

Bài 3: TRUY VẤN THỐNG KÊ

Nội dung

- Truy vấn thống kê là gì?
- Các hàm thống kê (aggregate functions)
- Mệnh đề GROUP BY trong truy vấn thống kê
- Mệnh đề HAVING và điều kiện đối với các biểu thức thống kê.

Truy vấn thống kê là gì?

- Là câu lệnh SELECT thực hiện việc truy vấn dữ liệu từ các bảng trong cơ sở dữ liệu; và dựa trên dữ liệu truy vấn được để tính toán các **biểu thức thống kê** dựa trên các **nhóm dữ liệu**.

Ví dụ:

- Mỗi một khoa có bao nhiêu sinh viên?
- Điểm trung bình tích lũy của mỗi sinh viên lớp Tin K43A là bao nhiêu?
- Nhiệt độ trung bình mỗi tháng trong năm 2019 tại Thừa Thiên Huế là bao nhiêu?
- Mỗi quốc gia hiện có bao nhiêu người nhiễm Covid-19, bao nhiêu người đã điều trị khỏi bệnh?
- ...

Truy vấn thống kê là gì?

HoTen	Lop	DiemThi
Nga	Tin K41	8.0
Tuấn	Tin K42	5.0
Mai	Tin K41	9.5
Hùng	Tin K41	8.5
Long	Tin K42	6.0
Hải	Tin K43	5.0
Thanh	Tin K42	7.0
Thủy	Tin K43	10.0
Sang	Tin K42	8.0
Hữu	Tin K42	3.0

Bảng DiemSV

- Có bao nhiêu sinh viên mỗi lớp?
- Trung bình điểm thi của sinh viên mỗi lớp là bao nhiêu?
- Điểm thấp nhất của mỗi lớp là mấy điểm?
- Điểm cao nhất của mỗi lớp là mấy điểm?

```
SELECT  Lop,
        COUNT(*) AS N'Số lượng',
        AVG(DiemThi) AS N'TB điểm thi',
        MIN(DiemThi) AS N'Diểm thấp nhất',
        MAX(DiemThi) AS N'Diểm cao nhất'
FROM    DiemSV
GROUP BY Lop
```

Lop	Số lượng	TB điểm thi	Điểm thấp nhất	Điểm cao nhất
Tin K41	3	8.67	8.0	9.5
Tin K42	5	5.80	3.0	8.0
Tin K43	2	7.50	5.0	10.0

Cách thực hiện một truy vấn thống kê?

HoTen	Lop	DiemThi
Nga	Tin K41	8.0
Tuấn	Tin K42	5.0
Mai	Tin K41	9.5
Hùng	Tin K41	8.5
Long	Tin K42	6.0
Hải	Tin K43	5.0
Thanh	Tin K42	7.0
Thủy	Tin K43	10.0
Sang	Tin K42	8.0
Hữu	Tin K42	3.0

GROUP BY Lop

HoTen	Lop	DiemThi
Nga	Tin K41	8.0
Mai	Tin K41	9.5
Hùng	Tin K41	8.5
Tuấn	Tin K42	5.0
Long	Tin K42	6.0
Thanh	Tin K42	7.0
Sang	Tin K42	8.0
Hữu	Tin K42	3.0
Hải	Tin K43	5.0
Thủy	Tin K43	10.0

Tính trung bình điểm thi của sinh viên mỗi lớp!

```
SELECT Lop,
       AVG(DiemThi) AS N'TB điểm thi'
```

```
FROM DiemSV
```

```
GROUP BY Lop
```

Biểu thức tính giá trị trung bình của DiemThi trên từng nhóm

Dữ liệu được nhóm dựa trên giá trị của cột Lop

AVG(DiemThi)

Lop	TB điểm thi
Tin K41	8.67
Tin K42	5.80
Tin K43	7.50

Những yếu tố cần xác định là gì?

Dữ liệu được sử dụng để thống kê là gì?

Dữ liệu được nhóm (group) dựa trên những cột nào?

Biểu thức thống kê dùng để tính toán trên mỗi nhóm dữ liệu là thế nào?



Hàm và biểu thức thống kê

- Việc tính toán các giá trị thống kê trong SQL được thực hiện thông qua các biểu thức thống kê, được xây dựng dựa trên các hàm thống kê (aggregate functions).
- SQL cung cấp 6 hàm thống kê cơ bản:

SUM([ALL DISTINCT] biểu_thức)	Tính tổng
AVG([ALL DISTINCT] biểu_thức)	Tính giá trị trung bình
COUNT([ALL DISTINCT] biểu_thức)	Đếm số lượng giá trị
COUNT(*)	Đếm số dòng dữ liệu
MIN(biểu_thức)	Tính giá trị nhỏ nhất
MAX(biểu_thức)	Tính giá trị lớn nhất

Lưu ý: từ khóa ALL là mặc định nên không cần thiết phải viết tường minh khi sử dụng hàm thống kê

Hàm và biểu thức thống kê

• Hàm SUM()

- SUM(A) = 11
- SUM(DISTINCT A) = 9

• Hàm AVG()

- AVG(A) = 2.75
- AVG(DISTINCT A) = 3

• Hàm COUNT()

- COUNT(A) = 4
- COUNT(DISTINCT A) = 3

A
2
4
NULL
3
2

Hàm và biểu thức thống kê

- Hàm COUNT(*)

▪ COUNT(*) = 5

- Hàm MIN()

▪ MIN(A) = 2

- Hàm MAX()

▪ MAX(A) = 4

A
2
4
NULL
3
2

Hàm và biểu thức thống kê

- Ngoại trừ hàm COUNT(*), các hàm thống kê còn lại bỏ qua giá trị NULL trong quá trình tính toán.
- Các hàm thống kê có thể được sử dụng kết hợp với nhau để tạo thành các biểu thức thống kê phức tạp. Tuy nhiên, các hàm thống kê không được viết lồng vào nhau.

