

# InfoPlanet

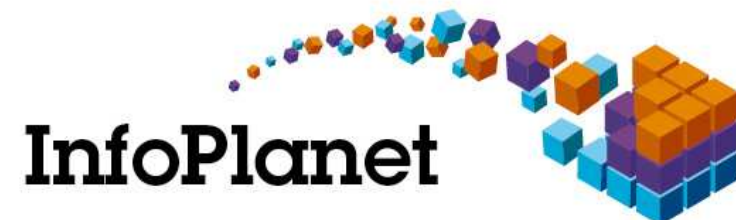
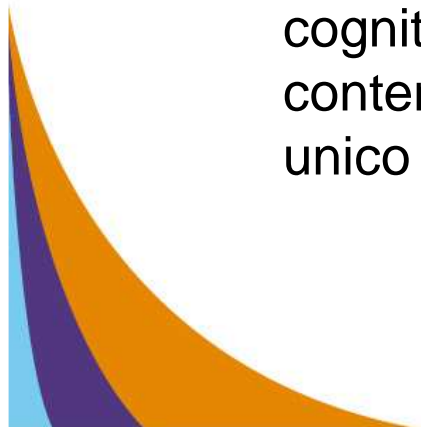
**Integrazione dei dati:  
una piattaforma completa**

**Brunello Bonanni, *Senior IT Architect, IBM Italia***



# Integrazione

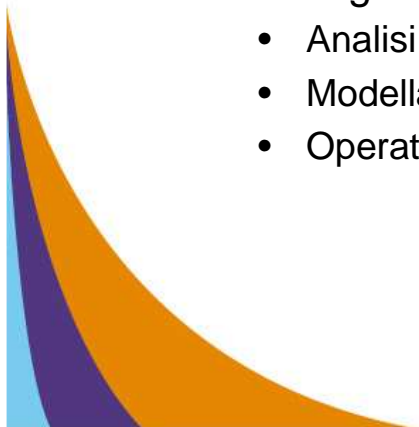
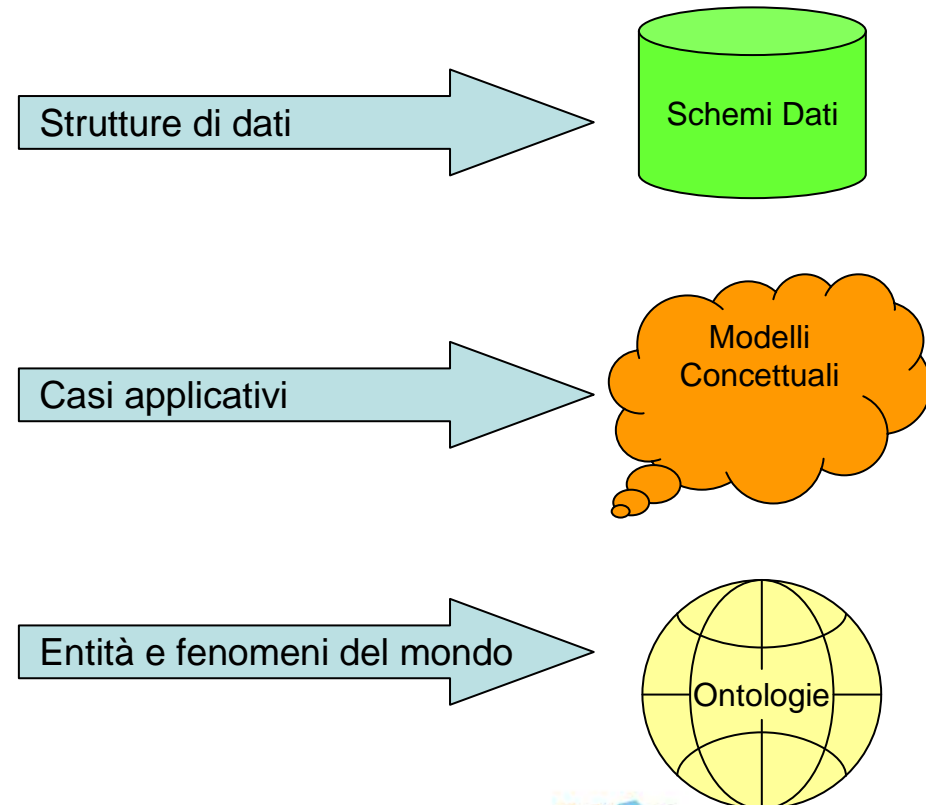
- Integrare più che una esigenza è diventata uno slogan che ricorre in numerosi contesti dell'IT.
- Nel dominio della gestione delle informazioni definiamo propriamente:
  - **Data Integration:** disciplina che indirizza tutte quelle attività che ricorrono lato back-end per unire dati, come ad esempio l'aggregazione di dati provenienti da sorgenti differenti in uno o molteplici repository centralizzati.
  - **Information Integration:** disciplina che indirizza i processi cognitivi che ciascun utente intraprende per sintetizzare contenuti anche parziali provenienti da differenti sorgenti in un unico concetto comprensibile



# Le tipologie dei dati

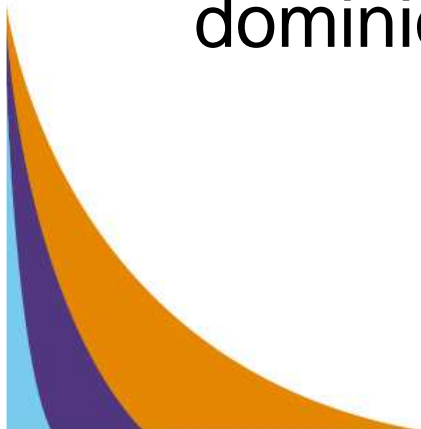
Il dato può essere visto da diverse prospettive

- Strutturale
  - Codifiche
  - Schemi logico-fisici
  - Protocolli e formati
  
- Concettuale
  - Analisi del dominio applicativo
  - Modellazione semi-formale
  - Operatività in contesti omogenei
  
- Ontologica
  - Analisi del dominio di business
  - Modellazione formale
  - Operatività in contesti eterogenei

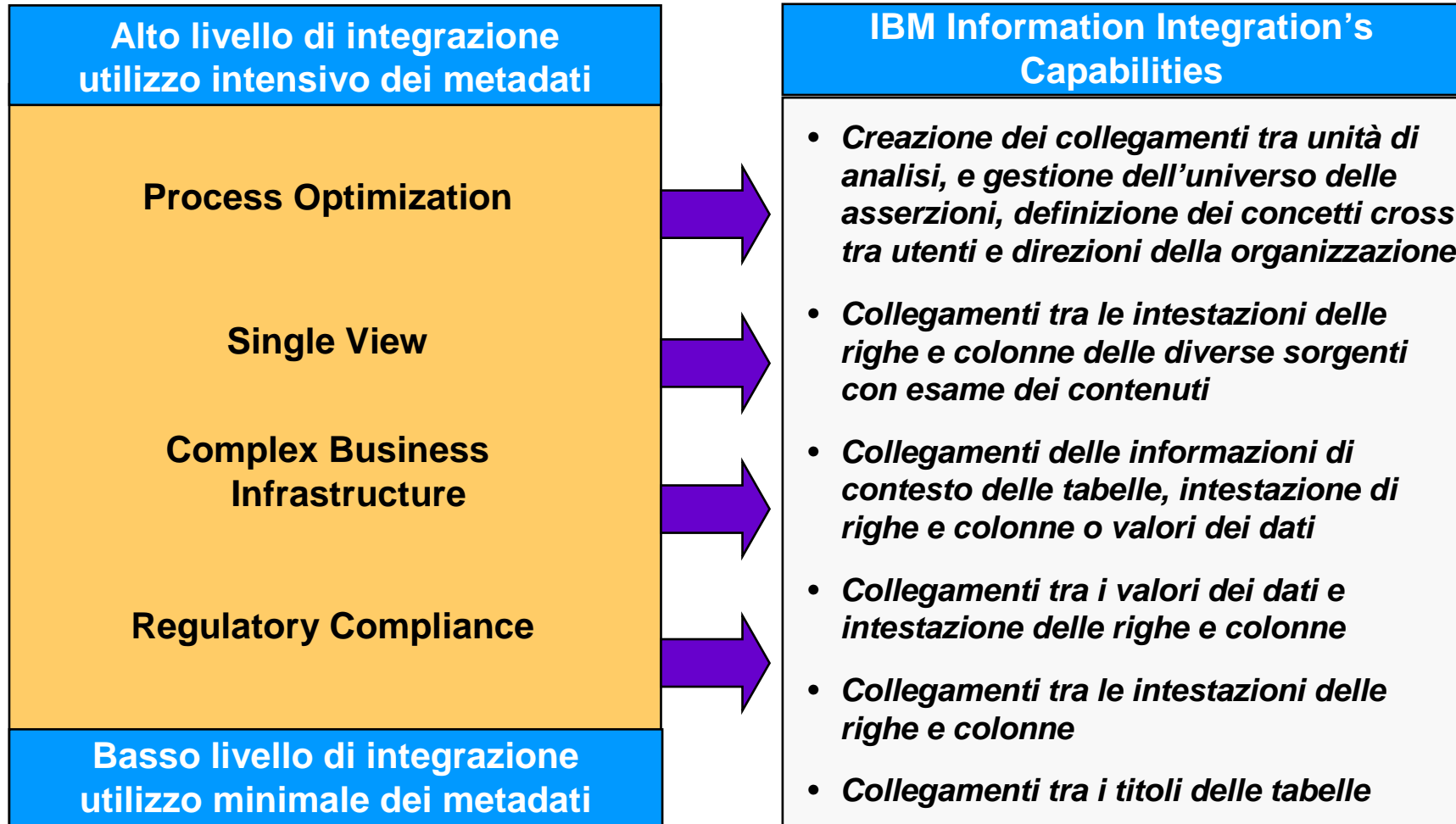


# Il ruolo dei metadati

- Favoriscono la **data integration** in quanto consentono agli agent coinvolti di raccogliere propriamente materiale dalle diverse sorgenti.
- Favoriscono anche la **Information Integration** consentendo agli utenti di raggruppare i dati secondo criteri basati sulla conoscenza del dominio di interesse.



