More SQL (and Relational Algebra)

More SQL

- Extended Relational Algebra
- Outerjoins, Grouping/Aggregation
- Insert/Delete/Update

The Extended Relational Algebra

δ = eliminate duplicates from bags.

τ = sort tuples.

γ = grouping and aggregation.

- Also: Outerjoin -- avoids losing "dangling tuples".
- Dangling tuple = tuple that does not join with anything.

Duplicate Elimination

- R1 := δ(R2).
- R1 consists of one copy of each tuple that appears in R2 one or more times.

Example: Duplicate Elimination

R =

A	B
1	2
3	4
1	2

$\delta(R)$ =

A	B
1	2
3	4

Sorting

- R1 := τ_L (R2).
  - L is a list of some of the attributes of R2.
- R1 is the list of tuples of R2 sorted first on the value of the first attribute on L, then on the second attribute of L, and so on.
  - Break ties arbitrarily.
- τ is the only operator whose result is neither a set nor a bag. (Why?)

Example: Sorting

R =

A	B
1	2
3	4
5	2

$\tau_B(R) = [(5,2), (1,2), (3,4)]$

Aggregation Operators

- Aggregation operators are not operators of relational algebra.
- Rather, they apply to entire columns of a table and produce a single result.
- The most important examples: SUM, AVG, COUNT, MIN, and MAX.

Example: Aggregation

R =

A	B
1	3
3	4
3	2

SUM(A) = 7
COUNT(A) = 3
MAX(B) = 4
AVG(B) = 3

Grouping Operator

- R1 := γ_L (R2). *L* is a list of elements that are either:
  1. Individual (*grouping*) attributes.
  2. AGG(*A*), where AGG is one of the aggregation operators and *A* is an attribute.
     - An arrow and a new attribute name renames the component.
     - Also **AS** is sometimes used.

Applying $\gamma_L(R)$

- Group *R* according to all the grouping attributes on list *L*.
  - That is: form one group for each distinct list of values for those attributes in *R*.
- Within each group, compute AGG(*A*) for each aggregation on list *L*.
- Result has one tuple for each group:
  1. The grouping attributes and
  2. Their group's aggregations.

Example: Grouping/Aggregation

R = (

A	B	C
1	2	3
4	5	6
1	2	5

)

$\gamma_{A,B,AVG(C)->X}(R) = ??$

First, group *R* by *A* and *B* :

A	B	C
1	2	3
1	2	5
4	5	6

Then, average *C* within groups:

A	B	X
1	2	4
4	5	6

Outerjoin

- Suppose we join $R \bowtie_C S$.
- A tuple of R that has no tuple of S with which it joins is said to be *dangling*.
  - Similarly for a tuple of S.
- Outerjoin preserves dangling tuples by padding them NULL.

Example: Outerjoin

R =

A	B
1	2
4	5

S =

B	C
2	3
6	7

(1,2) joins with (2,3), but the other two tuples are dangling.

R OUTERJOIN S =

A	B	C
1	2	3
4	5	NULL
NULL	6	7

Now --- Back to SQL

Each Operation Has a SQL Equivalent

Outer Join

- An extension of the join operation that avoids loss of information.
- Computes the join and then adds tuples from one relation that do not match tuples in the other relation to the result of the join.
- Uses *null* values:
  - NULL signifies that the value is unknown or does not exist

Outerjoins

- R OUTER JOIN S is the core of an outerjoin expression. It is modified by:
  - Optional NATURAL in front of OUTER.
  - Optional ON <condition> after JOIN.

    Only one of “NATURAL” or “ON <condition>” allowed.
  - Optional LEFT, RIGHT, or FULL before OUTER.
    - LEFT = pad dangling tuples of R only.
    - RIGHT = pad dangling tuples of S only.
    - FULL = pad both; this choice is the default.
- ON is used to specify a theta join.

