Carnegie Mellon University

DATABASE SYSTEMS

Modern SQL

LECTURE #02 » 15-445/645 FALL 2025 » PROF. ANDY PAVLO

ADMINISTRIVIA

Project #0 is due Sunday Sept 7th @ 11:59pm

Homework #1 is due Sunday Sept 7th @ 11:59pm

No in-class lecture next Wednesday Sept 3rd

LAST CLASS

We introduced the Relational Model as the superior data model for databases.

We then showed how Relational Algebra is the building blocks that will allow us to query and modify a relational database.

SQL HISTORY

In 1971, IBM created its first relational query language called SQUARE.

IBM then created "SEQUEL" in 1972 for IBM System R prototype DBMS.
→ Structured English Query Language

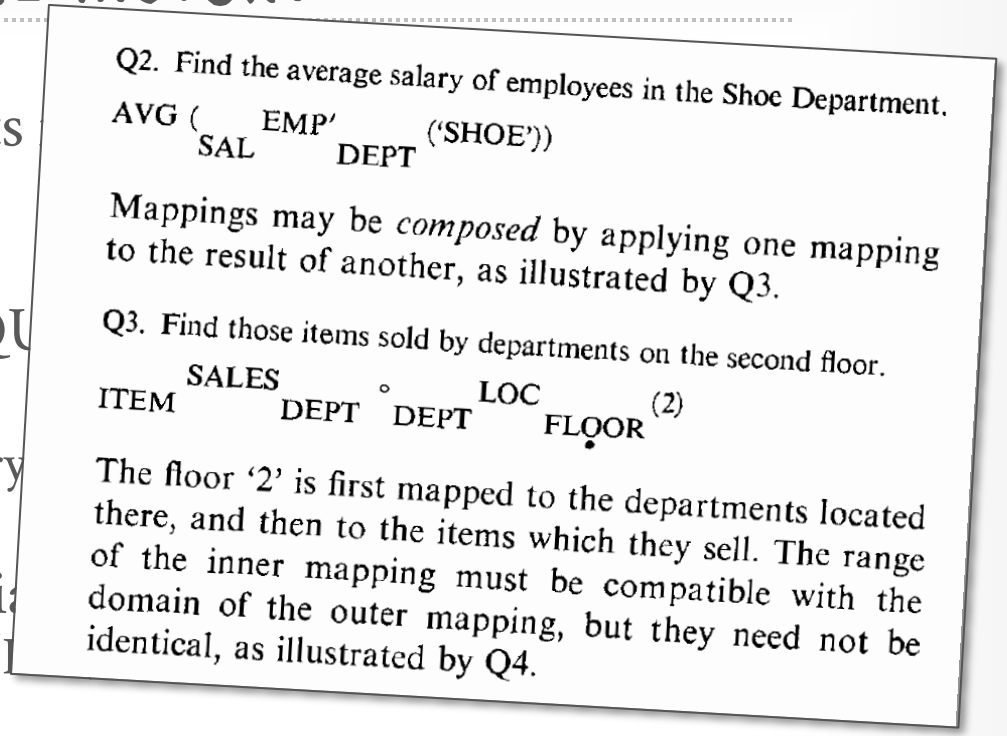
IBM releases commercial SQL-based DBMSs:
→ System/38 (1979), SQL/DS (1981), and DB2 (1983).

SQL HISTORY

In 1971, IBM created its
called SQUARE.

IBM then created "SEQU
prototype DBMS.
→ Structured English Query

IBM releases commerci
→ System/38 (1979), SQL/



Q2. Find the average salary of employees in the Shoe Department.

$$\text{AVG}\ (_{\text{SAL}}\ \text{EMP}'_{\text{DEPT}}\ (\text{'SHOE'}))$$

Mappings may be *composed* by applying one mapping to the result of another, as illustrated by Q3.

Q3. Find those items sold by departments on the second floor.

$$\text{ITEM}^{\text{SALES}}_{\text{DEPT}} \circ_{\text{DEPT}}^{\text{LOC}}{}_{\text{FLOOR}}\ (2)$$

The floor '2' is first mapped to the departments located there, and then to the items which they sell. The range of the inner mapping must be compatible with the domain of the outer mapping, but they need not be identical, as illustrated by Q4.

