



*Data Management*

# IBM Infosphere Discovery Overview

19. Januar 2010

## Important Disclaimer

**THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY.**

**WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED.**

**IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE.**

**IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION.**

**NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, OR SHALL HAVE THE EFFECT OF:**

- **CREATING ANY WARRANTY OR REPRESENTATION FROM IBM (OR ITS AFFILIATES OR ITS OR THEIR SUPPLIERS AND/OR LICENSORS); OR**
- **ALTERING THE TERMS AND CONDITIONS OF THE APPLICABLE LICENSE AGREEMENT GOVERNING THE USE OF IBM SOFTWARE.**

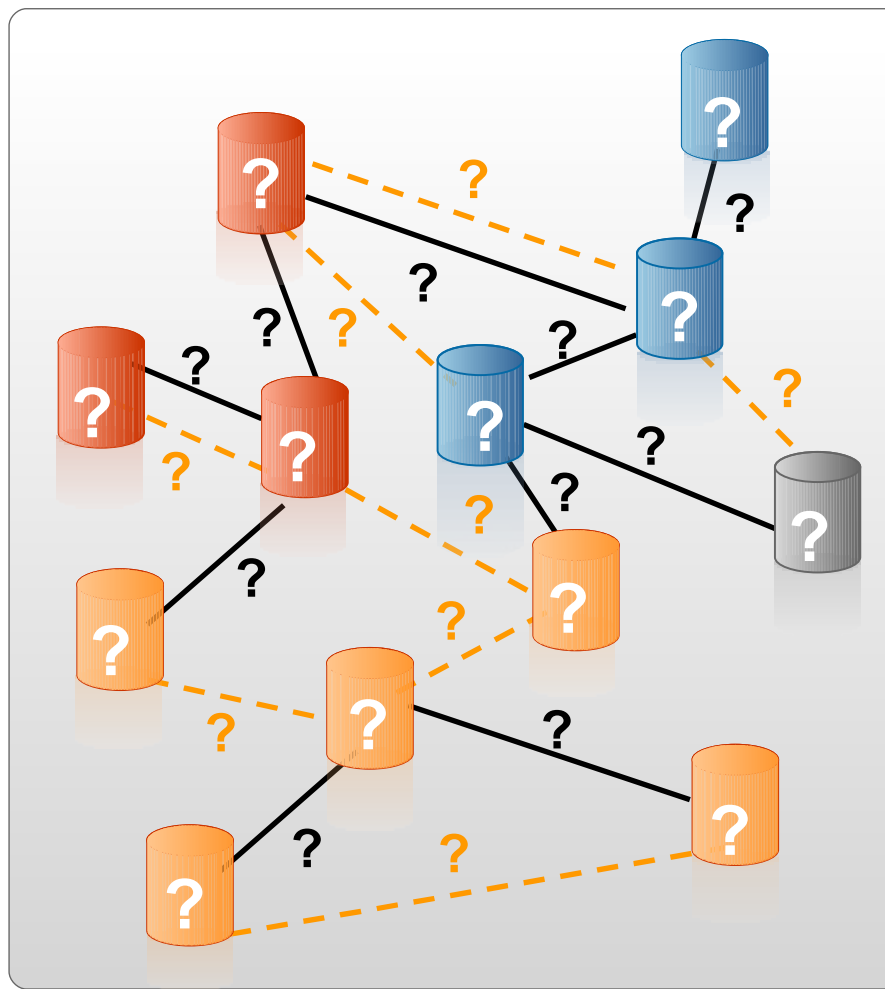
# You Can't Manage What you don't understand!

## • Distributed Data Situation:

- Grows exponentially
- Increasingly distributed
- Poorly understood

## • Problems:

- Poor understanding = Poor IT Agility
- Poor understanding = Poor Data Governance
- Poor understanding = Bad data = Bad business decisions



