

Database and data mining group, Politecnico di Torino
DBG

Data warehouse Analisi dei dati

Elena Baralis
Politecnico di Torino

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 1 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DBG

Ambiente controllato di query

- Sono definite
 - ricerche complesse con struttura prefissata (normalmente parametrica)
 - procedure specifiche di analisi
 - rapporti con struttura prefissata
- È possibile introdurre elementi specifici del settore economico considerato
- È necessario lo sviluppo di codice ad hoc
 - si utilizzano stored procedures, applicazioni contenute in packages, join e aggregazioni predefinite
 - sono disponibili strumenti flessibili per la gestione della reportistica, che permettono di definire layout, periodicità di pubblicazione, liste di distribuzione

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 4 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DBG

Operazioni di analisi dei dati

- Calcolo di funzioni aggregate lungo una o più dimensioni
 - necessità di fornire supporto a diversi tipi di funzione aggregata (esempi: media mobile, top ten)
- Operazioni di confronto, essenziali per confrontare l'andamento degli affari (esempio: confronto dei dati delle vendite in mesi diversi)
 - è difficile eseguire confronti utilizzando solo il linguaggio SQL
- Analisi dei dati mediante tecniche di data mining

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 2 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DBG

Ambiente di query ad hoc

- È possibile definire interrogazioni OLAP di tipo arbitrario, progettate al momento dall'utente
 - formulazione delle interrogazioni mediante tecniche point and click, che generano automaticamente istruzioni SQL
 - si possono definire interrogazioni (tipicamente) complesse
 - interfaccia basata sul paradigma dello spreadsheet
- Una sessione di lavoro OLAP permette raffinamenti successivi della stessa interrogazione
- Utile quando i rapporti predefiniti non sono adeguati

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 5 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DBG

Strumenti di interfaccia

L'utente può interrogare il data warehouse mediante strumenti di vario tipo:

- ambiente controllato di query
- strumenti specifici di query e generazione rapporti
- strumenti di data mining

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 3 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DBG

OLAP

Elena Baralis
Politecnico di Torino

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 6 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Analisi OLAP

- Operazioni di ricerca disponibili
 - roll up, drill down
 - slice and dice
 - pivot di tabelle
 - ordinamento
- Le operazioni possono essere
 - combinate tra loro nella stessa query
 - eseguite in una sequenza di raffinamenti successivi della stessa query che forma la sessione di lavoro OLAP

Elena Baralis
Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Roll up

Matrix	Customer	Region	Year	Quarter	Month	Product	Category
North-east	North-east	North-east	2000	Q1	Jan	Food	Food
North-east	North-east	North-east	2000	Q1	Jan	Food	Food
North-east	North-east	North-east	2000	Q1	Jan	Food	Food

↓

Matrix	Customer	Region	Year	Quarter	Month	Product	Category
North-east	North-east	North-east	2000	Q1	Jan	Food	Food
North-east	North-east	North-east	2000	Q1	Jan	Food	Food
North-east	North-east	North-east	2000	Q1	Jan	Food	Food

Elena Baralis
Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Roll up

- Riduzione di dettaglio dei dati mediante
 - la riduzione del livello di dettaglio di una delle dimensioni presenti, con l'aumento di livello in una gerarchia
 - esempio
group by negozio, mese → group by città, mese
 - l'eliminazione di una delle dimensioni presenti
 - esempio
group by prodotto, città → group by prodotto

Elena Baralis
Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Roll up

Matrix	Customer	Region	Year	Quarter	Month	Product	Category
North-east	North-east	North-east	2000	Q1	Jan	Food	Food
North-east	North-east	North-east	2000	Q1	Jan	Food	Food
North-east	North-east	North-east	2000	Q1	Jan	Food	Food

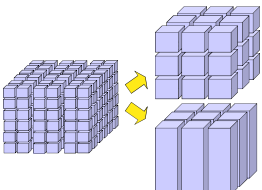
↓

Matrix	Customer	Region	Year	Quarter	Month	Product	Category
North-east	North-east	North-east	2000	Q1	Jan	Food	Food
North-east	North-east	North-east	2000	Q1	Jan	Food	Food
North-east	North-east	North-east	2000	Q1	Jan	Food	Food