SQL HISTORY

ANSI Standard in 1986. ISO in 1987
→ Structured Query Language

Current standard is **SQL:2023**
→ **SQL:2023** → Property Graph Queries, Muti-Dim. Arrays
→ **SQL:2016** → JSON, Polymorphic tables
→ **SQL:2011** → Temporal DBs, Pipelined DML
→ **SQL:2008** → Truncation, Fancy Sorting
→ **SQL:2003** → XML, Windows, Sequences, Auto-Gen IDs.
→ **SQL:1999** → Regex, Triggers, OO

The minimum language syntax a system needs to say that it supports SQL is **SQL-92**.

SQL HISTORY

...37

...Muti-Dim. Arrays

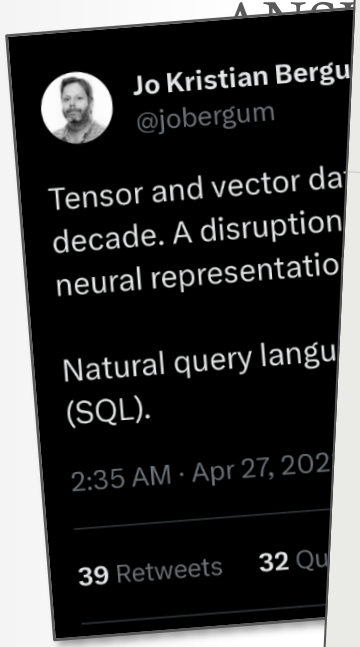
...DML

...es, Auto-Gen IDs.

The minimum language syntax a system needs to say that it supports SQL is **SQL-92**.

SQL HISTORY

The minimum langu
that it supports SQL



NEWS | COMPUTING

The Rise of SQL

It's become the second programming language everyone needs to know

BY RINA DIANE CABALLAR | 23 AUG 2022 | 3 MIN READ

ISTOCK

SHARE THIS STORY

TAGS

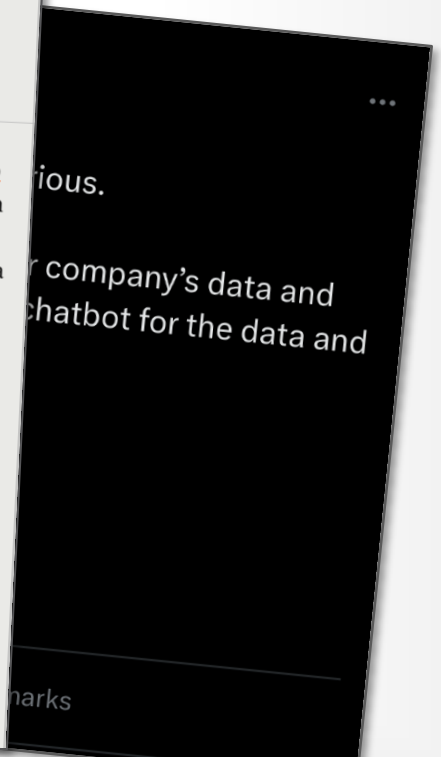
TOP PROGRAMMING LANGUAGES

SQL

SQL dominated the jobs ranking in *IEEE Spectrum*'s interactive rankings of the top programming languages this year. Normally, the top position is occupied by Python or other mainstays, such as C, C++, Java, and JavaScript, but the sheer number of times employers said they wanted developers with SQL skills, albeit in addition to a more general-purpose language, boosted it to No. 1.

So what's behind SQL's soar to the top? The ever-increasing use of databases, for one. SQL has become the primary query language for accessing and managing data stored in such databases—specifically relational databases, which represent data in table form with rows and columns. Databases serve as the foundation of many enterprise applications and are increasingly found in other places as well, for example taking the place of traditional file systems in smartphones.

"This ubiquity means that every software developer will have to interact with databases no matter the field, and SQL is the de facto standard for interacting with databases," says Andy Pavlo, a professor specializing in database management at the Carnegie Mellon University (CMU) School of Computer Science and a member of the CMU database group.



RELATIONAL LANGUAGES

Data Manipulation Language (DML)
Data Definition Language (DDL)
Data Control Language (DCL)

Also includes:
→ View definition
→ Integrity & Referential Constraints
→ Transactions

Important: SQL is based on **bags** (duplicates) not **sets** (no duplicates).

