

Università di Padova

Facoltà di Scienze Statistiche



Una proposta per l'organizzazione dei dati museali: il Data Warehouse

CORSO DI LAUREA IN STATISTICA E TECNOLOGIE INFORMATICHE

Relatore: Ch.mo Prof. Susi Dulli

Laureando: Renato Guarino

Anno Accademico 2003 - 2004

SOMMARIO

Prefazione	4
Introduzione	5
1 DataBase e DataWarehouse: due mondi differenti	7
1. Sistemi informativi e informatici	7
2. Evoluzione dei sistemi informatici	9
3. Le basi di dati	12
4. Passaggio di consegne	14
2 Il DataWarehouse	15
1. Generalità	15
2. Definizione di data warehouse	18
3. Applicazioni	19
4. Caratteristiche di base	23
5. Architettura dei data warehouse	27
6. Progettazione del data warehouse	30
6.1 Strategie di progettazione	31
6.2 Schema del DW	33
3 Scenari futuri	37
1. Introduzione	37
2. Dal Datawarehouse al Datawebhouse	38
2.1. Il Data Mining	40
3. Tendenze e sviluppi	42
Conclusioni	43
Glossario dei termini	46
Bibliografia	52

Prefazione

Questo lavoro è frutto di un'attività di stage offertomi dall'ufficio preposto della facoltà svolto presso il Settore dei Musei e monumenti civici d'arte di Verona con sede effettiva nel Museo di Castelvechio.

Questo stage aveva la finalità di effettuare un'analisi di fattibilità per la (eventuale) successiva progettazione e implementazione di un Data Warehouse per la gestione del settore in questione.

Alla fine del periodo di stage si è giunti a presentare questa analisi; nel frattempo si è scoperto che l'Istituto Centrale per la Catalogazione e la Documentazione (ICCD) stava a sua volta facendo, a livello nazionale, un'analisi dei requisiti per lo stesso tipo di progetto, a conferma del fatto che il Data Warehouse sta ottenendo sempre più consensi (da parte di utenti e specialisti) e una maggior diffusione, anche in ambiti diversi da quelli in cui si è inizialmente sviluppato.

Nel capitolo 1 è stata fatta una breve panoramica sui DataBase e un confronto con i Data Warehouse, poi, nel capitolo 2 si analizzerà la struttura e gli usi possibili di un Data Warehouse all'interno di un generico museo; successivamente, nel capitolo 3, si prenderanno in considerazione gli sviluppi futuri con le tecnologie connesse.

Introduzione

Quando si pensa ad un museo, la prima sensazione è quella di un'attività passiva, che si svolge lentamente con ritmi molto blandi e senza particolari sussulti. Questo è vero se di un museo si considera solo il lato espositivo e conservativo.

Anche gli attuali dizionari considerano il museo in questo modo: "luogo in cui sono raccolti, ordinati e custoditi oggetti d'interesse storico, artistico, scientifico, etnico e simili per il bene delle generazioni contemporanee e quelle future." Tale definizione richiama il concetto di museo "pubblico", nato intorno al Settecento e che identifica il museo come una collezione privata d'oggetti d'arte che, per volontà del proprietario, è aperta al pubblico.

Attualmente però le cose non stanno più in questo modo: infatti, una struttura museale è organizzata per svolgere una pluralità di funzioni, che mirano ad erogare attività e servizi culturali di tipo composito.

Le caratteristiche dei servizi erogati dai musei si sono modificate parallelamente all'evoluzione subita dalle finalità cui i musei sono stati destinati. Infatti, la funzione conservativa, attività tradizionale dei musei, si è progressivamente modificata: si segnala una drastica riduzione dei consumi culturali, a causa di un interesse crescente per attività rivolte all'intrattenimento, che hanno usufruito maggiormente dei progressi tecnologici.

Un argomento interessante riguarda l'incrocio di due mondi apparentemente opposti: il museo inteso come luogo di cultura dove si raccolgono e si conservano oggetti d'arte e il mondo delle nuove tecnologie.

L'applicazione dei sistemi informatici in un museo è molto varia e interessa diversi ambiti:

- il controllo, la tutela e la valorizzazione del patrimonio;
- la gestione nell'ambito economico e finanziario;
- l'esposizione delle opere, dove la sistemazione di supporti informatici nei diversi luoghi di un museo consente una comprensione più immediata della collezione con la possibilità di trovare percorsi alternativi e adatti a qualsiasi esigenza,
- l'informazione ai visitatori,
- il portale, che consente on line di prenotare biglietti, organizzare tour culturali, scambiare idee con gli addetti ai lavori, noleggiare opere d'arte, ecc.

Partendo da queste considerazioni, nel presente lavoro saranno delineate le premesse di un Data Warehouse per i musei che vorrebbe servire come punto di partenza per la realizzazione di progetti informatici, possibili fondamenta per la creazione di un sistema di supporto alle decisioni e per analisi "aziendali", volti ad una migliore gestione del Museo da parte del personale amministrativo e manageriale e che inoltre permettano una più vantaggiosa fruizione del patrimonio artistico da parte dei visitatori: un patrimonio ancora sottovalutato e poco sfruttato.

Capitolo I

DataBase e DataWarehouse: due mondi differenti

1. Sistemi informativi e informatici

Ogni organizzazione, per il proprio funzionamento, ha bisogno di disporre di informazioni, che ne costituiscono una delle principali risorse. Perciò ogni organizzazione contiene al proprio interno un sottosistema che si occupa di gestire queste risorse. Tale sottosistema è detto "sistema informativo dell'organizzazione" del quale si propone la seguente definizione:

Il sistema informativo di un'organizzazione è una combinazione di risorse, umane e materiali, e di procedure organizzate per la raccolta, l'archiviazione, l'elaborazione e lo scambio delle informazioni necessarie alle attività operative, alle attività di gestione e alle attività di programmazione, controllo e valutazione dell'organizzazione.

Con il termine *sistema* si evidenzia il fatto che esiste un insieme organizzato di elementi, di natura diversa, che interagiscono in modo coordinato, mentre con *informativo* si precisa che tutto ciò è finalizzato alla gestione dell'informazione e quindi le interazioni che preme evidenziare sono quelle dovute a scambi di informazione.

Le informazioni di un'organizzazione, una volta ridotte a dati mediante un processo di interpretazione, quantificazione e formalizzazione, possono essere trattate automaticamente con gli elaboratori elettronici. La riduzione dei costi della tecnologia informatica ha diffuso largamente questa possibilità, rendendo più accurate e rapide le procedure e potenziando i modi di elaborare le informazioni.

Si parla in questo caso di *sistema informatico* che si definisce come segue:

Un sistema informatico è l'insieme degli strumenti informatici impiegati per il trattamento automatico delle informazioni di una organizzazione al fine di agevolare le funzioni del suo sistema informativo.

Il sistema informatico prevede 4 componenti principali:

- il software e l'hardware di base
- una base informativa che contiene una rappresentazione delle informazioni memorizzate
- uno schema che viene usato dal sistema per garantire un uso corretto della base informativa.
- i programmi applicativi che forniscono i servizi agli utenti eseguendo un certo insieme di operazioni sulla base informativa.

L'architettura dei sistemi informatici per la gestione di informazioni è cambiata molto negli ultimi 30 anni con l'evolvere della tecnologia degli elaboratori elettronici e degli strumenti per la gestione di dati permanenti.

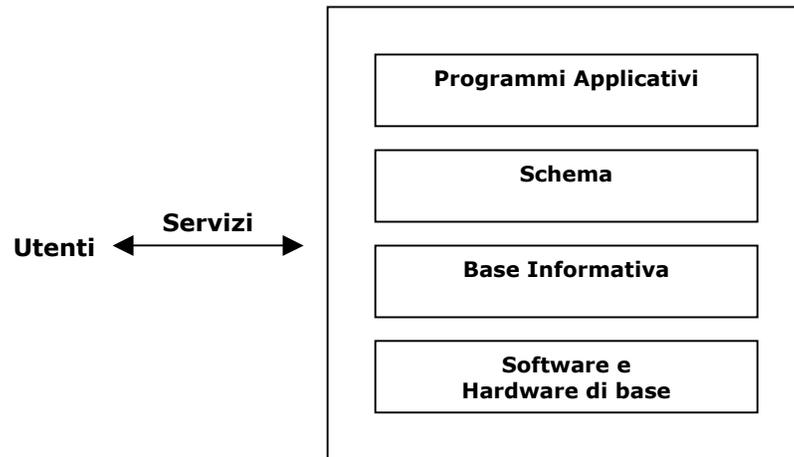


Figura 1. Sistema Informatico

2. Evoluzione dei sistemi informatici

Agli inizi degli anni '60 l'esigenza più sentita era una strumentazione per migliorare l'efficienza e la produttività di alcune parti dei processi operativi. Si ebbe così una diffusione della tecnologia informatica per applicazioni settoriali, soprattutto nell'amministrazione, con l'obiettivo di automatizzare quelle attività che richiedono l'elaborazione sistematica e ripetitiva di grandi quantità di dati. Questo modo di operare comportava indubbi vantaggi, ma presentava degli inconvenienti. I vantaggi riguardavano soprattutto la correttezza dei risultati, la riduzione dei costi e la maggior produttività settoriale. L'inconveniente principale riguardava invece la scarsa integrazione dei dati comuni ai diversi settori, con duplicazione di dati che comportavano il rischio di copie incoerenti, e una limitata possibilità di correlare dati settoriali per generare informazioni di interesse generale per l'organizzazione.

La tecnologia informatica impiegata in quegli anni era basata sull'unico strumento ancora oggi generalmente offerto dai sistemi di elaborazione per lo sviluppo di applicazioni: un linguaggio di programmazione che utilizza archivi (file), ovvero insiemi di registrazioni (record) che persistono all'esecuzione dei programmi che li creano e ne fanno uso.

