



Extending Your Mainframe For More Business Value

Deliver Business Insight With A
Data Warehouse On System Z

Get More Business Results Out of Your Data

Our branch offices have separate databases.
Each branch is analyzing customers and sales on their own.



**Service Oriented Finance
Marketing**

Looking at data in isolation can miss larger trends and opportunities

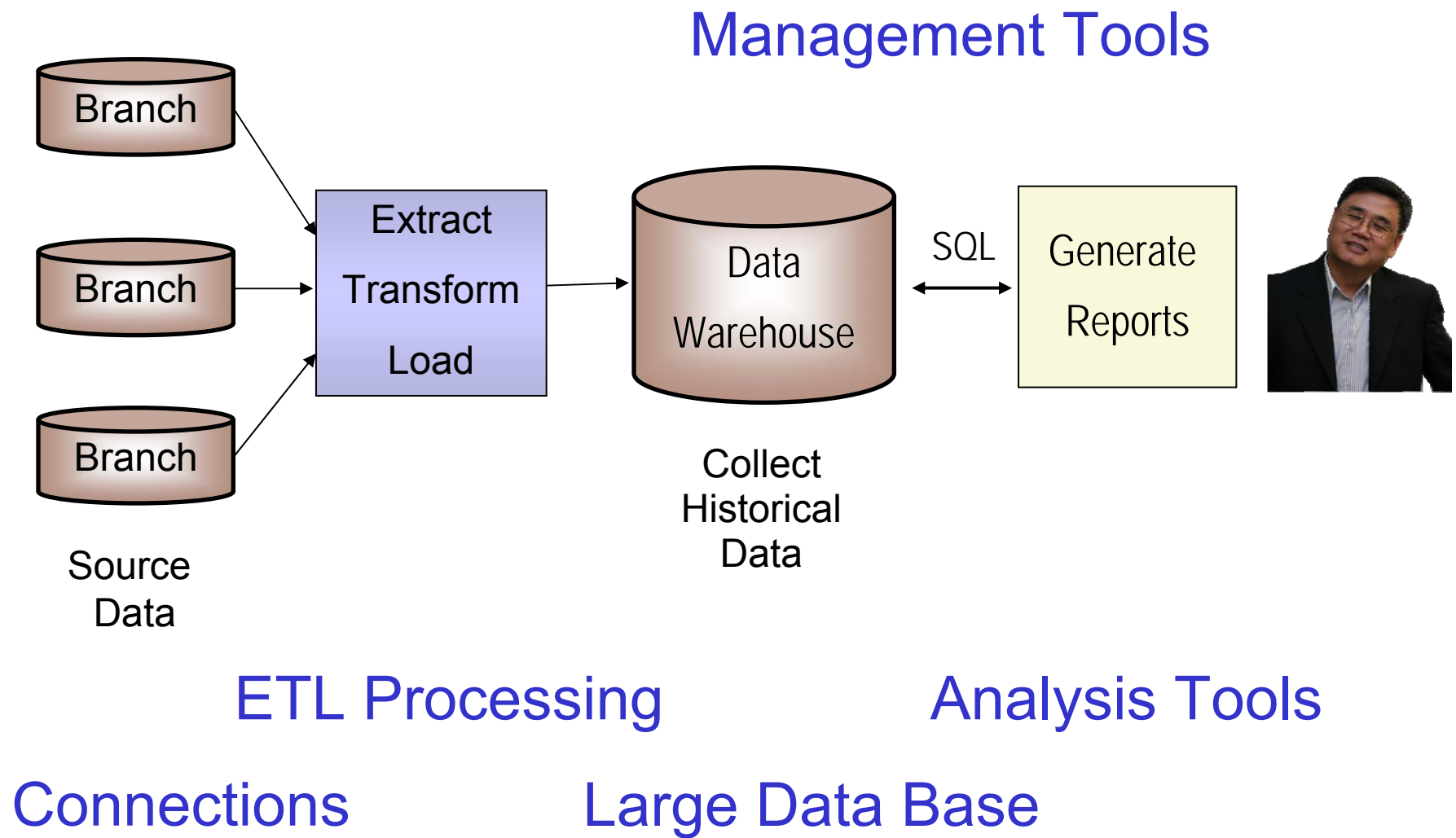


IBM

Service Oriented Finance Needs A Data Warehouse To Make Optimal Business Decisions

- Each branch is responsible for its own marketing campaign
- Corporate marketing gets reports from each of the branches based on local results
- Corporate marketing needs to spot trends to know what campaigns are most effective region-wide
- A corporate data warehouse would give marketing the data to easily do comparisons between the branches and promote best practices

Mainframe Extension Solution – Deliver Business Insight With A Data Warehouse



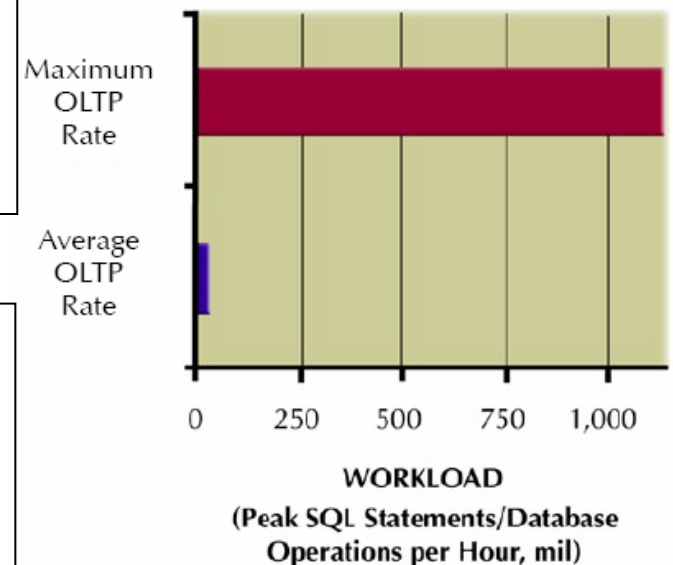
Data Warehouse Capabilities On System z

- Large Capacity Data Base
 - ▶ **DB2 for z/OS**
 - ▶ Parallel queries, Materialized Query Table, Star Join Enhancements
- Connect, Extract, Transform and Load
 - ▶ **IBM Information Server For System z**
 - ▶ Runs on zLinux (DataStage also runs on z/OS)
- Analysis Tools
 - ▶ **DataQuant and Cognos**
 - Solution for reporting, analysis dashboards and scorecards
 - Other tools (QMF, Alphablox, Hyperion, Business Objects, SAS, IBI)
 - ▶ **InfoSphere Master Data Management Server for Linux on System z**
 - More effectively manage high-value operational data
- Performance Monitoring
 - ▶ **IBM Tivoli Omegamon XE for DB2 Performance Expert on z/OS**
- Security and Compliance
 - ▶ **DB2 Data Archive Expert, DB2 Test Database Generator, DB2 Audit Manager Expert, IBM Encryption for DB2 and IMS Databases**
- Dedicated Query Hardware
 - ▶ zIIPs for parallel queries & remote access
 - ▶ Superior I/O bandwidth, multiple I/O paths

