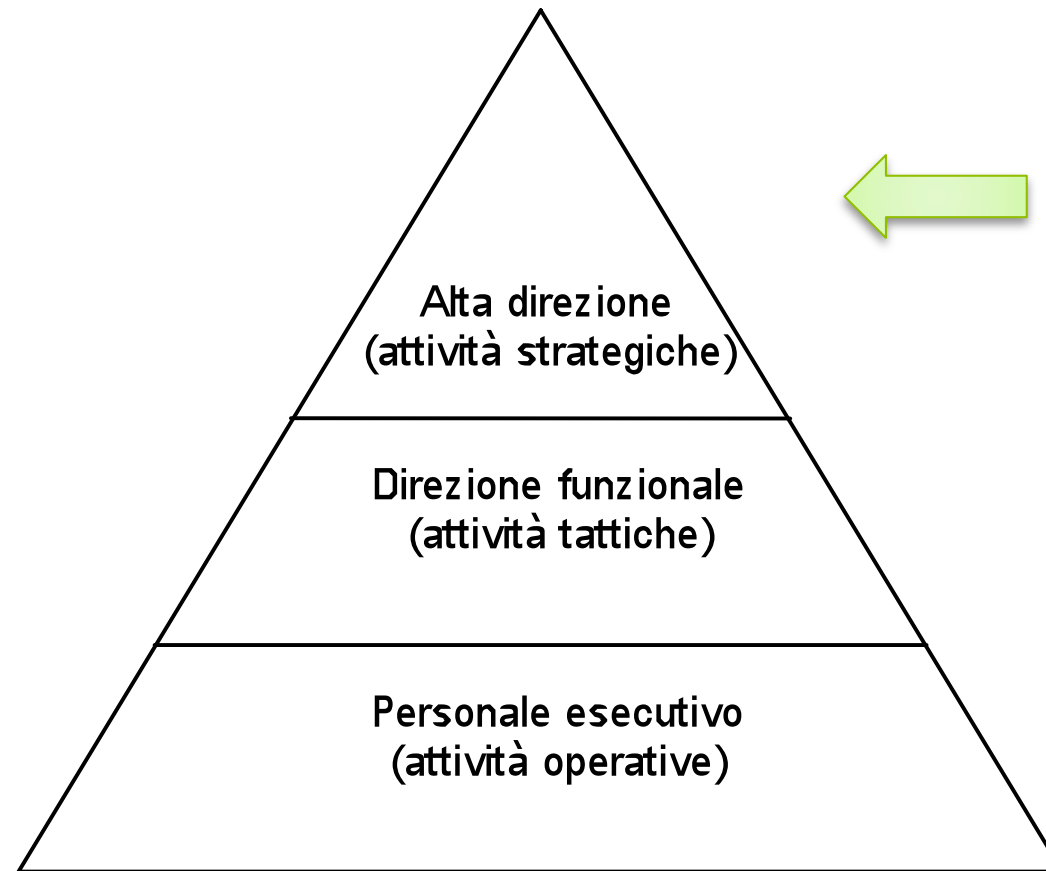


Filippo Geraci



# **I SISTEMI INFORMAZIONALI (DIREZIONALI)**

# Piramide di Antony (MEMO)



Sistema informativo  
informazionale



# Tattica e strategia (MEMO)

- **Strategia:** descrizione di un piano d'azione di lungo termine usato per impostare e successivamente coordinare le azioni tese a raggiungere uno scopo predeterminato.
- **Tattica:** metodo utilizzato per conseguire determinati obiettivi
- La strategia rappresenta il "cosa" fare mentre la tattica il "come" farlo.



# Obiettivi

- Sfruttare il patrimonio dei dati operativi per l'estrazione di informazioni utili nel processo decisionale
- Arricchire i dati operativi con altre fonti di informazione esterne all'azienda
- Superare i limiti dei primi strumenti di supporto alle decisioni
  - Report
  - Fogli di calcolo

# Limiti degli approcci tradizionali

- Capacità di elaborazione di modeste quantità di dati
- Reporting:
  - Staticità (per ogni approfondimento serve un nuovo report)
  - Parzialità (Mancanza di informazioni storiche)
  - difficoltà e lunghezza dell'iter di realizzazione



- Fogli di calcolo
  - macchinosità (difficoltà nell'estrarre i dato in riferimento alla struttura del db)
    - ERP esportano spesso viste del database in formato xls
  - scarsa controllabilità dei dati
  - limitatezza dello spazio dei dati



# Interrogazioni sui dati - sistemi operazionali

- Puntuali
- Prefissate, dipendenti dalle procedure implementate
- Piccole quantità di dati
- Operatore guidato dall'interfaccia del sistema
- Esempi:
  - Che tipo di modalità di pagamento è stata concordata per questo cliente?
  - Cosa deve produrre oggi il centro di lavoro I0?
  - Di cosa ho bisogno per produrre 12 sedie modello Tamigi blu?
  - Quali fatture fornitore devo ancora pagare?



# Interrogazioni sui dati - sistemi informazionali

- Complesse
- Grandi quantità di dati
- Guidate dal ragionamento di chi le usa
- Interfacce intuitive che permettano di esprimerne concetti di varia natura
- Esempi:
  - Di quanto è aumentato il margine operativo di questa linea di prodotti rispetto allo scorso anno?
  - Esiste una correlazione tra la zona in cui risiede un cliente, la sua scolarità e la sua propensione all'acquisto di prodotti di lusso?
  - Che rapporto c'è tra i costi delle spedizioni di peso inferiore ai 100kg e quelle di peso superiore o pari a 100 kg?



