

SQL × DML

64

Data Manipulation Language to

INSERT ×
UPDATE ×
DELETE ×
SELECT

from tables.

6
0

SQL \times DML \times **SELECT** (Relational Algebra) 65

Operations in relational algebra, i.e.,

Project(π)

Select(σ)

Rename(ρ)

Union(\cup)

Set Difference(\setminus)

Cartesian Product(\times)

Intersection(\cap), Division, Joins, ...

Only by one statement!

SQL \times DML \times **SELECT** \times Project(π)

66

2 **SELECT** ***** | *ColumnName1*,
ColumnName2,
...
ColumnNameN

1 **FROM** *TableName*

π *ColumnName1, ColumnName2, ..., ColumnNameN* (*TableName*)

SQL \times DML \times SELECT \times Select(σ)

70

3 SELECT * | ColumnName1,
ColumnName2,
...
ColumnNameN

1 FROM TableName

2 WHERE θ

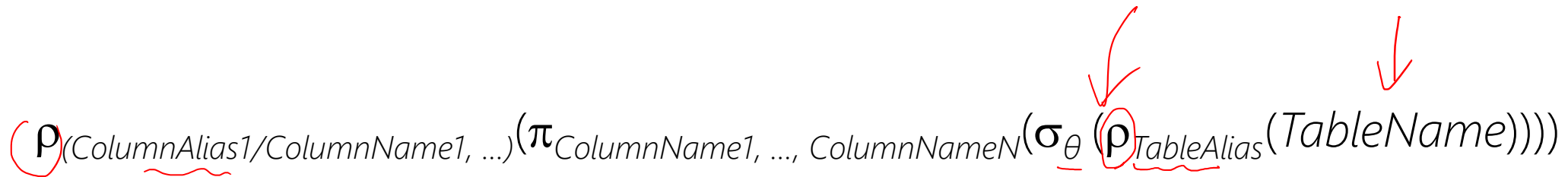
$\pi_{\text{ColumnName1, ColumnName2, ..., ColumnNameN}}(\sigma_{\theta}(\text{TableName}))$

SQL \times DML \times SELECT \times Rename(ρ)

75

- 3 SELECT * | ColumnName1 AS ColumnAlias1,
 ColumnName2 AS ColumnAlias2,
 ...
 ColumnNameN AS ColumnAliasN
- 1 FROM TableName AS TableAlias
- 2 WHERE θ

$(\rho_{(\text{ColumnAlias1/ColumnName1, ...})}(\pi_{\text{ColumnName1, ..., ColumnNameN}}(\sigma_{\theta}(\rho_{\text{TableAlias}}(\text{TableName}))))$



SQL × DML × **SELECT** × Union(∪)

77

- 1 (SELECT ...) AS A1
- 3 UNION
- 2 (SELECT ...) AS A2

SQL × DML × **SELECT** × Intersection(\cap) 81

1 (SELECT ...)
3 INTERSECT
2 (SELECT ...)

SQL × DML × **SELECT** × Set Diff(\)

84

① (SELECT ...) A
③ EXCEPT
② (SELECT ...) B

A \ B

SQL \times DML \times SELECT \times Product(\times)