Compare Scalability Achievements: Winter Corporation's "2005 Top Ten" Awards

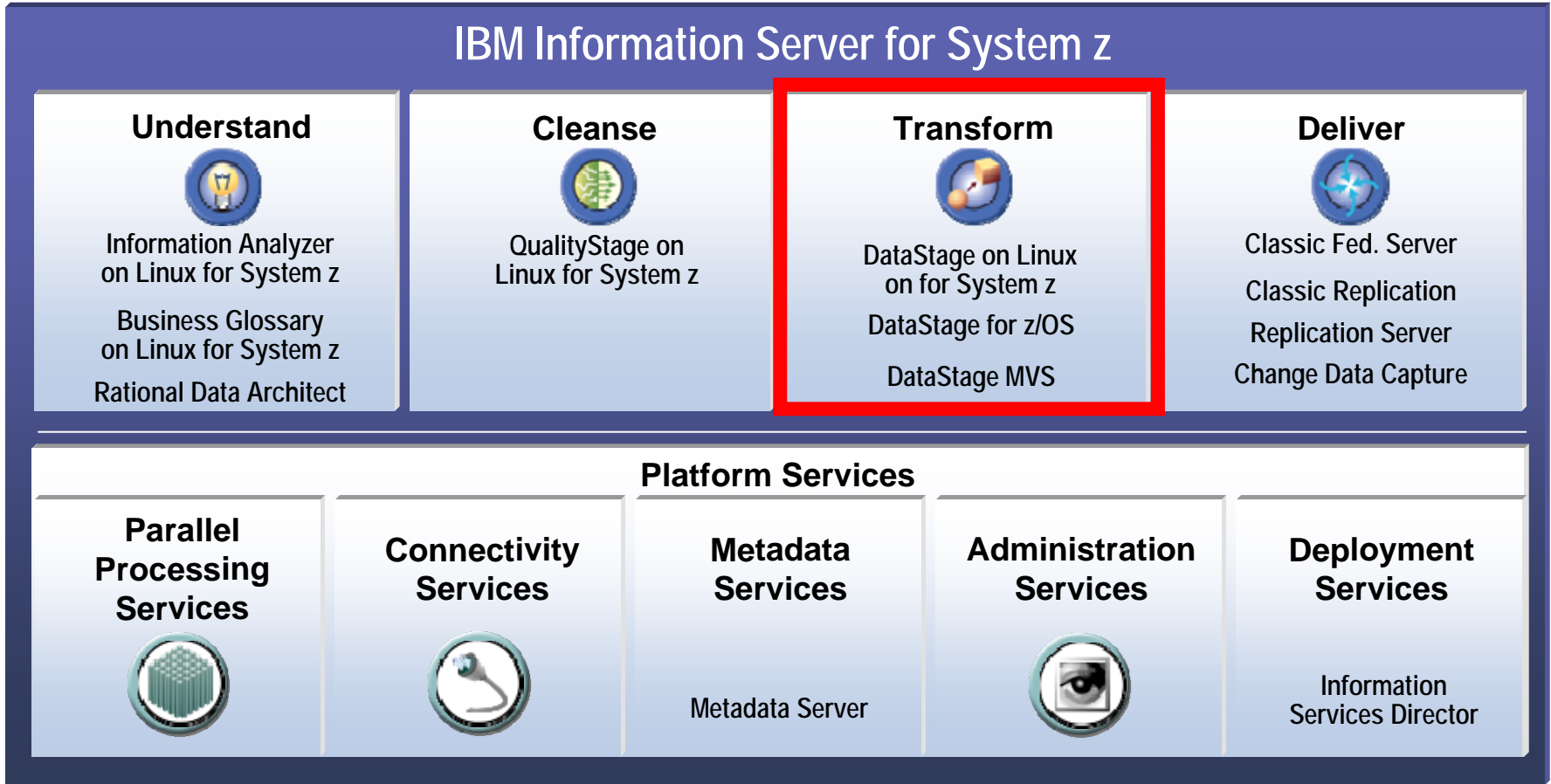
- “The highest performing transaction processing system in the 2005 program, a [DB2] z/OS implementation, executed **over one billion** SQL statements in an hour. The average for operational systems was **35 million** SQL statements or database operations per hour”
 - The study lists the largest known peak workload on **Oracle RAC** to be **8.6 million** SQL statements per hour
-
- “The largest transaction processing [database] in the program, **23 TB**, was hosted on [DB2] z/OS, as in the last program”
 - The study lists the largest transaction processing database on **Oracle RAC** to be **9.6 TB**

Leading OLTP System,
DB2 for z/OS, Processed
Over One *Billion* SQL
Statements per Hour