Elena Baralis
Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Roll up



Elena Baralis
Politecnico di Torino

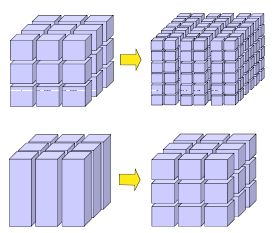
Database and data mining group, Politecnico di Torino
DMG

Drill down

- Aumento di dettaglio dei dati mediante
 - l'aumento del livello di dettaglio di una delle dimensioni presenti, con la riduzione di livello in una gerarchia
 - esempio: da raggruppamento per città e mese a raggruppamento per negozio e mese
 - l'aggiunta di una nuova dimensione
 - esempio: da raggruppamento per città, raggruppamento per città e prodotto
- Spesso il drill down opera su un sottoinsieme dei dati di partenza

Elena Baralis
Politecnico di Torino

Drill down



Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 13 Elena Baralis Politecnico di Torino

Slice and dice

- Riduzione del volume dei dati da analizzare
 - selezione di un sottoinsieme mediante predicati
 - slice: predicato di uguaglianza che seleziona una "fetta"
 - esempio: Anno=2005
 - dice: combinazione di predicati che seleziona un "cubetto"
 - esempio: Categoria='Alimentari' and Città='Torino'

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 16 Elena Baralis Politecnico di Torino

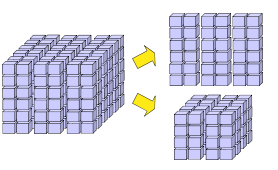
Drill down

Customer	Region	Year	Quarter	Product	Category	Sub-Category	Item	Quantity	Price	Discount	Amount
Customer1	Region1	2005	Q1	Product1	Category1	Sub-Category1	Item1	10	100	0	1000

Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 14 Elena Baralis Politecnico di Torino

Slice and dice



Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 17 Elena Baralis Politecnico di Torino

Drill down

Customer	Region	Year	Quarter	Product	Category	Sub-Category	Item	Quantity	Price	Discount	Amount
Customer1	Region1	2005	Q1	Product1	Category1	Sub-Category1	Item1	10	100	0	1000

Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 15 Elena Baralis Politecnico di Torino

Slice and dice

Customer	Region	Year	Quarter	Product	Category	Sub-Category	Item	Quantity	Price	Discount	Amount
Customer1	Region1	2005	Q1	Product1	Category1	Sub-Category1	Item1	10	100	0	1000

Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 18 Elena Baralis Politecnico di Torino

Slice and dice

Metrica	Customer	City	Alban	Albania	Albania	Albania	Albania	Albania	Albania	Albania	Albania
Subcategory	Electronics	Alban									
Year	1997										
Category	Electronics	1997	\$ 10,614	\$ 29,299	\$ 5,300	\$ 5,638	\$ 20,047	\$ 5,665	\$ 20,391	\$ 2,939	\$ 4,497
Subcategory	Electronics	1997	\$ 10,614	\$ 29,299	\$ 5,300	\$ 5,638	\$ 20,047	\$ 5,665	\$ 20,391	\$ 2,939	\$ 4,497

Tratto da Gofarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Pivot

Metrica	Customer	City	Alban	Albania	Albania	Albania	Albania	Albania	Albania	Albania	Albania
Subcategory	Electronics	Alban									
Year	1997										
Category	Electronics	1997	\$ 10,614	\$ 29,299	\$ 5,300	\$ 5,638	\$ 20,047	\$ 5,665	\$ 20,391	\$ 2,939	\$ 4,497
Subcategory	Electronics	1997	\$ 10,614	\$ 29,299	\$ 5,300	\$ 5,638	\$ 20,047	\$ 5,665	\$ 20,391	\$ 2,939	\$ 4,497

Tratto da Gofarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Pivot

- Riorganizzazione dell'orientamento della struttura multidimensionale senza variare il livello di dettaglio
 - permette una visualizzazione più chiara delle stesse informazioni
 - la rappresentazione dei dati multidimensionali rimane sotto forma di "griglia"
 - due dimensioni sono gli assi principali della griglia
 - varia la posizione delle dimensioni nella griglia

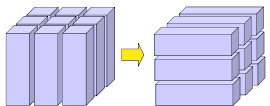
Tratto da Gofarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Pivot

