



IBM Information On Demand

Unlock The Business Value of Information

Building A Trusted View
Of Your Data With Information Server

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.



**Mortgage Line of Business
VP**

You need to consolidate your information in a Data Warehouse.

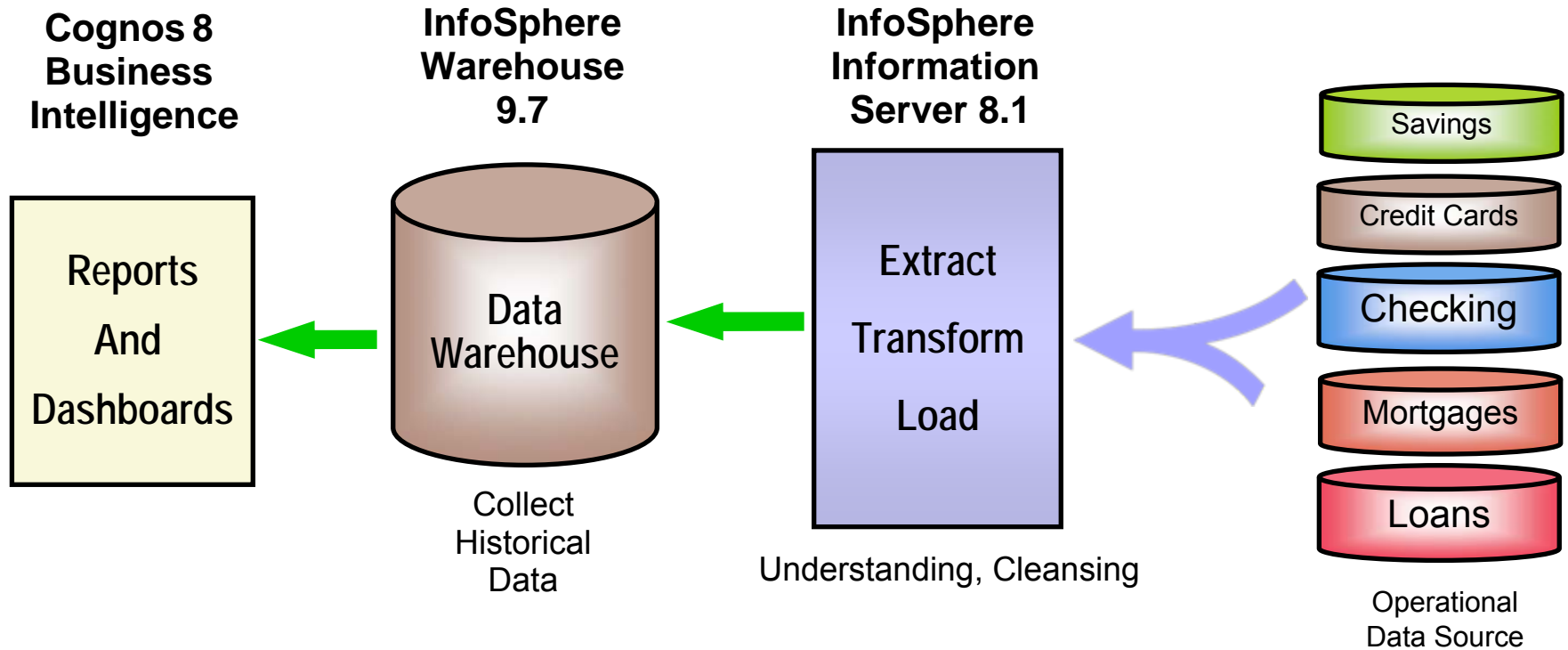


**Service Oriented Finance
CIO**

Service Oriented Finance Needs A Complete View Of Their Customers

- **Problem:** Segmented Lines of Business store their data in isolated silos
 - ▶ Savings, Checking, Mortgage, Consumer Loans
- **Problem:** Identify customer trends that span Lines of Business
- **Solution:** Create new intelligence by building a corporate data warehouse, where employees can analyze data in a holistic fashion

Create New Intelligence With IBM Information Management Software



Analysis Tools

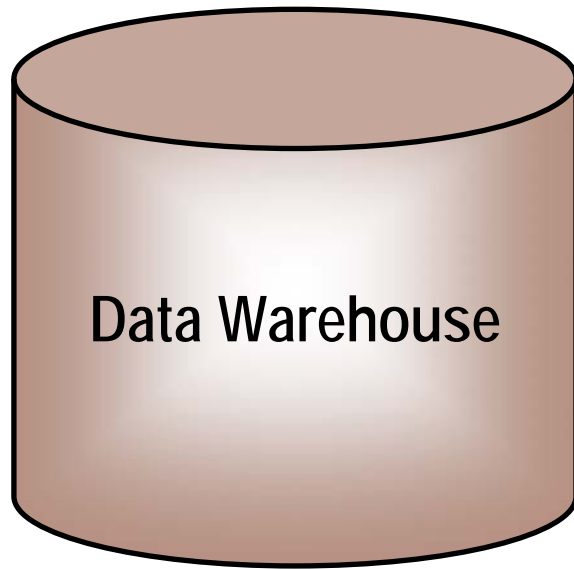
ETL Processing

Scalable Database

Connections

IBM InfoSphere Warehouse 9.7

InfoSphere Warehouse 9.7



Microsoft SQL Server is Limited by the scalability of Windows on Intel

- Use InfoSphere Data Warehouse to build your data warehouse
- Built on DB2 9.5
- Capabilities include:
 - ▶ Data partitioning (enables scale out)
 - ▶ Workload management
 - ▶ Data Compression to save storage
 - ▶ Basic analysis support
 - ▶ Data Studio Tooling based on Eclipse
- Runs on AIX, Redhat, SuSE, Windows, HP-UX, Solaris and System z

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

IBM 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 Can Load Your Data Warehouse

