Today’s Topics

- Relational Model and Keys
- Relational Algebra
  - Select, Project, Set Operators
  - Natural Join, Theta Join
  - Outer Join
  - Grouping, Aggregation
  - “Write” Operations: Insert, Delete, Update
Join Operators

- Join operators seen so far

- Natural Join
- Theta Join
- Left Outer Natural Join
- Right Outer Natural Join
- Full Outer Natural Join
What Are Outer Joins?

- Consider the relations defined below.
  - Let's do a natural-join between `player` and `playsFor` to return a set of players and the teams for which they play.

<table>
<thead>
<tr>
<th>pid</th>
<th>Name</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>Homer</td>
<td>60000</td>
</tr>
<tr>
<td>67</td>
<td>Bart</td>
<td>50000</td>
</tr>
<tr>
<td>46</td>
<td>Ned</td>
<td>25000</td>
</tr>
<tr>
<td>11</td>
<td>Lisa</td>
<td>70000</td>
</tr>
<tr>
<td>10</td>
<td>Marge</td>
<td>50000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Team</th>
<th>pid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavs</td>
<td>31</td>
</tr>
<tr>
<td>Heat</td>
<td>11</td>
</tr>
<tr>
<td>Wizards</td>
<td>46</td>
</tr>
</tbody>
</table>

- What does the result set contain?
What if we wanted to *preserve* all players, even if they didn’t have a natural join partner?

![Diagram of left outer natural join with tables for players and playsFor]

- Players: Homer, Bart, Ned, Lisa, Marge
- Salaries: Homer 60000, Bart 50000, Ned 25000, Lisa 70000, Marge 50000
- Teams: Cavs, Heat, Wizards
- PlaysFor:
  - Cavs: 31
  - Heat: 11
  - Wizards: 46

*(What are these NULLs?)*
The *Nice* Thing About NULLs...

- **NULL** values are useful when the data isn't available

- **2 cases** when **NULLs** make sense:
  1. When a value of the attribute is simply unknown
     - *e.g.*, Unknown birthday or eye color?
  2. When a value is not applicable
     - *e.g.*, Some people don’t have middle names

- **NULLs** resulting from outer-joins are of case (2)
The *Annoying Things about NULLs...*

- **Annoyance 1:** The result of `NULL` and any arithmetic operation (+,-,*,/) is `NULL`
  - e.g., `5 + NULL = NULL`

- **Annoyance 2:** The result of `NULL` and boolean operators is `UNKNOWN`
  - e.g., `(NULL || true) == UNKNOWN`

- **Annoyance 3:** The result of `NULL` and comparative operators is `UNKNOWN`
  - e.g., `(4 > NULL) is UNKNOWN`
  - e.g., `(attribute_name == NULL) is UNKNOWN`
  - **Important:** How to check if an attribute contains `NULL`?
    - Use boolean operator: `IS NULL`
Example 1: What are the names of free agents (players without a team) using a *left outer natural join*

Example 2: Same query, but use *set difference*. (Hint: first find all players who are on a team).

Example 3: Same query, but using a *sub-query*. (Hint: find pid of all players who are on a team, then use [non]membership operator).
Outer Natural Joins

- **Left outer natural join** $R_1 \bowtie_\left R_2$ produces a natural join, but retains all tuples in $R_1$, concatenating NULL for all attributes in $R_2$

- **Right outer natural join** $R_1 \bowtie_\right R_2$ is defined similarly to left-outer join, except we retain all tuples in $R_2$

- **Full outer natural join** $R_1 \bowtie_\full R_2$ is defined similarly, except we retain all tuples in both sides.
  - NULLs may appear in both sides of the join
Today’s Topics

- The Relational Model
- Relational Algebra
  - Select, Project, Set Operators
  - Natural Join, Theta Join
  - Outer Join
  - Rename
  - Grouping, aggregation
  - “Write” Operations: Insert, Delete, Update
Rename Operator (Primitive)

- Syntax: $\rho_{R_2(b_1, \ldots, b_n)}(R_1)$
  - **Meaning**: Renames $R_1(a_1, \ldots, a_n)$ to $R_2(b_1, \ldots, b_n)$ for the lifetime of a query. Not a permanent rename!
  - Both $R_2$ and $b_1, \ldots, b_n$ are *optional*, but you're required to specify at least one.

**Examples:**

$\rho_{users}(Students)$

$\rho(F,l\_name)(Students)$

$\rho_{users}(F,L)(Students)$
When to Use Rename?

