

Universidad Nacional del Nordeste



Facultad de Ciencias Exactas y Naturales y Agrimensura

Carrera: Licenciatura en Sistemas de Información

Cátedra: Base de Datos I

Tema: Optimización de consultas a través de índices

Año: 2024

Integrantes:

Almirón, Pedro Augusto. <u>DNI:</u> 42.791.957 <u>LU:</u> 54251

Indice

•	RESUMEN		
•	OBJETIV	O PRINCIPAL3	
	•	ENTENDER CONCEPTUALMENTE EL SIGNIFICADO 3	
	•	ENTENDER LOS DIFERENTES TIPOS DE ÍNDICES3	
	•	ENTENDER LA DIFERENCIA EN EL ACCESO A LOS DATOS A TRAVÉS DE ÍNDICES Y SIN ELLOS3	
•	CAPÍTUI	LO I: INTRODUCCIÓN	
	•	IMPORTANCIA DE LA OPTIMIZACIÓN DE BÚSQUEDAS EN BASES DE DATOS3	
	•	PAPEL CRUCIAL DE LOS ÍNDICES EN LA OPTIMIZACIÓN DEL RENDIMIENTO DE CONSULTAS3	
•	CAPÍTUI	LO II: MARCO CONCEPTUAL O REFERENCIAL4	
	•	INTRODUCCIÓN A LOS ÍNDICES EN LOS MOTORES DE BASES DE DATOS4	
	•	FUNDAMENTOS DE LOS ÍNDICES4	
	•	TIPOS DE ÍNDICES EN SQL SERVER4	
	•	ÍNDICES CLUSTERED Y NON-CLUSTERED5	
	•	DIFERENCIAS ENTRE ÍNDICES CLUSTERED Y NON-CLUSTERED5	
	•	ESTRATEGIAS DE USO Y CREACIÓN DE ÍNDICES5	
•	CAPÍTUI	LO III: METODOLOGÍA SEGUIDA6	
	•	DESCRIPCIÓN DE CÓMO SE REALIZÓ EL TRABAJO PRÁCTICO6	
	•	HERRAMIENTAS (INSTRUMENTOS Y PROCEDIMIENTOS)6	
	•	INSERCIÓN MASIVA DE DATOS7	
	•	GENERACIÓN DE CONSULTAS Y CREACIÓN DE ÍNDICES8	
•	CAPÍTUI	LO IV: DESARROLLO DEL TEMA/RESULTADOS12	
	•	RESULTADOS DE LAS PRUEBAS DE RENDIMIENTO CON DIFERENTES ÍNDICES12	
	•	COMPARACIÓN DE TIEMPOS DE RESPUESTA ENTRE CONSULTAS CON Y SIN ÍNDICES.	
	•	IMPACTO DE LOS ÍNDICES EN EL RENDIMIENTO DEL SISTEMA12	
•	CAPÍTUI	LO V: CONCLUSIÓN13	
	•	RESUMEN DE LOS HALLAZGOS Y CONCLUSIONES DE LAS PRUEBAS13	
	•	IMPORTANCIA DE LA IMPLEMENTACIÓN ADECUADA DE ÍNDICES EN LA OPTIMIZACIÓN DE BASES DE DATOS13	
•	CAPÍTUI	LO VI: BIBLIOGRAFÍA13	
	•	FUENTES CONSULTADAS PARA LA INVESTIGACIÓN Y DESARROLLO DEL TRABAJO.	

Resumen:

Los índices son una herramienta clave en el mundo del rendimiento de las bases de datos en general, permite hacer de forma eficiente consultas que a grande rasgos, si una base de datos no tiene una buena gestión de índices podría significar una perdida enorme de rendimiento y en algunos casos causar que la aplicación o sistema ni siguiera funcione.

Objetivo Principal:

- Entender conceptualmente el significado.
- Entender los diferentes tipos de índices.
- Entender la diferencia en el acceso a los datos a través de índices y sin ellos.

