



IBM Cognos Infoseminar
InfoSphere - Für vertrauenswürdige Informationen
17. Juni 2009

Lothar Vogelbacher, Technical Pre-Sales IBM Software Group Switzerland

Lothar.vogelbacher@ch.ibm.com



#### **Customer Business Issues**



#### Too much information and not knowing what's important

- Not using demand signals to drive supply chain
- Not using customer analysis to tailor marketing and sales
- Not leveraging valuable unstructured information



#### Multiple versions of the truth

- Problems managing customer, product and partner interactions
- Regulatory compliance inhibited by poor transparency



#### Lack of trusted information

- Incomplete, out-of-date, inaccurate, misinterpreted data
- Difficult to understand or control how information is used



#### Lack of agility

- Inability to take advantage of opportunities for innovation
- Escalating costs due to inflexible systems and changing needs



# Unlocking the Business Value of Information

Financial Risk Insight Workforce Optimization

Dynamic Supply Chain

Customer & Product Profitability

**Business Optimization** 

Multi-channel Marketing





Business Intelligence& Performance Management





Information Integration, Warehousing, & Management



Data Management

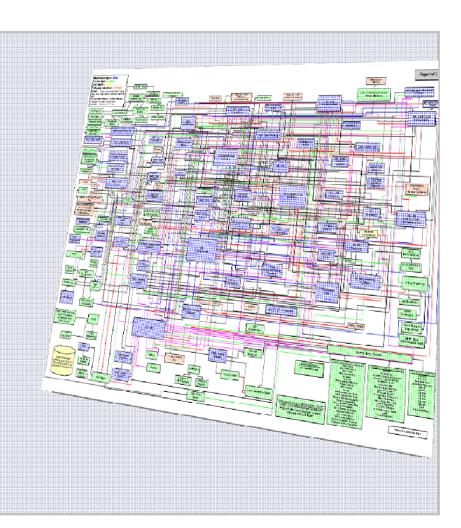


Information Assets



#### **Transformation Pain**

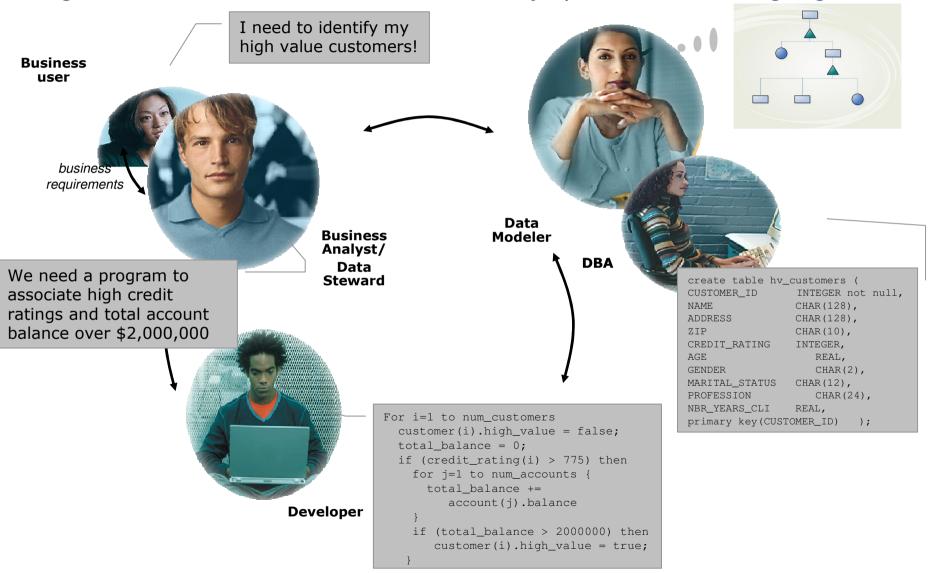
- Multiple sources for the same entity
- Lack of standards or consistent semantic meanings across systems
- Embedded business intelligence
- Evolving transformation requirements
- Need for batch and real-time and service oriented architectures
- Extreme data volumes!
- Business rules for resolving data conflicts
- Ownership and accountability
- Zero re-use of skills and processes



