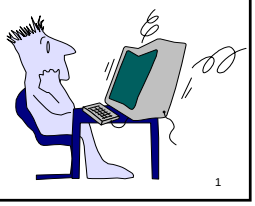
Modellazione Concettuale



Fasi di progettazione del DW

Fase	Ingresso	Uscita	Figure
Analisi e riconciliazione delle fonti dati	Schemi delle sorgenti	Schema riconciliato	Progettista; amministratori del db oper.
Analisi dei requisiti	Schema riconciliato	Fatti; carico di lavoro (I)	Progettista; utenti finali
Progettazione concettuale	Schema riconciliato; fatti; carico di lavoro (I)	Schemi di fatto	Progettista; utenti finali
Raffinamento carico di lavoro; validazione schema concettuale	Schemi di fatto; carico di lavoro (I)	Carico di lavoro (II); schemi di fatti validati	Progettista; utenti finali
Progettazione logica	Schemi di fatto; modello logico; carico di lavoro	Schema logico	Progettista
Progettazione dell'alimentazione	Schemi delle sorgenti; schema riconciliato; schema logico	Procedure di alimentazione	Progettista; amministratori del db oper.
Progettazione fisica	Schema logico; carico di lavoro	Schema fisico	Progettista

Modellazione Concettuale

- **Progettazione Concettuale**
  - Ha lo scopo di rappresentare le specifiche informali della realtà in una descrizione:
  - **Formale**: espressa in modo non ambiguo ma adeguato a catturare le caratteristiche fondamentali del mondo da descrivere
  - **Integrata**: la descrizione si riferisce alla totalità dell'ambiente (non settoriale)
  - **Indipendente** dalla realizzazione fisica
  - **Leggibile**: è possibile usare questa rappresentazione per comunicare con l'utente finale
  - Si concentra sul **COSA** rappresentare in una base di dati
  - Usa un modello concettuale per produrre uno schema concettuale

Modello concettuale

- Mentre è universalmente riconosciuto che un DW si appoggia sul modello multidimensionale, non c'è accordo sulla metodologia di progetto concettuale da seguire.
- Il modello Entity/Relationship è molto diffuso come formalismo per la descrizione concettuale e la documentazione dei sistemi informativi relazionali, ma *non è molto adatto per modellare il DW*

→ **Viene introdotto un apposito modello: DIMENSIONAL FACT MODEL (DFM)**

Dimensional Fact Model (DFM)

- Il ***Dimensional Fact Model*** (DFM) è un modello concettuale grafico concepito per fungere da supporto alla progettazione di data mart / DW.
- Il DFM è indipendente dal modello logico utilizzato (multidimensionale o relazionale).
- La rappresentazione generata dal DFM è detta Schema Dimensionale (DS) e consiste in un insieme di Schemi di fatto (FS) con i quali si modellano **fatti, misure, dimensioni e gerarchie.**

Dimensional Fact Model (DFM)

- Il DFM è un modello concettuale grafico per data mart, pensato per:
  - ✓ supportare efficacemente il progetto concettuale;

Dimensional Fact Model (DFM)

- Il DFM è un modello concettuale grafico per data mart, pensato per:
  - ✓ supportare efficacemente il progetto concettuale;
  - ✓ creare un ambiente su cui formulare in modo intuitivo le interrogazioni dell'utente;

Dimensional Fact Model (DFM)

- Il DFM è un modello concettuale grafico per data mart, pensato per:
  - ✓ supportare efficacemente il progetto concettuale;
  - ✓ creare un ambiente su cui formulare in modo intuitivo le interrogazioni dell'utente;
  - ✓ permettere il dialogo tra progettista e utente finale per raffinare le specifiche dei requisiti;

Dimensional Fact Model (DFM)

- Il DFM è un modello concettuale grafico per data mart, pensato per:
  - ✓ supportare efficacemente il progetto concettuale;
  - ✓ creare un ambiente su cui formulare in modo intuitivo le interrogazioni dell'utente;
  - ✓ permettere il dialogo tra progettista e utente finale per raffinare le specifiche dei requisiti;
  - ✓ creare una piattaforma stabile da cui partire per il progetto logico (*indipendentemente dal modello logico target*);

Dimensional Fact Model (DFM)

- Il DFM è un modello concettuale grafico per data mart, pensato per:
  - ✓ supportare efficacemente il progetto concettuale;
  - ✓ creare un ambiente su cui formulare in modo intuitivo le interrogazioni dell'utente;
  - ✓ permettere il dialogo tra progettista e utente finale per raffinare le specifiche dei requisiti;
  - ✓ creare una piattaforma stabile da cui partire per il progetto logico (*indipendentemente dal modello logico target*);
  - ✓ restituire una documentazione a posteriori espressiva e non ambigua.

FATTO

COS'E' UN FATTO?

COS'E' UNA MISURA?

COS'E' UNA DIMENSIONE?

FATTO

- Un ***fatto*** è un concetto di interesse per l'organizzazione ed è descritto da un insieme di *misure*.
- *Vedremo che il "fatto" esprime un'associazione molti-a-molti tra le dimensioni di analisi*

MISURA

Le ***misure*** sono attributi tipicamente numerici che descrivono aspetti quantitativi di interesse per l'analisi.

