

# Integrazione di basi di dati

## Caso di Studio

tratto da

M. Golfarelli, S. Rizzi.

Data Warehouse - Teoria e pratica della progettazione.

McGrawHill, 2002

# Il dominio applicativo

La società, che chiameremo *StraSport*, si occupa principalmente di vendita all'ingrosso; per gestire le proprie vendite si affida ad agenzie distribuite sul territorio italiano, ciascuna delle quali gestisce un certo numero di agenti cui è assegnata una propria zona di lavoro. La maggior parte dei clienti sono grandi catene di negozi con più sedi dislocate sul territorio.

Una vendita è rappresentata da un documento di vendita composto da una o più righe di dettaglio che rappresentano le unità di merce venduta. Gli articoli trattati da *StraSport* hanno un forte carattere stagionale, il che porta a un andamento disomogeneo delle vendite dei diversi prodotti durante l'anno e quindi a un'accentuata fluttuazione del fatturato. L'anno solare comprende principalmente due campagne, Autunno/Inverno e Primavera/Estate, che coprono mesi diversi con riferimento alla raccolta degli ordini e alla fatturazione.

Poco prima dell'inizio di una campagna viene stilato un listino comprendente tutti gli articoli che verranno presentati ai clienti per raccogliere gli ordini e successivamente fatturarli. I prodotti venduti vengono divisi in due categorie, nuovi e classici. I primi sono stati inseriti nella collezione per rinnovare il campionario in funzione delle nuove tendenze dettate dalla moda, i secondi vengono venduti con regolarità per diverse stagioni.

Il budget di esercizio viene stilato una volta all'anno, in estate, per pianificare gli articoli da vendere e gli importi da fatturare in ogni stagione; ovviamente può essere variato durante il corso dell'esercizio. Il budget viene poi affidato alla direzione commerciale e a quella del marketing, che successivamente si occupano di dividere i dati previsionali per cliente e per prodotto.

# I database

Il sistema informativo è gestito attraverso un sistema AS/400 su cui coesistono due distinti database relazionali:

**DB Amministrazione:** gestisce ordini e fatture, è utilizzato da oltre quindici anni per la gestione delle pratiche amministrative più direttamente legate alle attività di vendita.

<i>Nome relazione</i>	<i>Descrizione</i>
CLIENTE	dati anagrafici di base sui clienti
SCHEDA_CLIENTE	informazioni aggiuntive su clienti
FILIALE	informazioni sulle filiali di ciascun cliente
ARTICOLO	informazioni di base sugli articoli vendibili da StraSport
SCHEDA_ARTICOLO	informazioni aggiuntive sugli articoli
VENDITORE	agenti di vendita e loro agenzie
MOVIMENTO	testate delle fatture
MOVIMENTO_RIGHE	singole righe delle fatture
ORDINE	testate dei documenti di vendita
ORDINE_RIGHE	singole righe degli ordini (una tupla per ciascun articolo)

**DB Marketing:** viene utilizzato dalla direzione generale e dalla divisione marketing per la programmazione e la gestione delle campagne di vendita (*budgeting*) e dei risultati attesi per agenti di vendita e articoli.

<i>Nome relazione</i>	<i>Descrizione</i>
ARTICOLO	informazioni di base sugli articoli vendibili da StraSport
SCHEDA_ARTICOLO	informazioni aggiuntive sugli articoli
LISTINO	prezzi di vendita degli articoli
SOTTOSEGMENTO	sottosegmenti degli articoli ('costumi da gara', 'costumi da spiaggia' ecc.)
SOTTOCATEGORIA	sottocategorie degli articoli ('tessile', 'calzature' ecc.)
AGENTE	agenti e le loro agenzie
BUDGET_COMMERCIALE	dati di interesse relativamente al budget commerciale
BUDGET_MARKETING	dati di interesse relativamente al budget di marketing

Pur essendo realizzati sulla stessa piattaforma, i due sistemi sono scarsamente integrati e presentano sovrapposizioni e conflitti.

