

DATA WAREHOUSE DI ATENEO

INTRODUZIONE AL DATAWAREHOUSING

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

APPC – SETTORE PROGRAMMAZIONE
DI ATENEO E ANALISI DATI

1.	<i>Il Data Warehousing</i>	3
1.1	Definizione e caratteristiche.....	3
1.2	Il modello multidimensionale.....	4
1.2.1	Restrizione	5
1.2.2	Aggregazione	6
1.3	Accedere al DW	7
1.3.1	Reportistica	7
1.3.2	OLAP.....	7
2.	<i>Reportistica operativa</i>	10
2.1	Raggruppamento a pagine.....	10
2.2	Drilling.....	13

1. Il Data Warehousing

Il fenomeno del *data warehousing* nasce dall'enorme accumulo di dati registrato nell'ultimo decennio, e dalla pressante richiesta di utilizzare attivamente questi dati per scopi che superino quelli, di routine, legati all'elaborazione giornaliera. Uno scenario tipico è quello di una grande azienda, con numerose filiali, i cui dirigenti desiderano quantificare e valutare il contributo dato da ciascuna di esse al rendimento commerciale globale dell'impresa. Essendo i dati elementari sulle attività svolte disponibili nel database aziendale, un approccio possibile consiste nel chiedere ai tecnici che lo amministrano di formulare una interrogazione *ad hoc* che effettui i calcoli necessari sui dati (in genere aggregazioni). Quando i tecnici saranno riusciti a formulare l'interrogazione voluta (tipicamente in SQL, dopo avere a lungo consultato i cataloghi del database), e una volta terminata la sua elaborazione (il che richiederà probabilmente alcune ore, dato l'elevato volume dei dati, la complessità dell'interrogazione e la contemporanea incidenza sui dati delle interrogazioni facenti parte del normale carico di lavoro), ai dirigenti verrà restituito un "rapporto" sotto forma di foglio elettronico su cui basare le loro decisioni future. Già da parecchi anni si è capito che questa via è difficilmente percorribile, perché porta a un inutile consumo di tempo e risorse e al contempo non sempre produce il risultato desiderato. Tra l'altro, mescolare questo tipo di interrogazioni "analitiche" con quelle "transazionali" di routine porta a inevitabili rallentamenti che rendono insoddisfatti gli utenti di entrambe le categorie. L'idea alla base del *data warehousing* è allora quella di separare l'elaborazione di tipo analitico (OLAP, *On-Line Analytical Processing*) da quella legata alle transazioni (OLTP, *On-Line Transactional Processing*), costruendo un nuovo raccoglitore di informazioni che integri i dati elementari provenienti da sorgenti di varia natura, li organizzi in una forma appropriata e li renda quindi disponibili per scopi di analisi e valutazione finalizzate alla pianificazione e al processo decisionale.

1.1 Definizione e caratteristiche

Con il termine *data warehousing* intendiamo una collezione di metodi, tecnologie e strumenti di ausilio al cosiddetto "lavoratore della conoscenza" (*knowledge worker*: dirigente, amministratore, gestore, analista) per condurre analisi dei dati finalizzate all'attuazione di processi decisionali e al miglioramento del patrimonio informativo. Fattori distintivi e requisiti indispensabili del *processo di data warehousing*, ossia del complesso di attività che consentono di trasformare i dati operazionali in conoscenza a supporto delle decisioni, sono i seguenti:

- *accessibilità* a utenti con conoscenze limitate di informatica e strutture dati;
- *integrazione* dei dati sulla base di un modello standard dell'impresa;
- *flessibilità di interrogazione* per trarre il massimo vantaggio dal patrimonio informativo esistente;
- *sintesi* per permettere analisi mirate ed efficaci;
- *rappresentazione multidimensionale* per offrire all'utente una visione intuitiva ed efficacemente manipolabile delle informazioni;
- *correttezza e completezza* dei dati integrati.

Al centro del processo, il *data warehouse* (letteralmente, magazzino di dati) è un contenitore di dati che diventa garante dei requisiti esposti e presenta le seguenti caratteristiche:

- è orientato ai soggetti di interesse;
- è integrato e consistente;
- è rappresentativo dell'evoluzione temporale e non volatile.

Si intende che il DW è orientato ai soggetti perché si incentra sui concetti di interesse dell'azienda, quali i clienti, i prodotti, le vendite, gli ordini. Viceversa, i database operazionali sono organizzati intorno alle differenti applicazioni del dominio aziendale.

L'accento sugli aspetti di integrazione e consistenza è importante poiché il DW si appoggia a più fonti di dati eterogenee: dati estratti dall'ambiente di produzione, e quindi originariamente archiviati in basi di dati aziendali, o addirittura provenienti da sistemi informativi esterni all'azienda. Di tutti

questi dati il DW si impegna a restituire una visione unificata. In linea di massima, la costruzione di un sistema di data warehousing non comporta l'inserimento di nuove informazioni bensì la riorganizzazione di quelle esistenti, e implica pertanto l'esistenza di un sistema informativo.

Mentre i dati operazionali coprono un arco temporale di solito piuttosto limitato, poiché la maggior parte delle transazioni coinvolge i dati più recenti, il DW deve permettere analisi che spazino sulla prospettiva di alcuni anni. Per questo motivo, il DW è aggiornato a intervalli regolari a partire dai dati operazionali ed è in crescita continua. Volendo fare un paragone possiamo supporre che, a intervalli regolari, venga scattata una fotografia istantanea dei dati operazionali. La progressione delle fotografie scattate viene immagazzinata nel DW, dove genera un film che documenta la situazione aziendale da un istante zero fino al tempo attuale.

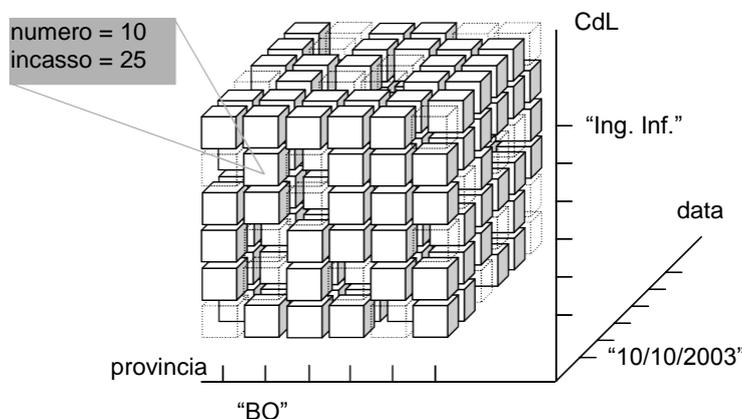
Le peculiari caratteristiche delle interrogazioni OLAP fanno sì che i dati nel DW siano normalmente rappresentati in forma *multidimensionale*. L'idea di base è quella di vedere i dati come punti in uno spazio le cui dimensioni corrispondono ad altrettante possibili dimensioni di analisi; ciascun punto, rappresentativo di un evento accaduto nell'azienda, viene descritto tramite un insieme di misure di interesse per il processo decisionale. Una descrizione più approfondita del modello multidimensionale è oggetto del prossimo paragrafo.

