Data Warehouse

PDF created with pdfFactory trial version www.pdffactory.com

Agenda

- Definizione
- Architetture per il Data Warehouse
- Gli strumenti di Extraction, Transformation e Loading
- I Metadati
- Accesso al DW

PDF created with pdfFactory trial version www.pdffactory.com

Definizione

PDF created with pdfFactory trial version www.pdffactory.com

Data Warehouse: una collezione di dati in supporto al processo decisionale del management…

- Orientata al soggetto
- Integrata
- Dipendente dal tempo
- Non volatile

Bill Inmon

PDF created with pdfFactory trial version www.pdffactory.com

ORIENTATA AL SOGGETTO:

- Perché il DWH è orientato a temi specifici dell'azienda (clienti, prodotti, ecc.) piuttosto che alle applicazioni o funzioni (quali ad esempio in un contesto bancario alle applicazioni transazionali).

- Nel DWH i dati vengono archiviati in modo che possano essere facilmente letti o elaborati dagli utenti cioè in modo da favorire la produzione di informazioni; viceversa i database operazionali sono organizzati intorno alle differenti applicazioni del dominio aziendale.

PDF created with pdfFactory trial version www.pdffactory.com

INTEGRATA:

- L'integrazione è un requisito fondamentale del DWH in quanto in esso confluiscono dati provenienti da più fonti;
- L'obiettivo dell'integrazione può essere raggiunto percorrendo differenti strade: mediante l'utilizzo di metodi di codifica uniformi, mediante il perseguimento di una omogeneità sistematica di tutte le variabili, mediante l'utilizzo delle stesse unità di misura,ecc.
- La costruzione di un sistema di Data Warehouse **non** comporta l'inserimento di nuove informazioni bensì la **riorganizzazione** di quelle esistenti, e implica pertanto l'esistenza di un sistema informativo.

PDF created with pdfFactory trial version www.pdffactory.com

DIPENDENTE DAL TEMPO:

- I dati archiviati all'interno di un DW hanno un orizzonte temporale molto più esteso rispetto a quelli archiviati in un sistema operazionali.
- Nel DW sono contenute una serie di informazioni relative all'area di interesse che colgono la situazione relativa ad un dato fenomeno in un determinato intervallo temporale piuttosto esteso.
- Ciò, tuttavia, comporta che dati contenuti in un DW sono aggiornati fino ad una certa data, che nella maggior parte dei casi, è antecedente a quella in cui l'utente interroga il sistema.
- Situazione differente si manifesta in un database transazionale in cui i dati corrispondono sempre ad una situazione costantemente aggiornata che tuttavia non fornisce un quadro storico del fenomeno.

PDF created with pdfFactory trial version www.pdffactory.com

NON VOLATILE:

- Il dato viene caricato ed acceduto fuori linea cioè non può essere modificato dall'utente (l'accesso è in sola lettura).
- Si evitano le possibili anomalie dovute ad aggiornamenti e tanto meno si ricorre a strumenti complessi per gestire l'integrità referenziale o per bloccare record a cui possono accedere altri utenti in fase di aggiornamento.

PDF created with pdfFactory trial version www.pdffactory.com

Perché accontentarsi di un Data Base quando si può avere un Data Warehouse?

Data Warehouse = Creare Conoscenza

Data Warehouse = Trasformare il dato in informazione.

Data Warehouse = Consolidazione dei dati operazionali da sistemi legacy diversi in un unico deposito informativo.

PDF created with pdfFactory trial version www.pdffactory.com

Perché accontentarsi di un Data Base quando si può avere un Data Warehouse?

- Mentre i dati operazionali coprono un arco temporale di solito piuttosto limitato, poiché la maggior parte delle transazioni coinvolge i dati più recenti, il DW deve permettere analisi che spaziano sulla prospettiva di alcuni anni.

- Per questo motivo, il DW è aggiornato a intervalli regolari a partire dai dati operazionali ed è in crescita continua.