# Sistema informazionale

- Base di dati
  - Modello intuitivo per l'utente
  - Modello efficiente per l'analisi
  - Fonti:
    - Db operazionale
    - Fonti esterne
  - fonti integrate, dati corretti e consistenti
  - aggiornamento tramite opportune procedure
- Algoritmi di analisi dei dati
  - Computazione on-line
    - Possibile pre-processing durante gli aggiornamenti del db
  - Sistemi di analisi interattiva guidata da ipotesi



# Sistema informazionale



# Terminologia

- Data Warehousing
  - Strumenti e tecniche di definizione, costruzione, mantenimento del data warehouse (“magazzino dei dati”)
- Decision Support System (DSS)
  - Sistemi informatici usati nel processo decisionale come supporto all'estrazione delle informazioni da basi di dati organizzate
- Data Mining
  - Strumenti e tecniche per l'estrazione dai dati di informazioni “nascoste”
- Business Intelligence
  - Attività di estrazione di informazione dai dati di business generati dai processi operativi aziendali
- Knowledge Management
  - Trattamento informatizzato della conoscenza aziendale
    - Dati strutturati e non strutturati



# Dati strutturati e non strutturati

- **Dato strutturato:**
  - Trattabile con un database (non necessariamente di tipo relazionale)
  - Localizzazione certa dell'informazione e del suo tipo
  - Possibili controlli ed ottimizzazioni
- **Dato non strutturato:**
  - Tipologie di dati molto diverse: testi, immagini, suoni, video
  - Localizzazione dell'informazione e del suo tipo difficile
  - Tecniche ad-hoc per ogni tipo di dato e contesto
- **Dato semi-strutturato:**
  - Parzialmente strutturato, parzialmente non strutturato

# Esempio dato semi strutturato

## Università degli Studi di Siena

Da Wikipedia, l'enciclopedia libera.

(Reindirizzamento da [Università di Siena](#))

L'**Università degli Studi di Siena** è una delle più antiche università [italiane](#).

Secondo una classifica stilata dal [Censis](#) nel 2009, è risultato il primo ateneo italiano per vivibilità e qualità dei servizi<sup>[2]</sup>. Nella graduatoria pubblicata annualmente da [Il Sole 24 Ore](#) l'Università occupa la 38ª posizione tra gli atenei statali<sup>[3]</sup>. Nella classifica [censis Repubblica 2010](#) si piazza al 2º posto.

### Indice [nascondi]

- 1 Storia
- 2 Laureati e docenti celebri
- 3 Facoltà (2010)
- 4 Radio
- 5 Note
- 6 Altri progetti
- 7 Collegamenti esterni

### Storia [modifica]

La prima notizia sull'istituzione risale al 1240: da un documento del tempo si apprende dell'esistenza di uno Studio "provinciale", finanziato dal Comune, costituito dalle scuole di diritto, di grammatica e di medicina. Ulteriori notizie certe dell'esistenza di una università senese finanziata dal Comune si hanno nel 1260. In quell'anno, un decreto del podestà Ildebrandino Cacciagontani imponeva a tutti coloro che affittavano alloggi agli studenti a pagare una tassa al Comune, con la quale si dovevano finanziare e stipendiare i professori dello "Studio Senese". Si sa che ne facessero parte una scuola medica, una di grammatica e una giuridica. Presso la scuola medica insegnò in quegli anni anche Pietro Ispano che nel 1276 sarebbe diventato [papa Giovanni XXI](#). Nel 1252 [papa Innocenzo IV](#) aveva concesso ai maestri e agli studenti l'immunità fiscale.

Nel 1321, a causa di una forte diatriba tra gli studenti dell'università [bolognese](#) "Alma Mater" e il podestà della città, Siena accolse numerosi studenti fuoriusciti dall'università felsinea, istigati anche da un locale lettore di legge Guglielmo Tolomei. Nel 1357 l'imperatore [Carlo IV](#) riconobbe lo studio senese come "Studium generale" e lo pose sotto la sua protezione. Nel 1392 si comincia a parlare della "Casa della Sapienza", una casa dello studente per accogliere i "fuori sede"; questa accoglierà i primi ospiti nel 1416, al prezzo di cinquanta fiorini.

In un periodo di decadenza della città, lo Studio fu uno dei pochi centri di sviluppo e di innovazione ancora attivi in città che riuscì a convogliare a Siena le innovazioni e le idee contemporanee. Anche dopo la caduta della città in mano [fiorentina](#) la città ottenne di conservare il proprio ateneo.

Nel 1808, con l'invasione francese, lo Studio senese venne chiuso, per riaprire i battenti solo con la [Restaurazione](#) Nel 1848 gli studenti senesi dimostrarono il loro patriottismo e parteciparono in grande numero alla [battaglia di Curtatone e Montanara](#) inquadrati nella "Compagnia della Guardia Universitaria", formata tra studenti e professori. Queste posizioni risorgimentali spinsero il Granduca di Toscana a chiudere la scuola medica, risparmiando

Università di Siena	
	
Nazione	<span><span></span></span> Italia
Città	Siena
Altre sedi	<a href="#">Arezzo</a> , <a href="#">Grosseto</a>
Nome latino	<i>Universitatis Senarum Studiorum</i>
Fondazione	1240
Tipo	statale
Facoltà	<a href="#">Economia</a> <span>↗</span> ; <a href="#">Farmacia</a> <span>↗</span> ; <a href="#">Giurisprudenza</a> <span>↗</span> ; <a href="#">Ingegneria</a> <span>↗</span> ; <a href="#">Lettere e Filosofia</a> <span>↗</span> ; <a href="#">Lettere e Filosofia, sede di Arezzo</a> <span>↗</span> ; <a href="#">Medicina e Chirurgia</a> <span>↗</span> ; <a href="#">Scienze matematiche,</a>



