

# UNIDAD PROFESIONAL INTERDISCIPLINARIA EN INGENIERÍA Y TECNOLOGÍAS AVANZADAS UPIITA

PROFESOR: Carlos De La Cruz Sosa

ASIGNATURA: Bases de Datos Distribuidas

**GRUPO:** 3TM3

# Práctica II

# **EQUIPO 5:**

- Bernal Aguilar Yuvia Abigail
- Contreras Jimenez Mariana Montserrat
  - Medina Gómez Jimena Zarahí

#### Introducción:

Cuando enviamos una consulta en lenguaje SQL, el motor de base de datos elabora una estrategia detallada: un mapa de ruta interno conocido como plan de ejecución. Estos planes actúan como una hoja de ruta que guía al motor en cómo acceder a las tablas, aplicar filtros, realizar uniones y, finalmente, entregar el conjunto de resultados deseado.

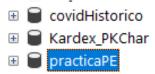
Al analizar estos planes, es posible comprender cómo se ejecuta realmente una consulta dentro del sistema. Esto permite identificar ineficiencias, optimizar el uso de recursos, detectar qué operaciones consumen más tiempo y aplicar mejoras clave en el rendimiento. Uno de los elementos más importantes en esta optimización son los índices.

Los índices están basados en estructuras que permiten localizar filas dentro de las tablas de forma más rápida y eficiente. Funcionan almacenando una copia ordenada de las columnas clave de una tabla, lo cual acelera significativamente el acceso a los datos más relevantes y mejora el rendimiento general de las consultas.

En conclusión, entender y aprovechar adecuadamente los planes de ejecución y los índices no solo permite optimizar el rendimiento de nuestras consultas, sino que también contribuye al diseño de bases de datos más extensas, escalables y eficientes a largo plazo.

#### Desarrollo:

1. Crear una base de datos con el nombre: practicaPE.



- 2. Copiar a la base de datos practicaPE las siguientes tablas de la base de datos AdventureWorks:
  - a) Sales.SalesOrderHeader

```
--- Tabla SalesOrderHeader
select * into order_header
from AdventureWorks2019.sales.SalesOrderHeader;
```

b) Sales.SalesOrderDetail

```
--- Tabla SalesOrderDetail
select * into order_detail
from AdventureWorks2019.sales.SalesOrderDetail;
```

c) Sales.Customer

```
-- Tabla Customer
select * into customer
from AdventureWorks2019.sales.Customer;
```

d) Sales.SalesTerritory

```
-- Tabla SalesTerritory
select * into SalesTerritory
from AdventureWorks2019.sales.SalesTerritory;
```

e) Production.Product

```
-- Tabla Product
select * into products
from AdventureWorks2019.Production.Product;
```

f) Production.ProductCategory

```
-- Tabla ProductCategory
select * into ProductCategory
from AdventureWorks2019.Production.ProductCategory;
```

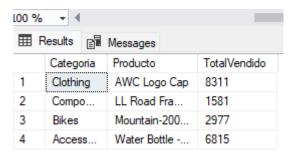
g) Production. Product Subcategory

```
-- Tabla ProductSubCategory
select * into ProductSubCategory
from AdventureWorks2019.Production.ProductSubCategory;
```

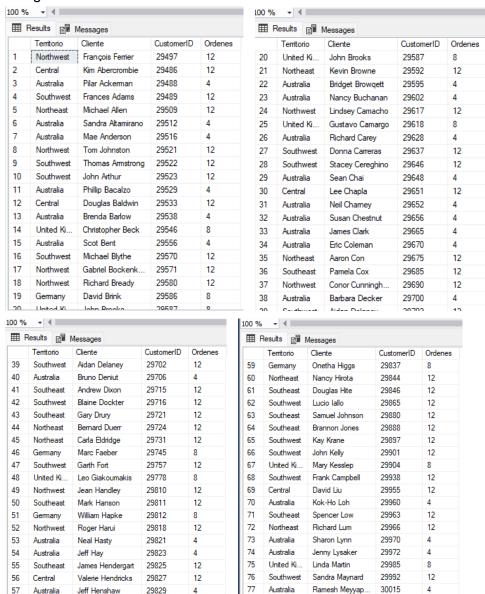
h) Person.Person

```
-- Tabla Person
--Lo hacemos de este modo ya que existen datos XML
CREATE TABLE person (
   BusinessEntityID int NOT NULL,
    PersonType nchar(2) NOT NULL,
   NameStyle bit NOT NULL,
   Title nvarchar(8),
   FirstName nvarchar(50) NOT NULL,
   MiddleName nvarchar(50).
   LastName nvarchar(50) NOT NULL,
   Suffix nvarchar(10),
   EmailPromotion int NOT NULL,
   AdditionalContactInfo xml,
   Demographics xml,
   rowguid uniqueidentifier NOT NULL,
   ModifiedDate datetime NOT NULL
);
  INSERT INTO person
  SELECT * FROM AdventureWorks2019.Person.Person;
```