- Volendo fare un paragone possiamo supporre che, a intervalli regolari, venga scattata una fotografia istantanea dei dati operazionali.

PDF created with pdfFactory trial version www.pdffactory.com

Perché accontentarsi di un Data Base quando si può avere un Data Warehouse?

- La progressione delle fotografie scattate viene immagazzinata nel DW, dove genera un film che documenta la situazione aziendale da un istante zero fino al tempo attuale.
- Proprio il fatto che, in linea di principio, non vengono mai eliminati dati dal DW e che gli aggiornamenti siano tipicamente eseguiti “a freddo”, ossia quando il DW è fuori linea, fa sì che un DW possa essere fondamentalmente considerato come un database a sola lettura.

PDF created with pdfFactory trial version www.pdffactory.com

Ulteriori e fondamentali differenze tra database operazionali e DW sono legate alle tipologie di interrogazioni

- Per i primi, le interrogazioni eseguono transazioni che in genere leggono e scrivono un ridotto numero di record da diverse tabelle legate da semplici relazioni: per esempio, si ricercano i dati di un cliente per inserire un suo nuovo ordine. Questo tipo di elaborazione viene comunemente detto On-Line Transactional Processing (**OLTP**).

- Al contrario, il tipo di elaborazione per cui nascono i DW viene detto On-Line Analytical Processing (**OLAP**), ed è caratterizzato da un'analisi dinamica e multidimensionale che richiede la scansione di un'enorme quantità di record per calcolare un insieme di dati numerici di sintesi che quantificano le prestazioni dell'azienda.

PDF created with pdfFactory trial version www.pdffactory.com

Differenze tra DW e Sistema Transazionale

	Data Warehouse	**Sistema Transazionale**
Utilizzo	*query - intensive*	*transaction - intensive*
Utenti	non molto numerosi	abbastanza numerosi
Storicità dei dati	dati integrati per soggetto	dati attuali
Integrazioni	dati integrati per soggetto	dati aggregati per attività o per processo
Database	dati aggiornati ad intervalli predefiniti, perciò *non-volatile*	dati aggiornati continuamente e quindi volatili
Modello dei dati	dati in genere denormalizzati; il modello è conforme alle dimensioni del soggetto	dati normalizzato; il modello è conforme alle esigenze che derivano dalle transazioni da supportare

PDF created with pdfFactory trial version www.pdffactory.com

Differenze tra DW e Sistema Transazionale

	Data Warehouse	**Sistema Transazionale**
Dominio Applicativo	i progetti di DW forniscono un'infrastruttura di supporto ai sistemi di supporto alle decisioni con caratteristiche di scalabilità di ampliamento e flessibilità	i sistemi transazionali sono definiti per un limitato dominio applicativo che si riferisce a una specifica applicazione
Sviluppo	i criteri di sviluppo rispondono a principi evolutivi e iterativi	sistemi OLTP sviluppati seguendo i requisiti del sistema esplicitati dagli utenti e il metodo di sviluppo a cassetta
Sponsorship	un progetto di DW richiede una forte sponsorizzazione a causa dell'ampiezza organizzativa dello stesso	i sistemi transazionali tendono ad essere sponsorizzati seguendo un processo che consente di individuare il responsabile che individua a sua volta anche le gerarchie organizzative

PDF created with pdfFactory trial version www.pdffactory.com

Cosa non è il Data Warehouse

Un prodotto che può essere acquistato e installato.

- Il Data Warehouse è un SISTEMA INFORMATIVO che va sviluppato da zero utilizzando, è vero, tecnologia software e hardware specifica, ma una volta acquistato il prodotto si è fatto si e no il 5% del lavoro, senza contare che molto spesso la scelta si rileva sbagliata o sottodimensionata perché non è stata fatta a fronte di un'attenta, seria e competente analisi delle esigenze di business.

PDF created with pdfFactory trial version www.pdffactory.com

Cosa non è il Data Warehouse

Una soluzione di business in sé e per sé.

