DB Module

NGEE ANN SCHOOL OF INFOCOMM TECHNOLOGY

SQL – Group By and SubQueries

Databases

Day 3

Prepared by Staff of School of ICT

Topics

- *SELECT* Statements with
  - *GROUP BY*
  - *HAVING*
- *SubQueries*

GROUP BY

How to list the number of staff for each branch?

```
SELECT BranchNo,
            COUNT(StaffID) AS “No. of Staff”
FROM Staff
GROUP BY BranchNo
```

- GROUP BY enables subtotal for subgroup

Step 1 : FROM

- Identify table

Example:

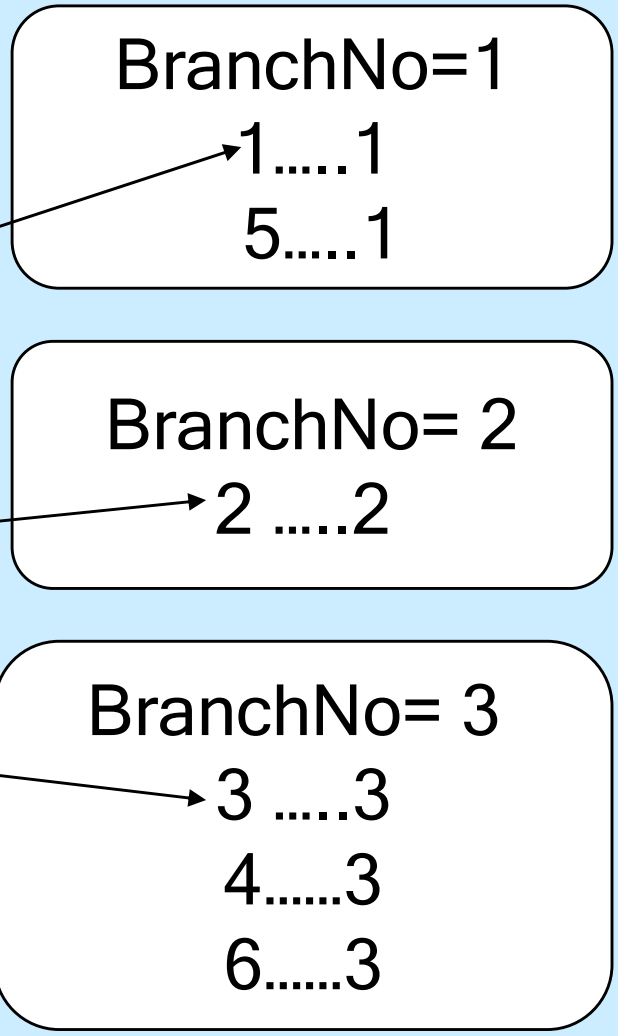
FROM Staff

Step 2: GROUP BY BranchNo

- Grouping column ?
- Grouping values ?

Rows with same value for grouping column are grouped together.

StaffID	….	BranchNo
1		1
2		2
3		3
4		3
5		1
6		3



STEP 3 : SELECT BranchNo, COUNT(StaffID)

For each group:

- Pick up column
  - *BranchNo*
- Apply aggregate function
  - *COUNT(StaffID)*

BranchNo	Count(StaffID)
1	?
2	?
3	?

BranchNo=1
1…..1
5…..1

BranchNo= 2
2 …..2

BranchNo= 3
3 …..3
4……3
6……3

GROUP BY

```
SELECT BranchNo, COUNT(StaffID) AS "No. of Staff"
FROM Staff
GROUP BY BranchNo
```

Results :

BranchNo	No. of Staff
1	2
2	1
3	3

- Can BranchNo be a SELECT column if it is not a GROUP BY column?

GROUP BY & HAVING

1. List number of staff for branches that has > 1 staff

```
SELECT BranchNo,
        COUNT(StaffID) AS “No. of Staff”
FROM Staff
GROUP BY BranchNo
HAVING COUNT(StaffID) > 1
```

Result: All branches shown as they each have 3 staff

2. List number of staff for branches that has total salary > 4900

```
SELECT BranchNo,
        COUNT(StaffID) AS “No. of Staff”, SUM(Salary)
FROM Staff
GROUP BY BranchNo
HAVING SUM(Salary) > 4900
```

Result: Only Branch 1 shown

GROUP BY & HAVING

- HAVING Specifies which group in GROUP BY to include in the results

HAVING COUNT(StaffID) > 1

- remove which group from the results?

Results:

BranchNo	No. of Staff
1	2
~~2~~	~~1~~
3	3

BranchNo=1
1…..1
5…..1

BranchNo= 2
2 ….2

BranchNo= 3
3 ….3
4……3
6……3

GROUP BY, HAVING, ORDER BY

List number of staff for branches that has more than 1 staff.

Display the results in *descending* order of number of staff.

```
SELECT BranchNo,
COUNT(StaffID) AS "No. of Staff"
FROM Staff
GROUP BY BranchNo
HAVING COUNT(StaffID) > 1
ORDER BY COUNT(StaffID) DESC
```

Results:

BranchNo	No.of Staff
3	3
1	2

GROUP BY Multiple Columns

List number of staff for each branch according to gender

```
SELECT BranchNo, Gender,
        COUNT(StaffID) AS "No. of Staff"
FROM Staff
GROUP BY BranchNo, Gender
```

BranchNo	Gender	No. of Staff
1	M	1
1	F	1

GROUP BY with WHERE

- List the number of female staff for each branch

```
SELECT BranchNo,
            COUNT(StaffID) AS “No. of Staff”
FROM Staff
WHERE Gender = ‘F’
GROUP BY BranchNo
```