- 3. Codificar las siguientes consultas:
- a) Listar el producto más vendido de cada una de las categorías registradas en la base de datos.



b) Listar el nombre de los clientes con más ordenes por cada uno de los territorios registrados en la base de datos.



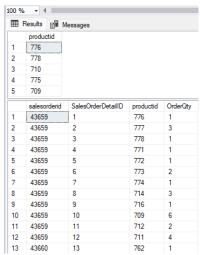
E0

20024

20010

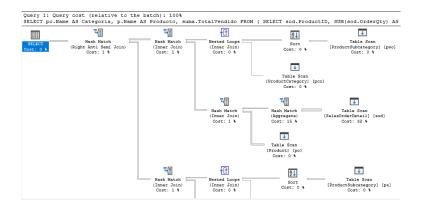


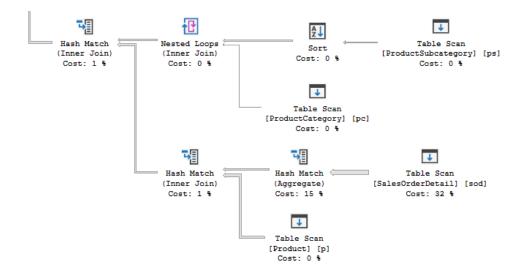
c) Listar los datos generales de las ordenes que tengan al menos los mismos productos de la orden con salesorderid = 43676.



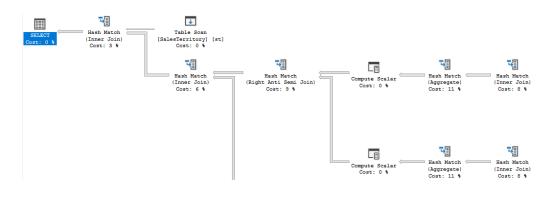
4. Generar los planes de ejecución de las consultas en las bases de datos practicaPE y proponer índices para mejorar el rendimiento de las consultas. -Planes de Ejecuciones:

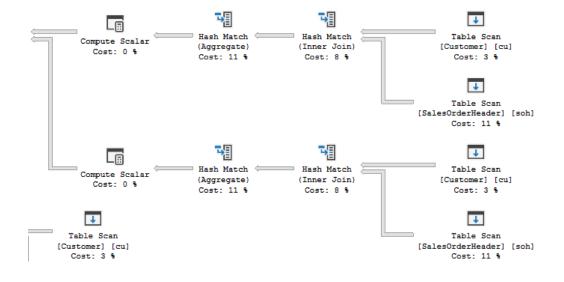
a)



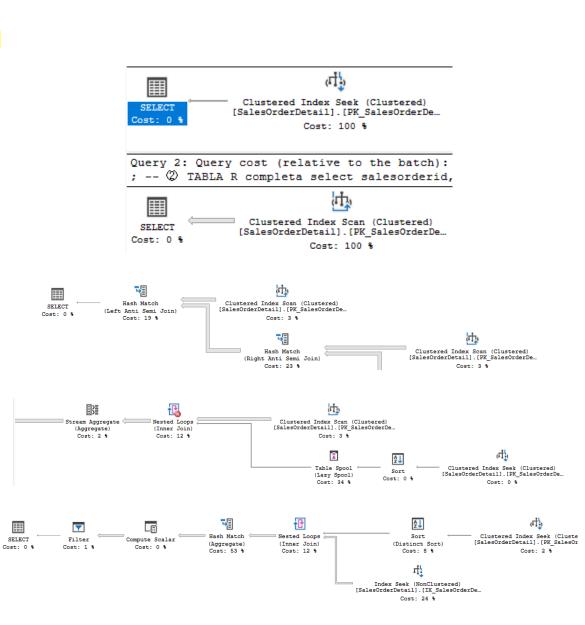


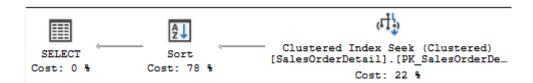
# b)





c)





