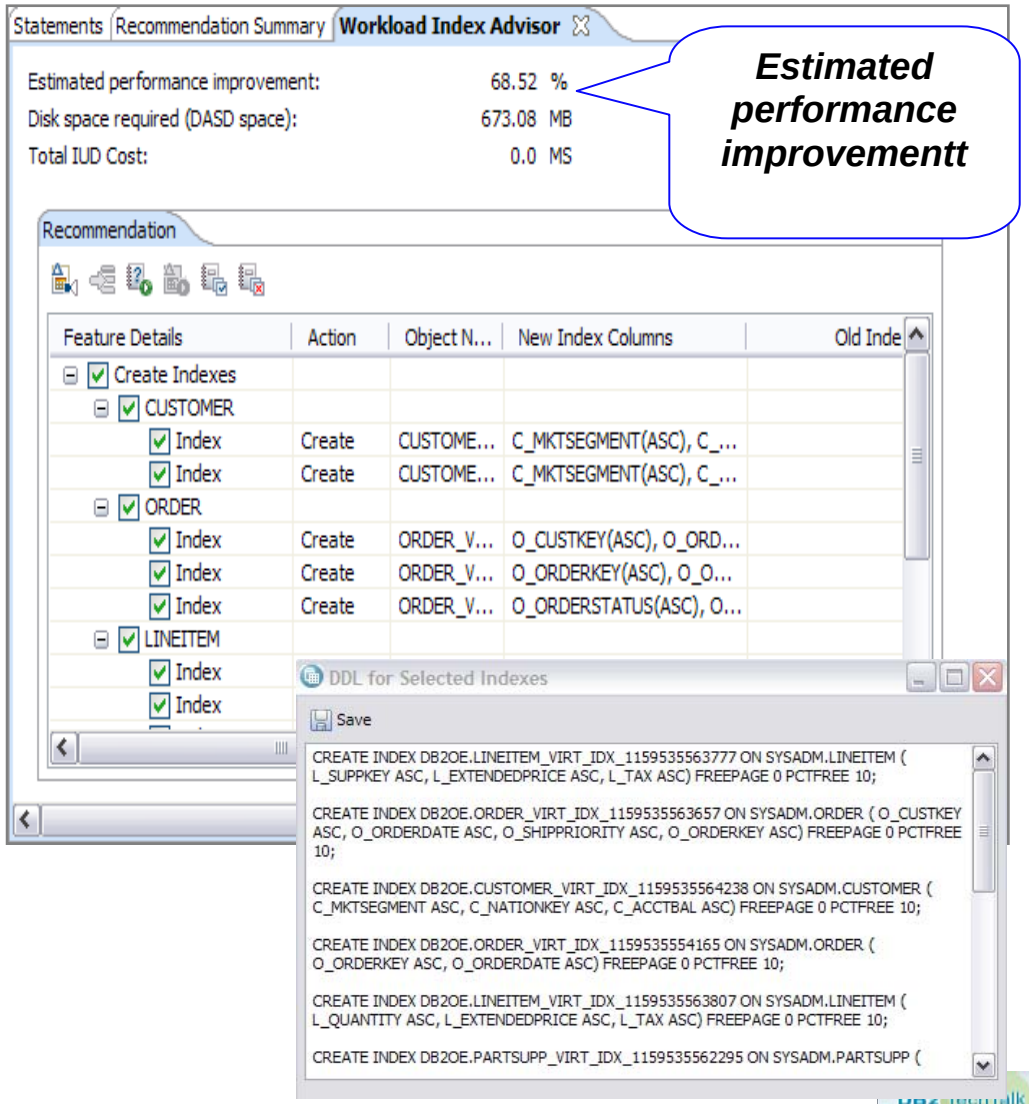
- **Challenge**
  - Data skew and join predicate filter [mis]calculation
    - WHERE salary+bonus > …
  - Overloaded dimension in a star join
    - Time dimension may contain 20 years of data but the transaction table keeps only the last 5 quarters of data. Uniform distribution is the wrong assumption after the join.
- **Solution: Generate statistical views**



© 2012 IBM Corporation

Indexing Advice to Improve Query Efficiency

- **Improve query efficiency**
  - Indexing foreign keys in queries that do not have indexes defined
  - Support for index only access
- **Simplify use**
  - Consolidate indexes and provide a single recommendation
  - Enables what-if analysis
  - Shows estimated improvement
- **Test before deployment**
  - Utilize virtual index capabilities built into the DB2 engine