- Il Data Warehouse non va progettato e creato in quanto tale. Un'azienda non affronta un progetto di 'data warehousing' in quanto tale, ma dovrebbe decidere in maniera seria e ragionata di risolvere una specifica esigenza di business (per esempio il Customer Care) utilizzando la tecnologia del data warehouse. Se un data warehouse parte con presupposti diversi da questi è destinato a fallire.

PDF created with pdfFactory trial version www.pdffactory.com

Cosa non è il Data Warehouse

Una 'pezza' ai problemi dei sistemi transazionali

- Il data warehouse è un'ottima risposta a esigenze di analisi e supporto decisionale, funzionalità supportate poco o male dai sistemi transazionali. Punto. Non deve essere utilizzato per risolvere altri problemi che non è deputato a risolvere. Alcuni esempi:

1. Calcolo dei costi di produzione;
2. Stampa delle etichette per il mailing;
3. Ribaltamenti e 'spalmature' della contabilità analitica;
4. Stampa del conto economico aziendale (civilistico);
5. Contact management;
6. Data entry manuale di informazioni;

PDF created with pdfFactory trial version www.pdffactory.com

Cosa non è il Data Warehouse

- Se qualcuno vi dice che intende usare il data warehouse per una o più di queste cose fermatelo prima che si faccia del male: spenderà tantissimi soldi e otterrà risultati scarsi. In realtà molto spesso chi crea confusione su cosa sia e come si fa un data warehouse sono i software vendor e le società di consulenza, gli uni nel tentativo di legare i propri prodotti a slogan e concetti che nel mercato 'tirano', le seconde nel costante tentativo di inventare qualcosa di nuovo da proporre ai clienti.
- Ecco che così un report writer qualsiasi diventa 'un prodotto OLAP' ed un database su PC usato per stampare conti economici per prodotto diventa 'il data warehouse dell'analisi della redditività'.

PDF created with pdfFactory trial version www.pdffactory.com

Cosa non è il Data Warehouse

- Costruire un data warehouse è un impresa lunga e complessa dove nulla va banalizzato e dove le competenze di tipo tecnologico contano esattamente quanto quelle funzionali.

PDF created with pdfFactory trial version www.pdffactory.com

Alcune osservazioni

- Non esiste una metodologia unica, consolidata e sempre valida per la realizzazione di un DW. Il motivo è semplice: realizzare sistemi di Decision Support non è proprio una scienza esatta, è piuttosto un percorso lungo il quale si effettuano spesso prove e s'impara dagli errori commessi.
- Arrivare ad DW "perfetto" (ammesso che esista) per una specifica azienda (o istituzione) è un obiettivo che si raggiunge per approssimazioni successive.

PDF created with pdfFactory trial version www.pdffactory.com

Imperativi-dogmi-assiomi!!!!

THINK BIG, START SMALL

- Non bisogna strafare, ma neanche avere un'ottica di breve. Per essere pratici: non bisogna iniziare un progetto di data warehousing con l'obiettivo di realizzare un data warehouse aziendale ma con la visione di arrivare a questo traguardo. Cosa significa? Significa che bisogna iniziare da aree analitiche relativamente semplici e conosciute in termini di dati e problematiche e costruire una serie di "storie di successo" che convincano l'organizzazione che il data warehouse è la strada giusta. Però è sbagliato compiere scelte (soprattutto da un punto di vista tecnologico) che non siano 'scalabile', cioè che richiedano di essere cambiate nel momento in cui la complessità ed il volume dei dati crescono. In sintesi quindi gli obiettivi devono essere a breve, la visione del progetto deve invece guardare necessariamente lontano.

PDF created with pdfFactory trial version www.pdffactory.com

Imperativi-dogmi-assiomi!!!!

Nel team inserisci solo amanti del data warehouse.