Ad esempio, ogni vendita è misurata dal suo incasso o dalla quantità venduta.

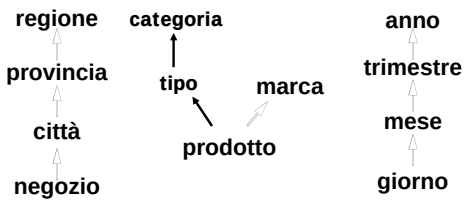
DIMENSIONE

- Le ***dimensioni*** sono le coordinate di analisi rispetto alle quali viene analizzato un fatto.
- Definiscono la granularità minima di rappresentazione dei fatti;

Es.: dimensioni tipiche per il fatto vendite sono prodotto, negozio, data.

Dimensioni e gerarchie di livelli

- Ciascuna dimensione è organizzata in una **gerarchia** che rappresenta i possibili livelli di aggregazione per i dati



GERARCHIA

Qual è il significato di una gerarchia?

Una *gerarchia* determina come le istanze di un fatto possono essere aggregate in modo significativo per il processo decisionale.

GERARCHIA

Come si rappresenta una gerarchia?

Algebricamente/graficamente si rappresenta con un **albero orientato**

- i cui **nodi** sono attributi dimensionali *(cioè attributi che costituiscono dimensioni di analisi)*
- i cui **archi** modellano relazioni molti-a-uno tra coppie di attributi dimensionali

GERARCHIA

Esempio:

sulla dimensione Prodotto potremo avere:

prodotto → tipo, tipo → categoria, categoria → reparto

e parallelamente

prodotto → marca

GERARCHIA

DIPENDENZA FUNZIONALE

Ogni arco di gerarchia modella una dipendenza funzionale.

Poiché per le dipendenze funzionali vale la **proprietà transitiva**, ogni cammino orientato all'interno della gerarchia rappresenta a sua volta una dipendenza funzionale tra l'attributo di partenza e quello di arrivo, cioè:

Se prodotto → tipo, tipo → categoria

allora prodotto → categoria

GERARCHIA

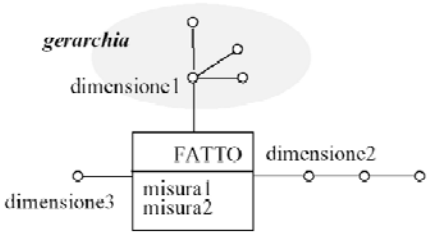
La presenza di una dipendenza funzionale stabilisce un'associazione molti-a-uno tra i valori assunti da ogni attributo dimensionale ad un livello e i valori assunti al livello successivo

Es. esiste un'associazione molti-a-uno tra i **prodotti** ed i loro **tipi**

(un prodotto è di un solo tipo mentre un tipo identifica un insieme di prodotti)

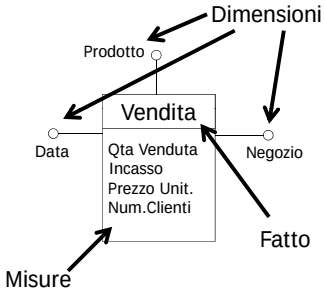
- La radice di una gerarchia dimensionale individua la granularità della dimensione.
- Agli altri attributi dimensionali corrispondono granularità via via crescenti.

Gerarchia

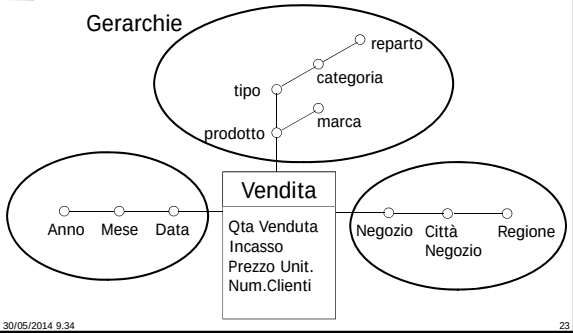


Un fatto esprime una associazione molti-a-molti tra le dimensioni.

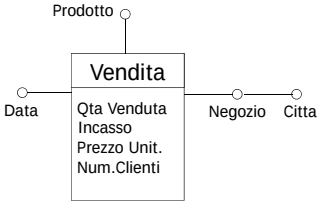
Schema di fatto



Schema di fatto



Schema di fatto

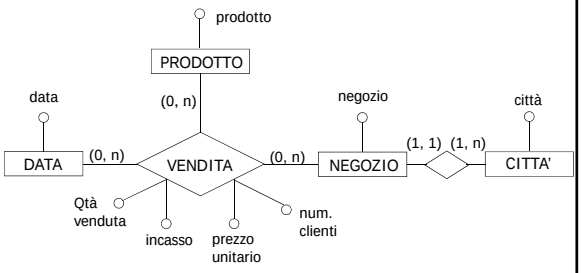


Lo possiamo rappresentare tramite lo schema E/R ?

