Course Administration

• Assignment #1 is due tomorrow (in class).
• Assignment #2 will be out on the home webpage.
  - It is due two weeks from today.
• Course slides will have black background
  - Printer friendly: set the printing color to “black-white”
• Next week reading:
  - Chapter 5 SQL
Reflection: DB design

• Step 1: Requirements Analysis
  – What data to store in the database?

• Step 2: Conceptual Database Design
  – Come up with the design: Entity-Relation (ER) model
  – Sketch the design with entity-relationship diagrams

• Step 3: Logical Database Design
  – Implement the design: relational data model
  – Map ER diagrams to relational tables
What’s next?

- How to ask questions about the relational database?
  - How much money in account XYZ?
  - Who are valuable customers \([\sum \text{ deposits} > 1M]\)?
- Two query languages
  - Relational algebra [CH4] : a Math query language
  - SQL [CH5] : a Real query Language

Relational Algebra

Chapter 4.1 – 4.2
Relational Query Languages

• What are query languages?
  – For asking questions about the database

• Relational Algebra
  – Mathematical Query Languages form the basis for “real” languages (e.g. SQL), and for implementation.
  – A relational query is composed using a small set of operators ($\pi$, $\sigma$, $\bowtie$, $\times$, …)
    • Like +, -, *, / operators in algebra

Preliminaries

• A query is applied to table(s), and the result of a query is also a table.
  – Schema of input table(s) for a query is fixed.
  – Schema for the result of a given query is also fixed! Determined by definition of query language constructs.

• Example of a relational algebra expression:
  – Find the names of sailors who have reserved boat 103
    $$\pi_{\text{sname}}((\sigma_{\text{bid} = 103} \text{Reserves}) \bowtie \text{Sailors})$$
Example Tables

- Sailors and Reserves are tables.
- Can refer to the fields by their positions or names:
- Assume names of fields in the result table are inherited from names of fields in input tables.

<table>
<thead>
<tr>
<th>sid</th>
<th>sname</th>
<th>rating</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>dustin</td>
<td>7</td>
<td>45.0</td>
</tr>
<tr>
<td>31</td>
<td>lubber</td>
<td>8</td>
<td>55.5</td>
</tr>
<tr>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sid</th>
<th>bid</th>
<th>day</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>101</td>
<td>10/10/96</td>
</tr>
<tr>
<td>58</td>
<td>103</td>
<td>11/12/96</td>
</tr>
</tbody>
</table>

Relational Algebra

- Basic relational algebra operators:
  - Selection (σ, pronounced sigma): Select a subset of rows from a table.
  - Projection (π): Delete unwanted columns from a table.
  - Cross-product (X): Combine two tables.
  - Set-difference (-): Tuples in table 1, but not in table 2.
  - Union (U): Tuples in tables 1 or 2.
- Each operator can take one or two input table(s), and returns one table.
Relational Algebra (more)

- Additional relational algebra operators:
  - Intersection (∩): Tuples in tables 1 and 2.
  - Join (∞): conditional cross product
  - Division (/):
  - Renaming (p, pronounced “row”)
- Operations can be composed to form a very complex query
  \[ \pi_{sid} (\sigma_{age > 20} \text{Sailors}) - \pi_{sid} (\sigma_{color = 'red'} \text{Boats}) \bowtie \text{Reserves} \bowtie \text{Sailors} \]

Relational Operators

- Projection
- Selection
- Union
- Intersection
- Set difference
- Cross product
- Rename operator
- Join
- Division
**Projection**

- Delete attributes not in projection list.
- Duplicates eliminated
- Find ages of sailors in S2
- Find names of sailors in S2

\[
\pi_{\text{name}, \text{rating}}(S2)
\]

\[
\begin{array}{|c|c|c|}
\hline
\text{sid} & \text{name} & \text{rating} \\
\hline
28 & yuppy & 9 \\
31 & lubber & 8 \\
44 & guppy & 5 \\
58 & rusty & 10 \\
\hline
\end{array}
\]

\[
\pi_{\text{age}}(S2)
\]

\[
\begin{array}{|c|}
\hline
\text{age} \\
\hline
35.0 \\
55.5 \\
\hline
\end{array}
\]

