# Information On Demand

## Unlock the Business Value of Information

IBM developerWorks Live!

Abbreviated Title Here

# Agenda

| | |
|---|---|
| The Business Value of New Intelligence | 30 min |
| Building a Trusted View of your Data with Information Server | 60 min |
| *** Break *** | 15 min |
| Understanding your Data with Cognos BI | 50 min |
| Video: Cognos BI and Information Server | 10 min |
| Why Choose DB2 as the Foundation for New Intelligence | 55 min |

IBM

**IBM Information On Demand
Unlock The Business Value Of Information**

# The Business Value of New Intelligence

IBM developerWorks Live!

Abbreviated Title Here

# Something Meaningful Is Happening...

*"Every human being, company, organization, city, nation, natural system and man-made system is becoming* **interconnected, instrumented and intelligent.** *This is leading to new savings and efficiency—but perhaps as important, new possibilities for progress."*

# More Information Than Ever Before

- **Volume of Digital Data\***

  ▸ 15 petabytes of new information generated every day

    ▪ 8x more than the information in all U.S. libraries

- **Variety of Information**

  ▸ Creating new types of data everyday

    ▪ Digital TV, sensor-based applications, social network updates, …

- **Velocity of Decision Making**

  ▸ Businesses are challenged to utilize complex information everyday

\*IDC White Paper – "The Diverse and Exploding Digital Universe", 2008

# Follow The Data !

- Old Intelligence
  - Experts make decisions based on what they know
    - Well known data
    - Experience

- New Intelligence
  - Employees make decisions based on changing data
  - Use tools to -
    - Find all data relevant to the task
    - Derive understanding from the data
    - Act on the data

# Businesses Need New Intelligence

## Businesses must use data in a strategic way to accomplish business goals

- Business Goals
  - ▶ Increase revenue and productivity
  - ▶ Reduce risk
  - ▶ Flexibility to adapt quickly to changing market conditions

- New Intelligence is a strategic way to get more business value from your data
  - ▶ Capture information on-line
  - ▶ Integrate information specific to the task
  - ▶ Analyze your data to create new intelligence
  - ▶ Deliver new intelligence to the people who need it

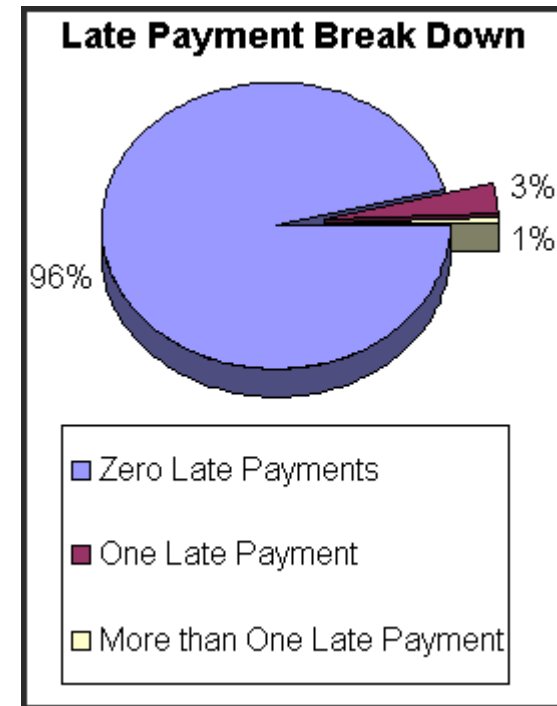# Service Oriented Finance Needs To Reduce Risk

I have an opportunity to resell our risky mortgages.

Our mortgage business seems to be fine but I know some of our customers are struggling.

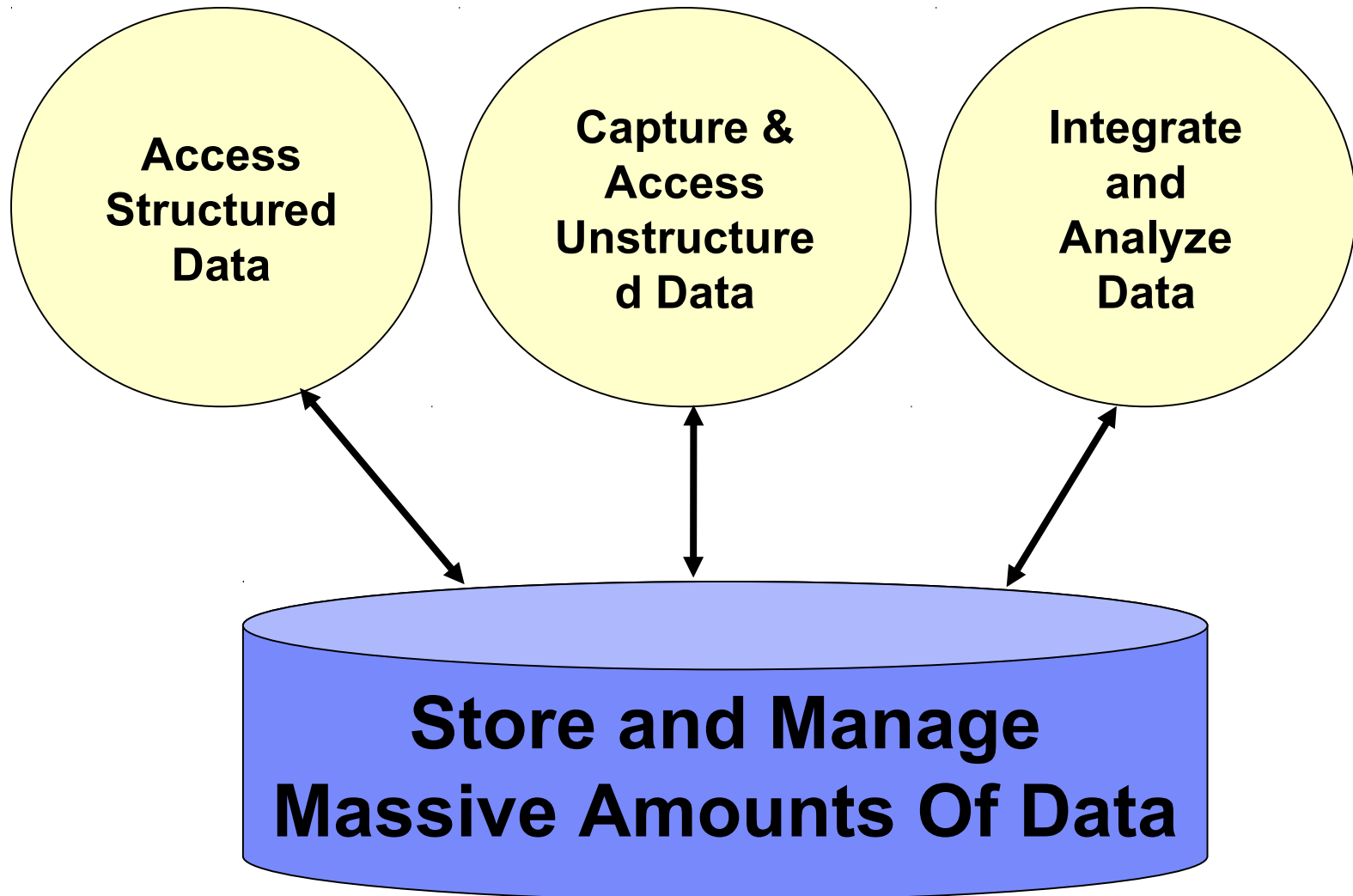I need to find out which ones I should remove from the balance sheet

**Service Oriented Finance CEO**

### Late Payment Break Down

96%

3%

1%

- Zero Late Payments
- One Late Payment
- More than One Late Payment

**Mortgage Business Spreadsheet Report**

**Unlock The Business Value Of Information**

# Create New Intelligence Through A Smarter Information Management Strategy

# First You Need To Efficiently Capture Your Data

- **Traditional transactional data is changing**

  ‣ Increased scale and performance is needed to handle growth

  ‣ Data growth also requires more efficient storage use

  ‣ New data type standards, required by business partners for easier integration

- **Growth of Digital Content (unstructured data)**

  ‣ Capture paper forms on-line for efficiency

  ‣ Agility to easily optimize/change current content flow for new business

  ‣ Manage content in compliance with regulatory laws to reduce risk

# Then Gather And Integrate All The Data Needed For A Specific Task

First you need to understand what data you have
- Need to make sure the data contains what you think it does
- And is described in your business terms

Then you can decide which data is needed
- Understand which data is the data of record for your task

And integrate all the various data sources for your task
- Either by accessing the data in its original location, or extracting it into a new system depending on your needs

# Finally You Need Tools To Analyze The Data And Deliver It To The People Who Need It

- Easy to use web tooling
  - Decreases development cost and supports easy deployment and delivery

- That creates common analysis and report artifacts
  - A common data model that works across the tools enables the greatest collaboration and reuse

- And produces output in most common standard formats
  - Makes sure the data is available to the people that need it in the format they can use it in

# Creating New Intelligence Requires Leveraging Data Across All Lines Of Business

**We already know the mortgages that have defaulted.**

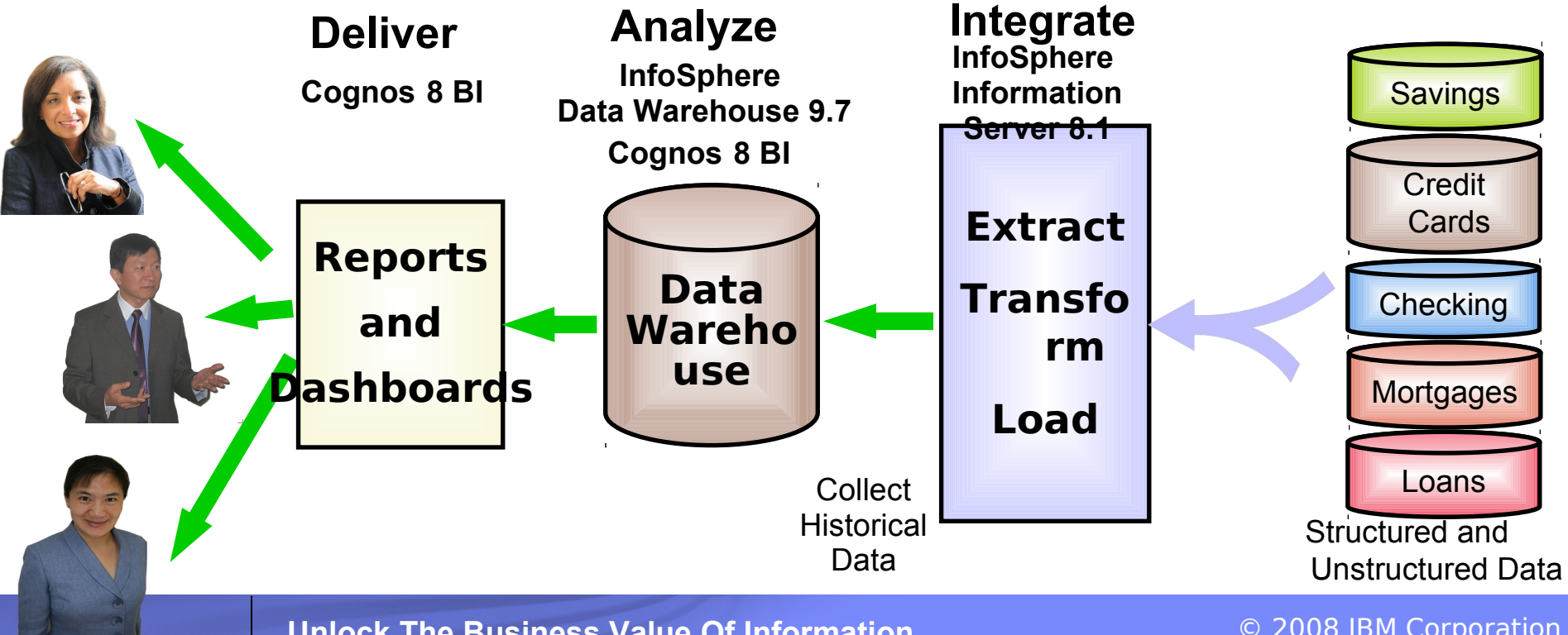**We need to predict which ones will default in the future.**

**Getting a complete picture of your customers is essential. Once you have that you'll be able to identify risky mortgages.**

**Service Orient Finance**
**Mortgage Line of BusinessVP**

**IBM**

# Service Oriented Finance Can Create New Intelligence With IBM Information Management Software

- Capture all required data online
  - ‣ Both unstructured and structured
- Gather (integrate) all relevant information for the task at hand
- Analyze relevant data for actionable insight
- Deliver it to the people who need it

**Capture**
**DB2  /Informix**

**Deliver**

**Cognos  8 BI**

**Analyze**
**InfoSphere**
**Data Warehouse 9.7**
**Cognos  8 BI**

**Integrate**
**InfoSphere**
**Information**
**Server 8.1**

**Reports and Dashboards**

**Data Warehouse**

**Extract Transform Load**

Collect Historical Data

Savings

Credit Cards

Checking

Mortgages

Loans

Structured and Unstructured Data

# Information Agenda

**Old Intelligence**



*Separate individual projects and initiatives*

**Information Agenda**

*Transform information into a trusted strategic asset that can be rapidly leveraged across applications, processes and decisions*

**New Intelligence**



*Sustained competitive advantage*

- A strategy to create new intelligence for your organization
  - ‣ Establish information driven strategy to enable business objectives
  - ‣ Prioritize information intensive projects aligned with the strategy
  - ‣ Use agile technology and leverage existing information for speed and flexibility
  - ‣ Design trusted information to sustain competitive advantage
- IBM can help you do this !

**IBM**

# IBM Information On Demand
## Unlock The Business Value Of Information

# Building A Trusted View
# Of Your Data With Information Server

IBM developerWorks **Live!**

Abbreviated Title Here

# Service Oriented Finance Needs To Leverage Customer Data Across All Lines Of Business



**We have customer data, but currently it is in different databases in different locations.**

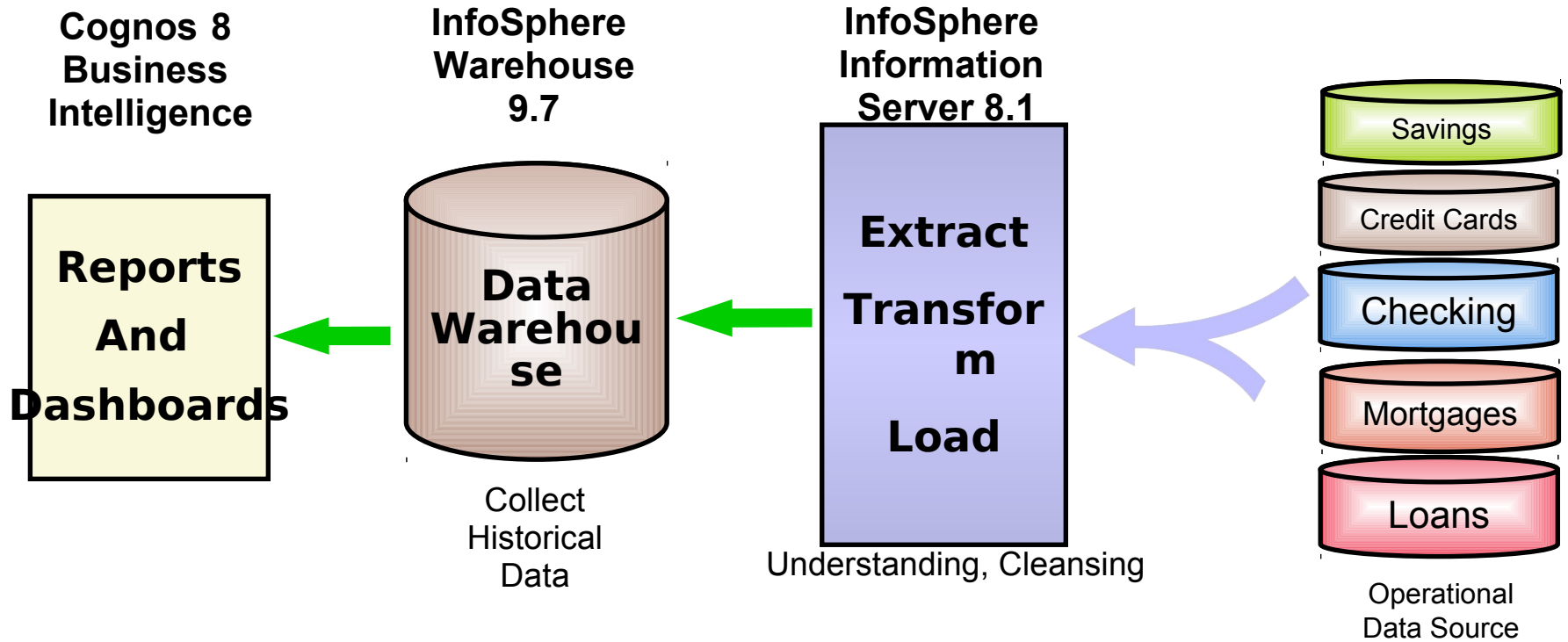**You need to consolidate your information in a Data Warehouse.**

**Mortgage Line of Business VP**

**Service Oriented Finance CIO**

# Service Oriented Finance Needs A Complete View Of Their Customers

- **Problem**: Segmented lines of business store their data in isolated silos
    - ▶ Banking, Credit Cards, Mortgage, Consumer Loans

- **Requirement**: Identify customer trends faster by viewing data from across all areas of business

- **Solution**: Create new intelligence by building an enterprise data warehouse containing a complete view of customer information

# Create New Intelligence With IBM Information Management Software

**Cognos 8 Business Intelligence**

**InfoSphere Warehouse 9.7**

**InfoSphere Information Server 8.1**

Savings

Credit Cards

**Reports**

**And**

**Dashboards**

**Data Warehouse**

**Extract**

**Transform**

**Load**

Checking

Mortgages

Loans

Collect Historical Data

Understanding, Cleansing

Operational Data Source

## Analysis Tools

## ETL Processing

## Scalable Database

## Connections

# InfoSphere Warehouse 9.7 Is Built For Data Warehouses

**InfoSphere Warehouse 9.7**

**Data Warehouse**
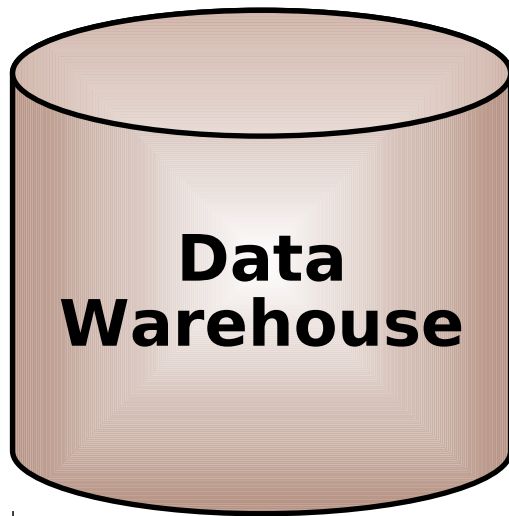
Microsoft SQL Server is limited by the scalability of Windows on Intel

- Built on DB2 9.7

- Capabilities include:
  - ▶ Data partitioning (enables scale out)
  - ▶ Workload management
  - ▶ Data Compression to save storage
  - ▶ Basic analysis support
  - ▶ Data Studio tools based on Eclipse

- Runs on AIX, Red Hat, SuSE, Windows, HP-UX, Solaris and System z

# The Competitors Can't Match IBM InfoSphere Warehouse On Power Systems

- Integration of IBM InfoSphere Warehouse, AIX, and Power Systems hardware gives IBM an advantage

- Oracle on HP or SUN can't beat IBM

  - ▸ Oracle is designed to run on commodity servers

  - ▸ No optimization specifically for HP servers

- SQL Server on Intel can't beat IBM

  - ▸ Microsoft SQL Server is restricted to the Windows platform

  - ▸ Some portions of SQL Server run in the windows kernel

**10TB TPC-H Performanc[e]
QphH @ 10000GB
(higher is better)**

| | |
|---|---|
| DB2 on P570 128 Cores | 343,551 |
| Oracle on HP Superdome 128 Cores | 208,457 |
| Oracle on Sun SunFire 144 Cores | 108,099 |
| SQL Server 2008 on Unisys 64 Cores | 80,173 |

*TPC-H results from tpc.org as of July 2009*
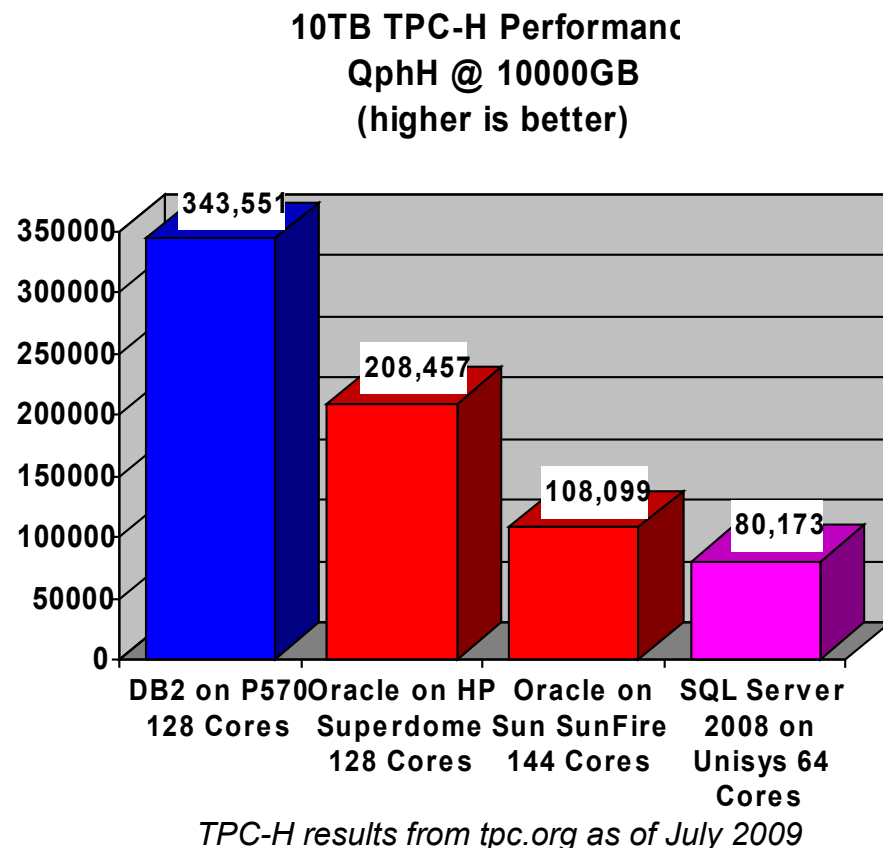
## *InfoSphere Scales Bigger and Performs Better than Competitors*

# Industry Data Models Help You Get Started



We need help to get started.

Service Oriented Finance
CIO

IBM industry data models can help you get started quickly.