- **Reason 1**: Prefer different attribute and relation names in result.
  - (Weak reason, but very common in real life)

- **Reason 2**: Resolving ambiguity of attribute or relation names!
  - Example on next slide.

- **Reason 3**: Sometimes, relations and attributes are *unnamed*!
  - (Usually after a query!)
Example: Find everyone named Adam (in a human readable report)

residents

<table>
<thead>
<tr>
<th>f</th>
<th>l</th>
<th>bd</th>
<th>sid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>Smith</td>
<td>2/26/70</td>
<td>1</td>
</tr>
<tr>
<td>Courtney</td>
<td>Thatcher</td>
<td>2/28/70</td>
<td>3</td>
</tr>
</tbody>
</table>

\[ \sigma_{f='Adam'}(residents) \]

(unnamed)

<table>
<thead>
<tr>
<th>f</th>
<th>l</th>
<th>bd</th>
<th>sid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>Smith</td>
<td>2/26/70</td>
<td>1</td>
</tr>
</tbody>
</table>
Example: Find everyone named Adam (in a human-readable report)

$$\sigma_{\text{first}=\text{Adam}} \left( \rho(\text{first}, \text{last}, \text{birthdate}, \text{ssn})(\text{residents}) \right)$$

(residents)

<table>
<thead>
<tr>
<th>first</th>
<th>last</th>
<th>birthdate</th>
<th>ssn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>Smith</td>
<td>2/26/70</td>
<td>1</td>
</tr>
<tr>
<td>Courtney</td>
<td>Thatcher</td>
<td>2/28/70</td>
<td>3</td>
</tr>
</tbody>
</table>


(unnamed)

<table>
<thead>
<tr>
<th>first</th>
<th>last</th>
<th>birthdate</th>
<th>ssn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>Smith</td>
<td>2/26/70</td>
<td>1</td>
</tr>
</tbody>
</table>
Example: Find all pairs of students' first names who share last names.

A hint: The result has two first names in attribute list... Is this close?

\[ \pi_{f\_name, f\_name} (\sigma_{l\_name=l\_name} (\text{Student} \times \text{Student})) \]

Solution? (It gets us closer...)

\[ \pi_{f\_1, f\_2} (\sigma_{l\_1=l\_2} (\rho_{f\_1, l\_1} (\text{Student}) \times \rho_{f\_2, l\_2} (\text{Student}))) \]
Rename Example

- Query: Find the fastest plane in the database.
  - Hint -- First find all the planes that's slower than another plane.

<table>
<thead>
<tr>
<th>tail_no</th>
<th>make</th>
<th>model</th>
<th>capacity</th>
<th>mph</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>2</td>
<td>Airbus</td>
<td>A350</td>
<td>270</td>
<td>580</td>
</tr>
<tr>
<td>3</td>
<td>McDonnel Douglas</td>
<td>DC10</td>
<td>380</td>
<td>610</td>
</tr>
</tbody>
</table>
Self Test: That Last Problem

Solution:

\[
\text{plane} \setminus \pi_{T.\text{tailno}, T.\text{make}, T.\text{model}, T.\text{capacity}, T.\text{mph}}(\sigma_{T.\text{mph} < \text{plane.mph}}(\rho_T(\text{plane}) \times \text{plane}))
\]

Expression tree:

```
    plane
   / \        π T.\text{tailno}, T.\text{make}, T.\text{model}, T.\text{capacity}, T.\text{mph}
   \    \                   σ T.\text{mph} < \text{plane.mph}
  \   \                       \ ×
 plane
```

Temporarily rename plane to T -->

--- What does this step produce? (next slide)
### Intermediate result of: $\rho_T(plane) \times plane$