**Relational Operators**

- Projection
- Selection
- Union
- Intersection
- Set difference
- Cross product
- Rename operator
- Join
- Division
Selection

- Selects rows satisfying selection condition.
  - with duplicate removal
- Result table can be fed into other operations
- Find (names, ratings) of sailors whose ratings are greater than 8.
- Find names of sailors whose ages are greater than 40.

\[
\pi_{\text{name}, \text{rating}}(\sigma_{\text{rating} > 8}(S2))
\]

\[
\sigma_{\text{rating} > 8}(S2)
\]

<table>
<thead>
<tr>
<th>sid</th>
<th>name</th>
<th>rating</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>yuppy</td>
<td>9</td>
<td>35.0</td>
</tr>
<tr>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
</tr>
</tbody>
</table>

Relational Operators

- Projection
- Selection
- Union
- Intersection
- Set difference
- Cross product
- Rename operator
- Join
- Division
**Union**

- Take two input tables, which must be union-compatible:
  - Same number of fields.
  - "Corresponding" fields have the same type.
- What is the schema of result?
- Find sailors in S1 or S2.

<table>
<thead>
<tr>
<th>sid</th>
<th>sname</th>
<th>rating</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>dustin</td>
<td>7</td>
<td>45.0</td>
</tr>
<tr>
<td>31</td>
<td>lubber</td>
<td>8</td>
<td>55.5</td>
</tr>
<tr>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
</tr>
<tr>
<td>44</td>
<td>guppy</td>
<td>5</td>
<td>35.0</td>
</tr>
<tr>
<td>28</td>
<td>yuppy</td>
<td>9</td>
<td>35.0</td>
</tr>
</tbody>
</table>

**Intersection**

<table>
<thead>
<tr>
<th>sid</th>
<th>sname</th>
<th>rating</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>dustin</td>
<td>7</td>
<td>45.0</td>
</tr>
<tr>
<td>31</td>
<td>lubber</td>
<td>8</td>
<td>55.5</td>
</tr>
<tr>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
</tr>
<tr>
<td>28</td>
<td>yuppy</td>
<td>9</td>
<td>35.0</td>
</tr>
<tr>
<td>31</td>
<td>lubber</td>
<td>8</td>
<td>55.5</td>
</tr>
<tr>
<td>44</td>
<td>guppy</td>
<td>5</td>
<td>35.0</td>
</tr>
<tr>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
</tr>
</tbody>
</table>
### Set-Difference

<table>
<thead>
<tr>
<th>sid</th>
<th>sname</th>
<th>rating</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>dustin</td>
<td>7</td>
<td>45.0</td>
</tr>
<tr>
<td>31</td>
<td>lubber</td>
<td>8</td>
<td>55.5</td>
</tr>
<tr>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sid</th>
<th>sname</th>
<th>rating</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>yuppy</td>
<td>9</td>
<td>35.0</td>
</tr>
<tr>
<td>31</td>
<td>lubber</td>
<td>8</td>
<td>55.5</td>
</tr>
<tr>
<td>44</td>
<td>guppy</td>
<td>5</td>
<td>35.0</td>
</tr>
<tr>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
</tr>
</tbody>
</table>

\[ S_1 - S_2 \]

### Relational Operators

- Projection
- Selection
- Union
- Intersection
- Set difference
- Cross product
- Rename operator
- Join
- Division
Cross-Product

- Each row of $S_1$ is paired with each row of $R_1$.
- Result schema has one field per field of $S_1$ and $R_1$, with field names ‘inherited’ if possible.
  - Conflict: Both $S_1$ and $R_1$ have a field called sid.

<table>
<thead>
<tr>
<th>sid</th>
<th>sname</th>
<th>rating</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>dustin</td>
<td>7</td>
<td>45.0</td>
</tr>
<tr>
<td>31</td>
<td>lubber</td>
<td>8</td>
<td>55.5</td>
</tr>
<tr>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sid</th>
<th>sname</th>
<th>rating</th>
<th>age</th>
<th>(sid)</th>
<th>bid</th>
<th>day</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>dustin</td>
<td>7</td>
<td>45.0</td>
<td>22</td>
<td>101</td>
<td>10/10/96</td>
</tr>
<tr>
<td>31</td>
<td>lubber</td>
<td>8</td>
<td>55.5</td>
<td>31</td>
<td>101</td>
<td>10/10/96</td>
</tr>
<tr>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
<td>58</td>
<td>103</td>
<td>11/12/96</td>
</tr>
</tbody>
</table>