**Distributed Data Landscape**



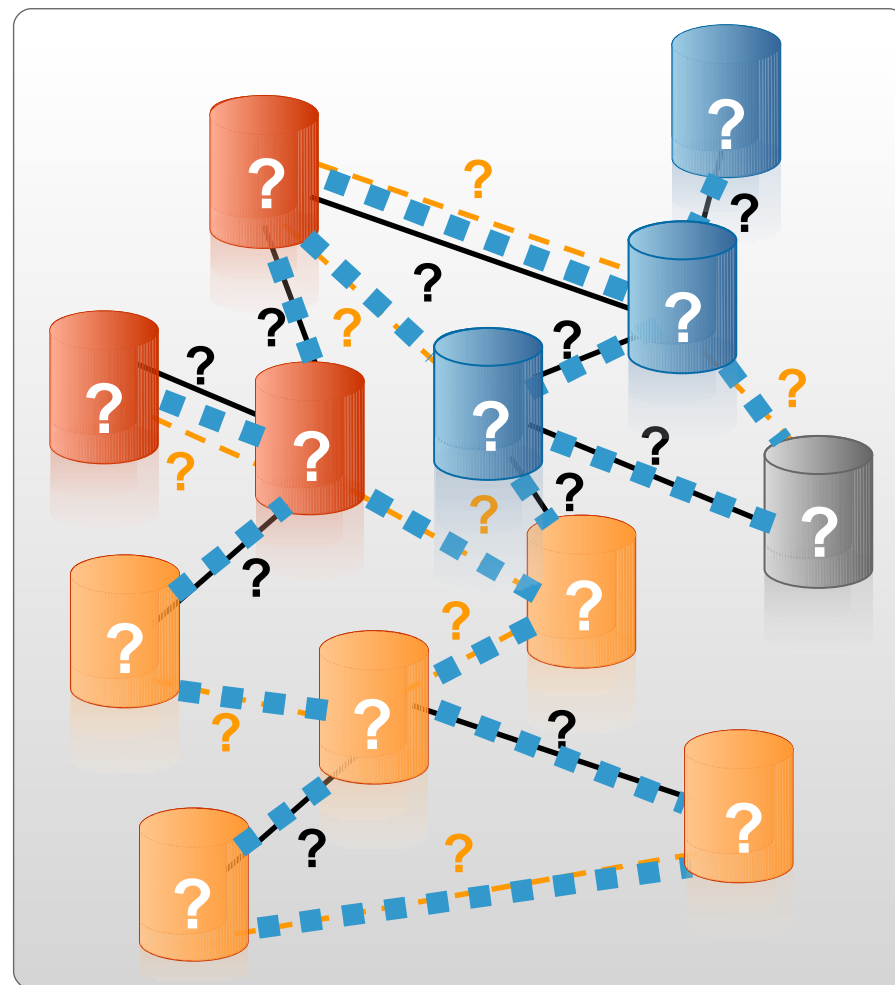
### Business Person

#### Applications

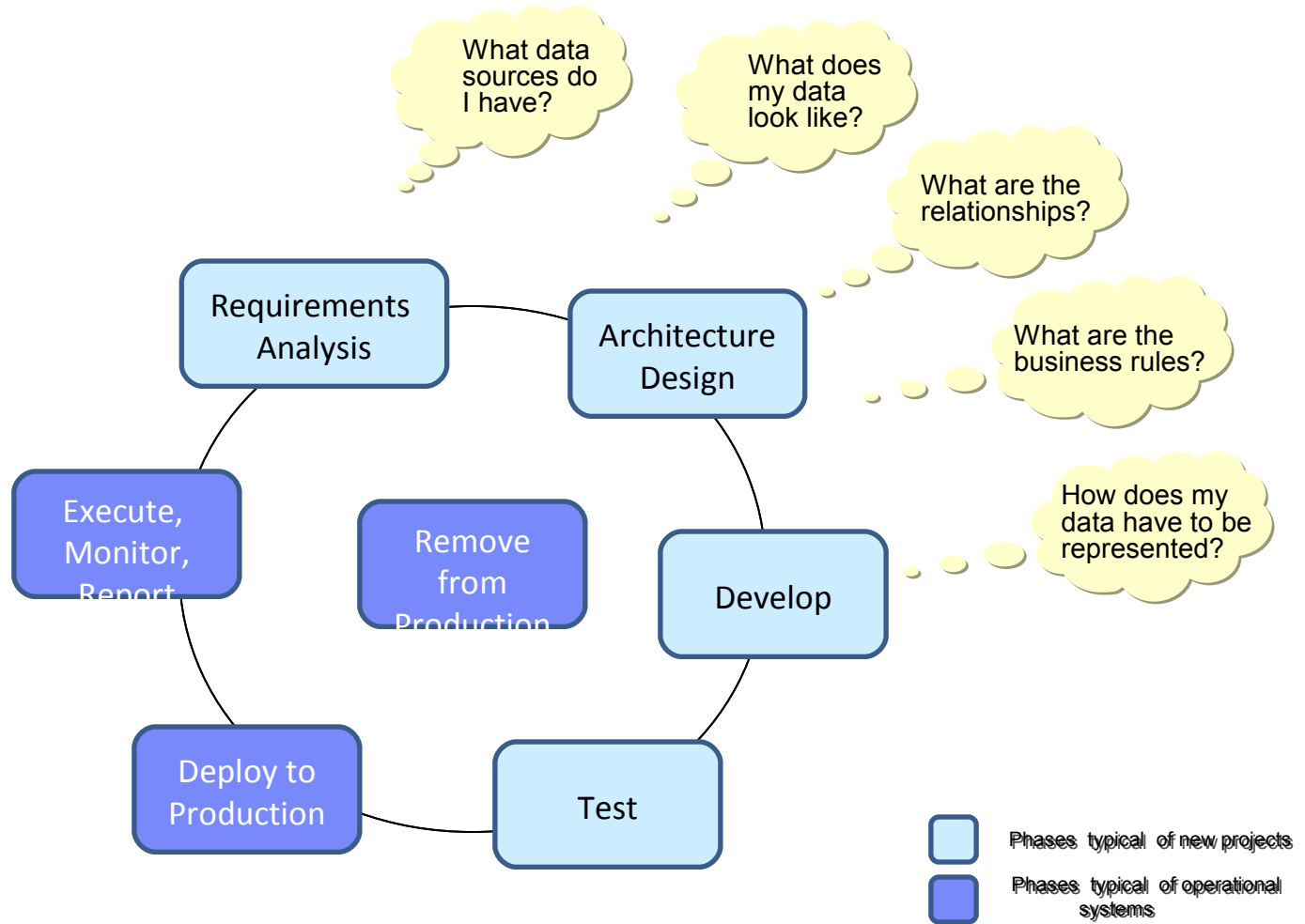
- CRM
- ERP
- Financials
- Risk Mgmt
- BI Reports
- Etc.