- Un progetto di DW si può fare solo e unicamente con gente che vive per costruire architetture di warehousing. Il ventaglio delle tematiche da affrontare in tale tipologia di progetto è talmente vasta che per costruire un esperto di data warehousing ci vogliono anni, non mesi. E' questa la ragione per cui chi affronta un progetto di data warehousing deve considerare ciò come l'unico scopo della propria vita.

PDF created with pdfFactory trial version www.pdffactory.com

Imperativi-dogmi-assiomi!!!!

Parti dall'esigenza, non dal prodotto.

- Come già detto, il DW non si compra. Diffidate di chiunque sostenga il contrario, pacchettizzare una piattaforma di decision support non è possibile ed in ogni caso ha delle controindicazione evidenti. Nello stesso tempo è scorretto selezionare un prodotto e poi costruire il DW con tale prodotto. La software selection può derivare unicamente da un'analisi attenta e seria delle esigenze funzionali da soddisfare e dell'infrastruttura tecnica e tecnologica di cui il sistema di DW sarà una componente.

PDF created with pdfFactory trial version www.pdffactory.com

Imperativi-dogmi-assiomi!!!!

Non banalizzare gli aspetti di business.

- A chiosa di quanto esposto al punto precedente va rimarcato quest'altro aspetto: la realizzazione di una DW deve assolutamente prendere mosse da un'analisi degli aspetti di business. Un DW non focalizzato su uno o più specifici aspetti di business sarà un DW senza utenti e quindi destinato a morire di inedia e solitudine.

PDF created with pdfFactory trial version www.pdffactory.com

Imperativi-dogmi-assiomi!!!!

Non banalizzare gli aspetti tecnologici

- Non bisogna però nemmeno eccedere nella direzione esattamente opposta. La realizzazione di un DW pone tali e tanti problemi di capacity planning e di definizione architetturale che un’attenzione meno che ossessiva a questo aspetto può decretare il fallimento del sistema. Non sono accettabili tempi di attesi lunghi, né interfacce poco user friendly, né continue failure di sistema dovute a mancanza di memoria, CPU,etc.

PDF created with pdfFactory trial version www.pdffactory.com

Imperativi-dogmi-assiomi!!!!

Attento alla qualità dei dati

- Secondo alcune ricerche circa il 50% degli insuccessi nei progetti di DW sono dovuti a problemi di scarsa qualità dei dati alimentati. E' opportuno, prima di partire con l'implementazione del DW, effettuare una verifica sulla qualità dei dati alimentati, onde evitare brutte sorprese in seguito. Per la verità molti problemi di data quality vengono scoperti durante il progetto ma un'analisi preliminare ben fatta permette di individuare eventuali aree particolarmente critiche o delle problematiche tipiche dei sistemi informativi alimentanti

PDF created with pdfFactory trial version www.pdffactory.com

Imperativi-dogmi-assiomi!!!!

- **Lascia perdere il report layout**
- L'approccio tradizionale alla progettazione dei sistemi di supporto decisionale prevede che si parta dai fabbisogni informativi, cioè dalle informazioni di cui gli utenti del sistema vorranno la disponibilità. Per questo motivo moltissima gente, invece di preoccuparsi di analizzare il modello dimensionale di business dell'azienda, si mette a realizzare improbabili e artistici report layout andando in giro a chiedere:"E' questa l'informazione che vuoi vedere?". Questo tipo di lavoro ha poco senso ed è pericoloso. Ha poco senso in
- Quanto l'utilizzo del DW avviene mediante dei report writer (mediante i quali l'utente stesso può disegnarsi i report) o attraverso degli OLAP browser (in cui il concetto di report non esiste). E' pericoloso perché può capitare che i report finali siano diversi da quelli presentati e inoltre perché da un punto di vista di layout qualsiasi foglio elettronico consente destrutturate perversioni artistiche, inimmaginabili per un OLAP viewer, tradizionalmente basato su uno schema dati di tipo rigorosamente multidimensionale

PDF created with pdfFactory trial version www.pdffactory.com