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.

You need to consolidate your information in a Data Warehouse.
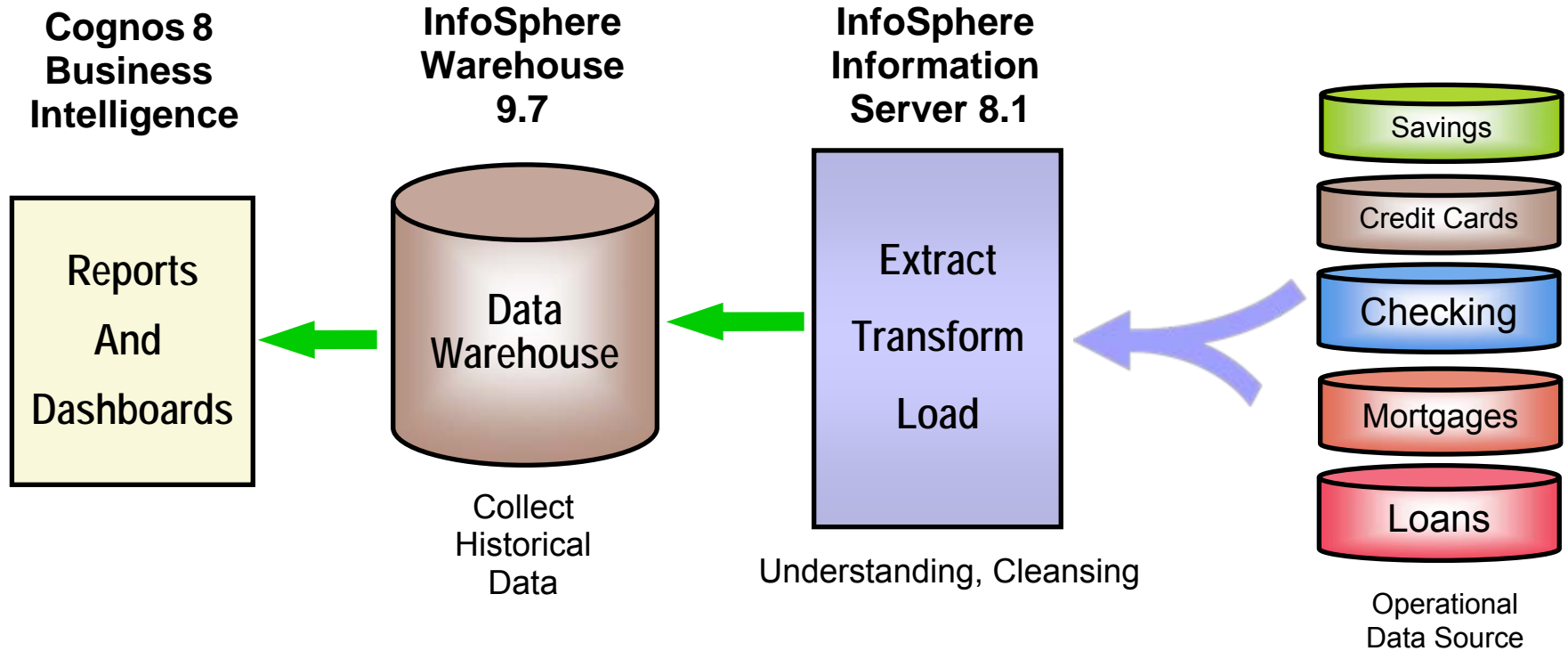
Mortgage Line of Business VP

