

**UNIVERSITÀ DEGLI STUDI DI PISA**  
**FACOLTÀ DI SCIENZE, MATEMATICHE, FISICHE E NATURALI**  
**CORSO DI LAUREA SPECIALISTICA IN INFORMATICA PER**  
**L'ECONOMIA E PER L'AZIENDA**

**TESI DI LAUREA**

**PROGETTAZIONE E REALIZZAZIONE DI UN**  
**DATAWAREHOUSE: IL CASO DI STUDIO DELLA**  
**COOPERATIVA AGRICOLA**

**RELATORE:**

Prof. Salvatore RUGGIERI

**Candidato:**

Michele DONATI

ANNO ACCADEMICO 2008-09



# RIASSUNTO

Questo lavoro consiste nella realizzazione di un datawarehouse per il supporto ai processi aziendali di controllo delle vendite, dell'inventario di magazzino e delle movimentazioni di cassa. Vengono esposte le diverse fasi della costruzione del datawarehouse: lo studio dei processi e l'analisi dei requisiti, la progettazione, l'analisi degli strumenti software utilizzati, la realizzazione delle procedure di estrazione, la trasformazione e il caricamento. Di ogni fase saranno presentate sia le problematiche di ordine generale, sia le soluzioni ai problemi riscontrati durante l'esperienza diretta con la realtà aziendale.

# Indice

<b>1</b>	<b>INTRODUZIONE</b>	<b>6</b>
1.1	Natura e fini della Cooperativa Agricola di Legnaia . . . . .	7
1.2	Contenuto della tesi . . . . .	9
<b>2</b>	<b>IL PROCESSO DI PROGETTAZIONE</b>	<b>11</b>
2.1	Le figure professionali coinvolte . . . . .	11
2.2	Modellazione dei requisiti . . . . .	12
2.2.1	Analisi dei requisiti . . . . .	12
2.2.2	Progettazione concettuale iniziale dei datamart . . . . .	12
2.2.3	Progettazione concettuale dei datamart dai dati operazionali	13
2.2.4	Progettazione concettuale finale dei datamart . . . . .	13
2.2.5	Progettazione logica dei datamart e del datawarehouse . .	13
<b>3</b>	<b>GLOSSARIO</b>	<b>14</b>
<b>4</b>	<b>DESCRIZIONE DEI PROCESSI AZIENDALI</b>	<b>17</b>
4.1	Processo di vendita . . . . .	17
4.2	Inventario di magazzino . . . . .	19
4.3	Analisi finanziarie . . . . .	22
<b>5</b>	<b>PROGETTAZIONE DEI DATA MART E DEL DATAWAREHOUSE</b>	<b>25</b>
5.1	I requisiti . . . . .	25
5.2	La base di dati . . . . .	26
5.3	Le specifiche . . . . .	28
5.4	Le tipologie di trattamento . . . . .	30
5.5	Progettazione concettuale iniziale dei data mart . . . . .	38
5.6	Progettazione concettuale dei data mart dai dati operazionali . .	40

5.6.1	Analisi dati operazionali . . . . .	40
5.6.2	Classificazione delle entità . . . . .	42
5.6.3	Definizione degli schemi concettuali candidati dei data mart . . . . .	42
5.7	Progettazione concettuale finale dei data mart . . . . .	44
<b>6</b>	<b>RAPPRESENTAZIONE LOGICA DEI DATA MART E DEL DATAWAREHOUSE</b>	<b>47</b>
<b>7</b>	<b>AMBIENTE DI SVILUPPO</b>	<b>51</b>
7.1	Indici utilizzabili per facilitare le interrogazioni e chiavi primarie	51
7.2	Software gestionale ForeOffice . . . . .	52
7.3	Scelta dell'ambiente di sviluppo . . . . .	53
<b>8</b>	<b>PROCEDURE DI ETL</b>	<b>55</b>
8.1	Il processo di ETL . . . . .	55
8.2	Scelta del software di ETL . . . . .	56
8.3	Area di staging . . . . .	57
8.4	Processo di trasformazione . . . . .	59
8.5	Processo di caricamento . . . . .	60
8.6	Esecuzione della procedura . . . . .	61
<b>9</b>	<b>ANALISI OLAP</b>	<b>62</b>
9.1	Struttura multidimensionale MOLAP . . . . .	63
9.1.1	Realizzazione del cubo . . . . .	64
9.1.2	Visualizzazione finale dei dati . . . . .	64
9.2	Struttura relazionale ROLAP . . . . .	66
9.3	Interrogazioni . . . . .	67
<b>10</b>	<b>CONCLUSIONI</b>	<b>72</b>
	<b>Riferimenti bibliografici</b>	<b>74</b>

# Capitolo 1

## INTRODUZIONE

Al giorno d'oggi le aziende di tutte le dimensioni sono sempre alla ricerca di strumenti, strategie e soluzioni per migliorare le performance e la propria competitività nei confronti della concorrenza.

Tuttavia ogni tentativo di miglioramento o di ottimizzazione ha bisogno di essere analizzato e valutato secondo degli indicatori che spesso il management o i responsabili non conoscono. Eppure lo sviluppo degli strumenti informatici all'interno di quasi tutte le realtà aziendali ha portato ad un aumento significativo delle informazioni a disposizione della catena gestionale.

Il problema consiste nel riuscire ad interpretare questa ingente mole di informazioni per ricavare indicatori chiari ed efficaci per la valutazione delle prestazioni di un determinato processo aziendale.

Lo scopo principale della tecnologia dei datawarehouse è proprio quello di riorganizzare e sintetizzare le informazioni immagazzinate dai sistemi operazionali permettendo di condurre analisi immediate sull'andamento di determinati processi.

Il datawarehouse oggetto della tesi è stato sviluppato in collaborazione con Neonevis.srl per la Cooperativa Agricola di Legnaia, società di medie dimensioni che opera in Toscana all'interno del mercato ortoflorovivaistico e ortofrutticolo.

Il budget a disposizione per lo sviluppo del datawarehouse era molto limitato. Non si è dunque tentato di sviluppare uno strumento che analizzasse tutta l'attività dell'azienda nel suo complesso ma, al contrario, si è scelto di lavorare sui processi ritenuti più significativi per le attività aziendali e che maggiormente avrebbero beneficiato di un sistema per il supporto alle decisioni.

Inoltre, sempre con l'obiettivo di mantenere i costi i più bassi possibili, si è scelto di utilizzare esclusivamente i software già di proprietà dell'azienda.

## 1.1 Natura e fini della Cooperativa Agricola di Legnaia

La Cooperativa Agricola di Legnaia è un'azienda di medie dimensioni in continua crescita che opera nel settore della vendita dei prodotti agricoli e nell'assistenza tecnica ai soci, nella commercializzazione dei prodotti per l'agricoltura (ad esempio prodotti per il lavoro nei campi), il giardinaggio e la vita in campagna. La mission dell'azienda è la seguente:

*Sostenere le imprese agricole, operando anche a difesa del territorio e delle sue produzioni per creare un rapporto sempre più stretto con la società e con i consumatori.*

La Cooperativa nasce agli inizi del '900, quando un frate cappuccino di nome Padre Pancrazio, figlio di contadini degli orti intorno a Firenze, riuniva gli agricoltori per parlare loro della cooperazione, della necessità di unirsi per rendere più vantaggioso, oltretutto più umano, il faticoso lavoro dei campi. Un gruppo di coraggiosi fece proprie quelle idee e nel 1903 fondò il primo nucleo di quella che è diventata la Cooperativa Agricola di Legnaia. Tanti, in breve tempo, li seguirono fino a far divenire la Cooperativa il principale punto di riferimento dei produttori del contado fiorentino. Oggi la Cooperativa Agricola di Legnaia è una realtà affermata con oltre 500 soci produttori, in gran parte dell'area fiorentina, ma anche di altre zone della Toscana e dell'Italia che coltivano oltre 4100 ettari di terreno, diventando così un centro all'avanguardia nella produzione ortofloro-vivaistica e ortofrutticola all'interno del centro Agroalimentare di Sollicciano e nel mercato di Novoli grazie a sette punti vendita.

Nel 1903 nasce "l'Unione Professionale Cattolica di Legnaia" fondata da padre Pancrazio Landini che aveva come obiettivo quello di portare mutua assistenza tra i contadini, gli operai della Unione Professionale e all'interno della comunità.

Nel 1907 nasce la "vera e propria" Cooperativa Agricola di Legnaia che può essere vista come la naturale prosecuzione dell'esperienza dell'Unione Cattolica fondata nel 1903. *Lo scopo dell'associazione era quello di acquistare e vendere ai propri soci zolfo, solfato, attrezzi rurali, macchine, scorte vive e morte, ed in genere tutto ciò che fosse necessario per l'esercizio dell'agricoltura. Inoltre*

*la Cooperativa intendeva promuovere e favorire l'istruzione civile ed agraria dei propri soci e patrocinare a ogni modo i loro interessi agricoli.*

Il 20 settembre del 1912 la Cooperativa per la prima volta partecipò ad un importante congresso a Roma in cui intervenne il presidente. Sempre in quest'anno vengono terminati i lavori della nuova sede sociale che fu successivamente inaugurata nel mese di giugno con una cerimonia significativa ma modesta, a carattere familiare.

Nel 1939 viene modificato lo statuto della Cooperativa con quello emesso dell'Ente Nazionale Fascista. *Con questo atto vengono cambiati gli scopi della società che ora si propone di gestire spacci per la distribuzione, anche al pubblico, di generi alimentari, articoli di abbigliamento e di uso casalingo. Di organizzare o assumere in appalto o in concessione servizi nell'interesse di tutti o in parte dei soci. Di raccogliere, trasformare e vendere prodotti dei soci o ritenuti utili a quest'ultimi e di provvedere a quant'altro ritenuto opportuno per il miglioramento economico degli associati.* Il campo di interessi della Cooperativa si era dunque allargato, perché le esigenze della famiglia contadina erano mutate e con queste la tipologia dei prodotti da vendere, facendo entrare nella Cooperativa anche coloro che non erano contadini ma artigiani e professionisti.

Il 1954 fu un anno di piccoli e grandi cambiamenti per la Cooperativa. Fu comprato il primo autoveicolo, vennero acquistate delle nuove bilance e altri macchinari per la vendita dei prodotti agricoli e per la gestione amministrativa come frigoriferi e macchine da scrivere.

Dal 1958 al 1963 il tasso medio annuo di crescita raggiunse il 6,3% e la produzione industriale nello stesso periodo risultò più che raddoppiata. Questa crescita fu dovuta al piano Marshall che permise ingenti investimenti da parte della Cooperativa per lo sviluppo e ammodernamento dei macchinari utilizzati per il lavoro nei campi.

A partire dagli anni '70 la Cooperativa iniziò l'ampliamento dei suoi punti vendita, comprando o affittando i fondi adiacenti ai propri magazzini ogni volta che si presentava l'occasione. Negli 1973 lo storico negozio fu trasformato in un negozio "self-service", nel quale i prodotti erano posti sugli scaffali e i clienti potevano rifornirsi da soli; questo cambiamento mise ulteriormente in evidenza come la Cooperativa, durante tutta la sua vita, sia stata sempre propensa a sperimentare soluzioni innovative e all'avanguardia.

Nel 1991 nasce il progetto di ampliamento del Centro di Sollicciano, con lo scopo di incrementare le attività e di crearne delle nuove. A questo fine fu acquistato un appezzamento di terreno contiguo alla prima area e fu predisposto



un progetto edilizio. Nel Novembre del 1993 la Regione Toscana approvò il progetto e lo finanziò in parte e nel 1994 il nuovo Centro per il Condizionamento Ortofrutticolo fu terminato.

Nel 1999 fu acquistato il fondo di Villamagna e fu presentato il nuovo progetto del nuovo centro commerciale dato che il precedente risaliva al 1991 e non era più adeguato alle attuali esigenze aziendali. Con un sostanzioso investimento l'area complessiva del centro arrivò a superare i 70.000 metri quadrati di cui circa 3.500 destinati all'attività commerciale e ai servizi, 3.000 per le serre di produzione, 3.400 per le aree pedonali, 3.500 per il centro di condizionamento e per i prodotti ortofrutticoli e 10.000 per il parcheggio. Il 27 settembre 2003 fu inaugurato il nuovo centro di Sollicciano definendo così il nuovo punto di partenza per le attività commerciali e un rinvigorismento per indirizzo strategico dell'azienda. Per maggiori informazioni consultare [Chiti, 03].

Questo *excursus* sull'evoluzione della Cooperativa fa capire la sua grande complessità dovuta al fatto che si trova ad operare in attività molto diverse tra loro; per fare un esempio basta pensare alla semplice vendita dei prodotti per il giardinaggio e "la vita della campagna": si spazia dalle sedie e tavoli da giardino fino ad arrivare a trattori e mezzi agricoli. Un'offerta così ampia rende la Cooperativa Agricola di Legnaia un punto di riferimento importante all'interno del territorio fiorentino e toscano in generale, generando però anche un'elevata complessità per quanto riguarda la gestione dei core business fondamentali per la sopravvivenza dell'azienda.

## 1.2 Contenuto della tesi

L'elaborato descrive le fasi di realizzazione di un datawarehouse:

- il Capitolo 2 descrive le fasi in cui si articola la progettazione del datawarehouse, vengono descritte le fasi che compongono le attività di progettazione della base di dati per il supporto alle decisioni;
- nel Capitolo 3 vengono descritti i soggetti che interagiscono con l'azienda, lo scopo è quello di descrivere i soggetti e termini utilizzati durante lo svolgimento della tesi;
- nel Capitolo 4 sono descritti i tre processi aziendali che sono monitorati tramite la base di dati per il supporto alle decisioni;

- il Capitolo 5 descrive la progettazione del datawarehouse, una volta definiti i requisiti e le specifiche si effettua la progettazione concettuale, si parte dalla progettazione concettuale iniziale, si passa alla progettazione concettuale utilizzando i dati operazionali e in fine si arriva alla progettazione concettuale finale che rappresenta quali conoscenze si possono estrarre dai dati;
- il Capitolo 6 descrive la rappresentazione logica della base di dati per il supporto alle decisioni realizzata;
- il Capitolo 7 descrive l'ambiente di sviluppo in cui è stato inserito il datawarehouse una volta implementato;
- il Capitolo 8 descrive le procedure di ETL *Extract, Transform and Load*, vengono descritte le principali problematiche affrontate e come sono state risolte;
- il Capitolo 9 descrive come accedere ai dati e visualizzarli utilizzando la struttura multidimensionale o relazionale implementate.

## Capitolo 2

# IL PROCESSO DI PROGETTAZIONE

La progettazione del datawarehouse inizia approfondendo le figure professionali coinvolte nei processi; questo permette di ottenere le informazioni ed i requisiti di analisi necessari per comprendere a fondo l'azienda e il suo funzionamento.

### 2.1 Le figure professionali coinvolte

Durante la fase di raccolta e studio dei requisiti si è dovuto interagire con molteplici figure professionali tra le quali esperti della base di dati relazionale di partenza, manager aziendali, conoscitori dei processi da modellare ed in ultimo esperti di modellazione di datawarehouse e data mining. Il team di lavoro è composto da due esperti della base di dati e del gestionale e un esperto di data mining e datawarehouse che hanno seguito lo sviluppo del progetto sin dalle prime fasi.

Per comprendere a fondo come i processi funzionino all'interno della Cooperativa e come interagiscano tra di loro si è reso necessario intervistare uno dei manager aziendali per comprendere i punti critici, quelli di forza e gli aspetti più importanti che l'azienda intendeva mettere in rilievo. I processi che vengono ritenuti strategicamente più importanti e che si vogliono osservare e studiare sono tre: le vendite, l'inventario di magazzino e i pagamenti e riscossioni di debiti e crediti. Successivamente ogni processo verrà dettagliatamente descritto così da comprendere a fondo la sua struttura e il suo funzionamento.

## 2.2 Modellazione dei requisiti

Una volta terminata la fase preliminare di studio e la comprensione dei processi aziendali si è passati alla modellazione dei requisiti precedentemente individuati, così da realizzare la progettazione del datawarehouse. Per effettuare questa modellazione si è deciso di utilizzare la metodologia studiata durante il corso di laurea e descritta dettagliatamente in [Albano, 09]. La progettazione concettuale avviene gradualmente in 5 fasi:

1. Analisi dei requisiti;
2. Progettazione concettuale iniziale dei datamart;
3. Progettazione concettuale dei datamart dai dati operazionali;
4. Progettazione concettuale finale dei datamart;
5. Progettazione logica dei datamart e del datawarehouse.

### 2.2.1 Analisi dei requisiti

In questa fase le interazioni all'interno del team di lavoro e con i manager aziendali sono state molteplici e importantissime, sia per comprendere la natura e i fini di quest'ultima, ma anche per conoscere a fondo le esigenze di analisi più adatte alla Cooperativa. Sono stati accuratamente analizzati i processi aziendali lavorando a stretto contatto con i responsabili di settore, comprendendo le esigenze e le priorità di quest'ultimi. Si è studiata la base di dati operazionale cercando di comprendere inizialmente la natura dei dati e, successivamente, se quest'ultimi fossero compatibili con le richieste di analisi fatte dai manager. In questa fase preliminare si è rivelato fondamentale l'interazione con gli esperti della base di dati e del gestionale per chiarire dubbi e perplessità analizzando l'elevata quantità di dati a disposizione. Una volta compresi tutti questi aspetti sono stati formalizzati e realizzati i requisiti di analisi.

