

Sommario

Introduzione	2
Integrazione dati	3
ETL.....	3
Modellazione.....	4
Cubo Olap.....	4
Schema a stella.....	4
Granularità	4
Datawarehouse strumenti	4
Visualizzazione dei dati	5
Fonti:	5

Introduzione

Un datawarehouse è una collezione o aggregazione di dati strutturati, provenienti da fonti interne operazionali ([DBMS](#)) ed esterne al sistema informativo aziendale, utili ad analisi e [rapporti informativi](#), prima adattati tramite strumenti appositi di trasformazione dei dati di tipo [ETL](#), e poi analizzati tramite strumenti di analisi

Un requisito fondamentale di un *data warehouse* è l'integrazione dei dati raccolti. Nel *data warehouse* confluiscono dati provenienti da più [sistemi transazionali](#) e da fonti esterne.

il data warehouse è orientato a temi aziendali specifici, alle applicazioni o alle funzioni. In un DW i dati vengono archiviati in modo da essere facilmente letti o elaborati dagli utenti.

I dati archiviati all'interno di un DW coprono un orizzonte temporale molto più esteso rispetto a quelli archiviati in un sistema operativo. Nel DW sono contenute una serie di informazioni relative alle aree di interesse che colgono la situazione relativa ad un determinato fenomeno in un determinato intervallo temporale piuttosto esteso.

Non volatile: tale caratteristica indica la non modificabilità dei dati contenuti nel DW, che consente accessi in sola lettura. Ciò comporta una semplicità di progettazione della banca dati rispetto a quella di un'applicazione transazionale.

I dati provenienti dai sistemi transazionali o sorgenti dati: sono quell'insieme di dati elaborati dai [sistemi transazionali](#) dell'azienda. Essi possono essere contenuti all'interno della stessa banca dati, provenienti da diverse [banche dati](#) o anche esterni all'azienda. Spesso l'architettura di un *data warehouse* prevede l'integrazione dei dati interni con quelli esterni.

Il data movement: tale componente è responsabile dell'estrazione dei dati dai sistemi transazionali, dell'integrazione tra dati aziendali e dati esterni, del pre-processing dei dati, del controllo della consistenza dei dati, della conversione delle strutture dati, e dell'aggiornamento dei dizionari dei dati.

Il data warehouse: i dati estratti dagli archivi transazionali vengono memorizzati internamente al *data warehouse*. Nel *data warehouse* l'accesso ai dati è consentito in sola lettura. Tali dati hanno una dimensione storica e sono riferiti a soggetti d'impresa. Essi possono essere memorizzati in un archivio centrale o in un [data mart](#). Il termine *data mart* identifica un *data warehouse* di dimensioni ridotte, specializzato per una particolare area di attività.

L'utente finale: i dati contenuti nel data warehouse vengono presentati all'utente finale, il quale dispone di un insieme di strumenti per effettuare elaborazioni e produrre informazioni appropriate. Gli strumenti a disposizione dell'utente possono essere semplici generatori di [query](#) e report, interfacce grafiche che consentono la rappresentazione dei dati o sistemi di analisi dati più complessi.

Il data warehouse è organizzato su quattro livelli architetturali:

1. *trasformazione dei dati*: è il livello che si occupa di acquisire i dati e validarli;
2. *preparazione e "stoccaggio" dati*: è il livello che fornisce i dati agli utenti e alle applicazioni analitiche;
3. *interpretazione e analisi dati*: è il livello, ad elevato valore aggiunto, che presiede alla trasformazione dei dati in informazioni aventi valore strategico;
4. *presentazione dati*: è il livello, a basso valore aggiunto, che presiede alla presentazione finale agli utenti delle informazioni e quindi delle risposte cercate.

Integrazione dati

Con il termine "Integrazione dei dati" si definisce il processo di unione di dati provenienti da più sorgenti differenti in una vista unificata: dall'assimilazione, alla pulizia, mappatura e trasformazione dei dati, fino all'elaborazione di intelligence più facilmente fruibile da parte di coloro che vi accedono.

Anche se un'azienda riceve tutti i dati di cui ha bisogno, questi spesso risiedono in diverse sorgenti distinte. Se volessimo ad esempio ottenere una visione a 360 gradi di una persona residente in un Comune dovremmo affidarci a diverse sorgenti dati quali il sistema di gestione in uso presso il Comune, il sistema di movimentazione della popolazione, il sistema di gestione delle contribuzioni, ecc. Le informazioni provenienti da tutte queste sorgenti devono essere unificate per esigenze di analisi o interventi operativi, un compito tutt'altro che semplice per gli sviluppatori e i tecnici dei dati incaricati di farlo.

L'integrazione dei dati è il processo di trasferimento regolare dei dati da un sistema a un altro. L'integrazione può essere programmata, ad esempio mensile o trimestrale, o essere attivata da un evento. I dati sono archiviati e gestiti sia nella sorgente che nella destinazione.

Una sorgente dati è un'applicazione verticale che genera dati in maniera transazionale, sono i luoghi dove vengono generati e salvati i dati principali di un'azienda. I dati possono essere salvati in svariati formati: su file, su database, ecc.

ETL

Il processo di estrazione/trasformazione/caricamento (Extract/Transform/Load), comunemente conosciuto come ETL, è un processo di Data Integration in cui le informazioni vengono prelevate dal sistema sorgente e depositate in un Data Warehouse. Si tratta del processo permanente utilizzato nei Data Warehouse per trasformare più sorgenti di dati in informazioni coerenti e utili per finalità di analisi e business intelligence.