CAPÍTULO I: INTRODUCCIÓN

En el vasto mundo de las bases de datos, la optimización de búsquedas es un tema fundamental que influye directamente en el rendimiento y la eficiencia de los sistemas de gestión de datos. La capacidad de recuperar información de manera rápida y precisa es esencial en cualquier aplicación o sistema que maneje grandes volúmenes de datos. Los índices juegan un papel crucial en este proceso, permitiendo acelerar las consultas al proporcionar acceso directo a los registros relevantes. Existen varios tipos de índices, en este trabajo en específico se tendrá especial énfasis en los tipos de índices **Clúster** y **No Clustered**

- a. **Tema.** El tema de este trabajo se centra en la optimización de búsqueda a través de índices. Este enfoque apunta a realizar consultas a determinadas tablas de manera eficiente, es decir requiriendo el menor costo posible del cpu (Unidad central del procesamiento)
 - b. Definición o planteamiento del problema. En este trabajo integrador exploraremos en detalle el concepto de optimización de búsquedas a través de índices en bases de datos. Analizaremos cómo los índices optimizan las consultas al mejorar la velocidad de acceso a los datos y reducir la carga en los sistemas. Además, examinaremos los diferentes tipos de índices disponibles en los sistemas de gestión de bases de datos, así como las estrategias y técnicas utilizadas para diseñar, implementar y mantener índices eficientes.
 - c. Objetivo del Trabajo Práctico. Proporcionar una comprensión integral de los índices de búsqueda, su definición, sus tipos y sus

aplicaciones para poder entender la importancia de trabajar con ellos. Para ello planteamos los siguientes objetivos:

i. Objetivos Generales:

• Comprender a fondo la administración de índices de búsqueda en bases de datos y su impacto en el rendimiento de los datos.

ii. Objetivos Específicos:

- Identificar los conceptos básicos relacionados con los índices, tipos, diferencias en una base de datos.
- Investigar cómo la gestión de índices contribuye al rendimiento y eficacia de las consultas en la base de datos.
- Analizar la diferencia entre los índices clúster y no clustered y la diferencia de trabajar con y sin ellos.

CAPITULO II: MARCO CONCEPTUAL o REFERENCIAL

Introducción a los índices en los motores de bases de datos

Los índices juegan un papel fundamental en la optimización del rendimiento de las consultas en SQL Server. Proporcionan una forma eficiente de acceder a los datos almacenados en tablas, permitiendo una recuperación rápida y efectiva de la información. En este trabajo, exploraremos en detalle el concepto de índices en SQL Server, su funcionamiento, tipos y estrategias de uso para mejorar el rendimiento de las consultas.

Fundamentos de los índices

Los índices se basan en estructuras de datos optimizadas, como árboles Binarios, que permiten una búsqueda rápida y eficiente de datos en función de los valores de las columnas indexadas. Al acceder a una tabla mediante un índice, el motor de base de datos puede evitar la necesidad de explorar todas las filas de la tabla en busca de los datos requeridos, lo que resulta en tiempos de respuesta más rápidos y un rendimiento general mejorado de las consultas.

Tipos de índices en Sql Server:

Los índices en SQL Server son estructuras que permiten una recuperación rápida de datos en una tabla, lo que mejora significativamente el rendimiento de las consultas. Existen varios tipos de índices, pero los más comunes son los índices clustered y non-clustered.

Índices Clustered:

Los índices clustered reorganizan físicamente los datos en la tabla según el orden de las claves de índice.

Cada tabla puede tener solo un índice clustered, ya que determina el orden físico de los datos.

Al ser el orden físico de los datos, el índice clustered también almacena los datos de la tabla, lo que significa que la clave del índice es la clave primaria de la tabla o, si no hay clave primaria, la primera columna definida en el índice.

Los índices clustered son especialmente eficientes para consultas que recuperan rangos de datos consecutivos o que devuelven una gran cantidad de columnas de la tabla.

Índices Non-Clustered:

Los índices non-clustered son una estructura de índice separada de los datos de la tabla.

Cada tabla puede tener varios índices non-clustered.

Los índices non-clustered contienen las claves de índice y un puntero a la ubicación física de los datos en la tabla.

