IBM Cognos Performance
Better Intelligence. Better Business.

Data Quality

Bjarne Jensen
IBM Information Management Technical Sales

© 2009 IBM Corporation

Agenda

- **Introductie**

Wat is Data Qualiteit en waarom is het belangrijk?

Het Informatie Landschap

Het Data Governance Platform

Verbeterde Data Qualiteit door Infosphere Information Server

Samenvatting

Transform Your Organization with Information

Volume of Digital Data

Variety of Information

Velocity of Decision Making

Lack of Insight

1 in 3 managers frequently make critical decisions without the information they need

Inefficient Access

1 in 2 don't have access to the information across their organization needed to do their jobs

Inability to Predict

3 in 4 business leaders say more predictive information would drive better decisions

Source: IBM Institute for Business Value

Agenda

Introductie

Wat is Data Qualiteit en waarom is het belangrijk?

Het Informatie Landschap

Het Data Governance Platform

Verbeterde Data Qualiteit door Infosphere Information Server

Samenvatting

Definitie van Data Qualiteit

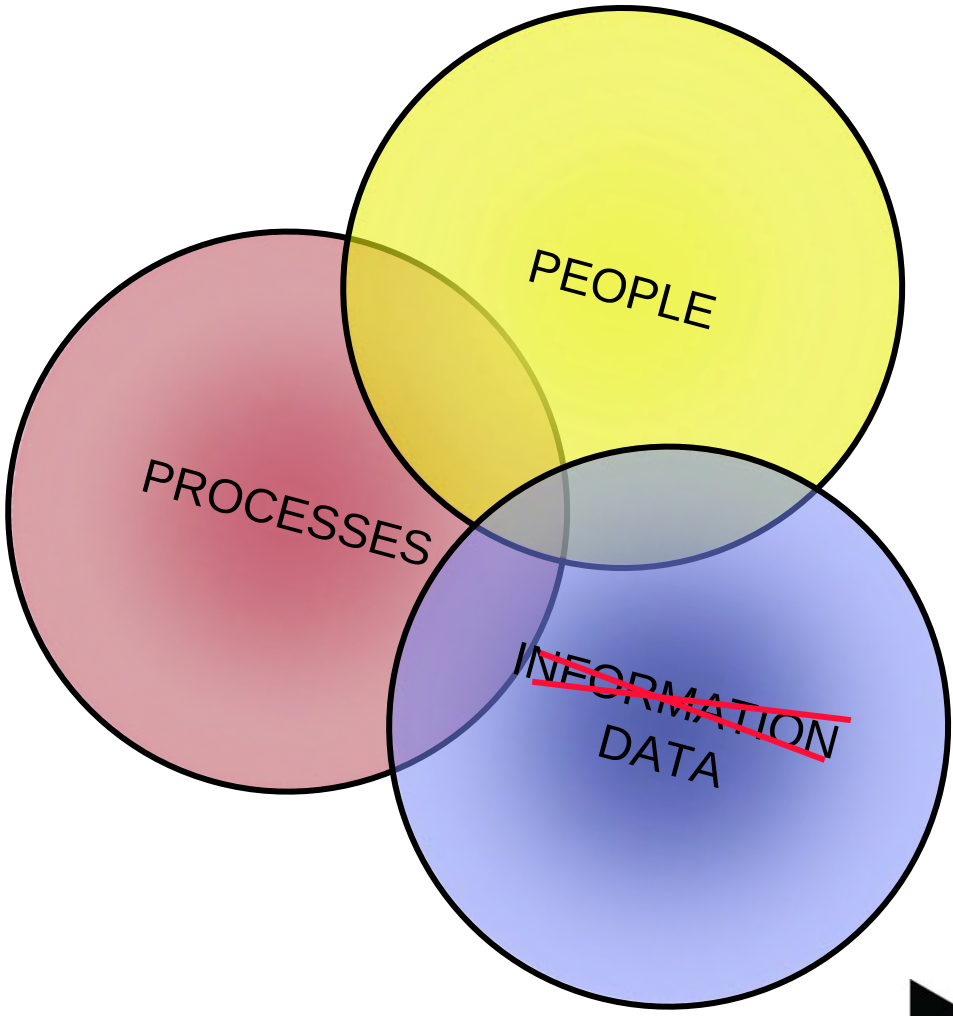
The state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use.

We have a customer information system, then my data quality is good, isn´t it?

- Schizophrenia
- Free Import & Typing mistakes
- Creativity
- Spill Over
- New releases
- Merging of customer info systems
  CRM systems are unique
  Difference in the operatability, usability, manageability, structure etc.
  Customers in multiple CRM systems

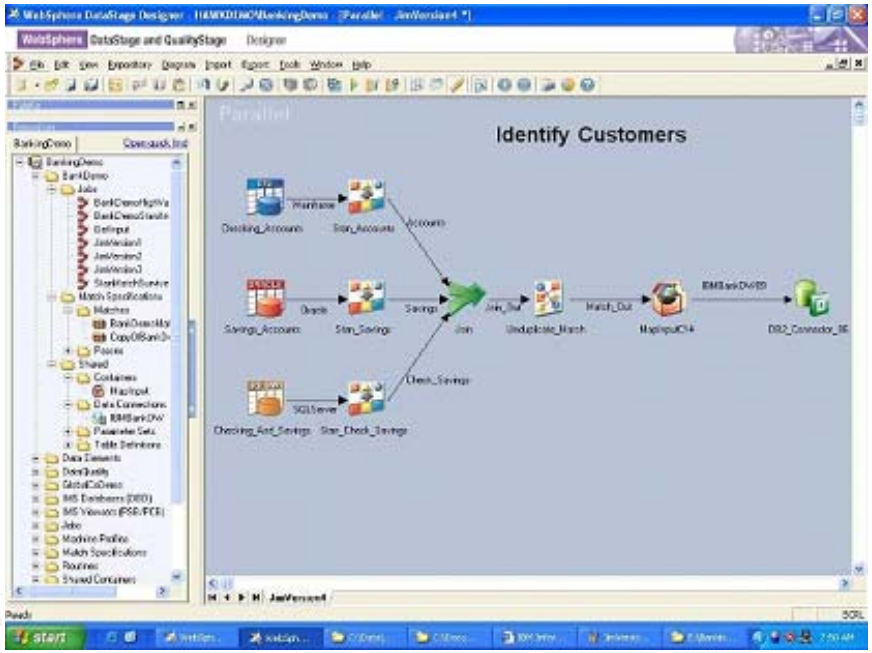
Conclusion:

Same problems, but then worse



Manually?