### Information Management Software

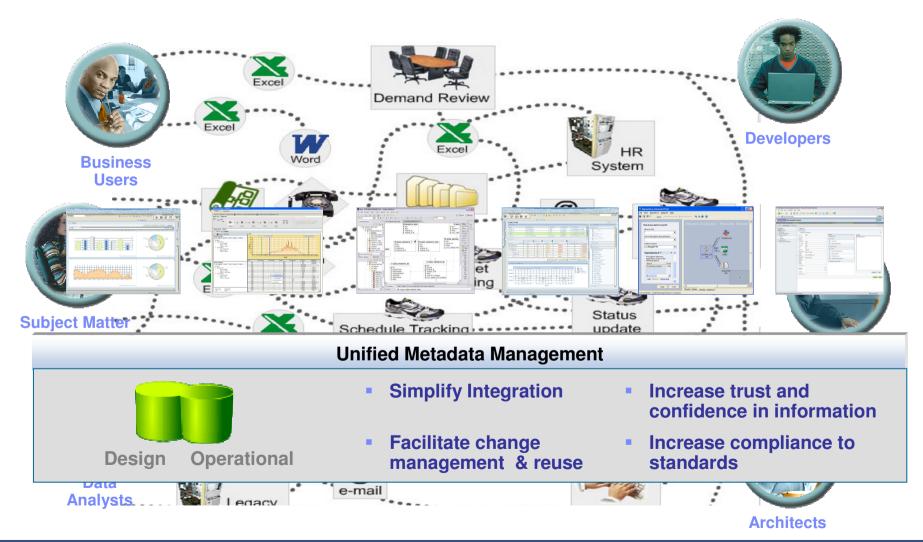


### Integration is a collaborative effort ...but they speak different languages ...





#### Role Based Tools with Shared Metadata

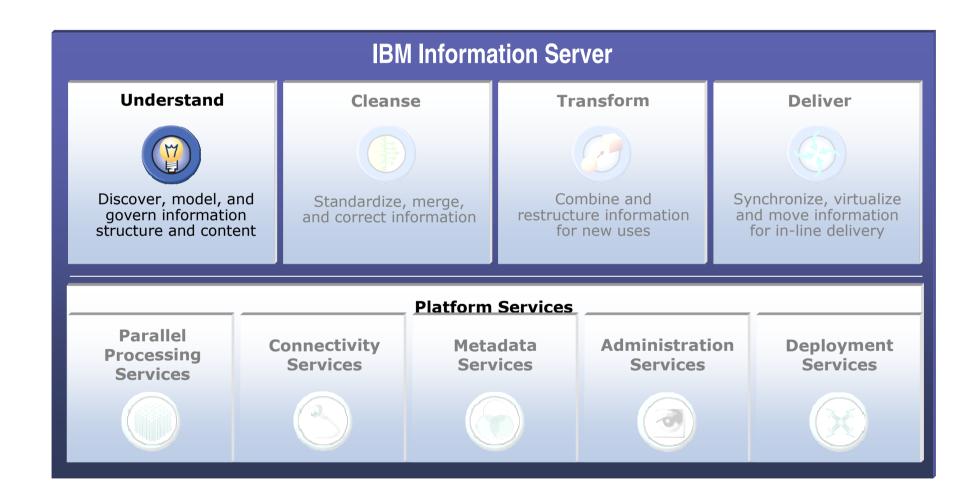




We have solutions for the collaborators... business terms **Business Analyst** Data Modeler **DBA** business rules data model data model ·extract ·transform load data flow application



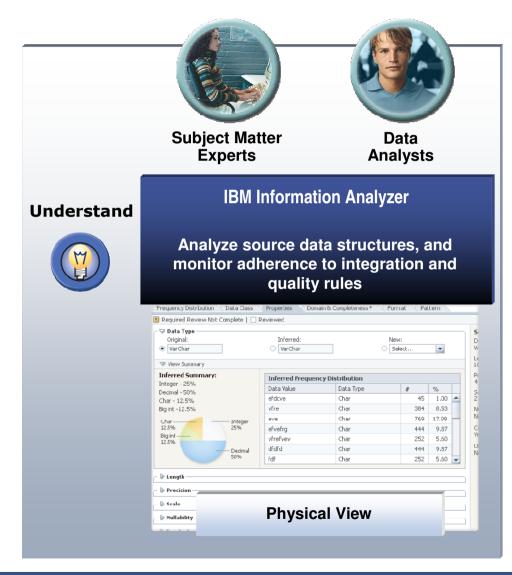
# The IBM Solution: IBM Information Server Delivering information you can trust





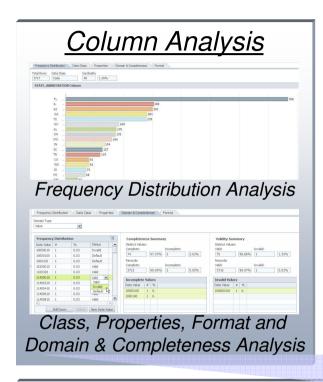
# Data Profiling & Monitoring: Information Analyzer

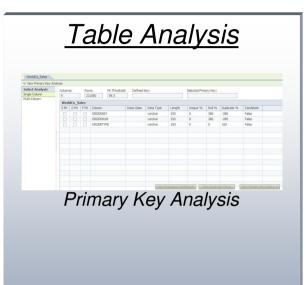
- Data-centric analysis of application, database and filebased sources for content, quality, and structure
- Secure, detailed profiling of fields, across fields, and across sources
  - Column Analysis
  - Primary Key Analysis
  - Foreign Key Analysis
  - Cross-Domain Analysis
  - Baseline Analysis
- Creation of metadata and reports from profiling results
- Results instantly promotable across IBM Information Server

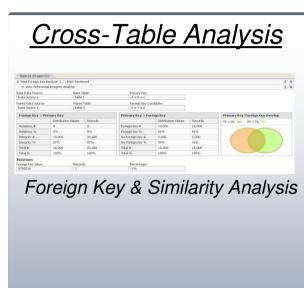


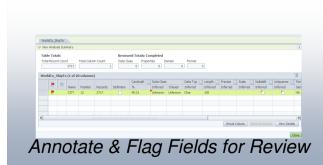


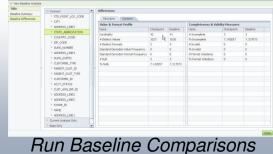
## How Does InfoSphere Information Analyzer Work?

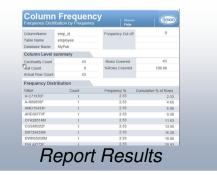












#### Information Management Software



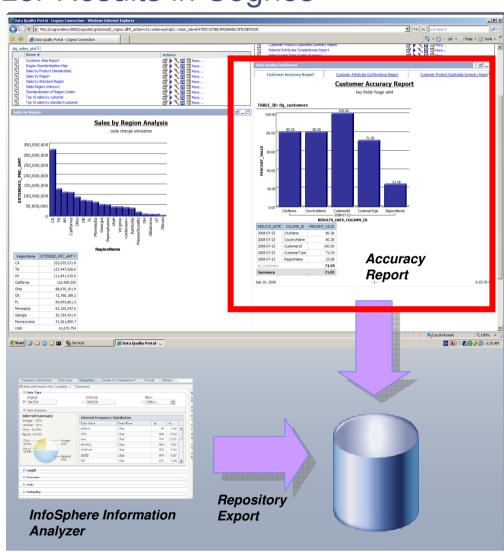
# Data Quality Dashboard: InfoSphere Information Analyzer Results in Cognos