# La documentazione I

La documentazione iniziale fornita dai gestori del sistema informativo per la ricognizione della base dati è una stampa, ottenuta dal programma di esportazione del DBMS. Il significato delle colonne è il seguente:

- *Da* indica la posizione di inizio dell'attributo.
- *A* indica il limite dell'attributo.
- *Nome* è il nome dell'attributo.
- *Tipo* può assumere due valori: 'A' significa che l'attributo è memorizzato normalmente, 'P' che il dato è rappresentato in formato compatto.
- *Lunghezza* indica quanti byte occupa l'attributo.
- *Descrizione* descrive brevemente il significato dell'attributo.

## Documentazione per alcuni attributi della relazione CLIENTE

<i>Da</i>	<i>A</i>	<i>Nome</i>	<i>Tipo</i>	<i>Lunghezza</i>	<i>Descrizione</i>
1	1	CLITRK	A	1	Tipo record D/A
2	6	CLIDIT	A	5	Codice ditta
7	12	CLICOD	A	6	Codice CLI/FOR
13	41	CLIRG1	A	29	Ragione sociale
42	71	CLIIND	A	30	Indirizzo

# La documentazione: DB amministrazione

## ORDINE

<i>Nome</i>	<i>Testo descrittivo</i>	<i>Dominio</i>	<i>Importato da</i>
<u>AAORCT</u>	Anno Ordine	Integer	
<u>NRORCT</u>	Numero Ordine	Integer	
<u>CLORCT</u>	Classe Ordine	String	
<u>CDDICT</u>	Codice ditta	String	FILIALE
<u>CDCLCT</u>	Codice cliente	String	FILIALE
<u>CDFICT</u>	Codice Filiale	String	FILIALE
<u>CDA1CT</u>	Venditore	String	VENDITORE
<u>CDPACT</u>	Pagamento	String	<i>Decodifica</i>
<u>DTCOCT</u>	Data Consegna	Date	
<u>SCPACT</u>	Sconto sul pagamento	Float	
<u>SCINT</u>	Sconto Incondizionato	Float	
<u>CDANCT</u>	Causale Annullamento	String	<i>Decodifica</i>
<u>DTPACT</u>	Data decorrenza pagamento	Date	
<u>DTIMCT</u>	Data immissione ordine	Date	
<u>QTTOCT</u>	Quantità ordine totale	Float	
<u>RU07CT</u>	Tipo pagamento	String	<i>Decodifica</i>
<u>RGRCT</u>	Stagione	String	

## ORDINE\_RIGHE

<i>Nome</i>	<i>Testo descrittivo</i>	<i>Dominio</i>	<i>Importato da</i>
<u>AAORCD</u>	Anno Ordine	Integer	ORDINE
<u>NRORCD</u>	Numero Ordine	Integer	ORDINE
<u>RIORCD</u>	Riga Ordine	Integer	
<u>CLORCD</u>	Classe Ordine	String	
<u>CDARCD</u>	Codice Articolo	String	ARTICOLO
<u>QTTOCD</u>	Quantità Ordinata	Integer	
<u>DTCOCD</u>	Data consegna merce	Date	
<u>CDANCD</u>	Causale Annullamento	String	<i>Decodifica</i>
<u>CDA1CD</u>	Venditore	String	VENDITORE

## ARTICOLO

<i>Nome</i>	<i>Testo descrittivo</i>	<i>Dominio</i>	<i>Importato da</i>
<u>CDARAR</u>	Codice Articolo	String	
<u>DEARAR</u>	Descrizione Articolo	String	
<u>DECOAR</u>	Descrizione Colore	String	
<u>CDMOAR</u>	Modello	String	
<u>CDIVAR</u>	Codice IVA	String	
<u>UBARAR</u>	Ubicazione in magazzino	String	
<u>CDFOAR</u>	Produttore	String	<i>Decodifica</i>
<u>LINEAR</u>	Linea	String	<i>Decodifica</i>

