

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN

Cơ sở dữ liệu nâng cao

Nhóm 3:

19126068 – Võ Minh Tuấn

19126047 – Nguyễn Hoàng Hùng

19126066 – Phạm Dung Thông

Thành phố Hồ Chí Minh
ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

BÁO CÁO ĐỒ ÁN

| Đề tài |

Đồ Án 2

| Giáo viên hướng dẫn |

Thầy Thái Lê Vinh

Cô Hồ Thị Hoàng Vy

Thành phố Hồ Chí Minh

LỜI CẢM ƠN

Để có thể hoàn thành đồ án này, chúng em xin gửi lời cảm ơn chân thành tới thầy Thái Lê Vinh và cô Hồ Thị Hoàng Vy khoa Công Nghệ Thông Tin, trường Đại học Khoa học Tự nhiên vì đã hỗ trợ tụi em trong suốt quãng thời gian làm đồ án, giúp đồ án này được hoàn thiện nhất có thể.

Do thời gian có hạn cũng như các kiến thức của chúng em vẫn có nhiều hạn chế, rất khó để tụi em tránh khỏi những sai sót khi hoàn thiện đồ án cũng như làm báo cáo. Vì vậy rất mong có thể nhận được những lời góp ý từ các thầy cô để chúng em có thể có thêm kiến thức để hoàn thành những đồ án tiếp theo.

Chúng em xin chân thành cảm ơn.

Mục Lục

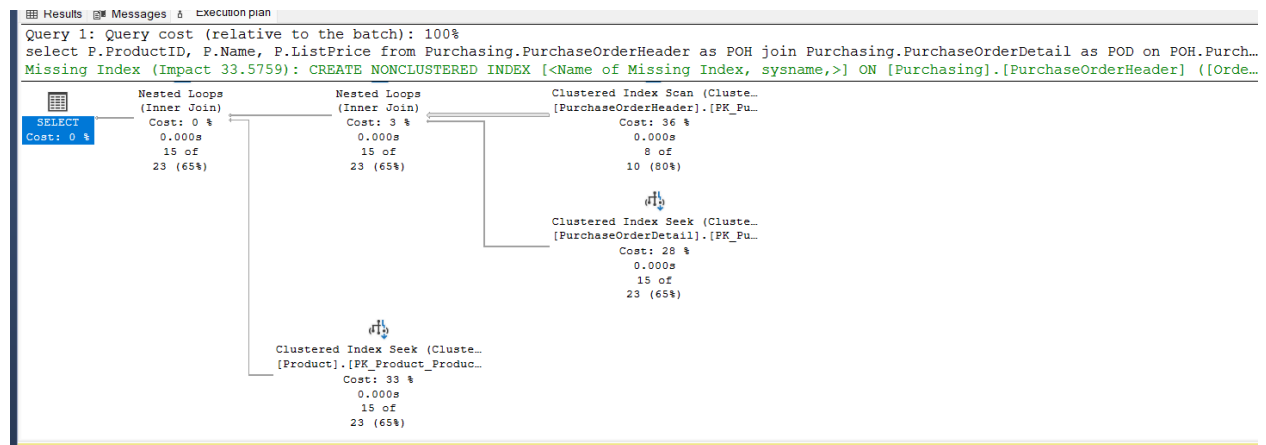
I. Quan sát Execution plan của các truy vấn trên	7
1. Query 1:	7
2. Query 2:	7
3. Query 3:	8
II. Đề xuất chỉ mục phù hợp làm tăng hiệu quả truy vấn. Quan sát và giải thích Execution plan trước và sau khi đề xuất giải pháp	10
1. Query 1.....	10
2. Query 2:	12
3. Query 3.....	14
III. Thành viên và bảng phân công:.....	16
1. Kế hoạch làm việc:	16
2. Source code	16

I. Quan sát Execution plan của các truy vấn trên

1. Query 1:

```
select P.ProductID, P.Name, P.ListPrice
from Purchasing.PurchaseOrderHeader as POH join
Purchasing.PurchaseOrderDetail as POD on POH.PurchaseOrderID =
POD.PurchaseOrderID
join Production.Product as P on P.ProductID = POD.ProductID
where POH.OrderDate between '2012-01-01' and '2013-01-01'
```