Example: Outerjoin

R =

A	B
1	2
4	5

S =

B	C
2	3
6	7

R LEFT OUTERJOIN S =

A	B	C
1	2	3
4	5	NULL

Aggregations

- SUM, AVG, COUNT, MIN, and MAX can be applied to a column in a SELECT clause to produce that aggregation on the column.
- Also, COUNT(*) counts the number of tuples.

Example: Aggregation

- From Sells(bar, beer, price), find the average price of Export:

```
SELECT AVG(price)
FROM Sells
WHERE beer = 'Export';
```

Eliminating Duplicates in an Aggregation

- Use DISTINCT inside an aggregation.
- Example: find the number of *different* prices charged for Export:

```
SELECT COUNT(DISTINCT price)
FROM Sells
WHERE beer = 'Export';
```

NULL are Ignored in Aggregation

- NULL never contributes to a sum, average, or count, and can never be the minimum or maximum of a column.
- But if there are no non-NULL values in a column, then the result of the aggregation is NULL.
  - Exception: COUNT of an empty set is 0.

Example: Effect of NULL's

```
SELECT count(*)
FROM Sells
WHERE beer = 'Export';
```

The number of bars that sell Export.

```
SELECT count(price)
FROM Sells
WHERE beer = 'Export';
```

The number of bars that sell Export at a known price.

Grouping

- We may follow a SELECT-FROM-WHERE expression by GROUP BY and a list of attributes.
- The relation that results from the FROM-WHERE is grouped according to the values of all those attributes, and any aggregation is applied only within each group.

Example: Grouping

- From Sells(bar, beer, price), find the average price for each beer:

```
SELECT beer, AVG(price)
FROM Sells
GROUP BY beer;
```

beer	AVG(price)
Export	2.33
…	…

Example: Grouping

- From Sells(bar, beer, price) and Frequents(customer, bar), find for each customer the average price of Export at the bars they frequent:

```
SELECT customer, AVG(price)
FROM Frequents, Sells
WHERE beer = 'Export' AND
          Frequents.bar = Sells.bar
GROUP BY customer;
```

Compute all customer-bar-price triples for Export.

Then group them by customer.

Restriction on SELECT Lists With Aggregation

- If any aggregation is used, then each element of the SELECT list must be either:
  1. Aggregated, or
  2. An attribute on the GROUP BY list.

Illegal Query Example

- From Sells(bar, beer, price), you might think you could find the bar that sells Export the cheapest by:

```
SELECT bar, MIN(price)
FROM Sells
WHERE beer = 'Export';
```

- But this query is illegal in SQL.