#### Value

- Provides a data source of Information Analyzer metrics
- Enables business users and BI analysts to understand the health of their data
- Information conformance, completeness and validity in Cognos 8 BI

#### Integration

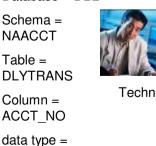
- Regular exports of Information Analyzer repository into reporting data mart
- Services offering includes IA installation, and templates for loading reporting data mart, and sample Cognos reports





# Business Metadata: InfoSphere Business Glossary

- Enables Web-based authoring, sharing & managing of business metadata
- Aligns the efforts of IT with the goals of the business
- Provides business context to information technology assets
- Establishes responsibility and accountability



Database = DB2





Technical

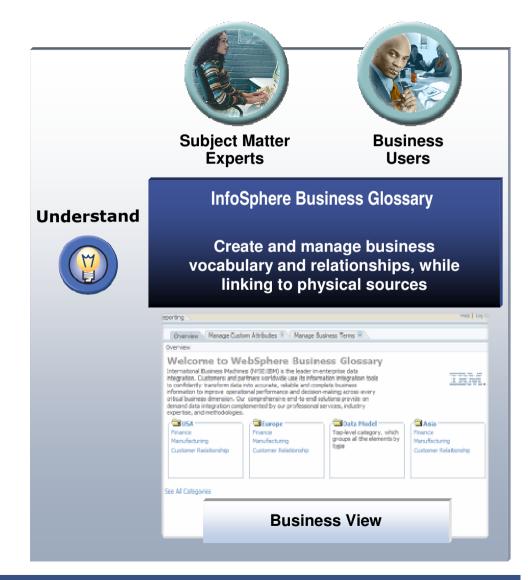
Business

The ten digit account number. Sometimes referred to as the account ID. This value is of the form L-

FIIIIVVVV.

**GL** Account

Number



char(11)



#### **Business Terms**

- Author & manage standard business terms (a controlled vocabulary)
- Agreed upon standard word/definitions (preferred term)
- Including examples, abbreviations, stewardship, synonyms, & related terms
- Connect them to more technical artifacts in the metadata repository
- Creates a translating semantic layer between the user's natural language and the highly structured information in the repository
- Import from XML

Account Number - The ten digit account number. Sometimes referred to as the account ID. This value is of the form L-FIIIIVVVV.

Owned By: Controller's Office, Mary Smith X3256

Synonyms: Account ID

See Also: Account Type



# **Business Categories**

- Build & manage business hierarchies (taxonomies)
- Use to categorize business terms
- Defines the browsing structure of the glossary
- Organizes search results
- Import from XML

### Finance

- United States
  - General Ledger
    - Account Number
    - Account Type



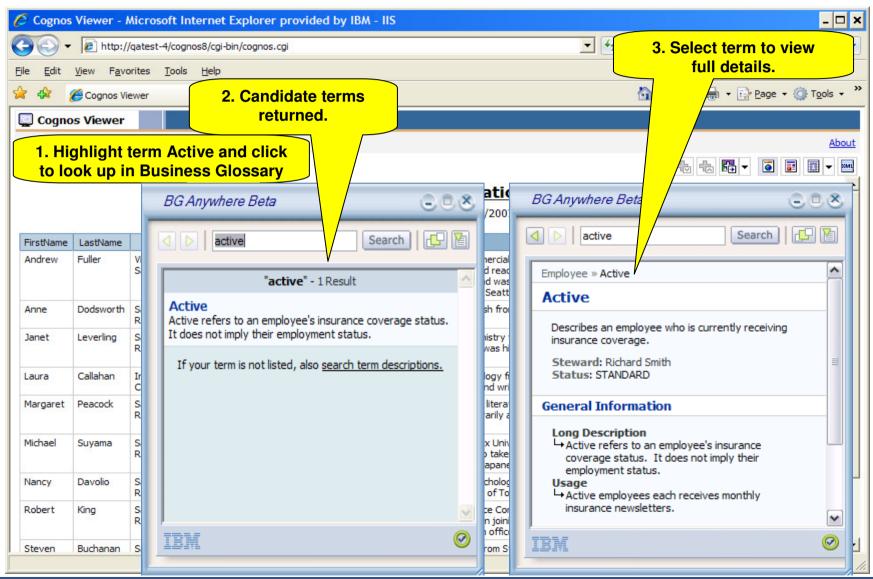
## Collaborative Authoring of Business Metadata

- Add Notes/Annotations to any object in the glossary
- Enables business users to enrich the metadata associated with information assets
- Similar to the user feedback section on many online storefront sites
- Immediately consumable by other users
- Monitored by administrator
- Provide feedback to system administrator

I've noticed that the last two digits of the GL Organizational Unit, which indicate the sub-department, are often blank.



# Business Glossary Anywhere with Cognos





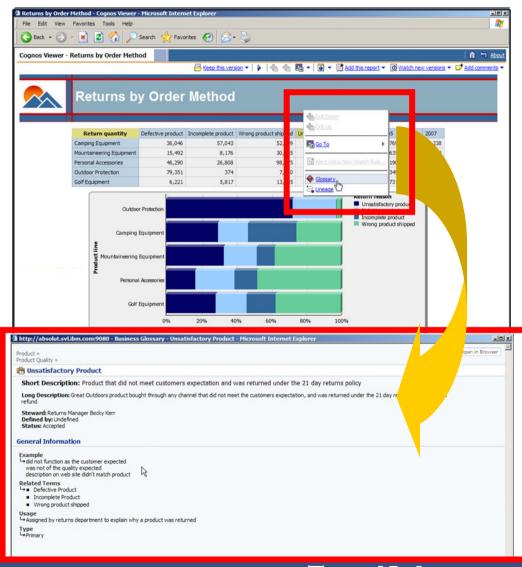
### Integration with Cognos: InfoSphere Business Glossary

#### Value

- Immediate access to what their information means and who can help
- Available from any Cognos report

