

# SQL03 - Columnstore Databases and the SQL Server Platform

**Marco Russo**  
DevLeap

[marco@devleap.com](mailto:marco@devleap.com)

Twitter: @marcorus

[http://sqlblog.com/blogs/marco\\_russo](http://sqlblog.com/blogs/marco_russo)

<http://blogs.devleap.com/marco>

#CDays13 – 27 e 28 febbraio 2013

# Grazie a

Sponsor



Technical  
Conferences



COMMUNITY DAYS 

# Agenda

- Database a colonne
- xVelocity
- PowerPivot
- Analysis Services Tabular
- Columnstore Index in SQL Server

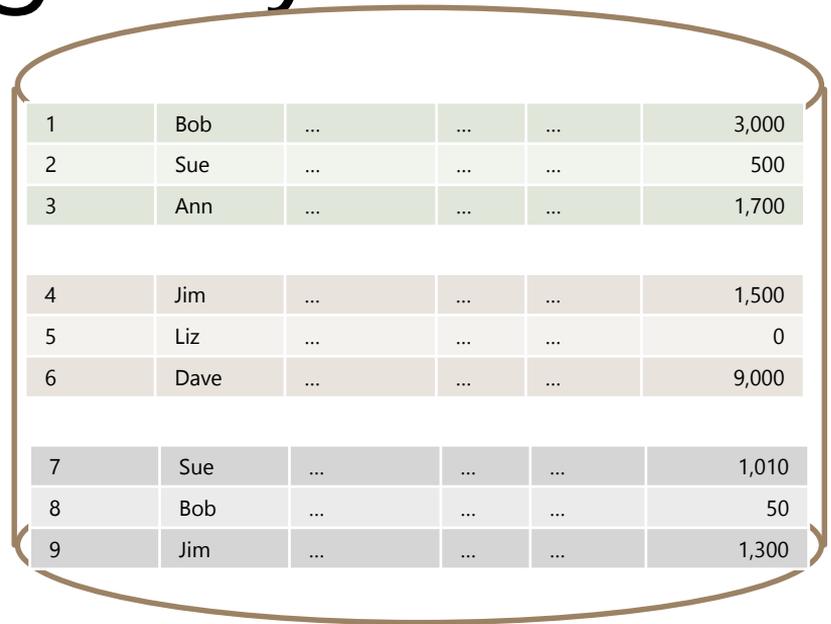
Perché memorizzare i dati per colonna anziché per riga?

# **DATABASE A COLONNE**

# Row Storage Layout

Customers Table

ID	Name	Address	City	State	Bal Due
1	Bob	...	...	...	3,000
2	Sue	...	...	...	500
3	Ann	...	...	...	1,700
4	Jim	...	...	...	1,500
5	Liz	...	...	...	0
6	Dave	...	...	...	9,000
7	Sue	...	...	...	1,010
8	Bob	...	...	...	50
9	Jim	...	...	...	1,300