1.2 Il modello multidimensionale

Il modello multidimensionale prende le mosse dalla constatazione che gli oggetti che influenzano il processo decisionale sono *fatti* che accadono nel mondo aziendale, quali per esempio le immatricolazioni, i movimenti contabili, le tesi di laurea, le progressioni di carriera. Le occorrenze di un fatto corrispondono a *eventi* accaduti. Per ciascun fatto, interessano in particolare i valori di un insieme di *misure* o *metriche* che descrivono quantitativamente gli eventi: l'importo di un mandato, il numero di esami svolti, l'importo lordo di uno stipendio.

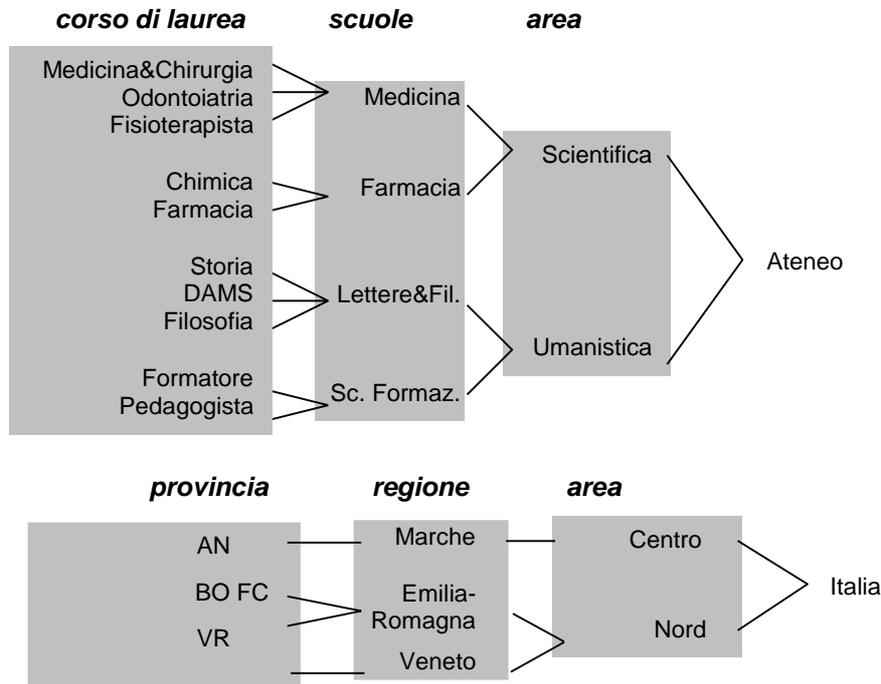
Gli eventi che accadono nell'azienda sono evidentemente tantissimi, troppi per poter essere analizzati singolarmente. Per poterli agevolmente selezionare e raggruppare si immagina allora di collocarli in uno spazio n -dimensionale i cui assi, chiamati appunto *dimensioni* di analisi, definiscono diverse prospettive per la loro identificazione. Per esempio, le immatricolazioni universitarie possono essere rappresentate in uno spazio tridimensionale le cui dimensioni sono i corsi di laurea, le province di provenienza degli studenti e le date.

È proprio il concetto di dimensione che ha dato origine alla metafora del *cubo* per la rappresentazione dei dati multidimensionali. Secondo questa metafora, gli eventi corrispondono a celle di un cubo i cui spigoli rappresentano le dimensioni di analisi (se le dimensioni sono più di tre, si tratta più propriamente di un *ipercubo*). Ogni cella del cubo contiene un valore per ciascuna misura. La figura seguente mostra una rappresentazione grafica intuitiva di un cubo in cui il fatto descritto sono le immatricolazioni. Le dimensioni di analisi sono CdL, provincia e data; un evento corrisponde all'immatricolazione a un certo corso di laurea di studenti provenienti da una certa provincia in un certo giorno, ed è descritto da due misure: il numero di immatricolazioni e l'incasso. La figura mette in evidenza il fatto che il cubo è sparso, ossia che molti eventi non si sono in effetti verificati: chiaramente, non da tutte le province ci si immatricula tutti i giorni in tutti i corsi di laurea!



Normalmente, ciascuna dimensione è associata a una *gerarchia* di livelli di aggregazione che ne raggruppa i valori in diversi modi. Chiameremo *attributi* i livelli che compongono una gerarchia. La

figura seguente propone un piccolo esempio di gerarchie sulle dimensioni CdL e provincia: i corsi di laurea sono raggruppati in scuole, ulteriormente suddivisi in aree; le province si trovano in regioni che a loro volta fanno parte di aree geografiche. In cima a ciascuna gerarchia si trova un livello fittizio che raggruppa tutti i valori relativi a una dimensione.