IBM

# IBM Industry Data Models Accelerate Your Data Warehouse Solution

- Industry Data Models are:
  - Best practices from over 400 IBM clients
  - Built on InfoSphere Information Server and InfoSphere Data Architect

- Industry Data Models include:
  - Enterprise Data Warehouse (EDW) Model
  - Business Terminology Data Model
  - Business Solution Templates (BST)

- Industry Data Model Business Benefits
  - 83% report their Data Warehouse is better aligned with business needs
  - Over 50% report that businesses are now getting the information they want

- Industry Data Model Development Benefits
  - 15-20% cost savings to build the warehouse
  - 20-25% decrease in the time spent in design phase
  - 30-40% decrease in time spent in the modeling phase

**- Source Hurwitz**

**Unlock The Business Value Of Information**

# IBM Has Industry Data Models For Many Enterprises

## Banking
**(Banking Data Warehouse)**

- Profitability
- Relationship Marketing
- Risk Management
- Asset and Liability Mgmt
- Compliance

## Financial Markets
**(Financial Markets Data Warehouse)**

- Risk Management
- Asset and Liability Mgmt
- Compliance

## Health Plan
**(Health Plan Data Warehouse)**

- Claims
- Medical Management
- Provider and Network
- Sales, Marketing and Membership
- Financials

## Insurance
**(Insurance Information Warehouse)**

- Customer centricity
- Claims
- Intermediary Performance
- Compliance
- Risk Management

## Retail
**(Retail Data Warehouse)**

- Customer centricity
- Merchandising Management
- Store Operations & Product Mgmt
- Supply Chain Management
- Compliance

## Telco
**(Telecommunications Data Warehouse)**

- Churn Management
- Relationship Mgmt & Segmentation
- Sales and Marketing
- Service Quality & Product Lifecycle
- Usage Profile

# InfoSphere Information Server Builds And Delivers Trusted Information

**Understand**

**InfoSphere** Information Analyzer

**Cleanse**

**InfoSphere** QualityStage

**Transform**

**InfoSphere** FastTrack
**InfoSphere** DataStage

**Deliver**

**InfoSphere** Change Data Capture

**Platform Services**
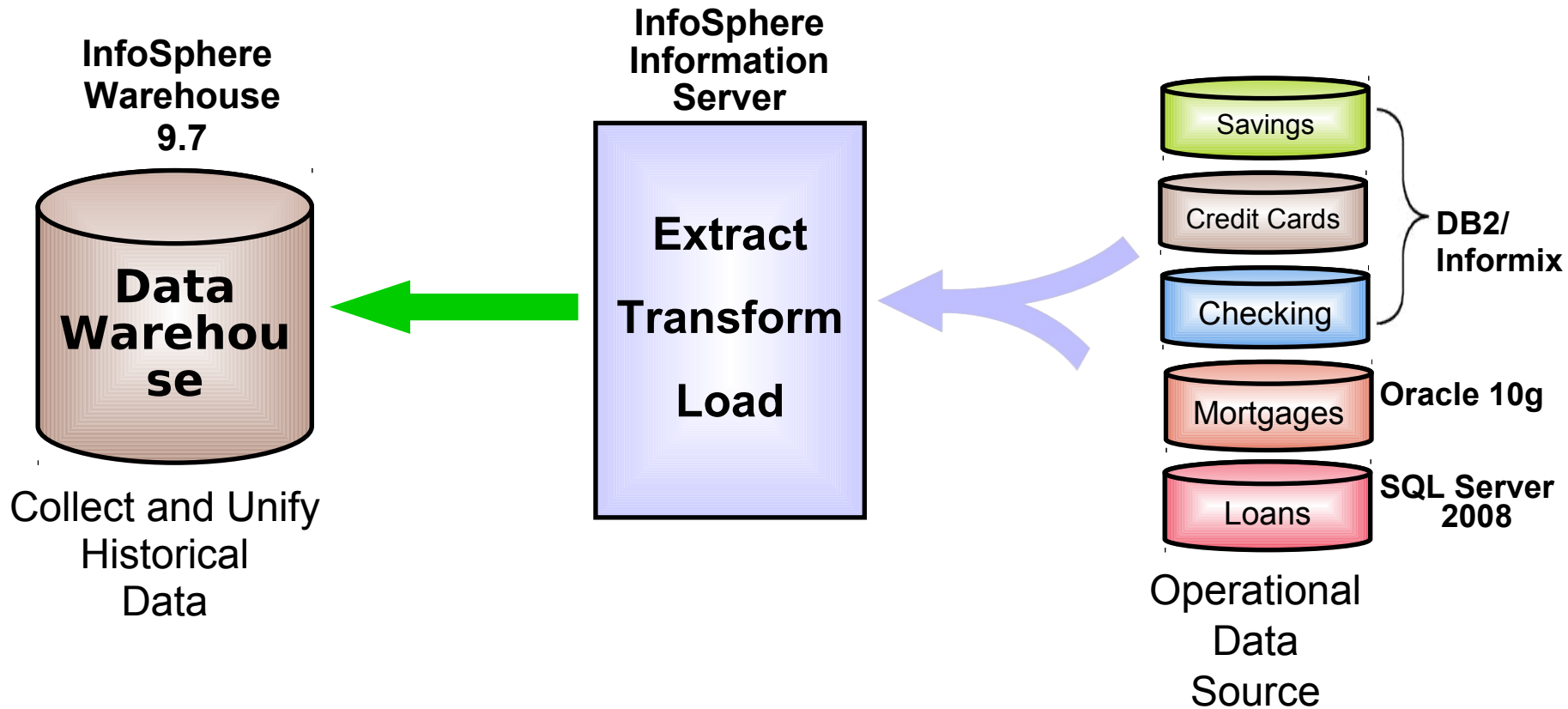
**Parallel Processing Services**

**Connectivity Services**

**Metadata Services**

**Administration Services**

**Deployment Services**

# InfoSphere Information Server Can Load Your Data Warehouse



**InfoSphere Warehouse 9.7**

**Data Warehouse**

Collect and Unify Historical Data

**InfoSphere Information Server**

**Extract**

**Transform**

**Load**

Savings

Credit Cards

Checking

**DB2/ Informix**

Mortgages

**Oracle 10g**

Loans

**SQL Server 2008**

Operational Data Source

# Extract, Transform, And Load (ETL) Jobs Map Data From Sources To Targets

**A few simple examples:**

**Mapping source columns to targets**

**Different column names and order**

**Generating new column values**

**Converting data types and formats**

| PROD D | CUST ID | SOURCE ID | QTY | BAL | SALEDATE |
|---|---|---|---|---|---|
| 000 101 | 100 | | 01 | 1 | |
| $10,000.00 | | 2007-02-28 | | | |
| 000 121 | 100 | | 01 | 3 | |
| $500.50 | | 2007-02-28 | | | |
| 000 102 | 101 | | 01 | 1 | $ |
| 20,000.00 | | 2007-03-01 | | | |

**Target:  Data Warehouse**

000 101   100      01          3
$20,000.00 2007-02-28

**Transform**

| PRODUCT | QTY | CUSTNO | BALANCE | DATE |
|---|---|---|---|---|
| 101 | 01 | 100 | | |
| 10000 | 02-28-2007 | | | |
| 121 | 03 | 100 | | |
| 500.50 | 02-28-2007 | | | |
| 102 | 01 | 100 | | |
| 20000 | 03-01-2007 | | | |

**Source:  Operational Data**

**A successful data integration project requires a detailed specification for the business goals and technical requirements!**

# InfoSphere FastTrack Creates Data Maps And Specifications For Your Data Integration Projects

- Create simplified data maps and transformations using drag and drop
  - ▶ Automatically discover source and target columns
    - Uses database introspection and Web 2.0-style tagging
    - Use business terms to accurately match source to target

- Data analysts and developers share project specifications
  - ▶ Collaboration and reuse improve productivity
  - ▶ Use metadata common to all Information Server tools
  - ▶ Standard formrats and centralized management for governance
    - Synchronize work across global teams

- Generate ETL code directly from job specifications
  - ▶ Reduces costs and errors in ETL job development

**Oracle & Microsoft don't offer any of these capabilities**

# InfoSphere FastTrack Automatically Discovers Data Mappings Using Business Term Tags

**Mortgage To Warehouse Mapping Specification**

| Source Discovery |
| :--- |
| Mortgage.Times_Past_Due |
| Checking.NSFCount |
| Loans.MissedPayments |

| Source | Target | Tag |
| :--- | :--- | :--- |
| Times_Past_Due | Credit_Events | Failure_To_Pay |
| Current_Balance | Ending_Balance | Period_Balance |
| Account_Num | Account_ID | Arrangement |
| Account_Holder | Customer_ID | Party_ID |

# InfoSphere DataStage Creates The Technical Implementation Of Data Integration Jobs

- ## Create jobs using a graphical, drag and drop interface
  - ▸ Select from hundreds of pre-built transformation and data quality functions
  - ▸ Batch & real-time operations

- ## Stores and retrieves metadata from Information Server
  - ▸ Allows easy reuse of integration work between projects

- ## Easily deploy data integration jobs as services for SOA applications
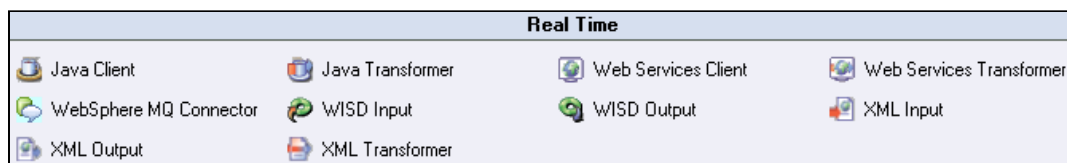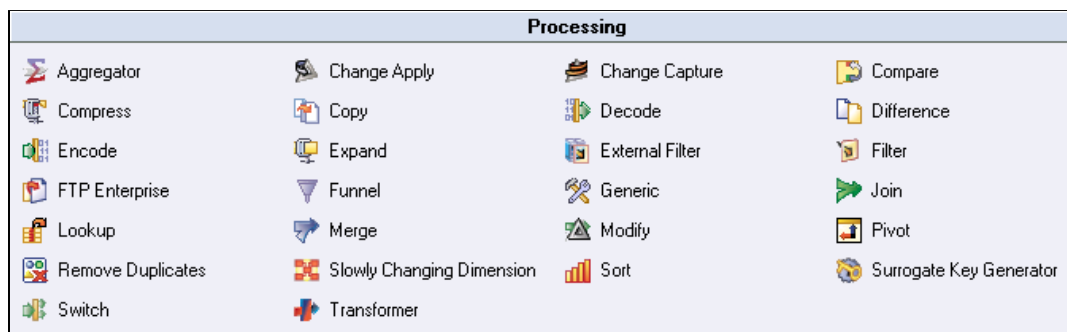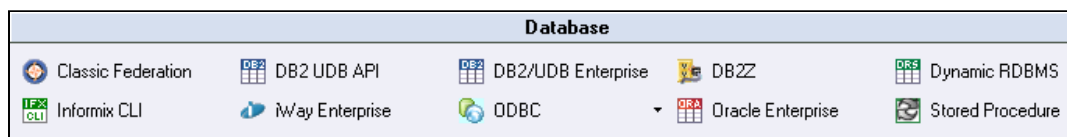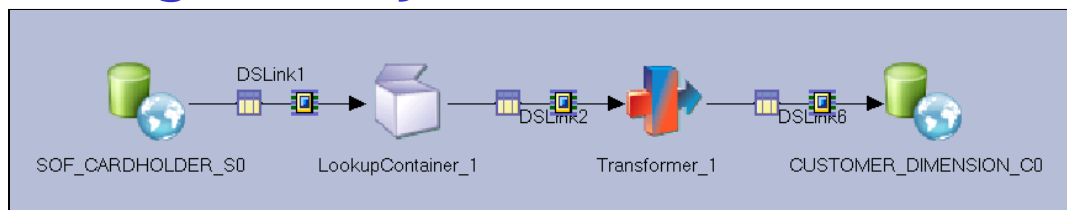
# IBM InfoSphere Information Server Connects To Almost All Sources Of Data

**RDBMS**
DB2 (on z, I, P or X series)
Oracle
Informix (IDS and XPS)
Ingres
MySQL
Netezza
Progress
RDB
RedBrick
SQL/DS
SQL Server
Sybase (ASE and IQ)
Teradata
Universe
UniData
NonStopSQL
And more…..
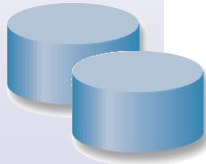
**General Access**
Sequential File
Complex Flat File
File / Data Sets
Named Pipe
FTP
Compressed / Encoded Data
External Command Call
Parallel/wrapped 3rd party apps
EMC InfoMover
Web logs
Unstructured: e-mail, docs, etc.
Content Management Systems
Life Sciences

**Enterprise Applications**
JDE/PeopleSoft EnterpriseOne
Oracle Applications
PeopleSoft Enterprise
SAS
SAP R/3 and BI
SAP XI
Siebel
JDA
Ariba
Manugistics
I2
And more…

**Standards and Real Time**
WebSphere MQ
Java Messaging Services (JMS)
Java
XML and XSL-T
EBXML
Web Services (SOAP)
Enterprise Java Beans (EJB)
EDI
FIX
SWIFT
HIPAA

**CDC / Replication**
DB2 (on z, I, P, X series)
Oracle
SQL Server
Sybase
Informix
IMS
VSAM
ADABAS
IDMS
NonStopSQL
Enscribe

**Legacy**
Allbase/SQL
C-ISAM
D-ISAM
Datacom/DB
DS Mumps
Enscribe
Essbase
FOCUS
IDMS/SQL
ImageSQL
Infoman
KSAM
M204
MS Analysis
Nomad
Nucleus
RMS S2000
Supra
TOTAL
TurboImage
Unify
And many more….

**Offering more connectivity than Oracle and Microsoft**

Unlock The Business Value Of Information

# InfoSphere DataStage Uses Parallel Processing for Enterprise Scalability
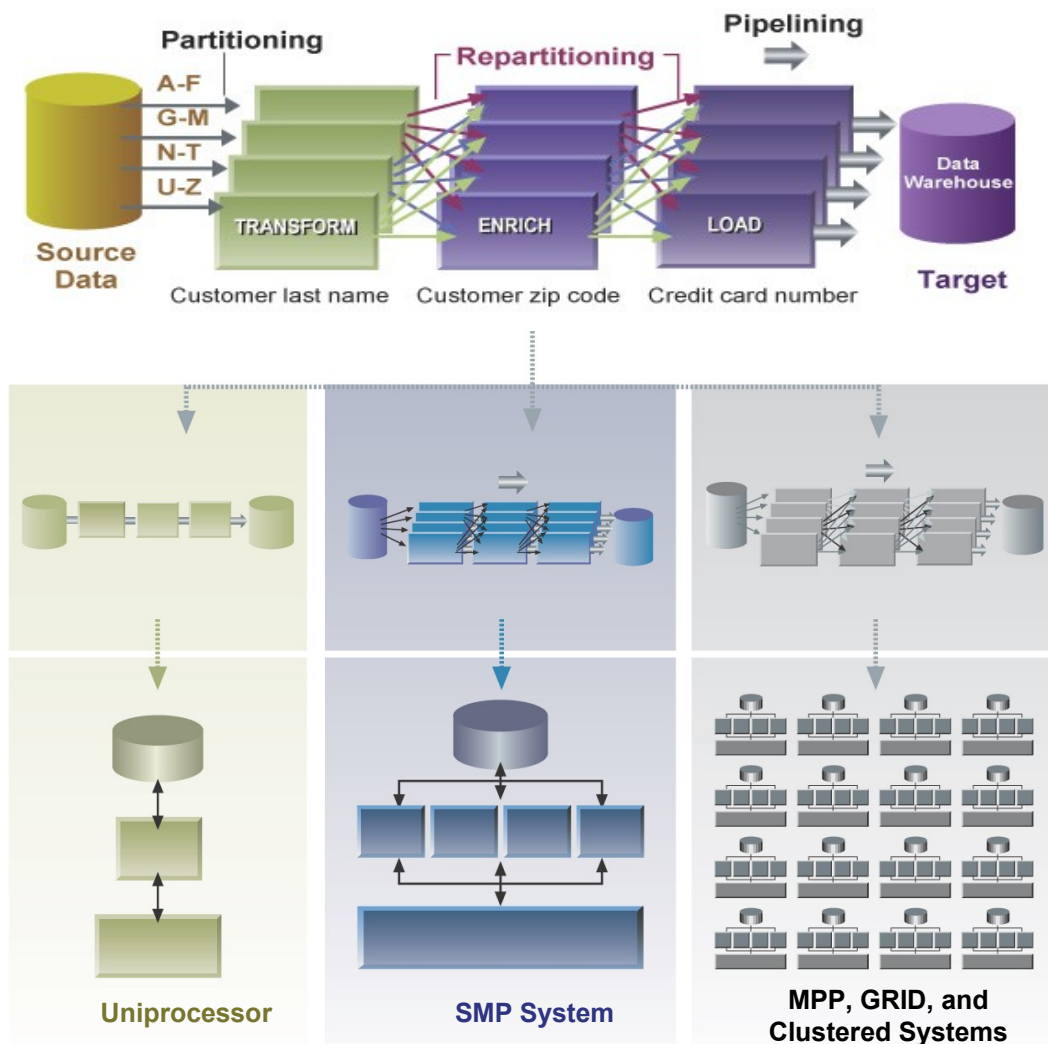


- **Supports major scalability options**
  - ▸ SMP, Clustered, GRID, and MPP platforms

- **Provides automatic dynamic data partitioning, repartitioning, and pipelining for optimal parallel performance**

- **Design integration processes without concern about underlying hardware architecture or number of processors**
  - ▸ Resources defined in a separate configuration file
  - ▸ Allows easy expansion to new hardware

# MGM Mirage Uses InfoSphere Information Server

## Customer Challenge

- MGM Mirage needed a single view of customer data from its many (26) casinos, hotels and other sources, including its new loyalty program (Players Club).
- They wanted to improve customer service, enable targeted marketing programs, and provide a critical basis for tactical and strategic decisions.
- Continued use of their custom-built, first generation EDW constrained their ability to change and expand the system as needed, and resulted in high costs.

## Solution

- Profiles source data, creates data quality processes, and integrates data into CRM warehouse, across all customers, properties and revenue lines.

- Web service process for real-time updates into warehouse for security, marketing and CRM activity.

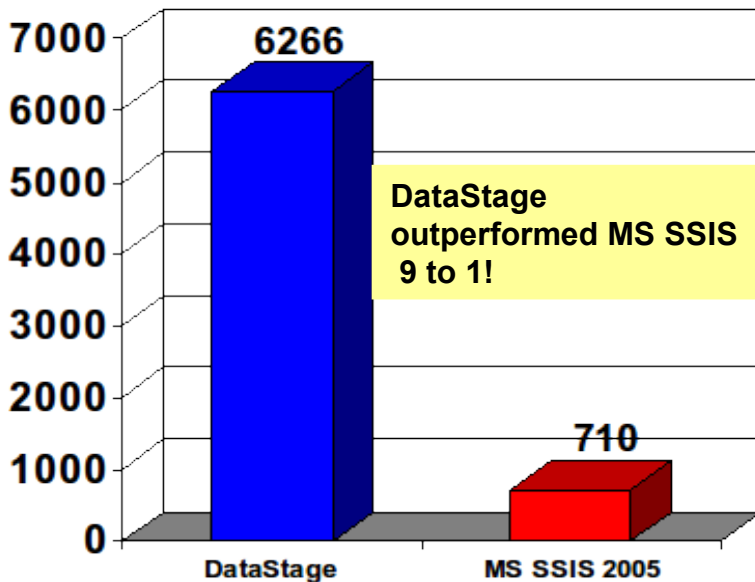**MGM&MIRAGE**

**PLAYERS CLUB**

## Benefits

- Achieved self-sufficiency in managing their EDW, eliminating reliance on 3rd party development and reducing maintenance costs.

- Improved customer service and gained ability to profile and promote its most profitable and active customers.

- Built a foundation that enables on-demand integration environment, with an up-to-date and authoritative customer database.

**Unlock The Business Value Of Information**

IBM

# DataStage Outperforms Microsoft SQL Server 2005 Integration Services

DB2 ← InfoSphere DataStage ← SQL Server

Microsoft
**SQL Server** 2005
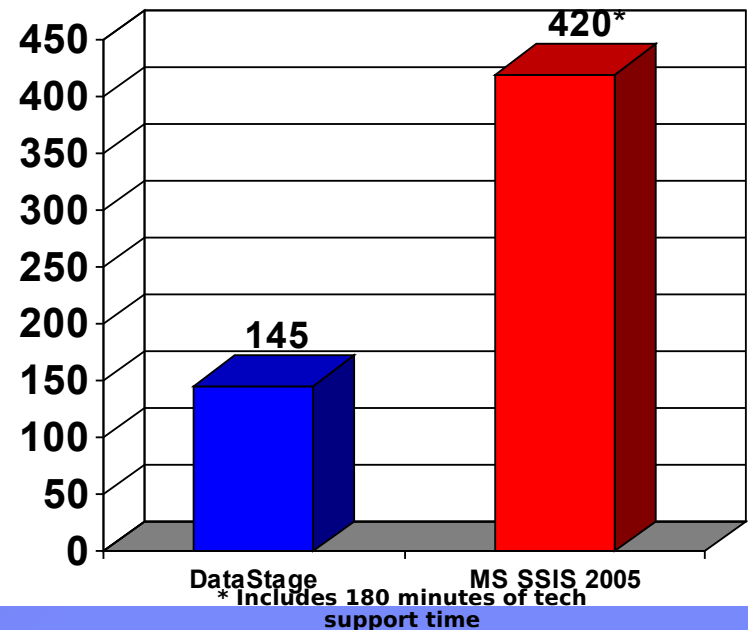**Integration Services**
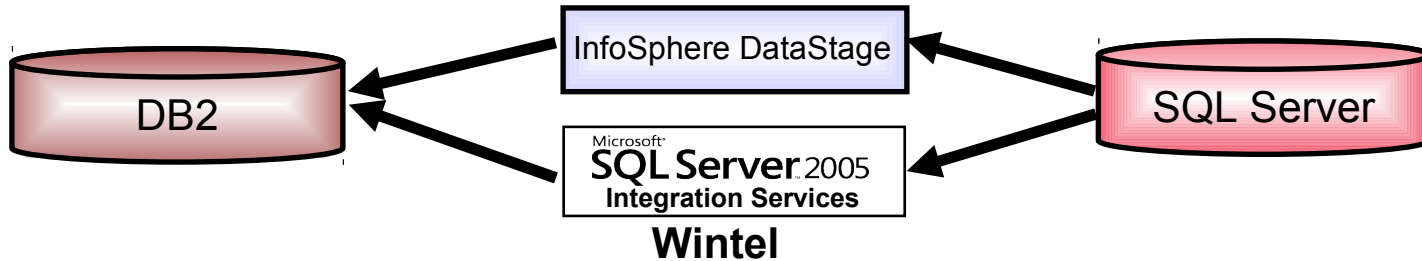
**Wintel**

**Results of a comparative study testing a simple ETL scenario involving database to database operations using Intel based servers**

## ETL Performance
**Records/Sec (bigger is better)**

| Value | Bar |
|-------|-----|
| 6266 | DataStage |
| 710 | MS SSIS 2005 |

7000 — 6000 — 5000 — 4000 — 3000 — 2000 — 1000 — 0

**DataStage outperformed MS SSIS 9 to 1!**

## Development Time
**Minutes (smaller is better)**

| Value | Bar |
|-------|-----|
| 145 | DataStage |
| 420* | MS SSIS 2005 |

450 — 400 — 350 — 300 — 250 — 200 — 150 — 100 — 50 — 0

*** Includes 180 minutes of tech support time**

# InfoSphere Information Server Overcomes Traditional Obstacles To Data Integration

**The success of this project depends on users trusting the information.  How can we be certain that we have the correct information?**

**IBM can provide an integrated information platform to address data quality *and* data movement to your warehouse.**

**Service Oriented Finance CIO**

**IBM**

# All Businesses Have Technical Issues That Inhibit Trusted Data

- **Multiple Data Sources**
  - ▸ Impact: Too much information coming from multiple systems, difficulty utilizing data

- **Different Data Formats**
  - ▸ Impact: Multiple versions of the truth

- **Understanding data and enabling reuse throughout different tools**
  - ▸ Impact: Lack of business agility, inability to take advantage of opportunities

- **Small time windows require high performance processing**
  - ▸ Impact: Incomplete, out-of-date, misinterpreted data

- **Service enable the warehouse to update frequently and access from variety of applications**
  - ▸ Impact: Limited ways for business to initiate data transfer

# A Consolidated Platform To Integrate Data

# InfoSphere Information Server Supports Metadata

Provides a common metadata repository for InfoSphere Information Server components, allowing tools to actively share information
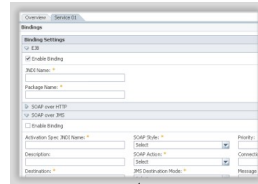
**InfoSphere Information Analyzer**
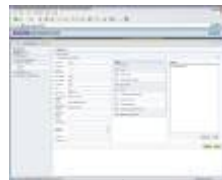
**InfoSphere Business Glossary**