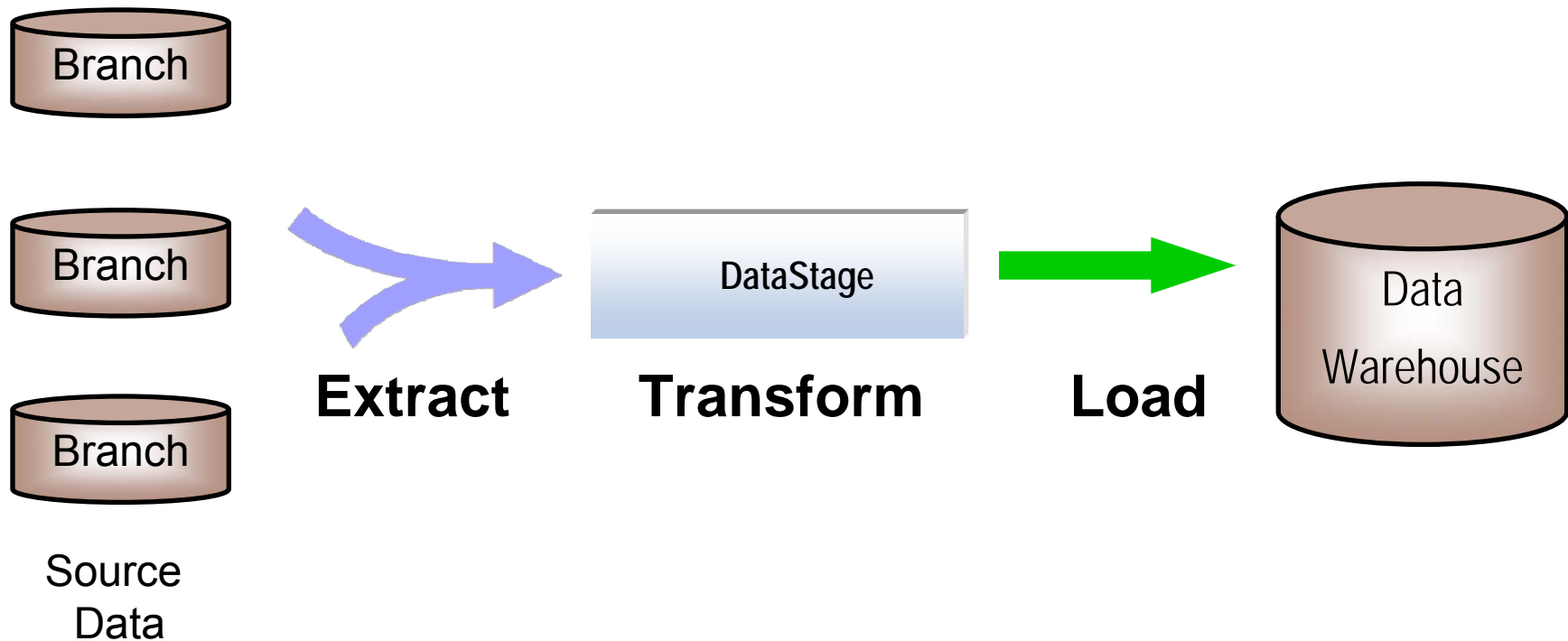


http://www.wintercorp.com/VLDB/2005_TopTen_Survey/TopTenWinners_2005.asp
http://www.wintercorp.com/WhitePapers/WC_TopTenWP.pdf

IBM Information Server For System z Can Load Your Data Warehouse



Load Your Data Warehouse With DataStage



Data Stage Transforms Data On The Fly

- Different field names
- Different field order
- Add Branch Identifier
- Different currency format



PROD ID	CUST ID	BRANCH ID	QTY	AMT	SALEDATE
000 101	100	01	01	10,000.00	2007-02-28
000 121	100	01	03	500.50	2007-02-28
000 101	101	01	01	20,000.00	2007-03-01

Data Warehouse



Transform

PRODUCT	QTY	CUSTNO	AMOUNT	DATE

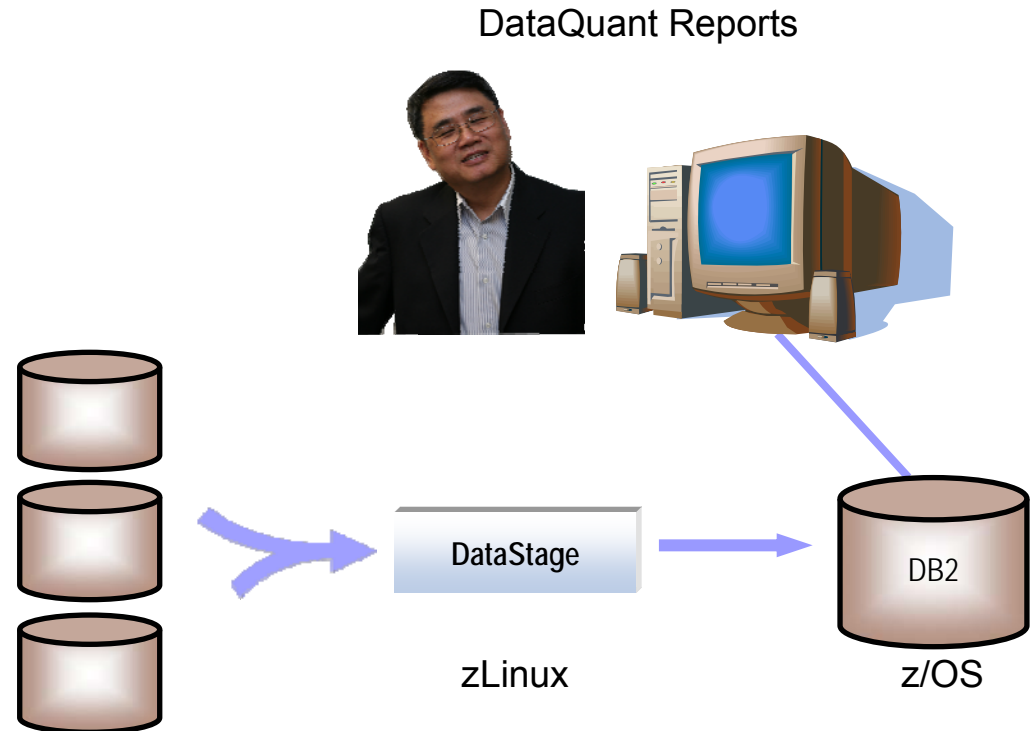
Branch Data

DEMO: Extract, Transform, Load

Use DataStage to load sales and customer data from a branch into your data warehouse

- ▶ ODBC (Input) – Branch Sales info from SQL Server
- ▶ DB2 (Update) – Corporate Sales Warehouse

Show how built-in stages make it easy to handle transformation and aggregations



BlueCross BlueShield Of Tennessee

Challenge

- In order to compete effectively for new business in the complex healthcare market, BCBST needed to differentiate itself from competitors with targeted offerings
- They needed a single view of information across their multiple LOBs with business intelligence capabilities

