Advanced SQL Puzzles

Scott Peters

www.advancedsqlpuzzles.com

Last updated 08/22/2021

Table of Contents

Section I

Puzzle #1 Dance Partners

Puzzle #2 Managers and Employees

Puzzle #3 Fiscal Year Pay Rates

Puzzle #4 Two Predicates

Puzzle #5 Phone Directory

Puzzle #6 Workflow Steps

Puzzle #7 Mission to Mars

Puzzle #8 Workflow Cases

Puzzle #9 Matching Sets

Puzzle #10 Mean, Median, Mode, and Range

Puzzle #11 Permutations

Puzzle #12 Average Days

Puzzle #13 Inventory Tracking

Puzzle #14 Indeterminate Process Log

Puzzle #15 Group Concatenation

Puzzle #16 Reciprocals

Puzzle #17 De-Grouping

Puzzle #18 Seating Chart

Puzzle #19 Back to the Future

Puzzle #20 Price Points

Puzzle #21 Average Monthly Sales

Puzzle #22 Occurrences

Puzzle #23 Divide in Half

Puzzle #24 Page Views

Puzzle #25 Top Vendors

Puzzle #26 Previous Year's Sales

Puzzle #27 Delete the Duplicates

Puzzle #28 Fill the Gaps

Puzzle #29 Count the Groupings

Puzzle #30 Select Star

Puzzle #31 Second Highest

Puzzle #32 First and Last

Puzzle #33 Deadlines

Puzzle #34 Specific Exclusion

Puzzle #35 International vs Domestic Sales

Puzzle #36 Traveling Salesman

Puzzle #37 Group Criteria Key

Puzzle #38 Reporting Elements

Puzzle #39 Prime Numbers

Puzzle #40 Sort Order

Answers to the puzzles are located in the following GitHub repository. https://github.com/smpetersgithub/AdvancedSQLPuzzles/tree/main/Advanced%20SQL%20Puzzles

I would be happy to receive corrections, additions, new tricks and techniques, and other suggestions. scottpeters1188@outlook.com.

The latest version of this document can be found at www.advancedsqlpuzzles.com

Dance Partners

You are tasked with providing a list of dance partners from the following table.

Provide an SQL statement that matches each Student ID with an individual of the opposite gender.

Note there is a mismatch in the number of students, as one female student will be left without a dance partner. Please include this individual in your list as well.

Student ID	Gender
1001	М
2002	М
3003	М
4004	М
5005	М
6006	F
7007	F
8008	F
9009	F

Managers and Employees

Given the following table, write an SQL statement that determines the level of depth each employee has from the president.

Employee ID	Manager ID	Job Title	Salary
1001		President	\$185,000
2002	1001	Director	\$120,000
3003	1001	Office Manager	\$97,000
4004	2002	Engineer	\$110,000
5005	2002	Engineer	\$142,000
6006	2002	Engineer	\$160,000

Here is the expected output.

Employee ID	Manager ID	Job Title	Salary	Depth
1001	NULL	President	\$185,000	0
2002	1001	Director	\$120,000	1
3003	1001	Office Manager	\$97,000	1
4004	2002	Engineer	\$110,000	2
5005	2002	Engineer	\$142,000	2
6006	2002	Engineer	\$160,000	2

Answers to the puzzles are located in the following GitHub repository.

https://github.com/smpetersgithub/AdvancedSQLPuzzles/tree/main/Advanced%20SQL%20Puzzles

Fiscal Year Pay Rates

For each standard fiscal year, a record exists for each employee that states their current pay rate for the specified year.

Can you determine all the constraints that can be applied to this table to ensure that it contains only correct information? Assume that no pay raises are given mid-year. There are quite a few of them, so think carefully!

```
CREATE TABLE #EmployeePayRecord (

EmployeeID INTEGER

FiscalYear INTEGER,

StartDate DATE,

EndDate DATE,

PayRate MONEY
);
```

Two Predicates

Write an SQL statement given the following requirements.

For every customer that had a delivery to California, provide a result set of the customer orders that were delivered to Texas.