# La gerarchia dell'Integrazione



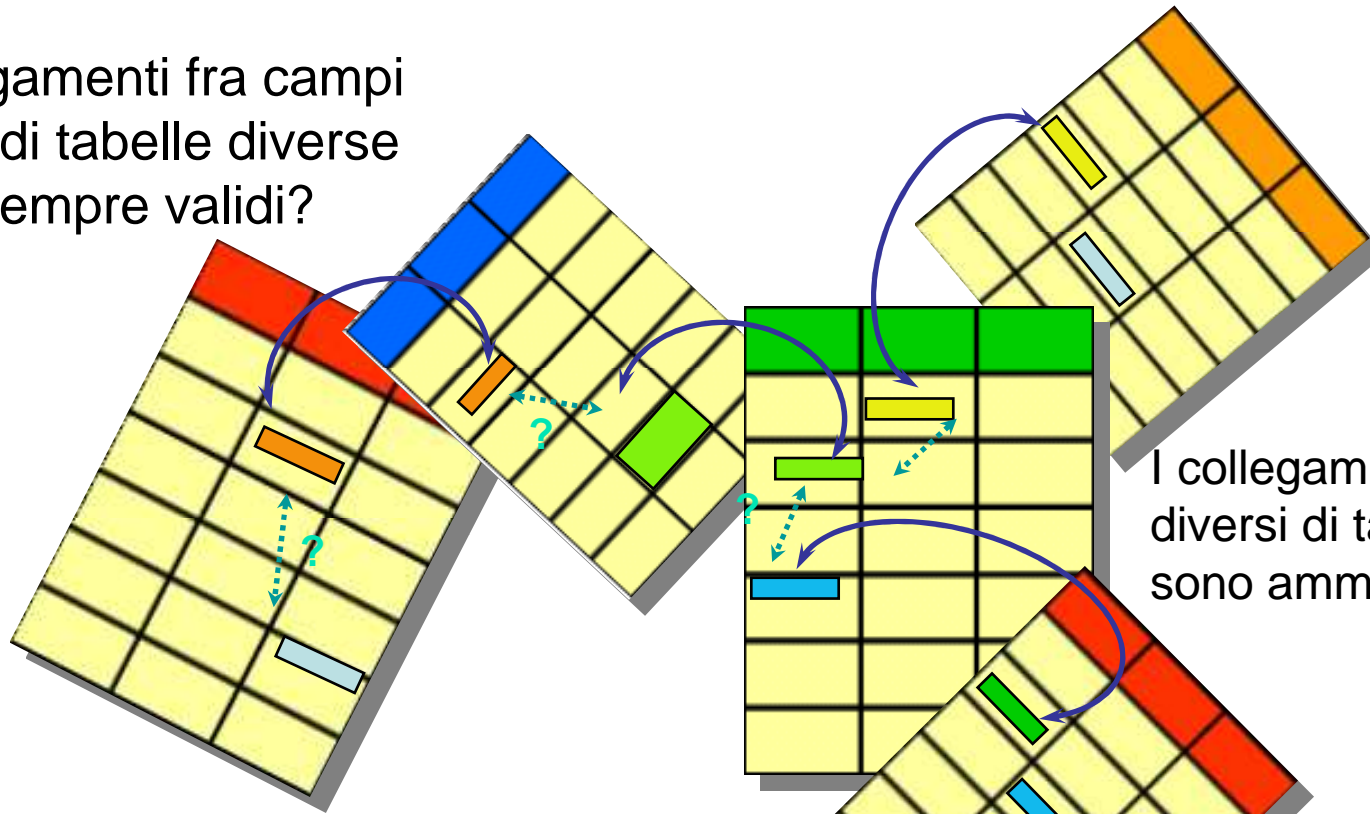
# La comprensione dell'Utente

- Lo scopo dell'utente nel ricevere informazioni integrate non è quello di ottenere un "Oggetto" composito ma di comprenderne il suo significato
- La comprensione si basa sulle relazioni:
  - Gli oggetti sono interpretabili attraverso il contesto\*
  - I contesti sono interpretabili attraverso gli oggetti
  - Gli oggetti sono interpretabili attraverso le categorie
  - Le categorie devono essere supportate da esempi
  - Le categorie possono anche basarsi su casi rari
- *Il fine ultimo è quello di acquisire conoscenza per produrre nuova conoscenza, fornire evidenze o fare nuove ipotesi: proporre nuove sfide per la propria azienda*

\* Principio del contesto di Frege - Introduzione ai Fondamenti dell'aritmetica- 1884

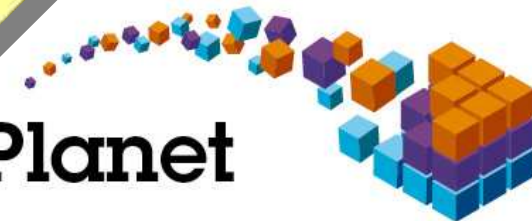
# Parliamo sempre della stessa cosa?

I collegamenti fra campi uguali di tabelle diverse sono sempre validi?



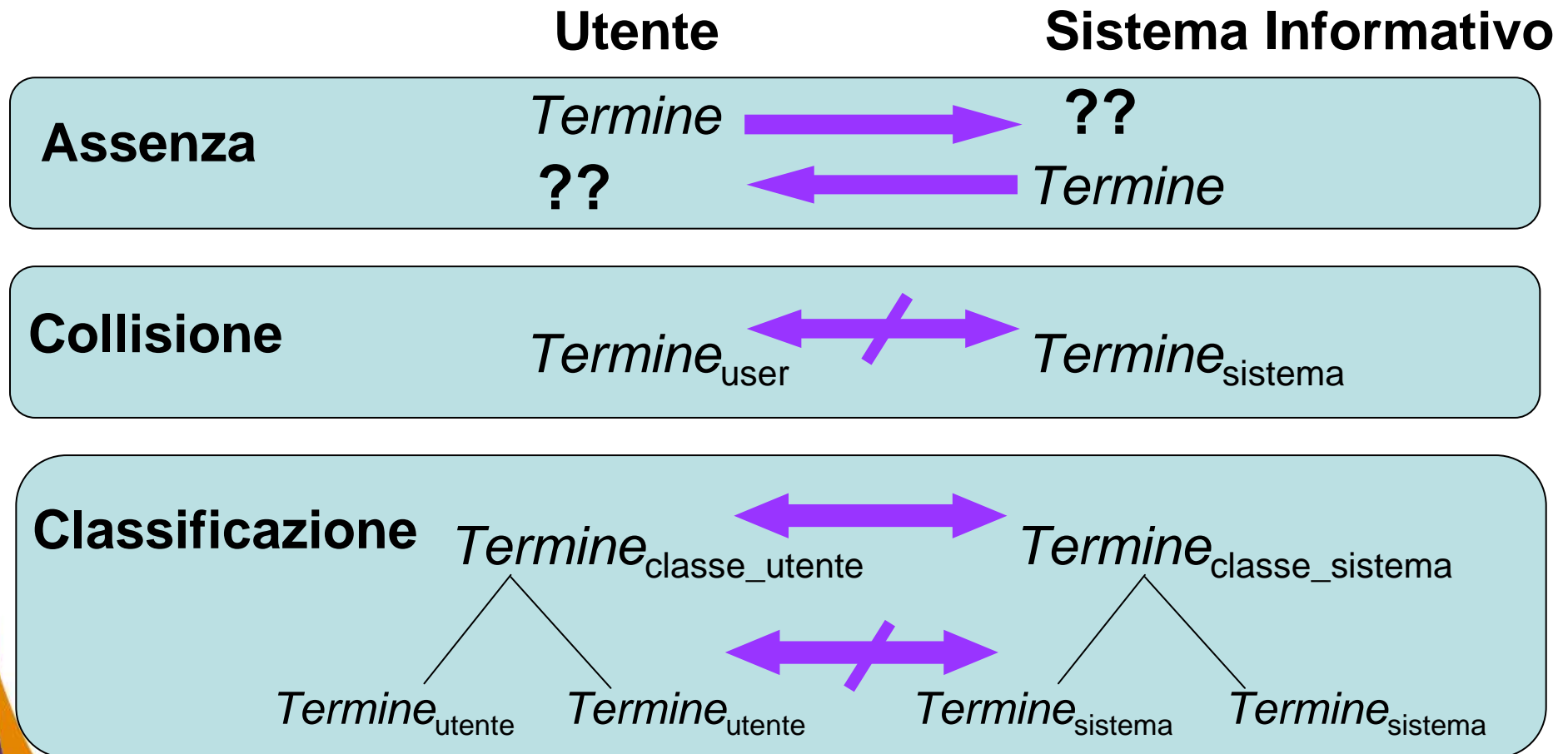
I collegamenti fra campi diversi di tabelle diverse sono ammissibili?