### 2.2.2 Progettazione concettuale iniziale dei datamart

In questa fase si è passati alla definizione degli schemi concettuali iniziali dei vari datamart: l'obiettivo era quello di creare degli schemi che spiegassero quello che si voleva, senza alcuna ambizione di completezza, ma al solo scopo di ottenere un punto di partenza per descrivere in maniera formale quello che si sarebbe realizzato.

### **2.2.3 Progettazione concettuale dei datamart dai dati operazionali**

Si sono analizzati i dati operazionali per definire possibili candidati degli schemi concettuali dei datamart che rappresentano che cosa si possa ottenere in più dai dati operazionali. L'obiettivo finale era quello di estrarre ulteriori utili informazioni da modellare così da arricchire gli schemi concettuali iniziali. Per realizzare questa progettazione si sono classificate le varie entità della base di dati ed eliminate quelle non ritenute significative. Le entità significative sono state a loro volta suddivise secondo la procedura descritta in seguito e alla fine sono stati proposti gli schemi concettuali dei datamart ottenuti.

### **2.2.4 Progettazione concettuale finale dei datamart**

Si è effettuato un confronto tra gli schemi concettuali iniziali e quelli ottenuti dai dati operazionali, con l'obiettivo di realizzare gli schemi concettuali finali dei datamart fondendo i differenti modelli ottenuti nelle precedenti fasi per ottenendo così modelli il più completi possibili. Questi schemi rappresentano cosa si può analizzare.

### **2.2.5 Progettazione logica dei datamart e del datawarehouse**

L'ultima fase consisteva nella creazione dello schema relazionale del datawarehouse. Si parla di schema relazionale perché è stato deciso di implementare il modello multidimensionale con un sistema Relational On-Line Analytical Processing (ROLAP): questo significa utilizzare un sistema relazionale esteso con funzionalità finalizzate a supportare efficientemente applicazioni OLAP (On-Line Analytical Processing).

Questa progettazione logica si realizza trasformando i vari schemi concettuali finali in uno schema relazionale, decidendo se costruire uno schema a stella o a fiocco di neve (nel caso specifico sono stati realizzati schemi a stella), e poi fondendo i vari schemi dei datamart in uno unico, ottenendo così il modello logico del datawarehouse.

## Capitolo 3

# GLOSSARIO

Si presentano di seguito i soggetti che interagiscono con l'azienda e gli impianti utilizzati nei processi aziendali studiati.

### **Articoli**

Gli articoli comprendono l'insieme di tutti i prodotti, le materie prime, le merci, vendute o acquistate dalla società durante tutto il processo produttivo. Possono dunque essere non soltanto fonte di ricavo, ma anche fonte di costo. Ogni articolo è caratterizzato da un codice che lo identifica univocamente. I prodotti con proprietà simili vengono raggruppati in una categoria merceologica e un prodotto deve appartenere ad una e una sola categoria merceologica.

### **Fornitori**

I fornitori sono tutti i soggetti che interagiscono con l'azienda per l'approvvigionamento delle materie prime, semilavorati o prodotti finiti e pronti per la commercializzazione. I fornitori possono essere divisi in 2 distinte categorie:

- coltivatori diretti o piccole aziende;
- aziende di medie e grandi dimensioni.

La prima categoria è composta da agricoltori diretti che producono in proprio i prodotti e li rivendono alla Cooperativa e da piccole o medie imprese. La seconda categoria è composta da medie e grandi aziende che riforniscono la

Cooperativa di grandi quantità di prodotti alimentari e anche di prodotti particolari che possono essere commercializzati solo da aziende di grandi dimensioni, come ad esempio costosi macchinari agricoli e pezzi di ricambio.

Di ogni fornitore vengono richieste numerose informazioni che risultano fondamentali per poter realizzare gli ordini, ad esempio viene riportato l'eventuale quantità minima ordinabile, il numero di ABI, di CAB e di conto corrente se la modalità di pagamento concordata tra fornitore e cooperativa è l'accredito su conto corrente. Vengono memorizzate anche informazioni su eventuali sconti applicabili o maggiorazioni dovute al trasporto o ad eventuali caratteristiche degli articoli acquistati.

Solitamente il pagamento dei prodotti acquistati non avviene al momento della consegna della merce, ma viene posticipato di 30, 60 o 90 giorni a seconda del potere contrattuale e dell'importanza del fornitore. Ogni fornitore a sua volta può essere anche un cliente registrato della Cooperativa.

### **Clienti registrati**

La registrazione di un cliente avviene tramite il rilascio di una tessera socio personale e ogni individuo può essere intestatario solamente di una tessera. I clienti registrati sono sia persone fisiche che figure giuridiche (esempio imprese) che acquistano uno o più articoli dalla Cooperativa Agricola di Legnaia. In futuro ogni volta che si parlerà di clienti si intenderà sempre e solamente clienti registrati. È doveroso fare questa precisazione in quanto tutte le vendite memorizzate nel database operativo sono già filtrate eliminando tutte le vendite effettuate ai clienti non registrati.

Questa selezione comporta una notevole riduzione dei dati analizzabili ma, d'altra parte, elimina molto rumore che avrebbe reso la comprensione dei risultati ottenuti estremamente complicata.

Ultima importante informazione da sapere sui clienti riguarda la loro classificazione. I clienti vengono divisi in due distinte categorie: *clienti finali* e *clienti con partita IVA*. I primi sono persone fisiche che al momento dell'acquisto pagano utilizzando contanti, bancomat o carta di credito; mentre i secondi sono aziende a cui viene rilasciata una fattura e solitamente pagano con un ritardo che può variare, dai 30 ai 90 giorni a seconda delle quantità acquistate e del potere contrattuale.

## **Agenti**

Sono dipendenti della Cooperativa che ricoprono il ruolo di intermediari tra i clienti e la Cooperativa durante la vendita di uno o più articoli. Solitamente gli agenti sono utilizzati quando vengono venduti prodotti che hanno un valore e un prezzo elevato come ad esempio macchinari agricoli. Il loro compito è quello di seguire il cliente durante la vendita fornendo le informazioni necessarie per chiarire eventuali dubbi. Un agente può essere anche un cliente della Cooperativa.

## **Magazzini**

Nel corso degli anni, come descritto in precedenza, è nata l'esigenza di creare diversi magazzini dislocati nel territorio così da soddisfare le esigenze aziendali dovute alla complessa catena di produzione ed alle elevate differenze tra i prodotti venduti nei numerosi punti vendita dislocati nel territorio. Ogni magazzino è identificato da un codice univoco di tre lettere eccetto il magazzino centrale che è identificato da un codice di 4 lettere: ad esempio il magazzino di Sollicciano è identificato dalla sigla *SOL*, mentre il magazzino della sede centrale è identificato dalla sigla *SEDE*. Alcuni articoli si possono trovare in più magazzini, mentre altri, a causa della difficoltà di trasporto, sono presenti in un unico magazzino. Per raggruppare i magazzini in fase di analisi si possono utilizzare anche il nome della zona in cui sorgono gli edifici (la zona corrisponde al nome della città) e il rispettivo cap.

## **Aree funzionali**

Sono aree che hanno lo scopo di raggruppare i clienti ed i fornitori dell'azienda. La definizione delle varie aree cambia se riferite ai clienti o ai fornitori. Le aree relative ai clienti sono identificate dai vari magazzini in cui vengono realizzate le vendite; mentre le aree relative ai fornitori sono definite dai manager con lo scopo di suddividere e raggruppare i fornitori tramite parametri ben precisi e ritenuti strategicamente importanti.



## Capitolo 4

# DESCRIZIONE DEI PROCESSI AZIENDALI

Dal colloquio iniziale è apparso evidente quali processi l'azienda giudica più importanti, questi processi sono tre: l'analisi delle vendite, analisi dell'inventario dei magazzini e l'analisi di natura finanziaria sui debiti e crediti dell'azienda.

### 4.1 Processo di vendita

Le vendite prese in esame sono quelle rivolte solamente a clienti che, come già detto in precedenza, possono essere divisi in due distinte categorie clienti finali o clienti con partita iva.

Ogni vendita, oltre che articolarsi in numerose fasi, può svolgersi in maniera diversa a seconda del tipo di cliente e del prodotto interessato. Nel caso di vendita effettuata ad un cliente finale lo svolgimento è molto semplice: il cliente sceglie i prodotti che vuole acquistare e, una volta arrivato alla cassa, effettua il pagamento in contanti, carta di credito o bancomat. Nel caso di vendita effettuata ad un cliente con partita iva lo svolgimento può essere più articolato: si può avere uno processo identico al precedente, con in aggiunta l'emissione di fattura, oppure totalmente differente. In questo secondo caso l'acquirente compra elevate quantità di prodotto e, conseguentemente, non è possibile effettuare il pagamento al momento dell'acquisto. In questa circostanza la vendita viene comunque completata e viene emessa una fattura con scadenza di pagamento a 30, 60 o 90 giorni a seconda del cliente.

Queste vendite, e in generale tutti i movimenti di magazzino, vengono classificati in diverse tipologie a seconda della loro natura. Tutti i movimenti, inizialmente, vengono divisi in due grandi gruppi: movimenti in entrata e movimenti in uscita dai magazzini. Successivamente ogni gruppo viene ulteriormente suddiviso a seconda della tipologia. I movimenti in uscita sono suddivisi in circa 30 tipologie, mentre le entrate sono suddivise in circa 15 tipologie. Questa dettagliata classificazione nasce dall'esigenza di identificare i vari movimenti in base al tipo di processo che lo ha generato.

Per riuscire a tenere traccia delle differenti tipologie di vendita effettuate viene utilizzato un codice identificativo di 3 lettere; alcuni esempi possono essere utili per chiarire la sua struttura: *V01* rappresenta una vendita su ordine, *OMA* rappresenta merce in omaggio e *RES* è una sostituzione.

All'interno del processo di vendita occorre prestare particolare attenzione ai resi che possono essere positivi e negativi. Questi resi si verificano nel caso di prodotto non conforme agli standard minimi richiesti e vengono trattati o come un nuovo acquisto di merce (in questo caso si tratta di resi positivi e sono effettuati dai clienti nei confronti della Cooperativa), o come una nuova vendita (in questo caso si tratta di resi negativi e sono effettuati dalla Cooperativa nei confronti dei fornitori). Questo non comporta la modifica delle relative quantità vendute o acquistate e delle eventuali fatture emesse. Per tenere traccia della differente natura di queste particolari vendite o acquisti sono utilizzati due particolari codici: *V51* (rettifica positiva) e *V52* (rettifica negativa).

Durante lo studio preliminare del processo di vendita il manager ha espresso la volontà di poter analizzare, per ogni articolo venduto, tre parametri che sono ritenuti fondamentali: le quantità vendute ad ogni singolo cliente, il valore delle quantità vendute e l'incasso effettuato dalla vendita. Cerchiamo di capire meglio il significato di queste grandezze.

**Quantità vendute** - Rappresenta la quantità di ogni singolo prodotto acquistato da un singolo cliente. Un esempio potrà chiarire meglio il suo significato. Se un cliente compra, in un solo acquisto, 10 volte lo stesso prodotto la cassa non emetterà uno scontrino con 10 battute ma effettuerà una sola battuta che riporterà come valore della quantità 10. Prendendo in esame ogni singola battuta emessa dalle casse risulta importante sapere le quantità acquistate di ogni prodotto per non commettere errori durante la successiva analisi.

**Valore delle quantità vendute** - Rappresenta il valore che l'azienda assegna al prodotto che sta vendendo. Questo valore viene calcolato moltiplicando la quantità venduta per il costo medio unitario del prodotto (CMP). Questo CMP

viene calcolato dividendo il valore delle quantità presenti in magazzino (valore ottenuto sommando il valore delle quantità acquistate durante l'anno e il valore delle esistenze iniziali) per le quantità presenti nei magazzini (valore ottenuto sommando le quantità acquistate e le quantità iniziali).

**Costo del venduto** - Rappresenta quanto è costato all'azienda i prodotti che vengono venduti, viene calcolato moltiplicando la quantità venduta per il prezzo unitario del prodotto; viene ritenuto importante osservare questo valore dato che permette di capire se il prezzo di vendita del prodotto è sufficiente a coprire il valore degli articoli acquistati.

Le misure fondamentali da monitorare per studiare l'andamento delle vendite sono:

- le quantità vendute;
- il valore delle quantità vendute in €;
- il costo del venduto in €.

## 4.2 Inventario di magazzino

Con il tempo l'elevata quantità di prodotti movimentati e la presenza di numerosi magazzini dislocati sul territorio, nati con lo scopo di rifornire i punti vendita, hanno fatto nascere l'esigenza di studiare l'inventario di magazzino così da monitorare i capitali immobilizzati e l'ammontare delle scorte.

Anche se l'azienda possiede numerosi magazzini con caratteristiche differenti ai fini dell'analisi verranno considerati tutti allo stesso modo.

Le quantità presenti all'interno di ogni magazzino non sono costanti nel corso del tempo poiché giornalmente molti prodotti sono movimentati: spostati da un magazzino ad un altro, trasportati ai punti vendita, acquistati dai fornitori o semplicemente venduti. Concettualmente, per lo studio dell'inventario di magazzino, tutti questi movimenti sono da considerarsi allo stesso livello e con la stessa importanza. È possibile effettuare questa semplificazione perché tutte le movimentazioni precedentemente descritte avvengono con l'emissione di documenti molto simili tra loro. Durante lo spostamento dei prodotti vengono monitorate sempre le quantità (sia acquistate che vendute) al netto dei resi. Le quantità acquistate sono quelle in ingresso nei magazzini (causa acquisto o movimentazione), mentre quelle vendute sono quelle in uscita dai magazzini (causa vendita o movimentazione).

Al fine dell'analisi i manager non ritengono necessario studiare i movimenti con scadenza giornaliera ma si ritiene sufficiente uno studio con scadenza mensile. Per comprendere a fondo le movimentazioni si è deciso di riportare un esempio di come un articolo venga movimentato, in un singolo magazzino, nei mesi di febbraio e di marzo Tabella 4.1.

<i>Data</i>	<i>Prodotto</i>	<i>Codice operazione</i>	<i>Quantità in uscita</i>	<i>Valore venduto</i>	<i>Quantità in ingresso</i>	<i>Valore acquistato</i>	<i>Fornitore Cliente</i>	<i>Categoria merceologica</i>
01/02/2009	034869	Vendita	10	20€			1254606	F0456
07/02/2009		Movimentazione in uscita	19	38€			1252781	
17/02/2009		Acquisto			100	200€	000001	
23/02/2009		Vendita	24	48€			1457800	
29/02/2009		Rettifica Negativa	0					
29/02/2009		Rettifica Positiva			0			
<b>TOTALE Febbraio</b>			<b>53</b>	<b>106€</b>	<b>100</b>	<b>200€</b>		
04/03/2009	034869	Vendita	2	4€			1252888	F0456
24/03/2009		Vendita	20	40€			1254678	
25/03/2009		Vendita	12	24€			1234354	
25/03/2009		Movimentazione in ingresso			50	100€	000001	
29/02/2009		Rettifica Negativa	10	20€				
29/02/2009		Rettifica Positiva			0			
<b>TOTALE Marzo</b>			<b>44</b>	<b>88€</b>	<b>50</b>	<b>100€</b>		

**Tabella 4.1:** Ipotesico movimento di un articolo in un magazzino.

La tabella evidenzia la gestione di un magazzino per un prodotto (*il prodotto 034869*): ogni mese viene redatto un documento che riassume i movimenti di un articolo in un mese. Nel caso specifico descritto in Tabella 4.1 si può notare come siano riportati tutti i movimenti in ingresso, con relativi valori, e tutti i movimenti in uscita, affiancati sempre dai relativi valori. Oltre a queste informazioni sono riportate le quantità di rettifica con i valori.

Oltre a queste informazioni i responsabili intervistati hanno espresso il desiderio di poter osservare le rimanenze mensili di ogni magazzino e le eventuali rettifiche. Monitorare le rimanenze renderebbe più facile decidere le quantità da ordinare in futuro o più semplicemente risulterebbe d'aiuto nel decidere una differente dislocazione dei prodotti all'interno dei magazzini. Le rettifiche (che possono essere sia positive che negative), rappresentano le quantità in eccesso o in difetto, degli articoli, individuate durante lo svolgimento delle normali attività lavorative. Queste modifiche delle quantità immagazzinate possono essere generate da differenti fenomeni: un errato conteggio durante l'inventario, un furto o da un'errata battitura durante una vendita. Una volta individuate le

quantità da correggere vengono realizzate delle vendite o acquisti ad hoc per eliminare la discrepanza tra rappresentazione digitale delle quantità immagazzinate e situazione reale. Questi movimenti vengono identificati utilizzando due particolari codici: rettifiche positive codice *I21*, rettifiche negative codice *I22*.