A partire dagli inizi degli anni '70 il progresso della tecnologia ha reso disponibili nuovi strumenti informatici, i sistemi per la gestione di basi di dati che, rendendo possibile una gestione integrata dei dati, hanno consentito la realizzazione di sistemi informatici di supporto ad ogni livello delle organizzazioni. Inizialmente erano di tipo generalizzato, ovvero la base di dati era gestita da un unico elaboratore.

Agli inizi degli anni '90 invece l'evoluzione della tecnologia ha reso possibile anche la realizzazione di sistemi per basi di dati distribuite, su rete locale o geografica, che consentono di avere una visione unica dei dati, indipendentemente da dove essi siano fisicamente memorizzati.

Nella tabella 1 sono riportati i passaggi più significativi dell'evoluzione della tecnologia delle basi di dati.

Periodo	Avvenimento
1960	Sistemi gerarchici (IMS e System 2000)
1970	Sistemi reticolari secondo la proposta CODASYL. E.F. Codd propone il modello relazionale
1980	Introduzione dei sistemi relazionali commerciali.
Fine anni '80	Introduzione dei Decision Support System (DataWarehouse)
1990	Sistemi relazionali distribuiti. Architetture Client/Server e parallele. Sistemi ad oggetti e relazionali ad oggetti.
Dal 1995	Integrazione con Internet

Tabella 1. Prospetto cronologico dell'evoluzione dei sistemi per basi di dati

3. Le basi di dati

Una base di dati è una collezione di dati, utilizzati per rappresentare le informazioni di interesse per un sistema informativo.

In questo lavoro verrà considerata con un'accezione molto più specifica: si prenderanno in considerazione solo basi di dati gestite da un elaboratore elettronico.

Una base di dati informatica deve garantire alcuni requisiti basilari:

- I dati sono strutturati, cioè hanno un formato predefinito, e il numero dei tipi di dati presenti è relativamente piccolo rispetto al numero degli esemplari di ognuno di essi;
- I dati sono raggruppati in insiemi omogenei, in relazione fra loro;
- I dati sono molti (da milioni a miliardi di caratteri) e sono memorizzati in una memoria permanente;
- I dati sono una risorsa condivisa e disponibile per usi molteplici che spesso hanno un'importanza relativa variabile nel tempo;
- I dati sono protetti da usi non autorizzati e da malfunzionamenti hardware e software.

Queste caratteristiche delle basi di dati sono garantite da un Data Base Management System (DBMS), che ha il controllo dei dati e li rende accessibili agli utenti autorizzati.

Un DBMS è quindi un sistema, centralizzato o distribuito, che consente di:

- a) Definire schemi di basi di dati
- b) Scegliere le strutture dati per la memorizzazione e l'accesso ai dati
- c) Memorizzare, recuperare e modificare i dati.

In figura 2 è mostrata una versione semplificata della struttura di un DBMS.

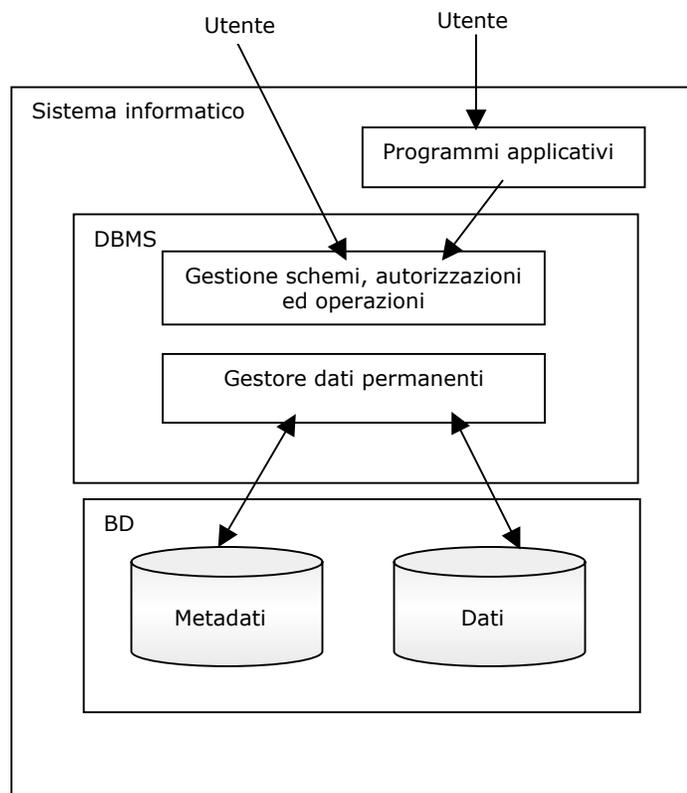


Figura 2

4. Passaggio di consegne

Dall'informatica, ci si aspetta sempre più una tecnologia in grado di rendere disponibili informazioni significative ed aggiornate su quelli che sono considerati i fattori critici di successo per il raggiungimento degli obiettivi aziendali.

Le informazioni così individuate dovranno poi essere trasformate dai responsabili in azioni, strategie, piani opportuni, grazie all'arricchimento indotto dalle esperienze e dalle competenze specifiche.

La diffusione del PC ha portato sulle scrivanie dei manager nuovi strumenti e nuove possibilità di analisi di dati aziendali. Inoltre la successiva comparsa di interfacce e tool per l'estrazione di dati dai sistemi legacy ha ridotto la distanza tra l'elaborazione e la fruizione dell'informazione.

Il DW, nato in risposta a reali esigenze, integra diverse tecniche e fa da tramite tra i dati (di per sé poco esplicativi) e gli utenti (non necessariamente esperti di informatica), tanto che gli utenti stessi sono stati i primi a rendersi conto delle reali, grandi potenzialità di questa nuova tecnologia.

Passare dal dato all'informazione per arrivare alla vera conoscenza è il sogno riposto nel cassetto di qualsiasi manager impegnato in decisioni strategiche per la propria organizzazione. Il DataWarehouse, fornisce un'articolata e intelligente soluzione in tal senso.

Capitolo II

Il DataWarehouse

1. Generalità

La tecnologia delle basi di dati è finalizzata alla gestione efficiente e affidabile di dati "in linea" (On Line Transaction Processing, OLTP); questi dati però potrebbero rivelarsi utili non solo per la gestione, ma anche per la pianificazione e il supporto alle decisioni. In pratica, i dati del presente e del passato possono consentire un'attività di analisi essenziale nella programmazione delle attività future di un'impresa.

Con l'inizio degli anni '90, parallelamente allo sviluppo delle reti e dei prodotti per la distribuzione dei dati, si sono imposte nuove architetture, caratterizzate dalla separazione degli ambienti: a fianco dei sistemi per OLTP si sono sviluppati sistemi dedicati all'elaborazione e analisi dei dati, cioè On Line Analytical Processing (OLAP). Questa sigla mette in evidenza che l'analisi dei dati avviene "in linea", tramite strumenti interattivi. L'elemento principale dell'architettura OLAP, è il **Data Warehouse** (come il DBMS lo è per OLTP).

Alla base del Data Warehouse (da intendersi in prima istanza come un Data Base di grandi dimensioni) c'è il modello dimensionale che serve per rappresentare l'oggetto d'osservazione.

Se ad esempio si è interessati ad avere informazioni di sintesi (reports) sui visitatori del museo, possiamo rappresentare i visitatori a seconda della sede e del tempo. Per riuscire a visualizzare questa struttura logica si disegna un cubo che rappresenta il Fatto Visitatori del museo, qualificato dalle dimensioni Sede, Biglietti e Tempo: si veda la figura 3. In particolare si osservi l'aggregazione dei dati sulla dimensione tempo tale da avere nel primo cubo una rappresentazione per mese, nel secondo per trimestre e infine nell'ultimo, per anno.

Il cubo può avere più di tre dimensioni (da cui il nome di cubo multidimensionale), ma il limite al numero delle dimensioni utilizzabili è dovuto alla comprensibilità della struttura.