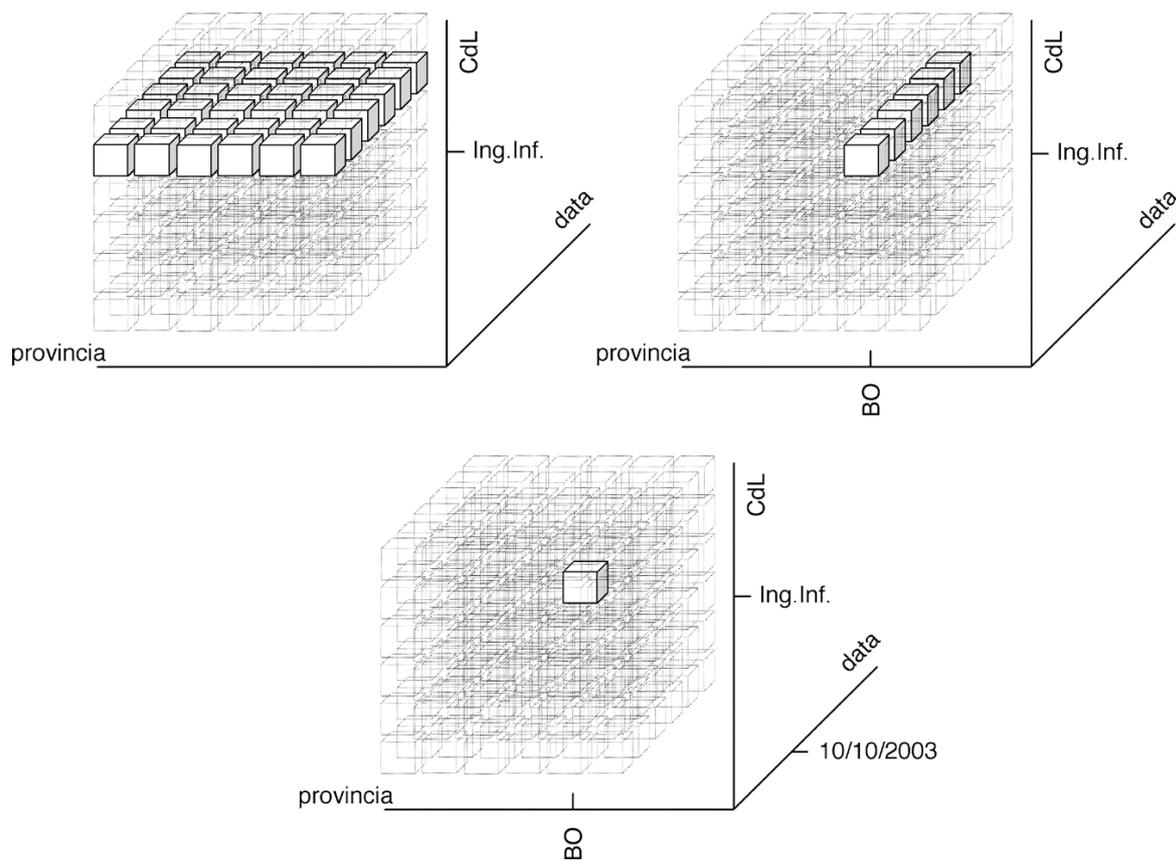
Riassumendo, un cubo multidimensionale è incentrato su un fatto di interesse per il processo decisionale. Esso rappresenta un insieme di eventi, descritti quantitativamente da misure numeriche. Ogni asse del cubo rappresenta una possibile *dimensione* di analisi; ciascuna dimensione può essere vista a più livelli di dettaglio individuati da *attributi* strutturati in gerarchie.

Osserviamo ora che le informazioni rappresentate nel cubo multidimensionale, pur costituendo di fatto una sintesi di quelle memorizzate nella base di dati operativa (dove per esempio vengono distinte le *single* immatricolazioni), sono ancora difficilmente fruibili dall'utente a causa della loro quantità. Le tecniche per ridurre la quantità di dati e ottenere così informazioni utili sono essenzialmente due: la *restrizione* e l'*aggregazione*; per entrambe, come vedremo nei due paragrafi successivi, la metafora del cubo offre un'agile e intuitiva chiave di interpretazione.

1.2.1 Restrizione

Restringere i dati significa ritagliare una porzione dal cubo circoscrivendo il campo di analisi.

La forma più semplice di restrizione è il cosiddetto *slicing* (letteralmente, affettatura) dei dati, illustrato in figura, in cui si riduce la dimensionalità del cubo fissando un valore per una o più dimensioni. Vediamo un esempio per il cubo delle immatricolazioni. Se si fissa un valore per una delle dimensioni, per esempio CdL='IngegneriaInformatica', si ottiene come risultato l'insieme degli eventi associati alle immatricolazioni effettuate presso il corso di laurea in Ingegneria Informatica: secondo la metafora si tratterà di un piano, ovvero una "fettina" di dati agevolmente visualizzabile all'interno di un foglio elettronico. Se vengono fissate due dimensioni, per esempio CdL='IngegneriaInformatica' e provincia='BO', il risultato sono tutte le immatricolazioni a Ingegneria Informatica effettuate dalla provincia di Bologna; graficamente, l'intersezione di due piani perpendicolari, ovvero una retta. Infine, se tutte le dimensioni vengono fissate, si identifica un unico evento corrispondente a un punto nello spazio tridimensionale delle immatricolazioni.

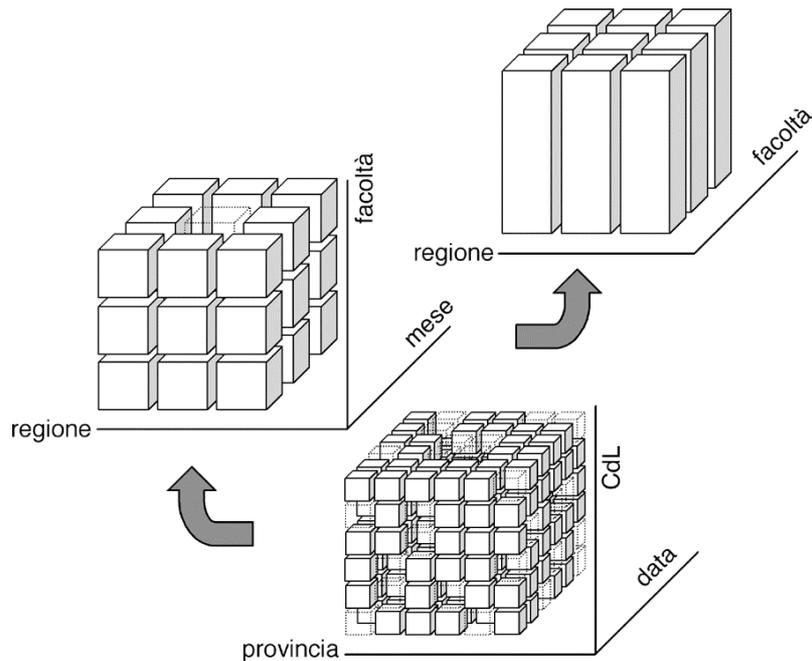


La *selezione* è una generalizzazione dello *slicing* in cui si riduce la grandezza del cubo esprimendo condizioni sugli attributi dimensionali. Per esempio, si possono selezionare le sole immatricolazioni da province dell'area centro a corsi di laurea della facoltà di Ingegneria nei giorni di ottobre 2003.