Los índice propuestos son:

b) 1. CREATE NONCLUSTERED

### INDEX IX\_Person\_BEID

ON dbo.Person (BusinessEntityID) INCLUDE (FirstName, LastName); GO

#### 2. CREATE NONCLUSTERED

#### INDEX IX\_SOH\_CustomerID

ON dbo.SalesOrderHeader (CustomerID); GO

→ Con el uso de estos índices ayudará a que a la hora de la ejecución no se recorran las tablas completas y solo se recorra lo que se utilizara. Esto se observo porque en las consultas se utilizaron comandos como "Join", "Group by" & "where"

SELECT		
Cached plan size	200 KB	
Estimated Operator Cost	0 (0%)	
Estimated Subtree Cost	4.28936	
Estimated Number of Rows for All Executions	0	
Estimated Number of Rows Per Execution	18342.7	
Statement  SELECT  st.Name AS Territorio, per.FirstName + ' ' + per.LastName AS Cliente, cu.CustomerlD, conteo.Ordenes  FROM (  SELECT  soh.CustomerlD, cu.TerritoryID, COUNT(*) AS Ordenes  FROM SalesOrderHeader soh JOIN Customer cu ON cu.CustomerlD = soh.Custo GROUP BY soh.CustomerlD, cu.TerritoryID ) AS conteo	omerlD	
JOIN Customer cu ON cu.CustomerID = conteo.Cus JOIN Person per ON per.BusinessEntityID = cu.Perso JOIN SalesTer		

Nota: En el inciso a) y c) no hay necesidad de proponer o en su defecto agregar algún otro índice ya que ambas consultas estan optimizadas en su funcionamiento.

#### Comprobación:

a)

SELECT	
Cached plan size	24 KB
Estimated Operator Cost	0 (0%)
Estimated Subtree Cost	0.0032917
Estimated Number of Rows for All Executions	0
Estimated Number of Rows Per Execution	8.8125
Statement ① PRODUCTOS del pedido 43676 select productid from AdventureWorks2022.sales.SalesOrderDetail where salesorderid = 43676	

```
Query 1: Query cost (relative to the batch): 0%
-- ① PRODUCTOS del pedido 43676 select productid from AdventureWorks2022.sales.SalesOrderDetail where salesorderid = 43676
```

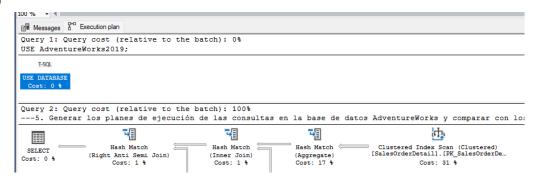
c)

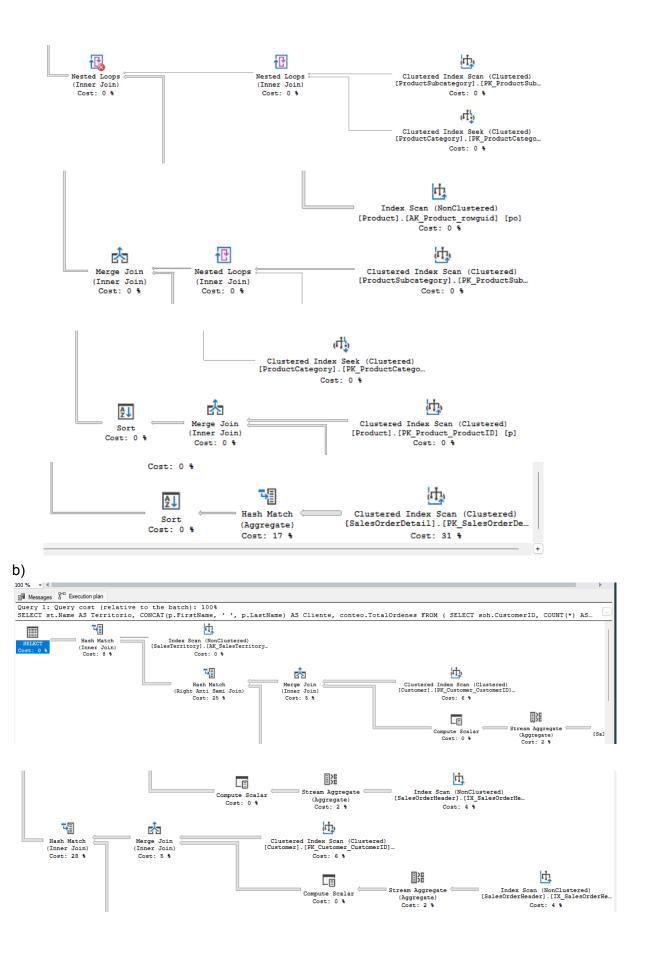
SELECT		
Cached plan size	24 KB	
Estimated Operator Cost	0 (0%)	
Estimated Subtree Cost	0.0148551	
Estimated Number of Rows for All Executions	0	
Estimated Number of Rows Per Execution	25.3052	
Statement; ;		