87

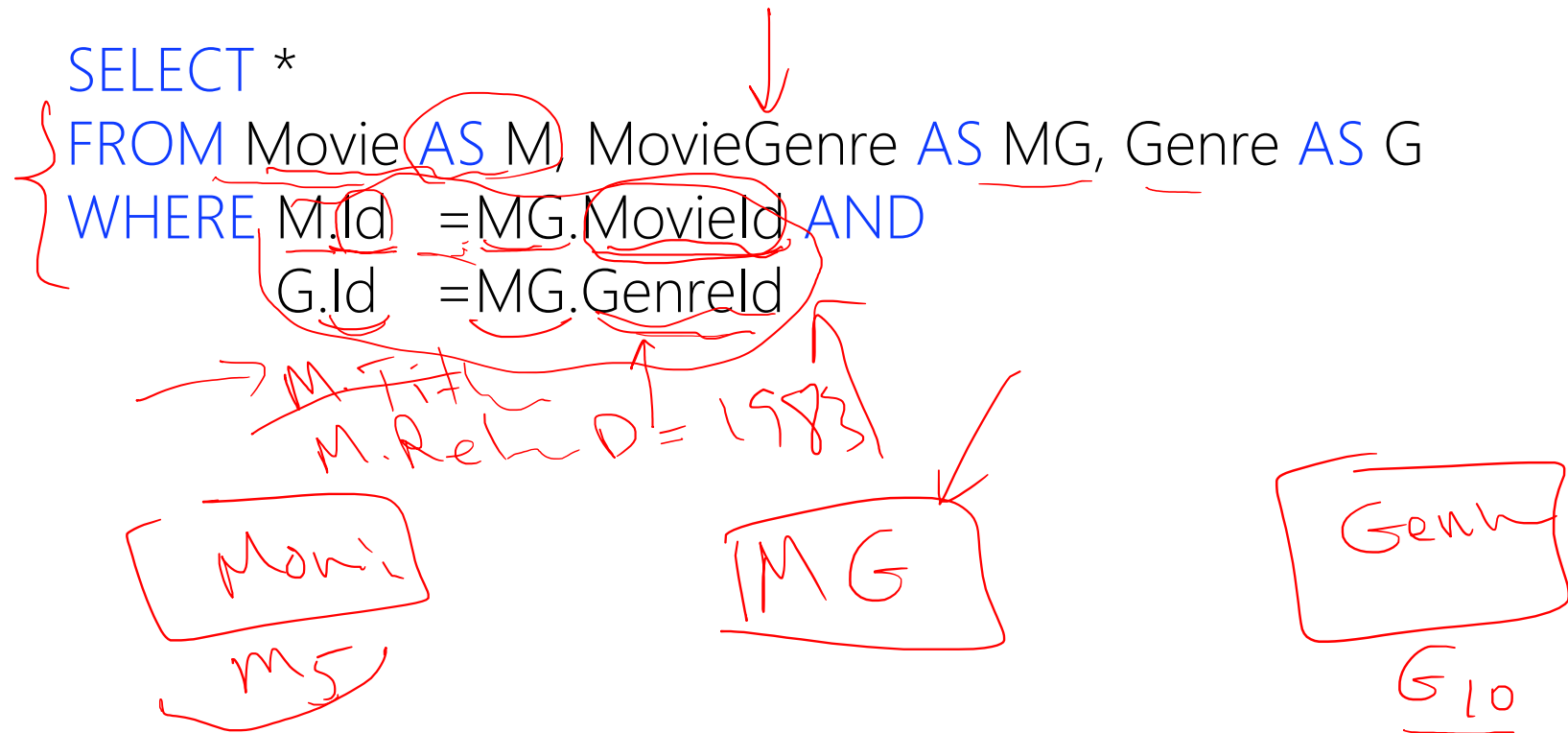
- 3 SELECT Column List
- 1 FROM TableName1, TableName2, ..., TableNameN
- 2 WHERE θ

$\pi_{\langle \text{Column List} \rangle} (\sigma_{\theta} (\text{TableName1} \times \text{TableName2} \times \dots \times \text{TableNameN}))$

Handwritten annotations: θ is circled in red. An arrow points from θ to $=$, and another from θ to AND . A third arrow points from θ to $>, <, \leq$. Four red arrows point upwards to the \times operators in the expression.

SQL × DML × **SELECT** × Product(×)

93





Advanced SQL × **SELECT**

A diagram showing the components of an SQL SELECT statement. The components are listed vertically, each preceded by a blue circle containing a number. Red handwritten annotations include: a large bracket on the left grouping items 1 through 5; a bracket on the right grouping items 1 through 4; a bracket on the right grouping items 6 and 7; and a bracket under the word 'LIMIT' in item 7. The text is as follows:

- 5 **SELECT DISTINCT** Columns
- 1 **FROM** Tables
- 2 [**WHERE** θ]
- 3 [**GROUP BY** Columns]
- 4 [**HAVING** θ']
- 6 [**ORDER BY** Columns [ASC | DESC]]
- 7 [**LIMIT** # [**OFFSET** #]]

The number 18 is written in black. A red arrow points from the top-left towards the number, and another red arrow points from the bottom-left towards the number. A small red bracket is drawn under the number 8.