1.2.2 Aggregazione

L'aggregazione è un meccanismo di importanza fondamentale nelle basi di dati multidimensionali. Si supponga di voler analizzare le immatricolazioni non nel loro dettaglio giornaliero, bensì a livello mensile; continuando la metafora del cubo, ciò significa raggruppare, per ciascun corso di laurea e provincia, tutti gli eventi relativi ai giorni di uno stesso mese in un unico macro-evento. Nel cubo così aggregato, ciascun evento conterrà una sintesi dei dati presenti negli eventi che esso aggrega: nel caso in esame, il numero totale di immatricolazioni effettuate nel mese e l'incasso complessivo calcolati sommando i valori elementari delle corrispondenti misure. Aggregando ulteriormente sul tempo, per ogni combinazione CdL-provincia si possono ottenere tre soli eventi: uno per ciascun anno. Al massimo livello di aggregazione sulla dimensione tempo, ciascuna combinazione corrisponde a un unico evento che riporta il numero totale di immatricolazioni a un corso di laurea da una provincia nei tre anni e l'incasso complessivo.

L'aggregazione può essere operata contemporaneamente su più dimensioni. Per esempio, come mostrato in figura, è possibile aggregare le immatricolazioni per regione, facoltà e mese, nonché solo per regione e facoltà. Inoltre, selezione e aggregazione possono essere combinate per permettere un processo di analisi mirato con precisione alle esigenze dell'utente.



1.3 Accedere al DW

L'ultimo livello comune a tutte le architetture di *data warehousing* è quello dell'analisi. Infatti, una volta che i dati sono stati ripuliti, integrati e trasformati, occorre capire come trarne il massimo vantaggio informativo. Esistono in sostanza tre approcci differenti, supportati da altrettante categorie di strumenti, all'interrogazione di un DW da parte degli utenti finali: *reportistica*, *OLAP* e *data mining*, che presenteremo nei prossimi paragrafi in ordine crescente di complessità.

1.3.1 Reportistica

Questo approccio è orientato agli utenti che hanno necessità di accedere, a intervalli di tempo predefiniti, a informazioni strutturate in modo pressoché invariabile. Per esempio, un ateneo deve consegnare al ministero rapporti annuali riepilogativi sul numero di laureati per genere, anno di prima immatricolazione, corso di studi e titolo di studio posseduto. Di questi rapporti è nota a priori la forma, che cambia solo a seguito di variazioni nella normativa vigente. Il progettista può allora disegnare l'interrogazione che genera il rapporto nella forma voluta e "congelarla" all'interno di un'applicazione perché possa essere eseguita sui dati correnti quando l'utente ne ha l'effettiva necessità.

1.3.2 OLAP

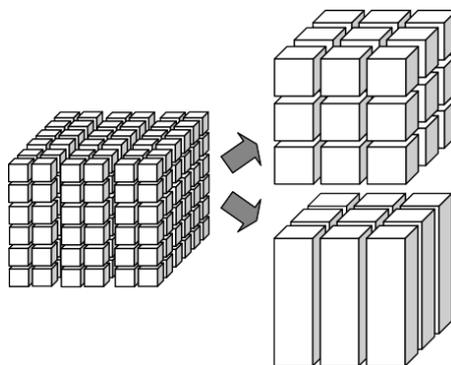
È forse la principale modalità di fruizione delle informazioni contenute in un DW, e consente a utenti le cui necessità di analisi non siano facilmente identificabili a priori di analizzare ed esplorare interattivamente i dati sulla base del modello multidimensionale. Mentre gli utenti degli strumenti di reportistica svolgono un ruolo essenzialmente passivo, gli utenti OLAP sono in grado di costruire attivamente una sessione di analisi complessa in cui ciascun passo effettuato è conseguenza dei risultati ottenuti al passo precedente.

Una sessione OLAP consiste in pratica in un *percorso di navigazione* che riflette il procedimento di analisi di uno o più fatti di interesse sotto diversi aspetti e a diversi livelli di dettaglio. Questo percorso si concretizza in una sequenza di interrogazioni che spesso non vengono formulate direttamente, ma per differenza rispetto all'interrogazione precedente. Il risultato delle interrogazioni è di tipo multidimensionale; poiché le capacità umane di ragionare in più di tre dimensioni sono molto limitate, gli strumenti OLAP rappresentano tipicamente i dati in modo tabellare evidenziando le

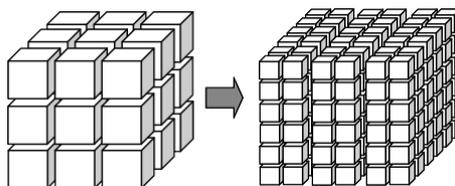
diverse dimensioni mediante intestazioni multiple, colori ecc.

Ogni passo della sessione di analisi è scandito dall'applicazione di un operatore OLAP che trasforma l'ultima interrogazione formulata in una nuova interrogazione. Gli operatori più comuni sono *roll-up*, *drill-down*, *slice-and-dice*, *pivoting*.

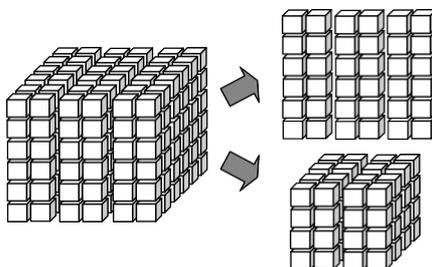
Roll-up significa letteralmente arrotolare o alzare, e induce un aumento nell'aggregazione dei dati eliminando un livello di dettaglio da una gerarchia. Il *roll-up* può anche portare alla diminuzione della dimensionalità del risultato, qualora tutti i dettagli di una gerarchia vengano eliminati. La figura seguente schematizza l'operazione di *roll-up*, con e senza diminuzione della dimensionalità, attraverso la metafora del cubo.



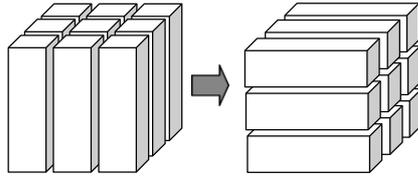
L'operatore di *drill-down* (letteralmente, trivellare) è duale al *roll-up*: infatti, come mostrato in figura, esso diminuisce l'aggregazione dei dati introducendo un ulteriore livello di dettaglio in una gerarchia.



Il termine *slice-and-dice* (letteralmente, tagliare a fette e cubetti) è uno dei più abusati nella letteratura sul *data warehousing*, anche se con differenti significati. Alcuni autori lo usano per denotare genericamente l'intero processo di navigazione OLAP, altri per indicare operazioni di selezione e proiezione sui dati. Chiameremo *slicing* l'operazione che riduce la dimensionalità del cubo fissando un valore per una delle dimensioni, *selezione* o *filtraggio* l'operazione che riduce l'insieme dei dati oggetto di analisi attraverso la formulazione di un criterio di selezione. La figura seguente mostra esempi di *slicing* e di selezione.



