Comp 5311 Database Management Systems

3. SQL 1

Structured Query Language - SQL

- Most common Query Language – used in all commercial systems
- Discussion is based on the SQL92 Standard.
  Commercial products have different features of SQL, but the basic structure is the same
- Data Manipulation Language
- Data Definition Language
- Constraint Specification
- Embedded SQL
- Transaction Management
- Security Management .....

Basic Structure

- SQL is based on algebra operations with certain modifications and enhancements
- A typical SQL query has the form:

  select A_1, A_2, …, A_n
  from R_1, R_2, …, R_m
  where P

  - A_i represent attributes
  - R_i represent relations
  - P is a predicate.
- This query is equivalent to the relational algebra expression:

$$\Pi_{A1, A2, \ldots, An}(\sigma_P(R_1 \times R_2 \times \ldots \times R_m))$$

- The result of an SQL query is a table (but may contain duplicates). SQL statements can be nested.

Projection

- The select clause corresponds to the projection operation of the relational algebra. It is used to list the attributes desired in the result of a query.
- Find the names of all branches in the *loan* relation

  select branch-name
  from loan

  Equivalent to: $\Pi_{\text{branch-name}}(\text{loan})$

- An asterisk in the select clause denotes "all attributes"

  select *
  from loan

- Note: for our examples we use the tables:
  - Branch (branch-name, branch-city, assets)
  - Customer (customer-name, customer-street, customer-city)
  - Loan (loan-number, amount, *branch-name*)
  - Account (account-number, balance, *branch-name*)
  - Borrower (*customer-name, loan-number*)
  - Depositor (*customer-name, account-number*)

Duplicate Removal

- SQL allows duplicates in relations as well as in query results. Use select distinct to force the elimination of duplicates.
  Find the names of all branches in the loan relation, and remove duplicates

```
select distinct branch-name
from loan
```

force the DBMS to remove duplicates

- The keyword all specifies that duplicates are not removed (optional for this query).

```
select all branch-name
from loan
```

force the DBMS not to remove duplicates

Arithmetic Operations on Retrieved Results

- The select clause can contain arithmetic expressions involving the operators,+,−,÷ and ×, and operating on constants or attributes of tuples.
- The query:

```
select branch-name, loan-number, amount * 100
from loan
```

would return a relation which is the same as the loan table, except that the attribute amount is multiplied by 100

The where Clause

- The where clause specifies conditions that tuples in the relations in the from clause must satisfy.
- Find all loan numbers for loans made at the Perryridge branch with loan amounts greater than $1200.

```
select loan-number
from loan
where branch-name="Perryridge" and amount >1200
```

- SQL allows logical connectives and, or, and not. Arithmetic expressions can be used in the comparison operators.
- Note: attributes used in a query (both select and where parts) must be defined in the relations in the from clause.

The where Clause (Cont.)

- SQL includes the between operator for convenience.
- Find the loan number of those loans with loan amounts between \$90,000 and \$100,000 (that is, ≥ \$90,000 and ≤ \$100,000)

```
select loan-number
from loan
where amount between 90000 and
100000
```

The from Clause

- The from clause corresponds to the Cartesian product operation of the relational algebra.
- Find the Cartesian product borrower × loan

```
select *
from borrower, loan
```

  It is rarely used without a where clause.
- Find the name and loan number of all customers having a loan at the Perryridge branch.

```
select distinct customer-name, borrower.loan-number
from borrower, loan
where borrower.loan-number = loan.loan-number and
branch-name = "Perryridge"
```

The Rename Operation

- Renaming relations and attributes using the as clause:
  old-name as new-name
- Find the name and loan number of all customers having a loan at the Perryridge branch; replace the column name loan-number with the name loan-id.

```
select distinct customer-name, borrower.loan-number as loan-id
from borrower, loan
where borrower.loan-number = loan.loan-number and
        branch-name = "Perryridge"
```

Tuple Variables/Alias

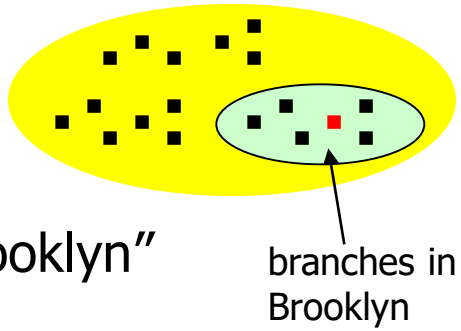
- Tuple variables are defined in the from clause via the use of the "as" clause.
- Find the customer names and their loan numbers for all customers having a loan at some branch.

```
select distinct customer-name, T.loan-number
from borrower as T, loan as S
where T.loan-number = S.loan-number
```

