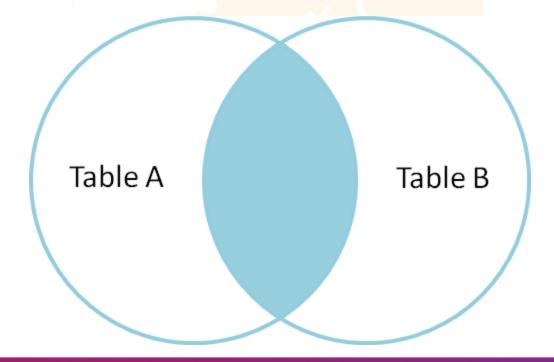


TRƯỜNG ĐÀO TẠO LẬP TRÌNH VIÊN VÀ QUẢN TRỊ MẠNG QUỐC TẾ BACHKHOA-APTECH

Bài 09 Truy vấn nâng cao và các phép kết nối (JOIN)

Tóm tắt

- Giới thiệu
- Nhóm và thống kê dữ liệu
- Truy vấn con
- Phép ghép nối (join)



Giới thiệu

- SQL Server 2014 đưa vào nhiều tính năng truy vấn mạnh mẽ giúp bạn có thể lấy dữ liệu nhanh chóng và hiệu quả.
- Dữ liệu có thể được nhóm và/hoặc tổng hợp (aggregated) cùng nhau để thể hiện tông tin tổng kết.
- Sử dụng khái niệm truy vấn con, tập kết quả của một câu lệnh SELECT có được sử dụng làm điều kiện cho câu lệnh SELECT hoặc truy vấn khác.
- Ghép nối (Joins) giúp bạn kết hợp các cột dữ liệu từ hai hay nhiều bảng dựa trên mối quan hệ giữa các bảng.



<u>Nhóm dữ liệu</u>:

- Mệnh đề GROUP BY phân vùng (partitions) tập kết quả thành một hoặc nhiều tập con. Mỗi tập kết quả có các giá trị và biểu thức chung.
- Từ khóa GROUP BY, theo sau nó là danh sách các cột được dùng để nhóm. Mỗi cột được nhóm giới hạn số lượng dòng của tập kết quả.
- Mỗi cột nhóm, chỉ có một dòng. Mệnh đề GROUP
 BY có thể nhiều hơn một cột nhóm.

Cú pháp:

```
SELECT <select list>
FROM 
WHERE <search condition>
GROUP BY <group by list>
```

Trong đó:

<group by list>: danh sách các cột được chọn để
nhóm các dòng.

Ví dụ: khách hàng có thể **mua nhiều sản phẩm**, truy vấn sau tới bảng dữ liệu hóa đơn sẽ **thống kê số sản phẩm** bằng lệnh GROUP BY

Code:

```
-- Truy vấn lấy về tổng số sản phẩm đã bán được NHÓM
theo KHÁCH HÀNG mua

SELECT tenkh, COUNT(idsp) AS 'Tổng số sản phẩm mua'
FROM tblBanHang, tblKhachHang
WHERE tblBanHang.idkh = tblKhachHang.idkh
GROUP BY tblKhachHang.tenkh -- nhóm kết quả truy vấn
theo tenkh
```



Kết quả: COUNT(idsp) AS 'Tổng số sản phẩm mua' Results Messages Tổng số sản phẩm mua tenkh Lukaku Messi Ronaldo

GROUP BY tblKhachHang.tenkh



GROUP BY với WHERE:

- Có thể sử dụng mệnh đề WHERE cùng với mệnh đề GROUP BY để giới hạn số dòng trước khi thực hiện nhóm.
- Các dòng thỏa mãn điều kiện sẽ được xem xét cho viêc nhóm.
- Các dòng dữ liệu không phù hợp với điều kiện trong mệnh đề WHERE sẽ bị loại bỏ trước khi công việc nhóm được thực hiện.

Ví dụ: vẫn với yêu cầu trên, giờ bổ sung thêm chỉ lọc giá bán sản phẩm có **giá lớn hơn \$200**.