TODAYS AGENDA

Aggregations + Group By

String / Date / Time Operations

Output Control + Redirection

Nested Queries

Lateral Joins

Common Table Expressions

Window Functions

⚡DB Flash Talk: dbt

EXAMPLE DATABASE

student(sid,name,login,gpa)

sid	name	login	age	gpa
53666	RZA	rza@cs	56	4.0
53688	Taylor	swift@cs	35	3.9
53655	Tupac	shakur@cs	25	3.5

course(cid,name)

cid	name
15-445	Database Systems
15-721	Advanced Database Systems
15-826	Data Mining
15-799	Special Topics in Databases

enrolled(sid,cid,grade)

sid	cid	grade
53666	15-445	C
53688	15-721	A
53688	15-826	B
53655	15-445	B
53666	15-721	C

AGGREGATES

Functions that return a single value from a bag of tuples:

→ **`AVG(col)`** → Return the average col value.
→ **`MIN(col)`** → Return minimum col value.
→ **`MAX(col)`** → Return maximum col value.
→ **`SUM(col)`** → Return sum of values in col.
→ **`COUNT(col)`** → Return # of values for col.

AGGREGATES

Aggregate functions can (almost) only be used in the **SELECT** output list.

Get # of students with a "@cs" login:

```
SELECT COUNT(login) AS cnt
  FROM student WHERE login LIKE '%@cs'
```

AGGREGATES

Aggregate functions can (almost) only be used in the **SELECT** output list.

Get # of students with a "@cs" login:

```
SELECT COUNT(login) AS cnt
  FROM student WHERE login LIKE '%@cs'
```

```
SELECT COUNT(*) AS cnt
  FROM student WHERE login LIKE '%@cs'
```

AGGREGATES

Aggregate functions can (almost) only be used in the **SELECT** output list.

Get # of students with a "@cs" login:

```
SELECT COUNT(login) AS cnt
  FROM student WHERE login LIKE '%@cs'
```

```
SELECT COUNT(*) AS cnt
  FROM student WHERE login LIKE '%@cs'
```

```
SELECT COUNT(1) AS cnt
  FROM student WHERE login LIKE '%@cs'
```

AGGREGATES

Aggregate functions can (almost) only be used in the **SELECT** output list.

Get # of students with a "@cs" login:

```
SELECT COUNT(login) AS cnt
  FROM student WHERE login LIKE '%@cs'
```

```
SELECT COUNT(*) AS cnt
  FROM student WHERE login LIKE '%@cs'
```

```
SELECT COUNT(1) AS cnt
  FROM student WHERE login LIKE '%@cs'
```

```
SELECT COUNT(1+1+1) AS cnt
  FROM student WHERE login LIKE '%@cs'
```

AGGREGATES

Output of other columns outside of an aggregate is undefined.

Get the average GPA of students enrolled in each course.

```
SELECT AVG(s.gpa), e.cid
  FROM enrolled AS e JOIN student AS s
    ON e.sid = s.sid
```

AVG(s.gpa)	e.cid
3.86	???

AGGREGATES

Output of other columns outside of an aggregate is undefined.

Get the average GPA of students enrolled in each course.

```
SELECT AVG(s.gpa), e.cid
  FROM enrolled AS e JOIN student AS s
    ON e.sid = s.sid
```

AVG(s.gpa)	e.cid
3.86	???

```
SELECT AVG(s.gpa), ANY_VALUE(e.cid)
  FROM enrolled AS e JOIN student AS s
    ON e.sid = s.sid
```

AVG(s.gpa)	e.cid
3.86	15-445

GROUP BY

Project tuples into subsets and calculate aggregates against each subset.

```
SELECT AVG(s.gpa), e.cid
  FROM enrolled AS e JOIN student AS s
    ON e.sid = s.sid
 GROUP BY e.cid
```

e.sid	s.sid	s.gpa	e.cid
53435	53435	2.25	15-721
53439	53439	2.70	15-721
56023	56023	2.75	15-826
59439	59439	3.90	15-826
53961	53961	3.50	15-826
58345	58345	1.89	15-445

GROUP BY

Project tuples into subsets and calculate aggregates against each subset.

```
SELECT AVG(s.gpa), e.cid
  FROM enrolled AS e JOIN student AS s
    ON e.sid = s.sid
 GROUP BY e.cid
```

e.sid	s.sid	s.gpa	e.cid
53435	53435	2.25	15-721
53439	53439	2.70	15-721
56023	56023	2.75	15-826
59439	59439	3.90	15-826
53961	53961	3.50	15-826
58345	58345	1.89	15-445