Renaming operator: $\rho(C(1 \rightarrow sid1, 5 \rightarrow sid2), S_1 \times R_1)$

Condition Joins

$S_1 \bowtie S_1 \text{sid} < R_1 \text{sid} R_1$

$R \bowtie_c S = \sigma_c(R \times S)$

- Cross-product, followed by a selection
- Result schema same as that of cross-product.
- Fewer tuples than cross-product, reduce tuples not meeting the condition.
Equi-Joins

- A special case of condition join where the condition $c$ contains only equalities.
- Result schema similar to cross-product, but only one copy of fields for which equality is specified.
- Natural Join ($\bowtie$): Equi-join on all common fields.

Example of Join

- Find the names of sailors who have reserved at least a boat.
  
  - $\pi_{\text{sname}}(R1 \bowtie S1)$
  - $\pi_{\text{sname}}(R1 \bowtie R1.\text{sid} = S1.\text{sid})$
  - $\pi_{\text{sname}}(R1 \bowtie \text{sid} S1)$

---

<table>
<thead>
<tr>
<th>sid</th>
<th>bid</th>
<th>day</th>
<th>sid</th>
<th>name</th>
<th>rating</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>101</td>
<td>10/10/96</td>
<td>22</td>
<td>dustin</td>
<td>7</td>
<td>45.0</td>
</tr>
<tr>
<td>58</td>
<td>103</td>
<td>11/12/96</td>
<td>31</td>
<td>lubber</td>
<td>8</td>
<td>55.5</td>
</tr>
<tr>
<td>58</td>
<td></td>
<td></td>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sid</th>
<th>bid</th>
<th>day</th>
<th>sid</th>
<th>name</th>
<th>rating</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>101</td>
<td>10/10/96</td>
<td>22</td>
<td>dustin</td>
<td>7</td>
<td>45.0</td>
</tr>
<tr>
<td>58</td>
<td>103</td>
<td>11/12/96</td>
<td>31</td>
<td>lubber</td>
<td>8</td>
<td>55.5</td>
</tr>
<tr>
<td>58</td>
<td></td>
<td></td>
<td>58</td>
<td>rusty</td>
<td>10</td>
<td>35.0</td>
</tr>
</tbody>
</table>

---
Relational Operators

- Projection
- Selection
- Union
- Intersection
- Set difference
- Cross product
- Rename operator
- Join
- Division

Division: analogy to integer division

- Two integers: A, B
  - A/B = Q, Q is the largest integer such that Q * B ≤ A
- How about relational tables A, B?
  - A/B = Q, Q is the largest table such that Q X B ≤ A
- Look at Q X B in A
  - Q must be a subset of attributes in A
  - Q’s attributes + B’s attributes = A’s attributes
  - A’s tuples must contain all pairings <q in Q, b in B>
Division

- Reserves(sailor_name, boat_id); Boats (boat_id)
  - Useful for expressing queries like:
    Find sailors who have reserved all boats => Reserves / Boats
- Let A have 2 fields, x and y; B have only field y:
  - $A/B = \{ \langle x \rangle | \exists \langle x, y \rangle \in A \land \forall \langle y \rangle \in B \}$
  - $A[xy]/B[y]$ contains all x tuples (sailor_name) such that for every y tuple (boat_id) in B, there is an xy tuple in A.

Examples of Division $A/B$

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Y1</td>
</tr>
<tr>
<td>X1</td>
<td>Y2</td>
</tr>
<tr>
<td>X1</td>
<td>Y3</td>
</tr>
<tr>
<td>X1</td>
<td>Y4</td>
</tr>
<tr>
<td>X2</td>
<td>Y1</td>
</tr>
<tr>
<td>X2</td>
<td>Y2</td>
</tr>
<tr>
<td>X3</td>
<td>Y2</td>
</tr>
<tr>
<td>X4</td>
<td>Y2</td>
</tr>
<tr>
<td>X4</td>
<td>Y3</td>
</tr>
<tr>
<td>X4</td>
<td>Y4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Y2</td>
</tr>
<tr>
<td>X2</td>
<td>Y2</td>
</tr>
<tr>
<td>X4</td>
<td>Y4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Y1</td>
</tr>
<tr>
<td>X1</td>
<td>Y2</td>
</tr>
<tr>
<td>X1</td>
<td>Y3</td>
</tr>
<tr>
<td>X4</td>
<td>Y4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Y1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Y1</td>
</tr>
<tr>
<td>X4</td>
<td>Y4</td>
</tr>
</tbody>
</table>
Expressing A/B Using Basic Operators