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

- **<u>Solution</u>**: 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

Reports And Dashboards

Data Warehouse

Collect Historical Data

Extract Transform Load

Understanding, Cleansing

Savings

Credit Cards

Checking

Mortgages

Loans

Operational Data Source

Analysis Tools

ETL Processing

Scalable Database

Connections

InfoSphere Warehouse 9.7 Is Purpose Built For Data Warehouses

InfoSphere Warehouse 9.7
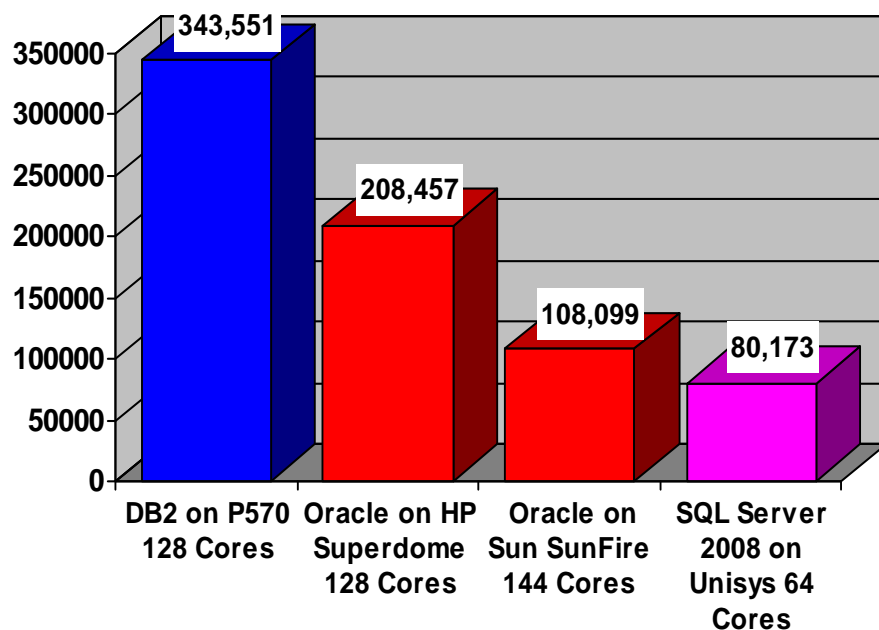
Data Warehouse
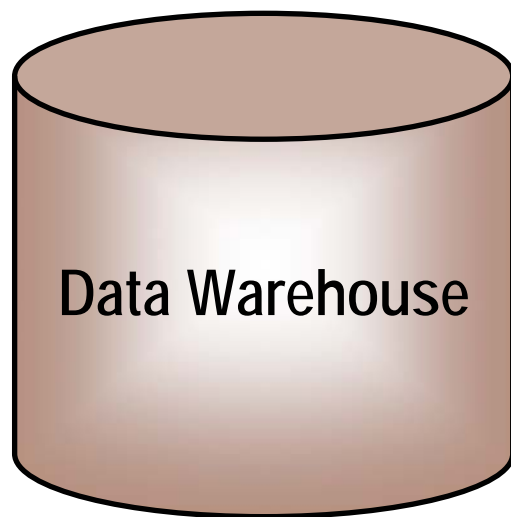
Microsoft SQL Server is limited by the scalability of Windows on Intel

- 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 Performance
QphH @ 10000GB
(higher is better)

	QphH
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.

IBM industry data models can help you get started quickly.

Service Oriented Finance CIO

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 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

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

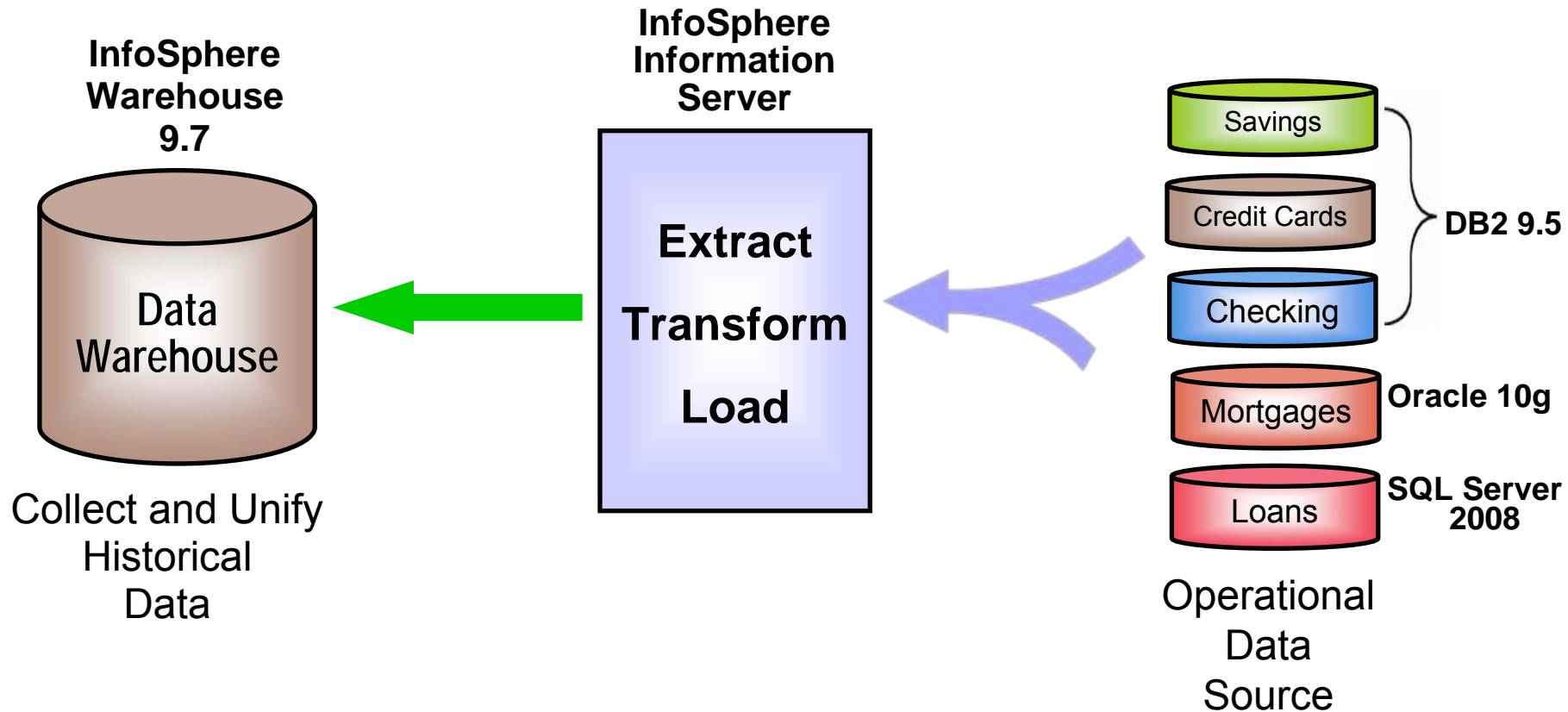**Metadata Services**
InfoSphere Metadata Server