Aunque los índices non-clustered son más versátiles que los índices clustered, ya que permiten múltiples índices y no alteran la organización física de los datos, pueden ser menos eficientes para consultas que recuperan grandes cantidades de datos o realizan operaciones de ordenación, ya que requieren acceso adicional a los datos de la tabla.

Diferencias entre Índices Clustered y Non-Clustered:

La principal diferencia entre los índices clustered y non-clustered radica en la forma en que se organizan físicamente los datos.

Los índices clustered reorganizan los datos en la tabla según el orden de las claves de índice, mientras que los índices non-clustered mantienen una estructura de índice separada.

Los índices clustered son más eficientes para consultas que recuperan rangos de datos consecutivos o devuelven una gran cantidad de columnas de la tabla, mientras que los índices non-clustered son más versátiles pero pueden ser menos eficientes para consultas que recuperan grandes cantidades de datos.

Estrategias de Uso y Creación de Índices:

Crear índices de manera efectiva es crucial para mejorar el rendimiento de las consultas en SQL Server. Aquí hay algunas estrategias básicas y consideraciones clave:

Identificación de Escenarios de Uso:

Analiza las consultas más frecuentes y las operaciones de búsqueda que se realizan con mayor regularidad en la base de datos.

Identifica las consultas que podrían beneficiarse de índices para mejorar su rendimiento.

Selección de Columnas a Indexar:

Indexa las columnas que se utilizan frecuentemente en las cláusulas WHERE, JOIN y ORDER BY de las consultas.

Considera indexar columnas que tienen alta cardinalidad (valores únicos) y que se utilizan para filtrar datos con frecuencia.

Evita indexar columnas que tienen baja selectividad o que se actualizan con frecuencia, ya que los índices pueden afectar el rendimiento de las operaciones de escritura.

Casos de Uso y Ejemplos Prácticos:

Estudio de Casos: Análisis de casos reales donde se aplicaron estrategias de índices en SQL Server para mejorar el rendimiento de consultas en entornos de producción.

Ejemplos Prácticos: Demostración de la creación y utilización de índices en SQL Server a través de ejemplos de código y escenarios de consulta comunes.

CAPÍTULO III: METODOLOGÍA SEGUIDA.

Las acciones llevadas a cabo para realizar este trabajo fueron:

a. Descripción de cómo se realizó el Trabajo Práctico.

Una vez que recibí el tema del Trabajo Práctico, asignado por el profesor Dario Villegas de la cátedra de Base de Datos I, empecé a investigar el tema de consulta a través de índices. El primer trabajo estuvo bastante incompleto, no porque estuviese mal sino porque las pruebas no se podían ver los resultados debido a que el lote de datos no era demasiado grande, en esta ocasión el profesor mencionado anteriormente me recomendó generar un script (demostración a posteriori) para realizar una inserción masiva de 1 millón de registros en la tabla gastos, donde se puede ver mejor el resultado y el impacto en de los índices en la base de datos.

Una de las dificultades que encontré durante este proyecto estuvo relacionada con la naturaleza de los índices, es decir existen muchos tipos y no son gratis utilizarlos, en términos de almacenamiento. También crear índices sin discriminación no garantiza un mejor rendimiento, muchas veces hace que el sistema sea más complejo innecesariamente y dificulta el escalado y el mantenimiento de la base de datos.

Para resolver esto, utilice muchas bibliografías, páginas de internet y también con ayuda de inteligencia artificial, donde recomiendan que índices utilizar, en que columnas implementar y sobre todo cuando vale la pena utilizar.

b. Herramientas (Instrumentos y procedimientos).

Es importante recordar, que estas consultas se debe hacer en un entorno local o en una base de pruebas para verificar que funciona, para luego asi poder implementarla en la producción

- 1. Abrimos el SQL Server Management Studio.
- 2. Generamos una nueva consulta, donde realizaremos la prueba.

