

Sistemi informativi

Obiettivi

I sistemi operazionali hanno fatto nascere l'esigenza di nuovi strumenti per i livelli dirigenziali dell'azienda.

I sistemi operazionali hanno dei grandi limiti per quanto riguarda la reportistica:

- I dati estratti sono molto statici.
- La difficoltà e la lentezza dei processi di interrogazione
- L'incompletezza

I vantaggi dell'essere altamente customizzabili permettono di ospitare molti tipi di dato, porta molto overhead nei fogli di calcolo:

- La macchinosità dell'estrazione dei dati
- Lo scarso controllo sui dati
- Molti strumenti di calcolo personalizzati e incontrollati
- L'enorme complessità strutturale

Oltre a tutto questo i sistemi operazionali non sono ottimizzati per analizzare enormi quantità di dati e ottimizzano le procedure interattive.

I sistemi informativi devono pertanto mettere a disposizione dell'utente un ambiente in cui analizzare i dati sia facile.

- La base di dati deve comprendere l'intero insieme dei dati interessanti per l'azienda.
- I dati devono essere riorganizzati per supportare al meglio le attività di analisi.
- La base di dati deve essere strutturata in modo semplice
- Le fonti dei dati devono essere integrate
- Devono essere disponibili strumenti di analisi
- Gli strumenti di analisi devono supportare grandi moli di dati

Concetti generali

- **Data warehouse:** è la base dei dati informativi
- **Data warehousing:** è l'insieme delle attività che definiscono, costruiscono e mantengono la struttura
- **Decision Support System (DSS):** i sistemi informatici che aiutano nel processo decisionale come il supporto all'estrazione delle informazioni
- **Data Mining:** l'insieme degli strumenti e delle tecniche usate per estrarre dati "nascosti"
- **Business intelligence:** l'insieme delle attività volte a estrarre informazioni dai dati di business come quelli generati dai processi operativi
- **Knowledge management:** i sistemi di gestione della conoscenza aziendale che non è mantenuta solo dai dati strutturati/strutturizzabili
- **Big Data:** sono i flussi continui di informazioni non generate dai processi, e che non richiedono costi aggiuntivi, per essere generate.

Il trattamento di questi dati ammette soluzioni diverse con elaborazione distribuita e algoritmi complessi.

Si è passati dal dato che permette di osservare gli andamenti al dato dato che può essere usato come materiale.

Caratteristiche dei dati informativi strutturati

Ovviamente le informazioni operazionali differiscono di molto da quelle informative, come per esempio in:

- **Finalità:** fornire al management il substrato di informazioni per conoscere l'azienda.
Descrivere il passato per aiutare a identificare i problemi e le loro cause, suggerire i cambiamenti da apportare per evitarli.
- **Struttura:** i dati sono articolati attorno a funzioni, procedure e eventi.
- **Utenza e strumenti:** sono utilizzati prevalentemente da personale esecutivo e manager.
- **Storicità dei dati:** vengono mantenute le informazioni con un arco temporale maggiore.
- **Dettaglio:** servono informazioni aggregate, è preferibile che esistano diversi livelli di aggregazione, dalla sintesi massima al dettaglio massimo.
- **Tipo di accesso:** è solitamente in sola lettura con aggiornamenti applicati nei periodi di inattività.

Modello multidimensionale

Uno stesso evento può essere analizzato mettendolo in relazione con diversi soggetti, da questa visione si intuisce che un dato può essere rappresentato come un punto in un *ipercubo*.

All'interno di questi ipercubi i punti verranno chiamati *fatti* mentre le coordinate *misure*, inoltre le coordinate di ciascun elemento costituiscono le *dimensioni* di analisi dai fatti.

1. Fatti

Sono gli eventi che accadono all'interno dell'azienda e si ha interesse a misurare.

Ogni fatto ha:

- Attributi che lo collocano nel tempo
- Almeno un dato numerico
- Informazioni descrittive

Un fatto è quindi identificato univocamente da una n-upla di coordinate.

2. Misure

Sono caratteristiche numeriche del fatto che ne descrive aspetti quantitativi rilevanti per l'analisi. Le misure possono essere:

- **Effettive:** contenute nella base di dati informazionale
- **Calcolate:** a run time utilizzando i valori già presenti

I fatti elementari possono essere aggregati, ovvero si può ottenere un insieme di dati sintetici raggruppando le misure con caratteristiche comuni.

Il processo di aggregazione è fatto con appositi operatori com: somma, media, max, min, ...

Gli operatori devono essere scelti con attenzione, esistono infatti operatori aggregabili in alcune dimensioni e non su altre.

Il caso più comune è l'aggregabilità con l'operatore somma detta *additività*.

Le misure possono quindi essere divise in:

- **Misure di livello:** che esprimono valori propri del fatto nel momento in cui è registrato. Non sono mai additive rispetto alla dimensione del tempo.
- **Misure unitarie:** ovvero i valori relativi a dei soggetti, valido nel momento in cui viene registrato. Non sono mai additive.
- **Misure di flusso:** valori propri del fatto rapportati ad un intervallo temporale di riferimento.
Additive lungo qualsiasi dimensione.

Esiste infine una categoria speciale di misure, le *misure proprie*.

Queste sono tutte le misure di cui ci interessa è che siano verificate o meno, quindi la presenza/assenza del fatto.

Solitamente si esprimono con valori booleani oppure 0/1.

Caratteristiche strutturali

- **Multidimensionalità:** ovvero il numero di parametri che occorre fornire per poter estrapolare un'informazione specifica.
- **Granularità:** il grado di aggregazione delle informazioni, se l'ipercubo ha granularità minima significa che ogni fatto corrisponde ad un solo dato di origine.
Spesso per limitare le dimensioni del warehouse i dati hanno già un'aggregazione di base.
- **Arco temporale:** il periodo temporale coperto dai dati presenti nel warehouse.
- **Profondità storica:** i dati possono variare nel tempo, questo parametro identifica quante informazioni storiche su un parametro si posseggono.

Caratteristiche funzionali

- **Integrazione dei dati:** la coerenza che hanno dati provenienti da diverse applicazioni e da differenti basi di dati.
- **Accessibilità:** facilità d'uso e velocità di risposta.
- **Flessibilità e sintesi:** capacità di assecondare l'utente nel fare le richieste, mettere in relazione misure con criteri non prefissati.
- **Correttezza:** aderenza con la realtà.

- **Completezza:** la capacità di rappresentare tutti gli eventi importanti, si riferisce sia alle istanze dei dati sia al modello.

Data warehouse e data mart

Il data warehouse è definito come il magazzino unico, completo e consistente delle informazioni aziendali.

Contiene articoli di interesse aziendale dai quali è possibile estrarre gli ipercubi, per questa sua utilità può raggiungere dimensioni molto elevate.

Operare sull'intero data warehouse sarebbe impensabile, vengono così introdotti i *data mart*, ovvero dei piccoli data warehouse tematici limitati ai fatti che riguardano un'area.

I data mart possono essere definiti con particolari proprietà:

- Estensioni temporali ridotte.
- Granularità minore.