

# **Sistemi informativi aziendali ERP e sistemi di data analysis**

## **I sistemi informazionali**

Maurizio Pighin, Anna Marzona



- Sfruttare il patrimonio dei dati operativi per l'estrazione di informazioni utili nel processo decisionale
- Superare i limiti dei primi strumenti di supporto alle decisioni
  - *Reporting*
    - staticità
    - difficoltà e pesantezza dell'iter di realizzazione
    - parzialità
  - *Fogli di calcolo*
    - macchinosità (complessità delle basi di dati; sequenza delle istruzioni da compiere per portare i dati sul foglio)
    - scarsa controllabilità dei dati
    - proliferazione di fogli di calcolo 'personali'
    - limitatezza dello spazio dei dati



# Interrogazioni sui dati

- Interrogazioni sui sistemi operazionali (puntuali)
  - *Che tipo di modalità di pagamento è stata concordata per questo cliente?*
  - *Cosa deve produrre oggi il centro di lavoro 10?*
  - *Di cosa ho bisogno per produrre 12 sedie modello Tamigi blu?*
  - *Quali fatture fornitore devo ancora pagare?*
- Interrogazioni sui sistemi informazionali (complesse)
  - *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?*



- Ambiente che facilita l'analisi dei dati
  - *Base di dati*
    - modello intuitivo per l'utente, efficiente per l'analisi
    - dati di interesse per l'analisi provenienti da BD operativa, BD specialistiche, fonti esterne
    - fonti integrate, dati corretti e consistenti
    - aggiornamento tramite opportune procedure
  - *Strumenti di analisi dei dati*
    - reporting
    - sistemi di analisi interattiva guidata da ipotesi
    - sistemi di data mining



- Data Warehouse e Data Warehousing
  - *Strumenti e tecniche di definizione, costruzione, mantenimento del data warehouse*
- 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 in senso ampio (anche da basi di dati non strutturate)*



- Big data
  - *Fonti di dati aggiuntive rispetto a quelle dei DW*
    - tutti gli ambiti in cui, per costruzione o per processo di interazione, la produzione del dato è automatica e non ha costi aggiuntivi
      - *ad esempio le informazioni di navigazione su Web*
    - flussi continui di informazione
  - *Il trattamento di questi dati necessita di soluzioni diverse, basate su tecniche di elaborazione distribuita e algoritmi complessi.*
  - *Passaggio dato che permette di “osservare” andamenti → dato che può essere usato per produrre nuovi servizi*



# Caratteristiche dei sistemi informazionali

- Finalità
  - *Fornire il substrato informativo per la conoscenza dell'azienda*
  - *Descrivere il passato, aiutare ad identificare i problemi e le loro cause*
  - *Suggerire i cambiamenti da apportare, fornire anticipazioni su scenari futuri*
- 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



# Caratteristiche dei dati ai fini informationali

- **Storicità**
  - *Arco temporale coperto molto più esteso di quella operativa*
  - *Mantenimento delle informazioni sull'evoluzione storica dei soggetti di interesse*
- **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*



# Esempio di analisi multi-dimensionale

## Soda Example

Time	\$ Sales
Q3	\$16,000
Q4	\$16,000
<b>Total</b>	<b>\$32,000</b>

Product	\$ Sales
Cola	\$8,000
Cherry	\$8,000
Grape	\$8,000
Lemon-Lime	\$8,000
<b>Total</b>	<b>\$32,000</b>

Geography	\$ Sales
Munich	\$8,000
Frankfurt	\$8,000
Cologne	\$8,000
Berlin	\$8,000
<b>Total</b>	<b>\$32,000</b>



# Esempio di analisi multi-dimensionale

Sistemi informativi aziendali  
ERP e sistemi di data analysis  
Cap. 11 - I sistemi informazionali  
Maurizio Pighin, Anna Marzona

## Soda Example

		Munich	Frankfurt	Cologne	Berlin	Total
Q3	Cola	\$ -	\$ -	\$2,500	\$1,500	\$4,000
	Cherry	\$ -	\$ -	\$2,000	\$2,000	\$4,000
	Grape	\$1,000	\$3,000	\$ -	\$ -	\$4,000
	Lem-Lime	\$2,000	\$2,000	\$ -	\$ -	\$4,000
	<b>Total Q3</b>	<b>\$3,000</b>	<b>\$5,000</b>	<b>\$4,500</b>	<b>\$3,500</b>	<b>\$16,000</b>
Q4	Cola	\$4,000	\$ -	\$ -	\$ -	\$4,000
	Cherry	\$1,000	\$3,000	\$ -	\$ -	\$4,000
	Grape	\$ -	\$ -	\$1,500	\$2,500	\$4,000
	Lem-Line	\$ -	\$ -	\$2,000	\$2,000	\$4,000
	<b>Total Q4</b>	<b>\$5,000</b>	<b>\$3,000</b>	<b>\$3,500</b>	<b>\$4,500</b>	<b>\$16,000</b>
<b>Total</b>		<b>\$8,000</b>	<b>\$8,000</b>	<b>\$8,000</b>	<b>\$8,000</b>	<b>\$32,000</b>