Non strutturato



Strutturato



# Data warehousing

- Data warehouse
  - Funzioni di popolamento, aggiornamento, pulizia, omologazione dei dati
  - Problema del “query transformation” per interrogare database con strutture diverse ed unire i risultati
- Strumenti
  - Analisi OLAP
  - Analisi interattiva guidata da ipotesi
  - Navigazione dei dati



# On Line Analytical Processing (OLAP)

- Obiettivo: fornire supporto efficiente per analisi multidimensionale delle informazioni
- Consente di scoprire
  - nuove relazioni tra le variabili,
  - identificare i parametri necessari alla gestione di grosse quantità di dati,
  - creare un numero illimitato di dimensioni
  - specificare condizioni ed espressioni che coinvolgono, contemporaneamente, più dimensioni.

# Schema Matching

- ***Query Transformation:***
  - trasformare una interrogazione su un db con una certa struttura in una interrogazione equivalente su un db con una struttura diversa
- ***Schema Mapping:***
  - definisce le regole per passare da una struttura all'altra
- ***Schema Matching:***
  - Il problema di “imparare” le regole di mapping

# Esempio

Docente	Facoltà	Insegnamento
Filippo Geraci	Ingegneria	SIA



Insegnante	Facoltà	Corso
Filippo Geraci	Ingegneria	SIA

- Mapping
  - Docente → Insegnante
  - Insegnamento → Corso





# Data Mining

- Estrarre informazione latente nei dati
- Analisi di grandi volumi
- Difficili da esaminare/analizzare interattivamente
- Tipi di dati diversi richiedono l'uso di tecniche diverse
  - Trattamento testi, immagini, video, audio
- Strumenti tipici
  - Classificazione, clustering



# Caratteristiche dei sistemi informativi

- Finalità
  - Fornire il substrato informativo per la conoscenza dell'azienda
    - Informazioni utili a livello operativo non sempre sono utili a livello informativo e vice versa
    - Esempio: numero di una fattura, partita iva di un cliente
  - Descrivere il passato, aiutare ad identificare i problemi e le loro cause
  - Suggestire i cambiamenti da apportare, fornire anticipazioni su scenari futuri



# Caratteristiche dei sistemi informativi

- **Struttura**
  - Dati articolati intorno a soggetti di cui si vuol conoscere l'apporto alla vita aziendale
- **Utenza**
  - Decisori, manager con visione e conoscenza ampia dell'azienda
- **richiedono strumenti di uso intuitivo che li seguano nei processi decisionali**
- **Storicità**
  - Arco temporale coperto molto più esteso di quella operativa
  - Mantenimento delle informazioni sull'evoluzione storica dei soggetti di interesse



# Caratteristiche dei dati ai fini informativi

- **Dettaglio**
  - Presentazione di dati in forma aggregata
  - Disponibilità di livelli diversi di aggregazione, dalla massima sintesi al dettaglio massimo
- **Accesso**
  - Accesso interattivo in sola lettura
  - Aggiornamenti applicati solo periodicamente ed in momenti in cui l'attività interattiva è nulla



# Modello multidimensionale

- Il processo di analisi
  - Descrizione e confronto di soggetti attraverso la misura degli eventi cui questi partecipano
    - un evento può essere in relazione con diversi soggetti
    - un soggetto può partecipare a più eventi
- Spazio delle informazioni come insieme di matrici multidimensionali
  - Ogni matrice rappresenta un tipo di evento
  - Ogni elemento della matrice rappresenta un singolo evento descritto da un insieme di coordinate
  - Ogni coordinata rappresenta un soggetto per le analisi da condurre su quella tipologia di evento

# Confronto caratteristiche tra operazionale ed informazionale

Caratteristica	Sistemi operazionali	Sistemi Informazionali
Tipo di dato	Corrente	Passato con possibilità di previsione futura
Utenza	Personale operativo	Manager
Processo di interazione	Guida operatore	Segue ragionamento
Tempificazione dati	Attuale	Storicizzata
Aggregazione dati	Puntuale	Aggregato
Tipo di accesso	Lettura/Scrittura	Lettura
Aggiornamento	Continuo	Periodico
Dimensionalità	Limitata	Elevata

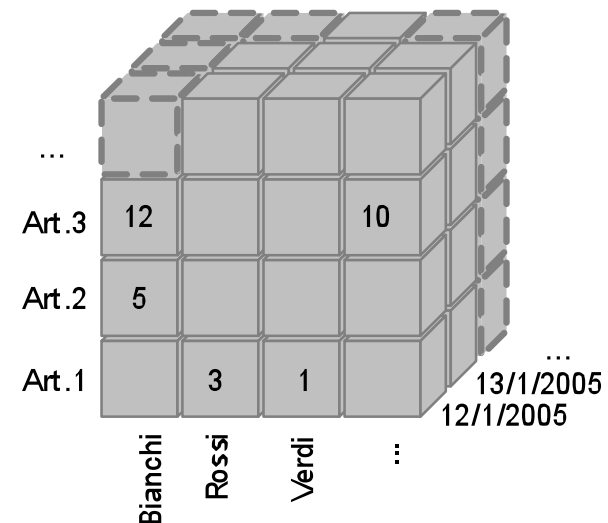
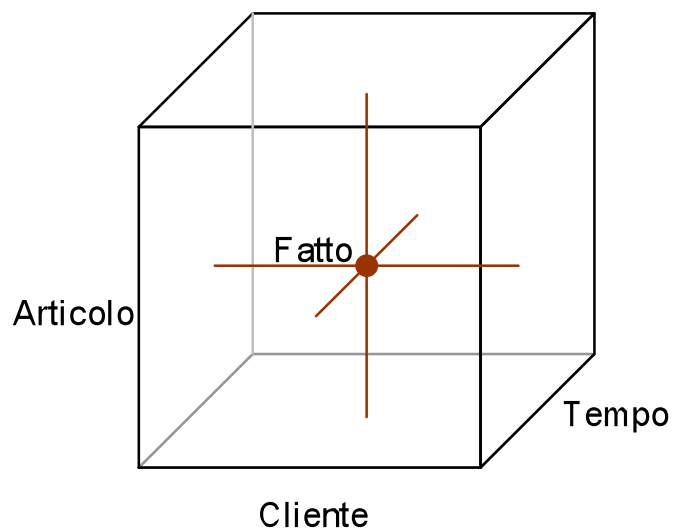


