



# IBM Information Server



14 Aprile 2008

Walter Aglietti

IBM SWG

Information Platforms & Solutions

**ON** DEMAND BUSINESS™

© 2007 IBM Corporation

# IBM Information Server

*Fornire informazioni affidabili*

## IBM Information Server

### Deployment unificato

#### Comprendere



Scoprire, modellare e gestire la struttura ed il contenuto dell'informazione

#### Bonificare



Standardizzare, riconciliare e correggere le informazioni

#### Trasformare



Combinare e ristrutturare le informazioni per nuovi utilizzi

#### Federare



Sincronizzare, spostare e virtualizzare le informazioni on-line

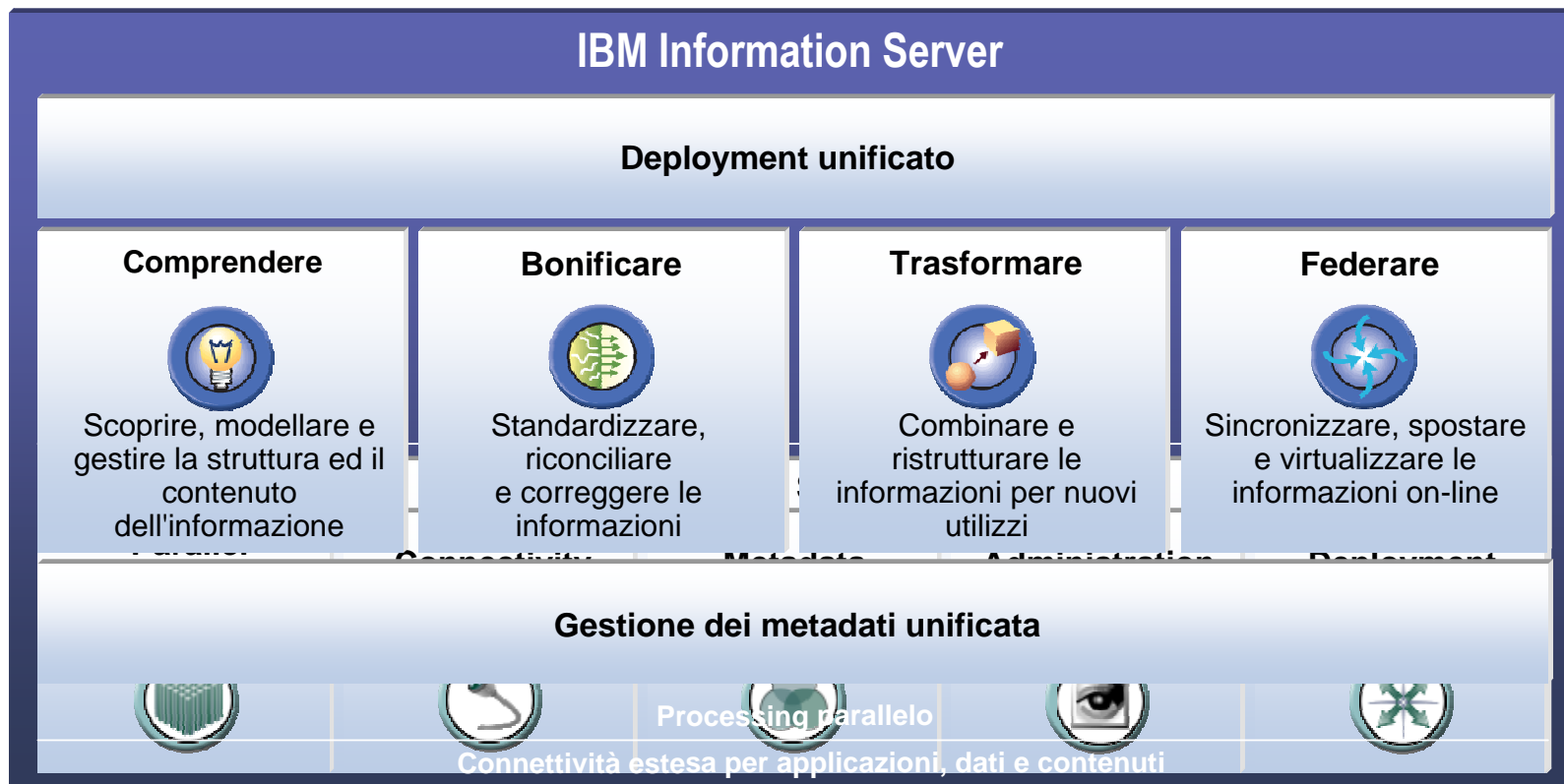
### Gestione dei metadati unificata

#### Processing parallelo

Connettività estesa per applicazioni, dati e contenuti

# IBM Information Server

*Fornire informazioni affidabili*



# Cos'è WebSphere Information Analyzer?

comprendere

- Fornisce un'analisi dettagliata dei sistemi esistenti**

- Analisi focalizzata su dati di applicazioni, database e files, per contenuto, qualità e struttura
- Affidabile e dettagliato profiling di campi e relazioni attraverso strutture e sorgenti multiple

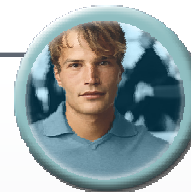
- Consente la misurazione ed il reporting della qualità dell'informazione**

- Crea i metadati che descrivono dove l'informazione è gestita attraverso i sistemi coinvolti**

- Consentendo l'analisi di specifiche fonti si mettono in evidenza i punti che costituiscono potenziali criticità



**Subject Matter Experts**



**Data Analysts**

**Comprendi**



## IBM Information Analyzer

Analizza strutture dati e ne controlla l'aderenza alle regole di integrazione e qualità