**InfoSphere FastTrack**

**InfoSphere DataStage**

**InfoSphere QualityStage**

**InfoSphere** Metadata Services

**metadata**

**Design Operational**

- **Data Profiling**
- **SOA deployment**
- **Logging/Reporting**

- **Data Qual**
- **ETL, Data Transformation**
- **Admin and Security**

- **Impact analysis**
- **Da** **eage**

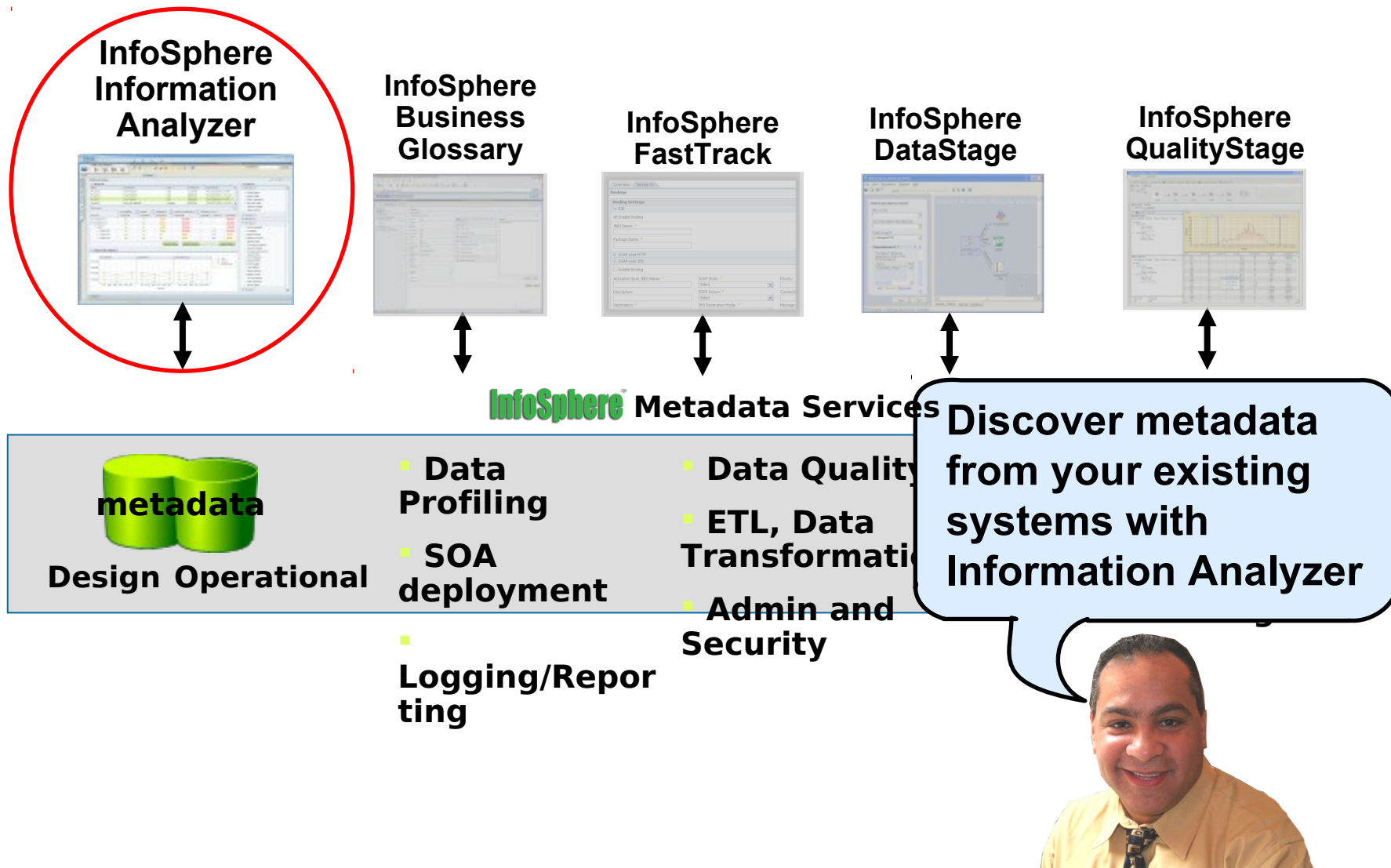Let's start by learning about metadata

# Introduction To Metadata

- Metadata is **information about data**

- Metadata facilitates the use and management of data
  - ▸ Enables intelligent code generation tools

- Examples of metadata:
  - ▸ A library has data in books, and metadata in a "card catalog"
  - ▸ A web search engine maintains metadata about the location and information contained in the sites it has indexed
  - ▸ Definitions of database tables are metadata

- InfoSphere Information Server shares metadata across products

**InfoSphere Metadata Services**

**metadata**

**Design  Operational**

- **Data Profiling**
- **SOA deployment**
- **Data Quality**
- **ETL, Data Transformation**
- **Admin and Security**
- **Metadata management**
- **Impact analysis**
- **Data lineage**

Logging/Repor
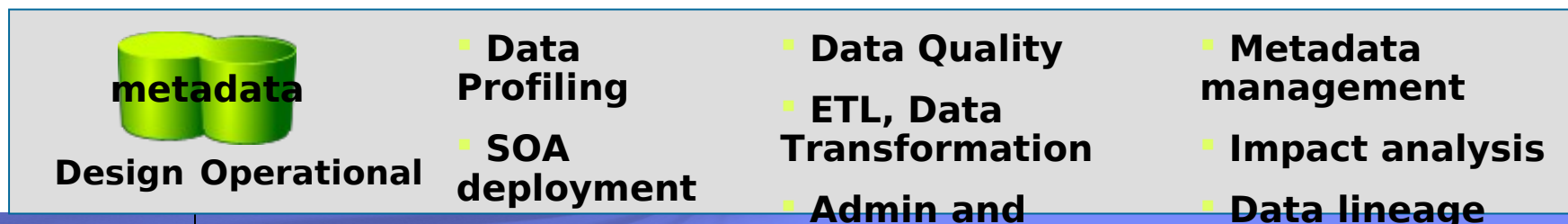
# InfoSphere Information Analyzer Discovers Metadata

**InfoSphere Information Analyzer**

**InfoSphere Business Glossary**

**InfoSphere FastTrack**

**InfoSphere DataStage**

**InfoSphere QualityStage**

**InfoSphere** **Metadata Services**

**metadata**

**Design  Operational**

- **Data Profiling**
- **SOA deployment**
- **Logging/Reporting**
- **Data Quality**
- **ETL, Data Transformation**
- **Admin and Security**

**Discover metadata from your existing systems with Information Analyzer**

# What Can InfoSphere Information Analyzer Do For You?

1. **What information do I already have?**

2. **Where is my information?**

3. **How do I get it when I need it?**

4. **What does it mean?**

5. **Can I trust it?**

6. **How do I get it in the form I need?**

7. **How do I get it where it needs to go?**

# InfoSphere Information Analyzer
## Discover, Analyze, And Store Technical Metadata

- **Discover and Understand Your Data**
  - ▸ Analysis of database and file-based sources

- **Data Profiling and Content Analysis**
  - ▸ Provides column, primary key, foreign key, and cross-domain analysis
  - ▸ Analyzes every data attribute and reverse engineers the true metadata of your source
  - ▸ Automates your data discovery process
  - ▸ Reduces time to analyze data by 70%

- **Creates and stores metadata describing information sources**
  - ▸ Identifies issues with structure and content of data sources
  - ▸ Metadata results accessible across InfoSphere Information Server components

**Unlock The Business Value Of Information**

# InfoSphere QualityStage Uses Metadata To Create Trusted Data
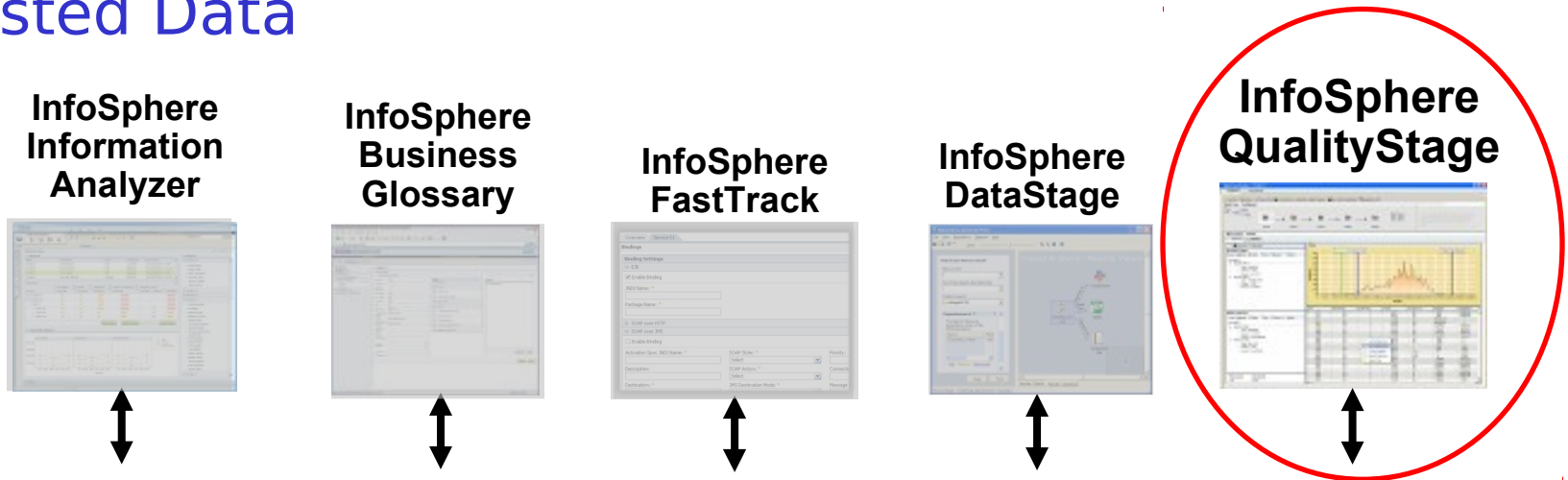
**InfoSphere Information Analyzer**

**InfoSphere Business Glossary**

**InfoSphere FastTrack**

**InfoSphere DataStage**

**InfoSphere QualityStage**

**InfoSphere Metadata Services**

**metadata**

**Design Operational**

- **Data Profiling**
- **SOA deployment**
- **Logging/Reporting**

- **Data Quality Metadata**
- **ETL, Data Transformation**
- **Admin and Security**

**Cleanse your data with QualityStage**

# Why Should I Care About Cleansing Information?



Low data quality costs companies $611 billion annually

25% of time is spent clarifying bad data

Scrap and rework Increased costs



Undetected defects will cost 10 to 100 times as much to fix upstream

Lack of consumer confidence

Lost opportunities

# Examples Of Data Quality Problems

- **Lack of data standards**
  - ▶ Different formats and structures across different systems

| Kate A. Roberts | 416 Columbus Ave #2, Boston, Mass 02116 |
| Catherine Roberts Four sixteen Columbus APT2, Boston, MA 02116 |
| Mrs. K. Roberts | 416 Columbus Suite #2, Suffolk County 02116 |

- **Data surprises in individual fields**
  - ▶ Data misplaced in the database

| Name | Tax ID | Telephone |
|---|---|---|
| J Smith DBA Lime Cons. | 228-02-1975 | 6173380300 |
| Williams & Co. C/O Bill | 025-37-1888 | 415-392-2000 |
| 1st Natl Provident | 34-2671434 | 3380321 |
| HP 15 State St. | 508-466-1200 | Orlando |

- **Information buried in free-form fields**

```
WING ASSY DRILL 4 HOLE USE 5J868A HEXBOLT 1/4 INCH

WING ASSEMBY, USE 5J868-A HEX BOLT .25" - DRILL FOUR HOLES

USE 4 5J868A BOLTS (HEX .25) - DRILL HOLES FOR EA ON WING ASSEM

RUDER, TAP 6 WHOLES, SECURE W/KL2301 RIVETS (10 CM)
```

- **Data myopia**
  - ▶ Lack of consistent identifiers inhibit a single view

| 19-84-103 | RS232 Cable 6' M-F CandS |
| CS-89641 | 6 ft. Cable Male-F, RS232 #87951 |
| C&SUCH6 | Male/Female 25 PIN 6 Foot Cable |

- **The redundancy nightmare**
  - ▶ Duplicate records with a lack of standards

| 90328574 | IBM | 187 N.Pk. Str. Salem NH 01456 |
| 90328575 | I.B.M. Inc. | 187 N.Pk. St. Salem NH 01456 |
| 90238495 | Int. Bus. Machines | 187 No. Park St Salem NH 04156 |
| 90233479 | International Bus. M. | 187 Park Ave Salem NH 04156 |
| 90233489 | Inter-Nation Consults | 15 Main Street Andover MA 02341 |
| 90345672 | I.B. Manufacturing | Park Blvd. Bostno MA 04106 |

# IBM InfoSphere QualityStage

- ## Ensures clean data by fixing data quality and consistency

  - ▶ Eliminates duplications
    - ▪ Matches against reference data

  - ▶ Provides visual tools for designing quality rules and matching logic
    - ▪ Integrated with DataStage (one platform, one user interface)

  - ▶ Deploy data quality processes
    - ▪ Within extract, transform, and load (ETL) jobs
    - ▪ As data quality web services (SOA)

<p style="text-align:center; color:red;">Microsoft only provides limited fuzzy search<br>Oracle relies on third parties (Trillium)</p>

# InfoSphere QualityStage Accurately Removes Duplicate Records

## Are these two records a match?

```
WILLIAM J      HOLDEN  128 MAIN  ST  02111
12/8/62        WILLAIM JOHN  HOLDEN  128 MAINE
AVE 02110 12/8/62
   B     B        A       A     B    D    B      A    =
```

```
 +5    +2    +20    +3    +4    -1    +7    +9    =   +49
```

***The Competition:*** **Deterministic Decisions Tables:**
- Fields are compared
- Letter grade assigned
- Combined letter grades are compared to a vendor delivered file
- Result: Match; Fail; Suspect

***InfoSphere QualityStage:*** **Probabilistic Record Linkage:**
- Fields are evaluated for degree-of-match
- Weight assigned: represents the **"information content"** by value
- Weights are summed to derive a total score
- Result: Statistical probability of a match

**Better Accuracy**

# Dynamically Cleanse Your Data While Loading Your Data Warehouse with QualityStage



**Cleanse**

**Discard Duplicates**

**Match/Compare against Reference Data**

**InfoSphere QualityStage**

**Data Warehouse**

**Load**

**InfoSphere DataStage**

**Transform**

**Extract**

Savings

Credit Cards

Checking

Mortgages

Loans

Operational Data Source
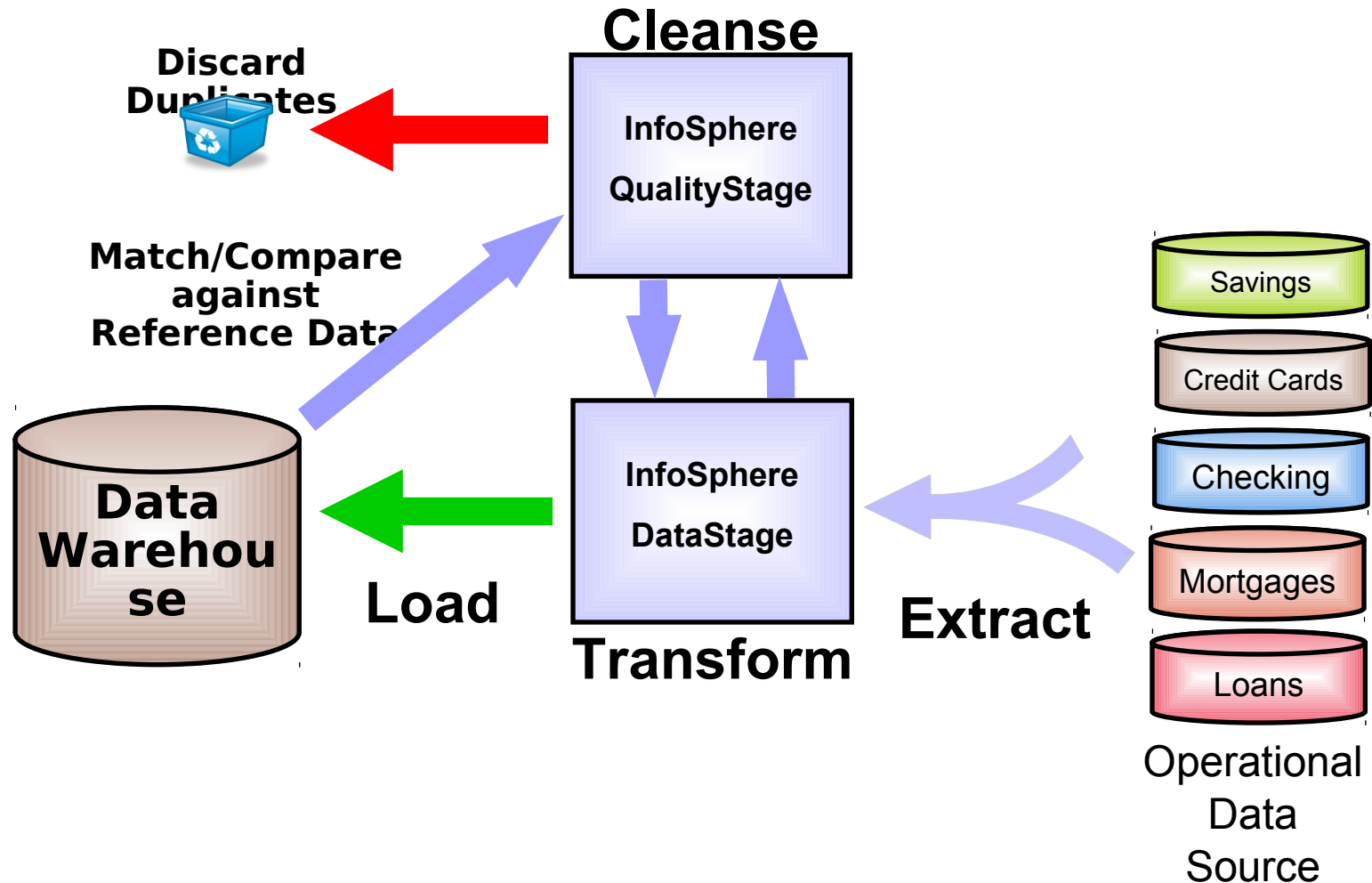
# Service Oriented Finance Wants A Dynamic Data Warehouse

The information our decision makers use needs to be more current.

IBM InfoSphere Change Data Capture together with InfoSphere Information Server can "trickle feed" data into your warehouse as it changes.

**Service Oriented Finance CIO**

**IBM**

# InfoSphere Change Data Capture and Information Server Enable Dynamic Data Warehousing

Data Warehouse  ←  **InfoSphere DataStage QualityStage**  ←  **InfoSphere Change Data Capture**  ←  Credit Cards / Checking / Mortgages

**Load**　　　**Integrate**　　　**Data Changes**
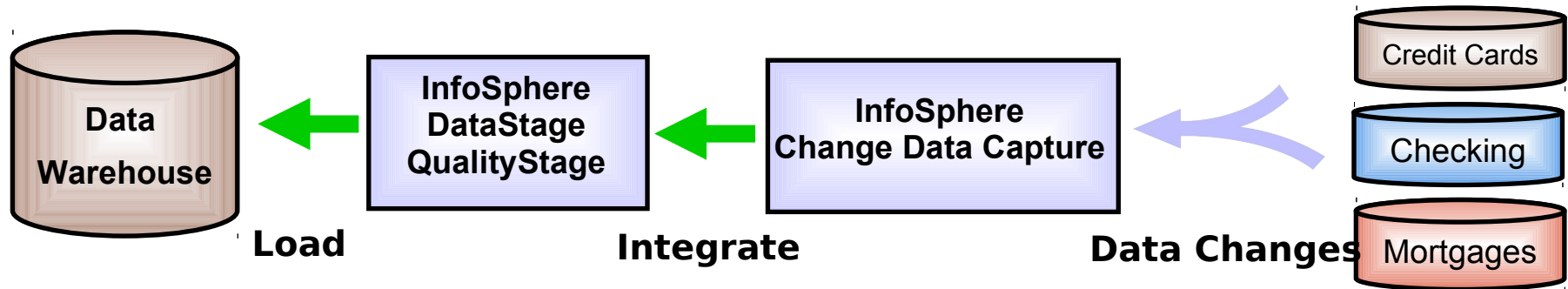
- Collects changes made to source databases
  - Monitors source database log files for data changes
  - Data changes can be sent to InfoSphere DataStage and QualityStage via **direct integration**
    - Data changes can also be sent directly to target databases
    - Data changes can also be delivered to message queues which can trigger events in an ESB

- Automates the process of loading the data warehouse
  - Improves the currency of data in the data warehouse
  - Alternative to statically scheduled updates  - which can lead to stale data and lost opportunities

- Complements InfoSphere DataStage and InfoSphere QualityStage

# InfoSphere Change Data Capture Is Efficient

- **Uses native database recovery logs to capture activity in databases**
  - ▸ **Most databases have a native log format that can be accessed**
    - ▪ DB2 and DB2 z/OS = DB2 Log, DB2(i5)=OS/400 Journal
    - ▪ Oracle = Re-Do Log
    - ▪ SQL Server = Transaction Log

- **No changes to existing applications or schemas required**

- **Very little performance impact to source application and system**
  - ▸ **0.05%** system resources required to process over 300 GB of data
  - ▸ Highly scalable and efficient - sends only the changes in the data
  - ▸ Avoids performance and other drawbacks of alternatives:
    - ▪ SQL Select, File Comparison, Database Triggers, Modifying Application

- **Wide range of RDBMS support**
  - ▸ DB2 (all platforms), Oracle, SQL Server, Sybase, and more

# IBM Leads In Data Integration

- Only InfoSphere Information Server delivers  unified metadata across all tools for collaboration and reuse
  - Oracle has no integration of metadata across products – manual  import/export required
- Model-Driven Design, with FastTrack and DataStage, speeds development
  - Neither Microsoft nor Oracle have any tool to help manage source to target mappings
- Integration built on open standards works in heterogeneous data environments
  - InfoSphere Information Server connects to, processes and cleanses more data from more sources than Oracle and Microsoft
- InfoSphere provides data analysis and lineage
  - Microsoft only has "fuzzy search" Oracle Data Integrator (Sunopsis) uses a 3rd party product (Trillium) for data quality and profiling

# Use IBM Cognos 8 BI To Make Better Business Decisions

**Cognos 8 BI**

**Reports and Dashboards**

**InfoSphere Warehouse**

**Data Warehouse**

Collect Historical Data

**InfoSphere Information Server**

**Extract Transform Load**

Savings

Credit Cards

Checking

Mortgages

Loans

**Operational Data Source**

Business Intelligence

ETL Processing

Scalable Database

Connections
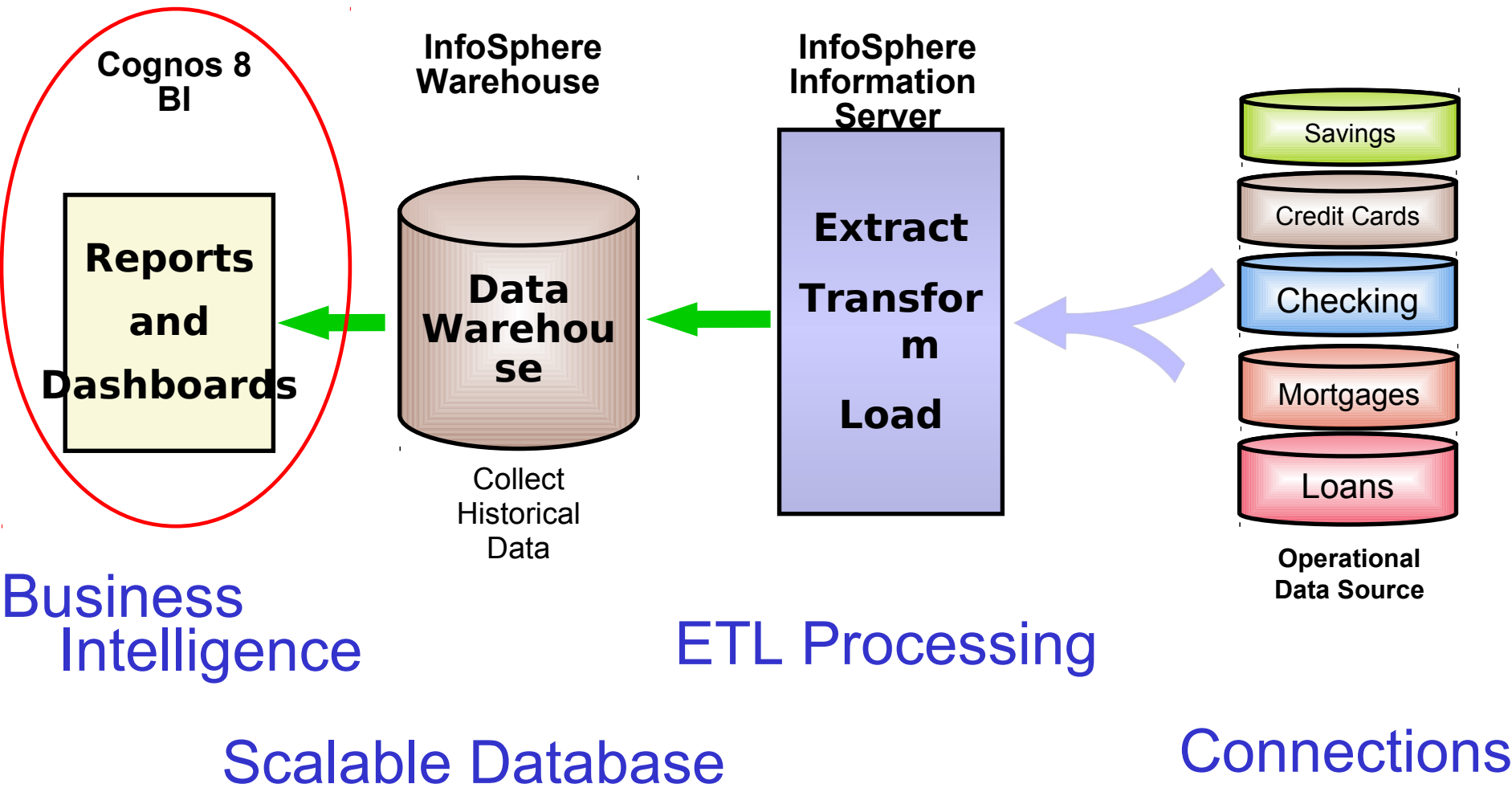
# Service Oriented Finance Needs New Insight Into Their Business

We need to improve decision making across our entire organization.