AVG(s.gpa)	e.cid
2.46	15-721
3.39	15-826
1.89	15-445

GROUPING SETS

Specify multiple groupings in a single query instead of using **UNION ALL** to combine the results of several individual **GROUP BY** queries.

```
SELECT c.name AS c_name, e.grade,
       COUNT(*) AS num_students
  FROM enrolled AS e
  JOIN course AS c ON e.cid = c.cid
 GROUP BY GROUPING SETS (
  (c.name, e.grade), -- By course and grade
  (c.name),          -- By course only
  ()                 -- Overall total
);
```

GROUPING SETS

Specify multiple groupings in a single query instead of using **UNION ALL** to combine the results of several individual **GROUP BY** queries.

```
SELECT c.name AS c_name, e.grade,
       COUNT(*) AS num_students
  FROM enrolled AS e
  JOIN course AS c ON e.cid = c.cid
 GROUP BY GROUPING SETS (
  (c.name, e.grade), -- By course a
  (c.name),          -- By course c
  ()                 -- Overall total
);
```

cid	grade	num_students
null	*null*	5
15-721	C	1
15-826	B	1
15-445	B	1
15-445	C	1
15-721	A	1
15-445	*null*	2
15-826	*null*	1
15-721	*null*	2

GROUPING SETS

Specify multiple groupings in a single query instead of using **UNION ALL** to combine the results of several individual **GROUP BY** queries.

```
SELECT c.name AS c_name, e.grade,
       COUNT(*) AS num_students
  FROM enrolled AS e
  JOIN course AS c ON e.cid = c.cid
 GROUP BY GROUPING SETS (
  (c.name, e.grade),  -- By course a
  (c.name),           -- By course c
  ()                  -- Overall total
);
```

cid	grade	num_students
null	*null*	5
15-721	C	1
15-826	B	1
15-445	B	1
15-445	C	1
15-721	A	1
15-445	*null*	2
15-826	*null*	1
15-721	*null*	2

HAVING

Filters results based on aggregation computation.

Like a **WHERE** clause for a **GROUP BY**

```
SELECT AVG(s.gpa) AS avg_gpa, e.cid
  FROM enrolled AS e, student AS s
 WHERE e.sid = s.sid
   AND avg_gpa > 3.9
 GROUP BY e.cid
```

HAVING

Filters results based on aggregation computation.

Like a **WHERE** clause for a **GROUP BY**

```
SELECT AVG(s.gpa) AS avg_gpa, e.cid
  FROM enrolled AS e, student AS s
 WHERE e.sid = s.sid
   AND avg_gpa > 3.9
 GROUP BY e.cid
```

HAVING

Filters results based on aggregation computation.

Like a **WHERE** clause for a **GROUP BY**

```
SELECT AVG(s.gpa) AS avg_gpa, e.cid
  FROM enrolled AS e, student AS s
 WHERE e.sid = s.sid
 GROUP BY e.cid
 HAVING avg_gpa > 3.9;
```

HAVING

Filters results based on aggregation computation.

Like a **WHERE** clause for a **GROUP BY**

```
SELECT AVG(s.gpa) AS avg_gpa, e.cid
  FROM enrolled AS e, student AS s
 WHERE e.sid = s.sid
 GROUP BY e.cid
 HAVING AVG(s.gpa) > 3.9;
```

AVG(s.gpa)	e.cid
3.75	15-415
3.950000	15-721
3.900000	15-826

avg_gpa	e.cid
3.950000	15-721

STRING OPERATIONS

	String Case	String Quotes
SQL-92	**Sensitive**	**Single Only**
Postgres	Sensitive	Single Only
MySQL	Insensitive	Single/Double
SQLite	Sensitive	Single/Double
MSSQL	Sensitive	Single Only
Oracle	Sensitive	Single Only

```
WHERE UPPER(name) = UPPER('TuPaC')
```
SQL-92

```
WHERE name = "TuPaC"
```
MySQL

STRING OPERATIONS

LIKE provides string matching with special match operators:

→ `'%'` Matches any substring (including empty strings).

→ `'_'` Match any one character

```
SELECT * FROM enrolled AS e
 WHERE e.cid LIKE '15-%';
```

```
SELECT * FROM student AS s
 WHERE s.login LIKE '%@c_';
```

SIMILAR TO allows for regular expression matching.

