

Relazione sul data warehouse e sul data mining

INTRODUZIONE

Inquadrandolo il sistema informativo aziendale automatizzato come *costituito dall'insieme delle risorse messe a disposizione della tecnologia, dalle persone che le utilizzano e dalle applicazioni che possono essere fatte con esse, con la finalità di automatizzare l'archiviazione, la produzione, l'elaborazione e la distribuzione dei dati aziendali*, possiamo dire che i dati da esso utilizzati non provengono solo dall'interno dell'azienda, ma anche dall'esterno, da una banca dati, ma anche semplicemente da un articolo di un quotidiano. Quindi un generico sistema informativo utilizza dati *strutturati*, come il record di un dipendente, oppure i registri IVA, e da dati *non strutturati*, come riunioni, telefonate, segnalazioni di un agente, etc. etc.

Intanto, lo sviluppo dei sistemi informatici nelle aziende, che procede insieme al progresso dell'hardware e del software, corrisponde alle esigenze di:

- rendere automatici i lavori ripetitivi, eliminando gli errori delle trascrizioni e nella ripetizione dei dati in schedari e documenti diversi;
- poter disporre dei dati che servono in tempi molto brevi, utilizzando supporti che possono memorizzare grandissime quantità di informazioni.

Dalla meccanografia alle reti di computer l'evoluzione del sistema informatico aziendale ha introdotto anche procedure standard di espletamento delle attività nella progettazione, nella produzione e nell'amministrazione.

I risultati che si ottengono sono quindi indipendenti dalle persone che utilizzano risorse informatiche, ed è possibile introdurre nel sistema informatico persone senza particolari requisiti di formazione specialistica, anche per la crescente facilità nell'uso delle apparecchiature informatiche.

In particolare i maggiori benefici e l'espansione più rilevante si è avuta nel livello operativo della piramide di Anthony, e in particolare nelle aree contabili e nell'automazione dei lavori d'ufficio.

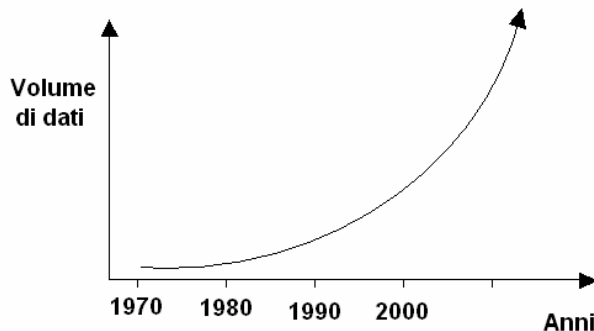
Esistono poi procedure e applicazioni riguardanti l'aspetto decisionale, come i prodotti software denominati DSS (Decision Support System), cioè programmi che consentono di raccogliere e rielaborare dati, anche provenienti da archivi e applicazioni diverse, che possono servire da supporto nelle decisioni aziendali.

In questo modo il sistema informatico si può sviluppare non solo in senso orizzontale nelle diverse aree, ma anche in senso verticale nei controlli.



LE BASI DI DATI AZIENDALI

Gli ultimi decenni hanno visto un grande incremento della qualità di informazioni e di dati memorizzati in formato elettronico. Questo accumulo di dati è avvenuto con tassi di crescita esplosivi. Si è stimato che la quantità di informazioni nel mondo raddoppia ogni 20 mesi e le dimensioni e il numero dei database aumentano ancora più velocemente. La disponibilità dei dati è favorita dall'uso di periferiche per l'acquisizione di dati, quali i POS (Point of Sale) negli esercizi commerciali, di terminali remoti e di connessione alle reti.



Dopo aver concentrato una così grande quantità di dati, il problema è stabilire quale valore si possa ottenere da queste risorse di informazione.

L'informazione è al centro delle operazioni di business e le persone che nell'impresa devono prendere le decisioni (decision makers), potrebbero usare i dati memorizzati per ottenere nuovi risultati nel mercato dell'azienda. **Il successo di un prodotto o di una strategia marketing dipendono infatti dalle informazioni che l'azienda dispone sulla propria clientela.**

I database management system DBMS facilitano l'accesso ai dati memorizzati, ma questo è solo una piccola parte di quello che si può ottenere dai dati.

I sistemi tradizionali di elaborazione delle transazioni on-line (OLTP, on-line transaction processing) consentono di immettere dati nei data base in modo veloce e con sicurezza ed efficienza, ma non sono adatti per produrre analisi significative.

Una società di assicurazione offre prodotti assicurativi insieme a molte altre società, un banca offre un mutuo come molti altri istituti di credito: la differenza che può determinare il successo di un prodotto è legata alle informazioni che l'azienda possiede sui clienti circa le motivazioni all'acquisto, le preferenze, la distribuzione geografica, le caratteristiche demografiche (come già visto a lezione).

L'analisi dei dati può fornire ulteriori conoscenze sul business andando al di là dei dati esplicitamente memorizzati per derivare conoscenze riguardo alla situazione dell'impresa e del mercato esterno: questo si chiama DATA MINING o scoperta di conoscenza estratta dalla miniera dei dati aziendali.

Il DATA MINING è l'estrazione non banale di informazione dai dati, implicita, sconosciuta in precedenza e potenzialmente utile; esso si realizza attraverso la ricerca delle associazioni tra i dati e degli indicatori globali che esistono nei grandi database, ma che sono nascosti all'interno della vastità dei dati.

Il risultato di questo lavoro fornisce informazioni e conoscenze che vengono messe a disposizione delle aree aziendali di supporto alle decisioni, previsioni e valutazioni.