Customer ID	Order ID	Delivery State	Amount
1001	Ord936254	CA	\$340
1001	Ord143876	TX	\$950
1001	Ord654876	TX	\$670
1001	Ord814356	TX	\$860
2002	Ord342176	WA	\$320
3003	Ord265789	CA	\$650
3003	Ord387654	CA	\$830
4004	Ord476126	TX	\$120

Here is the expected output.

Customer ID	Order ID	Delivery State	Amount
1001	Ord143876	TX	\$950
1001	Ord654876	TX	\$670
1001	Ord814356	TX	\$860

Customer ID 1001 would be in the expected output as this customer had deliveries to both California and Texas. Customer ID 3003 would not show in the result set as they did not have a delivery to Texas, and Customer ID 4004 would not appear in the result set as they did not have a delivery to California.

Phone Directory

Your customer phone directory table allows individuals to setup a home, cellular, or a work phone number.

Write an SQL statement to transform the following table into the expected output.

Customer ID	Туре	Phone Number
1001	Cellular	555-897-5421
1001	Work	555-897-6542
1001	Home	555-698-9874
2002	Cellular	555-963-6544
2002	Work	555-812-9856
3003	Cellular	555-987-6541

Here is the expected output.

Customer ID	Cellular	Work	Home
1001	555-897-5421	555-897-6542	555-698-9874
2002	555-963-6544	555-812-9856	
3003	555-987-6541		

Workflow Steps

Write an SQL statement that determines all workflows that have started but have not completed.

Workflow	Step Number	Completion Date
Alpha	1	7/2/2018
Alpha	2	7/2/2018
Alpha	3	7/1/2018
Bravo	1	6/25/2018
Bravo	2	
Bravo	3	6/27/2018
Charlie	1	
Charlie	2	7/1/2018

The expected output would be Bravo and Charlie, as they have a workflow that has started but has not completed.

Bonus: Write this query only using the COUNT function with no subqueries. Can you figure out the trick?

Mission to Mars

You are given the following tables that list the requirements for a space mission and a list of potential candidates.

Write an SQL statement to determine which candidates meet the requirements of the mission.

Candidates

Candidate ID	Description
1001	Geologist
1001	Astrogator
1001	Biochemist
1001	Technician
2002	Surgeon
2002	Machinist
3003	Cryologist
4004	Selenologist

Requirements

Description	
Geologist	
Astrogator	
Technician	

The expected output would be Candidate ID 1001, as this candidate has all the necessary skills for the space mission. Candidate ID 2002 and 3003 would not be in the output as they have some, but not all the required skills.

Workflow Cases

You have a report of all workflows and their case results.

A value of 0 signifies the workflow failed, and a value of 1 signifies the workflow passed.

Write an SQL statement that transforms the following table into the expected output.

Workflow	Case 1	Case 2	Case 3
Alpha	0	0	0
Bravo	0	1	1
Charlie	1	0	0
Delta	0	0	0

Here is the expected output.

Workflow	Passed
Alpha	0
Bravo	2
Charlie	1
Delta	0

Matching Sets

Write an SQL statement that matches an employee to all other employees that carry the same licenses.

Employee ID	License
1001	Class A
1001	Class B
1001	Class C
2002	Class A
2002	Class B
2002	Class C
3003	Class A
3003	Class D

Employee ID 1001 and 2002 would be in the expected output as they both carry a Class A, Class B, and a Class C license.

Mean, Median, Mode, and Range

The mean is the average of all numbers.

The median is the middle number in a sequence of numbers.

The mode is the number that occurs most often within a set of numbers.

The range is the difference between the highest and lowest values in a set of numbers.

Write an SQL statement to determine the mean, median, mode and range of the following set of integers.

```
CREATE TABLE #SampleData
(
IntegerValue INTEGER
);
GO

INSERT INTO #SampleData
VALUES(5),(6),(10),(10),(13),(14),(17),(20),(81),(90),(76);
GO
```

Permutations

You are given the following list of test cases and must determine all possible permutations.

Write an SQL statement that produces the expected output.

Test Case	
А	
В	
С	

Here is the expected output.

