Data Analysis Fundamentals : SQL

Global Engineering Challenge - Vaccine Distribution Plan

UNIVERSITY OF TORONTO

Slide by Sarah.L

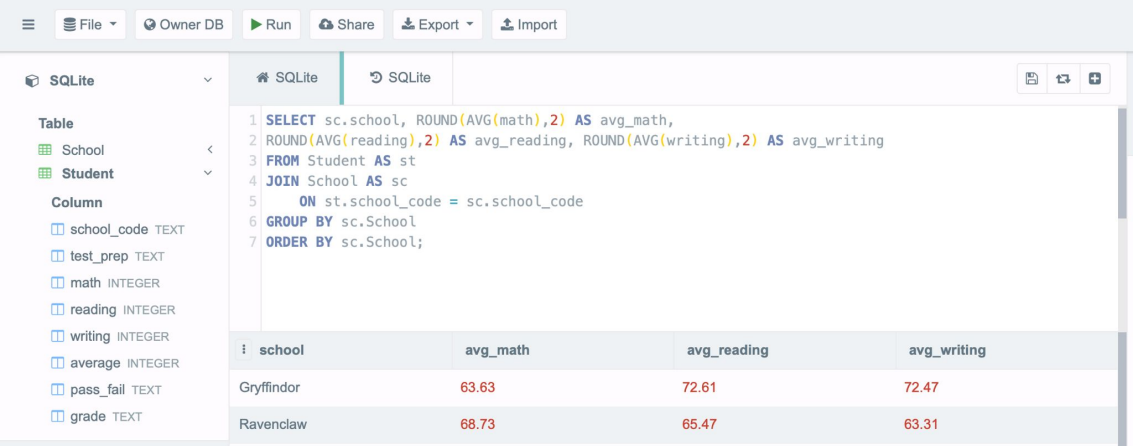
What is SQL?

- **S**tructured **Q**uery **L**anguage, **SQL**, is the standard and most widely **used** programming language for relational databases. It is **used** to manage and organize data in all sorts of systems in which various data relationships exist.

BigQuery

Basic Commands in SQL

- **SQL commands** are the instructions used to communicate with a database to perform tasks, functions, and queries with data
- Basic functions:
  - SELECT
  - FROM
  - WHERE
  - JOIN
  - GROUP BY
  - ORDER BY



UNIVERSITY OF TORONTO

Slide by Sarah.L

SQL Order Of Operations

ORDER	CLAUSE	FUNCTION
1	from	Choose and join tables to get base data.
2	where	Filters the base data.
3	group by	Aggregates the base data.
4	having	Filters the aggregated data.
5	select	Returns the final data.
6	order by	Sorts the final data.
7	limit	Limits the returned data to a row count.

https://www.sisense.com/blog/sql-query-order-of-operations/

SELECT

The **SELECT** statement is used to select data from a database.

Example:

SELECT *

SELECT test_prep, grade

Run Share Export Import

SQLite SQLite

```
SELECT *
FROM Student;
```

school_...	test_prep	math	reading	writing	average	pass_fail	grade
GF	None	72	72	74	73	Pass	B
GF	Completed	69	90	88	82	Pass	A

SQLite SQLite

```
SELECT test_prep, grade
FROM Student;
```

test_prep	grade
None	B
Completed	A

FROM

The **FROM** command is used to specify which table to select or delete data from

Example:

```
SELECT *
FROM Student;
```

Run Share Export Import

SQLite SQLite

```
SELECT *
FROM Student;
```

school_...	test_prep	math	reading	writing	average	pass_fail	grade
GF	None	72	72	74	73	Pass	B
GF	Completed	69	90	88	82	Pass	A
GF	None	90	95	93	93	Pass	A
RC	None	47	57	44	49	Fail	F
RC	None	76	78	75	76	Pass	B

WHERE

The **WHERE** clause is used to filter records

Example:

```
SELECT *
FROM Student
WHERE school_code = "GF"
```

SQLite | SQLite

```
1 SELECT *
2 FROM Student
3 WHERE school_code = "GF"
```

school_...	test_prep	math	reading
GF	None	72	72
GF	Completed	69	90
GF	None	90	95
GF	None	71	83

ORDER BY

The **ORDER BY** keyword is used to sort the result-set in ascending or descending order

