

## **Capitolo settimo**

### **La filosofia Data Warehouse**

---

## 7.1. Introduzione

*Il mercato di riferimento geografico di un'azienda non è più il paese o il continente, ma l'insieme dei grandi paesi industrializzati*<sup>1</sup>. Si parla, dunque, di globalizzazione dei mercati, intendendo con questo termine il fatto che i mercati locali, essendo fra loro interdipendenti, diventano un tutt'uno. Tutto ciò determina per le imprese, sia nuove opportunità, sia un aumento della pressione competitiva. La pressione competitiva aumenta anche a causa della deregolamentazione e liberalizzazione di tutti i settori. Linee aeree, ferrovie, telecomunicazioni, servizi di fornitura d'energia ed acqua hanno tutti assistito alla fine del supporto governativo delle posizioni monopolistiche permettendo a nuovi fornitori di entrare nei mercati e di scegliersi gli utenti. Le imprese, nell'ambito delle strategie di Business<sup>2</sup>, devono affrontare la competizione attraverso nuovi, più veloci, più attraenti modi che conquistarsi gli acquirenti<sup>3</sup>.

La forza della competizione è legata all'*efficienza* del mercato. L'efficienza del mercato dipende dalla facilità d'accesso dello stesso da parte dell'utente e dalla quantità d'informazioni disponibili al consumatore su tutti i prodotti del mercato.

Le nuove tecnologie ampliano sia l'accessibilità del mercato, sia la conoscenza dei prodotti da parte dell'utente, avvicinandolo sempre più all'accesso perfetto e alla perfetta informazione.

In una situazione di questo tipo la competizione tende ad essere più dura<sup>4</sup>.

---

<sup>1</sup> J.J. Lambin : "Marketing", McGraw-Hill.

<sup>2</sup> "La redditività di una impresa dipende da due fattori :la redditività del settore in cui opera e la capacità dell'impresa di stabilire un vantaggio competitivo rispetto ai suoi rivali. La scelta del settore o dei settori riguarda la strategia a livello corporate. Stabilire un vantaggio entro un particolare settore è ciò che si propone di stabilire la strategia a livello di business." R.M. Grant : L'analisi strategica nella gestione aziendale.

<sup>3</sup> Si possono portare alcuni esempi : assistenza diretta via telefono, home-banking e shopping via cavo o tramite Internet.. The second Age of Computing Data Warehousing with Oracle. R. Stewart.

<sup>4</sup> "Un esempio : Un giocatore di golf residente a N.Y. prenderà in considerazione (e quindi è limitato) per il suo equipaggiamento da golf solo i fornitori di N.Y.. Al contrario con un PC ed un equipaggiamento internet, una carta di credito il nostro giocatore di golf può velocemente acquistare da qualsiasi fornitore nel mondo che effettui pubblicità sulla rete.

Per avere un vantaggio competitivo in un ambiente caratterizzato dall'industria globale, innovativa, deregolamentata, sempre più aggressiva, le imprese devono migliorare la loro comprensione del business, dei loro clienti, dei loro mercati. Devono conoscere con esattezza chi sono i clienti, come fare a mantenerli, come e quando spendono il loro denaro, quali beni e servizi saranno acquisiti, da chi e così via . Le imprese devono focalizzare le loro energie su questi elementi in modo tale da poter sfruttare al massimo le diverse e mutevoli opportunità che si presentano. Se ciò non avverrà, i clienti andranno altrove.

Nella produzione di massa ha sempre più importanza la cosiddetta *personalizzazione del prodotto*. Oggi non è più sufficiente per un'azienda portare un qualsiasi prodotto di massa o servizio sul mercato maturo ed aspettarsi di venderlo abbastanza bene per un certo periodo di tempo. Il cliente chiede un prodotto più personalizzato, che soddisfi meglio i suoi bisogni.

Per sopravvivere in questi mercati le imprese avranno bisogno di identificare in modo più adeguato i bisogni dei consumatori, segmentando il mercato stesso, per offrire prodotti che si avvicinino il più possibile alle richieste di questo o di quel segmento. Anche se il prodotto in se stesso non può essere

---

Questo fatto rende la vita dei fornitori di attrezzature di golf di N.K. più difficile." The second Age of Computing Data Warehousing with Oracle. R. Stewart.