Time consuming
Bad quality
One time (4% year ´pollution´)
Uncontrollable



Automated process?

Consistent high quality
Re-usable
Fast and efficient
Controllable
Traceability

Agenda

Introductie

Wat is Data Qualiteit en waarom is het belangrijk?

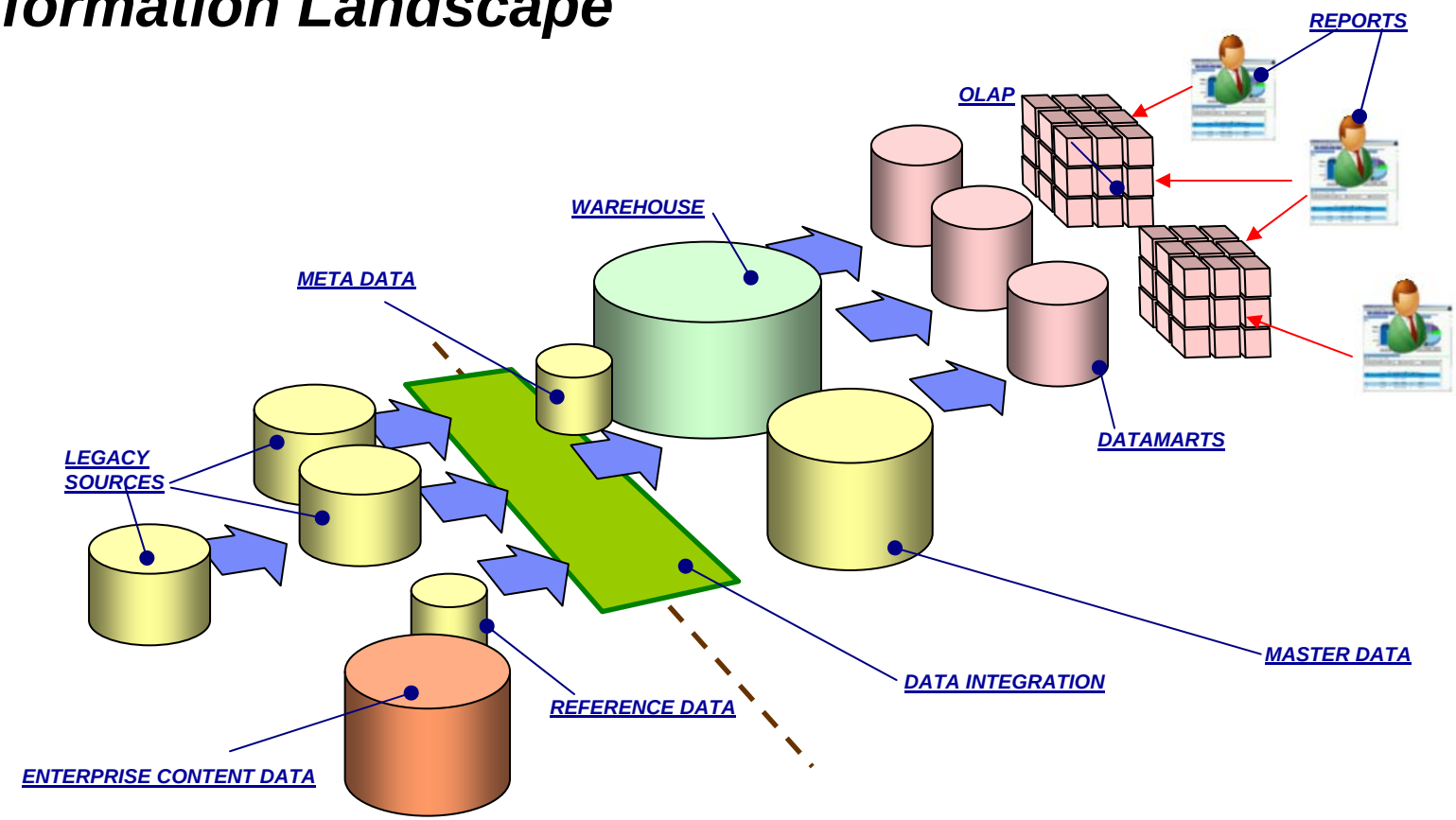
Het Informatie Landschap

Het Data Governance Platform

Verbeterde Data Qualiteit door Infosphere Information Server

Samenvatting

Information Landscape



Information On Demand

Better Business Outcomes

End-to-end Capabilities

Financial Risk Insight

Workforce Optimization

Dynamic Supply Chain

Customer & Product Profitability

Business Optimization

Multi-channel Marketing…

Leaders Investing to Optimize Information Assets
Clients Tell Us What They Are Looking For …

InfoSphere Capabilities

Insightful
Derive meaning from information changes

In Context
Real-time delivery of relevant information when and where it's needed

Complete
Related information reconciled into a single and holistic view

Accurate
Complex and disparate data transformed, cleansed and delivered

Agenda

Introductie

Wat is Data Qualiteit en waarom is het belangrijk?

Het Informatie Landschap

Het Data Governance Platform

Verbeterde Data Qualiteit door Infosphere Information Server

Samenvatting

IBM Cognos Performance

How do I know that that I have an accurate view and how do I know what this data means?

Business Users

Subject Matter Experts

How can I actively collaborate with developers, f.e. in the area of data mapping/business rules?