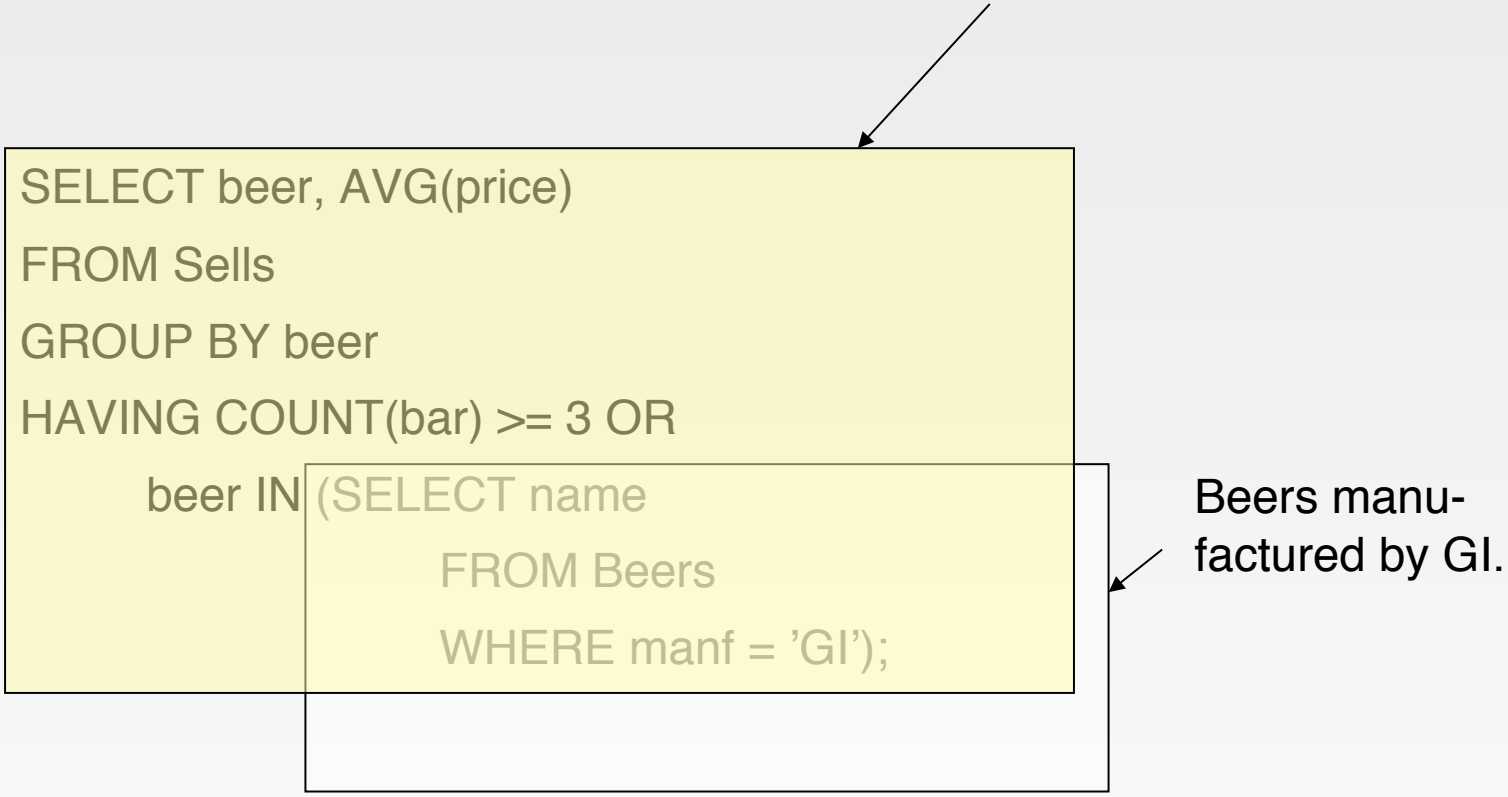
HAVING Clauses

- HAVING <condition> may follow a GROUP BY clause.
- If so, the condition is applied to each group, and groups not satisfying the condition are eliminated.
  - So HAVING is like a WHERE, but applied to groups
- Example:

  From Sells(bar, beer, price) and Beers(name, manf), find the average price of those beers that are either served in at least three bars or are manufactured by GI.

Solution

Beer groups with at least 3 non-NULL bars and also beer groups where the manufacturer is GI.



Beers manu-factured by GI.

Requirements on HAVING Conditions

- Anything goes in a subquery.
- Otherwise:
  1. An aggregation in a HAVING clause applies only to the tuples of the group being tested.
  2. Any attribute of relations in the FROM clause may be aggregated in the HAVING clause, but
  3. Only those attributes that are in the GROUP BY list may appear unaggregated in the HAVING clause.

  (same condition as for SELECT clauses with aggregation).

Database Modifications

- A *modification* command does not return a result (as a query does), but changes the database in some way.
- Three kinds of modifications:
  1. *Insert* a tuple or tuples.
  2. *Delete* a tuple or tuples.
  3. *Update* the value(s) of an existing tuple or tuples.

Insertion

- To insert a single tuple:

  INSERT INTO <relation>

  VALUES (<list of values>);

- Example: add to Likes(customer, beer) the fact that Sally likes Export.

  INSERT INTO Likes

  VALUES('Sally', 'Export');

- Also:

  BULK INSERT Courses

  FROM '\\cypress\userdata\<usersfolder>\Courses.txt'

  WITH(

  ROWTERMINATOR = '\n',

  FIELDTERMINATOR = ',');

Specifying Attributes in INSERT

- We may add to the relation name a list of attributes.
- Two reasons to do so:
  1. We forget the standard order of attributes for the relation.
  2. We don't have values for all attributes, and we want the system to fill in missing components with NULL or a default value.