<sup>4</sup> Caratteristiche dello scenario Industriale :

- ? Maggiore concorrenza
- ? Diffusione/Omogeneizzazione della Tecnologia
- ? Incertezza economica
- ? Ritmo di crescita più basso
- ? Minore prevedibilità dei clienti
- ? Riduzione del Time-to Market
- ? Re-engineering dei processi Aziendali
- ? Aggiornamento dell'offerta
- ? Valorizzazione della base dei Clienti Esistenti

Fonte :Data Warehouse, metodologia, Architettura e prodotti. Presentazione Datamat per il gruppo FIAT.

personalizzato, il modo di fornirlo e con cui si effettua il pagamento può esserlo.

Questa situazione genera una maggiore richiesta di informazioni per conoscere meglio i clienti, i mercati, la concorrenza.

Le fonti di queste informazioni si trovano all'esterno ma innanzi tutto all'interno dell'impresa. La tecnologia dei computer è stata applicata fin dai primi anni di attuazione ai processi operativi di un'impresa (sistemi di fatturazione, emissione di ricevuta, prenotazioni, gestione degli ordini di vendita, controllo del magazzino, finanza, ecc..).

Attraverso i sistemi operazionali attualmente in uso, una grande quantità di dati viene raccolta e conservata. Questi dati descrivono con molta precisione, e spesso implicitamente, gli aspetti critici del business: clienti, mercati, prodotti, vendita e le relazioni che intercorrono tra questi. Tali informazioni devono essere rese disponibili a coloro che ci lavorano sopra: direttori, lo staff di marketing, gli analisti di business e di prodotti, ecc..

Così come sono prodotti, questi dati "operazionali" non riescono a soddisfare le esigenze di chi deve lavorare con l'informazione attraverso particolari applicazioni DSS, MIS o altro. Questi dati, così come sono raccolti, sono destinati a servire i sistemi operazionali per i quali sono stati creati. E non potrebbe essere altrimenti; in altre parole non si potrebbe riconfigurare la raccolta di questi dati, che devono servire particolari operazioni per accogliere le esigenze informative, poiché verrebbe meno quello che è il loro scopo originario che deve essere senz'altro soddisfatto.

Una soluzione è quella di duplicare i dati liberandoli dai confini dei diversi sistemi per integrarli ed arricchirli con informazioni di provenienza esterna (ricerche di mercato, analisi industriali, notizie da agenzie, reti mondiali ed altro). Devono, inoltre, essere predisposti dei Tools (strumenti) che mettano in grado i vari utenti di analizzare le informazioni a loro disposizione.

Il Data Warehouse si occupa proprio di questo.

*In un'interpretazione più limitata un Data Warehouse è semplicemente il database o i database in cui i dati rilevanti dei sistemi operativi sono copiati e resi disponibili agli utilizzatori di informazioni attraverso strumenti analitici di indagine. In un senso più ampio rappresenta l'intera struttura di estrazione di dati grezzi dai sistemi operativi, di trasformazione in informazioni affidabili, di trasporto in apposito ambiente per l'utente delle informazioni e la fornitura degli strumenti on-line per analizzare questi dati (OLAP Tools, data mining, ecc..) al fine di fornire conoscenza alla Direzione del business<sup>5</sup>.*

## **7.2. Evoluzione dei sistemi di raccolta dati**

Le aziende hanno sempre cercato di raccogliere dati. In passato ciò era possibile solo tramite mezzi cartacei. Successivamente si è avuto l'avvento di mezzi meccanici ed elettronici che hanno facilitato questa attività.

In un primo momento (anni '60) venivano create solo applicazioni con dati legati alle applicazioni stesse. Questo faceva sì che per ogni applicazione si aveva un archivio, creando così evidenti problemi di ridondanza dei dati.

---

<sup>5</sup> . The second Age of Computing Data Warehousing with Oracle. R. Stewart.