IBM InfoSphere Information Server

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

InfoSphere Metadata Server

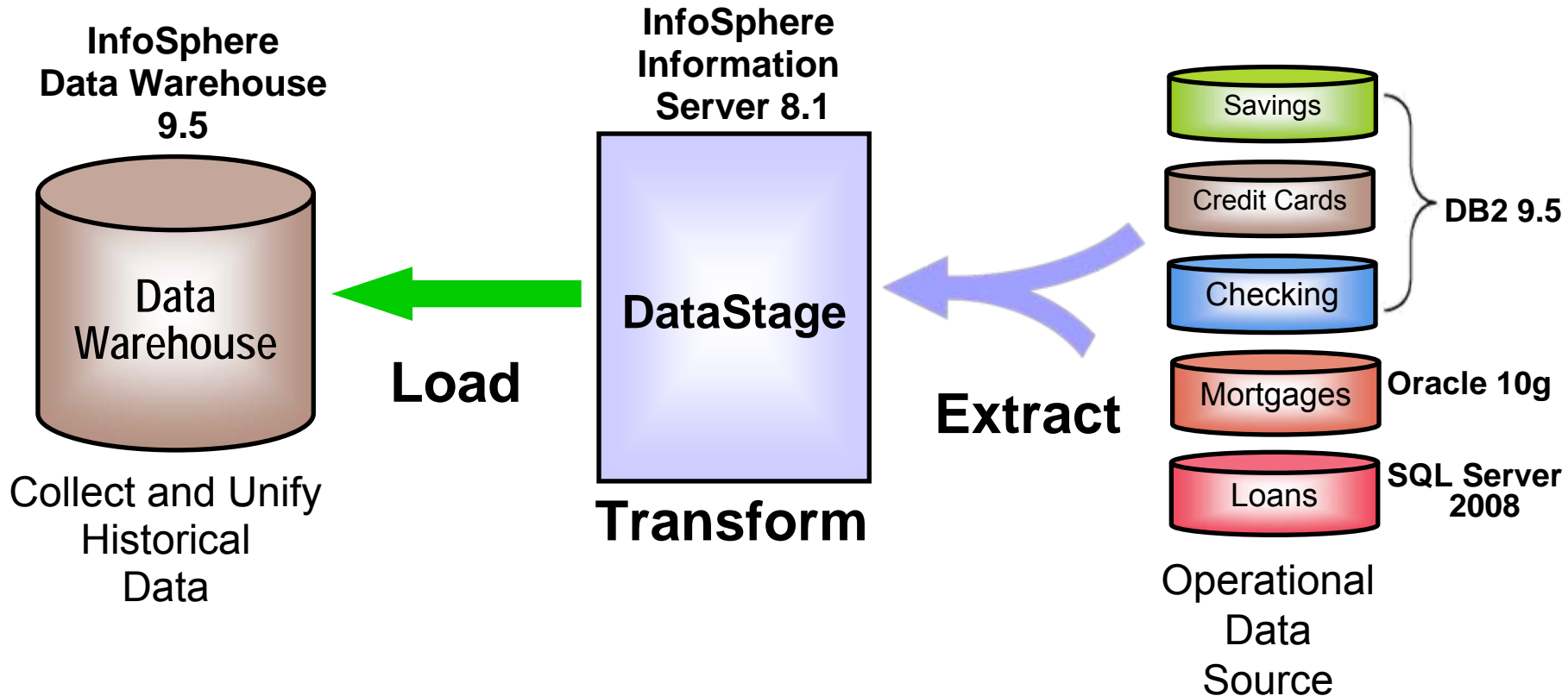
Administration Services



Deployment Services

InfoSphere
Information Services Director

Loading Your Data Warehouse With InfoSphere DataStage



InfoSphere DataStage Transforms Data On The Fly

- Different field names
- Different field order
- Add source Identifier
- Different field formats



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

Warehouse

000 101 100



Transform

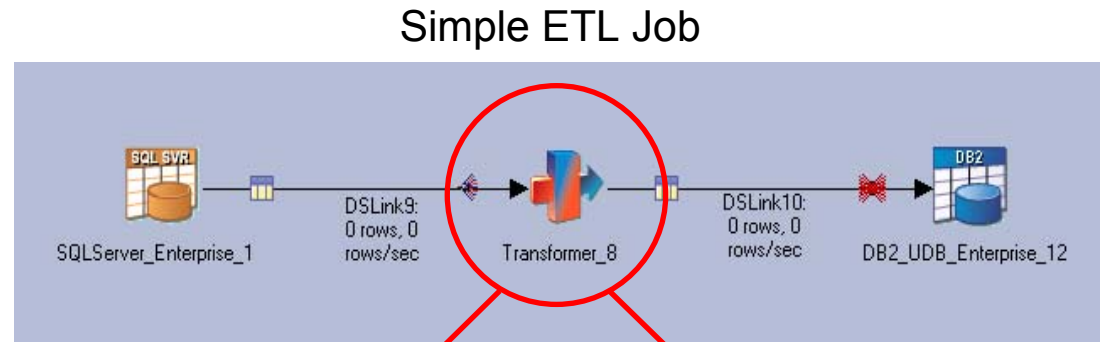
01 01 \$20,500.00 2007-02-28

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

Source Data

InfoSphere DataStage Creates Executable Extract, Transform And Load (ETL) Jobs

- Visually define extract, transform, and load jobs to load the data warehouse
- Jobs may be created from scratch or imported from FastTrack tool
 - ▶ Data Maps



Data Map
In Transform Node

DSLInk9	
STATE	
ZIP	
ACCOUNT_ID	
JOINT_ACCOUNT_HOLDER	
ACCOUNT_BRANCH	
BALANCE	
PAST_DUE_COUNT	
SOURCE	
LOAN_DATE	

DSLInk10	
Constraint:	
Derivation	Column Name
DSLInk9.STATE	STATE
DSLInk9.ZIP	ZIP
DSLInk9.ACCOUNT_ID	ACCOUNT_ID
DSLInk9.JOINT_ACCOUNT_HOLDER	JOINT_ACCOUNT_HO
DSLInk9.ACCOUNT_BRANCH	ACCOUNT_BRANCH
DSLInk9.BALANCE	BALANCE
DSLInk9.PAST_DUE_COUNT	PAST_DUE_COUNT
DSLInk9.SOURCE	SOURCE
(TimestampToDate[DSLInk9.LOAN_DATE])	LOAN_DATE

FastTrack Tool Creates Data Maps For Transform Nodes

- Solves a key problem of associating business terms with field names
 - ▶ Improve understanding
 - ▶ Reduce mistakes
 - ▶ Search for relevant fields using business glossary terms
 - ▶ “Tagging” for data fields
- Visual tool to create the data maps
- Publish data maps to DataStage Designer
- Oracle and Microsoft don’t address these ETL development steps
- Oracle nor Microsoft have any tool to help manage sources to get mappings

InfoSphere Business Glossary “Tags”
Find Fields Tagged as “**CUSTOMER**”