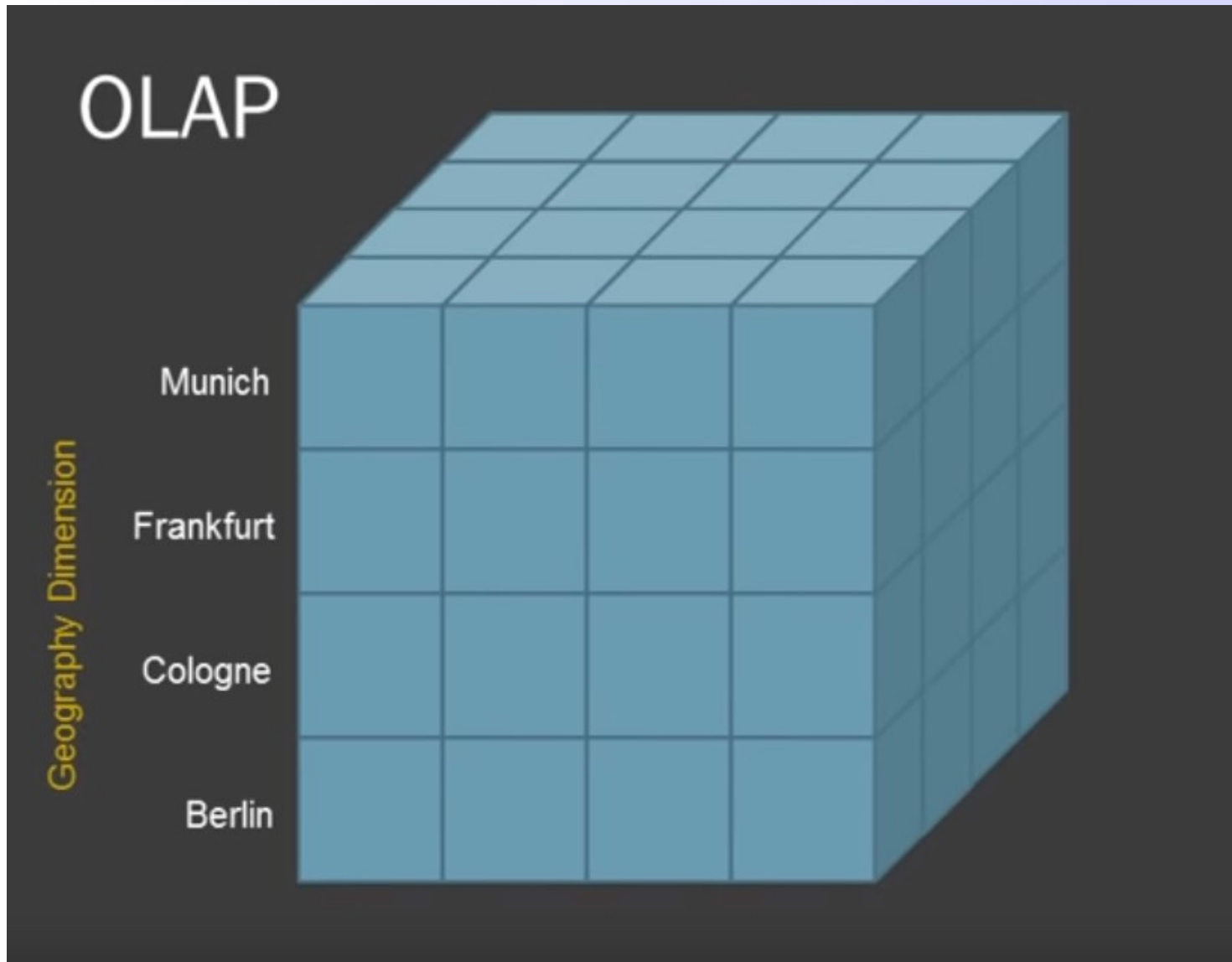
# II «pivoting»