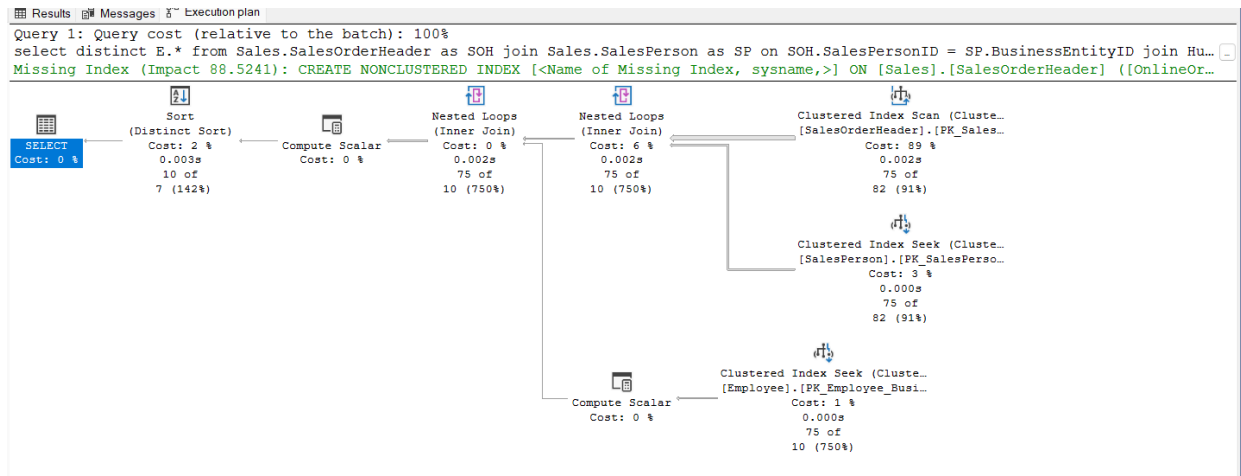
Ta được execution plan như sau:



2. Query 2:

```
select distinct E.*
from Sales.SalesOrderHeader as SOH join Sales.SalesPerson as SP on
SOH.SalesPersonID = SP.BusinessEntityID
join HumanResources.Employee as E on SP.BusinessEntityID =
E.BusinessEntityID
where OrderDate > '2011-06-30' and OrderDate < '2011-08-01' and
OnlineOrderFlag = '0'
```

Ta được execution plan như sau:



