SQL WORKSHOP

Association of Data Science

WHAT IS SQL?

SQL: Structured Query Language

- Used to communicate with databases
- Common tasks
  - Retrieve data
  - Insert new data
  - Update or delete records
  - Analyze data sets
- Widely used in data science analytics, and software development

DATABASES EXAMPLES

- Social media feeds
- Online shopping catalogs
- Banking systems
- University course registration

OUR TABLES

people table

name	age	occupation	college
John Peterson	34	Marketing director	UNC
Emily Johnson	27	Data scientist	UNC Charlotte
James Smith	25	Civil engineer	NC A&T
Jenna Shaw	36	Data scientist	UNC Charlotte
Diego Williams	24	Entrepreneur	NC State

OUR TABLES

college table

college	city	state
UNC	Chapel Hill	NC
UNC Charlotte	Charlotte	NC
NC A&T	Greensboro	NC
Duke University	Durham	NC

BASIC SQL SYNTAX

```
SELECT column1, column2
FROM table_name
WHERE condition
ORDER BY column1;
```

SELECT → What you want

FROM → Where it's stored

WHERE → filter results

ORDER BY → sort results

SELECT + FROM

```
SELECT name, age
FROM people;
```

name	age
John Peterson	34
Emily Johnson	27
James Smith	25
Jenna Shaw	36
Diego Williams	24

- Returns all rows, showing name and age columns
- “Show me all the name and age columns in this table”

SELECT & FROM

```
SELECT *
FROM people;
```

- Returns all rows, showing all columns
- “Show me all the columns in this table”

name	age	occupation	college
John Peterson	34	Marketing director	UNC
Emily Johnson	27	Data scientist	UNC Charlotte
James Smith	25	Civil engineer	NC A&T
Jenna Shaw	36	Data scientist	UNC Charlotte
Diego Williams	24	Entrepreneur	NC State

WHERE

```
SELECT *
FROM people;
WHERE age > 30;
```

- Comparison operators: =, >, <, >=, <=, !=
- Text comparison use quotes: WHERE college = ‘UNC Charlotte’

name	age	occupation	college
John Peterson	34	Marketing director	UNC
Jenna Shaw	36	Data scientist	UNC Charlotte

ORDER BY

```
SELECT *
FROM people;
ORDER BY age DESC;
```

- ASC = ascending (default)
- DESC = descending

name	age	occupation	college
Jenna Shaw	36	Data scientist	UNC Charlotte
John Peterson	34	Marketing director	UNC
Emily Johnson	27	Data scientist	UNC Charlotte

....

ALIASES & LIMIT

```
SELECT name AS employee, age
FROM people;
LIMIT 2;
```

name	age
John Peterson	34
Emily Johnson	27

- AS renames columns in the result
- LIMIT shows only a certain number of rows

AGGREGATION FUNCTIONS

```
SELECT AVG(age)
FROM people;
```

Returns:29.2

- COUNT() → number of rows
- SUM() → calculates sum of all non-NULL values
- AVG() → average of all non-NULL values
- MAX() →highest values
- MIN() → lowest values