# Analisi e riconciliazione delle fonti dati

- Nella relazione ORDINE il riferimento al venditore è ridondante poiché ogni cliente può essere seguito da un solo venditore che quindi è determinato univocamente fissato il cliente. La ridondanza è vieppiù ovvia nella relazione ORDINE\_RIGHE.
- La chiave utilizzata in ORDINE è in realtà una superchiave poiché la numerazione degli ordini nei diversi anni è indipendente dalla classe d'ordine; l'inconsistenza è stata individuata poiché ORDINE\_RIGHE importa solo i primi due campi della chiave di ORDINE. A riscontro di tale ipotesi, si è effettuata una interrogazione sui dati che ha confermato che le tuple nella relazione sono univocamente identificate dai campi AAORCT e NRORCT.
- Per *classe d'ordine* si intende la modalità con cui viene effettuato un ordine (per esempio 'telefonico', 'campionario') durante una stagione (per esempio 'primavera/estate', 'riassortimento'). Un possibile valore del campo CLORCT è 'PE2002 campionario', cui corrisponde nel campo RGRCT il valore 'PE'. La relazione ORDINE risulta pertanto denormalizzata a causa della dipendenza funzionale CLORCT→RGRCT.
- L'attributo CLORCD in ORDINE\_RIGHE è ridondante, poiché la classe d'ordine è determinata dall'ordine cui la riga appartiene.

# La documentazione: DB marketing I

## ARTICOLO

<i>Nome</i>	<i>Testo descrittivo</i>	<i>Dominio</i>	<i>Importato da</i>
<u>CDARAR</u>	Codice Articolo	String	
DEARAR	Descrizione Articolo	String	
DECOAR	Descrizione Colore	String	
CDMOAR	Modello	String	
CDTEAR	Tema	String	<i>Decodifica</i>
AAPRAR	Anno di Produzione	Date	
STPRAR	Stagione di Produzione	String	
AAATAR	Anno di listino attuale	Integer	LISTINO
STPRAR	Stagione di listino attuale	String	LISTINO
LISTAR	Codice di listino	String	LISTINO
RG10AR	Designer	String	<i>Decodifica</i>
PREMAR	Prezzo medio	Float	
CDFOAR	Produttore	String	<i>Decodifica</i>
SESSAR	Target	String	<i>Decodifica</i>
LINEAR	Linea	String	<i>Decodifica</i>

## LISTINO

<i>Nome</i>	<i>Testo descrittivo</i>	<i>Dominio</i>	<i>Importato da</i>
<u>AAATLT</u>	Anno di listino attuale	Integer	
<u>ATATAL</u>	Stagione di listino attuale	String	
<u>CDARLT</u>	Codice articolo	String	ARTICOLI
<u>CL01LT</u>	Classe Ordine	String	
PREILT	Prezzo unitario	Float	
CL02LT	Classe Ordine 2	String	
CL03LT	Classe Ordine 3	String	

## SOTTOSEGMENTO

<i>Nome</i>	<i>Testo descrittivo</i>	<i>Dominio</i>	<i>Importato da</i>
<u>CODSOS</u>	Codice Sottosegmento	String	
DESSOS	Descrizione Sottosegmento	String	

## SOTTOCATEGORIA

<i>Nome</i>	<i>Testo descrittivo</i>	<i>Dominio</i>	<i>Importato da</i>
<u>CODSOC</u>	Codice Sottocategoria	String	
DESSOC	Descrizione Sottocategoria	String	

## BUDGET\_COMMERCIALE

<i>Nome</i>	<i>Testo descrittivo</i>	<i>Dominio</i>	<i>Importato da</i>
<u>BUDCAN</u>	Anno di vendita	Integer	
<u>RGCLOC</u>	Stagione	String	
<u>AGCOM</u>	Agente	String	AGENTE
<u>CODSOC</u>	Sottocategoria	String	
BUDCPRR	Previsione resi	Integer	
BUDCQT	Previsione quantità totale	Integer	
BUDCPRA	Previsione annullamenti	Integer	
BUDCPRS	Previsione sconti	Integer	
BUDCVL	Previsione valore lordo	Integer	