3. Query 3:

```

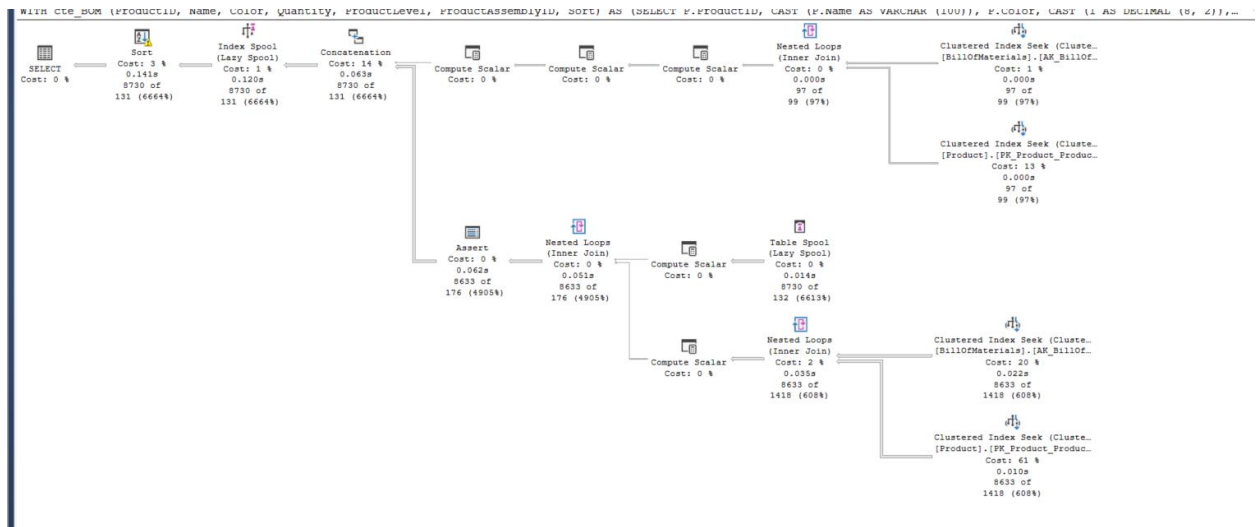
WITH cte_BOM (ProductID, Name, Color, Quantity, ProductLevel,
ProductAssemblyID, Sort)
AS (SELECT P.ProductID,
        CAST (P.Name AS VARCHAR (100)),P.Color,CAST (1 AS DECIMAL
(8, 2)),1,NULL,
        CAST (P.Name AS VARCHAR (100))
FROM    Production.Product AS P
INNER JOIN
Production.BillofMaterials AS BOM
ON BOM.ComponentID = P.ProductID
AND BOM.ProductAssemblyID IS NULL
AND (BOM.EndDate IS NULL
OR BOM.EndDate > GETDATE())

UNION ALL
SELECT P.ProductID,
        CAST (REPLICATE('-', cte_BOM.ProductLevel) + P.Name AS
VARCHAR (100)),
        P.Color,
        BOM.PerAssemblyQty,
        cte_BOM.ProductLevel + 1,
        cte_BOM.ProductID,
        CAST (cte_BOM.Sort + '\' + p.Name AS VARCHAR (100))
FROM    cte_BOM
INNER JOIN Production.BillofMaterials AS BOM
ON BOM.ProductAssemblyID = cte_BOM.ProductID
INNER JOIN Production.Product AS P
ON BOM.ComponentID = P.ProductID
AND (BOM.EndDate IS NULL
OR BOM.EndDate > GETDATE())
)
SELECT ProductID,
        Name,
        Color,

```

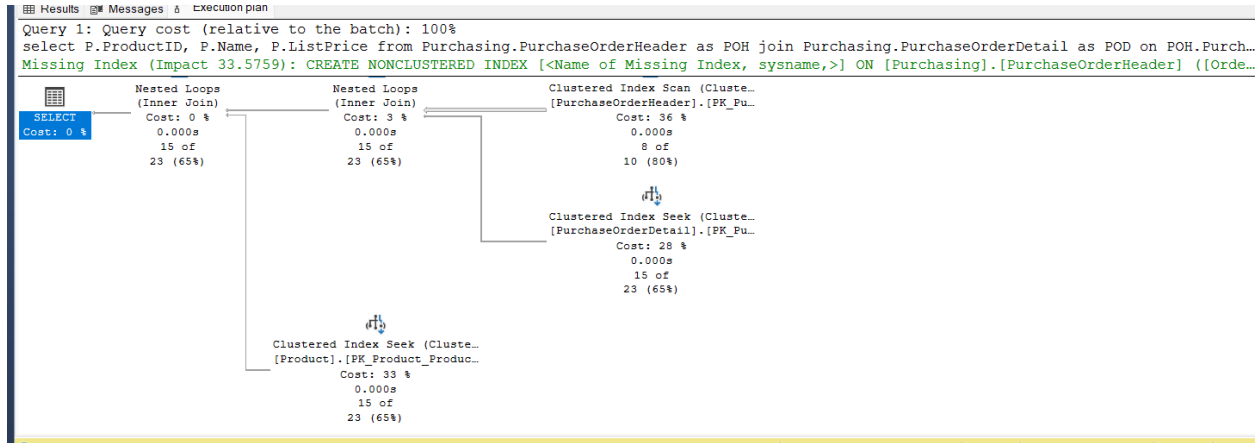
Quantity,
 ProductLevel,
 ProductAssemblyID,
 Sort
 FROM cte_BOM
 ORDER BY Sort;

Ta được execution plan như sau:

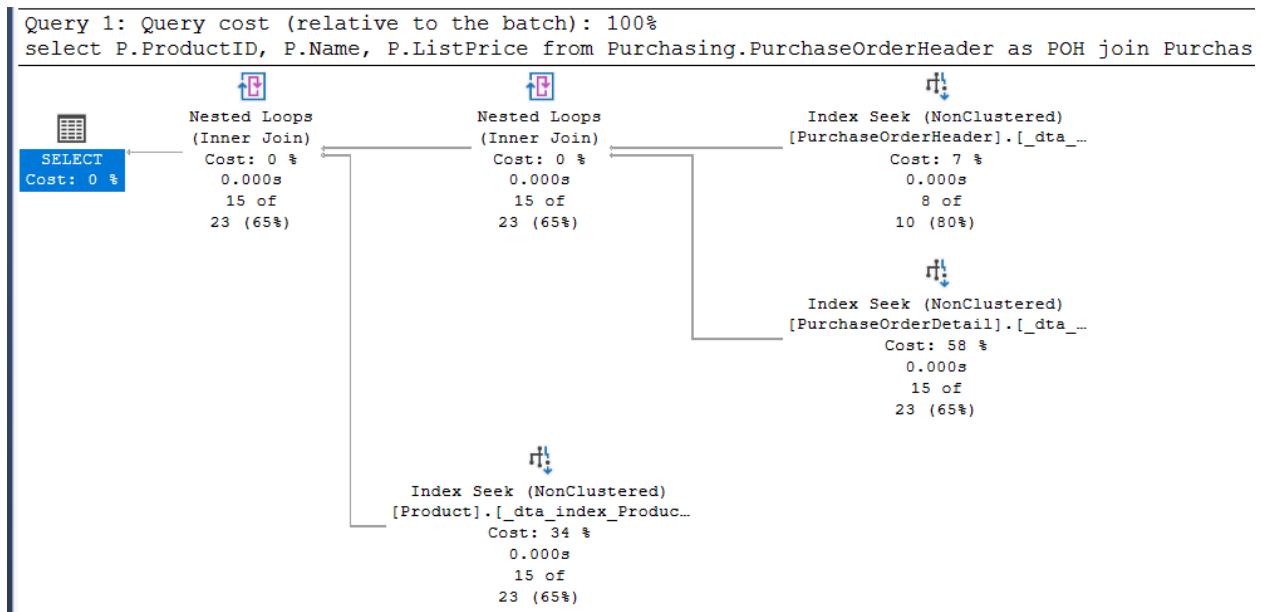


II. Đề xuất chỉ mục phù hợp làm tăng hiệu quả truy vấn. Quan sát và giải thích Execution plan trước và sau khi đề xuất giải pháp

1. Query 1



Sau khi tạo Index



Bảng [Product]

Number of Rows Read	15
Actual Number of Rows for All Executions	15
Actual Number of Batches	0
Estimated Operator Cost	0.0348753 (33%)
Estimated I/O Cost	0.003125
Estimated Subtree Cost	0.0348753
Estimated CPU Cost	0.0001581
Estimated Number of Executions	22.8629371
Number of Executions	15
Estimated Number of Rows for All Executions	22.8629371
Estimated Number of Rows to be Read	1
Estimated Number of Rows Per Execution	1
Estimated Row Size	73 B

Number of Rows Read	15
Actual Number of Rows for All Executions	15
Actual Number of Batches	0
Estimated Operator Cost	0.0149372 (34%)
Estimated I/O Cost	0.003125
Estimated Subtree Cost	0.0149372
Estimated CPU Cost	0.0001581
Estimated Number of Executions	22.8629371
Number of Executions	15
Estimated Number of Rows for All Executions	22.8629371
Estimated Number of Rows to be Read	1
Estimated Number of Rows Per Execution	1
Estimated Row Size	73 B

- ⇒ Từ 2 bảng so sánh, các chỉ số Cost sau khi cài Index đều giảm đi đáng kể
 - ⇒ Với câu query không có index thì câu query sẽ đi hết table để quét (**TABLE SCAN**) và sẽ tốn rất nhiều thời gian
- Khi thực hiện câu lệnh dưới đây, SQL Server duyệt qua chỉ mục (**Index Seek**) để xác định vị trí bản ghi, cách này thì nhanh hơn quét toàn bộ bảng.