Per poter schematizzare e rappresentare la movimentazione dei prodotti nei vari magazzini si è deciso di adottare il *modello ad istantanea periodica*; questo modello analizza, in maniera periodica (ad esempio giornalmente o mensilmente), le quantità di prodotti in magazzino. Per le nostre esigenze, come già detto in precedenza, un'osservazione mensile sarà più che adeguata. Tramite questa analisi si potranno creare modelli per analizzare le quantità acquistate e il loro valore, le quantità vendute e i rispettivi valori e le rimanenze di un articolo in magazzino.

Si riporta la descrizione dei valori richiesti dai responsabili per studiare l'inventario dei prodotti.

**Quantità acquistate** - Valore calcolato sommando in ogni magazzino e per ogni prodotto, le quantità acquistate o semplicemente trasferite nel deposito in esame.

**Valore dell'acquistato** - Valore calcolato facendo il prodotto tra le quantità acquistate per il costo medio unitario (CMP).

**Quantità vendute** - Valore calcolato sommando in ogni magazzino e per ogni prodotto, le quantità vendute o semplicemente trasferite all'esterno del deposito esaminato.

**Valore del venduto** - Valore calcolato facendo il prodotto tra le quantità vendute per il costo medio unitario (CMP).

**Rimanenze** - Valore calcolato effettuando la differenza tra le quantità acquistate e quelle vendute.

**Valore delle rimanenze** - Valore calcolato facendo il prodotto tra le quantità residue di ogni mese per il costo medio unitario (CMP).

**Rettifiche positive / negative** - Valore calcolato sommando rispettivamente le quantità in eccesso o in difetto dei prodotti nei vari magazzini.

Le misure fondamentali da monitorare per studiare l'inventario di magazzino sono:

- quantità vendute;
- valore del venduto;
- quantità acquistata;

- valore dell'acquistato;
- rettifiche positive / negative;
- rimanenze;
- valore delle rimanenze.

### 4.3 Analisi finanziarie

L'ultima analisi richiesta dal management della Cooperativa serve per studiare i debiti e crediti dell'azienda.

La prima cosa da comprendere consiste nella natura dei dati da studiare. Per monitorare debiti e crediti occorre osservare il flusso di cassa. Questo flusso non è altro che l'insieme di tutti i movimenti in entrata e in uscita di denaro dalla cassa della Cooperativa. Queste entrate sono composte dai pagamenti effettuati dai clienti finali e da quelli con partita iva, mentre le uscite sono composte dai vari pagamenti effettuati dall'azienda verso i fornitori o i pagamenti sostenuti durante le normali attività di gestione. Per poter operare con un adeguato dettaglio sul flusso di cassa si è deciso di prendere in esame ogni singola entrata o uscita di cassa.

Una volta compresa la natura del flusso si può prestare attenzione ai debiti e crediti dell'azienda. Questi debiti e crediti nascono in due differenti momenti:

- quando un cliente non paga al momento dell'acquisto, ma a 30 - 60 o 90 giorni;
- quando l'azienda compra dai fornitori e a sua volta paga a 30 - 60 o 90 giorni.

In entrambi i casi si verifica l'emissione di una fattura che potrà contenere differenti voci (chiamate anche scadenze) che generano debiti o crediti. Ogni singola voce riporta i seguenti valori:

**Importo della voce** - Consiste nell'importo totale che può essere pagato o incassato dalla Cooperativa.

**Importo alla scadenza** - Consiste nell'importo da pagare o incassare al momento della data di scadenza. A volte l'importo da movimentare viene suddiviso in più pagamenti (30, 60 o 90 giorni), in questo caso la voce riporta l'ammontare da pagare o incassare alla prima scadenza.

**Importo versato alla data di scadenza** - Questo valore è differente da zero quando il pagamento è rateizzato e rappresenta quanto dell'importo da movimentare alla scadenza sia stato effettivamente trasferito.

**Importo residuo da versare alla data di scadenza** - Consiste nell'importo della voce ancora da pagare, questo valore è differente da zero nel momento in cui l'importo non sia stato trasferito interamente con le precedenti movimentazioni.

**Importo residuo già versato** - Questo valore non è altro che la sommatoria dei precedenti versamenti o pagamenti effettuati più eventuali acconti, questo valore sarà differente da zero quando l'importo da trasferire è saldato in più soluzioni o vengono lasciati acconti.

Ogni voce, oltre a contenere i valori appena elencati può essere chiusa o aperta: si definisce chiusa la voce pagata o riscossa entro la data di scadenza, mentre si definisce aperta se non pagata o riscossa entro tale data.

Tutte le scadenze sono associate a un soggetto destinatario così da risalire al cliente o fornitore che li ha generati.

Le misure fondamentali da monitorare per studiare le analisi finanziarie sono:

- Importo delle voci;
- Importo alla scadenza;
- Importo versato alla data di scadenza;
- Importo residuo da versare alla data di scadenza;
- Importo residuo già versato.

Qualche esempio della gestione dei pagamenti può aiutare il lettore a comprendere meglio il comportamento delle scadenze e il mutare dei valori inseriti nelle differenti voci. Di seguito sono descritti tre esempi di come vengono gestite le scadenze: il primo esempio riguarda un pagamento in contanti, il secondo rappresenta un pagamento a rata unica a 30 giorni con preventivo acconto mentre l'ultimo è un pagamento a 60 giorni con due rate, la prima a 30 e la seconda a 60 giorni.

1. Se viene effettuato un pagamento in contanti il totale dell'importo è inserito nel campo *importo della voce* e tutte le altre voci sono a zero, questo è il caso più semplice.

2. Nel caso di un pagamento dilazionato (a 30 giorni) la situazione si complica leggente: ipotizziamo di dovere incassare un totale di 1000€, questo valore sarà inserito nel campo *importo della voce*, ipotizzando sempre che il cliente abbia lasciato un acconto pari a 300€ questo valore viene inserito nel campo *importo residuo già versato*. Ora avendo un pagamento a ratea unica (a 30 giorni), il campo *importo alla scadenza* conterrà il valore di 700€ cioè quanto ancora si deve incassare. Raggiunti i 30 giorni a pagamento avvenuto il campo *importo versato alla data di scadenza* diventerà conterrà i 700€ l'importo alla scadenza diventerà zero e *importo residuo già versato* conterrà 1000€.
3. L'ultimo esempio riguarda il caso di pagamento a 60 giorni in due rate: una prima rata a 30 giorni e una a 60 giorni. Inizialmente *l'importo della voce* conterrà il totale da pagare, per esempio, 1000€ e le altre voci saranno zero eccetto *importo alla scadenza* che sarà di 400€, cioè pari all'importo da incassare a 30 giorni. Arrivati alla prima scadenza *l'importo versato alla data di scadenza* diventa 400€, *l'importo alla scadenza* diventa 600€ (cioè pari all'importo da pagare alla successiva scadenza), *l'importo residuo da versare alla data di scadenza* diventa 600€ (1000€ - 400€) e *l'importo residuo già versato* passa a 400€. Quando l'intero importo sarà pagato a 60 giorni *l'importo residuo già versato* diventerà 1000€ e tutti gli altri campi diventeranno zero eccetto *l'importo della voce*.

Per quanto riguarda la lettura dei dati giornalmente le informazioni delle scadenze vengono aggiornate nel database gestionale e il 1° di ogni mese tutte le modifiche vengono selezionate e inserite nel datawarehouse.



## Capitolo 5

# PROGETTAZIONE DEI DATA MART E DEL DATAWAREHOUSE

Dopo aver analizzato i processi aziendali che la Cooperativa Agricola di Legnaia giudica più importanti e fondamentali, si può passare ad analizzare i requisiti per capire come schematizzare i processi e alla verifica dei dati operazionali. Nelle successive tabelle: Tabella 5.1, Tabella 5.2 e Tabella 5.3 vengono riportate le schematizzazioni realizzate.

### 5.1 I requisiti

N.	Requisito di analisi
1	Totale delle quantità vendute per cliente, magazzino ed agente che ha effettuato la vendita.
2	Totale del valore delle vendite per cliente, magazzino ed agente che ha effettuato la vendita.
3	Totale del costo del venduto per ogni vendita effettuata.

**Tabella 5.1:** Requisiti di analisi del processo di vendita.

N.	Requisito di analisi
4	Totale delle quantità acquistate e del loro valore suddividendole per fornitore, categoria merceologica del prodotto e magazzino.
5	Totale delle quantità vendute e del loro valore suddividendole per fornitore, categoria merceologica del prodotto e magazzino.
6	Totale delle Rettifiche positive e negative e del loro valore, suddivise per magazzino e prodotto.
7	Totale delle Rimanenze annue dei prodotti e del loro valore, suddivise per magazzino e prodotto.

**Tabella 5.2:** Requisiti di analisi del processo di Inventario di magazzino.

N.	Requisito di analisi
8	Totale degli importo dovuto delle scadenze dell'importo della scadenza dell'importo della scadenza versato dell'importo residuo e dell'importo residuo versato, per la data di emissione per la data di scadenza, per destinatario e per area funzionale.

**Tabella 5.3:** Requisiti di analisi delle analisi finanziarie.

## 5.2 La base di dati

La base di dati iniziale è composta da circa 400 tabelle, molte delle quali inutili per le analisi che si vogliono realizzare. Questa base di dati è gestita e implementata tramite un gestionale: ForeOffice, di cui alcuni dettagli sono illustrati nel capitolo 7.2.

In Figura 5.1 viene riportata solo una parte dello schema della base di dati, perché risulta molto difficile da rappresentare nella sua interezza e renderebbe complicata l'identificazione dei dati effettivamente utili e necessari. L'ultimo dettaglio da considerare riguarda i nomi di alcune tabelle e attributi che sono stati modificati per favorire la comprensione del database.

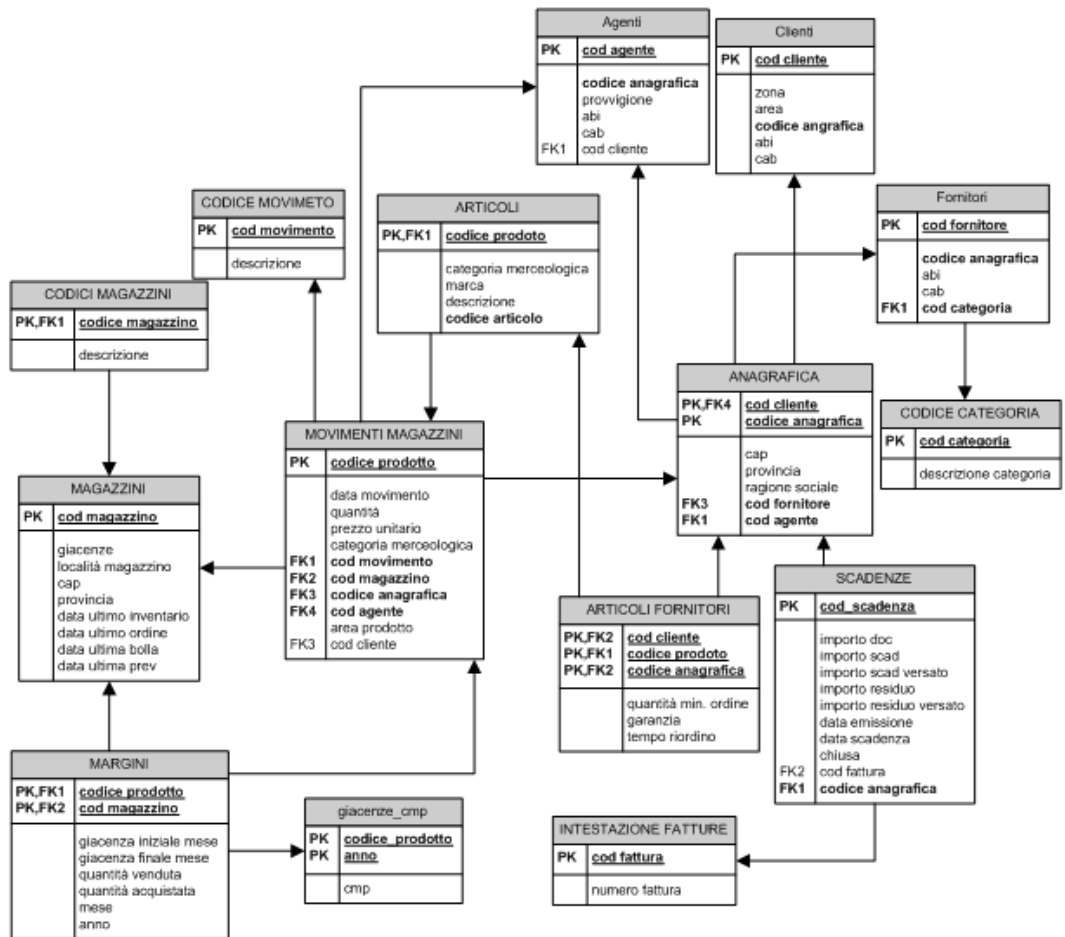


Figura 5.1: Schema relazionale base di dati operativo.

### 5.3 Le specifiche

Dall'analisi dei requisiti si ottengono le seguenti specifiche iniziali. Vengono riportate in Tabella 5.4 le granularità dei fatti.

Fatto	Descrizione
Vendite	Osservazione giornaliera delle singole voci di una vendita.
Inventario di magazzino	Osservazione mensile del singolo prodotto presente in ogni magazzino.
Scadenze	Osservazione giornaliera della singola scadenza.

**Tabella 5.4:** Granularità dei fatti.

Nelle successive tabelle Tabella 5.5, Tabella 5.6 e Tabella 5.7, per ogni fatto, vengono riportati e descritti i requisiti di analisi, le dimensioni (tra parentesi gli attributi) e le misure ritenute interessanti.

Vendita			
N	Requisito di analisi	Dimensioni	Misure
1	Somma della quantità venduta, per cliente (codice cliente), per data (anno, mese, giorno), per agente (codice agente), per magazzino (codice magazzino), per prodotto (codice prodotto).	Cliente ( ragione sociale, zona, cap), Data (giorno, mese, anno), Agente (cap, zona), Magazzino (zona, cap, provincia), Prodotto (descrizione, categoria merceologica, tipo articolo).	Quantità.
2	Somma del costo del venduto e del valore del venduto, per cliente (codice cliente), per data (anno, mese, giorno), per agente (codice agente), per magazzino (codice magazzino), per prodotto (codice prodotto).	Cliente ( ragione sociale, zona, cap), Data (giorno, mese, anno), Agente (cap, zona), Magazzino (zona, cap, provincia), Prodotto (descrizione, categoria merceologica, tipo articolo).	Costo, Valore del Venduto.

**Tabella 5.5:** Descrizione fatto vendite.

N	Requisito di analisi	Dimensioni	Misure
3	Somma delle quantità acquistate e delle quantità vendute, per fornitore (codice fornitore), per data (anno, mese), per magazzino (codice magazzino), per prodotto (codice prodotto).	Fornitore (tipo, ragione sociale), Data (mese, anno), Magazzino (zona, cap, provincia), Prodotto (descrizione, categoria merceologica, tipo articolo).	Quantità Acquistata, Quantità Venduta.
4	Somma del valore dell'acquistato del valore del venduto e del valore delle rimanenze, per fornitore (codice fornitore), per data (anno, mese), per magazzino (codice magazzino), per prodotto (codice prodotto).	Fornitore (tipo, ragione sociale), Data (mese, anno), Magazzino (codice magazzino), Prodotto (descrizione, categoria merceologica, tipo articolo).	Valore Acquistato, Valore Venduto, Valore Rimanenze.
5	Somma delle quantità rimanenti, per magazzino (codice magazzino), per data (anno, mese).	Magazzino (zona, cap, provincia), Data (mese, anno), Prodotto (descrizione, categoria merceologica, tipo articolo).	Quantità Rimanenti.
6	Somma delle quantità delle rettifiche positive e quantità delle rettifiche negative, per magazzino (codice magazzino), per data (anno, mese).	Magazzino (zona, cap, provincia), Data (mese, anno), Prodotto (descrizione, categoria merceologica, tipo articolo).	Quantità Rettifiche Positive, Quantità Rettifiche Negative.
7	Somma del valore delle rettifiche positive e valore delle rettifiche negative, per magazzino (codice magazzino), per data (anno, mese).	Magazzino (zona, cap, provincia), Data (mese, anno), Prodotto (descrizione, categoria merceologica, tipo articolo).	Valore Rettifiche Positive, Valore Rettifiche Negative.

**Tabella 5.6:** Descrizione fatto inventario di magazzino.

Scadenze			
N	Requisito di analisi	Dimensioni	Misure
8	Somma dell'importo dovuto alla scadenza dell'importo alla scadenza dell'importo versato alla scadenza dell'importo residuo e dell'importo residuo versato, per data emissione (anno, mese, giorno), per data scadenza (anno, mese, giorno), per destinatario (codice destinatario), per area funzionale (codice area funzionale), per numero fattura (codice fattura), per chiusura scadenza (codice chiusura).	Data Emissione (giorno, mese, anno), Data scadenza (giorno, mese, anno), Destinatario (cap, zona), Area Funzionale (area), Numero fattura (numero della fattura), Chiusa.	Importo Scadenza dovuto, Importo Scadenza, Importo Scad. Versato, Importo Residuo, Importo Residuo Versato.