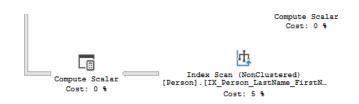
```
Query 1: Query cost (relative to the batch): 100%
SELECT st.Name AS Territorio, per.FirstName + ' ' + per.LastName AS Cliente, cu.CustomerID, conteo.Ordenes FROM ( SELECT sch.Cust...
```

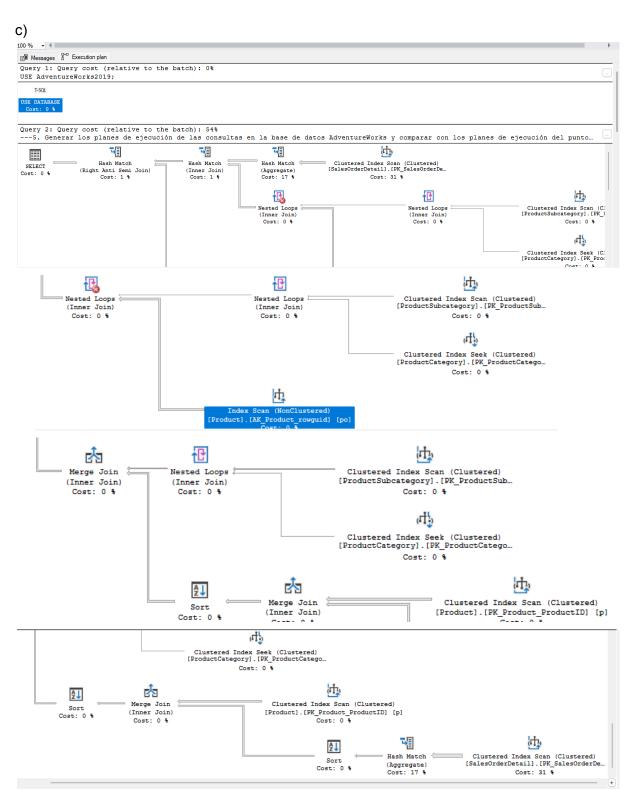
5. Generar los planes de ejecución de las consultas en la base de datos AdventureWorks y comparar con los planes de ejecución del punto 4.

a)









Query 3: Query cost (relative to the batch): 30%; ---Listar el nombre de los clientes con más ordenes por cada uno de los territorios registrados en la base de datos. SELECT st.Name AS Te. 瑁 ιŢ, Hash Match Index Scan (NonClustered)
[SalesTerritory].[AK\_SalesTerritory... SELECT Cost: 0 % 猖 ďψ, Å Merge Join (Inner Join) Cost: 5 % ered Index Scan (Clustered)
er].[PK\_Customer\_CustomerID] Hash Match t Anti Semi Join) Cost: 25 % Cost: 6 % 35 mpute Scalar Cost: 0 % CUSL. 3 T 35 4 Index Scan (NonClustered)
[SalesOrderHeader].[IX\_SalesOrderHe. Stream Aggregate Compute Scalar Cost: 0 % (Aggregate) Cost: 2 % Cost: 4 % 瑁 ďψ 么 Hash Match Merge Join Clustered Index Scan (Clustered) (Inner Join) Cost: 28 % (Inner Join) Cost: 5 % [Customer].[PK Customer CustomerID].. Cost: 6 % 猖 įψ, 乙 Hash Match Clustered Index Scan (Clustered)
[Customer].[PK\_Customer\_CustomerID].. Merge Join (Inner Join) (Inner Join) Cost: 28 % Cost: 5 % Cost: 6 % 38 4 Stream Aggregate Index Scan (NonClustered)
[SalesOrderHeader].[IX\_SalesOrderHe. Compute Scalar Cost: 0 % Cost: 4 % Index Scan (NonClustered)
[Person].[IX\_Person\_LastName\_FirstN...

