

# CS245: Databases

## SQL

Vijaya saradhi

Department of Computer Science and Engineering  
Indian Institute of Technology Guwahati

# Reading Data From Tables

- Selecting columns
- Selecting rows
- Select rows and columns
- Table operations
  - Union of two tables
  - Intersection between two tables
  - Difference of two tables
  - Cross product of two tables
  - Joining two tables
    - Natural join
    - Inner join (theta join)
    - Left outer join
    - Right outer join
    - Full outer join
  - Group by
  - Distinct rows/columns
  - Sort rows
  - Extended selection

## SELECT statement

- Is the most frequently used statement
- Is at the heart of the querying database tables
- Important as SELECT statement combines more than 9 relational algebraic operators
- We build from basics to advanced query structures

# SELECT statement structure

<b>SELECT</b>	list	the	<b>column names</b>
<b>FROM</b>	list	the	<b>table names</b>
<b>WHERE</b>	specify	the	condition
<b>GROUP BY</b>	list	the	<b>column names</b>
<b>HAVING</b>	specify	the	condition
<b>ORDER BY</b>	specify	the	<b>column names;</b>

## SELECT

- SELECT statement result in a table
- The result table will not be explicitly stored in the database
- Compose several SELECT statements to perform a required query
- Needed privileges to perform the select statemet!

## A quick list

- Select - columns
- Select - rows
- Select - rows & columns
- Select - remove duplicates
- Select - perform column sum, minimum, maximum, average, count
- Select - sort
- Select - group by specified column
- Select - create new columns by using expressoions/functions

# SELECT - columns

Select specified list of columns

Select rating from Sailors

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

=

New Table
7
1
8
8
10
7
10
9
3
3

Select specified list of columns

```
SELECT rating  
FROM Sailors ;
```

# SELECT - columns

Select specified list of columns

Select sid and rating

$$\left( \begin{array}{l} \text{Sailors} \\ \hline \text{sid} & \text{sname} & \text{rating} & \text{age} \\ \hline 22 & Dustin & 7 & 45.0 \\ 29 & Brutus & 1 & 33.0 \\ 31 & Lubber & 8 & 55.5 \\ 32 & Andy & 8 & 25.5 \\ 58 & Rusty & 10 & 35.0 \\ 64 & Horatio & 7 & 35.0 \\ 71 & Zorba & 10 & 16.0 \\ 74 & Horatio & 9 & 35.0 \\ 85 & Art & 3 & 25.5 \\ 95 & Bob & 3 & 63.5 \end{array} \right) = \begin{array}{l} \text{New Table} \\ \hline \text{sid} & \text{rating} \\ \hline 22 & 7 \\ 29 & 1 \\ 31 & 8 \\ 32 & 8 \\ 58 & 10 \\ 64 & 7 \\ 71 & 10 \\ 74 & 9 \\ 85 & 3 \\ 95 & 3 \end{array}$$

Select specified list of columns

```
SELECT sid , rating  
FROM Sailors ;
```

# SELECT - columns: order of selection

## order of list of columns

Order of columns need not be identical to the table stored in the database. Select rating, sid

Sailors				New Table	
sid	sname	rating	age	rating	sid
22	Dustin	7	45.0	7	22
29	Brutus	1	33.0	1	29
31	Lubber	8	55.5	8	31
32	Andy	8	25.5	8	32
58	Rusty	10	35.0	10	58
64	Horatio	7	35.0	7	64
71	Zorba	10	16.0	10	71
74	Horatio	9	35.0	9	74
85	Art	3	25.5	3	85
95	Bob	3	63.5	3	95

## Select specified list of columns

```
SELECT rating , sid  
FROM Sailors ;
```

# SELECT - columns: order of selection

list all columns

Order of columns need not be identical to the table stored in the database. Select rating, sid

The diagram illustrates the result of a SELECT query on the 'Sailors' table. On the left, the original 'Sailors' table is shown with columns: sid, sname, rating, and age. On the right, a 'New Table' is shown with columns: sid, sname, rating, and age. The 'New Table' contains the same data as the 'Sailors' table, but it only includes the 'rating' and 'sid' columns, as specified in the query.