I need a graphical environment to easily adapt to future data requirements (SID..?)

Developers

What about Governance, Security, Scalability?

Architects

Simplify administration, deployment, & maintenance

DBAs

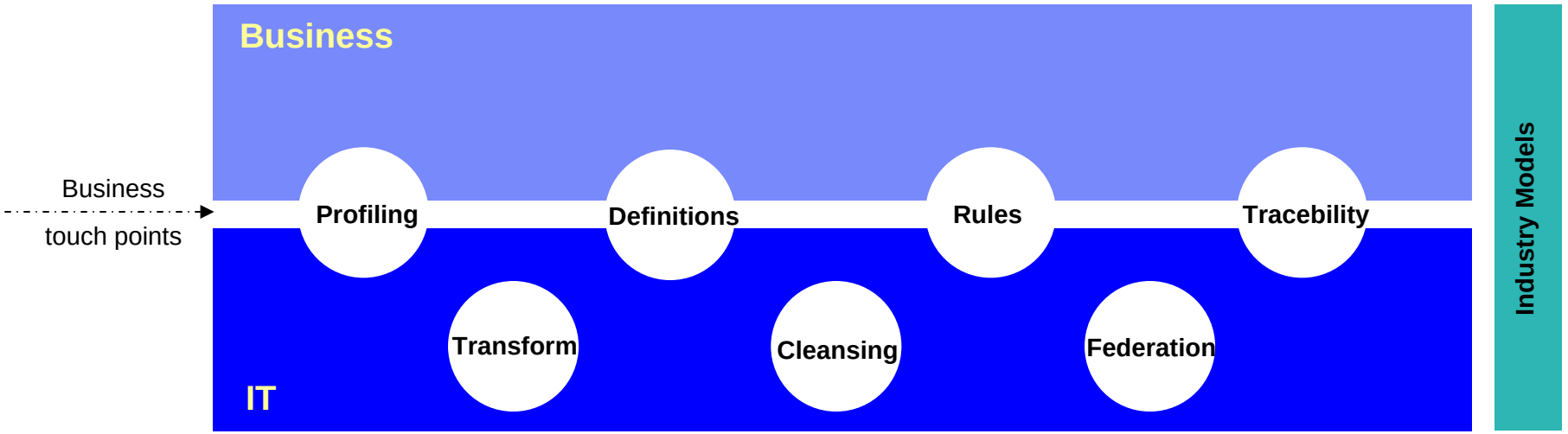
How can I quickly tell the business where the data is coming from?

Data Analysts

IBM

Business & IT alignment in information delivery

Offer tailored collaborative solutions for both Business & IT

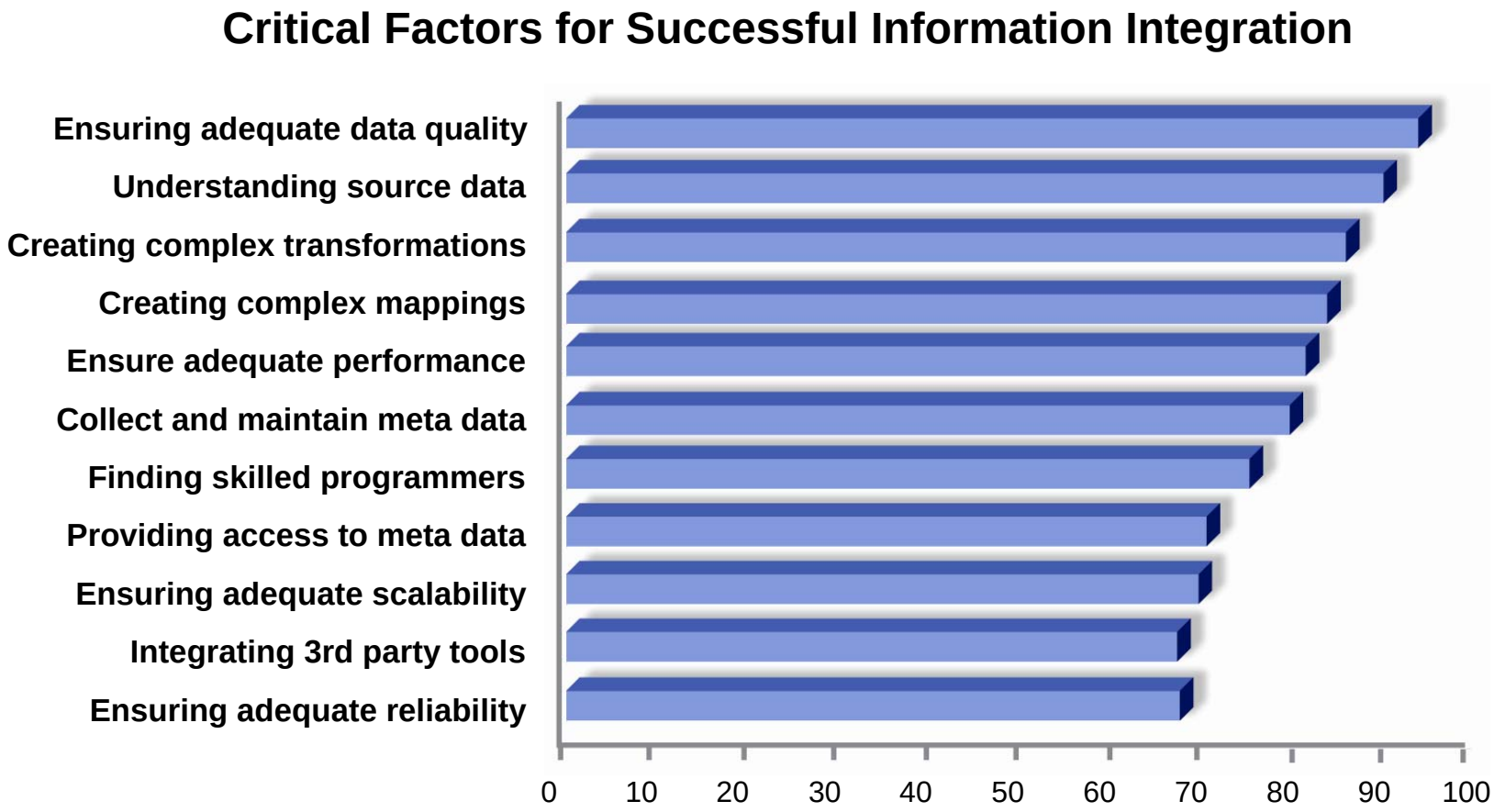


In order to take full control of information integration and enable the business to drive definition, rules and results a solution needs to deliver non-technical solutions to be able to:

- **Understand the nature of the data in all of its aspects**
- **Manage business definitions and terminology**
- **Manage and control source to target translations**
- **Leverage existing investments in data models**
- **Fully understand and control data traceability**

This can only be achieved with a fully integrated platform

Information Integration



Source: TDWI World Conference

An Enterprise Data Governance Platform should...

1. ... Deliver a broad range of data integration capabilities ...

2. ... tailored to the needs of both business & IT users in doing their part of the work effectively and efficiently ...

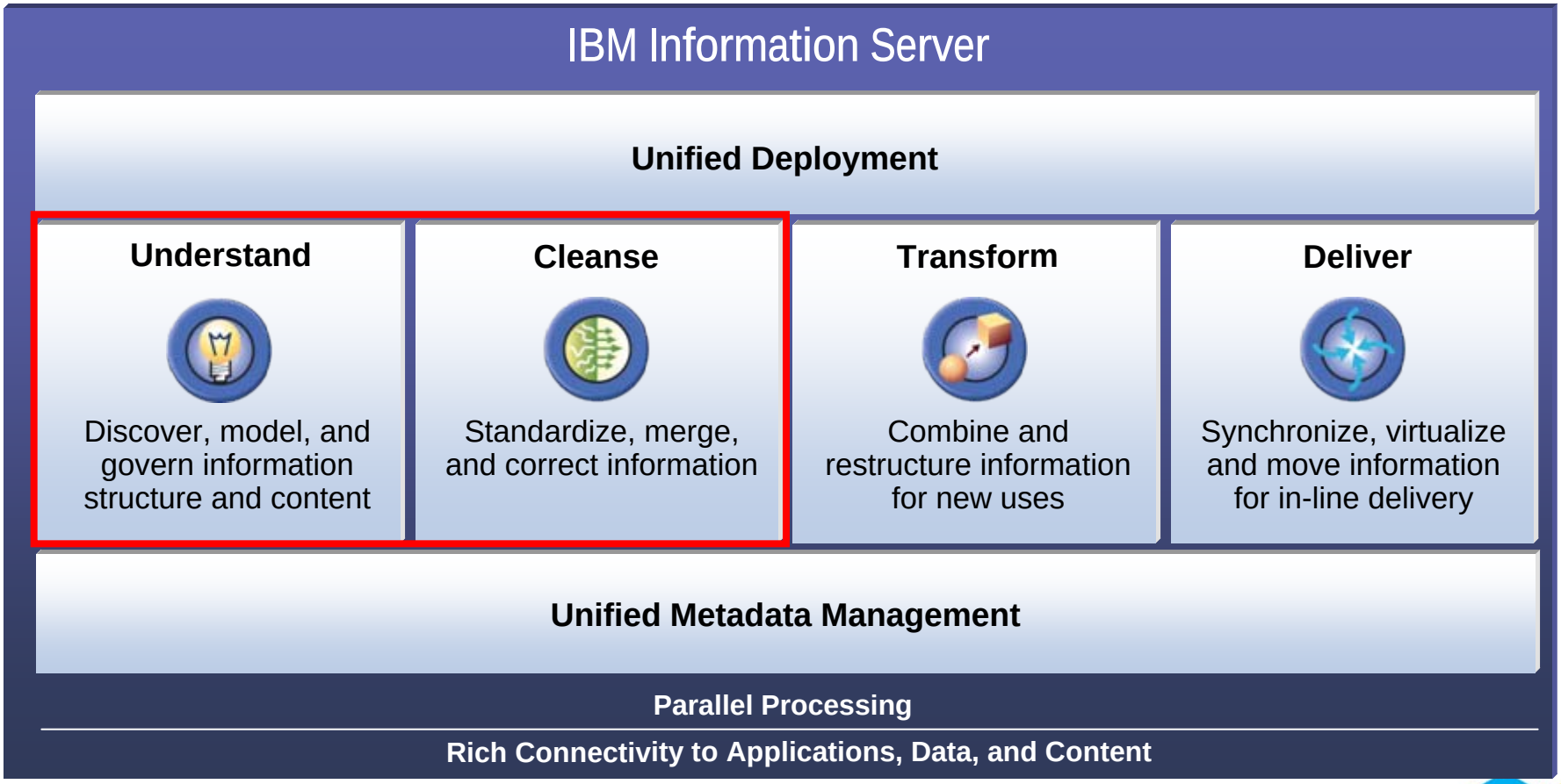
3. ... in an integrated way to enable business & IT alignment, traceability/auditability and re-useability of information assets for collaboration to ensure better business outcome.

Agenda

- Introductie
- Wat is Data Qualiteit en waarom is het belangrijk?
- Het Informatie Landschap
- Het Data Governance Platform
- **Verbeterde Data Qualiteit door Infosphere Information Server**
- Samenvatting

The IBM Solution: IBM Information Server

Delivering information you can trust



Information Analyzer - Data Profiling

ERP from acquisition

Mainframe systems

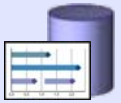
Parts BOM

External Lists

Distribution

Demographic

Contact


Billing / Accounts

Critical Problems:

- **Which data is *really* in your legacy systems?**
- **Sources are new and unknown, or have changed**

Why?