Metrica	Customer	Region	North-East	Mid-Atlantic	South-East	Central	South	North-West	England	France	Germany
Category	Electronics	1997	\$ 1,104	\$ 1,174	\$ 284	\$ 1,201	\$ 2,241	\$ 2,514	\$ 2,184	\$ 561	\$ 1,078
Subcategory	Electronics	1997	\$ 1,104	\$ 1,174	\$ 284	\$ 1,201	\$ 2,241	\$ 2,514	\$ 2,184	\$ 561	\$ 1,078
Year	1997										
Category	Electronics	1997	\$ 1,104	\$ 1,174	\$ 284	\$ 1,201	\$ 2,241	\$ 2,514	\$ 2,184	\$ 561	\$ 1,078

Tratto da Gofarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Pivot



Tratto da Gofarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Estensioni del linguaggio SQL

Elena Baralis
Politecnico di Torino

Tratto da Gofarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Database and data mining group, Politecnico di Torino
DMG

Estensioni del linguaggio SQL

- Gli strumenti di interfaccia richiedono
 - nuove funzioni aggregate
 - funzioni aggregate utilizzate per le analisi economiche (media mobile, mediana, ...)
 - posizione nell'ordinamento
 - funzioni per la generazione di rapporti
 - definizione di totali parziali e cumulativi
- Lo standard ANSI ha accettato la proposta di nuove funzioni OLAP
 - incorporate nei prodotti a partire da DB2 UDB 7.1, Oracle 8i v2

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 25 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Funzioni OLAP in SQL

- Nuova classe di funzioni aggregate (funzioni OLAP) caratterizzate da:
 - finestra di calcolo, all'interno di cui è possibile specificare il calcolo di funzioni aggregate
 - possibilità di calcolare totali cumulativi e media mobile
 - nuove funzioni aggregate per ricavare la posizione nell'ordinamento (ranking)

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 26 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Estensioni del linguaggio SQL

- Gli strumenti di interfaccia richiedono
 - operatori per il calcolo di più raggruppamenti (group by) diversi nello stesso momento
- Lo standard SQL-99 (SQL3) ha esteso la clausola group by di SQL

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 26 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Finestra di calcolo

- Nuova clausola **window** caratterizzata da:
 - *partizionamento*: divide le righe in gruppi, senza collassarle (diverso da **group by**)
 - assenza di partizionamento: un solo gruppo
 - *ordinamento delle righe* separatamente all'interno di ogni partizione (simile a **order by**)
 - *finestra di aggregazione*: definisce il gruppo di righe su cui l'aggregato è calcolato, per ciascuna riga della partizione

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 29 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Base di dati di esempio

Vendite (Città, Mese, Importo)

Città	Mese	Importo
Milano	7	110
Milano	8	10
Milano	9	70
Milano	10	90
Milano	11	35
Milano	12	135
Torino	7	70
Torino	8	35
Torino	9	80
Torino	10	95
Torino	11	50
Torino	12	120

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 27 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Esempio

- Visualizzare, per ogni città e mese
 - l'importo delle vendite
 - la media rispetto al mese corrente e ai due mesi precedenti, separatamente per ogni città

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 30 Elena Baralis Politecnico di Torino

Esempio

- Partizionamento in base alla città
 - il calcolo della media è azzerato ogni volta che cambia la città
- Ordinamento in base al mese per calcolare la media mobile sul mese corrente insieme ai due mesi precedenti
 - senza ordinamento, il calcolo sarebbe privo di significato
- Dimensione della finestra di calcolo: riga corrente e le due righe precedenti

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 31 Elena Baralis Politecnico di Torino

Risultato

Città	Mese	Importo	MediaMobile
Milano	7	110	110
Milano	8	10	60
Milano	9	90	70
Milano	10	80	60
Milano	11	40	60
Milano	12	140	90
Torino	7	70	70
Torino	8	30	50
Torino	9	80	60
Torino	10	100	70
Torino	11	50	60
Torino	12	150	100

} Partizione 1
} Partizione 2

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 34 Elena Baralis Politecnico di Torino

Esempio

```

SELECT Città, Mese, Importo,
       AVG(Importo) OVER Wavg AS MediaMobile
FROM Vendite
WINDOW Wavg AS (PARTITION BY Città
                ORDER BY Mese
                ROWS 2 PRECEDING)
    
```