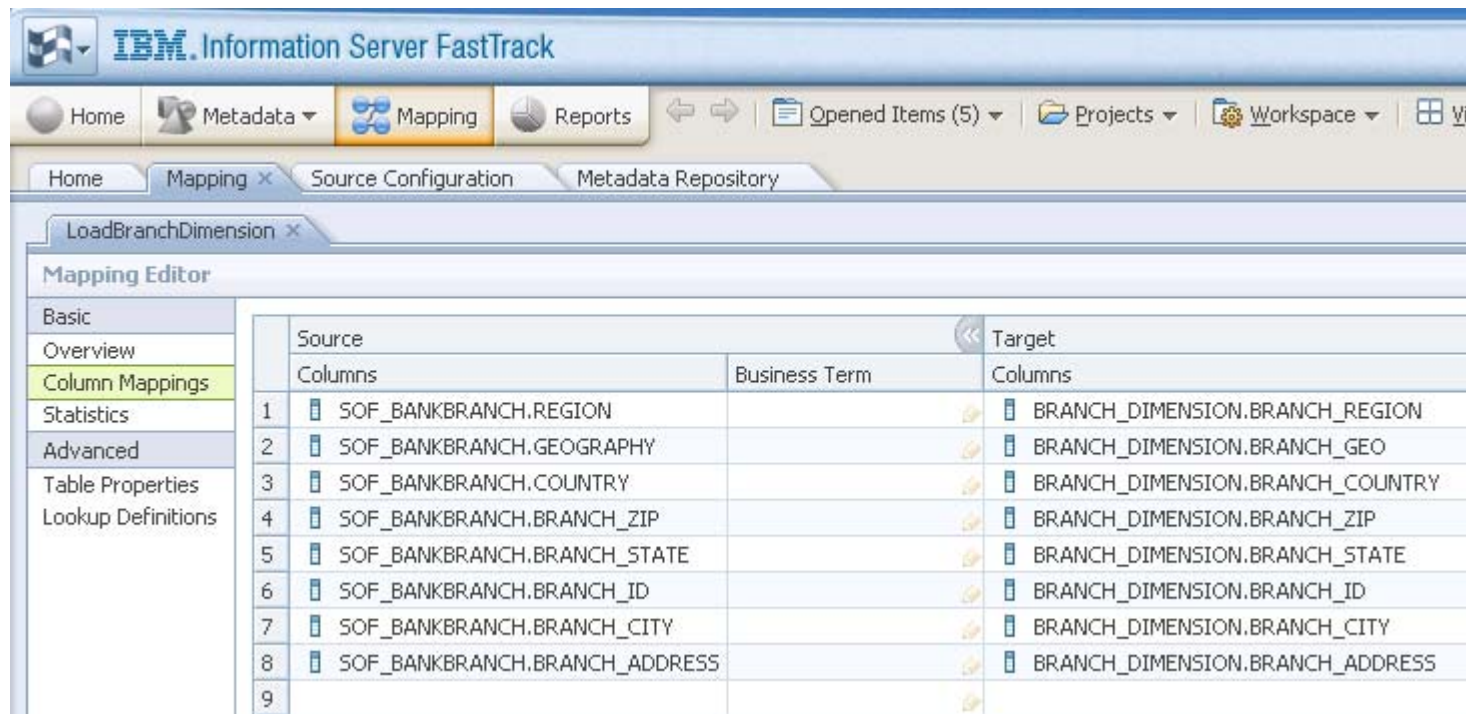
Business Glossary Tag	Column	Table
CUSTOMER	CUSTID	SOF_Banking.Accounts
CUSTOMER	CID	SOF_Loan.Borrower
CUSTOMER	CNUM	SOF_Mortgage.Borrower



Select correct source
for the data map

DEMO 1: Use FastTrack To Define Column Mappings

1. Quickly create source-to-target mappings in FastTrack
2. Use terms from Business Glossary to suggest possible mappings
3. Generate ETL Job for DataStage



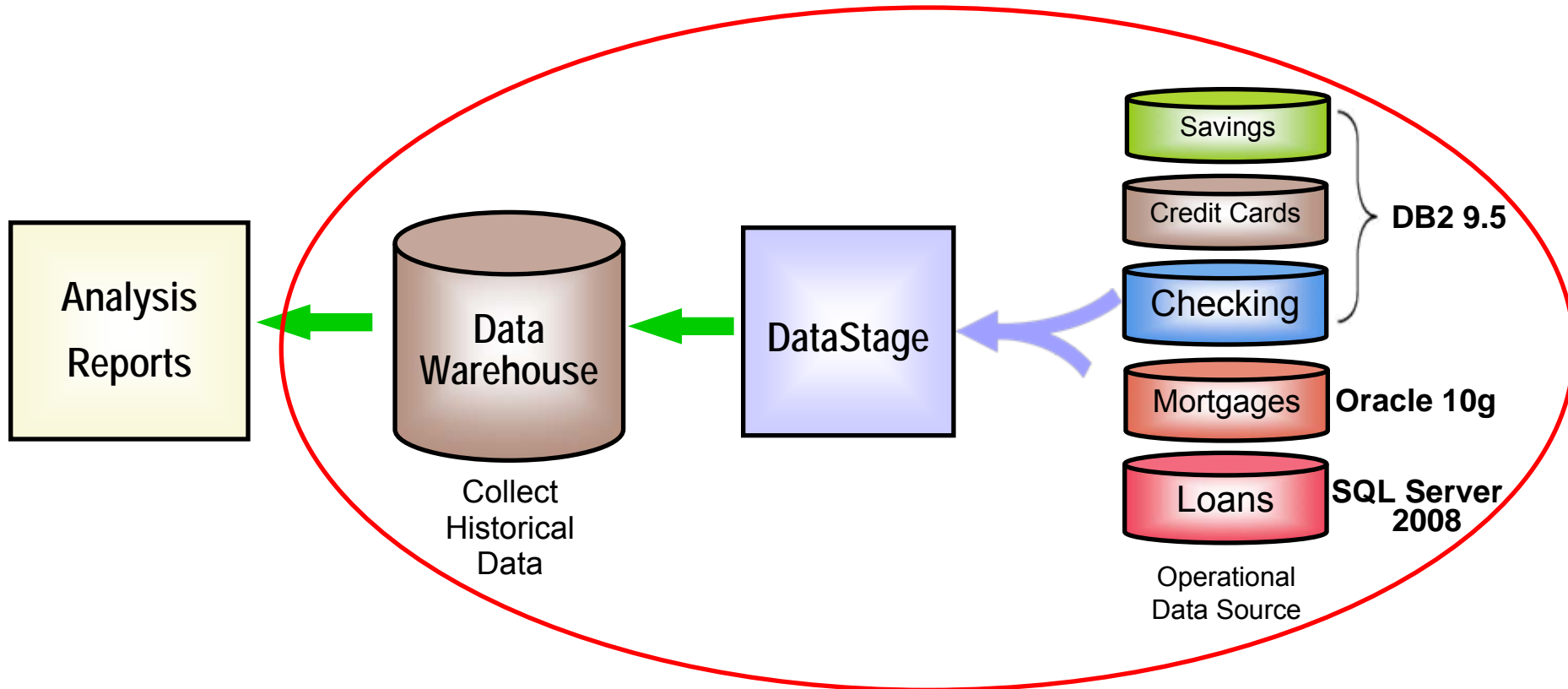
The screenshot displays the IBM Information Server FastTrack Mapping Editor interface. The window title is "IBM Information Server FastTrack". The main menu includes Home, Metadata, Mapping (selected), and Reports. Below the menu, there are tabs for Home, Mapping (selected), Source Configuration, and Metadata Repository. The current workspace is titled "LoadBranchDimension".

