

Università degli Studi di Modena e Reggio Emilia
Facoltà d'Ingegneria "Enzo Ferrari"

Corso di Laurea in Ingegneria Informatica (DM 270/04)

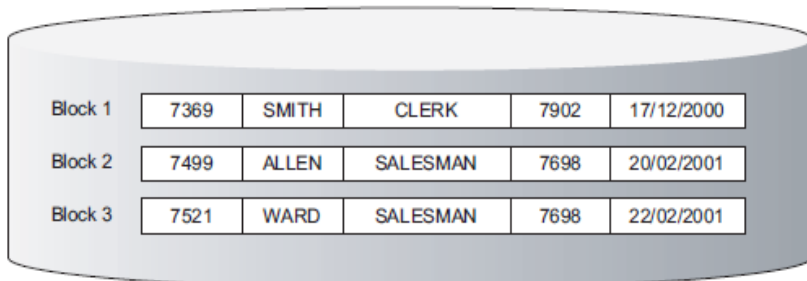
Confronto di prestazioni fra DBMS Relazionali e Verticali

Tutor:
Chiar.mo Prof Sonia Bergamaschi

Candidato:
Gianluca Farinelli

Anno Accademico 2012/2013

Row Oriented VS Column Oriented



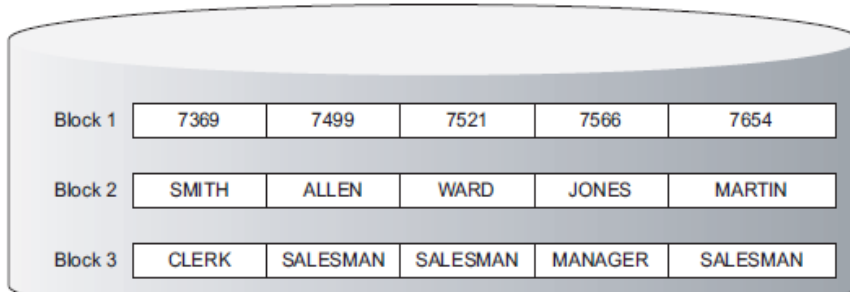
Row Database stores row values together

| EmpNo | EName | Job | Mgr | HireDate |
|-------|--------|----------|------|------------|
| 7369 | SMITH | CLERK | 7902 | 17/12/1980 |
| 7499 | ALLEN | SALESMAN | 7698 | 20/02/1981 |
| 7521 | WARD | SALESMAN | 7698 | 22/02/1981 |
| 7566 | JONES | MANAGER | 7839 | 2/04/1981 |
| 7654 | MARTIN | SALESMAN | 7698 | 28/09/1981 |
| 7698 | BLAKE | MANAGER | 7839 | 1/05/1981 |
| 7782 | CLARK | MANAGER | 7839 | 9/06/1981 |

Row Store Physical Layout

Logical Schema

Column Store physical layout



Column Database stores column values together

- Row Oriented salva i dati come sezione di righe
- Column Oriented salva i dati come sezione di colonne

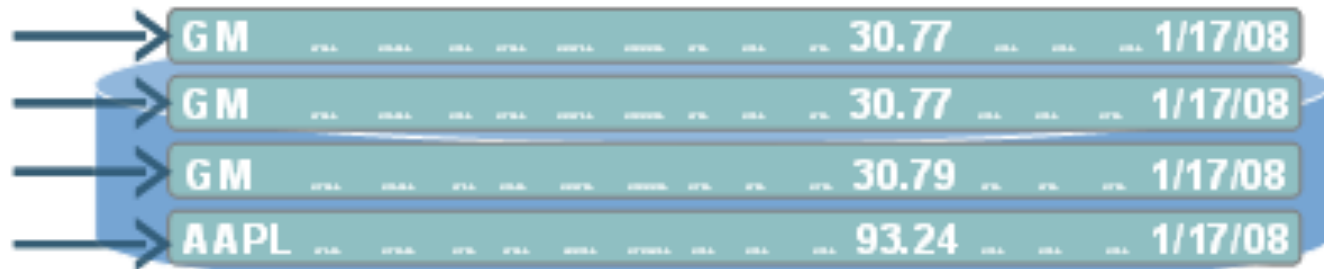
Column Store

Reads 3 columns



Row Store

Reads all columns



- Vantaggi nella lettura dei dati: evita di leggere attributi non utilizzati
- Svantaggio nell'inserimento e modifica dei dati