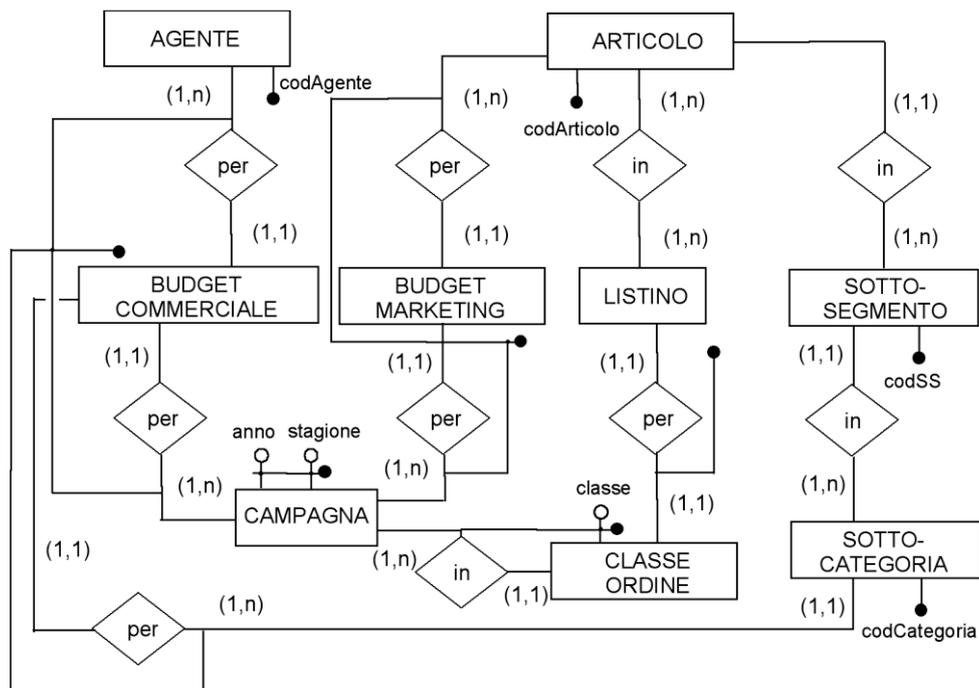
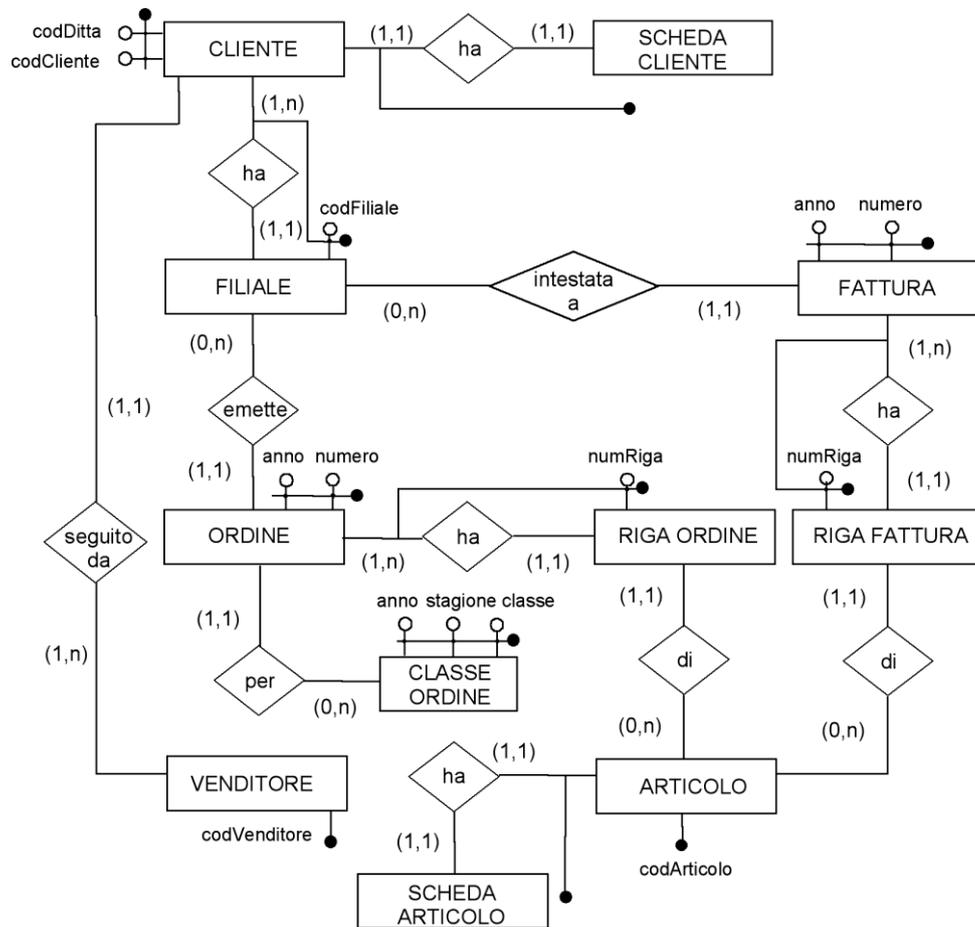
## BUDGET\_MARKETING