# Automate Discovery and Accelerate Information Understanding

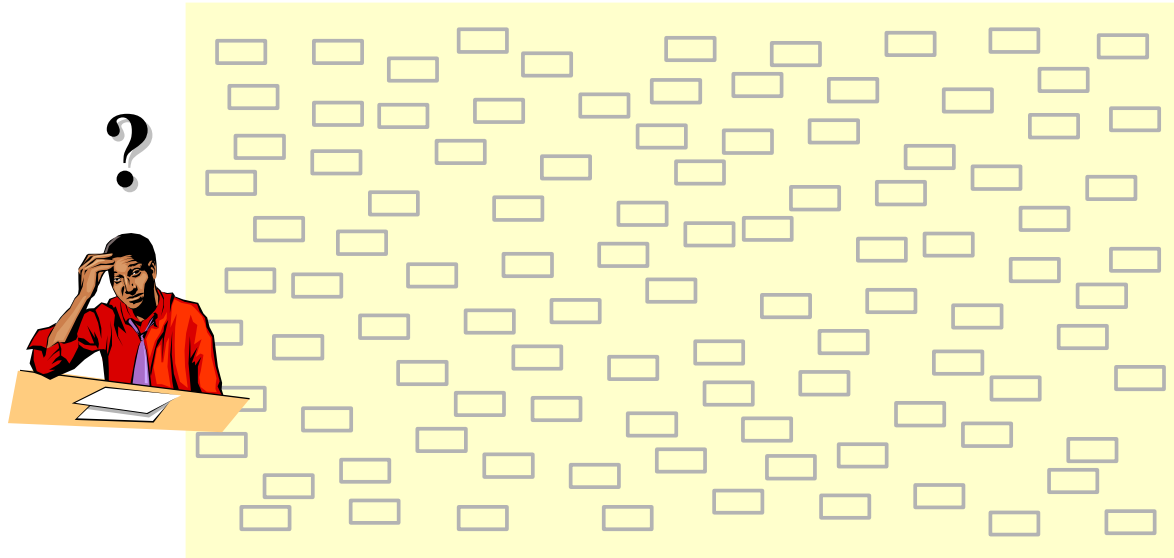
- **Significant Acceleration of Information Agenda projects**
  - Data Growth Management
  - Test Data Management
  - Sensitive Data De-identification
  - Application Consolidation and Migration
  - Application Decommissioning and Retirement
- **Answers the Questions:**
  - How does data interoperate across systems?
  - What are the inconsistencies?
  - What is part of the business entity?
  - Where is my Sensitive Data Element?
- **Why is this Different?**
  - Data-based discovery
  - Automate discovery of business entities, cross-source business rules & transformation logic
  - Evaluate multiple data sources simultaneously
  - Identify & remediate cross-system rules and inconsistencies