# Comparación con el plan de ejecución del punto 4.

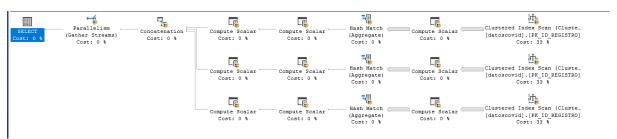
Las consultas ejecutadas en las bases de datos practicaPE y AdventureWorks2019 presentan diferencias marcadas, tanto en su estructura como en su rendimiento, esto lo podemos observar en sus respectivos planes de ejecución.

En la base de datos AdventureWorks2019, las consultas están mejor estructuradas, usando correctamente llaves foráneas, los nombres de las tablas son completamente calificados y vistas intermedias que facilitan la planificación eficiente del optimizador de consultas.

Por otro lado, la base de datos del punto 4 presenta algunos errores de lógica en los JOIN, así mismo hay uso incorrecto de columnas en la consulta del producto más vendido. Además, algunas vistas en la base del ejercicio 4 incluyen más columnas de las necesarias, lo que incrementa el uso de recursos al realizar operaciones como EXCEPT.

Sin embargo, se observa una mejora en la eficiencia del agrupamiento por territorio en la consulta de clientes más activos, donde se optimiza el GROUP BY desde la subconsulta, reduciendo la carga de filtrado posterior. En general, aunque ambas versiones cumplen con sus objetivos, las consultas de AdventureWorks2019 tienden a generar planes de ejecución más eficientes debido a una mejor organización y claridad en la lógica de los JOIN y subconsultas.

- 6. Generar los planes de ejecución de las consultas 3, 4 5 de la práctica de consultas en la base de datos Covid y propone índices para mejorar el rendimiento.
- Consulta 3:



SELECT  Cached plan size 120 KB Estimated Operator Cost 0 (0%) Estimated Subtree Cost 1177.14 Estimated Number of Rows for All Executions 0 Estimated Number of Rows Per Execution 3  Statement SELECT 'Diabetes' AS morbilidad,
Estimated Operator Cost 0 (0%) Estimated Subtree Cost 1177.14 Estimated Number of Rows for All Executions 0 Estimated Number of Rows Per Execution 3  Statement SELECT 'Diabetes' AS morbilidad,
Estimated Subtree Cost 1177.14  Estimated Number of Rows for All Executions 0  Estimated Number of Rows Per Execution 3  Statement  SELECT 'Diabetes' AS morbilidad,
Estimated Number of Rows for All Executions 0 Estimated Number of Rows Per Execution 3  Statement SELECT 'Diabetes' AS morbilidad,
Estimated Number of Rows Per Execution 3  Statement SELECT 'Diabetes' AS morbilidad,
Statement SELECT 'Diabetes' AS morbilidad,
SELECT 'Diabetes' AS morbilidad,
SELECT 'Diabetes' AS morbilidad,
'Diabetes' AS morbilidad,
•
(COUNT(CASE WHEN DIABETES = 1 AND
CLASIFICACION_FINAL = 3 THEN 1 END) * 100.0) / COUNT
(CASE WHEN CLASIFICACION_FINAL = 3 THEN 1 END) AS porcentaie
FROM dbo.datoscovid
UNION ALI
SELECT
'Obesidad' AS morbilidad,
(COUNT(CASE WHEN OBESIDAD = 1 AND
CLASIFICACION_FINAL = 3 THEN 1 END) * 100.0) / COUNT
(CASE WHEN CLASIFICACION_FINAL = 3 THEN 1 END) AS
porcentaje
FROM dbo.datoscovid
UNION ALL
SELECT
'Hipertensión' AS morbilidad,
(COU