Advanced SQL × DISTINCT

21

3 SELECT DISTINCT Columns ←
1 FROM Table
2 [WHERE θ]

To eliminate duplicate tuples, considering all columns.

Advanced SQL × Math Operation

27

3 **SELECT** Column {+, -, *, /, %, ...} Column | Constant
1 **FROM** Tables
2 [**WHERE** θ]

To apply a function on each value of a column.

Advanced SQL × Built-in Function

29

3 SELECT FUNCTION(Column), ...
1 FROM Tables
2 [WHERE θ]

To apply a function on each value of a column.

Advanced SQL × AGG Function

32

3 SELECT COUNT | SUM | MAX | MIN | AVG(Column)
1 FROM Tables
2 [WHERE θ]



To apply AGGregation functions on non-NULL values of one column and return a single value.

Advanced SQL × AGG Function

41

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	EN	1968	112

What's the shortest, longest, and average length of movies have been made in 1968?

title, Language, ReleaseDate

```
SELECT MIN(RunningTime) AS Min,  
       MAX(RunningTime) AS Max,  
       AVG(RunningTime) AS Avg,  
       SUM(RunningTime) AS Sum,  
       COUNT(RunningTime) AS Count,  
       COUNT(*)  
FROM Movie  
WHERE ReleaseDate = 1968
```

Min	Max	Avg	Sum	Count	Count(*)
112	142	127	254	2	3

127
142 + 112

Advanced SQL × AGG × DISTINCT

42

3 SELECT COUNT | SUM | MAX | MIN | AVG(DISTINCT Column)
1 FROM Tables
2 [WHERE θ]

To apply AGG functions on non-NULL values of one column,
after removing duplicates, and return a single value.

Advanced SQL × AGG × Math

44

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	EN	1968	112

What's the longest movie in hour which have been made in 1968?

```
SELECT MAX(RunningTime / 60) AS Max,  
FROM Movie  
WHERE ReleaseDate = 1968
```

Max
2

Advanced SQL × AGG × Built-in

45

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963 2022	119
4	Planet of the Apes	EN	1968	112

How many years old is the oldest movie?

```
SELECT MAX(STRFTIME('%Y', 'now') - ReleaseDate) AS Result
FROM Movie
WHERE ReleaseDate = 1968
```

Result
56

59

2022, 11, 8

Advanced SQL × WHERE

46

- 3 SELECT Columns
- 1 FROM Tables
- 2 [WHERE θ]

Advanced SQL × WHERE × Math

Advanced SQL × WHERE × Built-in

47

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	EN	1968	112

List all movies which are older than 55 years?

```
SELECT * FROM Movie
WHERE STRFTIME('%Y', 'now') - ReleaseDate > 55
```

Soln: $STRFTIME()$
From Movie

Advanced SQL × WHERE × AGG

48

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	EN	1968	112

What are the longest movies which have been made in 1968?

```
SELECT *  
FROM Movie  
WHERE ReleaseDate = 1968 AND  
RunningTime = MAX(RunningTime)
```

SELECT Max(RT) FROM



Advanced SQL × WHERE × ~~AGG~~

49

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	EN	1968	112

What is the oldest movie?

```
SELECT (*)  
FROM Movie  
WHERE ReleaseDate = MIN(ReleaseDate)
```



Advanced SQL × WHERE × NULL

50



NULL represents two facts about information

- I) No Value, e.g., non-American directors do not have SSN
- II) Missing | Unknow Value, e.g., PlaceOfBirth for a director might be missed, or unknown at the time of data entrance,

Advanced SQL × WHERE × NULL

51

Comparing with NULL value result in NULL, not FALSE, not TRUE!