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 32 Elena Baralis Politecnico di Torino

Osservazioni

- E' necessario specificare l'ordinamento, perché l'aggregazione richiesta utilizza le righe in modo ordinato
 - l'ordinamento indicato non corrisponde ad un ordine predefinito delle righe in output
- Quando la finestra è incompleta, il calcolo è effettuato sulla parte presente
 - è possibile specificare che, se la finestra è incompleta, il risultato deve essere **NULL**
- E' possibile specificare più finestre di calcolo diverse

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 35 Elena Baralis Politecnico di Torino

Esempio

```

SELECT Città, Mese, Importo,
       AVG(Importo) OVER (PARTITION BY Città
                        ORDER BY Mese
                        ROWS 2 PRECEDING)
       AS MediaMobile
FROM Vendite
    
```

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 33 Elena Baralis Politecnico di Torino

Finestra di aggregazione

- La finestra mobile su cui è effettuato il calcolo dell'aggregato può essere definita
 - a *livello fisico*, formando il gruppo mediante conteggio delle righe
 - esempio: la riga corrente e le due righe precedenti
 - a *livello logico*, formando il gruppo in base alla definizione di un intervallo intorno alla chiave di ordinamento
 - esempio: il mese corrente e i due mesi precedenti

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 36 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Definizione intervallo fisico

- Tra un estremo inferiore e la riga corrente
`ROWS 2 PRECEDING`
- Tra un estremo inferiore e uno superiore
`ROWS BETWEEN 1 PRECEDING AND 1 FOLLOWING`
`ROWS BETWEEN 3 PRECEDING AND 1 PRECEDING`
- Tra l'inizio (o la fine) della partizione e la riga corrente
`ROWS UNBOUNDED PRECEDING (o FOLLOWING)`

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 37 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Raggruppamento logico

- Adatto per dati "sparsi", che hanno interruzioni nella sequenza
 - esempio: manca un mese nella sequenza
 - non è possibile specificare più di una chiave di ordinamento
 - è possibile utilizzare solo tipi di dato numerici o data come chiave di ordinamento (consentono di scrivere espressioni aritmetiche)

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 40 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Raggruppamento fisico

- Adatto per dati che non hanno interruzioni nella sequenza
 - esempio: non manca nessun mese nella sequenza
 - è possibile specificare più di una chiave di ordinamento
 - il raggruppamento ignora le separazioni
 - esempio: ordinamento per mese e anno
 - non occorrono formule per specificare come calcolare la finestra

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 38 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Applicazioni

- Calcolo di aggregati mobili
 - l'aggregato è calcolato su una finestra che "scorre" sui dati
 - esempi: media mobile, somma mobile
- Calcolo di totali cumulativi
 - il totale (cumulativo) è incrementato aggiungendo una riga alla volta
- Confronto tra dati dettagliati e dati complessivi

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 41 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Definizione intervallo logico

- Si utilizza il costrutto **range**, con la stessa sintassi dell'intervallo fisico
- E' necessario definire la distanza tra gli estremi dell'intervallo e il valore corrente sulla chiave di ordinamento
- Esempio
`RANGE 2 MONTH PRECEDING`

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 39 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Calcolo di totali cumulativi

- Visualizzare, per ogni città e mese
 - l'importo delle vendite
 - l'importo cumulativo delle vendite al trascorrere dei mesi, separatamente per ogni città

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 42 Elena Baralis Politecnico di Torino

Calcolo di totali cumulativi

- Partizionamento in base alla città
 - il calcolo della somma cumulativa è azzerato ogni volta che cambia la città
- Ordinamento (crescente) in base al mese per calcolare la somma al passare dei mesi
 - senza ordinamento, il calcolo sarebbe privo di significato
- Dimensione della finestra di calcolo: dalla riga iniziale della partizione alla riga corrente

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 43 Elena Baralis Politecnico di Torino

Confronto tra dati dettagliati e dati complessivi

- Visualizzare, per ogni città e mese
 - l'importo delle vendite
 - l'importo totale delle vendite sul periodo completo per la città corrente

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 46 Elena Baralis Politecnico di Torino