# Modello multidimensionale

- Intuitivo
- Facile de interpretare
- Efficiente per ricerca ed aggregazione
- Basato sul concetto di evento e soggetto
- Un evento può essere messo in relazione con diversi soggetti
- Esempi:
  - Abitudini di acquisto dei clienti
  - Redditività di un articolo

# Modello multidimensionale

- Ipercubo: matrice multidimensionale che rappresenta una tipologia di eventi
  - Fatto Elementare: elemento ottenuto come incrocio delle possibili coordinati
  - Misura: valore numerico che quantifica un fatto
  - Dimensione: coordinata del fatto







# Fatti

- Evento che accade nell'ambito dell'attività e che si ha interesse a misurare
  - **Esempio:** le vendite, i reclami, le spedizioni
- Caratteristiche del fatto
  - Dimensioni che lo collocano nel tempo e nello spazio aziendale (quando e dove e' avvenuto)
  - Misure che lo quantificano
  - Informazioni descrittive
- Identificazione univoca del fatto e delle misure tramite l'ennupla di coordinate
  - Fatto : (dimensione1, ... dimensioneN)
  - Misura: (dimensione1, ... dimensioneN).Misura



# Misure

- Caratteristica numerica del fatto che ne descrive aspetti quantitativi rilevanti per l'analisi
- Ogni fatto può avere più misure
- Le misure possono essere
  - Effettive: memorizzate sulla base di dati
  - Calcolate run-time: usando valori misure effettive
  - Implicite: indicano la presenza (o l'assenza) di un fatto
- Bisogna valutare il tradeoff tra l'efficienza computazionale delle misure effettive e lo spazio risparmiato da quelle calcolate

# Aggregabilità

- Dai fatti elementari si possono derivare fatti sintetici, quando non si considera una o più coordinate
- Le misure dei fatti sintetici sintesi sono ottenute aggregando le misure dei fatti elementari tramite opportuni operatori
  - Somma, media, max, min, ...
- Per ogni coppia (misura, dimensione) possono essere definiti operatori di aggregazione diversi
  - Una misura può essere non aggregabile lungo una dimensione
  - Un operatore può essere utilizzato per aggregare lungo alcune dimensioni e non lungo altre



# Caratteristiche delle misure

- **Aggregabilità**
  - Possibilità di usare un operatore di aggregazione su una misura o su una coppia (misura, dimensione)
- **Additività**
  - Possibilità di usare l'operatore di aggregazione "Somma" su una misura o su una coppia (misura, dimensione)

# Tipi di misure

- Misura di livello
  - Valore proprio del fatto, valido nel momento in cui viene registrato il fatto
  - Mai additiva rispetto alla dimensione temporale
  - **Esempio**: totale fattura
- Misura unitaria
  - Valore relativo ad uno dei soggetti, valido nel momento in cui viene registrato il fatto
  - Mai additiva
  - **Esempio**: costo unitario di un articolo
- Misura di flusso (tipo più comune)
  - Valore proprio del fatto rapportato ad un intervallo temporale di riferimento
  - Additiva lungo qualsiasi dimensione
  - **Esempio**: fatturato

# Aggregabilità: esempio

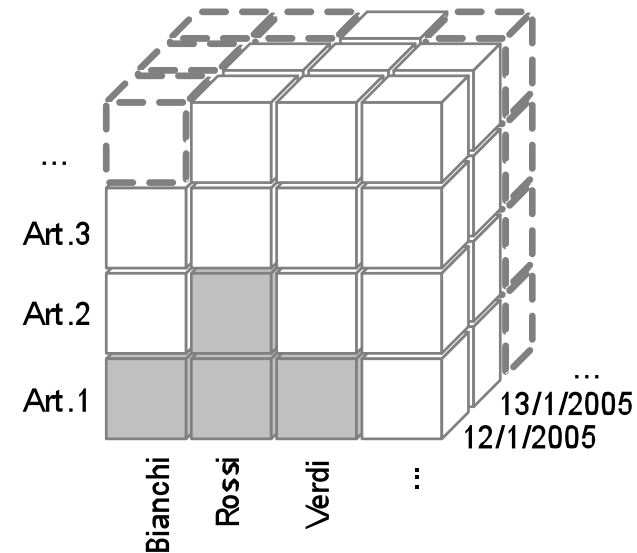
Articolo		Deposito	Data	Misura (Giacenza)
PP1007015	Pannello di Polistirolo 100x70x1,5	Materia Prima	13/02/05	100
PP1007015	Pannello di Polistirolo 100x70x1,5	Ricezione Merce	13/02/05	20
VA1010	Vite in acciaio 10mmx1	Materia Prima	13/02/05	24002
	...			
PP1007015	Pannello di Polistirolo 100x70x1,5	Materia Prima	14/02/05	110
PP1007015	Pannello di Polistirolo 100x70x1,5	Ricezione Merce	14/02/05	0
VA1010	Vite in acciaio 10mmx1	Materia Prima	14/02/05	23870
	...			