#### Los indice propuestos son:

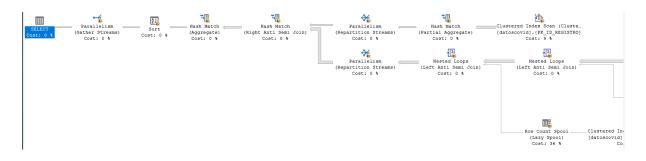
- CREATE NONCLUSTERED INDEX
   INDEX\_CLASIFICACION\_DIABETES
   on datoscovid(CLASIFICACION\_FINAL, DIABETES)
- CREATE NONCLUSTERED INDEX
   INDEX\_CLASIFICACION\_OBESIDAD on datoscovid(CLASIFICACION\_FINAL, OBESIDAD)
- CREATE NONCLUSTERED INDEX
   INDEX\_CLASIFICACION\_HIPERTENSION on datoscovid(CLASIFICACION\_FINAL, HIPERTENSION)

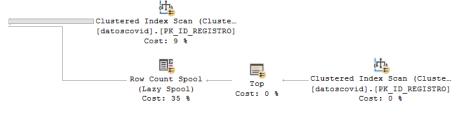
→ Se propusieron estos indices porque el programa buscara primero la clasificacion\_final=3 seguido de las enfermedades especificadas: obesidad, diabetes e hipertension.

Con indices propuestos:

SELECT	
Cached plan size	120 KE
Estimated Operator Cost	0 (0%
Estimated Subtree Cost	107.623
Estimated Number of Rows for All Executions	(
Estimated Number of Rows Per Execution	:
Statement	
SELECT	
'Diabetes' AS morbilidad,	
(COUNT(CASE WHEN DIABETES = 1 AND	
CLASIFICACION_FINAL = 3 THEN 1 END) * 100.0) / 0	COUNT
(CASE WHEN CLASIFICACION_FINAL = 3 THEN 1 EN	ID) AS
porcentaje	
FROM dbo.datoscovid	
UNION ALL	
SELECT	
'Obesidad' AS morbilidad,	
(COUNT(CASE WHEN OBESIDAD = 1 AND	
CLASIFICACION_FINAL = 3 THEN 1 END) * 100.0) / C	COUNT
(CASE WHEN CLASIFICACION_FINAL = 3 THEN 1 EN	ID) AS
porcentaje	
FROM dbo.datoscovid	
UNION ALL	
SELECT	
'Hipertensión' AS morbilidad,	
(COU	

# - Consulta 4:





\_\_Clustered Index Scan (Cluste... [datoscovid].[PK\_ID\_REGISTRO] Cost: 9 %

SELECT	
Cached plan size	88 KB
Estimated Operator Cost	0 (0%)
Estimated Subtree Cost	4320.25
Estimated Number of Rows for All Executions	0
Estimated Number of Rows Per Execution	347.136
Statement	

# Statement SELECT DISTINCT MUNICIPIO\_RES FROM dbo.datoscovid WHERE MUNICIPIO\_RES NOT IN ( SELECT DISTINCT MUNICIPIO\_RES FROM dbo.datoscovid WHERE CLASIFICACION\_FINAL = 3 -- Casos confirmados AND (DIABETES = 1 OR OBESIDAD = 1 OR TABAQUISMO = 1) ) ORDER BY MUNICIPIO\_RES

#### El indice propuesto es:

 CREATE NONCLUSTERED INDEX <u>INDEX\_CLASIFICACION\_FINAL</u> on datoscovid (CLASIFICACION\_FINAL) include (MUNICIPIO\_RES, DIABETES, OBESIDAD, TABAQUISMO)

→ Lo elegimos porque se necesita filtrar de la columna clasificacion final y ademas, debe incluirse los casos de diabetes, obesidad, tabaquismo y sus respectivos municipios.

