**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

MÔN CƠ SỞ DỮ LIỆU

NÂNG CAO



Sinh viên thực hiện: 1753087 – Nguyễn Tuấn Phùng

1753093 – Hồ Bùi Văn Quang

1753132 – Võ Khánh Vy

GV phụ trách: TS. Nguyễn Trần Minh Thư

Th.S Hồ Thị Hoàng Vy

ĐỒ ÁN THỰC HÀNH 01 – CƠ SỞ DỮ LIỆU NÂNG CAO

HỌC KỲ I – NĂM HỌC 2019-2020

**BẢNG THÔNG TIN CHI TIẾT NHÓM**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Mã nhóm:** | 04 | | | |
| **Tên nhóm:** | **Nhóm04** | | | |
| **Số lượng:** | **3** | | | |
| **MSSV** | **Họ tên** | **Email** | **Điện thoại** | **Hình ảnh** |
| 1753087 | Nguyễn Tuấn Phùng | 1753087@student.hcmus.edu.vn | 0944312225 | Ảnh có chứa người, ngoài trời, bầu trời, cây  Mô tả được tạo tự động |
| 1753093 | Hồ Bùi Văn Quang | 1753093@student.hcmus.edu.vn | 0355156808 | Ảnh có chứa người, bầu trời, ngoài trời, người đàn ông  Mô tả được tạo tự động |
| 1753132 | Võ Khánh Vy | 1753132@student.hcmus.edu.vn | 0329576841 | Ảnh có chứa cây, ngoài trời, hàng rào, tòa nhà  Mô tả được tạo tự động |

**Nội dung yêu cầu trong đồ án:**

1. Cho biết tổng doanh thu từ UnitPrice và tổng doanh thu từ UnitPriceDiscount của từng sản phẩm trong thời gian trước “2014-05-15”
2. Cho danh sách các nhân viên bán hàng chưa đạt quota trong năm ngoái.
3. Cho biết tổng thu từ bán hàng online, offline của năm 2014 và 2015
4. Cho danh sách các khách hàng đã mua sản phẩm 'Water Bottle - 30 oz.'
5. Cho danh sách các khách hàng đã mua sản phẩm 'Water Bottle - 30 oz.' trong thời gian giữa '2013-05-01' và '2013-08-01'
6. Chạy thực thi các truy vấn trên (câu 2a) và xem kết quả thực thi với execution plan, nhận xét kết quả.
7. Tạo các chỉ mục như yêu cầu dưới đây vào csdl Sale Management Index, chạy lại các truy vấn trên với csdl có chỉ mục và nhận xét kết quả:
   1. Tạo chỉ mục nonCluster Index cho OrderDate (truy vấn 2a)
   2. Tạo chỉ mục nonCluster Index cho product Name (truy vấn2d, 2e)
8. Từ kết quả câu f, ghi nhận lại index recommendation từ execution plan và đề xuất 1 chỉ mục phù hợp. Quan sát exection plan giải thích nguyên nhân, và ghi nhận kết quả sau khi thực hiện tạo chỉ mục đề xuất.
9. Tạo partition trên bảng OrderHeader theo Năm lập hoá đơn (trên hai csdl có index và không index). Thực hiện lại các kết quả truy vấn trên, nhận xét kết quả.

**Nội dung thực hiện trong báo cáo:**

1. Chạy thực thi các truy vấn trên (câu 2a) và xem kết quả thực thi với execution plan, nhận xét kết quả. (a với f)
2. Tạo các chỉ mục như yêu cầu dưới đây vào csdl SaleManagement Index, chạy lại các truy vấn trên với csdl có chỉ mục và nhận xét kết quả: (câu d, e, g)

- Tạo chỉ mục nonCluster Index cho OrderDate (truy vấn 2a)

- Tạo chỉ mục nonCluster Index cho product Name (truy vấn2d, 2e)