NULL = NULL

SELECT NULL=1 AS Result;

SELECT NULL<>1 AS Result;

SELECT NULL>1 AS Result;

SELECT NULL>=1 AS Result;

...

Result
NULL

Advanced SQL × WHERE × IS NULL

52

To compare with NULL value explicitly, IS NULL | IS NOT NULL:

```
SELECT NULL IS NULL           AS Result1,  
       NULL IS NOT NULL       AS Result2;
```

Result1	Result2
TRUE	FALSE

Advanced SQL × WHERE × IS NULL

53

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	EN	1968	112

How many movies whose running time is missing?

```
SELECT COUNT(*) AS IncompleteMovieCount
FROM Movie
WHERE RunningTime IS NULL
```

IncompleteMovieCount
1

Advanced SQL × WHERE × LIKE

54

VARCHAR CHAR

To compare with CHAR-based values symbol, SQL has LIKE | NOT LIKE operator and pattern matching symbols:

- I) %, Represents 0 or more of any CHAR, called wildcard
- II) _, Represents any single character

Title = 'Rosen'

Advanced SQL × WHERE × LIKE

55

e.g.,

Title LIKE 'H%' : matches values start with 'H' or 'h', the rest can be anything.

LIKE 'H____' : matches values with exactly four CHARs starting with 'H' or 'h'.

LIKE '%e' : matches values ends with 'e' or 'E'.

LIKE '%birds%': matches values containing 'birds', or 'BiRds'.

NOT LIKE 'H%': matches values whose first character is not 'H' or 'h'.

LIKE operator is not case sensitive for ASCII chars

Advanced SQL × WHERE × LIKE

56

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	EN	1968	112

Find the movie 'space odyssey'?

```
SELECT *  
FROM Movie  
WHERE Title LIKE '%space odyssey%'
```


Advanced SQL × WHERE × LIKE

57

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	EN	1968	112

Find the movie 'space odyssey'?

```
SELECT *  
FROM Movie  
WHERE Title = '%space odyssey%'
```



Advanced SQL × FROM

58

- 3 SELECT Columns
- 1 FROM Tables
- 2 [WHERE θ]

Advanced SQL \times FROM \times Product(\times) 59

- 3 SELECT Columns
- 1 FROM *Table1, Table2, ..., TableN*
- 2 WHERE θ

$$\pi_{\langle \text{Columns} \rangle}(\sigma_{\theta}(\text{Table1} \times \text{Table2} \times \dots \times \text{TableNameN}))$$

Advanced SQL \times FROM \times Product(\times) 60

Movie				MovieGenre		Genre	
Id	Title	Language	RunningTime	MovielId	GenrelId	Id	Title
1	2001: A Space Odyssey	English	142	1	1	1	Sci-fi
1	2001: A Space Odyssey	English	142	1	3	3	Adventure

Find genres of the movie '2001: A Space Odyssey'?

$\sigma_{\text{Genre.Id}=\text{GenrelId}}(\sigma_{\text{Movie.Id}=\text{MovielId} \text{ AND } \text{Title}='2001: A Space Odyssey'}(\text{Movie} \times \text{MovieGenre})) \times \text{Genre}$

Advanced SQL \times FROM \times Product(\times)

61

Movie				MovieGenre		Genre	
Id	Title	Language	RunningTime	MovieId	GenreId	Id	Title
1	2001: A Space Odyssey	English	142	1	1	1	Sci-fi
1	2001: A Space Odyssey	English	142	1	3	3	Adventure

Find genres of the movie '2001: A Space Odyssey'?