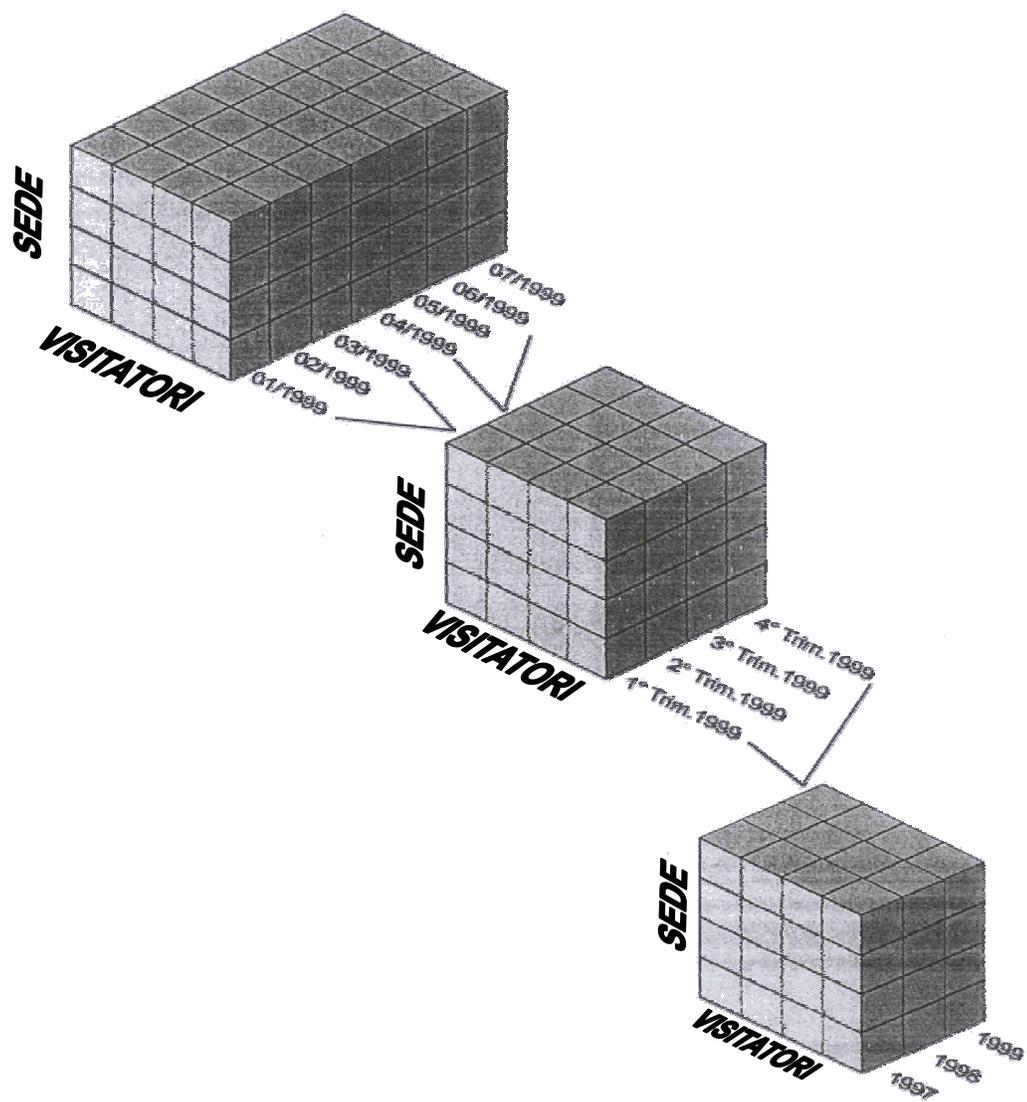


Figura 3

2. Definizione di data warehouse

Non è facile definire univocamente un DW, dato il proliferare di interpretazioni su cosa realmente s'intenda: un processo, una tecnica, oppure un sistema. In questo lavoro si farà riferimento principalmente alla definizione data da William Inmon: "Un insieme di dati orientati al soggetto, integrato, variante rispetto al tempo, non volatile costruito per supportare il processo decisionale". Questa definizione focalizza l'attenzione su alcuni aspetti peculiari del DW:

- Un DW è orientato al soggetto (gli utenti (fruitori) del museo, i servizi offerti, ...), cioè esclude tutti i dati che non saranno utilizzati nel processo di pianificazione delle strategie, raggruppandoli per aree o temi d'interesse e finalizzandoli a chi li utilizza più che a chi li genera;
- Un DW è integrato, ossia consistente rispetto ad un sistema concettuale globale dei dati e rispetto alle unità di misura e alle strutture di codifica. In pratica occorre che il significato di ciascun dato sia chiaro, concordato a livello aziendale e definito con precisione sia in termini applicativi che informatici;
- I dati di un DW hanno un orizzonte temporale molto ampio (5, 10 anni o più), per implementare il quale è necessario disporre di informazioni che contengano la dimensione temporale;
- Un DW è non volatile: il dato viene caricato e letto off line; una volta memorizzato correttamente può essere rilevato, ma non modificato dall'utente.

Una definizione più operativa di Data Warehouse è quella data da Ralph Kimball:

“A data warehouse is a copy of transaction data specifically structured for querying and reporting¹”.

3. Applicazioni

La tecnologia del Data Warehousing, nata per analizzare grandi quantità di dati e trasformarli in informazioni, può essere applicata anche ai musei: infatti, il personale direttivo deve continuamente prendere decisioni su come utilizzare le risorse disponibili per raggiungere gli obiettivi prefissati e come valutarne i risultati.

Queste operazioni sono quelle che consentono ad un museo di coniugare un'efficace impostazione strategica con un uso efficiente delle risorse a disposizione. La pianificazione di strategie può essere facilitata proprio con l'introduzione di un sistema di Data Warehouse (nel seguito denominata DW).

In altre parole, poiché l'obiettivo di un progetto DW è quello di consentire agli utenti la gestione più efficace ed efficiente l'attività dell'impresa, il primo passo nella sua costruzione è di individuare il tipo di dati che dovranno entrare a farvi parte e successivamente le tecniche di analisi più appropriate ai fini del supporto decisionale.

¹ Un Data Warehouse è una copia dei dati operativi specificamente strutturati per la richiesta d'informazioni e l'analisi
Ralph Kimball “The Data Warehouse Toolkit” Addison & Wesley

Una volta inseriti i dati, il DW permette di ottenere attraverso un'analisi incrociata l'ottimizzazione del loro utilizzo a vantaggio dell'impresa e di ridurre al minimo i margini di errore, tenendo sempre presente che il vantaggio principale ottenibile è sempre legato dall'abilità dell'utente di interrogare i dati in modo innovativo.

La trasformazione dei vari dati in informazioni congruenti può potenziare e/o ottimizzare le attività del museo, per es. attraverso:

- a) il rapporto con gli sponsor (del passato, del presente e possibili)
- b) l'offerta delle esposizioni
- c) la migliore gestione contabile - collegata ad a), b) e d)
- d) la miglior gestione del personale - collegata ad a), b) e c).

Esempio 1

Se al responsabile di un museo sono messi a disposizione in modo congruente le seguenti informazioni:

- a) c'è un revival delle opere del 1700
- b) nel museo ci sono opere di questo periodo:
 - già esposte
 - disponibili in magazzino
 - esistenti in altri musei o presso privati
- c) mese/i di maggior frequenza dei visitatori
- d) periodi di facile reperibilità di personale suppletivo
- e) studi già effettuati sull'argomento
- f) mostre, seminari o altro in corso, in preparazione, o terminati in altri musei
- g) somma disponibile per la manifestazione

h) gli sponsor che potrebbero intervenire

dall'analisi di questi dati il museo potrà decidere se:

- organizzare un'esposizione sulle opere del '700
- organizzare seminari per ricercatori, studenti, studiosi sull'argomento
- valutare il periodo ottimale
- valutare la possibilità di utilizzare la somma in bilancio o ricorrere a uno o più sponsor
- ecc.

Ecco che l'attività del museo può trarre vantaggi:

- economici
- d'immagine
- di efficienza

Esempio 2

Obiettivo: incrementare il numero di visitatori nei singoli musei, aumentare i ricavi e allargare la conoscenza ai musei meno conosciuti.

Dati a disposizione per la valutazione:

- prezzo attuale d'entrata ai singoli musei
- stima del numero visitatori per ogni singolo museo
- tipo di turismo (giornaliero, settimanale, festivo, ecc.)
- ubicazione musei e monumenti (distanze)
- tipo di museo (archeologico, arte medievale, rinascimentale , ecc)
- tipologia del visitatore medio per singolo museo

Mediante un'analisi realizzata tramite il DW si potrebbe stabilire:

- valutazione economica nella creazione di un biglietto unico
- la durata della validità del biglietto (giornaliero, settimanale, mensile, ecc)
- il formato del biglietto: unico, con tagliandi, a cartolina, ecc.
- tipo del biglietto: personale o utilizzabile da più persone.
- possibilità di scelta fra i musei e monumenti
- itinerari possibili da proporre, in base alle preferenze dei turisti
- possibili accordi con infrastrutture alberghiere, agenzie viaggi, trasporti urbani, FS, parcheggi, ecc.
- promozione e sponsorizzazione della iniziativa
- valutazione di eventuali convenzioni con musei di altra gestione (statali, curiali, privati).

Dopo aver valutato le varie possibilità, si potrebbe decidere di istituire un biglietto unico che permetta di visitare alcuni musei e monumenti in base alla loro vicinanza e ai turisti che li visitano. Si potrebbe stabilire un accordo con le FS e con i trasporti urbani. Ai visitatori che usufruissero di un mezzo proprio mettere a disposizione un parcheggio pubblico gratuito. Per aumentare l'afflusso dei visitatori si potrebbe promuovere una campagna pubblicitaria da concordare con eventuali sponsor. Il biglietto potrebbe essere una cartolina con le immagini dei musei e monumenti che potrebbe essere spedita dopo aver effettuato le visite.

Ovviamente questi sono solo degli esempi per dimostrare i possibili usi del DW applicato ai musei.

Per mantenere il proprio valore nel tempo, il DW deve adattarsi continuamente alle necessità dell'utente. Nuovi dati devono essere continuamente acquisiti e le informazioni aggregate e disaggregate per supportare nuove domande, strumenti, analisi.

Per cui successivamente, presa la decisione e resala operativa, si valuteranno i risultati ottenuti confrontandoli con i precedenti dati, e si potrà in base a nuove informazioni o a nuove situazioni apportare le necessarie modifiche per l'ottimizzazione continua dei risultati.

4. Caratteristiche di base

I dati contenuti nel DW dovranno poter essere sintetizzati in modo da fornire una visione globale caratterizzata da informazioni aggregate, le quali diventano dettagliate, al punto da risultare atomiche ai livelli più bassi, in modo da spiegare aspetti particolarmente critici durante l'analisi (si veda figura 4).