**Quando e Come** garantire che l'integrazione dei dati non solo è sostenibile ma soprattutto consistente?



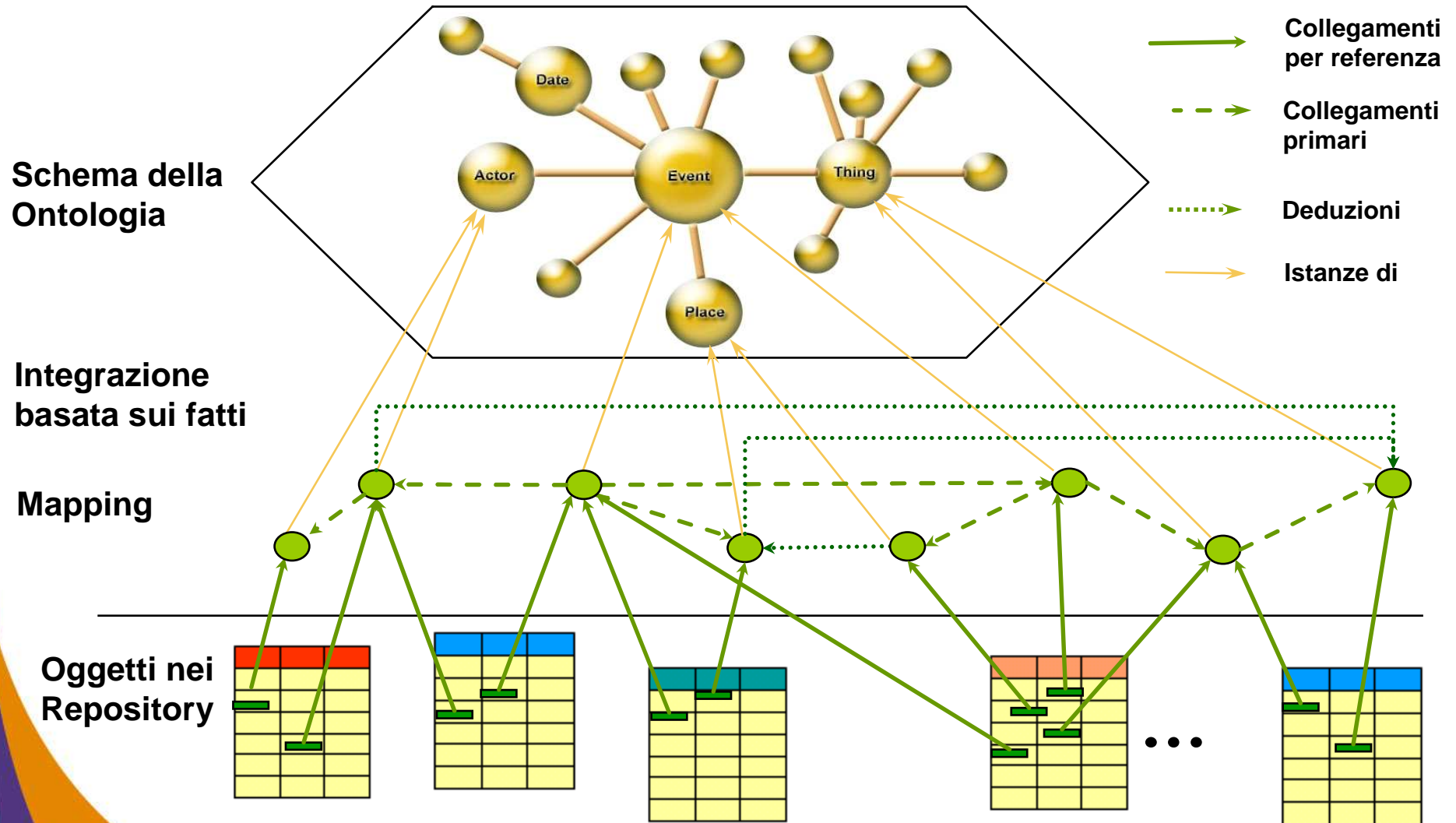


# I limiti della terminologia





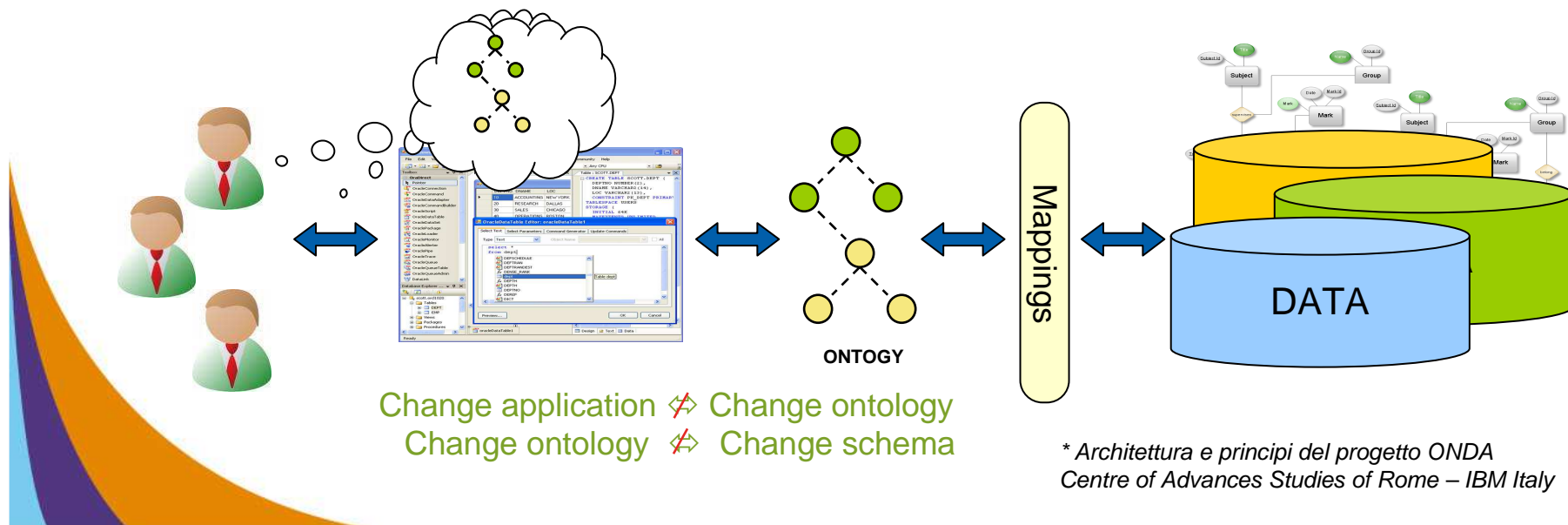
# Il supporto delle Ontologie



# L'integrazione basata sulla semantica

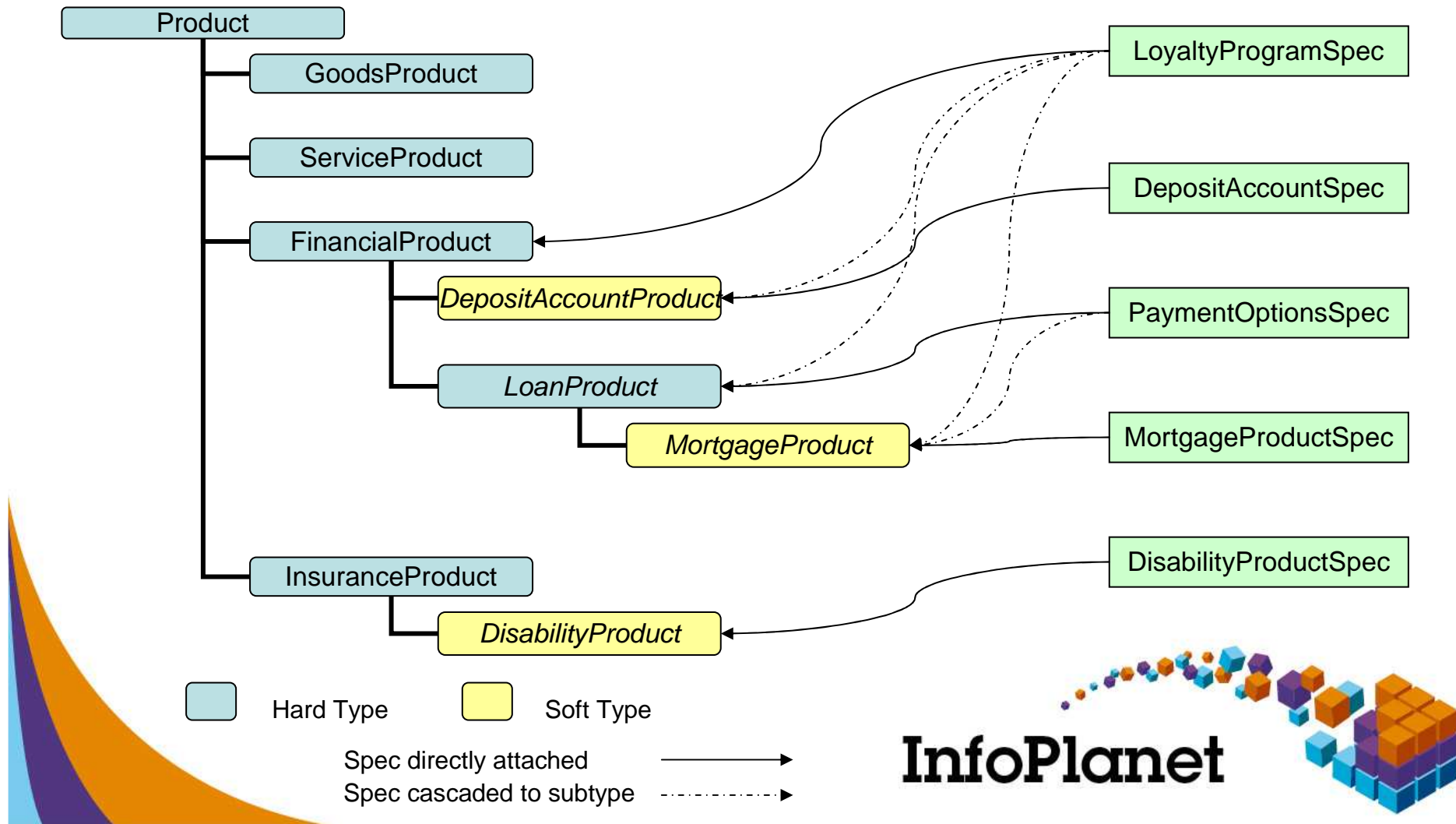
E' il processo di integrazione dei dati a livello concettuale