1. Từ kết quả câu f, ghi nhận lại index recommendation từ execution plan và đề xuất 1 chỉ mục phù hợp. Quan sát exection plan giải thích nguyên nhân, và ghi nhận kết quả sau khi thực hiện tạo chỉ mục đề xuất. (câu h)
2. Tạo partition trên bảng OrderHeader theo Năm lập hoá đơn (trên hai csdl có index và không index). Thực hiện lại các kết quả truy vấn trên, nhận xét kết quả.
3. **Chạy thực thi các truy vấn trên (câu 2a) và xem kết quả thực thi với execution plan, nhận xét kết quả.**

Khi bắt đầu thực hiện một câu truy vấn SQL Server sẽ lên một kế hoạch gồm các bước sẽ tiến hành để thực thi câu lệnh đó, gọi là kế hoạch thực thi (execution plan). Sử dụng tình năng “Include Actual Execution Plan” trên thanh công cụ, khi đó mỗi lần Execute câu truy vấn hệ thống sẽ vừa thực hiện câu lệnh vừa đồng thời trả lại kế hoạch thực thi mà nó đã dùng để thực hiện câu lệnh đó.

* Cho biết tổng doanh thu UnitPrice của từng sản phẩm trong thời gian trước ‘2014-05-15’

+ Câu truy vấn:

Select sod.ProductID, SUM(sod.UnitPrice) as TongTien

From dbo.n\_SalesOrderDetail as sod join dbo.n\_SalesOrderHeader as soh

on(sod.SalesOrderID = soh.SalesOrderID)

Where soh.OrderDate < '2014-05-15' and sod.ProductID =

(Select pro.ProductID

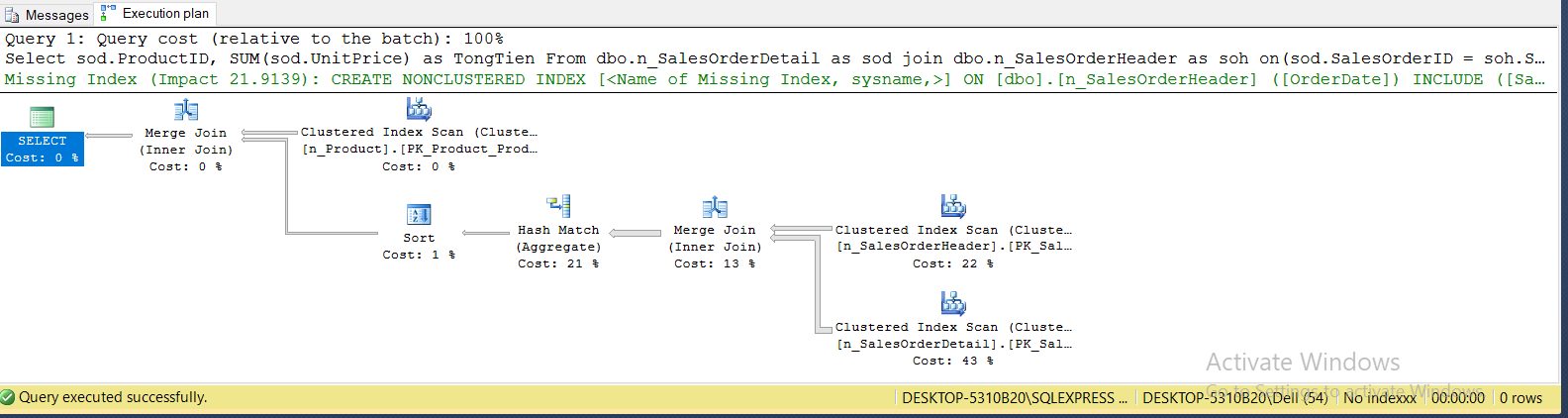
From dbo.n\_Product as pro

where sod.ProductID = pro.ProductID)

Group by sod.ProductID

Order by sod.ProductID ASC

+ Kết quả thực thi với execution plan:



* Cho biết tổng doanh thu UnitPriceDiscount của từng sản phẩm trong thời gian trước ‘2014-05-15’

+ Câu truy vấn:

Select sod.ProductID, SUM(sod.UnitPriceDiscount) as TongTien

From dbo.n\_SalesOrderDetail as sod join dbo.n\_SalesOrderHeader as soh

on(sod.SalesOrderID = soh.SalesOrderID)

Where soh.OrderDate < '2014-05-15' and sod.ProductID =

(Select pro.ProductID

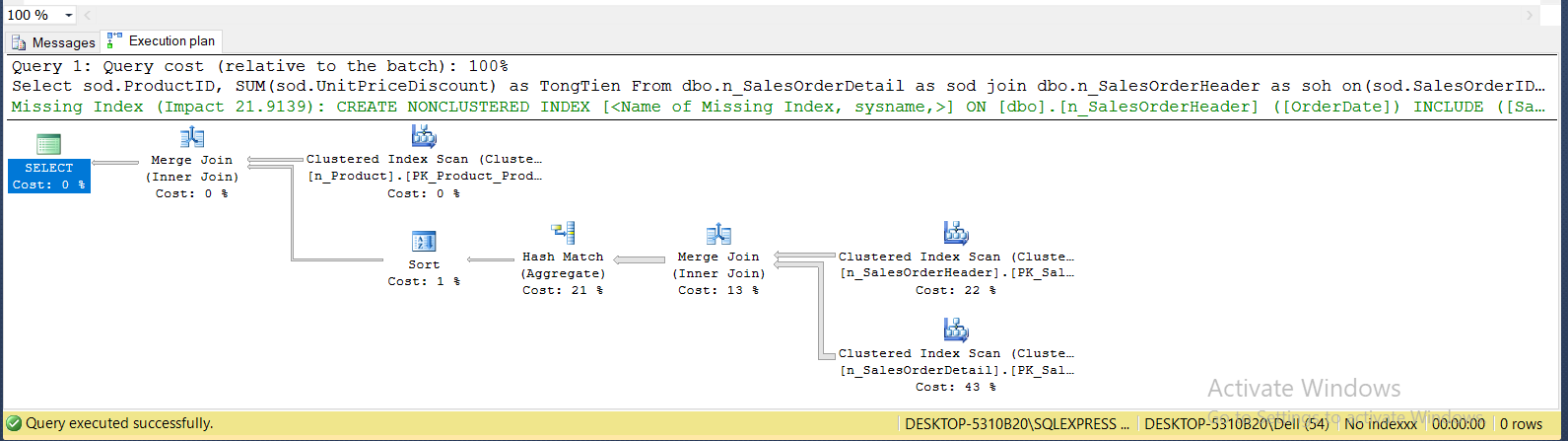
From dbo.n\_Product as pro

where sod.ProductID = pro.ProductID)

Group by sod.ProductID

Order by sod.ProductID ASC

+ Kết quả thực thi với execution plan:



* Nhận xét:

Kết quả sau khi chạy 2 câu truy vấn trên ta sẽ nhận được 266 dòng dữ liệu tương ứng với 266 produce\_id.

Hệ thống sẽ thực thi 2 câu truy vấn trên theo kế hoạch:

+ Clustered Index Scan: Quet chỉ mục, cụ thể ở đây là chỉ mục được tạo ra tự động do ràng buộc khóa chính là:

[PK\_SalesOrderHeader\_SalesOrderID]

[PK\_SalesOrderDetail\_SalesOrderID\_SalesOrderDetailID]

[PK\_Product\_ProductID]

+ Merge Join

+ Hash Match

+ Sort

Như vậy hệ thống sẽ thực hiện câu lệnh bằng cách duyệt qua cả bảng sau khi join các bảng một cách hợp lý và tìm ra các bảng ghi thỏa mãn yêu câu tìm kiếm và sau đó thực hiện phép cộng. Tức là hệ thống cần đọc tuần tự từng dòng từ đầu đến cuối để tìm ra kết quả, cụ thể ở đây câu truy vấn đã phải duyệt qua 21 6926 dòng để tìm được các dòng thỏa soh.OrderDate < '2014-05-15' (với ý 1), soh.OrderDate < '2014-05-15' (với ý 2).  Đây là một thao tác rất chậm vì nó phải xử lý tất cả các bản ghi trong bảng. Hệ thống sẽ không dừng lại khi nó tìm được bản ghi đầu tiên phù hợp, vì nó không biết liệu còn bản ghi nào khác có giá trị tương tự hay không, cho nên để chắc chắn trả lại kết quả đầy đủ hệ thống vẫn phải tiếp tục đọc các bản ghi còn lại. Có thể thấy chi phí của thao tác duyệt bảng tăng tuyến tính cùng với số lượng bản ghi trong bảng.