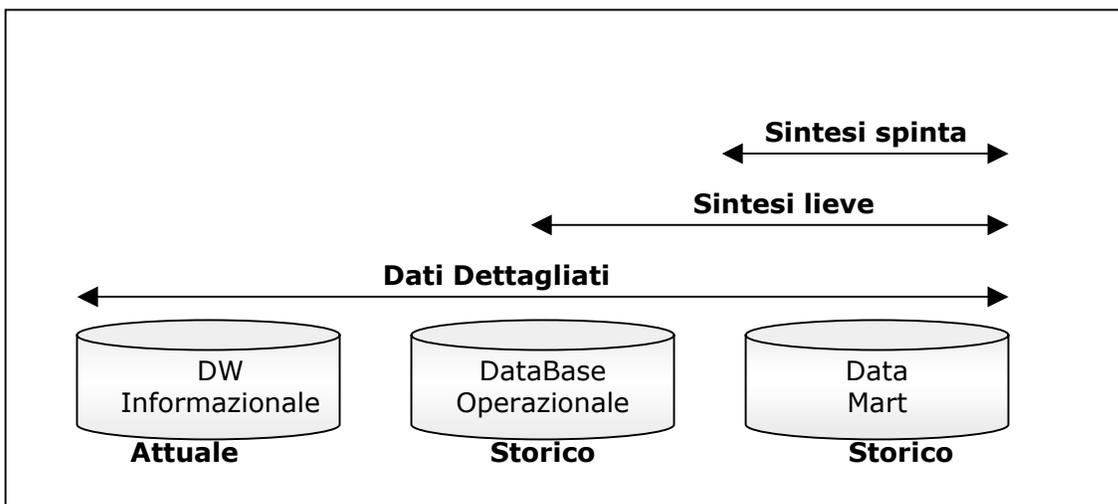


Figura 4

In un'ottica tecnica la disponibilità di dati dettagliati offre il vantaggio di rendere necessaria una sola estrazione dei dati dall'ambiente operativo, che quindi non vedrebbe penalizzate le proprie prestazioni, poiché la successiva attività di sintesi sui database del **Warehouse** (e l'eventuale distribuzione verso i Data Mart dipartimentali) impegnerebbe solo i sistemi informazionali. D'altra parte è evidente che alcuni dati di sintesi (specie quelli di più largo uso) se preaggregati consentono di elevare le prestazioni del DW.

La storicizzazione dei dati è un altro dei concetti cardine del DW. E' proprio la necessità di storicizzare una rilevante mole di dati che spinge ad operare una sintesi di questi su più livelli nel sistema informazionale, senza perdere completamente traccia del dato in dettaglio, pur rinunciando alla sua attualità. Ciò conduce ad un'architettura dei dati organizzata su più piani.

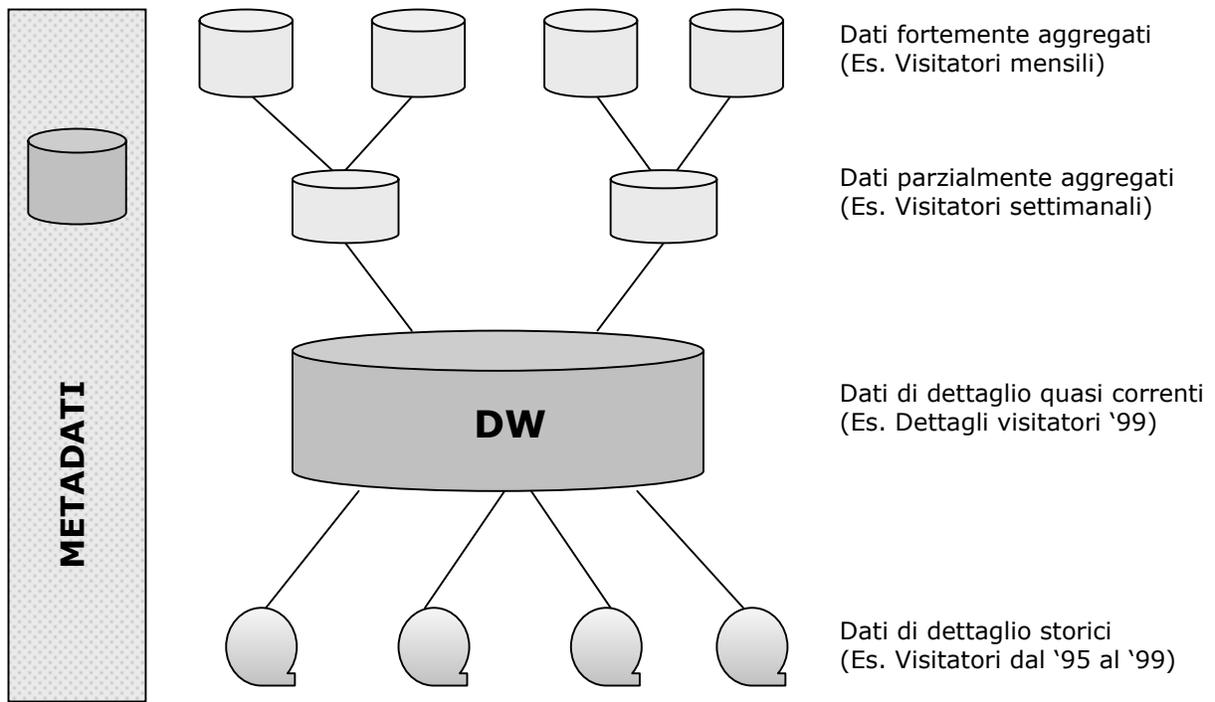


Figura 5

Nello strato più alto troviamo informazioni ad un livello di sintesi molto elevato (tanto da giustificare la loro allocazione su DW di tipo tematico, i **Data Mart**), che vengono dettagliate man mano che si scende nella struttura fino ad essere storicizzati al livello più basso.

L'utente dovrebbe essere in grado di accedere ed operare sui dati desiderati anche senza sapere dove essi risiedono, in quale forma siano stati memorizzati e quali tool provvedano alla loro consegna.

Evidentemente, a causa delle dimensioni in gioco, non tutte le informazioni possono essere tenute on line e tipicamente, i dati storici dettagliati saranno archiviati off line: quindi le risposte alle interrogazioni che riguardano i livelli di sintesi più alti saranno

senz'altro interattive, mentre analisi che si spingono fino al reperimento di dati storici archiviati determineranno risposte differite; ai livelli intermedi la scelta sarà legata alle caratteristiche dei sistemi e dei DBMS che ospitano e gestiscono il DW.

In ogni caso nel rispetto della filosofia stessa del Data Warehousing, la demarcazione tra dati on line e off line non deve ostacolare la possibilità dell'utente di navigare tra i percorsi indicati: in altri termini il sistema di Data Warehousing deve assicurare attraverso le funzioni di accesso ai dati la possibilità, a partire da un percorso intrapreso, di approfondire il grado di indagine sulla dimensione, in maniera diretta o differita secondo la disponibilità del dato dettagliato, o di regredire ad un più elevato livello di sintesi.

Per completare il quadro progettuale dell'architettura dati che caratterizza la soluzione di Data Warehousing, occorre individuare una tecnica che consenta di definire i contenuti del sistema, permettendo di includere all'interno del DW l'intero orizzonte dell'attività che si sta considerando.

La necessità di accedere ai dati storici è uno degli elementi caratteristici di un progetto di Data Warehousing poiché questo tipo di dati gioca un ruolo primario nell'alimentare tutti quei sistemi usati per analisi di tendenza che sono determinanti nel supporto decisionale.

5. Architettura dei Data Warehouse

L'architettura generale di un DW è mostrata nella figura seguente:

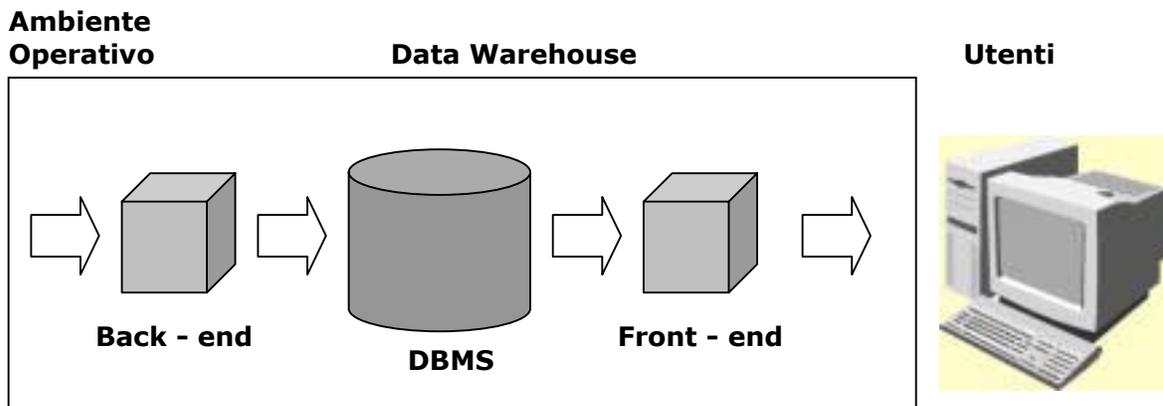


Figura 6

Vi si individuano tre componenti principali:

- il back-end, che gestisce il processo di estrazione, filtraggio, integrazione e trasformazione dei dati prelevati da vari ambienti operativi; esso attiva inoltre il processo di caricamento e di indicizzazione dell'informazione all'interno della base dati.
- Il sistema di gestione dei dati (DBMS) su cui risiedono i dati (e i meta dati²) del DW.
- Il front-end, l'interfaccia grafica che permette all'utente finale di interagire con i dati.

² Informazioni che descrivono il significato e la struttura dei dati aziendali, come sono creati, come vi si accede e come possono essere usati. In questa definizione troviamo diverse tipologie di meta dati: informazioni sui dati di origine, su come accedervi e trasformarli, sulla struttura del DW, su come utilizzare i dati da parte dell'utente.

Fermi restando i concetti finora espressi, la realizzazione di un sistema di Data Warehousing comporta diverse scelte progettuali. Qui verrà presentata l'architettura DW più completa ed esaminati i relativi vantaggi e svantaggi.

Presentiamo qui di seguito un'estensione della architettura generale di un DW, orientata al raggiungimento di un alto grado di flessibilità a costo di una ridondanza dei dati memorizzati nei vari livelli.

Come si vede in fig. 7, il DW è composto da database OLTP, datamart, OLAP-Server e Client.

I dati contenuti nei database subiscono delle operazioni di estrazione, trasposto e trasformazione per essere immessi nei datamart e nel DW. In ogni architettura vi possono essere diversi datamart ma un solo DW. I data mart condividono con il DW la stessa struttura, ma sono limitati nel loro scopo a fornire informazioni utili a una singola divisione aziendale (nel caso del museo i data mart sono finalizzati ai diversi reparti museali).