L'operazione di *pivoting* comporta un cambiamento nella modalità di presentazione con l'obiettivo di analizzare le stesse informazioni sotto diversi punti di vista. Seguendo la metafora multidimensionale, effettuare il *pivoting* significa ruotare il cubo in modo da riorganizzarne le celle secondo una nuova prospettiva, ossia portando in primo piano una differente combinazione di dimensioni (si veda la figura).



2. Reportistica operativa

2.1 Raggruppamento a pagine

Effettuare un raggruppamento a pagine con un attributo dimensionale coincide con l'esecuzione di precise operazioni OLAP; per capire quali sono utilizziamo l'esempio del "Report 1".

Il report analizza il **Numero Iscritti** (metrica basata sul fatto *Iscritto*) lungo gli attributi dimensionali **Anno**, **Genere** e **Regione** ognuno appartenente ad una specifica gerarchia. (Fig. 1).

Allo stato iniziale (Fig. 2) il report è paginato per **Genere** con valore "Donna"; in termini di operatori OLAP, rispetto al cubo iniziale, è stato fatto uno *slicing*. Cambiare il valore di **Genere** in "Uomo", per farlo bisogna selezionare tale valore dal menù a tendina che compare di fianco all'attributo, significa effettuare un nuovo *slicing* sempre sul medesimo attributo ma con valore fisso differente (Fig. 3).

Tra i valori selezionabili per l'attributo in pagina compare (non è detto che avvenga sempre) il valore "Total"; esso rappresenta l'unione di tutti i valori che l'attributo può assumere. Selezionare "Total" significa eliminare tutti i dettagli della gerarchia Persona cioè effettuare un *roll-up* con diminuzione della dimensionalità (Fig. 4).

Fig. 1 - Cubo di analisi

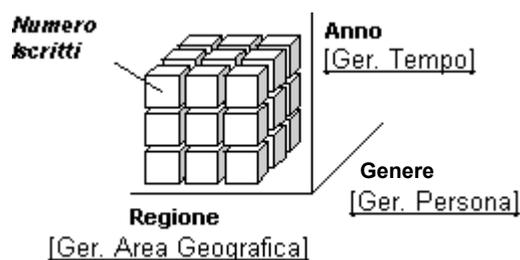


Fig. 2 - Report 1: stato iniziale con paginazione Genere="Donna" (slicing)

PAGE-BY: Genere:

Metrics Regione	Iscritti		
	2018	2017	...
PIEMONTE	***	***	
VALLE D'AOSTA	***	***	
LOMBARDIA	***	***	
TRENTINO ALTO ADIGE	***	***	
VENETO	***	***	
FRIULI-VENEZIA GIULIA	***	***	
LIGURIA	***	***	
EMILIA-ROMAGNA	***	***	***
TOSCANA	***	***	
UMBRIA	***	***	
MARCHE	***	***	
LAZIO	***	***	
ABRUZZO	***	***	
MOLISE	***	***	
CAMPANIA	***	***	
PUGLIA	***	***	
BASILICATA	***	***	
CALABRIA	***	***	
SICILIA	***	***	
SARDEGNA	***	***	

Fig. 3 - Report 1: cambio paginazione Genere= "Uomo" (nuovo slicing)

Genere:

Metrics Regione	Iscritti		
	2018	2017	...
PIEMONTE	***	***	
VALLE D'AOSTA	***	***	
LOMBARDIA	***	***	
TRENTINO ALTO ADIGE	***	***	
VENETO	***	***	
FRIULI-VENEZIA GIULIA	***	***	
LIGURIA	***	***	
EMILIA-ROMAGNA	***	***	***
TOSCANA	***	***	
UMBRIA	***	***	
MARCHE	***	***	
LAZIO	***	***	
ABRUZZO	***	***	
MOLISE	***	***	
CAMPANIA	***	***	
PUGLIA	***	***	
BASILICATA	***	***	
CALABRIA	***	***	
SICILIA	***	***	
SARDEGNA	***	***	

Fig. 4 - Report 1: cambio paginazione *Genere='Total'* (roll-up con diminuzione della dimensionalità)

Genere: Total ▼

Metrics Regione	Iscritti		
	2018	2017	...
PIEMONTE	***	***	
VALLE D'AOSTA	***	***	
LOMBARDIA	***	***	
TRENTINO ALTO ADIGE	***	***	
VENETO	***	***	
FRIULI-VENEZIA GIULIA	***	***	
LIGURIA	***	***	
EMILIA-ROMAGNA	***	***	***
TOSCANA	***	***	
UMBRIA	***	***	
MARCHE	***	***	
LAZIO	***	***	
ABRUZZO	***	***	
MOLISE	***	***	
CAMPANIA	***	***	
PUGLIA	***	***	
BASILICATA	***	***	
CALABRIA	***	***	
SICILIA	***	***	
SARDEGNA	***	***	
Total	***	***	***

2.2 Drilling

Il *drilling* (navigazione) consente di passare ai diversi livelli di informazioni di un report. È possibile eseguire il *drilling* da un determinato livello di informazioni a livelli con maggiori o minori informazioni seguendo specifici percorsi gerarchici di navigazione; questi sono costruiti *ad hoc* per soddisfare le analisi di interesse del report e associati allo stesso. Per capire meglio quali operazioni permette di eseguire un *drill* in Microstrategy Web basiamoci sull'esempio del "Report 2"; il cubo di analisi su cui si basa il report è quello del "Report 1" (Fig. 1). I percorsi di navigazione associati al report sono i seguenti:

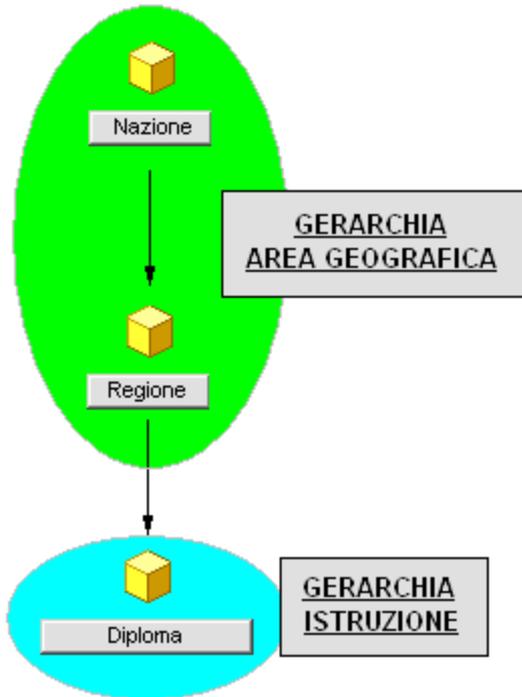


Fig. 5 - Percorso Area/Istruzione

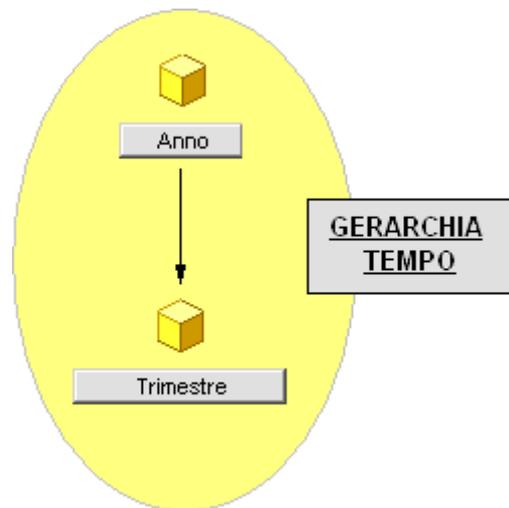


Fig. 6 – PercorsoTempo

Un percorso di navigazione può contenere attributi appartenenti a gerarchie differenti (fig. 5). Il report allo stato iniziale (fig. 7) presenta i seguenti attributi:

- Regione:
partendo da questo è possibile muoversi all'interno del percorso di fig. 5 ;
- Anno:
partendo da questo è possibile muoversi all'interno del percorso di fig. 6 ;
- Genere:
partendo da questo non è possibile andare in nessuna direzione in quanto i percorsi associati al report non contengono l'attributo.
In Microstrategy Web è possibile effettuare un'operazione di *drill* in diversi modi; per capirli meglio esamineremo di seguito un insieme di esempi che racchiudono la panoramica delle operazioni OLAP realizzabili con il *drilling*.

■ Drilling con il pulsante destro del mouse

Posizionandosi sugli attributi e premendo il tasto destro del mouse compare un menù a tendina che mostra le direzioni lungo le quali è possibile muoversi: la freccia ↓ indica che il nuovo attributo, all'interno del percorso utilizzato, si trova ad un livello più basso; la freccia ↑ indica invece che il nuovo attributo si trova ad un livello più alto del percorso; l'assenza di attributi indica che non è possibile muoversi in alcuna direzione (fig. 7).

Vediamo alcuni esempi:

● drill da Regione a ↑Nazione (fig. 8)

Posizionandosi sull'intestazione di **Regione** e selezionando **↑ Nazione** si ha come risultato quello di aggregare i dati per **Nazione**; ci si è spostati all'interno della gerarchia Area Geografica aumentando il livello di aggregazione: si è quindi effettuato un *roll-up*.

● drill da Regione a ↑ Nazione con mantenimento del padre (fig. 9)

In Microstrategy Web è possibile decidere se mantenere all'interno del nuovo report, risultato di un *drill*, l'attributo da cui è partita l'operazione stessa. Questa preferenza è impostata, in ogni report, dal progettista dello stesso e può essere modificata solo utilizzando il riquadro *Drill* (pag. 19). Tutte le altre modalità di *drill* utilizzano quindi il valore di default impostato.

Nell'esempio di fig. 8 questa preferenza non era impostata; questo esempio è identico a quello precedente ad eccezione proprio di questa preferenza che supponiamo impostata. Il risultato è quello di aggregare i dati per **Nazione** e per **Regione**; a causa del mantenimento di **Regione** il livello di aggregazione dei dati non è cambiato, si è solo acquisita un'informazione in più cioè la Nazione di appartenenza di ogni Regione.

● drill da "VALLE D'AOSTA" a ↓ Diploma (fig. 10)

Posizionandosi sul valore "**VALLE D'AOSTA**" di **Regione** e selezionando **↓ Diploma** si ha come risultato quello di aggregare i dati per **Diploma** visualizzando solo i diplomi relativi a tale regione; ci si è spostati verso la gerarchia Istruzione cambiando il livello di aggregazione e si è ristretto l'insieme dei dati di interesse effettuando una selezione.

● drill da "2001" a ↓ Trimestre (fig. 11)

Posizionandosi sul valore "**2001**" di **Anno** e selezionando **↓ Trimestre** si ha come risultato quello di aggregare i dati per **Trimestre** visualizzando solo i trimestri relativi a tale anno; ci si è spostati all'interno della gerarchia Tempo diminuendo il livello di aggregazione e si è ristretto l'insieme dei dati di interesse: si è quindi effettuato un *drill-down* con selezione.

Da notare in questo esempio che se anche si fosse mantenuto il padre si sarebbe comunque ottenuto il medesimo *drill-down* con selezione con in più l'informazione dell'anno di appartenenza dei trimestri: il 2001.

■ Drilling con collegamenti ipertestuali (fig. 7)

Se un attributo in tabella presenta un percorso di *drill* i suoi valori sono visualizzati come collegamenti ipertestuali, ovvero vengono sottolineati. Se si posiziona il puntatore del mouse sul collegamento viene visualizzato l'attributo di destinazione del *drill* che coincide con il primo presente nel menù a tendina utilizzato nella modalità di *drilling* precedentemente descritta. In breve facendo clic sui collegamenti ipertestuali si esegue un *drill* al livello immediatamente sottostante a quello di partenza presente nel percorso utilizzato e contemporaneamente una selezione basata sul valore cliccato; questo vuol dire seguire il *drilling* predefinito creato dal progettista del report.