$\sigma_{\text{Genre.Id}=\text{GenreId}}(\sigma_{\text{Movie.Id}=\text{MovieId} \text{ AND } \text{Title}='2001: A Space Odyssey'}(\text{Movie} \times \text{MovieGenre})) \times \text{Genre}$



I) Corollary: $\sigma_{\theta}(\sigma_{\theta'}(\sigma_{\theta''}(\sigma_{\theta'''}(R))) = \sigma_{\theta \text{ AND } \theta' \text{ AND } \theta'' \text{ AND } \theta'''}(R)$

II) Product has commutative property, i.e., $(R1 \times R2) \times R3 = R1 \times (R2 \times R3) = R1 \times R2 \times R3$

Advanced SQL × FROM × Product(×)

62

Movie				MovieGenre		Genre	
<u>Id</u>	Title	Language	RunningTime	<u>Movied</u>	<u>GenreId</u>	<u>Id</u>	Title
1	2001: A Space Odyssey	English	142	1	1	1	Sci-fi
1	2001: A Space Odyssey	English	142	1	3	3	Adventure

σ

(Movie × MovieGenre × Genre)

Movie.Id=Movied AND

Genre.Id=GenreId AND

Title='2001: A Space Odyssey'

Advanced SQL × FROM × Product(×)

63

Movie				MovieGenre		Genre	
Id	Title	Language	RunningTime	Movied	Genred	Id	Title
1	2001: A Space Odyssey	English	142	1	1	1	Sci-fi
1	2001: A Space Odyssey	English	142	1	3	3	Adventure

σ

(Movie × MovieGenre × Genre)

Movie.Id=Movied AND

Genre.Id=Genred AND

Title='2001: A Space Odyssey'

SELECT *

FROM Movie, MovieGenre, Genre

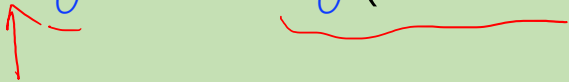
WHERE Movie.Id =Movied AND

Genre.Id =Genred AND

Title ='2001: A Space Odyssey'

Relational Algebra \times θ -Join

\bowtie_{θ} , θ -join, is product (\times) of relations followed by selection (σ)

$$R1 \bowtie_{\theta} R2 = \sigma_{\theta}(R1 \times R2)$$


Advanced SQL × FROM × INNER JOIN 64

3 SELECT Columns
1 FROM Table1, Table2
2 WHERE θ

~~σ_{θ}~~ (R1 × R2)

----- SAME AS -----

3 SELECT Columns
1 FROM Table1
2 INNER JOIN Table2 ON θ

R1 ⋈ $_{\theta}$ R2

Advanced SQL × FROM × INNER JOIN

65

Movie				MovieGenre		Genre	
Id	Title	Language	RunningTime	Movielid	Genrelid	Id	Title
1	2001: A Space Odyssey	English	142	1	1	1	Sci-fi
1	2001: A Space Odyssey	English	142	1	3	3	Adventure

$\sigma_{\text{Title}='2001: A Space Odyssey'}((\text{Movie} \bowtie_{\text{Movie.Id}=\text{Movielid}} \text{MovieGenre}) \bowtie_{\text{Genre.Id}=\text{Genrelid}} \text{Genre})$

SELECT *
 FROM Movie
 INNER JOIN MovieGenre ON Movie.Id = Movielid
 INNER JOIN Genre ON Genre.Id = Genrelid
 WHERE Title = '2001: A Space Odyssey'

Advanced SQL × FROM × INNER JOIN

66

Director						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	NULL	47
3	Clint	Eastwood	May 31, 1930	USA	NULL	35

Movie			
Id	Title	Language	RunningTime
1	2001: A Space Odyssey	English	142
2	Rosemary's Baby	English	NULL

What are directors' best movie name?

```
SELECT D.Id, FirstName, LastName, BestMovieId, M.Id, Title
FROM Director AS D, Movie AS M
WHERE M.Id = BestMovieId
```

3 × 6
↓
1

Advanced SQL × FROM × INNER JOIN

67

Director						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	NULL	47
3	Clint	Eastwood	May 31, 1930	USA	NULL	35

Movie			
Id	Title	Language	RunningTime
1	2001: A Space Odyssey	English	142
2	Rosemary's Baby	English	NULL

What are directors' best movie name?

```
SELECT D.Id, FirstName, LastName, BestMovieId, M.Id, Title
FROM Director AS D
INNER JOIN Movie AS M ON M.Id = BestMovieId
```