ESERCIZIO

- Rappresentare lo schema di fatto precedente tramite il modello E/R
- Trovare la rappresentazione relazionale del modello E/R precedente

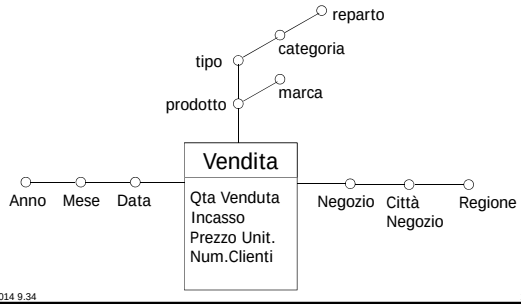
Schema E/R



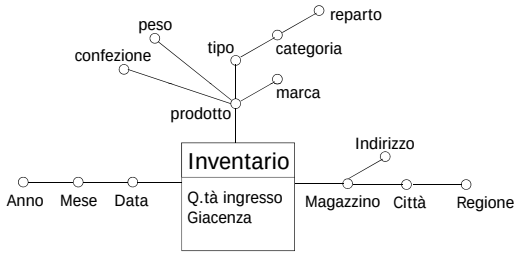
Schema E/R

- RIASSUMENDO:
  - Lo schema E/R corrispondente ad uno schema di fatto, consiste in un'associazione **n-aria** che modella la relazione tra il "fatto" e le "dimensioni di analisi".
  - Le misure sono attributi dell'associazione.
  - Ad ogni attributo dimensionale corrisponde un'entità che ha quell'attributo come identificatore.
  - Ogni arco delle gerarchie è rappresentato da un'associazione molti-a-uno.

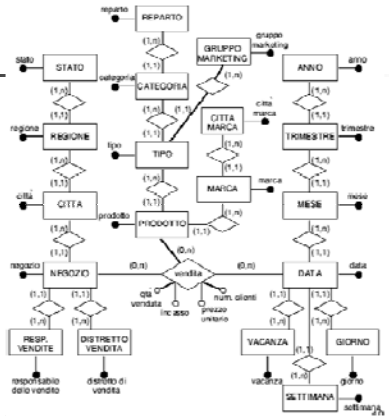
Schema di fatto



Schema di fatto



Schema di fatto

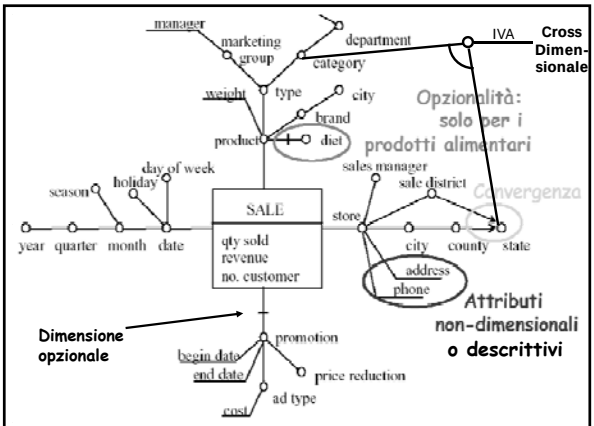


Attributi

- Li possiamo classificare in:
  - Attributi dimensionali (che abbiamo visto finora)
  - Attributi descrittivi
  - Attributi cross-dimensionali

ATTRIBUTI DESCRITTIVI

- Alcuni attributi sono detti ***attributi non dimensionali o descrittivi*** e non possono essere utilizzati per aggregare i dati.
- Spesso sono legati da un'associazione one-to-one con gli attributi dimensionali (ad es., associazione indirizzo-cliente).
- Nel DFM sono sempre nodi foglia e sono rappresentati con una linea orizzontale.
- Possono anche essere connessi direttamente al fatto se descrivono un evento primario, senza peraltro poter essere usati per aggregazioni o misure.



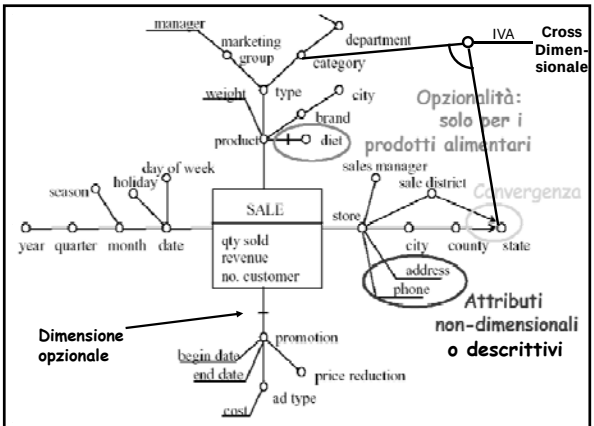
ATTRIBUTI DESCRITTIVI

- Un attributo descrittivo può essere anche connesso al fatto se descrive un evento primario, ma non è possibile utilizzarlo né per identificare i singoli eventi, né per effettuare i calcoli.
- Per collegare un attributo descrittivo al fatto occorre necessariamente che esso assuma, in corrispondenza di un attributo primario, un valore univoco

Attributi descrittivi connessi direttamente ai fatti

ATTRIBUTI CROSS-DIMENSIONALE

- Si tratta di un attributo il cui valore è determinato dalla combinazione di due o più attributi dimensionali anche appartenenti a gerarchie distinte.
- Es. Iva:
  - Dipende sia dalla categoria di appartenenza del prodotto e dallo Stato in cui il prodotto viene venduto (oppure ad es. dal cliente che può essere esente)
  - In tal caso la combinazione viene rappresentata da un attributo cross-dimensionale che unisce con un arco circolare gli archi che la determinano



Convergenza

- Il concetto riguarda la struttura delle gerarchie
- Le gerarchie possono non essere "veri alberi"
- Un attributo dimensionale può essere punto di convergenza di due o più cammini distinti.