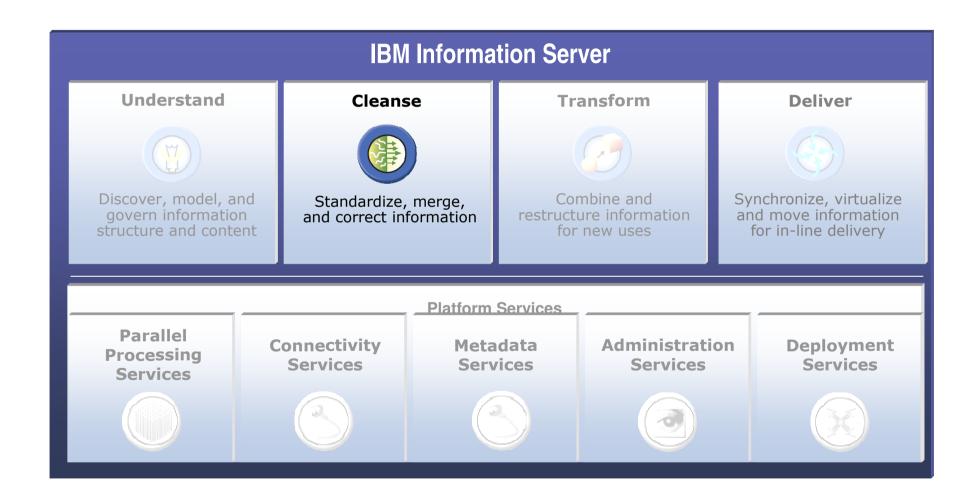
#### Integration

- Right click options for business glossary
- New Business Glossary web-client with Business Glossary Anywhere-like functionality





# The IBM Solution: IBM Information Server Delivering information you can trust



#### Information Management Software



# Typical Data Quality Challenges

Lack of information standards

 Different formats & 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

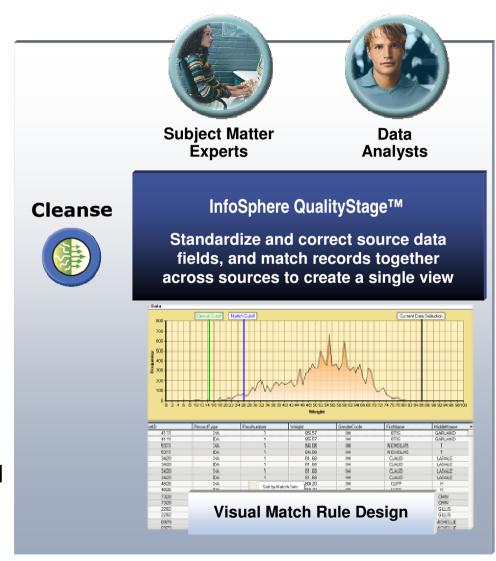
- The redundancy nightmare
  - Duplicate records with a lack of standards

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



# What is InfoSphere QualityStage?

- Provides specialized data quality processing
  - Ensures clean, standardized, deduplicated information
  - Enables a single version of the truth
  - Supports global postal verification
- Provides visual tools for designing quality rules and matching logic
  - Seamlessly integrated with DataStage (one engine, one metamodel, one UI)
  - Precisely calibrates matching rules
- Allows quality logic to be deployed seamlessly within ETL, or as shared services





# How Does QualityStage Work?



#### **Cleansing Process**

- 1. Data Investigation
- 2. Data Standardization
- 3. Data Matching
- 4. Data Survivorship



Accurate, cleansed data that drives critical decisions



Investigation - Word

123 St. Virginia St.

Parsing:

123 | St. | Virginia | St.

Separating multi-valued fields into individual pieces

Lexical analysis:

Type Type 123 | St. | Virginia | St.

Alpha

Determining business significance of individual pieces

Context Sensitive: 123 | St. Virginia | St.

House Street Number Street Name Type

Street

Identifying various data structures and content

"The instructions for handling the data are inherent within the data itself."

Number

Street



### Standardization – Name & Address

Address Line 1						Address Line 2								
639 N MILLS AVENUE						ORLANDO	, FLA 32803	3						
306 W MAIN STR, CUMMING, GA 30130														
3142 WEST CENTRAL AV						TOLEDO C	H 43606							
843 HEAR	D AVE					AUGUSTA-	GA-30904							
1139 GRE	ENE S	T ACC	T #12	234		AUGUSTA GEORGIA 30901								
		DAD SUI			_	GA 30809								
1775 RUS	SELL (	CIRCLE MIL	1775 RUSSELL CIRCLE MILLIS MASSACH					USETTS 02038						
Posult F	ilo													
Result Fi	i <b>le:</b> Dir	Str. Name	Туре	Unit	No.	NYSIIS	City	SOUNDEX	State	Zip	ACCT#			
		Str. Name	Type AVE		No.	<b>NYSIIS</b> MAL			<b>State</b> FL	<b>Zip</b> 32803	ACCT#			
House #	Dir				No.		City	O645			ACCT#			
<b>House</b> # 639	Dir N	MILLS	AVE ST		No.	MAL	<b>city</b> ORLANDO	O645	FL	32803	ACCT#			
House # 639 306	Dir N W	MILLS MAIN	AVE ST		No.	MAL MAN	city ORLANDO CUMMING	O645 C552	FL GA	32803 30130	ACCT#			
House # 639 306 3142	Dir N W	MILLS MAIN CENTRAL	AVE ST AVE		No.	MAL MAN CANTRAL	City ORLANDO CUMMING TOLEDO	O645 C552 T430	FL GA OH	32803 30130 43606				
House #  639  306  3142  843	Dir N W	MILLS MAIN CENTRAL HEARD	AVE ST AVE AVE			MAL MAN CANTRAL HAD	City  ORLANDO CUMMING TOLEDO AUGUSTA	O645 C552 T430 A223	FL GA OH GA	32803 30130 43606 30904				