Administration Services

Deployment Services
InfoSphere Information Services Director

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

DB2 9.5

Checking

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 ID	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   1 $10,000.00   2007-02-28
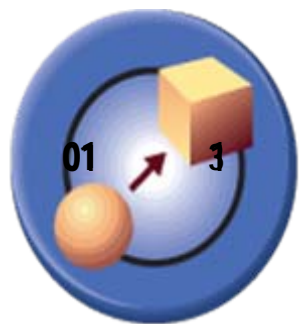
Transform

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 formats 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

DEMO: Use InfoSphere FastTrack To Create ETL Specification For Warehouse

- Use discovery feature to find source columns matching business term tags
- Generate ETL job for InfoSphere DataStage

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
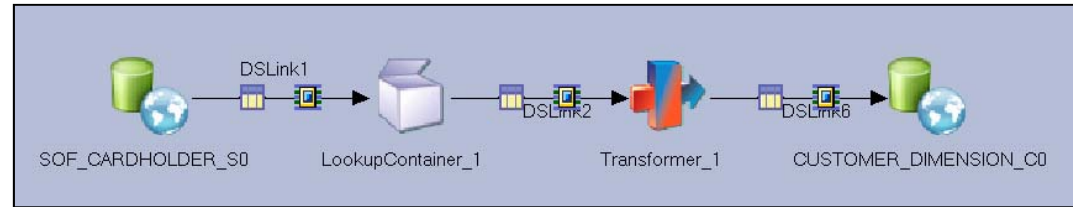
Database

Classic Federation	DB2 UDB API	DB2/UDB Enterprise	DB2Z		Dynamic RDBMS
Informix CLI	iWay Enterprise	ODBC	Oracle Enterprise		Stored Procedure

Processing

Aggregator	Change Apply	Change Capture	Compare
Compress	Copy	Decode	Difference
Encode	Expand	External Filter	Filter
FTP Enterprise	Funnel	Generic	Join
Lookup	Merge	Modify	Pivot
Remove Duplicates	Slowly Changing Dimension	Sort	Surrogate Key Generator
Switch	Transformer		

Real Time

Java Client	Java Transformer	Web Services Client	Web Services Transformer
WebSphere MQ Connector	WISD Input	WISD Output	XML Input
XML Output	XML Transformer		

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…..

Offering more connectivity than Oracle and Microsoft

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….