Convergenza

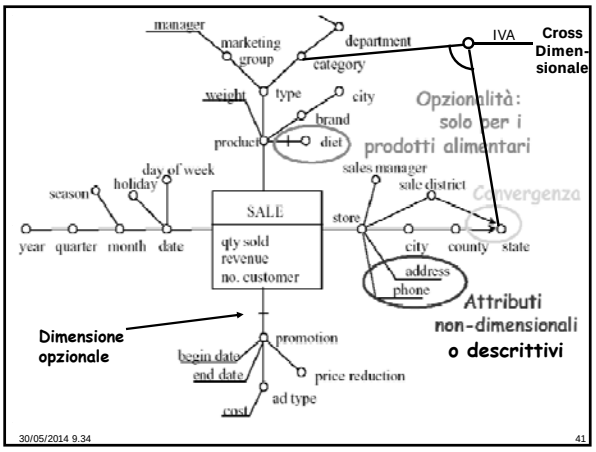
ESEMPIO:

negozio → città → regione → stato

negozio → distretto → stato

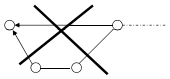
→ convergenza su stato

- Viene denotata tramite due o più archi che terminano nello stesso attributo dimensionale

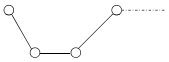


Convergenze ridondanti

- Esempio: uno dei percorsi alternativi non comprende attributi intermedi ➔ non ha ragione di esistere



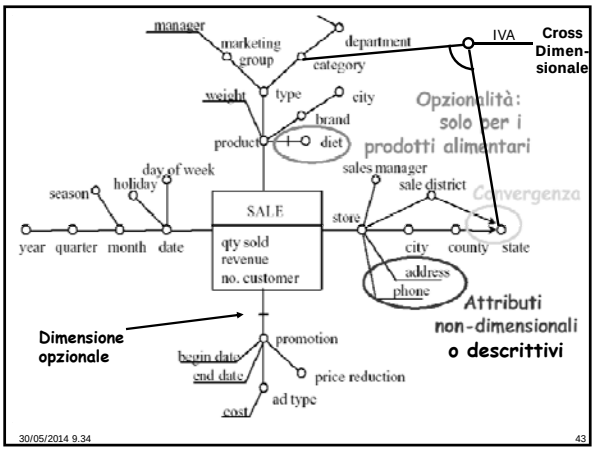
Convergenza ridondante



Convergenza non ridondante

La convergenza è ovvia in virtù della transitività delle dipendenze funzionali

30/05/2014 9.34 42



Opzionalità

- Ci possono essere rami opzionalmente presenti corrispondenti a dimensioni o attributi opzionali
- Gli archi opzionali sono denotati con un trattino

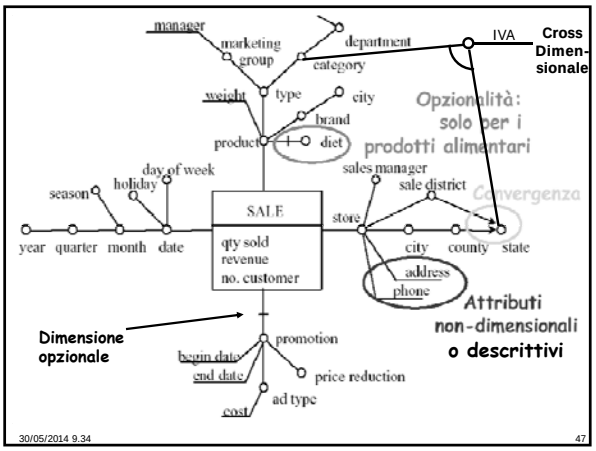
30/05/2014 9.34 44

Opzionalità

- Per le dimensioni:
  - si adopera l'opzionalità per modellare situazioni in cui un'associazione rappresentata nello schema di fatto non è definita per un certo sottoinsieme di eventi.