L'OLAP-Server permette di ottimizzare l'utilizzo delle risorse hardware in due modi distinti: da un lato trattando le operazioni a più alta richiesta di potenza elaborativa; dall'altro riduce i dati che passano dal server ai client, i computer utilizzati dal personale amministrativo.

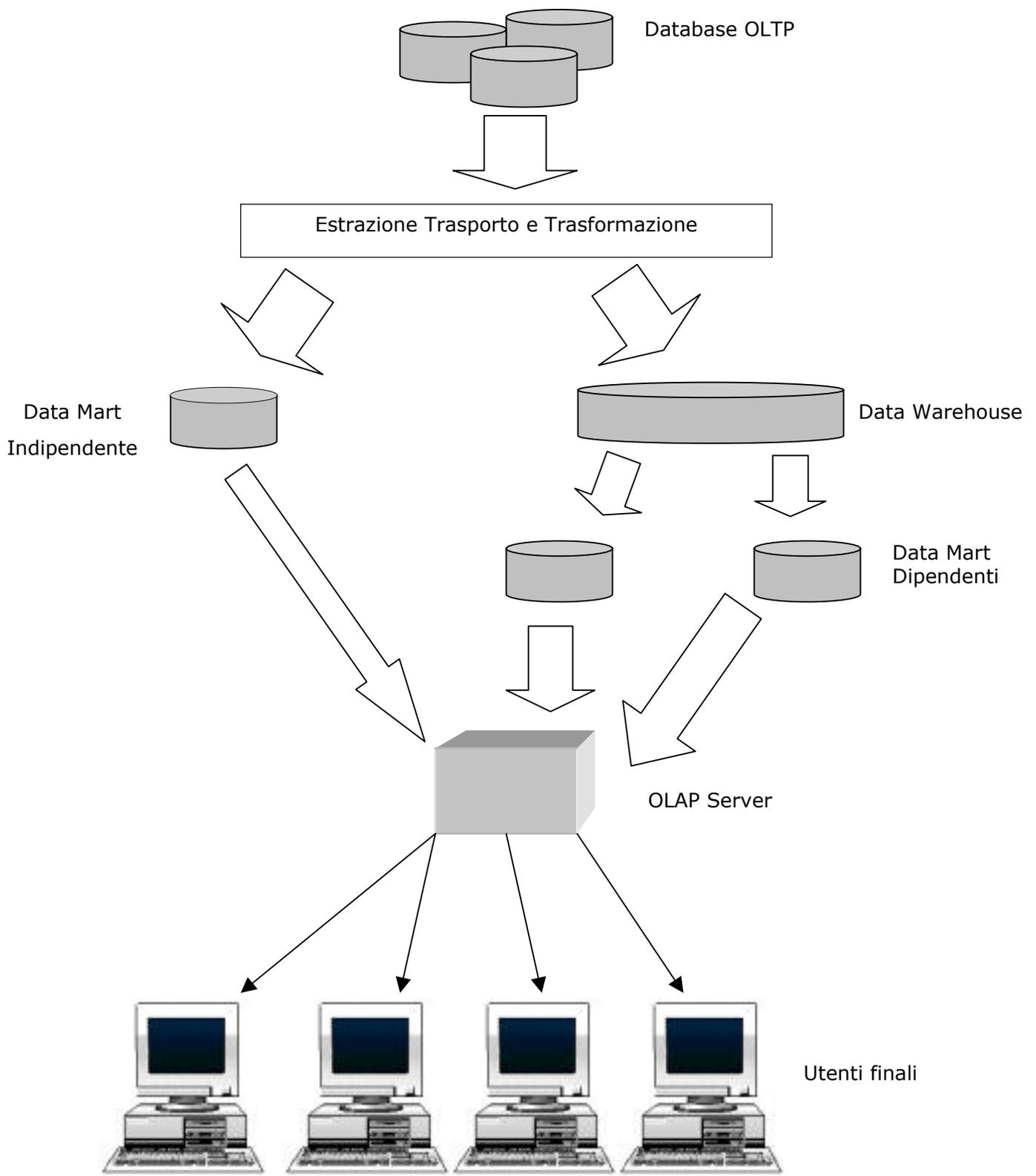


Figura 7: Architettura a più livelli (Multi - Tier)

Vista nell'ottica dei musei l'architettura mostrata precedentemente può essere interpretata come segue: partendo dal presupposto che siano già esistenti dei data base e/o dei sistemi legacy (antiquati), si potrebbe creare un DW per mezzo del quale migliorare il supporto decisionale del personale amministrativo e direttivo. Per poter realizzare un tale sistema, i dati devono essere in un certo qual modo "omogeneizzati" cioè trasformati in maniera tale da poter essere caricati nel DW e/o in un DM. Una volta effettuate tali operazioni i dati potranno essere riorganizzati e sintetizzati (si veda figura 3) per poter essere analizzati dal Server OLAP. Per esempio si potrebbero analizzare l'afflusso di visitatori in base al tempo o alla sede; naturalmente queste sono solo due possibilità: a seconda di cosa si vuole analizzare e capire si possono creare MDDB³ (si veda figura 3) con dimensioni differenti.

6. Progettazione del Data Warehouse

Ci sono due aspetti principali nella costruzione di un DW: la progettazione dell'interfaccia a partire dai sistemi operazionali e il progetto del DW in sé. In un certo senso, parlare di progettazione non descrive in modo preciso quello che avviene nel processo di realizzazione che procede in maniera euristica.

Infatti dapprima viene popolata una porzione del DW per essere utilizzata e verificata dall'analista. In seguito, sulla base delle indicazioni di ritorno da parte dell'utilizzatore finale, i dati sono modificati ed altri vengono aggiunti.

³Multi Dimensional Data Base

Il ciclo di retroazione continua lungo tutta la vita del DW. Quindi il DW non può essere costruito con il consueto modello di sviluppo guidato dai requisiti, tuttavia non bisogna commettere l'errore di pensare che non serva del tutto definire dei requisiti di partenza.

6.1 Strategie di progettazione

La scelta della strategia dovrà tener conto di una serie di fattori legati indissolubilmente alla realtà aziendale, in questo caso al tipo di museo.

La strategia Top-Down è stata storicamente la più utilizzata nell'implementazione di grandi DW: il "dipartimento Sistemi Informativi" è responsabile della catalogazione di tutte le informazioni disponibili e di quelle richieste dai vari dipartimenti.

Viene progettato un Modello (di business del Museo) della struttura organizzativa del Museo, al quale il DW corrisponderà e al quale verrà fatto riferimento per l'interpretazione delle informazioni disponibili. Questo fa in modo che qualunque decisione venga presa partendo da una base di conoscenza comune, condivisa e più estesa e significativa di quella normalmente in possesso dei singoli "dipartimenti".

I vantaggi principali della strategia top-down risiedono nel fatto che i dati vengono estratti, trasformati e caricati una sola volta, indipendentemente dal tipo di utilizzo e dal numero di dipartimenti che ne fanno uso, e nel fatto che l'architettura multi - tier permette di utilizzare il DW come fonti di dati comuni per più data mart dipartimentali.

Tuttavia gli svantaggi non sono da sottovalutare: il progetto e la costruzione di una simile struttura hanno, infatti, bisogno di notevoli sforzi in termini di modellazione, implementazione, istruzione e assistenza agli utilizzatori finali.

La strategia bottom up *si articola in due momenti: l'implementazione di una serie di data mart indipendenti, e successivamente la fusione dei singoli data mart in una visione globale, che corrisponde al DW.*

L'idea alla base di questa strategia è di avvicinare la costruzione del DW ai suoi fruitori finali.

Gli aspetti positivi di questa strategia, rispetto alla top down, si possono riassumere come segue:

- relativa facilità di implementazione
- tempo ridotti per ottenere i primi risultati parziali
- ritorno economico maggiormente controllabile
- maggiore semplicità per il fruitore finale che ha a disposizione tutte le informazioni di cui ha bisogno;
- minori requisiti iniziali in termini di hardware.

Naturalmente anche questa metodologia porta con sé alcuni punti deboli:

- Aumenta la ridondanza dei dati, in quanto più data mart indipendenti conservano più copie degli stessi dati.
- Aumenta il carico per il sistema OLTP (i data base tradizionali) per il moltiplicarsi di estrazioni degli stessi dati.
- Aumenta il pericolo di una proliferazione incontrollata di data mart che può portare a problemi di integrazione nel futuro DW.

6.2 Schema del DW

Anche nel caso di un Museo, il successo di un sistema di DW sta nel determinare un disegno globale dell'impresa, salvo poi, per ragioni tecniche, economiche e anche temporali, implementare dei sottoinsiemi, i Data Mart (DM), i quali risulteranno integrabili in qualunque successivo momento.

I dati di un DM sono organizzati secondo una semplice struttura, detta schema multidimensionale o più semplicemente schema a stella.

Il primo termine mette in luce la presenza di molteplici dimensioni di analisi; il secondo la struttura "stellare" dello schema, una volta che esso venga interpretato con il modello Entità-Relazioni classico.

La principale caratteristica dello schema a stella è la sua struttura regolare e indipendente dal problema considerato.

In ambiente informazionale è indispensabile ristrutturare i dati e configurare correttamente il sistema in modo da soddisfare le necessità degli utenti finali ai quali ci si rivolge. In tal modo si consente al DBMS di elaborare le query (le interrogazioni) che si rivolgono al sistema, tenendo bene a mente che il segreto del successo verso tale comunità è molto spesso la semplicità della struttura dati con la quale questa deve interfacciarsi. Il modello dimensionale va proprio incontro a questo obiettivo di semplicità.