- ✓ Consente l'integrazione di sorgenti dati eterogenee
- ✓ Non impone all'utente di confrontarsi con le strutture dati
- ✓ La gestione del mapping consente di collegare gli oggetti fisici allo schema ontologico
- ✓ Le funzioni di interazione sono offerte sui concetti del modello ontologico e sono risolte sugli oggetti fisici delle sorgenti dati



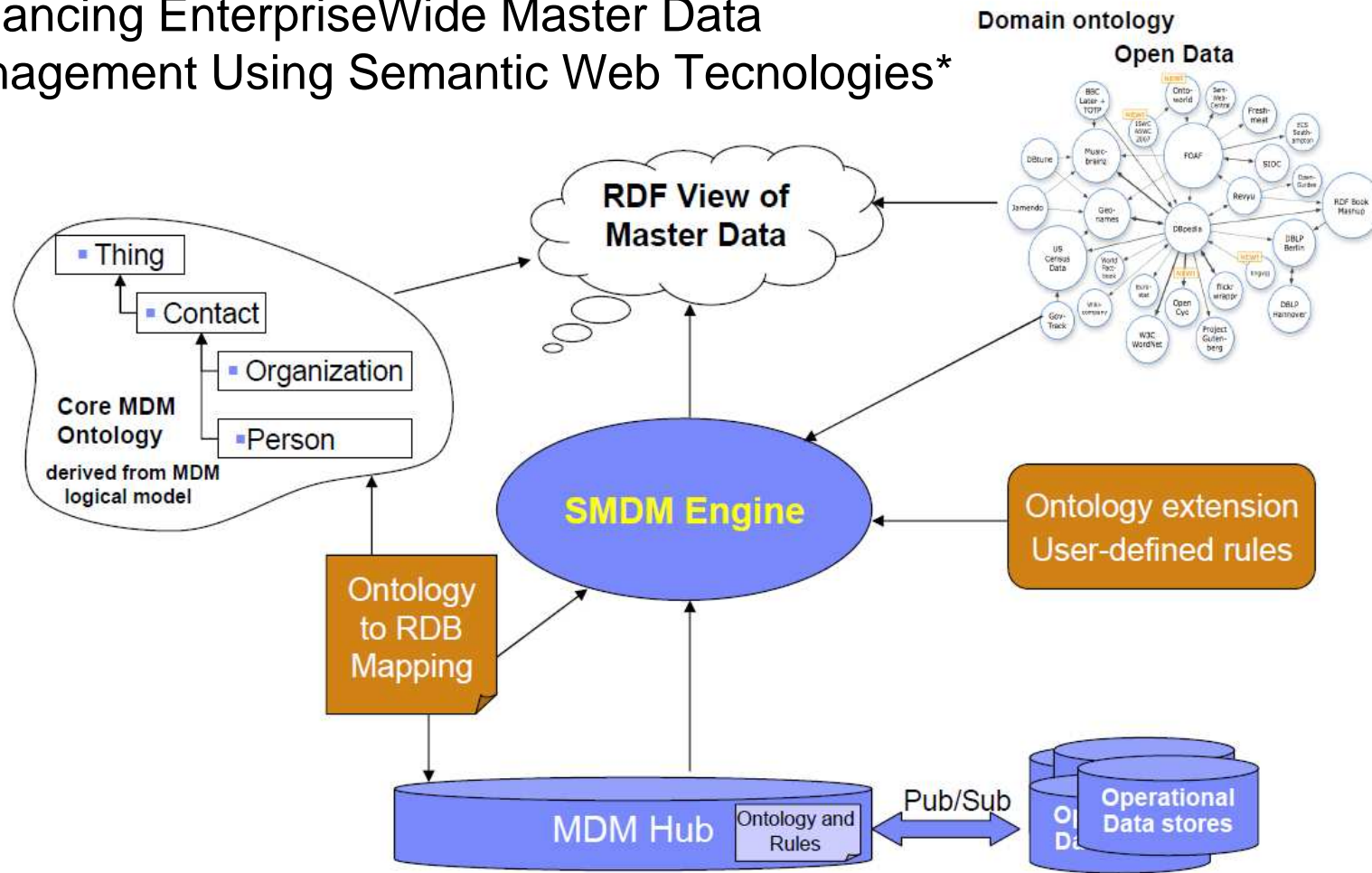
# MDM: un livello estensionale della ontologia

Infosphere MDM Server Hub verso l'integrazione basata sulla semantica



# Semantic Master Data Management

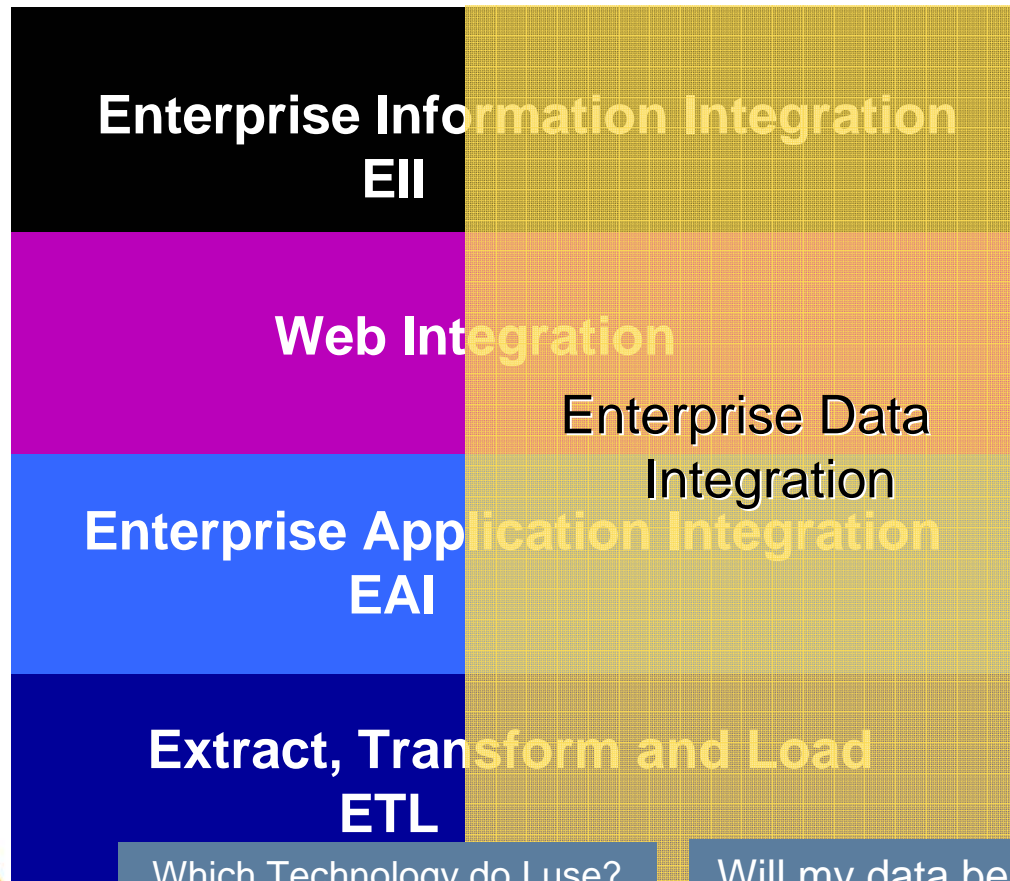
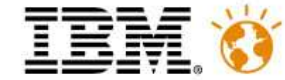
Enhancing EnterpriseWide Master Data Management Using Semantic Web Technologies\*



\* IBM China Research Lab  
 IBM Software Group, Canada  
 Shanghai Jiao Tong University, China



# Le tecnologie introdotte per indirizzare il problema



- Importanti e preziose funzionalità
- Ma non la reale soluzione all'integrazione dei dati

- Disegno ed organizzazione dei dati molto povera
- Gestione dei dati non efficiente
- Deterioramento della qualità
- Inadeguato delivery al cambiamento del business e della tecnologia

Which Technology do I use?