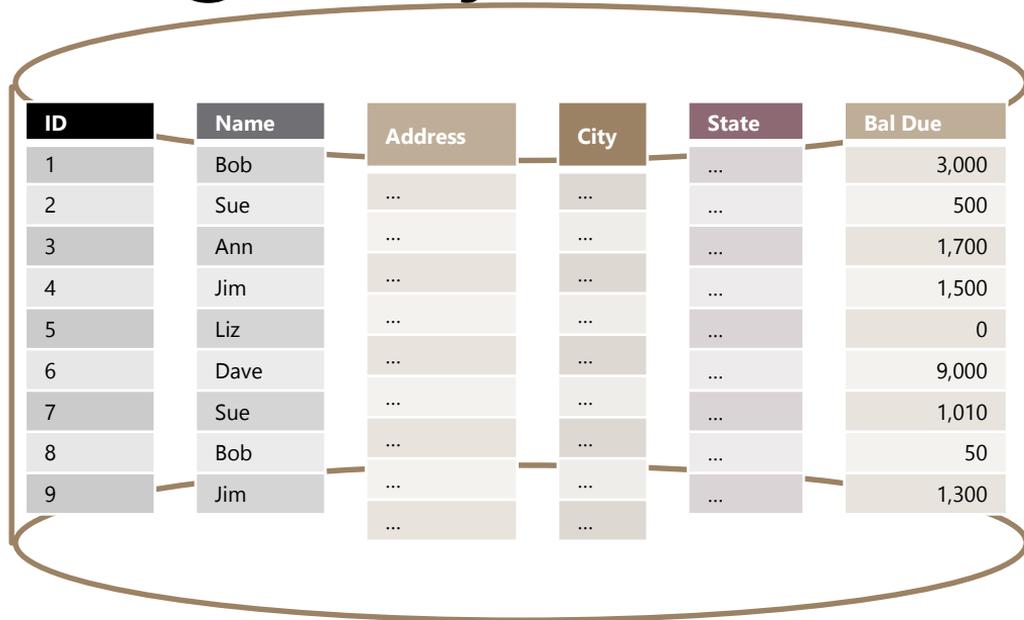


Niente di speciale, in genere i database relazionali lavorano così.  
Tecnicamente è un "row store"

# Column Storage Layout

Customers Table

ID	Name	Address	City	State	Bal Due
1	Bob	...	...	...	3,000
2	Sue	...	...	...	500
3	Ann	...	...	...	1,700
4	Jim	...	...	...	1,500
5	Liz	...	...	...	0
6	Dave	...	...	...	9,000
7	Sue	...	...	...	1,010
8	Bob	...	...	...	50
9	Jim	...	...	...	1,300



Le tabelle sono memorizzate "per colonna", con tutti i valori di ciascuna colonna memorizzati in un solo blocco.

# Column vs Row Storage

- Column Storage
  - Accesso veloce e una sola Colonna
  - Tempo richiesto per materializzare righe
  - Spende più CPU che I/O
- Row Storage
  - Accesso veloce a una singola riga
  - Nessuna materializzazione necessaria
  - Spende più I/O che CPU

