

CONTRASTO ALL'EVASIONE FISCALE

Formazione Specialistica

**L'utilizzo di Tecniche di data warehousing
per indirizzare le attività di Contrasto
all'Evasione**

Mondovì, 29 settembre 2017
Massimiliano Messina
CSI Piemonte



Piemonte

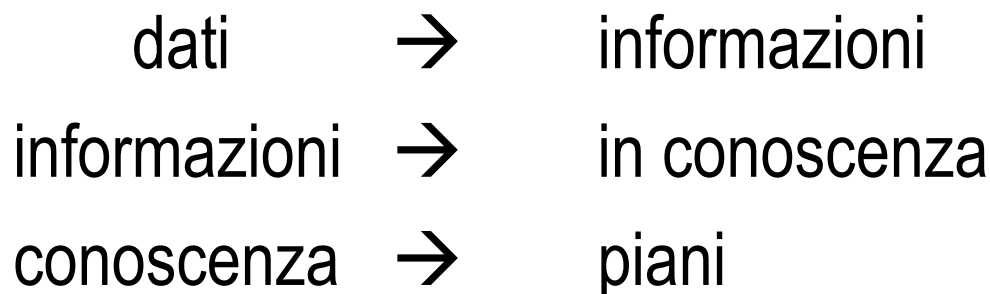
Utilizzo delle Banche Dati

Perché ?

Con quali strumenti ?

Business Intelligence

La business intelligence è il processo, e la tecnologia alla base, che permette la trasformazione di



che orientano il processo decisionale ai vari livelli dell'organizzazione.

Business intelligence

Per consentire ***analisi potenti e flessibili*** è necessario definire un'apposita infrastruttura hardware e software di supporto composta da:

- Hardware dedicato
- Infrastrutture di rete
- DBMS
- Software di back-end
- Software di front-end

Il ruolo chiave è ***la trasformazione dei dati in informazioni*** fruibili a diversi livelli di dettaglio

Dai dati alle informazioni

- L'informazione è un bene a valore crescente, necessario per pianificare e controllare le attività con efficacia □
- la disponibilità di troppi dati rende arduo, se non impossibile, estrapolare le informazioni veramente importanti

~~dati = informazione~~

Dai dati alle informazioni

È fondamentale poter disporre in maniera rapida e completa delle informazioni necessarie al processo decisionale:

- le indicazioni strategiche sono estrapolate principalmente dalla mole dei dati operazionali contenuti nei database,
- attraverso un procedimento di selezione e sintesi progressiva

Il nuovo ruolo dell'informatica

- L'aumento esponenziale del volume dei dati operazionali ha reso il calcolatore l'unico supporto adatto al processo decisionale .
- L'utilizzo massiccio di tecniche di analisi dei dati ha reso il sistema informativo un elemento strategico
- Il ruolo dell'informatica è passato da passivo strumento per la registrazione delle operazioni a fattore decisivo per la individuazione di elementi critici dell'organizzazione e di potenziali aree di business

I Data Warehouse

Un Data Warehouse è una collezione di dati di supporto per il processo decisionale che presenta le seguenti caratteristiche:

- *è orientato ai soggetti di interesse;*
- *è integrato e consistente;*
- *è rappresentativo dell'evoluzione temporale;*
- *non volatile.*

Caratteristiche del processo di warehousing

- **accessibilità** a utenti con conoscenze limitate di informatica e strutture dati;
- **integrazione** dei dati sulla base di un modello standard dell'impresa;
- **flessibilità** di interrogazione per trarre il massimo vantaggio dal patrimonio informativo esistente;

Caratteristiche del processo di warehousing

- **sintesi** per permettere analisi mirate ed efficaci;
- rappresentazione multidimensionale per offrire all'utente una **visione intuitiva ed efficacemente**
- **manipolabile** delle informazioni;
- correttezza e completezza dei dati integrati.

Acquisizione dei dati nel DWH

I dati vengono

- **Estratti dai sistemi sorgenti**
- **Ripuliti**
- **Integrati**

Pulitura

Si incarica di migliorare la qualità dei dati delle sorgenti

- dati duplicati
- inconsistenza tra valori logicamente associati
- dati mancanti
- uso non previsto di un campo
- valori impossibili o errati
- valori inconsistenti per la stessa entità dovuti a differenti convenzioni
- valori inconsistenti per la stessa entità dovuti a errori di battitura

Trasformazione

Converte i dati dal formato operativo sorgente a quello del DW.