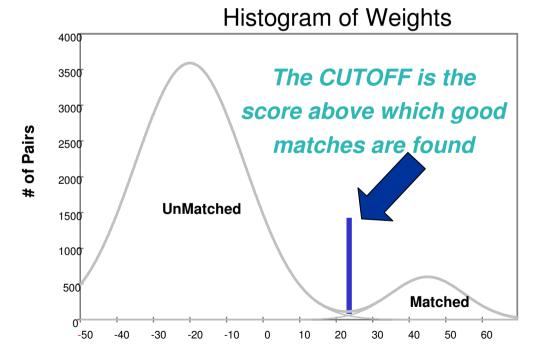
Results in strongly "typed" fixed fielded standardized data



# A Closer Look at Probabilistic Matching

WILLIA	M J	HOLDEN	128	MAIN	ST	02111	12/8/62	
WILLAI	M JOHN	HOLDEN	128	MAINE	AVE	02110	12/8/62	
+5	+2	+14	+5	+4	-1	+5	+11	= 49

The weighted score is a relative measure of the probability of a match; it expresses the amount of information content for all of the fields compared





# Survivorship – Name and Address

### Survivorship Input (Match Output)

1	A1367	Robert		Dickson	1500	JL	ROSS CLARK	CIR		
23 23 23	D689 A436 D352	William Billy William	A Alex	Obrian O'Brian Obrian	5901 5901 5901	SW SW	74TH 74TH 74	ST ST ST	STE #	202 202

#### Cross Populated Output

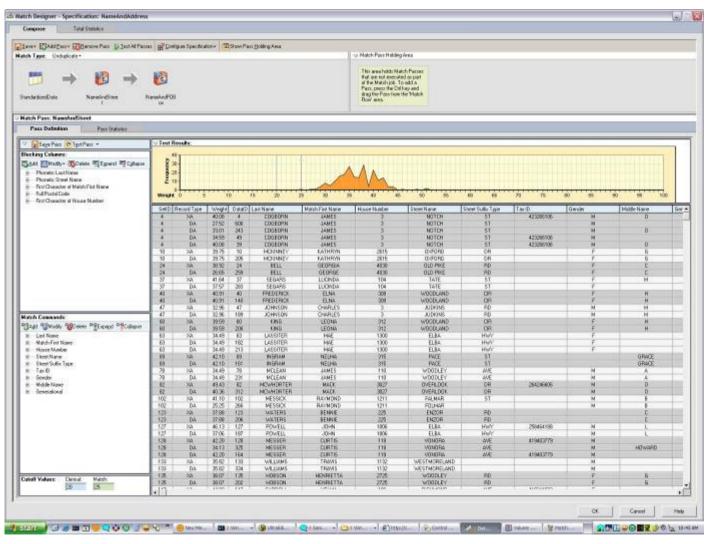
	Group	Legacy	First	Middle	Last	No.	Dir.	Str. Name	Туре	Unit	No.
	1	D150	Robert		Dickson	1500	SE	ROSS CLARK	CIR		
	1	A1367	Robert		Dickson	1500	SE	ROSS CLARK	CIR		
П											
	23	D689	William	Alex	O'Brian	5901	SW	74TH	ST	STE	202
ľ	2	1126	William	^ ! -	0/2	F001	CVV	7471	CT	STE	202
				Alex	O'Brian		SW	74TH		STE	

#### Consolidated Output

Group	Legacy	Group	First	Middle	Last	No.	Dir.	Str. Name	Туре	Unit	No.
1	D150 A1367	1	Robert		Dickson	1500	SE	ROSS CLARK	CIR		
23 23	D689 A436	23	William	Alex	O'Brian	5901	SW	74TH	ST	STE	202
23	D352										



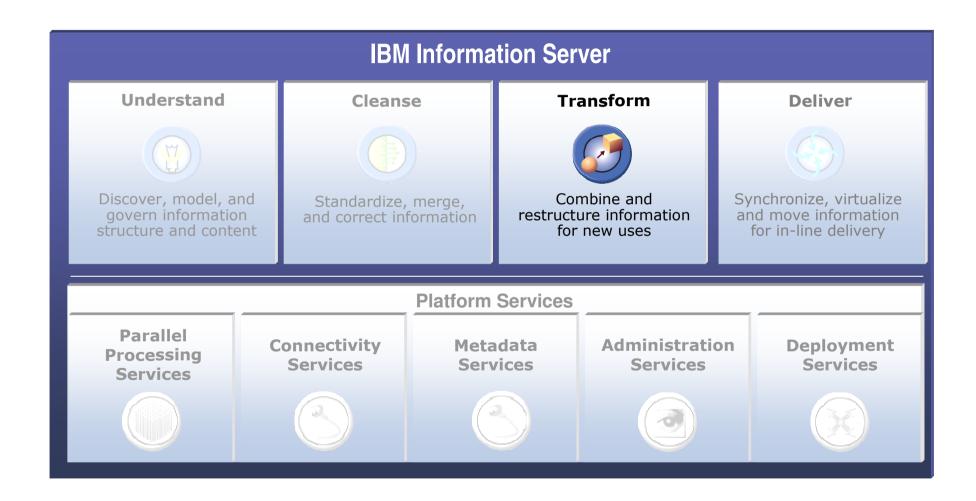
# **Industry Leading Parsing & Matching**



- Probabilistic Matching Engine
- Intuitive match designer, with data sampling, visual fine tuning and extensive reporting
- Parsing handles any number of fields & freeform fields
- Unified parallel processing framework



