



# CS 143 Discussion Session (Draft)

Week 2

# SQL

- Structured Query Language
  - Designed for RDBMS
  - Declarative Language
- Relational Algebra
  - SQL's mathematical interpretation

# General SQL

SELECT attributes, aggregates

FROM relations(tables)

WHERE conditions

GROUP BY attributes

HAVING conditions on aggregates

ORDER BY attributes, aggregates

# General SQL

- Evaluation order:
  - FROM → WHERE → GROUP BY → HAVING → ORDER BY → SELECT
- Relational Algebra Counterparts
  - FROM  $R_1, \dots, R_m$ :  $R_1 \times \dots \times R_m$
  - WHERE C:  $\sigma_C$
  - SELECT  $A_1, \dots, A_n$ :  $\pi_{A_1, \dots, A_n}$

# Important points

- Tables in FROM clause
  - cross product, not natural join
- SELECT clause
  - Projection, not selection in RA
- SQL uses bag semantics, duplicates not removed
  - SELECT distinct ...
- Tables/Attributes can also be renamed
  - GPA (AS) grade

# Set Operators

- $\cap$ : INTERSECT,  $\cup$ : UNION,  $-$ : EXCEPT
- Set operators should have the same schema for operands
  - In practice, it is okay to have just compatible types
- Set operators follow set semantics and remove duplicates
  - keep duplicates, use UNION ALL, INTERSECT ALL, EXCEPT ALL (bag semantics)

# Other set operators

- Set membership
  - IN, NOT IN
- Set comparison operator
  - > ALL, < SOME, = SOME, ..., etc.

# Aggregates

- Sum, Count, Avg, Min, Max, ...
- Combine multiple tuples into one
- Often used with GROUP BY
  - If not providing GROUP BY clause, treat all tuples as one single group.

# GROUP BY

- SELECT can have only attributes that have a single value in each group or aggregates
  - SELECT can contain **no more attributes than group by attributes** (group by keys)
  - SELECT can contain any number of **aggregates**

# Having

- Eg. Find students who take two or more classes

```
SELECT sid, count(*) as c  
FROM Enroll E  
GROUP BY sid  
HAVING c >= 2
```

# Having

- Condition tests on the groups
  - Executed after group by
- Conditions on aggregates should appear in the HAVING clause
- We can rewrite a query not to have a HAVING clause – Can be complicated!

# Order By

```
SELECT sid, GPA  
FROM Student  
ORDER BY GPA DESC, sid ASC
```