Example:

```
SELECT school_code,
pass_fail, grade
FROM Student
WHERE grade;
```

SQLite | SQLite

```
SELECT school_code, pass_fail, grade
FROM Student
ORDER BY grade;
```

school_code	pass_fail	grade
GF	Pass	A
GF	Pass	A
GF	Pass	A
RC	Pass	A
RC	Pass	A

GROUP BY

The **GROUP BY** statement groups rows that have the same values into summary rows

The **GROUP BY** statement is often used with aggregate functions (COUNT, MAX, MIN, SUM, AVG) to group the result-set by one or more columns

Example:

```
SELECT school_code,
ROUND(AVG(math),2) as avg_math,
ROUND(AVG(reading),2) as avg_reading
ROUND(AVG(writing),2) as avg_writing
FROM Student
GROUP BY school_code;
```

SQLite | SQLite

```
SELECT school_code, ROUND(AVG(math),2) AS avg_math,
ROUND(AVG(reading),2) AS avg_reading, ROUND(AVG(writing),2) AS avg_writing
FROM Student
GROUP BY school_code;
```

school_code	avg_math	avg_reading	avg_writing
GF	63.63	72.61	72.47
RC	68.73	65.47	63.31

JOIN

A **JOIN** clause is used to combine rows from two or more tables, based on a related column between them

Example:

```
SELECT sc.school, st.grade
FROM Student AS st
JOIN School AS sc
ON st.school_code = sc.school_code
WHERE school_code = “GF”
```

SQLite | SQLite

```
SELECT sc.school, st.grade
FROM Student AS st
JOIN School AS sc
ON st.school_code = sc.school_code
WHERE st.school_code = "GF"
```

school	grade
Gryffindor	B
Gryffindor	A
Gryffindor	A
Gryffindor	B

JOIN

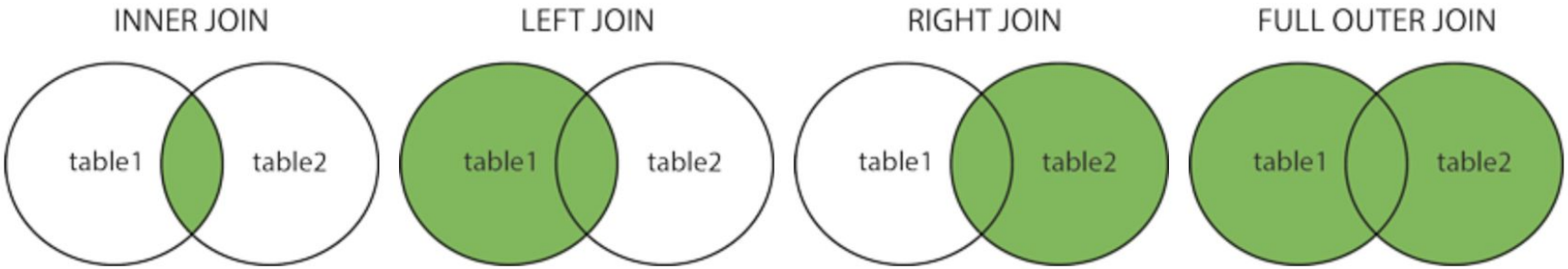
Different types of the JOINs in SQL:

(INNER) JOIN

Returns records that have matching values in both tables

FULL (OUTER) JOIN

Returns all records when there is a match in either left or right



https://www.w3schools.com/sql/sql_join.asp

JOIN

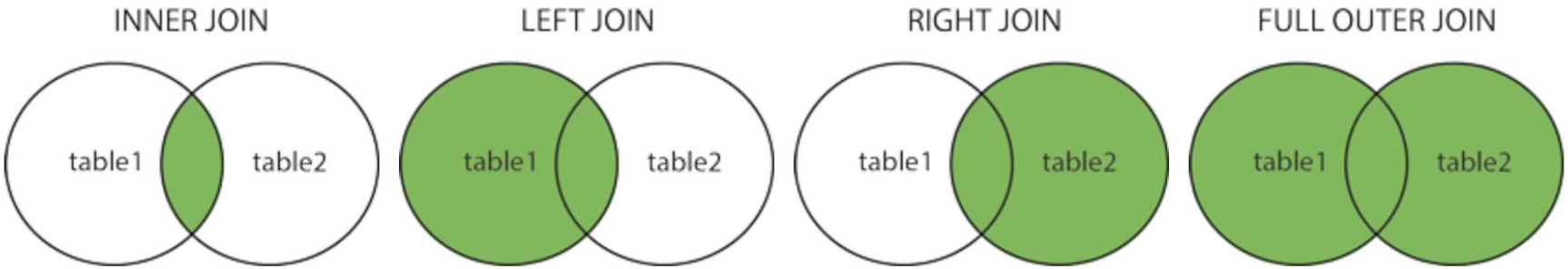
Different types of the JOINs in SQL:

LEFT (OUTER) JOIN:

Returns all records from the left table, and the matched records from the right table

RIGHT (OUTER) JOIN:

Returns all records from the right table, and the matched records from the left table



https://www.w3schools.com/sql/sql_join.asp

SQL Exercise

Download SQL Exercise Data
https://docs.google.com/spreadsheets/d/1ZVKyKP-xmVeXDzbRzprdKfE5g2c6XrSrscbUqYuXPoI/edit?usp=sharing

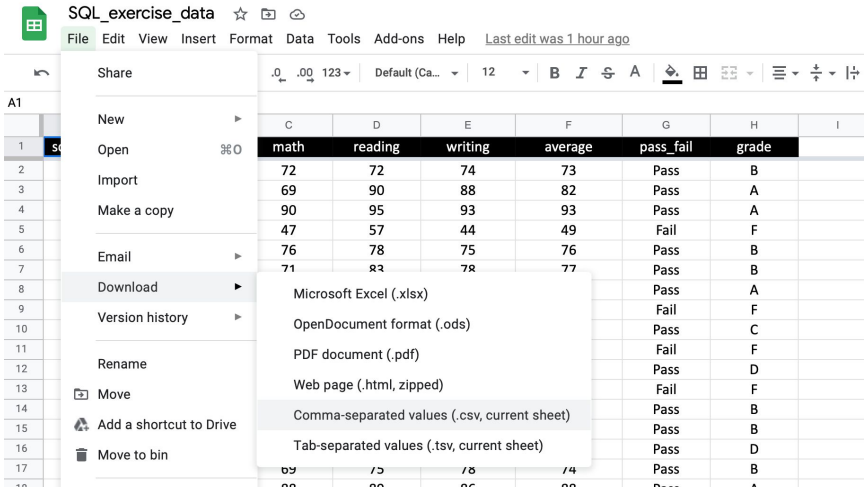
Download both Student and School Data as .csv file and import them on https://sqliteonline.com

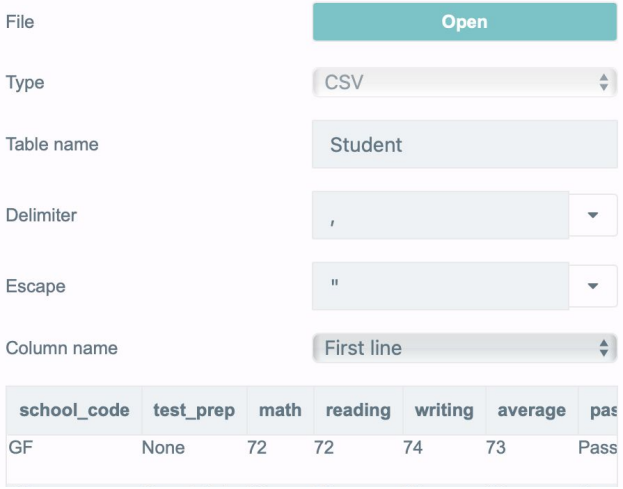
Sheet1: Student

- Contains detailed grade data for each student

Sheet2: School

- Contains data for school code and matching school name





SQL Exercise

1. Write a statement that will select the grade column from Student table
2. Select all records where they school_code column has the value “GF”
3. Select all records from the Student table, sort the result alphabetically by the column pass_fail
4. List the average math, average reading and average writing score group by school_code from Student table
5. Select all records from Student table order by grades (A, B, C, D, F)
6. Write a statement that will select average math, average reading and average writing score from Student table and group by school name from School table