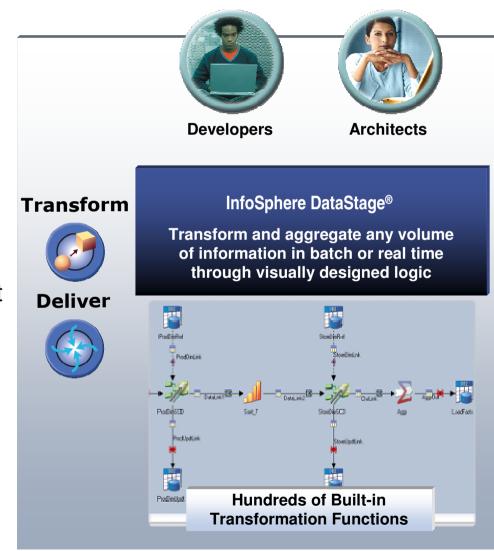
# The IBM Solution: IBM Information Server Delivering information you can trust





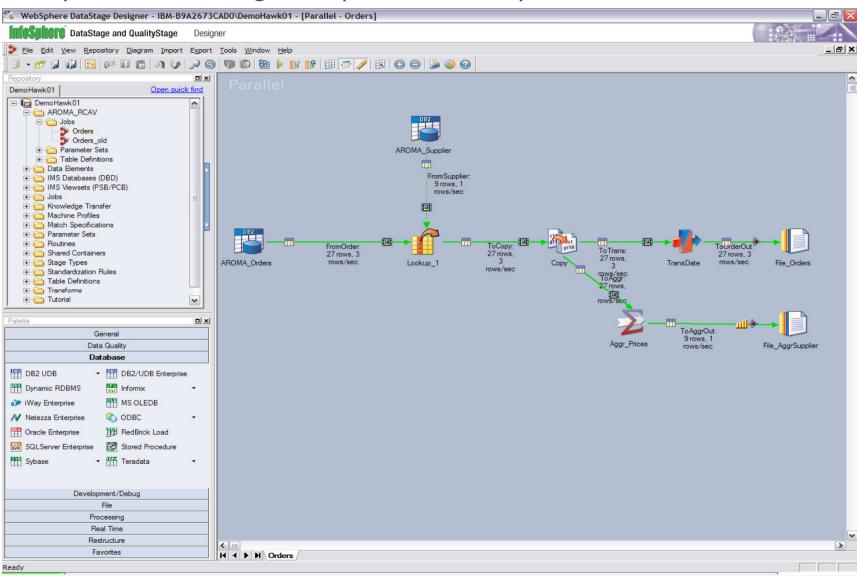
# What is InfoSphere DataStage?

- Provides codeless visual design of data flows with hundreds of built-in transformation functions
  - Optimized reuse of integration objects
  - Supports batch & real-time operations
  - Produces reusable components that can be shared across projects
- Complete ETL functionality with metadata-driven productivity
- Supports team-based development and collaboration
- Provides integration from across the broadest range of sources