# Typical Information Project - Lifecycle Phases

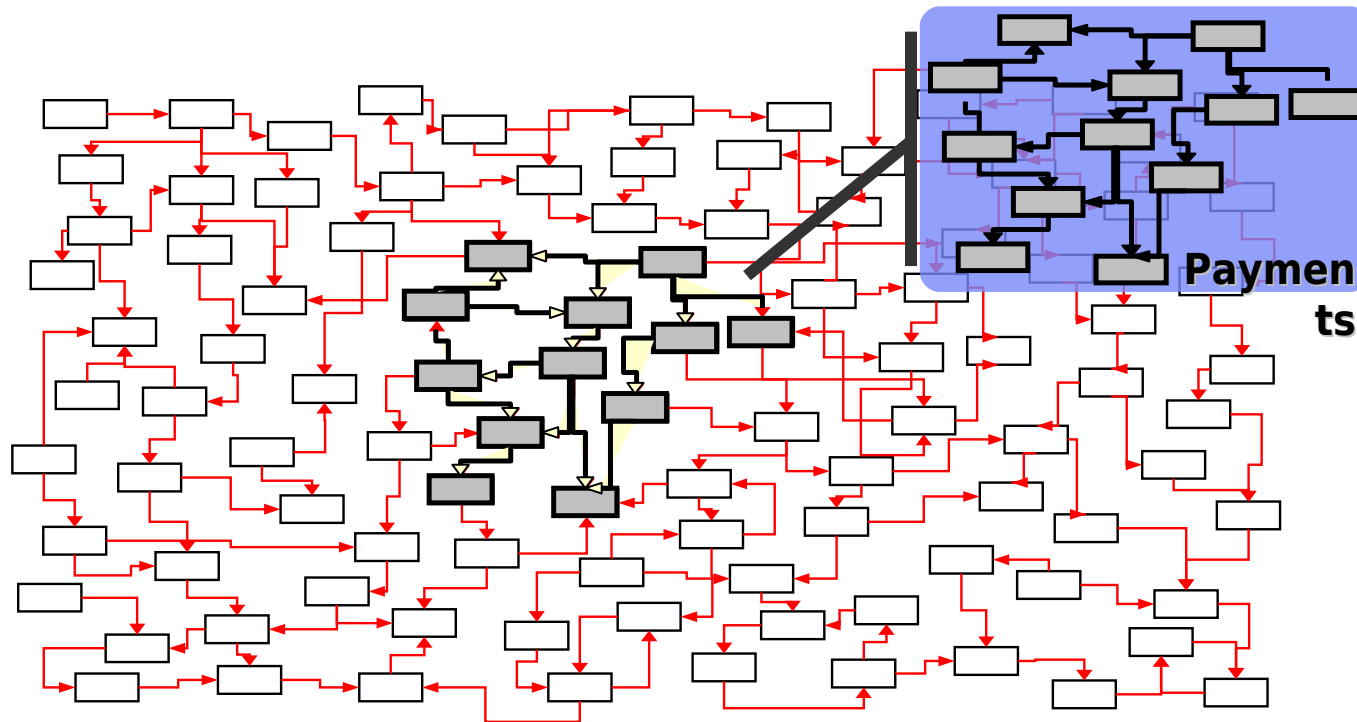


## The Challenge: How well do we know the data?



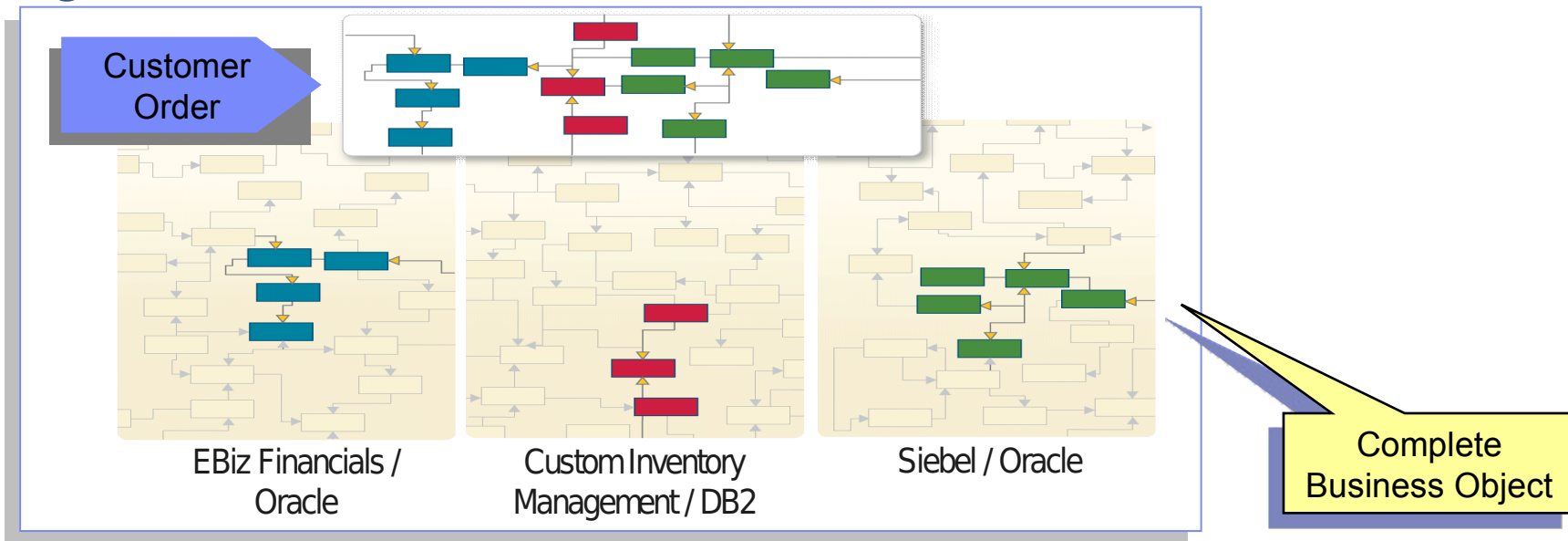
- **What tables should be in the business object?**
- **What data elements should be in the business object?**
- **How are the tables and data elements related (context)?**
- **What data elements are sensitive?**
- **How do I find them?**

## Complete Business Object



- Represents application data record – payment, invoice, customer
  - Referentially-intact subset of data across related tables and applications; includes metadata
- Provides “historical reference snapshot” of business activity
- Federated extract support across enterprise data stores