Solution

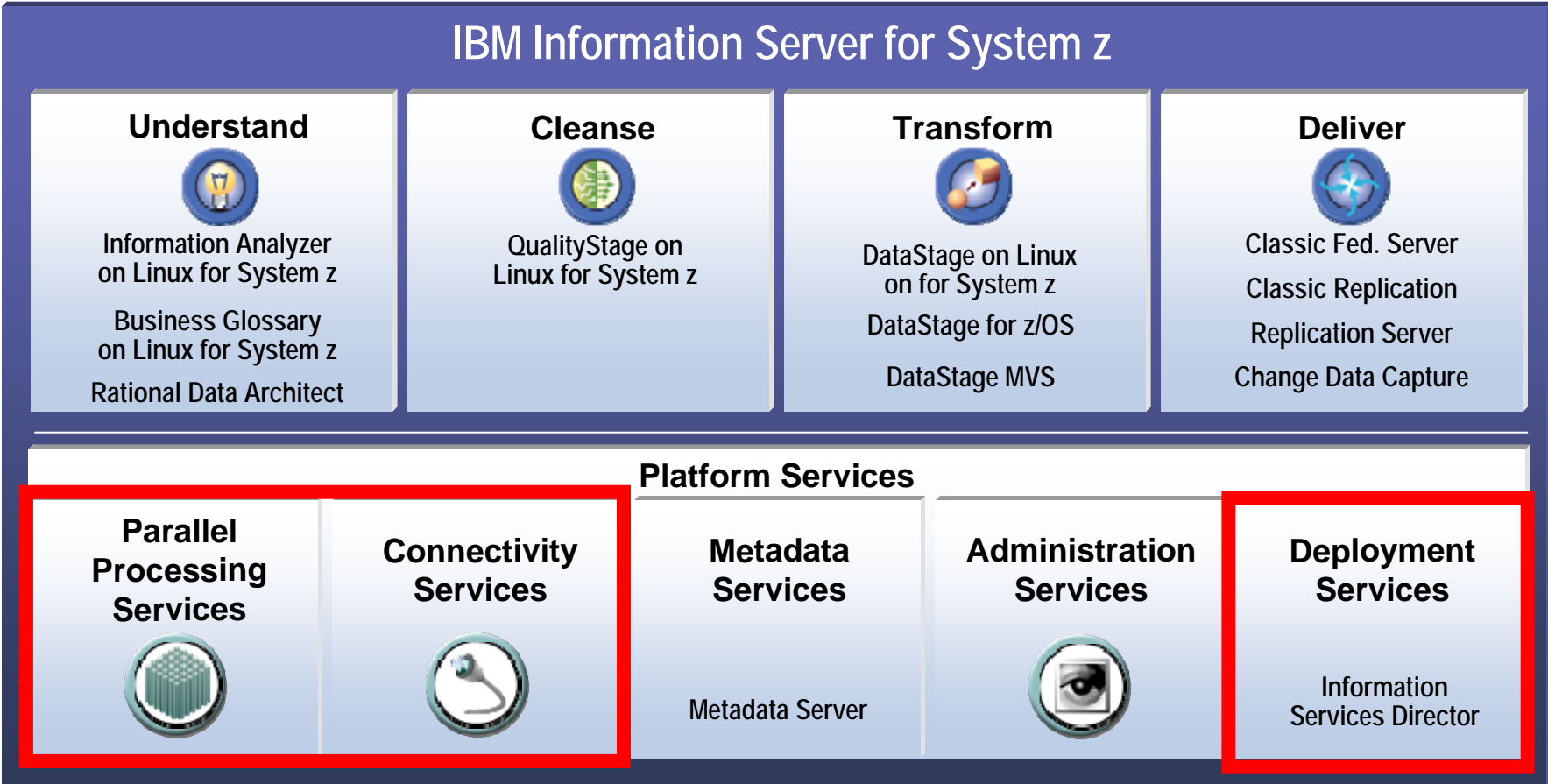
- BCBST is using **IBM Information Server** with **IBM DB2** to profile, transform, and load data to their enterprise data warehouse
- The solution also provides intelligent search capabilities for unstructured data using IBM OmniFind and IBM DB2 Content Manager



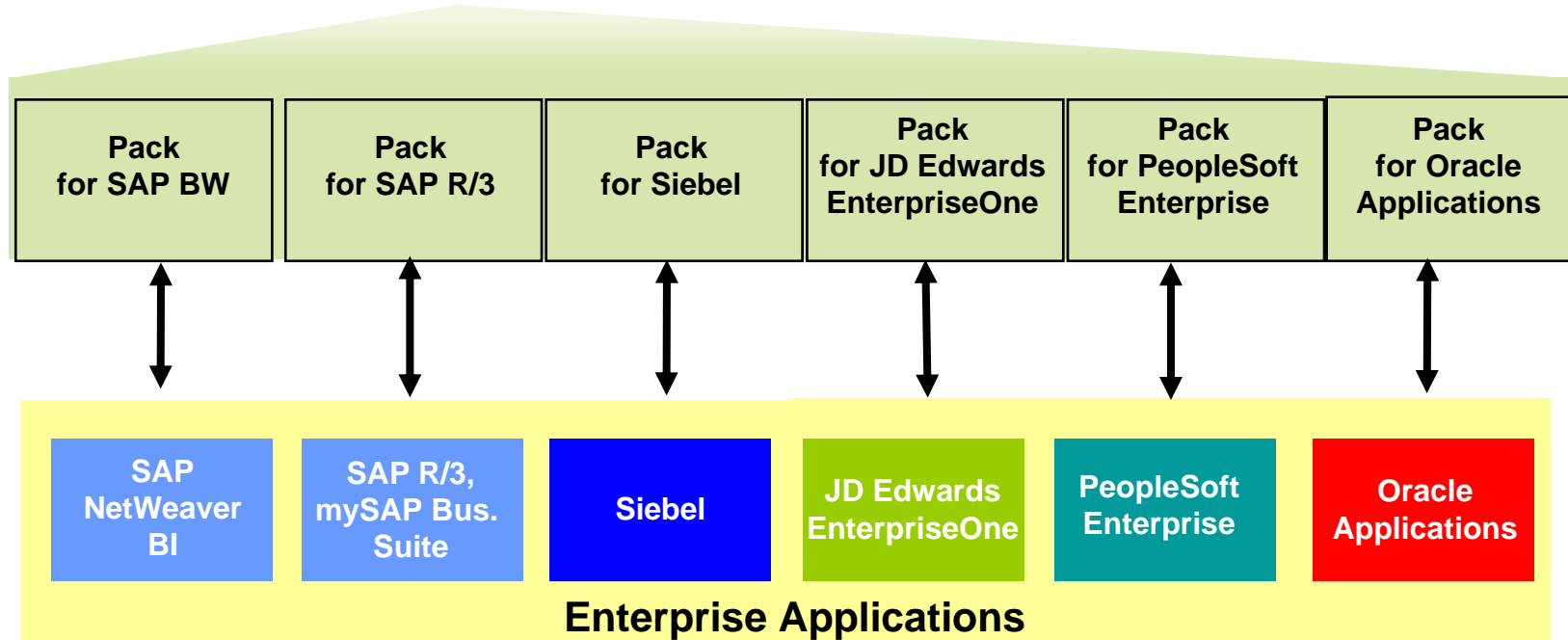
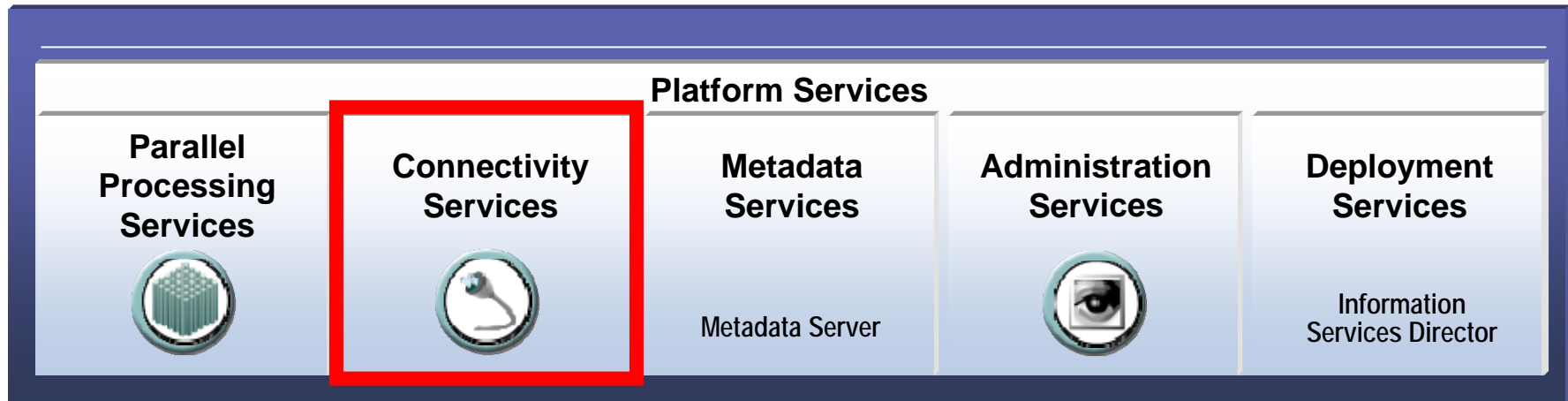
Business Benefits

- Improved customer and provider satisfaction by enabling them to run their own analytics and better manage their healthcare costs.