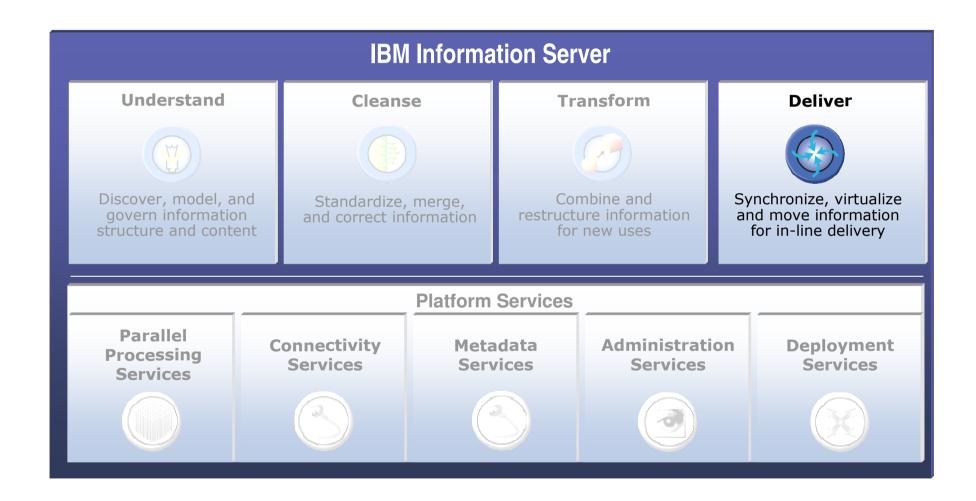


## InfoSphere DataStage Graphical Development Interface





# The IBM Solution: IBM Information Server Delivering information you can trust

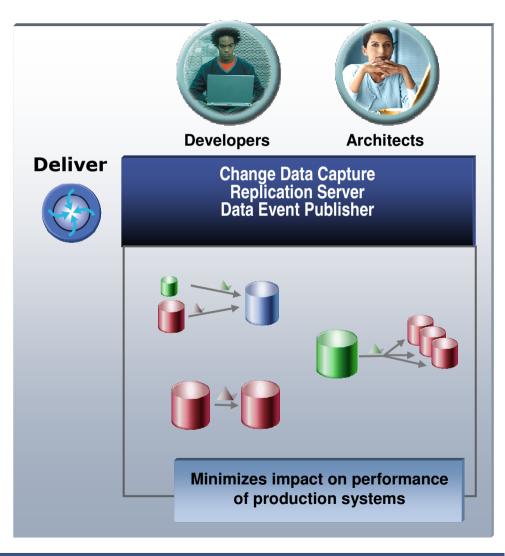


#### Information Management Software



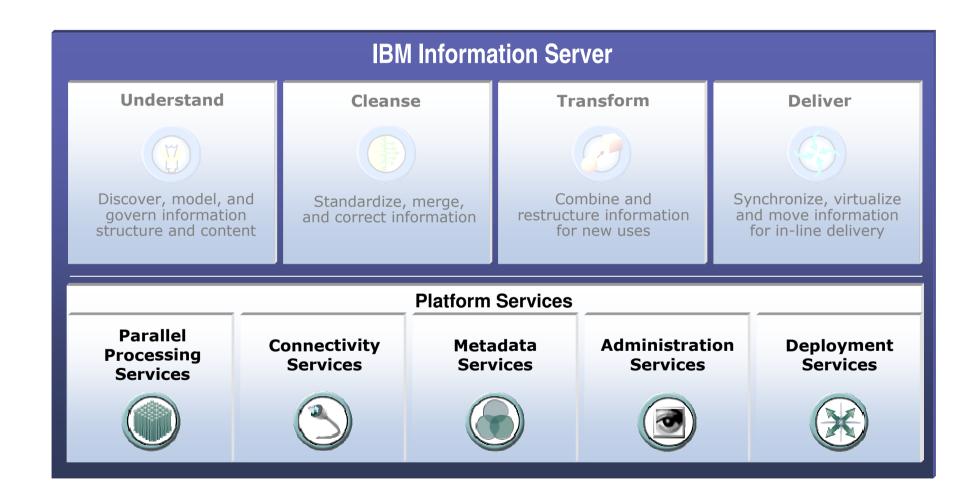
# Change Data Capture and Replication

- Provides real time changed-data capture and delivery for
  - Dynamic warehousing, eBusiness
  - Synchronization
  - Replication
- Provides high-volume, low-latency replication for
  - Business continuity
  - Workload distribution
  - Business integration scenarios
- Minimal impact on production systems thanks to log-based capturing
- High scalability and end-to-end performance
- Wide breadth of RDBMS support
  - Captures for DB2 (i/z/LUW), Oracle, SQL Server, Sybase and more....
- Target integration with DataStage/ QualityStage and WebServices





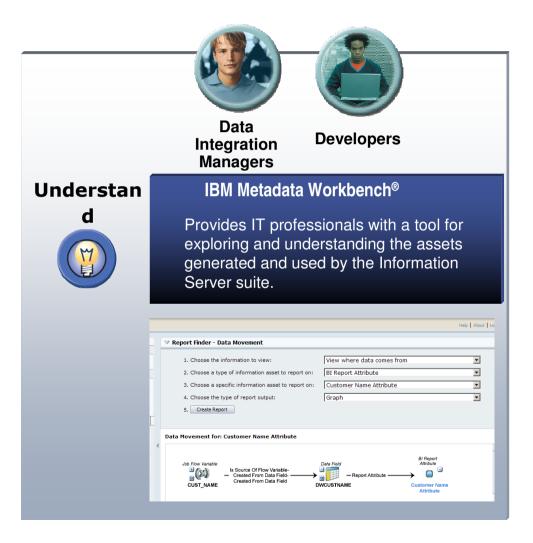
# The IBM Solution: IBM Information Server Delivering information you can trust





# IBM InfoSphere Metadata Workbench

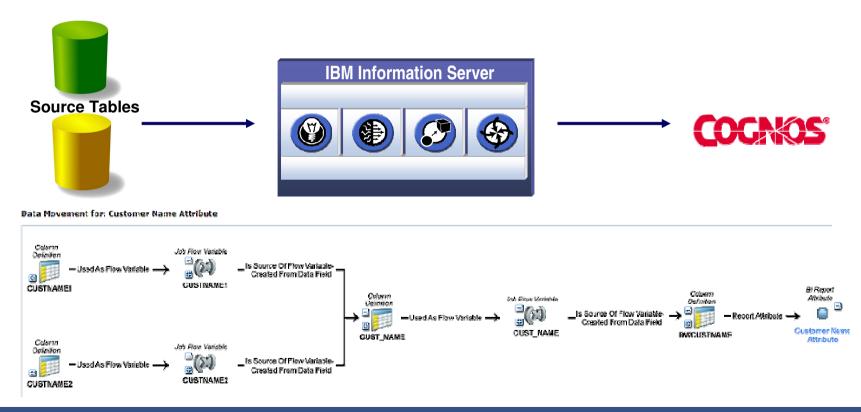
- Web-based exploration of Information Assets generated and used by Information Server applications
- Out of the box reporting on data movement, data lineage, business meaning, impact of changes and dependencies
- Tracing the data lineage of Business Intelligence Reports to provide basis for compliance with legislation such as Sarbanes-Oxley and Basel II





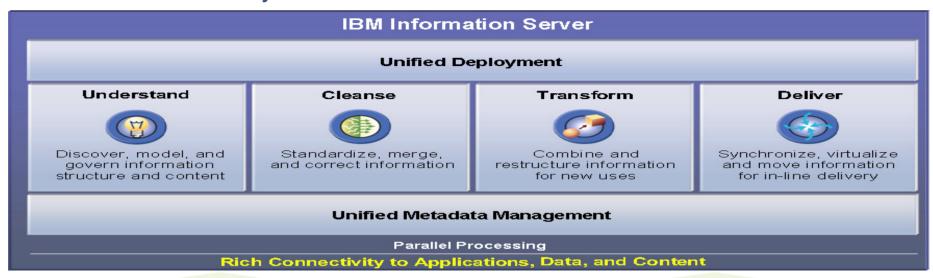
## Where does a field of data in this report come from?

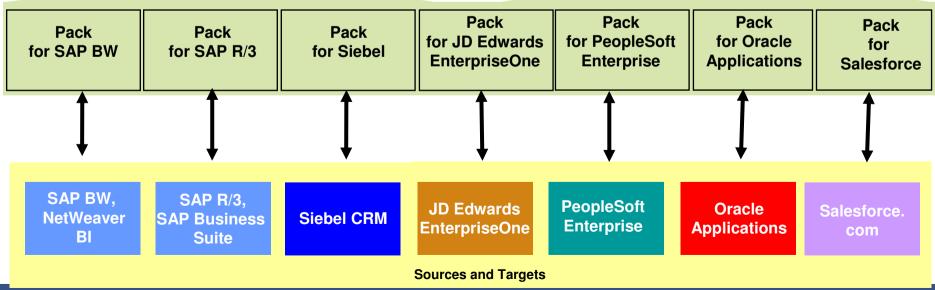
- Import & Browse Full BI Report Metadata
- Navigate through report attributes
- Visually navigate through data lineage across tools
- Increases trust and understanding of business information