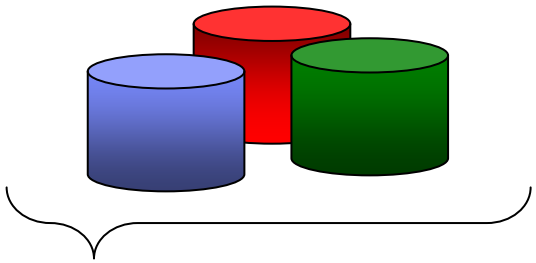
- **Data values and relationships are inconsistent and divergent from documented (?) rules**
- **Documentation, if it exists, is incomplete**
- **Data sources are never static. They change without warning**

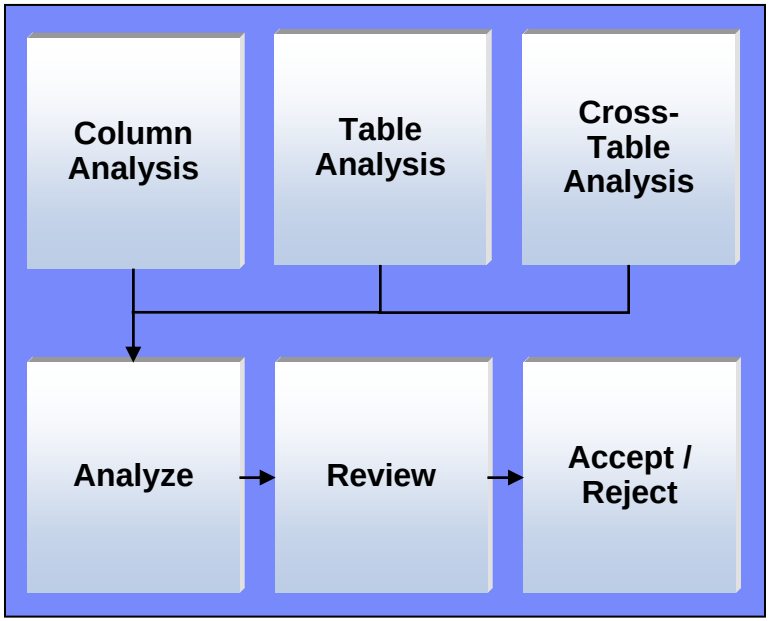
Typical Strategy

- **Labor intensive, resource devouring process**
- **Unable to review 100% of data elements**

IBM

Information Analyzer - Data Profiling



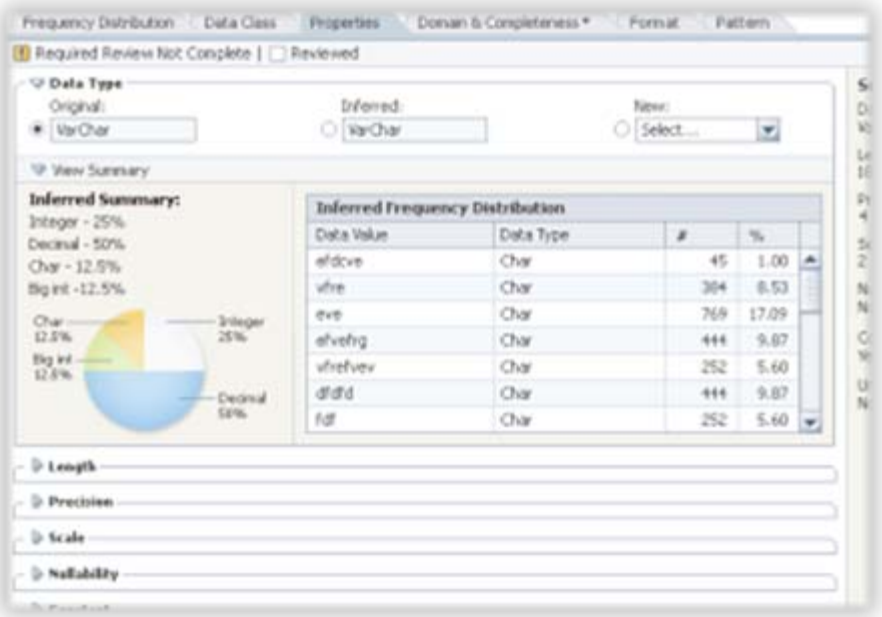


- Enables understanding of data ***before*** starting development
- **Automates the profiling effort**
  - **Column analysis**
    - Generates frequency distributions for all values in all columns
  - **Relationship analysis**
    - Discovers join possibilities
  - **Table analysis**
    - Identifies primary and foreign key candidates
- **Performs baseline comparisons**
- **Helps validate schemas**

InfoSphere Information Analyzer

Information Analyzer

Analyze source data structures, and monitor adherence to integration and quality rules



Requirements

- Discover and analyze source data structure, content and quality
- Identify patterns and relationships across systems
- Identify data re-engineering needs

Benefits

- Identify information gaps early to reduce project risks
- Monitor information quality over time
- Visibility to enterprise information assets

Why Should I Care About Cleansing Information?

- **Lack of information standards**
  - Different formats & structures across different systems
  - **Data surprises in individual fields**
  - Data misplaced in the database
- **Information buried in free-form fields (text)**
- **Data myopia**
  - Lack of consistent identifiers inhibit a single view
- **The redundancy nightmare**
  - Duplicate records with a lack of standards

```
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
```

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

```
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)
```

```
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
```

```
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
```

Data Cleansing: Infosphere QualityStage

- Specialized data quality functions seamlessly integrated with DataStage
- Visual tools for defining complex matching and survivorship logic
- Ensures clean, standardized, de-duplicated information
- Enables a single version of the truth