IBM Information Server For System Z Has Connectivity And Scalability Suitable For The Large Enterprise




Data From 3rd Party Systems Can Be Integrated Using Connectivity Services



IBM 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....



Remove Erroneous Data Before It Gets Into the Warehouse

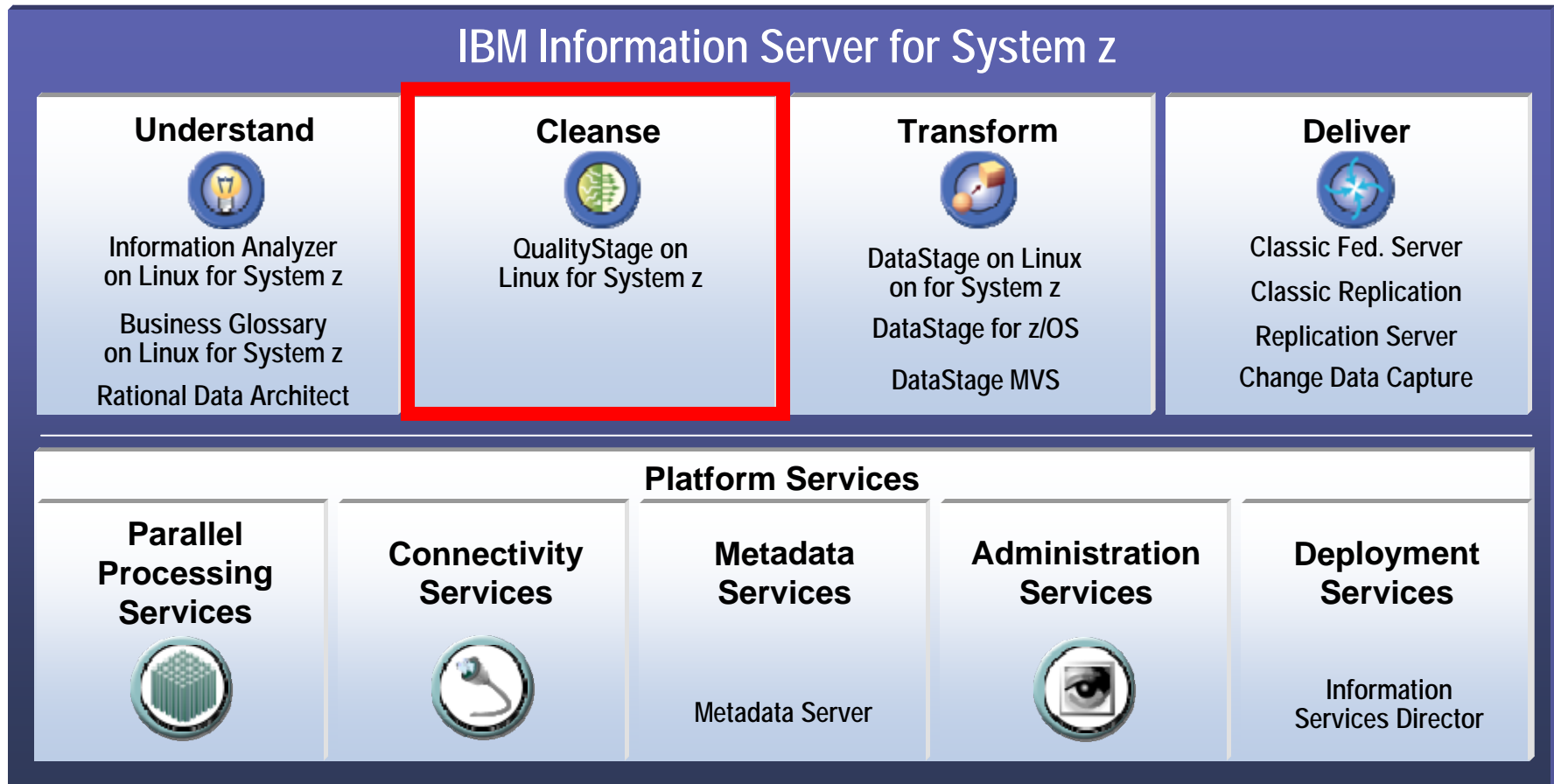
DataStage can Extract, Transform, and Load data into your data warehouse.

But we need to make sure the data is clean.



IBM

IBM Information Server for System z Can Load Your Data Warehouse



Why Should I Care About Cleansing Information?

- **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 data base

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

QualityStage Fixes Data Quality and Consistency

- Ensures clean, standardized information
 - ▶ Eliminates duplications
 - ▶ Matches against reference data
- Supports global postal address verification
- Provides visual tools for designing quality rules and matching logic
- Integrated with DataStage (one platform, one user interface)
- Data quality processes can be deployed within extract, transform, and load (ETL) jobs logic or separately as shared data quality services (SOA)



Two Methods To Decide A Match

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 = BBAABDBA

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

Deterministic Decisions Tables:

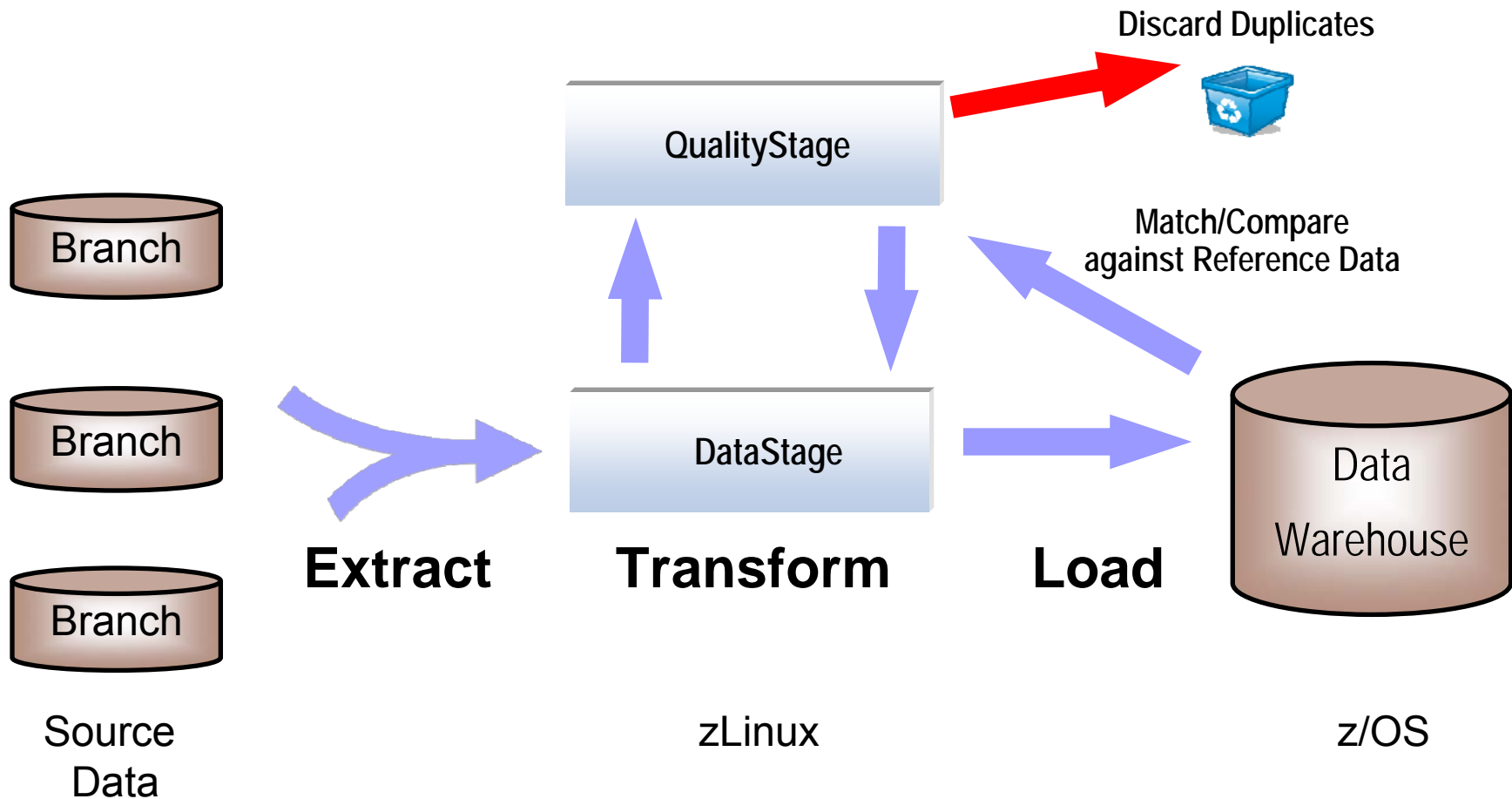
- ▶ Fields are compared
- ▶ Letter grade assigned
- ▶ Combined letter grades are compared to a vendor delivered file
- ▶ Result: Match; Fail; Suspect

Probabilistic Record Linkage:

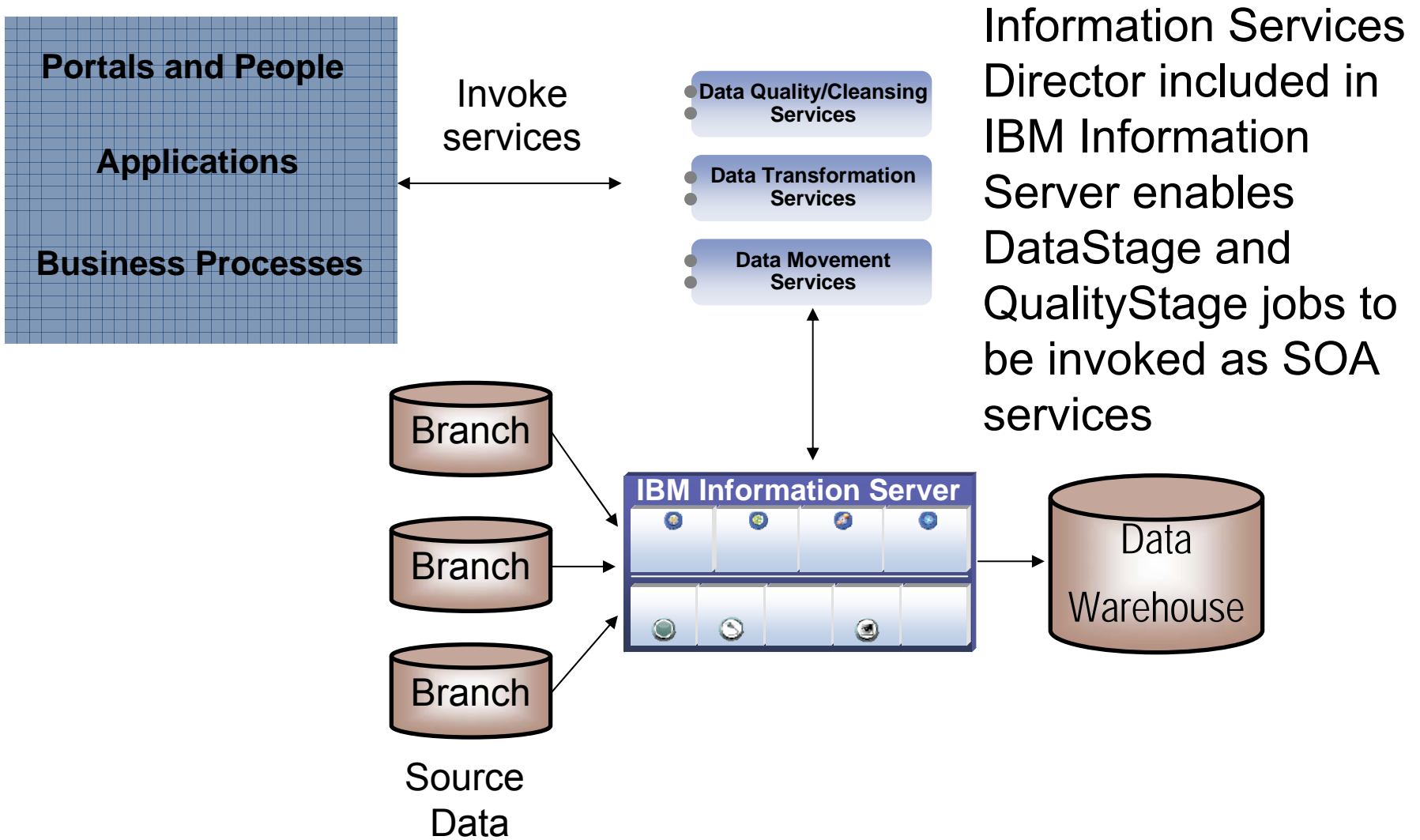
- ▶ Fields are evaluated for degree-of-match
- ▶ Weight assigned: represents the “**information content**” by value
- ▶ Weights are summed to derived a total score
- ▶ Result: Statistical probability of a match