DEMO: Use InfoSphere DataStage To Load The Data Warehouse

1. **Show the DataStage ETL job generated by FastTrack**
2. **Run the DataStage Job to populate the data warehouse fact table**
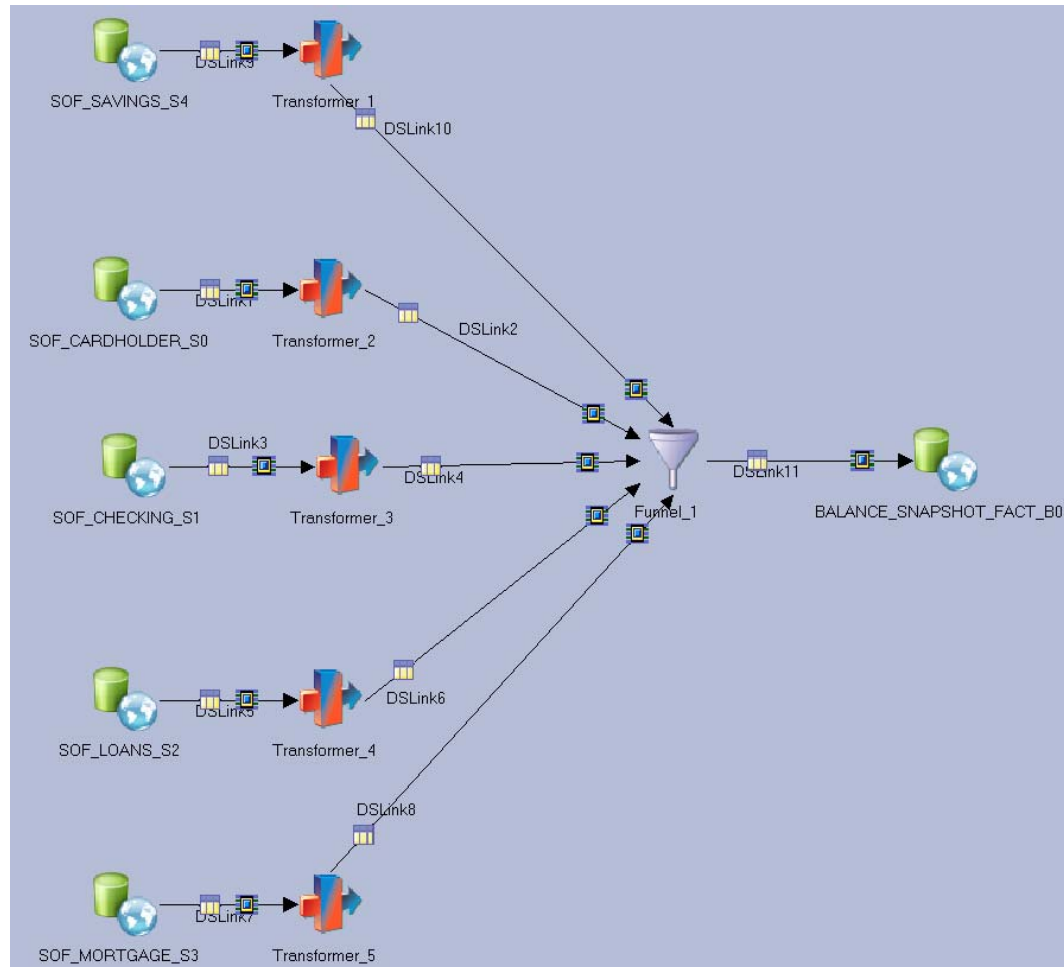
InfoSphere DataStage Uses Parallel Processing For Enterprise Scalability
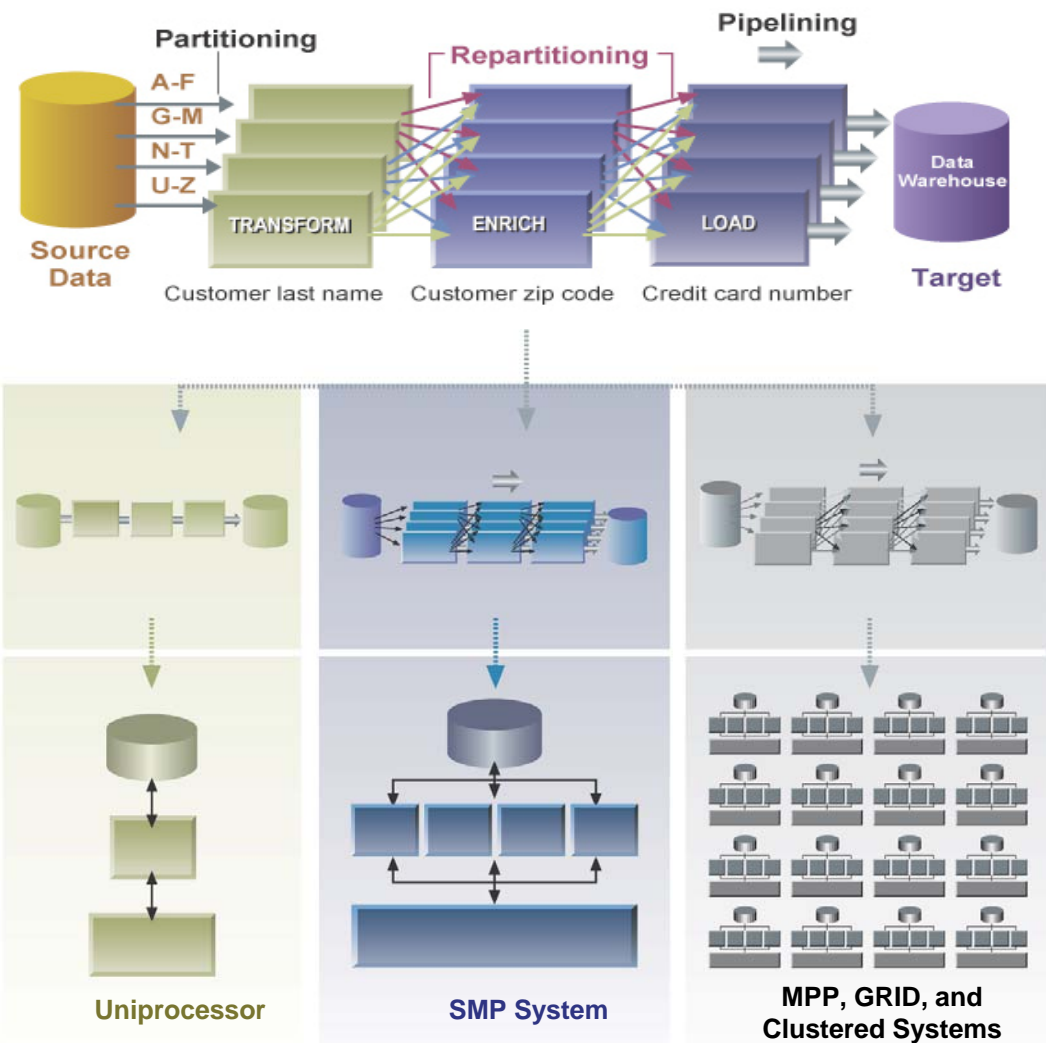
- **Supports major scalability options**
  - ▶ SMP, Clustered, GRID, and MPP platforms