Opzionalità

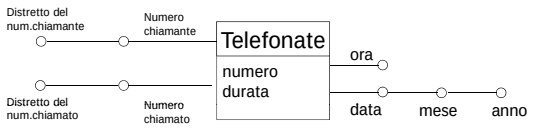
- Se r è l'arco opzionale si distinguono due casi notevoli:
  - r determina un attributo dimensionale a tramite un attributo b (ossia b→a)
    - Allora significa che a ed i suoi discendenti sono indefiniti in corrispondenza di uno o più valori di b (es. attributo "dieta" assume un valore solo per i prodotti alimentari)
  - r determina una dimensione d
    - Esistono alcuni eventi primari identificati solo dalle altre dimensioni (es. la dimensione "promozione" nello schema delle vendite è opzionale, ossia per certe combinazioni "prodotto-data-negozio" il valore della promozione sarà "nessuna promozione")



Gerarchie Condivise

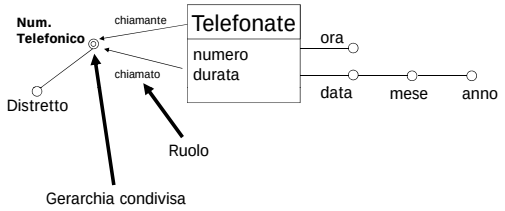
- Accade spesso che intere porzioni di gerarchie vengano replicate due o più volte
  - Es. della gerarchia geografica ← (es. città della marca, città del negozio che si gerarchizzano allo stesso modo (provincia, regione,...)
- Si introduce una notazione grafica abbreviata che evidenzia la condivisione della gerarchia

Schema di fatto



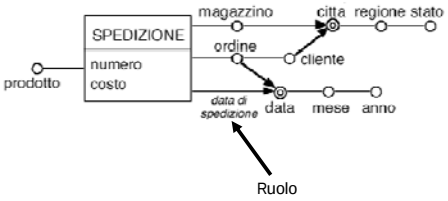
Schema di fatto senza condivisioni

Schema di fatto



Schema di fatto con condivisioni

Schema di fatto

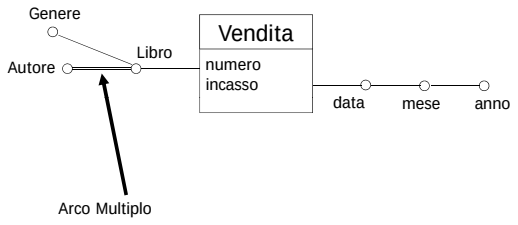


Schema di fatto con condivisioni

Archi Multipli

- Siano dati due attributi a e b di una gerarchia dimensionale
- Quando tra l'attributo a e l'attributo b esiste una associazione molti-a-molti si introducono i cosiddetti "archi multipli"

Schema di fatto



Misure ed operatori di aggregazione

Aggregazioni su misure

- Il workload attivabile su uno schema di fatto consiste in interrogazioni che comportano valutazioni di un fenomeno tramite le misure di interesse del fatto.
- Le valutazioni si fanno in base alle gerarchie dimensionali per cui occorre poter AGGREGARE e DISAGGREGARE, aumentando o diminuendo il livello di granularità delle interrogazioni stesse.

Aggregazioni su misure

- Ecco che assumono particolare importanza gli operatori di AGGREGAZIONE
- Tipici operatori sono:
  - SUM, AVG, MIN, MAX, ...

Aggregazioni su misure

- In questo contesto, assume particolare importanza il concetto di additività.
- Ci sono misure su cui è possibile applicare ADDITIVITA'; altre in cui non è possibile.
- Vediamo di cosa si tratta.....

Additività

DEF.

- Una misura è detta ADDITIVA su una dimensione se i suoi valori possono essere aggregati lungo la corrispondente gerarchia tramite l'operatore di SOMMA, altrimenti si dice NON-ADDITIVA.
- Una misura non-additiva è NON-AGGREGABILE se nessun operatore di aggregazione può essere usato su di essa.

- Per capirsi è necessario definire una classificazione delle misure.

Additività

- E' possibile classificare le misure in categorie diverse su cui è possibile effettuare tipi diversi di aggregazione, ossia per le quali sono validi certi operatori di aggregazione.
- Le categorie sono:
  - Misure di flusso ← Misura in un periodo
  - Misure di livello ← Misura in particolari istanti
  - Misure unitarie ←

Additività

- Misure di flusso:
  - Si riferiscono ad un periodo, al cui termine vengono valutate in modo cumulativo
  - Es.
    - il numero di prodotti venduti in un giorno
    - l'incasso mensile
    - il numero di nati in un anno

Additività

- Misure di livello:
  - vengono valutate in particolari istanti di tempo ed esprimono una valutazione di un livello
  - Es.
    - il numero di prodotti in inventario
    - il numero di abitanti di una città

Additività

- Misure unitarie:
  - vengono valutate in particolari istanti di tempo e sono espresse in termini relativi
  - Es.
    - il prezzo unitario di un prodotto
    - La percentuale di sconto
    - Il cambio di valuta

Additività

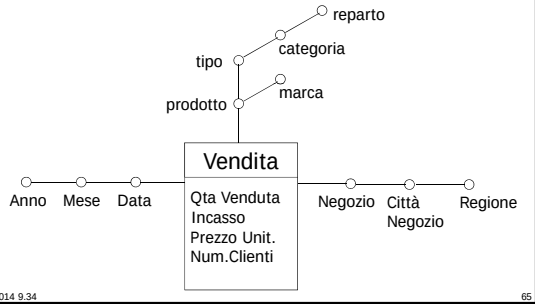
- In linea di massima abbiamo che:
  - le misure di flusso sono additive su tutte le dimensioni
  - le misure di livello sono non-additive sulle dimensioni temporali
  - le misure unitarie sono non-additive su tutte le dimensioni