The Mapping Editor is open, showing a table with columns for Source, Business Term, and Target. The Source column is further divided into Columns and Business Term. The Target column is divided into Columns. The table lists 8 mappings, each with a row number, a source column name, a target column name, and a small icon indicating the mapping status.

	Source	Business Term	Target
	Columns		Columns
1	SOF_BANKBRANCH.REGION		BRANCH_DIMENSION.BRANCH_REGION
2	SOF_BANKBRANCH.GEOGRAPHY		BRANCH_DIMENSION.BRANCH_GEO
3	SOF_BANKBRANCH.COUNTRY		BRANCH_DIMENSION.BRANCH_COUNTRY
4	SOF_BANKBRANCH.BRANCH_ZIP		BRANCH_DIMENSION.BRANCH_ZIP
5	SOF_BANKBRANCH.BRANCH_STATE		BRANCH_DIMENSION.BRANCH_STATE
6	SOF_BANKBRANCH.BRANCH_ID		BRANCH_DIMENSION.BRANCH_ID
7	SOF_BANKBRANCH.BRANCH_CITY		BRANCH_DIMENSION.BRANCH_CITY
8	SOF_BANKBRANCH.BRANCH_ADDRESS		BRANCH_DIMENSION.BRANCH_ADDRESS
9			

DEMO 2: Use DataStage To Load The Data Warehouse

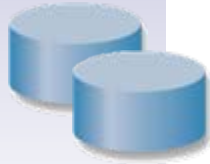
1. Execute DataStage Job to populate data warehouse fact table
2. Verify Data Load for Data Warehouse



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
Nettezza
Progress
RDB
RedBrick
SQL/DS
SQL Server
Sybase (ASE and IQ)
Teradata
Universe
UniData
NonStopSQL
And more.....



Offering more
connectivity
than Microsoft
or Oracle

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

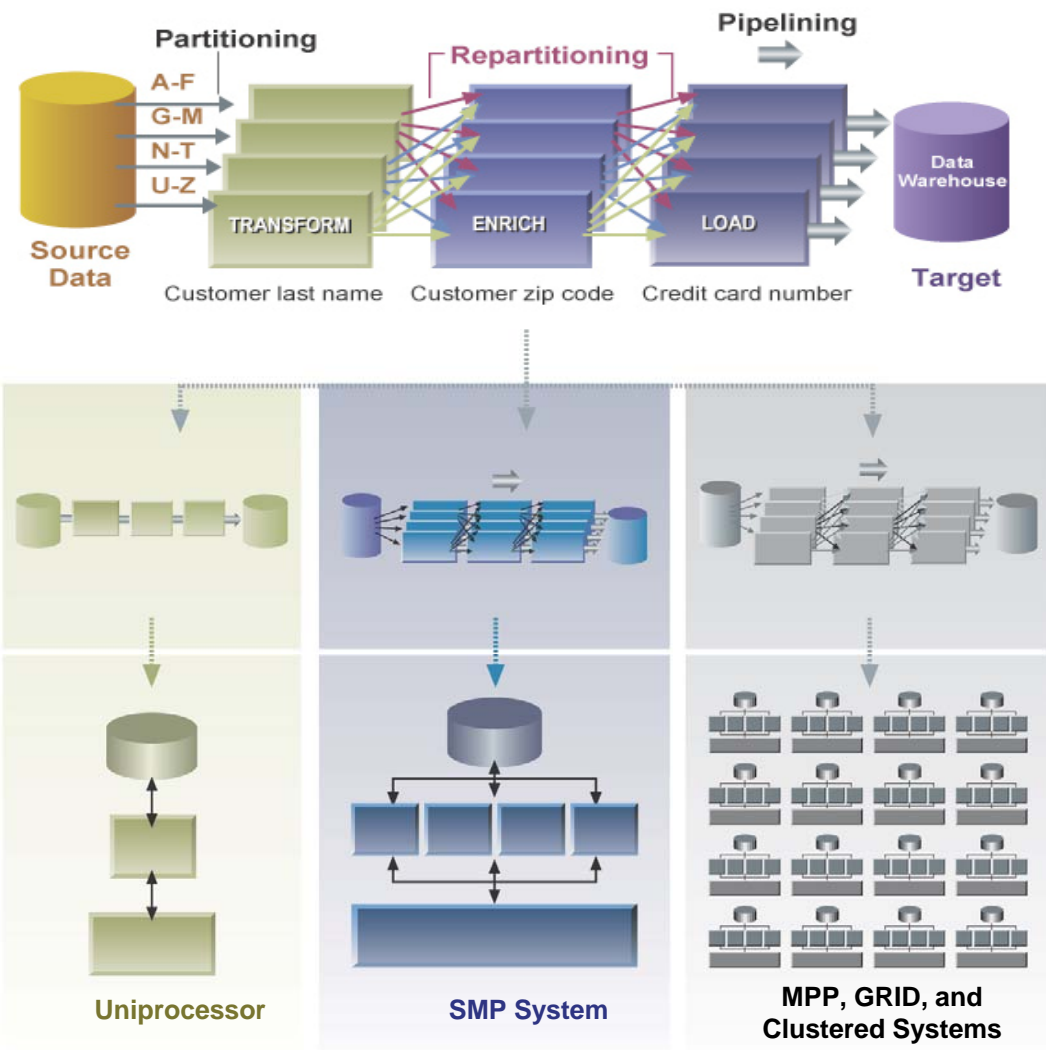
DB2 (on z, I, P, X series)
Oracle
SQL Server
Sybase
Informix
IMS
VSAM
ADABAS
IDMS
NonStopSQL
Enscribe

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....