Hai câu truy vấn trên chiếm tới 100% tổng chi phí.

1. Tạo các chỉ mục như yêu cầu dưới đây vào csdl SaleManagementIndex, chạy lại các truy vấn trên với csdl có chỉ mục và nhận xét kết quả:

- Tạo chỉ mục nonCluster Index cho OrderDate (truy vấn 2a)

+ Tạo chỉ mục:

CREATE NONCLUSTERED INDEX OOrderDate\_INDEX ON n\_SalesOrderHeader(OrderDate)

+ Kết quả chạy câu truy vấn 2a

Ảnh có chứa ảnh chụp màn hình

Mô tả được tạo tự động

Nhận đượcc kết quả gồm 266 dòng và có các giá trị giống hệt với kết quả nhận được khi thực hiện câu truy vấn không sử dụng chỉ mục (nonclustered Index).

- Tạo chỉ mục nonCluster Index cho product Name (truy vấn2d, 2e)

+ Tạo chỉ mục:

CREATE NONCLUSTERED INDEX pRoDuCtNaMe\_INDEX ON n\_product(Name)

+ Kết quả chạy câu truy vấn 2d, 2e

Ảnh có chứa ảnh chụp màn hình

Mô tả được tạo tự động

Ảnh có chứa ảnh chụp màn hình

Mô tả được tạo tự động

Nhận đượcc kết quả có các giá trị giống hệt với kết quả nhận được khi thực hiện câu truy vấn không sử dụng chỉ mục (nonclustered Index).

Từ Execution Plan được tạo ra khi thực hiện câu truy vấn có sử dụng chỉ mục (Nonclustered Index) có thể nhận thấy sự xuất hiện của Index Seek. Index Seek là hệ thống có thể nhảy đến được node trên cây index chứa khóa thỏa mãn yêu cầu tìm kiếm. Index là một cấu trúc dữ liệu có dạng B-tree, nên nó rất thích hợp với các thao tác tìm kiếm theo kiểu key=value, chỉ cần vài phép so sánh là hệ thống định vị được node chứa khóa cần tìm.

1. **Từ kết quả câu f, ghi nhận lại index recommendation từ execution plan và đề xuất 1 chỉ mục phù hợp. Quan sát exection plan giải thích nguyên nhân, và ghi nhận kết quả sau khi thực hiện tạo chỉ mục đề xuất.**

* Index recommendation từ execution plan:

/\*

Missing Index Details from SQLQuery2.sql - DESKTOP-5310B20\SQLEXPRESS.No indexxx (DESKTOP-5310B20\Dell (54))

The Query Processor estimates that implementing the following index could improve the query cost by 21.9139%.

\*/

/\*

USE [No indexxx]

GO

CREATE NONCLUSTERED INDEX [<Name of Missing Index, sysname,>]

ON [dbo].[n\_SalesOrderHeader] ([OrderDate])

INCLUDE ([SalesOrderID])