## Intelligent move of data is often more than a few tables



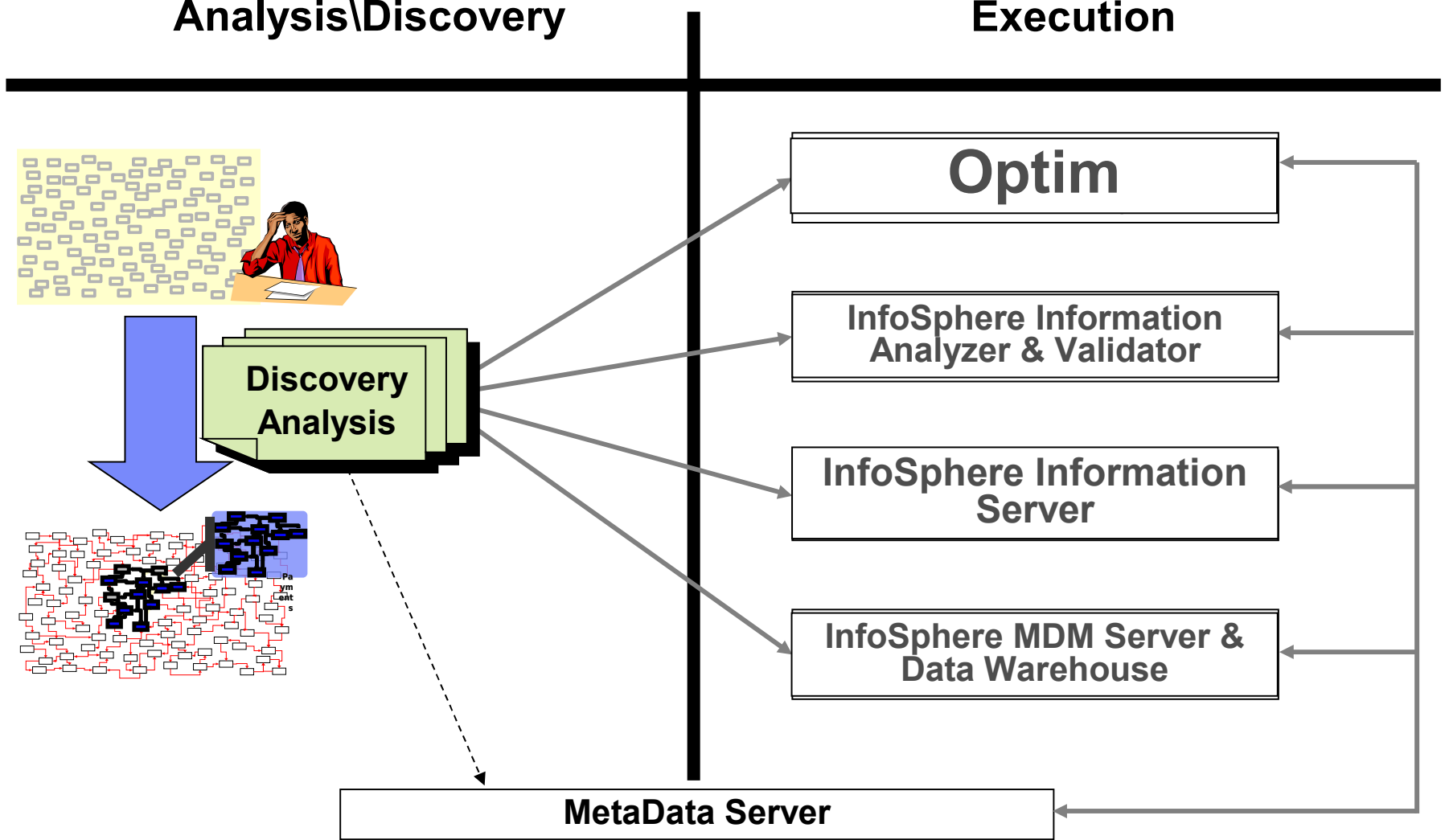
- Represents application data record – payment, invoice, customer
  - Referentially-intact subset of data across related tables and applications; includes metadata
- Provides “historical reference snapshot” of business activity
  - Responds to: why did we make that decision?
- Federated extract support across enterprise data stores



# Early Lifecycle Phase for All Information Projects

**Analysis\Discovery**

**Execution**



# Design-Time Data Discovery and Analysis

## • InfoSphere Discovery

### • Cross Source Profiling

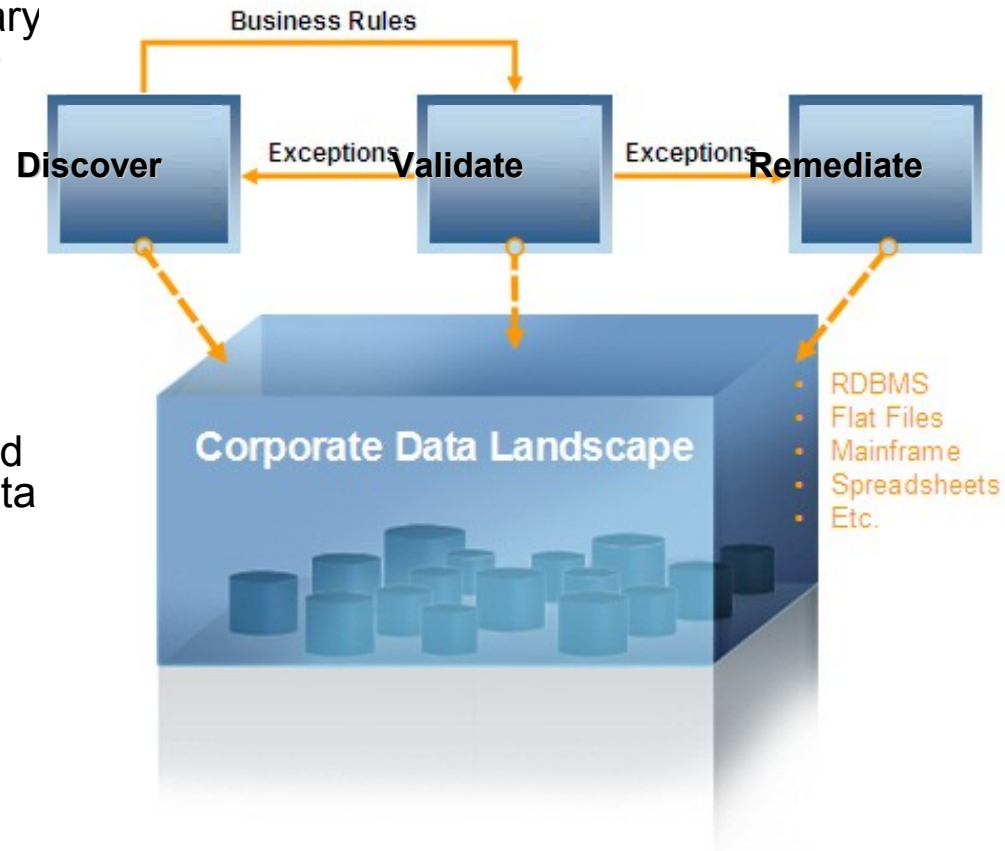
- Basic profiling plus automated primary foreign key, business entity & cross-source overlaps discovery