Report		Visualizza	Dati	Formato	Ultimo aggiornamento: 03/11/2003 14.03.00				
RAGGRUPPAMENTO A PAGINE: nessuno									
Regione	Metriche Anno	Numero Iscritti							
		2001		2002		2003		Total	
		Femmina	Maschio	Femmina	Maschio	Femmina	Maschio	Femmina	Total
ABRUZZO		418	413	353	23	20			1.563
BASILICATA		129	137	149	152	6	4		577
CALABRIA		383	403	424	382	14	29		1.635
CAMPANIA		183	174	317	209	24	35		942
EMILIA-ROMAGNA		4.722	5.728	5.305	6.403	312	536		23.006
FRIULI-VENEZIA GIULIA		81	107	121	86	10	6		411
LAZIO		178	148	345	121	24	12		828
LIGURIA		58	61	79	69	4	4		275
LOMBARDIA		304	388	586	418	24	36		1.756
MARCHE		586	703	610	794	22	53		2.768
MOLISE		102	81	76	95	3	2		359
PIEMONTE		49	56	182	68	7	14		376
PUGLIA		684	659	791	641	53	53		2.881
REGIONE STRANIERA		162	182	193	175	23	24		759
SARDEGNA		169	199	207	184	14	11		784
SICILIA		434	313	503	329	32	34		1.645
TOSCANA		187	242	265	228	22	30		974
TRENTINO ALTO ADIGE		131	156	163	165	2	4		621
UMBRIA		48	55	56	60	8	8		235
VALLE D'AOSTA		6	8	14	7		1		36
VENETO		454	586	576	493	35	37		2.181
Total		9.386	10.804	11.375	11.432	662	953		44.612

Fig. 7 - Report 2: stato iniziale

Report		Visualizza	Dati	Formato	Ultimo aggiornamento: 03/11/2003 16.00.52				
RAGGRUPPAMENTO A PAGINE: nessuno									
Nazione	Metriche Anno Sesso	Numero Iscritti						Total	
		2001		2002		2003			
		Maschio	Femmina	Maschio	Femmina	Maschio	Femmina		
ITALIA		9.224	10.622	11.182	11.257	639	929	43.853	
NAZIONE STRANIERA		162	182	193	175	23	24	759	
Total		9.386	10.804	11.375	11.432	662	953	44.612	

Fig. 8 - Report 2: drill da Regione a ↑ Nazione (roll-up lungo la gerarchia Area Geografica)

■ Drill da un grafico

Se il report è formattato come grafico, è possibile eseguire il *drilling* a livelli diversi di dati facendo clic direttamente sul grafico, sulle etichette o sulle intestazioni degli assi. In tal modo viene visualizzato un nuovo grafico basato sul *drilling* predefinito creato dal progettista del report. Se si fa clic su un valore del grafico o su un'etichetta si esegue un'operazione di *drilling* con selezione; facendo clic sulle intestazioni si esegue il *drill* senza alcuna selezione.

Se non è disponibile alcun percorso di *drilling* valido per la parte di grafico selezionata, viene visualizzato nuovamente il grafico originale.

Report		Visualizza	Dati	Formato	Ultimo aggiornamento: 10/11/2003 16.35.01						
RAGGRUPPAMENTO A PAGINE: nessuno											
Nazione	Regione	Metriche Anno Sesso	Numero Iscritti						Total		
			2001		2002		2003				
			Maschio	Femmina	Maschio	Femmina	Maschio	Femmina			
ITALIA	ABRUZZO		336	418	413	353	23	20		1.569	
	BASILICATA		129	137	149	152	6	4		577	
	CALABRIA		383	403	424	382	14	29		1.635	
	CAMPANIA		183	174	317	209	24	35		942	
	EMILIA-ROMAGNA		4.722	5.728	5.305	6.403	312	536		23.006	
	FRIULI-VENEZIA GIULIA		81	107	121	86	10	6		411	
	LAZIO		178	148	345	121	24	12		828	
	LIGURIA		58	61	79	69	4	4		275	
	LOMBARDIA		304	388	586	418	24	36		1.756	
	MARCHE		586	703	610	794	22	53		2.768	
	MOLISE		102	81	76	95	3	2		359	
PIEMONTE		49	56	182	68	7	14		376		
PUGLIA		684	659	791	641	53	53		2.881		
NAZIONE STRANIERA	REGIONE STRANIERA		162	182	193	175	23	24		759	
ITALIA	SARDEGNA		169	199	207	184	14	11		784	
	SICILIA		434	313	503	329	32	34		1.645	
	TOSCANA		187	242	265	228	22	30		974	
	TRENTINO ALTO ADIGE		131	156	163	165	2	4		621	
	UMBRIA		48	55	56	60	8	8		235	
	VALLE D'AOSTA		6	8	14	7		1		36	
	VENETO		454	586	576	493	35	37		2.181	
Total			9.386	10.804	11.375	11.432	662	953		44.612	

Fig. 9 - Report 2: drill da Regione a ↑ Nazione con mantenimento del padre (livello di aggregazione dei dati inalterato)

Report		Visualizza	Dati	Formato	Ultimo aggiornamento: 05/11/2003 15.25.07					
RAGGRUPPAMENTO A PAGINE: nessuno										
Regione	Metriche Anno Sesso	Numero Iscritti						Total		
		2001		2002		2003				
		Maschio	Femmina	Maschio	Femmina	Femmina				
NON DEFINITO		1					1			
MATURITA' CLASSICA				1			1			
MATURITA' SCIENTIFICA		4		3		4	11			
MATURITA' TECNICA COMMERCIALE		1	1	4	1	1	8			
MATURITA' TECNICA PER GEOMETRI				2			2			
MATURITA' MAGISTRALE SPERIMENT.						1	1			
MAT.MAGIS.LICEO SOCIO-PSICO-PED.				2		1	3			
PERITO ELETTRONICO				1			1			
DIP. IND. SOCIO PSICOPEDAGOGICO (PROG. "BROCCA")				2			2			
PERITO INDUSTRIALE CAPOTECNICO SPECIALIZZATO				2			2			
DIRIGENTE DI COMUNITA'				1			1			
MATURITA' LINGUISTICA				1			1			
RAGIONIERE E PERITO COMMER. SPEC.COMMERCIO ESTERO				1			1			
MATURITA' SOCIO/PSICO/PEDAGOGICA SPERIM. BROCCA		1		1			1			
Total		6	8	14	7	1	36			

Fig. 10 - Report 2: drill da Regione "VALLE D'AOSTA" a ↓ Diploma (spostamento verso la gerarchia Istruzione con selezione); il livello di aggregazione è cambiato