**Tabella 5.7:** Descrizione fatto scadenze.

## 5.4 Le tipologie di trattamento

Le possibili tipologie di trattamento delle modifiche ai dati sono 4:

- Tipo 1: si cambiano i valori degli attributi dell'entità della dimensione che ha subito modifiche. E' la soluzione più semplice e immediata, ma si alterano le analisi storiche;
- Tipo 2: si aggiunge una nuova riga alla tabella dimensionale, creando di fatto una entità nuova. Tutti i fatti precedenti alla modifica fanno riferimento alla vecchia entità, mentre tutti i fatti successivi alla modifica fanno riferimento a quella nuova. In questo modo il caricamento dei dati si complica e aumentano i dati della dimensione;
- Tipo 3: si prevede la possibilità di un cambiamento di attributo sostituendo l'attributo *Attr* con 3 attributi *Attr*, *NuovoAttr*, *DataModifica*;
- Tipo 4: per le dimensioni che cambiano molto di frequente si possono prevedere due tabelle dimensionali, una contenente gli attributi che rimangono immutati e una contenente gli attributi che variano. Per maggiori chiarimenti consultare [Albano, 09].

Di seguito in Tabella 5.8 si elenca la lista delle dimensioni ricavate evidenziandone la granularità e la metodologia da utilizzare per il trattamento delle modifiche. Nel nostro caso, la maggior parte delle dimensioni non ha bisogno di trattamento delle modifiche in quanto non variano nel tempo. Le uniche dimensioni che possono cambiare sono *Clienti*, *Agenti* e *Fornitori*. Bisogna però

precisare che non tutti gli attributi di queste tre dimensioni cambiano, a variare sono solo alcuni valori come, ad esempio, la *residenza* di agenti o clienti o la *ragione sociale* del fornitore. In questo caso nel momento in cui la residenza cambia questa modifica potrebbe generare un cambiamento di zona o cap. In questo caso l'aggiornamento delle informazioni contenute nel datawarehouse avviene utilizzando la seconda tipologia di trattamento delle modifiche: viene creato un nuovo record contenente sia i nuovi dati aggiornati che quelli non modificati.

Lo stesso procedimento è applicato quando un fornitore cambia ragione sociale, anche in questo si inserisce un nuovo record con i dati aggiornati.

Successivamente in forma tabellare vengono descritte le dimensioni appartenenti ad ogni fatto, specificando gli attributi che le compongono e le relative gerarchie.

Nome	Descrizione	Granularità	Parzialità	Trattamento Modifiche
Cliente	Contiene le informazioni sul singolo cliente che effettua l'acquisto.	Il cliente	SI	Tipo 2.
Data	Rappresenta la data delle misure dei fatti.	Giorno	NO	Non necessario.
Agente	Rappresenta l'agente che effettua la vendita.	L'agente	SI	Tipo 2.
Magazzino	Contiene il codice del magazzino in cui il prodotto è immagazzinato.	Il magazzino	NO	Non necessario.
Prodotto	Contiene informazioni sul prodotto che viene analizzato.	L'articolo	NO	Non necessario.
Fornitore	Rappresenta il fornitore che vende il prodotto all'azienda.	Il fornitore	NO	Tipo 2.
Data	Rappresenta la data che identifica il verificarsi di un fatto.	Il giorno	NO	Non necessario.
Data emissione	Rappresenta la data di quando è stata emessa la fattura.	Il giorno	NO	Non necessario.
Data scadenza	Rappresenta la data di quando scade la fattura.	Il giorno	NO	Non necessario.
Scadenze chiusa	Identifica se la scadenza è chiusa oppure no.	Si, No	NO	Non necessario.
Area funzionale	Identifica l'area funzionale a cui si riferisce una fattura.	L'area funzionale	NO	Non necessario.
Numero fattura	Identifica un insieme di una o più scadenze.	La fattura	NO	Non necessario.
Destinatario scadenze	Identifica il destinatario di una o più scadenze, solitamente il destinatario è un cliente.	Il destinatario	NO	Non necessario.

**Tabella 5.8:** Elenco dimensioni.



Per semplicità le descrizioni degli attributi elementari sono state tralasciate.

Cliente (Vendite)

Attributo	Descrizione
Ragione sociale	...
Zona	Rappresenta il comune o provincia
cap	...

Data (Vendite)

Attributo	Descrizione
Giorno	...
Mese	...
Anno	...

Agente (Vendite)

Attributo	Descrizione
Agente	Contiene il codice dell'agente.
Zona	...
cap	...

Magazzino (Vendite)

Attributo	Descrizione
Magazzino	Contiene il codice del magazzino.
Zona	...
cap	...
Provincia	...

Prodotto (Vendite)

Attributo	Descrizione
Prodotto	Contiene il codice del magazzino.
Descrizione	...
Cat. merceologica	...
Tipo articolo	...

## Magazzino (Inventario)

Attributo	Descrizione
Magazzino	Contiene il codice del magazzino.
Zona	...
Cap	...
Provincia	...

## Fornitore (Inventario)

Attributo	Descrizione
Fornitore	Contiene il codice del fornitore.
Ragione sociale	Contiene la ragione sociale del fornitore

## Data (Inventario)

Attributo	Descrizione
Mese	...
Anno	...

## Prodotto (Inventario)

Attributo	Descrizione
Prodotto	Contiene il codice del magazzino.
Descrizione	...
Cat. merceologica	...
Tipo articolo	...

## Data emissione (Scadenze)

Attributo	Descrizione
Giorno	...
Mese	...
Anno	...

## Data scadenza (Scadenze)

Attributo	Descrizione
Giorno	...
Mese	...
Anno	...

Area funzionale (Scadenze)

Attributo	Descrizione
Codice area funzionale	Codice dell'area funzionale a cui si riferisce la fattura.

Scadenza chiusa (Scadenze)

Attributo	Descrizione
Si	Scadenza chiusa.
No	Scadenza aperta.

Numero fattura (Scadenze)

Attributo	Descrizione
Numero fattura	Numero della fattura che contiene una o più scadenze.

Destinatario (Scadenze)

Attributo	Descrizione
Cap	Cap del destinatario della scadenza.
Zona	Zona di residenza del destinatario della scadenza.

Data(Vendite)

Descrizione gerarchia	Tipo gerarchia
Giorno → Mese → Anno	Bilanciata

Data(Inventario)

Descrizione gerarchia	Tipo gerarchia
Mese → Anno	Bilanciata

Data emissione (Scadenza)

Descrizione gerarchia	Tipo gerarchia
Giorno → Mese → Anno	Bilanciata

Data scadenza (Scadenze)

Descrizione gerarchia	Tipo gerarchia
Giorno → Mese → Anno	Bilanciata

Nella Tabella 5.9 viene visualizzato un riepilogo delle misure individuate, specificando il fatto di appartenenza, il nome, la descrizione e informazioni aggiuntive come aggregabilità e se derivata.

Misure				
Fatto	Misura	Descrizione	Aggregabilità	Derivata
Vendite	Quantità		Additiva	NO
Vendite	Costo	Quantità acquistata * CMP	Additiva	NO
Vendite	Valore del venduto in €	Quantità acquistata * Prezzo di vendita	Additiva	NO
Inventario	Quantità acquistata		Additiva	NO
Inventario	Quantità venduta		Additiva	NO
Inventario	Valore acquistato in €	Quantità acquistata * CMP	Additiva	NO
Inventario	Valore venduto in €	Quantità venduta * CMP	Additiva	NO
Inventario	Valore rimanenze in €	Rimanenze * CMP	Additiva	NO
Inventario	Rimanenze	Quantità acquistata – Quantità venduta	Semi additiva	SI
Inventario	Rettifiche positiva	Quantità di rettifica positiva	Semi additiva	NO
Inventario	Valore rettifiche positive in €	Quantità di rettifica positiva * CMP		
Inventario	Rettifiche negative	Quantità di rettifica negativa	Semi additiva	NO
Inventario	Valore rettifiche negative in €	Quantità di rettifica negativa * CMP		
Scadenze	Importo scadenza dovuto		Additiva	NO
Scadenze	Importo scadenza		Additiva	NO
Scadenze	Importo scad. versato		Additiva	NO
Scadenze	Importo residuo	Importo scadenza dovuto – Importo scadenza versato	Additiva	SI
Scadenze	Importo residuo versato		Additiva	NO

**Tabella 5.9:** Riepilogo misure.

Prima di passare alle successive fasi della progettazione, le dimensioni e le misure dei fatti vengono rappresentate nella seguente forma tabellare che mette in evidenza quali misure e dimensioni sono in comune a fatti diversi e quindi vanno uniformate o rinominate, Tabella 5.10 e Tabella 5.11.

Dimensione	Dimensioni dei fatti		
	Vendite	Rotazione	Analisi Finanziaria
Cliente	X		
Data	X	X	
Agente	X		
Magazzino	X	X	
Prodotto	X	X	
Fornitore		X	
Area funzionale			X
Scadenza chiusa			X
Data emissione			X
Data scadenza			X
Destinatario scadenza			X
Numero fattura			X

**Tabella 5.10:** Dimensione dei fatti.

Misura	Misure dei fatti		
	Vendite	Inventario	Analisi Finanziaria
Quantità	X		
Costo	X		
Valore del venduto in €	X		
Quantità acquistata		X	
Quantità venduta		X	
Valore acquistato		X	
Valore venduto		X	
Valore Rimanenze		X	
Rimanenze		X	
Rettifiche positive		X	
Valore rettifiche positive		X	
Rettifiche negative		X	
Valore rettifiche negative		X	
Importo Scadenza dovuto in €			X
Importo scadenza in €			X
Importo scad. versato in €			X
Importo residuo in €			X
Importo residuo versato in €			X

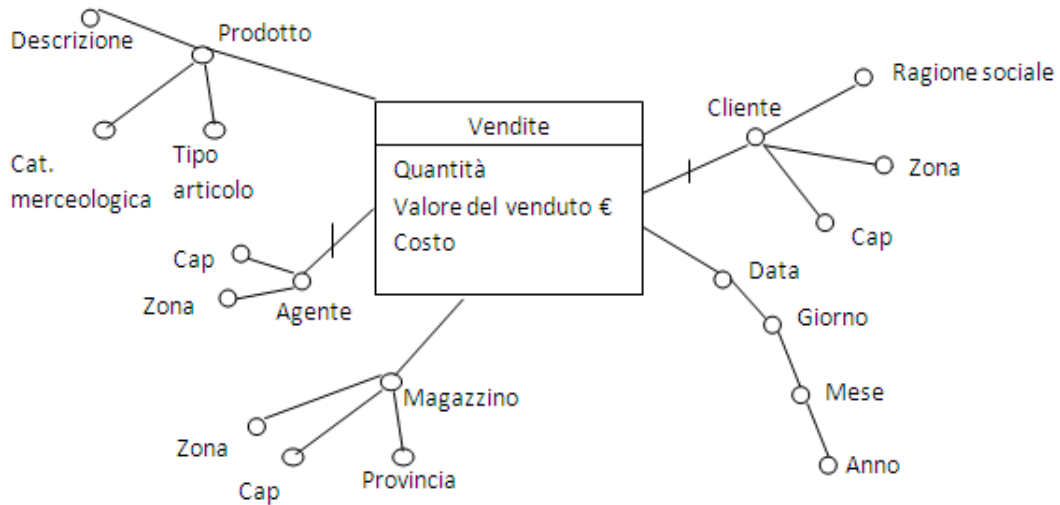
**Tabella 5.11:** Misure dei fatti.

## 5.5 Progettazione concettuale iniziale dei data mart

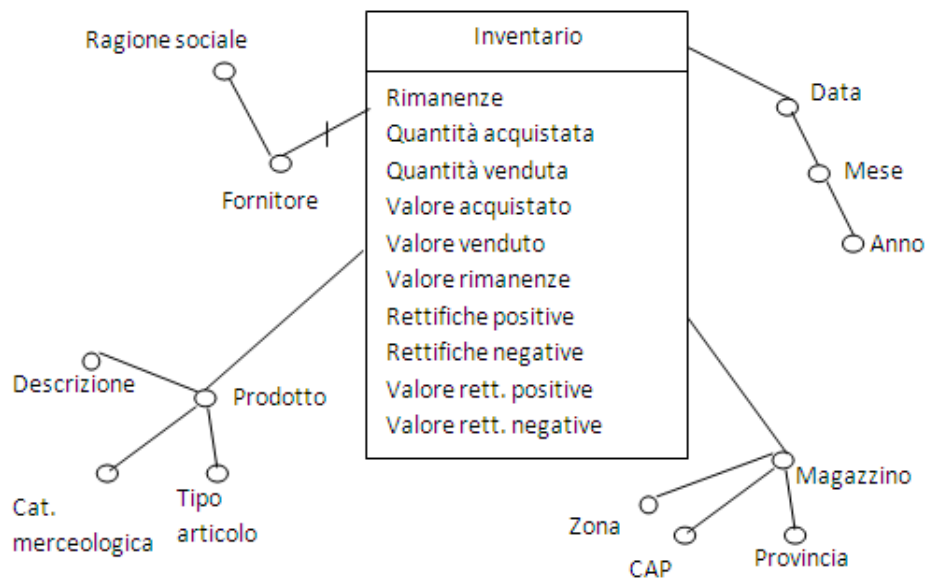
L'ottenere *Vendite*, *Inventario* e *Scadenze* come fatti ha portato alla decisione di creare 3 data mart, uno per ogni fatto. Gli attributi citati nelle analisi dei dati suggeriscono come possibili schemi concettuali iniziali quelli mostrati di seguito Figura 5.2, Figura 5.3 e Figura 5.4.

Gli schemi riportano le misure all'interno della tabella, mentre esternamente sono collocate le relative dimensioni. Questa rappresentazione mette in evidenza importanti dettagli già descritti in precedenza:

- *Gerarchie bilanciate*: si verificano quando i possibili livelli sono in numero predefinito e i valori degli attributi che ne fanno parte sono sempre definiti. Per esempio, gli attributi *Giorno*, *Mese* e *Anno* della *Data* fanno parte di una gerarchia bilanciata con tre livelli;
- *Attributi o dimensioni opzionali*: si verifica quando il valore di una dimensione o di un attributo può essere opzionale e sono modellati con archi tagliati. Per esempio, nel fatto in Figura 5.2 l'attributo *Agente* e *Cliente* sono attributi dimensionali opzionali;
- *Attributi descrittivi*: Le dimensioni e gli attributi dimensionali rappresentati con archi che terminano con un circoletto possono essere usati nelle operazioni di analisi per esprimere restrizioni sui loro valori e per fare raggruppamenti o aggregazioni. Quando invece si vogliono rappresentare attributi dimensionali, o attributi dei fatti diversi dalle misure, che non vanno usati nelle operazioni di analisi per fare raggruppamenti, essi vengono detti descrittivi e si rappresentano con archi privi del circoletto. Per esempio, nel fatto in Figura 5.4 l'attributo *Numero fattura* si rappresenta senza circoletto dato che non interessa usarlo nelle analisi per fare raggruppamenti [Albano, 09].



**Figura 5.2:** Schema concettuale iniziale del data mart Vendite



**Figura 5.3:** Schema concettuale iniziale del data mart Inventario

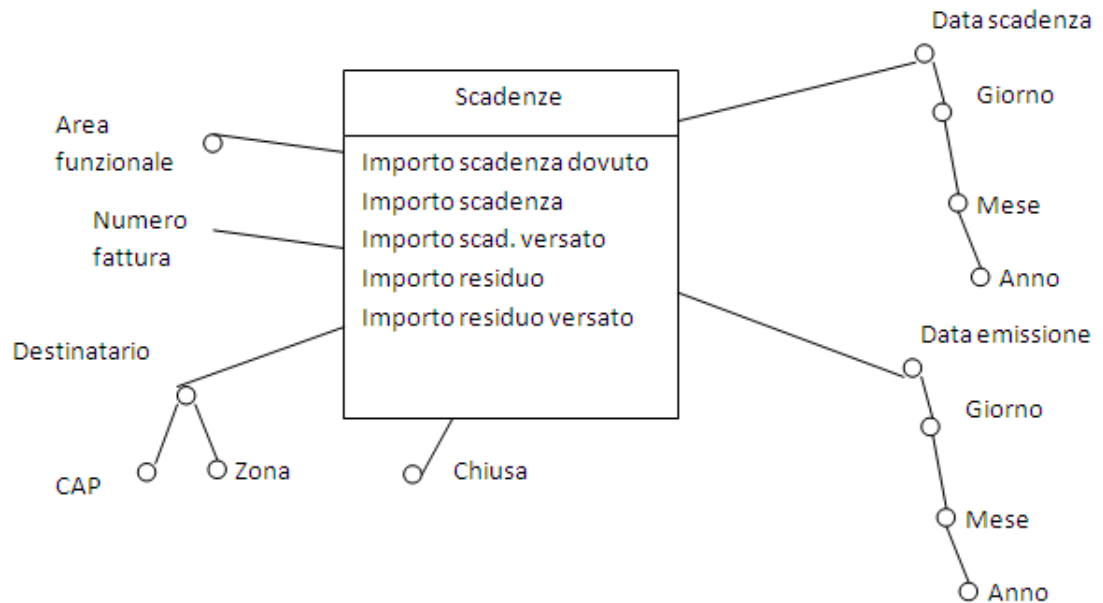


Figura 5.4: Schema concettuale iniziale del data mart Scadenze

## 5.6 Progettazione concettuale dei data mart dai dati operazionali

Si esamina lo schema relazionale per decidere quali tabelle e attributi sono interessanti.