Advanced SQL × FROM × INNER JOIN

68

Id	FirstName	LastName	BestMovieId	Id	Title
1	Stanley	Kubrick	1	1	2001: A Space Odyssey



What are directors' best movie name?

```
SELECT D.Id, FirstName, LastName, BestMovieId, M.Id, Title
FROM Director AS D
INNER JOIN Movie AS M ON M.Id = BestMovieId
```

Advanced SQL × FROM × LEFT JOIN

69

FirstName	LastName	BestMovieId	Id	Title
Stanley	Kubrick	1	1	2001: A Space Odyssey
Alfred	Hitchcock	NULL	NULL	NULL
Clint	Eastwood	NULL	NULL	NULL

What are directors' best movie name if any?

Director ⋈_{BestMovieId=Movie.Id} Movie

```
SELECT D.Id, FirstName, LastName, BestMovieId, M.Id, Title
FROM Director AS D
LEFT [OUTER] JOIN Movie AS M ON M.Id = BestMovieId
```

Advanced SQL × FROM × RIGHT JOIN

70

FirstName	LastName	BestMovieId	Id	Title
Stanley	Kubrick	1	1	2001: A Space Odyssey
NULL	NULL	NULL	2	Rosemary's Baby

List all movies and identify whether each one is the best of its director?

Director  BestMovieId=Movie.Id Movie

From Movie
Left Join

Dir

```
SELECT D.Id, FirstName, LastName, BestMovieId, M.Id, Title
FROM Director AS D
RIGHT [OUTER] JOIN Movie AS M ON M.Id = BestMovieId
```

Advanced SQL × FROM × FULL JOIN

71

FirstName	LastName	BestMovieId	Id	Title
Stanley	Kubrick	1	1	2001: A Space Odyssey
Alfred	Hitchcock	NULL	NULL	NULL
Clint	Eastwood	NULL	NULL	NULL
NULL	NULL	NULL	2	Rosemary's Baby

NULL
NULL
NULL
NULL

Director ⋈_{BestMovieId=Movie.Id} Movie

→ SELECT D.Id, FirstName, LastName, BestMovieId, M.Id, Title
① FROM Director AS D
② FULL [OUTER] JOIN Movie AS M ON M.Id = BestMovieId

NULL

→ WHERE Place of Birth = 'USA'

Advanced SQL × FROM × RIGHT JOIN 72

Advanced SQL × FROM × FULL JOIN

NOT currently supported in SQLite!

Think about a workaround & bring it with you next week.



Advanced SQL × SELECT

73

- 5 SELECT DISTINCT Columns
- 1 FROM Tables
- 2 [WHERE θ]
- 3 [GROUP BY Columns]
- 4 [HAVING θ']

AVG

IM NOW →

Advanced SQL × GROUP BY

74

To group tuples based on values on columns, usually followed by AGG functions.

Advanced SQL × GROUP BY

75

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	English	1968	112

How many movies have been made in each year?

```
SELECT ReleaseDate, COUNT(*)  
FROM Movie  
GROUP BY ReleaseDate
```

AVG RunningTime

ReleaseDate	COUNT(*)
1968	3
1963	1

$142 + \cancel{NULL} + 112$
AVG
Run
 $\frac{142 + 112}{3}$
119

Advanced SQL × GROUP BY

76

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	EN	1968	112

How many movies have been made in each year per language?

```
SELECT ReleaseDate, Language, COUNT(*)  
FROM Movie  
GROUP BY ReleaseDate, Language
```

ReleaseDate	Language	COUNT(*)
1968	English	2
1968	EN	1
1963	English	1

Advanced SQL × GROUP BY

77

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	EN	1968	112

How many movies have been made in each year per language?

```
SELECT Title, ReleaseDate, Language, COUNT(*)  
FROM Movie  
GROUP BY ReleaseDate, Language
```

ReleaseDate	Language	COUNT(*)
1968	English	2
1968	EN	1
1963	English	1

SELECT clause only accepts AGG or columns in GROUP BY list.