### • Prototyping a Unified Schema

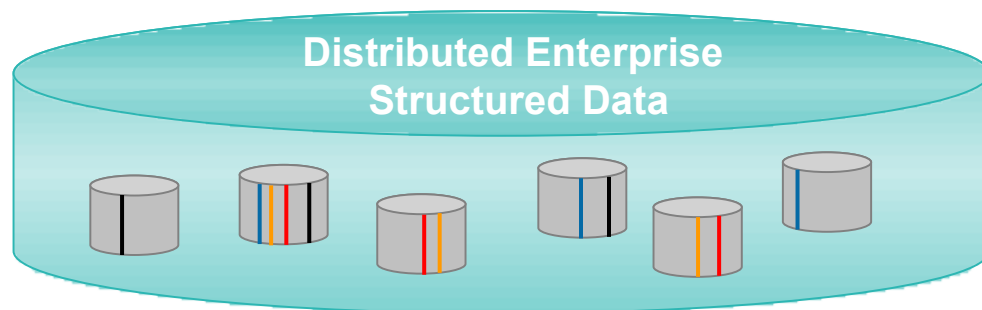
- Prototype empty targets from the combination of many data sources

### • Transformation Analyzer:

- Discover complex business rules and transformation logic between two data sources



# InfoSphere Discovery Cross-source Profiling

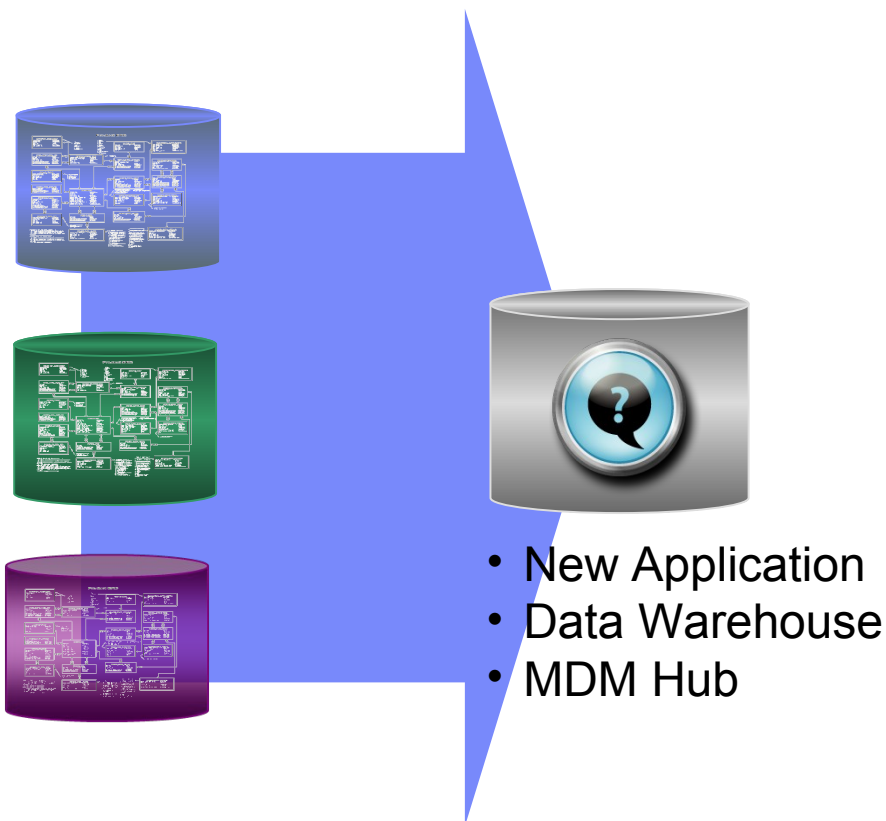


## What is unique?

Only solution on the market that automatically discovers primary foreign keys, business entities, and performs cross-source analysis

- **Cross-Source Profiling:**
  - Data profiling and cross-system overlap analysis of up to 20 systems simultaneously
  - Automated PF Key and Business Object discovery
  - Extremely easy to install and use
- **Value for Optim:**
  - Competitive Differentiator
  - Deployment Accelerator - faster time to value
- **Applicability**
  - Data Archiving
  - Test Data Management
  - Sensitive Data Discovery