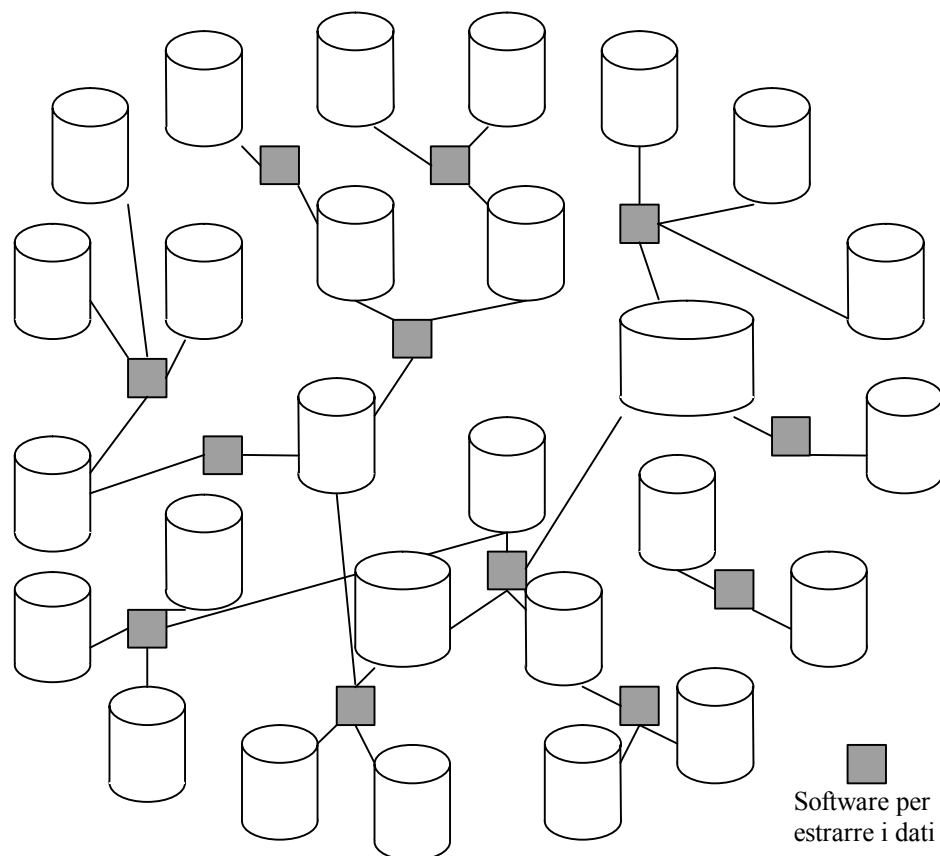
“il Data Warehouse è un processo, non un prodotto. Questa rappresenta una particolare tecnica per assemblare e gestire dati provenienti da varie sorgenti per realizzare una singola, dettagliata vista di una parte o di tutto il business.” White paper di NCR Corporation : Data Warehousing :Clearing the Confusion.

“Il D/W nasce dalla costituzione di un ambiente, fisicamente separato dal database operativo, ottimizzato per le interrogazioni e il supporto alle decisioni. C'è, quindi, una netta distinzione tra sistema operativo e sistema decisionale.....Sono differenti anche gli utenti dei due sistemi : nel primo caso è lo staff operativo dell'azienda (contabili, addetti agli sportelli..), mentre nel secondo sarà il top il middle manager a servirsi dei dati del D/W per l'analisi, l'accesso e la produzione di report.” P. Lombardi : Nuovi strumenti per cercare i dati, Zerouno 3/1995

B. Cortona nell'articolo di Zerouno del 4/1995 afferma :” La nascita del concetto di warehouse (magazzino) risale alla metà degli anni '80 ma solo dai primi anni '90 ha iniziato a trovare la piena attuazione. L'idea che sta alla base è molto semplice : estrarre dai database operazionali i dati significativi, compattarli e metterli infine a disposizione degli strumenti desktop, che effettuano le proprie interrogazioni sul warehouse e non più sui database operazionali.”

D. Sandri nell'articolo :La soluzione del SAS Institute di Zerouno del 4/1995 afferma : “Il Data Warehouse è in sostanza una soluzione tecnologica con la quale i dati vengono archiviati secondo un ben determinato formato che ne facilita l'accesso e la consultazione da parte degli utenti. In genere interessa non tutti gli utenti di una azienda, ma per lo più i decision maker, quelle persone che nell'ambito dell'impresa hanno bisogno di dati di sintesi provenienti dalle varie unità aziendali, sui quali operare ulteriori elaborazioni e grazie ai quali simulare scenari e prendere infine decisioni.”

Verso gli anni '70 cominciano a diffondersi i primi gestori di basi dati in cui si realizza una separazione fra le applicazioni e gli archivi dati. Allo stesso modo si diffondono nuovi supporti fisici per la registrazione (dischi magnetici). Una singola base di dati può servire più applicazioni. Verso gli anni '80 prendono il via le tecnologie distribuite, per cui le applicazioni che erano dirette ai decisori potevano essere spostate più vicino ad essi (attraverso l'uso dei PC). Si comincia a pensare di poter utilizzare i dati di questi database, costruiti però per servire specificatamente le transazioni, per sistemi MIS o DSS. Ma le difficoltà per realizzare questi sistemi erano molte. Sempre negli anni '80 si sono creati dei programmi capaci di estrarre i dati da una base di dati, secondo vari criteri, per trasferirli altrove. Questi dati a loro volta possono essere trasferiti ancora in altri database creando la possibilità di avere schemi come quelli in figura.