© 2012 IBM Corporation

IBM DB2 Advanced Recovery Solutions

Save time, reduce errors, and meet SLAs

DB2 Merge Backup

- *Improve speed of your backup & recovery processes*
- *Minimize application impact*

Backup

DB2 Recovery Expert

- *Recover faster with greater granularity while protecting your critical business data*
- *Eliminate data errors before they compound into costly business mistakes*
- *Track and report data changes in response to auditing requests*

Recover

Optim High Performance Unload

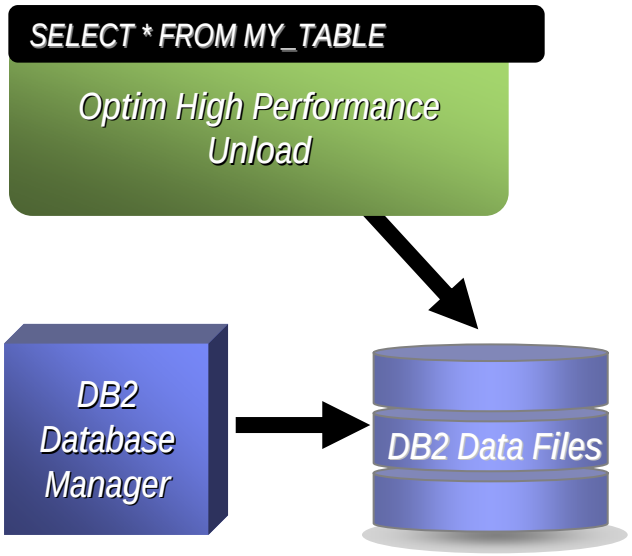
- *Extract large amounts of data quickly and with minimal impact on system productivity*
- *Perform full data and system migrations from one DB2 instance to another*

Unload

© 2012 IBM Corporation

InfoSphere Optim High Performance Unload

Rapidly unload, extract, and repartition data across enterprise systems



Total Data Unloaded:	2,905,294,900 rows
Complete database export time by usual methods	23 hours 54 minutes
Time Taken by DB2 HPU	1 hour 30 minutes

"That's about 16 times faster without using any special features, just out of the box performance."

Requirements

- *Quickly move large volumes of data between different sources without affecting the production system*
- *Reduce storage costs*
- *Fit more work into shrinking batch windows*

