SQL Queries: GROUP BY and HAVING

- **Aggregate functions and queries**
- **More clauses in the SELECT statement**
  - GROUP By Clause
  - HAVING Clause

Aggregate Function

- **A function that generates a single value from a group of values**

  - often used with Group By and Having clauses

  - a.k.a. set function

- **Examples:**

  - Avg, Count, Max, Min, and Sum

Examples of Aggregate Functions

Aggregate function	Description
AVG(*expr*)	Average of the values in a column. The column can contain only numeric data.
COUNT(*expr*), COUNT(*)	A count of the values in a column (if you specify a column name as *expr*) or of all rows in a table or group (if you specify *). COUNT(*expr*) ignores null values, but COUNT(*) includes them in the count.
MAX(*expr*)	Highest value in a column (last value alphabetically for text data types). Ignores null values.
MIN(*expr*)	Lowest value in a column (first value alphabetically for text data types). Ignores null values.
SUM(*expr*)	Total of values in a column. The column can contain only numeric data.

Aggregate Query

- **A query (SQL statement) that summarizes information from multiple rows by including an aggregate function such as Sum or Avg**

    - For example, you can create a query that averages the contents of a price column

        ```
        SELECT Avg(Price) AS AvgPrice
        FROM Book
        ```

Aggregate Queries (cont.)

- **Aggregate queries can also display subtotal information by creating groups of rows that have data in common**
  - An example would be a query that displays the average price of a book for each publisher
  - Use the GROUP BY clause

```
SELECT pub_id, Avg(price) As AvgPrice
FROM titles
GROUP BY pub_id
```

	pub_id	AvgPrice
1	0736	9.7840
2	0877	15.4100
3	1389	18.9760

GROUP BY Clause

- **Combines records with identical values in the specified field list into a single record.**

- **Syntax:**
```
SELECT <fieldlist>
FROM <tableList>
[WHERE <criteria>]
[GROUP BY <groupfieldlist>]
[HAVING <condition>]
```

- **A summary value is created for each record if you include an SQL aggregate function in the SELECT statement.**

- **For example, compute average pay per department:**
```
SELECT DeptId, Avg(PayRate) As AvgRate
FROM Employee
Group By DeptId
```

GROUP BY Clause (cont.)

- **Use the WHERE clause to exclude rows you don't want grouped, and use the HAVING clause to filter records after they've been grouped.**

- **For example, dept's averaging > $20 per hour:**

```
SELECT DeptId, Avg(PayRate) As AvgPay
FROM Employee
GROUP BY DeptId
HAVING Avg(PayRate)>20.00
```

- **Or, a count of each dept's employees making more than $40 per hour**

```
SELECT DeptId, Count(*) As HighWageEarners
FROM Employee
WHERE PayRate>40
GROUP BY DeptId
```
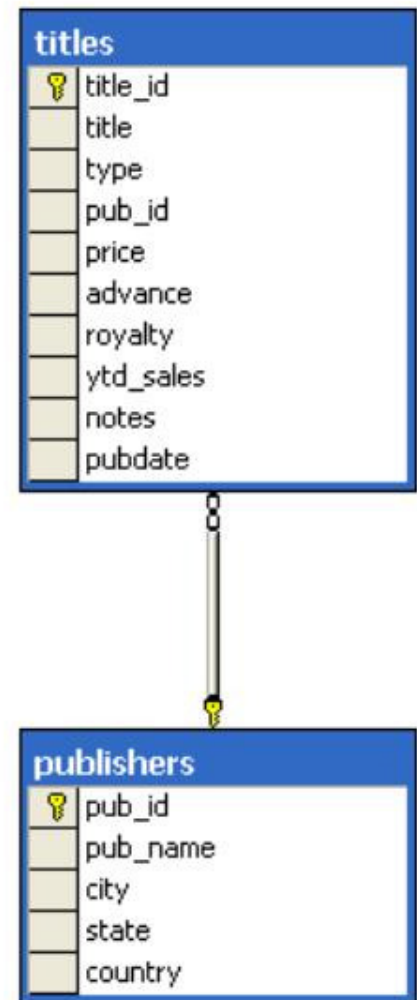
Aggregate Query Processing Example

Problem:

List publishers with more than two titles priced under $15.00

Solution:

```
SELECT pub_id,
       Count(*) AS OverFifteenTitles
FROM titles
WHERE price < 15.00
GROUP BY pub_id
HAVING Count(*) > 2
```