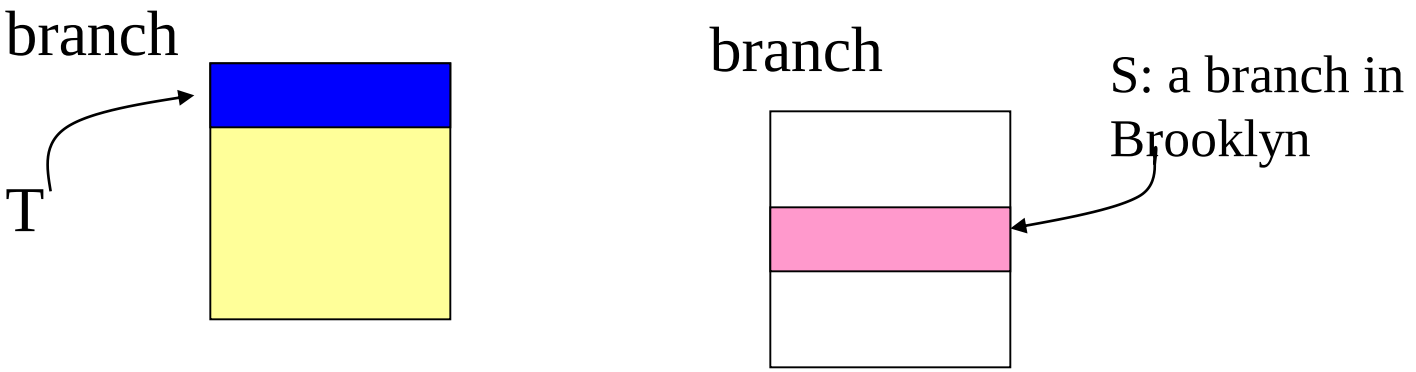
- Tuple variable/Alias can be used as short hand, but it is more than just a short hand (see next slide)

Tuple Variables/Alias

- Find the names of all branches that have greater assets than *some* branch located in Brooklyn.




```
select distinct T.branch-name
from branch as T, branch as S
where T.assets > S.assets and S.branch-city="Brooklyn"
```



Does it returns branches within Brooklyn?

String Operations

- Character attributes can be compared to a pattern:
  % matches any substring.
  _ matches any single character.
- Find the name of all customers whose street includes the substring 'Main'. (Eg Mainroad, Smallmain Road, AMainroad,…)

```
select customer-name
from customer
where customer-street like "%Main%"
```

Ordering the Display of Tuples

- List in alphabetic order the names of all customers having a loan at Perryridge branch

```
select distinct customer-name
from borrower, loan
where borrower.loan-number = loan.loan-number and
branch-name = "Perryridge"
order by customer-name
```

- order by customer-name desc, amount asc
  desc for descending order; asc for ascending order (default)
- SQL must perform a sort to fulfill an order by request. Since sorting a large number of tuples may be costly, it is desirable to sort only when necessary.

Set Operations

- The set operation union, intersect, and except operate on relations and correspond to the relational algebra operations $\cup$, $\cap$ and $-$.
- Each of the above operations automatically eliminates duplicates; to retain all duplicates use union all, intersect all and except all.
- Suppose a tuple occurs m times in r and n times in s, then, it occurs:
  - m + n times in r union all s
  - min(m,n) times in r intersect all s
  - max(0,m-n) times in r except all s

Set operations

- Find all customers who have a loan, an account, or both:

  (select customer-name from depositor)
  union
  (select customer-name from borrower)

- Find all customers who have both a loan and an account.

  (select customer-name from depositor)
  intersect
  (select customer-name from borrower)

- Find all customers who have an account but no loan.

  (select customer-name from depositor)
  except
  (select customer-name from borrower)

Example Database

Sailors (sid, sname),

Reserves (sid, bid, date),

Boats (bid,bname,color)

Sailors

sid	sname	rating	age
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

Reserves

sid	bid	day
22	101	10/10/96
58	103	11/12/96

Find the names of sailors who reserved bid=103

```
SELECT  S.sname
FROM    Sailors as S, Reserves as R
WHERE   S.sid=R.sid AND R.bid=103
```

(sid)	sname	rating	age	(sid)	bid	day
22	dustin	7	45.0	22	101	10/10/96
22	dustin	7	45.0	58	103	11/12/96
31	lubber	8	55.5	22	101	10/10/96
31	lubber	8	55.5	58	103	11/12/96
58	rusty	10	35.0	22	101	10/10/96
58	rusty	10	35.0	58	103	11/12/96

Find sid's of sailors who've reserved a red or a green boat

```
SELECT  R.sid
FROM  Boats as B, Reserves as R
WHERE  R.bid=B.bid
  AND (B.color='red' OR B.color='green')
```

Alternative

```
SELECT  R.sid
FROM  Boats as B, Reserves as R
WHERE  R.bid=B.bid
            AND B.color='red'
UNION
SELECT  R.sid
FROM  Boats as B, Reserves as R
WHERE  R.bid=B.bid
            AND B.color='green'
```

- If we replace OR by AND in the first version, what do we get?
- What do we get if we replace UNION by EXCEPT in the second version?

Find sid's of sailors who've reserved a red and a green boat

Key field!

```
SELECT  S.sid
FROM  Sailors as S, Boats as B, Reserves as R
WHERE  S.sid=R.sid AND R.bid=B.bid
          AND B.color='red'
INTERSECT
SELECT  S.sid
FROM  Sailors as S, Boats as B, Reserves as R
WHERE  S.sid=R.sid AND R.bid=B.bid
          AND B.color='green'
```

```
SELECT  S.sid
FROM  Sailors as S, Boats as B1, Reserves as R1,
        Boats as B2, Reserves as R2
WHERE  S.sid=R1.sid AND R1.bid=B1.bid
 AND B1.color='red' AND  S.sid=R2.sid
AND R2.bid=B2.bid AND B2.color='green'
```

- What if instead of the sid we want the sname? Would the queries be correct if we replace SELECT S.sid with S.sname?