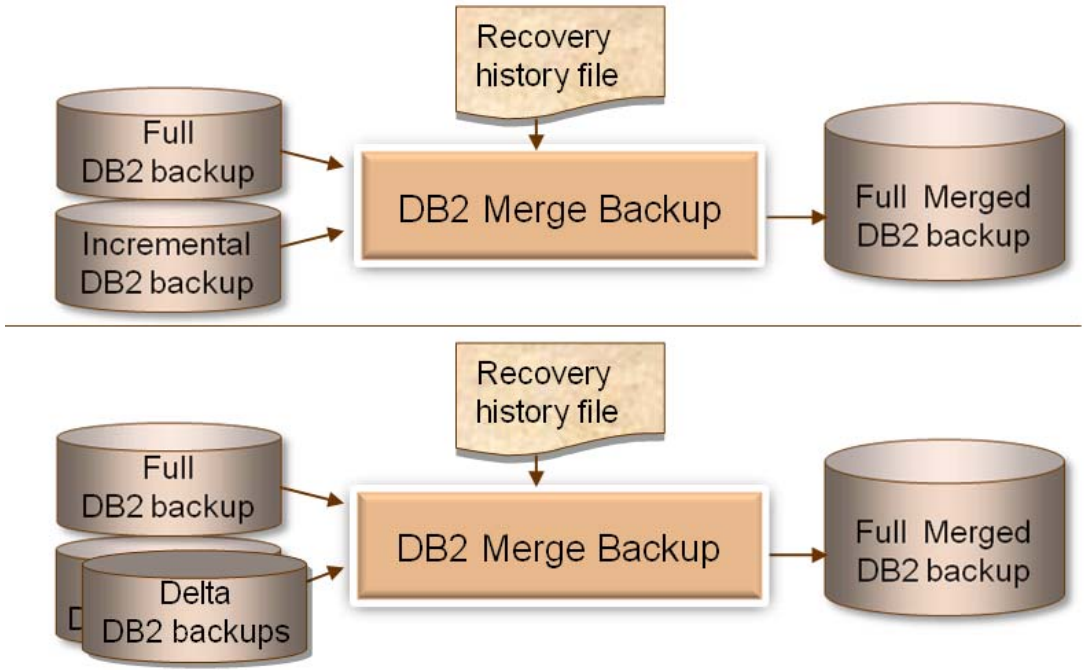
Benefits

- *Meet SLA's with high speed data unload capabilities*
- *Move data between systems while production systems are fully functioning*
- *Eliminate the need for intermediate storage*
- *Shorten data recovery time*

© 2012 IBM Corporation

DB2 Merge Backup

Accelerate DB2 backups for improved availability



Requirements

- *Meet SLAs for accuracy and responsiveness*
- *Perform backup/recovery with minimal impact*
- *Scaled approach for responding to unplanned service interruptions*
- *Ensure a complete, accurate DB2 backup*

Benefits

- *Perform backup and recovery faster*
- *Create incremental backup to minimize production impact*
- *Merge incremental with full backups to meet SLAs*

© 2012 IBM Corporation

DB2 Merge Backup Scenario

Scenario

If you have the following backup schedule

- Sunday – Full Backup
- Monday – Delta Backup, Create new full backup using DB2 Merge Backup
- Tuesday – Delta Backup, Create new full backup using DB2 Merge Backup
- Wednesday – Delta Backup, Create new full backup using DB2 Merge Backup
- Thursday – BANG. Table is accidentally dropped

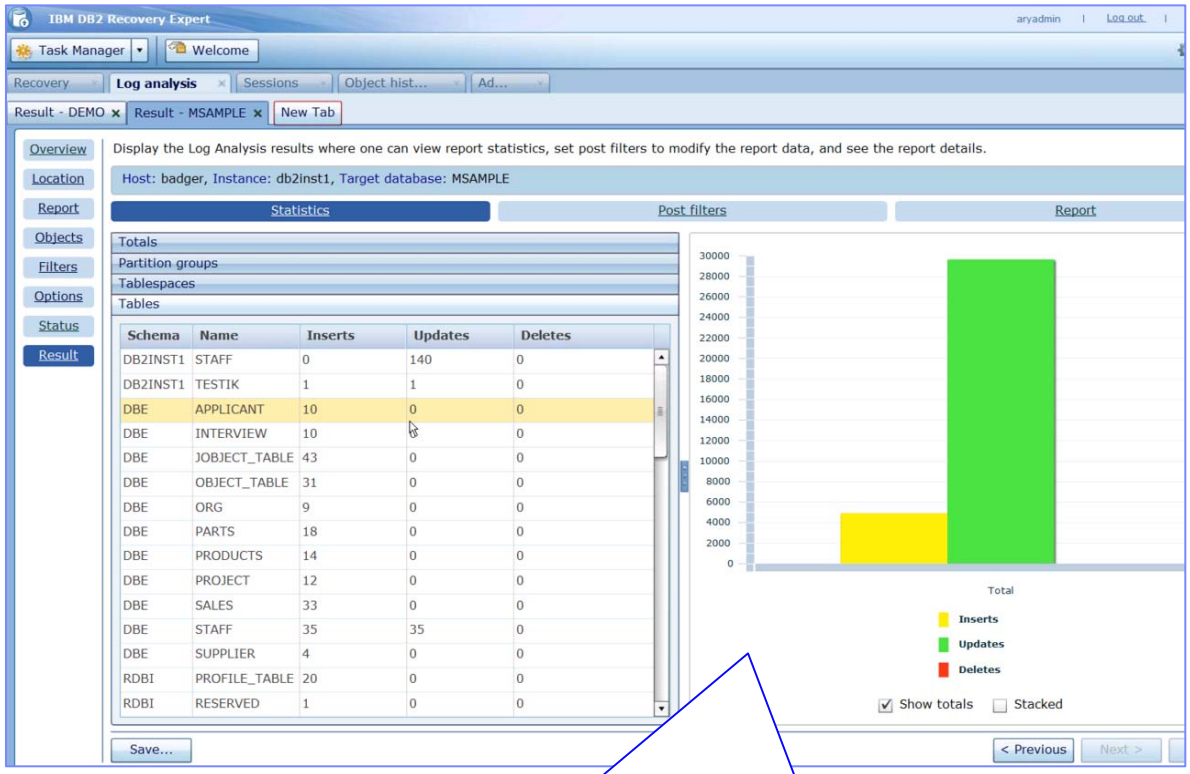
- **Use DB2 Merge Backup and the delta backup to use in the recovery**
- **Without DB2 Merge Backup, you would need to restore from the full backup from Sunday and 3 delta backups!**