→ In the SQL standard but not all systems support it.

→ Other systems also support POSIX-style regular expressions.

```
SELECT * FROM student AS s
 WHERE login SIMILAR TO
       '[\w]{3}@cs';
```

STRING OPERATIONS

SQL-92 defines string functions.

→ Many DBMSs also have their own unique functions

Can be used in either output and predicates:

```
SELECT SUBSTRING(name,1,5) AS abbrv_name
  FROM student WHERE sid = 53688
```

```
SELECT * FROM student AS s
 WHERE UPPER(s.name) LIKE 'KAN%'
```

STRING OPERATIONS

SQL standard defines the `||` operator for concatenating two or more strings together.

SQL-92

```
SELECT name FROM student
 WHERE login = LOWER(name) || '@cs'
```

MSSQL

```
SELECT name FROM student
 WHERE login = LOWER(name) + '@cs'
```

MySQL

```
SELECT name FROM student
 WHERE login = CONCAT(LOWER(name), '@cs')
```

DATE/TIME OPERATIONS

Operations to manipulate and modify **DATE**/**TIME** attributes.

Can be used in both output and predicates.

Support/syntax varies wildly…

Demo: Compute the number of days since the beginning of the year.

OUTPUT CONTROL

ORDER BY <column*> [ASC|DESC]

→ Sort tuples by the values in one or more of their columns.

```
SELECT sid, name FROM student
 WHERE login LIKE '%@cs'
 FETCH FIRST 10 ROWS ONLY;
```

FETCH {FIRST|NEXT} <#> ROWS ONLY
OFFSET <#> ROWS

→ Limit # of tuples returned in output.
→ Can set an offset to return a "range"

```
SELECT sid, name FROM student
 WHERE login LIKE '%@cs'
 ORDER BY gpa
OFFSET 5 ROWS
 FETCH FIRST 5 ROWS WITH TIES;
```

MSSQL

```
SELECT TOP 10 sid, name
  FROM student
 WHERE login LIKE '%@cs';
```

OUTPUT REDIRECTION

Store query results in another table:

→ Table must not already be defined.

→ Table will have the same # of columns with the same types as the input.

SQL-92

```
SELECT DISTINCT cid INTO CourseIds
  FROM enrolled;
```

MySQL

```
CREATE TABLE CourseIds (
  SELECT DISTINCT cid FROM enrolled);
```

OUTPUT REDIRECTION

Store query results in another table:

→ Table must not already be defined.

→ Table will have the same # of columns with the same types as the input.

SQL-92

```
SELECT DISTINCT cid INTO CourseIds
  FROM enrolled;
```

Postgres

```
SELECT DISTINCT cid
  INTO TEMPORARY CourseIds
  FROM enrolled;
```

MySQL

```
CREATE TABLE CourseIds (
  SELECT DISTINCT cid FROM enrolled);
```

NESTED QUERIES

Invoke a query inside of another query to compose more complex computations.

→ Inner queries can appear (almost) anywhere in query.

```
SELECT name FROM student WHERE
 sid IN (SELECT sid FROM enrolled)
```

NESTED QUERIES

Invoke a query inside of another query to compose more complex computations.

→ Inner queries can appear (almost) anywhere in query.

Outer Query

```
SELECT name FROM student WHERE
 sid IN (SELECT sid FROM enrolled)
```

Inner Query

```
SELECT sid,
       (SELECT name FROM student AS s
         WHERE s.sid = e.sid) AS name
  FROM enrolled AS e;
```

NESTED QUERIES

Get the names of students in '15-445'

```
SELECT name FROM student
 WHERE ...
```

sid in the set of people that take 15-445

NESTED QUERIES

Get the names of students in '15-445'

```
SELECT name FROM student
 WHERE ...
   SELECT sid FROM enrolled
    WHERE cid = '15-445'
```

NESTED QUERIES

Get the names of students in '15-445'

```
SELECT name FROM student
 WHERE sid IN (
   SELECT sid FROM enrolled
    WHERE cid = '15-445'
 )
```

NESTED QUERIES

Get the names of students in '15-445'

```
SELECT name FROM student
 WHERE sid IN (
   SELECT sid FROM enrolled
    WHERE cid = '15-445'
 )
```

NESTED QUERIES

ALL→ The expression must be true for all rows in the sub-query.