La corrispondenza con il livello sorgente è complicata dalla presenza di fonti distinte eterogenee, che richiede una complessa fase di integrazione

- presenza di testi liberi che nascondono informazioni importanti
- utilizzo di formati differenti per lo stesso dato

Trasformazione

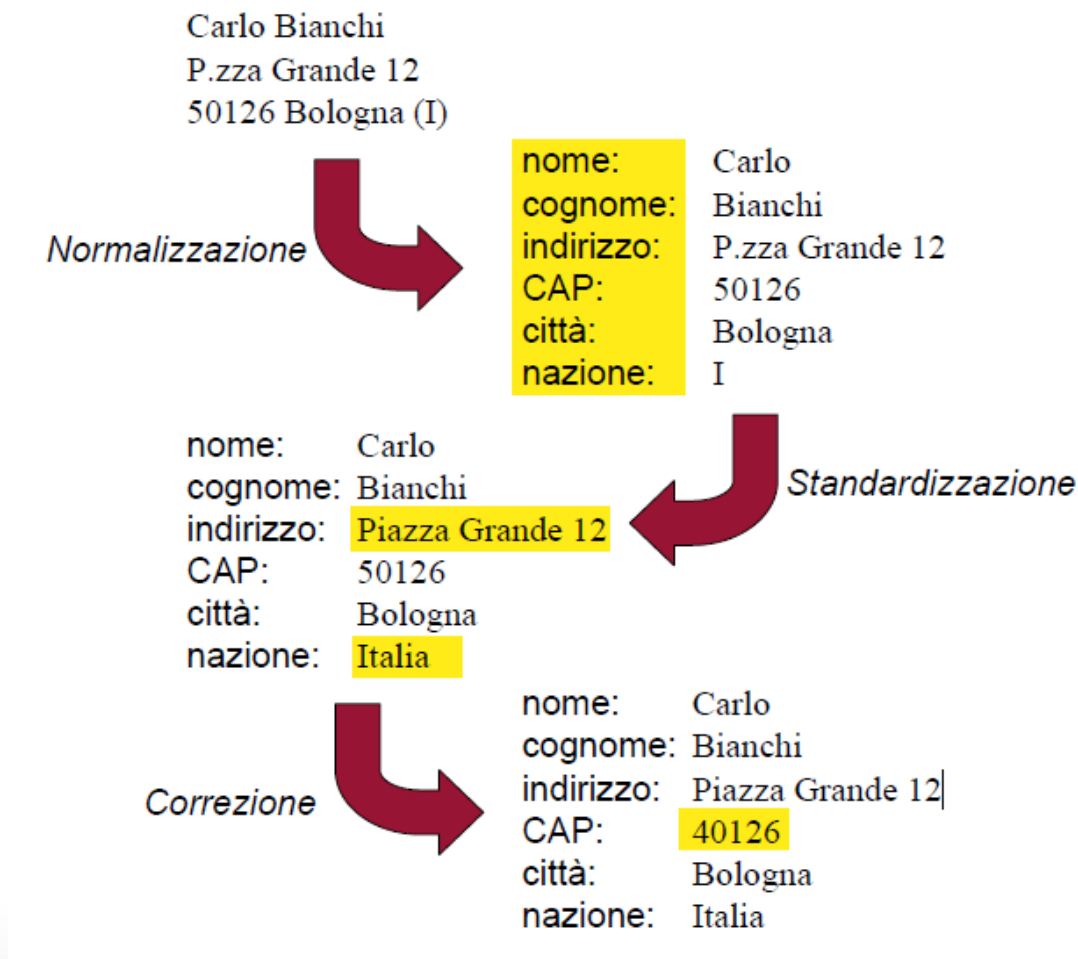
. Per l'alimentazione dei dati riconciliati:

- **conversione e normalizzazione** (operano a livello di formato di memorizzazione e di unità di misura per uniformare i dati)
- **matching** (stabilisce corrispondenze tra campi equivalenti in sorgenti diverse)
- selezione (riduce il numero di campi e di record rispetto alle sorgenti)

. Per l'alimentazione del DW:

- la **normalizzazione** è sostituita dalla denormalizzazione
- si introduce l'**aggregazione**, che realizza le opportune sintesi dei dati

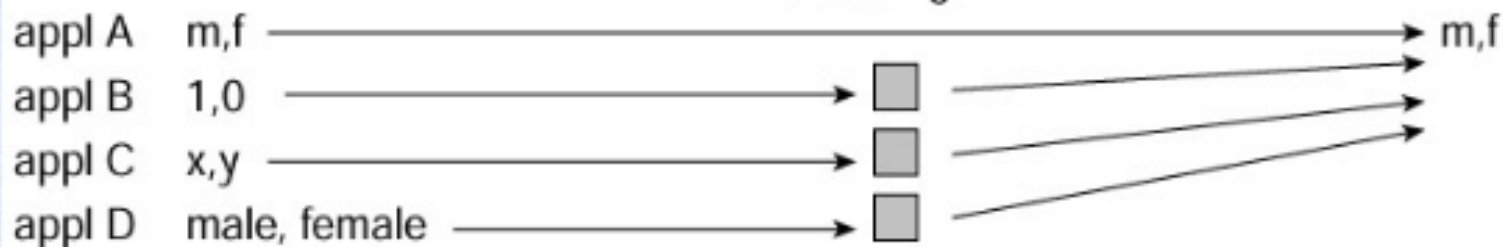
Pulitura e trasformazione



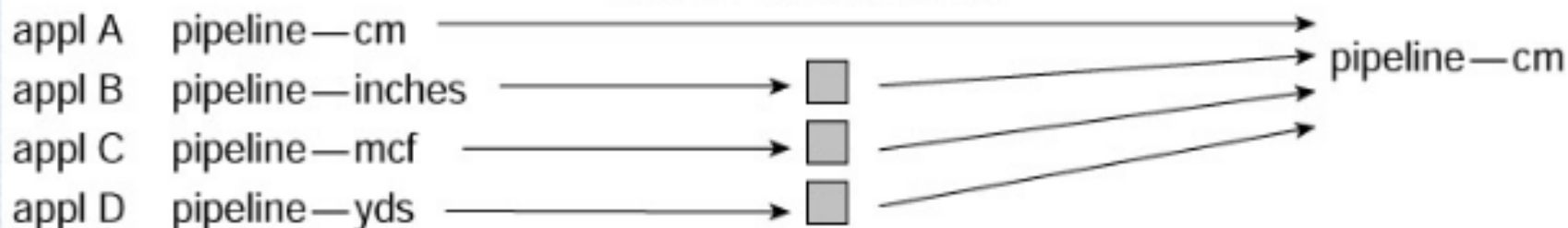
operational