Additività

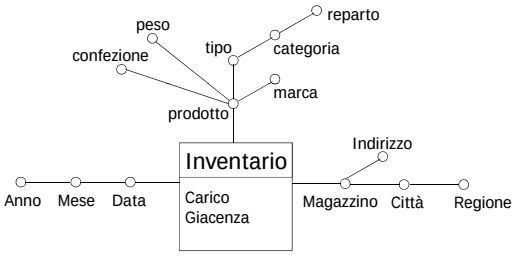
Operatori di aggregazione validi per i tre tipi di misure

	Gerarchie temporali	Gerarchie non temporali
Misure di flusso	**Sum, Avg, Min, Max**	**Sum, Avg, Min, Max**
Misure di livello	**Avg, Min, Max**	**Sum, Avg, Min, Max**
Misure unitarie	**Avg, Min, Max**	**Avg, Min, Max**

Additività



Additività



Additività

ESEMPI

- **Misura additiva**:
  - la quantità venduta (nello schema delle vendite per qualunque dimensione)
  - la giacenza di inventario (nello schema dell'inventario per le dimensioni magazzino e prodotto)
- **Misura non additiva di tipo livello**:
  - la giacenza di inventario (nello schema dell'inventario nel caso della dimensione tempo)
- **Misura non additiva di tipo unitario**:
  - il prezzo unitario (nello schema delle vendite), lo sconto, il cambio di valuta

Additività

ESEMPIO

- **Misura non aggregabile**:
- Supponiamo di valutare il N° clienti contando gli scontrini emessi in un certo giorno (data) e negozio. Se salgo a livello di "Reparto" nella gerarchia "Prodotto" non c'è modo di desumere il numero totale di clienti con qualche funzione di aggregazione.

  Poiché lo stesso scontrino può includere più prodotti, sommare o mediare i numeri di clienti per due o più prodotti porterebbe a delle inconsistenze

  ➔ *Il numero di clienti sulla dimensione prodotto (nello schema delle vendite) è una misura non aggregabile*

Additività

ESEMPIO

- **Misura non aggregabile**:
  Nel caso precedente, si noti come l'associazione tra scontrini e prodotti è di tipo molti-a-molti invece che molti-a-uno

Additività

Tabella delle additività per lo schema delle vendite

	Prodotto	Data	Negozio	Promozione
Q.tà venduta	SUM	SUM	SUM	SUM
Incasso	SUM	SUM	SUM	SUM
Num.Clienti	—	SUM	SUM	SUM
Prezzo Unitario	AVG	AVG	AVG	AVG

Misure Derivate

Misure derivate

Con gli esempi che seguono vogliamo evidenziare l'importanza di misure di supporto per la determinazione di misure derivate:

queste ultime sono calcolate in funzione di altre; questa loro dipendenza da altre misure comporta dei **rischi di inconsistenza** man mano che si procede nel processo di aggregazione.

Misure derivate

Nell'esempio seguente il valore dell'incasso è calcolabile a partire da una "vista" primaria che contiene 'prezzo unitario' e 'qta venduta'.

MA: operando su dati aggregati, ossia passando ad una "vista" secondaria ..... questa possibilità può perdersi !!!

Esempio

tipo	prodotto	Qtà	Prezzo
Latticino	Latte A	5	1,0
Latticino	Latte B	7	1,5
Bibita	Colissima	9	0,8

Incasso
5,0
10,5
7,2

Totale: 22,7

Aggregazione sul tipo

SUM SUM

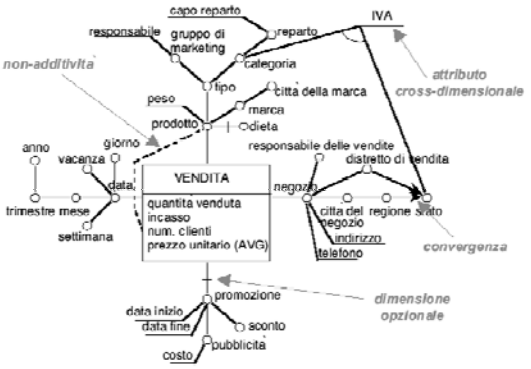
tipo	Qtà	Prezzo	Qtà*prezzo
Latticino	12	2,5	30,0
Bibita	9	0,8	7,2

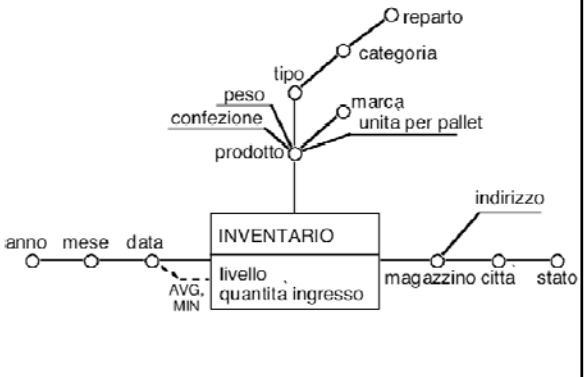
Totale: 37,2

PERCHE':