**Inferred Summary:**

Integer	25%
Decimal	50%
Char	12.5%
Big Int	12.5%

**Inferred Frequency Distribution:**

Data Value	Data Type	#	%
afdvce	Char	45	0.00
vfrs	Char	364	0.53
evr	Char	769	17.09
afvafvg	Char	444	9.87
vfrvfrv	Char	252	5.60
dfvdf	Char	444	9.87
fdf	Char	252	5.60

**View Summary**

Length, Precision, Scale, Reliability

**Vista fisica**

# Come lavora WebSphere Information Analyzer?

comprendere

## Column Analysis



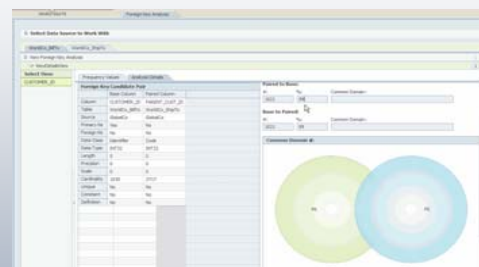
Analisi delle frequenze di distribuzione

Analisi di Classe, Proprietà, Formato, Dominio e Completezza

## Table Analysis

Analisi delle "Primary Keys"

## Cross-Table Analysis



Analisi delle "Foreign Keys" e delle analogie tra colonne

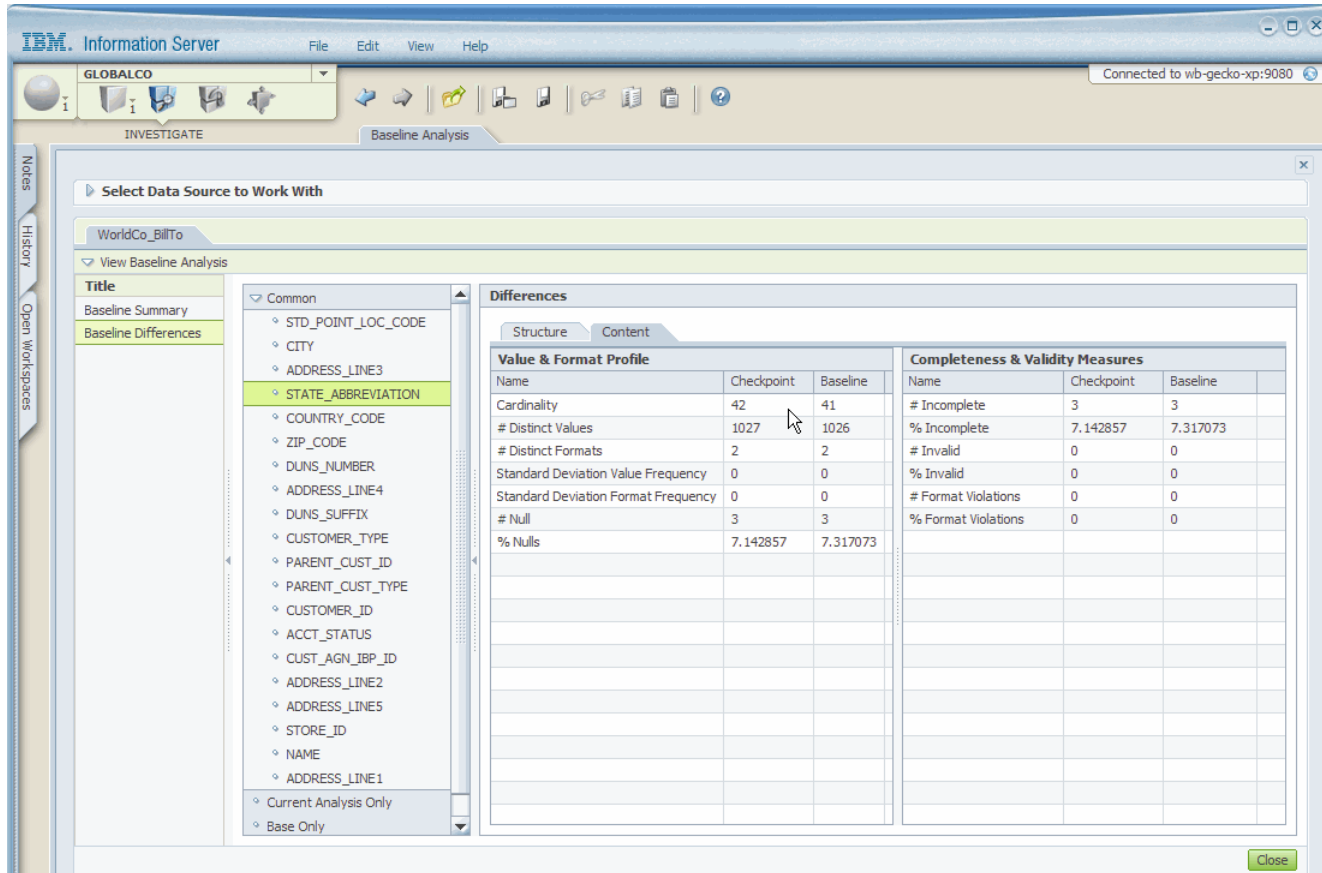
Annotazioni e evidenziazioni dei campi da controllare

Capacità di confronto rispetto a "baselines"

Report dei risultati



# Information Analyzer: misurazione della qualità dei dati



The screenshot shows the IBM Information Analyzer interface. The main window is titled 'WorldCo\_BillTo' and displays a 'Baseline Analysis' for the 'STATE\_ABBREVIATION' field. The interface includes a menu bar (File, Edit, View, Help), a toolbar, and a sidebar with 'Notes', 'History', and 'Open Workspaces'.

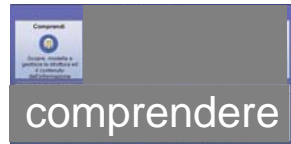
The 'View Baseline Analysis' section shows a tree view of fields under 'Common'. The 'STATE\_ABBREVIATION' field is selected. The 'Differences' section is divided into 'Structure' and 'Content' tabs. The 'Content' tab displays two tables comparing 'Checkpoint' and 'Baseline' data.