- La misura Giacenza è
  - Additiva rispetto alla dimensione deposito
  - Non additiva rispetto alla dimensione tempo
  - Non aggregabile rispetto alla dimensione Articolo

# Fatti senza misure proprie

- Di alcuni fatti l'unica cosa che interessa misurare è che si siano verificati
  - Valutazione della presenza
  - Valutazione della frequenza
- La misura (implicita) è la presenza o l'assenza del fatto
- Operatori di aggregazione
  - Conteggio
    - Presenza = 1
    - Assenza = 0
  - Operatori booleani
    - Presenza = True
    - Assenza = False

Articolo	Conteggio	Esistenza
Art.1	3	True
Art.2	2	True
Art. 3	0	False



# Nota: simulare gli operatori booleani

- Aggregazione avviene utilizzando formule nella logica booleana AND, OR, NOT
- Simulabili con operazioni algebriche
  - True  $\longrightarrow$  1
  - False  $\longrightarrow$  0

A and B	Min (A,B)
A or B	Abs (A-B)
Not A	1-A





# Dimensioni

- Soggetto a cui si vuole rapportare i fatti di analisi
- L'insieme delle dimensioni di un fatto
  - Determina la granularità con cui è possibile misurare il fatto
  - Influisce sulla complessità del modello informativo
  - Limita il dettaglio delle informazioni che si possono estrarre
  - Quando usate in maniera interattiva conviene ridurle per non generare confusione



# Dominio della dimensione

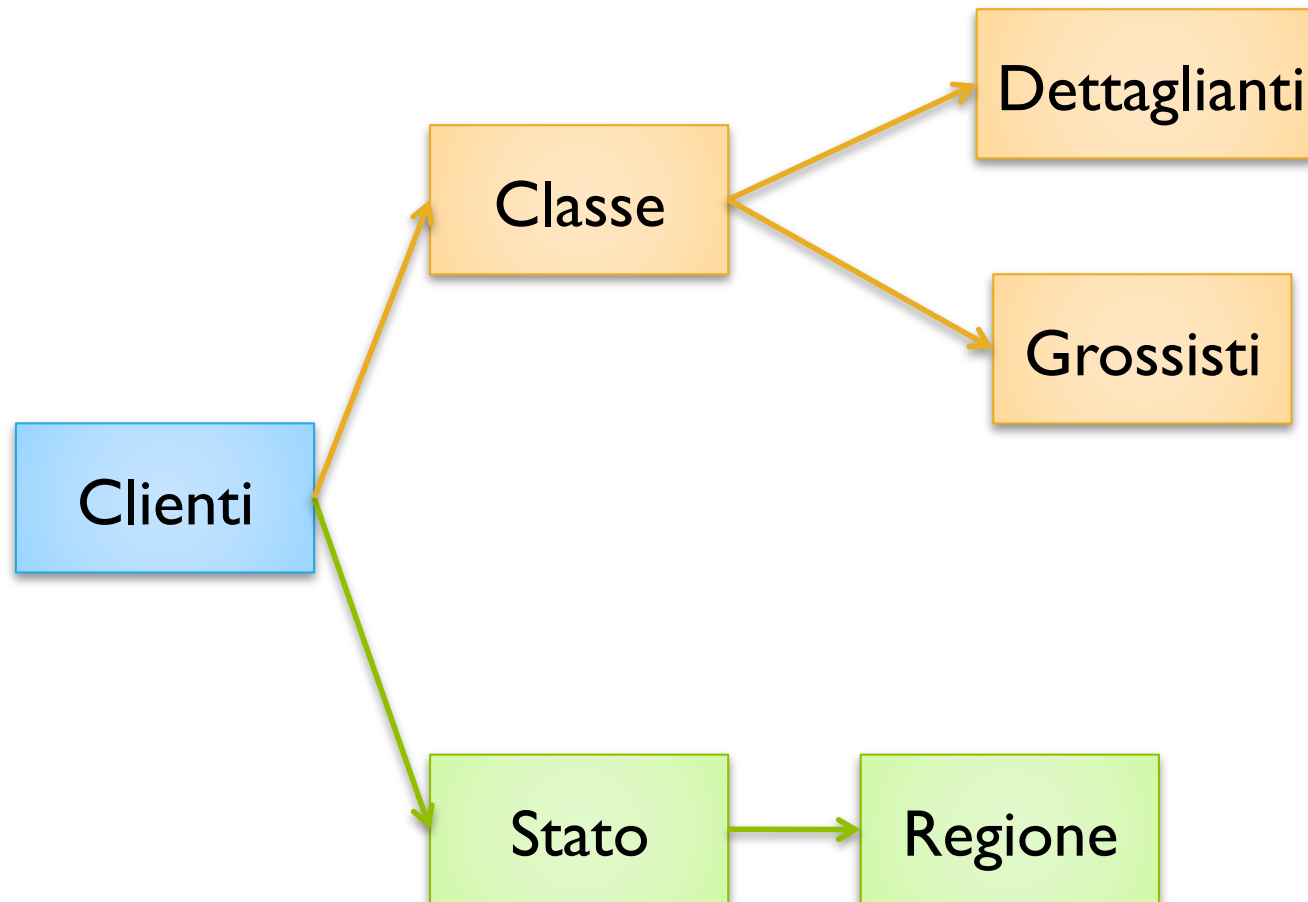
- Insieme dei diversi valori che una dimensione può assumere
- Il dominio deve essere:
  - Finito (limitato)
  - Discreto (non continuo)
- In caso contrario si devono usare strategie per rendere vere queste proprietà
  - Limitazione: limiti arbitrari al dominio
    - **Esempio:** intervallo sulla dimensione tempo
  - Quantizzazione: divisione dello spazio in “bucket” ed aggregazione di tutti i valori del bucket in uno unico
    - **Esempio:** quantizzazione del tempo in giorni