**Figura 7.1. Legacy system**

Questa "ragnatela" viene chiamata legacy system.

### 7.3. Problemi con la naturale evoluzione della architettura.

L'evoluzione di questa architettura pone alcuni problemi per l'utilizzo diretto di applicazioni DSS, EIS ed analisi<sup>6</sup>. I principali problemi sono :

- ? la credibilità dei dati
- ? la produttività
- ? l'incapacità di trasformare i dati in informazioni.

*Mancanza di credibilità dei dati.*

Il primo di questi problemi è quello della mancanza di credibilità dei dati. Inmon porta il caso di due dipartimenti che realizzano due report sullo stesso tema con i dati del legacy ottenendo due risultati differenti. Di fronte a queste incongruità il management non sa bene cosa fare. Ci sono alcune ragioni che possono causare questa incongruità :

- ? i report non si sono effettuati nello stesso tempo
- ? scelta di parametri differenti per l'analisi dei di dati
- ? problemi con i dati esterni
- ? utilizzo di fonti diverse

Realizzare i report in **istanti di tempo differenti** può portare a dei risultati diversi poiché nel tempo trascorso si sono verificate delle transazioni che possono aver mutato i dati. La **scelta di parametri differenti** per l'analisi dei di dati può allo stesso modo produrre risultati differenti qualora, ad esempio, si decida di prendere in considerazione tutte le polizze vita o solo quelle più vecchie. I **dati esterni** che spesso vengono utilizzati nei report non sono coordinati fra di loro e non c'è neppure coordinamento con i dati interni utilizzati. Ad esempio, per realizzare un report possono essere stati utilizzati dati del Wall Street Journal mentre in un altro i dati di Business Week. I dati dei due giornali, anche se riguardano lo stesso argomento, possono essere

---

<sup>6</sup> ".....Convinti delle promesse dei produttori che il loro sistema di gestione dati sarebbe stato in grado di soddisfare sia le necessità di produzione (l'OLTP) che di supporto alle decisioni, gli EDP aziendali hanno per qualche tempo scelto di far accedere gli utenti direttamente ai database operazionali. In breve hanno dovuto fare marcia indietro dopo aver constatato che le operazioni sui dati tipiche dell'analisi per il supporto decisionale sono molto diverse da quelle tipiche dell'OLTP e pesanti nell'esecuzione." B. Cortona :Come accedere ai patrimoni di dati, Zerouno 4/1995

diversi Più in generale poi, ci possono essere delle discrepanze nei risultati dovuti alla **scelta di fonti diverse** per realizzare lo stesso report.

#### *Il problema della produttività*

Un altro problema è quello relativo alla produttività, ovvero alla efficienza dovuta alla naturale evoluzione di questa struttura. Per realizzare i report occorrerà localizzare i dati che sono sparsi nel legacy, estrarli e realizzare programmi di analisi dei dati.

I dati che servono in un report sono sparsi nel legacy. Questi si sono stratificati nel tempo senza pensare che un giorno si sarebbero dovuti integrare. Per cui, stessi dati potranno avere nomi diversi come dati diversi potranno avere lo stesso nome. Quindi, localizzare i dati è un processo difficile e noioso.

Una volta localizzati, i dati devono essere estratti. Per ogni fonte di dati si dovrà realizzare un programma per l'estrazione. Infine, si dovrà realizzare il report. Tutte queste attività occupano tempo, ed un eccessivo impiego dello stesso può portare a realizzare dei report che non hanno nessun valore perché non sono più attuali.

#### *Dai dati all'informazione*

Un altro limite della architettura legacy è quello relativo alla capacità di fornire informazioni. I decisori hanno bisogno di informazioni di carattere complesso derivanti dall'aggregazione di più fonti. Spesso questa integrazione è difficile poiché i dati vengono estratti da database che non sono stati progettati per servire questi scopi, ma per supportare delle transazioni. Inoltre, un altro limite è rappresentato dall'orizzonte temporale dei dati. Mentre i decisori hanno bisogno di informazioni storiche che mostrino gli andamenti passati e le previsioni sul futuro, ai database operazionali interessa solo mantenere aggiornati i dati e, quindi, tendono ad avere la memoria più corta (perché si tengono in linea solo i dati correnti).