Will my data be siloed?

How Flexible Is My Design?

Integrate the Parts

Common Meta Data

Automation

Reusability

Profiling

Cleansing

SOA

# La tecnologia IBM InfoSphere

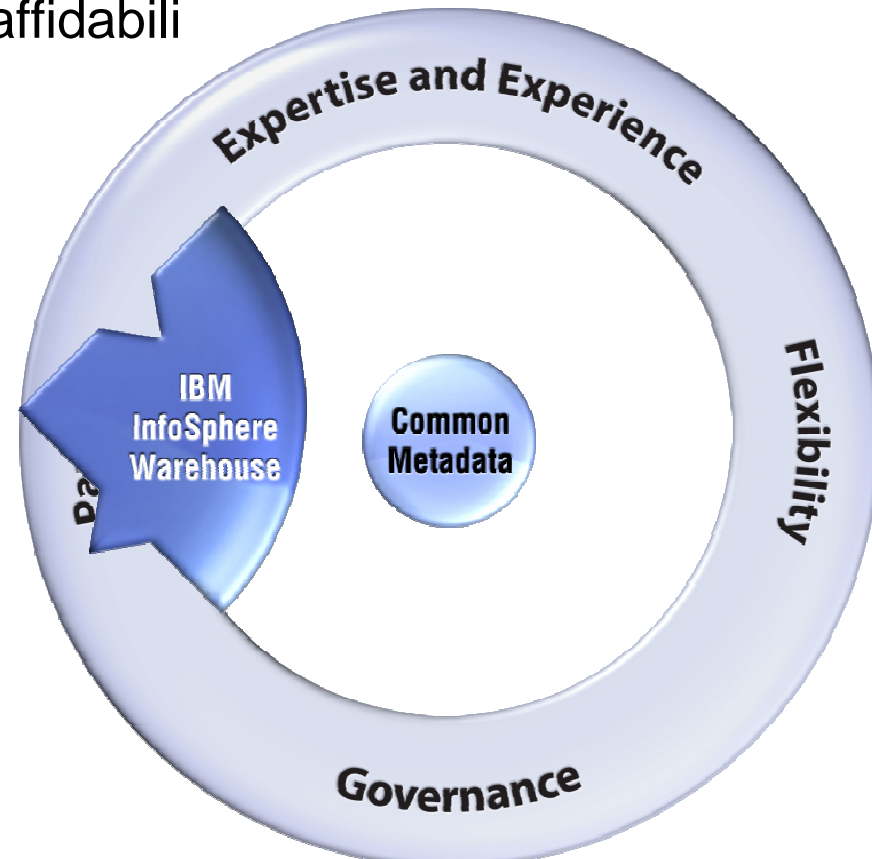
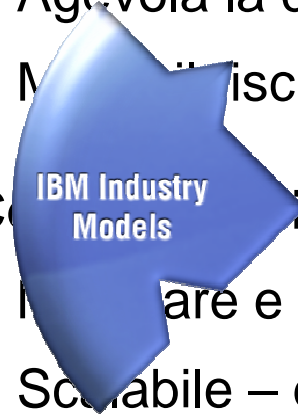


Una piattaforma per la gestione delle informazioni

- Semplifica il rilascio di informazioni affidabili
- Accelera il client value
- Agevola la collaborazione
- Minimizza il rischio

Caratteristiche importanti

- Integrale e Integrata
- Scalabile – dal progetto all'impresa



InfoSphere™

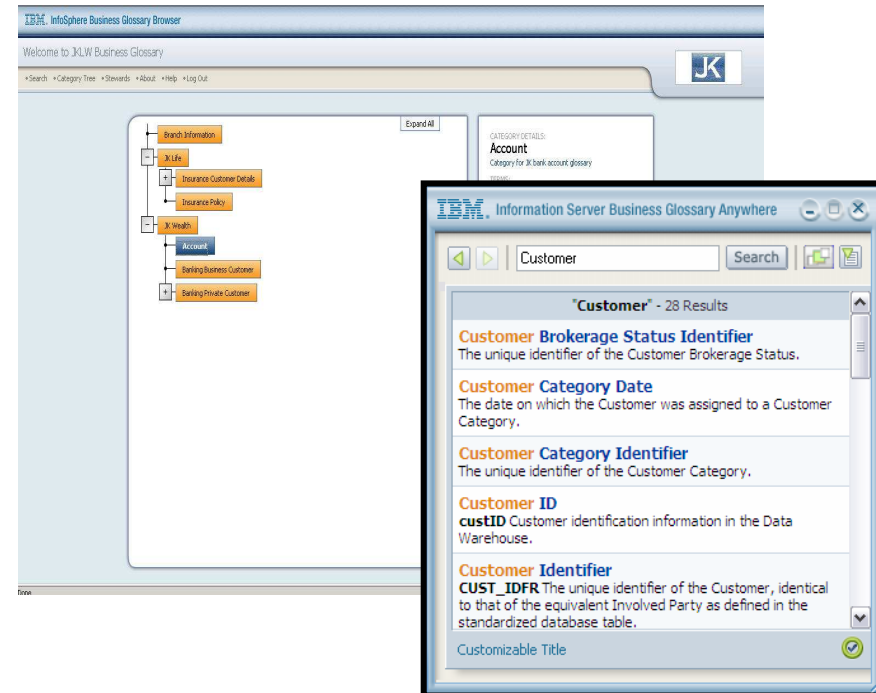
InfoPlanet





# Creare un Dizionario di Business

- Un tool web-based dedicato agli utenti di business che permette:
  - La creazione e gestione di un vocabolario controllato
  - Creazione e gestione di una tassonomia di business
  - Creazione di un repository dei metadati di business
- Un riferimento per conoscere il patrimonio informativo dell'impresa
  - Significato
  - Dipendenze
  - Utilizzo
  - Qualità
  - Ownership/Responsabilità



*Organizzato secondo gerarchie di business definite dalla tassonomia di business*

**InfoPlanet**





# Analizzare le informazioni e monitorarne la qualità



- 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

The diagram illustrates the workflow of the IBM Information Analyzer. It features two circular icons at the top: 'Subject Matter Experts' (a woman working on a laptop) and 'Data Analysts' (a man's portrait). Below these is a blue box with the text 'IBM Information Analyzer' and 'Analizza strutture dati e ne controlla l'aderenza alle regole di integrazione e qualità'. To the left, a light blue box labeled 'Comprendi' contains a lightbulb icon. At the bottom, a screenshot of the software interface shows a 'Data Type' section with 'VarChar' selected, an 'Inferred Summary' pie chart, and an 'Inferred Frequency Distribution' table. A light blue box labeled 'Vista fisica' is positioned at the bottom right of the interface.

**Subject Matter Experts**

**Data Analysts**

**IBM Information Analyzer**

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

**Comprendi**

**Vista fisica**

Data Value	Data Type	#	%
afslve	Char	45	1.00
wire	Char	384	8.53
evb	Char	769	17.09
afvefvg	Char	444	9.87
wirevew	Char	252	5.60
afvld	Char	444	9.87
fff	Char	252	5.60

# Dai requisiti alle specifiche



Fast Track è progettato in base quattro principi guida:

1. Ottimizzare il lavoro di collaborazione del team
2. Supportare la data governance e fornire capacità di auditing
  - Repository per i termini di business, le specifiche e i requirements di progetto e le loro correlazioni
  - Tracking e storicizzazione delle informazioni, dalla specifica allo sviluppo
3. Incrementare la produttività del team
4. Favorire il successo e la consegna puntuale del progetto



Data Analysts



Data Modelers



Developers

## Comprensione



## InfoSphere Fast Track

*Favorire la collaborazione tra Analisti e Sviluppatori per velocizzare la traduzione dei business requirements in progetti di data integration di successo*

The screenshot displays the InfoSphere Fast Track interface with several key components highlighted by callouts:

- Customizable spreadsheet view hosted in metadata repository:** Points to the main mapping specification table.
- Source column info:** Points to the 'Source' column header in the mapping table.
- Target column info:** Points to the 'Target' column header in the mapping table.
- Transformation rule and/or function:** Points to the 'Transformation' column in the mapping table.
- Details of source-to-target mapping:** Points to the 'Sources/Targets' and 'Annotation' tabs at the bottom.
- Drag&drop metadata browser:** Points to the right-hand pane showing a tree view of metadata.
- DataStage job generation:** Points to the 'Generate DataStage Job' button at the bottom right.