- **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

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
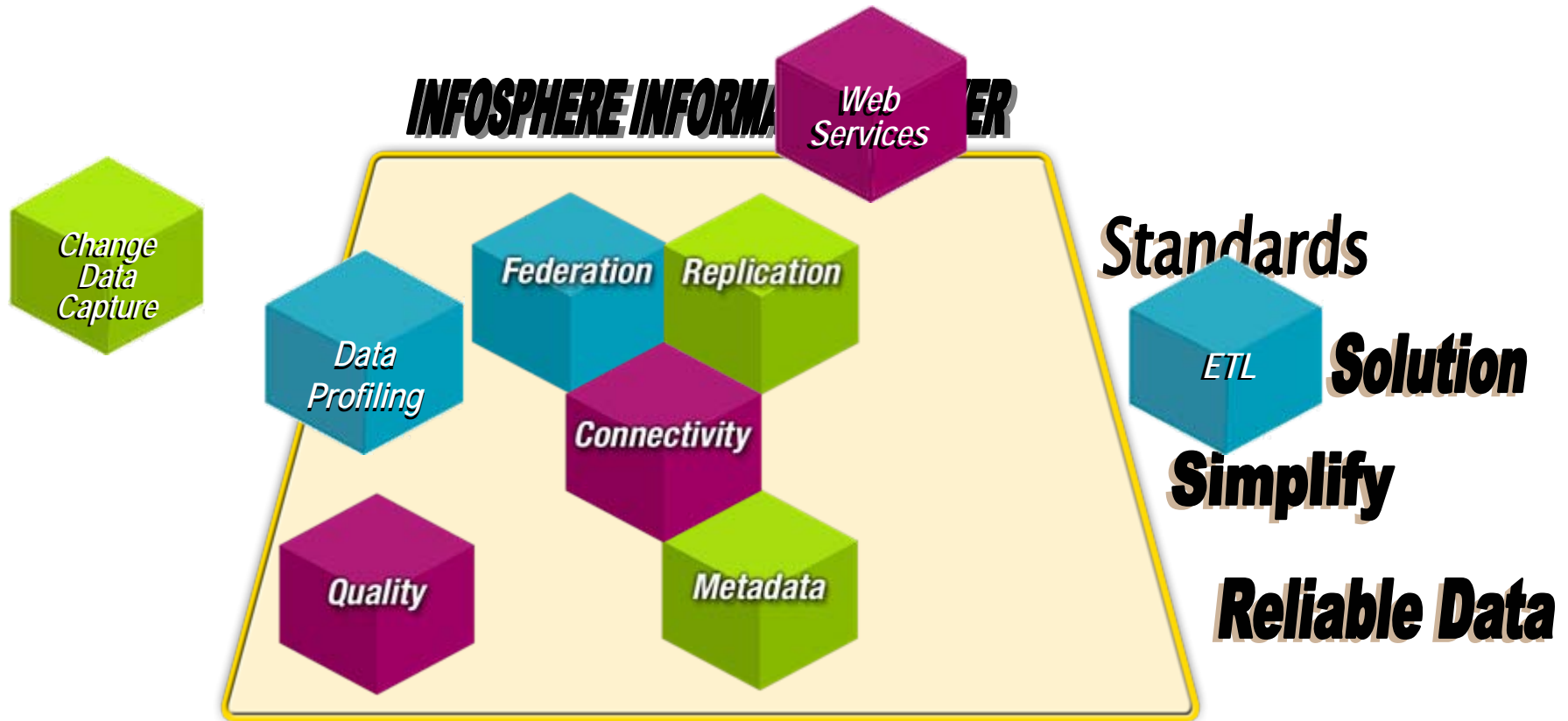
InfoSphere Information Server Supports Metadata