Cognos 8 BI provides easy to learn tools so that users can create the reports they need for themselves using only a web browser.

**Mortgage Line of Business VP**

**IBM**

# IBM Cognos Is An Integrated Platform Built On SOA

- Implemented in Java, runs on WebSphere

- 100% browser based access
  - ▶ Server side business intelligence
  - ▶ Users can access new intelligence from anywhere

- Easiest for IT to deploy and manage
  - ▶ Scales up and out across heterogeneous hardware and operating systems
  - ▶ Unified security
  - ▶ Unified administration

- Consistent user interface across tooling
  - ▶ Greater user satisfaction and increased business agility with lower IT costs

- Common metadata model
  - ▶ Author new intelligence assets once, consume anywhere
  - ▶ Common view enables open data strategy
  - ▶ Supports Unicode and multilingual features without recreating reports

**IBM Cognos**

User

Cognos Server

| Query | Report | OLAP |

Metadata Model

Data
(Multiple sources)

IBM Cognos

**Unlock The Business Value Of Information**
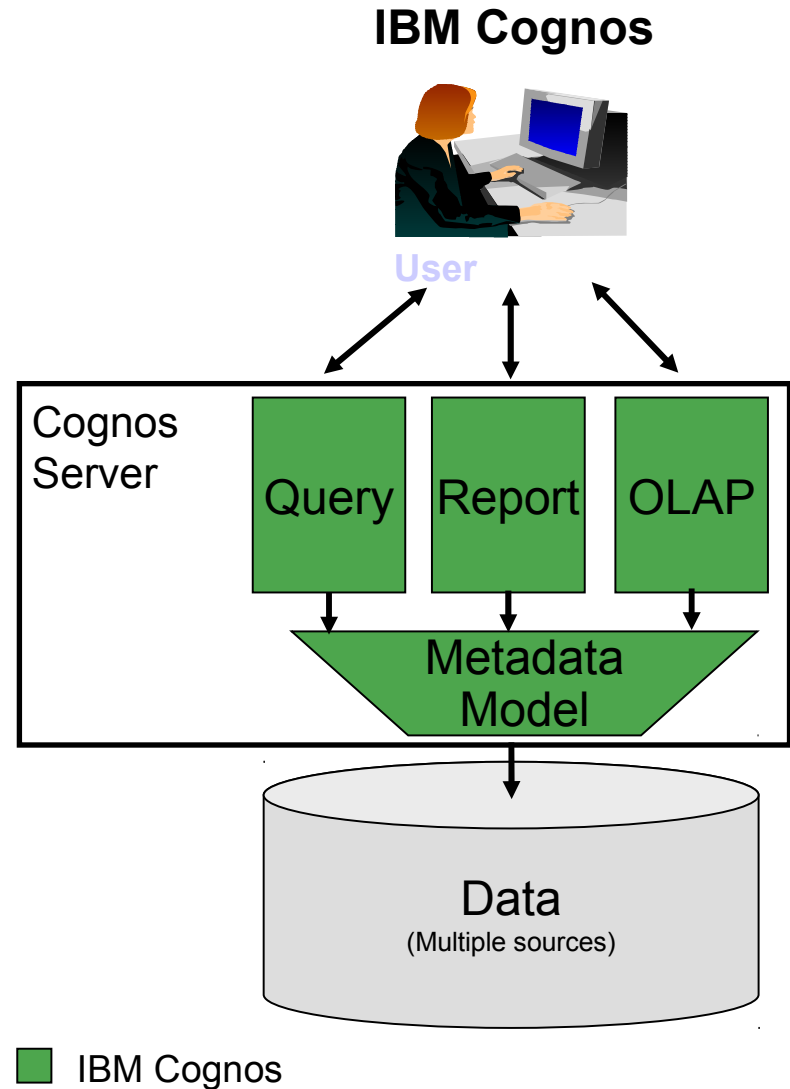
# Reuse Trusted New Intelligence Assets Across The Cognos 8 Platform
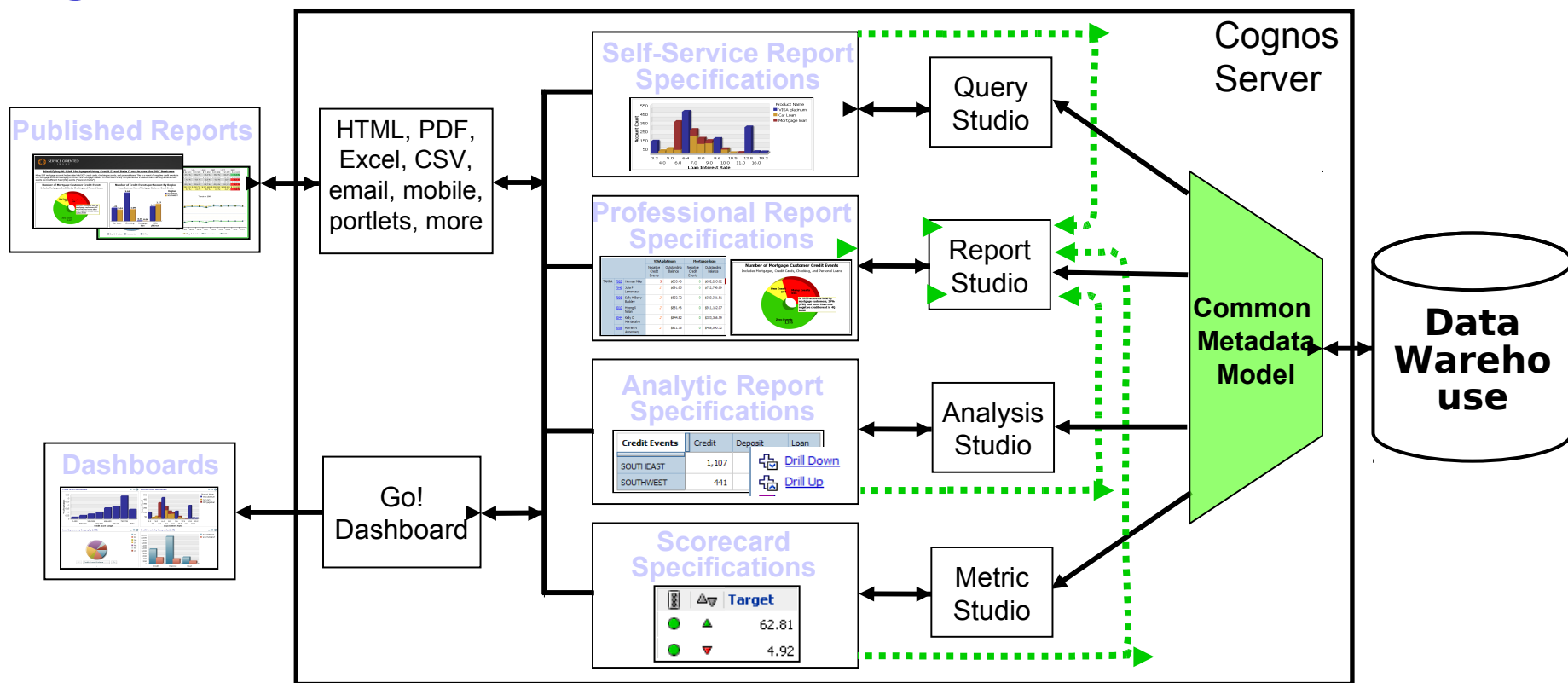


- **All new intelligence assets share a common metadata model and common report specification**
- **Author Once – Consume Anywhere**
- **Ensures consistent information and enables reuse across platform functions**

# Include Any Data Source In Your New Intelligence With The Cognos 8 Platform
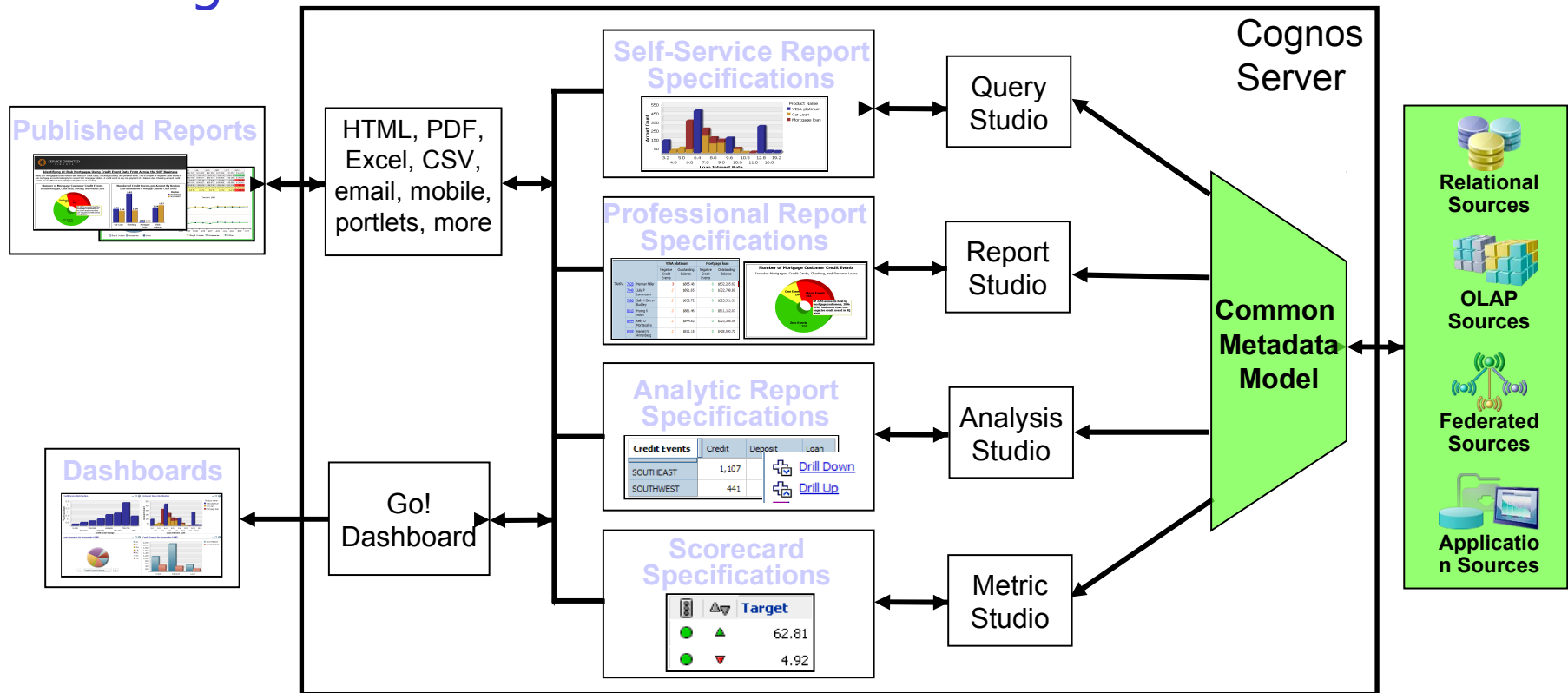


- **Open data strategy enables a common view across a variety of data sources**

- **Support for application data sources such as SAP ERP**

- **Combine relational, OLAP, federated, and other data sources in any tool**

- **All capabilities access a trusted set of information defined in the common metadata model**

- **As sources change, metadata model can control and identify impacts to report specifications**

# Cognos Go! Dashboard Enables You To Monitor Business Operations

- Cognos report specifications can be incorporated into dashboards using Cognos 8 Go! Dashboards

- What goes into a dashboard?
  - ▸ Self service reports
  - ▸ Professional reports
  - ▸ Analytical reports
  - ▸ Scorecards
  - ▸ RSS feeds, HTML, search, more

- Users can create their own dashboards from existing Cognos report assets

- Everything you need to monitor multiple areas of the business at the same time

- Provides dynamic and visually appealing capabilities by using Adobe Flash
  - ▸ Everything runs in a browser – only Adobe Flash is required
  - ▸ Easily change chart types and color palettes
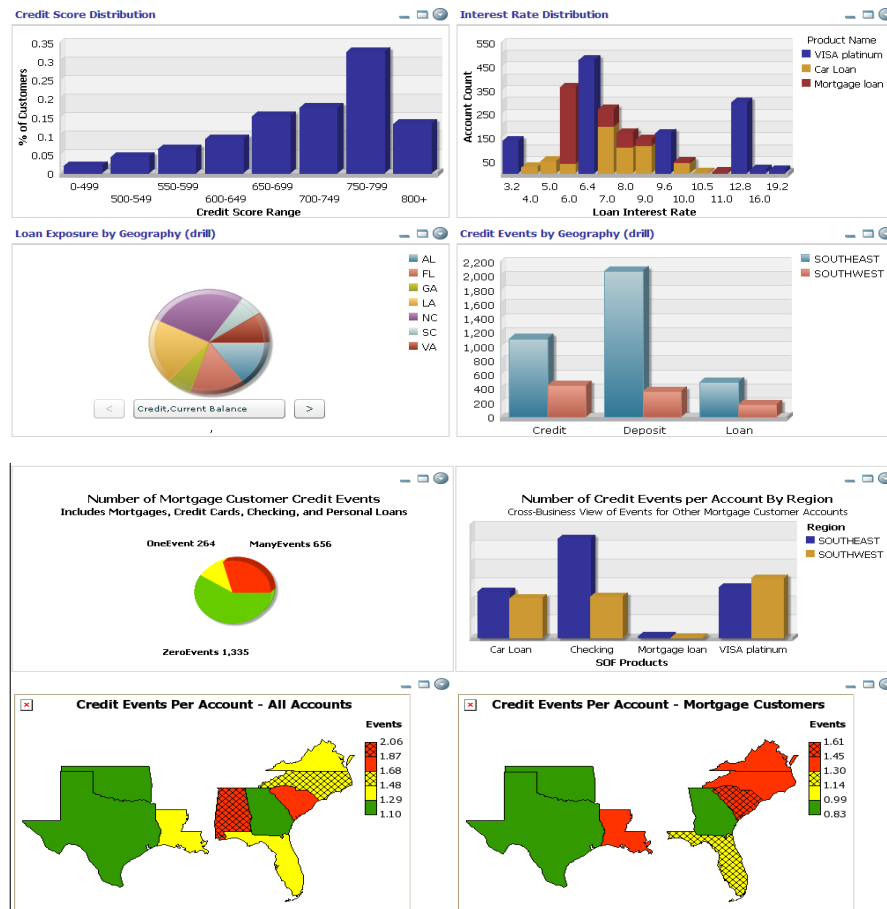
# Different Employee Roles Need Access To Different Information And Capabilities

**Deliver**

**Analyze**

**Integrate**
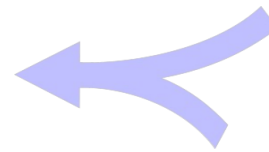
**Capture**



Reports

OLAP

Self Service

Data

Data

Data

Data

Data
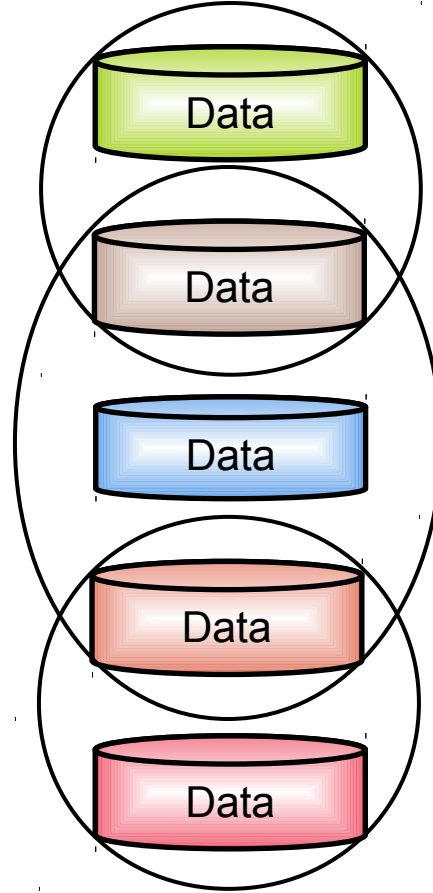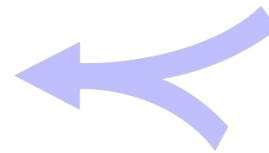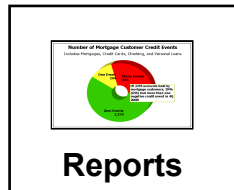
Task 1

Task 2

Task 3
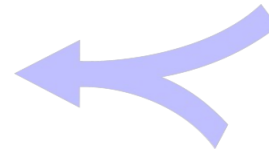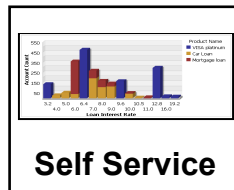
# Novice Users Can Create The Reports They Need Using Cognos 8 BI Query Studio

- **Easy to learn self-service reporting requiring minimal reporting knowledge**
  - ▸ Helps alleviate report authoring backlog

- **Query Studio allows users to create, modify, save, and publish self-service report specifications**
  - ▸ Use existing self-service reports to create a new reports
  - ▸ Modify the style and layout of self-service reports
  - ▸ Add or remove data using the Cognos common metadata model
  - ▸ Create filters, summaries, and calculations
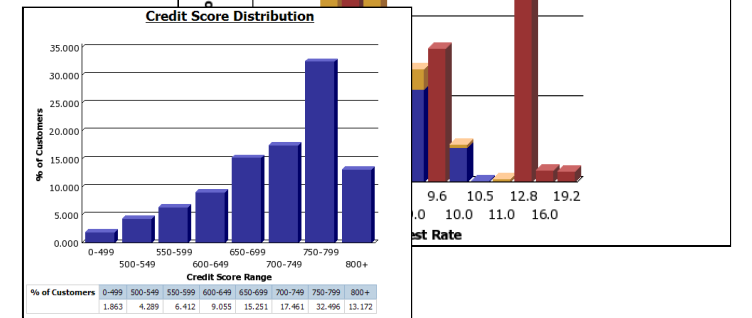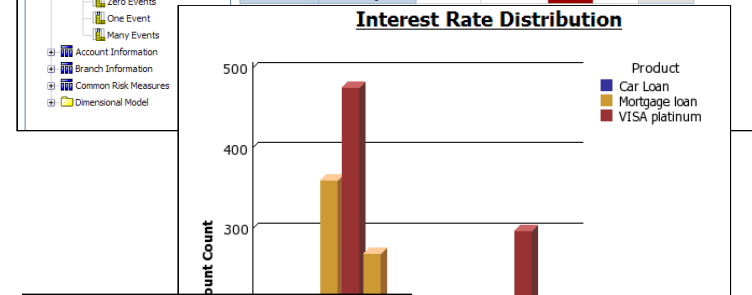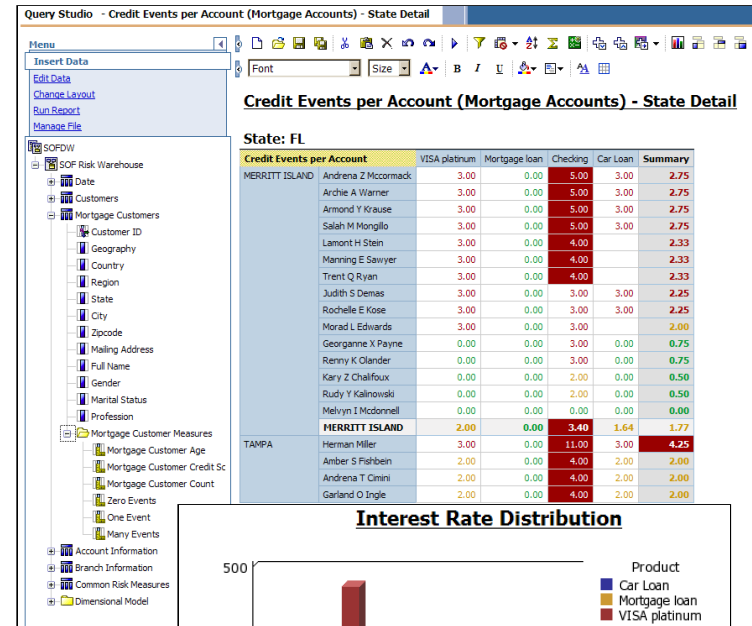
# Cognos 8 BI Report Studio Creates Professional Report Specifications To Run Your Business

- Advanced reporting tool to create any style of report like invoices, financial statements, inventory, payroll, etc.