Ví dụ:

- Cách viết đúng:

SUM(DiemThi * SoTinChi) / SUM(SoTinChi)

- Cách viết sai:

MIN(AVG(DiemThi))

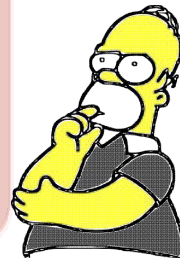


Mệnh đề GROUP BY

- Mệnh đề GROUP BY được sử dụng trong câu lệnh SELECT nhằm phân nhóm các dòng dữ liệu truy vấn được thành các nhóm dựa trên một hoặc nhiều cột.
- Mệnh đề GROUP BY sử dụng kết hợp với các biểu thức thống kê để thực hiện việc tính toán các giá trị thống kê trên từng nhóm dữ liệu.

Lưu ý:

Ngoại trừ những cột xuất hiện với vai trò là tham số của hàm thống kê, **tất cả các cột xuất hiện trong danh sách chọn của câu lệnh SELECT bắt buộc phải xuất hiện sau mệnh đề GROUP BY** của câu lệnh



Ví dụ minh họa

- Thống kê xem mỗi khoa hiện có bao nhiêu sinh viên

```
SELECT    k.MAKHOA,
          k.TENKHOA,
          COUNT(sv.MASV) AS SOSV
FROM      KHOA AS k
          LEFT JOIN LOP AS l
              ON k.MAKHOA = l.MAKHOA
          LEFT JOIN SINHVIEN AS sv
              ON l.MALOP = sv.MALOP
GROUP BY k.MAKHOA, k.TENKHOA
```

Ví dụ minh họa

- Cho biết điểm trung bình của các sinh viên lớp Tin K25 (lấy điểm thi lần 1)

```
SELECT      sv.MASV, sv.HODEM, sv.TEN,
            SUM(d.DIEMLAN1 * m.SODVHT) / SUM(m.SODVHT) AS DTB
FROM        LOP AS l
            JOIN SINHVIEN AS sv ON l.MALOP = sv.MALOP
            LEFT JOIN DIEMTHI as d ON sv.MASV = d.MASV
            LEFT JOIN MONHOC AS m ON d.MAMONHOC = m.MAMONHOC
WHERE       l.TENLOP = 'Tin K25'
GROUP BY    sv.MASV, sv.HODEM, sv.TEN
```

Điều kiện đối với biểu thức thống kê

- Mệnh đề HAVING được sử dụng trong câu lệnh SELECT trong trường hợp cần chỉ định điều kiện đối với các biểu thức thống kê.
- Trong câu lệnh SELECT, điều kiện truy vấn dữ liệu phải sử dụng đúng theo qui tắc:
 - Điều kiện đối với các cột hoặc các biểu thức sử dụng các hàm không phải là hàm thống kê phải được đặt sau mệnh đề WHERE
 - Điều kiện đối với các biểu thức có sử dụng hàm thống kê phải được đặt sau mệnh đề HAVING

Điều kiện đối với biểu thức thống kê

- **Ví dụ:** Hiển thị mã khoa, tên khoa và số lượng lớp của những khoa có số lượng lớp lớn hơn 2.

```
SELECT    k.MAKHOA,
          k.TENKHOA,
          COUNT(1.MALOP) AS SOLOP
FROM      KHOA AS k
          JOIN LOP AS l ON k.MAKHOA = l.MAKHOA
GROUP BY  k.MAKHOA, k.TENKHOA
HAVING    COUNT(MALOP) > 2
```

Truy vấn thống kê và phép nối

- Khi sử dụng phép nối trong truy vấn thống kê, cần chú ý cẩn thận đến việc sử dụng đúng loại phép nối (phép nối trong hay phép nối ngoài).
- Việc sử dụng sai loại phép nối có thể dẫn đến những nhận định và đánh giá sai về số liệu thống kê trên dữ liệu.

Bài tập thực hành

1. Hãy thống kê xem mỗi một loại chứng chỉ hiện đã mở được bao nhiêu lớp.
2. Hãy thống kê xem mỗi một loại chứng chỉ hiện đã mở được bao nhiêu lớp và đã có bao nhiêu lượt học viên đăng ký học.
3. Hãy thống kê xem trong tháng 9 năm 2010 mỗi một chứng chỉ có bao nhiêu lượt học viên đăng ký học.
4. Tổng số tiền học phí mà mỗi lớp dự kiến sẽ thu được là bao nhiêu?
5. Tổng số tiền học phí mà mỗi lớp đã thu được là bao nhiêu?

Bài tập thực hành

6. Những học viên nào đã nộp tiền học phí với tổng tiền lớn hơn 1 triệu đồng? Là đã nộp bao nhiêu tiền?
7. Hãy thống kê xem tổng số tiền học phí dự kiến sẽ thu và tổng số tiền học phí đã thu được của mỗi lớp là bao nhiêu?
8. Thống kê xem mỗi lớp có bao nhiêu học viên, trong đó có bao nhiêu học viên nam, bao nhiêu học viên nữ. Kết quả thống kê hiển thị theo mẫu:

MaLop	TenLop	SoLuongHV	SoHVNam	SoHVNũ