La somma dei prodotti non è uguale al prodotto della somma

(P1*Q1) + (P2*Q2) <> (P1+P2) * (Q1+Q2)





Misure derivate

Nell'esempio precedente

➔ ➔ ➔

l'attributo derivato (es. incasso=prezzo unitario*qtà venduta) deve essere memorizzato esplicitamente per evitare inconsistenze

Date	Inventory level
1/1/1999	100
10/2/1999	200
30/4/1999	60
5/6/1999	85
18/7/1999	125
31/12/1999	110
	113,33

Aggregare per quadrimestre determinando il livello di inventario medio ossia la giacenza media

Misura di supporto

Date	Inventory level	Count	Inventory level
4/1999	120	3	360
8/1999	105	2	210
12/1999	110	1	110

Media totale 111,66 113,33

Misure di supporto

Possono essere necessarie per calcolare dati aggregati a partire da altri dati aggregati

Nell'esempio precedente il valore della giacenza media totale su una vista aggregata ha bisogno del "supporto" della misura "conteggio" affinchè il calcolo sia corretto.

Operatori

Anche l'ordine di applicazione degli operatori di aggregazione può portare a risultati diversi:

→ è importante anche l'ordine di applicazione

Esempio

Città	Negozio	Prodotto	Qtà
RM	N1	P1	5
RM	N1	P2	6
RM	N2	P1	2
RM	N2	P2	8
MI	N3	P1	7
MI	N3	P2	9
MI	N4	P1	5
MI	N4	P2	3

Esempio

- Calcolare una vista che mostri, per ogni città, la media delle minime vendite di ogni negozio
- OPERATORI di aggregazione richiesti: MIN, AVG
- Supponiamo di applicare prima MIN, poi AVG

Città	Q3
RM	3.5
MI	5

Città, AVG(**Q1**) AS Q3

Città	Negozio	Q1
RM	N1	5
RM	N2	2
MI	N3	7
MI	N4	3

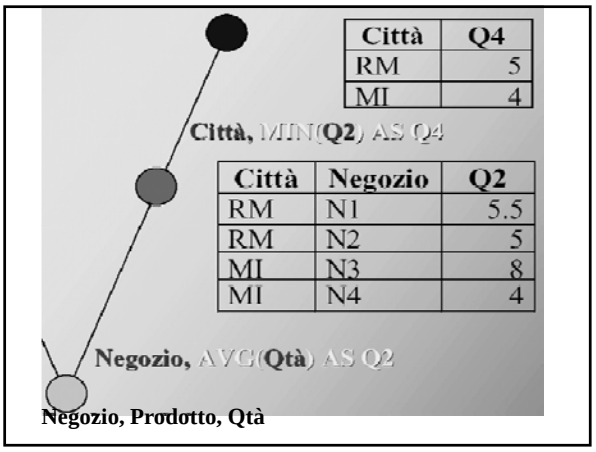
Negozio, MIN(**Qtà**) AS Q1

Negozio, Prodotto, Qtà

Esempio

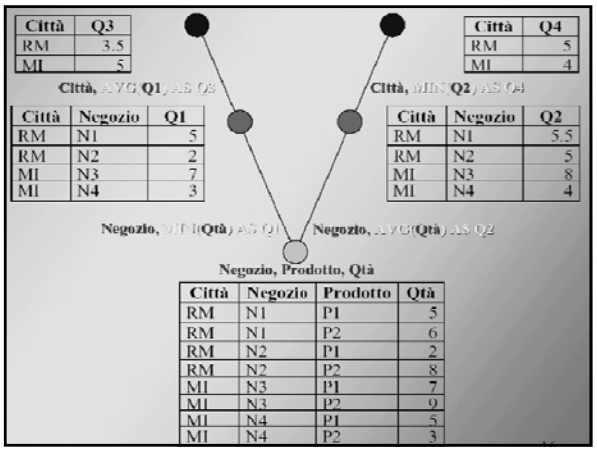
Esempio

- Calcolare una vista che mostri, per ogni città, il minimo della media delle vendite dei diversi negozi.
- OPERATORI di aggregazione richiesti: MIN, AVG
- Supponiamo di applicare prima MIN, poi AVG