Value & Format Profile			Completeness & Validity Measures		
Name	Checkpoint	Baseline	Name	Checkpoint	Baseline
Cardinality	42	41	# Incomplete	3	3
# Distinct Values	1027	1026	% Incomplete	7.142857	7.317073
# Distinct Formats	2	2	# Invalid	0	0
Standard Deviation Value Frequency	0	0	% Invalid	0	0
Standard Deviation Format Frequency	0	0	# Format Violations	0	0
# Null	3	3	% Format Violations	0	0
% Nulls	7.142857	7.317073			

A 'Close' button is located at the bottom right of the window.



# Cos'è il WebSphere Business Glossary?



## Un'applicazione *web-based* che

- Consente la condivisione e la gestione dei metadati di business
- Allinea le procedure IT con gli obiettivi di business
- Fornisce informazioni sul contesto del business agli asset informativi
- Stabilisce responsabilità e competenze sul dato

Database = DB2

Schema = NAACCT

Table = DLYTRANS

Column = ACCT\_NO

data type = char(11)



Technical



Business

GL Account Number

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



Subject Matter Experts



Business Users

Comprendi



## WebSphere Business Glossary

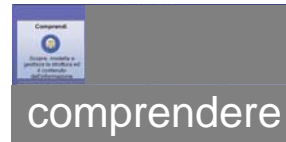
Consente la creazione e la gestione del business vocabulary, delle relazioni e dei collegamenti alle fonti fisiche del dato







# Come lavora WebSphere Business Glossary?



WebSphere Data Integration Console - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://localhost:9080/

**WebSphere. Data Integration Console**

Home Administration Glossary Reporting

**Navigation**

- Browse Search History
- Administration
  - Edit Overview
  - Manage Business Categories
  - Manage Business Terms
  - Manage Custom Attributes

Overview

**Welcome to WebSphere Business Glossary**

International Business Machines (NYSE:IBM) is the leader in data integration. Customers and partners worldwide use its information to confidently transform data into accurate, reliable and complete information to improve operational performance and decision critical business dimension. Our comprehensive end-to-end solution demand data integration complemented by our professional expertise, and methodologies.

**USA**

- Finance
- Manufacturing
- Customer Relationship

**Europe**

- Finance
- Manufacturing
- Customer Relationship

See All Categories

Naviga per Business Category

WebSphere Data Integration Console - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://localhost:9080/

**WebSphere. Data Integration Console**

Home Administration Glossary Reporting

**Navigation**

- Browse Search History
- Administration
  - Edit Overview
  - Manage Business Categories
  - Manage Business Terms
  - Manage Custom Attributes

Overview

Overview > All Categories

<b>Transaction</b> Metadata related to account transactions	<b>USA</b> Finance Manufacturing Customer Relationship	<b>Finance</b> General Ledger Transaction	<b>General Ledger</b> Metadata related to General Ledger accounts
<b>Manufacturing</b> Metadata related to manufacturing	<b>Customer Re...</b> Metadata related to CRM activities	<b>Europe</b> Finance Manufacturing Customer Relationship	<b>Data Model</b> Top level category, which groups all the elements by type
<b>Finance</b> Metadata related to Asian finances	<b>Manufacturing</b> Metadata related to manufacturing	<b>Finance</b> Metadata related to European finance	<b>Manufacturing</b> Metadata related to manufacturing
<b>Customer Re...</b> Metadata related to CRM activities	<b>Asia</b> Finance Manufacturing Customer Relationship	<b>Customer Re...</b> Metadata related to CRM activities	<b>All Terms</b> Special category containing all terms

Naviga per Business Category

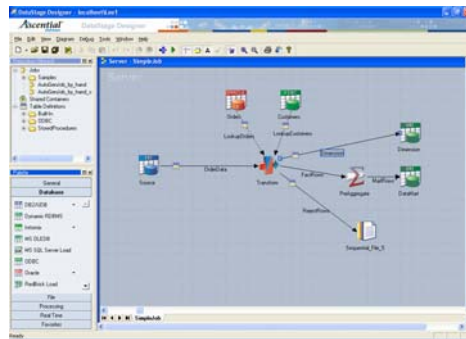
Database = DB2  
 Schema = NAACCT  
 Table = DLYTRANS  
 Column = ACCT\_NO  
 data type = char(11)



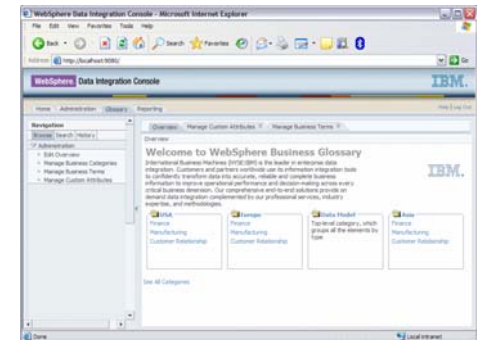
### GL Account Number

Un numero di 10 digit.  
 Compare anche come **account ID**. Pattern: L-FIIIIVVVV.