<u>Importante</u>: Las claves primarias (Primary key) generan un índice cluster por defecto en la mayoría de los motores de las bases de datos

```
--proyecto bases de datos 1 2024 - ALMIRON PEDRO AUGUSTO, OPTIMIZACION DE BUSQUEDA MEDIANTE INDICES

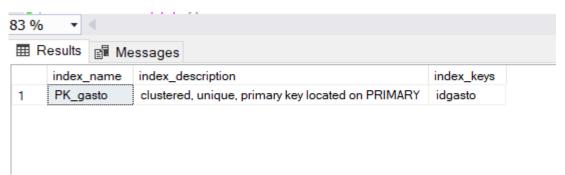
--use consorcio
--VERIFICAMOS QUE NO EXISTE NINGUN INDEX EN LA TABLA CONSORCIO MAS QUE LA PK QUE SE GENERA POR DEFECTO execute sp_helpindex 'gasto'

--activamos estadisticas de tiempo y de lectura

SET STATISTICS TIME ON;

SET STATISTICS IO ON;
```

- 3. La primer instrucción se utilizar para seleccionar la base de datos en la cual generaremos las consultas
- 4. El comando "execute sp_helpindex" nos devolverá en forma de registros cualquier tipo de índice que este creado en la tabla en este caso la tabla gasto:



Como era de esperarse la tabla gasto tiene un índice cluster generado por la PK gasto

i. Inserción Masiva de datos: Aquí generaremos el script que nos permitirá generar una inserción de 1 millón de registros en la tabla gasto de la base de datos consorcio, el script es el siguiente

```
SET NOCOUNT ON; -- Para mejorar el rendimiento al insertar una gran cantidad de registros
DECLARE @TotalRows INT = 1000000; -- Total de registros a insertar
DECLARE @Counter INT = 1;
DECLARE @FechaPago datetime = GETDATE(); -- Fecha de pago constante, se toma la fecha actual
DECLARE @Periodo INT = 1; -- Periodo fijo para todos los registros
WHILE @Counter <= @TotalRows
BEGIN
    INSERT INTO gasto (idprovincia, idlocalidad, idconsorcio, periodo, fechapago, idtipogasto, importe)
    VALUES (1,
            1,
            1,
            2,
            getdate(),
            1000); -- Mismo valor para el importe en cada registro
    SET @Counter = @Counter + 1;
FND:
```

Es un script que contiene lógica básica de programación simplemente es iterar el numero de veces necesaria la ejecución de la misma instrucción (en este caso el insert), los valores son exactamente los mismos para cada uno de los registros excepto obviamente la clave primaria, se puede modificar el script para obtener valores aleatorios por cada registro, pero eso podria afectar significativamente el tiempo de ejecución de la inserción, situación la cual no es el objetivo de este trabajo en particular.

j. Generamos una consulta a la tabla con el lote insertado:

A continuación, se generará una consulta a la tabla gasto con los registros insertados anteriormente y verificaremos el rendimiento de la misma con el plan de ejecución de Sql server

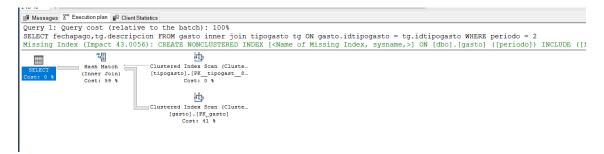
```
--realizamos la busqueda por periodo seleccionando fechade pago y tipo de gasto

SELECT fechapago,tg.descripcion FROM gasto

inner join tipogasto tg ON gasto.idtipogasto = tg.idtipogasto

WHERE periodo = 2
```

Plan de ejecución



El mismo sql server nos recomienda crear un índice para optimizar la consulta, pero muchas veces estos índices creados por defecto son inviables o "poco convenientes" es decir, no solo porque aparezca la recomendación debemos apresurarnos a crear el índice, muchas veces es importante revisar y contemplar otros aspectos antes de ejecutar el script

Aquí están los costos estimados de lo que necesita la consulta para ser ejecutada