Città	Q4
RM	5
MI	4

Città	Negozio	Q2
RM	N1	5.5
RM	N2	5
MI	N3	8
MI	N4	4



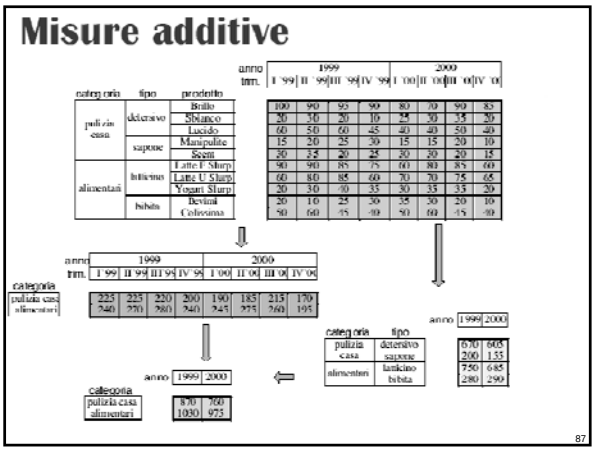
Città	Q3
RM	3.5
MI	5

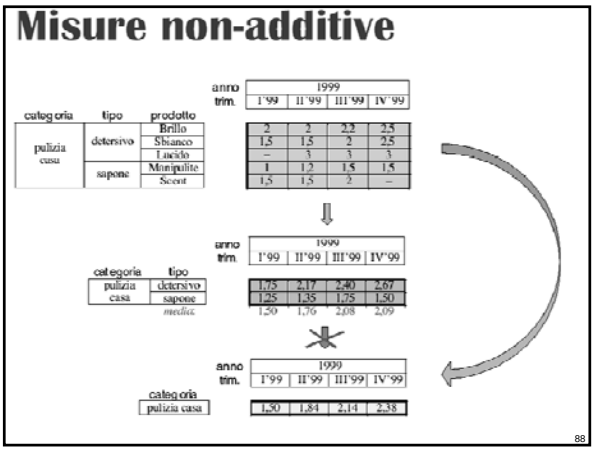
Città	Q4
RM	5
MI	4

Città	Negozio	Q1
RM	N1	5
RM	N2	2
MI	N3	7
MI	N4	3

Città	Negozio	Q2
RM	N1	5.5
RM	N2	5
MI	N3	8
MI	N4	4

Città	Negozio	Prodotto	Qtà
RM	N1	P1	5
RM	N1	P2	6
RM	N2	P1	2
RM	N2	P2	8
MI	N3	P1	7
MI	N3	P2	9
MI	N4	P1	5
MI	N4	P2	3





MORALE

1. Occorre predisporre eventuali misure di supporto oppure misure derivate che dovranno essere incluse nelle viste

   ➔ proliferazione dei campi nelle fact table

2. Dovrà essere preso in considerazione l'ordine di applicazione degli operatori di aggregazione

Classificazione degli operatori

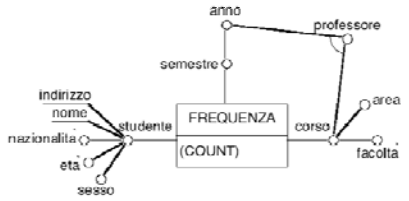
- Distributivi: permettono di calcolare dati aggregati a partire da dati parzialmente aggregati
  - Es.: count(), sum(), min(), max().
- Algebrici: richiedono un numero finito di informazioni aggiuntive (misure di supporto) per calcolare dati aggregati a partire da dati parzialmente aggregati
  - Es.: avg() ha bisogno del count()
- Olistici: non permettono di calcolare dati aggregati a partire da dati parzialmente aggregati tramite un numero finito di misure di supporto.
  - Es.: mediana

Classificazione degli operatori

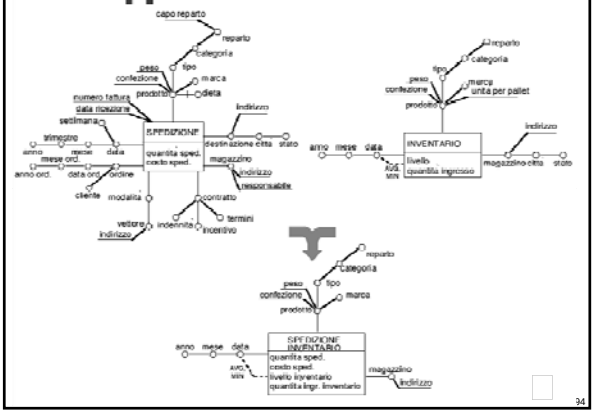
- Attualmente gli AGGREGATE NAVIGATOR dei sistemi commerciali, ossia i moduli preposti a riformulare le interrogazioni OLAP sulla "migliore" vista disponibile, gestiscono soltanto l'aggregazione tramite operatori distributivi
- Attività di altro tipo (es. Uso di misure di supporto) potranno essere coinvolte solo nelle interrogazioni scritte direttamente dall'utente

Schemi di fatto vuoti

- Uno schema di fatto si dice **vuoto** se non ha misure
  - ✓ In questo caso, il fatto registra solo il verificarsi di un evento



Sovrapposizione di schemi

- Una *query di drill-across* mette in relazione misure appartenenti a fatti diversi
- La presenza di query di drill-across richiede di definire quali misure e quali pattern di aggregazione possano essere utilizzati quando si interrogano 2 o più schemi di fatto. In uno schema sovrapposto:
  - ✓ Le misure sono l'unione delle misure presenti nei vari schemi
  - ✓ Le gerarchie includono solo gli attributi presenti in entrambe le gerarchie
  - ✓ Il dominio di ogni attributo è l'intersezione del dominio dei corrispondenti attributi sugli schemi di origine

Sovrapposizione di schemi