Report		Visualizza		Dati		Formato		Ultimo aggiornamento: 05/11/2003 15.46.54			
RAGGRUPPAMENTO A PAGINE: nessuno											
Regione	Metriche Trimestre Sesso	Numero Iscritti									
		1/2001		2/2001		3/2001		4/2001		Total	
		Maschio	Femmina	Maschio	Femmina	Maschio	Femmina	Maschio	Femmina		
ABRUZZO		5	5	4	5	223	299	104	109		754
BASILICATA		3	2			91	98	35	37		266
CALABRIA		6	8	2	4	255	253	120	138		786
CAMPANIA		4	6			82	85	97	83		357
EMILIA-ROMAGNA		51	112	63	89	3.486	4.020	1.122	1.507		10.450
FRIULI-VENEZIA GIULIA		4	7			47	67	30	33		188
LAZIO		2	11	1		57	77	118	60		326
LIGURIA		1				30	33	27	28		119
LOMBARDIA		7	12	3	2	194	252	100	122		692
MARCHE		4	16	8	9	443	501	131	177		1.289
MOLISE			2			76	53	26	26		183
PIEMONTE		2	2			27	37	20	17		105
PUGLIA		14	13	4	8	421	410	245	228		1.343
REGIONE STRANIERA				1	2	135	134	26	46		344
SARDEGNA		5	5	1		97	103	66	91		368
SICILIA		11	11	3	1	244	164	176	137		747
TOSCANA		10	6			103	144	74	92		429
TRENTINO ALTO ADIGE		1	2	1	1	89	121	40	32		287
UMBRIA		1				29	28	18	27		103
VALLE D'AOSTA						4	5	2	3		14
VENETO		19	24	1	2	273	369	161	191		1.040
Total		150	244	92	123	6.406	7.253	2.738	3.184		20.190

Fig. 11 - Report 2: drill da Anno "2001" a Trimestre (drill-down lungo la gerarchia Tempo con selezione); anche con mantenimento del padre l'operazione OLAP risultante

■ Drilling dal riquadro Drill

Questa modalità è senz'altro quella più completa a disposizione dell'utente; per aprire il riquadro *Drill* selezionare dal menù a tendina di <Dati> la voce <Drill>.

Nel riquadro (fig. 12) sono elencate tutte le possibili operazioni di *drill* eseguibili nel report e l'impostazione della preferenza di mantenimento del padre. Utilizzando le altre modalità di *drill* si è obbligati ad usare l'impostazione predefinita stabilita dal progettista del report, in questo caso la preferenza è modificabile a proprio piacimento.

Aperto il riquadro di *Drill* il layout della tabella cambia: di fianco al valore di ogni attributo compare una casella di controllo. La selezione di una o più caselle permette di abbinare alle operazioni di *drill* filtri contemporanei su più attributi selezionando più valori alla volta; nelle precedenti modalità era invece possibile abbinare ai *drill* al massimo una selezione su un solo valore di un solo attributo. Da notare che la deselegione di tutti i valori di un attributo equivale alla selezione di tutti gli stessi.

Vediamo un esempio di utilizzo di questa modalità di *drilling*: selezioniamo la caselle di controllo di "2001", "EMILIA-ROMAGNA" e "MARCHE"; deselegioniamo la preferenza di mantenimento del padre; selezioniamo dal menù a tendina che parte da **Anno** la voce **Trimestre** e facciamo clic sull'icona  posizionata di fianco.

Il risultato ottenuto (fig. 13) consiste nell'aver aggregato i dati per **Trimestre** visualizzando solo i trimestri e le regioni di interesse; ci si è spostati all'interno della gerarchia Tempo diminuendo il livello di aggregazione e si è ristretto, applicando diversi criteri, l'insieme dei dati oggetto dell'analisi: si è quindi effettuato un *drill-down* con selezione multipla.

Drill

Da: **Regione Residenza** A:
 Nazione Diploma

Da: **Anno** A:
 Trimestre

Mantieni il padre durante il drilling

Regione	Metriche Anno Sesso	Numero Iscritti						Total
		2001		2002		2003		
		Maschio	Femmina	Maschio	Femmina	Maschio	Femmina	
<input type="checkbox"/> ABRUZZO		336	418	413	353	23	20	1.563
<input type="checkbox"/> BASILICATA		129	137	149	152	6	4	577
<input type="checkbox"/> CALABRIA		383	403	424	382	14	29	1.635
<input type="checkbox"/> CAMPANIA		183	174	317	209	24	35	942
<input type="checkbox"/> EMILIA-ROMAGNA		4.722	5.728	5.305	6.403	312	536	23.006
<input type="checkbox"/> FRIULI-VENEZIA GIULIA		81	107	121	86	10	6	411
<input type="checkbox"/> LAZIO		178	148	345	121	24	12	828
<input type="checkbox"/> LIGURIA		58	61	79	69	4	4	275
<input type="checkbox"/> LOMBARDIA		304	388	586	418	24	36	1.756
<input type="checkbox"/> MARCHE		586	703	610	794	22	53	2.768
<input type="checkbox"/> MOLISE		102	81	76	95	3	2	359
<input type="checkbox"/> PIEMONTE		49	56	182	68	7	14	376
<input type="checkbox"/> PUGLIA		684	659	791	641	53	53	2.881
<input type="checkbox"/> REGIONE STRANIERA		162	182	193	175	23	24	759
<input type="checkbox"/> SARDEGNA		169	199	207	184	14	11	784
<input type="checkbox"/> SICILIA		434	313	503	329	32	34	1.645
<input type="checkbox"/> TOSCANA		187	242	265	228	22	30	974
<input type="checkbox"/> TRENTO ALTO ADIGE		131	156	163	165	2	4	621
<input type="checkbox"/> UMBRIA		48	55	56	60	8	8	235
<input type="checkbox"/> VALLE D'AOSTA		6	8	14	7		1	36
<input type="checkbox"/> VENETO		454	586	576	493	35	37	2.181
Total		9.386	10.804	11.375	11.432	662	953	44.612

Fig. 12 - Report 2: riquadro Drill dello stato iniziale

Report Visualizza Dati Formato Ultimo aggiornamento: 05/11/2003 16.11.29

RAGGRUPPAMENTO A PAGINE: nessuno

Regione	Metriche Trimestre Sesso	Numero Iscritti								Total
		1/2001		2/2001		3/2001		4/2001		
		Maschio	Femmina	Maschio	Femmina	Maschio	Femmina	Maschio	Femmina	
EMILIA-ROMAGNA		51	112	63	89	3.486	4.020	1.122	1.507	10.450
MARCHE		4	16	8	9	443	501	131	177	1.289
Total		55	128	71	98	3.929	4.521	1.253	1.684	11.739

Fig. 13 - Report 2: drill da Anno a Trimestre con selezione multipla (drill-down lungo la gerarchia Tempo + selezioni Anno=[2001] e Regione=[EMILIA-ROMAGNA,MARCHE])