Chapter 45: Window Functions

Section 45.1: Setting up a flag if other rows have a common property

Let's say I have this data:

Table items

id	name	tag
1	example	unique_tag
2	foo	simple
42	bar	simple
3	baz	hello
51	quux	world

I'd like to get all those lines and know if a tag is used by other lines

```
SELECT id, name, tag, COUNT(*) OVER (PARTITION BY tag) > 1 AS flag FROM items
```

The result will be:

id	name	tag	flag
1	example	unique_tag	false
2	foo	simple	true
42	bar	simple	true
3	baz	hello	false
51	quux	world	false

In case your database doesn't have OVER and PARTITION you can use this to produce the same result:

```
SELECT id, name, tag, (SELECT COUNT(tag) FROM items B WHERE tag = A.tag) > 1 AS flag FROM items A
```

Section 45.2: Finding "out-of-sequence" records using the LAG() function

Given these sample data:

ID	STATUS	STATUS_TIME	STATUS_BY
1	ONE	2016-09-28-19.47.52.501398	USER_1
3	ONE	2016-09-28-19.47.52.501511	USER_2
1	THREE	2016-09-28-19.47.52.501517	USER_3
3	TWO	2016-09-28-19.47.52.501521	USER_2
3	THREE	2016-09-28-19.47.52.501524	USER_4

Items identified by `ID` values must move from `STATUS` 'ONE' to 'TWO' to 'THREE' in sequence, without skipping statuses. The problem is to find users (`STATUS_BY`) values who violate the rule and move from 'ONE' immediately to 'THREE'.

The `LAG()` analytical function helps to solve the problem by returning for each row the value in the preceding row:

```
SELECT * FROM (
 SELECT
  t.*,
  LAG(status) OVER (PARTITION BY id ORDER BY status_time) AS prev_status
  FROM test t
) t1 WHERE status = 'THREE' AND prev_status != 'TWO'
```

In case your database doesn't have LAG() you can use this to produce the same result:

```
SELECT A.id, A.status, B.status as prev_status, A.status_time, B.status_time as prev_status_time
FROM Data A, Data B
WHERE A.id = B.id
AND   B.status_time = (SELECT MAX(status_time) FROM Data where status_time < A.status_time and id =
A.id)
AND   A.status = 'THREE' AND NOT B.status = 'TWO'
```

Section 45.3: Getting a running total

Given this data:

date	amount
2016-03-12	200
2016-03-11	-50
2016-03-14	100
2016-03-15	100
2016-03-10	-250

```
SELECT date, amount, SUM(amount) OVER (ORDER BY date ASC) AS running
FROM operations
ORDER BY date ASC
```

will give you

date	amount	running
2016-03-10	-250	-250
2016-03-11	-50	-300
2016-03-12	200	-100
2016-03-14	100	0
2016-03-15	100	-100

Section 45.4: Adding the total rows selected to every row

```
SELECT your_columns, COUNT(*) OVER() as Ttl_Rows FROM your_data_set
```

id	name	Ttl_Rows
1	example	5
2	foo	5
3	bar	5
4	baz	5
5	quux	5

Instead of using two queries to get a count then the line, you can use an aggregate as a window function and use the full result set as the window.

This can be used as a base for further calculation without the complexity of extra self joins.

Section 45.5: Getting the N most recent rows over multiple grouping

Given this data

User_ID	Completion_Date
1	2016-07-20
1	2016-07-21
2	2016-07-20
2	2016-07-21
2	2016-07-22

```
;with CTE as
(SELECT *,
        ROW_NUMBER() OVER (PARTITION BY User_ID
                           ORDER BY Completion_Date DESC) Row_Num
FROM    Data)
SELECT * FORM CTE WHERE Row_Num <= n
```

Using n=1, you'll get the one most recent row per `user_id`:

User_ID	Completion_Date	Row_Num
1	2016-07-21	1
2	2016-07-22	1