Code:

```
-- Truy vấn lấy về tổng số sản phẩm có giá > $200 đã bán được NHÓM theo KHÁCH HÀNG mua

SELECT tenkh, COUNT(idsp) AS 'Tổng số sản phẩm mua'

FROM tblBanHang, tblKhachHang, tblSanPham

WHERE

tblBanHang.idkh = tblKhachHang.idkh

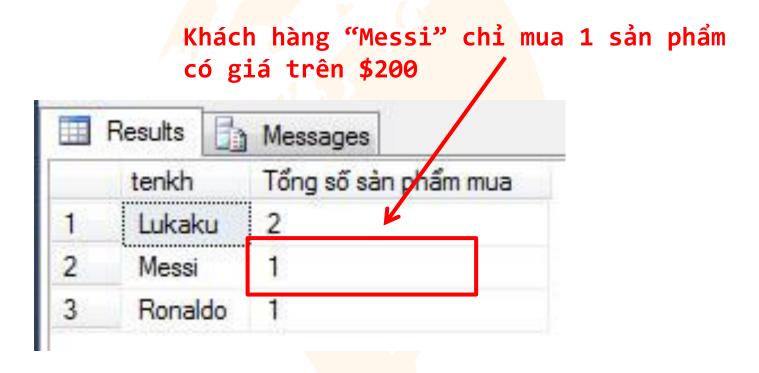
ANDtblBanHang.idsp = tblSanPham.id

AND tblSanPham.giaban > 200

GROUP BY tblKhachHang.tenkh -- nhóm kết quả truy vấn theo tenkh
```



Kết quả:





GROUP BY ALL:

 Có một số tình huống, dữ liệu sau khi bị lọc loại bỏ bởi mệnh đề WHERE vẫn muốn hiển thị trong tập kết quả trả về.

Ví dụ: nếu chỉ thống kê tổng sản phẩm có giá trên \$250 thì ta sẽ chỉ có 1 khách hàng là "Lukaku"





Còn muốn lấy tất cả:

```
-- Truy vấn lấy <mark>về tổng số sản phẩm có gi</mark>á > $250 đã bán
được NHÓM theo KHÁCH HÀNG mua, thống kê cả khách hàng
không thỏa mãn
SELECT tenkh, COUNT(idsp) AS 'Tổng số sản phẩm mua'
FROM tblBanHang, tblKhachHang, tblSanPham WHERE
tblBanHang.idkh = tblKhachHang.idkh
AND tblBanHang.idsp = tblSanPham.id
AND tblSanPham.giaban > 250
GROUP BY ALL tblKhachHang.tenkh -- nhóm kết quả truy vấn
theo tenkh
```



Kết quả:

Khách hàng mua sản phẩm không thỏa mãn điều kiện vẫn thống kê

Results 🛅		Messages
	tenkh	Tổng số sản phẩm mua
1	Lukaku	1
2	Messi	0
3	Ronaldo	0

GROUP BY với HAVING:

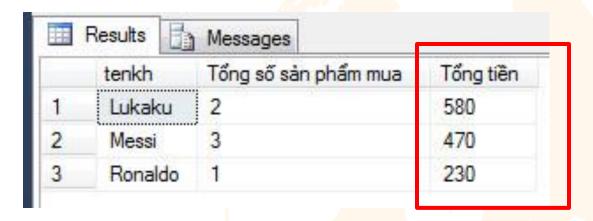
- Mệnh đề HAVING chỉ được sử dụng cùng với mệnh đề GROUP BY trong câu lệnh SELECT để chỉ ra điều kiện lọc dữ liệu sau khi đã tính toán theo nhóm.
- Mệnh đề HAVING có vai trò như mệnh đề WHERE, nhưng nó được dùng ở những nơi mà mệnh đề WHERE không thể sử dụng được với các hàm tổng hợp như SUM ().
- Sau khi bạn đã tạo ra các nhóm bằng mệnh đề GROUP BY, và bạn muốn tiếp tục lọc các nhóm kết quả nữa.
- Mệnh đề HAVING đóng vai trò như bộ lọc trên các nhóm, tương tự cách mệnh đề WHERE lọc trên các dòng được trả về bởi mệnh đề FROM.

Ví dụ: thống kê tổng số tiền khách hàng đã chi và chỉ lọc lấy khách hàng có **tổng tiền lớn hơn \$500**.