# Bonificare le Informazioni

- 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**



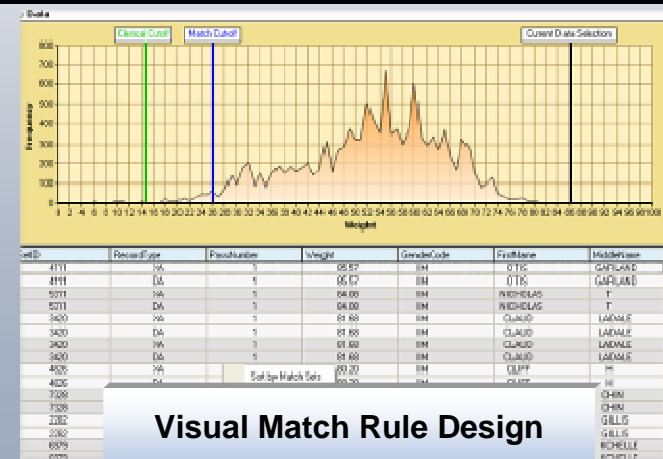
**Data Analysts**

**Bonifica**



**InfoSphere QualityStage™**

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

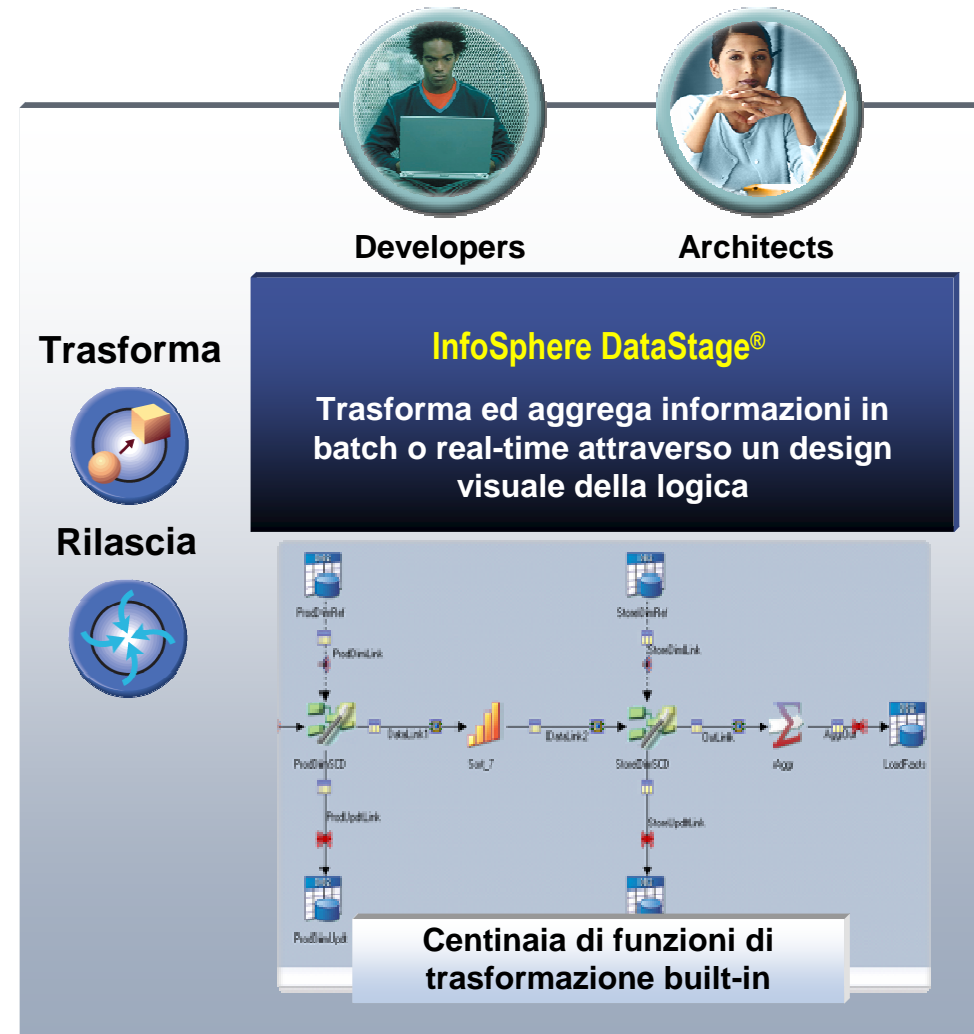


**Visual Match Rule Design**



# Trasformare ed aggregare informazioni in batch o real-time

- 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





# Lo sviluppo visuale delle trasformazioni

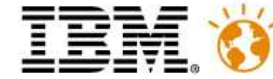
The screenshot displays the IBM DataStage Designer interface. The main workspace shows a job titled "Identify High Value Customers" in a "Parallel" environment. The job flow is as follows:

- Input Stage 1:** A "Mainframe" connector feeds into a "Star\_Accounts" stage, which is linked to a "Checking\_Accounts" data source.
- Input Stage 2:** An "Oracle" connector feeds into a "Star\_Savings" stage, which is linked to a "Savings\_Accounts" data source.
- Input Stage 3:** An "SQLServer" connector feeds into a "Star\_Check\_Savings" stage, which is linked to a "Checking\_And\_Savings" data source.
- Join Stage:** The three star stages feed into a "Join" stage.
- Intermediate Stage:** The "Join" stage feeds into a "Unduplicate\_Match" stage.
- Output Stage:** The "Unduplicate\_Match" stage feeds into a "Match\_Out" stage, which is linked to an "IBMBankDWES" data source.
- Final Stage:** The "Match\_Out" stage feeds into a "MapInputC14" stage, which is linked to a "DB2\_Connector\_06" data source.

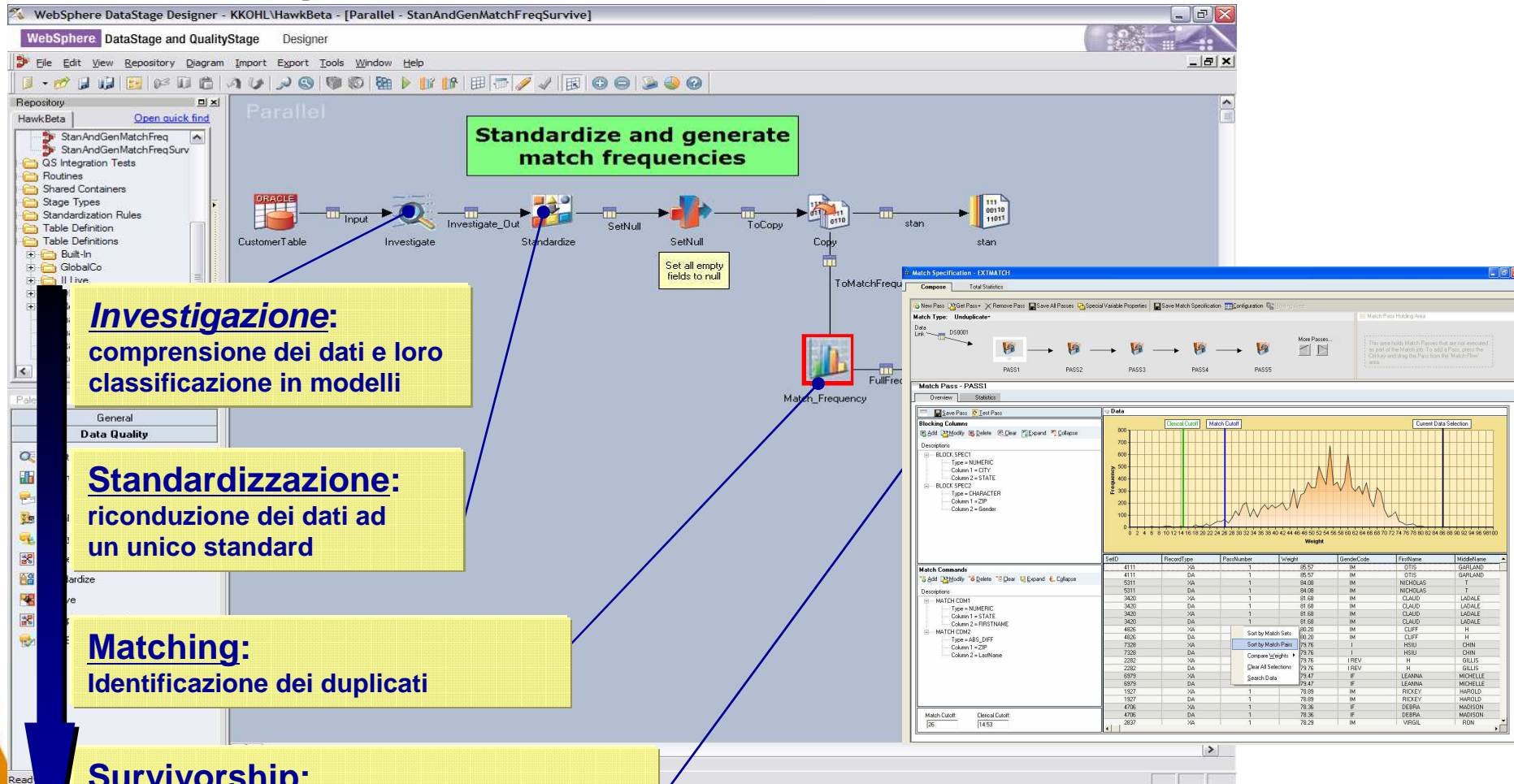
On the right side of the interface, a vertical toolbar contains several icons with labels: "Write Range Map", "Encode", "Decode", "Compress", and "Expand".

The left sidebar shows a project tree for "BankingDemo" with folders for "Jobs", "Match Specifications", "Matches", "Shared", "Containers", "Data Connections", "Parameter Sets", and "Table Definitions".

The bottom of the image shows a Windows taskbar with the "start" button and several open applications, including "Webph...", "C:\Data...", "C:\Docu...", "IBM Infor...", "Imkeren...", "E:\Bankin...", and a system clock showing "7:50 AM".