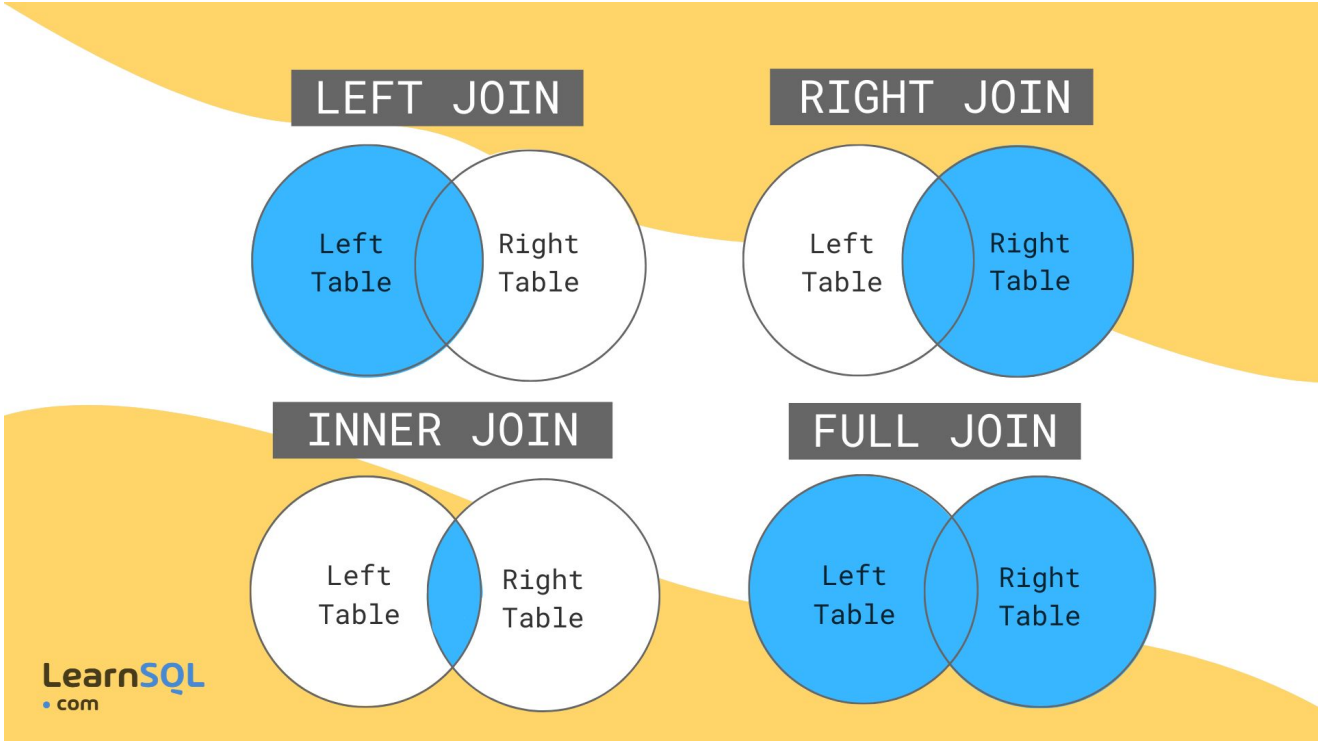
GROUP BY & HAVING

```
SELECT occupation,
COUNT(*) AS num_people

FROM people;

GROUP BY occupation

HAVING COUNT(*) > 1;
```

- Group by → groups rows with the same values
- Having → filters groups (after aggregation)

occupation	num_people
Data scientist	2

JOINS


LEFT JOIN
Left Table
Right Table
RIGHT JOIN
Left Table
Right Table
INNER JOIN
Left Table
Right Table
FULL JOIN
Left Table
Right Table
LearnSQL.com

INNER JOIN

```
SELECT p.name, c.college, c.city
FROM people p;
INNER JOIN colleges c
ON p.college = c.college;
```

name	college	city
John Peterson	UNC	Chapel Hill
Emily Johnson	UNC Charlotte	Charlotte
James Smith	NC A&T	Greensboro
Jenna Shaw	UNC Charlotte	Charlotte

LEFT JOIN

```
SELECT p.name, c.college,
c.city
FROM people p;
LEFT JOIN colleges c
ON p.college = c.college;
```

name	college	city
John Peterson	UNC	Chapel Hill
Emily Johnson	UNC Charlotte	Charlotte
James Smith	NC A&T	Greensboro
Jenna Shaw	UNC Charlotte	Charlotte
Diego Williams	NC state	NULL

RIGHT JOIN

```
SELECT p.name, c.college, c.city
FROM people p;
RIGHT JOIN colleges c
ON p.college = c.college;
```

name	college	city
John Peterson	UNC	Chapel Hill
Emily Johnson	UNC Charlotte	Charlotte
James Smith	NC A&T	Greensboro
Jenna Shaw	UNC Charlotte	Charlotte
NULL	Duke University	Durham

- Note, right joins are not very commonly used and can be converted to left joins

FULL OUTER JOIN

```
SELECT p.name, c.college,
c.city
FROM people p;
FULL OUTER JOIN colleges c
ON p.college = c.college;
```

name	college	city
John Peterson	UNC	Chapel Hill
Emily Johnson	UNC Charlotte	Charlotte
James Smith	NC A&T	Greensboro
Jenna Shaw	UNC Charlotte	Charlotte
Diego Williams	NC state	NULL
NULL	Duke University	Durham

PRACTICE

- We'll practice our skills using SQL Island! (https://sql-island.informatik.uni-kl.de/)

WRAP UP

Today we covered

- What SQL is
- How to write basic queries
- How to filter, sort, group, and join data
- Applying what we learned to a real database

Next Steps

- Escape SQL Island
- Try out online SQL sandboxes (SQL Island has one)