# InfoSphere Discovery – Prototyping a Unified Schema

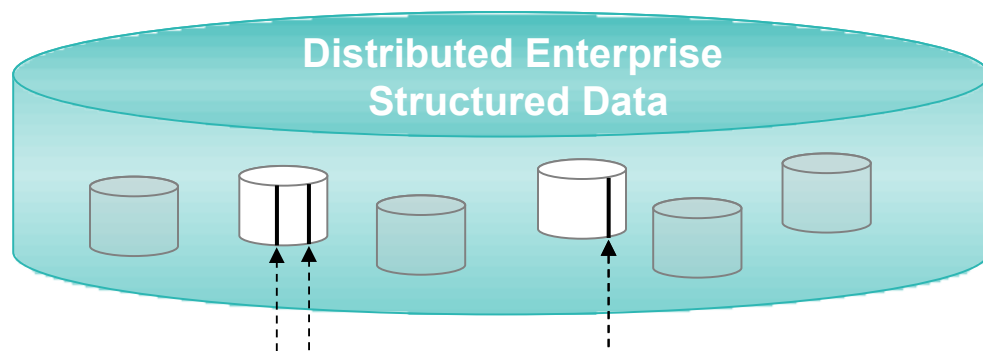


## What is unique?

Prototypes empty targets from existing source data (MDM, EDW, data migration)

- **Unified Schema Builder:**
  - Data analyst workbench for data consolidation projects
    - Profile data sources
    - Perform overlap analysis
    - Unify data models
    - Analyze Matching keys
    - Propose conflict resolution precedence
  - Cross source trouble-shooting workbench
- **Value for Optim**
  - Application/Data Consolidation, Migration & Retirement

# InfoSphere Discovery Transformation Analyzer



If age < 18 and Sex = M then 0  
 If age < 18 and Sex = F then 1  
 If age >= 18 and Sex = M then 2  
 If age >= 18 and Sex = F then 3

} = Demo1

## What is unique?

Discovers cross-system business rules, transformations and data exceptions by examining data values

## • Transformation Analyzer:

- Automates discovery of:
  - cross-system business rules and transformations
  - data inconsistencies
- Detailed data mapping between 2 data sources
- Discrepancy discovery
- Cross source trouble-shooting workbench

## • Value for Optim:

- Discovery cross-source business rules and transformations for data consolidation

# Transformation and Complex Business Rules Discovery

- **Scalar**
  - One to one
  - Substring
  - Concatenation
  - Constants
  - Tokens
- **Conditional Logic**
  - Case Statements
  - Equality/Inequality
  - Null Conditions
  - In/Not In
  - Conjunctions
- **Joins**
  - Inner
  - Left Outer
- **Aggregation**
  - Sum
  - Average
  - Minimum
  - Maximum
- **Column Arithmetic**
  - Add
  - Subtract
  - Multiply
  - Divide
- **Reverse Pivot**
  - **Cross-Reference**
  - **Custom Data Rules**



```

CASE WHEN HQ_EMP.STATUS in
('Current', 'Fired', 'Resigned')
THEN HQ_EMP.TERMINATION_DATE ELSE
HQ_EMP.RETURN_DATE END
    
```