Row Number	Output
1	A,B,C
2	A,C,B
3	B,A,C
4	B,C,A
5	C,A,B
6	C,B,A

Average Days

Write an SQL statement to determine the average number of days between executions for each workflow.

Workflow	Execution Date
Alpha	6/1/2018
Alpha	6/14/2018
Alpha	6/15/2018
Bravo	6/1/2018
Bravo	6/2/2018
Bravo	6/19/2018
Charlie	6/1/2018
Charlie	6/15/2018
Charlie	6/30/2018

Here is the expected output.

Workflow	Average Days
Alpha	7
Bravo	9
Charlie	14

Inventory Tracking

You work for a manufacturing company and need to track inventory adjustments from the warehouse.

Some days the inventory increases, on other days the inventory decreases.

Write an SQL statement that will provide a running balance of the inventory.

	Quantity	
Date	Adjustment	
7/1/2018	100	
7/2/2018	75	
7/3/2018	-150	
7/4/2018	50	
7/5/2018	-100	

Here is the expected output.

Date	Quantity Adjustment	Inventory
7/1/2018	100	100
7/2/2018	75	175
7/3/2018	-150	25
7/4/2018	50	75
7/5/2018	-50	25

Indeterminate Process Log

Your process log has several workflows broken down by step numbers with the possible status values of Complete, Running, or Error.

Your task is to write an SQL statement that creates an overall status based upon the following requirements.

- If all the workflow steps have a status of complete, set the overall status to complete. (ex. Bravo).
- If all the workflow steps have a status of error, set the overall status to error (ex. Foxtrot).
- If the workflow steps have the combination of error and complete, or error and running, the overall status should be indeterminate (ex. Alpha, Charlie, Echo).
- If the workflow steps have the combination of complete and running, the overall status should be running (ex. Delta).

Workflow	Step Number	Status
Alpha	1	Error
Alpha	2	Complete
Bravo	1	Complete
Bravo	2	Complete
Charlie	1	Complete
Charlie	2	Error
Delta	1	Complete
Delta	2	Running
Echo	1	Running
Echo	2	Error
Foxtrot	1	Error

Here is the expected output.

Workflow	Status	
Alpha	Indeterminate	
Bravo	Complete	
Charlie	Indeterminate	
Delta	Running	
Echo	Indeterminate	
Foxtrot	Error	

Group Concatenation

Write an SQL statement that can group concatenate the following values.

Sequence	Syntax
1	SELECT
2	Product
3	UnitPrice
4	EffectiveDate
5	FROM
6	Products
7	WHERE
8	UnitPrice
9	> 100

Here is the expected output.

Syntax
SELECT Product, UnitPrice, EffectiveDate FROM Products WHERE UnitPrice > 100

Reciprocals

You work for a software company that released a 2-player game and you need to tally the scores.

Given the following table, write an SQL statement to determine the reciprocals and calculate their aggregate score.

In the data below, players 3003 and 4004 have two valid entries, but their scores need to be aggregated together.

Player A	Player B	Score
1001	2002	150
3003	4004	15
4004	3003	125

Here is the expected output.

Player A	Player B	Score
1001	2002	150
3003	4004	140

De-Grouping

Write an SQL Statement to de-group the following data.

Product	Quantity
Pencil	3
Eraser	4
Notebook	2

Here is the expected output.

Product	Quantity
Pencil	1
Pencil	1
Pencil	1
Eraser	1
Notebook	1
Notebook	1

Answers to the puzzles are located in the following GitHub repository.

https://github.com/smpetersgithub/AdvancedSQLPuzzles/tree/main/Advanced%20SQL%20Puzzles

Seating Chart

Given the following set of integers, write an SQL statement to determine the expected outputs.

```
CREATE TABLE #SeatingChart
(
SeatNumber INTEGER
);
GO

INSERT INTO #SeatingChart VALUES
(7),(13),(14),(15),(27),(28),(29),(30),(31),(32),(33),(34),(35),(52),(53),(54);
GO
```

Here is the expected output.