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

Orange Telecom Uses InfoSphere Information Server

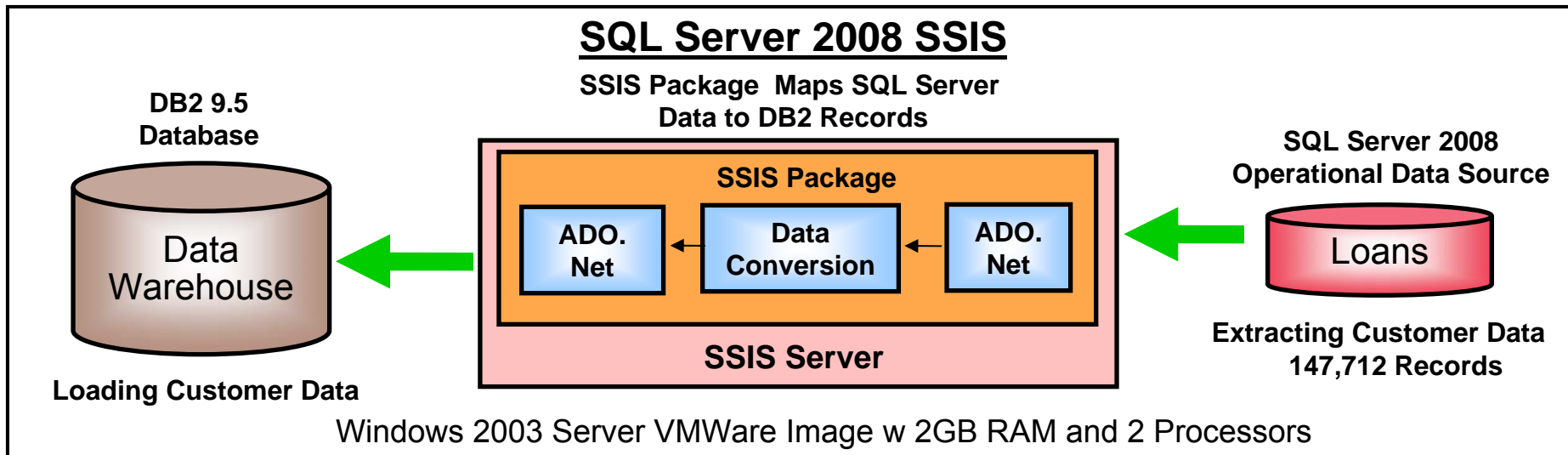
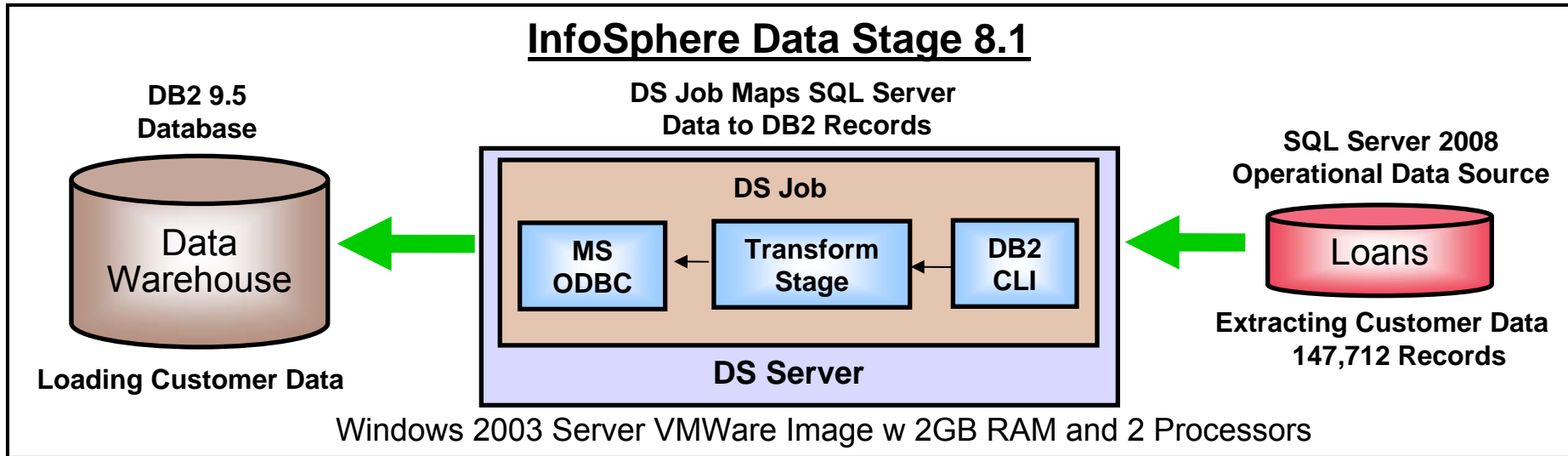
- Convergence is the big trend with telecom providers delivering voice, data and multimedia services in a single package. But what telecom companies are finding is that the success of converged services often referred to as “tripleplay” based on the number of products combined—can depend on the company’s information management strategy.”
- Overall, the company needed to gain a single view of customer information across all business units to lower operational costs and increase service quality for the customer.



“IBM Information Server will enable our development staff to test new concepts and migrate applications to new platforms without impacting our end users. This provides us with tremendous flexibility when delivering new services.”

- Françoise Gesbert, Applications Metadata Manager,
Orange France Telecom Group

DEMO 3: Compare InfoSphere DataStage 8.1 to MS SQL Server 2008 Integration Services



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?