- Designed to provide sophisticated capabilities and advanced features required by professional report authors

  - Provides "pixel-perfect" formatting with absolute control over visual layout and formatting

  - Large library of lists, crosstabs, charts, maps, operators, constants, functions, filters, more

  - Prompts add interactivity allowing users to customize reports

  - Bursting delivers reports with only the data relevant to the recipient

  - Express Authoring Mode is designed specifically for financial report authors

# Use Cognos 8 BI Analysis Studio To Gain Insight Into Your Business

- **Cognos Analysis Studio provides multidimensional OLAP analysis**

- **Use Analysis Studio to interactively explore data to find answers to business questions**

  ▸ Use drill up or drill down to "slice and dice" the context and view of data

  ▸ Identify trends, compare actual results to plans, or highlight the best or worst performers

  ▸ Create calculations using many built-in operators such as growth, rank, average, percentile
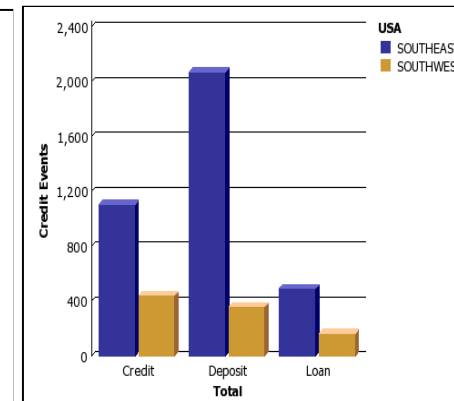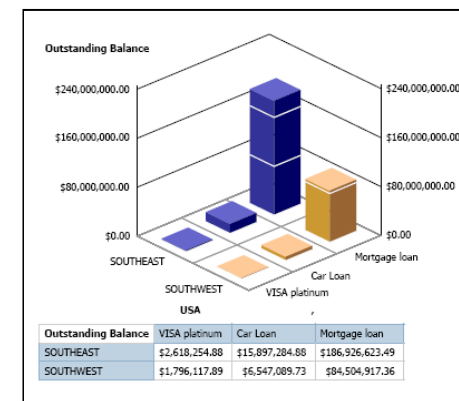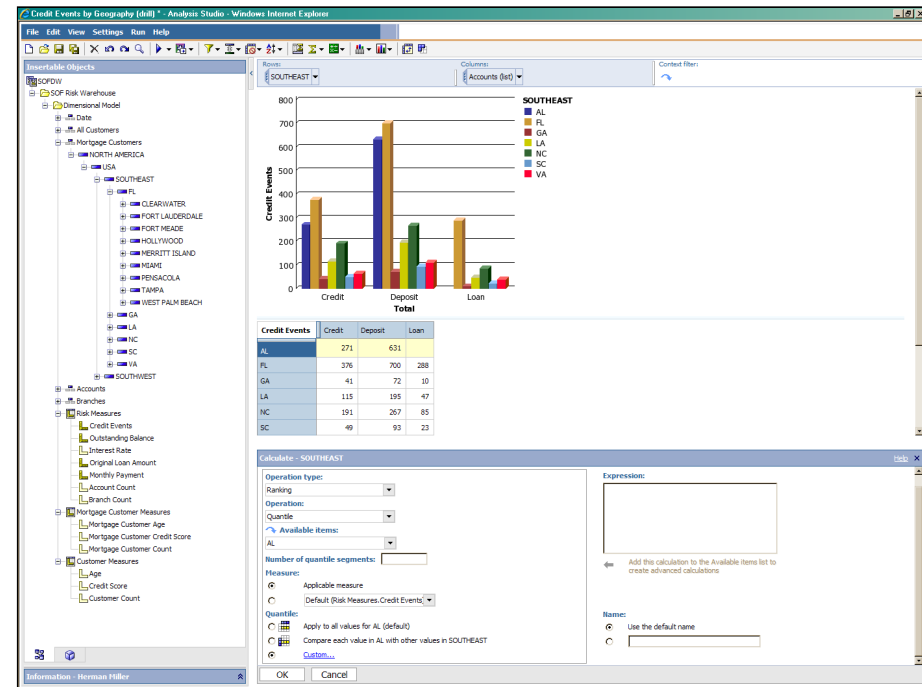
# Create Scorecards With Cognos 8 Metric Studio To Track Key Business Metrics

- **Cognos Metric Studio tracks progress toward business objectives**



- **Creates scorecard specifications using the Cognos common report format**
  - ▶ Scorecard specifications can be included within professional report specifications using Cognos 8 BI Report Studio

- **Metrics display results relative to targets**
  - ▶ Defined on any indicator of business performance
  - ▶ Actual results imported on a scheduled basis
  - ▶ Scorecards organize metrics logically
  - ▶ Diagrams organize metrics visually
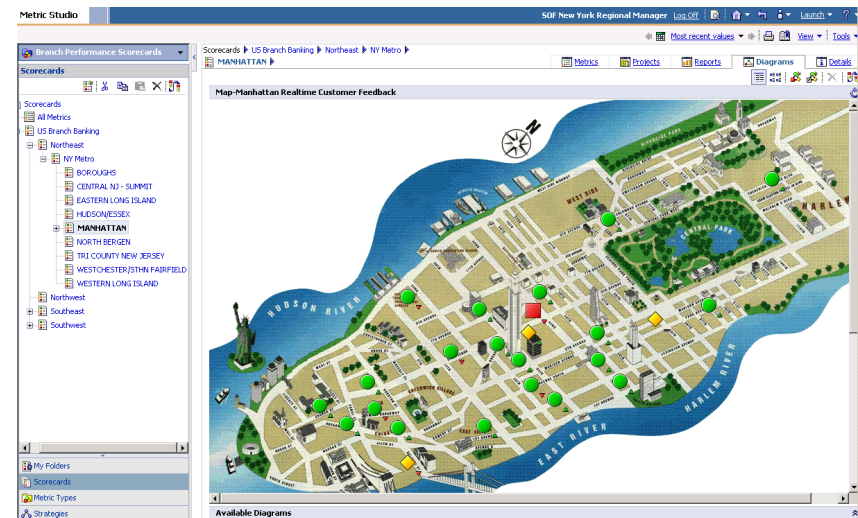
# All These Functions Access A Common Metadata Model And Create Assets In A Common Format

- All run in a browser, no need to install desktop applications

- All use trusted information from the Cognos common metadata model
  - Includes relational, OLAP, federated, application, and many other data sources

- All create report specifications using the Cognos common report format
  - Any report specification can be opened and extended into a professional report specifications using Cognos 8 BI Report Studio

- All report specifications share a single viewer and can be viewed in a browser or delivered in many formats and modes
  - Formats:  HTML, PDF, Excel, XML, CSV
  - Modes:  print, email, scheduled, browser, mobile, search, Microsoft Office, search

- All allow users to configure watch items, alerts, and watch rules to automatically receive updated reports via email or browser

- All Cognos report specifications can be incorporated into dashboards using Cognos 8 Go! Dashboard

**Unlock The Business Value Of Information**

# Service Oriented Finance Needs To Do More With Less

**This needs to be simple to deploy.**

**Cognos is an integrated platform that fits into your existing environment.**

**The competition is not so simple.**

**Service Oriented Finance CIO**

**IBM**

# Cognos 8 BI Fits Into Your Existing Environment

## Interoperable and Extensible

| | |
|---|---|
| *Standards* | SOAP, XML, WSDL, HTTP(S), OMG CWM, XML/XMI, REST |
| *Extend* | Open, Web Services API to programming language of choice including Java, Microsoft .NET Framework, COM & Perl |
| *Embed* | Industry Standard Portals including IBM, SAP, Plumtree, WSRP) |
| | Enterprise Search, BPM, Office, 3ʳᵈ party Applications |
| *Security* | Any combination of LDAP, NTLM, Active Directory, Netegrity, SAP, custom |

## Environments

| | |
|---|---|
| *App Servers* | Any J2EE application server Apache Tomcat, IBM WebSphere, SAP NetWeaver, BEA WebLogic, Oracle Application Server, Sun, JBoss |
| *Operating Systems* | 32 or 64 bit scale-out with any combination of Linux, Unix, or Windows |
| *Platforms* | IBM Power, IBM System z, Intel, HP, Sun |
| *Routers/ Firewalls* | All standard gateways supported |

**Cognos.**
software

## Metadata Integration

| | |
|---|---|
| *Modeling tools* | CA ERwin, Oracle Designer, Business Objects Designer, Sybase PowerDesigner, CA COOL, Rational Rose, XML/XMI, OMG CWM |
| *ETL & EII* | Cognos Data Manager, Informatica PowerCenter, IBM Information Server (Ascential) Metabroker, Composite, IBM II |
| *Source catalogs* | Relational databases, ERP |

## Open Data Access

| | |
|---|---|
| *Relational* | IBM DB2 UDB, Oracle, Microsoft SQL Server, Teradata, other… |
| *Modern* | XML, WSDL, LDAP, JDBC, JDBC… |
| *ERP* | SAP, PeopleSoft, Oracle, Siebel, salesforce… |
| *OLAP* | PowerCubes: High Performance Dimensional Cache |
| | Industry Standard OLAP Providers: Microsoft, IBM, SAP, Essbase |

# IBM Lab Tests Show That Cognos 8 BI Delivers Exceptional Scalability And Economies Of Scale

Testing demonstrated that Cognos 8 BI delivers faster response time when hardware capacity is increased

**Horizontal Scalability**

(Bar chart: Total Average Transaction Time (s) vs BI Processing CPUs)

| BI Processing CPUs | Total Average Transaction Time (s) |
|---|---|
| 4 | ~2170 |
| 8 | ~790 |
| 16 | ~270 |
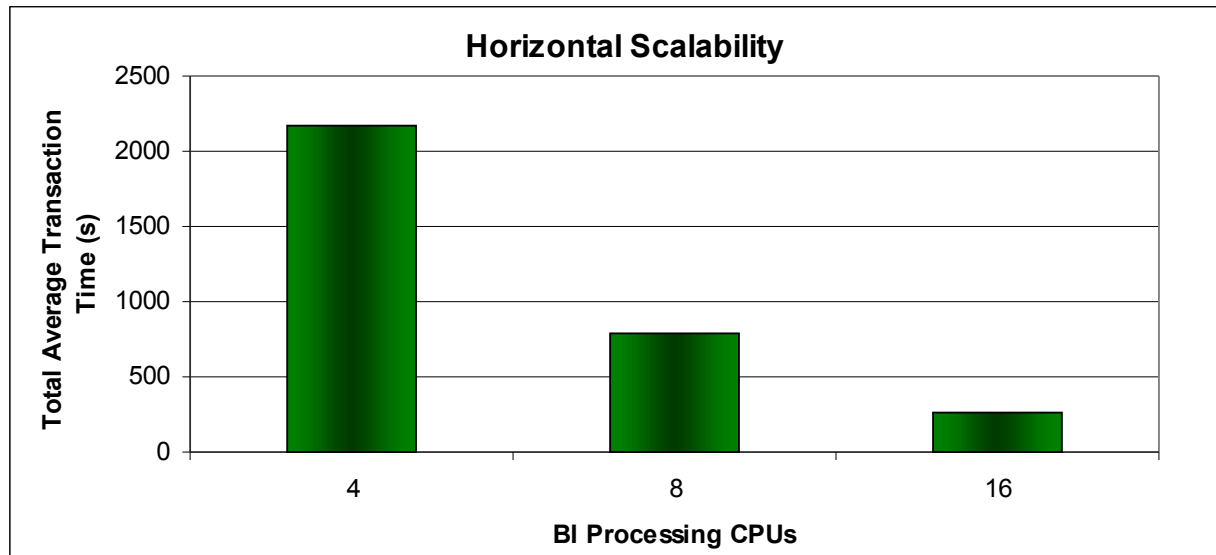
When hardware was doubled, response times were reduced by an average of **65%**

# IBM Cognos Is Also Integrated With InfoSphere Information Server And InfoSphere Warehouse

- Users can understand the meaning of business terms used in reports
  - ▶ Access InfoSphere Business Glossary from within Cognos tooling and reports

- Users can see the source of data used in Cognos reports
  - ▶ Access to InfoSphere Metadata Workbench from within Cognos tooling and reports

- Cognos easily connects to InfoSphere Warehouse Cubing Services
  - ▶ Integration allows simple setup and optimized performance when accessing cube data in the warehouse (ROLAP)

- InfoSphere Warehouse data mining and unstructured analytics results can be incorporated into Cognos reports

# Microsoft Has Separate Unintegrated Products And Desktop Applications

- **Multiple desktop tools**
  - ▸ Report Builder (v1, v2)
  - ▸ BI Development Studio (BIDS) Report Designer
  - ▸ Excel, Office

- **Fragmented metadata approach**
  - ▸ Direct query, Report Builder Models, BIDS datasets, MSAS cubes (data source views), …

- **Programming required**

- **Creates IT Burden**
  - ▸ Install, update, user support all happen at the desktop
  - ▸ Not SOA – Only works with Microsoft product stacks
  - ▸ Multiple security models

**Microsoft**

**User**

(Browser)

| Report Builder (Desktop) | BI Dev Studio (Desktop) | Excel (Deskt op) |

**Multiple Desktop Applications**

Sharepoint Portal | Reporting Services | Analysis Services

**Multiple Server Products**

Report Model Designer | Data Source Views

**Multiple View Designers**

SQL Server

■ MS Sharepoint  ☐ MS SQL Server  ■ MS Office
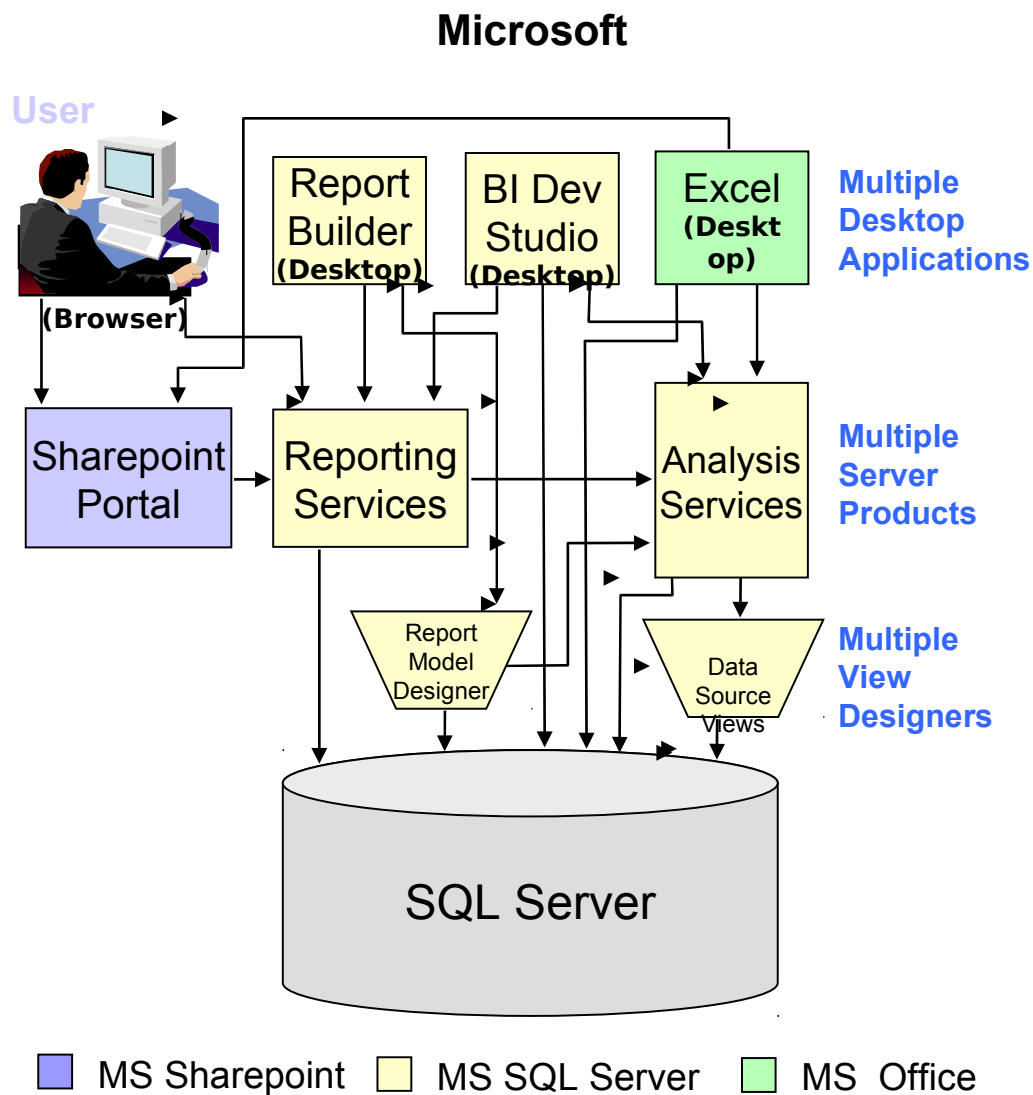
# Microsoft Lacks Common Metadata And Common Report Format So Tools Cannot Share Assets

- **Multiple Report Formats**
  - ▸ Report Definition Language (RDL) v2.0 not supported by all tools

- **Limited sharing of assets between Report Builder, BI Dev Studio Report Designer, Excel**

- **Version incompatibilities**
  - ▸ BI Dev Studio 2008 reports cannot be viewed using Visual Studio 2008 ReportViewer control. Waiting for updated ReportViewer in Visual Studio 2010…[1]

**Microsoft**

**User**

**(Browser)**

Report Builder **(Desktop)**

BI Dev Studio **(Desktop)**

Excel **(Desktop)**

**Multiple Desktop Applications**

Sharepoint Portal

Reporting Services

Analysis Services

**Multiple Server Products**

Report Model Designer

Data Source Views

**Multiple View Designers**

SQL Server
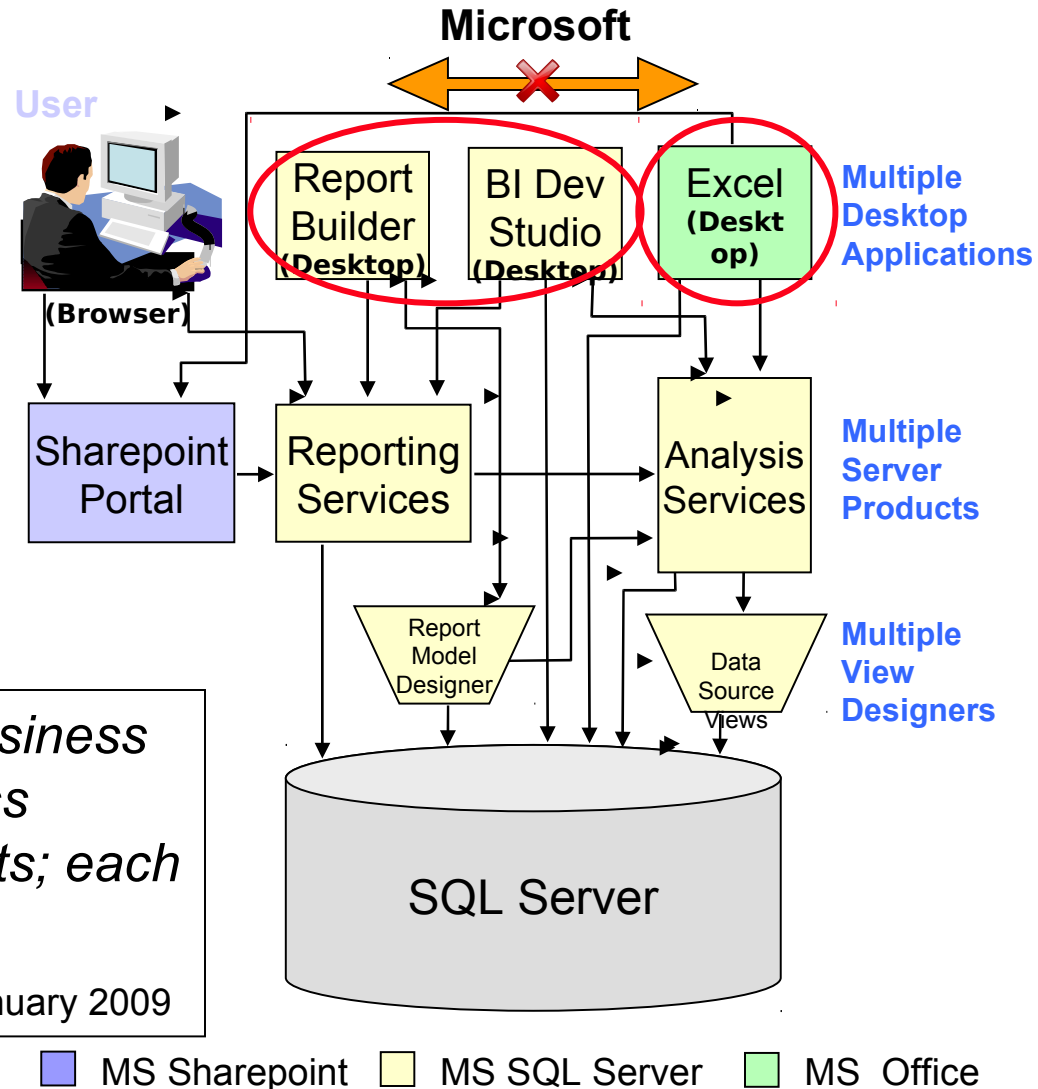
> "…*there is no single, enterprise business metadata layer or capability across Microsoft's BI platform components; each has its own metadata model,…*"
>
> -- Gartner Magic Quadrant For BI Platforms, January 2009

■ MS Sharepoint    ▢ MS SQL Server    ▢ MS Office

1 https://connect.microsoft.com/VisualStudio/feedback/ViewFeedback.aspx?FeedbackID=380314

http://www.gotreportviewer.com/

# Oracle Business Intelligence Enterprise Edition Is Complex And Is Not Integrated

- **Multiple Thick Client Tools**
  - ▸ Multiple report formats
  - ▸ Multiple metadata models
  - ▸ Creates more work

- **Too Many Products!**

- **Creates IT Burden**
  - ▸ Install, Patch, User Support all happen at the desktop

**Oracle BI Enterprise Edition (OBIEE)**



Legend: Oracle | Siebel | Hyperion | Brio
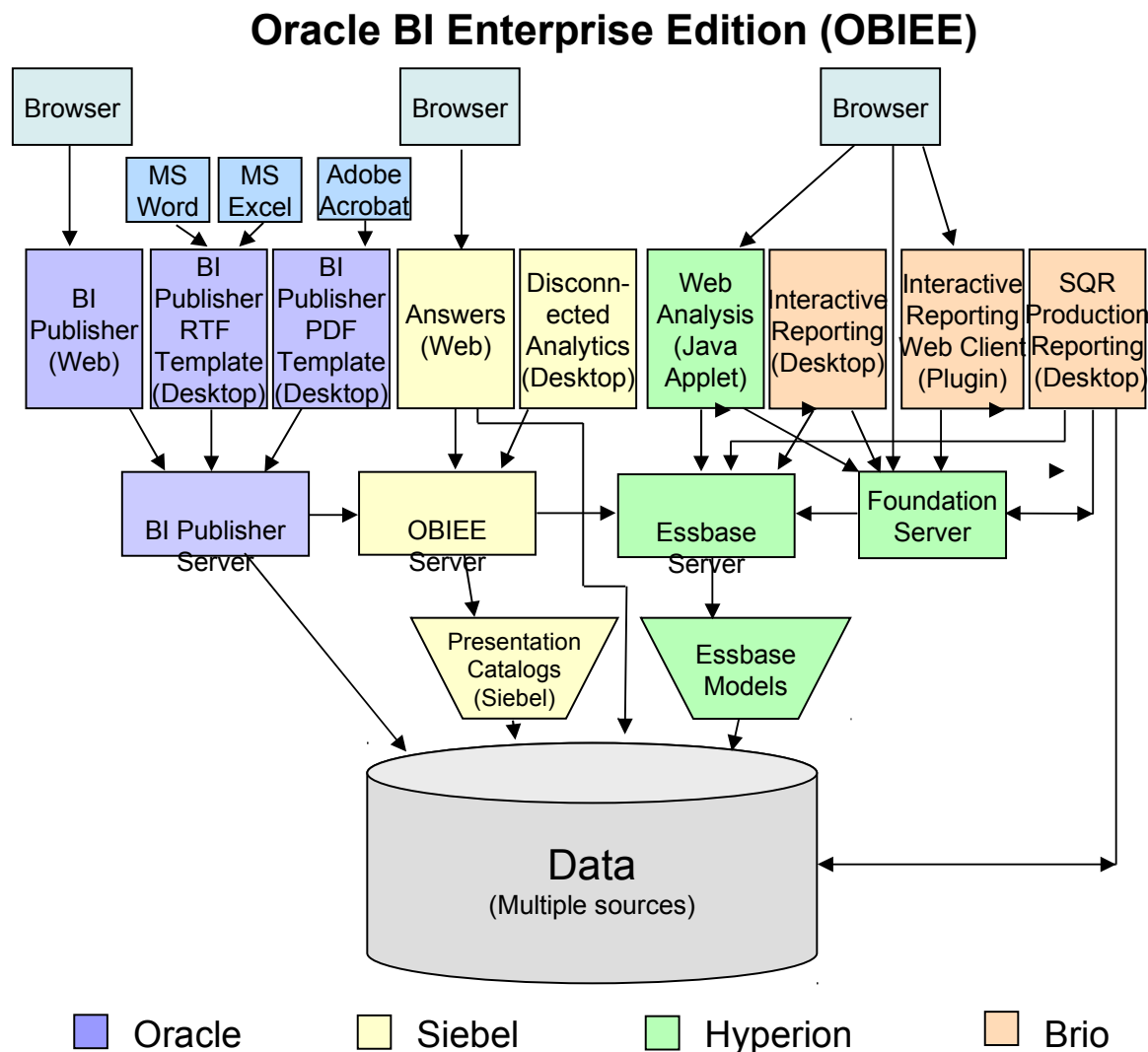
# Compare For Yourself: Cognos Is Simple, Oracle Is Complex

## IBM Cognos

**100% browser-based web access**

User

| Cognos Server | Query | Report | OLAP |
|---|---|---|---|

Metadata Model
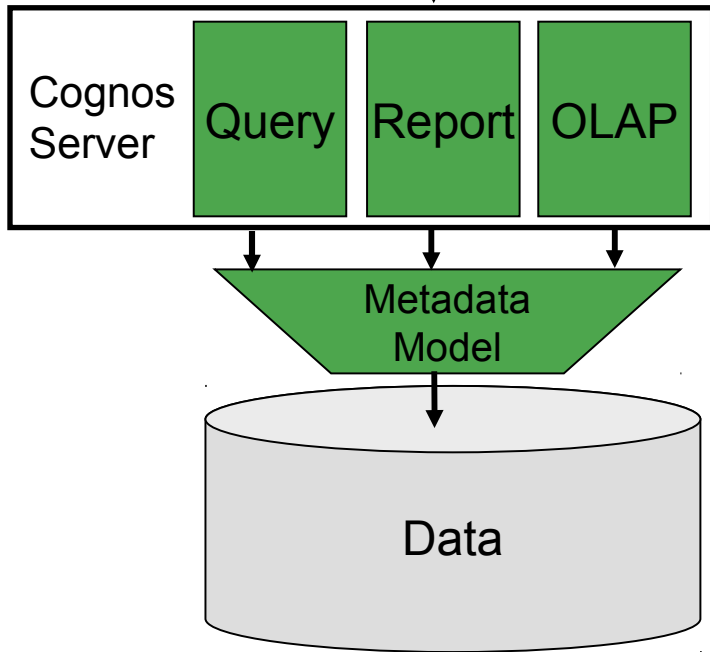
Data

■ IBM Cognos

## Oracle BI Enterprise Edition (OBIEE)

Browser

Browser

Browser

MS Word

MS Excel

Adobe Acrobat

BI Publisher (Web)