Bảng [PurchaseOrderDetail]

Number of Rows Read	15
Actual Number of Rows for All Executions	15
Actual Number of Batches	0
Estimated Operator Cost	0.0290925 (28%)
Estimated I/O Cost	0.003125
Estimated Subtree Cost	0.0290925
Estimated CPU Cost	0.0001594
Estimated Number of Executions	10.37037
Number of Executions	8
Estimated Number of Rows for All Executions	22.8629325
Estimated Number of Rows to be Read	2.20464
Estimated Number of Rows Per Execution	2.20464
Estimated Row Size	11 B

Number of Rows Read	15
Actual Number of Rows for All Executions	15
Actual Number of Batches	0
Estimated Operator Cost	0.025504 (58%)
Estimated I/O Cost	0.003125
Estimated Subtree Cost	0.025504
Estimated CPU Cost	0.0001594
Estimated Number of Executions	10.37037
Number of Executions	8
Estimated Number of Rows for All Executions	22.8629325
Estimated Number of Rows to be Read	2.20464
Estimated Number of Rows Per Execution	2.20464
Estimated Row Size	11 B

- ⇒ Từ 2 bảng so sánh, các chỉ số Cost sau khi cài Index đều giảm đi đáng kể
 - ⇒ Với câu query không có index thì câu query sẽ đi hết table để quét (**TABLE SCAN**) và sẽ tốn rất nhiều thời gian
- Khi thực hiện câu lệnh dưới đây, SQL Server duyệt qua chỉ mục (**Index Seek**) để xác định vị trí bản ghi, cách này thì nhanh hơn quét toàn bộ bảng.

Bảng [PurchaseOrderHeader]

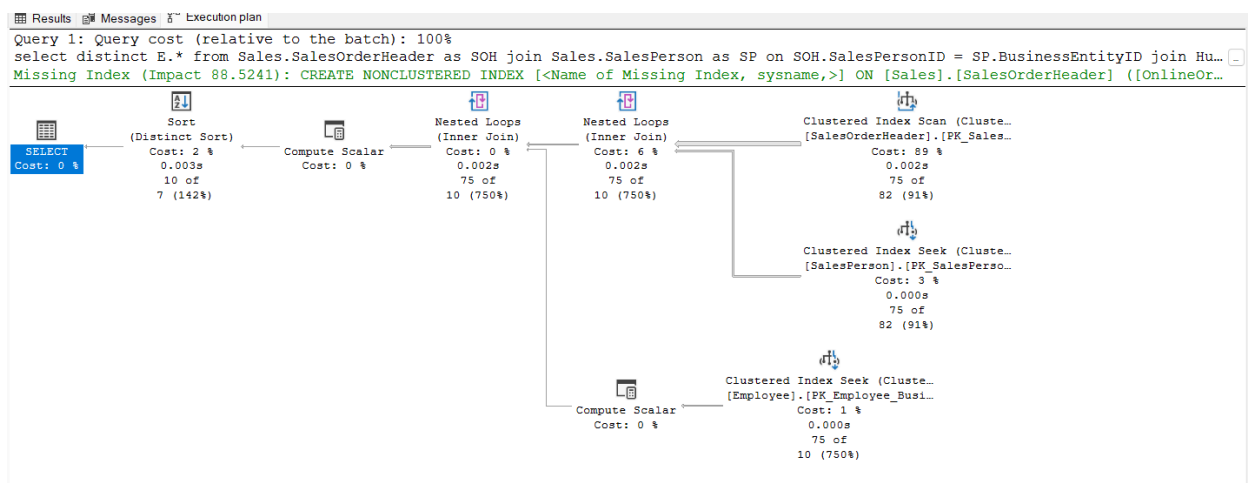
Number of Rows Read	4012
Actual Number of Rows for All Executions	8
Actual Number of Batches	0
Estimated I/O Cost	0.0334954
Estimated Operator Cost	0.0380656 (36%)
Estimated Subtree Cost	0.0380656
Estimated CPU Cost	0.0045702
Estimated Number of Executions	1
Number of Executions	1
Estimated Number of Rows for All Executions	10.3704
Estimated Number of Rows Per Execution	10.3704
Estimated Number of Rows to be Read	4012
Estimated Row Size	19 B