Advanced SQL × GROUP BY × HAVING 78

Movie				
Id	Title	Language	ReleaseDate	RunningTime
1	2001: A Space Odyssey	English	1968	142
2	Rosemary's Baby	English	1968	NULL
3	The Birds	English	1963	119
4	Planet of the Apes	English	1968	112

In which years more than 2 movies have been released?

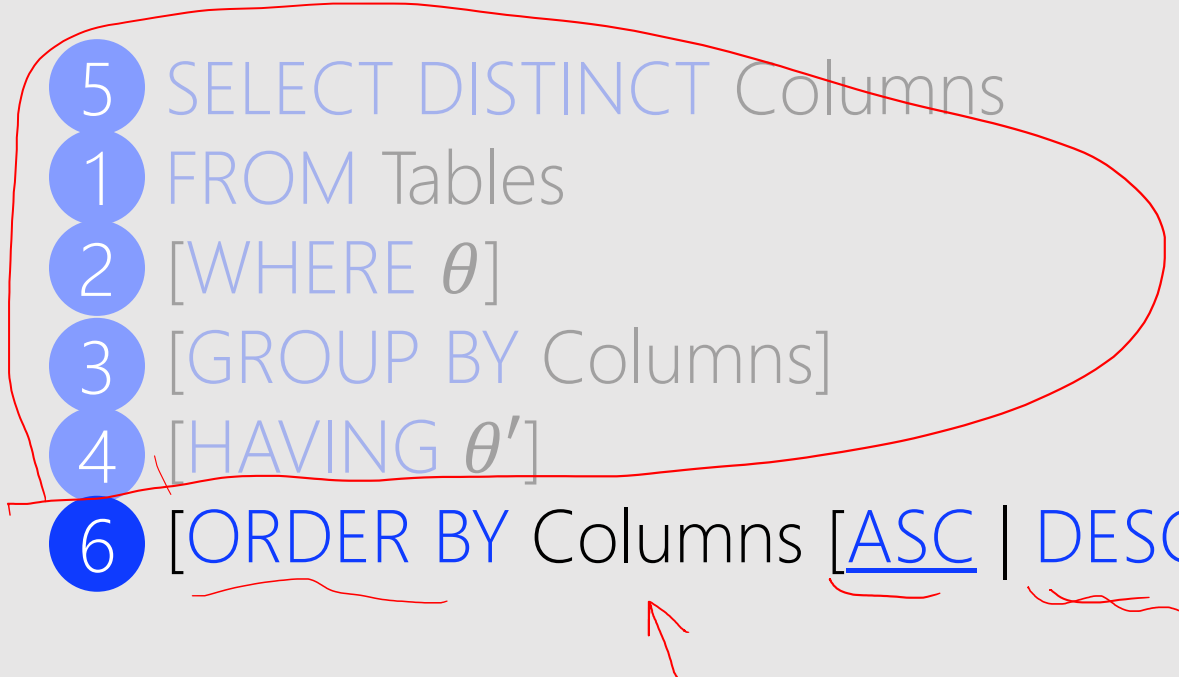
```
SELECT ReleaseDate, COUNT(*)  
FROM Movie  
GROUP BY ReleaseDate  
HAVING COUNT(*) > 2
```

ReleaseDate	COUNT(*)
1968	3

HAVING only accepts AGG or columns in GROUP BY list.

Advanced SQL × Sorting

79

- 
- 5 SELECT DISTINCT Columns
 - 1 FROM Tables
 - 2 [WHERE θ]
 - 3 [GROUP BY Columns]
 - 4 [HAVING θ']
 - 6 [ORDER BY Columns [ASC | DESC]]