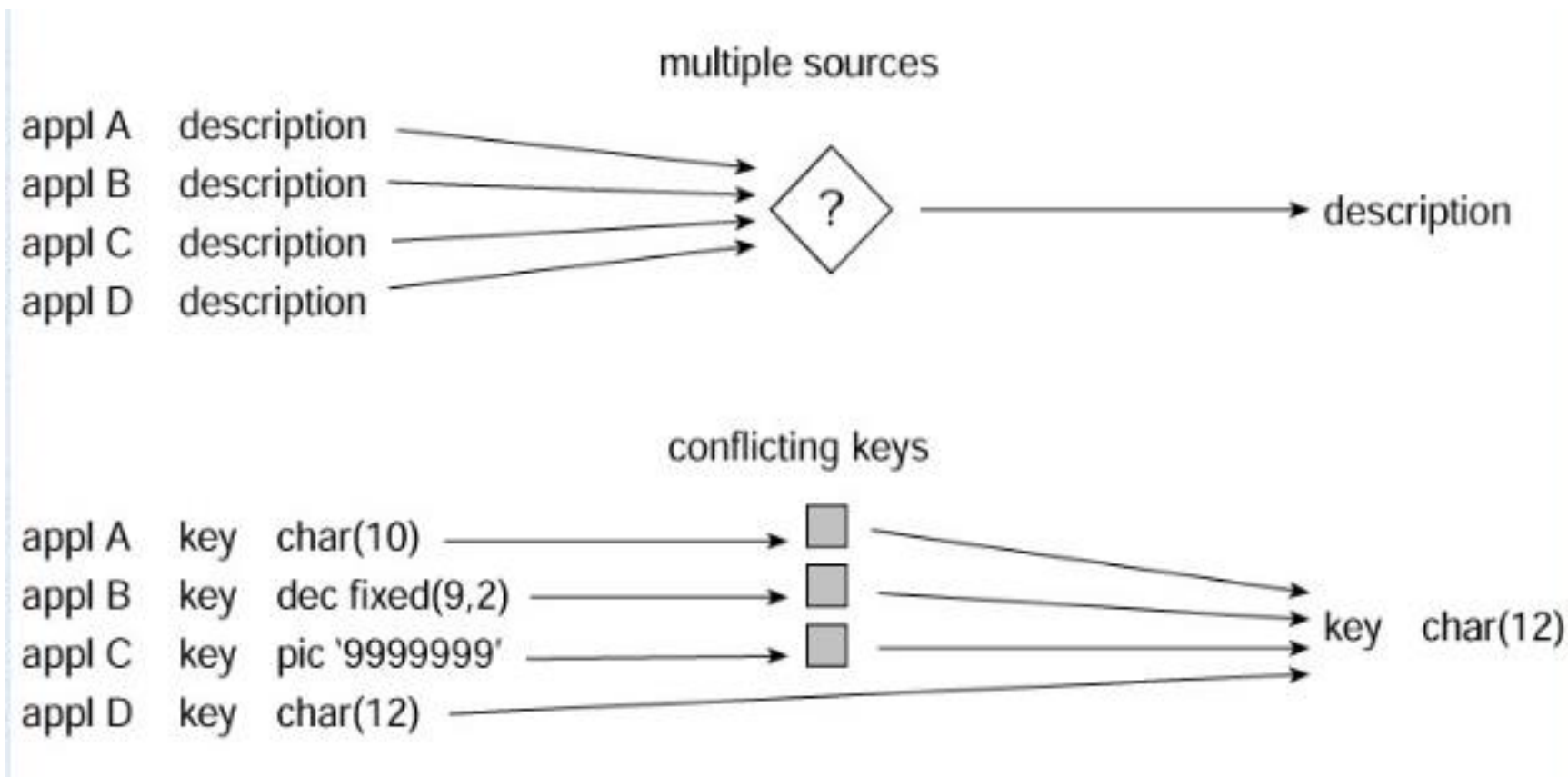
data warehouse

encoding



attribute measurement





Pulitura e trasformazione

Il sistema deve

- raffrontare le informazioni non solo per “identità di valori” ma anche per **“similitudine”**
- consentire l'individuazione di “record anagrafici corrispondenti” (record matching) anche solo per “mera approssimazione”.

Pulitura e trasformazione

Esempio 1

Da Tari : «via Giuseppe Verdi»

Da Catasto: «via G. verdi»

Esempio 2

Da Anagrafe :

«MSSMSM64C12L291W»«Messina»«Massimiliano»«Torino»«12/03/1964»

Da Catasto :

«MSSMSS64C12L219W»«Messina»«Massimiliano»«Torino»«12/03/1964»