#### Plan de ejecucion con indices:

SELECT	
Cached plan size	112 KB
Estimated Operator Cost	0 (0%)
Estimated Subtree Cost	3209.64
Estimated Number of Rows for All Executions	0
Estimated Number of Rows Per Execution	347.137
SELECT DISTINCT MUNICIPIO_RES FROM dbo.datoscovid WHERE MUNICIPIO_RES NOT IN (     SELECT DISTINCT MUNICIPIO_RES     FROM dbo.datoscovid     WHERE CLASIFICACION_FINAL = 3 Casos con     AND (DIABETES = 1 OR OBESIDAD = 1 OR TABE     = 1) ) ORDER BY MUNICIPIO_RES	

#### - Consulta 5:



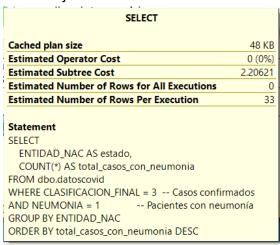
SELECT		
Cached plan size	40 KB	
Estimated Operator Cost	0 (0%)	
Estimated Subtree Cost	393.947	
Estimated Number of Rows for All Executions	0	
<b>Estimated Number of Rows Per Execution</b>	33	
Statement  SELECT ENTIDAD_NAC AS estado, COUNT(*) AS total_casos_con_neumonia  FROM dbo.datoscovid  WHERE CLASIFICACION_FINAL = 3 Casos confir AND NEUMONIA = 1 Pacientes con neu GROUP BY ENTIDAD_NAC  ORDER BY total_casos_con_neumonia DESC		

#### Los indice propuesto es:

1. CREATE NONCLUSTERED INDEX <a href="INDEX\_NEUMONIA">INDEX\_NEUMONIA</a> on datoscovid (neumonia, clasificacion\_final) include (entidad\_nac)

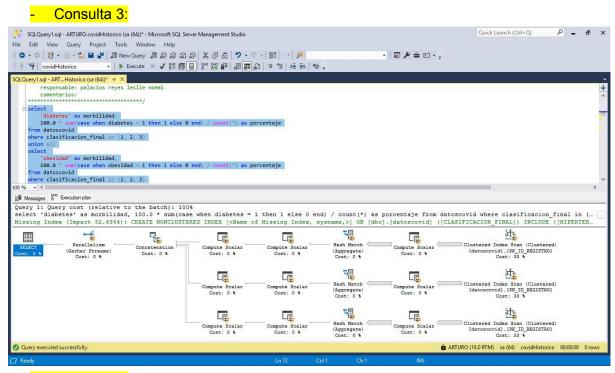
→ Fue propuesto porque se encontrara con mayor facilidad los casos de neumonia seguido de un filtrado de clasificacion final= 3

Plan de ejecucion con indices:

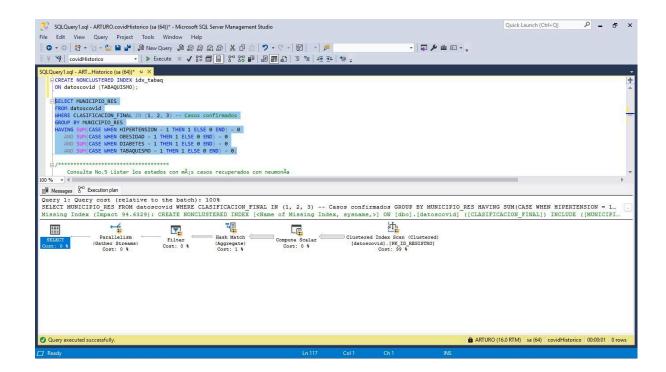


7. Comprobar los planes de ejercicio del punto 6 con los planes de ejecucion de otro equipo.

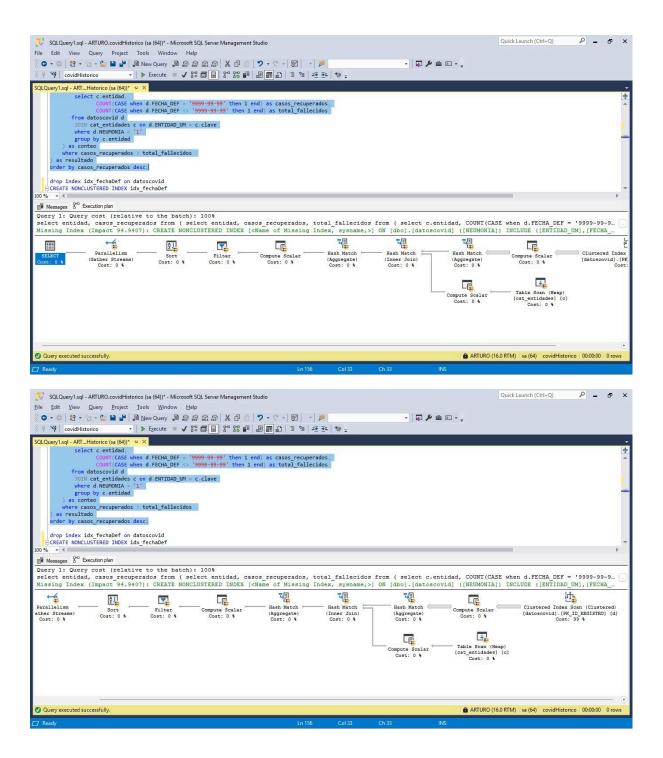
Comprobacion con el equipo 1.