## Obiettivo: Creare un vocabolario comune tra utenti di business ed utenti tecnici

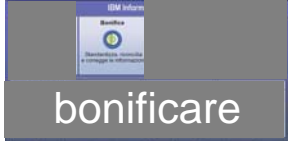


WebSphere DataStage



WebSphere Business Glossary

# Cos'è WebSphere QualityStage?



## Implementa processi specializzati sulla qualità del dato

- Assicura informazioni corrette, standardizzate, de-duplicate
- Fornisce un'unica versione valida del dato
- Supporta gli standard postali worldwide

## Fornisce strumenti visuali per il design delle regole di qualità e la logica di matching

- Completamente integrato con DataStage (un engine, un metamodello, una UI)
- Calibrazione fine delle regole di matching

## Consente il deployment della logica in procedure ETL o come servizi condivisi



Subject Matter Experts



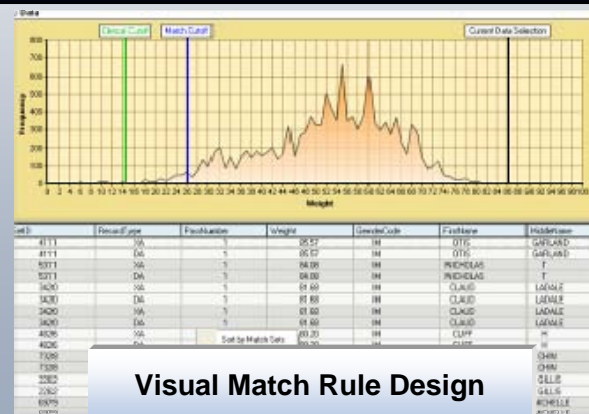
Data Analysts

Bonifica



### WebSphere QualityStage™

Standardizza e corregge dati, consente il matching di record da sorgenti diverse e la creazione di singole viste



Visual Match Rule Design

# Le sfide sulla qualità dei dati

## ■ Mancanza di standard informativi

- Diversi formati e strutture su diversi sistemi

Luca Di Lorenzo via Sciangai 53, 00144, Roma (IT)

Di Lorenzo, Luca v. Shangai, Roma Italia, 00144

Sig. Di Lorenzo L. 53, v. Shangai B4/3 Rome, Italy

## ■ Sorprese sui dati nei singoli campi

- Dati “fuori posto” nelle fonti dati

Name	CF	Telephone
Luca Di Lorenzo	LCDLRZ23GH501G	+39066170300
Sistemi S.r.l	15884773662	0039-02415-39200
SimCasa ag. 3	p.iva 126577488	nessuno
Marco Rinaldi	cod. fisc.	MRCRND12H23

## ■ Informazioni “annegate” in campi *free-form*

WING ASS. FORO 4 USO 5J868A DADO ESAGONALE 1/4 POLLICE  
 WING ASSEMBL, UTIL. 5J868-A DADO ESAG. .25" - DRILL FOUR HOLES  
 USE 4 5J868A DADO (HEX .25) - USO FORATURA PER ASSEMBL. ALA  
 RUDER, TAP 6 WHOLES, SECURE W/KL2301 RIVETTI (10 CM)

## ■ Consistenza dei dati

- Mancanza di identificatori consistenti non permette la realizzazione di una vista unica

19-84-103 Cavo RS232 6' M-F Cands

CS-89641 Cavo Maschio-F 6 ft. RS232 #87951

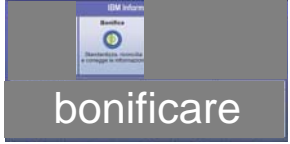
C&SUCH6 Maschio/Femmina 25 PIN 6 piedi, cavo

## ■ Ridondanza

- Record duplicati, mancanza di standard

90328574	IBM	via Sciangai 53, Roma IT 00144
90328575	I.B.M. Inc.	V. Sciangai 54 Roma IT 00144
90238495	Int. Bus. Machines	via Agrostemmi 21, Roma, Italia
90233479	International Bus. M.	v. Agrostemmi 21, Santa Palomba
90233489	Inter-Nation Consults	v.le Sciangai 53, Roma, 00144
90345672	I.B. Manufacturing	via Shangai 53, 00144 Roma,

# Come lavora WebSphere QualityStage?



**Standardize and generate match frequencies**

**Investigazione:**  
comprensione dei dati e loro classificazione in modelli

**Standardizzazione:**  
riconduzione dei dati ad un unico standard

**Matching:**  
Identificazione dei duplicati

**Survivorship:**  
i dati migliori, anche rispetto a più sorgenti, "sopravvivono" in un nuovo record

**Match Frequency**

Match Criteria	Record Count	Percentage	Usage	Gender Code	Profession	Headline
011	100	100.00	1	00	00	000000
012	100	100.00	1	00	00	000000
013	100	100.00	1	00	00	000000
014	100	100.00	1	00	00	000000
015	100	100.00	1	00	00	000000
016	100	100.00	1	00	00	000000
017	100	100.00	1	00	00	000000
018	100	100.00	1	00	00	000000
019	100	100.00	1	00	00	000000
020	100	100.00	1	00	00	000000
021	100	100.00	1	00	00	000000
022	100	100.00	1	00	00	000000
023	100	100.00	1	00	00	000000
024	100	100.00	1	00	00	000000
025	100	100.00	1	00	00	000000
026	100	100.00	1	00	00	000000
027	100	100.00	1	00	00	000000
028	100	100.00	1	00	00	000000
029	100	100.00	1	00	00	000000
030	100	100.00	1	00	00	000000
031	100	100.00	1	00	00	000000
032	100	100.00	1	00	00	000000
033	100	100.00	1	00	00	000000
034	100	100.00	1	00	00	000000
035	100	100.00	1	00	00	000000
036	100	100.00	1	00	00	000000
037	100	100.00	1	00	00	000000
038	100	100.00	1	00	00	000000
039	100	100.00	1	00	00	000000
040	100	100.00	1	00	00	000000
041	100	100.00	1	00	00	000000
042	100	100.00	1	00	00	000000
043	100	100.00	1	00	00	000000
044	100	100.00	1	00	00	000000
045	100	100.00	1	00	00	000000
046	100	100.00	1	00	00	000000
047	100	100.00	1	00	00	000000
048	100	100.00	1	00	00	000000
049	100	100.00	1	00	00	000000
050	100	100.00	1	00	00	000000
051	100	100.00	1	00	00	000000
052	100	100.00	1	00	00	000000
053	100	100.00	1	00	00	000000
054	100	100.00	1	00	00	000000
055	100	100.00	1	00	00	000000
056	100	100.00	1	00	00	000000
057	100	100.00	1	00	00	000000
058	100	100.00	1	00	00	000000
059	100	100.00	1	00	00	000000
060	100	100.00	1	00	00	000000
061	100	100.00	1	00	00	000000
062	100	100.00	1	00	00	000000
063	100	100.00	1	00	00	000000
064	100	100.00	1	00	00	000000
065	100	100.00	1	00	00	000000
066	100	100.00	1	00	00	000000
067	100	100.00	1	00	00	000000
068	100	100.00	1	00	00	000000
069	100	100.00	1	00	00	000000
070	100	100.00	1	00	00	000000
071	100	100.00	1	00	00	000000
072	100	100.00	1	00	00	000000
073	100	100.00	1	00	00	000000
074	100	100.00	1	00	00	000000
075	100	100.00	1	00	00	000000
076	100	100.00	1	00	00	000000
077	100	100.00	1	00	00	000000
078	100	100.00	1	00	00	000000
079	100	100.00	1	00	00	000000
080	100	100.00	1	00	00	000000
081	100	100.00	1	00	00	000000
082	100	100.00	1	00	00	000000
083	100	100.00	1	00	00	000000
084	100	100.00	1	00	00	000000
085	100	100.00	1	00	00	000000
086	100	100.00	1	00	00	000000
087	100	100.00	1	00	00	000000
088	100	100.00	1	00	00	000000
089	100	100.00	1	00	00	000000
090	100	100.00	1	00	00	000000
091	100	100.00	1	00	00	000000
092	100	100.00	1	00	00	000000
093	100	100.00	1	00	00	000000
094	100	100.00	1	00	00	000000
095	100	100.00	1	00	00	000000
096	100	100.00	1	00	00	000000
097	100	100.00	1	00	00	000000
098	100	100.00	1	00	00	000000
099	100	100.00	1	00	00	000000
100	100	100.00	1	00	00	000000