Better Accuracy

Add QualityStage To Cleanse The Data

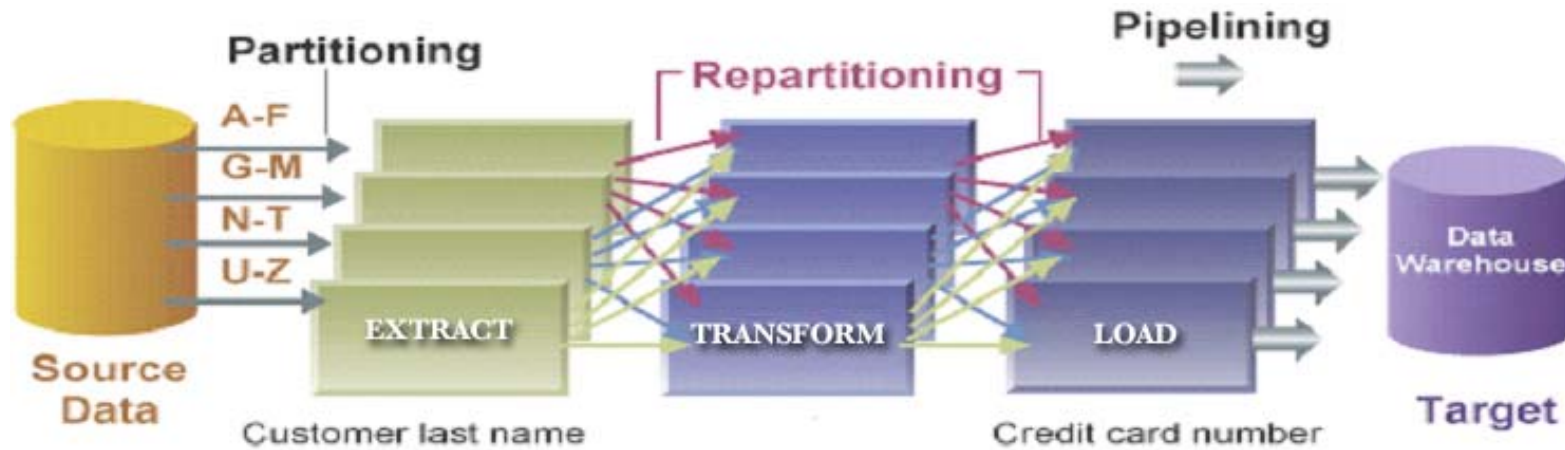


IBM Information Server Exposes Jobs As SOA Services



Information Services Director included in IBM Information Server enables DataStage and QualityStage jobs to be invoked as SOA services

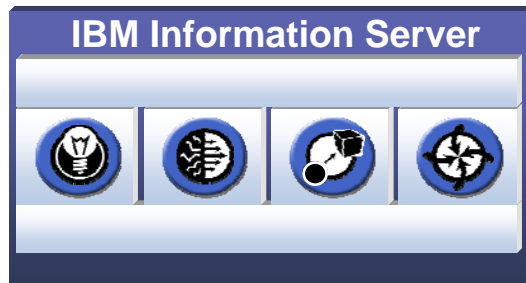
DataStage And QualityStage Utilize Parallel Processing Services For Extreme Scalability



- 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
- Benefits from processing capacity, I/O capacity, and Hipersockets on System z

IBM Beats The Competition In Data Warehouse Solutions

IBM



- Integrated data services platform
- Extensive connectivity
- Enterprise scalability
- Easily expose jobs as services

Oracle



- Two separate un-integrated products
- Warehouse Builder can only load Oracle databases
- Oracle Data Integrator (used to be Sunopsis) has no data quality capabilities
- Coding required to expose jobs as services

Microsoft



- Data quality limited to "Fuzzy Search" and MS SQL only
- Limited connectivity and limited support for non-Microsoft
- Lacks enterprise scalability
- Coding required to expose jobs as services

Service Oriented Finance Wants A *Dynamic* Data Warehouse

Our data is updated frequently,
We need the data in our data warehouse to be more current



Change Data Capture together with Information Server can “trickle feed” data into your data warehouse as it changes



IBM

IBM Information Server For System z Can Provide Near Real-Time Data Movement

IBM Information Server for System z

Understand



Information Analyzer
on Linux for System z

Business Glossary
on Linux for System z

Rational Data Architect

Cleanse



QualityStage on
Linux for System z

Transform



DataStage on Linux
on for System z
DataStage for z/OS

DataStage MVS

Deliver



Classic Fed. Server
Classic Replication
Replication Server
**Change Data Capture
for System z**

Platform Services

Parallel Processing Services



Connectivity Services



Metadata Services

Metadata Server

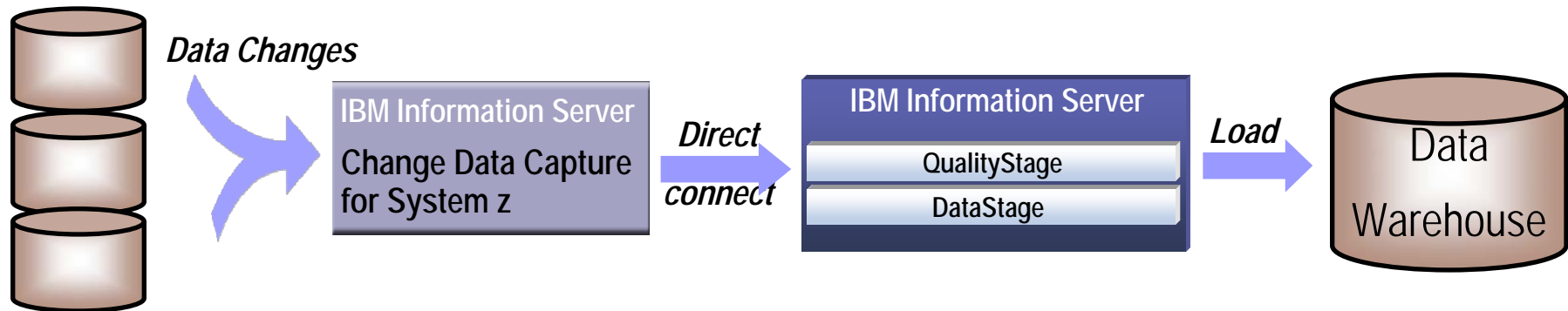
Administration Services



Deployment Services

Information
Services Director

Change Data Capture For System z



- Event is triggered automatically by a change in the data
 - ▶ Monitors DB2 log file for data changes
 - ▶ Sends message with data changes
- Message can initiate a business process
 - ▶ Example: After \$10,000 in sales, a service request is sent to WebSphere Information Services Director, invoking a DataStage job to load sales data for analysis
- 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

Log-Based Change Data Capture

- **Using existing native database recovery logs to capture table and row level activity in a relational database**
 - ▶ **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

- **Many advantages to log-based Change Data Capture**
 - ▶ No changes to existing applications or schemas required
 - ▶ Little performance impact to source application and system
 - 0.05% system resources required to process over 300 GB of data
 - Sends only the changes – efficient use of bandwidth
 - Scalable
 - ▶ Alternatives have many drawbacks
 - SQL Select, File Comparison, Database Triggers, Modifying Source Application

- **Wide range support**
 - ▶ Information Server CDC for - DB2 (all platforms), Oracle, SQL Server, Sybase, and more
 - ▶ Data Event Publisher for non relational sources

Where Should We Deploy This Solution?