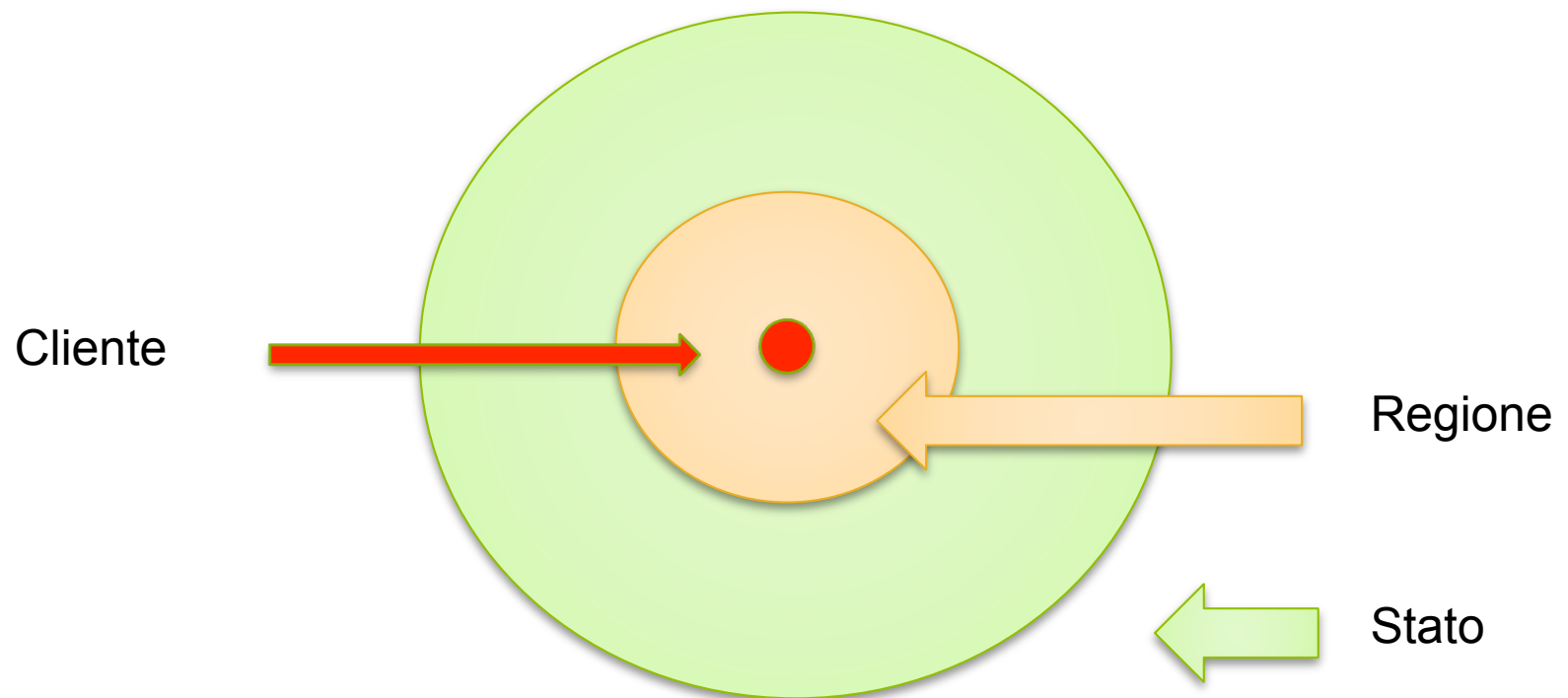
# Gerarchie

- Una dimensione può essere la radice di una o più gerarchie
  - Ogni gerarchia è un albero
  - Figli di uno stesso nodo non hanno elementi in comune
  - Livelli inferiori rappresentano specializzazioni (restrizioni) dei livelli superiori
  - I nodi si chiamano attributi dimensionali
- Gli attributi dimensionali
  - Sono usati per aggregare i fatti elementari
  - Sono determinati univocamente dal valore della dimensione
  - Rappresentano classificazioni della dimensione