## Pivoted Soda Data

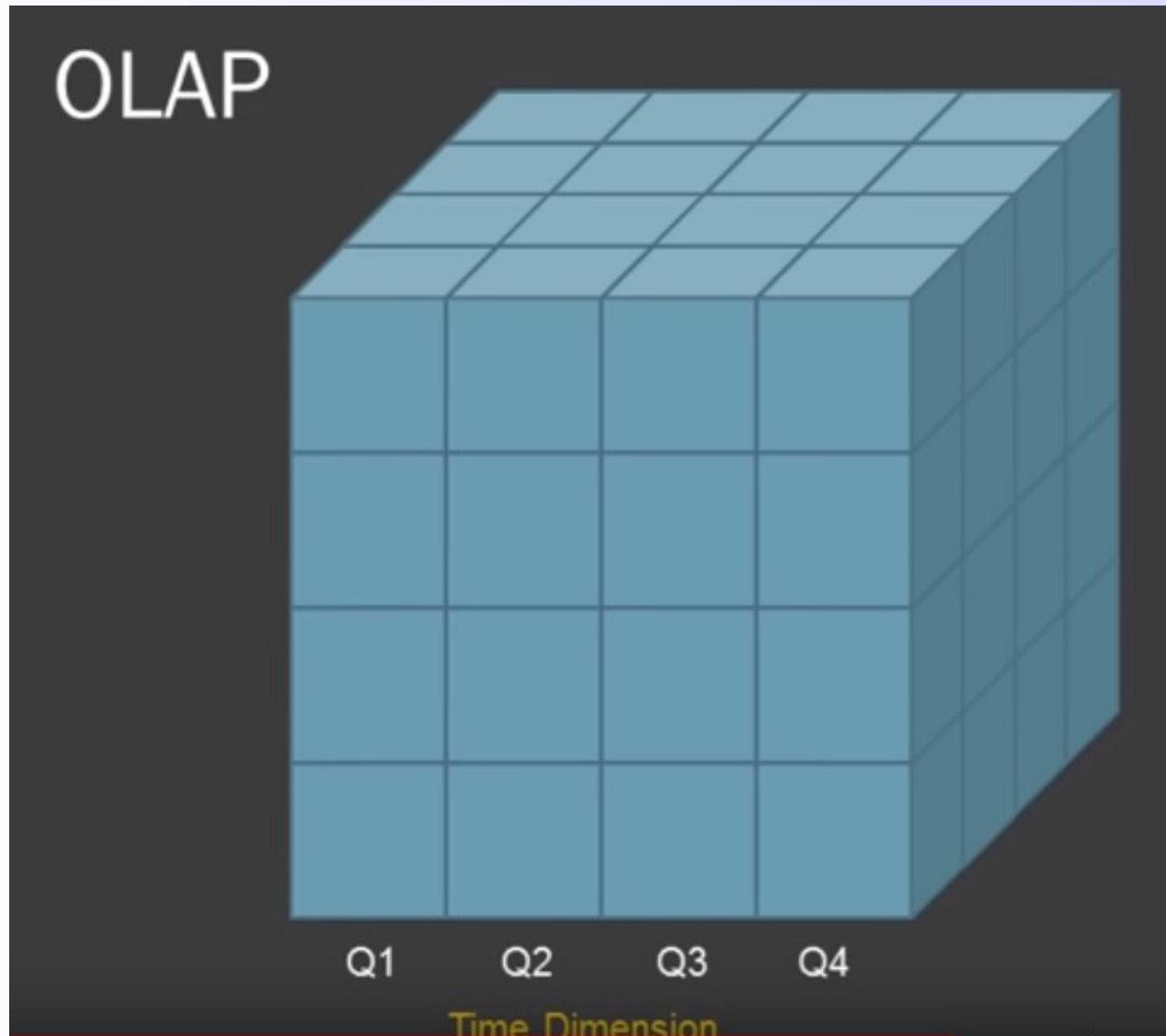
		Cola	Cherry	Grape	Lem-Lime	Total
Munich	Qtr 3	\$ -	\$ -	\$1,000	\$2,000	\$3,000
	Qtr 4	\$4,000	\$1,000	\$ -	\$ -	\$5,000
	<b>Total</b>	<b>\$4,000</b>	<b>\$1,000</b>	<b>\$1,000</b>	<b>\$2,000</b>	<b>\$8,000</b>
Frankfurt	Qtr 3	\$ -	\$ -	\$3,000	\$2,000	\$5,000
	Qtr 4	\$ -	\$3,000	\$ -	\$ -	\$3,000
	<b>Total</b>	<b>\$ -</b>	<b>\$3,000</b>	<b>\$3,000</b>	<b>\$2,000</b>	<b>\$8,000</b>
Cologne	Qtr 3	\$2,500	\$2,000	\$ -	\$ -	\$4,500
	Qtr 4	\$ -	\$ -	\$1,500	\$2,000	\$3,500
	<b>Total</b>	<b>\$2,500</b>	<b>\$2,000</b>	<b>\$1,500</b>	<b>\$2,000</b>	<b>\$8,000</b>
Berlin	Qtr 3	\$1,500	\$2,000	\$ -	\$ -	\$3,500
	Qtr 4	\$ -	\$ -	\$2,500	\$2,000	\$4,500
	<b>Total</b>	<b>\$1,500</b>	<b>\$2,000</b>	<b>\$2,500</b>	<b>\$2,000</b>	<b>\$8,000</b>
<b>Grand</b>	<b>Total</b>	<b>\$8,000</b>	<b>\$8,000</b>	<b>\$8,000</b>	<b>\$8,000</b>	<b>\$32,000</b>