#### 7.4. Il magazzino dei dati

Per superare l'empasse dei problemi visti fino ad ora si è trovata la soluzione Data Warehouse. Il cuore di questa architettura si basa su due tipi di dato. Quello **primitivo** residente nel legacy e quello **derivato** residente nel Warehouse. I dati **derivati** sono ricavati in tutto o in parte da quelli primitivi e posti nel Warehouse (o repository), vale a dire un grande database. Il Data Warehouse è di sola lettura e periodicamente viene aggiornato. Quindi, fra un aggiornamento e l'altro i dati sono sempre gli stessi, al contrario dei database operazionali che sono soggetti ad un continuo mutamento.

Si possono distinguere nel Data Warehouse quattro livelli.

Il primo è il livello operativo composto di dati primitivi del legacy ed eventualmente di altre fonti esterne. Il secondo è quello del Warehouse (o repository) in cui i dati, dopo una attività di acquisizione vengono inseriti a livello atomico. I dati non vengono letti direttamente da qui ma verranno ulteriormente estratti o sintetizzati nei Datamart (terzo livello, o livello dipartimentale)<sup>7</sup>, che rappresentano database destinati a servire particolari funzioni richieste dall'utente finale (analista DSS, data mining, ecc.). Il quarto livello è quello individuale, ovvero quello dell'analista o del decisore che con il suo PC utilizza delle applicazioni che vanno a "pescare" dati dal Datamart per effettuare l'attività di analisi.

Se la struttura logica di un Data Warehouse si dispone su questi quattro livelli, la struttura "fisica", o meglio hardware, è un **Client Server**. Infatti, l'analista opera con le sue applicazioni da PC mentre i vari database risiedono su server. Quindi, l'ambiente è quello di una rete con elaborazione distribuita<sup>8</sup>.

---

<sup>7</sup> Un Datamart rappresenta uno speciale subset di dati estratti da una unità centrale e tali dati vengono selezionati per servire una particolare funzione o applicazione." White Paper di NCR corporation : Data Warehousing :Clearing the Confusion

<sup>8</sup> L'architettura C/S può essere a due o a tre livelli. I sistemi a due livelli sono piuttosto semplici. Il client spesso definito Fat client perché svolge la maggior parte della elaborazione, comunica al server le richieste di accesso al database mediante SQL o attraverso una interfaccia come ODBC. Nei sistemi a tre livelli, l'elaborazione viene in buona parte affidata ad un application server distinto. In questo modo l'interfaccia utente, l'aspetto elaborativo e la gestione del database vengono separati. A. Dickman Zerouno 2/1996 (trasduzione da InformationWeek)