### 5.6.1 Analisi dati operazionali

Le tabelle *Codice movimento*, *Codici magazzini*, *Articoli fornitori* e *Codice categoria* non servono ai fini dell'analisi dei dati, pertanto non compariranno nel datawarehouse. Per quanto riguarda invece le tabelle utilizzate, valgono le seguenti considerazioni sugli attributi relativi:

- *Magazzini*: vengono memorizzate informazioni sulla località del magazzino, nello specifico il *cap* e *provincia*;
- *Articoli*: vengono selezionati solamente la *categoria merceologica*, la *descrizione* e il *tipo articolo* degli articoli che sono venduti e acquistati dall'azienda;



- *Anagrafica*: vengono selezionati il *cap*, *provincia* e *ragione sociale* degli individui registrati; questi dati saranno relativi non solo ai *clienti* ma anche ai *fornitori* e agli *agenti* che saranno distinti rispettivamente da un codice fornitore e un codice agente posto a *NULL* se l'anagrafica si riferisce a un cliente;
- *Movimenti magazzini*: questa è la tabella più importante dato che contiene informazioni su ogni singola transazione effettuata da qualunque magazzino. Di questa tabella vengono selezionati, dei prodotti movimentati, le *quantità* (che possono essere vendute o acquistate), la *data del movimento* (data in notazione americana), il *prezzo unitario* e *l'area prodotto*;
- *Agenti*: viene selezionata solo la *provvigione*;
- *Clienti*: viene selezionato solo la *zona* e *l'area di provenienza*;
- *Fornitore*: viene selezionato il *codice categoria*;
- *Intestazione fatture*: viene selezionato solo il *codice della fattura* così da rendere possibile il collegamento tra le scadenze e le fatture relative;
- *Margini*: vengono selezionati tutti i campi contenuti nella tabella;
- *Giacenze cmp*: viene selezionato solo il *cmp* di ogni prodotto per ogni anno, che moltiplicato per le quantità di prodotti genererà il relativo valore;
- *Scadenze*: vengono selezionati i vari importi: *importo documento* (rappresenta il valore nominale della scadenza), *importo scadenza* (contiene l'importo effettivamente riscosso), *importo scadenza versato* (contiene l'importo versato entro la scadenza), *importo residuo* (contiene l'eventuale importo che si deve ancora versare), *importo residuo versato* (contiene l'importo già versato), la *data di emissione della scadenza*, la *data di riscossione delle scadenze*, il *codice della fattura* a cui fanno riferimento le scadenze e *chiusa* (un campo Sì, No che identifica se la scadenza è chiusa oppure no).

In questa prima analisi sono state volutamente trascurate le chiavi delle tabelle, poiché verranno prese in considerazione in seguito. Una volta scelte le informazioni interessanti, si passa alla fase successiva della progettazione, ovvero la classificazione delle entità.

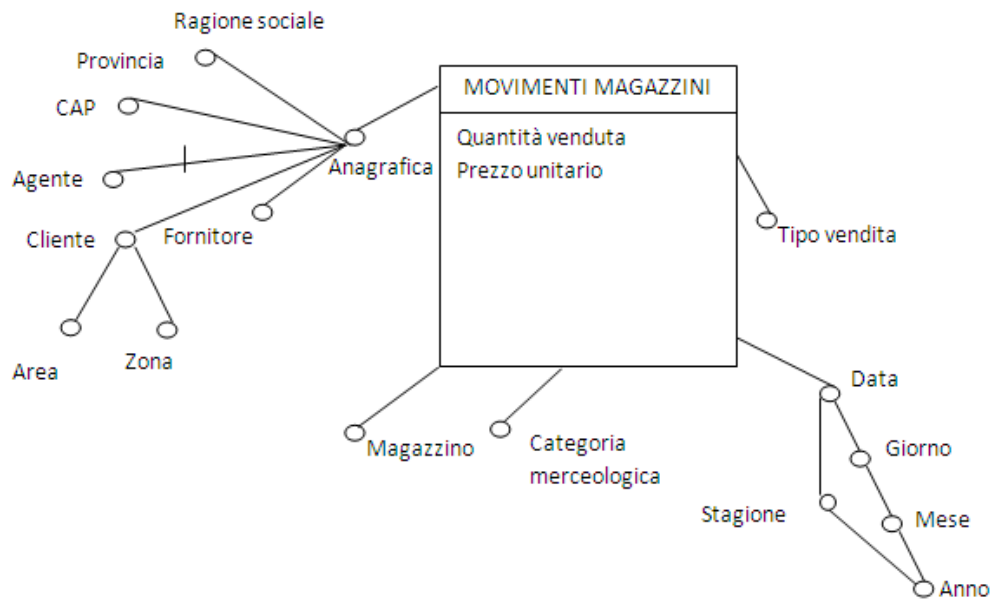
### 5.6.2 Classificazione delle entità

Vengono classificate le tabelle dello schema relazionale in base al loro contenuto e alle loro associazioni.

- *Entità evento*: ricordando la definizione, sono le entità che descrivono eventi che si verificano frequentemente a certe date e contengono attributi numerici che rappresentano possibili misure interessanti ai fini dell'analisi dei dati. Le tabelle che possono essere considerate *Entità evento* nel nostro caso sono: *Movimenti Magazzini*, *Margini* e *Scadenze*.
- *Entità componente*: sono le tabelle in relazione con un'entità evento e con un'associazione 1:N. Analizzando lo schema si nota come le *entità componenti* siano *Magazzini*, *Articoli*, *Anagrafica* e *Intestazione fatture*.
- *Entità di classificazione*: sono le tabelle in relazione con un'entità componente e con una catena di associazioni 1:N. Nel nostro caso queste entità sono costituite dalle tabelle *Agenti*, *Clienti* e *Fornitori*.

### 5.6.3 Definizione degli schemi concettuali candidati dei data mart

Avendo identificato tre entità evento interessanti, si procede con la definizione di tre schemi concettuali per i data mart con le relative dimensioni Figura 5.5, Figura 5.6 e Figura 5.7.



**Figura 5.5:** Schema data mart Movimenti Magazzini



**Figura 5.6:** Schema data mart Articolo Riepilogo Statico

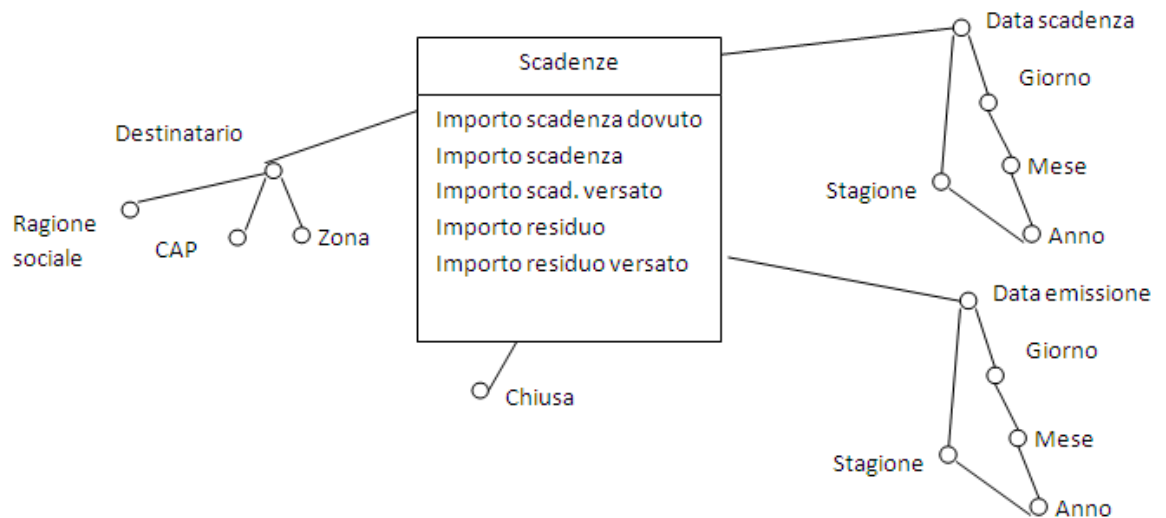


Figura 5.7: Schema data mart Scadenze

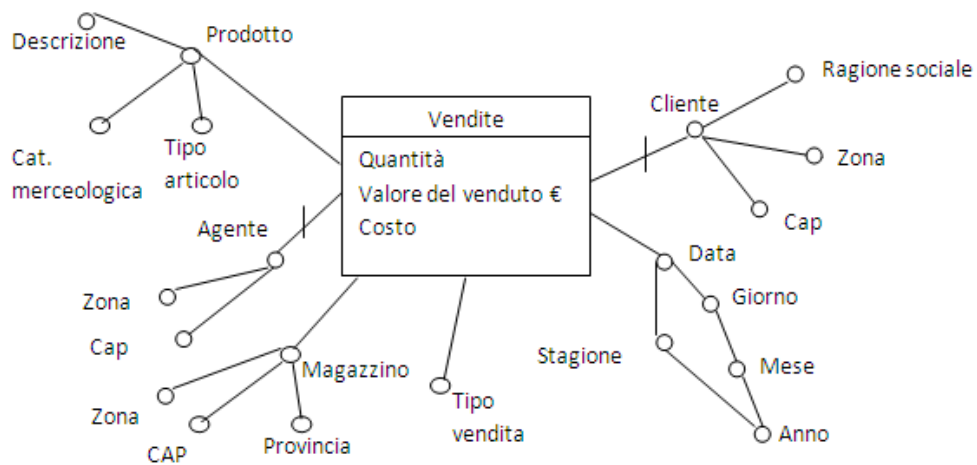
## 5.7 Progettazione concettuale finale dei data mart

Dall'analisi concettuale iniziale e dagli schemi candidati si estraggono i *data mart finali*, che rappresentano che cosa si può analizzare dai dati a disposizione. I data mart da realizzare sono tre e, in prevalenza, ricalcano quelli realizzati durante l'analisi concettuale fatta in precedenza con la modifica di alcune dimensioni.

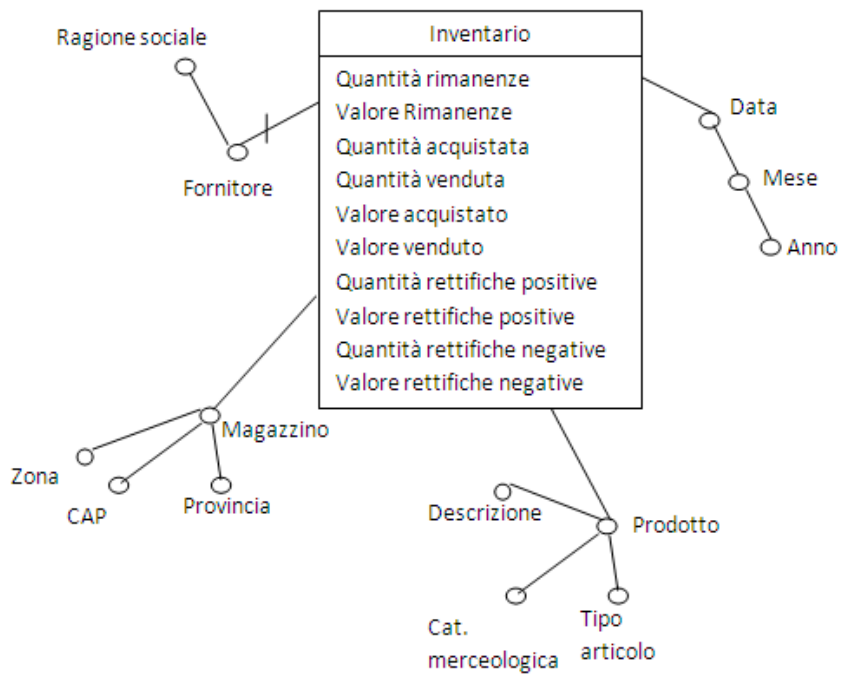
Alla dimensione *Data* è stato aggiunto l'attributo *Stagione*, questo perché molti prodotti vengono commercializzati solo in alcune stagioni e quindi si è ritenuto importante permettere l'analisi delle vendite non solo mensilmente o annualmente, ma anche per stagione.

La dimensione *Agente* è stata modificata inserendogli attributi *Zona* e *cap*, questo perché durante lo studio dei dati operazionali si sono individuati anche queste informazioni e si è deciso di inserirle per rendere possibile l'osservazione delle vendite effettuate dagli agenti in base alla Zona di residenza. Si ricorda che per gli Agenti la zona di residenza equivale all'area di lavoro.

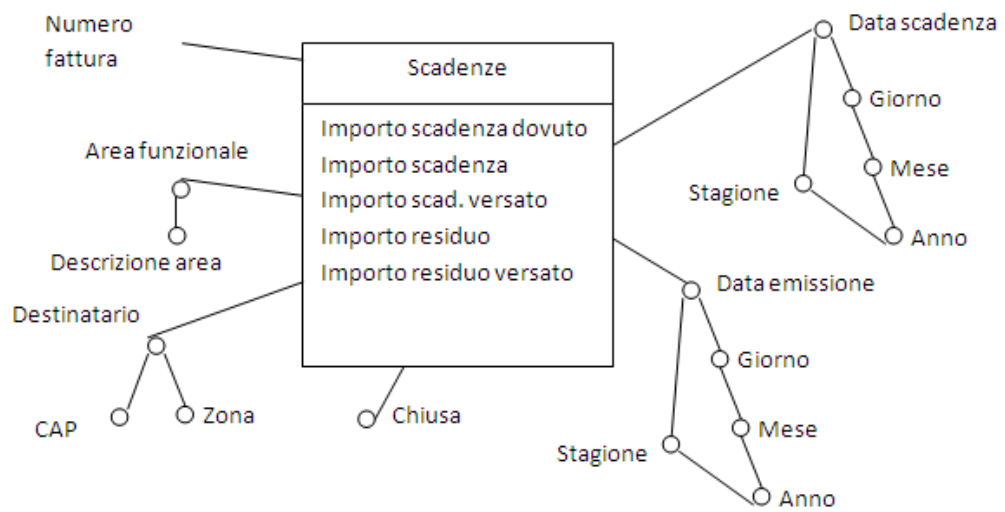
Anche gli altri due fatti *Inventario* e *Scadenze* hanno avuto delle lievi modifiche. Anche qui è stato inserito l'attributo dimensionale *Stagione* all'interno delle dimensioni *Data*, *Data scarico*, *Data carico*, *Data scadenza* e *Data emissione* Figura 5.8, Figura 5.9, Figura 5.10.



**Figura 5.8:** Schema finale data mart Vendite



**Figura 5.9:** Schema finale data mart Inventario



**Figura 5.10:** Schema finale data mart Scadenze

## Capitolo 6

# RAPPRESENTAZIONE LOGICA DEI DATA MART E DEL DATAWAREHOUSE

Una volta terminata la realizzazione concettuale del datawarehouse si può passare alla progettazione logica che consiste nella creazione dello schema relazionale del datawarehouse. La prima fase comprende la trasformazione di ogni schema concettuale finale dei data mart in uno schema relazionale decidendo se creare uno schema a stella o a fiocco di neve (nel caso specifico si è utilizzato lo schema a stella); successivamente i modelli dei data mart saranno integrati così da ottenere un unico schema che rappresenterà il datawarehouse Figura 6.1, Figura 6.2, Figura 6.3 e Figura 6.4.