As well: it's less error prone!

Example: Specifying Attributes

- Another way to add the fact that Sally likes Export to Likes(customer, beer):

```
INSERT INTO Likes(beer, customer)
VALUES('Export', 'Sally');
```

Adding Default Values

- In a CREATE TABLE statement, we can follow an attribute by DEFAULT and a value.
- When an inserted tuple has no value for that attribute, the default will be used.

Example: Default Values

```
CREATE TABLE customers (
        name CHAR(30) PRIMARY KEY,
        addr CHAR(50)
                DEFAULT '123 Sesame St.',
        phone CHAR(16)
);
```

Example: Default Values

```
INSERT INTO customers(name)
VALUES('Sally');
```

Resulting tuple:

name	address	phone
Sally	123 Sesame St	NULL

Inserting Many Tuples

- We may insert the entire result of a query into a relation, using the form:

  INSERT INTO <relation>

  (<subquery>);

Example: Insert a Subquery

- Using Frequents(customer, bar), enter into the new relation Compatriot(name), all of Sally's "compatriots," i.e., those customers who frequent at least one bar that Sally also frequents.

Solution

The other customer

Pairs of customer tuples where the first is for Sally, the second is for someone else, and the bars are the same.

```
INSERT INTO Compatriot
(SELECT d2.customer
 FROM Frequents d1, Frequents d2
 WHERE d1.customer = 'Sally' AND
     d2.customer <> 'Sally' AND
     d1.bar = d2.bar
);
```

Deletion

- To delete tuples satisfying a condition from some relation:

  DELETE FROM <relation>

  WHERE <condition>;

- Example:

  Delete from Likes(customer, beer) the fact that Sally likes Export:

  DELETE FROM Likes

  WHERE customer = 'Sally' AND beer = 'Export';

Example: Delete all Tuples

- Make the relation Likes empty:

```
DELETE FROM Likes;
```

- Note a WHERE clause is not needed.
- Q: What's the difference between

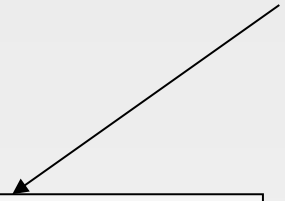
```
DELETE FROM Likes;
```

and

```
DROP TABLE Likes;  ?
```

Example: Delete Some Tuples

- Delete from Beers(name, manf) all beers for which there is another beer by the same manufacturer.



Beers with the same manufacturer and a different name from the name of the beer represented by tuple b.

```
DELETE FROM Beers b
WHERE EXISTS (
    SELECT name FROM Beers
    WHERE manf = b.manf AND
          name <> b.name);
```

Semantics of Deletion -- (1)

- Suppose Molson makes only Export and Canadian.
- Suppose we come to the tuple *b* for Export first.
- The subquery is nonempty, because of the Canadian tuple, so we delete Export.
- Now, when *b* is the tuple forCanadian, do we delete that tuple too?

Semantics of Deletion --- (2)

- **Answer**: we *do* delete Canadian as well.
- The reason is that deletion proceeds in two stages:
  1. Mark all tuples for which the WHERE condition is satisfied.
  2. Delete the marked tuples.

Updates

- To change certain attributes in certain tuples of a relation:

  UPDATE <relation>

  SET <list of attribute assignments>

  WHERE <condition on tuples>;

Example: Update

- Change customer Fred's phone number to 555-1212:

```
UPDATE customers
SET phone = '555-1212'
WHERE name = 'Fred';
```

Example: Update Several Tuples

- Make $4 the maximum price for beer:

```
UPDATE Sells
SET price = 4.00
WHERE price > 4.00;
```

- Increase the balance of accounts by 5% for accounts with a balance of more than $10000:

```
UPDATE Account
SET balance = balance * 1.05
WHERE balance >= 10000;
```

End: More SQL