Database Systems I

More SQL

Instructor: Ouldooz Baghban Karimi

CMPT 354 - Summer 2019
More SQL

✅ Joins
  • Inner Join
  • Outer Join

• Aggregation Queries
  • Simple Aggregations
  • Group By
  • Having
# Example Tables

## Students

<table>
<thead>
<tr>
<th>sid</th>
<th>name</th>
<th>school</th>
<th>age</th>
<th>gpa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td>Adam</td>
<td>SFU</td>
<td>23</td>
<td>3.2</td>
</tr>
<tr>
<td>1002</td>
<td>Aiden</td>
<td>UBC</td>
<td>19</td>
<td>3.5</td>
</tr>
<tr>
<td>1003</td>
<td>Alice</td>
<td>SFU</td>
<td>18</td>
<td>3.7</td>
</tr>
<tr>
<td>1004</td>
<td>Bob</td>
<td>UBC</td>
<td>22</td>
<td>3.1</td>
</tr>
<tr>
<td>1006</td>
<td>John</td>
<td>SFU</td>
<td>21</td>
<td>3.1</td>
</tr>
<tr>
<td>1007</td>
<td>Mary</td>
<td>UBC</td>
<td>21</td>
<td>3.4</td>
</tr>
<tr>
<td>1008</td>
<td>Mike</td>
<td>SFU</td>
<td>24</td>
<td>3.1</td>
</tr>
<tr>
<td>1009</td>
<td>Sarah</td>
<td>UBC</td>
<td>18</td>
<td>3.0</td>
</tr>
</tbody>
</table>

## Enrolled

<table>
<thead>
<tr>
<th>stid</th>
<th>cid</th>
<th>grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td>200</td>
<td>A</td>
</tr>
<tr>
<td>1001</td>
<td>295</td>
<td>A</td>
</tr>
<tr>
<td>1001</td>
<td>250</td>
<td>B+</td>
</tr>
<tr>
<td>1002</td>
<td>130</td>
<td>A</td>
</tr>
<tr>
<td>1002</td>
<td>125</td>
<td>A+</td>
</tr>
<tr>
<td>1003</td>
<td>120</td>
<td>A</td>
</tr>
<tr>
<td>1003</td>
<td>125</td>
<td>B</td>
</tr>
<tr>
<td>1003</td>
<td>150</td>
<td>B</td>
</tr>
</tbody>
</table>
Joins

```
SELECT name, cid 
FROM Students INNER JOIN Enrolled ON sid = stid;

SELECT name, cid 
FROM Students FULL OUTER JOIN Enrolled ON sid = stid;

SELECT name, cid 
FROM Students LEFT OUTER JOIN Enrolled ON sid = stid;

SELECT name, cid 
FROM Students RIGHT OUTER JOIN Enrolled ON sid = stid;
```
More SQL

• Joins
  • Inner Join
  • Outer Join

✓ Aggregation Queries
  • Simple Aggregations
  • Group By
  • Having
Simple Aggregation

```
SELECT AGG(column)
FROM <table name>
WHERE <conditions>
```

AGG = COUNT, SUM, AVG, MAX, MIN, etc

• Count applies to more than one attribute
• Other aggregations apply to a single attribute
Examples: Simple Aggregation

• `SELECT COUNT(*) FROM Students;`

• `SELECT SUM(gpa) FROM Students;`

• `SELECT AVG(gpa) FROM Students;`

• `SELECT MIN(gpa) FROM Students;`

• `SELECT MAX(gpa) FROM Students;`
Examples: Simple Aggregation

- `SELECT COUNT(DISTINCT gpa) FROM Students;`
- `SELECT SUM(DISTINCT gpa) FROM Students;`
- `SELECT AVG(gpa) FROM Students WHERE age = 19;`
Group By: The Need

• How to get AVG(gpa) for each age?

SELECT AVG(gpa) FROM Students WHERE age = 18;
SELECT AVG(gpa) FROM Students WHERE age = 19;
SELECT AVG(gpa) FROM Students WHERE age = 20;
SELECT AVG(gpa) FROM Students WHERE age = 21;
Grouping & Aggregation

\[
\text{SELECT } \text{agg(column)} \\
\text{FROM } <\text{table name}> \\
\text{WHERE } <\text{conditions}> \\
\text{GROUP BY } <\text{columns}>
\]

• How to get AVG(gpa) for each age?
  \[
  \text{SELECT AVG(gpa) FROM Students GROUP BY age;}
  \]
Grouping

• How is the following query processed?
  
  ```sql
  SELECT age, AVG(gpa)
  FROM Students
  WHERE gpa > 2.5
  GROUP BY age;
  ```

• Semantics of the query
  • Compute the `FROM` and `WHERE` clauses
  • Group by the attributes in the `GROUP BY`
  • Compute the `SELECT` clause: grouped attributes and aggregates
Grouping

• Special Cases
  • Selecting Attributes
    • Everything in SELECT must be either a GROUP-BY attribute, or an aggregate

• Empty Groups
Having

• Specify which groups you are interested in

```sql
SELECT agg(column)
FROM <table name>
WHERE <conditions>
GROUP BY <columns>
HAVING <columns>
```

• **HAVING** clause contains conditions on aggregates.
Example: Having

```
SELECT AVG(gpa), age
FROM Student
WHERE gpa > 2.5
GROUP BY age
HAVING COUNT(*) >= 2;
```
Having: Order of Evaluation

- Create the cross product of the tables in the FROM clause
- Remove rows not meeting the WHERE condition
- Divide records into groups by the GROUP BY clause
- Remove groups not meeting the HAVING clause
- Create one row for each group and remove columns not in the SELECT clause
Acknowledgements

I have used materials from the following resources in preparation of this course:

• Database Systems: The Complete Book
• Database Systems (Kifer, Bernstein, Lewis)
• Database System Concepts: https://www.db-book.com
• Course offerings
  • W 4111 (Eugene Wu - Columbia): https://w4111.github.io/
  • CS 245 (Matei Zaharia - Stanford): http://web.stanford.edu/class/cs245/
  • CS 186 (Joe Hellerstein - Berkeley): https://sites.google.com/site/cs186fall17/