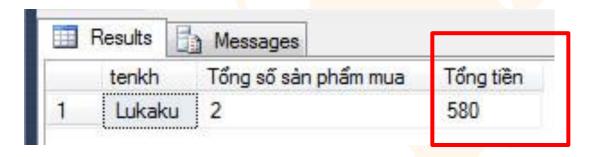
Code:

```
-- Truy vấn lấy về tổng số sản phẩm đã bán được
-- NHÓM theo KHÁCH HÀNG mua và tổng số tiền mà khách hàng đã chi
-- và lấy về KHÁCH HÀNG VIP có tổng số tiền lớn hơn 500$
SELECT tenkh, COUNT(idsp) AS 'Tổng số sản phẩm mua', SUM(giaban) AS
'Tổng tiền' FROM tblBanHang, tblKhachHang, tblSanPham
WHERE
tblBanHang.idkh = tblKhachHang.idkh AND
tblBanHang.idsp = tblSanPham.id
GROUP BY tblKhachHang.tenkh -- nhóm kết quả truy vấn theo tenkh
HAVING SUM(giaban) > 500
```

Kết quả:



Tổng chi phí của tất cả khách hàng



Loc khách hàng
có tổng hóa đơn
> \$500

Hàm thống kê:

- Các hàm thống kê thực hiện tính toán trên tập giá trị và trả về giá trị duy nhất
- Các hàm thống kê đều bỏ qua giá trị NULL, nhưng trừ hàm COUNT(*).
- Các hàm thống kê thường được sử dụng với mệnh đề GROUP BY của câu lệnh SELECT.
- Khi sử dụng hàm thóng kê trong phần danh sách cột của câu lệnh SELECT thường tạo ra một cột không có bí danh (alias), do vậy bạn có thể dùng mệnh đề AS cung cấp cho nó một bí danh.

Hàm thống kê thường dùng:

Tên hàm	Cú pháp	Mô tả
AVG	AVG(<mệnh đề="">)</mệnh>	Tính giá trị trung bình của dữ liệu số trên các dòng của cột chỉ định.
COUNT	COUNT(*)	Đếm tổng số dòng dữ liệu.
MAX	MAX(<mệnh đề="">)</mệnh>	Trả về giá trị số lớn nhất hoặc ngày gần nhất.
MIN	MIN(<mệnh đề="">)</mệnh>	Trả về giá trị số bé nhất hoặc ngày xa nhất.
SUM	SUM(<mệnh đề="">)</mệnh>	Tín <mark>h tổng giá tr</mark> ị của tất cả các dòng dữ liệu tron <mark>g cột</mark>

Hàm AVG():

SELECT AVG(gia_nhap) AS 'Giá trung bình' FROM

SAN_PHAM

	ma_sp	ten_sp	mo_ta	thong_tin	gia_nhap
1	1	Conserve Chuck 1	fake 1	NULL	600
2	2	Gucci SNN69	mới nhập về	NULL	120
3	3	Adidas	hàng xịn của TQ	NULL	1450
4	4	Nike MG21	nhập khẩu Malay	NULL	690
5	5	Tom 300	đồ đều	NULL	55
6	6	Lacoste 066	nhanh hỏng	NULL	99



Hàm COUNT():

SELECT COUNT(*) AS 'Tổng số sản phẩm' FROM

SAN_PHAM

	ma_sp	ten_sp	mo_ta	thong_tin	gia_nhap
1	1	Conserve Chuck 1	fake 1	NULL	600
2	2	Gucci SNN69	mới nhập về	NULL	120
3	3	Adidas	hàng xịn của TQ	NULL	1450
4	4	Nike MG21	nhập khẩu Malay	NULL	690
5	5	Tom 300	đồ đều	NULL	55
6	6	Lacoste 066	nhanh hòng	NULL	99