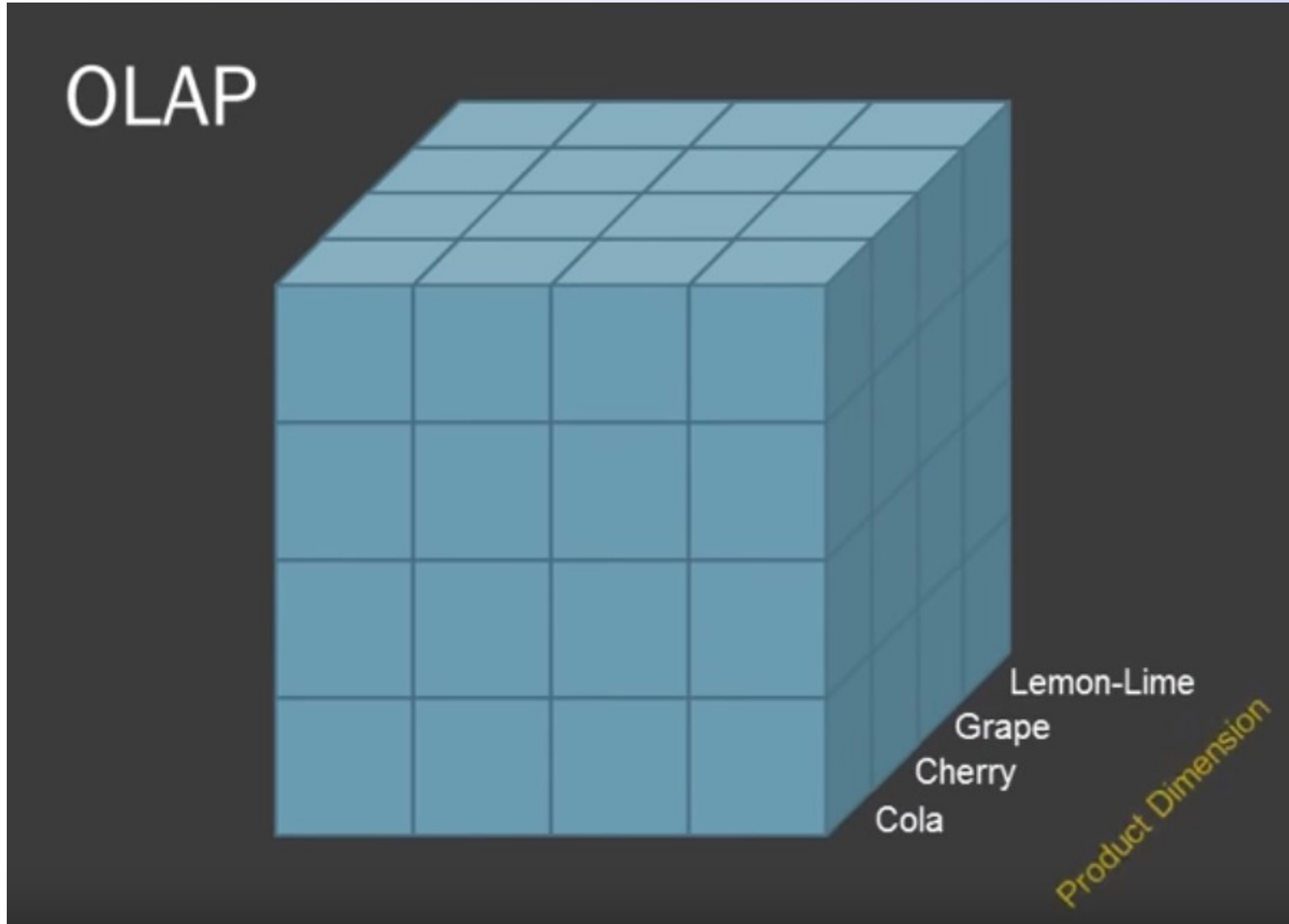
# La visione a cubo



# La visione a cubo

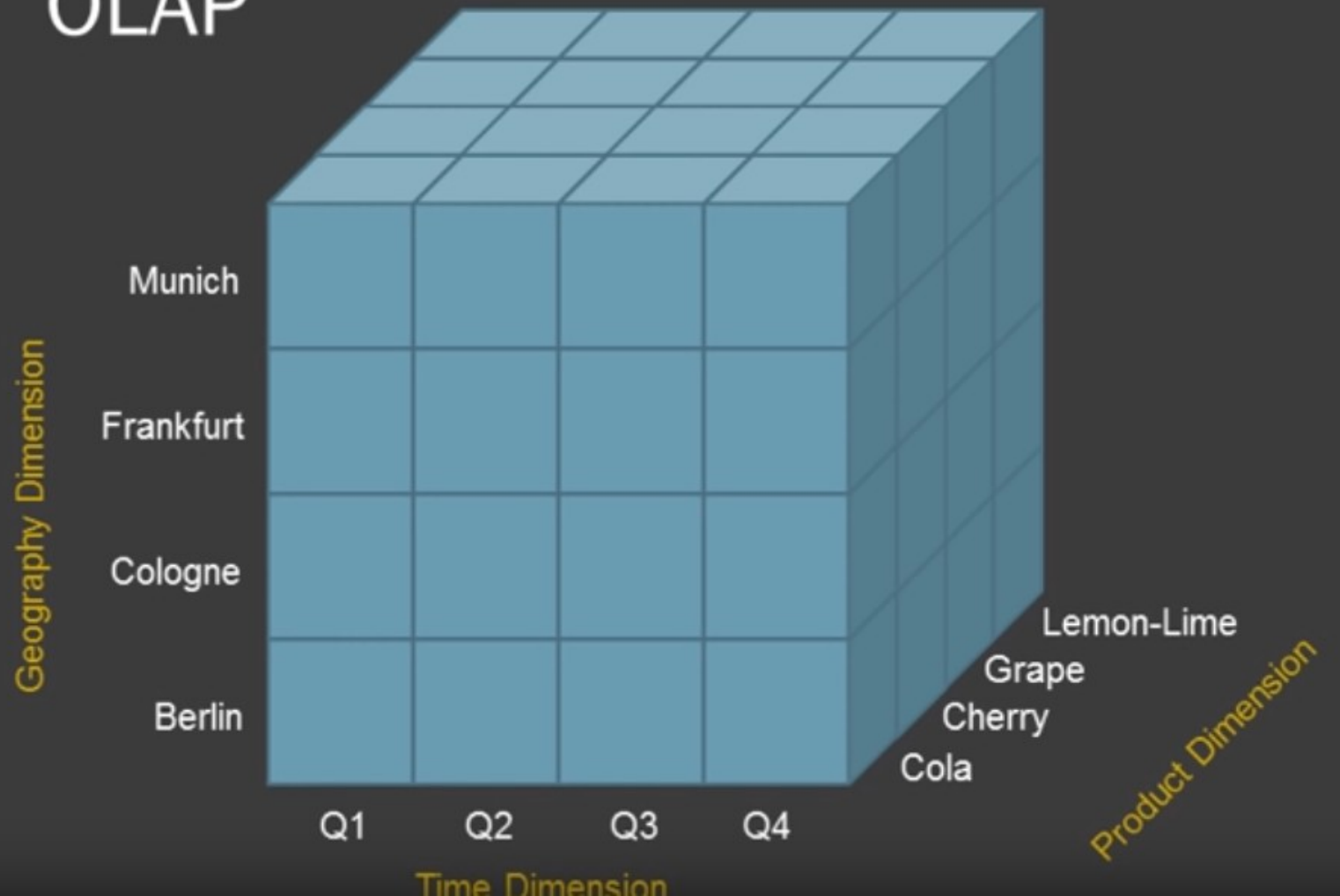


# La visione a cubo



# La visione a cubo

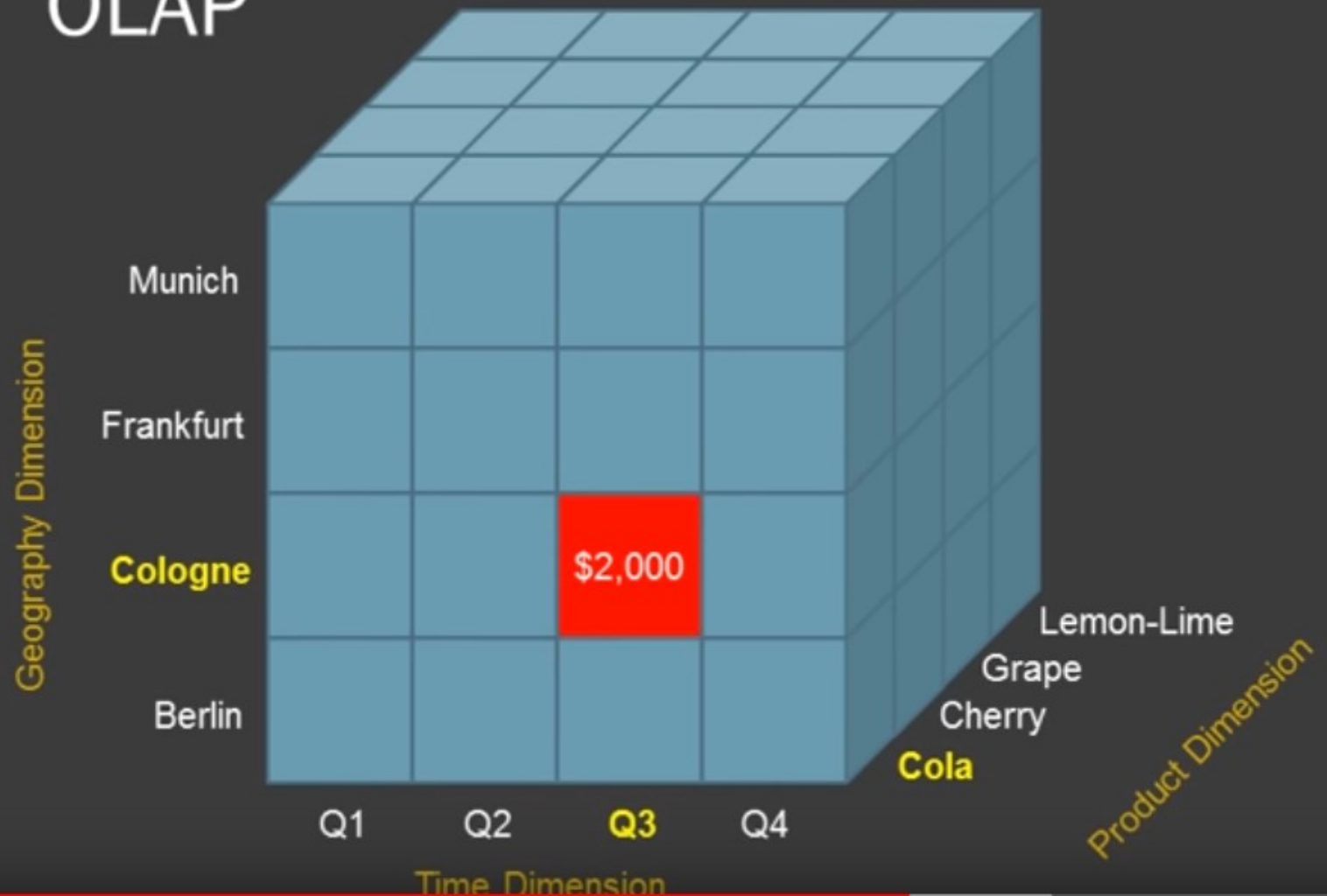
OLAP



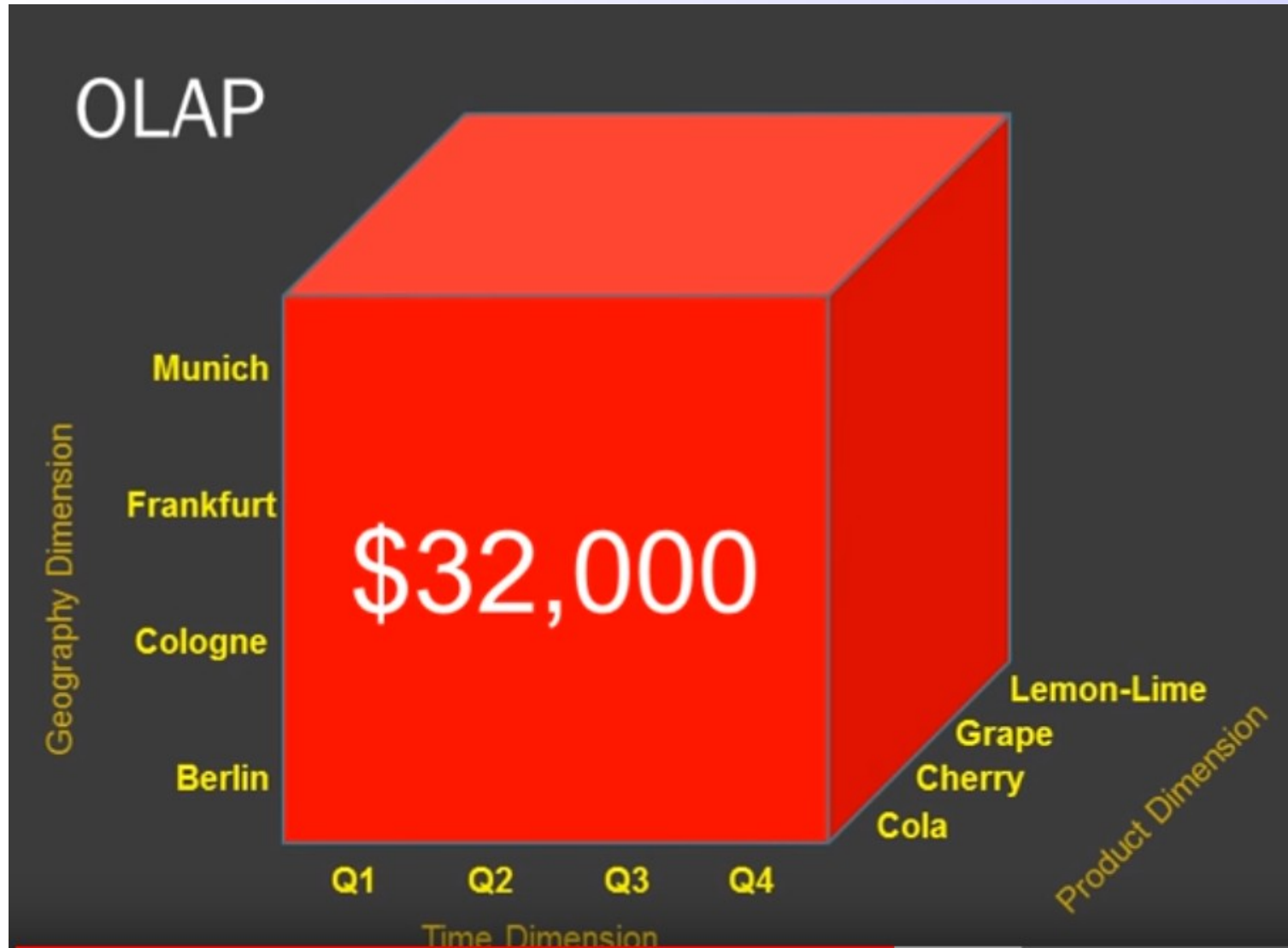


# La visione a cubo

## OLAP

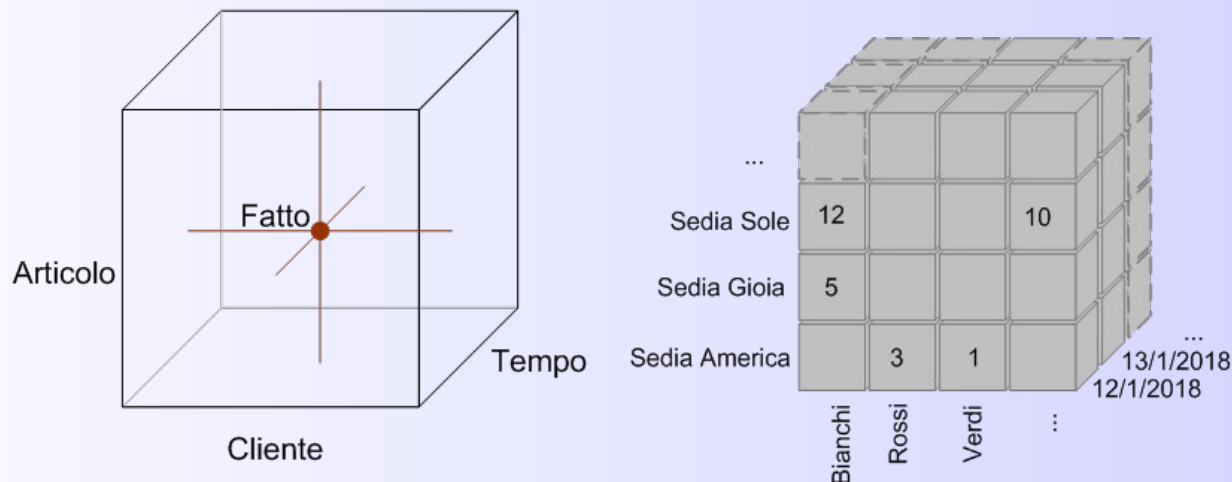


# La visione a cubo



# Modello multidimensionale

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



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



- Misura
  - *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 utilizzando i valori delle misure effettive*