<i>Nome</i>	<i>Testo descrittivo</i>	<i>Dominio</i>	<i>Importato da</i>
<u>BUDMAN</u>	Anno di vendita	Integer	
<u>RGCLOC</u>	Stagione	String	
<u>ARTBUDM</u>	Codice Articolo	String	ARTICOLO
<u>CODSOS</u>	Sottosegmento	String	
SESSAR	Target	String	
BUDMPRR	Previsione resi	Integer	
BUDMPRA	Previsione annullamenti	Integer	
BUDMPRS	Previsione sconti fine anno	Integer	
BUDMVL	Previsione valore lordo	Integer	
BUDMQT	Previsione quantità totale	Integer	
BUDMFT	Previsione sconto fattura	Integer	

In **ARTICOLO** le informazioni relative al listino non sono di nessuna utilità poiché **LISTINO** mantiene già lo storico di tutti i listini in cui un articolo è stato inserito.

- **LISTINO** modella in forma denormalizzata l'appartenenza di un articolo a tre listini di una stessa campagna; in realtà il numero di listini varia di campagna in campagna e può comunque essere superiore a tre.
- Sebbene da **ARTICOLO** manchino chiavi esterne riferite a **SOTTOSEGMENTO**, dal colloquio con gli utenti emerge che gli articoli appartengono a sottosegmenti (per esempio 'costumi da gara', 'costumi da spiaggia'), a loro volta raggruppati in sottocategorie (per esempio 'tessili mare', 'calzature palestra'). Inoltre, in **SOTTOSEGMENTO** manca una chiave esterna che faccia riferimento a **SOTTOCATEGORIA**.
- La relazione **BUDGET\_MARKETING** risulta denormalizzata a causa delle dipendenze funzionali tra articolo e target e tra articolo e sottosegmento.