Number of Rows Read	8
Actual Number of Rows for All Executions	8
Actual Number of Batches	0
Estimated Operator Cost	0.0032934 (7%)
Estimated I/O Cost	0.003125
Estimated Subtree Cost	0.0032934
Estimated CPU Cost	0.0001684
Estimated Number of Executions	1
Number of Executions	1
Estimated Number of Rows for All Executions	10.3704
Estimated Number of Rows to be Read	10.3704
Estimated Number of Rows Per Execution	10.3704
Estimated Row Size	11 B

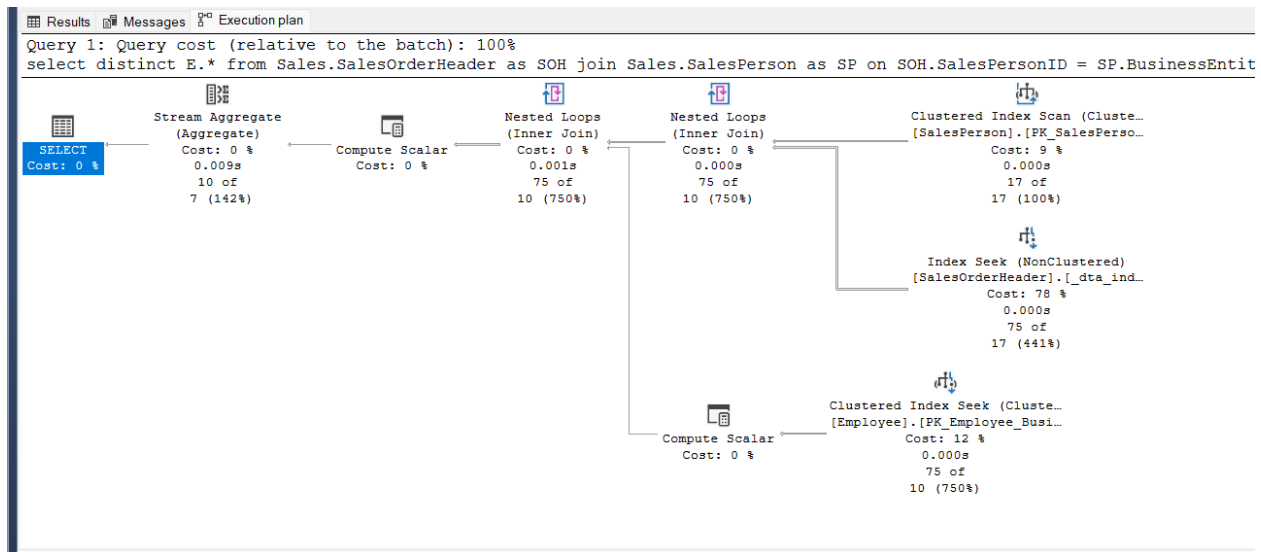
- ⇒ Số Number of Rows Read giảm từ 4012 xuống 8 tốn ít tài nguyên và thời gian hơn
 - ⇒ Từ 2 bảng so sánh, các chỉ số Cost sau khi cài Index đều giảm đi đáng kể
 - ⇒ Với câu query không có index thì câu query sẽ đi hết table để quét (**TABLE SCAN**) và sẽ tốn rất nhiều thời gian
- Khi thực hiện câu lệnh dưới đây, SQL Server duyệt qua chỉ mục (**Index Seek**) để xác định vị trí bản ghi, cách này thì nhanh hơn quét toàn bộ bảng.