**Service Oriented Finance
CIO**

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

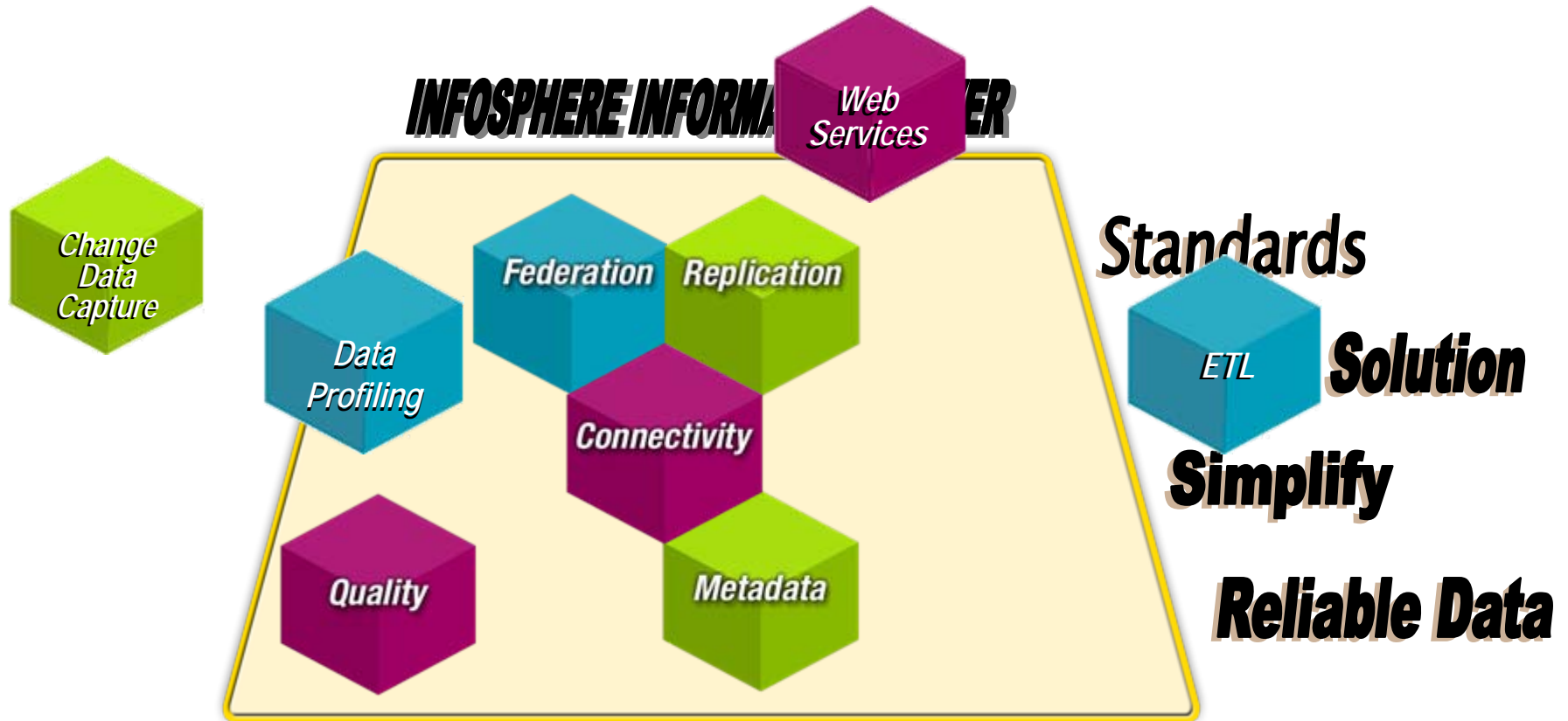


IBM

SOF Technical Issues That Inhibit Trusted Data

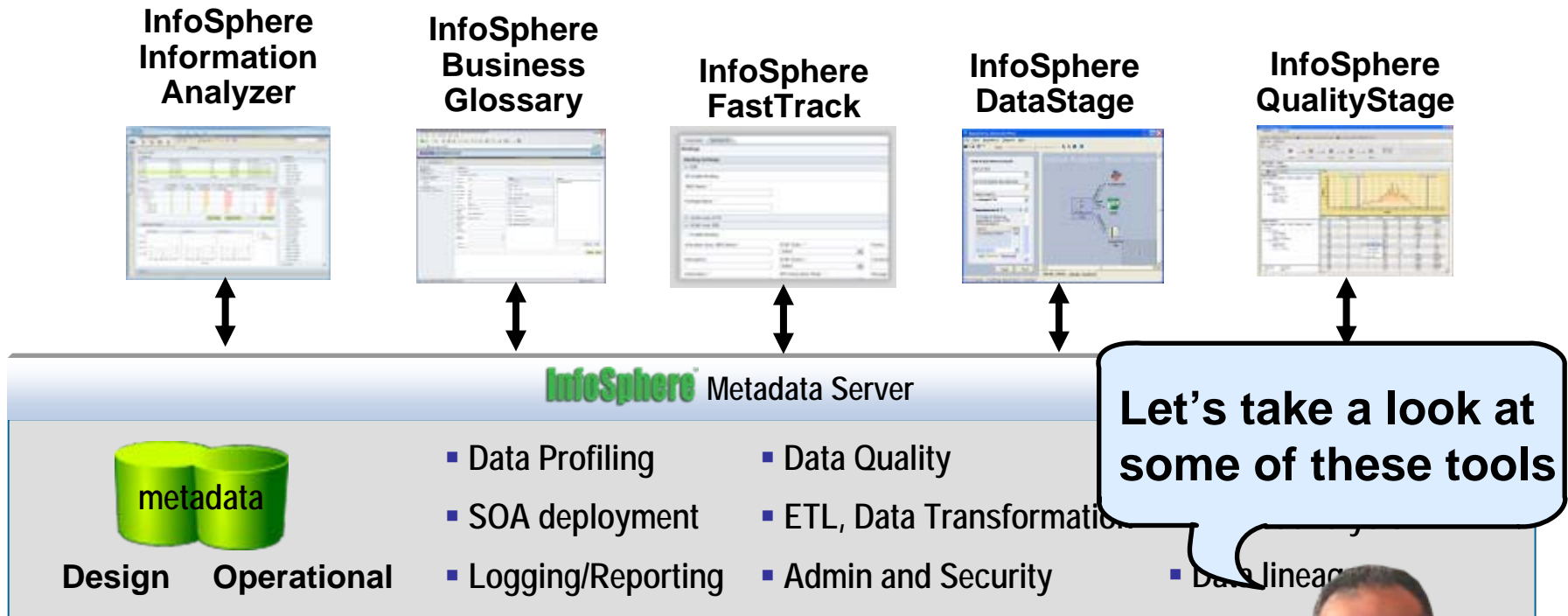
- Problem: Multiple Data Sources
 - ▶ Impact: Too much information coming from multiple systems, difficulty utilizing data
- Problem: Different Data Formats
 - ▶ Impact: Multiple versions of the truth,
- Problem: Understanding data and enabling reuse throughout different tools
 - ▶ Impact: Lack of business agility, inability to take advantage of opportunities
- Problem: Small time windows require high performance processing
 - ▶ Impact: Incomplete, out-of-date, misinterpreted data
- Problem: 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



Oracle has no integration of metadata across products:
Oracle Warehouse Builder, Oracle Data Integrator and
3rd Party Trillium. Requires manual import and export

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

What Else Can InfoSphere Information Server Do For You?