La figura 8 mostra un semplice esempio di schema a stella, al cui centro è stata posta la tabella dei fatti (fact table), mentre le entità satelliti più piccole sono dette tabelle delle dimensioni (dimension table). Al contrario degli schemi Entità-Relazione tradizionali, lo schema a stella è evidentemente asimmetrico: c'è una tabella

dominante di grandi dimensioni al centro del diagramma, ed è l'unica a possedere collegamenti multipli, che si concretizzano attraverso identificatori univoci (campi chiavi) in essa contenuti, con le altre tabelle; queste ultime hanno tutte un solo collegamento (join) alla tabella centrale, con lo scopo di minimizzare il numero di join richiesti per ciascuna query.

A volte durante la fase di progettazione del DW, non sempre è chiaro se un campo estratto dal Data Base operativa sia un attributo piuttosto che un fatto; spesso il dubbio può essere risolto osservando se il campo ha la caratteristica di variare il proprio valore con continuità (o meglio, ogni volta che viene rilevato), proponendosi come una potenziale misura dell'attività da inserire nella fact table; oppure se è più ragionevole considerarlo come la descrizione di qualcosa che è più o meno costante e quindi inserirlo come attributo dimensionale. Un ruolo fondamentale degli attributi delle tabelle dimensionali è quello di essere usati all'interno dei vincoli nelle interrogazioni sul DW o come intestazioni delle colonne negli output per gli utenti finali.

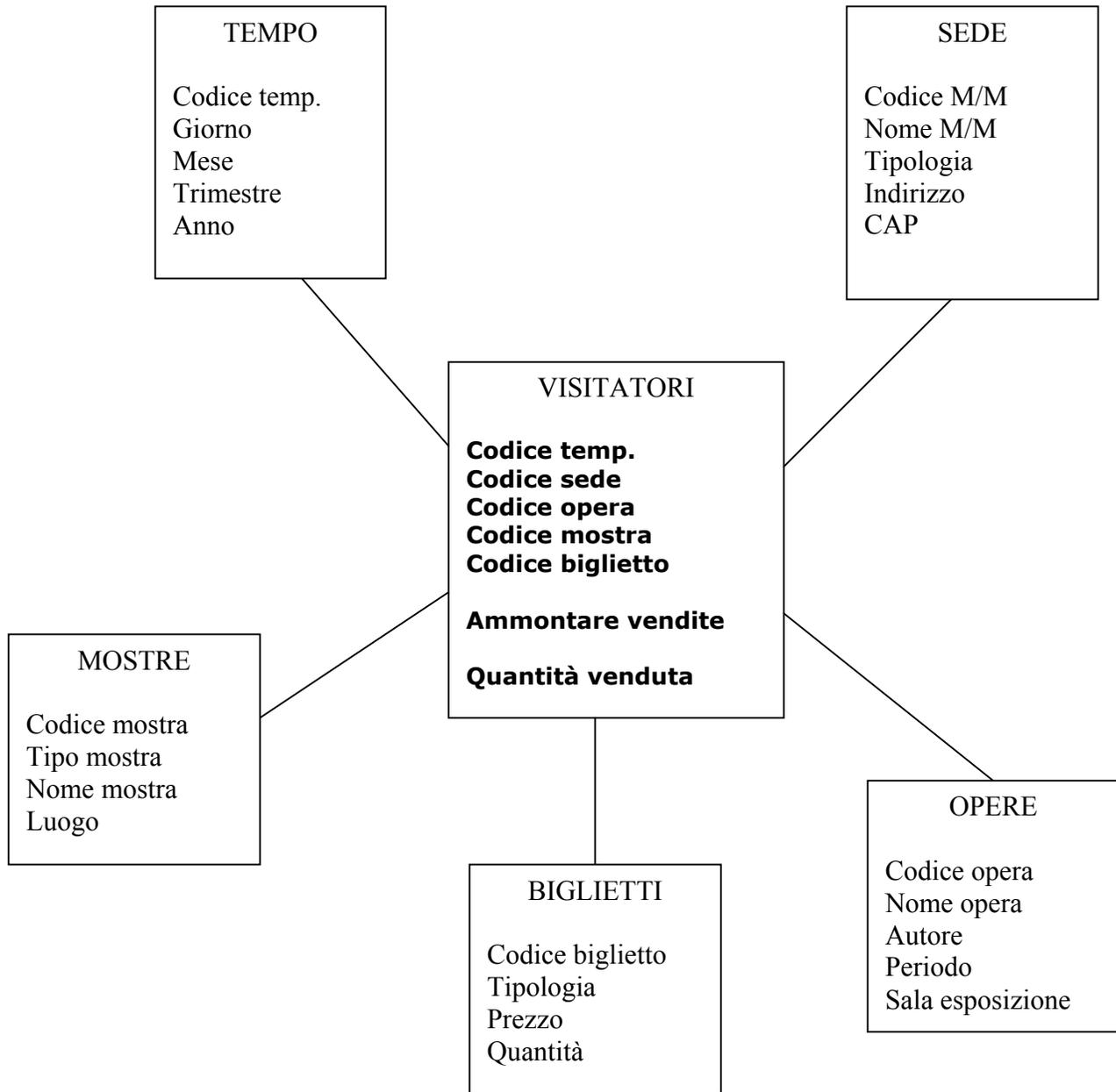


Figura 8

Nello schema a stella di figura 8 la fact table contiene i fatti: Ammontare delle vendite (dei biglietti d'ingresso) e Quantità venduta (dei biglietti), mentre le dimensional table ospitano ognuna la descrizione funzionale della dimensione del business presa in considerazione nella fact table, dettagliandone i parametri rilevanti ai fini dell'analisi. Per estendere l'esempio fatto precedentemente si potrebbero aggregare le vendite dei biglietti rispetto alle mostre per cercare di capire se le varie mostre hanno successo oppure no: esse potrebbero diventare un importante strumento di presentazione del museo al pubblico. Oppure si potrebbe, ancora una volta, aggregare i visitatori per tipologia di biglietto per verificare quali siano le classi di utenti (giovani, adulti, militari, comitive, ecc,.....) che maggiormente frequentano il museo per poi prendere, di conseguenza, decisioni appropriate. Ma naturalmente, le possibilità non si fermano qui.

Capitolo III

Scenari futuri

1. Introduzione

In questo capitolo si cercherà di fare il punto sullo stato dell'arte attuale del DataWarehouse.

Se da una parte si può infatti affermare che il processo è finalmente stato compreso, e accettato e numerosi sono gli esempi di successo, dall'altra occorre considerare che svariate sono le reali possibilità di implementazione di un DataWarehouse in funzione degli obiettivi che si intendono perseguire e della tecnologia scelta.

Si parlerà quindi di strumenti cosiddetti dinamici e del DataWarehousing nel contesto "rete", intendendo con questo termine la presenza di una Internet, Intranet o Extranet e di siti Web destinati alla richiesta e gestione delle informazioni di interesse.

Il tutto pur tenendo conto degli inevitabili e notevoli cambiamenti indotti dal tempo e dalla tecnologia.

2. Dal DataWarehouse al DataWebhouse

L'avvento del Web alla fine degli anni 90 ha enormemente influenzato lo sviluppo del DW.

La rivoluzione web è cresciuta oltre ogni aspettativa per quanto riguarda tutti i generi di informazione, tuttavia non rimpiazza il bisogno del DW.

Il DW deve essere il motore che controlla e analizza i dati ottenuti via Web, ma per questo la sua natura ha bisogno di essere in qualche modo modificata: deve prendere in considerazione un certo numero di nuove regole, abbandonare in maniera permanente alcune delle sue concezioni iniziali e pregiudizi e, in certo qual modo, reinventarsi: si ottiene così la rinascita del DW nel Datawebhouse.

Il Web ha condizionato milioni di persone ad accendere il loro computer e navigare per ottenere informazioni. Poiché la funzione primaria del DW è di generare informazioni, il Web è un invitante partner potenziale.

Indistintamente tutti gli utenti del web si aspettano determinate caratteristiche e funzioni, fra le quali troviamo, per esempio:

- il Web è un grande contenitore di informazioni
- il Web è pieno di contenuti interessanti
- tutti i clienti, i partner, gli impiegati sono collegati al Web

- il WEB va dovunque nel mondo
- il collegamento avviene per mezzo del browser così ognuno ha facilità di reperire il software di cui ha bisogno
- Il Web è facile da usare
- Si presuppone che sia veloce
- I siti web sono raggiungibili 24 ore al giorno e 7 giorni alla settimana
- Il Web trasmette in tempo reale
- Le interfacce Web sono grafiche e dinamiche
- Ogni pagina Web può essere personalizzata all'utente

Questo insieme di requisiti, che ogni utente si aspetta dal Web, può essere applicato anche al DW, in maniera da trasformarlo in un Datawebhouse. Infatti il Datawebhouse ha bisogno di essere particolarmente facile da usare, in grado di consegnare i suoi risultati in pochi secondi, disponibile 24 ore su 24, capace di creare risultati grafici personalizzati alle informazioni di cui abbiamo bisogno, ed inoltre deve essere altamente sicuro.

Il DW ha molte fonti d'informazione che provengono dai sistemi OLTP. Alcune di queste potrebbero venire dalle transazioni catturate via Web. Catturare transazioni è facile ed è quello che i progettisti del DW fanno da oltre 10 anni.

Catturare, analizzare e capire il comportamento degli utenti che arrivano al sito Web, può essere di grande vantaggio per il Museo. Il Web si presenta come una nuova fonte dati senza precedenti.

Ogni utente del web crea un flusso di dati che viene generalmente chiamato "clickstream". Il clickstream è la registrazione di ogni singolo click effettuato da ogni singolo utilizzatore, ed è quindi la fonte più

ricca e più espressiva, riguardo al comportamento e ai gusti dell'utente, di ogni altra fonte di dati abitualmente usata.