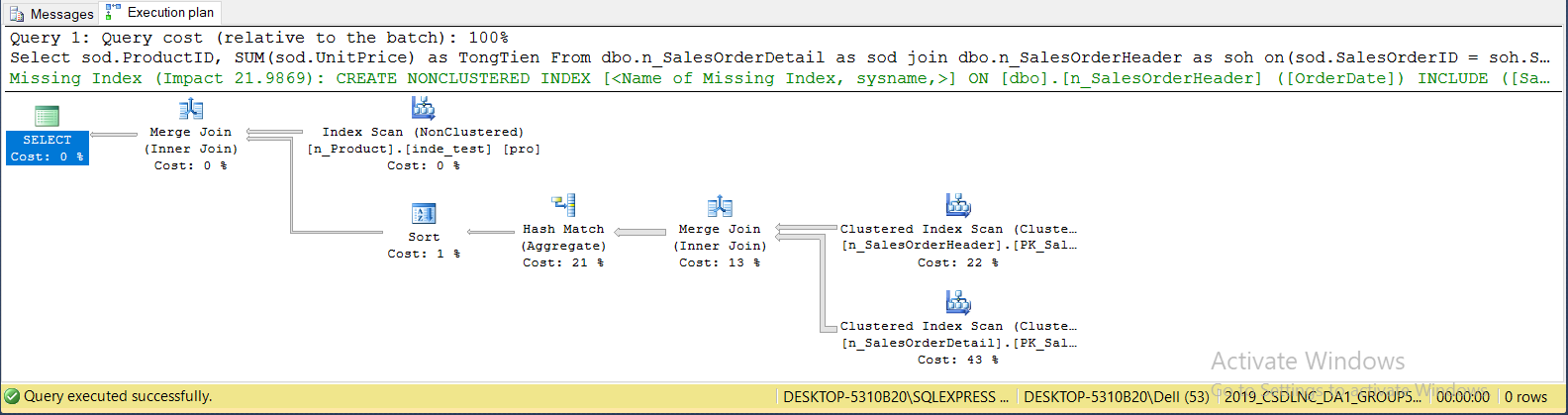
GO

\*/

* Đề xuất chỉ mục phù hợp:

CREATE NONCLUSTERED INDEX inde\_test ON [dbo].[n\_Product] ([ProductID])

* SKết quả sau khi tạo chỉ mục:



* Qua quan sát 2 cơ sở dữ liệu có index và không có index, thì ở cơ sở dữ liệu có index với chỉ mục đề xuất, nhận thấy được thời gian chạy của các thao tác SELECT, MERGE JOIN và INDEX SCAN (NONCLUSTERED) chạy nhanh hơn một chút dù chi phí vẫn là 100%.
* Không có chỉ mục:

A screenshot of a cell phone

Description automatically generatedA screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

* Có chỉ mục (với chỉ mục đề xuất):
* A screenshot of a cell phone

  Description automatically generated

A screenshot of a cell phone

Description automatically generatedA screenshot of a social media post

Description automatically generated

* Việc thêm chỉ mục sẽ làm tăng tốc độ truy vấn, giúp cho chương trình chạy nhanh hơn.

1. **Tạo partition trên bảng OrderHeader theo Năm lập hoá đơn (trên hai csdl có index và không index). Thực hiện lại các kết quả truy vấn trên, nhận xét kết quả.**

* Việc tạo partition chỉ có thể tạo được trên MS SQL version enterprise nên phần này em chỉ có query tạo và execute (còn có một cách khác là dùng Partitioned Views nhưng tách thuộc tính của một bảng nhiều quá làm không hết)
* Tạo Partition trên bảng OrderHeader theo Năm lập hóa đơn:

CREATE PARTITION FUNCTION partitionIndex(DATETIME) AS RANGE LEFT FOR VALUES ('20130501', '20130901')

GO

* Ta sẽ được 3 partition:
  + P1: p1 < 2013-05-01
  + P2: 2013-05-01 < p2 < 2013-09-01
  + P3: p3 > 2013-09-01
* Vì đề yêu cầu giữa 2013-05-01 và 2013-09-01 nên ta sẽ dùng p2, ta được query là:

SELECT snn.customerid, snn.modifieddate

FROM n\_salesorderheader snn JOIN n\_salesorderdetail snl ON(snn.salesorderid = snl.salesorderid)

WHERE (snl.productid =

(SELECT cdl.productid

FROM dbo.n\_product AS cdl WITH(PARTITION('p2'))

WHERE cdl.name LIKE 'Water Bottle - 30 oz.') )

GROUP BY snn.customerid, snn.ModifiedDate