		Hash Match	
Clustered Index Scan (Clus	stered)	Use each row from the top input to build a hash table, and each row	
Scanning a clustered index, entirely or only a ra	ange.	from the bottom input to probe into the hash table, outputting all	
		matching rows.	
Physical Operation	Clustered Index Scan		
Logical Operation	Clustered Index Scan	Physical Operation	Hash Match
Estimated Execution Mode	Row	Logical Operation	Inner Join
Storage	RowStore	Estimated Execution Mode	Row
Estimated Operator Cost	7.97978 (41%)	Estimated Operator Cost	11.5150325 (59%)
Estimated I/O Cost	6.32683	Estimated I/O Cost	0
Estimated Subtree Cost	7.97978	Estimated Subtree Cost	19.4981
Estimated CPU Cost	1.65296	Estimated CPU Cost	10.7938
Estimated Number of Executions	1	Estimated Number of Executions	1
Estimated Number of Rows to be Read	1502540	Estimated Number of Rows Per Execution	1412310
Estimated Number of Rows for All Executions	s 1412310	Estimated Number of Rows for All Executions	1412310
Estimated Number of Rows Per Execution	1412310	Estimated Row Size	44 B
Estimated Row Size	23 B	Node ID	0
Ordered	False		
Node ID	2	Output List	
		[proyecto_consorcio].[dbo].[gasto].fechapago, [pro	yecto_consorcio].
Predicate		[dbo].[tipogasto].descripcion	
[proyecto_consorcio].[dbo].[gasto].[periodo]=(2	2)	Hash Keys Probe	
Object		[proyecto consorcio].[dbo].[gasto].idtipogasto	
[proyecto_consorcio].[dbo].[gasto].[PK_gasto]		Probe Residual	
Output List		[proyecto_consorcio].[dbo].[tipogasto].[idtipogasto	ol as [tg].
[proyecto_consorcio].[dbo].[gasto].fechapago, [[proyecto_consorcio].	[idtipogasto]=[proyecto_consorcio].[dbo].[gasto].[_

Ejecutamos la consulta



Como se puede apreciar la consulta arroja un resultado de 1.404.543 registros con un tiempo de respuesta de 5815 ms lo que equivale a 5.815 segundos de respuesta

k. Creamos el índice para ejecutar la misma consulta:

(IMPORTANTE) En las consignas del trabajo se indica que se debe crear un índice clustered, pero debido a que solo se puede tener un solo índice agrupado por tabla (en este caso es la PK_GASTO) entonces crearemos un índice no agrupado para realizar la prueba

--creamos algunos index no clustered en la tabla gasto, para el campo periodo CREATE NONCLUSTERED INDEX IDX_GASTO_TIPOGASTO ON gasto(periodo);

```
Messages Client Statistics

SQL Server parse and compile time:

CPU time = 0 ms, elapsed time = 4 ms.

Table 'gasto'. Scan count 1, logical reads 8570, physical reads 0,

SQL Server Execution Times:

CPU time = 672 ms, elapsed time = 770 ms.

Completion time: 2024-02-15T16:12:16.8790236-03:00
```

El índice se crea con éxito, veamos el plan de ejecución de la consulta definida al principio

Query:

```
--ejecutamos los select con los indices no clustered

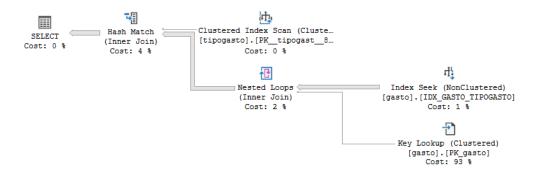
]SELECT fechapago, tg.descripcion

FROM gasto WITH(INDEX(IDX_GASTO_TIPOGASTO))

INNER JOIN tipogasto tg ON gasto.idtipogasto = tg.idtipogasto

WHERE periodo = 2;
```

Plan de ejecucion



Resultados de la consulta con el índice clúster IDX_GASTO_TIPO GASTO

```
SOL Server parse and compile time:

CPU time = 2 ms, elapsed time = 2 ms.

(1404543 rows affected)

Table 'Workfile'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server re
```

COMPAREMOS LOS RESULTADOS

QUERY SIN INDICE: 5.815 seg de respuesta

QUERY CON INDICE: 5.927 seg de respuesta

Es decir, en este caso, la creación del índice no fue conveniente y se provoco una perdida en el rendimiento del motor para realizar la consulta