ANY→ The expression must be true for at least one row in the sub-query.

IN→ Equivalent to '**=ANY()**' .

EXISTS→ At least one row is returned without comparing it to an attribute in outer query.

NESTED QUERIES

Find student record with the highest id that is enrolled in at least one course.

```
SELECT sid, name FROM student
 WHERE ...
```

"Is the highest enrolled sid"

NESTED QUERIES

Find student record with the highest id that is enrolled in at least one course.

```
SELECT sid, name FROM student
 WHERE sid =
   SELECT MAX(sid) FROM enrolled
```

NESTED QUERIES

Find student record with the highest id that is enrolled in at least one course.

```
SELECT sid, name FROM student
 WHERE sid IN (
   SELECT MAX(sid) FROM enrolled
);
```

sid	name
53688	Bieber

NESTED QUERIES

Find student record with the highest id that is enrolled in at least one course.

```
SELECT sid, name FROM student
 WHERE sid IN (
   SELECT sid FROM enrolled
    ORDER BY sid DESC FETCH FIRST 1 ROW ONLY
);
```

NESTED QUERIES

Find student record with the highest id that is enrolled in at least one course.

```
SELECT sid, name FROM student
 WHERE sid IN (
   SELECT MAX(sid) FROM enrolled
);
```

```
SELECT sid, name FROM student
 WHERE sid IN (
   SELECT MAX(sid) FROM enrolled
);
```

```
SELECT student.sid, name
  FROM student
  JOIN (SELECT MAX(sid) AS sid
          FROM enrolled) AS max_e
    ON student.sid = max_e.sid;
```

NESTED QUERIES

Find all courses that have no students enrolled in it.

```
SELECT * FROM course
 WHERE ...
```

"with no tuples in the enrolled table"

cid	name
15-445	Database Systems
15-721	Advanced Database Systems
15-826	Data Mining
15-799	Special Topics in Databases

sid	cid	grade
53666	15-445	C
53688	15-721	A
53688	15-826	B
53655	15-445	B
53666	15-721	C

NESTED QUERIES

Find all courses that have no students enrolled in it.

```
SELECT * FROM course
 WHERE NOT EXISTS(
   tuples in the enrolled table
);
```

NESTED QUERIES

Find all courses that have no students enrolled in it.

```
SELECT * FROM course
 WHERE NOT EXISTS(
   SELECT * FROM enrolled
    WHERE course.cid = enrolled.cid
);
```

cid	name
15-799	Special Topics in Databases

NESTED QUERIES

Find all courses that have no students enrolled in it.

```
SELECT * FROM course
 WHERE NOT EXISTS(
   SELECT * FROM enrolled
    WHERE course.cid = enrolled.cid
);
```

cid	name
15-799	Special Topics in Databases

LATERAL JOINS

The **LATERAL** operator allows a nested query to reference attributes in other nested queries that precede it (according to position in the query).

→ You can think of it like a **for** loop that allows you to invoke another query for each tuple in a table.

```
SELECT * FROM
  (SELECT 1 AS x) AS t1,
  LATERAL (SELECT t1.x+1 AS y) AS t2;
```

t1.x	t2.y
1	2

```
for x in [1]:
  for y in [x+1]:
    print(x,y)
```

LATERAL JOIN

Calculate the number of students enrolled in each course and the average GPA. Sort output by the courses' enrollment count in ascending order.

```
SELECT * FROM course AS c,
  For each course:
    → Compute the # of enrolled students

  For each course:
    → Compute the average gpa of enrolled students
```

LATERAL JOIN

Calculate the number of students enrolled in each course and the average GPA. Sort output by the courses' enrollment count in ascending order.

```
SELECT * FROM course AS c,
  LATERAL (SELECT COUNT(*) AS cnt FROM enrolled
             WHERE enrolled.cid = c.cid) AS t1,
  LATERAL (SELECT AVG(gpa) AS avg FROM student AS s
              JOIN enrolled AS e ON s.sid = e.sid
             WHERE e.cid = c.cid) AS t2
 ORDER BY cnt ASC;
```