- **Integrazione manuale dei dati** – Si tratta semplicemente del processo tramite il quale un singolo utente raccoglie manualmente i dati necessari dalle varie sorgenti, accedendo direttamente alle relative interfacce, quindi li ripulisce e li combina in un unico Data Warehouse. Si tratta di un metodo del tutto inefficiente e incoerente, non adatto alla maggior parte delle organizzazioni, se non a quelle più piccole con risorse di dati minime.
- **Integrazione dei dati tramite middleware** – In questo metodo di integrazione viene utilizzata un'applicazione middleware che funge da mediatore, favorendo la normalizzazione dei dati e la loro consegna al pool di dati master. (Si pensi agli adattatori per vecchi elettrodomestici con spine obsolete.) Le applicazioni legacy spesso non si integrano perfettamente con quelle più recenti. Il middleware entra in gioco quando un sistema di integrazione dei dati non è in grado di accedere autonomamente ai dati di una di queste applicazioni.
- **Integrazione basata sulle applicazioni** – Metodo di integrazione in cui le applicazioni software individuano, recuperano e integrano i dati. Durante l'integrazione, il software deve rendere i dati provenienti dai vari sistemi compatibili tra loro, in modo che possano essere trasferiti da una sorgente a un'altra.
- **Integrazione con accesso uniforme** – Metodo di Data Integration incentrato sulla creazione di un front-end che fa assumere ai dati un aspetto coerente quando gli utenti vi accedono dalle varie sorgenti. I dati, tuttavia, restano nella sorgente

originale. Con questo metodo, è possibile utilizzare sistemi di gestione di database strutturati in base all'oggetto per uniformare l'aspetto di database differenti.

- **Integrazione in un sistema di storage comune** – Si tratta del metodo di integrazione dei dati più utilizzato. Una copia dei dati acquisiti dalla sorgente originale viene conservata nel sistema integrato ed elaborata per una visione unificata. Questo metodo si contrappone a quello dell'accesso uniforme, che lascia i dati nella sorgente originale. L'integrazione in un sistema di storage comune è il principio dalla base delle tradizionali soluzioni di data warehousing.

Modellazione

Le funzionalità di analisi di un data warehouse si appoggiano su una tecnologia di tipo OLAP (On-Line Analytical Processing). Il termine OLAP designa un insieme di tecniche software per l'analisi interattiva e veloce di grandi quantità di dati, che è possibile esaminare in modalità piuttosto complesse. Questa è la componente tecnologica base del data warehouse e permette agli utenti finali di eseguire analisi ad hoc di dati in molteplici dimensioni.

L'OLAP è essenzialmente un approccio ai processi decisionali che si focalizza sull'analisi dimensionale. La creazione di un database OLAP (cubo multidimensionale) consiste nell'effettuare una fotografia di informazioni (ad esempio quelle di un database relazionale) in un determinato momento e trasformare queste singole informazioni in dati multidimensionali.

Cubo Olap

Un Cubo multidimensionale permette di eseguire delle analisi veloci utilizzando delle dimensioni che definiscono un processo di business. Ogni cella del cubo rappresenta un fatto o misura, ad esempio una registrazione contabile, una multa, una decisione municipale, ecc. Le dimensioni sono i punti di vista rispetto ai quali è interessante analizzare un processo di business.

Schema a stella

La definizione di cubo multidimensionale si trasporta nella pratica con uno schema database a forma di stella, questa è la rappresentazione più semplice per il modellamento dimensionale. Al centro di uno schema di questo tipo troviamo una o più tabelle dei fatti di grandi dimensioni. Le braccia della stella sono formate dalle tabelle delle dimensioni che ci danno un punto di vista analitico da cui guardare le misure dei processi di business.

Granularità

Per granularità si intende il livello di dettaglio dei dati salvati nel datawarehouse. Più alto è il livello di dettaglio e più bassa è la granularità e viceversa. In altre parole la granularità definisce la singola riga di una tabella dei fatti, riprendendo quanto detto in precedenza il livello di granularità in un sistema contabile può essere la singola registrazione contabile.

Datawarehouse strumenti

Al giorno d'oggi esistono sempre più strumenti che permettono di entrare nel campo dell'integrazione dati e della business intelligence. Una buona guida per capire quale strumento ricopre tutte le necessità di un certo caso d'uso è il quadrante magico di Gartner. Questo strumento è utile per capire quali sono i fornitori di servizi in ambito BI e come si posizionano a livello di visione e abilità di esecuzione. Per andare sul sicuro sulla scelta di un eventuale provider di servizi di BI conviene puntare sui prodotti considerati Leaders o su prodotti costruiti a partire da questi ultimi.

Grazie all'evoluzione degli strumenti di analisi e modellazione, in certi contesti semplici che non richiedono particolari skill tecniche, si può passare dalla corporate BI, fatta da sviluppatori professionisti, alla personal BI dove è l'utente finale a gestire tutti i passaggi visti in precedenza.

Visualizzazione dei dati

Per visualizzazione dei dati si intende il tentativo di comprendere il significato dei dati inserendoli in un contesto visivo. Pattern, tendenze e correlazioni possono non risaltare in dati testuali, ma possono essere riconosciuti in maniera più diretta se presentati in visivamente.

Esistono numerosi modi per dare una forma visuale ai dati e che porti ad una rapida interpretazione delle tendenze. Grazie all'utilizzo di forme visuali è possibile approfondire le analisi andando a ricercare laddove si notano delle variazioni, questo risulta difficile utilizzando la semplice forma tabellare.

Fonti:

- <https://www.talend.com/it/resources/what-is-data-integration/>
- <https://vitolavecchia.altervista.org/definizione-e-caratteristiche-di-un-olap-on-line-analytical-processing-in-informatica/>