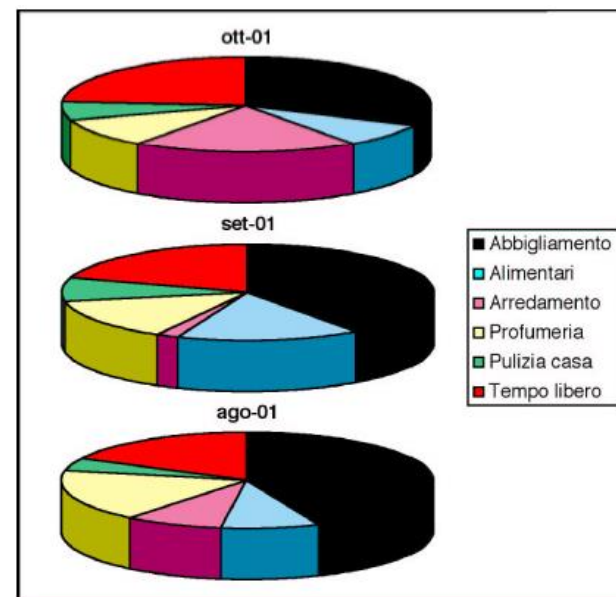
Tecniche di analisi dei dati

Esistono in sostanza **tre** approcci differenti,

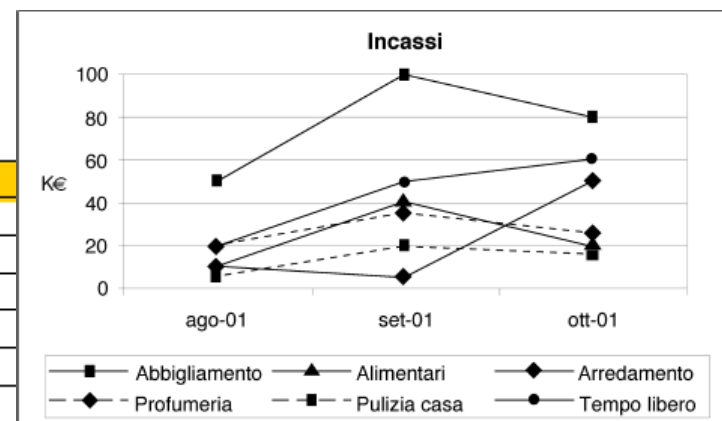
- ***reportistica***: non richiede conoscenze informatiche
- ***OLAP***: richiede all'utente di ragionare in modo multidimensionale e di conoscere l'interfaccia dello strumento grafico utilizzato
- ***data mining***: richiede all'utente la conoscenza dei principi che stanno alla base degli strumenti utilizzati

Reportistica

Orientato agli utenti che hanno
necessità di accedere, a
intervalli di tempo predefiniti, a
Informazioni strutturate in
modo pressoché invariabile



incassi (K€)	Ottobre 2001	Settembre 2001	Agosto 2001
Abbigliamento	80	100	50
Alimentari	20	40	10
Arredamento	50	5	10
Profumeria	25	35	20
Pulizia casa	15	20	5
Tempo libero	60	50	20



OLAP

Consente, a utenti le cui necessità di analisi non siano facilmente identificabili a priori, di **analizzare** ed **esplorare interattivamente** i dati sulla base del modello multidimensionale

- estemporaneità delle sessioni di lavoro
- richiesta approfondita conoscenza dei dati
- complessità delle interrogazioni formulabili
- orientamento verso utenti non esperti di informatica
interfaccia flessibile, facile da usare ed efficace

Data mining

Attività orientata a scoprire informazioni nascoste nei dati

In presenza di moli di dati molto elevate, l'utente non è sempre in grado di individuare tutti i pattern (modelli) significativi presenti

Data mining

Il data mining raccoglie tecniche di intelligenza artificiale e pattern recognition per aiutare l'utente nella ricerca di pattern: è sufficiente indicare cosa e dove si vuole ricercare

- Analisi delle abitudini
- Pianificazione
- **Rilevamento di attività fraudolente**
- Valutazione delle categorie di rischio
- Riconoscimento di similarità tra sequenze di eventi

Esempio : Locazioni

- Immobili locati senza Tari
 - «locazioni»
 - «anagrafe»
 - «redditi»
 - «tari»
- Immobili locati non dichiarati
 - «Tari»
 - «catasto»
 - «anagrafe»
 - «redditi»
 - «locazioni»

Prospettive future

Il web scraping o data scraping

Tecnica che consiste nell'estrarre informazioni e dati da siti web o pagine web grazie all'utilizzo di script, allo scopo di poter **riutilizzare le informazioni prelevate per i propri scopi.**

Grazie per l'attenzione

Massimiliano Messina
massimiliano.messina@csi.it

Direzione Governo Servizi e Soluzioni Applicative Per La P.A.
Fiscalità, Catasto e Polizia Municipale