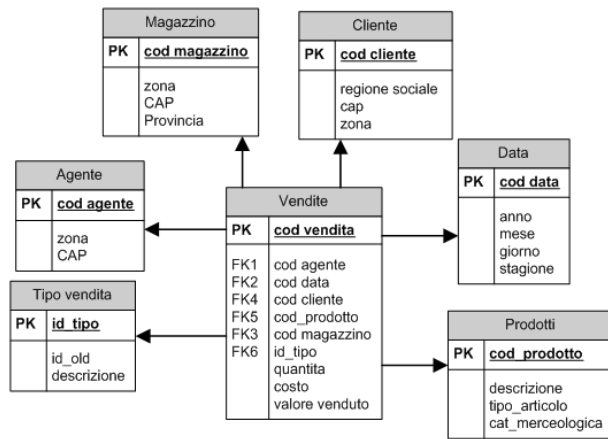


Figura 6.1: Schema relazionale data mart Vendite

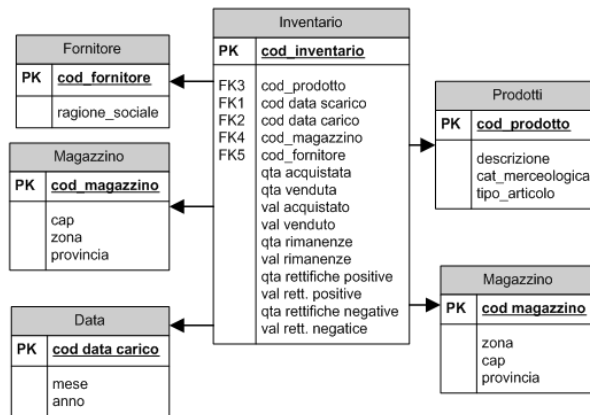
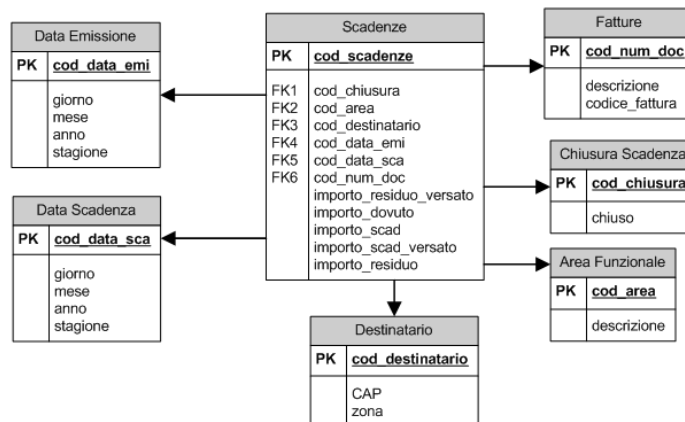


Figura 6.2: Schema relazionale data mart Inventario





**Figura 6.3:** Schema relazionale data mart Scadenze

Prima di mostrare lo schema relazionale finale ottenuto dall'integrazione dei precedenti tre modelli occorre effettuare alcune piccole precisazioni: per ottenere uno schema relazionale omogeneo sono state combinate le tabelle delle dimensioni condivise da due o più fatti, come ad esempio per le varie date. Anche le tabelle *Agente*, *Cliente* e *Venditore* sono state unite dato che le informazioni contenute nelle tre tabelle sono molto simili, in aggiunta è stato inserito un campo (*tipo\_individuo*) che identifica il ruolo di ogni persona: i valori possibili sono solamente tre A per *Agente*, C per *Cliente* e V per *Venditore*.

All'interno della tabella *Agente*, *Cliente* e *Venditore* è stato inserito l'attributo *data\_validita*. Questo attributo serve per poter realizzare le modifiche di Tipo 2 ai *Clienti*, *Agenti* e *Fornitori*. L'attributo ha il compito di riportare la data di validità del record: se un record è ancora valido la data sarà settata a *NULL*, mentre se un record non è più valido l'attributo riporterà il momento in cui il record è stato dichiarato non più valido a causa di modifiche con conseguente creazione del nuovo record.

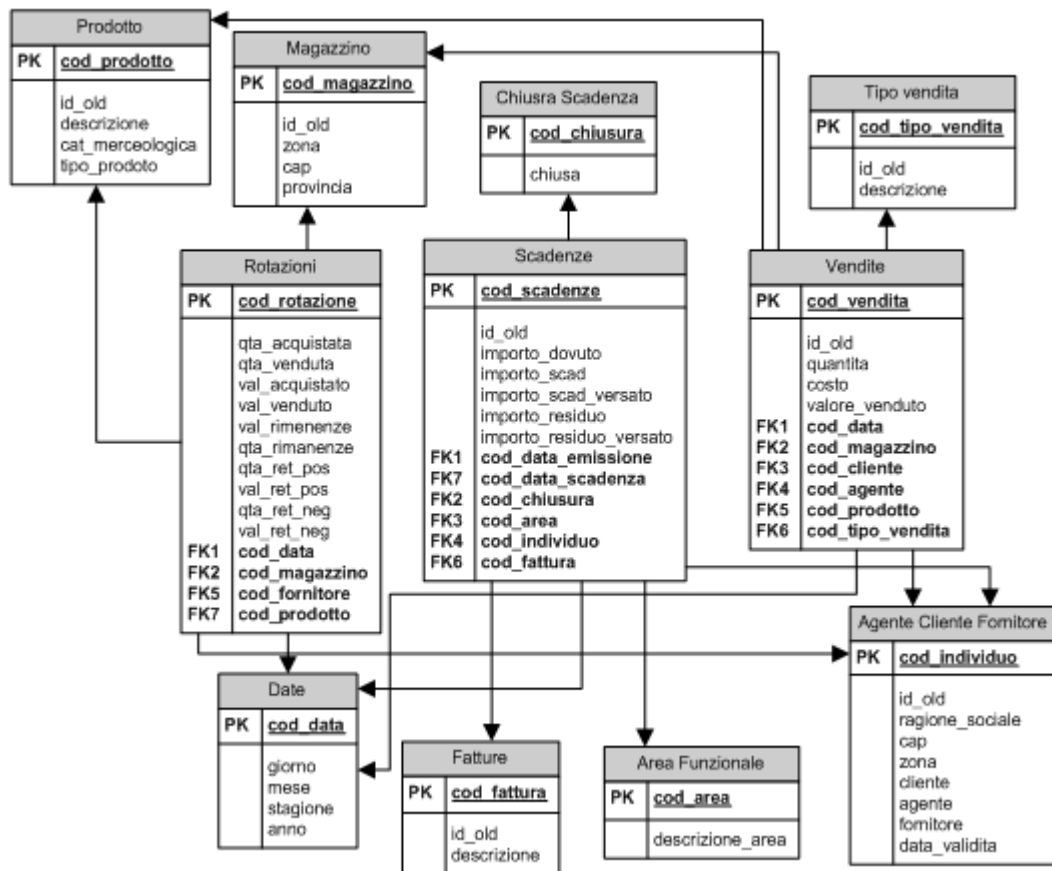


Figura 6.4: Schema relazionale finale datawarehouse

## Capitolo 7

# AMBIENTE DI SVILUPPO

Una volta compresa la struttura della base di dati per il supporto alle decisioni da realizzare si può passare allo studio dell'ambiente di sviluppo in cui si andrà a realizzare il datawarehouse e successivamente inserire.

All'interno dell'azienda viene utilizzato il seguente RDBMS (*Relational DataBase Management System*): Microsoft SQL Server 2005 per la gestione dei database aziendali.

Essendo già presente un valido programma per la gestione dei DataBase si è deciso di utilizzarlo per la gestione della base di dati per il supporto alle decisioni così da ottenere la massima interoperabilità tra i DataBase.

### 7.1 Indici utilizzabili per facilitare le interrogazioni e chiavi primarie

SQL Server 2005 è un prodotto molto completo per la gestione delle basi di dati operazionali, ma purtroppo non offre molte funzionalità orientate alla gestione di una base di dati per il supporto alle decisioni.

Gli indici sono strutture dati associate ad attributi di una tabella e volti a rendere più efficiente l'esecuzione di alcune interrogazioni. Nei sistemi dedicati ai datawarehouse o nei sistemi relazionali estesi per il funzionamento con

i datawarehouse esiste la possibilità di creare indici particolari che non sono di solito disponibili nei normali sistemi relazionali, come ad esempio [Albano, 09]:

- *indici a bitmap*: ad ogni possibile valore  $v$  di un attributo è associato un array di bit grande quanto il numero dei record della tabella. Il bit  $i$ -esimo è uguale a 1 se il record  $i$ -esimo ha proprio  $v$  come valore dell'attributo. E' un tipo di indice adatto all'uso con attributi poco selettivi e risulta essere particolarmente efficiente per trovare riferimenti a record che soddisfano una condizione booleana su più attributi;
- *indici di giunzione*: sullo schema a stella si definisce un indice di giunzione multiattributo fra le tabelle delle dimensioni  $D_1$  e  $D_2$  e la tabella dei fatti  $F$ . Un elemento dell'indice di giunzione contiene nell'ordine i RID (identificativi dei record) dei record delle  $D_i$  e di  $F$  che sono in giunzione;
- *foreign column join index*: utilizzati per agevolare interrogazioni sulla tabella dei fatti  $F$  con una restrizione su un attributo  $A$  della tabella dimensionale  $D$ . Si immagina di estendere  $F$  con gli attributi di  $D$  e di costruire un indice su  $A$  con elementi  $\langle i, r_i \rangle$ , dove  $a_i$  è un valore di  $A$  e  $r_i$  è un RID di  $F$ : si costruisce dunque un indice sull'attributo  $A$  per la tabella dei fatti  $F$ , nonostante  $A$  non sia un attributo di  $F$ .

Sql Server 2005 non offre la possibilità di creare nessuno degli indici appena presentati. Tuttavia si ha la possibilità di definire normali indici a B+albero, che pur non presentando gli stessi vantaggi di indici specializzati, offrono comunque un aumento prestazionale rispetto all'esecuzione di interrogazioni che non facciano uso di nessun indice.

### Uso di chiavi surrogate

Nella costruzione delle tabelle dimensionali di un datawarehouse è spesso conveniente l'utilizzo di chiavi primarie surrogate, ossia con valori numerici generati automaticamente. Essendo possibile definire chiavi autoincrementali tutte le chiavi primarie delle tabelle dimensionali sono chiavi surrogate.

## 7.2 Software gestionale ForeOffice

La Cooperativa Agricola usa un software di supporto all'attività operativa chiamato ForeOffice e sviluppato dalla ditta italiana Gruppo Software. L'appli-

cazione è realizzata per essere eseguita in ambienti Windows e si appoggia al RDBMS di SQL Server 2005.

Offre, come la maggioranza dei software gestionali, un'enorme mole di funzionalità per la gestione completa di tutte le attività dell'azienda. Si segnala, a titolo di esempio, la possibilità di gestire [GS, 09]:

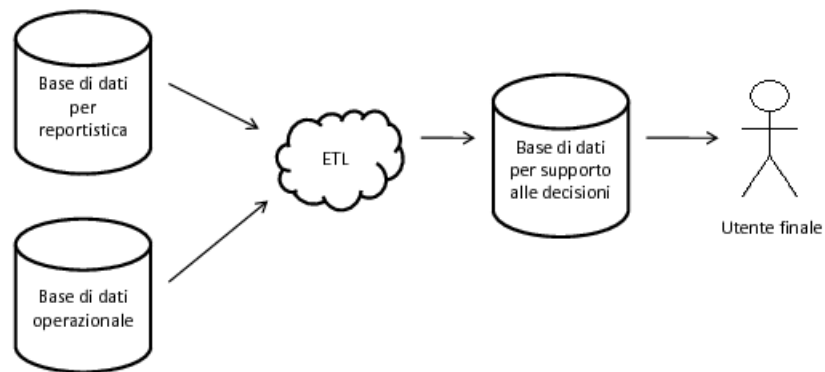
- le anagrafiche dei clienti, dei fornitori e degli agenti di vendita;
- le anagrafiche degli articoli presenti nei vari punti vendita e magazzini;
- la contabilità generale (ciclo attivo e ciclo passivo);
- la gestione del magazzino;
- le attività di produzione con controllo e commesse.

### 7.3 Scelta dell'ambiente di sviluppo

Il requisito fondamentale da rispettare nello sviluppo del datawarehouse aziendale era il contenimento dei costi, attraverso il riutilizzo dei software già a disposizione dell'azienda. Dal momento che l'applicativo gestionale era realizzato utilizzando SQL Server 2005 e, offrendo quest'ultimo valide soluzioni per la gestione delle basi di dati e per l'esplorazione dei dati (utilizzando SQL Server Business Intelligence), si è deciso di sviluppare il datawarehouse utilizzando questo RDBMS.

Prima di parlare dell'ambiente di lavoro è doveroso precisare alcuni dettagli riguardanti la base di operativa. Dopo aver studiato le esigenze dell'azienda e i dati di partenza ci si è resi conto che alcuni dati erano già presenti in un secondo database che viene utilizzato solamente per la reportistica interna.

Per alleggerire la procedura di ETL (*Extract, Transform and Load*), discussa nel Capitolo 8, si è dunque deciso di non ricalcolare alcuni valori ma di estrarli da questo "secondo" database e usarli assieme i dati dalla base di dati operativa per la creazione del datawarehouse, come schematizzato in Figura 7.1.



**Figura 7.1:** Schematizzazione fasi di sviluppo

## Capitolo 8

# PROCEDURE DI ETL

L'obiettivo è quello di descrivere il processo di estrazione, trasformazione e caricamento dei dati, presentando alcune delle problematiche più rilevanti. Per ogni problematica verranno riportate le possibili soluzioni suggerite in letteratura e sarà descritto come è stata affrontata nella realizzazione del datawarehouse per la Cooperativa Agricola di Legnaia.

### 8.1 Il processo di ETL

Il processo di ETL (*Extract-Transform-Load*) è il processo di estrazione, trasformazione e consolidamento di dati provenienti da sorgenti eterogenee e memorizzate in un datawarehouse [Ruggieri, 09].

Un processo di ETL ben realizzato estrae i dati da una o più sorgenti dati, si assicura che le informazioni siano conformi agli standard di qualità e di consistenza definiti durante la progettazione, rende i dati delle diverse sorgenti omogenei tra loro e, in fine, li carica nel datawarehouse.

Una volta comprese le operazioni realizzate in questa fase è immediato intuire come il processo di ETL influenzi il corretto funzionamento di una base di dati per il supporto alle decisioni e, allo stesso tempo, come possa essere una delle fasi più costose, in termini di risorse utilizzate, durante la realizzazione di un datawarehouse.

Dal momento che in un datawarehouse quello che conta sono i dati, bisogna considerare l'importanza della procedura di ETL, specialmente considerando quanto valore aggiunga ai dati, visto che:

- rimuove errori e corregge i dati mancanti;
- crea confidenza con i dati;
- rende omogenei i dati che provengono da sorgenti diverse affinché possano essere usati insieme;
- struttura i dati per essere usati da strumenti di analisi.

Anche durante la creazione del datawarehouse per la Cooperativa Agricola di Legnaia la realizzazione della procedura di ETL ha richiesto un elevato impiego di tempo e di risorse. Questo è dovuto anche al fatto che la procedura è stata realizzata in T-SQL (*Transact-SQL*) [I. Ben-Gan, 08] senza avvalersi di alcun software.

## 8.2 Scelta del software di ETL

La scelta del software per realizzare il processo di ETL è fondamentale, occorre conoscere la base di dati di partenza e il risultato che si vuole ottenere.

La prima valutazione consiste nel decidere se acquistare un programma commerciale oppure realizzare manualmente la procedura per il caricamento.

Ogni scelta avrà dei vantaggi e degli svantaggi. I vantaggi sono:

- riduzione dei tempi e degli investimenti per lo sviluppo;
- presenza di strumenti grafici, i software per ETL offrono pratiche e semplici interfacce progettate per rendere più semplice e visuale lo sviluppo di trasformazioni complesse. Propongono una visione grafica del flusso dati rendendo immediata l'interpretazione del processo di caricamento o di manipolazione dei dati;
- maggiore semplicità in fase di manutenzione, gli aspetti evidenziati al punto precedente consentono chiaramente non solo una più semplice implementazione dei progetti ma anche, e soprattutto, una loro manutenzione molto più rapida e immediata;
- possibile integrazione con strumenti di reporting, alcuni vendor di ETL propongono all'interno del loro catalogo anche strumenti di analisi dei dati e reportistica;



- disponibilità di tool per realizzare la documentazione, gli strumenti di ETL mettono tipicamente a disposizione molte opzioni per commentare le trasformazioni sviluppate a diversi livelli di dettaglio;
- presenza di strumenti per il debug, gli strumenti di ETL forniscono generalmente un ambiente di debug in cui è possibile testare le operazioni eseguite, verificando passo dopo passo le trasformazioni realizzate.

Le caratteristiche fin qui evidenziate permettono di capire come in molte realtà aziendali l'acquisto di un prodotto di questo genere sia velocemente ammortizzabile tramite un aumento di efficienza ed efficacia.

D'altra parte anche lo scrivere manualmente la procedura di ETL ha dei benefici:

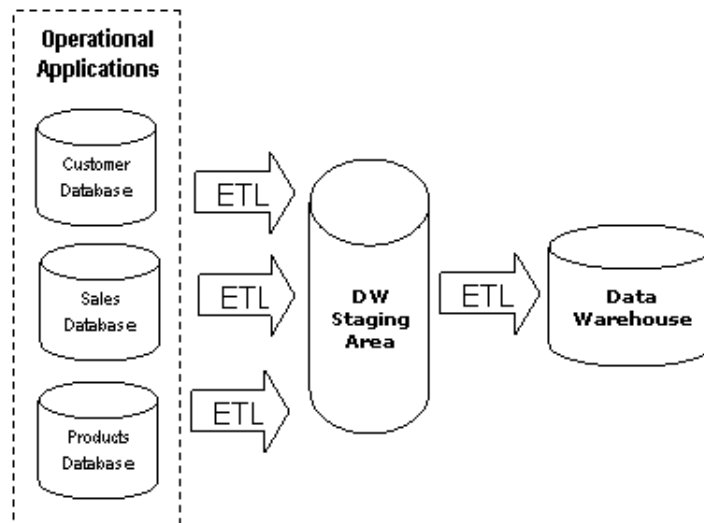
- ottenere la massima flessibilità e ottimizzazione, realizzando una procedura di ETL che si adatta perfettamente ai dati;
- risparmio un termini di costi utilizzando i software a disposizione;
- massima compatibilità nella comunicazione tra i database utilizzati;
- reale comprensione dei dati da trasformare con una conseguente riduzione di errori;
- test forzato dei dati estratti.