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 Server

metadata

Design   Operational

- Data Profiling
- SOA deployment
- Logging/Reporting

- Data Quality
- ETL, Data Transform
- Admin and Security

- Impact analysis
- Data

Let's start by learning about metadata
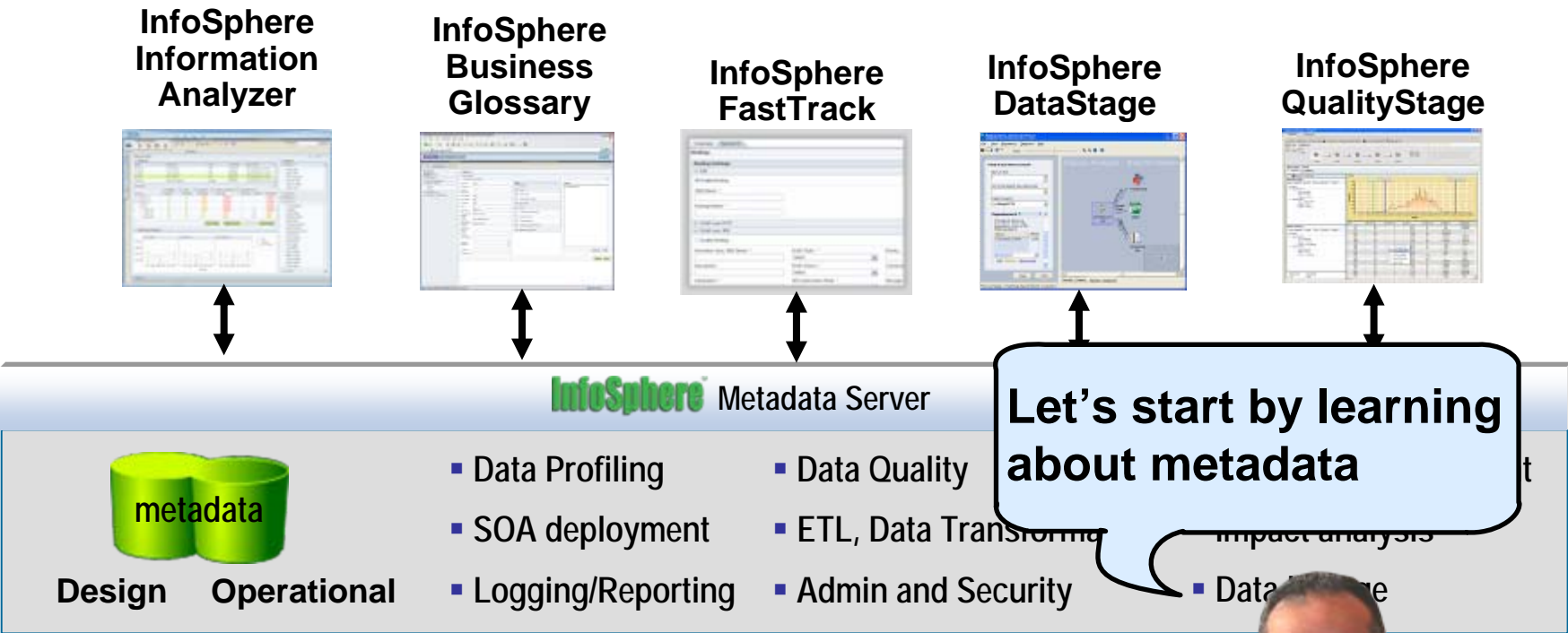
Introduction To Metadata

- 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 stores metadata in the Metadata Server

InfoSphere Metadata Server

metadata		
Design Operational	• Data Profiling	• Data Quality	• Metadata management
	• SOA deployment	• ETL, Data Transformation	• Impact analysis
	• Logging/Reporting	• Admin and Security	• Data lineage
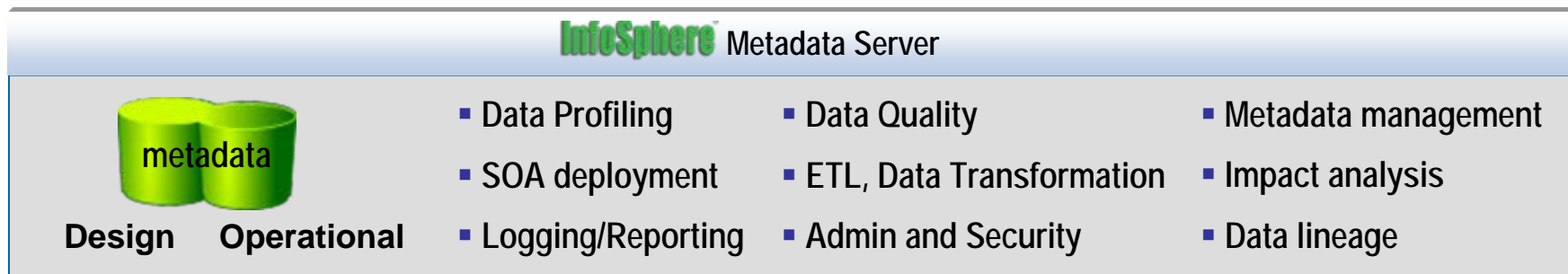
InfoSphere Information Analyzer Discovers Metadata

InfoSphere Information Analyzer

InfoSphere Business Glossary

InfoSphere FastTrack

InfoSphere DataStage

InfoSphere QualityStage

InfoSphere Metadata Server

metadata

Design **Operational**

- Data Profiling
- SOA deployment
- Logging/Reporting

- Data Quality
- ETL, Data Transf...
- Admin and Secur...