Calcolo di totali cumulativi

```
SELECT Città, Mese, Importo,
       SUM(Importo) OVER (PARTITION BY Città
                          ORDER BY Mese
                          ROWS UNBOUNDED PRECEDING)
       AS SommaCumul
FROM Vendite
```

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 44 Elena Baralis Politecnico di Torino

Confronto tra dati dettagliati e dati complessivi

- Partizionamento in base alla città
 - il calcolo del totale complessivo è azzerato ogni volta che cambia la città
- Non è necessario l'ordinamento
 - il totale complessivo è calcolato indipendentemente dall'ordinamento
- Non è necessaria la finestra di calcolo
 - è l'intera partizione

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 47 Elena Baralis Politecnico di Torino

Calcolo di totali cumulativi: risultato

Città	Mese	Importo	SommaCumul
Milano	7	110	110
Milano	8	10	120
Milano	9	90	210
Milano	10	80	290
Milano	11	40	330
Milano	12	140	470
Torino	7	70	70
Torino	8	30	100
Torino	9	80	180
Torino	10	100	280
Torino	11	50	330
Torino	12	150	480

Partizione 1 (Milano)
Partizione 2 (Torino)

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 45 Elena Baralis Politecnico di Torino

Confronto tra dati dettagliati e dati complessivi

```
SELECT Città, Mese, Importo,
       SUM(Importo) OVER (PARTITION BY Città)
       AS ImpTotale
FROM Vendite
```

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 48 Elena Baralis Politecnico di Torino

Confronto tra dati dettagliati e dati complessivi

Città	Mese	Importo	ImpTotale
Milano	7	110	470
Milano	8	10	470
Milano	9	90	470
Milano	10	80	470
Milano	11	40	470
Milano	12	140	470
Torino	7	70	480
Torino	8	30	480
Torino	9	80	480
Torino	10	100	480
Torino	11	50	480
Torino	12	150	480

Partizione 1 (Milano rows)
Partizione 2 (Torino rows)

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 49 Elena Baralis Politecnico di Torino

Confronto tra dati dettagliati e dati complessivi

```

SELECT Città, Mese, Importo
  Importo/SUM(Importo) OVER (
AS PercTotale
  Importo/SUM(Importo) OVER (PARTITION BY Città)
AS PercCittà
  Importo/SUM(Importo) OVER (PARTITION BY Mese)
AS PercMese
FROM Vendite
  
```

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 52 Elena Baralis Politecnico di Torino

Confronto tra dati dettagliati e dati complessivi

- Visualizzare, per ogni città e mese
 - l'importo
 - il rapporto tra l'importo della riga corrente per le vendite e il totale complessivo per città
 - il rapporto tra l'importo della riga corrente per le vendite e il totale complessivo per mese

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 50 Elena Baralis Politecnico di Torino

Confronto tra dati dettagliati e dati complessivi

Città	Mese	Importo	PercTotale	PercCittà	PercMese
Milano	7	110	110/950	110/470	110/180
Milano	8	10	10/950	10/470	10/40
Milano	9	90	90/950	90/470	90/170
Milano	10	80	80/950	80/470	80/180
Milano	11	40	40/950	40/470	40/90
Milano	12	140	140/950	140/470	140/290
Torino	7	70	70/950	70/480	70/180
Torino	8	30	30/950	30/480	30/40
Torino	9	80	80/950	80/480	80/170
Torino	10	100	100/950	100/480	100/180
Torino	11	50	50/950	50/480	50/90
Torino	12	150	150/950	150/480	150/290

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 53 Elena Baralis Politecnico di Torino

Confronto tra dati dettagliati e dati complessivi

- Tre finestre di calcolo diverse
 - totale complessivo: nessun partizionamento
 - totale per città: partizionamento per città
 - totale per mese: partizionamento per mese
- Non è necessario l'ordinamento per nessuna finestra
 - il totale complessivo è calcolato indipendentemente dall'ordinamento
- La finestra di calcolo è sempre l'intera partizione

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 51 Elena Baralis Politecnico di Torino

Group by e finestre

- È possibile abbinare l'uso di finestre con il raggruppamento eseguito dalla clausola **group by**
- La "tabella temporanea" generata dall'esecuzione della clausola **group by** (con eventuale calcolo di funzioni aggregate abbinato al **group by**) diviene l'operando a cui applicare le operazioni definite per la **window**

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 54 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Esempio