<table>
<thead>
<tr>
<th>tail_no</th>
<th>make</th>
<th>model</th>
<th>cap</th>
<th>mph</th>
<th>tail_no</th>
<th>make</th>
<th>model</th>
<th>cap</th>
<th>mph</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
<td>2</td>
<td>Airbus</td>
<td>A350</td>
<td>270</td>
<td>610</td>
</tr>
<tr>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
<td>3</td>
<td>MD</td>
<td>DC10</td>
<td>380</td>
<td>610</td>
</tr>
<tr>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
<td>2</td>
<td>Airbus</td>
<td>A350</td>
<td>270</td>
<td>610</td>
</tr>
<tr>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
<td>3</td>
<td>MD</td>
<td>DC10</td>
<td>380</td>
<td>610</td>
</tr>
<tr>
<td>2</td>
<td>Airbus</td>
<td>A350</td>
<td>270</td>
<td>580</td>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>2</td>
<td>Airbus</td>
<td>A350</td>
<td>270</td>
<td>580</td>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>2</td>
<td>Airbus</td>
<td>A350</td>
<td>270</td>
<td>580</td>
<td>2</td>
<td>Airbus</td>
<td>A350</td>
<td>270</td>
<td>610</td>
</tr>
<tr>
<td>2</td>
<td>Airbus</td>
<td>A350</td>
<td>270</td>
<td>580</td>
<td>3</td>
<td>MD</td>
<td>DC10</td>
<td>380</td>
<td>610</td>
</tr>
<tr>
<td>3</td>
<td>MD</td>
<td>DC10</td>
<td>380</td>
<td>610</td>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>747</td>
<td>570</td>
</tr>
<tr>
<td>3</td>
<td>MD</td>
<td>DC10</td>
<td>380</td>
<td>610</td>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>747</td>
<td>570</td>
</tr>
<tr>
<td>3</td>
<td>MD</td>
<td>DC10</td>
<td>380</td>
<td>610</td>
<td>2</td>
<td>Airbus</td>
<td>A350</td>
<td>580</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MD</td>
<td>DC10</td>
<td>380</td>
<td>610</td>
<td>3</td>
<td>MD</td>
<td>DC10</td>
<td>610</td>
<td></td>
</tr>
</tbody>
</table>

Pay attention to the $mph$ in both halves.
Self Test: That Last Problem (Cont.)

Solution:

\[
\text{plane} \setminus \pi_{T.\text{tailno}, T.\text{make}, T.\text{model}, T.\text{capacity}, T.\text{mph}}(\sigma_{T.\text{mph} < \text{plane.mph}}(\rho_T(\text{plane}) \times \text{plane}))
\]

Expression tree:

```
    plane
   /   \
  /     \   \     \
/       \  /      \  \\
\sigma_{T.\text{mph} < \text{plane.mph}} \rho_T \text{plane} \ni T.\text{tailno}, T.\text{make}, T.\text{model}, T.\text{capacity}, T.\text{mph}
```

(intermediate result on next slide) --->
Intermediate result of: $\sigma_{T.mph < plane.mph}(\rho_T(plane) \times plane)$

Observe: What tuples remain in the left-hand-side (T) of intermediate result?

Answer: These are all planes that are slower than some other plane (on the right)!
Solution:

\[
plane \setminus \pi_{T.tailno,T.make,T.model,T.capacity,T.mph}(\sigma_{T.mph < plane.mph}(\rho_T(plane) \times plane))
\]

Expression tree:

\[
\begin{aligned}
\text{plane} & \quad \pi_{T.tailno,T.make,T.model,T.capacity,T.mph} \\
\sigma_{T.mph < plane.mph} & \quad \rho_T \\
T & \quad \text{plane}
\end{aligned}
\]

← Project only the attributes in \(T\). (These are the slow planes)
Self Test: That Last Problem (Cont.)

- Intermediate result of:

\[
\pi_{T.t\text{ailno},T.m\text{ake},T.model,T.capacity,T.mph}(\sigma_{T.mph < plane.mph}(\rho_T(plane) \times plane))
\]

**from T (plane renamed)**

<table>
<thead>
<tr>
<th>tail_no</th>
<th>make</th>
<th>model</th>
<th>cap</th>
<th>mph</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>0</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>1</td>
<td>Boeing</td>
<td>747</td>
<td>525</td>
<td>570</td>
</tr>
<tr>
<td>2</td>
<td>Airbus</td>
<td>A350</td>
<td>270</td>
<td>580</td>
</tr>
</tbody>
</table>
Solution:

\[ plane \setminus \pi_{T\text{.tailno}, T\text{.make}, T\text{.model}, T\text{.capacity}, T\text{.mph}}(\sigma_{T\text{.mph} < \text{plane.mph}}(\rho_T(\text{plane}) \times \text{plane})) \]

Expression tree:

--- Finally, take set difference
Self Test: That Last Problem (Cont.)

Solution

\[
plane \setminus \pi_{T.\text{tailno}, T.\text{make}, T.\text{model}, T.\text{capacity}, T.\text{mph}}(\sigma_{T.\text{mph} < \text{plane.mph}}(\rho_T(\text{plane}) \times \text{plane}))
\]

- Take T away from original set of planes
  - And we have our answer

<table>
<thead>
<tr>
<th>tail_no</th>
<th>make</th>
<th>model</th>
<th>cap</th>
<th>mph</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>MD</td>
<td>DC10</td>
<td>380</td>
<td>610</td>
</tr>
</tbody>
</table>