La famiglia di tecnologie xVelocity in SQL Server 2012

**XVELOCITY**

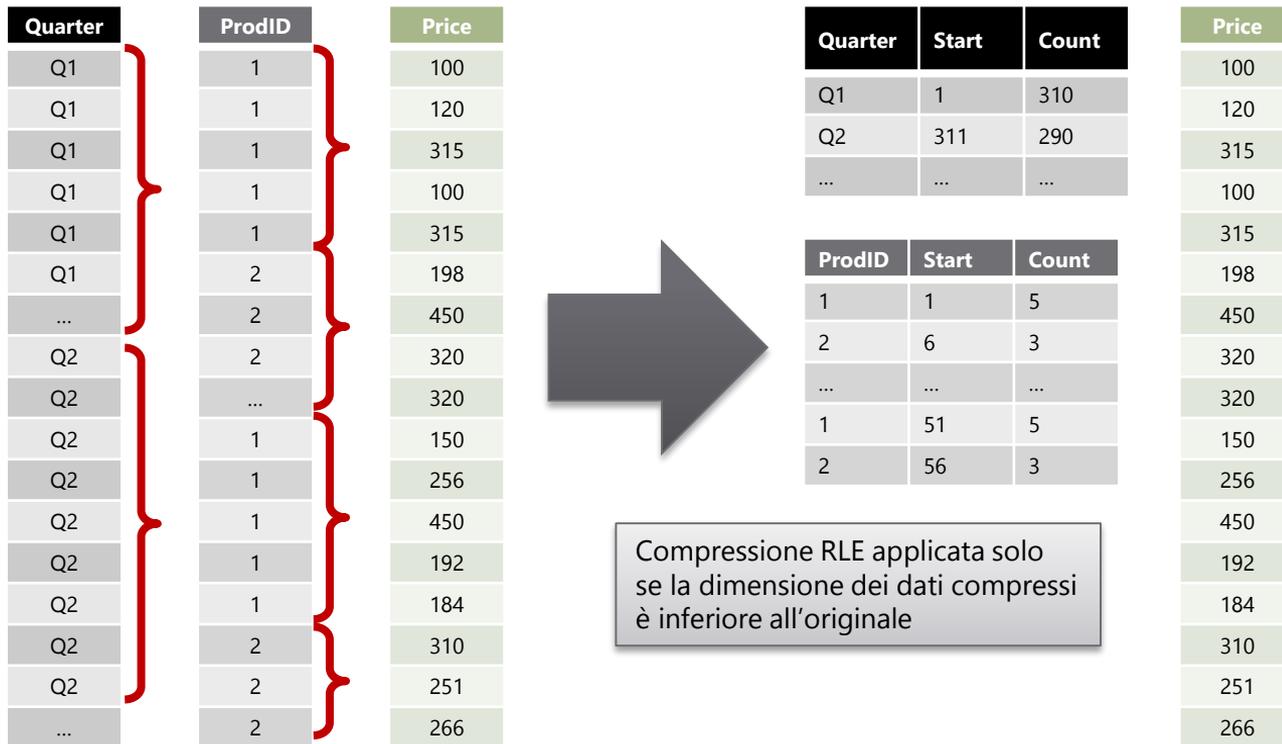
# xVelocity

- VertiPaq
  - xVelocity in-memory analytics engine
    - PowerPivot for Excel
    - Analysis Services Tabular Model
- Columnstore Index
  - xVelocity memory optimized columnstore index
    - SQL Server 2012 (solo licenza Enterprise)

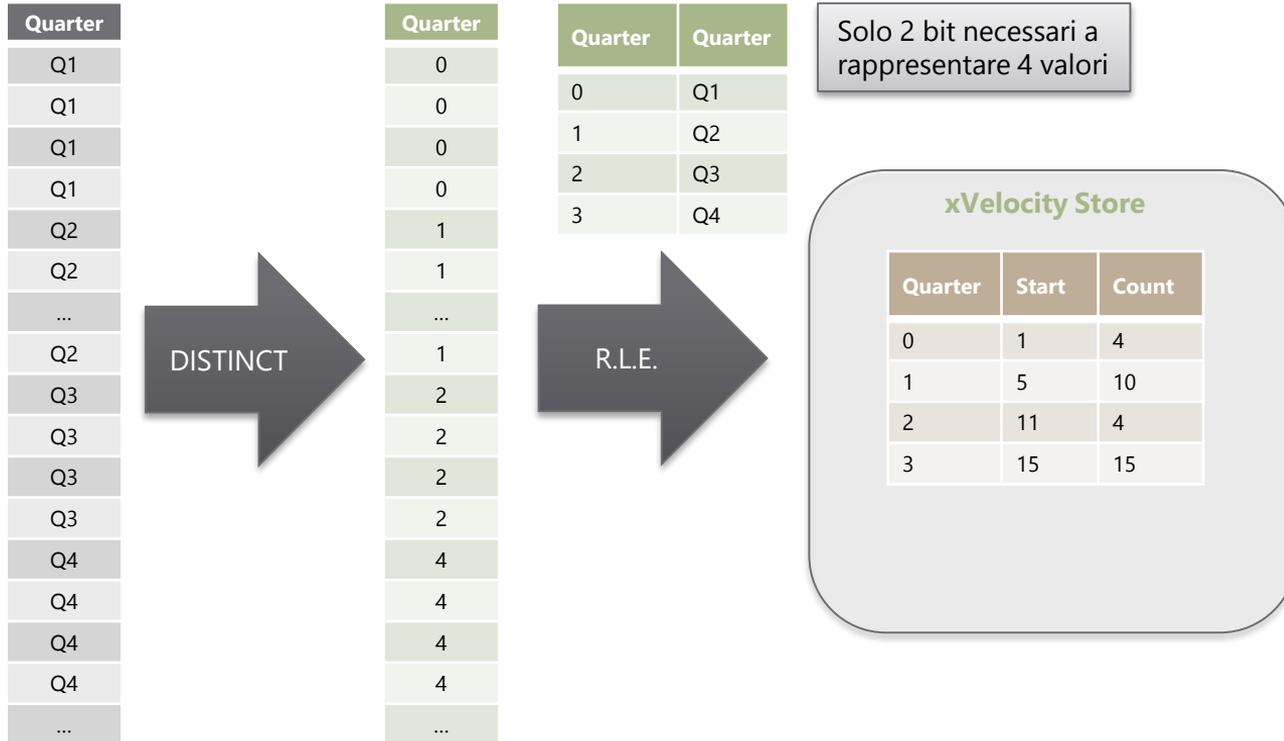
# xVelocity Internals

- Famiglia di tecnologie con base comune
  - Nasce come Storage Engine di PowerPivot
  - Scritta dal team di Analysis Services
  - Portata su SQL Server
    - Integrata con motore relazionale
    - Principi di compressione comuni