Hàm MAX() và MIN():

```
SELECT MAX(gia_nhap) AS 'Giá cao nhất',
```

MIN(gia_nhap) AS 'Giá thấp nhất' FROM SAN_PHAM

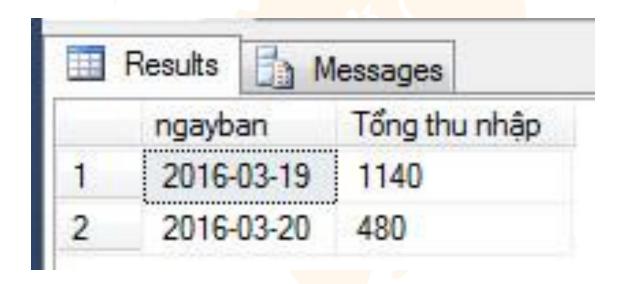
	ma_sp	ten_sp	mo_ta	thong_tin	gia_nhap
1	1	Conserve Chuck 1	fake 1	NULL	600
2	2	Gucci SNN69	mới nhập về	NULL	120
3	3	Adidas	hàng xịn của TQ	NULL	1450
4	4	Nike MG21	nhập khẩu Malay	NULL	690
5	5	Tom 300	đồ đều	NULL	55
6	6	Lacoste 066	nhanh hòng	NULL	99

Giá cao nhất	Giá thấp nhất	
1450	55	
	1450	



<u>Hàm SUM()</u>:

```
SELECT ngayban, SUM(giatien) AS 'Tổng thu nhập' FROM ThongKeBanHang GROUP BY ngayban
```





- Có thể sử dụng một câu lệnh SELECT hoạc một truy vấn trả về các bản ghi được để làm điều kiện cho câu truy lệnh SELECT hoặc truy vấn khác.
- Truy vấn lồng (inner query) còn được gọi là truy vấn cha, và truy vấn bên trong được gọi là truy vấn con. Mục đích của truy vấn con là để trả về kết quả cho truy vấn bên ngoài.
- Dạng đơn giản nhất của một truy vấn con là chỉ trả
 về một cột. Truy vấn cha có thể sử dụng kết quả của
 truy vấn con này bằng một dấu =

Ví dụ: lấy về các đơn hàng cuối cùng trong bảng hóa đơn

```
SELECT * FROM ThongKeBanHang
```

```
WHERE ngayban = (SELECT MAX(ngayban) FROM
```

ThongKeBanHang)

	sanpham	giatien	ngayban
1	Samsung Trend	120	2016-03-20
2	Nokia 1200	120	2016-03-20
3	Samsung Trend	120	2016-03-20
4	Nokia 1200	120	2016-03-20

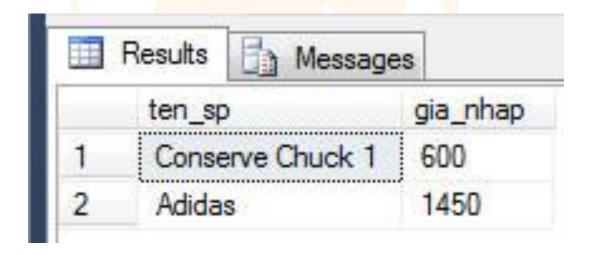


Truy vấn đa trị:

- Nếu toán tử = được sử dụng với truy vấn con, truy vấn con đó chỉ được trả về một giá trị vô hướng duy nhất.
- Nếu trả về nhiều hơn một giá trị, truy vấn sẽ bị lỗi và không được thực thi.
- Các từ khóa ANY, ALL, IN, và EXISTS có thể được sử dụng trong mệnh đề WHERE của câu lệnh SELECT, khi đó truy vấn trả về một cột hay có nhiều dòng.
- Những từ khóa này còn được gọi là các vị từ (predicates), được sử dụng với các truy vấn đa trị.

Ví dụ: lấy về tên và giá sản phẩm mà mã loại sản phẩm là 1.

```
SELECT ten_sp, gia_nhap FROM SAN_PHAM WHERE
ma_sp IN (SELECT ma_sp FROM SAN_PHAM WHERE
ma_lsp = 1)
```





Các phép ghép nối được sử dụng để **lấy dữ liệu từ hai hoặc nhiều bảng dựa trên mối quan hệ luận lý** (logical relationship) giữa các bảng. Nó định nghĩa cách thức hai bảng được quan hệ với nhau bằng cách:

- Chỉ ra cột từ mỗi bảng được sử dụng cho việc ghép nối.
 Một ghép nối đặc thù chỉ ra khóa ngoại từ một bảng kết hợp (associated) với khóa chính từ một bảng khác.
- Toán tử luận lý =, <> được sử dụng được sử dụng để so sánh giá trị từ các cột.