4 Steps to clean your data:
- Free form investigation
- Standardize
- Match
- Survivorship

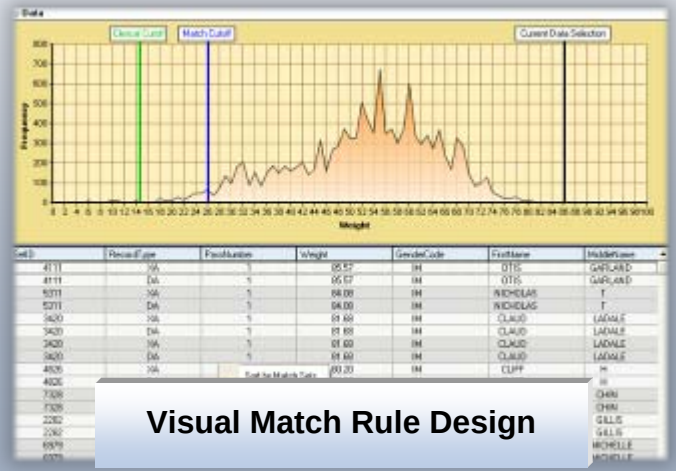
Subject Matter Experts

Data Analysts

Cleanse

Infosphere QualityStage™

Standardize and correct source data fields, and match records together across sources to create a single view