Il clickstream è una serie temporale di azioni microscopiche che possono essere assemblate. La traiettoria delle azioni che conducono a dei comportamenti a cui si è interessati possono essere analizzati e capiti: in questo modo possono essere prese decisioni migliori e più consone alla gestione del Museo.

2.1. Il Data Mining

Per scoprire correlazioni e regole comportamentali non note a priori in cui si faccia ricorso a diverse tecniche di analisi, da quelle statistiche più o meno tradizionali, all'utilizzo di algoritmi genetici, logica fuzzy, sistemi esperti, reti neurali, ecc., è facile prevedere una diffusione sempre maggiore del Data Mining.

Tecniche di Data mining sono state usate per anni prima dell'avvento del DataWarehousing in applicazioni specialistiche non commerciali, quali la ricerca geologica, medica, demografica, astronomica, ecc.

Tramite il DataWarehousing, il Data Mining viene trasportato dal mondo dell'analisi scientifica a quello dell'analisi commerciale, aprendo la via a una miriade di applicazioni di incalcolabile valore ai fini del supporto decisionale. Qui di seguito si cita solo qualche esempio:

- ricerche di mercato, per es. studio delle preferenze del cliente, analisi dello stile di vita, valutazione della soddisfazione dei clienti.

- studio dell'efficacia del marketing, per es. click-stream analysis per monitorare l'utilità di un sito web.
- segmentazione di mercato per identificare profili-cliente significativi al fine di fornire migliori servizi o per promozioni mirate.
- pianificazione per la gestione dei carichi di lavoro, il controllo dei costi e del budget, la previsione delle entrate, ecc.

Tramite il Data mining non occorre più effettuare manualmente l'analisi dei dati, ma è sufficiente indicare con una certa approssimazione cosa dove si vuole ricercare e lasciare che uno strumento automatico si sobbarchi il peso della computazione. Preme tuttavia sottolineare che, nonostante sia opinione diffusa che il Data mining sia la panacea che risolve tutti i problemi di analisi avanzata, in realtà esso permette solamente di analizzare automaticamente grosse quantità di dati: il successo della ricerca dipende principalmente dalla corretta esecuzione di tutti i passi del processo e dalle capacità dell'analista.

3. Tendenze e sviluppi

Non è facile fare previsioni per il futuro anche quando il campo è limitato, come in questo caso, ai DataWarehouse.

In questo paragrafo si cercherà di delineare brevemente quali sono gli sviluppi attualmente in atto e più promettenti.

In realtà le metodologie e tecnologie qui riportate sono ampiamente disponibili già nel presente, ma si parla del futuro come del tempo in cui queste saranno diffusamente adottate ed effettivamente utilizzate.

Si è già parlato di DataWebhouse come della naturale fusione tra Web e DataWarehouse. Ormai quasi tutti i software di tipo OLAP sono in grado di garantire un certo grado di interazione attraverso un Browser Web. In breve tempo aumenterà l'interattività tra utente e Data Webhouse tramite la creazione di pagine Web dinamiche (ASP, JSP, e in genere programmazione lato server), l'utilizzo di Applet Java o controlli ActiveX all'interno delle pagine web. L'uso della "Push technology" permetterà la distribuzione attraverso Internet o Intranet di informazioni non contestualmente richieste dall'utente, ma alle quali venga effettuata una semplice "sottoscrizione".

I recenti sviluppi della tecnologia WAP (Wireless Access Protocol) permettono di anticipare un uso di applicazioni WAP anche in campo OLAP che avverrà tramite la fusione tra PC e telefono cellulare.

La vera incognita sarà però la possibilità o meno di visualizzare dati e tabelle strutturate sullo schermo ridotto di un terminale WAP; probabilmente nell'immediato saranno SubNotebook e PDA (Personal Digital Assistant) a veicolare dati OLAP verso i "manager on-line".

CONCLUSIONI

Per scoprire a volte bisogna osare ("Considerate la vostra semenza: fatti non foste a viver come bruti, ma per seguire virtute e conoscenza", Dante, Inferno XXVI), ma la cultura classica ha sempre visto (perlomeno fino a poco tempo fa) la tecnologia con un po' di sospetto e timore. Ormai però i tempi sono maturi affinché il mondo della cultura e dell'arte, in questo caso, i musei, conviva con quello delle tecnologie informatiche, per cui non si impedisca il loro incontro, ma anzi si cerchi di trovare un modo affinché essi, non solo convivano, ma lavorino in sinergia per far sì che il loro incontro sia produttivo.

L'esplorazione dei dati e delle informazioni, tramite DW e Data Mining, permette di aprire e percorrere nuove strade fino ad ora inesplorate e di ampliare i nostri orizzonti. Il loro sistematico utilizzo consentirà una gestione a lungo raggio, non limitandosi ad una visione "giornaliera", ma attuando la trasformazione dei musei in spazi attivi, multimediali, luoghi totali e totalizzanti dove vivere l'esperienza "arte" con ausili tecnologici, pause rilassanti, acquisti che mantengano l'esperienza viva nella memoria. Il numero crescente di presenze nei musei che offrono questa tipologia e' l'esempio eloquente del gradimento che stanno riscuotendo queste strutture ovunque nel mondo vengano realizzate.

Il museo quindi cambia e si mostra come un centro di servizi e ciò parallelamente al mutamento della politica gestionale nei confronti del nostro patrimonio artistico attuata dal Ministero per i Beni Culturali e le Attività Culturali. L'ottica è muoversi nell'indirizzo delle più innovative strutture europee ed internazionali che da tempo hanno capito che il patrimonio artistico va gestito anche da un punto di vista manageriale, cioè economico.

Il management riguarda quindi l'organizzazione e la gestione di tutte le attività che gravitano attorno alla realtà museale e si concretizza nel far spuntare servizi aggiuntivi, in una promozione più attenta e vivace, dotata di nuove tecnologie, nella diversa qualificazione del merchandising culturale e nell'ottimizzazione della divulgazione del materiale editoriale informativo.

La definizione conservativa essenzialmente rivolta ai beni da salvaguardare si arricchisce di altre funzioni quali la valorizzazione e la comunicazione rivolte ad una sfera esterna ai beni stessi, non più orientata alla sola utenza ma anche al mercato. Ecco allora che il museo allarga a macchia d'olio la sua zona di influenza al territorio circostante, implicando un plus di strutture e di centri di servizi che producono valore aggiunto all'economia locale. Di conseguenza si viene a creare un contesto aziendale che produce grandi quantità di dati e informazioni relative a se stesso e all'ambiente nel quale si colloca; questi dati possono costituire un importante fattore di sviluppo se il loro potenziale (inteso come capacità di prevedere l'evoluzione di variabili di interesse o le tendenze dell'ambiente esterno) viene utilizzato.

I motivi che finora hanno rallentato il pieno utilizzo della grande mole di informazioni disponibili in ogni istituzione sono riconducibili essenzialmente

- da un lato alla frammentazione delle informazioni all'interno di differenti sistemi di archiviazioni non comunicanti fra loro con la conseguente inefficienza nell'organizzazione dei dati

- dall'altro, alla scarsa consapevolezza della potenzialità degli strumenti statistici ed informatici di elaborazione delle informazioni, che impedisce la produzione di sintesi e reporting efficaci e rilevanti, analogamente a quanto avviene nel contesto aziendale.

Da queste considerazioni nasce quindi la proposta di adottare una tecnologia ormai matura e consolidata in molti ambiti, soprattutto economico-finanziario, anche al contesto museale. Si affermerà questo strumento innovativo anche in altri settori diversi da quelli dove storicamente e' stato originato sulla spinta finalizzata ad ottenere vantaggi strategici d'impresa? Se la risposta sarà affermativa assisteremo ad un mutamento nel nostro modo di vedere e vivere il museo e, di conseguenza, l'arte.

Glossario dei termini

Aggregazione Nel modello multidimensionale, meccanismo per la sintesi dei dati che permette di raggruppare un insieme di eventi in un unico evento che li riassume.

Alimentazione Procedura attraverso la quale i dati presenti nelle sorgenti operazionali vengono integrati, resi consistenti e caricati nel data warehouse.

applicazione Le maschere e i report con cui l'utente interagisce.

associazione Un legame fra due o più entità.

Attributo Nel modello relazionale, nome dato a una colonna di una relazione. Nel modello multidimensionale e nel DFM, è un campo con dominio finito (tipicamente alfanumerico) che descrive una dimensione.

Business Intelligence (BI) Qualsiasi processo usato per estrarre ed analizzare dati aziendali. Questo termine sta avendo un notevole successo e sta soppiantando il suo sinonimo Decision Support System.

Campo la rappresentazione di un attributo in un data base

Caricamento Fase del processo di data warehousing durante la quale i dati operazionali ripuliti e trasformati vengono caricati nel data warehouse.

Cubo Metafora adottata per rappresentare i dati nel modello multidimensionale. Ogni evento accaduto viene visto come una cella di un cubo i cui spigoli rappresentano le dimensioni di analisi.

Data Mart (DM) Un sottoinsieme o un'aggregazione dei dati presenti nel data warehouse, contenente l'insieme delle informazioni rilevanti per una particolare area del business, una particolare divisione dell'azienda, una particolare categoria di soggetti.

Data mining Attività orientata a scoprire informazioni nascoste nei dati sotto forma di pattern o modelli quali regole associative, cluster e serie temporali.

Data warehouse Una collezione di dati di supporto per il processo decisionale orientata ai soggetti di interesse, integrata e consistente, rappresentativa dell'evoluzione temporale e non volatile.

Data warehousing Metodi, tecnologie e strumenti di ausilio al "lavoratore della conoscenza" per condurre analisi dei dati finalizzate all'attuazione di processi decisionali e al miglioramento del patrimonio informativo aziendale.

database La combinazione dello schema di database e dei dati memorizzati.