Gap Start	Gap End
1	6
8	12
16	26
36	51

Total Missing Numbers	
38	

Туре	Count
Even Numbers	8
Odd Numbers	9

Answers to the puzzles are located in the following GitHub repository.

https://github.com/smpetersgithub/AdvancedSQLPuzzles/tree/main/Advanced%20SQL%20Puzzles

Back to the Future

Here is one of the more difficult puzzles to solve with a declarative SQL statement.

Write an SQL statement to merge the overlapping time periods.

Start Date	End Date
1/1/2018	1/5/2018
1/3/2018	1/9/2018
1/10/2018	1/11/2018
1/12/2018	1/16/2018
1/15/2018	1/19/2018

Here is the expected output.

Start Date	End Date
1/1/2018	1/9/2018
1/10/2018	1/11/2018
1/12/2018	1/19/2018

Price Points

Write an SQL statement to determine the current price point for each product.

Product ID	Unit Price	Effective Date
1001	\$1.99	1/1/2018
1001	\$2.99	5/17/2018
1001	\$3.99	6/14/2018
2002	\$12.99	2/12/2018
2002	\$17.99	3/1/2018
2002	\$20.99	4/19/2018

Here is the expected output.

Product ID	Effective Date	Unit Price
1001	6/8/2018	\$3.99
2002	5/19/2018	\$2.99

Average Monthly Sales

Write an SQL statement that returns a list of states where customers have an average monthly sales value that is consistently greater than \$100.

Order ID	Customer ID	Order Date	Amount	State
Ord145332	1001	1/1/2018	\$100	TX
Ord657895	1001	1/1/2018	\$150	TX
Ord887612	1001	1/1/2018	\$75	TX
Ord654374	1001	2/1/2018	\$100	TX
Ord345362	1001	3/1/2018	\$100	TX
Ord912376	2002	2/1/2018	\$75	TX
Ord543219	2002	2/1/2018	\$150	TX
Ord156357	3003	1/1/2018	\$100	IA
Ord956541	3003	2/1/2018	\$100	IA
Ord856993	3003	3/1/2018	\$100	IA
Ord864573	4004	4/1/2018	\$100	IA
Ord654525	4004	5/1/2018	\$50	IA
Ord987654	4004	5/1/2018	\$100	IA

In this example, Texas would show in the result set as Customer ID 1001 and 2002 each have their average monthly value over \$100. Iowa would not show in the result set because Customer ID 3003 did not have an average monthly value over \$100 in May 2018.

Occurrences

Write an SQL statement that returns all distinct process log messages and the workflow where the message occurred the most often.

Workflow	Occurrences	Message
Alpha	5	Error: Conversion Failed
Alpha	8	Status Complete
Alpha	9	Error: Unidentified error occurred
Bravo	3	Error: Cannot Divide by 0
Bravo	1	Error: Unidentified error occurred
Charlie	10	Error: Unidentified error occurred
Charlie	7	Error: Conversion Failed
Charlie	6	Status Complete

Here is the expected output.

Workflow	Message
Alpha	Status Complete
Bravo	Error: Cannot Divide by 0
Charlie	Error: Conversion Failed
Charlie	Error: Unidentified error occurred

Divide in Half

You work for a gaming company and need to rank players by their score into two categories.

Players that rank in the top half must be given a value of 1; the remaining players must be given a value of 2.

Write an SQL statement that meets these requirements.

```
CREATE TABLE #PlayerScores (
    PlayerID VARCHAR(MAX),
    Score    INTEGER
);
    GO

INSERT INTO #PlayerScores VALUES (1001,2343), (2002,9432), (3003,6548), (4004,1054), (5005,6832);
    GO
```

Answers to the puzzles are located in the following $\operatorname{\sf GitHub}$ repository.

https://github.com/smpetersgithub/AdvancedSQLPuzzles/tree/main/Advanced%20SQL%20Puzzles

Page Views

Write an SQL statement that retrieves records 10 to 20 ordered by the RowID column. Here is the syntax to create and populate the table.

Answers to the puzzles are located in the following GitHub repository.

https://github.com/smpetersgithub/AdvancedSQLPuzzles/tree/main/Advanced%20SQL%20Puzzles

Top Vendors

Write an SQL statement that returns the vendor from which each customer has placed the most orders.