Consulta 4:



#### - Consulta 5:



#### **Conclusiones:**

- Bernal Aguilar Yuvia Abigail

Podemos confirmar que los planes de ejecución son muchas veces fundamentales para poder conocer mejor sobre el proceso que se lleva a cabo al realizar una consulta en una base de datos ya que nos ayudaron para saber como es el proceso, que parte de la consulta se consume más recursos, problemas de estadísticas y de identificar la falta de índices, los cuales son indispensables en una base de datos que contienen miles de información y es necesario acceder a esta información de forma optima. Entre estos planes, observamos unos conceptos como lo son nested loop,

hash match y merge join. El primero es una unión de bucle anidado en donde se escanea la primera tabla y después se unen las filas coincidentes de la segunda tabla, el segundo término, hash match, que se crea una tabla hash para que después, de igual forma buscar coincidencias con la otra tabla. Por último, tenemos merge join ésta es diferente a las anteriores ya que se necesita que las dos tablas deberán estar ordenadas para después escanearse al mismo tiempo y encontrar términos iguales para combinarlas.

#### - Medina Gómez Jimena Zarahí

Con esta práctica comprendí la utilidad de consultar los planes de ejecución, su importancia y los distintos datos que se pueden obtener de ellos. Estos planes facilitan la identificación de mejoras que no son evidentes al momento de compilar los queries. Todo este análisis se realiza con el objetivo de aumentar la velocidad de cada consulta y determinar si se requieren índices adicionales que permitan optimizar su ejecución.

Además, al comparar los distintos planes de ejecución en diferentes equipos, puede establecer estrategias que demuestran cómo muchas consultas pueden simplificarse sin alterar sus resultados, logrando una ejecución más eficiente y un menor consumo de recursos.

#### - Contreras Jiménez Mariana Montserrat

En esta práctica comprendí la importancia de ejecutar consultas junto con la revisión de sus planes de ejecución, ya que esto no solo permite verificar su funcionamiento, sino también evaluar su eficiencia y el uso de recursos, especialmente al trabajar con grandes volúmenes de datos. A través del análisis de cada plan, fue posible identificar operaciones costosas y detectar oportunidades para optimizar el rendimiento mediante el uso de índices. Los índices, correctamente aplicados, ayudan a dirigir las búsquedas de forma más precisa, especialmente en consultas que utilizan comandos como join o select. Incorporarlos de manera adecuada permite obtener respuestas más rápidas, claras y concisas, lo cual mejora significativamente la experiencia en el manejo de bases de datos.

#### Conclusión General:

Durante esta práctica comprendimos que los planes de ejecución son herramientas fundamentales para analizar y mejorar el rendimiento de las consultas en bases de datos. Estos planes nos permiten observar con detalle cómo se procesa una consulta, qué partes consumen más recursos, y nos ayudan a identificar problemas como estadísticas desactualizadas o la falta de índices, lo cual es crucial en bases de datos con grandes volúmenes de información. A través del análisis, detectamos conceptos técnicos como nested loop, hash match y merge join, cuya comprensión nos permitió conectar la teoría con el código SQL: por ejemplo, entendimos que nested loop escanea fila por fila para encontrar coincidencias, hash match crea una tabla hash para optimizar la comparación, y merge join requiere que las tablas están ordenadas para unirlas eficientemente. Al comparar los planes generados en distintos escenarios, también identificamos cómo reestructurar consultas sin cambiar sus

resultados, logrando una ejecución más simple, rápida y con menor carga de recursos. En definitiva, analizar los planes de ejecución no sólo valida el funcionamiento del código, sino que lo lleva a un nivel más eficiente y profesional.