AQP Eg: Look at base table(s)

```
SELECT title, pub_id, price
FROM titles
ORDER BY pub_id, price
```

	title	pub_id	price
1	You Can Combat Computer Stress!	0736	2.9900
2	Life Without Fear	0736	7.0000
3	Emotional Security: A New A...	0736	7.9900
4	Is Anger the Enemy?	0736	10.9500
5	Prolonged Data Deprivation:...	0736	19.9900
6	The Psychology of Computer ...	0877	NULL
7	The Gourmet Microwave	0877	2.9900
8	Fifty Years in Buckingham P...	0877	11.9500
9	Sushi, Anyone?	0877	14.9900
10	Silicon Valley Gastronomic ...	0877	19.9900
11	Onions, Leeks, and Garlic: ...	0877	20.9500
12	Computer Phobic AND Non-Pho...	0877	21.5900
13	Net Etiquette	1389	NULL
14	Cooking with Computers: Sur...	1389	11.9500
15	The Busy Executive's Databa...	1389	19.9900
16	Straight Talk About Computers	1389	19.9900
17	Secrets of Silicon Valley	1389	20.0000
18	But Is It User Friendly?	1389	22.9500

AQP Eg: Implement the WHERE condition

```
SELECT title, pub_id,
   price
FROM titles
WHERE price < 15.00
ORDER BY pub_id, price
```

	title	pub_id	price
1	You Can Combat Computer Stress!	0736	2.9900
2	Life Without Fear	0736	7.0000
3	Emotional Security: A New A...	0736	7.9900
4	Is Anger the Enemy?	0736	10.9500
5	Prolonged Data Deprivation:...	0736	19.9900
6	The Psychology of Computer ...	0877	NULL
7	The Gourmet Microwave	0877	2.9900
8	Fifty Years in Buckingham P...	0877	11.9500
9	Sushi, Anyone?	0877	14.9900
10	Silicon Valley Gastronomic ...	0877	19.9900
11	Onions, Leeks, and Garlic: ...	0877	20.9500
12	Computer Phobic AND Non-Pho...	0877	21.5900
13	Net Etiquette	1389	NULL
14	Cooking with Computers: Sur...	1389	11.9500
15	The Busy Executive's Databa...	1389	19.9900
16	Straight Talk About Computers	1389	19.9900
17	Secrets of Silicon Valley	1389	20.0000
18	But Is It User Friendly?	1389	22.9500

AQP Eg: Aggregate the resulting set

```
SELECT pub_id, Count(*)
FROM titles
WHERE price < 15.00
GROUP BY pub_id
```

	title	pub_id	price	
1	You Can Combat Computer Stress!	0736	2.9900	
2	Life Without Fear	0736	7.0000	4
3	Emotional Security: A New A...	0736	7.9900	
4	Is Anger the Enemy?	0736	10.9500	
5	The Gourmet Microwave	0877	2.9900	
6	Fifty Years in Buckingham P...	0877	11.9500	3
7	Sushi, Anyone?	0877	14.9900	
8	Cooking with Computers: Sur...	1389	11.9500	1

AQP Eg: Filter Aggregate Results with HAVING

```
SELECT pub_id,
        Count(*) AS
   OverFifteenTitles
FROM titles
WHERE price < 15.00
GROUP BY pub_id
HAVING Count(*) > 2
```

	pub_id	OverFifteenTitles
1	0736	4
2	0877	3
3	1389	1

Result→

	pub_id	OverFifteenTitles
1	0736	4
2	0877	3

Aggregation with Joins

- **What if we wanted to display a more descriptive value for department instead of the group by PK?**

- **Answer: Join employee to department and include the DeptName in the Group By clause**

```
SELECT Department.DeptId, DeptName,
       Avg(PayRate) As AvgRate
FROM Employee INNER JOIN
     Department ON Department.DeptId=Employee.DeptId
Group By Department.DeptId, DeptName
```

Example with Subgroups

- **Problem: What are the highest individual pay rates among both union and non-union employees for each department?**

- **Solution:**

```
SELECT DeptName, IsUnion, Max(PayRate) As MaxRate
FROM Employee INNER JOIN
     Department ON Department.DeptId=Employee.DeptId
Group By DeptName, IsUnion
```