Come già accennato in precedenza, in questo caso si è deciso di realizzare manualmente in T-SQL la procedura di caricamento ed elaborazione dei dati. Si è scelta questa soluzione perché, in fase di progettazione, si è ritenuto l'unico metodo per riuscire a comprendere a fondo i dati di partenza e allo stesso tempo ottenere i benefici sopra elencati.

### 8.3 Area di staging

Solitamente le procedure di estrazione, trasformazione e caricamento avvengono in più stadi. I risultati delle elaborazioni intermedie dei dati vengono memorizzati su disco in aree di memoria chiamate staging area Figura 8.1.

La staging area viene di solito creata su una base di dati, ma non di rado si ricorre all'uso di semplici file. E' possibile anche non utilizzare le staging area durante le fasi di ETL, in questo caso i dati vengono lavorati direttamente in memoria.



**Figura 8.1:** Schematizzazione processo di ETL con staging area

Tale scelta è consigliabile in processi di estrazione e trasformazione lineari in quanto lavorando direttamente in memoria si riducono i tempi e si ottimizza la procedura di ETL.

Tuttavia la staging area è ampiamente utilizzata in quasi tutte le realizzazioni di datawarehouse dal momento che offre dei vantaggi aggiuntivi non indifferenti:

- la possibilità, in caso di fallimento della procedura di ETL, di non ricominciare dall'inizio l'intera procedura ma soltanto dal punto in cui si è verificato il fallimento;
- la possibilità di verificare meglio, in fase di controllo del funzionamento della procedura, le diverse fasi del trattamento dei dati.

Nel datawarehouse in esame si è fatto uso della staging area per eseguire trasformazioni sui dati e stabilizzare le procedura di ETL. Dovendo alcune trasformazioni essere realizzate in una sequenza ben precisa, per memorizzare i valori intermedi delle elaborazioni si è utilizzata una staging area realizzata tramite tabelle temporanee. L'utilizzo delle tabelle è stato necessario anche per rendere più stabile il processo di ETL dal momento che il solo utilizzo della memoria con milioni di record rende il processo instabile e soggetto a fallimento.

## 8.4 Processo di trasformazione

Il processo di trasformazione consiste in numerose attività tra cui le seguenti [Ruggieri, 09]:

- codifiche e normalizzazioni dei dati al fine di risolvere i differenti formati;
- splitting o merging di attributi;
- creazione di chiavi surrogate;
- merge-purge cioè fusione di dati provenienti da diverse sorgenti;
- realizzare un controllo sulla qualità dei dati;
- join di dati provenienti da tabelle differenti;
- calcolo di attributi derivati ottenuti partendo da quelli disponibili.

Nella procedura di ETL realizzata sono state molte le trasformazioni eseguite, le più articolate e interessanti sono state:

- la ricostruzione della struttura dei pagamenti: l'individuazione delle voci relative di ogni nota di pagamento o di riscossione e l'identificazione del soggetto destinatario. L'operazione è stata al quanto articolata dato che le informazioni erano suddivise in diverse tabelle non collegate direttamente tra loro, quindi è stato necessario ricostruire le informazioni collegando varie tabelle e poi filtrare i risultati eliminando le informazioni non necessari;
- la determinazione dell'area merceologica dei venditori, i dati relativi a queste informazioni non erano inseriti all'interno del database e quindi sono stati in primo luogo ricostruiti e inseriti nella base di dati operativa e successivamente elaborati;
- la determinazione della stagione in cui è avvenuta una movimentazione di magazzino, essendo molti prodotti stagionali o venduti prevalentemente in determinati periodi, si è ritenuto importante estrarre la stagione dai movimenti di magazzino e dalle date di scadenze dei pagamenti.

## 8.5 Processo di caricamento

Esistono diversi tipi di caricamento dati in un datawarehouse e possono essere classificati in questo modo [Ruggieri, 09]:

- *Initial load* (caricamento iniziale): avviene alla prima esecuzione della procedura di ETL e consiste nel caricamento iniziale dei dati sul datawarehouse;
- *Incremental load* (caricamento incrementale): consiste nell'aggiunta di nuovi dati al datawarehouse (*append*), l'aggiornamento di dati esistenti con dei nuovi dati (*merge distruttivo*) e l'aggiunta di nuovi dati marcando quelli esistenti (*merge costruttivo*);
- *Full refresh* (ricaricamento dall'inizio): consiste nel caricamento di dati eliminando quelli esistenti.

Ovviamente, la procedura di ricaricamento dovrebbe essere evitata il più possibile: se ad ogni aggiornamento del datawarehouse fosse necessario un ricaricamento dall'inizio, mantenere i dati aggiornati potrebbe diventare proibitivo dal punto di vista computazionale.

Per riuscire ad evitare il ricaricamento è necessario riuscire a individuare i nuovi record della base di dati operazionale e caricarli nel datawarehouse tramite un caricamento incrementale. Sfortunatamente non è sempre immediato individuare i nuovi record; si presentano due possibili alternative per l'individuazione:

1. uso di colonne di controllo: in molti casi i database sorgenti hanno colonne di controllo, cioè colonne utilizzate per tenere traccia della data di inserimento o di modifica di un relativo record. Queste colonne di controllo possono essere utilizzate per capire quali sono i nuovi record e quindi quali debbano essere processati ed inseriti all'interno del datawarehouse.
2. analisi dei log: nel caso in cui la base di dati di partenza non contenesse colonne di controllo è possibile ricostruire l'insieme dei nuovi record da selezionare leggendo e analizzando il log del database. Utilizzando i log è infatti possibile ricostruire tutte le operazioni fatte sul database e, conseguentemente, capire quali dati sono stati inseriti tra una fase di caricamento incrementale e la successiva.

Nel datawarehouse realizzato non si è mai incontrata una effettiva difficoltà nel comprendere quali fossero le nuove righe da elaborare, questa facilità nell'operare è stata possibile grazie alla presenza di colonne di controllo in tutte le tabelle sorgenti.

## 8.6 Esecuzione della procedura

Le procedure di ETL, qualora non avviate in tempo reale, hanno bisogno di essere eseguite in un secondo momento così da effettuare l'aggiornamento del datawarehouse.

L'esecuzione può essere lanciata manualmente, ma solitamente viene controllata da uno *scheduler* che non si limita semplicemente ad eseguire la procedura a tempo debito ma gestisce le dipendenze e le relazioni tra i diversi processi rendendone così possibile la parallelizzazione; monitora l'esecuzione delle operazioni e in caso d'errore fa ripartire la procedura.

Nel caso specifico si è utilizzato lo *scheduler* di SQL Server 2005 eseguendo le operazioni di mantenimento mensilmente.

## Capitolo 9

# ANALISI OLAP

In questo capitolo verranno descritti gli strumenti utilizzati per accedere alla base di dati di supporto alle decisioni e le possibili modalità di esplorazione dei dati.

L'utente finale dispone di due differenti modalità per esplorare i dati memorizzati nel datawarehouse.

1. La prima modalità consiste nell'accedere direttamente al datawarehouse interrogando una base di dati ROLAP (*Relational Online Analytical Processing*).
2. La seconda modalità permette di interrogare una struttura MOLAP (*Multidimensional On-Line Analytical Processing*) creata appositamente partendo dalla base di dati per il supporto alle decisioni.

Queste differenti modalità di esplorazione dei dati sono state implementate e rese fruibili dalla Cooperativa, per migliorare le possibilità di analisi: utilizzare la struttura MOLAP per visualizzare ed elaborare elevate quantità di dati, adoperando computer non performanti, in alcuni casi risulta molto difficoltoso o, a volte, quasi impossibile in quanto Excel richiede molte risorse per elaborare i dati. Questo problema può essere risolto utilizzando la struttura ROLAP che permette di eseguire, senza eccessivi problemi, anche analisi su grandi quantità di dati utilizzando una quantità inferiore di risorse. In aggiunta la struttura ROLAP mette a disposizione degli amministratori della base di dati per il supporto alle decisioni, anche possibili implementazioni future: in qualunque momento è possibile realizzare cubi con porzioni di dati del datawarehouse implementando

strutture studiate appositamente per eseguire determinate analisi realizzando soluzioni mirate per ogni singolo manager.

Prima di illustrare le differenze tra le due modalità di accesso cerchiamo di comprendere meglio la struttura multidimensionale realizzata.

## 9.1 Struttura multidimensionale MOLAP

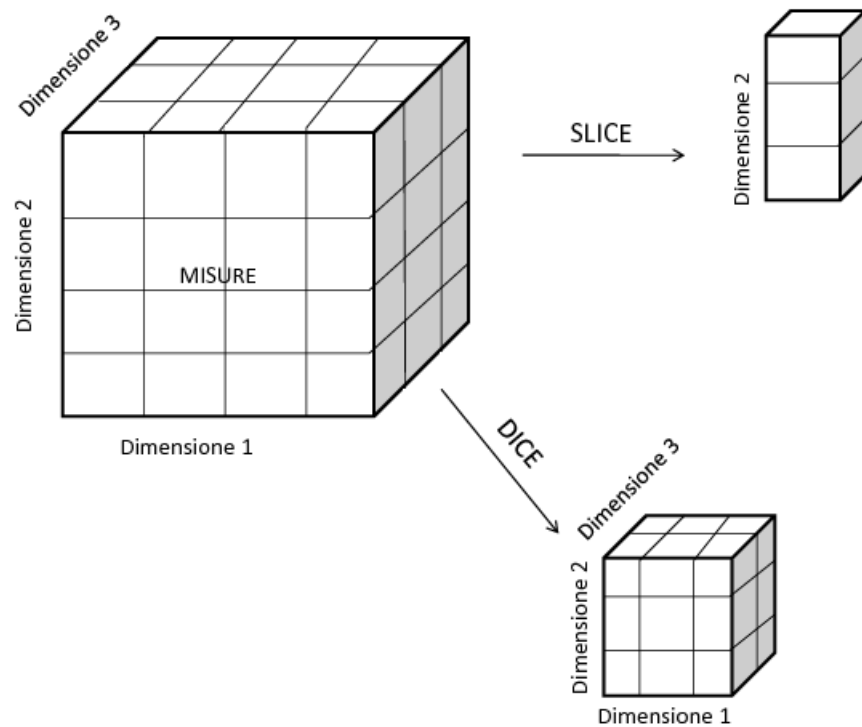
Si è deciso di affiancare a una struttura relazionale una struttura multidimensionale e, conseguentemente, anche un'alternativa modalità di esplorazione dei dati per offrire agli utilizzatori finali un prodotto il più completo possibile.

L'utilizzo di questa modalità di archiviazione offre diversi vantaggi: la struttura MOLAP è ottimizzata in modo da garantire le massime prestazioni durante l'esecuzione delle query. Poiché una copia dei dati di origine risiede nella struttura multidimensionale, le query possono essere eseguite senza accedere ai dati di origine. I tempi di risposta alle query sono in genere più veloci con la modalità di archiviazione MOLAP rispetto alle modalità di archiviazione ROLAP.

Oltre ai vantaggi appena elencati, si presentano anche degli svantaggi: poiché i dati di origine contenuti nel datawarehouse vengono modificati, gli oggetti nell'archivio MOLAP devono essere processati periodicamente in modo da comprendere tali modifiche e renderle disponibili agli utenti. L'intervallo di tempo tra un caricamento dei dati e quello successivo crea un periodo di latenza durante il quale i dati contenuti nella struttura MOLAP possono essere diversi da quelli della struttura ROLAP.

La realizzazione della struttura MOLAP consiste nella costruzione di un cubo con i dati contenuti nella base di dati per il supporto alle decisioni e, successivamente, accedere a questo cubo tramite interrogazioni. I dati utilizzati per la costruzione del cubo si dividono in 2 differenti gruppi: *misure* e *dimensioni*. I primi sono costituiti da dati contenuti nei fatti del datawarehouse, si troveranno all'interno del cubo e costituiranno le informazioni che verranno lette. I secondi sono costituiti da dati contenuti nelle dimensioni dei fatti del datawarehouse e andranno a formare le dimensioni del cubo.

E' possibile estrarre le informazioni dal cubo utilizzando due operazioni: la *SLICE* (taglia una fetta del cubo con una restrizione su una delle dimensioni) e la *DICE* (taglia un cubetto del cubo con una restrizione su due o più dimensioni) [Albano, 09]. La Figura 9.1 può facilitare la comprensione.



**Figura 9.1:** Operazioni SLICE e DICE

### 9.1.1 Realizzazione del cubo

Il cubo è stato realizzato utilizzando SQL Server Business Intelligence Development Studio. La procedura per la creazione è relativamente semplice e intuitiva: bisogna scegliere la sorgente dati, le misure e le dimensioni da utilizzare per la creazione del cubo e alla fine processare la struttura creata così da popolarla con i dati. La Figura 9.2 illustra i passi eseguiti durante il processo di creazione del cubo.

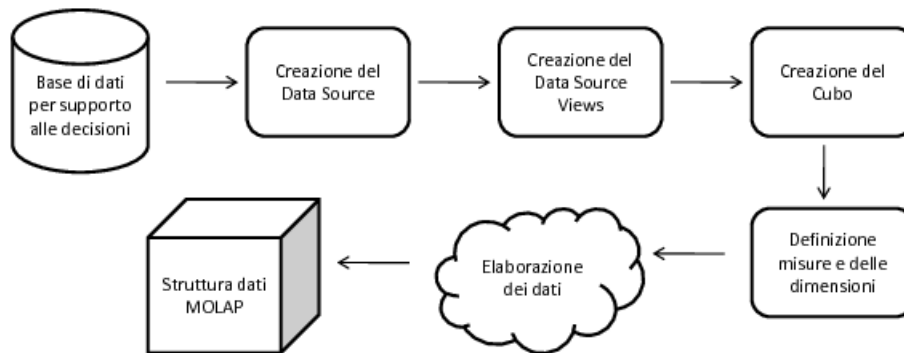
### 9.1.2 Visualizzazione finale dei dati

Per accedere ai dati l'utente finale dispone tre differenti strumenti:

- utilizzare SQL Server Business Intelligence Development Studio;
- utilizzare un foglio di calcolo (ed esempio Excel);



- scrivere direttamente interrogazioni MDX adoperando SQL Server Management Studio.



**Figura 9.2:** Processo di creazione del cubo

### SQL Server Business Intelligence Development Studio

SQL Server Business Intelligence Development Studio mette a disposizione un'interfaccia che permette l'esplorazione di cubi. Tramite questa interfaccia è possibile selezionare le misure e le dimensioni che si vogliono usare durante un'analisi definendo così la struttura della tabella che conterrà i dati desiderati.

L'utilizzo di questa interfaccia viene mostrato nella terza interrogazione della sezione 9.3.

### Excel

Anche Excel offre la possibilità di connessione a un'origine dati Analysis Services. Tramite questa interfaccia è possibile creare una connessione verso SQL Server Analysis Services ed importare i dati di un cubo sotto forma di tabella o grafico pivot.

Nel secondo esempio riportato nella sezione 9.3 viene mostrato come l'utente finale riesca a visualizzare i dati estratti. Anche se l'esempio riguarda la struttura ROLAP la visualizzazione risulta uguale per entrambi le strutture.

Anche in questo caso Excel mette a disposizione un'interfaccia semplice e intuitiva per selezionare le misure e dimensioni necessarie per le analisi che si vogliono eseguire.

### **SQL Server Management Studio**

L'ultima modalità di accesso messa a disposizione all'utente consiste nello scrivere manualmente le interrogazioni nel linguaggio *MDX* (MultiDimensional Expressions) [G. Spofford, 06].

Utilizzando questo linguaggio, realizzato appositamente per interrogare strutture dati multidimensionali, è possibile estrarre i dati dal cubo e visualizzarli sotto forma tabellare.

Scrivendo direttamente le interrogazioni al cubo è possibile ottimizzare ulteriormente le query e sfruttare al massimo le potenzialità offerte dalla struttura MOLAP. In Figura 9.3 viene riportata una possibile interrogazione realizzata in MDX.

L'interrogazione visualizza le quantità acquistate di ogni prodotto (il codice di ogni prodotto) negli anni. Come si può osservare la rappresentazione dei dati risulta essere più elementare e schematica rispetto ad Excel; le uniche azioni che si possono fare per migliorare la visualizzazione sono scegliere i valori da visualizzare sulle righe e quelli da inserire nelle colonne.

## **9.2 Struttura relazionale ROLAP**

Utilizzare la struttura relazionale consente all'utente finale di accedere e analizzare i dati contenuti nella base di dati per il supporto alle decisioni. Per la visualizzazione dei dati si è utilizzato Excel.