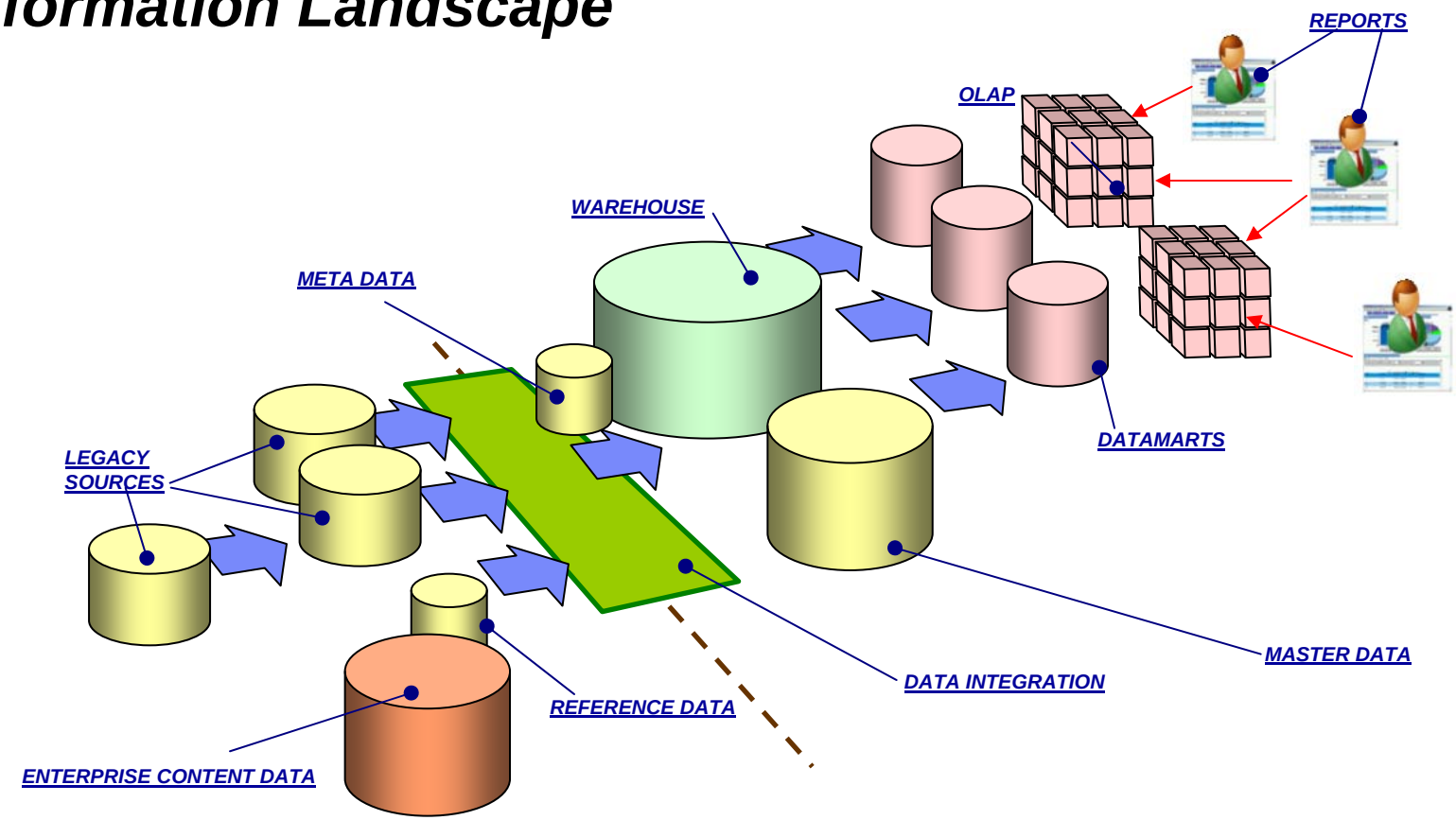


Visual Match Rule Design

Agenda

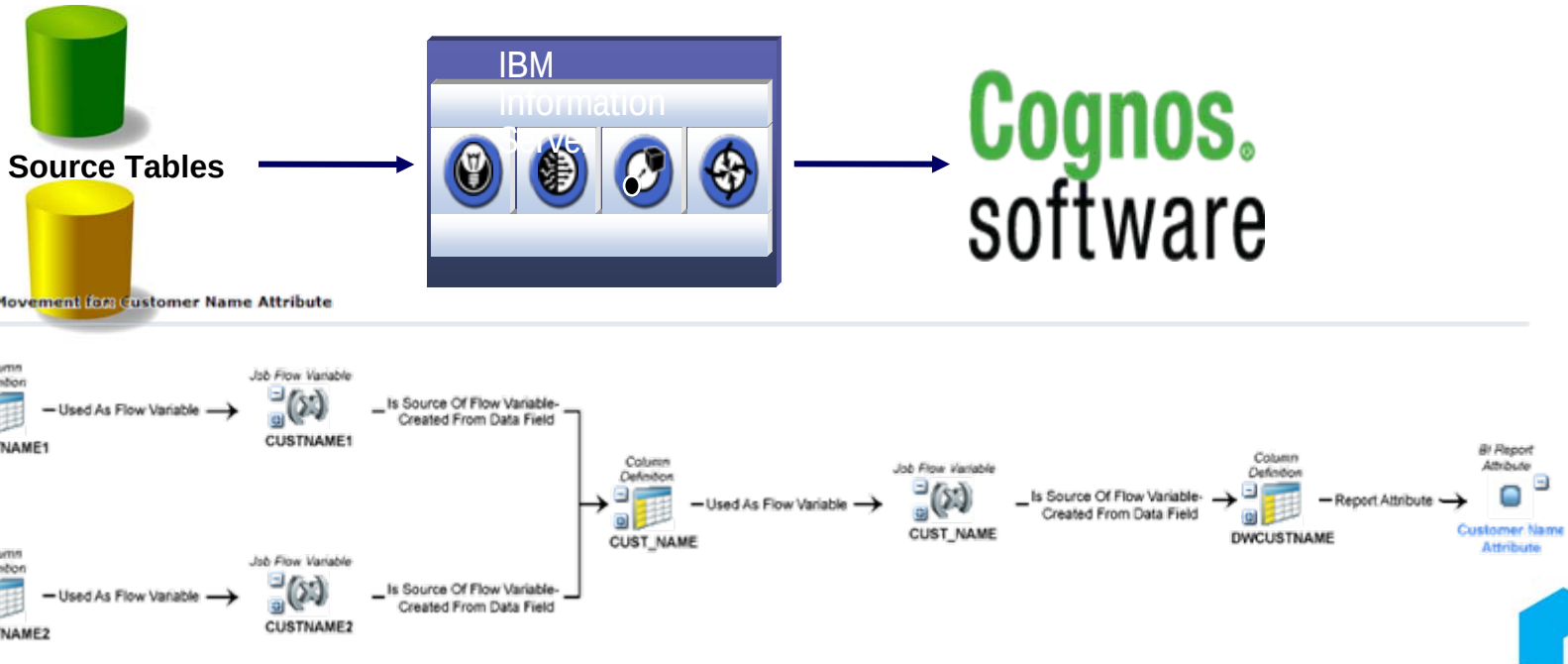
- Introductie
- Wat is Data Qualiteit en waarom is het belangrijk?
- Het Informatie Landschap
- Het Data Governance Platform
- Verbeterde Data Qualiteit door Infosphere Information Server
- **Samenvatting**

Information Landscape



Where does a field of data in this report come from?

- Import & Browse Full BI Report Metadata
- Navigate through report attributes
- Visually navigate through data lineage across tools
- **Increases trust and understanding of business information**



InfoSphere and Cognos

InfoSphere provides the foundation of trusted information for Cognos business intelligence & performance management