# Cos'è WebSphere DataStage?



trasformare

- **Consente la progettazione visuale delle procedure senza bisogno di codice, sfruttando centinaia di funzioni built-in**
  - Riutilizzo ottimizzato delle procedure
  - Supporto batch & real-time
  - Produce componenti riusabili tra progetti
- **Funzionalità ETL complete con approccio orientato ai metadati**
- **Supporto alla collaborazione ed al lavoro in team**
- **Fornisce accesso ad ogni genere di fonte dati**



Developers



Architects

Trasforma

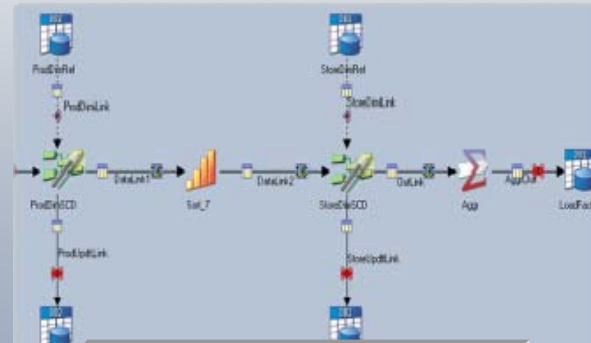


Rilascia



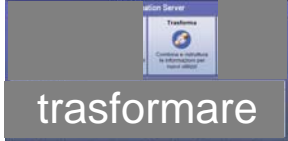
WebSphere DataStage®

Trasforma ed aggrega informazioni in batch o real-time attraverso un design visuale della logica



Centinaia di funzioni di trasformazione built-in

# Come lavora WebSphere DataStage?



The screenshot shows the WebSphere DataStage Designer interface. The main workspace displays a data flow diagram titled "Identify High Value Customers". The diagram starts with three input nodes: "Checking\_Accounts" (DB2), "Savings\_Accounts" (ORACLE), and "Checking\_And\_Savings" (SQLServer). These feed into intermediate nodes: "Mainframe" (to "Stan\_Accounts"), "Oracle" (to "Stan\_Savings"), and "SQLServer" (to "Stan\_Check\_Savings"). These three intermediate nodes converge into a "Join" node. The output of the "Join" node goes to "Unduplicate\_Match", then "Match\_Out", and finally "MapInputC14", which connects to a "DB2\_Connector\_06" target. A legend on the right side of the diagram lists the operations: "Write Range Map", "Encode", "Decode", "Compress", and "Expand".

On the left side of the interface, there is a "Pakete" (Repository) tree showing a project named "BankingDemo" with various sub-objects like "Jobs", "Match Specifications", "Containers", "Data Elements", etc. Below the repository tree are icons for "File S...", "Seq...", "Para Data", "SAS", and "DB2".

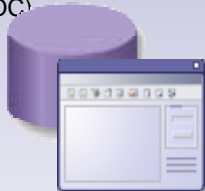
At the bottom of the screen, there is a Windows taskbar with the "start" button and several open applications. Below the taskbar, there are icons for "Oracle", "Data Set", "Modify", "Tail", and "Expand".



# WS Connettività DataStage

## Applicazioni Enterprise

JD Edwards  
Oracle Applications  
PeopleSoft  
SAP BW (BAPI, IDOC)  
SAP R/3 (ABAP, BAPI, IDOC)  
Siebel



## RDBMS

IBM DB2  
IBM IMS  
VSAM  
Oracle  
Informix  
RedBrick  
SQL Server  
Sybase  
Teradata  
U2 (Universe, UniData)  
Tandem NON-STOP SQL  
SAS



## Formati di Business Exchange

XMLS  
EXML  
EDI  
FIX  
SWIFT  
HIPAA



## Real-Time

WebSphere MQ  
SeeBeyond  
Java Messaging Services  
Java (Client & Transformer)  
XML (Read / Write)  
XSL-T XSL-T Transformer  
Web Services (SOAP)  
Enterprise Java Beans  
Tibco Rendez Vous  
Bea Tuxedo



## Flat File e accesso generico

VSAM  
VSAM CICS  
IDMS  
C-ISAM  
Sequential File  
Complex Flat File  
File Set  
Data Set  
Named Pipe  
FTP (standard, secure)  
Compressed / Encoded Data  
External Command Call  
Parallel Wrap 3<sup>rd</sup> party applications



...e molti altri !