**Obiettivo: valutare sperimentalmente
le prestazioni dei due DBMS**

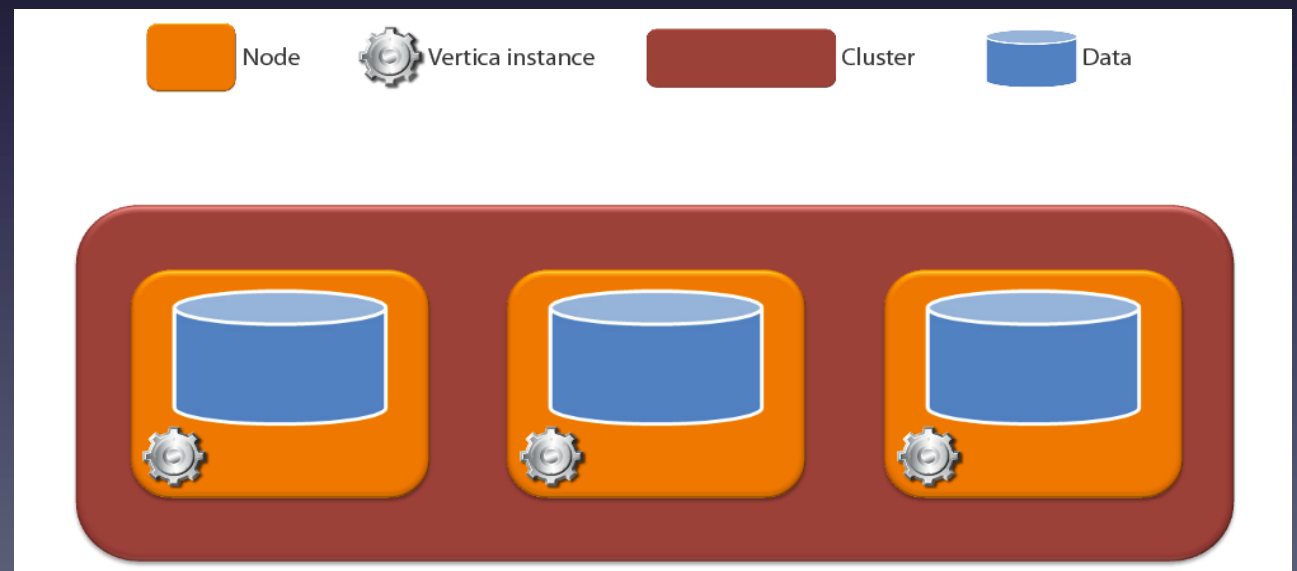
Column Oriented DBMS: VERTICA

Fondata nel 2005 da Stonebraker dal 2011
ad HP

Sviluppa DBMS e software analitico