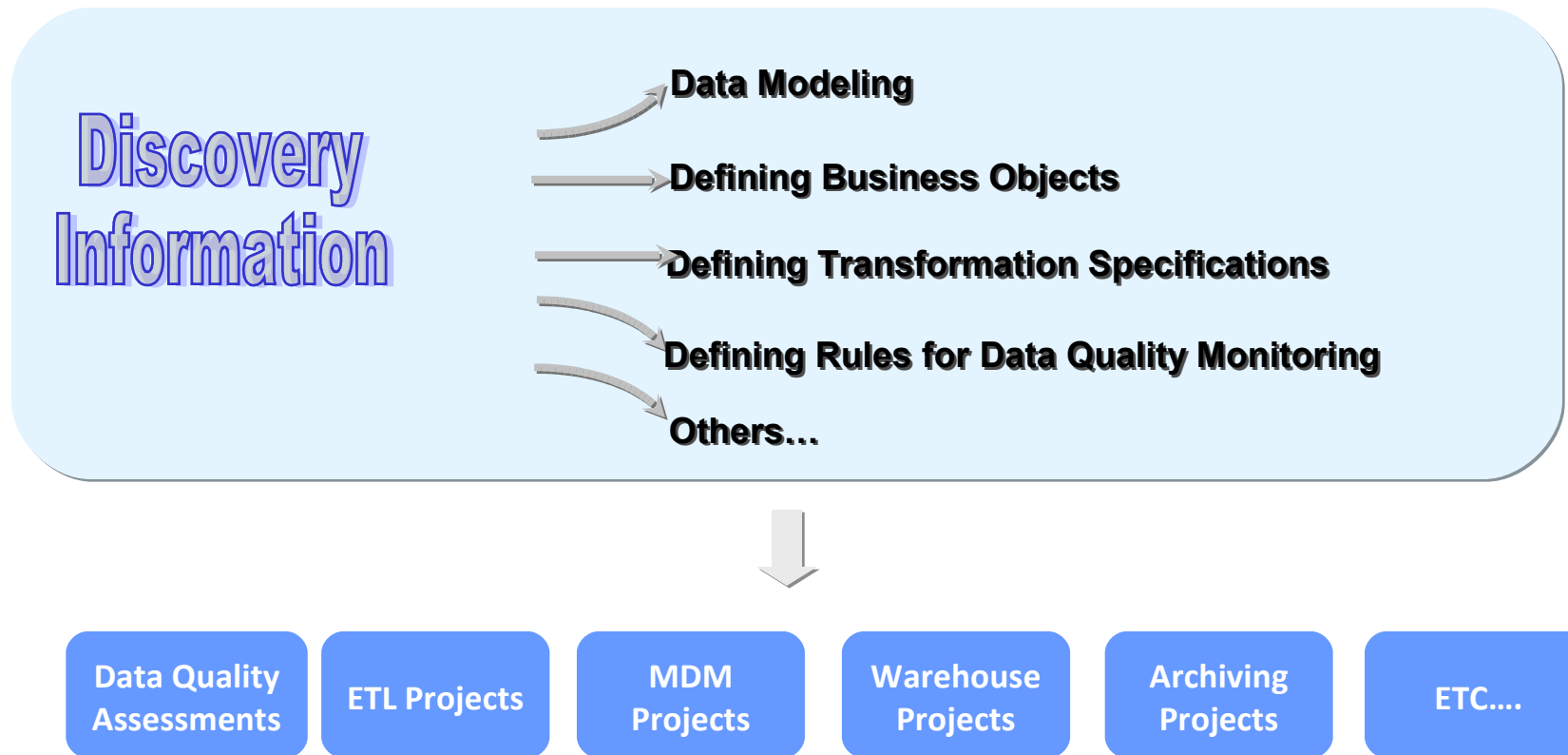
The screenshot shows the Mapping Studio interface for a map named 'DOMap\_DOMap1\_DO\_HQ\_EMP\_to\_DO\_W\_EMPS1\_W\_EMPS'. The 'Transformations' tab is active, displaying a table of target columns and their primary transformations. The 'END\_DATE' transformation is highlighted with a red box, showing the following logic:

```

CASE WHEN HQ_EMP.STATUS in ('Current', 'Fired', 'Resigned') THEN HQ_EMP.TERMINATION_DATE ELSE
HQ_EMP.RETURN_DATE END
    
```

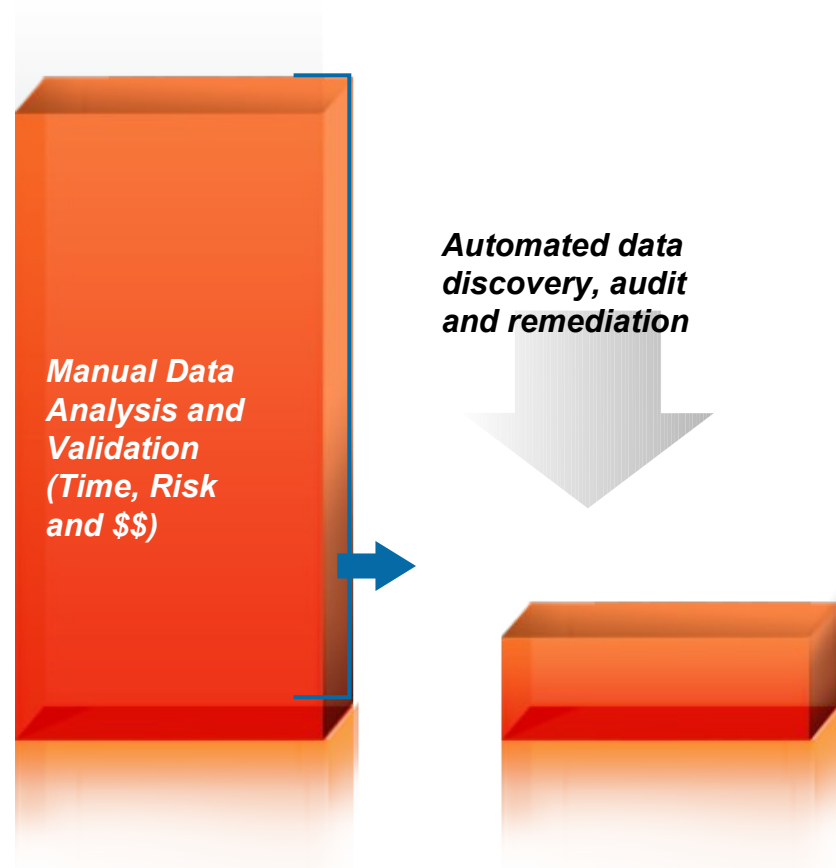
The interface also includes a 'Columns' pane on the right showing source columns for 'HQ\_EMP' and buttons for 'Validate Step', 'Approve Step', 'Re-Run Step...', and 'Run Next Steps...'.

# The Starting Point ....



## Value to Our Customers

- **Accelerate understanding your existing distributed data landscape for:**
  - Data Archiving
  - Test Data management
  - Sensitive Data
  - Application/Data Consolidation, Migration and Retirement
  - Master Data Management and Data Warehousing
- **10x reduction in risk, time and effort for the discovery phase of your project:**
  - Automated discovery of business entities, cross-source business rules & anomalies
  - Increased repeatability
  - Verifiable results





Thank  
YOU