# Cos'è IBM WebSphere Federation Server ?



- **Consente l'accesso a informazioni eterogenee e distribuite come se provenissero da un singolo sistema**
  - Query SQL singola per accedere fonti dati diverse
  - Tools visuali per la definizione di query federate
- **Include ottimizzatore delle query, single sign-on, viste unificate e compensazione delle funzioni**
- **Supporto per scritture transazionali su fonti eterogenee**
- **Consente il deployment in architettura SOA di servizi bidirezionali per l'accesso ai dati**

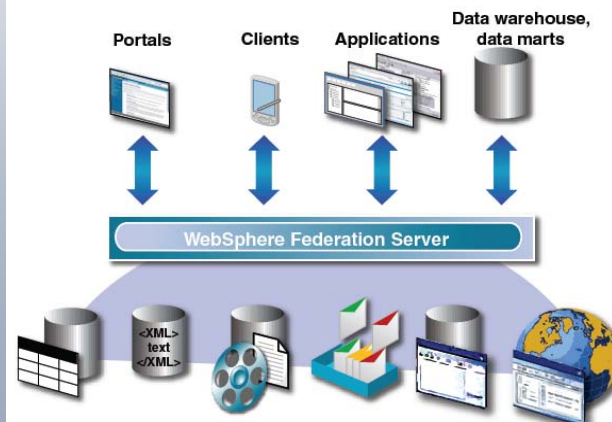
Federa



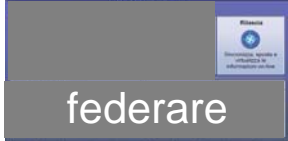
## IBM WebSphere Federation Server

Integra fonti dati eterogenee e ne consente l'accesso come se fossero una singola fonte dati omogenea

Consente di sfruttare le applicazioni già esistenti estendendone l'accesso in real-time



# Come lavora WebSphere Federation Server ?



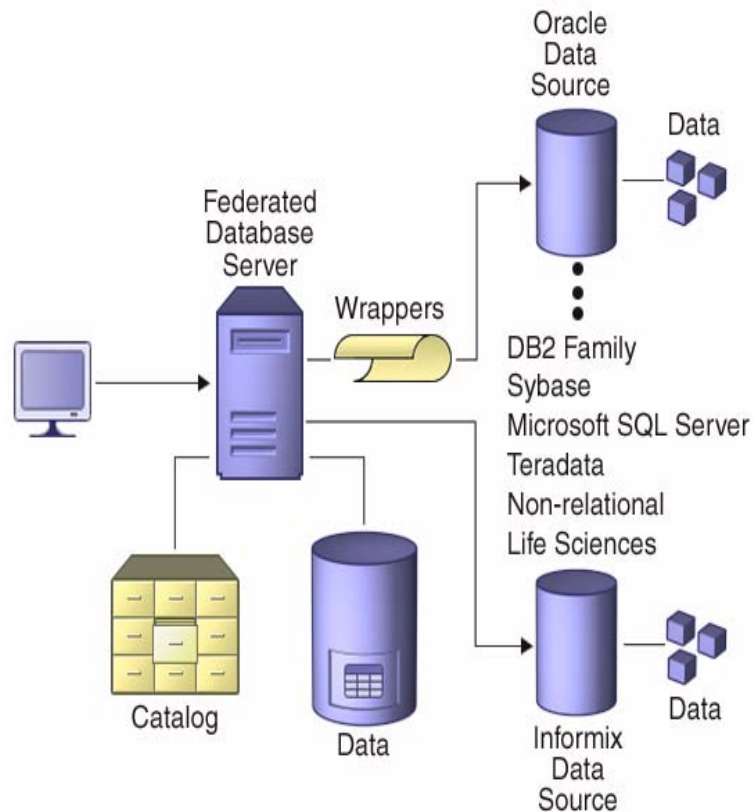
## Single SQL Query Joins:

```

SELECT      parameters_return_billto_key as BILL_TO_KEY,
            billto_company_name,
            parameters_return_shipto_key as SHIP_TO_KEY,
            CASES_SHIPPED,
            GROSS_SALES,
            URL

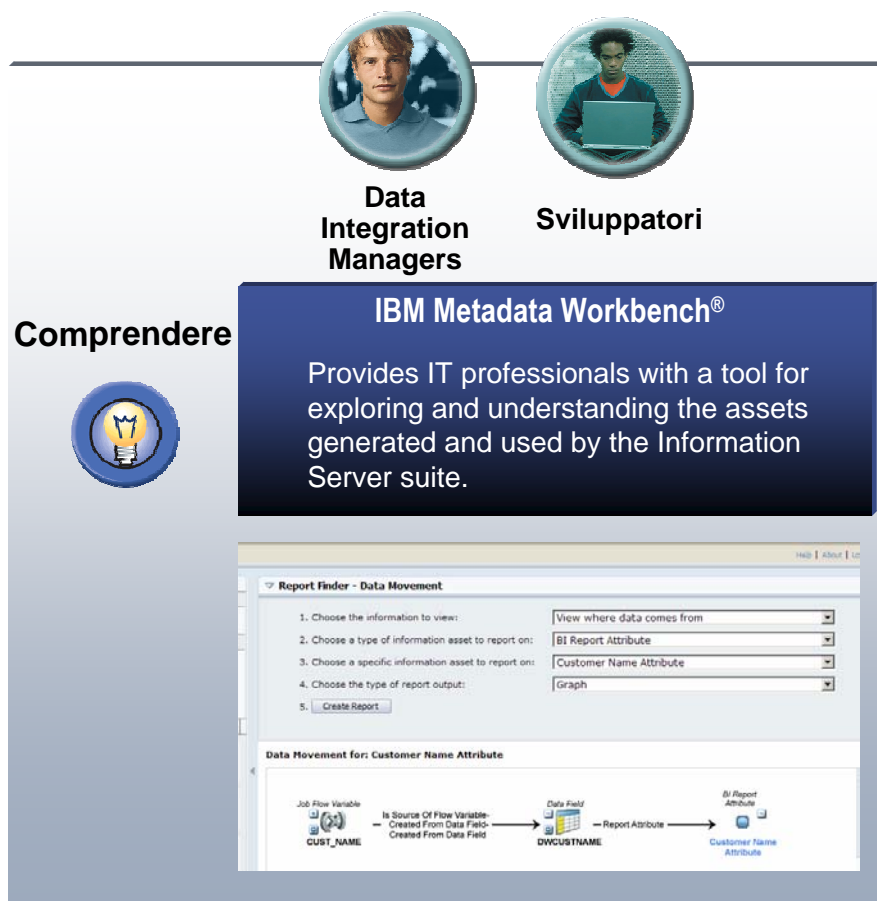
FROM        GETKEYSSOAP_KEYSREALTIME,           ← Web Service
            GLOBAL_SALES_TRAN_NN,              ← XML Documents
            BILLTO_DIMENSION,                  ← Data Warehouse
            URL_INVOICES                       ← Unstructured Data

WHERE      getkeysrealtime_ship_to_number = '13546'
and        getkeysrealtime_ship_to_number = URL_INVOICES.shipno
and        ltrim(rtrim(translate(ship_to_number, ',', x'0a')))
            = getkeysrealtime_ship_to_number
and        parameters_return_billto_key = billto_key
and        ltrim(rtrim(translate(sales_order_number, ',', x'0a')))
            = URL_INVOICES.orderno
    
```