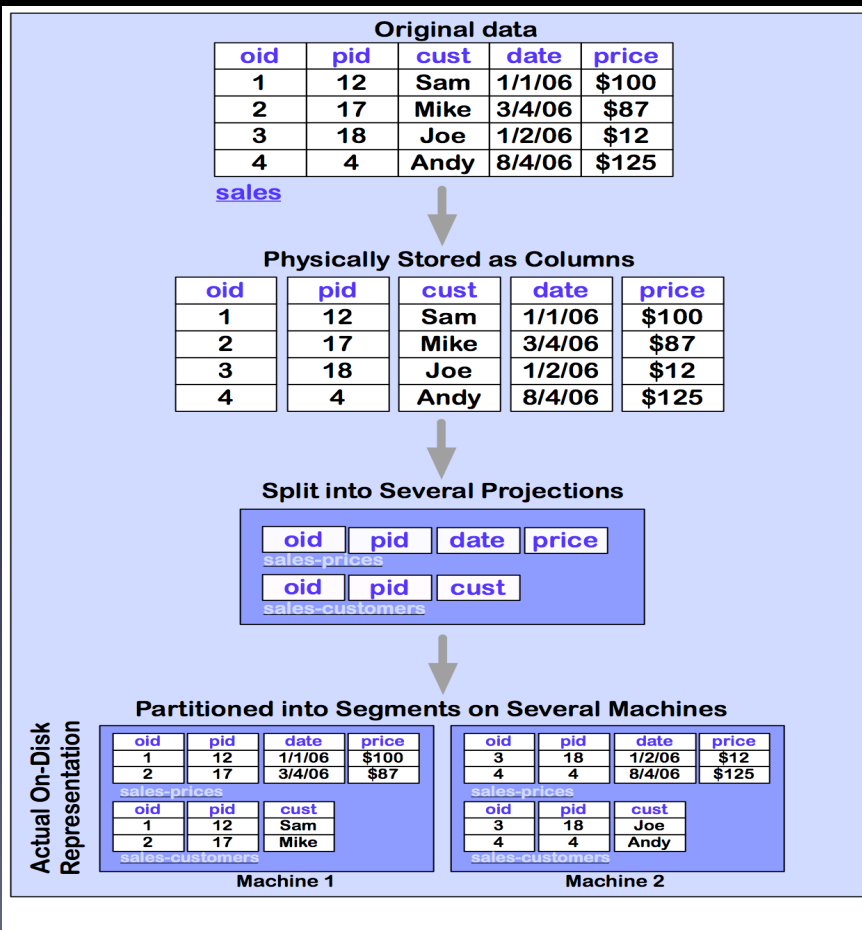
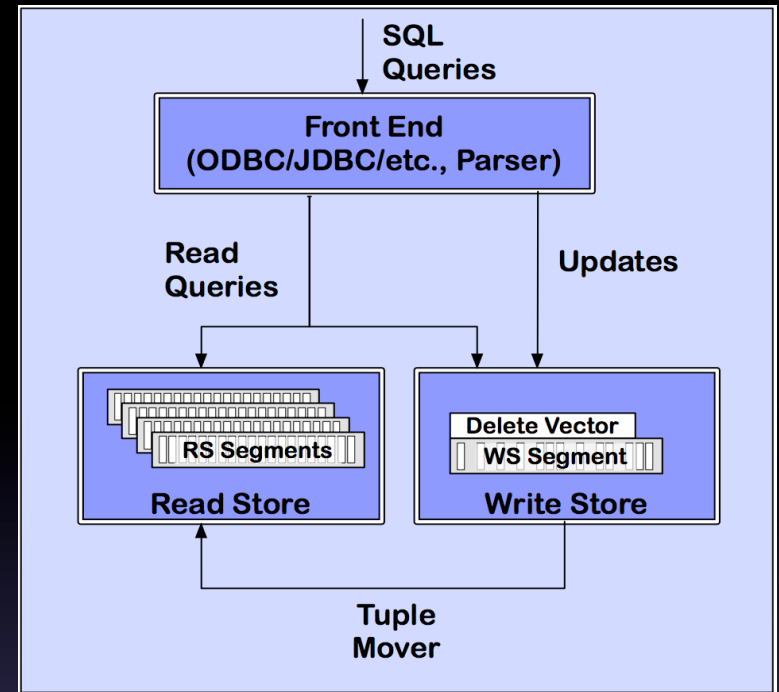


Disegnato per
distribuire
memoria fisica e
per sfruttare il
parallelismo delle
query



Diviso in due differenti strutture di memoria:

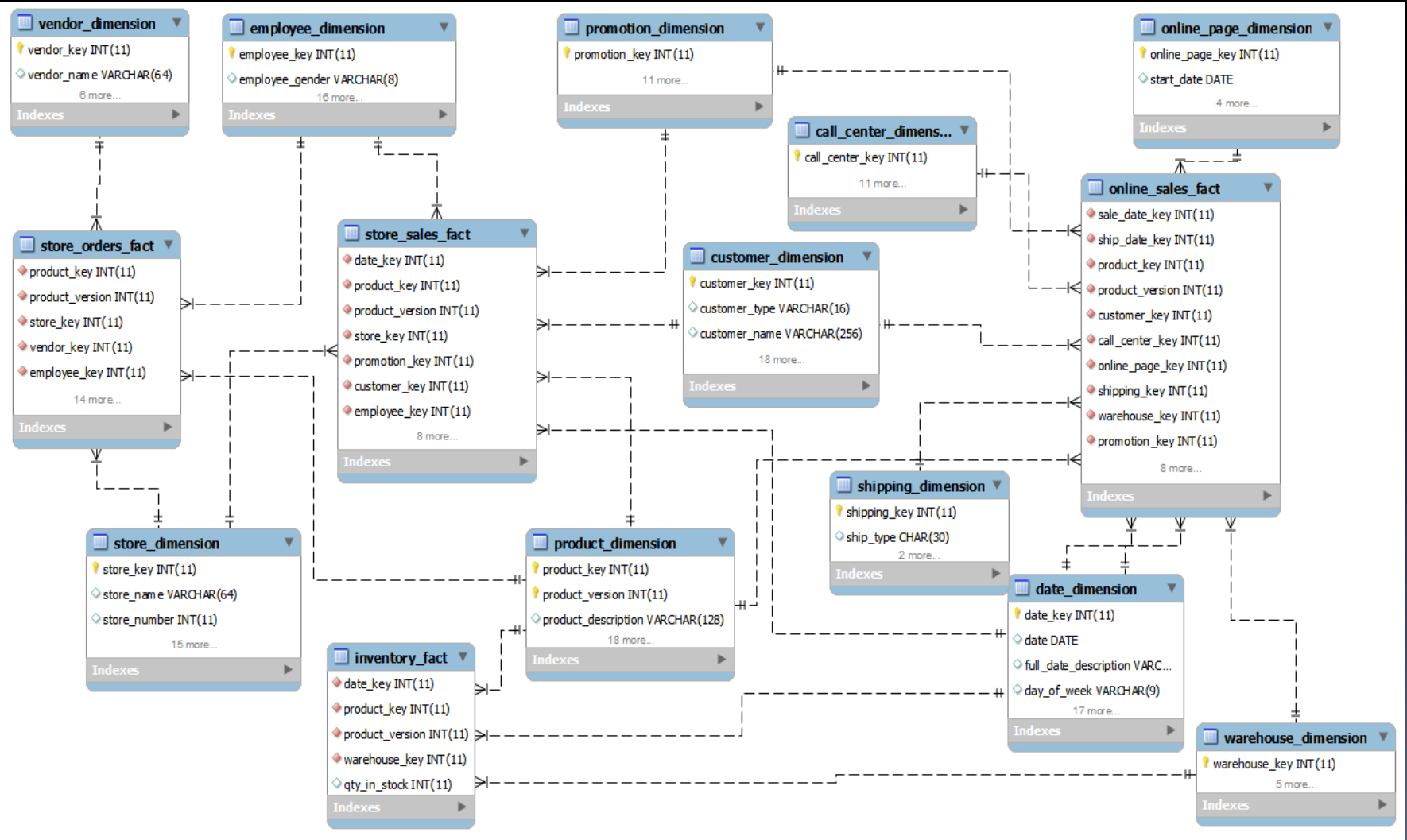
- Write-Optimized-Storage per un rapido inserimento dati
- Read-Optimized-Storage contiene tutti dati del database



Tuple Mover:

- MoveOut copia i dati dal WOS al ROS
- MergeOut riorganizza i dati nel ROS

DATA SET di Valutazione



DBMS a confronto

Row Oriented

- mySQL
- SO: Windows 7 x64
- RAM: 8GB
- HDD: 80GB
- CPU: 1 core

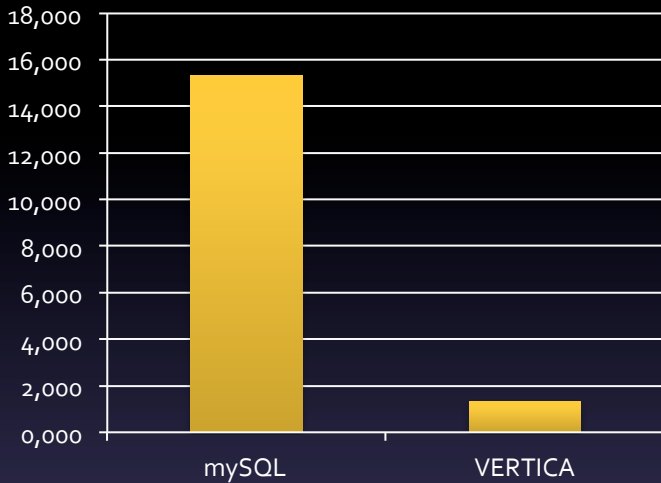


Column Oriented

- VERTICA
- SO: CentOS x64
- RAM: 8GB
- HDD: 80GB
- CPU: 1 core



Confronto di prestazioni tra un DBMS Verticale e un RDBMS



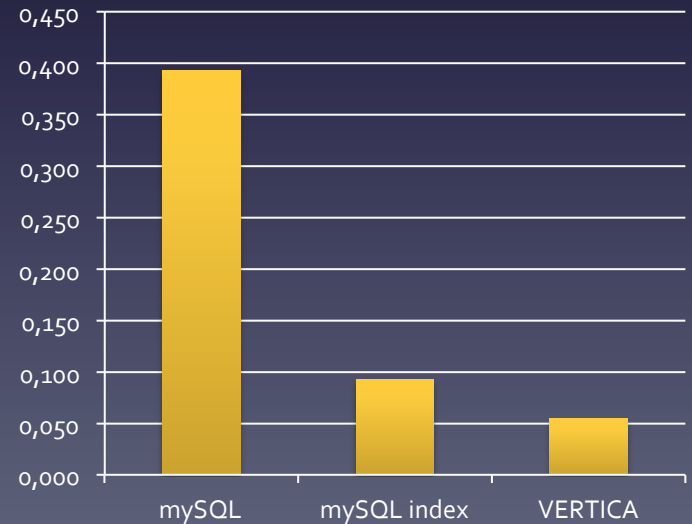
3. Selezione dei clienti col maggior incasso annuo per genere

```
SELECT customer_name, annual_income
FROM public.customer_dimension
WHERE (customer_gender, annual_income) IN
      (SELECT customer_gender, MAX(annual_income)
       FROM public.customer_dimension
       GROUP BY customer_gender);
```

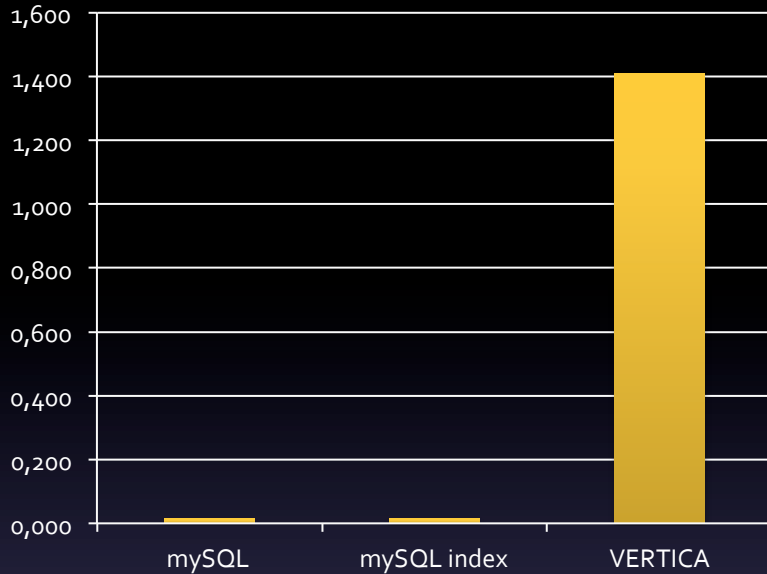
Indici su customer_gender, annual_income

4. Selezione delle vendite in uno stato

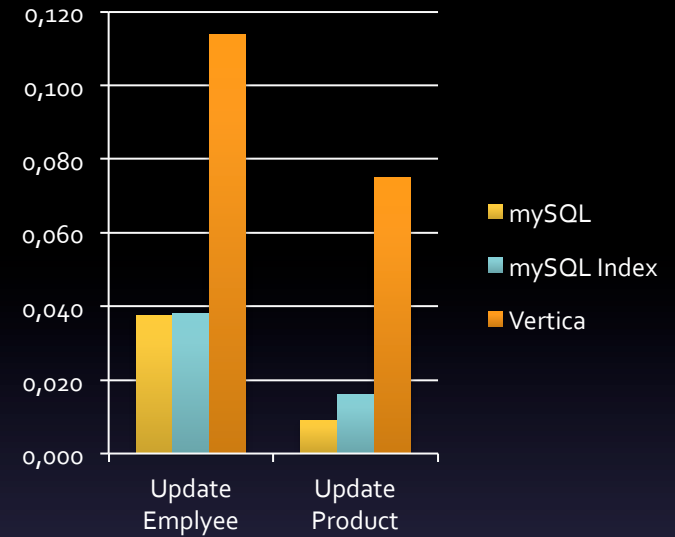
```
SELECT DISTINCT s.product_key, p.product_description
FROM store.store_sales_fact s, public.product_dimension p
WHERE s.product_key = p.product_key
AND s.product_version = p.product_version
AND s.store_key IN (
  SELECT store_key
  FROM store.store_dimension
  WHERE store_state = 'MA')
ORDER BY s.product_key;
```