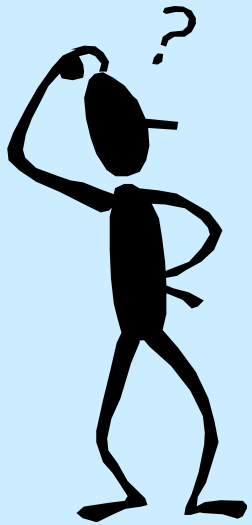
WHERE search condition is applied *before* GROUP BY

Interpreting GROUP BY Statement

SELECT BranchNo, COUNT(StaffID) AS “No. of Staff”
FROM Staff GROUP BY BranchNo

Steps :

1. FROM
   - identify table
2. GROUP BY
   - rows with same value for grouping column are grouped together
3. SELECT
   - Select column, apply aggregate function for each individual group



Subqueries

Using a Subquery with "="

Query 2.43: List the name and salary of staff in the branch in Rose Central.

```
SELECT Name, Salary
FROM Staff
WHERE BranchNo =
   (SELECT BranchNo
   FROM Branch
   WHERE Address LIKE '%Rose Central%')
```

Outer (main) query

Inner (sub) query

How to interpret a nested query?

- Step 1: Identify Outer (main) and Inner (sub) query.
- Step2: Ignore Outer query, execute Inner query first. An intermediate table (in this case consisting of one column and one row i.e. a singleton) is produced, which is then used to replace the Inner query.
- Step 3: Execute Outer query using the result of the Inner query.

Using a Subquery with "="

Query 2.43: List the name and salary of staff in the branch in Rose Central.

```
SELECT Name, Salary
FROM Staff
WHERE BranchNo = 3
   (SELECT BranchNo
   FROM Branch
   WHERE Address LIKE '%Rose Central%');
```

	Name	Salary
1	May May	1990.0000
2	Sadiah	1450.0000
3	Samuel	1250.0000

Returns 3

Using a Subquery with an Aggregate Function

Query 2.44: List the name and salary of those staff whose salary is greater than the average salary of all staff.

```
SELECT Name, Salary
FROM Staff
WHERE Salary > 1616.6666
   (SELECT AVG(Salary)
   FROM Staff)
```

	Name	Salary
1	Mary	1970.0000
2	Nana	2100.0000
3	May May	1990.0000

Returns 1616.6666

Using IN Operator in a Subquery

Query 2.46: List the name and salary of staff in all branches other than the branch in Rose Central.

```
SELECT Name, Salary
FROM Staff
WHERE BranchNo IN (1, 2)
   (SELECT BranchNo
    FROM Branch
    WHERE Address NOT LIKE '%Rose Central%')
```

Returns 1 & 2

	Name	Salary
1	Richard	1500.0000
2	John	1500.0000
3	Mary	1700.0000
4	Sun Sun	1300.0000
5	Jane	1950.0000
6	Nana	2100.0000

Ordering Data with Subqueries

- ORDER BY clause is not allowed in Inner (sub) queries.
- Only ONE ORDER BY clause for a SELECT statement and it must be only in the Outermost (main) query.

Ordering Data with Subqueries

Query 2.46a: List the name and salary of staff in all branches other than the branch in Rose Central. Within each branch, show staff with the highest salary first.

```
SELECT BranchNo, Name, Salary
FROM Staff
WHERE BranchNo IN
   (SELECT BranchNo
    FROM Branch
    WHERE Address NOT LIKE '%Rose Central%')
ORDER BY BranchNo, Salary DESC
```

	BranchNo	Name	Salary
1	1	Mary	1970.0000
2	1	Richard	1500.0000
3	1	John	1500.0000
4	2	Nana	2100.0000
5	2	Jane	1390.0000
6	2	Sun Sun	1300.0000

Using IN Operator in a Subquery

Query 2.46: List the name and salary of staff in all branches other than the branch in Rose Central.

```
SELECT Name, Salary
FROM Staff
WHERE BranchNo IN
    (SELECT BranchNo
     FROM Branch
     WHERE Address NOT LIKE '%Rose Central%')
```

	Name	Salary
1	Richard	1500.0000
2	John	1500.0000
3	Mary	1970.0000
4	Sun Sun	1300.0000
5	Jane	1390.0000
6	Nana	2100.0000

Using Join Vs using Subquery

Query 2.43:

```
SELECT Name, Salary FROM Staff
WHERE BranchNo =
    (SELECT BranchNo
    FROM Branch
    WHERE Address LIKE '%Rose Central%')
```

Query 2.47:

```
SELECT Name, Salary
FROM Staff s INNER JOIN Branch b
ON s.BranchNo = b.BranchNo
WHERE Address LIKE '%Rose Central%'
```

Using Join Vs using Subquery

- A join is necessary when the final table includes data from both tables.
- A subquery is necessary when comparison must be made with an aggregate function applied to the second table.

Join is necessary when...

Query 2.48: List the name and salary of staff and the telephone number of the branch in Rose Central.

```
SELECT Name, Salary, TelNo
FROM Staff s INNER JOIN Branch b
ON s.BranchNo = b.BranchNo
WHERE Address LIKE '%Rose Central%'
```

Subquery is necessary when...

Query 2.44: List the name and salary of those staff whose salary is greater than the average salary of all staff.

```
SELECT Name, Salary
FROM Staff
WHERE Salary >
   (SELECT AVG(Salary)
    FROM Staff)
```

Summary
Interpreting SELECT Statement

Steps :

1. FROM - identify table(s)
2. WHERE - retain rows that satisfy search condition(s)
3. GROUP BY - rows with same value(s) of grouping column(s) are grouped together
4. HAVING -retain group(s) that satisfy search condition(s)
5. SELECT - specify column(s) for output
6. ORDER BY - sort results for display ins ascending or descending order