BI Publisher RTF Template (Desktop)

BI Publisher PDF Template (Desktop)

Answers (Web)

Disconn-ected Analytics (Desktop)

Web Analysis (Java Applet)

Interactive Reporting (Desktop)

Interactive Reporting Web Client (Plugin)

SQR Production Reporting (Desktop)

BI Publisher Server

OBIEE Server

Essbase Server

Foundation Server

Presentation Catalogs (Siebel)

Essbase Models

Data

■ Oracle    ■ Siebel    ■ Hyperion    ■ Brio

# Oracle Lacks Common Metadata And Common Report Format So Tools Cannot Share Assets

- Limited sharing of assets between BI Publisher, Answers, Web Analysis, Interactive Reporting, etc

- Cannot promote self-service (Answers) reports into BI Publisher

  ▸ Different technologies, different formats

- Multiple metadata strategies

- Complexity = More Work

**Oracle BI Enterprise**

Browser — MS Word, MS Excel, Adobe Acrobat
BI Publisher (Web), BI Publisher RTF Template (Desktop), BI Publisher PDF Template (Desktop), Answers (Web), Disconnected Analytics (Desktop), Web Analysis (Java Applet), Interactive Reporting (Desktop), Interactive Reporting Web Client (Plugin), SQR Production Reporting (Desktop)

BI Publisher Server, OBIEE Server, Essbase Server, Foundation Server

Presentation Catalogs (Siebel), Essbase Models

Database

| | Oracle | | Siebel | | Hyperion | | Brio |

# Deliver New Intelligence Quickly With The IBM Smart Analytics System

**How can we get started quickly?**

**The IBM Smart Analytics System provides a complete ready-to-deploy system for delivering New Intelligence**

**Service Oriented Finance CIO**

**IBM**

# The IBM Smart Analytics System Delivers New Intelligence

A complete, ready-to-deploy system including hardware, software, storage, and network

- System is installed, tuned, and ready for your data

- Choose from standard configurations, add capacity with hardware and software modules

- Single point of contact for support

**Analytics**
- ☑ Cognos 8 Business Intelligence
- ☑ Cubing Services
- ☑ Text Analytics & Data Mining

**Data Warehouse Software**
- ☑ InfoSphere Warehouse
- ☑ Advanced Workload Managemer
- ☑ Tivoli System Automation

**Hardware/OS**
- ☑ AIX 6.1
- ☑ IBM Power 550
- ☑ IBM System Storage DS5300

**Faster Time To Value**

- Deploy in days instead of months

- Focus on creating business value, not installing hardware and software

**Costs less than building it yourself**

- Less staff and expertise required to implement, tune, and maintain

Find out more:  http://www-01.ibm.com/software/data/infosphere/smart-analytics-system/

**Unlock The Business Value Of Information**

# Blue Cross Blue Shield Delivers New Intelligence With The Smart Analytics System

## Challenge:

Give Employers, Consumers, and Providers:

- ☑ Decision Support
- ☑ Actionable Information
- ☑ Predictive Analysis

**BlueCross BlueShield Association**

Chicago, IL-based Blue Cross and Blue Shield Association (BCBSA), a national federation of 39 independent BCBS health insurance companies

## Solution:

Blue Cross and Blue Shield is using components of IBM's Smart Analytics System, (InfoSphere Warehouse and Cognos) as the backbone of the new "Blue Health Intelligence (BHI)," a unique resource that will help to improve healthcare quality. BHI is now able to **better understand and detect demographic trends**, such as comparing an employer's plans to those of companies in the same industry, to improve workforce health while better managing costs for the employee.

## Benefits:

- ☑ **Gain insight** into healthcare trends and clinical best practices
- ☑ **Deliver information about treatment and medical technology effectiveness**
- ☑ **Identify emerging trends** in healthcare practice and delivery

*"This is revolutionizing the way we deliver healthcare analytics while giving us faster access to the types of information we didn't have before."*

*Shirley Lady, Vice President, Blue Health Intelligence*

# IBM Cognos vs. Oracle And Microsoft – Summary

| | IBM | Microsoft | Oracle |
|---|---|---|---|
| Integrated, Comprehensive Performance Management Platform | IBM Cognos | **NO** Multiple unintegrated products tied to Microsoft OS | **NO** Multiple separate products (Oracle, Hyperion, Siebel, Brio) |
| 100% web-based zero footprint BI | IBM Cognos | **NO** Requires MS Office, BI Dev Studio, Report Builder, more | **NO** BI Publisher uses MS Word, Many BI functions require desktop apps and plug-ins |
| Common metadata model | IBM Cognos | **NO** Each component has its own metadata model | **NO** Multiple separate metadata strategies |
| Common report specification format | IBM Cognos | **NO** RDL not used by all tools, incompatibilities between versions | **NO** Multiple technologies using different report formats |
| Purpose-built architecture for IT simplification | IBM Cognos | **NO** Collection of many technologies | **NO** Collection of many technologies |

# IBM Information On Demand
## Unlock The Business Value Of Information

## Why DB2 Is The Right Database For New Intelligence

IBM developerWorks **Live!**

Abbreviated Title Here

# New Intelligence Demands A High Efficiency Foundation



Data warehousing, business intelligence and content management all put demands on the underlying database.

Your database software is the foundation for New Intelligence. DB2 is not only inexpensive but highly efficient.

**Service Oriented Finance CIO**

**IBM**

**Unlock The Business Value Of Information**

# DB2 Is A High-Efficiency Design

- **High Efficiency Scalability** keeps things simple and postpones scale-out

- DB2 has **High Efficiency Compression** that saves twice as much on storage  **\*\*ENHANCED\*\***

- DB2's **High Efficiency XML** storage takes less space and has much better performance  **\*\*ENHANCED\*\***

- New **High Efficiency tools** make developers and DBAs more productive

- **High Efficiency DB2 HADR** beats Oracle RAC

- DB2 exploits IBM hardware to improve performance and availability

**EPA Fuel Economy Estimates**

These estimates reflect new EPA methods beginning with 2008 models.

CITY MPG **47**  HIGHWAY MPG **52**

Combined Fuel Economy
This Vehicle **47**
10  47
All SUVs

Your actual mileage will vary depending on how you drive and maintain your vehicle

See the FREE Fuel Economy Guide at dealers or www.fueleconomy.gov

**\*\*ENHANCED\*\***

*Oracle 11g, Oracle RAC and Microsoft SQL Server require more resources to perform the same work!*

*DB2 9.7 makes this story even better*

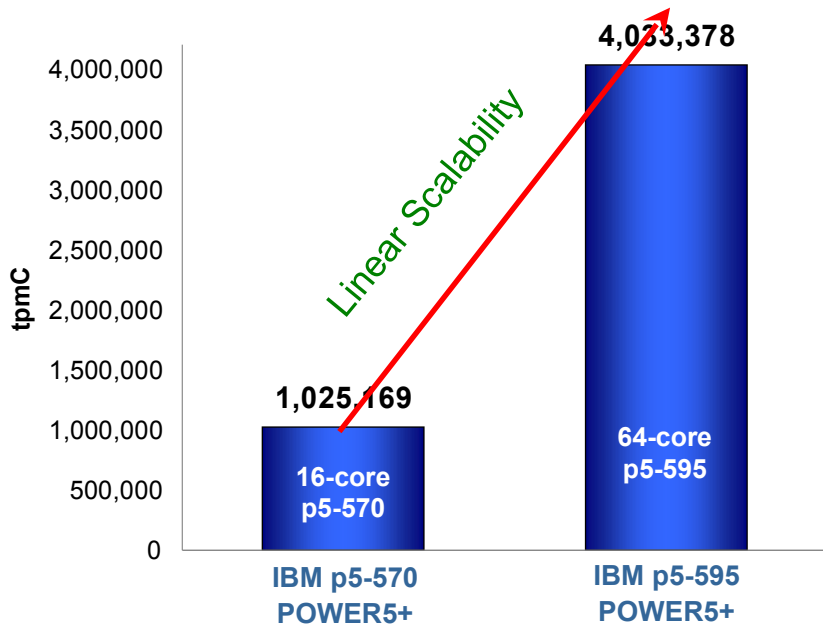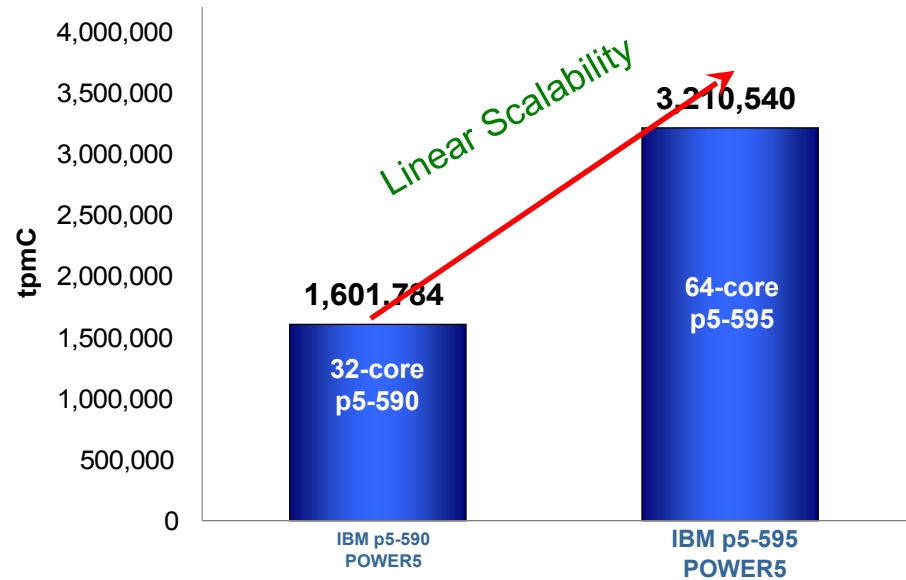# TPC-C Benchmarks Demonstrate DB2's High Efficiency Scaling On Power Systems

4x Cores, 3.93x Throughput

2x Cores, 2.00x Throughput



**DB2 Benchmarks on Power Systems POWER5+**

**DB2 Benchmarks on Power Systems POWER5**

# System z With DB2 Scales Further Than Best Oracle On HP Superdome Banking Benchmark

- **Kookmin Bank**
  - ▶ **IBM System z9 and DB2**
  - ▶ **TCS BaNCS (Cobol)**
  - ▶ **15,353** Transactions/second
  - ▶ **50** Million Accounts
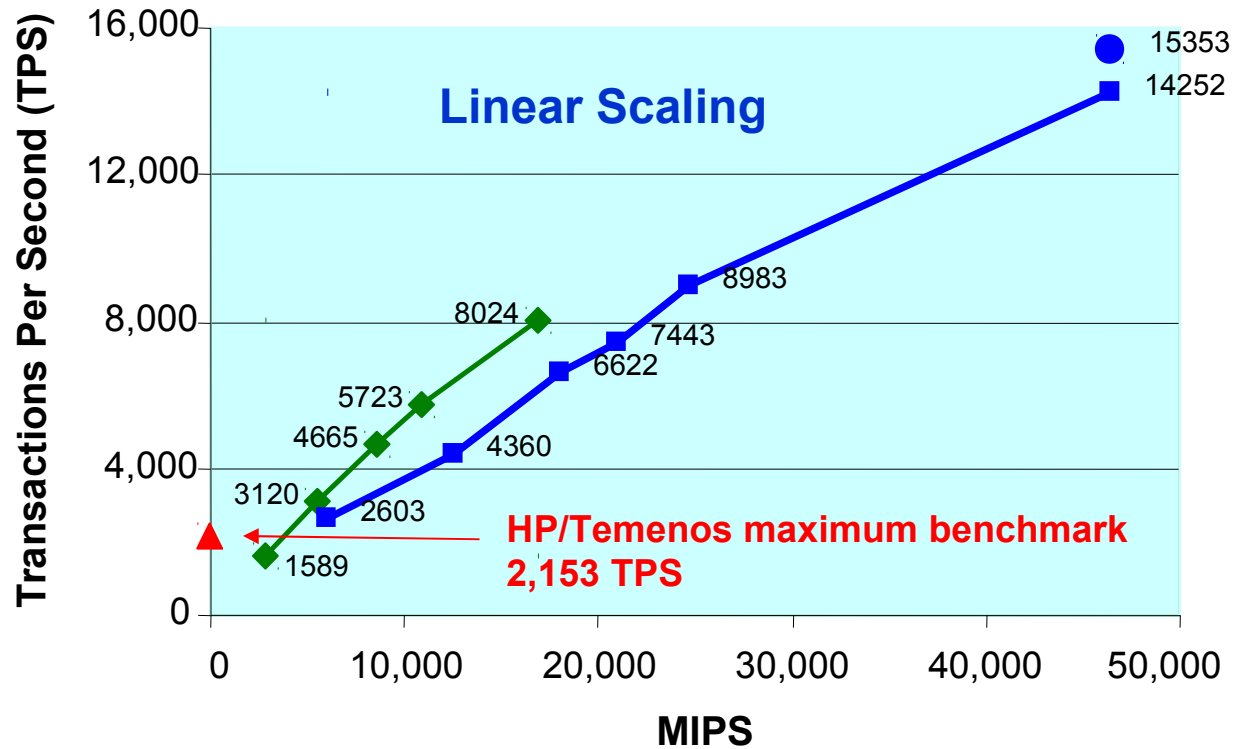  - ▶ IBM benchmark for customer

- **Bank of China ***
  - ‣ **IBM System z9 and DB2**
  - ‣ **TCS BaNCS (Cobol)**
  - ‣ **8024*** ** Transactions/second
  - ‣ **380** Million Accounts
  - ‣ IBM benchmark for customer

- **HP/Temenos ***
  - ‣ **HP Itanium and Oracle**
  - ‣ **Temenos T24 (Java)**
  - ‣ **2,153** Transactions/second
  - ‣ **13** Million Accounts
  - ‣ Largest banking benchmark performance claimed by HP

## System z and BaNCS Online Banking Benchmarks

Transactions Per Second (TPS) vs MIPS

Values plotted:
- 15353
- 14252
- **Linear Scaling**
- 8983
- 8024
- 7443
- 6622
- 5723
- 4665
- 4360
- 3120
- 2603
- 1589

**HP/Temenos maximum benchmark 2,153 TPS**

Y-axis: 0, 4,000, 8,000, 12,000, 16,000
X-axis (MIPS): 0, 10,000, 20,000, 30,000, 40,000, 50,000

* **SOURCE**: TEMENOS BENCHMARKS; http://h71028.www7.hp.com/enterprise/downloads/TemenosBenchmark.pdf
** **SOURCE**:http://www.enterprisenetworksandservers.com/monthly/art.php?2976  **Source**: InfoSizing FNS BANCS Scalability on IBM System z – Report Date: September 20, 2006
*** Standard benchmark configuration reached 8024 tps, a modified prototype reached 9445 tps

# What Is DB2 Storage Compression?

- Implemented by software

- Compression looks for repeating patterns <u>across the entire table</u>
    - Replaces repeating strings with a 12-bit symbol
    - Symbols are stored in a dictionary for fast lookup
    - Oracle and SQL Server use page level compression

- Data resides in compressed form in memory and on disk
    - **Significant disk storage cost savings**
    - **I/O bandwidth savings**
    - **Memory usage savings**
    - Some CPU costs
        - Rows must be decompressed before being processed for evaluation

# DB2 9.5 Compression Is More Efficient

**Data Warehouse Compression Tes**
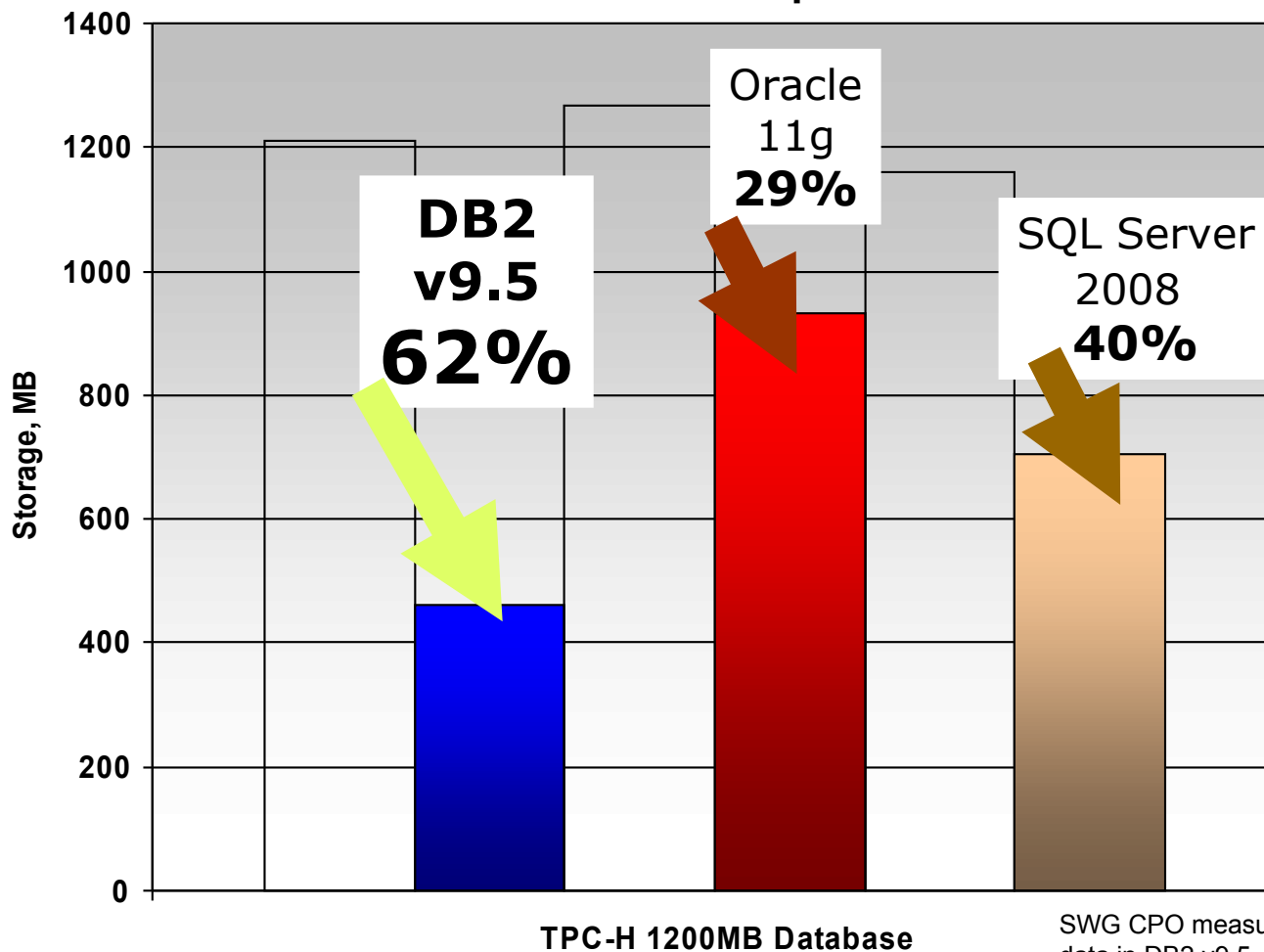


- **DB2 uses half the storage of Oracle**

- **Uses 2/3rds the storage of SQL Server**

**TPC-H 1200MB Database**

SWG CPO measurements of TPC-H benchmark
data in DB2 v9.5, Oracle 11g and SQL Server 2008

# DB2 9.7 Compression Is Even Better

- **DB2 9.7 adds more compression features**
  - ▸ Adds index compression
    - ▪ Row id (RID) list compression
    - ▪ Between 30 and 55% compression for SAP test indicies
  - ▸ Adds temp table compression
  - ▸ Adds inline LOBs
    - ▪ which can take advantage of page compression
  - ▸ Adds XML XDA page compression

- **Only DB2 does RID list and temp table compression**
  - ▸ Neither Microsoft or Oracle can do this

# XML – The Difference Is Fundamental

- ## Relational is a data model

  - Relations (tables)
  - Attributes (columns)
  - Set based w/some sequences
  - Strict schema

| SSN | CreditReportID | CreditDate |
|---|---|---|
| 111111111 | 1234 | Dec 12, 2007 |
| 111111111 | 4456 | Feb 8, 2008 |
| 123456789 | 2314 | Nov 30, 2007 |

| SSN | LastName | FirstName | Street | City | State | Zip |
|---|---|---|---|---|---|---|
| 111111111 | Haan | Brian | 1 Harry Rd | San Jose | CA | 95141 |
| 123456789 | Smith | Joe | 555 Bailey Ave | San Jose | CA | 95141 |

| CreditReportID | CreditBureau | CreditLiability | Rating |
|---|---|---|---|
| 1234 | ABC Credit | Collection | 649 |
| 1235 | ABC Credit | Collection | 687 |
| 2314 | TRW Reporting | Mortgage | 750 |

- ## XML is a data model

  - Hierarchical tree structure
  - Nodes (elements, attributes, comments, etc.)
  - Relationships between nodes
  - Sequence based w/ some sets
  - Flexible schema

```
<MISMOVersionID="2.3.1" ?>
<RespondingParty="ABC Credit">
  <RESPONSE_DATA>
    <CREDIT_RESPONSE_MISMOVersionID="2.3.1"
      CreditResponseID="CRResp0001"
      CreditRatingCodeType="Equifax">

    <CREDIT_BUREAU _Name="ABC Credit" _StreetAddress="..
     ...
    </CREDIT_BUREAU>
    <BORROWER BorrowerID="B1" _FirstName="Joe" _LastName="Smith"
     ....
    </BORROWER>
    <CREDIT_LIABILITY CreditLiablityID="CrL12923"
     ....
      <_CURRENT_RATING _Code="9" _Type="Collection/>
.......
</RESPONSE_DATA>
```

# DB2 9 Native XML Storage Is Called pureXML

- A "Hybrid" data base environment combining the relational and XML hierarchical data models

  ‣ Adds a new "XML" data type

- A new storage mechanism to efficiently manage XML data

  ‣ "Native" means that XML documents are stored on data base pages as parsed tree structures to reflect XML's hierarchical structure
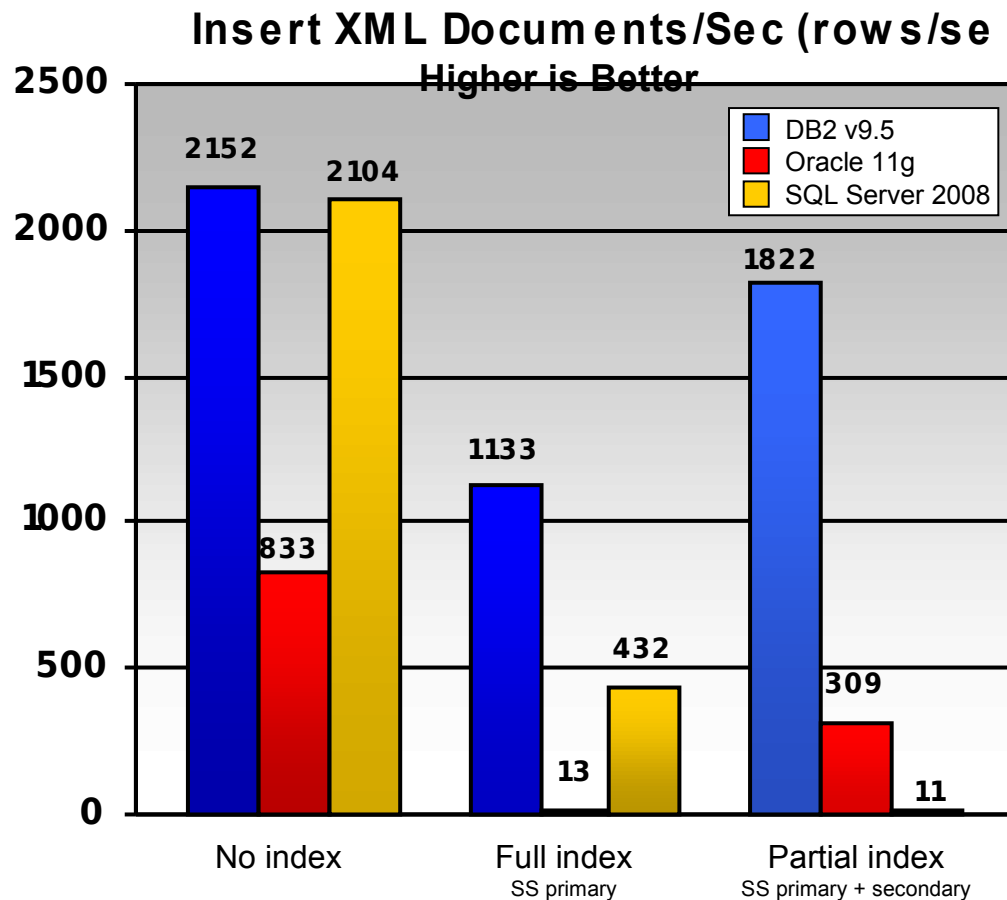
- This avoids conversions between XML and relational structures, and the corresponding limitations