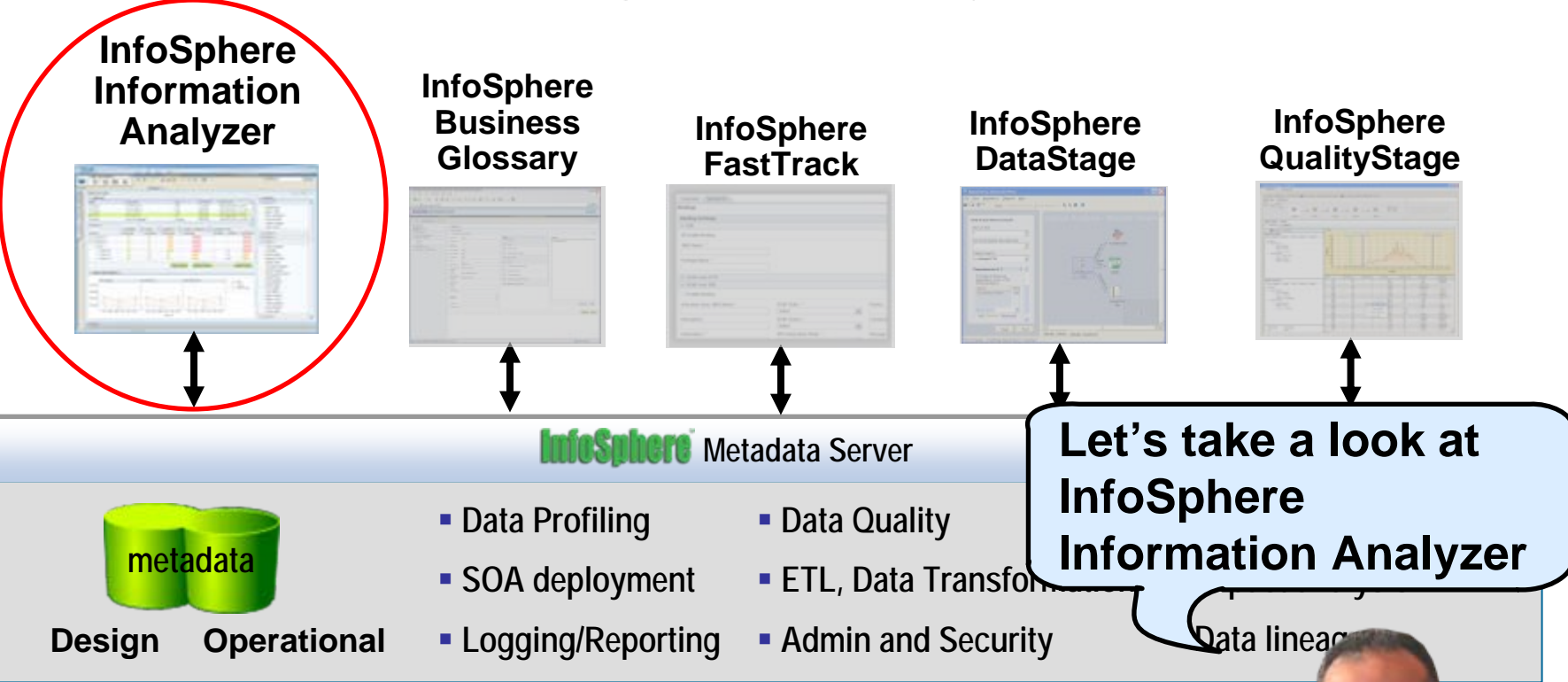
We can also help you discover and cleanse your existing data.



IBM

InfoSphere Information Server Supports Metadata

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



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

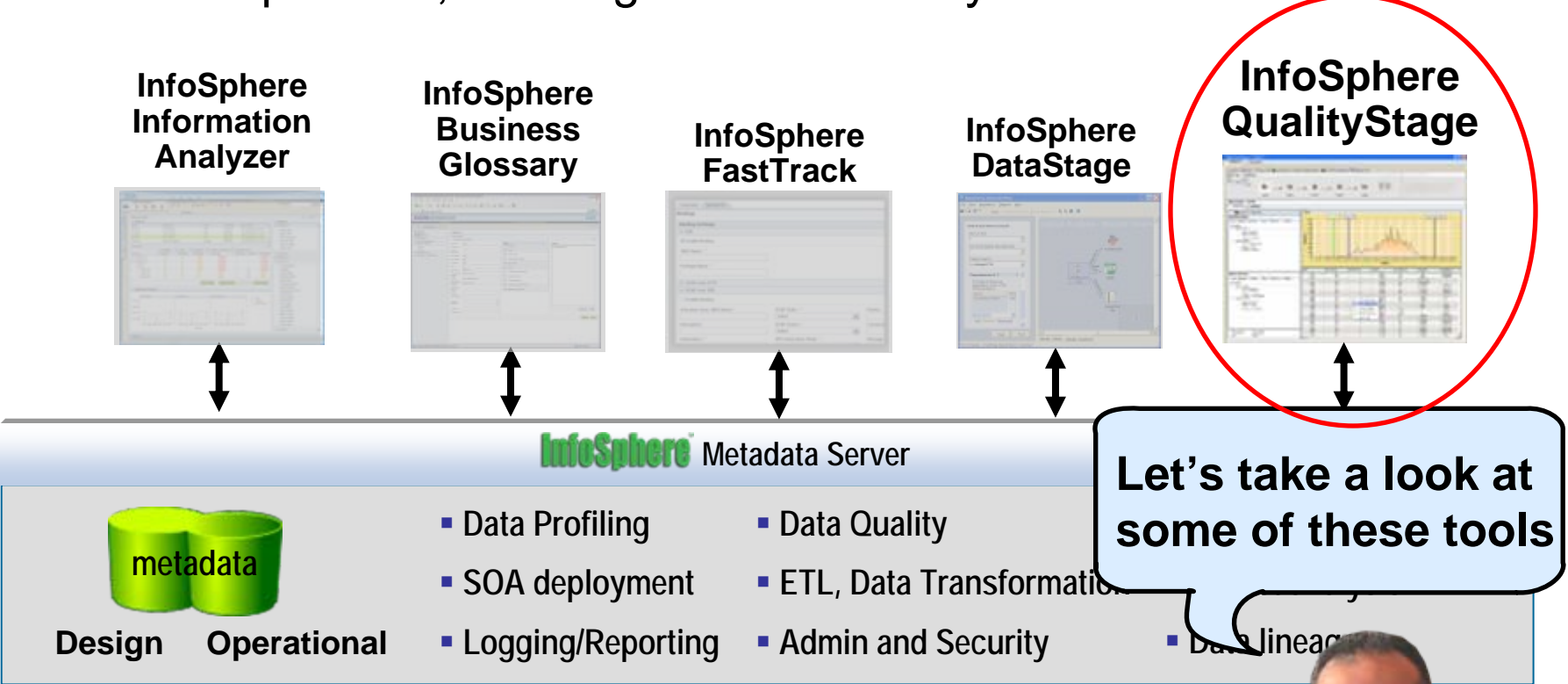
What Can InfoSphere Information Analyzer Do For You?