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?**

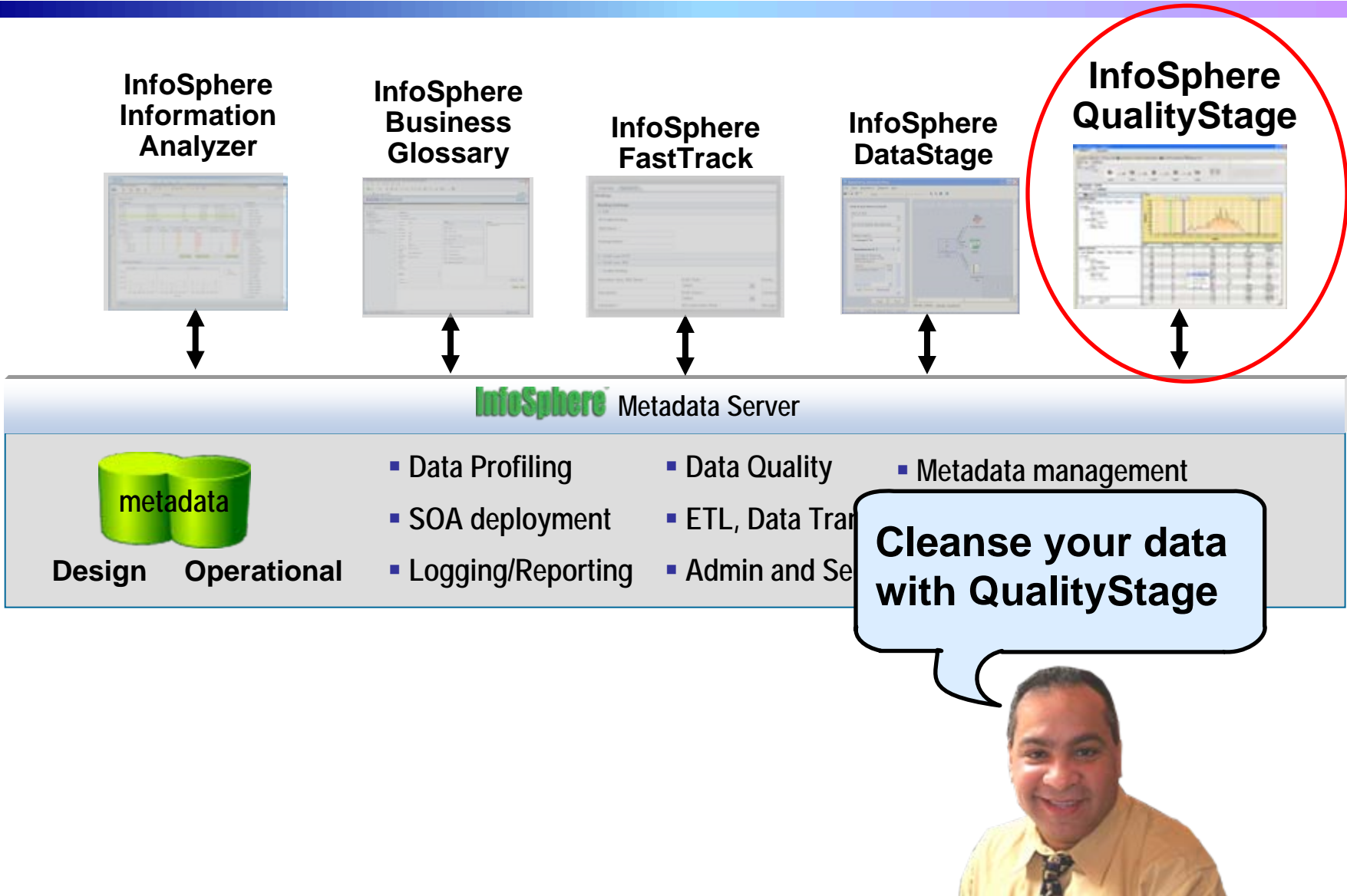7. **How do I get it where it needs to go?**

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

InfoSphere QualityStage Uses Metadata To Create Trusted Data

InfoSphere Information Analyzer

InfoSphere Business Glossary

InfoSphere FastTrack

InfoSphere DataStage

InfoSphere QualityStage

InfoSphere Metadata Server

metadata

Design   Operational

- Data Profiling
- SOA deployment
- Logging/Reporting

- Data Quality
- ETL, Data Tra...
- Admin and Se...

- Metadata management

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)