2. Query 2:

Trước khi cài Index



Sau khi cài Index

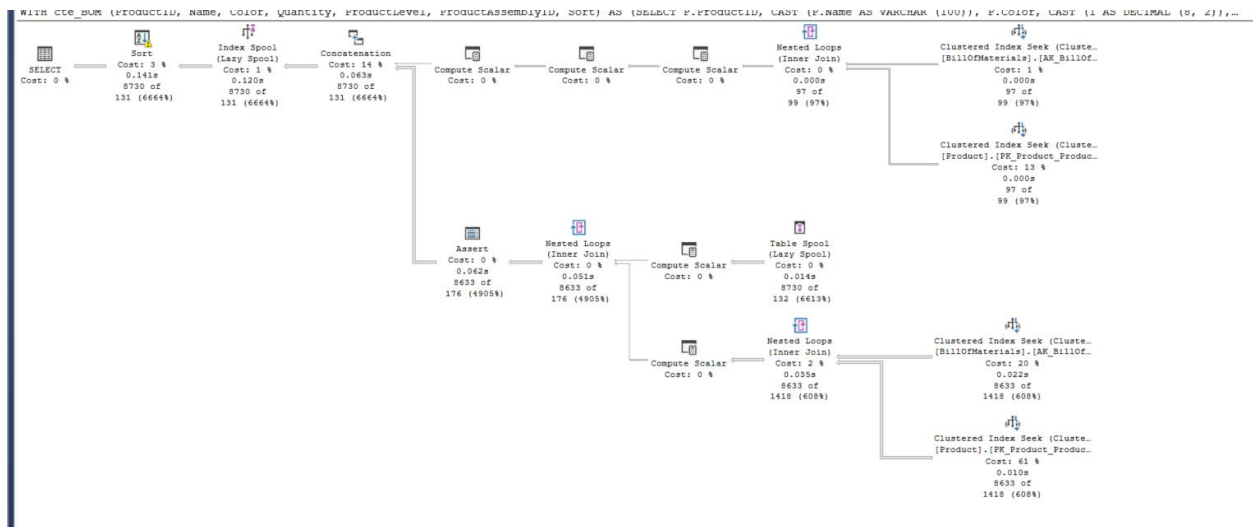


So sánh thông số tại bảng [SalesOrderHeader]

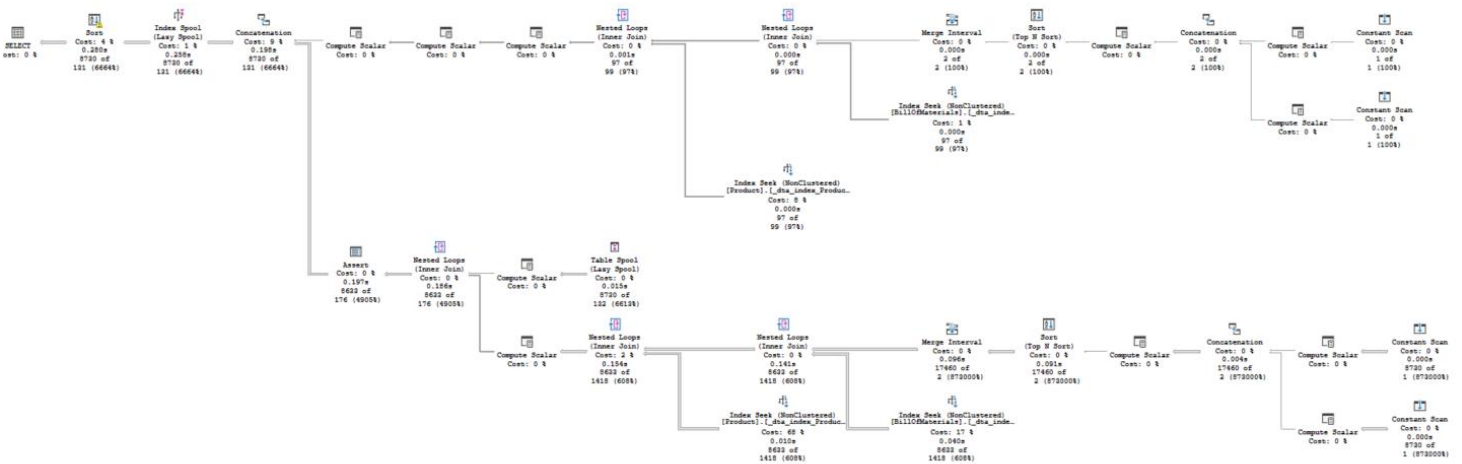
Number of Rows Read	31465	Number of Rows Read	17
Actual Number of Rows for All Executions	75	Actual Number of Rows for All Executions	17
Actual Number of Batches	0	Actual Number of Batches	0
Estimated I/O Cost	0.509792	Estimated I/O Cost	0.003125
Estimated Operator Cost	0.54456 (89%)	Estimated Operator Cost	0.0033007 (9%)
Estimated Subtree Cost	0.54456	Estimated CPU Cost	0.0001757
Estimated CPU Cost	0.0347685	Estimated Subtree Cost	0.0033007
Estimated Number of Executions	1	Number of Executions	1
Number of Executions	1	Estimated Number of Executions	1
Estimated Number of Rows for All Executions	81.824	Estimated Number of Rows for All Executions	17
Estimated Number of Rows Per Execution	81.824	Estimated Number of Rows Per Execution	17
Estimated Number of Rows to be Read	31465	Estimated Number of Rows to be Read	17
Estimated Row Size	20 B	Estimated Row Size	11 B