• Idea: For A/B, compute all x values that are not disqualified by some y value in B.
  – x value is disqualified if by attaching y value from B, we obtain an xy tuple that is not in A.
  – 1) Iterate through each x value
  – 2) Check: combined with each y value, xy in A? If not, disqualify.

• Disqualified x values: ______________________

A/B = ______________________

Examples of Division A/B

- $\mathcal{R}_x(A) \times B$
- above – A
- $\mathcal{R}_x$(above)
- $\mathcal{R}_x(A)$ - above
Practices with Relational Operators

(Q 1) Find names of sailors who’ve reserved boat #103

Reserves(sid, bid, day)
Sailors(sid, sname, rating, age)

• Solution 1:
  \( \pi_{\text{sname}}(\sigma_{\text{bid} = 103}(\text{Reserves} \bowtie \text{Sailors})) \)

• Solution 2 (more efficient)
  \( \pi_{\text{sname}}((\sigma_{\text{bid} = 103} \text{Reserves}) \bowtie \text{Sailors}) \)

• Solution 3 (using rename operator)
  \( \pi_{\text{sname}}(\text{Temp1}, \sigma_{\text{bid} = 103} \text{Reserves}) \)
  \( \pi_{\text{sname}}(\text{Temp2}, \text{Temp1} \bowtie \text{Sailors}) \)
(Q 2) Find names of sailors who’ve reserved a red boat

Reserves(sid, bid, day)
Sailors(sid, sname, rating, age)
Boats(bid, bname, color)

• Solution?
  \[ \pi_{sname}((\sigma_{\text{color} = \text{red}} \text{Boats}) \bowtie \text{Reserves} \bowtie \text{Sailors}) \]

• A more efficient solution (# of disk access)?
  \[ \pi_{sname}((\pi_{sid}(\pi_{bid}(\sigma_{\text{color} = \text{red}} \text{Boats}) \bowtie \text{Reserves}) \bowtie \text{Sailors})) \]

A query optimizer can find this, given the first solution!

(Q 3) Find the colors of boats reserved by Lubber

Reserves(sid, bid, day)
Sailors(sid, sname, rating, age)
Boats(bid, bname, color)

• Solution?
  \[ \pi_{\text{color}}((\sigma_{\text{sname} = \text{Lubber}} \text{Sailor}) \bowtie \text{Reserves} \bowtie \text{Boats}) \]
(Q 4) Find the names of sailors who have reserved at least one boat

Reserves(sid, bid, day)
Sailors(sid, sname, rating, age)
Boats(bid, bname, color)

• Solution?

\[ \pi_{\text{sname}}(\text{Sailor} \bowtie \text{Reserves}) \]

(Q 5) Find the names of sailors who have reserved a red or a green boat

Reserves(sid, bid, day)
Sailors(sid, sname, rating, age)
Boats(bid, bname, color)

• Solution?

\[ \pi_{\text{sname}}((\sigma_{\text{color} = \text{red}} \text{ or } \text{color} = \text{green}) \text{Boats} \bowtie \text{Reserves} \bowtie \text{Sailors}) \]
(Q 6) Find the names of sailors who have reserved a red and a green boat

Reserves(sid, bid, day)
Sailors(sid, sname, rating, age)
Boats(bid, bname, color)

• Wrong solution:
  \[ \pi_{\text{sname}} (\sigma_{\text{color} = \text{red}} \land \text{color} = \text{green}) \text{ Boats} \bowtie \text{Reserves} \bowtie \text{Sailors} \]

• What’s wrong with the above?