© 2012 IBM Corporation

DB2 Recovery Expert

Reduce recovery window to increase availability

enterprise systems



Use Log Analysis to figure out who is doing what with your database

Requirements

- *Recover lost data with minimal impact to the business*
- *Scaled approach for responding to accidental data loss or corruption*
- *Ensure a complete, accurate DB2 backup*

Benefits

- *Significantly reduce data recovery times*
- *Allows DBAs to quickly restore or correct erroneous data with less resource and minimal disruption*

© 2012 IBM Corporation

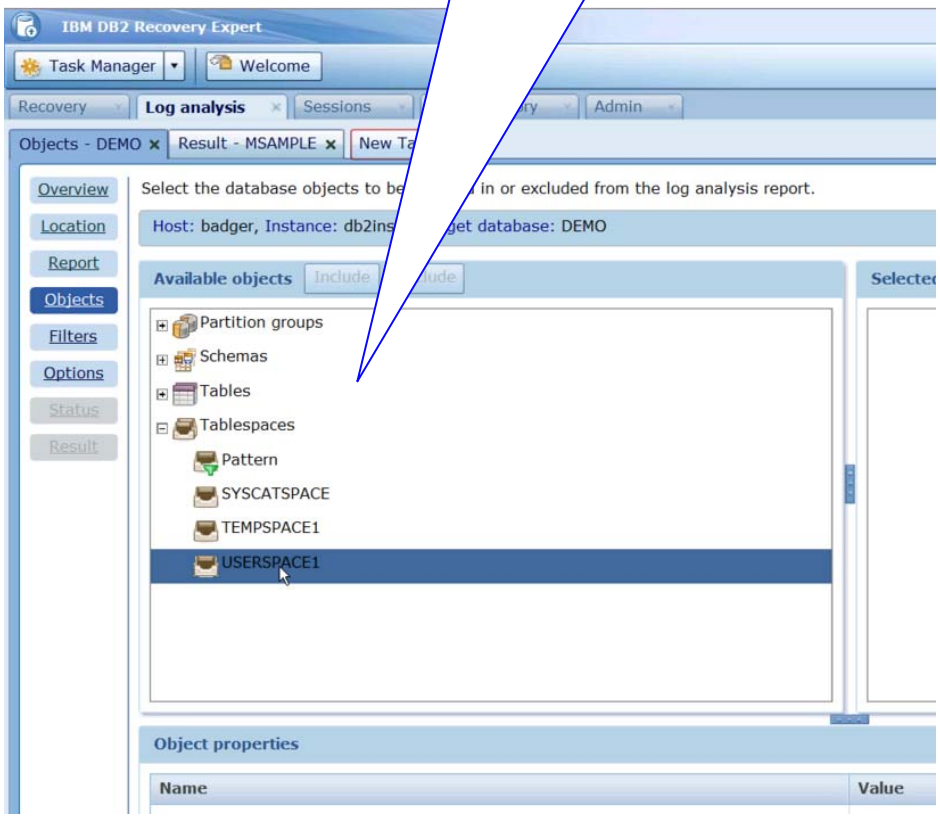
DB2 Recovery Expert

Scenarios

An accident occurs

- I ran the weekly batch job instead of the monthly job
- I dropped a production table instead of a test table
- I ran a delete with the wrong date parameters and ended up deleting too many rows