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 Server Supports Metadata

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



Why Should I Care About Cleansing Information?



25% of time is spent clarifying bad data

Low data quality costs companies \$611 billion annually

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

Scrap and rework
Increased costs



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)
 - ▶ Data quality processes can be deployed within extract, transform, and load (ETL) jobs or separately as shared data quality services (SOA)



Microsoft only provides limited fuzzy search
Oracle relies on third parties (Trillium)

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 = BBAABDBA

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

The Competition: Deterministic Decisions Tables:

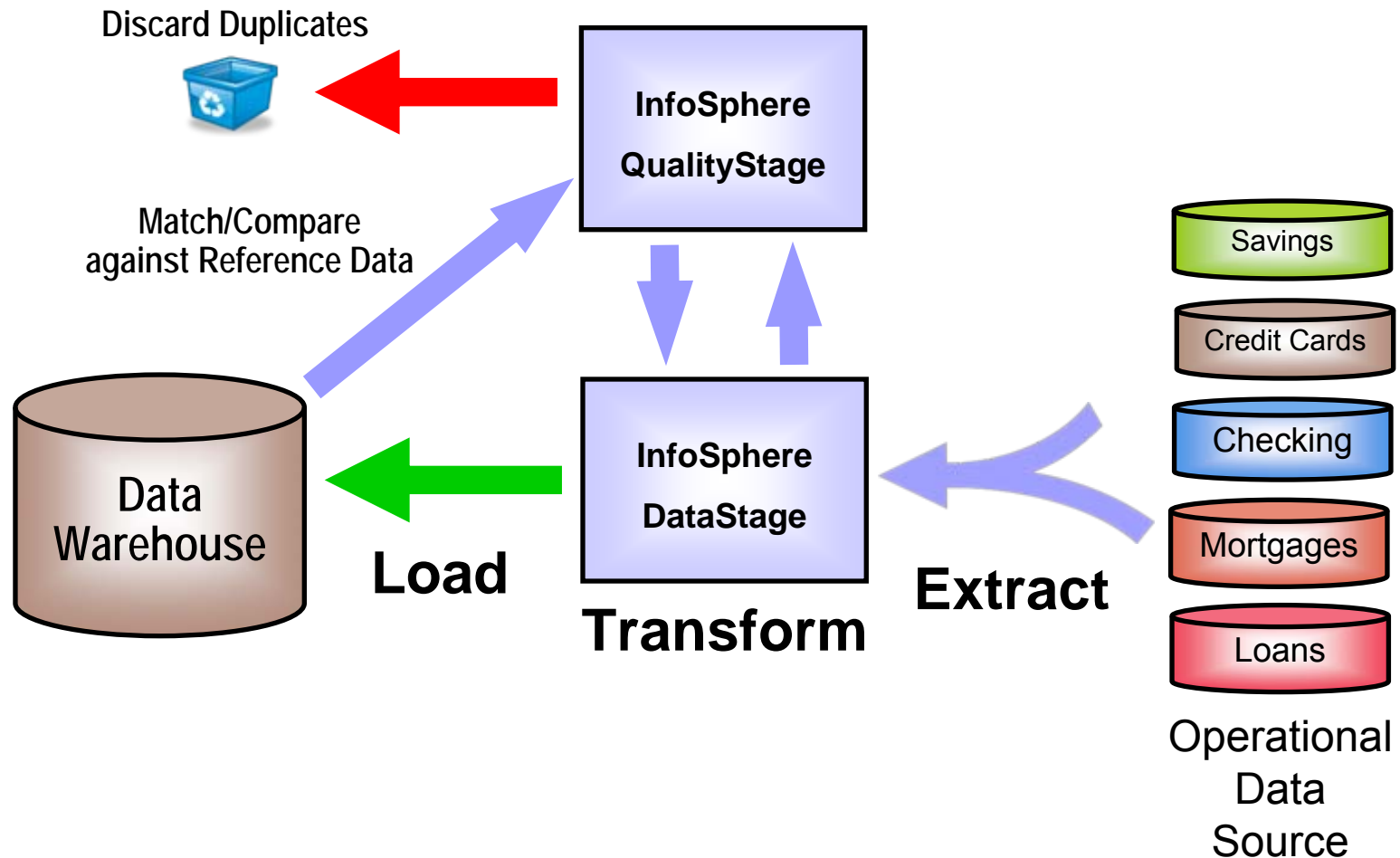
- ▶ Fields are compared
- ▶ Letter grade assigned
- ▶ Combined letter grades are compared to a vendor delivered file
- ▶ Result: Match; Fail; Suspect

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

Loading Your Data Warehouse with InfoSphere FastTrack And DataStage



Service Oriented Finance Wants A Dynamic Data Warehouse

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



**Service Oriented Finance
CIO**

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

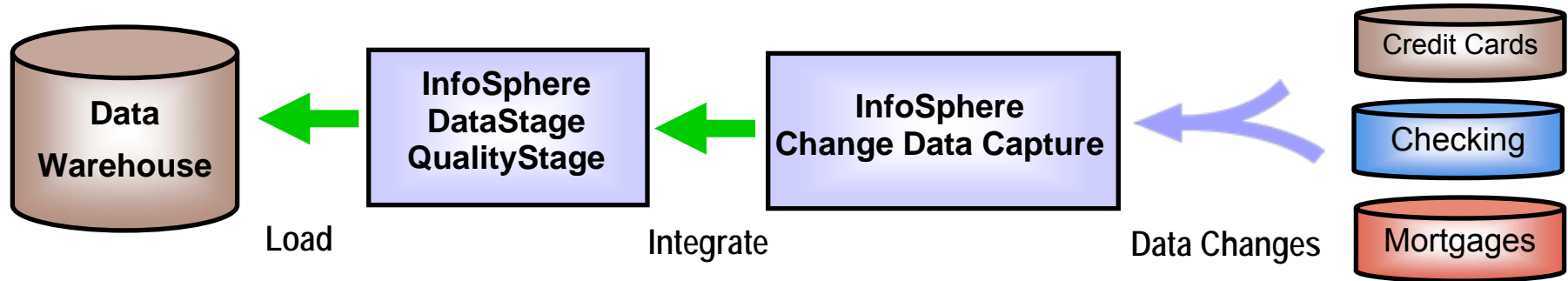


IBM

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

InfoSphere Change Data Capture and Information Server Enable Dynamic Data Warehousing



- 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

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 MS nor Oracle have any tool to help manage source to target mappings
- Integration (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
 - ▶ MS only has “Fuzzy Search” Oracle Data Integrator (Sunopsis) uses a 3rd party product (Trillium) for data quality and profiling.