• Correct solution?
  \[ \pi_{\text{sname}} (\sigma_{\text{color} = \text{red}} \text{ Boats} \bowtie \text{Reserves} \bowtie \text{Sailors}) \cap \]
  \[ \pi_{\text{sname}} (\sigma_{\text{color} = \text{green}} \text{ Boats} \bowtie \text{Reserves} \bowtie \text{Sailors}) \]

(Q 7) Find the names of sailors who have reserved at least two different boats

Reserves(sid, bid, day)
Sailors(sid, sname, rating, age)
Boats(bid, bname, color)

• Strategy?
  - Join a table (sid, bid): sailors reserving at least one boat
  - Cross-product the table with itself
  - Select sailors with two different boats reserved
  \[ P (\text{Reservations}, C(1->\text{sid1}, 2->\text{sid2}, 3->\text{bid1}, 4->\text{bid2})) \]
  \[ \pi_{\text{sid}, \text{sname}, \text{bid}} (\text{Sailors} \bowtie \text{Reserves}) \]
  \[ \pi_{\text{sname}} (\sigma_{(\text{sid1}=\text{sid2}) \land (\text{bid1} \neq \text{bid2})} \text{ Reservations} \bowtie \text{Reservations}) \]
(Q 8) Find the sids of sailors with age over 20 who have not reserved a red boat

\[ \text{Reserves}(\text{sid, bid, day}) \]
\[ \text{Sailors}(\text{sid, sname, rating, age}) \]
\[ \text{Boats}(\text{bid, bname, color}) \]

• **Strategy**
  - Find all sailors (sids) with age over 20
  - Find all sailors (sids) who have reserved a red boat
  - Take their set differences

\[ \pi_{\text{sid}} (\sigma_{\text{age}>20} \text{ Sailors}) - \pi_{\text{sid}} (\sigma_{\text{color}='\text{red'}} \text{ Boats}) \cap \text{Reserves} \]

(Q 9A) Find the names of sailors who have reserved all boats

\[ \text{Reserves}(\text{sid, bid, day}) \]
\[ \text{Sailors}(\text{sid, sname, rating, age}) \]
\[ \text{Boats}(\text{bid, bname, color}) \]

• **Strategy**
  - Use division; all = division
  - what divides by what? Solution

\[ \pi_{\text{sname}} ((\pi_{\text{sid, bid}} (\text{Reserves}) / \pi_{\text{bid}} (\text{Boats})) \cap \text{Sailors}) \]
(Q 9’) Find the names of sailors who have reserved boats with all different colors

Reserves(sid, bid, day)
Sailors(sid, sname, rating, age)
Boats(bid, bname, color)

• Strategy
  – what divides by what?
    • (sid, color) / (color)
• Solution
  \[ \pi_{sname} ((\pi_{sid,color} (\text{Reserves} \bowtie \text{Boats}) / \pi_{color} (\text{Boats})) \bowtie \text{Sailors}) \]

(Q 10) Find the names of sailors who have reserved all boats called “Interlake”

Reserves(sid, bid, day)
Sailors(sid, sname, rating, age)
Boats(bid, bname, color)

• Previous solution?
  \[ \pi_{sname} ((\pi_{sid,bid} (\text{Reserves}) / \pi_{bid} (\text{Boats})) \bowtie \text{Sailors}) \]
• How to modify it?
  \[ \pi_{sname} ((\pi_{sid,bid} (\text{Reserves}) / \pi_{bid} (\sigma_{bname='Interlake'} \text{Boats})) \bowtie \text{Sailors}) \]
More Exercises

employee(person-name, street, city)
works(person-name, company-name, salary)
company(company-name, city)
manages(person-name, manager-name)

(a) Find the names of all employees who work for First Bank Corporation.
(b) Find the names and cities of residence of all employees who work for First Bank Corporation.
(c) Find the names, street address, and cities of residence of all employees who work for First Bank Corporation and earn more than $10,000 per annum.
(d) Find the names of all employees in this database who live in the same city as the company locates for which they work.

(e) Find the names of all employees who live in the same city and on the same street as do their managers.

(f) Find the names of all employees in this database who do not work for the First Bank Corporation.

(g) Find the names of all employees who earn more than every employee of Small Bank Corporation.

(h) Assume the companies may be located in several cities. Find all companies located in every city in which Small Bank Corporation is located.

(i) Find the names of employees who work for more than 3 (included) companies.