I) Creamos un índice compuesto entre fecha pago y tipo gasto:

El motor de SQL server, posee una herramienta que nos especifica que al crear un índice compuesto podríamos mejorar la eficiencia de la ejecución de la consulta

Script de la generación del índice

```
--ahora creamos un indice compuesto

USE [proyecto_consorcio]

GO

|CREATE NONCLUSTERED INDEX [<IDX_fechapago_idtipogasto>]

ON [dbo].[gasto] ([periodo])

INCLUDE ([fechapago],[idtipogasto])

GO
```

Query:

```
--ejecutamos los select con los indices no clustered

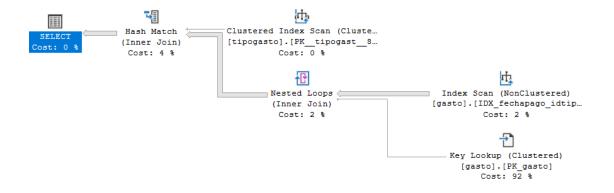
SELECT fechapago, tg.descripcion

FROM gasto with(INDEX(IDX_fechapago_idtipogasto))

INNER JOIN tipogasto tg ON gasto.idtipogasto = tg.idtipogasto

WHERE periodo = 2;
```

Plan de ejecución:



Resultados

CAPÍTULO IV: DESARROLLO DEL TEMA/RESULTADOS.

QUERY CON INDICE CLUSTER PK_GASTO: 5.815 seg de respuesta

QUERY CON INDICE NO CLUSTER: 5.927 seg de respuesta

QUERY CON INDICE NO CLUSTER COMPUESTO: 5.915 seg de respuesta

Durante nuestras pruebas de rendimiento en la base de datos del proyecto, hemos evaluado el impacto de diferentes tipos de índices en el tiempo de respuesta de las consultas.

Al analizar los resultados, podemos observar que la consulta utilizando el índice cluster (PK_GASTO) fue ligeramente más rápida en comparación con las consultas que utilizaron índices no cluster. Sin embargo, la diferencia en el tiempo de respuesta entre estas consultas fue mínima.

Hemos demostrado que, para la misma consulta, utilizando índices hemos podido ahorrarnos 12 milisegundos en tiempo de respuesta. Aunque la mejora puede parecer pequeña, en entornos de producción con una carga de trabajo más pesada y consultas más complejas, estos ahorros pueden acumularse y contribuir significativamente a mejorar el rendimiento del sistema en general.

CAPÍTULO IV: Conclusión.

En resumen, nuestras pruebas de rendimiento han demostrado la significativa influencia que los índices ejercen en el tiempo de respuesta de las consultas en SQL Server. Aunque las diferencias entre el uso de índices clustered y no clustered fueron mínimas en este caso específico, queda claro que la implementación adecuada de índices puede conducir a mejoras palpables en el rendimiento del sistema. Estos resultados subrayan la importancia de considerar cuidadosamente la creación de índices y su mantenimiento como parte integral de la optimización de bases de datos, especialmente en entornos donde la eficiencia y la capacidad de respuesta son críticas para el éxito del proyecto.

En segundo lugar, es esencial reconocer que, aunque los índices pueden proporcionar mejoras en el tiempo de respuesta de las consultas, su diseño y uso deben ser evaluados en función de las necesidades específicas del sistema. Esto implica una comprensión profunda de las consultas frecuentes, el patrón de acceso a los datos y la carga de trabajo esperada. Al equilibrar cuidadosamente estos factores y aplicar estrategias efectivas de creación y mantenimiento de índices, podemos aprovechar al máximo el potencial de los índices para optimizar el rendimiento de las bases de datos SQL Server.

CAPÍTULO IV: Bibliografía.

- https://www.sqlshack.com/es/vision-general-y-estrategias-de-indices-sql/
- https://use-the-index-luke.com/es/sql/%C3%ADndice-anatom%C3%ADa
- https://es.wikipedia.org/wiki/%C3%8Dndice (base de datos)
- https://www.ibm.com/docs/es/mam/7.6.0.8?topic=databases-database-indexing