

Índices Columnstore en SQL Server

Gerardo Rossel

2025

¿Por qué Columnstore?

- Diseñado para escenarios de **Data Warehouse** y grandes volúmenes de datos.
- Creciente cantidad de registros → compresión tradicional (row/page) no es suficiente.
- Procesadores modernos con muchos núcleos → se busca procesamiento paralelo eficiente.
- Columnstore ofrece:
 - Compresión muy alta.
 - Lectura de sólo las columnas necesarias.
 - Procesamiento por lotes (Batch Mode).

Rowstore vs Columnstore

Rowstore

- Datos guardados **fila por fila**.
- Fácil inserción y actualización.
- Al leer, se arrastran columnas que tal vez no se necesitan.

1	Jody	Philips	43.03
2	Mark	Johnson	37.08
3	Matt	Markensen	16.81
4	Gail	Lindberg	24.90

Data is stored on the disk tuple by tuple

Columnstore

- Datos guardados **columna por columna**.
- Permite leer solo las columnas requeridas.
- Alta compresión (valores repetidos se reducen).

1	Jody	Philips	43.03
2	Mark	Johnson	37.08
3	Matt	Markensen	16.81
4	Gail	Lindberg	24.90

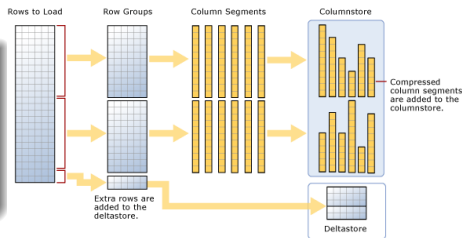
Data is stored on the disk column by column

Cómo se construye un Columnstore Index

- 1 **Data Shuffling:** reorganización de filas.
- 2 **Row Groups:** bloques de hasta 1 millón de filas.
- 3 **Segmentos:** cada columna se divide en segmentos.
- 4 **Compresión:** algoritmos aplicados sobre cada segmento.

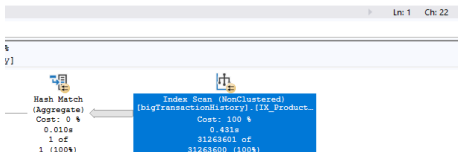
Segmentos

- Cada segmento \approx 1 millón de filas.
- Unidad lógica de operaciones de E/S.
- Permite "segment elimination"(ignorar bloques que no cumplen filtros).



Batch Mode Processing

- Se introduce en 2012 para trabajar con columnstore.
- En Row Mode: la ejecución es operador por operador, fila a fila.
- En Batch Mode: procesamiento vectorizado.
 - Se procesan entre 64 y 912 filas en un bloque.
 - Ganancias de rendimiento de hasta 10x o más.
- Exclusivo de Columnstore Indexes.
- Pensarlo como: **trabajo en fábrica (lotes)** vs **trabajo manual (fila a fila)**.
- A partir de SQL Server 2019 el procesamiento Batch Mode puede mejorar el rendimiento de las consultas analíticas en tablas que no tienen un índice columnstore, pero solo mejoraría el uso de CPU no el de E/S.



Captura de pantalla de la pestaña 'Misc' en SQL Server Enterprise Manager, mostrando estadísticas de ejecución en modo Batch.

Actual Execution Mode Batch	
Actual I/O Statistics	
Actual Number of Batches	34741
Actual Number of Rows for 31263601	
Actual Number of Rows Re: 31263601	
Actual Rebinds	0
Actual Rewinds	0
Actual Time Statistics	
Defined Values	
Description	Scan a nonclustered index, entire
Estimated CPU Cost	4.29876

- **SQL Server 2012:** Nonclustered, solo lectura.
- **SQL Server 2014:** Clustered (tablas completas), todavía limitaciones de actualización.
- **SQL Server 2016:** Columnstore actualizable, soporte en OLTP y DWH.
- **SQL Server 2017+:** mejoras de compatibilidad (replicación, índices filtrados, online rebuilds).