LATERAL JOIN

Calculate the number of students enrolled in each course and the average GPA. Sort output by the courses' enrollment count in ascending order.

```
SELECT * FROM course AS c,
  LATERAL (SELECT COUNT(*) AS cnt FROM enrolled
             WHERE enrolled.cid = c.cid) AS t1,
  LATERAL (SELECT AVG(gpa) AS avg FROM student AS s
             JOIN enrolled AS e ON s.sid = e.sid
             WHERE e.cid = c.cid) AS t2
 ORDER BY cnt ASC;
```

LATERAL JOIN

Calculate the number of students enrolled in each course and the average GPA. Sort output by the courses' enrollment count in ascending order.

```
SELECT * FROM course AS c,
  LATERAL (SELECT COUNT(*) AS cnt FROM enrolled
             WHERE enrolled.cid = c.cid) AS t1,
  LATERAL (SELECT AVG(gpa) AS avg FROM student AS s
             JOIN enrolled AS e ON s.sid = e.sid
             WHERE e.cid = c.cid) AS t2
 ORDER BY cnt ASC;
```

LATERAL JOIN

Calculate the number of students enrolled in each course and the average GPA. Sort output by the courses' enrollment count in ascending order.

```
SELECT * FROM course AS c,
  LATERAL (SELECT COUNT(*) AS cnt FROM enrolled
             WHERE enrolled.cid = c.cid) AS t1,
  LATERAL (SELECT AVG(gpa) AS avg FROM student AS s
             JOIN enrolled AS e ON s.sid = e.sid
             WHERE e.cid = c.cid) AS t2
 ORDER BY cnt ASC;
```

LATERAL JOIN

Calculate the number of students enrolled in each course and the average GPA. Sort output by the courses' enrollment count in ascending order.

```
SELECT * FROM course AS c,
  LATERAL (SELECT COUNT(*) A
             WHERE enrolled.cid = c.cid) AS t1,
  LATERAL (SELECT AVG(gpa) AS avg FROM student AS s
             JOIN enrolled AS e ON s.sid = e.sid
             WHERE e.cid = c.cid) AS t2
 ORDER BY cnt ASC;
```

cid	name	cnt	avg
15-799	Special Topics in Databases	0	null
15-826	Data Mining	1	3.9
15-445	Database Systems	2	3.75
15-721	Advanced Database Systems	2	3.95

COMMON TABLE EXPRESSIONS

Specify a temporary result set that can then be referenced by another part of that query.

→ Bind/alias output columns to names before the **AS** keyword.

Alternative to nested queries, views, and explicit temp tables.

```
WITH cteName (col1, col2) AS (
   SELECT 1, 2
)
SELECT col1 + col2 FROM cteName
```

COMMON TABLE EXPRESSIONS

Specify a temporary result set that can then be referenced by another part of that query.

→ Bind/alias output columns to names before the **AS** keyword.

Alternative to nested queries, views, and explicit temp tables.

```
WITH cteName (col1, col2) AS (
   SELECT 1, 2
)
SELECT col1 + col2 FROM cteName
```

COMMON TABLE EXPRESSIONS

Find student record with the highest id that is enrolled in at least one course.

```
WITH maxCTE (maxId) AS (
   SELECT MAX(sid) FROM enrolled
)
SELECT name FROM student AS s
  JOIN maxCTE ON s.sid = maxCTE.maxId;
```

COMMON TABLE EXPRESSIONS

Find student record with the highest id that is enrolled in at least one course.

```
WITH maxCTE (maxId) AS (
   SELECT MAX(sid) FROM enrolled
)
SELECT name FROM student AS s
  JOIN maxCTE ON s.sid = maxCTE.maxId;
```

WINDOW FUNCTIONS

Performs a calculation across a set of tuples that are related to the current tuple, without collapsing them into a single output tuple, to support running totals, ranks, and moving averages.

→ Like an aggregation but tuples are not grouped into a single output tuples.

How to "slice" up data
Can also sort tuples

```
SELECT FUNC-NAME(...) OVER (...)
  FROM tableName;
```

Aggregation Functions
Special Functions

WINDOW FUNCTIONS

Aggregation functions:

→ Anything that we discussed earlier

Special window functions:

→ **ROW_NUMBER()** → # of the current row

→ **RANK()** → Order position of the current row.

sid	cid	grade	row_num
53666	15-445	C	1
53688	15-721	A	2
53688	15-826	B	3
53655	15-445	B	4
53666	15-721	C	5

```
SELECT *, ROW_NUMBER() OVER () AS row_num
  FROM enrolled;
```

WINDOW FUNCTIONS

Aggregation functions:

→ Anything that we discussed earlier

Special window functions:

→ **ROW_NUMBER()** → # of the current row

→ **RANK()** → Order position of the current row.

sid	cid	grade	row_num
53666	15-445	C	1
53688	15-721	A	2
53688	15-826	B	3
53655	15-445	B	4
53666	15-721	C	5

```
SELECT *, ROW_NUMBER() OVER () AS row_num
  FROM enrolled;
```

WINDOW FUNCTIONS

The **OVER** keyword specifies how to group together tuples when computing the window function.