# La progettazione visuale



InfoPlanet

# Visione Integrata

The screenshot shows the IBM Information Server FastTrack Mapping Editor interface. The main window displays a mapping specification for 'Processo\_D\_DMA'. The 'Progress of Specification' section shows progress bars for 'Mapped' (100%), 'Reviewed' (0%), and 'Validated' (0%). A yellow callout box points to this section with the text 'Stato di lavorazione'. Below this is a 'Summary' table with columns 'Item', 'Count', and 'Percent'. A yellow callout box points to this table with the text 'Statistiche; entità e termini utilizzati'. At the bottom of the interface, a yellow callout box points to the 'Generate Job' button with the text 'Generazione job template DataStage'. The right-hand side of the interface shows a 'Validation' pane with a search bar and a list of database objects, and a 'Properties' pane with fields for 'Name Alias', 'Primary Key', and 'Referenced Tables'.

Item	Count	Percent
Source Tables	4 Tables	
THOR-COLL.ASCE.TEST_PROTOTIPO.MIGRAZIONETEST.D_EDW_EESF_ESERCIZI_FINANZIARI	1 of 18 Colu	5%
THOR-COLL.ASCE.TEST_PROTOTIPO.MIGRAZIONETEST.D_DMA_SSTS_STATUS	1 of 5 Colu	20%
THOR-COLL.ASCE.TEST_PROTOTIPO.MIGRAZIONETEST.D_EDW_EPCS_STRUTT_CAP_RAG_SP	1 of 11 Colu	9%
THOR-COLL.ASCE.TEST_PROTOTIPO.MIGRAZIONETEST.D_EDW_EPGC_PIANI_GEST_CAPITOLI	1 of 35 Colu	2%
Total	4 of 69 Colu	5%
Target Tables	1 Tables	
THOR-COLL.ASCE.TEST_PROTOTIPO.MIGRAZIONETEST.D_DMA_FCPG_PIANI_GEST	18 of 30 Col	60%
Total	18 of 30 Col	60%
Candidate Source Tables	0 Tables	
Candidate Target Tables	0 Tables	
Source Categories	1 Categories	
DW_RGS_90_CONC.Entità.Piano di Gestione	1 of 16 Ter	6%
Total	1 of 16 Ter	6%
Target Categories	0 Categories	
Lookups	4 Tables	
LK_FSPG	5 of 84 Colu	5%
LK_EPGC	2 of 35 Colu	5%
LK_FGPG	9 of 72 Colu	12%
LK_EDPC	7 of 53 Colu	13%
Total	23 of 244 C	9%
Candidate Terms		
Transformations		
Transformation Rules	18	
Transformation Functions	3	
Total	1	

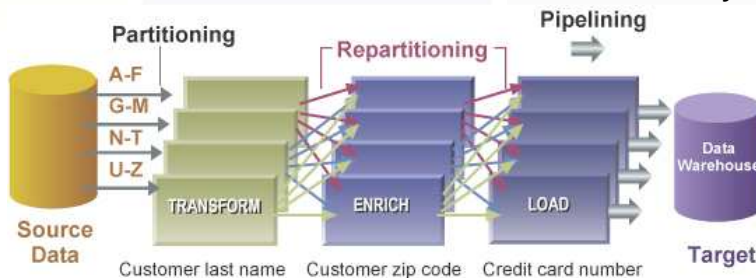
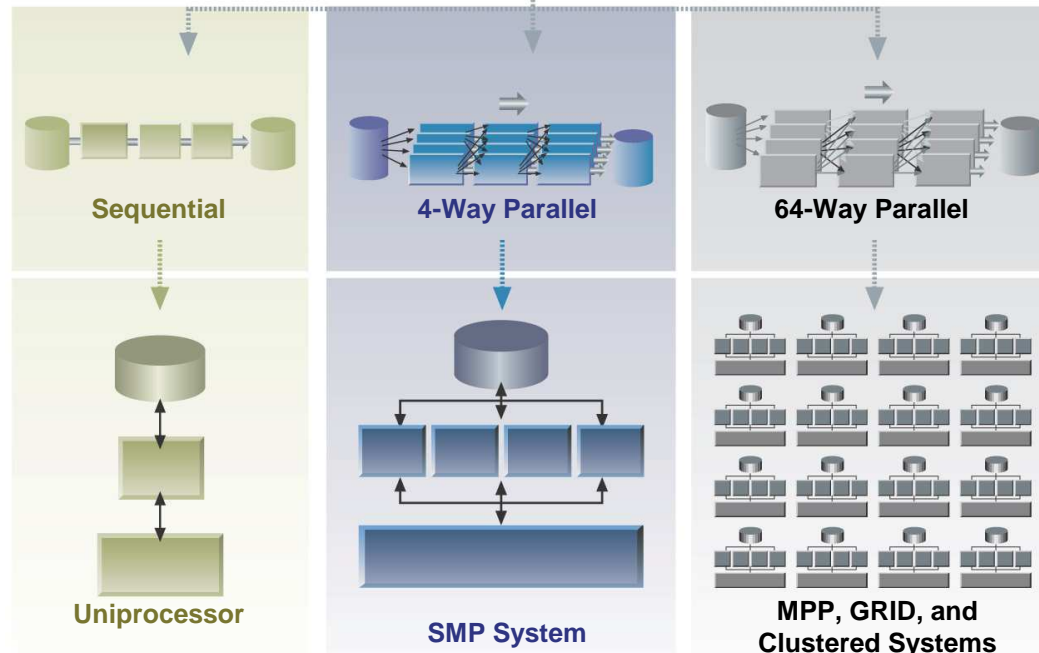


# Offrire le migliori performance

Assemblaggio della procedura: un singolo flusso



Esecuzione della procedura: sequenziale o parallela



## Perché IS DataStage ?

- Disegno sequenziale, deploy parallelo
- Scalabilità lineare
- Partizionamento dinamico dei dati e ripartizionamento “in-flight”
- Portabile su SMP, Cluster, GRID, Cloud e piattaforma MPP
- Supporto per RDBMS paralleli (IBM DB2 UDB, Oracle, Teradata, Informix)
- Parallelizzazione “codeless”
- Incorpora e rende parallele applicazioni esterne esistenti, senza scrittura di codice

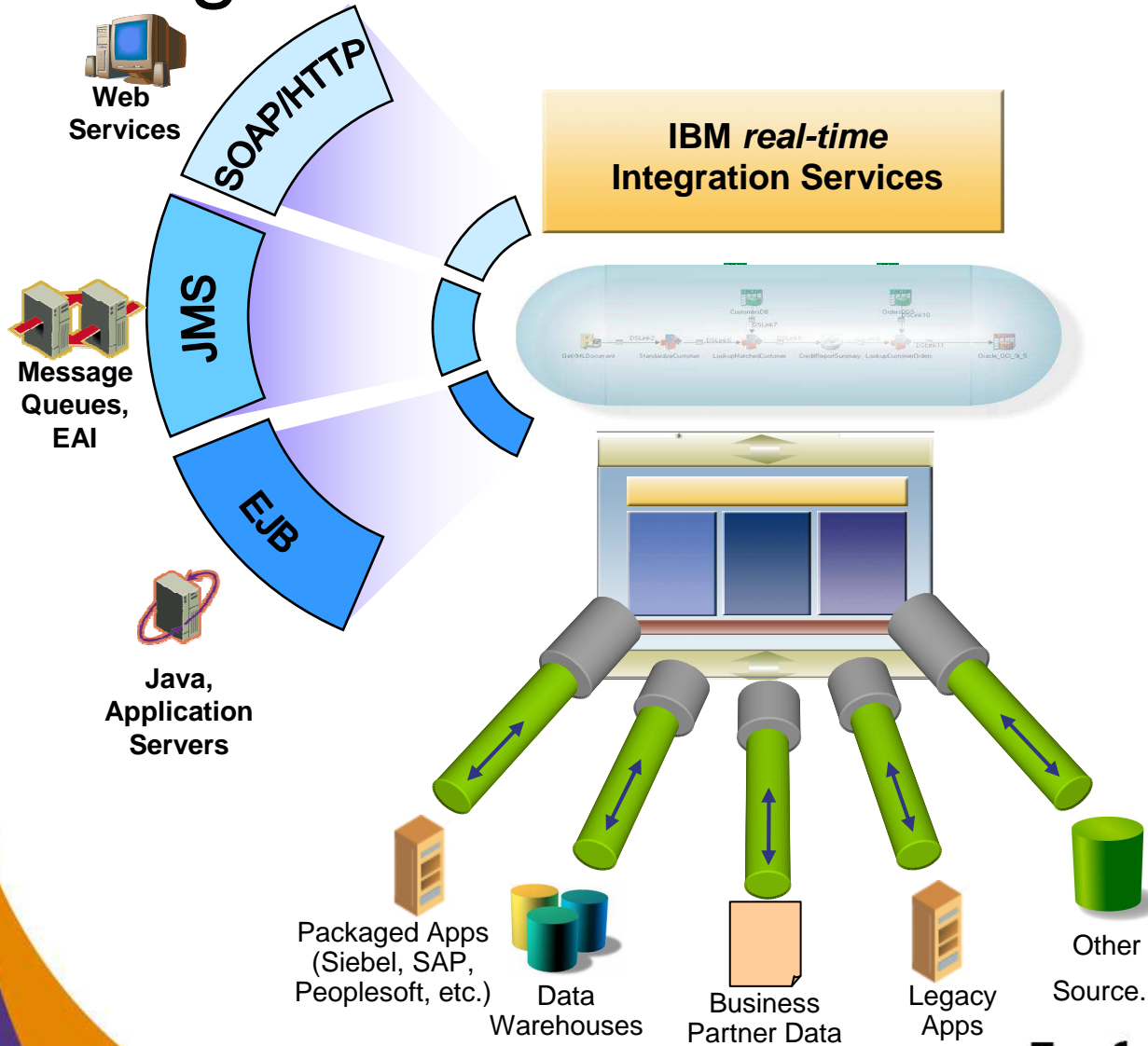
## Benefici per il business

- Elevata produttività
- Uso ottimale dell’hardware
- Flessibilità nell’esecuzione