Advanced SQL × Sorting

80

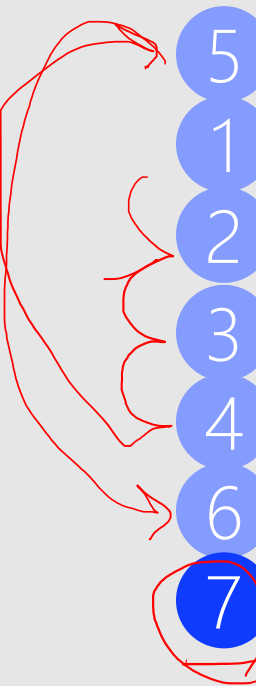
Movie				
Id	Title	Language	ReleaseDate	RunningTime
2	Rosemary's Baby	English	1968	NULL
4	Planet of the Apes	English	1968	112
1	2001: A Space Odyssey	English	1968	142
3	The Birds	English	1963	119

List movies sorted by release date & running time?

```
SELECT *  
FROM Movie  
ORDER BY ReleaseDate DESC, RunningTime
```


Advanced SQL × Paging

81

- 
- 5 SELECT DISTINCT Columns
- 1 FROM Tables
- 2 [WHERE θ]
- 3 [GROUP BY Columns]
- 4 [HAVING θ']
- 6 [ORDER BY Columns [ASC | DESC]]
- 7 [LIMIT # [OFFSET #]]
- The diagram illustrates the execution order of SQL clauses. Red arrows show dependencies: FROM (1) is the base for WHERE (2), GROUP BY (3), and HAVING (4). WHERE (2) also feeds into ORDER BY (6). GROUP BY (3) and HAVING (4) also feed into ORDER BY (6). ORDER BY (6) feeds into LIMIT/OFFSET (7). The LIMIT/OFFSET clause (7) is circled in red, and a red arrow points from it back to the top of the list.

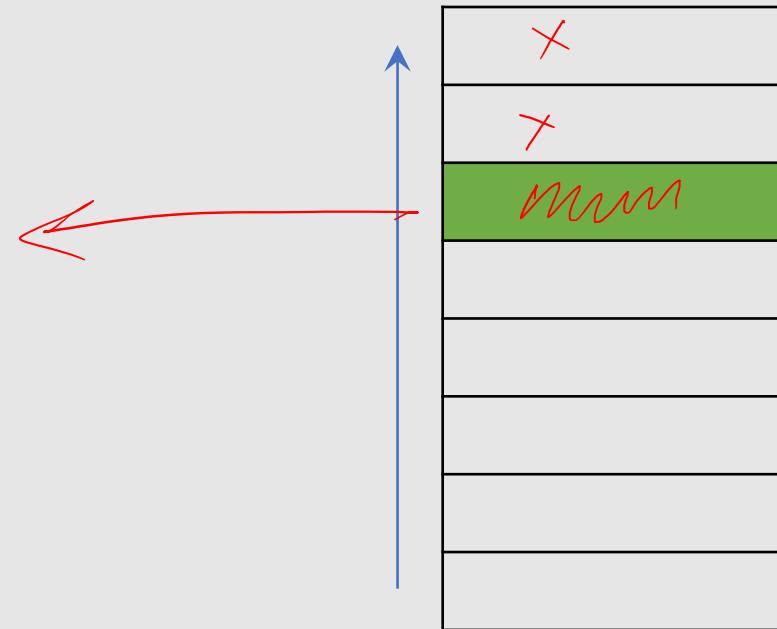
Advanced SQL × Paging

82

Page 3 of recent movies? (each page shows 10 movies)

Top-10 recent movies after skipping the top-2×10 most recent ones?

```
SELECT *  
FROM Movie  
ORDER BY ReleaseDate DESC  
LIMIT 10 OFFSET 2×10
```



Advanced SQL × Paging

83

Page n of recent items? (each page shows N items)
Top- N recent items after skipping the top- $(n-1) \times N$ most recent ones?

```
SELECT *  
FROM Items  
ORDER BY Date DESC  
LIMIT  $N$  OFFSET  $(n-1) \times N$ 
```

Advanced SQL × Paging

84

$i = 1 \rightarrow 100$

Page n of expensive items? (each page shows N items)

Top- N expensive items after skipping the top- $(n-1) \times N$ most expensive?

```
SELECT *  
FROM Items  
ORDER BY Price DESC  
LIMIT  $N$  OFFSET  $(n-1) \times N$ 
```