- Si supponga che la tabella **vendite** contenga informazioni sulle vendite con granularità giornaliera
- Visualizzare, per ogni città e mese
 - l'importo delle vendite
 - la media rispetto al mese corrente e ai due mesi precedenti, separatamente per ogni città

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 55 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Funzioni di ranking

- Funzioni per calcolare la posizione di un valore all'interno di una partizione
 - funzione **rank** () : calcola la posizione, lasciando intervalli vuoti successivi alla presenza di "pari merito"
 - esempio: 2 primi, subito dopo vi è il terzo nella graduatoria
 - funzione **denserank** () : calcola la posizione, senza lasciare intervalli vuoti successivi alla presenza di "pari merito"
 - esempio: 2 primi, subito dopo vi è il secondo nella graduatoria

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 56 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Esempio

- E' necessario raggruppare i dati per mese e calcolare l'importo totale per mese prima di effettuare il calcolo della media mobile
 - si usa la clausola group by per calcolare il totale mensile
- La tabella temporanea generata dalla prima aggregazione diviene l'operando su cui definire la finestra di calcolo

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 56 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Esempio

- Visualizzare, per ogni città nel mese di dicembre
 - l'importo delle vendite
 - la posizione nella graduatoria

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 59 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Esempio

```

SELECT Città, Mese, SUM(Importo) AS TotMese,
       AVG(SUM(Importo)) OVER (PARTITION BY Città
                               ORDER BY Mese
                               ROWS 2 PRECEDING)
       AS MediaMobile
FROM Vendite, ...
WHERE <cond. join>
GROUP BY Città, Mese
  
```

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 57 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DMG

Esempio

- Non occorre partizionamento
 - una sola partizione che include tutte le città
- Ordinamento in base all'importo per stilare la graduatoria
 - senza ordinamento, il calcolo sarebbe privo di significato
- La finestra di calcolo è l'intera partizione

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 60 Elena Baralis Politecnico di Torino

Esempio

```

SELECT Città, Importo,
       RANK() OVER (ORDER BY Importo DESC)
       AS Graduatoria
FROM Vendite
WHERE Mese = 12
    
```

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 61 Elena Baralis Politecnico di Torino

Esempio

```

SELECT Città, Importo,
       RANK() OVER (ORDER BY Importo DESC)
       AS Graduatoria
FROM Vendite
WHERE Mese = 12
ORDER BY Città
    
```

Città	Importo	Graduatoria
Milano	140	2
Torino	150	1

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 64 Elena Baralis Politecnico di Torino

Risultato

Città	Importo	Graduatoria
Torino	150	1
Milano	140	2

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 62 Elena Baralis Politecnico di Torino

Estensioni della clausola group by

- Gli spreadsheet multidimensionali richiedono più totali parziali "in un colpo solo"
 - somma delle vendite per mese e città
 - somma delle vendite per mese
 - somma delle vendite per città
- Per motivi di efficienza è opportuno evitare
 - letture multiple dei dati
 - ordinamenti ridondanti dei dati

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 65 Elena Baralis Politecnico di Torino

Ordinamento del risultato

- L'ordinamento del risultato è ottenuto mediante la clausola **order by**
 - può essere diverso dall'ordinamento delle finestre di calcolo
- Esempio: ordinare il risultato dell'esempio precedente in ordine alfabetico di città

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 63 Elena Baralis Politecnico di Torino

Estensioni della clausola group by

- Lo standard SQL-99 ha esteso la sintassi della clausola **group by**
 - **rollup** per calcolare le aggregazioni su tutti i gruppi ottenuti togliendo in ordine una colonna per volta dall'insieme specificato di colonne
 - **cube** per calcolare le aggregazioni su tutte le possibili combinazioni delle colonne specificate
 - **grouping sets** per specificare un elenco di raggruppamenti richiesti (diversi da quelli ottenibili con le due clausole precedenti)
 - () per richiedere totali generali (nessun raggruppamento)

Database and data mining group, Politecnico di Torino
Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 66 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DRG