Use **PARTITION BY** to specify group.

```
SELECT cid, sid,
    ROW_NUMBER() OVER (PARTITION BY cid)
  FROM enrolled
 ORDER BY cid;
```

WINDOW FUNCTIONS

The **OVER** keyword specifies how to group together tuples when computing the window function.

Use **PARTITION BY** to specify group.

cid	sid	row_number
15-445	53666	1
15-445	53655	2
15-721	53688	1
15-721	53666	2
15-826	53688	1

```
SELECT cid, sid,
    ROW_NUMBER() OVER (PARTITION BY cid)
  FROM enrolled
 ORDER BY cid;
```

WINDOW FUNCTIONS

You can also include an **ORDER BY** in the window grouping to sort entries in each group.

```
SELECT *,
    ROW_NUMBER() OVER (ORDER BY cid)
  FROM enrolled
 ORDER BY cid;
```

WINDOW FUNCTIONS

Find the student with second highest grade for each course.

Group tuples by cid
Then sort by grade

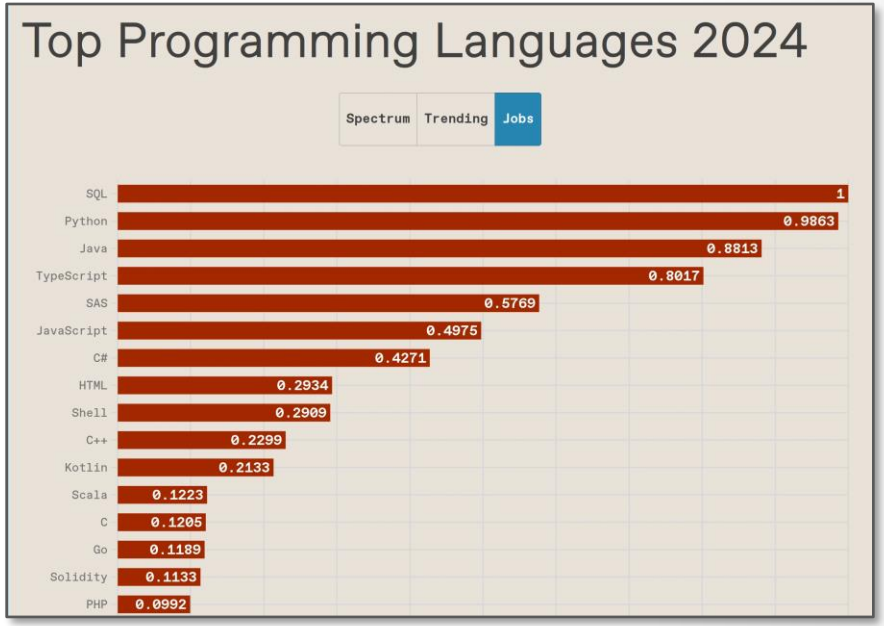
```
SELECT * FROM (
  SELECT *, RANK() OVER (PARTITION BY cid
            ORDER BY grade ASC) AS rank
    FROM enrolled) AS ranking
 WHERE ranking.rank = 2
```

CONCLUSION

SQL is a hot language.

→ Lots of NL2SQL tools, but writing SQL is not going away.

You should (almost) always strive to compute your answer as a single SQL statement.



https://spectrum.ieee.org/top-programming-languages-2024

HOMEWORK #1

Write SQL queries to perform basic data analysis.

→ Write the queries locally using DuckDB.
→ Submit them to Gradescope
→ You can submit multiple times and use your best score.

Due: Sunday Sept 7th @ 11:59pm

https://15445.courses.cs.cmu.edu/fall2025/homework1

NEXT CLASS

We will begin our journey to understanding the internals of database systems starting with Storage!

No In-Class Lecture!