Dopo aver studiato numerosi applicativi sia Open-Source che con Licenza d'uso si è presa questa decisione perché utilizzare Excel ha di fatto portato due vantaggi:

- un risparmio in termini monetari dato che l'azienda possiede già diverse licenze per l'utilizzo del software;
- si propone un programma con un'interfaccia semplice da utilizzare e già conosciuta dagli utenti finali.

```

select NONEMPTY({[Prodotti].[Id Old].[Id Old]}) on columns,
NONEMPTY({[Date].[Anno].[Anno]}) on rows
from [Krlie Dwh]

```

	3	43	44	45	46	47	48	49	50	51	52	53	54	55	57	58	59	61
2002	(null)	12	12	6	5412	33072	11262	372	30	(null)	0	0	1452	1356	762	0	0	0
2003	0	6	3	5	1584	34758	10590	119	84	144	36	55	1569	693	843	0	7	4
2004	(null)	0	0	0	(null)	18264	5274	0	0	(null)	0	0	294	160	0	0	0	0
2005	(null)	(null)	0	(...)	(null)	26638	9191	0	(null)	(null)	0	(null)	504	408	0	(null)	(null)	(null)
2006	(null)	(null)	(...)	(...)	(null)	22320	7928	(null)	(null)	(null)	(null)	(null)	492	0	(null)	(null)	(null)	(null)
2007	(null)	(null)	0	(...)	(null)	24308	10762	(null)	(null)	(null)	(null)	(null)	4271	2556	(null)	(null)	(null)	(null)
2008	(null)	(null)	11	(...)	(null)	26706	12066	(null)	(null)	(null)	0	(null)	2752	804	(null)	(null)	(null)	(null)

Figura 9.3: Interrogazione eseguita in MDX

## Excel

Per accedere ai dati tramite Excel è indispensabile creare una connessione verso una sorgente dati esterna. La connessione creata, a differenza della precedente, è realizzata verso SQL Server con lo scopo di accedere ai dati contenuti nel datawarehouse e visualizzarli con l'ausilio di tabelle e grafici pivot.

Una implementazione di questa visualizzazione è illustrata nel secondo esempio della sezione 9.3, nel caso specifico estraendo numerosi dati si è potuta realizzare solo la tabella pivot.

## 9.3 Interrogazioni

In questa sezione viene illustrato come sono state eseguite le interrogazioni al datawarehouse e al cubo e, successivamente, come i dati estratti possono essere visualizzati dall'utente finale.

I primi 2 esempi illustrano le possibili visualizzazioni realizzabili interrogando la struttura ROLAP, mentre nell'ultimo esempio interroga la struttura MOLAP e successivamente ne visualizza i dati estratti.

### Vendite

Il primo esempio fa relativo all'analisi delle vendite descritte in Tabella 5.5. L'analisi ha l'obiettivo di estrarre le quantità e il valore del venduto nel quinquennio 2004 - 2008, per anno, per prodotto e per cliente che ha effettuato l'acquisto.

Per problemi di privacy non sono riportati i nomi dei clienti ma viene mostrato solamente il codice identificativo di ogni cliente.

Questa analisi può essere realizzata in due differenti modi:

1. scrivendo una query che estrae le misure e le dimensioni che si vogliono inserire nel report

```
1 select anno, cod_cliente , des_prodotto ,  
2 sum(quantita)as quantita , sum(valore_venduto)as valore_venduto  
3 from vendite , prodotti , data  
4 where vendite.cod_data=data.cod_data  
5 and vendite.cod_prodotto= prodotti.cod_prodotto  
6 and anno IN(2004,2005,2006,2007,2008)  
7 group by anno, cod_cliente , des_prodotto
```

il risultato è osservabile in Figura 9.4. L'utilizzo di Excel per visualizzare i dati estratti dalla query permette anche di aggiungere ulteriori filtri (successivi all'estrazione) e marcatori per evidenziare le celle che contengono valori giudicati interessanti dall'utilizzatore finale;

2. sempre tramite Excel ci si può collegare al datawarehouse e scegliere, tra tutti gli attributi presenti nelle tabelle, le misure e le dimensioni da utilizzare per la realizzazione del report desiderato e decidere dove posizionarle: nelle colonne, nelle righe e quali misure inserire nei valori, senza effettuare una pre-selezione dei valori scelti tramite una query SQL. Anche in questo caso è possibile aggiungere successivamente filtri e marcatori per evidenziare le celle più interessanti. La visualizzazione dei dati estratti con questa modalità di accesso è illustrata nel successivo esempio.

Si può notare come, tramite il tasto più in prossimità dei codici clienti sia possibile espandere a piacimento il singolo individuo visualizzando i prodotti acquistati. Questa funzionalità, esclusiva delle gerarchie, è utilizzabile in questo caso in cui non c'è una relazione tra cliente e prodotto solo perché i dati vengono estratti tramite un interrogazione SQL.

## **Inventario**

Il seguente esempio ha lo scopo di realizzare alcune analisi descritte in Tabella 5.6. Per la realizzazione di questo report si è deciso di accedere direttamente

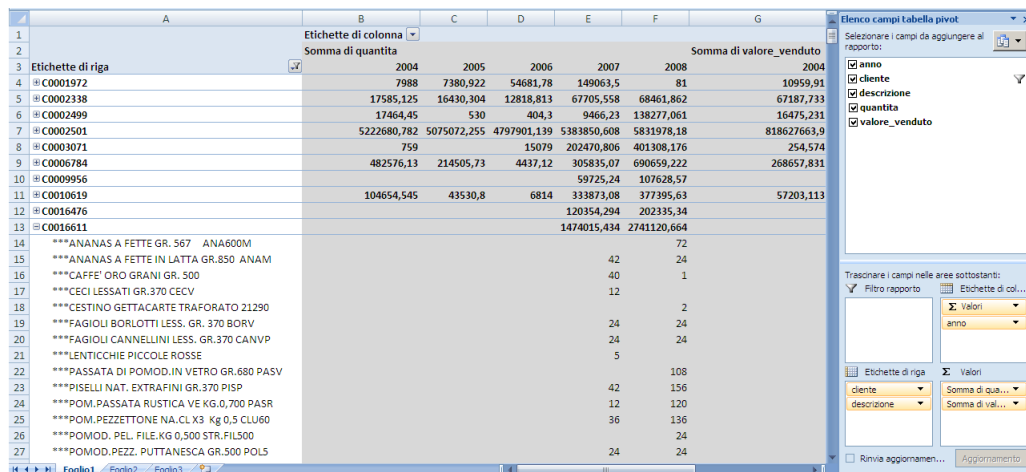


Figura 9.4: Visualizzazione dei dati Tabella 5.5

al datawarehouse selezionando le misure e le dimensioni dal menù di Excel, senza effettuare una preselezione tramite un'interrogazione SQL.

Il report in Figura 9.5 ha l'obiettivo di estrarre le quantità acquistate e quelle quantità vendute per anno per i magazzini e per i prodotti. Anche in questo caso per rispettare le esigenze della Cooperativa si è deciso di non visualizzare il nome dei magazzini ma solo il codice.

Come si può facilmente notare, le differenze sostanziali tra questa visualizzazione e la precedente consistono nelle informazioni contenute nel pannello sulla destra. Nel precedente caso l'accesso ai dati veniva realizzato tramite una interrogazione SQL e conseguentemente, nel pannello comparivano solo gli attributi selezionati tramite l'interrogazione. In questo esempio ci si collega direttamente al datawarehouse senza realizzare un preventivo filtraggio delle informazioni, questo fa sì che nel pannello vengano riportati tutte le misure e le dimensioni raggruppate per le tabelle di appartenenza.

Altra differenza consiste nell'assenza della possibilità di espandere o ridurre i prodotti dei magazzini utilizzando il tasto più. Ciò non è possibile perché le dimensioni *magazzino* e *prodotto* non appartengono alla stessa gerarchia.

La selezione dei campi interessati, come nel caso precedente, avviene trascinando il capo all'interno dell'area sottostante. Anche in questa visualizzazione è sempre possibile realizzare filtri sui dati visualizzati o applicare eventuali marcatori alle celle.

Etichette di riga	2002	2003	2004	2005	2006	2007	2008	Qta
BAR					6,356	302,373	19436,358	14537,875
***ANANAS A FETTE GR. 567 ANA600M								1
***CAFFE' BELCAFE' GR.250 M0006								1
***CANDELA SCALDAVIVANDE PZ. 10 A/139								1
***COLTELLO BLU PZ20 A/5794						0		0
***DOPPIO CONCENTR. POMOD.GR. 500 CON500						13		26
***L MORTADELLA BOLOGNA IGP GR.125 4560						6		
***LOMBATINA FIL/VITELLO LATTE ART.38						0,854		
***MACINATA SCELTA CONF. FAMIGLIA A/75						1,562		
***MORTADELLA IBIS						0,43		
***MORTADELLINA MOMO' 3x60 GR. 5408						2		
***MOZZARELLA BUFALA CAMPANA GR.250 46						3		
***PANCETTA AFFUMICATA CUB. GR.125 5907						1		
***PANINI LATTE GR.350 060879								2
***PASSATA BRIK KG. 1 DEL MONTE						441		27
***PASSATA DI POMOD.IN VETRO GR.680 PASV						1		5
***PIATTO DESS./FRUTTA PZ100								0
***PL CAFFE' GR 500 5V. ESPRESSO						7		
***PL COPPA GR. 125 4069						14		
***PL GRAPPA 40° AST.TONDO CL.50 413								2
***PL MERENDERO GR. 500						15		
***PL PANCETTA GR.125 4785								11
***PL PROSCIUTTO COTTO GR. 125 4559								5
***PL PROSCIUTTO CRUDO GR. 100 4562								15

Figura 9.5: Visualizzazione dei dati Tabella 5.6

## Scadenze

In questo ultimo report viene realizzata parte dell'analisi descritta in Tabella 5.7. Differentemente dalle due interrogazioni eseguite in precedenza questa analisi è realizzata interrogando la struttura MOLAP tramite l'interfaccia messa a disposizione da SQL Server Business Intelligence.

L'analisi riporta la somma dell'importo dovuto e dell'importo residuo alla data di scadenza (per anno e per mese), per agente. Anche in questa analisi per rispettare le esigenze dell'azienda si è deciso di visualizzare solo i codici degli agenti. La Figura 9.6 mostra il risultato del report.

Le lievi differenze tra l'utilizzo di questa interfaccia e l'utilizzo di Excel consistono nel non potere applicare ulteriori filtri o speciali marcatori ai dati estratti dal cubo. Tramite questo strumento è possibile infatti solo visualizzare e stampare i dati estratti. Altro interessante dettaglio è come vengano gestite le gerarchie: come per Excel anche questa interfaccia permette di espandere gli attributi appartenenti a una gerarchia: è possibile scendere in dettaglio in anno arrivando al mese e eventualmente al giorno.

The screenshot shows a BI tool interface with a data table. The table has columns for 'Agenti', 'Anno' (2003-2007), and 'Mese'. The data is organized into a pivot table structure. The 'Agenti' column lists various agent codes and names. The 'Anno' column is set to 'Mese' (Month). The table displays 'Importo Dovuto' (Due Amount) and 'Importo Residuo' (Residual Amount) for each agent across the years 2003 to 2007.

Agenti	2003	2004	2005	2006	2007					
(Blank)	45298045.3060001	63709.84	38455037.6819997	69924.8	35229530.3800002	463794.151	40621578.5440002	274287.554	55046286.26	65665.4
A0000001	2907.26	0	2035.27	0	572.18	0	777.89	0	1891.36	0
A0000002	84	0	0	0	0	0	0	0	1217.3	0
A0000003	91.74	0	0	0	0	0	0	0	0	0
A0000004	717.04	0	0	0	0	0	0	0	0	0
A0000005	51.04	0	678.74	0	0	0	0	0	0	0
A0000006	89.08	0	0	0	714.28	0	0	0	0	0
A0000007	42.8	0	0	0	0	0	0	0	0	0
A0000008	14.06	0	0	0	0	0	140	0	377.47	0
A0000009	202.48	0	0	0	0	0	0	0	0	0
A0000010	15.12	0	0	0	0	0	0	0	0	0
A0000013	287.34	0	100	0	950.8	0	2520.1	0	811.28	0
A0000014	74.5	0	0	0	0	0	0	0	0	0
A0000017	0	13942.36	0	0	730.3	0	0	0	378.1	0
A0000018	0	0	0	0	600.46	0	0	0	611.12	0
A0000019	0	0	0	0	639.92	0	0	0	0	0
A0000023	0	0	0	0	600	0	0	0	0	0
A0000024	0	0	0	0	0	0	0	0	0	0
A0000025	0	0	0	0	0	0	0	0	0	0

Figura 9.6: Visualizzazione dei dati Tabella 5.7

## Capitolo 10

# CONCLUSIONI

L'obiettivo del lavoro svolto era quello di progettare e realizzare un datawarehouse che soddisfacesse i seguenti obiettivi:

1. essere funzionante e utile alla Cooperativa Agricola di Legnaia, ossia non rimanere un semplice progetto su carta o avere una realizzazione approssimativa;
2. avere dei costi il più possibile contenuti, utilizzando software già in possesso dell'azienda o realizzando soluzioni scritte a mano;
3. avere delle buone prestazioni durante le interrogazioni.

Il primo obiettivo è stato sicuramente raggiunto in quanto il datawarehouse è stato testato ed è perfettamente funzionante. Le funzionalità offerte dalla base di dati per il supporto alle decisioni sono a tutti gli effetti sostitutive delle precedenti analisi che venivano eseguite su richiesta dei vertici aziendali ed erano studiate per estrarre informazioni ben precise.

Il secondo obiettivo è stato raggiunto in quanto non si è reso necessario acquistare nessun nuovo software privilegiando l'utilizzo dei programmi già in possesso dell'azienda.

Il terzo obiettivo è stato raggiunto in maniera soddisfacente in quanto i tempi di esecuzione delle query sono sempre all'interno di intervalli accettabili, anche se migliorabili realizzando indici specifici.

Per il futuro sarebbe interessante utilizzare i dati a disposizione in maniera alternativa: non limitarsi soltanto ad un'analisi multidimensionale ma applicare tecniche di data mining per approfondire la conoscenza dei dati e mettere in



luce ulteriori informazioni nascoste. Realizzare un'interfaccia che utilizzando dispositivi grafici renda più semplice, intuitiva e immediata la comprensione dei dati elaborati.

In conclusione, l'esperienza di lavoro è stata sicuramente positiva in quanto ha dato la possibilità di conoscere in prima persona le problematiche che si affrontano e che devono essere risolte durante la realizzazione di un datawarehouse e soprattutto, perché ha arricchito l'esperienza formativa offerta dal corso di laurea con conoscenze difficilmente apprendibili dai libri.

# Bibliografia

- [Chiti, 03] S. Chiti (2003), Una passione secolare, 1903-2003 I cento anni della Cooperativa Agricola di Legnaia, Edimedia srl, Firenze.
- [Albano, 09] A. Albano (2009), Dispense del corso di Basi di dati di supporto alle decisioni.
- [I. Ben-Gan, 08] I. Ben-Gan (2008), Inside Microsoft SQL Server 2008: T-SQL Querying, Microsoft Press, 2009.
- [G. Spofford, 06] G. Spofford (2006), MDX solutions with Microsoft SQL Server Analysis Services 2005 and Hyperion Essbase, Wiley Publishing INC, 2006.
- [Ruggieri, 09] S. Ruggieri (2009), Dispense del corso di Laboratorio di Sistemi Informativi Economico-Aziendali.
- [Kimball, 02] R. Kimball M. Ross (2002), The data warehouse toolkit: the complete guide to dimensional modeling, Wiley, New York, 2002.
- [SQLS, 05] G. Rajesh, SQL Server 2005 Express: panoramica, <<http://www.visual-basic.it/articoli/sfSQLServer2005ExpressEditionOverview.htm>>, Novembre 2005.
- [GS, 09] Documentazione ForeOffice, <<http://62.110.161.101/Documentazione/ForeOffice.pdf>>, 20 Agosto 2009.
- [D'Ercole, 07] N. D'Ercole (2007), *TESI DI LAUREA Progettazione e realizzazione di un datawarehouse per il supporto al controllo contabile e l'analisi delle vendite.*

# RINGRAZIAMENTI

Ringrazio tutti coloro che mi hanno seguito durante la realizzazione di questa tesi in modo particolare Antonella Ficini e Lorenzo Migliorino di Neonevis s.r.l. per avermi aiutato durante lo sviluppo del progetto.

La Cooperativa Agricola di Legnaia per avermi dato la possibilità di sviluppare il datawarehouse.

Il professore Salvatore Ruggieri per avermi seguito e consigliato durante la stesura della tesi.

Inoltre ringrazio tutti gli amici di Perugia che sono venuti ad assistere alla discussione confidando in uno striscione non troppo eccessivo ma allo stesso tempo adeguatamente "incisivo", tutti gli amici di Pisa che sono venuti solo con lo scopo di scroccare qualcosa: sicuramente Simone, Andrea e Emanuele. Il Dott. Rag. Mirco Rastelli con il quale potrò riniziare a guardare dall'alto verso il basso tutti i nostri amici che ancora non hanno conseguito la laurea specialistica.

In ultimo ringrazio Silvia per non avermi messo i bastoni tra le ruote e avermi sopportato in questi ultimi due mesi.