Order ID	Customer ID	Order Count	Vendor
Ord195342	1001	12	Direct Parts
Ord245532	1001	54	Direct Parts
Ord344394	1001	32	ACME
Ord442423	2002	7	ACME
Ord524232	2002	16	ACME
Ord645363	2002	5	Direct Parts

Here is the expected output.

Customer ID	Vendor
1001	Direct Parts
2002	ACME

Previous Year's Sales

Write an SQL statement that shows the current year's sales, along with the previous year's sales, and the sales from two years ago.

Year	Amount
2018	\$352,645
2017	\$165,565
2017	\$254,654
2016	\$159,521
2016	\$251,696
2016	\$111,894

Here is the expected output.

2018	2017	2016
\$352,645	\$420,219	\$411,217

Delete the Duplicates

Write an SQL statement that deletes the duplicate data.

```
CREATE TABLE #SampleData
(
IntegerValue INTEGER
);
GO

INSERT INTO #SampleData VALUES
(1),
(1),
(2),
(3),
(3),
(4);
GO
```

Fill the Gaps

The answer to this problem is often referred to as a "data smear" or a "flash fill".

Write an SQL statement to fill in the missing gaps.

Row Number	Workflow	Status
1	Alpha	Pass
2		Fail
3		Fail
4		Fail
5	Bravo	Pass
6		Fail
7		Fail
8		Pass
9		Pass
10	Charlie	Fail
11		Fail
12		Fail

Here is the expected output.

Row Number	Workflow	Status
1	Alpha	Pass
2	Alpha	Fail
3	Alpha	Fail
4	Alpha	Fail
5	Bravo	Pass
6	Bravo	Fail
7	Bravo	Fail
8	Bravo	Pass
9	Bravo	Pass
10	Charlie	Fail
11	Charlie	Fail
12	Charlie	Fail

Count the Groupings

Write an SQL statement that counts the consecutive values in the Status field.

Step Number	Status
1	Passed
2	Passed
3	Passed
4	Passed
5	Failed
6	Failed
7	Failed
8	Failed
9	Failed
10	Passed
11	Passed
12	Passed

Here is the expected outcome.

Order	Status	Consecutive Counts
1	Passed	4
2	Failed	5
3	Passed	3

Select Star

Your developers have many bad practices; the worst of them being they routinely deploy procedures that do not explicitly define which fields to return in their SELECT clause.

Modify the following table in such a way that the statement [SELECT * FROM Products] will return an error when executed.

```
CREATE TABLE #Products
(
ProductID INTEGER,
ProductName VARCHAR(MAX)
);
```

Second Highest

How many different SQL statements can you write that will return the second highest integer?

```
CREATE TABLE #SampleData
(
IntegerValue INTEGER
);
GO
INSERT INTO #SampleData VALUES
(3759),(3760),(3761),(3762),(3763);
GO
```

First and Last

Write an SQL statement that determines the most and least experienced Spaceman ID by their job description.

Spaceman ID	Job Description	Mission Count
1001	Astrogator	6
2002	Astrogator	12
3003	Astrogator	17
4004	Geologist	21
5005	Geologist	9
6006	Geologist	8
7007	Technician	13
8008	Technician	2
9009	Technician	7

Here is the expected output.

Job Description	Most Experienced	Least Experienced
Astrogator	3003	1001
Geologist	4004	6006
Technician	7007	8008

Deadlines

Write an SQL statement that determines if an order will be fulfilled by the requested delivery date. Is there a better SQL construct to use then the MAX function?

Orders

Order ID	Product	Delivery Date (Days)
Ord893456	Widget	7
Ord923654	Gizmo	3
Ord187239	Doodad	9

Manufacturing Time

Part	Product	Days to Manufacture
AA-111	Widget	7
BB-222	Widget	2
CC-333	Widget	3
DD-444	Widget	1
AA-111	Gizmo	7
BB-222	Gizmo	2
AA-111	Doodad	7
DD-444	Doodad	1

Here is the expected output.