# Cos'è IBM Metadata Workbench ?

- **Esplorazione *web-based* degli assets generati e utilizzati da Information Server o importati da tools esterni**
- **Funzionalità *out-of-the-box* di reporting sui dati, *data lineage*, significato di business, analisi di impatto e dipendenze**
- **Data lineage dai report di business intelligence come base per la governance e la compliance verso legislazioni come Sarbanes-Oxley, Basel II**



**Data Integration Managers**      **Sviluppatori**

**Comprendere**

**IBM Metadata Workbench®**

Provides IT professionals with a tool for exploring and understanding the assets generated and used by the Information Server suite.



The screenshot shows the 'Report Finder - Data Movement' interface. It includes a list of steps: 1. Choose the information to view: View where data comes from; 2. Choose a type of information asset to report on: BI Report Attribute; 3. Choose a specific information asset to report on: Customer Name Attribute; 4. Choose the type of report output: Graph; 5. Create Report. Below this, a data lineage diagram is shown for 'Customer Name Attribute', illustrating the flow from a Job Flow Variable (CUST\_NAME) to a Data Field (DWCUSTNAME) and finally to a BI Report Attribute (Customer Name Attribute).



# Overview delle funzionalità del Metadata Workbench

Esplorazione degli assets di Information Server:

Job, Report, Database, Modelli, Termini, Stewards, Sistemi

- ✓ Navigazione degli assets
- ✓ Ricerca semplice ed avanzata
- ✓ Vista *cross-suite* integrata degli assets
- ✓ Vista grafica delle relazioni tra assets
- ✓ Query builder *ad-hoc*

Analisi delle dipendenze e delle relazioni tra assets e report di BI di applicazioni esterne

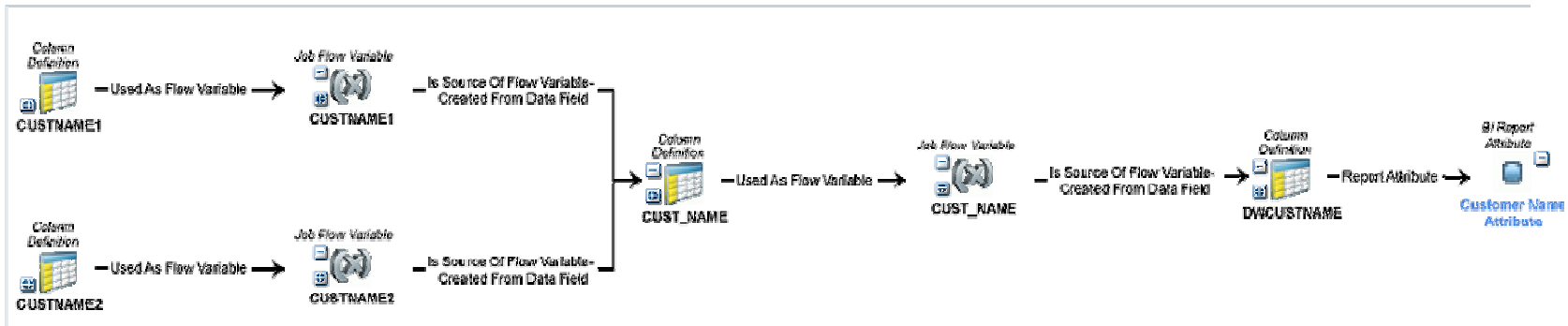
- ✓ Tracciamento dei movimenti dati da e per I database, job e report
- ✓ Comprensione del significato di business di colonne, tabelle ed altri assets
- ✓ Comprensione dell'impatto dei cambiamenti di una colonna o di un flusso attraverso tutta la piattaforma

Gestione dei metadati di Information Server a scopo analitico

- ✓ Modifica dei nomi e delle descrizioni degli assets
- ✓ Creazione di steward e loro assegnazione
- ✓ Riconciliazione di assets duplicati
- ✓ Accesso alle informazioni di *runtime*