- ⇒ Số Number of Rows Read giảm từ 31465 xuống 17 tốn ít tài nguyên và thời gian hơn
 - ⇒ Với truy vấn minh họa, kết quả trả về có 17 dòng thỏa điều kiện, và khi dùng index chỉ trở tới 17 dòng này (**Actual Number of Rows for All Executions**), mà không cần duyệt tất cả dòng như khi không dùng index (kiểm tra 17 dòng).
 - ⇒ Các chỉ số Cost sau khi cài Index đều giảm đi đáng kể
 - ⇒ Với câu query không có index thì câu query sẽ đi hết table để quét (TABLE SCAN) và sẽ tốn rất nhiều thời gian
- Khi thực hiện câu lệnh dưới đây, SQL Server duyệt qua chỉ mục (Index Seek) để xác định vị trí bản ghi, cách này thì nhanh hơn quét toàn bộ bảng.

Trước khi cài Index



Sau khi cài Index



So sánh thông số tại bảng [BillOfMaterials]

Number of Rows Read	103
Actual Number of Rows for All Executions	97
Actual Number of Batches	0
Estimated Operator Cost	0.0033953 (1%)
Estimated I/O Cost	0.003125
Estimated Subtree Cost	0.0033953
Estimated CPU Cost	0.0002703
Estimated Number of Executions	1
Number of Executions	1
Estimated Number of Rows for All Executions	99.1007
Estimated Number of Rows to be Read	103
Estimated Number of Rows Per Execution	99.1007
Estimated Row Size	19 B

Number of Rows Read	97
Actual Number of Rows for All Executions	97
Actual Number of Batches	0
Estimated Operator Cost	0.003391 (1%)
Estimated I/O Cost	0.003125
Estimated Subtree Cost	0.003391
Estimated CPU Cost	0.000266
Estimated Number of Executions	1
Number of Executions	2
Estimated Number of Rows for All Executions	99.1007
Estimated Number of Rows to be Read	99.1007
Estimated Number of Rows Per Execution	99.1007
Estimated Row Size	11 B

- ⇒ Với truy vấn minh họa, kết quả trả về có 97 dòng thỏa điều kiện, và khi dùng index chỉ trở tới 97 dòng này (**Actual Number of Rows for All Executions**), mà không cần duyệt tất cả dòng như khi không dùng index (kiểm tra 97 dòng).
- ⇒ Dễ thấy chi phí ước tính khi sử dụng index đem lại hiệu quả tốt (**Estimated Operator Cost / CPU Cost / Subtree Cost**).
- ⇒ Ghi gán index câu truy vấn sẽ thực hiện được nhanh hơn

BÁO CÁO

III. Thành viên và bảng phân công:

Họ và tên	Username Github	Phân công	Mức độ hoàn thành
Võ Minh Tuấn 19126068	votuan8425	2a, 2b và tối ưu hóa 2a, 2b	Mức độ hoàn thành: 100%
Nguyễn Hoàng Hùng 19126047	HungNguyen73	3a và 3b	Mức độ hoàn thành: 100%
Phạm Dung Thông 19126066	ThongP	2c và tối ưu hóa 2c	Mức độ hoàn thành 100%

1. Kế hoạch làm việc:

- Mỗi cuối tuần thứ 7 sẽ họp nhóm một lần và giao công việc tiếp theo cho tuần mới
- Mỗi thành viên sẽ cố gắng hoàn thành công việc được giao trong tuần để cuối tuần báo cáo đầy đủ
- Nếu có thành viên không hoàn thành công việc được giao trong tuần thì sẽ báo cáo và được các thành viên trong nhóm phụ giải quyết vấn đề mà bạn gặp phải

2. Source code

- Link source code Github: [votuan8425/DOAN2-CSDLNC \(github.com\)](https://github.com/votuan8425/DOAN2-CSDLNC)