# Normalizzazione



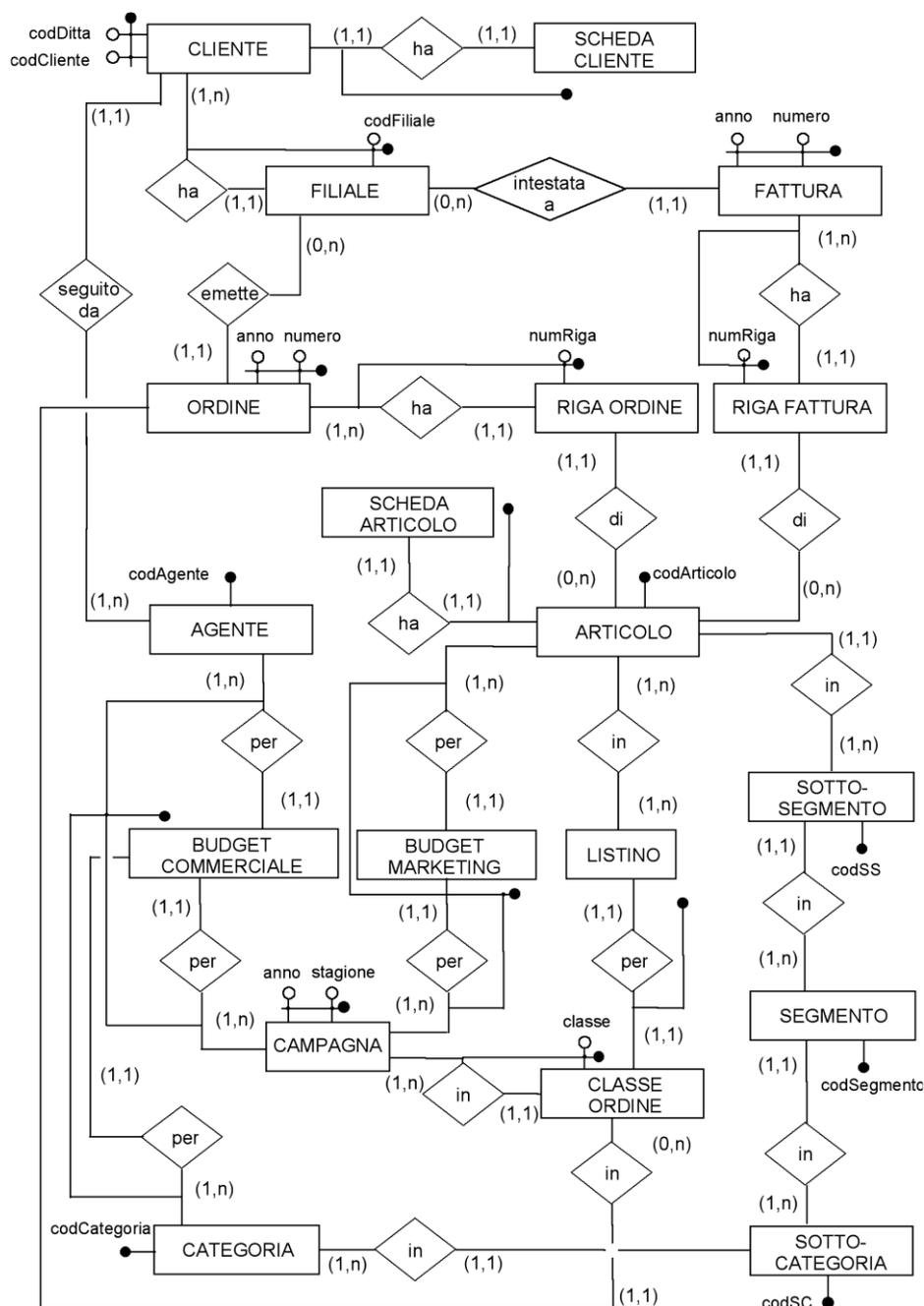
# Integrazione

La comparazione tra i due schemi evidenzia i seguenti conflitti:

- **Sinonimia** tra AGENTE e VENDITORE, che viene risolta a favore di AGENTE. Sebbene molti degli attributi comuni ai due schemi abbiano nomi diversi, le corrispondenze sono facilmente identificabili analizzandone la descrizione; anche in questi casi è necessario risolvere il conflitto per poter procedere alla sovrapposizione degli schemi.
- A seguito di una verifica sui dati, gli attributi CDFOAR delle tabelle ARTICOLO dei due database sono risultati essere **omonimi**. Infatti quello del DB Amministrazione memorizza i nomi dei fornitori del prodotto, mentre quello del DB Marketing memorizza l'effettivo produttore.
- Il concetto di campagna è espresso nello schema del DB Amministrazione mediante gli attributi stagione e anno dell'entità CLASSE ORDINE, nel DB Marketing tramite un'entità.
- Due delle entità comuni ai due schemi, AGENTE e ARTICOLO, descrivono lo stesso concetto da diversi punti di vista e perciò contengono diversi insiemi di attributi.