Database Management System (DBMS) Un Database Management System è un software, o un gruppo di software, attraverso il quale è possibile definire, creare e gestire una base di dati e regolarne l'accesso alle informazioni. La maggior parte dei DBMS utilizzati nel Data Warehousing sono basati sulla tecnologia relazionale mentre alcuni dei prodotti più recenti sfruttano direttamente una memorizzazione di tipo multidimensionale.

Database operativo Contiene i dati generati da operazioni, principalmente di carattere amministrativo, svolte all'interno dei processi gestionali.

Decision support system (DSS) Vedi sistemi di supporto alle decisioni

Dimension table Nello schema a stella, tabella relazionale denormalizzata contenente una chiave surrogata e l'insieme degli attributi facenti parte di una gerarchia.

Dimensional Fact Model (DFM) Modello concettuale per la progettazione di data warehouse.

Dimensione E' un attributo strutturale di un Array Multidimensionale formato da una lista di membri simili dal punto di vista dell'utente finale; ad esempio giorni, mesi, trimestri e anni fanno parte della dimensione temporale mentre comuni, province e regioni possono far parte di una dimensione geografica. Una dimensione può essere considerata un indice in base al quale identificare i valori in un Array Multidimensionale e permette di avere un modo conciso e intuitivo di organizzare e selezionare i dati da analizzare.

Drill-down Operatore OLAP che diminuisce l'aggregazione dei dati in un cubo introducendo un ulteriore livello di dettaglio.

Drill-through Operatore OLAP che permette il passaggio dai dati aggregati multidimensionali di un cubo ai dati operazionali presenti nelle sorgenti o nel livello riconciliato.

Entità Qualsiasi cosa su cui il sistema ha necessità di memorizzare informazioni.

Entity/Relationship Diffuso modello concettuale per la progettazione di basi di dati operazionali.

Estrazione Fase del processo di data warehousing durante la quale i dati rilevanti vengono estratti dalle sorgenti.

ETL - Extraction, Transformation e Loading Si tratta di strumenti che permettono di integrare schemi eterogenei, nonché di estrarre, trasformare, ripulire, validare, filtrare e caricare i dati dalle sorgenti nel data warehouse.

Nelle architetture a più livelli le operazioni svolte vengono spesso globalmente indicate con il termine *riconciliazione*.

Fact table (tabella dei fatti) Nello schema a stella, tabella relazionale contenente tutte le misure di un fatto. La fact table importa le chiavi di tutte le dimension table, che insieme ne formano la chiave primaria.

Fatto Concetto di interesse per il processo decisionale. Tipicamente modella un insieme di eventi che accadono nell'impresa.

Integrazione Procedimento attraverso il quale più schemi operazionali sorgenti possono essere combinati in un unico schema consistente.

Integrità dei dati Le regole usate da un database per assicurare che i dati siano, se non corretti, almeno plausibili.

Intervallo di storicizzazione Per ogni schema di fatto specifica l'ampiezza della finestra temporale durante la quale i dati vengono conservati.

Join Date due relazioni e un predicato booleano che coinvolge attributi di entrambe, l'operazione di join collega tutte le coppie di tuple prese dalle due relazioni per le quali il predicato è soddisfatto.

Knowledge Discovery in Databases (KDD) Un sinonimo di Data Mining.

Logica fuzzy La logica fuzzy si basa su enunciati che sono veri in una qualche misura tra 0 e 1. Pertanto non è necessario che la legge di Aristotele del terzo escluso *ha o non ha* valga al cento per cento.

Meta-dati Dati che descrivono altri dati. Nel contesto del data warehousing indicano le sorgenti, il valore, l'uso e le funzioni dei dati memorizzati nel data warehouse e descrivono come i dati sono alterati e trasformati durante il passaggio attraverso i diversi livelli dell'architettura.

Modello dei dati La descrizione concettuale di uno spazio del problema in termini relazionali.

Modello multidimensionale Modello alla base dei sistemi di data warehousing, vede i dati come punti in uno spazio le cui dimensioni corrispondono ad altrettante possibili dimensioni di analisi; ciascun punto, rappresentativo di un evento accaduto nell'azienda, viene descritto tramite un insieme di misure di interesse per il processo decisionale.

Modello relazionale Modello logico per la memorizzazione dei dati sotto forma di relazioni o tabelle.

MOLAP Implementazione del modello multidimensionale su un modello logico *ad hoc* nel quale i dati multidimensionali possono essere direttamente rappresentati sotto forma di matrici ad accesso posizionale.

Multi-dimensional Array (Cube) Sinonimi: Hypercube, Struttura Multidimensionale, Data Cube.

OLAP On-Line Analytical Processing, elaborazione interattiva dei dati orientata all'analisi dinamica e multidimensionale.

OLTP On-Line Transactional Processing, elaborazione interattiva dei dati basata su transazioni.

Pattern Insieme di attributi delle gerarchie che definiscono il livello di aggregazione per un'interrogazione. Nel data mining, modello significativo estratto dai dati grezzi.

Progettazione concettuale Fase mirata a produrre una descrizione astratta di un data mart, richiede di identificare i concetti di interesse e di costruire gli schemi di fatto.

Progettazione fisica Fase di sviluppo del data mart in cui viene definito lo schema fisico che include un insieme di indici e direttive sull'allocazione dei dati.

Progettazione logica Fase di sviluppo del data mart in cui vengono definiti gli schemi relazionali che modellano i fatti. Durante questa fase vengono applicate tecniche di ottimizzazione che incidono sugli schemi prodotti: materializzazione delle viste e frammentazione.

Pulitura Fase del processo di data warehousing durante la quale viene migliorata la qualità dei dati.

Regola aziendale Un vincolo di integrità che ha origine nello spazio del problema anziché nella teoria relazionale.

Relazione Un costrutto logico che organizza i dati in righe e colonne.

Report Rapporto riassuntivo di tipo statico destinato a utenti che hanno necessità di accedere periodicamente a informazioni strutturate in modo pressoché invariabile.

ROLAP Implementazione del modello multidimensionale sul modello logico relazionale, effettuata tipicamente attraverso uno schema a stella.

Roll up Operatore OLAP che aggrega un insieme di eventi in un cubo diminuendone il livello di dettaglio.

Schema a stella Schema base per la rappresentazione di dati multidimensionali attraverso il modello relazionale. Consiste di una fact table e un insieme di dimension table a essa collegate.

Sistemi di supporto alle decisioni L'insieme delle tecniche e degli strumenti informatici atti a estrapolare informazioni da un insieme di dati memorizzati su supporti elettronici.

Sistemi legacy Applicazioni esistenti in azienda, tipicamente su mainframe o minicomputer, che pur essendo correntemente utilizzate per la gestione operativa non rispondono ai requisiti architetturali moderni e agli standard attuali, risultando di conseguenza difficilmente accessibili e integrabili con sistemi di più recente realizzazione.

Slice Una Slice è composta da un sottoinsieme, o "fetta", di un Array Multidimensionale corrispondente a uno o più elementi di una o più dimensioni. Le dimensioni lungo le quali si effettua la selezione degli elementi non saranno presenti nel sottoinsieme stesso. Dal punto di vista dell'utente il termine Slice si riferisce di solito a una vista bidimensionale creata dal cubo multidimensionale.

Slice and Dice Il processo di navigazione dei dati da parte dell'utente finale, è effettuato per mezzo di visualizzazioni interattive prodotte da rotazioni e selezioni o da operazioni di Drill-Down/Roll-Up.

Slicing Operatore OLAP che riduce la dimensionalità di un cubo fissando un valore per una delle sue dimensioni.

Trasformazione Fase del processo di data warehousing durante la quale i dati vengono convertiti dal formato operativo sorgente a quello utilizzato nel livello riconciliato.

Tupla Nel modello relazionale, una singola riga di una tabella.

Vista. Fact table aggregata che consolida i risultati di un insieme di interrogazioni e permette pertanto di ridurre i costi.

Bibliografia

Susi Dulli, Vittorio Favero, *Modelli e strutture per il Data Warehouse*
Diade, 2000

Susi Dulli, Renato Guarino, *Uno strumento innovativo a supporto
dell'organizzazione e gestione delle attività museali: il Data Warehouse*
in *Conoscere l'arte per conoscere*, Luciano Pilotti (a cura di)

Angela Roncaccioli (a cura di), *Il Museo come azienda culturale*, 1996

Filippo la Noce, Luigi D'Ercole, *Data Warehouse – Dal dato
all'informazione* Franco Angeli, 1998

Ralph Kimball, *The Data Warehouse Toolkit*, Wiley second edition,
2001

Ralph Kimball, Richard Merz, *The Data Webhouse Toolkit – Building the
Web – Enabled Data Warehouse*, Wiley, 2000

Antonio Albano, Giorgio Ghelli, Renzo Orsini, *Basi di dati relazionali e a
oggetti*, Zanichelli 1997

Rebecca M. Riordan, *Progettare database relazionali*, Mondadori 1999

Matteo Golfarelli, Stefano Rizzi, *Data Warehouse – Teoria e pratica
della progettazione* McGraw-Hill, 2002

Sistemi Intelligenti, *Il Museo Multimediale*, Rivista trimestrale di
Scienze Cognitive e Intelligenza Artificiale (numero monografico), Il
Mulino, 1998

Ringraziamenti

A compimento di questo lavoro vorrei, prima di tutto, ringraziare la professoressa Dulli per tutto il tempo, l'impegno che mi ha dedicato e per avermi aiutato e sopportato fino adesso.

Anche se non direttamente coinvolti in questa tesi ringrazio di cuore Alberto Cammozzo, Silvia Sartorelli, Angelo Scarano i tecnici dell'aula ASID per avermi fornito conoscenze nelle loro rispettive aree di lavoro (e naturalmente per l'amicizia dimostratami).

Inoltre un ringraziamento speciale va a Tomaso Minelli, Nicola Margola e al professor Claudio Capiluppi per l'amicizia, l'aiuto e il supporto datomi in questi anni.