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

=

New Table			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Select specified list of columns

```
SELECT sid , sname , rating , age  
FROM Sailors ;
```

# SELECT - columns: Wild character

list all columns

Order of columns need not be identical to the table stored in the database. Select rating, sid

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

=

New Table			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Select specified list of columns

```
SELECT * -- Specify regular expression; will construct list of column names  
FROM   Sailors;
```

# SELECT - rows: one specific row

select all rows that meet specific condition

Select one specific row; Example: sid = 58

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

= **New Table**

sid	sname	rating	age
58	Rusty	10	35.0

Select specified list of columns

```
SELECT sid, sname, rating, age  
FROM Sailors  
WHERE sid = 58;
```

# SELECT - rows: one specific row

select all rows that meet specific condition

Select one specific row; Example: sid = 58

$$\left( \begin{array}{l} \text{Sailors} \\ \hline \text{sid} & \text{sname} & \text{rating} & \text{age} \\ \hline 22 & Dustin & 7 & 45.0 \\ 29 & Brutus & 1 & 33.0 \\ 31 & Lubber & 8 & 55.5 \\ 32 & Andy & 8 & 25.5 \\ 58 & Rusty & 10 & 35.0 \\ 64 & Horatio & 7 & 35.0 \\ 71 & Zorba & 10 & 16.0 \\ 74 & Horatio & 9 & 35.0 \\ 85 & Art & 3 & 25.5 \\ 95 & Bob & 3 & 63.5 \end{array} \right) = \begin{array}{l} \text{New Table} \\ \hline \text{sid} & \text{sname} & \text{rating} & \text{age} \\ \hline 58 & Rusty & 10 & 35.0 \end{array}$$

Select specified list of columns

```
SELECT * -- wild character; list all columns of row containing sid=58  
FROM Sailors  
WHERE sid = 58;
```

# SELECT - rows: several rows

select all rows that meet specific condition

Select one specific row; Example: sname = Horatio

The diagram illustrates the selection of rows from a table. On the left, there is a large bracketed table labeled "Sailors" with columns sid, sname, rating, and age. The rows are:

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

On the right, the result of the selection is shown as a new table labeled "New Table". It contains only the rows where sname is "Horatio".

New Table			
sid	sname	rating	age
64	Horatio	7	35.0
74	Horatio	9	35.0

## Select specified list of columns

```
SELECT sid, sname, rating, age  
FROM Sailors  
WHERE sname = 'Horatio';
```

# SELECT - rows: several rows

select all rows that meet specific condition

Selecting rows with complex Example: sailors whose rating more than 6 and name should not be Horatio

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

=

New Table			
sid	sname	rating	age
22	Dustin	7	45.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
71	Zorba	10	16.0

Select specified list of columns

```
SELECT sid , sname , rating , age  
FROM Sailors  
WHERE ( rating > 6 AND sname <> 'Horatio' );
```

# SELECT - rows & columns

select specified rows and columns of a given condition

Selecting sname and rating columns of sailors whose age is greater than or equal to 30 and name should not be Horatio

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

=

New Table	
sname	rating
Dustin	7
Brutus	1
Lubber	8
Rusty	10
Bob	3

Select specified list of columns

```
SELECT sname, rating
FROM Sailors
WHERE (age >= 30 AND sname <> 'Horatio');
```

# SELECT - remove duplicates

## Definition

$r_i[a_k] = r_j[a_k] \quad \forall i \neq j; \quad \forall k = 1, 2, \dots, \text{number of columns}$

## Removing duplicates

a1	a2	a3
1	2	3
1	2	3
1	2	3
1	2	4

Rows 1, 2 & 3 are duplicate; fourth row is not a duplicate

# SELECT - remove duplicates

## Remove Duplicates - SQL

```
SELECT DISTINCT a1 , a2 , a3  
FROM tableA ;
```

## Remove Duplicates - SQL

a1	a2	a3
1	2	3
1	2	4

# SELECT - perform column SUM - 01

## Aggregate operations - SQL

```
SELECT SUM( a1 ) , SUM( a2 ) , SUM( a3 )  
FROM tableA ;
```

## Aggregation operation - SUM

SUM(a1)	SUM(a2)	SUM(a3)
4	8	13

new result table; columns SUM(a1), ... created! data type same as column a1

## SELECT - perform column SUM - 02

a1	a2	a3
1	2	3
1	2	3
1	2	3
1	2	4

```
SELECT SUM( a1 ) , SUM( a2 ) , SUM( a3 )
FROM    tableA ;
```

SUM(a1)	SUM(a2)	SUM(a3)
4	8	13

new result table; columns SUM(a1), ... created! data type same as column a1

# SELECT - perform column MIN, MAX, AVG

a1	a2	a3
1	2	3
1	2	3
1	2	3
4	5	6
⊥	⊥	⊥

```
SELECT MIN( a1 ) , MAX( a2 ) , AVG( a3 )  
FROM tableA ;
```