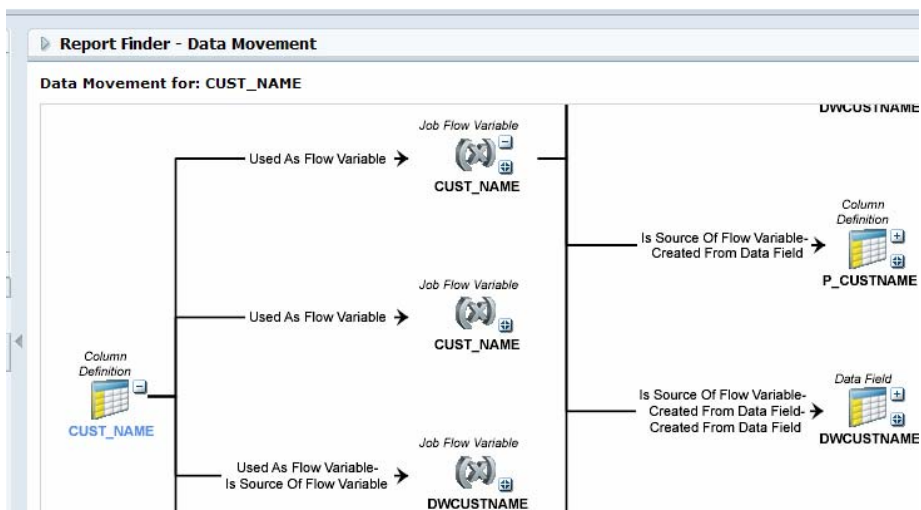
# Caso d'uso: *Data lineage* da un report di business

Data Movement for: Customer Name Attribute



Osservando un report che riporta l'attività di un conto, da quale fonte dati viene prelevato il nome del cliente ? E da dove viene il dato relativo al movimento ?

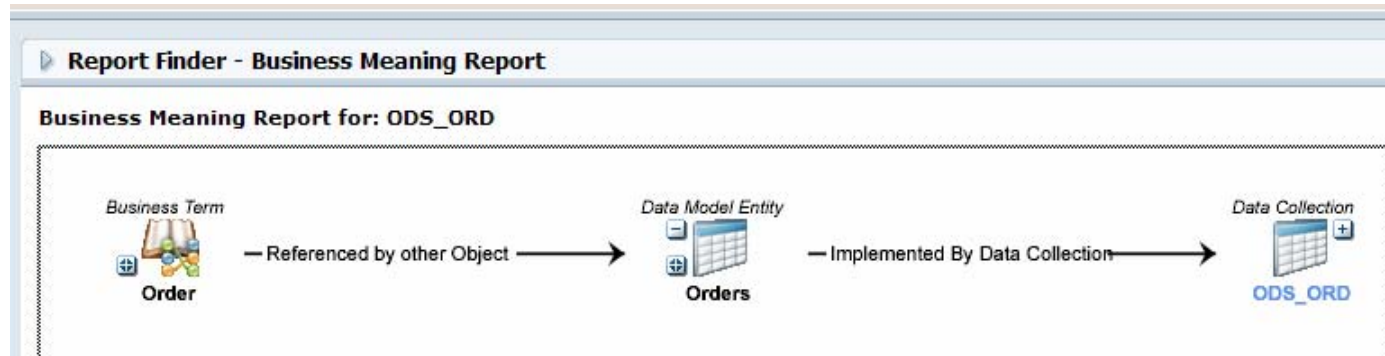
# Caso d'uso: Analisi di impatto da una colonna



Uno sviluppatore ETL deve cambiare la lunghezza della colonna NOME\_CLIENTE e deve sapere dove questa colonna viene utilizzata al fine di stimare correttamente i tempi ed i costi connessi alla modifica.



# Caso d'uso: Comprensione del significato di business



Quale genere di informazione è contenuta nella tabella ODS\_ORD ?

# Cos'è WebSphere Information Services Director?

- Un applicazione J2EE che incapsula le logiche di integrazione informativa in servizi che isolano gli sviluppatori dalle fonti dati fisiche
- Consente ai servizi di essere invocati come Enterprise Java Beans
- Implementa funzionalità di load balancing & fault tolerance per richieste verso clusters di Information Servers
- Implementa l'infrastruttura di base degli Information Services

**Developers**      **Architects**

**WebSphere Information Services Director**  
Consente la gestione ed il deploy dei servizi senza la scrittura di codice

Overview    Service 01

**Bindings**

**Binding Settings**

↳ EJB

Enable Binding

JNDI Name:

Package Name:

↳ SOAP over HTTP

↳ SOAP over JMS

Enable Binding

Activation Spec JNDI Name:     SOAP Style:  Select    Priority:

Description:     Connect:

Destination:     Message:

**Deployment SOA rapido**

# Cosa rende WISD differente ?

Deployment unificato dei servizi	Benefici
Funzionalità condivisa per la pubblicazione e la gestione dei servizi DataStage, QualityStage e Federation	<b>Riduzione dei costi di amministrazione e debugging</b>
Supporto alla pubblicazione diretta dei servizi per database DB2	<b>Flessibilità nell'integrazione</b>
Supporto per Classic Federation & IICE attraverso Federation Server	<b>Servizi per ogni tipo di dato</b>
Metadata Service	
Implementa la gestione di metadati operazionali e di design	<b>Riduce i costi di amministrazione</b>
Reportistica su impact analysis e lineage fino a livello di servizio	<b>Identifica potenziali rischi</b>
Deployment dei servizi semplificato	
Deployment rapido dei servizi	<b>Sviluppo più veloce a parità di skill</b>
Supporto per Web Services e bindings EJB con una singola definizione ed un unico punto di gestione	<b>Riutilizzo della logica esistente</b>
Flessibilità nella definizione delle interfacce dei servizi. Include supporto per arrays, defaults e overrides	<b>Flessibilità in fase di sviluppo</b>

# IBM®