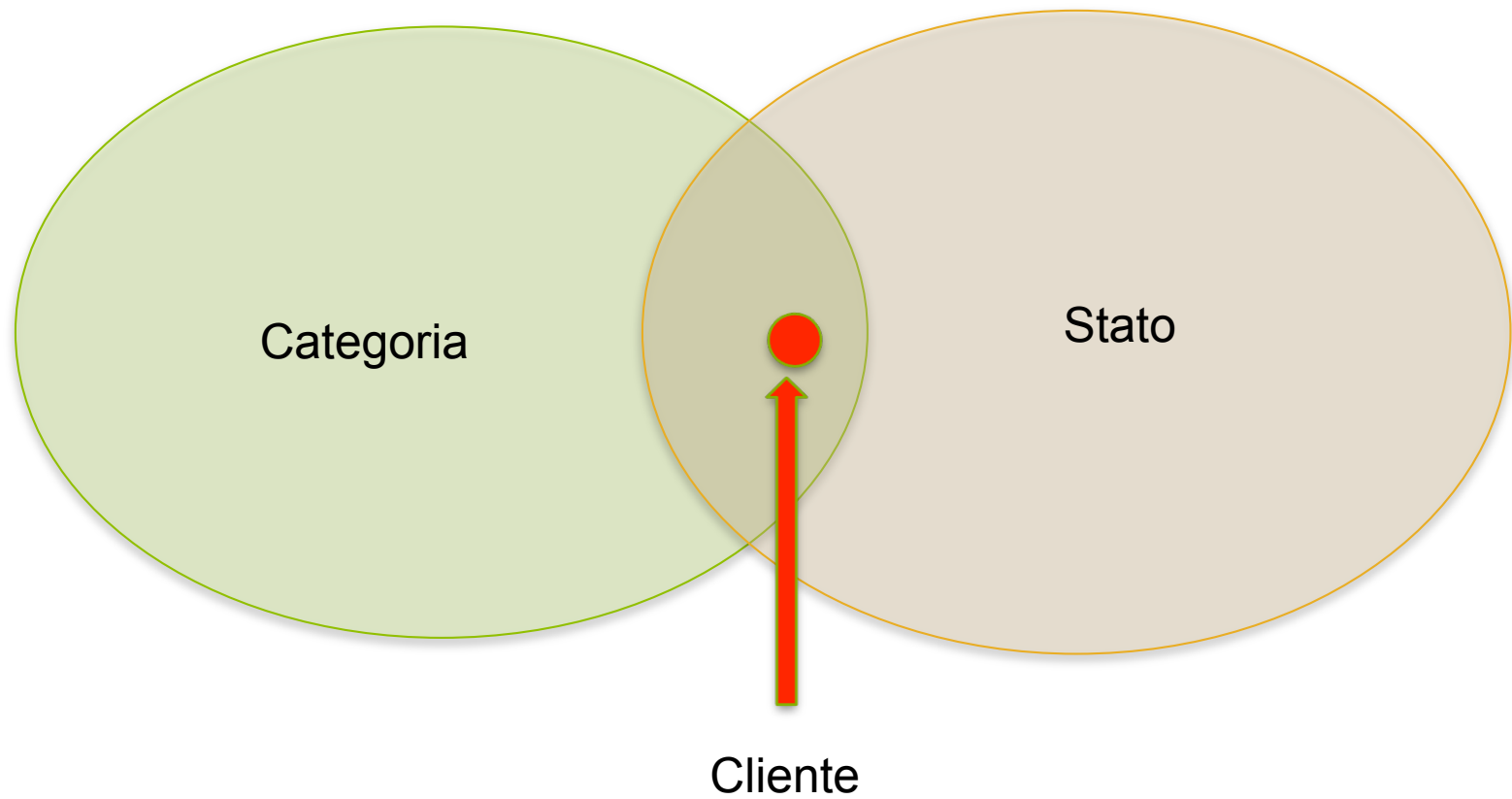
# Esempio



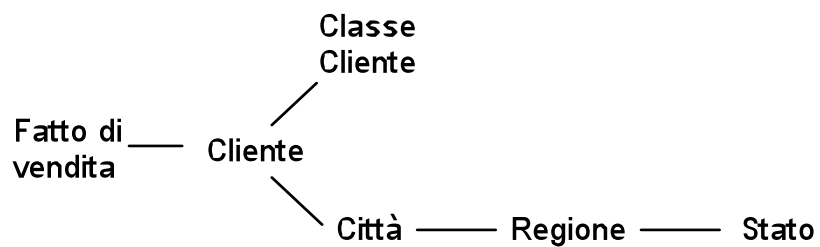
# Sottoinsiemi su linea gerarchica



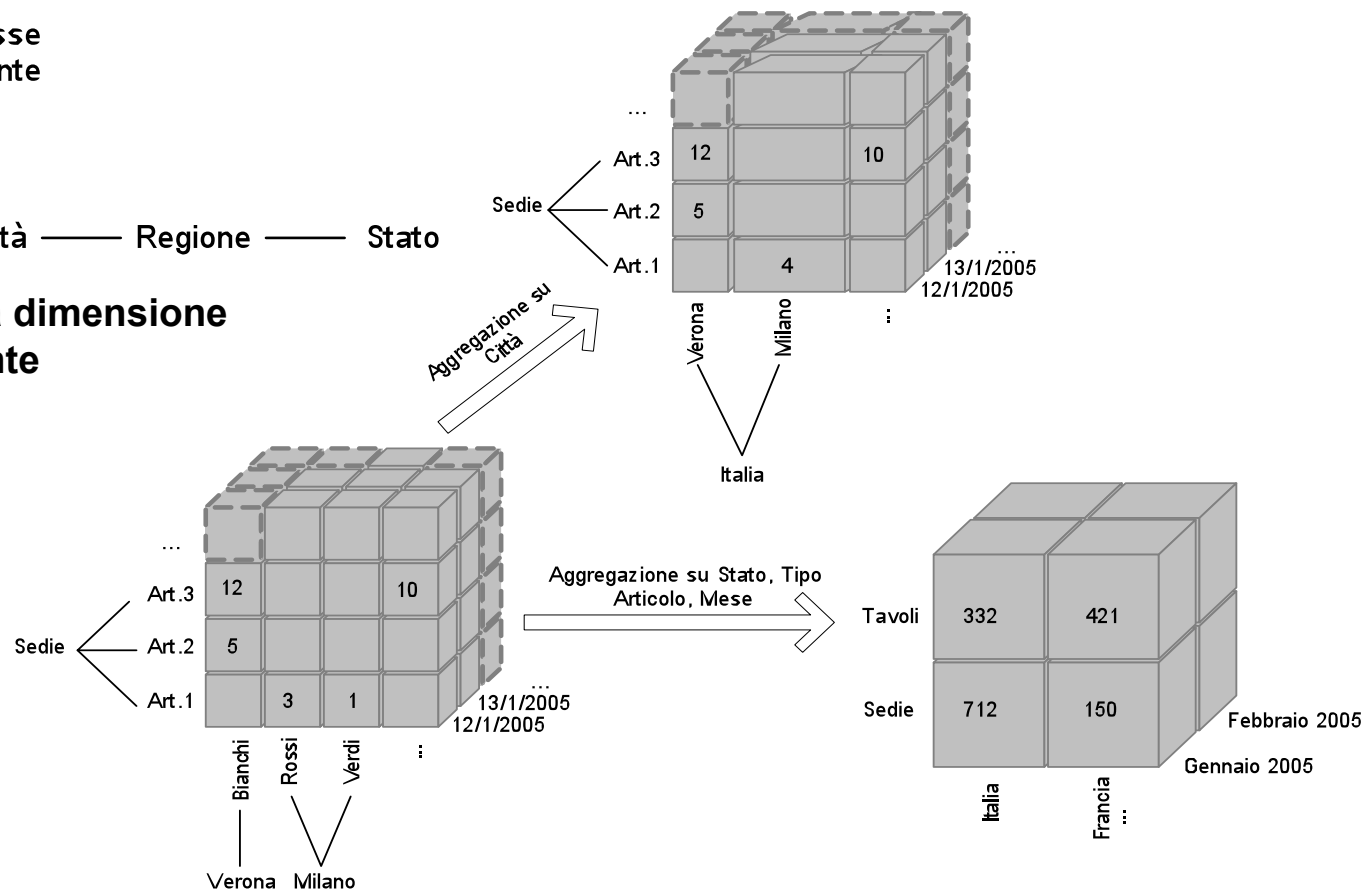
# Sottoinsiemi su gerarchie diverse



# Gerarchie: esempio



**Gerarchia sulla dimensione Cliente**



**Esempi di aggregazione lungo attributi dimensionali**



# Dimensioni

- Dimensione opzionale
  - Attributo dimensionale che può essere non definito per alcuni valori della dimensione
  - L'opzionalità deve essere nota per poter mantenere la consistenza della base di dati
- Gerarchia condivisa
  - Gerarchia utilizzata in fatti diversi per descrivere il ruolo di una certa classe di soggetti
  - Gerarchia utilizzata per descrivere nello stesso fatto ruoli diversi di una stessa classe di soggetti
  - Ottimizzabili a livello di data warehouse





# Caratteristiche strutturali

- **Multidimensionalità**
  - Accesso alle informazioni tramite ennuple di coordinate
- **Granularità**
  - Grado minimo di aggregazione dei dati nel data warehouse; corrisponde ad un fatto elementare
    - La granularità del data warehouse è minima quando ogni fatto corrisponde ad un solo evento origine
    - Maggiore granularità implica maggiore spazio e costo computazionale
    - Granularità troppo grossa implica possibile perdita di informazione interessante
- **Arco Temporale**
  - Intervallo temporale coperto dai dati nel data warehouse
- **Profondità storica**
  - Capacità di mantenere anche l'informazione storica sui soggetti, oltre al valore attuale



# Caratteristiche funzionali

- **Integrazione dei dati**
  - Livello di integrazione dei dati provenienti da fonti diverse
  - Schema mapping e schema matching
  - Controlli sui formati dei dati per mantenere omogeneità