  - *Implicite, indicano la presenza (o l'assenza) di un fatto*



- Dai fatti elementari si possono derivare fatti sintetici, quando non si considera una o più coordinate
- Le misure dei fatti sintetici 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*



- 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)*



- Misura di livello
  - *Valore proprio del fatto, valido nel momento in cui viene registrato il fatto*
  - *Mai additiva rispetto alla dimensione temporale*
- Misura unitaria
  - *Valore relativo ad uno dei soggetti, valido nel momento in cui viene registrato il fatto*
  - *Mai additiva*
- Misura di flusso
  - *Valore proprio del fatto rapportato ad un intervallo temporale di riferimento*
  - *Additiva lungo qualsiasi dimensione*





# Aggregabilità: esempio

Articolo		Deposito	Data	Misura (Giacenza)
PP1007015	Pannello di polistirolo 100x70x1,5	Materia prima	13/02/2018	100
PP1007015	Pannello di polistirolo 100x70x1,5	Ricezione merce	13/02/2018	20
VA1010	Vite in acciaio 10mmx1	Materia prima	13/02/2018	24.002
	...			
PP1007015	Pannello di polistirolo 100x70x1,5	Materia prima	14/02/2018	110
PP1007015	Pannello di polistirolo 100x70x1,5	Ricezione merce	14/02/2018	0
VA1010	Vite in acciaio 10mmx1	Materia prima	14/02/2018	23.870
	...			

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



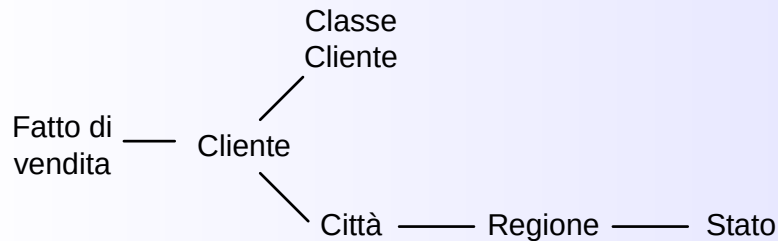
- Dimensione
  - *Soggetto a cui si vuole rapportare i fatti di analisi*
- L'insieme delle dimensioni di un fatto
  - *Determina la granularità con cui è possibile misurare il fatto*
  - *Influisce sulla complessità del modello informativo*
  - *Limita il dettaglio delle informazioni che si possono estrarre*
- Dominio della dimensione
  - *Insieme dei diversi valori che una dimensione può assumere*
  - *Il dominio deve essere finito (se il dominio è continuo o infinito deve essere reso discreto e finito)*