**InfoPlanet**



# Integrazione con i mondi *real-time*




- Consente la trasformazione dei flussi batch in flussi real-time senza scrittura di codice
- Presenta la logica come un servizio attraverso un layer che astrae il servizio dalla logica sottostante
- Basato su plug-ins
- Opzioni di deployment multiple: JMS, EJB, o Web services
- Implementa i meccanismi standard di sicurezza, logging e load-balancing tramite la compliance allo standard J2EE



# Gestire la filiera informativa

- Esplorazione *web-based* degli assets generati e utilizzati da InfoSphere 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**

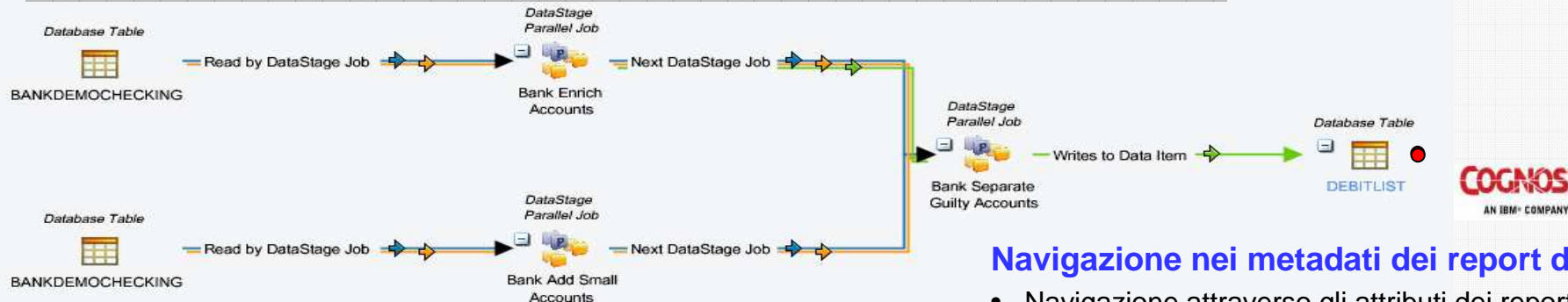


**IBM Metadata Workbench®**

Fornisce allo staff IT uno strumento per esplorare e comprendere il patrimonio generato e utilizzato dalla suite InfoSphere Information Server.



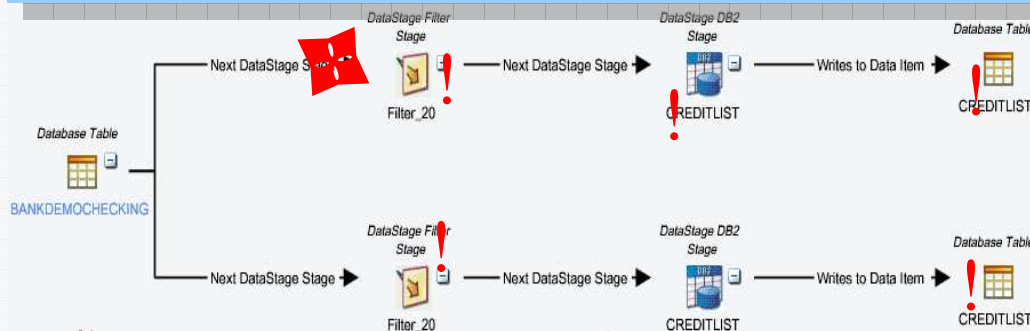
## Da dove provengono i dati relativi ad un campo di un report ?



### Navigazione nei metadati dei report di BI

- Navigazione attraverso gli attributi dei reports
- Data lineage anche rispetto a strumenti differenti
- Aumenta affidabilità e comprensione delle informazioni di business

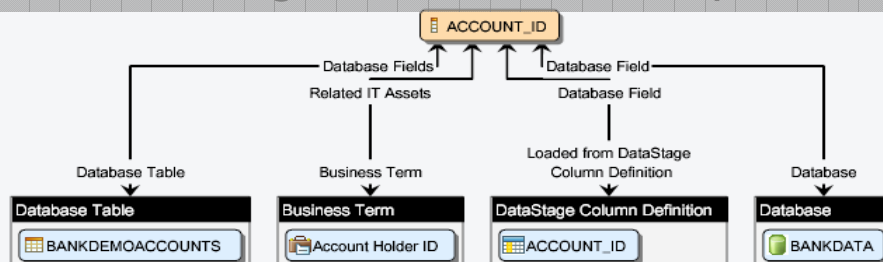
## Cosa accade modificando una colonna ?



### Analisi di impatto

- Mostra gli impatti in formato grafico o tabellare
- Anche su report degli strumenti di BI
- Riduce i costi associati ai cambiamenti

## Qual è il significato di un campo ?

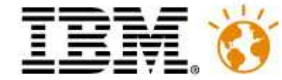


### Mostra le relazioni tra termini di business, entità di un modello dati, campi tecnici o di un report

- Permette la comprensione del significato dei campi
- Permette la comprensione delle relazioni tra termini di business
- Supporta la "data governance"

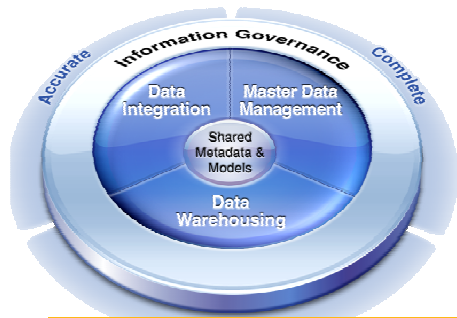






# IBM InfoSphere Information Server 8.5

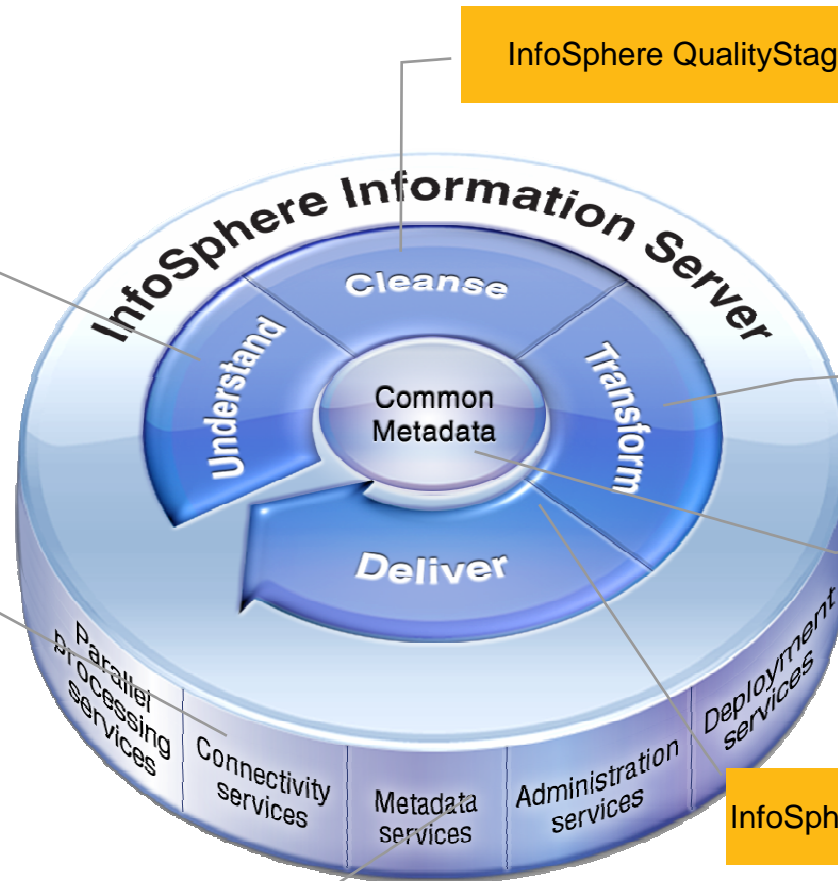
## Data Integration Semplificata



InfoSphere Business Glossary  
InfoSphere Information Analyzer  
InfoSphere FastTrack

Native connectivity  
(DB2, Informix, Oracle,  
Teradata...)

InfoSphere Metadata Workbench



InfoSphere QualityStage

InfoSphere DataStage

Business Glossary  
Business Glossary anywhere

InfoSphere Information Services Director

# InfoPlanet





# Thank You

**Brunello Bonanni**  
**IOD Certified Senior IT Architect**  
**IOD Center of Excellence Manager**  
**[brunello.bonanni@it.ibm.com](mailto:brunello.bonanni@it.ibm.com)**