Rollup: esempio

- Si considerino le seguenti tabelle
 Tempo (**Tkey**, **Giorno**, **Mese**, **Anno**, ...)
 Supermercato (**Skey**, **Città**, **Regione**, ...)
 Prodotto (**Pkey**, **NomeP**, **Marca**, ...)
 Vendite (**Skey**, **Tkey**, **Pkey**, **Importo**)
- Calcolare il totale delle vendite nel 2000 per le seguenti diverse combinazioni di attributi
 - prodotto, mese e città
 - mese, città
 - città

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 67 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DRG

Cube: esempio

- Calcolare il totale delle vendite nel 2000 per *tutte* le combinazioni dei seguenti attributi
 - prodotto, mese, città
- Si devono calcolare le seguenti aggregazioni:
 - prodotto, mese, città
 - prodotto, mese
 - mese, città
 - prodotto, città
 - prodotto
 - mese
 - città
 - nessun raggruppamento

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 70 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DRG

Rollup: esempio

```
SELECT Città, Mese, Pkey,
       SUM(Importo) AS TotVendite
FROM Tempo T, Supermercato S, Vendite V
WHERE T.Tkey = V.Tkey
      AND S.Skey = V.Skey
      AND Anno = 2000
GROUP BY ROLLUP (Città, Mese, Pkey)
```

- L'ordinamento delle colonne in **rollup** determina quali aggregati sono calcolati

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 68 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DRG

Cube: esempio

```
SELECT Città, Mese, Pkey,
       SUM(Importo) AS TotVendite
FROM Tempo T, Supermercato S, Vendite V
WHERE T.Tkey = V.Tkey
      AND S.Skey = V.Skey
      AND Anno = 2000
GROUP BY CUBE (Città, Mese, Pkey)
```

- L'ordinamento delle colonne in **cube** è ininfluente

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 71 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DRG

Rollup: risultato

Città	Mese	Pkey	TotVendite
Milano	7	145	110
Milano	7	150	10
Milano
Milano	7	NULL	8500
Milano	8
Milano	NULL	NULL	150000
Torino	150
Torino	...	NULL	2500
Torino	NULL	NULL	135000
..
NULL	NULL	NULL	25005000

- “superaggregati” sono rappresentati con **NULL**

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 69 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino
DRG

Calcolo del cubo

- Si considerano le proprietà distributive e algebriche delle funzioni aggregate
 - le funzioni aggregate *distributive* (**min**, **max**, **sum**, **count**) possono essere calcolate a partire da aggregazioni su un numero maggiore di attributi (con granularità maggiore)
 - Esempio: dall'importo totale su prodotto e mese, si calcola l'importo totale per mese
 - per le funzioni aggregate *algebriche* (**avg**, ...) è possibile il calcolo a partire da aggregazioni su un numero maggiore di attributi (con granularità maggiore), pur di memorizzare opportuni risultati intermedi
 - Esempio: per la media serve conoscere
 - il valore della media nel gruppo
 - il numero di elementi per gruppo

Copyright – Tutti i diritti riservati DATA WAREHOUSE: OLAP - 72 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino

Calcolo del cubo

- Per rendere più efficiente il calcolo del cubo, si usano le proprietà distributive/algebriche delle funzioni aggregate
 - si usano i risultati di **group by** già calcolati
 - l'operazione di **rollup** richiede una sola operazione di ordinamento
 - il cubo può essere visto come una combinazione di più operazioni di **rollup** (in ordine opportuno)
 - si sfruttano operazioni di sort già eseguite (anche parzialmente)
 - è possibile utilizzare l'ordinamento delle colonne (A,B) per ordinare (A,C)

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 73 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino

Grouping Set: esempio

- Calcolare il totale delle vendite nel 2000 per le seguenti combinazioni di attributi
 - mese
 - mese, città, prodotto
- Eseguire un rollup richiederebbe il calcolo di aggregati aggiuntivi

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 74 Elena Baralis Politecnico di Torino

Database and data mining group, Politecnico di Torino

Grouping Set: esempio

```

SELECT Città, Mese, Pkey,
       SUM(Importo) AS TotVendite
FROM Tempo T, Supermercato S, Vendite V
WHERE T.Tkey = V.Tkey
      AND S.Skey = V.Skey
      AND Anno = 2000
GROUP BY GROUPING SETS
         (Mese, (Città,Mese,Pkey))
  
```

Copyright - Tutti i diritti riservati DATA WAREHOUSE: OLAP - 75 Elena Baralis Politecnico di Torino