  ‣ Input and retrieval are faster, performance is better, and querying is better and faster

  ‣ With BLOBs and shredding, every operation (parsing, etc.) is expensive and there is a potential loss of data

  ‣ The XML document might be too complex to shred

# DB2 pureXML Stores XML Data Faster Than Oracle11g And Microsoft SQL Server 2008

- DB2 indexed inserts are more efficient
  - ▸ 6x – 85x faster than Oracle
  - ▸ 2.6x – 160x faster than SQL Server

- Oracle's poor insert performance gets worse

- SQL Server primary index similar to DB2 full index
  - ▸ Secondary index similar to DB2 partial index
  - ▸ Secondary index <u>requires</u> primary
  - ▸ 2 indexes slow performance

**Insert XML Documents/Sec (rows/se**
**Higher is Better**

Legend:
- ■ DB2 v9.5 (blue)
- ■ Oracle 11g (red)
- ■ SQL Server 2008 (yellow)

No index:
- DB2 v9.5: 2152
- Oracle 11g: 833
- SQL Server 2008: 2104

Full index (SS primary):
- DB2 v9.5: 1133
- Oracle 11g: 13
- SQL Server 2008: 432

Partial index (SS primary + secondary):
- DB2 v9.5: 1822
- Oracle 11g: 309
- SQL Server 2008: 11

SWG CPO study based on inserting TPOX benchmark data into DB2 v9.5, Oracle 11g and SQL Server 2008

**Unlock The Business Value Of Information** © 2008 IBM Corporation

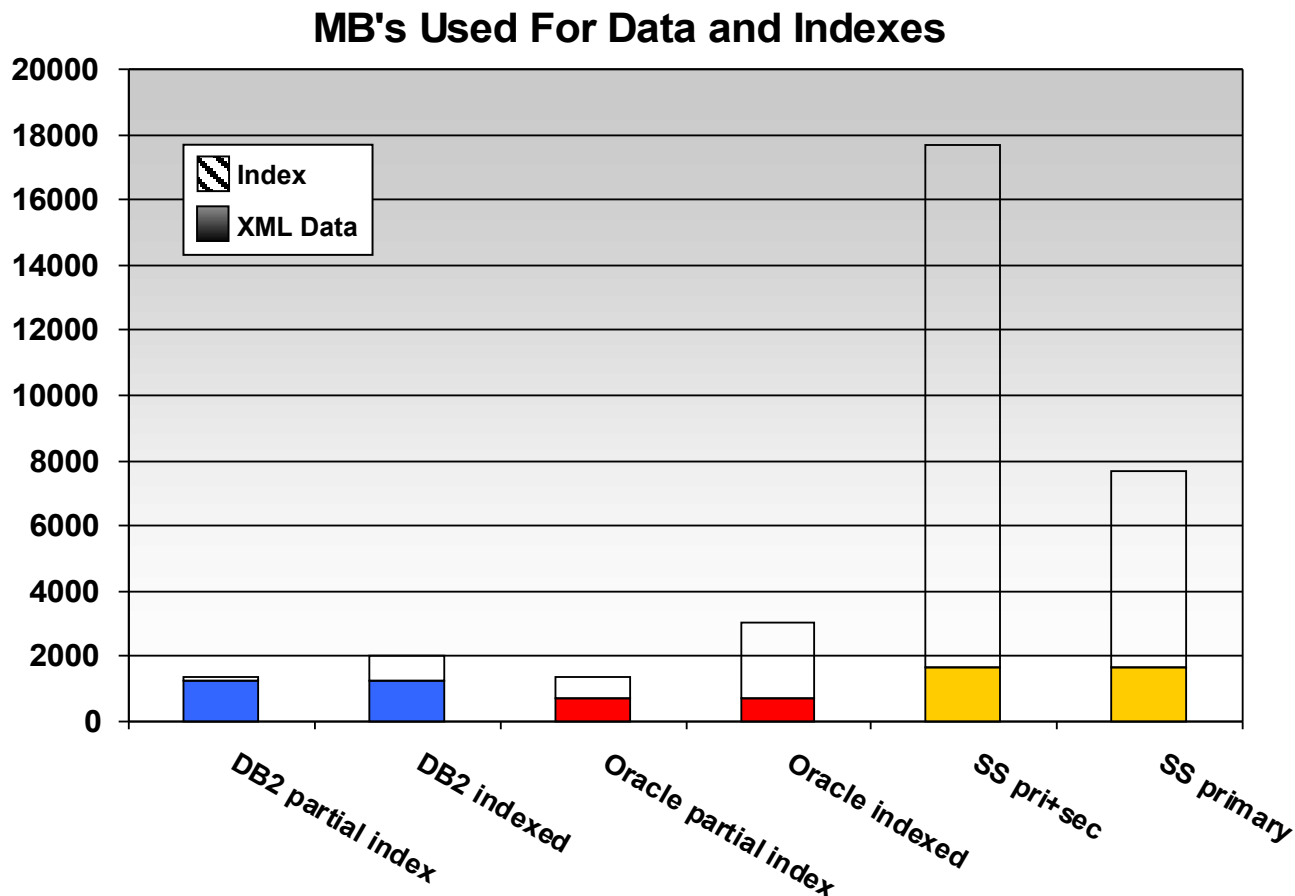# DB2 Partial Index XML Query Performance Blows Away Competitor Results!

- DB2 indexed query performance is efficient
  - 4x – 77x faster than Oracle 11g
  - 6x faster than SQL Server 2008

- DB2 partial index generated 18x performance improvement
  - Oracle full index generated 5x improvement
    - Still slower than DB2 without indexes
  - SQL Server performance slowed by half with primary index
    - Requires secondary index to generate 7.5x improvement

**Indexed XML Query Performance (queries/min)**

**Higher is Better**

| | DB2 | Oracle | SQL Server |
|---|---|---|---|
| Unindexed | 24 | 1.4 | 9.5 |
| Full Index | 29 | 7.7 | 4.6 |
| Partial Index | 451 | 5.8 | 71 |
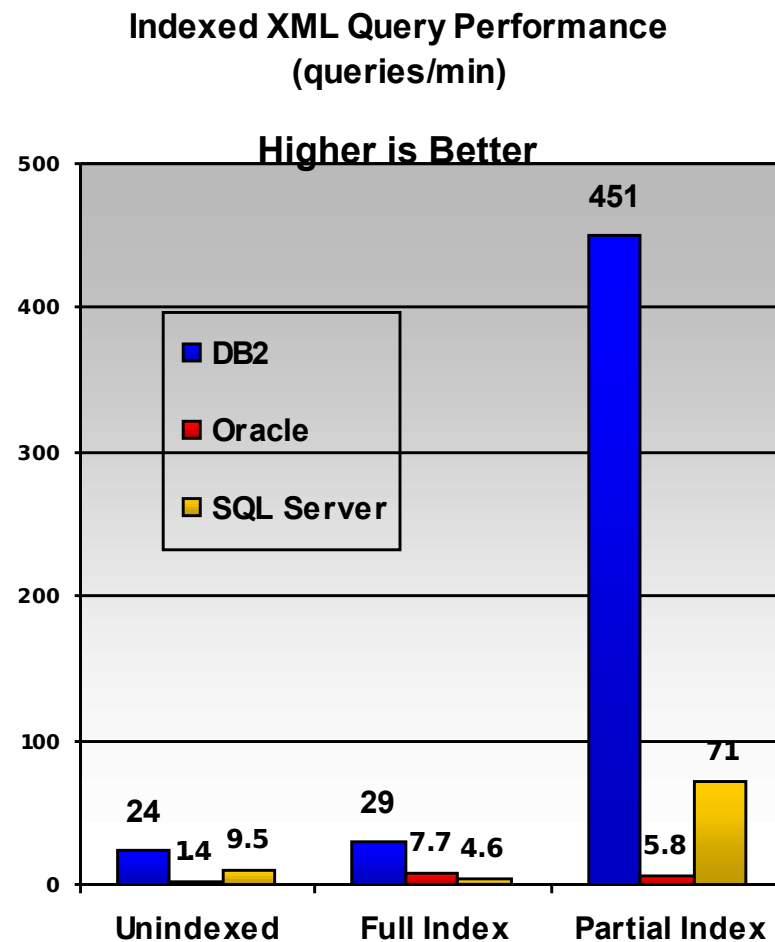
SWG CPO study based on querying 1GB TPOX benchmark data from DB2 v9.5, Oracle 11g and SQL Server 2008

**Unlock The Business Value Of Information**

# DB2 Uses Less Storage For XML Than Oracle And A Tiny Fraction Of What SQL Server Needs



**MB's Used For Data and Indexes**

Legend:
- Index
- XML Data

X-axis categories: DB2 partial index, DB2 indexed, Oracle partial index, Oracle indexed, SS pri+sec, SS primary

Y-axis: 0 to 20000

*50% less storage than Oracle 11g full indexes*
*35% - 90% less storage than SQL Server 2008*

Study based on storing TPOX benchmark data in DB2 v9.5 and SQL Server 2008

# Develop XQuery Web Services Applications 2.5x – 10x Faster With Optim Data Studio

| Project steps to expose XQuery data as a web service | IBM | Oracle | Microsoft |
|---|---|---|---|
| 1) Create Project, add database connection and test | 1m 18s | 1m 21s | 0m 45s |
| 2) Import and test XQuery | 1m 17s | 1m 42s | 1m 08s |
| 3) Create Web Service from XQuery | 1m 51s | 49m 25s | 11m 17s |
| 4) Test Web Service and package it for installation | 0m 56s | 0m 54s | 0m 34s |
| **TOTAL:** | **5m 22s** | **53m 22s** | **13m 44s** |

- **IBM Optim Data Studio v2.1 handles XML development tasks more efficiently than Oracle JDeveloper or Microsoft Visual Studio**

- **Wizards generate pureQuery code to improve application performance**

- **Rapid, iterative development of SQL or Java stored procedures**

| Step 3) Create Web Service from XQuery | | |
|---|---|---|
| IBM | Oracle | Microsoft |
| - Run a wizard | - Create service definition class and testing harness<br>- Write JDBC code XQuery<br>- Attach missing JAR files<br>- Test and debug<br>- Create web service from Java class | - Create new ASP.NET Web Service<br>- Write ADO .NET code for running XQuery<br>- Test and debug |
| **< 2 min** | **> 49 minutes** | **> 11 minutes** |

N

# DB2 pureXML Outperforms Oracle 11g And SQL Server

## Summary of test results

DB2 pureXML vs Oracle 11g
- 2.5x to over 85x faster inserts
  - 4x to 77x faster queries
  - 50% less storage used with full indexing
  - Develop XQuery Web Services 10x faster

DB2 pureXML vs SQL Server 2008
- 2.5x to over 150x faster inserts
  - 2.5x to 6x faster queries
  - 35% to 90% less storage used
  - Develop XQuery Web Services 2.5x faster

SWG CPO study based on measurements of 1GB TPOX benchmark
data from DB2 v9.5, Oracle 11g and SQL Server 2008

**Unlock The Business Value Of Information**

# DB2 9.7 Extends XML Support

- XML now fully supported in Multidimensional Clustering (MDC) tables
  - ▸ Dimensional indexes can be combined with XML indexes
  - ▸ Not clustering the XML Data itself
    - ▪ Storage for XML remains in XML storage objects (XDA objects)
  - ▸ XML column is not part of the clustering key

- XML now fully supported in Range Partitioned Tables
  - ▸ Partitions can contain XML columns
  - ▸ XML column is not part of the partitioning key

**Unlock The Business Value Of Information** © 2008 IBM Corporation

# DB2 Has Autonomic Capabilities To Make Administrators More Productive

## Everyday tasks are simply automatic with the DB2 Control Center!

- Automatic Configuration Parameters

- Automatic Statistics collection

- Automatic Backup

- Automatic Table reorganization

# IBM Optim Database Administrator 2.2 Improves Administrator Productivity

- **Simplifies promoting changes from development to test or test to production**

  ‣ Compare, synchronize, copy, clone, or merge database schema definitions from the source to the target

  ‣ Specify masks and ignores to simplify the comparison of tables

  ‣ Migrate objects from physical data models, live database connections, or from scripts

  ‣ Move objects and data with copy and paste, drag and drop, or directly from a comparison

  ‣ Make changes-in-place or develop scripts for later use

- **Eclipse-based**

- **Integrates with Rational ClearCase or other version control systems**

# Customers See The Benefits Of DB2's High Efficiency Administration

"There's far less administration involved with DB2 than with Oracle." **BOC runs "50-100 SAP systems," supported by 12 people "That is really quite extraordinary." -** Sheila Moran at BOC in UK

"DB2 requires significantly less database administration than Oracle. **We can now deploy our IT staff for more productive and business-critical needs" -** Zdenek Vosahlo, Head of IT at Precheza

# Ease Of Administration – Big Part Of TCO

## Solitaire Interglobal Study - Staffing
### *Real world study of 250 sites*



**Relative Staffing by Discipline**

50% less DBA staffing required

Total for DB2     15.6 FTE
Total for Oracle 22.1 FTE

DB2 requires 30% less overall staffing

FTE

Skill Category

DB2  Oracle

Source:  Solitaire Interglobal

**Unlock The Business Value Of Information**

© 2008 IBM Corporation
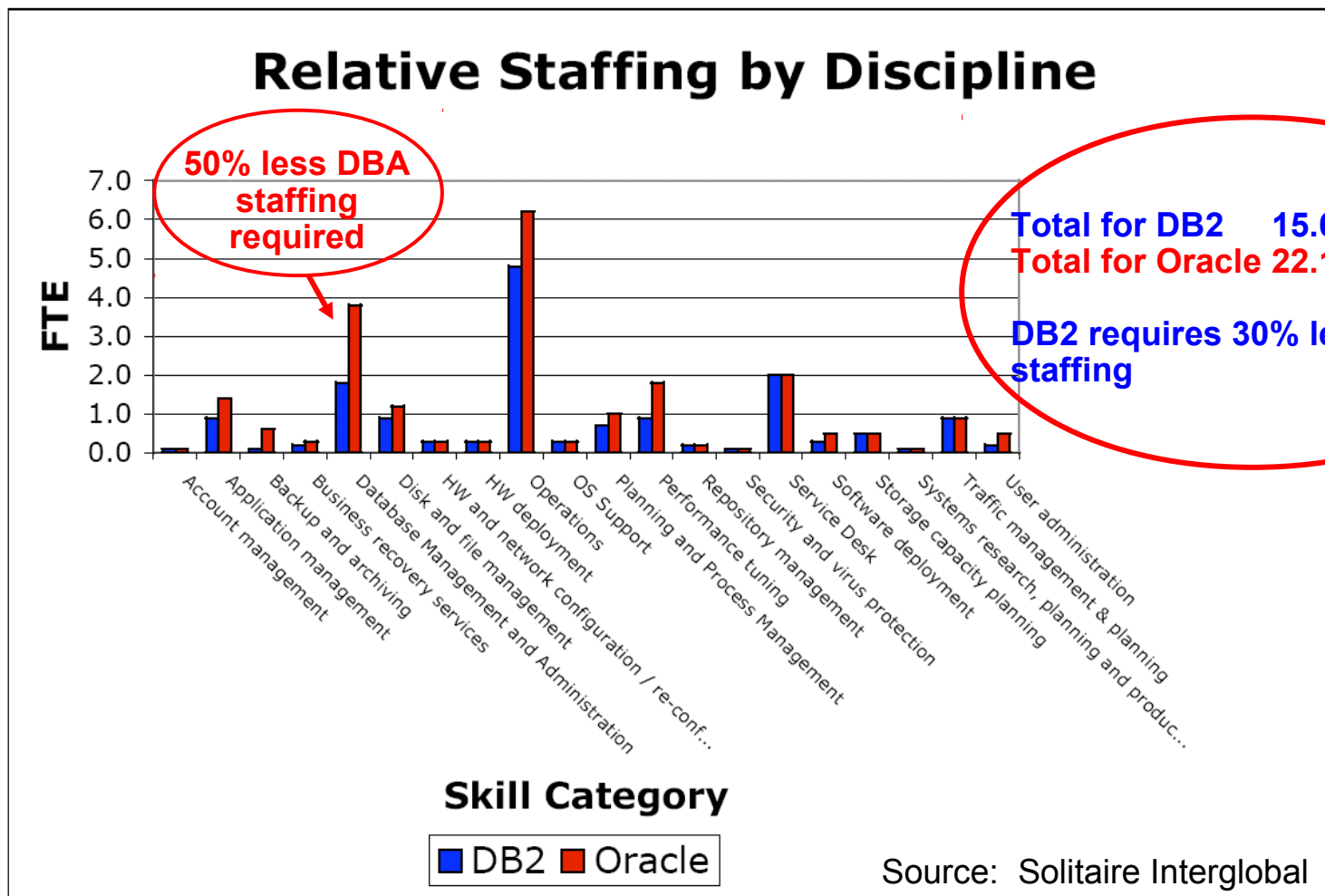
# Oracle RAC Does Not Scale Linearly



**Oracle 11g RAC Performance
(Hard Disk Storage)**

WHY?
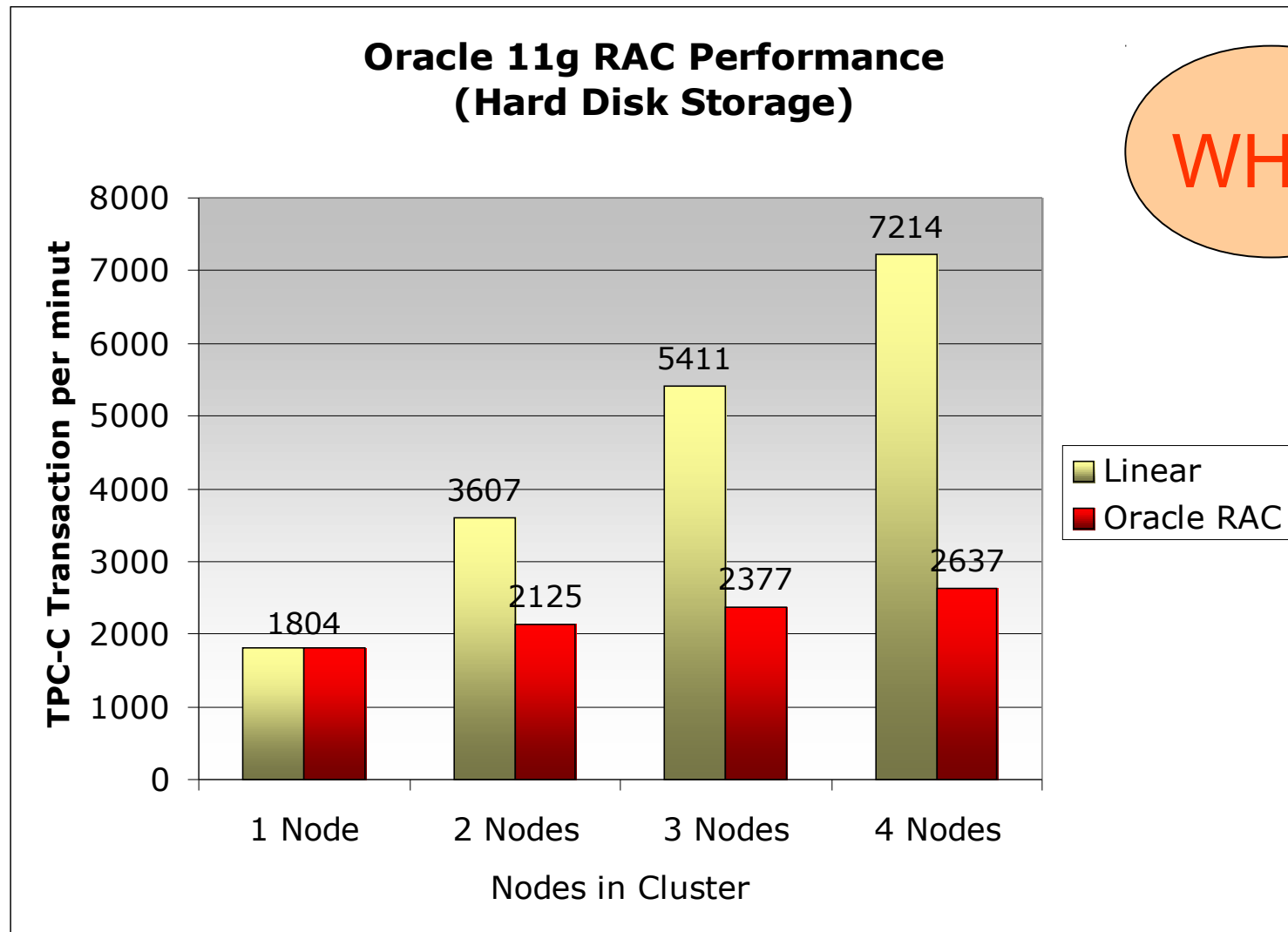
# Oracle RAC Cluster Model Is An Inefficient Design

- **Incoming requests are dispersed (sprayed) among the nodes**
  - ‣ RAC maintains locks, cache per node

- **Requires two separate networks**
  - ‣ A private network for internal traffic between the nodes and the database
  - ‣ A public network for external communication and incoming requests

- **Requires a single copy of the database in storage**

**Oracle RAC**

Network

| Locks | Locks | Locks | Locks |
|-------|-------|-------|-------|
| Cache | Cache | Cache | Cache |
| Oracle | Oracle | Oracle | Oracle |

Switch

Disks
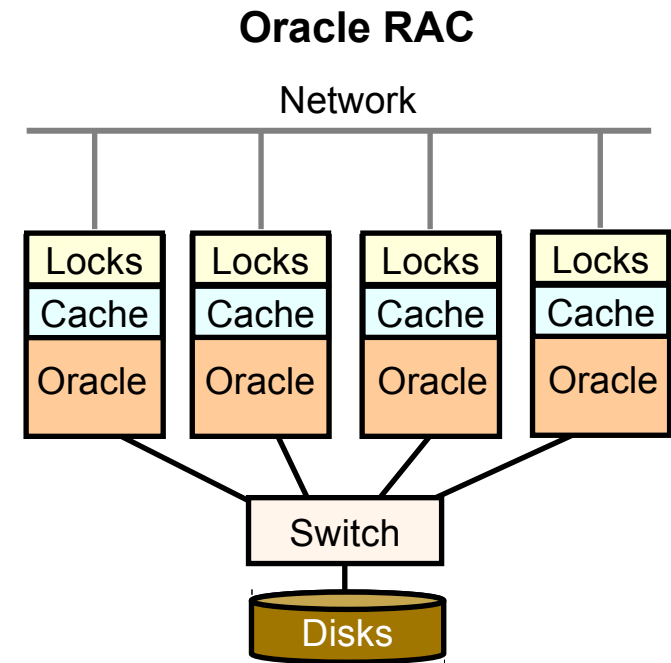
# Oracle RAC Requires More Packets Per Transaction . . .