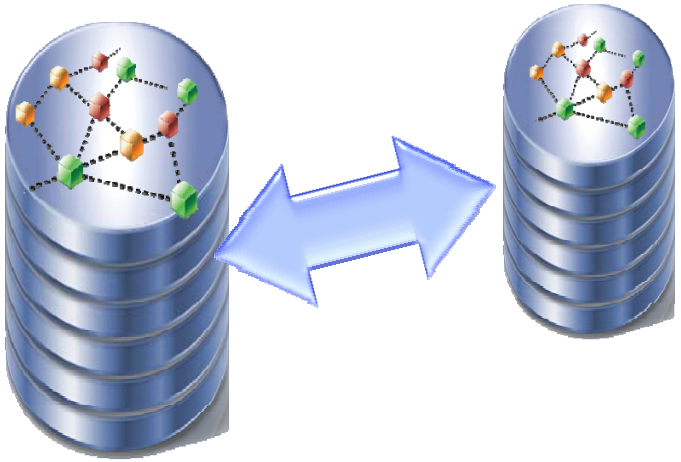
Web UI helps to easily navigate DB2 logs



© 2012 IBM Corporation

High Availability
InfoSphere Replication Server

- **Bi Directional Synchronisation**
  - Ability to synchronize across sites in real-time
  - Synchronize Entire Warehouse or Subsets
  - Extendable to multiple sites



- **Seamless Customer Failover**
  - Automatic Failover of Clients to Secondary Sites
  - Seamless to Applications

- **Rolling Upgrades**
  - Ability to Upgrade Warehouse environments sequentially without loss of service

- **Active – Active**

© 2012 IBM Corporation

IBM InfoSphere Warehouse

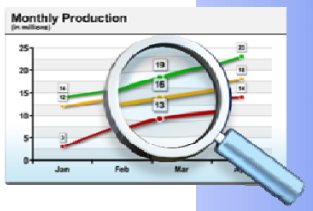
Solve problems – Create new opportunities

FLEXIBILITY

Everything to manage your data across its lifecycle

PERFORMANCE

DB2 9.7 with partitioning, compression, and workload management



SOPHISTICAED ANALYTICS

In database analytics and Cognos reporting

© 2012 IBM Corporation

IBM Smart Analytics System Offerings Deliver Analytics Value Without Forcing Migration

Faster Time to Value, Faster Business Results

Meeting clients where their information is…

5600

Based on System x

…Designed for business analytics workloads

…Optional Solid State Disk – reducing data latency

5710 New

Based on System x

…Cost-effective solutions for analytics and BI, reporting

…Compact, integrated single analytics solutions

…Available for mid-market

7700

Based on POWER7 Servers

…Scaling to hundreds of terabytes of data

…Extract insights from untapped information

7710 New

Based on Power7 Server

…A single server warehousing and analytics solution

…Built on POWER7 based servers and designed for production data warehouses - sizes under 10 TB,

…For development and non-production use.

9600

Based on Based on System z

…Advanced query /workload management

…Database designed and optimized for system

…Disk controller, optimized to reduce data latency

© 2012 IBM Corporation

Communities

- **On-line communities, User Groups, Technical Forums, Blogs, Social networks, and more**
  - Find the community that interests you…
    - **Information Management** ibm.com/software/data/community
    - **Business Analytics** ibm.com/software/analytics/community
    - **Enterprise Content Management** ibm.com/software/data/content-management/usernet.html

- **IBM Champions**
  - Recognizing individuals who have made the most outstanding contributions to Information Management, Business Analytics, and Enterprise Content Management communities
    - ibm.com/champion

© 2012 IBM Corporation

More information

- **Analyst report from ITG**
  - Bottom-line Advantages of IBM InfoSphere Warehouse
- **IBM White Paper**
  - IBM InfoSphere Warehouse 9.7: Deliver actionable business insights

Next DB2 Tech Talk Coming in March:

Database Archiving: Best practices to improve performance of your DB2 LUW and z/OS databases

Registration coming soon to DB2 Tech Talk page

© 2012 IBM Corporation

Questions

© 2012 IBM Corporation

Thank You and Follow-up!

ibm.com/db2/labchats

© 2012 IBM Corporation

Acknowledgements and Disclaimers:

Availability. References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates.

The workshops, sessions and materials have been prepared by IBM or the session speakers and reflect their own views. They are provided for informational purposes only, and are neither intended to, nor shall have the effect of being, legal or other guidance or advice to any participant. While efforts were made to verify the completeness and accuracy of the information contained in this presentation, it is provided AS-IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this presentation or any other materials. Nothing contained in this presentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.

© Copyright IBM Corporation 2011. All rights reserved.

- ***U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.***

IBM, the IBM logo, ibm.com, DB2, InfoSphere, and Optim are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml

Other company, product, or service names may be trademarks or service marks of others.