Cú pháp:

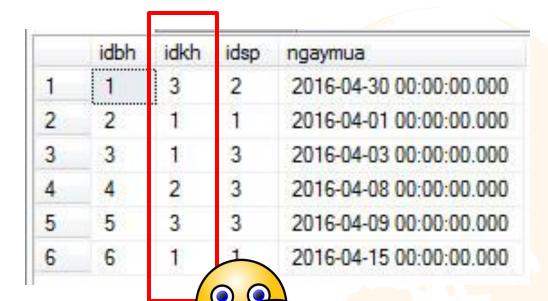
```
SELECT <ColumnName1>, <ColumnName2>...<ColumnNameN>
FROM Table_A AS Table_Alias_A
JOIN
Table_B AS Table_Alias_B
ON
Table_Alias_A.<CommonColumn> = Table_Alias_B.<CommonColumn>
```

Trong đó:

- <ColumnName1>, <ColumnName2>: là một danh sách cột được hiển thị
- Table_A: là tên của bảng phía bên trái của từ khóa
 JOIN.
- Table_B: là tên của bảng phía bên phải của từ khóa
 JOIN.



- AS Table_Alias: là cách đặt tên bí danh cho bảng. Có thể xác định một bí danh cho bảng trong truy vấn để không cần phải dùng đến tên đầy đủ của nó (Tên đầy đủ của bảng thường dài, nó gồm tên lược đồ với tên bảng).
- <CommonColumn>: là cột chung có trong hai cả bảng tham gia ghép nối. Trong trường hợp đó, ghép nói chỉ thành công nếu các cột có các giá trị khớp nhau



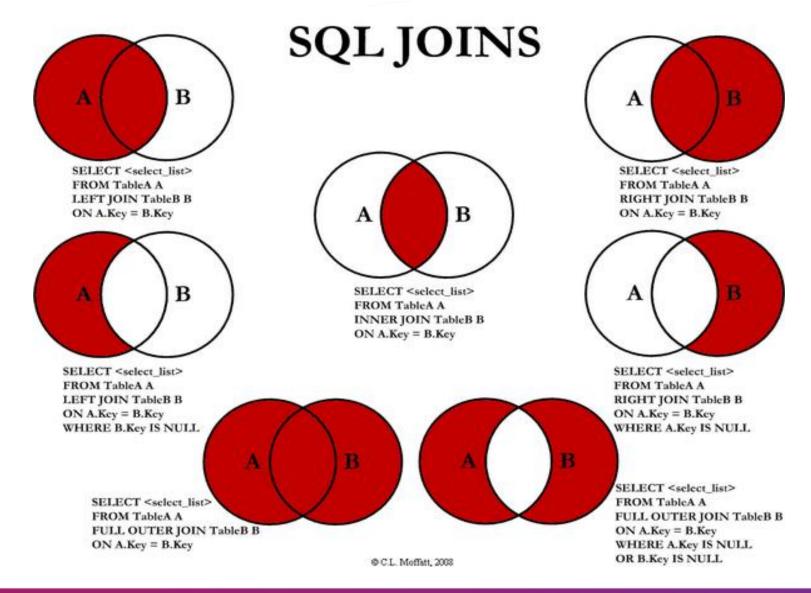
Bảng tblBanHang

Bảng tblKhachHang

	idkh	tenkh	namsinh
1	1	Messi	1986
2	2	Ronaldo	1983
3	3	Lukaku	1990

SELECT idbh AS 'Mã hóa đơn', tenkh AS 'Tên khách hàng', ngaymua AS 'Ngày mua'FROM tblBanHang bh JOIN tblKhachHang kh ON bh.idkh = kh.idkh Bí danh Điều kiện ghép nối Phép nối Tên khách hàng Mã hóa đơn Ngày mua Lukaku 2016-04-30 00:00:00.000 Messi 2016-04-01 00:00:00 000 Messi 2016-04-03 00:00:00.000 Ronaldo 2016-04-08 00:00:00.000 Lukaku 2016-04-09 00:00:00 000 Messi 2016-04-15 00:00:00.000