Order ID	Product
Ord893456	Widget
Ord187239	Doodad

Order ID Ord893456 and Ord187239 will be in the output as these orders have a promised delivery date that is equal to or greater than the days to manufacture.

Specific Exclusion

Write an SQL statement that returns all rows except where the Customer ID is 1001 and the Amount is \$50.

Customer ID	Order ID	Amount
1001	Ord143933	\$25
1001	Ord789765	\$50
2002	Ord345434	\$65
3003	Ord785633	\$50

Here is the expected output.

Customer ID	Order ID	Amount
1001	Ord143933	\$25
2002	Ord345434	\$65
3003	Ord785633	\$50

International vs Domestic Sales

You work in a sales office that sells widgets both domestically and internationally.

Write an SQL statement that shows all sales representatives who either had a domestic sale or an international sale, but not both.

Sales Rep ID	Invoice ID	Amount	Sales Type
1001	Inv345756	\$13,454	International
2002	Inv546744	\$3,434	International
4004	Inv234745	\$54,645	International
5005	Inv895745	\$234,345	International
7007	Inv006321	\$776	International
1001	Inv734534	\$4,564	Domestic
2002	Inv600213	\$34,534	Domestic
3003	Inv757853	\$345	Domestic
6006	Inv198632	\$6,543	Domestic
8008	Inv977654	\$67	Domestic

Sales Rep ID 3003, 4004, 5005 and 6006 would appear in the result set as they had either an international sale or a domestic sale, but not both.

Traveling Salesman

Here is a well-known problem that is called the Traveling Salesman among programmers.

Write an SQL statement that shows all the possible routes from Austin to Des Moines. Which route is the most expensive? Which route is the least expensive? Make any necessary assumptions to complete the puzzle.

Departure City	Arrival City	Cost
Austin	Dallas	\$100
Dallas	Memphis	\$200
Memphis	Des Moines	\$300
Dallas	Des Moines	\$400

Group Criteria Keys

Write an SQL statement that provides a key based upon the distinct combination of distributor, facility, and zone.

Order ID	Distributor	Facility	Zone	Amount
Ord156795	ACME	123	ABC	\$100
Ord826109	ACME	123	ABC	\$75
Ord342876	Direct Parts	789	XYZ	\$150
Ord994981	Direct Parts	789	XYZ	\$125

Here is the expected output.

Criteria ID	Order ID	Distributor	Facility	Zone	Amount
1	Ord156795	ACME	123	ABC	\$100
1	Ord826109	ACME	123	ABC	\$75
2	Ord342876	Direct Parts	789	XYZ	\$150
2	Ord994981	Direct Parts	789	XYZ	\$125

Reporting Elements

You must provide a report of all distributors and their sales by region. If a distributor did not have any sales for a region, provide a zero-dollar amount for that day. Assume there is at least one sale for each region.

Region	Distributor	Sales	
North	ACE	10	
South	ACE	67	
East	ACE	54	
North	Direct Parts	8	
South	Direct Parts	7	
West	Direct Parts	12	
North	ACME	65	
South	ACME	9	
East	ACME	1	
West	ACME	7	

Here is the expected output.

Region	Distributor	Sales
North	ACE	10
South	ACE	67
East	ACE	54
West	ACE	0
North	ACME	65
South	ACME	9
East	ACME	1
West	ACME	7
North	Direct Parts	8
South	Direct Parts	7
East	Direct Parts	0
West	Direct Parts	12

Answers to the puzzles are located in the following GitHub repository.

https://github.com/smpetersgithub/AdvancedSQLPuzzles/tree/main/Advanced%20SQL%20Puzzles

Prime Numbers

Write an SQL statement to determine which of the below numbers are prime numbers.

```
CREATE TABLE #SampleData
(
IntegerValue INTEGER
);
GO

INSERT INTO #SampleData VALUES
(1),(2),(3),(4),(5),(6),(7),(8),(9),(10);
GO
```

Sort Order

Write an SQL statement that sorts the following values into the expected output. Can you find the most elegant solution?

City
Atlanta
Baltimore
Chicago
Denver

Here is the expected output.

City
Baltimore
Denver
Atlanta
Chicago