MIN(a1)	MAX(a2)	AVG(a3)
1	5	3.75

# SELECT - perform column - row COUNT - 01

a1	a2	a3
1	2	3
1	2	3
1	2	3
4	5	6

```
SELECT count( a1 )
FROM    tableA ;
```

COUNT(a1)
4

# SELECT - perform column - row COUNT - 02

a1	a2	a3
1	2	3
1	2	3
1	2	3
4	5	6
⊥	⊥	⊥

```
SELECT count(a1)
FROM    tableA ;
```

COUNT(a1)
4

# SELECT - perform row COUNT

a1	a2	a3
1	2	3
1	2	3
1	2	3
4	5	6
⊥	⊥	⊥

```
SELECT count(*) -- wild character counts rows  
FROM tableA ;
```

COUNT(*)
5

# SELECT - sort - 01

sort specified columns in ascending order (by default)

## Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

## Sorting

SELECT \* FROM Sailors ORDER BY rating

sid	sname	rating	age
29	Brutus	1	33.0
85	Art	3	25.5
95	Bob	3	63.5
22	Dustin	7	45.0
64	Horatio	7	35.0
31	Lubber	8	55.5
32	Andy	8	25.5
74	Horatio	9	35.0
58	Rusty	10	35.0
71	Zorba	10	16.0

# SELECT - sort - 02

sort specified columns in descending order

## Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

## Sorting

SELECT \* FROM Sailors ORDER BY rating DESC

sid	sname	rating	age
58	Rusty	10	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
31	Lubber	8	55.5
32	Andy	8	25.5
22	Dustin	7	45.0
64	Horatio	7	35.0
85	Art	3	25.5
95	Bob	3	63.5
29	Brutus	1	33.0

# SELECT - sort - 03

sort multiple columns

## Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

## Sorting

SELECT \* FROM Sailors ORDER BY rating, age

sid	sname	rating	age
29	Brutus	1	33.0
85	Art	3	25.5
95	Bob	3	63.5
64	Horatio	7	35.0
22	Dustin	7	45.0
32	Andy	8	25.5
31	Lubber	8	55.5
74	Horatio	9	35.0
71	Zorba	10	16.0
58	Rusty	10	35.0

## Grouping on Department attribtue

Department
EEE
CSE
EEE
CSE
...
...
CSE
...

On grouping on  
Department attribtue

Department
CSE
CSE
CSE
EEE
EEE
...
...
...

On grouping on  
Department attribtue

Department
CSE
<del>CSE</del>
<del>CSE</del>
EEE
<del>EEE</del>
...
...
...

# Grouping - 02

- Partitions rows of table into groups on the given column (cid)
- Each group (cid) consists of all rows having one particular assignment of values
- If there are no grouping attributes, entire relation is one group
- For each group (cid) produce **one row** consisting of
  - The grouping attributes' values for that group and
  - The aggregations over all row of that group for the aggregated column list (cid)

# Grouping Example

Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping

SELECT * FROM Sailors GROUP BY rating			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

# Grouping Example

Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping

SELECT * FROM Sailors GROUP BY rating			
sid	sname	rating	age
22	Dustin	7	45.0
64	Horatio	7	35.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

# Grouping Example

Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping

SELECT * FROM Sailors GROUP BY rating			
sid	sname	rating	age
22	Dustin	7	45.0
64	Horatio	7	35.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

# Grouping Example

## Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

## Grouping

SELECT * FROM Sailors GROUP BY rating			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
58	Rusty	10	35.0
74	Horatio	9	35.0
85	Art	3	25.5

# AS - re-nameing columns - 01

Create new column X using B, C

table1		
A	B	C
0	1	2
0	1	2
3	4	5

A	X
0	3
0	3
3	9

SQL statement

```
SELECT A, (B + C) AS X  
FROM table1 ;
```

# AS - re-nameing columns - 02

Create new columns X, Y using B, A and C, B

table1		
A	B	C
0	1	2
0	1	2
3	4	5

X	Y
1	1
1	1
1	1

SQL statement

```
SELECT (B - A) AS X, (C - B) AS Y  
FROM      table1 ;
```

# AS - re-nameing table - 03

Create new table using SELECT

table1		
A	B	C
0	1	2
0	1	2
3	4	5

table2	
A	X
0	3
0	3
3	9

SQL statement

```
CREATE TABLE table2 AS (SELECT A, (B + C) AS X FROM table1);
```