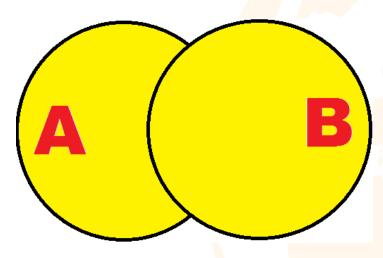


Toán tử UNION:

- Kết quả từ hai câu lệnh truy vấn khác nhau có thể được kết hợp thành một tập kết quả duy nhất bằng toán tử UNION.
- Các câu lệnh truy vấn phải tương thích về kiểu và số lượng các cột.
- Tên các cột trong mỗi câu lệnh có thể khác nhau, nhưng kiểu dữ liệu phải tương thích (giống nhau).

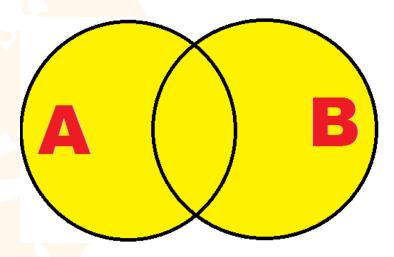
Query_Statement1 UNION [ALL] Query_Statement2

UNION



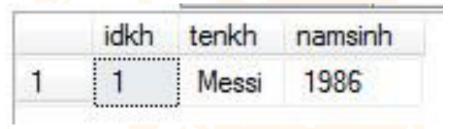
UNION sẽ hội kết quả của 2 câu truy vấn sau khi LOẠI BỞ kết quả trùng lặp

UNION ALL



UNION ALL sẽ KHÔNG LOẠI BỎ kết quả trùng lặp

(Lệnh A) SELECT * FROM tblKhachHang WHERE tenkh LIKE 'M%'



(Lênh B) SELECT * FROM tblKhachHang WHERE

namsinh >= 1986

	idkh	tenkh	namsinh
1	1	Messi	1986
2	3	Lukaku	1990



```
SELECT * FROM tblKhachHang WHERE tenkh LIKE 'M%'
UNION
```

SELECT * FROM tblKhachHang WHERE namsinh >= 1986

	idkh	tenkh	namsinh
1	1	Messi	1986
2	3	Lukaku	1990



SELECT * FROM tblKhachHang WHERE tenkh LIKE 'M%' UNION ALL

SELECT * FROM tblKhachHang WHERE namsinh >= 1986

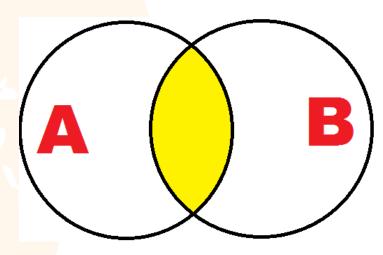
	idkh	tenkh	namsinh
1	1	Messi	1986
2	1	Messi	1986
3	3	Lukaku	1990



Toán tử INTERSERT: được sử dụng với hai câu lệnh truy vấn để trả về tập các dòng chung duy nhất của cả hai câu lệnh.

Cú pháp:

Query_statement1 INTERSECT Query_statement2



Nguyên tắc:

- Số lượng cột và thứ tự của các cột được chỉ ra trong hai lệnh truy vấn phải giống nhau.
- Kiểu dữ liệu của các cột phải phù hợp.



```
SELECT * FROM tblBanHang
```

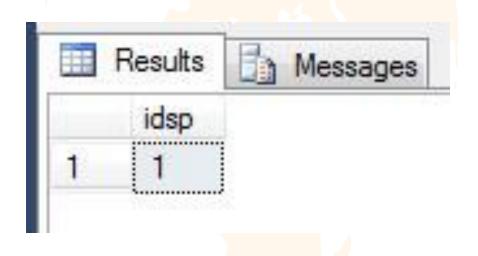
SELECT * FROM tblSanPham

	idbh	idkh	idsp	nga	ymua
1	1	3	2	20	16-04-30 00:00:00.000
2	2	1	1	20	16-04-01 00:00:00.000
3	3	1	3	20	16-04-03 00:00:00.000
4	4	2	3	20	16-04-08 00:00:00.000
5	5	3	3	20	16-04-09 00:00:00.000
6	6	1	1	20	16-04-15 00:00:00.000
	id t	ensp	g	iaban	mota
1	1	Nokia	1	120	Sản phẩm phổ thông
2	2	Samsun	g :	350	Cao cấp
3	3	LG	1	230	Rê tiền



SELECT idsp FROM tblBanHang WHERE ngaymua > '2016-04-05'
INTERSECT

SELECT id FROM tblSanPham WHERE giaban < 200





Toán tử PIVOT: được sử dụng với mục đích đảo dữ liệu hàng thành cột => dễ nhìn - dễ hiểu.

