CS 437/537 (Fall 2016)

Assignment 1

Published: Sep. 9, 2016
Due: Sep. 16, 2016 (11:59pm)

Total: 20 points

Please upload your solutions to classes*v2. To do so, please enter the classes*v2 page of CPSC 437/537, then click the “Assignment” button on your left-hand toolbar, and finally click “Assignment1” to upload your assignment.

If you know you are going to submit your assignment late, please let us know in advance (send an email to cs437ta@cs.yale.edu). Any and all resources may be used as long as you cite them, with the exception of collaborating with other people.

Do not copypasta your solutions from our slides, or the Internet. We do not really care if you do this, but you are not really learning anything.

If you have ANY questions, please do not hesitate to let us know (email, office hours, etc.)
Part 1: Short questions (1-4 line answers for each, 9 points)

1. What is a DBMS? (1 point)
2. Given the fact that we already had file systems, why we need database? (2 points)
3. How many levels of abstraction a typical DB system should hold? What’s the purpose of each of these abstraction levels? (2 points)
4. What’s the difference between relational algebra language and SQL? (2 points)
5. In relational algebra language, why we say set intersection operation and natural join operation are not necessary? (2 points)

Part 2: Multiple choice (5 x 1 points each = 5 points)

1. A superkey is:
   a. A minimal candidate key
   b. A set of attributes that uniquely identify each tuple
   c. A proper superset of the primary key
   d. All of the above

2. Relational algebra cannot be used to compute:
   a. Sum
   b. Equality
   c. Max
   d. None of the above

3. Which of the following is not a relational database:
   a. MySQL
   b. PostgreSQL
   c. Redis
   d. None of the above

4. Physical data independence is the ability to:
   a. Modify the logical schema without changing the physical schema
   b. Modify the physical schema without changing the logical schema
   c. Modify the data without changing the physical or logical schemas
   d. None of the above

5. Which of the following is not the feature of a typical database system (e.g., PostgreSQL):
   a. High availability
   b. Data consistency
   c. Data private to DB administrators
   d. None of the above

Part 3: Longer questions (3 x 2 points each = 6 points)

3.1. We have the following database schema:

Computer(vendor, model, machinetype)
where vendor indicates the manufacturer of a computer, and machinetype takes values such as “desktop”, “laptop” and “server”. Underline indicates the primary key. Following inclusion dependencies hold: vendor \(\subseteq\) name, and model \(\subseteq\) num. Express following queries in relational algebra:

1. Find all the vendors who make laptop(s).
2. Find the phone numbers of all the vendors who make desktops with speed = 2

3.2. For the relations \(s\) and \(r\) in the Page 14 of our chapter 2 slides (i.e., joining two relations), suppose we type a collection of relational algebra queries, and get the following results.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)</td>
<td>2</td>
<td>(\alpha)</td>
<td>10</td>
<td>b</td>
</tr>
<tr>
<td>(\beta)</td>
<td>2</td>
<td>(\alpha)</td>
<td>10</td>
<td>b</td>
</tr>
</tbody>
</table>

Please explain how this happened? If you think this result could happen in practice, please write the corresponding relational algebra queries; otherwise, please explain why.