- **Accessibilità**
  - Grado di facilità d'uso e velocità di risposta
  - Richiede algoritmi di calcolo on-line
  - Richiede interfacce intuitive
- **Flessibilità e sintesi**
  - Capacità di assecondare l'utente nell'articolare richieste, aggregare dati e mettere in relazione misure secondo criteri di selezione non prefissati
  - Fornire vari livello di aggregazione/dettaglio



# Caratteristiche funzionali

- **Correttezza**
  - Grado di aderenza dei dati alla realtà rappresentata
  - Fonti diverse possono rappresentare lo stesso dato in modo diverso
  - Controlli su dati non rilevanti estratti dal sistema operativo spesso poco controllati
    - Esempio: caratteristiche descrittive, classificazioni
- **Completezza**
  - Capacità del sistema informazionale di rappresentare tutti gli eventi importanti per le analisi
    - Si riferisce sia al modello sia alle istanze dei dati memorizzati
- **Decisioni prese in base a dati non completi o non corretti possono portare a scelte errate**



# Data warehouse e Data mart

- Data warehouse
  - Magazzino unico, completo e consistente dell'informazione aziendale
  - Contiene dati articolati attorno a tutti i fatti di interesse aziendale (tutti i possibili ipercubi)
  - Può raggiungere dimensioni estremamente elevate
- Troppi dati nell'analisi rischiano di confondere e rallentano la computazione
- Troppo pochi dati possono fare perdere informazioni
- Data mart: porzione del data warehouse che contiene
  - Tutti i dati di suo interesse
  - Solo i dati di suo interesse

# Data warehouse e Data mart

- Data warehouse tematico, derivato dal data warehouse aziendale
  - Comprende i soli fatti che riguardano una certa area d'indagine
  - Estensione temporale ridotta
  - Granularità dei fatti minore