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

Dynamically Cleanse Your Data While Loading Your Data Warehouse with QualityStage
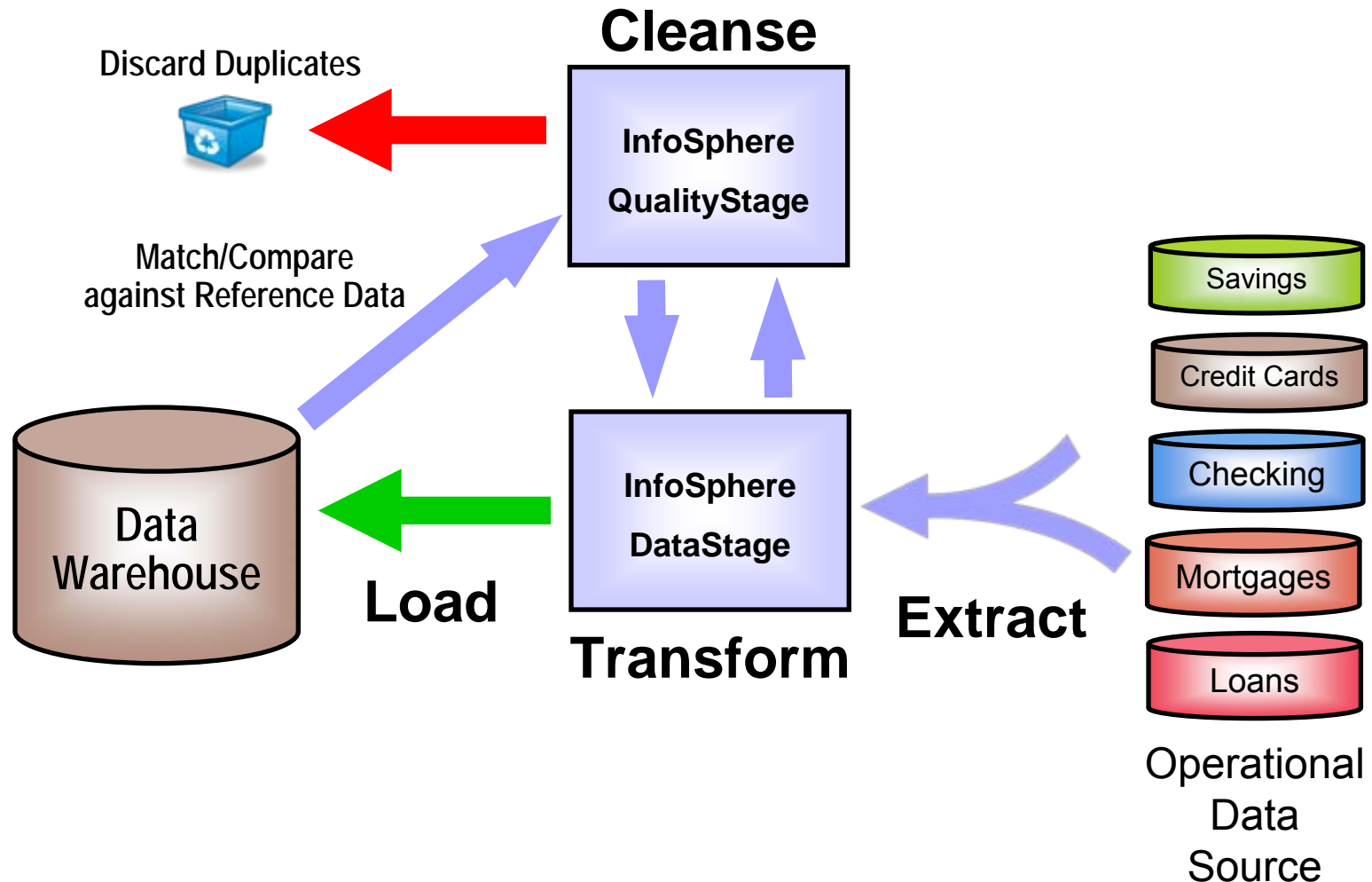
Service Oriented Finance Wants A Dynamic Data Warehouse

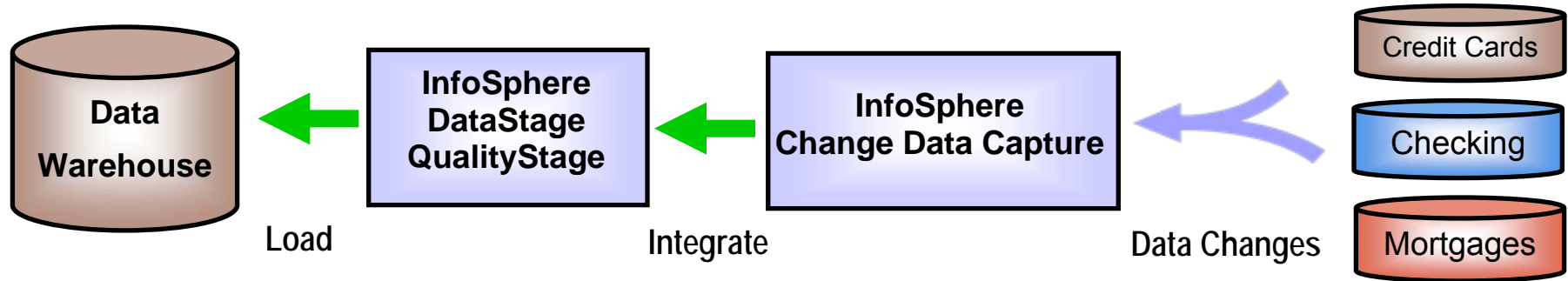The information our decision makers use needs to be more current.

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 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

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.