Capability is important, but cost is a big concern for us



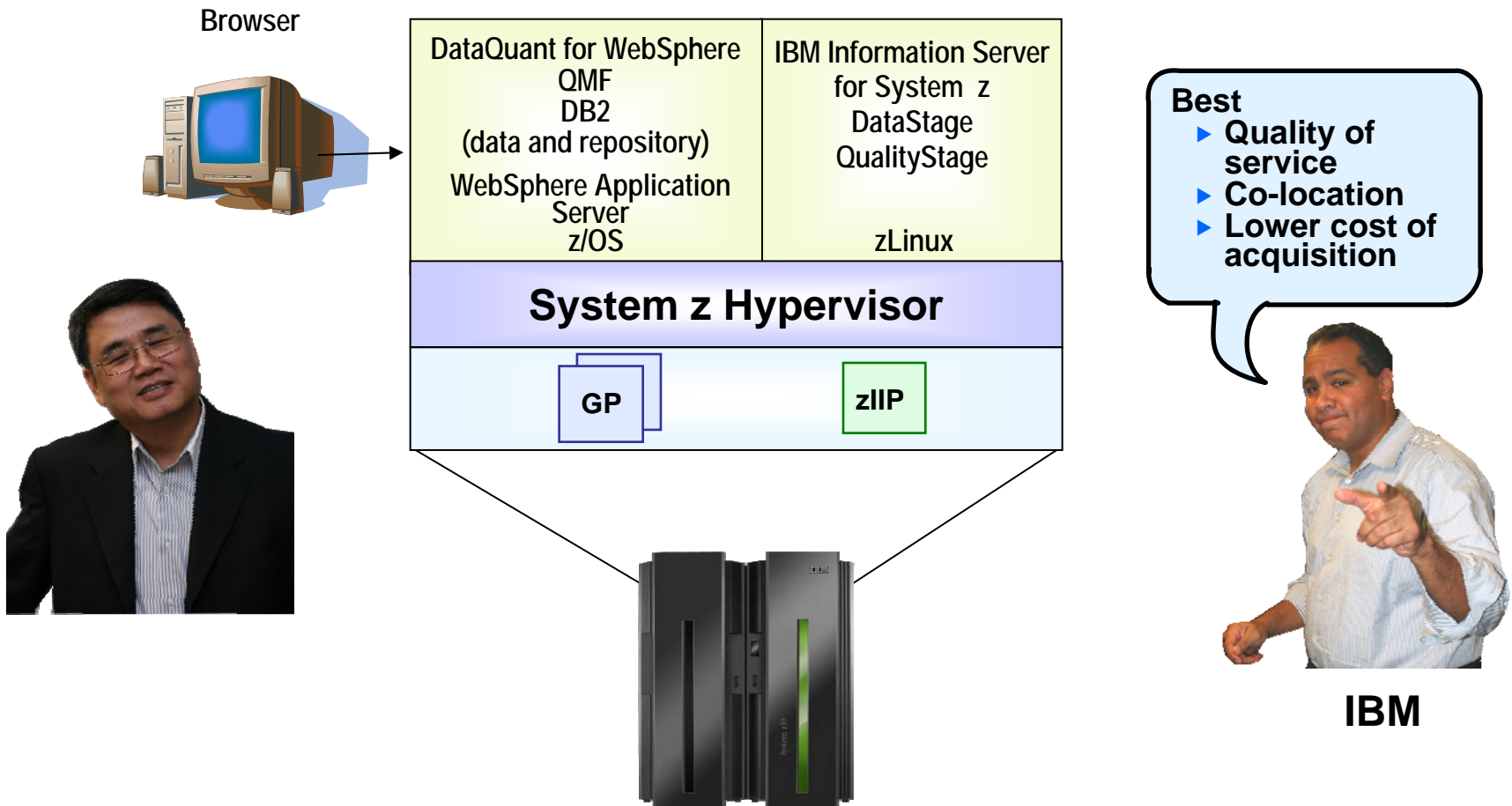
CIO

Deploying this solution on your System z will cost less than Oracle on distributed



IBM

Mainframe Extension Solution – Deliver Business Insight With A Data Warehouse



Storage Costs: DB2 Provides More Storage Savings Than Oracle

- DB2 for z/OS lowers TCO by reducing storage needed
 - ▶ TPC-H Benchmark: DB2 compression of 59% vs 29% for Oracle RAC
- Storage savings with DB2 vs. Oracle for a 10 TB data base

	Oracle	DB2 for z/OS*
Storage System	HP Enterprise Virtual Array 8100 Storage	IBM System Storage DS6800
Overall database compression ratio (using TPC-H benchmark results)	29%	59%
For 10 TB uncompressed data storage needed	7.5 TB of HP Storage	4.2 TB of IBM Storage
Cost of storage (3 year TCA)	$\$319,270 + \$15,113 \times 3$ = \$364,609	$\$234,101 + \$13,164 \times 2^{**}$ = \$260,429
With compression, storage for DB2 costs <u>29% less</u> than for Oracle		

*DB2 for z/OS achieves similar compression ratios to those of DB2 for LUW

**IBM storage maintenance fee for the first year is included in the warranty

US Retailer Improves Response Time By Co-locating Data Warehouse And Operational Data

- A major US retailer moved their 5.5TB data warehouse from distributed servers to System z
 - ▶ Operational data bases were already located on System z
- On average they reduced query processing times by 80% due to better query parallelism in DB2 for z/OS
 - ▶ (17 minutes to 3 minutes)
- They saved CPU cycles to load the data warehouse
 - ▶ Avoided network processing

The Value Of Business Insight

We learned the best promotion to maximize our business profits



Marketing VP

And I saved money by deploying our data warehouse solution on System z



CIO

Accelerate Development With Pre-built Data Models For Industry Data Warehouses

- Quick start your data warehouse design with pre-built **IBM Industry Data Models**
- Use **Enterprise Model Extender** to customize the data model
 - ▶ Eclipse plug-in for Rational Data Architect
- Models capture the best practices of over 400 IBM customers
 - ▶ Banking, finance, health, insurance, retail and Telco