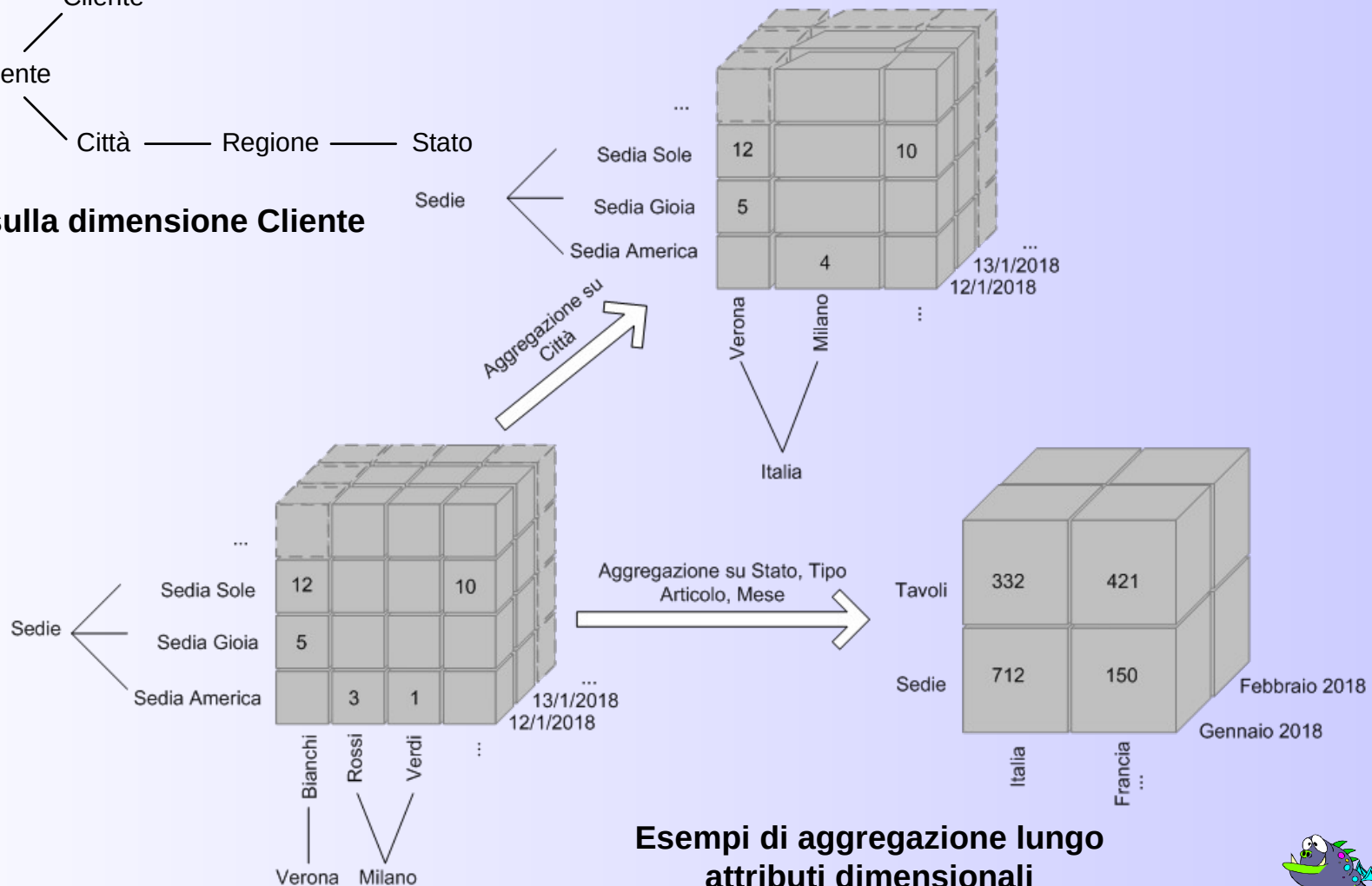
- Gerarchia
  - *Insieme di attributi (attributi dimensionali) collegati gerarchicamente ad una dimensione*
- Gli attributi dimensionali
  - *Sono usati per aggregare i fatti elementari*
  - *Sono determinati univocamente dal valore della dimensione*
  - *Rappresentano classificazioni della dimensione*
- L'analisi dei fatti di un cubo può essere condotta attraverso viste che utilizzano attributi dimensionali



# Gerarchie: esempio



## Gerarchia sulla dimensione Cliente



## Esempi di aggregazione lungo attributi dimensionali



- 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*



# 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
- 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*
- Accessibilità
  - *Grado di facilità d'uso e velocità di risposta*
- Flessibilità e sintesi
  - *Capacità del modello di assecondare l'utente nell'articolare richieste, aggregare dati e mettere in relazione misure secondo criteri di selezione non prefissati*
- Correttezza
  - *Grado di aderenza dei dati alla realtà rappresentata*
- 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



# 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*
- Nelle analisi l'utente opera su una porzione del data warehouse (data mart) 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*