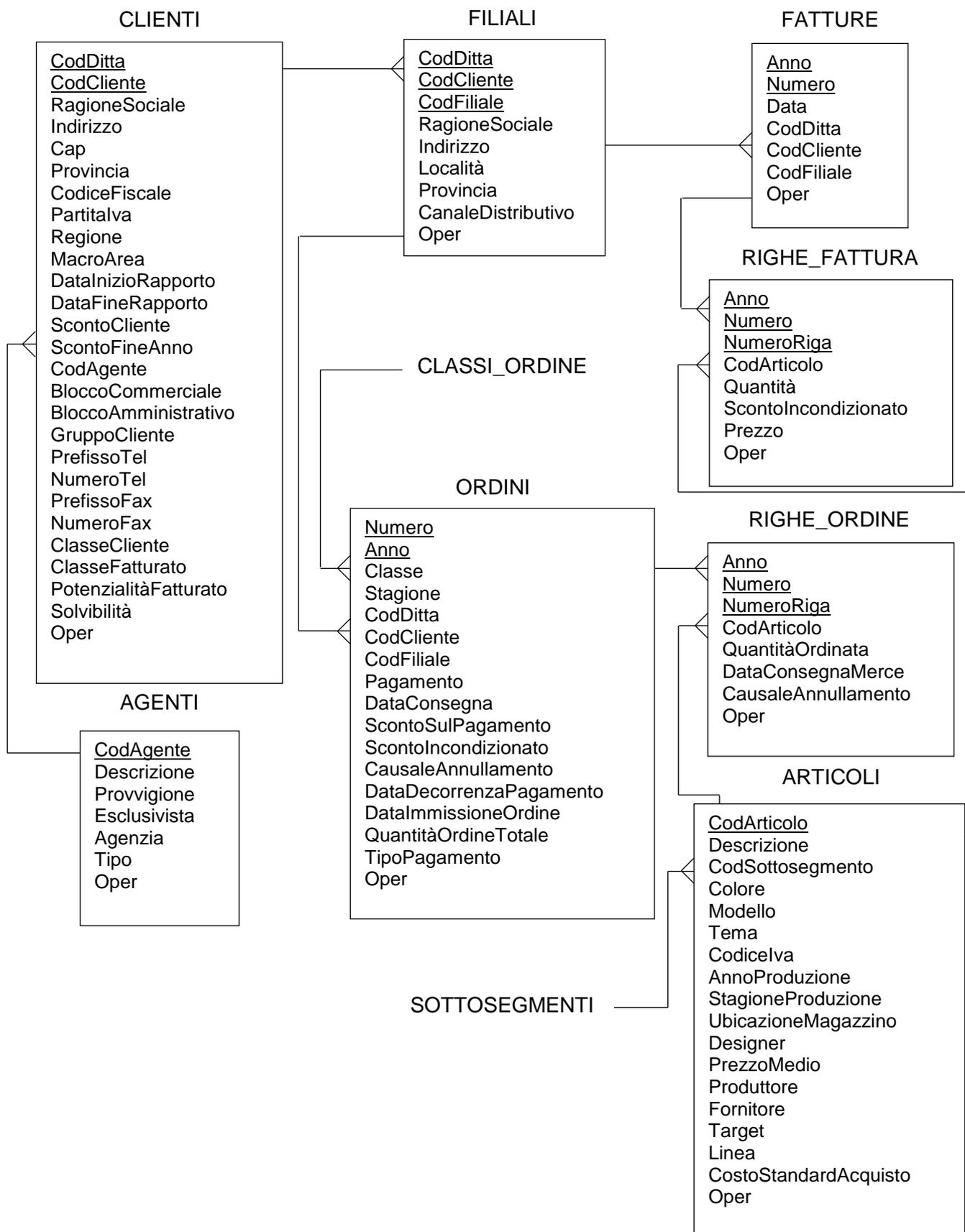
Oltre all'eliminazione delle sinonimie si rende necessario aggiungere nello schema del DB Amministrazione l'entità CAMPAGNA; l'omonimia evidenziatasi nelle entità ARTICOLO viene risolta rinominando l'attributo del DB Amministrazione. Per poter essere fuse, le entità AGENTE e ARTICOLO devono contenere lo stesso insieme di attributi, cioè l'unione dei due insiemi di attributi.