Ví dụ: trong bảng dữ liệu bán hàng, mã hóa đơn, mã nhân viên và hình thức thanh toán lưu trong mỗi dòng bản ghi.

	id	sanpham	giatien	ngayban	ma_nv	thanhtoan
1	1	Nokia 1200	100	2016-03-19	1	Tiền mặt
2	2	Samsung Trend	120	2016-03-19	2	Visa
3	3	HTC One	50	2016-03-19	1	Tiền mặt
4	4	HTC One	50	2016-03-19	2	Chuyển khoản
5	5	HTC One M8	100	2016-03-19	3	Tiền mặt
6	6	HTC One M9	150	2016-03-19	1	Tiền mặt
7	7	Samsung Trend	120	2016-03-20	3	Tiền mặt
8	8	Nokia 1200	120	2016-03-20	1	Chuyển khoản
9	9	Nokia 1200	100	2016-03-19	3	Tiền mặt
10	10	Samsung Trend	120	2016-03-19	3	Chuyển khoản
11	11	HTC One	50	2016-03-19	2	Tiền mặt
12	12	HTC One	50	2016-03-19	1	3
13	13	HTC One M8	100	2016-03-19	3	Tiền mặt
14	14	HTC One M9	150	2016-03-19	1	Visa
15	15	Samsung Trend	120	2016-03-20	3	Visa
16	16	Nokia 1200	120	2016-03-20	1	Visa

Lệnh sau chỉ lấy về mã hóa đơn, mã nhân viên và hình thức thanh toán:

SELECT id, ma_nv, thanhtoan FROM ThongKeBanHang

	id	ma_nv	thanhtoan
1	1	1	Tiền mặt
2	2	2	Visa
3	3	1	Tiền mặt
4	4	2	Chuyển khoản
5	5	3	Tiền mặt
6	6	1	Tiền mặt
7	7	3	Tiền mặt
8	8	1	Chuyển khoản
9	9	3	Tiền mặt
10	10	3	Chuyển khoản



Yêu cầu nghiệp vụ là: thống kê mỗi nhân viên đã thực hiện bao nhiều hình thức thanh toán. Ví du: nhân viên A - tiền mặt (3 lần) | chuyển khoản (6 lần).... Code: -- PIVOT SELECT ma_nv, [Tiền mặt], [Visa], [Chuyển khoản] FROM (SELECT id, ma_nv, thanhtoan FROM ThongKeBanHang) BangNguon PIVOT (COUNT(id) FOR thanhtoan IN ([Tiền mặt], [Visa], [Chuyển khoản])) BangDao

Kết quả:

- Thống kê được hình thức thanh toán tương ứng với nhân viên
- Hình thức thanh toán chuyển sang cột => dễ nhìn

	ma_nv	Tiền mặt	Visa	Chuyển khoản
1	1	3	2	1
2	2	1	1	1
3	3	4	1	1



Tóm tắt bài học

- Mệnh đề GROUP BY và cá hàm thống kê cho phép tổng hợp dữ liệu để trình bày thông tin tổng hợp.
- Truy vấn con cho phép tập kết quả của một câu lệnh SELECT này sử dụng để làm điều kiện cho một câu lệnh SELECT khác.
- Việc ghép nối (Joins) giúp bạn kết hợp cá cột dữ liệu từ hai hay nhiều bảng dựa trên mối quan hệ logic giữa các bảng.
- Các toán tử tập hợp giúp bạn kết hợp/gộp các dòng dữ liệu từ từ hai hay nhiều bảng



TRƯỜNG ĐÀO TẠO LẬP TRÌNH VIÊN VÀ QUẨN TRỊ MẠNG QUỐC TẾ BACHKHOA-APTECH

Thank for watching!