**Network Traffic: Packets Per Transa**
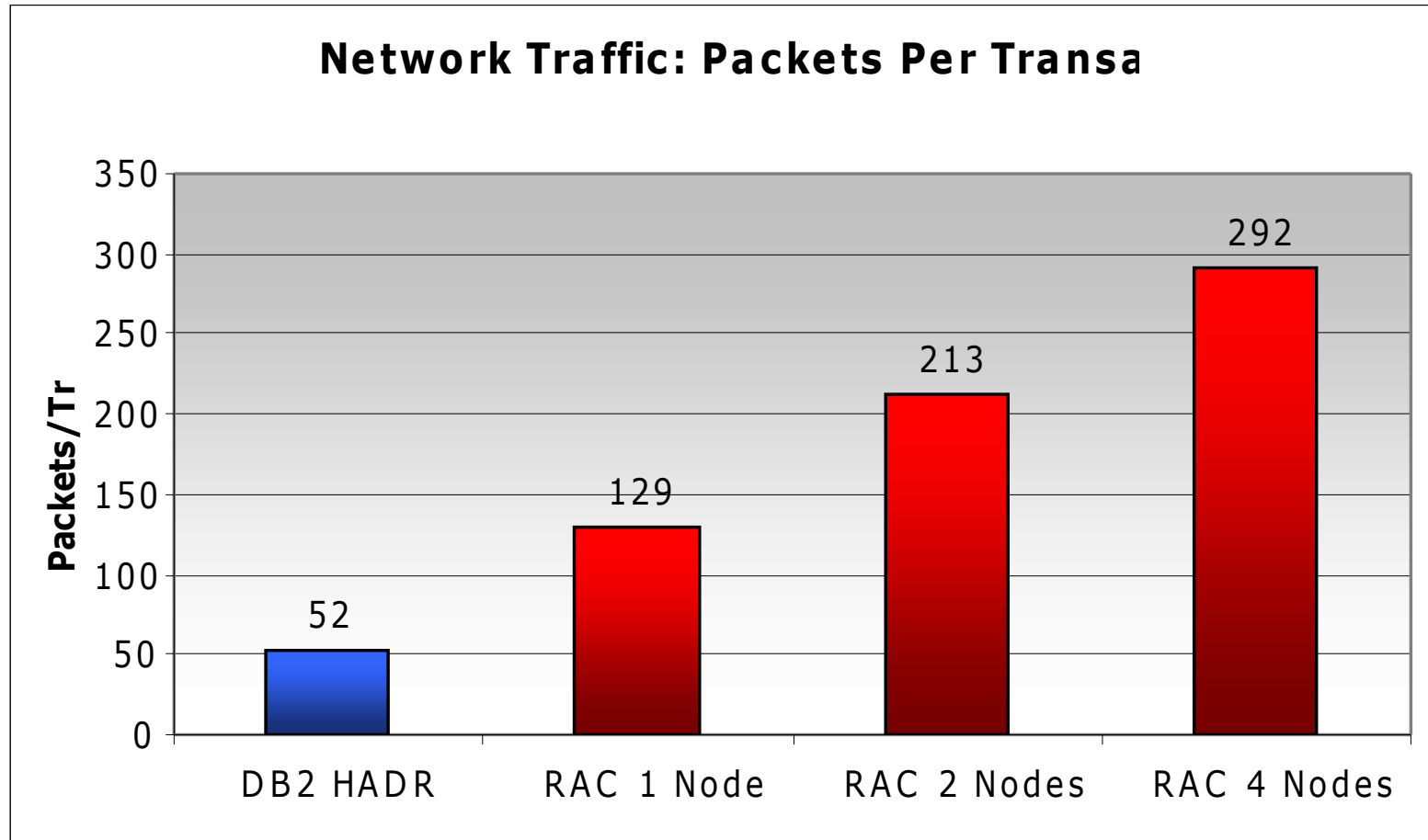
# . . . And Oracle RAC Consumes More CPU Time Per Transaction

- To process identical TPC-C transactions an Oracle 2-node RAC uses more than twice the CPU time than DB2 v9.5 HADR server (almost 3x more with 4 nodes)

- Calculated by dividing CPU busy time (mpstat: user+system) by the number of transactions processed (by this CPU)

- DB2 HADR results normalized to 100%

**CPU Time Per Transaction**

| | |
|---|---|
| 300% | |
| 250% | |
| 200% | |
| 150% | |
| 100% | |
| 50% | |
| 0% | |

DB2 HADR | Oracle RAC 2-node | Oracle RAC 4-node
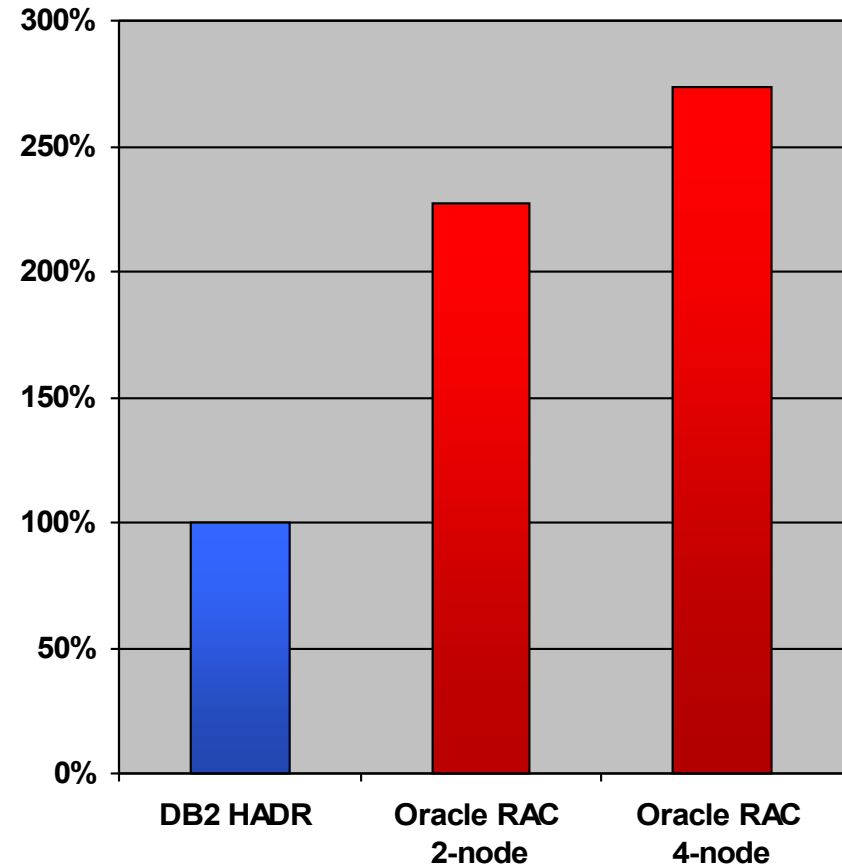
# . . . So Oracle RAC Processes Fewer Transactions Per Node As Nodes Are Added

**RAC Transactions Per Node**



Bar chart titled "RAC Transactions Per Node"

- Y-axis: TPC-C Transactions per minute (0 to 2000)
- X-axis: Nodes in Cluster

| Nodes in Cluster | TPC-C Transactions per minute |
|---|---|
| 1 Node | 1804 |
| 2 Nodes | 1062 |
| 3 Nodes | 792 |
| 4 Nodes | 659 |
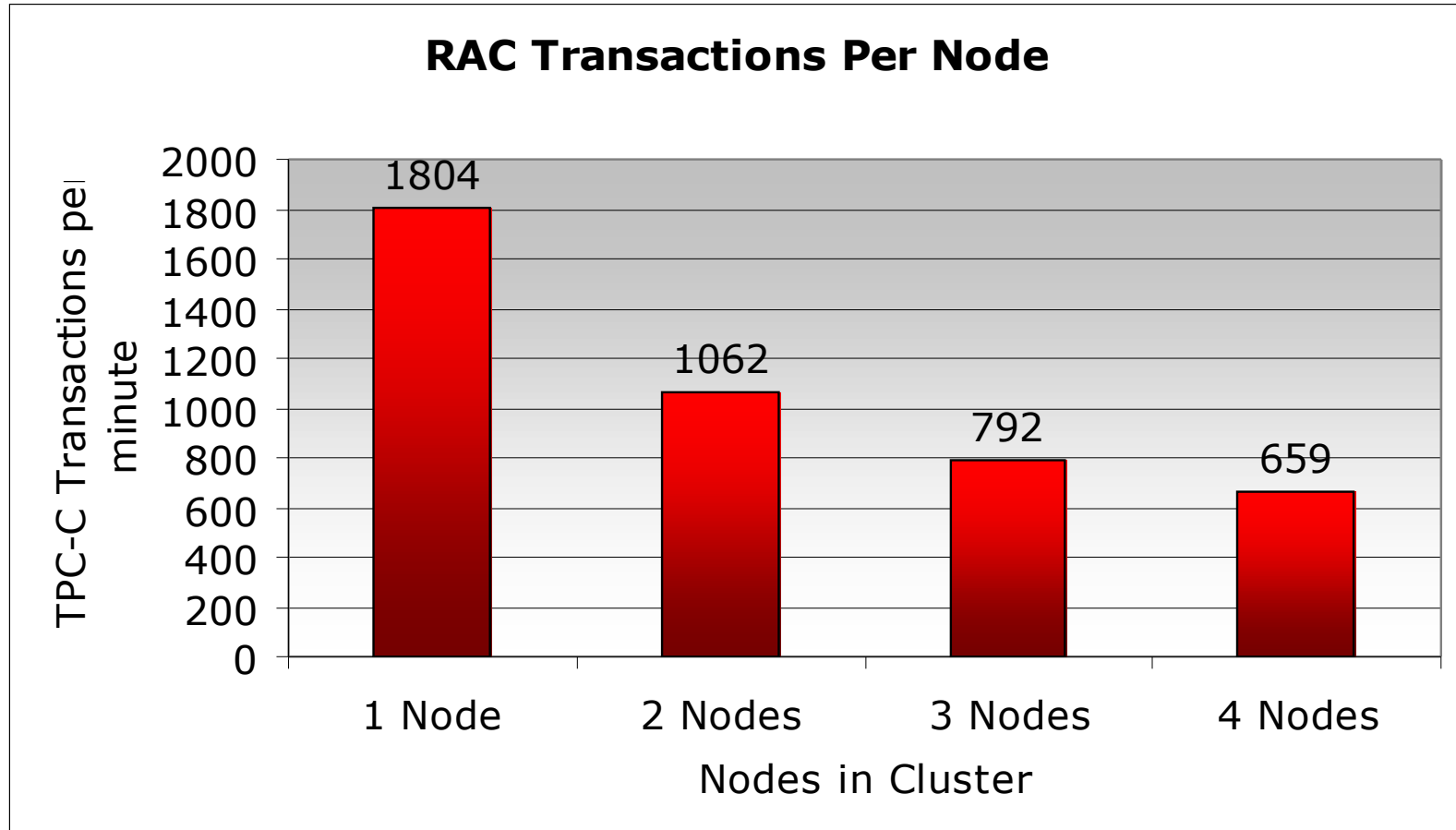
# HP Agrees!  Oracle Scale-Out Glass Ceiling

- DB2 provides near-linear scalability on Power Systems

- With Oracle RAC, overhead increases rapidly as additional nodes are added, and performance degrades significantly after only 4 to 6 nodes

Sources: "Scale-up versus scale-out using Oracle 10*g* with HP StorageWorks", Hewlett-Packard, 2005;

"Enterprise Data Base Clustering Solutions" ITG, October 2003;

Benchmark tests, IBM Toronto Labs and Systems and Technology Group, using TPC-C-like workload, 2.2 GHz POWER5+, 2006 Power Systems TPC Benchmarks

### DB2 and Oracle RAC 10g Scal

**Scaling Achieved** vs **Nodes**

- 16.00
- 15.70
- DB2 scales on SMP
- Oracle RAC "Glass Ceiling"
- 8.00
- 7.52
- 4.00
- 3.97
- 3.04
- 4.04
- 4.31
- 2.00

Legend: Oracle RAC — Linear Scale — DB2 on Power Systems Part

# DB2 High Availability And Disaster Recovery (HADR) Is Better Than Oracle RAC's

- DB2 HADR provides high availability and fast failover
- **Failover in less than 15 seconds**
- **Real SAP workload failover for 600 SAP users performed in 11 seconds**
- Storage mirror survives disaster
- Buffer pool primed on standby server with recent updates
- 100% performance after server failure
- Uses ordinary network and storage devices

- RAC failover is delayed due to remastering of distributed locks
- Only one copy of storage
- Degraded performance after server failure
- Pay for Oracle RAC for every node
- Specialized network and storage raises cost of ownership, erodes any savings from commodity servers

**Automatic Client Reroute**
Client application automatically resumes on Standby

Network Connection

**HADR**
Keeps the two servers in sync

Primary Server          Standby Server

**Oracle RAC**

Network

| Locks | Locks | Locks | Locks |
| Cache | Cache | Cache | Cache |
| Oracle | Oracle | Oracle | Oracle |

Switch

Disks

IBM
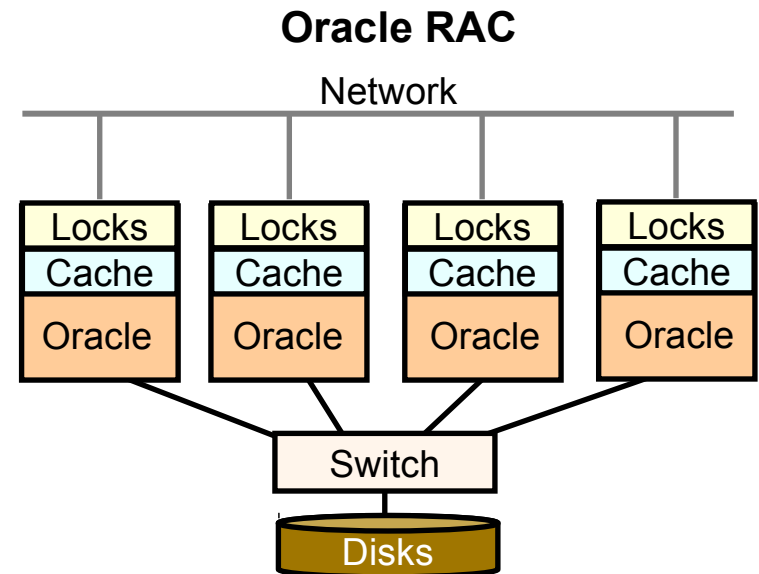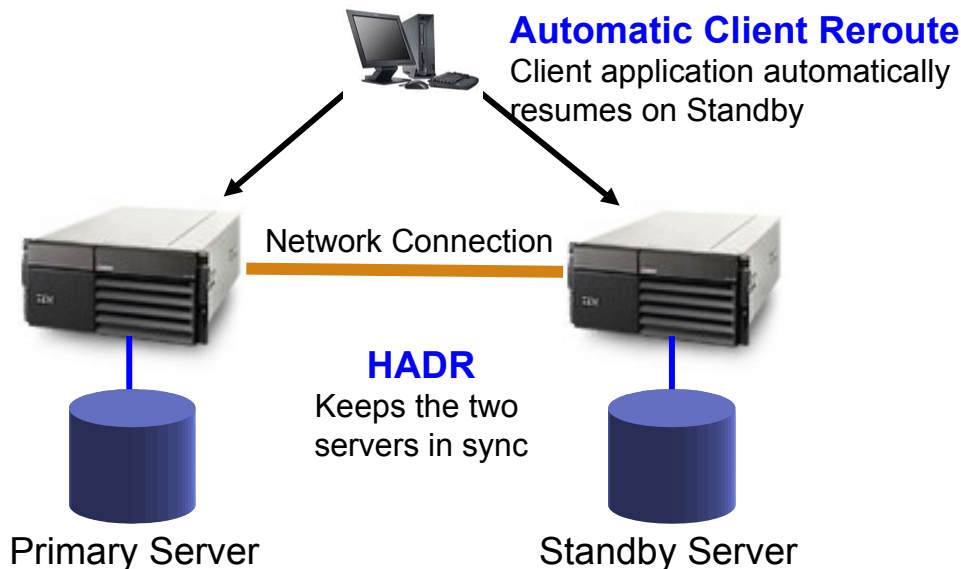
# Data Security And Compliance:
# DB2 For z/OS Has A Proven Track Record

## DB2 for z/OS Security

- 10 security related patches in the last 10 years

- Proven RACF and Multi Level Security

- End-to-end encryption via hardware assist

- Optim Test Data Management
  - Ensures anonymous access to data necessary for testing

- Optim Archiving Expert
  - Allows customers to easily archive and access data

- DB2 Audit Management Expert
  - Supports compliance requirements
  - Consul for enterprise wide audit

## Oracle's Security Exposures

- **Oracle.com – July 2009**
  30 security patches, including **12** for the database (with 3 being remotely exploitable WITHOUT authentication!)

- **Oracle.com – April 2009**
  43 security patches, including **16** for the database

- **eWeek.com – January 2009**
  41 security patches, including **10** for the database

- **eWeek.com – October 2008**
  23 security patches, including **7** for the database

In the last year Oracle has issued 137 security patches, **45** for the database

**Unlock The Business Value Of Information**

# DB2 Is Optimized To Take Full Advantage Of System z

- Coupling Facility enables unmatched scalability through centralized cache and lock management
- Hipersockets enable fast, secure communication between DB2 z/OS and zLinux applications
- Integrated with RACF and Multi Level Security
- Supports System z hardware encryption – Crypto Express 2
- DB2 Recovery Expert for automatic recovery and backup
- Offload up to 40% of workload onto zIIP processors to reduce licensing costs
- Hardware compression in addition to "Venom" deep compression can reduce storage up to 70%
- System z parallel sysplex supports rolling updates to running DB2

# DB2 Is Also Optimized To Take Full Advantage Of Power Systems And AIX

- Supports AIX 64KB, 16MB and 16GB large page sizes

- Optimized DB2 resource object alignment with Power Systems architecture

- DB2 takes advantage of AIX storage protection keys for security

- Deep integration between AIX workload management (WLM) and DB2 WLM
  - Helps meet service levels and maintain predictable performance via work priority settings and finer levels of monitoring

- Dynamic reconfiguration
  - Allows administrators to add and remove processors, memory and I/O adapters to and from LPARs, without disturbing operations or applications

- Recovery integration
  - DB2 recovery process with Power Systems autonomic computing technologies

- First Failure Data Capture (FFDC)
  - Provides failure analysis and automated recovery capabilities

- Oracle has a product philosophy of running on commodity servers

# DB2 v9.5 Costs 22% Less Than Oracle 11g

| | DB2 v9.5 Enterprise Edition | Oracle 11g Enterprise Edition |
|---|---|---|
| Base CPU license (dual core Intel/AMD) | $ 40,500 | $ 47,500 |

| Popular Options | | |
|---|---|---|
| XML | INCLUDED | INCLUDED |
| Compression | $ 15,300 | $ 11,500 |
| Advanced Security | $ 11,100 | $ 11,500 |

| | | |
|---|---|---|
| SUPPORT (First year) | INCLUDED | $ 15,510 ($10450 + $2530 + $2530) |

| | | |
|---|---|---|
| TOTAL | **$ 66,900** | **$ 86,010** |

Prices for Dual Core Intel Server

# Break Free From Oracle With DB2 9.7



That all sounds great BUT …

We have Oracle applications written in PL/SQL. I can't afford to port those applications and retrain my development staff.

Good news with DB2 9.7 you can use PL/SQL with DB2.

**Service Oriented Finance CIO**

**IBM**

**Unlock The Business Value Of Information**

# Consolidate Oracle Database Applications On DB2 9.7 With The SQL Compatibility Feature

- Run Oracle workloads in DB2 9.7 without porting or rewriting code
  - ‣ Easier for applications to use DB2
  - ‣ Easier for developers to use DB2
  - ‣ Easier to migrate from Oracle to DB2

**NEW!**

| What Needs To Change With DB2 9.7? | | |
|---|---|---|
| **Oracle** | | **DB2** |
| Oracle Data Types | → | Native support |
| Oracle SQL | → | Native support |
| Oracle PL/SQL | → | Native support |
| Oracle Functions and Built-in packages | → | Native support |
| Oracle Concurrency Control | → | Native support |
| JDBC | → | Native support |
| SQL*Plus Scripts | → | Native support |

**Unlock The Business Value Of Information**

# Native Execution Of All SQL Statements

- Built in PL/SQL native compiler
  - NOT an emulator or translator
- Source level debugging and profiling
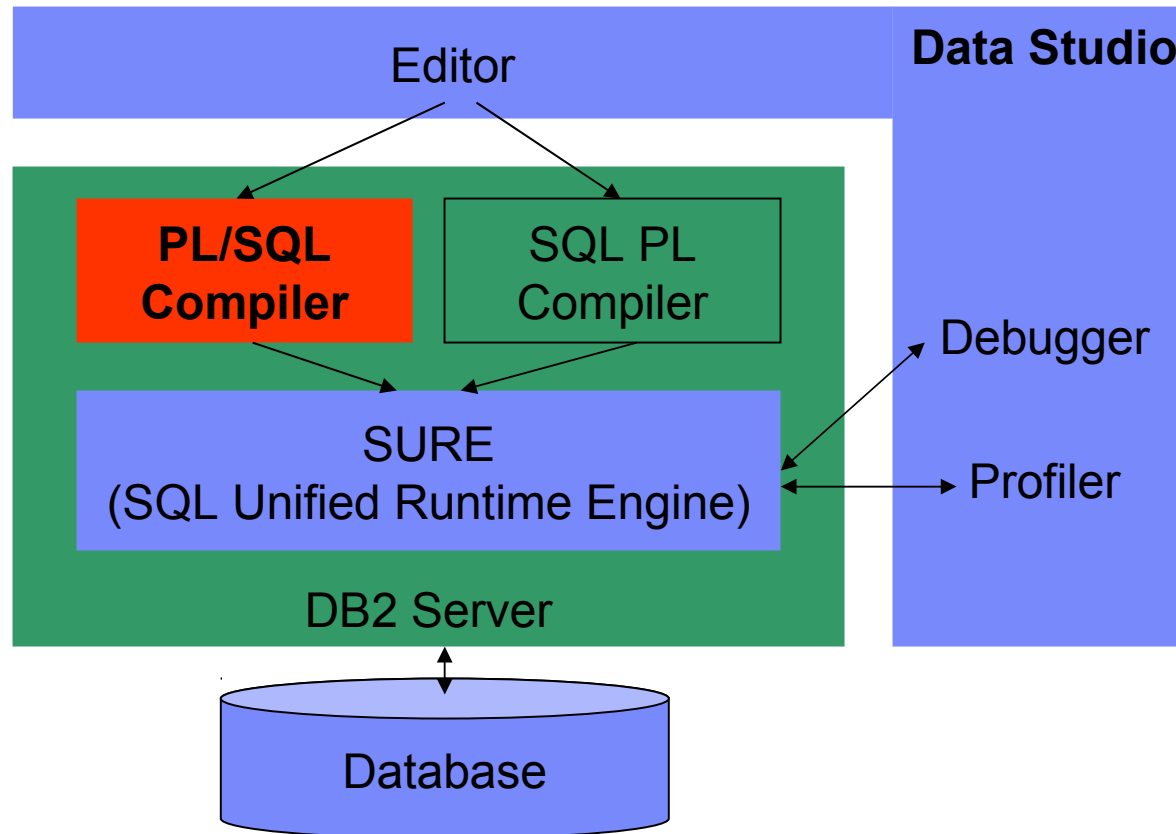
# Praise For SQL Compatibility

**Gartner.**

"The Oracle compatibility feature will **enable Oracle applications to run natively on DB2**. In discussions with Gartner, reference customers tell us that DB2 runs 95% or more of Oracle specific functionality found in SQL statements and natively runs PL/SQL, Oracle's stored procedure language. This **native functionality is not an emulator, nor does it require changes to the application code** (other than the 5%, which is mostly minor functionality, not found in many applications). Any provider offering applications that run with the Oracle DBMS can easily port its application to IBM's DB2…"
– Donald Feinberg, Gartner VP Distinguished Analyst

**Venedim**

"You no longer have to migrate and painfully transcode your Oracle code. Instead, there is a simple flag that will let you compile and run your PL/SQL as is .. **Developers can now choose to code in DB2 SQL/PL or Oracle PL/SQL language, depending on their skills**"
— Jean-Marc Blaise, Technical Architect at Venedim

**openbravo**
opening ERP's future!

"To move our application to a previous version of DB2 would have taken an **estimated two-year effort**. We were **thrilled to see it took only one week** to move it to the new version of DB2.." — Paolo Juvara, CTO of Openbravo

**\*\*The Openbravo application contains about 2 _MILLION_ lines of code\*\***

# High Efficiency DB2 Is The Winning Foundation For New Intelligence At Service Oriented Finance



Now I'm convinced!

We can leave our tired, old databases behind and move forward with DB2.

With DB2, InfoSphere, and Cognos, you have everything you need to make better business decisions and beat your competitors.

**Service Oriented Finance CIO**

**IBM**

My contact information in case you have follow up
 questions:


David Carew

carew@us.ibm.com


# Thank you !