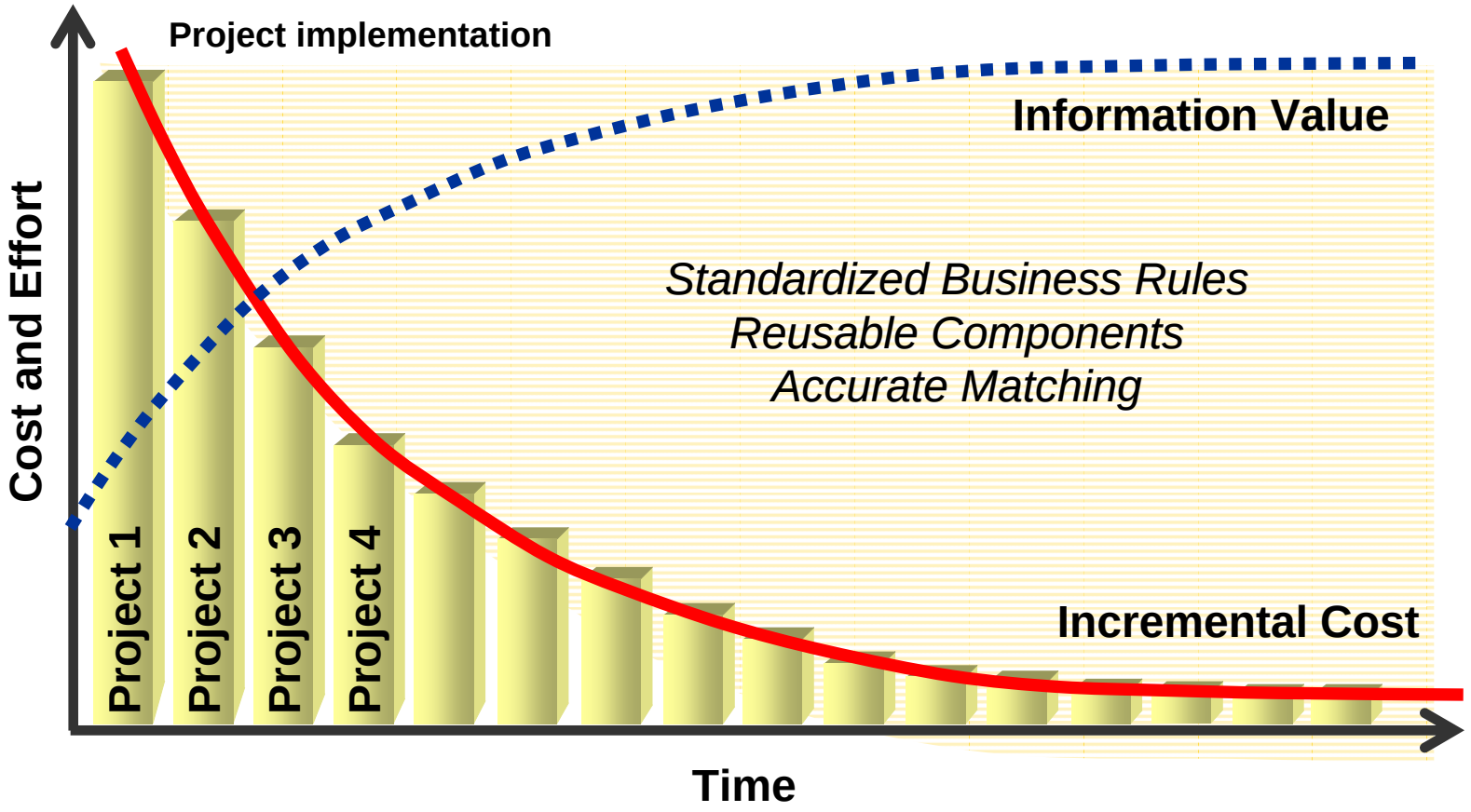
InfoSphere™ software

Trusted Information

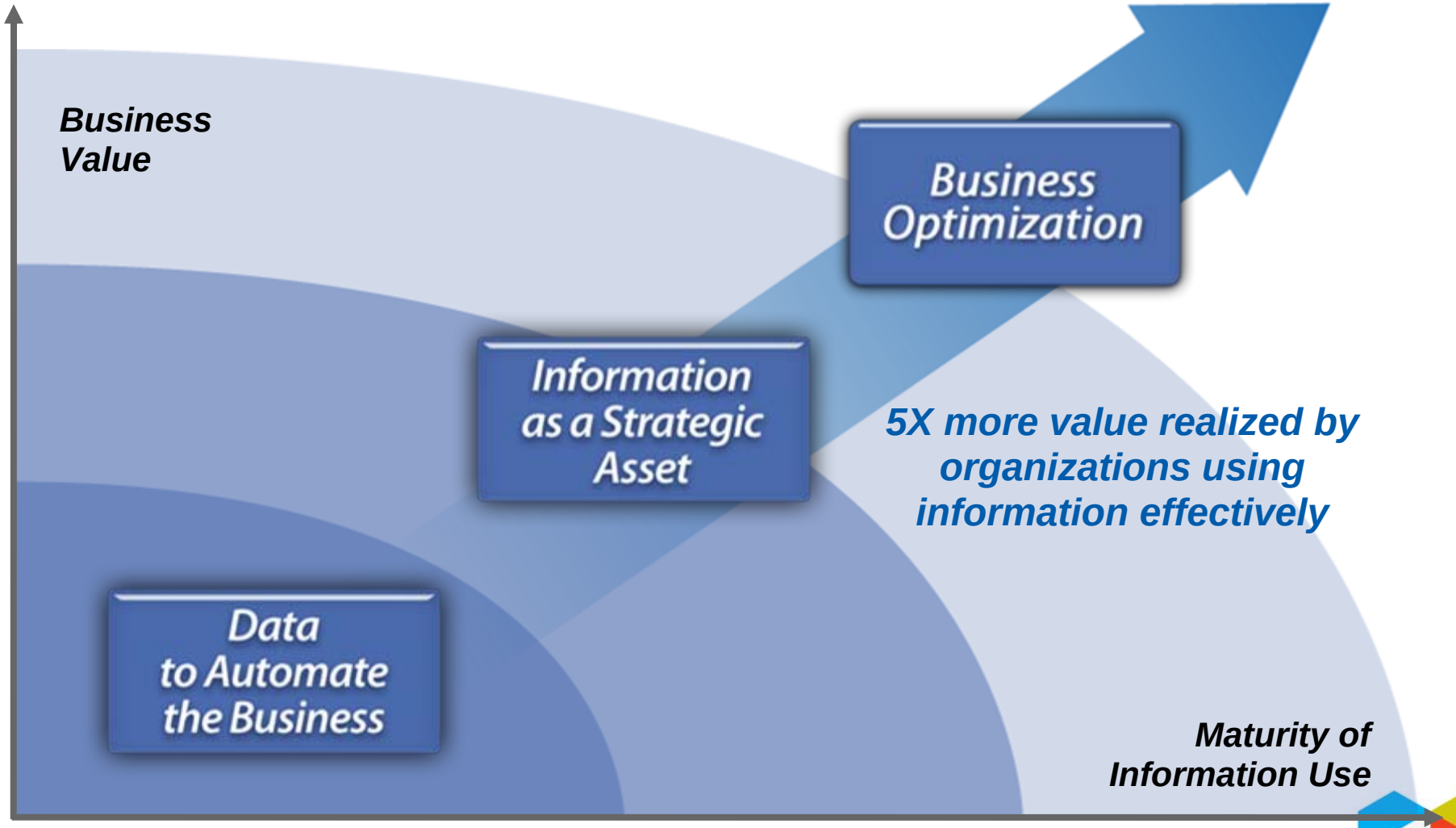
- ***Business metadata sharing***
- ***End-to-end data lineage***
- ***Data Profiling and Quality integration***
- ***InfoSphere DataStage with Cognos Data Manager***
- ***InfoSphere Change Data Capture with Cognos Now!***

IBM

Break-through Productivity With Information Server



Accelerating Your Journey to Business Optimization



Why Choose Information Server…

- **Complete, unified data integration capabilities**
  - *Removes the integration burden from you & allows for easy capability expansion*
- **End-to-end metadata management**
  - *Provides more insight and tracking of information – enables data governance*
- **Model-driven, metadata-driven development**
  - *Accelerates development, leverages industry assets, aligns business & technology*
- **Unified, market-leading data quality**
  - *Improves data, improves business trust, enables a single view*
- **Easy, visual, reuse-optimized development**
  - *Speeds time to value, lowers maintenance cost, Improves project to project leverage*
- **Snap-in, linear scalability through parallel processing**
  - *Handles any data volume, Optimizes hardware investment, Allows painless growth*
- **Extensive, native connectivity and change data capture**
  - *Allows all data to be integrated and accessed easily and efficiently with low impact to sources*
- **Extensive integration (Cognos, SOA, MDM, etc.)**
  - *Improves utility, Improves return on investments*

Questions?