# Lo schema concettuale riconciliato

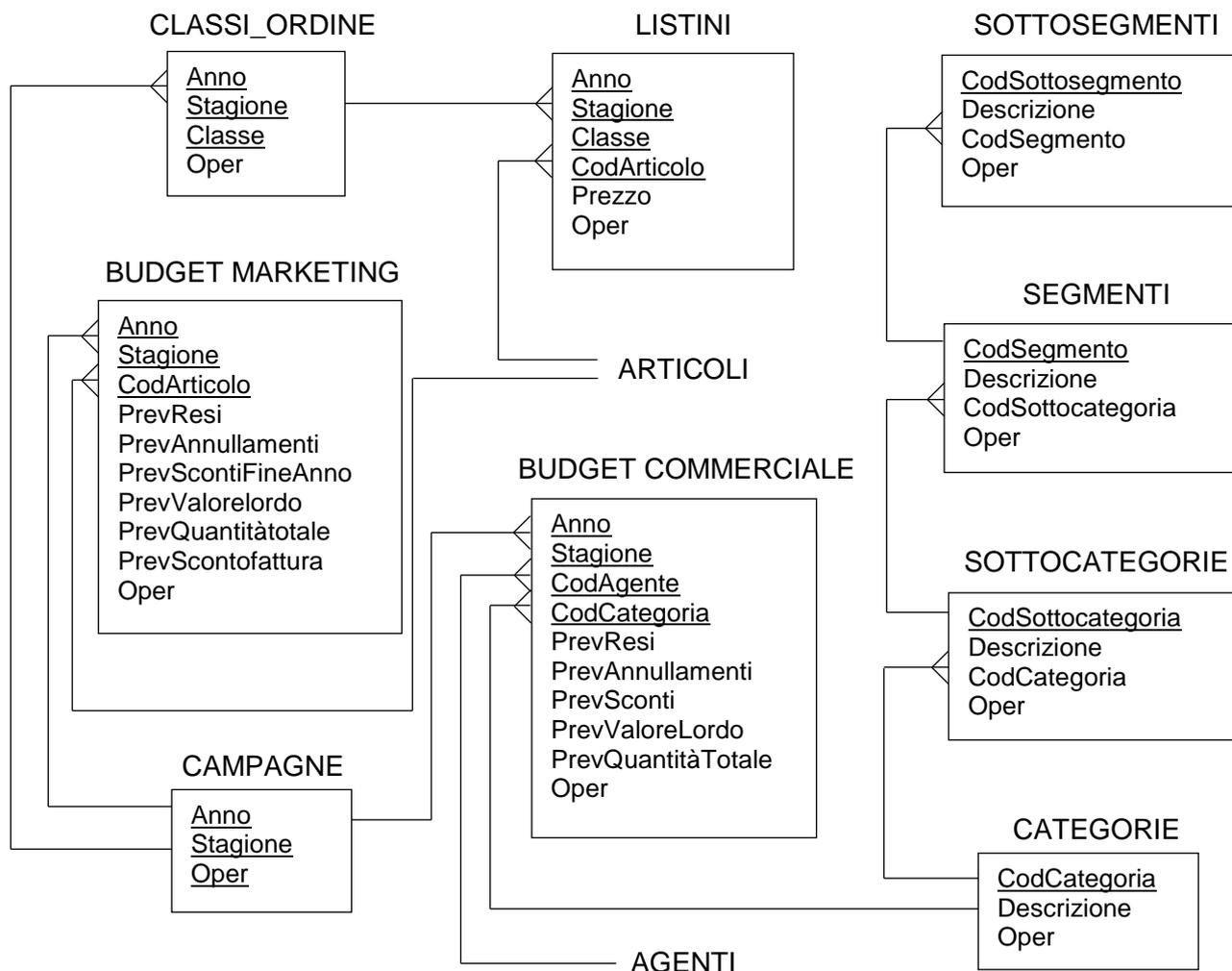


L'introduzione di CATEGORIA permette di modellare il fatto che, come dichiarato dagli utenti, il budget commerciale viene attualmente redatto per categoria piuttosto che per sottocategoria, come avveniva ai tempi della progettazione del DB Marketing. In effetti, da circa 10 anni nel campo CODSOC viene impropriamente memorizzata una categoria!

# Lo schema logico riconciliato I



# Lo schema logico riconciliato II



# Il mapping

Nel seguito è presentato un esempio di mapping tra schemi locali e livello riconciliato ottenuto con tecnica GAV

//Mapping dei clienti dal DB Amministrazione

```
CREATE VIEW CLIENTI AS
SELECT CL.CLIDIT, CL.CLICOD, CL.CLIRG1..., SC.PTELSC, SC.TELSC...
FROM CLIENTE AS CL, SCHEDA_CLIENTE AS SC
WHERE CL.CLIDIT= SC.CLIDIT AND CL.CLICOD = SC.CLICOD
```

// Mapping delle campagne dal DB Marketing

```
CREATE VIEW CAMPAGNE AS
SELECT BUDCAN, RGCLOC
FROM BUDGET_COMMERCIALE
UNION
SELECT BUDMAN, RGCLOC
FROM BUDGET_MARKETING
```