# Run Length Encoding (RLE)



# Dictionary Encoding



# xVelocity in-memory Compression

- Dictionary Encoding
  - Solo quando serve
    - Ampii range di interi non usano un dizionario
  - Dimensione tabelle non dipende dal tipo di dati
- RLE Encoding
  - Solo se dimensione dati compressi inferior a originale
  - Efficienza compressione dipende dall'ordinamento dei dati
  - Algoritmo per riordinamento automatico (non sempre efficace)

# Accesso ai dati

- Accesso basato su scansioni complete
  - Veloce assumendo che i dati siano in memoria
- Non ci sono indici
- Diverse implementazioni
  - SQL Server usa anche partizionamento

# Ottimizzare xVelocity

- Compressione tramite:
  - Column Store
  - Dictionary Encoding
  - RLE Encoding
- Meno RAM richiesta
- Scansione colonne più veloci
- 10x è un tipico rapporto di compressione
  - A volte si raggiunge 100x !

La prima implementazione di xVelocity

# POWERPIVOT

# PowerPivot

- Add-in gratuito per Excel 2010
- Integrato in Excel 2013
  - Data Model in tutte le versioni
  - Add-in solo in alcune versioni
    - Office Professional Plus
    - Excel 2013 standalone (volume licensing)
    - Office 365 ProPlus

[http://sqlblog.com/blogs/marco\\_russo/archive/2013/02/18/the-right-version-of-excel-2013-for-using-powerpivot-powerpivot-excel.aspx](http://sqlblog.com/blogs/marco_russo/archive/2013/02/18/the-right-version-of-excel-2013-for-using-powerpivot-powerpivot-excel.aspx)

PowerPivot for Excel

**DEMO**

# Quando usare PowerPivot

- Analisi di dati a livello personale
  - Anche log di siti, performance counter, ecc.
- Self-service BI (utenti avanzati di Excel)
  - Pubblicazione su SharePoint
- Difficile da integrare con VSTO
  - Ok refresh modelli dati
  - Limitazioni per creare nuovi modelli dati

Il motore di PowerPivot usato sul server

# **ANALYSIS SERVICES TABULAR**

# Analysis Services Tabular

- Con Microsoft SQL Server 2012
  - Analysis Services Multidimensional
    - Istanza che funziona come versioni precedenti
  - Analysis Services Tabular
    - Istanza che funziona con motore xVelocity
    - Usa stesso modello dati di PowerPivot

# Vantaggi rispetto a PowerPivot

- Funzionalità aggiuntive
  - Security row-level basata sui ruoli
  - Partizionamento dei dati nelle tabelle
- Maggior controllo su Process
- Si può usare tutta la RAM del server
  - SharePoint limita dimensione file a 2GB

Microsoft SQL Server Analysis Services 2012 - Tabular Model

**DEMO**

# Quando usare Tabular

- Indipendenza da SharePoint
- Controllo su Process
  - Refresh anche ogni minuto – quasi real-time
- Gestione centralizzata modello e security

La tecnologia xVelocity integrata in SQL Server 2012

# **COLUMNSTORE INDEX**

# Columnstore Index

- Indice nonclustered
  - Paginazione come per altri dati SQL Server
- Può agire su una selezione di colonne
  - Limitazioni sui tipi supportati
- Può essere usato dal query optimizer
  - Verificare in query plan

Microsoft SQL Server 2012 – Columnstore Index

**DEMO**

# Limitazioni Columnstore Index

- Tabella diventa read-only
  - Aggiornabile con switch-in partizioni
- Non tutte le query sfruttano l'indice
  - Adattare query SQL perché sfruttino indice
  - Migliorerà in futuro, dettagli in whitepaper

# Quando usare Columnstore Index

- Accesso a dati in sola lettura
- Numero troppo alto di indici «tradizionali»
- Non va bene per query real-time
  - Dividere dati consolidati da attuali in tabelle diverse e usare viste
  - Attenzione a limitazione query per uso indici

# Conclusioni

- xVelocity è una tecnologia da conoscere
- Utile già oggi in molti scenari
- Evolverà nel tempo
  - In SQL Server sarà possibile aggiornare le tabelle, con qualche compromesso prestazionale

# Link utili

- Columnstore Indexes

<http://msdn.microsoft.com/en-us/library/gg492088.aspx>

- VertiPaq vs ColumnStore Comparison

<http://www.sqlbi.com/articles/vertipaq-vs-columnstore/>

# Q&A

Tutto il materiale di questa sessione su  
<http://www.communitydays.it/>

#CDays13