### Seamless connectivity to Information Server





#### Information Management Software



Rapid SOA Deployment:
InfoSphere Information Services Director

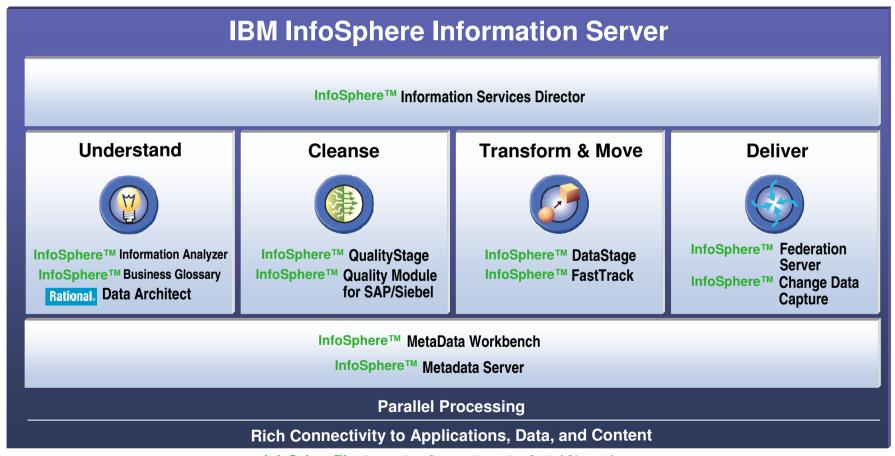
- Packages information integration logic as services that insulate developers from underlying sources
- Allows these services to be invoked as Enterprise Java Beans or Web services
- Provides load balancing & fault tolerance for requests across multiple Information Servers
- Provides foundation infrastructure for Information Services





### **IBM Information Server**

Delivering information you can trust

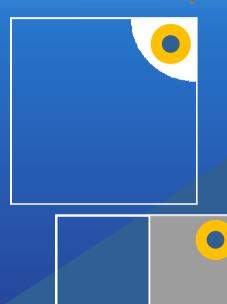


InfoSphere™ Information Server Pack for SAP / Siebel / ...



## Leading the way

# Industry Analysts Rank IBM as Leaders ....



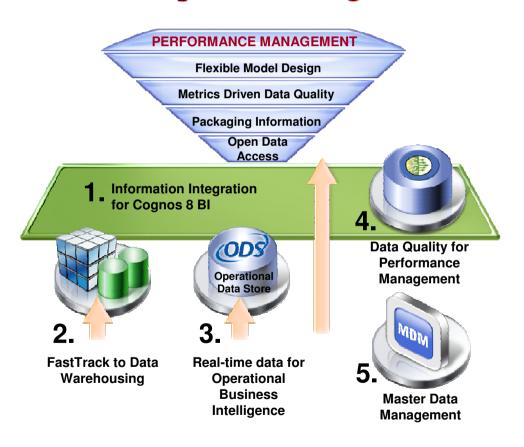
and many other firms...

- ETL Tools
- Business Intelligence Platforms
- Data Warehousing
- Data Quality Tools
- Data Integration Tools
- Customer Data Integration



# Cognos and IBM InfoSphere Information Server

#### Cognos 8 Cognos Now! Cognos Solutions



- Goal: Performance Management with Information you can Trust
- Fact: Performance Management is reliant on trusted information

 Fact: Maximizing the value of your trusted information requires Performance Management

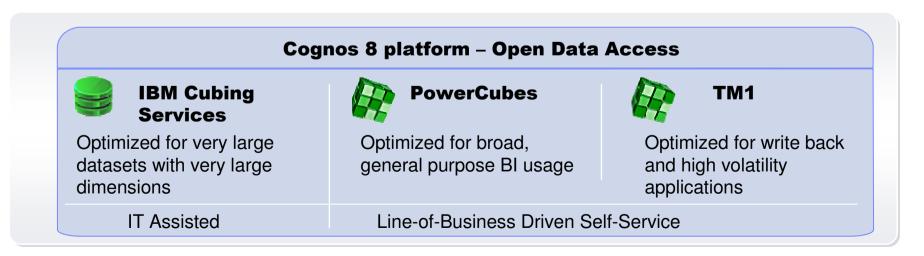


# Integration with InfoSphere Warehouse OLAP

**Challenge:** Deliver data optimized for analysis

#### **Solution:**

- InfoSphere Cubing Services as a data source
- Embed IBM Cognos OLAP technology into InfoSphere Warehouse to accelerate Cubing Services development





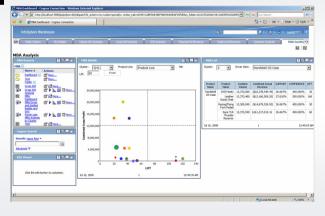
# Integration with InfoSphere Warehouse Data Mining and Text Analytics

Challenge: Enable more sophisticated and predictive BI applications

#### **Today**

 Deliver data mining results to a broader audience of users with Cognos 8 BI

 Text analytics to transform text to structured data for reporting and analysis with Cognos 8 BI



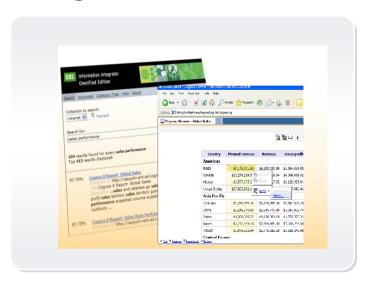


# Integration with OmniFind Search

**Challenge:** Discover BI content through search

#### **Today**

- Enable OmniFind to find and display Go! Search results as part of enterprise search results
- Expose OmniFind content results within Cognos Go! Search





# Accelerated End-to-End Deployment