Selezione di molti attributi appartenenti allo stesso record

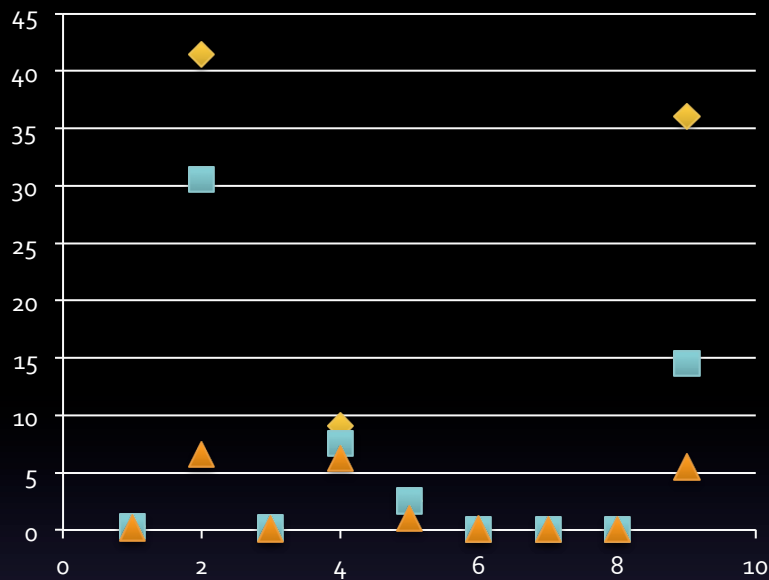


Aggiornamento di valori



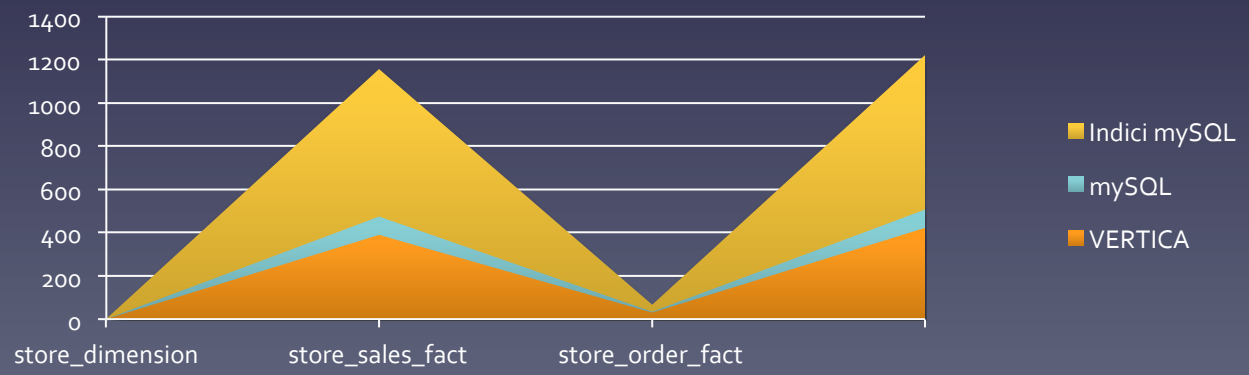
Inserimento di nuovi record nelle tabelle





Schema Public
 1 date_dimension
 2 product_dimension
 3 promotion_dimension
 4 customer_dimension
 5 employee_dimension
 6 shipping_dimension
 7 vendor_dimension
 8 warehouse_dimension
 9 inventory_fact

Gli algoritmi di compressione di VERTICA permettono di risparmiare il 25% di spazio in più su disco e fino al 71% nel caso siano presenti indici



Conclusioni

- I DBMS verticali sono avvantaggiati nella selezione di dati da poche colonne appartenenti a tabelle diverse e, grazie a specifici algoritmi di codifica, riescono ad ottimizzare lo spazio occupato dai dati e a limitare i cicli di I/O per operare su questi ultimi
- I DBMS relazionali sono avvantaggiati nelle operazioni di modifica dei dati e, se indicizzati correttamente, possono competere coi DBMS verticali ma necessitano di hardware più costoso per garantire prestazioni equivalenti per interrogazioni.

GRAZIE PER L'ATTENZIONE