Le applicazioni che vengono utilizzate in queste aree sono indicate con le sigle:

- DSS (Decision Support System), sistemi di supporto alle decisioni;
- OLAP (on-line Analytical Processing), elaborazioni di analisi on-line;
- EIS (Executive Information System), supporto alla direzione aziendale, per agevolare il controllo degli andamenti e degli indicatori di business.

Gli aspetti pratici del Data Mining possono essere:

- raggruppamenti di dati
- sommarizzazione di dati
- apprendimento di regole di classificazione
- ricerca di dipendenze tra i dati
- analisi di cambiamenti
- individuazione di anomalie
- analisi di comportamenti e di abitudini

IL DATA WAREHOUSE

Il Data Mining serve a vagliare grandi quantità di dati per produrre associazione tra i dati, per scoprire regole e fatti. Tutto questo lavoro diventa produttivo se i dati appropriati vengono raccolti e memorizzati in un Data Warehouse: un Data Warehouse può essere definito come il *database di tipo decisionale che diventa il deposito centralizzato dei dati e che viene interrogato per intervenire sul business dell'azienda; è il legame logico tra ciò che i manager vedono nelle loro applicazioni per il supporto delle decisioni e le attività operazionali dell'azienda.*

Le tecniche per i datawarehouse consentono di estrarre e trasformare i dati presenti nei database relazionali per caricarli nei database decisionali dei sistemi informativi direzionali, per l'analisi e il reporting degli utenti che devono prendere le decisioni: i responsabili aziendali possono in questo modo penetrare all'interno del vasto database dei dati operazionali dell'azienda, per determinare gli indicatori, rispondere all'andamento del mercato, facilitare gli sforzi di previsione e di pianificazione.

Ci sono quattro caratteristiche che descrivono un data warehouse:

1. **orientato ad un aspetto specifico**: i dati sono organizzati secondo un aspetto invece che un'applicazione; per esempio in una società di assicurazione un data



warehouse organizza i dati secondo i clienti o i premi invece che secondo i differenti prodotti assicurativi (auto, vita). I dati organizzati per argomento contengono solo le informazioni necessarie nei processi di supporto alle decisioni.

2. **integrato**: quando i dati risiedono in molte applicazioni separate nell'ambiente operativo, la codifica dei dati è spesso inconsistente. Per esempio, in una applicazione, il sesso delle persone è codificato con 'M' e 'F', in un'altra con '0' e '1'. Quando i dati passano dal sistema operativo al data warehouse, essi devono assumere un'uniformità nella codifica, per esempio trasformando il sesso in 'M' o 'F'.
3. **consolidato**: il data warehouse contiene i dati che sono precedenti da 5 a 10 anni, o più vecchi, che devono essere usati per confronti, tendenze e previsioni. Questi dati non possono essere modificati.
4. **non volatile**: i dati non possono essere modificati o cambiati in nessun modo una volta che sono entrati nel data warehouse, ma possono essere solo caricati e consultati.

La realizzazione di un data warehouse si compone di tre elementi fondamentali:

- ✓ strumenti per estrarre, trasformare e caricare dati provenienti dai database operazionali;
- ✓ archivi per memorizzare i dati in database decisionali;
- ✓ strumenti per accedere e operare l'analisi dei dati in archivio.

La seguente tabella illustra le differenze tra un database tradizionale e un data warehouse:

	DATABASE	DATA WAREHOUSE
Caratteristica	Operazionale	Decisionale
Obiettivo	Transazione giornaliere	Ritrovamento e analisi delle informazioni
Tipo di dati	Dati per gestire il business	Dati per analizzare il business
Stato dei dati	Modificabili	Storici



CAMPI DI APPLICAZIONE

Le applicazioni di data mining e di data warehouse sono potenzialmente importanti in vari campi per ottenere indicatori, controlli di andamenti, previsioni e pianificazioni, come si vede dai seguenti esempi:

Commercio e grande distribuzione	<ul style="list-style-type: none">● <i>indicatori negli acquisti dei clienti</i>● <i>associazioni tra le caratteristiche demografiche dei clienti</i>● <i>previsione di risposta a una campagna di pubblicità postale</i>● <i>analisi del contenuto del carrello della spesa</i>
---	---

Assicurazioni	<ul style="list-style-type: none">● <i>previsione sulle nuove polizze che verranno acquistate dai clienti</i>● <i>indicatori di comportamento dei clienti a rischio</i>● <i>individuazione di comportamenti fraudolenti</i>
----------------------	---

Trasporti	<ul style="list-style-type: none">● <i>percorsi ottimali per le consegne</i>● <i>caricamento degli automezzi in funzione degli itinerari</i>
------------------	---

Medicina	<ul style="list-style-type: none">● <i>comportamento dei pazienti per prevedere le visite ambulatoriali</i>● <i>terapie mediche di successo per differenti malattie</i>
-----------------	--



BANCHE	
	<ul style="list-style-type: none">● <i>indicatori di usi fraudolenti delle carte di credito</i>● <i>identificazione dei clienti fedeli</i>● <i>previsione su come i clienti cambieranno il tipo di carta di credito</i>● <i>modalità di utilizzo della carta di credito per gruppi di clienti</i>● <i>correlazioni nascoste tra differenti indicatori finanziari</i>● <i>identificazione delle regole del mercato attraverso i dati storici</i>

Per quanto concerne i prodotti commerciali per eseguire il Data Mining, si rimanda al seguente indirizzo web dal quale è possibile scaricare il documento “ **Stato dell’arte dei sistemi commerciali di Data Warehouse**” redatto da InterData:

www.dia.uniroma3.it/interdata/relaz98/MO.htm