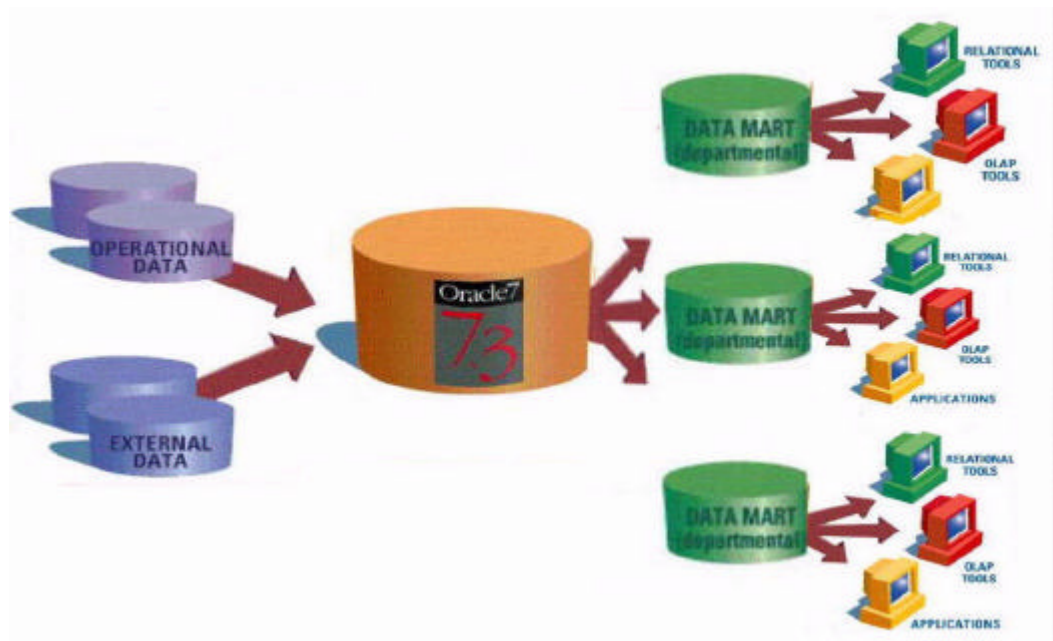


Figura 7.2. Livelli del Data Warehouse

### 7.5. Il costo del Data Warehouse

La valutazione dell'investimento in strutture Data Warehouse non può essere fatta tramite tecniche classiche come, ad esempio, il ROI. E' impossibile "pianificare" e verificare quali siano i reali costi e benefici che questo tipo di architettura determina <sup>9</sup>. *Fortunatamente il Data Warehouse è una architettura incrementale, per cui si può costruire un primo Data Warehouse con un*

<sup>9</sup> "Realizzare una soluzione di Data Warehousing ben dimensionata sulle esigenze aziendali comporta tutta una serie di difficoltà da superare; non soltanto nei criteri e nei metodi di implementazione ma anche nella capacità dell'azienda di supportare nel tempo questa scelta a fronte di una chiara valutazione dei costi diretti (software, hardware, reti) e soprattutto quelli indiretti, quali la definizione e il mantenimento della organizzazione preposta al ruolo di gestore del Data Warehouse (con figure professionali dedicate), con tutte le complesse problematiche relative all'aggiornamento periodico dei dati estratti dai data base operazionali e inseriti nel Data Warehouse in funzione delle differenti esigenze di supporto decisionale sparse nell'azienda" S. Uberti Foppa :Quando serve un D/W, Zerouno 4/1995  
Inmon in una intervista rilasciata a Giorgio Marras di Zerouno del giugno 1996 affermava che per ogni dollaro speso per soddisfare le esigenze dell'executive community se ne spendono almeno 9 per acquisire dati dalle applicazioni legacy e trasferirli all'ambiente EIS. Questo in assenza di un ambiente Data Warehouse. Attraverso il D/W questa componente di costo tende a scendere.

modesto impiego si denaro. Una volta che la prima porzione di D/W è stata costruita l'analista può esplorarne le possibilità<sup>10</sup>.

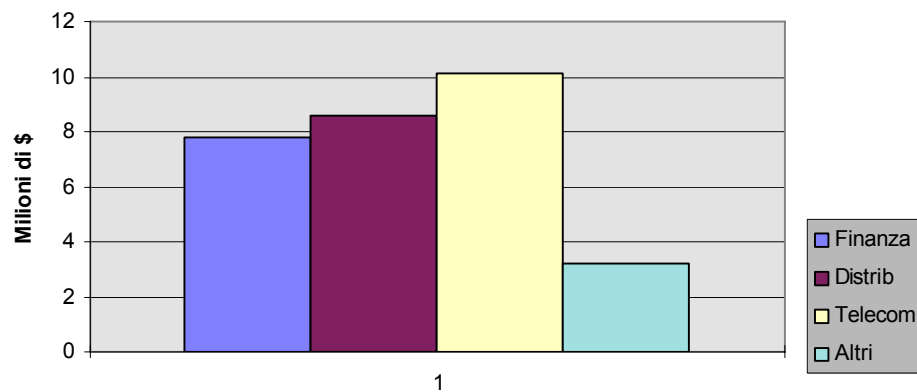
Comunque una società di ricerche, **Metrica**, ha realizzato una indagine su 117 aziende Europee (non ci sono aziende italiane) al fine di valutare quale fossero i vantaggi economici e commerciali del D/W legati a sistemi DSS.

Il 55% del campione ha già inserito sistemi di Data Warehouse e il 30% ha deciso di realizzarlo nei prossimi due anni.

Solo 31 delle 64 aziende che hanno implementato sistemi DSS e D/W hanno potuto o saputo quantificare in termini economici il vantaggio portato dalla scelta effettuata. Le altre non sono state capaci di definire i criteri di valutazione economica o di misura dei risultati.

Queste 31 aziende hanno valutato in media un guadagno attribuito all'uso di un DSS basato su D/W pari a 5,4 milioni di dollari. Il settore delle telecomunicazioni mostra un guadagno maggiore (10,1 milioni), seguito dalla distribuzione (8,6 milioni), dalla finanza (7,8 milioni) e dagli altri (3,2 milioni).

Il guadagno che un DSS può portare



<sup>10</sup> W.H. Inmon : "Building The Data Warehouse"